

Discrete approximations in stochastic rough path theory

Guy Flint

Green Templeton College

University of Oxford

A thesis submitted for the degree of
Doctor of Philosophy



Hilary Term 2016

Discrete approximations in stochastic rough path theory

A thesis submitted for the degree of Doctor of Philosophy

Hilary Term 2016

G. H. Flint

Green Templeton College

Abstract

We consider two discrete schemes for studying and approximating stochastic differential equations (SDEs) using the theory of rough paths.

The first part of the thesis introduces a canonical method for transforming a discrete sequential data set into an associated rough path made up of lead-lag increments. In particular, by sampling a d -dimensional continuous signal $X : [0, 1] \rightarrow \mathbb{R}^d$ at a set of times $D = \{t_i\}$, we construct a piecewise linear, axis-directed process $X^D : [0, 1] \rightarrow \mathbb{R}^{2d}$ comprised of a past and future component. We call such an object the Hoff process associated with the discrete data $\{X_t\}_{t_i \in D}$. The Hoff process can be lifted to its natural rough path enhancement and we consider the question of convergence as the sampling frequency increases. In the specific case where X is a semimartingale we prove that the Itô integral can be recovered from a sequence of random ODEs driven by the components of X^D . This is in contrast to the usual Stratonovich integral limit suggested by the classical Wong-Zakai theorem [119]. Such random ODEs have a natural interpretation in the context of mathematical finance.

The second part of the thesis introduces a new pathwise approximation scheme for SDEs driven by multidimensional Brownian motion. The work of Cameron, Clarke and Dickinson [17, 37] shows that N -step schemes cannot have a strong approximation error better than $O(N^{-1/2})$ unless higher iterated integrals of Brownian motion are used. On the other hand, our N -step scheme does not require Lévy area increments and if the vector fields of the SDE possess Stein-Lipschitz regularity of order $\gamma > 2$, we prove that the approximation error is $O(N^{-1+\varepsilon+2/\gamma})$ for all $\varepsilon > 0$ in the Wasserstein metric from optimal transport theory. By using techniques from rough path theory we avoid imposing any non-degenerate Hörmander or ellipticity conditions on the vector fields of the SDE, which is in contrast to previous pathwise schemes of Alfonsi, Davie, Malliavin et al in [1, 2, 32, 33, 34]. Our scheme is based on the log-ODE method with the Lévy area increments replaced by a quadratic polynomial of Gaussian random variables with the same covariance structure. The coupling construction is an extension of [33, 35] concerning a multidimensional variant of the Komlós-Major-Tusnády theorem and Wasserstein estimates for polynomial perturbations of Gaussian measures. These latter results require no rough path theory and may be of independent interest to probabilists.

Preface

This document constitutes my doctoral thesis. It is submitted in partial fulfilment for the requirements of the degree of Doctor of Philosophy in Mathematics at the Mathematical Institute of the University of Oxford.

Acknowledgements. I would like to thank my supervisor Terry Lyons. It has been a real privilege to spend many hours discussing mathematics with him. He has shown great levels of patience throughout all our arguments, both mathematical and otherwise. I would also like to thank my secondary supervisor Ben Hambly for all his help in providing a steadying counterbalance along with my college advisor Sarah Darby for offering well-timed dinner invitations to discuss anything but mathematics. I must thank Professor Sandy Davie at Edinburgh for answering many questions and patiently explaining his original proofs in [33, 35]. Thanks must also go to Peter Friz, Mike Giles, Zhongmin Qian and Danyu Yang for feedback on my transfer, confirmation and final theses. Thanks also to Greg Gyurkó and Weijun Xu for much help on rough path theory.

I gratefully acknowledge the financial support from the ERC¹, the Oxford-Man Institute in addition to funds from the G-Research DPhil prize and the IMU Itô travel award.

¹Under the European Union's 7th Framework Programme (FP7-IDEAS-ERC, ERC grant agreement No. 291244).

I am thankful to friends in Oxford, Cambridge and Sydney, especially Charlotte Koldewej, Katherine Fender, Nikola Novcic as well as Philippe Charmoy and Sean Ledger for sharing an office with me.

I am extremely grateful to my grandparents for sending a much-needed case of wine at the beginning of each Michaelmas term.

Finally, I would like to thank my parents along with James and George for their never-ending support all the way from Australia.

Oxford, England, May 30, 2016

Contents

Preface	5
Chapter 1. Introduction	11
1.1. Discrete approximations in stochastic rough path theory	11
1.2. Outline of the content and main results	14
1.3. Notation	16
Chapter 2. Elements of rough path theory	21
2.1. A primer on rough path theory	21
2.2. Algebraic preliminaries	23
2.3. Carnot-Carathéodory norm	26
2.4. Inhomogeneous metrics on rough path space	27
2.5. Geometric rough paths	28
2.6. Rough differential equations	29
2.7. Continuity of the Itô map	35
2.8. Stochastic differential equations as rough paths	36
2.9. Lipschitz estimates of rough differential equations	40
2.10. Log-ODE approximation scheme	42
Chapter 3. Discretely sampled signals and the rough Hoff path	47
3.1. Discrete streams of event data	47
3.2. Defining the Hoff process	49

3.3.	Rough Hoff process convergence	55
3.4.	Recovery of the Itô integral	59
3.5.	Relation to original work of Hoff	63
3.6.	Uniform convergence of the Hoff process	65
3.7.	Maximal p -variation bounds	70
3.8.	Possible extension to fractional Brownian motion	73
Chapter 4.	Pathwise approximation of stochastic differential equations	75
4.1.	Coupling and Wasserstein metrics	75
4.2.	Pathwise approximation of stochastic differential equations	79
4.3.	Previous pathwise research	83
4.4.	Komlós-Major-Tusnády theorem	89
4.5.	Pathwise approximation work of Davie	89
Chapter 5.	Pathwise approximation by coupling piecewise abelian rough paths	95
5.1.	New scheme and main theorem	96
5.2.	Kantorovich-Rubinstein duality formula	99
5.3.	Overview of the proof	101
5.4.	Piecewise abelian rough paths	104
5.5.	Coupling piecewise abelian rough paths	110
5.6.	Lifted rough path Hölder estimates	115
5.7.	Coupling solutions of stochastic differential equations	125
5.8.	Possible extensions	128
Chapter 6.	Polynomial perturbations of Gaussian distributions and the coupling construction	131

6.1.	Polynomial perturbations of Gaussian distributions	132
6.2.	Main coupling theorem	138
6.3.	A central limit theorem	142
6.4.	Proof of the coupling construction	147
Appendix A. Baker-Campbell-Hausdorff formula		155
A.1.	Baker-Campbell-Hausdorff formula	155
A.2.	Iterated Baker-Campbell-Hausdorff formula	156
Bibliography		159

CHAPTER 1

Introduction

1.1. Discrete approximations in stochastic rough path theory

Stochastic differential equations provide a mathematical framework for describing the evolution of multidimensional systems affected by some stochastic input signal. Formally, we study the response $y = (y_t)_{t \in [0,1]} \subset \mathbb{R}^q$ of a system $V : \mathbb{R}^q \rightarrow L(\mathbb{R}^d, \mathbb{R}^q)$ to a driving stochastic path $x = (x_t)_{t \in [0,1]} \subset \mathbb{R}^d$. This is written in the form of a controlled differential equation:

$$(1.1) \quad dy_t = V(y_t) dx_t, \quad y_0 = \xi \in \mathbb{R}^q.$$

As it is random, x is typically highly oscillatory and thus far from differentiable. This means classical calculus can be of no help in defining a solution to (1.1). If x has finite p -variation for some $p \in [1, 2)$ then Young's theory of integration [120] is applicable but in the canonical example of Brownian motion, traditionally one must use the stochastic calculus of Itô [67].

Developed during World War II in the context of control theory, Itô calculus has found applications in many areas of mathematics, most notably mathematical finance. However, it suffers from an important flaw: it is constructed via probabilistic techniques rather than a path-by-path method. The solution to (1.1) when x is Brownian motion W , denoted by $\pi(W) = (\pi_t(W))_{t \in [0,1]}$, is defined as the probabilistic limit of a sequence of forward Riemann sums. Fixing two (possibly correlated) sample paths of Brownian motion W_1, W_2 , this means

that one cannot make any quantitative statements about the difference in the output solutions:

$$(1.2) \quad \sup_{t \in [0,1]} |\pi_t(W_1) - \pi_t(W_2)|.$$

The problem is that the so-called Itô mapping $\pi : C([0, 1], \mathbb{R}^d) \rightarrow C([0, 1], \mathbb{R}^q)$ is not continuous in the uniform topology unless $d = 1$ or the vector fields making up V commute (in which case the classical Doss-Sussmann theorem [38, 110] can be employed). These latter conditions exclude interesting examples of SDEs with degenerate vector fields driven by correlated multidimensional stochastic signals.

Rough path theory solves this predicament. First introduced in [76], the key insight of Lyons was that the increments of the input path x are in general not enough to make quantitative estimates regarding the continuity of the output such as (1.2). Instead, the theory fixes a single realization of the driving signal x and lifts the path to an enhanced path $\mathbb{X}$ taking values in a high-dimensional Lie group. The object $\mathbb{X}$ encodes information about the iterated integrals of dx (up to some sufficient truncation) and is called the rough path lift of x . Given a rough path $\mathbb{X}$ above x , (1.1) can be rewritten as a rough differential equation (RDE) driven by $\mathbb{X}$:

$$(1.3) \quad dy_t = V(y_t) d\mathbb{X}_t, \quad y_0 = \xi \in \mathbb{R}^q,$$

and write $y = \pi(x) = \Xi(\mathbb{X})$. Let the stochastic processes x_1, x_2 have rough path lifts $\mathbb{X}_1, \mathbb{X}_2$ and denote the corresponding solutions of (1.3) by $\pi(x_i) = \Xi(\mathbb{X}_i)$. The power of the theory is that we can make precise statements of the form

$$(1.4) \quad \sup_{t \in [0,1]} |\pi_t(x_1) - \pi_t(x_2)| \leq C_V Z\rho(\mathbb{X}_1, \mathbb{X}_2).$$

Here, $Z > 0$ is a random variable independent of V and $\rho(\cdot)$ is some metric on rough path space. The vector field V only appears through the deterministic constant $C_V > 0$, while all the randomness enters through the terms $\rho(\mathbb{X}_1, \mathbb{X}_2)$ and Z (where the latter possesses exponential integrability in the case of semimartingales). Thus rough path theory has separated out the random and deterministic components of (1.1), (respectively x and V). In particular, the inequality (1.4) is a path-by-path statement, holding for all realizations of the stochastic rough path inputs.

Naturally such an estimate is extremely useful for numerical approximations of SDEs. For example, suppose we have an approximation $\mathbb{X}^N$ of the Brownian rough path $\mathbb{W}$ made up of N group-increments such that for some $\theta > 0$, we have

$$\mathbb{E}(\rho(\mathbb{X}^N, \mathbb{W})^p) = O(N^{-\theta p}) \text{ for all } p \in [1, \infty).$$

Then (1.4) implies that the scheme $\Xi(\mathbb{X}^N)$ approximates the true solution $\pi(W)$ with a strong error of $O(N^{-\theta})$ for every vector field V (assuming sufficient Stein-Lipschitz regularity). Importantly, V may be degenerate in possessing no Hörmander regularity. The point is that rough path theory provides a rigorous framework for developing new discrete approximation schemes with greater accuracy.

This thesis utilises rough path theory to introduce two new schemes for discretely approximating SDEs. The first one is concerned with sampling a semimartingale signal to recover the Itô integral from a sequence of random ODEs. Surprisingly, by producing an artificial lead-lag path and taking its rough path lift, we do not recover the Stratonovich stochastic integral, as suggested by classical theory.

The second scheme is concerned with the pathwise approximation of SDEs driven by multidimensional Brownian motion. This is an active area of research (see [1, 2, 32, 33, 34]) in which the error of the approximation is measured using the Wasserstein metric for probability measures from optimal transport theory. The novelty of this new algorithm is that it achieves a pathwise rate of convergence which is superior to the traditional Euler scheme but does not require higher iterated Brownian integrals (including Lévy area increments).

Rough path theory makes up the body but not the interface of the thesis. To present the schemes and the main results prior knowledge of rough paths is not required. Instead, the theory is used to establish the proofs.

1.2. Outline of the content and main results

Chapter 2 presents an overview of rough path theory. We place particular emphasis on the formulation of SDEs as rough differential equations as well as the log-ODE approximation scheme.

The main body of the thesis is split into two parts.

The first half consists of Chapter 3 in which we introduce the rough Hoff process as a canonical transformation of event stream data into a 2-geometric rough path [41]. Two main results (Theorems 3.3.1 and 3.4.1) are presented in the case where the event stream arises from a continuous semimartingale. The first theorem establishes the convergence of Hoff processes as the sampling frequency increases, showing that the quadratic variation of the semimartingale naturally appears in the rough path limit. As a result of this phenomenon, the second theorem proves the recovery of the Itô integral from a sequence of random ODEs driven by the Cauchy sequence of Hoff processes. The usual

Stratonovich correction term from the Wong-Zakai theorem [119] is cancelled out by the appearance of the quadratic variation term. The proof of the latter result is an almost immediate consequence of the universal limit theorem of rough path theory. We also discuss the connections between our contribution and the original work of Hoff in [66].

Chapter 4 marks the beginning of the second half of the thesis (the results of which can be found in the preprint [42]). We define the notion of pathwise approximation of SDEs and review previous research. Particular emphasis is placed on the work of Davie in [33] as we adapt his arguments and coupling constructions in Chapters 5 and 6. Moreover, by using the example of circular Brownian motion, we show that the N -step Euler scheme cannot perform better $O(N^{-1/2})$ in the Wasserstein metric $\mathcal{W}_p$.

In Chapter 5 we present a pathwise scheme based on the log-ODE method. The main result (Theorem 5.1.2) states that if the vector fields of the SDE enjoy Stein-Lipschitz regularity of order $\gamma > 2$, then our N -step scheme is of order $O(N^{-1+\varepsilon+2/\gamma})$ in every $\mathcal{W}_p$ -metric for arbitrary $\varepsilon > 0$. Importantly, our scheme does not require Lévy area yet performs better than the Euler scheme's optimal rate if $\gamma > 4$. The scheme replaces Lévy area increments in the log-ODE method with a quadratic polynomial of Gaussian random variables. We also show that the scheme has an optimal convergence rate of $O(N^{-1} \log N)$ via the SDE defining Lévy area. A discussion then follows on the connections with the previous work of Davie along with the Kantorovich-Rubinstein duality formula. The remainder of the chapter is dedicated to the proof of Theorem 5.1.2 using the coupling provided by Theorem 6.2.1. We use rough path theory to push this coupling of two area-valued random walks through the Itô map

in order to establish a Wasserstein bound between our scheme and the original SDE solution. The proof of Lipschitz estimates of the driving rough paths takes up most of the chapter.

In Chapter 6 we prove Theorem 6.2.1, which establishes a coupling of the two random walks respectively composed of Lévy area increments and a quadratic Gaussian approximation. The chapter contains no rough path theory and instead presents a series of probabilistic results in order to finally prove the theorem. An overview of polynomial perturbations of Gaussian measures is given along with an extension of [33, Corollary 3] from quartics to higher order polynomial perturbations by exploiting the recent preprint [35]. We also extend a central limit theorem in the Wasserstein metric in order to generalise the original coupling proof of Davie [33, Theorem 1] from a bound in $\mathcal{W}_2$ to $\mathcal{W}_p$ for general $p \in [1, \infty)$, thereby proving Theorem 6.2.1.

The thesis is concluded with a short appendix on the Baker-Campbell-Hausdorff formula and its iterated version.

1.3. Notation

Constants. Throughout the thesis, $C, c, \dots$ denote various deterministic constants (that may vary from line to line). Constants which are dependent upon a variable will have the dependency explicitly stated; for example, C_p denotes a constant dependent on p . If a constant C has many dependencies $\alpha, \beta, \gamma, \dots, p, q$, we will simply write $C = C(\alpha, \beta, \gamma, \dots, p, q)$. If we are working in the Euclidean space $\mathbb{R}^d$ we will always ignore dependencies on the dimension d (except for Section 6.1, where we repeat results in [35]).

Euclidean space. The standard basis of $\mathbb{R}^d$ is denoted by $\{e_k\}_{k=1}^d$ and the standard Euclidean norm is denoted by $|\cdot|$. The set of antisymmetric $d \times d$ matrices is denoted by $[\mathbb{R}^d, \mathbb{R}^d]$ and the usual bracket operation on $\mathbb{R}^d$ is given by

$$[x, y] = x \otimes y - y \otimes x = \sum_{1 \leq k < l \leq d} (x_k y_l - x_l y_k) [e_k, e_l],$$

where $[e_k, e_l] := e_k \otimes e_l - e_l \otimes e_k \in [\mathbb{R}^d, \mathbb{R}^d]$. Given a matrix $A \in \mathbb{R}^{d \times d} = (\mathbb{R}^d)^{\otimes 2}$, we write A^t and $\|A\|$ for its transpose and norm respectively.

General notation. The sets of strictly positive integers is denoted by $\mathbb{N} := \{1, 2, \dots\}$. For real numbers $x, y \in \mathbb{R}$, set $x \wedge y := \min\{x, y\}$ and $x \vee y := \max\{x, y\}$. Moreover, we set $\lfloor x \rfloor$ to be the largest integer less than or equal to x . Denote the set of all partitions of an interval $[s, t] \subseteq [0, 1]$ by $\mathcal{D}_{[s, t]}$ and define the mesh of $D = \{t_i\}_{i=1}^n \in \mathcal{D}_{[0, 1]}$ to be $|D| := \max_{t_i \in D} |t_{i+1} - t_i|$. We adopt the simplex notation $\Delta_{[s, t]} := \{(u, v) : s \leq u \leq v \leq t\}$. Given a complex number $z \in \mathbb{C}$, we write $\operatorname{Re}(z)$ and $\operatorname{Im}(z)$ for its real and imaginary parts respectively.

Big O notation. A function $f : \mathbb{N} \rightarrow \mathbb{R}$ is of order $O(g(N))$ if $|f(N)| / |g(N)|$ is bounded from above by a positive constant in the limit $N \rightarrow \infty$.

Differential calculus. The gradient of a continuously differentiable scalar function $\varphi \in C^1(\mathbb{R}^d, \mathbb{R})$ is given by

$$\nabla \varphi = \sum_{i=1}^d \frac{\partial \varphi}{\partial x_i} e_i.$$

We use $\log(\cdot)$ for the natural logarithm $\log_e(\cdot)$.

Flow notation. Given a (sufficiently regular) vector field $V : \mathbb{R}^q \rightarrow \mathbb{R}^q$, consider the classical ODE:

$$(1.5) \quad z_t = \xi_0 + \int_0^t V(z_s) ds, \quad t \in [0, 1].$$

Let us adopt classical flow notation by writing the time-1 value of the solution as $\exp(V)(\xi_0) := z_1$.

Stochastic processes. All stochastic processes are indexed over $[0, 1]$ unless explicitly stated otherwise. Moreover, all processes are assumed to live on some filtered probability space $(\Omega, \mathcal{F} = \{\mathcal{F}_t\}, \mathbb{P})$ satisfying the usual conditions. Given a random variable $X : \Omega \rightarrow \mathbb{R}$, we denote its L^p -norm for $p \geq 1$ by $\|X\|_{L^p} := \mathbb{E}(|X|^p)^{1/p}$. We will either use L^p or L^r (the latter when p is already being used for the p -variation of a path). The covariance matrix of a random vector $Z : \Omega \rightarrow \mathbb{R}^d$ is denoted by $\text{cov}(Z) \in \mathbb{R}^{d \times d}$, (in the case of $d = 1$, we simply use $\text{Var}(x)$). Given a σ -algebra $\mathcal{G} \subseteq \mathcal{F}$, we will write $\mathbb{E}_{\mathcal{G}}(\cdot)$ as shorthand for the conditional expectation operator with respect to $\mathcal{G}$. Brownian motion is always denoted by $W : [0, 1] \rightarrow \mathbb{R}^d$ and i.i.d. stands for independent and identically distributed.

RDE/SDEs. We use standard rough path notation which will be introduced in the next chapter. Importantly, we adopt the convention that all stochastic processes be denoted by capital letters while all deterministic paths and solutions to rough differential equations (RDEs) will be written in lower case. In the case of RDEs driven by semimartingales (that is, SDEs), we will continue to use lower case for the stochastic solutions in order to avoid confusion.

Semimartingales. We assume that the continuous semimartingale $X = (X^1, \dots, X^d)$ has the canonical decomposition $X = M + V$, where M is a local martingale and V is a process of bounded variation. We denote the quadratic variation of M by $\langle M \rangle := \{\langle M^i, M^j \rangle : i, j \in \{1, \dots, d\}\} : [0, 1] \rightarrow \mathbb{R}^{d \times d}$. Given $p \in [1, \infty)$, let $\mathcal{M}^p$ denote the class of L^p -bounded martingales $M : [0, 1] \mapsto$

$(M_t^1, \dots, M_t^d) \in \mathbb{R}^d$ which we equip with the seminorm

$$\|M\|_p := \left| \sup_{t \in [0,1]} |M_t - M_0| \right|_{L^p} = \mathbb{E} \left(\sup_{t \in [0,1]} |M_t - M_0|^p \right)^{1/p}.$$

Considering one particular dimension, say M^i , the Burkholder-Davis-Gundy inequality ([105, Theorem 42.1]) states that

$$\begin{aligned} c_p \|M^i\|_p^p &= c_p \mathbb{E} \left(\langle M^i \rangle_{0,1}^{p/2} \right) \leq \mathbb{E} \left(\sup_{t \in [0,1]} |M_t^i - M_0^i|^p \right) \leq C_p \mathbb{E} \left(\langle M^i \rangle_{0,1}^{p/2} \right) \\ &= C_p \|M^i\|_p^p \end{aligned}$$

for some universal constants $c_p < C_p$ dependent only on p .

Stratonovich integral. Let $M = (M^1, \dots, M^d)$ be a continuous semimartingale and suppose each vector field in the collection $\{V_k\}_{k=1}^d \subset C(\mathbb{R}^q, \mathbb{R}^q)$ is continuously differentiable. Adopting the notation $V_k = (V_k^1, \dots, V_k^q)$, it follows that the vector field

$$V_k V_l := \sum_{j=1}^q V_k^j \frac{\partial V_l}{\partial x_j}$$

is well-defined. We write the semimartingale solution $Y \in C([0, 1], \mathbb{R}^q)$ to the integral equation

$$Y_t = Y_0 + \sum_{k=1}^d \int_0^t V_k(Y_s) dM_s^k + \frac{1}{2} \sum_{1 \leq k, l \leq d} \int_0^t V_k V_l(Y_s) d\langle M^k, M^l \rangle_s$$

in the form of the Stratonovich stochastic integral

$$Y_t = Y_0 + \sum_{k=1}^d \int_0^t V_k(Y_s) \circ dM_s^k =: Y_0 + \int_0^t V(Y_s) \circ dM_s.$$

CHAPTER 2

Elements of rough path theory

This chapter provides a tailored overview of relevant elements of rough path theory needed for the thesis. Particular emphasis is placed on Lipschitz estimates for rough differential equations and the log-ODE method. We also take the opportunity to establish notation. For a detailed and precise overview of the theory we direct the reader to [45, 52, 74, 76, 77, 79, 81] among a multitude of others. For a brief sketch of the theory we recommend Lyons survey article [78] or the highly accessible lecture notes of Baudoin [8].

2.1. A primer on rough path theory

The theory of rough paths by Lyons burst into stochastic analysis in the mid-nineties, culminating in the landmark paper [77]. The theory is focused on capturing and making quantitatively precise the interactions between highly oscillatory and non-linear systems. The techniques draw particularly on the analysis of Young [120] and the geometric algebra of Chen [23]. With the theory built upon non-trivial aspects of Lie algebra theory and geometric machinery, the definition of a rough path was a curiously unfamiliar mix of abstract algebra and analysis to many probabilists at the time. Remarkably, the theory was able to show that Young's integration [120] and Itô's stochastic integration [67] were both special cases of the same calculus.

Since then rough path theory has evolved into an essential area of stochastic analysis, enabling one to perform differential calculus at a pathwise level on a wide variety of paths far rougher than the Brownian motion of Itô's original stochastic integration theory. Applications of the theory abound both inside and outside stochastic analysis. It has led to deep and powerful results in SDEs and SPDEs. In particular, rough paths provided the initial set of techniques that led Hairer to develop the more general theory of regularity structures [45, 63]. In turn, this allowed Hairer to give meaning to solutions of the famous Kardar-Parisi-Zhang equation [62], for which he was awarded the Fields medal in 2014. Most recently, the signature transform of rough path theory has provided new insights into efficient extraction of information from data and machine learning [41, 59, 75, 80, 93].

The focus of rough path theory is to give quantitative mathematical meaning to the concept of the evolving response $y = (y_t)_{t \in [0,1]} \subset \mathbb{R}^q$ of a system $f : \mathbb{R}^q \rightarrow L(\mathbb{R}^d, \mathbb{R}^q)$ driven by a signal $x = (x_t)_{t \in [0,1]} \subset \mathbb{R}^d$, which we usually write in the form of a controlled differential equation:

$$(2.1) \quad dy_t = f(y_t) dx_t, \quad y_0 = \xi \in \mathbb{R}^q.$$

Here, x can be highly oscillatory and thus non-differentiable.

One application of the theory is that we may now define solutions to the differential equation (2.1) in the case where the driving path $x = (x_t)$ is not of bounded variation but has a controlled roughness. A measure of this roughness is given by the p -variation of the path (see (2.2) below). For larger values of p we require more information about the structure of x ; specifically, we need to control the first $\lfloor p \rfloor$ levels of the signature of x .

2.2. Algebraic preliminaries

We introduce the necessary algebraic and geometric machinery in order to define rough path analysis. The foundation of the theory is given by the free tensor algebra and the free nilpotent Lie group embedded in it. Denote the space of continuous paths $x : [0, 1] \rightarrow \mathbb{R}^d$ by $C([0, 1], \mathbb{R}^d)$. We write $x_{s,t} := x_t - x_s$ as shorthand for the increment of x and let $\|x\|_\infty$ denote the uniform norm. For $p \geq 1$ define the p -variation norm of x by

$$(2.2) \quad \|x\|_{p\text{-var};[0,1]} := \sup_{D=(t_i) \subset [0,1]} \left(\sum_i |x_{t_i, t_{i+1}}|^p \right)^{1/p}.$$

Let us denote by $C^{p\text{-var}}([0, 1], \mathbb{R}^d)$ the linear subspace of $C([0, 1], \mathbb{R}^d)$ consisting of paths of finite p -variation. In the case of $x \in C^{p\text{-var}}([0, 1], \mathbb{R}^d)$ for $p \in [1, 2)$, the iterated integrals of x are canonically defined via Young integration [120]. The collection of all these integrals as an object in itself is called the signature of the path:

$$S(x)_{s,t} := 1 + \sum_{k=1}^{\infty} \int \dots \int_{s < t_1 < \dots < t_k < t} dx_{t_1} \otimes \dots \otimes dx_{t_k},$$

where $(s, t) \in \Delta_{[0,1]} := \{(s, t) : 0 \leq s \leq t \leq 1\}$. Adopting the convention that $(\mathbb{R}^d)^{\otimes 0} = \mathbb{R}$, we formally define the tensor algebras

$$T^{(\infty)}(\mathbb{R}^d) := \bigoplus_{k=0}^{\infty} (\mathbb{R}^d)^{\otimes k}, \quad T^{(n)}(\mathbb{R}^d) := \bigoplus_{k=0}^n (\mathbb{R}^d)^{\otimes k}.$$

We can see that the signature of x takes its values in $T^{(\infty)}(\mathbb{R}^d)$. Defining the canonical projections $\pi_n : T^{(\infty)}(\mathbb{R}^d) \rightarrow (\mathbb{R}^d)^{\otimes n}$ and $\pi_{0,n} : T^{(\infty)}(\mathbb{R}^d) \rightarrow T^{(n)}(\mathbb{R}^d)$, we can also consider the truncated signature:

$$\begin{aligned} S_n(x)_{s,t} &:= \pi_{0,n}(S(x)_{s,t}) = 1 + \sum_{k=1}^n \int \dots \int_{s < t_1 < \dots < t_k < t} dx_{t_1} \otimes \dots \otimes dx_{t_k} \\ &\in \bigoplus_{k=0}^n (\mathbb{R}^d)^{\otimes k} = T^{(n)}(\mathbb{R}^d). \end{aligned}$$

Hence we can view $S_n(\cdot)$ as a continuous mapping from $\Delta_{[0,1]}$ to $T^{(n)}(\mathbb{R}^d)$.

Given a coordinate $e_{i_1} \otimes \dots \otimes e_{i_k} \in (\mathbb{R}^d)^{\otimes k}$, we define the corresponding projection of the signature via the dual space. For example,

$$\langle e_{i_1}^* \otimes \dots \otimes e_{i_k}^*, S(x)_{s,t} \rangle = \int \dots \int_{s < t_1 < \dots < t_k < t} dx_{t_1}^{i_1} \dots dx_{t_k}^{i_k} \in \mathbb{R}.$$

We equip each $(\mathbb{R}^d)^{\otimes k}$ with the tensor algebra norm $\|\cdot\|_{(\mathbb{R}^d)^{\otimes k}}$ defined by

$$\|a\|_{(\mathbb{R}^d)^{\otimes k}} = \sqrt{\sum_{1 \leq i_1, \dots, i_k \leq d} |\langle e_{i_1}^* \otimes \dots \otimes e_{i_k}^*, a \rangle|^2},$$

and when no confusion arises we shall simply write $\|a\|$. This norm satisfies a compatibility relation between the tensor norms on the respective tensor levels: for all $0 \leq i \leq k \leq n$,

$$\|a \otimes b\|_{(\mathbb{R}^d)^{\otimes k}} = \|a\|_{(\mathbb{R}^d)^{\otimes i}} \|b\|_{(\mathbb{R}^d)^{\otimes (k-i)}} \quad \text{for all } (a, b) \in (\mathbb{R}^d)^{\otimes i} \times (\mathbb{R}^d)^{\otimes (k-i)}.$$

(For further details, see [52, §7.2]). We define a norm on $T^{(n)}(\mathbb{R}^d)$ by

$$\|g - h\|_{T^{(n)}(\mathbb{R}^d)} := \max_{k=0, \dots, n} \|g - h\|_{(\mathbb{R}^d)^{\otimes k}}, \quad g, h \in T^{(n)}(\mathbb{R}^d),$$

which turns $T^{(n)}$ into a Banach algebra.

It is a well-known fact that the level n signature $S_n(x)$ not only takes its values in $T^{(n)}(\mathbb{R}^d)$, but it lies in a special nilpotent Lie group embedded in the tensor algebra. To be precise, the level- n signature takes its values in the free step- n nilpotent Lie group with the generators $\{e_k\}_{k=1}^d$, which we denote by $G^{(n)}(\mathbb{R}^d)$. Indeed, defining the free step- n Lie algebra $\mathfrak{g}^{(n)}(\mathbb{R}^d)$ by

$$\mathfrak{g}^{(n)}(\mathbb{R}^d) := \mathbb{R}^d \oplus [\mathbb{R}^d, \mathbb{R}^d] \oplus \dots \oplus \underbrace{[\mathbb{R}^d, [\dots, [\mathbb{R}^d, \mathbb{R}^d]]]}_{(n-1) \text{ brackets}},$$

and the natural non-commutative exponential $\exp_n : T^{(n)}(\mathbb{R}^d) \rightarrow T^{(n)}(\mathbb{R}^d)$ by

$$\exp_n(a) := 1 + \sum_{k=1}^n \frac{a^{\otimes k}}{k!},$$

we have $G^{(n)}(\mathbb{R}^d) = \exp_n(\mathfrak{g}^{(n)}(\mathbb{R}^d))$. Again using a formal power series, we can also define the truncated logarithm on $T^{(n)}(\mathbb{R}^d)$:

$$\log_n(a) := \sum_{k=1}^n \frac{(-1)^k}{k} (1-a)^{\otimes k} \text{ for } a \in T^{(n)}(\mathbb{R}^d) \text{ such that } \pi_0(a) = 1.$$

Similarly, we define $\exp(a)$ and $\log(a) \in T^{(\infty)}(\mathbb{R}^d)$ using the infinite series expansions and we certainly have $\exp_n(a) = \pi_{0,n} \exp(a)$ and $\log_n(a) = \pi_{0,n}(\log(a))$. The following characterization summarises the situation (a modern proof can be found in [52, Theorem 7.30]).

Theorem 2.2.1 (Chow-Rashevskii, [24, 100]). *We have*

$$G^{(n)}(\mathbb{R}^d) := \{S_n(x)_{0,1} : x \in C^{1\text{-var}}([0, 1], \mathbb{R}^d)\}.$$

More abstractly, after fixing $p \geq 1$ and defining $\lfloor p \rfloor$ to be the smallest integer less than or equal to p , we can consider a continuous group-valued path $\mathbb{X} : [0, 1] \rightarrow G^{(\lfloor p \rfloor)}(\mathbb{R}^d)$:

$$\mathbb{X}_t = \left(1, \mathbb{X}_t^1, \dots, \mathbb{X}_t^{\lfloor p \rfloor}\right) \in G^{(\lfloor p \rfloor)}(\mathbb{R}^d) \text{ where } \pi_k(\mathbb{X}_t) = \mathbb{X}_t^k.$$

Importantly, the group structure provides a natural non-commutative notion of increment: $\mathbb{X}_{s,t} := \mathbb{X}_s^{-1} \otimes \mathbb{X}_t$. This multiplication operation is well-defined by Chen's theorem [79, Theorem 2.9].

Example 2.2.2. Take a path $x \in C^{1\text{-var}}([0, 1], \mathbb{R}^d)$ and set $\mathbb{X} := S_2(x) \in C([0, 1], G^{(2)}(\mathbb{R}^d))$ for its level-2 signature, which we call the (level-2) rough path lift of x . From the definition of the exponential map it is immediate that

$$\mathbb{X}_{s,t} = 1 + x_{s,t} + \frac{1}{2}x_{s,t} \otimes x_{s,t} + A_{s,t} = \exp_2(x_{s,t} + A_{s,t}),$$

where $A : \Delta_{[0,1]} \rightarrow [\mathbb{R}^d, \mathbb{R}^d]$ is the Lévy area process of x . Here we have used the truncated exponential notation defined above.

Remark 2.2.3. In fact, the signature $S(x)$ of a given path $x \in C^{1\text{-var}}([0, 1], \mathbb{R}^d)$ is uniquely determined by x up to the notion of so-called tree-like equivalence [64]. This theorem of Hambly and Lyons can be thought of as a non-commutative analogue of the well-known result in functional analysis that functions on the unit circle are determined up to Lebesgue null sets by their Fourier coefficients. We observe that the Fourier transform is a linear mapping, while the signature $S(\cdot)$ and its truncated version $S_n(\cdot)$ are not; $G^{(n)}(\mathbb{R}^d)$ possesses no linear structure.

2.3. Carnot-Carathéodory norm

We can now describe a symmetric and sub-additive norm on $G^{(\lfloor p \rfloor)}(\mathbb{R}^d)$ which is homogeneous with respect to the natural dilation operator on the tensor algebra, (see [52, Chapter VII] for details). This so-called Carnot-Carathéodory norm is given by

$$\|g\|_C := \inf \left\{ \int_0^1 |d\gamma| : \gamma \in C^{1\text{-var}}([0, 1], \mathbb{R}^d) \text{ such that } S_{\lfloor p \rfloor}(\gamma)_{0,1} = g \right\},$$

which is well-defined by the Chow-Rashevskii theorem. This in turn gives rise to the notion of a homogeneous metric on $G^{(n)}(\mathbb{R}^d)$:

$$d(g, h) = \|g^{-1} \otimes h\|_C.$$

Recalling the path x from Example 2.2.2, it can be shown that there exists a constant $C_d > 0$ such that

$$(2.3) \quad \frac{1}{C_d} \|S_2(x)_{s,t}\|_C \leq |x_{s,t}| \vee \sqrt{\|A_{s,t}\|_{(\mathbb{R}^d)^{\otimes 2}}} \leq C_d \|S_2(x)_{s,t}\|_C.$$

We will use this equivalence throughout the thesis. Moreover, this metric gives rise to the following p -variation metric between p -rough paths $\mathbb{X}, \mathbb{Y} \in$

$C([0, 1], G^{(\lfloor p \rfloor)}(\mathbb{R}^d))$:

$$d_{p\text{-var};[0,1]}(\mathbb{X}, \mathbb{Y}) := \sup_{D=(t_i) \subset [0,1]} \left(\sum_i \left\| \mathbb{X}_{t_i, t_{i+1}}^{-1} \otimes \mathbb{Y}_{t_i, t_{i+1}} \right\|_C^p \right)^{1/p}.$$

The p -variation norm is given by $\|\mathbb{X}\|_{p\text{-var};[0,1]} = d_{p\text{-var};[0,1]}(\mathbb{X}, 1)$. If this latter quantity is finite, then the function $\omega : (s, t) \in \Delta_{[0,1]} \mapsto \|\mathbb{X}\|_{p\text{-var};[s,t]}^p$ is a control; that is, a continuous bounded function, which vanishes on the diagonal $\{(t, t) : t \in [0, 1]\}$, and is super-additive in that for all $s \leq u \leq t$ in $[0, 1]$:

$$\omega(s, u) + \omega(u, t) \leq \omega(s, t).$$

Similarly we define the homogeneous $1/p$ -Hölder metric and norm by

$$d_{1/p\text{-Höl};[0,1]}(\mathbb{X}, \mathbb{Y}) := \sup_{0 \leq s < t \leq 1} \frac{\|\mathbb{X}_{s,t}^{-1} \otimes \mathbb{Y}_{s,t}\|_C}{|t - s|^{1/p}};$$

$$\|\mathbb{X}\|_{1/p\text{-Höl};[0,1]}(\mathbb{X}) = d_{1/p\text{-Höl};[0,1]}(\mathbb{X}, 1),$$

and define the rough path spaces:

$$C^{p\text{-var}}([0, 1], G^{(\lfloor p \rfloor)}(\mathbb{R}^d)) = \left\{ \mathbb{X} \in C([0, 1], G^{(\lfloor p \rfloor)}(\mathbb{R}^d)) : \|\mathbb{X}\|_{p\text{-var};[0,1]}(\mathbb{X}) < \infty \right\};$$

$$C^{1/p\text{-Höl}}([0, 1], G^{(\lfloor p \rfloor)}(\mathbb{R}^d)) = \left\{ \mathbb{X} \in C([0, 1], G^{(\lfloor p \rfloor)}(\mathbb{R}^d)) : \|\mathbb{X}\|_{1/p\text{-Höl};[0,1]}(\mathbb{X}) < \infty \right\}.$$

We stress that these spaces are not vector spaces; the addition of two rough paths, while being well-defined in the tensor algebra, may not sum up to a group element.

2.4. Inhomogeneous metrics on rough path space

The inhomogeneous p -variation and $1/p$ -Hölder metrics for p -rough paths are defined by ignoring the group structure of $G^{(\lfloor p \rfloor)}(\mathbb{R}^d)$, and instead using the

inherited norm from the tensor algebra:

$$\begin{aligned}\rho_{p\text{-var};[0,1]}(\mathbb{X}, \mathbb{Y}) &:= \max_{k=1,\dots,[p]} \rho_{p\text{-var};[0,1]}^{(k)}(\mathbb{X}, \mathbb{Y}); \\ \rho_{1/p\text{-H\"ol};[0,1]}(\mathbb{X}, \mathbb{Y}) &:= \max_{k=1,\dots,[p]} \rho_{1/p\text{-H\"ol};[0,1]}^{(k)}(\mathbb{X}, \mathbb{Y}),\end{aligned}$$

where

$$\begin{aligned}\rho_{p\text{-var};[0,1]}^{(k)}(\mathbb{X}, \mathbb{Y}) &:= \sup_{(t_i) \subset [0,1]} \left(\sum_i \left\| \pi_k(\mathbb{X}_{t_i, t_{i+1}} - \mathbb{Y}_{t_i, t_{i+1}}) \right\|_{(\mathbb{R}^d)^{\otimes k}}^{p/k} \right)^{k/p}; \\ \rho_{1/p\text{-H\"ol};[0,1]}^{(k)}(\mathbb{X}, \mathbb{Y}) &:= \sup_{0 \leq s < t \leq 1} \frac{\left\| \pi_k(\mathbb{X}_{s,t} - \mathbb{Y}_{s,t}) \right\|_{(\mathbb{R}^d)^{\otimes k}}}{|t - s|^{k/p}}.\end{aligned}$$

Given a control function $\omega : \Delta_{[0,1]} \rightarrow [0, \infty)$, we also define the metric

$$\begin{aligned}\rho_{p\text{-}\omega;[0,1]}(\mathbb{X}, \mathbb{Y}) &= \max_{k=1,\dots,[p]} \rho_{p\text{-}\omega;[0,1]}^{(k)}(\mathbb{X}, \mathbb{Y}) \\ \text{where } \rho_{p\text{-}\omega;[0,1]}^{(k)}(\mathbb{X}, \mathbb{Y}) &= \sup_{0 \leq s < t \leq 1} \frac{\left\| \pi_k(\mathbb{X}_{s,t} - \mathbb{Y}_{s,t}) \right\|_{(\mathbb{R}^d)^{\otimes k}}}{\omega(s, t)^{k/p}}.\end{aligned}$$

2.5. Geometric rough paths

The space of weakly geometric p -rough paths will be denoted by $WG\Omega_p(\mathbb{R}^d)$. This is the set of continuous paths $\mathbb{X}$ with values in $G^{(\lfloor p \rfloor)}(\mathbb{R}^d)$ such that $\|\mathbb{X}\|_{p\text{-var};[0,1]} < \infty$. A refinement of this notion is the space of geometric p -rough paths, denoted by $G\Omega_p(\mathbb{R}^d)$, which is the closure of

$$\{S_{[p]}(x) : x \in C^{1\text{-var}}([0, 1], \mathbb{R}^d)\}$$

with respect to the rough path metric $d_{p\text{-var}}(\cdot)$, (or equivalently $\rho_{p\text{-var}}(\cdot)$). Certainly we have the inclusion $G\Omega_p(\mathbb{R}^d) \subset WG\Omega_p(\mathbb{R}^d)$ and it turns out that this inclusion is strict. As described in [79, §3.2.2], this insignificant difference between geometric and weakly geometric rough paths can be compared to the difference between C^1 and Lipschitz functions.

We make a note that given $\mathbb{X} \in W\Omega_p(\mathbb{R}^d)$ and some $q \geq p$, there is a canonical extension of $\mathbb{X}$ to a q -rough path $S_{[q]}(\mathbb{X}) \in W\Omega_q(\mathbb{R}^d)$ such that $\pi_{0,[p]}(S_{[q]}(\mathbb{X})) = \mathbb{X}$. This so-called rough path lift operation is unique in the sense that the p -variation of the lift is Lipschitz continuous with respect to the original rough path (for a proof we refer to [52, Theorem 9.5]).

Theorem 2.5.1. *Let $q \geq p \geq 1$ and suppose $\mathbb{X} \in W\Omega_p(\mathbb{R}^d)$. There exists a constant $C_{p,q} > 0$ such that*

$$\|\mathbb{X}\|_{p\text{-var};[0,1]} \leq \|S_{[q]}(\mathbb{X})\|_{p\text{-var};[0,1]} \leq C_{p,q} \|\mathbb{X}\|_{p\text{-var};[0,1]}.$$

2.6. Rough differential equations

For now, let $x \in C^{1\text{-var}}([0, 1], \mathbb{R}^d)$, $h \in C^{1\text{-var}}([0, 1], \mathbb{R}^{d'})$. Define $y \in C([0, 1], \mathbb{R}^q)$ to be the solution of the (controlled) ordinary differential equation

$$\begin{aligned} dy_t &= V(y_t) dx_t + U(y_t) dh_t \\ &:= \sum_{k=1}^d V_k(y_t) dx_t^k + \sum_{k=1}^{d'} U_k(y_t) dh_t^k, \quad y_0 \in \mathbb{R}^q, \end{aligned}$$

which we summarise with the notation: $y = \pi_{(V,U)}(y_0, (x, h))$. Here, $\{V_k\}_{k=1}^d$ and $U = \{U_k\}_{k=1}^{d'}$ are collections of suitably regular vector fields $V_k, U_k : \mathbb{R}^q \rightarrow \mathbb{R}^q$.

Definition 2.6.1 (Rough differential equation with drift). *Fix $p \in [1, \infty)$ and $r \in [1, 2)$ such that $p^{-1} + r^{-1} > 1$. Let $\mathbb{X} \in W\Omega_p(\mathbb{R}^d)$ and $h \in C^{r\text{-var}}([0, 1], \mathbb{R}^{d'})$. We say that $y \in C([0, 1], \mathbb{R}^q)$ is a solution to the rough differential equation (RDE) driven by $\mathbb{X}$ with drift h along the collection of vector fields $\{V = \{V_k\}_{k=1}^d, U = \{U_k\}_{k=1}^{d'}\}$, and started from $y_0 \in \mathbb{R}^q$, if there exists a sequence*

$\{x_n, h_n\}_{n=1}^\infty \subset C^{1\text{-var}}([0, 1], \mathbb{R}^d) \times C^{1\text{-var}}([0, 1], \mathbb{R}^{d'})$ such that:

$$\sup_n \left\{ \|S_{[p]}(x_n)\|_{p\text{-var};[0,1]} + \|h_n\|_{q\text{-var};[0,1]} \right\} < \infty,$$

$$\lim_{n \rightarrow \infty} d_0(S_{[p]}(x_n), \mathbb{X}) = 0 \text{ and } \lim_{n \rightarrow \infty} \|h^n - h\|_\infty = 0,$$

and the sequence of ODE solutions $y_n := \pi_{(V,U)}(y_0, (x_n, h_n))$ satisfies:

$$\|y_n - y\|_\infty \rightarrow 0 \text{ as } n \rightarrow \infty.$$

We denote this situation with the (formal) equation:

$$dy_t = V(y_t)d\mathbb{X}_t + U(y_t)dh_t,$$

which we refer to as a rough differential equation (with drift), retaining the notation $y = \pi_{(V,U)}(y_0, (\mathbb{X}, h))$. In the special case of $h_t = t$, we write $y = \pi_{(V,U)}(y_0, \mathbb{X})$. Similarly, if $U = 0$ then we simply use $y = \pi_V(y_0, \mathbb{X})$.

Stein-Lipschitz vector fields. For unique solutions to classical ODEs we require the vector fields to be Lipschitz. In the rough differential case, from the point of view of existence and uniqueness results, the appropriate way to measure the regularity of the vector fields $V = \{V_k\}_{k=1}^d$ and $U = \{U_k\}_{k=1}^{d'}$ turns out to be the notion of γ -Lipschitz in the sense of Stein, (see [108, §VI.2]). To prepare for the precise definition, given a real non-integer $\gamma > 0$, we agree that $[\gamma]$ is the largest integer strictly smaller than γ so that $\gamma = [\gamma] + \{\gamma\}$ with $[\gamma] \in \mathbb{N}$ and $\{\gamma\} \in (0, 1)$.

Definition 2.6.2 (Stein-Lipschitz). *Let E, F be normed spaces. A map $V_k : E \rightarrow F$ is γ -Lipschitz (in the sense of Stein), in symbols*

$$V \in \text{Lip}^\gamma(E, F) \text{ or simply } V \in \text{Lip}^\gamma(E) \text{ if } E = F,$$

if V is $[\gamma]$ times continuous differentiable and there exists a constant $M \in [0, \infty)$ such that the supremum of its k th derivatives, $k = 0, \dots, [\gamma]$, and $\{\gamma\}$ -Hölder

norm of its $[\gamma]$ th derivative are bounded by M . The smallest M satisfying the above conditions is the γ -Lipschitz norm of V , denoted by $\|V\|_{\text{Lip}^\gamma}$.

Notation. Throughout the thesis we mostly consider $E = F = \mathbb{R}^q$, and write Lip^γ as shorthand for $\text{Lip}^\gamma(\mathbb{R}^q)$.

Remark 2.6.3. Informally, the definition states that the vector field can be approximated locally by a function taking values in polynomial functions [86, §2].

In [108, Chapter VI] Stein originally studied the general problem of extending smooth functions from subsets of Euclidean space to the whole space. To be precise, he considered the classical Whitney extension theorem [108, Theorem 6.2.4] which extends continuously the space $\text{Lip}^\gamma(F)$ of γ -Lipschitz functions on a compact set $F \subset \mathbb{R}^q$ to the whole Euclidean space.

Theorem 2.6.4 (Whitney extension theorem 1934). *Let $F \subset \mathbb{R}^q$ be a compact set. and suppose we have a function $V : F \rightarrow \mathbb{R}^q$. We have $V \in \text{Lip}^\gamma(F)$ if and only if V can be extended to an element $\bar{V} \in \text{Lip}^\gamma(\mathbb{R}^q)$. Moreover, in this case we have*

$$\|\bar{V}\|_{\text{Lip}^\gamma(\mathbb{R}^q)} \leq C \|V\|_{\text{Lip}^\gamma(F, \mathbb{R}^q)}$$

where $C > 0$ is some universal constant independent of F .

This result does not generalise to infinite dimensions.

As stated in [79], the definition of Stein-Lipschitz vector fields might leave the reader wondering what it means for a function to be Lip^γ for $\gamma > 1$, as the formal definition would lead one to the view that such functions are always constant. A useful fact to remember is that every bounded function with n

bounded derivatives is Lip^γ for all $\gamma \leq n$ [77, §1.2]. For more details on the notion of Stein-Lipschitz, we refer to [79, §1.4.2].

Existence and uniqueness of RDEs. Armed with the notion of Lipschitz vector fields, we state the following existence and uniqueness result for RDEs with drift taken from [52, Theorem 12.10].

Theorem 2.6.5. *Fix $\gamma > p \in [1, \infty)$ and $\beta > r \in [1, 2)$ such that $p^{-1} + r^{-1} > 1$. Let $\mathbb{X} \in WG\Omega_p(\mathbb{R}^d)$ with $h \in C^{r\text{-var}}([0, 1], \mathbb{R}^{d'})$. If $V = \{V_k\}_{k=1}^d \subset \text{Lip}^\gamma(\mathbb{R}^q)$ and $U = \{U_k\}_{k=1}^{d'} \subset \text{Lip}^\beta(\mathbb{R}^q)$, then the RDE with drift*

$$dy_t = V(y_t)d\mathbb{X}_t + U(y_t)dh_t,$$

has a unique solution for each initial point $y_0 \in \mathbb{R}^q$.

Intuitively, if the driving signal is rougher (that is, the greater the p -variation), then we need more structure ($[p]$ levels of its signature) in order to define the rough differential equation.

Remark 2.6.6. As will often be the case throughout the thesis, let us consider a polynomial vector field collection $V = \{V_k\}_{k=1}^d$. Certainly each $V_k \notin \text{Lip}^\gamma(\mathbb{R}^q)$ for any $\gamma > 1$ but for every compact set $F \subset \mathbb{R}^q$ its restriction to F is bounded and satisfies $V_k|_F \in \text{Lip}^\gamma(F, \mathbb{R}^q)$ for all $\gamma > 1$. We can use this observation together with Theorem 2.6.5 to prove the existence and uniqueness for RDEs driven by V .

Indeed, the extension is routine: we fix compact F big enough so that the vector fields of V are defined everywhere the solution can reach within 1 unit of time (certainly the range of the solution is compact if we impose non-explosion conditions on the RDE and apply a stopping time argument). Then using the Whitney extension theorem (Theorem 2.6.4) we can replace V globally by the

extension $\overline{V}$ of $V|_F$. The existence and uniqueness of the RDE driven by $\overline{V}$ are guaranteed by Theorem 2.6.5. Note that our quantitative estimates of the solution do not change because the extension satisfies

$$\|\overline{V}_k\|_{\text{Lip}^\gamma(\mathbb{R}^q)} \leq C \|V_k|_F\|_{\text{Lip}^\gamma(F, \mathbb{R}^q)} \text{ for all } k,$$

for some constant C independent of F .

To ensure the RDE solution does not explode in finite time, we use [58, Proposition 14] which requires the condition that the driving rough path is geometric and the original vector collection V satisfies the standard linear growth condition (see also [52, §10.7]). Interestingly, this regularity result cannot be generalised to non-geometric rough paths; a counterexample can be found in [58, Example 16]. Fortunately every RDE we encounter in this thesis is geometric and so we need not worry.

Definition 2.6.7 (Locally Stein-Lipschitz). *We say that $V : \mathbb{R}^q \rightarrow \mathbb{R}^q$ is locally γ -Stein-Lipschitz if for all compact $F \subset \mathbb{R}^q$, the restriction $V|_F$ is in $\text{Lip}^\gamma(F, \mathbb{R}^q)$. Moreover, we will also use the notation Lip^γ to denote the set of locally Stein-Lipschitz vector fields.*

Using the same notation for strict and locally Stein-Lipschitz vector fields will not cause confusion because we can define RDE solutions using the latter class of vector fields. Indeed, we repeat the argument given in Remark 2.6.6 by using the characterization in the definition to employ the Whitney extension theorem.

Central perturbations of RDEs. We conclude the section with an important RDE perturbation result. The theorem is stated in its complete form because it will be used throughout the thesis. Let us denote the centre of the

Lie algebra $\mathfrak{g}^{(n)}(\mathbb{R}^d)$ by $\mathcal{Z}^{(n)}(\mathbb{R}^d)$; in particular,

$$\mathcal{Z}^{(n)} := \mathfrak{g}^{(n)}(\mathbb{R}^d) \cap (\mathbb{R}^d)^{\otimes n} = \underbrace{[\mathbb{R}^d, [\dots, [\mathbb{R}^d, \mathbb{R}^d]]]}_{(n-1) \text{ brackets}}.$$

Theorem 2.6.8 (Central perturbations of RDEs). *Fix $p \in [1, \infty)$. Let $\mathbb{X} \in W\Omega G_p(\mathbb{R}^d)$ and suppose we have $\varphi \in C^{1\text{-var}}([0, 1], \mathcal{Z}^{(\lfloor p \rfloor)}(\mathbb{R}^d))$. We define the perturbation $\mathbb{X}^\varphi \in C([0, 1], G^{(\lfloor p \rfloor)}(\mathbb{R}^d))$ by*

$$\mathbb{X}_{s,t}^\varphi := \exp_{[p]}(\log(\mathbb{X}_{s,t}) + \varphi_{s,t}).$$

Then $\mathbb{X}^\varphi \in W\Omega G_p(\mathbb{R}^d)$. Assume $V = \{V_k\}_{k=1}^d \subset \text{Lip}^\gamma(\mathbb{R}^q)$ with $\gamma > p$ and $V_0 \in \text{Lip}^\beta(\mathbb{R}^q)$ with $\beta > 1$ so that the RDE with drift

$$dy_t = V(y_t) d\mathbb{X}^\varphi + V_0(y_t) dt, \quad y_0 = \xi_0 \in \mathbb{R}^q,$$

has a unique solution. Then there is a unique solution to the RDE with drift,

$$dz_t = V(z_t) d\mathbb{X}_t + V_0(z_t) dt + U(z_t) d\varphi_t, \quad z_0 = \xi_0,$$

where

$$U(z) d\varphi_t := \sum_{i_1, \dots, i_{\lfloor p \rfloor} \in \{1, \dots, d\}} [V_{i_1}, [\dots, [V_{i_{\lfloor p \rfloor - 1}}, V_{i_{\lfloor p \rfloor}}]]](z) \left\langle e_{i_1}^* \otimes \dots \otimes e_{i_{\lfloor p \rfloor}}^*, d\varphi_t \right\rangle.$$

Moreover, we have the equality:

$$y = \pi_{(V, V_0)}(\xi_0; (\mathbb{X}^\varphi, t)) = \pi_{(V, V_0 \oplus U)}(\xi_0; (\mathbb{X}, t \oplus \varphi_t)) = z.$$

The original proof of Theorem 2.6.8 was established in [46, Theorem 2] for the case of piecewise linear φ . This was then extended in [52, Theorem 12.14] to the case of bounded variation φ in the special case of $V_0 = 0$. The proof is generalised to the non-zero drift case (as stated here) with no extra effort so we omit the details.

2.7. Continuity of the Itô map

We introduce the fundamental notion of the Itô map in rough differential equations.

Definition 2.7.1 (Itô map). *Fix $p \in [1, \infty)$ and retain the notation of Theorem 2.6.5 with fixed initial point $y_0 \in \mathbb{R}^q$. The function*

$$\Xi : WG\Omega_p(\mathbb{R}^d) \rightarrow C([0, 1], \mathbb{R}^q) : \mathbb{X} \mapsto \pi_{(V, V_0)}(y_0, \mathbb{X}),$$

is known as the Itô map.

An important question is under what topology is Ξ continuous? Even when restricted to the case of bounded variation paths, the uniform topology is not strong enough. Unless the vector fields commute or $d = 1$, counterexamples are easy to construct. Moreover, it can be famously shown that the function mapping a bounded variation path in $\mathbb{R}^d = \mathbb{R}^2$ to its enclosed signed area is not a continuous functional with respect to the 2-variation metric (see [79, §1.5.2]). The initial raison d'être of rough path theory was that the information contained in the $\mathbb{R}^d$ -valued path $x := \pi_1(\mathbb{X})$ alone is not enough to ensure continuity of Ξ . We need the extra information contained in the rough path $\mathbb{X}$, along with the stronger topologies induced by the p -variation and $1/p$ -Hölder rough path metrics.

Theorem 2.7.2 (Universal limit theorem). [79, Theorem 5.3] *Under the conditions of Theorem 2.6.5, Ξ is continuous in the topologies induced by the metrics $d_{p\text{-var};[0,1]}(\cdot)$ or $d_{1/p\text{-Höl};[0,1]}(\cdot)$ on $WG\Omega_p(\mathbb{R}^d)$.*

Remark 2.7.3. We could have equivalently used the inhomogeneous rough path metrics $\rho_{p\text{-var};[0,1]}(\cdot)$, $\rho_{1/p\text{-Höl};[0,1]}(\cdot)$ in the previous theorem.

An additional power of the theory is that we can even establish quantitative Lipschitz estimates concerning this continuity, (which we will make precise in Section 2.9).

Remark 2.7.4. The universal limit theorem has enabled rough path theory to provide simple proofs of many classical results in stochastic analysis. Famous examples include the papers of Friz and Victoir [49] in addition to Ledoux, Qian and Zhang [73], which respectively prove the Stroock-Varadhan support theorem and the large deviation Freidlin-Wentzell estimates. These papers make elegant use of the Itô map to reduce these non-trivial results to simpler statements about Brownian motion and Lévy area in a rough path topology (see [52, Chapter IX]). In the next section (Theorem 2.8.6) we will also see how Theorem 2.7.2 can be used to establish the classical Wong-Zakai theorem in SDE theory.

2.8. Stochastic differential equations as rough paths

Now consider the case where the driving signal of a (suitably regular) RDE is a d -dimensional Brownian motion W . In the case of $d = 1$ or assuming the vector fields $V = \{V_k\}_{k=1}^d$ commute, the Doss-Sussmann theorem [38, 110] of classical stochastic analysis guarantees that the mapping

$$W = (W_t)_{t \in [0,1]} \in C([0,1], \mathbb{R}^d) \mapsto y \in C([0,1], \mathbb{R}^q)$$

is continuous in the uniform metric almost surely. If we now consider the case of $d > 1$ and remove the commutativity condition on the vector fields, this statement is no longer true. The canonical counterexample is the SDE defining the Lévy area of two-dimensional Brownian motion (see [116] for the original

proof by Watanabe). The problem is that Brownian motion has infinite p -variation for $p \in [1, 2]$ ([52, §13.9]). Instead, we must use the Itô map Ξ and consider the rough path lift (enhancement) of Brownian motion. As mentioned in [29, §1], in light of Theorem 2.7.2 rough path theory can be seen as a major extension of the Doss-Sussmann result.

Definition 2.8.1 (Enhanced Brownian motion). *Define enhanced Brownian motion as the rough path $\mathbb{W} \in C([0, 1], G^{(2)}(\mathbb{R}^d))$ given by*

$$\begin{aligned}
 \mathbb{W}_{s,t} &= 1 + W_{s,t} + \int_s^t W_{s,u} \otimes \circ dW_u \\
 &= \exp_2 \left(\sum_{k=1}^d W_k(s,t) e_k + \sum_{1 \leq k < l \leq d} A_{kl}(s,t) [e_k, e_l] \right) \\
 &= 1 + \sum_{k=1}^d W_k(s,t) e_k + \frac{1}{2} \sum_{1 \leq k, l \leq d} W_k(s,t) W_l(s,t) e_k \otimes e_l \\
 &\quad + \sum_{1 \leq k < l \leq d} A_{kl}(s,t) [e_k, e_l] \in G^{(2)}(\mathbb{R}^d),
 \end{aligned}
 \tag{2.4}$$

where $A_{kl} : \Delta_{[0,1]} \rightarrow [\mathbb{R}^d, \mathbb{R}^d]$ is the Lévy area of W :

$$A_{kl}(s,t) := \frac{1}{2} \int_s^t (W_k(s,u) dW_l(u) - W_l(s,u) dW_k(u)).$$

We know that $\mathbb{W} \in G\Omega_p(\mathbb{R}^d)$ almost surely by the following proposition, (taken from [52, Corollary 13.19]).

Proposition 2.8.2. *Let $\{D_n\}_{n=1}^\infty \subset \mathcal{D}_{[0,1]}$ such that $|D_n| \rightarrow 0$ as $n \rightarrow \infty$. For each $n \in \mathbb{N}$, let W^n denote the piecewise linear approximation of W with vertices $\{W_{t_i}\}_{t_i \in D_n}$. Then for all $p \in (2, \infty)$,*

$$d_{1/p\text{-Hö};[0,1]}(S_2(W^n), \mathbb{W}) \rightarrow 0 \text{ as } n \rightarrow \infty$$

almost surely and in L^r for all $r \in [1, \infty)$.

In fact, the p -variation (respectively $1/p$ -Hölder) norm of enhanced Brownian motion possesses exponential tails ([52, Corollary 13.15]).

Proposition 2.8.3. *Fix $p \in (2, \infty)$. There exists a constant $\eta > 0$ such that*

$$\mathbb{E} \left(\exp \left(\eta \|\mathbb{W}\|_{p\text{-var};[0,1]}^2 \right) \right) \leq \mathbb{E} \left(\exp \left(\eta \|\mathbb{W}\|_{1/p\text{-H\"ol};[0,1]}^2 \right) \right) < \infty.$$

The theorem linking SDEs with RDEs ([52, Theorem 17.3]) is stated below.

Theorem 2.8.4 (SDEs as RDEs). *Consider the following Stratonovich SDE driven by W :*

$$(2.5) \quad \begin{aligned} dx_t &= V(x_t) \circ dW_t + V_0(x_t) dt \\ &= \sum_{k=1}^d V_k(x_t) \circ dW_k(t) + V_0(x_t) dt, \quad x_0 = \xi_0 \in \mathbb{R}^q, \end{aligned}$$

where $V = \{V_k\}_{k=1}^d \subset \text{Lip}^\gamma(\mathbb{R}^q)$ and $V_0 \in \text{Lip}^\beta(\mathbb{R}^q)$. If $\gamma > 2$ and $\beta > 1$, then the solution x of (2.5) exists and is unique. Moreover, x coincides with the solution of the following RDE with drift:

$$dy_t = V(y_t) d\mathbb{W}_t + V_0(y_t) dt, \quad y_0 = \xi_0 \in \mathbb{R}^q.$$

Remark 2.8.5. As is standard practice in classical SDE theory, to prohibit the explosion of solutions we must make the standard assumption that V_0 and $\{V_k\}_{k=1}^d$ satisfy a global (classical) Lipschitz condition:

$$|V_k(y_1) - V_k(y_2)| \leq C_1 |y_1 - y_2| \quad \text{for } k = 0, 1, \dots, d,$$

for all $y_i \in \mathbb{R}^q$, where $C_1 > 0$ is a constant. In turn, this implies the linear growth condition: $|V_k(y)| \leq |V_k(0)| + |V_k(y) - V_k(0)| \leq C_2(1 + |y|)$ for some constant $C_2 > 0$. Using the standard existence and uniqueness SDE result of [70, Theorem 4.5.3] then guarantees that (2.5) has a unique solution $x = (x_t)_{t \in [0,1]}$ which is a process adapted to the filtration induced by the Brownian motion. Moreover, this solution satisfies $\mathbb{E}(|x_t|^q) < \infty$ for all $t \in [0, 1]$ and $q \in [1, \infty)$.

In terms of the Itô map of the previous subsection, Theorem 2.8.4 can be rewritten as

$$(2.6) \quad \Xi : \mathbb{W} \in G\Omega_p(\mathbb{R}^d) \mapsto \pi_{(V,V_0)}(\xi_0, \mathbb{W}) = y \in C([0, 1], \mathbb{R}^q).$$

Moreover, the universal limit theorem (Theorem 2.7.2) together with Proposition 2.8.2 immediately imply the classical Wong-Zakai theorem [119].

Theorem 2.8.6 (Wong-Zakai). *Let $p \in (2, \infty)$. Then under the same conditions as Proposition 2.8.2, we have*

$$d_{1/p\text{-var};[0,1]}(\Xi(S_2(W^n)), \Xi(\mathbb{W})) \rightarrow 0 \text{ as } n \rightarrow \infty$$

almost surely and in L^r for all $r \in [1, \infty)$.

Remark 2.8.7. Note that since each W^n is of bounded variation, the approximations $y^n := \Xi(S_2(W^n))$ can be written as the solutions of the random ODEs

$$dy_t^n = V(y_t^n) dW_t^n + V_0(y_t^n) dt, \quad y_0^n = \xi_0 \in \mathbb{R}^q.$$

Here, the solutions are well-defined by Young integration.

To summarise: because of the roughness of the Brownian sample paths, we must consider enhanced Brownian motion in order to make the SDE solution map continuous (the increment level W does not contain sufficient information on its own). This has been part of the success of rough path theory. We now have a powerful pathwise construction of stochastic calculus; we fix a sample path of Brownian motion and can define the solution of the SDE for this particular realization without using probability theory. This is in contrast to the construction of Itô's stochastic integral via the standard L^2 -isometry. The Itô map Ξ is purely deterministic, while the driving signal is stochastic.

2.9. Lipschitz estimates of rough differential equations

It is well-known that the Lipschitz constant of the Itô map Ξ for a RDE driven by a p -rough path $\mathbb{X}$ is of the order $O(\exp\{C \|\mathbb{X}\|_{p\text{-var};[0,1]}^p\})$. For a discussion on the optimality of the Lipschitz constant we refer to §4.2 of [46], (particularly Theorem 3).

Even for the well-studied case of Gaussian rough paths, the random variable $\exp\{C \|\mathbb{X}\|_{p\text{-var};[0,1]}^p\}$ fails to be finite in every L^r -norm, $r \geq 1$. Cass, Litterer and Lyons rectified this problem in [18] by refining the deterministic estimates for the Lipschitz constant. We start by recalling their definition of the so-called greedy p -variation partition.

Definition 2.9.1. *Let $\mathbb{X} \in WG\Omega_p(\mathbb{R}^d)$. For $\alpha > 0$ and $[s, t] \subseteq [0, 1]$, set $\tau_0(\alpha) = s$ then*

$$\tau_{n+1}(\alpha) = \inf \left\{ u : \|\mathbb{X}\|_{p\text{-var};[\tau_n, u]}^p \geq \alpha \text{ and } \tau_n(\alpha) < u \leq t \right\} \wedge t.$$

Define the integer

$$N_{\alpha,p}(\mathbb{X}, [s, t]) := \sup \{n \in \mathbb{N} \cup \{0\} : \tau_n(\alpha) < t\}.$$

Certainly $N_{\alpha,p}(\mathbb{X}, [s, t]) \leq C \|\mathbb{X}\|_{p\text{-var};[s,t]}^p$ for some constant $C > 0$, but more importantly the tail estimates for $N_{\alpha,p}(\mathbb{X}, [s, t])$ are significantly tighter than for $\|\mathbb{X}\|_{p\text{-var};[s,t]}^p$ when we consider Gaussian rough paths $\mathbb{X}$ (cf. [18, 47]). The following result is a slight variation of [19, Lemma 4.2].

Proposition 2.9.2. *Fix $\gamma > p \in [1, \infty)$ and $\beta > 1$. Suppose $\mathbb{X}^1, \mathbb{X}^2 \in WG\Omega_p(\mathbb{R}^d)$ with $\mathbb{X}^1 \neq \mathbb{X}^2$. Define the control $\omega : \Delta_{[0,1]} \rightarrow [0, \infty)$ by*

$$\omega(s, t) := |t - s| + \sum_{i=1}^2 \|\mathbb{X}^i\|_{p\text{-var};[s,t]}^p + \sum_{k=1}^{\lfloor p \rfloor} \left(\frac{\rho_{p\text{-var};[s,t]}(\mathbb{X}^1, \mathbb{X}^2)}{\rho_{p\text{-var};[0,1]}(\mathbb{X}^1, \mathbb{X}^2)} \right)^{p/k}.$$

Suppose $V_0 \in \text{Lip}^\beta$, $V = \{V_k\}_{k=1}^d \subset \text{Lip}^\gamma$ and fix $\lambda \geq \|V_0\|_{\text{Lip}^\beta} \vee \|V\|_{\text{Lip}^\gamma}$. Then the RDEs

$$dy_t^i = V(y_t^i) d\mathbb{X}_t^i + V_0(y_t^i) dt, \quad y_0^1 = y_0^2 \in \mathbb{R}^q,$$

have unique solutions. Moreover, for every $\alpha > 0$ there exists some constant $C = C(\alpha, \beta, \gamma, \lambda, p) > 0$ such that

$$\begin{aligned} \rho_{p\text{-var};[0,1]}(y^1, y^2) &\leq C\omega(0, 1) (1 \vee \omega(0, 1)^{\lfloor p \rfloor + 1}) \rho_{p\text{-var};[0,1]}(\mathbb{X}^1, \mathbb{X}^2) \\ &\quad \times \exp\left(C\left\{1 + N_{\alpha,p}(\mathbb{X}^1, [0, 1]) + N_{\alpha,p}(\mathbb{X}^2, [0, 1])\right\}\right). \end{aligned}$$

PROOF. The proof is obtained from following the argument of [19, Lemma 4.2] with some minor modifications. Indeed, the latter result guarantees that

$$\rho_{p\text{-}\omega;[0,1]}(y^1, y^2) \leq C\rho_{p\text{-}\omega;[0,1]}(\mathbb{X}^1, \mathbb{X}^2) \exp(CM_{\alpha,[0,1]}(\omega))$$

for some constant $C = C(\alpha, \beta, \gamma, \lambda, p) > 0$, where

$$M_{\alpha,[s,t]}(\omega) := \sup_{\substack{D=(t_i) \subset [s,t] \\ \omega(t_i, t_{i+1}) \leq \alpha}} \sum_i \omega(t_i, t_{i+1}).$$

It can be shown that $N_{\alpha,p}(\omega, [0, 1]) \leq M_{\alpha,[0,1]}(\omega) \leq 2N_{\alpha,p}(\omega, [0, 1]) + 1$. Moreover, by [11, Lemma 5] there exists a constant $C > 0$ such that

$$N_\alpha(\omega, [0, 1]) \leq C(1 + N_{\alpha,p}(\mathbb{X}^1, [0, 1]) + N_{\alpha,p}(\mathbb{X}^2, [0, 1])).$$

Therefore,

(2.7)

$$\rho_{p\text{-}\omega;[0,1]}(y^1, y^2) \leq C\rho_{p\text{-}\omega;[0,1]}(\mathbb{X}^1, \mathbb{X}^2) \exp\left(C\left\{1 + N_{\alpha,p}(\mathbb{X}^1, [0, 1]) + N_{\alpha,p}(\mathbb{X}^2, [0, 1])\right\}\right).$$

Under the assumption that $\omega(0, 1) \leq 1$, we can refer to the proof of [11, Theorem 4] to find that

$$\|\mathbb{X}^i\|_{p\text{-}\omega;[0,1]} \leq 1, \quad i = 1, 2 \quad \text{and} \quad \rho_{p\text{-}\omega;[0,1]}(\mathbb{X}^1, \mathbb{X}^2) \leq \rho_{p\text{-var};[0,1]}(\mathbb{X}^1, \mathbb{X}^2).$$

For the general case, a normalization argument gives

$$(2.8) \quad \rho_{p-\omega;[0,1]}(\mathbb{X}^1, \mathbb{X}^2) \leq (1 \vee \omega(0, 1)^{\lfloor p \rfloor}) \rho_{p-\text{var};[0,1]}(\mathbb{X}^1, \mathbb{X}^2).$$

As a consequence of the super-additivity of controls (see [52, §8.1]),

$$\rho_{p-\text{var};[0,1]}(y^1, y^2) \leq \omega(0, 1) \rho_{p-\omega;[0,1]}(y^1, y^2)$$

and combining this inequality with (2.7) and (2.8) concludes the proof. $\square$

2.10. Log-ODE approximation scheme

The log-ODE method is a powerful technique for numerically simulating the solutions of RDEs [51, 78]. Although not considered here, the method even holds for RDEs in infinite-dimensional Banach spaces [5, 13, 85]. For this thesis we are only concerned with the method in the context of approximating SDEs but before examining this special case, we first present the algorithm in full generality.

Log-ODE method. We give a tailored overview based on [61, 77, 84] and [107, §4]. The reader can also find a concise introduction in §7 of [78]. The method can also be found in the papers [4, 20, 21, 22, 40, 83], which were inspired by the pioneering work of Chen [23] and Magnus [87].

Fix $p \in [1, \infty)$ and let $\mathbb{X} \in WG\Omega_p(\mathbb{R}^d)$ be a p -rough path. Consider the RDE driven by $\mathbb{X}$:

$$(2.9) \quad dy_t = V(y_t) d\mathbb{X}_t, \quad y_0 \in \mathbb{R}^q.$$

As usual, we assume that $V = \{V_k\}_{k=1}^d \subset \text{Lip}^\gamma$, where $\gamma > p$, so that the RDE has a unique solution by Theorem 2.6.5. We consider the driftless case ($V_0 = 0$) because the inclusion of drift means we have to define the technical (but uninteresting) Lie algebra $\mathfrak{U}^{(n)}(\mathbb{R}, \mathbb{R}^d)$ given in [83, §4]. Essentially, the inclusion

of the drift complicates matters because scaling demands that the time coordinate $e_0 \in \mathbb{R}$ has multiplicity two in $\mathfrak{U}^{(n)}(\mathbb{R}, \mathbb{R}^d)$ whereas the spatial coordinates $\{e_k\}_{k=1}^d \subset \mathbb{R}^d$ each have multiplicity one. For our work with SDEs the inclusion of the drift component is an annoying albeit straightforward complication so we omit the details.

We begin by introducing a special Lie algebra homomorphism. A differentiable vector field $V : \mathbb{R}^q \rightarrow \mathbb{R}^q$ can be interpreted as a differential operator:

$$V(f) = \sum_{j=1}^q V^j \frac{\partial f}{\partial x_j}, \quad f \in C^1(\mathbb{R}^q, \mathbb{R}).$$

This allows us to define a new vector field

$$V_k V_l := \sum_{i,j=1}^q V_k^j \frac{\partial V_l^i}{\partial x_j} e_i$$

and in turn the Lie bracket operation on two continuously differentiable vector fields:

$$[V_k, V_l] := V_k V_l - V_l V_k = \sum_{i,j=1}^q \left(V_k^j \frac{\partial V_l^i}{\partial x_j} - V_l^j \frac{\partial V_k^i}{\partial x_j} \right) e_i.$$

Let $\text{Lip}^0(\mathbb{R}^q)$ denote the space of continuous vector fields on $\mathbb{R}^q$.

Definition 2.10.1. Fix $\gamma > \lfloor \gamma \rfloor \geq 1$. We define a Lie algebra homomorphism $\Phi : \mathfrak{g}^{(\lfloor \gamma \rfloor)}(\mathbb{R}^d) \rightarrow \text{Lip}^0(\mathbb{R}^q)$ by setting $\Phi(e_k) = V_k$ for $k \in \{1, \dots, d\}$.

Thus Φ maps Lie algebra elements in $\mathfrak{g}^{(\lfloor \gamma \rfloor)}(\mathbb{R}^d)$ to continuous vector fields on $\mathbb{R}^q$, while preserving Lie bracket operations. For example, if $\lfloor \gamma \rfloor \geq 2$,

$$\Phi(e_i + [e_j, e_k]) = V_i + [V_j, V_k].$$

We introduce the log-ODE method as presented in [85, Theorem 19]. Recall the classical flow notation of (1.5).

Theorem 2.10.2 (Log-ODE method). Fix $\gamma > p \geq 1$ and let $\mathbb{X} \in W\mathcal{G}\Omega_p(\mathbb{R}^d)$ and consider the RDE (2.9). We assume that $V = \{V_k\}_{k=1}^d \subset \text{Lip}^\gamma$ so that a

unique solution exists. Define the control $\omega : \Delta_{[0,1]} \rightarrow [0, \infty)$ by

$$(2.10) \quad \omega(s, t) := \|V\|_{\text{Lip}^\gamma}^p \|\mathbb{X}\|_{p\text{-var};[s,t]}^p.$$

Then for all integers $m \geq \lfloor p \rfloor$ there exists a constant $C_{\gamma,m,p} > 0$ such that, for every $(s, t) \in \Delta_{[0,1]}$, we have

$$|y_t - \exp \{ \Phi(\log_m(\mathbb{X}_{s,t})) \} (y_s)| \leq C_{\gamma,m,p} \left(\omega(s, t)^{\frac{m+1}{p}} \vee \omega(s, t)^m \right).$$

This is known as the log-ODE method at level m .

SDE approximation via the log-ODE method. We now specialise to the case of SDEs (considered as RDEs), and apply the log-ODE method to approximate the solution of the Stratonovich SDE

$$(2.11) \quad \begin{aligned} dx_t &= V(x_t) \circ dW_t + V_0(x_t) dt \\ &= \sum_{k=1}^d V_k(x_t) \circ dW_k(t) + V_0(x_t) dt, \quad t \in [0, 1]. \end{aligned}$$

Although Theorem 2.10.4 was presented without a drift term, we can bring the drift term V_0 back into the SDE/RDE with an additional drift vector field: $\frac{1}{N}V_0$. The details are omitted and can be found in [60, Chapter IX].

For this thesis we only apply the method at level $m = 2$, although the scheme can be generalised to arbitrary levels. Recalling (2.4), the level-2 log-signature of enhanced Brownian motion is given by

$$\log_2(\mathbb{W}_{s,t}) = \sum_{k=1}^d W_k(s, t) e_k + \sum_{1 \leq k < l \leq d} A_{kl}[e_k, e_l] \in \mathfrak{g}^{(2)}(\mathbb{R}^d).$$

The scheme is composed of iterated ODE approximations over successive intervals $[jh, (j+1)h]$ where $h = N^{-1}$.

Definition 2.10.3 (Log-ODE method for SDEs). *Set $x^{(0)} := x_0$, then*

$$x^{(j+1)} := \exp(F_j)(x^{(j)}), \quad j \in \{0, 1, \dots, N-1\},$$

where $F_j := \frac{1}{N}V_0 + \Phi(\log_2(\mathbb{W}_{jh, (j+1)h}))$

$$(2.12) \quad = \frac{1}{N}V_0 + \sum_{k=1}^d W_k(jh, (j+1)h) V_k + \sum_{1 \leq k < l \leq d} A_{kl}(jh, (j+1)h) [V_k, V_l].$$

This algorithm is the so-called (level 2) log-ODE method for the case of SDEs.

The N -step scheme has a strong approximation rate of order $O(N^{-1})$ [61, Theorem 4.1].

Theorem 2.10.4 (Log-ODE method for SDEs). *Fix $\gamma > 2$, $\beta > 1$, $r \in [1, \infty)$ and fix $\lambda \geq \|V_0\|_{\text{Lip}^\beta} \vee \|V\|_{\text{Lip}^\gamma}$. Then there exists a constant $C_{\beta, \gamma, \lambda, r} > 0$ such that*

$$\left\| \max_{j=1, \dots, N} |x^{(j)} - x_{jh}| \right\|_{L^r} \leq \frac{C_{\beta, \gamma, \lambda, r}}{N}.$$

Remark 2.10.5. The level 1 log-ODE scheme replaces each F_j of (2.12) with the simpler vector field

$$G_j := \frac{1}{N}V_0 + \sum_{k=1}^d W_k(jh, (j+1)h) V_k.$$

In common with the standard Euler scheme, Lévy area is absent and it can be shown that this scheme has a strong approximation error of $O(N^{-1/2})$.

Advantages of the log-ODE method. As noted in Remark 1.1 of [94], if one approximates each ODE (2.12) by its order-1 Taylor expansion, then we fall back on the traditional Euler scheme. Similarly, taking the better approximation offered by the order-2 Taylor expansion gives the Milstein scheme.

However, the log-ODE method has one important advantage over these more popular schemes. As outlined in [78, §7], the previous methods are based on Taylor expansions and thus can produce approximations whose law is not

absolutely continuous with respect to the measure of the solution on Wiener space. For example, we could consider the following SDE whose solution is constrained by its Stratonovich formulation to lie on the unit sphere at all times:

$$dx(t) = V(x(t)) \circ dW(t), \quad x_0 = (1, 0)^t, \quad \text{where } V(x) = (-x_2, x_1)^t.$$

Both the Euler and Milstein schemes are numerically unstable in that they will produce approximations which do not live on the sphere (see [98, §17.5]). By contrast, the log-ODE method only returns solutions which could have originated from an actual realization of the SDE, (assuming that the ODE solver used is sufficiently accurate over small time steps; for example, the Runge-Kutta method or an adaptive step-size method). This property is known as geometric stability in [60] and arises in the log-ODE method because the technique restricts the approximations to only flow along the vector fields (and their nested Lie brackets) of the original SDE. In turn, this ensures that feasibility constraints imposed on the law of the SDE solution will also be satisfied by the log-ODE output. A more complicated example could be a system with Hamiltonian vector fields [111] or the stochastic volatility example of [94, §3].

CHAPTER 3

Discretely sampled signals and the rough Hoff path

This chapter introduces a canonical method for transforming a discrete sequential data set into an associated rough path made up of lead-lag increments. In particular, by sampling a d -dimensional continuous signal $X : [0, 1] \rightarrow \mathbb{R}^d$ at a set of times $D = \{t_i\}$, we construct a piecewise linear, axis-directed process $X^D : [0, 1] \rightarrow \mathbb{R}^{2d}$ comprised of a past and future component. We call such an object the Hoff process associated with the discrete data $\{X_t\}_{t_i \in D}$. The Hoff process can be lifted to its natural rough path enhancement and we consider the question of convergence as the sampling frequency increases. In the specific case where X is a semimartingale we prove that the Itô integral can be recovered from a sequence of random ODEs driven by the components of X^D . This is in contrast to the usual Stratonovich integral limit suggested by the classical Wong-Zakai theorem [119]. Such random ODEs have a natural interpretation in the context of mathematical finance. The results presented here are taken from the paper [41].

3.1. Discrete streams of event data

A stream of event data is a time-ordered sequence of events where we have the time of the event as a real value and the data associated with the event typically taking some numerical value. In many applications, for example order books in finance or internet traffic, it is typical that these events come in a

variety of quite different forms at different frequencies, and that the realization of such streams can have a complex dependency structure between their coordinates. However, if we focus our interest in such event streams around their effects and the actions they trigger, then rough path theory provides an established and effective method (using the signature of the stream and the Itô map) to transform these event streams into much smaller data sets, while retaining adequate information to accurately approximate the primary effects of the data (see [78, §1-5]). As mentioned in Section 2.1, the focus of rough path theory is to give quantitative mathematical meaning to the concept of the evolving response $y = (y_t)_{t \in [0,1]} \subset \mathbb{R}^q$ of a system $f : \mathbb{R}^q \rightarrow L(\mathbb{R}^d, \mathbb{R}^q)$ driven by a signal $x = (x_t)_{t \in [0,1]} \subset \mathbb{R}^d$. We write this as the RDE:

$$(3.1) \quad dy_t = f(y_t) dx_t, \quad y_0 = \xi \in \mathbb{R}^q.$$

Classical calculus will fail to define a solution because x will be far from differentiable nor have p -variation for $p \in [1, 2]$, (for example, Brownian motion). On the other hand, rough path theory provides a systematic method of summarising the data x while retaining a good approximation of y . A reader familiar with differential equations might want to think about the case where x is smooth but highly oscillatory on normal scales. The meaning of the differential equation (3.1) is not in doubt. Indeed, rough path theory tackles the question of how to define solutions by exploiting the smoothness of the system f . The theory summarises x on normal scales so as to effectively capture its effects and predict y without knowing or caring further about x on microscopic scales.

This chapter introduces a canonical method of transforming discrete numerical event data $\{X_{t_i}\}_{t_i \in D} \subset \mathbb{R}^d$ into a geometric 2-rough path $\mathbb{X}^D$ which we call a Hoff process. Importantly, our transformation takes into account the order

of events and the effect of latency between different coordinates. Moreover, differential and integral equations driven by $\mathbb{X}^D$ have a classical as well as a rough path meaning, and we can consider the question of convergence as our sampling rate becomes finer by exploiting the Itô map from rough path theory.

With this in mind, this chapter examines the interplay between this way of looking at discrete data through the Hoff process specifically when the discrete event data comes from a multidimensional semimartingale signal. In effect, we prove the equivalent of the Wong-Zakai Theorem by demonstrating that, under certain regularity conditions, a sequence of Hoff processes is Cauchy in the standard p -variation rough path metric as the sampling frequency becomes finer. This is the content of the first main result of the chapter (Theorem 3.3.1). Before introducing the second result, let us in the next section formally define a Hoff process.

3.2. Defining the Hoff process

This section defines the Hoff process as a method of transforming discrete numerical event data in $\mathbb{R}^d$ into a geometric 2-rough path. Given a sample $\{X_{t_i}\}_{t_i \in D} \subset \mathbb{R}^d$ at times $D = \{t_i\}_{i=0}^n \subset [0, 1]$, the standard approach to transforming this into a continuous time trajectory is to define the càdlàg, piecewise constant path $Z : [0, 1] \rightarrow \mathbb{R}^d$ by $Z_t = X_{t_i}$ for $t \in [t_i, t_{i+1})$. Although this is a natural construction, the order of events across different coordinates is in some sense lost. Following the PhD thesis [66] of B. Hoff, we go one step further and consider the continuous process made up of lead and lag versions of this standard interpolation. We call the resultant path $X^D : [0, 1] \rightarrow \mathbb{R}^{2d}$ the Hoff process:

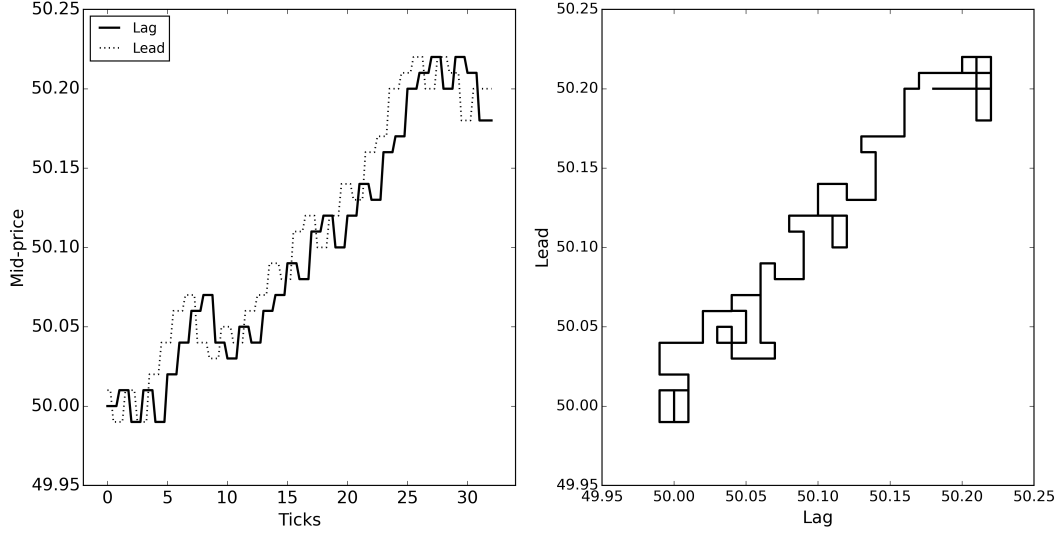


FIGURE 3.1. An example of a Hoff process trajectory

Definition 3.2.1. Let $D = \{t_i\}_{i=0}^n \subset [0, 1]$ be a sequence of times with $(t_0, t_n) = (0, 1)$. Define the path $\hat{X}_t : [0, 2n] \rightarrow \mathbb{R}^{2d}$ given by

$$\begin{aligned} \hat{X}_t^D &= \left(\hat{X}_t^{D,b}, \hat{X}_t^{D,f} \right) \\ &:= \begin{cases} (X_{t_k}, X_{t_{k+1}}) & t \in [2k, 2k+1); \\ (X_{t_k}, X_{t_{k+1}} + 2(t - (2k+1))(X_{t_{k+2}} - X_{t_{k+1}})) & t \in [2k+1, 2k+\frac{3}{2}); \\ (X_{t_k} + 2(t - (2k+\frac{3}{2}))(X_{t_{k+1}} - X_{t_k}), X_{t_{k+2}}) & t \in [2k+\frac{3}{2}, 2k+2); \end{cases} \end{aligned}$$

with $\hat{X}_n^D = (X_{t_{n-1}}, X_{t_n})$. We define the Hoff process associated with the sequential data $\{X_{t_i}\}_{t_i \in D}$ to be the path $X^D : [0, 1] \rightarrow \mathbb{R}^{2d}$ given by

$$X_t^D = (X_t^{D,b}, X_t^{D,f}) := \hat{X}_{2nt}^D.$$

The rough Hoff process $\mathbb{X}^D$ is the 2-rough path lift of X^D .

We call $X^{D,b}$ and $X^{D,f}$ the *lag* and *lead* components of the Hoff process respectively. Since lead and lag share the same first letter, we use *b* for *backward* and *f* for *forward*. At this point we find it helpful in building intuition about X^D to include a diagram of a possible trajectory (see Figure 3.1).

The definition states that $X_t = (X_{t_k}, X_{t_{k+1}})$ for $t \in [\frac{k}{n}, \frac{2k+1}{2n})$, and on the intervals $[\frac{2k+1}{2n}, \frac{k+1}{n})$, X is a special interpolation between $(X_{t_k}, X_{t_{k+1}})$ and $(X_{t_{k+1}}, X_{t_{k+2}})$. This interpolation is done by first updating the lead component, followed by the lag component. We note that other lead-lag interpolations are possible ([59, 75, 93]) but we use Definition 3.2.1 because of the following interpretation in the context of mathematical finance.

Suppose one is given the tick data of a collection of stocks $\{S_{t_i}\}_{i=0}^n \subset \mathbb{R}^d$ at certain times $D = \{t_i\}_{i=0}^n$. For simplicity we assume that $(t_0, t_n) = (0, 1)$. Using the sequential data $\{X_{t_i}\}_{i=0}^n := \{(t_i, S_{t_i})\}_{i=0}^n \subset [0, \infty) \times \mathbb{R}^d \simeq \mathbb{R}^{d+1}$, we could define the corresponding càdlàg function $Z : t \in [0, 1] \mapsto (Z_t^1, \dots, Z_t^d) \in \mathbb{R}^d$ as above. One could think of Z as representing the current known price for the given stock collection. Similarly, we can think of the previsible process

$$(3.2) \quad \tilde{Z}_t := \sum_{t_i \in D} Z_{t_i} \mathbf{1}\{t \in (t_{i+1}, t_{i+2}]\} \text{ where } Z_{t_i} \in \mathcal{F}_{t_i},$$

as recording the last known traded price for the collection. The performance of an agent's trading strategy could then be represented as the Itô integral:

$$(3.3) \quad \begin{aligned} Y^D &:= \int_0^1 f(\tilde{Z}_t) dZ_t = \sum_{k=1}^d \int_0^1 f_k(\tilde{Z}_t) dZ_t^k \\ &= \sum_{k=1}^d \sum_{t_i \in D} f_k(t_i, S_{t_i}) (S_{t_{i+2}}^k - S_{t_{i+1}}^k), \end{aligned}$$

where for $f = (f_1, \dots, f_d) : \mathbb{R} \times \mathbb{R}^d \rightarrow \mathbb{R}^d$. Here, $f_k(t_i, S_{t_i})$ represents the amount stock k held by the agent at time t_{i+1} , given that they hold the portfolio S_{t_i} at time t_i . Following the work of Friz in [43], we use (3.2) instead of the well-known process $Z_{t-} = \sum_{t_i \in D} Z_{t_i} \mathbf{1}\{t \in (t_i, t_{i+1}]\}$ because we want the agent to act on the market at time t_{i+1} based on their information at time t_i ; there is a lag between receiving market information and acting upon it. Alternatively,

one could interpret the model as representing a delay between placing an order on an exchange and the consequent change in the portfolio's value. Both events happen over disjoint time intervals.

On the other hand, let $X^D \rightarrow \mathbb{R}^{2(d+1)}$ denote the Hoff process associated with the same discrete data $\{X_{t_i}\}_{i=0}^n = \{(t_i, S_{t_i})\}_{i=0}^n \subset \mathbb{R}^{d+1}$, with $X^{D,0} = (X^{D,b,0}, X^{D,f,0})$ representing the time component. A simple computation gives

$$\begin{aligned} \sum_{t_i \in D} f_k(t_i, S_{t_i}) (S_{t_{i+2}}^k - S_{t_{i+1}}^k) &= \sum_{i=0}^{n-2} f_k(t_i, S_{t_i}) (S_{t_{i+2}}^k - S_{t_{i+1}}^k) \\ &= \sum_{i=0}^{n-2} \int_{2i+1}^{2i+\frac{3}{2}} f_k(\hat{X}_t^{D,b}) d\hat{X}_t^{D,f,k} \\ &= \sum_{i=0}^{n-2} \int_{2i}^{2i+2} f_k(\hat{X}_t^{D,b}) d\hat{X}_t^{D,f,k} \\ &= \int_0^1 f_k(X_t^{D,b}) dX_t^{D,f,k}. \end{aligned}$$

That is, the performance of the agent's portfolio (3.3) can be rewritten as an *exact integral* of the lag and lead components of the Hoff process. Importantly, since X^D is continuous and certainly of finite variation, the last integral is well-defined as a classical Riemann-Stieltjes integral. Thus by using the particular interpolation scheme above, via the Hoff process, we are able to recast the simple forward Riemann sums of (3.3) as an exact integral of a piecewise linear path. We note in passing that an alternative pathwise integral in the context of mathematical finance via rough path theory can be found in the preprint [97].

For the rest of this chapter we will assume that the underlying event data arises from sampling a continuous semimartingale

$$X : t \in [0, 1] \mapsto (X_t^1, \dots, X_t^d) \in \mathbb{R}^d.$$

Then a natural question to ask is what does the integral Y^D converge to as the sampling frequency increases; that is, as the mesh of our time partition D , $|D| := \max_{t_i \in D} |t_{i+1} - t_i|$, goes to 0. Such a question could arise in the situation where the agent has access to more data points, thus reducing the delay between the market behaviour and his strategy, (equivalently the delay between the forward and backward components of the Hoff process X^D).

To answer this question, we use the theory of rough paths. Specifically, we lift the Hoff process X^D to its natural rough path enhancement $\mathbb{X}^D$. The rough path lift $\mathbb{X}^D$, also known as the signature, is a well-defined transform in the sense that no information is lost (see [64]). The signed area enclosed by the different coordinates of X^D is contained in $\mathbb{X}^D$. In particular, the area between the forward and backward components of X^D is included and thus we have a natural candidate approximation for the quadratic variation of the original semimartingale X . Our first main result concerns the convergence of this area to this quadratic variation and the theorem more generally establishes the convergence of the rough Hoff process $\mathbb{X}^D$ to a rough path limit in the p -variation rough path topology as $|D| \rightarrow 0$. As a corollary of this theorem, our second result proves that the limit of the random integrals Y^D in (3.3) is the Itô integral not the Stratonovich integral as might be expected from the classical Wong-Zakai theorem (see the original paper [119] by Wong and Zakai, or the rough path versions in [30, 107]). This is almost an immediate consequence of the continuity of the Itô map of rough path theory. The basic idea of the proof is that the areas between the forward and backward components of the Hoff process, (captured by $\mathbb{X}^D$), introduce a correction factor in the stochastic integral limit, consequently allowing us to recover the Itô, not Stratonovich,

integral. To be precise, Theorem 3.4.1 implies the convergence

$$\sum_{k=1}^d \int_0^1 f_k \left(X_t^{D,b} \right) dX_t^{D,f,k} \rightarrow \int_0^1 f(X_t) dX_t := \sum_{k=1}^d \int_0^1 f_k(X_t) dX_t^k$$

as $|D| := \max_{t_i, t_{i+1} \in D} |t_{i+1} - t_i| \rightarrow 0,$

either in probability or some L^p -norm, (depending on the regularity imposed on X). We refer to [44, 46] and Theorem 17.20 of [52] for similar results concerning lead-lag driven random ODE convergence using rough path theory.

Remark 3.2.2. In a heuristic sense, this result demonstrates that the Itô integral makes good sense as a classical integral of the Hoff process and does not need a probabilistic underpinning. As a consequence, we feel that despite the focus of the present chapter in studying semimartingales as the underlying signal, our perspective and results undermine the concept to a significant extent. Taking mathematical finance as an example, almost all data presents itself initially in discrete form. The standard practice is to assume that the data arises from sampling a semimartingale signal so that Itô's theory of stochastic integration can be used to model hedging and investment strategies. It is identified in many contexts that this is a serious simplification. For example, momentum trading and long-range dependence may be better captured in a non-semimartingale context by modelling volatility using fractional Brownian motion with a Hurst parameter $H \neq \frac{1}{2}$ [10, 27, 54]. We can now see that by moving the data to the Hoff process, now regarded as a geometric rough path, we can capture our data in a model that further allows the simulation of various strategies, thus enabling simplification and summary, without making any unrealistic assumptions about the nature of the data as a semimartingale. The pathwise concept of volatility is replaced by *area* and is naturally included

in any second order description of the Hoff process as a rough path. Successful modelling of financial data needs to understand the statistics of the first few terms in the signature of the Hoff process. In general these will be incompatible with any semimartingale model.

The recent papers [59, 75, 80, 93] have also taken the approach of viewing time series data as samples of an underlying continuous process, which can in turn be lifted to a rough path. In [59] the authors have a very similar definition of a lead-lag path and apply standard regression analysis to linear functionals of the resultant rough path in order to extract interesting features of West Texas crude oil futures, (such as deciding whether a given sample trade was made in the morning or afternoon).

3.3. Rough Hoff process convergence

This section presents the first main result of the chapter. As in Section 3.2, we sample a continuous semimartingale X at a sequence of refining partitions $\{D_n\}$ and construct the corresponding sequence of Hoff processes $\{X^{D_n}\}$. In Theorem 3.3.1 we establish the convergence of the sequence of rough path lifts $\mathbb{X}^{D_n}$ of X^{D_n} to a rough path limit under the p -variation rough path topology.

Parameterization considerations. Although p -variation estimates and the limit of the Hoff processes are unchanged by reparameterization, the relative parameterization of the Hoff process with respect to the original signal is important when making precise statements about limits and convergence. Certainly the original signal is not invariant to reparameterization. To elaborate further, suppose we have a partition $D_n := \{t_i^n\}_{i=0}^n \subset [0, 1]$ and a sampled signal $X : [0, 1] \rightarrow \mathbb{R}^d$. The corresponding Hoff process $X^{D_n} = (X^{D_n, b}, X^{D_n, f})$

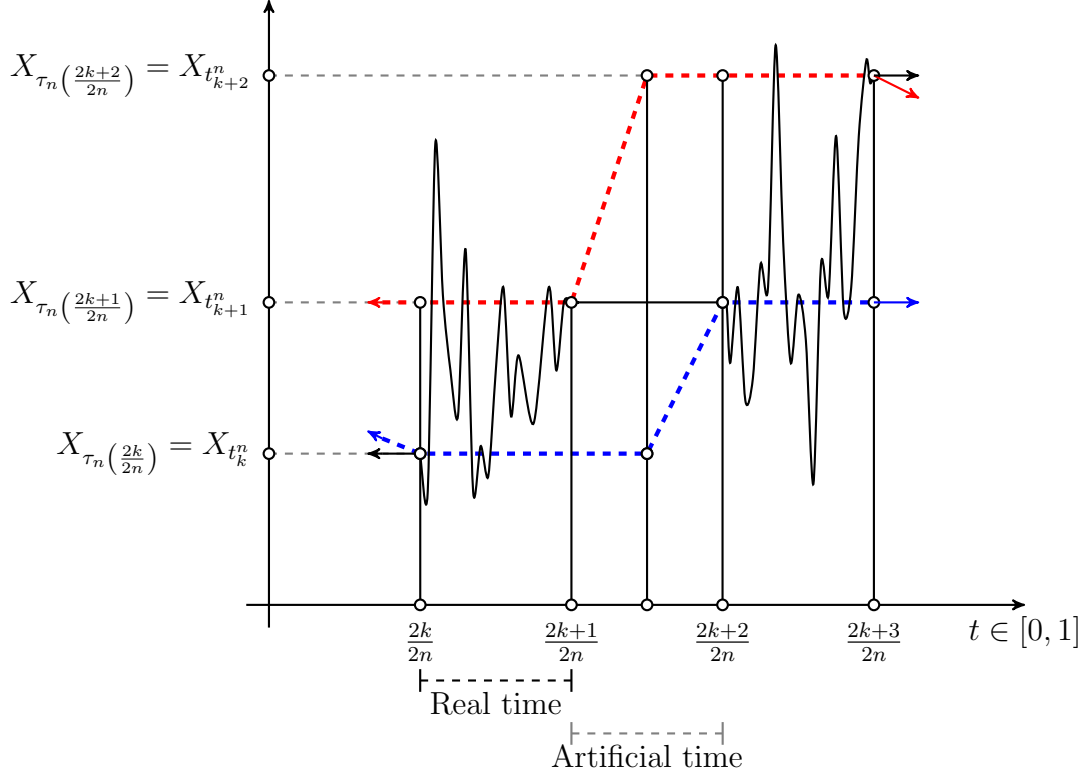


FIGURE 3.2. Parameterization of the Hoff process $X^{D_n} = (X^{D_n,b}, X^{D_n,f})$ with respect to the original signal $(X_{\tau_n(t)})_{t \in [0,1]}$ (in black), now in terms of τ_n . The lag and lead paths are in dashed blue and red lines respectively.

is not parameterized with respect to the real time of the signal but rather an artificial time; for example, we cannot say how close $X_{t_k^n}^{D_n,b}$ or $X_{t_k^n}^{D_n,f}$ are to the sampled signal point $X_{t_k^n}$ for each $t_k^n \in D_n$. Therefore we need a precise method of comparing the Hoff process to the original signal. To this end, for each partition D_n we define the corresponding time change $\tau_n : [0, 1] \rightarrow [0, 1]$ between artificial and real time by setting $\tau_n(\frac{2k}{2n}) = t_k^n$ and $\tau_n(\frac{2k+1}{2n}) = t_{k+1}^n$, with piecewise linear interpolation in between. We provide a simple example of the comparison between the original signal and the Hoff process provided by the time change τ_n in Figure 3.2, (where we assume $d = 1$ for simplicity).

Since τ_n is an increasing (continuous) function, we have no trouble in defining its generalised inverse as $\tau_n^{-1}(u) := \inf \{t \geq 0 : \tau_n(t) \geq u\}$. As intended, this time change allows us to relate the Hoff process back to the original signal:

$$X_{\frac{2k+1}{2n}}^{D_n} = X_{\tau_n^{-1}(t_k^n)}^{D_n} = (X_{t_k^n}, X_{t_{k+2}^n}) \quad \text{and} \quad X_{\tau_n(\frac{2k}{2n})} = X_{t_k^n}, \quad X_{\tau_n(\frac{2k+1}{2n})} = X_{t_{k+1}^n}.$$

A similar parameterization using artificial time was previously defined in [118] by D.R.E. Williams in the context of extending rough path theory to discontinuous Lévy processes (cf. [118, Definition 1.3]).

Rough Hoff convergence. We present both of the two main theorems of this chapter in terms of these time changes τ_n .

Theorem 3.3.1 (Rough Hoff process convergence). *Let $X = M + V : [0, 1] \rightarrow \mathbb{R}^d$ be a continuous semimartingale with Lévy area A , and define the associated Hoff processes X^{D_n} corresponding to the collection of sampling times $\{D_n\}_{n=1}^\infty \subset \mathcal{D}_{[0,1]}$. Then for all $p > 2$ we have*

$$d_{p\text{-var};[0,1]}(\mathbb{X}^{D_n}, \mathbb{X}_{\tau_n(\cdot)}^\infty) \rightarrow 0 \text{ in probability as } |D_n| \rightarrow 0,$$

where $\mathbb{X}^\infty \in C([0, 1], G^{(2)}(\mathbb{R}^{2d}))$ is defined by

$$\mathbb{X}_{s,t}^\infty = \exp_2 \left(\begin{pmatrix} X_{s,t} \\ X_{s,t} \end{pmatrix} + \begin{pmatrix} A_{s,t} & A_{s,t} - \frac{1}{2}\langle X \rangle_{s,t} \\ A_{s,t} + \frac{1}{2}\langle X \rangle_{s,t} & A_{s,t} \end{pmatrix} \right).$$

If in addition V is bounded in L^r for some $r \geq 1$ and $M \in \mathcal{M}^p$, then the convergence also holds in the $L^{p \wedge r}$ -norm.

PROOF. Throughout the proof we refer to technical results in Sections 3.6 and 3.7. First we prove the theorem for $V = 0$. Assuming $M \in \mathcal{M}^p \subseteq \mathcal{M}^q$, where $q \in (2, p)$, it follows from Proposition 3.7.1 that the rough path collection $\{\mathbb{M}^{D_n}\}_{n=1}^\infty$ satisfies uniform q -variation bounds in L^q :

$$(3.4) \quad \sup_n \left\| \|\mathbb{M}^{D_n}\|_{q\text{-var};[0,1]} \right\|_{L^q} = C_q < \infty.$$

Moreover, since each $\mathbb{M}_{\tau_n(\cdot)}$ is simply a reparameterization of the rough path lift of the original martingale $S_2(\mathbb{M})$, by [52, Proposition 14.9] we have

$$\sup_n \left\| \left\| \mathbb{M}_{\tau_n(\cdot)} \right\|_{q\text{-var};[0,1]} \right\|_{L^q} = \left\| \left\| \mathbb{M} \right\|_{q\text{-var};[0,1]} \right\|_{L^q} = C_q < \infty.$$

Recalling the interpolation result of [50, Proposition 12], for $2 < q < p$ it follows that

$$\begin{aligned} d_{p\text{-var};[0,1]}(\mathbb{M}^{D_n}, \mathbb{M}_{\tau_n(\cdot)}^\infty) &\leq d_{\infty;[0,1]}(\mathbb{M}^{D_n}, \mathbb{M}_{\tau_n(\cdot)}^\infty)^{1-q/p} \cdot d_{q\text{-var};[0,1]}(\mathbb{M}^{D_n}, \mathbb{M}_{\tau_n(\cdot)}^\infty)^{q/p} \\ &\leq d_{\infty;[0,1]}(\mathbb{M}^{D_n}, \mathbb{M}_{\tau_n(\cdot)}^\infty)^{1-q/p} \cdot \left(\left\| \mathbb{M}^{D_n} \right\|_{q\text{-var};[0,1]} + \left\| \mathbb{M}_{\tau_n(\cdot)} \right\|_{q\text{-var};[0,1]} \right)^{q/p}. \end{aligned}$$

Using Hölder's inequality with $\tilde{r} := \frac{p}{p-q} > 1$ and $s = \frac{p}{q} \in (1, 3/2)$ satisfying $\tilde{r}^{-1} + s^{-1} = 1$, we combine the maximal q -variation bounds (3.4) with the interpolation inequality to find that

$$\left| d_{p\text{-var};[0,1]}(\mathbb{M}^{D_n}, \mathbb{M}_{\tau_n(\cdot)}^\infty) \right|_{L^p} \leq C_{p,q} \left| d_{\infty;[0,1]}(\mathbb{M}^{D_n}, \mathbb{M}_{\tau_n(\cdot)}^\infty) \right|_{L^p}^{1/(p-q)}.$$

Thus the second claimed convergence in the L^p -norm follows from the uniform convergence result of Proposition 3.6.1. In the case of $M \notin \mathcal{M}^p$, we obtain convergence in probability by a simple localization argument (as done previously in [52, Theorem 14.16]).

We now turn to the general case where $V \neq 0$ and define $\bar{V} : [0, 1] \rightarrow \mathbb{R}^{2d}$ by $\bar{V}_t = (V_t, V_t)$. In contrast to the martingale rough path convergence, the corresponding convergence proof for the finite variation process is straightforward.

Lemma 3.3.2. *For every $\delta > 0$,*

$$d_{(1+\delta)\text{-var};[0,1]}(V^{D_n}, \bar{V}_{\tau_n(\cdot)}) \rightarrow 0 \text{ in probability as } |D_n| \rightarrow 0.$$

If in addition V is bounded in L^r for some $r \geq 1$ then the convergence also holds in L^r .

PROOF. The first convergence follows readily from [52, Proposition 1.28] and interpolation (see [52, §14.1] for details). The stronger convergence in the L^r -norm is a consequence of the uniform L^r -bound assumption ([25, Theorem 4.14]). $\square$

Returning to the proof of Theorem 3.3.1, it is a straightforward exercise in algebra to confirm that $\mathbb{X}_{s,t}^\infty = T_{\bar{V}}(\mathbb{X}_{s,t}^\infty)$, where $T_{\bar{V}}$ is the standard rough path translation operator defined in [52, §9.4.6]. The theorem then follows immediately from the continuity of the translation operator ([52, Corollary 9.35]). $\square$

Remark 3.3.3. The quadratic variation $\langle X \rangle = \langle M \rangle$ of the local martingale component M naturally appears as we consider the limiting area of the rough path lift $\mathbb{X}^D$. This non-linear phenomena was first observed by Hoff in the context of moving Brownian frames [66].

Remark 3.3.4. In the special case of the evenly spaced sampling sequence, that is $D_n = \{t_k^n := \frac{k}{n}\}_{k=0}^n$, we have $\tau_n(\frac{k}{n}) = \frac{k}{n}$ and $\tau_n(\frac{2k+1}{2n}) = \frac{k+1}{n}$. It follows immediately that we can drop the time change $\tau_n(\cdot)$ in Theorem 3.3.1 to arrive at the neater convergence result:

$$d_{p\text{-var};[0,1]}(\mathbb{X}^{D_n}, \mathbb{X}^\infty) \rightarrow 0 \text{ as } n \rightarrow \infty,$$

either in probability, or in $L^{p \wedge q}$, (depending on the regularity imposed on X).

3.4. Recovery of the Itô integral

We now consider the second main result of the chapter.

Theorem 3.4.1 (Recovery of the Itô integral). *Let $X : [0, 1] \rightarrow \mathbb{R}^d$ be a continuous semimartingale and let $\{D_n\}_{n=1}^\infty \subset \mathcal{D}_{[0,1]}$ be a sequence of times*

satisfying $|D_n| \rightarrow 0$. Let $f = (f_i)_{i=1}^d$ be a collection of $\text{Lip}^\gamma(\mathbb{R}^d, \mathbb{R}^q)$ mappings where $\gamma > 2$. For each n , set $X^n := X^{D_n}$ to be the associated Hoff process of the event stream $\{X_{t_i}\}_{t_i \in D_n}$. Lastly, let Y^n be the solution to the random integral equation

$$(3.5) \quad dY_t^n = f\left(X_t^{n,b}\right) dX_t^{n,f} = \sum_{i=1}^d f_i\left(X_t^{n,b}\right) dX_t^{n,f;i}, \quad Y_0^n = y_0 \in \mathbb{R}^q,$$

and set Y to be the standard Itô integral

$$dY_t = f(X_t) dX_t = \sum_{i=1}^d f_i(X_t) dX_t^i, \quad Y_0 = y_0 \in \mathbb{R}^q.$$

Then for all $p \in (2, \gamma)$,

$$(3.6) \quad d_{p\text{-var};[0,1]}(Y^n, Y_{\tau_n(\cdot)}) \rightarrow 0 \text{ in probability as } n \rightarrow \infty.$$

If in addition the finite variation process V of X is bounded in L^r for some $r \geq 1$ and $M \in \mathcal{M}^p$, then the convergence (3.6) also holds in the $L^{p \wedge r}$ -norm.

Before presenting the proof of Theorem 3.4.1 we set up some notation. Denote the rough path lift of $\bar{X}_t := (X_t, X_t)$ by $\bar{\mathbb{X}}$. Thus $\bar{\mathbb{X}} : \Delta_{[0,1]} \rightarrow G^{(2)}(\mathbb{R}^{2d})$ with

$$\mathbb{X}_{s,t} := 1 + \bar{X}_{s,t} + \frac{1}{2} \bar{X}_{s,t} \otimes \bar{X}_{s,t} + \begin{pmatrix} A_{s,t} & A_{s,t} \\ A_{s,t} & A_{s,t} \end{pmatrix},$$

where $A : \Delta_{[0,1]} \rightarrow [\mathbb{R}^d, \mathbb{R}^d]$ is the Lévy area process of X . Importantly we have $\mathbb{X}^\infty = \bar{\mathbb{X}}^\psi$; that is $\mathbb{X}^\infty$ is the rough path created from perturbing $\bar{\mathbb{X}}$ by the map $\psi : \Delta_{[0,1]} \rightarrow [\mathbb{R}^{2d}, \mathbb{R}^{2d}]$:

$$\begin{aligned} \mathbb{X}^\infty &= \bar{\mathbb{X}}_{s,t} + \psi_{s,t} = \exp_2 \left(\bar{X}_{s,t} + \begin{pmatrix} A_{s,t} & A_{s,t} \\ A_{s,t} & A_{s,t} \end{pmatrix} \right) + \psi_{s,t} \\ &= \exp_2 \left(\bar{X}_{s,t} + \begin{pmatrix} A_{s,t} & A_{s,t} \\ A_{s,t} & A_{s,t} \end{pmatrix} + \psi_{s,t} \right), \end{aligned}$$

where the path ψ is defined by

$$\psi_{s,t} := \begin{pmatrix} 0 & -\frac{1}{2} \langle X \rangle_{s,t} \\ \frac{1}{2} \langle X \rangle_{s,t} & 0 \end{pmatrix} \in [\mathbb{R}^{2d}, \mathbb{R}^{2d}].$$

The appearance of the covariation term ψ in the limit $\mathbb{X}^\infty$ of Theorem 3.3.1 allows us to recover the Itô integral from the Stratonovich integral limit. In particular, the drift terms from the Stratonovich correction cancel out. We reserve the notation $\rho : \mathbb{R}^d \oplus \mathbb{R}^q \rightarrow \mathbb{R}^q$ for the canonical projection.

PROOF OF THEOREM 3.4.1. We consider a more complicated SDE on the larger space $\mathbb{R}^d \oplus \mathbb{R}^q$ and denote the standard bases of $\mathbb{R}^d$ and $\mathbb{R}^q$ by $(\widehat{e}_i)_{i=1}^d$ and $(\bar{e}_j)_{j=1}^q$ respectively. Indeed, set $z = (\widehat{z}, \bar{z})$ to be a path in $\mathbb{R}^d \oplus \mathbb{R}^q$ and define vector fields on $\mathbb{R}^d \oplus \mathbb{R}^q$ by

$$Q_i(z) := (\widehat{e}_i, 0), \quad W_k(z) := \sum_{j=1}^q (0, \bar{e}_j) f_k^j(\widehat{z}),$$

for $i, k = 1, \dots, d$. Here we have used the standard notation convention: $f_i(x) = (f_i^1(x), \dots, f_i^q(x)) \in \mathbb{R}^q$ for $x \in \mathbb{R}^d$. Let $z^n = (\widehat{z}^{n,1}, \dots, \widehat{z}^{n,d}, \bar{z}^{n,1}, \dots, \bar{z}^{n,q})$ be the solution to the SDE

$$dz_t^n = \sum_{i=1}^d \left(Q_i(z_t^n) dX_t^{n,b;i} + W_i(z_t^n) dX_t^{n,f;i} \right), \quad z_0^n = (0, y_0) \in \mathbb{R}^d \oplus \mathbb{R}^q.$$

Since the axis path X^n is piecewise linear, (thus of finite variation), we could equivalently formulate the above SDE into its Stratonovich form without incurring a corrective drift term. It follows that the projection $\bar{z}^n := \rho(z^n)$ of z^n onto $\mathbb{R}^q$ satisfies the SDE

$$d\bar{z}_t^{n,j} = \sum_{i=1}^d W_i^j(z_t^n) dX_t^{n,f;i} = \sum_{i=1}^d f_i^j(\widehat{z}_t^n) dX_t^{n,f;i}.$$

Moreover, Theorem 17.3 of [52] tells us that $z^n = \pi_{(Q,W)}(0, (0, y_0); \mathbb{X}^n)$ almost surely.

A simple computation confirms that the only non-zero Lie brackets between the vector fields $Q_{i_1}, Q_{i_2}, W_{k_1}$ and W_{k_2} are of the form

$$[Q_i, W_k](\widehat{z}) = Q_i W_k(\widehat{z}) = \sum_{j=1}^q (0, \bar{e}_j) \frac{\partial f_k^j}{\partial x_i}(\widehat{z}),$$

for all $i, k = 1, \dots, d$. We denote this collection of vector fields by

$$L := (Q_i W_k : i, k \in \{1, \dots, d\}).$$

Next consider the SDE:

$$\begin{aligned} dz_t &= (Q + W)(z_t) dX_t = (Q + W)(z_t) \circ dX_t - \frac{1}{2} \sum_{i,k=1}^d Q_i W_k(z_t) d\langle X^i, X^k \rangle_t \\ &= (Q + W)(z_t) \circ dX_t - \frac{1}{2} \sum_{i,k=1}^d [Q_i, W_k](z_t) d\langle X^i, X^k \rangle_t, \end{aligned}$$

with initial condition $z_0 = (0, y_0) \in \mathbb{R}^d \oplus \mathbb{R}^q$. Again appealing to Theorem 17.3 of [52], it follows that almost surely

$$z = \pi_{(Q,W;L)}(0, (0, y_0); (\mathbb{X}, \psi)).$$

As noted above, $\overline{\mathbb{X}}^\psi = \mathbb{X}^\infty$, and so the perturbation theorem of [52, Theorem 12.14] guarantees the equality:

$$z = \pi_{(Q,W;L)}(0, (0, y_0); (\mathbb{X}, \psi)) = \pi_{(Q,W)}(0, (0, y_0); \overline{\mathbb{X}}^\psi) = \pi_{(Q,W)}(0, (0, y_0); \mathbb{X}^\infty).$$

In light of the convergence result of Theorem 3.3.1, we conclude that for all $p \in (2, \infty)$,

$$d_{p\text{-var};[0,1]}(z^n, z_{\tau_n(\cdot)}) \rightarrow 0 \text{ in probability (or in } L^{p \wedge r}) \text{ as } n \rightarrow \infty.$$

Indeed, as noted in the above remark, almost sure equality and convergence in probability (and L^p) are all preserved under continuous mappings. Since the Itô map is continuous with respect to the driving rough path signal under the p -variation topology, the proof is then finished by noting that $Y^n = \rho(z^n)$ and $Y = \rho(z)$. $\square$

The remainder of the chapter is devoted to establishing the technical results referred to in the proof of Theorem 3.3.1.

3.5. Relation to original work of Hoff

We now discuss the connections between the main results of this chapter and the original contribution of Hoff. As discussed in the previous section, the appearance of quadratic variation in various semimartingale rough path limits was first considered by Hoff in his thesis [66]. He used this non-linear phenomenon to prove the following convergence result concerning stochastic delay differential equations. We note that he did not formally state the result, rather it was first presented in [79, §3.3].

Theorem 3.5.1 (Hoff 2005). *Fix vector fields $V_1, V_2 \in \text{Lip}^2(\mathbb{R}^q)$. Given $\varepsilon > 0$, let W^ε be the ε -delayed Brownian motion defined by $W_t^\varepsilon := W_{t-\varepsilon} \mathbf{1}\{t \geq \varepsilon\}$. For every $\varepsilon > 0$, let Y^ε be the solution of the anticipating Stratonovich SDE*

$$dY_t^\varepsilon = V_1(Y_t^\varepsilon) \circ dW_t + V_2(Y_t^\varepsilon) \circ dW_t^\varepsilon, \quad Y_0^\varepsilon = y_0 \in \mathbb{R}^q.$$

Moreover, let Y be the solution to the (standard) Stratonovich SDE

$$dY_t = (V_1(Y_t) + V_2(Y_t)) \circ dW_t - \frac{1}{2}[V_1, V_2](Y_t) dt, \quad Y_0 = y_0.$$

Then for every $p \in (2, \infty)$, we have

$$\left| d_{1/p\text{-H\"{o}l};[0,1]}(Y^\varepsilon, Y) \right|_{L^r} \rightarrow 0 \text{ as } \varepsilon \rightarrow 0$$

for all $r \in [1, \infty)$.

This result is counterintuitive. Certainly $(W^\varepsilon, W) = W^\varepsilon \oplus W$ converges almost surely and in L^p for all $p \in [1, \infty)$ to $(W, W) = W \oplus W$ as $\varepsilon \rightarrow 0$. Thus it seems reasonable to expect the process Y^ε to converge to $\tilde{Y}$, where the latter solves the SDE

$$d\tilde{Y}_t = \left(V_1(\tilde{Y}_t) + V_2(\tilde{Y}_t) \right) \circ dW_t.$$

However, Hoff's original proof established that for every $p \in (2, \infty)$ we have

$$(3.7) \quad |d_{1/p\text{-H\"ol};[0,1]}(S_2(W^\varepsilon \oplus W), \mathbb{W}^\infty)|_{L^r} \rightarrow 0 \text{ as } \varepsilon \rightarrow 0,$$

for all $r \in [1, \infty)$. Here, $\mathbb{W}^\infty \in C([0, 1], G^{(2)}(\mathbb{R}^2))$ is defined by

$$\mathbb{W}_{s,t}^\infty := \exp_2 \left(\begin{pmatrix} W_{s,t} \\ W_{s,t} \end{pmatrix} + \begin{pmatrix} 0 & -\frac{1}{2}(t-s) \\ \frac{1}{2}(t-s) & 0 \end{pmatrix} \right).$$

Considered as rough path controls, the rough path lift of (W^ε, W) does not converge to the rough path lift of $W \oplus W$, namely $S_2(W \oplus W)_t = \exp_2(W_t \oplus W_t)$. The appearance of the quadratic variation term $\frac{1}{2} \langle W \rangle_{s,t} = \frac{1}{2}(t-s)$ in the limiting rough path is the key ingredient; this non-zero area term produces the additional drift term of $-\frac{1}{2}[V_1, V_2](Y_t) dt$ in the limiting SDE of Theorem 3.5.1.

In a heuristic sense, the two main theorems of this chapter concerning the convergence of rough Hoff paths can be thought of as a discrete analogue of Hoff's original results (3.7) and Theorem 3.5.1. Instead of delayed Brownian motion, we are considering Riemann sum approximations of lead-lag paths based on semimartingale signals.

Remark 3.5.2. The proof of Hoff's result given in [52, Theorem 17.20] by Friz and Victoir was the blueprint for the arguments of Theorems 3.3.1 and 3.4.1. Indeed, as in our argument, once the convergence (3.7) is established, Theorem 3.5.1 is an immediate corollary of the result of the universal limit theorem (Theorem 2.7.2) and the central perturbations of RDEs result of Theorem 2.6.8.

The argument to prove (3.7) is actually much shorter and straightforward because the underlying signal is Brownian motion and so Hölder rough path norms can be employed. We do not have this advantage in our semimartingale case and are forced to work with more difficult p -variation rough path metrics. As an additional complication, our proof requires non-trivial maximal

p -variation estimates for all sampling partitions, (which will be dealt with in Section 3.7).

The remainder of the chapter is devoted to establishing the auxiliary regularity results referred to in the proof of Theorem 3.3.1; namely, the uniform convergence of the rough Hoff process lift and the maximal p -variation norm bounds. Throughout all of the following proofs we make the simplifying assumption that $X_0 = 0$. Since the signature of a path is invariant under translation, this assumption comes at no cost to generality.

3.6. Uniform convergence of the Hoff process

In this section we establish the following uniform convergence in L^p of the Hoff process in the case that $X = M \in \mathcal{M}^p$.

Proposition 3.6.1. *Let $M \in \mathcal{M}^p$ for some $p > 2$ and let $\{D_n\}_{n=1}^\infty \subset \mathcal{D}_{[0,1]}$ be a sequence of partitions with $|D_n| \rightarrow 0$. Then*

$$\left| d_{\infty;[0,1]}(\mathbb{M}^{D_n}, \mathbb{M}_{\tau_n(\cdot)}) \right|_{L^p} \rightarrow 0 \text{ as } |D_n| \rightarrow 0.$$

Before presenting a proof, we introduce two useful technical lemmas concerning $\mathcal{M}^p$ -martingales. We make frequent use of the famous Burkholder-Davis-Gundy (BDG) inequality ([105, Theorem 42.1]).

Lemma 3.6.2. *Suppose $X : [0, 1] \rightarrow \mathbb{R}$ is a $\mathcal{M}^p$ -martingale for some $p > 2$. Then there exists a constant $C_p > 0$ such that for all $D = \{t_k\}_{k=0}^n \subset [0, 1]$, we have*

$$\mathbb{E} \left(\sup_{t_k \in D} \sup_{t \in [t_k, t_{k+1}]} |X_t - X_{t_k}|^p \right) \leq C_p \|X\|_p^2 \mathbb{E} \left(\sup_{t_k \in D} \langle X \rangle_{t_k, t_{k+1}}^{p/2} \right)^{\frac{p-2}{p}}.$$

PROOF. We first apply the BDG inequality to deduce that

$$\begin{aligned} \mathbb{E} \left(\sup_{t_k \in D} \sup_{t \in [t_k, t_{k+1}]} |X_t - X_{t_k}|^p \right) &\leq \sum_{t_k \in D} \mathbb{E} \left(\sup_{t \in [t_k, t_{k+1}]} |X_t - X_{t_k}|^p \right) \\ &\leq C_p \sum_{t_k \in D} \mathbb{E} \left(\langle X \rangle_{t_k, t_{k+1}}^{p/2} \right). \end{aligned}$$

Then we apply the Hölder inequality with $q = p/2$ and $r = p/(p-2)$, (so that $q^{-1} + r^{-1} = 1$), to find

$$\begin{aligned} \sum_{t_k \in D} \mathbb{E} \left(\langle X \rangle_{t_k, t_{k+1}}^{p/2} \right) &= \mathbb{E} \left(\sum_{t_k \in D} \langle X \rangle_{t_k, t_{k+1}} \cdot \sup_{t_k \in D} \langle X \rangle_{t_k, t_{k+1}}^{p/2-1} \right) \\ &= \mathbb{E} \left(\langle X \rangle_{0,1} \cdot \sup_{t_k \in D} \langle X \rangle_{t_k, t_{k+1}}^{p/2-1} \right) \\ &\leq |\langle X \rangle_{0,1}|_{L^q} \cdot \left| \sup_{t_k \in D} \langle X \rangle_{t_k, t_{k+1}}^{p/2-1} \right|_{L^r} \\ &= \mathbb{E} \left(\langle X \rangle_{0,1}^{p/2} \right)^{2/p} \mathbb{E} \left(\sup_{t_k \in D} \langle X \rangle_{t_k, t_{k+1}}^{p/2} \right)^{\frac{p-2}{p}}. \end{aligned}$$

Again using the BDG inequality, $\mathbb{E}(\langle X \rangle_{0,1}^{p/2}) \leq C_p \|X\|_p^p$, and the proof is complete. $\square$

Lemma 3.6.3. *Suppose $X, Y : [0, 1] \rightarrow \mathbb{R}$ are $\mathcal{M}^p$ -martingales for some $p > 2$. Then there exists a constant $C_p > 0$ such that for $D = \{t_k\}_{k=0}^n \subset [0, 1]$, the following inequalities hold:*

$$\begin{aligned} \mathbb{E} \left(\sup_{t_k \in D} \left| \sum_{t_i \leq t_k} X_{t_{i-1}} (Y_{t_i} - Y_{t_{i-1}}) - \int_0^{t_k} X_r dY_r \right|^{p/2} \right) \\ \leq C_p \|Y\|_p^{p/2} \|X\|_p \mathbb{E} \left(\sup_{t_k \in D} \langle X \rangle_{t_k, t_{k+1}}^{p/2} \right)^{\frac{p-2}{2p}}; \\ \mathbb{E} \left(\sup_{t_k \in D} \left| \sum_{t_i \leq t_k} (X_{t_{i+1}} - X_{t_i}) (Y_{t_{i+2}} - Y_{t_{i+1}}) \right|^{p/2} \right) \\ \leq C_p \|X\|_p^{p/2} \|Y\|_p \mathbb{E} \left(\sup_{t_k \in D} \langle Y \rangle_{t_k, t_{k+1}}^{p/2} \right)^{\frac{p-2}{2p}}. \end{aligned}$$

PROOF. Define the previsible process $Z_r := \sum_{0 \neq t_i \in D} X_{t_{i-1}} \mathbf{1}\{r \in (t_i, t_{i+1}]\}$. By the Kunita-Watanabe identity, along with the BDG inequality, we have

$$\begin{aligned} & \mathbb{E} \left(\sup_{t_k \in D} \left| \sum_{t_i \leq t_k} X_{t_{i-1}} (Y_{t_i} - Y_{t_{i-1}}) - \int_0^{t_k} X_r dY_r \right|^{p/2} \right) \\ &= \mathbb{E} \left(\sup_{t_k \in D} \left| \int_0^{t_k} (Z_r - X_r) dY_r \right|^{p/2} \right) \\ &\leq C_p \mathbb{E} \left(\left| \int_0^1 |Z_r - X_r|^2 d\langle Y \rangle_r \right|^{p/4} \right) \leq C_p \mathbb{E} \left(\sup_{r \in [0,1]} |Z_r - X_r|^{p/2} \langle Y \rangle_{0,1}^{p/4} \right). \end{aligned}$$

The Cauchy-Schwarz inequality then yields

$$\mathbb{E} \left(\sup_{r \in [0,1]} |Z_r - X_r|^{p/2} \langle Y \rangle_{0,1}^{p/4} \right) \leq \|Y\|_p^{p/2} \sqrt{\mathbb{E} \left(\sup_{r \in [0,1]} |Z_r - X_r|^p \right)}.$$

Note that

$$\begin{aligned} \mathbb{E} \left(\sup_{r \in [0,1]} |Z_r - X_r|^p \right) &= \mathbb{E} \left(\sup_{t_i \in D} \sup_{t \in [t_{i+1}, t_{i+2}]} |X_t - X_{t_i}|^p \right) \\ &\leq C_p \mathbb{E} \left(\sup_{t_i \in D} \sup_{t \in [t_i, t_{i+1}]} |X_t - X_{t_i}|^p \right). \end{aligned}$$

Thus Lemma 3.6.2 gives:

$$\mathbb{E} \left(\sup_{r \in [0,1]} |Z_r - X_r|^p \right) \leq C_p \|X\|_p^2 \mathbb{E} \left(\sup_{t_k \in D} \langle X \rangle_{t_k, t_{k+1}}^{p/2} \right)^{\frac{p-2}{p}},$$

and the first claimed inequality follows. The second inequality has an almost identical proof; we simply use $Z_r := \sum_{0 \neq t_i \in D} Y_{t_{i-1}} \mathbf{1}\{r \in (t_i, t_{i+1}]\}$ instead. $\square$

PROOF OF PROPOSITION 3.6.1. For notational convenience we drop the usual n superscript in t_k^n and just write $D_n = \{t_k\}_{k=0}^n$. We first consider the increment level. For $\theta \in \{b, f\}$, using Lemma 3.6.2 gives

$$\mathbb{E} \left(\sup_{s \neq t} \left| M_{s,t}^{D_n, \theta; i} - M_{\tau_n(s), \tau_n(t)}^i \right|^p \right) \leq C_p \|M^i\|_p^2 \mathbb{E} \left(\sup_{t_k \in D_n} \langle M^i \rangle_{t_k, t_{k+1}}^{p/2} \right)^{\frac{p-2}{p}}.$$

Since $M \in \mathcal{M}^p$, we use the Dominated Convergence Theorem (DCT) to deduce that

$$\mathbb{E} \left(\sup_{s \neq t} \left| M_{s,t}^{D_n, \theta; i} - M_{\tau_n(s), \tau_n(t)}^i \right|^p \right) \rightarrow 0 \text{ as } |D_n| \rightarrow 0.$$

We now turn our attention to the second level. Let us denote the Lévy area process associated with M^{D_n} by

$$\begin{aligned} A_{s,t}^{D_n, \theta, \lambda; i, j} &:= \frac{1}{2} \int_s^t (M_{s,r}^{D_n, \theta; i} dM_r^{D_n, \lambda; j} - M_{s,r}^{D_n, \lambda; j} dM_r^{D_n, \theta; i}) \\ (3.8) \quad &= \int_s^t M_r^{D_n, \theta; i} dM_r^{D_n, \lambda; j} - \frac{1}{2} M_{s,t}^{D_n, \theta; i} M_{s,t}^{D_n, \lambda; j}, \end{aligned}$$

where $\theta, \lambda \in \{b, f\}$ and $i, j \in \{1, \dots, d\}$. Since M^{D_n} is piecewise linear, (and thus has zero quadratic variation), we are able to switch between the Itô and Stratonovich integrals in (3.8) without incurring a correction term.

Recalling (2.3), we have

$$d_{\infty; [0,1]}(\mathbb{M}^{D_n}, \mathbb{M}_{\tau_n(\cdot)}) \leq C_d \sup_{s \neq t} \left(\|M_{s,t}^{D_n} - \overline{M}_{\tau_n(s), \tau_n(t)}\|^p + \|A_{s,t}^{D_n} - \widehat{A}_{\tau_n(s), \tau_n(t)}\|^{p/2} \right),$$

where $\widehat{A} = \{\widehat{A}^{\theta, \lambda; i, j}\}$ is defined by

$$\widehat{A}_{s,t}^{\theta, \lambda; i, j} := \begin{cases} A_{s,t}^{M; i, j} & \text{if } \theta = \lambda; \\ A_{s,t}^{M; i, j} - \frac{1}{2} \langle M^i, M^j \rangle_{s,t} & \text{if } \theta = b, \lambda = f; \\ A_{s,t}^{M; i, j} + \frac{1}{2} \langle M^i, M^j \rangle_{s,t} & \text{if } \theta = f, \lambda = b. \end{cases}$$

Thus to complete the proof it suffices to show that

$$(3.9) \quad \mathbb{E} \left(\sup_{s \neq t} \left| A_{s,t}^{D_n, \theta, \lambda; i, j} - \widehat{A}_{\tau_n(s), \tau_n(t)}^{\theta, \lambda; i, j} \right|^{p/2} \right) \rightarrow 0 \text{ as } |D_n| \rightarrow 0.$$

Given a bounded variation path $x : [0, 1] \rightarrow \mathbb{R}^d$ with area process $A : \Delta_{[0,1]} \rightarrow [\mathbb{R}^d, \mathbb{R}^d]$, we have the decomposition:

$$A_{s,t} = A_{0,t} - A_{0,s} - \frac{1}{2} [x_{0,s}, x_{s,t}] \text{ for all } (s, t) \in \Delta_{[0,1]}.$$

In light of this formula, the fact that straight line segments enclose zero area, and the previously established uniform convergence at level 1, in order to prove

(3.9) it is sufficient to show

$$\mathbb{E} \left(\sup_{t_k \in D_n} \left| A_{0, \frac{k}{n}}^{D_n, \theta, \lambda; i, j} - \widehat{A}_{0, \tau_n(\frac{k}{n})}^{\theta, \lambda; i, j} \right|^{p/2} \right) \rightarrow 0 \text{ as } |D_n| \rightarrow 0.$$

To this end, we first consider the case of $(\theta, \lambda) = (b, b)$, (the case of (f, f) is almost identical and thus omitted). Setting $X := M^i$, $Y := M^j$ allows us to free up (i, j) for index notation. A simple computation gives:

$$\begin{aligned} A_{0, \frac{k}{n}}^{D_n, b, b; i, j} &= \frac{1}{2} \sum_{i \leq k} \{X_{t_{i-1}}(Y_{t_i} - Y_{t_{i-1}}) - Y_{t_{i-1}}(X_{t_i} - X_{t_{i-1}})\}; \\ \widehat{A}_{0, \tau_n(\frac{k}{n})}^{b, b; i, j} &= A_{0, t_k}^{M; i, j} = \frac{1}{2} \int_0^{t_k} (X_r dY_r - Y_r dX_r). \end{aligned}$$

It follows that

$$\begin{aligned} &\mathbb{E} \left(\sup_{t_k \in D_n} \left| A_{0, \frac{k}{n}}^{D_n, b, b; i, j} - \widehat{A}_{0, \tau_n(\frac{k}{n})}^{b, b; i, j} \right|^{p/2} \right) \\ &\leq C_p \mathbb{E} \left(\sup_{t_k \in D_n} \left| \sum_{t_i \leq t_k} X_{t_{i-1}}(Y_{t_i} - Y_{t_{i-1}}) - \int_0^{t_k} X_r dY_r \right|^{p/2} \right) \\ &\quad + C_p \mathbb{E} \left(\sup_{t_k \in D_n} \left| \sum_{t_i \leq t_k} Y_{t_{i-1}}(X_{t_i} - X_{t_{i-1}}) - \int_0^{t_k} Y_r dX_r \right|^{p/2} \right). \end{aligned}$$

By applying Lemma 3.6.3 to both terms, we can then use the DCT to find that

$$\begin{aligned} &\mathbb{E} \left(\sup_{t_k \in D_n} \left| A_{0, \frac{k}{n}}^{D_n, b, b; i, j} - \widehat{A}_{0, \tau_n(\frac{k}{n})}^{b, b; i, j} \right|^{p/2} \right) \\ &\leq C_{p, M} \left(\mathbb{E} \left(\sup_{t_k \in D_n} \langle X \rangle_{t_k, t_{k+1}}^{p/2} \right)^{\frac{p-2}{2p}} + \mathbb{E} \left(\sup_{t_k \in D_n} \langle Y \rangle_{t_k, t_{k+1}}^{p/2} \right)^{\frac{p-2}{2p}} \right) \rightarrow 0 \text{ as } |D_n| \rightarrow 0. \end{aligned}$$

Since $A_{0, \frac{k}{n}}^{D_n, b, f; i, j} = -A_{0, \frac{k}{n}}^{D_n, f, b; j, i}$, it only remains to prove the claimed convergence for the case of $(\theta, \lambda) = (b, f)$. To this end, we first note that

$$(3.10) \quad A_{0, \frac{k}{n}}^{D_n, b, f; i, j} = \sum_{t_i \leq t_k} X_{t_i}(Y_{t_{i+2}} - Y_{t_{i+1}}) - \frac{1}{2} X_{0, t_k} Y_{t_1, t_{k+2}}.$$

We have the decomposition

$$\sum_{t_i \leq t_k} X_{t_i}(Y_{t_{i+2}} - Y_{t_{i+1}}) = \sum_{t_i \leq t_k} X_{t_{i+1}}(Y_{t_{i+2}} - Y_{t_{i+1}}) - \sum_{t_i \leq t_k} (X_{t_{i+1}} - X_{t_i})(Y_{t_{i+2}} - Y_{t_{i+1}}).$$

As before, we use the second inequality of Lemma 3.6.3 to deal with the second term:

$$\mathbb{E} \left(\sup_{t_k \in D_n} \left| \sum_{t_i \leq t_k} (X_{t_{i+1}} - X_{t_i}) (Y_{t_{i+2}} - Y_{t_{i+1}}) \right|^{p/2} \right) \leq C_{p,X,Y} \mathbb{E} \left(\sup_{t_k \in D_n} \langle Y \rangle_{t_k, t_{k+1}}^{p/2} \right)^{\frac{p-2}{2p}}.$$

The first inequality of the same lemma guarantees that

$$\mathbb{E} \left(\sup_{t_k \in D_n} \left| \sum_{t_i \leq t_k} X_{t_i} (Y_{t_{i+1}} - Y_{t_i}) - \int_0^{t_k} X_r dY_r \right|^{p/2} \right) \leq C_{p,X,Y} \mathbb{E} \left(\sup_{t_k \in D_n} \langle X \rangle_{t_k, t_{k+1}}^{p/2} \right)^{\frac{p-2}{2p}}.$$

Therefore,

$$\begin{aligned} & \mathbb{E} \left(\sup_{t_k \in D_n} \left| A_{0, \frac{k}{n}}^{D_n, b, f; i, j} - \left(\int_0^{t_k} X_r dY_r - \frac{1}{2} X_{t_k} Y_{t_k} \right) \right|^{p/2} \right) \\ (3.11) \quad & \leq C_{p,X,Y} 2^{p/2-1} \left(\mathbb{E} \left(\sup_{t_k \in D_n} \langle X \rangle_{t_k, t_{k+1}}^{p/2} \right)^{\frac{p-2}{2p}} + \mathbb{E} \left(\sup_{t_k \in D_n} \langle Y \rangle_{t_k, t_{k+1}}^{p/2} \right)^{\frac{p-2}{2p}} \right). \end{aligned}$$

Rearranging via the integration by parts formula for continuous (semi-)martingales M, N :

$$d(MN)_t = M_t dN_t + N_t dM_t + d\langle M, N \rangle_t,$$

we arrive at the identity

$$\begin{aligned} & \int_0^{t_k} X_r dY_r - \frac{1}{2} X_{t_k} Y_{t_k} \\ &= \int_0^{t_k} X_r dY_r - \frac{1}{2} \left(\int_0^{t_k} X_r dY_r + \int_0^{t_k} Y_r dX_r + \langle X, Y \rangle_{0, t_k} \right) \\ &= \frac{1}{2} \int_0^{t_k} (X_r dY_r - Y_r dX_r) - \frac{1}{2} \langle X, Y \rangle_{0, t_k} = A_{0, t_k}^{M; i, j} - \frac{1}{2} \langle X, Y \rangle_{0, t_k} = \widehat{A}_{0, \frac{k}{n}}^{b, f; i, j}. \end{aligned}$$

Thus the desired convergence (3.9) for the case of $(\theta, \lambda) = (b, f)$ is deduced by applying the DCT to (3.11). The proof is complete. $\square$

3.7. Maximal p -variation bounds

The objective of this section is to prove the following maximal p -variation bound.

Proposition 3.7.1 (Maximal p -variation bounds). *For all $p > 2$, there exists a constant $C_p > 0$ such that set of all rough Hoff paths $\mathbb{M}^D$ satisfies*

$$\sup_{D \in \mathcal{D}_{[0,1]}} \mathbb{E} \left(\left\| \mathbb{M}^D \right\|_{p\text{-var};[0,1]}^p \right) \leq C_p \|M\|_p^p < \infty.$$

We split the proof into two propositions concerning $M^{D,\theta,\lambda}$; the first dealing with $\theta = \lambda$ and the second $\theta \neq \lambda$, where $\theta, \lambda \in \{b, f\}$.

Proposition 3.7.2. *For all $p > 2$ and $\theta \in \{b, f\}$, there exists a constant $C_p > 0$ such that*

$$\sup_{D \in \mathcal{D}_{[0,1]}} \mathbb{E} \left(\left\| S_2 \left(M^{D,\theta,\theta} \right) \right\|_{p\text{-var};[0,1]}^p \right) \leq C_p \|M\|_p^p.$$

PROOF. For all $\theta \in \{b, f\}$, $M^{D,\theta}$ is a reparameterization of the standard piecewise linear approximation $\tilde{M}^D$ of M based on the partition D . Since p -variation is invariant under reparameterization, it follows from the Burkholder-Davis-Gundy rough path result of [52, Theorem 14.15], (using the moderate function $F(x) = x^p$), that we have

$$\mathbb{E} \left(\left\| S_2 \left(M^{D,\theta} \right) \right\|_{p\text{-var};[0,1]}^p \right) = \mathbb{E} \left(\left\| S_2 \left(\tilde{M}^D \right) \right\|_{p\text{-var};[0,1]}^p \right) \leq C_p \mathbb{E} \left(\langle M \rangle_{0,1}^{p/2} \right).$$

By the classical BDG inequality, $\mathbb{E}(\langle M \rangle_{0,1}^{p/2}) \leq C_p \|M\|_p^p$, and the proof is complete. $\square$

Remark 3.7.3. The use of [52, Theorem 14.15] in the proof of Proposition 3.7.2 critically requires that $p > 2$.

We now consider the case of $\theta \neq \lambda$. As in the proof of Proposition 3.6.1, by symmetry it suffices to only consider the case of $\theta = b \neq \lambda = f$.

Proposition 3.7.4. *For all $p > 2$, there exists a constant $C_p > 0$ such that*

$$\sup_{D \in \mathcal{D}_{[0,1]}} \mathbb{E} \left(\left\| S_2 \left(M^{D,b,f} \right) \right\|_{p\text{-var};[0,1]}^p \right) \leq C_p \|M\|_p^p.$$

PROOF. Since Proposition 3.7.2 has already established that

$$\sup_{D \in \mathcal{D}_{[0,1]}} \mathbb{E} \left(\left\| \pi_1 \left(S_2 \left(M^{D,b,f} \right) \right) \right\|_{p\text{-var};[0,1]}^p \right) \leq C_p \|M\|_p^p,$$

it is sufficient to show that for all $D \in \mathcal{D}_{[0,1]}$, there exists a constant C_p independent of D such that

$$\mathbb{E} \left(\sup_{P \in \mathcal{D}_{[0,1]}} \sum_{s_j \in P} \left| A_{s_j, s_{j+1}}^{D,b,f;i,j} \right|^{p/2} \right) = \mathbb{E} \left(\left| A^{D,b,f;i,j} \right|_{p/2\text{-var};[0,1]}^{p/2} \right) \leq C_p \|M\|_p^p.$$

We begin by fixing a partition $D = \{t_k\}_{k=0}^n \in \mathcal{D}_{[0,1]}$. As done previously, let $X := M^i$ and $Y := M^j$. By summing up the absolute area of the triangles enclosing the piecewise axis path $(M^{D,b;i}, M^{D,f;j})$ over each interval $[\frac{2k}{2n}, \frac{2k+1}{2n}]$, we find that

$$\begin{aligned} \left| A^{D,b,f;i,j} \right|_{p/2\text{-var};[0,1]} &\leq \left| A^{D,b,f;i,j} \right|_{1\text{-var};[0,1]} \\ &= \sum_{k=0}^{n-1} \left| A^{D,b,f;i,j} \right|_{1\text{-var};[\frac{2k}{2n}, \frac{2k+2}{2n}]} = \frac{1}{2} \sum_{t_k \in D} \left| X_{t_k, t_{k+1}} \right| \cdot \left| Y_{t_{k+1}, t_{k+2}} \right|. \end{aligned}$$

Therefore, using the inequalities $ab \leq \frac{1}{2}(a^2 + b^2)$ and $(a + b)^r \leq 2^{r-1}(a^r + b^r)$ for $a, b \geq 0$, $r \geq 1$, we find

$$\begin{aligned} \mathbb{E} \left(\left| A^{D,b,f;i,j} \right|_{p/2\text{-var};[0,1]}^{p/2} \right) &\leq \mathbb{E} \left(\left| A^{D,b,f;i,j} \right|_{1\text{-var};[0,1]}^{p/2} \right) \\ &= 2^{-p/2} \mathbb{E} \left(\left| \sum_{t_k \in D} \left| X_{t_k, t_{k+1}} \right| \cdot \left| Y_{t_{k+1}, t_{k+2}} \right| \right|^{p/2} \right) \\ &\leq \frac{1}{2} \mathbb{E} \left(\left| \sum_{t_k \in D} \left| X_{t_k, t_{k+1}} \right|^2 \right|^{p/2} + \left| \sum_{t_k \in D} \left| Y_{t_k, t_{k+1}} \right|^2 \right|^{p/2} \right). \end{aligned}$$

From the discrete Burkholder-Davis-Gundy inequality ([16, Theorem 1.1]), followed by the continuous BDG inequality, we deduce that

$$\begin{aligned} \mathbb{E} \left(\left| \sum_{t_k \in D} \left| X_{t_k, t_{k+1}} \right|^2 \right|^{p/2} \right) &\leq C_p \mathbb{E} \left(\sup_{t_k \in D} |X_{t_k} - X_0|^p \right) \leq C_p \mathbb{E} \left(\sup_{t \in [0,1]} |X_t - X_0|^p \right) \\ &\leq C_p \|M\|_p^p. \end{aligned}$$

An identical inequality holds for the sum of squares term involving Y . As this bound is independent of the original partition choice of D , the proof is complete. $\square$

3.8. Possible extension to fractional Brownian motion

Instead of a semimartingale as our underlying signal X , we ask whether it is possible to extend the results of this chapter to a Hoff process defined from a fractional Brownian motion $X^H : [0, 1] \rightarrow \mathbb{R}^d$ with Hurst parameter $H \neq \frac{1}{2}$. As discussed in [65], such a model is desirable in mathematical finance; fractional Brownian motion retains much of the simplicity of ordinary Brownian motion (stationarity, continuity, Gaussian increments) but dispenses with independence of increments. This last property allows for “mean-reversion” effects (that is, negatively correlated increments) in the case of $H < \frac{1}{2}$, and conversely allows for “momentum effects” (positively correlated increments) when $H > \frac{1}{2}$.

The case of $H > \frac{1}{2}$. In this case X^H has finite p -variation for all $p > H^{-1}$ and zero quadratic variation almost surely ([81, Chapter IV]). Let $X^{H,D}$ denote the Hoff process corresponding to a partition $D \in \mathcal{D}_{[0,1]}$. It is straightforward to establish the convergence

$$\left| d_{1/p\text{-HöL};[0,1]}(X^{H,D}, X^H \oplus X^H) \right|_{L^q} \rightarrow 0 \text{ as } |D| \rightarrow 0,$$

for all $q \in [1, \infty)$. Since we can choose $p \in (H^{-1}, 2)$ so that $\lfloor p \rfloor = 1$, the universal limit theorem applies and we conclude that the limit of differential equations driven by lead-lag paths converges to the corresponding differential equation driven by X^H . Thus construction of the Hoff process is superfluous.

The case of $H < \frac{1}{2}$. For simplicity, let us assume that $d = 1$. Then for $p > 2$,

$$\left| \sum_{t_k \in D} \left| X_{t_k, t_{k+1}}^H \right|^2 \right|_{L^{p/2}} \geq \left| \sum_{t_k \in D} \left| X_{t_k, t_{k+1}}^H \right|^2 \right|_{L^1} = \sum_{t_k \in D} |t_{k+1} - t_k|^{2H} \rightarrow \infty$$

as $|D| \rightarrow 0$. Thus the Burkholder-Davis-Gundy-based argument in the proof of Proposition 3.7.4, (which was used to bound the individual area triangles enclosed by the Hoff path $(X^{D,b}, X^{D,f})$), cannot be generalised to the case where the underlying signal is a fractional Brownian motion (at least when $H < \frac{1}{2}$). This is the first obstacle we encounter when attempting to extend the convergence results of Theorems 3.3.1 and 3.4.1 to this non-semimartingale setting. Perhaps other techniques exploiting the properties of Gaussian rough paths are needed (see [52, Chapter XV]), but the author is uncertain on whether such an extension is even possible; recall that the quadratic variation of fractional Brownian motion with Hurst index $H < \frac{1}{2}$ is infinite [104, §2].

CHAPTER 4

Pathwise approximation of stochastic differential equations

In this chapter we present the notion of approximating SDEs in a pathwise sense. To do this we first define the Wasserstein metric from optimal transport theory and introduce the idea of a coupling between two random variables. We then discuss strong SDE approximations and how these schemes are limited in their convergence rates by the results of Cameron and Clark [17] unless they employ higher iterated integrals of Brownian motion. Next, we review previous research on pathwise approximation and study schemes which do not require the computationally expensive simulation of the Lévy area increments yet achieve superior rates of convergence in the Wasserstein metric. In order to do this, these papers impose strict Hörmander conditions on the vector fields of the SDE being approximated. We place particular emphasis on the recent work of Davie [33, 34] as we extend his arguments in Chapters 5 and 6 to construct our own pathwise scheme using rough path theory. Importantly, by using rough path theory our scheme does not require any Hörmander or ellipticity assumptions.

4.1. Coupling and Wasserstein metrics

Before introducing the notion of pathwise approximation, we need to define the Wasserstein metric from optimal transport theory. We present a brief primer in this section. For more details we refer to Villani's surveys in [114, 115].

Since we only work with probability measures on Euclidean space throughout the thesis, we restrict our study of the Wasserstein metric to the space of probability measures on $\mathbb{R}^d$, denoted by $\mathcal{P}(\mathbb{R}^d)$. The theory can be extended to general Polish spaces.

Given $p \in [1, \infty)$ let $\mathcal{P}_p(\mathbb{R}^d) \subset \mathcal{P}(\mathbb{R}^d)$ denote the subspace of measures with finite p th moment:

$$\mathcal{P}_p(\mathbb{R}^d) := \left\{ \mu \in \mathcal{P}(\mathbb{R}^d) : \int_{\mathbb{R}^d} |x|^p \mu(dx) < \infty \right\}.$$

Definition 4.1.1. Fix $p \in [1, \infty)$ and let ρ be a metric on $\mathbb{R}^d$. Given $\mu, \nu \in \mathcal{P}_p(\mathbb{R}^d)$, define the set of transport plans of μ to ν by

$$\mathcal{M}(\mu, \nu) := \left\{ \Phi : \mathbb{R}^d \rightarrow \mathbb{R}^d \mid \Phi \text{ measurable and } \Phi_*(\mu) = \nu \right\}.$$

Here, $\Phi_*(\mu)$ denotes the pushforward measure of μ under Φ :

$$\Phi_*(\mu)(A) = \mu(\Phi^{-1}(A)) \text{ for all Borel sets } A \subset \mathbb{R}^d.$$

The p th Wasserstein distance on $\mathcal{P}_p(\mathbb{R}^d)$ (using ρ) is defined as

$$(4.1) \quad \mathcal{W}_p(\mu, \nu) := \left(\inf_{\Phi \in \mathcal{M}(\mu, \nu)} \int_{\mathbb{R}^d} \rho(x, \Phi(x))^p \mu(dx) \right)^{1/p}.$$

Certainly we have $\mathcal{W}_q(\mu, \nu) \leq \mathcal{W}_p(\mu, \nu)$ for all $1 \leq q \leq p < \infty$. This is analogous to the relation $L^q \subseteq L^p$ for Lebesgue L^p -spaces.

It can be shown that $\mathcal{W}_p$ is a genuine metric on $\mathcal{P}_p(\mathbb{R}^d)$ ([114, Theorem 7.3]). Moreover, the infimum (4.1) is actually a minimum under mild regularity conditions on the measures μ, ν (for instance, a sufficient condition is that one of the measures is absolutely continuous with respect to Lebesgue measure). The minimizing transport plan $\hat{\Phi}$ is unique but it is difficult to explicitly find $\hat{\Phi}$ because $\mathcal{M}(\mu, \nu)$ possesses no convexity or linear structure in general. However, in the case of $d = 1$ classical optimal transport theory reduces the Wasserstein

distance to an elegant formulation. While we will not use this result in the thesis, its inclusion may be useful for general orientation.

Proposition 4.1.2. [99, Theorem 3.1.2] *Let F and G be distribution functions on $\mathbb{R}$ corresponding to probability measures μ and ν respectively. Suppose that F is continuous with density f and use G^{-1} to denote the generalised inverse of G . Then for all $p \in [1, \infty)$,*

$$\mathcal{W}_p(\mu, \nu) = \left(\int_{\mathbb{R}} \rho(G^{-1} \circ F(x), x)^p f(x) dx \right)^{1/p}.$$

In particular, for each p the minimizing transport plan $\hat{\Phi} \in \mathcal{M}(\mu, \nu)$ is given by $\hat{\Phi} := G^{-1} \circ F$.

An equivalent definition of $\mathcal{W}_p$ is as follows.

Proposition 4.1.3. [115, Theorem 9.4] *Fix $p \in [1, \infty)$ and let $\mu, \nu \in \mathcal{P}_p(\mathbb{R}^d)$ with one of the measures being absolutely continuous with respect to Lebesgue measure. Then*

$$\mathcal{W}_p(\mu, \nu) = \inf \{ |\rho(X, Y)|_{L^p} : \mathcal{L}(X) = \mu, \mathcal{L}(Y) = \nu \}.$$

Specifically, the infimum is taken over all distributions of the $\mathbb{R}^d$ -valued random variables X and Y with marginal distributions μ and ν respectively.

This proposition inspires the following definition.

Definition 4.1.4 (Coupling). *A pair of random variables (X, Y) with the correct marginals μ, ν is known as a coupling of the two probability measures.*

Thus the Wasserstein distance is given by the L^p -distance (equipped with the metric ρ) between the optimal coupling. That is, if $\mathcal{L}(X) = \mu, \mathcal{L}(Y) = \nu$ then certainly $\mathcal{W}_p(\mu, \nu) \leq |\rho(X, Y)|_{L^p}$.

Remark 4.1.5. The topology induced by $\mathcal{W}_p$ on $\mathcal{P}_p(\mathbb{R}^d)$ is slightly stronger than the weak topology; namely, convergence of a sequence $\{\mu_n\}_{n=1}^\infty \subset \mathcal{P}_p(\mathbb{R}^d)$ to a measure $\mu \in \mathcal{P}_p(\mathbb{R}^d)$ in $\mathcal{W}_p$ is equivalent to weak convergence plus a uniform bound on the p th moments of the measures $\{\mu_n\}_{n=1}^\infty$ ([115, Theorem 6.9]). In symbols,

$$\mathcal{W}_p(\mu_n, \mu) \rightarrow 0 \iff \mu_n \rightarrow \mu \text{ weakly and } \sup_n \int_{\mathbb{R}^d} |x|^p \mu_n(dx) < \infty.$$

Remark 4.1.6. The metric $\mathcal{W}_p$ originates from the Monge-Kantorovich mass transportation problem, first introduced by Monge in 1781 [91], then rediscovered many times in various forms since by Kantorovich [69], Lévy among others. The first modern definition was given by the algebraic K -theorist Vaserstein in his sole paper in probability theory [113]. Vaserstein, (anglicized as *Wasserstein* from the Cyrillic alphabet), used the letter $\mathcal{K}$ for the Wasserstein metric in honour of Kantorovich's original contribution of [69]. (Amusingly, by total coincidence Kantorovich's work conversely used the letter $\mathcal{W}$). Throughout this chapter we have used $\mathcal{W}$ and *Wasserstein*, in agreement with the modern literature. For more historical details we refer to [88, 115] and [34, §12].

We conclude the section with a useful lemma [114, Proposition 7.10]. Given two measures μ, ν on $\mathbb{R}^d$ with respective densities f, g with respect to Lebesgue measure, we use the notation: $|\mu - \nu|(x) := |f(x) - g(x)|$.

Lemma 4.1.7. *Let μ and ν be two probability measures on $\mathbb{R}^d$. For $p \in [1, \infty)$,*

$$\mathcal{W}_p(\mu, \nu) \leq 2^{1-1/p} \left\{ \int_{\mathbb{R}^d} |x|^p d|\mu - \nu|(dx) \right\}^{1/p}.$$

As mentioned in [34], this is quite a good bound for $p = 1$ but less good for larger p . We will use this lemma extensively throughout Chapter 6 in order to establish $\mathcal{W}_p$ -estimates.

4.2. Pathwise approximation of stochastic differential equations

We consider the Stratonovich SDE

$$(4.2) \quad \begin{aligned} dx_t &= V(x_t) \circ dW(t) + V_0(x_t) dt \\ &= \sum_{k=1}^d V_k(x_t) \circ dW_k(t) + V_0(x_t) dt, \quad x_0 \in \mathbb{R}^q, \quad t \in [0, 1]. \end{aligned}$$

Throughout this chapter we always assume that the vector fields satisfy the standard non-explosive conditions stated in Remark 2.8.5 so that a solution to (4.2) is well-defined. Given $N \in \mathbb{N}$, we split the unit interval $[0, 1]$ into N pieces $[jh, (j+1)h]$ where $h := N^{-1}$. We will always write x_{jh} instead of the cumbersome $x_{\frac{j}{N}}$.

Definition 4.2.1 (Strong approximation). *An N -step approximation scheme of (4.2) is an algorithm whose input is the initial point $x_0 \in \mathbb{R}^q$ and whose output is a sequence of points $\hat{x} = \{x_j\}_{j=1}^N \subset \mathbb{R}^q$ that approximates the solution of (4.2) (started at x_0) at the times $t \in \{jh : j = 1, \dots, N\}$. We say $\hat{x}$ is a (L^p) -strong approximation scheme of order $\theta > 0$ if*

$$(4.3) \quad \left| \max_{j=1, \dots, N} |x_{jh} - x_j| \right|_{L^p} = O(N^{-\theta}).$$

Example 4.2.2. Under standard assumptions (for example, the conditions of Remark 2.8.5 are sufficient), the standard N -step Euler and Milstein schemes have strong errors of $O(N^{-1/2})$ and $O(N^{-1})$ respectively. Let us introduce notation for the Brownian and Lévy area increments as follows: given $N \in \mathbb{N}$ (and $h = N^{-1}$), we set

$$(4.4) \quad \begin{aligned} W_k^{(j)} &:= W_k(jh, (j+1)h) := W_k((j+1)h) - W_k(jh); \\ A_{kl}^{(j)} &:= \frac{1}{2} \int_{jh}^{(j+1)h} \{W_k(jh, t) dW_l(t) - W_l(jh, t) dW_k(t)\} = A_{kl}(jh, (j+1)h) \end{aligned}$$

for $j \in \{0, 1, \dots, N-1\}$. The Euler scheme $\{x^{(j)}\}_{j=0}^N$ is given by

$$(4.5) \quad x^{(0)} := x_0 \text{ and } x^{(j+1)} := x^{(j)} + \frac{1}{N} V_0(x^{(j)}) + \sum_{k=1}^d W_k^{(j)} V_k(x^{(j)}).$$

Similarly, recalling the differentiable vector field notation introduced in Section 2.10, the Milstein scheme $\{y^{(j)}\}_{j=0}^N$ is defined by $y^{(0)} := x_0$ and

$$(4.6) \quad \begin{aligned} y^{(j+1)} := & y^{(j)} + \frac{1}{N} V_0(y^{(j)}) + \sum_{k=1}^d W_k^{(j)} V_k(x^{(j)}) + \sum_{1 \leq k < l \leq d} A_{kl}^{(j)} [V_k, V_l](y^{(j)}) \\ & + \frac{1}{2} \sum_{1 \leq k, l \leq d} \left(W_k^{(j)} W_l^{(j)} - \frac{1}{N} \delta_{kl} \right) V_k V_l(y^{(j)}). \end{aligned}$$

See [89, 90] for the original proofs of Maruyama and Milstein.

By the famous paper [17] of Cameron and Clark, to construct strong L^2 -approximations of (4.2) of order $\theta > \frac{1}{2}$ requires the ability to generate iterated integrals of Brownian motion. In fact Dickinson later proved that this restriction holds for every N -dimensional linear functional of the Wiener process [37, Proposition 17]. This latter work implies that wavelets and Fourier transforms (which are themselves linear methods) cannot be used to escape the speed limit of $O(N^{-1/2})$.

In the case of $d = 1$ it is a trivial exercise to generate iterated Brownian integrals but for $d \geq 2$, this task becomes a hard problem. In the case of $d = 2$ efficient algorithms based on Fourier expansions are available for generating double integrals (that is, Lévy area increments $A^{(j)}$) but they involve a significant computational cost [53, 106, 117]. The general case of $d > 2$ is still out of reach. The different coordinates of Lévy area are uncorrelated but not independent (as we will prove in Lemma 4.5.1), which makes joint simulation extremely difficult.

With this obstacle as motivation, the papers [1, 2, 3, 26, 31, 32, 33, 34, 36, 55, 56, 68] and [99, §II.9], among many more, have studied N -step SDE approximation schemes which do not require Lévy area increments, but achieve an order of convergence greater than $O(N^{-1/2})$. Instead of measuring the error in the strong sense (using L^p -norms), these papers use the Wasserstein metric introduced in the previous section. In particular, one constructs a probabilistic coupling between the SDE solution x and a scheme $\{x_j\}_{j=1}^N$ such that the error is measured in $\mathcal{W}_p$ (using some particular metric ρ). It is important to remember that since we are producing an approximation which is not necessarily measurable with respect to the same filtration as the solution, the rate of convergence is immune from the limit of $O(N^{-1/2})$ imposed by the results of Cameron, Clark and Dickinson.

We now formally define the notion of pathwise approximations after defining some additional notation. For a given N -step approximation scheme $\hat{x} = \{x_j\}_{j=1}^N \subset \mathbb{R}^q$ for (4.2), we can think of $\hat{x}$ as taking values in $\mathbb{R}^{qN}$. We equip the latter Euclidean space with the metric

$$(4.7) \quad \rho(x, y) := \max_{j=1, \dots, N} |x^j - y^j|,$$

where $x = (x^1, \dots, x^N)$, $x^j \in \mathbb{R}^q$, with identical notation for y . This particular metric is used because (4.3) can be written succinctly as $|\rho(\{x_{jh}\}_{j=1}^N, \hat{x})|_{L^p} = O(N^{-\theta})$. Moreover, it possesses the nice property that the size $\rho(z, 0)$ of the vector $z = (c, \dots, c) \in \mathbb{R}^{qN}$, where $c \in \mathbb{R}$, does not grow as $N \rightarrow \infty$.

Definition 4.2.3 (Pathwise approximation). *Fix $p \in [1, \infty)$, $N \in \mathbb{N}$ and $x_0 \in \mathbb{R}^q$. Let μ_N denote the law of $\{x_{jh}\}_{j=1}^N$ on $\mathbb{R}^{qN}$, where x is the solution of (4.2) started at x_0 . We say that the N -step scheme $\hat{x} = \{x_j\}_{j=1}^N$ is a (L^p) -pathwise*

approximation scheme of order $\theta > 0$ if its law ν_N satisfies:

$$\inf_{\Phi \in \mathcal{M}(\mu_N, \nu_N)} \int_{\mathbb{R}^{qN}} \max_{j=1, \dots, N} |x_j - \Phi(x)_j|^p \mu_N(dx) = O(N^{-\theta p}).$$

Equipping $\mathcal{W}_p$ with the metric (4.7), we rewrite this compactly as:

$$\mathcal{W}_p(\mu_N, \nu_N) = O(N^{-\theta}).$$

Remark 4.2.4. Unfortunately other parts of SDE and rough path theory use the term pathwise approximation in the sense of an approximation holding almost surely (for example, [92]). In contrast, we follow the precedence set by the papers [1, 2, 33, 34] by using the term pathwise in the Wasserstein metric sense.

As discussed in [34, §12], one could argue that a pathwise approximation scheme (with an error measured by $\mathcal{W}_p$) is in fact equivalent to a standard strong approximation. Indeed, suppose we have a scheme $\hat{x} = \{x_j\}_{j=1}^N$ such that its law μ_N satisfies $\mathcal{W}_p(\mu_N, \nu_N) = O(N^{-\theta})$ for some $p \in [1, \infty)$. The Wasserstein bound implies that on the same probability space as the SDE solution x , there exists a $\mathbb{R}^{qN}$ -valued random vector $\tilde{x} = \{\tilde{x}_j\}_{j=1}^N$ with law ν_N such that

$$(4.8) \quad \left| \max_{j=1, \dots, N} |x_{jh} - \tilde{x}_j| \right|_{L^p} = O(N^{-\theta}).$$

When we implement the scheme and generate the points $\hat{x}$, we can regard the process as generating $\tilde{x}$, which approximates the solution in a strong sense. However, the important point is that we cannot describe the joint distribution of $(\tilde{x}, x)$. They are highly non-independent in order for (4.8) to hold for non-zero θ . By simulating an individual solution trajectory in a Wasserstein sense $\mathcal{W}_p$, we have produced a random variable $\tilde{x}$ which is close to an actual realization of a solution of the SDE (4.2). This is why we call such schemes pathwise approximations.

4.3. Previous pathwise research

In this section we review previous research into the pathwise approximation of SDEs. As before, given $N \in \mathbb{N}$, the law of the SDE solution $\{x_{jh}\}_{j=1}^N$ on $\mathbb{R}^{qN}$ is denoted by μ_N . We begin with a theorem of Kanagawa [68] which to the author's knowledge represents the first contribution to the area.

Theorem 4.3.1 (Kanagawa 1989). *Fix $N \in \mathbb{N}$ and let $\{U^{(j)}\}_{j=0}^{N-1} \subset \mathbb{R}^d$ denote a sequence of i.i.d. random vectors, each with mean zero and covariance matrix $N^{-1}I_d$ such that $\mathcal{L}(U^{(j)}) \in \mathcal{P}_{2+\delta}(\mathbb{R}^d)$ for some $\delta \in (0, 1]$. Let $\hat{x}_1^N$ be the terminal point of the N -step Euler scheme (4.5) to the SDE (4.2) with each Brownian increment $W^{(j)}$ replaced with $U^{(j)}$. Then for each $p \in [1, 2 + \delta)$ there exists a constant $C_{\delta,p} > 0$ such that*

$$\mathcal{W}_p(\mathcal{L}(x_1), \mathcal{L}(\hat{x}_1^N)) \leq C_{\delta,p} \frac{(\log N)^\varepsilon}{N^{\frac{1}{2} - \frac{1}{\delta+2}}}$$

for all $\varepsilon > \frac{1}{2}$.

In the case of $\delta = 1$, Kanagawa's scheme achieves a Wasserstein error of $O(N^{-\frac{1}{6}+\varepsilon})$ for arbitrary $\varepsilon > 0$. Note this scheme is only approximating the terminal point of the SDE rather than the entire trajectory.

Alfonsi, Jourdain and Kohatsu-Higa. The authors' first paper [2] considered the Euler scheme in $d = q = 1$ and established a Wasserstein bound under an ellipticity assumption. As in Definition 4.2.3, given $N \in \mathbb{N}$, μ_N denotes the law of $\{x_{jh}\}_{j=1}^N$ on $\mathbb{R}^{qN}$.

Theorem 4.3.2 (AJK-H 2014). [2, Theorem 3.2] *Let ν_N denote the law of the N -step Euler scheme for the SDE (4.2) for the case of $d = q = 1$. Suppose $V_0 \in C^3(\mathbb{R}, \mathbb{R})$ and $V_1 \in C_b^4(\mathbb{R}, \mathbb{R})$ is uniformly elliptic:*

$$\text{there exists } c > 0 \text{ such that for all } x \in \mathbb{R} : V_1(x)^2 \geq c.$$

Then for all $p \in [1, \infty)$ there exists a constant $C_p > 0$ such that

$$\mathcal{W}_p(\mu_N, \nu_N) \leq \frac{C_p}{N^{2/3-\varepsilon}}$$

for all $\varepsilon > 0$.

Their proof relies upon the Lamperti transform [72] which cannot be extended to the case of $d > 1$ unless the vector field collection commutes. In fact, their theorem cannot be generalised to non-elliptic SDEs in the case of $(d, q) = (1, 2)$. We give the details of the counterexample sketched in [34, §11].

Example 4.3.3. Consider the following system describing Brownian motion on the unit disc:

$$dx_1 = x_2 dW, \quad dx_2 = -x_1 dW, \quad x(0) = (1, 0)^t.$$

In this example the vector field is about as regular as possible without being non-trivial; certainly $V(x) = (x_2, -x_1)^t$ is linear and satisfies the non-explosive conditions of Remark 2.8.5. Setting $S_t := x_1(t)^2 + x_2(t)^2$, we find that S satisfies the deterministic differential equation $dS_t = S_t dt$ with $S_0 = 1$, and so $S_t = e^t$. The N -step Euler scheme is given by $x_1^{(j+1)} = x_1^{(j)} + x_2^{(j)} W^{(j)}$ and $x_2^{(j+1)} = x_2^{(j)} - x_1^{(j)} W^{(j)}$ where $W^{(j)} \sim N(0, N^{-1})$. Writing $S^{(j)} := (x_1^{(j)})^2 + (x_2^{(j)})^2$, it follows that $S^{(j+1)} = S^{(j)}(1 + (W^{(j)})^2)$, and so

$$S^{(N)} = \prod_{j=0}^{N-1} (1 + (W^{(j)})^2).$$

We claim that there exists a universal constant $C_1 > 0$ such that

$$(4.9) \quad \|\sqrt{e} - x^{(N)}\|_{L^2} \geq \frac{C_1}{\sqrt{N}}.$$

Indeed, using the inequality $|e^a - e^b| = \left| \int_b^a e^x dx \right| \geq |a - b|$ for $a, b \geq 0$, and the fact that $|\log(1+x) - x| = O(x^2)$ for $|x| \leq \frac{1}{2}$, we have

$$\begin{aligned} \mathbb{E}(|e - S^{(N)}|) &\geq \mathbb{E}(|1 - \log(S^{(N)})|) = \left| 1 - \sum_{j=1}^N \log\left(1 + \frac{1}{N} Z_j^2\right) \right|_{L^1} \\ &\geq \left| 1 - \frac{1}{N} \sum_{j=1}^N Z_j^2 \right|_{L^1} - \frac{C_2}{N}, \end{aligned}$$

where the $\{Z_j\}_{j=1}^N$ are independent $N(0, 1)$ random variables. Markov's inequality and the central limit theorem guarantee that

$$\begin{aligned} \sqrt{N} \mathbb{E}(|e - S^{(N)}|) &\geq \sqrt{N} \left| 1 - \frac{1}{N} \sum_{j=1}^N Z_j^2 \right|_{L^1} + O(N^{-1/2}) \\ &\geq \mathbb{P}\left(\sqrt{N} \left(\frac{1}{N} \sum_{j=1}^N Z_j^2 - 1 \right) \geq 1\right) + O(N^{-1/2}) \rightarrow \Phi(1) > 0, \end{aligned}$$

and (4.9) follows. The Cauchy-Schwarz inequality then implies that

$$\begin{aligned} \frac{C_1}{\sqrt{N}} &\leq |e - S^{(N)}|_{L^1} = \left| e - |x^{(N)}|^2 \right|_{L^1} = |(\sqrt{e} + |x^{(N)}|)(\sqrt{e} - |x^{(N)}|)|_{L^1} \\ (4.10) \qquad \qquad \qquad &\leq \|\sqrt{e} - |x^{(N)}|\|_{L^2} \cdot \|\sqrt{e} + |x^{(N)}|\|_{L^2} \end{aligned}$$

for all $p \in [2, \infty)$. Next, note that as $N \rightarrow \infty$,

$$\mathbb{E}(S^{(N)}) = \mathbb{E}\left(\prod_{j=0}^{N-1} (1 + W^{(j)})^2\right) = \prod_{j=0}^{N-1} \mathbb{E}((1 + W^{(j)})^2) = \left(1 + \frac{1}{N}\right)^N \rightarrow e,$$

and so there exists some constant $C_3 > 0$ (independent of n) such that

$$(4.11) \qquad \|\sqrt{e} + |x^{(N)}|\|_{L^2} \leq \sqrt{e} + \|x^{(N)}\|_{L^2} = \sqrt{e} + \sqrt{\mathbb{E}(S^{(N)})} \leq C_3.$$

Combining (4.10) and (4.11) yields (4.9). As $|x(1)|^2 = S_1 = e$ holds independently of the driving Brownian motion, (4.9) shows that the error of the Euler approximation cannot be less than $O(N^{-1/2})$ no matter what coupling is used.

Specifically, there exists a constant $c > 0$ such that for all $p \in [2, \infty)$,

$$\frac{c}{\sqrt{N}} \leq \mathcal{W}_2(\mathcal{L}(x(1)), \mathcal{L}(x^{(N)})) \leq \mathcal{W}_2(\mu_N, \nu_N) \leq \mathcal{W}_p(\mu_N, \nu_N).$$

On the other hand, since $d = 1$, Lévy area is absent and hence the Milstein and (level 2) log-ODE schemes both give strong convergence errors of order $O(N^{-1})$. Moreover, recalling Remark 2.10.5, the lack of Lévy area means the level 1 and 2 log-ODE methods coincide.

In their follow-up paper [1] Alfonsi et al were able to extend their results to the elliptic multidimensional case by using Malliavin calculus, albeit in a weaker form of the Wasserstein metric [1, Theorem 2.1].

Theorem 4.3.4 (AJK-H 2015). *Consider the SDE (4.2) for the case of $d = q \geq 1$ and let ν_N denote the law of the N -step Euler scheme on $\mathbb{R}^{qN}$. Assume $\{V_k\}_{k=0}^d \subset C_b^2(\mathbb{R}^q, \mathbb{R}^q)$ and the collection $V = \{V_k\}_{k=1}^d$ is uniformly elliptic: there exists $c > 0$ such that:*

$$\text{for all } x \in \mathbb{R}^q : \sum_{k=1}^d V_k(x) V_k(x)^t - cI_q \text{ is positive semidefinite.}$$

Then for all $p \in [1, \infty)$ there exists a constant $C_p > 0$ such that

$$\begin{aligned} \widehat{\mathcal{W}}_p(\mu_N, \nu_N) &:= \inf_{\Psi \in \mathcal{M}(\mu_N, \nu_N)} \max_{j=1, \dots, N} \left(\int_{\mathbb{R}^{qN}} |x_j - \Psi(x_j)|^p d\mu_N(x) \right)^{1/p} \\ &\leq C_p \frac{\sqrt{\log N}}{N}. \end{aligned}$$

Certainly $\widehat{\mathcal{W}}_p(\mu_N, \nu_N) \leq \mathcal{W}_p(\mu_N, \nu_N)$, which is analogous to the inequality: $\max_i \mathbb{E}(|X_i - Y_i|^p) \leq \mathbb{E}(\max_i |X_i - Y_i|^p)$.

Cruzeiro, Malliavin, Thalmaier. To describe the modified Milstein scheme of Cruzeiro et al in [32], we first need to set up some extra notation. They also consider the case of $d = q \geq 1$ and assume the vectors fields $\{V_k\}_{k=0}^d$ are smooth with the subset $\{V_k\}_{k=1}^d$ satisfying the ellipticity condition:

$$(4.12) \quad \text{for all } x \in \mathbb{R}^d \text{ the vectors } \{V_1(x), \dots, V_d(x)\} \text{ form a basis of } \mathbb{R}^d.$$

Under this assumption, for a given vector field $F : \mathbb{R}^q \rightarrow \mathbb{R}^q$ we use the notation

$$F(x) = \sum_{k=1}^d \langle F, V_k \rangle_x V_k(x)$$

to denote the change of basis. The N -step scheme $z = \{z_j\}_{j=1}^N$ is defined iteratively by $z_0 := x_0 \in \mathbb{R}^d$, then

$$\begin{aligned} z_{j+1} := & z_j + \frac{1}{N} V_0(z_j) + \sum_{k=1}^d W_k^{(j)} V_k(z_j) \\ & + \frac{1}{2} \sum_{1 \leq k, l, m \leq d} \left(W_k^{(j)} W_l^{(j)} - \frac{1}{N} \delta_{kl} \right) \left(V_k V_l(z_j) + V_m(z_j) \langle [V_l, V_m], V_k \rangle_{z_j} \right) \end{aligned}$$

for $j \in \{0, 1, \dots, N-1\}$. We can see that the Lévy area term in the Milstein scheme (4.6) has disappeared.

Theorem 4.3.5 (CMT 2004). [32, Theorem 2.1] *Suppose the smooth vector fields $\{V_k\}_{k=0}^d$ along with their first three derivatives are bounded and assume the ellipticity condition of (4.12) holds. Let λ_N be the law of the scheme $z = \{z_j\}_{j=1}^N$ on $\mathbb{R}^{dN}$. Then there exists a constant $C > 0$ such that*

$$\mathcal{W}_2(\mu_N, \lambda_N) \leq \frac{C}{N}.$$

The idea of the proof is to start with the usual Milstein scheme but perform a change of basis of $\mathbb{R}^d$ in terms of the vectors $\{V_k(z_j)\}_{1 \leq k \leq d}$ at each step. In turn, this induces an orthogonal rotation of Brownian motion (whose law is rotationally invariant) and this action eliminates the Lévy area term. If the condition (4.12) fails, the change of basis is not possible and so the proof cannot be generalised to the degenerate vector field case.

Davie. In [34] Davie presents a similar variant of the Milstein scheme by perturbing the increments by a Gaussian term so as to replace the Lévy area increments. By carefully matching moments and using polynomial perturbations results (similar to those contained in Chapter 6), he is able to construct a

pathwise scheme of order $O(N^{-1})$ in the metric $\mathcal{W}_p$ which does not require Lévy area (see [34, Theorem 4]). However, his main theorem requires the assumption that the vector field collection $\{V_k\}_{k=1}^d$ satisfies the non-degeneracy condition:

$$(4.13) \quad \text{for all } x \in \mathbb{R}^q \text{ the collection } \{V_k(x), [V_k, V_l](x)\}_{1 \leq k \leq l \leq d} \text{ spans } \mathbb{R}^q.$$

This is a very strong condition. It excludes the SDE which defined Lévy area (see (5.3)) as well as every SDE in the case of $d = q = 2$ whose coordinates are not independent. It also does not cover the SDE defining circular Brownian motion (Example 4.3.3). As noted in [31, Remark 2], in this case $(d, q) = (2, 1)$ and the rank of the associated Lie bracket is 1. The coupling construction in this paper is interesting because it is local; that is, the couplings are between increments for a single step, so that the pathwise scheme $\{\tilde{x}_j\}_{j=1}^N$ and the original solution $\{x_{jh}\}_{j=1}^N$ are adapted to the same filtration. This makes the proof of the global error of $O(N^{-1})$ immediate from local errors of $O(N^{-3/2})$ by using [70, Theorem 10.6.3].

Remark 4.3.6. For completeness, we mention that [34] concludes with an argument showing the perturbed Milstein scheme has an error of order $O(N^{-1})$ in the metric $\widehat{\mathcal{W}}_2$ without placing any assumptions on the vector fields. The proof follows the previous work of Kloeden, Platen and Wright in [71] by employing a truncation of the Fourier series expansion of Lévy area.

Before presenting the second of Davie's approximation schemes in the subsequent paper [33], we use the next section to introduce the Komlós-Major-Tusnády (KMT) theorem.

4.4. Komlós-Major-Tusnády theorem

The classical KMT result is a form of the simultaneous central limit theorem using couplings.

Theorem 4.4.1 (Komlós-Major-Tusnády 1975). *Suppose μ is a probability measure on $\mathbb{R}$ with mean zero, variance 1 and zero third moment. Assume that μ has finite exponential moments: $\mathbb{E}(e^{\alpha|X|}) < \infty$ for some $\alpha > 0$, where $\mathcal{L}(X) = \mu$. Then there exists positive constants C, K and λ (dependent on α) such that the following holds: for every $N \in \mathbb{N}$, it is possible to construct a sequence of i.i.d. random variables $\{X_i\}_{i=1}^N$ with law μ and a corresponding sequence of i.i.d. $N(0, 1)$ variables $\{Y_i\}_{i=1}^N$ such that for all $x \geq 0$,*

$$\mathbb{P}(\Delta_N \geq C \log N + x) \leq K e^{-\lambda x} \text{ where } \Delta_N := \max_{1 \leq n \leq N} \left| \sum_{i=1}^n (X_i - Y_i) \right|.$$

The exponential tail decay implies that for all $p \in [1, \infty)$, there exists a constant $C_p > 0$ such that $|\Delta_N|_{L^p} \leq C_p \log N$.

The original KMT theorem was later extended to vector random variables by Einmahl [39] and then Zaitsev established the result for the case of non-identical distributions which are uniformly non-degenerate in a series of papers [121]. For approximating certain non-Gaussian distributions, Zaitsev's extension obtains the best possible rate of convergence [122, 123].

4.5. Pathwise approximation work of Davie

Following Zaitsev's work, Davie proves his own variation of the KMT theorem which allows the underlying distributions to be random themselves. In Chapter 6 we prove our own extension of Davie's coupling argument so that coupling holds in $\mathcal{W}_p$ for $p \in [1, \infty)$ rather than just $p = 2$.

In [33] it is assumed for simplicity that there is no drift in the SDE (5.1) and so for this section only we also assume $V_0 = 0$. As discussed earlier, the problem with the Milstein scheme is that it requires the ability to generate the Lévy area increments $A^{(j)}$ defined in (4.4). To combat this problem, Davie substitutes each $A^{(j)}$ with a suitable quadratic polynomial of Gaussian random variables denoted by $B^{(j)}$. The construction is such that $A^{(j)}$ and $B^{(j)}$ share the same mean and covariance structure along with the same underlying Brownian increment $W^{(j)}$. To introduce $B^{(j)}$ we must first closely examine the original Brownian Lévy area increments. The following lemma gives a simple decomposition of $A^{(j)}$ into parts dependent and independent of the corresponding Brownian increment $W^{(j)} := W(jh, (j+1)h)$.

Lemma 4.5.1. *For all $1 \leq k < l \leq d$:*

$$(4.14) \quad A_{kl}^{(j)} = \zeta_k^{(j)} W_l^{(j)} - \zeta_l^{(j)} W_k^{(j)} + K_{kl}^{(j)},$$

where the $\zeta_k^{(j)}$: $k = 1, \dots, d$, $K_{kl}^{(j)}$: $1 \leq k < l \leq d$, are mutually uncorrelated (but not independent), independent of $W^{(j)}$ and have mean zero. Moreover,

$$\text{Var}(\zeta_k^{(j)}) = \frac{1}{12N} \quad \text{and} \quad \text{Var}(K_{kl}^{(j)}) = \frac{1}{12N^2}.$$

PROOF. We suppose $j = 0$ for simplicity and begin by decomposing the $A^{(0)}$ increment into parts dependent and independent of the Brownian increment $W(h)$. To this end, following [34, §7] we can write

$$W_k(t) = h^{1/2} B_k(t/h) + th^{-1/2} Z_k \quad t \in [0, h],$$

where $B_1, \dots, B_d$ are independent standard Brownian bridges on $[0, 1]$ and $Z_k = h^{-1/2} W_k(h)$ are independent $N(0, 1)$ random variables (which are independent

of the B_j). Also write $B_0(t) = t$ and set

$$K_\alpha := \int_0^1 \int_0^{t_l} \dots \int_0^{t_2} dB_{j_1}(t_1) \dots dB_{j_l}(t_l)$$

for an index $\alpha = (j_1, \dots, j_l) \in \{0, 1, \dots, d\}^l$. For such an index it can be shown that (see [34, §7])

$$I_\alpha := \int_0^h \int_0^{t_l} \dots \int_0^{t_2} dW_{j_1}(t_1) \dots dW_{j_l}(t_l) = h^{(l(\alpha) + n(\alpha))/2} \sum_{\beta=(i_1, \dots, i_l)} K_\beta \prod_{k: i_k < j_k} Z_{j_k},$$

where the sum is over all $\beta = (i_1, \dots, i_l)$ such that for each $k \in \{1, \dots, l\}$ we have either $i_k = j_k$ or $i_k = 0 < j_k$. Here we have used $l(\alpha)$ and $n(\alpha)$ to denote the length and number of zero entries of α respectively. Noting the antisymmetry $K_{kl} = -K_{lk}$ for $0 \leq k < l$, it follows that

$$A_{12}^{(0)} = \frac{1}{2} (I_{12} - I_{21}) = h (K_{10}Z_2 - K_{20}Z_1 + K_{12}),$$

where

$$K_{12} = \int_0^1 B_1(t) dB_2(t) \text{ and } K_{j0} = \int_0^1 B_j(t) dt \text{ for } j \in \{1, 2\}.$$

Thus $\zeta_j^{(0)} := h^{1/2} K_{j0}^{(0)}$ for $j = 1, 2$, and $K_{12}^{(0)} := h K_{12}$ gives the claimed decomposition. The variances follow from Itô's isometry. For the calculations we refer to Lemma 7 of [34]. $\square$

The fact that for fixed j , the increments $\zeta^{(j)}$ and $K^{(j)}$ are not independent makes them (and consequently $A^{(j)}$) very difficult to simulate numerically. A natural solution would be to approximate these two variables with normal random variables $z^{(j)}$, $\lambda^{(j)}$ with the correct mean and moments, to produce a Gaussian approximation $B^{(j)}$ for $A^{(j)}$. Since uncorrelated Gaussian random variables are necessarily independent, simulation is much easier. This is precisely what Davie proposes in [33]; define the following independent normal

random variables

$$\begin{aligned} W^{(j)} &\sim N\left(0, \frac{1}{N}I_d\right), \quad z^{(j)} \sim N\left(0, \frac{1}{12N}I_d\right), \\ \lambda^{(j)} &= (\lambda_{kl}^{(j)})_{1 \leq k < l \leq d} \sim N\left(0, \frac{1}{12N^2}I_{\frac{d(d-1)}{2}}\right), \end{aligned}$$

for each $j \in \{0, 1, \dots, N-1\}$. Then set $\{B^{(j)}\}_{1 \leq k < l \leq d}$ to be quadratic polynomial

$$(4.15) \quad B_{kl}^{(j)} := z_k^{(j)}W_l^{(j)} - z_l^{(j)}W_k^{(j)} + \lambda_{kl}^{(j)}.$$

Davie's N -step scheme $\tilde{x} = \{\tilde{x}^{(j)}\}_{j=0}^N$ is defined as the Milstein scheme (4.6) but with the $A^{(j)}$ increments replaced with the $B^{(j)}$. To be explicit: $\tilde{x}^{(0)} := x_0$ and

$$(4.16) \quad \begin{aligned} \tilde{x}^{(j+1)} &:= \tilde{x}^{(j)} + \sum_{1 \leq k < l \leq d} B^{(j)}[V_k, V_l](\tilde{x}^{(j)}) \\ &\quad + \frac{1}{2} \sum_{1 \leq k, l \leq d} \left(W_k^{(j)}W_l^{(j)} - \frac{1}{N}\delta_{kl} \right) V_k V_l(\tilde{x}^{(j)}). \end{aligned}$$

Under a non-degeneracy condition on the vector fields $\{V_k\}_{k=1}^d$, [33, Theorem 1] establishes the following coupling error under the weaker metric $\widehat{\mathcal{W}}_2$ of $\mathcal{W}_2$ (as defined in Theorem 4.3.4). To formulate the non-degeneracy condition, for each $x \in \mathbb{R}^d$, we define the linear mapping $L_x : \mathbb{R}^d \oplus [\mathbb{R}^d, \mathbb{R}^d] \rightarrow \mathbb{R}^q$ by

$$L_x(r, s) = \sum_{k=1}^d r_k V_k(x) + \sum_{1 \leq k < l \leq d} s_{kl} [V_k, V_l](x)$$

for $r = (r_k) \in \mathbb{R}^d$ and $s = (s_{kl}) \in [\mathbb{R}^d, \mathbb{R}^d]$. As Davie writes, the assumption on the vector fields can be thought of as a strengthened Hörmander condition with uniformity in x . While still restrictive, in contrast to (4.12) and (4.13), this condition is satisfied by the SDEs defining Lévy area and circular Brownian motion.

Theorem 4.5.2 (Davie 2014). *Suppose the first and second derivatives of $\{V_k\}_{k=1}^d$ are bounded and that there exist constants $\delta > 0$ and $K > 0$ such*

that for each $x \in \mathbb{R}^q$, the image under L_x of the unit ball in $\mathbb{R}^d \oplus [\mathbb{R}^d, \mathbb{R}^d]$ contains the ball of radius $\delta(1 + |x|)^{-K}$ in $\mathbb{R}^q$. Denote the law of (4.16) on $\mathbb{R}^{qN}$ by λ_N . Then there exists a constant $C_{\delta,K} > 0$ such that

$$\widehat{\mathcal{W}}_2(\mu_N, \lambda_N) \leq \frac{C_{\delta,K}}{N}.$$

The proof is an adaptation of Zaitsev's series of papers [121] concerning an extension of the Komlós-Major-Tusnády theorem (Theorem 4.4.1) to random distributions. Indeed, Davie constructs an explicit coupling to establish the Wasserstein error by coupling the scheme (4.16) with the true solution $\{x_{jh}\}_{j=1}^N$ using the same dyadic argument of Zaitsev. However, in contrast to Zaitsev, the coupling is performed conditionally on the Brownian increments $W^{(j)}$ of the driving Brownian motion W of (4.2). This means the distribution of the increments $A^{(j)}$ and $B^{(j)}$ are themselves random and matters are complicated by the fact that the coupling procedure does not guarantee the Brownian increments $\tilde{W}^{(j)} \sim N(0, N^{-1}I_d)$ composing each $B^{(j)}$ are necessarily equal to the $W^{(j)}$ used to define $A^{(j)}$. When we adapt Davie's construction in Chapter 6 we take great pains to ensure that $W^{(j)} = \tilde{W}^{(j)}$ for all j , (the reason for this will be made clear in Section 5.5 when we introduce piecewise abelian rough paths).

Davie recursively constructs the conditional coupling by using Lemma 4.1.7 along with a version of a central limit theorem concerning polynomial perturbations of Gaussian distributions. We do not repeat the latter limit result since we prove an extension of the original [33, Lemma 4] from quartics to higher order polynomials in Lemma 6.3.1. For now we do not discuss the proof further; the entire Chapter 6 can be viewed as a modification of Davie's original coupling construction using the polynomial perturbation result of [35, Proposition 1].

CHAPTER 5

Pathwise approximation by coupling piecewise abelian rough paths

In this chapter we present a new pathwise approximation scheme for the SDE considered in the previous chapter:

$$(5.1) \quad \begin{aligned} dx_t &= V(x_t) \circ dW(t) + V_0(x_t) dt \\ &= \sum_{k=1}^d V_k(x_t) \circ dW_k(t) + V_0(x_t) dt, \quad x_0 \in \mathbb{R}^q, \quad t \in [0, 1]. \end{aligned}$$

Importantly, our scheme does not require the generation of Lévy area but exhibits a Wasserstein convergence rate which is superior to the N -step Euler scheme's strong error rate of $O(N^{-\frac{1}{2}})$. Moreover, by using rough path theory we avoid imposing any non-degenerate Hörmander or ellipticity assumptions on the vector fields of the SDE. This is in contrast to the previously considered research in [1, 2, 32, 33, 34]. Our scheme is based on the log-ODE method of Section 2.10 with the Lévy area increments replaced by Gaussian approximations such that the covariance structure is maintained. The actual Wasserstein coupling construction is postponed to Chapter 6 in which we extend the arguments of Davie in [33, 35]. The results presented here are taken from the preprint [42].

Since we are using rough path theory, throughout the chapter we make the standard assumption that $V_0 \in \text{Lip}^\beta$ and $V = \{V_k\}_{k=1}^d \subset \text{Lip}^\gamma$ for some constants $\beta > 1$ and $\gamma > 2$ so that Theorem 2.8.4 guarantees the existence of

a unique solution to (5.1) almost surely. In addition, we always assume that the vector fields satisfy the standard non-explosive conditions stated in Remark 2.8.5. As in the previous chapter, given $N \in \mathbb{N}$, we reserve the notation μ_N for the probability measure on $\mathbb{R}^{qN}$ given by the law of the solution x to the SDE (5.1) at the points $\{x_{jh}\}_{j=1}^N \subset \mathbb{R}^q$.

5.1. New scheme and main theorem

In this section we introduce our approximation scheme for (5.1). Recalling the definition (4.15) in Section 4.5, we set $\{B^{(j)}\}_{1 \leq k < l \leq d}$ to be the quadratic polynomials:

$$B_{kl}^{(j)} := z_k^{(j)} \tilde{W}_l^{(j)} - z_l^{(j)} \tilde{W}_k^{(j)} + \lambda_{kl}^{(j)},$$

where we have the independent Gaussian random variables:

$$\tilde{W}_k^{(j)} \sim N\left(0, \frac{1}{N}\right), \quad z_k^{(j)} \sim N\left(0, \frac{1}{12N}\right), \quad \lambda_{kl}^{(j)} \sim N\left(0, \frac{1}{12N^2}\right).$$

Definition 5.1.1 (New pathwise approximation scheme). *Given $N \in \mathbb{N}$, the scheme $\{x_j^N\}_{j=0}^N$ is defined iteratively by $x_0^N := x_0$ and*

$$(5.2) \quad x_{j+1}^N := \exp\left(\frac{1}{N}V_0 + \sum_{k=1}^d \tilde{W}_k^{(j)}V_k + \sum_{1 \leq k < l \leq d} B_{kl}^{(j)}[V_k, V_l]\right)(x_j^N)$$

for $j \in \{0, 1, \dots, N-1\}$. Let ν_N denote the law of $\{x_j^N\}_{j=1}^N$ on $\mathbb{R}^{qN}$.

We now state the main result of the chapter.

Theorem 5.1.2. *Fix $p \in [1, \infty)$, $\beta > 1$ and $\gamma > 2$. Suppose $V_0 \in \text{Lip}^\beta(\mathbb{R}^q)$, $V = \{V_k\}_{k=1}^d \subset \text{Lip}^\gamma(\mathbb{R}^q)$ and fix $\lambda \geq \|V_0\|_{\text{Lip}^\beta} \vee \|V\|_{\text{Lip}^\gamma}$. Then there exists a constant $C_{\beta, \gamma, \lambda, p} > 0$ such that*

$$\mathcal{W}_p(\mu_N, \nu_N) \leq \frac{C_{\beta, \gamma, \lambda, p}}{N^{1 - \frac{2}{\gamma} - \varepsilon}}$$

for all $\varepsilon > 0$.

We construct an explicit coupling in order to prove Theorem 5.1.2. Importantly, the coupling of μ_N, ν_N uses the same underlying Brownian motion W to produce the solution x and define the Brownian increments $\{\tilde{W}^{(j)}\}_{j=0}^{N-1}$ needed to define x^N ; that is, $W^{(j)} = \tilde{W}^{(j)}$ for all j .

The actual coupling construction is postponed to the next chapter (Theorem 6.2.1) as its arguments are non-trivial and require extensions of recent results contained in [33, 35] concerning polynomial perturbations of Gaussian distributions and central limit theorems. For now, we discuss Theorem 5.1.2 and dedicate the remaining sections of the chapter to its proof under the assumption that Theorem 6.2.1 holds.

Remark 5.1.3. Theorem 5.1.2 implies that if the vector fields are sufficiently regular (that is, $\gamma > 4$) then the scheme performs better than the $O(N^{-1/2})$ strong convergence rate of the Euler scheme. In particular, if the vector fields have Stein-Lipschitz regularity of arbitrary order then by setting $\gamma > \frac{2}{\varepsilon}$ for $\varepsilon > 0$, we have $\mathcal{W}_p(\mu_N, \nu_N) = O(N^{-1+\varepsilon})$.

One may ask whether the Wasserstein convergence rate of Theorem 5.1.2 is tight. We have the following upper bound on the performance.

Proposition 5.1.4. *Consider the SDE defining Lévy area in $d = 2$:*

$$(5.3) \quad dx_1 = dW_1, \quad dx_2 = dW_2, \quad dx_3 = \frac{1}{2}(x_1 dW_2 - x_2 dW_1), \quad x(0) = 0 \in \mathbb{R}^3.$$

Let μ_N, ν_N be the corresponding measures on $\mathbb{R}^{qN} = \mathbb{R}^{3N}$. Then there exists a constant $c > 0$ such that

$$\mathcal{W}_1(\mu_N, \nu_N) \geq c \frac{\log N}{N}.$$

PROOF. For this SDE the schemes of Davie and Definition 5.1.1 coincide (denoted by $\{\tilde{x}_j^N\}_{j=1}^N$ and $\{x_j^N\}_{j=1}^N$ respectively). It was established in [33, Example

1] that there exists a constant $c > 0$ such that for every $N \in \mathbb{N}$ and any coupling between the random variables $\{\tilde{x}_j^N\}_{j=1}^N$ and $\{x(jh)\}_{j=1}^N$, we have

$$\mathbb{P} \left(\max_{j=1, \dots, N} |\tilde{x}_j^N - x(jh)| \geq c \frac{\log N}{N} \right) > \frac{1}{2}.$$

Applying Markov's inequality finishes the proof. $\square$

Note that the vector fields of (5.3) have arbitrary Lipschitz regularity. We conclude that in the case of Lévy area, the coupling used in Theorem 5.1.2 is almost optimal up to a logarithmic factor: for all $p \in [1, \infty)$, there exist constants $0 < c < C_p$ for which

$$c \frac{\log N}{N} \leq \mathcal{W}_1(\mu_N, \nu_N) \leq \mathcal{W}_p(\mu_N, \nu_N) \leq \frac{C_p}{N^{1-\varepsilon}},$$

where $\varepsilon > 0$ is arbitrary.

Remark 5.1.5. We can view Davie's original scheme (4.16) (for the case of $V_0 = 0$) as an $O(N^{-1})$ Taylor approximation to the ODEs within the scheme of Definition 5.1.1. Consequently, if vector fields $V = \{V_k\}_{k=1}^d$ possess Stein-Lipschitz regularity of arbitrarily high order, Theorem 5.1.2 implies that Davie's scheme achieves a Wasserstein convergence rate of $O(N^{-1+\varepsilon})$ for $\varepsilon > 0$. In the final remarks of [33] Davie conjectured that it might be possible to remove the non-degeneracy condition of Theorem 4.5.2, yet have his scheme maintain an error of order $O(N^{-1})$. In a sense, this corollary of Theorem 5.1.2 is a partial answer to this conjecture; we make no assumptions on the ellipticity or non-degeneracy on the vector field collection but achieve a rate of convergence which is arbitrarily close to $O(N^{-1})$. Moreover, we use the Wasserstein metric $\mathcal{W}_p$ rather than the weaker version $\widehat{\mathcal{W}}_p$ defined in Theorem 4.3.4.

5.2. Kantorovich-Rubinstein duality formula

An elegant feature of the particular Wasserstein metric $\mathcal{W}_1$ is its primal representation via functionals ([115, Theorem 5.10 and Remark 5.16]).

Theorem 5.2.1 (Kantorovich-Rubinstein duality formula). *Let X, Y be random variables with respective laws $\mu, \nu \in \mathcal{P}_1(\mathbb{R}^d)$. Then*

$$\mathcal{W}_1(\mu, \nu) = \sup_{\substack{\psi \in C(\mathbb{R}^d, \mathbb{R}) \\ \text{Lip}(\psi) \leq 1}} \mathbb{E}(\psi(X) - \psi(Y)),$$

where

$$\text{Lip}(\psi) := \sup_{x \neq y} \frac{|\psi(x) - \psi(y)|}{|x - y|}$$

denotes the Lipschitz constant of ψ in the classical (not Stein) sense.

This classical theorem of optimal transport theory shows how the metric $\mathcal{W}_1$ can be used to establish weak approximation rates for schemes. Indeed, combining Theorem 5.1.2 with the Kantorovich-Rubinstein duality formula gives the following.

Corollary 5.2.2. *Fix $\beta > 1$ and $\gamma > 2$. Suppose $V_0 \in \text{Lip}^\beta$, $V = \{V_k\}_{k=1}^d \subset \text{Lip}^\gamma$ and fix $\lambda \geq \|V_0\|_{\text{Lip}^\beta} \vee \|V\|_{\text{Lip}^\gamma}$. Then there exists a constant $C_{\beta, \gamma, \lambda} > 0$ such that*

$$\sup_{\substack{\psi \in C(\mathbb{R}^{qN}, \mathbb{R}) \\ \text{Lip}(\psi) \leq 1}} \left| \mathbb{E}(\psi(\{x_{jh}\}_{j=1}^N)) - \mathbb{E}(\psi(x^N)) \right| \leq \frac{C_{\beta, \gamma, \lambda}}{N^{1 - \frac{2}{\gamma} - \varepsilon}}$$

for all $\varepsilon > 0$.

Our scheme does not demand any (algebraic) ellipticity conditions on the vector fields, only sufficient (analytic) Stein-Lipschitz regularity. This is in contrast to the weak approximation schemes considered in [1, 2, 57].

In the context of option pricing, we can interpret Theorem 5.2.1 as a measure of the performance of a given scheme for approximating the expectation of certain exotic Asian options instead of just the terminal value of the path (that is, European options). For example, in the case of $q = 1$ we could consider the discrete Asian option payoff function:

$$\psi(\{x_{jh}\}_{j=1}^N) := \max \left\{ 0, \frac{1}{N} \sum_{j=1}^N x_{jh} - K \right\} \text{ for some strike } K > 0.$$

Then Corollary 5.2.2 implies that our scheme exhibits a weak approximation error of order $O(N^{-1+\varepsilon+2/\gamma})$ for the payoff ψ .

Remark 5.2.3. As an aside, we note that if we wanted to approximate $\mathbb{E}(\varphi(x_1))$ for some $\varphi \in C_b^\infty(\mathbb{R}^q, \mathbb{R})$ then the algorithm of Ninomiya and Victoir [94] does not require Lévy area increments but it possesses an error of $O(N^{-2})$. In common with our scheme, theirs is also a sequential ODE-based scheme which, given $N \in \mathbb{N}$ solves a series of N ODEs to produce a point $\hat{x}_1^N$ satisfying:

$$(5.4) \quad |\mathbb{E}(\varphi(x_1)) - \mathbb{E}(\varphi(\hat{x}_1^N))| = O(N^{-2}).$$

This is at least a whole order better than the best possible rate of our scheme. In fact, the proof of (5.4) views the log-ODE as a random walk, corresponding to our notion of a piecewise abelian rough path (to be introduced in Section 5.4).

Note that in general, the standard Euler and Milstein schemes have a weak approximation error of $O(N^{-1})$ for European payoffs φ . Indeed, by the work of Talay and Tubaro [112, Theorem 1], this is the case for when V_0 and V are non-zero, smooth vector fields with bounded derivatives of all orders and the payoff $\varphi \in C^\infty(\mathbb{R}^q, \mathbb{R})$, together with all its derivatives, has polynomial growth. We stress that our scheme was designed for pathwise approximation in that

its output is meant to approximate an actual realization of the solution path rather than the expectation of a given functional.

5.3. Overview of the proof

Before going into specific details, we use this section to provide a brief sketch of how the components of the proof of Theorem 5.1.2 go together.

In common with the algorithm of [94], our scheme $\{x_j^N\}_{j=1}^N$ is based on the level-2 method of the log-ODE method from rough path theory. Recalling Section 2.10, the latter scheme also consists of solving a sequence of ODEs to produce a set of points $\{x^{(j)}\}_{j=1}^N \subset \mathbb{R}^q$. In particular, $x^{(0)} := x_0$ and

$$(5.5) \quad x^{(j+1)} := \exp \left(\frac{1}{N} V_0 + \sum_{k=1}^d W_k^{(j)} V_k + \sum_{1 \leq k < l \leq d} A_{kl}^{(j)} [V_k, V_l] \right) (x^{(j)})$$

for $j \in \{0, 1, \dots, N-1\}$. This algorithm requires the ability to generate Lévy area increments $A^{(j)} \in [\mathbb{R}^d, \mathbb{R}^d]$. For our new scheme $\{x_j^N\}_{j=1}^N$ we replace these increments with the Gaussian random variables $B^{(j)}$ of (4.15). Recalling Theorem 2.8.4, the theory of rough paths allows us to rewrite the original SDE (5.1) as the solution of the following RDE with drift:

$$dx_t = V(x_t) d\mathbb{W}_t + V_0(x_t) dt,$$

Here $\mathbb{W} \in C([0, 1], G^{(2)}(\mathbb{R}^d))$ is the enhanced Brownian rough path (Definition 2.8.1). It turns out that the schemes $\{x^{(j)}\}_{j=1}^N$ and $\{x_j^N\}_{j=1}^N$ can also be written as solutions of two RDEs with drift terms. To be precise; $x^{(j)} = y_{jh}$ and $x^{(j)} = z_{jh}$ for all j , where $y, z \in C([0, 1], \mathbb{R}^q)$ solve the RDEs:

$$dy_t = V(y_t) d\mathbb{W}_t^h + V_0(y_t) dt, \quad y_0 = x_0,$$

$$dz_t = V(z_t) d\mathbb{X}_t^h + V_0(z_t) dt, \quad z_0 = x_0.$$

Here, $\mathbb{W}^h$ and $\mathbb{X}^h$ are members of a special class of rough paths which we call *piecewise abelian*. The notion of piecewise abelian rough paths can be thought of as the natural non-commutative (that is, group-valued) analogue of piecewise linear approximations of paths with values in the abelian group $G^{(1)} = \mathbb{R}^d$.

Both $\mathbb{W}^h$ and $\mathbb{X}^h$ share the same first level given by the standard N -step piecewise linear approximation W^h of the Brownian motion W (recall $W^{(j)} = \tilde{W}^{(j)}$). Hence $\mathbb{W}^h$ and $\mathbb{X}^h$ can be interpreted as two different 2-rough path lifts of the same underlying path W^h . Their difference at level 2 is given by the continuous interpolations of the two discrete random walks composed of the increments $A^{(j)}$ and $B^{(j)}$ respectively:

$$\pi_2(\mathbb{W}_{0,rh}^h - \mathbb{X}_{0,rh}^h) = \sum_{j=0}^{r-1} (A^{(j)} - B^{(j)}).$$

Theorem 6.2.1 constructs a probabilistic coupling of the two random walks (conditional on the underlying Brownian increments $\{W^{(j)}\}_{j=0}^{N-1}$ of W^h) and this induces a coupling of $\mathbb{W}^h$ and $\mathbb{X}^h$ in the space $G\Omega_p(\mathbb{R}^d)$ of geometric p -rough paths, where $p \in [2, 3)$. Using the Lipschitz-continuity of the Itô map Ξ of rough path theory in the inhomogeneous p -variation metric (Proposition 2.9.2), this action induces a further coupling of the RDE solutions y and z in $C([0, 1], \mathbb{R}^q)$. The situation can be described by the following diagram (Figure 5.1), where dashed arrows represent couplings.

The previous pathwise approximation research studied in the last chapter coupled the SDE solution and the approximation scheme directly (that is, on $\mathbb{R}^{qN}$). On the other hand, we perform the initial coupling at the input side of the Itô map Ξ and so our construction is completely independent of the vector fields of the original SDE (5.1) being approximated. The vector fields are only

$$\begin{array}{ccccccc}
\{A^{(j)}\}_{j=0}^{N-1} & \longrightarrow & \mathbb{W}^h & \hookrightarrow & S_\kappa(\mathbb{W}^h) & \hookrightarrow & y \longrightarrow \{x^{(j)}\}_{j=1}^N \\
\uparrow \text{dashed} & \nwarrow & \uparrow \text{dashed} & & \uparrow \text{dashed} & & \uparrow \text{dashed} \\
& & W^h & & & & \\
\downarrow \text{dashed} & \swarrow & \downarrow \text{dashed} & & \downarrow \text{dashed} & & \downarrow \text{dashed} \\
\{B^{(j)}\}_{j=0}^{N-1} & \longrightarrow & \mathbb{X}^h & \hookrightarrow & S_\kappa(\mathbb{X}^h) & \hookrightarrow & z \longrightarrow \{x_j^N\}_{j=1}^N
\end{array}$$

$$\begin{array}{ccccccc}
[\mathbb{R}^d, \mathbb{R}^d]^{\times N} & \longrightarrow & G\Omega_2(\mathbb{R}^d) & \xrightarrow{S_\kappa(\cdot)} & G\Omega_\kappa(\mathbb{R}^d) & \xrightarrow{\Xi} & C([0, 1], \mathbb{R}^q) \longrightarrow \mathbb{R}^{qN} \\
& \nwarrow & \nearrow & & & & \\
& & (\mathbb{R}^d)^{\times N} & & & &
\end{array}$$

FIGURE 5.1. The induced couplings

relevant once we push the coupling through the Itô map and therefore we do not need to impose any non-degeneracy conditions on the vector field collection $\{V_k\}_{k=1}^d$. The Lipschitz estimates for RDEs are important because they enforce the Wasserstein error at the output to have the same order of convergence as the original coupling of the (lifted) rough paths at the input side of Ξ .

Remark 5.3.1. Other papers exploiting the separation of the input and output of SDE flows provided by the Itô map Ξ of rough path theory include the ε -strong SDE simulation paper [12] by Blanchet, Chen and Dong and the SDE quantization paper of [95] by Pagès and Sellami. By using Ξ instead of the classical Lamperti transform, these papers are able to generalise previous results in $d = 1$ to greater dimensions including non-commutative vector field collections. Another example of a paper exploiting Ξ is [102] by Riedel, where he establishes connections between Gaussian rough paths and transportation-cost inequalities from optimal transport theory.

The remainder of the chapter is devoted to the proof of Theorem 5.1.2, whilst the proof of the initial coupling of the increments $\{A^{(j)}\}$ and $\{B^{(j)}\}$ is, as mentioned, postponed to the next chapter (Theorem 6.2.1).

5.4. Piecewise abelian rough paths

In this section we introduce the notion of piecewise abelian rough paths. Such a class of rough paths allows us to recast the log-ODE technique introduced in Section 2.10 from the perspective of rough path theory; namely, as the solutions of rough differential equations driven by piecewise abelian rough paths. As before, given $N \in \mathbb{N}$, the unit interval $[0, 1]$ is partitioned into intervals $[jh, (j+1)h]$ of equal length $h = N^{-1}$. First we define a piecewise abelian rough path.

Definition 5.4.1. *We call a p -rough path $\mathbb{X} \in C([0, 1], G^{(\lfloor p \rfloor)}(\mathbb{R}^d))$ a piecewise abelian p -rough path if the identity*

$$(5.6) \quad \mathbb{X}_{s_1, t_1} \otimes \mathbb{X}_{s_2, t_2} = \mathbb{X}_{s_2, t_2} \otimes \mathbb{X}_{s_1, t_1}$$

holds for all pairs $(s_1, t_1), (s_2, t_2) \in \Delta_{[jh, (j+1)h]}$ and each $j \in \{0, 1, \dots, N-1\}$.

In the case of $\lfloor p \rfloor = 1$, the definition is trivial since the group $G^{(1)}(\mathbb{R}^d) = \mathbb{R}^d$ is abelian. However, suppose we have a bounded variation path $X : [0, 1] \rightarrow \mathbb{R}^d$ such that its level-2 enhancement $\mathbb{X} := S_2(X)$ is a piecewise abelian 2-rough path. Then (5.6) implies that

$$0 = [X_{jh, u}, X_{u, (j+1)h}] = [X_{jh, u}, X_{jh, u} + X_{u, (j+1)h}] = [X_{jh, u}, X_{jh, (j+1)h}],$$

and so for every $t \in [jh, (j+1)h]$, $X_{jh, t}$ and $X_{jh, (j+1)h}$ are parallel vectors. Since both vectors start from X_{jh} , we conclude that X is piecewise linear over the increment $[jh, (j+1)h]$. Due to this observation, heuristically we can think of

piecewise abelian rough paths as the natural non-commutative (that is group-valued), equivalent of piecewise linear paths in $\mathbb{R}^d$.

Log-ODE revisited with piecewise abelian rough paths. We now return to our original task of rewriting the log-ODE method in terms of this new class of rough paths. Recalling Theorem 2.8.4, the original SDE (5.1) can be rewritten as the RDE with drift

$$(5.7) \quad dx_t = V(x_t) d\mathbb{W}_t + V_0(x_t) dt, \quad x_0 \in \mathbb{R}^q,$$

where $\mathbb{W} \in C^{p\text{-var}}([0, 1], G^{(\lfloor p \rfloor)}(\mathbb{R}^d))$, $p \in (2, 3)$, denotes enhanced Brownian motion. From this perspective, the underlying idea of the log-ODE method is to approximate solutions of the SDE/RDE (5.1/5.7) by replacing $\mathbb{W}$ with its much simpler N -step piecewise abelian approximation $\mathbb{W}^h$. This is analogous to replacing W with the piecewise interpolation W^h of Brownian motion in the Wong-Zakai theorem (Theorem 2.8.6).

Definition 5.4.2. Define $\mathbb{W}^h \in C^{p\text{-var}}([0, 1], G^{(2)}(\mathbb{R}^d))$ to be the piecewise abelian rough path given iteratively by $\mathbb{W}_0^h := 1$, then

$$\mathbb{W}_{0,t}^h := \mathbb{W}_{0,jh}^h \otimes \exp_2 \left(\frac{t-jh}{h} \xi^{(j)} \right), \quad t \in [jh, (j+1)h],$$

where

$$\xi^{(j)} := \sum_{k=1}^d W_k^{(j)} e_k + \sum_{1 \leq k < l \leq d} A_{kl}^{(j)} [e_k, e_l] \in \mathbb{R}^d \oplus [\mathbb{R}^d, \mathbb{R}^d] = \mathfrak{g}^{(2)}(\mathbb{R}^d).$$

Chen's theorem and the identity $\mathbb{W}_{s,t}^h = (\mathbb{W}_{jh,s}^h)^{-1} \otimes \mathbb{W}_{jh,t}^j$ give

$$\mathbb{W}_{s,t}^h = \exp_2((t-s)\xi^{(j)}) \text{ for all } (s, t) \in \Delta_{[jh, (j+1)h]}.$$

Then the Baker-Campbell-Hausdorff formula (Theorem A.1.1) confirms that $\mathbb{W}^h$ actually is a piecewise abelian rough path.

Remark 5.4.3. One can also think of $\mathbb{W}^h$ as the N -step random walk in the 2-step nilpotent Lie group $G^{(2)}(\mathbb{R}^d)$ with i.i.d. increments $\exp_2(\xi^{(j)}) \in G^{(2)}(\mathbb{R}^d)$ (see [14, 96]). Here the corresponding increments $\xi^{(j)}$ live in the 2-nilpotent Lie algebra $\mathfrak{g}^{(2)}(\mathbb{R}^d)$. Previous research has studied this random walk interpretation of what we have called piecewise abelian rough paths. In [15] Breuillard, Friz and Huesmann used the central limit theorem in nilpotent Lie groups to prove a Donsker-type weak limit theorem for similar random walks converging to enhanced Brownian motion in a particular rough path Hölder topology.

The rough path $\mathbb{W}^h$ is certainly a 2-geometric rough path (take the standard piecewise linear approximation of W and add helical paths of decreasing radius mapping out the correct areas). On the other hand, note that $\mathbb{W}^h$ is (almost surely) not the canonical lift of an actual path in $\mathbb{R}^d$; that is, $\mathbb{W}^h \neq S_2(Y)$ for some continuous stochastic process $Y : [0, 1] \rightarrow \mathbb{R}^d$. Indeed if otherwise, Y would enclose non-zero area over each increment yet be piecewise linear, which is a contradiction. Given an increment $\xi^{(0)} \in G^{(2)}(\mathbb{R}^d)$, the interesting isoperimetry problem of finding a helical path $\gamma \in C([0, 1], \mathbb{R}^d)$ with minimal length such that $S_2(\gamma) = \xi^{(0)}$ is known as *reconstruction* (see [7, §5] and [82]).

Having defined $\mathbb{W}^h$, we now reformulate the log-ODE method in terms of a rough differential equation via the central RDE perturbation result of Theorem 2.6.8.

Proposition 5.4.4. *Fix $\beta > 1$ and $\gamma > 2$. Let $y \in C([0, 1], \mathbb{R}^q)$ be the solution of the RDE with drift*

$$(5.8) \quad dy_t = V(y_t) d\mathbb{W}_t^h + V_0(y_t) dt, \quad y_0 = x_0 \in \mathbb{R}^q,$$

where $V_0 \in \text{Lip}^\beta$ and $V = \{V_k\}_{k=1}^d \subset \text{Lip}^\gamma$. Moreover, let $\tilde{x}^h \in C([0, 1], \mathbb{R}^q)$ denote the entire path of the approximation produced by the log-ODE method given by (2.12). Then y and $\tilde{x}^h$ coincide at all times: $y_t = \tilde{x}_t^h$ for all $t \in [0, 1]$.

PROOF. First set $\mathbb{X} \in W\Omega_p(\mathbb{R}^d)$ to be the rough path lift of the standard piecewise linear approximation of W : $\mathbb{X}_{s,t} := S_2(W^h)_{s,t}$, and define the piecewise linear path $\varphi \in C^{1\text{-var}}([0, 1], [\mathbb{R}^d, \mathbb{R}^d])$ by

$$\varphi_{jh,t} := \frac{t - jh}{h} \sum_{1 \leq k < l \leq d} A_{kl}^{(j)}[e_k, e_l], \quad t \in [jh, (j+1)h].$$

Note that φ takes its values in the centre $\mathcal{Z}^{(2)}(\mathbb{R}^d) = [\mathbb{R}^d, \mathbb{R}^d]$ of the Lie algebra $\mathfrak{g}^{(2)}(\mathbb{R}^d)$. Then the perturbed rough path $\mathbb{X}^\varphi \in W\Omega_p(\mathbb{R}^d)$ is given by

$$\mathbb{X}_{s,t}^\varphi := \exp_2(\log(\mathbb{X}_{s,t}) + \varphi_{s,t}) = \exp_2(\log(S_2(W^h)_{s,t}) + \varphi_{s,t}) = \mathbb{W}_{s,t}^h.$$

Next, consider the RDE with drift

$$(5.9) \quad dz_t = V(z_t) d\mathbb{X}_t + V_0(z_t) dt + U(z_t) d\varphi_t, \quad z_0 = x_0,$$

where

$$\begin{aligned} U(z_t) d\varphi_t &= \sum_{1 \leq k, l \leq d} [V_k, V_l](z_t) \langle e_k^* \otimes e_l^*, d\varphi_t \rangle \\ &= \sum_{1 \leq k < l \leq d} [V_k, V_l](z_t) \langle e_k^* \otimes e_l^* - e_l^* \otimes e_k^*, d\varphi_t \rangle \\ &= \sum_{1 \leq k < l \leq d} [V_k, V_l](z_t) \langle [e_k^*, e_l^*], d\varphi_t \rangle. \end{aligned}$$

Since $\mathbb{X} = S_2(W^h)$, the RDE (5.9) can be rewritten as the ODE

$$dz_t = V(z_t) dW_t^h + V_0(z_t) dt + U(z_t) d\varphi_t, \quad z_0 = x_0.$$

Moreover, as W^h and φ are piecewise linear we can explicitly write the solution as

$$\begin{aligned} z_t &= z_{jh} + \int_{jh}^t \{V(z_s) dW_s^h + V_0(z_s) ds + U(z_s) d\varphi_s\} \\ &= z_{jh} + \frac{1}{h} \int_{jh}^t \left\{ \sum_{k=1}^d W_k^{(j)} V_k + hV_0 + \sum_{1 \leq k < l \leq d} A_{kl}^{(j)} [V_k, V_l] \right\} (z_s) ds \\ &= \exp \left(\frac{t - jh}{h} \left\{ \sum_{k=1}^d W_k^{(j)} V_k + hV_0 + \sum_{1 \leq k < l \leq d} A_{kl}^{(j)} [V_k, V_l] \right\} \right) (z_{jh}), \end{aligned}$$

for all $t \in [jh, (j+1)h]$. In other words, z is precisely the log-ODE method given by (2.12). However, the central perturbation result of Theorem 2.6.8 guarantees that z coincides precisely with the RDE (5.8). The proof is complete. $\square$

Remark 5.4.5. To prove the weaker statement:

$$y_{jh} = \tilde{x}_{jh}^h = x^{(j)} \text{ for all } j \in \{1, \dots, N\},$$

a more intuitive proof would be to compare the stochastic Taylor expansions of y_{jh} and $x^{(j)}$. One finds that the expansions agree at each time $t \in \{jh\}_{j=1}^N$ for every order, hence the approximations must coincide. For more details, see [46, Theorem 2] and [51].

We can also consider the rough path lift of $\mathbb{W}^h$ to level $\kappa > 2$:

$$S_\kappa(\mathbb{W}^h) \in C^{p\text{-var}}([0, 1], G^{(\kappa)}(\mathbb{R}^d)).$$

This multiplicative functional is the unique lift of $\mathbb{W}^h$ in the sense that the p -variation of $S_\kappa(\mathbb{W}^h)$ is equal, (up to multiplicative constants), to that of $\mathbb{W}^h$. Indeed, by the Lipschitz-continuity of the lift operator (Theorem 2.5.1), given $p \in [2, 3)$ there exists a constant $C_{\kappa,p} > 0$ such that $\|\mathbb{W}^h\|_{p\text{-var};[0,1]} \leq \|S_\kappa(\mathbb{W}^h)\|_{p\text{-var};[0,1]} \leq C_{\kappa,p} \|\mathbb{W}^h\|_{p\text{-var};[0,1]}$. But what is $S_\kappa(\mathbb{W}^h)$ precisely? To

answer this we embed the increments

$$\xi^{(j)} = W^{(j)} + A^{(j)} \in \mathbb{R}^d \oplus [\mathbb{R}^d, \mathbb{R}^d] = \mathfrak{g}^{(2)}(\mathbb{R}^d)$$

as increments $\widehat{\xi}^{(j)}$ in the larger free κ -step nilpotent Lie algebra $\mathfrak{g}^{(\kappa)}(\mathbb{R}^d)$. Then we can define a piecewise abelian κ -rough path $\mathbb{Z}^h \in C([0, 1], G^{(\kappa)}(\mathbb{R}^d))$ by $\mathbb{Z}_0^h = 1$ and

$$\mathbb{Z}_{0,t}^h = \mathbb{Z}_{0,jh}^h \otimes \exp_{\kappa} \left(\frac{t - jh}{h} \widehat{\xi}^{(j)} \right), \quad t \in [jh, (j+1)h].$$

Lemma 5.4.6. *We have $S_{\kappa}(\mathbb{W}^h) = \mathbb{Z}^h$.*

PROOF. To prove this, first note that the first two levels of $\mathbb{W}^h$ and $\mathbb{Z}^h$ agree:

$$\pi_{0,2}(\mathbb{Z}_{s,t}^h - \mathbb{W}_{s,t}^h) = 0 \text{ for all } (s, t) \in \Delta_{[0,1]}.$$

Moreover, by its construction $\mathbb{Z}^h$ is certainly a multiplicative functional. By the rough path extension theorem of [79, Theorem 3.7] it remains to show that the p -variation of the higher tensor level increments of $\mathbb{Z}^h$ are controlled by that of $\mathbb{W}^h$. By Chen's theorem, it suffices to consider the p -variation over the single interval $[0, h]$. Exploiting the piecewise geodesic nature of $\mathbb{W}^h$ and $\mathbb{Z}^h$, we know from the equivalence given by (2.3) that there exists constants $C_i = C_i(\kappa) > 0$, ($i = 1, 2$), such that

$$\begin{aligned} \|\mathbb{W}^h\|_{p\text{-var};[0,h]} &= \|\xi^{(0)}\|_C \leq C_1 \left(|W^{(0)}| \vee \sqrt{\|A^{(0)}\|_{(\mathbb{R}^d)^{\otimes 2}}} \right) \\ &= C_2 \left\| \widehat{\xi}^{(0)} \right\|_C = C_2 \|\mathbb{Z}^h\|_{p\text{-var};[0,h]}. \end{aligned}$$

Here, the Carnot-Carathéodory norms are taken over $G^{(2)}(\mathbb{R}^d)$ and $G^{(\kappa)}(\mathbb{R}^d)$ in the first and second lines respectively. We conclude that $S_{\kappa}(\mathbb{W}^h) = \mathbb{Z}^h$. $\square$

Remark 5.4.7. Our interest in the enhancement $S_{\kappa}(\mathbb{W}^h)$ comes from the following observation that will become critical in the proof of our main result: if we replace the driving rough path $\mathbb{W}^h$ by $S_{\kappa}(\mathbb{W}^h)$ in the RDE (5.8), the solution

of Proposition 5.4.4 remains the same. Strictly speaking, we must assume the stronger Stein-Lipschitz condition: $V = \{V_k\}_{k=1}^d \subset \text{Lip}^\gamma$ for $\gamma > \kappa > 2$ in order for the new RDE to remain well-defined.

To see this directly, we note that by their construction the log-signatures of $\mathbb{W}^h$ and $S_\kappa(\mathbb{W}^h)$ coincide over each increment $[jh, (j+1)h]$:

$$\log(S_\kappa(\mathbb{W}^h)_{jh, (j+1)h}) = \log(\mathbb{W}^h_{jh, (j+1)h}).$$

Therefore, by the central perturbation result of Theorem 2.6.8, we conclude that RDEs driven by $S_\kappa(\mathbb{W}^h)$ are precisely the same as that driven by the original $\mathbb{W}^h$. An important consequence of this is that the log-ODE technique produces the same approximation points whether we use $\mathbb{W}^h$ or $S_\kappa(\mathbb{W}^h)$ to drive the RDE in Proposition 5.4.4. One can see that this phenomenon holds for arbitrary rough paths and their corresponding lifts.

5.5. Coupling piecewise abelian rough paths

In practice it is difficult to drive RDEs with the piecewise abelian rough path $\mathbb{W}^h$ because we must be able to generate Lévy area increments $A^{(j)}$. As mentioned previously, if $d > 2$ then the task of accurately generating these increments is an open problem while if $d = 2$ then the existing Fourier expansion techniques are computationally expensive. Following Davie's scheme introduced in Section 4.5, we propose another piecewise abelian 2-rough path $\mathbb{X}^h$ which substitutes each $A^{(j)}$ with the $B^{(j)}$ defined in (4.15).

Definition 5.5.1. *In the same fashion as the construction of $\mathbb{W}^h$ in Definition 5.4.2, we define $\mathbb{X}^h \in C([0, 1], G^{(2)}(\mathbb{R}^d))$ to be the piecewise abelian, geometric*

2-rough path given by $\mathbb{X}_0^h := 1$ and

$$\mathbb{X}_{0,t}^h := \mathbb{X}_{0,jh}^h \otimes \exp_2 \left(\frac{t-jh}{h} \eta^{(j)} \right), \quad t \in [jh, (j+1)h],$$

where the increments $\eta^{(j)} \in \mathfrak{g}^{(2)}(\mathbb{R}^d)$ are defined as:

$$\eta^{(j)} := \sum_{k=1}^d W_k^{(j)} e_k + \sum_{1 \leq k < l \leq d} B_{kl}^{(j)} [e_k, e_l].$$

As with $\mathbb{W}^h$, we may consider the κ -lift $S_\kappa(\mathbb{X}^h)$ of $\mathbb{X}^h$ and, as before, RDEs driven by $\mathbb{X}^h$ and $S_\kappa(\mathbb{X}^h)$ coincide (assuming $\gamma > \kappa$). Moreover, again using Theorem 2 of [46] we can prove that the approximation scheme $\{x_j^N\}_{j=0}^N$ given by (5.2) coincides with the solution of the following RDE with drift:

$$dz_t = V(z_t) d\mathbb{X}_t^h + V_0(z_t) dt, \quad z_0 = x_0 \in \mathbb{R}^q.$$

Importantly, $z_{jh} = x_j^N$ for $j \in \{1, \dots, N\}$.

The coupling of $(A^{(j)})$ and $(B^{(j)})$ conditional on $(W^{(j)})$ provided by Theorem 6.2.1 in the next chapter automatically induces a coupling between $\mathbb{W}^h$ and $\mathbb{X}^h$. Of critical importance is that fact that this coupling uses the same underlying Brownian increments $W^{(j)}$ for both $\mathbb{W}^h$ and $\mathbb{X}^h$. As we will see, this significantly reduces the algebraic complexity when considering the lifted rough paths via the Baker-Campbell-Hausdorff formula and allows us to establish useful Hölder estimates for the lifted rough paths in Section 5.6.

We use the remainder of this section to study how well this coupling performs in the topology of the rough path Hölder metric. Since we are interested in numerically approximating SDEs using these rough paths, we need to employ the useful RDE Lipschitz estimates of [11, 18, 47] in the form of Proposition 2.9.2. The latter estimate uses the inhomogeneous $1/p$ -Hölder rough path metric and so this is the metric we work with.

Establishing Hölder bounds. Next, we focus on establishing Hölder bounds for our original coupling of $\mathbb{W}^h$ and $\mathbb{X}^h$ provided by Theorem 6.2.1.

First, it can be shown that the common first level satisfies

$$|\pi_1(\mathbb{W}_{s,t}^h)|_{L^r} = |\pi_1(\mathbb{X}_{s,t}^h)|_{L^r} = |W_{s,t}^h|_{L^r} = C_r |t-s|^{1/2} \text{ for all } r \in [1, \infty)$$

and by scaling the second levels satisfy

$$\left\| \left\| \pi_2(\mathbb{W}_{s,t}^h) \right\|_{(\mathbb{R}^d)^{\otimes 2}} \right\|_{L^r} + \left\| \left\| \pi_2(\mathbb{X}_{s,t}^h) \right\|_{(\mathbb{R}^d)^{\otimes 2}} \right\|_{L^r} \leq C_r |t-s|.$$

Thus the standard Kolmogorov regularity result of [52, Theorem A.12] implies that for all $\alpha \in [0, \frac{1}{2})$,

$$|\rho_{\alpha\text{-Höl};[0,1]}(\mathbb{W}^h)|_{L^r} + |\rho_{\alpha\text{-Höl};[0,1]}(\mathbb{X}^h)|_{L^r} \leq C_r.$$

We now consider the Hölder distance between $\mathbb{W}^h$ and $\mathbb{X}^h$. Just by scaling,

$$\begin{aligned} \left\| \left\| \pi_2(\mathbb{W}_{s,t}^h - \mathbb{X}_{s,t}^h) \right\|_{(\mathbb{R}^d)^{\otimes 2}} \right\|_{L^r} &= \left\| \left\| \frac{t-s}{h} (A^{(j)} - B^{(j)}) \right\|_{(\mathbb{R}^d)^{\otimes 2}} \right\|_{L^r} \\ &\leq C_r |t-s| h^{-1} h = C_r |t-s| \end{aligned}$$

for all $(s, t) \in \Delta_{[jh, (j+1)h]}$. Recall the first levels of $\mathbb{W}^h$ and $\mathbb{X}^h$ agree:

$$\pi_1(\mathbb{W}_{jh, (j+1)h}^h) = W^{(j)} = \tilde{W}^{(j)} := \pi_1(\mathbb{X}_{jh, (j+1)h}^h).$$

Thus the Baker-Campbell-Hausdorff formula reduces the second level difference to the coupling error of the two $[\mathbb{R}^d, \mathbb{R}^d]$ -valued random walks:

$$\begin{aligned} \mathbb{W}_{r_1 h, r_2 h}^h - \mathbb{X}_{r_1 h, r_2 h}^h &= \sum_{j=r_1}^{r_2-1} (A^{(j)} - B^{(j)}) + \frac{1}{2} \sum_{r_1 \leq i < j < r_2} ([W^{(i)}, W^{(j)}] - [\tilde{W}^{(i)}, \tilde{W}^{(j)}]) \\ &= \sum_{j=r_1}^{r_2-1} (A^{(j)} - B^{(j)}). \end{aligned}$$

Thus for larger increments $|t-s| \geq N^{-1}$, the coupling of Theorem 6.2.1 ensures that for all $r \in [1, \infty)$, there exists a constant $C_r > 0$ such that

$$\left\| \left\| \pi_2(\mathbb{W}_{s,t}^h - \mathbb{X}_{s,t}^h) \right\|_{(\mathbb{R}^d)^{\otimes 2}} \right\|_{L^r} \leq \max_{0 \leq r_1 < r_2 \leq N-1} \left\| \left\| \sum_{j=r_1}^{r_2-1} (A^{(j)} - B^{(j)}) \right\|_{(\mathbb{R}^d)^{\otimes 2}} \right\|_{L^r} \leq C_r \frac{\log N}{N}.$$

We summarise the situation with the next lemma.

Lemma 5.5.2. *For all $r \in [1, \infty)$ there exists a constant $C_r > 0$ such that*

$$\left\| \left\| \pi_2(\mathbb{W}_{s,t}^h - \mathbb{X}_{s,t}^h) \right\|_{(\mathbb{R}^d)^{\otimes 2}} \right\|_{L^r} \leq C_r \left(\frac{\log N}{N} \wedge |t - s| \right) \text{ for all } (s, t) \in \Delta_{[0,1]}.$$

In turn, these estimates quickly establish a corresponding Hölder bound.

Proposition 5.5.3. *For all $r \in [1, \infty)$ there exists a constant $C_r > 0$ such that for all $\alpha \in [0, \frac{\theta}{2})$ we have*

$$\left| \rho_{\alpha\text{-Höl};[0,1]}(\mathbb{W}^h, \mathbb{X}^h) \right|_{L^r} \leq C_r \left(\frac{\log N}{N} \right)^{1-\theta}.$$

PROOF. Since $\mathbb{W}^h$ and $\mathbb{X}^h$ share a common first level,

$$\rho_{\alpha\text{-Höl};[0,1]}(\mathbb{W}^h, \mathbb{X}^h) = \rho_{\alpha\text{-Höl};[0,1]}^{(2)}(\mathbb{W}^h, \mathbb{X}^h) = \sup_{0 \leq s < t \leq 1} \frac{\left\| \pi_2(\mathbb{W}_{s,t}^h - \mathbb{X}_{s,t}^h) \right\|_{(\mathbb{R}^d)^{\otimes 2}}}{|t - s|^{2\alpha}}.$$

Lemma 5.5.2 guarantees that for every $\theta \in [0, 1]$,

$$\left\| \left\| \pi_2(\mathbb{W}_{s,t}^h - \mathbb{X}_{s,t}^h) \right\| \right\|_{L^r} \leq C_r \left(\frac{\log N}{N} \wedge |t - s| \right) \leq C_r \left(\frac{\log N}{N} \right)^{1-\theta} |t - s|^\theta.$$

Here, we have used the inequality $a \wedge b \leq a^{1-\theta} b^\theta$ for $a, b \geq 0$. Appealing to the Kolomogorov result for rough paths of [52, Theorem A.14], we arrive at the claim. $\square$

It may seem that for our approximation $\mathbb{X}^h$ of $\mathbb{W}^h$ to be of any use in RDE Lipschitz estimates, we need the previous quantity $(\frac{\log N}{N})^{1-\theta}$ to be less than $O(N^{-1/2})$. Otherwise we might as well have used the level-1 log-ODE method which has order $O(N^{-1/2})$ and neither needs Lévy area increments (in common with $\mathbb{X}^h$). However this requires $\theta < \frac{1}{2}$, which in turn demands that $\alpha \in [0, \frac{1}{4})$. Unfortunately for $\alpha < \frac{1}{4}$, $\rho_{\alpha\text{-Höl}}(\cdot)$ is no longer a rough path metric for 2-rough paths. Therefore to make the Hölder bound of Proposition 5.5.3 useful at all we would need to not only compute *but control* the p -variation,

(where $p = \alpha^{-1} > 4$), of at least the first 4 levels of the lifts $S_\kappa(\mathbb{W}^h)$ and $S_\kappa(\mathbb{X}^h)$, ($\kappa \geq 4$), and the corresponding difference at each tensor level:

$$\max_{k=2,\dots,\kappa} \left\| \pi_k \left(S_\kappa(\mathbb{W}^h)_{s,t} - S_\kappa(\mathbb{X}^h)_{s,t} \right) \right\|_{(\mathbb{R}^d)^{\otimes k}} \Big|_{L^r}.$$

This is precisely what we will establish in the next section. Importantly, we know from the discussion in Remark 5.4.7 that lifting the piecewise abelian rough paths $\mathbb{W}^h, \mathbb{X}^h$ does not change the RDE defining the log-ODE method and our approximation scheme.

Remark 5.5.4. This technique of lifting the rough paths $\mathbb{W}^h$ and $\mathbb{X}^h$ to higher levels $\kappa > 2$ in order to get a better p -variation distance bound is inspired by the same technique for dealing with the convergence rates of Gaussian rough paths by Friz, Riedel and Xu in [48, 103] (see specifically the remarks made after Corollary 1 in [48] and [103, Remark 5.2]). The idea is that by lifting the driving rough path to a higher level κ , p is allowed to be larger, and hence a better Hölder estimate may be hoped for. This is akin to the fact that there exist examples of paths which are far apart in the 1-Hölder (1-variation) metric but close in the $1/p$ -Hölder (p -variation) metric for very large $p \in [1, \infty)$.

In order for the RDE to still be well-defined and have a unique solution, we are penalised by requiring that the Stein-Lipschitz order of the vector fields is of order $\gamma > p$. In their particular applications, the papers [48, 103] recover (almost) optimal rates of convergence for Wong-Zakai approximations of RDEs driven by various Gaussian rough path lifts. This includes the case of fractional Brownian motion with Hurst index $H \in (\frac{1}{3}, 1]$, (see the final remark after Corollary 1 of [48, Corollary 1]).

5.6. Lifted rough path Hölder estimates

The aim of this section is to prove that the lifting mechanism discussed in the previous section will give the inhomogeneous $1/p$ -Hölder estimates needed to establish Corollary 5.7.1 in the next section. This is the content of the final result of this section, Theorem 5.6.6.

As our first step towards proving the theorem, we consider an error estimate for our lifted rough path coupling.

Proposition 5.6.1. *Fix an integer $\kappa > 2$. Then for all $r \in [1, \infty)$ there exists a constant $C_{\kappa,r} > 0$ such that for each integer $2 \leq m \leq \kappa$,*

$$\left\| \pi_m (S_\kappa(\mathbb{W}^h)_{s,t} - S_\kappa(\mathbb{X}^h)_{s,t}) \right\|_{(\mathbb{R}^d)^{\otimes m}} \Big|_{L^r} \leq C_{\kappa,r} |t - s|^{\frac{m}{2}-1} \left(\frac{\log N}{N} \wedge |t - s| \right)$$

for all $(s, t) \in \Delta_{[0,1]}$.

Before giving the proof we present some useful technical results. We begin with a small lemma concerning the tensor norm of log-signature increments (cf. §3 of [84]).

Lemma 5.6.2. *Fix an integer $\kappa > 2$, $p \in [2, 3)$ and let $\mathbb{X} \in C^{1/p\text{-Hö}}([0, 1], G^{(2)}(\mathbb{R}^d))$. Then there exists a constant $C_{\kappa,p} > 0$ such that for all $m = 1, \dots, \kappa$,*

$$\left\| \pi_m \left(\log S_\kappa(\mathbb{X})_{s,t} \right) \right\|_{(\mathbb{R}^d)^{\otimes m}} \leq C_{\kappa,p} |t - s|^{m/p} \|\mathbb{X}\|_{1/p\text{-Hö};[s,t]}^m.$$

PROOF. The function

$$g \in G^{(\kappa)}(\mathbb{R}^d) \mapsto \max_{k=1,\dots,\kappa} \|\pi_k(\log g)\|_{(\mathbb{R}^d)^{\otimes k}}^{1/k}$$

is a homogeneous norm on $G^{(\kappa)}(\mathbb{R}^d)$ ([52, Exercise 7.38]). Therefore by the equivalence of such norms on $G^{(\kappa)}(\mathbb{R}^d)$ given by [52, Theorem 7.44], there exists

a constant $C_\kappa \geq 1$ such that

$$\begin{aligned} \|\pi_m(\log S_\kappa(\mathbb{X})_{s,t})\|_{(\mathbb{R}^d)^{\otimes m}} &\leq \left(\max_{k=1,\dots,\kappa} \|\pi_k(\log S_\kappa(\mathbb{X})_{s,t})\|_{(\mathbb{R}^d)^{\otimes k}}^{1/k} \right)^k \\ &\leq C_\kappa \|S_\kappa(\mathbb{X})_{s,t}\|_C^k \leq C_\kappa \|S_\kappa(\mathbb{X})\|_{p\text{-var};[s,t]}^k. \end{aligned}$$

The Lipschitz-continuity of the rough path lift in p -variation (Theorem 2.5.1) guarantees the existence of a constant $C_{\kappa,p} > 0$ such that $\|S_\kappa(\mathbb{X})\|_{p\text{-var};[s,t]} \leq C_{\kappa,p} \|\mathbb{X}\|_{p\text{-var};[s,t]}$. Moreover, a simple consequence of the super-additivity of controls ([52, §8.1]) gives $\|\mathbb{X}\|_{p\text{-var};[s,t]} \leq |t-s|^{1/p} \|\mathbb{X}\|_{1/p\text{-Höl};[s,t]}$. Putting the last three inequalities together completes the proof. $\square$

We apply Lemma A.2.2 with

$$(5.10) \quad x_{j+1} = \xi^{(j)} = W^{(j)} + A^{(j)}, \quad y_{j+1} = \eta^{(j)} = W^{(j)} + B^{(j)},$$

for $j \in \{0, 1, \dots, N-1\}$. Fix $n \in \{1, \dots, N\}$ and define

$$(5.11) \quad \begin{aligned} g_i &:= \pi_i(H(x_1, \dots, x_n)) = \pi_i(\log S_\kappa(\mathbb{W}^h)_{0,nh}), \\ h_i &:= \pi_i(H(y_1, \dots, y_n)) = \pi_i(\log S_\kappa(\mathbb{X}^h)_{0,nh}) \in \mathfrak{g}^{(k)}(\mathbb{R}^d). \end{aligned}$$

Then, exploiting the compatibility of the tensor algebra norm, for all $m \leq \kappa$ we have

$$\begin{aligned} &\left\| \pi_m(S_\kappa(\mathbb{W}^h)_{0,nh} - S_\kappa(\mathbb{X}^h)_{0,nh}) \right\|_{(\mathbb{R}^d)^{\otimes m}} \\ &= \left\| \pi_m(e^{x_1} \otimes \dots \otimes e^{x_n} - e^{y_1} \otimes \dots \otimes e^{y_n}) \right\|_{(\mathbb{R}^d)^{\otimes m}} \\ &\leq \sum_{k=1}^m \frac{1}{k!} \sum_{\substack{i_1, \dots, i_k > 0 \\ i_1 + \dots + i_k = m}} \left\| \sum_{j=1}^k g_{i_1} \otimes \dots \otimes g_{i_{j-1}} \otimes (g_{i_j} - h_{i_j}) \otimes h_{i_{j+1}} \otimes \dots \otimes h_{i_k} \right\|_{(\mathbb{R}^d)^{\otimes m}} \\ &\leq \sum_{k=1}^m \frac{1}{k!} \sum_{\substack{i_1, \dots, i_k > 0 \\ i_1 + \dots + i_k = m}} \sum_{j=1}^k \|g_{i_1} \otimes \dots \otimes g_{i_{j-1}}\| \|g_{i_j} - h_{i_j}\| \|h_{i_{j+1}} \otimes \dots \otimes h_{i_k}\|. \end{aligned}$$

Since the increments of $\mathbb{W}^h$ and $\mathbb{X}^h$ agree it follows that $g_1 - h_1 = 0$ and

$$g_2 - h_2 = \sum_{j=0}^{n-1} (A^{(j)} - B^{(j)}) \in [\mathbb{R}^d, \mathbb{R}^d].$$

Moreover, by Lemma 5.6.2, given $p \in (2, 3)$, we have

$$\|g_i\|_{(\mathbb{R}^d)^{\otimes i}} = \|\pi_i(\log S_\kappa(\mathbb{W}^h)_{0,nh})\|_{(\mathbb{R}^d)^{\otimes i}} \leq C_{\kappa,p} \left(\frac{n}{N}\right)^{i/p} \|\mathbb{W}^h\|_{1/p\text{-Höl};[0,nh]}^i$$

and an identical bound holds for h_i with $\mathbb{X}^h$. Therefore the previous inequality becomes

$$\begin{aligned} & \|\pi_m(S_\kappa(\mathbb{W}^h)_{0,nh} - S_\kappa(\mathbb{X}^h)_{0,nh})\|_{(\mathbb{R}^d)^{\otimes m}} \\ & \leq C_{\kappa,p} \sum_{k=1}^m \frac{1}{k!} \sum_{\substack{i_1, \dots, i_k > 0 \\ i_1 + \dots + i_k = m}} \sum_{j=1}^k \left(\frac{n}{N}\right)^{\frac{1}{p}(m-i_j)} \\ & \quad \times \|\mathbb{W}^h\|_{1/p\text{-Höl};[0,nh]}^{\frac{1}{p}(i_1 + \dots + i_{j-1})} \|\mathbb{X}^h\|_{1/p\text{-Höl};[0,nh]}^{\frac{1}{p}(i_{j+1} + \dots + i_k)} \|g_{i_j} - h_{i_j}\|_{(\mathbb{R}^d)^{\otimes i_j}} \\ & \leq C_{\kappa,p} \sum_{k=1}^m \frac{1}{k!} \sum_{\substack{i_1, \dots, i_k > 0 \\ i_1 + \dots + i_k = m}} \sum_{j=1}^k \left(\frac{n}{N}\right)^{\frac{1}{p}(m-i_j)} \\ & \quad \times \left(\|\mathbb{W}^h\|_{1/p\text{-Höl};[0,nh]} \vee \|\mathbb{X}^h\|_{1/p\text{-Höl};[0,nh]} \right)^{m-i_j} \|g_{i_j} - h_{i_j}\|. \end{aligned} \tag{5.12}$$

So it remains to bound $\|g_i - h_i\|_{(\mathbb{R}^d)^{\otimes i}}$ in L^r for each integer $3 \leq i \leq \kappa$. This is the content of Lemma 5.6.5 below, which takes particular care in separating out the dependencies on the variables m and n . To build intuition on the iterated Baker-Campbell-Hausdorff formula, we refer to the specific calculations of $g_m - h_m$ for $m \leq 5$ in (A.1) found in the appendix at the end of the thesis. But first we need the following consequence of symmetry.

Lemma 5.6.3. *Unless the indices $(j_1, \dots, j_{2n}) \in \{0, 1, \dots, N-1\}^{2n}$ are all present in pairs, we have*

$$\mathbb{E} \left(W^{(j_1)} \otimes \dots \otimes W^{(j_{2n})} \mid \{A^{(j)}, B^{(j)}\}_{j=0}^{N-1} \right) = 0.$$

PROOF. If the index j_* does not belong to a pair then there must be some odd number of $W^{(j_*)}$ increment terms in the tensor product $W^{(j_1)} \otimes \dots \otimes W^{(j_{2n})}$. Hence in every coordinate projection of $(\mathbb{R}^d)^{\otimes(2n)}$

$$\langle e_{c_1}^* \otimes \dots \otimes e_{c_{2n}}^*, W^{(j_1)} \otimes \dots \otimes W^{(j_{2n})} \rangle,$$

one particular term $W_k^{(j_*)}$ (for some $k \in \{c_1, \dots, c_{2n}\}$) appears an odd number of times. By symmetry it follows that

(5.13)

$$\mathbb{E} \left(\langle e_{c_1}^* \otimes \dots \otimes e_{c_{2n}}^*, W^{(j_1)} \otimes \dots \otimes W^{(j_{2n})} \rangle \mid \{A^{(j)}, B^{(j)}\}_{j=0}^{N-1}, \{W^{(j_i)}\}_{j_i \neq j_*} \right) = 0.$$

Indeed, recalling (4.14) and (4.15), for each j there exists a matrix $Z^{(j)} \in \mathbb{R}^{d(d-1) \times d}$ with entries taking values in $\{\pm z_k^{(j)}, \pm \zeta_k^{(j)}\}_{k=1}^d$ such that

$$\begin{pmatrix} A^{(j)} \\ B^{(j)} \end{pmatrix} = Z^{(j)} W^{(j)} + \begin{pmatrix} K^{(j)} \\ \lambda^{(j)} \end{pmatrix},$$

where $A^{(j)} = (A_{12}^{(j)}, \dots, A_{d-1,d}^{(j)})^t \in \mathbb{R}^{\frac{d}{2}(d-1)}$ with similar definitions for $B^{(j)}, K^{(j)}$ and $\lambda^{(j)}$. Certainly we have the equality of laws:

$$\mathcal{L} \left(\begin{pmatrix} A^{(j)} \\ B^{(j)} \end{pmatrix} \right) = \mathcal{L} \left(Z^{(j)} (-W^{(j)}) + \begin{pmatrix} K^{(j)} \\ \lambda^{(j)} \end{pmatrix} \right),$$

from which (5.13) immediately follows. Applying the tower property of conditional expectation concludes the proof. $\square$

For convenience, let us define the operation $*$ on $\mathfrak{g}^{(m)}(\mathbb{R}^d)$ by setting

$$(5.14) \quad x_1 * x_2 * \dots * x_{k-1} * x_k := [x_1, [x_2, [\dots, [x_{k-1}, x_k] \dots]]].$$

Lemma 5.6.4. *There exists a universal constant $C_m > 0$ such that for every $x_1, \dots, x_k \in \mathfrak{g}^{(m)}(\mathbb{R}^d)$, we have*

$$\|\pi_m(x_1 * \dots * x_k)\|_{(\mathbb{R}^d)^{\otimes m}} \leq C_m \sum_{1 \leq j_1, \dots, j_m \leq d} |\langle e_{j_1}^* \otimes \dots \otimes e_{j_m}^*, \pi_m(x_1 \otimes \dots \otimes x_k) \rangle|.$$

PROOF. If $k > m$ then the inequality is trivial since both sides are zero by the truncation operator π_m . So we fix $k \leq m$. Unfolding the nested Lie brackets of $x_1 * \dots * x_k$, there will be at most $2^k \leq 2^m$ non-zero terms in the resultant expansion with each term taking the form

$$\pm x_{\sigma(1)} \otimes x_{\sigma(2)} \otimes \dots \otimes x_{\sigma(k)} \text{ for some } \sigma \in S_k,$$

where S_k is the symmetric group of permutations on $\{1, \dots, k\}$, (see Lemma A.2.3). In other words, every x_i will be present exactly once in each term in the expansion of $x_1 * \dots * x_k$. By symmetry,

$$\|\pi_m(x_{\sigma(1)} \otimes \dots \otimes x_{\sigma(k)})\|_{(\mathbb{R}^d)^{\otimes m}} = \|\pi_m(x_1 \otimes \dots \otimes x_k)\|_{(\mathbb{R}^d)^{\otimes m}} \text{ for all } \sigma \in S_k.$$

Since $k \leq m$, it follows that there exists a constant $C_m > 0$ such that

$$\begin{aligned} & \|\pi_m(x_1 * \dots * x_k)\|_{(\mathbb{R}^d)^{\otimes m}} \\ & \leq C_m \|\pi_m(x_1 \otimes \dots \otimes x_k)\|_{(\mathbb{R}^d)^{\otimes m}} \\ & = C_m \sqrt{\sum_{1 \leq j_1, \dots, j_m \leq d} |\langle e_{j_1}^* \otimes \dots \otimes e_{j_m}^*, \pi_m(x_1 \otimes \dots \otimes x_k) \rangle|^2}. \end{aligned}$$

Using the inequality $\|a\|_{l^2} \leq \|a\|_{l^1}$ for $a \in \mathbb{R}^s$, ($s = d^m$), yields

$$\|\pi_m(x_1 \otimes \dots \otimes x_k)\|_{(\mathbb{R}^d)^{\otimes m}} \leq \sum_{1 \leq j_1, \dots, j_m \leq d} |\langle e_{j_1}^* \otimes \dots \otimes e_{j_m}^*, \pi_m(x_1 \otimes \dots \otimes x_k) \rangle|.$$

The claim follows. $\square$

Lemma 5.6.5. *For $n \in \{1, \dots, N\}$ and $m \in \{2, \dots, \kappa\}$ define:*

$$g_m = \pi_m(\log S_\kappa(\mathbb{W}^h)_{0,nh}), \quad h_m = \pi_m(\log S_\kappa(\mathbb{X}^h)_{0,nh}) \in \mathfrak{g}^{(m)}(\mathbb{R}^d).$$

Then for each m and $r \in [1, \infty)$ there exists a constant $C_{m,r} > 0$ such that

$$(5.15) \quad \left\| \|g_m - h_m\|_{(\mathbb{R}^d)^{\otimes m}} \right\|_{L^r} \leq C_{m,r} \left(\frac{n}{N} \right)^{\frac{m}{2}-1} \frac{\log N}{N}.$$

PROOF. The case of $m = 2$ is precisely the content of our coupling result in Theorem 6.2.1 so we restrict the proof to $3 \leq m \leq \kappa$. We define x_i, y_i as in (5.10)

so that our definition of g_m, h_m agrees with (5.11). Let $\tau_k : \mathfrak{g}^{(m)}(\mathbb{R}^d) \rightarrow \mathfrak{g}^{(m)}(\mathbb{R}^d)$ denote the canonical projection onto nested Lie brackets terms of length k . Note that $\tau_k(g_m - h_m) = 0$ for all $k \geq m$. Indeed, for $k > m$ this is a consequence of truncation, while in the case of $k = m$, the nested Lie brackets will only consist of $W^{(j)}$ increment terms, which are common to both expressions.

Using Lemma A.2.3 for rearranging nested Lie brackets into brackets of the form of (5.14), the Baker-Campbell-Hausdorff formula (BCHF) gives index sets $\Lambda_k \subset \{1, \dots, N-1\}^k$ and corresponding coefficients c_λ such that

$$\tau_k(g_m) = \sum_{\lambda \in \Lambda_k} c_\lambda \sum_{(i_1, \dots, i_k) \in \lambda} \pi_m(x_{i_1} * \dots * x_{i_k}).$$

Setting $z_i := x_i - y_i = A^{(i-1)} - B^{(i-1)}$, we can exploit the property that a nested bracket of length k is a k -multilinear map and employ a telescoping sum to find that

$$\begin{aligned} \tau_k(g_m - h_m) &= \sum_{\lambda \in \Lambda_k} c_\lambda \sum_{(i_1, \dots, i_k) \in \lambda} \pi_m(x_{i_1} * \dots * x_{i_k} - y_{i_1} * \dots * y_{i_k}) \\ &= \sum_{\lambda \in \Lambda_k} c_\lambda \sum_{(i_1, \dots, i_k) \in \lambda} \pi_m(z_{i_1} * x_{i_2} * \dots * x_{i_k} \\ &\quad + y_{i_1} * z_{i_2} * x_{i_3} * \dots * x_{i_k} + \dots + y_{i_1} * \dots * y_{i_{k-1}} * z_{i_k}) \\ &= \sum_{s=1}^k \sum_{\lambda \in \Lambda_k} c_\lambda \sum_{(i_1, \dots, i_k) \in \lambda} \pi_m(y_{i_1} * \dots * y_{i_{s-1}} * z_{i_s} * x_{i_{s+1}} * \dots * x_{i_k}). \end{aligned}$$

We fix $1 \leq s \leq k$ and $\lambda \in \Lambda_k$ then consider the corresponding term

$$(5.16) \quad \sum_{(i_1, \dots, i_k) \in \lambda} \pi_m(y_{i_1} * \dots * y_{i_{s-1}} * z_{i_s} * x_{i_{s+1}} * \dots * x_{i_k}).$$

There are $N_{k,m} \leq N_m$ such terms in the above sum, where N_m is an integer (independent of n) determined by the iterated BCHF. Therefore, to prove the

lemma it suffices to show that

$$(5.17) \quad \left\| \left\| \sum_{(i_1, \dots, i_k) \in \lambda} \pi_m(y_{i_1} * \dots * y_{i_{s-1}} * z_{i_s} * x_{i_{s+1}} * \dots * x_{i_k}) \right\|_{(\mathbb{R}^d)^{\otimes m}} \right\|_{L^r} \leq C_{m,r} \left(\frac{n}{N} \right)^{\frac{m}{2}-1} \frac{\log N}{N}.$$

To this end, we note that each nested Lie bracket term in (5.16) will contain $N_1 = N_1(s, \lambda)$ increment terms (that is, $W^{(i)}$) and $N_2 = N_2(s, \lambda)$ area terms (with each term either $A^{(i)}, B^{(i)}$ for some i). Moreover, $N_1 + N_2 = k$ and $N_1 + 2N_2 = m$. For each pair (s, λ) , there exists an index set $\theta \subset \{1, \dots, n\}^{s-1}$ and a collection $\{J_I\}_{I \in \theta} \subset \{1, \dots, n\}$ such that

$$\lambda = \bigcup_{I=(i_1, \dots, i_{s-1}, i_{s+1}, \dots, i_k) \in \theta} \{(i_1, \dots, i_{s-1}, i_s, i_{s+1}, \dots, i_k) : i_s \in J_I\}$$

Consequently,

$$(5.18) \quad \begin{aligned} & \sum_{(i_1, \dots, i_k) \in \lambda} \pi_m(y_{i_1} * \dots * y_{i_{s-1}} * z_{i_s} * x_{i_{s+1}} * \dots * x_{i_k}) \\ &= \sum_{I=(i_1, \dots, i_{s-1}, i_{s+1}, \dots, i_k) \in \theta} \sum_{j \in J_I} \pi_m(y_{i_1} * \dots * y_{i_{s-1}} * z_{i_s} * x_{i_{s+1}} * \dots * x_{i_k}) \\ &= \sum_{I=(i_1, \dots, i_{s-1}, i_{s+1}, \dots, i_k) \in \theta} \pi_m \left(y_{i_1} * \dots * y_{i_{s-1}} * \left(\sum_{j \in J_I} z_j \right) * x_{i_{s+1}} * \dots * x_{i_k} \right). \end{aligned}$$

For each $I \in \theta$, the set J_I consists of consecutive elements. Indeed, the induction proof of the iterated Baker-Campbell-Hausdorff formula (A.1) and the bilinearity of nested Lie brackets guarantee that for any collection of fixed indices $i_1, \dots, i_{s-1}, i_{s+1}, \dots, i_k \in \{1, \dots, n\}$, the set $\{j : (i_1, \dots, i_{s-1}, j, i_{s+1}, \dots, i_k) \in \lambda\}$ is either empty or made up of consecutive elements. Since the indices of J_I are consecutive we can use Lemma 5.5.2 to arrive at the estimate

$$(5.19) \quad \left\| \left\| \sum_{j \in J_I} z_j \right\|_{(\mathbb{R}^d)^{\otimes 2}} \right\|_{L^r} = \left\| \left\| \sum_{j \in J_I} (A^{(j-1)} - B^{(j-1)}) \right\|_{(\mathbb{R}^d)^{\otimes 2}} \right\|_{L^r} \leq C_r \frac{\log N}{N}.$$

Now let Z be the projection of (5.18) onto an arbitrary coordinate $(c_1, \dots, c_m)$ of $(\mathbb{R}^d)^{\otimes m}$. Specifically, $Z := \sum_{I \in \theta} \varphi_I$, where we define

$$\varphi_I := \left\langle e_{c_1}^* \otimes \dots \otimes e_{c_m}^*, \pi_m \left(y_{i_1} \otimes \dots \otimes y_{i_{s-1}} \otimes \left(\sum_{j \in J_I} z_j \right) \otimes x_{i_{s+1}} \otimes \dots \otimes x_{i_k} \right) \right\rangle$$

for each index $I = (i_1, \dots, i_{s-1}, i_{s+1}, \dots, i_k) \in \theta$. By the equivalence of norms established in Lemma 5.6.4, in order to establish (5.17) it suffices to prove that

$$|Z|_{L^r} \leq C_{m,r} \left(\frac{n}{N} \right)^{\frac{m}{2}-1} \frac{\log N}{N}.$$

We note that each φ_I can be written as

$$\varphi_I = W_{a_1}^{(r_1)} \dots W_{a_{N_1}}^{(r_{N_1})} R_{b_1}^{(s_1)} \dots R_{b_{N_2-1}}^{(s_{N_2-1})} \left\langle e_{\tilde{c}_1}^* \otimes e_{\tilde{c}_2}^*, \sum_{j \in J_I} z_j \right\rangle$$

where each $R_{b_i}^{(s_j)}$ denotes either $A_{b_i}^{(s_i)}$ or $B_{b_i}^{(s_j)}$, and $a_i, b_i \in \{1, \dots, d\}$ and $r_j, s_j \in \{0, 1, \dots, N-1\}$ are the appropriate coordinate and increment labels respectively (as determined by $(c_1, \dots, c_m)$).

Without loss of generality we may assume $r = 2q$ for some integer $q \in \mathbb{N}$.

Then

$$(5.20) \quad |Z|_{L^r}^r = |Z|_{L^{2q}}^{2q} = \mathbb{E} \left(\left(\sum_{I \in \theta} \varphi_I \right)^{2q} \right) = \sum_{I_1 \in \theta} \dots \sum_{I_{2q} \in \theta} \mathbb{E} (\varphi_{I_1} \dots \varphi_{I_{2q}}).$$

The $rN_1 = 2qN_1$ increment terms in the expression $\varphi_{I_1} \dots \varphi_{I_{2q}}$ must be paired to each other, otherwise the expectation $\mathbb{E} (\varphi_{I_1} \dots \varphi_{I_{2q}})$ will be zero by Lemma 5.6.3 and the tower property of conditional expectation. There are $(rN_1 - 1)!! \leq C_{m,r}$ possible perfect matchings amongst the rN_1 increment indices in the set $\{I_1, \dots, I_{2q}\}$, (where $(2k+1)!! := \frac{(2k)!}{2^k k!}$ for $k \in \mathbb{N}$). Each matching is summed over at most n possible indices and the paired (that is, squared) Brownian increments each have scaling $O(N^{-1})$ in L^r for all $r \in [1, \infty)$. The summation (5.20) also includes $2q$ terms of the form $\left\langle e_{\tilde{c}_1}^* \otimes e_{\tilde{c}_2}^*, \sum_{j \in J_I} z_j \right\rangle$. Recalling (5.19), each of these terms are of the order $O(N^{-1} \log N)$ in L^r . The remaining $2(qN_2 -$

1) area sums are taken over at most n indices each, with each constituent area term having scaling $O(N^{-1})$ in L^r . By combining these L^r -bounds using the Cauchy-Schwarz inequality we conclude that

$$\begin{aligned} |Z|_{L^{2q}}^{2q} &\leq C_{m,r} \left(\frac{n}{N}\right)^{qN_1} \left(\frac{n}{N}\right)^{2q(N_2-1)} \left(\frac{\log N}{N}\right)^{2q} \\ &= C_{m,r} \left(\frac{n}{N}\right)^{q(N_1+2N_2-2)} \left(\frac{\log N}{N}\right)^{2q} = C_{m,r} \left(\frac{n}{N}\right)^{q(m-2)} \left(\frac{\log N}{N}\right)^{2q}. \end{aligned}$$

The proof is complete. $\square$

PROOF OF PROPOSITION 5.6.1. For $|t-s| \leq h$ the result is immediate from scaling so we restrict ourselves to $|t-s| \geq h$. Since $\mathbb{W}^h$ and $\mathbb{X}^h$ are both piecewise abelian over each interval $[jh, (j+1)h]$ it suffices to prove the claim for $(s, t) = (0, \frac{n}{N})$.

To this end we combine (5.12) with the estimates of Lemma 5.6.5 using the Cauchy-Schwarz inequality and find

$$\begin{aligned} &\left\| \pi_m (S_\kappa(\mathbb{W}^h)_{0,nh} - S_\kappa(\mathbb{X}^h)_{0,nh}) \right\|_{(\mathbb{R}^d)^{\otimes m}} \Big|_{L^r} \\ &\leq C_{\kappa,q} \sum_{k=1}^m \frac{1}{k!} \sum_{\substack{i_1, \dots, i_k > 0 \\ i_1 + \dots + i_k = m}} \sum_{j=1}^k \left(\frac{n}{N}\right)^{\frac{1}{p}(m-i_j)} \\ &\quad \times \left(\left\| \mathbb{W}^h \right\|_{1/p\text{-HöL};[0,nh]}^{m-i_j} \Big|_{L^{2r}} + \left\| \mathbb{X}^h \right\|_{1/p\text{-HöL};[0,nh]}^{m-i_j} \Big|_{L^{2r}} \right) \left\| g_{i_j} - h_{i_j} \right\|_{L^{2r}} \\ &\leq C_{\kappa,p,r} \sum_{k=1}^m \frac{1}{k!} \sum_{\substack{i_1, \dots, i_k > 0 \\ i_1 + \dots + i_k = m}} \sum_{j=1}^k \left(\frac{n}{N}\right)^{\frac{1}{p}(m-i_j) + \frac{m}{2} - 1} \frac{\log N}{N}. \end{aligned}$$

The last line uses the integrability of the random quantities

$$\exp \left(\alpha \left\| \mathbb{W}^h \right\|_{1/p\text{-HöL};[0,1]}^2 \right), \quad \exp \left(\alpha \left\| \mathbb{X}^h \right\|_{1/p\text{-HöL};[0,1]}^2 \right)$$

for some $\alpha > 0$ in any given L^r -norm. We refer to Corollary 13.14 of [52] for the case of $\mathbb{W}^h$ and the corresponding argument for $\mathbb{X}^h$ is identical. Since

$nh \leq Nh = 1$, we have

$$\begin{aligned} & \left\| \pi_m (S_\kappa(\mathbb{W}^h)_{0,nh} - S_\kappa(\mathbb{X}^h)_{0,nh}) \right\|_{(\mathbb{R}^d)^{\otimes m}} \Big|_{L^r} \\ & \leq C_{\kappa,p,r} \sum_{k=1}^m \frac{1}{k!} \sum_{\substack{i_1, \dots, i_k > 0 \\ i_1 + \dots + i_k = m}} \sum_{j=1}^k \left(\frac{n}{N} \right)^{\frac{m}{2}-1} \frac{\log N}{N} \leq C_{\kappa,p,r} \left(\frac{n}{N} \right)^{\frac{m}{2}-1} \frac{\log N}{N}. \end{aligned}$$

The proof is complete. $\square$

We are now in a position to prove the main result of this section, namely the claimed lifted rough path $1/p$ -Hölder estimates.

Theorem 5.6.6. *Fix an integer $\kappa > 2$ and set $p \geq \kappa$ such that $\lfloor p \rfloor = \kappa$. Then for all $r \in [1, \infty)$ there exists a constant $C_{\kappa,p,r} > 0$ such that*

$$\left| \rho_{1/p\text{-Hölder}[0,1]} (S_\kappa(\mathbb{W}^h), S_\kappa(\mathbb{X}^h)) \right|_{L^r} \leq \frac{C_{\kappa,p,r}}{N^{1-\frac{2}{p}-\varepsilon}}$$

for all $\varepsilon > 0$.

PROOF. Fix $\theta \in [0, 1]$ and note that

$$\begin{aligned} |t - s|^{\frac{m}{2}-1} (N^{-1} \log N \wedge |t - s|) & \leq (N^{-1} \log N)^{1-\theta} |t - s|^{\frac{m}{2}-1+\theta} \\ & \leq (N^{-1} \log N)^{1-\theta} |t - s|^{\frac{m\theta}{2}}, \end{aligned}$$

since $\frac{m}{2}-1+\theta \geq \frac{m\theta}{2}$ for $m \geq 2$. Then combining the fact that $\pi_1(\mathbb{W}_{s,t}^h - \mathbb{X}_{s,t}^h) = 0$ with Proposition 5.6.1 guarantees that

$$\max_{m=1, \dots, \kappa} \left\| \pi_m (S_\kappa(\mathbb{W}^h)_{s,t} - S_\kappa(\mathbb{X}^h)_{s,t}) \right\|_{(\mathbb{R}^d)^{\otimes m}} \Big|_{L^r} \leq C_{\kappa,r} (N^{-1} \log N)^{1-\theta} |t - s|^{\frac{m\theta}{2}}.$$

Appealing to the rough path Kolmogorov regularity theorem of [52, Theorem A.13], it follows that for all $\gamma < \frac{\theta}{2}$,

$$\max_{m=1, \dots, \kappa} \left| \sup_{0 \leq s < t \leq 1} \frac{\left\| \pi_m (S_\kappa(\mathbb{W}^h)_{s,t} - S_\kappa(\mathbb{X}^h)_{s,t}) \right\|_{(\mathbb{R}^d)^{\otimes m}}}{|t - s|^{m\gamma}} \right|_{L^r} \leq C_{\kappa,r} \left(\frac{\log N}{N} \right)^{1-\theta}.$$

Therefore, so long as $p = \gamma^{-1} > \frac{2}{\theta}$, we have

$$\left| \rho_{1/p\text{-Hölder}[0,1]} (S_\kappa(\mathbb{W}^h), S_\kappa(\mathbb{X}^h)) \right|_{L^r} \leq C_{\kappa,p,r} (N^{-1} \log N)^{1-\theta}.$$

The condition $p > \frac{2}{\theta}$ is satisfied by taking $\theta = \frac{2}{p} + \varepsilon$ for some small constant $\varepsilon > 0$. Since this latter constant is arbitrary, we may drop the $\log N$ factor from the estimate and the result follows. $\square$

Remark 5.6.7. Although Theorem 5.6.6 requires that $p > 2$, as a quick heuristic check we note that if $p \in [1, 2)$ then the right-hand side of its bound explodes as $N = h^{-1} \rightarrow \infty$, (as was indicated by our previous calculations in Section 5.5). Moreover, if $p \in [2, 4)$ then, as predicted earlier, the convergence rate is worse than $O(N^{-1/2})$.

5.7. Coupling solutions of stochastic differential equations

In this section, we immediately use the inhomogeneous $1/p$ -Hölder estimate of Theorem 5.6.6 in order to derive Wasserstein convergence rates between the solutions of the previous RDEs (with drift) driven by $\mathbb{W}^h$ and $\mathbb{X}^h$ (or, equivalently, their rough path lifts). Recall that the solutions coincide with the approximation scheme $\{x_j^N\}_{j=1}^N$ and the output of the log-ODE method. By coupling the rough paths $\mathbb{W}^h, \mathbb{X}^h$, (or more precisely their lifts), the following result is an almost immediate corollary of the Lipschitz-continuity of the Itô map (Proposition 2.9.2). In turn, the following Corollary 5.7.1 establishes the main result of the chapter, Theorem 5.1.2.

Corollary 5.7.1. *Fix $\gamma > p \in (2, \infty)$ and $\beta > 1$. Let x denote the solution to the SDE (2.11) where we assume that $V_0 \in \text{Lip}$ and $V = \{V_k\}_{k=1}^d \subset \text{Lip}^\gamma$. Fix $\lambda \geq \|V_0\|_{\text{Lip}^\beta} \vee \|V\|_{\text{Lip}^\gamma}$ and let y be the unique solution to the following RDE driven by $\mathbb{X}^h$:*

$$(5.21) \quad dy_t = V(y_t) d\mathbb{X}_t^h + V_0(y_t) dt, \quad y_0 = x_0 \in \mathbb{R}^q.$$

Then for all $r \in [1, \infty)$ there exists a constant $C_{\beta, \gamma, \lambda, p, r} > 0$ such that for all $\varepsilon > 0$,

$$\left| \max_{j=1, \dots, N} \|x_{jh} - y_{jh}\|_{\mathbb{R}^q} \right|_{L^r} \leq \frac{C_{\beta, \gamma, \lambda, p, r}}{N^{1-\frac{2}{p}-\varepsilon}}.$$

PROOF. Let $\kappa \geq 2$ be the unique integer satisfying $[p] = \kappa$. Recalling Remark 5.4.7, we can rewrite the RDE (5.21) as

$$dy_t = V(y_t) dS_\kappa(\mathbb{X}^h)_t + V_0(y_t) dt, \quad y_0 = x_0 \in \mathbb{R}^q.$$

A corresponding statement holds for the pair $(\mathbb{W}^h, S_\kappa(\mathbb{W}^h))$; let z be the solution to the equivalent RDEs with drift:

$$dz_t = V(z_t) d\mathbb{W}_t^h + V_0(z_t) dt = V(z_t) dS_\kappa(\mathbb{W}^h)_t + V_0(z_t) dt, \quad z_0 = x_0.$$

From Theorem 2.10.4 and Proposition 5.4.4 we know

$$\left| \max_{j=1, \dots, N} \|x_{jh} - z_{jh}\|_{\mathbb{R}^q} \right|_{L^r} \leq \frac{C}{N},$$

and so it suffices to prove $\|y - z\|_\infty|_{L^r} \leq CN^{-1+\varepsilon+2/p}$.

To this end, we employ the RDE Lipschitz estimate of Proposition 2.9.2 to find that for each $\alpha > 0$,

$$\begin{aligned} & \|y - z\|_\infty \\ & \leq \rho_{p\text{-var};[0,1]}(y, z) \\ & \leq C_1 \rho_{p\text{-var};[0,1]}(S_\kappa(\mathbb{W}^h), S_\kappa(\mathbb{X}^h)) \left(1 + \|S_\kappa(\mathbb{W}^h)\|_{p\text{-var};[0,1]}^{[p]+1} + \|S_\kappa(\mathbb{X}^h)\|_{p\text{-var};[0,1]}^{[p]+1} \right) \\ (5.22) \quad & \times \exp(C_1 \{1 + N_{\alpha,p}(S_\kappa(\mathbb{W}^h), [0, 1]) + N_{\alpha,p}(S_\kappa(\mathbb{X}^h), [0, 1])\}) \end{aligned}$$

for some (deterministic) constant $C_1 = C_1(\alpha, \beta, \gamma, \lambda, p)$. We first bound the moments of the p -variation terms in (5.22).

By the Lipschitz-continuity of the rough path lift (Theorem 2.5.1) and the fact that $\mathbb{W}^h$ is the piecewise abelian approximation of enhanced Brownian

motion $\mathbb{W}$,

$$\begin{aligned} \|S_\kappa(\mathbb{W}^h)\|_{p\text{-var};[0,1]} &\leq \|S_\kappa(\mathbb{W}^h)\|_{(2+\delta)\text{-var};[0,1]} \leq C_{\delta,\kappa} \|\mathbb{W}^h\|_{(2+\delta)\text{-var};[0,1]} \\ (5.23) \qquad \qquad \qquad &\leq C_{\delta,\kappa} \|\mathbb{W}\|_{(2+\delta)\text{-var};[0,1]}, \end{aligned}$$

for arbitrary $0 < \delta < 1 \wedge (p - 2)$. The last quantity is integrable in L^r for all $r \in [1, \infty)$ ([52, Corollary 13.14]), with an identical statement holding for $S_\kappa(\mathbb{X}^h)$.

Next we establish the integrability of the exponential term in (5.22) by following the proofs of Theorems 10 and 13 in [11] along with [47, Corollary 5]. First let $q^{-1} + s^{-1} > 1$ for $q \in (2, 3 \wedge p)$. Using the results of [18, Remark 6.4] and [47, Corollary 2 and Remark 1], there exist constants $\alpha = \alpha(q, s) > 0$ and $c_1 = c_1(q, s) > 0$ such that the tail estimate:

$$\mathbb{P}(N_{\alpha,q}(\mathbb{W}, [0, 1]) > u) \leq \exp(-c_1 \alpha^{2/q} u^{2/s})$$

holds for all $u > 0$ and step-sizes $h > 0$. Since $\|\mathbb{W}^h\|_{q\text{-var};[s,t]} \leq \|\mathbb{W}\|_{q\text{-var};[s,t]}$, Lemma 2 of [47] gives $N_{\alpha,q}(\mathbb{W}^h, [0, 1]) \leq N_{\alpha,q}(\mathbb{W}, [0, 1])$, and so:

$$\mathbb{P}(N_{\alpha,q}(\mathbb{W}^h, [0, 1]) > u) \leq \mathbb{P}(N_{\alpha,q}(\mathbb{W}, [0, 1]) > u).$$

Then the Lipschitz-continuity of $S_\kappa(\cdot)$ and [47, Lemma 2] imply that

$$\mathbb{P}(N_{\alpha,q}(S_\kappa(\mathbb{W}^h), [0, 1]) > u) \leq \mathbb{P}(N_{\alpha,q}(\mathbb{W}^h, [0, 1]) > u) \leq \exp(-c_1 \alpha^{2/q} u^{2/s})$$

also holds for all $u > 0$, $h > 0$, and for a possibly smaller $\alpha > 0$ (as in the proof of Theorem 13 of [11]). As $p > q$, the interpolation result of Lemma 8.16 of [52] tells us that

$$\omega_1(s, t) := \|S_\kappa(\mathbb{W}^h)\|_{p\text{-var};[s,t]}^p \leq \omega_2(s, t) \cdot \sup_{s \leq u < v \leq t} \|S_\kappa(\mathbb{W}^h)_{u,v}\|_C^{p-q},$$

where $\omega_2(s, t) := \|S_\kappa(\mathbb{W}^h)\|_{q\text{-var};[s,t]}^q$. Thus taking $(u, v) \in \Delta_{[s,t]}$ such that $\omega_2(u, v) \leq \alpha$ gives $\omega_1(u, v) \leq \alpha^{p-q} \omega_2(u, v)$. Again using [47, Lemma 2], we

have

$$N_{\alpha^{p-q+1},p}(S_\kappa(\mathbb{W}^h), [s, t]) \leq N_{\alpha,q}(S_\kappa(\mathbb{W}^h), [s, t]).$$

If $\alpha \leq 1$ then $N_{\alpha,p}(S_\kappa(\mathbb{W}^h), [0, 1]) \leq N_{\alpha,q}(S_\kappa(\mathbb{W}^h), [0, 1])$ and so

$$(5.24) \quad \mathbb{P}(N_{\alpha,p}(S_\kappa(\mathbb{W}^h), [0, 1]) > u) \leq \exp(-c_1 \alpha^{2/q} u^{2/s}).$$

On the other hand, if $\alpha > 1$ then Lemma 3 of [47] gives

$$\begin{aligned} N_{\alpha,p}(S_\kappa(\mathbb{W}^h), [0, 1]) &\leq \alpha^{p-q} (1 + 2N_{\alpha^{p-q+1},p}(S_\kappa(\mathbb{W}^h), [0, 1])) \\ &\leq \alpha^{p-q} (1 + 2N_{\alpha,q}(S_\kappa(\mathbb{W}^h), [0, 1])), \end{aligned}$$

which in turn implies the tail estimate:

$$(5.25) \quad \mathbb{P}(N_{\alpha,p}(S_\kappa(\mathbb{W}^h), [0, 1]) > u) \leq c_2 \exp(-c_3 \alpha^{2/q+2(q-p)/s} u^{2/s})$$

for some constants $c_i = c_i(\alpha, q, s) > 0$, $i = 1, 2$. The tail estimates (5.24) and

(5.25), together with the fact that $s \in (1, 2)$ (so that $2/s > 1$), guarantee that

for all $r \in [1, \infty)$,

$$\sup_{h>0} |\exp(C_1 N_{\alpha,p}(S_\kappa(\mathbb{W}^h), [0, 1]))|_{L^r} \leq C_{\alpha,\kappa,r} < \infty.$$

By an identical argument, corresponding uniform estimates with $S_\kappa(\mathbb{X}^h)$ instead

of $S_\kappa(\mathbb{W}^h)$ can be established. Consequently for all $r \in [1, \infty)$,

$$\sup_{h>0} |\exp(C_1 \{1 + N_{\alpha,p}(S_\kappa(\mathbb{W}^h), [0, 1]) + N_{\alpha,p}(S_\kappa(\mathbb{X}^h), [0, 1])\})|_{L^r} \leq C_{\alpha,\kappa,r}.$$

In light of this uniform estimate along with that of (5.23), we apply the Cauchy-

Schwarz inequality to (5.22) to arrive at

$$|||y - z|||_\infty|_{L^r} \leq C |\rho_{p\text{-var};[0,1]}(y, z)|_{L^{2r}} \leq C |\rho_{1/p\text{-Höl};[0,1]}(y, z)|_{L^{2r}}.$$

The proof is then concluded by Theorem 5.6.6. $\square$

5.8. Possible extensions

We discuss the possibility of two extensions of Theorem 5.1.2.

Fractional Brownian motion. A natural question to ask is whether Theorem 5.1.2 can be generalised to the case where the driving Brownian motion of the SDE (2.11) is replaced with fractional Brownian motion with Hurst index $H \in (\frac{1}{3}, \frac{1}{2})$, (thus turning (2.11) into a strict RDE). Recall that Brownian motion corresponds to $H = \frac{1}{2}$. The log-ODE method remains perfectly valid for this RDE, as does our piecewise abelian interpretation since the Lévy area of fractional Brownian motion is well-defined for $H > \frac{1}{4}$. (This is not the case for $H \leq \frac{1}{4}$ [92, §1]; even the lift of the standard piecewise linear interpolation does not converge in p -variation under L^1 for $p > H^{-1} \geq 4$ [81, §4.5]).

Instead, the main obstacle of this extension is the need to reproduce the main coupling result of Theorem 6.2.1. This is problematic because the proof of Davie in [33] inherently relies upon the independence of increments in the Brownian motion case in order to perform the inductive coupling over finer dyadic intervals of $[0, 1]$. Similarly, the original Komlós-Major-Tusnády theorem and the modern extensions of Zaitsev also critically rely upon the independence of increments of the random walk to be coupled with a Gaussian approximation. Since this is no longer the case when $H \neq \frac{1}{2}$, the author sees no way around this at present.

In the even more extreme case of $H \in (\frac{1}{4}, \frac{1}{3}]$, for the log-ODE method to converge we require the first $\lfloor H^{-1} \rfloor = 3$ levels of the log-signature. The task of coupling these higher order terms is difficult, even in the Brownian case (as we now discuss).

Higher order log-ODE approximations. This chapter has dealt exclusively with the log-ODE method at level $m = 2$; that is, the log-signature of the Brownian motion has been truncated to its first two levels. Thus we have

only needed to couple the Lévy area increments with a Gaussian approximation, (conditional on the underlying Brownian increments). A natural extension would be to couple the higher order terms of the log-signature with Gaussian approximations, thus enabling us to use a higher order version of the log-ODE method. If this coupling were successful then we could expect a better convergence rate in the Wasserstein metric for our resultant approximation scheme. Indeed, a truncation of the log-signature to level m produces a log-ODE scheme with strong convergence in L^p of order $O(N^{-m/2})$ ([61, Theorem 4.1]).

One difficulty is that one would need to extend the proof of Theorem 6.2.1 to couple not only the Lévy area increments, but also the third order iterated integrals, (again conditional on the underlying Brownian increments). This may not be possible without violating the matrix non-degeneracy conditions needed for Davie's coupling proof ([33, Theorem 1]); it is critical that the matrix H_E defined in (6.8) of the next chapter is strictly positive-definite.

A more pronounced obstacle is the task of establishing the necessary lifted rough path estimates of Section 5.6. The difference of the iterated Baker-Campbell-Hausdorff expansions of each piecewise abelian rough path would be significantly more complex because levels 2 and 3 of the group increments would not be equal. Again we see no solution at present.

CHAPTER 6

Polynomial perturbations of Gaussian distributions and the coupling construction

In this chapter we prove the previously claimed coupling result of Theorem 6.2.1 concerning the random walks composed of Lévy area increments $A^{(j)}$ and the quadratic Gaussian approximations $B^{(j)}$. Unlike previous chapters, there is no mention of rough path theory. Instead, the chapter presents purely probabilistic results in order to establish the main theorem. Hopefully these results are of interest in their own right. We begin in the next section by presenting a short overview of polynomial perturbations of Gaussian measures. Utilising the main results of the recent preprint [35], we then prove an extension of [33, Corollary 3] from quartics to higher order polynomial perturbations. The final section of the chapter then uses this extension together with a central limit theorem to generalise the original coupling proof of [33, Theorem 1] from a bound in the 2nd Wasserstein metric $\mathcal{W}_2$ to $\mathcal{W}_p$ for general $p \in [1, \infty)$, thereby establishing Theorem 6.2.1.

Notation. Before proceeding further we establish some special notation for the chapter. Given a covariance matrix $\Sigma \in \mathbb{R}^{d \times d}$, let ϕ_Σ denote the density function of $N(0, \Sigma)$:

$$\phi_\Sigma(x) = \frac{1}{(2\pi)^{d/2} (\det \Sigma)^{1/2}} \exp \left(-\frac{1}{2} x^t \Sigma^{-1} x \right).$$

In the case of $\Sigma = I_d$ (the identity matrix) we simply write $\phi(x) = \phi_{I_d}(x)$. Given an arbitrary probability measure μ on $\mathbb{R}^d$, we will commit a slight abuse of notation by using $\mathcal{W}_p(\mu, \phi_\Sigma)$ to denote the Wasserstein distance $\mathcal{W}_p(\mu, \mathcal{L}(Z))$ where $Z \sim N(0, \Sigma)$.

Let P denote the space of all real-valued polynomials on $\mathbb{R}^d$ (which we will write as $P(\mathbb{R}^d)$ when we want to specify the dimension d) and define the subspace

$$P_\Sigma := \left\{ p \in P : \int_{\mathbb{R}^d} p(x) \phi_\Sigma(x) dx = 0 \right\}.$$

Let P^d denote the space of $\mathbb{R}^d$ -valued polynomial functions on $\mathbb{R}^d$. To be precise, $p \in P^d$ means $p(x) = (p^1(x), \dots, p^d(x))$ for some $p^j \in P$. Given a polynomial p^j , let $\deg(p^j)$ denote the highest degree of its terms. Similarly, for a polynomial function $p = (p^1, \dots, p^d) \in P^d$, set $\deg(p) := \max_{j=1, \dots, d} \deg(p^j)$. Lastly, given the dimension d , we set $d_1 := \frac{d}{2}(d+1)$ and $d_2 := \frac{d}{2}(d-1)$ to avoid cumbersome sub/superscript notation.

6.1. Polynomial perturbations of Gaussian distributions

In this section we consider signed measures on $\mathbb{R}^d$ with a density given by a polynomial perturbation of the standard Gaussian distribution $\phi(x) = (2\pi)^{-d/2} e^{-|x|^2/2}$. Our main aim is to prove that if such a signed measure is close to a probability measure μ (in the form of an estimate similar to that of Lemma 4.1.7), then we can expect the distance $\mathcal{W}_p(\mu, \phi)$ to be bounded by the magnitude of the perturbation. This is the content of the following proposition.

Proposition 6.1.1. *Fix $n \in \mathbb{N}$ and $\varepsilon \in (0, 1)$. Suppose $\{S_j\}_{j=1}^n \subset P$ is a sequence of polynomials with the absolute values of its coefficients bounded by a universal constant $M > 0$ and set $s := \max_{j=1, \dots, n} \deg(S_j)$. Fix some integer*

$1 \leq n_0 \leq n$ and let $\nu_{\varepsilon,n}$ denote the signed measure on $\mathbb{R}^d$ with density

$$\phi(y) \left(1 + \sum_{j=n_0}^n \varepsilon^j S_j(y) \right).$$

If μ is a probability measure on $\mathbb{R}^d$ such that

$$(6.1) \quad \int_{\mathbb{R}^d} (1 + |y|)^p d|\mu - \nu_{\varepsilon,n}|(y) \leq \delta,$$

then for all $p \in [1, \infty)$ we have

$$\mathcal{W}_p(\mu, \phi) \leq C_{d,M,n,p,s} \left(\varepsilon^{n_0} + \delta^{1/p} + \varepsilon^{\frac{n+1}{p}} \right).$$

The proposition is a strengthened version of [33, Corollary 3]. Instead of dealing with only quartic perturbations, Proposition 6.1.1 can handle polynomials of arbitrarily high order. The proof was suggested to the author by Professor Davie in private communications and the argument relies upon the results in his preprint [35]. Thus before presenting its proof we summarise the work of the latter paper.

Remark 6.1.2. To be precise, the original polynomial perturbation result of Davie actually considered quartic perturbations of ϕ_Σ for some arbitrary covariance matrix $\Sigma \in \mathbb{R}^{d \times d}$. One may ask whether Proposition 6.1.1 can also be generalised to the case of arbitrary ϕ_Σ rather than just $\Sigma = I_d$. The author conjectures that this is possible since the technical tools from [35] remain valid. However, the precise dependencies of the norms $\|\Sigma\|$ and $\|\Sigma\|^{-1}$ in the bound of $\mathcal{W}_p(\mu, \phi_\Sigma)$ are complicated and it is a non-trivial task to keep track of these quantities throughout the proof. Fortunately we only need the case of $\Sigma = I_d$ for the coupling arguments of this thesis.

We now summarise the main contributions of [35] and begin by characterising the subspace $P_\Sigma \subset P$ as follows.

Lemma 6.1.3. *Define the linear mapping $\mathcal{L}_\Sigma : P^d \rightarrow P$ by $\mathcal{L}_\Sigma p(x) = \nabla \cdot p(x) - x^t \Sigma^{-1} p(x)$. The space P_Σ is precisely the range of $\mathcal{L}_\Sigma$. Moreover, any element $p \in P_\Sigma$ can be expressed as $\mathcal{L}_\Sigma \nabla u$ for some $u \in P$ (and the u is unique up to an additive constant).*

The proof is by induction and the divergence theorem (see [34, Lemma 1]). One consequence of the lemma is that if we define P_G^d to be the set of $p \in P^d$ of the form $p = \nabla u$ with $u \in P$, we have that $\mathcal{L}_\Sigma$ is bijective from $P_G^d \rightarrow P_\Sigma$ and we can define the inverse linear mapping $\mathcal{L}_\Sigma^{-1} : P_\Sigma \rightarrow P_G^d$.

Next, suppose we have a sequence of polynomial functions $\{p_j\}_{j=1}^n \subset P_G^d$. For each $\varepsilon > 0$, we define the polynomial perturbation mapping $\rho_\varepsilon : \mathbb{R}^d \rightarrow \mathbb{R}^d$ by

$$(6.2) \quad \rho_\varepsilon(x) = x + \sum_{j=1}^n \varepsilon^j p_j(x).$$

We are interested in the distribution of $\rho_\varepsilon(Z)$ for small ε , where $Z \sim N(0, \Sigma)$. If we assume that ρ_ε is bijective then this distribution has a density given by

$$f_\varepsilon(y) = \det(D\rho_\varepsilon^{-1}(y)) \phi_\Sigma(\rho_\varepsilon^{-1}(y)).$$

As Davie notes in the introduction of [35], in general bijectivity will only hold on some bounded region of $\mathbb{R}^d$ (which will be large if ε is small). It actually turns out that bijectivity is a sufficient condition for (6.2) to imply the following asymptotic expansion of the density:

$$(6.3) \quad f_\varepsilon(y) = \phi_\Sigma(y) \left(1 + \sum_{j=1}^{\infty} \varepsilon^j S_j(y) \right),$$

for some sequence of polynomials $\{S_j\}_{j=1}^{\infty} \subset P_\Sigma$.

In fact, we can explicitly construct the sequence $\{S_j\}_{j=1}^{\infty}$ from the polynomial functions $\{p_j\}_{j=1}^n$ via a bijection. We introduce the notation $\mathcal{P}$ for the set of all sequences $(u_1, u_2, \dots)$ with $u_j \in P$ with similar definitions for $\mathcal{P}^d, \mathcal{P}_\Sigma, \mathcal{P}_G^d$.

By using the inverse linear mapping $\mathcal{L}_\Sigma^{-1} : P_\Sigma \rightarrow P_G^d$, Davie inductively constructs a bijection $\mathcal{S}_\Sigma : \mathcal{P}_G^d \rightarrow \mathcal{P}_\Sigma$ such that

$$\mathcal{S}_\Sigma(p_1, \dots, p_n, 0, 0, \dots) = (S_1, S_2, \dots).$$

Since each S_k is dependent only on $p_1, \dots, p_k$, this can be rewritten in the more succinct form of the truncated mapping: $\mathcal{S}_\Sigma^{(n)}(p_1, \dots, p_n) = (S_1, \dots, S_n)$. For the explicit definition of $\mathcal{S}_\Sigma$ we refer to [35, Lemma 2]. We are now in a position to state the main result of [35] in a simplified form for the special case of $\Sigma = I_d$.

Proposition 6.1.4. [35, Proposition 1] *Let $\{S_j\}_{j=1}^n \subset P_I$ and define the corresponding sequence $\{p_j\}_{j=1}^n \subset P_G^d$ via the truncated bijection*

$$\mathcal{S}_I^{(n)}(p_1, \dots, p_n) = (S_1, \dots, S_n).$$

Set R to be an upper bound on the absolute values of the coefficients of $p_1, \dots, p_n$. Using these sequences, define the mapping ρ_ε as in (6.2) and let $\nu_{\varepsilon,n}$ be the signed measure on $\mathbb{R}^d$ with density (6.3). Finally, let μ_ε be the law of $\rho_\varepsilon(Z)$ where $Z \sim N(0, I)$. Then for all $p \in [1, \infty)$ we have

$$\int_{\mathbb{R}^d} (1 + |y|)^p d|\mu_\varepsilon - \nu_{\varepsilon,n}|(y) \leq C_p \varepsilon^{n+1},$$

for some constant $C_p > 0$ depending only on d, n, R and the maximum degree of $p_1, \dots, p_n$.

Armed with the latter result, we are now able to present the proof of Proposition 6.1.1.

PROOF OF PROPOSITION 6.1.1. Following the same technique in [33, Corollary 2], we first show that we may assume that $S_j \in P_\Sigma$ for all $j \in \{n_0, \dots, n\}$.

Indeed, let

$$\beta_j := \varepsilon^j \int_{\mathbb{R}^d} S_j(y) \phi(y) dy \text{ and } \beta := \sum_{j=n_0}^n \beta_j.$$

Then we have

$$\begin{aligned}
|\beta| &= \left| \sum_{j=n_0}^n \varepsilon^j \int_{\mathbb{R}^d} S_j(y) \phi(y) dy \right| \\
&= \left| \int_{\mathbb{R}^d} \phi(y) \left(1 + \sum_{j=n_0}^n \varepsilon^j S_j(y) \right) dy - \int_{\mathbb{R}^d} \mu_\varepsilon(y) dy \right| \\
&= \left| \int_{\mathbb{R}^d} (\nu_{\varepsilon,n} - \mu_\varepsilon)(y) dy \right| \\
&\leq \int_{\mathbb{R}^d} d|\nu_{\varepsilon,n} - \mu_\varepsilon|(y) \leq \int_{\mathbb{R}^d} (1 + |y|)^p d|\nu_{\varepsilon,n} - \mu_\varepsilon|(y) \leq \delta.
\end{aligned}$$

Next, define a new signed measure $\tilde{\nu}_{\varepsilon,n}$ on $\mathbb{R}^d$ with density

$$\phi(y) \left(1 + \sum_{j=n_0}^n (\varepsilon^j S_j(y) - \beta_j) \right) = \phi(y) \left(1 + \sum_{j=n_0}^n \varepsilon^j S_j(y) \right) - \beta \phi(y).$$

Consequently,

$$\begin{aligned}
\int_{\mathbb{R}^d} (1 + |y|)^p d|\tilde{\nu}_{\varepsilon,n} - \mu_\varepsilon|(y) &\leq \int_{\mathbb{R}^d} (1 + |y|)^p \{d|\nu_{\varepsilon,n} - \mu_\varepsilon|(y) + |\beta| \phi(y) dy\} \\
&\leq \delta + |\beta| \int_{\mathbb{R}^d} (1 + |y|)^p \phi(y) dy \leq C_p \delta.
\end{aligned}$$

Therefore, by replacing each S_j by $S_j - \beta_j$ we may assume that $S_j \in P_\Sigma$.

Since we can assume each $S_j \in P_I$, we can apply the bijection $\mathcal{S}_I$ of [35, Lemma 2] to find a corresponding sequence $\{p_j\}_{j=1}^n \subset P_G^d$ such that

$$\mathcal{S}_I^{(n)}(p_1, \dots, p_n) = (S_1, \dots, S_n).$$

As stated in the proof of [33, Corollary 3], if $g_0 = \mathcal{L}_I p_0$ for $p_0 \in P^d$, $g_0 \in P$, then $\deg(p_0) \leq \deg(g_0)$ and the absolute values of the coefficients of p_0 are bounded by those of g_0 up to a universal multiplicative constant. Thus by the recursive construction of the map $\mathcal{S}_I$ using $\mathcal{L}_I^{-1}$ in [35], we can bound the absolute values of the coefficients of $p_1, \dots, p_n$ by some function of those of $S_1, \dots, S_n$. Similarly, the degree of each p_j can be bounded by a function of $s = \max_j \deg(S_j)$. A simple consequence is the bound $|p_j(Z)|_{L^p} \leq C_{d,M,p,s}$

for each j , where $Z \sim N(0, I)$. Taking the trivial coupling of $\mathcal{L}(Z)$ and μ_ε , specifically $(Z, \rho_\varepsilon(Z))$, this leads to the estimate

$$\begin{aligned} \mathcal{W}_p(\mu_\varepsilon, \phi) &\leq |Z - \rho_\varepsilon(Z)|_{L^p} \leq \sum_{j=n_0}^n \varepsilon^j |p_j(Z)|_{L^p} \\ (6.4) \qquad \qquad \qquad &\leq C_{d,M,p,s} \sum_{j=n_0}^n \varepsilon^j = C_{d,M,n,p,s} \varepsilon^{n_0}. \end{aligned}$$

As before, define the polynomial perturbation ρ_ε by (6.2) and let μ_ε be the law of $\rho_\varepsilon(Z)$ for $Z \sim N(0, I)$. Then Proposition 6.1.4 guarantees that for all $p \in [1, \infty)$,

$$\int_{\mathbb{R}^d} (1 + |y|)^p d|\mu_\varepsilon - \nu_{\varepsilon,n}|(y) \leq C_{d,M,n,p,s} \varepsilon^{n+1}.$$

Combined with (6.1), this estimate gives

$$\int_{\mathbb{R}^d} (1 + |y|)^p d|\mu - \mu_\varepsilon|(y) \leq \delta + C_{d,M,n,p,s} \varepsilon^{n+1},$$

and so Lemma 4.1.7 yields the Wasserstein estimate

$$(6.5) \qquad \mathcal{W}_p(\mu, \mu_\varepsilon) \leq C_{d,M,n,p,s} (\delta + \varepsilon^{n+1})^{1/p} \leq C_{d,M,n,p,s} \left(\delta^{1/p} + \varepsilon^{\frac{n+1}{p}} \right).$$

Finally, combining (6.4) and (6.5) via the triangle inequality, we conclude that

$$\mathcal{W}_p(\mu, \phi) \leq \mathcal{W}_p(\mu, \mu_\varepsilon) + \mathcal{W}_p(\mu_\varepsilon, \phi) \leq C_{d,M,n,p,s} \left(\varepsilon^{n_0} + \delta^{1/p} + \varepsilon^{\frac{n+1}{p}} \right).$$

The proof is complete. □

Remark 6.1.5. The proof of Proposition 6.1.1 reveals the reason why we insist on bounding the quantity $\int_{\mathbb{R}^d} (1 + |y|)^p d|\mu - \nu|(y)$ rather than the simpler integral

$$\int_{\mathbb{R}^d} |y|^p d|\mu - \nu|(y).$$

At first glance the latter quantity is all that is needed for Lemma 4.1.7 to bound $\mathcal{W}_p(\mu, \nu)$. However, in order to generalise the proof to polynomials not

contained in P_Σ we use the simple inequality

$$\int_{\mathbb{R}^d} d|\nu_{n,\varepsilon} - \mu|(y) \leq \int_{\mathbb{R}^d} (1 + |y|)^p d|\nu_{n,\varepsilon} - \mu|(y).$$

This bound does not remain true when we replace the right-hand side with $\int_{\mathbb{R}^d} |y|^p d|\nu_{n,\varepsilon} - \mu|(y)$. For instance, consider the case when the supports of $\nu_{n,\varepsilon}$ and μ are contained within the open unit ball around the origin of $\mathbb{R}^d$.

6.2. Main coupling theorem

We state the main coupling theorem of the chapter. Then, after establishing some notation and performing a linear transformation, we show that it suffices to prove the simpler statement of Proposition 6.2.3.

Theorem 6.2.1. *For every $p \in [1, \infty)$, there exists a constant $C_p > 0$ such that the following holds: given $N \in \mathbb{N}$, we can construct a coupling between the i.i.d. sequence $\{A^{(j)}\}_{j=0}^{N-1}$ and i.i.d. $\{B^{(j)}\}_{j=0}^{N-1}$ random variables, with each pair $(A^{(j)}, B^{(j)})$ defined by (4.5.1) and (4.15) using a common Brownian increment $W^{(j)} \sim N(0, N^{-1}I_d)$, such that*

$$\max_{r=1, \dots, N} \left\| \left\| \sum_{j=0}^{r-1} (A^{(j)} - B^{(j)}) \right\|_{(\mathbb{R}^d)^{\otimes 2}} \right\|_{L^p} \leq C_p \frac{\log N}{N}.$$

Before presenting the proof in the Section 6.4, we first set up some notation, then consider a version of the central limit theorem in the next section. We fix a sequence of independent Brownian increments $\{W^{(j)}\}_{j=0}^{N-1}$ with $W^{(j)} \sim N(0, N^{-1}I_d)$, and let $\mathcal{G}$ denote the σ -algebra generated by these variables.

For each $r \in \{0, 1, \dots, N-1\}$ define the random vector $X^{(r)} \in \mathbb{R}^{d_1}$ by

$$\begin{aligned} X_k^{(r)} &:= \sqrt{12N} \zeta_k^{(r)} \text{ for } 1 \leq k \leq d; \\ X_{\frac{k}{2}(2d-k-1)+l}^{(r)} &:= \sqrt{12N^2} K_{kl}^{(r)} \text{ for } 1 \leq k < l \leq d. \end{aligned}$$

Then, conditional on $\mathcal{G}$, $X^{(r)}$ has mean zero and covariance matrix I_{d_2} . We can then write $A^{(r)} = N^{-1}G_r X^{(r)}$, where G_r is a $d_2 \times d_1$ matrix defined in terms of the $W^{(j)}$. Specifically,

$$G_r = \frac{1}{\sqrt{12}} \left(M_r \mid I_{d_2} \right),$$

where M_r is the $d_2 \times d$ matrix defined by the rows

$$(M_r)_{\frac{k}{2}(2d-k-1)+(l-d)} = \sqrt{N} \left(W_l^{(r)} e_k - W_k^{(r)} e_l \right).$$

This makes M_r have the form:

$$M_r = \sqrt{N} \begin{pmatrix} W_2^{(r)} & -W_1^{(r)} & 0 & \cdots & 0 & 0 \\ W_3^{(r)} & 0 & -W_1^{(r)} & \cdots & 0 & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & 0 & \cdots & W_d^{(r)} & -W_{d-1}^{(r)} \end{pmatrix}.$$

In the same way we can write $B^{(r)} = N^{-1}G_r \tilde{X}^{(r)}$, where $\tilde{X}^{(r)}$ has the normal distribution $N(0, I_{d_1})$.

By scaling we can see that to complete the proof it is sufficient to construct a coupling of the random walks composed of the vectors $X^{(r)}$ and $\tilde{X}^{(r)}$, conditional on $\mathcal{G}$, such that

$$(6.6) \quad \max_{r=1, \dots, N} \left\| \left\| \sum_{j=0}^{r-1} G_j \left(X^{(j)} - \tilde{X}^{(j)} \right) \right\|_{(\mathbb{R}^d)^{\otimes 2}} \right\|_{L^p} \leq C_p \log N$$

for some constant $C_p > 0$. To this end, we first note that without loss of generality we may assume $N = h^{-1} = 2^m$ for some $m \in \mathbb{N}$. We define a dyadic set to be a subset $E \subseteq \{0, 1, \dots, 2^m - 1\}$ of the form

$$E = \{k2^n, k2^n + 1, \dots, (k+1)2^n - 1\},$$

for some integers $n \in \{0, 1, \dots, m\}$ and $k \in \{0, 1, \dots, 2^{m-n} - 1\}$. Since a consecutive set $F \subset \{0, 1, \dots, 2^m - 1\}$ can be expressed as the disjoint union of

at most $\log_2 N$ dyadic subsets $E_1, \dots, E_k$ of different sizes, we need only prove

$$(6.7) \quad \mathbb{E} \left(\left\| \sum_{r \in E} G_r \left(X^{(r)} - \tilde{X}^{(r)} \right) \right\|_{(\mathbb{R}^d)^{\otimes 2}}^p \right) \leq C_p$$

for all dyadic sets E in order to establish (6.6).

Next, for each dyadic set E of size 2^n let us define the matrix

$$(6.8) \quad H_E := 2^{-n} \sum_{r \in E} G_r G_r^t$$

along with the random variables

$$Y_E := 2^{-n/2} \sum_{r \in E} G_r X^{(r)} \text{ and } Z_E := 2^{-n/2} \sum_{r \in E} G_r \tilde{X}^{(r)}.$$

Since, conditional on $\mathcal{G}$, the random variables $\{A^{(j)}\}_{j=0}^{N-1}$ are independent, H_E is the (conditional) covariance matrix of Y_E . Similarly, H_E is also the (conditional) covariance matrix of Z_E . Note that H_E^{-1} is well-defined since each product $G_r G_r^t$ is a positive-definite symmetric matrix. Indeed, block matrix multiplication confirms that

$$G_r G_r^t = \frac{1}{12} (I_{d_2} + M_r M_r^t).$$

It also follows that the eigenvalues of $G_r G_r^t$ are bounded below by $\frac{1}{12}$ and so $\|(G_r G_r^t)^{-1}\| \leq 12$. Similarly, $\|H_E^{-1}\| \leq 12$ for all dyadic E . Moreover, since $\sqrt{N}W_k^{(r)} \sim N(0, 1)$, certainly $\mathbb{E}(\|G_r\|^p) \leq C_p$ for all $p \geq 1$. In fact, $\|G_r\|^2$ possesses exponential tails (a property which we will exploit later in the coupling construction of Section 6.4).

Lemma 6.2.2. *For all $\alpha \in (0, \frac{1}{48d})$ there exists a constant $C_\alpha > 0$ such that*

$$\max_{r=0,1,\dots,N-1} \mathbb{E} \left(e^{\alpha \|G_r\|^2} \right) \leq C_\alpha$$

Similarly, $\mathbb{E} (e^{\alpha \|H_E\|}) \leq C_\alpha$ for all consecutive sets E .

PROOF. Recall that the matrix norm $\|M_r\|$ is bounded by the l_2 -norm of its entries $\{M_r(i, j)\}_{i,j}$. Hence,

$$\|M_r\|^2 \leq \sum_{i,j} |M_r(i, j)|^2 = 2h^{-1} \sum_{k=1}^d |W_k^{(r)}|^2,$$

and so for all $\alpha \in (0, \frac{1}{4d})$ we have

$$\mathbb{E} \left(e^{\alpha \|M_r\|^2} \right) \leq \prod_{k=1}^d \mathbb{E} \left(e^{2\alpha h^{-1} |W_k^{(r)}|^2} \right) = \mathbb{E} \left(e^{2\alpha d h^{-1} |W_k^{(r)}|^2} \right) = C_\alpha < \infty.$$

Thus for small enough $\alpha \in (0, \frac{1}{48d})$,

$$\mathbb{E} \left(e^{\alpha \|G_r\|^2} \right) \leq \mathbb{E} \left(e^{\frac{\alpha}{12} (\|M_r\|+1)^2} \right) \leq e^{\frac{\alpha}{6}} \mathbb{E} \left(e^{\frac{\alpha}{6} \|M_r\|^2} \right) \leq C_\alpha.$$

The second statement then follows:

$$\begin{aligned} \mathbb{E} \left(e^{\alpha \|H_E\|^2} \right) &\leq \mathbb{E} \exp \left(\alpha 2^{-n} \sum_{r \in E} \|G_r\|^2 \right) = \prod_{r \in E} \mathbb{E} \left(e^{\alpha 2^{-n} \|G_r\|^2} \right) \\ &= \mathbb{E} \left(e^{\alpha \|G_r\|^2} \right) = C_\alpha. \end{aligned}$$

The proof is complete. $\square$

The aim is to prove the following proposition, from which (6.7) follows immediately, thereby establishing Theorem 6.2.1.

Proposition 6.2.3. *With the notation above, for every $p \in [1, \infty)$ there exists a constant $C_p > 0$ and a coupling of $\{Y_E\}_{E \subseteq E_0}$ and $\{Z_E\}_{E \subseteq E_0}$, conditional on $\mathcal{G}$, such that*

$$\left| \|Y_E - Z_E\|_{(\mathbb{R}^d)^{\otimes 2}} \right|_{L^p} \leq C_p 2^{-n/2}$$

for every dyadic set E of size 2^n , where $n \leq m = \log_2 N$.

We copy the coupling construction of [33, Theorem 1] except for modifications which are needed to establish general L^p -estimates. This is because the original result only established the coupling for $p \in [1, 4)$. In fact, we can follow the argument of [33] verbatim to establish Proposition 6.2.3 for this range of

p . Unfortunately for Theorem 6.2.1 to be useful in a rough path context we need the coupling to hold in L^p for arbitrary large $p \in [1, \infty)$. Thus we require the higher order polynomial perturbation result of Proposition 6.1.1 along with Lemma 6.3.1 of the next section. Hence we postpone the proof of Proposition 6.2.3 to Section 6.4.

Remark 6.2.4. It is a common practice in Gaussian rough path theory to use Wiener-Itô chaos to establish general L^p -bounds from the $p = 2$ case ([52, Theorem D.8]). Certainly each $B^{(j)}$ increment lives in the second inhomogeneous Wiener chaos (being a quadratic polynomial of Gaussian random variables), and it is well-known that Lévy area also lives in a (possibly different) second Wiener chaos [52, Proposition 15.19]. One could ask whether we could use this theory to immediately get all L^p -estimates from the $p = 2 \in [1, 4)$ case. Alas our coupling argument does not necessarily guarantee that the random variables $\{A^{(j)}\}_{j=0}^{N-1}$ and $\{B^{(j)}\}_{j=0}^{N-1}$ belong to the same Wiener chaos and so we cannot apply the theory to their difference.

6.3. A central limit theorem

In light of the previous polynomial perturbation extension of Proposition 6.1.1, we need a modified version of [33, Lemma 4] for the proof of Proposition 6.2.3 in the next section. This is contained in the following lemma. It can be viewed as a version of the central limit theorem, stating that the density of Y_E is close to the (Gaussian) density of Z_E as the size of the dyadic set E increases.

Lemma 6.3.1. *Let E be a dyadic set of size 2^n and let f_E be the density function of Y_E , conditional on $\mathcal{G}$. Fix $\eta \in (0, \frac{1}{20})$ and an integer $\kappa \geq 2$. Then, provided that $\|G_r\| \leq 2^{n\eta}$ and $\|(G_r G_r^t)^{-1}\| \leq 2^{2n\eta}$, there exists a constant $C_\kappa > 0$*

such that the following holds: for each $r \in E$ we have

$$\left| f_E - \phi_{H_E} \left(1 + \sum_{k=2}^{\kappa} 2^{nk(2\eta-\frac{1}{2})} S_{E,2k} \right) \right| (v) \leq C_{\kappa} 2^{n\kappa(10\eta-\frac{1}{2})} \phi_{H_E}(v)$$

for all $|v| \leq 2^{n\eta}$. Here, each $S_{E,2k} \in P(\mathbb{R}^{d_2})$ is respectively a polynomial of degree $2k$ with coefficients whose absolute values are bounded by some universal constant $C > 0$ independent of E and k .

Note the original [33, Lemma 4] can be recovered by setting $\kappa = 2$.

PROOF. The bounds on G_r imply that $\|H_E\| \leq 2^{2n\eta}$ and $\|H_E^{-1}\| \leq 2^{2n\eta}$.

Let ψ be the characteristic function of $X^{(r)}$ (which is independent of r). An explicit expression for ψ can be found in [117]. Note that ψ is real-valued and even on $\mathbb{R}^{d_1}$ and extends to a complex-analytic function of a strip $\{x + iy : x, y \in \mathbb{R}^{d_1}, |y| < a\}$ for some $a > 0$. In a neighbourhood of 0 in $\mathbb{C}^{d_1}$, $\log \psi$ has a convergent expansion

$$\log \psi(z) = -\frac{1}{2} |z|^2 + c_4(z) + c_6(z) + \dots,$$

where $c_k(z)$ is a homogeneous polynomial of degree k satisfying $|c_k(z)| \leq (C|z|)^k$ for even $k \geq 4$. Thus,

$$\psi(z) = \exp \left(-\frac{1}{2} |z|^2 + \chi(z) \right) \text{ where } \chi(z) := \sum_{k=2}^{\infty} c_{2k}(z).$$

From this it follows that there exists some $\delta > 0$ such that

$$(6.9) \quad \text{if } x, y \in \mathbb{R} \text{ with } 2|y| \leq |x| < \delta \text{ then } |\psi(x + iy)| \leq e^{-\frac{1}{6}|x|^2}.$$

Using the decay of $\psi(z)$ as $x = \operatorname{Re}(z) \rightarrow \infty$ and the fact that $|\psi(x)| < 1$ for $0 \neq x \in \mathbb{R}^{d_1}$, we can find $\gamma \in (0, 1)$ and $\delta' > 0$ so that

$$(6.10) \quad \begin{array}{l} \text{if } x, y \in \mathbb{R}^{d_1} \text{ with } |x| \geq \delta \text{ and } |y| \leq \delta' \\ \text{then } |\psi(x + iy)| \leq \min(\gamma, C|x|^{-1}). \end{array}$$

Now let Ψ be the characteristic function of Y_E ; $\Psi(u) = \prod_{r \in E} \psi(2^{-n/2} G_r^t u)$.

Recall that Ψ is the Fourier transform of the density f_E of Y_E , and taking the inverse Fourier transform we obtain the expression

$$f_E(v) = (2\pi)^{-d_2/2} \int_{\mathbb{R}^{d_2}} e^{-iu^t v} \Psi(u) du.$$

By translating the subspace of integration in $\mathbb{C}^{d_2}$ by $-iH_E^{-1}v$ we can rewrite this as

$$\begin{aligned} f_E(v) &= (2\pi)^{-d_2/2} e^{-v^t H_E^{-1}v} \int_{\mathbb{R}^{d_2}} e^{-iu^t v} \Psi(u - iH_E^{-1}v) du \\ &= (\det H_E)^{1/2} \phi_{H_E}(v) \int_{\mathbb{R}^{d_2}} e^{-iu^t v} \Psi(u - iH_E^{-1}v) du. \end{aligned}$$

Note that

$$(6.11) \quad \Psi(u - iH_E^{-1}v) = \prod_{r \in E} \psi(2^{-n/2} G_r^t u - i2^{-n/2} G_r^t H_E^{-1}v).$$

If $|u| \geq 2^{4n\eta+1}$ then using (6.9) and (6.10) we see that each term in the product is bounded by either $\min(\gamma, C2^{n(\eta+1/2)} |u|^{-1})$ or $\exp(-\frac{1}{6}2^{-n(1+2\eta)} |u|^2)$. Consequently, the product (6.11) is bounded by

$$|\Psi(u - iH_E^{-1}v)| \leq \min(\gamma, C2^{n(\eta+1/2)} |u|^{-1})^{2^n} + \exp\left(-\frac{1}{6}2^{-2n\eta} |u|^2\right)$$

for all $|u| \geq 2^{4n\eta+1}$. It then follows that

$$\int_{\{|u| \geq 2^{4n\eta+1}\}} |\Psi(u - iH_E^{-1}v)| du \leq C \{2^{nm} \gamma^{2^n} + \exp(-2^{6n\eta-1})\}.$$

To consider the case of $|u| \leq 2^{4n\eta+1}$, we first set $w = u - iH_E^{-1}v$. We then have

$$\begin{aligned}
e^{-iu^tv}\Psi(w) &= e^{-iu^tv} \prod_{r \in E} \psi(2^{-n/2}G_r^t w) \\
&= e^{-iu^tv} \exp \left(\sum_{r \in E} \left\{ -\frac{1}{2}2^{-n} |G_r^t w|^2 + \chi(2^{-n/2}G_r^t w) \right\} \right) \\
&= e^{-iu^tv} \exp \left(\frac{1}{2}v^t H_E^{-1}v - \frac{1}{2}u^t H_E u + iu^tv + \Lambda(w) \right) \\
&= \exp \left(\frac{1}{2}v^t H_E^{-1}v - \frac{1}{2}u^t H_E u + \Lambda(w) \right),
\end{aligned}$$

where

$$\Lambda(w) = \sum_{r \in E} \chi(2^{-n/2}G_r^t w) = \sum_{k=2}^{\infty} T_{2k}(w) \text{ with } T_{2k}(w) = 2^{-kn} \sum_{r \in E} c_{2k}(G_r^t w).$$

We see that T_{2k} is a homogeneous polynomial of degree $2k$ and satisfies

$$|T_{2k}(w)| \leq C 2^{n(1+k(2\eta-1))} |w|^{2k}.$$

Next, we approximate $e^{\Lambda(w)}$ by an inhomogeneous polynomial of degree 2κ composed of $\{T_{2k}\}_{k=2}^{\kappa}$ and their powers (as mentioned previously, the original proof of [33] set $\kappa = 2$). First note that

$$\begin{aligned}
\left| \sum_{k=\kappa+1}^{\infty} T_{2k}(w) \right| &\leq \sum_{k=\kappa+1}^{\infty} |T_{2k}(w)| \leq C \sum_{k=\kappa+1}^{\infty} 2^{n(1+k(2\eta-1))} |w|^{2k} \\
&= C 2^n \sum_{k=\kappa+1}^{\infty} (2^{(2\eta-1)n} |w|^2)^k \\
&\leq C 2^n (2^{(2\eta-1)n} |w|^2)^{2(\kappa+1)} \\
&= C 2^{n(1+2(\kappa+1)(2\eta-1))} |w|^{4(\kappa+1)}
\end{aligned}$$

and so the simple Taylor approximation $e^x = 1 + O(x)$ for small $x \geq 0$ gives

$$\begin{aligned}
 e^{\Lambda(w)} &= \prod_{k=2}^{\infty} e^{T_{2k}(w)} = \prod_{k=2}^{\kappa} e^{T_{2k}(w)} \cdot \exp \left(\sum_{k=\kappa+1}^{\infty} T_{2k}(w) \right) \\
 (6.12) \quad &= \prod_{k=2}^{\kappa} e^{T_{2k}(w)} \left\{ 1 + O \left(2^{n(1+2(\kappa+1)(2\eta-1))} |w|^{4(\kappa+1)} \right) \right\}.
 \end{aligned}$$

We now consider each $e^{T_{2k}(w)}$ term in the product. For each k and integer $l \geq 1$,

$$\begin{aligned}
 e^{T_{2k}(w)} - 1 &= \sum_{i=1}^{\infty} \frac{1}{i!} T_{2k}(w)^i = \sum_{i=1}^l \frac{1}{i!} T_{2k}(w)^i + O \left(|T_{2k}(w)|^{l+1} \right) \\
 &= \sum_{i=1}^l \frac{1}{i!} T_{2k}(w)^i + O \left(2^{n(1+k(2\eta-1))(l+1)} |w|^{2k(l+1)} \right).
 \end{aligned}$$

Since $k \geq 2$, $1 - k \leq -\frac{k}{2}$ then $(1 + k(2\eta - 1)) \leq k(2\eta - \frac{1}{2})$. Hence, if for each k we choose l_k to be the smallest integer such that $k(l_k + 1) \geq \kappa$, we have

$$\begin{aligned}
 e^{T_{2k}(w)} &= 1 + \sum_{i=1}^{l_k} \frac{1}{i!} T_{2k}(w)^i + O \left(2^{nk(2\eta-\frac{1}{2})(l_k+1)} |w|^{2k(l_k+1)} \right) \\
 (6.13) \quad &= 1 + \sum_{i=1}^{l_k} \frac{1}{i!} T_{2k}(w)^i + O \left(2^{n\kappa(2\eta-\frac{1}{2})} |w|^{2\kappa} \right).
 \end{aligned}$$

Combining (6.12) and (6.13), it follows that

$$e^{\Lambda(w)} = 1 + \sum_{k=2}^{\kappa} q_{2k}(w) + O \left(2^{n\kappa(2\eta-\frac{1}{2})} |w|^{2\kappa} \right)$$

for some homogeneous polynomials $\{q_{2k}\}_{k=2}^{\kappa}$, where each q_{2k} has degree $2k$ and coefficients bounded by $C2^{nk(2\eta-\frac{1}{2})}$. Consequently, setting $Q_{2\kappa}(w) := \sum_{k=2}^{\kappa} q_{2k}(w)$, we have

$$\begin{aligned}
 &e^{-\frac{1}{2}v^t H_E^{-1}v} \int_{\{|u| \leq 2^{4n\eta+1}\}} \left| e^{-iu^t v} \Psi(u - iH_E^{-1}v) - (1 + Q_{2\kappa}(u - iH_E^{-1}v)) e^{-\frac{1}{2}u^t H_E u} \right| du \\
 &= e^{-\frac{1}{2}v^t H_E^{-1}v} \int_{\{|u| \leq 2^{4n\eta+1}\}} \left| e^{\Lambda(u - iH_E^{-1}v)} - 1 - Q_{2\kappa}(u - iH_E^{-1}v) \right| e^{-\frac{1}{2}u^t H_E u} du \\
 &\leq C2^{n\kappa(2\eta-\frac{1}{2})} e^{-\frac{1}{2}v^t H_E^{-1}v} \int_{\{|u| \leq 2^{4n\eta+1}\}} e^{-\frac{1}{2}u^t H_E u} |u - iH_E^{-1}v|^{2\kappa} du \\
 &= C2^{n\kappa(2\eta-\frac{1}{2})+8n\kappa} = C2^{n\kappa(10\eta-\frac{1}{2})}
 \end{aligned}$$

where we have used the inequality $|iH_E^{-1}v| \leq \|H_E^{-1}\| |v| \leq 12|v|$. Moreover,

$$\int_{\{|u| \geq 2^{4n\eta+1}\}} |1 + Q_{2\kappa}(u - iH_E^{-1}v)| e^{-\frac{1}{2}u^t H_E u} du \leq C e^{-2^{2n\eta}}.$$

Collecting these bounds, the lemma then follows by setting

$$S_{E,2k} := 2^{-nk(2\eta-\frac{1}{2})} R_{E,2k} \text{ where } R_{E,2k}(v) := \int_{\mathbb{R}^{d_2}} q_{2k}(u - iH_E^{-1}v) e^{-\frac{1}{2}u^t H_E u} du,$$

for each $k \in \{2, \dots, \kappa\}$. The polynomial $R_{E,2k}$ is of degree $2k$ with coefficients bounded by $C2^{nk(2\eta-\frac{1}{2})}$. The proof is complete. $\square$

6.4. Proof of the coupling construction

We are now in a position to prove Proposition 6.2.3, which in turn establishes Theorem 6.2.1. Throughout the proof we treat $Y_E, Z_E \in [\mathbb{R}^d, \mathbb{R}^d] \subset (\mathbb{R}^d)^{\otimes 2}$ as vectors in $\mathbb{R}^{d_2} = \mathbb{R}^{\frac{d}{2}(d-1)}$ so as to avoid the bulky notation of $\|Y_E - Z_E\|_{L^p}$ and $\|Y_E - Z_E\|_{(\mathbb{R}^d)^{\otimes 2}}|_{L^p}$. Instead, we will simply write $|Y_E - Z_E|_{L^p}$.

PROOF OF PROPOSITION 6.2.3. The idea of Davie's original proof in [33] is to construct a coupling of Y_E and Z_E recursively, starting with the base case $E_0 = \{0, 1, \dots, 2^m - 1\}$ and proceeding by successive bisection to smaller dyadic sets.

To begin the construction, we fix a constant $\eta \in (0, \frac{1}{44})$ and let κ be the smallest integer such that $\kappa \geq 2p$. It follows that any constant dependent upon both κ and p can be made solely dependent on p . Intuitively, the inequality $\kappa \geq 2p \geq 2$ makes sense. If we desire a stronger L^p -bound, the degree (2κ) of the polynomial approximation produced by Lemma 6.3.1 will have to increase.

Initial step. We start the construction by finding a coupling between Y_{E_0} and Z_{E_0} . To this end, let $\mathcal{E}_0$ be the event $\{\|G_r\| \leq 2^{m\eta} : \text{for all } r \in E_0\}$. Note that under $\mathcal{E}_0$, $\|H_{E_0}\| \leq 2^{2m\eta}$ (and recall that $\|H_{E_0}^{-1}\| \leq 12$ regardless). Thus

provided $\mathcal{E}_0$ holds, we apply Lemma 6.3.1 to find that for all $|y| \leq 2^{m\eta}$,

$$(6.14) \quad \left| \frac{f_{E_0}(y)}{\phi_{H_{E_0}}(y)} - \{1 + r_{E_0}\}(y) \right| \leq C_\kappa 2^{m\kappa(10\eta - \frac{1}{2})},$$

where

$$r_{E_0} := \sum_{k=2}^{\kappa} 2^{nk(2\eta - \frac{1}{2})} S_{E_0, 2k}.$$

Here, each $S_{E_0, 2k} \in P(\mathbb{R}^{d_2})$ is a polynomial of degree $2k$ for which the absolute values of its coefficients are bounded by some constant independent of η, H_{E_0}, k . We write $y = H_{E_0}^{1/2}u$ and define the probability density $h(u) = (\det H_{E_0})^{1/2} f_{E_0}(H_{E_0}^{1/2}u)$. To convince ourselves that h is a probability density, we observe that the Jacobian matrix of the linear transformation $y \mapsto H_{E_0}^{-1/2}y$ is the matrix $H_{E_0}^{-1/2}$ itself and so the chain rule gives

$$f_{E_0}(y) = \left(\det H_{E_0}^{-1/2} \right) h \left(H_{E_0}^{-1/2}y \right) = (\det H_{E_0}^{1/2})^{-1} h(u) = (\det H_{E_0})^{-1/2} h(u).$$

The matrix transformation also gives

$$\phi_{H_{E_0}}(y) = (\det H_{E_0})^{-1/2} \frac{1}{(2\pi)^{d_2/2}} \exp \left(-\frac{1}{2} y^t H_{E_0}^{-1} y \right) = (\det H_{E_0})^{-1/2} \phi(u)$$

and this change of variables immediately implies the inequality

$$(6.15) \quad \mathcal{W}_p(f_{E_0}, \phi_{H_{E_0}}) \leq \|H_{E_0}^{1/2}\| \mathcal{W}_p(h, \phi).$$

Setting $\mathcal{A} = \{u \in \mathbb{R}^{d_2} : |H_{E_0}^{1/2}u| \leq 2^{m\eta}\}$, it follows from (6.14) that

$$(6.16) \quad \begin{aligned} \int_{\mathcal{A}} \left| h(u) - \left\{ 1 + r_{E_0} \left(H_{E_0}^{1/2}u \right) \right\} \phi(u) \right| du &\leq C_\kappa 2^{m\kappa(10\eta - \frac{1}{2})} \int_{\mathcal{A}} (1 + |u|^p) \phi(u) du \\ &\leq C_p 2^{m\kappa(10\eta - \frac{1}{2})}. \end{aligned}$$

The exponential tail property of the Gaussian distribution ensures

$$\begin{aligned} \int_{\mathcal{A}^c} (1 + |u|)^p \left\{ 1 + r_{E_0} \left(H_{E_0}^{1/2}u \right) \right\} \phi(u) du &\leq C_p (1 + \|H_{E_0}\|^\kappa) \int_{\mathcal{A}^c} |u|^{2\kappa} \phi(u) du \\ &\leq C_p 2^{2m\kappa\eta} e^{-\alpha 2^{m\eta}} \leq C_p 2^{-\beta 2^{m\eta}} \end{aligned}$$

for some constants $\alpha > \beta > 0$. Similarly, the density h conditional on $\mathcal{G}$ possesses exponential tails and so $\int_{\mathcal{A}^c} (1 + |u|)^p h(u) du \leq C_p 2^{-\beta 2^{m\eta}}$. Thus the integral of (6.16) can be taken over all of $\mathbb{R}^{d_2}$ with the inequality remaining true.

In light of (6.16), (now taken over $\mathbb{R}^{d_2}$), we now apply Proposition 6.1.1 with $\Sigma = I_{d_2}$, $\delta = C 2^{m\kappa(10\eta - \frac{1}{2})}$, $K \leq C$, $n_0 = 2$, $n = \kappa$ and $\varepsilon = 2^{m(2\eta - \frac{1}{2})}$ to find that conditional of $\mathcal{E}_0$,

$$\mathcal{W}_p(h, \phi) \leq C_{\kappa, p} \left(2^{2m(2\eta - \frac{1}{2})} + 2^{m\frac{\kappa}{p}(10\eta - \frac{1}{2})} + 2^{m\frac{\kappa+1}{p}(2\eta - \frac{1}{2})} \right) \leq C_{\kappa, p} 2^{m(21\eta - 1)}.$$

Here, we have used the fact that $\frac{\kappa}{p} \geq 2$. Recalling (6.15), we deduce that

$$(6.17) \quad \mathcal{W}_p(f_{E_0}, \phi_{H_{E_0}}) \leq \|H_{E_0}^{1/2}\| \mathcal{W}_p(h, \phi) \leq C_{\kappa, p} 2^{m(22\eta - 1)}$$

since under $\mathcal{E}_0$ we have $\|H_{E_0}\| \leq 2^{2m\eta}$. To finish the initial step, we recall the exponential tail bounds of $\|G_r\|^2$ given by Lemma 6.2.2 to deduce that for all $\alpha \in (0, \frac{1}{48d})$:

$$\begin{aligned} \mathbb{P}(\mathcal{E}_0^c) &\leq \sum_{r \in E_0} \mathbb{P}(\|G_r\| \geq 2^{m\eta}) \leq 2^m \mathbb{P}\left(e^{\alpha\|G_0\|^2} \geq e^{\alpha 2^{2m\eta}}\right) \\ &\leq 2^m e^{-\alpha 2^{2m\eta}} \mathbb{E}\left(e^{\alpha\|G_0\|^2}\right) \\ &= C_\alpha \exp\left(-\alpha 2^{2m\eta} + m \log 2\right). \end{aligned}$$

If $\mathcal{E}_0$ fails we simply construct independent copies of Y_{E_0} and Z_{E_0} . Hence, from (6.17) and Hölder's inequality, we can find a coupling of Y_{E_0} and Z_{E_0} such that unconditionally

$$\mathbb{E}(|Y_{E_0} - Z_{E_0}|^p) \leq \mathbb{E}(|Y_{E_0} - Z_{E_0}|^p \mathbf{1}_{\mathcal{E}_0}) + \mathbb{E}(|Y_{E_0} + Z_{E_0}|^p \mathbf{1}_{\mathcal{E}_0^c}) \leq C_{\kappa, p} 2^{mp(22\eta - 1)}$$

for some constant $C_{\kappa, p} > 0$.

Recursive step. Let E be a dyadic set of size 2^n . Then we can write $E = F \cup G$ where F and G are disjoint dyadic sets of size 2^{n-1} . Note that

$$Y_F + Y_G = 2^{1/2}Y_E \text{ and } Z_F + Z_G = 2^{1/2}Z_E.$$

We suppose a coupling between Y_E and Z_E has been defined, conditional on $\mathcal{G}$. In other words, for each choice of $\{W^{(j)}\}_{j=0}^{N-1}$, we have a joint distribution of Y_E and Z_E with the correct conditional marginal distributions. We wish to extend this coupling to a coupling between (Y_F, Y_G) and (Z_F, Z_G) .

For each $x \in \mathbb{R}^{d_2}$, let f_x denote the density of Y_E conditional on $Y_E = x$ and on $\mathcal{G}$. Similarly, let g_x be the corresponding density for Z_E . As noted in [33], the conditional distribution of Z_F , given $Z_E = x$ and $\mathcal{G}$, is $N(Jx, H)$ where $J = H_F H_E^{-1}$ and $H = \frac{1}{2} H_F H_E^{-1} H_G$. Thus g_x is the density of $N(Jx, H)$.

We need to find a coupling between Y_F and Z_F conditional on $Y_E = x$ and $Z_E = \tilde{x}$. To do this we need a coupling between the distributions with densities f_x and $g_{\tilde{x}}$. However, we shall instead construct a coupling between f_x and g_x , then use the fact that $g_{\tilde{x}}$ is just g_x translated by $J(x - \tilde{x})$.

We begin by noting that $f_x(y) = \frac{2^{1/2} f_F(y) f_G(2^{1/2}x - y)}{f_E(x)}$. Then provided the event $\mathcal{E} = \{\|G_r\| \leq 2^{n\frac{\eta}{2p}} : \text{for all } r \in E\}$ holds, we apply Lemma 6.3.1 to each of E, F, G to find:

$$(6.18) \quad \left| \frac{f_x(y)}{g_x(y)} - \{1 + r_x(y)\} g_x(y) \right| \leq C_\kappa 2^{n\kappa(10\eta - \frac{1}{2})} \text{ for all } |x|, |y| \leq 2^{n\eta/p},$$

where $r_x(y) = r_F(y) + r_G(2^{1/2}x - y) - r_E(x)$. Note that

$$r_E(y) = \sum_{k=2}^{\kappa} 2^{nk(2\eta - \frac{1}{2})} S_{E,2k},$$

where each $S_{E,2k} \in P(\mathbb{R}^{d_2})$ is a polynomial of degree $2k$ with the absolute value of its coefficients bounded by some constant independent of η, H, k . Corresponding decompositions and associated properties hold for r_F and r_G .

Next, we define the set

$$\Omega := \left\{ x \in \mathbb{R}^{d_2} : \mathbb{E} \left(|Y_F|^p \mathbf{1}_{\{|Y_F| \geq 2^{n\eta/p}\}} \mid Y_E = x \text{ and } \mathcal{G} \right) > 2^{-n\kappa/2} \right\}.$$

Utilising the inequalities of Markov and Hölder gives

$$\begin{aligned} \mathbb{P}(Y_E \in \Omega \mid \mathcal{G}) &\leq 2^{n\kappa/2} \mathbb{E}_{\mathcal{G}} \left(|Y_F|^p \mathbf{1}_{\{|Y_F| \geq 2^{n\eta/p}\}} \right) \\ &\leq 2^{n\kappa/2} \mathbb{E}_{\mathcal{G}} (|Y_F|^{pq})^{1/q} \mathbb{P} \left(|Y_F|^s \geq 2^{n\eta \frac{s}{p}} \mid \mathcal{G} \right)^{1/r} \\ &\leq 2^{n\kappa/2} \mathbb{E}_{\mathcal{G}} (|Y_F|^{pq})^{1/q} 2^{-n\eta \frac{s}{pr}} \mathbb{E}_{\mathcal{G}} (|Y_F|^s)^{1/r} \\ &\leq C 2^{n(\frac{\kappa}{2} - \eta \frac{s}{pr})} \|H_F\|^{\frac{1}{2}(p + \frac{s}{r})}. \end{aligned}$$

where $s \geq 1$ and $q, r > 1$ satisfy $q^{-1} + r^{-1} = 1$. Under the assumption of $\mathcal{E}$ we have $\|H_F\| \leq 2^{\frac{n}{2} \cdot \frac{\eta}{p}}$ (recall F is of size $2^{n/2}$), and so,

$$\mathbb{P}(Y_E \in \Omega \mid \mathcal{G}) \leq C 2^{n(\frac{\kappa}{2} - \eta \frac{s}{pr}) + \eta \frac{n}{4}(1 + \frac{s}{pr})} = C 2^{\frac{n}{4}(2\kappa + \eta(1 - \frac{3s}{pr}))}.$$

Taking $s \geq \frac{pr}{3}(1 + \frac{2}{\eta}(\kappa + 4p))$ ensures that $\frac{n}{4}(2\kappa + \eta(1 - \frac{3s}{pr})) \leq -2pn$ and we conclude that

$$(6.19) \quad \mathbb{P}(Y_E \in \Omega \mid \mathcal{G}) \leq C 2^{-2np}.$$

Define the event $\widehat{\mathcal{E}} := \mathcal{E} \cap \{Y_E \notin \Omega\}$. Note that under $\widehat{\mathcal{E}}$, we have

$$\|H\| \leq \frac{1}{2} \|H_F\| \|H_E^{-1}\| \|H_G\| \leq \frac{12}{6} \left(2^{\frac{n}{2} \cdot \frac{\eta}{p}} \right)^2 = C 2^{n\eta/p}.$$

Let us write $x = Y_E$ for shorthand. In contrast to the initial step, we cannot directly apply the polynomial perturbation result of Proposition 6.1.1 to the conditional distribution of Y_F because $g_x \sim N(Jx, H)$ is not centred. Instead, we need need to make the change of variable $y := Jx + H^{1/2}u$, so that now we have $g_x(y) = (\det H)^{-1/2} \phi(u)$. Similarly, we define $h_x(u) := (\det H)^{1/2} f_x(Jx + H^{1/2}u)$.

Then provided $\widehat{\mathcal{E}}$ holds, (6.18) implies the estimate

$$\int_{\mathcal{A}} (1 + |u|)^p |h_x(u) - \{1 + r_x(Jx + H^{1/2}u)\} \phi(u)| du \leq C_{\kappa,p} 2^{n\kappa(10\eta - \frac{1}{2})},$$

where $\mathcal{A} = \{u \in \mathbb{R}^{d_2} : |Jx + H^{1/2}u| \leq 2^{n\eta/p}\}$. If $u \notin \mathcal{A}$, that is $|y| \geq 2^{n\eta/p}$, then $|H^{1/2}u| \leq 2|y|$ and so $|u| \leq C2^{n\frac{\eta}{2p}}|y|$. As previously argued in [33], using the fact that $x \notin \Omega$ under $\widehat{\mathcal{E}}$ it follows that

$$\int_{\mathcal{A}^c} (1 + |u|)^p h_x(u) du \leq C_{\kappa,p} 2^{n(\eta - \frac{\kappa}{2})}.$$

Moreover, the exponential tail property of the Gaussian distribution ensures that

$$\int_{\mathcal{A}^c} (1 + |u|)^p |1 + r_x(Jx + H^{1/2}u)| \phi(u) du \leq C_{\kappa,p} 2^{-n\kappa/2}.$$

Combining the last three estimates yields

$$\int_{\mathbb{R}^{d_2}} (1 + |u|)^p |h_x(u) - \{1 + r_x(Jx + H^{1/2}u)\} \phi(u)| du \leq C_{\kappa,p} 2^{n\kappa(10\eta - \frac{1}{2})}.$$

As in the initial case, in light of the previous estimate we apply Proposition 6.1.1 with $\Sigma = I_{d_2}$, $\delta = C2^{n\kappa(10\eta - \frac{1}{2})}$, $K \leq C$, $n_0 = 2$, $n = \kappa$ and $\varepsilon = 2^{n(10\eta - \frac{1}{2})}$ to find that, conditional on $\widehat{\mathcal{E}}$,

$$\mathcal{W}_p(h_x, \phi) \leq C_{\kappa,p} 2^{n(21\eta - 1)}.$$

Again we have used the fact $\kappa \geq 2p$. Assuming $\widehat{\mathcal{E}}$, $\|H^{1/2}\| \leq C2^{n\frac{\eta}{2p}}$ and so

$$\mathcal{W}_p(f_x, g_x) \leq \|H^{1/2}\| \mathcal{W}_p(h_x, \phi) \leq C_{\kappa,p} 2^{n(22\eta - 1)}.$$

As Davie writes in [33], the situation can be summarised as follows: conditional on $Y_E = x$ and assuming $\widehat{\mathcal{E}}$, we can find a random variable Z_F^* with density g_x such that $|Z_F^* - Y_F|_{L^p} \leq C_{\kappa,p} 2^{n(22\eta - 1)}$. If $\widehat{\mathcal{E}}$ fails then we generate an independent random variable Z^* with density g_x and set $Z_F^* := Z^*$. As in the initial case, the exponential tail bounds of Lemma 6.2.2 imply that $\mathbb{P}(\mathcal{E}^c) \leq C2^{-2np}$

and so together with (6.19) we certainly have $\mathbb{P}(\widehat{\mathcal{E}}^c) \leq C2^{-2np}$. Thus taking expectations over $\mathcal{G}$ and Y_E and applying the tower property, we find that unconditionally,

$$(6.20) \quad \mathbb{E}(|Z_F^* - Y_F|^p) \leq C_{\kappa,p} 2^{np(22\eta-1)}.$$

We can now complete the recursive step by defining

$$(6.21) \quad Z_F := Z_F^* + H_F H_E^{-1}(Z_E - Y_E),$$

which has the correct conditional density $g_{\tilde{x}}$ with $\tilde{x} = Z_E$. Then we must have $Z_G = 2^{1/2}Z_E - Z_F$. Moreover, setting $Z_F^* := 2^{1/2}Y_E - Z_F^*$, (6.20) and (6.21) hold with F replaced with G .

Conclusion of the proof. Consider a given dyadic set E of size 2^n . We can uniquely write the expansion $E = E_k \subseteq E_{k-1} \subseteq \dots \subseteq E_0$ where $k = m - n$ and, for each j , E_j is a dyadic set of size 2^{m-j} . From (6.21) we obtain

$$Z_E - Y_E = \sum_{j=1}^k H_{E_k} H_{E_j}^{-1}(Z_{E_j}^* - Y_{E_j}) + H_{E_k} H_{E_0}^{-1}(Z_{E_0} - Y_{E_0}).$$

By the Hölder inequality, the exponential tail estimates of $\|H_E\|$ and the fact that $\|H_E^{-1}\| \leq 12$, for all $r \in [1, p)$ we have

$$\begin{aligned} |Z_E - Y_E|_{L^r} &\leq 12 \sum_{j=1}^k \|H_{E_k}\|_{L^q} \left| Z_{E_j}^* - Y_{E_j} \right|_{L^p} + 12 \|H_{E_k}\|_{L^q} |Z_{E_0} - Y_{E_0}|_{L^p} \\ &\leq C_{\kappa,p} \sum_{j=0}^k 2^{(m-j)(22\eta-1)} \leq C_{\kappa,p} 2^{(m-k)(22\eta-1)} = C_{\kappa,p} 2^{n(22\eta-1)}. \end{aligned}$$

Here, $q = \frac{pr}{p-r}$ (so that $q^{-1} + r^{-1} = 1$). Taking suitably small $\eta \in (0, \frac{1}{44})$ ensures that $22\eta - 1 < -\frac{1}{2}$ and the proof is complete. $\square$

APPENDIX A

Baker-Campbell-Hausdorff formula

We use this short appendix to collect some useful formulas for the Baker-Campbell-Hausdorff formula and its iterated version.

A.1. Baker-Campbell-Hausdorff formula

The Baker-Campbell-Hausdorff formula links the structure of a Lie group with the corresponding structure on its associated Lie algebra. It does this by expressing the logarithm of the product of two Lie group elements as a Lie algebra element using only Lie algebraic operations.

Theorem A.1.1 (Baker-Campbell-Hausdorff formula). *Let $\mathcal{G}$ be a Lie group with group product $\star$ and the corresponding Lie algebra $\mathfrak{g}$ defined over any field of characteristic 0. Let $\exp : \mathfrak{g} \rightarrow \mathcal{G}$ be the exponential map. Then for every pair $x, y \in \mathfrak{g}$,*

$$H(x, y) := \log (\exp(x) \star \exp(y)) ,$$

can be written as a formal infinite sum of elements of $\mathfrak{g}$. To be precise:

$$H(x, y) = \sum_{n=1}^{\infty} \frac{(-1)^{n+1}}{n} H_n(x, y)$$

where

$$H_n(x, y) := \sum_{\substack{\{r_i, s_i\}_{i=1}^n \subset \mathbb{Z}_+ \\ r_i + s_i > 0}} \frac{1}{r_1! s_1! \dots r_n! s_n! \sum_{i=1}^n (r_i + s_i)} \\ \times \underbrace{[x, [x, \dots [x}_{r_1 \text{ times}} \underbrace{, [y, [y, \dots [y}_{s_1 \text{ times}} \underbrace{, \dots [x, [x, \dots [x}_{r_n \text{ times}} \underbrace{, [y, [y, \dots y] \dots]}_{s_n \text{ times}}].$$

For a proof of the Baker-Campbell-Hausdorff formula, see [28, Theorem 1.2.1]. The first terms of order less than or equal to 5 are given by

$$\begin{aligned} H(x, y) = & x + y + \frac{1}{2}[x, y] + \frac{1}{12}([x, [x, y]] + [y, [y, x]]) - \frac{1}{24}[y, [x, [x, y]]] \\ & - \frac{1}{720}([[[[x, y], y], y], y] + [[[[y, x], x], x], x]) \\ & + \frac{1}{360}([[[[x, y], y], y], x] + [[[[y, x], x], x], y]) \\ & + \frac{1}{120}([[[[y, x], y], x], y] + [[[[x, y], x], y], x]) + \dots \end{aligned}$$

A.2. Iterated Baker-Campbell-Hausdorff formula

We can also consider the iterated product:

$$H(x_1, \dots, x_n) := \log(e^{x_1} \star \dots \star e^{x_n}).$$

This is the iterated version of the Baker-Campbell-Hausdorff formula. It was first used in the proof of the Chen-Strichartz development formula, where the latter formula gives an explicit expression for the logarithm of the Brownian signature lift $\log S_N(\mathbb{W})_{s,t}$ (see [23, 109]).

Theorem A.2.1 (Iterated Baker-Campbell-Hausdorff formula). *The iterated Hausdorff coefficient has the form:*

$$H(x_1, \dots, x_n) = \sum_{k=1}^{\infty} \frac{(-1)^{k-1}}{k} \sum_{P \in B_k} \frac{1}{\overline{P} P!} x^P,$$

where B_k , $\overline{P}$, $P!$ and X^P are given by the following expressions:

$$\begin{aligned} B_k &= \left\{ (p_i^j)_{\substack{i \in \{1, \dots, n\} \\ j \in \{1, \dots, k\}}}, p_i^j \in \mathbb{N} : \forall j \in \{1, \dots, k\}, \sum_{i=1}^n p_i^j > 0 \right\}, \\ \overline{P} &= \sum_{j=1}^k \sum_{i=1}^n p_i^j, \quad P! = \prod_{j=1}^k \prod_{i=1}^n p_i^j!, \\ x^P &= \underbrace{[x_1 \dots [x_1 \dots]}_{p_1^1 \text{ times}} \underbrace{[x_n \dots [x_n \dots]}_{p_n^1 \text{ times}} \underbrace{[x_1 \dots [x_1 \dots]}_{p_1^k \text{ times}} \underbrace{[x_n \dots [x_n, x_n] \dots]}_{p_n^k \text{ times}} \dots \end{aligned}$$

We refer to [9, Appendix B], [20, §3.2] and [101, Theorem 3.11] for further details. By induction (see Example 3.2 of [6]) it can be shown that the first terms of the expansion up to nested Lie brackets of length 4 are given by

(A.1)

$$\begin{aligned}
& H(x_1, \dots, x_n) \\
&= \sum_{i=1}^n x_i + \frac{1}{2} \sum_{1 \leq i < j \leq n} [x_i, x_j] \\
&+ \frac{1}{4} \sum_{1 \leq i < j < k \leq n} [[x_i, x_j], x_k] + \frac{1}{12} \sum_{i,j} \sum_{k > i \vee j} [x_i, [x_j, x_k]] + \frac{1}{12} \sum_{1 \leq i < j \leq n} [x_j, [x_j, x_i]] \\
&+ \frac{1}{8} \sum_{1 \leq i < j < k < l \leq n} [[[x_i, x_j], x_k], x_l] + \frac{1}{24} \sum_{i,j} \sum_{l > k > i \vee j} [[x_i, [x_j, x_k]], x_l] \\
&+ \frac{1}{24} \sum_{1 \leq i < j < k \leq n} [[x_j, [x_j, x_i]], x_k] + \frac{1}{24} \sum_{1 \leq i < j < k \leq n} [x_k, [x_k, [x_i, x_j]]] \\
&- \frac{1}{24} \sum_{i,j} \sum_{k > i \vee j} [x_k, [x_i, [x_j, x_k]]] + \dots
\end{aligned}$$

We now specialise to the case of the free κ -nilpotent Lie group and algebra:

$$(\mathcal{G}, \star, \mathfrak{g}) := (G^{(\kappa)}(\mathbb{R}^d), \otimes, \mathfrak{g}^{(\kappa)}(\mathbb{R}^d)).$$

Next, we present a useful technical lemma for expressing the difference of two iterated BCHF expansions as the global difference at the Lie algebra level, rather than the local difference at the Lie group level.

Lemma A.2.2. *Fix sequences $\{x_j\}_{j=1}^n, \{y_j\}_{j=1}^n \in G^{(\kappa)}(\mathbb{R}^d)$ and set*

$$g_i = \pi_i(H(x_1, \dots, x_n)), \quad h_i = \pi_i(H(y_1, \dots, y_n)) \in \mathfrak{g}^{(\kappa)}(\mathbb{R}^d).$$

Then for every integer $m \leq \kappa$,

$$\begin{aligned}
& \|\pi_m(e^{x_1} \otimes \dots \otimes e^{x_n} - e^{y_1} \otimes \dots \otimes e^{y_n})\|_{(\mathbb{R}^d)^{\otimes m}} \\
& \leq \sum_{k=1}^m \frac{1}{k!} \sum_{\substack{i_1, \dots, i_k > 0 \\ i_1 + \dots + i_k = m}} \left\| \sum_{j=1}^k g_{i_1} \otimes \dots \otimes g_{i_{j-1}} \otimes (g_{i_j} - h_{i_j}) \otimes h_{i_{j+1}} \otimes \dots \otimes h_{i_k} \right\|_{(\mathbb{R}^d)^{\otimes m}}.
\end{aligned}$$

PROOF. Recall the well-known non-commutative identity ([7, §4]),

$$(A.2) \quad \bigotimes_{j=1}^n a_j - \bigotimes_{j=1}^n b_j = \sum_{j=1}^n \bigotimes_{i=1}^{j-1} a_i (a_j - b_j) \bigotimes_{i=j+1}^n b_i,$$

for any sequences $\{a_j\}, \{b_j\}$, with the convention that $\bigotimes_{i=1}^0 a_i = 1$. It follows that for every positive integer m ,

$$\begin{aligned} & \|\pi_m(e^{x_1} \otimes \dots \otimes e^{x_n} - e^{y_1} \otimes \dots \otimes e^{y_n})\|_{(\mathbb{R}^d)^{\otimes m}} \\ &= \|\pi_m(e^{H(x_1, \dots, x_n)} - e^{H(y_1, \dots, y_n)})\|_{(\mathbb{R}^d)^{\otimes m}} \\ &= \left\| \pi_m \left(\sum_{k=0}^m \frac{1}{k!} (H(x_1, \dots, x_n)^{\otimes k} - H(y_1, \dots, y_n)^{\otimes k}) \right) \right\|_{(\mathbb{R}^d)^{\otimes m}} \\ &\leq \sum_{k=1}^m \frac{1}{k!} \|\pi_m(H(x_1, \dots, x_n)^{\otimes k} - H(y_1, \dots, y_n)^{\otimes k})\|_{(\mathbb{R}^d)^{\otimes m}} \\ &= \sum_{k=1}^m \frac{1}{k!} \left\| \sum_{\substack{i_1, \dots, i_k > 0 \\ i_1 + \dots + i_k = m}} (g_{i_1} \otimes \dots \otimes g_{i_k} - h_{i_1} \otimes \dots \otimes h_{i_k}) \right\|_{(\mathbb{R}^d)^{\otimes m}} \\ &\leq \sum_{k=1}^m \frac{1}{k!} \sum_{\substack{i_1, \dots, i_k > 0 \\ i_1 + \dots + i_k = m}} \left\| \sum_{j=1}^k g_{i_1} \otimes \dots \otimes g_{i_{j-1}} \otimes (g_{i_j} - h_{i_j}) \otimes h_{i_{j+1}} \otimes \dots \otimes h_{i_k} \right\|_{(\mathbb{R}^d)^{\otimes m}}. \end{aligned}$$

The proof is complete. $\square$

We conclude the appendix with a useful technical lemma ([6, Lemma 3.1]).

Lemma A.2.3. *Any Lie bracket of elements $x_1, \dots, x_n \in \mathfrak{g}$ is a linear combination with coefficients of ± 1 of nested commutators of the form*

$$x_{i_1} * x_{i_2} * \dots * x_{i_{n-1}} * x_{i_n} := [x_{i_1}, [x_{i_2}, [\dots [x_{i_{n-1}}, x_{i_n}] \dots]]].$$

PROOF. The proof relies upon the Jacobi identity and induction on the length of the commutator. The induction basis is the identity $[[x_1, x_2], [x_3, x_4]] = [x_4, [x_3, [x_1, x_2]]] - [x_3, [x_2, [x_1, x_2]]]$. $\square$

Bibliography

- [1] A. Alfonsi, B. Jourdain, and A. Kohatsu-Higa. Pathwise optimal transport bounds between a one-dimensional diffusion and its Euler scheme. *Ann. Appl. Probab.*, 24(3):1049–1080, 2014.
- [2] A. Alfonsi, B. Jourdain, and A. Kohatsu-Higa. Optimal transport bounds between the time-marginals of a multidimensional diffusion and its Euler scheme. *Electron. J. Probab.*, 2015.
- [3] C.J.S. Alves and A.B. Cruzeiro. Monte Carlo simulation of stochastic differential systems - a geometrical approach. *Stochastic Processes. Appl.*, 118(3):346–367, 2008.
- [4] G. Arous. Flots et séries de Taylor stochastiques. *Probab. Theory Related Fields*, 81(1):29–77, 1989.
- [5] I. Bailleul. Flows driven by Banach space-valued rough paths. In *Séminaire de Probabilités XLVI*, pages 195–205. Springer, 2014.
- [6] Z. Balogh, R. Berger, R. Monti, and J. Tyson. Exceptional sets for self-similar fractals in Carnot groups. In *Math. Proc. Cambridge Philos. Soc.*, volume 149, pages 147–172. Cambridge University Press, 2010.
- [7] R.F. Bass, B.M. Hambly, and T.J. Lyons. Extending the wong-zakai theorem to reversible markov processes. *Eur. Math. Soc. (JEMS)*, 4(3):237–269, 2002.
- [8] F. Baudoin. Rough paths theory. *Lecture notes*, 2013.
- [9] Fabrice Baudoin. *An introduction to the geometry of stochastic flows*. Imperial College Press, 2004.
- [10] C. Bayer, P. Friz, and J. Gatheral. Pricing under rough volatility. *Preprint*, 2015.
- [11] C. Bayer, P. Friz, S. Riedel, and J. Schoenmakers. From rough path estimates to multilevel Monte Carlo. *arXiv preprint arXiv:1305.5779*, 2013.
- [12] J. Blanchet, X. Chen, and J. Dong. Epsilon-Strong Simulation for Multidimensional Stochastic Differential Equations via Rough Path Analysis. *arXiv preprint arXiv:1403.5722*, 2014.
- [13] Y. Boutaib, L.G. Gyurkó, T.J. Lyons, and D. Yang. Dimension-free Euler estimates of rough differential equations. *Rev. Roumanine Math. Pures Appl.*, 2013.
- [14] E. Breuillard. Local limit theorems and equidistribution of random walks on the Heisenberg group. *Geom. Funct. Anal.*, 15(1):35–82, 2005.
- [15] E. Breuillard, P. Friz, and M. Huesmann. From random walks to rough paths. *Proc. Amer. Math. Soc.*, 137(10):3487–3496, 2009.
- [16] D.L. Burkholder, B.J. Davis, and R.F. Gundy. Integral inequalities for convex functions of operators on martingales. In *Proc. Sixth Berkeley Symp. Math. Statist. Prob.*, volume 2, pages 223–240, 1972.
- [17] R.J. Cameron and J.M.C. Clark. *The maximum rate of convergence of discrete approximations for stochastic differential equations*. Springer, 1980.
- [18] T. Cass, C. Litterer, and T.J. Lyons. Integrability and tail estimates for Gaussian rough differential equations. *Ann. Probab.*, 41(4):3026–3050, 2013.

- [19] T. Cass and T.J. Lyons. Evolving communities with individual preferences. *Proc. Lond. Math. Soc.*, pages 83–107, 2014.
- [20] F. Castell. Asymptotic expansion of stochastic flows. *Probab. Theory Related Fields*, 96(2):225–239, 1993.
- [21] F. Castell and J. Gaines. An efficient approximation method for stochastic differential equations by means of the exponential Lie series. *Math. Comput. Simulation*, 38(1):13–19, 1995.
- [22] F. Castell and J. Gaines. The ordinary differential equation approach to asymptotically efficient schemes for solution of stochastic differential equations. In *Ann. Inst. Henri Poincaré Probab. Stat.*, volume 32, pages 231–250, 1996.
- [23] K.T. Chen. Integration of paths, geometric invariants and a generalized Baker-Hausdorff formula. *Ann. of Math.*, pages 163–178, 1957.
- [24] W.L. Chow. Über Systeme von linearen partiellen Differentialgleichungen erster Ordnung. *Math. Ann.*, pages 98–105, 1939.
- [25] K.L. Chung. *A course in probability theory*. Academic Press, 2001.
- [26] E. Clément and A. Gloter. An application of the KMT construction to the pathwise weak error in the Euler approximation of the geometric Brownian motion. *Preprint*, 2015.
- [27] F. Comte and E. Renault. Long memory in continuous-time stochastic volatility models. *J. Math. Fin.*, 8(4):291–323, 1998.
- [28] L. Corwin and Frederick P. Greenleaf. *Representations of Nilpotent Lie Groups and Their Applications. Part I*, volume 18. Cambridge University Press, 2004.
- [29] L. Coutin, P. Friz, and N. Victoir. Good rough path sequences and applications to anticipating stochastic calculus. *Ann. Probab.*, 35(3):1172–1193, 2007.
- [30] L. Coutin and A. Lejay. Semimartingales and rough paths theory. *Electron. J. Probab.*, 10(23):761–785, 2005.
- [31] A.B. Cruzeiro and P. Malliavin. Numerical approximation of diffusions in $\mathbb{R}^d$ using normal charts of a Riemannian manifold. *Stochastic Processes. Appl.*, 116(7):1088–1095, 2006.
- [32] A.B. Cruzeiro, P. Malliavin, and A. Thalmaier. Geometrization of Monte-Carlo numerical analysis of an elliptic operator: strong approximation. *C. R. Math. Acad. Sci. Soc. Paris*, 338(6):481–486, 2004.
- [33] A.M. Davie. KMT theory applied to approximations of SDE. In *Stochastic Analysis and Applications 2014*, pages 185–201. Springer, 2014.
- [34] A.M. Davie. Pathwise approximation of stochastic differential equations using coupling. *preprint*, 2014.
- [35] A.M. Davie. Polynomial perturbations of normal distributions. *preprint*, 2015.
- [36] A. Deya, A. Neuenkirch, and S. Tindel. A Milstein-type scheme without Lévy area terms for SDEs driven by fractional Brownian motion. In *Ann. Inst. Henri Poincaré Probab. Stat.*, volume 48, pages 518–550. Institut Henri Poincaré, 2012.
- [37] A.S. Dickinson. Optimal Approximation of the Second Iterated Integral of Brownian Motion. *Stochastic Processes. Appl.*, 25(5):1109–1128, 2007.
- [38] H. Doss. Liens entre équations différentielles stochastiques et ordinaires. In *Ann. Inst. Henri Poincaré Probab. Stat.*, volume 13, pages 99–125, 1977.
- [39] U. Einmahl. Extensions of results of Komlós, Major, and Tusnády to the multivariate case. *J. Multivariate Anal.*, 28(1):20–68, 1989.
- [40] D. Flies M., Normand-Cyrot. Algèbres de Lie nilpotentes, formule de Baker-Campbell-Hausdorff et intégrals itérées de K.T. Chen. *Séminaire de Probabilités LNM*, 920, 1982.
- [41] G. Flint, B. Hambly, and T.J. Lyons. Discretely sampled signals and the rough Hoff path. *To appear in Stochastic Processes. Appl.*, 2016.

- [42] G. Flint and T.J. Lyons. Pathwise approximation of SDEs by coupling piecewise abelian rough paths. *arXiv preprint arXiv:1505.01298*, 2015.
- [43] P. Friz. Examples from physics and economics where rough paths matter. *Oberwolfach report*, 9(3), 2012.
- [44] P. Friz, P. Gassiat, and T.J. Lyons. Physical Brownian motion in magnetic field as a rough path. *Trans. Amer. Math. Soc.*, 2015.
- [45] P. Friz and M. Hairer. *A course on rough paths: with an introduction to regularity structures*. Springer, 2014.
- [46] P. Friz and H. Oberhauser. Rough path limits of the Wong–Zakai type with a modified drift term. *J. Funct. Anal.*, 256(10):3236–3256, 2009.
- [47] P. Friz and S. Riedel. Integrability of (non-) linear rough differential equations and integrals. *Stochastic Processes. Appl.*, 31(2):336–358, 2013.
- [48] P. Friz and S. Riedel. Convergence rates for the full Gaussian rough paths. In *Ann. Inst. Henri Poincaré Probab. Stat.*, volume 50, pages 154–194. Institut Henri Poincaré, 2014.
- [49] P. Friz and N. Victoir. Approximations of the Brownian rough path with applications to stochastic analysis. In *Ann. Inst. Henri Poincaré Probab. Stat.*, volume 41, pages 703–724, 2005.
- [50] P. Friz and N. Victoir. A note on the notion of geometric rough paths. *Probab. Theory Related Fields*, 136(3):395–416, 2006.
- [51] P. Friz and N. Victoir. Euler estimates for rough differential equations. *J. Differential Eqn.*, 244(2):388–412, 2008.
- [52] P. Friz and N. Victoir. *Multidimensional stochastic processes as rough paths: theory and applications*, volume 120. Cambridge University Press, 2010.
- [53] J.G. Gaines and T.J. Lyons. Random generation of stochastic area integrals. *SIAM J. Appl. Math.*, 54(4):1132–1146, 1994.
- [54] J. Gatheral, T. Jaisson, and M. Rosenbaum. Volatility is rough. *Preprint*, 2014.
- [55] M. Gelbrich. Simultaneous time and chance discretization for stochastic differential equations. *J. Comput. Appl. Math.*, 58(3):255–289, 1995.
- [56] M. Gelbrich and S. Rachev. Discretization for stochastic differential equations, Lp Wasserstein metrics, and econometrical models. *Springer*, pages 97–119, 1996.
- [57] E. Gobet. Weak approximation of killed diffusion using Euler schemes. *Stochastic Processes. Appl.*, 87(2):167–197, 2000.
- [58] M. Gubinelli and A. Lejay. Global existence for rough differential equations under linear growth conditions. *arXiv preprint arXiv:0905.2399*, 2009.
- [59] G. Gyurkó, T.J. Lyons, M. Kontkowski, and J. Field. Extracting information from the signature of a financial data stream. Technical report, arXiv. org, 2013.
- [60] L.G. Gyurkó. *Numerical methods for approximating solutions to rough differential equations*. PhD thesis, University of Oxford, 2008.
- [61] L.G. Gyurkó and T.J. Lyons. Rough paths based numerical algorithms in computational finance. *Mathematics in Finance: UIMP-RSME Lluis A. Santaló Summer School*, 2008.
- [62] M. Hairer. Solving the KPZ equation. *Ann. of Math.*, 178:559–664, 2013.
- [63] M. Hairer. A theory of regularity structures. *Invent. Math.*, 198(2):269–504, 2014.
- [64] B. Hambly and T.J. Lyons. Uniqueness for the signature of a path of bounded variation and the reduced path group. *Ann. of Math.*, 171(1):109–167, 2010.
- [65] A. Hammond and S. Sheffield. Power law Pólya’s urn and fractional Brownian motion. *Probab. Theory Related Fields*, 157(3-4):691–719, 2013.
- [66] B. Hoff. *The Brownian frame process as a rough path*. PhD thesis, University of Oxford, 2005.
- [67] K. Itô. Stochastic integral. *Proc. Imperial Acad.*, 20(8):519–524, 1944.

- [68] S. Kanagawa. The rate of convergence for approximate solutions of stochastic differential equations. *Tokyo J. Math.*, 12(1), 1989.
- [69] L.V. Kantorovich. On a problem of Monge (Russian). *Uspekhi Mat. Nauk.*, 3:225–226, 1948.
- [70] P.E. Kloeden and E. Platen. *Numerical solution of stochastic differential equations*, volume 23. Springer, 1992.
- [71] P.E. Kloeden, E. Platen, and I. Wright. The approximation of multiple stochastic integrals. *Stochastic Processes. Appl.*, 10(4):431–441, 1992.
- [72] J. Lamperti. A simple construction of certain diffusion processes. *J. Math. Kyoto*, 4(1):161–170, 1964.
- [73] M. Ledoux, Z. Qian, and T. Zhang. Large deviations and support theorem for diffusion processes via rough paths. *Stochastic Processes. Appl.*, 102(2):265–283, 2002.
- [74] A. Lejay and T.J. Lyons. On the importance of the Lévy area for studying the limits of functions of converging stochastic processes. Application to homogenization. In *Current trends in potential theory*, volume 7. The Theta foundation/American Mathematical Society, 2003.
- [75] D. Levin, T.J. Lyons, and H. Ni. *Learning from the past, predicting the statistics for the future, learning an evolving system*. preprint, 2013.
- [76] T.J. Lyons. Differential equations driven by rough signals. I. An extension of an inequality of Young. *Math. Res. Lett.*, 1(4):451–464, 1994.
- [77] T.J. Lyons. Differential equations driven by rough signals. *Rev. Mat. Iberoam.*, 14(2):215–310, 1998.
- [78] T.J. Lyons. Rough paths, Signatures and the modelling of functions on streams. *arXiv preprint arXiv:1405.4537*, 2014.
- [79] T.J. Lyons, M. Caruana, and T. Lévy. Differential equations driven by rough paths. *École d’Été de Probabilités de Saint-Flour XXXIV*, pages 1–93, 2004.
- [80] T.J. Lyons, H. Ni, and H. Oberhauser. A feature set for streams and an application to high-frequency financial tick data. In *Proceedings of the 2014 International Conference on Big Data Science and Computing*, page 5. ACM, 2014.
- [81] T.J. Lyons and Z. Qian. *System control and rough paths*. Oxford University Press, 2002.
- [82] T.J. Lyons and N. Sidorova. Sound compression: a rough path approach. In *Proceedings of the 4th international symposium on Information and communication technologies*, pages 223–228. Trinity College Dublin, 2005.
- [83] T.J. Lyons and N. Victoir. Cubature on Wiener space. *Proc. Roy. Soc. London*, 460(2041):169–198, 2004.
- [84] T.J. Lyons and N. Victoir. An extension theorem to rough paths. In *Ann. Inst. Henri Poincaré Probab. Stat.*, volume 24, pages 835–847, 2007.
- [85] T.J. Lyons and D. Yang. Rough differential equations in Banach space driven by weak geometric p-rough paths. *arXiv preprint arXiv:1402.2900*, 2014.
- [86] T.J. Lyons and D. Yang. The theory of rough paths via one-forms and the extension of an argument of Schwartz to RDEs. *J. Math. Soc. Japan*, 2015.
- [87] W. Magnus. On the exponential solution of differential equations for a linear operator. *Comm. Pure Appl. Math.*, 7(4):649–673, 1954.
- [88] F. Malrieu. Convergence to equilibrium for granular media equations and their Euler schemes. *Ann. Appl. Probab.*, 13(2):540–560, 2003.
- [89] G. Maruyama. Continuous Markov processes and stochastic equations. *Rend. Circ. Math. Palermo*, 4(1):48–90, 1955.
- [90] G.N. Milshstein. Approximate integration of stochastic differential equations (Russian). *Teor. Veroyatnost. i Primenen.*, 19(3):583–588, 1974.

- [91] G. Monge. Mémoire sur la théorie des déblais et des remblais. *Mémoires de l'Académie Royale des Sciences*, XVIII-XIX:666–704, 1781.
- [92] A. Neuenkirch, S. Tindel, and J. Unterberger. Discretizing the fractional Lévy area. *Stochastic Processes. Appl.*, 120(2):223–254, 2010.
- [93] H. Ni. A multidimensional stream and its signature representation. *Preprint*, 2015.
- [94] S. Ninomiya and N. Victoir. Weak approximation of stochastic differential equations and application to derivative pricing. *Appl. Math. Finance*, 15(2):107–121, 2008.
- [95] G. Pagès and A. Sellami. Convergence of multi-dimensional quantized SDEs. In *Séminaire de probabilités XLIII*, pages 269–307. 2011.
- [96] G. Pap. Central limit theorems on nilpotent Lie groups. *Probab. Math. Stat*, 14(2):287–312, 1993.
- [97] N. Perkowski and D.J. Prömel. Pathwise stochastic integrals for model free finance. *arXiv preprint arXiv:1311.6187*, 2013.
- [98] W.H. Press, S.A. Teukolsky, W.T. Vetterling, and B.P. Flannery. *Numerical recipes 3rd edition: The art of scientific computing*. Cambridge University Press, 2007.
- [99] S. Rachev and L. Ruschendorf. *Mass transportation problems, I and II: theory and applications*. Springer, 1998.
- [100] P.K. Rashevskii. About connecting two points of complete holomorphic space by admissible curves. *Uch Zapiski Ped. Inst. Libknekhta*, 2(83–94), 1938.
- [101] C. Reutenauer. *Free Lie algebras*. London Math. Soc. Monogr. Ser., 1993.
- [102] S. Riedel. Talagrand’s transportation-cost inequality and applications to (rough) path spaces. *arXiv preprint arXiv:1403.2585*, 2014.
- [103] S. Riedel and W. Xu. A simple proof of distance bounds for Gaussian rough paths. *Electron. J. Probab*, 18(108):1–22, 2013.
- [104] L.C.G. Rogers. Arbitrage with fractional Brownian motion. *J. Math. Fin.*, 7(1):95–105, 1997.
- [105] L.C.G. Rogers and D. Williams. *Diffusions, Markov processes and martingales: Volume 2, Itô calculus*. Cambridge University Press, 2000.
- [106] T. Rydén and M. Wiktorsson. On the simulation of iterated Itô integrals. *Stochastic Processes. Appl.*, 91(1):151–168, 2001.
- [107] E.M. Sipiläinen. *A pathwise view on solutions of stochastic differential equations*. PhD thesis, University of Edinburgh, 1993.
- [108] E.M. Stein. *Singular integrals and differentiability properties of functions*, volume 2. Princeton University Press, 1970.
- [109] R.S. Strichartz. The Campbell-Baker-Hausdorff-Dynkin formula and solutions of differential equations. *J. Funct. Anal.*, 72(2):320–345, 1987.
- [110] H.J. Sussmann. On the gap between deterministic and stochastic ordinary differential equations. *Ann. Probab.*, 6(1):19–41, 1978.
- [111] D. Talay. Stochastic Hamiltonian systems: exponential convergence to the invariant measure, and discretization by the implicit Euler scheme. *Markov Process. Related Fields*, 8(2):163–198, 2002.
- [112] D. Talay and L. Tubaro. Expansion of the global error for numerical schemes solving stochastic differential equations. *Stochastic Processes. Appl.*, 8(4):483–509, 1990.
- [113] L.N. Vaserstein. Markov processes over denumerable products of spaces describing large system of automata (Russian). *Problemy Peredaci Informacii*, 5:64–72, 1969.
- [114] C. Villani. *Topics in Optimal Transportation*. Number 58. American Math. Soc., 2003.
- [115] C. Villani. *Optimal transport: old and new*, volume 338. Springer, 2008.
- [116] S. Watanabe. *Lectures on stochastic differential equations and Malliavin calculus*. Springer, 1984.

- [117] M. Wiktorsson. Joint characteristic function and simultaneous simulation of iterated Itô integrals for multiple independent Brownian motions. *Ann. Appl. Probab.*, pages 470–487, 2001.
- [118] D.R.E. Williams. Path-wise solutions of stochastic differential equations driven by Lévy processes. *Rev. Mat. Iberoam.*, 17(2):295–330, 2001.
- [119] E. Wong and M. Zakai. On the relation between ordinary and stochastic differential equations. *Int. J. Eng. Sci.*, 3(2):213–229, 1965.
- [120] L.C. Young. An inequality of the Hölder type, connected with Stieltjes integration. *Acta Math.*, 67(1):251–282, 1936.
- [121] A.Y. Zaitsev. Multidimensional version of a result of Sakhanenko in the invariance principle for vectors with finite exponential moments. I,II,III. *Teor. Veroyatnost. i Primenen* **45** (2000), 718-738; **46** (2001), 535-561 and 744-769. Translations in *Theory Probab. Appl.* **45** (2002), 626-641; **46** (2003), 490-514 and 676-698.
- [122] A.Y. Zaitsev. Estimates for the quantiles of smooth conditional distributions and the multidimensional invariance principle. *Sib. Math. J.*, 37(4):706–729, 1996.
- [123] A.Y. Zaitsev. Multidimensional version of the results of Komlós, Major and Tusnády for vectors with finite exponential moments. *ESAIM: Probab. Stat.*, 2:41–108, 1998.