Numerical approximations for stochastic differential equations

James Matthew Foster
Worcester College
University of Oxford

A thesis submitted for the degree of
Doctor of Philosophy

2020

Acknowledgements

I would like to express my gratitude and appreciation to my supervisors Prof. Terry Lyons and Prof. Harald Oberhauser for their guidance and invaluable support. Throughout our mathematical discussions, they have shown tremendous patience and being their student has been a privilege.

I am especially grateful to my examiners, Prof. Christoph Reisinger and Prof. Christian Bayer, for their assessment of this thesis and constructive feedback. In particular, I appreciate their insights into the applications of this research for Quasi-Monte Carlo methods. I am also very grateful to Prof. Nick Trefethen, Prof. Zhongmin Qian and Dr. Danyu Yang for all of their comments regarding my transfer and confirmation theses. In addition, I would like to thank Prof. Trefethen for making a beautiful Matlab demo of the "Brownian polynomials" using the Chebfun software package.

I am indebted to my friends and colleagues who have made my time at Oxford so enjoyable. I wish to thank my officemates Dr. Sina Nejad, Dr. Imanol Pérez Arribas and Cristopher Salvi. In many ways, Sina was a "third supervisor" who provided fantastic help with rough path theory and I owe Cris for getting me involved with neural differential equations.

Along with Cris, I am grateful to Patrick Kidger and James Morrill for allowing me to put some rough analysis in their machine learning papers. In addition, Patrick's work on the log-ODE method for general SDEs provided the necessary motivation that led to the results in section 7.4.

I would also like to acknowledge Dr. Vlad Margarint, who introduced me to the Schramm-Loewner evolution (SLE) and the challenges regarding its simulation. It has been very satisfying to collaborate on this problem (which became a refreshing side project during the writing of this thesis).

Finally, I am deeply grateful to my family who have been wonderfully supportive during these past few years.

Abstract

In this thesis, we consider problems related to the numerical simulation of stochastic differential equations (SDEs). In particular, we are interested in methods that use both increments and areas of the Brownian path. For example, time integrals of Brownian motion carry useful information whilst being normally distributed and thus straightforward to generate. We will make precise the "path information" contained by such integrals and present a polynomial approximation theorem for Brownian motion. Since the Gaussian coefficients given by this expansion are independent, we can express Brownian motion as a polynomial with additional noise. We shall then use this simple observation to develop high order methods.

The majority of these methods approximate SDEs through the use of ODEs, which can be accurately discretized using state-of-the-art solvers. To numerically demonstrate these approaches, we will simulate four SDEs: Inhomogeneous Geometric Brownian Motion, Cox-Ingersoll-Ross model, underdamped Langevin equation and the Lévy area of Brownian motion. The code for these examples can be found at github.com/james-m-foster.

We also study problems related to variable step size methods for SDEs. Most notably, we investigate which numerical methods and variable step size strategies can ensure pathwise convergence to the true SDE solution. Our analysis is based on rough path theory and requires methods to be locally close to an ODE driven by a piecewise "Brownian polynomial". This result can therefore be applied to the Heun and midpoint methods. In order to prove convergence, we will require variable step size strategies to produce a nested sequence of partitions with mesh size tending to zero.

Contents

List of Figures

List of Tables

Chapter 1

A brief outline of the thesis

The main result of this thesis is a polynomial representation of Brownian motion, which is presented in Chapter 4 (along with some applications to SDE simulation).

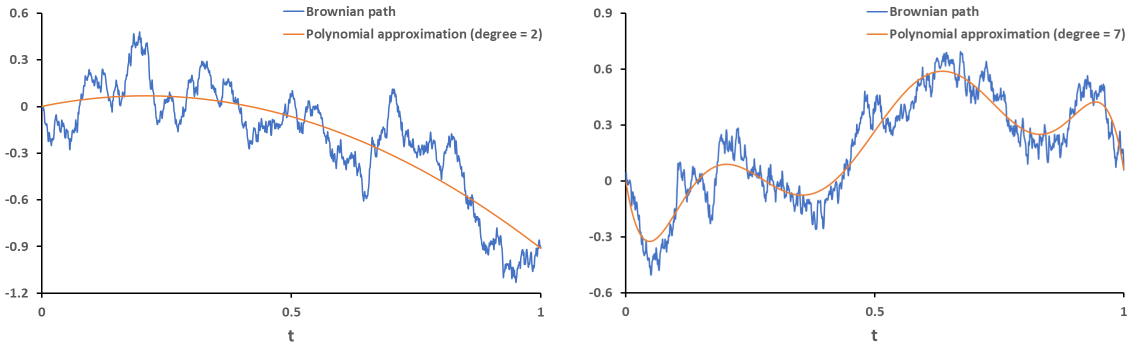

Figure 1.1: Sample paths of Brownian motion with their polynomial approximations.

In particular, this polynomial approximation (Theorem 4.1.12) can be applied to develop estimators for non-Gaussian iterated integrals of Brownian motion such as

$$\int_s^t \int_s^u \circ dW_v^{(i)} \circ dW_u^{(j)} \quad \text{and} \quad \int_s^t \int_s^u \int_s^v \circ dW_r^{(i)} \circ dW_v^{(j)} du. \tag{1.0.1}$$

These estimators are obtained by taking the moments of (1.0.1) conditional on certain Gaussian iterated integrals of Brownian motion (which can be generated exactly):

$$\int_s^t \circ dW_u \quad \text{and} \quad \int_s^t \int_s^u \circ dW_v \, du. \tag{1.0.2}$$

These integrals appear naturally in stochastic Taylor expansions and are of great importance to the design and analysis of high order numerical methods for SDEs. The study of Brownian iterated integrals will be a long-running focus of the thesis and we present results concerning their conditional moments in Chapters 4, 5 and 7. (More specifically, these are Theorems 4.2.8, 4.4.1, 5.1.2, 5.1.3, 7.1.1, 7.2.1 and 7.2.2).

The polynomial approximation of Brownian motion is introduced in Chapter 4 and its immediate applications to iterated integrals are given in sections 4.1 and 4.2. In Chapters 4 & 5, we will develop numerical methods using the integral estimators:

- **A log-ODE method for SDEs driven by a single Brownian motion**
  In section 4.3, we shall present a new version of the asymptotically efficient log-ODE (or Castell-Gaines) method that utilizes a certain integral estimator. We discuss the rough path formulation of the log-ODE method in section 3.4.

- **A stochastic Runge-Kutta method for SDEs with additive noise**
  In section 4.4, we will derive and solve order conditions for a type of explicit three-stage Runge-Kutta method that approximates SDEs with additive noise.

- **ODE methods based on "high order" piecewise linear paths**
  In sections 5.1 and 5.2, we use the integral estimators to design piecewise linear paths that accurately approximate Brownian motion (with its iterated integrals).

Each research chapter (i.e. 4–7) has a section with numerical experiments. These are:

- **Inhomogeneous Geometric Brownian Motion** (section 4.5)
  $dy_t = a(b - y_t)\,dt + \sigma y_t\,dW_t$.

- **The underdamped Langevin equation** (section 5.3)
  $dQ_t = P_t\,dt$,
  $dP_t = -\nabla U(Q_t)\,dt - \nu P_t\,dt + \sqrt{\frac{2\nu}{\beta}}\,dW_t$.

- **The Cox-Ingersoll-Ross model** (section 6.3)
  $dy_t = a(b - y_t)\,dt + \sigma\sqrt{y_t}\,dW_t$.

- **The Lévy area between independent Brownian motions** (section 7.3)
  $dx_t = W_t^{(i)}\,dW_t^{(j)}$.

Each numerical method for these SDEs is based on the theory developed in its chapter.

In Chapter 6, we consider approximations for SDEs that use variable step sizes. Since the methods proposed in Chapter 5 require generating a third random variable along with (1.0.2), we extend Lévy's construction to certain triples in section 6.1. This allows us to reduce step sizes conditional on these (Brownian) random variables. In section 6.2, we shall apply rough path theory (which is discussed in Chapter 3) to establish sufficient conditions that ensure a variable step size method converges. This result applies to a large class of numerical methods, including Milstein's method, but does not apply to the Euler-Maruyama method (which may not converge [39]). Chapters 2 and 3 are introductory and give the necessary background for the thesis.

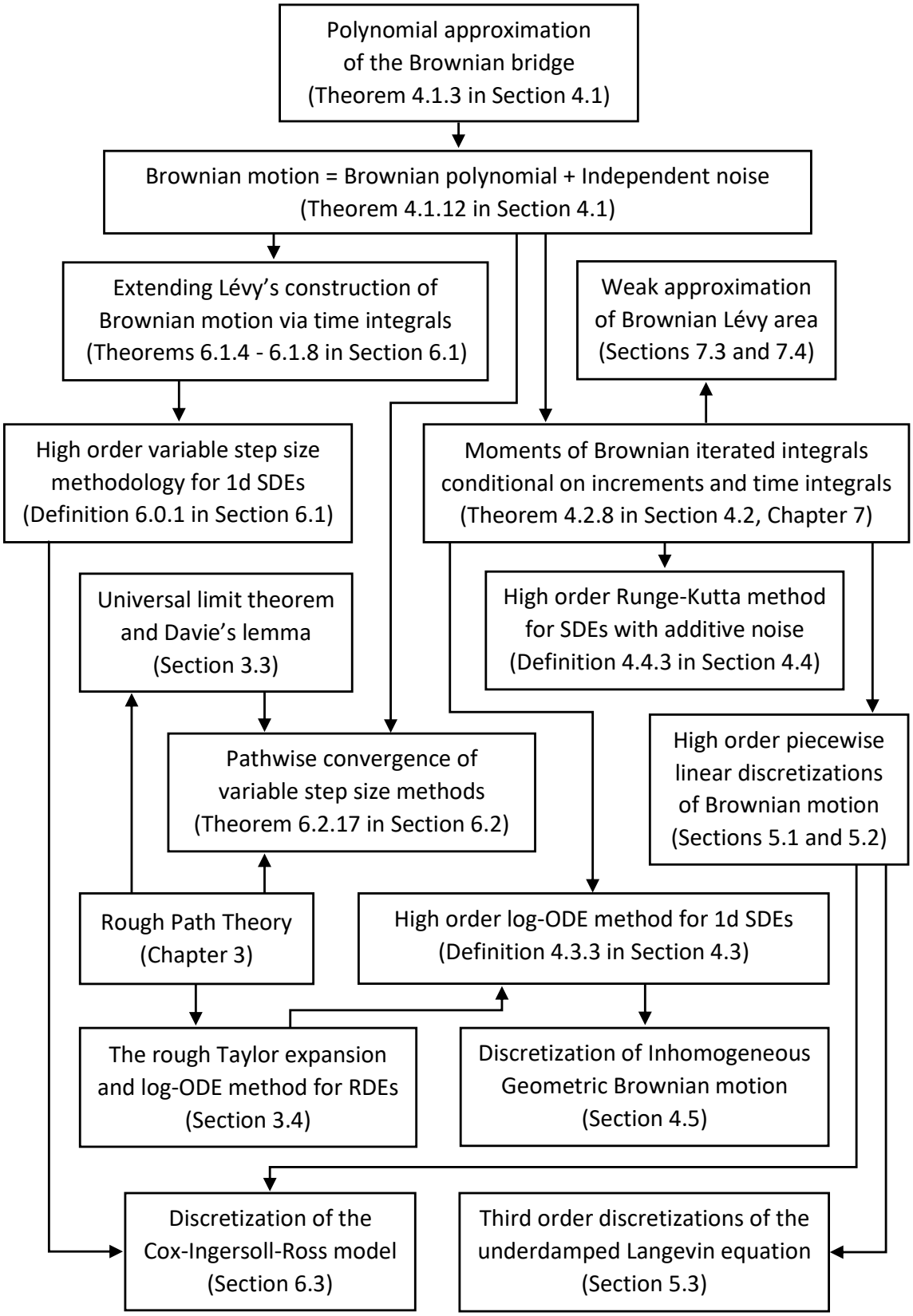

Figure 1.2: Diagram showing how the various topics within the thesis are connected.

Chapter 2

Introduction to numerical methods for SDEs

Numerical methods for stochastic differential equations (SDEs) have be studied over the past few decades and thus an extensive literature has developed on the subject. However, for the purposes of this thesis, it will be enough to introduce some of the key concepts underlying SDE numerics before applying them in subsequent chapters. This introductory chapter is primarily based on the book of Kloeden & Platen [62] (currently the highest cited text on the subject) and a review paper of Higham [54].

Many real-world phenomena (such as particle systems [76], financial markets [11] and population dynamics [55]) can be modelled using the following class of Itô SDEs:

$$\begin{aligned} dy_t &= f_0(y_t)\,dt + \sum_{i=1}^{d} f_i(y_t)\,dW_t^{(i)}, \\ y_0 &= \xi, \end{aligned} \tag{2.0.1}$$

where the initial condition ξ and solution process $\{y_t\}_{t\geq 0}$ takes their values in $\mathbb{R}^e$, each $f_i : \mathbb{R}^e \to \mathbb{R}^e$ is a vector field on $\mathbb{R}^e$ and W is a d-dimensional Brownian motion. In practice, one is often interested in computing some averaged quantity from (2.0.1):

$$\bar{\varphi} := \mathbb{E}\big[\varphi(y) : y_0 = \xi\big], \tag{2.0.2}$$

where φ is a functional on path space, e.g. $\varphi(y)$ could be a function of the point y_T. In certain situations, the expected functional (2.0.2) can be approximated by solving a parabolic PDE (such as via the Feynman-Kac formula). However, standard finite difference methods for these PDEs will suffer from the "curse of dimensionality". Furthermore, there are many problems (such as pricing exotic financial derivatives) where φ depends on the entire trajectory of y and a PDE formulation is not possible. Hence, in this thesis we shall focus on computing (2.0.2) by Monte Carlo simulation.

Therefore, the general strategy for approximating (2.0.2) that we will consider is

Definition 2.0.1 (Monte Carlo simulation of SDEs)**.** For $N \geq 1$, define the estimator:

$$\bar{\varphi}^N := \sum_{n=1}^{N} w_i^N \varphi(Y^{n,N}), \tag{2.0.3}$$

where $\{w_n^N\}_{1 \leq n \leq N}$ are a collection of weights (typically $w_n^N = \frac{1}{N}$) and each $Y^{n,N}$ is a numerical solution of (2.0.1), usually obtained from a discretized sample path of W.

Remark 2.0.2. When the weights are uniform and the numerical solutions $Y^{n,N}$ are independent and identically distributed, the Monte Carlo estimator (2.0.3) becomes

$$\bar{\varphi}^N = \frac{1}{N} \sum_{n=1}^{N} \varphi(Y^n). \tag{2.0.4}$$

So if $\varphi(Y^n)$ has a finite second moment, it follows from the central limit theorem that

$$\sqrt{N}\left(\bar{\varphi}^N - \mathbb{E}\big[\varphi(Y)\big]\right) \xrightarrow{d} \mathcal{N}\left(0, \mathrm{Var}\big(\varphi(Y)\big)\right).$$

In other words, the estimator $\bar{\varphi}^N$ converges to its expectation with a rate of $O\big(\frac{1}{\sqrt{N}}\big)$. It should be noted that $\mathbb{E}\big[\varphi(Y)\big]$ will differ from the true expectation $\bar{\varphi} = \mathbb{E}\big[\varphi(y)\big]$ due to the (weak) approximation error of the numerical solution. For standard methods, such as the Euler-Maruyama method, this error is $O(h)$ where h denotes the step size.

It is reasonable to ask whether there are smarter ways of designing the Monte Carlo estimator (2.0.3) than the standard approach described by the above remark. One popular technique is called Multilevel Monte Carlo (introduced by Giles [43]) which is a variance reduction strategy inspired by control variates and takes the form

$$\bar{\varphi}^{N_0,\cdots,N_L} := \sum_{\ell=0}^{L} \widehat{\varphi}^{N_\ell}, \quad \text{with} \quad \widehat{\varphi}^{N_\ell} := \frac{1}{N_\ell} \sum_{n=1}^{N_\ell} \big(\varphi\big(Y_\ell^{n,\ell}\big) - \varphi\big(Y_{\ell-1}^{n,\ell}\big)\big), \tag{2.0.5}$$

where $\varphi\big(Y_{-1}\big) \equiv 0$ and each pair of numerical solutions $\big(Y_\ell^{n,\ell}, Y_{\ell-1}^{n,\ell}\big)$ can be coupled. The level ℓ often indicates the step size used to compute each numerical solution Y_ℓ (e.g. Y_0 is obtained cheaply with large steps whilst each $Y_L^{n,L}$ is a fine discretization). By computing $\varphi\big(Y_\ell^{n,\ell}\big)$ and $\varphi\big(Y_{\ell-1}^{n,\ell}\big)$ from the same sample paths of Brownian motion it is possible to reduce the variance of the estimator $\frac{1}{N_\ell}\sum_{n=1}^{N_\ell}\big(\varphi\big(Y_\ell^{n,\ell}\big) - \varphi\big(Y_{\ell-1}^{n,\ell}\big)\big)$. Thus, by choosing each N_ℓ appropriately, the multilevel Monte Carlo estimator (2.0.5) will exhibit a significantly lower variance than standard estimator given by (2.0.4).

Theorem 2.0.3 (Complexity of Multilevel Monte Carlo, Theorem 1 in [44])**.** *Suppose that the independent estimators $\widehat{\varphi}^{N_\ell}$ based on N_ℓ samples given by (2.0.5) each costs C_ℓ and there exists positive constants $\alpha, \beta, \gamma, c_1, c_2, c_3$ such that $\alpha \geq \frac{1}{2}\min(\beta, \gamma)$ and*

- $\left|\mathbb{E}\left[\varphi\left(Y_\ell^\ell\right)\right] - \mathbb{E}[\varphi(y)]\right| \leq c_1 2^{-\alpha\ell}$,
- $\mathrm{Var}(\widehat{\varphi}^{N_\ell}) \leq c_2 N_\ell^{-1} 2^{-\beta\ell}$,
- $\mathbb{E}[C_\ell] \leq c_3 2^{\gamma\ell}$.

Then there exists a positive constant c_4 such that for any $\varepsilon < 1$ there exist $L \geq 0$ and $\{N_\ell\}_{0\leq\ell\leq L}$ for which the multilevel estimator $\bar{\varphi}^{N_0,\cdots,N_L}$ given by (2.0.5) obtains a mean squared error $\mathbb{E}\left[\left(\bar{\varphi}^{N_0,\cdots,N_L} - \mathbb{E}[\varphi(y)]\right)^2\right] < \varepsilon^2$ with expected computational cost

$$C \leq \begin{cases} c_4\varepsilon^{-2}, & \beta > \gamma, \\ c_4\varepsilon^{-2}(\log\varepsilon)^2, & \beta = \gamma, \\ c_4\varepsilon^{-2-\frac{(\gamma-\beta)}{\alpha}}, & \beta < \gamma. \end{cases} \tag{2.0.6}$$

Remark 2.0.4. Suppose that we discretized the SDE (2.0.1) using a numerical method with a "weak" convergence rate (see definition 2.2.1) of $O(h^r)$ where h is the step size. Then in order to achieve a mean squared accuracy of ε^2, the standard Monte Carlo estimator (2.0.4) would require $O(\varepsilon^{-2})$ simulations (due to the Monte Carlo error) and each simulation would require $O(\varepsilon^{-\frac{1}{r}})$ steps (due to the discretization error). Therefore the standard Monte Carlo estimator (2.0.4) has a complexity of $O(h^{-2-\frac{1}{r}})$. On the other hand, when $\beta > \gamma$, the multilevel estimator (2.0.5) achieves the optimal Monte Carlo complexity of $O(\varepsilon^{-2})$. That is, the average simulation has an $O(1)$ cost.

In the multilevel Monte Carlo paradigm, three different rates appear: α, β and γ. The rate α may be viewed simply as the "weak" convergence rate of the SDE solver, whereas the rate β can be related to the "strong" convergence of the numerical solver (these two types of convergence for SDE approximations are discussed in section 2.2). For example, when the functional φ is assumed to be Lipschitz continuous, we have

$$\mathrm{Var}(\widehat{\varphi}^{N_\ell}) \leq N_\ell^{-1}\mathbb{E}\left[\left\|\varphi\left(Y_\ell^\ell\right) - \varphi\left(Y_{\ell-1}^\ell\right)\right\|^2\right] \leq N_\ell^{-1}\|\varphi\|^2_{\text{Lipschitz}}\,\mathbb{E}\left[\left\|Y_\ell^\ell - Y_{\ell-1}^\ell\right\|^2\right],$$

where the mean squared error $\mathbb{E}\left[\left\|Y_\ell^\ell - Y_{\ell-1}^\ell\right\|^2\right]$ depends on the strong convergence of the method. The rate γ relates to how the computational cost increases with level.

Whilst the design of efficient Monte Carlo estimators is tremendously important for SDE simulation, this thesis shall focus on the methodologies for computing the numerical solutions Y accurately from discretized sample paths of Brownian motion. That said, the majority of numerical methods that are proposed by this thesis will have high convergence rates and should be compatible with the multilevel framework. Incorporating the ideas of the thesis into more sophisticated Monte Carlo estimators would be a step towards tackling real-world problems and thus is a goal of the author.

2.1 Some examples of numerical methods

In this section, we will present a variety of well-known methods for discretizing (2.0.1). For simplicity, these methods will be defined using a uniform partition $\triangle_N$ of $[0, T]$.

Definition 2.1.1 (**The Euler-Maruyama method**)**.** For $N \geq 1$, we can construct a numerical solution $\{Y_k\}_{0\leq k\leq N}$ of (2.0.1) by setting $Y_0 := \xi$ and for $k \in [0\,..\,N-1]$ defining Y_{k+1} as

$$Y_{k+1} := Y_k + f_0(Y_k)h + \sum_{i=1}^{d} f_i(Y_k)\left(W^{(i)}_{t_{k+1}} - W^{(i)}_{t_k}\right), \tag{2.1.1}$$

where $h = \frac{T}{N}$ and $t_k = \frac{kT}{N}$.

Definition 2.1.2 (**The Milstein method**)**.** For a fixed $N \geq 1$, we can construct a numerical solution $\{Y_k\}_{0\leq k\leq N}$ of (2.0.1) by setting $Y_0 := \xi$ and for $k \in [0\,..\,N-1]$ defining Y_{k+1} as

$$\begin{aligned} Y_{k+1} := Y_k + f_0(Y_k)h + \sum_{i=1}^{d} f_i(Y_k)\left(W^{(i)}_{t_{k+1}} - W^{(i)}_{t_k}\right) \qquad (2.1.2)\\ + \sum_{i,j=1}^{d} f_i'(Y_k)\, f_j(Y_k)\left(\frac{1}{2}\left(W^{(i)}_{t_{k+1}} - W^{(i)}_{t_k}\right)\left(W^{(j)}_{t_{k+1}} - W^{(j)}_{t_k}\right) + A^{(i,j)}_{t_k,t_{k+1}} - \frac{1}{2}h\delta_{ij}\right), \end{aligned}$$

where the antisymmetric matrix A is known as the **Lévy area** of Brownian motion:

$$A^{(i,j)}_{t_k,t_{k+1}} := \frac{1}{2}\left(\int_{t_k}^{t_{k+1}} \left(W^{(i)}_u - W^{(i)}_{t_k}\right) dW^{(j)}_u - \int_{t_k}^{t_{k+1}} \left(W^{(j)}_u - W^{(j)}_{t_k}\right) dW^{(i)}_u\right),$$

for $i, j \in \{1, \cdots, d\}$, $h = \frac{T}{N}$, $t_k = \frac{kT}{N}$ and δ is the Kronecker delta. So when $d = 1$, the Milstein method becomes

$$Y_{k+1} := Y_k + f_0(Y_k)h + f_1(Y_k)\, W_{t_k,t_{k+1}} + \frac{1}{2} f_1'(Y_k)\, f_1(Y_k)\left(W^2_{t_k,t_{k+1}} - h\right), \tag{2.1.3}$$

where $W_{t_k,t_{k+1}} = W_{t_{k+1}} - W_{t_k}$ is the increment of the Brownian motion over $[t_k, t_{k+1}]$.

The methods (2.1.1) and (2.1.3) are both straightforward to implement since the Brownian motion W has independents increments and coordinate processes $\{W^{(i)}\}$. So to propagate these numerical solutions we generate the Gaussian random variables

$$W^{(i)}_{t_k,t_{k+1}} \sim \mathcal{N}(0, h), \tag{2.1.4}$$

independently for $i \in \{1, \cdots, d\}$ and $k \geq 0$.

If the Brownian motion is multidimensional, the Milstien method will in general[1] require one to generate the Lévy area $A_{t_k,t_{k+1}}$ in addition to the increment $W_{t_k,t_{k+1}}$. This is known to be difficult problem [30] and algorithms for exact simulation have only been proposed when $d = 2$ [38]. In Chapter 7, we will discuss Lévy area further.

As well as requiring a Jacobian matrix $f_i' : \mathbb{R}^e \to L\big(\mathbb{R}^e, \mathbb{R}^e\big)$ for each $i \in \{1, \cdots, d\}$, the Milstein method has a higher strong convergence rate than Euler-Maruyama and, for example, is known to outperform Euler-Maruyama for SDEs with scalar noise [54]. However in the general setting, the Euler-Maruyama and Milstein methods will both achieve the same orders of convergence when the area $A_{s,t}$ is omitted from (2.1.2). The convergence rates for both of these methods will be discussed in the next section.

In Chapter 6, we shall analyse the convergence of variable step size methods using rough path theory. Whilst the Euler-Maruyama method generally will not converge to the true SDE solution when variable steps are used ([39] gives an example of this), we will show that the "no-area" Milstein method (that is (2.1.2) without $A_{s,t}$) does converge under certain assumptions on the vector fields $\{f_i\}$ and variable step sizes.

However, since this analysis relies on certain key theorems from rough path theory, it will be necessary to write the Itô SDE (2.0.1) in the following Stratonovich form:

$$\begin{aligned} dy_t &= \tilde{f}_0\left(y_t\right) dt + \sum_{i=1}^{d} f_i\left(y_t\right) \circ dW_t^{(i)}, \\ y_0 &= \xi, \end{aligned} \tag{2.1.5}$$

where $\tilde{f}_0(y) := f_0(y) - \frac{1}{2}\sum_{i=1}^{d} f_i'(y) f_i(y)$ is the drift with the Itô-Stratonovich correction.

Since it is easier to develop high order numerical methods for Stratonovich SDEs, the approximations presented in this thesis will directly apply to the SDE (2.1.5). Below is a list of some low order numerical methods for the above Stratonovich SDE:

Definition 2.1.3 (**The no-area Milstein method for Stratonovich SDEs**).

$$\begin{aligned} Y_{k+1} := Y_k + \tilde{f}_0\left(Y_k\right)h + \sum_{i=1}^{d} f_i\left(Y_k\right)\Big(W_{t_{k+1}}^{(i)} - W_{t_k}^{(i)}\Big) \\ + \frac{1}{2}\sum_{i,j=1}^{d} f_i'\left(Y_k\right) f_j\left(Y_k\right)\Big(W_{t_{k+1}}^{(i)} - W_{t_k}^{(i)}\Big)\Big(W_{t_{k+1}}^{(j)} - W_{t_k}^{(j)}\Big). \end{aligned} \tag{2.1.6}$$

[1] If the vector fields $\{f_i\}_{1\le i\le d}$ commute according to (2.2.3) then the $A_{t_k,t_{k+1}}$ terms cancel out.

Definition 2.1.4 (**The Euler-Heun method**).

$$\widetilde{Y_k} := Y_k + \tilde{f}_0(Y_k)h + \sum_{i=1}^{d} f_i(Y_k)\left(W^{(i)}_{t_{k+1}} - W^{(i)}_{t_k}\right),$$
$$Y_{k+1} := Y_k + \tilde{f}_0\big(Y_k\big)h + \frac{1}{2}\sum_{i=1}^{d}\Big(f_i\big(Y_k\big) + f_i\big(\widetilde{Y_k}\big)\Big)\Big(W^{(i)}_{t_{k+1}} - W^{(i)}_{t_k}\Big).$$

Definition 2.1.5 (**Heun's method (or the explicit trapezoidal rule)**).

$$Y_{k+1} := Y_k + \frac{1}{2}\Big(\tilde{f}_0\big(Y_k\big) + \tilde{f}_0\big(\widetilde{Y_k}\big)\Big)h + \frac{1}{2}\sum_{i=1}^{d}\Big(f_i\big(Y_k\big) + f_i\big(\widetilde{Y_k}\big)\Big)\Big(W^{(i)}_{t_{k+1}} - W^{(i)}_{t_k}\Big).$$

Definition 2.1.6 (**The explicit midpoint method**).

$$Y_{k+1} := Y_k + \tilde{f}_0\Big(\frac{1}{2}\Big(Y_k + \widetilde{Y_k}\Big)\Big)h + \frac{1}{2}\sum_{i=1}^{d} f_i\Big(\frac{1}{2}\Big(Y_k + \widetilde{Y_k}\Big)\Big)\Big(W^{(i)}_{t_{k+1}} - W^{(i)}_{t_k}\Big).$$

Except possibly Euler-Heun, these methods can be obtained by discretizing an ODE goverened by the same vector fields but driven by a linear path $\widehat{W}$ on $[t_k, t_{k+1}]$ with a second order ODE method. Since these methods match the $O(\sqrt{h})$ and $O(h)$ terms in the Taylor expansion of this ODE, they converge with variable step sizes (Theorem 6.2.17). Finally, we shall present two high order (in a weak sense) schemes.

The first scheme (originally proposed by Talay in [88]) is based on the idea that it is straightforward to generate a random variable with the same mean and covariance as Brownian Lévy area and exploits the that (iterated) Itô integrals have mean zero.

Definition 2.1.7 (**The Talay scheme**). For a fixed $N \geq 1$, we can construct a numerical solution $\{Y_k\}_{0\leq k\leq N}$ of (2.0.1) by setting $Y_0 := \xi$ and for $k \in [0\,..\,N-1]$ defining Y_{k+1} as

$$\begin{aligned} Y_{k+1} := Y_k &+ \sum_{i=0} f_i(Y_k)\,\overline{W}^{(i)}_{t_k,t_{k+1}} \\ &+ \sum_{i,j=0}^{d} f_i'(Y_k)\,f_j(Y_k)\left(\frac{1}{2}\overline{W}^{(i)}_{t_k,t_{k+1}}\,\overline{W}^{(j)}_{t_k,t_{k+1}} - \overline{A}^{(i,j)}_{t_k,t_{k+1}} - \frac{1}{2}h\delta_{ij}\right), \end{aligned}$$

where $\overline{W}^{(0)}_{t_k,t_{k+1}} = h$ and $\big\{\overline{W}^{(i)}_{t_k,t_{k+1}}\big\}_{1\leq i\leq d}$ is a collection of independent and identically distributed random variables with either $\overline{W}^{(i)}_{t_k,t_{k+1}} \sim \mathcal{N}(0,h)$ or

$$\mathbb{P}\Big(\overline{W}^{(i)}_{t_k,t_{k+1}} = \pm\sqrt{3h}\Big) = \frac{1}{6}, \quad \mathbb{P}\Big(\overline{W}^{(i)}_{t_k,t_{k+1}} = 0\Big) = \frac{2}{3},$$

and $\overline{A}_{t_k,t_{k+1}}$ is an antisymmetric matrix with $\overline{A}^{(i,0)}_{t_k,t_{k+1}} = \overline{A}^{(0,i)}_{s,t} = 0$ for $1 \leq i \leq d$ and

$$\mathbb{P}\Big(\overline{A}^{(i,j)}_{t_k,t_{k+1}} = \pm\frac{1}{2}\sqrt{h}\Big) = \frac{1}{2},$$

for $1 \leq i < j \leq d$.

For the Itô SDE (2.0.1), Talay's scheme obtains $\big|\mathbb{E}[\varphi(y_T)] - \mathbb{E}[\varphi(Y_N)]\big| = O(N^{-2})$ provided the functional φ is six times continuously differentiable and has derivatives with polynomial growth. However the "Stratonovich" version of the Talay scheme (i.e. without the Kronecker delta) will not achieve a high rate of weak convergence. In particular, it will not capture the expectation of the certain triple iterated integrals in its Taylor expansion (approximations of such integrals are discussed in section 4.2).

For Stratonovich SDEs, the Ninomiya-Victoir scheme (originally proposed in [81]) is a well-known method that can be viewed as an extension of the Strang splitting.

Definition 2.1.8 (**The Ninomiya-Victoir scheme**)**.** For a fixed $N \geq 1$, we can construct a numerical solution $\{Y_k\}_{0\leq k\leq N}$ of (2.1.5) by setting $Y_0 := \xi$ and for each $k \in [0\,..\,N-1]$ defining Y_{k+1} as

$$Y_{k+1} := \begin{cases} \widehat{Y}_{k+1}, & \text{if } \ n_k = 1, \\ \overleftarrow{Y}_{k+1}, & \text{if } \ n_k = -1, \end{cases}$$

where n_k is an independent Rademacher random variable and $\widehat{Y}_{k+1}, \overleftarrow{Y}_{k+1}$ are given by

$$\widehat{Y}_{k+1} := \exp\Big(\frac{1}{2}hf_0\Big)\exp\Big(W^{(1)}_{t_k,t_{k+1}}f_1\Big)\cdots\exp\Big(W^{(d)}_{t_k,t_{k+1}}f_d\Big)\exp\Big(\frac{1}{2}hf_0\Big)Y_k,$$

$$\overleftarrow{Y}_{k+1} := \exp\Big(\frac{1}{2}hf_0\Big)\exp\Big(W^{(d)}_{t_k,t_{k+1}}f_d\Big)\cdots\exp\Big(W^{(1)}_{t_k,t_{k+1}}f_1\Big)\exp\Big(\frac{1}{2}hf_0\Big)Y_k,$$

with $\exp\big(Cf_i\big)$ denoting the map from $x \in \mathbb{R}^e$ to the solution at $u = 1$ of the ODE,

$$\frac{dz}{du} = Cf_i(z), \quad z(0) = x.$$

Remark 2.1.9. The high order weak convergence rate of this method can be achieved if each ODE is discretized using an explicit fourth order Runge-Kutta method [41].

It was shown in [5] that if the SDE (2.1.5) has smooth bounded vector fields satisfying an ellipticity condition, then the Ninomiya-Victoir scheme converges in total variation distance with order 2. That is, for $t \in (0, T]$ there exists a constant $C_t < \infty$ such that for any given bounded measurable function $\varphi : \mathbb{R}^e \to \mathbb{R}$, we have

$$\sup_{t_k\geq t}\big|\mathbb{E}\big[\varphi\big(Y_k\big)\big] - \mathbb{E}\big[\varphi\big(y_{t_k}\big)\big]\big| \leq \frac{C_t\|\varphi\|_\infty}{N^2}, \quad \forall N \geq 1.$$

Since this thesis primarily focuses on strong approximations, we refer the reader to Chapters 14 and 15 of [62] for an overview of weak numerical methods for SDEs.

2.2 Types of convergence for numerical methods

In this section, we consider how the convergence of SDE methods can be quantified. Since we are usually interested in computing the expectation of a functional (2.0.2), a natural measure of convergence for SDE numerical methods is "weak" convergence:

Definition 2.2.1 (**Weak convergence**)**.** A numerical solution Y for (2.0.1) is said to converge in a weak sense with order β if for any polynomial p there exists $C_p > 0$ such that

$$|\,\mathbb{E}\,[\,p\,(Y_N)] - \mathbb{E}\,[\,p\,(y_T)]\,| \leq C_p\, h^\beta, \tag{2.2.1}$$

for all sufficiently small step sizes $h = \frac{T}{N}$.

On the other hand, there are situations where we would like the entire trajectory of the numerical solution Y to be accurate and a notation of "pathwise" accuracy should also help reduce the variance of the Multilevel Monte Carlo estimator (2.0.5). The most popular measure for "strong" convergence is defined using the $L^2(\mathbb{P})$ norm:

Definition 2.2.2. We define the $L^2(\mathbb{P})$ norm of an $\mathbb{R}^e$-valued random variable Z as

$$\|Z\|_{L^2(\mathbb{P})} := \sqrt{\mathbb{E}\big[\|Z\|_2^2\big]}. \tag{2.2.2}$$

We shall now define the order of strong convergence for SDE numerical methods.

Definition 2.2.3 (**Strong convergence**)**.** A numerical solution Y for (2.0.1) is said to converge in a strong sense with order α if there exists a constant $C > 0$ such that

$$\|Y_N - y_T\|_{L^2(\mathbb{P})} \leq Ch^\alpha,$$

for all sufficiently small step sizes $h = \frac{T}{N}$.

The convergence rates for the Euler-Maruyama and Milstein methods are given below:

Theorem 2.2.4. *Let y denote the true solution of (2.0.1) over the interval $[0, T]$ where $f_0, f_1, \cdots, f_d$ are bounded smooth vector fields on $\mathbb{R}^e$ with bounded derivatives. Let Y^{eul} and Y^{mil} denote the numerical solutions of (2.0.1) on $[0, T]$ obtained from the initial condition y_0 using the Euler-Maruyama and Milstein methods respectively. Then Y^{eul} converges strongly with order 0.5, Y^{mil} converges strongly with order 1.0 and both methods will converge weakly with order 1.0. Without the Lévy area term, the Milstein method converges strongly with order 0.5 and weakly with order 1.0. Furthermore, if the "noise" vector fields $\{f_i\}_{1\leq i\leq d}$ satisfy the commutativity condition*

$$f_i'(x)f_j(x) = f_j'(x)f_i(x), \qquad \forall x \in \mathbb{R}^e, \tag{2.2.3}$$

for $i, j \in \{1, \cdots, d\}$, the no-area Milstein method converges strongly with order 1.0.

As the Euler-Maruyama and no-area Milstein methods have a strong convergence rate of $O\big(\sqrt{h}\big)$ for general SDEs, it is reasonable to ask if there exist approximations defined from the increments of W that achieve higher orders of strong convergence. In 1980, Cameron and Clark [21] addressed this question with the following example:

Theorem 2.2.5 (An example of a slow best rate)**.** *Consider the following Itô SDE:*

$$\begin{aligned} dx_t &= y_t \, dW_t^{(1)}, \\ dy_t &= dW_t^{(2)}, \end{aligned} \tag{2.2.4}$$

where $W^{(1)}$ and $W^{(2)}$ are two independent Brownian motions and $x_0 = 0, y_0 = 0$. For $T > 0$, $N \geq 1$, Let P_N be the σ-algebra generated by $\left\{W^{(1)}_{\frac{kT}{N}}, W^{(2)}_{\frac{kT}{N}}\right\}_{0 \leq k \leq N}$. Then

$$\big\| x_T - \mathbb{E}\big[x_T | P_N\big] \big\|_{L^2(\mathbb{P})} = \frac{T}{2}\left(\frac{1}{N}\right)^{\frac{1}{2}}. \tag{2.2.5}$$

Corollary 2.2.6. *For a general Itô SDE, any numerical method which uses only increments of the Brownian motion cannot converge strongly with an order above* 0.5*.*

Proof. Since the Milstein method converges strongly with order 1.0 (Theorem 2.2.4), it is enough to argue that no "increment only" method can approximate the Milstein numerical solution Y^{mil} with an order above 0.5. This follows immediately from Theorem 2.2.5 as Milstein's method requires iterated integrals of Brownian motion, (or equivalently Lévy area) which requires one to solve the Itô SDE (2.2.4). □

In addition, it was shown in [30] that any approximation of the SDE (2.2.4) which is measurable with respect to N Gaussian random variables (which are expressible as linear functionals on W) cannot have a strong convergence rate better than $O\big(N^{-\frac{1}{2}}\big)$.

Since "increment only" methods for (2.0.1) usually have a strong convergence rate of $O\big(\sqrt{h}\big)$, the Euler-Maruyama method is appealing as it's computationally cheap. On the other hand, these complexity results do not give upper bounds for the weak convergence of numerical methods. For example, the Talay and Ninomiya-Victoir schemes both achieve a second order weak convergence rate (see previous section). In fact, methods can be constructed to have higher orders of weak convergence [62].

Numerical methods can also achieve convergence rates better than $O\big(\sqrt{h}\big)$ in the Wasserstein metric (which is defined by the $L^p(\mathbb{P})$ norm and a probabilistic coupling). Such an approximation was proposed by Davie in [27] and further analysed in [34]. In Chapter 7, we consider possible improvements to Davie's approximation of (2.2.4).

Fortunately, there are SDEs used in practice which satisfy the commutativity condition (such as SDEs driven by a single Brownian motion or with additive noise). In Chapters 4 and 5, we shall consider the $d = 1$ case and be primarily interested in numerical methods that can achieve high orders of both strong and weak convergence. In addition, we will develop methods for additive noise SDEs in sections 4.4 and 5.3. Our approaches are based on a new polynomial representation of Brownian motion (Theorem 4.1.12) and its direct applications to approximating stochastic integrals. In Chapter 6, we incorporate some of these ideas into a variable step size framework.

2.3 The stochastic Taylor expansion

Just as for ODEs, numerical methods for SDEs can be analysed via Taylor expansions. Therefore, the general strategy used to develop and analyse methods in this thesis is

Step 1. We first consider a high order Taylor approximation of the SDE (2.1.5) when the commutativity condition (2.2.3) is satisfied. In Chapters 4 and 5, we focus on Taylor expansions containing terms with an $L^2(\mathbb{P})$ norm that is at least $O(h^2)$. When we assume the SDE has bounded smooth vector fields with bounded derivatives, it will follow that the approximation has a local error that is $O\big(h^{\frac{5}{2}}\big)$ in an $L^2(\mathbb{P})$ sense (by Theorem 4.3.1).

Step 2. The Taylor approximation described in Step 1 contains certain integrals:

$$\int_s^t \circ\, dW_u, \int_s^t \int_s^u \circ\, dW_v\, du, \int_s^t \int_s^u dv \circ dW_u, \tag{2.3.1}$$

$$\int_s^t \int_s^u \int_s^v \circ\, dW_r \circ dW_v\, du, \int_s^t \int_s^u \int_s^v \circ\, dW_r\, dv \circ dW_u, \int_s^t \int_s^u \int_s^v dr \circ dW_v \circ dW_u.$$

Whilst the integrals in the first line are Gaussian and can be generated, the integrals given in the second line are non-Gaussian and currently can only be approximated. Two key results in the thesis (Theorems 4.2.8 and 5.1.2) give the mean and variance of these non-Gaussian triple iterated integrals conditional on the Gaussian integrals. When the Brownian motion is multidimensional, there are also integrals of the form

$$\int_s^t \int_s^u \int_s^v \circ\, dW_r^{(i)} \circ dW_v^{(j)}\, du, \qquad (i \neq j)$$

We present the conditional expectations of these iterated integrals in Theorem 4.4.1.

Step 3. By applying these new unbiased integrals estimators for (2.3.1) to the Taylor approximation in Step 1, we can derive an explicit stochastic Taylor method.

We then wish to design efficient and practical numerical methods that are locally $O\big(h^{\frac{5}{2}}\big)$ close to this high order stochastic Taylor approximation (in an $L^2(\mathbb{P})$ sense). In Chapters 4 and 5, we shall be considering the following types of numerical method:

- **Log-ODE methods** (definition 4.3.4)
- **Stochastic Runge-Kutta methods** (definition 4.4.3)
- **Piecewise linear ODE methods** (definitions 5.1.6, 5.1.11 and 5.2.2)

It is worth noting that the analysis of above methods also requires Taylor expansions. Therefore, we will now present the Stratonovich-Taylor expansion for the SDE (2.1.5). To do this, we shall first identify vector fields with differential operators as follows:

Definition 2.3.1 (**Vector fields as differential operators**)**.** Let $V : \mathbb{R}^n \to \mathbb{R}^n$ be a vector field on $\mathbb{R}^n$. Then we can identify V with the first order differential operator

$$\sum_{i=1}^{n} V_i(x) \frac{\partial}{\partial x^i}, \tag{2.3.2}$$

where $V = (V_1, V_2, \cdots, V_n)$ and the variables $\{x^i\}$ come from a choice of basis of $\mathbb{R}^n$.

Definition 2.3.2. For vector fields U, V on $\mathbb{R}^n$, we can define the differential operator

$$\begin{aligned} UV &:= \left(\sum_{i=1}^{n} U_i(x) \frac{\partial}{\partial x^i}\right)\left(\sum_{j=1}^{n} V_j(x) \frac{\partial}{\partial x^j}\right) \\ &= \sum_{i,j=1}^{n} U_i(x) \frac{\partial V_j}{\partial x^i}(x) \frac{\partial}{\partial x^j} + \sum_{i,j=1}^{n} U_i V_j \frac{\partial^2}{\partial x^i \partial x^j}. \end{aligned} \tag{2.3.3}$$

Definition 2.3.3 (**Vector field Lie bracket**)**.** Let U and V be vector fields on $\mathbb{R}^n$. We shall define the Lie bracket of U and V as the first order differential operator:

$$[U, V] := UV - VU = \sum_{i,j=1}^{n} \left(U_i(x) \frac{\partial V_j}{\partial x^i}(x) - V_i(x) \frac{\partial U_j}{\partial x^i}(x)\right) \frac{\partial}{\partial x^j}. \tag{2.3.4}$$

Remark 2.3.4. The Lie bracket $[U, V]$ does not depend on the choice of basis of $\mathbb{R}^n$.

Whilst the stochastic Taylor expansion of (2.1.5) is detailed in Chapter 5 of [62], it will be more convenient to use the notation that is given by Proposition 1.1 of [6].

Definition 2.3.5. Let $A^* := \bigcup_{k\geq 0}\{0, 1, \cdots, d\}^k$ denote the set of all multi-indices. For a multi-index $I = (i_1, \cdots, i_k) \in A^*$, we define the degree of I by $\deg(\emptyset) = 0$ and

$$\deg(I) := k + \big|\{j \in \{1, \cdots, k\} : i_j = 0\}\big|. \tag{2.3.5}$$

Theorem 2.3.6 (**Stratonovich-Taylor expansion**). *Let $m \geq 0$ be a fixed constant. Suppose that $F : \mathbb{R}^e \to \mathbb{R}^e$ is an $(m+2)$-times continuously differentiable function. Let y be the unique solution of (2.1.5) where the $d+1$ vector fields $\{f_i\}_{0 \leq i \leq d}$ on $\mathbb{R}^e$ are smooth, bounded and with bounded derivatives. Then $F(y_t)$ can be expanded as*

$$F(y_t) = \sum_{\substack{I=(i_1,\cdots,i_k)\in A^* \\ \deg(I)\leq m,\, k\geq 0}} f_{i_1} \cdots f_{i_k} F(y_0) W_t^I + R_m(t, y_0, F), \tag{2.3.6}$$

where for each multi-index $I = (i_1, \cdots, i_k)$, the Stratonovich integral W_t^I is defined as

$$W_t^I := \int_0^t \int_0^{t_k} \cdots \int_0^{t_2} \circ\, dW_{t_1}^{(i_1)} \cdots \circ\, dW_{t_k}^{(i_k)}, \tag{2.3.7}$$

with $W_t^{\emptyset} \equiv 1$ and the zero coordinate of W denoting time (i.e. $W_t^{(0)} := t$ for $t \geq 0$). Furthermore, there exists a constant $C > 0$ such that the remainder term R_m satisfies

$$\sup_{y_0 \in \mathbb{R}^e} \|R_m(t, y_0, F)\|_{L^2(\mathbb{P})} \leq C t^{\frac{1}{2}\left(m+1+\mathbf{1}_{\{t>1\}}\right)} \sup_{\substack{I=(i_1,\cdots,i_k)\in A^* \\ m<\deg(I)\leq m+2,\, k\geq 0}} \|f_{i_1} \cdots f_{i_k} F\|_{\infty}. \tag{2.3.8}$$

In particular, the remainder term $R_m(t)$ is $O\big(t^{\frac{1}{2}(m+1)}\big)$ in an $L^2(\mathbb{P})$ sense for $t < 1$.

In Chapters 4 and 5, we will apply this stochastic Taylor expansion when $F(y) = y$, $m = 4$ and the vector fields $\{f_i\}_{1 \leq i \leq d}$ satisfy the commutativity condition (2.2.3). When $d = 1$, we explicitly write (2.3.6) in terms of $\{f_i, f_i', f_i''\}_{0 \leq i \leq d}$ (Theorem 4.3.1).

Throughout the thesis we assume the vector fields $\{f_i\}$ are smooth and bounded with bounded derivatives. However, it is possible to establish convergence rates for strong approximations of Stratonovich SDEs under significantly weaker assumptions. We informally describe the conditions required by Theorem 10.7.1 from [62] below:

- (Lipschitz condition) $\big|f_{i_1} \cdots f_{i_k}\mathrm{Id}(t,x) - f_{i_1} \cdots f_{i_k}\mathrm{Id}(t,y)\big| \leq K_1|x-y|$ for all $x, y \in \mathbb{R}^e$ and multi-indices $I = (i_1, \cdots, i_k)$ with $\deg(I) \leq m$.
- (Technical conditions) These are regularity and integrability conditions on each $f_{i_1} \cdots f_{i_k}\mathrm{Id}$ so that $\big\| \int_0^t \int_0^{t_k} \cdots \int_0^{t_2} f_{i_1} \cdots f_{i_k}\mathrm{Id}(t_1, y_{t_1}) \circ dW_{t_1}^{(i_1)} \cdots \circ\, dW_{t_k}^{(i_k)} \big\|_{L^2(\mathbb{P})}$ exists and is finite (for all multi-indices $I = (i_1, \cdots, i_k)$ with $\deg(I) \leq m$).
- (Linear growth condition) $\big|f_{i_1} \cdots f_{i_k}\mathrm{Id}(t,x)\big| \leq K_2\big(1 + |x|\big)$ for all multi-indices $I = (i_1, \cdots, i_k)$ with $\deg(I) \leq m+2$.

At the very least, we expect that the methods proposed in Chapters 4 and 5 will achieve their high convergence rates when f_0 is three times differentiable, $\{f_i\}_{i \geq 1}$ are four times differentiable and all the vector fields and their derivatives are bounded. Analysing our proposed methods under weaker conditions will be left as future work.

Chapter 3

Elements of rough path theory

In this chapter, we shall give a brief overview of rough path theory. In particular, we will focus on the fundamental concepts that will be most relevant to the thesis. Detailed accounts of rough path theory are given by the books of Friz & Hairer [36] and Friz & Victoir [37] as well as the lecture notes of Lyons, Caruana and Lévy [72].

The theory of rough paths was introduced in 1998 by Terry Lyons is his paper "Differential equations driven by rough signals" (see [70] for further information). Rough path theory provides an integration theory for controlled differential equations, which are a natural generalization of SDEs/ODEs and can be expressed succinctly as

$$dY_t = f(Y_t)\, dX_t, \tag{3.0.1}$$

where the solution Y takes values in a Banach space W, the path X takes values in a Banach space V and $f : W \to L(V, W)$ denotes a collection of vector fields on W. The crucial observation of rough path theory is that when X is sufficiently irregular, additional information about the path X is required for the solution Y to be defined. In particular, this information can be viewed as the iterated integrals of the path X. For example, if the driving path is Brownian motion (possibly coupled with time), this information can encode whether (3.0.1) is defined in an Itô or Stratonovich sense. By viewing an SDE as a "rough differential equation", we can study the solution Y in a pathwise manner. This interpretation complements the classical theory of SDEs and gives a topology where Y is well-posed (and thus depends continuously on X). An important object within this theory is the so-called enhanced Brownian motion:

Definition 3.0.1. Let W be a d-dimensional Brownian motion on the interval $[0, T]$. For $T > 0$, let $\Delta_T := \big\{(s, t) : 0 \le s \le t \le T\big\}$. Using Itô or Stratonovich integration, we shall define a stochastic process $\boldsymbol{W}$ taking values in $T^{(2)}\big(\mathbb{R}^d\big) := \mathbb{R} \oplus \mathbb{R}^d \oplus \big(\mathbb{R}^d\big)^{\otimes 2}$. We will refer to the process $\boldsymbol{W} : \Delta_T \to T^{(2)}\big(\mathbb{R}^d\big)$ as an **enhanced Brownian motion**

in the Itô sense if, almost surely, we have the following increment for $(s,t) \in \Delta_T$,

$$\boldsymbol{W}_{s,t} = \left(1, \int_s^t dW_u\,, \int_s^t \int_s^u dW_v \otimes dW_u\right). \tag{3.0.2}$$

Likewise, an **enhanced Brownian motion in the Stratonovich sense** is given by

$$\boldsymbol{W}_{s,t} = \left(1, \int_s^t \circ\, dW_u\,, \int_s^t \int_s^u \circ\, dW_v \otimes \circ\, dW_u\right). \tag{3.0.3}$$

Remark 3.0.2. It can be show that (3.0.2) and (3.0.3) are p-rough paths with $p > 2$. The classical argument is to construct a sequence of discrete approximations of $\boldsymbol{W}$ and consider the limiting object. For example, we could approximate (3.0.3) using piecewise linear discretizations of W and then apply a correction term to get (3.0.2). These "rough path lifts" of Brownian motion will be discussed further in section 3.2. (For a comprehensive account of enhanced Brownian motion, see Chapter 13 of [37]).

Another key observation provided by rough path theory is that when the driving path X has controlled roughness (i.e. finite p-variation), the governing vector fields must compensate for this and have sufficient regularity for (3.0.1) to be well posed. This notion of regularity is referred to as $\mathrm{Lip}(\gamma)$ and is originally due to E. Stein [87].

Definition 3.0.3. A function $f : \mathbb{R}^e \to \mathbb{R}^e$ is said to be $\mathrm{Lip}(\gamma)$ with $\gamma > 1$ if it is bounded with $\lfloor \gamma \rfloor$ bounded derivatives, the last being Hölder continuous with exponent $(\gamma - \lfloor \gamma \rfloor)$. We say that f is $\mathrm{Lip}(1)$ if it is bounded and Lipschitz continuous.

In addition to the analytic results concerning controlled differential equations, there are also algebraic aspects to rough path theory. These arise immediately when one considers the increments and iterated integrals of a bounded variation path X:

$$\int_s^t X_{s,v}\, dX_v = \int_s^u X_{s,v}\, dX_v + X_{s,u}\, X_{u,t} + \int_u^t X_{u,v}\, dX_v\,, \tag{3.0.4}$$

where these integrals are defined in the Riemann-Stieltjes sense and $X_{s,v} := X_v - X_s$. The above equality follows from standard properties of integration and implies that

$$\begin{aligned} &\left(1, \int_s^t dX_v\,, \int_s^t \int_s^v dX_r \otimes dX_v\right) \\ &\quad = \left(1, \int_s^u dX_v\,, \int_s^u \int_s^v dX_r \otimes dX_v\right) \otimes \left(1, \int_u^t dX_v\,, \int_u^t \int_u^v dX_r \otimes dX_v\right), \end{aligned} \tag{3.0.5}$$

where the middle $\otimes$ denotes the tensor product on $T^{(2)}\big(\mathbb{R}^d\big) := \mathbb{R} \oplus \mathbb{R}^d \oplus \big(\mathbb{R}^d\big)^{\otimes 2}$. The identity (3.0.5) can be seen as an example of Chen's relation (Theorem 3.1.20), which is a fundamental property that is used to define rough paths more generally. We begin the chapter by giving an overview of the algebraic side to rough path theory.

3.1 Signatures and tensor algebras

One of the central objects within rough path theory is the so-called "path signature". In order to define the signature, it will be necessary to consider the following spaces:

Definition 3.1.1. We say that $T\big(\mathbb{R}^d\big) := \bigoplus_{k\geq 0} \big(\mathbb{R}^d\big)^{\otimes k}$ is the tensor algebra of $\mathbb{R}^d$ and $T\big(\big(\mathbb{R}^d\big)\big) := \Big\{\boldsymbol{a} = \big(a_0, a_1, \cdots\big) : a_k \in \big(\mathbb{R}^d\big)^{\otimes k} \ \forall k \geq 0\Big\}$ is the set of formal series of tensors of $\mathbb{R}^d$. Moreover, $T\big(\mathbb{R}^d\big)$ and $T\big(\big(\mathbb{R}^d\big)\big)$ can be endowed with the operations of addition and multiplication. Given $\boldsymbol{a} = (a_0, a_1, \cdots)$ and $\boldsymbol{b} = (b_0, b_1, \cdots)$, we have

$$\boldsymbol{a} + \boldsymbol{b} = \big(a_0 + b_0, a_1 + b_1, \cdots\big), \tag{3.1.1}$$

$$\boldsymbol{a} \otimes \boldsymbol{b} = \big(c_0, c_1, c_2, \cdots\big), \tag{3.1.2}$$

where for $n \geq 0$, the n-th term $c_n \in \big(\mathbb{R}^d\big)^{\otimes n}$ can be written using tensor notation as

$$c_n := \sum_{k=0}^{n} a_k \otimes b_{n-k}. \tag{3.1.3}$$

The operation $\otimes$ given by (3.1.2) is called the "tensor product".

Remark 3.1.2. With these operations, we have that $T\big(\mathbb{R}^d\big)$ and $T\big(\big(\mathbb{R}^d\big)\big)$ are algebras.

Due to the structure of $T\big(\mathbb{R}^d\big)$, we can study certain finite dimensional subspaces:

Definition 3.1.3. For $N \geq 0$, we say that $T^{(N)}\big(\mathbb{R}^d\big) := \bigoplus_{k=0}^{N} \big(\mathbb{R}^d\big)^{\otimes k}$ is the N-step truncated tensor algebra of $\mathbb{R}^d$, which we can endow with addition and multiplication. For $\boldsymbol{a}, \boldsymbol{b} \in T^{(N)}\big(\mathbb{R}^d\big)$, we define $\boldsymbol{a} \otimes \boldsymbol{b} := P_N(\iota_N(\boldsymbol{a}) \otimes \iota_N(\boldsymbol{b}))$ where the functions $P_N : T\big(\big(\mathbb{R}^d\big)\big) \to T^{(N)}\big(\mathbb{R}^d\big)$ and $\iota_N : T^{(N)}\big(\mathbb{R}^d\big) \to T\big(\big(\mathbb{R}^d\big)\big)$ are the natural projection and inclusion maps. In particular, $T^{(N)}\big(\mathbb{R}^d\big)$ is not a subalgebra of $T\big(\mathbb{R}^d\big)$ or $T\big(\big(\mathbb{R}^d\big)\big)$.

Remark 3.1.4. There are natural isomorphisms between the algebras $T\big(\big(\mathbb{R}^d\big)\big)$, $T\big(\mathbb{R}^d\big)$, $T^{(N)}\big(\mathbb{R}^d\big)$ and certain spaces of non-commutative polynomials (that have d variables). Let A be an alphabet with d letters and $\mathbb{R}\langle\langle A\rangle\rangle$ denote the space of all formal series over $\mathbb{R}$ in non-commutative variables from A. Similarly, let $\mathbb{R}\langle A\rangle$ be the associated subspace of polynomials (i.e formal series with a finite number of non-zero terms). It is worth noting that the standard notions of addition and multiplication in $\mathbb{R}\langle A\rangle$ immediately extend to $\mathbb{R}\langle\langle A\rangle\rangle$. In other words, $\mathbb{R}\langle\langle A\rangle\rangle$ and $\mathbb{R}\langle A\rangle$ are both algebras. Moreover, it is then a straightforward exercise to establish the identifications below.

$$\begin{aligned} T\big(\big(\mathbb{R}^d\big)\big) &\simeq \mathbb{R}\langle\langle A\rangle\rangle, \\ T\big(\mathbb{R}^d\big) &\simeq \mathbb{R}\langle A\rangle, \\ T^{(N)}\big(\mathbb{R}^d\big) &\simeq \{p \in \mathbb{R}\langle A\rangle : \deg p \leq N\}. \end{aligned}$$

In rough path theory, there are three important functions defined on these spaces.

Definition 3.1.5 (The exponential of a formal series)**.** We define the function exp as

$$\exp : T\big(\big(\mathbb{R}^d\big)\big) \to T_1\big(\big(\mathbb{R}^d\big)\big), \tag{3.1.4}$$

$$\boldsymbol{a} \mapsto \sum_{n=0}^{\infty} \frac{\boldsymbol{a}^{\otimes n}}{n!}, \tag{3.1.5}$$

where $\boldsymbol{a} = (a_0, a_1, a_2, \cdots)$ is an element of $T\big(\big(\mathbb{R}^d\big)\big)$ and the set $T_1\big(\big(\mathbb{R}^d\big)\big)$ is given by

$$T_1\big(\big(\mathbb{R}^d\big)\big) := \big\{\boldsymbol{a} \in T\big(\big(\mathbb{R}^d\big)\big) : a_0 = 1\big\}. \tag{3.1.6}$$

Remark 3.1.6. The series (3.1.5) is defined as the limit of the partial sums $\big(\sum_{n=0}^{N} \frac{\boldsymbol{a}^{\otimes n}}{n!}\big)$. It is straightforward to verify that this limit is well-defined in the coordinatewise sense.

Definition 3.1.7 (The logarithm of a formal series)**.** For $\boldsymbol{a} = (a_0, a_1, \cdots) \in T\big(\big(\mathbb{R}^d\big)\big)$ with $a_0 > 0$, define $\log(\boldsymbol{a})$ to be the element of $T\big(\big(\mathbb{R}^d\big)\big)$ given by the following series:

$$\log(\boldsymbol{a}) := \log(a_0) + \sum_{n=1}^{\infty} \frac{(-1)^n}{n} \Big(\mathbf{1} - \frac{\boldsymbol{a}}{a_0}\Big)^{\otimes n}, \tag{3.1.7}$$

where $\mathbf{1} = (1, 0, \cdots)$ is the unit element of $T\big(\big(\mathbb{R}^d\big)\big)$ and $\log(a_0)$ is viewed as $\log(a_0)\mathbf{1}$.

Remark 3.1.8. The series (3.1.7) is locally finite, and thus defined purely algebraically. (That is, for any k, the entries with degree k can be written as a finite sum of terms)

Definition 3.1.9 (The inverse of a formal series)**.** For $\boldsymbol{a} = (a_0, a_1, \cdots) \in T\big(\big(\mathbb{R}^d\big)\big)$ with $a_0 \neq 0$, we define $\boldsymbol{a}^{-1}$ to be the element of $T\big(\big(\mathbb{R}^d\big)\big)$ given by the following series:

$$\boldsymbol{a}^{-1} := a_0^{-1} \sum_{n=0}^{\infty} \Big(\mathbf{1} - \frac{\boldsymbol{a}}{a_0}\Big)^{\otimes n}. \tag{3.1.8}$$

Remark 3.1.10. As before, the series (3.1.8) is locally finite and therefore well-defined.

Similar to (3.1.6), we will also consider the subset of $T\big(\big(\mathbb{R}^d\big)\big)$ with leading term zero.

$$T_0\big(\big(\mathbb{R}^d\big)\big) := \big\{\boldsymbol{a} \in T\big(\big(\mathbb{R}^d\big)\big) : a_0 = 0\big\}. \tag{3.1.9}$$

Given these functions, one can immediately prove a number of elementary properties. In particular, when the domain of the exponential function (3.1.5) is restricted to $T_0\big(\big(\mathbb{R}^d\big)\big)$ and the domain of the logarithm function (3.1.7) is restricted to $T_1\big(\big(\mathbb{R}^d\big)\big)$, it is straightforward to directly verify that exp and log are inverses of one another. Similarly, it can be shown that the formula (3.1.8) for $\boldsymbol{a}^{-1}$ satisfies $\boldsymbol{a}^{-1}\boldsymbol{a} = \boldsymbol{a}\boldsymbol{a}^{-1} = \mathbf{1}$. As one would expect, there is also an identity concerning the inverse of an exponential. We will give these results in the below theorem (see chapter 2 of [72] for more details).

Theorem 3.1.11 (Basic properties of the exponential, logarithm and inverse maps)**.** *The maps* $\exp : T_0\big(\big(\mathbb{R}^d\big)\big) \to T_1\big(\big(\mathbb{R}^d\big)\big)$ *and* $\log : T_1\big(\big(\mathbb{R}^d\big)\big) \to T_0\big(\big(\mathbb{R}^d\big)\big)$ *are inverses of each other. In addition, there exists an inverse map* $\boldsymbol{a} \mapsto \boldsymbol{a}^{-1}$ *on the subset* $T\big(\big(\mathbb{R}^d\big)\big) \setminus T_0\big(\big(\mathbb{R}^d\big)\big)$ *given by the series (3.1.8). We also have the following identities:*

$$\exp(\mathbf{0}) = \mathbf{1},$$
$$\log(\mathbf{1}) = \mathbf{0},$$
$$\exp(\boldsymbol{a})^{-1} = \exp(-\boldsymbol{a}),$$
$$\log\left(\boldsymbol{a}^{-1}\right) = -\log(\boldsymbol{a}).$$

In addition, these maps can be defined on the truncated tensor algebra $T^{(N)}\big(\mathbb{R}^d\big)$. To do this, we will first consider the projection and inclusion maps on these algebras.

Definition 3.1.12 (Projection and inclusion maps)**.** Let $P_N : T\big(\big(\mathbb{R}^d\big)\big) \to T^{(N)}\big(\mathbb{R}^d\big)$ be the natural projection map and $\iota_N : T^{(N)}\big(\mathbb{R}^d\big) \to T\big(\big(\mathbb{R}^d\big)\big)$ be the inclusion map. Let $\pi_N : T\big(\big(\mathbb{R}^d\big)\big) \to \big(\mathbb{R}^d\big)^{\otimes N}$ denote the projection onto the N^{th} tensor power of $\mathbb{R}^d$.

We can now define truncated versions of exp, log and $(\cdot)^{-1}$ using the above maps.

Definition 3.1.13 (The exponential, logarithm and inverse functions on $T^{(N)}\big(\mathbb{R}^d\big)$)**.** Consider the subset of $T^{(N)}\big(\mathbb{R}^d\big)$ given by $T_+^{(N)}\big(\mathbb{R}^d\big) := \big\{\boldsymbol{a} \in T^{(N)}\big(\mathbb{R}^d\big) : \pi_0(\boldsymbol{a}) > 0\big\}$. We can define the exponential and logarithm maps on these (truncated) domains as

$$\begin{aligned} \exp^{(N)} : T^{(N)}\big(\mathbb{R}^d\big) &\to T^{(N)}\big(\mathbb{R}^d\big), \\ \boldsymbol{a} &\mapsto \big(P_N \circ \exp \circ \iota_N\big)(\boldsymbol{a}), \end{aligned} \tag{3.1.10}$$

$$\begin{aligned} \log^{(N)} : T_+^{(N)}\big(\mathbb{R}^d\big) &\to T^{(N)}\big(\mathbb{R}^d\big), \\ \boldsymbol{a} &\mapsto \big(P_N \circ \log \circ \iota_N\big)(\boldsymbol{a}), \end{aligned} \tag{3.1.11}$$

and for $\boldsymbol{a} = (a_0, a_1, \cdots, a_n) \in T^{(N)}\big(\mathbb{R}^d\big)$ with $a_0 \neq 0$, we can define the inverse $\boldsymbol{a}^{-1}$ as

$$\boldsymbol{a}^{-1} := P_N\big(\iota_N(\boldsymbol{a})^{-1}\big). \tag{3.1.12}$$

Remark 3.1.14. The above maps will satisfy a "truncated version" of Theorem 3.1.11.

Throughout this chapter, we shall see that there are additional algebraic structures within $T^{(N)}\big(\mathbb{R}^d\big)$. In particular, there is a free nilpotent Lie group which will enable us to study certain group-valued rough paths that have desirable analytic properties. The logarithm map (3.1.11) then maps this Lie group to a certain free Lie algebra. This connection with Lie theory will become relevant in the next section, where we introduce the geometric rough paths (which satisfy an integration by parts formula).

We will now define the signature of a finite variation path X taking values in $\mathbb{R}^d$.

Definition 3.1.15. The path $X : [0, T] \to \mathbb{R}^d$ is said to have finite variation if

$$\|X\|_{1\text{-var};[0,T]} := \sup_{\mathcal{D}} \sum_k \|X_{t_{k+1}} - X_{t_k}\| < \infty, \tag{3.1.13}$$

where $\|\cdot\|$ is a norm on $\mathbb{R}^d$ and the supremum is taken over all partitions $\mathcal{D}$ of $[0, T]$.

Definition 3.1.16. The **signature** of a finite variation path $X : [0, T] \to \mathbb{R}^d$ over the interval $[s, t]$ is defined as the following collection of Riemann-Stieltjes integrals:

$$S_{s,t}\big(X\big) := \Big(1, X^{(1)}_{s,t}, X^{(2)}_{s,t}, X^{(3)}_{s,t}, \cdots\Big) \in T\big(\big(\mathbb{R}^d\big)\big), \tag{3.1.14}$$

where for $n \geq 1$,

$$X^{(n)}_{s,t} := \int\limits_{s<u_1<\cdots<u_n<t} \cdots \int dX_{u_1} \otimes \cdots \otimes dX_{u_n} \in \big(\mathbb{R}^d\big)^{\otimes n}. \tag{3.1.15}$$

Similarly, we can define the N-step (or truncated) signature of the path X on $[s, t]$.

$$S^{(N)}_{s,t}\big(X\big) := \left(1, \int_{s<u_1<t} dX_u, \cdots, \int\limits_{s<u_1<\cdots<u_N<t} \cdots \int dX_{u_1} \otimes \cdots \otimes dX_{u_N}\right) \in T^{(N)}\big(\mathbb{R}^d\big). \tag{3.1.16}$$

The signature gives a high level description of the path and has seen a number of recent applications within machine learning [71]. One of the main reasons for this, is that $S_{s,t}\big(X\big)$ uniquely determines the path X on the interval $[s, t]$ modulo certain equivalences [52] (for example $S_{s,t}\big(X\big)$ does not depend on the parametrization of X). The following theorem indicates that the signature is related to the exponential map.

Theorem 3.1.17 (The signature of a linear path)**.** *Suppose that $X_u = X_s + \lambda(u - s)$ for $u \in [s, t]$ where $\lambda \in \mathbb{R}^d$ is some constant. Then the signature of X over $[s, t]$ is*

$$S_{s,t}\big(X\big) = \exp\Big(\iota_1\big(0, \lambda(t - s)\big)\Big), \tag{3.1.17}$$

where exp *is the exponential given by (3.1.5) and ι_1 is an inclusion map into $T\big(\big(\mathbb{R}^d\big)\big)$.*

Remark 3.1.18. With a slight abuse of notation, we will no longer display the map ι_1. Hence the (truncated) signature of the linear path X can be concisely expressed as

$$S_{s,t}\big(X\big) = \exp\big(X_{s,t}\big), \tag{3.1.18}$$

$$S^N_{s,t}\big(X\big) = \exp^{(N)}\big(X_{s,t}\big). \tag{3.1.19}$$

Definition 3.1.19. The **log-signature** of a path X is $\log\big(S_{s,t}(X)\big) \in T_0\big(\big(\mathbb{R}^d\big)\big)$. Similar to before, we can also define the N-step log-signature as $\log^{(N)}\big(S^{(N)}_{s,t}(X)\big)$.

By applying the logarithm, we see the log-signature of a linear path X is $X_{s,t}$. Moreover, we will see that the log-signature is an important quantity in rough path theory and leads to the "log-ODE method" (which will be presented in section 3.4).

To conclude this section, we shall discuss a key property of signatures known as Chen's relation, which allows us to interpret the signature map as a homomorphism.

Theorem 3.1.20 (**Chen's relation**). *Let $X : [0,T] \to \mathbb{R}^d$ be a finite variation path. Then for $s,u,t \in [0,T]$ with $s \le u \le t$, the signatures on $[s,u]$, $[u,t]$, $[s,t]$ satisfy*

$$S_{s,t}(X) = S_{s,u}(X) \otimes S_{u,t}(X), \tag{3.1.20}$$

where $\otimes$ is the tensor product given by definition 3.1.1. Similarly, we have

$$S^{(N)}_{s,t}(X) = S^{(N)}_{s,u}(X) \otimes S^{(N)}_{u,t}(X). \tag{3.1.21}$$

Proof. The result follows directly from the additivity and linearity of integration. □

Remark 3.1.21. Suppose that X and Y are finite variation paths on the intervals $[s,u]$ and $[u,t]$ respectively (both taking values in $\mathbb{R}^d$). Then (3.1.20) can be expressed as

$$S_{s,t}(X * Y) = S_{s,u}(X) \otimes S_{u,t}(Y), \tag{3.1.22}$$

where $*$ is path concatenation. In this sense, we see that the signature transform $X \mapsto S(X)$ is a homomorphism from the monoid of finite variation paths to $T_1\big(\big(\mathbb{R}^d\big)\big)$.

Moreover, we can take the view that the signature map is a homomorphism further. Concatenating a path X with its reverse path $\overleftarrow{X}$ does not result in the "zero path". However, the resulting path $X * \overleftarrow{X}$ has no total increment or any enclosed area over the interval and thus can be viewed as a "tree-like" excursion from the zero path. This then implies that $X * \overleftarrow{X}$ has trivial iterated integrals and so $S\big(X * \overleftarrow{X}\big) = \mathbf{1}$.

Theorem 3.1.22 (The inverse signature of a path). *Let X be a finite variation path over the interval $[s,t]$. Then the inverse signature of the path X can be expressed as*

$$S_{s,t}(X)^{-1} = S_{s,t}\big(\overleftarrow{X}\big), \tag{3.1.23}$$

where $\overleftarrow{X} := \big\{X_{t-(u-s)}\big\}_{u\in[s,t]}$ is the path X run backwards. Similarly, we also have

$$S^{(N)}_{s,t}(X)^{-1} = S^{(N)}_{s,t}\big(\overleftarrow{X}\big). \tag{3.1.24}$$

Remark 3.1.23. In this sense, the signature map can be viewed as a homomorphism from the group of finite variation paths (modulo tree-like excursions) into $T_1\big(\big(\mathbb{R}^d\big)\big)$.

3.2 Rough paths and the p-variation metric

Having previously discussed certain algebraic structures defined using tensors of $\mathbb{R}^d$, we shall now turn our attention to some of the analytic aspects of rough path theory. To begin, we will equip the tensor algebra $T\big(\mathbb{R}^d\big)$ with the following family of norms:

Definition 3.2.1 (Admissible norms on tensor products)**.** We say that the tensor powers of $\mathbb{R}^d$ are endowed with admissible norms if the below conditions are satisfied:

1. The n-th norm $\|\cdot\|$ on $\big(\mathbb{R}^d\big)^{\otimes n}$ is invariant under the symmetric group S_n, i.e.

$$\|\sigma v\| = \|v\|, \quad \forall v \in \big(\mathbb{R}^d\big)^{\otimes n}, \ \forall \sigma \in S_n. \tag{3.2.1}$$

2. The tensor product has an operator norm of 1. That is, for $n, m \geq 1$, we have

$$\|v \otimes w\| \leq \|v\|\|w\|, \quad \forall v \in \big(\mathbb{R}^d\big)^{\otimes n}, \ \forall w \in \big(\mathbb{R}^d\big)^{\otimes m}. \tag{3.2.2}$$

Definition 3.2.2 (Examples of admissible norms on $T\big(\mathbb{R}^d\big)$)**.** Let $v \in \big(\mathbb{R}^d\big)^{\otimes n}$. Then

- the projective norm of v is

$$\|v\|_1 := \sum_{i_1, i_2, \cdots, i_n \in \{1, \cdots, d\}} \left|v^{i_1, i_2, \cdots, i_n}\right|, \tag{3.2.3}$$

- the injective norm of v is

$$\|v\|_\infty := \sup_{i_1, i_2, \cdots, i_n \in \{1, \cdots, d\}} \left|v^{i_1, i_2, \cdots, i_n}\right|, \tag{3.2.4}$$

- the L^2 norm of v is

$$\|v\|_2 := \sqrt{\sum_{i_1, i_2, \cdots, i_n \in \{1, \cdots, d\}} \left|v^{i_1, i_2, \cdots, i_n}\right|^2}, \tag{3.2.5}$$

As indicated in the previous section, the truncated signature of a finite variation path can itself be viewed as a path taking values in a certain group within $T^{(N)}\big(\mathbb{R}^d\big)$. This is referred to as the "N-step free nilpotent group" $G^N\big(\mathbb{R}^d\big)$, and is defined below.

Definition 3.2.3 (The N-step free nilpotent group)**.** Consider the subset of $T^{(N)}\big(\mathbb{R}^d\big)$,

$$G^N\big(\mathbb{R}^d\big) := \Big\{ S_{0,1}^{(N)}\big(X\big) : X \text{ is a finite variation path taking values in } \mathbb{R}^d \Big\}. \tag{3.2.6}$$

Then by Chen's relation, $G^N\big(\mathbb{R}^d\big)$ becomes a group under the tensor product (3.1.2).

Associated with the Lie group $G^N(\mathbb{R}^d)$ is the "N-step free Lie algebra" $\mathfrak{g}^N(\mathbb{R}^d)$.

Definition 3.2.4 (The N-step free Lie algebra)**.** Consider the commutator Lie bracket

$$[\boldsymbol{a}, \boldsymbol{b}] := \boldsymbol{a} \otimes \boldsymbol{b} - \boldsymbol{b} \otimes \boldsymbol{a}, \tag{3.2.7}$$

for $\boldsymbol{a}, \boldsymbol{b} \in T\big((\mathbb{R}^d)\big)$. Then for subspaces U and V of $T\big((\mathbb{R}^d)\big)$, we can define the space

$$[U, V] := \mathbb{R}\big\langle \big\{[\boldsymbol{a}, \boldsymbol{b}] : \boldsymbol{a} \in U, \boldsymbol{b} \in V\big\}\big\rangle \subseteq T\big((\mathbb{R}^d)\big). \tag{3.2.8}$$

The N-step free Lie algebra $\mathfrak{g}^N(\mathbb{R}^d)$ is the smallest subalgebra of $T^{(N)}(\mathbb{R}^d)$ that contains $\mathbb{R}^d$ and is closed under the Lie bracket (3.2.7). That is, $\mathfrak{g}^N(\mathbb{R}^d)$ is given by

$$\mathfrak{g}^N(\mathbb{R}^d) = \mathbb{R}^d \oplus \big[\mathbb{R}^d, \mathbb{R}^d\big] \oplus \cdots \oplus \underbrace{\big[\mathbb{R}^d, \big[\ldots, \big[\mathbb{R}^d, \mathbb{R}^d\big]\big]\big]}_{(N-1) \text{ brackets}}. \tag{3.2.9}$$

Using the exponential map, we can give the key relationship between G^N and $\mathfrak{g}^N$.

Theorem 3.2.5. *The N-step free nilpotent group $G^N(\mathbb{R}^d)$ can also be expressed as*

$$G^N(\mathbb{R}^d) = \exp^{(N)}\Big(\mathfrak{g}^N(\mathbb{R}^d)\Big), \tag{3.2.10}$$

and

$$G^N(\mathbb{R}^d) = \left\{ \bigotimes_{i=1}^{m} \exp^{(N)}(x_i) : x_1, \cdots, x_m \in \mathbb{R}^d \right\}. \tag{3.2.11}$$

As one would expect, we can extend the above definitions and hence construct the full nilpotent Lie group $G(\mathbb{R}^d)$ and free Lie algebra $\mathfrak{g}(\mathbb{R}^d)$ as subsets of $T\big((\mathbb{R}^d)\big)$.

Definition 3.2.6 (The full Lie group and Lie algebra)**.** We define the set $G(\mathbb{R}^d)$ as

$$G(\mathbb{R}^d) := \Big\{ S_{0,1}(X) : X \text{ is a finite variation path taking values in } \mathbb{R}^d \Big\}. \tag{3.2.12}$$

Then by Chen's relation, $G(\mathbb{R}^d)$ becomes a group under the tensor product.
We can also define the free Lie algebra $\mathfrak{g}(\mathbb{R}^d)$ as the smallest subalgebra of $T\big((\mathbb{R}^d)\big)$ that contains $\mathbb{R}^d$ and is closed under the Lie bracket (3.2.7). Thus, $\mathfrak{g}(\mathbb{R}^d)$ is given by

$$\mathfrak{g}(\mathbb{R}^d) = \Big\{ \boldsymbol{a} \in T\big((\mathbb{R}^d)\big) : P_N(\boldsymbol{a}) \in \mathfrak{g}^N(\mathbb{R}^d), \quad \forall N \geq 0 \Big\}. \tag{3.2.13}$$

Remark 3.2.7. In the remainder of this chapter, we may refer to elements of the algebra $\mathfrak{g}(\mathbb{R}^d)$ as Lie series and say that elements of $\bigcup_{N \geq 0} \mathfrak{g}^N(\mathbb{R}^d)$ are Lie polynomials.

Remark 3.2.8. For $N \geq 0$, we have $G^N(\mathbb{R}^d) = P_N\big(G(\mathbb{R}^d)\big)$ and $\mathfrak{g}^N(\mathbb{R}^d) = P_N\big(\mathfrak{g}(\mathbb{R}^d)\big)$.

As before, the exponential map gives a key connection between $G\big(\mathbb{R}^d\big)$ and $\mathfrak{g}\big(\mathbb{R}^d\big)$.

Theorem 3.2.9. *The free Lie group $G\big(\mathbb{R}^d\big)$ and Lie algebra $\mathfrak{g}\big(\mathbb{R}^d\big)$ can be related by*

$$G\big(\mathbb{R}^d\big) = \exp\big(\mathfrak{g}\big(\mathbb{R}^d\big)\big)\,. \tag{3.2.14}$$

A useful concept within Chapter 6 will be the "homogeneous" norm on $G^N\big(\mathbb{R}^d\big)$.

Definition 3.2.10. A homogeneous norm on the N-step free nilpotent group is a function $\|\cdot\| : G^N\big(\mathbb{R}^d\big) \to [0,\infty)$ that is continuous and satisfies the below conditions:

1. $\|\boldsymbol{g}\| = 0$ if and only if $\boldsymbol{g}$ is the unit element $\mathbf{1} \in G^N\big(\mathbb{R}^d\big)$.
2. For any element $\boldsymbol{g} \in G^N\big(\mathbb{R}^d\big)$ and scalar $\lambda \in \mathbb{R}$, we have the following equality:

$$\|\delta_\lambda(\boldsymbol{g})\| = |\lambda| \cdot \|\boldsymbol{g}\|, \tag{3.2.15}$$

   where $\delta_\lambda : T\big(\mathbb{R}^d\big) \to T\big(\mathbb{R}^d\big)$ is the dilation map with the defining property that

$$\pi_n\big(\delta_\lambda(\boldsymbol{g})\big) = \lambda^n \pi_n(\boldsymbol{g}), \tag{3.2.16}$$

   for $n \geq 0$.

In addition, a homogeneous norm on $G^N\big(\mathbb{R}^d\big)$ is said to be symmetric if $\|\boldsymbol{g}\| = \|\boldsymbol{g}^{-1}\|$ and subadditive if $\|\boldsymbol{g} \otimes \boldsymbol{h}\| \leq \|\boldsymbol{g}\| + \|\boldsymbol{h}\|$ (where $\otimes$ denotes the usual tensor product).

Remark 3.2.11. The classic example of a homogeneous norm on $G^N\big(\mathbb{R}^d\big)$ is given by

$$\|\boldsymbol{g}\|_G := \max_{k=1,2,\cdots,N} \|\pi_k(\boldsymbol{g})\|^{\frac{1}{k}}, \tag{3.2.17}$$

for $\boldsymbol{g} \in G^N\big(\mathbb{R}^d\big)$, where the above $\|\cdot\|$ form a family of admissible norms on $T^{(N)}\big(\mathbb{R}^d\big)$. In general, the homogeneous norm given by (3.2.17) is not symmetric or subadditive.

Remark 3.2.12. Just as for norms on $\mathbb{R}^d$ and admissible norms on $T\big(\mathbb{R}^d\big)$, we have that all the homogeneous norms on the free nilpotent group $G^N\big(\mathbb{R}^d\big)$ are equivalent. So for homogeneous norms $\|\cdot\|_1$ and $\|\cdot\|_2$, there exists a constant $C \geq 1$ such that

$$\frac{1}{C}\|\boldsymbol{g}\|_1 \leq \|\boldsymbol{g}\|_2 \leq C\|\boldsymbol{g}\|_1, \tag{3.2.18}$$

for all $\boldsymbol{g} \in G^N\big(\mathbb{R}^d\big)$.

Given the above definitions, we would like to explicitly find a homogeneous norm on the N-step free nilpotent group $G^N\big(\mathbb{R}^d\big)$ that is both symmetric and subadditive. Such a homogeneous norm can then be used to define a metric on the group $G^N\big(\mathbb{R}^d\big)$. This will be possible using the Carnot-Caratheodory distance (Theorem 7.32 in [37]).

Definition 3.2.13 (**Carnot-Caratheodory distance**)**.** For $\boldsymbol{g} \in G^N\big(\mathbb{R}^d\big)$, we define

$$\|\boldsymbol{g}\| := \inf\left\{\int_0^1 |d\gamma| : \gamma \text{ is a finite variation path in } \mathbb{R}^d \text{ with } S_{0,1}^{(N)}(\gamma) = \boldsymbol{g}\right\}.$$

Then $\|\cdot\|$ gives a well-defined homogeneous norm on $G^N\big(\mathbb{R}^d\big)$ that is symmetric, subadditive and continuous. Moreover, it induces the following metric on the group:

$$d\big(\boldsymbol{g}, \boldsymbol{h}\big) := \big\|\boldsymbol{g}^{-1} \otimes \boldsymbol{h}\big\|. \tag{3.2.19}$$

The above metric is commonly referred to as the "Carnot-Caratheodory distance". Moreover, the Carnot-Caratheodory metric (3.2.19) is left-invariant. In other words,

$$d\big(\boldsymbol{g} \otimes \boldsymbol{h}, \boldsymbol{g} \otimes \boldsymbol{k}\big) = d\big(\boldsymbol{h}, \boldsymbol{k}\big),$$

for $\boldsymbol{g}, \boldsymbol{h}, \boldsymbol{k} \in G^N\big(\mathbb{R}^d\big)$.

Another key concept within rough path theory is the multiplicative functional, which extends Chen's relation (Theorem 3.1.20) to paths taking values in $T^{(N)}\big(\mathbb{R}^d\big)$. We can then define rough paths as multiplicative functionals with certain regularity.

Definition 3.2.14 (Rescaled simplex in $\mathbb{R}^2$)**.** We define the simplex $\triangle_T \subset [0,T]^2$ as

$$\triangle_T := \big\{(s,t) \in [0,T]^2 : 0 \le s \le t \le T\big\}.$$

Definition 3.2.15 (**Multiplicative functional**)**.** Let $N \ge 1$ be a positive integer and let $X : \triangle_T \to T^{(N)}\big(\mathbb{R}^d\big)$ be a continuous function. For $(s,t) \in \triangle_T$, we use the notation $X_{s,t}$ to denote the image $X(s,t)$. We will refer to $X_{s,t}$ as the path increment:

$$X_{s,t} = \Big(X_{s,t}^{(0)}, X_{s,t}^{(1)}, \cdots, X_{s,t}^{(N)}\Big) \in T^{(N)}\big(\mathbb{R}^d\big).$$

The function X is called a multiplicative functional of degree N in $\mathbb{R}^d$ if $X_{s,t}^{(0)} = 1$ and

$$X_{s,u} \otimes X_{u,t} = X_{s,t}, \tag{3.2.20}$$

for all $(s,u), (u,t) \in \triangle_T$.

Remark 3.2.16. Although we have the Chen property (3.2.20), it does not follow that a multiplicative functional X of degree N in $\mathbb{R}^d$ takes its values in the group $G^N\big(\mathbb{R}^d\big)$.

We shall now introduce the p-variation metric, which will extend the notion of finite variation given by (3.1.13) to more irregular paths (such as Brownian motion). Within rough path theory, there are two popular definitions for the p-variation metric.

Definition 3.2.17 (**The p-variation metric for multiplicative functionals**). For multiplicative functionals X, Y of degree N, we define the p-variation metric as

$$d_{p\text{-var};[s,t]}(X,Y) := \max_{n=1,2,\cdots,N} \sup_{\mathcal{D}} \left(\sum_k \left\| X^{(n)}_{t_k,t_{k+1}} - Y^{(n)}_{t_k,t_{k+1}} \right\|^{\frac{p}{n}} \right)^{\frac{n}{p}}, \tag{3.2.21}$$

where supremum is taken over all partitions $\mathcal{D}$ of $[s,t]$ and $\|\cdot\|$ are admissible norms. Similarly, we can define the p-variation norm of X by removing Y from the above:

$$\|X\|_{p\text{-var};[s,t]} := \max_{n=1,2,\cdots,N} \sup_{\mathcal{D}} \left(\sum_k \left\| X^{(n)}_{t_k,t_{k+1}} \right\|^{\frac{p}{n}} \right)^{\frac{n}{p}}. \tag{3.2.22}$$

Definition 3.2.18 (**The homogeneous p-variation metric for paths in G^N**). For multiplicative functionals X and Y of degree N and taking values in the N-step free nilpotent group $G^N\big(\mathbb{R}^d\big)$, we can define the homogeneous p-variation metric as

$$\tilde{d}_{p\text{-var};[s,t]}(X,Y) := \left(\sup_{\mathcal{D}} \sum_k d\big(X_{t_k,t_{k+1}}, Y_{t_k,t_{k+1}}\big)^p \right)^{1/p}, \tag{3.2.23}$$

where d is the Carnot-Caratheodory metric.

Another metric that can be defined on the space of multiplicative functionals is the α-Hölder metric and is intimately related to the p-variation metric given by (3.2.21).

Definition 3.2.19 (**The α-Hölder metric for multiplicative functionals**). For multiplicative functionals X and Y of degree N, the α-Hölder metric is defined as

$$d_{\alpha\text{-Höl};[s,t]}(X,Y) := \max_{n=1,2,\cdots,N} \sup_{s\le u<v\le t} \frac{\|X^{(n)}_{u,v} - Y^{(n)}_{u,v}\|}{|v-u|^{n\alpha}}, \tag{3.2.24}$$

where the $\|\cdot\|$ are a family of admissible norms on the truncated algebra $T^{(N)}\big(\mathbb{R}^d\big)$. Just as before, we can define the α-Hölder norm of X by removing Y from the above:

$$\|X\|_{\alpha\text{-Höl};[s,t]} := \max_{n=1,2,\cdots,N} \sup_{s\le u<v\le t} \frac{\|X^{(n)}_{u,v}\|}{|v-u|^{n\alpha}}. \tag{3.2.25}$$

Definition 3.2.20 (**The homogeneous α-Hölder metric for paths in G^N**). For multiplicative functionals X and Y of degree N and taking values in the N-step free nilpotent group $G^N\big(\mathbb{R}^d\big)$, we can define the homogeneous α-Hölder metric as

$$\tilde{d}_{\alpha\text{-Höl};[s,t]}(X,Y) := \sup_{s\le u<v\le t} \frac{d\big(X_{u,v}, Y_{u,v}\big)}{|v-u|^{\alpha}}, \tag{3.2.26}$$

where d is the Carnot-Caratheodory metric.

We can view the p-variation norm as a parametrization invariant version of the α-Hölder norm. Furthermore, we have a simple estimate that relates the two norms.

Theorem 3.2.21. *Let X, Y be multiplicative functionals with finite α-Hölder norm. Then on any interval $[s,t]$ with $|t-s|<1$, we have*

$$d_{p\text{-var};[s,t]}(X,Y) \leq d_{\alpha\text{-Höl};[s,t]}(X,Y)(t-s)^{\alpha}. \tag{3.2.27}$$

We will now define the most important object in the chapter, the p-rough path.

Definition 3.2.22 (**The p-rough path**)**.** Let $p \geq 1$ be a constant. Then a p-rough path in $\mathbb{R}^d$ is a multiplicative functional of degree $\lfloor p \rfloor$ in $\mathbb{R}^d$ with finite p-variation.

For the purposes of this thesis (which primarily focuses on Stratonovich SDEs), we shall consider a special class of rough path that behave like finite variation paths.

Definition 3.2.23 (**Geometric p-rough path**)**.** A p-rough path $\boldsymbol{X}$ is said to be geometric if there exists a sequence of finite variation paths (taking values in $\mathbb{R}^d$) $\{X_n\}_{n\geq 1}$ which converges to $\boldsymbol{X}$ in the p-variation metric on the group $G^{\lfloor p \rfloor}\big(\mathbb{R}^d\big)$, i.e

$$d_{p\text{-var};[0,T]}\left(S_{0,T}^{(\lfloor p \rfloor)}\big(X_n\big), \boldsymbol{X}\right) \to 0, \tag{3.2.28}$$

as $n \to \infty$. We shall denote the space of geometric p-rough paths in $\mathbb{R}^d$ by $G\Omega_p\big(\mathbb{R}^d\big)$.

We now give the main example of a p-rough path, the enhanced Brownian motion.

Theorem 3.2.24 (**Enhanced Brownian motion as a geometric p-rough path**)**.** *Let W be a standard Brownian motion on $[0,1]$ and let $\widehat{W}^n$ denote its piecewise linear discretization on the grid $\mathcal{D}_n = \{t_0 < t_1 < \cdots < t_{2^n}\}$ with $t_k = k2^{-n}$. Then for $p > 2$,*

$$d_{p\text{-var};[0,1]}\left(S_{0,1}^{(2)}\left(\widehat{W}^n\right), \boldsymbol{W}\right) \to 0, \tag{3.2.29}$$

as $n \to \infty$ almost surely, where $\boldsymbol{W}$ is the Stratonovich enhanced Brownian motion:

$$\boldsymbol{W}_{s,t} = \left(1, \int_s^t \circ\, dW_u\,, \int_s^t \int_s^u \circ\, dW_v \otimes \circ\, dW_u\right). \tag{3.2.30}$$

Remark 3.2.25. By applying the Itô-Stratonovich correction to the iterated integrals,

$$\int_s^t \int_s^u dW_v^{(i)} \otimes dW_u^{(j)} = \int_s^t \int_s^u \circ\, dW_v^{(i)} \otimes \circ\, dW_u^{(j)} - \frac{1}{2}(t-s)\delta_{ij}\,,$$

where $i,j \in \{1,\cdots,d\}$, we can construct the **Itô enhanced Brownian motion** as

$$\boldsymbol{W}_{s,t}^{\text{Itô}} = \left(1, \int_s^t dW_u\,, \int_s^t \int_s^u dW_v \otimes dW_u\right). \tag{3.2.31}$$

The Itô enhanced Brownian motion is a p-rough path ($p > 2$) but is not geometric.

In this thesis, we are primarily interested in approximating Stratonovich SDEs. As a result, we will not use the Itô enhanced Brownian motion given by (3.2.31). Instead, we will extend (3.2.29) to a wider class of pathwise approximations of W (see Theorem 6.2.14 in Chapter 6). This application of rough path theory will then lead to sufficient conditions that ensure the convergence of variable step size methods, which can be challenging to study when step sizes depend on the future values of W.

Since the signature of a finite variation path is defined using Riemann-Stieltjes integration, its various integrals can be related via the integration by parts formula. By (3.2.28), it follows that integration by parts will apply to geometric rough paths. These relationships can be encoded using an operation called the "shuffle product". To begin with, we can consider the integration by parts for second iterated integrals.

$$\int_s^t dX_u^i \int_s^t dX_u^j = \int_s^t \int_s^u dX_v^i \, dX_u^j + \int_s^t \int_s^u dX_v^j \, dX_u^i, \tag{3.2.32}$$

for $i, j \in \{1, \cdots, d\}$. In order to place (3.2.32) within a wider algebraic framework, we will first identify the iterated integrals of X with words built from an alphabet A.

Definition 3.2.26. Let $X : [0, T] \to \mathbb{R}^d$ be a finite variation path and let A denote an alphabet consisting of d letters. For a word $w = w_1 \cdots w_n$ with $w_i \in A$, we define

$$\begin{aligned} I : A^* \times \triangle_T &\to \mathbb{R}, \\ I_w(s,t) &:= \int\limits_{s<u_1<\cdots<u_n<t} \cdots \int dX_{u_1}^{w_1} \cdots dX_{u_n}^{w_n}. \end{aligned} \tag{3.2.33}$$

In addition, the map I can be naturally extended to $\mathbb{R}\langle A \rangle \times \triangle_T$ by linearity. That is

$$I_{\lambda_1 u + \lambda_2 v}(s,t) := \lambda_1 I_u(s,t) + \lambda_2 I_v(s,t),$$

for scalars $\lambda_1, \lambda_2 \in \mathbb{R}$ and words $u, v \in A^*$.

Using the notation given above, integration by parts (3.2.32) can be written as

$$\begin{aligned} I_a(s,t) \cdot I_b(s,t) &= I_{ab}(s,t) + I_{ba}(s,t) \\ &= I_{ab+ba}(s,t), \end{aligned}$$

for letters $a, b \in A$. Ideally, we would like to generalize the above to $I_u \cdot I_v = I_{u*v}$, where $*$ is an operation defined on the algebra of noncommutative polynomials $\mathbb{R}\langle A \rangle$. We will see that the operation with this desired property is the shuffle product $\sqcup\!\sqcup$. As $\sqcup\!\sqcup$ encodes the integration by parts formula, we expect it to be defined recursively.

Definition 3.2.27 (**The shuffle product on noncommutative polynomials**). The shuffle product is the unique bilinear map $\sqcup\!\sqcup : \mathbb{R}\langle A\rangle \times \mathbb{R}\langle A\rangle \to \mathbb{R}\langle A\rangle$ satisfying

$$\begin{aligned} ua \sqcup\!\sqcup vb &= (u \sqcup\!\sqcup vb)a + (ua \sqcup\!\sqcup v)b, \quad && \forall a, b \in A,\ \forall u, v \in A^*, \\ u \sqcup\!\sqcup e &= e \sqcup\!\sqcup u = u, && \forall u \in A^*, \end{aligned} \tag{3.2.34}$$

where e denotes the empty word.

The shuffle product between words $u = u_1 \cdots u_n$ and $v = v_1 \cdots v_m$ can be viewed as the sum of all words w constructed from the $n+m$ letters $\{u_1, \cdots, u_n, v_1, \cdots, v_m\}$ such that u can be obtained from w by removing the letters of v and vice versa for v. Hence, each word w can be defined from u and v using a riffle shuffle permutation. We shall give some examples to illustrate this interpretation of the shuffle product.

Example 3.2.28.
$$\begin{aligned} ab \sqcup\!\sqcup xy &= (a \sqcup\!\sqcup xy)b + (ab \sqcup\!\sqcup x)y \\ &= ((e \sqcup\!\sqcup xy)a + (a \sqcup\!\sqcup x)y)b + ((a \sqcup\!\sqcup x)b + (ab \sqcup\!\sqcup e)x)y \\ &= xyab + (ax + xa)yb + (ax + xa)by + abxy \\ &= xyab + axyb + xayb + axby + xaby + abxy. \end{aligned}$$

Example 3.2.29.
$$\begin{aligned} a \sqcup\!\sqcup (w_1 \cdots w_n) &= (e \sqcup\!\sqcup (w_1 \cdots w_n))a + (a \sqcup\!\sqcup (w_1 \cdots w_{n-1}))w_n \\ &= (w_1 \cdots w_n)a + (w_1 \cdots w_{n-1})aw_n \\ &\qquad + (a \sqcup\!\sqcup (w_1 \cdots w_{n-2}))w_{n-1}w_n \\ &\ \ \vdots \\ &= (w_1 \cdots w_n)a + \sum_{k=1}^{n-1} (w_1 \cdots w_k)a(w_{k+1} \cdots w_n) \\ &\qquad + a(w_1 \cdots w_n). \end{aligned}$$

The shuffle product is important in rough path theory due to the below theorem:

Theorem 3.2.30 (**Shuffle product encodes the integration by parts formula**). *Let $X : [0, T] \to \mathbb{R}^d$ be a finite variation path. Then with the above notation, we have*

$$I_u(s, t) \cdot I_v(s, t) = I_{u \sqcup\!\sqcup v}(s, t), \tag{3.2.35}$$

for all $u, v \in \mathbb{R}\langle A\rangle$. In addition, (3.2.35) extends naturally to geometric rough paths.

Since the signature of "space-time" Brownian motion $X_t^W := (t, W_t)$ can be viewed as a geometric rough path, we will be able to use the shuffle product to simplify calculations whenever a product of iterated integrals between W and time appear. This shall be particularly helpful in Chapter 4 (Theorem 4.2.5) where we are interested in finding efficient methods for approximating certain triple iterated integrals of X^W.

3.3 The universal limit theorem and Davie's lemma

In previous sections, we have presented some of the key objects in rough path theory but have yet to connect these objects to the controlled differential equation (3.0.1). Therefore, we shall begin this section by outlining two prominent definitions for the solution of a "rough differential equation" driven by a geometric p-rough path $\boldsymbol{X}$. Whilst RDEs can be defined more generally, the definition due to Davie [26] will suffice for studying Stratonovich SDEs driven by multidimensional Brownian motion.

Definition 3.3.1 (**Davie's definition of RDE solution**). Let $\boldsymbol{X} = \big(1, X^{(1)}, X^{(2)}\big)$ be a p-rough path in $\mathbb{R}^d$ with $p \in [2,3)$ and for each $i \in \{1, \cdots d\}$ let $f_i : \mathbb{R}^e \to \mathbb{R}^e$ denote a continuously differentiable vector field. We use the map $f : \mathbb{R}^e \to \mathcal{L}\big(\mathbb{R}^d, \mathbb{R}^e\big)$ to represent the collection of vector fields, $f(y)x := \sum_i f_i(y)x^i$ for $y \in \mathbb{R}^e$ and $x \in \mathbb{R}^d$. We say that $Y : [0,T] \to \mathbb{R}^e$ is a solution to the rough differential equation (or RDE)

$$\begin{aligned} dY_t &= f(Y_t)\, d\boldsymbol{X}_t, \\ Y_0 &= \xi, \end{aligned} \tag{3.3.1}$$

if there exits a non-negative function θ on $[0,\infty)$ such that $\theta(h) = o(h)$ as $h \to 0$ and

$$\Big\| Y_t - Y_s - f(Y_s)X^{(1)}_{s,t} - Df(Y_s)f(Y_s)X^{(2)}_{s,t} \Big\| \leq \theta(t-s), \tag{3.3.2}$$

for all $s,t \in \triangle_T$, where $Df : \mathbb{R}^e \to \mathcal{L}\big(\mathbb{R}^e, \mathcal{L}\big(\mathbb{R}^d, \mathbb{R}^e\big)\big)$ is the Fréchet derivative of f and $(Df)f : \mathbb{R}^e \to \mathcal{L}\big(\big(\mathbb{R}^d\big)^{\otimes 2}, \mathbb{R}^e\big)$.

In addition, we shall also describe the original approach due to Lyons [70, 72]. That said, we will not give the key details as certain concepts have not be introduced within this chapter (such as almost rough paths and integration along rough paths). As well as applying to SDEs driven by Brownian motion, the Lyon's definition also "works" in the rougher setting (i.e. when $\boldsymbol{X}$ only has finite p-variation for $p \geq 3$).

Remark 3.3.2 (**Lyon's definition of RDE solution**). We will briefly outline the approach taken by Lyons to define the solution of a rough differential equation [70, 72]. As one might expect, the RDE (3.3.1) could be formulated as an integral equation. To do this, we require a theory of rough integration that can give precise meaning to

$$\int_0^{\cdot} \varphi(Z)\, d\boldsymbol{Z}, \tag{3.3.3}$$

where $Z = \pi_1(\boldsymbol{Z})$ and φ is sufficiently smooth (i.e. $\mathrm{Lip}(\gamma)$, as in definition 3.0.3). In particular, when $\boldsymbol{Z} = S^{\lfloor p \rfloor}_{0,T}(Z)$ for some finite variation path $Z : [0,T] \to \mathbb{R}^d$, we

would like (3.3.3) to coincide with the usual Riemann-Stieltjes integral $\int_0^{\cdot} \varphi(Z)\, dZ$. With a notion of rough integration, one can make sense of (3.3.1) by defining $\boldsymbol{Z}$ to be the geometric p-rough path in $\mathbb{R}^d \oplus \mathbb{R}^e$ satisfying the following integral equation:

$$\boldsymbol{Z}_{0,\cdot} = \int_0^{\cdot} \varphi(Z)\, d\boldsymbol{Z}, \tag{3.3.4}$$

where $Z = \pi_1(\boldsymbol{Z}) = (X, Y)$ and $\varphi(Z) = \begin{pmatrix} I_d & 0 \\ f(Y) & 0 \end{pmatrix}$.
The solution Y can then be defined as a projection of Z onto its second component.

With Lyons' definition of RDE solution, it is possible to present one of the main results within rough path theory, the Universal Limit Theorem (Theorem 5.3 in [72]).

Theorem 3.3.3 (**Universal Limit Theorem**). *Let $p \geq 1$ and $f : \mathbb{R}^e \to \mathcal{L}\big(\mathbb{R}^d, \mathbb{R}^e\big)$ be* $\mathrm{Lip}(\gamma)$ *function with $\gamma > p$. Then for all $\boldsymbol{X} \in G\Omega_p\big(\mathbb{R}^d\big)$ and $\xi \in \mathbb{R}^e$, the equation*

$$\begin{aligned} dY_t &= f(Y_t)\, d\boldsymbol{X}_t, \\ Y_0 &= \xi, \end{aligned} \tag{3.3.5}$$

admits a unique solution $\boldsymbol{Z} \in G\Omega_p\big(\mathbb{R}^d \oplus \mathbb{R}^e\big)$ which on a path level is $\pi_1(\boldsymbol{Z}) = (X, Y)$ (as outlined by remark 3.3.2 and given by definition 5.1 in [72]) and the solution map

$$\begin{aligned} \Phi_f : G\Omega_p\big(\mathbb{R}^d\big) \times \mathbb{R}^e &\to G\Omega_p\big(\mathbb{R}^e\big), \\ \big(\boldsymbol{X}, \xi\big) &\mapsto \boldsymbol{Y}, \end{aligned}$$

is the unique extension of the standard solution map (i.e. when X has finite variation) that is continuous in the p-variation topology.

For the purposes of this thesis (where we study Brownian motion and $p \in (2, 3)$), it suffices to consider Davie's definition due to the result below (Theorem 3.3 in [26]):

Theorem 3.3.4 (**RDE solutions when $p < 3$**). *Suppose that $\boldsymbol{X}$ is a geometric p-rough path with $p \in [2, 3)$ and $f : \mathbb{R}^e \to \mathcal{L}\big(\mathbb{R}^d, \mathbb{R}^e\big)$ is a* $\mathrm{Lip}(\gamma)$ *function with $\gamma > p$. Then the Davie and Lyons definitions for the solution of RDE (3.0.1) are equivalent.*

In Chapter 6, the universal limit theorem will enable us to establish pathwise convergence for a certain class of SDE approximation (namely the solutions of ODEs that are driven by piecewise polynomials and constructed using variable step sizes). However in practice, one would not expect a numerical method to exactly solve these "polynomial driven ODEs". On the other hand, it is possible to develop numerical methods that are close to a polynomial-ODE (local errors are $o(h)$, similar to (3.3.2)). To show convergence for this larger, and more practical, class of variable step size approximation, we should quantify how the flow of an RDE propagates local errors.

This analysis will be made possible due to another key result in rough path theory, Davie's lemma (also called Davie's estimate). In addition to showing that an RDE solution depends continuously on its initial condition, vector field and driving path, Davie's lemma tells us that all these dependencies are Lipschitz (in the usual sense). Before introducing these estimates, we will define a norm on the $\mathrm{Lip}(\gamma)$ vector fields.

Definition 3.3.5. For $\gamma \geq 1$, we define the norm of a $\mathrm{Lip}(\gamma)$ function $f : \mathbb{R}^e \to \mathbb{R}^e$,

$$\|f\|_{\mathrm{Lip}(\gamma)} := \max\left\{ \sup_{x,y\in\mathbb{R}^e} \frac{\|D^{\lfloor\gamma\rfloor}f(y) - D^{\lfloor\gamma\rfloor}f(x)\|}{\|x-y\|^{\{\gamma\}}}, \max_{k=0,1,\cdots,\lfloor\gamma\rfloor-1}\left\{ \sup_{y\in\mathbb{R}^e} \|D^k f(y)\| \right\} \right\},$$

where $\{\gamma\} := 1$ for $\gamma \in \mathbb{N}$ and $\{\gamma\} := \gamma - \lfloor\gamma\rfloor$ otherwise.

Using this norm, we shall now present Davie's lemma (Theorem 10.29 from [37]).

Theorem 3.3.6 (**Davie's lemma**). *Suppose that $\gamma > p \geq 1$ and consider the RDEs:*

$$\begin{aligned} dY_t^1 &= f^1\big(Y_t^1\big)\, d\boldsymbol{X}_t^{\boldsymbol{1}}, \\ Y_0^1 &= \xi_1, \\ dY_t^2 &= f^2\big(Y_t^2\big)\, d\boldsymbol{X}_t^{\boldsymbol{2}}, \\ Y_0^2 &= \xi_2, \end{aligned}$$

where $\boldsymbol{X^1}$ and $\boldsymbol{X^2}$ denote geometric p-rough paths in $\mathbb{R}^d$ with $\|\boldsymbol{X^i}\|_{\frac{1}{p}\text{-Höl};[0,T]} \leq 1$, $f^1, f^2 : \mathbb{R}^e \to \mathcal{L}\big(\mathbb{R}^d, \mathbb{R}^e\big)$ are collections of d $\mathrm{Lip}(\gamma)$ vector fields, $\xi_1, \xi_2 \in \mathbb{R}^e$ are the initial conditions for the solutions $Y^1, Y^2 : [0,T] \to \mathbb{R}^e$ and k is an upper bound on the norms $\|f^1\|_{\mathrm{Lip}(\gamma)}$ and $\|f^2\|_{\mathrm{Lip}(\gamma)}$. Then the RDE solutions Y^1, Y^2 are well-defined (as geometric p-rough paths in $\mathbb{R}^e$) and there exists a constant $C(\gamma, p) > 0$ such that

$$\begin{aligned} d_{\frac{1}{p}\text{-Höl};[0,T]}\big(Y^1, Y^2\big) \leq C\Big(k\|Y_0^1 - Y_0^2\| + \|f^1 - f^2\|_{\mathrm{Lip}(\gamma-1)} + k\, d_{\frac{1}{p}\text{-Höl};[0,T]}\big(\boldsymbol{X^1}, \boldsymbol{X^2}\big)\Big) \\ \cdot \exp(Ck^pT). \end{aligned} \tag{3.3.6}$$

Remark 3.3.7. As one might expect, there is a version of Davie's lemma in p-variation:

$$\begin{aligned} d_{p\text{-var};[0,T]}\big(Y^1, Y^2\big) \leq C\big(k\|Y_0^1 - Y_0^2\| + \|f^1 - f^2\|_{\mathrm{Lip}(\gamma-1)} + k\, d_{p\text{-var};[0,T]}\big(\boldsymbol{X^1}, \boldsymbol{X^2}\big)\big) \\ \cdot \exp(Ck^pv^p), \end{aligned} \tag{3.3.7}$$

where v is a bound on $\|\boldsymbol{X^1}\|_{p\text{-var};[0,T]}$ and $\|\boldsymbol{X^2}\|_{p\text{-var};[0,T]}$.

In Chapter 6, we are interested in quantifying the propagation of local errors. Therefore, we only require a Lipschitz property for the RDE flow and will apply Davie's lemma (3.3.6) in the case when $f^1 = f^2$ and $\boldsymbol{X^1} = \boldsymbol{X^2}$ (see Theorem 6.2.18).

3.4 The Taylor expansion and log-ODE method for RDEs

Just as for SDEs, Taylor expansions are an important tool within rough path theory. Moreover, such RDE expansions naturally lead to a canonical ODE approximation known as the log-ODE method, which we study in the SDE setting in Chapter 4. The below expansion (Theorem 1.1 of [9]) also inspired the methods in Chapter 5.

Theorem 3.4.1 (**Rough Taylor expansion**). *Let $\boldsymbol{X}$ be a geometric p-rough path in $\mathbb{R}^d$ with $p \in [2,3)$ and let $f : \mathbb{R}^e \to \mathcal{L}\big(\mathbb{R}^d, \mathbb{R}^e\big)$ denote a $\mathrm{Lip}(\gamma)$ function with $\gamma > p$. Let Y be the unique solution of (3.3.1) defined in either the Davie or Lyons sense. Then Y_t can be expanded as a function of Y_s along with a remainder term as follows:*

$$Y_t = Y_s + \sum_{k=1}^{\lfloor\gamma\rfloor} \sum_{i_1,\cdots,i_k \in \{1,\cdots,d\}} f_{i_1} \cdots f_{i_k} Id\big(Y_s\big) \int\cdots\int_{s<u_1<\cdots<u_k<t} dX_{u_1}^{i_1} \cdots dX_{u_k}^{i_k} \tag{3.4.1}$$
$$+ R_{\lfloor\gamma\rfloor}(Y_s, \boldsymbol{X}, [s,t]).$$

where

$$R_N(Y_s, \boldsymbol{X}, [s,t]) \tag{3.4.2}$$
$$= \sum_{i_1,\cdots,i_N \in \{1,\cdots,d\}} \int\cdots\int_{s<u_1<\cdots<u_N<t} f_{i_1} \cdots f_{i_N} Id\big(Y_{u_1}\big) - f_{i_1} \cdots f_{i_N} Id\big(Y_s\big)\, dX_{u_1}^{i_1} \cdots dX_{u_N}^{i_N},$$

with the differential operators $f_{i_1} \cdots f_{i_N}$ defined by (2.3.3) from the previous chapter. In addition, there exists a positive constant $C_p > 0$, that depends only on p, such that

$$\big\|R_N(Y_s, \boldsymbol{X}, [s,t])\big\| \le \frac{1}{\big(\frac{\lfloor\gamma\rfloor}{p}\big)!} \beta^{\lfloor\gamma\rfloor} M_{p,\gamma} \|f\|_{\circ\gamma} \|\boldsymbol{X}\|_{p\text{-var};[s,t]}^{\gamma}, \tag{3.4.3}$$

where we use the gamma function Γ to define the factorial $\big(\frac{\lfloor\gamma\rfloor}{p}\big)! := \Gamma\big(\frac{\lfloor\gamma\rfloor}{p} + 1\big)$ and

$$M_{p,\gamma} := 2C_p\big(\|f\|_{\mathrm{Lip}((\gamma-1)\wedge\lfloor p\rfloor)} \vee 1\big)^{\lfloor p\rfloor+1} \big(\|\boldsymbol{X}\|_{p\text{-var};[s,t]} \vee 1\big)^{\lfloor p\rfloor+1},$$
$$\|f\|_{\circ\gamma} := \max_{\lfloor\gamma\rfloor-\lfloor p\rfloor+1 \le k \le \lfloor\gamma\rfloor} \|f^{\circ k}\|_{\mathrm{Lip}(\min(\gamma-k,1))}^{\min(\gamma-m,1)},$$
$$\beta := p\left(1 + \sum_{r=2}^{\infty} \left(\frac{2}{r-1} \wedge 1\right)^{\frac{\lfloor p\rfloor+1}{p}}\right),$$

and the derivatives $f^{\circ k} : \mathbb{R}^e \to \mathcal{L}\big((\mathbb{R}^d)^{\otimes k}, \mathbb{R}^e\big)$ are defined inductively for $k \ge 1$ by

$$f^{\circ 1}(y) := f(y),$$
$$f^{\circ(k+1)}(y) := D\big(f^{\circ k}\big)(y) f(y),$$

for $y \in \mathbb{R}^e$.

Remark 3.4.2. It is sometimes more convenient to write (3.4.1) in the following form:

$$Y_t = Y_s + \sum_{k=1}^{\lfloor\gamma\rfloor} f^{\circ k}(Y_s)\boldsymbol{X}_{s,t}^{(k)} + R_{\lfloor\gamma\rfloor}(Y_s, \boldsymbol{X}, [s,t]), \tag{3.4.4}$$

where

$$\boldsymbol{X}_{s,t}^{(k)} := \int\cdots\int_{s<u_1<\cdots<u_k<t} dX_{u_1}\otimes\cdots\otimes dX_{u_k} \in \big(\mathbb{R}^d\big)^{\otimes k},$$

for $k \geq 1$.

Whilst there is clearly a very detailed analysis required to prove Theorem 3.4.1, we can make a simple observation concerning the (truncated) rough Taylor expansion

$$Y_s + \sum_{k=1}^{\lfloor\gamma\rfloor} f^{\circ k}(Y_s)\boldsymbol{X}_{s,t}^{(k)}. \tag{3.4.5}$$

Namely that (3.4.5) is expressible as a linear functional on the signature $S_{s,t}^{(\lfloor\gamma\rfloor)}\big(\boldsymbol{X}\big)$. Thus, we define a homomorphism from $T\big(\mathbb{R}^d\big)$ into the space of differential operators.

Theorem 3.4.3. *Let $\{f_1,\cdots,f_d\}$ denote a collection of smooth vector fields on $\mathbb{R}^e$ and let $\{f^1,\cdots,f^d\}$ denote the corresponding differential operators given by (2.3.2). Then there exists a unique algebra homomorphism Φ from the tensor algebra $T\big(\mathbb{R}^d\big)$ (endowed with the tensor product $\otimes$) into the space of differential operators (which is an algebra under the composition operation given by (2.3.3)) such that $\Phi(\mathbf{1}) = Id$ and*

$$\Phi(e_i) = f^i, \tag{3.4.6}$$

for basis vectors $\{e_i\}$ of $\mathbb{R}^d$ (viewed as elements of $T\big(\mathbb{R}^d\big)$ using the inclusion map).

Remark 3.4.4. Using this homomorphism, the Taylor expansion (3.4.5) can be written

$$\Phi\big(S_{s,t}^{(\lfloor\gamma\rfloor)}\big(\boldsymbol{X}\big)\big)\mathrm{Id}(Y_s) = Y_s + \sum_{k=1}^{\lfloor\gamma\rfloor} f^{\circ k}(Y_s)\boldsymbol{X}_{s,t}^{(k)}. \tag{3.4.7}$$

Since $S_{s,t}^{(\lfloor\gamma\rfloor)}\big(\boldsymbol{X}\big) \in G^{\lfloor\gamma\rfloor}\big(\mathbb{R}^d\big)$, we can take the above remark further using exp and log.

$$\begin{aligned}
\Phi\big(S_{s,t}^{(\lfloor\gamma\rfloor)}\big(\boldsymbol{X}\big)\big) &= \Phi\big(\exp^{(\lfloor\gamma\rfloor)}\log^{(\lfloor\gamma\rfloor)} S_{s,t}^{(\lfloor\gamma\rfloor)}\big(\boldsymbol{X}\big)\big)\\
&= \Phi\left(\mathbf{1} + \sum_{k=1}^{\lfloor\gamma\rfloor}\frac{1}{k!}\Big(\log^{(\lfloor\gamma\rfloor)} S_{s,t}^{(\lfloor\gamma\rfloor)}\big(\boldsymbol{X}\big)\Big)^{\otimes k}\right)\\
&= \mathrm{Id} + \sum_{k=1}^{\lfloor\gamma\rfloor}\frac{1}{k!}\Big(\Phi\Big(\log^{(\lfloor\gamma\rfloor)} S_{s,t}^{(\lfloor\gamma\rfloor)}\big(\boldsymbol{X}\big)\Big)\Big)^{k}.
\end{aligned}$$

Thus, we can view a Taylor expansion as an "exponential" of a differential operator.

$$Y_s + \sum_{k=1}^{\lfloor\gamma\rfloor} f^{\circ k}(Y_s)\boldsymbol{X}_{s,t}^{(k)} = \exp^{(\lfloor\gamma\rfloor)}\Phi\big(\log S_{s,t}^{(\lfloor\gamma\rfloor)}\big(\boldsymbol{X}\big)\big)\mathrm{Id}(Y_s), \tag{3.4.8}$$

where $\log S_{s,t}^{(\lfloor\gamma\rfloor)}\big(\boldsymbol{X}\big)$ is the $\lfloor\gamma\rfloor$-step log-signature of the geometric p-rough path $\boldsymbol{X}$.

Since the log-signature of $\boldsymbol{X}$ is an element of the $\lfloor\gamma\rfloor$-step free Lie algebra $\mathfrak{g}^{\lfloor\gamma\rfloor}\big(\mathbb{R}^d\big)$, it can be expressed uniquely with respect to a basis $\mathcal{B}$ of $\mathfrak{g}^{\lfloor\gamma\rfloor}\big(\mathbb{R}^d\big)$. Thus, we will write

$$\log S_{s,t}^{(\lfloor\gamma\rfloor)}\big(\boldsymbol{X}\big) = \sum_{\ell\in\mathcal{B}} L_{s,t}^{\ell}\,\ell. \tag{3.4.9}$$

The next key observation is that $\Phi(\ell)$ is a <u>first order</u> differential operator for $\ell \in \mathcal{B}$.

Theorem 3.4.5. *Let $\boldsymbol{a}, \boldsymbol{b} \in T\big(\mathbb{R}^d\big)$ give first order differential operators $\Phi(\boldsymbol{a}), \Phi(\boldsymbol{b})$. Then $\Phi\big([\boldsymbol{a}, \boldsymbol{b}]\big)$ is also first order (and hence corresponds to a vector field by (2.3.2)).*

Proof. The result follows from the definition of vector field Lie bracket, (2.3.4). □

Therefore $\Phi\big(\log S_{s,t}^{(\lfloor\gamma\rfloor)}\big(\boldsymbol{X}\big)\big)\mathrm{Id}(\cdot)$ can be viewed as the following vector field on $\mathbb{R}^e$:

$$F(\cdot) := \sum_{\ell\in\mathcal{B}} L_{s,t}^{\ell}\,\Phi(\ell)\mathrm{Id}(\cdot). \tag{3.4.10}$$

As vector field Lie brackets do not depend on a choice of basis, F is uniquely determined by the RDE vector fields $\{f_1, \cdots, f_d\}$ and the geometric rough path $\boldsymbol{X}$. Hence by (3.4.8), a natural approximation of Y_t would be the exponential of F at Y_s. These observations motivate the so-called "log-ODE method" for RDEs [57, 71, 10].

Definition 3.4.6 (**The log-ODE method**)**.** Consider $\boldsymbol{X}$, f, Y from Theorem 3.4.1. Then, given Y_s, we can approximate Y_t by the solution at $u = 1$ of the below ODE:

$$\begin{aligned} \frac{dY^{s,t}}{du} &= F\big(Y_u^{s,t}\big), \\ Y_0^{s,t} &= Y_s, \end{aligned} \tag{3.4.11}$$

where $F : \mathbb{R}^e \to \mathbb{R}^e$ is the vector field (3.4.10) defined using the log-signature of $\boldsymbol{X}$.

Theorem 3.4.7 (Lemma 15 in [10])**.** *There exists a positive constant $C_{p,\gamma}$ such that*

$$\|Y_t - Y_1^{s,t}\| \le C_{p,\gamma}\|f\|_{\mathrm{Lip}(\gamma)}^{\gamma}\|\boldsymbol{X}\|_{p\text{-var};[s,t]}^{\gamma}. \tag{3.4.12}$$

Remark 3.4.8. The solution at $u = 1$ of the log-ODE (3.4.11) can be viewed as the solution of an RDE driven by the geodesic path $\boldsymbol{Z}$ with the defining property that $S_{s,t}\big(\boldsymbol{Z}\big) = \exp\big(\log S_{s,t}^{(\lfloor\gamma\rfloor)}\big(\boldsymbol{X}\big)\big)$. The existence of this path follows by Chow's theorem. This makes the log-ODE method well-suited for approximating RDEs on manifolds.

Chapter 4

A polynomial approximation of Brownian motion and applications

In this chapter, we will introduce a strong (or pathwise) approximation of standard Brownian motion by polynomials. As Brownian motion is continuous almost surely, the Weierstrass approximation theorem [67] shows that polynomials are a sufficiently rich class of functions for approximating Brownian sample paths on closed intervals. This was studied in [63], where random Bernstein polynomials were constructed by interpolating a finite number of points along the Brownian motion. Although this method converges almost surely, it is not suitable for numerical modelling due to the large computational cost needed to evaluate Bernstein polynomials with high degree.

In [48], the authors were motivated by the idea of Lévy's construction which allows one to discretize Brownian motion as a piecewise linear function using Haar wavelets. In the paper, they show that this can be extended by discretizing Brownian motion as a piecewise polynomial using Alpert-Rokhlin multiwavelets (see [8]). However, explicit formulae is only derived in the piecewise linear and piecewise parabola cases. Since we expect these wavelets to be more practical than the Bernstein polynomials, it is the approach of [48] that leads to the main result (Theorem 4.1.3) of this chapter.

In the first section, we will prove a "polynomial version" of the Karhunen-Loève theorem for the Brownian bridge. This enables us to express Brownian motion as a weighted sum of orthogonal polynomials where the weights are independent Gaussian random variables. We shall then explore the relationship between these orthogonal polynomials and the Alpert-Rokhlin wavelets. In the second section, the polynomial approximation will be used for studying certain triple iterated integrals of Brownian motion and time. In particular, we shall derive new optimal estimators for these third order integrals and compute their respective $L^2(\mathbb{P})$ approximation errors. These new

estimators can be directly incorporated into high order numerical methods for SDEs. This will be demonstrated in the final sections where we develop a Runge-Kutta method for SDEs with multidimensional additive noise as well as a state-of-the-art discretization of Inhomogeneous Geometric Brownian Motion (IGBM). The majority of the results that are presented in this chapter have already been published in [35].

4.1 A polynomial based Karhunen-Loève theorem for the Brownian bridge

Before we state and prove the main theorem of this chapter, we shall review the approach detailed in [48]. In the paper, Brownian motion is generated in terms of Alpert-Rokhlin multiwavelets (see [8]), which are defined as piecewise polynomials.

Definition 4.1.1 (Alpert-Rokhlin wavelets)**.** For $q \geq 1$, define the q functions $\phi^{q,1}, \cdots, \phi^{q,q} : [0,1] \to \mathbb{R}$ as piecewise polynomials of degree $q-1$ with pieces on $[0, \frac{1}{2}]$, $[\frac{1}{2}, 1]$ and which satisfies the following conditions for $1 \leq p \leq q$ and $t \in [0, \frac{1}{2})$:

$$\phi^{q,p}(t) = (-1)^{q+p-1} \phi^{q,p}(1-t), \tag{4.1.1}$$

$$\int_0^1 \phi^{q,p}(t)\phi^{q,r}(t)\, dt = \delta_{qr}, \quad \text{for } 1 \leq r \leq q, \tag{4.1.2}$$

$$\int_0^1 t^k \phi^{q,p}(t)\, dt = 0, \quad \text{for } 0 \leq k \leq q-1. \tag{4.1.3}$$

The Alpert-Rokhlin multiwavelets of order q can now be generated by translating and scaling the mother functions $\phi^{q,p}$.

$$\phi^{q,p}_{nk}(t) := \frac{1}{\sqrt{2^n}}\, \phi^{q,p}(2^n t - k), \tag{4.1.4}$$

for $n \geq 0$ and $k \in \{0, \cdots, 2^n - 1\}$.

Brownian motion can then be represented using the integrals of the above wavelets,

$$h^{q,p}_{00}(t) := \int_0^t \phi^{q,p}(s)\, ds,$$

with $h^{q,p}_{nk}$ defined by translating and scaling $h^{q,p}_{00}$ in precisely the same way as (4.1.4),

$$h^{q,p}_{nk}(t) := \frac{1}{\sqrt{2^n}}\, h^{q,p}_{nk}(2^n t - k).$$

For example, the following representation of Brownian motion was established in [48].

Theorem 4.1.2 (Brownian motion expressed in terms of quadratic polynomials)**.**

$$W(t) = a_0 t + a_1\sqrt{3}t(1-t) + \sum_{n=0}^{\infty}\sum_{k=0}^{2^n-1}\sum_{p=1}^{2} a_{nk}^{(p)} h_{nk}^{2,p}(t), \quad \textit{for } t \in [0,1], \tag{4.1.5}$$

where the weights a_0, a_1, $a_{nk}^{(p)}$ *are independent standard normal random variables and*

$$h_{00}^{2,1}(t) = \begin{cases} \sqrt{3}\,t(1-2t), & \textit{for } \ t \in \left[0, \frac{1}{2}\right] \\ \sqrt{3}(1-t)(1-2t), & \textit{for } \ t \in \left[\frac{1}{2}, 1\right] \end{cases},$$

$$h_{00}^{2,2}(t) = \begin{cases} t(3t-1), & \textit{for } \ t \in \left[0, \frac{1}{2}\right] \\ (t-1)(3t-2), & \textit{for } \ t \in \left[\frac{1}{2}, 1\right] \end{cases}.$$

It is also highlighted in [48] that the random coefficient a_1 is the integral of the basis function $\sqrt{3}(2t-1)$ with respect to the Brownian path. Since a_0 corresponds to the endpoint W_1, it is straightforward to produce a graph comparing a finely discretized Brownian sample path against the approximation we obtain from truncating (4.1.5).

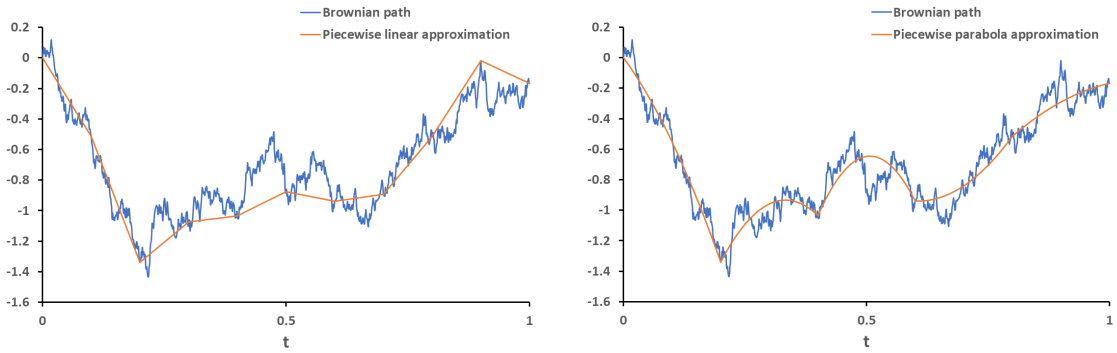

Figure 4.1: Brownian motion approximated by piecewise linear and parabola paths. Both discretizations require ten independent Gaussian random variables to generate.

The above figure seems to indicate that the piecewise parabola may be a more effective approximant for Brownian motion than the standard piecewise linear path. As we shall see, the reason for this is that parabolas can capture the "space-time" area of a path. This fact will be exploited in one of our numerical methods for SDEs.

The main result of this chapter provides a representation of Brownian motion in terms of Jacobi-like polynomials. The proof follows the standard argument for the Karhunen-Loève theorem, but relies on the new discovery that a class of orthogonal polynomials are eigenfunctions of an integral operator defined from the Brownian bridge covariance function. After establishing a defining property for these basis

functions, we shall give an explicit formula for them in terms of the well-studied Jacobi polynomials. In addition, these polynomials can then be used to interpret the geometrical features that certain iterated integrals encode about the Brownian path.

Theorem 4.1.3 (A polynomial Karhunen-Loève theorem for the Brownian bridge)**.** *Let B denote a Brownian bridge on $[0,1]$ and consider the Borel measure μ given by*

$$\mu(a,b) := \int_a^b \frac{1}{x(1-x)}\,dx, \quad \textit{for all open intervals } (a,b) \subset [0,1].$$

Then there exists a family of orthogonal polynomials $\{e_k\}_{k\geq 1}$ with $\deg(e_k) = k+1$ and

$$\int_0^1 e_i\, e_j\, d\mu = \delta_{ij},$$

with δ_{ij} denoting the Kronecker delta, such that B admits the following representation

$$B = \sum_{k=1}^{\infty} I_k e_k, \tag{4.1.6}$$

where $\{I_k\}$ is the collection of independent centered Gaussian random variables with

$$I_k := \int_0^1 B_t \cdot \frac{e_k(t)}{t(1-t)}\,dt, \tag{4.1.7}$$

and

$$\mathrm{Var}(I_k) = \frac{1}{k(k+1)}.$$

Moreover, $\{e_k\}$ is an optimal orthonormal basis of $L^2([0,1],\mu)$ for approximating B using truncated series expansions with respect to the following weighted $L^2(\mathbb{P})$ norm

$$\|X\|_{L^2_\mu(\mathbb{P})} := \sqrt{\mathbb{E}\left[\int_0^1 (X_s)^2\, d\mu(s)\right]},$$

where X is a square μ-integrable process.

Proof. Our argument is that of the Karhunen-Loève theorem in general L^2 spaces. Note that B is a square μ-integrable process as

$$\mathbb{E}\left[\int_0^1 (B_s)^2\, d\mu(s)\right] = \int_0^1 \mathbb{E}\big[(B_s)^2\big]\, d\mu(s) = \int_0^1 s(1-s)\cdot\frac{1}{s(1-s)}\,ds = 1 < \infty.$$

Let K_B denote the covariance function for the standard Brownian bridge on $[0,1]$. One can show by direct calculation that

$$\|K_B\|^2_{L^2([0,1]^2,\,\mu^2)} = \int_0^1\int_0^1 (\min(s,t)-st)^2\, d\mu(s)\, d\mu(t) = \frac{1}{3}\pi^2 - 3 < \infty.$$

Hence, it follows that the integral operator $T_K : L^2([0,1],\, \mu) \to L^2([0,1],\, \mu)$ given by

$$(T_K f)(t) := \int_0^1 K_B(s,t) f(s)\, d\mu(s),$$

is well-defined and continuous.

Furthermore, the function k_B on $[0,1]$ defined as $k_B(x) := K_B(x,x)$ is μ-integrable as

$$\int_0^1 |k_B(x)|\, d\mu(x) = \int_0^1 x(1-x) \cdot \frac{1}{x(1-x)}\, dx = 1 < \infty.$$

It is now possible to apply Mercer's theorem for kernels on general L^2 spaces (see [18]). It then follows from Mercer's theorem that there exists an orthonormal set $\{e_k\}_{k\geq 1}$ of $L^2([0,1],\, \mu)$ consisting of eigenfunctions of T_K such that the corresponding sequence of eigenvalues $\{\lambda_k\}_{k\geq 1}$ is non-negative. Moreover, the eigenfunctions corresponding to non-zero eigenvalues are continuous on $[0,1]$ and the kernel K_B has the representation

$$K_B(s,t) = \sum_{k=1}^{\infty} \lambda_k e_k(s) e_k(t), \tag{4.1.8}$$

where the series converges absolutely and uniformly on the compact subsets of $[0,1]$. In the next part of the proof, we'll show that each e_k is a polynomial of degree $k+1$. As each e_k is an eigenfunction of T_K, we have

$$\int_0^1 \frac{\min(s,t) - st}{s(1-s)}\, e_k(s)\, ds = \lambda_k e_k(t). \tag{4.1.9}$$

Since $e_k \in L^2([0,1],\, \mu)$, it follows that $e_k(0) = 0$ and $e_k(1) = 0$ for each $k \geq 1$. Therefore by using the Leibniz integral rule to twice differentiate both sides of (4.1.9) and then multiplying by $t(1-t)$, we observe that e_k satisfies the differential equation,

$$t(1-t)\lambda_k e_k''(t) + e_k(t) = 0. \tag{4.1.10}$$

Since $e_k \neq 0$, we have that $\lambda_k \neq 0$. Thus we can write the above in the standard form

$$e_k''(t) + \frac{1}{t(1-t)} \lambda_k^{-1} e_k(t) = 0. \tag{4.1.11}$$

Since the function $t^{-1}(1-t)^{-1}$ only has order 1 poles, the above ODE will have regular singular points at 0 and 1. Therefore we can apply the Frobenius method to (4.1.11) (see Chapter 4 of [90]). That is, we write e_k using it power series expansion at zero,

$$e_k(t) = \sum_{n=1}^{\infty} a_n^{(k)} t^n, \tag{4.1.12}$$

and then substitute it into the second order ODE (4.1.10). This gives the following:

$$\sum_{n=0}^{\infty}(n+1)(n+2)a_{n+2}^{(k)}t^{n+1}-\sum_{n=0}^{\infty}(n+1)(n+2)a_{n+2}^{(k)}t^{n+2}+\lambda_k^{-1}\sum_{n=0}^{\infty}a_{n+1}^{(k)}t^{n+1}=0.$$

Thus, by equating the coefficients of each monomial, we have

$$n(n+1)a_{n+1}^{(k)}=\big(n(n-1)-\lambda_k^{-1}\big)a_n^{(k)}, \tag{4.1.13}$$

for $n\geq 1$ and so each coefficient is given by

$$a_n^{(k)}=\left(\prod_{i=1}^{n-1}\frac{i(i-1)-\lambda_k^{-1}}{i(i+1)}\right)a_1^{(k)}.$$

We assume for a contradiction that λ_k^{-1} is not of the form $i(i-1)$ for some $i>1$. Then $\lim\limits_{n\to\infty}a_n^{(k)}$ exists and is non-zero when $a_1^{(k)}\neq 0$. So by the limit comparison test, we see that for each $t\in[0,1]$,

$$\sum_{n=1}^{\infty}a_n^{(k)}t^n \text{ converges if and only if } \sum_{n=1}^{\infty}t^n \text{ converges.}$$

Thus the series (4.1.12) does not converge at 1 which gives a contraction as $e_k(1)=0$. Hence $\lambda_k^{-1}=i(i+1)$ for some $i>1$. Without loss of generality, we can let $\{\lambda_k^{-1}\}_{k\geq 1}$ be an increasing sequence with $\lambda_k^{-1}=k(k+1)$. Then by the series expansion (4.1.12) and the recurrence relation (4.1.13), we see that e_k is a polynomial with degree $k+1$.

Differentiating both sides of ODE (4.1.10) yields

$$t(1-t)\frac{d^2}{dt^2}(e_k')+(1-2t)\frac{d}{dt}(e_k')+k(k+1)e_k'(t)=0.$$

For $x\in[-1,1]$, we define the function

$$y_k(x):=e_k'\left(\frac{1}{2}(1+x)\right).$$

Thus y_k satisfies the following differential equation

$$(1-x^2)y_k''(x)-2xy_k'(x)+k(k+1)y_k(x)=0. \tag{4.1.14}$$

Remarkably, this is the Legendre differential equation [31] and so y_k is proportional the k-th Legendre polynomial (as we have already established that e_k' is a polynomial). Hence e_k is a constant multiple of the integral of the k-th shifted Legendre polynomial.

We now define the following integrals for $k \geq 1$,

$$I_k := \int_0^1 B_t \cdot \frac{e_k(t)}{t(1-t)}\, dt.$$

It follows from Fubini's theorem that

$$\begin{aligned}
\mathbb{E}[I_k] &= 0, \\
\mathbb{E}[I_i I_j] &= \mathbb{E}\left[\int_0^1 \int_0^1 B_s B_t\, e_i(s)\, e_j(t)\, d\mu(s)\, d\mu(t)\right] \\
&= \int_0^1 \int_0^1 \mathbb{E}[B_s B_t]\, e_i(s)\, e_j(t)\, d\mu(s)\, d\mu(t) \\
&= \int_0^1 e_j(t) \left(\int_0^1 K_B(s,t)\, e_i(s)\, d\mu(s)\right) d\mu(t) \\
&= \lambda_i \delta_{ij}.
\end{aligned}$$

Since each I_k is defined by a linear functional on the same Gaussian process B, we see from the above that $\{I_k\}$ is a collection of uncorrelated (and therefore independent) Gaussian random variables with

$$\begin{aligned}
\mathbb{E}[I_k] &= 0, \\
\mathrm{Var}(I_k) &= \frac{1}{k(k+1)}.
\end{aligned}$$

Finally, the $L^2(\mathbb{P})$ convergence we require follows as

$$\begin{aligned}
\mathbb{E}\left[\left(B_t - \sum_{k=1}^N I_k e_k(t)\right)^2\right] &= k_B(t) + \mathbb{E}\left[\sum_{i,j=1}^N I_i I_j e_i(t) e_j(t)\right] - 2\mathbb{E}\left[B_t \sum_{k=1}^N I_k e_k(t)\right] \\
&= k_B(t) + \sum_{i=1}^N \lambda_i e_i^2(t) - 2\mathbb{E}\left[\sum_{i=1}^N \int_0^1 B_s B_t\, e_i(s) e_i(t)\, d\mu(s)\right] \\
&= k_B(t) - \sum_{i=1}^N \lambda_i e_i^2(t),
\end{aligned}$$

which converges to 0 by Mercer's theorem (4.1.8).

All that remains is to prove optimality for the truncated series expansions of (4.1.6). Let $\{f_k\}_{k\geq 1}$ denote an orthonormal basis of $L^2([0,1],\mu)$ such that

$$B = \sum_{k=1}^\infty J_k f_k, \quad \text{where} \quad J_k := \int_0^1 B_t\, f_k(t)\, d\mu(t), \quad \forall k \geq 1.$$

For each $n \geq 1$, we will consider the associated error process $r_n := \sum_{k=n+1}^\infty J_k f_k$.

Then the square $L^2(\mathbb{P})$ norm of the error process admits the following expansion,

$$\begin{aligned}\|r_n(t)\|^2_{L^2(\mathbb{P})} &= \mathbb{E}\left[\sum_{i=n+1}^{\infty}\sum_{j=n+1}^{\infty} J_i J_j f_i(t) f_j(t)\right] \\ &= \sum_{i=n+1}^{\infty}\sum_{j=n+1}^{\infty} \mathbb{E}\left[\int_0^1\int_0^1 B_s B_t\, f_i(s) f_j(t)\, d\mu(s)\, d\mu(t)\right] f_i(t) f_j(t) \\ &= \sum_{i=n+1}^{\infty}\sum_{j=n+1}^{\infty} \left(\int_0^1\int_0^1 K_B(s,t)\, f_i(s) f_j(t)\, d\mu(s)\, d\mu(t)\right) f_i(t) f_j(t).\end{aligned}$$

Integrating the above with respect to μ and using the orthogonality of $\{f_k\}$ gives

$$\|r_n\|^2_{L^2_\mu(\mathbb{P})} = \int_0^1 \|r(t)\|^2_{L^2(\mathbb{P})}\, d\mu(t) = \sum_{k=n+1}^{\infty}\int_0^1\int_0^1 K_B(s,t) f_k(s) f_k(t)\, d\mu(s)\, d\mu(t).$$

Note that an optimal orthonormal basis of $L^2([0,1],\mu)$ solves the following problem:

$$\min_{f_k} \|r_n\|^2_{L^2_\mu(\mathbb{P})} \quad \text{subject to} \quad \|f_k\|_{L^2([0,1],\mu)} = 1.$$

By introducing Lagrange multipliers ν_k, we wish to find functions $\{f_k\}$ that minimize

$$E_n[\{f_k\}] := \sum_{k=n+1}^{\infty}\int_0^1\int_0^1 K_B(s,t) f_k(s) f_k(t)\, d\mu(s) d\mu(t) - \nu_k\left(\int_0^1 \big(f_k(s)\big)^2\, d\mu(s) - 1\right).$$

We will now consider the following square integrable functions, defined for $s,t \in (0,1)$:

$$\tilde{f}_k(t) := f_k(t)\cdot\frac{1}{\sqrt{t(1-t)}}, \qquad \tilde{K}_B(s,t) := K_B(s,t)\cdot\frac{1}{\sqrt{s(1-s)}}\cdot\frac{1}{\sqrt{t(1-t)}}.$$

Therefore it is enough to find a family of functions $\{\tilde{f}_k\}$ in $L^2([0,1])$ that minimizes

$$\tilde{E}_n[\{\tilde{f}_k\}] := \sum_{k=n+1}^{\infty}\int_0^1\int_0^1 \tilde{K}_B(s,t)\tilde{f}_k(s)\tilde{f}_k(t)\, ds\, dt - \nu_k\left(\int_0^1 \big(\tilde{f}_k(s)\big)^2\, ds - 1\right).$$

To find a minimizer, we set the functional derivative of $\tilde{E}_n$ with respect to $\tilde{f}_k$ to zero.

$$\frac{\partial \tilde{E}_n}{\partial \tilde{f}_k(t)} = 2\int_0^1 \tilde{K}_B(s,t)\tilde{f}_k(s)\, ds - 2\nu_k \tilde{f}_k(t) = 0.$$

By using the definitions of $\tilde{f}_k$ and $\tilde{K}_B$, it is trivial to show the above is equivalent to

$$\int_0^1 K_B(s,t) f_k(s)\, d\mu(s) = \nu_k f_k(t),$$

which is satisfied if and only if f_k are eigenfunctions of T_K. □

As Brownian motion is expressible as the sum of its straight line approximation and an independent Brownian bridge, the above theorem can be naturally extended.

Corollary 4.1.4 (A polynomial decomposition of Brownian motion)**.** *Let W be a standard Brownian motion on $[0,1]$ and let B denote the associated bridge process. Then by Theorem 4.1.3, we have the following series representation for W:*

$$W = W_1 e_0 + \sum_{k=1}^{\infty} I_k e_k, \tag{4.1.15}$$

where $e_0(t) := t$ for $t \in [0,1]$, and the random variables $\{I_k\}$ are independent of W_1.

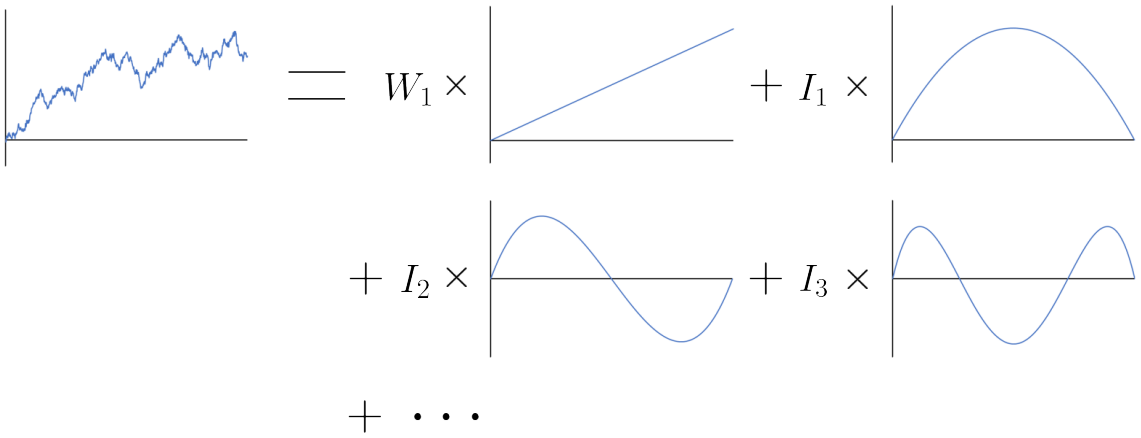

Figure 4.2: Brownian motion as a sum of polynomials with independent weights. The polynomials are orthogonal and capture different time integrals of the path.

Since $e_k \in L^2([0,1],\mu)$ for each $k \geq 1$, it follows that e_k has roots at the endpoints, $e_k(0) = e_k(1) = 0$. This implies that $e_k \cdot \frac{1}{t(1-t)}$ is a polynomial but with degree $k-1$. Thus we can use integration by parts to express the stochastic integrals $\{I_k\}$ defined in Theorem 4.1.6 in terms of the following iterated integrals of the Brownian bridge:

$$\int_{0<s_1<1} B_{s_1}\, ds_1, \int_{0<s_1<s_2<1} B_{s_1}\, ds_1\, ds_2, \cdots, \int_{0<s_1<\cdots<s_n<1} B_{s_1}\, ds_1 \cdots ds_n, \cdots$$

These are special coordinate integrals that appear in the signature of a space-time Brownian bridge $X_t^B := (t, B_t)$. Theorem 4.1.3 will help us to interpret these integrals and what they encode about the trajectory of the Brownian path. In essence, the first n iterated time integrals describe a polynomial approximant with degree $n+1$. In order to establish this general idea as a theorem, we require the following lemma.

Lemma 4.1.5. *For each $n \geq 1$, there exists a lower triangular $n \times n$ matrix M_n with non-zero diagonal entries such that*

$$\begin{pmatrix} I_1 \\ \vdots \\ I_n \end{pmatrix} = M_n \begin{pmatrix} \int_{0<s_1<1} B_{s_1}\, ds_1 \\ \vdots \\ \int_{0<s_1<\cdots<s_n<1} B_{s_1}\, ds_1 \cdots ds_n \end{pmatrix}. \tag{4.1.16}$$

Proof. It has already been shown that $e_k \cdot \frac{1}{t(1-t)}$ is a polynomial with degree $k-1$. Hence we can apply the integration by parts formula $k-1$ times to express I_k as a linear combination of integrals in $\left\{\int_{0<s_1<1} B_{s_1}\, ds_1, \cdots, \int_{0<s_1<\cdots<s_k<1} B_{s_1}\, ds_1 \cdots ds_k\right\}$. Importantly, the coefficient that appears in front of $\int_{0<s_1<\cdots<s_k<1} B_{s_1}\, ds_1 \cdots ds_k$ is non-zero as the $(k-1)$-th derivative of $e_k \cdot \frac{1}{t(1-t)}$ is non-zero. □

Since M_n is a lower triangular matrix with a non-zero diagonal, it is invertible. Therefore both of the column vectors in (4.1.16) carry precisely the same information about the Brownian bridge. As a result, we can establish the following connection between Brownian motion, polynomials and certain coordinate integrals appearing in the signature of $X_t^W := (t, W_t)$.

Theorem 4.1.6. *Consider the following unbiased estimator of a one-dimensional Brownian motion over the interval $[0,1]$,*

$$W_t^n := \mathbb{E}\left[W_t \,\middle|\, W_1, \int_{0<s_1<1} W_{s_1}\, ds_1, \cdots, \int_{0<s_1<\cdots<s_{n-1}<1} W_{s_1}\, ds_1 \cdots ds_{n-1}\right]. \tag{4.1.17}$$

Then W^n is the unique polynomial of degree n with a root at 0 such that $W_1^n = W_1$ and $\int_{0<s_1<\cdots<s_k<1} W_{s_1}^n\, ds_1 \cdots ds_k = \int_{0<s_1<\cdots<s_k<1} W_{s_1}\, ds_1 \cdots ds_k$ for $1 \leq k \leq n-1$.

Proof. It is a direct consequence of lemma 4.1.5 that $W_t^n = \mathbb{E}[W_t \,|\, W_1, I_1, \cdots, I_{n-1}]$. Therefore by Corollary 4.1.4 and independence of the random variables $\{W_1, I_1, \cdots\}$, we have that

$$W^n = W_1 e_0 + \sum_{k=1}^{n-1} I_k e_k. \tag{4.1.18}$$

Hence W^n is a polynomial of degree n with a root at 0 and has the same increment as the Brownian path. Without loss of generality, we shall be assuming that $n \geq 2$. It now suffices to show that W^n matches the $n-1$ iterated integrals given in (4.1.17). Recall that the stochastic integrals $\{I_k\}$ are defined in terms of a Brownian bridge as

$$I_k = \int_0^1 B_t \cdot \frac{e_k(t)}{t(1-t)}\, dt.$$

Using the orthogonality of $\{e_k\}$, it is a consequence of (4.1.18) that for $1 \leq k \leq n-1$:

$$\begin{aligned} I_k &= \int_0^1 \Big(W_t - W_1 e_0(t)\Big) \cdot \frac{e_k(t)}{t(1-t)}\, dt \\ &= \int_0^1 \left(W_t^n + \sum_{m=n}^{\infty} I_m e_m(t) - W_1 e_0(t)\right) \cdot \frac{e_k(t)}{t(1-t)}\, dt \\ &= \int_0^1 \Big(W_t^n - W_1 e_0(t)\Big) \cdot \frac{e_k(t)}{t(1-t)}\, dt + \sum_{m=n}^{\infty} I_m \int_0^1 \frac{e_m(t)\, e_k(t)}{t(1-t)}\, dt \\ &= \int_0^1 \Big(W_t^n - W_1 e_0(t)\Big) \cdot \frac{e_k(t)}{t(1-t)}\, dt. \end{aligned}$$

Therefore W^n matches the integrals of Brownian motion against polynomials with degree at most $n-1$. The result now follows from the integration by parts argument used in the proof of lemma 4.1.5. □

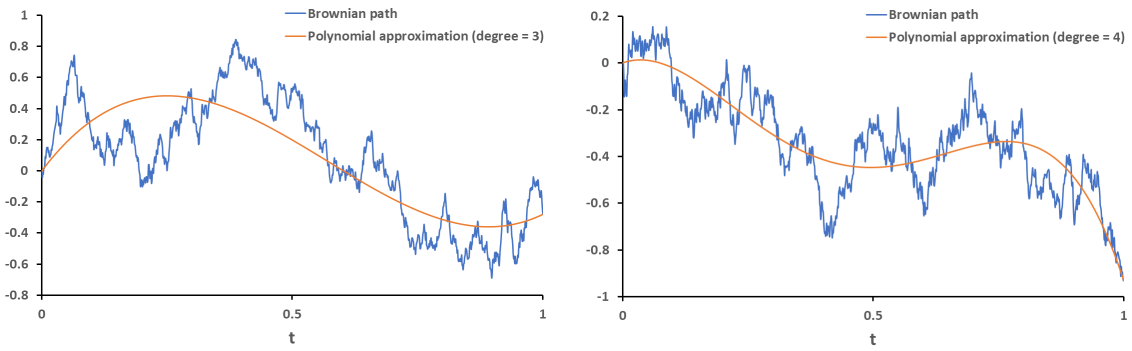

Figure 4.3: Sample paths of Brownian motion with their polynomial approximations.

Whilst the approximation theorems we have seen are interesting from a theoretical viewpoint, there has been no explicit formula for the orthogonal polynomials $\{e_k\}$. Fortunately, it was identified in the proof of Theorem 4.1.3 that the derivative e_k' is proportional to the k-th shifted Legendre polynomial. As a result, we would expect that each e_k is a (normalized) shifted (α, β)-Jacobi polynomial but with $\alpha = \beta = -1$. Since Jacobi polynomials are traditionally studied with $\alpha, \beta > -1$, it is important to check that there are no issues appearing as both of these parameters tend to -1. After establishing the existence of this non-standard family of Jacobi polynomials, we shall derive explicit constructions for them and hence also for the functions $\{e_k\}$.

Lemma 4.1.7. *Let $P_k^{(\alpha,\beta)}$ denote the k-th degree (α, β)-Jacobi polynomial on $[-1, 1]$. Then for $k \geq 2$, there exists a real-valued polynomial P_k such that $\left\|P_k - P_k^{(\alpha,\beta)}\right\|_\infty \to 0$ as $\alpha, \beta \to -1^+$.*

Proof. Consider the following integral relationship for (α,β)-Jacobi polynomials with $\alpha,\beta > -1$ (see [96]).

$$P_k^{(\alpha,\beta)}(x) = \frac{k+\alpha+\beta+1}{2}\int_{-1}^{x} P_{k-1}^{(\alpha+1,\beta+1)}(u)\,du, \quad \text{for all } k \geq 2. \tag{4.1.19}$$

Define a real-valued polynomial P_k on $[-1,1]$ by

$$P_k(x) := \frac{k-1}{2}\int_{-1}^{x} P_{k-1}^{(0,0)}(u)\,du, \quad \text{for all } k \geq 2. \tag{4.1.20}$$

Since the terms of (α,β)-Jacobi polynomials depend continuously on (α,β), we have $P_n^{(0,0)} = \lim_{\alpha,\beta\to 0} P_n^{(\alpha,\beta)}$ for $n \geq 1$. The result follows from (4.1.19) and (4.1.20). □

Hence, we can take the limit $\alpha,\beta \to -1^+$ so that the definition below makes sense.

Definition 4.1.8. For $k \geq 2$, the k-th degree $(-1,-1)$-Jacobi polynomial $P_k^{(\text{-}1,\text{-}1)}$ is

$$P_k^{(\text{-}1,\text{-}1)} := \lim_{\alpha,\beta\to -1^+} P_k^{(\alpha,\beta)}.$$

Using the above definition, it is now straightforward to produce an explicit formula for the family of orthogonal polynomials $\{e_k\}$ originally appearing in Theorem 4.1.3.

Theorem 4.1.9. *Suppose each e_k has a positive leading coefficient. Then for $k \geq 1$,*

$$e_k(t) = \frac{1}{k}\sqrt{k(k+1)(2k+1)}\,P_{k+1}^{(\text{-}1,\text{-}1)}(2t-1), \quad \forall t \in [0,1].$$

Proof. The following result is stated in [96]:

$$\int_{-1}^{1}(1-x)^\alpha(1+x)^\beta\big(P_n^{(\alpha,\beta)}(x)\big)^2\,dx = \frac{2^{\alpha+\beta+1}}{2n+\alpha+\beta+1}\frac{\Gamma(n+\alpha+1)\,\Gamma(n+\beta+1)}{\Gamma(n+\alpha+\beta+1)\,n!},$$

for $n \geq 1$ and $\alpha,\beta > -1$. Applying the change of variables, $t := \frac{1}{2}(x+1)$, we have

$$\int_0^1 t^\beta(1-t)^\alpha\big(P_n^{(\alpha,\beta)}(2t-1)\big)^2 dt = \frac{1}{2n+\alpha+\beta+1}\frac{\Gamma(n+\alpha+1)\,\Gamma(n+\beta+1)}{\Gamma(n+\alpha+\beta+1)\,n!},$$

for $n \geq 1$ and $\alpha,\beta > -1$. By definition 4.1.8, taking the limit $\alpha,\beta \to -1^+$ will yield

$$\begin{aligned}\int_0^1 \frac{1}{t(1-t)}\left(P_n^{(\text{-}1,\text{-}1)}(2t-1)\right)^2 dt &= \frac{1}{2n-1}\frac{(n-1)!\,(n-1)!}{n!\,(n-2)!}\\ &= \frac{1}{2n-1}\frac{n-1}{n}, \quad \text{for all } n \geq 2.\end{aligned}$$

Therefore by setting $k := n-1$, we have

$$\int_0^1 \frac{1}{t(1-t)}\left(\frac{1}{k}\sqrt{k(k+1)(2k+1)}\,P_n^{(\text{-}1,\text{-}1)}(2t-1)\right)^2 dt = 1, \quad \text{for all } k \geq 1.$$

Note that the construction of P_k in the proof of lemma 4.1.7 implies that each $e_k(t)$ is proportional to $P_{k+1}^{(\text{-}1,\text{-}1)}(2t-1)$. The result now follows from the above calculations. □

In practice, there are at least two different methodologies for computing the Jacobi-like polynomials $P_k^{(-1,-1)}$. These approaches are outlined by the theorems below.

Theorem 4.1.10 (A recurrence relation for constructing Jacobi-like polynomials)**.** *For each $n \geq 2$, we have*

$$n(n+2)P_{n+2}^{(-1,-1)}(x) = (n+1)(2n+1)xP_{n+1}^{(-1,-1)}(x) - n(n+1)P_n^{(-1,-1)}(x), \qquad (4.1.21)$$

where the initial polynomials are given by

$$P_2^{(-1,-1)}(x) = \frac{1}{4}(x-1)(x+1),$$
$$P_3^{(-1,-1)}(x) = \frac{1}{2}x(x-1)(x+1).$$

Proof. The following recurrence relation for Jacobi polynomials is given in [96],

$$\begin{aligned} &2(k+1)(k+\alpha+\beta+1)(2k+\alpha+\beta)P_{k+1}^{(\alpha,\beta)}(x) \\ &= (2k+\alpha+\beta+1)\left((2k+\alpha+\beta)(2k+\alpha+\beta+2)x+\alpha^2-\beta^2\right)P_k^{(\alpha,\beta)}(x) \\ &\quad - 2(k+\alpha)(k+\beta)(2k+\alpha+\beta+2)\,P_{k-1}^{(\alpha,\beta)}(x), \end{aligned}$$

for $k \geq 1$ and $\alpha, \beta > -1$. By definition 4.1.8, it is possible to take the limit $\alpha, \beta \to 1^+$ provided that $k \geq 3$. Therefore, taking this limit and setting $n = k-1 \geq 2$ produces

$$4n^2(n+2)P_{n+2}^{(-1,-1)}(x) = 4n(n+1)(2n+1)xP_{n+1}^{(-1,-1)}(x) - 4n^2(n+1)\,P_n^{(-1,-1)}(x),$$

for $n \geq 2$. Dividing the above by $4n$ gives the required recurrence relation (4.1.21). Finally the below formula, stated in [96], can be used to compute $P_2^{(-1,-1)}$ and $P_3^{(-1,-1)}$:

$$P_n^{(\alpha,\beta)}(x) = \frac{(-1)^n}{2^n n!}(1-x)^{-\alpha}(1+x)^{-\beta}\frac{d^n}{dx^n}\left((1-x)^{\alpha+n}(1+x)^{\beta+n}\right), \quad \text{for } n \geq 2.$$

As before we take $\alpha, \beta \to 1^+$ in the above to obtain an explicit formula for $P_n^{(-1,-1)}$. □

In addition to computing these polynomials via a recurrence relation, it is also possible to express each $P_n^{(-1,-1)}$ as the difference of two rescaled Legendre polynomials. Since the Legendre polynomials already have efficient implementations in the majority of high-level programming languages, this second approach is particularly appealing.

Theorem 4.1.11 (Relationship between the Jacobi-like and Legendre polynomials)**.** *For $n \geq 1$, we have*

$$P_{n+1}^{(-1,-1)}(x) = \frac{n}{4n+2}\left(Q_{n+1}(x) - Q_{n-1}(x)\right),$$

where Q_k denotes the degree k Legendre polynomial defined on $[-1,1]$.

Proof. Recall that $\frac{d}{dx}(P_{n+1}^{(-1,-1)}) = \frac{n}{2}P_n^{(0,0)}$ for $n \geq 1$, where $P_n^{(0,0)}(= Q_n)$ denotes the degree n Legendre polynomial. Therefore differentiating both sides of (4.1.21) yields

$$\frac{1}{2}n(n+1)(n+2)Q_{n+1}(x) = (n+1)(2n+1)P_{n+1}^{(-1,-1)}(x) + \frac{1}{2}n(n+1)(2n+1)xQ_n(x) - \frac{1}{2}n(n-1)(n+1)Q_{n-1}(x).$$

Hence by simplifying and rearranging the above, we have that for $n \geq 1$,

$$(2n+1)P_{n+1}^{(-1,-1)}(x) = \frac{1}{2}n\left(Q_{n+1}(x) - Q_{n-1}(x)\right) + \frac{1}{2}n\left((n+1)Q_{n+1}(x) - (2n+1)xQ_n(x) + n\,Q_{n-1}(x)\right).$$

We see the last term is zero by a recurrence relation for Legendre polynomials [31]. □

Using the above theorem, one can generate plots of these "Brownian polynomials".

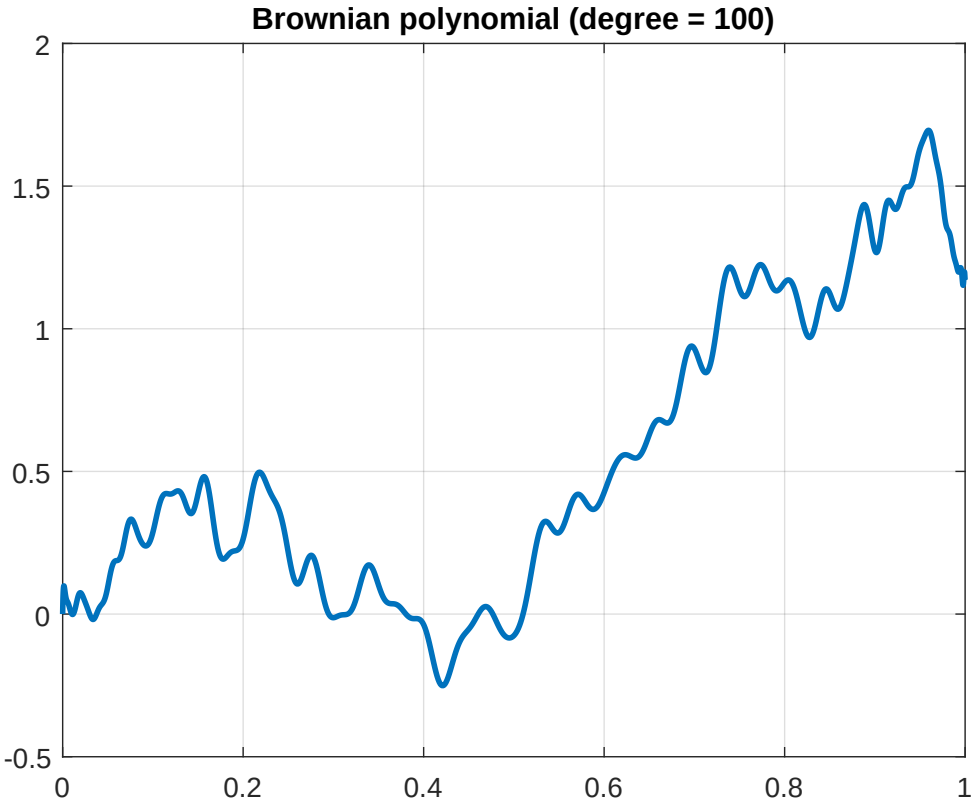

Figure 4.4: A polynomial sample path of Brownian motion that has degree 100.

Matlab code for producing such plots can be found at the following web page:

www.chebfun.org/examples/stats/RandomPolynomials.html

The first four orthogonal polynomials are given below for $t \in [0,1]$:

$$\begin{aligned}
e_1(t) &= \sqrt{6}\, t(t-1), \\
e_2(t) &= \sqrt{30}\, t(t-1)(2t-1), \\
e_3(t) &= \frac{5\sqrt{21}}{2}\, t(t-1)\left((2t-1)^2 - \frac{1}{5}\right), \\
e_4(t) &= \frac{21\sqrt{5}}{2}\, t(t-1)(2t-1)\left((2t-1)^2 - \frac{3}{7}\right).
\end{aligned}$$

In addition to viewing the polynomials $\{e_k\}$ as orthogonal with respect to the weight function $w(x) := \frac{1}{x(1-x)}$, we can characterize them via their iterated integrals. In particular, for $1 \leq k \leq n-1$, we have

$$\begin{aligned}
\int_{0<s_1<\cdots<s_k<1} e_n(s_1)\, ds_1 \cdots ds_k &= \int_0^1 e_n(s)\, d\left(\frac{1}{k!}s^k\right) \\
&= \frac{1}{(k-1)!}\int_0^1 s^{k-1} e_n(s)\, ds \\
&= -\frac{1}{k!}\int_0^1 s^k e_n'(s)\, ds \\
&= 0, \qquad \text{(by the orthogonality of } e_k'\text{)}.
\end{aligned}$$

Thus for $k \geq 1$, e_k is a polynomial with degree $k+1$ that has zeroes at 0 and 1 as well as $k-1$ trivial iterated integrals against time. By additionally specifying the k-th iterated time integral, it is then possible to characterize the k-th polynomial e_k.

We shall now address the relationship between the Jacobi-like polynomials $\{e_k\}$ and the Alpert-Rokhlin multiwavelets given by definition 4.1.1. Recall that the family of polynomials $\{e_k'\}$ is orthogonal with respect to the standard $L^2([0,1])$ inner product. This orthogonality is exactly what is needed to satisfy conditions (4.1.2) and (4.1.3). So for any $q \geq 1$ there exists an Alpert-Rokhlin mother function of order q that is a piecewise polynomial where both pieces can be translated and rescaled to give e_{q-1}'. Since it is the integrals of Alpert-Rokhlin wavelets that form the approximants for the Brownian path, we see that both of the approaches discussed in this chapter will create the same family of piecewise polynomial approximations of Brownian motion.

Finally, we shall study the error processes corresponding to these polynomial approximations. In particular, we will focus on the two simplest non-trivial cases (when the polynomial approximant is quadratic or cubic). It will be shown that these error processes are independent and Gaussian with interesting variance profiles.

Theorem 4.1.12. *Let W denote a standard real-valued Brownian motion on $[0,1]$. Let W^n be the unique n-th degree random polynomial with a root at 0 and satisfying*

$$\int_0^1 u^k \, dW_u^n = \int_0^1 u^k \, dW_u, \quad \text{for } k = 0, 1, \cdots, n-1.$$

Then $W = W^n + Z^n$, where Z^n is a centered Gaussian process independent of W^n. Furthermore, Z^n has the following covariance function:

$$\mathrm{Cov}(Z_s^n, Z_t^n) = K_B(s,t) - \sum_{k=1}^{n-1} \lambda_k e_k(s) e_k(t), \quad \text{for } s,t \in [0,1].$$

where K_B denotes the standard Brownian bridge covariance function and $\{\lambda_k\}$, $\{e_k\}$ are the eigenvalues and eigenfunctions that were defined in the proof of Theorem 4.1.3.

Proof. It follows directly from the integration by parts formula that W^n matches the increment and $n-1$ iterated time integrals of Brownian motion appearing in (4.1.17). Hence W^n is also the polynomial defined in Theorem 4.1.6 and $W = W^n + Z^n$ where

$$W^n = W_1 e_0 + \sum_{k=1}^{n-1} I_k e_k,$$

$$Z^n = \sum_{k=n}^{\infty} I_k e_k.$$

So by Theorem 4.1.3, Z^n is a centered Gaussian process that is independent of W^n. In addition, the covariance function defining Z^n can be directly computed as follows:

$$\begin{aligned}
\mathrm{Cov}(Z_s^n, Z_t^n) &= \mathrm{Cov}\left(\sum_{i=n}^{\infty} I_i e_i(s), \sum_{j=n}^{\infty} I_j e_j(t) \right) \\
&= \sum_{k=n}^{\infty} \lambda_k e_k(s) e_k(t) \\
&= K_B(s,t) - \sum_{k=1}^{n-1} \lambda_k e_k(s) e_k(t), \quad \text{for } s,t \in [0,1].
\end{aligned}$$

Note that the final line is achieved using the representation of K_B given by (4.1.8). □

The above theorem has a simple yet striking conclusion, namely that polynomials can be unbiased approximants of Brownian motion for which the error process is independent and can be uniformly estimated in an $L^2(\mathbb{P})$ sense. Since we shall be investigating some applications of this theorem, we will use the following definitions:

Definition 4.1.13. The standard **Brownian parabola** $\widehat{W}$ is the unique quadratic polynomial on $[0,1]$ with a root at 0 and satisfying

$$\widehat{W}_1 = W_1, \qquad \int_0^1 \widehat{W}_u \, du = \int_0^1 W_u \, du.$$

Definition 4.1.14. The standard **Brownian arch** Z is the process $Z := W - \widehat{W}$. By Theorem 4.1.12 is the centered Gaussian process on $[0,1]$ with covariance function

$$K_Z(s,t) = \min(s,t) - st - 3st(1-s)(1-t), \quad \text{for } s,t \in [0,1].$$

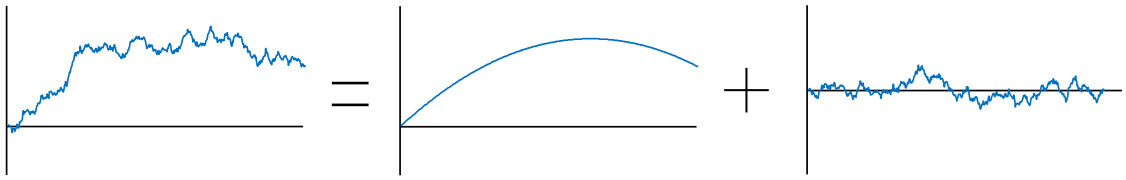

Figure 4.5: Brownian motion can be expressed as a parabola plus independent noise. In fact, this parabola has the same increment and time integral as the original path.

Definition 4.1.15. The standard ***n*-th degree Brownian polynomial** W^n is the n-th degree random polynomial with a root at 0 that was defined in Theorem 4.1.12.

Definition 4.1.16. The standard ***n*-th degree Brownian bridge** Z^n is the centered Gaussian process on $[0,1]$ given as $Z^n := W - W^n$ and constructed in Theorem 4.1.12.

In the case $n = 3$, the above bridge process has the following covariance function:

$$\begin{aligned} K_{Z^3}(s,t) = \min(s,t) - st - 3st(1-s)(1-t) \\ - 5st(1-s)(1-t)(1-2s)(1-2t), \quad \text{for } s,t \in [0,1]. \end{aligned}$$

By applying the natural scaling of Brownian motion, the above definitions can be extended to define processes on any interval $[s,t]$ with finite length $h = t - s$. Although the Brownian arch can be viewed in a similar light to the Brownian bridge, there are clear qualitative and quantitative differences in their covariance functions. In particular, the Brownian arch has less variance at its midpoint compared to most points in $[s,t]$ $\big($by which we mean that $|\{u \in [s,t] : \mathrm{Var}(Z_u) \leq \mathrm{Var}(Z_{\frac{1}{2}(s+t)})\}| < \frac{1}{2}h\big)$. This is in contrast to the Brownian bridge, which has most variance at its midpoint. Moreover, the Brownian parabola gives a fairly uniform estimate of the original path.

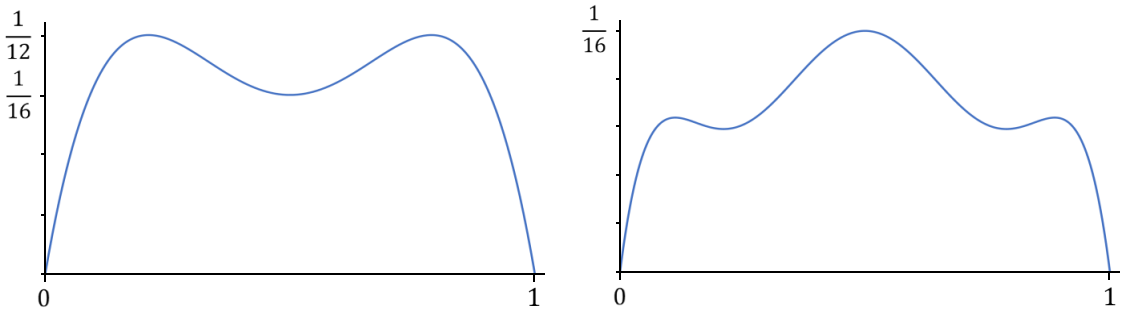

Figure 4.6: Variance profiles of the Brownian arch and third degree bridge over $[0, 1]$.

The polynomial expansion of Brownian motion presented in this section was also obtained by Habermann in [50] and expressed using shifted Legendre polynomials. Furthermore, it was shown that the n-th degree bridge process Z^n converges to zero in an $L^2(\mathbb{P})$ sense at a rate of $O\left(\frac{1}{\sqrt{n}}\right)$ and has a variance profile approaching a semicircle.

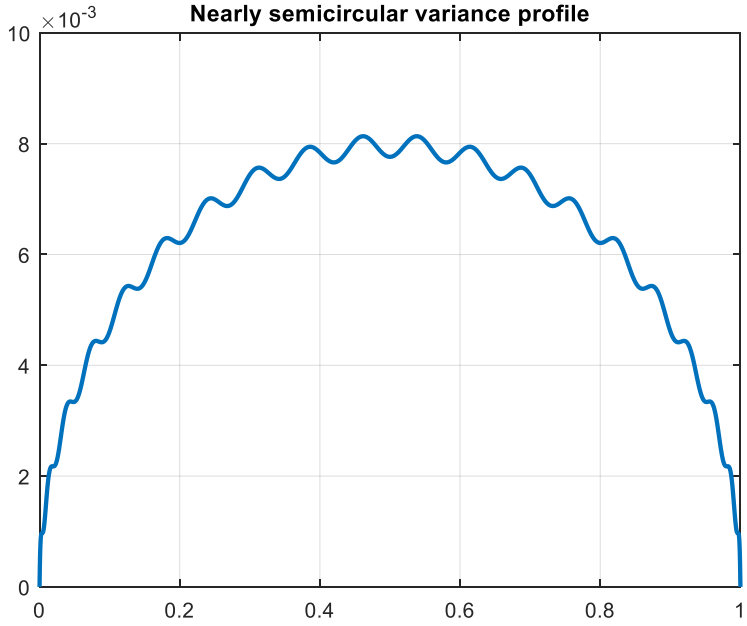

Figure 4.7: Variance profile of the 20th degree Brownian bridge over the unit interval.

It is straightforward to extend all the results in this section to a d-dimensional Brownian motion $W = \left(W^{(1)}, \cdots, W^{(d)}\right)$. For example, we can approximate W by taking d independent polynomial approximations for each coordinate process $W^{(i)}$. There may also be possible extensions of these results to fractional Brownian motion, however this will not be explored within the thesis. In Chapter 7, we will consider the multidimensional setting and use the polynomials to study Brownian Lévy area.

Since we have a polynomial expansion of Brownian motion, it is natural to compare it to the standard Karhunen–Loève expansion, which involves trigonometric functions. Truncating the trigonometric expansion may give a better approximation of W as it is optimal with respect to the standard $L^2(\mathbb{P})$ norm for stochastic processes on $[0,1]$,

$$\|X\|_{L^2(\mathbb{P})} := \sqrt{\mathbb{E}\left[\int_0^1 (X_s)^2 \, ds\right]},$$

whereas the polynomial approach is optimal with respect to a weighted $L^2(\mathbb{P})$ norm. Therefore we believe the advantage of the polynomial approach is theoretical in that it leads to new approximations for iterated integrals of Brownian motion and SDEs. In fact, the majority of approximations proposed within this thesis are based on this polynomial expansion (one exception being the shifted ODE method in section 5.3).

Although we have not found an application for it, we would also like to present a "rational construction" of Brownian motion that is associated to the polynomials e_k. To derive this rational construction, we will first note the following standard result:

Proposition 4.1.17 (Brownian motion as a rescaled time-changed Brownian bridge)**.** *Let B denote a Brownian bridge on $[0,1]$. Then the stochastic process $W' = \{W'_t\}_{t\geq 0}$ given by*

$$W'_t := (1+t)B_{\frac{t}{1+t}},$$

has the same law as a standard Brownian motion defined on $[0,\infty)$.

Proof. To show the result, it suffices to verify that $\mathrm{Cov}(W'_s, W'_t) = \min(s,t)$. □

By expressing B in terms of the polynomials $\{e_k\}$, we obtain the following theorem:

Theorem 4.1.18 (Construction of Brownian motion using rational functions)**.** *Let $\{Z_k\}_{k\geq 1}$ be a sequence of independent normal random variables with $Z_k \sim \mathcal{N}\big(0, \frac{1}{k(k+1)}\big)$ and $\{e_k\}_{k\geq 1}$ be the family of orthogonal polynomials given by Theorem 4.1.3 and define*

$$W'_t := \sum_{k=1}^{\infty} Z_k \widetilde{e}_k(t)$$

where $\widetilde{e}_k(t) := (1+t)e_k\big(\frac{t}{1+t}\big)$ for $t \geq 0$ (note that each $\widetilde{e}_k$ is a rational function). Then W' has the same law as a standard Brownian motion defined on $[0,\infty)$.

Proof. The result follows immediately from Theorem 4.1.3 and Proposition 4.1.17. □

Moreover, we conjecture that $\{\widetilde{e}_k\}_{k\geq 1}$ is a family of Jacobi-like rational functions on $[0,\infty)$ which are orthogonal with respect to the weight function $w(x) := \frac{1}{x(1+x)^2}$.

4.2 Conditional moments of third order iterated Brownian time integrals

It was shown in the previous section that certain iterated integrals of Brownian motion and time have a centered Gaussian distribution. In particular, these iterated integrals are expressible using multiple dt integrators but with only a single dW integrator. Thus, they correspond to certain linear functionals evaluated on the Brownian path. It is therefore practical to generate such integrals for the purpose of SDE simulation. The following integral is not Gaussian and currently can only be approximated [89],

$$\int_s^t \int_s^u \int_s^v \circ\, dW_r \circ\, dW_v\, du.$$

In this section, we will derive new optimal and unbiased estimators for the above quantity and understand their associated $L^2(\mathbb{P})$ errors. These estimators and their $L^2(\mathbb{P})$ errors can then be utilized in numerical methods for SDEs. For example, they can be directly incorporated into the stochastic Taylor method and the log-ODE or Castell-Gaines method (see [19, 74]). To achieve this, we use the following definitions:

Definition 4.2.1. The rescaled **space-time Lévy area** of Brownian motion over the interval $[s, t]$ encodes the signed area of the associated bridge process with time,

$$H_{s,t} := \frac{1}{h}\int_s^t W_{s,u} - \frac{u-s}{h} W_{s,t}\, du, \tag{4.2.1}$$

where $h = t - s$. Since $e_1(t) = \sqrt{6}\, t(t-1)$, one can see that $H_{0,1}$ corresponds to $\frac{\sqrt{6}}{6} I_1$ as defined in Theorem 4.1.3. Therefore $H_{s,t} \sim \mathcal{N}\left(0, \frac{1}{12}h\right)$ and is independent of $W_{s,t}$.

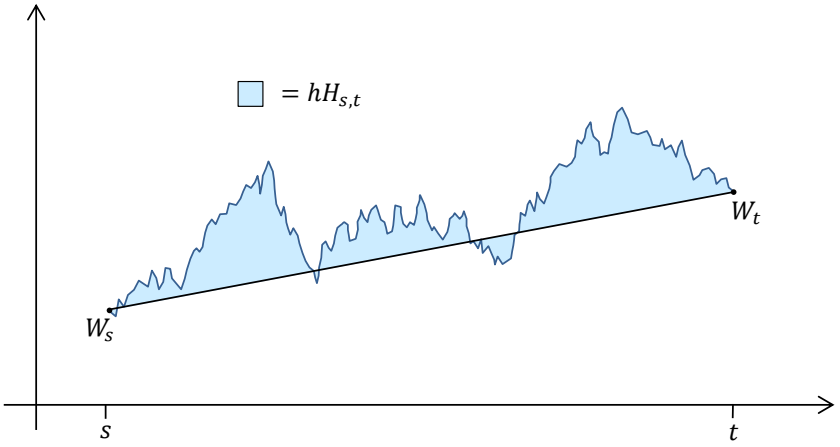

Figure 4.8: Space-time Lévy area as the region enclosed by a path and its chord.

Remark 4.2.2. It is straightforward to prove that $\int_s^t \int_s^u \circ\, dW_v\, du = \frac{1}{2} hW_{s,t} + hH_{s,t}$ and $\int_s^t \int_s^u dv \circ dW_u = \frac{1}{2} hW_{s,t} - hH_{s,t}$. This fact will be used by the log-ODE method.

Definition 4.2.3. The rescaled **space-time-time Lévy area** of Brownian motion over an interval $[s,t]$ is the below estimator for skewness of the associated arch process.

$$K_{s,t} := \frac{1}{h^2}\int_s^t \left(W_{s,u} - \frac{u-s}{h}W_{s,t}\right)\left(\frac{1}{2}h - (u-s)\right)du, \tag{4.2.2}$$

where $h = t-s$. Since $e_2(t) = \sqrt{30}\,t(t-1)(2t-1)$, $K_{0,1}$ will be associated to $\frac{1}{2\sqrt{30}}I_2$ as defined in Theorem 4.1.3. So $K_{s,t} \sim \mathcal{N}\left(0, \frac{1}{720}h\right)$ and is independent of $(W_{s,t}, H_{s,t})$.

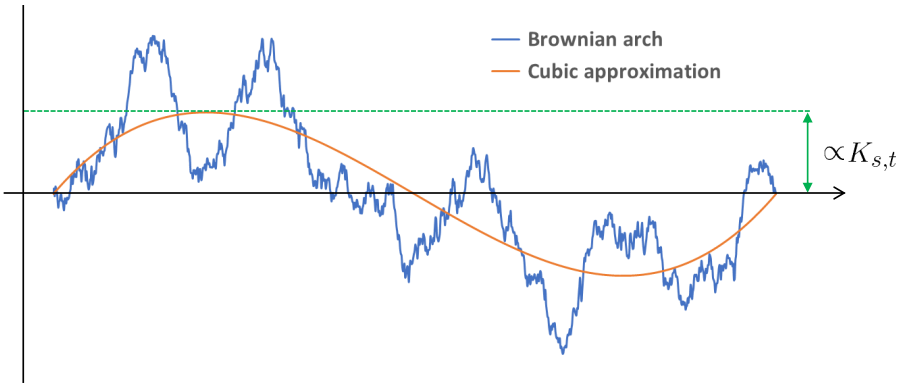

Figure 4.9: Space-time-time Lévy area gives the amplitude of the cubic approximant.

Definition 4.2.4. The **space-space-time Lévy area** of Brownian motion over an interval $[s,t]$ is defined as the following linear combination of triple iterated integrals.

$$\begin{aligned} L_{s,t} := \frac{1}{6}\Bigg(&\int_s^t\int_s^u\int_s^v \circ dW_r \circ dW_v\,du - 2\int_s^t\int_s^u\int_s^v \circ dW_r\,dv \circ dW_u \\ &+ \int_s^t\int_s^u\int_s^v dr \circ dW_v \circ dW_u\Bigg). \end{aligned} \tag{4.2.3}$$

Although it is difficult to describe space-space Lévy area through an illustration, the lemma below shows why it is an important object for studying iterated integrals.

Lemma 4.2.5. *Let $H_{s,t}$ and $L_{s,t}$ denote the Lévy areas of Brownian motion given by definitions 4.2.1 and 4.2.4 respectively. Then the following integral relationships hold,*

$$\begin{aligned} \int_s^t\int_s^u\int_s^v \circ dW_r \circ dW_v\,du &= \frac{1}{6}hW_{s,t}^2 + \frac{1}{2}hW_{s,t}H_{s,t} + L_{s,t}, \\ \int_s^t\int_s^u\int_s^v \circ dW_r\,dv \circ dW_u &= \frac{1}{6}hW_{s,t}^2 - 2L_{s,t}, \\ \int_s^t\int_s^u\int_s^v dr \circ dW_v \circ dW_u &= \frac{1}{6}hW_{s,t}^2 - \frac{1}{2}hW_{s,t}H_{s,t} + L_{s,t}, \end{aligned}$$

where $h = t-s$.

Proof. By the Markov property of Brownian motion, it is enough to show the result in the case $s = 0$. We use the following notation to represent various iterated integrals.

$$I_\tau(t) := \int_0^t du, \qquad I_w(t) := \int_0^t \circ\, dW_u,$$

$$I_{\tau w}(t) := \int_0^t \int_0^u dv \circ\, dW_u, \qquad I_{w\tau}(t) := \int_0^t \int_0^u \circ\, dW_v\, du,$$

$$I_{\tau ww}(t) := \int_0^t \int_0^u \int_0^v dr \circ dW_v \circ dW_u, \qquad I_{w\tau w}(t) := \int_0^t \int_0^u \int_0^v \circ\, dW_r\, dv \circ dW_u,$$

$$I_{ww\tau}(t) := \int_0^t \int_0^u \int_0^v \circ\, dW_r \circ\, dW_v\, du.$$

The result then follows by direct computation using the integration by parts formula (or equivalently the shuffle product on words with letters from the alphabet $\{\tau, w\}$).

$$\begin{aligned}
\frac{1}{6}tW_t^2 + \frac{1}{2}tW_tH_t + L_t &= \frac{1}{6}I_\tau I_w I_w + \frac{1}{4}I_w\left(I_{w\tau} - I_{\tau w}\right) + \frac{1}{6}\left(I_{ww\tau} - 2I_{w\tau w} + I_{\tau ww}\right) \\
&= \frac{1}{3}\left(I_{\tau ww} + I_{w\tau w} + I_{ww\tau}\right) + \frac{1}{4}\left(2I_{ww\tau} + I_{w\tau w}\right) \\
&\quad - \frac{1}{4}\left(I_{w\tau w} + 2I_{\tau ww}\right) + \frac{1}{6}I_{ww\tau} - \frac{1}{3}I_{w\tau w} + \frac{1}{6}I_{\tau ww} \\
&= I_{ww\tau}.
\end{aligned}$$

$$\begin{aligned}
\frac{1}{6}tW_t^2 - 2L_t &= \frac{1}{6}I_\tau I_w I_w - \frac{1}{3}\left(I_{ww\tau} - 2I_{w\tau w} + I_{\tau ww}\right) \\
&= \frac{1}{3}\left(I_{\tau ww} + I_{w\tau w} + I_{ww\tau}\right) - \frac{1}{3}\left(I_{ww\tau} - 2I_{w\tau w} + I_{\tau ww}\right) \\
&= I_{w\tau w}.
\end{aligned}$$

$$\begin{aligned}
\frac{1}{6}tW_t^2 - \frac{1}{2}tW_tH_t + L_t &= \frac{1}{6}I_\tau I_w I_w - \frac{1}{4}I_w\left(I_{w\tau} - I_{\tau w}\right) + \frac{1}{6}\left(I_{ww\tau} - 2I_{w\tau w} + I_{\tau ww}\right) \\
&= \frac{1}{3}\left(I_{\tau ww} + I_{w\tau w} + I_{ww\tau}\right) - \frac{1}{4}\left(2I_{ww\tau} + I_{w\tau w}\right) \\
&\quad + \frac{1}{4}\left(I_{w\tau w} + 2I_{\tau ww}\right) + \frac{1}{6}I_{ww\tau} - \frac{1}{3}I_{w\tau w} + \frac{1}{6}I_{\tau ww} \\
&= I_{\tau ww}.
\end{aligned}$$

□

It is also possible to derive similar formulae for the rescaled space-time-time Lévy area defined previously. These integral relationships can be interpreted using rough path theory and will be vital to our error analysis for a new version of the log-ODE (or Castell-Gaines) method. In the last section of this chapter, we shall utilize this method to accurately discretize Inhomogeneous Geometric Brownian Motion (IGBM).

Lemma 4.2.6. *Let $H_{s,t}$ and $K_{s,t}$ denote the rescaled Lévy areas of Brownian motion given by definitions 4.2.1 and 4.2.3. Then we have the following integral relationships,*

$$\int_s^t \int_s^u \int_s^v \circ dW_r \; dv \; du = \frac{1}{6}h^2 W_{s,t} + \frac{1}{2}h^2 H_{s,t} + h^2 K_{s,t},$$
$$\int_s^t \int_s^u \int_s^v dr \circ dW_v \, du = \frac{1}{6}h^2 W_{s,t} - 2h^2 K_{s,t},$$
$$\int_s^t \int_s^u \int_s^v dr \, dv \circ dW_u = \frac{1}{6}h^2 W_{s,t} - \frac{1}{2}h^2 H_{s,t} + h^2 K_{s,t}.$$

Proof. It suffices to prove the result when $s = 0$. It follows by direct calculation that

$$\begin{aligned}
\int_0^t \int_0^u \int_0^v dr \, dv \circ dW_u &= \int_0^t \frac{1}{2}u^2 \circ dW_u \\
&= \frac{1}{2}t^2 W_t - \int_0^t W_u \, d\left(\frac{1}{2}u^2\right) && \text{(integration by parts)} \\
&= \frac{1}{2}t^2 W_t - \int_0^t u W_u \, du \\
&= \frac{1}{2}t^2 W_t + \int_0^t \left(W_u - \frac{u}{t}W_t\right)\left(\frac{1}{2}t - u\right) du \\
&\quad - \frac{1}{2}t\int_0^t W_u \, du + \int_0^t \frac{u}{t} W_t \left(\frac{1}{2}t - u\right) du \\
&= \frac{1}{6}t^2 W_t - \frac{1}{2}t^2 H_t + t^2 K_t.
\end{aligned}$$

$$\begin{aligned}
\int_0^t \int_0^u \int_0^v \circ dW_r \; dv \; du &= \int_0^t \int_0^u W_v \, dv \, du \\
&= t\int_0^t W_u \, du - \int_0^t u \, d\left(\int_0^u W_v \, dv\right) && \text{(integration by parts)} \\
&= t\int_0^t W_u \, du - \int_0^t u W_u \, du \\
&= \frac{1}{6}t^2 W_t + \frac{1}{2}t^2 H_t + t^2 K_t. && \text{(by the previous calculations)}
\end{aligned}$$

$$\begin{aligned}
\int_0^t \int_0^u \int_0^v dr \circ dW_v \, du &= \int_0^t \left(u W_u - \int_0^u \int_0^v \circ dW_r \, dv\right) du && \text{(integration by parts)} \\
&= \int_0^t u W_u \, du - \int_0^t \int_0^u \int_0^v \circ dW_r \; dv \; du \\
&= \frac{1}{6}t^2 W_t - 2t^2 K_t. && \text{(by the previous calculations)} \quad \square
\end{aligned}$$

Using the notation presented in the proof of lemma 4.2.5, we can express these Brownian integrals in terms of the commutator Lie bracket $[\cdot, \cdot]$ on the algebra $\mathbb{R}\langle \tau, w \rangle$.

Theorem 4.2.7. *Let $H_{s,t}$, $L_{s,t}$ and $K_{s,t}$ denote the Lévy areas of Brownian motion over the interval $[s,t]$ given by definitions 4.2.1, 4.2.1 and 4.2.3 respectively. Then*

$$\begin{aligned} W_{s,t} &= I_w\,, \\ hH_{s,t} &= \frac{1}{2} I_{[w,\tau]}\,, \\ L_{s,t} &= \frac{1}{6} I_{[w,[w,\tau]]}\,, \\ h^2 K_{s,t} &= \frac{1}{6} I_{[[w,\tau],\tau]}\,, \end{aligned}$$

where for any word $z \in \{\tau, w\}^*$ and we define the integral $I_z := I_z(s,t)$ recursively as

$$I_z(s,r) := \begin{cases} 1, & \text{if the word } z \text{ has zero length} \\ \int_s^r I_x(s,u)\,du, & \text{if } z = x\tau \text{ for some word } x \in \{\tau, w\}^* \\ \int_s^r I_x(s,u) \circ dW_u, & \text{if } z = xw \text{ for some word } x \in \{\tau, w\}^* \end{cases},$$

and for words $x, y \in \{\tau, w\}^*$ and $\lambda_1, \lambda_2 \in \mathbb{R}$, we can consider the linear combinations

$$\begin{aligned} I_{\lambda_1 x + \lambda_2 y} &:= \lambda_1 I_x + \lambda_2 I_y\,, \\ I_{[x,y]} &:= I_{xy-yx}\,. \end{aligned}$$

Proof. This is a consequence of remark 4.2.2, definition 4.2.4 and lemma 4.2.6. □

The above theorem has an interpretation within rough path theory. It shows that $W_{s,t}, H_{s,t}, L_{s,t}$ and $K_{s,t}$ can all be viewed as (rescaled) coefficients appearing in the Stratonovich log-signature of space-time Brownian motion $X^W := \{(u, W_u)\}_{u\in[s,t]}$. Specifically, they appear as (rescaled) coefficients of the log-signature with respect to the Lyndon basis for the free real Lie algebra generated by the alphabet $\{w, \tau\}$. Log-signatures, Lyndon words and their respective computations are detailed in [83].

Rough path theory also informs us that the log-signature is an efficient method of encoding information about a path [71]. In the thesis, this will be utilized for the numerical simulation of SDEs. It is first worth noting that $W_{s,t}, H_{s,t}, K_{s,t}$ are all independent Gaussian random variables and thus it is straightforward to generate or approximate these quantities in practice. On the other hand, the Lévy area $L_{s,t}$ poses a significant challenge as its distribution is non-Gaussian and currently it can only be approximately generated [89]. This motivates us to study the relationship between $L_{s,t}$ and the other "low order" coefficients of the space-time Brownian log-signature. In particular, the following theorem establishes various conditional moments of $L_{s,t}$ that give new unbiased estimators for the truncated log-signature of Brownian motion.

Theorem 4.2.8 (Conditional moments of Brownian space-space-time Lévy area)**.** *Let $H_{s,t}$, $L_{s,t}$ and $K_{s,t}$ denote the Brownian Lévy areas we previously defined. Then*

$$\mathbb{E}\left[L_{s,t} \,\middle|\, W_{s,t}, H_{s,t}\right] = \frac{1}{30}h^2 + \frac{3}{5}hH_{s,t}^2\,, \tag{4.2.4}$$

$$\operatorname{Var}\left(L_{s,t} \,\middle|\, W_{s,t}, H_{s,t}\right) = \frac{11}{25200}h^4 + h^3\left(\frac{1}{720}W_{s,t}^2 + \frac{1}{700}H_{s,t}^2\right), \tag{4.2.5}$$

$$\mathbb{E}\left[L_{s,t} \,\middle|\, W_{s,t}, H_{s,t}, K_{s,t}\right] = \frac{3}{140}h^2 + \frac{3}{5}hH_{s,t}^2 - hW_{s,t}K_{s,t} + \frac{60}{7}hK_{s,t}^2\,, \tag{4.2.6}$$

$$\operatorname{Var}\left(L_{s,t} \,\middle|\, W_{s,t}, H_{s,t}, K_{s,t}\right) = \frac{11}{88200}h^4 + h^3\left(\frac{1}{700}H_{s,t}^2 + \frac{1}{49}K_{s,t}^2\right), \tag{4.2.7}$$

Proof. By the natural Brownian scaling it is enough to prove the result on $[0,1]$. Recall that $W = \widehat{W} + Z$ where the parabola $\widehat{W}$ is completely determined by (W_1, H_1) and Z is independent of (W_1, H_1). This immediately leads to the following calculation,

$$\begin{aligned}
\mathbb{E}\left[\int_0^1 W_u^2\,du \,\middle|\, W_1, H_1\right] &= \mathbb{E}\left[\int_0^1 \left(\widehat{W}_u + Z_u\right)^2 du \,\middle|\, W_1, H_1\right] \\
&= \mathbb{E}\left[\int_0^1 \widehat{W}_u^2\,du + 2\int_0^1 \widehat{W}_u\,Z_u\,du + \int_0^1 Z_u^2\,du \,\middle|\, W_1, H_1\right] \\
&= \int_0^1 \widehat{W}_u^2\,du + 2\int_0^1 \widehat{W}_u\,\mathbb{E}\left[Z_u\right]du + \int_0^1 \mathbb{E}\left[Z_u^2\right]du \\
&= \int_0^1 (uW_1 + 6u(1-u)H_1)^2\,du + \int_0^1 u - u^2 - 3u^2(1-u)^2\,du \\
&= \int_0^1 u^2W_1^2 + 12u^2(1-u)W_1H_1 + 36u^2(1-u)^2H_1^2\,du \\
&\qquad + \int_0^1 u - u^2 - 3u^2(1-u)^2\,du \\
&= \frac{1}{3}W_1^2 + W_1H_1 + \frac{6}{5}H_1^2 + \frac{1}{15}.
\end{aligned}$$

The first conditional expectation (4.2.4) now follows from the above and lemma 4.2.5. Deriving the first conditional variance (4.2.5) requires more extensive computations.

$$\begin{aligned}
\operatorname{Var}\left(\int_0^1 W_u^2\,du \,\middle|\, W_1, H_1\right) &= \operatorname{Var}\left(\int_0^1 W_u^2\,du - \int_0^1 \widehat{W}_u^2\,du \,\middle|\, W_1, H_1\right) \\
&= \operatorname{Var}\left(2\int_0^1 \widehat{W}_u\,Z_u\,du + \int_0^1 Z_u^2\,du \,\middle|\, W_1, H_1\right) \\
&= 4\operatorname{Var}\left(\int_0^1 \widehat{W}_u\,Z_u\,du \,\middle|\, W_1, H_1\right) + \operatorname{Var}\left(\int_0^1 Z_u^2\,du\right) \\
&\qquad + 4\operatorname{Cov}\left(\int_0^1 \widehat{W}_u\,Z_u\,du, \int_0^1 Z_u^2\,du \,\middle|\, W_1, H_1\right).
\end{aligned}$$

Recall $Z = \sum_{k=2}^{\infty} I_k e_k$ where $\{I_k\}$ are independent centered Gaussian random variables. In particular, this means that Z and $-Z$ have the same law. Therefore, we have that

$$\mathbb{E}\left[\int_0^1 \widehat{W}_u Z_u \, du \,\middle|\, W_1, H_1\right] = 0,$$

and

$$\operatorname{Cov}\left(\int_0^1 \widehat{W}_u Z_u \, du, \int_0^1 Z_u^2 \, du \,\middle|\, W_1, H_1\right) = 0.$$

The remaining two terms were resolved with assistance from Wolfram Mathematica.

$$\begin{aligned}
&\operatorname{Var}\left(\int_0^1 \widehat{W}_u Z_u \, du \,\middle|\, W_1, H_1\right) \\
&\quad = \int_0^1 \int_0^1 \widehat{W}_u \widehat{W}_v \, \mathbb{E}\left[Z_u Z_v \,|\, W_1, H_1\right] \, du \, dv \\
&\quad = \int_0^1 \int_0^1 \widehat{W}_u \widehat{W}_v \left(\min(u,v) - uv - 3uv(1-u)(1-v)\right) \, du \, dv \\
&\quad = 2\int_0^1 \widehat{W}_v \int_0^v u \widehat{W}_u \, du \, dv - \left(\int_0^1 u \widehat{W}_u \, du\right)^2 - 3\left(\int_0^1 u(1-u)\widehat{W}_u \, du\right)^2 \\
&\quad = 2\left(\frac{1}{15}W_1^2 + \frac{13}{60}W_1 H_1 + \frac{13}{70}H_1^2\right) - \left(\frac{1}{3}W_1 + \frac{1}{2}H_1\right)^2 - 3\left(\frac{1}{12}W_1 + \frac{1}{5}H_1\right)^2 \\
&\quad = \frac{1}{720}W_1^2 + \frac{1}{700}H_1^2.
\end{aligned}$$

$$\begin{aligned}
\operatorname{Var}\left(\int_0^1 Z_u^2 \, du\right) &= \mathbb{E}\left[\left(\int_0^1 Z_u^2 \, du\right)^2\right] - \left(\mathbb{E}\left[\int_0^1 Z_u^2 \, du\right]\right)^2 \\
&= \int_0^1 \int_0^1 \mathbb{E}\left[Z_u^2 Z_v^2\right] \, du \, dv - \left(\mathbb{E}\left[\int_0^1 Z_u^2 \, du\right]\right)^2 \\
&= \int_0^1 \int_0^1 \mathbb{E}\left[Z_u^2\right] \mathbb{E}\left[Z_v^2\right] + 2\left(\mathbb{E}\left[Z_u Z_v\right]\right)^2 \, du \, dv - \left(\mathbb{E}\left[\int_0^1 Z_u^2 \, du\right]\right)^2 \\
&= 4\int_0^1 \int_0^v \left(u - uv - 3uv(1-u)(1-v)\right)^2 \, du \, dv \\
&= \frac{11}{6300}.
\end{aligned}$$

By lemma 4.2.5, we can derive an explicit formula for the conditional variance (4.2.5).

$$\operatorname{Var}\left(L_1 \,\middle|\, W_1, H_1\right) = \operatorname{Var}\left(\frac{1}{2}\int_0^1 W_u^2 \, du \,\middle|\, W_1, H_1\right) = \frac{11}{25200} + \frac{1}{720}W_1^2 + \frac{1}{700}H_1^2.$$

The last two moments (4.2.6) and (4.2.7) will be shown using a very similar strategy.

Recall that $W = \widetilde{W} + \widetilde{Z}$ where the cubic polynomial $\widetilde{W}$ is completely determined by (W_1, H_1, K_1) and $\widetilde{Z}$ is a centered Gaussian process independent of (W_1, H_1, K_1). Moreover, $\widetilde{Z}$ is a third degree Brownian bridge and has the below covariance function:

$$\begin{aligned} K_{\widetilde{Z}}(s,t) = \min(s,t) - st - 3st(1-s)(1-t) \\ - 5st(1-s)(1-t)(1-2s)(1-2t), \quad \text{for } s,t \in [0,1]. \end{aligned}$$

As before, we consider a conditional expectation for the integral of W^2 against time.

$$\begin{aligned} \mathbb{E}\left[\int_0^1 W_u^2\,du \,\middle|\, W_1, H_1, K_1\right] &= \mathbb{E}\left[\int_0^1 \left(\widetilde{W}_u + \widetilde{Z}_u\right)^2 du \,\middle|\, W_1, H_1, K_1\right] \\ &= \mathbb{E}\left[\int_0^1 \widetilde{W}_u^2\,du + 2\int_0^1 \widetilde{W}_u \widetilde{Z}_u\,du + \int_0^1 \widetilde{Z}_u^2\,du \,\middle|\, W_1, H_1, K_1\right] \\ &= \int_0^1 \widetilde{W}_u^2\,du + 2\int_0^1 \widetilde{W}_u\,\mathbb{E}[\widetilde{Z}_u]\,du + \int_0^1 \mathbb{E}[\widetilde{Z}_u^2]\,du \\ &= \int_0^1 \left(uW_1 + 6u(1-u)H_1 + 60u(1-u)(1-2u)K_1\right)^2 du \\ &\quad + \int_0^1 u - u^2 - 3u^2(1-u)^2 - 5u^2(1-u)^2(1-2u)^2\,du \\ &= \frac{1}{3}W_1^2 + W_1H_1 + \frac{6}{5}H_1^2 - 2W_1K_1 + \frac{120}{7}K_1^2 + \frac{3}{70}. \end{aligned}$$

The second conditional expectation (4.2.6) follows from the above and lemma 4.2.5. Finally, we will consider the second conditional variance (4.2.7) for the Lévy area L_1.

$$\begin{aligned} \mathrm{Var}\left(\int_0^1 W_u^2\,du \,\middle|\, W_1, H_1, K_1\right) &= 4\,\mathrm{Var}\left(\int_0^1 \widetilde{W}_u \widetilde{Z}_u\,du \,\middle|\, W_1, H_1, K_1\right) + \mathrm{Var}\left(\int_0^1 \widetilde{Z}_u^2\,du\right) \\ &\quad + 4\,\mathrm{Cov}\left(\int_0^1 \widetilde{W}_u \widetilde{Z}_u\,du, \int_0^1 \widetilde{Z}_u^2\,du \,\middle|\, W_1, H_1, K_1\right). \end{aligned}$$

Since $\widetilde{Z} = \sum_{k=3}^{\infty} I_k e_k$, we see that $\widetilde{Z}$ and $-\widetilde{Z}$ have the same law. Hence this implies that

$$\mathbb{E}\left[\int_0^1 \widetilde{W}_u \widetilde{Z}_u\,du \,\middle|\, W_1, H_1, K_1\right] = 0,$$

and

$$\mathrm{Cov}\left(\int_0^1 \widetilde{W}_u \widetilde{Z}_u\,du, \int_0^1 \widetilde{Z}_u^2\,du \,\middle|\, W_1, H_1, K_1\right) = 0.$$

Therefore the above conditional variance can be simplified to the following expression:

$$\mathrm{Var}\left(\int_0^1 W_u^2\,du \,\middle|\, W_1, H_1, K_1\right) = 4\mathrm{Var}\left(\int_0^1 \widetilde{W}_u \widetilde{Z}_u\,du \,\middle|\, W_1, H_1, K_1\right) + \mathrm{Var}\left(\int_0^1 \widetilde{Z}_u^2\,du\right)$$

The remaining two terms were resolved with assistance from Wolfram Mathematica.

$$
\begin{aligned}
&\operatorname{Var}\left(\int_0^1 \widetilde{W}_u \widetilde{Z}_u \, du \,\middle|\, W_1, H_1, K_1\right) \\
&\quad = \int_0^1 \int_0^1 \widetilde{W}_u \widetilde{W}_v \, \mathbb{E}\left[\widetilde{Z}_u \widetilde{Z}_v \,|\, W_1, H_1, K_1\right] \, du \, dv \\
&\quad = \int_0^1 \int_0^1 \widetilde{W}_u \widetilde{W}_v \left(\min(u,v) - uv - 3uv(1-u)(1-v)\right. \\
&\qquad\qquad \left. -5uv(1-u)(1-v)(1-2u)(1-2v)\right) \, du \, dv \\
&\quad = 2\int_0^1 \widetilde{W}_v \int_0^v u\widetilde{W}_u \, du \, dv - \left(\int_0^1 u\widetilde{W}_u \, du\right)^2 - 3\left(\int_0^1 u(1-u)\widetilde{W}_u \, du\right)^2 \\
&\qquad\quad - 5\left(\int_0^1 u(1-u)(1-2u)\widetilde{W}_u \, du\right)^2 \\
&\quad = 2\left(\frac{1}{15}W_1^2 + \frac{13}{60}W_1H_1 - \frac{5}{14}W_1K_1 + \frac{13}{70}H_1^2 - \frac{1}{2}H_1K_1 + \frac{5}{7}K_1^2\right) \\
&\qquad\quad - \left(\frac{1}{3}W_1 + \frac{1}{2}H_1 - K_1\right)^2 - 3\left(\frac{1}{12}W_1 + \frac{1}{5}H_1\right)^2 - 5\left(\frac{2}{7}K_1 - \frac{1}{60}W_1\right)^2 \\
&\quad = \frac{1}{700}H_1^2 + \frac{1}{49}K_1^2.
\end{aligned}
$$

$$
\begin{aligned}
\operatorname{Var}\left(\int_0^1 \widetilde{Z}_u^2 \, du\right) &= \mathbb{E}\left[\left(\int_0^1 \widetilde{Z}_u^2 \, du\right)^2\right] - \left(\mathbb{E}\left[\int_0^1 \widetilde{Z}_u^2 \, du\right]\right)^2 \\
&= \int_0^1 \int_0^1 \mathbb{E}[\widetilde{Z}_u^2 \widetilde{Z}_v^2] \, du \, dv - \left(\mathbb{E}\left[\int_0^1 \widetilde{Z}_u^2 \, du\right]\right)^2 \\
&= \int_0^1 \int_0^1 \mathbb{E}[\widetilde{Z}_u^2]\,\mathbb{E}[\widetilde{Z}_v^2] + 2\left(\mathbb{E}[\widetilde{Z}_u \widetilde{Z}_v]\right)^2 du \, dv - \left(\mathbb{E}\left[\int_0^1 \widetilde{Z}_u^2 \, du\right]\right)^2 \\
&= 4\int_0^1 \int_0^v \left(u - uv - 3uv(1-u)(1-v)\right. \\
&\qquad\qquad \left. -5uv(1-u)(1-v)(1-2u)(1-2v)\right)^2 \, du \, dv \\
&= \frac{11}{22050}.
\end{aligned}
$$

By lemma 4.2.5, we can derive an explicit formula for the conditional variance (4.2.7).

$$
\operatorname{Var}\left(L_1 \,\middle|\, W_1, H_1, K_1\right) = \operatorname{Var}\left(\frac{1}{2}\int_0^1 W_u^2 \, du \,\middle|\, W_1, H_1, K_1\right) = \frac{11}{88200} + \frac{1}{700}H_1^2 + \frac{1}{49}K_1^2.
$$

□

Remark 4.2.9. The methodology presented in the above proof can be extended to computing means and variances of $L_{s,t}$ conditional on $\{W_1, I_1, I_2, \cdots, I_n\}$ for $n > 2$. However, this would be quite challenging due to the involved calculations for large n.

Theorem 4.2.8 is a central result in the thesis and utilized in upcoming chapters. For the remainder of this section, we shall outline these applications for the theorem.

Most notably, the conditional expectations (4.2.4) and (4.2.6) can be used to approximate $L_{s,t}$ and are thus applicable to high order numerical methods for SDEs. Although there are many numerical methods for any given SDE, we will see that these unbiased estimators are especially applicable within the following classes of methods:

- **Stochastic Taylor methods**

  Unsurprisingly, it is possible to incorporate estimators for iterated integrals into stochastic Taylor methods. Chapter 2 gives a brief overview of these methods. Just as for ODEs, there are potentially three disadvantages of Taylor methods:

  - They require one to explicitly compute or estimate vector fields derivatives
  - They can exhibit instability (such as for SDEs on manifolds or stiff SDEs)
  - High order Taylor methods can perform poorly if large step sizes are used

  For the above reasons, we will present more suitable methods for the numerical examples covered in the thesis. On the other hand, it is worth noting that the Milstein scheme is a type of stochastic Taylor method and is particularly applicable to multidimensional SDEs (see [42] and [45] for concrete examples). Moreover, the approximations of Brownian Lévy area developed in Chapter 7 can be naturally incorporated into the Milstein scheme. That said, it is still ongoing work to demonstrate and quantify the effectiveness of this methodology.

- **The log-ODE method**

  The log-ODE method is a way of approximating an SDE by numerically solving an ODE constructed using iterated integrals along with vector field Lie brackets. Compared to Taylor methods, this approach is more stable and performs well when larger step sizes are used (provided the ODE solver has these properties).

  Unfortunately, this method has the disadvantage of requiring derivatives of vector fields and the "log-ODE" can be stiff and difficult to solve numerically. Chapter 3 contains an overview of this method from the rough path perspective.

  In the next section, we shall analyse a version of the log-ODE method that uses the estimator (4.2.4) for SDEs driven by a one-dimensional Brownian motion. The log-ODE method with (4.2.4) is then used further in this chapter to design a state-of-the-art discretization of Inhomogeneous Geometric Brownian Motion.

- **Stochastic Runge-Kutta methods**

  A popular class of numerical methods for ODEs are the Runge-Kutta methods, which have seen significant development in the past century (see [15] for details). The general idea is to match the Taylor expansion of the differential equation, up to some order, by weighted sums of nested compositions of the vector field. As SDEs have multiple vector fields and admit Taylor expansions which contain stochastic integrals, it is often difficult to design Runge-Kutta methods for them.

  In the penultimate section of this chapter, we will exploit the structure that comes from the parabola approximation of Brownian motion to design a high order three-stage stochastic Runge-Kutta method for SDEs with additive noise.

- **Piecewise linear or Wong-Zakai methods**

  A classical approach (first introduced in [97] by E. Wong and M. Zakai) is to discretize the driving Brownian motion as the piecewise linear path that agrees on discretization points. Along each such piece, the SDE then reduces to an ODE that can be approximated using a sufficiently accurate numerical method. It is worth noting that this is precisely the "derivative-free" log-ODE method.

  In Chapter 5, we develop a piecewise linear discretization of Brownian motion that captures the "shape" of the path (i.e. matches certain iterated integrals). Here the estimators (4.2.4) and (4.2.6) play a critical role in designing the paths. These new piecewise linear discretizations can be applied to SDE simulation just as before, but are expected to give higher orders of strong and weak convergence. Specific numerical examples for these methods are given in sections 5.3 and 6.3.

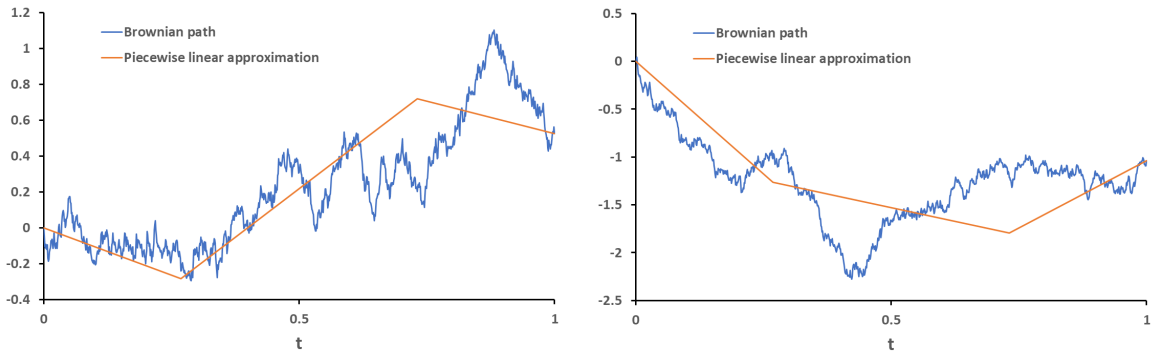

Figure 4.10: Brownian motion approximated by "high order" piecewise linear paths. In practice, the above piecewise linear paths can be generated using two independent Gaussian random variables and a further independent Rademacher random variable.

In addition, the conditional variances given by (4.2.5) and (4.2.7) enable one to compute the $L^2(\mathbb{P})$ errors corresponding to the estimators (4.2.4) and (4.2.6). Since the iterated integrals associated with $L_{s,t}$ appear naturally in stochastic Taylor expansions and can be the largest source of numerical error in high order methods, this provides a methodology for estimating the $L^2(\mathbb{P})$ error of each one-step approximation. Estimating such $L^2(\mathbb{P})$ errors is especially important within the context of variable step size methods which are very well suited for tackling non-homogeneous problems.

Unlike for ODEs, it is not entirely trivial to simply "take a smaller step size". The reason is that the decision to reduce a step size usually comes after information about the Brownian path has been generated (with respect to the larger step size). Fortunately, Lévy's construction of Brownian motion is precisely what is needed to generate path increments within an interval conditional on the endpoint information. In [13], it was explicitly shown how to extend this approach to generate both the Brownian increment and time area conditional on the same pair over a larger interval.

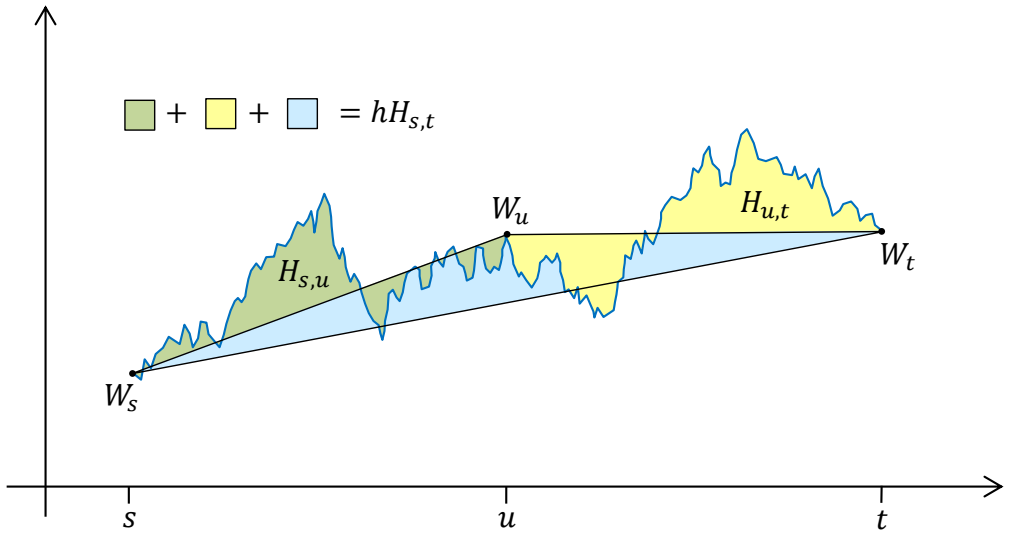

Figure 4.11: Brownian motion and its space-time area can be generated at a midpoint.

In Chapter 6, we shall apply Theorem 4.2.8 for determining when a step size is large enough to cause the local numerical error to exceed at a prespecified tolerance. For example, if the path's increment or space-time Lévy area is particularly large (which is possible as they have normal distributions), then the conditional variance computed using (4.2.5) and thus the estimated $L^2(\mathbb{P})$ error will also be considerable. In section 6.3, we will combine the new piecewise linear method for SDEs, a diagonally implicit Runge-Kutta method for ODEs and this variable step size methodology to accurately discretize the Cox-Ingersoll-Ross model (and thus also the Bessel process).

4.3 Application of conditional moments to parabola and log-ODE methods

For this section, we shall be considering the Stratonovich SDE on the interval $[0, T]$:

$$\begin{aligned} dy_t &= f_0(y_t)dt + f_1(y_t) \circ dW_t, \\ y_0 &= \xi, \end{aligned} \tag{4.3.1}$$

where $\xi \in \mathbb{R}^e$ and f_i denote bounded C^∞ vector fields on $\mathbb{R}^e$ with bounded derivatives. It follows from the standard Picard iteration or fixed point argument that there exists a unique strong solution y to the above SDE (see chapter 8 of [66] for further details). As discussed in section 2.3, an important tool in the numerical analysis of this solution is the stochastic Taylor expansion (see chapter 5 of [62] for a comprehensive review). For the purposes of this section, we only need the following specific Taylor expansion:

Theorem 4.3.1 (High order Stratonovich-Taylor expansion)**.** *Let y denote the unique strong solution to (4.3.1) and let $0 \le s \le t$. Then y_t can be Taylor expanded as follows:*

$$\begin{aligned} y_t = y_s &+ f_0(y_s)h + f_1(y_s)W_{s,t} + \frac{1}{2}f_1'(y_s)f_1(y_s)\,W_{s,t}^2 + \frac{1}{2}f_0'(y_s)f_0(y_s)h^2 \\ &+ f_0'(y_s)f_1(y_s)\int_s^t\int_s^u \circ dW_v\,du + f_1'(y_s)f_0(y_s)\int_s^t\int_s^u dv \circ dW_u \\ &+ \frac{1}{6}\big(f_1'(y_s)f_1'(y_s)f_1(y_s) + f_1''(y_s)(f_1(y_s), f_1(y_s))\big)W_{s,t}^3 \\ &+ (f_0'(y_s)f_1'(y_s)f_1(y_s) + f_0''(y_s)(f_1(y_s), f_1(y_s)))\int_s^t\int_s^u\int_s^v \circ dW_r \circ dW_v\,du \\ &+ (f_1'(y_s)f_0'(y_s)f_1(y_s) + f_1''(y_s)(f_0(y_s), f_1(y_s)))\int_s^t\int_s^u\int_s^v \circ dW_r\,dv \circ dW_u \\ &+ (f_1'(y_s)f_1'(y_s)f_0(y_s) + f_1''(y_s)(f_1(y_s), f_0(y_s)))\int_s^t\int_s^u\int_s^v dr \circ dW_v \circ dW_u \\ &+ \frac{1}{24}\big(f_1'(y_s)f_1'(y_s)f_1'(y_s)f_1(y_s) + f_1'(y_s)f_1''(y_s)(f_1(y_s), f_1(y_s)) \\ &\qquad + 3f_1''(y_s)(f_1'(y_s)f_1(y_s), f_1(y_s)) + f_1'''(y_s)(f_1(y_s), f_1(y_s), f_1(y_s))\big)W_{s,t}^4 \\ &+ R_4(h, y_s) + R_5(h, y_s), \end{aligned} \tag{4.3.2}$$

where $h := t - s$ and the remainder terms have the below uniform estimates for $h < 1$,

$$\mathbb{E}\left[R_4(h, y_s)\right] = 0, \tag{4.3.3}$$

$$\sup_{y_s \in \mathbb{R}^e} \|R_4(h, y_s)\|_{L^2(\mathbb{P})} \le C_1\, h^{\frac{5}{2}}, \tag{4.3.4}$$

$$\sup_{y_s \in \mathbb{R}^e} \|R_5(h, y_s)\|_{L^2(\mathbb{P})} \le C_2\, h^3, \tag{4.3.5}$$

and the constants C_1, C_2 depend only on the vector fields of the differential equation.

Remark 4.3.2. The derivatives appearing in the above Taylor expansion can be defined as Fréchet derivatives, and thus viewed as maps into the space of multilinear functions:

$$f^n : \mathbb{R}^e \to L\left((\mathbb{R}^e)^{\otimes n}, \mathbb{R}^e\right), \quad \text{for } n \geq 0.$$

In addition to being multilinear, each $f^n(y)$ is symmetric in its n arguments for $n \geq 2$.

Remark 4.3.3. As discussed in section 2.3, the above Taylor expansion holds when a finite order of smoothness is assumed on the vector fields (for example, it suffices to assume boundedness for the first three derivatives of f_0 and four derivatives of f_1). Moreover, stochastic Taylor expansions can be used under weaker assumptions such as linear growth (instead of boundedness) and this will be considered in future work.

From a numerical perspective, the most challenging terms appearing within (4.3.2) are those with non-trivial third order iterated integrals of Brownian motion and time. Fortunately, Theorem 4.2.8 gives a natural method for approximating these integrals. Therefore in order to propagate a numerical solution of (4.3.1) over an interval $[s, t]$, we will generate $(W_{s,t}, H_{s,t})$ exactly and then approximate $L_{s,t}$ using Theorem 4.2.8. Recall that by lemma 4.2.5, we can recover iterated integrals from $(h, W_{s,t}, H_{s,t}, L_{s,t})$. In this section, we shall be considering the following two numerical methods for (4.3.1):

Definition 4.3.4. (High order log-ODE method). For a fixed number of steps N we can construct a numerical solution $\{Y_k\}_{0 \leq k \leq N}$ of (4.3.1) by setting $Y_0 := \xi$ and for each $k \in [0 \,..\, N-1]$, defining Y_{k+1} to be the solution at $u = 1$ of the following ODE:

$$\begin{aligned} \frac{dz}{du} &= f_0(z)h + f_1(z)W_{t_k,t_{k+1}} + [f_1, f_0](z) \cdot hH_{t_k,t_{k+1}} \\ &\qquad + [f_1, [f_1, f_0]](z) \cdot \left(\frac{3}{5}hH^2_{t_k,t_{k+1}} + \frac{1}{30}h^2\right), \\ z_0 &= Y_k, \end{aligned} \tag{4.3.6}$$

where $h := \frac{T}{N}$, $t_k := kh$ and $[\cdot, \cdot]$ denotes the standard Lie bracket of vector fields.

Definition 4.3.5. (The parabola-ODE method). For a fixed number of steps N we can construct a numerical solution $\{Y_k\}_{0 \leq k \leq N}$ of (4.3.1) by setting $Y_0 := \xi$ and for each $k \in [0 \,..\, N-1]$, defining Y_{k+1} to be the solution at $u = 1$ of the following ODE:

$$\begin{aligned} \frac{dz}{du} &= f_0(z)h + f_1(z)\left(W_{t_k,t_{k+1}} + (6 - 12u)H_{t_k,t_{k+1}}\right), \\ z_0 &= Y_k, \end{aligned} \tag{4.3.7}$$

where $h := \frac{T}{N}$ and $t_k := kh$.

In the above methods, the true solution y at time t_k can be approximated by Y_k. Whilst there are different ways of interpolating between the successive approximations Y_k and Y_{k+1}, for the thesis we will simply interpolated between these points linearly. Before giving an error analysis for these numerical methods, we shall discuss how the high order log-ODE method relates to the rough path theory presented in Chapter 3.

The (Stratonovich) signature of space-time Brownian motion $X^W := \{(u, W_u)\}_{u\in[s,t]}$ is given as the following element in the vector space of all formal series of tensors, $T((\mathbb{R}^2)) := \{\boldsymbol{a} = (a_0, a_1, \cdots) : a_n \in (\mathbb{R}^2)^{\otimes n} \ \ \forall n \geq 0\}$:

$$S_{s,t}\left(X^W\right) = \left(1,\, (I_\tau, I_w)\,,\, \begin{pmatrix} I_{\tau\tau} & I_{\tau w} \\ I_{w\tau} & I_{ww} \end{pmatrix},\, \left(\begin{pmatrix} I_{\tau\tau\tau} & I_{\tau\tau w} \\ I_{\tau w\tau} & I_{\tau ww} \end{pmatrix}, \begin{pmatrix} I_{w\tau\tau} & I_{w\tau w} \\ I_{ww\tau} & I_{www} \end{pmatrix}\right), \cdots\right),$$

where we use notation presented in Theorem 4.2.7. Using the results of the previous section, we see that taking an expectation of $S_{s,t}\left(X^W\right)$ conditional on $W_{s,t}, H_{s,t}$ yields:

$$\begin{aligned}
&\mathbb{E}\left[S_{s,t}\left(X^W\right) \,\middle|\, W_{s,t}, H_{s,t}\right] \\
&= \left(1,\, (h, W_{s,t})\,,\, \begin{pmatrix} \frac{1}{2}h^2 & \frac{1}{2}hW_{s,t} - hH_{s,t} \\ \frac{1}{2}hW_{s,t} + hH_{s,t} & \frac{1}{2}W_{s,t}^2 \end{pmatrix},\right. \\
&\qquad \left(\begin{pmatrix} \frac{1}{6}h^3 & \frac{1}{6}h^2W_{s,t} - \frac{1}{2}h^2H_{s,t} \\ \frac{1}{6}h^2W_{s,t} & \frac{1}{6}hW_{s,t}^2 - \frac{1}{2}hW_{s,t}H_{s,t} + \frac{3}{5}H_{s,t}^2 + \frac{1}{30}h^2 \end{pmatrix},\right. \\
&\qquad\quad \left.\left.\begin{pmatrix} \frac{1}{6}h^2W_{s,t} + \frac{1}{2}h^2H_{s,t} & \frac{1}{6}hW_{s,t}^2 - \frac{6}{5}H_{s,t}^2 - \frac{1}{15}h^2 \\ \frac{1}{6}hW_{s,t}^2 + \frac{1}{2}hW_{s,t}H_{s,t} + \frac{3}{5}H_{s,t}^2 + \frac{1}{30}h^2 & \frac{1}{6}W_{s,t}^3 \end{pmatrix}\right), \cdots\right).
\end{aligned}$$

Since $\mathbb{E}\left[S_{s,t}\left(X^W\right) \,\middle|\, W_{s,t}, H_{s,t}\right]$ is an element of $T((\mathbb{R}^2))$ with first entry equal to 1, we can apply the logarithm map $\log(\boldsymbol{a}) := \log(a_0) + \sum_{n=1}^{\infty} \frac{(-1)^{n-1}}{n}\left(\frac{1}{a_0}\,\boldsymbol{a} - \mathbf{1}\right)^n$ to give

$$\begin{aligned}
&\log\left(\mathbb{E}\left[S_{s,t}\left(X^W\right) \,\middle|\, W_{s,t}, H_{s,t}\right]\right) \\
&= \left(0,\, (h, W_{s,t})\,,\, hH_{s,t}\begin{pmatrix} 0 & -1 \\ 1 & 0 \end{pmatrix},\, \left(\frac{3}{5}H_{s,t}^2 + \frac{1}{30}h^2\right)\left(\begin{pmatrix} 0 & 0 \\ 0 & 1 \end{pmatrix}, \begin{pmatrix} 0 & -2 \\ 1 & 0 \end{pmatrix}\right), \cdots\right) \\
&= h\, e_\tau + W_{s,t}\, e_w + hH_{s,t}\, e_{[w,\tau]} + \left(\frac{3}{5}H_{s,t}^2 + \frac{1}{30}h^2\right) e_{[w,[w,\tau]]} + 0\, e_{[[w,\tau],\tau]} + \cdots,
\end{aligned}$$

where e_τ, e_w, $e_{[w,\tau]}$, $e_{[w,[w,\tau]]}$, $e_{[[w,\tau],\tau]}$ are the first five Lyndon basis vectors in the free Lie algebra $\mathfrak{g}(\mathbb{R}^2) \subset T((\mathbb{R}^2))$ (see [83] for further details). From the above calculations, we can now give a rough path interpretation of the new log-ODE method:

Remark 4.3.6. The log-ODE (4.3.6) is constructed from the "log-expected-signature" of space-time Brownian motion conditional on $(W_{s,t}, H_{s,t})$ and truncated at level 3.

Moreover, the basis used to express this log-signature coincides with the Lyndon basis and the log-ODE true solution, z_1, is equal to the solution at time $r = h$ of the RDE:

$$dx_r = f_0\,(x_r)\,d\gamma_r^{(0)} + f_1\,(x_r)\,d\gamma_r^{(1)}, \quad x_0 = y_s. \tag{4.3.8}$$

where the rough path γ is a so-called "geodesic in the 3-step nilpotent group" with

$$\log(S_{0,h}(\gamma)) = \log\left(\mathbb{E}\left[S^3_{s,t}\left(X^W\right) \,\middle|\, W_{s,t}, H_{s,t}\right]\right).$$

This leads us to an important remark concerning the practicality of these methods.

Remark 4.3.7. Both methods can be viewed as solutions to RDEs driven by either a geodesic (in the log-ODE case) or a space-time parabola (in the parabola-ODE case). However, whilst these methods involve computing a numerical solution of an ODE, the parabola method does not require one to explicit resolve vector field derivatives. Therefore, the parabola-ODE method may be applicable to a wider class of problems.

In order to estimate the numerical errors experienced by these methods, we will compute their respective (standard) Taylor expansions up to a sufficiently high order.

Theorem 4.3.8. *Let Y^{log} be the one-step approximation defined by the log-ODE method on the interval $[s,t]$ with initial value $Y_0^{log} = y_s$. Then for sufficiently small h*

$$\begin{aligned}
Y_1^{log} = y_s &+ f_0(y_s)h + f_1(y_s)W_{s,t} + \frac{1}{2}f_1'\,f_1(y_s)W_{s,t}^2 + \frac{1}{2}f_0'\,f_0(y_s)h^2 \\
&+ f_0'\,f_1(y_s)\left(\frac{1}{2}hW_{s,t} + hH_{s,t}\right) + f_1'\,f_0(y_s)\left(\frac{1}{2}hW_{s,t} - hH_{s,t}\right) \\
&+ \frac{1}{6}(f_1'\,f_1'\,f_1 + f_1''(f_1,f_1))(y_s)W_{s,t}^3 \\
&+ (f_0'\,f_1'\,f_1 + f_0''(f_1,f_1))(y_s)\left(\frac{1}{6}hW_{s,t}^2 + \frac{1}{2}hW_{s,t}H_{s,t} + \frac{3}{5}H_{s,t}^2 + \frac{1}{30}h^2\right) \\
&+ (f_1'\,f_0'\,f_1 + f_1''(f_0,f_1))(y_s)\left(\frac{1}{6}hW_{s,t}^2 - \frac{6}{5}H_{s,t}^2 - \frac{1}{15}h^2\right) \\
&+ (f_1'\,f_1'\,f_0 + f_1''(f_1,f_0))(y_s)\left(\frac{1}{6}hW_{s,t}^2 - \frac{1}{2}hW_{s,t}H_{s,t} + \frac{3}{5}H_{s,t}^2 + \frac{1}{30}h^2\right) \\
&+ \frac{1}{24}\big(f_1'\,f_1'\,f_1'\,f_1 + f_1'\,f_1''(f_1,f_1) + 3f_1''(f_1'\,f_1,f_1) + f_1'''(f_1,f_1,f_1)\big)(y_s)W_{s,t}^4 \\
&+ R^{log}(h,y_s) + O(h^3),
\end{aligned} \tag{4.3.9}$$

where $\|R^{log}\|_{L^2(\mathbb{P})} \sim O(h^{\frac{5}{2}})$ satisfies the same $L^2(\mathbb{P})$ estimate as (4.3.4) and $\mathbb{E}\big[R^{log}\big] = 0$. In the expansion, $O(h^3)$ denotes a remainder term $R(h,y_s)$ which can be estimated as

$$\sup_{y_s\in\mathbb{R}^e} \|R(h,y_s)\|_{L^2(\mathbb{P})} \le C_R\,h^3.$$

where the constant C_R depends only on the vector fields of the differential equation.

Remark 4.3.9. Throughout this thesis (except section 6.2), we use the notation $O(h^\alpha)$ to refer to terms with an $L^2(\mathbb{P})$ norm that can be bounded by Ch^α for some $C > 0$.

Proof. In order to derive (4.3.9), we shall Taylor expand of the ODE (4.3.6) on $[0, 1]$. Let F denote the vector field defined in (4.3.6) that was constructed from f_0 and f_1:

$$\begin{aligned} F(y) &:= f_0(y)h + f_1(y)W_{s,t} + [f_1, f_0](y)\, hH_{s,t} + [f_1, [f_1, f_0]](y)\left(\frac{3}{5}H_{s,t}^2 + \frac{1}{30}h^2\right), \\ &= f_0(y)h + f_1(y)W_{s,t} + \big(f_0' f_1 - f_1' f_0\big)(y)\, hH_{s,t} + \big(f_0' f_1' f_1 - 2f_1' f_0' f_1 + f_1' f_1' f_0 \\ &\qquad + f_0''(f_1, f_1) - 2f_1''(f_0, f_1) + f_1''(f_1, f_0)\big)(y)\left(\frac{3}{5}H_{s,t}^2 + \frac{1}{30}h^2\right). \end{aligned}$$

Therefore, we see that F is a bounded C^∞ vector field on $\mathbb{R}^e$ with bounded derivatives. As F is smooth, we can apply the classical Taylor's theorem to the log-ODE (4.3.6).

$$\begin{aligned} Y_1^{\log} = y_s &+ F(y_s) + \frac{1}{2}F'F(y_s) + \frac{1}{6}F'F'F(y_s) + \frac{1}{6}F''(F,F)(y_s) + \frac{1}{24}F'F'F'F(y_s) \\ &+ \frac{1}{24}F'F''(F,F)(y_s) + \frac{1}{8}F''(F'F,F)(y_s) + \frac{1}{24}F'''(F,F,F)(y_s) \\ &+ \frac{1}{120}F'F'F'F'F(y_s) + \frac{1}{120}F'F'F''(F',F')(y_s) + \frac{1}{120}F'F'''(F,F,F)(y_s) \\ &+ \frac{1}{40}F'F''(F'F,F)(y_s) + \frac{1}{40}F''(F'F,F'F)(y_s) + \frac{1}{30}F''(F''(F,F),F)(y_s) \\ &+ \frac{1}{30}F''\left(F'F'F,F\right)(y_s) + \frac{1}{20}F'''(F'F,F,F)(y_s) + \frac{1}{120}F''''(F,F,F,F)(y_s) \\ &+ \textit{higher order terms.} \end{aligned}$$

One can define the degree of each term in the above Taylor expansion by counting the number of times functions from $\{F, F', F'', \cdots\}$ appear. Hence we can see that the higher order terms are precisely the terms with degree strictly greater than five. Since the largest component of F is $f_1(\cdot)W_{s,t}$, both F and its derivatives are $O(h^{\frac{1}{2}})$. Thus the "higher order terms" in the above Taylor expansion are $O(h^3)$ as in (4.3.5). Moreover, it follows from the definition of F that the $O(h^{\frac{5}{2}})$ component in the Taylor expansion can be expressed as a linear combination of the following random variables:

$$W_{s,t}^5,\ hW_{s,t}^3,\ h^2W_{s,t},\ hW_{s,t}^2H_{s,t},\ h^2H_{s,t},\ W_{s,t}^3H_{s,t}^2,\ hW_{s,t}H_{s,t}^2,\ hH_{s,t}^3,\ W_{s,t}H_{s,t}^4.$$

As $W_{s,t}$ and $H_{s,t}$ are independent centered Gaussian random variables, all of these terms have zero expectation. Since f_0, f_1 and their derivatives are globally bounded, we have shown that $Y_1^{\log} = M^{log} + R^{\log} + O(h^3)$ where M^{log} consists entirely of terms that are at least $O(h^2)$ and the remainder $R^{\log} + O(h^3)$ satisfies the required estimates.

As the terms of degree four that are not $O(h^{\frac{5}{2}})$ are those involving $W_{s,t}^4$, it is enough to expand only the first four terms in the expansion of $Y_1^{\log}$ using the definition of F:

$$\begin{aligned}
Y_1^{\log} = y_s &+ f_0(y_s)h + f_1(y_s)W_{s,t} + [f_1,f_0](y_s)\,hH_{s,t} + [f_1,[f_1,f_0]](y_s)\left(\frac{3}{5}H_{s,t}^2 + \frac{1}{30}h^2\right)\\
&+ \frac{1}{2}\big(f_0' h + f_1' W_{s,t} + [f_1,f_0]' hH_{s,t}\big)\big(f_0 h + f_1 W_{s,t} + [f_1,f_0]\,hH_{s,t}\big)(y_s)\\
&+ \frac{1}{6}\big(f_0' h + f_1' W_{s,t}\big)\big(f_0' h + f_1' W_{s,t}\big)\big(f_0 h + f_1 W_{s,t}\big)(y_s)\\
&+ \frac{1}{6}\big(f_0'' h + f_1'' W_{s,t}\big)\big(f_0 h + f_1 W_{s,t}, f_0 h + f_1 W_{s,t}\big)(y_s)\\
&+ \frac{1}{24}\big(f_1' f_1' f_1' f_1 + f_1' f_1''(f_1,f_1) + 3f_1''(f_1' f_1, f_1) + f_1'''(f_1,f_1,f_1)\big)(y_s)\,W_{s,t}^4\\
&+ R_1^{\log}(h,y_s) + O(h^3),
\end{aligned}$$

where $R_1^{\log}(h,y_s)$ has zero mean and satisfies the same $L^2(\mathbb{P})$ estimate as $R^{\log}(h,y_s)$. The result can now be shown by directly expanding and rearranging the above terms. For example, the coefficient which appears in front of $hW_{s,t}H_{s,t}$ can be computed as

$$\begin{aligned}
&\frac{1}{2}\big([f_1,f_0]' f_1 + f_1'[f_1,f_0]\big)(y_s)\\
&\quad= \frac{1}{2}\big((f_0' f_1 - f_1' f_0)' f_1 + f_1'(f_0' f_1 - f_1' f_0)\big)(y_s)\\
&\quad= \frac{1}{2}\big(f_0''(f_1,f_1) - f_1''(f_1,f_0) + f_0' f_1' f_1 - f_1' f_0' f_1 + f_1' f_0' f_1 - f_1' f_1' f_0\big)(y_s)\\
&\quad= \frac{1}{2}\big(f_0' f_1' f_1 + f_0''(f_1,f_1) - f_1' f_1' f_0 - f_1''(f_1,f_0)\big)(y_s)\,.
\end{aligned}$$

As the higher order terms obtained after expanding are either $O(h^{\frac{5}{2}})$ with mean zero or $O(h^3)$, it follows that $Y_1^{\log}$ has the required Taylor expansion given by (4.3.8). □

Remark 4.3.10. In practice, it is often difficult or even impossible to solve the log-ODE (4.3.6) exactly. In which case, it must be approximated using an appropriate solver. It is clear from the proof of Theorem 4.3.8 that a one-step approximation from any ODE numerical method of fourth order will produce the desired Taylor expansion and if the noise is additive (i.e. f_1 is constant) then a third order method is sufficient. Due of the structure of the log-ODE's vector field, splitting methods can be applied.

Fundamentally the idea behind this proof is an entirely algebraic one and can be extended to the general rough path setting (see Chapter 3 for further details). The exponential map of a smooth vector field V is defined as the flow obtained from solving $y' = V(y)$ over the unit interval. Thus the log-ODE method corresponds to the exponential of a vector field F that is constructed from a truncated log-signature.

In particular, the Taylor expansion for z_1 where $z' = F(z)$ can be viewed as the standard algebraic series representation of $\exp(F)$. Similarly, applying the algebraic series representation of the exponential map to the log-signature gives the signature. Since there is a natural algebra isomorphism between the space of formal tensor series $T((\mathbb{R}^d))$ and a space of differential operators generated using $V = (V_1, \cdots, V_d)$, $\exp(F)(y_s)$ can be expressed as the output of a linear function applied to the signature.

In fact, the algebra works out exactly and $\exp(F)(y_s)$ coincides precisely with the Stochastic/Rough Taylor expansion obtained from the vector fields $V = (V_1, \cdots, V_d)$ and the exponential of the input truncated log-signature in $T((\mathbb{R}^d))$. Combining this result with Chow's theorem (see section 7.4 of [37] for a statement and proof of this) then leads to the geodesic interpretation of the log-ODE method seen in remark 4.3.6.

We shall now compute a similar Taylor expansion for the parabola-ODE method.

Theorem 4.3.11. *Let Y^{para} be the one-step approximation for the parabola-ODE method on an interval $[s,t]$ with initial value $Y_0^{para} = y_s$. Then for sufficiently small h*

$$\begin{aligned} Y_1^{para} = y_s &+ f_0(y_s)h + f_1(y_s)W_{s,t} + \frac{1}{2}f_1' f_1(y_s)W_{s,t}^2 + \frac{1}{2}f_0' f_0(y_s)h^2 \\ &+ f_0' f_1(y_s)\left(\frac{1}{2}hW_{s,t} + hH_{s,t}\right) + f_1' f_0(y_s)\left(\frac{1}{2}hW_{s,t} - hH_{s,t}\right) \\ &+ \frac{1}{6}(f_1' f_1' f_1 + f_1''(f_1,f_1))(y_s)W_{s,t}^3 \\ &+ (f_0' f_1' f_1 + f_0''(f_1,f_1))(y_s)\left(\frac{1}{6}hW_{s,t}^2 + \frac{1}{2}hW_{s,t}H_{s,t} + \frac{3}{5}H_{s,t}^2\right) \\ &+ (f_1' f_0' f_1 + f_1''(f_0,f_1))(y_s)\left(\frac{1}{6}hW_{s,t}^2 - \frac{6}{5}H_{s,t}^2\right) \\ &+ (f_1' f_1' f_0 + f_1''(f_1,f_0))(y_s)\left(\frac{1}{6}hW_{s,t}^2 - \frac{1}{2}hW_{s,t}H_{s,t} + \frac{3}{5}H_{s,t}^2\right) \\ &+ \frac{1}{24}\big(f_1' f_1' f_1' f_1 + f_1' f_1''(f_1,f_1) + 3f_1''(f_1' f_1, f_1) + f_1'''(f_1,f_1,f_1)\big)(y_s)W_{s,t}^4 \\ &+ R^{para}(h, y_s) + O(h^3), \end{aligned} \tag{4.3.10}$$

where $\|R^{para}\|_{L^2(\mathbb{P})} \sim O(h^{\frac{5}{2}})$ has the same $L^2(\mathbb{P})$ estimate as (4.3.4) and $\mathbb{E}\big[R^{para}\big] = 0$. In the expansion, $O(h^3)$ denotes a remainder term $R(h, y_s)$ which can be estimated as

$$\sup_{y_s \in \mathbb{R}^e} \|R(h, y_s)\|_{L^2(\mathbb{P})} \leq C_R\, h^3.$$

where the constant C_R depends only on the vector fields of the differential equation.

Proof. Using the substitution $X_u = z_{\frac{1}{h}(u-s)}$ for $u \in [s,t]$, the ODE (4.3.7) can be rewritten as

$$\begin{aligned} dX_u &= f_0\left(X_u\right) du + f_1\left(X_u\right) d\widehat{W}_u, \\ X_s &= y_s, \end{aligned} \tag{4.3.11}$$

where $\widehat{W}$ denotes the Brownian parabola defined by $(W_{s,t}, H_{s,t})$ on the interval $[s,t]$. By emulating the derivation of the Stratonovich-Taylor expansion (4.3.2), it is possible to Taylor expand (4.3.11) in the same fashion. The only difference is that Stratonovich integrals with respect to W are replaced by Riemann-Stieltjes integrals against $\widehat{W}$. Note that this type is argument is common in rough path theory, and leads to the so-called rough Taylor expansion (see [9] for a recent account of this expansion that gives a uniform factorial decay estimate for the remainder in the p-variation metric).

In our setting, the rough path driving the differential equation is the "space-time" Brownian parabola $X^{\widehat{W}} := \{(u, \widehat{W}_u)\}_{u\in[s,t]}$. However, we are interested in estimating the remainder term in an $L^2(\mathbb{P})$ sense and not a rough path sense (such as p-variation). Thus, the Taylor expansion of the ODE (4.3.11) can be written in the following form:

$$\begin{aligned} Y_1^{\text{para}} = y_s &+ f_0(y_s)h + f_1(y_s)\widehat{W}_{s,t} + \frac{1}{2} f_1' \, f_1(y_s)\widehat{W}_{s,t}^2 + \frac{1}{2} f_0' \, f_0(y_s)h^2 \\ &+ f_0' \, f_1(y_s) \int_s^t \int_s^u d\widehat{W}_v \, du + f_1' \, f_0(y_s) \int_s^t \int_s^u dv \, d\widehat{W}_u \\ &+ \frac{1}{6}\big(f_1' \, f_1' \, f_1 + f_1''(f_1, f_1)\big)(y_s)\widehat{W}_{s,t}^3 \\ &+ (f_0' \, f_1' \, f_1 + f_0''(f_1, f_1))(y_s) \int_s^t \int_s^u \int_s^v d\widehat{W}_r \, d\widehat{W}_v \, du \\ &+ (f_1' \, f_0' \, f_1 + f_1''(f_0, f_1))(y_s) \int_s^t \int_s^u \int_s^v d\widehat{W}_r \, dv \, d\widehat{W}_u \\ &+ (f_1' \, f_1' \, f_0 + f_1''(f_1, f_0))(y_s) \int_s^t \int_s^u \int_s^v dr \, d\widehat{W}_v \, d\widehat{W}_u \\ &+ \frac{1}{24}\big(f_1' \, f_1' \, f_1' \, f_1 + f_1' \, f_1''(f_1, f_1) + 3f_1''(f_1' \, f_1, f_1) + f_1'''(f_1, f_1, f_1)\big)(y_s)\widehat{W}_{s,t}^4 \\ &+ R^{\text{para}}(h, y_s) + \textit{higher order terms}, \end{aligned} \tag{4.3.12}$$

where $R^{\text{para}}(h, y_s)$ consists of terms corresponding to the following iterated integrals:

$$\begin{aligned} &\widehat{W}_{s,t}^5, \int_s^t \int_s^u \int_s^v dr \, dv \, d\widehat{W}_u, \int_s^t \int_s^u \int_s^v dr \, d\widehat{W}_v \, du, \int_s^t \int_s^u \int_s^v d\widehat{W}_r \, dv \, du, \\ &\int_s^t \int_s^u \int_s^v \int_s^r dw \, d\widehat{W}_r \, d\widehat{W}_v \, d\widehat{W}_u, \int_s^t \int_s^u \int_s^v \int_s^r d\widehat{W}_w \, dr \, d\widehat{W}_v \, d\widehat{W}_u, \\ &\int_s^t \int_s^u \int_s^v \int_s^r d\widehat{W}_w \, d\widehat{W}_r \, dv \, d\widehat{W}_u, \int_s^t \int_s^u \int_s^v \int_s^r d\widehat{W}_w \, d\widehat{W}_r \, d\widehat{W}_v \, du. \end{aligned}$$

Since $\widehat{W}_u$ can be expressed as a linear combination of $W_{s,t}$ and $H_{s,t}$, it follows that all of the above iterated integrals have zero expectation. Therefore, $\mathbb{E}\left[R^{\text{para}}(h, y_s)\right] = 0$. Moreover, by the natural Brownian scaling, we see that these integrals are all $O(h^{\frac{5}{2}})$. This implies that $R^{\text{para}} \sim O(h^{\frac{5}{2}})$ satisfies the same type of $L^2(\mathbb{P})$ estimate as (4.3.4). By lemma 4.2.5 and some calculations from the proof of Theorem 4.2.8, we have that

$$\begin{aligned}
\int_s^t \int_s^u \int_s^v d\widehat{W}_r \, d\widehat{W}_v \, du &= \frac{1}{6} h W_{s,t}^2 + \frac{1}{2} h W_{s,t} H_{s,t} + \frac{3}{5} H_{s,t}^2, \\
\int_s^t \int_s^u \int_s^v d\widehat{W}_r \, dv \, d\widehat{W}_u &= \frac{1}{6} h W_{s,t}^2 - \frac{6}{5} H_{s,t}^2, \\
\int_s^t \int_s^u \int_s^v dr \, d\widehat{W}_v \, d\widehat{W}_u &= \frac{1}{6} h W_{s,t}^2 - \frac{1}{2} h W_{s,t} H_{s,t} + \frac{3}{5} H_{s,t}^2.
\end{aligned}$$

Finally, it follows from Brownian scaling that the "higher order terms" are $O(h^3)$. □

Remark 4.3.12. The key difference between the Taylor expansions of the two methods are the terms corresponding to certain triple integrals of Brownian motion and time. In particular, after comparing the expansions for the log-ODE and parabola-ODE methods with the Stratonovich-Taylor expansion of the original SDE (4.3.1), we have

$$Y_1^{\log} = y_t - [f_1, [f_1, f_0]](y_s)\left(L_{s,t} - \mathbb{E}\left[L_{s,t} \middle| W_{s,t}, H_{s,t}\right]\right) + E^{\log}(h, y_s), \tag{4.3.13}$$

$$Y_1^{\text{para}} = y_t - [f_1, [f_1, f_0]](y_s)\left(L_{s,t} - \frac{3}{5} h H_{s,t}^2\right) + E^{\text{para}}(h, y_s), \tag{4.3.14}$$

with high order errors satisfying $E^{\log}, E^{\text{para}} \sim O(h^{\frac{5}{2}})$ and $\mathbb{E}\left[E^{\log}\right], \mathbb{E}\left[E^{\text{para}}\right] \sim O(h^3)$.

The above remark shows that both methods give a one-step local error of $O(h^2)$. This means the log-ODE and parabola-ODE methods are both locally high order; however there is a significant difference in how the methods propagate local errors. The reason is that the $O(h^2)$ parts of the log-ODE local errors give a martingale, whilst the $O(h^2)$ part for each parabola-ODE local error has non-zero expectation. Thus the log-ODE method is globally high order whilst the parabola method is not. However, since the parabola-ODE method is straightforward to implement and locally high order, one could expect it to perform well compared to other low order methods. In the numerical example, we shall see that the parabola method has the same order of convergence as the standard piecewise linear approach but gives much smaller errors.

We will now present the convergence rates for the parabola and log-ODE methods. Recall that the notions of weak and strong convergence (with order α) for numerical solutions of SDEs were given in Chapter 2 by definitions 2.2.1 and 2.2.3 respectively.

Theorem 4.3.13 (Orders of convergence)**.** *For a general SDE (4.3.1), the log-ODE method converges in a strong sense with order 1.5 and a weak sense with order 2.0. The parabola-ODE method converges in both a strong and weak sense with order 1.0.*

Proof. Taylor expansions for these methods are given in Theorems 4.3.8 and 4.3.11. The strong convergence can then be shown as in the proof of Theorem 11.5.1 in [62]. Moreover, the proof of Theorem 11.5.1 also provides the orders of strong convergence. Since these methods converge strongly to the true solution, they also converge weakly. Additionally, the weak convergence (with rates) follow from the Taylor expansions (4.3.9), (4.3.10) and can be established as in the proof of Theorem 14.5.2 in [62]. Although we have not presented a detailed self-contained proof (which would require further estimates such as Gröwall's inequality), we will highlight the following ideas:

1. Since the largest part of the log-ODE local error is $O(h^2)$ and has mean zero, the local errors propagate like a random walk and give a strong error of $O(h^{\frac{3}{2}})$.

2. In addition, the $O(h^{\frac{5}{2}})$ component of the log-ODE local error has zero mean. Therefore each moment of the one-step approximation has an error of $O(h^3)$. These errors propagate at least linearly and provide the $O(h^2)$ weak convergence.

3. For the parabola-ODE method, the expected value of the local error is $O(h^2)$ provided that the two vectors fields do not satisfy the commutativity condition:

$$[f_1, [f_1, f_0]] = 0.$$

   The $O(h^2)$ bias in each one-step approximation causes local errors to propagate linearly and produces the $O(h)$ convergence rates for the parabola-ODE method.

It is also worth noting that these arguments (and theorems in [62]) still apply if the initial condition ξ is a random variable that satisfies certain moment conditions. □

Remark 4.3.14. This result can be extended naturally to the multidimensional case:

$$\begin{aligned} dy_t &= f_0(y_t)\, dt + \sum_{i=1}^{d} f_i(y_t) \circ dW_t^{(i)}, \\ y_0 &= \xi, \end{aligned}$$

where W is a standard d-dimensional Brownian motion and the smooth vector fields f_i are assumed to have bounded derivatives and satisfy the commutativity conditions

$$[f_i, f_j] = 0, \quad \forall\, i, j \in \{1, \cdots, d\}. \tag{4.3.15}$$

Note that this includes the SDEs with additive noise (i.e. f_i is constant for $i \geq 1$). In this setting, each step of the numerical methods require generating d independent copies of $(W_{s,t}, H_{s,t})$ for each coordinate of W. The parabola-ODE and log-ODE methods can then be extended by replacing the any term constructed from f_1 with the corresponding sum of d terms that involve $\{f_1, \cdots, f_d\}$ and $\left\{\left(W_{s,t}^{(i)}, H_{s,t}^{(i)}\right) : 1 \leq i \leq d\right\}$. Due to the condition (4.3.15), the convergence rates given in Theorem 4.3.13 still hold.

Although the parabola-ODE method has a low order of convergence, it is in some sense "close" to being high order. Specifically, if one corrects for the local bias of $\frac{1}{30}[f_1, [f_1, f_0]]h^2$ after each step then the resulting numerical method would enjoy high orders of both strong and weak convergence. That said, this approach would require computing vector field derivatives (which was a key feature of the parabola method). This suggests that whilst the parabola gives an unbiased and effective approximation of a Brownian path, it may not be the most suitable path for general SDE simulation.

In Chapter 5, we construct piecewise linear paths that compensate for this $O(h^2)$ bias and thus solving the ODE driven by these paths gives high order methodologies. As one would expect, this also extends to the commutative multidimensional setting.

To conclude this section, we shall briefly discuss some topics related to this work.

Interplay between ODE approximations of SDEs and numerical ODE solvers

A central theme throughout this thesis is that SDEs can be well approximated by solutions of (random) ODEs. However for this to lead to practical methodologies, one needs to also consider the numerical method that is used to discretize each ODE.

In the case where vector fields are smooth, bounded and with bounded derivatives we have that the random coefficients of each ODE have finite moments of all orders (since any polynomial function of Gaussian random variables has finite moments). Therefore, when we apply a Runge-Kutta method to the random ODE (and consider its Taylor expansion), we should obtain an error term which has a finite $L^2(\mathbb{P})$ norm. Moreover any fourth order Runge-Kutta method will have a Taylor expansion with the desired terms and therefore should produce an error term that is $O(h^{\frac{5}{2}})$ in $L^2(\mathbb{P})$.

On the other hand, for certain problems these random ODEs could become stiff and thus standard explicit Runge-Kutta methods may lead to numerical instability. In Chapters 5 and 6, we give two numerical examples of SDEs where this is the case:

- The Langevin diffusion moving in a locally Lipschitz potential (section 5.3). Here, the SDE can become stiff whenever the gradient of the potential is large.

- The Cox-Ingersoll-Ross and squared Bessel processes (section 6.3). These SDEs become stiff near zero as the derivatives of the square root vector field blow-up.

In the ODE literature, implicit methods and variable step size methods are used for tackling stiff problems [15]. Similarly, such ideas apply to the SDE setting [62]. However, in our setting there are two ways in which variable step sizes can be used:

1. We discretize the SDE using variable step sizes. This would involve generating increments (and areas) of the Brownian motion over a non-uniform partition. The problem of generating W in an adaptive manner is discussed in Chapter 6.

2. We discretize each random ODE using variable step sizes (for the ODE solver). This does not require generating further information about the Brownian path.

It is worth noting that these approaches are not mutually exclusive and can be simultaneously applied to tackle the same problem. That said, we take the view that (1) is the more important type of step size and so we prefer to avoid the use of (2). For example, when discretizing the CIR model, we use variable step sizes for the SDE but solve each ODE using one step of an implicit third order Runge-Kutta method. We prefer step sizes of type (1) simply because it intuitively feels more appealing to generate further information about the Brownian path. Thus understanding the trade-off between these two different types of variable step size is an open problem.

Although it is not a central focus of this thesis, the use of implicit ODE solvers is an important consideration for stiff problems. In particular, implicit Runge-Kutta methods can be "A-stable" (that is, they exhibit the correct long-term behaviour when applied to $y' = -\lambda y$ with $\lambda > 0$) and are therefore desirable for stiff ODEs [15]. For the CIR model, we apply an A-stable diagonally implicit Runge-Kutta method (in addition, this method also preserves the non-negativity of the numerical solution).

Comparison to well-known SDE schemes such as stochastic Taylor methods

It is reasonable to compare the numerical methods proposed in this thesis with other strong order 1.5 methods (such as those given in [84] and section 11.2 of [62]). Whilst all these methods achieve the same order of strong convergence, previously studied methods do not utilize expectations of space-space-time Lévy area that are conditional on both increment and integral information about the Brownian path. So, everything else being equal, we expect previous methodologies to exhibit greater local error for each one-step approximation (asymptotically and in an $L^2(\mathbb{P})$ sense). Quantifying the improvement in accuracy gained from these new (optimal) estimators of $L_{s,t}$ as well as further comparisons with existing schemes is a topic of future research.

Extension to more complicated models with piecewise smooth vector fields

In mathematical finance, it is reasonable to consider SDEs with vector fields that have discontinuities in time. In particular, so-called "term structure" models for interest rates may use piecewise constant or piecewise linear coefficients - which allows them to be calibrated against multiple bond prices with different maturity dates [11]. For example, in [56], the authors propose the following extension of the CIR model:

$$dr_t = \big(a(t)(b - r_t) + c(t)\big)\, dt + \sigma(t)\sqrt{r_t}\, dW_t,$$

where the mean reversion $a(t), c(t)$ and volatility $\sigma(t)$ coefficients can be time-varying. For problems where vector fields (or their derivatives) have discontinuities in time, we would simply compute a numerical solution between each point of discontinuity. Letting $\{\tau_n\}$ be the sequence of times where a discontinuity occurs and provided the SDE's vector fields are sufficiently smooth on each interval (τ_n, τ_{n+1}), we may apply a high order approximation. For each such interval, we would continuously extend the vector fields from (τ_n, τ_{n+1}) to $[\tau_n, \tau_{n+1}]$ before defining our ODE approximation. As usual, a high order Runge-Kutta method can then be applied to solve each ODE.

4.4 A Runge-Kutta method for the parabola-ODE when noise is additive

In this section, we shall develop a numerical method for SDEs of the following form:

$$dy_t = f(y_t)\, dt + \sigma\, dW_t, \tag{4.4.1}$$
$$y_0 = \xi,$$

where $\xi \in \mathbb{R}^d$, W is a standard d-dimensional Brownian motion, $\sigma > 0$ is a constant and f will denote a bounded smooth vector field on $\mathbb{R}^d$ that has bounded derivatives.

In [84], the author develops stochastic Runge-Kutta methods for (4.4.1) which converge strongly with order 1.5 and do not require us to evaluate derivatives of f. However whilst these methods use the increment $W_{s,t}$ and space-time Lévy area $H_{s,t}$ of the Brownian motion, they do not optimally approximate (in an $L^2(\mathbb{P})$ sense) higher iterated integrals using the pair $(W_{s,t}, H_{s,t})$. In this section, we will develop a Runge-Kutta method for (4.4.1) based on some optimal estimators of such integrals.

Due to the structure of these integral estimators (see Theorem 4.4.1), the proposed Runge-Kutta method also requires evaluations/approximations of the Laplacian Δf. Whilst this is not ideal, Δf is easier to compute than f'' and scales linearly with d, which may be important for applications such as the overdamped Langevin equation.

To discretize (4.4.1), we consider a three-stage Runge-Kutta method of the form:

$$Y_t := Y_s + b_1 k_1 + b_2 k_2 + b_3 k_3, \tag{4.4.2}$$

where $b_1, b_2, b_3 \in \mathbb{R}$ and the stages that we use to compute the numerical solution are

$$k_1 := h f(Y_s) + \sigma W_{s,t} + d_1 \sigma H_{s,t}, \tag{4.4.3}$$

$$k_2 := h f(Y_s + a_{21} k_1) + \sigma W_{s,t} + d_2 \sigma H_{s,t}, \tag{4.4.4}$$

$$k_3 := h f(Y_s + a_{31} k_1 + a_{32} k_2) + \sigma W_{s,t} + d_3 \sigma H_{s,t}, \tag{4.4.5}$$

where $W_{s,t}$ and $H_{s,t}$ are now defined as vectors in $\mathbb{R}^d$ and the coefficients are all real. One can express the method (4.4.2) more concisely using the Butcher tableau below:

$$\begin{array}{c|ccc|c} 0 & & & & d_1 \\ c_2 & a_{21} & & & d_2 \\ c_3 & a_{31} & a_{32} & & d_3 \\ \hline & b_1 & b_2 & b_3 & \end{array}$$

As the noise in (4.4.1) is additive, it can be interpreted in a Stratonovich sense. Therefore, we have the following high order Stratonovich-Taylor expansion for (4.4.1):

$$\begin{aligned} y_t = y_s + \sigma W_{s,t} + f(y_s)h + \frac{1}{2} f'(y_s) f(y_s) h^2 + \sigma f'(y_s) \int_s^t \int_s^u \circ dW_v \, du \\ + \sigma^2 f''(y_s) \int_s^t \int_s^u \int_s^v \circ dW_r \otimes \circ dW_v \, du \\ + \sigma f''(y_s) \left(f(y_s) \otimes \int_s^t \int_s^u \int_s^v \circ dW_r \, dv \, du \right) \\ + \sigma f'(y_s) f'(y_s) \int_s^t \int_s^u \int_s^v \circ dW_r \, dv \, du \\ + \sigma f''(y_s) \left(f(y_s) \otimes \int_s^t \int_s^u \int_s^v dr \circ dW_v \, du \right) + R_4(h, y_s) + R_5(h, y_s), \end{aligned} \tag{4.4.6}$$

where the two remainder terms will satisfy the below (uniform) estimates for $h < 1$,

$$\begin{aligned} \mathbb{E}\left[R_4(h, y_s) \right] = 0, \\ \sup_{y_s \in \mathbb{R}^d} \| R_4(h, y_s) \|_{L^2(\mathbb{P})} \leq C_1 \, h^{\frac{5}{2}}, \\ \sup_{y_s \in \mathbb{R}^d} \| R_5(h, y_s) \|_{L^2(\mathbb{P})} \leq C_2 \, h^3, \end{aligned}$$

and the constants C_1 and C_2 depend only on the smooth vector field f as well as σ. Tensor notation can be used here since the derivative $f''(\,\cdot\,)$ is a bilinear form on $\mathbb{R}^d$. (see chapter 5 of [62] for a review of multidimensional stochastic Taylor expansions)

As in the previous section, we would like to design numerical methods that match the conditional expectation of the Taylor expansion (4.4.6) up to a certain high order. Therefore, our first step is to derive the unbiased estimators for the various integrals.

Theorem 4.4.1. (Conditional expectations of multidimensional Brownian integrals)

$$\int_s^t W_{s,u}\,du = \frac{1}{2}hW_{s,t} + hH_{s,t}, \tag{4.4.7}$$

$$\mathbb{E}\left[\int_s^t\int_s^u\int_s^v \circ dW_r \otimes \circ dW_v\,du \,\middle|\, W_{s,t}, H_{s,t}\right] = \frac{1}{6}hW_{s,t}^{\otimes 2} + \frac{1}{2}hH_{s,t}\otimes W_{s,t} + \frac{3}{5}hH_{s,t}^{\otimes 2} + \frac{1}{30}h^2 I_d, \tag{4.4.8}$$

$$\mathbb{E}\left[\int_s^t\int_s^u\int_s^v \circ dW_r\,dv\,du \,\middle|\, W_{s,t}, H_{s,t}\right] = \frac{1}{6}h^2W_{s,t} + \frac{1}{2}h^2H_{s,t}, \tag{4.4.9}$$

$$\mathbb{E}\left[\int_s^t\int_s^u\int_s^v dr\ \circ dW_v\,du \,\middle|\, W_{s,t}, H_{s,t}\right] = \frac{1}{6}h^2W_{s,t}, \tag{4.4.10}$$

where I_d denotes the $d\times d$ identity matrix.

Proof. The first equality was addressed by remark 4.2.2 and used in previous sections. The last two equalities follow by lemma 4.2.6 and the independence of $(W_{s,t}, H_{s,t}, K_{s,t})$. The second conditional expectation has already been proved in the one-dimensional case by Theorem 4.2.8 (along with lemma 4.2.5). Thus, all that remains is to compute

$$\mathbb{E}\left[\int_s^t\int_s^u\int_s^v \circ dW_r^{(i)} \circ\ dW_v^{(j)}\,du \,\middle|\, W_{s,t}, H_{s,t}\right],$$

when $i\neq j$.

Similar to the proof of Theorem 4.2.8, we shall use the decomposition $W = \widehat{W} + Z$, where $\widehat{W}$ is the (multidimensional) Brownian parabola and Z is the associated arch. So by the linearity of conditional expectations and Stratonovich integration, we have

$$\begin{aligned}
&\mathbb{E}\left[\int_s^t\int_s^u\int_s^v \circ dW_r^{(i)} \circ\ dW_v^{(j)}\,du \,\middle|\, W_{s,t}, H_{s,t}\right]\\
&= \mathbb{E}\left[\int_s^t\int_s^u\int_s^v \circ d\left(\widehat{W}_r^{(i)} + Z_r^{(i)}\right) \circ\ d\left(\widehat{W}_r^{(j)} + Z_r^{(j)}\right)\,du \,\middle|\, W_{s,t}, H_{s,t}\right]\\
&= \int_s^t\int_s^u\int_s^v d\widehat{W}_r^{(i)}\,d\widehat{W}_v^{(j)}\,du + \mathbb{E}\left[\int_s^t\int_s^u\int_s^v \circ dZ_r^{(i)} \circ\ dZ_r^{(j)}\,du \,\middle|\, W_{s,t}, H_{s,t}\right]\\
&\quad + \mathbb{E}\left[\int_s^t\int_s^u\int_s^v d\widehat{W}_r^{(i)} \circ\ dZ_r^{(j)}\,du + \int_s^t\int_s^u\int_s^v \circ dZ_r^{(i)}\,d\widehat{W}_r^{(j)}\,du \,\middle|\, W_{s,t}, H_{s,t}\right].
\end{aligned}$$

As Z is symmetric and $\widehat{W}$ is determined by $(W_{s,t}, H_{s,t})$, the last line gives zero. The other expectation is zero as $Z^{(i)}$ and $Z^{(j)}$ are independent symmetric processes. Therefore, the result follows by explicitly computing the remaining iterated integral.

$$\begin{aligned}
&\mathbb{E}\left[\int_s^t \int_s^u \int_s^v \circ dW_r^{(i)} \circ dW_v^{(j)}\, du \,\middle|\, W_{s,t}, H_{s,t} \right] \\
&\quad = \int_s^t \int_s^u \int_s^v d\widehat{W}_r^{(i)}\, d\widehat{W}_v^{(j)}\, du \\
&\quad = \frac{1}{6} h W_{s,t}^{(i)} W_{s,t}^{(j)} + \frac{1}{h^3} W_{s,t}^{(i)} H_{s,t}^{(j)} \int_s^t \int_s^u (v-s)\, d\big(6(t-v)(v-s)\big)\, du \\
&\qquad + \frac{1}{h^3} H_{s,t}^{(i)} W_{s,t}^{(j)} \int_s^t \int_s^u 6(t-v)(v-s)\, dv\, du \\
&\qquad + \frac{1}{h^4} H_{s,t}^{(i)} H_{s,t}^{(j)} \int_s^t \int_s^u 6(t-v)(v-s)\, d\big(6(t-v)(v-s)\big)\, du \\
&\quad = \frac{1}{6} h W_{s,t}^{(i)} W_{s,t}^{(j)} + \frac{1}{2} h H_{s,t}^{(i)} W_{s,t}^{(j)} + \frac{3}{5} h H_{s,t}^{(i)} H_{s,t}^{(j)},
\end{aligned}$$

for $i, j \in \{1, \cdots d\}$ with $i \neq j$. □

Therefore, we would like the method (4.4.2) to have the following Taylor expansion:

$$\begin{aligned}
Y_t = Y_s &+ \sigma W_{s,t} + f(Y_s)h + \frac{1}{2} f'(Y_s) f(Y_s) h^2 + \sigma f'(Y_s)\left(\frac{1}{2} h W_{s,t} + h H_{s,t}\right) \\
&+ \sigma^2 f''(Y_s)\left(\frac{1}{6} h W_{s,t}^{\otimes 2} + \frac{1}{2} h H_{s,t} \otimes W_{s,t} + \frac{3}{5} h H_{s,t}^{\otimes 2} + \frac{1}{30} h^2 I_d\right) \\
&+ \sigma f''(Y_s)\left(f(Y_s) \otimes \left(\frac{1}{3} h^2 W_{s,t} + \frac{1}{2} h^2 H_{s,t}\right)\right) \\
&+ \sigma f'(Y_s) f'(Y_s)\left(\frac{1}{6} h^2 W_{s,t} + \frac{1}{2} h^2 H_{s,t}\right) \\
&+ \textit{higher order terms}.
\end{aligned} \tag{4.4.11}$$

Since $\frac{1}{30} h^2 \sigma^2 f''(Y_s)(I_d)$ is a challenging term that can be computed afterwards using the stages k_1, k_2, k_3, the Runge-Kutta method will actually have the expansion:

$$\begin{aligned}
Y_t = Y_s &+ \sigma W_{s,t} + f(Y_s)h + \frac{1}{2} f'(Y_s) f(Y_s) h^2 + \sigma f'(Y_s)\left(\frac{1}{2} h W_{s,t} + h H_{s,t}\right) \\
&+ \sigma^2 f''(Y_s)\left(\frac{1}{6} h W_{s,t}^{\otimes 2} + \frac{1}{4} h H_{s,t} \otimes W_{s,t} + \frac{1}{4} h W_{s,t} \otimes H_{s,t} + \frac{3}{5} h H_{s,t}^{\otimes 2}\right) \\
&+ \sigma f''(Y_s)\left(f(Y_s) \otimes \left(\frac{1}{6} h^2 W_{s,t} + \frac{1}{4} h^2 H_{s,t}\right) + \left(\frac{1}{6} h^2 W_{s,t} + \frac{1}{4} h^2 H_{s,t}\right) \otimes f(Y_s)\right) \\
&+ \sigma f'(Y_s) f'(Y_s)\left(\frac{1}{6} h^2 W_{s,t} + \frac{1}{2} h^2 H_{s,t}\right) + \textit{higher order terms}.
\end{aligned} \tag{4.4.12}$$

Since $f''(\,\cdot\,)$ is symmetric, the rearrangements that were performed are acceptable. By Taylor expanding the stages (4.4.4) and (4.4.5) using tensor notation, we obtain

$$
\begin{aligned}
k_1 &= \sigma W_{s,t} + d_1\sigma H_{s,t} + hf(Y_s),\\
k_2 &= \sigma W_{s,t} + d_2\sigma H_{s,t} + hf(Y_s + a_{21}k_1)\\
&= \sigma W_{s,t} + d_2\sigma H_{s,t} + hf(Y_s) + hf'(Y_s)(a_{21}k_1) + \frac{1}{2}hf''(Y_s)(a_{21}k_1)^{\otimes 2}\\
&\qquad + \textit{higher order terms}\\
&= \sigma W_{s,t} + d_2\sigma H_{s,t} + hf(Y_s) + a_{21}hf'(Y_s)(\sigma W_{s,t} + d_1\sigma H_{s,t} + hf(Y_s))\\
&\qquad + \frac{1}{2}a_{21}^2 hf''(Y_s)\left(\sigma W_{s,t} + d_1\sigma H_{s,t} + hf(Y_s)\right)^{\otimes 2} + \textit{higher order terms}.\\
k_3 &= \sigma W_{s,t} + d_3\sigma H_{s,t} + hf(Y_s + a_{31}k_1 + a_{32}k_2)\\
&= \sigma W_{s,t} + d_3\sigma H_{s,t} + hf(Y_s) + hf'(Y_s)(a_{31}k_1 + a_{32}k_2) + \frac{1}{2}hf''(Y_s)(a_{31}k_1 + a_{32}k_2)^{\otimes 2}\\
&\qquad + \textit{higher order terms}\\
&= \sigma W_{s,t} + d_3\sigma H_{s,t} + hf(Y_s)\\
&\qquad + hf'(Y_s)((a_{31}+a_{32})(\sigma W_{s,t} + hf(Y_s)) + (a_{31}d_1 + a_{32}d_2)\sigma H_{s,t})\\
&\qquad + a_{32}a_{21}h^2 f'(Y_s)f'(Y_s)(\sigma W_{s,t} + d_1\sigma H_{s,t})\\
&\qquad + \frac{1}{2}hf''(Y_s)((a_{31}+a_{32})(\sigma W_{s,t} + hf(Y_s)) + (a_{31}d_1 + a_{32}d_2)\sigma H_{s,t})^{\otimes 2}\\
&\qquad + \textit{higher order terms}.
\end{aligned}
$$

Since the expansion of (4.4.2) will have 16 individual terms (and Y_s), there are initially 16 order conditions that we would like to be satisfied by the 9 coefficients. That said, there are several terms in these expansions that give the same conditions.

After removing such redundancies, we are left with the following order conditions:

$$b_1 + b_2 + b_3 = 1, \tag{4.4.13}$$

$$b_1 d_1 + b_2 d_2 + b_3 d_3 = 0, \tag{4.4.14}$$

$$b_2 a_{21} + b_3(a_{31} + a_{32}) = \frac{1}{2}, \tag{4.4.15}$$

$$b_2 a_{21} d_1 + b_3(a_{31}d_1 + a_{32}d_2) = 1, \tag{4.4.16}$$

$$b_2 a_{21}^2 + b_3(a_{31} + a_{32})^2 = \frac{1}{3}, \tag{4.4.17}$$

$$b_2 a_{21}^2 d_1 + b_3(a_{31} + a_{32})(a_{31} d_1 + a_{32} d_2) = \frac{1}{2}, \tag{4.4.18}$$

$$b_2 a_{21}^2 d_1^2 + b_3(a_{31} d_1 + a_{32} d_2)^2 = \frac{6}{5}, \tag{4.4.19}$$

$$b_3 a_{32} a_{21} = \frac{1}{6}, \tag{4.4.20}$$

$$b_3 a_{32} a_{21} d_1 = \frac{1}{2}. \tag{4.4.21}$$

As the number of equations and variables are the same, we hope that a solution exists.

Theorem 4.4.2. *The conditions (4.4.13)–(4.4.21) have the following unique solution*

$$\begin{array}{c|cccc} 0 & & & & 3 \\ \frac{3}{7} & \frac{3}{7} & & & \frac{12}{7} \\ 1 & -\frac{13}{15} & \frac{28}{15} & & -\frac{36}{5} \\ \hline & \frac{1}{9} & \frac{49}{72} & \frac{5}{24} & \end{array} \tag{4.4.22}$$

Proof. It is clear that the two order conditions (4.4.20) and (4.4.21) imply that $d_1 = 3$. To make the computations easier to handle, we shall use the change of variables below.

$$c_3 := a_{31} + a_{32}, \tag{4.4.23}$$

$$x := 3a_{31} + a_{32} d_2, \tag{4.4.24}$$

With the above variables, the four order conditions (4.4.15)–(4.4.18) will simplify to

$$b_2 a_{21} + b_3 c_3 = \frac{1}{2}, \tag{4.4.25}$$

$$b_2 a_{21}^2 + b_3 c_3^2 = \frac{1}{3}, \tag{4.4.26}$$

$$3b_2 a_{21} + b_3 x = 1, \tag{4.4.27}$$

$$3b_2 a_{21}^2 + b_3 c_3 x = \frac{1}{2}. \tag{4.4.28}$$

By multiplying the first two conditions by 3 and then subtracting the others, we have

$$3b_3 c_3 - b_3 x = \frac{1}{2},$$

$$3b_3 c_3^2 - b_3 c_3 x = \frac{1}{2}.$$

Dividing the second equation by the first gives $c_3 = 1$ and therefore $b_3 = \frac{1}{2(3-x)}$. Substituting these values for c_3 and b_3 into the equations (4.4.25) and (4.4.26) yields

$$b_2 a_{21} + \frac{1}{2(3-x)} = \frac{1}{2},$$

$$b_2 a_{21}^2 + \frac{1}{2(3-x)} = \frac{1}{3}.$$

By multiplying the top equation by a_{21}, we can derive the following equation for a_{21}.

$$\frac{1}{2(3-x)}(a_{21}-1) = \frac{1}{2}a_{21} - \frac{1}{3}.$$

Rearranging the above shows that $a_{21} = \frac{2x-3}{3(x-2)}$ and therefore gives $b_2 = \frac{3(x-2)^2}{2(x-3)(2x-3)}$. We can now substitute the values for b_3 and a_{21} into condition (4.4.20) to obtain a_{32}.

$$a_{32} = \frac{(x-3)(x-2)}{3-2x}.$$

As $c_3 = 1$, it follows by (4.4.23) that $a_{31} = 1 - a_{32}$. Thus, we have $a_{31} = \frac{x^2-3x+3}{2x-3}$. Using these values along with the definition of x, we can derive the following equation,

$$d_2 = \frac{3-2x}{(x-3)(x-2)}\left(x - 3 \cdot \frac{x^2-3x+3}{2x-3}\right) = \frac{x-3}{x-2}.$$

Conditions (4.4.13) and (4.4.14) now allow us to determine b_1 and d_3 in terms of x.

$$b_1 = \frac{x-1}{2(2x-3)}, \quad d_3 = 3(x-3).$$

Substituting x and the values of a_{21}, b_2, b_3, d_1 into the order condition (4.4.19) yields

$$\frac{27(x-2)^2}{2(x-3)(2x-3)}\left(\frac{2x-3}{3(x-2)}\right)^2 + \frac{1}{2(3-x)}x^2 = \frac{6}{5}.$$

It is now straightforward to see that the unique solution to the above is $x = \frac{3}{5}$. □

The Runge-Kutta method that is defined by the Butcher tableau (4.4.22) will now produce the Taylor expansion given by (4.4.12). In particular, it is worth noting that this coincides precisely with the expansion of the parabola-ODE up to the third level. Hence, we have derived a three-stage Runge-Kutta method for discretizing the ODE:

$$dY_t = f(Y_t)\,dt + \sigma\,d\widehat{W}_t, \tag{4.4.29}$$

Therefore, we would expect this three-stage Runge-Kutta method to have the same low order of convergence as the parabola-ODE method did in the previous section. On the other hand, by accounting for the local $O(h^2)$ bias in each step, the numerical method should achieve the same high orders of convergence as the log-ODE method.

Recall that the $O(h^2)$ correction term that should be computed after each step is

$$\frac{1}{30}h^2\sigma^2 f''(Y_s)(I_d) = \frac{1}{30}h^2\sigma^2\Delta f(Y_s), \tag{4.4.30}$$

which can be estimated by taking finite differences of the stages within the method:

$$f_i'(Y_s) \approx \frac{f_i\big(Y_s + \frac{3}{7}k_1^{(i)}\big) - f_i\big(Y_s\big)}{\frac{3}{7}k_1^{(i)}},$$

$$f_i'\Big(Y_s + \frac{3}{7}k_1^{(i)}\Big) \approx \frac{f_i\big(Y_s + \frac{3}{7}k_1^{(i)}\big) - f_i\big(Y_s - \frac{13}{15}k_1^{(i)} + \frac{28}{15}k_2^{(i)}\big)}{\frac{136}{105}k_1^{(i)} - \frac{28}{15}k_2^{(i)}},$$

$$\Delta f_i(Y_s) \approx \frac{f_i'\big(Y_s + \frac{3}{7}k_1^{(i)}\big) - f_i'(Y_s)}{\frac{3}{7}k_1^{(i)}},$$

where $f_i : \mathbb{R}^d \to \mathbb{R}$ is the i-th component of f and $k_j^{(i)} \in \mathbb{R}$ is the i-th coordinate of k_j. Therefore, the proposed three-stage Runge-Kutta method for SDE (4.4.1) is given by

Definition 4.4.3. (A stochastic Runge-Kutta method for additive noise). For a fixed number of steps N, we can construct a numerical solution $\{Y_i\}_{0\le i\le N}$ of (4.4.29) by setting $Y_0 := \xi$ and for each $i \in [0\,..\,N-1]$, defining Y_{i+1} with the formula

$$Y_{i+1} := Y_i + \frac{1}{9}k_1 + \frac{49}{72}k_2 + \frac{5}{24}k_3 + \frac{1}{30}h^2\sigma^2\Delta f(Y_i), \tag{4.4.31}$$

where the Laplacian $\Delta f(Y_i)$ can be discretized through finite differences of the stages

$$k_1 := hf(Y_i) + \sigma W_{t_i,t_{i+1}} + 3\sigma H_{t_i,t_{i+1}},$$

$$k_2 := hf\Big(Y_i + \frac{3}{7}k_1\Big) + \sigma W_{t_i,t_{i+1}} + \frac{12}{7}\sigma H_{t_i,t_{i+1}},$$

$$k_3 := hf\Big(Y_i - \frac{13}{15}k_1 + \frac{28}{15}k_2\Big) + \sigma W_{t_i,t_{i+1}} - \frac{36}{5}\sigma H_{t_i,t_{i+1}}.$$

Using the same error analysis as in the previous section, we see that the above method converges in a strong sense with order 1.5 and in a weak sense with order 2.0. It is also worth noting that in the zero noise setting (i.e. when the constant $\sigma = 0$), it is straightforward to verify that (4.4.31) gives a third order Runge-Kutta method. Therefore, we expect this method to perform especially well when the noise is small.

We have not yet conducted numerical experiments for our proposed method but intend to apply it to the overdamped Langevin diffusion. We would compare against the following SRK method given in [77] and studied as an MCMC algorithm in [69]:

$$Y_{i+1} := Y_i + \frac{1}{2}h\Big(f\big(Y_i^+\big) + f\big(Y_i^-\big)\Big) + \sigma W_{t_i,t_{i+1}},$$

$$Y_i^+ := Y_i + \sigma\left(\Big(\frac{1}{2} + \frac{1}{\sqrt{6}}\Big)W_{t_i,t_{i+1}} + H_{t_i,t_{i+1}}\right),$$

$$Y_i^- := Y_i + f(Y_i)h + \sigma\left(\Big(\frac{1}{2} - \frac{1}{\sqrt{6}}\Big)W_{t_i,t_{i+1}} + H_{t_i,t_{i+1}}\right).$$

4.5 Discretizations of Inhomogeneous Geometric Brownian Motion (IGBM)

We shall demonstrate some of the core ideas presented in this chapter through various discretizations of Inhomogeneous Geometric Brownian Motion (IGBM)

$$dy_t = a(b - y_t)\,dt + \sigma y_t\,dW_t, \tag{4.5.1}$$

where $a \geq 0$ and $b \in \mathbb{R}$ are the mean reversion parameters and $\sigma \geq 0$ is the volatility. As the vector fields are smooth, the SDE (4.5.1) is expressible in Stratonovich form:

$$dy_t = \tilde{a}(\tilde{b} - y_t)\,dt + \sigma y_t \circ dW_t, \tag{4.5.2}$$

where $\tilde{a} := a + \frac{1}{2}\sigma^2$ and $\tilde{b} := \frac{2ab}{2a+\sigma^2}$ denote the "adjusted" mean reversion parameters.

IGBM is an example of a one-factor short rate model and has seen recent attention in the mathematical finance literature as an alternative to popular models [17, 100]. Since the solution of the SDE (4.5.1) is both mean-reverting and non-negative, IGBM is suitable for modelling interest rates, stochastic volatilities and default intensities. In addition, it was used in [1] as a stochastic model for coal and natural gas prices. IGBM is also one of the simplest SDEs that has no known method of exact simulation.

By incorporating the ideas provided by the main result (Theorem 4.1.3) into the log-ODE method, we will design a state-of-the-art numerical method for IGBM. Whilst the vector fields are unbounded, they have linear growth and our numerical evidence indicates the method converges with a strong rate of $O(h^{\frac{3}{2}})$ and a weak rate of $O(h^2)$. It is known that (4.5.2) admits a unique strong solution (see [99]) satisfying

$$y_t = e^{-\tilde{a}h + \sigma W_{s,t}} \left(y_s + ab \int_s^t e^{\tilde{a}(u-s) - \sigma W_{s,u}}\,du \right). \tag{4.5.3}$$

This representation will be utilized for approximating the parabola-ODE accurately. Another appealing feature of (4.5.2) is that the following Lie brackets are constants:

$$\begin{aligned}
[f_1, f_0](y) &= \big(f_0' f_1 - f_1' f_0\big)(y) \\
&= -\tilde{a}\sigma y - \tilde{a}\sigma(\tilde{b} - y) \\
&= -ab\sigma\,,
\end{aligned}$$

$$\begin{aligned}
[f_1, [f_1, f_0]](y) &= \big(f_0' f_1' f_1 - 2 f_1' f_0' f_1 + f_1' f_1' f_0 \\
&\qquad + f_0''(f_1, f_1) - 2 f_1''(f_0, f_1) + f_1''(f_1, f_0)\big)(y) \\
&= -\tilde{a}\sigma^2 y + 2\tilde{a}\sigma^2 y + \tilde{a}\sigma^2(\tilde{b} - y) \\
&= ab\sigma^2.
\end{aligned}$$

Therefore the high order log-ODE (4.3.6) that was given by definition 4.3.4 becomes

$$\frac{dz}{du} = \tilde{a}(\tilde{b} - z)h + \sigma z\, W_{s,t} - ab\sigma h H_{s,t} + ab\sigma^2 \left(\frac{3}{5} h H_{s,t}^2 + \frac{1}{30} h^2\right). \tag{4.5.4}$$

From the above, we will see that the log-ODE method for IGBM can be written in terms of a closed-form expression that guarantees positivity for the numerical solution. We shall investigate the strong and weak convergence rates of the following methods:

1. **Log-ODE method** (see definition 4.3.4)
   Since the vector fields of (4.5.2) give constant Lie brackets, this method becomes

$$\begin{aligned} Y_{k+1}^{\log} &:= Y_k^{\log} e^{-\tilde{a}h + \sigma W_{t_k,t_{k+1}}} \\ &\quad + abh\left(1 - \sigma H_{t_k,t_{k+1}} + \sigma^2\left(\frac{3}{5} h H_{t_k,t_{k+1}}^2 + \frac{1}{30} h\right)\right) \frac{e^{-\tilde{a}h + \sigma W_{t_k,t_{k+1}}} - 1}{-\tilde{a}h + \sigma W_{t_k,t_{k+1}}}, \\ Y_0^{\log} &:= y_0. \end{aligned}$$

2. **Parabola-ODE method** (see definition 4.3.5)
   By emulating the proof of the formula (4.5.3), this method can be expressed as

$$\begin{aligned} Y_{k+1}^{\text{para}} &:= e^{-\tilde{a}h + \sigma W_{t_k,t_{k+1}}} \left(Y_k^{\text{para}} + ab \int_{t_k}^{t_{k+1}} e^{\tilde{a}(s-t_k) - \sigma \widehat{W}_{t_k,s}}\, ds\right), \\ Y_0^{\text{para}} &:= y_0, \end{aligned}$$

   where $\widehat{W}$ denotes the Brownian parabola constructed using $(W_{t_k,t_{k+1}}, H_{t_k,t_{k+1}})$. The integral above will be approximated by 3-point Gauss-Legendre quadrature.

3. **Piecewise linear method** (see [97] for a definition with proof of convergence)
   Just as above, this method can be simplified to give a straightforward formula.

$$\begin{aligned} Y_{k+1}^{\text{lin}} &:= Y_k^{\text{lin}} e^{-\tilde{a}h + \sigma W_{t_k,t_{k+1}}} + abh\, \frac{e^{-\tilde{a}h + \sigma W_{t_k,t_{k+1}}} - 1}{-\tilde{a}h + \sigma W_{t_k,t_{k+1}}}, \\ Y_0^{\text{lin}} &:= y_0. \end{aligned}$$

4. **Milstein method** (see section 6 of [54] and section 10.3 of [62] for overviews)
   For this method, we shall be taking the positive part to guarantee non-negativity.

$$\begin{aligned} Y_{k+1}^{\text{mil}} &:= \left(Y_k^{\text{mil}} + \tilde{a}(\tilde{b} - Y_k^{\text{mil}})h + \sigma\, Y_k^{\text{mil}}\, W_{t_k,t_{k+1}} + \frac{1}{2}\sigma^2\, Y_k^{\text{mil}}\, W_{t_k,t_{k+1}}^2\right)^+, \\ Y_0^{\text{mil}} &:= y_0. \end{aligned}$$

5. **Euler-Maruyama method** (see sections 4, 5 of [54] and section 10.2 of [62])
   Just as before, we take the positive part of each step and ensure non-negativity.

$$\begin{aligned} Y_{k+1}^{\text{eul}} &:= \left(Y_k^{\text{eul}} + a(b - Y_k^{\text{eul}})h + \sigma\, Y_k^{\text{eul}}\, W_{t_k,t_{k+1}}\right)^+, \\ Y_0^{\text{eul}} &:= y_0. \end{aligned}$$

The Euler-Maruyama and Milstein methods are included in the numerical experiment as benchmarks for testing how the proposed methods compare to well-known methods.

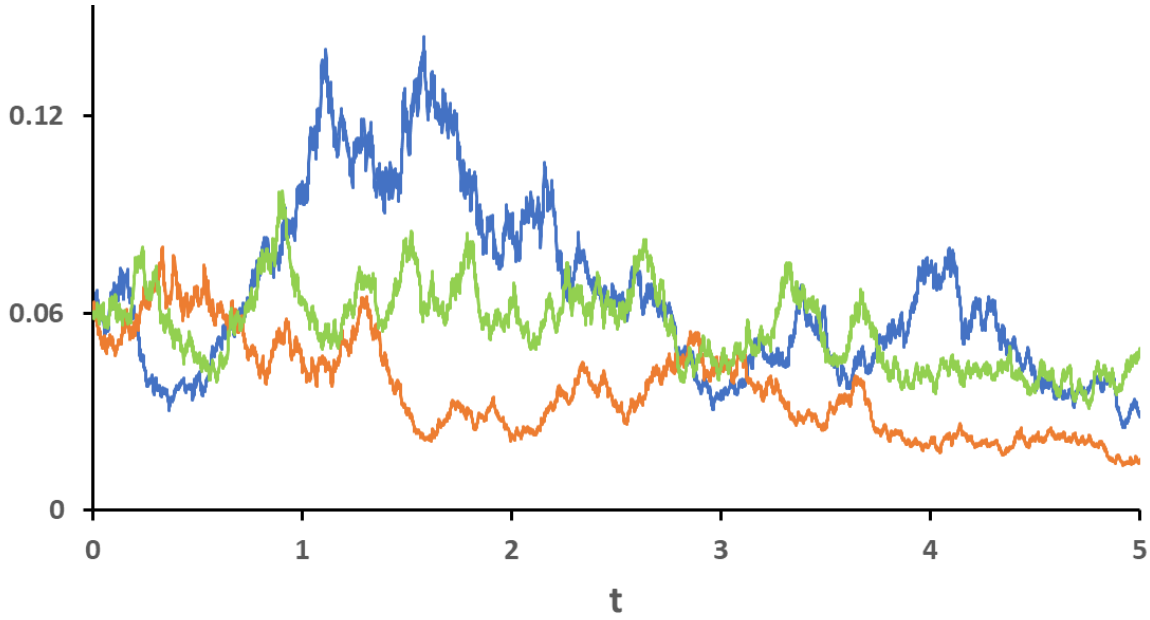

Figure 4.12: Log-ODE sample paths of IGBM where $a = 0.1$, $b = 0.04$ and $\sigma = 0.6$. The sample paths experience larger fluctuations the further away from zero they are.

We shall now present the estimators used for quantifying strong and weak errors. As before, we shall be discretizing the SDE on a uniform partition with mesh size h. The functional chosen for testing weak convergence is $p(y) = (y - b)^+$ which is the payoff of a European call option with strike b and well approximated by polynomials.

Definition 4.5.1 (Strong and weak error estimators)**.** For each $N \geq 1$, let Y_N denote a numerical solution of (4.5.1) computed at time T using the fixed step size $h = \frac{T}{N}$. We can define the following estimators for quantifying strong and weak convergence:

$$S_N := \sqrt{\mathbb{E}\left[\left(Y_N - Y_T^{\text{fine}}\right)^2\right]}, \tag{4.5.5}$$

$$E_N := \left|\mathbb{E}\left[\left(Y_N - b\right)^+\right] - \mathbb{E}\left[\left(Y_T^{\text{fine}} - b\right)^+\right]\right|, \tag{4.5.6}$$

where the above expectations are approximated by standard Monte-Carlo simulation and Y_T^{fine} is the numerical solution of (4.5.1) obtained at time T using the log-ODE method with a "fine" step size of $\min\left(\frac{h}{10}, \frac{T}{1000}\right)$. The fine step size is chosen so that the $L^2(\mathbb{P})$ error between Y_T^{fine} and the true solution y is negligible compared to S_N. Note that Y_N and Y_T^{fine} are both computed with respect to the same Brownian paths.

In this numerical example, we shall use the same parameter values as in [17], namely $a = 0.1$, $b = 0.04$, $\sigma = 0.6$ and $y_0 = 0.06$. The time horizon is fixed at $T = 5$.

We will now present our results for the numerical experiment described above.

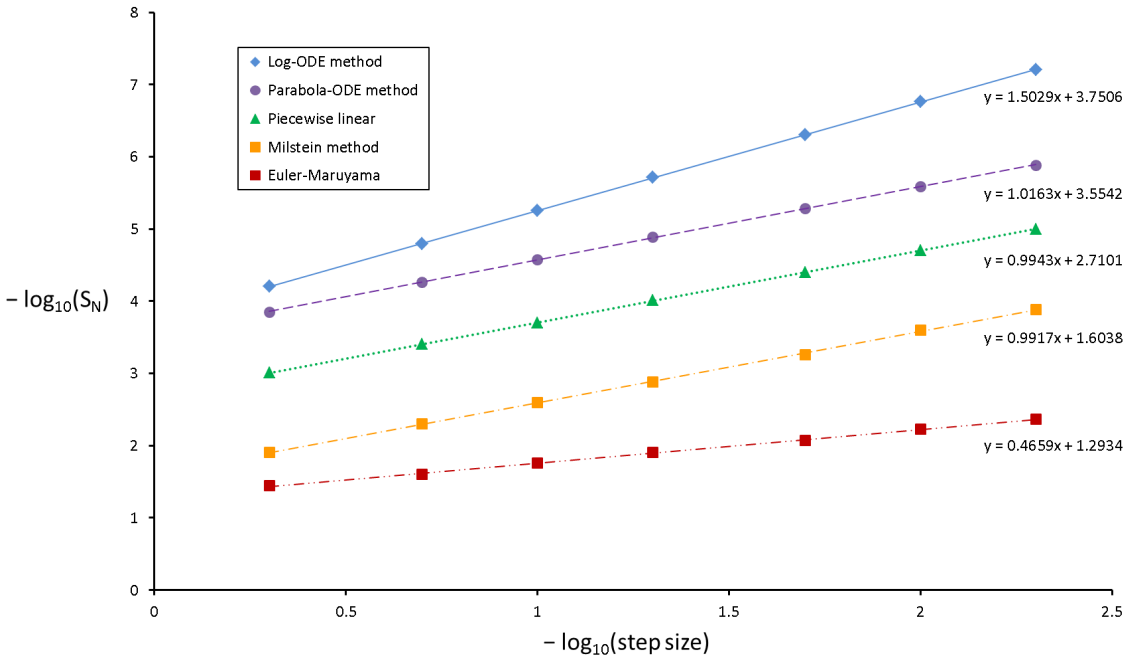

Figure 4.13: S_N computed with 100,000 sample paths as a function of step size $h = \frac{T}{N}$.

From the above graph we see that the log-ODE method is by far the most accurate. This is epitomized by the fact that the numerical error produced by 100 steps of the log-ODE method is comparable to the error of the parabola method with 1000 steps. In addition, whilst there are three methods that share the same order of convergence it is evident there are magnitudes of difference between their respective accuracies. For example, the parabola method is seven times more accurate than piecewise linear. As one might expect, the Euler-Maruyama and Milstein schemes both perform poorly. Nevertheless, in order to truly measure the performance of these numerical methods, we should consider the computational costs required for achieving a specified accuracy.

Table 4.1: Estimated simulation times for computing 100,000 sample paths that achieve a given accuracy by a single-threaded C++ program on a desktop computer.

	Log-ODE	Parabola	Linear	Milstein	Euler
Estimated time to achieve an accuracy of $S_N = 10^{-4}$	0.179 (s)	0.405 (s)	1.47 (s)	15.4 (s)	0.437 (days)
Estimated time to achieve an accuracy of $S_N = 10^{-5}$	0.827 (s)	3.90 (s)	14.9 (s)	157 (s)	61.2 (days)

The above times are estimated using the graph in figure 4.13 with the following table:

Table 4.2: Simulation times for computing 100,000 sample paths with 100 steps per path using a single-threaded C++ program on the same desktop computer.

	Log-ODE	Parabola	Linear	Milstein	Euler
Computation time (s)	2.44	2.95	1.48	1.18	1.17

Finally, we will investigate the rates of weak convergence for these numerical methods.

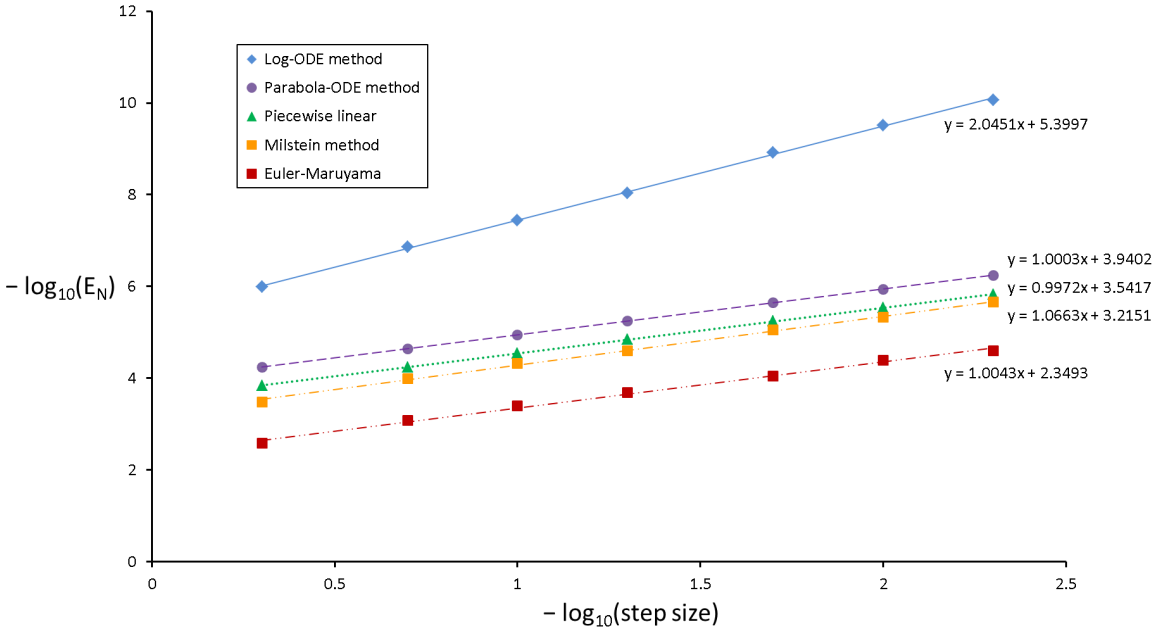

Figure 4.14: E_N computed with 500,000 sample paths as a function of step size $h = \frac{T}{N}$.

The above graph demonstrates that the log-ODE method is especially well-suited for weak approximation as it achieves a second order convergence rate in this example. As well as achieving this high order of weak convergence, the method is accurate even when using larger step sizes. For example, the log-ODE method produces less weak error with a step size of 0.5 than the other methods produce using a step size of 0.005.

Surprisingly, the middle three methods all perform with a similar level of accuracy. In particular, this gives numerical evidence that the parabola-ODE method has low convergence rates (in both senses) but may perform better as a strong approximation.

We expect the log-ODE and parabola-ODE methods to have about twice the computational cost as other methods since each step requires two random variables. Table 4.2 confirms this and thus sampling may be a bottleneck for these methods. So overall, the numerical evidence supports our claim that the high order log-ODE method is currently a state-of-the-art method for the strong approximation of IGBM (though we have yet to compare with the splitting schemes recently proposed in [92]).

In conclusion, there are primarily four results that are established in this chapter:

- **An efficient strong polynomial approximation of Brownian motion** The main result allows one to construct a "smoother" Brownian motion as a finite sum of (-1, -1)-Jacobi polynomials with independent Gaussian weights. Moreover, it was shown that the approximation is optimal in a weighted $L^2(\mathbb{P})$ sense and the surrounding noise is an independent centered Gaussian process.
- **Unbiased approximation of third order Brownian iterated integrals** Iterated integrals of Brownian motion and time are important objects in the study of SDEs as they appear naturally within stochastic Taylor expansions. We have derived the $L^2(\mathbb{P})$-optimal estimator for a class of such integrals that is measurable with respect to the path's increment and space-time Lévy area.
- **A parabola-based Runge-Kutta method for SDEs with additive noise** Runge-Kutta methods are popular for numerically solving differential equations since they only require evaluations of the vector field (and not its derivatives). We have developed a three-stage Runge-Kutta for SDEs with additive noise that gives a Taylor expansion which correctly uses the new integral estimators.
- **Simulation of Inhomogeneous Geometric Brownian Motion (IGBM)** IGBM is a mean-reverting short rate model used in mathematical finance and also one of the simplest SDEs that has no known method of exact simulation. By incorporating a new iterated integral estimator into the log-ODE method we have developed a high order and state-of-the-art numerical method for IGBM.

Moreover, the results of this chapter immediately lead to the following open problems:

- Is it possible to generalize the main theorem for a fractional Brownian motion?
- What are the most efficient Runge-Kutta methods for general one-dimensional SDEs that correctly utilize the new estimators for third order iterated integrals?
- Does this polynomial expansion give an optimal approximation of Lévy area? (This problem was studied for the standard Karhunen-Loève expansion in [30])
- Is there a connection between the stochastic integrals $\{I_k\}$ appearing in the polynomial expansion and the "space-time" log-signature of Brownian motion? (We have already shown that I_1 and I_2 are (rescaled) log-signature coefficients)
- Does the polynomial approximation theorem extend to the Brownian sheet? (This would correspond to the space-time white noise commonly used in SPDEs)

Chapter 5

High order piecewise linear discretizations of Brownian motion

In the previous chapter, we presented a high order version of the log-ODE method based on new unbiased estimators for certain iterated integrals of Brownian motion. In particular, this numerical method was designed for Stratonovich SDEs of the form:

$$\begin{aligned} dy_t &= f_0\left(y_t\right) dt + f_1\left(y_t\right) \circ dW_t, \\ y_0 &= \xi, \end{aligned} \tag{5.0.1}$$

where $\xi \in \mathbb{R}^e$ and f_i denote bounded C^∞ vector fields on $\mathbb{R}^e$ with bounded derivatives. Although it is accurate, there are primarily two disadvantages of the log-ODE method.

1. The method requires one to explicitly evaluate or approximate the Lie brackets:

$$\begin{aligned} [f_1, f_0] &= f_0' \, f_1 - f_1' \, f_0, \\ [f_1, [f_1, f_0]] &= f_0' \, f_1' \, f_1 - 2 f_1' \, f_0' \, f_1 + f_1' \, f_1' \, f_0 \\ &\qquad + f_0''(f_1, f_1) - 2 f_1''(f_0, f_1) + f_1''(f_1, f_0). \end{aligned}$$

2. It is possible for the "log-ODE" itself to be stiff and difficult to solve numerically.

This motivates us to consider alternative numerical methods that can discretize SDEs using ODE solvers, but do not necessarily require knowledge of vector field derivatives. The methods that are presented in this chapter are all examples of piecewise linear or Wong-Zakai approximations ([97] has the first account of this standard approach). The idea is to discretize the Brownian motion as a piecewise linear path that agrees on discretization points. Along each such piece, the SDE can then be approximated by an associated ODE that has the same vector fields but is driven by a linear path. In practice, each step of this numerical method requires one to discretize an ODE with a sufficiently accurate solver (such as Runge-Kutta or linear multistep methods).

Before we start designing piecewise linear discretizations in sections 5.1 and 5.2, we will give an overview of our motivation, general strategy and numerical examples. To begin, we recall the piecewise linear paths used in the Wong-Zakai approximation.

Definition 5.0.1. (Standard piecewise linear method). Let $\widehat{W}$ be the piecewise linear approximant for a Brownian motion W that agrees on N equally spaced points,

$$\widehat{W}_u := W_{t_k} + \frac{u - t_k}{h} W_{t_k, t_{k+1}}, \quad \text{for } u \in [t_k, t_{k+1}] \text{ and } k = 0, 1, \cdots, N-1,$$

where $h := \frac{T}{N}$ and $t_k := kh$. We construct a numerical solution $\{Y_k\}_{0 \le k \le N}$ for (5.0.1) by setting $Y_0 := \xi$ and defining Y_{k+1} to be the solution at $u = 1$ of the following ODE:

$$\begin{aligned} \frac{dz}{du} &= f_0(z)h + f_1(z)\widehat{W}_{t_k, t_{k+1}}, \\ z_0 &= Y_k, \end{aligned} \tag{5.0.2}$$

for $k \in [0 \,..\, N-1]$ where $\widehat{W}_{t_k, t_{k+1}} := \widehat{W}_{t_{k+1}} - \widehat{W}_{t_k}$ is an increment of the path $\widehat{W}$. Alternatively, we can view the above as simply the "derivative-free" log-ODE method.

Remark 5.0.2. The path $\widehat{W}$ is the best $L^2(\mathbb{P})$ approximation of the Brownian motion that is measurable with respect to its increments on uniformly spaced grid points [21],

$$\widehat{W}_u = \mathbb{E}\big[W_u \,\big|\, W_{t_0}, W_{t_1}, \cdots W_{t_N}\big] \quad \text{for } u \in [0, T].$$

Remark 5.0.3. For a general SDE (5.0.1), this method exhibits strong and weak convergence rates of $O(h)$. Thus the standard piecewise linear approach is low order.

In this chapter, we shall be constructing piecewise linear paths so that solving the associate sequence of ODEs (5.0.2) gives a high order numerical method for (5.0.1). In order to achieve this, we design these piecewise linear discretizations of Brownian motion to capture the "local shape" of the path (i.e. match certain iterated integrals).

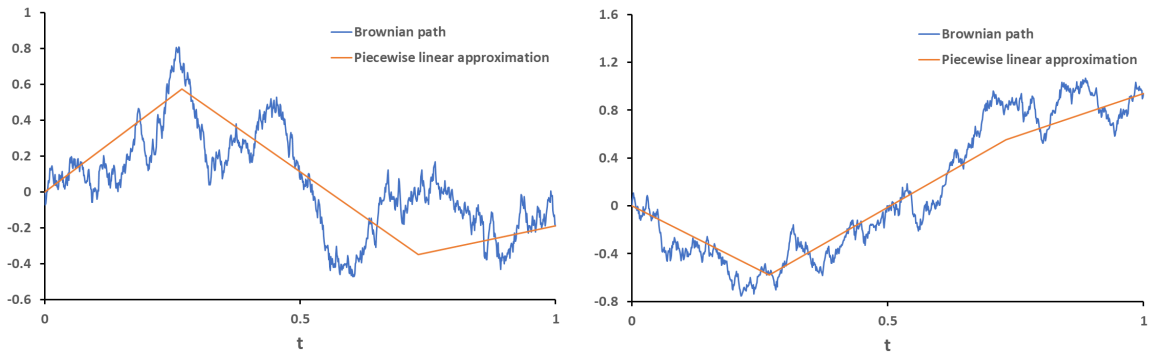

Figure 5.1: Brownian motion with piecewise linear paths that match certain integrals. In practice, the above piecewise linear paths can be generated using two independent Gaussian random variables and a further independent Rademacher random variable.

In addition to the above geometrical interpretation, we can view these piecewise linear approximations as specific compositions of derivative-free log-ODE methods. It is still ongoing research to develop and analyse new methodologies for composing log-ODE methods containing vector field Lie brackets or multiple Brownian motions.

As one would expect, over each interval $[s,t]$ we will construct a piecewise linear path γ using the increment $W_{s,t}$ and space-time Lévy area $H_{s,t}$ of the Brownian path. After generating γ, we can then consider an ODE defined with the same vector fields:

$$dz_u = f_0\left(z_u\right) du + f_1\left(z_u\right) d\gamma_u. \tag{5.0.3}$$

Recall that in the proof of Theorem 4.3.11, we analysed the above differential equation but with γ equal to the Brownian parabola corresponding to W over the interval $[s,t]$. In a similar fashion, the ODE (5.0.3) yields the (rough) Taylor expansion for small h,

$$\begin{aligned} z_t = z_s &+ f_0(z_s)\, h + f_1(z_s)\, \gamma_{s,t} + \frac{1}{2}\, f_1' \, f_1(z_s)\, \gamma_{s,t}^2 + \frac{1}{2}\, f_0' \, f_0(z_s)\, h^2 \\ &+ f_0' \, f_1(z_s) \int_s^t \int_s^u d\gamma_v \, du + f_1' \, f_0(z_s) \int_s^t \int_s^u dv \, d\gamma_u \\ &+ \frac{1}{6}(f_1' \, f_1' \, f_1 + f_1''(f_1, f_1))(z_s)\, \gamma_{s,t}^3 \\ &+ \left(f_0' \, f_1' \, f_1 + f_0''(f_1, f_1)\right)(z_s) \int_s^t \int_s^u \int_s^v d\gamma_r \, d\gamma_v \, du \\ &+ \left(f_1' \, f_0' \, f_1 + f_1''(f_0, f_1)\right)(z_s) \int_s^t \int_s^u \int_s^v d\gamma_r \, dv \, d\gamma_u \\ &+ \left(f_1' \, f_1' \, f_0 + f_1''(f_1, f_0)\right)(z_s) \int_s^t \int_s^u \int_s^v dr \, d\gamma_v \, d\gamma_u \\ &+ \frac{1}{24}\big(f_1' \, f_1' \, f_1' \, f_1 + f_1' \, f_1''(f_1, f_1) + 3 f_1''(f_1' \, f_1, f_1) + f_1'''(f_1, f_1, f_1)\big)(z_s)\, \gamma_{s,t}^4 \\ &+ \textit{higher order terms}. \end{aligned} \tag{5.0.4}$$

Thus for the numerical method defined by γ to perform comparably to the high order log-ODE method, we want (5.0.4) to be close to the Taylor expansion (4.3.9). Therefore as a minimum requirement, the path γ should match both $W_{s,t}$ and $H_{s,t}$. So if γ has two pieces, it must belong to the following one-parameter family of paths:

$$\gamma^\theta := \begin{cases} W_s + \frac{u-s}{h}\, W_{s,t} + \frac{2(u-s)}{\theta h}\, H_{s,t} & \text{for } \ u \in [s, s+\theta h] \\ W_s + \frac{u-s}{h}\, W_{s,t} + \frac{2(t-u)}{(1-\theta) h}\, H_{s,t} & \text{for } \ u \in [s+\theta h, t] \end{cases}, \ \text{ where } \ \theta \in (0,1).$$

In this case, it is possible to analytically compute the following integrals of the path, (The calculations required to derive these expressions are presented in Appendix A)

$$\begin{aligned}\int_s^t \gamma^\theta_{s,u}\,du &= \int_s^t W_{s,u}\,du,\\ \int_s^t \big(\gamma^\theta_{s,u}\big)^2\,du &= \frac{1}{3}hW^2_{s,t} + \frac{2}{3}(1+\theta)hW_{s,t}H_{s,t} + \frac{4}{3}hH^2_{s,t},\\ \int_s^t (u-s)\gamma^\theta_{s,u}\,du &= \frac{1}{3}h^2W_{s,t} + \frac{1}{3}(1+\theta)h^2H_{s,t}.\end{aligned}$$

On the other hand, the integral moments below were derived in the previous chapter.

$$\begin{aligned}\mathbb{E}\bigg[\int_s^t W^2_{s,u}\,du\,\Big|\,W_{s,t},H_{s,t}\bigg] &= \frac{1}{3}\,hW^2_{s,t} + hW_{s,t}H_{s,t} + \frac{6}{5}hH^2_{s,t} + \frac{1}{15}h^2,\\ \mathbb{E}\bigg[\int_s^t (u-s)W_{s,u}\,du\,\Big|\,W_{s,t},H_{s,t}\bigg] &= \frac{1}{3}h^2W_{s,t} + \frac{1}{2}h^2H_{s,t}.\end{aligned}$$

It is clear that the paths in the one-parameter family $\{\gamma^\theta\}_{\theta\in(0,1)}$ are not sufficient to capture $(W_{s,t},H_{s,t})$ along with the above estimators for third order iterated integrals. Thus, the piecewise linear paths developed in this chapter have at least three pieces. To begin with, we shall choose the paths from the following three-parameter family:

$$\gamma^{a,b,c} := \begin{cases} W_s + \frac{b}{ah}\,(u-s) & \text{for } u\in[s,s+ah]\\ W_s + b + \frac{c-b}{(1-2a)h}\,(u-(s+ah)) & \text{for } u\in[s+ah,t-ah]\\ W_s + c + \frac{W_{s,t}-c}{ah}\,(u-(t-ah)) & \text{for } u\in[t-ah,t]\end{cases}$$

where $a\in\big(0,\frac{1}{2}\big)$ and $b,c\in\mathbb{R}$. As before, certain integrals can be computed explicitly. (The calculations required to derive these expressions are presented in Appendix A)

$$\int_s^t \gamma^{a,b,c}_{s,u}\,du = \frac{1}{2}\,abh + \frac{1}{2}(1-2a)(b+c)h + \frac{1}{2}a\big(c+W_{s,t}\big)h, \tag{5.0.5}$$

$$\int_s^t \big(\gamma^{a,b,c}_{s,u}\big)^2\,du = \frac{1}{3}(1-2a)\big(b^2+bc+c^2\big)h + \frac{1}{3}a\big(b^2+c^2+cW_{s,t}+W^2_{s,t}\big)h, \tag{5.0.6}$$

$$\begin{aligned}\int_s^t (u-s)\gamma^{a,b,c}_{s,u}\,du = \frac{1}{3}a^2bh^2 &+ \frac{1}{6}(1-2a)((1+a)b+(2-a)c)h^2\\ &+ \frac{1}{6}a\big((3-2a)c+(3-a)W_{s,t}\big)h^2.\end{aligned} \tag{5.0.7}$$

For simplicity, we will first consider the choice of $\gamma^{a,b,c}$ when $W_{s,t}=0$ and $H_{s,t}=0$. In this scenario, the parameters (a,b,c) are chosen so that the Taylor expansion (5.0.4)

matches the log-ODE expansion up to $O(h^2)$ terms and hence satisfy the equations,

$$\frac{1}{2}abh + \frac{1}{2}(1-2a)(b+c)h + \frac{1}{2}ach = 0,$$
$$\frac{1}{3}ab^2h + \frac{1}{3}(1-2a)\big(b^2+bc+c^2\big)h + \frac{1}{3}ac^2h = \frac{1}{15}h^2.$$

For any fixed choice of $a \in \big(0, \frac{1}{2}\big)$, there are two distinct solutions to the above system:

$$(b,c) = \pm\left(\frac{1}{\sqrt{5}}\sqrt{h}\,, -\frac{1}{\sqrt{5}}\sqrt{h}\right).$$

Although our aim is to design the path $\gamma^{a,b,c}$ solely for the purpose of SDE simulation, we would ideally want this approximation to resemble the driving Brownian motion. In the above case, both choices for $\gamma^{a,\cdot}$ are equally poor approximants (by symmetry).

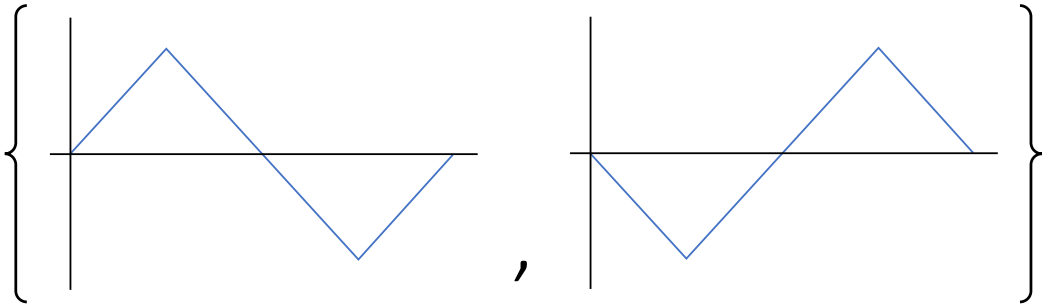

Figure 5.2: Possible piecewise linear paths that can approximate the Brownian arch.

This motivates one to generate addition information about the Brownian path. A natural candidate for this could be the space-time-time Lévy area $K_{s,t}$ as it is independent of $(W_{s,t}, H_{s,t})$ and estimates the skewness of the associated arch process. In this setting, we can choose the path $\gamma^{a,b,c}$ that additionally minimizes the quantity

$$\left| \int_s^t (u-s)\gamma^{a,b,c}_{s,u}\,du - \int_s^t (u-s)W_{s,u}\,du \right|$$

The construction of piecewise linear approximations that are measurable with respect to $(W_{s,t}, H_{s,t}, K_{s,t})$ shall be discussed in the first two sections of this chapter. On the other hand, for SDEs where the two vector fields are cheap to evaluate, generating an additional Gaussian random variable during each step may be excessive. As a result, we will define the following quantities to be cheaper alternatives to $K_{s,t}$.

Definition 5.0.4. The **space-time orientation** $n_{s,t}$ of Brownian motion over an interval $[s,t]$ represents the relative size of two space-time subareas and is defined as

$$n_{s,t} := \operatorname{sgn}\left(H_{s,u} - H_{u,t}\right),$$

where $u := \frac{1}{2}(s+t)$.

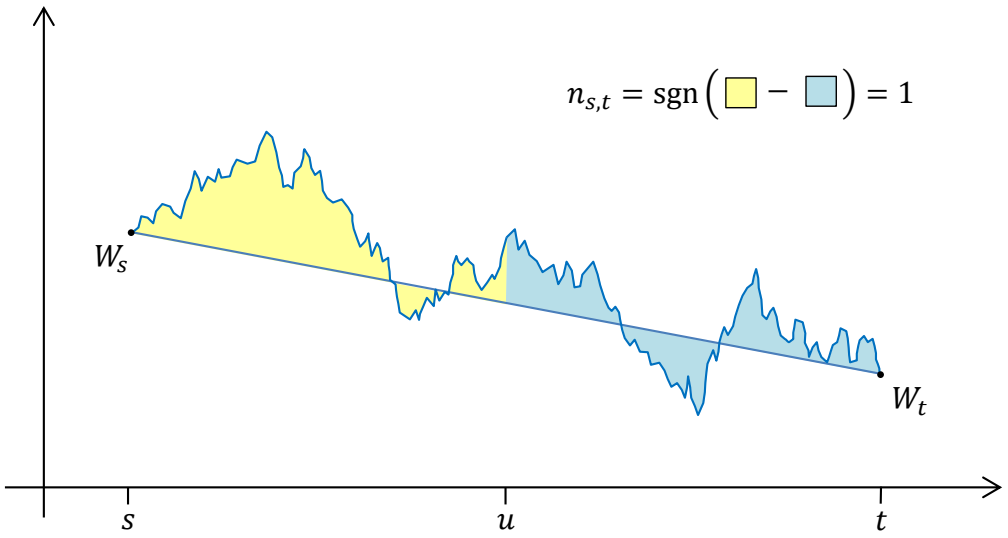

Figure 5.3: A sample path of Brownian motion with positive space-time orientation.

Definition 5.0.5. The **sign of the skew** of Brownian motion over $[s,t]$ is given by

$$k_{s,t} := \mathrm{sgn}\left(K_{s,t}\right).$$

It is clear that $n_{s,t}$ and $k_{s,t}$ have Rademacher distributions and therefore require less computational effort to generate than $K_{s,t}$. Whilst it appears that these random variables can be used interchangeably with one another, we shall see that it is more straightforward to incorporate $n_{s,t}$ into variable step size and multilevel Monte Carlo methodologies that use both increments and areas. That said, in order for $n_{s,t}$ or $k_{s,t}$ to be used in a numerical scheme, they must be generatable along with $W_{s,t}$ and $H_{s,t}$.

Theorem 5.0.6. *The space-time orientation $n_{s,t}$ is independent of $(W_{s,u}, W_{u,t}, H_{s,t})$.*

Proof. It was established in the previous chapter that $H_{s,t}$ is independent of $W_{s,t}$. Since (s,u) and (u,t) are disjoint intervals and Brownian motion has independent increments, it follows that $(W_{s,u}, W_{u,t})$ and $(H_{s,u}, H_{u,t})$ are independent of each other. As $\{W_{s,r}\}_{r\in[s,t]}$ and $\{W_{s,t} - W_{r,t}\}_{r\in[s,t]}$ have the same law, we have by symmetry that

$$\mathrm{Cov}(H_{s,u}, H_{s,t}) = \mathrm{Cov}(H_{u,t}, H_{s,t}). \tag{5.0.8}$$

Note that $H_{s,u} - H_{u,t}$ and $H_{s,t}$ can be expressed as linear functionals on the same Brownian motion and are therefore jointly normal random variables. By the above covariance, we see that $H_{s,u} - H_{u,t}$ and $H_{s,t}$ are uncorrelated (and hence independent). The result immediately follows since $n_{s,t}$ is a deterministic function of $H_{s,u} - H_{u,t}$. □

Theorem 5.0.7. *The sign of the skew* $\operatorname{sgn}(K_{s,t})$ *is independent of* $(W_{s,u}, W_{u,t}, H_{s,t})$.

Proof. Without loss of generality, we can consider the Brownian motion over $[0,1]$. It has already been established in Chapter 4 that $K_{s,t}$ corresponds to the integral I_2, as defined in Theorem 4.1.3, and is therefore independent of both $W_{s,t}$ and $H_{s,t}$. Corollary 4.1.4 expresses W using the independent Gaussian random variables $\{I_k\}$,

$$W_t = W_1 e_0(t) + \sum_{k=1}^{\infty} I_k \, e_k(t).$$

We can see from the above that $W_{\frac{1}{2}}$ and I_2 are independent as $e_2(t) = 0$ when $t = \frac{1}{2}$. The result now follows since $k_{s,t} := \operatorname{sgn}(K_{s,t})$ is a deterministic function of $K_{s,t}$. □

In the first section of this chapter, we will construct piecewise linear paths for discretizing the Stratonovich SDE (5.0.1). Our general methodology is outlined below:

1. Identify the expectation of $L_{s,t}$ and $K_{s,t}$ conditional on certain random variables. In our case, we shall be generating one of the following triples over each interval.

   • $(W_{s,t}, H_{s,t}, K_{s,t})$ • $(W_{s,t}, H_{s,t}, n_{s,t})$ • $(W_{s,t}, H_{s,t}, k_{s,t})$

2. Using (5.0.5) and (5.0.6), we can derive a system of equations for (a, b, c) from

$$\int_s^t \gamma_{s,u}^{a,b,c} \, du = \int_s^t W_{s,u} \, du,$$
$$\int_s^t \left(\gamma_{s,u}^{a,b,c}\right)^2 du = \mathbb{E}\left[\int_s^t W_{s,u}^2 \, du \,\middle|\, (W, H, \cdot)_{s,t} \right].$$

   As already seen, for any fixed a, there may be two solutions to the above system.

3. So once a is fixed, the solution for (b, c) is then chosen to additionally minimize

$$\left| \int_s^t (u-s)\gamma_{s,u}^{a,b,c} \, du - \mathbb{E}\left[\int_s^t (u-s) W_{s,u} \, du \,\middle|\, (W, H, \cdot)_{s,t} \right] \right|.$$

   Currently the choice of a is somewhat arbitrary and may require further work.

4. Propagate a numerical solution to (5.0.1) over the time interval $[s, t]$ using the ODE that is governed by the vector fields f_0, f_1 but is driven by the path $\gamma^{a,b,c}$. This method then converges strongly with order 1.5 and weakly with order 2.0.

The numerical examples that we will use to demonstrate this approach are the Cox-Ingersoll-Ross and squared Bessel processes (which are presented in Chapter 6).

- **Cox-Ingersoll-Ross (CIR) process** (see [23] for the original paper on this)

$$dy_t = a(b - y_t)\,dt + \sigma\sqrt{y_t}\,dW_t.$$

- **Squared δ-dimensional Bessel process** ([47] gives a survey of both processes)

$$dz_t = \delta\,dt + 2\sqrt{z_t} \circ dW_t.$$

These SDEs are not further discussed in this chapter as they are discretized using a piecewise linear approach in a variable step size framework (see section 6.3 for details). The motivation for varying the step sizes used within the numerical approximations of these processes is that one of the governing vector fields is not globally Lipschitz. More specifically, the derivatives of $\sqrt{\cdot}$ "blow-up" whenever the process is near zero. So we generally would like to take smaller step sizes if the process gets closer to zero.

In the second section of this chapter, we shall construct piecewise linear paths for discretizing the SDE (5.0.1) when the vector fields satisfy the commutativity condition

$$[f_1, [f_1, f_0]] = 0. \tag{5.0.9}$$

The above condition implies that the $O(h^2)$ component of the SDE's stochastic Taylor expansion does not depend on the Brownian space-space-time Lévy area $L_{s,t}$. As a result, we do not need to approximate this quantity when constructing $\gamma^{a,b,c}$. Instead, we can generate the triple $(W_{s,t}, H_{s,t}, K_{s,t})$ and consider the three-parameter piecewise linear path $\gamma^{a,b,c}$ on $[s,t]$ which is defined according to the following recipe:

- Once a is fixed, the pair (b,c) can be uniquely determined so that $\gamma^{a,b,c}$ satisfies

$$\int_s^t \gamma_{s,u}^{a,b,c}\,du = \int_s^t W_{s,u}^{a,b,c}\,du,$$
$$\int_s^t (u-s)\gamma_{s,u}^{a,b,c}\,du = \int_s^t (u-s)W_{s,u}^{a,b,c}\,du.$$

- We will choose this constant to be $a = \frac{1}{2} - \frac{1}{2\sqrt{5}}$ as it produces the equality below

$$\int_s^t (u-s)^2\gamma_{s,u}^{a,b,c}\,du = \mathbb{E}\left[\int_s^t (u-s)^2 W_{s,u}\,du \,\middle|\, W_{s,t}, H_{s,t}, K_{s,t}\right].$$

For Stratonovich SDEs that satisfy the condition (5.0.9), we expect the numerical method associated with this path to converge strongly and weakly with order 2.0. Whilst the above methodology is fairly general, we have a specific example in mind:

The Langevin equation (Chapter 6 of [82] gives a detailed analysis of this SDE)

$$dQ_t = P_t\,dt,$$
$$dP_t = -\nabla U(Q_t)\,dt - \nu P_t\,dt + \sqrt{\frac{2\nu}{\beta}}\,dW_t,$$

where $Q, P \in \mathbb{R}^d$ represent the position and momentum of a point particle that moves in a scalar potential U according to Newton's second law but is subject to additional frictional and stochastic forces. In the above SDE, the linear dissipation term $-\nu P$ is governed by the friction coefficient ν and the noise term is determined by a d-dimensional Brownian motion W and inverse temperature $\beta = \big(k_B T\big)^{-1}$ where k_B is the Boltzmann constant and T denotes the absolute temperature of the system.

The Langevin equation appears in statistical mechanics as the fundamental model of a Brownian particle that moves under the influence of an external force along with frictional and stochastic forces ([22] is a comprehensive textbook on this equation). In the third section of this chapter, we shall discretize the Langevin equation using the above methodology and demonstrate that it achieves a convergence rate of $O(h^3)$.

5.1 Piecewise linear methods for SDEs with non-commuting vector fields

In this section, we will define piecewise linear paths for numerically solving SDE (5.0.1) where the vector fields f_0, f_1 are assumed to not satisfy the condition (5.0.9). As stated previously, these paths shall be measurable with respect to $(W_{s,t}, H_{s,t})$ along with an additional random variable that encodes the "skew" of the Brownian motion. To follow the methodology outlined earlier in this chapter, we use the lemma below.

Lemma 5.1.1. *Consider a piecewise linear path $\gamma^{a,b,c}$ in the three-parameter family. The rescaled Lévy areas of the space-time process $\{(u, \gamma_u^{a,b,c})\}_{u\in[s,t]}$ can be defined as*

$$\widehat{H}_{s,t} := \frac{1}{h}\int_s^t \gamma_{s,u}^{a,b,c} - \frac{u-s}{h}\gamma_{s,t}^{a,b,c}\,du, \tag{5.1.1}$$

$$\widehat{K}_{s,t} := \frac{1}{h^2}\int_s^t \Big(\gamma_{s,u}^{a,b,c} - \frac{u-s}{h}\gamma_{s,t}^{a,b,c}\Big)\Big(\frac{1}{2}h - (u-s)\Big)\,du. \tag{5.1.2}$$

Then for any choice of $(a,b,c) \in \big(0,\frac{1}{2}\big)\times\mathbb{R}^2$ and any Brownian increment $W_{s,t}\in\mathbb{R}$,

$$\widehat{H}_{s,t} = \frac{1}{2}(1-a)\big(b+c-W_{s,t}\big), \tag{5.1.3}$$

$$\widehat{K}_{s,t} = \frac{1}{12}(1-a)\big(b-c+(1-2a)W_{s,t}\big), \tag{5.1.4}$$

and the following triple iterated integral of $\gamma^{a,b,c}$ and time can be computed explicitly:

$$\int_s^t \left(\gamma_{s,u}^{a,b,c}\right)^2 du = \frac{1}{3}hW_{s,t}^2 + hW_{s,t}\widehat{H}_{s,t} - 2hW_{s,t}\widehat{K}_{s,t} + \frac{3-4a}{3(1-a)^2}h\widehat{H}_{s,t}^2 + \frac{12}{(1-a)^2}h\widehat{K}_{s,t}^2. \tag{5.1.5}$$

Proof. Expressing the integrals in (5.0.5) and (5.0.7) in terms of $\widehat{H}_{s,t}$ and $\widehat{K}_{s,t}$ yields

$$\begin{aligned}\frac{1}{2}hW_{s,t} + h\widehat{H}_{s,t} &= \frac{1}{2}abh + \frac{1}{2}(1-2a)(b+c)h + \frac{1}{2}a\big(c + W_{s,t}\big)h,\\ \frac{1}{3}h^2W_{s,t} + \frac{1}{2}h^2\widehat{H}_{s,t} - h^2\widehat{K}_{s,t} &= \frac{1}{3}a^2bh^2 + \frac{1}{6}(1-2a)((1+a)b + (2-a)c)h^2\\ &\quad + \frac{1}{6}a\big((3-2a)c + (3-a)W_{s,t}\big)h^2.\end{aligned}$$

Simplifying the LHS and RHS of these equations and dividing by h (and h^2) produces

$$\begin{aligned}\frac{1}{2}W_{s,t} + \widehat{H}_{s,t} &= \frac{1}{2}(1-a)(b+c) + \frac{1}{2}aW_{s,t},\\ \frac{1}{12}W_{s,t} + \frac{1}{4}(1-a)(b+c) + \frac{1}{4}aW_{s,t} - \widehat{K}_{s,t} &= \frac{1}{6}(1-a)b + \frac{1}{3}(1-a)c + \frac{1}{6}a(3-a)W_{s,t}.\end{aligned}$$

Solving this system for $\widehat{H}_{s,t}$ and $\widehat{K}_{s,t}$ is straightforward and gives (5.1.3) and (5.1.4).

Moreover, after further rearranging, we obtain explicit formulae for $b+c$ and $b-c$.

$$\begin{aligned}b + c &= W_{s,t} + \frac{2}{1-a}\widehat{H}_{s,t},\\ b - c &= \frac{12}{1-a}\widehat{K}_{s,t} - (1-2a)W_{s,t}.\end{aligned}$$

We can derive (5.1.5) by substituting the above expressions into the RHS of (5.0.6).

$$\begin{aligned}\int_s^t \left(\gamma_{s,u}^{a,b,c}\right)^2 du &= \frac{1}{3}ab^2h + \frac{1}{3}(1-2a)\big(b^2 + bc + c^2\big)h + \frac{1}{3}a\big(c^2 + cW_{s,t} + W_{s,t}^2\big)h\\ &= \frac{1}{12}(3-4a)(b+c)^2h + \frac{1}{12}(b-c)^2h + \frac{1}{3}achW_{s,t} + \frac{1}{3}ahW_{s,t}^2\\ &= \frac{1}{12}(3-4a)\Big(W_{s,t} + \frac{2}{1-a}\widehat{H}_{s,t}\Big)^2h + \frac{1}{12}\Big(\frac{12}{1-a}\widehat{K}_{s,t} - (1-2a)W_{s,t}\Big)^2h\\ &\quad + \frac{1}{6}a\Big(W_{s,t} + \frac{2}{1-a}\widehat{H}_{s,t} - \frac{12}{1-a}\widehat{K}_{s,t} + (1-2a)W_{s,t}\Big)hW_{s,t} + \frac{1}{3}ahW_{s,t}^2.\end{aligned}$$

The result now follows by simplifying the various terms in the above equation. □

The second ingredient that we use for constructing the piecewise linear paths are the conditional moments of Lévy area $L_{s,t}$. Recall that Theorem 4.2.8 established the

mean and variance of $L_{s,t}$ conditional on the tuples $(W_{s,t}, H_{s,t})$ and $(W_{s,t}, H_{s,t}, K_{s,t})$. Since we intend on using the Rademacher random variables $n_{s,t}$ and $k_{s,t}$ as cheaper alternatives to $K_{s,t}$, we will derive the corresponding moments of both $L_{s,t}$ and $K_{s,t}$.

Theorem 5.1.2 (Additional conditional moments of space-space-time Lévy area)**.** *Let W be a standard Brownian motion and $H_{s,t}, L_{s,t}$ be the previously defined areas. Let $n_{s,t}$ and $k_{s,t}$ be the discrete random variables given by definitions 5.0.4 and 5.0.5. Then the conditional means and variances of $L_{s,t}$ associated with this information are*

$$\mathbb{E}\big[L_{s,t} \,\big|\, W_{s,t}, H_{s,t}, n_{s,t}\big] = \frac{1}{30}h^2 + \frac{3}{5}hH_{s,t}^2 - \frac{1}{8\sqrt{6\pi}}\, n_{s,t}\, h^{\frac{3}{2}} W_{s,t}\,, \tag{5.1.6}$$

$$\begin{aligned}\mathrm{Var}\big(L_{s,t} \,\big|\, W_{s,t}, H_{s,t}, n_{s,t}\big) = {}& \frac{11}{25200}h^4 + \left(\frac{1}{720} - \frac{1}{384\pi}\right) h^3 W_{s,t}^2 + \frac{1}{700}h^3 H_{s,t}^2 \\ & - \frac{1}{320\sqrt{6\pi}}\, n_{s,t}\, h^{\frac{7}{2}} W_{s,t}\,,\end{aligned} \tag{5.1.7}$$

$$\mathbb{E}\big[L_{s,t} \,\big|\, W_{s,t}, H_{s,t}, k_{s,t}\big] = \frac{1}{30}h^2 + \frac{3}{5}hH_{s,t}^2 - \frac{1}{6\sqrt{10\pi}}\, k_{s,t}\, h^{\frac{3}{2}} W_{s,t}\,, \tag{5.1.8}$$

$$\begin{aligned}\mathrm{Var}\big(L_{s,t} \,\big|\, W_{s,t}, H_{s,t}, k_{s,t}\big) = {}& \frac{11}{25200}h^4 + \left(\frac{1}{720} - \frac{1}{360\pi}\right) h^3 W_{s,t}^2 + \frac{1}{700}h^3 H_{s,t}^2 \\ & - \frac{1}{252\sqrt{10\pi}}\, k_{s,t}\, h^{\frac{7}{2}} W_{s,t}\,,\end{aligned} \tag{5.1.9}$$

where $h = t - s$.

Proof. In order to show (5.1.6) and (5.1.7), we shall use the following random variable,

$$N_{s,t} := H_{s,u} - H_{u,t}, \tag{5.1.10}$$

where $u := s + \frac{1}{2}h$ is the midpoint of the interval.

Our strategy is to divide $[s,t]$ in half and apply Theorem 4.2.8 on both subintervals. The result then follows after a further conditional expectation is taken (by Tower law). It is worth noting that Theorem 4.1.12 allows us to express the Brownian motion as

$$W_{s,r} = \frac{r-s}{h}\, W_{s,t} + \frac{6(r-s)(t-r)}{h^2}\, H_{s,t} + Z_{s,r}\,, \quad \forall r \in [s,t], \tag{5.1.11}$$

where Z denotes a Brownian arch process on $[s,t]$ that is independent of $(W_{s,t}, H_{s,t})$.

Recall that in the proof of Theorem 5.0.6, it was shown that $N_{s,t}$ is independent of $(W_{s,u}, W_{u,t}, H_{s,t})$. In particular, this implies that $W_{s,t}, H_{s,t}, N_{s,t}, Z_{s,u}$ are independent. By considering the subregions which contribute to the space-time Lévy area, we have

$$hH_{s,t} = \frac{1}{2}hH_{s,u} + \frac{1}{2}hH_{u,t} + \frac{1}{2}h\Big(W_{s,u} - \frac{1}{2}W_{s,t}\Big). \tag{5.1.12}$$

Using equations (5.1.10), (5.1.11) and (5.1.12), it is straightforward to express the Gaussian random variables $\{W_{s,u}, W_{u,t}, H_{s,u}, H_{u,t}\}$ in terms of $\{W_{s,t}, H_{s,t}, N_{s,t}, Z_{s,u}\}$.

$$W_{s,u} = \frac{1}{2}W_{s,t} + \frac{3}{2}H_{s,t} + Z_{s,u}, \tag{5.1.13}$$

$$W_{u,t} = \frac{1}{2}W_{s,t} - \frac{3}{2}H_{s,t} - Z_{s,u}, \tag{5.1.14}$$

$$H_{s,u} = \frac{1}{4}H_{s,t} - \frac{1}{2}Z_{s,u} + \frac{1}{2}N_{s,t}, \tag{5.1.15}$$

$$H_{u,t} = \frac{1}{4}H_{s,t} - \frac{1}{2}Z_{s,u} - \frac{1}{2}N_{s,t}. \tag{5.1.16}$$

In a similar fashion to (5.1.12), the third order iterated integrals of Brownian motion and time can be written as a linear combination of integrals over the two half intervals.

$$\int_s^t W_{s,r}^2\,dr = \int_s^u W_{s,r}^2\,dr + \int_u^t W_{s,u}^2\,dr + 2\int_u^t W_{s,u}W_{u,r}\,dr + \int_u^t W_{u,r}^2\,dr. \tag{5.1.17}$$

Since W has independent increments, applying Theorem 4.2.8 to the above will give

$$\begin{aligned}\mathbb{E}\left[\int_s^t W_{s,r}^2\,dr \,\middle|\, W_{s,u}, W_{u,t}, H_{s,u}, H_{u,t}\right] &= \frac{1}{6}hW_{s,u}^2 + \frac{1}{2}hW_{s,u}H_{s,u} + \frac{1}{60}h^2 + \frac{3}{5}hH_{s,u}^2 \\ &\quad + \frac{1}{2}hW_{s,u}^2 + \frac{1}{2}hW_{s,u}W_{u,t} + hW_{s,u}H_{u,t} \\ &\quad + \frac{1}{6}hW_{u,t}^2 + \frac{1}{2}hW_{u,t}H_{u,t} + \frac{1}{60}h^2 + \frac{3}{5}hH_{u,t}^2.\end{aligned}$$

Before taking the expectation of the above conditional on $(W_{s,t}, H_{s,t}, n_{s,t})$, we shall express the LHS in terms of $\{W_{s,t}, H_{s,t}, N_{s,t}, Z_{s,u}\}$ since only $N_{s,t}$ is dependent on $n_{s,t}$. So after the four equations (5.1.13) – (5.1.16) are substituted into the above, we have

$$\begin{aligned}\mathbb{E}\left[\int_s^t W_{s,r}^2\,dr \,\middle|\, W_{s,u}, W_{u,t}, H_{s,u}, H_{u,t}\right] &= \frac{1}{3}hW_{s,t}^2 + hW_{s,t}H_{s,t} + \frac{1}{30}h^2 \\ &\quad + \frac{6}{5}hH_{s,t}^2 + \frac{2}{15}hZ_{s,u}^2 + \frac{3}{10}hN_{s,t}^2 \\ &\quad + \frac{1}{5}hH_{s,t}Z_{s,u} - \frac{1}{4}hW_{s,t}N_{s,t}.\end{aligned} \tag{5.1.18}$$

Since $n_{s,t} = \operatorname{sgn}(N_{s,t})$, we see that $N_{s,t}|n_{s,t}$ has a half-normal distribution (times $n_{s,t}$). Moreover, as $N_{s,t} \sim \mathcal{N}(0, \frac{1}{12}h)$, the following moments can be computed analytically:

$$\mathbb{E}\big[N_{s,t} \,\big|\, n_{s,t}\big] = \frac{1}{\sqrt{6\pi}}\,n_{s,t}h^{\frac{1}{2}}, \tag{5.1.19}$$

$$\mathbb{E}\big[N_{s,t}^2 \,\big|\, n_{s,t}\big] = \frac{1}{12}h, \tag{5.1.20}$$

$$\mathbb{E}\big[N_{s,t}^3 \,\big|\, n_{s,t}\big] = \frac{1}{6\sqrt{6\pi}}\, n_{s,t} h^{\frac{3}{2}}, \tag{5.1.21}$$

$$\mathbb{E}\big[N_{s,t}^4 \,\big|\, n_{s,t}\big] = \frac{3}{144} h^2. \tag{5.1.22}$$

(The raw moments of a half-normal distribution on $[0,\infty)$ are explicitly given in [95])

Since $Z_{s,u}$ is independent of $(W_{s,t}, H_{s,t}, N_{s,t})$ and $N_{s,t}$ is independent of $(W_{s,t}, H_{s,t})$, the expectation of (5.1.18) conditional on $(W_{s,t}, H_{s,t}, n_{s,t})$ can simply be computed as

$$\begin{aligned}
&\mathbb{E}\left[\int_s^t W_{s,r}^2\, dr \,\middle|\, W_{s,t}, H_{s,t}, n_{s,t}\right] \\
&\quad = \frac{1}{3} h W_{s,t}^2 + h W_{s,t} H_{s,t} + \frac{1}{30} h^2 + \frac{6}{5} h H_{s,t}^2 \\
&\qquad\qquad + \frac{2}{15} h \mathbb{E}\big[Z_{s,u}^2\big] + \frac{3}{10} h \mathbb{E}\big[N_{s,t}^2 \,\big|\, n_{s,t}\big] \\
&\qquad\qquad + \frac{1}{5} h H_{s,t} \mathbb{E}\big[Z_{s,u}\big] - \frac{1}{4} h W_{s,t} \mathbb{E}\big[N_{s,t} \,\big|\, n_{s,t}\big] \\
&\quad = \frac{1}{3} h W_{s,t}^2 + h W_{s,t} H_{s,t} + \frac{1}{15} h^2 + \frac{6}{5} h H_{s,t}^2 - \frac{1}{4\sqrt{6\pi}}\, n_{s,t}\, h^{\frac{3}{2}} W_{s,t},
\end{aligned}$$

where we use the moments $\mathbb{E}\big[Z_{s,u}\big] = 0$ and $\mathbb{E}\left[Z_{s,u}^2\right] = \frac{1}{16} h$ (see definition 4.1.14). The conditional expectation (5.1.6) now follows by applying lemma 4.2.5 to the above.

The conditional variance (5.1.7) can also be computed via the decomposition (5.1.17).

$$\begin{aligned}
&\mathrm{Var}\left(\int_s^t W_{s,r}^2\, dr \,\middle|\, W_{s,u}, W_{u,t}, H_{s,u}, H_{u,t}\right) \\
&= \mathrm{Var}\left(\int_s^u W_{s,r}^2\, dr + \int_u^t \left(W_{s,u}^2 + 2 W_{s,u} W_{u,r}\right) dr + \int_u^t W_{u,r}^2\, dr \,\middle|\, W_{s,u}, W_{u,t}, H_{s,u}, H_{u,t}\right) \\
&= \mathrm{Var}\left(\int_s^u W_{s,r}^2\, dr \,\middle|\, W_{s,u}, W_{u,t}, H_{s,u}, H_{u,t}\right) + \mathrm{Var}\left(\int_u^t W_{u,r}^2\, dr \,\middle|\, W_{s,u}, W_{u,t}, H_{s,u}, H_{u,t}\right).
\end{aligned}$$

As before, we will apply Theorem 4.2.8 and change variables using (5.1.13) - (5.1.16).

$$\begin{aligned}
&\mathrm{Var}\left(\int_s^t W_{s,r}^2\, dr \,\middle|\, W_{s,u}, W_{u,t}, H_{s,u}, H_{u,t}\right) \\
&\quad = \frac{11}{50400} h^4 + h^3 \left(\frac{1}{1440} W_{s,u}^2 + \frac{1}{1400} H_{s,u}^2 + \frac{1}{1440} W_{u,t}^2 + \frac{1}{1400} H_{u,t}^2\right) \\
&\quad = \frac{11}{50400} h^4 + h^3 \left(\frac{1}{2880} W_{s,t}^2 + \frac{9}{2800} H_{s,t}^2 + \frac{2}{525} H_{s,t} Z_{s,u} + \frac{11}{6300} Z_{s,u}^2 + \frac{1}{2800} N_{s,t}^2\right).
\end{aligned}$$

So the second moment of the triple integral conditional on $(W_{s,u}, W_{u,t}, H_{s,u}, H_{u,t})$ is

$$\begin{aligned}
&\mathbb{E}\left[\left(\int_s^t W_{s,r}^2\,dr\right)^2 \,\middle|\, W_{s,u}, H_{s,u}, W_{u,t}, H_{u,t}\right] \\
&= \left(\mathbb{E}\left[\int_s^t W_{s,r}^2\,dr \,\middle|\, W_{s,u}, W_{u,t}, H_{s,u}, H_{u,t}\right]\right)^2 + \operatorname{Var}\left(\int_s^t W_{s,r}^2\,dr \,\middle|\, W_{s,u}, W_{u,t}, H_{s,u}, H_{u,t}\right) \\
&= \frac{11}{50400}h^4 + h^3\left(\frac{1}{2880}W_{s,t}^2 + \frac{9}{2800}H_{s,t}^2 + \frac{2}{525}H_{s,t}Z_{s,u} + \frac{11}{6300}Z_{s,u}^2 + \frac{1}{2800}N_{s,t}^2\right) \\
&\qquad + h^2\left(\frac{1}{3}W_{s,t}^2 + W_{s,t}H_{s,t} + \frac{1}{30}h + \frac{6}{5}H_{s,t}^2\right. \\
&\qquad\qquad \left. + \frac{2}{15}Z_{s,u}^2 + \frac{3}{10}N_{s,t}^2 + \frac{1}{5}H_{s,t}Z_{s,u} - \frac{1}{4}W_{s,t}N_{s,t}\right)^2 \\
&= \frac{1}{2800}h^3N_{s,t}^2 - \frac{1}{2}h^2W_{s,t}N_{s,t}\left(\frac{1}{3}W_{s,t}^2 + W_{s,t}H_{s,t} + \frac{1}{30}h + \frac{6}{5}H_{s,t}^2 + \frac{2}{15}Z_{s,u}^2 + \frac{1}{5}H_{s,t}Z_{s,u}\right) \\
&\qquad - \frac{3}{20}h^2W_{s,t}N_{s,t}^3 + \frac{1}{16}h^2W_{s,t}^2N_{s,t}^2 + \frac{9}{100}h^2N_{s,t}^4 + \textit{additional terms}.
\end{aligned}$$

Since the additional terms do not contain $N_{s,t}$ or $N_{s,t}^3$, they will be independent of $n_{s,t}$. Thus we can compute the expectation of these terms conditional on $(W_{s,t}, H_{s,t}, n_{s,t})$.

$$\begin{aligned}
&\mathbb{E}\Big[\textit{additional terms} \,\Big|\, W_{s,t}, H_{s,t}, n_{s,t}\Big] \\
&\quad = \mathbb{E}\left[\left(\int_s^t W_{s,r}^2\,dr\right)^2 - \frac{1}{2800}h^3N_{s,t}^2 - \frac{1}{16}h^2W_{s,t}^2N_{s,t}^2 - \frac{9}{100}h^2N_{s,t}^4 \,\middle|\, W_{s,t}, H_{s,t}\right].
\end{aligned}$$

Applying Theorem 4.2.8 and noting that $N_{s,t}^2$ and $N_{s,t}^4$ are independent of $n_{s,t}$ gives

$$\begin{aligned}
&\mathbb{E}\left[\left(\int_s^t W_{s,r}^2\,dr\right)^2 \,\middle|\, W_{s,t}, H_{s,t}, n_{s,t}\right] \\
&= \left(\frac{1}{3}hW_{s,t}^2 + hW_{s,t}H_{s,t} + \frac{6}{5}hH_{s,t}^2 + \frac{1}{15}h^2\right)^2 + \frac{11}{6300}h^4 + h^3\left(\frac{1}{180}W_{s,t}^2 + \frac{1}{175}H_{s,t}^2\right) \\
&- \frac{1}{2}h^2W_{s,t}\,\mathbb{E}\big[N_{s,t} \,\big|\, n_{s,t}\big]\left(\frac{1}{3}W_{s,t}^2 + W_{s,t}H_{s,t} + \frac{1}{24}h + \frac{6}{5}H_{s,t}^2\right) - \frac{3}{20}h^2W_{s,t}\,\mathbb{E}\big[N_{s,t}^3 \,\big|\, n_{s,t}\big].
\end{aligned}$$

Finally, we can derive an explicit formula for the conditional variance of this integral.

$$\begin{aligned}
&\operatorname{Var}\left(\int_s^t W_{s,r}^2\,dr \,\middle|\, W_{s,t}, H_{s,t}, n_{s,t}\right) \\
&\quad = \mathbb{E}\left[\left(\int_s^t W_{s,r}^2\,dr\right)^2 \,\middle|\, W_{s,t}, H_{s,t}, n_{s,t}\right] - \left(\mathbb{E}\left[\int_s^t W_{s,r}^2\,dr \,\middle|\, W_{s,t}, H_{s,t}, n_{s,t}\right]\right)^2.
\end{aligned}$$

Substituting in the above formulae for the (conditional) moments of $\int_s^t W_{s,r}^2\,dr$ gives

$$\begin{aligned}
&\operatorname{Var}\left(\int_s^t W_{s,r}^2\,dr \,\middle|\, W_{s,t}, H_{s,t}, n_{s,t}\right)\\
&\quad= \left(\frac{1}{3}hW_{s,t}^2 + hW_{s,t}H_{s,t} + \frac{1}{15}h^2 + \frac{6}{5}hH_{s,t}^2\right)^2 + \frac{11}{6300}h^4 + h^3\left(\frac{1}{180}W_{s,t}^2 + \frac{1}{175}H_{s,t}^2\right)\\
&\qquad- \frac{1}{2\sqrt{6\pi}}\,n_{s,t}\,h^{\frac{5}{2}}W_{s,t}\left(\frac{1}{3}W_{s,t}^2 + W_{s,t}H_{s,t} + \frac{1}{24}h + \frac{6}{5}H_{s,t}^2\right) - \frac{1}{40\sqrt{6\pi}}\,n_{s,t}\,h^{\frac{7}{2}}W_{s,t}\\
&\qquad- \left(\frac{1}{3}hW_{s,t}^2 + hW_{s,t}H_{s,t} + \frac{1}{15}h^2 + \frac{6}{5}hH_{s,t}^2 - \frac{1}{4\sqrt{6\pi}}\,n_{s,t}\,h^{\frac{3}{2}}W_{s,t}\right)^2\\
&\quad= \frac{11}{6300}h^4 + \left(\frac{1}{180} - \frac{1}{96\pi}\right)h^3W_{s,t}^2 + \frac{1}{175}h^3H_{s,t}^2 - \frac{1}{80\sqrt{6\pi}}\,n_{s,t}\,h^{\frac{7}{2}}W_{s,t}.
\end{aligned}$$

The conditional variance (5.1.7) is now a consequence of the above with lemma 4.2.5.

To derive (5.1.8), we do not need subintervals and can directly apply Theorem 4.2.8.

$$\begin{aligned}
\mathbb{E}\big[L_{s,t}\,\big|\,W_{s,t}, H_{s,t}, k_{s,t}\big] &= \mathbb{E}\Big[\,\mathbb{E}\big[L_{s,t}\,\big|\,W_{s,t}, H_{s,t}, K_{s,t}\big]\,\Big|\,W_{s,t}, H_{s,t}, k_{s,t}\Big]\\
&= \mathbb{E}\left[\frac{3}{140}h^2 + \frac{3}{5}hH_{s,t}^2 - hW_{s,t}K_{s,t} + \frac{60}{7}hK_{s,t}^2\,\middle|\,W_{s,t}, H_{s,t}, k_{s,t}\right]\\
&= \frac{1}{30}h^2 + \frac{3}{5}hH_{s,t}^2 - \frac{1}{6\sqrt{10\pi}}\,k_{s,t}\,h^{\frac{3}{2}}W_{s,t},
\end{aligned}$$

where we use the half-normal moments $\mathbb{E}\big[K_{s,t}\,\big|\,k_{s,t}\big] = \frac{\sqrt{h}k_{s,t}}{6\sqrt{10\pi}}$ and $\mathbb{E}\big[K_{s,t}^2\,\big|\,k_{s,t}\big] = \frac{1}{720}h$, with the independence of $(W_{s,t}, H_{s,t})$ and $k_{s,t}$ (since the latter only depends on $K_{s,t}$).

All that remains is to compute the final conditional variance (5.1.9) using the above. As before, we will begin by obtaining the second moment of $L_{s,t}$ from Theorem 4.2.8.

$$\begin{aligned}
&\mathbb{E}\Big[L_{s,t}^2\,\Big|\,W_{s,t}, H_{s,t}, K_{s,t}\Big]\\
&\quad= \operatorname{Var}\Big(L_{s,t}\,\Big|\,W_{s,t}, H_{s,t}, K_{s,t}\Big) + \Big(\mathbb{E}\big[L_{s,t}\,\big|\,W_{s,t}, H_{s,t}, K_{s,t}\big]\Big)^2\\
&\quad= \frac{11}{88200}h^4 + h^3\left(\frac{1}{700}H_{s,t}^2 + \frac{1}{49}K_{s,t}^2\right) + \left(\frac{3}{140}h^2 + \frac{3}{5}hH_{s,t}^2 - hW_{s,t}K_{s,t} + \frac{60}{7}hK_{s,t}^2\right)^2\\
&\quad= \frac{103}{176400}h^4 + h^3\left(\frac{19}{700}H_{s,t}^2 + \frac{19}{49}K_{s,t}^2\right) + \frac{9}{25}h^2H_{s,t}^4 + h^2W_{s,t}^2K_{s,t}^2 + \frac{72}{7}h^2H_{s,t}^2K_{s,t}^2\\
&\qquad\qquad+ \frac{3600}{49}h^2K_{s,t}^4 - \frac{3}{70}h^3W_{s,t}K_{s,t} - \frac{6}{5}h^2W_{s,t}H_{s,t}^2K_{s,t} - \frac{120}{7}h^2W_{s,t}K_{s,t}^3.
\end{aligned}$$

Substituting the moments $\mathbb{E}\big[K_{s,t}^3 \,|\, k_{s,t}\big] = \frac{k_{s,t}}{2160\sqrt{10\pi}} h^{\frac{3}{2}}$ and $\mathbb{E}\big[K_{s,t}^4 \,|\, k_{s,t}\big] = \frac{1}{1728000} h^2$, (which can be calculated using [95]) after a conditional expectation is taken produces

$$\mathbb{E}\Big[L_{s,t}^2 \,\big|\, W_{s,t}, H_{s,t}, k_{s,t}\Big] = \frac{13}{8400} h^4 + h^3\Big(\frac{1}{720} W_{s,t}^2 + \frac{29}{700} H_{s,t}^2\Big) + \frac{9}{25} h^2 H_{s,t}^4 - \frac{19}{1260\sqrt{10\pi}}\, k_{s,t}\, h^{\frac{7}{2}} W_{s,t} - \frac{1}{5\sqrt{10\pi}}\, k_{s,t}\, h^{\frac{7}{2}} W_{s,t} H_{s,t}^2.$$

To conclude the proof, we explicitly compute the conditional variance given by (5.1.9).

$$\begin{aligned}
\mathrm{Var}\Big(L_{s,t} \,\big|\, W_{s,t}, H_{s,t}, k_{s,t}\Big) &= \mathbb{E}\Big[L_{s,t}^2 \,\big|\, W_{s,t}, H_{s,t}, k_{s,t}\Big] - \Big(\mathbb{E}\big[L_{s,t} \,\big|\, W_{s,t}, H_{s,t}, k_{s,t}\big]\Big)^2 \\
&= \frac{13}{8400} h^4 + h^3\Big(\frac{1}{720} W_{s,t}^2 + \frac{29}{700} H_{s,t}^2\Big) + \frac{9}{25} h^2 H_{s,t}^4 \\
&\quad - \frac{19}{1260\sqrt{10\pi}}\, k_{s,t}\, h^{\frac{7}{2}} W_{s,t} - \frac{1}{5\sqrt{10\pi}}\, k_{s,t}\, h^{\frac{7}{2}} W_{s,t} H_{s,t}^2 \\
&\quad - \Big(\frac{1}{30} h^2 + \frac{3}{5} h H_{s,t}^2 - \frac{1}{6\sqrt{10\pi}}\, k_{s,t}\, h^{\frac{3}{2}} W_{s,t}\Big)^2 \\
&= \frac{11}{25200} h^4 + \Big(\frac{1}{720} - \frac{1}{360\pi}\Big) h^3 W_{s,t}^2 + \frac{1}{700} h^3 H_{s,t}^2 \\
&\quad - \frac{1}{252\sqrt{10\pi}}\, k_{s,t}\, h^{\frac{7}{2}} W_{s,t}.
\end{aligned}$$

□

To derive formulae for the piecewise linear paths, we also use the moments of $K_{s,t}$.

Theorem 5.1.3 (Conditional moments of the rescaled space-time-time Lévy area)**.** *Consider a standard Brownian motion W with space-time Lévy area $H_{s,t}$ over $[s,t]$. Let $n_{s,t}$ and $k_{s,t}$ be the discrete random variables given by definitions 5.0.4 and 5.0.5. Then the conditional means and variances of $K_{s,t}$ associated with this information are*

$$\mathbb{E}\Big[K_{s,t} \,\big|\, W_{s,t}, H_{s,t}, n_{s,t}\Big] = \frac{1}{8\sqrt{6\pi}} n_{s,t} \sqrt{h}\,, \tag{5.1.23}$$

$$\mathrm{Var}\Big(K_{s,t} \,\big|\, W_{s,t}, H_{s,t}, n_{s,t}\Big) = \frac{1}{720} h^2\,, \tag{5.1.24}$$

$$\mathbb{E}\Big[K_{s,t} \,\big|\, W_{s,t}, H_{s,t}, k_{s,t}\Big] = \frac{1}{6\sqrt{10\pi}} k_{s,t} \sqrt{h}\,, \tag{5.1.25}$$

$$\mathrm{Var}\Big(K_{s,t} \,\big|\, W_{s,t}, H_{s,t}, k_{s,t}\Big) = \frac{1}{720} h^2\,, \tag{5.1.26}$$

where $h = t - s$.

Proof. As in the previous proof, we shall introduce the Gaussian random variable $N_{s,t} := H_{s,u} - H_{u,t}$, where $u := s + \frac{1}{2}h$ denotes the midpoint of the interval, so that

$$\begin{aligned}
&\mathbb{E}\bigg[\int_s^t (r-s)^2 \, dW_r \,\Big|\, W_{s,u}, W_{u,t}, H_{s,u}, H_{u,t} \bigg] \\
&\quad = \mathbb{E}\bigg[\int_s^u (r-s)^2 \, dW_r + \int_u^t (r-s)^2 \, dW_r \,\Big|\, W_{s,u}, W_{u,t}, H_{s,u}, H_{u,t} \bigg] \\
&\quad = \mathbb{E}\bigg[\int_s^u (r-s)^2 \, dW_r \,\Big|\, W_{s,u}, H_{s,u} \bigg] + \mathbb{E}\bigg[\int_u^t (r-s)^2 \, dW_r \,\Big|\, W_{u,t}, H_{u,t} \bigg] \\
&\quad = \frac{1}{12} h^2 W_{s,u} - \frac{1}{4} h^2 H_{s,u} + \frac{7}{12} h^2 W_{u,t} - \frac{3}{4} h^2 H_{u,t} \\
&\quad = \frac{1}{3} h^2 W_{s,t} - h^2 H_{s,t} + \frac{1}{4} h^2 N_{s,t},
\end{aligned}$$

where the above was computed by applying Theorem 4.1.12 on the two half intervals. Taking the expectation conditional on $(W_{s,t}, H_{s,t}, n_{s,t})$ and using lemma 4.2.6 yields

$$\mathbb{E}\big[K_{s,t} \,\big|\, W_{s,t}, H_{s,t}, n_{s,t} \big] = \mathbb{E}\bigg[\frac{1}{8} N_{s,t} \,\Big|\, W_{s,t}, H_{s,t}, n_{s,t} \bigg] = \frac{1}{8\sqrt{6\pi}} n_{s,t} \sqrt{h} \,.$$

Clearly the value of $K_{s,t}^2$ does not change if the Brownian motion is multiplied by -1. On the other hand, the value of $n_{s,t}$ would be flipped if W were replaced by $-W$. Therefore, by this symmetry, $K_{s,t}^2$ and $n_{s,t}$ are independent random variables and so

$$\mathbb{E}\big[K_{s,t}^2 \,\big|\, W_{s,t}, H_{s,t}, n_{s,t} \big] = \mathbb{E}\big[K_{s,t}^2 \,\big|\, n_{s,t} \big] = \mathbb{E}\big[K_{s,t}^2 \big] = \frac{1}{720} h.$$

The moments (5.1.25) and (5.1.26) are elementary to derive as $k_{s,t} = \operatorname{sgn}(K_{s,t})$. □

Finally, we can now incorporate the previous two theorems (5.1.2 and 5.1.3) with lemma 5.1.1 to define high order piecewise linear paths in the three-parameter family.

Theorem 5.1.4. *Let $[s,t]$ be a fixed interval and $(W, H, \cdot)_{s,t}$ denote one of the triples:*

- $(W_{s,t}, H_{s,t}, K_{s,t})$
- $(W_{s,t}, H_{s,t}, n_{s,t})$
- $(W_{s,t}, H_{s,t}, k_{s,t})$

For $a \in (0, \frac{1}{2})$, suppose there exists a three-parameter piecewise linear path satisfying

$$b + c = W_{s,t} + \frac{2}{1-a} H_{s,t}, \tag{5.1.27}$$

$$\begin{aligned}
b - c = a W_{s,t} + \varepsilon \Big(&(1-a)^2 W_{s,t}^2 - 24 W_{s,t} \, \mathbb{E}\big[K_{s,t} \,\big|\, (W, H, \cdot)_{s,t} \big] \\
&+ \frac{72a^2 - 64a + 12}{5(1-a)^2} H_{s,t}^2 + \frac{1440}{7} \, \mathbb{E}\Big[K_{s,t}^2 \,\Big|\, (W, H, \cdot)_{s,t} \Big] + \frac{18}{35} h \Big)^{\frac{1}{2}},
\end{aligned} \tag{5.1.28}$$

where

$$\varepsilon = \begin{cases} 1 & \text{if } \mathbb{E}\big[K_{s,t}\big|\,(W,H,\cdot)_{s,t}\,\big] \geq \frac{(1-a)^2}{12} W_{s,t} \\ -1 & \text{if } \mathbb{E}\big[K_{s,t}\big|\,(W,H,\cdot)_{s,t}\,\big] < \frac{(1-a)^2}{12} W_{s,t} \end{cases}.$$

Then

1. $\int_s^t \gamma_{s,u}^{a,b,c}\,du = \int_s^t W_{s,u}\,du$.
2. $\int_s^t \big(\gamma_{s,u}^{a,b,c}\big)^2\,du = \mathbb{E}\Big[\int_s^t W_{s,u}^2\,du\,\Big|\,(W,H,\cdot)_{s,t}\,\Big]$.
3. *For any three-parameter piecewise linear path $\gamma^{a,b',c'}$ on $[s,t]$ with properties 1-2 and the same value of a as $\gamma^{a,b,c}$, we have the following inequality*

$$\begin{aligned} &\Big|\int_s^t (u-s)\gamma_{s,u}^{a,b,c}\,du - \mathbb{E}\Big[\int_s^t (u-s)W_{s,u}\,du\,\Big|\,(W,H,\cdot)_{s,t}\,\Big]\Big| \\ &\qquad \leq \Big|\int_s^t (u-s)\gamma_{s,u}^{a,b',c'}\,du - \mathbb{E}\Big[\int_s^t (u-s)W_{s,u}\,du\,\Big|\,(W,H,\cdot)_{s,t}\,\Big]\Big|. \end{aligned}$$

Proof. Let $\gamma^{a,b,c}$ have space-time Lévy area $\widehat{H}_{s,t}$ and space-time-time Lévy area $\widehat{K}_{s,t}$. Then by lemma 5.1.1, we can explicitly compute these areas from (5.1.27) and (5.1.28).

$$\widehat{H}_{s,t} = \frac{1}{2}(1-a)\big(b+c-W_{s,t}\big) = H_{s,t}, \tag{5.1.29}$$

$$\widehat{K}_{s,t} = \frac{1}{12}(1-a)\Big((1-a)W_{s,t} + \varepsilon\,(\cdots)^{\frac{1}{2}}\Big). \tag{5.1.30}$$

Therefore $\gamma^{a,b,c}$ satisfies the first property and applying (5.1.5) in lemma 5.1.1 yields

$$\begin{aligned} \int_s^t \big(\gamma_{s,u}^{a,b,c}\big)^2\,du &= \frac{1}{3}hW_{s,t}^2 + hW_{s,t}H_{s,t} - 2hW_{s,t}\widehat{K}_{s,t} + \frac{3-4a}{3(1-a)^2}hH_{s,t}^2 + \frac{12}{(1-a)^2}h\widehat{K}_{s,t}^2 \\ &= \frac{1}{3}hW_{s,t}^2 + hW_{s,t}H_{s,t} - \frac{1}{12}(1-a)^2hW_{s,t}^2 + \frac{3-4a}{3(1-a)^2}hH_{s,t}^2 \\ &\qquad + \frac{1}{12}\bigg((1-a)^2W_{s,t}^2 - 24W_{s,t}\,\mathbb{E}\big[K_{s,t}\big|\,(W,H,\cdot)_{s,t}\,\big] + \frac{18}{35}h \\ &\qquad\qquad + \frac{72a^2-64a+12}{5(1-a)^2}H_{s,t}^2 + \frac{1440}{7}\,\mathbb{E}\Big[K_{s,t}^2\Big|\,(W,H,\cdot)_{s,t}\,\Big]\bigg) \\ &= \frac{1}{3}hW_{s,t}^2 + hW_{s,t}H_{s,t} + \frac{6}{5}hH_{s,t}^2 + \frac{3}{70}h^2 - 2W_{s,t}\mathbb{E}\big[K_{s,t}\big|\,(W,H,\cdot)_{s,t}\,\big] \\ &\qquad + \frac{120}{7}\,\mathbb{E}\Big[K_{s,t}^2\Big|\,(W,H,\cdot)_{s,t}\,\Big]. \end{aligned}$$

By Theorems 4.2.8, 5.1.2 and 5.1.3, it follows that $\gamma^{a,b,c}$ satisfies the second property.

In order to show the third property, we should identify all the three-parameter piecewise linear paths with the same value for a as $\gamma^{a,b,c}$ and the first two properties. Once a is fixed, it is clear from (5.1.29) that the first property will determine $b+c$. The second property can then be rewritten as a quadratic equation in $\widehat{K}_{s,t}$ by (5.1.5). Hence by (5.1.4), there are at most two three-parameter paths with properties 1 - 2. Moreover, it follows by direct calculation that the roots of this quadratic equation correspond to the formula (5.1.28) for $b-c$ where $\varepsilon \in \{-1, 1\}$ is determined by $\widehat{K}_{s,t}$. Since $\widehat{H}_{s,t} = H_{s,t}$, it is evident that the third property is equivalent to minimizing $\left|\widehat{K}_{s,t} - \mathbb{E}[K_{s,t} \big| (W, H, \cdot)_{s,t}]\right|$ and the required value of ε can be seen using (5.1.30). $\square$

The above theorem explicitly gives a method for constructing piecewise linear paths that capture certain iterated integrals of the original Brownian path and time. That said, there are still two important details that Theorem 5.1.4 does not address.

- What type of information $(K_{s,t}, n_{s,t}, k_{s,t})$ should be used in an SDE simulation?
- What value of a should be chosen for the three-parameter piecewise linear path?

Whilst these questions are not fully resolved, there is plenty that can be discussed. If a vector field in the SDE is expensive to evaluate, then the choice of random variables we generate per step makes little difference to the simulation's complexity. Therefore in this case, it makes sense to use $K_{s,t}$ to construct the piecewise linear path as it results in a superior estimator for $L_{s,t}$ when compared to either $n_{s,t}$ or $k_{s,t}$. Ironically, the effectiveness of this estimator can cause additional challenges if one is discretizing an SDE using the variable step size methodology proposed in Chapter 6. This is because the main term in the local error expansion is independent of $W_{s,t}$ as

$$\operatorname{Var}\Big(L_{s,t} \,\Big|\, W_{s,t}, H_{s,t}, K_{s,t}\Big) = \frac{11}{88200} h^4 + h^3 \left(\frac{1}{700} H_{s,t}^2 + \frac{1}{49} K_{s,t}^2\right).$$

So in this regime, it is unclear how one should control the size of the increment $W_{s,t}$.

Alternatively, if the vector fields governing the SDE are all cheap to compute, then the sampling of random variables during each step may be become a bottleneck. Hence in this setting, both $n_{s,t}$ and $k_{s,t}$ can be seen as cheaper alternatives than $K_{s,t}$. Theorem 5.1.2 shows that $k_{s,t}$ gives a slightly better approximation of $L_{s,t}$ than $n_{s,t}$. That said, we shall see in the next chapter that $n_{s,t}$ is much more straightforward to incorporate into variable step size methods and can be used to control $(W_{s,t}, H_{s,t})$ as

$$\begin{aligned}&\operatorname{Var}\Big(L_{s,t} \,\Big|\, W_{s,t}, H_{s,t}, n_{s,t}\Big) \\ &\quad = \frac{11}{25200} h^4 + \left(\frac{1}{720} - \frac{1}{384\pi}\right) h^3 W_{s,t}^2 + \frac{1}{700} h^3 H_{s,t}^2 - \frac{1}{320\sqrt{6\pi}}\, n_{s,t}\, h^{\frac{7}{2}} W_{s,t}.\end{aligned}$$

A natural choice for a is $\frac{8-\sqrt{10}}{18}$ since it ensures that $b-c$ is independent of $H_{s,t}$. This may be desirable since Brownian motion has this type of independence property:

Proposition 5.1.5. *The process $\{W_{s,r}-W_{r,t}\}_{r\in[s,t]}$ is independent of the area $H_{s,t}$.*

Proof. By Theorem 4.1.12, we have that $W_{s,r}-W_{r,t}=Z_{s,r}-Z_{r,t}$, where Z denotes the Brownian arch process. The result now follows as Z and $H_{s,t}$ are independent. □

Using this choice of a, we can now define the paths corresponding to $n_{s,t}$ and $k_{s,t}$.

Definition 5.1.6. Consider the unique three-parameter path $\gamma^{a,b,c}$ on $[s,t]$ satisfying

$$\begin{aligned} a &= \frac{8-\sqrt{10}}{18}, \\ b+c &= W_{s,t}+\frac{2}{1-a}H_{s,t}, \\ b-c &= aW_{s,t}+\varepsilon\sqrt{(1-a)^2W_{s,t}^2-C_{s,t}W_{s,t}+\frac{4}{5}h}, \end{aligned}$$

where

$$\varepsilon := \begin{cases} 1 & \text{if } \ C_{s,t}\geq 2(1-a)^2W_{s,t} \\ -1 & \text{if } \ C_{s,t}<2(1-a)^2W_{s,t} \end{cases},$$

and

$$C_{s,t} := \begin{cases} \frac{3}{\sqrt{6\pi}}n_{s,t}\sqrt{h} & \text{if } n_{s,t} \text{ is used} \\ \frac{4}{\sqrt{10\pi}}k_{s,t}\sqrt{h} & \text{if } k_{s,t} \text{ is used} \end{cases}.$$

Then $\gamma^{a,b,c}$ is a high order piecewise linear approximant of W given by Theorem 5.1.4.

Ideally, we would like to choose the above value for a when constructing piecewise linear paths that are $(W_{s,t},H_{s,t},K_{s,t})$-measurable due to this independence property. However, it is now possible for the expression in the square root to become negative.

Theorem 5.1.7. *When $a=\frac{8-\sqrt{10}}{18}$, there is a non-zero probability that a high order piecewise linear path cannot be constructed from $(W_{s,t},H_{s,t},K_{s,t})$ as in Theorem 5.1.4. Moreover, the probability of this event occurring is small and can be bounded above by*

$$\begin{aligned} &\mathbb{P}\left(|X|>\frac{3\sqrt{\frac{60+28\sqrt{5}}{430070-283500\sqrt{2}+331128\sqrt{5}-37525\sqrt{10}}}\left(3545\sqrt{14}+700\sqrt{35}+2268\sqrt{70}\right)}{140\sqrt{2}+98\sqrt{10}}\right) \\ &\quad\approx 0.0022, \end{aligned}$$

where $X\sim\mathcal{N}(0,1)$ is a standard normal random variable.

Proof. It is clear that the construction of the piecewise linear path fails if and only if

$$\frac{55+10\sqrt{10}}{162}W_{s,t}^2 - 24W_{s,t}K_{s,t} + \frac{1440}{7}K_{s,t}^2 + \frac{18}{35}h < 0.$$

Since this quadratic defines a hyperbola, the above event has a non-zero probability. In order to estimate this probability, we aim to find the largest constant C such that

$$\frac{55+10\sqrt{10}}{162}W_1^2 - 24W_1K_1 + \frac{1440}{7}K_1^2 + \frac{18}{35} < 0 \implies |W_1 + \sqrt{720}K_1| > C.$$

The constant C can be obtained by solving the below non-linear optimization problem.

$$\text{minimize } |x+y| \text{ subject to } \frac{55+10\sqrt{10}}{162}x^2 - \frac{2\sqrt{5}}{5}xy + \frac{2}{7}y^2 + \frac{18}{35} = 0.$$

With assistance from Wolfram Mathematica, the constant is explicitly computable as

$$C = \frac{3\sqrt{\frac{60+28\sqrt{5}}{430070-283500\sqrt{2}+331128\sqrt{5}-37525\sqrt{10}}}\left(3545\sqrt{14}+700\sqrt{35}+2268\sqrt{70}\right)}{140+98\sqrt{5}}$$

which is approximately 4.321. □

Remark 5.1.8. Monte Carlo simulations indicate this probability is less than 0.0005.

The above theorem shows that there can be additional complications when one considers piecewise linear paths that are measurable with respect to $(W_{s,t}, H_{s,t}, K_{s,t})$. Since we cannot always construct the three-parameter path with $a = \frac{8-\sqrt{10}}{18}$, we use

$$a = \frac{10-\sqrt{70}}{10},$$

as it is the largest constant in $(0, \frac{1}{2})$ that guarantees non-negativity for the following:

$$(1-a)^2W_{s,t}^2 - 24W_{s,t}K_{s,t} + \frac{1440}{7}K_{s,t}^2.$$

This then ensures the existence of the piecewise linear path given by Theorem 5.1.4.

A potential downside of this approach is that $b-c$ is not independent of $H_{s,t}$. Therefore, it may be worthwhile to consider constructing paths that have four pieces.

Definition 5.1.9. For $a \in (0, \frac{1}{2})$ and $b, c, d \in \mathbb{R}$, we define the path $\gamma^{a,b,c,d}$ on $[s,t]$ as

$$\gamma^{a,b,c,d} := \gamma^{a,b,c} + d\widehat{\gamma}^a,$$

where the path $\gamma^{a,b,c}$ belongs to the three-parameter family and $\widehat{\gamma}^a$ is the tent function

$$\widehat{\gamma}^a := \begin{cases} 0 & \text{for } u \in [s, s+ah] \cup [t-ah, t] \\ \frac{2(u-(s+ah))}{(1-2a)h} & \text{for } u \in [s+ah, s+\frac{1}{2}h] \\ \frac{2((t-ah)-u)}{(1-2a)h} & \text{for } u \in [s+\frac{1}{2}h, t-ah] \end{cases}.$$

By extending lemma 5.1.1, we can derive formulae for iterated integrals of $\gamma^{a,b,c,d}$.

Theorem 5.1.10. *Let $\gamma^{a,b,c,d}$ be a piecewise linear path in the four-parameter family. The rescaled Lévy areas of the space-time process $\{(u, \gamma_u^{a,b,c,d})\}_{u\in[s,t]}$ can be defined as*

$$\widehat{H}_{s,t} := \frac{1}{h}\int_s^t \gamma_{s,u}^{a,b,c,d} - \frac{u-s}{h}\gamma_{s,t}^{a,b,c,d}\,du, \tag{5.1.31}$$

$$\widehat{K}_{s,t} := \frac{1}{h^2}\int_s^t \Big(\gamma_{s,u}^{a,b,c,d} - \frac{u-s}{h}\gamma_{s,t}^{a,b,c,d}\Big)\Big(\frac{1}{2}h - (u-s)\Big)\,du. \tag{5.1.32}$$

Then for any choice of $(a,b,c,d) \in \big(0,\frac{1}{2}\big)\times\mathbb{R}^3$ and any Brownian increment $W_{s,t} \in \mathbb{R}$,

$$\widehat{H}_{s,t} := \frac{1}{2}(1-a)\big(b+c-W_{s,t}\big) + \frac{1}{2}(1-2a)d, \tag{5.1.33}$$

$$\widehat{K}_{s,t} = \frac{1}{12}(1-a)\big(b-c+(1-2a)W_{s,t}\big), \tag{5.1.34}$$

and the following triple iterated integral of $\gamma^{a,b,c,d}$ and time can be computed explicitly:

$$\begin{aligned}\int_s^t \big(\gamma_{s,u}^{a,b,c,d}\big)^2\,du = \frac{1}{3}hW_{s,t}^2 + hW_{s,t}\widehat{H}_{s,t} - 2hW_{s,t}\widehat{K}_{s,t} + \frac{3-4a}{3(1-a)^2}h\widehat{H}_{s,t}^2 \\ + \frac{12}{(1-a)^2}h\widehat{K}_{s,t}^2 + \frac{a(1-2a)}{3(1-a)^2}hd\widehat{H}_{s,t} + \frac{1-2a}{12(1-a)^2}hd^2.\end{aligned} \tag{5.1.35}$$

Proof. It is straightforward to explicitly compute the following time integrals of $\widehat{\gamma}_{s,u}^a$.

$$\begin{aligned}\frac{1}{h}\int_s^t \widehat{\gamma}_{s,u}^a - \frac{u-s}{h}\widehat{\gamma}_{s,t}^a\,du &= \frac{1}{2}(1-2a),\\ \frac{1}{h^2}\int_s^t \Big(\widehat{\gamma}_{s,u}^a - \frac{u-s}{h}\widehat{\gamma}_{s,t}^a\Big)\Big(\frac{1}{2}h-(u-s)\Big)\,du &= 0.\end{aligned}$$

Hence, by lemma 5.1.1 and the linearity of integration, we have (5.1.33) and (5.1.34). Similarly, we can compute the final integral using lemma 5.1.1 along with the above.

$$\begin{aligned}\int_s^t \big(\gamma_{s,u}^{a,b,c,d}\big)^2\,du &= \int_s^t \big(\gamma_{s,u}^{a,b,c}\big)^2\,du + 2d\int_s^t \gamma_{s,u}^{a,b,c}\cdot\widehat{\gamma}_{s,u}^a\,du + d^2\int_s^t \big(\widehat{\gamma}_{s,u}^a\big)^2\,du\\ &= \frac{1}{3}hW_{s,t}^2 + hW_{s,t}\Big(\widehat{H}_{s,t} - \frac{1}{2}(1-2a)d\Big) - 2hW_{s,t}\widehat{K}_{s,t}\\ &\quad + \frac{3-4a}{3(1-a)^2}h\Big(\widehat{H}_{s,t} - \frac{1}{2}(1-2a)d\Big)^2 + \frac{12}{(1-a)^2}h\widehat{K}_{s,t}^2\\ &\quad + \frac{1}{2}hd(1-2a)\Big(W_{s,t} + \frac{2}{1-a}\widehat{H}_{s,t} - \frac{1-2a}{1-a}d\Big) + \frac{1}{3}(1-2a)hd^2.\end{aligned}$$

The result then follows by rearranging and simplifying the RHS of this equation. □

We can set $a = \frac{10-\sqrt{70}}{10}$ for the above path and find b, c, d to match the integrals:

$$\begin{aligned}
\int_s^t \gamma_{s,u}^{a,b,c,d}\, du &= \int_s^t W_{s,u}\, du, \\
\int_s^t (u-s)\gamma_{s,u}^{a,b,c,d}\, du &= \int_s^t (u-s) W_{s,u}\, du, \\
\int_s^t \left(\gamma_{s,u}^{a,b,c,d}\right)^2 du &= \mathbb{E}\left[\int_s^t W_{s,u}^2\, du \,\middle|\, W_{s,t}, H_{s,t}, K_{s,t} \right].
\end{aligned}$$

It is then clear from Theorem 5.1.10 that this yields a quadratic equation for d.

$$\frac{-10+4\sqrt{70}}{21} h H_{s,t}^2 + \frac{-8+\sqrt{70}}{7} h d H_{s,t} + \frac{-5+\sqrt{70}}{42} h d^2 = \frac{3}{70} h^2 + \frac{6}{5} h H_{s,t}^2.$$

Solving this equation produces two distinct real solutions, which are given below.

$$d = \frac{-10+\sqrt{70}}{5} H_{s,t} \pm \sqrt{\frac{490-28\sqrt{70}}{225} H_{s,t}^2 + \frac{5+\sqrt{70}}{25} h}.$$

Since the cubic polynomial approximation of the Brownian path is $\frac{1}{2} W_{s,t} + \frac{3}{2} H_{s,t}$ at its midpoint, we will choose the value of d to be a close as possible to $\left(\frac{3}{2} - \frac{1}{7}\sqrt{70}\right) H_{s,t}$. Therefore, we can now define the $(W_{s,t}, H_{s,t}, K_{s,t})$-measurable piecewise linear paths.

Definition 5.1.11. Using the triple $(W_{s,t}, H_{s,t}, K_{s,t})$, we have derived two piecewise linear paths that can each be used to approximate the non-commutative SDE (5.0.1).

- **Three pieces with guaranteed existence**

$$\begin{aligned}
a &= \frac{10-\sqrt{70}}{10}, \\
b + c &= W_{s,t} + \frac{2}{1-a} H_{s,t}, \\
b - c &= a W_{s,t} + \varepsilon \Bigg((1-a)^2 W_{s,t}^2 - 24 W_{s,t} K_{s,t} + \frac{2112-240\sqrt{70}}{105} H_{s,t}^2 \\
&\qquad + \frac{1440}{7} K_{s,t}^2 + \frac{18}{35} h \Bigg)^{\frac{1}{2}},
\end{aligned}$$

where in the above piecewise linear path, we define ε to minimize $|K_{s,t} - \widehat{K}_{s,t}|$,

$$\varepsilon := \begin{cases} 1 & \text{if } \ K_{s,t} \geq \frac{1}{12}(1-a)^2 W_{s,t} \\ -1 & \text{if } \ K_{s,t} < \frac{1}{12}(1-a)^2 W_{s,t} \end{cases}.$$

Then the resulting path $\gamma^{a,b,c}$ satisfies all the properties given by Theorem 5.1.4.

- **Four pieces with an independence property and guaranteed existence**

$$
\begin{aligned}
a &= \frac{10-\sqrt{70}}{10}, \\
b+c &= W_{s,t} + \frac{2}{1-a} H_{s,t} - \frac{1-2a}{1-a} d, \\
b-c &= \frac{12}{1-a} K_{s,t} - (1-2a) W_{s,t}, \\
d &= \frac{-10+\sqrt{70}}{5} H_{s,t} + \operatorname{sgn}(H_{s,t}) \sqrt{\frac{490-28\sqrt{70}}{225} H_{s,t}^2 + \frac{5+\sqrt{70}}{25} h}.
\end{aligned}
$$

The piecewise linear path $\gamma^{a,b,c,d}$ defined by these parameters has the integrals:

1. $\int_s^t \gamma_{s,u}^{a,b,c,d}\, du = \int_s^t W_{s,u}\, du$.
2. $\int_s^t \left(\gamma_{s,u}^{a,b,c,d}\right)^2 du = \mathbb{E}\left[\int_s^t W_{s,u}^2\, du \,\middle|\, W_{s,t}, H_{s,t}, K_{s,t} \right]$.
3. $\int_s^t (u-s)\gamma_{s,u}^{a,b,c,d}\, du = \int_s^t (u-s) W_{s,u}\, du$.

Remark 5.1.12. In addition, we have also derived a piecewise linear path with three pieces and the same independence property as the paths given by definition 5.1.6.

$$
\begin{aligned}
a &= \frac{8-\sqrt{10}}{18}, \\
b+c &= W_{s,t} + \frac{2}{1-a} H_{s,t}, \\
b-c &= aW_{s,t} + \varepsilon \sqrt{(1-a)^2 W_{s,t}^2 - 24 W_{s,t} K_{s,t} + \frac{1440}{7} K_{s,t}^2 + \frac{18}{35} h}\,,
\end{aligned}
$$

However, since there is a small probability (< 0.0022) that this path can not be constructed, it will require further analysis to justify its use for approximating SDEs.

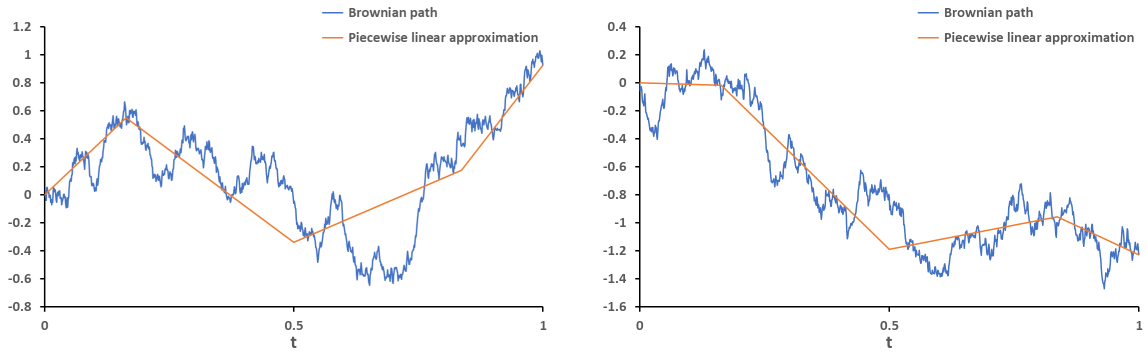

Figure 5.4: Brownian motion approximated with "high order" piecewise linear paths. The above paths are generated using three independent Gaussian random variables.

Therefore in this section, we have developed four different piecewise linear paths that can be used to discretize the non-commutative SDE (5.0.1) using ODE solvers.

These piecewise linear paths are explicitly described in definitions 5.1.6 and 5.1.11. We shall now conclude this section by giving the convergence rates of these methods.

Theorem 5.1.13. *For a fixed number of steps N, we can construct a numerical solution $\{Y_k\}_{0\leq k\leq N}$ of (5.0.1) by setting $Y_0 := \xi$ and for each $k \in [0 \,..\, N-1]$, defining Y_{k+1} to be the true solution at time $u = t_{k+1}$ of the following (random) ODE:*

$$\begin{aligned} dz_u &= f_0(z_u)\,du + f_1(z_u)\,d\gamma_u, \\ z_{t_k} &= Y_k, \end{aligned} \tag{5.1.36}$$

where $h := \frac{T}{N}$, $t_k = kh$ and γ is a piecewise linear path from definition 5.1.6 or 5.1.11. Then Y converges in a strong sense with order 1.5 and a weak sense with order 2.0.

Proof. Just as in the analysis of parabola-ODE (see Theorem 4.3.11), we can derive a (rough) Taylor expansion of (5.1.36) in terms of the iterated integrals of γ and time. By the construction of γ, the lower order terms in the expansion are matched exactly, the $O(h^2)$ terms have the correct mean and the expectation of the remainder is $O(h^3)$. Therefore, the result follows using the standard convergence arguments in [62]. □

Remark 5.1.14. In order to approximate the ODE (5.1.36) over the interval $[t_k, t_{k+1}]$, we must discretize a sequence of several ODEs (equal to the number of pieces γ has). Each intermediate ODE requires at least one step of a high order numerical method. In general, one should use fourth order ODE solvers, however for an SDE without the $W_{s,t}^4$ term in its Taylor expansion, it is enough to use a third order numerical method.

In Chapter 6, we will apply the $(W_{s,t}, H_{s,t}, n_{s,t})$-measurable piecewise linear path given by definition 5.1.6 to accurately discretize the CIR and squared Bessel processes. There are two key features of this path that make it well suited for both of these SDEs:

- Since $n_{s,t}$ is a Rademacher random variable, it is faster to generate than $K_{s,t}$. This is appealing as the vector fields governing these SDEs are cheap to evaluate.
- It is possible to generate $(W, H, n)_{s,u}$ and $(W, H, n)_{u,t}$ conditional on $(W, H, n)_{s,t}$. This property enables variable step size methods (and multilevel Monte Carlo), which is important as the CIR and Bessel SDEs have non-Lipschitz vector fields.

Whilst the paths presented in this section are designed for SDEs driven by a single Brownian motion, it is straightforward to extend them to the multidimensional case since the width of each "linear piece" does not depend on the Brownian path itself. That said, it seems unlikely that this captures the mean and variance of Lévy area. Hence it is still ongoing research to construct piecewise linear paths for general SDEs.

5.2 Piecewise linear methods for SDEs satisfying a commutativity condition

In this section, we shall consider two piecewise linear paths for discretizing the SDE (5.0.1) when the governing vector fields satisfy the following commutativity condition:

$$[f_1, [f_1, f_0]] = 0. \tag{5.2.1}$$

Due to the above condition, it is no longer necessary for the path to capture the space-space-time Lévy area $L_{s,t}$ as it does not appear in Taylor expansion of (5.0.1). Therefore, we can now reconsider the paths that belong to the one-parameter family.

Theorem 5.2.1. *The path $\gamma^{\frac{1}{2}}$ is the only path with two pieces and the below integrals.*

$$\int_s^t \gamma^{\frac{1}{2}}_{s,u}\, du = \int_s^t W_{s,u}\, du,$$

$$\int_s^t \left(\gamma^{\frac{1}{2}}_{s,u}\right)^2 du = \frac{1}{3} h W_{s,t}^2 + h W_{s,t} H_{s,t} + \frac{4}{3} h H_{s,t}^2,$$

$$\int_s^t (u-s)\gamma^{\frac{1}{2}}_{s,u}\, du = \mathbb{E}\left[\int_s^t (u-s) W_{s,u}\, du \,\middle|\, W_{s,t}, H_{s,t} \right].$$

Proof. The result follows directly from the calculations presented in Appendix A. □

Recall that the disadvantage of using this path for discretizing non-commutative SDEs is that it approximates the Lévy area $L_{s,t}$ with $\frac{4}{3} h H_{s,t}^2$ (which is then unbiased). If the SDE (5.0.1) also has additive noise, then it is worthwhile to generate $K_{s,t}$. Hence, the piecewise linear path defined below will match the triple $(W_{s,t}, H_{s,t}, K_{s,t})$.

Definition 5.2.2. Consider the unique three-parameter path $\gamma^{a,b,c}$ on $[s,t]$ satisfying

$$a = \frac{1}{2} - \frac{1}{2\sqrt{5}},$$

$$b + c = W_{s,t} + \frac{2}{1-a} H_{s,t},$$

$$b - c = \frac{12}{1-a} K_{s,t} - (1-2a) W_{s,t}.$$

Then, provided (5.2.1) holds, driving (5.0.1) by $\gamma^{a,b,c}$ gives a high order approximation.

Remark 5.2.3. It is straightforward to extend the above paths to the multidimensional setting by viewing $W_{s,t}$, $H_{s,t}$ and $K_{s,t}$ as vectors in $\mathbb{R}^d$ with independent coordinates. Whilst it is still ongoing research to study piecewise linear methods for general SDEs, we shall see that the multidimensional version of path given by definition 5.2.2 is particularly well suited for numerically solving the (underdamped) Langevin equation.

Theorem 5.2.4. *The piecewise linear path $\gamma^{a,b,c}$ given by definition 5.2.2 has integrals*

$$\begin{aligned}
\int_s^t \gamma_{s,u}^{a,b,c}\,du &= \int_s^t W_{s,u}\,du,\\
\int_s^t \big(\gamma_{s,u}^{a,b,c}\big)^2\,du &= \frac{1}{3}hW_{s,t}^2 + hW_{s,t}H_{s,t} + \widehat{L}_{s,t},\\
\int_s^t (u-s)\gamma_{s,u}^{a,b,c}\,du &= \int_s^t (u-s)W_{s,u}\,du,\\
\int_s^t (u-s)^2\gamma_{s,u}^{a,b,c}\,du &= \mathbb{E}\bigg[\int_s^t (u-s)^2 W_{s,u}\,du \,\bigg|\, W_{s,t}, H_{s,t}, K_{s,t}\bigg],
\end{aligned}$$

where

$$\widehat{L}_{s,t} := \frac{5+\sqrt{5}}{6}hH_{s,t}^2 - 2hW_{s,t}K_{s,t} + \big(90-30\sqrt{5}\big)hK_{s,t}^2.$$

Proof. The first two integrals can be shown using (5.1.3) and (5.1.4) in lemma 5.1.1. In a similar manner, the formula for the third integral directly follows from (5.1.5). In order to derive the last integral, we will first compute the conditional expectation:

$$\begin{aligned}
&\mathbb{E}\bigg[\int_s^t (u-s)^2 W_{s,u}\,du \,\bigg|\, W_{s,t}, H_{s,t}, K_{s,t}\bigg]\\
&\quad= \int_s^t (u-s)^2\,\mathbb{E}[W_{s,u}\,|\,W_{s,t}, H_{s,t}, K_{s,t}]\,du\\
&\quad= h^3\int_0^1 u^2\big(uW_{s,t} + 6u(1-u)H_{s,t} + 60u(1-u)(1-2u)K_{s,t}\big)\,du\\
&\quad= \frac{1}{4}h^3W_{s,t} + \frac{3}{10}h^3H_{s,t} - h^3K_{s,t},
\end{aligned}$$

where the penultimate line uses Theorem 4.1.6 as well as integration by substitution. The corresponding time integral for $\gamma^{a,b,c}$ can also be resolved by a direct calculation.

$$\begin{aligned}
\int_s^t (u-s)^2\gamma_{s,u}^{a,b,c}\,du &= \frac{1}{4}a^3bh^3 + \frac{1}{12}(1-2a)\big((2a^2+1)b + (2a^2-4a+3)c\big)h^3\\
&\qquad + \frac{1}{12}a\big((3a^2-8a+6)c + (a^2-4a+6)W_{s,t}\big)h^3\\
&= \frac{1}{12}(1-a)(a^2-a+2)(b+c)h^3 - \frac{1}{12}(1-a)(b-c)h^3\\
&\qquad + \frac{1}{12}a(a^2-4a+6)h^3W_{s,t}\\
&= \frac{1}{4}h^3W_{s,t} + \frac{1}{6}(a^2-a+2)h^3H_{s,t} - h^3K_{s,t}.
\end{aligned}$$

Therefore, the integrals for $\gamma^{a,b,c}$ and W will coincide if and only if $a = \frac{1}{2} \pm \frac{1}{2\sqrt{5}}$. □

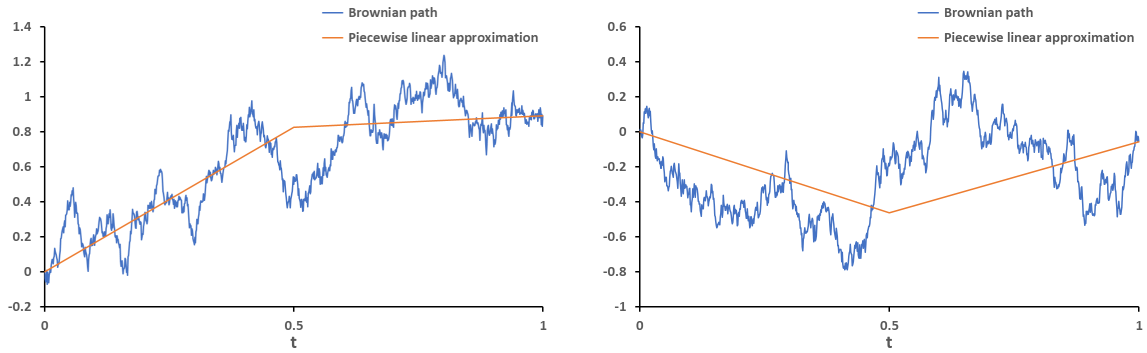

Figure 5.5: Brownian motion approximated with the simpler piecewise linear paths. The above paths were generated using two independent Gaussian random variables.

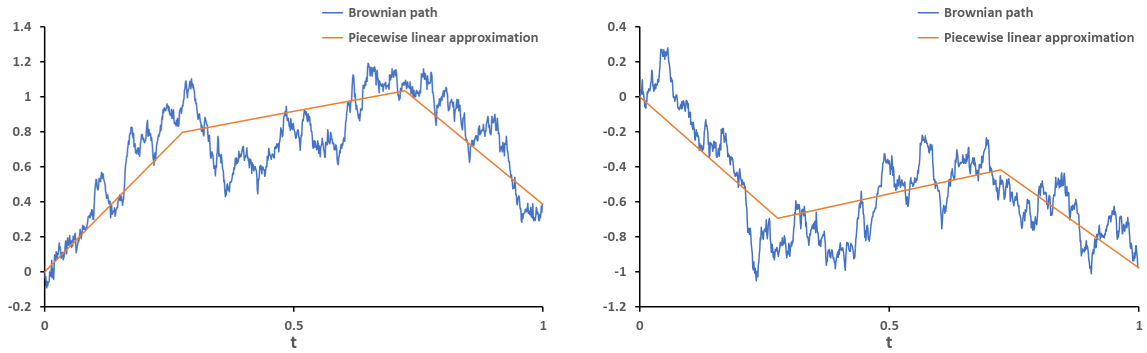

Figure 5.6: Brownian motion along with the associated linear three-parameter paths. The above paths are generated using three independent Gaussian random variables.

To conclude the section, we shall present a convergence theorem for these paths.

Theorem 5.2.5. *For a fixed number of steps $N \geq 1$, we construct a numerical solution $\{Y_k\}_{0\leq k\leq N}$ of (5.0.1) by setting $Y_0 := \xi$ and for each $k \in [0 \,..\, N-1]$, defining Y_{k+1} to be the true solution at time $u = t_{k+1}$ of the following (random) ODE:*

$$dz_u = f_0(z_u)\,du + f_1(z_u)\,d\gamma_u, \tag{5.2.2}$$

$$z_{t_k} = Y_k,$$

where $h := \frac{T}{N}$, $t_k = kh$ and γ is a piecewise linear path considered in this section. Then provided (5.2.1) holds, Y converges in a strong and weak sense with order 2.0.

Proof. The result follows using essentially the same analysis as for Theorem 5.1.13. The key difference is that, due to the vectors fields commuting according to (5.2.1), space-space-time Lévy area no longer appears in the expansions of (5.0.1) and (5.2.2). This means that the Taylor expansions now coincide up to the terms of size $O(h^2)$. Since the expectation of the $O(h^{\frac{5}{2}})$ remainder is still $O(h^3)$, we can use the standard arguments presented in [62] to establish the strong and weak convergence rates. □

5.3 Third order numerical methods for the underdamped Langevin equation

In this section, we shall apply the piecewise linear path given by definition 5.2.2 with a Runge-Kutta-Nyström method to discretize the (underdamped) Langevin equation:

$$dQ_t = P_t\,dt, \tag{5.3.1}$$

$$dP_t = -\nabla U(Q_t)\,dt - \nu P_t\,dt + \sqrt{\frac{2\nu}{\beta}}\,dW_t, \tag{5.3.2}$$

where $Q, P \in \mathbb{R}^d$ will represent the position and momentum of a point particle, $U : \mathbb{R}^d \to \mathbb{R}$ denotes a scalar potential, $\nu > 0$ is a friction coefficient, $\beta = \left(k_B T\right)^{-1}$ is the inverse temperature of the system (k_B denotes Boltzmann's constant, T is the system's absolute temperature) and W is a standard d-dimensional Brownian motion.

The underdamped Langevin equation can be viewed as an extension of Newton's second law which takes additional frictional and stochastic forces into consideration. Since its conception, it has become well established within statistical mechanics as the fundamental model of a Brownian particle that exhibits noise and moves under the influence of an external force ([22] is a comprehensive textbook on this equation).

In the numerical analysis of the underdamped Langevin diffusion (ULD), one typically assumes that the potential $U : \mathbb{R}^d \to \mathbb{R}$ is twice continuously differentiable, m-strongly convex and has an M-Lipschitz gradient. That is, for $x, y \in \mathbb{R}^d$, we have

$$\left\|\nabla U(x) - \nabla U(y)\right\| \leq M\|x - y\|, \tag{5.3.3}$$

and

$$U(y) \geq U(x) + \left\langle \nabla U(x), y - x \right\rangle + \frac{1}{2} m\|x - y\|^2. \tag{5.3.4}$$

We refer to the conditions (5.3.3) and (5.3.4) as the "standard assumptions" for ULD.

More recently, the underdamped Langevin equation has been applied to problems in Bayesian statistics as it is intimately related to Hamiltonian Monte Carlo (HMC). Whilst HMC was originally proposed in 1987 by Duane et al. [32], it was the 2011 papers of Neil [80] along with Girolami and Calderhead [46] that spurred the recent developments of the approach. In addition, underdamped Langevin Monte Carlo has also progressed due to a non-asymptotic convergence result by Dalalyan [25] in 2017. In these statistical problems, U is obtained as the logarithm of the target density and for many real-world applications, it will satisfy the standard assumptions for ULD. For example, in Logistic regression, U is strongly convex and infinity differentiable (with each derivative being Lipschitz continuous).

To the best of the authors knowledge, the current state-of-the-art methods for underdamped Langevin MCMC are the "Randomized Midpoint Method" [85] and "OBABO Scheme" [14, 78, 86] (which is a stochastic version of the symplectic velocity Verlet algorithm). The convergence rates for these methods are summarized below:

- The randomized midpoint method achieves an $O\big(\sqrt{d}h^{1.5}\big)$ 2-Wasserstein error under standard assumptions. This rate was later shown to be optimal in [16].
- The OBABO scheme achieves an $O\big(\sqrt{d}h\big)$ Wasserstein error under standard assumptions. In addition, if the Hessian $\nabla^2 U$ is Lipschitz continuous, then the OBABO scheme achieves an $O\big(dh^2\big)$ Wasserstein error. It is also worth noting that a Metropolis-adjusted version of the OBABO scheme is given in [78, 86].

In addition, Strang splittings are known to perform well on Hamiltonian SDEs. In [12], the authors propose an ABC algorithm (Approximate Bayesian Computation) for a stochastic neural mass model and show that splitting schemes will significantly outperform the Euler-Maruyama method at capturing the SDE's stationary measure.

The underdamped Langevin diffusion has also been studied when the gradient ∇U is superlinear but satisfies certain growth conditions (such as Condition 3.1 in [75]). For example, Theorem 3.2 in [75] guarantees the existence of the stationary measure for ULD when U has polynomial growth. In this setting, ∇U is only locally Lipschitz and the Euler-Maruyama method may not produce an ergodic numerical solution. Hence to address this problem, researchers have proposed implicit numerical methods, such as the split-step backward Euler method [75] or variable step size methods [64].

Since the discretizations of (5.3.1) presented in this section were discovered during the writing of this thesis, it is still ongoing research to prove existence and then establish convergence results for the stationary measure of the numerical solution (which is needed to quantify the method's effectiveness for statistical applications).

Therefore, the primary focus of this section will instead be to demonstrate the high orders of convergence of the proposed numerical methods over a finite time horizon. In particular, the piecewise linear approaches for (5.3.1) described in this section can achieve convergence rates of $O(h^3)$ provided each step uses a suitable ODE solver. That said, if ∇U is only locally Lipschitz and with polynomial growth, we would not expect convergence for our methodology without the use of either variable step sizes or implicit ODE solvers. This is likely to be a topic for future research. Another advantage of using ODE methods is that they allow one to employ symplectic integrators, which can achieve state-of-the-art performance on Hamiltonian systems.

To simplify the notation, we shall rewrite the underdamped Langevin equation as

$$dQ_t = P_t\,dt, \tag{5.3.5}$$

$$dP_t = f(Q_t)\,dt - \nu P_t\,dt + \sigma\,dW_t, \tag{5.3.6}$$

where $f := -\nabla U$ is a smooth bounded vector field on $\mathbb{R}^d$ with bounded derivatives and $\sigma := \sqrt{\frac{2\nu}{\beta}}$. We can expand the above in terms of iterated time integrals of W.

Theorem 5.3.1 (Stochastic Taylor expansion for underdamped Langevin equation)**.** *Let (Q, P) denote the solution to (5.3.5) over $[s, t]$. Then (Q, P) can be expanded as*

$$\begin{aligned} Q_t = Q_s + P_s h + \frac{1}{2}\big(f(Q_s) - \nu P_s\big)h^2 + \frac{1}{6}\big(\nu^2 P_s + f'(Q_s)P_s - \nu f(Q_s)\big)h^3 \\ + \sigma\int_s^t W_{s,u}\,du - \sigma\nu\int_s^t\int_s^u W_{s,v}\,dv\,du + \sigma\nu^2\int_s^t\int_s^u\int_s^v W_{s,r}\,dr\,dv\,du \\ + \sigma f'(Q_s)\int_s^t\int_s^u\int_s^v W_{s,r}\,dr\,dv\,du + R^Q(h,(Q_s,P_s)), \end{aligned} \tag{5.3.7}$$

$$\begin{aligned} P_t = P_s + \big(f(Q_s) - \nu P_s\big)h + \sigma W_{s,t} \\ - \sigma\nu\int_s^t W_{s,u}\,du + \frac{1}{2}\big(\nu^2 P_s + f'(Q_s)P_s - \nu f(Q_s)\big)h^2 \\ + \frac{1}{6}\Big(\big(f'(Q_s) + \nu^2\big)\big(f(Q_s) - \nu P_s\big) - \nu f'(Q_s)P_s + f''(Q_s)P_s^{\otimes 2}\Big)h^3 \\ + \big(f'(Q_s) + \nu^2\big)\bigg(\sigma\int_s^t\int_s^u W_{s,v}\,dv\,du - \sigma\nu\int_s^t\int_s^u\int_s^v W_{s,r}\,dr\,dv\,du\bigg) \\ - \sigma\nu f'(Q_s)\int_s^t\int_s^u\int_s^v W_{s,r}\,dr\,dv\,du + \sigma f''(Q_s)\bigg(P_s\otimes\int_s^t\int_s^u\int_s^v W_{s,r}\,dr\,dv\,du\bigg) \\ + \sigma f''(Q_s)\bigg(P_s\otimes\int_s^t\int_s^u\int_s^v (r-s)\,dW_r\,dv\,du\bigg) + R^P(h,(Q_s,P_s)), \end{aligned} \tag{5.3.8}$$

where $h := t - s$ and the remainder terms have the below uniform estimates for $h < 1$,

$$\sup_{Q_s,P_s\in\mathbb{R}^d}\big\|R^Q(h,(Q_s,P_s))\big\|_{L^2(\mathbb{P})} \le C_1\,h^4, \tag{5.3.9}$$

$$\sup_{Q_s,P_s\in\mathbb{R}^d}\big\|R^P(h,(Q_s,P_s))\big\|_{L^2(\mathbb{P})} \le C_2\,h^4, \tag{5.3.10}$$

and the constants C_1, C_2 depend only on the vector fields of the differential equation.

Proof. The result follows by applying the standard Stratonovich-Taylor expansion (see Proposition 1.1 in [6] or Theorem 5.6.1 in [62]) to the Langevin equation (5.3.5). Although the various coefficients within the expansion can be computed individually, it is possibly more straightforward to derive the entire Taylor expansion by hand. □

Perhaps the most significant feature of the Taylor expansions (5.3.7) and (5.3.8) is that lower order terms only depend on "linear" iterated integrals of Brownian motion. As a result, it is then clear from Theorem 5.2.4 that the piecewise linear path given by definition 5.2.2 is particularly well suited to the underdamped Langevin equation. Since the path $\gamma^{a,b,c}$ will capture these time integrals, we have the following theorem:

Theorem 5.3.2. *For a fixed number of steps N, we construct a numerical solution $\{Y_k\}_{0\leq k\leq N}$ of the SDE (5.3.1) by setting $Y_0 := (Q_0, P_0)$ and for each $k \in [0\,..\,N-1]$, defining Y_{k+1} to be the true solution at time $u = t_{k+1}$ of the following (random) ODE:*

$$dx_u = z_u\, du, \tag{5.3.11}$$

$$dz_u = -\nabla U(x_u)\, du - \nu z_u\, du + \sqrt{\frac{2\nu}{\beta}}\, d\gamma_u, \tag{5.3.12}$$

$$(x_{t_k}, z_{t_k}) = Y_k,$$

where $h := \frac{T}{N}$, $t_k = kh$ and γ is the piecewise linear path given by definition 5.2.2. Then Y_N converges to the solution (P_T, Q_T) in strong and weak senses with order 3.0.

Proof. As we have already noted, the path γ was constructed so that it matches the iterated time integrals of Brownian motion that are Gaussian and larger than $O(h^3)$. So the Taylor expansions of (5.3.11) and (5.3.12) will match those in Theorem 5.3.1 up to and including the $O(h^3)$ terms and therefore give local errors that are $O(h^{\frac{7}{2}})$. By linearity of expectation, we see that the iterated integrals below have mean zero:

$$\int_s^t\int_s^u\int_s^v W_{s,r}\, dr\, dv\, du, \quad \int_s^t\int_s^u\int_s^v (r-s)\, dW_r\, dv\, du,$$
$$\int_s^t\int_s^u\int_s^v \gamma_{s,r}\, dr\, dv\, du, \quad \int_s^t\int_s^u\int_s^v (r-s)\, d\gamma_r\, dv\, du.$$

Hence, the expectation of the local error between the two Taylor expansions is $O(h^4)$. The result now follows using the standard argument (see Theorem 10.7.1 in [62]). □

Remark 5.3.3. Since the odd moments of a standard multivariate Gaussian random variable are all zero, the $O(h^{\frac{7}{2}})$ terms in Taylor expansions always have zero mean. Therefore it is sufficient to discretize (5.3.11) and (5.3.12) using three steps of a third order numerical method to achieve the strong and weak convergence rates of $O(h^3)$.

Remark 5.3.4. From the Taylor expansions (5.3.7) and (5.3.8) for ULD, we would conjecture that the above theorem holds when U is three times differentiable with each derivative being Lipschitz continuous. In particular, we do not expect that the boundedness assumptions on U and ∇U are needed to obtain the local error estimates. That said, the analysis required is likely to be involved and is thus left as future work.

Since the piecewise linear path γ given by definition 5.2.2 has three pieces, the corresponding SDE numerical method requires one to discretize three ODEs per step. It is therefore reasonable to question whether there exist cheaper ODE methods that can achieve a third order convergence rate for the underdamped Langevin diffusion. We conjecture that by setting $a = 0$ in definition 5.2.2, we can obtain such a method. That is, we consider a piecewise linear path with discontinuities (or "vertical pieces").

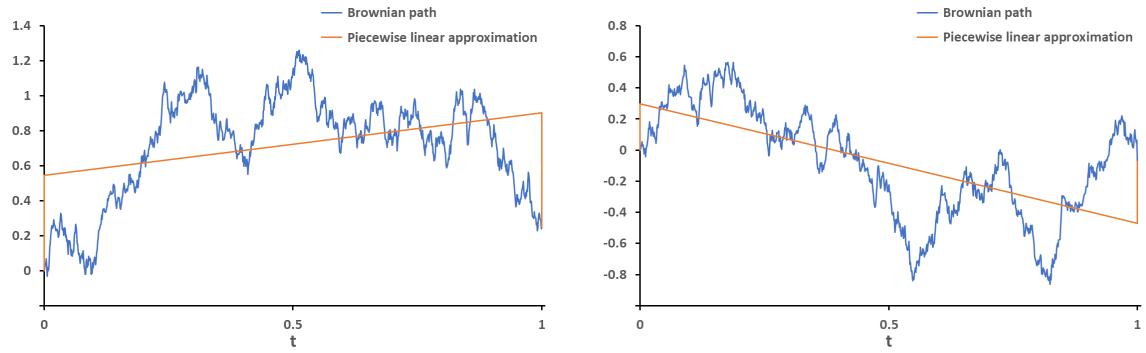

Figure 5.7: Brownian motion approximated using paths defined with vertical pieces. The above paths are generated using three independent Gaussian random variables.

Since the Langevin equation has additive noise, each vertical pieces corresponds to a trivial ODE (with a constant vector field). Hence, the resulting numerical method only requires one ODE to be approximated per step. We define this method below:

Definition 5.3.5 (A shifted ODE method for the underdamped Langevin diffusion)**.** For a fixed number of steps N, we construct a numerical solution $\{Y_k\}_{0\leq k\leq N}$ of the SDE (5.3.1) by setting $Y_0 := (Q_0, P_0)$ and for each $k \in [0\,..\,N-1]$, defining Y_{k+1} as

$$Y_{k+1} := \begin{pmatrix} x_{t_{k+1}} \\ z_{t_{k+1}} \end{pmatrix} - \sqrt{\frac{2\nu}{\beta}} \begin{pmatrix} 0 \\ H_{t_k,t_{k+1}} - 6K_{t_k,t_{k+1}} \end{pmatrix}, \tag{5.3.13}$$

where $\{(x, z)\}_{u\in[t_k,t_{k+1}]}$ is the solution to the following (random) ODE:

$$\frac{dx}{du} = z, \tag{5.3.14}$$

$$\frac{dz}{du} = -\nabla U(x) - \nu z + \sqrt{\frac{2\nu}{\beta}} \left(\frac{W_{t_k,t_{k+1}} - 12K_{t_k,t_{k+1}}}{h} \right), \tag{5.3.15}$$

with initial condition

$$\begin{pmatrix} x_{t_k} \\ z_{t_k} \end{pmatrix} := Y_k + \sqrt{\frac{2\nu}{\beta}} \begin{pmatrix} 0 \\ H_{t_k,t_{k+1}} + 6K_{t_k,t_{k+1}} \end{pmatrix}.$$

Remark 5.3.6. By setting $K_{t_k,t_{k+1}} = 0$ and applying a change of variable to the velocity, we see that the above reduces to the log-ODE method (see definition 4.3.4).

For both the "piecewise linear" and "shifted" ODE methods, it is necessary to discretize ODEs of the form (5.3.11), (5.3.12) using a high order numerical method. However, before doing so, it is worth noting that each "linear" ODE is expressible as

$$\frac{d^2x}{du^2} = -\nabla U(x) - \nu\frac{dx}{du} + \sqrt{\frac{2\nu}{\beta}}\frac{d\gamma}{du}. \tag{5.3.16}$$

As this ODE is second order, it is natural to consider Runge-Kutta-Nyström methods:

Definition 5.3.7 (Runge-Kutta-Nyström method)**.** Consider the second order ODE

$$\begin{aligned} y'' &= F(t,y), \\ y(0) &= y_0, \quad y'(0) = y_0', \end{aligned} \tag{5.3.17}$$

where $y_0, y_0' \in \mathbb{R}^d$ and the time-varying vector field $F : \mathbb{R} \times \mathbb{R}^d \to \mathbb{R}^d$ does not use y'. An explicit m-stage Runge-Kutta-Nyström (RKN) method for (5.3.17) is defined as

$$\begin{aligned} Y_{k+1} &:= Y_k + hY_k' + h^2\sum_{i=1}^{m} b_i F(t_k + c_i h, \tilde{y}_{k,i}), \\ Y_{k+1}' &:= Y_k' + h\sum_{i=1}^{m} d_i F(t_k + c_i h, \tilde{y}_{k,i}), \\ \tilde{y}_{k,i} &:= Y_k + c_i hY_k' + h^2\sum_{j=1}^{i-1} a_{ij} F(t_k + c_i kh, \tilde{y}_{k,j}), \end{aligned}$$

where $\left(Y_0, Y_0'\right) := \left(y_0, y_0'\right)$, $t_k := kh$ and the coefficients a_{ij}, b_i, c_i, d_i are real numbers. It is usually more convenient to express the above method using the Butcher tableau:

$$\begin{array}{c|cccc} c_1 & & & & \\ c_2 & a_{21} & & & \\ \vdots & \vdots & \ddots & & \\ c_m & a_{m1} & \cdots & a_{m(m-1)} & \\ \hline & b_1 & \cdots & b_{m-1} & b_m \\ & d_1 & \cdots & d_{m-1} & d_m \end{array}$$

Although the theory of Runge-Kutta-Nyström methods is separate from that of standard Runge-Kutta methods, in both cases it is essential to derive order conditions. In [91], order conditions were derived using the Mathematicia programming Language. A detailed reference for these methods is also given by section 14 in Chapter 2 of [51].

However in order to apply a Runge-Kutta-Nyström method, we require the ODE to have the form (5.3.17). Therefore, we shall apply the following change of variable:

$$\tilde{x}_u := e^{\frac{1}{2}\nu(u-t_k)} x_u, \qquad \forall u \in [t_k, t_{k+1}],$$

to remove the $\dfrac{dx}{du}$ term in the RHS of (5.3.16). Thus, we rewrite the ODE (5.3.11) as

$$\frac{d^2\tilde{x}}{du^2} = \frac{1}{4}\nu^2 \tilde{x}_u - e^{\frac{1}{2}\nu(u-t_k)} \nabla U\big(e^{-\frac{1}{2}\nu(u-t_k)}\tilde{x}_u\big) + e^{\frac{1}{2}\nu(u-t_k)} \sqrt{\frac{2\nu}{\beta}} \frac{d\gamma}{du}, \tag{5.3.18}$$

$$\tilde{x}_{t_k} = x_{t_k}, \quad \tilde{x}'_{t_k} = \frac{1}{2}\nu x_{t_k} + z_{t_k}.$$

Since the time-varying vector fields in (5.3.18) depend only on $\tilde{x}$ and not its derivative, we can discretize the above ODE using any high order Runge-Kutta-Nyström method.

In the numerical example, we will use the three-stage RKN method described in [28] as it is third order as well as symplectic (when applied to a Hamiltonian system).

Definition 5.3.8 (Third order three-stage symplectic Runge-Kutta-Nyström method)**.** This RKN method is presented in [28] and defined by the following Butcher tableau:

0			
0.630847693	0.164217030		
0.536704894	0.139710559	−0.103005664	
	0.260311692	0.4039053382	−0.164217030
	0.260311692	1.094142798	−0.354454490

Of course, the analysis of numerical methods for ODEs is a well developed research area and thus we expect that our approach to discretizing (5.3.11) can be improved as we would ideally like the ODE solver to be specifically designed for this problem. One possibility (that is left as future work) is to use a high order splitting method. To apply a splitting method, we write the ODE's vector field as a sum of operators:

$$d\begin{pmatrix} x_t \\ z_t \end{pmatrix} = \underbrace{\begin{pmatrix} z_t \\ 0 \end{pmatrix}}_{=:A} + \underbrace{\begin{pmatrix} 0 \\ -\nabla U(x_t) - \nu z_t + \sqrt{\frac{2\nu}{\beta}}\frac{d\gamma}{du} \end{pmatrix}}_{=:B}.$$

In this case, the ODEs governed by A and B have the following closed-form solutions:

$$\varphi_t^A \begin{pmatrix} x \\ z \end{pmatrix} := \begin{pmatrix} x + zt \\ z \end{pmatrix},$$

$$\varphi_t^B \begin{pmatrix} x \\ z \end{pmatrix} := \begin{pmatrix} x \\ e^{-\nu t} z + \Big(-\nabla U(x) + \sqrt{\frac{2\nu}{\beta}}\frac{d\gamma}{du}\Big)\Big(\frac{1-e^{-\nu t}}{\nu}\Big) \end{pmatrix}.$$

For example, the fourth order splitting detailed in [98] gives the approximation

$$\begin{pmatrix} x_{t_{k+1}} \\ v_{t_{k+1}} \end{pmatrix} := \varphi^B_{(\frac{1}{2}+\phi)h} \circ \varphi^A_{(1+2\phi)h} \circ \varphi^B_{-\phi h} \circ \varphi^A_{(-1-4\phi)h} \circ \varphi^B_{-\phi h} \circ \varphi^A_{(1+2\phi)h} \circ \varphi^B_{(\frac{1}{2}+\phi)h} \begin{pmatrix} x_{t_k} \\ v_{t_k} \end{pmatrix},$$

where $\phi := \frac{-1+\sqrt[3]{2}}{2(2-\sqrt[3]{2})}$. Since $\nabla U(x_{t_k})$ is computed in the first operation and $\nabla U(x_{t_{k+1}})$ is computed in the last, we see that this method only requires 3 additional evaluations of ∇U per step (with the only exception being the initial step – which requires four). In addition, when $\nu = 0$, this fourth order splitting method will become symplectic.

For the numerical experiment, we will consider a scalar double-well potential along with the same friction and temperature parameters used in the first example of [93]:

$$U(q) = (q^2 - 1)^2, \quad \nu = 1, \quad \beta = 3, \quad \big(Q_0, P_0\big) = 0.$$

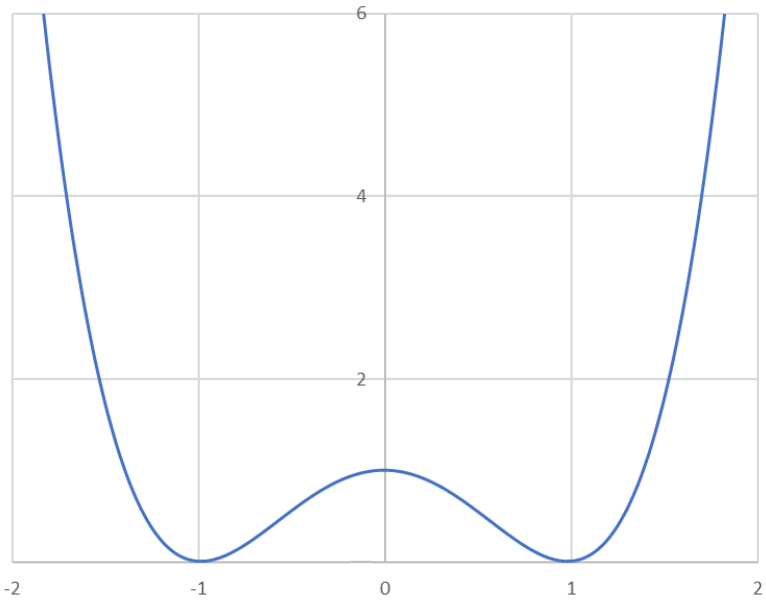

Figure 5.8: The scalar double-well potential given by the polynomial $U(q) = (q^2-1)^2$.

Although this is a simple example, the potential U does not satisfy the standard assumptions for ULD. Moreover, U and its derivatives are not even globally Lipschitz. Thus, it may be possible that the proposed ODE method does not converge to the SDE solution. At the very least, further analysis would be required to show otherwise. A double-well potential example is also included in [64], where the authors compared adaptive Euler and sympelctic schemes (both of which appeared to converge weakly). In addition, it is possible for implicit schemes, such as the split-step backward Euler method in [75], to converge for this example as U satisfies a one-sided Lipschitz condition. Hence we conjecture that the piecewise linear and shifted ODE methods can converge provided they are either adaptive or use an implicit solver for each ODE.

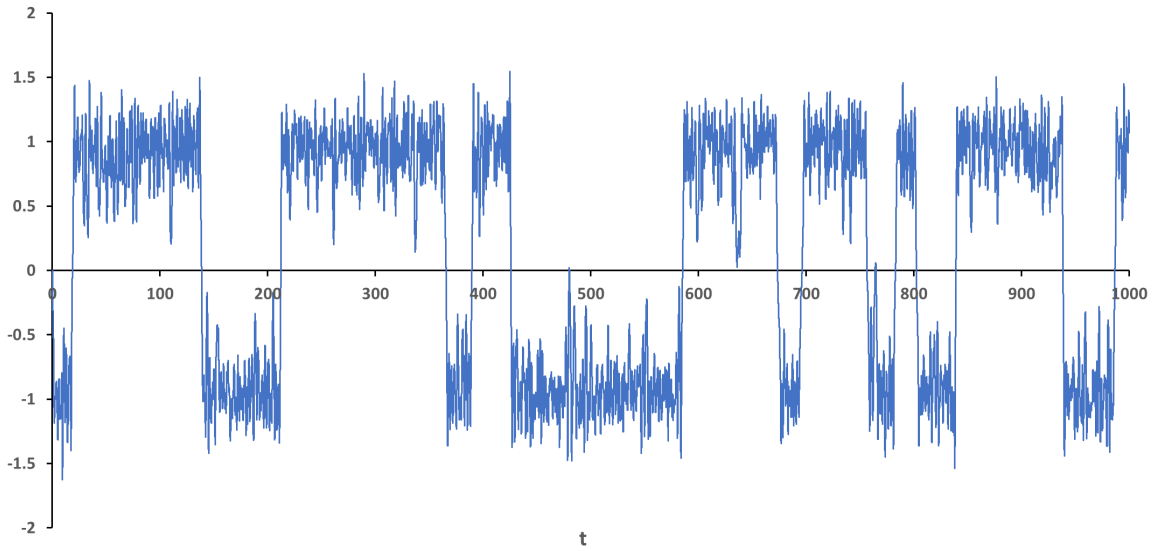

Figure 5.9: A sample path of a Brownian particle moving in a double-well potential. The path was computed by applying the RKN method in definition 5.3.8 to (5.3.18).

Due to the symmetry of the double-well potential (and the initial conditions), the particle is expected to spend the same amount of time above and below zero. Furthermore, the sign of the Brownian particle will change less frequently when the friction coefficient ν is decreased or the inverse temperature parameter β is increased.

In the numerical experiment, we will approximate (5.3.1) at a fixed time using the piecewise linear and shifted ODE methods on a uniform partition with step size h. We shall use similar error estimators as those presented in the previous chapter to demonstrate the strong and weak convergence properties of the numerical method. The functional chosen for testing weak convergence is q^2 and estimates the variance.

Definition 5.3.9 (Strong and weak error estimators)**.** For each $N \geq 1$, let Q_N denote a numerical solution of (5.3.1) computed at time T using the fixed step size $h = \frac{T}{N}$. We can define the following estimators for quantifying strong and weak convergence:

$$S_N := \sqrt{\mathbb{E}\left[\left(Q_N - Q_T^{\text{fine}}\right)^2\right]}, \tag{5.3.19}$$

$$E_N := \left|\mathbb{E}\left[Q_N^2\right] - \mathbb{E}\left[\left(Q_T^{\text{fine}}\right)^2\right]\right|, \tag{5.3.20}$$

where the above expectations are approximated by Monte-Carlo simulation and Q_T^{fine} is the numerical solution of (5.3.1) obtained at time T using the piecewise linear ODE method with a "fine" step size of $\min\left(\frac{h}{25}, \frac{T}{50000}\right)$. The fine step size is chosen so that the $L^2(\mathbb{P})$ error between Q_T^{fine} and the true solution is negligible compared to S_N. Note that Q_N and Q_T^{fine} are both computed with respect to the same Brownian paths.

Below are our results for this numerical experiment with the time horizon $T = 10$.

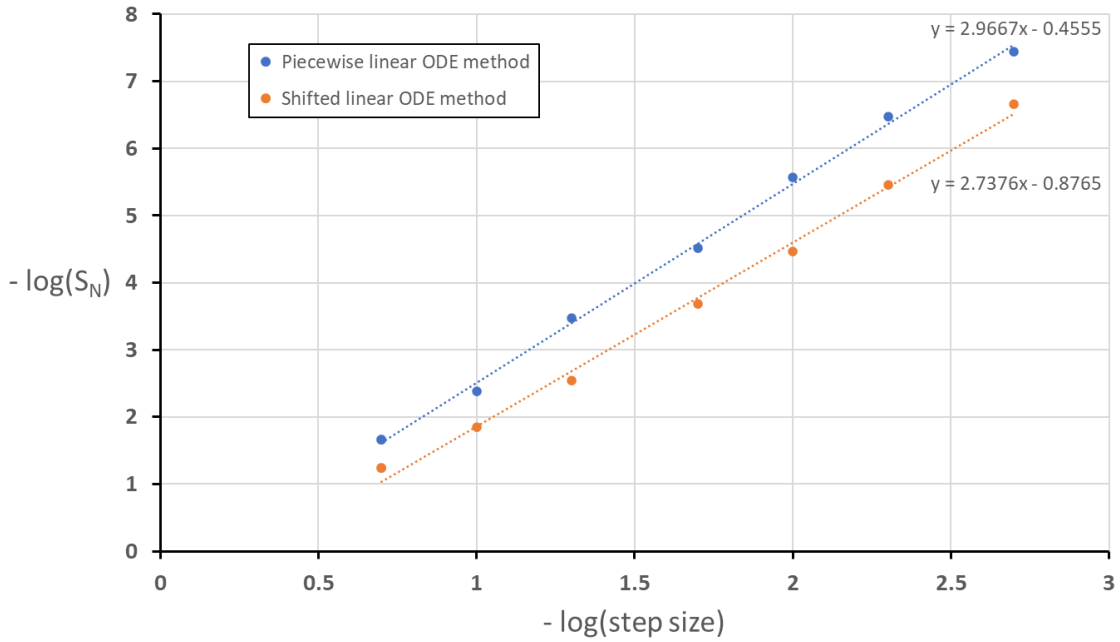

Figure 5.10: S_N computed with 100,000 sample paths as a function of step size $h = \frac{T}{N}$.

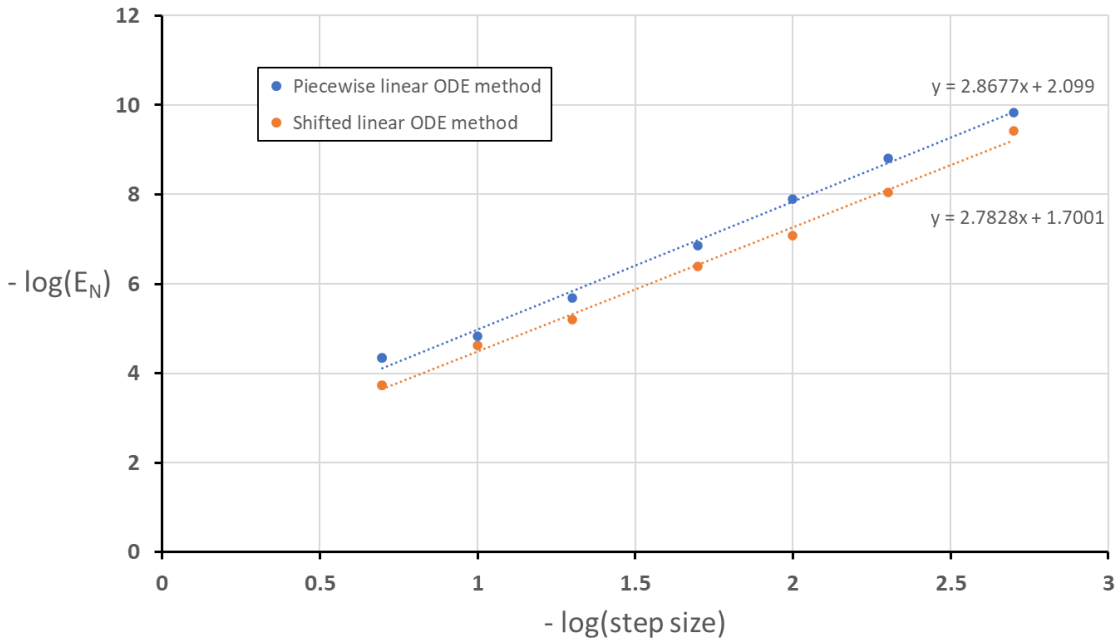

Figure 5.11: E_N computed with 100,000 sample paths as a function of step size $h = \frac{T}{N}$.

Although there is additional error due to the Monte Carlo estimation of S_N and E_N, these graphs support the hypothesis that the methods converge with a rate of $O(h^3)$.

In conclusion, there are two results that have been discussed within this chapter.

- **High order pathwise discretizations of space-time Brownian motion**

  One classical technique for approximating SDEs is to discretize the driving Brownian motion as a piecewise linear path and then solve the resulting ODE. This was introduced in [97] by Wong and Zakai, however their piecewise linear paths were constructed to match only the increments of the Brownian motion. By designing piecewise linear paths that match additional iterated integrals, the subsequent ODE methods can achieve higher orders of strong convergence. In this chapter, we developed several such paths that can be used to compute accurate pathwise discretizations of SDEs driven by a single Brownian motion.

- **Third order approximations of the underdamped Langevin equation**

  The underdamped Langevin equation appears in statistical mechanics as a model of a particle moving in a potential under the effect of stochastic forces. Recently this equation has been applied to computational statistics as, under mild conditions (see [20]), the process admits a stationary measure with density

  $$\varphi(q,p) \propto \exp\left(-\beta\left(U(q) + \frac{1}{2}\|p\|^2\right)\right),$$

  and thus gives a method for approximately sampling from a target distribution. We have proposed piecewise linear methods for the underdamped Langevin equation which can achieve both strong and weak convergence rates of $O(h^3)$. These numerical methods also enable one to apply state-of-the-art ODE solvers.

Therefore we have the following open questions relating to this chapter's research:

- What are the most effective piecewise linear discretizations of Brownian motion that lead to high order numerical methods for general multidimensional SDEs?
- How might this approach be used to design effective cubature paths? (see [73])
- Under what conditions on the potential do the proposed ODE approximations of the underdamped Langevin diffusion admit a unique stationary distribution?
- Would an ODE method for ULD exhibit exponential Wasserstein contractivity?
- Does the stationary measure of the ODE solution converge to that of the SDE? If so, could one quantify the rate of this convergence in a 2-Wasserstein sense?
- Similarly, the overdamped Langevin equation has applications within statistics. Would the Runge-Kutta method in section 4.4 be practical for this? (see [69])

Chapter 6

Variable step size methods for stochastic differential equations

It is well established within the ODE literature that for certain problems, there are significant advantages to propagating numerical solutions using variable step sizes. Concrete examples demonstrating this can be found in the textbooks [51] and [15]. On the other hand, we have only been discretizing SDEs using constant step sizes. Whilst this approach seems to be effective for our current numerical examples, it is possible that a fixed step size is inappropriate for some (non-homogeneous) systems. In addition, since the increments of Brownian motion are all normally distributed, there is always a non-zero probability that the path will experience large excursions. These excursions could then result in further numerical errors in an SDE simulation.

In this chapter, we shall investigate variable step size approaches which aim to control the (local) errors experienced by the high order methods studied previously. For these methodologies to be developed, we require the following two key ingredients:

- **A refinement procedure for increments and areas of a Brownian path**

  Reducing the step size is perhaps the most basic strategy for controlling errors. However the decision to take "a smaller step size" often comes after information about the Brownian path has been generated (with respect to a larger step size). Fortunately, Lévy's construction of Brownian motion enables us to generate increments of the path in an interval conditional on the endpoint information. It was shown in [13] that this construction can be extended to generate both increments and space-time Lévy areas of the Brownian path in a similar manner. In the first section of this chapter, we will further extend these constructions to generate the triples $(W_{s,t}, H_{s,t}, n_{s,t})$ and $(W_{s,t}, H_{s,t}, K_{s,t})$ on dyadic intervals. This allows the piecewise linear methods from Chapter 5 to use variable steps.

- **Estimation of local $L^2(\mathbb{P})$ errors conditional on the path's information**

  As the motivation for using variable step sizes is to reduce numerical errors, it is necessary to determine when a one-step approximation is insufficiently accurate. It can be seen from the Taylor expansions of our proposed numerical methods that the leading term in the variance expansion of the local error has the form:

$$\big([f_1,[f_1,f_0]](y_s)\big)^2 \operatorname{Var}\left(L_{s,t} \,\middle|\, W_{s,t}, H_{s,t}, \cdot\right),$$

  where the above variance would correspond to either (4.2.5), (4.2.7) or (5.1.7). A potential disadvantage of using this quantity to estimate local $L^2(\mathbb{P})$ errors is that it requires one to compute or approximate certain vector field derivatives. Furthermore, there is an additional complication if the triple $(W_{s,t}, H_{s,t}, K_{s,t})$ is used as the conditional variance would not depend on the path increment $W_{s,t}$. That said, we will see that this approach is effective for our numerical example.

In this chapter, we shall only consider strategies defined over dyadic partitions. The main reason for this is that the new "refinement procedures" established in the first section are limited to generating information at the midpoint of a given interval.

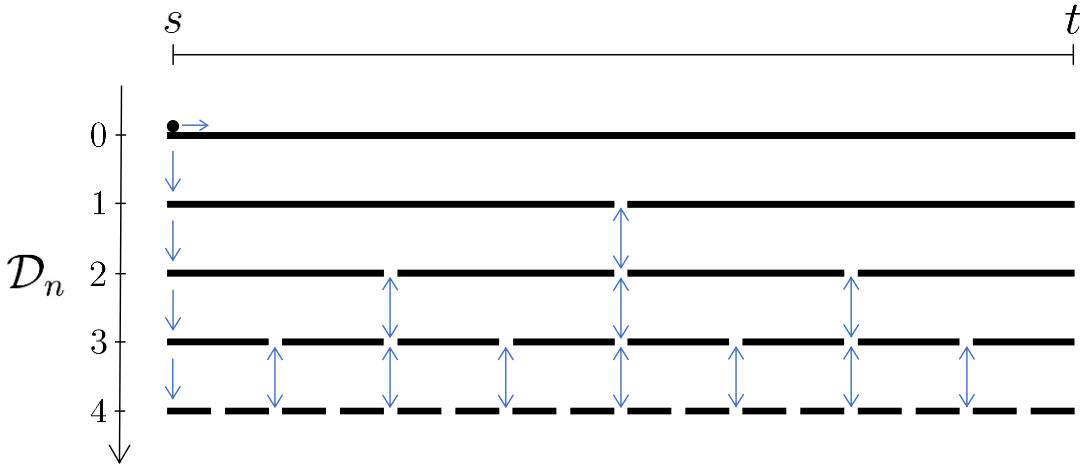

Figure 6.1: The particle can always take a smaller step, but not always a larger one.

Once an initial step size has been fixed, we can propagate the numerical solution over the whole interval or the two half intervals depending on our local error estimate. It is then possible to recursively apply this process for both of the smaller intervals. Proceeding in this way gives a variable step size methodology using dyadic partitions. The above diagram illustrates a potential limitation of this approach since it only allows step sizes to be "refined" (and therefore step sizes can not always be increased). That said, it will be shown in section 6.2 that the requirement for methods to only refine intervals is helpful and enables us to prove convergence using rough path theory.

Definition 6.0.1. The proposed variable step size methodology for SDEs of the form:

$$dy_t = f_0(y_t)\,dt + f_1(y_t) \circ dW_t, \tag{6.0.1}$$
$$y_0 = \xi,$$

is defined by the following steps:

Step 1. Fix an initial step size, $h = t - s$, for computing the numerical solution.

Step 2. Generate some information describing the Brownian path, $\big(W_{s,t}, H_{s,t}, \bullet\big)$.

Step 3. If h is below a prespecified constant $h_{\min}$, approximate (6.0.1) over $[s, t]$.

Step 4. Otherwise, check that the discretization error is small using the condition

$$\big([f_1, [f_1, f_0]](Y_s)\big)^2 \operatorname{Var}\big(L_{s,t} \,\big|\, W_{s,t}, H_{s,t}, \bullet\big) \leq Ch, \tag{6.0.2}$$

where Y_s denotes the approximation of y_s and $C > 0$ is some prespecified constant.

Step 5. If (6.0.2) is satisfied, approximate (6.0.1) over the interval $[s, t]$ using an appropriate numerical method $\big($i.e. one using the estimator $\mathbb{E}\big[L_{s,t} \,\big|\, W_{s,t}, H_{s,t}, \bullet\big]\big)$.

Step 6. If (6.0.2) is not satisfied, apply Steps 1-6 on the intervals $[s, u]$ and $[u, t]$, where $u = s + \frac{1}{2}h$ is the midpoint and Step 2 is now done by refinement procedure.

Remark 6.0.2. In numerical simulations, we would ideally like to avoid using Step 3, which was only included to ensure that the algorithm uses a finite number of steps. Depending on the problem, it may also be better to modify the condition (6.0.2) as an alternative way of controlling the (expected) computational cost of the algorithm.

Remark 6.0.3. The above variable step size methodology drew some inspiration from the paper of Gaines and Lyons [39] where a step size control was proposed for SDEs and aimed for steps to have uniform contribution to the variance of the global error. Since variances can "add up", the condition (6.0.2) was designed with this intention.

As discussed in [39], one of the challenges associated with variable step size methodologies is establishing convergence of the approximation to the SDE solution. The difficulty is due to the resulting partitions depending on the Brownian motion. In addition, there may be complications if the SDE is driven by a multidimensional Brownian path as local Taylor expansions can contain the following iterated integral:

$$\int_s^t \int_s^u \circ dW_v \otimes \circ dW_u. \tag{6.0.3}$$

Since there is currently no refinement procedure that exactly generates this quantity, it must be approximated – which gives an $O(h)$ error by the natural Brownian scaling.

In the second section of this chapter, we shall establish pathwise convergence for a large class of variable step size methods using key ideas from rough path theory. The core of this argument is showing that the piecewise polynomial approximants given in Chapter 4 can be "lifted" to an enhanced Brownian motion in a rough path sense along certain sequences of partitions $\{D_n\}$ with $\text{mesh}(D_n) \to 0$ almost surely. This result (along with its proof) was inspired by the following theorem given in [37].

Theorem 6.0.4 (Rough path convergence of nested piecewise linear approximants)**.** *Let $\{D_n\}$ denote a fixed sequence of nested partitions of $[0,T]$, that is $D_n \subset D_{n+1}$ for all n, such that $\text{mesh}(D_n) \to 0$ as $n \to \infty$. Let W^{D_n} be the (standard) piecewise linear approximation that agrees with the Brownian motion W on the partition D_n. Then W^{D_n} can be lifted to a geometric 2-rough path $\boldsymbol{W}^{D_n} := S^{(2)}_{0,T}\big(W^{D_n}\big)$ and we have*

$$d_{\alpha\text{-Höl};[0,T]}\big(\boldsymbol{W}^{D_n}, \boldsymbol{W}\big) \to 0,$$

almost surely where $d_{\alpha\text{-Höl};[0,T]}$ is the α-Hölder metric given by definition 3.2.19 with $\alpha \in \big(0, \frac{1}{2}\big)$ and $\boldsymbol{W}$ is the enhanced Brownian motion obtained via a Stratonovich lift.

Once this lift has been established, it then follows from the universal limit theorem (see Theorem 3.3.3) that the solution of an ODE with sufficiently regular vector fields and driven by a piecewise polynomial on D_n converges to the desired SDE solution. Moreover, we can argue that variable step size methods only need to be "locally close" to these polynomial driven ODE solutions to have guaranteed pathwise convergence. The key condition is that each one-step approximation should utilize an estimator for the iterated integral (6.0.3) that can be expressed as a conditional expectation since

$$\int_s^t \int_s^u dW_v^n \otimes dW_u^n = \mathbb{E}\left[\int_s^t \int_s^u \circ dW_v \otimes \circ dW_u \,\middle|\, \big\{ \cdots \big\} \right], \quad \forall n \geq 1, \tag{6.0.4}$$

where $\big\{ \cdots \big\}$ is the σ-algebra generated by $W_{s,t}$ along with $n-1$ time integrals of W.

In the last section of this chapter, we shall incorporate the high order piecewise linear paths developed in Chapter 5 into the above variable step size framework and numerically demonstrate the resulting method for the following (non-Lipschitz) SDEs:

- **Cox-Ingersoll-Ross (CIR) process** (see [23] for the original paper on this)

$$dy_t = a(b - y_t)\, dt + \sigma\sqrt{y_t}\, dW_t.$$

- **Squared δ-dimensional Bessel process** ([47] gives a survey of both processes)

$$dz_t = \delta\, dt + 2\sqrt{z_t}\, dW_t.$$

6.1 Extending Lévy's construction of Brownian motion via time integrals

As previously stated, the first ingredient that a variable step size method requires is a "refinement procedure" that can generate a Brownian path over subintervals of $[s,t]$. If we wish to generate increments of the Brownian motion, this can be done using the well-known Lévy's construction (see, for example, the proof of Theorem 1.3 in [79]):

Theorem 6.1.1 (Paul Lévy's construction of Brownian motion)**.** *For each $n \geq 0$, consider the random piecewise linear path $\widehat{W}_n$ defined on $[0,1]$ inductively as follows:*

$$\widehat{W}_0(1) \sim \mathcal{N}(0,1),$$
$$\widehat{W}_0(t) := t\,\widehat{W}_0(1),$$

and for each $k \in [0\,..\,2^n - 1]$, we can generate $\widehat{W}_{n+1}$ at equidistant dyadic points by

$$\widehat{W}_{n+1}\Big(\Big(k+\frac{1}{2}\Big)2^{-n}\Big) \sim \mathcal{N}\Big(\frac{1}{2}\big(\widehat{W}_n\big(k2^{-n}\big) + \widehat{W}_n\big((k+1)2^{-n}\big)\big), \frac{1}{4}\cdot 2^{-n}\Big),$$

where these normal random variables are independent (once their means are fixed), and we can extend $\widehat{W}_{n+1}$ to the whole interval by linearly interpolating between points,

$$\widehat{W}_{n+1}\big((k+t)2^{-n}\big) := \begin{cases} (1-2t)\widehat{W}_n\big(k2^{-n}\big) + 2t\,\widehat{W}_{n+1}\big(\big(k+\frac{1}{2}\big)2^{-n}\big), & \text{for } t \in \big[0,\frac{1}{2}\big] \\ (2-2t)\widehat{W}_{n+1}\big(\big(k+\frac{1}{2}\big)2^{-n}\big) + (2t-1)\,\widehat{W}_n\big((k+1)2^{-n}\big), & \text{for } t \in \big[\frac{1}{2},1\big] \end{cases}.$$

Then $\big\{\widehat{W}_n\big\}$ converges uniformly as $n \to \infty$ (almost surely) and the limiting process is distributed by the Wiener measure (i.e. it has the same law as Brownian motion).

As well as giving the standard argument for the existence of Brownian motion, the above theorem can be used to establish some basic continuity properties of W. In particular, Brownian paths are locally Hölder continuous (Corollary 1.20 in [79]):

Theorem 6.1.2. *Let α be a positive real number with $\alpha < \frac{1}{2}$. Then, almost surely, Brownian motion is everywhere locally α-Hölder continuous so that for all $s \in [0,\infty)$,*

$$\exists \epsilon > 0,\ \exists c > 0,\ \ |W_s - W_t| \leq c|s-t|^\alpha,\ \ \forall t \geq 0 \quad \text{with} \quad |s-t| < \epsilon.$$

The key idea behind Lévy's construction is that one can always generate W at the midpoint of a fixed interval using elementary properties of the Brownian bridge. Since we defined "higher degree" bridge processes in Chapter 4 (see definition 4.1.16), we should be able to extend this construction using integrals such as $H_{s,t}$ and $K_{s,t}$.

The diagram below illustrates how this process would refine a linear path on $[s, t]$.

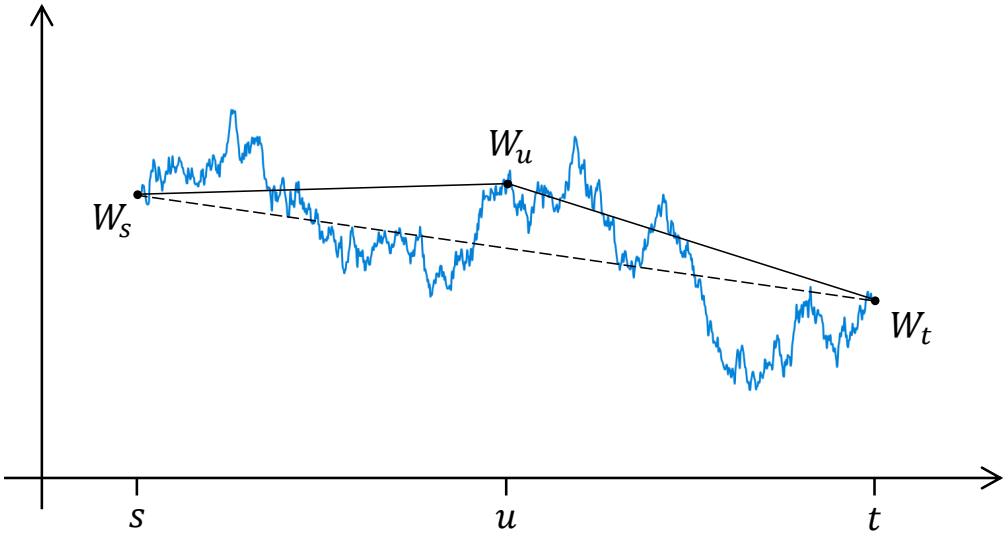

Figure 6.2: Brownian motion can be generated conditional on its values at endpoints.

Proposition 6.1.3. *For any value of $W_{s,t}$, we have $W_{s,u} \sim \mathcal{N}\left(\frac{u-s}{h} W_{s,t}, \frac{(u-s)(t-u)}{h}\right)$.*

Proof. The result follows using the covariance function of the Brownian bridge. □

Since two of the central objects studied in this thesis are space-time Lévy area and the Brownian arch, the above theorem naturally leads to the following question:

What is the joint distribution of $\left(W_{s,u}, W_{u,t}, H_{s,u}, H_{u,t}\right)$ conditional on $\left(W_{s,t}, H_{s,t}\right)$?

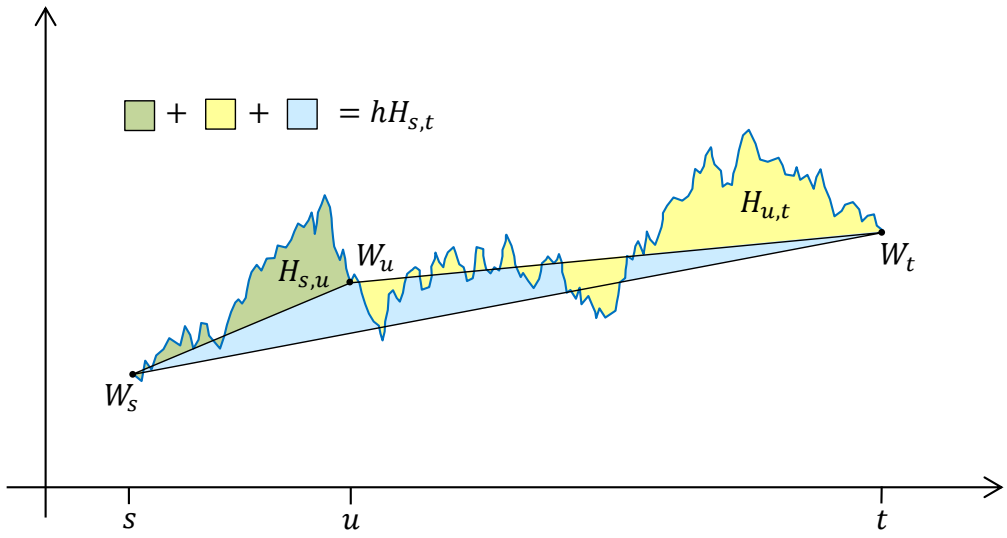

Figure 6.3: Brownian motion and its space-time area can be generated in an interval.

Recall that the distribution of $W_{s,u}$ conditional on the pair $\left(W_{s,t}, H_{s,t}\right)$ can be deduced from the $n = 2$ case of Theorem 4.1.12 using the natural Brownian scaling. Therefore all that remains is to establish the conditional distribution of $\left(H_{s,u}, H_{u,t}\right)$. Fortunately, this distribution was determined (albeit with different notation) in [13]. As this result is crucial for high order variable step size methods, it is presented below:

Theorem 6.1.4. *Let W denote a standard Brownian motion. Then for $u \in (s,t)$,*

$$\begin{aligned}
W_{s,u} &= \frac{u-s}{h} W_{s,t} + \frac{6(u-s)(t-u)}{h^2} H_{s,t} + \frac{2(a+b)}{h} X_1, \\
W_{u,t} &= \frac{t-u}{h} W_{s,t} - \frac{6(u-s)(t-u)}{h^2} H_{s,t} - \frac{2(a+b)}{h} X_1, \\
H_{s,u} &= \frac{(u-s)^2}{h^2} H_{s,t} - \frac{a}{u-s} X_1 + \frac{c}{u-s} X_2, \\
H_{u,t} &= \frac{(t-u)^2}{h^2} H_{s,t} - \frac{b}{t-u} X_1 - \frac{c}{t-u} X_2,
\end{aligned}$$

where $W_{s,t}, H_{s,t}, X_1, X_2$ are independent normally distributed random variables and

$$\begin{aligned}
X_1, X_2 &\sim \mathcal{N}(0,1), \\
a &:= \frac{(u-s)^{\frac{7}{2}} (t-u)^{\frac{1}{2}}}{2h\sqrt{(u-s)^3 + (t-u)^3}}, \\
b &:= \frac{(u-s)^{\frac{1}{2}} (t-u)^{\frac{7}{2}}}{2h\sqrt{(u-s)^3 + (t-u)^3}}, \\
c &:= \frac{\sqrt{3}\,(u-s)^{\frac{3}{2}} (t-u)^{\frac{3}{2}}}{6\sqrt{(u-s)^3 + (t-u)^3}}.
\end{aligned}$$

Proof. It follows by the independence of both Brownian increments and bridges that

$$\begin{aligned}
W_{s,t} &\sim \mathcal{N}\big(0, h\big), \\
W_{s,u} - \frac{u-s}{h} W_{s,t} &\sim \mathcal{N}\left(0, \frac{(u-s)(t-u)}{h}\right), \\
H_{s,u} &\sim \mathcal{N}\left(0, \frac{1}{12}(u-s)\right), \\
H_{u,t} &\sim \mathcal{N}\left(0, \frac{1}{12}(t-u)\right),
\end{aligned}$$

are independent random variables.

Consider the following 3×3 matrix:

$$M := \begin{pmatrix} \frac{6(u-s)(t-u)}{h^2} & \frac{2(a+b)}{h} & 0 \\ \frac{(u-s)^2}{h^2} & -\frac{a}{u-s} & \frac{c}{u-s} \\ \frac{(t-u)^2}{h^2} & -\frac{b}{t-u} & -\frac{c}{t-u} \end{pmatrix}, \tag{6.1.1}$$

where the constants $a, b, c > 0$ were defined in the statement of this theorem and

$$\begin{aligned} \det M = {} & \frac{6(u-s)(t-u)}{h^2} \left(\frac{a}{u-s} \cdot \frac{c}{t-u} + \frac{c}{u-s} \cdot \frac{b}{t-u} \right) \\ & + \frac{2\,(a+b)}{h} \left(\frac{(u-s)^2}{h^2} \cdot \frac{c}{t-u} + \frac{c}{u-s} \cdot \frac{(t-u)^2}{h^2} \right) > 0. \end{aligned}$$

Therefore M is invertible, and we can define the random variables X_0, X_1, X_2 by

$$X = \begin{pmatrix} X_0 \\ X_1 \\ X_2 \end{pmatrix} := M^{-1} \begin{pmatrix} W_{s,u} - \frac{u-s}{h} W_{s,t} \\ H_{s,u} \\ H_{u,t} \end{pmatrix}. \tag{6.1.2}$$

It is clear that X_0, X_1, X_2 are jointly normal with zero mean as they can be expressed as a linear combination of independent centered normal random variables. This means that the random vector X has a centered trivariate normal distribution. In addition, it is worthwhile noting that X is independent of the path increment $W_{s,t}$.

Using the linearity of expectations, the covariance matrix for X can be computed as

$$\begin{aligned} \Sigma &:= \mathbb{E}\big[X X^T \big] \\ &= M^{-1} \begin{pmatrix} \frac{(u-s)(t-u)}{h} & 0 & 0 \\ 0 & \frac{1}{12}(u-s) & 0 \\ 0 & 0 & \frac{1}{12}(t-u) \end{pmatrix} \big(M^{-1}\big)^T. \end{aligned}$$

In order to explicitly compute these covariances, we shall define the following matrices:

$$C := \begin{pmatrix} \frac{1}{12}h & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix}, \qquad D := MCM^T. \tag{6.1.3}$$

It can then be verified by a direct calculation that D is the following diagonal matrix:

$$\begin{aligned} D &= \begin{pmatrix} \frac{6(u-s)(t-u)}{h^2} & \frac{2(a+b)}{h} & 0 \\ \frac{(u-s)^2}{h^2} & -\frac{a}{u-s} & \frac{c}{u-s} \\ \frac{(t-u)^2}{h^2} & -\frac{b}{t-u} & -\frac{c}{t-u} \end{pmatrix} \begin{pmatrix} \frac{1}{12}h & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} \frac{6(u-s)(t-u)}{h^2} & \frac{(u-s)^2}{h^2} & \frac{(t-u)^2}{h^2} \\ \frac{2(a+b)}{h} & -\frac{a}{u-s} & -\frac{b}{t-u} \\ 0 & \frac{c}{u-s} & -\frac{c}{t-u} \end{pmatrix}, \\ &= \begin{pmatrix} \frac{(u-s)(t-u)}{h} & 0 & 0 \\ 0 & \frac{1}{12}(u-s) & 0 \\ 0 & 0 & \frac{1}{12}(t-u) \end{pmatrix}. \end{aligned}$$

Therefore, the matrix D can be used to simplify the computation of Σ as follows,

$$\begin{aligned}\Sigma &= M^{-1}D\left(M^{-1}\right)^T \\ &= M^{-1}\left(MCM^T\right)\left(M^{-1}\right)^T \\ &= C \\ &= \begin{pmatrix} \frac{1}{12}h & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix}.\end{aligned}$$

This means that X_0, X_1, X_2 are jointly normal and uncorrelated random variables. In particular, the above covariance matrix implies that they must be independent. By considering the different regions within the area $hH_{s,t}$ (see Figure 6.3), we have

$$\begin{aligned} hH_{s,t} &= (u-s)H_{s,u} + (t-u)H_{u,t} + \frac{1}{2}h\Big(W_{s,u} - \frac{u-s}{h}W_{s,t}\Big) \qquad (6.1.4)\\ &= \left(\frac{(u-s)^3}{h^2}X_0 - aX_1 + cX_2\right) + \left(\frac{(t-u)^3}{h^2}X_0 - bX_1 - cX_2\right) \\ &\quad + \left(\frac{3(u-s)(t-u)}{h}X_0 + (a+b)X_1\right) \\ &= \frac{1}{h^2}\left((u-s)^3 + 3(u-s)^2(t-u) + 3(u-s)(t-u)^2 + (t-u)^3\right)X_0 \\ &= hX_0. \end{aligned}$$

Thus $X_0 = H_{s,t}$ and so

$$\begin{pmatrix} W_{s,u} - \frac{u-s}{h}W_{s,t} \\ H_{s,u} \\ H_{u,t} \end{pmatrix} = \begin{pmatrix} \frac{6(u-s)(t-u)}{h^2} & \frac{2(a+b)}{h} & 0 \\ \frac{(u-s)^2}{h^2} & -\frac{a}{u-s} & \frac{c}{u-s} \\ \frac{(t-u)^2}{h^2} & -\frac{b}{t-u} & -\frac{c}{t-u} \end{pmatrix}\begin{pmatrix} H_{s,t} \\ X_1 \\ X_2 \end{pmatrix},$$

where $W_{s,t}, H_{s,t}, X_1, X_2$ are independent random variables with $X_1, X_2 \sim \mathcal{N}(0,1)$. □

Remark 6.1.5. The above gives an alternative proof of Theorem 4.1.12 when $n = 2$. Moreover, we can identify the Gaussian random variable X_1 with a Brownian arch as

$$Z_{s,u} \sim \frac{2(a+b)}{h}X_1.$$

In this way, we can interpret X_1 as a normalized version of the stochastic process Z,

$$X_1(u) := \left(\sqrt{\frac{(u-s)(t-u)}{h} - \frac{3(u-s)^2(t-u)^2}{h^3}}\right)^{-1} Z_{s,u}.$$

When u is the midpoint $u = s + \frac{1}{2}h$, X_2 becomes proportional to $N_{s,t} := H_{s,u} - H_{u,t}$.

Using the above result (and remark), we can extend Proposition 6.1.3 and derive a dyadic refinement procedure for generating increments and space-time Lévy areas.

Theorem 6.1.6. *Let u denote the midpoint $u = s + \frac{1}{2}h$ and $N_{s,t} := H_{s,u} - H_{u,t}$. Then*

$$\begin{aligned}
W_{s,u} &= \frac{1}{2}W_{s,t} + \frac{3}{2}H_{s,t} + Z_{s,u}, \\
W_{u,t} &= \frac{1}{2}W_{s,t} - \frac{3}{2}H_{s,t} - Z_{s,u}, \\
H_{s,u} &= \frac{1}{4}H_{s,t} - \frac{1}{2}Z_{s,u} + \frac{1}{2}N_{s,t}, \\
H_{u,t} &= \frac{1}{4}H_{s,t} - \frac{1}{2}Z_{s,u} - \frac{1}{2}N_{s,t},
\end{aligned}$$

where $W_{s,t}$, $H_{s,t}$, $Z_{s,u}$, $N_{s,t}$ are independent normally distributed random variables with

$$\begin{aligned}
W_{s,t} &\sim \mathcal{N}\left(0, h\right), \\
H_{s,t} &\sim \mathcal{N}\left(0, \frac{1}{12}h\right), \\
Z_{s,u} &\sim \mathcal{N}\left(0, \frac{1}{16}h\right), \\
N_{s,t} &\sim \mathcal{N}\left(0, \frac{1}{12}h\right).
\end{aligned}$$

Proof. The result follows using the same strategy as in the proof of Theorem 6.1.4. □

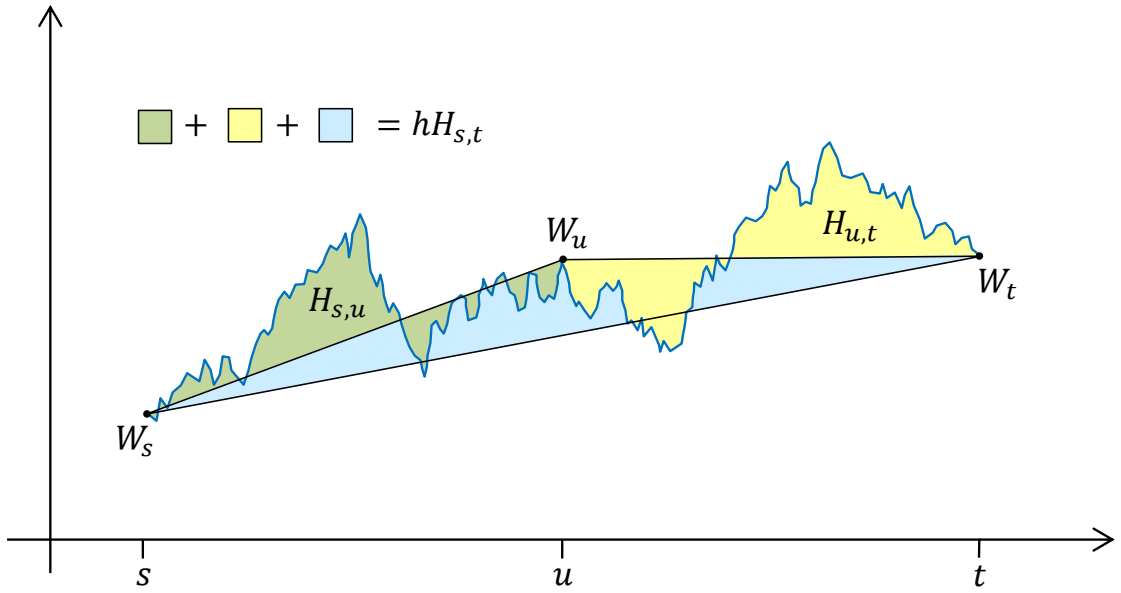

Figure 6.4: Brownian motion and its space-time area can be generated at a midpoint.

Recall that in the previous chapter, we studied numerical methods that require an additional random variable (along with the increment and space-time Lévy area). Therefore, we wish to extend Theorem 6.1.6 to refine triples of the form $(W_{s,t}, H_{s,t}, \cdot)$. It is clear that this is possible when the space-time orientation $n_{s,t} := \operatorname{sgn}(H_{s,u} - H_{u,t})$ is used due to the independence of the normal random variables: $W_{s,t}, H_{s,t}, Z_{s,u}, N_{s,t}$.

Corollary 6.1.7 (A dyadic refinement procedure for generating $(W_{s,t}, H_{s,t}, n_{s,t})$)**.** *Let u be the midpoint $u = s + \frac{1}{2}h$. Then, using the same notation as before, we have*

$$\begin{aligned}
W_{s,u} &= \frac{1}{2}W_{s,t} + \frac{3}{2}H_{s,t} + Z_{s,u}, \\
W_{u,t} &= \frac{1}{2}W_{s,t} - \frac{3}{2}H_{s,t} - Z_{s,u}, \\
H_{s,u} &= \frac{1}{4}H_{s,t} - \frac{1}{2}Z_{s,u} + \frac{1}{2}n_{s,t}\big|N_{s,t}\big|, \\
H_{u,t} &= \frac{1}{4}H_{s,t} - \frac{1}{2}Z_{s,u} - \frac{1}{2}n_{s,t}\big|N_{s,t}\big|,
\end{aligned}$$

where $\big|N_{s,t}\big|$ is a half-normal random variable independent of $\big(W_{s,t}, H_{s,t}, Z_{s,u}, n_{s,t}\big)$. Furthermore, $n_{s,u}$ and $n_{u,t}$ can be independently generated as Rademacher variables.

Proof. Since $n_{s,t} = \operatorname{sgn}(N_{s,t})$, the result is a direct consequence of Theorem 6.1.6. □

In section 6.3, this procedure will be used with a piecewise linear ODE method (from Chapter 5) to discretize the Cox-Ingersoll-Ross model using a variable step size. We shall now turn our attention to generating space-time-time Lévy area on dyadics. Most notably, this refinement procedure could lead to high order variable step size methods for the underdamped Langevin equation (see section 5.3 for further details).

Theorem 6.1.8 (A dyadic refinement procedure for generating $(W_{s,t}, H_{s,t}, K_{s,t})$)**.** *Let u denote the midpoint $u = s + \frac{1}{2}h$. Then, the following system of equations hold:*

$$\begin{aligned}
W_{s,u} &= \frac{1}{2}W_{s,t} + \frac{3}{2}H_{s,t} + Z_{s,u}, \\
W_{u,t} &= \frac{1}{2}W_{s,t} - \frac{3}{2}H_{s,t} - Z_{s,u}, \\
H_{s,u} &= \frac{1}{4}H_{s,t} + \frac{15}{4}K_{s,t} - \frac{1}{2}Z_{s,u} + X_1, \\
H_{u,t} &= \frac{1}{4}H_{s,t} - \frac{15}{4}K_{s,t} - \frac{1}{2}Z_{s,u} - X_1, \\
K_{s,u} &= \frac{1}{8}K_{s,t} - \frac{1}{2}X_1 + X_2, \\
K_{u,t} &= \frac{1}{8}K_{s,t} - \frac{1}{2}X_1 - X_2,
\end{aligned}$$

where $W_{s,t}, H_{s,t}, K_{s,t}, Z_{s,u}, X_1, X_2$ *are independent Gaussian random variables with*

$$X_1 \sim \mathcal{N}\Big(0, \frac{1}{768}h\Big),$$

$$X_2 \sim \mathcal{N}\Big(0, \frac{1}{2880}h\Big).$$

Proof. The random variables X_1 and X_2 will be explicitly defined using the formulae:

$$X_1 := \frac{15}{16}K_{s,u} + \frac{15}{16}K_{u,t} - \frac{1}{32}\big(H_{s,u} - H_{u,t}\big),$$
$$X_2 := \frac{1}{2}\big(K_{s,u} - K_{u,t}\big).$$

As $H_{s,u}, K_{s,u}, H_{u,t}, K_{u,t}$ are independent, we see (by symmetry) that $\mathrm{Cov}(X_1, X_2) = 0$. Thus, X_1 and X_2 are independent random variables with the required distributions. Similarly, it is straightforward to argue that X_2 is independent of $\big(W_{s,t}, H_{s,t}, K_{s,t}\big)$. Furthermore, we can use a symmetry argument to compute the following covariance.

$$\begin{aligned}\mathrm{Cov}\big(X_2, Z_{s,u}\big) &= \mathrm{Cov}\Big(\frac{1}{2}\big(K_{s,u} - K_{u,t}\big), W_{s,u} - \frac{1}{2}W_{s,t} - \frac{3}{2}H_{s,t}\Big)\\ &= \mathrm{Cov}\Big(\frac{1}{2}\big(K_{u,t} - K_{s,u}\big), \frac{1}{2}W_{s,t} + \frac{3}{2}H_{s,t}\Big)\\ &= 0.\end{aligned}$$

We have therefore established that X_2 is independent of $\big(W_{s,t}, H_{s,t}, K_{s,t}, Z_{s,u}, X_1\big)$. Showing the independence of X_1 is more involved but will follow a similar argument. Firstly, we shall express $K_{s,t}$ using the random variables defined on the half intervals. To simply the calculation, we introduce the following integral of the Brownian path:

$$M_{s,t} := \frac{1}{h^2}\int_s^t (r-s)W_{s,r}\,dr. \tag{6.1.5}$$

We can view (6.1.5) as a "shifted" version of the space-time-time Lévy area $K_{s,t}$ since

$$\begin{aligned}M_{s,t} &= \frac{1}{2h}\int_s^t W_{s,r}\,dr - \frac{1}{h^2}\int_s^t \Big(\frac{1}{2}h - (r-s)\Big)W_{s,r}\,dr\\ &= \frac{1}{2h}\Big(\frac{1}{2}hW_{s,t} + hH_{s,t}\Big) - \frac{1}{h^2}\int_s^t \Big(\frac{1}{2}h - (r-s)\Big)\Big(\frac{r-s}{h}W_{s,t}\Big)\,dr\\ &\qquad - \frac{1}{h^2}\int_s^t \Big(\frac{1}{2}h - (r-s)\Big)\Big(W_{s,r} - \frac{r-s}{h}W_{s,t}\Big)\,dr\\ &= \frac{1}{2}\Big(\frac{1}{2}W_{s,t} + H_{s,t}\Big) + \frac{1}{12}W_{s,t} - \frac{1}{h^2}\int_s^t \Big(\frac{1}{2}h - (r-s)\Big)\Big(W_{s,r} - \frac{r-s}{h}W_{s,t}\Big)\,dr\\ &= \frac{1}{3}W_{s,t} + \frac{1}{2}H_{s,t} - K_{s,t}.\end{aligned}$$

So decomposing $M_{s,t}$ over the intervals $[s,u]$ and $[u,t]$ will give an expression for $K_{s,t}$.

$$\begin{aligned}
M_{s,t} &= \frac{1}{h^2}\int_s^u (r-s)W_{s,r}\,dr + \frac{1}{h^2}\int_u^t (r-s)W_{s,r}\,dr \\
&= \frac{1}{4}M_{s,u} + \frac{1}{h^2}\int_u^t (u-s)W_{s,u}\,dr + \frac{1}{h^2}\int_u^t (r-u)W_{s,u}\,dr \\
&\qquad + \frac{1}{h^2}\int_u^t (u-s)W_{u,r}\,dr + \frac{1}{h^2}\int_u^t (r-u)W_{u,r}\,dr \\
&= \frac{1}{4}M_{s,u} + \frac{1}{4}W_{s,u} + \frac{1}{8}W_{s,u} + \frac{1}{2h}\left(\frac{1}{4}hW_{u,t} + \frac{1}{2}hH_{u,t}\right) + \frac{1}{4}M_{u,t} \\
&= \frac{1}{4}M_{s,u} + \frac{1}{4}M_{u,t} + \frac{3}{8}W_{s,u} + \frac{1}{8}W_{u,t} + \frac{1}{4}H_{u,t}.
\end{aligned}$$

By substituting the previous formulae $\big(M = \frac{1}{3}W + \frac{1}{2}H - K\big)$ into the above, we have

$$\begin{aligned}
\frac{1}{3}W_{s,t} + \frac{1}{2}H_{s,t} - K_{s,t} &= \frac{1}{12}W_{s,u} + \frac{1}{8}H_{s,u} - \frac{1}{4}K_{s,u} + \frac{1}{12}W_{u,t} + \frac{1}{8}H_{u,t} - \frac{1}{4}K_{u,t} \\
&\qquad + \frac{3}{8}W_{s,u} + \frac{1}{8}W_{u,t} + \frac{1}{4}H_{u,t} \\
&= \frac{1}{3}W_{s,t} + \frac{1}{2}\left(\frac{1}{2}H_{s,u} + \frac{1}{2}H_{u,t} + \frac{1}{2}\left(W_{s,u} - \frac{1}{2}W_{s,t}\right)\right) \\
&\qquad - \frac{1}{4}K_{s,u} - \frac{1}{4}K_{u,t} - \frac{1}{8}H_{s,u} + \frac{1}{8}H_{u,t} \\
&= \frac{1}{3}W_{s,t} + \frac{1}{2}H_{s,t} - \frac{1}{4}K_{s,u} - \frac{1}{4}K_{u,t} - \frac{1}{8}H_{s,u} + \frac{1}{8}H_{u,t},
\end{aligned}$$

where the final line is a direct consequence of (6.1.4), or equivalently Theorem 6.1.6. Therefore, we have derived the following decomposition of $K_{s,t}$ over the half intervals.

$$K_{s,t} = \frac{1}{4}K_{s,u} + \frac{1}{4}K_{u,t} + \frac{1}{8}\big(H_{s,u} - H_{u,t}\big). \tag{6.1.6}$$

Using this expression, it is possible to compute the covariance between X_1 and $K_{s,t}$.

$$\begin{aligned}
\mathrm{Cov}\big(X_1, K_{s,t}\big) &= \mathrm{Cov}\left(\frac{15}{16}K_{s,u} + \frac{15}{16}K_{u,t} - \frac{1}{32}N_{s,t}, \frac{1}{4}K_{s,u} + \frac{1}{4}K_{s,u} + \frac{1}{8}N_{s,t}\right) \\
&= \frac{15}{64}\mathrm{Var}\big(K_{s,u}\big) + \frac{15}{64}\mathrm{Var}\big(K_{u,t}\big) - \frac{1}{256}\mathrm{Var}\big(N_{s,t}\big) \\
&= \frac{15}{64}\cdot\frac{1}{1440} + \frac{15}{64}\cdot\frac{1}{1440} - \frac{1}{256}\cdot\frac{1}{12} \\
&= 0.
\end{aligned}$$

Hence, it is clear that the random variable X_1 is independent of $\big(W_{s,t}, H_{s,t}, K_{s,t}, X_2\big)$.

Therefore, the final covariance that we need to calculate is between X_1 and $Z_{s,u}$.

$$\begin{aligned}
\operatorname{Cov}\big(X_1, Z_{s,u}\big) &= \operatorname{Cov}\Big(\frac{15}{16}K_{s,u} + \frac{15}{16}K_{u,t} - \frac{1}{32}N_{s,t}, W_{s,u} - \frac{1}{2}W_{s,t} - \frac{3}{2}H_{s,t}\Big) \\
&= -\frac{45}{32}\operatorname{Cov}\big(K_{s,u} + K_{u,t}, H_{s,t}\big) \\
&= -\frac{45}{32}\operatorname{Cov}\Big(K_{s,u} + K_{u,t}, \frac{1}{2}H_{s,u} + \frac{1}{2}H_{u,t} + \frac{1}{2}\Big(W_{s,u} - \frac{1}{2}W_{s,t}\Big)\Big) \\
&= 0.
\end{aligned}$$

Hence the random variables $W_{s,t}$, $H_{s,t}$, $K_{s,t}$, $Z_{s,u}$, X_1, X_2 are all independent and

$$\begin{aligned}
W_{s,t} &= W_{s,u} + W_{u,t}, \\
H_{s,t} &= \frac{1}{2}H_{s,u} + \frac{1}{2}H_{u,t} + \frac{1}{4}\big(W_{s,u} - W_{u,t}\big), \\
K_{s,t} &= \frac{1}{4}K_{s,u} + \frac{1}{4}K_{s,u} + \frac{1}{8}\big(H_{s,u} - H_{u,t}\big), \\
Z_{s,u} &= \frac{1}{8}W_{s,u} - \frac{1}{8}W_{u,t} - \frac{3}{4}H_{s,u} - \frac{3}{4}H_{s,u}, \\
X_1 &= \frac{15}{16}K_{s,u} + \frac{15}{16}K_{u,t} - \frac{1}{32}\big(H_{s,u} - H_{u,t}\big), \\
X_2 &= \frac{1}{2}\big(K_{s,u} - K_{u,t}\big).
\end{aligned}$$

The result can now be obtained by solving this system of linear equations. □

We now have dyadic refinement procedures (Corollary 6.1.7 and Theorem 6.1.8) for the triples $(W_{s,t}, H_{s,t}, n_{s,t})$ and $(W_{s,t}, H_{s,t}, K_{s,t})$ studied in the previous chapter. Along with these random variables, we also considered generating $k_{s,t} := \operatorname{sgn}\big(K_{s,t}\big)$. Although $k_{s,t}$ may give a more accurate approximation than $n_{s,t}$, it has additional limitations as we cannot always deduce $k_{s,t}$ from $(W_{s,u}, H_{s,u}, k_{s,u})$, $(W_{s,u}, H_{u,t}, k_{u,t})$. This means that there is no "exact coupling" between the triple $\big(W, H, k\big)$ over the whole interval $[s,t]$ and the same variables over the two half intervals $[s,u]$, $[u,t]$. Thus, it is difficult to construct an efficient refinement procedure for $(W_{s,t}, H_{s,t}, k_{s,t})$. Since $n_{s,t}$ can be refined by Corollary 6.1.7, it is more straightforward to incorporate into variable step size methods than $k_{s,t}$ and shall be used in the numerical example.

The coupling of random variables over the intervals $[s,t]$, $[s,u]$, $[u,t]$ is also essential for the multilevel Monte Carlo method introduced by Giles (see [42, 43, 45]). Whilst it is ongoing research to incorporate the methods proposed in this thesis into a multilevel framework, it should be easier to use $n_{s,t}$ or $K_{s,t}$ instead of $k_{s,t}$ as they admit exact couplings for $\big(W_{s,t}, H_{s,t}, \cdot\big)$. For this line of research, we would be interested in studying the variance reduction exhibited by methods with a strong order of 1.5.

6.2 Pathwise convergence of polynomial based variable step size methods

In this section, we will present sufficient conditions for ensuring that a variable step size method converges (in a pathwise sense) to the solution of the Stratonovich SDE:

$$\begin{aligned} dy_t &= f_0(y_t)\,dt + \sum_{i=1}^{d} f_i(y_t) \circ dW_t^{(i)}, \\ y_0 &= \xi, \end{aligned} \tag{6.2.1}$$

where $\xi \in \mathbb{R}^e$, the vector fields $f_i : \mathbb{R}^e \to \mathbb{R}^e$ are all assumed to be $\mathrm{Lip}(\gamma)$ with $\gamma > 2$ (as detailed in definition 3.0.3) and W is a standard d-dimensional Brownian motion.

The first step in our analysis will be to consider the following two approximations:

Definition 6.2.1 (Discretization of Brownian motion using a piecewise polynomial)**.** For each $\omega \in \Omega$, let $D(\omega) = \{t_0 < t_1 < \cdots < t_N\}$ denote a partition of $[0, T]$ and let $\widetilde{W}^{n,D}(\omega) : [0, T] \to \mathbb{R}^d$ be the continuous path constructed from $W(\omega)$ and $D(\omega)$ as

$$\widetilde{W}^{n,D}\Big|_{[t_k, t_{k+1}]} := W^n_{[t_k, t_{k+1}]},$$

for each $k \in [0 \,..\, N-1]$ where $n \geq 1$ denotes a fixed integer and $W^n_{[t_k, t_{k+1}]}$ is defined as the unique n-th degree polynomial on $[t_k, t_{k+1}]$ satisfying $W^n_{[t_k, t_{k+1}]}(t_k) = W_{t_k}$ and

$$\int_{t_k}^{t_{k+1}} (u - t_k)^m \, dW^n_{[t_k, t_{k+1}]}(u) = \int_{t_k}^{t_{k+1}} (u - t_k)^m \, dW_u, \quad \text{for } m = 0, 1, \cdots, n-1.$$

Definition 6.2.2 (Polynomial driven ODE approximation for Stratonovich SDEs)**.** For $n \geq 1$ and $\omega \in \Omega$, let $\widetilde{Y}^{n,D}_t(\omega)$ be the unique solution of the below ODE on $[0, T]$,

$$\begin{aligned} d\widetilde{Y}^{n,D}_t &= f_0\big(\widetilde{Y}^{n,D}_t\big)\,dt + \sum_{i=1}^{d} f_1\big(\widetilde{Y}^{n,D}_t\big)\, d\widetilde{W}^{n,D}_t, \\ \widetilde{Y}^{n,D}_0 &= \xi, \end{aligned} \tag{6.2.2}$$

where ξ, f_i are from (6.2.1) and $\widetilde{W}^{n,D}$ is the piecewise polynomial in definition 6.2.1.

Given a discrete numerical solution of (6.2.1), there is a corresponding partition of $[0, T]$ that is obtained from the discretization points used in the SDE simulation. Therefore the action of "taking smaller step sizes" will add points to this partition. On the other hand, any sequence of partitions naturally corresponds with a sequence of piecewise polynomial approximations that are constructed using definition 6.2.2.

Thus, we will study the convergence of (6.2.2) along certain sequences of partitions. Since the decision to reduce a step can depend on the Brownian path itself, we shall represent a given variable step size methodology using the (deterministic) map below.

Definition 6.2.3 (Decision function for variable step size methods)**.** For $N \geq 1$, let $\mathcal{D}_N$ denote the set of partitions of $[0,T]$ with $N+1$ points (including 0 and T) and

$$\mathcal{D} := \bigcup_{N\geq 1} \mathcal{D}_N,$$
$$\widetilde{\mathcal{D}}^n := \bigcup_{N\geq 1} \mathcal{D}_N \times \left(\left(\mathbb{R}^d\right)^n\right)^N,$$

where $n \geq 1$. A decision function for a method using n pieces of information is a map

$$\Delta^n : \widetilde{\mathcal{D}}^n \to \mathcal{D}.$$

We wish to find conditions so that Δ^n gives partitions on which (6.2.2) converges. In order to express these conditions, we shall require the two definitions given below.

Definition 6.2.4 (Partially enhanced Brownian motion)**.** For $n \geq 1$ and $0 \leq s < t$, we say $\overline{W}^n_{s,t} := \left\{ \int_s^t (u-s)^k \, dW_u \right\}_{0\leq k\leq n-1}$ is a partially enhanced Brownian increment.

Definition 6.2.5 (Map induced sequence of partitions)**.** For a partition $D \in \mathcal{D}$, let

$$\overline{W}^n_D := \left\{ \overline{W}^n_{s,t} : s \text{ and } t \text{ are neighbouring points in } D \right\}.$$

Then, for an initial partition $D_0 \in \mathcal{D}$, we can define a sequence $\{D_m\}$ inductively as

$$D_{m+1} := \Delta^n \left(D_m, \overline{W}^n_{D_m} \right),$$

for $m \geq 0$.

Using these definitions, we can consider an important class of decision functions.

Definition 6.2.6 (Refining decision function for variable step size methodologies)**.** We say that $\Delta^n : \widetilde{\mathcal{D}}^n \to \mathcal{D}$ is refining if for every $D_0 \in \mathcal{D}$ we have $D_0 \subset \Delta^n\big(D_0, \cdot\big)$ and the induced sequence of partitions $\{D_m\}$ satisfies $\text{mesh}(D_m) \to 0$ almost surely.

Throughout this section, we will assume that the decision function Δ^n is refining as it naturally induces a filtration (which allows us to use certain martingale arguments).

Theorem 6.2.7. *Let $\Delta^n : \widetilde{\mathcal{D}}^n \to \mathcal{D}$ be a refining decision function for some $n \geq 1$. So for any fixed partition $D_0 \in \mathcal{D}$, Δ^n induces a sequence of nested partitions $\{D_m\}$. Letting $\overline{W}_t^n := \overline{W}_{0,t}^n$ denote the partially enhanced Brownian motion, we shall define*

$$\mathcal{F}_m^n := \sigma\big(\overline{W}_t^n : t \in D_m\big). \tag{6.2.3}$$

Then $\big\{\mathcal{F}_m^n\big\}_{m\geq 0}$ is a family of σ-algebras that is increasing in m, i.e. $\mathcal{F}^n$ a filtration.

Proof. Since D_0 is a fixed partition, it is clear that $\mathcal{F}_0^n$ is a well-defined σ-algebra. Suppose that for some $m \geq 0$, $\big\{\mathcal{F}_k^n\big\}_{0\leq k\leq m}$ is a family of σ-algebras increasing in k. As $D_{m+1} = \Delta^n\left(D_m, \overline{W}_{D_m}^n\right)$, we see that D_{m+1} is a $\mathcal{F}_m^n$-measurable random variable and therefore does not encode any additional information about the Brownian path. Since Δ^n is assumed to be refining, it follows that $D_m(\omega) \subset D_{m+1}(\omega)$ for all $\omega \in \Omega$. Thus $\sigma\big(\overline{W}_t^n : t \in D_m\big) \subseteq \sigma\big(\overline{W}_t^n : t \in D_{m+1}\big)$ and the result follows by induction. □

Remark 6.2.8. By viewing $\overline{W}^n$ as part of the signature of space-time Brownian motion $X^W := \{(\,\cdot, W_\cdot)\}$, we can apply Chen's relation (given by Theorem 3.1.20) to recover the increment $\overline{W}_{u,t}^n$ from the pair $\big(\overline{W}_{s,u}^n, \overline{W}_{s,t}^n\big)$ where $u \in (s,t)$. On the other hand, given the pair $\big(\overline{W}_{s,u}^n, \overline{W}_{u,t}^n\big)$, we can use Chen's relation to obtain the increment $\overline{W}_{s,t}^n$. Hence, this correspondence gives an alternative characterization of the filtration $\mathcal{F}^n$.

$$\mathcal{F}_m^n = \sigma\big(\overline{W}_{s,t}^n : s \text{ and } t \text{ are neighbouring points in } D_m\big). \tag{6.2.4}$$

Having defined the filtration $\mathcal{F}^n$, we shall highlight two of its key properties.

Theorem 6.2.9. *Let $\mathcal{F}^n$ be the filtration given by (6.2.3) and consider the σ-algebra*

$$\mathcal{F}_\infty^n := \sigma\left(\bigcup_{m\geq 0} \mathcal{F}_m^n\right).$$

Then, up to null sets, we have $\mathcal{F}_\infty^n = \sigma\big(\overline{W}_t^n : t \in [0,T]\big)$.

Proof. Recall that the decision function Δ^n was assumed to be refining, and thus $\operatorname{mesh}(D_m) \to 0$ almost surely. In addition, when the sample path $W(\omega)$ is continuous, so are the integrals $\big(\int_0^t u^k W_u\, du\big)(\omega)$ and we therefore have continuity of $t \mapsto \overline{W}_t^n(\omega)$. The result immediately follows since Brownian motion is continuous almost surely. □

Theorem 6.2.10. *Let $\{D_m\}$ be the sequence of (random) partitions associated with the filtration $\mathcal{F}^n$ defined in Theorem 6.2.7. Then the conditional expectation of W is*

$$\mathbb{E}\big[W_t \,|\, \mathcal{F}_m^n\big] = \widetilde{W}_t^{n,D_m}, \tag{6.2.5}$$

for $t \in [0,T]$.

Proof. Since D_0 is fixed, the result when $m = 0$ is a consequence of Theorem 4.1.12 along with the scaling and independent increment properties of Brownian motion. Recall that the partition D_{m+1} is $\mathcal{F}_m^n$-measurable and thus only encodes information obtained by evaluating the partially enhanced Brownian motion $\overline{W}^n(\omega)$ on $D_m(\omega)$. Assume there exists $m \geq 0$ so that whenever s and t are neighbouring points in $D_m(\omega)$, we have that $\{W_u - \widetilde{W}_u^{n,D_m} \,|\, \mathcal{F}_m^n\}_{u\in[s,t]}$ is an independent n-th degree Brownian bridge. Then by the independent increments of Brownian motion and the $\mathcal{F}_m^n$-measurability of D_{n+1}, it follows that when s and t are neighbouring points in $D_{m+1}(\omega)$, we still have that $\{W_u - \widetilde{W}_u^{n,D_{m+1}} \,|\, \mathcal{F}_{m+1}^n\}_{u\in[s,t]}$ is an independent n-th degree Brownian bridge. Hence by induction, we see that for each $m \geq 1$, the Brownian motion $W \,|\, \mathcal{F}_m^n$ has this "independent Brownian bridge" property and the result follows as each n-th degree Brownian bridge is an independent centered Gaussian process (with mean zero). $\square$

In order to incorporate ideas from rough path theory into our analysis, it will be necessary to lift the piecewise polynomial $\widetilde{W}^{n,D}$ to a path in a certain tensor algebra.

Definition 6.2.11. Let $\widetilde{W}^{n,D}$ denote the piecewise polynomial approximation of Brownian motion given by definition 6.2.1 and obtained using the partition $D(\omega) \in \mathcal{D}$. As it has finite variation, we can enhance $\widetilde{W}^{n,D}(\omega)$ to a path $\widetilde{\boldsymbol{W}}^{n,D}(\omega)$ in $T^{(2)}(\mathbb{R}^d)$ by

$$\begin{aligned}\widetilde{\boldsymbol{W}}^{n,D} &:= S_{0,T}^{(2)}(\widetilde{W}^{n,D}) \\ &= \left\{\left(1, \widetilde{W}_{s,t}^{n,D}, \int_s^t \int_s^u d\widetilde{W}_v^{n,D} \otimes d\widetilde{W}_u^{n,D}\right)\right\}_{0\leq s\leq t\leq T},\end{aligned}$$

where the above iterated integral of $\widetilde{W}^{n,D}$ is understood in a Riemann-Stieltjes sense.

We will now present our results concerning the pathwise convergence of $\{\widetilde{\boldsymbol{W}}^{n,D_m}\}$. In the rest of this section, $\boldsymbol{W}$ shall denote the Stratonovich lift of Brownian motion.

Theorem 6.2.12. *For any $0 \leq s \leq t \leq T$ we have that $\widetilde{\boldsymbol{W}}_{s,t}^{n,D_m} \to \boldsymbol{W}_{s,t}$ as $m \to \infty$ almost surely and in an $L^2(\mathbb{P})$ sense (where the convergence is viewed coordinatewise).*

Proof. By Theorems 6.2.7 and 6.2.9, we see that $\{\mathcal{F}_m^n\}_{m\geq 0}$ is a filtration and therefore

$$\{\widetilde{W}_{s,t}^{n,D_m}\}_{m\geq 0} = \{\mathbb{E}[W_{s,t} \,|\, \mathcal{F}_m^n]\}_{m\geq 0},$$

is a martingale. We can now apply Doob's martingale convergence theorem to give:

$$\mathbb{E}[W_{s,t} \,|\, \mathcal{F}_m^n] \to \mathbb{E}[W_{s,t} \,|\, \mathcal{F}_\infty^n],$$

almost surely and in $L^2(\mathbb{P})$. By Theorem 6.2.9, we have that $\mathcal{F}_\infty^n = \sigma\left(\overline{W}_t^n : t \in [0,T]\right)$ up to null sets (recall that the proof of this required $\text{mesh}(D_m) \to 0$ almost surely). Hence, it follows that $\widetilde{W}_{s,t}^{n,D_m} \to W_{s,t}$ almost surely and in an $L^2(\mathbb{P})$ sense as $m \to \infty$.

All that remains is to establish convergence for the iterated integrals of $\widetilde{W}^{n,D_m}$. Let $i, j \in \{1, \cdots, d\}$ with $i \neq j$ so that $W^{(i)}$ and $W^{(j)}$ are independent of each other. Recall from the proof of Theorem 6.2.10 that whenever s and t are neighbouring points in $D_m(\omega)$, we have that $\{W_u - \widetilde{W}_u^{n,D_m} \,|\, \mathcal{F}_m^n\}_{u\in[s,t]}$ is an independent n-th degree Brownian bridge. In particular, this means that $\big(W - \widetilde{W}^{n,D_m}\big)^{(i)}$ and $\big(W - \widetilde{W}^{n,D_m}\big)^{(j)}$ are independent when $i \neq j$. It is also worth noting that $\widetilde{W}^{n,D_m}$ is $\mathcal{F}_m^n$-measurable. Therefore, we can simplify the conditional expectation of the following Riemann sum:

$$
\begin{aligned}
&\mathbb{E}\bigg[\sum_{t_l\in\triangle_k} W^{(i)}_{s,t_l} W^{(j)}_{t_l,t_{l+1}} \,\Big|\, \mathcal{F}_m^n \bigg] \\
&\quad= \sum_{t_l\in\triangle_k} \mathbb{E}\big[W^{(i)}_{s,t_l} W^{(j)}_{t_l,t_{l+1}} \,|\, \mathcal{F}_m^n \big] \\
&\quad= \sum_{t_l\in\triangle_k} \mathbb{E}\Big[\big(W_{s,t_l} - \widetilde{W}^{n,D_m}_{s,t_l}\big)^{(i)} \big(W_{t_l,t_{l+1}} - \widetilde{W}^{n,D_m}_{t_l,t_{l+1}}\big)^{(j)} \,\Big|\, \mathcal{F}_m^n \Big] \\
&\qquad+ \sum_{t_l\in\triangle_k} \mathbb{E}\Big[W^{(i)}_{s,t_l} \big(\widetilde{W}^{n,D_m}_{t_l,t_{l+1}}\big)^{(j)} \,\Big|\, \mathcal{F}_m^n \Big] + \sum_{t_l\in\triangle_k} \mathbb{E}\Big[\big(\widetilde{W}^{n,D_m}_{s,t_l}\big)^{(i)} W^{(j)}_{t_l,t_{l+1}} \,\Big|\, \mathcal{F}_m^n \Big] \\
&\qquad- \sum_{t_l\in\triangle_k} \mathbb{E}\Big[\big(\widetilde{W}^{n,D_m}_{s,t_l}\big)^{(i)} \big(\widetilde{W}^{n,D_m}_{t_l,t_{l+1}}\big)^{(j)} \,\Big|\, \mathcal{F}_m^n \Big] \\
&\quad= \sum_{t_l\in\triangle_k} \mathbb{E}\Big[\big(W_{s,t_l} - \widetilde{W}^{n,D_m}_{s,t_l}\big)^{(i)} \,\Big|\, \mathcal{F}_m^n \Big] \mathbb{E}\Big[\big(W_{t_l,t_{l+1}} - \widetilde{W}^{n,D_m}_{t_l,t_{l+1}}\big)^{(j)} \,\Big|\, \mathcal{F}_m^n \Big] \\
&\qquad+ \sum_{t_l\in\triangle_k} \big(\widetilde{W}^{n,D_m}_{s,t_l}\big)^{(i)} \mathbb{E}\big[W^{(j)}_{t_l,t_{l+1}} \,\big|\, \mathcal{F}_m^n \big] + \sum_{t_l\in\triangle_k} \big(\widetilde{W}^{n,D_m}_{t_l,t_{l+1}}\big)^{(j)} \mathbb{E}\big[W^{(i)}_{s,t_l} \,\big|\, \mathcal{F}_m^n \big] \\
&\qquad- \sum_{t_l\in\triangle_k} \big(\widetilde{W}^{n,D_m}_{s,t_l}\big)^{(i)} \big(\widetilde{W}^{n,D_m}_{t_l,t_{l+1}}\big)^{(j)} \\
&\quad= \sum_{t_l\in\triangle_k} \big(\widetilde{W}^{n,D_m}_{s,t_l}\big)^{(i)} \big(\widetilde{W}^{n,D_m}_{t_l,t_{l+1}}\big)^{(j)},
\end{aligned}
$$

where $\{\triangle_k\}$ denotes a sequence of partitions on $[s,t]$ with $\mathrm{mesh}(\triangle_k) \to 0$ as $k \to \infty$.

Note that Itô integrals of W can be defined as the limit of their Riemann sum approximations where the convergence taking place is in the standard $L^2(\mathbb{P})$ sense. As $\{\triangle_k\}$ is a fixed sequence with $\mathrm{mesh}(\triangle_k) \to 0$, we can apply (13.3) in [37] to give

$$
\mathbb{E}\left[\left(\sum_{t_l\in\triangle_k} W^{(i)}_{s,t_l} W^{(j)}_{t_l,t_{l+1}} - \int_s^t W^{(i)}_{s,u}\, dW^{(j)}_u \right)^2 \right] \to 0,
$$

as $k \to \infty$. Using Jensen's inequality and the Tower law for $\mathbb{E}[\,\cdot\,|\,\mathcal{F}_m^n]$, it follows that

$$\mathbb{E}\left[\left(\mathbb{E}\left[\sum_{t_l\in\triangle_k} W_{s,t_l}^{(i)} W_{t_l,t_{l+1}}^{(j)} \,\middle|\, \mathcal{F}_m^n\right] - \mathbb{E}\left[\int_s^t W_{s,u}^{(i)}\, dW_u^{(j)} \,\middle|\, \mathcal{F}_m^n\right]\right)^2\right]$$
$$\leq \mathbb{E}\left[\left(\sum_{t_l\in\triangle_k} W_{s,t_l}^{(i)} W_{t_l,t_{l+1}}^{(j)} - \int_s^t W_{s,u}^{(i)}\, dW_u^{(j)}\right)^2\right] \to 0.$$

Since $W^{(i)}$ and $W^{(j)}$ are independent, there is no Itó-Stratonovich correction term for

$$\int_s^t W_{s,u}^{(i)}\, dW_u^{(j)} = \int_s^t W_{s,u}^{(i)} \circ dW_u^{(j)}.$$

Hence, taking the limit as $k \to \infty$ of the Riemann sum's conditional expectation gives

$$\mathbb{E}\left[\int_s^t W_{s,u}^{(i)} \circ dW_u^{(j)} \,\middle|\, \mathcal{F}_m^n\right] = \int_s^t \big(\widetilde{W}_{s,u}^{n,D_m}\big)^{(i)} d\big(\widetilde{W}_u^{n,D_m}\big)^{(j)}. \tag{6.2.6}$$

As before, since $\{\mathcal{F}_m^n\}_{m\geq 0}$ is a filtration, it follows that the above gives a martingale. So by applying Doob's martingale convergence theorem to these integrals, we have

$$\int_s^t \big(\widetilde{W}_{s,u}^{n,D_m}\big)^{(i)} d\big(\widetilde{W}_u^{n,D_m}\big)^{(j)} \to \mathbb{E}\left[\int_s^t W_{s,u}^{(i)} \circ dW_u^{(j)} \,\middle|\, \mathcal{F}_\infty^n\right] = \int_s^t W_{s,u}^{(i)} \circ dW_u^{(j)},$$

almost surely and in an $L^2(\mathbb{P})$ sense. Finally, for each $i \in \{1, \cdots, d\}$, we note that

$$\int_s^t \big(\widetilde{W}_{s,u}^{n,D_m}\big)^{(i)} d\big(\widetilde{W}_u^{n,D_m}\big)^{(i)} = \frac{1}{2}\Big(\big(\widetilde{W}_{s,t}^{n,D_m}\big)^{(i)}\Big)^2, \tag{6.2.7}$$

and

$$\int_s^t W_{s,u}^{(i)} \circ dW_u^{(i)} = \frac{1}{2}\Big(W_{s,t}^{(i)}\Big)^2.$$

The above is simply a consequence of the integration by parts formula. We can now apply the same martingale convergence argument to these "diagonal" integrals. □

In order to establish convergence of $\big\{\widetilde{\boldsymbol{W}}^{n,D_m}\big\}_{m\geq 0}$ in a rough path sense, we will first show the sequence is bounded using the α-Hölder norm given by definition 3.2.19.

Theorem 6.2.13. *For every $\alpha \in \big(0, \frac{1}{2}\big)$ there exists a positive random variable C, with $C < \infty$ almost surely, which is an α-Hölder upper bound for $\boldsymbol{W}$ and $\big\{\widetilde{\boldsymbol{W}}^{n,D_m}\big\}$.*

$$\|\boldsymbol{W}\|_{\alpha\text{-Höl};[0,T]} \leq C, \qquad \sup_{m\geq 0} \big\|\widetilde{\boldsymbol{W}}^{n,D_m}\big\|_{\alpha\text{-Höl};[0,T]} \leq C. \tag{6.2.8}$$

Proof. By corollary 13.15 in [37], there exists $c > 0$ (depending on α and T) such that

$$\mathbb{E}\Big[\exp\Big(c\,\|\boldsymbol{W}\|_{\alpha\text{-Höl};[0,T]}\Big)\Big] < \infty.$$

In particular, this implies that there exists a non-negative random variable C_1 with $\|C_1\|_{L^2(\mathbb{P})} < \infty$ so that, for almost all $\omega \in \Omega$, the increments of the path $\boldsymbol{W}(\omega)$ satisfy

$$\|\boldsymbol{W}_{s,t}\|^2 \leq C_1|t-s|^{2\alpha},$$

for all $0 \leq s < t \leq T$ where $\|\cdot\|$ is a homogeneous norm with the form (3.2.17). Hence, there exists a non-negative random variable C_2 with $\|C_2\|_{L^2(\mathbb{P})} < \infty$ such that

$$-C_2|t-s|^{2\alpha} \leq \int_s^t W_{s,u}^{(i)} \circ dW_u^{(j)} \leq C_2|t-s|^{2\alpha}.$$

Taking the expectation of the above inequality conditional on the σ-algebra $\mathcal{F}_m^n$ gives

$$-C_3|t-s|^{2\alpha} \leq \mathbb{E}\left[\int_s^t W_{s,u}^{(i)} \circ dW_u^{(j)} \,\middle|\, \mathcal{F}_m^n\right] \leq C_3|t-s|^{2\alpha},$$

where

$$C_3 := \sup_{m\geq 0} \mathbb{E}\big[C_2|\mathcal{F}_m^n\big].$$

Since $\big\{\mathbb{E}\big[C_2|\mathcal{F}_m^n\big]\big\}_{m\geq 0}$ is a martingale, we can apply Doob's maximal inequality,

$$\left\|\sup_{0\leq m\leq k} \mathbb{E}\big[C_2|\mathcal{F}_m^n\big]\right\|_{L^2(\mathbb{P})} \leq 2\big\|\mathbb{E}\big[C_2|\mathcal{F}_k^n\big]\big\|_{L^2(\mathbb{P})} \leq 2\|C_2\|_{L^2(\mathbb{P})}.$$

Therefore $\|C_3\|_{L^2(\mathbb{P})} \leq 2\|C_2\|_{L^2(\mathbb{P})} \leq \infty$. In particular, C_3 will be finite almost surely. It then follows from (6.2.6) and (6.2.7) that each iterated integral of $\widetilde{W}^{n,D_m}$ satisfies

$$-C_3|t-s|^{2\alpha} \leq \int_s^t \big(\widetilde{W}_{s,u}^{n,D_m}\big)^{(i)} d\big(\widetilde{W}_u^{n,D_m}\big)^{(j)} \leq C_3|t-s|^{2\alpha},$$

where the constant C_3 is independent of m. If necessary, replacing C_3 by d^2C_3 gives

$$\sup_{m\geq 0}\left\|\int_s^t \int_s^u d\widetilde{W}_v^{n,D_m} \otimes d\widetilde{W}_u^{n,D_m}\right\| \leq C_3|t-s|^{2\alpha}.$$

Similarly, we can apply the same martingale argument to obtain the below estimate:

$$\sup_{m\geq 0}\big\|\widetilde{W}_{s,t}^{n,D_m}\big\| \leq C_3|t-s|^{\alpha}.$$

Therefore, by the equivalence of homogeneous norms (3.2.18), it directly follows that

$$\|\widetilde{\boldsymbol{W}}_{s,t}^{n,D_m}\| \lesssim \max\left(\|\widetilde{W}_{s,t}^{n,D_m}\|, \left\|\int_s^t \int_s^u d\widetilde{W}_v^{n,D_m} \otimes d\widetilde{W}_u^{n,D_m}\right\|^{1/2}\right) \leq C_3|t-s|^{\alpha}, \tag{6.2.9}$$

where C_3 is a non-negative random variable, independent of m and almost surely finite. Note that (6.2.9) is precisely what is required to obtain an estimate for $\|\cdot\|_{\alpha\text{-Höl};[0,T]}$. The result now follows by setting the random variable C equal to $C_1 + C_3$. □

Using the above two theorems, we can establish pathwise convergence for $\{\widetilde{\boldsymbol{W}}^{n,D_m}\}$.

Theorem 6.2.14. *Let $\{D_m\}$ denote the sequence of partitions induced by a refining decision function Δ^n and let $\{\widetilde{\boldsymbol{W}}^{n,D_m}\}$ be the associated piecewise polynomials. Then*

$$d_{\alpha\text{-Höl};[0,T]}\big(\widetilde{\boldsymbol{W}}^{n,D_m}, \boldsymbol{W}\big) \to 0,$$

almost surely where $d_{\alpha\text{-Höl};[0,T]}$ denotes the α-Hölder metric given by definition 3.2.19 and $\boldsymbol{W}$ is the enhanced Brownian motion obtained from W using a Stratonovich lift.

Proof. By Theorem 6.2.12, the sequence $\{\widetilde{\boldsymbol{W}}^{n,D_m}\}_{m\geq 0}$ will converge almost surely (and in an $L^2(\mathbb{P})$ sense) to $\boldsymbol{W}$ when restricted to any uniform partition $\triangle_N$ of $[0,T]$. Using the uniform α-Hölder bound (6.2.8), we can estimate the difference between each $\widetilde{\boldsymbol{W}}^{n,D_m}$ and its piecewise linear discretization $\widehat{\boldsymbol{W}}^{n,D_m}$ that agrees on the grid $\triangle_N$.

$$\left\|\pi_k\Big(\widetilde{\boldsymbol{W}}_{s,t}^{n,D_m} - \boldsymbol{W}_{s,t}\Big)\right\| \leq \left\|\pi_k\Big(\widetilde{\boldsymbol{W}}_{s,t}^{n,D_m} - \widehat{\boldsymbol{W}}_{s,t}^{n,D_m}\Big)\right\| + \left\|\pi_k\Big(\widehat{\boldsymbol{W}}_{s,t}^{n,D_m} - \boldsymbol{W}_{s,t}\Big)\right\|.$$

So for fixed N and sufficiently large m, the RHS terms can be estimated by $\big(\frac{T}{N}\big)^{k\alpha}$. As this estimate is uniform, it follows that $\{\widetilde{\boldsymbol{W}}^{n,D_m}\}$ converges uniformly to $\boldsymbol{W}$ a.s. Note that for α-Hölder continuous rough paths $\boldsymbol{X}$ and $\boldsymbol{Y}$ $\big($with $0<\alpha<\beta\big)$, we have

$$\frac{\|\pi_k\big(\boldsymbol{X}_{s,t} - \boldsymbol{Y}_{s,t}\big)\|}{|t-s|^{k\alpha}} \leq \left(\frac{\|\pi_k\big(\boldsymbol{X}_{s,t} - \boldsymbol{Y}_{s,t}\big)\|}{|t-s|^{k\beta}}\right)^{\frac{\alpha}{\beta}} \left(\sup_{0\leq s\leq t\leq T} \big\|\pi_k\big(\boldsymbol{X}_{s,t} - \boldsymbol{Y}_{s,t}\big)\big\|\right)^{1-\frac{\alpha}{\beta}}.$$

The result follows as $\{\widetilde{\boldsymbol{W}}^{n,D_m}\}$ and $\boldsymbol{W}$ can be bounded in the β-Hölder sense. □

Corollary 6.2.15 (Convergence of piecewise polynomials in p-variation for $p>2$)**.**

$$d_{p\text{-var};[0,T]}\big(\widetilde{\boldsymbol{W}}^{n,D_m}, \boldsymbol{W}\big) \to 0,$$

almost surely where $d_{p\text{-var};[0,T]}$ denotes the p-variation metric given by definition 3.2.17.

Proof. The result immediately follows since $\|\boldsymbol{X}\|_{\frac{1}{\alpha}\text{-var};[s,t]} \leq \|\boldsymbol{X}\|_{\alpha\text{-Höl};[s,t]}(t-s)^{\alpha}$. □

Now that we have established that the piecewise polynomials $\{\widetilde{\boldsymbol{W}}^{n,D_m}\}$ converge in a rough path sense to the enhanced Brownian motion $\boldsymbol{W}$, we can apply Lyons' universal limit theorem (Theorem 3.3.3) to show convergence for the associated ODEs.

Theorem 6.2.16 (Convergence of polynomial driven ODEs along variable step sizes)**.** *Let y denote the true solution of the Stratonovich SDE (6.2.1) over the interval $[0,T]$, where the vector fields $f_i : \mathbb{R}^e \to \mathbb{R}^e$ are each assumed to be Lip(γ) for some $\gamma > 2$. Suppose that $\{D_m\}$ is a sequence of partitions induced by a refining decision function. For each $m \geq 0$, let $\widetilde{Y}^{n,D_m}$ denote the solution of the ODE driven by the piecewise polynomial $\widetilde{W}^{n,D_m}$ and governed by the same vector fields (see definition 6.2.2). Then*

$$d_{p\text{-var};[0,T]}\left(\widetilde{Y}^{n,D_m}, y\right) \to 0,$$

almost surely where $2 < p < \gamma$.

Proof. The result follows from the universal limit theorem and Corollary 6.2.15. □

In practice, it is unlikely that we can solve each "polynomial driven" ODE exactly. However some numerical methods can be viewed as being close to the ODE (6.2.2). For example, the Heun and midpoint methods are close to "piecewise linear" ODEs. In the rest of this section, we shall make these ideas precise and give our main result:

Theorem 6.2.17 (Pathwise convergence of variable step size numerical methods)**.** *Let y denote the true solution of the Stratonovich SDE (6.2.1) over the interval $[0,T]$, where the vector fields $f_i : \mathbb{R}^e \to \mathbb{R}^e$ are each assumed to be Lip(γ) for some $\gamma > 2$. Suppose that $\{D_m\}$ is a sequence of partitions induced by a refining decision function. For $m \geq 0$, let Y^{n,D_m} denote a numerical solution of (6.2.2) with $Y_0^{n,D_m} = y_0$ and*

$$\left\| Y_{t_{k+1}}^{n,D_m} - \Phi_{t_k,t_{k+1}}^n\left(Y_{t_k}^{n,D_m}\right) \right\| \leq w(t_{k+1} - t_k), \tag{6.2.10}$$

for each pair of consecutive points $t_k, t_{k+1} \in D_m$ where $\Phi_{t_k,t_{k+1}}^n\left(Y_{t_k}^{n,D_m}\right)$ denotes the solution at time $t = t_{k+1}$ of the below ODE driven by the Brownian polynomial $\widetilde{W}^{n,D_m}$:

$$\begin{aligned} d\widetilde{Y}_t &= f_0\left(\widetilde{Y}_t\right) dt + \sum_{i=1}^{d} f_1\left(\widetilde{Y}_t\right) d\widetilde{W}_t^{n,D_m}, \\ \widetilde{Y}_{t_k} &= Y_{t_k}^{n,D_m}, \end{aligned}$$

and $w(h) = o(h)$ as $h \to 0$ almost surely. Then $\left\{Y^{n,D_m}\right\}$ converges pathwise, that is

$$\sup_{t_k \in D_m} \|Y_{t_k}^{n,D_m} - y_{t_k}\| \to 0,$$

as $m \to \infty$ almost surely.

Proof. Our argument is inspired by the proof of Theorem 4.3 in [39] and will require a corollary of Davie's lemma (which establishes a Lipschitz property for RDE flows):

Lemma 6.2.18 (ODEs driven by Brownian polynomials produce Lipschitz flows)**.** *There exists a positive random variable K with $K < \infty$ almost surely so that for each piecewise polynomial approximation $\widetilde{W}^{n,D_m}$ of the path W and $s, t \in [0,T]$, we have*

$$\big\|\widetilde{X}_t^{n,D_m} - \widetilde{Y}_t^{n,D_m}\big\| \leq K\big\|\widetilde{X}_s^{n,D_m} - \widetilde{Y}_s^{n,D_m}\big\|, \tag{6.2.11}$$

almost surely, where $\widetilde{X}^{n,D_m}$ and $\widetilde{Y}^{n,D_m}$ denote two solutions of the ODE (6.2.2) that have different initial conditions. The constant K depends on the vector fields $\{f_i\}$, the time horizon $T > 0$ and the positive random variable C given by Theorem 6.2.13.

Proof. The result immediately follows by applying Davie's lemma (see Theorem 3.3.6) to the family of ODEs corresponding to the piecewise polynomials $\big\{\widetilde{W}^{n,D_m}\big\}$. This is possible due to the uniform α-Hölder bound that was given by Theorem 6.2.13. □

For each $k \geq 0$, let $\big\{\widetilde{Y}_{k,t}^{n,D_m}\big\}_{t \geq t_k}$ denote the solution the polynomial driven ODE:

$$\begin{aligned} d\widetilde{Y}_{k,t}^{n,D_m} &= f_0\big(\widetilde{Y}_{k,t}^{n,D_m}\big)\,dt + \sum_{i=1}^{d} f_1\big(\widetilde{Y}_{k,t}^{n,D_m}\big)\,d\widetilde{W}_t^{n,D_m}, \\ \widetilde{Y}_{k,t_k}^{n,D_m} &= Y_{t_k}^{n,D_m}, \end{aligned} \tag{6.2.12}$$

where $t_k(\omega)$ is the endpoint of the k-th interval in the partition $D_m(\omega)$, for $\omega \in \Omega$. Then by the original assumption (6.2.10) on the approximation Y^{n,D_m}, it follows that

$$\Big\|Y_{t_{k+1}}^{n,D_m} - \widetilde{Y}_{k,t_{k+1}}^{n,D_m}\Big\| \leq w(t_{k+1} - t_k).$$

So by lemma 6.2.18, we obtain the following estimate for each $t \in D_m$ with $t \geq t_{k+1}$:

$$\Big\|\widetilde{Y}_{k+1,t}^{n,D_m} - \widetilde{Y}_{k,t}^{n,D_m}\Big\| \leq K\,\Big\|Y_{t_{k+1}}^{n,D_m} - \widetilde{Y}_{k,t_{k+1}}^{n,D_m}\Big\| \leq Kw(t_{k+1} - t_k).$$

Expressing $\widetilde{Y}_{k,t_k}^{n,D_m} - \widetilde{Y}_{0,t_k}^{n,D_m}$ as a telescoping sum and using the triangle inequality gives

$$\Big\|\widetilde{Y}_{k,t_k}^{n,D_m} - \widetilde{Y}_{0,t_k}^{n,D_m}\Big\| \leq K\sum_{i<k} w(t_{i+1} - t_i).$$

Since $\widetilde{Y}_{k,t_k}^{n,D_m} = Y_{t_k}^{n,D_m}$ and $\big\{\widetilde{Y}_{0,t}^{n,D_m}\big\}_{t \geq 0}$ has initial condition $\widetilde{Y}_{0,0}^{n,D_m} = y_0$, we have that

$$\sup_{t_k \in D_m} \big\|Y_{t_k}^{n,D_m} - \widetilde{Y}_{t_k}^{n,D_m}\big\| \leq K\sum_{i} w(t_{i+1} - t_i). \tag{6.2.13}$$

It is now worth recalling that (for almost all $\omega \in \Omega$), $\mathrm{mesh}(D_m) \to 0$ as $m \to \infty$ and $w(h) = o(h)$ as $h \to 0$. Hence the RHS of (6.2.13) converges to zero as $m \to \infty$ a.s. The result now immediately follows from the above estimate and Theorem 6.2.16. □

Remark 6.2.19. This result already has an interesting application in the $n = 1$ case. In this setting, $\widetilde{W}^{1,D}$ is a standard piecewise linear approximation of W and therefore

$$\int_{t_k}^{t_{k+1}} \int_{t_k}^{u} d\widetilde{W}_v^{1,D} \otimes d\widetilde{W}_u^{1,D} = \frac{1}{2} W_{t_k,t_{k+1}}^{\otimes 2},$$

for each pair of consecutive discretization points $t_k, t_{k+1} \in D$. So Theorem 6.2.17 would apply to numerical methods with each one-step approximation taking the form:

$$Y_t = Y_s + f_0(y_s)h + \sum_{i=1}^{d} f_i(y_s) W_{s,t}^{(i)} + \frac{1}{2} \sum_{i,j=1}^{d} f_i'(y_s) f_j(y_s) W_{s,t}^{(i)} W_{s,t}^{(j)} + o(h),$$

where $h = t - s$. In particular, this means that the Heun and midpoint methods will converge when incorporated into a variable step size framework (provided that the underling decision function is refining and the vector fields are sufficiently regular). On the other hand, this result clearly does not apply to the Euler-Maruyama method.

Remark 6.2.20. In order to apply Davie's lemma and the universal limit theorem, we required the vector fields $\{f_i\}$ to all be $\mathrm{Lip}(\gamma)$ with $\gamma > 2$. It may be possible that the conditions on the f_0 vector field can be weakened due to time having finite variation. Thus we conjecture that Theorem 6.2.17 would still hold if f_0 were only $\mathrm{Lip}(1)$, and the proof may require the theory of Π-rough paths (as introduced by Gyurkó in [49]).

Although Theorem 6.2.17 establishes pathwise convergence for a large class of numerical methods, it is qualitative and does not give explicit rates of convergence. That said, quantitative results are known for piecewise linear approximations of W:

Theorem 6.2.21 (Corollary 13.22 from [37])**.** *Let D denote a fixed partition of $[0, T]$. Then for $\alpha \in [0, \frac{1}{2})$ and $\eta \in (0, \frac{1}{2} - \alpha)$, there exists a constant C depending only on (α, η, T) such that for $k \in \{1, 2\}$ and $q \geq 1$,*

$$\left\| \sup_{s,t \in [0,T]} \frac{\left\| \pi_k\big(\boldsymbol{W}_{s,t} - S_{s,t}^{(2)}(W^D)\big) \right\|}{|t-s|^{k\alpha}} \right\|_{L^q(\mathbb{P})} \leq C q^{\frac{k}{2}} \big(\mathrm{mesh}(D)\big)^{\eta}, \tag{6.2.14}$$

where $W^D : [0, T] \to \mathbb{R}^d$ is the piecewise linear path that coincides with W on D and $S_{s,t}^{(2)}(W^D)$ is the 2-step truncated signature of W^D on $[s, t]$ given by definition 3.1.16.

So when the step sizes do not depend on the Brownian motion, we can apply the above theorem to obtain an α-Hölder convergence rate of $O\big(h_{\max}^{\frac{1}{2}-\varepsilon}\big)$ from (6.2.14) (where $h_{\max}$ denotes the largest step size used). By Davie's lemma (Theorem 3.3.6), we have the same α-Hölder convergence rate for the associated ODE approximations provided the vector fields are $\mathrm{Lip}(\gamma)$. It is still an open problem to understand whether these convergence rates extend to the setting when the variable step sizes used come from a refining decision function and therefore may depend on the Brownian motion.

6.3 Simulation of the Cox-Ingersoll-Ross (CIR) model and Bessel process

In this section, we shall use the ideas discussed in previous chapters to present a new (variable step size) numerical discretization of the Cox-Ingersoll-Ross (CIR) process:

$$dy_t = a(b - y_t)\,dt + \sigma\sqrt{y_t}\,dW_t, \tag{6.3.1}$$

where $a, b > 0$ are the mean reversion speed/level parameters and $\sigma \geq 0$ is volatility.

The CIR process was first introduced in [23] and is now a well-established and popular one-factor short rate model used within the mathematical finance literature (see [11]). Since the solution of (6.3.1) is both mean-reverting and non-negative, the CIR process is suitable for modelling interest rates, stochastic volatilities and default intensities. As the vector fields are smooth on $(0, \infty)$, we can write (6.3.1) in Stratonovich form:

$$dy_t = a(\tilde{b} - y_t)\,dt + \sigma\sqrt{y_t} \circ dW_t, \tag{6.3.2}$$

provided that the "adjusted" mean reversion level parameter $\tilde{b} := b - \frac{1}{4a}\sigma^2$ is positive. Therefore, we impose the following constraints on the parameters of the CIR model:

$$\theta := \frac{2ab}{\sigma^2} > \frac{1}{2}. \tag{6.3.3}$$

The quantity θ is often referred to as the Feller ratio. In particular when $\theta \geq 1$, the CIR process is almost surely positive whereas if $\theta < 1$, the process touches zero with a non-zero probability [58]. Therefore the condition (6.3.3) allows for this latter case. If $\theta \leq \frac{1}{2}$, then it could be problematic to discretize the Stratonovich SDE (6.3.2) near zero since the drift $\tilde{b} \leq 0$ would no longer ensure that the diffusion is non-negative. Thus, for our proposed method to be non-negative, we require that (6.3.3) is satisfied.

In any case, increments of the CIR process follow a closed-form distribution given by

$$y_t | y_s = \frac{1}{c} X,$$

where $c = \frac{4a}{1 - e^{-a(t-s)}}$ and X is a random variable that follows a non-central chi-squared distribution with $\frac{4ab}{\sigma^2}$ degrees of freedom and non-centrality parameter $ce^{-a(t-s)}y_s$. Since it is possible to sample from this distribution, y_t can be generated exactly. However in practice, it can be computationally expensive to do so (see Table 1 in [2]). This provides the motivation for developing numerical methods for the CIR diffusion. That said, there are two significant challenges one faces when approximating (6.3.1):

- The solution y is non-negative almost surely. Thus, we would like the proposed method for the CIR model to always produce non-negative numerical solutions.
- The square root vector field is not globally Lipschitz as its derivatives "blow-up" near the origin. For example, this makes it is difficult to successfully apply the log-ODE method (see definition 4.3.4) to the CIR model for small values of y.

In [53], the authors establish an upper bound for the $L^1(\mathbb{P})$ error of approximations that use just the increments of the driving Brownian path (with uniform time steps).

Theorem 6.3.1 (Theorem 1 in [53])**.** *Let y denote the solution to the SDE (6.3.1) on the interval $[0,T]$ with the mean reverse speed and volatility parameters satisfying $2a < \sigma^2$. Then there exists a positive constant $c > 0$ such that for all $N \geq 1$, we have*

$$\inf_{\substack{\varphi : \mathbb{R}^N \to \mathbb{R} \\ \textit{Borel measurable}}} \mathbb{E}\Big[\big| y_T - \varphi(W_{\frac{T}{N}}, W_{\frac{2T}{N}}, \cdots, W_T) \big| \Big] \geq c \Big(\frac{1}{N} \Big)^{\frac{2a}{\sigma^2}}.$$

For the CIR process, the majority of (strong) numerical methods proposed in the literature are based on either an explicit or implicit Euler method and use a fixed step size (we refer the reader to [2], [4], [7], [24] and [29] for examples of such methods). However, it follows from Theorem 6.3.1 that these methods will exhibit arbitrary slow convergence rates in the "higher volatility" settings. Thus, the authors of [53] argue that there is a need to develop more sophisticated approximations of the CIR process.

This leads one to consider either high order weak approximations (such as in [3]) and/or variable step size methods (an adaptive scheme was recently proposed in [59]). The below table summarizes convergence results for some recent Euler-type methods:

Numerical method	L^p convergence rate and parameter assumptions
Truncated Euler [24]	$O\big(\sqrt{h}\big)$, $\quad 2 \leq p < \frac{2ab}{\sigma^2} - 1.$
Drift-implicit Euler [2, 4, 29]	$\begin{cases} O\big(\sqrt{h \log h}\big), & 1 \leq p < \frac{2ab}{\sigma^2}. \\ O(h), & 1 \leq p < \frac{4ab}{3\sigma^2}. \end{cases}$
Symmetrized Euler method [7]	$O\big(\sqrt{h}\big)$, $\ \frac{2ab}{\sigma^2} > 1 + \sqrt{8}\Big(\frac{\sqrt{a(16p-1)}}{\sigma} \vee (16p-2) \Big).$
Adaptive Euler method [59]	$O\big(\sqrt{h_{\max}}\big)$, $\quad ab \geq 2\sigma^2,\ p = 2.$

Table 6.1: Convergence rates for Euler-based approximations of the CIR process.

We will simulate the CIR model using a high order piecewise linear discretization of Brownian motion (given in Chapter 5) and variable step sizes (see definition 6.0.1). Importantly, this is outside the scope of [53] and thus Theorem 6.3.1 does not apply.

The proposed method can also be applied to the squared δ-dimensional Bessel process:

$$dz_t = \delta\, dt + 2\sqrt{z_t} \circ dW_t, \tag{6.3.4}$$

provided that $\delta > 0$. The Bessel process is an important one-dimensional diffusion which arises when one studies $\{\|W_t\|_2\}_{t\geq 0}$ (for a d-dimensional Brownian motion W). That said, we will focus our attention on the CIR model as it has significant real-world applications and presents a slightly greater challenge from a numerical perspective. Moreover, since Bessel processes are more analytically tractable than the CIR model, it may be the case that we should design a numerical method specifically for (6.3.4).

Throughout this section, we shall assume that $\sigma^2 < 4ab$ so that $\tilde{b} > 0$ in (6.3.2). Since our approach is based on driving the CIR model by a piecewise linear path $\widehat{W}$, we will require an appropriate numerical method for discretizing the following ODE:

$$d\widehat{y}_t = a\big(\tilde{b} - \widehat{y}_t\big)\, dt + \sigma\sqrt{\widehat{y}_t}\, d\widehat{W}_t, \tag{6.3.5}$$

which, along each linear piece of $\widehat{W}$, can then be expressed as the solution at $t = 1$ of

$$\frac{d\widehat{y}}{dt} = a\Big(\tilde{b} - \widehat{y}\Big)\Delta t + \sigma\sqrt{\widehat{y}}\,\Delta\widehat{W}. \tag{6.3.6}$$

The numerical method that we will apply to (6.3.6) belongs to the following family.

Definition 6.3.2 (Diagonally implicit Runge-Kutta methods)**.** Consider the ODE:

$$\begin{aligned} y' &= F(t, y), \\ y(0) &= y_0, \end{aligned} \tag{6.3.7}$$

where $y_0 \in \mathbb{R}^d$ and $F : [0, \infty) \times \mathbb{R}^d \to \mathbb{R}^d$ denotes a time-varying vector field on $\mathbb{R}^d$. A diagonally implicit m-stage Runge-Kutta method is defined using the recurrence:

$$\begin{aligned} Y_{k+1} &:= Y_k + h_k \sum_{i=1}^{m} b_i k_i, \\ Y_0 &:= y_0, \end{aligned} \tag{6.3.8}$$

where $h_k := t_{k+1} - t_k$, $b_1, \cdots, b_m \in \mathbb{R}$ and each stage $k_i \in \mathbb{R}^d$ satisfies the equation

$$k_i = F\left(t_k + c_i h_k, Y_k + h_k \sum_{j=1}^{i} a_{ij} k_j\right), \tag{6.3.9}$$

with real coefficients $\{a_{ij}\}$ and $\{c_i\}$. Note that (6.3.9) may not have a unique solution.

It is usually more convenient to express the above method using the Butcher tableau:

$$\begin{array}{c|cccc} c_1 & a_{11} & & & \\ c_2 & a_{21} & a_{22} & & \\ \vdots & \vdots & \vdots & \ddots & \\ c_m & a_{m1} & a_{m2} & \cdots & a_{mm} \\ \hline & b_1 & b_2 & \cdots & b_m \end{array}$$

Using (6.3.9), we can apply the standard Taylor theorem for F to expand (6.3.8). This can then be directly compared with the Taylor expansion of the ODE (6.3.7). The theory of rooted trees provide an effective tool called "B-series" (see [51] or [15]) for computing these expansions which enables one to derive order conditions on the coefficients $\{a_{ij}\}$, $\{b_i\}$ and $\{c_i\}$. The use of rooted trees in the theory of ODEs can be viewed as a precursor to the algebraic theory of rough paths and SPDEs (see [36]).

Applying the change of variable $\widehat{z} := \sqrt{\widehat{y}}$ to (6.3.6) results in the below equation.

$$\frac{d\widehat{z}}{dt} = \frac{a}{2}\left(\frac{\tilde{b}}{\widehat{z}} - \widehat{z}\right)\Delta t + \frac{1}{2}\sigma\Delta\widehat{W}. \tag{6.3.10}$$

For discretizing the (square root) CIR process, we shall use the following method:

Definition 6.3.3 (Third order A-stable diagonally implicit Runge-Kutta method)**.** This three-stage method is presented in [60] and defined by the below Butcher tableau.

$$\begin{array}{c|ccc} 0 & 0 & & \\ \frac{3+\sqrt{3}}{3} & \frac{3+\sqrt{3}}{6} & \frac{3+\sqrt{3}}{6} & \\ 1 & \frac{3+\sqrt{3}}{12} & \frac{1-\sqrt{3}}{4} & \frac{3+\sqrt{3}}{6} \\ \hline & \frac{3+\sqrt{3}}{12} & \frac{1-\sqrt{3}}{4} & \frac{3+\sqrt{3}}{6} \end{array} \tag{6.3.11}$$

Although the vector field of (6.3.10) is not globally Lipschitz (or even bounded), we shall see that the numerical method given above is particularly well suited for it.

Theorem 6.3.4. *When applied to the ODE (6.3.10), the Runge-Kutta method (6.3.11) reduces to solving two quadratic equations which will always have positive solutions, provided $\widehat{Z}_0 > 0$ and $\sigma^2 < 4ab$. The corresponding method for the CIR model is then*

$$Y_{t+\Delta t} := \left(\frac{C_1 + \sqrt{C_1^2 + \big(1 + ca\Delta t\big)\big(ca\tilde{b}\Delta t\big)}}{1 + ca\Delta t}\right)^2, \tag{6.3.12}$$

where Y_t is the numerical solution of (6.3.2) computed at time t (with $Y_0 := y_0$) and

$$
\begin{aligned}
c &:= \frac{3+\sqrt{3}}{12}, \\
C_1 &:= \frac{1}{2}\left(\sqrt{Y_t} + cF\Big(\sqrt{Y_t}\Big) + \frac{1-\sqrt{3}}{4}F\Big(\widetilde{Z}_t\Big) + c\sigma\Delta W\right), \\
\widetilde{Z}_t &:= \frac{C_2 + \sqrt{C_2^2 + \big(1+ca\Delta t\big)\big(c\tilde{b}\Delta t\big)}}{1+ca\Delta t}, \\
C_2 &:= \frac{1}{2}\left(\sqrt{Y_t} + 2cF\Big(\sqrt{Y_t}\Big) + c\sigma\Delta W\right),
\end{aligned}
$$

with $F : (0,\infty) \to \mathbb{R}$ denoting the vector field that governs the square root process $\widehat{Z}$,

$$
F(z) := \frac{a}{2}\left(\frac{\tilde{b}}{z} - z\right)\Delta t + \frac{1}{2}\sigma\Delta\widehat{W}. \tag{6.3.13}
$$

Proof. Applying the Runge-Kutta method (6.3.11) to the ODE (6.3.10) on $[0,1]$ gives

$$
\widehat{Z}_1 := \widehat{Z}_0 + \frac{3+\sqrt{3}}{12}k_1 + \frac{1-\sqrt{3}}{4}k_2 + \frac{3+\sqrt{3}}{6}k_3, \tag{6.3.14}
$$

where

$$
\begin{aligned}
k_1 &= F\big(\widehat{Z}_0\big), \\
k_2 &= F\left(\widehat{Z}_0 + \frac{3+\sqrt{3}}{6}k_1 + \frac{3+\sqrt{3}}{6}k_2\right), \\
k_3 &= F\left(\widehat{Z}_0 + \frac{3+\sqrt{3}}{12}k_1 + \frac{1-\sqrt{3}}{4}k_2 + \frac{3+\sqrt{3}}{6}k_3\right).
\end{aligned}
$$

Let $\widetilde{Z} := \widehat{Z}_0 + \frac{3+\sqrt{3}}{6}k_1 + \frac{3+\sqrt{3}}{6}k_2$ so that $k_1 = F\big(\widehat{Z}_0\big)$, $k_2 = F\big(\widetilde{Z}\big)$ and $k_3 = F\big(\widehat{Z}_1\big)$. Substituting the values for $\{k_i\}$ into the definitions of $\widetilde{Z}$ and $\widehat{Z}_1$ gives the equations:

$$
\widetilde{Z} = \widehat{Z}_0 + \frac{3+\sqrt{3}}{6}F\big(\widehat{Z}_0\big) + \frac{3+\sqrt{3}}{6}F\big(\widetilde{Z}\big), \tag{6.3.15}
$$

$$
\widehat{Z}_1 = \widehat{Z}_0 + \frac{3+\sqrt{3}}{12}F\big(\widehat{Z}_0\big) + \frac{1-\sqrt{3}}{4}F\big(\widetilde{Z}\big) + \frac{3+\sqrt{3}}{6}F\big(\widehat{Z}_1\big). \tag{6.3.16}
$$

As Y_t is the numerical solution of (6.3.2) computed at time t, we will set $\widehat{Z}_0 = \sqrt{Y_t}$. Similarly, we shall define $Y_{t+\Delta t}$ to be $\widehat{Z}_1^2$. Since $F(z)$ has the form $Az + B + Cz^{-1}$, both (6.3.15) and (6.3.16) can be expressed as quadratic equations. The result now follows by using the quadratic formula to find the positive roots for $\widetilde{Z}$ and $\widehat{Z}_1$. □

Remark 6.3.5. In general, each step of an implicit Runge-Kutta method requires a root-finding algorithm (such as the Newton-Raphson method) to compute each stage. Depending on the problem, this could be challenging or computationally expensive. Fortunately, due to the analytic tractability of the vector fields that govern (6.3.10), we have found a positive solution $Y_{t+\Delta t}$ that can be computed quickly and accurately.

Having chosen a method for discretizing the ODE (6.3.5) along each piece of $\widehat{W}$, the next step is to decide which "high order" piecewise linear path should be used. However, before doing so, it worth noting the Taylor expansion for the CIR process.

Theorem 6.3.6 (Stochastic Taylor expansion for the Cox-Ingersoll-Ross process)**.** *Let y be the solution of the SDE (6.3.2). Applying the Taylor expansion (4.3.2) gives*

$$\begin{aligned} y_t = y_s &+ \sigma\sqrt{y_s}W_{s,t} + a\big(\tilde{b} - y_s\big)h + \frac{1}{4}\sigma^2 W_{s,t}^2 \\ &- a\sigma\sqrt{y_s}\int_s^t\int_s^u \circ dW_v\, du + \frac{a\sigma}{2\sqrt{y_s}}\big(\tilde{b} - y_s\big)\int_s^t\int_s^u dv \circ dW_u \\ &- \frac{1}{2}a^2\big(\tilde{b} - y_s\big)h^2 - \frac{1}{2}a\sigma^2\int_s^t\int_s^u\int_s^v \circ dW_r \circ dW_v\, du \\ &- \Big(\frac{1}{2}a\sigma^2 + \frac{a\sigma^2}{4y_s}\big(\tilde{b} - y_s\big)\Big)\int_s^t\int_s^u\int_s^v \circ dW_r\, dv \circ dW_u \\ &+ \textit{remainder terms}, \end{aligned} \tag{6.3.17}$$

where the remainder would be $O\big(h^{\frac{5}{2}}\big)$ in the "Lipschitz setting" (as in Theorem 4.3.1).

Proof. The result follows by setting $f_0(y) = a\big(\tilde{b} - y\big)$ and $f_1(y) = \sigma\sqrt{y}$ in (4.3.2). □

Remark 6.3.7. There is no $W_{s,t}^4$ term appearing within the above Taylor expansion. Therefore, a third order numerical method would be sufficient for discretizing (6.3.5).

Remark 6.3.8. The SDE (6.3.2) has non-commuting vector fields (see section 5.1) as

$$\begin{aligned} [f_1,[f_1,f_0]](y) &= \big(f_0' f_1' f_1 - 2f_1' f_0' f_1 + f_1' f_1' f_0 \\ &\qquad + f_0''(f_1,f_1) - 2f_1''(f_0,f_1) + f_1''(f_1,f_0)\big)(y) \\ &= -\frac{1}{2}a\sigma^2 + a\sigma^2 + \frac{\sigma^2}{4y}\cdot a\big(\tilde{b} - y\big) + \frac{\sigma^2}{2y}\cdot a\big(\tilde{b} - y\big) - \frac{\sigma^2}{4y}\cdot a\big(\tilde{b} - y\big) \\ &= \frac{a\tilde{b}\sigma^2}{2y}. \end{aligned}$$

Since the vector fields of the CIR model do not commute with each other and we aim to have variable step sizes, we shall use the $(W_{s,t}, H_{s,t}, n_{s,t})$-measurable piecewise linear path given by definition 5.1.6 (with Theorem 5.1.4 detailing its key properties).

Since we are discretizing the CIR process using the $(W_{s,t}, H_{s,t}, n_{s,t})$-measurable approximation of Brownian motion from Chapter 5 and the third order Runge-Kutta method given by (6.3.11), the leading term in the (squared) $L^2(\mathbb{P})$ error expansion is

$$\left(\frac{a\tilde{b}\sigma^2}{2y_s}\right)^2 \mathrm{Var}\left(L_{s,t} \,\middle|\, W_{s,t}, H_{s,t}, n_{s,t}\right). \tag{6.3.18}$$

So in order to control local errors, we shall use the strategy given by definition 6.0.1.

Definition 6.3.9. The proposed variable step size methodology for the CIR process

$$\begin{aligned} dy_t &= a\big(\tilde{b} - y_t\big)\,dt + \sigma\sqrt{y_t} \circ dW_t, \\ y_0 &> 0, \end{aligned}$$

over $[0, T]$ is defined recursively (starting from $Y_0 := y_0$) by the following procedure:

Step 0. Assume the numerical solution Y_s at time s has already been computed.

Step 1. Fix an initial step size, $h = t - s$, for computing the numerical solution. After the first iteration of this procedure, h is obtained by halving a previous step.

Step 2. Generate the triple $\big(W_{s,t}, H_{s,t}, n_{s,t}\big)$ for describing the Brownian motion. Recall that this information is straightforward to generate due to Theorem 5.0.6.

Step 3. If h is below a prespecified constant $h_{\min}$, approximate (6.3.2) over $[s, t]$.

The numerical method we shall use is obtained by approximating the ODE (6.3.5) with $\widehat{W}$ denoting the $\big(W_{s,t}, H_{s,t}, n_{s,t}\big)$-measurable piecewise linear path given by definition 5.1.6. Along each linear piece of $\widehat{W}$, we will discretize (6.3.5) using one step of the third order implicit Runge-Kutta method (6.3.11). By Theorem 6.3.4, this method is practical and guaranteed to give a positive numerical solution Y_t.

Step 4. Otherwise, check that the discretization error is small using the condition

$$\left(\frac{a\tilde{b}\sigma^2}{2Y_s}\right)^2 \mathrm{Var}\left(L_{s,t} \,\middle|\, W_{s,t}, H_{s,t}, n_{s,t}\right) \leq Ch, \tag{6.3.19}$$

where $C > 0$ is a prespecified constant and the variance is given by Theorem 5.1.3.

Step 5. If (6.3.19) is satisfied, approximate the SDE (6.3.2) over the interval $[s, t]$ using the piecewise linear ODE method described above (with initial value Y_s).

Step 6. If (6.3.19) is not satisfied, apply Steps 1-6 on the intervals $[s, u]$ and $[u, t]$ where $u = s + \frac{1}{2}h$ is the midpoint and Step 2 is now done by using the refinement procedure given by Corollary 6.1.7.

Applying the above procedure to the CIR model on the whole interval $[0, T]$ gives a discretization where each step satisfies the condition (6.3.19) or has size $h \leq h_{\min}$. A worthwhile sanity check for the strategy is to estimate the step sizes for small y. Since the derivatives of $\sqrt{\cdot}$ blow-up near zero, we would want to use smaller steps. For any fixed Y_s, the LHS of the condition (6.3.19) will scale with h as $O(h^4)$ and so

$$h \propto Y^{\frac{2}{3}}.$$

In particular, this means the step size h tends to zero as Y approaches zero if $h_{\min} = 0$. That said, the step sizes taken by the proposed strategy can also be viewed as being asymptotically large when compared to the numerical solution (provided it is small).

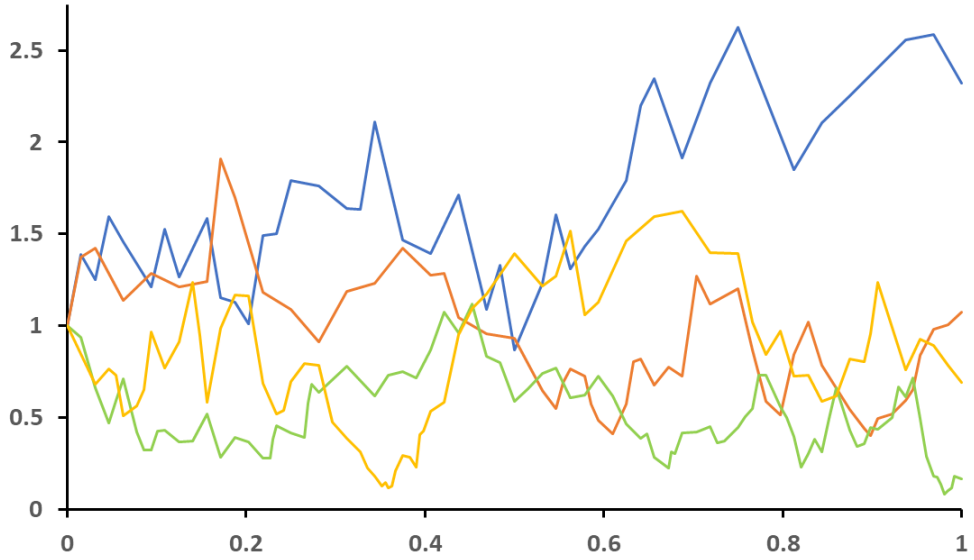

Figure 6.5: CIR sample paths with $a = 1$, $b = 1$, $\sigma^2 = 3$ and the same control C. When close to zero, the sample paths exhibit smaller fluctuations and step sizes.

For the numerical experiment, we compare the proposed variable step size method (definition 6.3.9) with the best-performing implicit scheme proposed by Alfonsi in [2].

Definition 6.3.10 (Drift-implicit Euler scheme [2, 4, 29])**.** We construct a numerical solution of (6.3.2) by setting $Y_0 := y_0$ and for each $k \in [0 \,..\, N-1]$, defining Y_{k+1} as

$$Y_{k+1} := \left(\frac{\frac{1}{2}\sigma W_{t_k,t_{k+1}} + \sqrt{Y_k} + \sqrt{\left(\frac{1}{2}\sigma W_{t_k,t_{k+1}} + \sqrt{Y_k}\right)^2 + 2a\tilde{b}h\left(1 + \frac{1}{2}ah\right)}}{2\left(1 + \frac{1}{2}ah\right)} \right)^2,$$

with $t_k := kh$ and $h := \frac{T}{N}$. The solution can then be approximated at time t_k by Y_k.

This method can be derived by applying the implicit Euler method to (6.3.10), but with $\widehat{W}$ now denoting the standard piecewise discretization of Brownian motion.

The drift-implicit Euler method has been well studied in the literature and has known strong convergence rates. We refer the reader to Table 6.3 or [4, 29] for further details.

We shall now present the estimators used for quantifying strong and weak errors. The functional chosen for testing weak convergence is $p(y) = (y-b)^+$ which is the payoff of a European call option with strike b and well approximated by polynomials.

Definition 6.3.11 (Strong and weak error estimators)**.** Let Y_N be the numerical solution of (6.3.2) computed at time T using the proposed variable step size method, as described in definition 6.3.9, with N steps (where N is allowed to be random). The first step size will be chosen to be equal to the time horizon T and $h_{\min} := \frac{1}{2^{20}}\,T$. We can define the following estimators for quantifying strong and weak convergence:

$$S_N := \sqrt{\mathbb{E}\left[\left(Y_N - Y_T^{\text{fine}}\right)^2\right]}, \tag{6.3.20}$$

$$E_N := \left|\mathbb{E}\left[\left(Y_N - b\right)^+\right] - \mathbb{E}\left[\left(Y_T^{\text{fine}} - b\right)^+\right]\right|, \tag{6.3.21}$$

where the above expectations are approximated by standard Monte-Carlo simulation and Y_T^{fine} denotes the numerical solution of (6.3.2) obtained at time T using the proposed piecewise linear approximation but with a "fine" step size of $\min\left(\frac{h}{64}, \frac{T}{4096}\right)$, with h denoting the step size used by the variable step size method to produce Y_N. Note that Y_N and Y_T^{fine} are both computed with respect to the same Brownian paths.

With a slight abuse of notation, we also use Y_N to denote the numerical solution of (6.3.2) obtained at time T using the drift-implicit Euler method (with N steps). We shall emulate the numerical experiments done in [2], and use a constant step size of $h = \frac{T}{N}$ to discretize (6.3.2) using the drift-implicit Euler method (definition 6.3.10). In this case, we will also use S_N and E_N to denote strong and weak error estimators. The only difference is that Y_T^{fine} is computed with fixed step size $h = \min\left(\frac{h}{10}, \frac{T}{5000}\right)$. For both numerical methods, the fine step size is chosen so that the $L^2(\mathbb{P})$ error between the true solution y_T and approximation Y_T^{fine} is negligible compared to S_N.

In this numerical example, we shall use the same parameter values as in [2], namely

$$\textbf{Low volatility case:} \quad a = 1,\ b = 1,\ \sigma^2 = 1,\ y_0 = 1,\ T = 1.$$
$$\textbf{High volatility case:} \quad a = 1,\ b = 1,\ \sigma^2 = 3,\ y_0 = 1,\ T = 1.$$

We will use 100,000 sample paths to compute the strong/weak error estimators. In addition, we can average the number of steps that are taken in each simulation. This "expected" number of steps can then be used to measure the computational cost of the method. As step sizes are fixed for the Euler method, this simply becomes N.

We will now present our results for this numerical experiment using log-log plots. Each data point for the proposed scheme is obtained with a fixed step size control C.

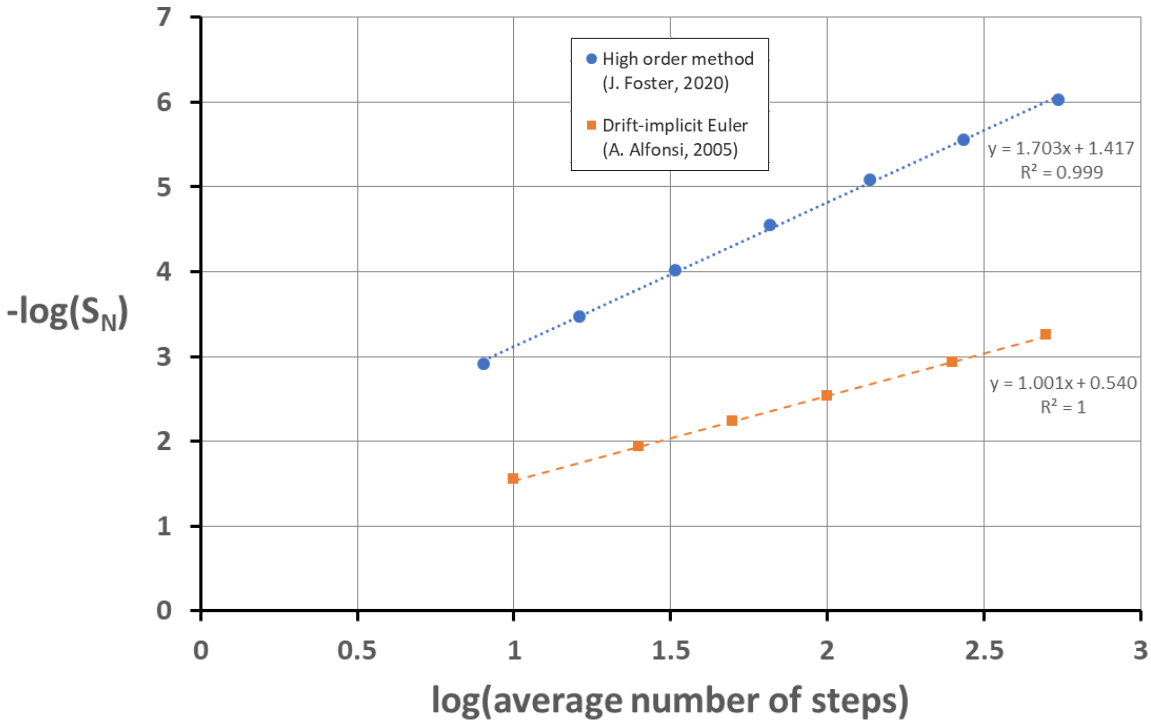

Figure 6.6: Strong convergence rates for both methods in the low volatility setting.

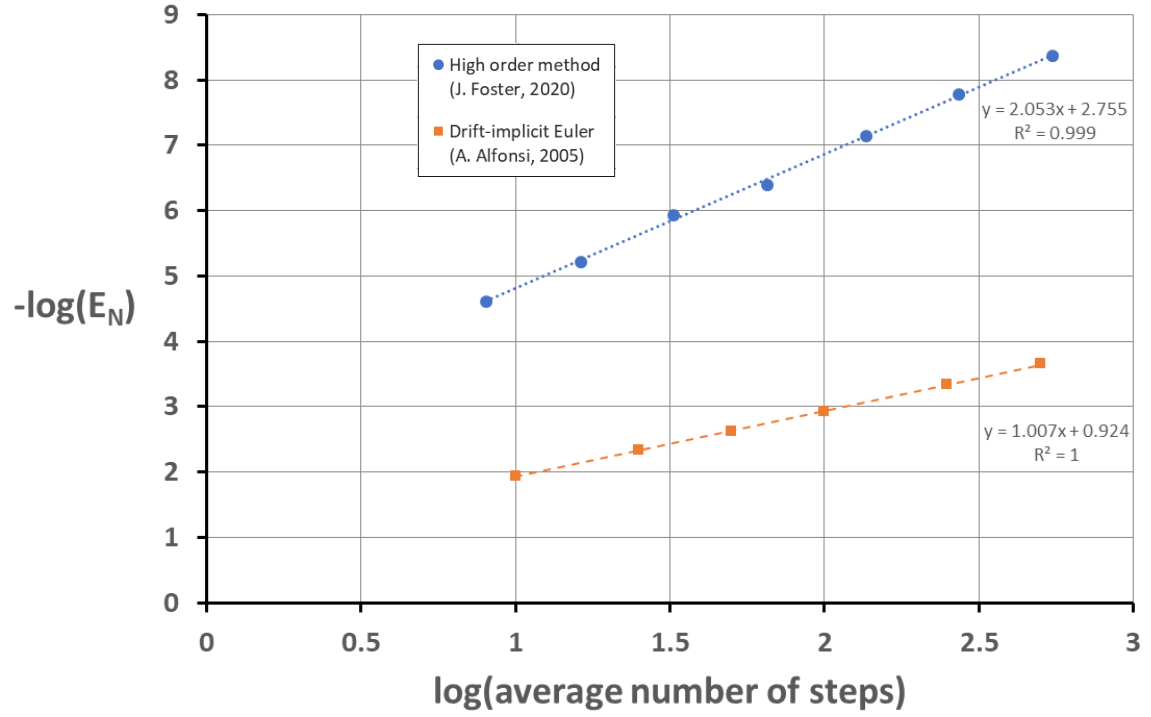

Figure 6.7: Weak convergence rates for both methods in the low volatility setting.

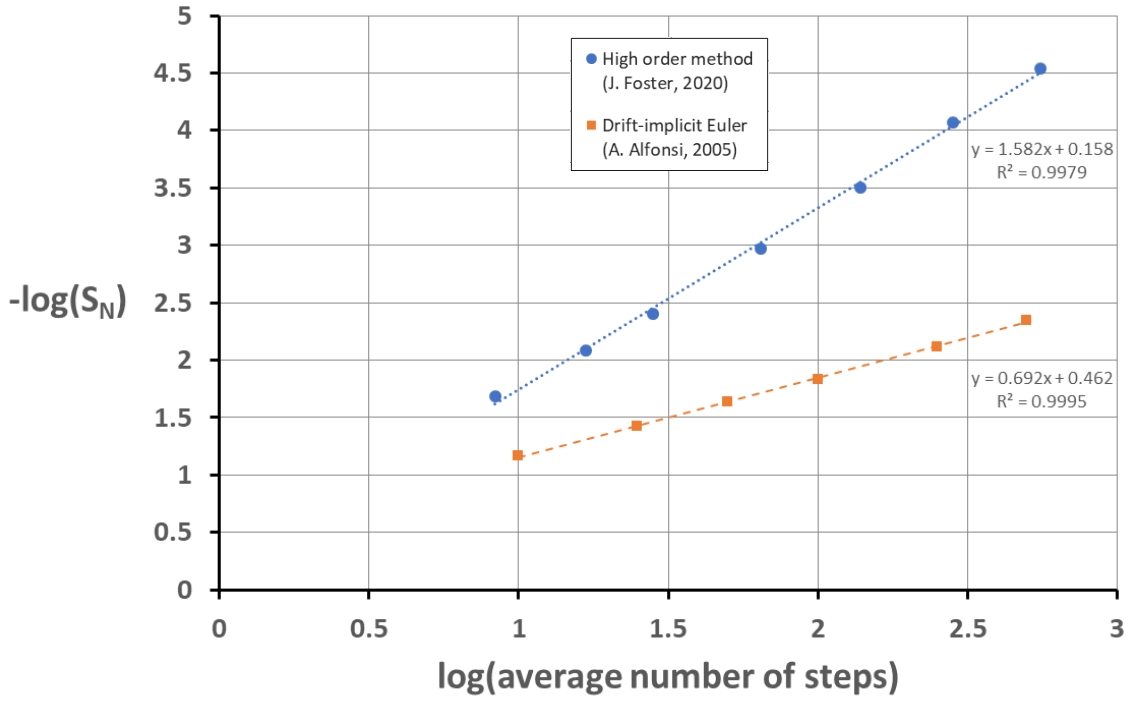

Figure 6.8: Strong convergence rates for both methods in the high volatility setting.

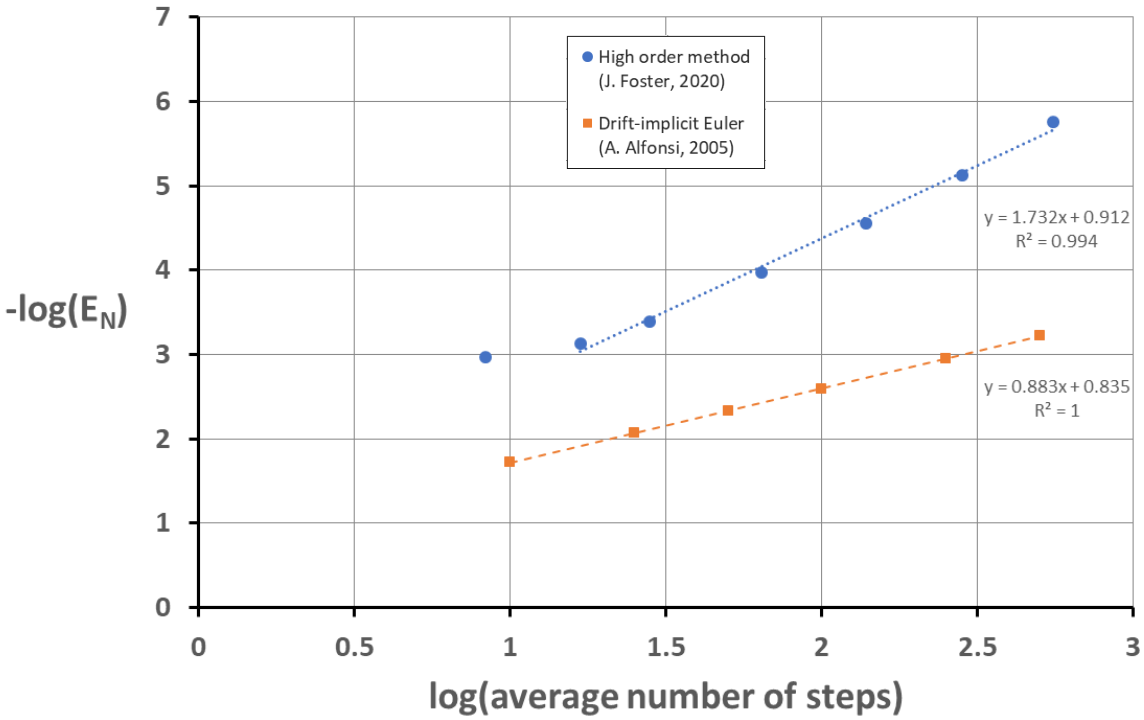

Figure 6.9: Weak convergence rates for both methods in the high volatility setting.

In each of the above graphs, we see that the proposed variable step size method achieves higher convergence rates than the (fixed step) drift-implicit Euler method. In addition, increasing the volatility σ appears to reduce the orders of convergence. This phenomenon is unsurprising as, when σ is large, the CIR process can have long excursions near zero which give numerical errors due to the non-Lipschitz vector field. In the high volatility setting, the proposed method exhibits a high convergence rate whilst the drift-implicit Euler method appears to converge with a low order (< 1.0). Nevertheless, in order to truly measure the performance of these numerical methods, we should consider the computational costs required for achieving a specified accuracy.

Table 6.2: Estimated simulation times for computing 100,000 sample paths that achieve a given accuracy using a single-threaded C++ program on a laptop computer.

	Estimated time to achieve an accuracy of		High/Low
	$S_N = 10^{-3}$ (s)	$E_N = 10^{-4}$ (s)	volatility
High order method	1.0	0.5	Low
(variable step size)	7.4	7.2	High
Drift-implicit Euler	2.9	11.6	Low
(fixed step size)	47.0	38.8	High

The above times are estimated using figures 6.6 – 6.9 along with the following table:

Table 6.3: Simulation times for computing 100,000 sample paths with an average of 100 steps per path using a single-threaded C++ program on a laptop.

	Computational time (s)	Average number of steps	High/Low volatility
High order method	11.76	100.1	Low
(variable step size)	11.89	100.2	High
Drift-implicit Euler	1.02	100	Low
(fixed step size)	1.01	100	High

Table 6.3 shows that the proposed method is an entire order of magnitude slower than the drift-implicit Euler method when the average number of steps used is 100. Our numerical evidence suggests that this difference in simulation time does not change significantly when the average number of steps used is less than 1000 (with 1000 steps, our method is about 12.5 times more expensive than the Euler method). It should also be noted that with a constant step sizes, the ODE method is only four times slower than the Euler method. Therefore, evaluating the condition (6.3.19) and refining the triples $(W_{s,t}, H_{s,t}, n_{s,t})$ has a noticeable impact on the simulation time.

On the other hand, the proposed method appears to be an order of magnitude more accurate than the Euler method (even with a small average number of steps). This is demonstrated in Table 6.2, which shows that the variable step method can obtain the same accuracy as the drift-implicit Euler method in a fraction of the time. For example, our method is between five and six times more efficient in the high volatility setting. For smaller error tolerances, we would expect the proposed method to substantially outperform the Euler method (since it appears to be higher order).

At this point, it is natural to question which part of the proposed method has more importance: the high order piecewise linear ODE method? or the variable step sizes? If the volatility is sufficiently small, then the high order ODE method should have the greater impact and we would conjecture that the resulting performance increase would be comparable to that demonstrated in the IGBM example (see section 4.5). On the other hand, when the volatility is high, the CIR process is more likely to have excursions away from the mean reversion level b (and, in particular, close to zero). Since the derivatives of $\sqrt{\cdot}$ blow-up near zero, the proposed high order ODE method would only be accurate when the step size used is sufficiently small. Moreover, we expect that Theorem 6.3.1 can be extended to methods which use integral information in addition to Brownian increments (and thus quantify the limitations of fixed steps).

So when the volatility is sufficiently high, the variable step sizes should have the greater impact. Of course, the use of a high order method is still likely to be beneficial. In particular, for excursions away from zero, the derivatives of $\sqrt{\cdot}$ can be bounded and thus we expect the proposed method to outperform the drift-implicit Euler method just as the log-ODE method outperformed Milstein's method for IGBM (section 4.5). Another benefit of employing a high order ODE method is that it is possible to estimate the local $L^2(\mathbb{P})$ error of each one-step approximation, which naturally leads to the variable step size strategy given by definition 6.0.1. If a lower order method were used, then it is possible that another criterion should be used instead of (6.3.19).

Whilst it would be interesting to see how the importance of the variable step sizes depends on the Feller ratio $\frac{2ab}{\sigma^2}$, a more pressing issue is simply proving convergence. In addition, it is not known whether the high convergence rates observed in the numerical experiments are sensible. Since each piecewise linear ODE produces a local error that is $O(h^2)$ in $L^2(\mathbb{P})$ and the noise vector field is not globally Lipschitz, it is a little surprising that the method could exhibit an $O(N^{-1.5})$ strong convergence rate (Recall that we would expect $O(h^{1.5})$ convergence for SDEs with "nice" vector fields). Thus, we wish to better understand how variable step sizes impact convergence rates.

In conclusion, there are primarily three key results that are presented in this chapter:

- **Generating Brownian increments and integrals on dyadic partitions** To develop variable step size methods for SDEs, it can be necessary to generate the Brownian path within an interval conditional on its endpoint information. Inspired by Lévy's construction of Brownian motion and the paper [13], we give algorithms that generate the triples $(W, H, n)_{s,t}$ and $(W, H, K)_{s,t}$ in this manner.
- **Proving convergence of variable step size methods using rough paths** The standard error analysis of SDE numerical methods requires deterministic step sizes and so variable step size methodologies can be challenging to study. Using rough path theory, we have established sufficient conditions which ensure a variable step size method converges to the SDE solution (in a pathwise sense). In particular, this result can be applied to both the Heun and midpoint methods.
- **Discretization of the Cox-Ingersoll-Ross model (and Bessel process)** It can be difficult to accurately simulate the CIR process as it is non-negative and governed by a square root vector field with derivatives that blow-up at zero. By incorporating a high order piecewise linear ODE method from Chapter 5 into a variable step size framework, we derive a new scheme for the CIR model. Our numerical experiments suggest this scheme has high orders of convergence.

Furthermore, the research discussed in this chapter raises some interesting questions:

- Can we extend Lévy's construction to generate increments and n time integrals?
- What improvements could be made to the step size control in definition 6.0.1?
- Given an explicit variable step size strategy (using a refining decision function), is it possible to strengthen Theorem 6.2.17 and quantify the rate of convergence?
- How might the log-ODE discretization of IGBM be improved by variable steps?
- Can we incorporate the ideas within this chapter into Multilevel Monte Carlo?
- How should one design variable step size strategies that do not use derivatives?

  For example, we could estimate local errors by comparing the "one-step" and "two-step" approximations. This quantity would have the form $Y_{s,t} - Y_{s,u} - Y_{u,t}$ and so may fit into the Stochastic Sewing Lemma framework developed in [65].
- Could we prove the ODE approximation (6.3.5) converges to the CIR process?

Chapter 7

Approximation and conditional moments of Brownian Lévy area

Up to this point, the majority of the thesis has focused on developing high order numerical approximations for SDEs driven by a one-dimensional Brownian motion. In this setting, most terms within the stochastic Taylor expansion (4.3.2) can be generated exactly. However, in order to develop high order numerical methodologies, it was necessary to first consider approximations of the "space-space-time" Lévy area:

$$L_{s,t} := \frac{1}{6}\Bigg(\int_s^t \int_s^u \int_s^v \circ dW_r \circ dW_v \, du - 2 \int_s^t \int_s^u \int_s^v \circ dW_r \, dv \circ dW_u + \int_s^t \int_s^u \int_s^v dr \circ dW_v \circ dW_u \Bigg).$$

In Chapters 4 and 5, we approximated $L_{s,t}$ using its conditional expectation $\mathbb{E}\big[L_{s,t} | W_{s,t}, H_{s,t}, \cdot\big]$, where $H_{s,t}$ is the space-time Lévy area of the path W over $[s,t]$.

As one might expect, the numerical analysis of SDEs driven by multidimensional Brownian motion is significantly more challenging than for the one-dimensional SDEs. In particular, we have the following classic result due to Cameron and Clark in [21].

Theorem 7.0.1 (An example of a slow best rate)**.** *Consider the following Itô SDE:*

$$\begin{aligned} dx_t &= y_t \, dW_t^{(1)}, \\ dy_t &= dW_t^{(2)}, \end{aligned} \tag{7.0.1}$$

where $W^{(1)}$ and $W^{(2)}$ are two independent Brownian motions and $x_0 = 0, y_0 = 0$. For $T > 0$ and $n \geq 1$, let P_n be the σ-algebra generated by $\left\{ W^{(1)}_{\frac{kT}{n}}, W^{(2)}_{\frac{kT}{n}} \right\}_{0 \leq k \leq n}$. Then

$$\big\| x_T - \mathbb{E}\big[x_T | P_n\big] \big\|_{L^2(\mathbb{P})} = \frac{T}{2} \left(\frac{1}{n} \right)^{\frac{1}{2}}.$$

Remark 7.0.2. This result implies that any P_n-measurable approximation of (7.0.1) cannot achieve a strong convergence rate better than $O\big(\sqrt{h}\big)$, where h is the step size.

Remark 7.0.3. It was shown by Dickinson in [30] that any approximation of x_T that is measurable which respect to n Gaussian random variables (which are expressible as linear functionals on W) cannot have a strong convergence rate better than $O\big(n^{-\frac{1}{2}}\big)$.

More generally, we would not expect better convergence rates when approximating multidimensional SDEs that are governed by non-commutative vector fields. That is,

$$\begin{aligned} dy_t &= f_0(y_t)\,dt + \sum_{i=1}^{d} f_i(y_t)\,dW_t^{(i)}, \\ y_0 &= \xi, \end{aligned} \tag{7.0.2}$$

where $\xi \in \mathbb{R}^e$, the vector fields $\{f_i\}$ do not satisfy the below commutativity condition:

$$[f_i, f_j] = 0, \quad \forall i, j \in \{1, \cdots, d\}. \tag{7.0.3}$$

When the condition (7.0.3) is not satisfied, then certain iterated integrals of W (Brownian Lévy area) in the Taylor expansion of (7.0.2) result in local errors of $O(h)$. If these errors propagate like a martingale, we obtain the $O\big(\sqrt{h}\big)$ convergence rate. Thus, we will consider the following "low order" stochastic Taylor expansion of (7.0.2). (see chapter 5 of [62] for a comprehensive review that presents the below expansion)

Theorem 7.0.4 (Low order Itô-Taylor expansion)**.** *Let y denote the unique strong solution to (7.0.2) where the vector fields $\{f_i\}$ are assumed to be smooth and bounded with bounded derivatives. Then for $0 \le s \le t$, y_t can be Taylor expanded as follows:*

$$\begin{aligned} y_t = y_s + f_0(y_t)\,h + \sum_{i=1}^{d} f_i(y_t)\,W_{s,t}^{(i)} + \sum_{i,j=1}^{d} f_j'(y_t) f_i(y_t) \int_s^t \int_s^u dW_v^{(i)}\,dW_u^{(j)} \\ + R_2(h, y_s), \end{aligned} \tag{7.0.4}$$

where $h := t - s$ and the remainder term R_2 has the below uniform estimate for $h < 1$,

$$\sup_{y_s \in \mathbb{R}^e} \|R_2(h, y_s)\|_{L^2(\mathbb{P})} \le C_1\, h^{\frac{3}{2}}, \tag{7.0.5}$$

and the constants C_1, C_2 depend only on the vector fields of the differential equation.

Remark 7.0.5. Each iterated integral in (7.0.4) will satisfy an SDE of the form (7.0.1), and may be hard to approximate by the results of Cameron & Clark and Dickinson.

Remark 7.0.6. Milstein's method can be obtained precisely as the first line of (7.0.4).

So when the commutativity condition (7.0.3) is not satisfied, in order to exactly generate the first line of this Taylor expansion, we require the below stochastic area.

Definition 7.0.7. The **Lévy area** of a d-dimensional Brownian motion over $[s,t]$ is

$$A_{s,t}^{(i,j)} := \frac{1}{2}\left(\int_s^t W_{s,u}^{(i)}\, dW_u^{(j)} - \int_s^t W_{s,u}^{(j)}\, dW_u^{(i)} \right), \tag{7.0.6}$$

for $i,j \in \{1,\cdots,d\}$.

Remark 7.0.8. $A_{s,t}$ can be viewed as an antisymmetric matrix since $A_{s,t}^{(i,j)} = -A_{s,t}^{(j,i)}$.

Remark 7.0.9. By integration by parts, we have $\displaystyle\int_s^t W_{s,u}^{(i)} \circ dW_u^{(j)} = \frac{1}{2} W_{s,t}^{(i)} W_{s,t}^{(j)} + A_{s,t}^{(i,j)}$.

Despite its non-trivial joint distribution, a method for generating $\left(W_{s,t}^{(1)}, W_{s,t}^{(j)}, A_{s,t}^{(1,2)}\right)$ was proposed in [38]. That said, this approach was not extended to higher dimensions. Designing methods that exactly generate $\left(W_{s,t}, A_{s,t}\right)$ is an open problem when $d > 2$.

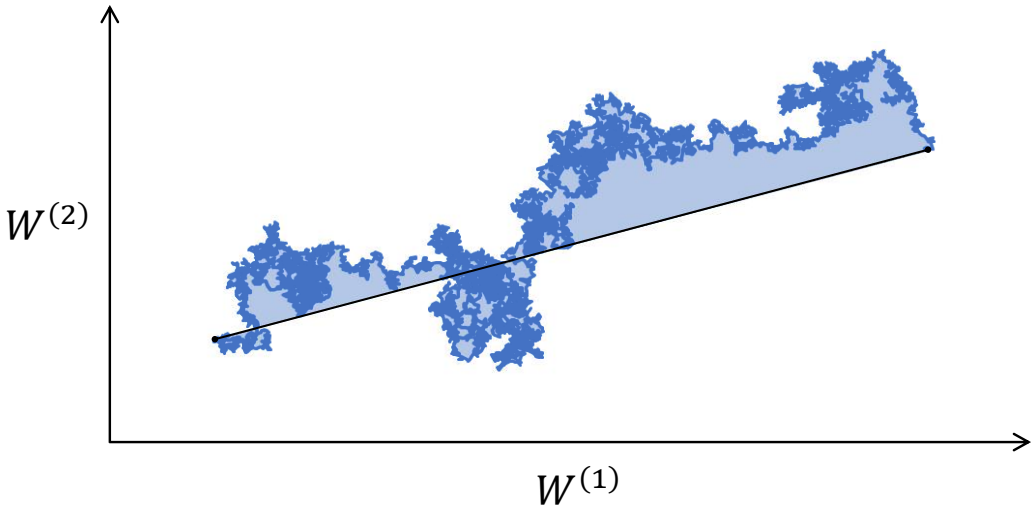

Figure 7.1: Lévy area is the chordal area between independent Brownian motions.

In this chapter, we will study the Lévy area of multidimensional Brownian motion. Just as for $L_{s,t}$, we shall compute moments of $A_{s,t}$ conditional on the pair $(W_{s,t}, H_{s,t})$. The motivation for investigating the relationship between $H_{s,t}$ and $A_{s,t}$ comes from the research of Davie [27] and Flint & Lyons [34] who proposed the below approximation:

$$\widehat{A}_{s,t}^{(i,j)} := H_{s,t}^{(i)} W_{s,t}^{(j)} - W_{s,t}^{(i)} H_{s,t}^{(j)} + \widehat{b}_{s,t}^{(i,j)}, \tag{7.0.7}$$

where $\left\{ \widehat{b}_{s,t}^{(i,j)} : 1 \le i < j \le d \right\}$ is a collection of $\frac{1}{2}d(d-1)$ independent and identically distributed random variables with $\widehat{b}_{s,t}^{(i,j)} \sim \mathcal{N}\left(0, \frac{1}{12}h^2\right)$ and $\widehat{b}_{s,t}^{(j,i)} = -\widehat{b}_{s,t}^{(i,j)}$ for all i,j. In particular, it was shown that discretizing Brownian Lévy area using the random variables $\left\{\widehat{A}_{t_k,t_{k+1}}\right\}$ will converge (in the 2-Wasserstein sense) with an order of $O(h)$.

One advantage of using the 2-Wasserstein metric (definition 7.3.9) is that it is now possible for methods without Lévy area to converge faster than $O\big(\sqrt{h}\big)$ for general SDEs. At the same time, the 2-Wasserstein error of an approximation can be viewed as an $L^2(\mathbb{P})$ error (specifically between random variables that are optimally coupled). The proof of this $O(h)$ convergence requires Komlós-Major-Tusnády (or KMT) theory but ultimately relies on the approximation $\widehat{A}_{s,t}$ having the same covariance structure as true Lévy area once the Brownian motion's increment $W_{s,t}$ has been generated. More concretely, the following theorem is used in the proof of Proposition 6.2 in [34].

Theorem 7.0.10 (Matching conditional moments of multidimensional Lévy area)**.**

$$\mathbb{E}\Big[\widehat{A}_{s,t}^{(i,j)}\,\Big|\,W_{s,t}\Big] = \mathbb{E}\Big[A_{s,t}^{(i,j)}\,\Big|\,W_{s,t}\Big]\,, \tag{7.0.8}$$

$$\mathbb{E}\Big[\widehat{A}_{s,t}^{(i,j)}\widehat{A}_{s,t}^{(k,l)}\,\Big|\,W_{s,t}\Big] = \mathbb{E}\Big[A_{s,t}^{(i,j)}A_{s,t}^{(k,l)}\,\Big|\,W_{s,t}\Big]\,, \tag{7.0.9}$$

for $i,j,k,l \in \{1,\cdots,d\}$.

Remark 7.0.11. The conditional moments (7.0.8) and (7.0.9) are all zero except for

$$\mathbb{E}\Big[\Big(A_{s,t}^{(i,j)}\Big)^2\,\Big|\,W_{s,t}\Big] = \frac{1}{12}h^2 + \frac{1}{12}h\left(\Big(W_{s,t}^{(i)}\Big)^2 + \Big(W_{s,t}^{(j)}\Big)^2\right). \tag{7.0.10}$$

Since the space-time area $H_{s,t}$ must be generated to compute $\widehat{A}_{s,t}$, it is natural to consider approximations of Lévy area that match $(W_{s,t}, H_{s,t})$-conditional moments:

Theorem 7.0.12 (Non-trivial conditional moments of two-dimensional Lévy area)**.**

$$\mathbb{E}\Big[A_{s,t}^{(i,j)}\,\Big|\,W_{s,t}, H_{s,t}\Big] = H_{s,t}^{(i)}\,W_{s,t}^{(j)} - W_{s,t}^{(i)}\,H_{s,t}^{(j)}\,, \tag{7.0.11}$$

$$\mathrm{Var}\Big(A_{s,t}^{(i,j)}\,\Big|\,W_{s,t}, H_{s,t}\Big) = \frac{1}{20}h^2 + \frac{1}{5}h\left(\Big(H_{s,t}^{(i)}\Big)^2 + \Big(H_{s,t}^{(j)}\Big)^2\right), \tag{7.0.12}$$

for $i,j \in \{1,\cdots,d\}$.

The above conditional moments will be derived in the first section of this chapter. As one might expect, the proof of this result uses the decomposition $W = \widehat{W} + Z$ where $\widehat{W}$ is a Brownian parabola and Z is a Brownian arch (see Theorem 4.1.12). Although the expectation (7.0.11) is already matched by the random variable $\widehat{A}_{s,t}$, the approximation does not capture the non-trivial conditional variance given by (7.0.12). Hence a potential improvement could be made to $\widehat{A}_{s,t}$ by rescaling each $\widehat{b}_{s,t}^{(i,j)}$ so that

$$\Big(\widehat{b}_{s,t}^{(i,j)}\Big|W_{s,t}, H_{s,t}\Big) \sim \mathcal{N}\left(0, \frac{1}{20}h^2 + \frac{1}{5}h\left(\Big(H_{s,t}^{(i)}\Big)^2 + \Big(H_{s,t}^{(j)}\Big)^2\right)\right).$$

In the last section of this chapter, we will show that this "change of variance" does not impact the $O(h)$ 2-Wasserstein convergence rate of the Lévy area approximation. It is still ongoing research to apply these weak approximations to real-world problems.

This chapter is outlined as follows. In the first section, we shall compute the first four conditional moments of two-dimensional Lévy area using the theorems below.

Theorem 7.0.13 (Decompositions of Lévy area using polynomial approximations)**.** *Let $H_{s,t}$ and $K_{s,t}$ denote the Brownian space-time areas defined in Chapter 4. Then*

$$A_{s,t} = H_{s,t} \otimes W_{s,t} - W_{s,t} \otimes H_{s,t} + b_{s,t}, \tag{7.0.13}$$

$$A_{s,t} = H_{s,t} \otimes W_{s,t} - W_{s,t} \otimes H_{s,t} + 12\big(K_{s,t} \otimes H_{s,t} - H_{s,t} \otimes K_{s,t}\big) + a_{s,t}, \tag{7.0.14}$$

where $b_{s,t}$ is the Lévy area of the Brownian bridge B on $[s,t]$ that corresponds to W,

$$b_{s,t}^{(i,j)} := \int_s^t B_{s,u}^{(i)} \circ dB_u^{(j)}, \tag{7.0.15}$$

and $a_{s,t}$ denotes the Lévy area of the Brownian arch Z on $[s,t]$ corresponding to W,

$$a_{s,t}^{(i,j)} := \int_s^t Z_{s,u}^{(i)} \circ dZ_u^{(j)}, \tag{7.0.16}$$

for $i,j \in \{1,\cdots,d\}$.

Proof. By the natural Brownian scaling, it is enough to prove the result on $[0,1]$. Suppose that $i,j \in \{1,\cdots,d\}$. Then by the decomposition $W_t = tW_1 + B_t$, we have

$$\begin{aligned}
A_{0,1}^{(i,j)} &= \int_0^1 W_t^{(i)} \circ dW_t^{(j)} - \frac{1}{2} W_1^{(i)} W_1^{(j)} \\
&= \int_0^1 \Big(tW_1^{(i)} + B_t^{(i)}\Big) \circ d\Big(tW_1^{(j)} + B_t^{(j)}\Big) - \frac{1}{2} W_1^{(i)} W_1^{(j)} \\
&= \frac{1}{2} W_1^{(i)} W_1^{(j)} + W_1^{(j)} \int_0^1 B_t^{(i)} dt + W_1^{(i)} \int_0^1 t\, dB_t^{(j)} + \int_0^1 B_t^{(i)} \circ dB_t^{(j)} - \frac{1}{2} W_1^{(i)} W_1^{(j)} \\
&= W_1^{(j)} H_1^{(i)} - W_1^{(i)} H_1^{(j)} + b_{0,1}^{(i,j)}.
\end{aligned}$$

Similarly, the decomposition $B_t = 6t(1-t)H_1 + Z_t$ (Theorem 4.1.3) can be applied.

$$\begin{aligned}
A_{0,1}^{(i,j)} &= W_1^{(j)} H_1^{(i)} - W_1^{(i)} H_1^{(j)} + \int_0^1 B_t^{(i)} \circ dB_t^{(j)} \\
&= W_1^{(j)} H_1^{(i)} - W_1^{(i)} H_1^{(j)} + \int_0^1 \Big(6t(1-t)H_1^{(i)} + Z_t^{(i)}\Big) \circ d\Big(6t(1-t)H_1^{(j)} + Z_t^{(j)}\Big) \\
&= W_1^{(j)} H_1^{(i)} - W_1^{(i)} H_1^{(j)} + 36 H_1^{(i)} H_1^{(j)} \int_0^1 \big(t - t^2\big)\, d\big(t - t^2\big) \\
&\qquad + 6H_1^{(j)} \int_0^1 Z_t^{(i)} d\big(t - t^2\big) + 6H_1^{(i)} \int_0^1 \big(t - t^2\big)\, dZ_t^{(j)} + \int_0^1 Z_t^{(i)} \circ dZ_t^{(j)} \\
&= W_1^{(j)} H_1^{(i)} - W_1^{(i)} H_1^{(j)} - 12 H_1^{(j)} \int_0^1 Z_t^{(i)} d\left(\frac{1}{2}t^2\right) - 12 H_1^{(i)} \int_0^1 \frac{1}{2} t^2\, dZ_t^{(j)} + a_{0,1}^{(i,j)}.
\end{aligned}$$

The result now follows by applying integration by parts and then Theorem 4.2.6. □

Remark 7.0.14. The quantity $\lambda_{s,t}$ in (7.0.7) approximates the "bridge Lévy area" $b_{s,t}$.

Theorem 7.0.15. *The two-dimensional Lévy area of the Brownian bridge process has a logistic distribution with mean* 0 *and variance* $\frac{1}{12}h^2$. *So for* $i, j \in \{1, \cdots, d\}$ *with* $i \neq j$,

$$b_{s,t}^{(i,j)} \sim \text{Logistic}\left(0, \frac{1}{\pi}h\right).$$

Proof. It is known that the joint density function for $x = W_1^{(i)}$, $y = W_1^{(j)}$, $z = A_1^{(i,j)}$ is

$$f(x,y,z) = \frac{1}{2\pi^2}\int_0^\infty \frac{u}{\sinh(u)} \exp\left(-\frac{(x^2+y^2)\,u}{2\tanh(u)}\right)\cos(zu)\,du,$$

(see Lévy [68] and Lyons [38]). So when the increment W_1 is zero, the density of z is

$$f(z) = \frac{1}{2\pi^2}\int_0^\infty \frac{u}{\sinh(u)}\cos(zu)\,du. \tag{7.0.17}$$

We shall now evaluate the above integral. It is first worth noting that for all $u \geq 0$,

$$\frac{1}{\sinh(u)} = \frac{2}{e^u - e^{-u}} = 2e^{-u}\cdot\frac{1}{1-e^{-2u}} = 2e^{-u}\sum_{k=0}^\infty e^{-2ku} = 2\sum_{k=0}^\infty e^{-(2k+1)u}.$$

Since this decays exponentially fast, we can swap the sum and integral operations in

$$\begin{aligned}
\int_0^\infty \frac{\sin(zu)}{\sinh(u)}\,du &= 2\int_0^\infty \sin(zu)\sum_{k=0}^\infty e^{-(2k+1)u}\,du \\
&= 2\sum_{k=0}^\infty \int_0^\infty \sin(zu)e^{-(2k+1)u}\,du \\
&= 2\sum_{k=0}^\infty \text{Im}\left(\int_0^\infty e^{(-(2k+1)+iz)u}\,du\right) \\
&= 2\sum_{k=0}^\infty \frac{-z}{(2k+1)^2+z^2}\,.
\end{aligned} \tag{7.0.18}$$

By the Weierstrass factorization theorem, there is a product representation of $\sinh(z)$:

$$\sinh(z) = z\prod_{k=1}^\infty \left(1+\left(\frac{z}{k\pi}\right)^2\right).$$

Taking the logarithm of the above and differentiating gives a representation of coth:

$$\coth(z) = \frac{1}{z} + \sum_{k=1}^\infty \frac{2z}{(k\pi)^2+z^2}\,.$$

This has the same structure to the infinite series (7.0.18). In addition, it implies that

$$\int_0^\infty \frac{\sin(zu)}{\sinh(u)}\,du = \pi\coth(\pi z) - \frac{1}{2}\pi\coth\left(\frac{1}{2}\pi z\right).$$

Using the double angle formulae for hyperbolic trigonometric functions, we have

$$\tanh(z) + \coth(z) = \frac{\sinh(z)}{\cosh(z)} + \frac{\cosh(z)}{\sinh(z)} = \frac{\sinh^2(z) + \cosh^2(z)}{\cosh(z)\sinh(z)} = 2\coth(2z),$$

and so

$$\int_0^\infty \frac{\sin(zu)}{\sinh(u)}\, du = \frac{1}{2}\pi \tanh\left(\frac{1}{2}\pi z\right). \tag{7.0.19}$$

Differentiating both sides of (7.0.19) and applying the Leibniz integral rule gives

$$\int_0^\infty \frac{u\cos(zu)}{\sinh(u)}\, du = \frac{1}{4\pi^2}\operatorname{sech}^2\left(\frac{1}{2}\pi z\right).$$

By substituting this into (7.0.17), we see that $b_1^{(i,j)}$ has the desired distribution. The result now follows as b_1 and $hb_{s,t}$ have the same law (by Brownian scaling). □

In the second section of this chapter, we shall compute additional moments for d-dimensional Brownian Lévy area when $d > 2$ (conditional on the increment $W_{s,t}$). In this setting, we will identify all conditional moments with degree at most four. Most notably, the following two moments demonstrate the intricate structure of $A_{s,t}$.

Theorem 7.0.16. *Let $b_{s,t}$ denote the Lévy area of a Brownian bridge on $[s,t]$. Then*

$$\mathbb{E}\Big[\Big(b_{s,t}^{(i,j)}\Big)^2\Big(b_{s,t}^{(j,k)}\Big)^2\Big] = \frac{7}{720}h^4,$$
$$\mathbb{E}\Big[H_{s,t}^{(i)}H_{s,t}^{(k)}b_{s,t}^{(i,j)}b_{s,t}^{(j,k)}\Big] = -\frac{1}{720}h^3,$$

where $i,j,k \in \{1,\cdots,d\}$ are distinct.

In the third section, we will develop new approximations for Brownian Lévy area in the low dimensional cases ($d = 2, 3$) by matching these new conditional moments. In addition to matching the first four moments of Brownian Lévy area, the proposed random variables are fast to generate and straightforward to implement in practice. Furthermore, we shall demonstrate the accuracy of this approach in the $d = 2$ case by numerically estimating the 2-Wasserstein distance between the proposed "weak" approximation $\widetilde{A}$ and the true Brownian Lévy area. Numerical evidence suggests that

$$W_2\big(\mu_A, \widetilde{\mu}_A\big) \leq 0.01h.$$

In the final section, we shall compare our approximation with that of Davie's. There may also be additional applications of the research discussed in this chapter. For example, the conditional variances (7.0.10) or (7.0.12) could be used to quantify local $L^2(\mathbb{P})$ errors and develop variable step size methods for multidimensional SDEs.

7.1 Moments conditional on the path increment and space-time Lévy area

In this section, we shall compute the first four moments of two-dimensional Brownian Lévy area conditional on the increment $W_{s,t}$ and space-time Lévy area $H_{s,t}$ of the path.

Theorem 7.1.1. *Let $A_{s,t}$ denote the Lévy area of Brownian motion over $[s,t]$. Then*

$$A_{s,t}^{(i,j)} = H_{s,t}^{(i)} W_{s,t}^{(j)} - W_{s,t}^{(i)} H_{s,t}^{(j)} + b_{s,t}^{(i,j)}, \tag{7.1.1}$$

for $i,j \in \{1,\cdots,d\}$ where the Brownian bridge Lévy area $b_{s,t}$ is independent of $W_{s,t}$ and a symmetric random variable. Then if $i \neq j$, we have the conditional moments:

$$\mathbb{E}\Big[A_{s,t}^{(i,j)} \,\Big|\, W_{s,t}, H_{s,t}\Big] = H_{s,t}^{(i)} W_{s,t}^{(j)} - W_{s,t}^{(i)} H_{s,t}^{(j)}, \tag{7.1.2}$$

$$\mathrm{Var}\Big(A_{s,t}^{(i,j)} \,\Big|\, W_{s,t}, H_{s,t}\Big) = \frac{1}{20}h^2 + \frac{1}{5}h\left(\left(H_{s,t}^{(i)}\right)^2 + \left(H_{s,t}^{(j)}\right)^2\right), \tag{7.1.3}$$

$$\mathrm{Skew}\Big(A_{s,t}^{(i,j)} \,\Big|\, W_{s,t}, H_{s,t}\Big) = 0, \tag{7.1.4}$$

$$\mathrm{Kurt}\Big(A_{s,t}^{(i,j)} \,\Big|\, W_{s,t}, H_{s,t}\Big) = 3 + \frac{\frac{3}{1400}h^2 + \frac{3}{175}h\left(\left(H_{s,t}^{(i)}\right)^2 + \left(H_{s,t}^{(j)}\right)^2\right)}{\left(\frac{1}{20}h + \frac{1}{5}\left(\left(H_{s,t}^{(i)}\right)^2 + \left(H_{s,t}^{(j)}\right)^2\right)\right)^2}. \tag{7.1.5}$$

Proof. The area decomposition (7.1.1) was established as part of Theorem 7.0.13. Since $b_{s,t}$ depends only on a Brownian bridge process, it must be independent of $W_{s,t}$. Using Theorem 7.0.13, we have the following decomposition of the bridge Lévy area,

$$b_{s,t}^{(i,j)} = 12\left(K_{s,t}^{(i)} H_{s,t}^{(j)} - H_{s,t}^{(i)} K_{s,t}^{(j)}\right) + a_{s,t}^{(i,j)}, \tag{7.1.6}$$

where $K_{s,t}$ denotes the "space-time-time" Lévy area given by definition 4.2.4 and $a_{s,t}$ is the Lévy area of a Brownian arch process (and thus also independent of $H_{s,t}$). Since the Gaussian weights in the polynomial representation of the Brownian bridge (Theorem 4.1.3) are all independent with zero mean, $b_{s,t}$ and $a_{s,t}$ must be symmetric. It is enough for us to compute the moments of $b_{s,t}$ conditional on the pair $(W_{s,t}, H_{s,t})$. Since $(K_{s,t}, a_{s,t})$ is independent of $W_{s,t}$ and $H_{s,t}$, it follows directly from (7.1.6) that

$$\mathbb{E}\Big[b_{s,t}^{(i,j)} \,\Big|\, W_{s,t}, H_{s,t}\Big] = 12 H_{s,t}^{(j)}\,\mathbb{E}\Big[K_{s,t}^{(j)}\Big] - 12 H_{s,t}^{(i)}\,\mathbb{E}\Big[K_{s,t}^{(i)}\Big] + \mathbb{E}\Big[a_{s,t}^{(i,j)}\Big] = 0,$$

and

$$\begin{aligned}\mathbb{E}\Big[\left(b_{s,t}^{(i,j)}\right)^2 \,\Big|\, W_{s,t}, H_{s,t}\Big] &= 144\left(H_{s,t}^{(i)}\right)^2 \mathbb{E}\Big[\left(K_{s,t}^{(j)}\right)^2\Big] + 144\left(H_{s,t}^{(j)}\right)^2 \mathbb{E}\Big[\left(K_{s,t}^{(i)}\right)^2\Big] \\ &\quad + 24 H_{s,t}^{(j)}\,\mathbb{E}\Big[K_{s,t}^{(i)} a_{s,t}^{(i,j)}\Big] - 24 H_{s,t}^{(i)}\,\mathbb{E}\Big[K_{s,t}^{(j)} a_{s,t}^{(i,j)}\Big] \\ &\quad - 288 H_{s,t}^{(i)} H_{s,t}^{(j)}\,\mathbb{E}\Big[K_{s,t}^{(i)} K_{s,t}^{(j)}\Big] + \mathbb{E}\Big[\left(a_{s,t}^{(i,j)}\right)^2\Big].\end{aligned}$$

By reflecting the i-th coordinate of the Brownian arch, we see that $\mathbb{E}\Big[K_{s,t}^{(j)}\, a_{s,t}^{(i,j)}\Big] = 0$. Since $K_{s,t} \sim \mathcal{N}\big(0, \frac{1}{720}h\big)$, the above expression can be simplified to give the variance:

$$\mathbb{E}\Big[\Big(b_{s,t}^{(i,j)}\Big)^2 \,|\, W_{s,t}, H_{s,t}\Big] = \mathbb{E}\Big[\Big(a_{s,t}^{(i,j)}\Big)^2\Big] + \frac{1}{5}h\left(\Big(H_{s,t}^{(i)}\Big)^2 + \Big(H_{s,t}^{(j)}\Big)^2\right). \tag{7.1.7}$$

By taking the expectation of both sides in (7.1.7) and using the Tower law, we have

$$\mathbb{E}\Big[\Big(a_{s,t}^{(i,j)}\Big)^2\Big] = \mathbb{E}\Big[\Big(b_{s,t}^{(i,j)}\Big)^2\Big] - \frac{1}{30}h^2 = \frac{1}{20}h^2, \tag{7.1.8}$$

by Theorem 7.0.15 $\Big($where it was shown that $\mathbb{E}\Big[\Big(b_{s,t}^{(i,j)}\Big)^2\Big] = \frac{1}{12}h^2\Big)$. Thus, we have

$$\mathbb{E}\Big[\Big(b_{s,t}^{(i,j)}\Big)^2 \,|\, W_{s,t}, H_{s,t}\Big] = \frac{1}{20}h^2 + \frac{1}{5}h\left(\Big(H_{s,t}^{(i)}\Big)^2 + \Big(H_{s,t}^{(j)}\Big)^2\right). \tag{7.1.9}$$

Hence the first two conditional moments, (7.1.2) and (7.1.3), have been established. To compute the skewness, we will first show that the following moments are all zero.

$$\begin{aligned}
\mathbb{E}\Big[\Big(a_{s,t}^{(i,j)}\Big)^3\Big] &= 0, \quad \Big(\text{by the symmetry of } a_{s,t}^{(i,j)}\Big), \\
\mathbb{E}\Big[K_{s,t}^{(i)}\Big(a_{s,t}^{(i,j)}\Big)^2\Big] &= 0, \quad \Big(\text{by the symmetry of } Z^{(i)}\Big), \\
\mathbb{E}\Big[\Big(K_{s,t}^{(i)}\Big)^2 a_{s,t}^{(i,j)}\Big] &= 0, \quad \Big(\text{by the symmetry of } Z^{(j)}\Big), \\
\mathbb{E}\Big[K_{s,t}^{(i)}\, K_{s,t}^{(j)}\, a_{s,t}^{(i,j)}\Big] &= 0, \quad \Big(\text{since } a_{s,t}^{(i,j)} = -a_{s,t}^{(j,i)}\Big).
\end{aligned}$$

Therefore expanding $\Big(b_{s,t}^{(i,j)}\Big)^3$ using (7.1.6) and taking a conditional expectation gives

$$\mathbb{E}\Big[\Big(b_{s,t}^{(i,j)}\Big)^3 \,|\, W_{s,t}, H_{s,t}\Big] = 0. \tag{7.1.10}$$

Recall that by Theorem 4.1.3, we can represent the Brownian arch Z on $[0,1]$ using a family of orthogonal polynomials $\{e_k\}_{k\geq 2}$ and independent Gaussian weights $\{I_k\}_{k\geq 2}$.

$$Z = \sum_{k=2}^{\infty} I_k e_k = I_2 e_2 + \widetilde{Z},$$

where $\widetilde{Z}$ denotes the "third degree" Brownian bridge (see definition 4.1.16) given by

$$\widetilde{Z} = \sum_{k=3}^{\infty} I_k e_k.$$

Then the Lévy area $a_{0,1}$ of the Brownain arch Z can be further decomposed as follows,

$$\begin{aligned}
a_{0,1}^{(i,j)} &= \int_0^1 Z_t^{(i)} \circ dZ_t^{(j)} \\
&= I_2^{(i)} I_2^{(j)} \int_0^1 e_2(t)\, de_2(t) + I_2^{(i)} \int_0^1 e_2(t)\, d\widetilde{Z}_t^{(j)} + I_2^{(j)} \int_0^1 \widetilde{Z}_t^{(i)} de_2(t) + \int_0^1 \widetilde{Z}_t^{(i)} \circ d\widetilde{Z}_t^{(j)}.
\end{aligned}$$

Therefore, by the orthogonality of the Jacobi-like polynomials $\{e_k\}$, this simplifies to

$$a_{0,1}^{(i,j)} = I_2^{(i)} I_3^{(j)} \int_0^1 e_2(t)\, de_3(t) + I_2^{(j)} I_3^{(i)} \int_0^1 e_3(t)\, de_2(t) + c_{0,1}^{(i,j)}, \tag{7.1.11}$$

where $c_{s,t}$ is the Lévy area of a third degree Brownian bridge over the interval $[s,t]$,

$$c_{s,t}^{(i,j)} := \int_s^t \widetilde{Z}_{s,u}^{(i)} \circ d\widetilde{Z}_u^{(j)}. \tag{7.1.12}$$

Recall that the polynomials $\{e_k\}$ can be constructed using Theorems 4.1.9 and 4.1.10. In particular, it is straightforward to explicitly compute the polynomials e_2 and e_3.

$$\begin{aligned} e_2(t) &= \sqrt{30}\, t(t-1)(2t-1), \\ e_3(t) &= \frac{5\sqrt{21}}{2}\, t(t-1)\Big((2t-1)^2 - \frac{1}{5}\Big), \end{aligned}$$

Thus, the integral of e_3 against e_2 can be directly computed using elementary calculus.

$$\int_0^1 e_3(t)\, de_2(t) = \int_0^1 e_3(t) \cdot \sqrt{30}\big(6t^2 - 6t + 1\big)\, dt = -3\sqrt{\frac{2}{35}}\,.$$

Furthermore, it follows from definition 4.2.3 that $K_{0,1} = \frac{1}{2\sqrt{30}} I_2$. So (7.1.11) becomes

$$a_{0,1}^{(i,j)} = 12\sqrt{\frac{3}{7}}\left(K_{0,1}^{(i)} I_3^{(j)} - K_{0,1}^{(j)} I_3^{(i)}\right) + c_{0,1}^{(i,j)}, \tag{7.1.13}$$

where $I_3 \sim \mathcal{N}\big(0, \frac{1}{12}\big)$ and $c_{0,1}^{(i,j)}$ are both independent of the space-time-time area $K_{0,1}$. Taking the squared $L^2(\mathbb{P})$ norm of (7.1.13) and simplifying the terms as before gives

$$\begin{aligned} \mathbb{E}\Big[\Big(a_{0,1}^{(i,j)}\Big)^2\Big] &= \frac{432}{7}\,\mathbb{E}\Big[\Big(K_{0,1}^{(i)} I_3^{(j)} - K_{0,1}^{(j)} I_3^{(i)}\Big)^2\Big] + \mathbb{E}\Big[\Big(c_{0,1}^{(i,j)}\Big)^2\Big] \\ &\quad + 24\sqrt{\frac{3}{7}}\,\mathbb{E}\Big[\Big(K_{0,1}^{(i)} I_3^{(j)} - K_{0,1}^{(j)} I_3^{(i)}\Big)\Big(c_{0,1}^{(i,j)}\Big)\Big] \\ &= \frac{1}{70} + \mathbb{E}\Big[\Big(c_{0,1}^{(i,j)}\Big)^2\Big]. \end{aligned}$$

The squared $L^2(\mathbb{P})$ norm of $a_{s,t}^{(i,j)}$ is given by (7.1.8). So by Brownian scaling, we have

$$\mathbb{E}\Big[\Big(c_{s,t}^{(i,j)}\Big)^2\Big] = \frac{1}{28} h^2. \tag{7.1.14}$$

Using (7.1.13) and (7.1.14), we shall calculate the following (fourth degree) moment:

$$\mathbb{E}\Big[\Big(K_{0,1}^{(i)}\Big)^2 \Big(a_{0,1}^{(i,j)}\Big)^2\Big] = \mathbb{E}\left[\Big(K_{0,1}^{(i)}\Big)^2 \left(12\sqrt{\frac{3}{7}}\left(K_{0,1}^{(i)} I_3^{(j)} - K_{0,1}^{(j)} I_3^{(i)}\right) + c_{0,1}^{(i,j)}\right)^2\right].$$

Expanding the RHS and evaluating the resulting moments (using independence) gives

$$\begin{aligned}
\mathbb{E}\Big[\big(K_{0,1}^{(i)}\big)^2\big(a_{0,1}^{(i,j)}\big)^2\Big] &= \frac{432}{7}\,\mathbb{E}\Big[\big(K_{0,1}^{(i)}\big)^2\big(K_{0,1}^{(i)}I_3^{(j)} - K_{0,1}^{(j)}I_3^{(i)}\big)^2\Big] + \mathbb{E}\Big[\big(K_{0,1}^{(i)}\big)^2\big(c_{0,1}^{(i,j)}\big)^2\Big] \\
&\quad + 24\sqrt{\frac{3}{7}}\,\mathbb{E}\Big[\big(K_{0,1}^{(i)}\big)^2\big(K_{0,1}^{(i)}I_3^{(j)} - K_{0,1}^{(j)}I_3^{(i)}\big)\big(c_{0,1}^{(i,j)}\big)\Big] \\
&= \frac{432}{7}\,\mathbb{E}\Big[\big(K_{0,1}^{(i)}\big)^4\big(I_3^{(j)}\big)^2\Big] + \frac{432}{7}\,\mathbb{E}\Big[\big(K_{0,1}^{(i)}\big)^2\big(K_{0,1}^{(j)}\big)^2\big(I_3^{(i)}\big)^2\Big] \\
&\quad + \mathbb{E}\Big[\big(K_{0,1}^{(i)}\big)^2\Big]\,\mathbb{E}\Big[\big(c_{0,1}^{(i,j)}\big)^2\Big] \\
&= \frac{432}{7}\times 4\times\left(\frac{1}{720}\right)^2\times\frac{1}{12} + \frac{432}{7}\times\left(\frac{1}{720}\right)^2\times\frac{1}{12} + \frac{1}{720}\times\frac{1}{28} \\
&= \frac{1}{11200}\,.
\end{aligned}$$

So by the natural scaling of Brownian motion, it follows that

$$\mathbb{E}\Big[\big(K_{s,t}^{(i)}\big)^2\big(a_{s,t}^{(i,j)}\big)^2\Big] = \frac{1}{11200}h^3. \tag{7.1.15}$$

By reflecting the j-th coordinate of the Brownian arch process Z, we can deduce that

$$\begin{aligned}
\mathbb{E}\Big[K_{s,t}^{(i)}\big(a_{s,t}^{(i,j)}\big)^3\Big] &= 0, \\
\mathbb{E}\Big[K_{s,t}^{(i)}\,K_{s,t}^{(j)}\big(a_{s,t}^{(i,j)}\big)^2\Big] &= 0, \\
\mathbb{E}\Big[\big(K_{s,t}^{(i)}\big)^3 a_{s,t}^{(i,j)}\Big] &= 0, \\
\mathbb{E}\Big[K_{s,t}^{(i)}\big(K_{s,t}^{(j)}\big)^2 a_{s,t}^{(i,j)}\Big] &= 0.
\end{aligned}$$

Using the above moments and the bridge Lévy area decomposition (7.1.6), we have

$$\begin{aligned}
\mathbb{E}\Big[\big(b_{s,t}^{(i,j)}\big)^4\,\big|\,W_{s,t},H_{s,t}\Big] &= 12^4\big(H_{s,t}^{(j)}\big)^4\,\mathbb{E}\Big[\big(K_{s,t}^{(i)}\big)^4\Big] + 12^4\big(H_{s,t}^{(i)}\big)^4\,\mathbb{E}\Big[\big(K_{s,t}^{(j)}\big)^4\Big] \\
&\quad + 6\times 12^4\big(H_{s,t}^{(i)}\big)^2\big(H_{s,t}^{(j)}\big)^2\,\mathbb{E}\Big[\big(K_{s,t}^{(i)}\big)^2\big(K_{s,t}^{(j)}\big)^2\Big] \\
&\quad + 6\times 144\big(H_{s,t}^{(j)}\big)^2\,\mathbb{E}\Big[\big(K_{s,t}^{(i)}\big)^2\big(a_{s,t}^{(i,j)}\big)^2\Big] \\
&\quad + 6\times 144\big(H_{s,t}^{(i)}\big)^2\,\mathbb{E}\Big[\big(K_{s,t}^{(j)}\big)^2\big(a_{s,t}^{(i,j)}\big)^2\Big] + \mathbb{E}\Big[\big(a_{s,t}^{(i,j)}\big)^4\Big] \\
&= \frac{3}{25}h^2\Big(\big(H_{s,t}^{(i)}\big)^4 + \big(H_{s,t}^{(j)}\big)^4\Big) + \frac{6}{25}h^2\big(H_{s,t}^{(i)}\big)^2\big(H_{s,t}^{(j)}\big)^2 \\
&\quad + \frac{27}{350}h^3\Big(\big(H_{s,t}^{(i)}\big)^2 + \big(H_{s,t}^{(j)}\big)^2\Big) + \mathbb{E}\Big[\big(a_{s,t}^{(i,j)}\big)^4\Big].
\end{aligned}$$

Therefore, all that remains is to compute the final term in the above calculation. By Theorem 7.0.15, $b_{s,t}^{(i,j)}$ has a logistic distribution with mean 0 and variance $\frac{1}{12}h^2$. As the kurtosis of the logistic distribution is known to be $\frac{21}{5}$, it directly follows that

$$\mathbb{E}\left[\left(b_{s,t}^{(i,j)}\right)^4\right] = \frac{7}{240}h^4. \tag{7.1.16}$$

By taking expectations in the previous expression and using the Tower law, we have

$$\mathbb{E}\left[\left(a_{s,t}^{(i,j)}\right)^4\right] = \mathbb{E}\left[\left(b_{s,t}^{(i,j)}\right)^4\right] - \frac{41}{2100}h^4 = \frac{27}{2800}h^4. \tag{7.1.17}$$

Substituting this into the expression for the fourth conditional moment of $a_{s,t}^{(i,j)}$ gives

$$\begin{aligned}&\mathbb{E}\left[\left(b_{s,t}^{(i,j)}\right)^4 \middle| W_{s,t}, H_{s,t}\right] \\ &\quad = \frac{27}{2800}h^4 + \frac{27}{350}h^3\left(\left(H_{s,t}^{(i)}\right)^2 + \left(H_{s,t}^{(j)}\right)^2\right) + \frac{3}{25}h^2\left(\left(H_{s,t}^{(i)}\right)^2 + \left(H_{s,t}^{(j)}\right)^2\right)^2.\end{aligned} \tag{7.1.18}$$

Dividing this by the squared conditional variance of $b_{s,t}^{(i,j)}$ yields the required kurtosis:

$$\mathrm{Kurt}\left(b_{s,t}^{(i,j)} \middle| W_{s,t}, H_{s,t}\right) = 3 + \frac{\frac{3}{1400}h^2 + \frac{3}{175}h\left(\left(H_{s,t}^{(i)}\right)^2 + \left(H_{s,t}^{(j)}\right)^2\right)}{\left(\frac{1}{20}h + \frac{1}{5}\left(\left(H_{s,t}^{(i)}\right)^2 + \left(H_{s,t}^{(j)}\right)^2\right)\right)^2}. \tag{7.1.19}$$

□

Remark 7.1.2. $\left(A_{s,t} \middle| W_{s,t}, H_{s,t}\right)$ always has more kurtosis than a normal distribution.

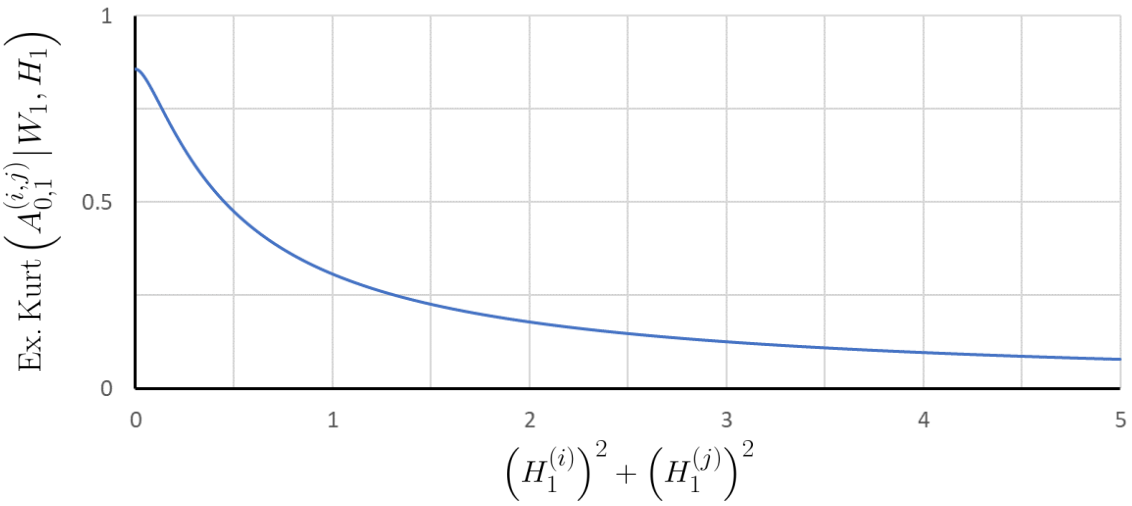

Figure 7.2: The kurtosis of Brownian Lévy area as a function of space-time area.

7.2 Additional moments of Lévy area conditional on the path's increment

In this section, we will apply the previous theorem to identify further moments of multidimensional Brownian Lévy area conditional on the increment of the path $W_{s,t}$. Since our motivation is to develop numerical approximations of the form (7.0.7), it suffices to consider moments of the multidimensional Brownian bridge Lévy area $b_{s,t}$. We shall begin by establishing a family of cross-moments of $\big(H_{s,t}, b_{s,t}\big)$ that are zero.

Theorem 7.2.1. *Let $H_{s,t}$ be the space-time Lévy area of a Brownian motion on $[s,t]$. Let $b_{s,t}$ denote the Lévy area of the associated Brownian bridge. Consider the moment*

$$\mu_{I,J} := \mathbb{E}\Big[\Big(H_{s,t}^{(i_1)} \cdots H_{s,t}^{(i_m)}\Big)\Big(b_{s,t}^{(j_1,j_2)} \cdots b_{s,t}^{(j_{n-1},j_n)}\Big)\Big], \tag{7.2.1}$$

where $I = \big(i_1, \cdots, i_m\big)$ and $J = \big(j_1, \cdots, j_n\big)$ are multi-indices with $i_k, j_k \in \{1, \cdots, d\}$. For each index $k \in \{1, \cdots, d\}$, we can count its number of appearances in I and J.

$$N_k := \big|\, p : i_p = k,\, i_p \in I \,\big| + \big|\, q : j_q = k,\, j_q \in J \,\big|. \tag{7.2.2}$$

Suppose there exists an index $k \in \{1, \cdots, d\}$ such that N_k is odd. Then $\mu_{I,J} = 0$.

Proof. We can construct a new process $\overline{B}$ on $[s,t]$ from the Brownian bridge B as

$$\overline{B}^{(i)} := \begin{cases} B^{(i)}, & \text{if } \ i \neq k, \\ -B^{(i)}, & \text{if } \ i = k. \end{cases}$$

Then $\overline{B}$ has the same law as a Brownian bridge and we can define the corresponding Lévy areas $\overline{H}_{s,t}$ and $\overline{b}_{s,t}$. By the linearity of integration, it immediately follows that

$$\overline{H}_{s,t}^{(i)} = \begin{cases} H_{s,t}^{(i)}, & \text{if } \ i \neq k, \\ -H_{s,t}^{(i)}, & \text{if } \ i = k. \end{cases} \quad \text{and} \quad \overline{b}_{s,t}^{(i,j)} = \begin{cases} b_{s,t}^{(i,j)}, & \text{if } \ k \notin \{i,j\} \ \text{ or } \ i = j, \\ -b_{s,t}^{(i,j)}, & \text{otherwise.} \end{cases}$$

As B and $\overline{B}$ have the same law, counting the number of k's appearing in (7.2.1) gives

$$\begin{aligned} \mu_{I,J} &= \mathbb{E}\Big[\Big(\overline{H}_{s,t}^{(i_1)} \cdots \overline{H}_{s,t}^{(i_m)}\Big)\Big(\overline{b}_{s,t}^{(j_1,j_2)} \cdots \overline{b}_{s,t}^{(j_{n-1},j_n)}\Big)\Big] \\ &= \big(-1\big)^{N_k}\, \mathbb{E}\Big[\Big(H_{s,t}^{(i_1)} \cdots H_{s,t}^{(i_m)}\Big)\Big(b_{s,t}^{(j_1,j_2)} \cdots b_{s,t}^{(j_{n-1},j_n)}\Big)\Big] \\ &= \big(-1\big)^{N_k} \mu_{I,J}. \end{aligned}$$

The result now follows as N_k is assumed to be odd for some index $k \in \{1, \cdots, d\}$. $\square$

In addition, there are certain cross-moments of $\left(H_{s,t}, b_{s,t}\right)$ that can be computed using the conditional moments obtained in Theorem 7.1.1 along with the Tower law (and possibly the fact that Brownian motion has independent coordinates processes). For example, the moment below is zero but falls outside the scope of Theorem 7.2.1:

$$\mathbb{E}\left[H_{s,t}^{(i)} H_{s,t}^{(j)} b_{s,t}^{(i,j)}\right] = \mathbb{E}\left[H_{s,t}^{(i)} H_{s,t}^{(j)} \mathbb{E}\left[b_{s,t}^{(i,j)} \middle| W_{s,t}, H_{s,t}\right]\right] = 0,$$

by the Tower law and the conditional expectation (7.1.2) detailed in Theorem 7.1.1.

We will now present the main result of this section, which gives some non-trivial cross-moments of $\left(H_{s,t}, b_{s,t}\right)$ with degrees at most four and even numbers of indices.

Theorem 7.2.2. *Let $b_{s,t}$ denote the Lévy area of a Brownian bridge on $[s,t]$. Then*

$$\mathbb{E}\left[b_{s,t}^{(i,j)} b_{s,t}^{(j,k)} b_{s,t}^{(k,i)}\right] = 0, \tag{7.2.3}$$

$$\mathbb{E}\left[\left(b_{s,t}^{(i,j)}\right)^2 \left(b_{s,t}^{(j,k)}\right)^2\right] = \frac{7}{720} h^4, \tag{7.2.4}$$

$$\mathbb{E}\left[H_{s,t}^{(i)} H_{s,t}^{(k)} b_{s,t}^{(i,j)} b_{s,t}^{(j,k)}\right] = -\frac{1}{720} h^3, \tag{7.2.5}$$

$$\mathbb{E}\left[b_{s,t}^{(i,j)} b_{s,t}^{(j,k)} b_{s,t}^{(k,l)} b_{s,t}^{(l,i)}\right] = \frac{1}{720} h^4, \tag{7.2.6}$$

where $i, j, k, l \in \{1, \cdots, d\}$ are distinct.

Proof. Since each coordinate of B is independent and identically distributed, we have

$$\mathbb{E}\left[b_{s,t}^{(i,j)} b_{s,t}^{(j,k)} b_{s,t}^{(k,i)}\right] = \mathbb{E}\left[b_{s,t}^{(k,j)} b_{s,t}^{(j,i)} b_{s,t}^{(i,k)}\right]. \tag{7.2.7}$$

On the other hand, $b_{s,t}$ can be viewed as an antisymmetric matrix as $b_{s,t}^{(i,j)} = -b_{s,t}^{(j,i)}$. Therefore, switching the indices of all the Lévy area terms in the RHS of (7.2.7) gives

$$\mathbb{E}\left[b_{s,t}^{(i,j)} b_{s,t}^{(j,k)} b_{s,t}^{(k,i)}\right] = -\mathbb{E}\left[b_{s,t}^{(j,k)} b_{s,t}^{(i,j)} b_{s,t}^{(k,i)}\right].$$

It then follows that the moment (7.2.3) must be zero. To derive (7.2.4), we note that

$$\widetilde{B}^{(i,k)} := \frac{\sqrt{2}}{2}\left(B^{(i)} + B^{(k)}\right), \tag{7.2.8}$$

is a process with the same law as a one-dimensional Brownian bridge on $[s,t]$. Hence

$$\frac{\sqrt{2}}{2}\left(b_{s,t}^{(i,j)} - b_{s,t}^{(j,k)}\right) = \frac{\sqrt{2}}{2}\left(b_{s,t}^{(i,j)} + b_{s,t}^{(k,j)}\right) = \int_s^t \frac{\sqrt{2}}{2}\left(B_{s,u}^{(i)} + B_{s,u}^{(k)}\right) \circ dB_u^{(j)},$$

has the same law as $b_{s,t}^{(i,j)}$. In particular, we can apply (7.1.16) to compute the moment

$$\mathbb{E}\left[\left(\frac{\sqrt{2}}{2}\left(b_{s,t}^{(i,j)} - b_{s,t}^{(j,k)}\right)\right)^4\right] = \mathbb{E}\left[\left(b_{s,t}^{(i,j)}\right)^4\right] = \frac{7}{240} h^4.$$

Expanding the LHS of the above and removing the terms with zero expectation gives

$$\frac{1}{4}\,\mathbb{E}\Big[\Big(b_{s,t}^{(i,j)}\Big)^4\Big]+\frac{3}{2}\,\mathbb{E}\Big[\Big(b_{s,t}^{(i,j)}\Big)^2\Big(b_{s,t}^{(j,k)}\Big)^2\Big]+\frac{1}{4}\,\mathbb{E}\Big[\Big(b_{s,t}^{(j,k)}\Big)^4\Big]=\frac{7}{240}h^4.$$

So by applying (7.1.16) to the first and third terms, we obtain the required moment.

$$\mathbb{E}\Big[\Big(b_{s,t}^{(i,j)}\Big)^2\Big(b_{s,t}^{(j,k)}\Big)^2\Big]=\frac{7}{720}h^4. \tag{7.2.9}$$

To compute the cross-moment (7.2.5), we will use the decomposition (7.1.6) for $b_{s,t}$.

$$\begin{aligned}\mathbb{E}\Big[H_{s,t}^{(i)}\,H_{s,t}^{(k)}\,b_{s,t}^{(i,j)}\,b_{s,t}^{(j,k)}\Big]=\mathbb{E}\Big[H_{s,t}^{(i)}\,H_{s,t}^{(k)}\times\Big(12\Big(K_{s,t}^{(i)}\,H_{s,t}^{(j)}-H_{s,t}^{(i)}K_{s,t}^{(j)}\Big)+a_{s,t}^{(i,j)}\Big)\\ \times\Big(12\Big(K_{s,t}^{(j)}\,H_{s,t}^{(k)}-H_{s,t}^{(j)}K_{s,t}^{(k)}\Big)+a_{s,t}^{(j,k)}\Big)\Big].\end{aligned}$$

When the RHS of the above is expanded, the majority of the resulting terms have zero expectation since both $K_{s,t}$ and $a_{s,t}$ are independent of $H_{s,t}$. This leaves us with

$$\mathbb{E}\Big[H_{s,t}^{(i)}\,H_{s,t}^{(k)}\,b_{s,t}^{(i,j)}\,b_{s,t}^{(j,k)}\Big]=-144\,\mathbb{E}\Big[\Big(H_{s,t}^{(i)}\Big)^2\Big(H_{s,t}^{(k)}\Big)^2\Big(K_{s,t}^{(j)}\Big)^2\Big]=-\frac{1}{720}h^3.$$

Therefore all that remains is to compute the final non-trivial cross-moment, (7.2.6). By considering the process $\widetilde{B}^{(i,k)}$ $\big($which has the same law as $B^{(i)}\big)$, we can see that

$$\left(\frac{\sqrt{2}}{2}\left(b_{s,t}^{(i,j)}+b_{s,t}^{(k,j)}\right),\ \frac{\sqrt{2}}{2}\left(b_{s,t}^{(l,i)}+b_{s,t}^{(l,k)}\right)\right)$$

has the same law as $\big(b_{s,t}^{(i,j)},\,b_{s,t}^{(l,i)}\big)$. Hence both pairs have the same fourth moments.

$$\mathbb{E}\left[\left(\frac{\sqrt{2}}{2}\left(b_{s,t}^{(i,j)}-b_{s,t}^{(j,k)}\right)+\frac{\sqrt{2}}{2}\left(b_{s,t}^{(l,i)}-b_{s,t}^{(k,l)}\right)\right)^4\right]=\mathbb{E}\Big[\Big(b_{s,t}^{(i,j)}+b_{s,t}^{(l,i)}\Big)^4\Big].$$

As before, we shall expand both sides and remove the moments that are equal to zero.

$$\begin{aligned}&\mathbb{E}\Big[\Big(b_{s,t}^{(i,j)}\Big)^4\Big]+6\,\mathbb{E}\Big[\Big(b_{s,t}^{(i,j)}\Big)^2\Big(b_{s,t}^{(l,i)}\Big)^2\Big]+\mathbb{E}\Big[\Big(b_{s,t}^{(l,i)}\Big)^4\Big]\\ &=\frac{1}{4}\Big(4\times\mathbb{E}\Big[\Big(b_{s,t}^{(i,j)}\Big)^4\Big]+6\times4\times\mathbb{E}\Big[\Big(b_{s,t}^{(i,j)}\Big)^2\Big(b_{s,t}^{(j,k)}\Big)^2\Big]+6\times2\times\mathbb{E}\Big[\Big(b_{s,t}^{(i,j)}\Big)^2\Big(b_{s,t}^{(k,l)}\Big)^2\Big]\\ &\qquad+\ 24\times\mathbb{E}\Big[b_{s,t}^{(i,j)}b_{s,t}^{(j,k)}b_{s,t}^{(k,l)}b_{s,t}^{(l,i)}\Big]\Big).\end{aligned}$$

Note that the cross-moments we removed can be seen to be zero by Theorem 7.2.1. Since the variance and kurtosis of $b_{s,t}^{(i,j)}$ are both known, the above can be simplified:

$$\mathbb{E}\Big[b_{s,t}^{(i,j)}b_{s,t}^{(j,k)}b_{s,t}^{(k,l)}b_{s,t}^{(l,i)}\Big]=\frac{1}{6}\Big(\mathbb{E}\Big[\Big(b_{s,t}^{(i,j)}\Big)^4\Big]-3\,\mathbb{E}\Big[\Big(b_{s,t}^{(i,j)}\Big)^2\Big]\,\mathbb{E}\Big[\Big(b_{s,t}^{(k,l)}\Big)^2\Big]\Big)=\frac{1}{720}h^4,$$

by Theorem 7.0.15 $\big($where it was shown that $b_{s,t}^{(i,j)}$ follows a logistic distribution$\big)$. □

Remark 7.2.3. The results of this section allow us to compute any conditional moment (with degree at most four) of Brownian Lévy area conditional on the increment $W_{s,t}$.

7.3 Efficient weak approximation of Lévy area in low-dimensional settings

In this section, we shall present approximations of Brownian Lévy area with the form

$$\widetilde{A}_{s,t} = H_{s,t} \otimes W_{s,t} - W_{s,t} \otimes H_{s,t} + \widetilde{b}_{s,t}, \tag{7.3.1}$$

where $\widetilde{b}_{s,t}$ approximates the Brownian bridge Lévy area $b_{s,t}$ in the "weak" sense. More specifically, the random variable $\widetilde{b}_{s,t}$ will be constructed to capture the various moments of $b_{s,t}$ established in the previous sections (see Theorems 7.1.1 and 7.2.2). We shall develop numerical approximations in the low-dimensional cases ($d = 2, 3$) that are fast to generate and match the moments of $A_{s,t}$ with degree at most four. Due to the non-trivial moment (7.2.6), it is ongoing research to construct $\widetilde{b}_{s,t}$ in the general multidimensional setting ($d > 3$) that matches the cokurtosis structure of $b_{s,t}$. We shall conclude the section by numerically investigating the $d = 2$ approximation:

Definition 7.3.1 (Weak approximation of two-dimensional Brownian Lévy area)**.** Let $\widetilde{A}_{s,t}$ denote the antisymmetric matrix with the following entries (for $i, j \in \{1, 2\}$),

$$\widetilde{A}_{s,t}^{(i,j)} := H_{s,t}^{(i)} W_{s,t}^{(j)} - W_{s,t}^{(i)} H_{s,t}^{(j)} + \widetilde{b}_{s,t}^{(i,j)}, \tag{7.3.2}$$

where, conditional on the area $H_{s,t}$, the random variable $\widetilde{b}_{s,t}$ will be defined as follows:

$$\widetilde{b}_{s,t}^{(i,j)} := \begin{cases} \sigma_{s,t}\, \xi_{s,t}, & \text{if } (i,j) = (1,2), \\ -\sigma_{s,t}\, \xi_{s,t}, & \text{if } (i,j) = (2,1), \\ 0, & \text{if } i = j, \end{cases} \tag{7.3.3}$$

$$\xi_{s,t} \sim \begin{cases} \text{Logistic}\left(0, \frac{\sqrt{3}}{\pi}\right), & \text{with probability } p_{s,t}, \\ \text{Normal}\big(0, 1\big), & \text{with probability } 1 - p_{s,t}, \end{cases} \tag{7.3.4}$$

$$\sigma_{s,t} := \sqrt{\frac{1}{20} h^2 + \frac{1}{5} h \left(\left(H_{s,t}^{(1)} \right)^2 + \left(H_{s,t}^{(2)} \right)^2 \right)}, \tag{7.3.5}$$

$$p_{s,t} := \frac{1}{\sigma_{s,t}^4} \left(\frac{1}{560} h^4 + \frac{1}{70} h^3 \left(\left(H_{s,t}^{(1)} \right)^2 + \left(H_{s,t}^{(2)} \right)^2 \right) \right), \tag{7.3.6}$$

$$h := t - s.$$

As it may not be clear, the random variable $\xi_{s,t}$ can be generated via the procedure:

Step 0. Generate the space-time Lévy areas $H_{s,t}^{(1)}$ and $H_{s,t}^{(2)}$ on the interval $[s, t]$. Using $H_{s,t}$, we can define the variables $\sigma_{s,t}$ and $p_{s,t}$ according to (7.3.5) and (7.3.6).

Step 1. As $p_{s,t} < 1$, we can generate an independent variable $X \sim \text{Bernoulli}\big(p_{s,t}\big)$.

Step 2. If $X = 1$, define $\xi_{s,t}$ to be an independent logistic random variable with mean 0 and variance 1.

Step 3. If $X = 0$, define $\xi_{s,t}$ to be an independent normal random variable with mean 0 and variance 1.

Remark 7.3.2. We can see that $0 < p_{s,t} < 1$ (and therefore $\widetilde{A}_{s,t}$ is well-defined) since

$$\begin{aligned} p_{s,t}\,\sigma_{s,t}^4 &= \frac{1}{560}h^4 + \frac{1}{70}h^3\left(\left(H_{s,t}^{(1)}\right)^2 + \left(H_{s,t}^{(2)}\right)^2\right) \\ &< \frac{1}{400}h^4 + \frac{1}{50}h^3\left(\left(H_{s,t}^{(1)}\right)^2 + \left(H_{s,t}^{(2)}\right)^2\right) + \frac{1}{25}h^2\left(\left(H_{s,t}^{(1)}\right)^2 + \left(H_{s,t}^{(2)}\right)^2\right)^2 = \sigma_{s,t}^4. \end{aligned}$$

Remark 7.3.3. It is clear that $\big(\xi_{s,t}\big|W_{s,t}, H_{s,t}\big)$ always has zero mean and unit variance. The above approximation is designed to match the following moments of Lévy area.

Theorem 7.3.4. *Let $\widetilde{A}_{s,t}$ denote the approximation given by definition 7.3.2. Then*

$$\mathbb{E}\Big[\Big(\widetilde{A}_{s,t}^{(i,j)}\Big)^n \,\Big|\, W_{s,t}, H_{s,t}\Big] = \mathbb{E}\Big[\Big(A_{s,t}^{(i,j)}\Big)^n \,\Big|\, W_{s,t}, H_{s,t}\Big], \tag{7.3.7}$$

for $i, j \in \{1, 2\}$ and $1 \le n \le 4$.

Proof. To begin, we shall compute the moments of $\xi_{s,t}$ conditional on $W_{s,t}$ and $H_{s,t}$.

$$\begin{aligned} \mathbb{E}\big[(\xi_{s,t})^n \,\big|\, W_{s,t}, H_{s,t}\big] &= \mathbb{E}\big[(\xi_{s,t})^n \,\big|\, W_{s,t}, H_{s,t}, (X = 0)\big]\,\mathbb{P}\big(X = 0\big) \\ &\quad + \mathbb{E}\big[(\xi_{s,t})^n \,\big|\, W_{s,t}, H_{s,t}, (X = 1)\big]\,\mathbb{P}\big(X = 1\big). \end{aligned}$$

As the normal and logistic distributions have the same first three moments, we have

$$\begin{aligned} \mathbb{E}\big[\xi_{s,t}\,\big|\, W_{s,t}, H_{s,t}\big] &= 0, \\ \mathbb{E}\Big[(\xi_{s,t})^2 \,\Big|\, W_{s,t}, H_{s,t}\Big] &= 1, \\ \mathbb{E}\Big[(\xi_{s,t})^3 \,\Big|\, W_{s,t}, H_{s,t}\Big] &= 0. \end{aligned}$$

We will also compute the fourth moment of $\xi_{s,t}$ conditional on the pair $(W_{s,t}, H_{s,t})$.

$$\begin{aligned} \mathbb{E}\Big[(\xi_{s,t})^4 \,\Big|\, W_{s,t}, H_{s,t}\Big] &= 3\big(1 - p_{s,t}\big) + \frac{21}{5}p_{s,t} \\ &= 3 + \frac{6}{5}p_{s,t} \\ &= 3 + \frac{1}{\sigma_{s,t}^4}\left(\frac{3}{1400}h^4 + \frac{3}{175}h^3\left(\left(H_{s,t}^{(1)}\right)^2 + \left(H_{s,t}^{(2)}\right)^2\right)\right). \end{aligned}$$

So by Theorem 7.1.1, $\xi_{s,t}$ has the same conditional skewness and kurtosis as $A_{s,t}^{(i,j)}$ when $i \neq j$. The result now follows since $\widetilde{A}_{s,t}^{(i,j)} = H_{s,t}^{(i)}W_{s,t}^{(j)} - W_{s,t}^{(i)}H_{s,t}^{(j)} \pm \sigma_{s,t}\,\xi_{s,t}$. □

Using the additional cross-moments established by Theorems (7.2.1) and (7.2.2), we can also construct a high order weak approximation of $A_{s,t}$ with dimension $d = 3$.

Definition 7.3.5 (Weak approximation of three-dimensional Brownian Lévy area)**.** Let $\widetilde{A}_{s,t}$ denote the antisymmetric tensor with the following entries (for $i, j \in \{1, 2, 3\}$),

$$\widetilde{A}_{s,t}^{(i,j)} := H_{s,t}^{(i)} W_{s,t}^{(j)} - W_{s,t}^{(i)} H_{s,t}^{(j)} + 12\left(K_{s,t}^{(i)} H_{s,t}^{(j)} - H_{s,t}^{(i)} K_{s,t}^{(j)}\right) + \widetilde{a}_{s,t}^{(i,j)}, \tag{7.3.8}$$

where, conditional on the space-time-time Lévy area $K_{s,t}$, the random variable $\widetilde{a}_{s,t}$ is

$$\widetilde{a}_{s,t}^{(i,j)} := \begin{cases} \sigma_{s,t}^{(i,j)}\, \xi_{s,t}^{(i,j)}, & \text{if } i < j, \\ -\sigma_{s,t}^{(i,j)}\, \xi_{s,t}^{(i,j)}, & \text{if } i > j, \\ 0, & \text{if } i = j, \end{cases} \tag{7.3.9}$$

with the independent random variables $\xi_{s,t}^{(i,j)}$ and $\sigma_{s,t}^{(i,j)}$ defined for $i < j$ according to

$$\xi_{s,t}^{(i,j)} \sim \begin{cases} \text{Uniform}\left[-\sqrt{3}, \sqrt{3}\right], & \text{with probability } p \\ \text{Rademacher}, & \text{with probability } 1 - p, \end{cases} \tag{7.3.10}$$

and

$$\sigma_{s,t}^{(i,j)} := \sqrt{\frac{3}{28}\big(C^{(i)} + c\big)\big(C^{(j)} + c\big)h^2 + \frac{1}{28}h\left(\left(12K_{s,t}^{(i)}\right)^2 + \left(12K_{s,t}^{(j)}\right)^2\right)}, \tag{7.3.11}$$

where each $C^{(i)}$ is an independent and identically distributed random variable with

$$C^{(i)} \sim \text{Exponential}\left(\frac{15}{8}\right), \tag{7.3.12}$$

and the constants c and p are defined as

$$c := \frac{\sqrt{3}}{3} - \frac{8}{15},$$
$$p := \frac{21130}{25621}.$$

The above approximation is designed to match the following moments of Lévy area.

Theorem 7.3.6. *Let $\widetilde{A}_{s,t}$ denote the approximation given by definition 7.3.8. Then*

$$\mathbb{E}\left[\left(\widetilde{A}_{s,t}^{(i,j)}\right)^{n_0} \middle| W_{s,t}, H_{s,t}\right] = \mathbb{E}\left[\left(A_{s,t}^{(i,j)}\right)^{n_0} \middle| W_{s,t}, H_{s,t}\right], \tag{7.3.13}$$

and

$$\begin{aligned} &\mathbb{E}\left[\left(\widetilde{A}_{s,t}^{(i,j)}\right)^{n_1}\left(\widetilde{A}_{s,t}^{(j,k)}\right)^{n_2}\left(\widetilde{A}_{s,t}^{(k,i)}\right)^{n_3} \middle| W_{s,t}\right] \\ &\qquad = \mathbb{E}\left[\left(A_{s,t}^{(i,j)}\right)^{n_1}\left(A_{s,t}^{(j,k)}\right)^{n_2}\left(A_{s,t}^{(k,i)}\right)^{n_3} \middle| W_{s,t}\right], \end{aligned} \tag{7.3.14}$$

for distinct $i, j, k \in \{1, 2, 3\}$, $n_0, n_1, n_2, n_3 \geq 0$ with $n_0 \leq 4$ and $n_1 + n_2 + n_3 \leq 4$.

Proof. To compute the conditional moments of $\widetilde{a}_{s,t}^{(i,j)}$, we note that by independence

$$\mathbb{E}\Big[\Big(\widetilde{a}_{s,t}^{(i,j)}\Big)^n \,\Big|\, W_{s,t}, H_{s,t}\Big] = \mathbb{E}\Big[\Big(\sigma_{s,t}^{(i,j)}\Big)^n\Big]\,\mathbb{E}\Big[\Big(\xi_{s,t}^{(i,j)}\Big)^n\Big], \tag{7.3.15}$$

for $i < j$. As $\xi_{s,t}$ is a mixture of uniform and Rademacher random variables, we have

$$\begin{aligned}
\mathbb{E}\Big[\xi_{s,t}^{(i,j)}\Big] &= 0,\\
\mathbb{E}\Big[\Big(\xi_{s,t}^{(i,j)}\Big)^2\Big] &= 1,\\
\mathbb{E}\Big[\Big(\xi_{s,t}^{(i,j)}\Big)^3\Big] &= 0,\\
\mathbb{E}\Big[\Big(\xi_{s,t}^{(i,j)}\Big)^4\Big] &= \frac{9}{5}p + \big(1-p\big) = \frac{42525}{25621}.
\end{aligned}$$

Here we used the fact that the kurtosis of the uniform distribution is known to be $\frac{9}{5}$. Using the mean and variance of each $C^{(i)}$, we can compute the even moments of $\sigma_{s,t}^{(i,j)}$.

$$\begin{aligned}
\mathbb{E}\Big[\Big(\sigma_{s,t}^{(i,j)}\Big)^2\Big] &= \mathbb{E}\Big[\frac{3}{28}\big(C^{(i)}+c\big)\big(C^{(j)}+c\big)h^2 + \frac{1}{28}h\Big(\Big(12K_{s,t}^{(i)}\Big)^2 + \Big(12K_{s,t}^{(j)}\Big)^2\Big)\Big],\\
&= \left(\frac{3}{28}\times\frac{\sqrt{3}}{3}\times\frac{\sqrt{3}}{3} + \frac{1}{28}\times\frac{2}{5}\right)h^2 = \frac{1}{20}h^2, \qquad (7.3.16)\\
\mathbb{E}\Big[\Big(\sigma_{s,t}^{(i,j)}\Big)^4\Big] &= \mathbb{E}\left[\left(\frac{3}{28}\big(C^{(i)}+c\big)\big(C^{(j)}+c\big)h^2 + \frac{1}{28}h\Big(\Big(12K_{s,t}^{(i)}\Big)^2 + \Big(12K_{s,t}^{(j)}\Big)^2\Big)\right)^2\right]\\
&= \left(\frac{3}{28}\right)^2\mathbb{E}\Big[\big(C^{(i)}+c\big)^2\Big]\,\mathbb{E}\Big[\big(C^{(j)}+c\big)^2\Big]\,h^4\\
&\qquad + 2\left(\frac{3}{28}\times\frac{1}{3}\times\frac{1}{28}\times\frac{2}{5}\right)h^4 + \left(\frac{1}{28}h\right)^2\times 8\times\left(\frac{1}{5}h\right)^2\\
&= \left(\frac{3}{28}\right)^2\left(2\left(\frac{8}{15}\right)^2 + \frac{16}{15}\left(\frac{\sqrt{3}}{3}-\frac{8}{15}\right) + \left(\frac{\sqrt{3}}{3}-\frac{8}{15}\right)^2\right)^2 h^4 + \frac{1}{700}h^4\\
&= \frac{19321}{4410000}h^4 + \frac{1}{700}h^4 = \frac{25621}{4410000}h^4. \qquad (7.3.17)
\end{aligned}$$

Therefore by (7.3.16) and (7.3.17), along with the first four moments of $\xi_{s,t}^{(i,j)}$, we have

$$\begin{aligned}
\mathbb{E}\Big[\widetilde{a}_{s,t}^{(i,j)}\Big] &= 0,\\
\mathbb{E}\Big[\Big(\widetilde{a}_{s,t}^{(i,j)}\Big)^2\Big] &= \frac{1}{20}h^2,\\
\mathbb{E}\Big[\Big(\widetilde{a}_{s,t}^{(i,j)}\Big)^3\Big] &= 0,\\
\mathbb{E}\Big[\Big(\widetilde{a}_{s,t}^{(i,j)}\Big)^4\Big] &= \frac{27}{2800}h^4.
\end{aligned}$$

We shall now define the "Brownian bridge" Lévy area $\widetilde{b}_{s,t}$ corresponding to $\widetilde{A}_{s,t}$.

$$\widetilde{b}_{s,t} := \widetilde{A}_{s,t} - \big(H_{s,t} \otimes W_{s,t} - W_{s,t} \otimes H_{s,t}\big). \tag{7.3.18}$$

Hence, it is enough to compare the conditional moments of $\widetilde{b}_{s,t}$ and $b_{s,t}$. Recall that

$$\widetilde{b}_{s,t}^{(i,j)} = 12\big(K_{s,t}^{(i)} H_{s,t}^{(j)} - H_{s,t}^{(i)} K_{s,t}^{(j)}\big) + \widetilde{a}_{s,t}^{(i,j)}, \tag{7.3.19}$$

where $\widetilde{a}_{s,t}^{(i,j)}$ and the space-time-time Lévy area $K_{s,t}$ are independent of $(W_{s,t}, H_{s,t})$. Using the above, the first three moments of $\widetilde{b}_{s,t}^{(i,j)}$ are then straightforward to compute

$$\begin{aligned}
\mathbb{E}\Big[\widetilde{b}_{s,t}^{(i,j)} \,\big|\, W_{s,t}, H_{s,t}\Big] &= 0, \\
\mathbb{E}\Big[\big(\widetilde{b}_{s,t}^{(i,j)}\big)^2 \,\big|\, W_{s,t}, H_{s,t}\Big] &= \big(H_{s,t}^{(j)}\big)^2 \,\mathbb{E}\Big[\big(12K_{s,t}^{(i)}\big)^2\Big] + \big(H_{s,t}^{(i)}\big)^2 \,\mathbb{E}\Big[\big(12K_{s,t}^{(j)}\big)^2\Big] \\
&\qquad + \mathbb{E}\Big[\big(\widetilde{a}_{s,t}^{(i,j)}\big)^2\Big], \\
&= \frac{1}{20}h^2 + \frac{1}{5}h\Big(\big(H_{s,t}^{(i)}\big)^2 + \big(H_{s,t}^{(j)}\big)^2\Big), \\
\mathbb{E}\Big[\big(\widetilde{b}_{s,t}^{(i,j)}\big)^3 \,\big|\, W_{s,t}, H_{s,t}\Big] &= 0.
\end{aligned}$$

Therefore $\widetilde{A}_{s,t}^{(i,j)}$ and $A_{s,t}^{(i,j)}$ have the same conditional moments up to degree three. It is also worth noting the following non-zero cross-moment between $\widetilde{a}_{s,t}^{(i,j)}$ and $K_{s,t}^{(i)}$.

$$\begin{aligned}
&\mathbb{E}\Big[\big(K_{s,t}^{(i)}\big)^2 \big(\widetilde{a}_{s,t}^{(i,j)}\big)^2\Big] \\
&= \mathbb{E}\Big[\big(K_{s,t}^{(i)}\big)^2 \big(\sigma_{s,t}^{(i,j)}\big)^2\Big] \,\mathbb{E}\Big[\big(\xi_{s,t}^{(i,j)}\big)^2\Big] \\
&= \mathbb{E}\Big[\big(K_{s,t}^{(i)}\big)^2 \Big(\frac{3}{28}\big(C^{(i)} + c\big)\big(C^{(j)} + c\big)h^2 + \frac{1}{28}h\Big(\big(12K_{s,t}^{(i)}\big)^2 + \big(12K_{s,t}^{(j)}\big)^2\Big)\Big)\Big] \\
&= \left(\frac{1}{720} \times \frac{3}{28} \times \frac{1}{3} + \frac{1}{28} \times 144 \times 3 \times \left(\frac{1}{720}\right)^2 + \frac{1}{28} \times 144 \times \left(\frac{1}{720}\right)^2\right) h^3 \\
&= \frac{1}{11200}h^3.
\end{aligned} \tag{7.3.20}$$

By the construction of $\widetilde{a}_{s,t}$, it is also clear the below fourth degree moments are zero.

$$\begin{aligned}
\mathbb{E}\Big[K_{s,t}^{(i)} \big(\widetilde{a}_{s,t}^{(i,j)}\big)^3\Big] &= 0, \\
\mathbb{E}\Big[K_{s,t}^{(i)} K_{s,t}^{(j)} \big(\widetilde{a}_{s,t}^{(i,j)}\big)^2\Big] &= 0, \\
\mathbb{E}\Big[\big(K_{s,t}^{(i)}\big)^3 \widetilde{a}_{s,t}^{(i,j)}\Big] &= 0, \\
\mathbb{E}\Big[K_{s,t}^{(i)} \big(K_{s,t}^{(j)}\big)^2 \widetilde{a}_{s,t}^{(i,j)}\Big] &= 0.
\end{aligned}$$

Expanding the fourth power of $\widetilde{b}_{s,t}^{(i,j)}$ and substituting in the above moments produces

$$\begin{aligned}
\mathbb{E}\Big[\Big(\widetilde{b}_{s,t}^{(i,j)}\Big)^4 \,\big|\, W_{s,t}, H_{s,t}\Big] &= 12^4\Big(H_{s,t}^{(j)}\Big)^4\,\mathbb{E}\Big[\Big(K_{s,t}^{(i)}\Big)^4\Big] + 12^4\Big(H_{s,t}^{(i)}\Big)^4\,\mathbb{E}\Big[\Big(K_{s,t}^{(j)}\Big)^4\Big] \\
&\quad + 6\times 12^4\Big(H_{s,t}^{(i)}\Big)^2\Big(H_{s,t}^{(j)}\Big)^2\,\mathbb{E}\Big[\Big(K_{s,t}^{(i)}\Big)^2\Big(K_{s,t}^{(j)}\Big)^2\Big] \\
&\quad + 6\times 144\Big(H_{s,t}^{(j)}\Big)^2\,\mathbb{E}\Big[\Big(K_{s,t}^{(i)}\Big)^2\Big(\widetilde{a}_{s,t}^{(i,j)}\Big)^2\Big] \\
&\quad + 6\times 144\Big(H_{s,t}^{(i)}\Big)^2\,\mathbb{E}\Big[\Big(K_{s,t}^{(j)}\Big)^2\Big(\widetilde{a}_{s,t}^{(i,j)}\Big)^2\Big] + \mathbb{E}\Big[\Big(\widetilde{a}_{s,t}^{(i,j)}\Big)^4\Big] \\
&= \frac{3}{25}h^2\Big(\Big(H_{s,t}^{(i)}\Big)^4 + \Big(H_{s,t}^{(j)}\Big)^4\Big) + \frac{6}{25}h^2\Big(H_{s,t}^{(i)}\Big)^2\Big(H_{s,t}^{(j)}\Big)^2 \\
&\quad + \frac{27}{350}h^3\Big(\Big(H_{s,t}^{(i)}\Big)^2 + \Big(H_{s,t}^{(j)}\Big)^2\Big) + \frac{27}{2800}h^4.
\end{aligned}$$

This matches the kurtosis given by Theorem 7.1.1, and so we have verified (7.3.13). We will now compute fourth degree moments of $\widetilde{b}_{s,t}$ with the forms (7.2.4) and (7.2.5).

$$\begin{aligned}
&\mathbb{E}\Big[\Big(\widetilde{b}_{s,t}^{(i,j)}\Big)^2\Big(\widetilde{b}_{s,t}^{(j,k)}\Big)^2\Big] \qquad (7.3.21)\\
&\quad = 12^4\,\mathbb{E}\Big[\Big(K_{s,t}^{(i)}H_{s,t}^{(j)} - H_{s,t}^{(i)}K_{s,t}^{(j)}\Big)^2\Big(K_{s,t}^{(j)}H_{s,t}^{(k)} - H_{s,t}^{(j)}K_{s,t}^{(k)}\Big)^2\Big] \\
&\qquad + \big(2\times 12^3\big)\,\mathbb{E}\Big[\Big(K_{s,t}^{(i)}H_{s,t}^{(j)} - H_{s,t}^{(i)}K_{s,t}^{(j)}\Big)\Big(K_{s,t}^{(j)}H_{s,t}^{(k)} - H_{s,t}^{(j)}K_{s,t}^{(k)}\Big)^2\widetilde{a}_{s,t}^{(i,j)}\Big] \\
&\qquad + \big(2\times 12^3\big)\,\mathbb{E}\Big[\Big(K_{s,t}^{(i)}H_{s,t}^{(j)} - H_{s,t}^{(i)}K_{s,t}^{(j)}\Big)^2\Big(K_{s,t}^{(j)}H_{s,t}^{(k)} - H_{s,t}^{(j)}K_{s,t}^{(k)}\Big)\widetilde{a}_{s,t}^{(j,k)}\Big] \\
&\qquad + 24^2\,\mathbb{E}\Big[\Big(K_{s,t}^{(i)}H_{s,t}^{(j)} - H_{s,t}^{(i)}K_{s,t}^{(j)}\Big)\Big(K_{s,t}^{(j)}H_{s,t}^{(k)} - H_{s,t}^{(j)}K_{s,t}^{(k)}\Big)\widetilde{a}_{s,t}^{(i,j)}\widetilde{a}_{s,t}^{(j,k)}\Big] \\
&\qquad + 144\,\mathbb{E}\Big[\Big(K_{s,t}^{(j)}H_{s,t}^{(k)} - H_{s,t}^{(j)}K_{s,t}^{(k)}\Big)^2\Big(\widetilde{a}_{s,t}^{(i,j)}\Big)^2\Big] \\
&\qquad + 144\,\mathbb{E}\Big[\Big(K_{s,t}^{(i)}H_{s,t}^{(j)} - H_{s,t}^{(i)}K_{s,t}^{(j)}\Big)^2\Big(\widetilde{a}_{s,t}^{(j,k)}\Big)^2\Big] \\
&\qquad + 24\,\mathbb{E}\Big[\Big(K_{s,t}^{(i)}H_{s,t}^{(j)} - H_{s,t}^{(i)}K_{s,t}^{(j)}\Big)\widetilde{a}_{s,t}^{(i,j)}\Big(\widetilde{a}_{s,t}^{(j,k)}\Big)^2\Big] \\
&\qquad + 24\,\mathbb{E}\Big[\Big(K_{s,t}^{(j)}H_{s,t}^{(k)} - H_{s,t}^{(j)}K_{s,t}^{(k)}\Big)\Big(\widetilde{a}_{s,t}^{(i,j)}\Big)^2\widetilde{a}_{s,t}^{(j,k)}\Big] \\
&\qquad + \mathbb{E}\Big[\Big(\widetilde{a}_{s,t}^{(i,j)}\Big)^2\Big(\widetilde{a}_{s,t}^{(j,k)}\Big)^2\Big].
\end{aligned}$$

As $\xi_{s,t}^{(i,j)}, \xi_{s,t}^{(j,k)}$ are both independent and symmetric, the above expansion simplifies to

$$\begin{aligned}
\mathbb{E}\Big[\Big(\widetilde{b}_{s,t}^{(i,j)}\Big)^2\Big(\widetilde{b}_{s,t}^{(j,k)}\Big)^2\Big] &= 12^4\,\mathbb{E}\Big[\Big(K_{s,t}^{(i)}H_{s,t}^{(j)} - H_{s,t}^{(i)}K_{s,t}^{(j)}\Big)^2\Big(K_{s,t}^{(j)}H_{s,t}^{(k)} - H_{s,t}^{(j)}K_{s,t}^{(k)}\Big)^2\Big] \\
&\quad + 144\,\mathbb{E}\Big[\Big(\Big(K_{s,t}^{(j)}H_{s,t}^{(k)}\Big)^2 + \Big(H_{s,t}^{(j)}K_{s,t}^{(k)}\Big)^2\Big)\Big(\widetilde{a}_{s,t}^{(i,j)}\Big)^2\Big] \\
&\quad + 144\,\mathbb{E}\Big[\Big(\Big(K_{s,t}^{(i)}H_{s,t}^{(j)}\Big)^2 + \Big(H_{s,t}^{(i)}K_{s,t}^{(j)}\Big)^2\Big)\Big(\widetilde{a}_{s,t}^{(j,k)}\Big)^2\Big] \\
&\quad + \mathbb{E}\Big[\Big(\widetilde{a}_{s,t}^{(i,j)}\Big)^2\Big(\widetilde{a}_{s,t}^{(j,k)}\Big)^2\Big].
\end{aligned}$$

The middle terms will be equal as each coordinate of W is identically distributed. We shall now compute the first term by expanding the brackets inside the expectation.

$$\begin{aligned}
&12^4\,\mathbb{E}\Big[\big(K^{(i)}_{s,t}\,H^{(j)}_{s,t}-H^{(i)}_{s,t}K^{(j)}_{s,t}\big)^2\big(K^{(j)}_{s,t}\,H^{(k)}_{s,t}-H^{(j)}_{s,t}\,K^{(k)}_{s,t}\big)^2\Big]\\
&\quad=12^4\Big(\mathbb{E}\Big[\big(K^{(i)}_{s,t}\,H^{(j)}_{s,t}\big)^2\big(K^{(j)}_{s,t}\,H^{(k)}_{s,t}\big)^2\Big]+\mathbb{E}\Big[\big(K^{(i)}_{s,t}\,H^{(j)}_{s,t}\big)^2\big(H^{(j)}_{s,t}\,K^{(k)}_{s,t}\big)^2\Big]\\
&\qquad\qquad+\mathbb{E}\Big[\big(H^{(i)}_{s,t}\,K^{(j)}_{s,t}\big)^2\big(K^{(j)}_{s,t}\,H^{(k)}_{s,t}\big)^2\Big]+\mathbb{E}\Big[\big(H^{(i)}_{s,t}\,K^{(j)}_{s,t}\big)^2\big(H^{(j)}_{s,t}\,K^{(k)}_{s,t}\big)^2\Big]\Big)\\
&\quad=2\times12^4\left(\left(\frac{1}{12}\right)^2\left(\frac{1}{720}\right)^2+3\left(\frac{1}{12}\right)^2\left(\frac{1}{720}\right)^2\right)h^4=\frac{1}{450}h^4.
\end{aligned}\tag{7.3.22}$$

For the middle terms, we use the cross-moment between $\widetilde{a}_{s,t}$ and $K_{s,t}$ given by (7.3.20).

$$\begin{aligned}
&144\,\mathbb{E}\Big[\Big(\big(K^{(j)}_{s,t}\,H^{(k)}_{s,t}\big)^2+\big(H^{(j)}_{s,t}\,K^{(k)}_{s,t}\big)^2\Big)\big(\widetilde{a}^{(i,j)}_{s,t}\big)^2\Big]\\
&\quad=144\left(\frac{1}{12}\times\frac{1}{11200}+\frac{1}{12}\times\frac{1}{720}\times\frac{1}{20}\right)h^4=\frac{1}{525}h^4.
\end{aligned}\tag{7.3.23}$$

The final term can be computed using the moments of the exponential distribution.

$$\begin{aligned}
&\mathbb{E}\Big[\big(\widetilde{a}^{(i,j)}_{s,t}\big)^2\big(\widetilde{a}^{(j,k)}_{s,t}\big)^2\Big]\\
&\quad=\mathbb{E}\Big[\Big(\frac{3}{28}\big(C^{(i)}+c\big)\big(C^{(j)}+c\big)h^2+\frac{1}{28}h\Big(\big(12K^{(i)}_{s,t}\big)^2+\big(12K^{(j)}_{s,t}\big)^2\Big)\Big)\\
&\qquad\times\Big(\frac{3}{28}\big(C^{(j)}+c\big)\big(C^{(k)}+c\big)h^2+\frac{1}{28}h\Big(\big(12K^{(j)}_{s,t}\big)^2+\big(12K^{(k)}_{s,t}\big)^2\Big)\Big)\Big]\\
&\quad=\left(\left(\frac{3}{28}\right)^2\times\frac{1}{3}\times\frac{139}{225}+2\times\frac{3}{28}\times\frac{1}{28}\times\frac{1}{3}\times\frac{2}{5}+\left(\frac{1}{28}\right)^2\times6\times\left(\frac{1}{5}\right)^2\right)h^4\\
&\quad=\frac{31}{8400}h^4.
\end{aligned}\tag{7.3.24}$$

Putting all this all together produces the required fourth degree cross-moment of $\widetilde{b}_{s,t}$.

$$\mathbb{E}\Big[\big(\widetilde{b}^{(i,j)}_{s,t}\big)^2\big(\widetilde{b}^{(j,k)}_{s,t}\big)^2\Big]=\frac{1}{450}h^4+\frac{2}{525}h^4+\frac{31}{8400}h^4=\frac{7}{720}h^4.\tag{7.3.25}$$

Computing the moment of $\big(H_{s,t},\widetilde{b}_{s,t}\big)$ corresponding to (7.2.5) is straightforward as

$$\begin{aligned}
&\mathbb{E}\Big[H^{(i)}_{s,t}\,H^{(k)}_{s,t}\widetilde{b}^{(i,j)}_{s,t}\widetilde{b}^{(j,k)}_{s,t}\Big]\\
&=\mathbb{E}\Big[H^{(i)}_{s,t}\,H^{(k)}_{s,t}\Big(12\Big(K^{(i)}_{s,t}\,H^{(j)}_{s,t}-H^{(i)}_{s,t}K^{(j)}_{s,t}\Big)+\widetilde{a}^{(i,j)}_{s,t}\Big)\Big(12\Big(K^{(j)}_{s,t}\,H^{(k)}_{s,t}-H^{(j)}_{s,t}\,K^{(k)}_{s,t}\Big)+\widetilde{a}^{(j,k)}_{s,t}\Big)\Big]\\
&=-144\,\mathbb{E}\Big[\big(H^{(i)}_{s,t}\big)^2\big(H^{(k)}_{s,t}\big)^2\big(K^{(j)}_{s,t}\big)^2\Big]=-\frac{1}{720}h^3,
\end{aligned}\tag{7.3.26}$$

where we have used the fact that $H_{s,t}$, $K_{s,t}$, $\xi^{(i,j)}_{s,t}$, $\xi^{(j,k)}_{s,t}$ are independent and symmetric.

By Theorem 7.2.1 and (7.2.3), we have that all the remaining cross-moments of $\big(H_{s,t}, b_{s,t}\big)$ with degree at most four are zero. On the other hand, by the symmetry and independence of $H_{s,t}$, $K_{s,t}$ and $\xi_{s,t}$, these moments will also be zero for $\big(H_{s,t}, \widetilde{b}_{s,t}\big)$. Thus, it follows that $A_{s,t}\big|W_{s,t}$ and $\widetilde{A}_{s,t}\big|W_{s,t}$ have the same cokurtosis structure. □

Remark 7.3.7. We can interpret $C^{(i)}$ as being the "size" of the Brownian arch $Z^{(i)}$.

Remark 7.3.8. In practice, we can generate $12K_{s,t}^{(i)} \sim \mathcal{N}\big(0, \frac{1}{5}h\big)$ as its own variable.

To conclude this section, we shall numerically demonstrate that the proposed two-dimensional Lévy area approximation is accurate in the 2-Wasserstein metric. (See Chapter 6 of [94] for an account of the Wasserstein distance and its properties)

Definition 7.3.9 (Wasserstein metric)**.** Let μ and ν be probability measures on $\mathbb{R}^d$. Then for any $p \in [1, \infty)$, the p-Wasserstein distance between μ and ν is defined as

$$W_p\big(\mu, \nu\big) := \inf_{(X,Y)\sim(\mu\times\nu)} \Big(\mathbb{E}\big[d(X,Y)^p\big]\Big)^{\frac{1}{p}}, \tag{7.3.27}$$

where $d(\cdot,\cdot)$ is a metric on $\mathbb{R}^d$. The above infimum is taken over all couplings of random variables X and Y with distributions $X \sim \mu$ and $Y \sim \nu$. For our purposes, we will consider the 2-Wasserstein metric associated with the standard $L^2(\mathbb{P})$ norm:

$$W_2\big(\mu, \nu\big) := \inf_{(X,Y)\sim(\mu\times\nu)} \Big(\mathbb{E}\Big[\|X - Y\|_2^2\Big]\Big)^{\frac{1}{2}}. \tag{7.3.28}$$

Recently, there has been significant progress in approximating the Wasserstein distance between empirical measures due to its applications within image processing and machine learning (well-documented open source software can be found at [33]). In particular, if μ and ν are distributions on $\mathbb{R}$ and have cumulative distribution functions F and G, we can consider the following empirical estimator studied in [61]:

Definition 7.3.10 (2-Wasserstein distance estimator for continuous distributions)**.** Let $\{X_i\}_{1\le i\le N}$ and $\{Y_i\}_{1\le i\le N}$ be independent and identically distributed samples from μ and ν respectively (with cdfs F and G). We define the following estimator for $W_2\big(\mu, \nu\big)$:

$$W_2^N := \sqrt{\frac{1}{N}\sum_{i=1}^{N} \big\|X_{(i)} - Y_{(i)}\big\|_2^2}\,, \tag{7.3.29}$$

where $X_{(i)}$ and $Y_{(i)}$ denote the i-th order statistics from the samples $\{X_i\}$ and $\{Y_i\}$.

Remark 7.3.11. The estimator W_2^N is precisely the 2-Wasserstein distance between the empirical distributions corresponding to the samples $\{X_i\}_{1\le i\le N}$ and $\{Y_i\}_{1\le i\le N}$.

As one might expect, this estimator is consistent and will converge almost surely to the true 2-Wasserstein distance $W_2(\mu, \nu)$ almost surely (see Theorem 2.1 in [61]).

Theorem 7.3.12 (Convergence of the empirical 2-Wasserstein estimator)**.** *Suppose that μ and ν are square-integrable, i.e. $\mathbb{E}_{X\sim\mu}[X^2] < \infty$ and $\mathbb{E}_{Y\sim\nu}[Y^2] < \infty$. Then*

$$W_2^N \to W_2(\mu, \nu), \tag{7.3.30}$$

almost surely as $N \to \infty$.

Moreover, with extra assumptions, one can establish an asymptotic convergence rate.

Theorem 7.3.13 (Central limit theorem for the empirical 2-Wasserstein estimator)**.** *Under certain mild conditions on F and G, there exists a constant $\sigma \geq 0$ such that*

$$\sqrt{N}\left(\left(W_2^N\right)^2 - W_2^2(\mu, \nu)\right) \to \mathcal{N}(0, \sigma^2), \tag{7.3.31}$$

as $N \to \infty$, where the above convergence is in distribution. In other words, the estimator W_2^N converges to the true 2-Wasserstein distance weakly with rate $O\left(N^{-\frac{1}{2}}\right)$.

Remark 7.3.14. The above theorem is the main result of [61], where precise conditions on F and G are given along with an explicit formula for the asymptotic variance σ^2.

In practice, we would approximate Lévy area after the increment of the Brownian path has been generated since both quantities appear in the Taylor expansion (7.0.4). Hence in the numerical example, we will investigate the Lévy area approximation (7.3.2) on the interval $[0, 1]$ for a number of different path increments $W_1^{(1)}$ and $W_1^{(2)}$. As the standard deviation of each increment is 1, we shall use the range $\{-2, \cdots, 2\}$.

Table 7.1: Table showing estimated 2-Wasserstein distances between finely discretized Lévy area and the approximation (7.3.2) conditional on certain Brownian increments. Each estimator W_2^N is obtained using $N = 1,000,000$ samples. The discretized Lévy area is computed using the standard piecewise linear approximation with 256 points.

		$W_1^{(2)}$				
		-2	-1	0	1	2
$W_1^{(1)}$	-2	0.0024	0.0019	0.0022	0.0019	0.0028
	-1	0.0018	0.0021	0.0022	0.0018	0.0018
	0	0.0020	0.0019	0.0030	0.0022	0.0020
	1	0.0023	0.0022	0.0020	0.0024	0.0021
	2	0.0026	0.0020	0.0019	0.0023	0.0029

The above results were obtained using a Python implementation (Jupyter Notebook).

Whilst there is significant variation in the above estimated Wasserstein distances, it is clear that they are all have similar orders of magnitude $\big($between 10^{-3} and $10^{-2}\big)$. To give some perspective on this, we recall that the $L^2(\mathbb{P})$ norm of two-dimensional Lévy area $A_{0,1}$, conditional on W_1, is $\|A_{0,1}\|_{L^2(\mathbb{P})} = \sqrt{\frac{1}{12} + \frac{1}{12}\|W_1\|_2^2}$ (remark 7.0.11). Therefore, the 2-Wasserstein distance between the distribution of Brownian Lévy area and the distribution of its best W_1-measurable strong approximation is at least 0.28.

It should be noted that since estimators for Lévy area and 2-Wasserstein distances have slow $O\big(N^{-\frac{1}{2}}\big)$ convergence rates, it is possible that our results have inaccuracies. That said, the numerical evidence is promising and indicates that the proposed weak approximation produces samples that are statistically similar to Brownian Lévy area.

7.4 Comparison with Davie's weak approximation of Brownian Lévy area

In this section, we will show that our approximation of Lévy area is "close" to the approach originally proposed by Davie [27] and later analysed using rough path theory by Flint and Lyons [34]. We first recall Davie's approximation of Brownian Lévy area.

Definition 7.4.1 (Davie's approximation of multidimensional Brownian Lévy area)**.** Let $[s,t]$ be an interval with size $h = t - s$. We define the random matrix $\widehat{A}_{s,t}$ as

$$\widehat{A}_{s,t} = H_{s,t} \otimes W_{s,t} - W_{s,t} \otimes H_{s,t} + \widehat{b}_{s,t}, \tag{7.4.1}$$

where $\widehat{b}_{s,t}$ is an antisymmetric matrix with the following independent entries:

$$\widehat{b}_{s,t}^{(i,j)} \sim \mathcal{N}\bigg(0, \frac{1}{12}h^2\bigg),$$

for $1 \leq i < j \leq d$.

Although Davie's approximation of Lévy area matches the conditional mean (7.1.2) exactly, it uses a constant variance as a proxy for the conditional variance (7.1.3). Thus, we propose the following approximation for the Lévy area of Brownian motion:

Definition 7.4.2 (Weak approximation of multidimensional Brownian Lévy area)**.** Let $[s,t]$ be an interval with size $h = t - s$. We define the random matrix $\widetilde{A}_{s,t}$ as

$$\widetilde{A}_{s,t} = H_{s,t} \otimes W_{s,t} - W_{s,t} \otimes H_{s,t} + \widetilde{b}_{s,t}, \tag{7.4.2}$$

where $\widetilde{b}_{s,t}$ is an antisymmetric matrix with the following independent entries:

$$\widetilde{b}_{s,t}^{(i,j)} \sim \mathcal{N}\Big(0, \frac{1}{20}h^2 + \frac{1}{5}h\Big(\big(H_{s,t}^{(i)}\big)^2 + \big(H_{s,t}^{(j)}\big)^2\Big)\Big),$$

for $1 \le i < j \le d$.

We will show that this choice of variance results in a small $O(h)$ 2-Wasserstein error when we discretize Lévy area on $[0,1]$ using a constant step size h. To this end, we shall construct a probabilistic coupling between Davie's approximation and our's.

Definition 7.4.3 (Coupling sums of approximate Lévy areas over dyadic intervals)**.** For each $N \ge 0$, let $D_N := \{t_k\}_{0 \le k \le 2^N}$ denote the uniform partition of $[0,1]$ with $t_k := kh$ and mesh size $h := 2^{-N}$. We say that a subset $E \subset D_N$ is *dyadic* if $E = \{t_k\}_{m2^{-n} \le k < (m+1)2^{-n}}$ for some integers $n \in \{0, \cdots, N\}$ and $m \in \{0, \cdots, 2^n - 1\}$. Given a dyadic subset E of D_N, we can then define the following sums:

$$\widehat{S}_E^{(i,j)} := \sum_{k \in E} \widehat{A}_{t_k,t_{k+1}}^{(i,j)} \sim \mathcal{N}\Big(\widehat{S}_E^{(i,j)}, \big(\widehat{\sigma}_E^{(i,j)}\big)^2\Big), \tag{7.4.3}$$

$$\widetilde{S}_E^{(i,j)} := \sum_{k \in E} \widetilde{A}_{t_k,t_{k+1}}^{(i,j)} \sim \mathcal{N}\Big(\widehat{S}_E^{(i,j)}, \big(\widetilde{\sigma}_E^{(i,j)}\big)^2\Big), \tag{7.4.4}$$

where

$$\widehat{S}_E^{(i,j)} := \sum_{k \in E} \Big(H_k^{(i)} W_k^{(j)} - W_k^{(i)} H_k^{(j)}\Big), \tag{7.4.5}$$

$$\widehat{\sigma}_E^{(i,j)} := \frac{\sqrt{3}}{6}|E|^{\frac{1}{2}}h, \tag{7.4.6}$$

$$\widetilde{\sigma}_E^{(i,j)} := \left(\frac{1}{20}|E|h^2 + \frac{1}{5}h\sum_{k \in E}\Big(\big(H_{s,t}^{(i)}\big)^2 + \big(H_{s,t}^{(j)}\big)^2\Big)\right)^{\frac{1}{2}}. \tag{7.4.7}$$

for $1 \le i < j \le d$ where W_k and H_k will simply be notation for $W_{t_k,t_{k+1}}$ and $H_{t_k,t_{k+1}}$. When $|E| = 2^N$, we will couple the sums $\widehat{S}_E^{(i,j)}$ and $\widetilde{S}_E^{(i,j)}$ via a standard normal random variable $X^{(i,j)}$. That is, for $1 \le i < j \le d$, we shall generate the pair $\big(\widehat{S}_E^{(i,j)}, \widetilde{S}_E^{(i,j)}\big)$ as

$$\widehat{S}_E^{(i,j)} = \widehat{S}_E^{(i,j)} + \widehat{\sigma}_E^{(i,j)} X^{(i,j)}, \quad \widetilde{S}_E^{(i,j)} = \widehat{S}_E^{(i,j)} + \widetilde{\sigma}_E^{(i,j)} X^{(i,j)}, \tag{7.4.8}$$

where each $X^{(i,j)} \sim \mathcal{N}(0,1)$ is independent. We extend this coupling recursively as

Step 1. Suppose that we have generated the pair $\big(\widehat{S}_E, \widetilde{S}_E\big)$ for some dyadic subset E of D_N with $|E| \ge 2$. Then E is the union of two disjoint dyadic subsets, which we denote by E^- and E^+.

Step 2. For all $1 \le i < j \le d$, we sample an independent normal random variable $X_E^{(i,j)} \sim \mathcal{N}(0,1)$.

Step 3. We generate the pairs $\big(\widehat{S}_{E^-}, \widetilde{S}_{E^-}\big)$ and $\big(\widehat{S}_{E^+}, \widetilde{S}_{E^+}\big)$ as

$$\widehat{S}_{E^-}^{(i,j)} = \widehat{S}_{E^-}^{(i,j)} + \frac{\big(\widehat{\sigma}_{E^-}^{(i,j)}\big)^2}{\big(\widehat{\sigma}_E^{(i,j)}\big)^2}\Big(\widehat{S}_E^{(i,j)} - \widehat{S}_E^{(i,j)}\Big) + \frac{\big(\widehat{\sigma}_{E^-}^{(i,j)}\big)\big(\widehat{\sigma}_{E^+}^{(i,j)}\big)}{\widehat{\sigma}_E^{(i,j)}} X_E^{(i,j)}, \tag{7.4.9}$$

$$\widehat{S}_{E^+}^{(i,j)} = \widehat{S}_{E^+}^{(i,j)} + \frac{\big(\widehat{\sigma}_{E^+}^{(i,j)}\big)^2}{\big(\widehat{\sigma}_E^{(i,j)}\big)^2}\Big(\widehat{S}_E^{(i,j)} - \widehat{S}_E^{(i,j)}\Big) - \frac{\big(\widehat{\sigma}_{E^-}^{(i,j)}\big)\big(\widehat{\sigma}_{E^+}^{(i,j)}\big)}{\widehat{\sigma}_E^{(i,j)}} X_E^{(i,j)}, \tag{7.4.10}$$

$$\widetilde{S}_{E^-}^{(i,j)} = \widehat{S}_{E^-}^{(i,j)} + \frac{\big(\widetilde{\sigma}_{E^-}^{(i,j)}\big)^2}{\big(\widetilde{\sigma}_E^{(i,j)}\big)^2}\Big(\widetilde{S}_E^{(i,j)} - \widehat{S}_E^{(i,j)}\Big) + \frac{\big(\widetilde{\sigma}_{E^-}^{(i,j)}\big)\big(\widetilde{\sigma}_{E^+}^{(i,j)}\big)}{\widetilde{\sigma}_E^{(i,j)}} X_E^{(i,j)}, \tag{7.4.11}$$

$$\widetilde{S}_{E^+}^{(i,j)} = \widehat{S}_{E^+}^{(i,j)} + \frac{\big(\widetilde{\sigma}_{E^+}^{(i,j)}\big)^2}{\big(\widetilde{\sigma}_E^{(i,j)}\big)^2}\Big(\widetilde{S}_E^{(i,j)} - \widehat{S}_E^{(i,j)}\Big) - \frac{\big(\widetilde{\sigma}_{E^-}^{(i,j)}\big)\big(\widetilde{\sigma}_{E^+}^{(i,j)}\big)}{\widetilde{\sigma}_E^{(i,j)}} X_E^{(i,j)}, \tag{7.4.12}$$

for all $1 \le i < j \le d$. We note that the above constructions of $\big(\widehat{S}_{E^-}, \widehat{S}_{E^+}\big)$ and $\big(\widetilde{S}_{E^-}, \widetilde{S}_{E^+}\big)$ clearly have the desired covariance structure and thus follow the normal distributions (7.4.3) and (7.4.4).

Step 4. If $|E| > 2$, we continue the coupling and apply Steps 1-3 to the dyadic subsets E^- and E^+. (There is nothing left to generate if $|E| = 2$).

It is worth noting that the variances for Davie's approximation are constant, $\widehat{\sigma}_{E^-} = \widehat{\sigma}_{E^+} = \frac{\sqrt{2}}{2}\widehat{\sigma}_E$.

Before quantifying the 2-Wasserstein error between our approximation and Davie's, we will estimate the difference between $\widetilde{\sigma}$ and $\widehat{\sigma}$. Hence we first recall the following result, which shows that $x \mapsto \sqrt{x}$ is Lipschitz continuous on $[a, \infty)$ whenever $a > 0$.

Proposition 7.4.4. *For $a > 0$ and $x, y \ge 0$, we have*

$$\big|\sqrt{a+x} - \sqrt{a+y}\,\big|^2 \le \frac{1}{4a}|x-y|^2.$$

Proof. Without loss of generality, we assume that $x \ge y$. Then

$$\sqrt{a+x} - \sqrt{a+y} = \frac{x-y}{\sqrt{a+x}+\sqrt{a+y}} \le \frac{1}{2\sqrt{a}}(x-y),$$

and the result follows. □

Theorem 7.4.5. *Let E, E^-, E^+ be dyadic subsets of D_N with E^- and E^+ disjoint and of equal size. Then for all $1 \le i < j \le d$, we have*

$$\mathbb{E}\Big[\big(\widetilde{\sigma}_E^{(i,j)} - \widehat{\sigma}_E^{(i,j)}\big)^2\Big] \le \frac{1}{180}h^2, \tag{7.4.13}$$

$$\mathbb{E}\Big[\big(\widetilde{\sigma}_{E^-}^{(i,j)} - \widetilde{\sigma}_{E^+}^{(i,j)}\big)^2\Big] \le \frac{1}{90}h^2. \tag{7.4.14}$$

Proof. We shall establish the first inequality by direct calculation. More concretely, we will express the LHS of (7.4.13) using the formulae for $\widetilde{\sigma}$ and $\widehat{\sigma}$ in definition 7.4.3 and then apply Proposition 7.4.4.

$$\begin{aligned}
\mathbb{E}\Big[\big(\widetilde{\sigma}_E^{(i,j)} - \widehat{\sigma}_E^{(i,j)}\big)^2\Big] &= h^2\mathbb{E}\left[\left(\left(\frac{1}{20}|E| + \frac{1}{5h}\sum_{k\in E}\Big(\big(H_{s,t}^{(i)}\big)^2 + \big(H_{s,t}^{(j)}\big)^2\Big)\right)^{\frac{1}{2}} - \left(\frac{1}{12}|E|\right)^{\frac{1}{2}}\right)^2\right] \\
&\leq 5|E|^{-1}h^2\,\mathbb{E}\left[\left(\frac{1}{20}|E| + \frac{1}{5h}\sum_{k\in E}\Big(\big(H_{s,t}^{(i)}\big)^2 + \big(H_{s,t}^{(j)}\big)^2\Big) - \frac{1}{12}|E|\right)^2\right] \\
&= 5|E|^{-1}\mathrm{Var}\left(\frac{1}{5}\sum_{k\in E}\Big(\big(H_k^{(i)}\big)^2 + \big(H_k^{(j)}\big)^2\Big)\right) \\
&\leq \frac{1}{5|E|}\sum_{k\in E}\mathrm{Var}\Big(\big(H_k^{(i)}\big)^2\Big) + \mathrm{Var}\Big(\big(H_k^{(j)}\big)^2\Big) \\
&= \frac{1}{180}h^2.
\end{aligned}$$

Since $\widetilde{\sigma}_{E^-}$ and $\widetilde{\sigma}_{E^+}$ are independent and identically distributed, the second inequality follows as

$$\begin{aligned}
\mathbb{E}\Big[\big(\widetilde{\sigma}_{E^-}^{(i,j)} - \widehat{\sigma}_{E^+}^{(i,j)}\big)^2\Big] &= \mathrm{Var}\big(\widetilde{\sigma}_{E^-}^{(i,j)} - \widetilde{\sigma}_{E^+}^{(i,j)}\big) \\
&= \mathrm{Var}\big(\widetilde{\sigma}_{E^-}^{(i,j)}\big) + \mathrm{Var}\big(\widetilde{\sigma}_{E^+}^{(i,j)}\big) \\
&\leq \mathbb{E}\Big[\big(\widetilde{\sigma}_{E^-}^{(i,j)} - \widehat{\sigma}_{E^-}^{(i,j)}\big)^2\Big] + \mathbb{E}\Big[\big(\widetilde{\sigma}_{E^+}^{(i,j)} - \widehat{\sigma}_{E^+}^{(i,j)}\big)^2\Big] \\
&\leq \frac{1}{90}h^2.
\end{aligned}$$

Alternatively, we could have show the second estimate (7.4.14) by slightly adapting the argument used for (7.4.13). □

We are now in a position to prove the main result of this section, which shows that our method for generating Lévy area can be closely coupled to Davie's approximation.

Theorem 7.4.6. *Let $N \geq 0$ and suppose that we have constructed the approximate Lévy areas $\big\{\widetilde{A}_{t_k,t_{k+1}}, \widehat{A}_{t_k,t_{k+1}}\big\}_{0\leq k<2^N}$ using the coupling procedure in definition 7.4.3. Then for any dyadic subset E of D_N, we have*

$$\left\|\sum_{k\in E}\widetilde{A}_{t_k,t_{k+1}}^{(i,j)} - \sum_{k\in E}\widehat{A}_{t_k,t_{k+1}}^{(i,j)}\right\|_{L^2(\mathbb{P})} \leq \left(\sqrt{\frac{3+2\sqrt{15}}{270}} + \frac{\sqrt{5}}{30}\right)h, \tag{7.4.15}$$

for $1 \leq i < j \leq d$, where $\|\cdot\|_{L^2(\mathbb{P})} := \big(\mathbb{E}\big[|\cdot|^2\big]\big)^{\frac{1}{2}}$ denotes the standard $L^2(\mathbb{P})$ norm.

Proof. Let $\mathcal{F}_N$ denote the σ-algebra that is generated by $\{W_{t_k,t_{k+1}}, H_{t_k,t_{k+1}}\}_{0\le k<2^N}$. When $|E| = 2^N$, we have

$$\begin{aligned}\big\|\widetilde{S}_E^{(i,j)} - \widehat{S}_E^{(i,j)}\big\|_{L^2(\mathbb{P})}^2 &= \mathbb{E}\Big[\big(\widetilde{\sigma}_E^{(i,j)}X^{(i,j)} - \widehat{\sigma}_E^{(i,j)}X^{(i,j)}\big)^2\Big]\\ &= \mathbb{E}\Big[\big(\widetilde{\sigma}_E^{(i,j)} - \widehat{\sigma}_E^{(i,j)}\big)^2\,\mathbb{E}\Big[\big(X^{(i,j)}\big)^2\big|\mathcal{F}_N\Big]\Big]\\ &= \mathbb{E}\Big[\big(\widetilde{\sigma}_E^{(i,j)} - \widehat{\sigma}_E^{(i,j)}\big)^2\Big]\\ &\le \frac{1}{180}h^2.\end{aligned}$$

by the Tower law and Theorem 7.4.5. Now suppose that E is a dyadic subset of D_N with $|E| \ge 2$. Therefore we have $E = E^- \cup E^+$ where E^- and E^+ are disjoint dyadic subsets of equal size and just as in definition 7.4.3, we construct the random variables $\widetilde{S}_{E^-}^{(i,j)}, \widehat{S}_{E^-}^{(i,j)}, \widetilde{S}_{E^+}^{(i,j)}, \widehat{S}_{E^+}^{(i,j)}$ using a coupling. Before we derive estimates for the differences $\widetilde{S}_{E^-}^{(i,j)} - \widehat{S}_{E^-}^{(i,j)}$ and $\widetilde{S}_{E^+}^{(i,j)} - \widehat{S}_{E^+}^{(i,j)}$, we shall first consider the following quantities:

$$\begin{aligned}\mathcal{E}_{E^-}^{(i,j)} &:= \left(\frac{\big(\widetilde{\sigma}_{E^-}^{(i,j)}\big)^2}{\big(\widetilde{\sigma}_E^{(i,j)}\big)^2}\Big(\widetilde{S}_E^{(i,j)} - \widehat{S}_E^{(i,j)}\Big) + \frac{\big(\widetilde{\sigma}_{E^-}^{(i,j)}\big)\big(\widetilde{\sigma}_{E^+}^{(i,j)}\big)}{\widetilde{\sigma}_E^{(i,j)}}X_E^{(i,j)}\right)\\ &\quad - \left(\frac{\big(\widehat{\sigma}_{E^-}^{(i,j)}\big)^2}{\big(\widehat{\sigma}_E^{(i,j)}\big)\big(\widetilde{\sigma}_E^{(i,j)}\big)}\Big(\widetilde{S}_E^{(i,j)} - \widehat{S}_E^{(i,j)}\Big) + \frac{\big(\widehat{\sigma}_{E^-}^{(i,j)}\big)\big(\widehat{\sigma}_{E^+}^{(i,j)}\big)}{\widehat{\sigma}_E^{(i,j)}}X_E^{(i,j)}\right),\\ \mathcal{E}_{E^+}^{(i,j)} &:= \left(\frac{\big(\widetilde{\sigma}_{E^+}^{(i,j)}\big)^2}{\big(\widetilde{\sigma}_E^{(i,j)}\big)^2}\Big(\widetilde{S}_E^{(i,j)} - \widehat{S}_E^{(i,j)}\Big) - \frac{\big(\widetilde{\sigma}_{E^-}^{(i,j)}\big)\big(\widetilde{\sigma}_{E^+}^{(i,j)}\big)}{\widetilde{\sigma}_E^{(i,j)}}X_E^{(i,j)}\right)\\ &\quad - \left(\frac{\big(\widehat{\sigma}_{E^+}^{(i,j)}\big)^2}{\big(\widehat{\sigma}_E^{(i,j)}\big)\big(\widetilde{\sigma}_E^{(i,j)}\big)}\Big(\widetilde{S}_E^{(i,j)} - \widehat{S}_E^{(i,j)}\Big) - \frac{\big(\widehat{\sigma}_{E^-}^{(i,j)}\big)\big(\widehat{\sigma}_{E^+}^{(i,j)}\big)}{\widehat{\sigma}_E^{(i,j)}}X_E^{(i,j)}\right).\end{aligned}$$

To estimate $\mathcal{E}_{E^-}^{(i,j)}$ and $\mathcal{E}_{E^+}^{(i,j)}$, we note that

$$\mathrm{Var}\big(\mathcal{E}_{E^-}^{(i,j)}\big) + \mathrm{Var}\big(\mathcal{E}_{E^+}^{(i,j)}\big) = \mathrm{Var}\big(\mathcal{E}_{E^-}^{(i,j)} + \mathcal{E}_{E^+}^{(i,j)}\big) - 2\,\mathrm{Cov}\big(\mathcal{E}_{E^-}^{(i,j)}, \mathcal{E}_{E^+}^{(i,j)}\big). \tag{7.4.16}$$

The first term is straightforward to estimate using the Tower law and Theorem 7.4.5.

$$\begin{aligned}\mathrm{Var}\big(\mathcal{E}_{E^-}^{(i,j)} + \mathcal{E}_{E^+}^{(i,j)}\big) &= \mathbb{E}\left[\left(\frac{\big(\widetilde{\sigma}_{E^-}^{(i,j)}\big)^2 + \big(\widetilde{\sigma}_{E^+}^{(i,j)}\big)^2}{\big(\widetilde{\sigma}_E^{(i,j)}\big)^2} - \frac{\big(\widehat{\sigma}_{E^-}^{(i,j)}\big)^2 + \big(\widehat{\sigma}_{E^+}^{(i,j)}\big)^2}{\big(\widehat{\sigma}_E^{(i,j)}\big)\big(\widetilde{\sigma}_E^{(i,j)}\big)}\right)^2\Big(\widetilde{S}_E^{(i,j)} - \widehat{S}_E^{(i,j)}\Big)^2\right]\\ &= \mathbb{E}\left[\left(\frac{\big(\widetilde{\sigma}_E^{(i,j)}\big)^2}{\big(\widetilde{\sigma}_E^{(i,j)}\big)^2} - \frac{\big(\widehat{\sigma}_E^{(i,j)}\big)^2}{\big(\widehat{\sigma}_E^{(i,j)}\big)\big(\widetilde{\sigma}_E^{(i,j)}\big)}\right)^2\mathbb{E}\left[\Big(\widetilde{S}_E^{(i,j)} - \widehat{S}_E^{(i,j)}\Big)^2\Big|\mathcal{F}_N\right]\right]\\ &= \mathbb{E}\left[\left(1 - \frac{\widehat{\sigma}_E^{(i,j)}}{\widetilde{\sigma}_E^{(i,j)}}\right)^2\big(\widetilde{\sigma}_E^{(i,j)}\big)^2\right]\\ &= \mathbb{E}\Big[\big(\widetilde{\sigma}_E^{(i,j)} - \widehat{\sigma}_E^{(i,j)}\big)^2\Big],\end{aligned}$$

and so by Theorem 7.4.5, we have

$$\mathrm{Var}\big(\mathcal{E}_{E^-}^{(i,j)} + \mathcal{E}_{E^+}^{(i,j)}\big) = \mathbb{E}\Big[\big(\widetilde{\sigma}_E^{(i,j)} - \widehat{\sigma}_E^{(i,j)}\big)^2\Big] \leq \frac{1}{180}h^2.$$

We can simplify the covariance term in (7.4.16) since $\big(\widetilde{S}_E^{(i,j)}, \widehat{S}_E^{(i,j)}\big)$ and $X_E^{(i,j)}$ are uncorrelated conditional on $\mathcal{F}_N$.

$$\begin{aligned}
&\mathrm{Cov}\big(\mathcal{E}_{E^-}^{(i,j)}, \mathcal{E}_{E^+}^{(i,j)}\big) \\
&\quad= \mathbb{E}\Bigg[\mathbb{E}\Bigg[\Bigg(\Bigg(\frac{\big(\widetilde{\sigma}_{E^-}^{(i,j)}\big)^2}{\big(\widetilde{\sigma}_E^{(i,j)}\big)^2} - \frac{\big(\widehat{\sigma}_{E^-}^{(i,j)}\big)^2}{\big(\widehat{\sigma}_E^{(i,j)}\big)\big(\widetilde{\sigma}_E^{(i,j)}\big)}\Bigg)\Big(\widetilde{S}_E^{(i,j)} - \widehat{S}_E^{(i,j)}\Big)^2 \\
&\qquad\qquad + \Bigg(\frac{\big(\widetilde{\sigma}_{E^-}^{(i,j)}\big)\big(\widetilde{\sigma}_{E^+}^{(i,j)}\big)}{\widetilde{\sigma}_E^{(i,j)}} - \frac{\big(\widehat{\sigma}_{E^-}^{(i,j)}\big)\big(\widehat{\sigma}_{E^+}^{(i,j)}\big)}{\widehat{\sigma}_E^{(i,j)}}\Bigg)X_E^{(i,j)}\Bigg) \\
&\qquad\cdot\Bigg(\Bigg(\frac{\big(\widetilde{\sigma}_{E^+}^{(i,j)}\big)^2}{\big(\widetilde{\sigma}_E^{(i,j)}\big)^2} - \frac{\big(\widehat{\sigma}_{E^+}^{(i,j)}\big)^2}{\big(\widehat{\sigma}_E^{(i,j)}\big)\big(\widetilde{\sigma}_E^{(i,j)}\big)}\Bigg)\Big(\widetilde{S}_E^{(i,j)} - \widehat{S}_E^{(i,j)}\Big)^2 \\
&\qquad\qquad - \Bigg(\frac{\big(\widetilde{\sigma}_{E^-}^{(i,j)}\big)\big(\widetilde{\sigma}_{E^+}^{(i,j)}\big)}{\widetilde{\sigma}_E^{(i,j)}} - \frac{\big(\widehat{\sigma}_{E^-}^{(i,j)}\big)\big(\widehat{\sigma}_{E^+}^{(i,j)}\big)}{\widehat{\sigma}_E^{(i,j)}}\Bigg)X_E^{(i,j)}\Bigg)\Bigg|\mathcal{F}_N\Bigg]\Bigg] \\
&\quad= \mathbb{E}\Bigg[\Bigg(\frac{\big(\widetilde{\sigma}_{E^-}^{(i,j)}\big)^2}{\big(\widetilde{\sigma}_E^{(i,j)}\big)^2} - \frac{\big(\widehat{\sigma}_{E^-}^{(i,j)}\big)^2}{\big(\widehat{\sigma}_E^{(i,j)}\big)\big(\widetilde{\sigma}_E^{(i,j)}\big)}\Bigg)\Bigg(\frac{\big(\widetilde{\sigma}_{E^+}^{(i,j)}\big)^2}{\big(\widetilde{\sigma}_E^{(i,j)}\big)^2} - \frac{\big(\widehat{\sigma}_{E^+}^{(i,j)}\big)^2}{\big(\widehat{\sigma}_E^{(i,j)}\big)\big(\widetilde{\sigma}_E^{(i,j)}\big)}\Bigg) \\
&\qquad\cdot\mathbb{E}\Big[\Big(\widetilde{S}_E^{(i,j)} - \widehat{S}_E^{(i,j)}\Big)^2\Big|\mathcal{F}_N\Big]\Bigg] \\
&\qquad- \mathbb{E}\Bigg[\Bigg(\frac{\big(\widetilde{\sigma}_{E^-}^{(i,j)}\big)\big(\widetilde{\sigma}_{E^+}^{(i,j)}\big)}{\widetilde{\sigma}_E^{(i,j)}} - \frac{\big(\widehat{\sigma}_{E^-}^{(i,j)}\big)\big(\widehat{\sigma}_{E^+}^{(i,j)}\big)}{\widehat{\sigma}_E^{(i,j)}}\Bigg)^2\mathbb{E}\Big[\big(X_E^{(i,j)}\big)^2\big|\mathcal{F}_N\Big]\Bigg] \\
&\quad= \mathbb{E}\Bigg[\Bigg(\frac{\big(\widetilde{\sigma}_{E^-}^{(i,j)}\big)^2}{\widetilde{\sigma}_E^{(i,j)}} - \frac{1}{2}\widehat{\sigma}_E^{(i,j)}\Bigg)\Bigg(\frac{\big(\widetilde{\sigma}_{E^+}^{(i,j)}\big)^2}{\widetilde{\sigma}_E^{(i,j)}} - \frac{1}{2}\widehat{\sigma}_E^{(i,j)}\Bigg) - \Bigg(\frac{\big(\widetilde{\sigma}_{E^-}^{(i,j)}\big)\big(\widetilde{\sigma}_{E^+}^{(i,j)}\big)}{\widetilde{\sigma}_E^{(i,j)}} - \frac{1}{2}\widehat{\sigma}_E^{(i,j)}\Bigg)^2\Bigg] \\
&\quad= -\frac{1}{2}\widehat{\sigma}_E^{(i,j)}\,\mathbb{E}\Bigg[\frac{\big(\widetilde{\sigma}_{E^-}^{(i,j)} - \widetilde{\sigma}_{E^+}^{(i,j)}\big)^2}{\widetilde{\sigma}_E^{(i,j)}}\Bigg].
\end{aligned}$$

It directly follows from definition 7.4.3 that $\widetilde{\sigma}_E^{(i,j)} \geq \frac{\sqrt{5}}{10}|E|^{\frac{1}{2}}h$ and $\widehat{\sigma}_E^{(i,j)} = \frac{\sqrt{3}}{6}|E|^{\frac{1}{2}}h$. Thus, by Theorem 7.4.5, we have

$$-2\,\mathrm{Cov}\big(\mathcal{E}_{E^-}^{(i,j)}, \mathcal{E}_{E^+}^{(i,j)}\big) = \widehat{\sigma}_E^{(i,j)}\,\mathbb{E}\Bigg[\frac{\big(\widetilde{\sigma}_{E^-}^{(i,j)} - \widetilde{\sigma}_{E^+}^{(i,j)}\big)^2}{\widetilde{\sigma}_E^{(i,j)}}\Bigg] \leq \frac{\sqrt{15}}{3}\mathbb{E}\Big[\big(\widetilde{\sigma}_{E^-}^{(i,j)} - \widetilde{\sigma}_{E^+}^{(i,j)}\big)^2\Big] \leq \frac{\sqrt{15}}{270}h^2,$$

Plugging these estimates into (7.4.16) gives

$$\mathrm{Var}\big(\mathcal{E}_{E^-}^{(i,j)}\big) + \mathrm{Var}\big(\mathcal{E}_{E^+}^{(i,j)}\big) \leq \frac{3 + 2\sqrt{15}}{540}h^2.$$

Since $\mathcal{E}_{E^-}^{(i,j)}$ and $\mathcal{E}_{E^+}^{(i,j)}$ have the same symmetric distribution, we obtain the estimates:

$$\left\|\mathcal{E}_{E^-}^{(i,j)}\right\|_{L^2(\mathbb{P})} = \left\|\mathcal{E}_{E^+}^{(i,j)}\right\|_{L^2(\mathbb{P})} \leq \sqrt{\frac{3+2\sqrt{15}}{1080}}h. \tag{7.4.17}$$

We can now estimate $\widetilde{S}_{E^-}^{(i,j)} - \widehat{S}_{E^-}^{(i,j)}$ using the coupling in definition 7.4.3 followed by Minkowski's inequality.

$$\begin{aligned}
&\left\|\widetilde{S}_{E^-}^{(i,j)} - \widehat{S}_{E^-}^{(i,j)}\right\|_{L^2(\mathbb{P})} \\
&\quad= \left\| \left(\frac{\left(\widetilde{\sigma}_{E^-}^{(i,j)}\right)^2}{\left(\widetilde{\sigma}_{E}^{(i,j)}\right)^2}\left(\widetilde{S}_{E}^{(i,j)} - \widehat{S}_{E}^{(i,j)}\right) + \frac{\left(\widetilde{\sigma}_{E^-}^{(i,j)}\right)\left(\widetilde{\sigma}_{E^+}^{(i,j)}\right)}{\widetilde{\sigma}_{E}^{(i,j)}} X_E^{(i,j)} \right) \right. \\
&\qquad\qquad \left. - \left(\frac{\left(\widehat{\sigma}_{E^-}^{(i,j)}\right)^2}{\left(\widehat{\sigma}_{E}^{(i,j)}\right)^2}\left(\widehat{S}_{E}^{(i,j)} - \widehat{S}_{E}^{(i,j)}\right) + \frac{\left(\widehat{\sigma}_{E^-}^{(i,j)}\right)\left(\widehat{\sigma}_{E^+}^{(i,j)}\right)}{\widehat{\sigma}_{E}^{(i,j)}} X_E^{(i,j)} \right) \right\|_{L^2(\mathbb{P})} \\
&\quad\leq \left\| \left(\frac{\left(\widetilde{\sigma}_{E^-}^{(i,j)}\right)^2}{\left(\widetilde{\sigma}_{E}^{(i,j)}\right)^2}\left(\widetilde{S}_{E}^{(i,j)} - \widehat{S}_{E}^{(i,j)}\right) + \frac{\left(\widetilde{\sigma}_{E^-}^{(i,j)}\right)\left(\widetilde{\sigma}_{E^+}^{(i,j)}\right)}{\widetilde{\sigma}_{E}^{(i,j)}} X_E^{(i,j)} \right) \right. \\
&\qquad\qquad \left. - \left(\frac{\left(\widehat{\sigma}_{E^-}^{(i,j)}\right)^2}{\left(\widehat{\sigma}_{E}^{(i,j)}\right)\left(\widetilde{\sigma}_{E}^{(i,j)}\right)}\left(\widetilde{S}_{E}^{(i,j)} - \widehat{S}_{E}^{(i,j)}\right) + \frac{\left(\widehat{\sigma}_{E^-}^{(i,j)}\right)\left(\widehat{\sigma}_{E^+}^{(i,j)}\right)}{\widehat{\sigma}_{E}^{(i,j)}} X_E^{(i,j)} \right) \right\|_{L^2(\mathbb{P})} \\
&\qquad + \left\| \frac{\left(\widehat{\sigma}_{E^-}^{(i,j)}\right)^2}{\left(\widehat{\sigma}_{E}^{(i,j)}\right)\left(\widetilde{\sigma}_{E}^{(i,j)}\right)}\left(\widetilde{S}_{E}^{(i,j)} - \widehat{S}_{E}^{(i,j)}\right) - \frac{\left(\widehat{\sigma}_{E^-}^{(i,j)}\right)^2}{\left(\widehat{\sigma}_{E}^{(i,j)}\right)^2}\widetilde{S}_{E}^{(i,j)} \right\|_{L^2(\mathbb{P})} \\
&\qquad + \left\| \frac{\left(\widehat{\sigma}_{E^-}^{(i,j)}\right)^2}{\left(\widehat{\sigma}_{E}^{(i,j)}\right)^2}\left(\widetilde{S}_{E}^{(i,j)} - \widehat{S}_{E}^{(i,j)}\right) - \frac{\left(\widehat{\sigma}_{E^-}^{(i,j)}\right)^2}{\left(\widehat{\sigma}_{E}^{(i,j)}\right)^2}\widehat{S}_{E}^{(i,j)} \right\|_{L^2(\mathbb{P})} \\
&\quad= \left\|\mathcal{E}_{E^-}^{(i,j)}\right\|_{L^2(\mathbb{P})} + \left\| \frac{\widehat{\sigma}_{E}^{(i,j)}}{2\widetilde{\sigma}_{E}^{(i,j)}}\left(\widetilde{S}_{E}^{(i,j)} - \widehat{S}_{E}^{(i,j)}\right) - \frac{1}{2}\left(\widetilde{S}_{E}^{(i,j)} - \widehat{S}_{E}^{(i,j)}\right) \right\|_{L^2(\mathbb{P})} \\
&\qquad\qquad + \frac{1}{2}\left\|\widetilde{S}_{E}^{(i,j)} - \widehat{S}_{E}^{(i,j)}\right\|_{L^2(\mathbb{P})}.
\end{aligned}$$

The second term can be simplified as

$$\begin{aligned}
\left\| \frac{\widehat{\sigma}_{E}^{(i,j)}}{2\widetilde{\sigma}_{E}^{(i,j)}}\widetilde{S}_{E}^{(i,j)} - \frac{1}{2}\widetilde{S}_{E}^{(i,j)} \right\|_{L^2(\mathbb{P})}^2 &= \mathbb{E}\left[\left(\frac{\widehat{\sigma}_{E}^{(i,j)}}{2\widetilde{\sigma}_{E}^{(i,j)}} - \frac{1}{2} \right)^2 \mathbb{E}\left[\left(\widetilde{S}_{E}^{(i,j)}\right)^2 \middle| \mathcal{F}_N \right] \right] \\
&= \mathbb{E}\left[\left(\frac{\widehat{\sigma}_{E}^{(i,j)}}{2\widetilde{\sigma}_{E}^{(i,j)}} - \frac{1}{2} \right)^2 \left(\widetilde{\sigma}_{E}^{(i,j)}\right)^2 \right] \\
&= \frac{1}{4}\mathbb{E}\left[\left(\widehat{\sigma}_{E}^{(i,j)} - \widetilde{\sigma}_{E}^{(i,j)}\right)^2 \right].
\end{aligned}$$

Therefore by Theorem 7.4.5 along with the previous estimates (7.4.17), we have that

$$\left\|\widetilde{S}_{E^-}^{(i,j)} - \widehat{S}_{E^-}^{(i,j)}\right\|_{L^2(\mathbb{P})} \leq \frac{1}{2}\left\|\widetilde{S}_{E}^{(i,j)} - \widehat{S}_{E}^{(i,j)}\right\|_{L^2(\mathbb{P})} + \left(\sqrt{\frac{3+2\sqrt{15}}{1080}} + \frac{\sqrt{5}}{60}\right)h.$$

Similarly, the above estimate will also hold for $\left\|\widetilde{S}_{E^+}^{(i,j)} - \widehat{S}_{E^+}^{(i,j)}\right\|_{L^2(\mathbb{P})}$.
Recall that when $|E| = 2^N$, we have $\left\|\widetilde{S}_{E}^{(i,j)} - \widehat{S}_{E}^{(i,j)}\right\|_{L^2(\mathbb{P})}^2 \leq \frac{1}{180}h^2$ and so

$$\left\|\widetilde{S}_{E}^{(i,j)} - \widehat{S}_{E}^{(i,j)}\right\|_{L^2(\mathbb{P})} \leq 2\left(\sqrt{\frac{3+2\sqrt{15}}{1080}} + \frac{\sqrt{5}}{60}\right)h,$$

for all dyadic subsets E of D_N. □

Hence by combining the above result with Theorem 1 from [27], we can establish an $O(h)$ 2-Wasserstein convergence rate for our approximation of Brownian Lévy area.

Corollary 7.4.7. *There exists a constant $C > 0$ such that for each $N \geq 0$, we can define the independent random variables $\{\widetilde{A}_{t_k,t_{k+1}}\}_{0\leq k<2^N}$ given by definition 7.4.2 (with $t_k := kh$ and $h := 2^{-N}$) on the same probability space as a d-dimensional Brownian motion W so that*

$$\left\|\sum_{k=0}^{2^N-1}\left(\frac{1}{2}\left(W_{t_k}^{(i)} + W_{t_{k+1}}^{(i)}\right)W_{t_k,t_{k+1}}^{(j)} + \widetilde{A}_{t_k,t_{k+1}}^{(i,j)}\right) - \int_0^1 W_t^{(i)}dW_t^{(j)}\right\|_{L^2(\mathbb{P})} \leq Ch, \quad (7.4.18)$$

for $1 \leq i < j \leq d$.

Proof. Recall that the $y_t = \int_0^t W_s \otimes dW_s$ is the solution to the following Itô SDE:

$$dx_t = dW_t,$$
$$dy_t = x_t \otimes dW_t,$$

with $x_0 = y_0 = 0$. By Theorem 1 in [27], there exists a constant $\widehat{C} > 0$ such that for each $N \geq 0$, we can define the independent random variables $\{\widehat{A}_{t_k,t_{k+1}}\}_{0\leq k<2^N}$ given by definition 7.4.1 (with $t_k := kh$ and $h := 2^{-N}$) on the same probability space as a d-dimensional Brownian motion W so that

$$\left\|Y_1^N - \int_0^1 W_t \otimes dW_t\right\|_{L^2(\mathbb{P})} \leq \widehat{C}h, \quad (7.4.19)$$

where $\left(X^N, Y^N\right)$ is the following Milstein approximation,

$$\begin{pmatrix} X_{t_{k+1}}^N \\ Y_{t_{k+1}}^N \end{pmatrix} = \begin{pmatrix} X_{t_k}^N \\ Y_{t_k}^N \end{pmatrix} + \begin{pmatrix} 0 & 0 \\ 0 & X_{t_k}^N \end{pmatrix} \otimes W_{t_k,t_{k+1}} + \begin{pmatrix} 0 \\ \frac{1}{2}\left(W_{t_k,t_{k+1}} \otimes W_{t_k,t_{k+1}} - hI_d\right) + \widehat{A}_{t_k,t_{k+1}} \end{pmatrix},$$

for $k \geq 0$ with $X_0^N = Y_0^N = 0$. It then follows that Y_1^N can be expressed as

$$\left(Y_1^N\right)^{(i,j)} = \sum_{k=0}^{2^N-1} \left(\frac{1}{2} \left(W_{t_k}^{(i)} + W_{t_{k+1}}^{(i)} \right) W_{t_k,t_{k+1}}^{(j)} + \widehat{A}_{t_k,t_{k+1}}^{(i,j)} \right).$$

Let $\left\{ \widetilde{A}_{t_k,t_{k+1}} \right\}_{0 \leq k < 2^N}$ be coupled with $\left\{ \widehat{A}_{t_k,t_{k+1}} \right\}_{0 \leq k < 2^N}$ using the procedure given by definition 7.4.3. The result follows by Theorem 7.4.6 and Minkowski's inequality. □

In conclusion, there are two key results that have been discussed in this chapter.

- **Conditional moments of multidimensional Brownian Lévy area**
  The second iterated integral of Brownian motion is an important object in the study of SDEs since it appears naturally within stochastic Taylor expansions. We have a computed a number of moments for this iterated integral conditional on certain path information (such as its increment and space-time Lévy area).

- **Weak approximation of multidimensional Brownian Lévy area**
  By generating normal random variables that match such conditional moments, we can approximate Lévy area with an $O(h)$ rate (in the 2-Wasserstein metric). In addition, we have developed approximations for low-dimensional Brownian Lévy area ($d = 2, 3$) that also capture the conditional cokurtosis structure. In the $d = 2$ setting, we have numerically demonstrated the accuracy of this approach using the natural empirical estimator for the 2-Wasserstein distance.

Furthermore, this research leads to the following open questions related to Lévy area.

- Can we show similar numerical Wasserstein errors for the $d = 3$ approximation?
- Is it possible to establish high order 2-Wasserstein convergence for $\widetilde{A}$ directly? (i.e. without having to compare with Davie's weak approximation $\widehat{A}$ from [27])
- How might one construct a piecewise linear path $\gamma : [s, t] \to \mathbb{R}^d$ that matches an approximate Lévy area and gives a high order numerical method for SDEs?
- Which conditional expectations can be computed for triple iterated integrals?
- Since $(W_{s,t}, H_{s,t})$ and $(W_{s,t}, H_{s,t}, K_{s,t})$ can be refined, would it be reasonable to use the proposed Lévy area approximations in a variable step size method?

  Is this approach worthwhile and when would such an approximation converge?
- Could this research be used for PDE numerics? (see "Group Laplacian" in [40])

Appendix A

Computing iterated integrals of piecewise linear paths

In this appendix, we will present some basic calculations that are used by Chapter 5.

Lemma A.0.1. *Let X be an affine path that is defined on the interval $[s,t]$. Then,*

$$\int_s^t X_u\,du = \frac{1}{2}\big(X_s + X_t\big)(t-s),$$
$$\int_s^t X_u^2\,du = \frac{1}{3}\big(X_s^2 + X_sX_t + X_t^2\big)(t-s),$$
$$\int_s^t (u-s)X_u\,du = \frac{1}{6}\big(X_s + 2X_t\big)(t-s)^2.$$

Proof. Since $X_u = X_s + \frac{u-s}{t-s}(X_t - X_s)$, we can directly compute the above integrals.

$$\begin{aligned}\int_s^t X_u^2\,du &= \int_s^t \left(X_s + \frac{u-s}{t-s}(X_t - X_s)\right)^2 du \\ &= \left(X_s^2 + X_s(X_t - X_s) + \frac{1}{3}(X_t - X_s)^2\right)(t-s).\end{aligned}$$

$$\begin{aligned}\int_s^t (u-s)X_u\,du &= \int_s^t (u-s)\left(X_s + \frac{u-s}{t-s}(X_t - X_s)\right) du \\ &= \left(\frac{1}{2}X_s + \frac{1}{3}(X_t - X_s)\right)(t-s)^2.\end{aligned}$$

□

Using the above lemma, we shall compute integrals of paths with two linear pieces.

Definition A.0.2. For $a \in (0,1)$ and $b \in \mathbb{R}$, we will define the path $\gamma^{a,b}$ on $[s,t]$ by

$$\gamma^{a,b} := \begin{cases} \gamma_s + \frac{b}{ah}(u-s) & \text{for } u \in [s, s+ah] \\ \gamma_s + b + \frac{\gamma_{s,t}^{a,b} - b}{(1-a)h}(u - (s+ah)) & \text{for } u \in [s+ah, t] \end{cases}$$

where $h = t - s$.

Theorem A.0.3. *Let $\gamma^{a,b}$ be a piecewise linear path given by definition A.0.2. Then*

$$\int_s^t \gamma_{s,u}^{a,b}\,du = \frac{1}{2}bh + \frac{1}{2}(1-a)\gamma_{s,t}^{a,b}h, \tag{A.0.1}$$

$$\int_s^t \left(\gamma_{s,u}^{a,b}\right)^2 du = \frac{1}{3}ab^2h + \frac{1}{3}(1-a)\left(b^2 + b\gamma_{s,t}^{a,b} + \left(\gamma_{s,t}^{a,b}\right)^2\right)h, \tag{A.0.2}$$

$$\int_s^t (u-s)\gamma_{s,u}^{a,b}\,du = \frac{1}{3}a^2bh^2 + \frac{1}{6}(1-a)(b+2ab+(2+a)\gamma_{s,t})h^2, \tag{A.0.3}$$

where $h = t - s$.

Proof. Clearly, the first two integrals of $\gamma^{a,b}$ follow immediately from lemma A.0.1. Hence, all that remains is to compute the third integral. By lemma A.0.1, we have

$$\begin{aligned}\int_s^t (u-s)\gamma_{s,u}^{a,b}\,du &= \int_s^{s+ah} (u-s)\left(\frac{b}{ah}(u-s)\right)du \\ &\quad + ah\int_{s+ah}^t b + \frac{\gamma_{s,t}^{a,b}-b}{(1-a)h}(u-(s+ah))\,du \\ &\quad + \int_{s+ah}^t (u-(s+ah))\left(b + \frac{\gamma_{s,t}^{a,b}-b}{(1-a)h}(u-(s+ah))\right)du \\ &= \frac{1}{3}a^2bh^2 + \frac{1}{2}(b+\gamma_{s,t})a(1-a)h^2 + \frac{1}{6}(b+2\gamma_{s,t})(1-a)^2h^2.\end{aligned}$$

The result now follows by simplifying the above terms. □

Corollary A.0.4. *The "space-time" and "space-time-time" Lévy areas of $\gamma^{a,b}$ are*

$$\int_s^t \gamma_{s,u}^{a,b} - \frac{u-s}{h}\gamma_{s,t}^{a,b}\,du = \frac{1}{2}\left(b - a\gamma_{s,t}^{a,b}\right)h,$$

$$\int_s^t \left(\frac{1}{2}h - (u-s)\right)\left(\gamma_{s,u}^{a,b} - \frac{u-s}{h}\gamma_{s,t}^{a,b}\right)du = \frac{1}{12}(1-2a)\left(b - a\gamma_{s,t}^{a,b}\right)h^2.$$

Proof. The first integral follows immediately from (A.0.1). The second integral is

$$\begin{aligned}&\int_s^t \left(\frac{1}{2}h - (u-s)\right)\left(\gamma_{s,u}^{a,b} - \frac{u-s}{h}\gamma_{s,t}^{a,b}\right)du \\ &= \frac{1}{4}\left(b - a\gamma_{s,t}^{a,b}\right)h^2 - \left(\frac{1}{3}a^2bh^2 + \frac{1}{6}(1-a)(b+2ab+(2+a)\gamma_{s,t})h^2\right) + \frac{1}{3}\gamma_{s,t}^{a,b}h^2 \\ &= \frac{1}{12}bh^2 - \frac{1}{12}a\gamma_{s,t}^{a,b}h^2 - \frac{1}{6}abh^2 + \frac{1}{6}a^2\gamma_{s,t}^{a,b}h^2.\end{aligned}$$

The result now follows by factorizing the above terms. □

In particular, the above calculation show that if $\gamma^{a,b}$ has no space-time-time Lévy area and is not a constant path, then a must equal $\frac{1}{2}$ (which proves Theorem 5.2.1).

We will now perform similar calculations for paths γ in the three-parameter family:

Definition A.0.5. For $a \in (0, \frac{1}{2})$ and $b, c \in \mathbb{R}$, we define the path $\gamma^{a,b,c}$ on $[s,t]$ by

$$\gamma^{a,b,c} := \begin{cases} \gamma_s + \frac{b}{ah}(u-s) & \text{for } u \in [s, s+ah] \\ \gamma_s + b + \frac{c-b}{(1-2a)h}(u-(s+ah)) & \text{for } u \in [s+ah, t-ah] \\ \gamma_s + c + \frac{\gamma_{s,t}-c}{ah}(u-(t-ah)) & \text{for } u \in [t-ah, t] \end{cases}$$

where $h = t - s$.

Theorem A.0.6. *Let γ be a piecewise linear path given by the above definition. Then*

$$\int_s^t \gamma_{s,u}\,du = \frac{1}{2}abh + \frac{1}{2}(1-2a)(b+c)h + \frac{1}{2}a\big(c+\gamma_{s,t}\big)h, \tag{A.0.4}$$

$$\int_s^t (\gamma_{s,u})^2\,du = \frac{1}{3}(1-2a)\big(b^2+bc+c^2\big)h + \frac{1}{3}a\big(b^2+c^2+c\gamma_{s,t}+\gamma_{s,t}^2\big)h, \tag{A.0.5}$$

$$\int_s^t (u-s)\gamma_{s,u}\,du = \frac{1}{3}a^2bh^2 + \frac{1}{6}(1-2a)((1+a)b+(2-a)c)h^2 + \frac{1}{6}a\big((3-2a)c+(3-a)\gamma_{s,t}\big)h^2. \tag{A.0.6}$$

Proof. As before, the first two integrals of γ follow immediately from lemma A.0.1. The third integral of γ can be directly computed using (A.0.1) and (A.0.3) as follows:

$$\begin{aligned} \int_s^t (u-s)\gamma_{s,u}\,du &= \int_s^{s+ah} (u-s)\Big(\frac{b}{ah}(u-s)\Big)\,du \\ &\quad + \int_{s+ah}^{t-ah} \big(u-(s+ah)\big)\Big(b + \frac{c-b}{(1-2a)h}(u-(s+ah))\Big)\,du \\ &\quad + ah\int_{s+ah}^{t-ah} b + \frac{c-b}{(1-2a)h}(u-(s+ah))\,du \\ &\quad + \int_{t-ah}^{t} \big(u-(t-ah)\big)\Big(c + \frac{\gamma_{s,t}-c}{ah}(u-(t-ah))\Big)\,du \\ &\quad + (1-a)h\int_{t-ah}^{t} c + \frac{\gamma_{s,t}^{a,b}-c}{ah}(u-(t-ah))\,du \\ &= \frac{1}{3}a^2bh^2 + \frac{1}{6}(b+2c)(1-2a)^2h^2 + \frac{1}{2}(b+c)a(1-2a)h^2 \\ &\quad + \frac{1}{6}(c+2\gamma_{s,t})a^2h^2 + \frac{1}{2}(c+\gamma_{s,t})a(1-a)h^2 \\ &= \frac{1}{3}a^2bh^2 + \frac{1}{6}(1-2a)\big((b+2c)(1-2a)+3(b+c)a\big)h^2 \\ &\quad + \frac{1}{6}a\big((c+2\gamma_{s,t})a+3(c+\gamma_{s,t})(1-a)\big)h^2. \end{aligned}$$

The result now follows by simplifying the above terms. □

Bibliography

[1] L. Abadie and J. M. Chamorro. Valuing flexibility: The case of an Integrated Gasification Combined Cycle power plant. *Energy Economics*, 30(4):1850–1881, 2008.

[2] A. Alfonsi. On the discretization schemes for the CIR (and Bessel squared) processes. *Monte Carlo Methods and Applications*, 11(4):355–384, 2005.

[3] A. Alfonsi. High order discretization schemes for the CIR process: application to Affine Term Structure and Heston models. *Mathematics of Computation*, 79(269):209–237, 2010.

[4] A. Alfonsi. Strong order one convergence of a drift implicit Euler scheme: Application to the CIR process. *Statistics and Probability Letters*, 83(2):602–607, 2013.

[5] V. Bally and C. Rey. Approximation of Markov semigroups in total variation distance. *Electronic Journal of Probability*, 21, 2016.

[6] C. Bayer. *The geometry of iterated Stratonovich integrals.* 2006.

[7] A. Berkaoui, M. Bossy, and A. Diop. Euler scheme for SDEs with non-Lipschitz diffusion coefficient: strong convergence. *ESAIM: Probability and Statistics*, 12:1–11, 2008.

[8] G. Beylkin, R. Coifman, and V. Rokhlin. *Wavelets in Numerical Analysis.* in Wavelets and Their Applications, ed. by Ruskai, Beylkin, Coifman, Daubechies, Mallat, Mayer, and Raphael (Jones and Bartlett, Boston), 1992.

[9] H. Boedihardjo, T. Lyons, and D. Yang. Uniform factorial decay estimates for controlled differential equations. *Electronic Communications in Probability*, vol. 20, paper no. 94, 2015.

[10] Y. Boutaib, L. G. Gyurkó, T. Lyons, and D. Yang. Dimension-free Euler estimates of rough differential equations. *Revue Roumaine de Mathématiques Pures et Appliquées*, 59(1):25–53, 2014.

[11] D. Brigo and F. Mercurio. *Interest Rate Models: Theory and Practice*. Springer, Berlin, 2001.

[12] E. Buckwar, M. Tamborrino, and I. Tubikanec. Spectral density-based and measure-preserving ABC for partially observed diffusion processes. An illustration on Hamiltonian SDEs. *Statistics and Computing*, 30(3):627–648, 2020.

[13] P. M. Burrage and K. Burrage. A Variable Stepsize Implementation for Stochastic Differential Equations. *SIAM Journal on Scientific Computing*, 24(3):848–864, 2002.

[14] G. Bussi and M. Parrinello. Accurate sampling using Langevin dynamics. *Physical Review E*, 75(5):056707, 2007.

[15] J. C. Butcher. *Numerical Methods for Ordinary Differential Equations*. John Wiley and Sons, 2016.

[16] Y. Cao, J. Lu, and L. Wang. *Complexity of randomized algorithms for underdamped Langevin dynamics*. https://arxiv.org/abs/2003.09906, 2020.

[17] L. Capriotti, Y. Jiang, and G. Shaimerdenova. Approximation Methods for Inhomogeneous Geometric Brownian Motion. *International Journal of Theoretical and Applied Finance*, 22(2), 2019.

[18] C. Carmeli, A. De Vito, and E. Toigo. Vector valued reproducing kernel Hilbert spaces of integrable functions and Mercer theorem. *Analysis and Applications*, 4(4):377–408, 2006.

[19] F. Castell and J. G. Gaines. The ordinary differential equation approach to asymptotically efficient schemes for solution of stochastic differential equations. *Annales de l'Institut Henri Poincaré Probabilités et Statistiques*, 32(2):231–250, 1996.

[20] X. Cheng, N. S. Chatterji, P. L. Bartlett, and M. I. Jordan. Underdamped Langevin MCMC: A non-asymptotic analysis. In *Proceedings of the 31st Conference On Learning Theory*, volume 75 of *Proceedings of Machine Learning Research*, pages 300–323, 2018.

[21] J. M. C. Clark and R. J. Cameron. *The maximum rate of convergence of discrete approximations for stochastic differential equations*. in Stochastic Differential Systems Filtering and Control, ed. by Grigelionis (Springer, Berlin), 1980.

[22] W. T. Coffey, Y. P. Kalmykov, and J. T. Waldron. *The Langevin Equation: With Applications to Stochastic Problems in Physics, Chemistry and Electrical Engineering*. World Scientifc, 2012.

[23] J. C. Cox, J. E. Ingersoll, and S. A. Ross. A theory of term structure of interest rates. *Econometrica*, 53(2):385–407, 1985.

[24] A. Cozma and C. Reisinger. Strong order 1/2 convergence of full truncation Euler approximations to the Cox–Ingersoll–Ross process. *IMA Journal of Numerical Analysis*, 40(1):358–376, 2020.

[25] A. S. Dalalyan. Theoretical guarantees for approximate sampling from smooth and log-concave densities. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 79(3):651–676, 2017.

[26] A. Davie. *Differential Equations Driven by Rough Paths: An Approach via Discrete Approximation*. Applied Mathematics Research eXpress, 2008.

[27] A. Davie. *KMT theory applied to approximations of SDE*. in Stochastic Analysis and Applications, ed. by Crisen, Hambly and Zariphopoulou (Springer Proceedings in Mathematics and Statistics), 2014.

[28] M. A. Demba, N. Senu, and F. Ismail. A symplectic explicit trigonometrically-fitted Runge-Kutta-Nyström method for the numerical solution of periodic problems. *International Journal of Applied Engineering Research*, 11(11):7495–7500, 2016.

[29] S. Dereich, A. Neuenkirch, and L. Szpruch. An Euler-type method for the strong approximation of the Cox-Ingersoll-Ross process. *Proceedings of the Royal Society A: Mathematical, Physical and Engineering Sciences*, 468(2140):1105–1115, 2012.

[30] A. S. Dickinson. Optimal Approximation of the Second Iterated Integral of Brownian Motion. *Stochastic Analysis and Applications*, 25(5):1109–1128, 2007.

[31] B. G. S. Doman. *The Classical Orthogonal Polynomials*. World Scientific, 2016.

[32] S. Duane, A. D. Kennedy, B. J. Pendleton, and D. Roweth. Hybrid Monte Carlo. *Physics Letters B*, 195(2), 1987.

[33] R. Flamary and N. Courty. *POT Python Optimal Transport library*. https://github.com/rflamary/POT, 2017.

[34] G. Flint and T. Lyons. *Pathwise approximation of SDEs by coupling piecewise abelian rough paths*. https://arxiv.org/abs/1505.01298, 2015.

[35] J. Foster, T. Lyons, and H. Oberhauser. An optimal polynomial approximation of Brownian motion. *SIAM Journal on Numerical Analysis*, 58(3):1393–1421, 2020.

[36] P. Friz and M. Hairer. *A Course on Rough Paths: With an Introduction to Regularity Structures*. Springer Universitext, Springer, 2014.

[37] P. Friz and N. Victoir. *Multidimensional Stochastic Processes as Rough Paths: Theory and Applications*. Cambridge University Press, 2010.

[38] J. G. Gaines and T. Lyons. Random Generation of Stochastic Area Integrals. *SIAM Journal on Applied Mathematics*, 54(4):1132–1146, 1994.

[39] J. G. Gaines and T. Lyons. Variable step size control for stochastic differential equations. *SIAM Journal on Applied Mathematics*, 57(5):1455–1484, 1997.

[40] P. Gassiat, H. Oberhauser, and C. Zou. *A free boundary characterisation of the Root barrier for Markov processes*. https://arxiv.org/abs/1905.13174, 2019.

[41] A. Gerbi, B. Jourdain, and E. Clément. *Ninomiya-Victoir scheme: strong convergence properties and discretization of the involved ordinary differential equations*. https://arxiv.org/abs/1410.5093, 2016.

[42] M. B. Giles. *Improved multilevel Monte Carlo convergence using the Milstein scheme*. in Monte Carlo and Quasi-Monte Carlo Methods, ed. by Keller, Heinrich and Niederreiter (Springer, Berlin), 2006.

[43] M. B. Giles. Multilevel Monte Carlo path simulation. *Operations Research*, 56(3):607–617, 2008.

[44] M. B. Giles. Multilevel Monte Carlo methods. *Acta Numerica*, 24:259–328, 2015.

[45] M. B. Giles and L. Szpruch. Antithetic multilevel Monte Carlo estimation for multi-dimensional SDEs without Lévy area simulation. *Annals of Applied Probability*, 24(4):1585–1620, 2014.

[46] M. Girolami and B. Calderhead. Riemann manifold Langevin and Hamiltonian Monte Carlo methods. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 73(2):123–214, 2011.

[47] A. Göing-Jaechke and M. Yor. A survey and some generalizations of Bessel processes. *Bernoulli*, 9(2):313–349, 2003.

[48] D. S. Grebenkov, D. Belyaev, and P. W. Jones. A multiscale guide to Brownian motion. *Journal of Physics A*, 49(4), 2015.

[49] L. G. Gyurkó. Differential Equations Driven by Π-Rough Paths. *Proceedings of the Edinburgh Mathematical Society*, 59(3):741–758, 2016.

[50] K. Habermann. *A semicircle law and decorrelation phenomena for iterated Kolmogorov loops.* Journal of the London Mathematical Society (in press), 2020.

[51] E. Hairer, S. P. Nørsett, and G. Wanner. *Solving Ordinary Differential Equations I. Nonstiff Problems.* Springer, Berlin, 1993.

[52] B. Hambly and T. Lyons. Differential equations driven by rough signals. *Annals of Mathematics*, 171:109–167, 2010.

[53] M. Hefter and A. Jentzen. On arbitrarily slow convergence rates for strong numerical approximations of Cox-Ingersoll-Ross processes and squared Bessel processes. *Finance and Stochastics*, 23(1):139–172, 2019.

[54] D. J. Higham. An algorithmic introduction to numerical simulation of stochastic differential equations. *SIAM Review*, 43(3):525–546, 2001.

[55] T. Huillet. On Wright-Fisher diffusion and its relatives. *Journal of Statistical Mechanics: Theory and Experiment*, 11, 2007.

[56] J. Hull and A. White. Pricing Interest-Rate-Derivative Securities. *Review of Financial Studies*, pages 573–592, 1990.

[57] A. Janssen. *Order book models, signatures and numerical approximations of rough differential equations.* PhD thesis, University of Oxford, 2011.

[58] I. Karatzas and S. E. Shreve. *Brownian Motion and Stochastic Calculus.* Second edition, Springer, Berlin, 1991.

[59] C. Kelly, G. Lord, and H. Maulana. *Hybrid, adaptive, and positivity preserving numerical methods for the Cox-Ingersoll-Ross model.* https://arxiv.org/abs/2002.10206, 2020.

[60] C. A. Kennedy and M. H. Carpenter. *Diagonally Implicit Runge-Kutta Methods For Ordinary Differential Equations: A Review.* NASA Scientific and Technical Information, Hampton, 2016.

[61] T. Klein, J. Fort, and P. Berthet. *Convergence of an estimator of the Wasserstein distance between two continuous probability distributions.* hal-01526879, 2017.

[62] P. E. Kloeden and E. Platen. *Numerical Solution of Stochastic Differential Equations.* Springer, Berlin, 1992.

[63] E. Kowalski. Bernstein polynomials and Brownian motion. *American Mathematical Monthly*, 113(10), 2006.

[64] H. Lamba, J. C. Mattingly, and A. M. Stuart. An adaptive Euler-Maruyama scheme for SDEs: convergence and stability. *IMA Journal of Numerical Analysis*, 27(3):479–506, 2007.

[65] K. Lê. A stochastic sewing lemma and applications. *Electronic Journal of Probability*, vol. 25, paper no. 38, 2020.

[66] J. Le Gall. *Brownian Motion, Martingales, and Stochastic Calculus.* Springer, 2016.

[67] K. Levasseur. A probabilistic proof of the Weierstrass approximation theorem. *American Mathematical Monthly*, 91(4):249–250, 1984.

[68] P. Lévy. Wiener's Random Function, and Other Laplacian Random Functions. In *Proceedings of the Berkeley Symposium on Mathematical Statistics and Probability*, pages 171–187. University of California Press, 1951.

[69] X. Li, D. Wu, L. Mackey, and M. A. Erdogdu. Stochastic Runge-Kutta Accelerates Langevin Monte Carlo and Beyond. *Advances in Neural Information Processing Systems*, 2019.

[70] T. Lyons. Differential equations driven by rough signals. *Revista Matemática Iberoamericana*, 14:215–310, 1998.

[71] T. Lyons. Rough paths, signatures and the modelling of functions on streams. In *Proceedings of the International Congress of Mathematicians*, 2014.

[72] T. Lyons, M. Caruana, and T. Lévy. *Differential equations driven by rough paths.* In École d'été de probabilités de Saint-Flour XXXIV-2004, edited by J. Picard in Volume 1908 of Lecture Notes in Mathematics, Springer, Berlin, 2007.

[73] T. Lyons and N. Victoir. Cubature on Wiener space. *Proceedings of the Royal Society A: Mathematical, Physical and Engineering Sciences*, 460(2041):169–198, 2004.

[74] S. J. A. Malham and A. Wiese. Stochastic Lie group integrators. *SIAM Journal on Scientific Computing*, 30(2):597–617, 2008.

[75] J. C. Mattingly, A. M. Stuart, and D. J. Higham. Ergodicity for SDEs and approximations: locally Lipschitz vector fields and degenerate noise. *Stochastic Processes and their Applications*, 101(2):185–232, 2002.

[76] G. Milstein and M. Tretyakov. Quasi-symplectic methods for Langevin-type equations. *IMA Journal of Numerical Analysis*, 23(4):593–626, 2003.

[77] G. Milstein and M. Tretyakov. *Stochastic Numerics for Mathematical Physics*. Springer Science & Business Media, 2013.

[78] P. Monmarché. *High-dimensional MCMC with a standard splitting scheme for the underdamped Langevin diffusion*. https://arxiv.org/pdf/2007.05455, 2020.

[79] P. Mörters and Y. Peres. *Brownian motion*. Cambridge University Press, 2010.

[80] R. M. Neal. *MCMC using Hamiltonian dynamics*. in Handbook of Markov Chain Monte Carlo, ed. by Brooks, Gelman, Jones and Meng (Chapman and Hall/CRC Handbooks of Modern Statistical Methods), 2011.

[81] S. Ninomiya and N. Victoir. Weak Approximation of Stochastic Differential Equations and Application to Derivative Pricing. *Applied Mathematical Finance*, 15(2):107–121, 2008.

[82] G. A. Pavliotis. *Stochastic Processes and Applications: Diffusion Processes, the Fokker-Planck and Langevin Equations*. Springer, New York, 2014.

[83] J. Reizenstein. *Calculation of Iterated-Integral Signatures and Log Signatures*. https://arxiv.org/abs/1712.02757, 2015.

[84] A. Rößler. Runge-Kutta Methods for the Strong Approximation of Solutions of Stochastic Differential Equations. *SIAM Journal on Numerical Analysis*, 8(3):922–952, 2010.

[85] R. Shen and Y. T. Lee. The Randomized Midpoint Method for Log-Concave Sampling. *Advances in Neural Information Processing Systems*, 2019.

[86] Z. Song and Z. Tan. *Hamiltonian Assisted Metropolis Sampling*. https://arxiv.org/pdf/2005.08159, 2020.

[87] E. Stein. *Singular integrals and differentiability properties of functions*. Princeton Mathematical Series, No. 30. Princeton University Press, Princeton, NJ, 1970.

[88] D. Talay. Efficient numerical schemes for the approximation of expectations of functionals of the solution of an SDE and applications. In *Filtering and Control of Random Processes*, pages 294–313. Springer, 1984.

[89] X. Tang and A. Xiao. Asymptotically optimal approximation of some stochastic integrals and its applications to the strong second-order methods. *Advances in Computational Mathematics*, 45:813–846, 2019.

[90] G. Teschl. *Ordinary Differential Equations and Dynamical Systems*. Graduate Studies in Mathematics, vol. 140, American Mathematical Society, Providence, 2012.

[91] C. Tsitouras and I. T. Famelis. Symbolic derivation of Runge-Kutta-Nyström order conditions. *Journal of Mathematical Chemistry*, 46(3):896–912, 2009.

[92] I. Tubikanec, M. Tamborrino, P. Lansky, and E. Buckwar. *Qualitative properties of numerical methods for the inhomogeneous geometric Brownian motion*. https://arxiv.org/abs/2003.10193, 2020.

[93] N. Vercauteren. *Numerical investigation of solutions of Langevin equations*. Master's thesis, Ecole Polytechnique Federale de Lausanne, Switzerland, 2005.

[94] C. Villani. *Optimal transport: Old and New*. Springer, Berlin, 2009.

[95] E. W. Weisstein. *Half-Normal Distribution*. From MathWorld – A Wolfram Web Resource, https://mathworld.wolfram.com/Half-NormalDistribution.html.

[96] E. W. Weisstein. *Jacobi Polynomial*. From MathWorld – A Wolfram Web Resource, https://mathworld.wolfram.com/JacobiPolynomial.html.

[97] E. Wong and M. Zakai. On the Convergence of Ordinary Integrals to Stochastic Integrals. *Annals of Mathematical Statistics*, 36(5):1560–1564, 1965.

[98] H. Yoshida. Construction of higher order symplectic integrators. *Physics Letters A*, 150(5-7):262–268, 1990.

[99] B. Zhao. Inhomogeneous geometric Brownian motion. *SSRN Electronic Journal*, 2009.

[100] B. Zhao, C. Yan, and S. Hodges. Three One-Factor Processes for Option Pricing with a Mean-Reverting Underlying: The Case of VIX. *Financial Review*, 54(1):165–199, 2019.