AUTHOR: **David Kelly** DEGREE: **Ph.D.**

TITLE: **Itô corrections in stochastic equations**

DATE OF DEPOSIT: ....

Itô corrections in stochastic equations

by

David Kelly

Thesis

Submitted to the University of Warwick

for the degree of

Doctor of Philosophy

Mathematics Institute

September 2012

THE UNIVERSITY OF WARWICK

Contents

Acknowledgments

My Ph.D was supported by a Warwick postgraduate research scholarship scheme and I am grateful for their generousity.

Starting from the top, I would like to thank Martin Hairer for being an excellent advisor. His child-like enthusiasm and wizard-like mind have been a constant inspiration to me. My research has been pushed forward by the insight of many staff from Warwick and abroad, including Andrew Stuart, Max Gubinelli, Takis Souganidis, Jérémie Unterberger, Robin Hudson, John Jones, David Elworthy and Hendrik Weber. Not to forget the little people - the grad students - who make life as a Ph.D student enjoyable, to name a few Dayal Strub, Sebastian Vollmer, Sergios Agapiou, Yuxin Yang, Martin Klimmek, Dintle Kagiso and Cyril Labbé. Despite living on the other side of the world, my parents have shown me great support for the past three years and their interest in my work has been convincing to say the least. Finally, I can't imagine how much harder it all would have been without my wonderful girlie Emma.

Declarations

I declare that, to the best of my knowledge, all material contained in this thesis is my own original work, unless otherwise stated, cited or commonly known.

Abstract

This thesis contains two "projects", both concerning the emergence of Itô corrections in stochastic equations. In the first project, we study Itô corrections in stochastic PDEs with multiscale structure. Namely, we show that a certain class of homogenisation systems display a correction of Itô type, when perturbed by a sufficiently irregular additive noise. In the second project, we look at Itô corrections for a general class of finite dimensional equations known as rough differential equations. Using a non-geometric theory of rough paths, we prove a generalised Itô-Stratonovich correction formula as well as a generalised Itô formula.

Chapter 1

Introduction

During the second world war, in a mathematically isolated Japan, K. Itô produced one of the most important scientific concepts of the twentieth century - the Itô integral. The question of how to integrate against Brownian motion had been asked (and to some extent answered) before, most notably by N. Wiener. However, these attempts never lead to a definition that would accommodate the notion of a differential equation perturbed by white noise. In contrast, Itô's theory's was more than accommodating, providing a platform for stochastic differential equations (SDEs) and what is more broadly referred to as the Itô calculus.

At the heart of Itô calculus, in fact any result concerning SDEs, is Itô's celebrated change of variables formula. It would perhaps be naive to expect that a change of variables formula involving Brownian motion, an object that is of infinite total variation on every compact interval, should obey the laws of calculus for objects of finite total variation. One quickly learns that the Itô calculus is not obliging to the naive. Indeed, for a function $F \in \mathcal{C}^2$, Itô proved that

$$F(W_t) = F(W_s) + \int_s^t F'(W_r)dW_r + \frac{1}{2}\int_s^t F''(W_r)dr \ , \tag{1.1}$$

where W is a $\mathbb{R}$ valued Brownian motion the "dW" integral is the Itô integral. The identity (1.1) is known as *Itô's formula* (or Lemma) and the "extra term" $\frac{1}{2}\int_s^t F''(W_r)dr$ is known as the *Itô correction*. Itô's formula and the associated correction term are prevalent in many major fields of modern mathematics, from differential geometry to the theory of option pricing.

A common variant of (1.1) is the *Itô-Stratonovich correction*, which is more or less a corollary of the Itô formula. One of the more striking features of stochastic integrals is that different Riemann-type integration schemes result in different limits. Most famously, a left-point integration scheme

$$\int_0^t Z_r dW_r \stackrel{\text{def}}{=} \lim_{\pi_{[0,t]}\downarrow 0} \sum_{[u,v]\in\pi_{[0,t]}} Z_u(W_v - W_u) \ ,$$

defines the Itô integral (for a suitable class of Z), where $\pi_{[0,t]}$ is some partition of $[0,t]$. On the other hand, a left-right average scheme

$$\int_0^t Z_r \circ dW_r \stackrel{\text{def}}{=} \lim_{\pi_{[0,t]}\downarrow 0} \sum_{[u,v]\in\pi_{[0,t]}} \frac{Z_u + Z_v}{2}(W_v - W_u) \ ,$$

defines the *Stratonovich integral*. The nicest feature of the Stratonovich integral is that is *does* obey the rules of ordinary calculus including the chain rule and integration by parts. It is easy to show that

$$\int_0^t Z_r dW_r = \int_0^t Z_r \circ dW_r - \frac{1}{2}[Z,W]_t \;, \tag{1.2}$$

where $[Z,W]$ is the quadratic covariation between Z and W. This yields what is known as the Itô-Stratonovich correction. The correction most famously arises when one attempts to approximate the solution to an SDE by discretising the underlying Brownian motion. The Wong-Zakai theorem tells us that the approximation is no good and the limit converges to a different SDE, given by interpreting the Itô integral in the original SDE as a Stratonovich integral, or in other words, by adding the Itô-Stratonovich correction.

In this thesis, we will investigate Itô corrections in a much broader scope, but where the underlying phenomenon is unchanged. Namely that corrections arise due to some kind of approximation happening in an equation that is perturbed by some highly irregular noise term. The first half of the thesis concentrates on a stochastic correction in the homogenisation of stochastic PDEs. The second half of the thesis looks at applying the powerful tools of rough path theory to develop an algebraic basis for Itô corrections.

1.1 Itô corrections in stochastic PDEs

A good example of this is found in [HM12], where the authors examine a class of "Burgers-like" stochastic PDEs of the form

$$\partial_t u = \nu \partial_x^2 u + \nabla G(u)\partial_x u + F(u) + \xi \;, \tag{1.3}$$

where $u : [0,T] \times [0,2\pi] \to \mathbb{R}^d$ with periodic boundary conditions, $F, G : \mathbb{R}^d \to \mathbb{R}^d$ are as smooth as required and ξ denotes space-time white noise.

In the classical theory of PDEs, there is a commonly held belief that any well-posed equation should be stable under all natural approximations of the equation. The authors questions this hypothesis by considering the discretisation

$$\partial_t u_\varepsilon = \nu \partial_x^2 u_\varepsilon + g(u_\varepsilon) D_\varepsilon u_\varepsilon + f(u_\varepsilon) + \xi \;, \tag{1.4}$$

where

$$D_\varepsilon u_\varepsilon(t,x) = \frac{u_\varepsilon(t,x+\varepsilon) - u_\varepsilon(t,x)}{\varepsilon} \;.$$

The main result of [HM12] states that $u_\varepsilon \Rightarrow \bar{u}$ where

$$\partial_t \bar{u} = \nu \partial_x^2 \bar{u} + \nabla G(\bar{u})\partial_x \bar{u} + \bar{F}(\bar{u}) + \xi \;, \tag{1.5}$$

where

$$\bar{F}(u) = F(u) - \frac{1}{4\nu}\Delta G(u) \ .$$

Thus, the SPDE is less well-posed than one might think. This result can be understood as an Itô correction on the equation (1.3). Indeed, one can check that the path $u(t, \cdot) : [0, 2\pi] \to \mathbb{R}^d$, for any fixed $t \in [0, T]$, has a finite (local) quadratic variation given by $1/(2\nu)$. Hence, the correction term in $\frac{1}{4\nu}\Delta G(u)$ is precisely $1/2$ times the quadratic varitation of u and $\nabla G(u)$, and appears in exactly the same manner as in the usual Itô-Stratonovich correction (1.2).

This phenomenon has all the features of an Itô correction, but appears due to a discretisation of the actual equation rather than the driving path. Nevertheless, the result should be considered as a Wong-Zakai type result. For an intuitive picture of the phenomenon, one should think of the correction term as being an "interference" term, arising due to the equation being perturbed by a highly oscillatory additive noise term. When the discretisation parameter ε is very small, any reasonable function will look roughly constant across the discretisation step. On the other hand, space-time white noise has no scale, and will oscillate just as rapidly on the ε scale as any other. This leads to a sort of resonance effect between the noise and the discretisation, from which the correction term arises.

In the first half of this thesis, we investigate Itô corrections for stochastic PDEs in a new setting, namely in the field of *homogenisation*. The theory of homogenisation was built to model the behaviour of dynamical systems living in a *multi-scale* material. By multi-scale, we specifically mean two scales - one large, one small. The prototypical example considers a diffusion on a material that has been partitioned into cells of scale ε, the drift and diffusion coefficients depend not just on the position in the material but also the position within each cell. More precisely, the diffusion is governed by the parabolic equation

$$\partial_t u_\varepsilon = \frac{1}{\varepsilon} B(x, x/\varepsilon) \cdot \nabla u_\varepsilon + \frac{1}{2} A(x, x/\varepsilon) : \nabla^2 u_\varepsilon \ , \tag{1.6}$$

where $B : \mathbb{R}^d \times \mathbb{R}^d \to \mathbb{R}$ and $A : \mathbb{R}^d \times \mathbb{R}^d \to \mathbb{R}^{d\times d}$ are periodic in the second variable and B is required to satisfy a *centering* condition. For instance, in the $d = 1$ case the domain $[0, L]$ is split into ε^{-1} cells of size $2\varepsilon L$. The theory has been very successful at establishing the behaviour of (1.6) as $\varepsilon \to 0$, under a broad range of assumption on the regularity of the coefficients B, A [BLP78, PSV77]. The typical result states that $u_\varepsilon \to \bar{u}$, where $\bar{u}$ sovles the equation

$$\partial_t \bar{u} = \frac{1}{2}\bar{A} : \nabla^2 \bar{u} \ ,$$

where $\bar{A}$ can be written down explicitly in terms of the functions A and B, typically involving some kind of averaging over the coefficients.

It is natural to consider the influence of additive or multiplicative noise to the system (1.6), for instance to model some kind of thermal fluctuation in the underlying material. This has been studied broadly in the past [BMP05, WCD07, WD07], typically one assumes a *nice* infinite dimensional noise, whose covariance is trace-class. With additive noise, we obtain an SPDE with a small scale parameter ε. It makes sense to ask: do Itô corrections occur in such systems when $\varepsilon \to 0$? Of course, the ε does not arise due to a discretisation as in (1.3), but rather the equation itself is an approximation of an "ideal" situation $\varepsilon = 0$, which makes no sense mathematically.

The particular case studied in this thesis concerns the one dimensional SPDE

$$du_\varepsilon = \left(\frac{1}{\varepsilon}b(x/\varepsilon)\partial_x + \frac{1}{2}\sigma(x/\varepsilon)\partial_x^2\right) u_\varepsilon dt + d\xi_\varepsilon \ , \tag{1.7}$$

where $u_\varepsilon : [0,T] \times [0,2\pi] \to \mathbb{R}$, $b, \sigma : \mathbb{R} \to \mathbb{R}$ are as smooth as required and where periodic boundary conditions are assumed. We use the shorthand $\mathcal{L}_\varepsilon = \frac{1}{\varepsilon}b(x/\varepsilon)\partial_x + \frac{1}{2}\sigma(x/\varepsilon)\partial_x^2$ and $\mathcal{L} = b(x)\partial_x + \frac{1}{2}\sigma(x)\partial_x^2$, which is the generator of the diffusion if the whole system consisted of just one cell. Note that the system depends periodically on the microscopic scale and does not depend on the macroscopic scale at all. The term ξ_ε will denote any infinite dimensional noise that satisfies a *cell translation invariance* property. Namely, that the law of $\{d\xi_\varepsilon(t, x_k)/dt\}_k$ for a finite collection of point x_k is invariant under translation of the points x_k by an integer number of cells. Intuitively, this means that the source of noise may change within each cell, but at a fixed point in two different cells, the source is the same. A good example to keep in mind is when ξ_ε is simply space-time white noise, which is translation invariant and doesn't depend on ε at all.

The main result of the homogenisation part of the thesis states that (1.7) does indeed exhibit something like an Itô correction, but not in the sense of a quadratic variation as in (1.3). However, the source of the correction is very similar to (1.3). For the sake of exposition, we will assume $\xi_\varepsilon = W$, the cylindrical Wiener process on $L^2[0, 2\pi]$, whose derivative is space-time white noise. Hence we study the solution to the SPDE

$$du_\varepsilon = \mathcal{L}_\varepsilon u_\varepsilon dt + dW \ , \tag{1.8}$$

with periodic boundary conditions and zero initial condition. All previous results in homogenisation theory point to the limiting equation of (1.8) being

$$d\bar{u} = \frac{\mu}{2}\partial_x^2 \bar{u} dt + dW \ , \tag{1.9}$$

for some constant $\mu > 0$. In particular, if one replaced W with something smoother, like for instance any continuous random process $f : [0,T] \to L^2[0, 2\pi]$, then one can easily show that

$$(du_\varepsilon = \mathcal{L}_\varepsilon u_\varepsilon dt + df) \quad \to \quad \left(d\bar{u} = \frac{\mu}{2}\bar{u}dt + df\right) \ . \tag{1.10}$$

The heuristic explanation for this hypothesis is that diffusion generated by $\mathcal{L}_\varepsilon$ can be shown to converge to Brownian motion, at the level of Markov kernels. However, our results show this guess to be false for (1.8), due to a correction term. In particular, we show that u_ε converges weakly (in both the PDE and probabilistic sense) to $\bar{u}$, where $\bar{u}$ satisfies

$$d\bar{u} = \frac{\mu}{2}\partial_x^2 \bar{u} dt + \|\rho\| d\widehat{W} \ , \tag{1.11}$$

where $\widehat{W}$ is a cylindrical Wiener process on $L^2[0, 2\pi]$, defined on a different probability space to W. We use $\|\cdot\|$ to denote the norm corresponding to the scalar product $\langle f, g\rangle = \frac{1}{2\pi}\int_0^{2\pi} fg dx$ and ρ is the invariant density of the operator $\mathcal{L}$, normalised in such a way that $\langle \rho, 1\rangle = 1$.

On the face of it, this is not an Itô correction in the usual sense, since there is no bounded variation correction term, corresponding to the quadratic variation of two quantities. However, it does arise via a similar mechanism to that found in (1.3). There is still an interference occurring at the ε scale, but in this context it is an interference between the semigroup S_ε of $\mathcal{L}_\varepsilon$ and the noise dW. To illustrate this, we write down the mild solution

$$u_\varepsilon(t, x) = \int_0^t S_\varepsilon(t-s) dW(s, x) = \sum_{k\in\mathbb{Z}} \int_0^t S_\varepsilon(t-s) e_k(x) dW_k(s) \ ,$$

where we use the Fourier expansion $W(t, x) = \sum_{k\in\mathbb{Z}} e_k(x) W_k(t)$ for independent Brownian motions W_k, satisfying $W_k^* = W_{-k}$ and where $e_k(x) = e^{ikx}$. For a more regular source of noise, one would usually proceed in the following way. Fix $m \in \mathbb{Z}$ and consider the m-th mode of the solution

$$\langle u_\varepsilon(t), e_m\rangle = \sum_{k\in\mathbb{Z}} \int_0^t \langle S_\varepsilon(t-s) e_k, e_m\rangle dW_k(s) \ .$$

One would like to proceed by using an approximation $S_\varepsilon(t-s)e_k \approx e_k(x) e^{-\mu k^2(t-s)}$ when k is of order 1 and then showing that $S_\varepsilon(t-s)e_k$ goes like some inverse power of k when k is large. However, the decay obtained on $S_\varepsilon(t-s)e_k$ is simply not sufficient to sum up and the approach fails. Another approach is to instead consider approximating the adjoint semigroup in the expression

$$\sum_{k\in\mathbb{Z}} \int_0^t \langle e_k, S_\varepsilon^*(t-s) e_m\rangle dW_k(s) \ .$$

In particular, we can show that $S_\varepsilon(t-s)e_m(x) \approx \rho(x/\varepsilon)e_m(x)e^{-\mu m^2(t-s)}$ and hence we obtain

$$\langle u_\varepsilon(t), e_m\rangle \approx \sum_{k\in\mathbb{Z}} \langle e_k, \rho(\cdot/\varepsilon)e_m\rangle \int_0^t e^{-\mu m^2(t-s)} dW_k(s) \; .$$

Moreover, by expanding $\rho(x) = \sum_{p\in\mathbb{Z}}\langle e_p,\rho\rangle e_p(x)$ and eliminating those modes that vanish, we obtain

$$\begin{aligned}\langle u_\varepsilon(t), e_m\rangle &\approx \sum_{p\in\mathbb{Z}} \langle e_p, \rho\rangle \int_0^t e^{-\mu m^2(t-s)} dW_{m+p/\varepsilon}(s) \\ &\stackrel{\text{dist}}{=} \int_0^t e^{-\mu m^2(t-s)} dW_m(s) + (\|\rho\|^2-1)^{1/2} \int_0^t e^{-\mu m^2(t-s)} d\widehat{W}_m(s) \; ,\end{aligned}$$

where $\widehat{W}$ is obtained by summing up all the $p \neq 0$ terms. Notice that the first term in the line above is exactly the m-th Fourier mode of the "naive" limiting solution (1.9) and the second term therefore plays the role of the Itô correction. Note that in the limiting equation (1.11), we adopt the more convenient approach of writing the two terms as one, using the fact that

$$\begin{aligned}&\int_0^t e^{-\mu m^2(t-s)} dW_m(s) + (\|\rho\|-1)^{1/2} \int_0^t e^{-\mu m^2(t-s)} d\widehat{W}_m(s) \\ &\quad\stackrel{\text{dist}}{=} \|\rho\| \int_0^t e^{-\mu m^2(t-s)} d\widehat{W}_m(s) \; ,\end{aligned}$$

which is the m-th Fourier mode of (1.11).

Another justification for calling this an Itô correction is as follows. If we were to replace W with anything smoother, then the correction disappears and we obtain a result along the lines of (1.10). For instance, if the additive noise in (1.7) was given by $\Lambda W = \sum_{k\in\mathbb{Z}} \lambda_k e^{ikx} W_k(t)$, then as long as $\lambda_k \to 0$ when $k \to \infty$, then we obtain the "naive" limit

$$\partial_t \bar{u} = \frac{\mu}{2}\partial_x^2 \bar{u} dt + \Lambda dW \; ,$$

with no correction term. We can actually prove a somewhat more insightful result, namely that if $|\lambda_k| \lesssim |k|^{-\alpha}$ as $k \to \infty$, then the correction term is of order ε^α. This sheds light on the transition between the classical "naive" case and the correction case.

1.2 Itô corrections and rough path theory

In the late 1990s, T. Lyons introduced a more algebraic approach to SDEs, known as the theory of rough paths [Lyo98]. The theory is concerned with equations of the type

$$dY_t = f(Y_t)dX_t \ , \tag{1.12}$$

where X and Y are paths taking values in the Banach spaces V and U respectively and $f : U \to L(V, U)$. The driving path $X : [0,T] \to V$ is assumed to be Hölder continuous with some exponent $\gamma \in (0,1)$. Typically, one must assume something probabilistic about the path X in order to proceed, such as requiring that X be distributed as fractional Brownian motion. The theory of rough paths tells us that we can proceed without making any assumption on the distribution of X, instead we can work with a *fixed* path X. The catch is that we must be able to *lift* the path X to a larger path $\mathbf{X}$ living in a much higher dimensional space. The object $\mathbf{X}$ encodes information about the iterated integrals of dX and is called a *rough path* above X. With a given rough path $\mathbf{X}$ above X, the equation (1.12) is called a rough differential equation (RDE) driven by $\mathbf{X}$.

To treat any $\gamma \in (0,1)$, Lyons' theory requires that the integrals be interpreted in a Stratonovich-like sense, leading to what is called a *geometric* rough path. Such objects are defined as paths in the dual space of a tensor algebra $T(V)$, with a special "group-like" property that generalises the idea that the ordinary integration by parts formula is satisfied. The downside of the theory is that it cannot treat integrations schemes for which the ordinary calculus doesn't hold, in other words, it cannot handle non-geometric rough paths. Since we are looking to treat Itô corrections in a more general setting, we certainly want a non-geometric theory.

More recently, M Gubinelli has proposed a non-geometric theory of rough paths, where the rough path $\mathbf{X}$ has many more components than a geometric rough path [Gub10]. These non-geometric rough paths are known as *brached rough paths* and take values in the *Connes-Kreimer Hopf algebra* $\mathcal{H}(V)$ generated by V. Broadly speaking, $\mathcal{H}(V)$ is an extension of the tensor product algebra $T(V)$ and is generated by rooted trees rather than tensors. For a branched rough path $\mathbf{X}$, one can view solutions to (1.12) as objects that locally "look like" linear combinations of the components of $\mathbf{X}$. This is the basis of the theory of *controlled rough paths*, first introduced in [Gub04]. .

The rough path section of the thesis contains two main results concerning Itô corrections. The first result is a generalisation of the Itô-Stratonovich correction. In particular, we show that every branched rough path $\mathbf{X}$ above a path X can be encoded in a geometric rough path $\bar{\mathbf{X}}$ living above a larger path $\bar{X}$. The geometric rough path is defined on a tensor algebra $T(\bar{V})$, where the vector space $\bar{V}$ is obtained through the process of extending V

by adding new basis that correspond to the tree components of $\mathcal{H}(V)$. When we say "encoded", we mean that all the components of $\mathbf{X}$ are stored throughout the components of $\bar{\mathbf{X}}$. This is described by

$$\langle \mathbf{X}, \tau \rangle = \langle \bar{\mathbf{X}}, \psi(\tau) \rangle ,$$

for each component (or rooted tree) τ, where $\psi(\tau)$ is a linear combination of components of $T(\bar{V})$. This allows us to re-write an objected controlled by $\mathbf{X}$ as an object controlled by $\bar{\mathbf{X}}$. The most important consequence of this fact is the following: Y is a controlled rough path solution to the equation

$$dY_t = f(Y_t) dX_t ,$$

driven by $\mathbf{X}$ if and only if it is also a controlled rough path solution to

$$dY_t = \bar{f}(Y_t) d\bar{X} , \tag{1.13}$$

driven by $\bar{\mathbf{X}}$, where the new vector fields $\bar{f}$ can be written down explicitly in terms of f. In the case of Brownian motion, this does indeed agree with the usual Itô-Stratonovich correction.

The second of the two main results in the rough path section concerns the Itô formula, from a rough path perspective. Specifically, if X is some path with a branched rough path $\mathbf{X}$ above it and $F : \mathbb{R}^d \to \mathbb{R}^d$ is a smooth function, we would like to write

$$dF(X) = DF(X) dX + \text{"correction terms"} ,$$

where the "correction terms" consist of differentials of "brackets" between components of X. The brackets are more regular objects than X and in some sense correspond to "variations" of X, like the quadratic variation for instance. The rough path framework is ideal for this problem, since we can easily build the leading integral $\int DF(X)dX$ and then attempt to form the bracket integrals from the leftover terms.

To build an Itô formula, one needs to know what these bracket terms are and moreover how to integrate objects against them. This information is stored in what we refer to as a *bracket extension* $\widehat{\mathbf{X}}$. The statement of our rough path Itô formula is as follows, for a given branched rough path $\mathbf{X}$ above X and a given bracket extension $\widehat{\mathbf{X}}$ of $\mathbf{X}$, we have that

$$dF(X) = DF(X) dX + \sum_{n=2}^{N} \sum_{\alpha_1, \dots, \alpha_n = 1}^{d} \frac{\partial^{\alpha_1} \dots \partial^{\alpha_n} F(X)}{n!} d\widehat{X}^{(\alpha_1, \dots, \alpha_n)} , \tag{1.14}$$

where ∂^i denotes partial differentiation in the i-th component. The bracket paths $\widehat{X}^{(\alpha_1, \dots, \alpha_n)}$ can be read off as the path elements of the bracket extension $\widehat{\mathbf{X}}$. The bracket extension also tells us how the differential $d\widehat{X}^{(\alpha_1 \cdots \alpha_n)}$ leads to the definition of integrals.

The theory of controlled rough paths allows us to extend this result to obtain something more like a usual Itô correction, finding a formula for the increment $dF(Y)$ where Y solves the RDE (1.12). We obtain a similar result to (1.14), but containing more *bizarre* brackets corresponding to variations between any components of $\mathbf{X}$, rather than just the path components.

1.3 Outline of thesis

The thesis is split sharply into two "halves". The first half, consisting solely of Chapter 2, contains all results concerning SPDEs, namely the homogenisation limit theorems discussed above. We also develop several tools that should prove useful when applying semigroup theory to multiscale problems. The second half, consisting of Chapters 3, 4 and 5 contains all results relating to rough path theory. In Chapter 3 we provide an overview of the theory of branched and controlled rough paths, we also show that branched rough paths can be defined in a similar way to geometric rough paths, namely as γ-Hölder paths in a nilpotent group. In Chapter 4 we prove the generalised Itô-Stratonovich correction formula. To achieve this, we first give an overview of geometric rough paths, we then prove the crucial conversion formula, which states that every branched rough path can be encoded in a geometric rough path. In Chapter 5 we prove a change of variables (or Itô) formula for branched rough paths. We first prove the result in the *simple case*, where we obtain a formula for the differential $dF(X)$ when X is some γ-Hölder path with a branched rough path above it. We then prove the result in the *general case*, where we obtain a formula for $dF(Y)$, with Y being the solution to a rough differential equation. Of course, the general case implies the simple case, but we include both for the sake of clarity.

We warn the reader that the two halves of the paper are only *morally* related to one another and should really be appreciated as two separate projects. We do not suspect that the rough path results can be applied to the SPDEs found in Chapter 2, since this is really a different kind of correction. Moreover, the thesis is not just split in two based on the nature of the results, but also in the method of proof. In particular, the first half of the paper is almost entirely analytic whereas the second half is almost entirely algebraic. Nevertheless, we hope that the readers can appreciate the common *theme* of the two halves.

Chapter 2

Stochastic PDEs with multiscale structure

2.1 Introduction

In the material sciences, there is a significant interest towards objects that contain one structure at a macroscopic scale, overlaying a totally different structure on a microscopic scale. Examples range from everyday life, such as concrete and fibreglass, to the cutting edge of science, such as the cloaking devices implemented by meta-materials. Composite materials pose an important mathematical problem. Given a system with certain dynamics on a macroscopic scale and separate, but not necessarily independent, dynamics on a microscopic scale, approximate the effective dynamics of the whole system when the microscopic scale is small. Such problems can be formulated, and dealt with, using homogenisation theory, see for example [Fre64, PSV77, ELVE04, TM05, PS05], as well as the monographs [BLP78, PS08] and references therein.

The following is the prototypical homogenisation problem. Take a Markov process X on $\mathbb{R}$ with generator

$$\mathcal{L} = b(x)\partial_x + \frac{1}{2}\sigma^2(x)\partial_x^2 \ , \tag{2.1}$$

where b and σ are suitably smooth functions, periodic on $[0, 2\pi]$. Consider then the diffusively rescaled process $X_\varepsilon(t) = \varepsilon X(t/\varepsilon^2)$, with generator given by

$$\mathcal{L}_\varepsilon = \frac{1}{\varepsilon}b(x/\varepsilon)\partial_x + \frac{1}{2}\sigma^2(x/\varepsilon)\partial_x^2 \ . \tag{2.2}$$

We also require that σ is bounded away from zero and that the "centering condition" $\int_0^{2\pi} b(v)/\sigma^2(v)dv = 0$ is satisfied.

One example to keep in mind is the when $\sigma = 1$ and

$$V(x/\varepsilon) = -\int^{x/\varepsilon} b(v)dv \ .$$

The centering condition guarantees that $\int_0^{2\pi} b(v)dv = 0$, so that $V(x/\varepsilon)$ itself is $2\pi\varepsilon$ periodic. In this case, the diffusion X_ε provides a simple model for diffusion in a one-dimensional composite material, where the material is composed of cells of size $2\pi\varepsilon$ and the dynamics in each cell is governed by the potential $V(x/\varepsilon)$.

It is a classical result that

$$X_\varepsilon(t) \Rightarrow \mu B(t) \ , \tag{2.3}$$

where $B(t)$ is a Brownian motion on $\mathbb{R}$, $\mu > 0$ is a constant determined by b and σ, and $\Rightarrow$ denotes convergence in distribution on the space of continuous functions [BLP78]. This result is powerful when analysing parabolic PDEs of the following type

$$\partial_t u_\varepsilon(x,t) = \mathcal{L}_\varepsilon u_\varepsilon(x,t) + f(x,t)\;, \tag{2.4}$$

with some forcing term f. We will assume $u_\varepsilon(x,0) = 0$ as we are more interested in the forcing term. Duhamel's principle then states that

$$u_\varepsilon(x,t) = \int_0^t \mathbb{E}[f(X_\varepsilon(t-s),s)|X_\varepsilon(0)=x]\,ds\;,$$

where $\mathbb{E}$ averages over the paths X_ε (but not any possible randomness in the forcing term). If f is sufficiently regular, it follows from (2.3) that $u_\varepsilon \to u$ as $\varepsilon \to 0$, where u satisfies the PDE

$$\partial_t u(x,t) = \frac{\mu}{2}\partial_x^2 u(x,t) + f(x,t)\;. \tag{2.5}$$

Such results have been widely generalised in both the forcing terms considered and also the structural assumptions placed on the generator $\mathcal{L}_\varepsilon$, see for example [Par99, Del04, HP08, SRP09]. The article [SRP09] contains a brief but recent overview of the field. On the other hand, one can find only very few results in the literature treating the case of stochastic PDEs where both the noise term and the linear operator exhibit a multiscale structure, and this is the main focus of this chapter. In some situations where the limiting noisy term is sufficiently regular, the previously mentioned results have been extended to the stochastic case, see for example [Ich04, WCD07, WD07]. In this chapter, our aim is to provide a preliminary understanding of the type of phenomena that can arise in the situation where the limiting equation is driven by very rough noise, so that resonance effects can also play an important role.

Over the last few decades, there has been much progress towards making sense of solutions to stochastic PDEs, where the forcing term may be a highly irregular Gaussian signal taking values in spaces of rather irregular distributions, see for example [DPZ92, Hai09] for introductory texts on the subject. It is therefore natural to ask whether asymptotic results for PDEs like (2.4) can be extended to the case where f is a random, distribution-valued process. To give an idea of the type of results obtained in this thesis, let ξ be space-time white noise, which is the distribution-valued Gaussian process formally satisfying $\mathbb{E}\xi(s,x)\xi(t,y) = \delta(s-t)\delta(x-y)$. For fixed $\varepsilon > 0$, one can easily show that

$$\partial_t u_\varepsilon = \mathcal{L}_\varepsilon u_\varepsilon + \xi \tag{2.6}$$

has a unique solution u_ε with almost surely continuous sample paths in $L^2[0,2\pi]$. By analogy with the classical theory outlined above and since ξ does not show any explicit

ε-dependence, one might guess that u_ε has a limit u, satisfying

$$\partial_t u = \mu \partial_x^2 u + \xi \; . \tag{2.7}$$

It turns out that this is not the case. Instead, we will show that the true limit solves

$$\partial_t u = \mu \partial_x^2 u + \|\rho\| \xi \; , \tag{2.8}$$

where $\|\cdot\|$ denotes the $L^2[0, 2\pi]$ norm (normalised such that the corresponding scalar product is given by $\langle f, g\rangle = \frac{1}{2\pi}\int_0^{2\pi} f(x)g(x)\,dx$) and ρ is the invariant measure for the process with generator $\mathcal{L}$, normalised to satisfy $\langle \rho, 1\rangle = 1$.

Remark 2.1.1. By the Cauchy-Schwartz inequality, one always has $\|\rho\| \geq 1$, with equality if and only if ρ is constant. As a consequence, (2.8) differs from (2.7) as soon as $\mathcal{L}$ is not in divergence form. Furthermore, the effect of the noise is always enhanced by non-trivial choices of $\mathcal{L}$, which is a well-known fact in different contexts [PS08].

The crucial fact is of course the lack of regularity of ξ. Since the law of the process X_ε generated by $\mathcal{L}_\varepsilon$ will vary with x/ε, its law will typically have large Fourier components at wave numbers close to integer multiples of $1/\varepsilon$. The difference between (2.8) and (2.7) can then be understood, at least at an intuitive level, as coming from the resonances between these Fourier modes and the corresponding Fourier modes of the driving noise. Such resonances would be negligible for more regular noises, but turn out to lead to non-negligible contributions in the case of space-time white noise.

The aim of this chapter is to investigate this phenomenon for SPDEs of the type (2.6), but replacing ξ with a more general Gaussian forcing term. In particular, we treat noise that exhibits spatial structure at the microscopic scale. We can always (formally) write such signals as

$$\zeta(x, x/\varepsilon, t) = \sum_{k\in\mathbb{Z}} q^k(x, x/\varepsilon)\dot{W}_k(t) \; , \tag{2.9}$$

where the W_k are i.i.d. complex-valued Brownian motions, save for the condition $W_{-k} = W_k^\star$ ensuring that the overall signal is real-valued. Throughout this chapter, we will require the additional assumption that the noise ζ is *cell-translation invariant*, in the sense that its distribution is unchanged by translations by multiples of $2\pi\varepsilon$. This assumption reflects the idea that the underlying material has the same structure in each cell. At the level of the representation (2.9), this invariance is enforced by assuming that one has

$$q^k(x, x/\varepsilon) = q_k(x/\varepsilon)e^{ikx} \; , \tag{2.10}$$

for each $k \in \mathbb{Z}$, where $\{q_k\}$ is a collection of 2π-periodic functions.

To see that this leads to the claimed invariance property, notice that, for x, y satisfying $x - y = 2\pi\varepsilon n$, we have that

$$\begin{aligned}\sum_{k\in\mathbb{Z}} q_k(y/\varepsilon)e^{iky}\dot{W}_k(t) &= \sum_{k\in\mathbb{Z}} q_k(x/\varepsilon)e^{ikx}e^{2\pi ik\varepsilon n}\dot{W}_k(t)\\ &\stackrel{d}{=} \sum_{k\in\mathbb{Z}} q_k(x/\varepsilon)e^{ikx}\dot{W}_k(t)\ .\end{aligned}$$

Indeed, since W_k is a complex Brownian motion, rotating it by $2\pi k\varepsilon n$ does not change its distribution. Conversely, cell-translation invariance of the noise is equivalent to the fact that its covariance operator C_ε commutes with the translation operator T_ε given by $T_\varepsilon f(x) = f(x+2\pi\varepsilon)$. The spectrum of T_ε consists of $\{e^{ik\varepsilon} : k \in \mathbb{Z}\}$, with corresponding eigenspaces given by $V_k = \{q(x/\varepsilon)e^{ikx}\}$, where q is periodic with period 2π. As a consequence, there is no loss of generality in assuming the representation (2.10).

Thus, we restrict our attention to the following class of SPDEs, written in the notation of [DPZ92]:

$$du_\varepsilon(x,t) = \mathcal{L}_\varepsilon u_\varepsilon(x,t)dt + \sum_{k\in\mathbb{Z}} q_k(x/\varepsilon)e^{ikx}dW_k(t)\ . \tag{2.11}$$

Again, we will always assume that u_ε satisfies periodic boundary conditions on $[0, 2\pi]$. By linearity, we can and will restrict ourselves to the case of vanishing initial conditions. We will always assume certain regularity conditions on b and σ, as well as a centering condition, which is a standard requirement of homogenisation problems. This is detailed in Assumption 2.2.2 below.

Remark 2.1.2. Unlike several recent studies [WCD07, WD07] we do not consider periodically perforated spatial domains. Instead, we assume that our domain $[0, 2\pi]$ has been split into cells of size $2\pi\varepsilon$ and that diffusions behave identically in each cell. This is implemented through the periodicity of b, σ and q_k. Thus, all composite-type geometry comes through the periodicity of the generator $\mathcal{L}_\varepsilon$ and the infinite dimensional noise; the spatial domain $[0, 2\pi]$ does not depend on ε in any way. However, we do require that the domain be partitioned in to cells of size $2\pi\varepsilon$. It is therefore natural to require that $\varepsilon^{-1} \in \mathbb{N}$ so that $[0, 2\pi]$ contains an integer number of cells.

We have already seen that taking $q_k = 1$ results in the surprising limit (2.8). However, if we chose $q_k = |k|^{-1}$ then the forcing term would be a continuous Gaussian process in $L^2[0, 2\pi]$, and by classical results u_ε would converge to the unsurprising limit, as in (2.5). We would like to classify those choices of q_k that result in the surprising limit, and those that result in the unsurprising limit.

Firstly, we will identify a large class of signals that result in the unsurprising limit. In particular, these signals need not be continuous processes in $L^2[0, 2\pi]$. To guarantee the unsurprising limit, we need some control over the coefficients of the noise q_k when k is large, as well as a suitable regularity assumption. If we assume that the coefficients decay algebraically as $k \to \infty$, then we are able to show that solutions converge to the correct limit and that this convergence occurs in $L^2(P)$. In particular, the quantity $\|q_k\|$ must decay like $|k|^{-\alpha}$ as $k \to \infty$, for some $\alpha \in (0, 1)$. The precise condition is detailed in Assumption 2.2.5. With these conditions in place, we will prove the following.

Theorem 2.1.3. *Suppose the SPDE* (2.11) *satisfies Assumptions 2.2.2 and 2.2.5. Then the solutions u_ε converge to the solutions of*

$$du(x,t) = \mu \partial_x^2 u(x,t)dt + \sum_{k \in \mathbb{Z}} \langle q_k, \rho \rangle e^{ikx} dW_k(t) \ , \tag{2.12}$$

in the sense that there exists $C_T > 0$ and $\theta > 0$ such that

$$\mathbb{E}\mathrm{sup}_{t\in[0,T]} |\langle u_\varepsilon(t) - u(t), \varphi \rangle|^2 \leq C_T \varepsilon^\theta \ ,$$

for all $\varphi \in H^s$ with $\|\varphi\|_{H^s} \leq 1$ and for large enough s.

Remark 2.1.4. Past results [Ich04, WCD07] rely on the noise being Hilbert-Schmidt in the sense that

$$\sum_k \|q_k\|^2 < \infty \ .$$

It is important to note that this condition does not imply our condition on the $\|q_k\|$. Indeed, one can easily exhibit a sufficiently sparse sequence $\|q_k\|$ that is square summable but which only converges logarithmically to zero. On the other hand, there are many situations where the noise is not Hilbert-Schmidt, that do fall into our framework. With only the Hilbert-Schmidt assumption, one can still prove via a tightness argument that the SPDE (2.11) has a weak limit and apply homogenisation techniques, similar to those found in [WCD07], to show that the limiting SPDE is indeed (2.12). However, we will not treat this case as it is somewhat incongruous with the existing framework.

Remark 2.1.5. Although not immediately clear, this is indeed the unsurprising limit in the sense of (2.5). To see this, pick $q_k(x/\varepsilon) = \hat{q}_k |k|^{-\alpha}$. It is easy to see that, since $\langle \rho, 1 \rangle = 1$, the noise in the limiting SPDE (2.12) is the same as the original noise, as was the case in the classical result (2.5).

This result is reminiscent of previous results [WCD07, WD07], but stronger in the sense that genuine mean-squared convergence is obtained. Moreover, the result comes with rates of convergence. These are some of the perks enjoyed by a Fourier analytic framework, which we employ in place of the tightness arguments usually found in homogenisation problems. Of course, we still have weak convergence in a variational sense.

There are some important things to note concerning the limiting SPDE (2.12). Firstly, it is a stochastic heat equation with additive noise, and that noise comes with the same spatial regularity as the noise in the original SPDE. That is, the coefficients of W_k decay with the same rate. Secondly, if we choose the noise to satisfy the centering condition $\langle q_k, \rho \rangle = 0$ for each $k \in \mathbb{Z}$, then the solution u_ε will converge strongly to zero as $\varepsilon \to 0$. In other words, the presence of noise will have vanishingly small effect on the system (2.11) when ε is small. It is natural to ask whether we can find the largest vanishing term as $\varepsilon \to 0$. To obtain this term, we scale up the solution u_ε by some cleverly chosen inverse factor of ε and then seek a non-zero solution. For this procedure to work, we need to have very precise control over the coefficients q_k when k is large. Namely, we require that there exists some $\alpha \in (0,1)$ and a sufficiently regular function $\bar{q}$ such that $|k|^\alpha q_k \to \bar{q}$ in $L^2[0,2\pi]$ as $|k| \to \infty$. One can check that these assumptions imply those made for the previous theorem. The precise assumptions are detailed in Assumption 2.2.6. With these conditions, we can prove the following.

Theorem 2.1.6. *Suppose the SPDE* (2.11) *satisfies Assumptions 2.2.2 and 2.2.6 for some decay exponent* $\alpha \in (0,1)$ *and* $\langle q_k, \rho \rangle = 0$ *for all* $k \in \mathbb{Z}$*. Then there exists a process* $\hat{u}_\varepsilon$ *equal in law to* u_ε *but defined on a different probability space, such that the rescaled solutions* $\varepsilon^{-\alpha}\hat{u}_\varepsilon$ *converge to the solutions of*

$$dv(x,t) = \mu \partial_x^2 v(x,t)dt + \|\bar{q}\rho\|_{-\alpha} \sum_k e^{ikx} d\hat{W}_k(t) \tag{2.13}$$

in the sense that

$$\lim_{\varepsilon \to 0} \mathbb{E}\sup_{t\in[0,T]} |\langle \varepsilon^{-\alpha}\hat{u}_\varepsilon(t) - v(t), \varphi \rangle|^2 = 0 \ ,$$

for all $\varphi \in H^s$ *with large enough* s*.*

Here the convergence result is weak in both a variational and probabilistic sense. In general, nothing stronger is possible. Although the result looks like convergence in mean-squared, it is merely *disguised* convergence in law since we must define the limiting solution on a different probability space to the original SPDE. Such results are often obtained artificially using the Skorokhod embedding theorem. In our case however, this is the natural way to write down the result. In particular, for fixed $\varepsilon > 0$, the dependencies of $\hat{W}_m$ can be traced back to the original BMs. It is worth mentioning that the scaling factor required in order to find this term is in fact $\varepsilon^{-\alpha}$, which is precisely the amount of decay placed on the coefficients q_k. In the limiting SPDE (2.13), we use the notation

$$\|f\|_{-\alpha} = \left(\sum_{k\in\mathbb{Z}} |k|^{-2\alpha}|\langle f, e_k\rangle|^2\right)^{1/2} ,$$

where $e_k(x) = e^{ikx}$.

As before, there are several things to note about the SPDE (2.13). Firstly, it is again a stochastic heat equation with additive noise, but now all contributions from the original driving noise come from the very high modes, as indicated by the factor $\|\bar{q}\rho\|_{-\alpha}$. Thus, the coefficients q_k with low k have no bearing at all on the limit. In particular, if one wanted to approximate the noise by cutting off the sum at a large value of k, they would be making a drastic mistake! Moreover, this suggests that v arises due to constructive interference occurring in the very high modes of the noise. The second observation to make is that no matter what spatial regularity is possessed by the noise in the original SPDE, the limiting SPDE is always driven by space-time white noise. As one might guess, the factor $\varepsilon^{-\alpha}$ essentially scales away the decay on the coefficients q_k and hence destroys the regularity of the driving noise.

The previous theorem may seem a bit off topic, as we are trying to determine how choices of q_k affect the limiting SPDE. However, the following theorem tells us that the second order term found in Theorem 2.1.6 acts as the *bridge* between the surprising limit and the unsurprising limit. In particular, we will show that the surprising limit occurs precisely when this second order term becomes non-vanishing. We can see in (2.6) that space-time white noise falls into the '$\alpha = 0$ class', in the context of the previous theorems, since obviously $q_k = 1$ does not decay. Since the second order term was shown to be $O(\varepsilon^\alpha)$, one would expect this term to become $O(1)$ and hence contribute to the limit in the space-time white noise case. This suggests that the second order term is precisely the difference between the surprising limit and the unsurprising limit. The following theorem proves this to be the case not just for (2.6) but for all SPDEs driven by noise in the $\alpha = 0$ class.

The only added requirement for noise to be in this class is that there exists $\bar{q} \in H^1$ such that $q_k \to \bar{q}$ as $k \to \infty$ and that this convergence happens with fast enough rate. The precise conditions are found in Assumption 2.2.7. We have that following result.

Theorem 2.1.7. *Suppose the SPDE* (2.11) *satisfies Assumptions 2.2.2 and 2.2.7. Then there exists $\hat{u}_\varepsilon$ equal in law to u_ε, but defined on a different probability space, such that $\hat{u}_\varepsilon$ converges to the solutions of*

$$d\hat{u}(x,t) = \mu\partial_x^2\hat{u}(x,t)dt + \sum_{k\in\mathbb{Z}}(|\langle q_k, \rho\rangle|^2 - |\langle \bar{q}, \rho\rangle|^2 + \|\bar{q}\rho\|^2)^{1/2} e^{ikx} d\hat{W}_k(t) , \quad (2.14)$$

in the sense that

$$\lim_{\varepsilon\to 0} \mathbb{E}\mathrm{sup}_{t\in[0,T]} |\langle \hat{u}_\varepsilon(t) - \hat{u}(t), \varphi\rangle|^2 = 0 \quad \textit{for all } \varphi \in H^s$$

for large enough s*.*

As one might expect, this result is almost a combination of the two previous results, only a few extra ingredients are needed to prove it. In the $\|\cdot\|_{-\alpha}$ notation of Theorem 2.1.6, we have that

$$-|\langle \bar{q}, \rho\rangle|^2 + \|\bar{q}\rho\|^2 = \|\bar{q}\rho\|_0^2 \ ,$$

which is precisely the contribution from the second order term (squared), so that (2.14) really is a combination of the first order limit in (2.12) and the second order limit in (2.14). Note that instead of the noise being comprised of the sum of the first order and second order terms, we have the square-root of the sum of the squares. This is simply because we want to write each term in the noise as a single Gaussian, rather than a sum of two independent Gaussians. Just as in Theorem 2.1.6, the BMs $\hat{W}_m$ are, for fixed $\varepsilon > 0$ defined in terms of the original BMs.

The second order term $-|\langle \bar{q}, \rho\rangle|^2 + \|\bar{q}\rho\|^2$ constitutes what we referred to in Chapter 1 as different type of Itô correction. It is clearly not a classical Itô correction, since the additive term is not of bounded variation, instead it is more like a new source of noise. However, we believe we are justified in calling it an Itô correction, as it arises due to an interference occurring in the very high modes of the noise. Moreover, as with the traditional Itô corrections, as soon as the noise is slightly more regular (by introducing any algebraic decay), the correction term disappears.

To prove these three convergence results, we develop several tools that are useful when dealing with any SPDE whose underlying diffusion is driven by $\mathcal{L}_\varepsilon$. Firstly, we develop a relationship between the interpolation spaces generated by $\mathcal{L}_\varepsilon$ and the usual Sobolev spaces. This is useful in determining which function spaces contain our solutions (uniformly in ε) and furthermore determining where convergence occurs. Secondly, we show that the effect of the semigroup S_ε generated by $\mathcal{L}_\varepsilon$ on a certain class of functions is approximated well by the heat semigroup. This is akin to the well-known fact that $\mathcal{L}_\varepsilon \Rightarrow \mu\partial_x^2$, as discussed earlier.

The chapter is structured in the following way. In Section 2.2, we give a precise formulation of the main SPDE and detail the structural assumptions. In Section 2.3 we develop some tools necessary for the proof of the convergence theorems. In Section 2.4 we rigorously state and prove all three convergence theorems.

2.2 Formulation of the SPDE and some notation

Recall that $L^2[0,2\pi]$ denotes the complex L^2 space with its inner product normalised as

$$\langle f, g\rangle = \frac{1}{2\pi}\int_0^{2\pi} f g^* dx \ ,$$

and corresponding norm $\|\cdot\|$. We denote elements of the orthonormal Fourier basis by $e_k(x) = e^{ikx}$. We will also denote the usual L^∞ norm by $\|\cdot\|_\infty$. We define $\mathcal{C}_b^2$ as the subspace of $L^2[0,2\pi]$ of bounded, continuous functions with two bounded, continuous derivatives. We measure regularity through the Sobolev spaces H^s which we define as the completion of $L^2[0,2\pi]$ under the norm

$$\|\cdot\|_{H^s} = \|(1-\partial_x^2)^{s/2}\cdot\| \ ,$$

for any $s \in \mathbb{R}$. We shall also make use of the following Sobolev-like semi-norm

$$\|f\|_{-s} = \left(\sum_{k\in\mathbb{Z}} |k|^{-2s}|\langle f, e_k\rangle|^2\right)^{1/2} \ , \tag{2.15}$$

which can only be defined on f with $\langle f, 1\rangle = 0$. One can therefore think of this semi-norm as the norm $\|(-\partial_x^2)^{-s}\cdot\|$ defined on the space of mean-zero functions. We denote by $\|\cdot\|_{\mathrm{HS}}$ the Hilbert-Schmidt norm on linear operators that map $L^2[0,2\pi]$ into itself. As a shorthand we will write

$$f^\varepsilon(x) = f(x/\varepsilon) \ ,$$

when we want to omit the function's dependence on x. Finally, we will use the notation $f \lesssim g$ to imply that $|f/g|$ can be bounded by some constant that is independent of parameters involved in the expression. The precise independence will be clear from the context.

2.2.1 Formulation of the equation

Let b and σ be twice continuously differentiable 2π-periodic functions and define the differential operator $\mathcal{L}_\varepsilon$ as in (2.2) and likewise define the unscaled operator $\mathcal{L}$ as in (2.1). Following [PSV77, BLP78], we require some conditions on the generator $\mathcal{L}_\varepsilon$ for the homogenization problem to have a limit.

Assumption 2.2.2. Assume that $b, \sigma \in \mathcal{C}_b^2$ and that the centering condition

$$\int_0^{2\pi} \frac{b(x)}{\sigma^2(x)} dx = 0 \ , \tag{2.16}$$

is satisfied. Furthermore, σ is uniformly elliptic, namely

$$0 < \delta < \sigma(x) < \delta' < \infty \; , \tag{2.17}$$

for some fixed δ and δ'.

Remark 2.2.3. One can check that the centering condition implies that

$$\int_0^{2\pi} b(x)\rho(x)dx = 0 \; , \tag{2.18}$$

where ρ is the solution to $\mathcal{L}^*\rho = 0$ with periodic boundary conditions and satisfying $\langle \rho, 1\rangle = 1$. We will call ρ the invariant density for $\mathcal{L}$, despite the fact that it is not normalised to be a probability measure. This centering condition serves the same purpose as subtracting the mean when trying to obtain a central limit theorem.

Remark 2.2.4. The smoothness of b and σ, combined with the ellipticity condition, are sufficient to guarantee that $\rho \in \mathcal{C}_b^2$ and similarly for all positive and negative powers of ρ.

Our main object of interest is the following SPDE, defined on finite temporal and spatial domains

$$du_\varepsilon(x,t) = \mathcal{L}_\varepsilon u_\varepsilon(x,t)dt + \sum_{k\in\mathbb{Z}} q_k(x/\varepsilon)e_k(x)dW_k(t) \qquad (x,t) \in [0,2\pi]\times(0,T] \tag{2.19}$$

$$u_\varepsilon(0,t) = u_\varepsilon(2\pi,t) \qquad t \in [0,T] \tag{2.20}$$

$$u_\varepsilon(x,0) = 0 \qquad x \in [0,2\pi] \; . \tag{2.21}$$

Each $q_k(\cdot)$ is a continuous 2π-periodic element of $L^2[0,2\pi]$, taking values in $\mathbb{R}$ and we require that $q_{-k} = q_k$ for each $k \in \mathbb{Z}$. As stated in Remark 2.1.2, the microscopic parameter $\varepsilon \in (0,1)$ must satisfy $\varepsilon^{-1} \in \mathbb{N}$. We define the sequence of Brownian motions $\{W_k\}_{k\in\mathbb{Z}}$ in the following way: W_0 is a $\mathbb{R}$-valued BM, where as $\{W_k\}_{k\geq 1}$ are $\mathbb{C}$-valued BMs, and $\{W_k\}_{k\geq 0}$ are pairwise independent; we then set $W_{-k} = W_k^\star$, where $(\cdot)^\star$ denotes complex conjugation. Every bi-infinite sequence of Brownian motions considered in the sequel will satisfy this conjugation property. As stated, we assume periodic boundary conditions and take the initial condition to be identically zero. We choose this initial condition as we are only interested in the evolution of the noise through the system. Determining the evolution with a non-trivial initial condition is equivalent to adding the solution to the noiseless problem, which has been well studied [BLP78, PSV77, PS08].

For convenience we introduce the linear operator on $L^2[0, 2\pi]$ by

$$Q_\varepsilon e_k(x) = q_k(x/\varepsilon)e_k(x) \; , \tag{2.22}$$

and one can then represent the noise in (2.19) as $Q_\varepsilon dW$ where dW denotes space-time white noise. We shall now list the assumptions needed to prove Theorems 2.1.3, 2.1.6, 2.1.7 respectively. Firstly, we require the following condition to prove Theorem 2.1.3.

Assumption 2.2.5. There exists $\alpha \in (0, 1)$ such that

$$\|q_k\| \lesssim 1 \wedge |k|^{-\alpha} \; , \tag{2.23}$$

for each $k \in \mathbb{Z}$. Moreover, if $\alpha \in (0, 1/2]$ then we additionally require that

$$\sup_{k\in\mathbb{Z}} \|\bar{q}_k\|_{H^1} < \infty \; , \tag{2.24}$$

where $\bar{q}_k = q_k/\|q_k\|$.

To prove Theorem 2.1.6, we need slightly different assumptions to those required for Theorem 2.1.3. Namely, we need the following.

Assumption 2.2.6. There exists $\alpha \in (0, 1)$ and $\bar{q} \in L^2[0, 2\pi]$ such that

$$\lim_{k\to\pm\infty} \| |k|^\alpha q_k - \bar{q}\| = 0 \; . \tag{2.25}$$

Moreover, if $\alpha \in (0, 1/2]$ then we additionally require that

$$\sup_{k\in\mathbb{Z}} \|\bar{q}_k\|_{H^1} < \infty \; . \tag{2.26}$$

Note that (2.25) guarantees that the bound

$$\|q_k\| \lesssim 1 \wedge |k|^{-\alpha}$$

holds for all $k \in \mathbb{Z}$ and therefore Assumption 2.2.6 implies Assumption 2.2.5. Unlike in Theorem 2.1.3, having a rate of decay on q_k does not suffice, we now need precise control over how q_k tends to zero as $k \to \infty$.

Recall that Theorem 2.1.7 deals with those SPDEs that converge to the so called wrong limit. We claimed that this wrong limit occurred when the limit from Theorem 2.1.3 combined with the limit from Theorem 2.1.6, by formally taking $\alpha = 0$. Since Assumption 2.2.6 implies Assumption 2.2.5, our condition on the noise for Theorem 2.1.7 should look like Assumption 2.2.6, with $\alpha = 0$. Actually, we need a tiny bit more than this.

Assumption 2.2.7. We require that there exists $\bar{q} \in H^1$ and $\eta \in [0, 1)$ such that

$$\sum_{k\in\mathbb{Z}} (1 \wedge |k|^{-\eta}) \|q_k - \bar{q}\|^2_{H^1} < \infty \; . \tag{2.27}$$

At first glance this looks quite a bit stronger than Assumption 2.2.6 with $\alpha = 0$. However, Assumption 2.2.6 with $\alpha = 0$ implies that $\|q_k - \bar{q}\|_{H^s} \to 0$ for every $s < 1$, since the convergence is true in $L^2[0, 2\pi]$ and the sequence $\{q_k\}$ is uniformly bounded in H^1. And since η can be arbitrarily close to 1, Assumption 2.2.6 almost implies Assumption 2.2.7, but not quite. Note that the uniform boundedness condition on $\|q_k\|_{H^1}$ is not implicitly stated, but it is implied by the listed assumptions. The parameter η will affect the strength of the convergence result in Theorem 2.1.7, namely, larger η leads to weaker convergence.

Remark 2.2.8. Another sufficient condition for Theorem 2.1.7 is that

$$\sum_k \|q_k - \bar{q}\|^2 < \infty \; , \tag{2.28}$$

with $\bar{q} \in H^1$. Actually, we could also replace the regularity condition in Assumption 2.2.5 with (2.28). However we consider the regularity assumption to be a more natural choice.

We define solutions to (2.19) using the mild formulation

$$u_\varepsilon(x, t) = \int_0^t S_\varepsilon(t - s) Q_\varepsilon dW(s) = \sum_{k\in\mathbb{Z}} \int_0^t S_\varepsilon(t - s) q_k(x/\varepsilon) e_k(x) dW_k(s) \; , \tag{2.29}$$

where $S_\varepsilon(t)$ is the semigroup generated by $\mathcal{L}_\varepsilon$. It is easy to check, using techniques introduced in the next section, that for fixed $\varepsilon > 0$, the semigroup $S_\varepsilon(t)$ is a $\mathcal{C}_0$-semigroup. In this case, one can check that weak and mild solutions coincide [Hai09, DPZ92], so the mild solution is indeed the correct one to look at. We also have the following regularity result

Proposition 2.2.9. *Suppose Assumptions 2.2.2, 2.2.5 or 2.2.2, 2.2.7 hold true. Then, for fixed $\varepsilon \in (0, 1)$, the solution u_ε to* (2.19) *has almost surely continuous sample paths in* $L^2[0, 2\pi]$.

Proof. Using standard results for linear SPDEs [Hai09, DPZ92] we need only check that

$$\|S_\varepsilon(t) Q_\varepsilon\|_{\mathrm{HS}} < \infty \; ,$$

for every $t \in (0, T]$ and that there exists $\beta \in (0, 1/2)$ such that

$$\int_0^T t^{-2\beta} \|S_\varepsilon(t) Q_\varepsilon\|^2_{\mathrm{HS}} dt < \infty \; .$$

In Lemma 2.4.9 below, we show that Assumption 2.2.5 implies that

$$\|S_\varepsilon(t)Q_\varepsilon\|_{\mathrm{HS}} \lesssim \varepsilon^{-4\gamma}|t|^{-\gamma}\left(\sum_{k\in\mathbb{Z}}(1\wedge|k|^{-4\gamma})\|q_k\|_{H^1}^2\right)^{1/2},$$

for any $\gamma \in (0,1/2)$. In Lemma 2.4.16, we show that Assumption 2.2.7 implies a similar estimate. The result follows immediately. □

Remark 2.2.10. Note that although the decay assumption on $\|q_k\|$ was not needed to show regularity of the solutions, it is necessary when proving convergence as $\varepsilon \to 0$. It furthermore allows us to fine tune our results so that we can find the optimal space in which convergence occurs.

2.3 Preliminary Results

In this section we shall develop a few tools necessary for the proof of the main results. In Section 2.3.1, we start with some standard results concerning the semigroups generated by one dimensional Itô diffusions. In Section 2.3.4, we develop a relationship between the interpolation spaces of $\mathcal{L}_\varepsilon$ and the Sobolev spaces. Finally, in Section 2.3.7, we go on to approximate the effect of the adjoint semigroup $S_\varepsilon^*(t)$ on trigonometric polynomials.

2.3.1 Properties of the diffusion

We recall some basic results concerning the semigroup $S_\varepsilon(t)$ generated by $\mathcal{L}_\varepsilon$. Firstly, we have the following smoothing properties.

Lemma 2.3.2. *For any $t \in [0,T]$ we have that*

$$\|S_\varepsilon(t)\| \le C_T\;. \tag{2.30}$$

Moreover, for any $\gamma \in [0,1)$ we have that

$$\|(1-\mathcal{L}_\varepsilon)^\gamma S_\varepsilon(t)\| \lesssim t^{-\gamma}\;. \tag{2.31}$$

Finally, the same results hold true with $S_\varepsilon(t)$ and $\mathcal{L}_\varepsilon$ replaced with their adjoints $S_\varepsilon^*(t)$ and $\mathcal{L}_\varepsilon^*$.

Proof. We shall only prove (2.31) since (2.30) follows as a special case. If $\mathcal{L}_\varepsilon$ were self-adjoint, then the result would follow easily from the spectral theorem [Hai09]. $\mathcal{L}_\varepsilon$ is self-adjoint if the domain of the operator is taken to be the weighted space $L^2(\rho_\varepsilon)$ with norm

$\|f\|_{\rho_\varepsilon} = \|f\rho_\varepsilon^{1/2}\|$ and corresponding inner product, where ρ_ε is the invariant density for $\mathcal{L}_\varepsilon$. The spectral theorem therefore implies that

$$\|(1-\mathcal{L}_\varepsilon)^\gamma S_\varepsilon(t) f\|_{\rho_\varepsilon} \lesssim t^{-\gamma}\|f\|_{\rho_\varepsilon} \ .$$

Furthermore, one can easily show that $\rho_\varepsilon = \rho(x/\varepsilon)$ where ρ is the invariant density of $\mathcal{L}$, which we assumed in (2.17) to be bounded above and away from zero. We therefore have that

$$\begin{aligned}\|(1-\mathcal{L}_\varepsilon)^\gamma S_\varepsilon(t) f\| &\le \|\rho^{-1/2}\|_\infty \|(1-\mathcal{L}_\varepsilon)^\gamma S_\varepsilon(t) f\|_{\rho_\varepsilon} \\ &\lesssim t^{-\gamma}\|\rho^{-1/2}\|_\infty \|f\|_{\rho_\varepsilon} \le t^{-\gamma}\|\rho^{-1/2}\|_\infty\|\rho^{1/2}\|_\infty\|f\| \lesssim t^{-\gamma}\|f\| \ ,\end{aligned}$$

which proves the results for $S_\varepsilon(t)$. The results for $S^*_\varepsilon(t)$ follow from the dual representation $\|S^*_\varepsilon(t) f\| = \sup_{\|g\|=1}|\langle f, S_\varepsilon(t) g\rangle|$. □

We now recall some standard estimates on the adjoint of the semigroup $S(t)$ generated by $\mathcal{L}$.

Lemma 2.3.3. *Let $S^*(t)$ denote the adjoint of $S(t)$. For any $t \in (0,T]$, we have that*

$$\|S^*(t)\| \le C_T \ , \tag{2.32}$$

$$\|\partial_x S^*(t)\| \lesssim |t|^{-1/2} \ . \tag{2.33}$$

Moreover, there exists $\omega > 0$ such that

$$\|S^*(t)\,(1-\rho(x))\,\| \lesssim \exp(-\omega t) \ . \tag{2.34}$$

Proof. The first result follows from Lemma 2.3.2 with $\varepsilon = 1$. The second result follows if we can show that the interpolation spaces of $(1-\mathcal{L})$ are the same as the Sobolev spaces interpolated by $(1-\partial_x^2)$. Firstly, one can find a change of variables Q such that

$$Q\mathcal{L}Q^{-1} = V(x)\partial_x + \partial_x^2$$

where Q and its inverse are bounded from H^s into itself for any s and V is bounded. This change of variables can be found in Lemma 2.3.5. Hence, the interpolation spaces of $(1-\mathcal{L})$ are the same as the interpolation spaces of $(-V(x)\partial_x + 1 - \partial_x^2)$. Furthermore, we have the following fact: if L_0 generates an analytic semigroup on $\mathcal{B}$ and has interpolation spaces $\mathcal{B}^0_\gamma$, then $B + L_0$ has the same interpolation spaces, whenever B is a bounded operator from $\mathcal{B}^0_\gamma$ into $\mathcal{B}$, for some $\gamma \in [0,1)$ by [Hai09]. It follows that $B + L_0 = (1 - Q\mathcal{L}Q^{-1})$ has the same interpolation spaces as $L_0 = (1-\partial_x^2)$, which proves the claim. The third result follows using standard machinery from spectral theory, similar to those used in Lemma 2.3.2. $\square$

Since it will not affect any of our future estimates, we will assume from this point on that $\omega = 1$. Notice that the semigroup $S_\varepsilon(t)$ satisfies the following rescaling identity

$$S_\varepsilon(t) f^\varepsilon(x) = (S(t/\varepsilon^2) f)(x/\varepsilon) \; . \tag{2.35}$$

One can therefore think of the semigroup as zooming in on the highly oscillatory parts, evolving them (according to the diffusion generated by $\mathcal{L}$) to very large times, and then zooming back out. In particular, combining this identity with Lemma 2.3.3 gives

$$\|S_\varepsilon^*(t)\,(1 - \rho(x/\varepsilon))\| \lesssim \exp(-\omega t/\varepsilon^2) \; , \tag{2.36}$$

which will prove useful in the sequel.

2.3.4 Interpolation Results

In order to prove convergence results in particular Sobolev spaces, we need to know the smoothing properties of the semigroup $S_\varepsilon(t)$. Estimates from analytic semigroup theory tell us which interpolation spaces of $\mathcal{L}_\varepsilon$ the solutions will live in. We would therefore like to obtain some embedding result between these interpolation spaces and the usual Sobolev spaces. It would be futile to look for an embedding result uniformly in ε, the best we can do is the following lemma, which, for a price, grants us the ability to switch back and forth between interpolation spaces and Sobolev spaces.

Lemma 2.3.5. *One has the following two inequalities*

$$\|(1 - \partial_x^2)^\gamma f\| \lesssim \varepsilon^{-2\gamma} \|(1 - \mathcal{L}_\varepsilon)^\gamma f\| \tag{2.37}$$

$$\|(1 - \mathcal{L}_\varepsilon)^{-\gamma} f\| \lesssim \varepsilon^{-2\gamma} \|(1 - \partial_x^2)^{-\gamma} f\| \tag{2.38}$$

for any $\gamma \in [0, 1]$ and any f for which the two norms are finite.

Proof. We start by proving the first inequality, the second will follow with a simple argument. To prove the first claim we apply the Caldéron-Lions interpolation theorem [RS75] to obtain a relationship between the interpolation spaces (in the notation of [RS75]) given by

$$\|\cdot\|_X^{(0)} = \|\cdot\| \; , \qquad \|\cdot\|_X^{(1)} = \|(1 - \mathcal{L}_\varepsilon) \cdot\| \; ,$$
$$\|\cdot\|_Y^{(0)} = \|\cdot\| \; , \qquad \|\cdot\|_Y^{(1)} = \|(1 - \partial_x^2) \cdot\| \; .$$

It guarantees that, for the identity operator I, one has

$$\|I\|_{L(X^{(\gamma)},Y^{(\gamma)})} \le \|I\|^{1-\gamma}_{L(X^{(0)},Y^{(0)})}\|I\|^{\gamma}_{L(X^{(1)},Y^{(1)})} , \tag{2.39}$$

where $X^{(\gamma)}$ and $Y^{(\gamma)}$ are the interpolation spaces given by completing $L^2[0, 2\pi]$ with respect to the norms $\|(1 - \mathcal{L}_\varepsilon)^\gamma \cdot \|$ and $\|(1 - \partial_x^2)^\gamma \cdot \|$ respectively.
It is clear that

$$\|I\|_{L(X^{(0)},Y^{(0)})} = 1 ,$$

since this is just the norm of the identity operator in $L^2[0, 2\pi]$. The first claim thus follows if we can show that

$$\|I\|_{L(X^{(1)},Y^{(1)})} \lesssim \varepsilon^{-2} ,$$

which is equivalent to proving that

$$\|(1 - \partial_x^2)(1 - \mathcal{L}_\varepsilon)^{-1} f\| \lesssim \varepsilon^{-2}\|f\| . \tag{2.40}$$

We will achieve this by simplifying the operator $\mathcal{L}_\varepsilon$ through two transformations. Firstly, for the generator $\mathcal{L}$, one can easily find a change of variables $z = \phi(x)$ with inverse $x = \psi(z)$ such that

$$\mathcal{L}f(x) = \left(B(\psi(z))\partial_z + \partial_z^2\right)(f \circ \psi)(z) , \tag{2.41}$$

where $B = \sqrt{2}\frac{b}{\sigma} - \frac{1}{\sqrt{2}}\sigma'$ and ϕ solves the ordinary differential equation

$$\phi'(x) = \frac{1}{\sqrt{2}}\sigma(\phi(x)) , \tag{2.42}$$

with boundary condition $\phi(0) = 0$. Given this change of variables, it is easy to find the corresponding change of variables for $\mathcal{L}_\varepsilon$, in fact, if we set $z = \varepsilon\phi(x/\varepsilon)$ we have that

$$\mathcal{L}_\varepsilon f(x) = \left(\frac{1}{\varepsilon}B(\psi(z/\varepsilon))\partial_z + \partial_z^2\right)(f \circ \psi_\varepsilon)(z) , \tag{2.43}$$

where $\psi_\varepsilon(\cdot) = \varepsilon\psi(\cdot/\varepsilon)$. Secondly, we hope to make the operator self-adjoint. To do this, we weight our space using the invariant measure of the underlying generator. Let $g(y)$ be the invariant density for the generator $\left(\frac{\sqrt{2}B(y)}{\sigma(y)}\partial_y + \partial_y^2\right)$. One can show that

$$\mathcal{L}_\varepsilon f(x) = g(x/\varepsilon)^{-1/2}(\mathcal{A}_\varepsilon u)(\varepsilon\phi(x/\varepsilon)) , \tag{2.44}$$

where $u = g(\psi(\cdot/\varepsilon))^{1/2} f \circ \psi_\varepsilon$. The Schrödinger operator $\mathcal{A}_\varepsilon$ is defined by

$$\mathcal{A}_\varepsilon u(z) = \frac{1}{\varepsilon^2} W(\psi(z/\varepsilon))u(z) + \partial_z^2 u(z)$$

where $W = g^{1/2} \left(\frac{\sqrt{2}B}{\sigma} \partial_y + \partial_y^2 \right) g^{-1/2}$. We then have that

$$\begin{aligned} \|(1-\partial_x^2)(1-\mathcal{L}_\varepsilon)^{-1} f(x)\| \leq \varepsilon^{-2} &\|(g^{-1/2})''\|_\infty \|(1-\mathcal{A}_\varepsilon)^{-1} u(\varepsilon\phi(x/\varepsilon))\| \\ &+ \varepsilon^{-1} \|(g^{-1/2})'\|_\infty \|\partial_x (1-\mathcal{A}_\varepsilon)^{-1} u(\varepsilon\phi(x/\varepsilon))\| \\ &+ \|g^{-1/2}\|_\infty \|\partial_x^2 (1-\mathcal{A}_\varepsilon)^{-1} u(\varepsilon\phi(x/\varepsilon))\| \ . \end{aligned}$$

One can easily deduce the boundedness of $g^{-1/2}$ and its derivatives from Assumption 2.2.2. Moreover, we have that

$$\begin{aligned} \|\partial_x (1-\mathcal{A}_\varepsilon)^{-1} u(\varepsilon\phi(x/\varepsilon))\|^2 &= \| \left((1-\mathcal{A}_\varepsilon)^{-1} u)'(\varepsilon\phi(x/\varepsilon)\right) \phi'(x/\varepsilon)\|^2 \\ &= \frac{1}{2\pi} \int_0^{2\pi} | \left((1-\mathcal{A}_\varepsilon)^{-1} u)'(\varepsilon\phi(x/\varepsilon)\right) \phi'(x/\varepsilon)|^2 dx \\ &= \frac{1}{2\pi} \int_0^{\varepsilon\phi(2\pi/\varepsilon)} |\partial_z (1-\mathcal{A}_\varepsilon)^{-1} u(z)|^2 |\phi'(\psi(z/\varepsilon))| dz \\ &\leq \|\phi'\|_\infty \|\partial_z (1-\mathcal{A}_\varepsilon)^{-1} u\|_\phi^2 \ , \end{aligned}$$

where $\|\cdot\|_\phi$ denotes the usual L^2 norm but over the interval $[0, \varepsilon\phi(2\pi/\varepsilon)]$ as in the integral above. We can similarly show that

$$\begin{aligned} \|\partial_x^2 (1-\mathcal{A}_\varepsilon)^{-1} u(\varepsilon\phi(x/\varepsilon))\| \leq \|(\phi')^3\|_\infty^{1/2} &\|\partial_z^2 (1-\mathcal{A}_\varepsilon)^{-1} u\|_\phi \\ &+ \varepsilon^{-1} \|\frac{(\phi'')^2}{\phi}\|_\infty^{1/2} \|\partial_z (1-\mathcal{A}_\varepsilon)^{-1} u\|_\phi \ . \end{aligned}$$

We can deduce the boundedness of the above expressions involving ϕ using (2.42) and Assumption 2.2.2. We therefore have the bound

$$\begin{aligned} \|(1-\partial_x^2)(1-\mathcal{L}_\varepsilon)^{-1} f\| \lesssim \varepsilon^{-2} &\|(1-\mathcal{A}_\varepsilon)^{-1} u\|_\phi + \varepsilon^{-1} \|\partial_z (1-\mathcal{A}_\varepsilon)^{-1} u\|_\phi \\ &+ \|\partial_z^2 (1-\mathcal{A}_\varepsilon)^{-1} u\|_\phi \ . \end{aligned}$$

We now claim the following bounds to hold, as operator norms from $L^2_\phi \to L^2_\phi$ in the sense of the norm defined above:

$$\|(1-\mathcal{A}_\varepsilon)^{-1}\|_\phi \leq 1 \ , \tag{2.45}$$

$$\|\partial_z^2 (1-\mathcal{A}_\varepsilon)^{-1}\|_\phi \lesssim \varepsilon^{-2} \ . \tag{2.46}$$

Note that these bounds immediately imply $\|\partial_z(1-\mathcal{A}_\varepsilon)^{-1}\|_\phi \lesssim \varepsilon^{-1}$ which follows from the Cauchy-Schwartz inequality. These three operator bounds are enough to prove (2.40), since by changing back to the x variables, we have that

$$\|u\|_\phi = \|g(x/\varepsilon)^{1/2} f(x)(\phi'(x/\varepsilon))^{1/2}\| \le \|g\|_\infty^{1/2}\|\phi'\|_\infty^{1/2}\|f\| \ .$$

Hence we need only prove the claimed bounds. To prove (2.45), we utilise the identity

$$\mathrm{spec}(1-\mathcal{A}_\varepsilon) = \mathrm{spec}\,(1-\mathcal{L}_\varepsilon) \ ,$$

which follows from the fact that $\mathcal{A}_\varepsilon$ and $\mathcal{L}_\varepsilon$ are conjugated via a bounded operator with bounded inverse. Since $\mathcal{L}_\varepsilon$ generates a Markov semigroup, elements in its spectrum have positive real part. Since $(1-\mathcal{A}_\varepsilon)$ is self-adjoint in the Hilbert space generated by the norm $\|\cdot\|_\phi$ with the corresponding inner product, it thus follows that

$$\|(1-\mathcal{A}_\varepsilon)^{-1}\|_\phi \le 1$$

using the spectral theorem [Hai09]. By writing ∂_z^2 in terms of $\mathcal{A}_\varepsilon$ and W, we also have that

$$\begin{aligned}\|\partial_z^2(1-\mathcal{A}_\varepsilon)^{-1}\|_\phi &\le 1 + \|\left(1+\frac{1}{\varepsilon^2}W(\psi(\cdot/\varepsilon))\right)(1-\mathcal{A}_\varepsilon)^{-1}\|_\phi \\ &\lesssim 1+\varepsilon^{-2}(1+\|W\|_\infty)\|(1-\mathcal{A}_\varepsilon)^{-1}\|_\phi \ ,\end{aligned}$$

which proves (2.46) and hence (2.37). To prove the second part of the lemma, just as in (2.40) it is sufficient to show that

$$\|(1-\mathcal{L}_\varepsilon)^{-1}(1-\partial_x^2)f\| \le C\varepsilon^{-2}\|f\| \ .$$

But we can use the fact that the operator norm is preserved under taking the adjoint, so that

$$\|(1-\mathcal{L}_\varepsilon)^{-1}(1-\partial_x^2)\| = \|(1-\partial_x^2)(1-\mathcal{L}_\varepsilon^*)^{-1}\| \ .$$

It is therefore sufficient to prove (2.40) with $\mathcal{L}_\varepsilon$ replaced with its adjoint $\mathcal{L}_\varepsilon^*$. An easy calculation shows that

$$\mathcal{L}_\varepsilon^* = \widetilde{\mathcal{L}}_\varepsilon + \frac{1}{\varepsilon^2}U(x/\varepsilon)$$

where

$$\widetilde{\mathcal{L}}_\varepsilon = \frac{1}{\varepsilon}\tilde{b}(x/\varepsilon)\partial_x + \frac{1}{2}\sigma^2(x/\varepsilon)\partial_x^2 \ .$$

We can reduce $\widetilde{\mathcal{L}}_\varepsilon$ to a Schrödinger operator with potential $\hat{W}$ in the same way that we did for $\mathcal{L}_\varepsilon$, and hence reduce $\mathcal{L}_\varepsilon^*$ to a Schrödinger operator with potential $\hat{W}+U$. The second claim then follows similarly to the first. □

Remark 2.3.6. We would like to briefly comment on the sharpness of the two estimates obtained in Lemma 2.3.5. The second estimate (2.38) is sharp. In fact, in the case $\sigma = 1$, upon rewriting the estimate in the adjoint setting, as done in the proof, it is clear that taking $f = \rho(x/\varepsilon)$ will prove sharpness. Unfortunately, this argument does not work for the first estimate (2.37). This comes down to the unlucky fact that the zero eigenvector of $\mathcal{L}_\varepsilon$ is the constant function (and not $\rho(x/\varepsilon)$), which of course does not yield powers of ε when integrated. In fact, we believe that estimate (2.37) is not sharp. However, improving the estimate would not considerably improve the strength of results in the sequel, so we do not attempt to do so.

2.3.7 Estimating the semigroup

A key ingredient in proving all three convergence results is an estimate on the low Fourier modes of the mild solution to (2.19), that is

$$\langle u_\varepsilon(t), e_m \rangle = \sum_k \int_0^t \langle S_\varepsilon(t-s) q_k^\varepsilon e_k, e_m \rangle dW_k(s) \ ,$$

for $|m| < \varepsilon^{-1}$, recalling the notation $q_k^\varepsilon(x) = q_k(x/\varepsilon)$. This could be achieved by estimating $S_\varepsilon(t-s)q_k^\varepsilon e_k$. However, this becomes troublesome when k is large. It is more convenient to exploit the fact that

$$\langle u_\varepsilon(t), e_m \rangle = \sum_k \int_0^t \langle q_k^\varepsilon e_k, S_\varepsilon^*(t-s) e_m \rangle dW_k(s)$$

and estimate $S_\varepsilon^*(t-s)e_m$, with m fixed. We will prove that

$$S_\varepsilon^*(t) e_m(x) \approx \rho(x/\varepsilon) e_m(x) e^{-\mu m^2 t} + f_\varepsilon^{\mathrm{BL}}(x,t) \ ,$$

uniformly in $t \in [0,T]$. As before, ρ is the invariant density of the $\mathcal{L}$ and we define the "boundary layer" $f_\varepsilon^{\mathrm{BL}}$ as a term that corrects the approximation when $t = O(\varepsilon^2)$ and converges rapidly to zero when $t > \varepsilon^2$. Such results can be obtained in the setting of martingale problems [PSV77] however, as we would like to obtain a bit of control over rates of convergence, we take the approach used in [BLP78, PS08].

Let us set $f_\varepsilon(x,t) = S_\varepsilon^*(t)e_m(x)$. We would then like to find an approximate solution to the PDE

$$\partial_t f_\varepsilon(x,t) = \mathcal{L}_\varepsilon^* f_\varepsilon(x,t) \ , \quad f_\varepsilon(x,0) = e_m(x) \ , \tag{2.47}$$

where the adjoint generator $\mathcal{L}_\varepsilon^*$ has periodic boundary conditions on $[0, 2\pi]$. The standard approach to problems of this kind is to rewrite (2.47) in the new variables $\widetilde{x} = x$ and $\widetilde{y} = x/\varepsilon$ and separate the macroscopic dynamics from the microscopic dynamics. One can then obtain an approximate solution by introducing a power series expansion

$$f_\varepsilon(\widetilde{x}, \widetilde{y}, t) = f_0(\widetilde{x}, \widetilde{y}, t) + \varepsilon f_1(\widetilde{x}, \widetilde{y}, t) + \varepsilon^2 f_2(\widetilde{x}, \widetilde{y}, t) + \dots$$

into the PDE (2.47) and solving for f_0, f_1, f_2 by matching powers of ε. Under this procedure, one obtains

$$\begin{aligned} f_0(x, x/\varepsilon, t) &= \rho(x/\varepsilon) e_m(x) e^{-\mu m^2 t} \ , \\ f_1(x, x/\varepsilon, t) &= \Phi_1(x/\varepsilon) \partial_x e_m(x) e^{-\mu m^2 t} \ , \\ f_2(x, x/\varepsilon, t) &= \Phi_2(x/\varepsilon) \partial_x^2 e_m(x) e^{-\mu m^2 t} \ , \end{aligned}$$

where $\Phi_1, \Phi_2 \in \mathcal{C}_b^2$. This approach encounters a small problem in that the approximation breaks down when $t = O(\varepsilon^2)$. The problem is averted by introducing a temporal boundary layer term, also known as a corrector, which we define as

$$f_\varepsilon^{\mathrm{BL}}(x,t) = \left(S_\varepsilon^*(t) \left(1 - \rho(x/\varepsilon)\right)\right) e_m(x) \ .$$

One can see that the boundary layer term corrects the discrepancy in the initial condition of the approximation $S_\varepsilon^*(t) e_m(x) \approx \rho(x/\varepsilon) e_m(x) e^{-\mu m^2 t}$, indeed, the boundary layer term's sole purpose is to correct the approximation for small times t. We therefore define the remainder term r_ε by setting

$$f_\varepsilon(x,t) = f_0(\widetilde{x}, \widetilde{y}, t) + \varepsilon f_1(\widetilde{x}, \widetilde{y}, t) + \varepsilon^2 f_2(\widetilde{x}, \widetilde{y}, t) + f_\varepsilon^{\mathrm{BL}}(x,t) + r_\varepsilon(x,t) \tag{2.48}$$

Note that our definition of the remainder depends explicitly on the wavenumber m, however, for convenience we omit this from the notation. Using the method described above, one can write down the following convenient expression for the remainder.

Lemma 2.3.8. *If $\varepsilon|m| < 1$ and r_ε is the remainder defined in* (2.48) *then we can write*

$$r_\varepsilon(x,t) = S_\varepsilon^*(t) r_\varepsilon(x,0) + \ \varepsilon \int_0^t S_\varepsilon^*(t-s) F_1(x, x/\varepsilon, s)\, ds \tag{2.49}$$

$$+ \varepsilon^2 \int_0^t S_\varepsilon^*(t-s)F_2(x,x/\varepsilon,s)ds + \int_0^t S_\varepsilon^*(t-s)(\partial_s - \mathcal{L}_\varepsilon^*)f_\varepsilon^{\mathrm{BL}}(x,s)\,ds\;,$$

where the functions F_1 and F_2 satisfy the bounds

$$\|F_1(t)\| \lesssim (1\vee|m|^3)e^{-\mu m^2 t} \quad \textit{and} \quad \|F_2(t)\| \lesssim (1\vee|m|^4)e^{-\mu m^2 t}\;, \tag{2.50}$$

where $\mu > 0$ is a constant determined by $\mathcal{L}$.

Proof. The method of proof is described above. One can find similar calculations in [BLP78, PS08]. □

Each term in (2.49) can be bounded without too much trouble, except for the boundary layer term, which we shall treat separately.

Lemma 2.3.9. *If $\varepsilon|m| < 1$, then for any $t \in [0,T]$, we have that*

$$\|f_\varepsilon^{\mathrm{BL}}(t)\| \lesssim \exp(-t/\varepsilon^2)\;. \tag{2.51}$$

Furthermore, for any $s \in [0,t]$ we have that

$$\|S_\varepsilon^*(t-s)(\mathcal{L}_\varepsilon^* - \partial_s)f_\varepsilon^{\mathrm{BL}}(x,s)\| \lesssim \frac{m}{\varepsilon}\exp(-s/\varepsilon^2)\;. \tag{2.52}$$

In both cases, the proportionality constants are independent of m, provided that $\varepsilon|m| \le 1$.

Proof. For the sake of brevity, throughout this proof and the next we will simply write m instead of $1\vee|m|$. We also introduce the shorthand

$$\hat\rho_{t/\varepsilon^2}(x/\varepsilon) := \big(S^*(t/\varepsilon^2)\,(1-\rho)\big)(x/\varepsilon) = \big(S_\varepsilon^*(t)\,(1-\rho^\varepsilon)\big)(x)$$

where the last identity follows from the rescaling property (2.35), recalling that $\rho^\varepsilon(x) = \rho(x/\varepsilon)$. We then have that

$$\|f_\varepsilon^{\mathrm{BL}}(t)\| = \|\hat\rho^\varepsilon_{t/\varepsilon^2}e_m\| = \|\hat\rho^\varepsilon_{t/\varepsilon^2}\| \lesssim \exp(-t/\varepsilon^2)\;,$$

which follows from (2.36). For the second result, notice that

$$(\mathcal{L}_\varepsilon^* - \partial_s)f_\varepsilon^{\mathrm{BL}}(x,s) = -\frac{1}{\varepsilon}b(x/\varepsilon)\hat\rho_{\frac{s}{\varepsilon^2}}(x/\varepsilon)\partial_x e_m(x) + \partial_x\left(\sigma^2(x/\varepsilon)\hat\rho_{\frac{s}{\varepsilon^2}}(x/\varepsilon)\right)\partial_x e_m(x) + \frac{1}{2}\sigma^2(x/\varepsilon)\hat\rho_{\frac{s}{\varepsilon^2}}(x/\varepsilon)\partial_x^2 e_m(x)\;.$$

Therefore, the quantity

$$\|S_\varepsilon(t-s)(\mathcal{L}_\varepsilon^* - \partial_s) f_\varepsilon^{\mathrm{BL}}(x,s)\| \lesssim \|(\mathcal{L}_\varepsilon^* - \partial_s) f_\varepsilon^{\mathrm{BL}}(x,s)\|$$

is bounded by

$$\frac{m}{\varepsilon}\|b\hat{\rho}_{s/\varepsilon^2}\| + \frac{m}{\varepsilon^2}\|\partial_x\left(\sigma^2\hat{\rho}_{s/\varepsilon^2}\right)\| + m^2\|\sigma^2\hat{\rho}_{s/\varepsilon^2}\| \ . \tag{2.53}$$

We furthermore have the bound

$$\begin{aligned}\|\partial_x\left(\sigma^2\hat{\rho}_{s/\varepsilon^2}\right)\| &\lesssim \|\partial_x\sigma^2\|_\infty\|\hat{\rho}_{s/\varepsilon^2}\| + \|\sigma^2\|_\infty\|\partial_x\hat{\rho}_{s/\varepsilon^2}\| \\ &\lesssim \left(\|\partial_x\sigma^2\|_\infty + \|\sigma^2\|_\infty\right)\exp(-s/\varepsilon^2)\ ,\end{aligned}$$

where we have used the bound

$$\|\partial_x\hat{\rho}_{s/\varepsilon^2}\| \lesssim \exp(-s/\varepsilon^2)\ . \tag{2.54}$$

which we will prove shortly. Therefore, we can bound (2.53) by

$$\begin{aligned}\frac{m}{\varepsilon}\|b\|_\infty\exp(-s/\varepsilon^2) + \frac{m}{\varepsilon}(\|\partial_x\sigma^2\|_\infty + \|\sigma^2\|_\infty)\exp(-s/\varepsilon^2) + m^2\|\sigma^2\|_\infty\exp(-s/\varepsilon^2) \\ \lesssim \frac{m}{\varepsilon}\exp(-s/\varepsilon^2)\ .\end{aligned}$$

Here we have used Assumption 2.2.2 to obtain the required bounds on b and σ and also the assumption $\varepsilon|m| < 1$. This proves the bounds stated in the lemma. To prove the claimed bound (2.54), first assume $s > \varepsilon^2$, then

$$\begin{aligned}\|\partial_x\hat{\rho}_{s/\varepsilon^2}\| &= \|\partial_x S^*(1)S^*(s/\varepsilon^2 - 1)(1-\rho)\| \\ &\lesssim \|\partial_x S^*(1)\|\|S^*(s/\varepsilon^2-1)(1-\rho)\| \lesssim \exp(-s/\varepsilon^2)\ ,\end{aligned}$$

where we have used Lemma 2.3.3. If $s \le \varepsilon^2$ then

$$\begin{aligned}\|\partial_x\hat{\rho}_{s/\varepsilon^2}\| &= \|\partial_x(\mathcal{L}^*)^{-1}S^*(s/\varepsilon^2)\mathcal{L}^*(1-\rho)\| \\ &\lesssim \|\partial_x(\mathcal{L}^*)^{-1}\|\|\mathcal{L}^*1\| < \infty\ .\end{aligned}$$

The boundedness of $\|\partial_x(\mathcal{L}^*)^{-1}\|$ follows from the proof of Lemma 2.3.3, where we showed that $\mathcal{L}$ and ∂_x^2 share the same interpolation spaces. We can therefore bound $\|\partial_x\hat{\rho}_{s/\varepsilon^2}\|$ uniformly for $s \in [0,\varepsilon^2]$, which, together with the bound for $s > \varepsilon^2$, implies (2.54). □

Note that r_ε contains extra terms f_1 and f_2 that are only in place to facilitate the proof of Lemma 2.3.8. We therefore define the following new remainder for the approximation that we actually use

$$S_\varepsilon^*(t)e_m(x) = \rho(x/\varepsilon)e_m(x)e^{-\mu m^2 t} + f_\varepsilon^{\mathrm{BL}}(x,t) + R_\varepsilon(x,t)\ .$$

We now obtain the estimates on R_ε.

Lemma 2.3.10. *If $\varepsilon|m| < 1$ then we have the estimates*

$$\sup_{t\in[0,T]}\|R_\varepsilon(t)\| \lesssim \varepsilon(1\vee|m|) \quad \textit{and} \quad \int_0^T \|R_\varepsilon(t)\|_{H^1}dt \lesssim (1\vee|m|)\ . \tag{2.55}$$

We also have that

$$\sup_{t\in[0,T]}\|\partial_t R_\varepsilon(t)\| \lesssim \frac{1\vee|m|^2}{\varepsilon^2}\ . \tag{2.56}$$

Proof. We will first prove the bound for $\|R_\varepsilon(t)\|$. From the definition of the remainder R_ε, we have that

$$R_\varepsilon(t) = r_\varepsilon(t) + \varepsilon f_1(t) + \varepsilon^2 f_2(t)\ \ ,$$

where $f_1(t) = im\Phi_1(x/\varepsilon)e_m(x)e^{-\mu m^2 t}$ and $f_2(t) = -m^2\Phi_2(x/\varepsilon)e_m(x)e^{-\mu m^2 t}$. As a consequence, we obtain

$$\|R_\varepsilon(t)\| \lesssim \|r_\varepsilon(t)\| + \varepsilon\|f_1(t)\| + \varepsilon^2\|f_2(t)\| \lesssim \|r_\varepsilon(t)\| + \varepsilon m\ .$$

From Lemma 2.3.8 we have that

$$\begin{aligned}\|r_\varepsilon(t)\| \lesssim \|S_\varepsilon^*(t)r_\varepsilon(0)\| &+ \varepsilon\int_0^t \|S_\varepsilon^*(t-r)F_1(r)\|dr \\ &+ \varepsilon^2\int_0^t \|S_\varepsilon^*(t-r)F_2(r)\|dr + \int_0^t \|S_\varepsilon^*(t-r)(\partial_r - \mathcal{L}_\varepsilon^*)f_\varepsilon^{\mathrm{BL}}(r)\|dr\ .\end{aligned}$$

Each of the above terms shall now be bounded separately. Using the uniform boundedness of the semigroup, we have that

$$\|S_\varepsilon^*(t)r_\varepsilon(0)\| \lesssim \|r_\varepsilon(0)\| \lesssim \varepsilon m\ ,$$

which follows from (2.48). If we use the bound on $\|F_1\|$ given in Lemma 2.3.8 we have that

$$\begin{aligned}\varepsilon\int_0^t \|S_\varepsilon^*(t-r)F_1(r)\|dr &\lesssim \varepsilon\int_0^t \|F_1(r)\|dr \\ &\lesssim \varepsilon\int_0^t m^3\exp(-\mu m^2 r)dr \lesssim \varepsilon m\ .\end{aligned}$$

Similarly, we have that

$$\varepsilon^2 \int_0^t \|S_\varepsilon^*(t-r)F_2(r)\| dr \lesssim \varepsilon^2 m^2 \lesssim \varepsilon m \; .$$

Finally, from Lemma 2.3.9 we have that

$$\int_0^t \|S_\varepsilon^*(t-r)(\partial_r - \mathcal{L}_\varepsilon^*) f_\varepsilon^{\mathrm{BL}}(r)\| dr \lesssim \int_0^t \frac{m}{\varepsilon} e^{-r/\varepsilon^2} dr \lesssim \varepsilon m \; .$$

Putting all this together, we have that

$$\|R_\varepsilon(t)\| \lesssim \varepsilon m \; ,$$

whenever $\varepsilon|m| < 1$. We now seek the bound on $\|R_\varepsilon(t)\|_{H^1}$. We have that

$$\begin{aligned} \|R_\varepsilon(t)\|_{H^1} &\lesssim \|r_\varepsilon(t)\|_{H^1} + \varepsilon\|f_1(t)\|_{H^1} + \varepsilon^2\|f_2(t)\|_{H^1} \\ &\lesssim \|(1-\partial_x^2)^{1/2}(1-\mathcal{L}_\varepsilon)^{-1/2}\|\|(1-\mathcal{L}_\varepsilon)^{1/2} r_\varepsilon(t)\| + m + \varepsilon m^2 \\ &\lesssim \varepsilon^{-1}\|(1-\mathcal{L}_\varepsilon)^{1/2} r_\varepsilon(t)\| + m \; . \end{aligned}$$

Here we have used Lemma 2.3.5 to switch between the the $\mathcal{L}_\varepsilon$ and ∂_x^2 interpolation spaces. We have from Lemma 2.3.8 that

$$\begin{aligned} \|(1-\mathcal{L}_\varepsilon)^{1/2} r_\varepsilon(t)\| \lesssim \|S_\varepsilon^*(t)(1-\mathcal{L}_\varepsilon)^{1/2} r_\varepsilon(0)\| &+ \varepsilon \int_0^t \|S_\varepsilon^*(t-r)(1-\mathcal{L}_\varepsilon)^{1/2} F_1(r)\| dr \\ &+ \varepsilon^2 \int_0^t \|S_\varepsilon^*(t-r)(1-\mathcal{L}_\varepsilon)^{1/2} F_2(r)\| dr \\ &+ \int_0^t \|S_\varepsilon^*(t-r)(1-\mathcal{L}_\varepsilon)^{1/2}(\partial_r - \mathcal{L}_\varepsilon^*) f_\varepsilon^{\mathrm{BL}}(r)\| dr \; . \end{aligned}$$

From Lemma 2.3.3, we have that

$$\|S_\varepsilon^*(t)(1-\mathcal{L}_\varepsilon)^{1/2}\| \lesssim |t|^{-1/2} \; ,$$

for any $t \in (0, T]$. Therefore, we have that

$$\|S_\varepsilon^*(t)(1-\mathcal{L}_\varepsilon)^{1/2} r_\varepsilon(0)\| \lesssim |t|^{-1/2}\|r_\varepsilon(0)\| \lesssim \varepsilon m |t|^{-1/2} \; .$$

Furthermore, we have that

$$\begin{aligned}\varepsilon\int_0^t \|S_\varepsilon^*(t-r)(1-\mathcal{L}_\varepsilon)^{1/2}F_1(r)\|dr &\lesssim \varepsilon\int_0^t |t-r|^{-1/2}\|F_1(r)\|dr\\ &\lesssim \varepsilon\int_0^t m^3|t-r|^{-1/2}\exp(-\mu m^2 r)dr\\ &\lesssim \varepsilon m\left(|t|^{-1/2}+m^2\exp(-\mu m^2 t)\right)\ .\end{aligned}$$

Here we have bounded the above integral by splitting the range of integration in half. Similarly, we have that

$$\varepsilon^2\int_0^t \|S_\varepsilon^*(t-r)(1-\mathcal{L}_\varepsilon)^{1/2}F_2(r)\|dr \lesssim \varepsilon m\left(|t|^{-1/2}+m^2\exp(-\mu m^2 t)\right)\ .$$

Finally, from Lemma 2.3.9 we have that

$$\begin{aligned}&\int_0^t \|S_\varepsilon^*(t-r)(1-\mathcal{L}_\varepsilon)^{1/2}(\partial_r-\mathcal{L}_\varepsilon^*)f_\varepsilon^{\mathrm{BL}}(r)\|dr\\ &\qquad\lesssim \frac{m}{\varepsilon}\int_0^t |t-r|^{-1/2}\exp(-r/\varepsilon^2)dr\\ &\qquad\lesssim \varepsilon m\left(|t|^{-1/2}+\frac{\exp(-t/\varepsilon^2)}{\varepsilon^2}\right)\ .\end{aligned}$$

Putting this all together, along with the fact that $\varepsilon|m|<1$, we have the bound

$$\|R_\varepsilon(t)\|_{H^1}\lesssim m\left(1+|t|^{-\gamma/2}+m^2\exp(-\mu m^2 t)+\frac{\exp(-t/\varepsilon^2)}{\varepsilon^2}\right)\ ,$$

and the requested bound on $\int_0^T\|R_\varepsilon(t)\|_{H^1}dt$ follows. For the final estimate, we use the definition

$$R_\varepsilon(t)=S_\varepsilon^*(t)e_m(x)-\rho(x/\varepsilon)e_m(x)e^{-\mu m^2 t}-\hat\rho_{t/\varepsilon^2}(x/\varepsilon)e_m(x)\ .$$

We then have

$$\begin{aligned}\sup_{t\in[0,T]}\|\partial_t R_\varepsilon(t)\| &\lesssim \sup_{t\in[0,T]}\|\partial_t S_\varepsilon^*(t)e_m\|+m^2\|\rho\|+\sup_{t\in[0,T]}\frac{\|\partial_t\hat\rho\|}{\varepsilon^2}\\ &\lesssim \sup_{t\in[0,T]}\|\partial_t S_\varepsilon^*(t)e_m\|+\frac{m^2}{\varepsilon^2}\ ,\end{aligned}$$

since the boundedness of $\sup_{t\in[0,T]}\|\partial_t\hat\rho\|$ and $\|\rho\|$ are guaranteed by the smoothness of b

and σ. Due to the uniform boundedness of the semigroup $S_\varepsilon(t)$, we also have that

$$\sup_{t\in[0,T]}\|\partial_t S_\varepsilon^*(t)e_m\| = \sup_{t\in[0,T]}\|S_\varepsilon^*(t)\mathcal{L}_\varepsilon^* e_m\| \lesssim \|\mathcal{L}_\varepsilon^* e_m\| \lesssim \frac{m^2}{\varepsilon^2} ,$$

where the last inequality follows from the smoothness assumptions placed on b and σ. This proves the result. □

2.4 Convergence results

In this section, we shall state the precise formulation of the main results and then provide their proofs in full detail. The first convergence result is as follows.

Theorem 2.4.1. *Suppose u_ε satisfies* (2.19) *and the conditions given in Assumptions 2.2.2, 2.2.5 hold true. Suppose furthermore that u solves the stochastic heat equation*

$$du(x,t) = \mu\partial_x^2 u(x,t)dt + \sum_k \langle q_k,\rho\rangle e_k(x)dW_k(t) , \tag{2.57}$$

with $u(x,0)=0$. Let $s_\alpha = 0 \vee \frac{3}{2}(1-2\alpha)$, then for any $s > s_\alpha$ there exists $\theta_0(s) > 0$ such that

$$\mathbb{E}\sup_{t\in[0,T]}\|u_\varepsilon(t) - u(t)\|_{H^{-s}}^2 \lesssim \varepsilon^\theta , \tag{2.58}$$

for any $\theta < \theta_0(s)$.

Remark 2.4.2. For the interested reader, the rate of decay θ_0 given by our proof is

$$\theta_0(s) = 2\alpha \wedge \frac{4}{3}(s - s_\alpha) .$$

As stated in the introduction, the next theorem deals with the second order term of the solution u_ε, obtained by subtracting the first order term (or in our case, setting $\langle q_k,\rho\rangle = 0$) and scaling the noise up by some inverse factor of ε. We have the following result.

Theorem 2.4.3. *Suppose u_ε satisfies* (2.19) *with $\langle q_k,\rho\rangle = 0$ for all $k\in\mathbb{Z}$ and the conditions given in Assumptions 2.2.2, 2.2.6 hold true for a given $\alpha\in(0,1)$.*

Then, there exists a probability space with a sequence of Wiener processes $\{\hat{W}_k\}$ and processes $\{\hat{u}_\varepsilon\}$ that are equal in law to $\{u_\varepsilon\}$, such that

$$\lim_{\varepsilon\to0}\mathbb{E}\sup_{t\in[0,T]}\|\varepsilon^{-\alpha}\hat{u}_\varepsilon(t) - v(t)\|_{H^{-s}}^2 = 0 , \tag{2.59}$$

where v is the solution to

$$dv(x,t) = \mu\partial_x^2 v(x,t)dt + \|\bar{q}\rho\|_{-\alpha} \sum_k e_k(x) d\hat{W}_k(t) \ , \tag{2.60}$$

with $v(x,0) = 0$. The convergence (2.59) holds for any $s > \frac{3}{2}\big(\alpha \vee (1-\alpha)\big)$.

The two preceding theorems always require some decay on the coefficients q_k, in particular the results do no treat SPDEs driven by space-time white noise, where $q_k = 1$ for each $k \in \mathbb{Z}$. We know that in the space-time white noise case, the solution converges to the so-called wrong limit. The following result generalises this phenomena to a broad class of driving noise processes.

Theorem 2.4.4. *Suppose u_ε satisfies* (2.19) *and that the conditions given in Assumption 2.2.2, 2.2.7 hold true. Then, there exists a probability space with a sequence of Wiener processes $\{\hat{W}_k\}$ and processes $\{\hat{u}_\varepsilon\}$ that are equal in law to $\{u_\varepsilon\}$, such that*

$$\lim_{\varepsilon\to 0} \mathbb{E}\sup_{t\in[0,T]} \|\hat{u}_\varepsilon(t) - \hat{u}(t)\|^2_{H^{-s}} = 0 \ , \tag{2.61}$$

where $\hat{u}$ satisfies the stochastic heat equation

$$d\hat{u}(x,t) = \mu\partial_x^2 \hat{u}(x,t)dt + \sum_k (|\langle q_k, \rho\rangle|^2 - |\langle \bar{q}, \rho\rangle|^2 + \|\bar{q}\rho\|^2)^{1/2} e_k(x) d\hat{W}_k(t) \ , \tag{2.62}$$

with $\hat{u}(x,0) = 0$. The convergence (2.61) holds for any $s > s_\eta$, where

$$s_\eta = \begin{cases} 1, & \text{if } \eta \in [0,1/2] \ , \\ \frac{3}{2}(2-\eta)^{-1}, & \text{if } \eta \in [1/2,1) \ . \end{cases}$$

(Here, η is the constant appearing in Assumption 2.2.7.)

Remark 2.4.5. If one assumes that the driving noise does not depend on ε, as is the case for space-time white noise, then the assumptions can be loosened. In particular, one can easily modify the proof of Theorem 2.4.1 to show the following. Suppose u_ε satisfies (2.19) with q_k constants and that u satisfies (2.57), then

$$\lim_{\varepsilon\to 0} \mathbb{E}\sup_{t\in[0,T]} \|u_\varepsilon(t) - u(t)\|^2_{H^{-s}} = 0 \ ,$$

for s large enough. Hence, we can still prove the limit, but at the expense of the rate of convergence. A similar result holds for Theorem 2.4.4, in the case of constant q_k, in that we can weaken the assumption to just $q_k \to \bar{q}$, and still prove the limit (2.61).

One might ask what happens if we approximate the noise by a smoother infinite dimensional Gaussian process, say W_ε, which, for nonzero ε falls into the class of the classical (unsurprising) case, but as ε tends to zero, approaches something as irregular as space-time white noise, for instance. To this end, let φ be a smooth test function on $\mathbb{R}$ with compact support and $\varphi(0) = 1$. We define the *smoothened version* of (2.19) by

$$du_\varepsilon(t) = \mathcal{L}_\varepsilon u_\varepsilon(t)dt + \sum_k \varphi(\varepsilon k) q_k(x/\varepsilon) e_k dW_k(t) \; .$$

This smoothening procedure consists in taking the convolution of the noise with a scaled version of the function $\tilde{\varphi}$, where $\tilde{\varphi}$ is the inverse Fourier transform of φ. The following corollary illustrates the transition between the classical case and the surprising case.

Corollary 2.4.6. Suppose u_ε satisfies the smoothened version of (2.19), as defined above and that Assumptions 2.2.2, 2.2.7 hold true. Suppose furthermore that

$$d\hat{u}(t) = \mu \partial_x^2 \hat{u}(t)dt + \sum_k \left(|\langle q_k, \rho\rangle|^2 - |\langle \bar{q}, \rho\rangle|^2 + \|(\bar{q}\rho) \star \tilde{\varphi}\|^2 \right)^{1/2} e_k d\hat{W}_k(t) \qquad (2.63)$$

Then $u_\varepsilon \to \hat{u}$ in precisely the same sense as claimed in Theorem 2.4.4.

Remark 2.4.7. If we take $\tilde{\varphi} = 1$, then we recover Theorem 2.4.1. If on the other hand, we take $\varphi = 1$ (so that $\tilde{\varphi} = \delta$), then we recover Theorem 2.4.4, so that we can view this corollary as an interpolation between the two theorems.

The proof of Corollary 2.4.6 is given on page 60 below. Before proving these results, we need a few specialised lemmas. The first technical lemma that we require will essentially provide us with a bound on the norm of the multiplication operator from H^{-s} to H^{-s}, where the multiplier function is highly oscillatory.

Lemma 2.4.8. *For any $f \in H^1$ we have that*

$$\|(1 - \partial_x^2)^{-s/2} f^\varepsilon (1 - \partial_x^2)^{s/2}\|_{L^2 \to L^2} \lesssim \varepsilon^{-s} \|f\|_{H^1} \; , \qquad (2.64)$$

where $f^\varepsilon(x) = f(x/\varepsilon)$ denotes the corresponding multiplication operator.

Proof. We will equivalently prove that

$$\|f^\varepsilon u\|_{H^{-s}} \lesssim \varepsilon^{-s} \|u\|_{H^{-s}} \|f\|_{H^1} \; ,$$

this is done once more using Caldéron-Lions interpolation theorem [RS75]. For $s = 0$, the claim holds simply because

$$\|f^\varepsilon u\| \lesssim \|f\|_{L^\infty} \|u\| \lesssim \|f\|_{H^1} \|u\| \; ,$$

which follows from a standard Sobolev embedding. For $s = 1$ we also have the simple result for negative Sobolev norms

$$\|f^\varepsilon u\|_{H^{-1}} \leq \|f^\varepsilon\|_{H^1}\|u\|_{H^{-1}} \lesssim \frac{1}{\varepsilon}\|f\|_{H^1}\|u\|_{H^{-1}} \ .$$

The Caldéron-Lions theorem then implies that the multiplication operator has norm

$$\|f^\varepsilon\|_{H^{-s}\to H^{-s}} \lesssim (\|f\|_{H^1})^{1-s}(\frac{1}{\varepsilon}\|f\|_{H^1})^s = \varepsilon^{-s}\|f\|_{H^1} \ ,$$

which proves the lemma. □

In the next lemma, we obtain a control on the variance of the Gaussian process u_ε in the space of continuous functions taking values in $L^2[0, 2\pi]$. This will be useful in deciding which Sobolev spaces contain the solutions uniformly in ε and hence determining where convergence occurs.

Lemma 2.4.9. *Suppose u_ε satisfies* (2.19) *and the conditions given in Assumptions 2.2.2, 2.2.5 hold true. If $\alpha \in (1/2, 1)$ then we have that*

$$\mathbb{E}\mathrm{sup}_{t\in[0,T]}\|u_\varepsilon(t)\|^2 \leq C_T \ . \tag{2.65}$$

Otherwise, if $\alpha \in (0, 1/2]$ we have that

$$\mathbb{E}\mathrm{sup}_{t\in[0,T]}\|u_\varepsilon(t)\|^2 \lesssim \varepsilon^{4\alpha-2-\delta} \ , \tag{2.66}$$

for any $\delta \in (0, 2)$.

Proof. We utilise the fact that the semigroup $S_\varepsilon(t)$ is a contraction semigroup when the domain is taken to be $L^2(\rho_\varepsilon)$ with the corresponding norm and inner product, as introduced in Lemma 2.3.2. This follows from the fact that the generator $\mathcal{L}_\varepsilon$ is self-adjoint in this weighted space combined with the fact that the generator has non-positive spectrum. One can therefore apply standard martingale-type inequalities for stochastic convolutions [DPZ92] to obtain

$$\begin{aligned}\mathbb{E}\mathrm{sup}_{t\in[0,T]}\|u_\varepsilon(t)\|^2_{\rho_\varepsilon} &= \mathbb{E}\mathrm{sup}_{t\in[0,T]}\Big\|\int_0^t S_\varepsilon(t-s)Q_\varepsilon dW(s)\Big\|^2_{\rho_\varepsilon} \\ &\lesssim \int_0^T \|S_\varepsilon(t)Q_\varepsilon\|^2_{HS,\rho_\varepsilon}dt \ ,\end{aligned}$$

where $\|\cdot\|_{HS,\rho_\varepsilon}$ denotes the Hilbert-Schmidt norm for operators mapping $L^2(\rho_\varepsilon)$ into itself. We have already seen in Lemma 2.3.2 that the norms $\|\cdot\|$ and $\|\cdot\|_{\rho_\varepsilon}$ are equivalent with their ratios bounded uniformly in $\varepsilon \in (0,1)$. One can easily show that the same is true for the Hilbert-Schmidt norms $\|\cdot\|_{\mathrm{HS}}$ and $\|\cdot\|_{HS,\rho_\varepsilon}$. Hence we have that

$$\mathbb{E}\mathrm{sup}_{t\in[0,T]}\|u_\varepsilon(t)\|^2 \lesssim \int_0^T \|S_\varepsilon(t)Q_\varepsilon\|_{\mathrm{HS}}^2 dt \; . \tag{2.67}$$

Since $\alpha \in (1/2,1)$ implies that the noise is Hilbert-Schimdt, the result (2.65) follows immediately from (2.67). Now suppose $\alpha \in (0,1/2]$, then

$$\begin{aligned}\|S_\varepsilon(t)Q_\varepsilon\|_{\mathrm{HS}}^2 &= \sum_{k\in\mathbb{Z}} \|S_\varepsilon(t)q_k^\varepsilon e_k\|^2 \\ &\lesssim \sum_{k\in\mathbb{Z}} (1\wedge |k|^{-2\alpha})\|S_\varepsilon(t)\bar{q}_k^\varepsilon e_k\|^2 \; ,\end{aligned}$$

where $\bar{q}_k = q_k/\|q_k\|$ and $\bar{q}_k^\varepsilon = \bar{q}_k(\cdot/\varepsilon)$. However, we can trade the smoothness of the $\bar{q}_k$ to obtain a little more decay as k gets large. In particular, we can write

$$\|S_\varepsilon(t)\bar{q}_k^\varepsilon e_k\|^2 = (1+k^2)^{-\nu}\|S_\varepsilon(t)\bar{q}_k^\varepsilon(1-\partial_x^2)^{\nu/2}e_k\|^2 \; ,$$

and using estimates from Lemmas 2.3.5 and 2.4.8 we have that

$$\begin{aligned}\|S_\varepsilon(t)\bar{q}_k^\varepsilon(1-\partial_x^2)^{\nu/2}e_k\|^2 &\le \|S_\varepsilon(t)(1-\mathcal{L}_\varepsilon)^{\nu/2}\|^2\|(1-\mathcal{L}_\varepsilon)^{-\nu/2}(1-\partial_x^2)^{\nu/2}\|^2 \\ &\quad \times \|(1-\partial_x^2)^{-\nu/2}\bar{q}_k^\varepsilon(1-\partial_x^2)^{\nu/2}e_k\|^2 \\ &\lesssim (t^{-\nu})(\varepsilon^{-2\nu})(\varepsilon^{-2\nu}\|\bar{q}_k\|_{H^1}^2) \; .\end{aligned}$$

Therefore, we have that

$$\|S_\varepsilon(t)Q_\varepsilon\|_{\mathrm{HS}} \lesssim \varepsilon^{-2\nu}t^{-\nu/2}\left(\sum_{k\in\mathbb{Z}}(1\wedge|k|^{-2\alpha-2\nu})\|\bar{q}_k\|_{H^1}^2\right)^{1/2} \; ,$$

for any $\nu \in [0,1)$. If we set $\nu = 1/2 - \alpha + \delta$ then, given the uniform boundedness of $\|\bar{q}_k\|_{H^1}$, the sum over $k \in \mathbb{Z}$ is clearly convergent and upon substitution into (2.67), the result (2.66) follows. □

The following lemma is simply a restatement of the Kolmogorov continuity criterion [RY99].

Lemma 2.4.10. *Suppose $\{\phi(t)\}_{t\in[0,T]}$ is a complex valued stochastic process, such that for* *every $q > 2$ there exists K_q satisfying*

$$(\mathbb{E}|\phi(t)|^q)^{1/q} \leq K_q \left(\mathbb{E}|\phi(t)|^2\right)^{1/2} \ ,$$
$$(\mathbb{E}|\phi(t)-\phi(s)|^q)^{1/q} \leq K_q \left(\mathbb{E}|\phi(t)-\phi(s)|^2\right)^{1/2} \ ,$$

for any $s,t \in [0,T]$. Suppose furthermore that there exists $\delta > 0$, $K_0 > 0$ such that

$$\mathbb{E}|\phi(t)-\phi(s)|^2 \leq K_0|t-s|^\delta \ ,$$

for any $s,t \in [0,T]$, where the constant K_0 depends only on the sequence K_q. Then for any $p > 0$ there exists $C > 0$ such that

$$\mathbb{E}\mathrm{sup}_{t\in[0,T]}|\phi(t)|^p \leq C(K_0 + \mathbb{E}|\phi(0)|^2)^{p/2} \ .$$

The next and final result is needed in order to trade some regularity of a pair of functions for some extra decay on the Fourier modes of products of those functions.

Lemma 2.4.11. *Suppose $f, g \in H^1$ taking values in $\mathbb{R}$, then for any $\nu \in [0,1]$ and each $k \in \mathbb{Z}$, we have that*

$$|\langle fe_k, g\rangle|^2 \lesssim (1 \wedge |k|^{-2\nu}) \left(\|f\|\|g\|\right)^{2-2\nu} \left(\|f\|_{H^1}\|g\|_{H^1}\right)^{2\nu} \ . \tag{2.68}$$

Proof. We have that

$$\begin{aligned}
|\langle fe_k, g\rangle|^2 &= |\langle fe_k, g\rangle|^{2-2\nu}|\langle fg, e_{-k}\rangle|^{2\nu} \\
&= (1+k^2)^{-\nu}|\langle fe_k, g\rangle|^{2-2\nu}|\langle(1-\partial_x^2)^{1/2}(fg), e_{-k}\rangle|^{2\nu} \\
&\lesssim (1 \wedge |k|^{-2\nu})\|fe_k\|^{2-2\nu}\|g\|^{2-2\nu}\|fg\|_{H^1}^{2\nu} \\
&\lesssim (1 \wedge |k|^{-2\nu})\|f\|^{2-2\nu}\|g\|^{2-2\nu}\|f\|_{H^1}^{2\nu}\|g\|_{H^1}^{2\nu} \ .
\end{aligned}$$

In the last inequality we have used the fact that H^1 is a Banach algebra [AF03]. This proves the lemma. □

We now have all the necessary machinery to prove our first theorem.

Proof of Theorem 2.4.1. To start off, we take the object we wish to bound and split it into two parts. Using the identity

$$\|\cdot\|_{H^{-s}}^2 = \sum_{m\in\mathbb{Z}} |\langle\cdot, e_m\rangle|^2(1+m^2)^{-s} \ ,$$

We obtain

$$
\begin{aligned}
\mathbb{E}\sup_{t\in[0,T]}\|u_\varepsilon(t)-u(t)\|_{H^{-s}}^2 \lesssim \sum_{|m|<\varepsilon^{-\beta}} &\mathbb{E}\sup_{t\in[0,T]}|\langle u_\varepsilon(t)-u(t),e_m\rangle|^2(1+m^2)^{-s} \\
+\mathbb{E}\sup_{t\in[0,T]} &\sum_{|m|\geq\varepsilon^{-\beta}} |\langle u_\varepsilon(t)-u(t),e_m\rangle|^2(1+m^2)^{-s}
\end{aligned}
$$

for any $\beta \in (0,1)$. The idea of the proof is to use standard homogenisation techniques for the low modes ($|m| < \varepsilon^{-\beta}$), while using rather soft *a priori* bounds for the high modes ($|m| \geq \varepsilon^{-\beta}$). We then choose β in the right way to balance the two contributions. We shall bound the low modes first. Here, we use the fact that

$$
\langle u_\varepsilon(t), e_m\rangle = \sum_k \int_0^t \langle q_k^\varepsilon e_k, S_\varepsilon^*(t-s)e_m\rangle\, dW_k(s)\ ,
$$

and then, using the results of Subsection 2.3.7, approximate the semigroup as follows

$$
S_\varepsilon^*(t-s)e_m = \rho(x/\varepsilon)e_m(x)e^{-\mu m^2(t-s)} + \hat{\rho}_{(t-s)/\varepsilon^2}(x/\varepsilon)e_m(x) + R_\varepsilon(x,t-s)\ ,
$$

so that

$$
\begin{aligned}
\langle u_\varepsilon(t), e_m\rangle = &\sum_k \langle q_k^\varepsilon e_k, \rho^\varepsilon e_m\rangle \int_0^t e^{-\mu m^2(t-s)} dW_k(s) \\
&+ \sum_k \int_0^t \langle q_k^\varepsilon e_k, \hat{\rho}^\varepsilon_{(t-s)/\varepsilon^2} e_m\rangle dW_k(s) \\
&+ \sum_k \int_0^t \langle q_k^\varepsilon e_k, R_\varepsilon(x,t-s)\rangle dW_k(s)\ ,
\end{aligned}
$$

where $\rho^\varepsilon(x) = \rho(x/\varepsilon)$ and similarly for all other instances of the superscript ε. We can simplify the terms above using the fact that, for fixed $|m| < \varepsilon^{-\beta} \ll \varepsilon^{-1}$ and varying $k \in \mathbb{Z}$ the expression $\langle q_k^\varepsilon e_k, \rho^\varepsilon e_m\rangle$ is zero, unless $k = m + l/\varepsilon$ for some $l \in \mathbb{Z}$. We can see this, for example, by performing a Fourier expansion on both q_k and ρ. Moreover,

$$
\sum_{k\in\mathbb{Z}} \langle q_k^\varepsilon e_k, \rho^\varepsilon e_m\rangle F_k = \sum_{l\in\mathbb{Z}} \langle q_{m+l/\varepsilon} e_l, \rho\rangle F_{m+l/\varepsilon}\ ,
$$

for any sequence $\{F_k\}_{k\in\mathbb{Z}}$. Therefore,

$$\sum_k \langle q_k^\varepsilon e_k, \rho^\varepsilon e_m\rangle \int_0^t e^{-\mu m^2(t-s)} dW_k(s)$$
$$= \langle q_m, \rho\rangle \int_0^t e^{-\mu m^2(t-s)} dW_m(s) + \sum_{l\neq 0} \langle q_{m+l/\varepsilon} e_l, \rho\rangle \int_0^t e^{-\mu m^2(t-s)} dW_{m+l/\varepsilon}(s)\ .$$

Similarly, we can write

$$\sum_k \int_0^t \langle q_k^\varepsilon e_k, \hat\rho^\varepsilon_{(t-s)/\varepsilon^2} e_m\rangle dW_k(s)$$
$$= \sum_l \int_0^t \langle q_{m+l/\varepsilon} e_l, \hat\rho_{(t-s)/\varepsilon^2}\rangle dW_{m+l/\varepsilon}(s)\ .$$

It is easy to see that $\langle u(t), e_m\rangle = \langle q_m, \rho\rangle \int_0^t e^{-\mu m^2(t-s)} dW_m(s)$ and we can therefore write

$$\begin{aligned}\langle u_\varepsilon(t) - u(t), e_m\rangle &= \sum_{l\neq 0} \langle q_{m+l/\varepsilon} e_l, \rho\rangle \int_0^t e^{-\mu m^2(t-s)} dW_{m+l/\varepsilon}(s) \\ &\quad + \sum_l \int_0^t \langle q_{m+l/\varepsilon} e_l, \hat\rho_{(t-s)/\varepsilon^2}\rangle dW_{m+l/\varepsilon}(s) \\ &\quad + \sum_k \int_0^t \langle q_k^\varepsilon e_k, R_\varepsilon(t-s)\rangle dW_k(s)\ .\end{aligned}$$

We then bound separately each of the three sums in this expression. In order to streamline the presentation, we state these bounds as separate lemmas, the proof of which is given below.

Lemma 2.4.12. *For $\varepsilon|m| < 1/2$, one has the bound*

$$\mathbb{E}\sup_{t\in[0,T]} \left|\sum_{l\neq 0} \langle q_{m+l/\varepsilon} e_l, \rho\rangle \int_0^t e^{-\mu m^2(t-s)} dW_{m+l/\varepsilon}(s)\right|^2 \lesssim \frac{\varepsilon^{2\alpha}}{1\vee m^2}\ , \qquad (2.69a)$$

for any $\alpha > 0$.

Lemma 2.4.13. *For $\varepsilon|m| < 1/2$, one has the bound*

$$\mathbb{E}\sup_{t\in[0,T]} \left|\sum_{l\in\mathbb{Z}} \int_0^t \langle q_{m+l/\varepsilon} e_l, \hat\rho_{(t-s)/\varepsilon^2}\rangle dW_{m+l/\varepsilon}(s)\right|^2 \lesssim \varepsilon^{2-2\delta}\ , \qquad (2.69b)$$

for any sufficiently small $\delta > 0$.

Lemma 2.4.14. *For $\varepsilon|m| < 1/2$, the bound*

$$\mathbb{E}\sup_{t\in[0,T]} \left|\sum_k \int_0^t \langle q_k^\varepsilon e_k, R_\varepsilon(t-s)\rangle dW_k(s)\right|^2 \lesssim \frac{\varepsilon^{4\alpha}+\varepsilon^2}{\varepsilon^{7\delta}}(1\vee m^{2+\delta}) , \tag{2.69c}$$

holds for any sufficiently small $\delta > 0$ and for any $\alpha > 0$

We now use these bounds to prove the claim made in the statement of the theorem in the case $\alpha \in (0, 1/2]$, and the case $\alpha \in (1/2, 1)$ will follow similarly. Inserting the bounds above into

$$\begin{aligned}\sum_{|m|<\varepsilon^{-\beta}} &\mathbb{E}\sup_{t\in[0,T]} |\langle u_\varepsilon(t) - u(t), e_m\rangle|^2 (1+m^2)^{-s} \\ &\lesssim \varepsilon^{2\alpha} \sum_{|m|<\varepsilon^{-\beta}} \frac{(1+m^2)^{-s}}{1\vee m^2} + \varepsilon^{2-2\delta} \sum_{|m|<\varepsilon^{-\beta}} (1+m^2)^{-s} \\ &\quad + \varepsilon^{4\alpha-7\delta} \sum_{|m|<\varepsilon^{-\beta}} (1\vee m^{2+\delta})(1+m^2)^{-s} \\ &\lesssim \varepsilon^{2\alpha} + \varepsilon^{2-\beta-2\delta} + \varepsilon^{4\alpha-(3-2s)\beta-(2\beta+7)\delta} ,\end{aligned} \tag{2.70}$$

for any $s > 0$. For the high modes on the other hand, we have the straightforward bound

$$\begin{aligned}\mathbb{E}\sup_{t\in[0,T]} \sum_{|m|\geq\varepsilon^{-\beta}} &|\langle u_\varepsilon(t) - u(t), e_m\rangle|^2 (1+m^2)^{-s} \\ &\lesssim \varepsilon^{2\beta s}\left(\mathbb{E}\sup_{t\in[0,T]}\|u_\varepsilon(t)\|^2 + \mathbb{E}\sup_{t\in[0,T]}\|u(t)\|^2\right) \lesssim \varepsilon^{2\beta s+4\alpha-2-\delta'} ,\end{aligned} \tag{2.71}$$

where we have used Lemma 2.4.9 combined with the fact that

$$\mathbb{E}\sup_{t\in[0,T]}\|u(t)\|^2 \lesssim 1 ,$$

which is easily verified. Since δ and δ' can be chosen arbitrarily small and since $\beta \in (0, 1)$, both the low modes and high modes will be bounded by a multiple of ε^θ, where $\theta < \theta_0$ and

$$\theta_0 = \min\{2\alpha, 1+2\alpha-\beta, 4\alpha-(3-2s)\beta, 2\beta s+4\alpha-2\} .$$

Since $\alpha > 0$ and $\beta \in (0, 1)$ we will find $\theta_0 > 0$ provided that $4\alpha - (3-2s)\beta > 0$ and $2\beta s + 4\alpha - 2 > 0$ are both satisfied. That is, the result (2.58) will hold for $s > 0$ if we can find $\beta \in (0, 1)$ such that

$$\frac{1-2\alpha}{s} < \beta < \frac{4\alpha}{3-2s} . \tag{2.72}$$

A simple diagram verifies that, for fixed $\alpha \in (0, 1/2]$ we can always find such a β provided

$s > s_\alpha$ where

$$s_\alpha = 0 \vee \frac{3}{2}(1-2\alpha) ,$$

as in the statement of the theorem. Moreover, one can also show that the optimal value of θ is given by

$$\theta_0(s,\alpha) = 2\alpha \wedge \left(4\alpha - 2 + \frac{4s}{3}\right) = 2\alpha \wedge \left(\frac{4}{3}(s - s_\alpha)\right) .$$

which only takes positive values when $s > s_\alpha$.

The case $\alpha \in (1/2, 1)$ is actually slightly easier, and we obtain the same bounds on the low and high modes as in (2.70) and (2.71), but with α replaced by $1/2$ and $\delta' = 0$. Hence, the result (2.58) will hold for $s > 0$ if we can find $\beta \in (0,1)$ such that

$$0 < \beta < \frac{2}{3-2s} . \tag{2.73}$$

One can always find such a β, provided $s > 0$ is small enough. Moreover, one can also show that the optimal value of θ is given in this case by

$$\theta_0(s) = 1 \wedge \frac{4}{3}s .$$

This proves the claims made in the statement of the theorem. □

It thus remain to show that the bounds (2.69) hold.

Proof of Lemma 2.4.12. Starting with (2.69a), we have that

$$\begin{aligned}
&\mathbb{E}\mathrm{sup}_{t\in[0,T]} \left| \sum_{l\neq 0} \langle q_{m+l/\varepsilon} e_l, \rho\rangle \int_0^t e^{-\mu m^2(t-s)} dW_{m+l/\varepsilon}(s) \right|^2 \\
&\qquad = \left(\sum_{l\neq 0} |\langle q_{m+l/\varepsilon} e_l, \rho\rangle|^2 \right) \mathbb{E}\mathrm{sup}_{t\in[0,T]} \left| \int_0^t e^{-\mu m^2(t-s)} dB(s) \right|^2 \\
&\qquad \lesssim \sum_{l\neq 0} \frac{|\langle q_{m+l/\varepsilon} e_l, \rho\rangle|^2}{1 \vee m^2} .
\end{aligned}$$

If $\alpha \in (1/2, 1)$ then

$$\sum_{l\neq 0} |\langle q_{m+l/\varepsilon} e_l, \rho\rangle|^2 \lesssim \sum_{l\neq 0} \|q_{m+l/\varepsilon}\|^2 \|\rho\|_\infty^2 \lesssim \sum_{l\neq 0} 1 \wedge |m + l/\varepsilon|^{-2\alpha} .$$

Assume for now that $m \geq 0$, the case $m < 0$ will follow similarly. Recalling that $\varepsilon|m| <$

$1/2$ by assumption, we can bound the above by

$$\varepsilon^{2\alpha} \sum_{l \neq 0} |\varepsilon m + l|^{-2\alpha} \lesssim \varepsilon^{2\alpha} \left(\sum_{l \geq 1} |l|^{-2\alpha} + \sum_{l \geq 1} |l - 1/2|^{-2\alpha} \right) \lesssim \varepsilon^{2\alpha} \ .$$

Now suppose $\alpha \in (0, 1/2]$. Using Lemma 2.4.11 with $\nu = 1$, we have the following bound

$$\begin{aligned} \sum_{l \neq 0} \langle q_{m+l/\varepsilon} e_l, \rho \rangle^2 &\lesssim \sum_{l \neq 0} (1 \wedge |l|^{-2}) \|q_{m+l/\varepsilon}\|^2 \|\bar{q}_{m+l/\varepsilon}\|_{H^1}^2 \|\rho\|_{H^1}^2 \\ &\lesssim \sum_{l \neq 0} |l|^{-2} \|q_{m+l/\varepsilon}\|^2 \ . \end{aligned}$$

The boundedness of $\|\rho\|_{H^1}$ is guaranteed by Assumption 2.2.2 and the uniform boundedness of $\|\bar{q}_k\|_{H^1}$ is guaranteed by Assumption 2.2.5. Moreover, we have that

$$\sum_{l \neq 0} |l|^{-2} \|q_{m+l/\varepsilon}\|^2 \lesssim \sum_{l \neq 0} |l|^{-2} |m + l/\varepsilon|^{-2\alpha} \ .$$

We will now show that this sum decays like $\varepsilon^{2\alpha}$. Since $\varepsilon|m| < 1/2$ it follows that $|\varepsilon m + l|^{-2\alpha} \leq |l - 1/2|^{-2\alpha}$ for $|l| \geq 1$. Therefore

$$\sum_{l \neq 0} |l|^{-2} |m + l/\varepsilon|^{-2\alpha} = \varepsilon^{2\alpha} \sum_{l \neq 0} |l|^{-2} |\varepsilon m + l|^{-2\alpha} \lesssim \varepsilon^{2\alpha} \sum_{l \neq 0} |l|^{-2} \ .$$

This proves (2.69a). □

Proof of Lemma 2.4.13. For both (2.69b) and (2.69c) we are trying to bound objects of the form

$$\phi(t) = \sum_k \int_0^t f_k(t - r) dw_k(r) \ ,$$

where the w_k are independent Brownian motions and each f_k takes values in $\mathbb{C}$. Since $\phi(t)$ is a Gaussian process, we may apply Lemma 2.4.10. Thus, if we can show that

$$\mathbb{E}|\phi(t) - \phi(s)|^2 \leq K_\delta(\varepsilon)|t - s|^\delta \ ,$$

then it follows that

$$\mathbb{E}\mathrm{sup}_{t \in [0,T]} |\phi(t)|^2 \lesssim K_\delta(\varepsilon) \ .$$

In general, we have that

$$\mathbb{E}|\phi(t)-\phi(s)|^2 = \mathbb{E}\left|\sum_k \int_s^t f_k(t-r)dw_k(r) + \int_0^s (f_k(t-r)-f_k(s-r))dw_k(r)\right|^2$$
$$\lesssim \sum_k \int_s^t |f_k(t-r)|^2 dr + \sum_k \int_0^s |f_k(t-r)-f_k(s-r)|^2 dr \;.$$

Note that the Brownian motions w_k are not truly independent due to the requirement $W_k = W^*_{-k}$. However, one can easily check that the above bound still holds. We then have that

$$\sum_k \int_s^t |f_k(t-r)|^2 dr = \sum_l \int_s^t |\langle q_{m+l/\varepsilon} e_l, \hat{\rho}_{(t-r)/\varepsilon^2}\rangle|^2 dr \;. \tag{2.74}$$

If $\alpha \in (1/2, 1)$ then we can bound the above by

$$\sum_l \|q_{m+l/\varepsilon}\|^2 \int_s^t \|\hat{\rho}_{(t-r)/\varepsilon^2}\|^2 dr \;. \tag{2.75}$$

From Lemma 2.3.3 we have that

$$\|\hat{\rho}_{r/\varepsilon^2}\| = \|S^*(r/\varepsilon^2)(1-\rho)\| \lesssim \exp(-r/\varepsilon^2) \;. \tag{2.76}$$

Moreover, since the sum over l is finite when $\alpha \in (1/2, 1)$ we can apply Hölder's inequality to (2.75) to obtain

$$\sum_l \|q_{m+l/\varepsilon}\|^2 \int_s^t \|\hat{\rho}_{(t-r)/\varepsilon^2}\|^2 dr \lesssim |t-s|^\delta \left(\int_s^t (\exp(-r/\varepsilon^2))^{2/(1-\delta)} dr\right)^{1-\delta}$$
$$\lesssim \varepsilon^{2-2\delta}|t-s|^\delta \;. \tag{2.77}$$

Now suppose $\alpha \in (0, 1/2]$. Using Lemma 2.4.11 with $\nu = 1$ we can bound (2.74) by

$$\sum_l (1 \wedge |l|^{-2}) \|q_{m+l/\varepsilon}\|^2_{H^1} \int_s^t \|\hat{\rho}_{(t-r)/\varepsilon^2}\|^2_{H^1} dr \tag{2.78}$$

Since $\|\bar{q}_k\|_{H^1}$ is bounded uniformly in k, the sum over l is finite. Furthermore, from Lemma 2.3.3 we see that

$$\|\hat{\rho}_{r/\varepsilon^2}\|_{H^1} = \left\|(1-\partial_x^2)^{1/2} S^*(1) S^*(r/\varepsilon^2 - 1)(1-\rho)\right\| \lesssim \exp(-r/\varepsilon^2) \;. \tag{2.79}$$

Therefore, with an application of Hölder's inequality, we can bound (2.78) by

$$|t-s|^{\delta}\left(\int_s^t \|\hat{\rho}_{(t-r)/\varepsilon^2}\|_{H^1}^{2/(1-\delta)}dr\right)^{1-\delta} \lesssim |t-s|^{\delta}\varepsilon^{2-2\delta} .$$

We also have that

$$\sum_k \int_0^s |f_k(t-r)-f_k(s-r)|^2 dr = \sum_l \int_0^s |\langle q_{m+l/\varepsilon} e_l, \hat{\rho}_{(t-r)/\varepsilon^2} - \hat{\rho}_{(s-r)/\varepsilon^2}\rangle|^2 dr . \quad (2.80)$$

If $\alpha \in (1/2, 1)$ then, as in the estimation of (2.74) we can bound the above by

$$\begin{aligned}\int_0^s \|\hat{\rho}_{(t-r)/\varepsilon^2} - \hat{\rho}_{(s-r)/\varepsilon^2}\|^2 dr &\lesssim \frac{|t-s|^{\delta}}{\varepsilon^{2\delta}}\Big(\sup\nolimits_{t\in[0,T]}\|\partial_t\hat{\rho}_t\|\Big)^{\delta} \\ &\qquad \times \int_0^s \|\hat{\rho}_{(t-r)/\varepsilon^2} - \hat{\rho}_{(s-r)/\varepsilon^2}\|^{2-\delta} dr \\ &\lesssim |t-s|^{\delta}\varepsilon^{-2\delta}\int_0^T \|\hat{\rho}_{r/\varepsilon^2}\|^{2-\delta}dr \lesssim \varepsilon^{2-2\delta}|t-s|^{\delta} .\end{aligned}$$

Here we have used the fact that

$$\|\hat{\rho}_{(t-r)/\varepsilon^2} - \hat{\rho}_{(s-r)/\varepsilon^2}\|^{\delta} \leq |t-s|^{\delta}\sup\nolimits_{t\in[0,T]}\|\partial_t\hat{\rho}_{t/\varepsilon^2}\|^{\delta} \lesssim \frac{|t-s|^{\delta}}{\varepsilon^{2\delta}}\sup\nolimits_{t\in[0,T]}\|\partial_t\hat{\rho}_t\|^{\delta} .$$

The fact that $\|\partial_t\hat{\rho}_t\|$ is bounded uniformly in time follows from

$$\|\partial_t\hat{\rho}_t\| = \|\mathcal{L}S(t)(1-\rho)\| = \|S(t)\mathcal{L}(1-\rho)\| \lesssim \|\mathcal{L}(1-\rho)\| \lesssim \|\rho\|_{\mathcal{C}_b^2} ,$$

which, by Remark 2.2.4, is finite. Now suppose $\alpha \in (0, 1/2]$. Using Lemma 2.4.11 with $\nu = 3/4$ and arguments similar to those used in the estimation of (2.74) we can bound (2.80) by

$$\begin{aligned}&\int_0^s \|\hat{\rho}_{(t-r)/\varepsilon^2} - \hat{\rho}_{(s-r)/\varepsilon^2}\|^{1/2}\|\hat{\rho}_{(t-r)/\varepsilon^2} - \hat{\rho}_{(s-r)/\varepsilon^2}\|_{H^1}^{3/2} dr \\ &\quad\lesssim \frac{|t-s|^{\delta}}{\varepsilon^{2\delta}}\Big(\sup\nolimits_{t\in[0,T]}\|\partial_t\hat{\rho}_t\|\Big)^{\delta} \\ &\qquad\times \int_0^s \|\hat{\rho}_{(t-r)/\varepsilon^2} - \hat{\rho}_{(s-r)/\varepsilon^2}\|^{1/2-\delta}\|\hat{\rho}_{(t-r)/\varepsilon^2} - \hat{\rho}_{(s-r)/\varepsilon^2}\|_{H^1}^{3/2} dr \\ &\quad\lesssim \varepsilon^{2-2\delta}|t-s|^{\delta} .\end{aligned}$$

To bound the integral term, we have used estimates (2.76) and (2.79). Putting this all together, we have that $K_\delta(\varepsilon) = \varepsilon^{2-2\delta}$, which proves estimate (2.69b). □

Proof of Lemma 2.4.14. We use the same strategy as in the proof of Lemma 2.4.13. We see that,

$$\sum_k \int_s^t |f_k(t-r)|^2 dr = \sum_k \int_s^t |\langle q_k^\varepsilon e_k, R_\varepsilon(t-r)\rangle|^2 dr \; . \tag{2.81}$$

If $\alpha \in (1/2, 1)$ then we can bound the above by

$$\left(\sum_k \|q_k\|^2\right) \int_s^t \|R_\varepsilon(t-r)\|^2 dr \lesssim \varepsilon^2 m^2 |t-s| \; .$$

Here we have used the finiteness of the sum over k as well as Lemma 2.3.10 to bound the remainder term uniformly in time. Suppose that $\alpha \in (0, 1/2]$. Using the Lemma 2.4.11, we can bound (2.81) by

$$\begin{aligned} &\sum_k (1 \wedge |k|^{-2\nu}) \|q_k\|^2 \|\bar{q}_k^\varepsilon\|_{H^1}^{2\nu} \int_s^t \|R_\varepsilon(t-r)\|^{2-2\nu} \|R_\varepsilon(t-r)\|_{H^1}^{2\nu} dr \\ &\quad \lesssim \sum_k (1 \wedge |k|^{-2\nu}) \frac{\|q_k\|^2}{\varepsilon^{2\nu}} |t-s|^\delta \left(\int_s^t \|R_\varepsilon(t-r)\|^{\frac{2-2\nu}{1-\delta}} \|R_\varepsilon(t-r)\|_{H^1}^{\frac{2\nu}{1-\delta}} dr\right)^{1-\delta} \; , \end{aligned}$$

for any $\nu \in [0, 1]$. Here we have used the fact that $\|\bar{q}_k^\varepsilon\|_{H^1} \leq \varepsilon^{-1} \|\bar{q}_k\|_{H^1} \lesssim \varepsilon^{-1}$ and then applied Hölder's inequality to the integral. Choose $\nu \in (0, 1/2)$ such that $\alpha + \nu > 1/2$, to guarantee that the above sum is bounded. Using the estimates on the remainder R_ε given in Lemma 2.3.10 we have that

$$\begin{aligned} &\int_s^t \|R_\varepsilon(t-r)\|^{(2-2\nu)/(1-\delta)} \|R_\varepsilon(t-r)\|_{H^1}^{2\nu/(1-\delta)} dr \\ &\qquad \lesssim (\varepsilon m)^{(2-2\nu)/(1-\delta)} \int_0^T \|R_\varepsilon(r)\|^{2\nu/(1-\delta)} dr \; . \end{aligned}$$

For any $\nu \in [0, 1/2)$, we can choose δ small enough that $2\nu/(1-\delta) < 1$ and hence, by Jensen's inequality

$$\int_0^T \|R_\varepsilon(r)\|^{2\nu/(1-\delta)} dr \leq \left(\int_0^T \|R_\varepsilon(r)\| dr\right)^{2\nu/(1-\delta)} \lesssim m^{2\nu/(1-\delta)} \; ,$$

which follows from Lemma 2.3.10. Therefore, we can bound (2.81) by

$$\varepsilon^{-2\nu} |t-s|^\delta \varepsilon^{2-2\nu} m^2 \lesssim \varepsilon^{2-4\nu} m^2 |t-s|^\delta \; .$$

We then substitute $\nu = 1/2 - \alpha + \delta$ and ensure δ is small enough so that all the above conditions on ν are satisfied.

We also have that

$$\sum_k \int_0^s |f_k(t-r) - f_k(s-r)|^2 dr = \sum_k \int_0^s |\langle q_k^\varepsilon e_k, R_\varepsilon(t-r) - R_\varepsilon(s-r)\rangle|^2 dr \;. \quad (2.82)$$

If $\alpha \in (1/2, 1)$ then, as in the previous step we can bound the above by a multiple of

$$\int_0^s \|R_\varepsilon(t-r) - R_\varepsilon(s-r)\|^2 dr \lesssim |t-s|^\delta \sup_{t\in[0,T]} \|\partial_t R_\varepsilon(t)\|^\delta \int_0^s \|R_\varepsilon(t-r) - R_\varepsilon(s-r)\|^{2-\delta} dr \;.$$

Using the estimates on R_ε given in Lemma 2.3.10, we can bound this by a constant multiple of

$$\varepsilon^{2-3\delta} m^{2+\delta} |t-s|^\delta \;.$$

If $\alpha \in (0, 1/2]$ on the other hand, we can bound (2.82) by

$$\sum_k (1 \wedge |k|^{-2\nu}) \|q_k\|^2 \varepsilon^{-2\nu} \times \int_0^s \|R_\varepsilon(t-r) - R_\varepsilon(s-r)\|^{2-2\nu} \|R_\varepsilon(t-r) - R_\varepsilon(s-r)\|_{H^1}^{2\nu} \, dr \;.$$

As before, we choose $\nu \in (0, 1/2)$ such that $\alpha + \nu > 1/2$, this guarantees the above sum is bounded. Moreover, we can bound the above integral by

$$|t-s|^\delta \sup_{t\in[0,T]} \|\partial_t R_\varepsilon(t)\|^\delta \int_0^s \|R_\varepsilon(t-r) - R_\varepsilon(s-r)\|^{2-2\nu-\delta} \|R_\varepsilon(t-r) - R_\varepsilon(s-r)\|_{H^1}^{2\nu} dr \;.$$

Using the estimates on R_ε given in Lemma 2.3.10, we can bound this by a constant multiple of

$$\varepsilon^{2-2\nu-3\delta} m^{2-2\nu+\delta} |t-s|^\delta \int_0^T \|R_\varepsilon(r)\|^{2\nu} dr \;.$$

And, by Jensen's inequality, since $2\nu < 1$, we can bound the above by

$$\varepsilon^{2-2\nu-3\delta} m^{2-2\nu+\delta} |t-s|^\delta \left(\int_0^T \|R_\varepsilon(r)\| dr \right)^{2\nu} \lesssim \varepsilon^{2-2\nu-3\delta} m^{2+\delta} |t-s|^\delta \;.$$

We then substitute $\nu = 1/2 - \alpha + \delta$ and ensure δ is small enough so that the above condition on ν are satisfied. Hence, we have that

$$K_\delta(\varepsilon) = \varepsilon^{2-4\nu-3\delta} m^{2+\delta} = \varepsilon^{4\alpha-7\delta} m^{2+\delta} \;,$$

which proves estimate (2.69c). □

We now concentrate on the second convergence theorem, where we assume that the noise satisfies $\langle q_k, \rho \rangle = 0$ for all $k \in \mathbb{Z}$. Before proving the theorem, we give a formal argument to describe how the proof works. It is clear from the proof of the previous theorem that we can formally write

$$\langle u_\varepsilon(t), e_m \rangle = \sum_{l \neq 0} \langle q_{m+l/\varepsilon} e_l, \rho \rangle \int_0^t e^{-\mu m^2 (t-s)} dW_{m+l/\varepsilon}(s) + O(\varepsilon^\theta)$$

for some $\theta > 0$, provided m is not too large. The previous theorem tells us that the first term above will decay with ε to zero. However, with Assumption 2.2.6 in place, we have precise control over how this term tends to zero. In fact, we have that

$$\begin{aligned}\langle \varepsilon^{-\alpha} u_\varepsilon(t), e_m \rangle &= \sum_{l \neq 0} \varepsilon^{-\alpha} \langle q_{m+l/\varepsilon} e_l, \rho \rangle \int_0^t e^{-\mu m^2 (t-s)} dW_{m+l/\varepsilon}(s) + O(\varepsilon^{\theta - \alpha}) \\ &= \sum_{l \neq 0} \varepsilon^{-\alpha} (m + l/\varepsilon)^{-\alpha} \langle (m + l/\varepsilon)^\alpha q_{m+l/\varepsilon} e_l, \rho \rangle \\ &\quad \times \int_0^t e^{-\mu m^2 (t-s)} dW_{m+l/\varepsilon}(s) + O(\varepsilon^{\theta - \alpha}) \, ,\end{aligned}$$

and all the terms in the sum are no longer decaying with ε. Now, since a convergent sum of complex OU processes is a complex OU process, we can find a Brownian motion $\hat{W}_m$ such that the above is equal in distribution to

$$\Lambda_{\varepsilon,m} \int_0^t e^{-\mu m^2 (t-s)} d\hat{W}_m(s) + O(\varepsilon^{\theta - \alpha})$$

where we denote

$$\Lambda_{\varepsilon,m} = \left(\sum_{l \neq 0} \varepsilon^{-2\alpha} |m + l/\varepsilon|^{-2\alpha} |\langle (m + l/\varepsilon)^\alpha q_{m+l/\varepsilon} e_l, \rho \rangle|^2 \right)^{1/2} .$$

If we can justify taking the limit inside the above sum then it is clear that

$$\lim_{\varepsilon \to 0} \Lambda_{\varepsilon,m} = \left(\sum_{l \neq 0} |l|^{-2\alpha} |\langle \bar{q} \rho, e_{-l} \rangle|^2 \right)^{1/2} = \|\bar{q} \rho\|_{-\alpha} \, ,$$

recalling that $|k|^\alpha q_k \to \bar{q}$ in $L^2[0, 2\pi]$. If we can also adjust our estimates on the remainder to ensure that $\theta > \alpha$, so that $\varepsilon^{\theta - \alpha}$ does indeed decay, then formally we have shown that

$\langle u_\varepsilon(t), e_m\rangle$ is equal in distribution to a process that converges to

$$\|\bar{q}\rho\|_{-\alpha} \int_0^t e^{-\mu m^2(t-s)} d\hat{W}_m(s) \;,$$

which is the m-th Fourier mode of the solution to the limiting SPDE (2.60). Of course, there are several caveats with this argument. Most importantly, the Brownian motions $\hat{W}_m$ are defined in such a way that their distribution changes as ε tends to zero and consequently, the limit above does not make sense. The correct way to proceed is actually *backwards*. That is, we fix a sequence of Brownian motions $\hat{W}_m$ that are used to construct the limiting SPDE (2.60). We then construct a sequence of processes $\hat{u}_\varepsilon$ equal in law to u_ε defined in such a way that when we perform the above calculations, the resulting OU process (driven by $\hat{W}_m$) does not depend on ε. This is made rigorous below.

Remark 2.4.15. It is clear from the preceding argument that no stronger type of convergence is possible in the context of Theorem 2.4.3. In particular, we see that the limiting term in $\langle \varepsilon^{-\alpha} u_\varepsilon, e_m\rangle$ is an OU process determined by $\{W_{m+l/\varepsilon}\}$ for each $l \in \mathbb{Z}$. Hence, even when ε is near zero, the contributing BMs are always changing; we will never be able to pin down the limiting process to a fixed location of our probability space so convergence in probability is not possible.

Proof of Theorem 2.4.3. The process $\hat{u}_\varepsilon$ will be defined using two sequences of BMs, namely $\{\hat{W}_m\}_{m\in\mathbb{Z}}$ and $\{B_k^\varepsilon\}_{k\in\mathbb{Z}}$, that live on a different probability space. Given a sequence $\{\hat{W}_m\}_{m\in\mathbb{Z}}$ of i.i.d. complex-valued Wiener processes (modulo the reality condition $\hat{W}_m = \hat{W}_{-m}^\star$), we construct a sequence $\{B_k^\varepsilon\}_{k\in\mathbb{Z}}$ of i.i.d. complex-valued Wiener processes (again modulo the corresponding reality condition) such that $(\hat{W}, B^\varepsilon)$ are jointly Gaussian with the covariance structure given by

$$\mathbb{E}\hat{W}_m(t)B_k^\varepsilon(s) = \begin{cases} \frac{\lambda_{\varepsilon,m}^l}{\Lambda_{\varepsilon,m}}(t\wedge s) & \text{if } k = m + l/\varepsilon \text{ for some } l \in \mathbb{Z}, \\ 0 & \text{otherwise,} \end{cases}$$

where $\lambda_{\varepsilon,m}^l = \varepsilon^{-\alpha}\langle q_{m+l/\varepsilon} e_l, \rho\rangle$. Such a construction is possible due to the fact that $\Lambda_{\varepsilon,m}^2 = \sum_l |\lambda_{\varepsilon,m}^l|^2$ by definition. In the new probability space, one should view the sequence $\{B_k^\varepsilon\}$ as playing the role of the sequence $\{W_k\}$ in the old space. We can now define $\hat{u}_\varepsilon$ by its Fourier coefficients. For $|m| < \varepsilon^{-\beta}$ set

$$\begin{aligned}\langle \hat{u}_\varepsilon(t), e_m\rangle = \varepsilon^\alpha \Lambda_{\varepsilon,m} \int_0^t e^{-\mu m^2(t-s)} d\hat{W}_m(s) + \sum_k \int_0^t \langle q_k^\varepsilon e_k, R_\varepsilon(t-s)\rangle dB_k^\varepsilon(s) \\ + \sum_k \int_0^t \langle q_k^\varepsilon e_k, \hat{\rho}_{(t-s)/\varepsilon^2}^\varepsilon\rangle dB_k^\varepsilon(s) \;.\end{aligned}$$

For $|m| \geq \varepsilon^{-\beta}$ on the other hand, we simply set

$$\langle \hat{u}_\varepsilon(t), e_m \rangle = \langle w_\varepsilon(t), e_m \rangle \ ,$$

where w_ε solves the SPDE (2.19) with $\{W_k\}$ replaced by $\{B_k^\varepsilon\}$. One can verify that $u_\varepsilon \overset{\text{law}}{=} \hat{u}_\varepsilon$ by checking that

$$\mathbb{E}\langle u_\varepsilon(t), e_m \rangle \langle u_\varepsilon(s), e_n \rangle = \mathbb{E}\langle \hat{u}_\varepsilon(t), e_m \rangle \langle \hat{u}_\varepsilon(s), e_n \rangle$$

for all choices of $t, s \in [0, T]$ and $n, m \in \mathbb{Z}$. We define $v(t)$ as the mild solution to SPDE (2.60). In particular, we have that

$$\langle v(t), e_m \rangle = \|\bar{q}\rho\|_{-\alpha} \int_0^t e^{-\mu m^2(t-s)} d\hat{W}_m(s) \ ,$$

for each $m \in \mathbb{Z}$.

We shall now prove that $\varepsilon^{-\alpha}\hat{u}_\varepsilon \to v$ in the required sense. Firstly, we split the problem into high and low modes

$$\begin{aligned} &\mathbb{E}\mathrm{sup}_{t\in[0,T]}\|\varepsilon^{-\alpha}\hat{u}_\varepsilon(t) - v(t)\|_{H^{-s}}^2 \\ &\qquad \lesssim \sum_{|m|<\varepsilon^{-\beta}} \mathbb{E}\mathrm{sup}_{t\in[0,T]}|\langle \varepsilon^{-\alpha}\hat{u}_\varepsilon(t) - v(t), e_m \rangle|^2 (1+m^2)^{-s} \\ &\qquad\quad + \mathbb{E}\mathrm{sup}_{t\in[0,T]} \sum_{|m|\geq\varepsilon^{-\beta}} |\langle \varepsilon^{-\alpha}\hat{u}_\varepsilon(t) - v(t), e_m \rangle|^2 (1+m^2)^{-s} \ . \end{aligned}$$

We can bound the low modes in the following way

$$\begin{aligned} &\mathbb{E}\mathrm{sup}_{t\in[0,T]}|\langle \varepsilon^{-\alpha}\hat{u}_\varepsilon(t) - v(t), e_m \rangle|^2 \\ &\qquad \lesssim \mathbb{E}\mathrm{sup}_{t\in[0,T]}\left|(\Lambda_{\varepsilon,m} - \|\bar{q}\rho\|_{-\alpha}) \int_0^t e^{-\mu m^2(t-s)} d\hat{W}_m(s)\right|^2 \\ &\qquad\quad + \mathbb{E}\mathrm{sup}_{t\in[0,T]}\left|\varepsilon^{-\alpha}\sum_k \int_0^t \langle q_k^\varepsilon e_k, R_\varepsilon(t-s)\rangle dB_k^\varepsilon(s)\right|^2 \\ &\qquad\quad + \mathbb{E}\mathrm{sup}_{t\in[0,T]}\left|\varepsilon^{-\alpha}\sum_l \int_0^t \langle q_{m+l/\varepsilon} e_l, \hat{\rho}_{(t-s)/\varepsilon^2}\rangle dB_k^\varepsilon(s)\right|^2 \ . \end{aligned}$$

However, it is clear that

$$\mathbb{E}\mathrm{sup}_{t\in[0,T]}\left|(\Lambda_{\varepsilon,m} - \|\bar{q}\rho\|_{-\alpha}) \int_0^t e^{-\mu m^2(t-s)} d\hat{W}_m(s)\right|^2 \lesssim |\Lambda_{\varepsilon,m} - \|\bar{q}\rho\|_{-\alpha}|^2 (1 \wedge m^{-2}) \ .$$

And from Lemmas 2.4.13 and 2.4.14 we have that the two estimates

$$\mathbb{E}\sup_{t\in[0,T]}\left|\sum_k \int_0^t \langle q_k^\varepsilon e_k, R_\varepsilon(t-s)\rangle dB_k^\varepsilon(s)\right|^2 \lesssim (\varepsilon^{4\alpha}\vee\varepsilon^2)\varepsilon^{-7\delta}m^{2+\delta}$$

$$\mathbb{E}\sup_{t\in[0,T]}\left|\sum_l \int_0^t \langle q_{m+l/\varepsilon} e_l, \hat{\rho}_{(t-s)/\varepsilon^2}\rangle dB_{m+l/\varepsilon}^\varepsilon(s)\right|^2 \lesssim \varepsilon^{2-\delta}$$

hold for sufficiently small $\delta > 0$. Using these estimates, when $|m| < \varepsilon^{-\beta}$, we have that

$$\begin{aligned}\sum_{|m|<\varepsilon^{-\beta}} &\mathbb{E}\sup_{t\in[0,T]}|\langle \varepsilon^{-\alpha}\hat{u}_\varepsilon(t) - v(t), e_m\rangle|^2(1+m^2)^{-s} \\ &\lesssim \sum_{|m|<\varepsilon^{-\beta}} |\Lambda_{\varepsilon,m} - \|\bar{q}\rho\|_{-\alpha}|^2(1\wedge m^{-(2+2s)}) \\ &\quad + \varepsilon^{2-2\alpha-\delta}\sum_{|m|<\varepsilon^{-\beta}}(1\wedge m^{-2s}) + (\varepsilon^{2\alpha}\vee\varepsilon^{2-2\alpha})\varepsilon^{-7\delta}\sum_{|m|<\varepsilon^{-\beta}}(1\wedge m^{2-2s+\delta})\,.\end{aligned} \tag{2.83}$$

Firstly, we would like to show that the first sum in the expression above tends to zero as $\varepsilon\to 0$, by taking the limit inside the sum over m. Now, since $\|q_k\| \lesssim 1\wedge|k|^{-\alpha}$ for each $k\in\mathbb{Z}$, we have that

$$\begin{aligned}\Lambda_{\varepsilon,m}^2 = \varepsilon^{-2\alpha}\sum_{l\neq 0}|\langle q_{m+l/\varepsilon}\rho, e_{-l}\rangle|^2 &\lesssim \sum_{l\neq 0}|\varepsilon m + l|^{-2\alpha}\langle \bar{q}_{m+l/\varepsilon}\rho, e_{-l}\rangle^2 \\ \lesssim \sum_{l\neq 0}|\varepsilon m + l|^{-2\alpha}|l|^{-2\nu}\|\bar{q}_{m+l/\varepsilon}\|_{H^1}^{2\nu}\,,\end{aligned}$$

where the last inequality follows from Lemma 2.4.11 and the smoothness of ρ. If $\alpha\in(1/2,1)$, then set $\nu = 0$, if $\alpha\in(0,1/2]$, then set $\nu = 1$. In either case, the above sum is bounded uniformly in ε and m, as long as $|m| < \varepsilon^{-1}/2$. For $|m| < \varepsilon^{-\beta}$, we therefore have that

$$\begin{aligned}\lim_{\varepsilon\to 0}\varepsilon^{-2\alpha}\sum_{l\neq 0}\langle q_{m+l/\varepsilon}\rho, e_{-l}\rangle^2 &= \sum_{l\neq 0}\lim_{\varepsilon\to 0}\varepsilon^{-2\alpha}\langle q_{m+l/\varepsilon}\rho, e_{-l}\rangle^2 \\ &= \sum_{l\neq 0}\lim_{\varepsilon\to 0}\varepsilon^{-2\alpha}|m+l/\varepsilon|^{-2\alpha}\langle |m+l/\varepsilon|^{\alpha}q_{m+l/\varepsilon}\rho, e_{-l}\rangle^2 \\ &= \sum_{l\neq 0}|l|^{-2\alpha}\langle\bar{q}\rho, e_{-l}\rangle^2 = \|\bar{q}\rho\|_{-\alpha}^2\,.\end{aligned}$$

For the first sum in (2.83), it is now clear that if $s > 0$ then

$$\sum_{|m|<\varepsilon^{-\beta}}(\Lambda_{\varepsilon,m} - \|\bar{q}\rho\|_{-\alpha})^2(1\wedge m^{-(2+2s)}) \lesssim \sum_m(1\wedge m^{-(2+2s)})\,,$$

and is therefore bounded uniformly in ε. Hence, we have that

$$\lim_{\varepsilon\to 0}\sum_{|m|<\varepsilon^{-\beta}}(\Lambda_{\varepsilon,m}-\|\bar{q}\rho\|_{-\alpha})^2(1\wedge m^{-(2+2s)})$$
$$=\sum_{|m|<\varepsilon^{-\beta}}\lim_{\varepsilon\to 0}(\Lambda_{\varepsilon,m}-\|\bar{q}\rho\|_{-\alpha})^2(1\wedge m^{-(2+2s)})=0\ .$$

For the second sum in (2.83), we have that

$$\varepsilon^{2-2\alpha-\delta}\sum_{|m|<\varepsilon^{-\beta}}(1\wedge m^{-2s})\lesssim\varepsilon^{2-2\alpha-\delta}(1\vee\varepsilon^{-(1-2s)\beta})\ .$$

For the third sum in (2.83), we have that

$$(\varepsilon^{2\alpha}\vee\varepsilon^{2-2\alpha})\varepsilon^{-7\delta}\sum_{|m|<\varepsilon^{-\beta}}(1\wedge m^{2-2s+\delta})$$
$$\lesssim(\varepsilon^{2\alpha}\vee\varepsilon^{2-2\alpha})\varepsilon^{-7\delta}(1\vee\varepsilon^{-(3-2s+\delta)\beta})\ ,$$

provided $s>0$. For the high modes, we have that

$$\begin{aligned}
&\mathbb{E}\mathrm{sup}_{t\in[0,T]}\sum_{|m|\geq\varepsilon^{-\beta}}|\langle\varepsilon^{-\alpha}\hat{u}_\varepsilon(t)-v(t),e_m\rangle|^2(1+m^2)^{-s}\\
&\lesssim\mathbb{E}\mathrm{sup}_{t\in[0,T]}\sum_{m\in\mathbb{Z}}|\langle\hat{u}_\varepsilon(t),e_m\rangle|^2\varepsilon^{2\beta s-2\alpha}+\mathbb{E}\mathrm{sup}_{t\in[0,T]}\sum_{m\in\mathbb{Z}}|\langle v(t),e_m\rangle|^2\varepsilon^{2\beta s}\\
&\lesssim\varepsilon^{2\beta s-2\alpha}\mathbb{E}\mathrm{sup}_{t\in[0,T]}\|\hat{u}_\varepsilon(t)\|^2+\varepsilon^{2\beta s}\mathbb{E}\mathrm{sup}_{t\in[0,T]}\|v(t)\|^2\\
&\lesssim\varepsilon^{2\beta s-2\alpha}(1\vee\varepsilon^{4\alpha-2-\delta})+\varepsilon^{2\beta s}\ .
\end{aligned}$$

Here we have used Lemma 2.4.9 as well as the clear fact that

$$\mathbb{E}\mathrm{sup}_{t\in[0,T]}\|v(t)\|^2\lesssim 1\ .$$

If $\alpha\in(1/2,1)$, then for both the low and high modes to converge to zero for some $s>0$, we need to find $\beta\in(0,1)$ such that

$$\frac{\alpha}{s}<\beta<\frac{2-2\alpha}{3-2s}\ .$$

A simple diagram confirms that we can always find such a β provided $s>\frac{3}{2}\alpha$. If $\alpha\in(0,1/2]$, then for both the low and high modes to converge to zero for some $s>0$, we need to find $\beta\in(0,1)$ such that

$$\frac{1-\alpha}{s}<\beta<\frac{2\alpha}{3-2s}\ .$$

A simple diagram confirms that we can always find such a β, provided $s > \frac{3}{2}(1-\alpha)$. This concludes the proof of the theorem. $\square$

Before proving Theorem 2.4.4, we need a new a priori bound on the solution u_ε, given that we are working with new assumptions on the noise.

Lemma 2.4.16. *Suppose u_ε satisfies* (2.19) *and the conditions given in Assumptions 2.2.2, 2.2.7 hold true, then we have that*

$$\mathbb{E}\sup_{t\in[0,T]}\|u_\varepsilon(t)\|^2 \lesssim \varepsilon^{-2-\delta} , \tag{2.84}$$

for arbitrarily small $\delta > 0$.

Proof. From Lemma 2.4.9 we know that

$$\mathbb{E}\sup_{t\in[0,T]}\|u_\varepsilon(t)\|^2 \lesssim \int_0^T \|S_\varepsilon(t)Q_\varepsilon\|_{\mathrm{HS}}^2 \, dt .$$

We can bound the Hilbert-Schmidt norm using Assumption 2.2.7. We have that

$$\|S_\varepsilon(t)Q_\varepsilon\|_{\mathrm{HS}}^2 = \sum_k \|S_\varepsilon(t)q_k^\varepsilon e_k\|^2 \lesssim \sum_k \|S_\varepsilon(t)(q_k^\varepsilon - \bar{q}^\varepsilon)e_k\|^2 + \sum_k \|S_\varepsilon(t)\bar{q}^\varepsilon e_k\|^2 .$$

But the first term can be bounded

$$\sum_k \|S_\varepsilon(t)(q_k^\varepsilon - \bar{q}^\varepsilon)e_k\|^2 \lesssim \varepsilon^{-4\nu}|t|^{-\nu} \sum_k |k|^{-2\nu}\|q_k - \bar{q}\|_{H^1}^2 ,$$

for any $\nu \in [0,1)$ using the same argument found in Lemma 2.4.9. By assumption, the sum over k is finite if we set $2\nu = \eta$. For the second term, we similarly know that

$$\sum_k \|S_\varepsilon(t)\bar{q}^\varepsilon e_k\|^2 \lesssim \varepsilon^{-4\gamma}|t|^{-\gamma}\|\bar{q}\|_{H^1} \sum_k |k|^{-2\gamma} ,$$

for any $\gamma \in (0,1)$. If we set $\gamma = 1/2 + \delta/4$, for arbitrarily small $\delta > 0$, then the sum over k will converge. Since $2\eta < 2$, the $\varepsilon^{-4\gamma}$ term will be the dominant one. It follows that

$$\int_0^T \|S_\varepsilon(t)Q_\varepsilon\|_{\mathrm{HS}} dt \lesssim \varepsilon^{-4\gamma} = \varepsilon^{-2-\delta} .$$

This proves the lemma. $\square$

Proof of Theorem 2.4.4. As in the proof of Theorem 2.4.3, we construct sequences $\{\hat{W}_m\}$

and $\{B_k^\varepsilon\}$ of Brownian motions with correlations

$$\mathbb{E}\hat{W}_m(t)B_k^\varepsilon(s) = \begin{cases} \frac{\lambda_{\varepsilon,m}^l}{\Lambda_{\varepsilon,m}}(t \wedge s), & \text{if } k = m + l/\varepsilon \text{ for some } l \in \mathbb{Z} , \\ 0, & \text{otherwise} , \end{cases} \tag{2.85}$$

where $\lambda_{\varepsilon,m}^l = \langle q_{m+l/\varepsilon} e_l, \rho \rangle$ and, as before, $\Lambda_{\varepsilon,m} = \left(\sum_{l\in\mathbb{Z}} |\lambda_{\varepsilon,m}^l|^2\right)^{1/2}$. We then define $\hat{u}_\varepsilon$ through its Fourier modes as follows For $|m| \le \varepsilon^{-\beta}$, we set

$$\langle \hat{u}_\varepsilon(t), e_m \rangle = \Lambda_{\varepsilon,m} \int_0^t e^{-\mu m^2(t-s)} d\hat{W}_m(s) ,$$

while for $|m| > \varepsilon^{-\beta}$, we set

$$\langle \hat{u}_\varepsilon(t), e_m \rangle = \langle w_\varepsilon(t), e_m \rangle ,$$

where w_ε solves (2.19) with W_k replaced with B_k^ε for each $k \in \mathbb{Z}$. This is identical to the construction given in the proof of Theorem 2.4.3, with the sole difference being that now $\lambda_{\varepsilon,m}^0 \neq 0$, in general. The proof proceeds identically to the previous theorem. We only need a few more ingredients to ensure that this proof will work just like the last. First, we need that

$$\Lambda_{\varepsilon,m} - (|\langle q_m, \rho \rangle|^2 - |\langle \bar{q}, \rho \rangle|^2 + \|\bar{q}\rho\|^2)^{1/2}$$

converges to zero as $\varepsilon \to 0$. But this is true by construction of the series $\Lambda_{\varepsilon,m}$, using the same arguments as previously employed to pass the limit inside the sum. Secondly, we need some bound on the remainder terms of the low modes. We cannot use the previous bounds (2.69b) and (2.69c), since we are effectively using $\alpha = 0$. However, just as in Lemma 2.4.16 we can use Assumption 2.2.7 instead. We claim the following bounds to be true and prove them in the sequel. For $|m| \le \varepsilon^{-\beta}$, we have that

$$\mathbb{E}\mathrm{sup}_{t\in[0,T]} \left| \sum_k \int_0^t \langle q_k^\varepsilon e_k, \hat{\rho}_{(t-s)/\varepsilon^2}^\varepsilon e_m \rangle dB_k^\varepsilon(s) \right|^2 \lesssim \varepsilon^{2-\eta-2\delta} |m|^\eta , \tag{2.86}$$

$$\mathbb{E}\mathrm{sup}_{t\in[0,T]} \left| \sum_k \int_0^t \langle q_k^\varepsilon e_k, R_\varepsilon(t-s) \rangle dB_k^\varepsilon(s) \right|^2 \lesssim \varepsilon^{2-2\eta-3\delta} |m|^{2+\delta} , \tag{2.87}$$

for arbitrarily small $\delta > 0$. From Lemma 2.4.16 we have that

$$\mathbb{E}\mathrm{sup}_{t\in[0,T]} \|\hat{u}_\varepsilon(t)\|^2 \lesssim \varepsilon^{-2-\delta} .$$

Moreover, one can easily show that

$$\mathbb{E}\sup_{t\in[0,T]}\|\hat{u}(t)\|^2 \le C_T \ .$$

We can then apply the exact arguments used in Theorem 2.4.3 to show that both high and low modes will converge to zero as $\varepsilon \to 0$ if we can choose $\beta \in (0,1)$ in such a way that

$$\frac{1}{s} < \beta < \frac{2-2\eta}{3-2s} \ .$$

It is easy to show that one can always find such a β provided $s > s_\eta$, where

$$s_\eta = 1 \vee \frac{3}{2(2-\eta)} \ .$$

This proves (2.61). We now prove the claimed bounds. For (2.86) and (2.87) we apply the Kolmogorov criterion from Lemma 2.4.10 just as we did to bound (2.69b) and (2.69c) respectively. This involves proving four estimates (two for each claim). For the first claim, we wish to find $K_\varepsilon(\delta)$ such that

$$\mathbb{E}\left|\sum_k \int_s^t \langle q_k^\varepsilon e_k, \hat{\rho}^\varepsilon_{(t-r)/\varepsilon^2} e_m\rangle dB_k^\varepsilon(r)\right|^2 \le K_\varepsilon(\delta)|t-s|^\delta \ , \tag{2.88}$$

and

$$\mathbb{E}\left|\sum_k \int_0^t \langle q_k^\varepsilon e_k, (\hat{\rho}^\varepsilon_{(t-r)/\varepsilon^2} - \hat{\rho}^\varepsilon_{(t-s)/\varepsilon^2}) e_m\rangle dB_k^\varepsilon(r)\right|^2 \le K_\varepsilon(\delta)|t-s|^\delta \ , \tag{2.89}$$

for some $\delta \in (0,1)$. Clearly, we can bound the left hand side of (2.88) by a constant multiple of

$$\sum_k \int_s^t |\langle (q_k^\varepsilon - \bar{q}^\varepsilon) e_k, \hat{\rho}^\varepsilon_{(t-r)/\varepsilon^2} e_m\rangle|^2 dr + \sum_k \int_s^t |\langle \bar{q}^\varepsilon e_k, \hat{\rho}^\varepsilon_{(t-r)/\varepsilon^2} e_m\rangle|^2 dr \ .$$

Applying Lemma 2.4.11 (with $2\nu = \eta$) to the first term and using the fact that, for every m, one has $\sum_k |\langle e_k e_{-m}, f\rangle|^2 = \|f\|^2$ for the second term, we can bound this by

$$\varepsilon^{-\eta}\left(\sum_k |m-k|^{-\eta}\|q_k - \bar{q}\|_{H^1}^2\right)\int_s^t \|\hat{\rho}_{(t-r)/\varepsilon^2}\|_{H^1}^\eta \|\hat{\rho}_{(t-r)/\varepsilon^2}\|^{2-\eta} dr$$
$$+ \int_s^t \|\bar{q}^\varepsilon \hat{\rho}^\varepsilon_{(t-r)/\varepsilon^2}\|^2 dr \ ,$$

$$\lesssim \varepsilon^{-\eta}|m|^\eta \left(\sum_k |k|^{-\eta}\|q_k - \bar q\|^2_{H^1}\right) \int_s^t \|\hat\rho_{(t-r)/\varepsilon^2}\|^\eta_{H^1}\|\hat\rho_{(t-r)/\varepsilon^2}\|^{2-\eta} dr$$
$$+ \|\bar q\|_\infty \int_s^t \|\hat\rho_{(t-r)/\varepsilon^2}\|^2 dr\ .$$

By Assumption 2.2.7 the sum over k is finite and, by a Sobolev embedding, $\|\bar q\|_\infty$ is also finite. The integral terms can be bounded exactly as in the proof of estimate (2.69b) to obtain $K_\varepsilon(\delta) = \varepsilon^{2-\eta-2\delta}|m|^\eta$. We then treat (2.89), and also the two respective estimates required to prove (2.87) in the same way, by first splitting q_k into $(q_k - \bar q) + \bar q$ and then applying the results from the proof of (2.69b) and (2.69c). The estimates (2.86) and (2.87) follow. □

Proof of Corollary 2.4.6. The proof follows in the same way as that of Theorem 2.4.4, except we now have $\lambda^l_{\varepsilon,m} = \varphi(\varepsilon m + l)\langle q_{m+l/\varepsilon} e_l, \rho\rangle$. Moreover, we now need to show that

$$\Lambda_{\varepsilon,m} - (|\langle q_m, \rho\rangle|^2 - |\langle \bar q, \rho\rangle|^2 + \|(\bar q\rho) \star \tilde\varphi\|^2)^{1/2} \tag{2.90}$$

converges to zero as $\varepsilon \to 0$, where $\Lambda_{\varepsilon,m}$ is defined as above, using the new $\lambda^l_{\varepsilon,m}$. But it is clear that

$$\Lambda^2_{\varepsilon,m} = |\varphi(\varepsilon m)|^2|\langle q_m, \rho\rangle|^2 + \sum_{l\neq 0} |\varphi(\varepsilon m + l)|^2|\langle q_{m+l/\varepsilon} e_l, \rho\rangle|^2$$
$$\to |\langle q_m, \rho\rangle|^2 + \sum_{l\neq 0} |\varphi(l)|^2|\langle \bar q\rho, e_l\rangle|^2\ ,$$

where the boundedness of φ in combination with previous arguments allows us to take the limit inside the sum over l. Since $\|(\bar q\rho) \star \tilde\varphi\|^2 = \sum_{l\in\mathbb{Z}} |\varphi(l)|^2|\langle \bar q\rho, e_l\rangle|^2$, we have proven (2.90). The remainder of the proof follows in exactly the same way as Theorem 2.4.4, and since φ is bounded, all corresponding estimates still hold. □

Chapter 3

Non-geometric rough paths

3.1 Introduction

The theory of rough paths suggests that one can understand the dynamics of a stochastic differential equation (SDE) by studying the dynamics on a Lie group lying above the state space of the SDE. This is indeed the geometric description offered in the groundbreaking work [Lyo98] and in the standard texts [FV10b, LV07]. For example, let X be a γ-Hölder continuous path taking values in some Banach space V and consider the SDE

$$dY_t = f(Y_t)dX_t \,, \tag{3.1}$$

where $f : U \to L(V, U)$, with $t \in [0, T]$ and U is another Banach space. In this thesis we will only consider the case where U, V are finite dimensional, hence we can essentially consider $U = \mathbb{R}^e$ and $V = \mathbb{R}^d$. In the framework of [Lyo98], one can solve (3.1) using what is called a *geometric rough path*. The aim of this chapter is to review a technique which uses *non-geometric rough paths*, first introduced in [Gub10]. Before understanding this technique, we first present the idea behind the geometric approach.

To solve (3.1) as a rough differential equation (RDE), the first step is to build a rough path $\mathbf{X}$ over X. Formally speaking, $\mathbf{X}$ is a path taking values in a free nilpotent group built over V, that is a *lift* of X, in the sense that $\pi_V(\mathbf{X}) = X$. A good example is when $X = B$, a Brownian motion taking values in $\mathbb{R}^2$. In this case to construct the lift $\mathbf{B}$ of B, all we need is the object

$$\int_0^t \int_0^{v_1} dB^j_{v_2} dB^i_{v_1} \,, \tag{3.2}$$

for each $i, j \in \{1, 2\}$ and $t \in [0, T]$. A rough path $\mathbf{B}$ is defined as a path taking values in $\mathbb{R}^2 \oplus (\mathbb{R}^2 \otimes \mathbb{R}^2)$ and satisfying

$$\langle \mathbf{B}_t, e_i \rangle = B^i_t \quad \text{and} \quad \langle \mathbf{B}_t, e_j \otimes e_i \rangle = \int_0^t \int_0^{v_1} dB^j_{v_2} dB^i_{v_1} \,,$$

where e_i is the i-th canonical basis vector in $\mathbb{R}^2$. However, the integral in (3.2) cannot be defined in the usual Lebesgue-Stiltjes sense, since Brownian motion is almost surely a path of infinite total variation. A stochastic theory of integration could be used to define (3.2), however the construction is not "well-posed". In particular, two equally valid numerical constructions, Itô and Stratonovich, result in two different definitions for (3.2). In fact, we could essentially choose for (3.2) any path taking values in $\mathbb{R}^2 \otimes \mathbb{R}^2$. However, there is a *canonical* choice for (3.2), corresponding to Stratonovich integration, and resulting in what is known as a *geometric* rough path. There are several reasons justifying this choice of $\mathbf{B}$. Firstly, it is stable under approximations, in that for any reasonable smooth approximation $B(\varepsilon)$ of B, the well-defined integral $\int\int dB(\varepsilon) \otimes dB(\varepsilon)$ converges to (3.2) under the appropriate metric [FV10a, FV10b]. This would certainly not be the case if (3.2) were defined under an Itô scheme. Secondly, in a more general setting, the algebraic structure of a geometric rough path seems to be more suitable for solving equations like (3.1) in the setting of [Lyo98].

This choice of rough path has a quite beautiful geometric interpretation, namely $\mathbf{B}$ is a path taking values in the *Heisenberg group* $(G, \otimes) \subset \mathbb{R}^2 \oplus (\mathbb{R}^2 \otimes \mathbb{R}^2)$, where the product $\otimes$ is simply the (truncated) tensor product. This group valued property turns out to be the key ingredient in defining geometric rough paths in an arbitrary setting. Returning now to the general case of (3.1), suppose N is the largest integer such that $N\gamma \leq 1$. A *(weak) geometric rough path* $\mathbf{X}$ over X is a path taking values in $G^{(N)}(V)$, the step N free nilpotent group over V. The group $(G^{(N)}(V), \otimes)$ is a subset of the truncated tensor product algebra $T^{(N)}(V) = V \oplus V^{\otimes 2} \oplus \cdots \oplus V^{\otimes N}$ and is a Lie group, described by

$$G^{(N)}(V) = \exp \mathcal{G}^{(N)}(V) \ ,$$

where $\mathcal{G}^{(N)}(V)$ is the step-N free Lie algebra over V and $\exp$ is the tensor exponential. For the path X, *evolution* from s to t is described by the increment $\delta X_{st} \stackrel{\text{def}}{=} X_t - X_s$. For the rough path $\mathbf{X}$, it is described by the more complicated increment

$$\mathbf{X}_{st} \stackrel{\text{def}}{=} \mathbf{X}_s^{-1} \otimes \mathbf{X}_t \ , \tag{3.3}$$

where the inverse denotes the group inverse. In particular, this allows us to define

$$\int_s^t \int_s^{v_1} \cdots \int_s^{v_{n-1}} dX_{v_n}^{i_n} \dots dX_{v_2}^{i_2} dX_{v_1}^{i_1} \stackrel{\text{def}}{=} \langle \mathbf{X}_{st}, e_{i_n} \otimes \cdots \otimes e_{i_2} \otimes e_{i_1} \rangle \ .$$

When solving RDEs, it is often more convenient to work with the increment $\mathbf{X}_{st}$ rather than the original path element. The group $G^{(N)}(V)$ can equally be defined as the step-N truncation of *group-like elements* of $T(V)$ or the *characters* with respect to the shuffle product. This ensures that

$$\langle \mathbf{X}_{st}, e_{i_1 \dots i_n} \rangle \langle \mathbf{X}_{st}, e_{j_1 \dots j_m} \rangle = \langle \mathbf{X}_{st}, e_{i_1 \dots i_n} ⧢ e_{j_1 \dots j_m} \rangle \ , \tag{3.4}$$

where $⧢$ denotes the *shuffle product* on $T(V)$ and we use the shorthand $e_{i_1 \dots i_n} = e_{i_1} \otimes \cdots \otimes e_{i_n}$ for the basis tensors of $T(V)$. The identity (3.4) is *only* satisfied by a geometric choice of rough path. One can think of (3.4) as a condition that ensures the “usual calculus” applies to $\mathbf{X}$, in particular, the statement itself is a generalised integration by parts formula. These algebraic concepts will be made precise in Section 4.2.

To solve the RDE (3.1), we adopt the approach pioneered in [Gub04]; the key observation is that Y is *locally controlled* by the rough path $\mathbf{X}$. We will illustrate this by assuming that $1/4 < \gamma \leq 1/3$, so that $N = 3$. As usual, we assume that $V = \mathbb{R}^d$, $U = \mathbb{R}^e$ and that $f(Y)dX = \sum_{i=1}^d f_i(Y)dX^i$, where the vector fields $f_i : \mathbb{R}^e \to \mathbb{R}^e$ are smooth. We will denote by f_i^α the α-th coordinate of the vector field f_i. Then (3.1) can be written in the integral form

$$Y_t - Y_s = \int_s^t f_i(Y_v) dX_v^i \,, \tag{3.5}$$

where we omit the sum notation. If we perform a Taylor expansion of f_i around Y_s and repeatedly substitute (3.5) back in to itself, then we formally obtain

$$\begin{aligned} Y_t - Y_s \approx{}& f_i(Y_s) \int_s^t dX_{v_1}^i + f_j^{\alpha_1}(Y_s)\partial^{\alpha_1} f_i(Y_s) \int_s^t \int_s^{v_1} dX_{v_2}^j dX_{v_1}^i \\ &+ f_k^{\alpha_1}(Y_s)\partial^{\alpha_1} f_j^{\alpha_2}(Y_s)\partial^{\alpha_2} f_i(Y_s) \int_s^t \int_s^{v_2} \int_s^{v_1} dX_{v_3}^k dX_{v_2}^j dX_{v_1}^i \\ &+ \frac{1}{2} f_k^{\alpha_1}(Y_s) f_j^{\alpha_2}(Y_s)\partial^{\alpha_1\alpha_2} f_i(Y_s) \int_s^t \left(\int_s^{v_3} dX_{v_1}^k\right)\left(\int_s^{v_3} dX_{v_2}^j\right) dX_{v_3}^i \,, \end{aligned} \tag{3.6}$$

where the error is of order $|t-s|^{4\gamma}$ and hence $o(|t-s|)$ for $|t-s| \ll 1$. Now, all of the above integrals are components of $\mathbf{X}_{st}$. For instance,

$$\int_s^t dX_{v_1}^i = \langle \mathbf{X}_{st}, e_i \rangle \quad , \quad \int_s^t \int_s^{v_1} dX_{v_2}^j dX_{v_1}^i = \langle \mathbf{X}_{st}, e_{ji} \rangle$$
$$\int_s^t \int_s^{v_2} \int_s^{v_1} dX_{v_3}^k dX_{v_2}^j dX_{v_1}^i = \langle \mathbf{X}_{st}, e_{kji} \rangle \,,$$

where we use the shorthand $e_{ji} = e_j \otimes e_i$ and $e_{kji} = e_k \otimes e_j \otimes e_i$. The non-trivial term must be understood using the shuffle product. Indeed, the identity (3.4) guarantees that

$$\begin{aligned} \left(\int_s^{v_3} dX_{v_1}^k\right)\left(\int_s^{v_3} dX_{v_2}^j\right) &\stackrel{\text{def}}{=} \langle \mathbf{X}_{st}, e_k \rangle \langle \mathbf{X}_{st}, e_j \rangle = \langle \mathbf{X}_{st}, e_i \sqcup\!\sqcup e_j \rangle \\ &= \langle \mathbf{X}_{st}, e_{kj} \rangle + \langle \mathbf{X}_{st}, e_{jk} \rangle \,, \end{aligned}$$

and hence we define

$$\int_s^t \left(\int_s^{v_3} dX_{v_1}^k\right)\left(\int_s^{v_3} dX_{v_2}^j\right) dX_{v_3}^i \stackrel{\text{def}}{=} \langle \mathbf{X}_{st}, e_{kji} \rangle + \langle \mathbf{X}_{st}, e_{jki} \rangle \,. \tag{3.7}$$

It should then be clear that Y looks locally like $\mathbf{X}$, in the sense that

$$Y_t - Y_s \approx \sum_{e_{i_1 \dots i_n}} F_{e_{i_1 \dots i_n}}(Y_s) \langle \mathbf{X}_{st}, e_v \rangle \,,$$

where we sum over all basis elements $e_{i_1 \dots i_n} \in T^{(N)}(V)$ and where $F_{e_{i_1 \dots i_n}} : \mathbb{R}^e \to \mathbb{R}^e$ are the coefficients from (3.6). One then constructs Y over all of $[0, T]$ by *sewing together* the increments $Y_t - Y_s$ over small intervals.

In certain situations, the geometric framework is not an appropriate model for a stochastic system. For example, in some financial models, an Itô type integral is more appropriate than Stratonovich, since the latter scheme requires one to "look in to the future". Furthermore, it is usually the case that natural approximations to stochastic integrals *do not* converge to objects for which the usual change of variables formula holds. The most famous example of this is of course the Itô integral, however the phenomenon has also spread to the world of non semi-martingales [BM96, ER00, GNRV05, BS10]. Thus, the limiting objects from discretisation schemes are often *non*-geometric. However, as soon as $\gamma \le 1/3$, the geometric assumption is crucial in solving (3.1) in the setting of [Lyo98]. In particular, without (3.4), we would have no way of defining (3.7) and hence would have no local description of Y. In [Gub10], a new approach was proposed, allowing one to solve (3.1) using a non-geometric rough path. The basic idea is to add (3.7) as a separate component of the rough path, defined independently of the other components. In the geometric setting, the rough path $\mathbf{X}$ was an object indexed by tensors; in a non-geometric setting however, a rough path is indexed by *trees*. Hence, these non-geometric rough paths are known as *branched rough paths*.

Whereas a geometric rough path lives in a tensor product algebra generated by $\mathbb{R}^d$, a branched rough path lives in a *Hopf algebra*, generated by the set of rooted, labelled trees $\mathcal{T}$ with vertex decorations from the set $\{1, \dots, d\}$. This space is known as the *Connes-Kreimer Hopf algebra* and was famously used in [CK98] in the context of renormalization theory. This Hopf algebra consists of a vector space $\mathcal{H}$, equipped with a product $\cdot$ and a coproduct Δ. The product $\cdot$ is the commutative polynomial product and the basis elements for the vector space $\mathcal{H}$ are simply the monomials in the variables $\mathcal{T}$, under the polynomial product. We will frequently omit the product $\cdot$ from the notation, for instance writing xy for the polynomial product of x and y. The coproduct Δ is the dual of a more interesting product $\star$ known as *convolution*. Much like the deconcatenation coproduct describes all ways of cutting apart a tensor, the coproduct Δ describes all ways of cutting apart a tree. The standard text on Hopf algebra theory is [Swe69], for an introduction aimed more towards the Connes-Kreimer case, see the monograph [Man04].

A *branched rough path of roughness* γ is defined in [Gub10] as a map
$\mathbf{X} : [0, T] \times [0, T] \to \mathcal{H}^*$ satisfying the following three conditions

1. $\langle \mathbf{X}_{st}, xy \rangle = \langle \mathbf{X}_{st}, x \rangle \langle \mathbf{X}_{st}, y \rangle$, for every $x, y \in \mathcal{H}$.
2. $\mathbf{X}_{st} = \mathbf{X}_{su} \star \mathbf{X}_{ut}$ or equivalently $\langle \mathbf{X}_{st}, h \rangle = \sum_{(h)} \langle \mathbf{X}_{su}, h^{(1)} \rangle \langle \mathbf{X}_{ut}, h^{(2)} \rangle$,

where $\Delta h = \sum_{(h)} h^{(1)} \tilde{\otimes} h^{(2)}$.

3. $\sup_{s\neq t} |\langle \mathbf{X}_{st}, \tau \rangle| / |t-s|^{\gamma|\tau|} < \infty$,
   for every $\tau \in \mathcal{H}$ with $|\tau| \leq N$, where $|\tau|$ counts the number of vertices in τ.

Condition 1 confirms that the polynomial product plays the role of the shuffle product in $\mathcal{H}$. That is, it picks out some object xy so that $\langle \mathbf{X}, xy\rangle = \langle \mathbf{X}, x\rangle\langle \mathbf{X}, y\rangle$. Condition 2 is a natural requirement of any iterated integral. Indeed, no matter how one defines an integral, it should always be linear with respect to the integrand, and satisfy $\int_s^t = \int_s^u + \int_u^t$. These two assumptions imply an algebraic identity known as Chen's property [Che77]. For a geometric rough path, Chen's property takes the form

$$\mathbf{X}_{st} = \mathbf{X}_{su} \otimes \mathbf{X}_{ut} \ ,$$

whereas, for a branched rough path, Chen's property takes the form of Condition 2. Condition 3 reflects the fact that the integral $\langle \mathbf{X}_{st}, \tau\rangle$ should be $|\tau|$ times as regular as the underlying path X; it is a purely analytic condition, as opposed to the first two purely algebraic conditions.

We will now illustrate the definition with the $1/4 < \gamma \leq 1/3$ example. Here, we would have

$$\int_s^t dX^i_{v_1} = \langle \mathbf{X}_{st}, \bullet_i \rangle \ , \ \int_s^t \int_s^{v_1} dX^j_{v_2} dX^i_{v_1} = \langle \mathbf{X}_{st}, \begin{smallmatrix}\bullet_j\\ \bullet_i\end{smallmatrix} \rangle$$

$$\text{and} \quad \int_s^t \int_s^{v_1} \int_s^{v_2} dX^k_{v_3} dX^j_{v_2} dX^i_{v_1} = \langle \mathbf{X}_{st}, \begin{smallmatrix}\bullet_k\\ \bullet_j\\ \bullet_i\end{smallmatrix} \rangle \ ,$$

as well as the *branched* object

$$\int_s^t \left(\int_s^{v_3} dX^k_{v_1}\right)\left(\int_s^{v_3} dX^j_{v_2}\right) dX^i_{v_3} = \langle \mathbf{X}_{st}, \begin{smallmatrix}\bullet_k & & \bullet_j\\ & \bullet_i & \end{smallmatrix} \rangle \ .$$

In this example, the only additional objects in a non-geometric rough path are the components corresponding to $\begin{smallmatrix}\bullet & & \bullet\\ & \bullet & \end{smallmatrix}$. However, as N increases (or γ decreases), a branched rough path becomes much larger than a geometric rough path. For $\tau = \begin{smallmatrix}\bullet_j\\ \bullet_i\end{smallmatrix}$, Condition (3) becomes the familiar identity for the Levy area

$$\langle \mathbf{X}_{st}, \begin{smallmatrix}\bullet_j\\ \bullet_i\end{smallmatrix} \rangle = \langle \mathbf{X}_{su}, \begin{smallmatrix}\bullet_j\\ \bullet_i\end{smallmatrix} \rangle + \langle \mathbf{X}_{ut}, \begin{smallmatrix}\bullet_j\\ \bullet_i\end{smallmatrix} \rangle + \langle \mathbf{X}_{su}, \bullet_j \rangle \langle \mathbf{X}_{ut}, \bullet_i \rangle \ ,$$

or in the language of the coproduct

$$\Delta \begin{smallmatrix}\bullet_j\\ \bullet_i\end{smallmatrix} = \begin{smallmatrix}\bullet_j\\ \bullet_i\end{smallmatrix} \tilde{\otimes} 1 + 1 \tilde{\otimes} \begin{smallmatrix}\bullet_j\\ \bullet_i\end{smallmatrix} + \bullet_j \tilde{\otimes} \bullet_i \ .$$

Let us again consider the solution to (3.1), now driven by a branched rough path $\mathbf{X}$ with $1/4 < \gamma \leq 1/3$. From (3.6), we would have

$$Y_t - Y_s \approx \sum_{\tau} f_\tau(Y_s)\langle \mathbf{X}_{st}, \tau\rangle \tag{3.8}$$

where we sum over all $\tau \in \mathcal{T}_3$, or in the case of arbitrary γ, all $\tau \in \mathcal{T}_N$, the set of $\tau \in \mathcal{T}$ with $|\tau| \leq N$. Hence, the idea of viewing the solution to (3.1) as an object that locally "looks like" $\mathbf{X}$ carries through nicely to the framework of non-geometric rough paths. The coefficients f_τ are known as the *Butcher coefficients*, in honour of J. Butcher who was the first to represent solutions to ODEs as a series indexed by trees [But72, HW74].

After defining branched rough paths rigorously, the next objective of this chapter is to cast branched rough paths in a similar light to geometric rough paths. For a geometric rough path, Chen's property is not a definition, but is a corollary from the definition $\mathbf{X}_{st} = \mathbf{X}_s^{-1} \otimes \mathbf{X}_t$. However, for a branched rough path, this is considered part of the definition. We will show that a branched rough path can equivalently be defined as a a path $\mathbf{X} : [0,T] \to G_N$, where $(G_N, \star)$ is the Lie group of step-N truncated group-like elements (or characters) in the Connes-Kreimer Hopf algebra, satisfying

$$\langle g, xy\rangle = \langle g, x\rangle\langle g, y\rangle \ ,$$

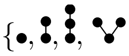
for all $x, y \in \mathcal{H}$. This allows us to define $\mathbf{X}_{st} = \mathbf{X}_s^{-1} \star \mathbf{X}_t$ and hence guarantee Chen's property from the definition. The Lie group $(G_N, \star)$ bears great similiarity to the step-N free nilpotent group, since it is the truncated set of characters in $\mathcal{H}$, and the step-N free nilpotent group is the truncated set of characters in $T(V)$. Moreover, one obtains G_N as the $\exp_\star$ of the Lie algebra of so-called primitive elements, where $\exp_\star$ is simply the tensor exponential, with tensor products replaced with $\star$ products.

The outline of the chapter is as follows. In Section 3.2 we define the algebraic concepts underlying branched rough paths, including the Connes-Kreimer Hopf algebra. We then provide a definition of branched rough paths, equivalent to that given in [Gub10], that is more in line with the concept of a geometric rough path. In Section 3.3, we define solutions to RDEs driven by branched rough paths, via the idea of controlled rough paths.

3.2 Hopf algebras and branched rough paths

3.2.1 The Connes-Kreimer Hopf algebra

Let $\mathcal{T}$ be the set of all rooted trees with finitely many vertices, whose vertices are decorated by labels from the alphabet $\{1, \dots, d\}$. Every element in $\mathcal{T}$ can be constructed recursively by attaching a collection of trees (of lower order) to a new root. For example, the set of (undecorated) trees with three vertices or less is given by

$$\mathcal{T}_3 = \{$$  $$\} \ .$$

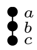
We can then construct all single vertex trees by attaching the empty tree 1 to a new root. We denote this by

$$[1]_a = \bullet_a \ ,$$

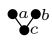
for any a from the alphabet. All trees of two vertices can be constructed by attaching these trees to a new root

$$[\bullet_a]_b =$$  .

For the trees of three vertices, we similarly have

$[$  $]_c \quad = \quad$  .

The remaining tree in $\mathcal{T}_3$ is obtained by attaching a pair of single vertex trees to a root

$$[\ \bullet_a \bullet_b\ ]_c =$$  .

Indeed, every element in $\mathcal{T}$ can be written recursively as

$$[\tau_1 \tau_2 \dots \tau_m]_a \ , \tag{3.9}$$

for some smaller trees $\tau_1, \dots, \tau_m \in \mathcal{T} \cup \{1\}$ and some a from the alphabet. We will always assume that the order of the branches in each tree does not matter, in the sense that $[\tau_1 \dots \tau_n]_i = [\tau_{\sigma(1)} \dots \tau_{\sigma(n)}]_i$ for all permutations σ of $\{1, \dots, n\}$. For each $[\tau_1 \dots \tau_n]_i$, only one such representation appears in the set $\mathcal{T}$.

Remark 3.2.2. In the rough path setting, rearranging branches in a tree corresponds to rearranging real-valued factors in an integrand. Hence, this is quite a natural assumption to make.

The *Connes-Kreimer Hopf algebra* $(\mathcal{H}, \cdot, \Delta, \mathcal{S})$ is the commutative polynomial algebra generated by the variables $\mathcal{T}$, equipped with a coproduct $\Delta : \mathcal{H} \to \mathcal{H} \tilde{\otimes} \mathcal{H}$ and an antipode $\mathcal{S} : \mathcal{H} \to \mathcal{H}$. Alternatively, we can view the set $\mathcal{H}$ as a real vector space whose basis is the commutative monoid $\mathcal{F} \cup \{1\}$ where $\mathcal{F}$ is given by

$$\mathcal{F} = \{\tau_1 \dots \tau_n : \tau_i \in \mathcal{T},\ n \in \mathbb{N}^+\} \ .$$

Each monomial $\tau_1 \dots \tau_n$ can be thought of as an *unordered forest*, since the polynomial product is commutative. Hence, a typical element of $\mathcal{H}$ is for example

$$[[\bullet_1]_2]_3 + 6\, \bullet_3 [\bullet_1]_2 - \sqrt{2}\, \bullet_3 [\bullet_3 \bullet_2]_1 .$$

Remark 3.2.3. We could equally construct the Connes-Kreimer Hopf algebra $\mathcal{H}(\mathcal{A})$, using any countable alphabet $\mathcal{A}$ in place of $\{1, \dots, d\}$. However, since $\{1, \dots, d\}$ this is the most commonly used choice, we reserve the notation $\mathcal{H}$ for this particular alphabet.

The coproduct Δ is defined recursively. We first set $\Delta 1 = 1 \tilde{\otimes} 1$, then for any $[\tau_1 \dots \tau_m]_a \in \mathcal{T}$ we set

$$\Delta[\tau_1 \dots \tau_m]_a = [\tau_1 \dots \tau_m]_a \tilde{\otimes} 1 + \sum_{(\tau_1)\dots(\tau_m)} (\tau_1^{(1)} \dots \tau_m^{(1)}) \tilde{\otimes} [\tau_1^{(2)} \dots \tau_m^{(2)}]_a \,, \tag{3.10}$$

where we use the Sweedler notation $\Delta x = \sum_{(x)} x^{(1)} \tilde{\otimes} x^{(2)}$. In the sequel, we will often omit the summation sign and simply write $\Delta x = x^{(1)} \tilde{\otimes} x^{(2)}$. We then extend Δ to all polynomials by requiring that it be linear and also a morphism with respect to polynomial multiplication, that is

$$\Delta(\tau_1 \dots \tau_n) = \Delta\tau_1 \dots \Delta\tau_n \,,$$

for every $\tau_i \in \mathcal{T}$. It is often useful to consider the *reduced coproduct* Δ' defined by $\Delta' x = \Delta x - 1 \tilde{\otimes} x - x \tilde{\otimes} 1$. In any coalgebra, the coproduct is required to be *coassociative*, which means that

$$(\Delta \tilde{\otimes} \mathrm{Id})\Delta = (\mathrm{Id} \tilde{\otimes} \Delta)\Delta \,.$$

One can check that this is true for both the coproduct and the reduced coproduct described above.

Remark 3.2.4. It is natural to ask why one needs to consider polynomials of $\mathcal{T}$ rather than just the set of trees. Indeed, for non-geometric rough paths, the trees are the important ingredients when solving an RDE. The reason we require polynomials is that we would like to define a rough path as a functional on some algebra, and this algebra must be big enough to include an element xy such that

$$\langle \mathbf{X}, x \rangle \langle \mathbf{X}, y \rangle = \langle \mathbf{X}, xy \rangle \,.$$

This, in particular, allows us to write Chen's property as a fundamental operation on the algebra $\mathcal{H}$, described by the coproduct Δ, rather than just an identity on the tree indexed components of $\mathbf{X}$.

Remark 3.2.5. Let $\mathcal{B}$ be the vector space spanned by the set of trees $\mathcal{T}$. This is clearly a subspace of $\mathcal{H}$. Then the tensor product algebra $T(\mathbb{R}^d)$ can easily be identified with the subspace of $\mathcal{B}$, and hence $\mathcal{H}$, spanned by the linear trees. This is achieved by identifying

$$e_a \otimes e_b \cong \substack{\bullet_a \\ \bullet_b} \quad , \quad e_a \otimes e_b \otimes e_c \cong \substack{\bullet_a \\ \bullet_b \\ \bullet_c} \; ,$$

and so forth, for any $a, b, c = 1 \dots d$. In the sequel, we will refer to this identification via the inclusion map $\iota : T(\mathbb{R}^d) \to \mathcal{H}$. In light of this, we should think of the Hopf algebra $\mathcal{H}$ as being an extension of the tensor product algebra over the same index set. As discussed in the introduction, the extra *branched* objects are required to encode a non-trivial product that cannot be described by objects in the tensor product algebra alone.

Remark 3.2.6. The definition (3.10) is indeed quite a natural one. If X were a smooth path in $\mathbb{R}^d$ then we could build the branched rough path $\mathbf{X}$ canonically, by setting

$$\langle \mathbf{X}_{st}, \bullet_i \rangle = \delta X^i_{st} \quad \text{and} \quad \langle \mathbf{X}_{st}, [\tau_1 \dots \tau_n]_i \rangle = \int_s^t \langle \mathbf{X}_{sr}, \tau_1 \rangle \dots \langle \mathbf{X}_{sr}, \tau_n \rangle dX^i_r \; .$$

Using the properties of a path integral, namely, linearity with respect to the integrand and the adjacent interval property $\int_s^t = \int_s^u + \int_u^t$, one can recursively show that

$$\begin{aligned}\langle \mathbf{X}_{st}, [\tau_1 \dots \tau_n]_i \rangle &= \langle \mathbf{X}_{su}, [\tau_1 \dots \tau_n]_i \rangle \\ &\quad + \sum_{(\tau_1) \dots (\tau_n)} \left(\langle \mathbf{X}_{su}, \tau_1^{(1)} \rangle \dots \langle \mathbf{X}_{su}, \tau_n^{(1)} \rangle \right) \int_u^t \langle \mathbf{X}_{ur}, \tau_1^{(2)} \rangle \dots \langle \mathbf{X}_{ur}, \tau_n^{(2)} \rangle dX^i_r\end{aligned}$$

or in other words,

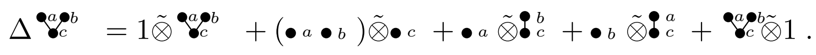
$$\langle \mathbf{X}_{st}, [\tau_1 \dots \tau_n]_i \rangle = \langle \mathbf{X}_{su} \tilde{\otimes} \mathbf{X}_{ut}, \Delta [\tau_1 \dots \tau_n]_i \rangle \; , \tag{3.11}$$

with Δ satisfying (3.10). Hence, (3.11) is an extension of Chen's property to more complicated looking integrals.

Remark 3.2.7. When restricted to linear trees (or tensor products), the coproduct Δ is known as *deconcatenation*, since it decomposes tensor products into subproducts that can be concatenated in to the original expression. There is a similar interpretation for Δ on all of $\mathcal{H}$, which is described by *cuts* of a tree. We will say that the pair $(\tau_1 \dots \tau_m) \tilde{\otimes} \tau_0$ is an admissible cut of $\tau \in \mathcal{T}$, if one can obtain τ by *attaching* the trees $\tau_1, \dots, \tau_m$ to the nodes of τ_0. We then have the interpretation

$$\Delta \tau = \sum_{(\tau)} \tau^{(1)} \tilde{\otimes} \tau^{(2)} \; ,$$

where we sum over all admissible cuts $\tau^1 \tilde{\otimes} \tau^2$, with τ^1 and τ^2 playing the roles of $(\tau_1 \dots \tau_m)$ and τ_0 respectively. For example, we have that



In particular, we always have that $\Delta\tau = 1\tilde{\otimes}\tau + \tau\tilde{\otimes}1 + \tau^1\tilde{\otimes}\tau^2$, where for each term $\tau^1\tilde{\otimes}\tau^2 \in \mathcal{F}\tilde{\otimes}\mathcal{T}$, we have that $|\tau^1| + |\tau^2| = |\tau|$, recalling that $|\cdot|$ simply counts the number of vertices in a forest or tree. This observation will be crucial in the sequel.

Remark 3.2.8. Although both $\otimes$ and $\tilde{\otimes}$ are tensor products, we reserve the former for the product in the tensor product algebra $T(V)$ and the latter simply to discriminate between the left and the right part of a coproduct. If x, y are two elements in some algebra and f, g are two maps on that algebra, then we use the convention $(f\tilde{\otimes}g)(x\tilde{\otimes}y) = f(x)\tilde{\otimes}g(y)$.

In any Hopf algebra, the *antipode* $\mathcal{S} : \mathcal{H} \to \mathcal{H}$ is a morphism of bialgebras satisfying

$$M(\mathrm{Id}\tilde{\otimes}\mathcal{S})\Delta x = M(\mathcal{S}\tilde{\otimes}\mathrm{Id})\Delta x = x \ ,$$

for any $x \in \mathcal{H}$, where M is the multiplication map $M(x\tilde{\otimes}y) = xy$. The existence of an antipode for $\mathcal{H}$ follows from standard results for bialgebras, as detailed in Remark 3.2.9. For the Connes-Kreimer Hopf algebra the antipode has been explicitly constructed in [CK98].

Remark 3.2.9. It is well known that if a bialgebra is irreducible as a coalgebra, then it is a Hopf algebra [Swe69, Theorem 9.2.2]. Irreducibility simply means that every two subcoalgebras have a non-empty intersection. This is guaranteed for any graded bialgebra, whose zeroth component is the field. Moreover, if a bialgebra has an antipode, then it is unique [DNR01]. Once we have defined the grading, this will clearly be the case for the Connes-Kreimer Hopf algebra. The role of the antipode will become clear in the next subsection.

The bialgebra $(\mathcal{H}, \cdot, \Delta)$ gives rise to a dual bialgebra $(\mathcal{H}^*, \star, \delta, \mathcal{S}^*)$. Since $\mathcal{H}$ is a countable vector space, the elements in $\mathcal{H}^*$ can be identified with formal series of elements in $\mathcal{H}$. In particular, we identify elements in the basis $\mathcal{F}$ with elements in $\mathcal{H}^*$ by the natural pairing $\langle \sigma_1, \sigma_2 \rangle = \mathbf{1}(\sigma_1 = \sigma_2)$ for $\sigma_1, \sigma_2 \in \mathcal{F}$. The co-unit $1 \in \mathcal{H}^*$ is the map satisfying $\langle 1, 1 \rangle = 1$ and $\langle 1, \tau_1 \dots \tau_n \rangle = 0$ for all $\tau_1 \dots \tau_n \in \mathcal{F}$.

Remark 3.2.10. In the sequel, our notation does not distinguish between the unit and the co-unit, nor the basis $\mathcal{F}$ and its dual $\mathcal{F}^*$ (and likewise $\mathcal{T}$ and $\mathcal{T}^*$). However, it will always be clear from the context which we are referring to.

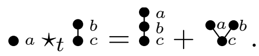
The product $\star : \mathcal{H}^*\tilde{\otimes}\mathcal{H}^* \to \mathcal{H}^*$, often referred to as *convolution*, is the dual of Δ, that is

$$\langle h_1 \star h_2, x\rangle \stackrel{\text{def}}{=} \langle h_1 \tilde{\otimes} h_2, \Delta x\rangle = \sum_{(x)} \langle h_1, x^1\rangle\langle h_2, x^2\rangle \ ,$$

for any $h_1, h_2 \in \mathcal{H}^*$ and $x \in \mathcal{H}$. It follows from the properties of the coproduct Δ (namely, coassociativity) that $\star$ provides $\mathcal{H}$ with an associative algebra structure. Let $\mathcal{T}^*$ denote those elements in $\mathcal{H}^*$ that correspond to dual elements of $\mathcal{T}$. Then for $\tau_1, \tau_2 \in \mathcal{T}^*$, the product $\tau_1 \star \tau_2$ can be interpreted as attaching τ_1 to τ_2. In particular, we have that

$$\tau_1 \star \tau_2 = \tau_1\tau_2 + \tau_1 \star_t \tau_2 \ , \tag{3.12}$$

where $\tau_1 \star_t \tau_2$ is the sum of all trees in $\mathcal{T}^*$ obtained by growing τ_1 from a vertex of τ_2. For example,



This is often referred to as the *Grossman-Larson product*, and was first discussed in [GL89]. The antipode $\mathcal{S}$ plays the role of an inverse with respect to $\star$ in the space $\mathcal{H}^*$, this will become clear in the next subsection. The dual coproduct $\delta : \mathcal{H}^* \to \mathcal{H}^* \tilde{\otimes} \mathcal{H}^*$ is likewise the dual of polynomial multiplication

$$\langle \delta\tau, x\tilde{\otimes} y\rangle = \langle \tau, xy\rangle \ .$$

Just as above, this endows $\mathcal{H}^*$ with a coassociative coalgebra structure and it is a nice exercise to check that δ is a morphism with repsect to $\star$, as every coproduct should be.

The trees $\mathcal{T}$ give rise to a natural *grading* on $\mathcal{H}$. For each $\tau \in \mathcal{T}$, we define $|\tau|$ to be the number of vertices in τ. We extend $|\cdot|$ to products by

$$|\tau_1 \dots \tau_n| = |\tau_1| + \dots + |\tau_n| \ ,$$

for any $\tau_i \in \mathcal{T}$. If we let $\mathcal{F}_{(k)}$ denote the set of $\tau_1 \dots \tau_m \in \mathcal{F}$ with $|\tau_1 \dots \tau_m| = k$ and $\mathcal{H}_{(k)}$ denote the real vector space spanned by $\mathcal{F}_{(k)}$, with $\mathcal{H}_{(0)} = \mathbb{R}$, then we clearly have

$$\mathcal{H} = \bigoplus_{k=0}^{\infty} \mathcal{H}_{(k)} \ ,$$

so that $|\cdot|$ turns $\mathcal{H}$ into a graded Hopf algebra. We will also make use of the truncated algebra

$$\mathcal{H}_n = \bigoplus_{k=0}^{n} \mathcal{H}_{(k)}$$

and its basis elements $\mathcal{F}_n$, containing all $m \in \mathcal{F}$ with $|m| \le n$. Keeping in line with this notation, we also define $\mathcal{T}_{(n)}$ as the set of $\tau \in \mathcal{T}$ with $|\tau| = n$ and $\mathcal{T}_n$ as the set of $\tau \in \mathcal{T}$ with $|\tau| \leq n$.

3.2.11 Group-like and primitive elements

We will denote by $\mathrm{Hom}(\mathcal{H}, \mathbb{R})$ those elements in $\mathcal{H}^*$ that are also homomorphisms with respect to polynomial multiplication $\cdot$, that is, $h \in \mathrm{Hom}(\mathcal{H}, \mathbb{R})$ if and only if

$$\langle h, xy \rangle = \langle h, x \rangle \langle h, y \rangle \ . \tag{3.13}$$

These are also known as the *characters* of $\mathcal{H}$. It is easy to check that $\mathrm{Hom}(\mathcal{H}, \mathbb{R})$ can be identified with the *group-like elements*, defined by

$$G(\mathcal{H}) = \{g \in \mathcal{H}^* : \delta g = g \tilde{\otimes} g\} \ .$$

In particular, the equality (3.13) holds if and only if

$$\langle \delta h, x \tilde{\otimes} y \rangle = \langle h, xy \rangle = \langle h, x \rangle \langle h, y \rangle = \langle h \tilde{\otimes} h, x \tilde{\otimes} y \rangle \ ,$$

for all $x, y \in \mathcal{H}$. The reason $G(\mathcal{H})$ is called the set of group-like elements is because it is indeed a group. For the Connes-Kreimer Hopf algebra, this is often referred to as the *Butcher group* [HW74].

Proposition 3.2.12. *The pair $(G(\mathcal{H}), \star)$ is a group with inverses given by $g^{-1} \stackrel{\mathrm{def}}{=} \mathcal{S}^* g$, where $\mathcal{S}^*$ is the adjoint of the antipode.*

Proof. Standard result for Hopf algebras and an easy exercise. □

The group property of $\mathrm{Hom}(\mathcal{H}, \mathbb{R})$ is one of the main motivations behind Hopf algebras and in particular explains the role of the antipode. Indeed, the concept of a Hopf algebra is often introduced as the linearisation of a group.

If we were to replace $\mathcal{H}$ with the tensor product space $T(V)$ over the vector space $V = \mathbb{R}^d$, then we could equivalently characterise each group-like elements as the exponential of a Lie polynomial [Reu93, Theorem 1.4]. Remarkably, the same construction works in this setting too. We define the bracket $[\cdot, \cdot]_\star : \mathcal{H}^* \times \mathcal{H}^* \to \mathcal{H}^*$ by

$$[x, y]_\star = x \star y - y \star x \ , \tag{3.14}$$

which one can easily check is a Lie bracket. We define the set of δ-primitives as

$$P(\mathcal{H}) = \{h \in \mathcal{H}^* : \delta h = 1 \tilde{\otimes} h + h \tilde{\otimes} 1\} \ , \tag{3.15}$$

where 1 is the co-unit in $\mathcal{H}^*$. In the context of Lie algebras, this condition is often stated as $\langle h, xy\rangle = \langle 1, x\rangle\langle h, y\rangle + \langle h, x\rangle\langle 1, y\rangle$ and the elements are known as *derivations*. As suggested by the notation, $P(\mathcal{H})$ is a Lie algebra with respect to $[\cdot,\cdot]_\star$ and has a very natural basis in $\mathcal{H}^*$.

Proposition 3.2.13. *The set $P(\mathcal{H})$ is a Lie algebra with bracket $[\cdot,\cdot]_\star$ and moreover*

$$P(\mathcal{H}) = \mathcal{B} ,$$

where $\mathcal{B}$ is the real vector space spanned by the dual trees $\mathcal{T}^*$.

Proof. Let $\tau \in \mathcal{T}^*$, then by definition

$$\delta\tau = \sum_{x_1,x_2\in\mathcal{F}^0} \langle \tau, x_1x_2\rangle x_1\tilde{\otimes}x_2 ,$$

where $\mathcal{F}^0 = \mathcal{F} \cup \{1\}$. But clearly, $\langle\tau, x_1x_2\rangle = 0$ unless $x_1 = 1$ or $x_2 = 1$ (but not both). It follows that

$$\delta\tau = \sum_{x_1\in\mathcal{F}^0} \langle\tau, x_1\rangle\big(x_1\tilde{\otimes}1 + 1\tilde{\otimes}x_1\big) = \tau\tilde{\otimes}1 + 1\tilde{\otimes}\tau ,$$

and hence $\mathcal{B} \subseteq P(\mathcal{H})$. To prove the reverse statement, suppose $h \in P(\mathcal{H})$ and that

$$h = u + v ,$$

where $u \in \mathcal{B}$ and $v \in \mathcal{B}^\perp$, which is the vector space spanned by 1 and all non-trivial products $\tau_1 \dots \tau_n \in \mathcal{F}^*$ with $n \geq 2$. Since $u \in P(\mathcal{H})$, it follows that $v = h - u \in P(\mathcal{H})$. Thus,

$$1\tilde{\otimes}v + v\tilde{\otimes}1 = \delta v = \langle v, 1\rangle 1\tilde{\otimes}1 + \sum_{\tau_1\dots\tau_n} \langle v, \tau_1\dots\tau_n\rangle \delta(\tau_1\dots\tau_n) ,$$

where we only sum over those $\tau_1 \dots \tau_n \in \mathcal{F}^*$ with $n \geq 2$. By definition of $\delta(\tau_1 \dots \tau_n)$, this equals

$$\langle v, 1\rangle 1\tilde{\otimes}1 + \sum_{\tau_1\dots\tau_n} \langle v, \tau_1\dots\tau_n\rangle \sum_{(i,j)} \tau_{i_1}\dots\tau_{i_p}\tilde{\otimes}\tau_{j_1}\dots\tau_{j_q} , \tag{3.16}$$

where we sum over all subsets $\{i_1, \dots, i_p\}$, $\{j_1, \dots, j_q\}$ of $\{1, \dots, n\}$. However, each term $\tau_{i_1}\dots\tau_{i_p}\tilde{\otimes}\tau_{j_1}\dots\tau_{j_q}$ (with $p, q \neq 0$) can only appear once in the expression (3.16), hence there can be no cancellations. Since these terms (as well as $1\tilde{\otimes}1$) are basis elements of $\mathcal{H}^*\tilde{\otimes}\mathcal{H}^*$, we must have that $\langle v, 1\rangle = 0$ and $\langle v, \tau_1\dots\tau_n\rangle = 0$ for all $\tau_1 \dots \tau_n \in \mathcal{F}^*$ with $n \geq 2$. It follows that $P(\mathcal{H}) \subseteq \mathcal{B}$. ☐

Let $\mathfrak{h}$ be the space of all $h \in \mathcal{H}^*$ with $\langle h, 1\rangle = 0$ and let $H = 1 + \mathfrak{h}$. Just as in the tensor product algebra case, the spaces $\mathfrak{h}$ and H are diffeomorphic via the exponential map $\exp_\star : \mathfrak{h} \to H$ given by

$$\exp_\star h = \sum_{k\geq 0} \frac{h^{\star k}}{k!} \ ,$$

where $h^{\star k} = h \star h^{\star(k-1)}$. Likewise we can define its inverse, the logarithmic map by

$$\log_\star(1 + h) = \sum_{k\geq 1}(-1)^{k-1}\frac{h^{\star k}}{k} \ ,$$

for any $1+h \in H$. See [Man04] for further details. This allows us to classify the group-like elements as being the exponential of a Lie element.

Proposition 3.2.14. *For any $g \in \mathfrak{h}$, we have that $g \in G(\mathcal{H})$ if and only if $g = \exp_\star h$ for some $h \in P(\mathcal{H})$.*

Proof. The proof is identical to the tensor product algebra case [Reu93, Theorem 3.2], but we will include it for completeness. Suppose $y \in P(\mathcal{H})$, we will show that $\delta \exp_\star y = \exp_\star y \tilde{\otimes} \exp_\star y$. Since δ is a morphism with respect to $\star$, we have that

$$\delta \exp_\star y = \exp_\star \delta y = \exp_\star(1\tilde{\otimes} y + y\tilde{\otimes} 1) \ .$$

Since $1\tilde{\otimes}y$ and $y\tilde{\otimes}1$ commute under $\star$, the above is equal to

$$\exp_\star(1\tilde{\otimes}y) \star \exp_\star(y\tilde{\otimes}1) = (1\tilde{\otimes} \exp_\star(y)) \star (\exp_\star(y)\tilde{\otimes}1) = \exp_\star(y)\tilde{\otimes} \exp_\star(y) \ .$$

Now let $x \in G(\mathcal{H})$. To prove that

$$\delta \log_\star x = 1\tilde{\otimes} \log_\star x + \log_\star x\tilde{\otimes}1 \ ,$$

we simply take $\exp_\star$ on both sides of the expression and the result follows using the same technique as above. □

3.2.15 Branched rough paths

We define the truncated group-like elements $G_N(\mathcal{H})$, obtained from $G(\mathcal{H})$ by quotienting the ideal

$$\bigoplus_{k=N+1}^{\infty} \mathcal{H}^*_{(k)} \ ,$$

hence we identify all elements $\tau_1 \dots \tau_n \in \mathcal{F}^*$ such that $|\tau_1 \dots \tau_n| \geq N + 1$, with zero. From Proposition 3.2.14, it follows that $G_N(\mathcal{H})$ is diffeomorphic to the real vector space

$\mathcal{T}_N$ and is therefore a Lie group. This Lie group plays precisely the same role as the step N free nilpotent group in the geometric theory of rough paths. Indeed, the definition for branched rough paths follows naturally from that of geometric rough paths.

Let $X = (X^i)$ be a path in $\mathbb{R}^d$ with Hölder regularity $\gamma \in (0,1)$. Throughout the sequel, we will reserve the constant N as the largest integer such that $N\gamma \leq 1$.

Definition 3.2.16. A map $\mathbf{X} : [0,T] \to G_N(\mathcal{H})$ is called a γ-Hölder *branched rough path* if it satisfies

$$\sup_{s\neq t} \frac{|\langle \mathbf{X}_{st}, \tau\rangle|}{|t-s|^{\gamma|\tau|}} < \infty \ , \tag{3.17}$$

for every $\tau \in \mathcal{H}_N$ and where $\mathbf{X}_{st} \stackrel{\text{def}}{=} \mathbf{X}_s^{-1} \star \mathbf{X}_t$. If $\langle \mathbf{X}_{st}, \bullet_i\rangle = \delta X^i_{st}$ for each $i = 1 \dots d$ then we call $\mathbf{X}$ a branched rough path *above* X.

We see that the generalised version of Chen's property, or Condition (3) of the introduction, is immediate from the definition, since we have

$$\mathbf{X}_{su} \star \mathbf{X}_{ut} = (\mathbf{X}_s^{-1} \star \mathbf{X}_u) \star (\mathbf{X}_u^{-1} \star \mathbf{X}_t) = \mathbf{X}_{st} \ . \tag{3.18}$$

Moreover, Definition 3.2.16 is clearly equivalent to the original definition in [Gub10] and also stated in the introduction. In particular, Condition 1 from the original definition can be reformulated as $\mathbf{X}_{st} \in G_N(\mathcal{H})$ for each $s,t \in [0,T]$.

Remark 3.2.17. As we shall see, the solution to an RDE only depends on the increment $\mathbf{X}_{st}$ rather than the path $\mathbf{X}_t$, hence there is no need to specify the initial value of the path $\mathbf{X}_0$.

Remark 3.2.18. In Definition 3.2.16, to justify calling $\mathbf{X}$ a γ-Hölder path, it should satisfy $d(\mathbf{X}_s, \mathbf{X}_t) \leq C|t-s|^\gamma$ for some metric d. This can be achieved using *homogeneous norms*. For the step N free nilpotent group, as with any Carnot group, one can show that all "norms" that are sub-additive and homogeneous with respect to the natural dilation of the group are equivalent [LV07]. This does not quite work with $G_N(\mathcal{H})$, since it is not a Carnot group with respect to the right dilation. To be precise, we see that

$$\mathcal{G}_N(\mathcal{H}) = \exp_\star\left(\bigoplus_{k=0}^{N} \mathcal{B}_{(k)}\right) \ ,$$

where $\mathcal{B}_{(k)}$ is the vector space spanned by $\mathcal{T}_{(k)}$. If $y_k \in \mathcal{B}_{(k)}$, then the natural dilation on $G_N(\mathcal{H})$ is given by

$$\delta_t \exp_\star (y_1 + \cdots + y_N) = \exp_\star \left(t y_1 + t^2 y_2 + \cdots + t^N y_N\right) \ .$$

In particular, in the case of a smooth path X whose branched rough path $\mathbf{X}$ is given by the corresponding iterated integrals, if we multiplied X by t, then we would obtain a factor of $t^{|\tau|}$ infront of $\langle \mathbf{X}, \tau\rangle$, so it is clear that this is the right choice of dilation. On the other hand, the only way $G_N(\mathcal{H})$ could be a Carnot group is if we let all of $\mathcal{B}_N$ have the same grade, which would lead to useless norms. The correct notion is to view $G_N(\mathcal{H})$ as a *homogeneous group*, as defined in [FS82]. A homogeneous group G is a Lie group whose Lie algebra is graded, and hence comes with a natural dilation. A (non-smooth) homogeneous norm on G is then a map $\|\cdot\| : G \to [0,\infty)$ that is continuous with respect to the manifold topology of G and satisfies the homogeneity property $\|\delta_t g\| = |t|\|g\|$, where δ_t is the natural dilation of G (along with other standard conditions). It is easy to show that all homogeneous norms on G are equivalent. In the case of $G_N(\mathcal{H})$, one can show that all homogeneous norms are equivalent to the natural homogeneous norm

$$\|g\|_{G_N(\mathcal{H})} = \sum_{\tau\in\mathcal{T}_N} |\langle \log_\star g, \tau\rangle|^{1/|\tau|} \ .$$

Moreover, it is easy to verify that the map $\mathbf{X} : [0,T] \to \big(G_N(\mathcal{H}), \|\cdot\|_{G_N(\mathcal{H})}\big)$ is γ-Hölder continuous if and only if condition (3.17) is satisfied. We can therefore equivalently define a branched rough path as a γ-Hölder path taking values in $\big(G_N(\mathcal{H}), \|\cdot\|_{G_N(\mathcal{H})}\big)$.

As with classical rough paths, one can show that every branched rough path $\mathbf{X}$ can be canonically extended to a γ-Hölder continuous path taking values in $G(\mathcal{H})$, courtesy of the *sewing map* [Gub04, Gub10]. In more generality, branched rough paths also extend the idea of an almost multiplicative functional, in the following way. One calls $\tilde{\mathbf{X}}$ an *almost branched rough path* if

$$|\langle \tilde{\mathbf{X}}_{st} - \tilde{\mathbf{X}}_{su} \star \tilde{\mathbf{X}}_{ut}\rangle| = o(|t-s|) \ ,$$

for all $s,t \in [0,T]$, with $|t-s| \ll 1$. And moreover, we have the following

Proposition 3.2.19. *For every almost branched rough path $\tilde{\mathbf{X}}$, there exists a unique branched rough path $\mathbf{X}$ of roughness γ such that*

$$|\langle \mathbf{X}_{st} - \tilde{\mathbf{X}}_{st}, \tau\rangle| = o(|t-s|) \ ,$$

for all $s,t \in [0,T]$ and $\tau \in \mathcal{T}_N$.

Proof. See [Gub10, Theorem 7.7]. □

Although we will not explicitly use the notion of an almost branched rough path, we include the definition to illustrate that all of the important tools for multiplicative functionals are still present in the setting of branched rough paths.

3.3 Controlled rough paths and solving RDEs

In this section we recall the definition of a *controlled rough path*, first defined in [Gub04] and later extended to branched rough paths in [Gub10]. We show how one can define rough integrals and moreover solutions to RDEs using this simple concept.

3.3.1 Controlled rough paths

A crucial step in the theory of geometric rough paths is defining the integral of a one-form along a geometric rough path [Lyo98]. For $\alpha : \mathbb{R}^d \to L(\mathbb{R}^d, \mathbb{R})$ and a geometric rough path $\mathbf{X}$ above $X \in \mathbb{R}^d$, in order to define $\int \alpha(X)dX$ one needs to impose a $\mathrm{Lip}(\beta)$ condition on α, which states that for $j = 1 \ldots N$, there exists $\alpha^j : \mathbb{R}^d \to L((\mathbb{R}^d)^{\otimes j}, \mathbb{R})$ such that $\alpha^1 = \alpha$ and

$$\alpha^j(X_t) = \sum_{i=0}^{N-j} \alpha^{i+j}(X_s)(\mathbf{X}^i_{st}) + R^j(X_s, X_t) , \tag{3.19}$$

where $\mathbf{X}^i_{st}$ is the component of $\mathbf{X}_{st}$ in $(\mathbb{R}^d)^{\otimes i}$ and the remainders R^j satisfy $|R^j(\xi, \eta)| \leq M|\xi - \eta|^{\beta - j}$. In particular, from the $j = 1$ case we see that

$$\alpha(X_t) - \alpha(X_s) = \sum_{i=1}^{N-1} \alpha^{i+1}(X_s)(\mathbf{X}^i_{st}) + R^1(X_s, X_t) , \tag{3.20}$$

and hence the increment of $\alpha(X)$ is (locally) controlled by $\mathbf{X}$. The expression (3.20) leads directly to a definition of an almost multiplicative functional $\widetilde{\mathbf{Y}}$ which is subsequently extended to define $\int \alpha(X)dX$. The conditions on the higher order α^j given in (3.19) are required to ensure that $\widetilde{\mathbf{Y}}$ actually is an almost multiplicative functional and thus prove that the map $\mathbf{X} \mapsto \int \alpha(X)dX$ is continuous in the p-var topology.

In the theory of controlled rough paths, the construction of integrals is more-or-less the same, except for that fact that one-forms are replaced with any object that satisfies a condition like (3.19). In particular let $\mathbf{X}$ be a branched rough path above X and suppose $Z : [0, T] \to \mathbb{R}$ satisfies

$$\delta Z_{st} = \sum_{h \in \mathcal{F}_{N-1}} Z^h_s \langle \mathbf{X}_{st}, h \rangle + R^Z_{st} , \tag{3.21}$$

where $|R^Z_{st}| \leq C|t - s|^{N\gamma}$ and the coefficients $Z^h : [0, T] \to \mathbb{R}$. It is clear that the integral

$\int_s^t Z_r dX_r^i$ should be approximated by the expression

$$\tilde{Z}_{st} = Z_s \langle \mathbf{X}_{st}, \bullet_i \rangle + \sum_{h \in \mathcal{F}_{N-1}} Z_s^h \langle \mathbf{X}_{st}, [h]_i \rangle = \sum_{h \in \mathcal{F}_{N-1}^0} Z_s^h \langle \mathbf{X}_{st}, [h]_i \rangle \ , \tag{3.22}$$

for $|t-s| \ll 1$ where we denote $Z_s^1 = Z_s$ and $\mathcal{F}_{N-1}^0 = \mathcal{F}_{N-1} \cup \{1\}$, since one would expect

$$\int_s^t dX_r^i = \langle \mathbf{X}_{st}, \bullet_i \rangle \quad \text{and} \quad \int_s^t \langle \mathbf{X}_{sr}, h \rangle dX_r^i = \langle \mathbf{X}_{st}, [h]_i \rangle \ .$$

This idea is formalised by the *sewing map*. The sewing map is essential the same as the map which extends an almost multiplicative functional to an (approximately equal) multiplicative functional. I

Lemma 3.3.2 (Sewing Map)**.** *For any* $\tilde{Z} : [0,T] \times [0,T] \to \mathbb{R}$*, if*

$$|\tilde{Z}_{st} - \tilde{Z}_{su} - \tilde{Z}_{ut}| \leq C|t-u|^p |u-s|^q \ , \tag{3.23}$$

for some $p+q > 1$*, then there exists a unique remainder terms* $r : [0,T] \times [0,T] \to \mathbb{R}$ *such that* $\tilde{Z}_{st} + r_{st}$ *is the increment of a path and* $|r_{st}| = o(|t-s|)$*. That is, there is some* $Y : [0,T] \to \mathbb{R}$ *such that*

$$\delta Y_{st} = \tilde{Z}_{st} + r_{st} \ .$$

for all $s, t \in [0,T]$*.*

Just as in (3.19), one needs conditions on the coefficients Z_s^h to ensure that $\tilde{Z}$ defined in (3.22) satisfies (3.23). The most convenient way of defining these controlled objects Z along with their coefficients Z^h is to consider them as one object $\mathbf{Z} : [0,T] \to \mathcal{H}_{N-1}$, by setting

$$\langle 1, \mathbf{Z}_t \rangle = Z_t \quad \text{and} \quad \langle h, \mathbf{Z}_t \rangle = Z_t^h \ ,$$

for all $h \in \mathcal{F}_{N-1}$. In the sequel we use the notation $\mathcal{F}_n^0 = \mathcal{F}_n \cup \{1\}$ and similarly for $\mathcal{T}_n^0$.

Definition 3.3.3. Let $\mathbf{X}$ be a γ-Hölder branched rough path. An $\mathbf{X}$-controlled rough path is a path $\mathbf{Z} : [0,T] \to \mathcal{H}_{N-1}$ satisfying

$$\langle h, \mathbf{Z}_t \rangle = \langle \mathbf{X}_{st} \star h, \mathbf{Z}_s \rangle + R_{st}^h \ , \tag{3.24}$$

for each $h \in \mathcal{F}_{N-1}^0$, where $|R_{st}^h| \leq C|t-s|^{(N-|h|)\gamma}$. When $\langle 1, \mathbf{Z}_t \rangle = Z$, we say that $\mathbf{Z}$ is a controlled rough path above Z.

Note that when $h = 1$ and $\langle 1, \mathbf{Z} \rangle = Z$, the expression (3.24) can be written

$$\delta Z_{st} = \sum_{h \in \mathcal{F}_{N-1}} \langle h, \mathbf{Z}_s \rangle \langle \mathbf{X}_{st}, h \rangle + R^1_{st} \ ,$$

just as suggested in (3.21). It is clear that (3.24) is simply the $\mathcal{H}$ counterpart of the $\mathrm{Lip}(\beta)$ condition (3.19).

Remark 3.3.4. We can easily adapt this to the situation in which the coefficients of the 'controlled object' take values in $\mathbb{R}^e$ rather than $\mathbb{R}$. In this case we have $\mathbf{Z} : [0, T] \to (\mathcal{H}_{N-1})^e$ where $(\mathcal{H}_{N-1})^e$ denotes the e-th cartesian power of $\mathcal{H}_{N-1}$. Hence, the coefficients $\langle h, \mathbf{Z} \rangle$ take values in $\mathbb{R}^e$ and we denote the i-th component by $\langle h, \mathbf{Z} \rangle_i$.

Let $\mathbf{Z}$ be an $\mathbf{X}$-controlled rough path above Z, then we can use $\mathbf{Z}$ to define the integral $\int Z dX^i$, for any $1 \leq i \leq d$. It is an easy exercise to check that the condition (3.24) ensures that

$$\tilde{Z}_{st} = \sum_{h \in \mathcal{F}^0_{N-1}} \langle h, \mathbf{Z}_s \rangle \langle \mathbf{X}_{st}, [h]_i \rangle$$

satisfies (3.23). From the sewing lemma, it follows that there exists a unique remainder r with $|r_{st}| = o(|t-s|)$ such that $\tilde{Z}_{st} + r_{st}$ is the increment of a path. Naturally, this increment is chosen as a definition of the integral

$$\int_s^t Z_r dX^i_r \stackrel{\text{def}}{=} \tilde{Z}_{st} + r_{st} = \lim_{\pi \to 0} \sum_{[u,v] \in \pi} \tilde{Z}_{uv} \ , \tag{3.25}$$

for any partition π of $[s, t]$. Hence, we have defined a map which sends a controlled rough path $\mathbf{Z}$ to a path $\int Z dX^i$. This map can be extended to $\mathcal{I} : \mathbf{Z} \mapsto \int \mathbf{Z} dX^i$, where $\int \mathbf{Z} dX^i$ is a controlled rough path above $\int Z dX^i$. To define $\int \mathbf{Z} dX^i$, we simply specify $\langle x, \int \mathbf{Z} dX^i \rangle$ for all dual basis elements $x \in \mathcal{F}^*_N \cup \{1\}$. Firstly, we let $\langle 1, \int_0^t \mathbf{Z}_r dX^i_r \rangle$ be the unique (up to an additive constant) path satisfying

$$\left\langle 1, \int_0^t \mathbf{Z}_r dX^i_r \right\rangle - \left\langle 1, \int_0^s \mathbf{Z}_r dX^i_r \right\rangle = \int_s^t Z_r dX^i_r \ ,$$

and then define the coefficients by

$$\left\langle [\tau_1 \dots \tau_n]_i, \int_0^t \mathbf{Z}_r dX^i_r \right\rangle = \langle \tau_1 \dots \tau_n, \mathbf{Z}_t \rangle \ ,$$

for all $[\tau_1 \dots \tau_n]_i \in \mathcal{T}_{N-1}$ with i fixed and

$$\left\langle \tau_1 \dots \tau_n, \int_0^t \mathbf{Z}_t dX_r^i \right\rangle = 0 \quad \text{otherwise} .$$

More generally, if $\mathbf{Z} = (\mathbf{Z}^1, \dots, \mathbf{Z}^d)$ where each $\mathbf{Z}^i$ is an $\mathbf{X}$-controlled rough path above Z^i, then we can define an $\mathbf{X}$-controlled rough path $\int \mathbf{Z} \cdot dX$ above $\int Z \cdot dX$, where $Z = (Z^1, \dots, Z^d)$. To do this, we set

$$\left\langle 1, \int_s^t \mathbf{Z}_r \cdot dX_r \right\rangle = \sum_{i=1}^d \left\langle 1, \int_s^t \mathbf{Z}_r^i dX_r^i \right\rangle ,$$

with coefficients

$$\langle [\tau_1 \dots \tau_n]_i, \int_0^t \mathbf{Z}_r \cdot dX_r \rangle = \langle h, \mathbf{Z}_t^i \rangle$$

for all $[\tau_1 \dots \tau_n]_i \in \mathcal{T}_{N-1}$ and each $1 \le i \le d$ and

$$\langle \tau_1 \dots \tau_n, \int_0^t \mathbf{Z}_r \cdot dX_r \rangle = 0 \quad \text{otherwise} .$$

For verification that $\int \mathbf{Z} dX^i$ satisfies (3.24) and hence actually is a controlled rough path, see [Gub10, Theorem 8.5].

Remark 3.3.5. Since the definition of $\int Z dX^i$ depends on how we define a controlled rough path above Z, it makes more sense to use the controlled rough path notation $\int \mathbf{Z} dX^i$.

Not only are controlled rough paths stable under the integration map, but they are also stable under composition by smooth functions. We will demonstrate this for a controlled rough path $\mathbf{Z} : [0,T] \to (\mathcal{H}_{N-1})^e$ and a smooth function $\phi : \mathbb{R}^e \to \mathbb{R}^e$. We first introduce the notation

$$D^n \phi(u) : (v_1, \dots, v_n) = \sum_{\alpha_1, \dots, \alpha_n = 1}^{e} \partial^{\alpha_1} \dots \partial^{\alpha_n} \phi(u) v_1^{\alpha_1} \dots v_n^{\alpha_n} ,$$

where $u, v_i \in \mathbb{R}^e$, v_i^j denotes the j-th component of v_i. We define a controlled rough path $\phi(\mathbf{Z}) : [0,T] \to (\mathcal{H}_{N-1})^e$ above $\phi(Z)$ using a Taylor expansion. In particular, we have that

$$\phi(Z_t) - \phi(Z_s) = \sum_{n=1}^{N-1} \frac{1}{n!} D^n \phi(Z_s) : (\langle h_1, \mathbf{Z}_s \rangle, \dots, \langle h_n, \mathbf{Z}_s \rangle) \langle \mathbf{X}_{st}, h_1 \dots h_n \rangle + R_{st}^{\phi} , \tag{3.26}$$

where we sum over all $h_i \in \mathcal{F}$ with $|h_1| + \dots + |h_n| \le N - 1$ and $|R_{st}^{\phi}| \le C|t-s|^{N\gamma}$. It is clear that the controlled rough path $\phi(\mathbf{Z})$ should have $\langle 1, \phi(\mathbf{Z}_s)\rangle = \phi(Z_s)$ and coefficients

$$\langle h, \phi(\mathbf{Z})_s\rangle = \sum_{n=1}^{N-1} \sum_{h_1\dots h_n = h} \frac{1}{n!} D^n\phi(Z_s) : (\langle h_1, \mathbf{Z}_s\rangle, \dots, \langle h_n, \mathbf{Z}_s\rangle) ,$$

where we sum over all $h_1, \dots, h_n$ appearing in (3.26) such that $h_1 \dots h_n = h$. For verification that $\phi(\mathbf{Z})$ satisfies (3.24), see [Gub10].

Example 3.3.6. As an exercise, we will calculate $\int_s^t F(X)dX^i$, where $F : \mathbb{R}^d \to \mathbb{R}^d$ is a smooth function and X has a branched rough path $\mathbf{X}$ above it. Firstly, since $\mathbf{X}$ is clearly an $\mathbf{X}$-controlled rough path, we can define $F(\mathbf{X})$. We set $\langle 1, F(\mathbf{X}_t)\rangle = F(X_t)$ and

$$\langle \bullet_{\beta_1} \cdots \bullet_{\beta_m}, F(\mathbf{X}_t)\rangle = \partial^{\beta_1} \dots \partial^{\beta_m} F(X_t) ,$$

for all $\bullet_{\beta_1} \cdots \bullet_{\beta_m} \in \mathcal{F}_{N-1}$ and $\langle h, F(\mathbf{X}_t)\rangle = 0$ otherwise. We then have

$$\begin{aligned}\int_s^t F(X_r)dX_r^i &= F(X_s)\langle \mathbf{X}_{st}, \bullet_i\rangle \\ &\quad + \sum_{m=1}^{N-1} \sum_{\bullet_{\beta_1}\cdots\bullet_{\beta_m}\in\mathcal{F}_{N-1}} \langle \bullet_{\beta_1} \cdots \bullet_{\beta_m}, F(\mathbf{X}_s)\rangle \, \langle \mathbf{X}_{st}, [\bullet_{\beta_1} \cdots \bullet_{\beta_m}]_i\rangle + o(|t-s|) \\ &= \sum_{m=0}^{N} \sum_{\beta_1,\dots,\beta_m=1}^{d} \frac{\partial^{\beta_1} \dots \partial^{\beta_m} F(X_s)}{m!} \langle \mathbf{X}_{st}, [\bullet_{\beta_1} \cdots \bullet_{\beta_m}]_i\rangle + o(|t-s|) ,\end{aligned}$$

where in the last line we have used the symmetry of the expression to replace $\sum_{\bullet_{\beta_1}\cdots\bullet_{\beta_m}\in\mathcal{F}_N}$, the unordered sum, with $\sum_{\beta_1,\dots,\beta_m=1}^{d} 1/m!$.

The set of $\mathbf{X}$-controlled rough paths is easily seen to be a vector space. One can turn it into a Banach space, denoted $\mathcal{Q}_{\mathbf{X}}(\mathbb{R}^e)$, by introducing the norm

$$\|\mathbf{Z}\|_{\mathcal{Q}_{\mathbf{X}}(\mathbb{R}^e)} = |\mathbf{Z}_0| + \sum_{h\in\mathcal{F}_{N-1}} \|R^h\|_{(N-|h|)\gamma} ,$$

where $\|f\|_{(N-|h|)\gamma} = \sup_{s\neq t} \frac{|f_{st}|}{|t-s|^{(N-|h|)\gamma}}$. The Banach space $\mathcal{Q}_{\mathbf{X}}(\mathbb{R}^e)$ turns out to be the right environment in which to solves RDEs.

3.3.7 Solving Rough DEs

The foremost example of a controlled rough path is the solution to an RDE. We will consider the equation

$$\delta Y_{st} = \int_s^t f(Y_r) \cdot dX_r \quad \text{with} \quad Y_0 = \xi , \tag{3.27}$$

for every $s, t \in [0, T]$, where $X = (X^i) \in \mathbb{R}^d$ and $f(Y) \cdot dX = \sum_{i=1}^d f_i(Y) dX^i$. The vector fields $f_i : \mathbb{R}^e \to \mathbb{R}^e$ are assumed to be as smooth as required. To solve this RDE, we must specify a branched rough path $\mathbf{X}$ above X. In [Gub10], solutions to (3.27) are defined by lifting the problem to the space of $\mathbf{X}$-controlled rough paths.

Definition 3.3.8. A path $Y : [0, T] \to \mathbb{R}^e$ with $Y_0 = \xi$ is a solution to (3.27) if and only if there exists an $\mathbf{X}$-controlled rough path $\mathbf{Y}$ above Y satisfying

$$\mathbf{Y}_t - \mathbf{Y}_s = \int_s^t f(\mathbf{Y}_r) \cdot dX_r \tag{3.28}$$

for every $s, t \in [0, T]$.

One can define the fixed point map $\mathcal{M} : \mathcal{Q}_{\mathbf{X}}(\mathbb{R}^e) \to \mathcal{Q}_{\mathbf{X}}(\mathbb{R}^e)$ by $(\mathcal{M}\mathbf{Y})_t = \int_0^t f(\mathbf{Y}_r) \cdot dX_r$. Since $\mathcal{Q}_{\mathbf{X}}(\mathbb{R}^e)$ is a Banach space, we can apply standard fixed point arguments on $\mathcal{M}$ to obtain existence and uniqueness results for (3.28). In particular, if the vector fields f_i have N continuous and bounded derivatives, then global solutions exist for any initial condition. Moreover, if the vector fields have $N + 1$ continuous and bounded derivatives then the solution is unique [Gub10, Theorem 8.8]. Throughout the sequel, we will always assume the vector fields are smooth enough to guarantee existence and uniqueness of solutions.

In this thesis we are more concerned with the structure of RDEs, and would like an explicit representation of the controlled rough path solution to (3.28). In particular, it is easy to see that a controlled rough path $\mathbf{Y}$ is a solution if and only if

$$\delta Y_{st} = \left\langle 1, \int_s^t f(\mathbf{Y}_r) \cdot dX_r \right\rangle$$

and

$$\langle [\tau_1 \dots \tau_n]_i, \mathbf{Y}_s \rangle = \left\langle [\tau_1 \dots \tau_n]_i, \int_0^s f(\mathbf{Y}_r) \cdot dX_r \right\rangle ,$$

for all $[\tau_1 \dots \tau_n]_i \in \mathcal{T}^*_{N-1}$ with $i = 1 \dots d$ and

$$\langle \tau_1 \dots \tau_m, \mathbf{Y}_s \rangle = 0 ,$$

for all non-trivial products $\tau_1 \dots \tau_m \in \mathcal{F}^\star_N \setminus \mathcal{T}^\star_N$. Using the definition of $\int f(\mathbf{Y}) \cdot dX$, we can refine the condition on the coefficients to $\langle \bullet_i, \mathbf{Y}_s \rangle = f_i(Y_s)$ and

$$\begin{aligned}
\langle [\tau_1 \dots \tau_n]_i, \mathbf{Y}_s \rangle &= \langle \tau_1 \dots \tau_n, f_i(\mathbf{Y}_s) \rangle \\
&= \sum_{\sigma \in \mathrm{Sym}(n)} \frac{D^n f_i(Y_s)}{n!} : \big(\langle \tau_{\sigma(1)}, \mathbf{Y}_s \rangle, \dots, \langle \tau_{\sigma(n)}, \mathbf{Y}_s \rangle \big) \\
&= D^n f_i(Y_s) : \big(\langle \tau_1, \mathbf{Y}_s \rangle, \dots, \langle \tau_n, \mathbf{Y}_s \rangle \big) \ ,
\end{aligned} \tag{3.29}$$

where we sum over all permutations σ of $\{1, \dots, n\}$. It follows that we can always write the coefficients as $\langle \tau, \mathbf{Y}_t \rangle = f_\tau(Y_t)$ where $f_\tau : \mathbb{R}^e \to \mathbb{R}^e$ is some smooth function determined by f and its derivatives. For instance,

$$\langle [\bullet_j \bullet_k]_i, \mathbf{Y}_s \rangle = D^2 f_i(Y_s) : \big(\langle \bullet_j, \mathbf{Y}_s \rangle, \langle \bullet_k, \mathbf{Y}_s \rangle \big) = D^2 f_i : \big(f_j, f_k \big)(Y_s) \ .$$

In the sequel, we will always reserve $\{f_\tau\}_{\tau \in \mathcal{T}^*}$ for the family of functions satisfying the recurrence

$$f_{[\tau_1 \dots \tau_n]_i} = D^n f_i : (f_{\tau_1}, \dots, f_{\tau_n}) \ , \tag{3.30}$$

for some specified $\{f_{\bullet_i}\}_{i=1\dots d}$. We also extend the family to any $h \in \mathcal{H}^*$ by

$$f_h \stackrel{\text{def}}{=} \langle h, 1 \rangle \mathrm{Id} + \sum_{\tau \in \mathcal{T}} \langle h, \tau \rangle f_\tau \ , \tag{3.31}$$

where $\mathrm{Id} : \mathbb{R}^e \to \mathbb{R}^e$ is the identity map. It also follows that

$$f_{\tau_1 \dots \tau_n} = 0 \ ,$$

for all non-trivial products $\tau_1 \dots \tau_n \in \mathcal{F}^*_N \setminus \mathcal{T}^*_N$. Moreover, if $f_{\bullet_i} = f_i$, then we have that $f_h(Y_t) = \langle h, \mathbf{Y}_t \rangle$ for all $h \in \mathcal{H}^*_N$.

Remark 3.3.9. There is a conflict of notation here, if $h = 1$ is the co-unit then from (3.31) we have $f_1 = \mathrm{Id}$, so that $f_1(Y_t) = \langle 1, \mathbf{Y}_t \rangle$. This is not to be confused with the vector field f_1 in the original RDE. Since never actually refer to f_1 for the co-unit 1, the reader should not be concerned. Indeed, it is simply included to make the definition (3.31) consistent.

By exploiting some algebraic properties of the coefficients f_τ, we can obtain an explicit formula for $\langle 1, \mathbf{Y} \rangle = Y$. In the following proposition we define a controlled rough path $\mathbf{Y} : [0, T] \to (\mathcal{H}_N)^e$, with an extra layer of components $\langle \tau_1 \dots \tau_n, \mathbf{Y} \rangle$ for $|\tau_1 \dots \tau_n| = N$, these extra components serve no purpose other than to facilitate the definition of $\langle 1, \mathbf{Y} \rangle$. It is not hard to see that these extra components become important when considering the fixed point equation $\mathbf{Y} = \mathcal{M}\mathbf{Y}$.

Proposition 3.3.10. $\mathbf{Y} : [0,T] \to (\mathcal{H}_N)^e$ *with* $\langle 1, \mathbf{Y} \rangle = Y$ *is the unique controlled rough path solution to* (3.28) *if and only if*

$$\delta Y_{st} = \sum_{\tau \in \mathcal{T}_N} f_\tau(Y_s)\langle \mathbf{X}_{st}, \tau \rangle + r_{st} \tag{3.32}$$

where $|r_{st}| = o(|t-s|)$ *and the coefficients of* $\mathbf{Y}$ *are given by* $\langle \tau_1 \dots \tau_n, \mathbf{Y}_t \rangle = f_{\tau_1 \dots \tau_n}(Y_t)$ *for all* $\tau_1 \dots \tau_n \in \mathcal{F}_N^*$, *with* $f_{\bullet_i} = f_i$ *for* $i = 1 \dots d$.

In order to show that $\mathbf{Y}$ constructed by (3.32) with $\langle \tau_1 \dots \tau_n, \mathbf{Y}_t \rangle = f_{\tau_1 \dots \tau_n}(Y_t)$ is a solution, we must first show that it is a controlled rough path.

Lemma 3.3.11. *Suppose* $\mathbf{Y} : [0,T] \to (\mathcal{H}_N)^e$ *with* $\langle 1, \mathbf{Y} \rangle = Y$ *satisfies* (3.32) *and* $\langle \tau_1 \dots \tau_n, \mathbf{Y}_t \rangle = f_{\tau_1 \dots \tau_n}(Y_t)$ *for all* $\tau_1 \dots \tau_n \in \mathcal{F}_N^*$ *with* $f_{\bullet_i} = f_i$. *Then* $\mathbf{Y}$ *is an* $\mathbf{X}$-*controlled rough path.*

Proof of Lemma 3.3.11. We must check the consistency condition (3.24) to ensure that $\mathbf{Y}$ is a controlled rough path. The assumption (3.32) ensures the condition holds for $h = 1$, so it is sufficient to prove the condition for all $\tau \in \mathcal{T}_N$, since the coefficients vanish on non-trivial products. We will assume the consistency condition holds for all of $\mathcal{T}_k$ and prove the condition for $\tau = [\tau_1 \dots \tau_n]_i$ where $n \geq 0$ and $\tau_i \in \mathcal{T}_k$. We have that

$$\begin{aligned}
&\langle [\tau_1 \dots \tau_n]_i, \mathbf{Y}_t \rangle - \langle \mathbf{X}_{st} \star [\tau_1 \dots \tau_n]_i, \mathbf{Y}_s \rangle \\
&\qquad = D^n f_i(Y_t) : \big(\langle \tau_1, \mathbf{Y}_t \rangle, \dots, \langle \tau_n, \mathbf{Y}_t \rangle\big) - \sum_{\rho \in \mathcal{F}_N^0} \langle \mathbf{Y}_s, \rho \rangle \langle \mathbf{X}_{st} \tilde{\otimes} [\tau_1 \dots \tau_n]_i, \Delta\rho \rangle \\
&\qquad = D^n f_i : \big(f_{\tau_1}, \dots, f_{\tau_n}\big)(Y_t) - \sum_{\rho \in \mathcal{T}_N^0} f_\rho(Y_s) \langle \mathbf{X}_{st} \tilde{\otimes} [\tau_1 \dots \tau_n]_i, \Delta\rho \rangle \ .
\end{aligned}$$

Now, by a Taylor expansion on $D^n f_i$, we obtain

$$\begin{aligned}
&D^n f_i(Y_t) : \big(f_{\tau_1}, \dots, f_{\tau_n}\big)(Y_t) \\
&\qquad = \sum_{m=n}^{N} \frac{1}{(m-n)!} D^m f_i(Y_s) : \big(f_{\tau_1}(Y_t), \dots, f_{\tau_n}(Y_t), \delta Y_{st}, \dots, \delta Y_{st}\big) + R_{st}^f \ ,
\end{aligned} \tag{3.33}$$

where the term δY_{st} appears $m - n$ times and

$$|R_{st}^f| \leq C|\delta Y_{st}|^{N-m} \leq C|t-s|^{(N-n)\gamma} \ .$$

Now, by the inductive hypothesis we have that

$$f_{\tau_j}(Y_t) = \sum_{\sigma_j \in \mathcal{T}_N^0} f_{\sigma_j}(Y_s) \langle \mathbf{X}_{st} \star \tau_j, \sigma_j \rangle + R_{st}^{\tau_j} \ ,$$

where $|R^{\tau_j}_{st}| \le C|t-s|^{(N-|\tau_j|)\gamma}$ and by assumption we have that

$$\delta Y_{st} = \sum_{\lambda_j \in \mathcal{T}_N} f_{\lambda_j}(Y_s)\langle \mathbf{X}_{st}, \lambda_j\rangle + r^j_{st}\;,$$

where $|r^j_{st}| = o(|t-s|)$. If we substitute these into (3.33), we obtain

$$\begin{aligned}\sum_{m=n}^{N} \frac{1}{(m-n)!} D^m f_i(Y_s) : \big(f_{\sigma_1}(Y_s),\ldots,f_{\sigma_n}(Y_s), f_{\lambda_1}(Y_s),\ldots,f_{\lambda_{m-n}}(Y_s)\big)\\ \times \langle \mathbf{X}_{st}\star\tau_1,\sigma_1\rangle\ldots\langle \mathbf{X}_{st}\star\tau_n,\sigma_n\rangle\langle \mathbf{X}_{st},\lambda_1\ldots\lambda_{m-n}\rangle + R^\tau_{st}\;,\end{aligned} \tag{3.34}$$

where we sum over all $\sigma_j \in \mathcal{T}_N$ (since $\sigma_j = 1$ vanishes) and $\lambda_j \in \mathcal{T}_N$ and where R^τ_{st} is the sum of all terms that contain at least one factor from the set $\{R^{\tau_1}_{st},\ldots,R^{\tau_n}_{st}, R^f_{st}, r^1_{st},\ldots,r^{m-n}_{st}\}$. Hence,

$$|R^\tau_{st}| \le C \max_{1\le j\le n}\Big(|t-s|^{(N-|\tau_j|)\gamma}\Big) + C|t-s|^{(N-n)\gamma} \le C|t-s|^{(N-|[\tau_1\ldots\tau_n]_i|)\gamma}\;, \tag{3.35}$$

where the bound on the second term follows from the fact that $n \le |\tau_1| + \cdots + |\tau_n| \le |[\tau_1\ldots\tau_n]_i|$. On the other hand, we have that

$$\begin{aligned}&\sum_{\rho\in\mathcal{T}^0_N} f_\rho(Y_s)\langle \mathbf{X}_{st}\tilde{\otimes}[\tau_1\ldots\tau_n]_i, \Delta\rho\rangle\\ &\qquad= \sum_{m=n}^{N-1}\sum_{\rho_1,\ldots,\rho_m}\frac{1}{m!} f_{[\rho_1\ldots\rho_m]_i}(Y_s)\langle \mathbf{X}_{st}\tilde{\otimes}[\tau_1\ldots\tau_n]_i, \Delta[\rho_1\ldots\rho_m]_i\rangle\end{aligned}$$

where we sum over $\rho_i \in \mathcal{T}_N$ with $|\rho_1| + \cdots + |\rho_m| \le N-1$, since only those $\rho \in \mathcal{T}$ with $\rho = [\rho_1\ldots\rho_m]_i$ for $m \ge n$ will not vanish. Note that the factor of $1/m!$ appears since all rearrangements of the ρ_i in $[\rho_1\ldots\rho_m]$ produce the same ρ. Using the recurrence (3.30), this expands to

$$\sum_{m=n}^{N-1}\sum_{\rho_1,\ldots,\rho_m}\frac{1}{m!} D^m f_i(Y_s) : \big(f_{\rho_1}(Y_s),\ldots,f_{\rho_m}(Y_s)\big)\langle \mathbf{X}_{st}\tilde{\otimes}[\tau_1\ldots\tau_n]_i, \Delta[\rho_1\ldots\rho_m]_i\rangle\;.$$

But we also have that

$$\langle \mathbf{X}_{st}\tilde{\otimes}[\tau_1\ldots\tau_n]_i, \Delta[\rho_1\ldots\rho_m]_i\rangle = \langle \mathbf{X}_{st}, \rho^{(1)}_1\ldots\rho^{(1)}_m\rangle\langle \tau_1\ldots\tau_n, \rho^{(2)}_1\ldots\rho^{(2)}_m\rangle\;.$$

It follows that

$$
\begin{aligned}
&\sum_{\rho_1,\dots,\rho_m} \frac{1}{m!} D^m f_i(Y_s) : \big(f_{\rho_1}(Y_s),\dots,f_{\rho_m}(Y_s)\big) \langle \mathbf{X}_{st} \tilde{\otimes} [\tau_1 \dots \tau_n]_i, \Delta[\rho_1 \dots \rho_m]_i \rangle \\
&= \sum_{\rho_1,\dots,\rho_m} \frac{1}{m!} D^m f_i(Y_s) : \big(f_{\rho_1}(Y_s),\dots,f_{\rho_m}(Y_s)\big) \\
&\qquad\qquad \times \langle \mathbf{X}_{st}, \rho_1^{(1)} \dots \rho_m^{(1)} \rangle \langle \tau_1 \dots \tau_n, \rho_1^{(2)} \dots \rho_m^{(2)} \rangle \\
&= \sum_{\rho_1,\dots,\rho_m} \frac{1}{m!} D^m f_i(Y_s) : \big(f_{\rho_1}(Y_s),\dots,f_{\rho_m}(Y_s)\big) \\
&\qquad\qquad \times \binom{m}{n} \langle \mathbf{X}_{st}, \rho_1^{(1)} \dots \rho_n^{(1)} \rho_{n+1} \dots \rho_m \rangle \langle \tau_1 \dots \tau_n, \rho_1^{(2)} \dots \rho_n^{(2)} \rangle \ .
\end{aligned} \tag{3.36}
$$

In the last equality we have used the fact that each term in $\langle \tau_1 \dots \tau_n, \rho_1^{(2)} \dots \rho_m^{(2)} \rangle$ will vanish unless $\rho_j^{(2)} = 1$ for some choice of $m-n$ terms from the product $\rho_1^{(2)} \dots \rho_m^{(2)}$. But since the function $D^m f_i(Y_s) : \big(f_{\rho_1},\dots,f_{\rho_m}\big)$ is symmetric in $\rho_1,\dots,\rho_m$ and we are summing over all $\rho_1,\dots,\rho_m$, we can assume without loss of generality that $\rho_j^{(2)} = 1$ for $n+1 \leq j \leq m$, provided we include the combinatorial factor $\binom{m}{n}$. Of course, it follows that for each $n+1 \leq j \leq m$, the only term remaining from the sum $\rho_j^{(1)} \tilde{\otimes} \rho_j^{(2)}$ is $\rho_j \tilde{\otimes} 1$.

Now, since $\langle \tau_1 \dots \tau_n, \rho_1^{(2)} \dots \rho_n^{(2)} \rangle = 1$ if and only if $\langle \tau_j, \rho_{i_j}^{(2)} \rangle = 1$ for any permutation $(i_1,\dots,i_n)$ of $\{1,\dots,n\}$ and every $j = 1 \dots n$, we can write (3.36) as

$$
\begin{aligned}
&\sum_{\rho_1,\dots,\rho_m} \frac{1}{m!} D^m f_i : \big(f_{\rho_1},\dots,f_{\rho_m}\big)(Y_s) \\
&\qquad \times \binom{m}{n} \langle \mathbf{X}_{st}, \rho_1^{(1)} \dots \rho_n^{(1)} \rho_{n+1} \dots \rho_m \rangle \sum_{(i_1,\dots,i_n)} \langle \tau_1, \rho_{i_1}^{(2)} \rangle \dots \langle \tau_n, \rho_{i_n}^{(2)} \rangle \\
&= \sum_{\rho_1,\dots,\rho_m} \frac{n!}{m!} D^m f_i : \big(f_{\rho_1},\dots,f_{\rho_m}(Y_s)\big) \\
&\qquad \times \binom{m}{n} \langle \mathbf{X}_{st}, \rho_1^{(1)} \dots \rho_n^{(1)} \rho_{n+1} \dots \rho_m \rangle \langle \tau_1, \rho_1^{(2)} \rangle \dots \langle \tau_n, \rho_n^{(2)} \rangle \ ,
\end{aligned} \tag{3.37}
$$

where in the first line we sum over all permutations $(i_1,\dots,i_n)$. If we set $(\rho_1,\dots,\rho_m) = (\sigma_1,\dots,\sigma_n,\lambda_1,\dots,\lambda_{m-n})$, then by comparing (3.34) with (3.37) and using the fact that $\frac{n!}{m!}\binom{m}{n} = \frac{1}{(m-n)!}$ we see that

$$
\langle [\tau_1 \dots \tau_n]_i, \mathbf{Y}_t \rangle - \langle \mathbf{X}_{st} \star [\tau_1 \dots \tau_n]_i, \mathbf{Y}_s \rangle = R_{st}^{\tau} \ ,
$$

and the estimate (3.35) proves the result. □

Proof of Proposition 3.3.10. We will first prove the 'if' statement. Define $\mathbf{Y}$ as in Lemma 3.3.11, then $\mathbf{Y}$ is indeed an $\mathbf{X}$-controlled rough path. Now, from (3.29), we have that $\langle \tau, \mathbf{Y}_t \rangle = \langle \tau, (\mathcal{M}\mathbf{Y})_t \rangle$ for all $\tau \in \mathcal{T}_N^*$, so to show $\mathbf{Y}$ is the unique solution, it suffices to show that

$$\delta Y_{st} = \left\langle 1, \int_s^t f(\mathbf{Y}_r) \cdot dX_r \right\rangle .$$

By the definition, we have that

$$\begin{aligned} &\left\langle 1, \int_s^t f(\mathbf{Y}_r) \cdot dX_r \right\rangle \qquad (3.38) \\ &= \sum_{i=1}^d \sum_{\tau_1 \dots \tau_n \in \mathcal{F}_{N-1}} D^n f_i(Y_s) : (\langle \tau_1, \mathbf{Y}_s \rangle, \dots, \langle \tau_n, \mathbf{Y}_s \rangle) \langle \mathbf{X}_{st}, [\tau_1 \dots \tau_n]_i \rangle + \tilde{r}_{st} \\ &= \sum_{i=1}^d \sum_{\tau_1 \dots \tau_n \in \mathcal{F}_{N-1}} D^n f_i : (f_{\tau_1}, \dots, f_{\tau_n})(Y_s) \langle \mathbf{X}_{st}, [\tau_1 \dots \tau_n]_i \rangle + \tilde{r}_{st} \\ &= \sum_{i=1}^d \sum_{\tau_1 \dots \tau_n \in \mathcal{F}_{N-1}} f_{[\tau_1 \dots \tau_n]_i}(Y_s) \langle \mathbf{X}_{st}, [\tau_1 \dots \tau_n]_i \rangle + \tilde{r}_{st} \ , \end{aligned}$$

where $|\tilde{r}_{st}| = o(|t-s|)$. However, by setting $\tau = [\tau_1 \dots \tau_n]_i$, we can rewrite the sum

$$\sum_{\tau \in \mathcal{T}_N} = \sum_{i=1}^d \sum_{\tau_1 \dots \tau_n \in \mathcal{F}_{N-1}} .$$

We obtain

$$\left\langle 1, \int_s^t f(\mathbf{Y}_r) \cdot dX_r \right\rangle = \sum_{\tau \in \mathcal{T}_N} f_\tau(Y_s) \langle \mathbf{X}_{st}, \tau \rangle + \tilde{r}_{st} \ . \qquad (3.39)$$

But from (3.32), it follows that

$$\left\langle 1, \int_s^t f(\mathbf{Y}_r) \cdot dX_r \right\rangle - \delta Y_{st} = \tilde{r}_{st} - r_{st} \ ,$$

and since the left hand side is an increment and the right hand side is $o(|t-s|)$, the left hand side must be identically zero.

For the 'only if' statement, suppose $\mathbf{Y}$ is the unique solution to (3.28) with $\langle 1, \mathbf{Y} \rangle = Y$. Then since $\langle [\tau_1 \dots \tau_n]_i, \mathbf{Y}_t \rangle = \langle [\tau_1 \dots \tau_n]_i, (\mathcal{M}\mathbf{Y})_t \rangle$, it follows from (3.29) (and the preceding argument) that $\langle \tau_1 \dots \tau_n, \mathbf{Y}_t \rangle = f_{\tau_1 \dots \tau_n}(Y_t)$ for all $\tau_1 \dots \tau_n \in \mathcal{F}_N^*$ with $f_{\bullet_i} = f_i$ for $i = 1 \dots d$. Note that (3.39) also holds for the solution $\mathbf{Y}$, since the identity only relies on the coefficients $\langle\tau, \mathbf{Y}\rangle$ satisfying (3.30). It follows that

$$\begin{aligned}\delta Y_{st} &= \left\langle 1, \int_s^t f(\mathbf{Y}_r) \cdot dX_r \right\rangle \\ &= \sum_{\tau\in\mathcal{T}_N} f_\tau(Y_s)\langle \mathbf{X}_{st}, \tau\rangle + \tilde{r}_{st} \ ,\end{aligned}$$

This proves (3.32) and hence completes the proof. □

Example 3.3.12. Let us consider the RDE with linear vector fields,

$$\delta Y_{st} = \sum_{i=1}^{d} \int_s^t V_i Y_r dX_r^i \ , \tag{3.40}$$

where $V_i \in L(\mathbb{R}^e, \mathbb{R}^e)$. Since the vector fields are smooth, the solution Y must take the form (3.32), where $Y = \langle 1, \mathbf{Y}\rangle$ and the coefficients satisfy $\langle [\tau]_i, \mathbf{Y}_s\rangle = V_i\langle\tau, \mathbf{Y}_s\rangle$ for any $[\tau]_i \in \mathcal{T}_N$ and $\langle \tau_1 \dots \tau_m, \mathbf{Y}_s\rangle = 0$ for any non-trivial product of $\tau_i \in \mathcal{T}$. Hence, we have that

$$\delta Y_{st} = \sum_{n=1}^{N}\sum_{a}(V_{a_n}\dots V_{a_1}Y_s)\langle \mathbf{X}_{st}, \tau^{a_n\dots a_1}\rangle + r_{st} \ ,$$

where we sum over all vectors $a = (a_1, \dots, a_n) \in \{1, \dots, d\}^n$ and we use the shorthand for linear trees $\tau^{a_n\dots a_1} = [\tau^{a_n\dots a_2}]_{a_1}$. If we use the injection map $\iota : T(\mathbb{R}^d) \to \mathcal{H}$, then we have

$$\delta Y_{st} = \sum_{n=1}^{N}\sum_{a}(V_{a_n}\dots V_{a_1}Y_s)\langle \mathbf{X}_{st}, \iota\big(e_{a_n}\otimes\dots\otimes e_{a_1}\big)\rangle + r_{st} \ ,$$

which coincides with the standard Davie solution [Dav07, FV10b], defined in the case of a geometric rough path. In our case, the "branched" components only influence the solution through terms involving second order derivatives of the vector field, which always vanish.

Chapter 4

Itô-Stratonovich correction for non-geometric rough paths

4.1 Introduction

Now that we have been introduced to branched and controlled rough paths, we can start to think about Itô corrections in a branched rough path framework. The main objective of this chapter is to prove an Itô-Stratontovich correction formula for controlled rough path solutions to RDEs. In order to obtain this formula, we must be able to *translate* between geometric and branched rough paths.

It is easy to show that a geometric rough path is a type of branched rough path, and moreover obtain a simple test to determine when a branched rough path over a path X is geometric over that same path X. The main result of the chapter provides a surprising converse statement, namely that every branched rough path over a path X can be rewritten as a geometric rough path over an extended path $\bar{X}$. The path $\bar{X}$ will take values in $\mathcal{B}_N$, where we define $\mathcal{B}_n$ as the real vector space spanned by the set $\mathcal{T}_n$. One can think of X as taking values in $\mathcal{B}_1 \cong \mathbb{R}^d$, under this interpretation, $\bar{X}$ is an extension of X in the sense that $\pi_{\mathcal{B}_1}(\bar{X}) = X$, where π_V denotes projection onto V. The geometric rough path $\bar{\mathbf{X}}$ lives in the truncated tensor product space

$$T^{(N)}(\mathcal{B}_N) = \operatorname{span}\{\tau_1 \otimes \cdots \otimes \tau_n : \tau_i \in \mathcal{T}_N \text{ and } 1 \leq n \leq N\} \ .$$

Thus, since τ is a basis vector of the underlying vector space $\mathcal{B}_N$, the object $\langle \bar{\mathbf{X}}_{st}, \tau \rangle$ will actually denote a *path* component of $\bar{\mathbf{X}}$, in that

$$\langle \bar{\mathbf{X}}_{st}, \tau \rangle = \delta \bar{X}^{\tau}_{st} \ ,$$

for all $\tau \in \mathcal{T}_N$, as opposed to the original $\langle \mathbf{X}_{st}, \tau \rangle$ which must be interpreted as a integral component, indexed by the tree τ. Moreover, the component $\langle \bar{\mathbf{X}}_{st}, \tau_1 \otimes \cdots \otimes \tau_n \rangle$ must be interpreted as an iterated integral of the form $\int_s^t \ldots \int_s^{v_2} d\bar{X}^{\tau_1}_{v_1} \ldots d\bar{X}^{\tau_n}_{v_n}$. We will prove the following result. As always, $\gamma \in (0, 1)$ and N is the largest integer such that $N\gamma \leq 1$.

Theorem 4.1.1. *Let $X = (X^i)_{i=1\ldots d}$ be a path in $\mathbb{R}^d$ and $\mathbf{X}$ a γ-Hölder branched rough path in $\mathcal{H}$ such that $\langle \mathbf{X}_{st}, \bullet_i \rangle = \delta X^i_{st}$. Then there exists*

1. *a path $\bar{X} = (\bar{X}^{\tau})_{\tau \in \mathcal{T}_N}$ taking values in $\mathcal{B}_N$, with $\pi_{\mathcal{B}_1}(\bar{X}) = X$,*
2. *a γ-Hölder geometric rough path $\bar{\mathbf{X}}$ in $T^{(N)}(\mathcal{B}_N)$ satisfying $\langle \bar{\mathbf{X}}_{st}, \tau \rangle = \delta \bar{X}^{\tau}_{st}$ for each $\tau \in \mathcal{T}_N$ and*
3. *a graded morphism of Hopf algebras $\psi : \mathcal{H} \to T(\mathcal{B}_N)$,*

such that

$$\langle \mathbf{X}_{st}, h \rangle = \langle \bar{\mathbf{X}}_{st}, \psi(h) \rangle \ , \tag{4.1}$$

for every $h \in \mathcal{H}$.

Before adding a few remarks, we will illustrate the result with the first non-trivial example.

Example 4.1.2. Consider the case where $X \in \mathbb{R}^d$ with Hölder exponent $1/3 < \gamma \leq 1/2$, so that $N = 2$. The important components of the branched rough path $\mathbf{X}$ above X are $\langle \mathbf{X}, \bullet_i \rangle$ and $\langle \mathbf{X}, \overset{\bullet_k}{\bullet_j} \rangle$, for all $i, j, k = 1 \dots d$. The theorem tells us that there exists a path

$$\bar{X} = (\bar{X}^{\bullet_i}, \bar{X}^{\overset{\bullet_k}{\bullet_j}})_{i,j,k=1\dots d}$$

where $\bar{X}^{\bullet_i} = X^i$ for all $i = 1 \dots d$ and moreover there exists a geometric rough path $\bar{\mathbf{X}}$ above $\bar{X}$. Since $\mathcal{B}_2$ is the real vector space spanned by the elements $\{\bullet_i, \overset{\bullet_k}{\bullet_j}\}_{i,j,k=1\dots d}$, we can see that $\bar{\mathbf{X}}$ is defined on the (truncated) tenor product space $\mathcal{B}_2 \oplus \mathcal{B}_2^{\otimes 2}$. The map ψ tells us how to write $\mathbf{X}$ in terms of $\bar{\mathbf{X}}$, for instance we have $\psi(\bullet_i) = \bullet_i$ and $\psi(\overset{\bullet_j}{\bullet_i}) = \bullet_j \otimes \bullet_i + \overset{\bullet_j}{\bullet_i}$ and therefore

$$\langle \mathbf{X}_{st}, \bullet_i \rangle = \langle \bar{\mathbf{X}}_{st}, \bullet_i \rangle \quad \text{and} \quad \langle \mathbf{X}_{st}, \overset{\bullet_j}{\bullet_i} \rangle = \langle \bar{\mathbf{X}}_{st}, \bullet_j \otimes \bullet_i + \overset{\bullet_j}{\bullet_i} \rangle \ ,$$

or in the more formal language

$$\delta X^i_{st} = \delta \bar{X}^{\bullet_i}_{st} \quad \text{and} \quad \int_s^t \int_s^{v_1} dX^j_{v_2} dX^i_{v_1} = \int_s^t \int_s^{v_1} \circ d\bar{X}^{\bullet_j}_{v_2} \circ d\bar{X}^{\bullet_i}_{v_1} + \delta \bar{X}^{\overset{\bullet_j}{\bullet_i}}_{st} \ ,$$

for all $i, j, k = 1 \dots d$. We use the $\circ d\bar{X}$ notation to remind the reader that this is a geometric (but not necessarily Stratonovich) integral.

This result relies on the Lyons-Victoir extension theorem of [LV07], which shows that every γ-Hölder path in a quotient of the free nilpotent group $G^{(N)}(V)$ can be extended to a γ-Hölder path in $G^{(N)}(V)$. This extension theorem relies on the axiom of choice and is therefore not constructive, hence we cannot read off the components of $\bar{\mathbf{X}}$, we only know that such an $\bar{\mathbf{X}}$ exists. Since the extension theorem of [LCL07] is non-unique, the path $\bar{\mathbf{X}}$ is also non-unique. Moreover, there is a great deal of redundancy in $\bar{\mathbf{X}}$, since it has many more components than $\mathbf{X}$, however, this is the most convenient way to build a geometric rough path containing all the information of $\mathbf{X}$. The map ψ describes how the components of $\mathbf{X}$ should be divided into the tensor product algebra $T^{(N)}(\mathcal{B}_N)$. As we shall see, the fact that ψ is a Hopf algebra morphism is crucial not only when obtaining $\bar{\mathbf{X}}$, but also when applying (4.1) further down the line.

The main motivation behind Theorem 4.1.1 is that it allows us to rewrite an expression controlled by a branched rough path as an expression controlled by a geometric rough path. In particular, we can use this to show that every RDE driven by a branched rough path can be rewritten as another RDE driven by a geometric rough path.

Theorem 4.1.3 (Generalised Itô-Stratonovich correction)**.** *Let Y solve* (3.1)*, driven by a branched rough path $\mathbf{X}$. Let $\bar{X}$ and $\bar{\mathbf{X}}$ be as defined in Theorem 4.1.1. Then Y is also a solution to*

$$dY_t = \bar{f}(Y_t)d\bar{X}_t \,, \tag{4.2}$$

driven by the geometric rough path $\bar{\mathbf{X}}$, where $\bar{f}(Y)d\bar{X} = \sum_{\tau\in\mathcal{T}_N} f_\tau(Y)d\bar{X}^\tau$ and f_τ is defined by (3.30) *with $f_{\bullet_i} = f_i$.*

Example 4.1.4. Returning back to the $1/4 < \gamma \leq 1/3$ example, if Y solves (3.1) driven by some $\mathbf{X}$ then we also have

$$\begin{aligned} dY_t = {} & f_i(Y_t)d\bar{X}_t^{\bullet_i} + (f_i^\alpha\partial^\alpha f_j)(Y_t)d\bar{X}^{[\bullet_j]_i} + (f_k^\alpha\partial^\alpha f_j^\beta\partial^\beta f_i)(Y_t)d\bar{X}^{[[\bullet_k]_j]_i} \\ & + \frac{1}{2}(f_k^\alpha f_j^\beta\partial^\alpha\partial^\beta f_i)(Y_t)d\bar{X}^{[\bullet_k\bullet_j]_i} \,, \end{aligned}$$

driven by the geometric rough path $\bar{\mathbf{X}}$ found in Theorem 4.1.1, where we sum over all $i, j, k = 1 \dots d$ and $\alpha, \beta = 1 \dots e$, noting that $\bar{X}^{[\bullet_k\bullet_j]_i} = \bar{X}^{[\bullet_j\bullet_k]_i}$. Even though $\bar{X}^{\bullet_i} = X^i$, one must distinguish between $f_i(Y_t)dX_t^i$ and $f_i(Y_t)d\bar{X}_t^{\bullet_i}$, since the former is driven by $\mathbf{X}$ and the latter is driven by $\bar{\mathbf{X}}$.

Remark 4.1.5. Although we call this a generalised Itô-Stratonovich correction, it is really more like a "Any non-geometric integral"-"Particular class of geometric integral" correction. However, we are quite justified in giving it this name. Suppose X was a non semi-martingale path for which there exists a branched rough path $\mathbf{X}$ above it and also some kind of "Stratonovich" rough path $\bar{\mathbf{X}}^{(1)}$ above it, fractional Brownian motion with Hurst parameter $H > 1/4$ is a good example [CQ02]. As will be clear in the proof of Theorem 4.1.1, we can actually choose $\bar{\mathbf{X}}$ such that the components above X are given by $\bar{\mathbf{X}}^{(1)}$ (or indeed any geometric rough path above X). Hence, the formula can tell us what correction we get if we take an RDE driven by $\mathbf{X}$ and rewrite it using "Stratonovich" integrals, just as in the usual Itô-Stratonovich correction formula.

The outline of this chapter is as follows. In Section 4.2 we review the definition and properties of geometric rough paths. We subsequently show that every geometric rough path is a branched rough path and also the converse statement, Theorem 4.1.1. In Section 4.3 we first characterise solutions to RDEs driven by geometric rough paths, before proving the generalised Itô-Stratonovich correction formula.

4.2 Geometric rough paths

Let V be some real Banach space. Let $T(V) = \bigoplus_{i=0}^{\infty} V^{\otimes i}$ be the tensor product algebra of V, with the convention $V^{\otimes 0} = \mathbb{R}$. We will call $T^{(n)}(V) = \bigoplus_{i=0}^{n} V^{\otimes i}$ the *step-n truncated tensor algebra*. The vector space $T(V)$ can be viewed as a Hopf algebra, by adding the *shuffle* product $⧢$ and the *deconcatenation* coproduct $\bar{\Delta}$. The existence of an antipode for this bialgebra is guaranteed by Remark 3.2.9. The shuffle product is defined in the following way, let $e_a = e_{a_1} \otimes \cdots \otimes e_{a_n}$ and $e_b = e_{b_1} \otimes \cdots \otimes e_{b_m}$ then

$$e_a ⧢ e_b = \sum_{c \in \mathrm{Shuf}(a,b)} e_c \ ,$$

where $c \in \mathrm{Shuf}(a,b)$ if and only if c is a permutation of the index sequence $(a,b) = (a_1, \ldots, a_n, b_1, \ldots, b_m)$ which preserves the original ordering of the index sequences a and b respectively. The coproduct $\bar{\Delta}$ is defined by

$$\bar{\Delta} e_c = \sum_{(a,b)=c} e_a \tilde{\otimes} e_b \ .$$

The dual Hopf algebra $T((V))$ is the space of formal series of tensors, equipped with the concatenation product $\otimes$ and the coproduct $\bar{\delta}$, that are dual to $\bar{\Delta}$ and $⧢$ respectively. We likewise have $T((V)) = \bigoplus_{i=0}^{\infty} (V^*)^{\otimes i}$ and the truncation, $T^{(n)}((V))$ which can clearly be identified with $T^{(n)}(V)$. More details on the above construction can be found, for instance, in [Reu93].

We define a Lie bracket on $T((V))$ using the commutator

$$[x, y]_{\otimes} = x \otimes y - y \otimes x \ , \tag{4.3}$$

for any $x, y \in T((V))$. The *free Lie algebra* generated from V is then defined by

$$\mathcal{G}(V) = \bigoplus_{i=1}^{\infty} W_i(V) \ ,$$

where $W_1(V) = V$ and $W_{i+1}(V) = [V, W_i(V)]_{\otimes}$. We similarly denote the *step-n free Lie algebra* by

$$\mathcal{G}^{(n)}(V) = \bigoplus_{i=1}^{n} W_i(V) \ .$$

We define the *step-n free nilpotent group* $G^{(N)}(V)$ as the image of the step n free Lie algebra, under the exponential map

$$G^{(n)}(V) = \exp(\mathcal{G}^{(n)}(V)) ,$$

where $\exp$ denotes the tensor exponential

$$\exp(x) = \sum_{k\geq 0} \frac{x^{\otimes k}}{k!} .$$

It is well known that $G^{(n)}(V)$ coincides with the (truncated) group-like objects, in that $g \in G^{(n)}(V)$ if and only if $g \in T^{(n)}(V)$ and $\bar{\delta} g = g \tilde{\otimes} g$, the proof of this statement can be found in [Reu93]. Since $\bar{\delta}$ is dual to $\sqcup\!\sqcup$, this group-like property can be equivalently stated as

$$\langle g, x\rangle\langle g, y\rangle = \langle g, x \sqcup\!\sqcup y\rangle , \tag{4.4}$$

for every $g \in G^{(n)}(V)$ and $x, y \in T(V)$. The group $G^{(n)}(V)$ can be equipped with a *subadditive homogeneous norm* $\|\cdot\|_{G^{(n)}(V)} : G^{(n)}(V) \to [0, \infty)$ as defined in [LV07]. One can show that all such norms are equivalent, and in particular all such norms are equivalent to

$$\rho(x) = \sum_{k=1}^{n} \|\ell_k\|^{1/k} ,$$

where $x = \exp(\ell_1 + \cdots + \ell_n)$ with $\ell_i \in W_i(V)$ and where $\|\cdot\|$ denotes the Euclidean norm [LV07]. A path $\mathbf{X} : [0, T] \to G^{(n)}(V)$ is called γ-Hölder continuous if and only if

$$\sup_{s\neq t} \frac{\|\mathbf{X}_{st}\|_{G^{(n)}(V)}}{|t - s|^{\gamma}} < \infty , \tag{4.5}$$

where $\mathbf{X}_{st} = \mathbf{X}_s^{-1} \otimes \mathbf{X}_t$. If $n = N$, the largest integer such $N\gamma \leq 1$, then such a path is called a *weak geometric rough path* of Hölder exponent γ. By the equivalence of norms, the regularity condition (4.5) is synonymous with the statement

$$\sup_{s\neq t} \frac{|\langle \mathbf{X}_{st}, v_1 \otimes \cdots \otimes v_k\rangle|}{|t - s|^{k\gamma}} < \infty , \tag{4.6}$$

for any $v_i \in V$. Hence, the regularity condition for a geometric rough path is identical to that of a branched rough path.

Remark 4.2.1. There is a subtle difference between weak geometric rough paths and geometric rough paths [FV06]. In this thesis we only refer to the weak kind and will henceforth omit the prefix.

It turns out that every γ-Hölder path in a Banach space V can be extended to a path

$\bar{\mathbf{X}}$ taking values in $G^{(N)}(V)$. That is, every path has a weak geometric rough path lying above it. This is a particular case of the following theorem proved in [LV07]. If K is a normal subgroup of $G^{(N)}(V)$, we define the quotient homogeneous norm on the quotient group $G^{(N)}(V)/K$ by

$$\|g \otimes K\|_{G^{(N)}(V)/K} = \inf_{k \in K} \|g \otimes k\|_{G^{(N)}(V)} .$$

Theorem 4.2.2 (Lyons-Victoir extension)**.** *Let $\gamma \in (0,1)$ such that $\gamma^{-1} \notin \mathbb{N} \setminus \{0,1\}$. Suppose K is a normal subgroup of $G^{(N)}(V)$. If $\mathbf{X}$ is a γ-Hölder continuous path in the quotient $G^{(N)}(V)/K$, then there exists a γ-Hölder continuous path $\bar{\mathbf{X}}$ taking values in $G^{(N)}(V)$ and satisfying*

$$\pi_{G^{(N)}(V)/K}\left(\bar{\mathbf{X}}\right) = \mathbf{X} ,$$

where π denotes the projection map.

Remark 4.2.3. The restriction $\gamma^{-1} \notin \mathbb{N} \setminus \{0,1\}$ is a necessary one and a counter example can be found in [Vic04]. Hence, all our results in this chapter actually assume $\gamma \in (0,1)$ with $\gamma^{-1} \notin \mathbb{N}$.

Example 4.2.4. To give an idea of the type of situation in which this theorem applies, let $\mathbf{X}$ be a geometric rough path in $T^{(N)}(\mathbb{R}^d)$, lying above a path $X \in \mathbb{R}^d$ and suppose we would like to add a new path component X^{d+1} to X, by setting $\bar{X} = (X, X^{d+1})$. The extension theorem tells us that there exists a geometric rough path $\bar{\mathbf{X}}$ above $\bar{X}$ that agrees with $\mathbf{X}$ on the subspace $T^{(N)}(\mathbb{R}^d) \subset T^{(N)}(\mathbb{R}^{d+1})$. To be precise, we set

$$\widehat{\mathbf{X}}_t = \exp\left(\log \mathbf{X}_t + x_t^{d+1} e_{d+1}\right) .$$

This is an element in $T^{(N)}(\mathbb{R}^{d+1})$ and one can easily check that it is γ-Hölder in the quotient space $G^{(N)}(\mathbb{R}^{d+1})/K$, where $K = \exp L$ and L is the Lie ideal generated by

$$[e_{d+1}, \mathbb{R}^d]_\otimes = \operatorname{span}\{e_{d+1} \otimes e_j - e_j \otimes e_{d+1} : j = 1 \ldots d\}. \tag{4.7}$$

In particular, under the quotient norm we can effectively ignore all bracket terms involving e_{d+1}, and the γ-Hölder property then follows from the fact that $\mathbf{X}$ is γ-Hölder in $G^{(N)}(\mathbb{R}^d)$. Theorem 4.2.2 tells us that we can *add* the missing e_{d+1} components to obtain a geometric rough path $\bar{\mathbf{X}}$ on $T^{(N)}(\mathbb{R}^{d+1})$.

Remark 4.2.5. It is worth pointing out that, although the extension map $\mathbf{X} \mapsto \bar{\mathbf{X}}$ relies on the axiom of choice, it is always possible to choose a measurable version. This follows from an easy application of the measurable selection theorem [Wag77]. In a probabilistic setting, one would like to view $\bar{\mathbf{X}}$ as a random variable, so measurability is clearly a necessity.

4.2.6 Geometric rough paths are branched rough paths

It should be no surprise that a geometric rough path is a special kind of branched rough path. As mentioned in Remark 3.2.5 the tensor algebra $T(\mathbb{R}^d)$ can be identified with the subspace of $\mathcal{H}$ spanned by the linear trees. Given a geometric rough path $\bar{\mathbf{X}}$, the idea is to extend $\bar{\mathbf{X}}$ from this subspace of linear trees to a branched rough path $\mathbf{X}$ defined on the whole of $\mathcal{H}$. To perform this extension, we simply replace $\mathcal{H}$ products with $\sqcup\!\sqcup$ products. That is, we set

$$\langle \mathbf{X}_{st}, h\rangle = \langle \bar{\mathbf{X}}_{st}, \phi_g(h)\rangle \ , \tag{4.8}$$

for every $h \in \mathcal{H}_N$, where the map $\phi_g : \mathcal{H} \to T(\mathbb{R}^d)$ is defined by the rules $\phi_g(1) = 1$,

$$\phi_g([h]_i) = \phi_g(h) \otimes e_i \quad \text{and} \quad \phi_g(h_1 h_2) = \phi_g(h_1) \sqcup\!\sqcup \phi_g(h_2) \ ,$$

for $h, h_1, h_2 \in \mathcal{H}$ and e_i is the i-th canonical basis vector in $\mathbb{R}^d$. For example, we have

$$\phi_g(\bullet_a \bullet_b) = e_a \otimes e_b + e_b \otimes e_a \quad \text{and} \quad \phi_g([\bullet_a \bullet_b]_c) = e_a \otimes e_b \otimes e_c + e_b \otimes e_a \otimes e_c.$$

Proposition 4.2.7. *If $\bar{\mathbf{X}}$ is a γ-Hölder geometric rough path defined on $T^{(N)}(\mathbb{R}^d)$ and $\mathbf{X}$ is defined by* (4.8)*, then $\mathbf{X}$ is a γ-Hölder branched rough path on $\mathcal{H}$.*

This also provides a way to test the geometricity of a branched rough path. In particular, a branched rough path is geometric if and only if the identity

$$\langle \mathbf{X}_{st}, h\rangle = \langle \mathbf{X}_{st}, \iota\phi_g(h)\rangle \ ,$$

holds for every $h \in \mathcal{H}_N$, where $\iota : T(\mathbb{R}^d) \to \mathcal{H}$ is the inclusion map that identifies each tensor in $T(\mathbb{R}^d)$ with its corresponding linear tree in $\mathcal{H}$. Before proving the proposition, we need an important lemma. The map ϕ_g is clearly a morphism from $\cdot$ to $\sqcup\!\sqcup$. What is less clear is that it is also a morphism of coproducts Δ and $\bar{\Delta}$ and hence a Hopf algebra morphism. This is crucial in guaranteeing that $\mathbf{X}$ constructed above satisfies the right algebraic conditions. For the following, recall that $\mathcal{F}_{(n)}$ is all $\tau_1 \dots \tau_k \in \mathcal{F}$ with $|\tau_1| + \cdots + |\tau_k| = n$ and that $\mathcal{H}_{(n)}$ is the vector space spanned by $\mathcal{F}_{(n)}$ and also that $\mathcal{F}_n$ is all $\tau_1 \dots \tau_k \in \mathcal{F}$ with $|\tau_1| + \cdots + |\tau_k| \le n$.

Lemma 4.2.8. *We have that*

$$\bar{\Delta}\phi_g(h) = (\phi_g \tilde{\otimes} \phi_g)\Delta h \ , \tag{4.9}$$

for every $h \in \mathcal{H}$.

Proof. When applied to any $h \in \mathcal{H}_{(1)}$, the identity (4.9) is clear, so assume the claim holds on all $h \in \mathcal{H}_{(n)}$, we will prove that the claim holds for $\mathcal{F}_{(n+1)}$ and hence $\mathcal{H}_{(n+1)}$. If $h \in \mathcal{F}_{(n+1)}$, then $h = [h_1]_i$ for some $h_1 \in \mathcal{F}_{(n)}$ or $h = h_1 h_2$ for $h_1, h_2 \in \mathcal{F}_n$. In the first case,

$$\bar{\Delta}\phi_g([h_1]_i) = \bar{\Delta}\big(\phi_g(h_1) \otimes e_i\big) = \big(\phi_g(h_1) \otimes e_i\big)\tilde{\otimes}1 + (\bar{\Delta}\phi_g(h_1)) \otimes (1\tilde{\otimes}e_i)\ .$$

By the inductive assumption, $\bar{\Delta}\phi_g(h_1) = (\phi_g\tilde{\otimes}\phi_g)\Delta h_1$. If $\Delta h_1 = h_1^{(1)}\tilde{\otimes}h_1^{(2)}$, then we obtain

$$\begin{aligned}&\phi_g([h_1]_i)\tilde{\otimes}1 + (\phi_g\tilde{\otimes}\phi_g)(h_1^{(1)}\tilde{\otimes}h_1^{(2)}) \otimes (1\tilde{\otimes}e_i)\\ &\quad = (\phi_g\tilde{\otimes}\phi_g)\big([h_1]_i\tilde{\otimes}1 + h_1^{(1)}\tilde{\otimes}[h_1^{(2)}]_i\big) = (\phi_g\tilde{\otimes}\phi_g)\Delta[h_1]\ .\end{aligned}$$

In the second case,

$$\bar{\Delta}\phi_g(h_1h_2) = \bar{\Delta}\big(\phi_g(h_1) \sqcup\!\sqcup \phi_g(h_2)\big) = (\bar{\Delta}\phi_g(h_1)) \sqcup\!\sqcup (\bar{\Delta}\phi_g(h_2))\ ,$$

where we have used the fact that $\bar{\Delta}$ is a morphism with respect to $\sqcup\!\sqcup$. By the inductive assumption, we obtain

$$(\phi_g\tilde{\otimes}\phi_g)(\Delta h_1) \sqcup\!\sqcup (\phi_g\tilde{\otimes}\phi_g)(\Delta h_2) = (\phi_g\tilde{\otimes}\phi_g)(\Delta h_1 \cdot \Delta h_2) = (\phi_g\tilde{\otimes}\phi_g)\Delta(h_1h_2)\ ,$$

where, in the first equality we have used the fact that ϕ_g is a $\sqcup\!\sqcup$ morphism and Δ is a $\cdot$ morphism. This proves (4.9). □

Proof of Proposition 4.2.7. From (4.8), the path $\mathbf{X}$ is only defined through the incremental object $\mathbf{X}_{st}$. Hence, we must first check that $\mathbf{X}_{tt} = 1$ and that

$$\mathbf{X}_{st} = \mathbf{X}_{su} \star \mathbf{X}_{ut}, \tag{4.10}$$

for every $s, u, t \in [0, T]$. The first claim follows from the fact that $\bar{\mathbf{X}}_{tt} = 1$ and that $\phi_g^* 1 = 1$, where ϕ_g^* is the adjoint of ϕ_g and 1 is the counit. To check (4.10), notice that

$$\langle \mathbf{X}_{su} \star \mathbf{X}_{ut}, h\rangle = \langle \mathbf{X}_{su}\tilde{\otimes}\mathbf{X}_{ut}, \Delta h\rangle = \langle \bar{\mathbf{X}}_{su}\tilde{\otimes}\bar{\mathbf{X}}_{ut}, (\phi_g\tilde{\otimes}\phi_g)\Delta h\rangle\ .$$

Applying Lemma 4.2.8, the above equals

$$\langle \bar{\mathbf{X}}_{su}\tilde{\otimes}\bar{\mathbf{X}}_{ut}, \bar{\Delta}\phi_g(h)\rangle = \langle \bar{\mathbf{X}}_{su} \otimes \bar{\mathbf{X}}_{ut}, \phi_g(h)\rangle = \langle \bar{\mathbf{X}}_{st}, \phi_g(h)\rangle = \langle \mathbf{X}_{st}, h\rangle\ .$$

The regularity condition (3.17) for a branched rough path follows easily from the fact that $\phi_g(\tau)$ is, for every $\tau \in \mathcal{T}$, a linear combination in $(\mathbb{R}^d)^{\otimes|\tau|}$. Hence, the regularity of

$\langle \mathbf{X}_{st}, \tau \rangle$ will follow from (4.6). We finally check that $\mathbf{X}_t \stackrel{\text{def}}{=} \mathbf{X}_{0t}$ takes values in the truncated group-like elements $G_N(\mathcal{H})$. Since ϕ_g is a morphism with respect to $\cdot$ and $\sqcup\!\sqcup$, we have that

$$\langle \mathbf{X}_t, h_1 h_2 \rangle = \langle \bar{\mathbf{X}}_t, \phi_g(h_1 h_2) \rangle = \langle \bar{\mathbf{X}}_t, \phi_g(h_1) \sqcup\!\sqcup \phi_g(h_2) \rangle \ ,$$

for any $h_1, h_2 \in \mathcal{H}$. Since $\bar{\mathbf{X}}$ is geometric and hence group-like, (4.4) yields

$$\langle \bar{\mathbf{X}}_t, \phi_g(h_1) \sqcup\!\sqcup \phi_g(h_2) \rangle = \langle \bar{\mathbf{X}}_t, \phi_g(h_1) \rangle \langle \bar{\mathbf{X}}_t, \phi_g(h_2) \rangle = \langle \mathbf{X}_t, h_1 \rangle \langle \mathbf{X}_t, h_2 \rangle \ .$$

Hence, $\mathbf{X}$ takes values in $G_N(\mathcal{H})$. □

4.2.9 Branched rough paths are geometric rough paths

The main result of this subsection provides a converse to Proposition 4.2.7, namely, for a given branched rough path $\mathbf{X}$ lying above a path X, we can construct a geometric rough path $\bar{\mathbf{X}}$ lying above a a higher dimensional path $\bar{X}$, in such a way that $\bar{\mathbf{X}}$ contains all the information of $\mathbf{X}$. Hence, every branched rough path can be viewed as a geometric rough path, living in an extended space.

Before stating the main result, we first need some notation. As above, let $\mathcal{B}_n$ be the real vector space spanned by $\mathcal{T}_n$, we can then define the tensor product algebra $T(\mathcal{B}_n)$ exactly as above. Similarly, we can define $T(\mathcal{B})$, where $\mathcal{B}$ is the real (and infinite dimensional) vector space spanned by $\mathcal{T}$. In $T(\mathcal{B})$, the elements of $\mathcal{T}$ are indivisible objects with respect to the coproduct $\bar{\Delta}$, that is

$$\bar{\Delta}\tau = 1 \tilde{\otimes} \tau + \tau \tilde{\otimes} 1 \ ,$$

for any $\tau \in \mathcal{T}$. Moreover, basis elements of $T(\mathcal{B})$ are tensors of the form $\tau_1 \otimes \cdots \otimes \tau_k$, for $\tau_i \in \mathcal{T}$. As usual, we denote the truncated tensor algebra by

$$T^{(N)}(\mathcal{B}_n) = \bigoplus_{k=0}^{N} \mathcal{B}_n^{\otimes k} \ . \tag{4.11}$$

Every tensor product space has the usual grading defined which counts the number of non-trivial factors in each tensor product. However, we equip $T(\mathcal{B})$ (and $T(\mathcal{B}_n)$) with a grading that does not ignore the individual grading of the trees. That is, we have

$$|\tau_1 \otimes \cdots \otimes \tau_n| = |\tau_1| + \cdots + |\tau_n| \ , \tag{4.12}$$

where $|\tau_i|$ is the $\mathcal{H}$ grading that counts the number of vertices in τ_i. Hence, we have the decomposition

$$T(\mathcal{B}) = \bigoplus_{m=0}^{\infty} T(\mathcal{B})_{(m)} \ ,$$

where $T(\mathcal{B})_{(m)}$ is the vector space spanned by the tensors $\tau_1 \otimes \cdots \otimes \tau_k$ for $\tau_i \in \mathcal{T}$ with $|\tau_1| + \cdots + |\tau_k| = m$, with $T(\mathcal{B})_{(0)} = \mathbb{R}$.

We will construct a path $\bar{X}$ will take values in the vector space $\mathcal{B}_N$. Since $\mathcal{B}_1 \cong \mathbb{R}^d$, to say that $\bar{X}$ is an extension of X means that $\pi_{\mathcal{B}_1}(\bar{X}) = X$. The geometric rough path $\bar{\mathbf{X}}$ will be built in the space $T^{(N)}(\mathcal{B}_N)$ defined by (4.11), satisfying $\langle \bar{\mathbf{X}}_{st}, \tau \rangle = \delta \bar{X}^{\tau}_{st}$ for each $\tau \in \mathcal{T}_N$. Recall that $\mathcal{H}$ has the decomposition

$$\mathcal{H} = \bigoplus_{m=0}^{\infty} \mathcal{H}_{(m)} \ ,$$

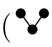
where $\mathcal{H}_{(m)}$ is the vector space spanned by $\mathcal{F}_{(m)}$, the set of all $\tau_1 \dots \tau_k \in \mathcal{F}$ with $|\tau_1| + \cdots + |\tau_k| = m$. The construction of $\bar{\mathbf{X}}$ relies on the following graded morphism of Hopf algebras, that is, a linear map $\psi : \mathcal{H}_{(m)} \to T(\mathcal{B})_{(m)}$ for each $m \in \mathbb{N}$, which is a morphism with respect to products and coproducts.

Lemma 4.2.10. *There exists a graded morphism of Hopf algebras* $\psi : (\mathcal{H}, \cdot, \Delta) \to (T(\mathcal{B}), ш, \bar{\Delta})$ *satisfying*

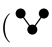
$$\psi(\tau) = \tau + \psi_{n-1}(\tau) \ , \tag{4.13}$$

for any $\tau \in \mathcal{T}_n$*, where* ψ_{n-1} *denotes the projection of* ψ *onto* $T(\mathcal{B}_{n-1})$*.*

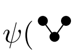
To illustrate the property (4.13), consider the following example. In the unlabelled case $d = 1$, we will see that

$\psi($  $) =$  $+ 2 \bullet \otimes \bullet \otimes \bullet + 2 \bullet \otimes$  .

Thus, we have

$\psi($  $) =$  $+ \psi_2($  $)$,

where

$\psi_2($  $) = \pi_{T(\mathcal{B}_2)} \psi($  $) = \pi_{T(\mathcal{B}_2)}\big($  $+ 2 \bullet \otimes \bullet \otimes \bullet + 2 \bullet \otimes$  $\big)$

$= 2 \bullet \otimes \bullet \otimes \bullet + 2 \bullet \otimes$  .

Proof of Lemma 4.2.10. We will construct ψ on each $\mathcal{H}_{(n)}$. For $n = 1$, the condition (4.13) forces

$$\psi(\bullet_a) = \bullet_a,$$

for each $a = 1 \dots d$. Hence, $\psi : \mathcal{H}_{(1)} \to \mathcal{B}_1 = T(\mathcal{B})_{(1)}$, and it is trivial to check that ψ is a morphism of Hopf algebras. Suppose that we have constructed such a map on $\mathcal{H}_{(k)}$, for all $1 \leq k \geq n-1$. We will now construct an extension of ψ to $\mathcal{F}_{(n)}$ and hence $\mathcal{H}_{(n)}$. Elements in $\mathcal{F}_{(n)}$ are either $\tau \in \mathcal{T}_n$ or products of elements in $\mathcal{F}_{(p)}$ and $\mathcal{F}_{(q)}$ for $p+q=n$. We will firstly extend ψ to $\mathcal{T}_n$.

Let $\tau \in \mathcal{T}_n$ with $\Delta\tau = \tau^1 \tilde{\otimes} \tau^2 + 1\tilde{\otimes}\tau + \tau\tilde{\otimes}1$, for some $\tau \in \mathcal{T}_n$, where we sum over the non-trivial parts τ^1, τ^2. We define

$$\psi_{n-1}(\tau) = \psi(\tau^1) \otimes \tau^2 \ . \tag{4.14}$$

We then set $\psi(\tau) = \psi_{n-1}(\tau) + \tau$. To complete the extension we set

$$\psi(h_1 h_2) = \psi(h_1) \sqcup \psi(h_2) \ ,$$

for $h_1 h_2 \in \mathcal{F}_{(n)}$ with $h_1 \in \mathcal{F}_{(p)}$ and $h_2 \in \mathcal{F}_{(q)}$. By construction, ψ satisfies (4.13) and is a graded morphism of algebras on $\mathcal{F}_{(n)}$, hence we only need that

$$(\psi\tilde{\otimes}\psi)\Delta h = \bar{\Delta}\psi(h) \ , \tag{4.15}$$

for all $h \in \mathcal{F}_{(n)}$. For $\tau \in \mathcal{T}_n$, we have that

$$\bar{\Delta}\psi(\tau) = \bar{\Delta}(\psi(\tau^1) \otimes \tau^2 + \tau) = \bar{\Delta}(\psi(\tau_1) \otimes \tau^2) + \tau\tilde{\otimes}1 + 1\tilde{\otimes}\tau \ . \tag{4.16}$$

It is easy to see that

$$\bar{\Delta}(\psi(\tau^1) \otimes \tau^2) = (\psi(\tau^1) \otimes \tau^2)\tilde{\otimes}1 + (\bar{\Delta}\psi(\tau^1)) \otimes (1\tilde{\otimes}\tau^2) \ .$$

Since $\tau^1 \in \mathcal{F}_{(n-1)}$, the inductive hypothesis implies that (4.16) equals

$$(\psi(\tau^1) \otimes \tau^2 + \tau)\tilde{\otimes}1 + (\psi\tilde{\otimes}\psi)(\Delta\tau^1) \otimes (1\tilde{\otimes}\tau^2) + 1\tilde{\otimes}\tau \ .$$

Using the notation, $(\Delta'\tilde{\otimes}\mathrm{Id})\Delta'\tau = \tau^{11}\tilde{\otimes}\tau^{12}\tilde{\otimes}\tau^2$, the above equals

$$\psi(\tau)\tilde{\otimes}1 + \psi(\tau^{11})\tilde{\otimes}(\psi(\tau^{12}) \otimes \tau^2) + 1\tilde{\otimes}(\psi(\tau^1) \otimes \tau^2) + \psi(\tau^1)\tilde{\otimes}\tau^2 + 1\tilde{\otimes}\tau \ . \tag{4.17}$$

On the other hand, using the notation $(\mathrm{Id}\tilde{\otimes}\Delta')\Delta'\tau = \tau^1\tilde{\otimes}\tau^{21}\tilde{\otimes}\tau^{22}$, we have that

$$(\psi\tilde{\otimes}\psi)\Delta\tau = \psi(\tau)\tilde{\otimes}1 + 1\tilde{\otimes}\psi(\tau) + \psi(\tau^1)\tilde{\otimes}(\psi(\tau^{21}) \otimes \tau^{22} + \tau^2) \ .$$

Hence, it is sufficient to check that

$$\psi(\tau^{11})\tilde{\otimes}(\psi(\tau^{12})\otimes\tau^{2}) = \psi(\tau^{1})\tilde{\otimes}\psi(\tau^{21})\otimes\tau^{22}\ . \tag{4.18}$$

But, from the coassociativity of the coproduct (and hence the reduced coproduct), we have that

$$\tau^{11}\tilde{\otimes}\tau^{12}\tilde{\otimes}\tau^{2} = (\Delta'\tilde{\otimes}\mathrm{Id})\Delta'\tau = (\mathrm{Id}\tilde{\otimes}\Delta')\Delta'\tau = \tau^{1}\tilde{\otimes}\tau^{21}\tilde{\otimes}\tau^{22}$$

and (4.18) clearly follows. The fact that (4.15) holds for the product h_1h_2 follows easily from the inductive hypothesis, and the fact that $\bar{\Delta}$ and Δ are morphisms with respect to $\sqcup\!\sqcup$ and $\cdot$ respectively. □

We can now state the main result of this section.

Theorem 4.2.11. *Let $X = (X^i)_{i=1\ldots d}$ be a path in $\mathbb{R}^d$ and $\mathbf{X}$ a γ-Hölder continuous branched rough path satisfying $\langle \mathbf{X}_t, \bullet_i\rangle = X^i_t$. Then there exists*

1. *a path $\bar{X} = (\bar{X}^\tau)_{\tau\in\mathcal{T}_N}$ taking values in the vector space $\mathcal{B}_N$ and satisfying $\pi_{\mathcal{B}_1}(\bar{X}) = X$,*
2. *a γ-Hölder geometric rough path $\bar{\mathbf{X}}$ in $T^{(N)}(\mathcal{B}_N)$ satisfying $\langle \bar{\mathbf{X}}_{st}, \tau\rangle = \delta\bar{X}^\tau_{st}$ for each $\tau\in\mathcal{T}_N$,*

such that

$$\langle \mathbf{X}_{st}, h\rangle = \langle \bar{\mathbf{X}}_{st}, \psi(h)\rangle\ , \tag{4.19}$$

for every $h\in\mathcal{H}_N$ and where ψ is the map constructed in Lemma 4.2.10.

The idea behind the proof is to construct $\bar{\mathbf{X}}$ iteratively, using the extension theorem 4.2.2. The first part of the iteration is to extend the path X. To start the iteration, we define the intermediate extension $\widehat{\mathbf{X}}^{(1)} : [0,T]\to G^{(N)}(\mathcal{B}_1)$ by

$$\widehat{\mathbf{X}}^{(1)}_t = \exp\left(\sum_{i=1}^{d}\langle \mathbf{X}_t, \bullet_i\rangle\ \bullet_i\right). \tag{4.20}$$

Hence, we have that $\langle \widehat{\mathbf{X}}^{(1)}_t, \bullet_i\rangle = \langle \mathbf{X}_t, \bullet_i\rangle$ and

$$\langle \widehat{\mathbf{X}}^{(1)}_t, \bullet_{a_1}\otimes\cdots\otimes\bullet_{a_k}\rangle = \frac{1}{k!}\langle \mathbf{X}_t, \bullet_{a_1}\cdots\bullet_{a_k}\rangle\ ,$$

for $a_i = 1\ldots d$ and $k\le N$. All we have done is extend X by adding the purely symmetric tensor components. Let K_1 be the normal subgroup of $G^{(N)}(\mathcal{B}_1)$ defined by

$$K_1 = \exp\left(W_2(\mathcal{B}_1)\oplus\cdots\oplus W_N(\mathcal{B}_1)\right)\ ,$$

or equivalently, let $K_1 = \exp L_1$, where L_1 is the Lie ideal generated by

$$[\mathcal{B}_1, \mathcal{B}_1]_\otimes = \operatorname{span}\{\bullet_i \otimes \bullet_j - \bullet_j \otimes \bullet_i : i, j = 1 \dots d\}. \tag{4.21}$$

In general, the path $\widehat{\mathbf{X}}_t^{(1)}$ is not a γ-Hölder continuous path in the group $G^{(N)}(\mathcal{B}_1)$, but it is in the quotient group $G^{(N)}(\mathcal{B}_1)/K_1$. Indeed, we have that

$$\|(\widehat{\mathbf{X}}_s^{(1)})^{-1} \otimes \widehat{\mathbf{X}}_t^{(1)}\|_{G^{(N)}(\mathcal{B}_1)/K_1} = \inf_{k \in K_1} \|(\widehat{\mathbf{X}}_s^{(1)})^{-1} \otimes \widehat{\mathbf{X}}_t^{(1)} \otimes k\|_{G^{(N)}(\mathcal{B}_1)} , \tag{4.22}$$

and by the Baker-Campbell-Hausdorff formula,

$$\begin{aligned}
(\widehat{\mathbf{X}}_s^{(1)})^{-1} \otimes \widehat{\mathbf{X}}_t^{(1)} &= \exp\left(\sum_{i=1}^d \big(\langle \mathbf{X}_t, \bullet_i\rangle - \langle \mathbf{X}_s, \bullet_i\rangle\big)\bullet_i\right) \otimes \exp(\ell) \\
&= \exp\left(\sum_{i=1}^d \langle \mathbf{X}_{st}, \bullet_i\rangle \bullet_i\right) \otimes \exp(\ell) ,
\end{aligned}$$

where $\ell \in L_1$. Hence, taking $k = \exp(-\ell)$, we can bound (4.22) by

$$\| \exp\left(\sum_{i=1}^d \langle \mathbf{X}_{st}, \bullet_i\rangle \bullet_i\right) \|_{G^{(N)}(\mathcal{B}_1)} \le C \sum_{i=1}^d |\langle \mathbf{X}_{st}, \bullet_i\rangle| \le C|t-s|^\gamma ,$$

which proves the claim for $\widehat{\mathbf{X}}^{(1)}$. We can therefore apply the extension theorem to $\widehat{\mathbf{X}}^{(1)}$, in particular it follows that there exists a γ-Hölder continuous path $\bar{\mathbf{X}}^{(1)} \in G^{(N)}(\mathcal{B}_1)$ such that

$$\pi_{G^{(N)}(\mathcal{B}_1)/K_1}(\bar{\mathbf{X}}^{(1)}) = \widehat{\mathbf{X}}^{(1)} ,$$

which simply means that $\langle \bar{\mathbf{X}}_{st}^{(1)}, \bullet_i\rangle = \langle \widehat{\mathbf{X}}_{st}^{(1)}, \bullet_i\rangle = \delta X_{st}$ for all $i = 1 \dots d$.

Remark 4.2.12. We should mention that one can actually choose *any* geometric rough path $\bar{\mathbf{X}}^{(1)}$ above X. We only use the extension theorem as it will work for every X.

The second part of the iteration relies on a generalisation of the following well-known (and easily verified) fact. Namely, that the difference between two area processes over a common path is equal to the increment of another path. In our case, for each $a, b = 1 \dots d$ there exists a path

$$\bar{X}^{\substack{\bullet a \\ \bullet b}} : [0,T] \to \mathbb{R} \quad \text{such that} \quad \delta \bar{X}_{st}^{\substack{\bullet a \\ \bullet b}} = \langle \mathbf{X}_{st}, {\substack{\bullet a \\ \bullet b}}\rangle - \langle \bar{\mathbf{X}}_{st}^{(1)}, \bullet_a \otimes \bullet_b\rangle,$$

where $\bar{\mathbf{X}}_{st}^{(1)} = (\bar{\mathbf{X}}_s^{(1)})^{-1} \otimes \bar{\mathbf{X}}_t^{(1)}$ and the path is unique up to an additive constant. We add

$\bar{X}$ as another component of a new path $\widehat{\mathbf{X}}^{(2)} : [0,T] \to T^{(N)}(\mathcal{B}_2)$. To be precise, we define

$$\widehat{\mathbf{X}}_t^{(2)} = \exp\left(\log \bar{\mathbf{X}}_t^{(1)} + \sum_{a,b=1}^{d} \bar{X}_t^{\begin{smallmatrix}\bullet\, a\\ |\\ \bullet\, b\end{smallmatrix}} \begin{smallmatrix}\bullet\, a\\ |\\ \bullet\, b\end{smallmatrix}\right) .$$

Hence, $\widehat{\mathbf{X}}^{(2)}$ satisfies

$$\langle \widehat{\mathbf{X}}_t^{(2)}, \begin{smallmatrix}\bullet\, a\\ |\\ \bullet\, b\end{smallmatrix} \rangle = \bar{X}_t^{\begin{smallmatrix}\bullet\, a\\ |\\ \bullet\, b\end{smallmatrix}}, \tag{4.23}$$

for all $a, b = 1 \dots d$,

$$\langle \widehat{\mathbf{X}}_t^{(2)}, \tau_1 \otimes \cdots \otimes \tau_k \rangle = \frac{1}{k!} \langle \widehat{\mathbf{X}}_t^{(2)}, \tau_1 \rangle \dots \langle \widehat{\mathbf{X}}_t^{(2)}, \tau_k \rangle ,$$

for all tensors $\tau_1 \otimes \cdots \otimes \tau_k \in T^{(N)}(\mathcal{B}_2) \setminus T^{(N)}(\mathcal{B}_1)$, and $\widehat{\mathbf{X}}^{(2)}$ is an extension of $\bar{\mathbf{X}}_1$, in the sense that

$$\pi_{T^{(N)}(\mathcal{B}_1)}(\widehat{\mathbf{X}}^{(2)}) = \bar{\mathbf{X}}^{(1)} .$$

We then repeat the first step, by finding the right quotient group and re-applying the extension theorem. To this end, for any integer $1 \leq n \leq N$, we define L_n as the Lie ideal generated by the set

$$[\mathcal{B}_{(n)}, \mathcal{B}_n]_{\otimes} = \operatorname{span}\{\tau_1 \otimes \tau_2 - \tau_2 \otimes \tau_1 : \tau_1, \tau_2 \in \mathcal{T} \text{ with } |\tau_1| = n \text{ and } |\tau_2| \leq n \} \tag{4.24}$$

in the free Lie algebra $\mathcal{G}^{(N)}(\mathcal{B}_n)$. In particular, L_n contains all brackets in $\mathcal{G}^{(N)}(\mathcal{B}_n)$ with at least one factor from $\mathcal{B}_{(n)}$. In order to construct meaningful quotients, we require the following Lemma.

Lemma 4.2.13. *For each $1 \leq n \leq N$, $K_n = \exp(L_n)$ is a normal subgroup of $G^{(N)}(\mathcal{B}_n)$.*

Proof. The statement is an elementary result in the theory of Lie algebras, see [Kir08, Theorem 3.22], for instance. □

Proof of Theorem 4.2.11. Throughout the proof, we will denote $\bar{\mathbf{X}}_{st}^{(n)} = (\bar{\mathbf{X}}_s^{(n)})^{-1} \otimes \bar{\mathbf{X}}_t^{(n)}$. Proceeding by induction, we will prove that, for each integer $n \geq 1$, there exists a γ-Hölder continuous path $\bar{\mathbf{X}}^{(n)} : [0,T] \to G^{(N)}(\mathcal{B}_n)$ such that

$$\langle \mathbf{X}_{st}, h \rangle = \langle \bar{\mathbf{X}}_{st}^{(n)}, \psi(h) \rangle , \tag{4.25}$$

for every $h \in \mathcal{H}_n$. For $n = 1$, we know from the introductory argument that such a construction is possible. Hence, assume the claim holds for some $n \geq 1$. We will now construct $\bar{\mathbf{X}}^{(n+1)}$ and show that (4.25) holds in the $n+1$ case. We will first show that for every $\tau \in \mathcal{T}_{(n+1)}$ there exists a path $\bar{X}^\tau : [0,T] \to \mathbb{R}$ such that

$$\delta \bar{X}^\tau_{st} = \langle \mathbf{X}_{st}, \tau \rangle - \langle \bar{\mathbf{X}}^{(n)}_{st}, \psi_n(\tau) \rangle \;, \tag{4.26}$$

unique up to an additive constant, this path will allow us to define $\bar{\mathbf{X}}^{(n+1)}$. Without loss of generality, let $\tau = [h]$, for some $h \in \mathcal{F}_{(n)}$, where we omit the label of the root. We have that

$$\langle \mathbf{X}_{st}, [h] \rangle = \langle \mathbf{X}_{su}, [h] \rangle + \langle \mathbf{X}_{ut}, [h] \rangle + \langle \mathbf{X}_{su}, h^1 \rangle \langle \mathbf{X}_{ut}, [h^2] \rangle \;, \tag{4.27}$$

where $\Delta h = h^1 \tilde{\otimes} h^2 + 1 \tilde{\otimes} h + h \tilde{\otimes} 1$ and we omit the summation. By hypothesis, we have that

$$\langle \mathbf{X}_{su}, h^1 \rangle \langle \mathbf{X}_{ut}, [h^2] \rangle = \langle \bar{\mathbf{X}}^{(n)}_{su}, \psi(h^1) \rangle \langle \bar{\mathbf{X}}^{(n)}_{ut}, \psi([h^2]) \rangle \;,$$

since h^1 and $[h^2]$ are elements of $\mathcal{H}_n$. Moreover, by definition of ψ, we have that

$$\psi(h^1) \tilde{\otimes} \psi([h^2]) = (\psi \tilde{\otimes} \psi) \Delta'[h] = \Delta' \psi([h]) \;,$$

where Δ' is the reduced coproduct. This yields the identity

$$\langle \mathbf{X}_{su}, h^1 \rangle \langle \mathbf{X}_{ut}, [h^2] \rangle = \langle \bar{\mathbf{X}}^{(n)}_{st}, \psi_n([h]) \rangle - \langle \bar{\mathbf{X}}^{(n)}_{su}, \psi_n([h]) \rangle - \langle \bar{\mathbf{X}}^{(n)}_{ut}, \psi_n([h]) \rangle \;,$$

combining this with (4.27), we obtain

$$\langle \mathbf{X}_{st}, [h] \rangle - \langle \bar{\mathbf{X}}^{(n)}_{st}, \psi_n([h]) \rangle = \langle \mathbf{X}_{su}, [h] \rangle - \langle \bar{\mathbf{X}}^{(n)}_{su}, \psi_n([h]) \rangle + \langle \mathbf{X}_{ut}, [h] \rangle - \langle \bar{\mathbf{X}}^{(n)}_{ut}, \psi_n([h]) \rangle \;.$$

Setting $\tau = [h]$, this implies the existence of $\bar{X}^\tau$ for each $\tau \in \mathcal{T}_{(n+1)}$, satisfying (4.26). We include this path in our construction by defining the intermediate extension $\widehat{\mathbf{X}}^{(n+1)}$ of $\bar{\mathbf{X}}^{(n)}$, setting

$$\widehat{\mathbf{X}}^{(n+1)}_t = \exp\left(\log \bar{\mathbf{X}}^{(n)}_t + \sum_{\tau \in \mathcal{T}_{(n+1)}} \bar{X}^\tau_t \tau \right) \;. \tag{4.28}$$

Hence, $\widehat{\mathbf{X}}^{(n+1)} : [0,T] \to G^{(N)}(\mathcal{B}_{n+1})$ and satisfies $\langle \widehat{\mathbf{X}}^{(n+1)}_t, \tau \rangle = \bar{X}^\tau_t$ for all $\tau \in \mathcal{T}_{(n+1)}$,

$$\langle \widehat{\mathbf{X}}^{(n+1)}_t, \tau_1 \otimes \cdots \otimes \tau_m \rangle = \frac{1}{m!} \langle \widehat{\mathbf{X}}^{(n+1)}_t, \tau_1 \rangle \ldots \langle \widehat{\mathbf{X}}^{(n+1)}_t, \tau_m \rangle \;,$$

for all $\tau_1 \otimes \cdots \otimes \tau_m \in T^{(N)}(\mathcal{B}_{n+1}) \setminus T^{(N)}(\mathcal{B}_n)$ and $\widehat{\mathbf{X}}^{(n+1)}$ is an extension of $\bar{\mathbf{X}}^{(n)}$ is the sense that

$$\pi_{T^{(N)}(\mathcal{B}_n)}(\widehat{\mathbf{X}}^{(n+1)}) = \bar{\mathbf{X}}^{(n)} \;.$$

We then have the following crucial fact, which we shall verify in the sequel.

Lemma 4.2.14. *For each $n \leq N-1$, the intermediate extension $\widehat{\mathbf{X}}^{(n+1)}$ is a γ-Hölder continuous path in the quotient group $G^{(N)}(\mathcal{B}_{n+1})/K_{n+1}$.*

Thus, from the extension theorem 4.2.2, we know that there exists a γ-Hölder path $\bar{\mathbf{X}}^{(n+1)} : [0,T] \to G^{(N)}(\mathcal{B}_{n+1})$ satisfying

$$\pi_{G^{(N)}(\mathcal{B}_{n+1})/K_{n+1}}(\bar{\mathbf{X}}^{(n+1)}) = \widehat{\mathbf{X}}^{(n+1)} .$$

We will now check that $\bar{\mathbf{X}}^{(n+1)}$ satisfies (4.25) for the basis elements $\mathcal{F}_{n+1}$ and hence $\mathcal{H}_{n+1}$. Firstly, suppose $h \in \mathcal{F}_n$, then $\psi(h) \in T^{(N)}(\mathcal{B}_n)$, which follows from the fact that ψ is graded. Moreover, since $\bar{\mathbf{X}}^{(n+1)}$ agrees with $\bar{\mathbf{X}}^{(n)}$ on $T^{(N)}(\mathcal{B}_n)$, we have that

$$\langle \mathbf{X}_{st}, h\rangle = \langle \bar{\mathbf{X}}^{(n)}_{st}, \psi(h)\rangle = \langle \bar{\mathbf{X}}^{(n+1)}_{st}, \psi(h)\rangle ,$$

which proves the claim for $\mathcal{F}_n$. It is clear that every element in $\mathcal{F}_{(n+1)}$ is is either a tree $[h]$ for some $h \in \mathcal{F}_{(n)}$ or a product h_1h_2 for $h_1, h_2 \in \mathcal{F}_n$. For the tree case, we have the identity

$$\langle \bar{\mathbf{X}}^{(n+1)}_{st}, [h]\rangle = \langle (\bar{\mathbf{X}}^{(n+1)}_s)^{-1} \otimes \bar{\mathbf{X}}^{(n+1)}_t, [h]\rangle = \langle (\widehat{\mathbf{X}}^{(n+1)}_s)^{-1} \otimes \widehat{\mathbf{X}}^{(n+1)}_t, [h]\rangle = \delta \bar{X}^{[h]}_{st} , \tag{4.29}$$

where we have used the facts that $\bar{\mathbf{X}}^{(n+1)}$ and $\widehat{\mathbf{X}}^{(n+1)}$ coincide on $[h]$ and that $\langle \widehat{\mathbf{X}}^{(n+1)}_t, [h]\rangle = \bar{X}^{[h]}_t$ and $\langle (\widehat{\mathbf{X}}^{(n+1)}_s)^{-1}, [h]\rangle = -\bar{X}^{[h]}_s$. And by definition,

$$\delta \bar{X}^{[h]}_{st} = \langle \mathbf{X}_{st}, [h]\rangle - \langle \bar{\mathbf{X}}^{(n)}_{st}, \psi_n([h])\rangle = \langle \mathbf{X}_{st}, [h]\rangle - \langle \bar{\mathbf{X}}^{(n+1)}_{st}, \psi_n([h])\rangle ,$$

where the last equality follows from the fact that $\psi_n([h]) \in T^{(N)}(\mathcal{B}_n)$, on which $\bar{\mathbf{X}}^{(n+1)}$ and $\bar{\mathbf{X}}^{(n)}$ agree. Combining this with (4.29), the claim follows from the condition $\psi([h]) = [h] + \psi_n([h])$. For the product case,

$$\langle \mathbf{X}_{st}, h_1h_2\rangle = \langle \mathbf{X}_{st}, h_1\rangle\langle \mathbf{X}_{st}, h_2\rangle = \langle \bar{\mathbf{X}}^{(n)}_{st}, \psi(h_1)\rangle\langle \bar{\mathbf{X}}^{(n)}_{st}, \psi(h_2)\rangle .$$

Since $\bar{\mathbf{X}}^{(n)}$ is geometric, the above equals

$$\langle \bar{\mathbf{X}}^{(n)}_{st}, \psi(h_1) \sqcup\!\sqcup\, \psi(h_2)\rangle = \langle \bar{\mathbf{X}}^{(n+1)}_{st}, \psi(h_1) \sqcup\!\sqcup\, \psi(h_2)\rangle = \langle \bar{\mathbf{X}}^{(n+1)}_{st}, \psi(h_1h_2)\rangle .$$

where the first equailty follows from the fact that $\psi(h_1) \sqcup\!\sqcup\, \psi(h_2) \in T^{(N)}(\mathcal{B}_n)$, on which $\bar{\mathbf{X}}^{(n)}$ and $\bar{\mathbf{X}}^{(n+1)}$ coincide and the second follows from the fact that ψ is a morphism with respect to multiplication. □

Proof of Lemma 4.2.14. By the Baker-Campbell-Hausdorff formula, we have that

$$\begin{aligned}
\widehat{\mathbf{X}}_{st}^{(n+1)} &:= (\widehat{\mathbf{X}}_s^{(n+1)})^{-1} \otimes \widehat{\mathbf{X}}_t^{(n+1)} \\
&= \exp\Bigg(\sum_{\tau \in \mathcal{T}_{(n+1)}} -\bar{X}_s^\tau \tau + \ell_1\Bigg) \otimes (\bar{\mathbf{X}}_s^{(n)})^{-1} \otimes \bar{\mathbf{X}}_t^{(n)} \otimes \exp\Bigg(\sum_{\tau \in \mathcal{T}_{(n+1)}} \bar{X}_t^\tau \tau + \ell_2\Bigg) \\
&= \exp\Bigg(\sum_{\tau \in \mathcal{T}_{(n+1)}} -\bar{X}_s^\tau \tau + \ell_1\Bigg) \otimes \bar{\mathbf{X}}_{st}^{(n)} \otimes \exp\Bigg(\sum_{\tau \in \mathcal{T}_{(n+1)}} \bar{X}_t^\tau \tau + \ell_2\Bigg) \\
&= \bar{\mathbf{X}}_{st}^{(n)} \otimes \exp\Bigg(\sum_{\tau \in \mathcal{T}_{(n+1)}} \delta\bar{X}_{st}^\tau \tau\Bigg) \otimes \exp(\ell_3)\ ,
\end{aligned}$$

where ℓ_1, ℓ_2 are linear combinations of brackets between $\log \bar{\mathbf{X}}^{(n)}$ and $\mathcal{T}_{(n+1)}$ and are therefore in the ideal L_{n+1}, and where ℓ_3 is a linear combination of brackets between $\log \bar{\mathbf{X}}^{(n)}$, $\mathcal{T}_{(n+1)}$, ℓ_1 and ℓ_2 and is therefore also in L_{n+1}. By taking $k = \exp(-\ell_3)$, we therefore have that

$$\begin{aligned}
\|\widehat{\mathbf{X}}_{st}^{(n+1)}\|_{G^{(N)}(\mathcal{B}_{n+1})/K_{n+1}} &= \inf_{k \in K_{n+1}} \|\bar{\mathbf{X}}_{st}^{(n)} \otimes \exp\Bigg(\sum_{\tau \in \mathcal{T}_{(n+1)}} \delta\bar{X}_{st}^\tau \tau\Bigg) \otimes \exp(\ell_3) \otimes k\|_{G^{(N)}(\mathcal{B}_{n+1})} \\
&\le \|\bar{\mathbf{X}}_{st}^{(n)} \otimes \exp\Bigg(\sum_{\tau \in \mathcal{T}_{(n+1)}} \delta\bar{X}_{st}^\tau \tau\Bigg)\|_{G^{(N)}(\mathcal{B}_{n+1})} \\
&\le \|\bar{\mathbf{X}}_{st}^{(n)}\|_{G^{(N)}(\mathcal{B}_{n+1})} + \|\exp\Bigg(\sum_{\tau \in \mathcal{T}_{(n+1)}} \delta\bar{X}_{st}^\tau \tau\Bigg)\|_{G^{(N)}(\mathcal{B}_{n+1})}
\end{aligned}$$

where in the last inequality we have used the sub-additivity property of $\|\cdot\|_{G^{(N)}(\mathcal{B}_{n+1})}$. For the first term, using the equivalence of norms on $G^{(N)}(\mathcal{B}_n)$, we have that

$$\begin{aligned}
\|\bar{\mathbf{X}}_{st}^{(n)}\|_{G^{(N)}(\mathcal{B}_{n+1})} &\le C \sum_{m=1}^N \sum_{\{\tau_1,\dots,\tau_m\} \subset \mathcal{T}_{n+1}} |\langle \log \bar{\mathbf{X}}_{st}^{(n)}, \tau_1 \otimes \cdots \otimes \tau_m\rangle|^{1/m} \\
&= C \sum_{m=1}^N \sum_{\{\tau_1,\dots,\tau_m\} \subset \mathcal{T}_{n}} |\langle \log \bar{\mathbf{X}}_{st}^{(n)}, \tau_1 \otimes \cdots \otimes \tau_m\rangle|^{1/m} \\
&\le C \|\bar{\mathbf{X}}_{st}^{(n)}\|_{G^{(N)}(\mathcal{B}_n)} \le C|t-s|^\gamma\ ,
\end{aligned}$$

since $\bar{\mathbf{X}}^{(n)}$ is a γ-Hölder continuous path in $G^{(N)}(\mathcal{B}_n)$. For the second term, we have that

$$\|\exp\Bigg(\sum_{\tau \in \mathcal{T}_{(n+1)}} \delta\bar{X}_{st}^\tau \tau\Bigg)\|_{G^{(N)}(\mathcal{B}_{n+1})} \le C \sum_{\tau \in \mathcal{T}_{(n+1)}} |\delta\bar{X}_{st}^\tau|\ .$$

And by definition,

$$
\begin{aligned}
|\delta \bar{X}^\tau_{st}| = |\langle \mathbf{X}_{st}, \tau\rangle - \langle \bar{\mathbf{X}}^{(n)}_{st}, \psi_n(\tau)\rangle| &\le |\langle \mathbf{X}_{st}, \tau\rangle| + |\langle \bar{\mathbf{X}}^{(n)}_{st}, \psi_n(\tau)\rangle| \\
&\le |\langle \mathbf{X}_{st}, \tau\rangle| + C\|\bar{\mathbf{X}}^{(n)}_{st}\|_{G^{(N)}(\mathcal{B}_n)} \le C|t-s|^\gamma \ .
\end{aligned}
$$

This completes the proof. □

Remark 4.2.15. Throughout the construction, we have ignored the fact that the path elements $\langle \bar{\mathbf{X}}, \tau\rangle$ actually have $\gamma|\tau|$-Hölder regularity, rather than just γ. Hence, for each component $\langle \bar{\mathbf{X}}, \tau_1 \otimes \cdots \otimes \tau_n\rangle$ with $|\tau_1| + \cdots + |\tau_n| > N$, there will be a *canonical* choice, given by defining the component as a Young integral.

If branched rough paths can be written as geometric rough paths, then we should be able to import some of the tools from geometric rough paths to the world of branched rough paths. The following result tells us that the extension theorem 4.2.2 can also be used on branched rough paths, for a special but very useful class of extension. Namely, if we have a branched rough path $\mathbf{X}^1$ above a path $X = (X^1, \ldots, X^d)$ and an *extended* path $\bar{X} = (X^1, \ldots, X^d, \bar{X}^{d+1}, \ldots, \bar{X}^e)$, then there exists a branched rough path $\mathbf{X}^2$ above $\bar{X}$ which agrees with $\mathbf{X}^1$ on the X components.

Corollary 4.2.16. Let $\mathcal{H}^1, \mathcal{H}^2$ be the Connes-Kreimer Hopf algebras generated by the alphabets $\mathcal{A}_1$ and $\mathcal{A}_2$ respectively, where $\mathcal{A}_1 \subset \mathcal{A}_2$, so that $\mathcal{H}^1$ is a sub Hopf algebra of $\mathcal{H}^2$. Let $X = (X^i)_{i\in\mathcal{A}_1}$ and $\bar{X} = (\bar{X}^i)_{i\in\mathcal{A}_2}$ be two γ-Hölder continuous paths with $\bar{X}^i = X^i$ when $i \in \mathcal{A}_1$. Let $\mathbf{X}^1$ be a branched rough path on $\mathcal{H}^1$ with $\langle \mathbf{X}^1_{st}, \bullet_i\rangle = \delta X^i_{st}$ for each $i \in \mathcal{A}_1$. Then there exists a branched rough path $\mathbf{X}^2$ on $\mathcal{H}^2$ with $\langle \mathbf{X}^2_{st}, \bullet_i\rangle = \delta \bar{X}^i_{st}$ for each $i \in \mathcal{A}_2$ and $\mathbf{X}^2$ is an extension of $\mathbf{X}^1$ in the sense that

$$
\langle \mathbf{X}^2_{st}, h\rangle = \langle \mathbf{X}^1_{st}, h\rangle \ , \tag{4.30}
$$

for every $h \in \mathcal{H}^1$.

Proof. Without loss of generality, assume $\mathcal{A}_1 = \{1, \ldots, d\}$ and $\mathcal{A}_2 = \{1, \ldots, d+1\}$, so that $X = (X^1, \ldots, X^d)$ and $\bar{X} = (X^1, \ldots, X^d, X^{d+1})$. Let $\mathcal{B}^1_N$ and $\mathcal{B}^2_N$ be the vector spaces spanned by the trees $|\tau| \le N$, with vertex decorations from $\mathcal{A}_1$, $\mathcal{A}_2$ respectively. Let $\psi^1 : \mathcal{H}^1 \to T(\mathcal{B}^1_N)$ be the map constructed in Lemma 4.2.10 and similarly for $\psi^2 : \mathcal{H}^2 \to T(\mathcal{B}^2_N)$. Clearly, we have that $\psi^1(h) = \psi^2(h)$ for $h \in \mathcal{H}^1$. From Theorem 4.2.11, we know that there exists a geometric rough path $\bar{\mathbf{X}}^1$ on $T^{(N)}(\mathcal{B}^1_N)$ satisfying

$$
\langle \mathbf{X}^1_{st}, h\rangle = \langle \bar{\mathbf{X}}^1_{st}, \psi^1(h)\rangle \ .
$$

Now define $\widehat{\mathbf{X}}^1 : [0,T] \to T^{(N)}(\mathcal{B}_N^2)$ by

$$\widehat{\mathbf{X}}_t^1 = \exp\left(\log \bar{\mathbf{X}}_t^1 + \bar{X}_t^{d+1} \bullet_{d+1}\right) \ .$$

Using the same techniques employed in Lemma 4.2.14, one can show that $\widehat{\mathbf{X}}^1$ is a γ-Hölder continuous path in the quotient group $G^{(N)}(\mathcal{B}_N^2)/K$, where $K = \exp L$ and L is the Lie ideal in $T^{(N)}(\mathcal{B}_N^2)$ generated by $[\mathcal{B}_N^2, \bullet_{d+1}]$. From the Lyons-Victoir extension theorem 4.2.2, there exists a γ-Hölder path $\bar{\mathbf{X}}^2 : [0,T] \to G^{(N)}(\mathcal{B}_N^2)$ satisfying

$$\langle \bar{\mathbf{X}}_{st}^2, u\rangle = \langle \widehat{\mathbf{X}}_{st}^1, u\rangle \ ,$$

for all $u \in T^{(N)}(\mathcal{B}_N^1)$ and $\langle \bar{\mathbf{X}}_{st}^2, \bullet_{d+1}\rangle = \delta X_{st}^{d+1}$. We then define $\mathbf{X}^2$ in $\mathcal{H}^2$ by

$$\langle \mathbf{X}_{st}^2, h\rangle = \langle \bar{\mathbf{X}}_{st}^2, \psi^2(h)\rangle \ ,$$

It follows from the properties and ψ^2 that $\mathbf{X}^2$ is indeed a branched rough path. Now, let $h \in \mathcal{H}^1$ then we have that

$$\langle \mathbf{X}_{st}^2, h\rangle = \langle \bar{\mathbf{X}}_{st}^2, \psi^2(h)\rangle = \langle \bar{\mathbf{X}}_{st}^2, \psi^1(h)\rangle = \langle \widehat{\mathbf{X}}_{st}^1, \psi^1(h)\rangle = \langle \bar{\mathbf{X}}_{st}^1, \psi^1(h)\rangle = \langle \mathbf{X}_{st}, h\rangle \ ,$$

which proves (4.30). Moreover, because $\psi^2(\bullet_{d+1}) = \bullet_{d+1}$, we have that

$$\langle \mathbf{X}_{st}^2, \bullet_{d+1}\rangle = \langle \bar{\mathbf{X}}_{st}^2, \bullet_{d+1}\rangle = \delta \bar{X}^{d+1} \ ,$$

which shows that $\mathbf{X}^2$ is a branched rough path above $\bar{X}$ and hence completes the proof. □

4.3 Conversion formula

If $\mathbf{Y}$ is the solution to the controlled rough path equation (3.28) with $\langle 1, \mathbf{Y}\rangle = Y$, then from Proposition 3.3.10 we have that

$$\delta Y_{st} = \sum_{\tau \in \mathcal{T}_N} f_\tau(Y_s)\langle \mathbf{X}_{st}, \tau\rangle + r_{st} \ , \tag{4.31}$$

where the coefficients $f_\tau(Y_s) = \langle \tau, \mathbf{Y}_s\rangle$ are determined by (3.30) with $f_{\bullet_i} = f_i$. In Section 4.2, we saw that for every branched rough $\mathbf{X}$, there exists a geometric rough path $\bar{\mathbf{X}}$ taking values in $T^{(N)}(\mathcal{B}_N)$ and satisfying

$$\langle \mathbf{X}_{st}, \tau\rangle = \langle \bar{\mathbf{X}}_{st}, \psi(\tau)\rangle \ ,$$

where ψ is the map derived in Lemma 4.2.10. If we apply this transformation to (4.31), we see that

$$\sum_{\tau\in\mathcal{T}_N} f_\tau(Y_s)\langle \mathbf{X}_{st},\tau\rangle = \sum_{\sigma\in U_{N,N}} f_{\psi^*(\sigma)}(Y_s)\langle \bar{\mathbf{X}}_{st},\sigma\rangle \tag{4.32}$$

where $\psi^* : T((\mathcal{B}_N)) \to \mathcal{H}^*$ is the adjoint of ψ, where $f_{\psi^*(\sigma)} = \sum_\tau \langle \psi^*(\sigma),\tau\rangle f_\tau$ and where

$$U_{N,n} = \{\tau_1 \otimes \cdots \otimes \tau_k : \tau_i \in \mathcal{T}_n \text{ with } k \leq N\} \tag{4.33}$$

is the set of basis tensors for $T^{(N)}(\mathcal{B}_n)$. Since Y appears to be controlled by $\bar{\mathbf{X}}$, it is natural to ask whether Y solves an RDE driven by the geometric rough path $\bar{\mathbf{X}}$, providing a generalised Itô-Stratonovich conversion formula. In Subsection 4.3.1 we provide a criterion to determined when expressions controlled by a geometric rough path are solutions to RDEs driven by that geometric rough path. In Subsection 4.3.8, namely in Theorem 4.3.9 we derive the Itô-Stratonovich conversion formula.

4.3.1 Geometric RDEs

Let $\bar{\mathbf{X}}$ be a branched rough path above $\bar{X} \in \mathbb{R}^d$ satisfying

$$\langle \bar{\mathbf{X}}_{st}, h\rangle = \langle \bar{\mathbf{X}}_{st}, \iota\phi_g(h)\rangle \ , \tag{4.34}$$

for each $h \in \mathcal{H}_N$, where $\iota : T(\mathbb{R}^d) \to \mathcal{H}$ is the inclusion map. Hence, $\bar{\mathbf{X}}$ is a geometric (branched) rough path. Let Y be a controlled rough path solution to the RDE

$$dY_t = f(Y_t)\cdot d\bar{X}_t \ , \tag{4.35}$$

driven by a geometric rough path $\bar{\mathbf{X}}$, where $f(Y)\cdot d\bar{X} = \sum_{i=1}^d f_i(Y)d\bar{X}^i$ and the vector fields $f_i : \mathbb{R}^e \to \mathbb{R}^e$ are smooth. From Proposition 3.3.10, we have that

$$\delta Y_{st} = \sum_{\tau\in\mathcal{T}_N} f_\tau(Y_s)\langle \mathbf{X}_{st},\tau\rangle + r_{st} \ , \tag{4.36}$$

where $|r_{st}| = o(|t-s|)$ and the coefficients f_τ satisfy the recurrence relation (3.30) with $f_{\bullet_i} = f_i$. The geometric constraint (4.34) allows us to rewrite δY_{st} as an expression controlled by only the linear trees in $\mathcal{H}_N$, which we identify with the basis elements of $T^{(N)}(\mathbb{R}^d)$. To be precise,

$$\sum_{\tau\in\mathcal{T}_N} f_\tau(Y_s)\langle \bar{\mathbf{X}}_{st},\tau\rangle = \sum_{\sigma\in U_{N,1}} f_{\phi_g^*(\sigma)}(Y_s)\langle \bar{\mathbf{X}}_{st},\iota\sigma\rangle \ ,$$

where $\phi_g^* : T((\mathbb{R}^d)) \to \mathcal{H}^*$ is the adjoint of ϕ_g, where $f_{\phi_g^*(\sigma)} = \sum_\tau \langle \phi_g^*(\sigma), \tau \rangle f_\tau$ and

$$U_{N,1} = \{e_{v_1} \otimes \cdots \otimes e_{v_k} : v_i = 1 \ldots d \text{ and } k \leq N\} \tag{4.37}$$

denotes the basis tensors of $T^{(N)}(\mathbb{R}^d)$. Note that only those terms in the subspace $T^{(N)}(\mathbb{R}^d)$ appear, since all branched trees are in the kernel of ϕ_g^*.

Remark 4.3.2. From (4.33), we have

$$\begin{aligned} U_{N,1} &= \{\bullet_{v_1} \otimes \cdots \otimes \bullet_{v_k} : v_i = 1 \ldots d \text{ and } k \leq N\} \\ &\cong \{e_{v_1} \otimes \cdots \otimes e_{v_k} : v_i = 1 \ldots d \text{ and } k \leq N\} , \end{aligned}$$

so that (4.37) is not an abuse of notation.

We can use this representation to develop another recurrence formula, to characterise those expressions controlled by geometric rough paths that are solutions to a given RDE.

Proposition 4.3.3. *Let $\bar{\mathbf{X}}$ be a geometric (branched) rough path above $\bar{X}$. Then $\bar{\mathbf{Y}}$ with $\langle 1, \bar{\mathbf{Y}} \rangle = Y$ is the controlled rough path solution to*

$$dY_t = f(Y_t) \cdot d\bar{X}_t , \tag{4.38}$$

driven by $\bar{\mathbf{X}}$ if and only if $\langle \tau, \bar{\mathbf{Y}}_t \rangle = f_\tau(Y_t)$ as defined above and

$$\delta Y_{st} = \sum_{\sigma \in U_{N,1}} F_\sigma(Y_s) \langle \bar{\mathbf{X}}_{st}, \iota\sigma \rangle + r_{st} , \tag{4.39}$$

where $|r_{st}| = o(|t-s|)$ and where the coefficients F_σ are defined by the recurrence $F_{e_i} = f_i$ and

$$F_{e_{v_1} \otimes \cdots \otimes e_{v_n}} = F_{e_{v_1}} \cdot DF_{e_{v_2} \otimes \cdots \otimes e_{v_n}} , \tag{4.40}$$

for any $v_i = 1, \ldots, d$ and any $n \leq N$.

Remark 4.3.4. Since each $F_{e_{v_1} \otimes \cdots \otimes e_{v_k}} : \mathbb{R}^e \to \mathbb{R}^e$, the identity (4.40) should be interpreted as

$$F_{e_{v_1} \otimes \cdots \otimes e_{v_n}}(Y)_i = F_{e_{v_1}}(Y)_j \partial^j F_{e_{v_2} \otimes \cdots \otimes e_{v_n}}(Y)_i ,$$

for each $i = 1 \ldots e$, where $F_\sigma(Y)_i$ denotes the i-th component.

Remark 4.3.5. One can also define $\bar{\mathbf{X}}$-controlled rough paths for a geometric $\bar{\mathbf{X}}$ above a path $\bar{X} \in \mathbb{R}^d$. These are similarly defined as paths $\bar{\mathbf{Y}} : [0, T] \to T^{(N-1)}(\mathbb{R}^d)$ satisfying the consistency condition

$$\langle v, \bar{\mathbf{Y}}_t \rangle = \langle \bar{\mathbf{X}}_{st} \otimes v, \bar{\mathbf{Y}}_s \rangle + R^v_{st} \ ,$$

for every tensor v and where $|R^v_{st}| \leq C|t-s|^{(N-|v|)\gamma}$. The new recurrence condition (4.40) is then simply the analogue of the branched rough path recurrence (3.30) in a geometric controlled rough path setting. In paricular, one could also read Proposition 4.3.3 as: The geometric controlled rough path $\bar{\mathbf{Y}} : [0,T] \to T^N(\mathbb{R}^d)$ is a controlled rough path solution to (4.38) with $\langle 1, \bar{\mathbf{Y}} \rangle = Y$ if and only if Y satisfies (4.39) where the coefficients $\langle \sigma, \bar{\mathbf{Y}}_t \rangle = F_\sigma(Y_t)$ are determined by the recurrence (4.40) with $F_{e_i} = f_i$. However, since we can already define geometric rough paths as a special class of branched rough paths, we see no need for this extra definition.

Remark 4.3.6. Naturally, we can apply Proposition 4.3.3 to any geometric (branched) rough path $\bar{\mathbf{X}}$ above a path $\bar{X}$, where $\bar{X}$ takes values in an arbitrary vector space V. For instance, in the next subsection we will have $\bar{X}$ taking values in $\mathcal{B}_N$, as constructed in Theorem 4.2.11. In this case, the condition (4.39) looks like

$$\delta Y_{st} = \sum_{\sigma \in U_{N,N}} F_\sigma(Y_s) \langle \bar{\mathbf{X}}_{st}, \sigma \rangle + r_{st} \ ,$$

where $U_{N,N}$ is defined by (4.33) and where F_σ satisfy

$$F_{\tau_1 \otimes \cdots \otimes \tau_n} = F_{\tau_1} \otimes DF_{\tau_2 \otimes \cdots \otimes \tau_n} \ ,$$

for all $\tau_1 \otimes \cdots \otimes \tau_n \in U_{N,N}$.

Before proving the proposition, we need the following lemma, which highlights a useful property of the functions f_τ. This lemma will be used in both this chapter and the next. As usual, we will use the notation $f_h = \langle h, 1 \rangle \mathrm{Id} + \sum_\tau \langle h, \tau \rangle f_\tau$ for any $h \in \mathcal{H}^*$.

Lemma 4.3.7. *We have that*

$$D^q f_h : (f_{\lambda_1}, \ldots, f_{\lambda_q}) = f_{(\lambda_1 \ldots \lambda_n) \star h} \ ,$$

for any $\lambda_1, \ldots, \lambda_q \in \mathcal{T}^*$ *and any* $h \in \mathcal{H}^*$.

Proof of Proposition 4.3.3. We will first prove the 'only if' statement. From Proposition (3.3.10), we know that the controlled rough path solution $\bar{\mathbf{Y}}$ to (4.38) with $\langle 1, \bar{\mathbf{Y}} \rangle = Y$ satisfies

$$\delta Y_{st} = \sum_{\tau \in \mathcal{T}_N} f_\tau(Y_s) \langle \bar{\mathbf{X}}_{st}, \tau \rangle + r_{st} \ ,$$

where $|r_{st}| = o(|t-s|)$ and has coefficients $\langle \tau, \bar{\mathbf{Y}}_t \rangle = f_\tau(Y_t)$. Since $\bar{\mathbf{X}}$ is geometric, we also know that

$$\delta Y_{st} = \sum_{\sigma \in U_{N,1}} f_{\phi_g^*(\sigma)}(Y_s) \langle \bar{\mathbf{X}}_{st}, \iota\sigma \rangle + r_{st} \ .$$

Therefore, since $f_{\phi_g^*(e_i)} = f_{\bullet_i} = f_i$, it suffices to check that $f_{\phi_g^*(\sigma)}(Y_s)$ satisfies (4.40) for each tensor $\sigma \in U_{N,1}$. Firstly, from Lemma 4.2.8, we know that $(\phi_g \tilde{\otimes} \phi_g)\Delta = \bar{\Delta}\phi_g$. Using the dual of this expression, we obtain

$$\phi_g^*(\sigma_1 \otimes \sigma_2) = \phi_g^*(\sigma_1) \star \phi_g^*(\sigma_2) \ ,$$

for any $\sigma_1, \sigma_2 \in T^{(N)}(\mathbb{R}^d)$. In particular,

$$\phi_g^*(e_{v_1} \otimes \cdots \otimes e_{v_n}) = \phi_g^*(e_{v_1}) \star \phi_g^*(e_{v_2} \otimes \cdots \otimes e_{v_n}) = e_{v_1} \star \phi_g^*(e_{v_2} \otimes \cdots \otimes e_{v_n}) \ .$$

Combining this with Lemma 4.3.7, we obtain

$$\begin{aligned} f_{\phi_g^*(e_{v_1} \otimes \cdots \otimes e_{v_n})} &= f_{e_{v_1} \star \phi_g^*(e_{v_2} \otimes \cdots \otimes e_{v_n})} \\ &= f_{e_{v_1}} \cdot D f_{\phi_g^*(e_{v_2} \otimes \cdots \otimes e_{v_n})} \\ &= f_{\phi_g^*(e_{v_1})} \cdot D f_{\phi_g^*(e_{v_2} \otimes \cdots \otimes e_{v_n})} \ . \end{aligned}$$

This proves the claimed recurrence. For the 'if' statement, suppose Y satisfies (4.39), with coefficients F_σ satisfying the recurrence (4.40). Let f_τ be the coefficients defined by (3.30) with $f_{\bullet_i} = f_i$. Since both F_σ and $f_{\phi_g^*(\sigma)}$ satisfy (4.40), with $F_{e_i} = f_{\phi_g^*(e_i)}$ we must have $F_\sigma = f_{\phi_g^*(\sigma)}$ for all $\sigma \in U_{N,1}$. Then, using the same calculation as above, we have that

$$\begin{aligned} \delta Y_{st} - \sum_{\tau \in \mathcal{T}_N} f_\tau(Y_s) \langle \bar{\mathbf{X}}_{st}, \tau \rangle &= \delta Y_{st} - \sum_{\sigma \in U_{N,1}} f_{\phi_g^*(\sigma)}(Y_s) \langle \bar{\mathbf{X}}_{st}, \iota\sigma \rangle \\ &= \delta Y_{st} - \sum_{\sigma \in U_{N,1}} F_\sigma(Y_s) \langle \bar{\mathbf{X}}_{st}, \iota\sigma \rangle = r_{st} \ . \end{aligned}$$

It follows from Proposition 3.3.10 that $\bar{\mathbf{Y}}$ is the controlled rough path solution to (4.38). Hence, upon proving Lemma 4.3.7, this completes the proof. □

Proof of Lemma 4.3.7. First note that if h is a non-trivial product then certainly $(\lambda_1 \dots \lambda_q) \star h$ is a linear combination of non-trivial products. Hence, we can restrict our attention to those $h \in \mathcal{H}^*$ that are linear combinations of elements in $\mathcal{T}^*$, since $f_{\tau_1 \dots \tau_n} = 0$ all non-trivial products $\tau_1 \dots \tau_n \in \mathcal{F}^* \setminus \mathcal{T}^*$. Moreover, since $f_{\tau_1 + \cdots + \tau_n} = f_{\tau_1} + \cdots + f_{\tau_n}$, the claim will follow from

$$D^q f_\tau : (f_{\lambda_1}, \dots, f_{\lambda_q}) = f_{(\lambda_1 \dots \lambda_n) \star \tau} \ ,$$

for all $\lambda_1, \dots, \lambda_q, \tau \in \mathcal{T}$.

We will prove the claim by induction. The claim clearly holds for $\tau = \bullet_i$ and any $\lambda_1, \dots, \lambda_q \in \mathcal{T}^*$, since this is simply the recurrence (3.30). Suppose the claim holds for $\tau \in \mathcal{T}^*_m$ and all $\lambda_1, \dots, \lambda_q \in \mathcal{T}^*$ we will prove the claim for $\tau \in \mathcal{T}^*_{m+1}$ and all $\lambda_1, \dots, \lambda_q \in \mathcal{T}^*$. Without loss of generality, let $\tau = [\tau_1 \dots \tau_n]_i$ for $\tau_j \in \mathcal{T}_m$. Firstly, by the recurrence (3.30), we have that

$$D^q f_{[\tau_1 \dots \tau_n]_i} : (f_{\lambda_1}, \dots, f_{\lambda_q}) = D^q \Big(D^n f_i : (f_{\tau_1} \dots f_{\tau_n}) \Big) : (f_{\lambda_1}, \dots, f_{\lambda_q}) \ .$$

If we apply the Leibniz formula then the above equals

$$\sum_{p, p_1, \dots, p_n} \binom{q}{p, p_1, \dots, p_n} D^{p+n} f_i : \Big(f_{\lambda_1}, \dots, f_{\lambda_p}, u_1^{p_1}, \dots, u_n^{p_n} \Big) \ , \tag{4.41}$$

where $\binom{q}{p, p_1, \dots, p_n} = \frac{q!}{p! p_1! \dots p_n!}$ and where we sum over all partitions $p + p_1 + \dots + p_n = q$ and where

$$u_i^{p_i} = D^{p_i} f_{\tau_i} : (f_{\lambda_i^1}, \dots, f_{\lambda_i^{p_i}}) \ .$$

Since $\tau_i \in \mathcal{T}_m$, it follows by the induction hypothesis that $u_i^{p_i} = f_{(\lambda_1^i \dots \lambda_{p_i}^i) \star \tau_i}$. Hence, (4.41) equals

$$\begin{aligned} &\sum_{\tilde{\lambda}_1, \dots, \tilde{\lambda}_n} \Big(\langle (\lambda_1^1 \dots \lambda_{p_1}^1) \star \tau_1, \tilde{\lambda}_1 \rangle \dots \langle (\lambda_1^n \dots \lambda_{p_1}^n) \star \tau_n, \tilde{\lambda}_n \rangle \Big) \\ &\qquad\qquad \binom{q}{p, p_1, \dots, p_n} \times D^{p+n} f_i : \Big(f_{\lambda_1}, \dots, f_{\lambda_p}, f_{\tilde{\lambda}_1}, \dots, f_{\tilde{\lambda}_n} \Big) \\ &= \sum_{\tilde{\lambda}_1, \dots, \tilde{\lambda}_n} \Big(\langle (\lambda_1^1 \dots \lambda_{p_1}^1) \star \tau_1, \tilde{\lambda}_1 \rangle \dots \langle (\lambda_1^n \dots \lambda_{p_1}^n) \star \tau_n, \tilde{\lambda}_n \rangle \Big) && (4.42) \\ &\qquad\qquad \times \binom{q}{p, p_1, \dots, p_n} f_{[\lambda_1 \dots \lambda_p \tilde{\lambda}_1 \dots \tilde{\lambda}_n]_i} \ , && (4.43) \end{aligned}$$

where we sum over all partitions $p + p_1 + \dots + p_n = q$ and all $\tilde{\lambda}_i \in \mathcal{T}$. On the other hand, we have that

$$f_{(\lambda_1 \dots \lambda_q) \star [\tau_1 \dots \tau_n]_i} = \sum_{\sigma \in \mathcal{T}} \langle (\lambda_1 \dots \lambda_q) \star [\tau_1 \dots \tau_n]_i, \sigma \rangle f_\sigma \ . \tag{4.44}$$

By eliminating those terms that will vanish, this equals

$$\sum_{\rho_1,\ldots,\rho_{n+p}} \frac{1}{(n+p)!}\langle(\lambda_1\ldots\lambda_q)\tilde{\otimes}[\tau_1\ldots\tau_n]_i, \Delta[\rho_1\ldots\rho_{n+p}]_i\rangle f_{[\rho_1\ldots\rho_{n+p}]_i} \ ,$$

where we sum over all $\rho_i \in \mathcal{T}_N$ and the factor $\frac{1}{(n+p)!}$ appears due to the fact that $[\rho_1\ldots\rho_{n+p}]_i$ is identical for different arrangements of the ρ_i. By definition, we have that

$$\sum_{p=0}^{N}\langle(\lambda_1\ldots\lambda_q)\tilde{\otimes}[\tau_1\ldots\tau_n]_i, \Delta[\rho_1\ldots\rho_{n+p}]_i\rangle$$
$$= \langle\lambda_1\ldots\lambda_q, \rho_1^{(1)}\ldots\rho_{n+p}^{(1)}\rangle\langle\tau_1\ldots\tau_n, \rho_1^{(2)}\ldots\rho_{n+p}^{(2)}\rangle \ .$$

Now, each of these terms will vanish unless $\rho_i^{(2)} = 1$ for exactly p factors in $\rho_1^{(2)}\ldots\rho_{n+p}^{(2)}$. Since we are summing over all ρ_i, the expression must be symmetric in the ρ_i. In particular, we can assume that $\rho_i^{(2)} = 1$ for $1 \le i \le p$, provided we include the combinatorial factor $\binom{n+p}{p}$. Of course, this implies $\rho_i^{(1)}\tilde{\otimes}\rho_i^{(2)} = \rho_i\tilde{\otimes}1$ for $1 \le i \le p$. Hence, (4.44) equals

$$\sum_{\rho_1,\ldots,\rho_{n+p}} \frac{1}{(n+p)!}\binom{n+p}{p}\langle\lambda_1\ldots\lambda_q, \rho_1\ldots\rho_p\rho_{p+1}^{(1)}\ldots\rho_{p+n}^{(1)}\rangle\langle\tau_1\ldots\tau_n, \rho_{p+1}^{(2)}\ldots\rho_{p+n}^{(2)}\rangle f_{[\rho_1\ldots\rho_{n+p}]} \ .$$

By the symmetry of the expression, we can also simplify this to

$$\sum_{\rho_1,\ldots,\rho_{n+p}} \frac{n!}{(n+p)!}\binom{n+p}{p}\langle\lambda_1\ldots\lambda_q, \rho_1\ldots\rho_p\rho_{p+1}^{(1)}\ldots\rho_{p+n}^{(1)}\rangle\langle\tau_1, \rho_{p+1}^{(2)}\rangle\ldots\langle\tau_n, \rho_{p+n}^{(2)}\rangle f_{[\rho_1\ldots\rho_{n+p}]}$$
$$= \sum_{\rho_1,\ldots,\rho_{n+p}} \frac{1}{p!}\langle\lambda_1\ldots\lambda_q, \rho_1\ldots\rho_p\rho_{p+1}^{(1)}\ldots\rho_{p+n}^{(1)}\rangle\langle\tau_1, \rho_{p+1}^{(2)}\rangle\ldots\langle\tau_n, \rho_{p+n}^{(2)}\rangle f_{[\rho_1\ldots\rho_{n+p}]}$$

Repeating the same idea on the other factors, this equals

$$\sum_{\rho_1,\ldots,\rho_{n+p}} \frac{1}{p!}\binom{q}{p}\langle\lambda_1\ldots\lambda_p, \rho_1\ldots\rho_p\rangle\langle\lambda_{p+1}\ldots\lambda_q, \rho_{p+1}^{(1)}\ldots\rho_{p+n}^{(1)}\rangle\langle\tau_1, \rho_{p+1}^{(2)}\rangle\ldots\langle\tau_n, \rho_{p+n}^{(2)}\rangle f_{[\rho_1\ldots\rho_{n+p}]}$$
$$= \sum_{\rho_1,\ldots,\rho_{n+p}} \binom{q}{p}\langle\lambda_1, \rho_1\rangle\ldots\langle\lambda_p, \rho_p\rangle\langle\lambda_{p+1}\ldots\lambda_q, \rho_{p+1}^{(1)}\ldots\rho_{p+n}^{(1)}\rangle\langle\tau_1, \rho_{p+1}^{(2)}\rangle\ldots\langle\tau_n, \rho_{p+n}^{(2)}\rangle f_{[\rho_1\ldots\rho_{n+p}]}$$

Let $p_1+\cdots+p_n = q-p$ be some partition and let $(\lambda_{p+1},\ldots,\lambda_q) = (\lambda_1^1,\ldots,\lambda_{p_1}^1,\ldots,\lambda_1^n,\ldots,\lambda_{p_n}^n)$.

Then it also follows from the symmetry of the expression that the above equals

$$\sum_{\rho_1,\ldots,\rho_{n+p}} \binom{q}{p}\binom{q-p}{p_1,\ldots,p_n} \langle\lambda_1,\rho_1\rangle\ldots\langle\lambda_p,\rho_p\rangle\langle\lambda^1_1\ldots\lambda^1_{p_1},\rho^{(1)}_{p+1}\rangle\ldots\langle\lambda^n_1\ldots\lambda^n_{p_n},\rho^{(1)}_{p+n}\rangle$$
$$\times\langle\tau_1,\rho^{(2)}_{p+1}\rangle\ldots\langle\tau_n,\rho^{(2)}_{p+n}\rangle f_{[\rho_1\cdots\rho_{n+p}]}$$
$$= \sum_{\rho_{p+1},\ldots,\rho_{n+p}} \binom{q}{p,p_1,\ldots,p_n}\langle(\lambda^1_1,\ldots,\lambda^1_{p_1})\star\tau_1,\rho_{p+1}\rangle\ldots\langle(\lambda^n_1,\ldots,\lambda^n_{p_n})\star\tau_n,\rho_{p+n}\rangle$$
$$\times f_{[\lambda_1\ldots\lambda_p\rho_{p+1}\ldots\rho_{p+n}]}$$

where we sum over all partitions $p+p_1+\cdots+p_n=q$. We see that (4.44) equals (4.43), which proves the induction. □

4.3.8 Itô-Stratonovich correction

We can now state and prove the generalised correction formula. In the following, let $f(Y)\cdot dX=\sum_{i=1}^d f_i(Y)dX^i$ for smooth vector fields $f_i:\mathbb{R}^e\to\mathbb{R}^e$. As usual, let f_τ be defined by the recurrence (3.30) with $f_{\bullet_i}=f_i$ and $f_h=\langle h,1\rangle\mathrm{Id}+\sum_{\tau\in\mathcal{T}}\langle h,\tau\rangle f_\tau$ for any $h\in\mathcal{H}^*$. Finally, let $\bar{X},\bar{\mathbf{X}}$ be as in Theorem 4.2.11

Theorem 4.3.9. *Let $\mathbf{Y}$ with $\langle 1,\mathbf{Y}\rangle=Y$ be the controlled rough path solution to the RDE*

$$dY_t=f(Y_t)\cdot dX_t\ , \tag{4.45}$$

driven by a branched rough path $\mathbf{X}$ over X. Then $\mathbf{Y}$ also solves the RDE

$$dY_t=\bar{f}(Y_t)\cdot d\bar{X}_t\ , \tag{4.46}$$

driven by $\bar{\mathbf{X}}$, where $\bar{f}(Y)\cdot d\bar{X}=\sum_{\tau\in\mathcal{T}_N} f_\tau(Y)d\bar{X}^\tau$.

Proof. As in (4.32), we have that

$$\sum_{\tau\in\mathcal{T}_N} f_\tau(Y_s)\langle\mathbf{X}_{st},\tau\rangle=\sum_{\sigma\in U_{N,N}} f_{\psi^*(\sigma)}(Y_s)\langle\bar{\mathbf{X}}_{st},\sigma\rangle\ .$$

Hence, if we can show that the coefficients $f_{\psi^*(\sigma)}$ satisfy

$$f_{\psi^*(\tau_1\otimes\cdots\otimes\tau_n)}=f_{\psi^*(\tau_1)}\cdot Df_{\psi^*(\tau_2\otimes\cdots\otimes\tau_n)}\ ,$$

for all $\tau_1\otimes\cdots\otimes\tau_n\in U_{N,N}$ then Proposition 4.3.3 implies that $\mathbf{Y}$ also solves the RDE

$$dY_t=\sum_{\tau\in\mathcal{T}_N} f_{\psi^*(\tau)}(Y_t)d\bar{X}^\tau_t\ , \tag{4.47}$$

driven by the geometric rough path $\bar{\mathbf{X}}$. From the definition of ψ found in (4.14), one can easily check that $\psi^*(\tau) = \tau$, so that (4.46) and (4.47) are indeed the same RDE.

From Lemma 4.2.10, we know that $(\psi \otimes \psi)\Delta = \bar{\Delta}\psi$, the dual of this statement implies that

$$\begin{aligned}
\psi^*(\tau_1 \otimes \cdots \otimes \tau_n) &= \psi^*(\tau_1) \star \psi^*(\tau_2 \otimes \cdots \otimes \tau_n) \\
&= \sum_{\sigma_1 \in \mathcal{T}_N} \langle \psi(\sigma_1), \tau_1 \rangle \big(\sigma_1 \star \psi^*(\tau_2 \otimes \cdots \otimes \tau_n) \big) \; .
\end{aligned}$$

In light of this, the theorem follows almost immediately from Lemma 4.3.7, where we take $\lambda_1 \dots \lambda_q = \sigma_1$ and $h = \sigma_2 \star \cdots \star \sigma_n$. We have that

$$\begin{aligned}
f_{\psi^*(\tau_1 \otimes \cdots \otimes \tau_n)} &= \sum_{\sigma_1 \in \mathcal{T}_N} \langle \psi(\sigma_1), \tau_1 \rangle f_{\sigma_1 \star \psi^*(\tau_2 \otimes \cdots \otimes \tau_n)} \\
&= \sum_{\sigma_1 \in \mathcal{T}_N} \langle \psi(\sigma_1), \tau_1 \rangle f_{\sigma_1} \cdot D f_{\psi^*(\tau_2 \otimes \cdots \otimes \tau_n)} \\
&= f_{\psi^*(\tau_1)} \cdot D f_{\psi^*(\tau_2 \otimes \cdots \otimes \tau_n)} \; .
\end{aligned}$$

This proves the recurrence (4.40) and hence completes the proof. $\square$

Remark 4.3.10. As with the geometric rough path $\bar{\mathbf{X}}$, there is some redundancy in the extended vector field $\bar{f}$. In fact, it is in general possible to choose another geometric rough path $\widehat{\mathbf{X}}$, such that Y also solves an RDE driven by $\widehat{\mathbf{X}}$, and features fewer vector fields. For example, if $1/3 < \gamma \le 1/2$, we have that

$$\begin{aligned}
\delta Y_{st} = f_i(Y_s) \langle \bar{\mathbf{X}}_{st}, \bullet_i \rangle &+ f_j^\alpha(Y_s) \partial^\alpha f_i(Y_s) \langle \bar{\mathbf{X}}_{st}, \bullet_j \otimes \bullet_i \rangle \\
&+ f_j^\alpha(Y_s) \partial^\alpha f_i(Y_s) \langle \bar{\mathbf{X}}_{st}, \genfrac{}{}{0pt}{}{\bullet_j}{\bullet_i} \rangle \; + r_{st} \; ,
\end{aligned}$$

for $|r_{st}| = o(|t-s|)$. Let us now define $\langle \widehat{\mathbf{X}}, \bullet_i \rangle = \langle \bar{\mathbf{X}}, \bullet_i \rangle$ and

$$\langle \widehat{\mathbf{X}}, \bullet_j \otimes \bullet_i \rangle = \langle \bar{\mathbf{X}}, \bullet_j \otimes \bullet_i + \genfrac{}{}{0pt}{}{\bullet_j}{\bullet_i} - \genfrac{}{}{0pt}{}{\bullet_i}{\bullet_j} \rangle \; .$$

for all $i, j = 1, \dots, d$, and finally, we set

$$\langle \widehat{\mathbf{X}}, \bullet_{ij} \rangle = \langle \bar{\mathbf{X}}, \genfrac{}{}{0pt}{}{\bullet_j}{\bullet_i} + \genfrac{}{}{0pt}{}{\bullet_i}{\bullet_j} \rangle$$

for all $i \le j$. Since we have only changed the higher order components of $\bar{\mathbf{X}}$ by adding an anti-symmetric 2γ-Hölder path, $\widehat{\mathbf{X}}$ remains geometric. Moreover, we see that

$$\begin{aligned}\delta Y_{st} &= f_i(Y_s)\langle \widehat{\mathbf{X}}_{st}, \bullet_i \rangle + f_j^\alpha(Y_s)\partial^\alpha f_i(Y_s)\langle \widehat{\mathbf{X}}_{st}, \bullet_j \otimes \bullet_i \rangle \\ &\quad + \frac{1}{2}\big(f_l^\alpha(Y_s)\partial^\alpha f_k(Y_s) + f_k^\alpha(Y_s)\partial^\alpha f_l(Y_s)\big)\langle \widehat{\mathbf{X}}_{st}, \bullet_{kl} \rangle + r_{st} \ .\end{aligned}$$

Hence, Y solves the RDE

$$dY_t = f_i(Y_t)d\widehat{X}_t^i + \frac{1}{2}\big(f_l^\alpha(Y_t)\partial^\alpha f_k(Y_t) + f_k^\alpha(Y_t)\partial^\alpha f_l(Y_t)\big)d\widehat{X}^{kl} \ ,$$

where we only sum over $k \leq l$. This is clearly a simpler RDE than the one obtained in Theorem (4.3.9), since we sum over a smaller index set than $\mathcal{T}_2$. It is also more reminiscent of the usual Itô-Stratonovich correction. The simplification is quite easy in the case $N = 2$, but the procedure becomes much more complicated for larger N, and we can see no natural method of generalising this simplification for larger N.

To put this another way, suppose X is a Brownian motion, or indeed any path for which there is a *canonical* geometric rough path lying above it [FV10a, FV10b]; in the case of Brownian motion, this corresponds to constructing Stratonovich integrals. If the extension $\bar{\mathbf{X}}$ were constructed using this canonical geometric rough path above X, then we would recover the classical Itô-Stratonovich correction. In particular, the antisymmetric part $\frac{1}{2}\langle \bar{\mathbf{X}}_{st}, \substack{\bullet_j \\ \bullet_i} - \substack{\bullet_i \\ \bullet_j} \rangle$ would vanish and the symmetric part $\frac{1}{2}\langle \bar{\mathbf{X}}_{st}, \substack{\bullet_j \\ \bullet_i} + \substack{\bullet_i \\ \bullet_j} \rangle$ would be the quadratic variation.

Chapter 5

Itô's formula for non-geometric rough paths

5.1 Introduction

The centrepiece of Itô's stochastic calculus is his change of variables formula, known as Itô's formula. Since its inception, the Itô formula has been extended in a multitude of directions, most famously to semi-martingales. In the recent past, there have been attempts to extend the formula to more ad hoc situations, including stochastic processes for which no sufficient calculus exists. These include fractional Brownian motion [GRV03, GNRV05], 4-stable procceses [BM96], solutions to stochastic PDEs [BS10] and the so called "finite n-variation processes" [ER00, ER03]. All of these examples fit nicely into the framework of branched rough paths, in that each stochastic process X has (almost surely) *some* Hölder exponent $\gamma > 0$. In this chapter we will show that branched rough paths are the perfect tool for building change of variables formula for any path X with some Hölder exponent $\gamma > 0$ and a branched rough path $\mathbf{X}$ living above X.

5.1.1 Motivation

Before proceeding, we must ask: what should a generalised change of variables formula look like? For a path $X : [0,T] \to \mathbb{R}^d$ and a smooth function $F : \mathbb{R}^d \to \mathbb{R}^d$, any such formula should resemble

$$F(X_t) - F(X_s) = \int_s^t DF(X_r) \cdot dX_r + \text{ "higher order terms"},$$

where the "higher order terms" are integrals driven by paths more regular than X, that in some sense correspond to "brackets" between components of X. To illustrate this idea, we will cite two nice examples of change-of-variables formulae for non semi-martingales.

Example 5.1.2. The first example, which is the subject of [BS10], considers the case in which $X_t = u(t,x)$ for a fixed x, where u is the solution to the stochastic heat equation

$$\partial_t u = \partial_x^2 u + \dot{W} \ ,$$

and where $\dot{W}$ is space-time white noise. It is shown in [Swa07] that X has infinite quadratic variation, which is unsurprising given that one expects u to be no better than "almost $1/4$-differentiable" in time. In particular, the ordinary stochastic calculus for semi-martingales cannot be applied. Nevertheless, it is shown that the one can still define an integral using the mid-point Riemann sum

$$\int_s^t F'(X_r) d_{\text{mid}} X_r \stackrel{\text{def}}{=} \lim_{\pi \downarrow 0} \sum_{[u,v]\in\pi} F'(X_{(u+v)/2}) \delta X_{uv} \ ,$$

for a regular partition π of $[s,t]$, where the limit is taken in law. Using this fact, the authors construct the change-of-variables formula

$$\delta F(X)_{st} = \int_s^t F'(X_r) d_{\text{mid}} X_r + \frac{1}{2}\int_s^t F''(X_r) d[X,X]_r \ , \tag{5.1}$$

where $[X,X]$ is a (constant multiple of) Brownian motion (on the same probability space as X but independent of X) and where the $d[X,X]$ integral is of Itô type. Note that (5.1) must be interpreted on the level of probability measures, in particular, it only claims that the law of the left hand side is equal to the law of the right hand side.

This looks very similar to a standard Itô formula, except that the correction term is now a Brownian motion and is certainly no longer a BV-process. The path $[X,X]$ is constructed similarly to the ordinary bracket term in Itô's formula, in that it is defined using cancellations in the squares of the path X.

Example 5.1.3. The second example concerns an Itô formula for the class of "finite 3-variation" processes, discussed in [ER00]. A stochastic process X is said to be of finite 3-variation if the limit

$$\operatorname*{ucplim}_{\varepsilon\to 0} \frac{1}{\varepsilon}\int_0^{\cdot} (X_{r+\varepsilon} - X_r)^3 dr$$

exists. The limiting path $[X,X,X]$ is called the 3-variation of X. As shown in [ER00], whenever $[X,X,X]$ exists, it is a BV-process. An important example of 3-variation processes is given by fractional Brownian motion. The authors prove the change-of-variables formula

$$\delta F(X)_{st} = \int_s^t F'(X_r) d^{\circ} X_r - \frac{1}{12}\int_s^t F^{(3)}(X_r) d[X,X,X]_r \ , \tag{5.2}$$

where the first integral is known as the *symmetric integral*, and is defined by

$$\int_0^t Y_r d^{\circ} X_r = \operatorname*{ucplim}_{\varepsilon\to 0} \frac{1}{\varepsilon}\int_0^t Y_r (X_{r+\varepsilon} - X_r) dr \ .$$

In the case where X is a semi-martingale, the symmetric integral is simply the Stratonovich integral.

There are a few remarks to be made concerning the two change of variables formulae (5.1) and (5.2).

1. In the former example, the correction term arises as a quadratic variation type process, where as in the latter example, a correction terms arises as a cubic variation type process. For a general change of variables formula, one should therefore expect the correction integrals to involve generalised brackets of the form $[X^{i_1}, \ldots, X^{i_n}]$, where X^j denotes the j-th component of X.

2. The choice of the leading integral is important. In both examples, the integral was chosen because it was known to be a natural choice for which the discretisation schemes actually converge. However, if one had several choices for the definition of the leading integral, one would expect that they would lead to different definitions of the brackets.

3. In the first example, the path driving the correction integral $[X, X]$ was not of bounded variation and therefore the integral could be defined in many ways. Clearly, if one defined the integral against $d[X, X]$ in a different sense, this would lead to a possibly different change of variables formula. Of course, any other choice would be less natural than the choice taken in the example, but it is still important to note that one must consider how to define integrals driven by the bracket paths.

Branched rough paths provides the perfect platform to obtain the type of formulae seen above, but in a more general setting. In Section 5.2 we will treat the case where X is some γ-Hölder path with a branched rough path $\mathbf{X}$ above X. Our results generalise those given in the examples, in the sense that *every* change of variables formula obtained by some reasonable discretisation of integrals must be a special case of our formulae. This is due to the simple fact that every reasonable discretisation for which integrals converge will lead to the definition of a branched rough path. In Section 5.3 we obtain a change of variables formula for $F(Y)$, where Y is the solution to an RDE. This is more in line with the usual Itô formula for the solution of an SDE. Of course, the $F(X)$ case is implied by the $F(Y)$ case, but we include the simple case independently because the explainations and proofs are somewhat more intuitive than those for the general case.

5.2 The simple case

5.2.1 Outline of the rough path approach

Our aim is to prove a change of variables formula, similar to those given in the examples, but for an arbitrary γ-Hölder path $X : [0, T] \to \mathbb{R}^d$ with a branched rough path $\mathbf{X}$ above it. In particular, we will develop a formula for the differential $dF(X_t)$.

In the rough path setting, instead of choosing a discretisation scheme for the leading integral, we simply fix a branched rough path $\mathbf{X}$. For any smooth $F : \mathbb{R}^d \to \mathbb{R}^d$, we can define

$$\int_s^t DF(X_r) \cdot dX_r \stackrel{\text{def}}{=} \left\langle 1, \int_s^t DF(\mathbf{X}_r) \cdot dX_r \right\rangle ,$$

using the fact that $\mathbf{X}$ is certainly an $\mathbf{X}$-controlled rough path. The information on how to interpret integrals driven by brackets, and hence the definition of the brackets themselves, is contained in what we call the *bracket extension* $\widehat{\mathbf{X}}$ of $\mathbf{X}$. A bracket extension is obtained by first defining a path

$$\widehat{X} = X \oplus \{\widehat{X}^{(\alpha_1 \dots \alpha_n)} : \text{ with } \alpha_i = 1 \dots d \text{ and } n \leq N\} ,$$

containing the original path X as well as the bracket paths. These bracket paths are generalisations of the quadratic and cubic variations seen in the examples, in that they arise as corrections to integration by parts formulas. One should think of $\widehat{X}^{(\alpha_1 \dots \alpha_n)}$ as a shorthand for the n-covariation $[X^{\alpha_1}, \dots, X^{\alpha_n}]$ seen in Examples 5.1.2 and 5.1.3.

The bracket extension $\widehat{\mathbf{X}}$ is a branched rough path above $\widehat{X}$ that agrees with $\mathbf{X}$ on the components above X. Thus, the bracket extension tells us how to define the bracket path $\widehat{X}$ and also how to integrate objects against it. Since $\widehat{X}$ lives in the extended vector space $\mathbb{R}^d \oplus \mathbb{R}^{d^2} \oplus \dots \oplus \mathbb{R}^{d^N}$, the bracket extension $\widehat{\mathbf{X}}$ will live in an extended Hopf algebra $\widehat{\mathcal{H}}$, where the vertices of the generating trees have multi-index labels $(\alpha_1 \dots \alpha_n)$ in addition to the usual labels from $\{1, \dots, d\}$.

We will briefly illustrate the bracket extension by looking at the first few steps taken to construct it. For any $\mathbf{X}$, we can define a "quadratic variation" path $\widehat{X}^{(ij)}$ as the correction to the integration by parts formula

$$\delta \widehat{X}^{(ij)}_{st} = \delta X^i_{st} \delta X^j_{st} - \int_s^t \delta X^i_{sr} dX^j_r - \int_s^t \delta X^j_{sr} dX^i_r , \tag{5.3}$$

where we use the formal "integral" notation for the components of $\mathbf{X}$. In Example 5.1.2, by choosing $F(x) = x^2$ and rearranging the terms in (5.1) a bit, we obtain

$$(\delta X_{st})^2 = 2 \int_s^t \delta X_{sr} d_{\text{mid}} X_r + \delta [X, X]_{st} ,$$

which is precisely (5.3) with $\widehat{X}^{(ij)} = [X^i, X^j]$. Hence, this bracket records how "non-geometric" the integral components of $\mathbf{X}$ are. One might ask if we can obtain a bracket from the product of *three* paths. In Example 5.1.2, by choosing $F = x^3$, we obtain

$$(\delta X_{st})^3 = 3 \int_s^t (\delta X_{sr})^2 d_{\text{mid}} X_r + 3 \int_s^t \delta X_{sr} d[X, X]_r . \tag{5.4}$$

The bracket $[X, X, X]$ does not appear in (5.4) (indeed we will see that it is identically zero), however the expression still provides a hint. Returning now to the general X case, suppose we knew the integrals

$$\int_s^t \int_s^{r_1} dX^k_{r_2} d\widehat{X}^{(ij)}_{r_1} ,$$

for all $i, j, k = 1 \ldots d$. Then one can check that the expression

$$\delta X^i_{st}\delta X^j_{st}\delta X^k_{st} - \int_s^t \delta X^j_{sr_2}\delta X^k_{sr_2} dX^i_{r_1} - \int_s^t \delta X^j_{sr_2}\delta X^k_{sr_2} dX^i_{r_1} - \int_s^t \delta X^j_{sr_2}\delta X^k_{sr_2} dX^i_{r_1}$$
$$- \int_s^t \int_s^{r_1} dX^i_{r_2} d\widehat{X}^{(jk)}_{r_1} - \int_s^t \int_s^{r_1} dX^j_{r_2} d\widehat{X}^{(ik)}_{r_1} - \int_s^t \int_s^{r_1} dX^k_{r_2} d\widehat{X}^{(ij)}_{r_1} ,$$

is the increment of a path. One might call this path the "cubic-variation" $\widehat{X}^{(ijk)}$, since it arises in a similar fashion to the quadratic variation. Indeed, returning to Example 5.1.2, one can see from (5.4) that we obtain $\delta\widehat{X}^{(ijk)}_{st} = 0$, moreover, in Example 5.1.3 we similarly obtain $\delta\widehat{X}^{(ijk)}_{st} = [X^i, X^j, X^k]$. As we shall see, the cubic-variation plays a role in the change of variables formula as soon as $\gamma \leq 1/3$. Although it looks as if we just plucked the above expression out of this air, it actually arises quite naturally in the construction of an Itô formula. This will be evident in the examples given towards the end of this subsection.

Another way of viewing these bracket paths and integrals above bracket paths is that they provide a mechanism for writing down products of components of X as a linear combination of integrals. For example, the expression above yields

$$\delta X^i_{st}\delta X^j_{st}\delta X^k_{st} = \int_s^t \delta X^j_{sr_2}\delta X^k_{sr_2} dX^i_{r_1} + \int_s^t \delta X^j_{sr_2}\delta X^k_{sr_2} dX^i_{r_1} + \int_s^t \delta X^j_{sr_2}\delta X^k_{sr_2} dX^i_{r_1}$$
$$+ \int_s^t \int_s^{r_1} dX^i_{r_2} d\widehat{X}^{(jk)}_{r_1} + \int_s^t \int_s^{r_1} dX^j_{r_2} d\widehat{X}^{(ik)}_{r_1} + \int_s^t \int_s^{r_1} dX^k_{r_2} d\widehat{X}^{(ij)}_{r_1} + \delta\widehat{X}^{(ijk)}_{st} .$$

Moreover, within the integrals $\int_s^t \delta X^j_{sr_2}\delta X^k_{sr_2} dX^i_{r_1}$ we can replace the product $\delta X^j_{sr_2}\delta X^k_{sr_2}$ with a sum, using (5.3). Hence we can write $\delta X^i_{st}\delta X^j_{st}\delta X^k_{st}$ as a linear combination of components of the *signature* of the path $\widehat{X}$. This idea, that one can write products as a linear combination of iterated integrals, is of course also a feature of geometric rough paths, found in the shuffle product rule, and is precisely what makes them so useful. In our case, this property will be crucial in developing the change of variables formula.

We can see the recipe for more general "n-variations" starting to emerge: suppose we have all n-variation paths, then we can build the $(n+1)$-variation paths provided we know how to integrate against the n-variation paths. This is precisely the idea used in the sequel to "build" a bracket extension in Subsection 5.2.8. The steps we take, in a nutshell, are as follows

1. Suppose we have the bracket paths

   $$\widehat{X}^{(n)} = X \oplus \{\widehat{X}^{(\alpha_1 \ldots \alpha_k)} : \text{for all } \alpha_i = 1 \ldots d \text{ and } 2 \leq k \leq n\} .$$

   and a branched rough path $\widehat{\mathbf{X}}^{(n)}$ above $\widehat{X}^{(n)}$ that is an extension of $\mathbf{X}$.

2. We can then uniquely define the next bracket paths $\widehat{X}^{(\alpha_1\dots\alpha_{n+1})}$ using the components of $\widehat{\mathbf{X}}^{(n)}$.

3. Add these new brackets paths to the underlying path

$$\widehat{X}^{(n+1)} = \widehat{X}^{(n)} \oplus \{\widehat{X}^{(\alpha_1\dots\alpha_{n+1})} : \text{for all } a_i = 1\dots d\} \ .$$

4. Find a branched rough path $\widehat{\mathbf{X}}^{(n+1)}$ above $\widehat{X}^{(n+1)}$ that is an extension of $\widehat{\mathbf{X}}^{(n)}$.

5. Repeat until we obtain $\widehat{X} = \widehat{X}^{(N)}$ and $\widehat{\mathbf{X}} = \widehat{\mathbf{X}}^{(N)}$.

The resulting object $\widehat{\mathbf{X}}$ contains all the information on how to define and integrate against bracket paths, moreover, a given $\widehat{\mathbf{X}}$ uniquely determines a change of variables formula. Of course, there are many ways of defining the integral $\int X d\widehat{X}^{(ij)}$, so in general, there will be many ways of defining a bracket extension $\widehat{\mathbf{X}}$ for a given $\mathbf{X}$. Since the choice of $\widehat{\mathbf{X}}$ affects how we define integrals against brackets, this will affect the change of variables formula we obtain.

In the simple case, we obtain the following change of variables formula.

Theorem 5.2.2 (Simple version)**.** *For a given branched rough path $\mathbf{X}$ and bracket extension $\widehat{\mathbf{X}}$ of $\mathbf{X}$, we have that*

$$\delta F(X)_{st} = \int_s^t DF(X_r)\cdot dX_r + \sum_{n=2}^{N} \int_s^t \frac{D^n F(X_r)}{n!} : d\widehat{X} \ ,$$

where we use the shorthand

$$D^n F(X_r) : d\widehat{X} = \sum_{\alpha_1,\dots,\alpha_n=1}^{d} \partial^{\alpha_1}\dots\partial^{\alpha_n} F(X_r) d\widehat{X}^{(\alpha_1\dots\alpha_n)} \tag{5.5}$$

with $\widehat{X}^{(\alpha_1\dots\alpha_n)}$ being the bracket paths given by $\delta\widehat{X}^{(\alpha_1\dots\alpha_n)}_{st} = \langle \widehat{\mathbf{X}}_{st}, \bullet_{(\alpha_1\dots\alpha_n)}\rangle$ (these objects will be defined properly in Subsection 5.2.8).

Remark 5.2.3. To be clear, the bracket integrals are defined as

$$\int_s^t \partial^{\alpha_1}\dots\partial^{\alpha_n} F(X_r) d\widehat{X}^{(\alpha_1\dots\alpha_n)} \stackrel{\text{def}}{=} \left\langle 1, \int_s^t \partial^{\alpha_1}\dots\partial^{\alpha_n} F(\mathbf{X}_r) d\widehat{X}^{(\alpha_1\dots\alpha_n)} \right\rangle \ ,$$

since $\mathbf{X}$ is an $\mathbf{X}$-controlled rough path and hence an $\widehat{\mathbf{X}}$-controlled rough path.

The rough path construction of the general Itô formula is best illustrated by looking at the simplest non-trivial cases.

Example 5.2.4. We first look at the Brownian motion case, or indeed any path X in $\mathbb{R}^d$ with Hölder exponent $1/3 < \gamma \leq 1/2$ and with a branched rough path $\mathbf{X}$ above X. Let $F : \mathbb{R}^d \to \mathbb{R}^d$ be some sufficiently smooth function. As usual, we start the Itô formula with a Taylor expansion

$$\begin{aligned} \delta F(X)_{st} &= DF(X_s) : \delta X_{st} + \frac{1}{2} D^2 F(X_s) : (\delta X_{st}, \delta X_{st}) + R^F_{st} \\ &= \partial^i F(X_s) \langle \mathbf{X}_{st}, \bullet_i \rangle + \frac{1}{2} \partial^i \partial^i F(X_s) \langle \mathbf{X}_{st}, \bullet_i \bullet_j \rangle + o(|t-s|) \ , \end{aligned}$$

where $|R^F_{st}| \leq C|\delta Y_{st}|^3 = o(|t-s|)$ and we sum over all $i, j = 1 \ldots d$. The second step in building an Itô formula is to construct the leading integral term. It is clear that $\partial^i F(X_s) \langle \mathbf{X}_{st}, \bullet_i \rangle = \partial^i F(X_s) \delta X^i_{st}$ is the lowest order term of this leading integral. By definition, we have that

$$\begin{aligned} \int_s^t \partial^i F(X_s) dX^i_r &\overset{\text{def}}{=} \left\langle 1, \int_s^t \partial^i F(\mathbf{X}_s) dX^i_r \right\rangle \\ &= \partial^i F(X_s) \langle \mathbf{X}_{st}, \bullet_i \rangle + \partial^i \partial^j F(X_s) \langle \mathbf{X}_{st}, \overset{\bullet_j}{\bullet_i} \rangle + r_{st} \ , \end{aligned} \tag{5.6}$$

where i is fixed and we sum over $j = 1 \ldots d$ and where $|r_{st}| = o(|t-s|)$. Hence, we obtain

$$\begin{aligned} &\delta F(Y)_{st} - \int_s^t DF(Y_r) \cdot dX_r \\ &= \sum_{i,j=1}^{d} \left(\frac{1}{2} \partial^i \partial^j F(X_s) \langle \mathbf{X}_{st}, \bullet_i \bullet_j \rangle - \partial^i \partial^j F(X_s) \langle \mathbf{X}_{st}, \overset{\bullet_j}{\bullet_i} \rangle \right) + o(|t-s|) \\ &= \sum_{i,j=1}^{d} \frac{1}{2} \partial^i \partial^j F(X_s) \left\langle \mathbf{X}_{st}, \bullet_i \, \bullet_j - \overset{\bullet_j}{\bullet_i} - \overset{\bullet_i}{\bullet_j} \right\rangle + o(|t-s|) \ . \end{aligned} \tag{5.7}$$

In the last line above, we have replaced $\overset{\bullet_j}{\bullet_i}$ with its symmetric part, this is permitted by the fact that we sum over all $i, j = 1 \ldots d$. The remaining term plays the role of the 'bracket'. Indeed, if we let

$$\langle\!\langle ij \rangle\!\rangle \overset{\text{def}}{=} \bullet_i \, \bullet_j - \overset{\bullet_j}{\bullet_i} - \overset{\bullet_i}{\bullet_j} \ ,$$

then we obtain an integration by parts formula,

$$\left\langle \mathbf{X}_{st}, \overset{\bullet_j}{\bullet_i} + \overset{\bullet_i}{\bullet_j} \right\rangle = \langle \mathbf{X}_{st}, \bullet_i \bullet_j \rangle - \langle \mathbf{X}_{st}, \langle\!\langle ij \rangle\!\rangle \rangle \ ,$$

or in the more traditional (but formal) notation

$$\int_s^t \delta X^j_{sr} dX^i_r + \int_s^t \delta X^i_{sr} dX^j_r = \delta X^i_{st} \delta X^j_{st} - \langle \mathbf{X}_{st}, \langle\!\langle ij \rangle\!\rangle \rangle \ .$$

In the case where X is Brownian motion and $\mathbf{X}$ is constructed using Itô integrals, it is clear that $\langle \mathbf{X}_{st}, \langle\!\langle ij \rangle\!\rangle \rangle$ is precisely the quadratic variation of (X^i, X^j). In fact, one can easily check that $\langle \mathbf{X}_{st}, \langle\!\langle ij \rangle\!\rangle \rangle$ is the increment of a path for any X and any branched rough path $\mathbf{X}$. This follows from the simple calculation

$$\langle \mathbf{X}_{st}, \langle\!\langle ij \rangle\!\rangle \rangle = \langle \mathbf{X}_{su}, \langle\!\langle ij \rangle\!\rangle \rangle + \langle \mathbf{X}_{ut}, \langle\!\langle ij \rangle\!\rangle \rangle \ ,$$

for every $s \leq u \leq t$. We then define the generalised quadratic variation (or bracket) path $\widehat{X}^{(ij)}$ by

$$\delta \widehat{X}^{(ij)}_{st} \stackrel{\text{def}}{=} \langle \mathbf{X}_{st}, \langle\!\langle ij \rangle\!\rangle \rangle \ ,$$

and it is clear from the definition that $|\delta \widehat{X}^{(ij)}_{st}| \leq C|t-s|^{2\gamma}$. Let us define a path $\widehat{X} : [0,T] \to \mathbb{R}^d \oplus \mathbb{R}^{d^2}$ by

$$\widehat{X} = X \oplus (\widehat{X}^{(ij)})_{i,j=1\ldots d} \ ,$$

so that $\widehat{X}$ contains the original path X along with the new bracket elements. Now, to complete the Itô formula, we need to turn the remaining term

$$\frac{1}{2} \partial^i \partial^j F(X_s) \delta \widehat{X}^{(ij)}_{st}$$

into an integral. This means we must know how to compute $\int \mathbf{Z} d\widehat{X}$ where $\mathbf{Z}$ is an $\mathbf{X}$-controlled rough path and $d\widehat{X}$ denotes the increment of any component of $\widehat{X}$. This is achieved by building a branched rough path $\widehat{\mathbf{X}}$ over $\widehat{X}$, that is an extension of $\mathbf{X}$. That is, $\widehat{\mathbf{X}}$ is a branched rough path built over the Connes-Kreimer Hopf algebra $\widehat{\mathcal{H}}^{(2)}$ generated by the extended alphabet

$$\{1, \ldots, d\} \cup \{(ij) : i,j = 1, \ldots, d\} \ ,$$

such that

1. $\langle \widehat{\mathbf{X}}_{st}, h \rangle = \langle \mathbf{X}_{st}, h \rangle$ when $h \in \mathcal{H}$,
2. $\langle \widehat{\mathbf{X}}_{st}, \bullet_{(ij)} \rangle = \delta \widehat{X}^{(ij)}_{st}$ for $i,j = 1 \ldots d$.

We call $\widehat{\mathbf{X}}$ a *bracket extension* of $\mathbf{X}$. In this example, the extension is somewhat trivial since for an $\mathbf{X}$-controlled rough path $\mathbf{Z}$, we can always define $\int \mathbf{Z} d\widehat{X}^{(ij)}$ as a Young integral [You36], due to the fact that $\widehat{X}^{(ij)}$ is 2γ-Hölder and $\gamma + 2\gamma > 1$. Thus, all the extra components $\int \delta\widehat{X} d\widehat{X}$ can be canonically defined. In particular we have that

$$\int_s^t \frac{1}{2}\partial^i\partial^i F(X_r) d\widehat{X}_r^{(ij)} \stackrel{\text{def}}{=} \frac{1}{2}\partial^i\partial^i F(X_r)\langle \widehat{\mathbf{X}}_{st}, \bullet_{(ij)}\rangle + o(|t-s|) \qquad (5.8)$$

$$= \lim_{\pi\downarrow 0} \sum_{[u,v]\in\pi} \frac{1}{2}\partial^i\partial^i F(X_u)\langle \widehat{\mathbf{X}}_{uv}, \bullet_{(ij)}\rangle \; .$$

We can now complete the formula, combining (5.7) and (5.8) we have that

$$\delta F(X)_{st} - \int_s^t DF(X_r)\cdot dX_r - \sum_{i,j=1}^d \int_s^t \frac{1}{2}\partial^i\partial^j F(X_r) d\widehat{X}_r^{(ij)} = o(|t-s|) \; .$$

But since the left hand side is an increment and the right hand side is $o(|t-s|)$, the left hand side must be identically zero. We therefore recover the change of variables formula

$$\delta F(X)_{st} = \int_s^t DF(X_r)\cdot dX_r + \int_s^t \frac{1}{2} D^2 F(X_r) : d\widehat{X}_r \; , \qquad (5.9)$$

where we use the shorthand (5.5). Clearly this coincides, at least structurally, with the usual Itô formula. In the case where X is a BM and $\mathbf{X}$ is the Itô rough path above X, we would of course recover the usual Itô formula. Note that (5.26) is simply the controlled rough path analogue of the Itô formula given in [LQ02] for non-geometric rough paths of finite p-variation for $2 \leq p < 3$.

Before looking at the next example, we first make a few observations

1. When $1/3 < \gamma \leq 1/2$, the bracket extension $\widehat{\mathbf{X}}$ is determined uniquely by $\mathbf{X}$. This means that there is only one possible change of variables formula once we have decided how to define the leading order integral. For smaller values of γ, In the next example, we will see that there are lots of ways of defining $\widehat{\mathbf{X}}$ and hence many different change of variables formulae.

2. The bracket polynomial $\langle\langle ij \rangle\rangle$ and path $\widehat{X}^{(ij)}$ arise quite naturally in the construction. In particular, it is the only term remaining in (5.7) and therefore must be the bracket driving the correction integral. Surprisingly, we will see that even when γ is smaller, the bracket still arises naturally, by simply matching derivatives of F.

Example 5.2.5. We will now consider the case $1/4 < \gamma \leq 1/3$, so that $N = 3$. Hence, the terms that are $O(|t-s|^{3\gamma})$ can no longer be ignored. The Taylor expansion takes the form

$$\begin{aligned}\delta F(X)_{st} &= \partial^i F(X_s)\langle \mathbf{X}_{st}, \bullet_i\rangle + \frac{1}{2}\partial^i\partial^j F(X_s)\langle \mathbf{X}_{st}, \bullet_i\bullet_j\rangle \\ &\quad + \frac{1}{6}\partial^i\partial^j\partial^k F(X_s)\langle \mathbf{X}_{st}, \bullet_i\,\bullet_j\,\bullet_k\rangle + o(|t-s|)\ ,\end{aligned}$$

where we sum over all $i,j,k = 1\ldots d$. The leading order integral is defined by

$$\begin{aligned}\int_s^t \partial^i F(X_r) dX_r^i &\overset{\text{def}}{=} \left\langle 1, \int_s^t \partial^i F(\mathbf{X}_r) dX_r^i \right\rangle \\ &= \partial^i F(X_s)\,\langle \mathbf{X}_{st}, \bullet_i\rangle + \partial^i\partial^j F(X_s)\left\langle \mathbf{X}_{st}, [\bullet_j]_i \right\rangle \\ &\quad + \frac{1}{2}\partial^i\partial^j F(X_s)\langle \mathbf{X}_{st}, [\bullet_j\bullet_k]_i\ \rangle + o(|t-s|)\ .\end{aligned}$$

For a fixed i and where we sum over all $j,k = 1\ldots d$. Hence, when we isolate the leading integral, the ∂F terms cancel out, leaving

$$\begin{aligned}\delta F(X)_{st} - &\int_s^t DF(X_r)\cdot dX_r \\ &= \frac{1}{2}\partial^i\partial^j F(X_s)\langle \mathbf{X}_{st}, \langle\!\langle ij\rangle\!\rangle\rangle \\ &\quad + \frac{1}{6}\partial^i\partial^j\partial^k F(X_s)\langle \mathbf{X}_{st}, \bullet_i\,\bullet_j\,\bullet_k - [\bullet_i\bullet_j]_k\ - [\bullet_i\bullet_k]_j\ - [\bullet_j\bullet_k]_i\ \ \rangle + o(|t-s|)\ ,\end{aligned}$$

where we sum over $i,j,k = 1\ldots d$ and we replace $1/2\,[\bullet_j\bullet_k]_i$ with the symmetrised version. As in the previous example, we extend the path X to $\widehat{X}^{(2)}$,

$$\widehat{X}^{(2)} = X \oplus (\widehat{X}^{(ij)})_{i,j=1\ldots d}\ ,$$

by adding the bracket path elements $\delta\widehat{X}^{(ij)_{st}} = \langle \mathbf{X}_{st}, \langle\!\langle ij\rangle\!\rangle\rangle$. Again following the previous example, we wish to find a branched rough path $\widehat{\mathbf{X}}^{(2)}$ above $\widehat{X}^{(2)}$ by extending $\mathbf{X}$. Thus, $\widehat{\mathbf{X}}^{(2)}$ is a branched rough path in the Hopf algebra $\widehat{\mathcal{H}}^{(2)}$ defined in the previous example. Now, since $\gamma + 2\gamma \leq 1$, the object

$$\langle \widehat{\mathbf{X}}_{st}, [\bullet_k]_{(ij)}\rangle = \int_s^t \delta X_{sr}^k d\widehat{X}_r^{(ij)}$$

cannot be defined canonically, which means that $\widehat{\mathbf{X}}^{(2)}$, if it exists, can be defined in many different ways. Actually, it follows from Corollary 4.2.16 that $\widehat{\mathbf{X}}^{(2)}$ must exist, since $\widehat{\mathbf{X}}^{(2)}$ is an extension of $\mathbf{X}$ obtained by adding components to the underlying path. This will be made clear in Proposition 5.2.14.

Suppose we are given such an $\widehat{\mathbf{X}}^{(2)}$. Since a smooth function of an $\mathbf{X}$-controlled rough path is an $\widehat{\mathbf{X}}^{(2)}$-controlled rough path, we can define

$$\begin{aligned}\int_s^t \frac{1}{2}\partial^i\partial^j F(X_r)d\widehat{X}_r^{(ij)} &\stackrel{\text{def}}{=} \left\langle 1, \int_s^t \frac{1}{2}\partial^i\partial^j F(\mathbf{X}_r)d\widehat{X}_r^{(ij)}\right\rangle \\ &= \frac{1}{2}\partial^i\partial^j F(X_s)\langle \mathbf{X}_{st}, \langle\!\langle ij\rangle\!\rangle\rangle \\ &\quad + \frac{1}{2}\partial^i\partial^j\partial^k F(X_s)\left\langle \widehat{\mathbf{X}}_{st}, \overset{\bullet_k}{\bullet_{(ij)}}\right\rangle + o(|t-s|)\ ,\end{aligned}$$

where i, j are fixed and we sum over $k = 1 \ldots d$. Thus, we obtain

$$\begin{aligned}&\delta F(X)_{st} - \int_s^t DF(X_r)\cdot dX_r - \sum_{i,j=1}^d \int_s^t \frac{1}{2}\partial^i\partial^j F(X_r)d\widehat{X}_r^{(ij)} \\ &= \sum_{i,j,k=1}^d \frac{1}{6}\partial^i\partial^j\partial^k F(X_s) \\ &\times \left(\left\langle \mathbf{X}_{st}, \bullet_i\bullet_j\bullet_k - \overset{\bullet_i\bullet_j}{\bullet_k} - \overset{\bullet_i\bullet_k}{\bullet_j} - \overset{\bullet_j\bullet_k}{\bullet_i}\right\rangle - \left\langle \widehat{\mathbf{X}}_{st}, \overset{\bullet_k}{\bullet_{(ij)}} - \overset{\bullet_i}{\bullet_{(jk)}} - \overset{\bullet_j}{\bullet_{(ik)}}\right\rangle\right) \\ &+ o(|t-s|)\ .\end{aligned}$$

And using the fact that $\mathbf{X}$ and $\widehat{\mathbf{X}}^{(2)}$ agree on $h \in \mathcal{H}$, this equals

$$\begin{aligned}&\sum_{i,j,k=1}^d \frac{1}{6}\partial^i\partial^j\partial^k F(X_s) \\ &\times \left\langle \widehat{\mathbf{X}}_{st}^{(2)}, \bullet_i\bullet_j\bullet_k - \overset{\bullet_i\bullet_j}{\bullet_k} - \overset{\bullet_i\bullet_k}{\bullet_j} - \overset{\bullet_j\bullet_k}{\bullet_i} - \overset{\bullet_k}{\bullet_{(ij)}} - \overset{\bullet_i}{\bullet_{(jk)}} - \overset{\bullet_j}{\bullet_{(ik)}}\right\rangle \\ &+ o(|t-s|)\ .\end{aligned}$$

The remaining terms become integrals driven by new bracket processes, so we set

$$\langle\!\langle ijk\rangle\!\rangle \stackrel{\text{def}}{=} \bullet_i\bullet_j\bullet_k - \overset{\bullet_i\bullet_j}{\bullet_k} - \overset{\bullet_i\bullet_k}{\bullet_j} - \overset{\bullet_j\bullet_k}{\bullet_i} - \overset{\bullet_k}{\bullet_{(ij)}} - \overset{\bullet_i}{\bullet_{(jk)}} - \overset{\bullet_j}{\bullet_{(ik)}}\ .$$

As with the previous brackets, one can check that

$$\langle \widehat{\mathbf{X}}_{st}^{(2)}, \langle\!\langle ijk\rangle\!\rangle\rangle = \langle \widehat{\mathbf{X}}_{su}^{(2)}, \langle\!\langle ijk\rangle\!\rangle\rangle\ , + \langle \widehat{\mathbf{X}}_{ut}^{(2)}, \langle\!\langle ijk\rangle\!\rangle\rangle$$

for all $s \le u \le t$. We can therefore define a 3γ-Hölder path $\widehat{X}^{(ijk)}$ by

$$\delta\widehat{X}_{st}^{(ijk)} \stackrel{\text{def}}{=} \langle \widehat{\mathbf{X}}_{st}^{(2)}, \langle\!\langle ijk\rangle\!\rangle\rangle\ .$$

We now repeat the previous step, by defining

$$\widehat{X}^{(3)} = \widehat{X}^{(2)} \oplus (\widehat{X}^{(ijk)})_{i,j,k=1\dots d}$$

and let $\widehat{\mathbf{X}}^{(3)}$ be an extension of $\widehat{\mathbf{X}}^{(2)}$, lying above the path $\widehat{X}^{(3)}$. That is, $\widehat{\mathbf{X}}^{(3)}$ is a branched rough path on the Connes-Kreimer Hopf algebra $\widehat{\mathcal{H}}^{(3)}$ generated by the alphabet

$$\{1,\dots,d\} \cup \{(ij) : i,j = 1,\dots,d\} \cup \{(ijk) : i,j,k = 1\dots d\} \ ,$$

satisfying

1. $\langle \widehat{\mathbf{X}}^{(3)}_{st}, h\rangle = \langle \mathbf{X}^{(2)}_{st}, h\rangle$ when $h \in \widehat{\mathcal{H}}^{(2)}$
2. $\langle \widehat{\mathbf{X}}_{st}, \bullet_{(ijk)}\rangle = \delta\widehat{X}^{(ijk)}_{st}$ for $i,j,k = 1\dots d$.

Now, since $\gamma + 3\gamma > 1$, objects of the form

$$\langle \widehat{\mathbf{X}}^{(3)}_{st}, \overset{\bullet^{l}}{\bullet}_{(ijk)}\rangle = \int_s^t \delta X^l_{sr} d\widehat{X}^{(ijk)}_r$$

can be constructed canonically as Young integrals and likewise for all new components of $\widehat{\mathbf{X}}^{(3)}$. That is, given $\widehat{\mathbf{X}}^{(2)}$ we can canonically construct $\widehat{\mathbf{X}}^{(3)}$. Since we don't need to add any more bracket terms to our bracket extension, we set $\widehat{X} = \widehat{X}^{(3)}$ and $\widehat{\mathbf{X}} = \widehat{\mathbf{X}}^{(3)}$. Proceeding as above, we obtain the Itô formula

$$\delta F(X)_{st} = \int_s^t DF(X_r)\cdot dX_r + \int_s^t \frac{1}{2}D^2F(X_r) : d\widehat{X}_r + \int_s^t \frac{1}{6}D^3F(X_r) : d\widehat{X}_r \ ,$$

where we use the notation (5.5).

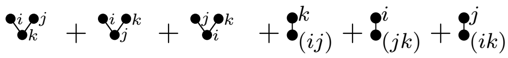
Remark 5.2.6. In the first example, we saw that the bracket extension $\widehat{\mathbf{X}}$ is constructed canonically from $\mathbf{X}$, which implies that there is a *unique* (canonical) Itô formula for a given $\mathbf{X}$. However, in the second example the bracket extension can be constructed in several ways, and in particular there are many ways of interpreting the integral

$$\int_s^t \frac{1}{2}\partial^i\partial^j F(X_r) d\widehat{X}^{(ij)}_r \ ,$$

which was also a feature of the change of variables formula given in Example 5.1.2.

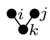
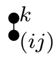
Remark 5.2.7. The bracket extension contains components that are never used in the Itô formula. In fact, we only ever need to compute integrals where the integrand is controlled by $\mathbf{X}$, hence we never need to know $\int \delta\widehat{X}^{(ij)} dX^k$ for instance. However, we find the notion of a bracket extension more natural than a condition that only requires the existence of certain components of $\widehat{\mathbf{X}}$.

The bracket polynomials $\langle\!\langle ij \rangle\!\rangle$ and $\langle\!\langle ijk \rangle\!\rangle$ fit into a general formula quite easily. For instance, the expression



can we written as

$$J_2\left(\bullet_i \bullet_j \tilde{\otimes} \bullet_k + \bullet_i \bullet_k \tilde{\otimes} \bullet_j + \bullet_i \bullet_j \tilde{\otimes} \bullet_k + \bullet_k \tilde{\otimes} \bullet_i \bullet_j + \bullet_i \tilde{\otimes} \bullet_j \bullet_k + \bullet_j \tilde{\otimes} \bullet_i \bullet_k \right) , \quad (5.10)$$

where J_2 is a linear map satisfying $J_2(\bullet_i \bullet_j \tilde{\otimes} \bullet_k) =$  and $J_2(\bullet_k \tilde{\otimes} \bullet_i \bullet_j) =$  . Hence, J_2 is the labelling map. Moreover, the product inside the brackets in (5.10) is simply the sum over all ways of dividing the objects from $\{\bullet_i, \bullet_j, \bullet_k\}$ into two non-trivial sets. One can check that this is also described by the reduced coproduct, in that

$$\Delta'(\bullet_i \bullet_j \bullet_k) = \bullet_i \bullet_j \tilde{\otimes} \bullet_k + \bullet_i \bullet_k \tilde{\otimes} \bullet_j + \bullet_i \bullet_j \tilde{\otimes} \bullet_k + \bullet_k \tilde{\otimes} \bullet_i \bullet_j + \bullet_i \tilde{\otimes} \bullet_j \bullet_k + \bullet_j \tilde{\otimes} \bullet_i \bullet_k ,$$

hence, we can write

$$\langle\!\langle ijk \rangle\!\rangle = \bullet_i \bullet_j \bullet_k - J_2 \Delta'(\bullet_i \bullet_j \bullet_k) .$$

This idea will be made precise below.

5.2.8 The bracket extension of X

Before defining bracket extensions, we must define the Hopf algebra in which they live. As seen in Example 5.2.5, if $N = 3$ then this Hopf algebra must contain labels (ij) and (ijk), in addition to the usual single index labels. Thus we define $\widehat{\mathcal{H}}^{(n)}$ to be the Connes-Kreimer Hopf algebra generated by the alphabet

$$\mathcal{A}_n = \{1, \dots, d\} \cup \{(a_1 \dots a_k) : a_i = 1 \dots d \text{ and } 2 \leq k \leq n\} .$$

We then set $\widehat{\mathcal{H}} = \widehat{\mathcal{H}}^{(N)}$.

Remark 5.2.9. In the simple case, it is perhaps more natural to define

$$\mathcal{A}_n = \{(a_1 \dots a_k) : a_i = 1 \dots d \text{ and } 1 \leq k \leq n\} ,$$

and simply identify the label (i) with i. However, in the general case we will see that it is important to distinguish between the single and multi-index labels as they range over different values.

This allows us to define the bracket polynomials

$$\langle\!\langle \gamma_1 \dots \gamma_n \rangle\!\rangle \stackrel{\text{def}}{=} \bullet_{\gamma_1} \cdots \bullet_{\gamma_n} - \sum_{(a,b)} [\bullet_{a_1} \cdots \bullet_{a_k}]_{(b_1 \dots b_{n-k})} , \tag{5.11}$$

where we sum over all ways of splitting the set $\{\gamma_1, \dots, \gamma_n\}$ into two non-empty sets $\{a_1, \dots, a_k\}$ and $\{b_1, \dots, b_{n-k}\}$ for any $1 \leq k \leq n-1$. As mentioned above, it is easy to check that this yields the correct expressions for $\langle\!\langle ij \rangle\!\rangle$ and $\langle\!\langle ijk \rangle\!\rangle$. There is an equivalent way of defining bracket extensions that is computationally more useful. Firstly, we define $J_2(\bullet_{\alpha_1} \cdots \bullet_{\alpha_n} \tilde{\otimes} \bullet_{\beta_1} \cdots \bullet_{\beta_m}) = [\bullet_{\alpha_1} \cdots \bullet_{\alpha_n}]_{(\beta_1 \dots \beta_m)}$, and then define J_2 on sums of terms like $\bullet_{\alpha_1} \cdots \bullet_{\alpha_n} \tilde{\otimes} \bullet_{\beta_1} \cdots \bullet_{\beta_m}$ by imposing linearity. Secondly, one can easily check that

$$\Delta'(\bullet_{\gamma_1} \cdots \bullet_{\gamma_n}) = \sum_{(a,b)} \bullet_{a_1} \cdots \bullet_{a_k} \tilde{\otimes} \bullet_{b_1} \cdots \bullet_{b_{n-k}} .$$

It follows that

$$\langle\!\langle \gamma_1 \dots \gamma_n \rangle\!\rangle = \bullet_{\gamma_1} \cdots \bullet_{\gamma_n} - J_2 \Delta'(\bullet_{\gamma_1} \cdots \bullet_{\gamma_n}) . \tag{5.12}$$

We will use both formulations in the sequel, depending on which is more convenient.

Remark 5.2.10. The definition of $\langle\!\langle \gamma_1 \dots \gamma_n \rangle\!\rangle$ may seem a bit arbitrary at present, but it is actually quite natural in light of the procedure seen in Examples 5.2.4 and 5.2.5. In fact, all the terms in $\langle\!\langle \gamma_1 \dots \gamma_n \rangle\!\rangle$ arise by simply grouping together the polynomial of $\widehat{\mathbf{X}}$ that has the coefficient $\partial^{\gamma_1} \dots \partial^{\gamma_n} F(X_s)$ This will be clear in the proof of the change of variables formula.

The following properties of bracket polynomials follow easily from the definiton.

1. The bracket polynomial is *symmetric*, since, according to (5.12), it is simply a function of the product $\bullet_{\gamma_1} \cdots \bullet_{\gamma_n}$, which is commutative.
2. We have that $\langle\!\langle \gamma_1 \dots \gamma_n \rangle\!\rangle \in \widehat{\mathcal{H}}^{(n-1)}$ for every $n \geq 2$.

The following property is less easy to check, but crucial in defining a bracket extension.

Lemma 5.2.11. *We have that*

$$\begin{aligned} \Delta \langle\!\langle \gamma_1 \dots \gamma_n \rangle\!\rangle &= \langle\!\langle \gamma_1 \dots \gamma_n \rangle\!\rangle \tilde{\otimes} 1 + 1 \tilde{\otimes} \langle\!\langle \gamma_1 \dots \gamma_n \rangle\!\rangle \\ &\quad + \sum_{(a,b)} (\bullet_{a_1} \cdots \bullet_{a_k}) \tilde{\otimes} \Big(\langle\!\langle b_1 \dots b_{n-k} \rangle\!\rangle - \bullet_{(b_1 \dots b_{n-k})} \Big) , \end{aligned}$$

where we sum over all (a,b) *as in* (5.11).

In some sense, this lemma tells us how the coproduct Δ commutes with the map $\langle\!\langle \cdot \rangle\!\rangle$. We can now define a bracket extension.

Definition 5.2.12. We define a *bracket extension* of $\mathbf{X}$ as any branched rough path $\widehat{\mathbf{X}}$ on $\widehat{\mathcal{H}}$ that is an extension of $\mathbf{X}$ in the sense that $\langle \widehat{\mathbf{X}}_{st}, h\rangle = \langle \mathbf{X}_{st}, h\rangle$ for any $h \in \mathcal{H} \subset \widehat{\mathcal{H}}$ and satisfies

$$\left\langle \widehat{\mathbf{X}}_{st}, \bullet_{(\gamma_1 \dots \gamma_n)} \right\rangle = \left\langle \widehat{\mathbf{X}}_{st}, \langle\!\langle \gamma_1 \dots \gamma_n \rangle\!\rangle \right\rangle \ , \tag{5.13}$$

for all $\gamma_i = 1 \dots d$ and $2 \leq n \leq N$.

It follows from the above properties of $\langle\!\langle \gamma_1 \dots \gamma_n \rangle\!\rangle$ that the bracket paths

$$\widehat{X}^{(\gamma_1 \dots \gamma_n)}_{st} = \langle \widehat{\mathbf{X}}_{st}, \bullet_{(\gamma_1 \dots \gamma_n)} \rangle \ ,$$

are symmetric in the multi-index, which one would expect for a bracket. Hence, is natural to make this assumption on the higher order components too, by requiring that for $\tau \in \widehat{\mathcal{H}}$ the object $\langle \widehat{\mathbf{X}}_{st}, \tau \rangle$ is unchanged by choosing a vertex of τ and permuting the entries in the multi-index label. Secondly, since $\langle\!\langle \gamma_1 \dots \gamma_n \rangle\!\rangle \in \widehat{\mathcal{H}}^{(n-1)}$, it follows that $\widehat{X}^{(\gamma_1 \dots \gamma_n)}$ is determined by the restriction of $\widehat{\mathbf{X}}$ to $\widehat{\mathcal{H}}^{(n-1)}$. Hence, the bracket extension has quite a lot of redundancy in it. This property will be important when we "construct" a bracket extension.

Remark 5.2.13. An important thing to notice is that having a bracket extension allows us to replace products of components of $\mathbf{X}$ as sums of components of $\widehat{\mathbf{X}}$. More specifically, we have that

$$\langle \mathbf{X}_{st}, \bullet_{\gamma_1} \cdots \bullet_{\gamma_n} \rangle = \left\langle \widehat{\mathbf{X}}_{st}, \sum_{(a,b)} [\bullet_{a_1} \cdots \bullet_{a_k}]_{(b_1 \dots b_{k-n})} + \bullet_{(\gamma_1 \dots \gamma_n)} \right\rangle \ ,$$

and the right hand side clearly has no products of components of $\widehat{\mathbf{X}}$, but is instead a linear combination of components of $\widehat{\mathbf{X}}$. This is akin to the geometric property of rough paths, which allows one to write products of components of $\mathbf{X}$ as a sum of the linear tree components of $\mathbf{X}$. One can also think of this as a generalised *integration by parts* formula. This will clearly be useful when building an Itô formula, since we can rewrite the products that appear in a Taylor expansion as $\widehat{\mathbf{X}}$ integrals.

Of course, it is not obvious that the definition is even well-posed, let alone whether or not we can find such an extension. The following proposition shows that we can always find bracket extensions.

Proposition 5.2.14. *A bracket extension $\widehat{\mathbf{X}}$ of $\mathbf{X}$ exists.*

Proof. We will proceed by induction. Suppose $\{\widehat{\mathbf{X}}^{(m)}\}_{1\leq m\leq n}$ is a sequence of branched rough paths on $\{\widehat{\mathcal{H}}^{(m)}\}_{1\leq m\leq n}$ such that, for each $1 \leq m \leq n-1$, we have

$$\langle \widehat{\mathbf{X}}^{(m+1)}_{st}, \bullet_{(\gamma_1\cdots\gamma_k)} \rangle = \langle \widehat{\mathbf{X}}^{(m)}_{st}, \langle\!\langle \gamma_1 \dots \gamma_k \rangle\!\rangle \rangle \ , \tag{5.14}$$

for $\tau_i = 1 \dots d$ with for $2 \leq k \leq m$ and

$$\langle \widehat{\mathbf{X}}^{(m+1)}, h \rangle = \langle \widehat{\mathbf{X}}^{(m)}_{st}, h \rangle$$

for $h \in \widehat{\mathcal{H}}^{(m)}$. In particular, each element $\widehat{\mathbf{X}}^{(m)}$ in the sequence satisfies (5.13) for all $(\gamma_1 \dots \gamma_k)$ with $k \leq m$. We will now construct the next element in the sequence $\widehat{\mathbf{X}}^{(n+1)}$. We first set

$$\langle \widehat{\mathbf{X}}^{(n+1)}_{st}, \bullet_{(\gamma_1\cdots\gamma_{n+1})} \rangle = \langle \widehat{\mathbf{X}}^{(n)}_{st}, \langle\!\langle \gamma_1 \dots \gamma_{n+1} \rangle\!\rangle \rangle \ . \tag{5.15}$$

For this to be a valid component of a branched rough path, we require that

$$\langle \widehat{\mathbf{X}}^{(n)}_{st}, \langle\!\langle \gamma_1 \dots \gamma_{n+1} \rangle\!\rangle \rangle = \langle \widehat{\mathbf{X}}^{(n)}_{su}, \langle\!\langle \gamma_1 \dots \gamma_{n+1} \rangle\!\rangle \rangle + \langle \widehat{\mathbf{X}}^{(n)}_{ut}, \langle\!\langle \gamma_1 \dots \gamma_{n+1} \rangle\!\rangle \rangle \ . \tag{5.16}$$

We will delay the verification of this fact for a moment. This implies the existence of paths $\widehat{X}^{(\gamma_1\cdots\gamma_{n+1})}$ defined by

$$\delta \widehat{X}^{(\gamma_1\cdots\gamma_{n+1})}_{st} = \langle \widehat{\mathbf{X}}^{(n)}_{st}, \langle\!\langle \gamma_1 \dots \gamma_{n+1} \rangle\!\rangle \rangle$$

for all such $(\gamma_1 \dots \gamma_{n+1})$. Since $\widehat{\mathcal{H}}^{(n)}, \widehat{\mathcal{H}}^{(n+1)}$ are generated by the alphabets $\mathcal{A}_n, \mathcal{A}_{n+1}$, and $\mathcal{A}_n \subset \mathcal{A}_{n+1}$, it follows from Corollary 4.2.16 that there exists a branched rough path $\widehat{\mathbf{X}}^{(n+1)}$ on $\widehat{\mathcal{H}}^{(n+1)}$ obtained from $\widehat{\mathbf{X}}^{(n)}$ by adding the path elements $\widehat{X}^{(\gamma_1\cdots\gamma_{n+1})}$. More precisely, $\widehat{\mathbf{X}}^{(n+1)}$ satisfies

$$\langle \widehat{\mathbf{X}}^{(n+1)}_{st}, \bullet_{(\gamma_1\cdots\gamma_{n+1})} \rangle = \delta \widehat{X}^{(\gamma_1\cdots\gamma_{n+1})}_{st} \ ,$$

for all such $(\gamma_1 \dots \gamma_{n+1})$ and

$$\langle \widehat{\mathbf{X}}^{(n+1)}_{st}, h \rangle = \langle \widehat{\mathbf{X}}^{(n)}_{st}, h \rangle$$

for all $h \in \widehat{\mathcal{H}}^{(n)}$. This proves the induction. If we set $\widehat{\mathbf{X}} = \widehat{\mathbf{X}}^{(N)}$, then the claim follows upon proving (5.16). But this follows from Lemma 5.2.11. Indeed, using the definition

(5.11), we have that

$$\left\langle \widehat{\mathbf{X}}^{(n+1)}_{st}, \bullet_{(\gamma_1\ldots\gamma_{n+1})} \right\rangle = \left\langle \widehat{\mathbf{X}}^{(n)}_{st}, \langle\!\langle \gamma_1 \ldots \gamma_{n+1} \rangle\!\rangle \right\rangle = \left\langle \widehat{\mathbf{X}}^{(n)}_{su} \tilde{\otimes} \widehat{\mathbf{X}}^{(n)}_{ut}, \Delta \langle\!\langle \gamma_1 \ldots \gamma_{n+1} \rangle\!\rangle \right\rangle$$
$$= \sum_{(a,b)} \left\langle \widehat{\mathbf{X}}^{(n)}_{su} \tilde{\otimes} \widehat{\mathbf{X}}^{(n)}_{ut}, \bullet_{a_1} \cdots \bullet_{a_k} \tilde{\otimes} \Big(\langle\!\langle b_1 \ldots b_{n+1-k} \rangle\!\rangle - \bullet_{(b_1\ldots b_{n+1-k})} \Big) \right\rangle$$
$$+ \left\langle \widehat{\mathbf{X}}^{(n)}_{su}, \langle\!\langle \gamma_1 \ldots \gamma_{n+1} \rangle\!\rangle \right\rangle + \left\langle \widehat{\mathbf{X}}^{(n)}_{ut}, \langle\!\langle \gamma_1 \ldots \gamma_{n+1} \rangle\!\rangle \right\rangle$$
$$= \left\langle \widehat{\mathbf{X}}^{(n)}_{su} \tilde{\otimes} \widehat{\mathbf{X}}^{(n)}_{ut}, \bullet_{a_1} \cdots \bullet_{a_k} \tilde{\otimes} \Big(\langle\!\langle b_1 \ldots b_{n+1-k} \rangle\!\rangle - \bullet_{(b_1\ldots b_{n+1-k})} \Big) \right\rangle$$
$$+ \left\langle \widehat{\mathbf{X}}^{(n+1)}_{su}, \bullet_{(\gamma_1\ldots\gamma_{n+1})} \right\rangle + \left\langle \widehat{\mathbf{X}}^{(n+1)}_{ut}, \bullet_{(\gamma_1\ldots\gamma_{n+1})} \right\rangle \ .$$

And (5.31) follows from the fact that

$$\left\langle \widehat{\mathbf{X}}^{(n)}_{su} \tilde{\otimes} \widehat{\mathbf{X}}^{(n)}_{ut}, \bullet_{a_1} \cdots \bullet_{a_k} \tilde{\otimes} \Big(\langle\!\langle b_1 \ldots b_{n+1-k} \rangle\!\rangle - \bullet_{(b_1\ldots b_{n+1-k})} \Big) \right\rangle$$
$$= \left\langle \widehat{\mathbf{X}}^{(n)}_{su}, \bullet_{a_1} \cdots \bullet_{a_k} \right\rangle \left\langle \widehat{\mathbf{X}}^{(n)}_{ut}, \langle\!\langle b_1 \ldots b_{n+1-k} \rangle\!\rangle - \bullet_{(b_1\ldots b_{n+1-k})} \right\rangle = 0 \ ,$$

which is a consequence of the fact that $n+1-k \leq n$ so by the inductive hypothesis

$$\left\langle \widehat{\mathbf{X}}^{(n)}_{ut}, \langle\!\langle b_1 \ldots b_{n+1-k} \rangle\!\rangle \right\rangle = \left\langle \widehat{\mathbf{X}}^{(n)}_{ut}, \bullet_{(b_1\ldots b_{n+1-k})} \right\rangle \ .$$

This proves the result. □

Proof of Lemma 5.2.11. We have that

$$\Delta \langle\!\langle \gamma_1 \ldots \gamma_n \rangle\!\rangle = \Delta(\bullet_{\gamma_1} \cdots \bullet_{\gamma_n}) - \sum_{(a,b)} \Delta[\bullet_{a_1} \cdots \bullet_{a_k}]_{(b_1\ldots b_{n-k})}$$
$$= 1 \tilde{\otimes} \langle\!\langle \gamma_1 \ldots \gamma_n \rangle\!\rangle + \langle\!\langle \gamma_1 \ldots \gamma_n \rangle\!\rangle \tilde{\otimes} 1$$
$$+ \Delta' \Big(\bullet_{\gamma_1} \cdots \bullet_{\gamma_n} - \sum_{(a,b)} [\bullet_{a_1} \cdots \bullet_{a_k}]_{(b_1\ldots b_{n-k})} \Big) \ . \tag{5.17}$$

But by definition we have that

$$\Delta'(\bullet_{\gamma_1} \cdots \bullet_{\gamma_n}) = \sum_{(a,b)} \bullet_{a_1} \cdots \bullet_{a_k} \tilde{\otimes} \bullet_{b_1} \cdots \bullet_{b_{n-k}} \ . \tag{5.18}$$

We also have that

$$\sum_{(a,b)} \Delta'[\bullet_{a_1}\cdots\bullet_{a_k}]_{(b_1\dots b_{n-k})} = \sum_{(a,b)} \Bigg((\bullet_{a_1}\cdots\bullet_{a_k})^1 \tilde{\otimes} [(\bullet_{a_1}\cdots\bullet_{a_k})^2]_{(b_1\dots b_{n-k})} + \bullet_{a_1}\cdots\bullet_{a_k} \tilde{\otimes} \bullet_{(b_1\dots b_{n-k})} \Bigg) ,$$

where we use the notation $\Delta'(\bullet_{a_1}\cdots\bullet_{a_k}) = (\bullet_{a_1}\cdots\bullet_{a_k})^1 \tilde{\otimes} (\bullet_{a_1}\cdots\bullet_{a_k})^2$. By the coassociativity of Δ (and hence Δ') we have that

$$\begin{aligned}
&\sum_{(a,b)} (\bullet_{a_1}\cdots\bullet_{a_k})^1 \tilde{\otimes} (\bullet_{a_1}\cdots\bullet_{a_k})^2 \tilde{\otimes} \bullet_{b_1}\cdots\bullet_{b_{n-k}} \\
&\qquad = (\Delta' \tilde{\otimes} \mathrm{Id})\Delta'(\bullet_{\gamma_1}\cdots\bullet_{\gamma_n}) = (\mathrm{Id} \tilde{\otimes} \Delta')\Delta'(\bullet_{\gamma_1}\cdots\bullet_{\gamma_n}) \\
&\qquad = \sum_{(a,b)} \bullet_{a_1}\cdots\bullet_{a_k} \tilde{\otimes} (\bullet_{b_1}\cdots\bullet_{b_{n-k}})^1 \tilde{\otimes} (\bullet_{b_1}\cdots\bullet_{b_{n-k}})^2 .
\end{aligned}$$

Hence, we have

$$\begin{aligned}
&\sum_{(a,b)} \Delta'[\bullet_{a_1}\cdots\bullet_{a_k}]_{(b_1\dots b_{n-k})} \\
&= \sum_{(a,b)} \Big((\bullet_{a_1}\cdots\bullet_{a_k})^1 \tilde{\otimes} J_2\big((\bullet_{a_1}\cdots\bullet_{a_k})^2 \otimes \bullet_{b_1}\cdots\bullet_{b_{n-k}}\big) + \bullet_{a_1}\cdots\bullet_{a_k} \tilde{\otimes} \bullet_{(b_1\dots b_{n-k})} \Big) \\
&= \sum_{(a,b)} \Big(\bullet_{a_1}\cdots\bullet_{a_k} \tilde{\otimes} J_2\big((\bullet_{b_1}\cdots\bullet_{b_{n-k}})^1 \otimes (\bullet_{b_1}\cdots\bullet_{b_{n-k}})^2\big) + \bullet_{a_1}\cdots\bullet_{a_k} \tilde{\otimes} \bullet_{(b_1\dots b_{n-k})} \Big) .
\end{aligned}$$

Combining this and (5.18) with (5.17), we obtain

$$\begin{aligned}
\Delta\langle\!\langle \gamma_1 \dots \gamma_n \rangle\!\rangle &= 1 \tilde{\otimes} \langle\!\langle \gamma_1 \dots \gamma_n \rangle\!\rangle + \langle\!\langle \gamma_1 \dots \gamma_n \rangle\!\rangle \tilde{\otimes} 1 \\
&\quad + \sum_{(a,b)} \bullet_{a_1}\cdots\bullet_{a_k} \tilde{\otimes} \Big(\bullet_{b_1}\cdots\bullet_{b_{n-k}} - J_2(\Delta'(\bullet_{b_1}\cdots\bullet_{b_{n-k}})) - \bullet_{(b_1\dots b_{n-k})} \Big) ,
\end{aligned}$$

and the result follows from (5.12) □

5.2.15 The change of variables formula

We now have all the machinery to prove our generalised Itô formula. The proof is relatively simple, and relies on nothing more than a Taylor expansion and the properties of the bracket extension.

Theorem 5.2.16 (Simple version)**.** *Let* $\mathbf{X}$ *be a branched rough path above* X*. Let* $\widehat{\mathbf{X}}$ *be a* *bracket extension of* $\mathbf{X}$*. Then*

$$\delta F(X)_{st} = \int_s^t DF(X_r) \cdot dX_r + \sum_{n=2}^{N} \int_s^t \frac{D^n F(X_r)}{n!} : d\widehat{X}_r \tag{5.19}$$

where we use the notation (5.5)*.*

Before proceeding with the proof, we will explicitly define the controlled rough path integrals seen above. Let $g : \mathbb{R}^d \to \mathbb{R}^d$ be some smooth function, since $\mathbf{X}$ is an $\widehat{\mathbf{X}}$-controlled rough path, we can define $g(\mathbf{X})$ as an $\widehat{\mathbf{X}}$-controlled rough path. We have that $\langle 1, g(\mathbf{X}_t)\rangle = g(X_t)$,

$$\langle \bullet_{\beta_1} \cdots \bullet_{\beta_m}, g(\mathbf{X}_t)\rangle = \partial^{\beta_1} \ldots \partial^{\beta_m} g(X_s) \ ,$$

for all $\bullet_{\beta_1} \cdots \bullet_{\beta_m} \in \mathcal{F}_N^*$ and $\langle h, g(\mathbf{X}_t)\rangle = 0$ otherwise. By definition of controlled rough path integrals, we have that

$$\begin{aligned}
\int_s^t g(X_r) d\widehat{X}_r^{(\alpha_1 \ldots \alpha_n)} &= g(X_s)\langle \widehat{\mathbf{X}}_{st}, \bullet_{(\alpha_1 \ldots \alpha_n)}\rangle \qquad (5.20)\\
&+ \sum_{m=1}^{N-1} \sum_{\bullet_{\beta_1} \cdots \bullet_{\beta_m} \in \mathcal{F}_N} \langle \bullet_{\beta_1} \cdots \bullet_{\beta_m}, g(\mathbf{X}_s)\rangle \left\langle \widehat{\mathbf{X}}_{st}, [\bullet_{\beta_1} \cdots \bullet_{\beta_m}]_{(\alpha_1 \ldots \alpha_n)} \right\rangle + o(|t-s|)\\
&= \sum_{m=0}^{N} \sum_{\beta_1, \ldots, \beta_m = 1}^{d} \frac{\partial^{\beta_1} \ldots \partial^{\beta_m} g(X_s)}{m!} \left\langle \widehat{\mathbf{X}}_{st}, [\bullet_{\beta_1} \cdots \bullet_{\beta_m}]_{(\alpha_1 \ldots \alpha_n)} \right\rangle + o(|t-s|) \ ,
\end{aligned}$$

where in the last line we have used the symmetry of the expression to replace $\sum_{\bullet_{\beta_1} \cdots \bullet_{\beta_m} \in \mathcal{F}_N}$, the unordered sum, with $\sum_{\beta_1, \ldots, \beta_m = 1}^{d} 1/m!$.

Proof. For convenience, in the sequel we will perform our arithmetic modulo $o(|t-s|)$. On the one hand, a Taylor expansion tells us that

$$\delta F(X)_{st} = \sum_{k=1}^{N} \frac{\partial^{\gamma_1} \ldots \partial^{\gamma_k} F(X_s)}{k!} \langle \mathbf{X}_{st}, \bullet_{\gamma_1} \cdots \bullet_{\gamma_k}\rangle \ ,$$

where we sum over all $\gamma_i = 1 \ldots d$. On the other hand, from (5.20) we have that

$$\begin{aligned}
\int_s^t \frac{\partial^{\alpha_1} \ldots \partial^{\alpha_n}}{n!} F(X_r) d\widehat{X}_r^{(\alpha_1 \ldots \alpha_n)} &\overset{\text{def}}{=} \left\langle 1, \int_s^t \frac{\partial^{\alpha_1} \ldots \partial^{\alpha_n}}{n!} F(\mathbf{X}_r) d\widehat{X}_r^{(\alpha_1 \ldots \alpha_n)} \right\rangle\\
&= \sum_{m=0}^{N-m} \frac{\partial^{\beta_1} \ldots \partial^{\beta_m}}{m!} \left(\frac{\partial^{\alpha_1} \ldots \partial^{\alpha_n} F(X_s)}{n!}\right) \left\langle \widehat{\mathbf{X}}_{st}, [\bullet_{\beta_1} \cdots \bullet_{\beta_m}]_{(\alpha_1 \ldots \alpha_n)} \right\rangle \ ,
\end{aligned}$$

where we sum over all $\beta_i = 1 \dots d$, with the convention $[x]_{(\alpha_1)} = [x]_{\alpha_1}$ for the original single index labels. Hence, by expanding the integrals on the right hand side of (5.19), we obtain

$$\int_s^t \partial^{\alpha_i} F(X_r) dX_r^{\alpha_i} + \sum_{n=2}^{N} \int_s^t \frac{\partial^{\alpha_1} \dots \partial^{\alpha_n}}{n!} F(X_r) d\widehat{X}_r^{(\alpha_1 \dots \alpha_n)}$$
$$= \sum_{n=1}^{N} \sum_{m=0}^{N-m} \frac{\partial^{\beta_1} \dots \partial^{\beta_m}}{m!} \left(\frac{\partial^{\alpha_1} \dots \partial^{\alpha_n} F(X_s)}{n!} \right) \left\langle \widehat{\mathbf{X}}_{st}, [\bullet_{\beta_1} \cdots \bullet_{\beta_m}]_{(\alpha_1 \dots \alpha_n)} \right\rangle ,$$

where we sum over all $\alpha_i, \beta_i = 1 \dots d$. By substituting $k = n + m$ and $(\gamma_1, \dots, \gamma_k) = (\beta_1, \dots, \beta_m, \alpha_1, \dots, \alpha_n)$, the above comes to

$$\sum_{k=1}^{N} \sum_{\gamma_1, \dots, \gamma_k = 1}^{d} \frac{\partial^{\gamma_1} \dots \partial^{\gamma_k} F(X_s)}{k!} \left(\sum_{n=1}^{k} \binom{k}{n} \left\langle \widehat{\mathbf{X}}_{st}, [\bullet_{\gamma_1} \cdots \bullet_{\gamma_{k-n}}]_{(\gamma_{k-n+1} \cdots \gamma_k)} \right\rangle \right)$$
$$= \sum_{k=1}^{N} \sum_{\{\gamma_1, \dots, \gamma_k\}} \frac{\partial^{\gamma_1} \dots \partial^{\gamma_k} F(X_s)}{k!} \left(\sum_{n=1}^{k} \binom{k}{n} \left\langle \widehat{\mathbf{X}}_{st}, \mathrm{Sym}[\bullet_{\gamma_1} \cdots \bullet_{\gamma_{k-n}}]_{(\gamma_{k-n+1} \cdots \gamma_k)} \right\rangle \right) ,$$

where Sym is the linear operator satisfying

$$\mathrm{Sym} \left\langle \widehat{\mathbf{X}}_{st}, [\bullet_{\gamma_1} \cdots \bullet_{\gamma_{k-n}}]_{(\gamma_{k-n+1} \cdots \gamma_k)} \right\rangle \stackrel{\text{def}}{=} \sum_{\sigma \in \mathrm{Sym}(k)} \left\langle \widehat{\mathbf{X}}_{st}, [\bullet_{\gamma_{\sigma(1)}} \cdots \bullet_{\gamma_{\sigma(k-n)}}]_{(\gamma_{\sigma(k-n+1)} \cdots \gamma_{\sigma(k)})} \right\rangle ,$$

and where $\sum_{\{\gamma_1, \dots, \gamma_k\}}$ denotes the sum over all $(\gamma_1, \dots, \gamma_k)$ that are not rearrangements of one another. However, it is not hard to convince oneself that

$$\sum_{n=1}^{k} \binom{k}{n} \mathrm{Sym} \left\langle \widehat{\mathbf{X}}_{st}, [\bullet_{\gamma_1} \cdots \bullet_{\gamma_{k-n}}]_{(\gamma_{k-n+1} \cdots \gamma_k)} \right\rangle$$
$$= \mathrm{Sym} \left\langle \widehat{\mathbf{X}}_{st}, \sum_{(a,b)} \left([\bullet_{a_1} \cdots \bullet_{a_{k-n}}]_{(b_1 \dots b_n)} \right) + \bullet_{(\gamma_1 \dots \gamma_k)} \right\rangle ,$$

where we sum over choices $(a_1, \dots, a_{k-n})$, $(b_1, \dots, b_n)$ from $(\gamma_1, \dots, \gamma_k)$ for $1 \leq n \leq k$ as in (5.11). By definition of bracket extension $\widehat{\mathbf{X}}$, we have that

$$\sum_{\{\gamma_1, \dots, \gamma_k\}} \mathrm{Sym} \left\langle \widehat{\mathbf{X}}_{st}, \sum_{(a,b)} \left([\bullet_{a_1} \cdots \bullet_{a_{k-n}}]_{(b_1 \dots b_n)} \right) + \bullet_{(\gamma_1 \dots \gamma_k)} \right\rangle$$
$$= \sum_{\{\gamma_1, \dots, \gamma_k\}} \mathrm{Sym} \langle \mathbf{X}_{st}, \bullet_{\gamma_1} \cdots \bullet_{\gamma_k} \rangle = \sum_{\gamma_1, \dots, \gamma_k = 1}^{d} \langle \mathbf{X}_{st}, \bullet_{\gamma_1} \cdots \bullet_{\gamma_k} \rangle .$$

It follows that

$$\delta F(X)_{st} - \int_s^t DF(X_r) \cdot dX_r - \sum_{n=2}^{N} \int_s^t \frac{\partial^{\alpha_1} \dots \partial^{\alpha_n}}{n!} F(X_r) d\widehat{X}_r^{(\alpha_1 \dots \alpha_n)} = o(|t-s|) \ ,$$

where we sum over $i = 1 \dots d$ and $\alpha_j = 1 \dots d$. But since the left hand side in an increment, it must be identically zero. □

5.3 The general case

5.3.1 Outline of the rough path approach

In the general case, we aim to develop a change of variables formula for $F(Y_t)$ where $\mathbf{Y} : [0,T] \to (\mathcal{H}_N)^e$ with $\langle 1, \mathbf{Y} \rangle = Y$ is the controlled rough path solution to the equation

$$dY_t = f(Y_t) \cdot dX_t \ , \tag{5.21}$$

driven by a branched rough path $\mathbf{X}$, where the vector fields $f = (f_1, \dots, f_d)$ and each $f_i : \mathbb{R}^e \to \mathbb{R}^e$ is smooth. Of course, by choosing $f_i^j(Y) = \delta_{i,j}$ we obtain the simple case. For a smooth $F : \mathbb{R}^e \to \mathbb{R}^e$, we can define the leading integral in the change of variables formula

$$\int_s^t DF : \big(f(Y_r) \cdot dX_r\big) \stackrel{\text{def}}{=} \left\langle 1, \int_s^t DF : \big(f(\mathbf{Y}_r) \cdot dX_r\big) \right\rangle \ ,$$

where we use the shorthand

$$DF : (f(Y) \cdot dX) = \sum_{i=1}^{d} \big(DF : f_i(Y)\big) dX^i \ .$$

As in the simple case, the formula we obtain contains Itô correction terms that involve integrals against a path $\widehat{X}$. But in the general case $\widehat{X}$ contains many more components than in the simple case. In addition to the brackets between components of X, we also have brackets between components of $\mathbf{X}$. In the simple case, the bracket $\widehat{X}^{(a_1 \dots a_n)}$ arose from the product $\langle \mathbf{X}, \bullet_{a_1} \cdots \bullet_{a_n} \rangle$, but in the general case we obtain a bracket $\widehat{X}^{(\tau_1 \dots \tau_n)}$ arising from the product $\langle \mathbf{X}, \tau_1 \dots \tau_n \rangle$ for any $\tau_i \in \mathcal{T}$ with $|\tau_1| + \cdots + |\tau_n| \leq N$ and $n \geq 2$. For instance, the expression

$$\delta X_{st}^i \int_s^t \delta X_{sr}^k dX_r^j - \int_s^t \delta X_{sr}^i \delta X_{sr}^k dX_r^j - \int_s^t \int_s^{r_1} \int_s^{r_2} dX_{r_3}^j dX_{r_2}^j dX_{r_1}^i - \int_s^t \int_s^{r_1} dX_{r_2}^k d\widehat{X}_{r_1}^{(ij)}$$

defines the increment of a path $\widehat{X}^{(\bullet_i \overset{\bullet k}{\bullet_j})}$, this bracket arises from the product $\langle \mathbf{X}_{st}, \bullet_i \overset{\bullet k}{\bullet_j} \rangle$. This bracket shows up in the change of variables formula as soon as $\gamma \leq 1/3$. For this reason, we will henceforth denote the brackets $\widehat{X}^{(a_1 \dots a_n)}$ as $\widehat{X}^{(\bullet_{a_1} \cdots \bullet_{a_n})}$, to be consistent with the notation for the general brackets $\widehat{X}^{(\tau_1 \dots \tau_n)}$. This allows us to write

$$\widehat{X} = X \oplus (\widehat{X}^{(\tau_1 \dots \tau_n)} : \tau_i \in \mathcal{T} \text{ with } |\tau_1| + \cdots + |\tau_n| \leq N \text{ and } n \geq 2) \ .$$

The bracket extension $\widehat{\mathbf{X}}$ will therefore take values on the Connes-Kreimer Hopf algebra generated by trees with labels of the form $i, (\tau_1\tau_2), (\tau_1\tau_2\tau_3)$ and so forth. This will be detailed in Subsection 5.3.6

The construction of a bracket extension $\widehat{\mathbf{X}}$ above $\widehat{X}$ follows a similar iterative procedure as found in the simple case.

1. Suppose we have the bracket paths

$$\widehat{X}^{(n)} = X \oplus \left(\widehat{X}^{(\tau_1 \dots \tau_k)} : \tau_i \in \mathcal{T} \text{ with } |\tau_1| + \cdots + |\tau_k| \leq n \text{ and } k \geq 2\right) \ .$$

   and a branched rough path $\widehat{\mathbf{X}}^{(n)}$ above $\widehat{X}^{(n)}$ that is an extension of $\mathbf{X}$.

2. We can then uniquely define the next bracket paths $\widehat{X}^{(\tau_1 \dots \tau_k)}$ for any $|\tau_1| + \cdots + |\tau_k| = n$, using the components of $\widehat{\mathbf{X}}^{(n)}$.

3. Add these new brackets paths to the underlying path

$$\widehat{X}^{(n+1)} = \widehat{X}^{(n)} \oplus \left(\widehat{X}^{(\tau_1 \dots \tau_k)} : \tau_i \in \mathcal{T} \text{ with } |\tau_1| + \cdots + |\tau_k| = n+1 \text{ and } k \geq 2\right) \ .$$

4. Find a branched rough path $\widehat{\mathbf{X}}^{(n+1)}$ above $\widehat{X}^{(n+1)}$ that is an extension of $\widehat{\mathbf{X}}^{(n)}$.

5. Repeat until we obtain $\widehat{\mathbf{X}} = \widehat{\mathbf{X}}^{(N)}$.

If Y is the solution to (5.21) ie.

$$\delta Y_{st} = \sum_{\tau \in \mathcal{T}_N} f_\tau(Y_s) \langle \mathbf{X}_{st}, \tau \rangle + r_{st} \ ,$$

where $|r_{st}| = o(|t-s|)$ then we obtain the following change of variables formula

Theorem 5.3.2 (General version)**.** *For a given branched rough path $\mathbf{X}$ and bracket extension $\widehat{\mathbf{X}}$ of $\mathbf{X}$, we have that*

$$\delta F(Y)_{st} = \int_s^t DF : (f(Y_r) \cdot dX_r) + \sum_{n=2}^{N} \int_s^t \frac{D^n F}{n!} : (f_{\tau_1}, \dots, f_{\tau_n})(Y_r) d\widehat{X}^{(\tau_1 \dots \tau_n)} \ ,$$

where we sum over all $\tau_i \in \mathcal{T}_N$ with $|\tau_1| + \cdots + |\tau_n| \le N$ and where $\widehat{X}^{(\tau_1 \ldots \tau_n)}$ are the bracket paths given by $\delta\widehat{X}^{(\tau_1\ldots\tau_n)}_{st} = \langle \widehat{\mathbf{X}}_{st}, \bullet_{(\tau_1\ldots\tau_n)}\rangle$ (to be defined properly in Subsection 5.3.6).

Remark 5.3.3. To be clear, all integrals listed are to be interpreted in a controlled rough path sense. In particular, $\mathbf{Y}$ is an $\mathbf{X}$-controlled rough path so it must also be an $\widehat{\mathbf{X}}$-controlled rough path, hence we can always define

$$\int_s^t \frac{D^nF}{n!} : (f_{\tau_1}, \ldots, f_{\tau_n})(Y_r) d\widehat{X}^{(\tau_1\ldots\tau_n)} \stackrel{\text{def}}{=} \left\langle 1, \int_s^t \frac{D^nF}{n!} : (f_{\tau_1}, \ldots, f_{\tau_n})(\mathbf{Y}_r) d\widehat{X}^{(\tau_1\ldots\tau_n)} \right\rangle \; ,$$

using the theory from Chapter 3.

As in the simple case, we will illustrate the construction of the formula by looking at the two simplest non-trivial cases.

Example 5.3.4. We first look at the Brownian motion case, or indeed any path X in $\mathbb{R}^d$ with Hölder exponent $1/3 < \gamma \le 1/2$ and with a branched rough path $\mathbf{X}$ above X. Let $\mathbf{Y}$ with $\langle 1, \mathbf{Y}\rangle = Y$ be the solution to

$$dY_t = f(Y_t) \cdot dX_t \; , \tag{5.22}$$

driven by $\mathbf{X}$. Then, from Proposition 3.3.10 we have that

$$\begin{aligned} \delta Y_{st} &= f_{\bullet_i}(Y_s)\langle \mathbf{X}_{st}, \bullet_i\rangle + f_{\overset{\bullet i}{\bullet j}}\langle \mathbf{X}_{st}, \overset{\bullet i}{\bullet j}\rangle + r_{st} \\ &= f_i(Y_s)\langle \mathbf{X}_{st}, \bullet_i\rangle + Df_i(Y_s) : f_j(Y_s)\langle \mathbf{X}_{st}, \overset{\bullet j}{\bullet i}\rangle + r_{st} \; , \end{aligned} \tag{5.23}$$

where we sum over $i, j, k = 1 \ldots d$, where $r_{st} = o(|t-s|)$. Let $F : \mathbb{R}^e \to \mathbb{R}^e$ be some sufficiently smooth function. As usual, we start the Itô formula with a Taylor expansion

$$\delta F(Y)_{st} = DF(Y_s) : \delta Y_{st} + \frac{1}{2} D^2F(Y_s) : (\delta Y_{st}, \delta Y_{st}) + R^F_{st} \; ,$$

where $|R^F_{st}| \le C|\delta Y_{st}|^3 = o(|t-s|)$. Using (5.23), this simplifies to

$$DF : f_i(Y_s)\langle \mathbf{X}_{st}, \bullet_i\rangle + DF : \big(Df_i : f_j\big)(Y_s)\langle \mathbf{X}_{st}, \overset{\bullet j}{\bullet i}\rangle + \frac{1}{2}D^2F : \big(f_i, f_j\big)(Y_s)\langle \mathbf{X}_{st}, \bullet_i \bullet_j\rangle + o(|t-s|) \; .$$

The second step in building an Itô formula is to construct the leading integral term. It is clear that $DF : f_i(Y_s)\langle \mathbf{X}_{st}, \bullet_i\rangle = DF : f_i(Y_s)\delta X^i_{st}$ is the lowest order term of this leading integral. From the controlled rough path definition of an integral we have that

$$\int_s^t DF : f_i(Y_r)dX_r^i \overset{\text{def}}{=} \left\langle 1, \int_s^t DF : f_i(\mathbf{Y}_r)dX_r^i \right\rangle \tag{5.24}$$

$$= DF : f_i(Y_s)\langle \mathbf{X}_{st}, \bullet_i \rangle + D(DF : f_i(Y_r)) : f_j(Y_r)\langle \mathbf{X}_{st}, \genfrac{}{}{0pt}{}{\bullet_j}{\bullet_i} \rangle + r_{st} \;, \tag{5.25}$$

where i is fixed and we sum over $j = 1 \ldots d$ and where $|r_{st}| = o(|t-s|)$. If we apply Leibniz formula, we see that

$$D\big(DF : f_i(Y_s)\big) : f_j(Y_s) = DF(Y_s) : \big(Df_i(Y_s) : f_j(Y_s)\big) + D^2F : \big(f_i(Y_s)), f_j(Y_s)\big) \;.$$

Hence, the DF terms in δF and the integral (5.25) will cancel out, leaving

$$\begin{aligned} \delta F(Y)_{st} - \int_s^t DF : (f(Y_r) \cdot dX_r) \\ &= \sum_{i,j=1}^d \bigg(\frac{1}{2} D^2F : (f_i, f_j)(Y_s)\langle \mathbf{X}_{st}, \bullet_i \bullet_j \rangle \\ &\quad - D^2F : (f_i), f_j)(Y_s)\langle \mathbf{X}_{st}, \genfrac{}{}{0pt}{}{\bullet_j}{\bullet_i} \rangle \bigg) + o(|t-s|) \\ &= \sum_{i,j=1}^d \frac{1}{2} D^2F : (f_i, f_j)(Y_s)\langle \mathbf{X}_{st}, \bullet_i \bullet_j - \genfrac{}{}{0pt}{}{\bullet_j}{\bullet_i} - \genfrac{}{}{0pt}{}{\bullet_i}{\bullet_j} \rangle + o(|t-s|) \;. \end{aligned}$$

In the last line above, we have replaced $\genfrac{}{}{0pt}{}{\bullet_i}{\bullet_j}$ with its symmetric part, this is permitted by the fact that we sum over all $i, j = 1 \ldots d$. As in the simple case, we set

$$\langle\!\langle \bullet_i \bullet_j \rangle\!\rangle \overset{\text{def}}{=} \bullet_i \bullet_j - \genfrac{}{}{0pt}{}{\bullet_j}{\bullet_i} - \genfrac{}{}{0pt}{}{\bullet_i}{\bullet_j} \;.$$

We likewise define the generalised quadratic variation (or bracket) path $\widehat{X}^{(\bullet_i \bullet_j)}$ by

$$\delta \widehat{X}_{st}^{(\bullet_i \bullet_j)} \overset{\text{def}}{=} \langle \mathbf{X}_{st}, \langle\!\langle \bullet_i \bullet_j \rangle\!\rangle \rangle \;,$$

and it is clear from the definition that $|\delta \widehat{X}_{st}^{(\bullet_i \bullet_j)}| \le C|t-s|^{2\gamma}$. The rest of the construction works precisely the same as in the simple case. We set

$$\widehat{X} = X \oplus (\widehat{X}^{(\bullet_i \bullet_j)})_{i,j=1\ldots d} \;.$$

We then define the bracket extension $\widehat{\mathbf{X}}$ on the Connes-Kreimer Hopf algebra generated by the extended alphabet

$$\{1, \ldots, d\} \cup \{(\bullet_i \bullet_j) : i, j = 1, \ldots, d\} \;.$$

As in the simple case, $\widehat{\mathbf{X}}$ is canonically defined. To complete the formula, we have that

$$\int_s^t \frac{1}{2} D^2 F : (f_i, f_j)(Y_r) d\widehat{X}_r^{(\bullet_i \bullet_j)}$$
$$\stackrel{\text{def}}{=} \frac{1}{2} D^2 F : (f_i, f_j)(Y_s) \langle \mathbf{X}_{st}, \langle\!\langle \bullet_i \, \bullet_j \rangle\!\rangle \rangle + o(|t-s|)$$

for fixed i, j, where we sum over $k = 1 \dots d$. Hence, we obtain

$$\delta F(Y)_{st} - \int_s^t DF : (f(Y_r) \cdot dX_r) - \sum_{i,j=1}^d \int_s^t \frac{1}{2} D^2 F : \big(f_i, f_j\big)(Y_r) d\widehat{X}_r^{(\bullet_i \bullet_j)}$$
$$= o(|t-s|) \ .$$

But since the left hand side is an increment and the right hand side is $o(|t-s|)$, the left hand side must be identically zero. We therefore recover the change-of-variables formula

$$\delta F(Y)_{st} = \int_s^t DF : (f(Y_r) \cdot dX_r) + \sum_{i,j=1}^d \int_s^t \frac{1}{2} D^2 F : \big(f_i, f_j\big)(Y_r) d\widehat{X}_r^{(\bullet_i \bullet_j)} \ , \quad (5.26)$$

which coincides, at least structurally, with the usual Itô formula. In the case where X is a BM and $\mathbf{X}$ is the Itô rough path above X, we would of course recover the usual Itô formula.

In the above example, the construction is only slightly more complicated than in the simple case. However, as soon as $\gamma \leq 1/3$, we encounter the brackets between trees, including $\langle\!\langle \bullet_i \, {}^{\bullet_k}_{\bullet_j} \rangle\!\rangle$. As we shall see, these brackets arise just as naturally as the brackets in the simple case, that is, by simply matching coefficients.

Example 5.3.5. We will now consider the case $1/4 < \gamma \leq 1/3$, so that $N = 3$. Hence, the terms that are $O(|t-s|^{3\gamma})$ can no longer be ignored. As above, suppose Y solves (5.22) so that

$$\delta Y_{st} = f_i(Y_s) \langle \mathbf{X}_{st}, \bullet_i \rangle + f_{{}^{\bullet_j}_{\bullet_i}}(Y_s) \langle \mathbf{X}_{st}, {}^{\bullet_j}_{\bullet_i} \rangle$$
$$+ f_{{}^{\bullet_k}_{\bullet_j \bullet_i}}(Y_s) \langle \mathbf{X}_{st}, {}^{\bullet_k}_{\bullet_j \bullet_i} \rangle + f_{{}^{\bullet_j \bullet_k}_{\bullet_i}}(Y_s) \langle \mathbf{X}_{st}, {}^{\bullet_j \bullet_k}_{\bullet_i} \rangle + r_{st} \ ,$$

where we sum over $i, j, k = 1 \dots d$ (that is, if the term has i, j in it, we sum over $i, j = 1 \dots d$ and so forth) and where we have

$$f_{{}^{\bullet_j}_{\bullet_i}}(Y_s) = Df_i(Y_s) : f_j(Y_s) \ , \quad f_{{}^{\bullet_k}_{\bullet_j \bullet_i}}(Y_s) = Df_i(Y_s) : \big(Df_j(Y_s) : f_k(Y_s)\big) \ ,$$
$$f_{{}^{\bullet_j \bullet_k}_{\bullet_i}}(Y_s) = \frac{1}{2} D^2 f_i : (f_j, f_k)(Y_s) \ .$$

Using this expression, the Taylor expansion takes the form

$$\begin{aligned}
\delta F(Y)_{st} &= DF : f_i(Y_s)\langle \mathbf{X}_{st}, \bullet_i\rangle + DF : f_{[\bullet_j]_i}(Y_s)\langle \mathbf{X}_{st}, [\bullet_j]_i\rangle \\
&+ DF : f_{[[\bullet_k]_j]_i}(Y_s)\langle \mathbf{X}_{st}, [[\bullet_k]_j]_i\rangle + DF : f_{[\bullet_j\bullet_k]_i}(Y_s)\langle \mathbf{X}_{st}, [\bullet_j\bullet_k]_i\rangle \\
&+ \frac{1}{2}D^2F : (f_i(Y_s), f_j(Y_s))\langle \mathbf{X}_{st}, \bullet_i\bullet_j\rangle \\
&+ D^2F : (f_i, f_{[\bullet_k]_j})(Y_s)\langle \mathbf{X}_{st}, \bullet_i[\bullet_k]_j\rangle \\
&+ \frac{1}{6}D^3F : (f_i, f_j, f_k)(Y_s)\langle \mathbf{X}_{st}, \bullet_i\bullet_j\bullet_k\rangle + o(|t-s|)\ ,
\end{aligned}$$

where we sum over $i, j, k = 1 \dots d$. The leading order integral is defined by

$$\begin{aligned}
\int_s^t DF : f_i(Y_r)dX_r^i &\stackrel{\text{def}}{=} \left\langle 1, \int_s^t DF : f_i(\mathbf{Y}_r)dX_r^i \right\rangle \\
&= DF : f_i(Y_s)\langle \mathbf{X}_{st}, \bullet_i\rangle + D(DF : f_i(Y_s)) : f_j(Y_s)\langle \mathbf{X}_{st}, [\bullet_j]_i\rangle \\
&+ D(DF : f_i(Y_s)) : f_{[\bullet_k]_j}(Y_s)\langle \mathbf{X}_{st}, [[\bullet_k]_j]_i\rangle \\
&+ \frac{1}{2}D^2(DF : f_i(Y_s)) : (f_j, f_k)(Y_s)\langle \mathbf{X}_{st}, [\bullet_j\bullet_k]_i\rangle + o(|t-s|)\ ,
\end{aligned}$$

where i is fixed and we sum over $j, k = 1 \dots d$. If we apply Leibniz rule to each of the terms in the expression for the above integral, we see that the DF terms appearing are precisely those found in the Taylor expansion. Hence, when we isolate the leading integral, these terms will cancel out, leaving

$$\begin{aligned}
\delta F(Y)_{st} - &\int_s^t DF : (f(Y_r)\cdot dX_r) \\
&= \sum_{i,j=1}^{d} \left(\frac{1}{2}D^2F : (f_i, f_j)(Y_s)\langle \mathbf{X}_{st}, \langle\!\langle \bullet_i\,\bullet_j\rangle\!\rangle\rangle\right) \\
&+ \sum_{i,j,k=1}^{d} \Bigg(D^2F : (f_i, f_{[\bullet_k]_j}))(Y_s)\langle \mathbf{X}_{st}, \bullet_i[\bullet_k]_j - [[\bullet_k]_j]_i - [\bullet_i\bullet_k]_j\rangle \\
&+ \frac{1}{6}D^3F : (f_i, f_j, f_k)(Y_s)\langle \mathbf{X}_{st}, \bullet_i\,\bullet_j\,\bullet_k - [\bullet_i\bullet_k]_j - [\bullet_i\bullet_j]_k - [\bullet_j\bullet_k]_i\rangle \Bigg) \\
&+ o(|t-s|)\ .
\end{aligned}$$

As in the simple case, we set

$$\widehat{X}^{(2)} = X \oplus (\widehat{X}^{(\bullet_i\bullet_j)})_{i,j=1\dots d}\ ,$$

We then define $\widehat{\mathbf{X}}^{(2)}$ on the Hopf algebra $\widehat{\mathcal{H}}^{(2)}$ also as in the simple case. Using $\widehat{\mathbf{X}}^{(2)}$, we can define

$$
\begin{aligned}
&\int_s^t \frac{1}{2} D^2F : (f_i, f_j)(Y_r) d\widehat{X}_r^{(\bullet_i\bullet_j)} \\
&\qquad \stackrel{\text{def}}{=} \left\langle 1, \int_s^t \frac{1}{2} D^2F : (f_i, f_j)(\mathbf{Y}_r) d\widehat{X}_r^{(\bullet_i\bullet_j)} \right\rangle \\
&\qquad = \frac{1}{2} D^2F : (f_i, f_j)(Y_s) \left\langle \widehat{\mathbf{X}}^{(2)}_{st}, \bullet_{(\bullet_i\bullet_j)} \right\rangle \\
&\qquad + \frac{1}{2} (D(D^2F : (f_i, f_j)) : f_k)(Y_s) \left\langle \widehat{\mathbf{X}}^{(2)}_{st}, \genfrac{}{}{0pt}{}{\bullet^k}{\bullet_{(\bullet_i\bullet_j)}} \right\rangle + o(|t-s|) \\
&\qquad = \frac{1}{2} D^2F : (f_i, f_j)(Y_s) \left\langle \mathbf{X}_{st}, \langle\!\langle \bullet_i \bullet_j \rangle\!\rangle \right\rangle \\
&\qquad + \frac{1}{2} (D(D^2F : (f_i, f_j)) : f_k)(Y_s) \left\langle \widehat{\mathbf{X}}^{(2)}_{st}, \genfrac{}{}{0pt}{}{\bullet^k}{\bullet_{(\bullet_i\bullet_j)}} \right\rangle + o(|t-s|) \,,
\end{aligned}
$$

where i, j are fixed and we sum over $k = 1 \ldots d$. If we apply Leibiz formula to this expression, a simple calculation yields

$$
\begin{aligned}
&\delta F(Y)_{st} - \int_s^t DF : (f(Y_r) \cdot dX_r) - \sum_{i,j=1}^d \int_s^t \frac{1}{2} D^2F : (f_i, f_j)(Y_r) \cdot d\widehat{X}_r^{(\bullet_i\bullet_j)} \\
&\quad = \sum_{i,j,k=1}^d D^2F(Y_s) : (f_i, f_{\genfrac{}{}{0pt}{}{\bullet^k}{\bullet_j}})(Y_s) \Bigg(\left\langle \mathbf{X}_{st}, \bullet_i \genfrac{}{}{0pt}{}{\bullet^k}{\bullet_j} - \genfrac{}{}{0pt}{}{\bullet^k}{\genfrac{}{}{0pt}{}{\bullet^j}{\bullet_i}} - \overset{\bullet_i\,\bullet_k}{\vee_j} \right\rangle - \left\langle \widehat{\mathbf{X}}^{(2)}_{st}, \genfrac{}{}{0pt}{}{\bullet^k}{\bullet_{(\bullet_i\bullet_j)}} \right\rangle \Bigg) \\
&\quad + \frac{1}{6} D^3F(Y_s) : (f_i, f_j, f_k)(Y_s) \\
&\quad \times \Bigg(\left\langle \mathbf{X}_{st}, \bullet_i \bullet_j \bullet_k - \overset{\bullet_i\,\bullet_k}{\vee_j} - \overset{\bullet_i\,\bullet_j}{\vee_k} - \overset{\bullet_j\,\bullet_k}{\vee_i} \right\rangle - \left\langle \widehat{\mathbf{X}}^{(2)}_{st}, \genfrac{}{}{0pt}{}{\bullet^k}{\bullet_{(\bullet_i\bullet_j)}} - \genfrac{}{}{0pt}{}{\bullet^i}{\bullet_{(\bullet_j,\bullet_k)}} - \genfrac{}{}{0pt}{}{\bullet^j}{\bullet_{(\bullet_i\bullet_k)}} \right\rangle \Bigg) \\
&\quad + o(|t-s|) \,.
\end{aligned}
$$

And using the fact that $\mathbf{X}$ and $\widehat{\mathbf{X}}^{(2)}$ agree on $h \in \mathcal{H}$, this equals

$$
\begin{aligned}
&\sum_{i,j,k=1}^d D^2F(Y_s) : (f_i, f_{\genfrac{}{}{0pt}{}{\bullet^k}{\bullet_j}})(Y_s) \left\langle \widehat{\mathbf{X}}^{(2)}_{st}, \bullet_i \genfrac{}{}{0pt}{}{\bullet^k}{\bullet_j} - \genfrac{}{}{0pt}{}{\bullet^k}{\genfrac{}{}{0pt}{}{\bullet^j}{\bullet_i}} - \overset{\bullet_i\,\bullet_k}{\vee_j} - \genfrac{}{}{0pt}{}{\bullet^k}{\bullet_{(\bullet_i\bullet_j)}} \right\rangle \\
&\qquad\qquad + \frac{1}{6} D^3F(Y_s) : (f_i, f_j, f_k)(Y_s) \\
&\qquad\qquad \times \left\langle \widehat{\mathbf{X}}^{(2)}_{st}, \bullet_i \bullet_j \bullet_k - \overset{\bullet_i\,\bullet_k}{\vee_j} - \overset{\bullet_i\,\bullet_j}{\vee_k} - \overset{\bullet_j\,\bullet_k}{\vee_i} - \genfrac{}{}{0pt}{}{\bullet^k}{\bullet_{(\bullet_i\bullet_j)}} - \genfrac{}{}{0pt}{}{\bullet^i}{\bullet_{(\bullet_j\bullet_k)}} - \genfrac{}{}{0pt}{}{\bullet^j}{\bullet_{(\bullet_i\bullet_k)}} \right\rangle \\
&\qquad\qquad + o(|t-s|) \,.
\end{aligned}
$$

The remaining terms become integrals driven by new bracket processes, so we set

$$
\langle\!\langle \bullet_i \genfrac{}{}{0pt}{}{\bullet^k}{\bullet_j} \rangle\!\rangle \stackrel{\text{def}}{=} \bullet_i \genfrac{}{}{0pt}{}{\bullet^k}{\bullet_j} - \genfrac{}{}{0pt}{}{\bullet^k}{\genfrac{}{}{0pt}{}{\bullet^j}{\bullet_i}} - \overset{\bullet_i\,\bullet_k}{\vee_j} - \genfrac{}{}{0pt}{}{\bullet^k}{\bullet_{(\bullet_i\bullet_j)}} \,,
$$

and

$$\langle\!\langle \bullet_i \bullet_j \bullet_k \rangle\!\rangle \stackrel{\text{def}}{=} \bullet_i \bullet_j \bullet_k - \overset{\bullet_i \bullet_k}{\bullet_j} - \overset{\bullet_i \bullet_j}{\bullet_k} - \overset{\bullet_j \bullet_k}{\bullet_i} - \overset{\bullet_k}{\bullet}_{(\bullet_i \bullet_j)} - \overset{\bullet_i}{\bullet}_{(\bullet_j \bullet_k)} - \overset{\bullet_j}{\bullet}_{(\bullet_i \bullet_k)} .$$

As with the previous brackets, one can check that each of these objects satisfy

$$\langle \widehat{\mathbf{X}}^{(2)}_{st}, \langle\!\langle \bullet_i \overset{\bullet_k}{\bullet_j} \rangle\!\rangle \rangle = \langle \widehat{\mathbf{X}}^{(2)}_{su}, \langle\!\langle \bullet_i \overset{\bullet_k}{\bullet_j} \rangle\!\rangle \rangle + \langle \widehat{\mathbf{X}}^{(2)}_{ut}, \langle\!\langle \bullet_i \overset{\bullet_k}{\bullet_j} \rangle\!\rangle \rangle$$

and

$$\langle \widehat{\mathbf{X}}^{(2)}_{st}, \langle\!\langle \bullet_i \bullet_j \bullet_k \rangle\!\rangle \rangle = \langle \widehat{\mathbf{X}}^{(2)}_{su}, \langle\!\langle \bullet_i \bullet_j \bullet_k \rangle\!\rangle \rangle + \langle \widehat{\mathbf{X}}^{(2)}_{ut}, \langle\!\langle \bullet_i \bullet_j \bullet_k \rangle\!\rangle \rangle$$

for all $s \leq u \leq t$. We can therefore define 3γ-Hölder paths by

$$\delta \widehat{X}^{(\bullet_i \overset{\bullet_k}{\bullet_j})}_{st} \stackrel{\text{def}}{=} \langle \widehat{\mathbf{X}}^{(2)}_{st}, \langle\!\langle \bullet_i \overset{\bullet_k}{\bullet_j} \rangle\!\rangle \rangle \text{ and } \delta \widehat{X}^{(\bullet_i \bullet_j \bullet_k)}_{st} \stackrel{\text{def}}{=} \langle \widehat{\mathbf{X}}^{(2)}_{st}, \langle\!\langle \bullet_i \bullet_j \bullet_k \rangle\!\rangle \rangle .$$

We then define

$$\widehat{X}^{(3)} = \widehat{X}^{(2)} \oplus (\widehat{X}^{(\bullet_i \overset{\bullet_k}{\bullet_j})}, \widehat{X}^{(\bullet_i \bullet_j \bullet_k)})_{i,j,k=1\ldots d}$$

and let $\widehat{\mathbf{X}}^{(3)}$ be a bracket extension of $\widehat{\mathbf{X}}^{(2)}$, lying above the path $\widehat{X}^{(3)}$. That is, $\widehat{\mathbf{X}}^{(3)}$ is a branched rough path on the Connes-Kreimer Hopf algebra $\widehat{\mathcal{H}}^{(3)}$ generated by the alphabet

$$\{1, \ldots, d\} \cup \{(\bullet_i \bullet_j) : i,j = 1, \ldots, d\} \cup \{(\bullet_i \overset{\bullet_j}{\bullet_k}), (\bullet_i \bullet_j \bullet_k) : i,j,k = 1 \ldots d\} ,$$

satisfying

1. $\langle \widehat{\mathbf{X}}^{(3)}_{st}, h \rangle = \langle \mathbf{X}^{(2)}_{st}, h \rangle$ when $h \in \widehat{\mathcal{H}}^{(2)}$
2. $\langle \widehat{\mathbf{X}}_{st}, \bullet_{(\bullet_i \overset{\bullet_k}{\bullet_j})} \rangle = \delta \widehat{X}^{(\bullet_i \overset{\bullet_k}{\bullet_j})}_{st}$ for $i,j = 1 \ldots d$.
3. $\langle \widehat{\mathbf{X}}_{st}, \bullet_{(\bullet_i \bullet_j \bullet_k)} \rangle = \delta \widehat{X}^{(\bullet_i \bullet_j \bullet_k)}_{st}$ for $i,j = 1 \ldots d$.

Now, since $\gamma + 3\gamma > 1$, objects of the form

$$\langle \widehat{\mathbf{X}}^{(3)}_{st}, \overset{\bullet_l}{\bullet}_{(\bullet_i \bullet_j \bullet_k)} \rangle = \int_s^t \delta X^l_{sr} d\widehat{X}^{(\bullet_i \bullet_j \bullet_k)}_r \quad \text{and} \quad \langle \widehat{\mathbf{X}}^{(3)}_{st}, \overset{\bullet_l}{\bullet}_{(\bullet_i \overset{\bullet_k}{\bullet_j})} \rangle = \int_s^t \delta X^l_{sr} d\widehat{X}^{(\bullet_i \overset{\bullet_k}{\bullet_j})}_r$$

can be constructed canonically as Young integrals and likewise for all 'new' components of $\widehat{\mathbf{X}}^{(3)}$. That is, given $\widehat{\mathbf{X}}^{(2)}$ we can canonically construct $\widehat{\mathbf{X}}^{(3)}$. Since we don't need to add any more bracket terms to our bracket extension, we set $\widehat{\mathbf{X}} = \widehat{\mathbf{X}}^{(3)}$. Proceeding as above, we obtain the Itô formula

$$\delta F(Y)_{st} = \int_s^t DF : (f(Y_r) \cdot dX_r) + \sum_{i,j=1}^{d} \int_s^t \frac{1}{2} D^2 F : (f_i, f_j)(Y_r) d\widehat{X}_r^{(\bullet_i \bullet_j)}$$

$$+ \sum_{i,j,k=1}^{d} \int_s^t D^2 F : (f_i, f_{[\bullet_k]_j}))(Y_r) d\widehat{X}_r^{(\bullet_i [\bullet_k]_j)} + \sum_{i,j,k=1}^{d} \int_s^t \frac{1}{6} D^3 F : (f_i, f_j, f_k)(Y_r) d\widehat{X}_r^{(\bullet_i \bullet_j \bullet_k)} .$$

As in the simple case, we can also define a general formula for the bizarre brackets $\langle\langle \bullet_i [\bullet_k]_j \rangle\rangle$, but it is slightly less natural. We will see that the same formula *almost* works, but with a minor clause. In particular, from the last example we saw that

$$\langle\langle \bullet_i [\bullet_k]_j \rangle\rangle = \bullet_i [\bullet_k]_j - [[\bullet_k]_j]_i - [\bullet_i \bullet_k]_j - [\bullet_k]_{(\bullet_i \bullet_j)} .$$

On the other hand, if we attempt to apply the definition from the simple case, we have that

$$\bullet_i [\bullet_k]_j - J_2 \Delta'(\bullet_i [\bullet_k]_j) = \bullet_i [\bullet_k]_j - J_2 \Big([\bullet_k]_j \tilde{\otimes} \bullet_i + \bullet_i \bullet_k \tilde{\otimes} \bullet_j + \bullet_k \tilde{\otimes} \bullet_i \bullet_j + \bullet_i \tilde{\otimes} [\bullet_k]_j \Big)$$

$$= \bullet_i [\bullet_k]_j - [[\bullet_k]_j]_i - [\bullet_i \bullet_k]_j - [\bullet_k]_{(\bullet_i \bullet_j)} + [\bullet_i]_{([\bullet_k]_j)} .$$

But the final term $[\bullet_i]_{([\bullet_k]_j)}$ is not defined since labels of the form $\bullet_{([\bullet_k]_j)}$ do not exist in the Hopf algebra $\widehat{\mathcal{H}}$. To get around this, we simply set such objects to zero.

5.3.6 The bracket extension of X

As explained in Example 5.3.5, the bracket extension must be defined on a larger Hopf algebra than in the simple case. We now define $\widehat{\mathcal{H}}^{(n)}$ as the Connes-Kreimer Hopf algebra generated by the alphabet

$$\mathcal{A}_n = \{1, \ldots, d\} \cup \{(\tau_1 \ldots \tau_k) : \tau_i \in \mathcal{T} \text{ with } |\tau_1| + \cdots + |\tau_k| \leq n \text{and } k \geq 2\} . \quad (5.27)$$

We then define $\widehat{\mathcal{H}} = \widehat{\mathcal{H}}^{(N)}$. The bracket polynomial $\langle\langle \tau_1 \ldots \tau_n \rangle\rangle$ will be defined similarly to the simple case. Recall the map $J_2(\bullet_{\alpha_1} \cdots \bullet_{\alpha_n} \tilde{\otimes} \bullet_{\beta_1} \cdots \bullet_{\beta_m}) = [\bullet_{\alpha_1} \cdots \bullet_{\alpha_n}]_{(\beta_1 \ldots \beta_m)}$, which was used to define the simple brackets. For the general brackets, we must extend this map. We define a bilinear map $J_2 : \mathcal{H} \tilde{\otimes} \mathcal{H}_N \to \widehat{\mathcal{H}}$ by the following rules. For any

$\sigma_1 \dots \sigma_m \in \mathcal{F} \cup \{1\}$ we have $J_2(\sigma_1 \dots \sigma_m \tilde{\otimes} 1) = 0$ and moreover

$$
\begin{aligned}
&J_2(\sigma_1 \dots \sigma_m \tilde{\otimes} \bullet_i) = [\sigma_1 \dots \sigma_m]_i && \text{for } i = 1 \dots d \ , \\
&J_2(\sigma_1 \dots \sigma_m \tilde{\otimes} \tau) = 0 && \text{for } \tau \in \mathcal{T}_N \text{ with } |\tau| \geq 2 \ , \\
&J_2(\sigma_1 \dots \sigma_m \tilde{\otimes} \tau_1 \dots \tau_n) = [\sigma_1 \dots \sigma_m]_{(\tau_1 \dots \tau_n)} && \text{for } |\tau_1| + \dots + |\tau_n| \leq N \text{ and } n \geq 2 \ .
\end{aligned}
$$

Hence, this is simply the "labelling map" $J_2(x \tilde{\otimes} y) = [x]_y$, with the exception that $[x]_\tau = 0$, since there does not exist a label $\bullet_\tau$ in $\widehat{\mathcal{H}}$. We could equally define $J_2(x \tilde{\otimes} y) = [x]_y$ without the clause for $y = \tau$, and simply quotient out all labels that do not appear in $\mathcal{A}_N$. Note that the third rule ensures that this agrees with the map used in the simple case, provided we identify the label $(\bullet_{\alpha_1} \cdots \bullet_{\alpha_n})$ with the previous notation $(\alpha_1 \dots \alpha_n)$.

We define the *bracket polynomial map* $\langle\!\langle \cdot \rangle\!\rangle : \mathcal{H}_N \to \widehat{\mathcal{H}}$ by

$$
\langle\!\langle h \rangle\!\rangle \stackrel{\text{def}}{=} h - J_2 \Delta'(h) \ ,
$$

for every $h \in \mathcal{H}_N$, where as usual Δ' is the reduced coproduct. Since this J_2 is an extension of the J_2 from the simple case, it is clear that $\langle\!\langle \bullet_{\gamma_1} \cdots \bullet_{\gamma_n} \rangle\!\rangle$ agrees with $\langle\!\langle \gamma_1 \dots \gamma_n \rangle\!\rangle$ in the simple case.

The general bracket map has all the properties of the simple version, and a few more.

1. The bracket polynomial is *symmetric*, since it is simply a function of the product $\tau_1 \dots \tau_n$, which is commutative.
2. We clearly have that $\langle\!\langle \bullet_i \rangle\!\rangle = \bullet_i$ and it is a nice exercise to check that $\langle\!\langle \tau \rangle\!\rangle = 0$ for any $\tau \in \mathcal{T}_N$ with $|\tau| \geq 2$.
3. If $|\tau_1| + \dots + |\tau_k| \leq n$, then $\langle\!\langle \tau_1 \dots \tau_k \rangle\!\rangle \in \widehat{\mathcal{H}}^{(n-1)}$.

The following Lemma is precisely the new version of Lemma 5.2.11.

Lemma 5.3.7. *If $h \in \mathcal{H}_N$, then the bracket map satisfies*

$$
\Delta \langle\!\langle h \rangle\!\rangle = \langle\!\langle h \rangle\!\rangle \tilde{\otimes} 1 + 1 \tilde{\otimes} \langle\!\langle h \rangle\!\rangle + h^1 \tilde{\otimes} \Big(\langle\!\langle h^2 \rangle\!\rangle - \bullet_{(h^2)} \Big)
$$

where we denote $\Delta' h = h^1 \tilde{\otimes} h^2$.

We can now define a bracket extension.

Definition 5.3.8. We define a *bracket extension* of $\mathbf{X}$ as any branched rough path $\widehat{\mathbf{X}}$ on $\widehat{\mathcal{H}}$ that is an extension of $\mathbf{X}$ in the sense that $\langle \widehat{\mathbf{X}}_{st}, h \rangle = \langle \mathbf{X}_{st}, h \rangle$ for any $h \in \mathcal{H} \subset \widehat{\mathcal{H}}$ and satisfies

$$\left\langle \widehat{\mathbf{X}}_{st}, \bullet_{(\tau_1 \dots \tau_k)} \right\rangle = \left\langle \widehat{\mathbf{X}}_{st}, \langle\!\langle \tau_1 \dots \tau_k \rangle\!\rangle \right\rangle \ , \tag{5.28}$$

for all $\tau_i \in \mathcal{T}$ with $|\tau_1| + \cdots + |\tau_k| \leq N$ and $k \geq 2$.

As in the simple case, the definition is well posed.

Proposition 5.3.9. *A bracket extension* $\widehat{\mathbf{X}}$ *of* $\mathbf{X}$ *exists.*

The proofs for both Lemma 5.3.7 and Proposition 5.3.9 follow in precisely the same way as in the simple case. However, since the setting is considerably different, we include both proofs in full.

Proof. We will proceed by induction. Suppose $\{\widehat{\mathbf{X}}^{(m)}\}_{1 \leq m \leq n}$ is a sequence of branched rough paths on $\{\widehat{\mathcal{H}}^{(m)}\}_{1 \leq m \leq n}$ such that, for each $1 \leq m \leq n-1$, we have

$$\langle \widehat{\mathbf{X}}_{st}^{(m+1)}, \bullet_{(\tau_1 \dots \tau_k)} \rangle = \langle \widehat{\mathbf{X}}_{st}^{(m)}, \langle\!\langle \tau_1 \dots \tau_k \rangle\!\rangle \rangle \ , \tag{5.29}$$

for $\tau_i \in \mathcal{T}$ with $|\tau_1| + \cdots + |\tau_k| \leq m+1$ for $k \geq 2$ and

$$\langle \widehat{\mathbf{X}}^{(m+1)}, h \rangle = \langle \widehat{\mathbf{X}}_{st}^{(m)}, h \rangle$$

for $h \in \widehat{\mathcal{H}}^{(m)}$. In particular, each element $\widehat{\mathbf{X}}^{(m)}$ in the sequence satisfies (5.28) for $|\tau_1| + \cdots + |\tau_k| \leq m$. We will now construct the next element in the sequence $\widehat{\mathbf{X}}^{(n+1)}$. Suppose $|\tau_1| + \cdots + |\tau_k| = n+1$, then we first set

$$\langle \widehat{\mathbf{X}}_{st}^{(n+1)}, \bullet_{(\tau_1 \dots \tau_k)} \rangle = \langle \widehat{\mathbf{X}}_{st}^{(n)}, \langle\!\langle \tau_1 \dots \tau_k \rangle\!\rangle \rangle \ . \tag{5.30}$$

For this to be a valid component of a branched rough path, we require that

$$\langle \widehat{\mathbf{X}}_{st}^{(n)}, \langle\!\langle \tau_1 \dots \tau_k \rangle\!\rangle \rangle = \langle \widehat{\mathbf{X}}_{su}^{(n)}, \langle\!\langle \tau_1 \dots \tau_k \rangle\!\rangle \rangle + \langle \widehat{\mathbf{X}}_{ut}^{(n)}, \langle\!\langle \tau_1 \dots \tau_k \rangle\!\rangle \rangle \ . \tag{5.31}$$

We will delay the verification of this fact for a moment. This implies the existence of paths $\widehat{X}^{(\tau_1 \dots \tau_k)}$ defined by

$$\delta \widehat{X}_{st}^{(\tau_1 \dots \tau_k)} = \langle \widehat{\mathbf{X}}_{st}^{(n)}, \langle\!\langle \tau_1 \dots \tau_k \rangle\!\rangle \rangle$$

for all such $\tau_1, \dots \tau_k$ with $k \geq 2$ and $|\tau_1| + \cdots + |\tau_k| = n+1$. Since $\widehat{\mathcal{H}}^{(n)}, \widehat{\mathcal{H}}^{(n+1)}$ are generated by the alphabets $\mathcal{A}_n, \mathcal{A}_{n+1}$, and $\mathcal{A}_n \subset \mathcal{A}_{n+1}$, it follows from Lemma 4.2.16 that there exists a branched rough path $\widehat{\mathbf{X}}^{(n+1)}$ on $\widehat{\mathcal{H}}^{(n+1)}$ obtained from $\widehat{\mathbf{X}}^{(n)}$ by adding the path elements $\widehat{X}^{(\tau_1 \dots \tau_k)}$. More precisely, $\widehat{\mathbf{X}}^{(n+1)}$ satisfies

$$\langle \widehat{\mathbf{X}}_{st}^{(n+1)}, \bullet_{(\tau_1 \dots \tau_k)} \rangle = \delta \widehat{X}_{st}^{(\tau_1 \dots \tau_k)} \ ,$$

for all such $\tau_1, \dots \tau_k$ with $k \geq 2$ and $|\tau_1| + \cdots + |\tau_k| = n + 1$ and

$$\langle \widehat{\mathbf{X}}^{(n+1)}_{st}, h \rangle = \langle \widehat{\mathbf{X}}^{(n)}_{st}, h \rangle$$

for all $h \in \widehat{\mathcal{H}}^{(n)}$. This proves the induction. If we set $\widehat{\mathbf{X}} = \widehat{\mathbf{X}}^{(N)}$, then the claim follows upon proving (5.31). But this follows from Lemma 5.3.7. Indeed, as in the left hand side of (5.31), using the notation $\Delta'(\tau_1 \dots \tau_k) = (\tau_1 \dots \tau_k)^1 \tilde{\otimes} (\tau_1 \dots \tau_k)^2$, we have that

$$\begin{aligned}
\langle \widehat{\mathbf{X}}^{(n+1)}_{st}, \bullet_{(\tau_1 \dots \tau_k)} \rangle &= \Big\langle \widehat{\mathbf{X}}^{(n)}_{st}, \langle\!\langle \tau_1 \dots \tau_k \rangle\!\rangle \Big\rangle = \Big\langle \widehat{\mathbf{X}}^{(n)}_{su} \tilde{\otimes} \widehat{\mathbf{X}}^{(n)}_{ut}, \Delta \langle\!\langle \tau_1 \dots \tau_k \rangle\!\rangle \Big\rangle \\
&= \Big\langle \widehat{\mathbf{X}}^{(n)}_{su} \tilde{\otimes} \widehat{\mathbf{X}}^{(n)}_{ut}, (\tau_1 \dots \tau_k)^1 \tilde{\otimes} \Big(\langle\!\langle (\tau_1 \dots \tau_k)^2 \rangle\!\rangle - \bullet_{(\tau_1 \dots \tau_k)^2} \Big) \Big\rangle \\
&+ \Big\langle \widehat{\mathbf{X}}^{(n)}_{su}, \langle\!\langle \tau_1 \dots \tau_k \rangle\!\rangle \Big\rangle + \Big\langle \widehat{\mathbf{X}}^{(n)}_{ut}, \langle\!\langle \tau_1 \dots \tau_k \rangle\!\rangle \Big\rangle \\
&= \Big\langle \widehat{\mathbf{X}}^{(n)}_{su} \tilde{\otimes} \widehat{\mathbf{X}}^{(n)}_{ut}, (\tau_1 \dots \tau_k)^1 \tilde{\otimes} \Big(\langle\!\langle (\tau_1 \dots \tau_k)^2 \rangle\!\rangle - \bullet_{(\tau_1 \dots \tau_k)^2} \Big) \Big\rangle \\
&+ \Big\langle \widehat{\mathbf{X}}^{(n+1)}_{su}, \bullet_{(\tau_1 \dots \tau_k)} \Big\rangle + \Big\langle \widehat{\mathbf{X}}^{(n+1)}_{ut}, \bullet_{(\tau_1 \dots \tau_k)} \Big\rangle \ .
\end{aligned}$$

And (5.31) follows from the fact that

$$\begin{aligned}
&\Big\langle \widehat{\mathbf{X}}^{(n)}_{su} \tilde{\otimes} \widehat{\mathbf{X}}^{(n)}_{ut}, (\tau_1 \dots \tau_k)^1 \tilde{\otimes} \Big(\langle\!\langle (\tau_1 \dots \tau_k)^2 \rangle\!\rangle - \bullet_{(\tau_1 \dots \tau_k)^2} \Big) \Big\rangle \\
&= \Big\langle \widehat{\mathbf{X}}^{(n)}_{su}, (\tau_1 \dots \tau_k)^1 \Big\rangle \Big\langle \widehat{\mathbf{X}}^{(n)}_{ut}, \langle\!\langle (\tau_1 \dots \tau_k)^2 \rangle\!\rangle - \bullet_{(\tau_1 \dots \tau_k)^2} \Big\rangle = 0 \ ,
\end{aligned}$$

which is a consequence of the fact that each term in the sum $(\tau_1 \dots \tau_n)^2$ is a forest $\sigma_1 \dots \sigma_m$ with $|\sigma_1| + \cdots + |\sigma_m| \leq n$ and by the inductive assumption (5.29) we have

$$\Big\langle \widehat{\mathbf{X}}^{(n)}_{st}, \langle\!\langle \sigma_1 \dots \sigma_m \rangle\!\rangle - \bullet_{(\sigma_1 \dots \sigma_m)} \Big\rangle = 0 \ ,$$

which proves the result. □

Proof of Lemma 5.3.7. We have that

$$\begin{aligned}
\Delta \langle\!\langle h \rangle\!\rangle &= \Delta h - \Delta J_2 \big(h^1 \tilde{\otimes} h^2 \big) = \Delta h - \Delta [h^1]_{(h^2)} \\
&= 1 \tilde{\otimes} \Big(h - J_2 \big(h^1 \tilde{\otimes} h^2 \big) \Big) + \Big(h - J_2 \big(h^1 \tilde{\otimes} h^2 \big) \Big) \tilde{\otimes} 1 + h^1 \tilde{\otimes} h^2 - \Delta' [h^1]_{(h^2)} \\
&= 1 \tilde{\otimes} \langle\!\langle h \rangle\!\rangle + \langle\!\langle h \rangle\!\rangle \tilde{\otimes} 1 + h^1 \tilde{\otimes} h^2 - \Big(h^{1(1)} \otimes [h^{1(2)}]_{(h^2)} - 1 \tilde{\otimes} [h^1]_{(h^2)} \Big) \ , \qquad (5.32)
\end{aligned}$$

where we denote $\Delta h^1 = h^{1(1)} \otimes h^{1(2)}$. Now, we also have that

$$
\begin{aligned}
h^{1(1)} \otimes [h^{1(2)}]_{(h^2)} &= (\mathrm{Id}\tilde{\otimes}J_2)\Big(h^{1(1)}\tilde{\otimes}h^{1(2)}\tilde{\otimes}h^2\Big) \\
&= (\mathrm{Id}\tilde{\otimes}J_2)\Big(h^{11}\tilde{\otimes}h^{12}\tilde{\otimes}h^2 + h^1\tilde{\otimes}1\tilde{\otimes}h^2 + 1\tilde{\otimes}h^1\tilde{\otimes}h^2\Big) , \qquad (5.33)
\end{aligned}
$$

where we denote $\Delta'(\tau^1) = \tau^{11}\tilde{\otimes}\tau^{12}$. But from the coassociativity of the coproduct (and hence reduced coproduct) we have that

$$
h^{11}\tilde{\otimes}h^{12}\tilde{\otimes}h^2 = (\Delta'\tilde{\otimes}\mathrm{Id})\Delta' h = (\mathrm{Id}\tilde{\otimes}\Delta')\Delta' h = h^1\tilde{\otimes}h^{21}\tilde{\otimes}h^{22} .
$$

Hence, (5.33) equals

$$
\begin{aligned}
&(\mathrm{Id}\tilde{\otimes}J_2)\Big(h^1\tilde{\otimes}h^{21}\tilde{\otimes}h^{22} + h^1\tilde{\otimes}1\tilde{\otimes}h^2 + 1\tilde{\otimes}h^1\tilde{\otimes}h^2\Big) \\
&\qquad = h^1\tilde{\otimes}\Big([h^{21}]_{(h^{22})} + \bullet_{(h^2)}\Big) + 1\tilde{\otimes}[h^1]_{(h^2)} .
\end{aligned}
$$

Substituting this back into (5.32), we obtain

$$
\Delta\langle\!\langle h\rangle\!\rangle = 1\tilde{\otimes}\langle\!\langle h\rangle\!\rangle + \langle\!\langle h\rangle\!\rangle\tilde{\otimes}1 + h^1\tilde{\otimes}\left(h^2 - [h^{21}]_{(h^{22})} - \bullet_{(h^2)}\right) ,
$$

which proves the claim. □

5.3.10 The change of variables formula

We can now prove the general version of the Itô formula. The proof is slightly more difficult than the simple case, and relies on some of the more subtle properties of the coefficients f_τ of the solution Y. As usual, we let $\mathbf{Y}$ be the controlled rough path solution to (5.21) with $\langle 1, \mathbf{Y}\rangle = Y$ and $\langle \tau, \mathbf{Y}\rangle = f_\tau(Y)$.

Theorem 5.3.11 (General version)**.** *Let $\mathbf{X}$ be a branched rough path above X. Let $\widehat{\mathbf{X}}$ be a bracket extension of $\mathbf{X}$. Then*

$$
\delta F(Y)_{st} = \int_s^t DF(Y_r) : (f(Y_r)\cdot dX_r) + \sum_{n=2}^{N}\int_s^t \frac{D^nF(Y_r)}{n!} : \big(f_{\tau_1}, \dots, f_{\tau_n}\big)(Y_r)\, d\widehat{X}_r^{(\tau_1\dots\tau_n)} , \qquad (5.34)
$$

where we sum over all $\tau_i \in \mathcal{T}$ such that $|\tau_1| + \dots + |\tau_n| \le N$.

As in the simple case, before proceeding with the proof we will first compute the controlled rough path integrals seen above. Let $g : \mathbb{R}^e \to \mathbb{R}^e$ be some smooth function, since $\mathbf{Y}$ is an $\widehat{\mathbf{X}}$-controlled rough path, we can define $g(\mathbf{Y})$ as an $\widehat{\mathbf{X}}$-controlled rough path. We have that $\langle 1, g(\mathbf{Y}_t)\rangle = g(Y_t)$ and

$$\langle \sigma_1 \dots \sigma_m, g(\mathbf{Y}(Y_t))\rangle = D^m g : \big(f_{\sigma_1}, \dots, f_{\sigma_m}\big)(Y_t) \ ,$$

since $f_{\sigma_i}(Y_t) = \langle \sigma_i, \mathbf{Y}_t\rangle$ are the coefficients of $\mathbf{Y}_t$. By definition of controlled rough path integrals, we have that

$$\begin{aligned}\int_s^t g(Y_r) d\widehat{X}_r^{(\tau_1\dots\tau_n)} &= g(Y_s)\langle \widehat{\mathbf{X}}_{st}, \bullet_{(\tau_1\dots\tau_n)}\rangle \\ &+ \sum_{m=1}^{N-1} \sum_{\lambda_1\dots\lambda_m\in\mathcal{F}_N} \langle \lambda_1 \dots \lambda_m, g(\mathbf{Y}_s)\rangle \left\langle \widehat{\mathbf{X}}_{st}, [\lambda_1 \dots \lambda_m]_{(\tau_1\dots\tau_n)}\right\rangle + o(|t-s|) \\ &= \sum_{m=0}^{N} \sum_{\lambda_1,\dots,\lambda_m} \frac{D^m g : \big(f_{\lambda_1}, \dots, f_{\lambda_m}\big)(Y_s)}{m!} \left\langle \widehat{\mathbf{X}}_{st}, [\lambda_1 \dots \lambda_m]_{(\tau_1\dots\tau_n)}\right\rangle + o(|t-s|) \ ,\end{aligned} \tag{5.35}$$

where in the last line we have used the symmetry of the expression to replace the unordered sum over *products* $\sum_{\lambda_1\dots\lambda_m\in\mathcal{F}_N}$ with the sum over all $\lambda_i \in \mathcal{T}_N$ denoted by $\sum_{\lambda_1,\dots,\lambda_m} 1/m!$.

Proof. As in the simple case, we perform arithmetic modulo $o(|t-s|)$. On the one hand, by performing a Taylor expansion and substituting in the controlled rough path expression for δY, we have that

$$\delta F(Y)_{st} = \sum_{k=1}^{N} \frac{D^k F : (f_{\sigma_1}, \dots, f_{\sigma_k})(Y_s)}{k!} \langle \mathbf{X}_{st}, \sigma_1 \dots \sigma_k\rangle \ , \tag{5.36}$$

where we sum over all $\sigma_i \in \mathcal{T}$ such that $|\sigma_1| + \dots + |\sigma_n| \leq N$. On the other hand, by (5.35) we have that

$$\begin{aligned}&\int_s^t DF(Y_r) : (f(Y_r) \cdot dX_r) \\ &\quad = \sum_{m=0}^{N-1} \frac{D^m\big(DF : f_i\big) : (f_{\lambda_1}, \dots, f_{\lambda_m})(Y_s)}{m!} \left\langle \widehat{\mathbf{X}}_{st}, [\lambda_1 \dots \lambda_m]_i \right\rangle \ ,\end{aligned}$$

where we sum over $i = 1 \dots d$ and all $\lambda_i \in \mathcal{T}$ with $|\lambda_1| + \dots + |\lambda_m| \leq N-1$. We also have that

$$\begin{aligned}&\int_s^t \frac{D^n F(Y_r)}{n!} : \big(f_{\tau_1}, \dots, f_{\tau_n}\big)(Y_r) d\widehat{X}_r^{(\tau_1\dots\tau_n)} \\ &= \sum_{m=0}^{N-1} \frac{D^m\big(D^n F : (f_{\tau_1}, \dots, f_{\tau_n}) : f_{\lambda_1}, \dots, f_{\lambda_m}\big)(Y_s)}{m!n!} \left\langle \widehat{\mathbf{X}}_{st}, [\lambda_1 \dots \lambda_m]_{(\tau_1\dots\tau_n)} \right\rangle \ ,\end{aligned}$$

where τ_i are fixed and we sum over all $\lambda_i \in \mathcal{T}$ with $|\lambda_1|+\cdots+|\lambda_m|+|\tau_1|+\cdots+|\tau_n| \leq N$. Hence, the right hand side of (5.34) can be written

$$\sum_{m=0}^{N-1} \frac{D^m\big(DF : f_i\big) : (f_{\lambda_1}, \ldots, f_{\lambda_m})(Y_s)}{m!} \left\langle \widehat{\mathbf{X}}_{st}, [\lambda_1 \ldots \lambda_m]_i \right\rangle$$
$$+ \sum_{n=2}^{N} \sum_{m=0}^{N-1} \frac{D^m\big(D^n F : (f_{\tau_1}, \ldots, f_{\tau_n}) : f_{\lambda_1}, \ldots, f_{\lambda_m}\big)(Y_s)}{m!n!} \left\langle \widehat{\mathbf{X}}_{st}, [\lambda_1 \ldots \lambda_m]_{(\tau_1 \ldots \tau_n)} \right\rangle \ , \tag{5.37}$$

where in the first term we sum over $i = 1 \ldots d$ all $\lambda_i \in \mathcal{T}$ with $|\lambda_1|+\cdots+|\lambda_m| \leq N-1$ and in the subsequent terms we sum over all $\lambda_i, \tau_i \in \mathcal{T}$ with $|\lambda_1|+\cdots+|\lambda_m|+|\tau_1|+\cdots+|\tau_n| \leq N$.

Let $p + p_1 + \cdots + p_n = m$ be an arbitrary partition of m and write

$$(\lambda_1, \ldots, \lambda_m) = (\lambda_1, \ldots, \lambda_p, \lambda_1^1, \ldots, \lambda_{p_1}^1, \ldots\ldots, \lambda_1^n, \ldots, \lambda_{p_n}^n) \ .$$

By Leibniz rule, we have that

$$\begin{aligned} &D^m(D^n F : (f_{\tau_1}, \ldots, f_{\tau_n}) : f_{\lambda_1}, \ldots, f_{\lambda_m}) \\ &= \sum_{p, p_1, \ldots, p_n} \binom{m}{p, p_1, \ldots, p_n} D^{p+n} F : (f_{\lambda_1}, \ldots, f_{\lambda_p}, u_1^{p_1}, \ldots, u_n^{p_n}) \ , \end{aligned} \tag{5.38}$$

where $\binom{m}{p,p_1,\ldots,p_n} = \frac{m!}{p!p_1!\ldots p_n!}$ and

$$u_i^{p_i} = D^{p_i} f_{\tau_i} : (f_{\lambda_1^i}, \ldots, f_{\lambda_{p_i}^i}) \ ,$$

and on the right hand side, we sum over all partitions $p + p_1 + \cdots + p_n = m$. Lemma 4.3.7 allows us to simplify this considerably. In particular, it implies that

$$u_i^{p_i} = f_{(\lambda_1^i \ldots \lambda_{p_i}^i) \star \sigma_i} \ .$$

Hence, if we substitute $k = p + n$, then (5.37) reduces to

$$\sum_{k=1}^{N-1} \sum_{m=k-1}^{N-1} \frac{D^k F : \big(f_{\lambda_1}, \dots f_{\lambda_{k-1}}, f_{(\lambda_1^1 \dots \lambda_{p_1}^1) \star \bullet_i}\big)}{k!} \frac{1}{p_1!} \binom{k}{1} \left\langle \widehat{\mathbf{X}}_{st}, [\lambda_1 \dots \lambda_m]_i \right\rangle$$
$$+ \sum_{n=2}^{N} \sum_{k=1}^{N-1} \sum_{m=k-n}^{N-1} \frac{D^k F : \big(f_{\lambda_1}, \dots, f_{\lambda_{k-n}}, f_{(\lambda_1^1 \dots \lambda_{p_1}^1) \star \tau_1}, \dots, f_{(\lambda_1^n \dots \lambda_{p_n}^n) \star \tau_n}\big)}{k!}$$
$$\times \frac{1}{p_1! \dots p_n!} \binom{k}{n} \left\langle \widehat{\mathbf{X}}_{st}, [\lambda_1 \dots \lambda_m]_{(\tau_1 \dots \tau_n)} \right\rangle \,, \tag{5.39}$$

where we sum over all partitions $p_1 + \cdots + p_n = m - k + n$ and all λ_i, τ_i as before. Now, recall that $J_2(x \otimes \bullet_i) = [x]_i$ and $J_2(x \otimes \tau) = 0$ when $\tau \in \mathcal{T}$ with $\tau \neq \bullet_i$ for any $i = 1 \dots d$. Then we can simplify (5.39) to

$$\sum_{k=1}^{N-1} \sum_{n=1}^{N-1} \sum_{m=k-n}^{N-1} \frac{D^k F : (f_{\lambda_1}, \dots, f_{\lambda_{k-n}}, f_{\left(\lambda_1^1 \dots \lambda_{p_1}^1\right) \star \tau_1}, \dots, f_{(\lambda_1^n \dots \lambda_{p_n}^n) \star \tau_n}\big)}{k!}$$
$$\times \frac{1}{p_1! \dots p_n!} \binom{k}{n} \left\langle \widehat{\mathbf{X}}_{st}, J_2\big(\lambda_1 \dots \lambda_m \tilde{\otimes} \tau_1 \dots \tau_n\big) \right\rangle \,, \tag{5.40}$$

where we sum over all partitions $p_1 + \cdots + p_n = m - k + n$ and all λ_i, τ_i as before, since all the unwanted terms of the form $[x]_\tau$ for $\tau \neq \bullet_i$ disappear. The expression (5.40) can be further simplified to

$$\sum_{k=1}^{N-1} \sum_{\sigma_1, \dots, \sigma_k} \frac{D^k F : (f_{\sigma_1}, \dots, f_{\sigma_k})}{k!} \langle \widehat{\mathbf{X}}_{st}, \theta(\sigma_1, \dots, \sigma_k) \rangle \,, \tag{5.41}$$

where we sum over $\sigma_i \in \mathcal{T}$ with $|\sigma_1| + \cdots + |\sigma_k| \leq N$ and where

$$\theta(\sigma_1, \dots, \sigma_k) = \sum_{n=1}^{N-1} \sum_{m=k-n}^{N-1} \sum_{\lambda} \sum_{\tau} \binom{k}{n} \left(\prod_{i=1}^{k-n} \langle \lambda_i, \sigma_i \rangle \right) \left(\prod_{j=1}^{n} \langle (\lambda_1^j \dots \lambda_{p_j}^j) \star \tau_j, \sigma_{k-n+j} \rangle \right)$$
$$\times J_2\big(\lambda_1 \dots \lambda_m \tilde{\otimes} \tau_1 \dots \tau_n\big) \,,$$

where we sum over all partitions $p_1 + \cdots + p_n = m - k + n$ and all λ_i, τ_i as before. The final step of the proof is to apply the following Lemma.

Lemma 5.3.12. *Let*

$$\mathrm{Sym}\,(\phi(\sigma_1, \dots, \sigma_k)) = \sum_{i \in \mathrm{Sym}(k)} \phi(\sigma_{i_1}, \dots, \sigma_{i_k}) \,,$$

where we sum over all permutations $(i_1, \dots, i_k)$ *of* $(1, \dots, k)$*. Then we have that*

$$\mathrm{Sym}\left(\theta(\sigma_1, \dots, \sigma_k)\right) = \mathrm{Sym}\left(J_2 \Delta'(\sigma_1 \dots \sigma_k) + \bullet_{(\sigma_1 \dots \sigma_k)}\right) ,$$

for any $\sigma_1, \dots, \sigma_k \in \mathcal{T}$ *with* $|\sigma_1| + \dots + |\sigma_k| \le N$*.*

In the following, we denote by $\sum_{\sigma_1 \dots \sigma_n \in \mathcal{F}_N}$ the sum over all unordered products in $\mathcal{F}_N$. Hence, since the function $D^k F : (f_{\sigma_1}, \dots, f_{\sigma_k})$ is symmetric under permutations of the elements σ_i, we have that (5.39) equals

$$\begin{aligned}
&\sum_{k=1}^{N-1} \sum_{\sigma_1, \dots, \sigma_k} \frac{D^k F : (f_{\sigma_1}, \dots, f_{\sigma_k})}{k!} \left\langle \widehat{\mathbf{X}}_{st}, \theta(\sigma_1, \dots, \sigma_k) \right\rangle \\
&= \sum_{k=1}^{N-1} \sum_{\sigma_1 \dots \sigma_k \in \mathcal{F}_N} \frac{D^k F : (f_{\sigma_1}, \dots, f_{\sigma_k})}{k!} \left\langle \widehat{\mathbf{X}}_{st}, \mathrm{Sym}\left(\theta(\sigma_1, \dots, \sigma_k)\right) \right\rangle \\
&= \sum_{k=1}^{N-1} \sum_{\sigma_1 \dots \sigma_k \in \mathcal{F}_N} \frac{D^k F : (f_{\sigma_1}, \dots, f_{\sigma_k})}{k!} \left\langle \widehat{\mathbf{X}}_{st}, \mathrm{Sym}\left(J_2 \Delta'(\sigma_1 \dots \sigma_k) + \bullet_{(\sigma_1 \dots \sigma_k)}\right) \right\rangle \\
&= \sum_{k=1}^{N-1} \sum_{\sigma_1, \dots, \sigma_k} \frac{D^k F : (f_{\sigma_1}, \dots, f_{\sigma_k})}{k!} \left\langle \widehat{\mathbf{X}}_{st}, J_2 \Delta'(\sigma_1 \dots \sigma_k) + \bullet_{(\sigma_1 \dots \sigma_k)} \right\rangle \\
&= \sum_{k=1}^{N-1} \sum_{\sigma_1, \dots, \sigma_k} \frac{D^k F : (f_{\sigma_1}, \dots, f_{\sigma_k})}{k!} \left\langle \widehat{\mathbf{X}}_{st}, \sigma_1 \dots \sigma_k \right\rangle ,
\end{aligned}$$

where the last line follows from the fact that $\widehat{\mathbf{X}}$ is a bracket extension of $\mathbf{X}$. The claim follows using the same argument employed at the end of the proof for the simple version.

□

Proof of Lemma 5.3.12. As always, we use the notation $\Delta \sigma_i = \sigma_i^{(1)} \tilde{\otimes} \sigma_i^{(2)}$, omitting the sum. To start the proof, notice that for $h \in \mathcal{H}_N$ we have that

$$\Delta' h = \sum_{m=1}^{N} \sum_{n=1}^{N} \frac{1}{n! m!} \langle \lambda_1 \dots \lambda_m \tilde{\otimes} \tau_1 \dots \tau_n, \Delta h \rangle \lambda_1 \dots \lambda_m \tilde{\otimes} \tau_1 \dots \tau_n ,$$

where we sum over all $\lambda_i, \sigma_i \in \mathcal{T}$ with $|\lambda_1| + \dots + |\lambda_m| + |\tau_1| + \dots + |\tau_n| \le N$. We also have that

$$\bullet_{(\sigma_1 \dots \sigma_k)} = \sum_{n=1}^{k} \frac{1}{n!} \langle 1 \otimes \tau_1 \dots \tau_n, \Delta(\sigma_1 \dots \sigma_n) \rangle J_2(1 \tilde{\otimes} \tau_1 \dots \tau_k) ,$$

where we sum over all τ_i with $|\tau_1| + \cdots + |\tau_n| \leq N$. In particular, we have that

$$
\begin{aligned}
&J_2\Delta'(\sigma_1 \ldots \sigma_k) + \bullet_{(\sigma_1\ldots\sigma_k)} \\
&\quad = \sum_{n=1}^{k}\sum_{m=0}^{N} \frac{1}{n!m!}\langle \lambda_1 \ldots \lambda_m \tilde{\otimes} \tau_1 \ldots \tau_n, \Delta(\sigma_1 \ldots \sigma_k)\rangle J_2(\lambda_1 \ldots \lambda_m \tilde{\otimes} \tau_1 \ldots \tau_n) \\
&\quad = \sum_{n=1}^{k}\sum_{m=1}^{N} \frac{1}{n!m!}\langle \lambda_1 \ldots \lambda_m, \sigma_1^{(1)} \ldots \sigma_k^{(1)}\rangle \langle \tau_1 \ldots \tau_n, \sigma_1^{(2)} \ldots \sigma_k^{(2)}\rangle J_2(\lambda_1 \ldots \lambda_m \tilde{\otimes} \tau_1 \ldots \tau_n)\ ,
\end{aligned}
$$

where we sum over all $\lambda_i, \tau_i \in \mathcal{T}$ with $|\lambda_1| + \cdots + |\lambda_m| + |\tau_1| + \cdots + |\tau_n| \leq N$. Since each $\sigma_i^{(2)} \in \mathcal{T} \cup \{1\}$, the terms $\langle \tau_1 \ldots \tau_n, \sigma_1^{(2)} \ldots \sigma_k^{(2)}\rangle$ will disappear unless $\sigma_j^{(2)} = 1$ for exactly $k-n$ of the factors in $\sigma_1^{(2)} \ldots \sigma_k^{(2)}$. If we apply Sym to the above expression, then we can assume that $\sigma_j^{(2)} = 1$ for $1 \leq j \leq k-n$, provided we include the combinatorial factor $\binom{k}{n}$. Of course, this implies that for $1 \leq j \leq k-n$, the only remaining term in the sum $\sigma_j^{(1)} \tilde{\otimes} \sigma_j^{(2)}$ will be $\sigma_j \tilde{\otimes} 1$. That is,

$$
\begin{aligned}
&\mathrm{Sym}\bigg(J_2\Delta'(\sigma_1 \ldots \sigma_k) + \bullet_{(\sigma_1\ldots\sigma_k)}\bigg) \\
&\quad = \mathrm{Sym}\bigg(\sum_{n=1}^{k}\sum_{m=1}^{N} \Big\langle \lambda_1 \ldots \lambda_m, \sigma_1 \ldots \sigma_{k-n}\sigma_{k-n+1}^{(1)} \ldots \sigma_k^{(1)}\Big\rangle \Big\langle \tau_1 \ldots \tau_n, \sigma_{k-n+1}^{(2)} \ldots \sigma_k^{(2)}\Big\rangle \\
&\qquad\qquad \times \frac{1}{m!n!}\binom{k}{n} J_2(\lambda_1 \ldots \lambda_m \tilde{\otimes} \tau_1 \ldots \tau_n)\bigg)\ , \qquad (5.42)
\end{aligned}
$$

where we sum over all $\lambda_i, \tau_i \in \mathcal{T}$ with $|\lambda_1| + \cdots + |\lambda_m| + |\tau_1| + \cdots + |\tau_n| \leq N$. Due to the symmetry in σ_i, we can make three replacements in (5.42). Firstly, we have

$$
\begin{aligned}
&\Big\langle \lambda_1 \ldots \lambda_m, \sigma_1 \ldots \sigma_{k-n}\sigma_{k-n+1}^{(1)} \ldots \sigma_k^{(1)}\Big\rangle \\
&= \binom{m}{k-n} \langle \lambda_1 \ldots \lambda_{k-n}, \sigma_1 \ldots \sigma_{k-n}\rangle \Big\langle \lambda_{k-n+1} \ldots \lambda_m, \sigma_{k-n+1}^{(1)} \ldots \sigma_k^{(1)}\Big\rangle \\
&= (k-n)!\binom{m}{k-n}\Big(\prod_{i=1}^{k-n} \langle \lambda_i, \sigma_i\rangle\Big) \Big\langle \lambda_{k-n+1} \ldots \lambda_m, \sigma_{k-n+1}^{(1)} \ldots \sigma_k^{(1)}\Big\rangle\ .
\end{aligned}
$$

Secondly we have

$$
\Big\langle \tau_1 \ldots \tau_n, \sigma_{k-n+1}^{(2)} \ldots \sigma_k^{(2)}\Big\rangle = n!\prod_{j=1}^{n} \Big\langle \tau_j, \sigma_{k-n+j}^{(2)}\Big\rangle
$$

And finally we have

$$\begin{aligned}&\left\langle\lambda_{k-n+1}\dots\lambda_m,\sigma^{(1)}_{k-n+1}\dots\sigma^{(1)}_k\right\rangle\\&\qquad=\sum_{p_1,\dots,p_n}\binom{m-k+n}{p_1,\dots,p_n}\left\langle\lambda^1_1\dots\lambda^1_{p_1},\sigma^{(1)}_{k-n+1}\right\rangle\dots\left\langle\lambda^n_1\dots\lambda^n_{p_n},\sigma^{(1)}_k\right\rangle\ ,\end{aligned}$$

where we sum over all partitions $p_1+\dots+p_n=m-k+n$. Putting this altogether, and using the fact that

$$\langle\lambda^j_1\dots\lambda^j_{p_j},\sigma^{(1)}_{k-n+j}\rangle\langle\tau_j,\sigma^{(2)}_{k-n+j}\rangle=\langle(\lambda^j_1\dots\lambda^j_{p_j})\star\tau_j,\sigma_{k-n+j}\rangle\ ,$$

we obtain

$$\begin{aligned}\mathrm{Sym}\Bigg(\sum_{n=1}^{k}\sum_{m=k-n}^{N}\Big(\prod_{i=1}^{k-n}\langle\lambda_i,\sigma_i\rangle\Big)\Big(\prod_{j=1}^{n}\left\langle(\lambda^j_1\dots\lambda^j_{p_j})\star\tau_j,\sigma_{k-n+j}\right\rangle\Big)&\\\times\binom{k}{n}J_2(\lambda_1\dots\lambda_m\tilde{\otimes}\tau_1\dots\tau_n)\Bigg)&\ ,\end{aligned}$$

where we sum over all λ_i,τ_i as above and all partitions $p_1+\dots+p_n=m-k+n$. This completes the proof. □

Bibliography

[AF03] R. A. ADAMS and J. J. F. FOURNIER. *Sobolev spaces*, vol. 140 of *Pure and Applied Mathematics (Amsterdam)*. Elsevier/Academic Press, Amsterdam, second ed., 2003.

[BLP78] A. BENSOUSSAN, J.-L. LIONS, and G. PAPANICOLAOU. *Asymptotic analysis for periodic structures*, vol. 5 of *Studies in Mathematics and its Applications*. North-Holland Publishing Co., Amsterdam, 1978.

[BM96] K. BURDZY and A. MPOLHKADRECKI. Itô formula for an asymptotically 4-stable process. *Ann. Appl. Probab.* **6**, no. 1, (1996), 200–217.

[BMP05] A. BENCHÉRIF-MADANI and É. PARDOUX. Homogenization of a diffusion with locally periodic coefficients. In *Séminaire de Probabilités XXXVIII*, vol. 1857 of *Lecture Notes in Math.*, 363–392. Springer, Berlin, 2005.

[BS10] K. BURDZY and J. SWANSON. A change of variable formula with Itô correction term. *Ann. Probab.* **38**, no. 5, (2010), 1817–1869.

[But72] J. C. BUTCHER. An algebraic theory of integration methods. *Math. Comp.* **26**, (1972), 79–106.

[Che77] K. T. CHEN. Iterated path integrals. *Bull. Amer. Math. Soc.* **83**, no. 5, (1977), 831–879.

[CK98] A. CONNES and D. KREIMER. Hopf algebras, renormalization and noncommutative geometry. *Comm. Math. Phys.* **199**, no. 1, (1998), 203–242.

[CQ02] L. COUTIN and Z. QIAN. Stochastic analysis, rough path analysis and fractional Brownian motions. *Probab. Theory Related Fields* **122**, no. 1, (2002), 108–140.

[Dav07] A. M. DAVIE. Differential equations driven by rough paths: an approach via discrete approximation. *Appl. Math. Res. Express.* , no. 2(2007).

[Del04] F. DELARUE. Auxiliary SDEs for homogenization of quasilinear PDEs with periodic coefficients. *Ann. Probab.* **32**, no. 3B, (2004), 2305–2361.

[DNR01] S. DĂSCĂLESCU, C. NĂSTĂSESCU, and S. RAIANU. *Hopf algebras: An introduction*, vol. 235 of *Monographs and Textbooks in Pure and Applied Mathematics*. Marcel Dekker Inc., New York, 2001.

[DPZ92] G. DA PRATO and J. ZABCZYK. *Stochastic equations in infinite dimensions*, vol. 44 of *Encyclopedia of Mathematics and its Applications*. Cambridge University Press, Cambridge, 1992.

[ELVE04] W. E, X. LI, and E. VANDEN-EIJNDEN. Some recent progress in multiscale modeling. In *Multiscale modelling and simulation*, vol. 39 of *Lect. Notes Comput. Sci. Eng.*, 3–21. Springer, Berlin, 2004.

[ER00] M. ERRAMI and F. RUSSO. n-covariation and symmetric sdes driven by finite cubic variation processes. *Preprint Prépublicatons Mathématiques de lUniversité Paris* **13**, no. 2000, (2000), 9.

[ER03] M. ERRAMI and F. RUSSO. n-covariation, generalized Dirichlet processes and calculus with respect to finite cubic variation processes. *Stochastic Process. Appl.* **104**, no. 2, (2003), 259–299.

[Fre64] M. I. FREĬDLIN. The Dirichlet problem for an equation with periodic coefficients depending on a small parameter. *Teor. Verojatnost. i Primenen.* **9**, (1964), 133–139.

[FS82] G. B. FOLLAND and E. M. STEIN. *Hardy spaces on homogeneous groups*, vol. 28 of *Mathematical Notes*. Princeton University Press, Princeton, N.J., 1982.

[FV06] P. FRIZ and N. VICTOIR. A note on the notion of geometric rough paths. *Probab. Theory Related Fields* **136**, no. 3, (2006), 395–416.

[FV10a] P. FRIZ and N. VICTOIR. Differential equations driven by Gaussian signals. *Ann. Inst. Henri Poincaré Probab. Stat.* **46**, no. 2, (2010), 369–413.

[FV10b] P. FRIZ and N. VICTOIR. *Multidimensional stochastic processes as rough paths*, vol. 120 of *Cambridge Studies in Advanced Mathematics*. Cambridge University Press, Cambridge, 2010. Theory and applications.

[GL89] R. GROSSMAN and R. G. LARSON. Hopf-algebraic structure of families of trees. *J. Algebra* **126**, no. 1, (1989), 184–210.

[GNRV05] M. GRADINARU, I. NOURDIN, F. RUSSO, and P. VALLOIS. m-order integrals and generalized Itô's formula: the case of a fractional Brownian motion with any Hurst index. *Ann. Inst. H. Poincaré Probab. Statist.* **41**, no. 4, (2005), 781–806.

[GRV03] M. GRADINARU, F. RUSSO, and P. VALLOIS. Generalized covariations, local time and Stratonovich Itô's formula for fractional Brownian motion with Hurst index $H \geq \frac{1}{4}$. *Ann. Probab.* **31**, no. 4, (2003), 1772–1820.

[Gub04] M. GUBINELLI. Controlling rough paths. *J. Funct. Anal.* **216**, no. 1, (2004), 86–140.

[Gub10] M. GUBINELLI. Ramification of rough paths. *J. Differential Equations* **248**, no. 4, (2010), 693–721.

[Hai09] M. HAIRER. *An Introduction to Stochastic PDEs.* Lecture Notes, 2009.

[HM12] M. HAIRER and J. MAAS. A spatial version of the ito-stratonovich correction. *Ann. Probab.* **40**, no. 4, (2012), 1675–1714.

[HP08] M. HAIRER and E. PARDOUX. Homogenization of periodic linear degenerate PDEs. *J. Funct. Anal.* **255**, no. 9, (2008), 2462–2487.

[HW74] E. HAIRER and G. WANNER. On the Butcher group and general multi-value methods. *Computing (Arch. Elektron. Rechnen)* **13**, no. 1, (1974), 1–15.

[Ich04] N. ICHIHARA. Homogenization problem for stochastic partial differential equations of Zakai type. *Stoch. Stoch. Rep.* **76**, no. 3, (2004), 243–266.

[Kir08] A. KIRILLOV, JR. *An introduction to Lie groups and Lie algebras*, vol. 113 of *Cambridge Studies in Advanced Mathematics*. Cambridge University Press, Cambridge, 2008.

[LCL07] T. LYONS, M. CARUANA, and T. LÉVY. *Differential equations driven by rough paths*, vol. 1908 of *Lecture Notes in Mathematics*. Springer, Berlin, 2007. Lectures from the 34th Summer School on Probability Theory held in Saint-Flour, July 6–24, 2004, With an introduction concerning the Summer School by Jean Picard.

[LQ02] T. LYONS and Z. QIAN. *System control and rough paths*. Oxford Mathematical Monographs. Oxford University Press, Oxford, 2002. Oxford Science Publications.

[LV07] T. LYONS and N. VICTOIR. An extension theorem to rough paths. *Ann. Inst. H. Poincaré Anal. Non Linéaire* **24**, no. 5, (2007), 835–847.

[Lyo98] T. LYONS. Differential equations driven by rough signals. *Rev. Mat. Iberoamericana* **14**, no. 2, (1998), 215–310.

[Man04] D. MANCHON. Hopf algebras, from basics to applications to renormalization. *ArXiv Mathematics e-prints* (2004). arXiv:arXiv:math/0408405.

[Par99] É. PARDOUX. Homogenization of linear and semilinear second order parabolic PDEs with periodic coefficients: a probabilistic approach. *J. Funct. Anal.* **167**, no. 2, (1999), 498–520.

[PS05] G. A. PAVLIOTIS and A. M. STUART. Periodic homogenization for inertial particles. *Phys. D* **204**, no. 3-4, (2005), 161–187.

[PS08] G. A. PAVLIOTIS and A. M. STUART. *Multiscale methods*, vol. 53 of *Texts in Applied Mathematics*. Springer, New York, 2008. Averaging and homogenization.

[PSV77] G. C. PAPANICOLAOU, D. STROOCK, and S. R. S. VARADHAN. Martingale approach to some limit theorems. In *Papers from the Duke Turbulence Conference (Duke Univ., Durham, N.C., 1976), Paper No. 6*, ii+120 pp. Duke Univ. Math. Ser., Vol. III. Duke Univ., Durham, N.C., 1977.

[Reu93] C. REUTENAUER. *Free Lie algebras*, vol. 7 of *London Mathematical Society Monographs. New Series*. The Clarendon Press Oxford University Press, New York, 1993. Oxford Science Publications.

[RS75] M. REED and B. SIMON. *Methods of modern mathematical physics. II. Fourier analysis, self-adjointness*. Academic Press [Harcourt Brace Jovanovich Publishers], New York, 1975.

[RY99] D. REVUZ and M. YOR. *Continuous martingales and Brownian motion*, vol. 293 of *Grundlehren der Mathematischen Wissenschaften [Fundamental Principles of Mathematical Sciences]*. Springer-Verlag, Berlin, third ed., 1999.

[SRP09] A. B. SOW, R. RHODES, and É. PARDOUX. Homogenization of periodic semilinear parabolic degenerate PDEs. *Ann. Inst. H. Poincaré Anal. Non Linéaire* **26**, no. 3, (2009), 979–998.

[Swa07] J. SWANSON. Variations of the solution to a stochastic heat equation. *Ann. Probab.* **35**, no. 6, (2007), 2122–2159.

[Swe69] M. E. SWEEDLER. *Hopf algebras*. Mathematics Lecture Note Series. W. A. Benjamin, Inc., New York, 1969.

[TM05] R. TEMAM and A. MIRANVILLE. *Mathematical modeling in continuum mechanics*. Cambridge University Press, Cambridge, second ed., 2005.

[Vic04] N. VICTOIR. Levy area for the free Brownian motion: existence and non-existence. *J. Funct. Anal.* **208**, no. 1, (2004), 107–121.

[Wag77] D. H. WAGNER. Survey of measurable selection theorems. *SIAM J. Control Optimization* **15**, no. 5, (1977), 859–903.

[WCD07] W. WANG, D. CAO, and J. DUAN. Effective macroscopic dynamics of stochastic partial differential equations in perforated domains. *SIAM J. Math. Anal.* **38**, no. 5, (2006/07), 1508–1527.

[WD07] W. WANG and J. DUAN. Homogenized dynamics of stochastic partial differential equations with dynamical boundary conditions. *Comm. Math. Phys.* **275**, no. 1, (2007), 163–186.

[You36] L. C. YOUNG. An inequality of the Hölder type, connected with Stieltjes integration. *Acta Math.* **67**, no. 1, (1936), 251–282.