An Introduction to the Theory of Rough Paths

Xi Geng

Abstract

These are extended lecture notes for a short summer course I gave in August, 2021. We give a self-contained introduction to Lyons' rough path theory from an equivalent perspective of Gubinelli. The core result is the existence, uniqueness and continuity of solutions to rough differential equations (Lyons' universal limit theorem).

In Chapter 1, we discuss the motivation of rough path theory and construct the essential rough path structures. In Chapter 2, we develop the main theory of rough differential equations. In Chapter 3, we study the algebraic basis of rough path theory. In Chapter 4, we discuss a few applications. The first two chapters are relatively self-contained and they form the core materials of the course. The third chapter is purely algebraic and essentially no analysis is involved. The last chapter focuses on conveying essential ideas and is thus highly non-technical.

Contents

1 Motivation and preliminary notions

We begin by describing the motivation of rough path theory and some of its fundamental points. The precise mathematics starts from Section 1.2 where we construct the essential rough path structures that are needed for the study of rough differential equations in the next chapter.

1.1 The philosophy of rough paths

Rough path theory, originally developed by Terry Lyons in his seminal work [Lyo98] in 1998, is an analytic theory of differential equations driven by multidimensional irregular paths (e.g. Brownian motion). Its development is partly motivated from the following re-examination of Itô's stochastic calculus from a pathwise/analytic perspective.

In the spirit of Itô's theory, solutions to stochastic differential equations (SDEs)

$$dX_t = \sigma(t, X_t)dB_t + b(t, X_t)dt,$$

where B_t is a Brownian motion, can be constructed in a probabilistic way by means of martingale methods (the martingale structure is hidden in the construction of stochastic integration and related properties). By realising the Brownian motion on the Wiener space (the path space equipped with the law of Brownian motion), one can view $X_t : \omega \mapsto X_t(\omega)$ as an a.s. well-defined measurable function of Brownian sample paths ω. However, this viewpoint does not bring up new insights since the solution is not obtained by solving the equation with a given fixed Brownian trajectory ω.

In a more generic form, one can raise the following basic question from a pathwise viewpoint.

Question. How can one make sense of a differential equation

$$dy_t = F(y_t)dx_t$$

where $x : [0, T] \to \mathbb{R}^d$ is a *deterministic* continuous path, e.g. a generic Brownian sample path?

Since differential equations are often interpreted and solved in integral form, it is natural to first address the question of constructing the integral

$$I_t(x, y) = \int_0^t y_s dx_s$$

where x, y come from a suitable class of continuous paths that is at least rich enough to include generic Brownian sample paths. If x, y have bounded total variation, the integral $I_t(x, y)$ can be defined in the classical Lebesgue-Stieltjes sense. If x, y are α-Hölder continuous with $\alpha > 1/2$, $I_t(x, y)$ can be constructed in the sense of Young [You36]. Essentially, one defines $I_t(x, y)$ through the Riemann sum approximation:

$$I_t(x, y) = \lim_{|\mathcal{P}| \to 0} \sum_{t_i \in \mathcal{P}} y_{t_{i-1}} (x_{t_i} - x_{t_{i-1}}), \tag{1.1}$$

whose limit can be shown to exist in both cases. Here $\mathcal{P}$ is an arbitrary finite partition of $[0, t]$ and $|\mathcal{P}|$ denotes its mesh size. However, the regularity regime of $\alpha > 1/2$ is unfortunately insufficient to cover the Brownian motion case as generic Brownian paths are only α-Hölder continuous when $\alpha < 1/2$. It is tempting to ask, when $\alpha \leqslant 1/2$, if the approximation (1.1) is sufficient for defining the integral $I_t(x, y)$? The answer is no. Indeed, (1.1) is only a *first order* approximation that is not accurate enough to yield convergence in the rougher regime $\alpha \leqslant 1/2$!

There is another natural attempt for a suitable definition of $I_t(x, y)$ from a functional analytic perspective. One knows that $(x, y) \mapsto I_t(x, y)$ is well-defined when x, y are smooth. Is it possible to construct the mapping I_t by taking completion of smooth paths with respect to a suitable topology on path space? A natural candidate of path topology is the uniform topology. However, the following negative example shows that I_t fails to be continuous with respect to the uniform topology.

Example 1.1. Let

$$(x_t^{(n)}, y_t^{(n)}) \triangleq \big(\frac{1}{n} \sin n^2 t, \frac{1}{n} \cos n^2 t\big), \quad 0 \leqslant t \leqslant T.$$

It is clear that $x^{(n)}, y^{(n)}$ both converge to zero uniformly. However, from explicit calculation one finds that

$$I_t(x^{(n)}, y^{(n)}) = \int_0^t y_s^{(n)} dx_s^{(n)} = \frac{t}{2} + \frac{1}{4n^2} \sin 2n^2 t,$$

which does not converge to the zero path.

As suggested by Young's integration theory, the α-Hölder topology with $\alpha \in (1/2, 1]$ does work. However, the completion of smooth paths with respect to this topology is not rich enough to at least cover the Brownian motion case. Is there a clever choice of path topology which on the one hand ensures the continuity of I_t and on the other hand is weak enough to contain Brownian sample paths in the completion of smooth paths? Unfortunately, the following negative result (cf. Friz-Hairer [FH14]) indicates that the answer is essentially *no*.

Proposition 1.2. *Let $\mathcal{W}$ denote the space of continuous paths $x : [0,T] \to \mathbb{R}^d$. There does not exist a norm $\|\cdot\|$ on the space $E^\infty \subseteq \mathcal{W}$ of smooth paths such that:*

(i) the closure E of E^∞ under $\|\cdot\|$ contains almost all Brownian rough paths (i.e. $\mu(E) = 1$ where μ is the law of Brownian motion);
(ii) the restriction of I_t on $E^\infty \times E^\infty$ extends continuously to $E \times E$ with respect to $\|\cdot\|$.

The failure of the convergence in (1.1) and the result of Proposition 1.2 suggest that something more fundamental than the classical viewpoint is missing. As we will see, the missing point is a suitable way of looking at paths: a rough path should be an *enhanced object* in which the original trajectory $x : [0,T] \to \mathbb{R}^d$ is embedded as a first level structure.

The following formal calculation reveals why paths need to be enhanced to include higher order structure that is not encoded in the original trajectory x. Recall that we wish to define the integral $\int_0^t y_s dx_s$. Let us assume for now that y_s has the form $y_s = F(x_s)$ where F is a smooth function. By a formal Taylor expansion of F, one can write

$$\begin{aligned}
\int_s^t F(x_u)dx_u &= F(x_s)(x_t - x_s) + \int_s^t (F(x_u) - F(x_s))dx_u \\
&= F(x_s)(x_t - x_s) + \int_s^t \Big(\int_s^u DF(x_v)dx_v\Big)dx_u \\
&= F(x_s)(x_t - x_s) + DF(x_s)\int_s^t \int_s^u dx_v dx_u \\
&\quad + \int_s^t \int_s^u \big(DF(x_v) - DF(x_s)\big)dx_v dx_u + \cdots \\
&= F(x_s)(x_t - x_s) + DF(x_s)\int_s^t \int_s^u dx_v dx_u \\
&\quad + D^2F(x_s)\int_s^t \int_s^u \int_s^v dx_r dx_v dx_u + \cdots .
\end{aligned}$$

From this expansion, the accurate definition of $\int_s^t F(x_u)dx_u$ should depend on the quantities

$$x_t - x_s,\ \int_s^t \int_s^u dx_v dx_u,\ \int_s^t \int_s^u \int_s^v dx_r dx_v dx_u,\ \cdots \tag{1.2}$$

It is important to note that the above notation is multidimensional. Assuming that x takes values in $\mathbb{R}^d$, the product $dx_v dx_u$ is not the usual scalar multiplication. Indeed, the integral $\int_s^t \int_s^u dx_v dx_u$ consists of $d \times d$ coordinate components

$$\int_s^t \int_s^u dx_v^i dx_u^j \quad (i, j = 1, \cdots, d).$$

A proper way of encoding this information is through the notion of tensor products (cf. Example 1.3 in Section 1.2.1). Let us not bother with this at the moment.

If dimension $d = 1$, the product $dx_v dx_u$ is indeed the (commutative) real multiplication. The integrals in (1.2) can all be evaluated (or defined) explicitly as

$$\int_s^t \int_s^u dx_v dx_u = \frac{1}{2}(x_t - x_s)^2,\ \int_s^t \int_s^u \int_s^v dx_r dx_v dx_u = \frac{1}{6}(x_t - x_s)^3 \text{ etc}$$

regardless of the regularity of x. If dimension $d > 1$ and if x has Hölder regularity $\alpha > 1/2$, the integrals in (1.2) are all canonically well-defined in the sense of Young [You36]. In both case (cf. (1.1) for the latter case), the values of these integrals are all uniquely determined by the original path x, more precisely, by the information encoded in the family of increments

$$\{x_v - x_u : 0 \leqslant u < v \leqslant T\}.$$

However, for a generic multidimensional α-Hölder continuous path x with $\alpha \leqslant 1/2$, the iterated integrals in (1.1) are no longer (and cannot be!) well-defined in any reasonable sense. These integrals need to be *specified* together with the original path in the definition of a rough path.

Summarised concisely, the total number N of iterated integrals that need to be specified is $N = \lfloor 1/\alpha \rfloor$ (the integer part of $1/\alpha$). A *rough path* $\mathbf{X}$ with roughness α (measured by the Hölder regularity α) is a continuous functional

$$\mathbf{X} : (s, t) \mapsto (X_{s,t}^1, X_{s,t}^2, \cdots, X_{s,t}^N) \in G$$

that is α-Hölder continuous with respect to a suitable metric on G in a proper sense. Here the component $X_{s,t}^n$ formally resembles the n-th order iterated integral

$$X_{s,t}^n \doteq \int_{s<t_1<\cdots<t_n<t} dx_{t_1} \cdots dx_{t_n} \tag{1.3}$$

and G is a natural algebraic space in which these iterated integrals live. The space G is defined through the algebraic constraints that are intrinsically satisfied by the classical iterated integrals. The precise mathematical definitions are given in Section 1.2.2.

Once a rough path $\mathbf{X}$ is properly defined, the path integral $\int_0^t F(\mathbf{X}_s)d\mathbf{X}_s$ can now be constructed through an enhanced Riemann sum approximation scheme. Mathematically, one can show that, when $\mathbf{X} = (X^1, \cdots, X^N)$ is an α-Hölder rough path, the limit

$$\lim_{|\mathcal{P}|\to 0} \sum_{t_i \in \mathcal{P}} \left(F(X^1_{0,t_{i-1}})X^1_{t_{i-1},t_i} + DF(X^1_{0,t_{i-1}})X^2_{t_{i-1},t_i} + \cdots + D^{N-1}F(X^1_{0,t_{i-1}})X^N_{t_{i-1},t_i} \right) \tag{1.4}$$

exists and can be taken as the definition of the integral $\int_0^t F(\mathbf{X})d\mathbf{X}$.

The construction of a more general integral "$\int_0^t y_s dx_s$" requires extra effort. Inspired by (1.4), an essential point is that the integrand y cannot be interpreted as a single path either. Instead, it also needs to be defined as a multi-level object

$$\mathcal{Y}_t = (Y^0_t, Y^1_t, \cdots, Y^{N-1}_t),$$

where Y^0_t represents the original path y_t, and $Y^1, \cdots, Y^{N-1}$ are "derivative paths" of Y^0 with respect to $\mathbf{X}$. These derivative paths are all prespecified along with Y^0 in the definition of $\mathcal{Y}$. It turns out that the limit of the Riemann sum approximation

$$\lim_{|\mathcal{P}|\to 0} \sum_{t_i \in \mathcal{P}} \left(Y^0_{t_{i-1}} X^1_{t_{i-1},t_i} + \cdots + Y^{N-1}_{t_{i-1}} X^N_{t_{i-1},t_i} \right) \tag{1.5}$$

exists and defines the integral $\int_0^t \mathcal{Y}d\mathbf{X}$. Once this part is developed in a proper way, differential equations of the form

$$d\mathcal{Y} = F(\mathcal{Y})d\mathbf{X}$$

can naturally be interpreted in integral form and their solutions can be constructed by means of fixed point arguments (as fixed points of the integral transformation $\mathcal{Y} \mapsto Y_0 + \int F(\mathcal{Y})d\mathbf{X}$).

The above philosophical discussion outlines the essence of rough path theory. The theory reduces to ordinary calculus when x is smooth. The following fundamental result, which was originally due to Lyons [Lyo98], shows that rough path theory is a natural extension of classical ODE theory.

The Universal Limit Theorem. The solution map, which sends the triple $(Y_0, F, \mathbf{X})$ consisting of the initial condition Y_0, the coefficient function F and the driving path $\mathbf{X}$ to the solution path $\mathcal{Y}$, is jointly continuous with respect to suitable topologies on the relevant objects.

The development of rough path theory has enormous applications in stochastic analysis, partly due to the following two reasons:

(i) Many continuous-time stochastic processes can be lifted as rough paths in a canonical way;
(ii) To some extent, rough path theory provides robust analytic tools that simplify/complement probabilistic considerations and overcome difficulties arising from the probabilistic side.

In these notes, we develop the program towards the universal limit theorem in a precise mathematical way. We follow the controlled rough path approach developed by Gubinelli [Gub04]. This is essentially equivalent to Lyons' original approach but is technically simpler to some extent (e.g. some algebraic considerations are simplified).

Organisation. In Section 1.2, we introduce the core rough path spaces that are relevant for our study. In Chapter 2, we develop the main ingredients towards the theory of rough differential equations and the universal limit theorem. In Chapter 3, we study the algebraic foundation of rough paths from the perspective of free Lie algebras. In Chapter 4, we discuss a few applications of rough path theory.

1.2 Rough path spaces

In this section, we define the core rough path spaces. Following Gubinelli's approach, there are two fundamental types of rough paths that are relevant to us: weakly geometric rough paths and controlled rough paths. A weakly geometric rough path $\mathbf{X}$ plays the role of a driving path while a controlled rough path $\mathcal{Y}$ acts as an integrand: a rough integral $\int \mathcal{Y} d\mathbf{X}$ is to be formed.

1.2.1 Tensor product spaces

The precise definition of rough paths relies critically on the notion of tensor products, as the latter provides an effective way of capturing the non-commutativity of iterated integrals. Before introducing rough path structures, we shall first recall the relevant algebraic concepts.

Tensor products and admissible tensor norms

Let V be a real vector space. The *algebraic tensor product* $V \otimes_{\mathrm{a}} V$ (also denoted as $V^{\otimes_{\mathrm{a}}2}$) is the vector space generated by the symbolic monomials $v \otimes w$ $(v, w \in V)$ modulo the following relations:

$$\begin{cases} (av_1 + v_2) \otimes w = a(v_1 \otimes w) + v_2 \otimes w, \\ v \otimes (aw_1 + w_2) = a(v \otimes w_1) + v \otimes w_2, \end{cases} \quad \forall a \in \mathbb{R}, v, v_1, v_2, w, w_1, w_2 \in V. \tag{1.6}$$

In other words, a generic element ξ in $V \otimes_{\mathrm{a}} V$ has the (non-unique!) form

$$\xi = \sum_{i=1}^{r} c_i v_i \otimes w_i,$$

where we identify different expressions according to the relations (1.6). For instance, $(2v_1 + 6v_2) \otimes w$ and $2v_1 \otimes w + (3v_2) \otimes (2w)$ represent the same element in $V \otimes_{\mathrm{a}} V$.

Example 1.3. Let $V = \mathbb{R}^d$ be equipped with the standard basis $\{e_1, \cdots, e_d\}$. Then $V \otimes_{\mathrm{a}} V$ is a d^2-dimensional vector space with basis

$$\{e_i \otimes e_j : 1 \leqslant i, j \leqslant d\}.$$

Every element ξ in $V \otimes_{\mathrm{a}} V$ has a unique representation

$$\xi = \sum_{i,j=1}^{d} c_{ij} e_i \otimes e_j, \quad c_{ij} \in \mathbb{R}.$$

If $x : [0, T] \to \mathbb{R}^d$ is a smooth path, one has

$$\int_{s<u<v<t} dx_u \otimes dx_v = \sum_{i,j=1}^{d} \Big(\int_{s<u<v<t} \dot{x}_u^i \dot{x}_v^j du dv \Big) e_i \otimes e_j,$$

where $\dot{x}$ denotes the derivative of x.

For each $n \geqslant 1$, one can define the n-th algebraic tensor product space $V^{\otimes_{\mathrm{a}} n}$ in a similar way. Given $\xi \in V^{\otimes_a m}$ and $\eta \in V^{\otimes_a n}$, one can form the tensor product $\xi \otimes \eta \in V^{\otimes_a (m+n)}$. This tensor multiplication is associative:

$$(\xi \otimes \eta) \otimes \zeta = \xi \otimes (\eta \otimes \zeta).$$

The above definition is purely algebraic. To perform analysis, we shall assume from now on that V is a (real) Banach space and equip the tensor product spaces with suitable norms. A family of *admissible tensor norms* on $(V^{\otimes_{\mathrm{a}} n})_{n=1}^{\infty}$ is a family of norms $\{|\cdot|_n\}$, one for each of $V^{\otimes_{\mathrm{a}} n}$, such that the following two properties hold true.

(i) For any $\xi \in V^{\otimes_{\mathrm{a}} m}$ and $\eta \in V^{\otimes_{\mathrm{a}} n}$,

$$|\xi \otimes \eta|_{m+n} \leqslant |\xi|_m \cdot |\eta|_n.$$

(ii) For any permutation σ of order n, let $P_\sigma : V^{\otimes_{\mathrm{a}} n} \to V^{\otimes_{\mathrm{a}} n}$ denote the linear transformation induced by

$$P_\sigma(v_1 \otimes \cdots \otimes v_n) \triangleq v_{\sigma(1)} \otimes \cdots \otimes v_{\sigma(n)}, \quad v_1, \cdots, v_n \in V. \tag{1.7}$$

Then

$$|P_\sigma(\xi)|_n = |\xi|_n \quad \forall \xi \in V^{\otimes_{\mathrm{a}} n}.$$

Suppose that a family of admissible tensor norms $\{|\cdot|_n : n \geqslant 1\}$ are given fixed. The n-th *tensor product* $V^{\otimes n}$ is defined to be the completion of $V^{\otimes_{\mathrm{a}} n}$ under $|\cdot|_n$. It is clear that Properties (i), (ii) remain valid over $V^{\otimes n}$.

Remark 1.4. If $\dim V < \infty$, the algebraic tensor product and its completion are the same thing. Since all norms are equivalent in finite dimensions, the caution about choosing tensor norms is only needed in the infinite dimensional case.

Example 1.5. Let $V = \mathbb{R}^d$. Suppose that V is equipped with the Euclidean norm. An example of admissible tensor norms is the Hilbert-Schmidt norm, namely the norm on $V^{\otimes n}$ is induced by the inner product

$$\langle v_1 \otimes \cdots \otimes v_n, w_1 \otimes \cdots \otimes w_n \rangle_{\mathrm{HS}} \triangleq \langle v_1, w_1 \rangle_{\mathbb{R}^d} \cdots \langle v_n, w_n \rangle_{\mathbb{R}^d}, \quad v_i, w_j \in \mathbb{R}^d.$$

If $\{e_1, \cdots, e_d\}$ is an orthonormal basis of V, then

$$\{e_{i_1} \otimes \cdots \otimes e_{i_n} : 1 \leqslant i_1, \cdots, i_n \leqslant d\}$$

is an orthonormal basis of $V^{\otimes n}$.

Suppose on the other hand that V is equipped with the l^1-norm:

$$|x|_{l^1} \triangleq |x_1| + \cdots + |x_d|, \quad x = \sum_{i=1}^{d} x_i e_i \in \mathbb{R}^d.$$

Another example of admissible tensor norms is the induced l^1-norm:

$$|\xi|_{l^1} \triangleq \sum_{i_1, \cdots, i_n = 1}^{d} |c_{i_1 \cdots i_n}| \quad \text{where } \xi = \sum c_{i_1 \cdots i_n} e_{i_1} \otimes \cdots \otimes e_{i_n} \in V^{\otimes n}.$$

Throughout the rest of the notes, we will always assume that a family of admissible tensor norms are given fixed and work on the completions $\{V^{\otimes n} : n \geqslant 1\}$.

Truncated tensor algebras and basic operations

Let $N \geqslant 1$ be given. The *N-th truncated tensor algebra* over V is defined by

$$T^N(V) \triangleq \bigoplus_{n=0}^{N} V^{\otimes n} = \{\xi = (\xi_0, \xi_1, \cdots, \xi_N) : \xi_n \in V^{\otimes n},\ 0 \leqslant n \leqslant N\} \tag{1.8}$$

with the convention that $V^{\otimes 0} \triangleq \mathbb{R}$. It is clear that $T^N(V)$ is a vector space. In addition, the tensor product $\otimes$ extends to a natural multiplication on $T^N(V)$: given $\xi = (\xi_0, \cdots, \xi_N)$ and $\eta = (\eta_0, \cdots, \eta_N)$ in $T^N(V)$, one defines $\xi \otimes \eta \in T^N(V)$ by

$$(\xi \otimes \eta)_n \triangleq \sum_{k=0}^{n} \xi_k \otimes \eta_{n-k}, \quad n = 0, \cdots, N. \tag{1.9}$$

It follows that $(T^N(V), +, \otimes)$ is an algebra with unit $\mathbf{1} \triangleq (1, 0, \cdots, 0)$. Note that not every element in $T^N(V)$ has an inverse under $\otimes$. Nonetheless, if $\xi = (1, \xi_1, \cdots, \xi_N)$, then ξ is invertible with inverse

$$\xi^{-1} = \sum_{n=0}^{N} (-1)^n (\xi - \mathbf{1})^{\otimes n}. \tag{1.10}$$

Remark 1.6. The roles of $V^{\otimes n}$ and $T^N(V)$ is that an n-th order iterated path integral (1.3) takes values in $V^{\otimes n}$ and a generic rough path takes values in $T^N(V)$ with a suitable N depending on the Hölder regularity of the path.

Exercise 1.7. Prove (1.10) for any $\xi = (1, \xi_1, \cdots, \xi_N) \in T^N(V)$.

1.2.2 Weakly geometric rough paths

As we explained in the introduction, a rough path should be an object consisting of a first level path together with prespecified "iterated integral paths" up to a certain order. We now make the definition precise. Let V be a Banach space. For each $N \geqslant 1$, define

$$T_1^N(V) \triangleq \{\xi = (\xi_0, \cdots, \xi_N) \in T^N(V) : \xi_0 = 1\}.$$

One knows from (1.10) that $(T_1^N(V),\otimes)$ is a group. We also set

$$\Delta_T \triangleq \{(s,t) : 0 \leqslant s \leqslant t \leqslant T\}.$$

Definition 1.8. Let

$$\mathbf{X}_{\cdot,\cdot} = (1, X^1_{\cdot,\cdot}, \cdots, X^N_{\cdot,\cdot}) : \Delta_T \to T_1^N(V)$$

be a given continuous mapping. We say that $\mathbf{X}$ is a *multiplicative functional* if it satisfies the following so-called *Chen's identity*:

$$\mathbf{X}_{s,u} = \mathbf{X}_{s,t} \otimes \mathbf{X}_{t,u} \quad \text{for all } s \leqslant t \leqslant u \text{ in } [0,T]. \tag{1.11}$$

Definition 1.9. Let $\alpha \in (0,1]$ be given fixed and denote $N_\alpha \triangleq \lfloor 1/\alpha \rfloor$ (the integer part of $1/\alpha$). An *α-Hölder rough path* over V is a multiplicative functional $\mathbf{X} : \Delta_T \to T_1^{N_\alpha}(V)$ that is α-Hölder continuous in the following sense:

$$\|X^n\|_{n\alpha} \triangleq \sup_{0\leqslant s<t\leqslant T} \frac{|X^n_{s,t}|}{(t-s)^{n\alpha}} < \infty, \quad \text{for each } n = 1, \cdots, N_\alpha. \tag{1.12}$$

Remark 1.10. Since the role of a rough path $\mathbf{X}$ is to drive an integral or differential equation, the only relevant information is the increment $\mathbf{X}_{s,t}$ rather than the actual path indexed by a single t. Nonetheless, by using the group multiplication $\otimes$ one can easily switch between the two objects: given increments $\mathbf{X}_{s,t}$ one can construct a path by $\mathbf{X}_t \triangleq \mathbf{X}_{0,t}$, while given a path $\mathbf{X}_t \in T_1^N(V)$ one can construct the increments $\mathbf{X}_{s,t} \triangleq \mathbf{X}_s^{-1} \otimes \mathbf{X}_t$.

The following example provides an important motivation of Definition 1.9.

Example 1.11. Let $x : [0,T] \to V$ be a smooth path. Then for any $\alpha \in (0,1]$, x can be lifted as an α-Hölder rough path in a canonical way. Indeed, one defines $\mathbf{X}_{s,t} = (1, X^1_{s,t}, \cdots, X^{N_\alpha}_{s,t})$ by

$$X^n_{s,t} = \int_{s<t_1<\cdots<t_n<t} dx_{t_1} \otimes \cdots \otimes dx_{t_n} = \int_{s<t_1<\cdots<t_n<t} \dot{x}_{t_1} \otimes \cdots \otimes \dot{x}_{t_n} dt_1 \cdots dt_n \tag{1.13}$$

for each $n = 1, \cdots, N_\alpha$. Note that $X^1_{s,t} = x_t - x_s$ and all the X^n's are determined by the first level path X^1 in this case. Chen's identity (1.11) can be justified as follows:

$$
\begin{aligned}
X^n_{s,u} &= \int_{s<t_1<\cdots<t_n<u} dx_{t_1}\otimes\cdots\otimes dx_{t_n} \\
&= \sum_{k=0}^{n} \int_{\substack{s<t_1<\cdots<t_k<t \\ t<t_{k+1}<\cdots<t_n<u}} dx_{t_1}\otimes\cdots\otimes dx_{t_k}\otimes dx_{t_{k+1}}\otimes\cdots\otimes dx_{t_n} \\
&= \sum_{k=0}^{n} \Big(\int_{s<t_1<\cdots<t_k<t} dx_{t_1}\otimes\cdots\otimes dx_{t_k}\Big)\otimes\Big(\int_{t<t_{k+1}<\cdots<t_n<u} dx_{t_{k+1}}\otimes\cdots\otimes dx_{t_n}\Big) \\
&= \sum_{k=0}^{n} X^k_{s,t}\otimes X^{n-k}_{t,u}.
\end{aligned}
$$

In addition, one has

$$
|X^n_{s,t}| \leqslant \int_{s<t_1<\cdots<t_n<t} |\dot{x}_{t_1}|\cdots|\dot{x}_{t_n}|dt_1\cdots dt_n \leqslant \frac{\|\dot{x}\|_\infty^n}{n!}|t-s|^n,
$$

from which the regularity condition (1.12) follows trivially.

Remark 1.12. The iterated integral (1.13) is well-defined in the classical sense of Young [You36] when x is α-Hölder continuous with $\alpha > 1/2$. However, it becomes formal when $\alpha \leqslant 1/2$. In this case, the components $X^2, \cdots, X^{N_\alpha}$ are no longer determined by the first level path X^1. The N_α-tuple $(X^1, \cdots, X^{N_\alpha})$ has to be *given* as a whole object in the definition of $\mathbf{X}$. From the formal representation (1.13), one also sees why the regularity condition (1.12) needs to be in place: when s, t are close one should heuristically expect that

$$
\Big|\int_{s<t_1<\cdots<t_n<t} dX^1_{t_1}\otimes\cdots\otimes dX^1_{t_n}\Big| \lesssim |t-s|^{n\alpha},
$$

since each "$dX^1_{t_i}$" is formally of order $|t-s|^\alpha$ by the α-Hölder continuity of X^1.

The space of rough paths can be metrised in the following way. Let $\mathbf{X}, \tilde{\mathbf{X}}$ be given α-Hölder rough paths. Define the metric

$$
\rho_\alpha(\mathbf{X}, \tilde{\mathbf{X}}) \triangleq \sum_{n=1}^{N_\alpha} \|X^n - \tilde{X}^n\|_{n\alpha} \triangleq \sum_{n=1}^{N_\alpha} \sup_{0\leqslant s<t\leqslant T} \frac{|X^n_{s,t} - \tilde{X}^n_{s,t}|}{(t-s)^{n\alpha}}. \tag{1.14}
$$

We also denote

$$
\|\mathbf{X}\|_\alpha \triangleq \rho_\alpha(\mathbf{X}, \mathbf{1}) = \sum_{n=1}^{N_\alpha} \|X^n\|_{n\alpha}.
$$

Through approximation by regular paths under the metric ρ_α, one can construct a natural and wide class of rough paths.

Definition 1.13. An α-Hölder rough path $\mathbf{X} : \Delta_T \to V$ is *geometric* if there exists a sequence of continuous paths $x^{(m)} : [0,T] \to V$ with bounded total variation, such that

$$\rho(\mathbf{X}^{(m)}, \mathbf{X}) \to 0 \quad \text{as } m \to \infty,$$

where $\mathbf{X}^{(m)} : \Delta_T \to T_1^{N_\alpha}(V)$ denotes the canonical lifting of $x^{(m)}$ through iterated integrals as defined in Example 1.11.

Remark 1.14. The class of geometric rough paths is sufficient for many applications in stochastic analysis. For instance, under suitable conditions, semimartingales, Markov processes, Gaussian processes can be lifted as (random) geometric rough paths in a canonical way. However, it may not be rich enough for some applications in stochastic partial differential equations (cf. Gubinelli [Gub10]).

The shuffle product formula

In the regime of $\alpha \in (1/3, 1/2]$ ($N_\alpha = 2$), Definition 1.9 is sufficient to yield a complete theory of rough differential equations (without $\mathbf{X}$ being geometric!). However, when $\alpha \leqslant 1/3$ the geometric property becomes relevant. Indeed, it is the following algebraic property of geometric rough paths that plays an essential role in this regularity regime.

Lemma 1.15. *Let $\mathbf{X} : \Delta_T \to T_1^{N_\alpha}(V)$ be a geometric rough path. Then*

$$X_{s,t}^m \otimes X_{s,t}^n = \sum_{\sigma \in \mathcal{S}(m,n)} P_\sigma(X_{s,t}^{m+n}) \quad \text{for all } m, n = 1, \cdots, N_\alpha. \tag{1.15}$$

Here $\mathcal{S}(m,n)$ denotes the set of (m,n)-shuffles, i.e. permutations σ of order $m+n$ that satisfy

$$\sigma(1) < \cdots < \sigma(m), \ \sigma(m+1) < \cdots < \sigma(m+n),$$

and $P_\sigma : V^{\otimes(m+n)} \to V^{\otimes(m+n)}$ is the tensor permutation operator induced by (1.7).

Proof. It is enough to consider the case when $\mathbf{X}$ is the canonical lifting of a continuous path x with bounded total variation. In this case, one has

$$
\begin{aligned}
X_{s,t}^{m} \otimes X_{s,t}^{n} &= \int_{\substack{s<t_1<\cdots<t_m<t \\ s<t_{m+1}<\cdots<t_{m+n}<t}} dx_{t_1} \otimes \cdots \otimes dx_{t_m} \otimes dx_{t_{m+1}} \otimes \cdots \otimes dx_{t_{m+n}} \\
&= \sum_{\sigma \in \mathcal{S}(m,n)} \int_{s<t_{\sigma^{-1}(1)}<\cdots<t_{\sigma^{-1}(m+n)}<t} dx_{t_1} \otimes \cdots \otimes dx_{t_{m+n}} \\
&= \sum_{\sigma \in \mathcal{S}(m,n)} \int_{s<t_1<\cdots<t_{m+n}<t} dx_{t_{\sigma(1)}} \otimes \cdots \otimes dx_{t_{\sigma(m+n)}} \\
&= \sum_{\sigma \in \mathcal{S}(m,n)} P_{\sigma}\big(X_{s,t}^{m+n}\big).
\end{aligned}
$$

□

Remark 1.16. If $V = \mathbb{R}^d$ with basis $\{e_1, \cdots, e_d\}$, under the canonical tensor basis (cf. Example 1.5) the formula (1.15) reads

$$
X^{m;i_1\cdots i_m} X^{n;j_1\cdots j_n} = \sum_{\sigma \in \mathcal{S}(m,n)} X^{m+n;k_{\sigma^{-1}(1)}\cdots k_{\sigma^{-1}(m+n)}},
$$

where $X^{m;i_1\cdots i_m}$ denotes the coordinate of X^m with respect to the basis element $e_{i_1} \otimes \cdots \otimes e_{i_m}$, and $(k_1, \cdots, k_{m+n}) \triangleq (i_1, \cdots, i_m, j_1, \cdots, j_n)$.

Equation (1.15) is known as the *shuffle product formula*. It is such an algebraic property, rather than being geometric, that is needed in the study of rough differential equations (when $\alpha \leqslant 1/3$). We therefore separate this out to introduce the following definitions.

Definition 1.17. Let $N \geqslant 1$. The *N-th order free nilpotent group* over V is the subgroup of $T_1^N(V)$ defined by

$$
G^N(V) \triangleq \big\{\xi = (1, \xi_1, \cdots, \xi_N) \in T_1^N(V) : \xi_m \otimes \xi_n = \sum_{\sigma \in \mathcal{S}(m,n)} P_{\sigma}(\xi_{m+n}) \ \forall m, n\big\}.
$$

Definition 1.18. A α-Hölder rough path is said to be *weakly geometric* if it takes values in the group $G^N(V)$.

Remark 1.19. It is not obvious at all that $G^N(V)$ is a group. This can be seen by a theorem of Chen [Che57] which asserts that $G^N(V)$ is the exponential of Lie polynomials.

Remark 1.20. When $\dim V < \infty$, there is no essential difference between geometric and weakly geometric rough paths. Indeed, a geometric rough path is weakly geometric as seen by Lemma 1.15. Conversely, in finite dimensions it is known that every α-Hölder weakly geometric rough path is β-Hölder geometric for all $\beta < \alpha$ (cf. Friz-Victoir [FV10]).

Remark 1.21. The shuffle product formula asserts that the product of two iterated integrals (over the same region) can be expressed as a linear combination of higher order iterated integrals. This property suggests that the structure of functions on $G^N(V)$ is particularly simple: polynomial functions on $G^N(V)$ are always linear. Such a linearisation property has far-reaching implications (cf. Section 4.3 for one aspect of applications).

Exercise 1.22. Let

$$T_0^N(V) \triangleq \{\xi = (\xi_0, \cdots, \xi_N) : \xi_0 = 0\}.$$

We define two functions

$$\exp : T_0^N(V) \to T_1^N(V), \ \log : T_1^N(V) \to T_0^N(V)$$

by

$$\exp(\xi) \triangleq \sum_{n=0}^{N} \frac{\xi^{\otimes n}}{n!}, \ \log(\xi) \triangleq \sum_{n=1}^{N} (-1)^{n-1} \frac{(\xi - \mathbf{1})^{\otimes n}}{n}$$

respectively.

(i) Show that exp and log are inverse to each other.
(ii) Compute the dimension of $G^N(\mathbb{R}^2)$ and identify a basis of $\log(G^N(\mathbb{R}^2))$ in the cases of $N = 2, 3, 4$.
(iii) Let $x_t \in \mathbb{R}^2$ be a two-dimensional smooth path and define its canonical lifting $\mathbf{X}_{s,t} = (1, X_{s,t}^1, X_{s,t}^2)$ as in Example 1.11 with $N = 2$. Give a geometric interpretation of the second level component of $\log \mathbf{X}_{s,t}$.
(iv) Let P be an element in $\log G^N(V)$. Show that

$$\mathbf{X}_t \triangleq \exp(tP) \in G^N(V)$$

is $1/N$-Hölder continuous (and thus weakly geometric) in the sense of (1.12).

1.2.3 Controlled rough paths

The previous section defines the structure for the driving path $\mathbf{X}$ in the formal integral $\int \mathcal{Y} d\mathbf{X}$ or differential equation $d\mathcal{Y} = F(\mathcal{Y})d\mathbf{X}$. In this section, we introduce the structure for the integrand or solution path $\mathcal{Y}$.

Inspired by the Riemann sum approximation (1.5), in order to construct the rough integral $\int \mathcal{Y} d\mathbf{X}$ one should also view $\mathcal{Y}$ as a multi-level object consisting of a (zeroth level) usual path Y^0 together with a collection of "derivative paths" $Y^1, \cdots, Y^{N-1}$ up to a certain order. These paths should be subject to suitable remainder regularity conditions arising from Taylor type expansions.

We now give the precise definition. Let U, V be Banach spaces. Let $\mathbf{X}$ be an α-Hölder rough path over V with given fixed $\alpha \in (0,1]$. We denote $\mathcal{L}(V^{\otimes i}; U)$ as the space of continuous linear operators from $V^{\otimes i}$ to U and also set $N \triangleq \lfloor 1/\alpha \rfloor$.

Definition 1.23. Let $\mathcal{Y}_t = (Y_t^0, Y_t^1, \cdots, Y_t^{N-1})$ $(0 \leqslant t \leqslant T)$ be a collection of continuous paths where $Y_t^0 \in U$ and $Y_t^i \in \mathcal{L}(V^{\otimes i}; U)$ for $1 \leqslant i \leqslant N-1$. We say that $\mathcal{Y}$ is an α-Hölder rough path *controlled by* $\mathbf{X}$, if the "remainders" defined by

$$\mathcal{R}\mathcal{Y}_{s,t}^i \triangleq \begin{cases} Y_t^i - Y_s^i - \sum_{j=1}^{N-1-i} Y_s^{i+j} X_{s,t}^j, & i = 0, \cdots, N-2; \\ Y_t^{N-1} - Y_s^{N-1}, & i = N-1 \end{cases} \tag{1.16}$$

satisfy the following regularity condition:

$$\|\mathcal{R}\mathcal{Y}^i\|_{(N-i)\alpha} \triangleq \sup_{0 \leqslant s < t \leqslant T} \frac{|\mathcal{R}\mathcal{Y}_{s,t}^i|}{|t-s|^{(N-i)\alpha}} < \infty \quad \text{for each } i = 0, \cdots, N-1. \tag{1.17}$$

Remark 1.24. Heuristically, Y^i is the i-th "derivative" of Y^0 with respect to $\mathbf{X}$. As a result, (1.16) is like a Taylor type expansion with respect to $\mathbf{X}$. The regularity requirement (1.17) becomes natural from this perspective. Note that Y^i is given along with Y^0 in the definition of $\mathcal{Y}$. It is though possible to show that the derivative paths Y^i $(i \geqslant 1)$ are unique if $\mathbf{X}$ is "truly α-Hölder rough" in a certain sense (cf. Exercise 1.25 below).

Let $\mathcal{D}_{\mathbf{X};\alpha}(U)$ denote the space of U-valued α-Hölder rough paths controlled by $\mathbf{X}$. We define a semi-norm $\|\cdot\|_{\mathbf{X};\alpha}$ on $\mathcal{D}_{\mathbf{X};\alpha}(U)$ by

$$\|\mathcal{Y}\|_{\mathbf{X};\alpha} \triangleq \sum_{i=0}^{N-1} \|\mathcal{R}\mathcal{Y}^i\|_{(N-i)\alpha}$$

and a norm $|||\cdot|||_{\mathbf{X};\alpha}$ by

$$|||\mathcal{Y}|||_{\mathbf{X};\alpha} \triangleq \|\mathcal{Y}\|_{\mathbf{X};\alpha} + \sum_{i=0}^{N-1} |Y_0^i|.$$

As a consequence of Lemma 1.26 below, it is easily seen that $(\mathcal{D}_{\mathbf{X};\alpha}(U), |||\cdot|||_{\mathbf{X};\alpha})$ is a Banach space. Note that this space depends on $\mathbf{X}$.

For the purpose of continuity estimates, it is also important to measure the distance between different controlled rough paths with respect to different $\mathbf{X}$'s. Let $\mathbf{X}, \tilde{\mathbf{X}}$ be α-Hölder rough paths and let $\mathcal{Y}, \tilde{\mathcal{Y}}$ be controlled by $\mathbf{X}, \tilde{\mathbf{X}}$ respectively. We define the functional

$$d_{\mathbf{X},\tilde{\mathbf{X}};\alpha}(\mathcal{Y}, \tilde{\mathcal{Y}}) \triangleq \sum_{i=0}^{N-1} \|\mathcal{R}\mathcal{Y}^i - \mathcal{R}\tilde{\mathcal{Y}}^i\|_{(N-i)\alpha}.$$

Note that this is not a distance function in the usual sense as $\mathcal{Y}, \tilde{\mathcal{Y}}$ live in different spaces.

Exercise 1.25. (i) Suppose that there exists a dense subset $\mathcal{S} \subseteq [0, T]$ such that

$$\overline{\lim_{t\to s^+}} \frac{|X^i_{s,t}|}{|t-s|^{(i+1)\alpha}} = +\infty \quad \forall s \in \mathcal{S},\ i = 1, \cdots, N-1.$$

Let $\mathcal{Y} = (Y^0, \cdots, Y^{N-1})$ and $\tilde{\mathcal{Y}} = (\tilde{Y}^0, \cdots, \tilde{Y}^{N-1})$ be two paths in $\mathcal{D}_{\mathbf{X};\alpha}(U)$ such that $Y^0 = \tilde{Y}^0$. Show that $Y^i = \tilde{Y}^i$ for all $i \geqslant 1$.
(ii) Find an example of $\mathcal{Y}, \tilde{\mathcal{Y}} \in \mathcal{D}_{\mathbf{X};\alpha}(U)$ such that $Y^0 = \tilde{Y}^0$ but $Y^i \neq \tilde{Y}^i$ for $i = 1, \cdots, N-1$.

Hölder estimates for controlled rough paths

The following lemma tells us how to estimate the α-Hölder norm of Y^i in terms of the control norm. This estimate will be useful for later purposes.

Lemma 1.26. *(i) Let $\mathcal{Y} \in \mathcal{D}_{\mathbf{X};\alpha}(U)$. Then for each i, the path Y^i is α-Hölder continuous and one has*

$$\|Y^{N-i}\|_\alpha \leqslant M(T, \|\mathbf{X}\|_\alpha, \max_{1\leqslant j\leqslant i-1} |Y_0^{N-j}|, \max_{1\leqslant j\leqslant i} \|\mathcal{R}\mathcal{Y}^{N-j}\|_\alpha). \tag{1.18}$$

(ii) Let $\mathbf{X}, \tilde{\mathbf{X}}$ be α-Hölder rough paths and let $\mathcal{Y}, \tilde{\mathcal{Y}}$ be controlled by $\mathbf{X}, \tilde{\mathbf{X}}$ respectively. Denote

$$\delta X^i \triangleq X^i - \tilde{X}^i,\ \delta Y^i \triangleq Y^i - \tilde{Y}^i,\ \delta\mathcal{R}^i \triangleq \mathcal{R}\mathcal{Y}^i - \mathcal{R}\tilde{\mathcal{Y}}^i.$$

Then for each $2 \leqslant i \leqslant N$, one has

$$\|\delta Y^{N-i}\|_\alpha \leqslant M\big(T, \|\mathbf{X}\|_\alpha, \|\tilde{\mathbf{X}}\|_\alpha, \max_{1\leqslant j\leqslant i-1} |Y_0^{N-j}|, \max_{1\leqslant j\leqslant i-1} \|\mathcal{R}\mathcal{Y}^{N-j}\|_{j\alpha}\big)$$
$$\big(\rho_\alpha(\mathbf{X}, \tilde{\mathbf{X}}) + \sum_{j=1}^{i} \|\delta\mathcal{R}^{N-j}\|_{j\alpha} + \sum_{j=1}^{i-1} |\delta Y_0^{N-j}|\big). \tag{1.19}$$

In both parts, $M(\cdots)$ denotes a universal function (indeed, polynomial) that is increasing in each of its variables.

Proof. The proof is lengthy and tedious. It is no deeper than the use of the triangle inequality as well as the elementary inequality

$$|ab - \tilde{a}\tilde{b}| \leqslant |a - \tilde{a}| \cdot |b| + |\tilde{a}| \cdot |b - \tilde{b}|.$$

We provide the fine details here but will no longer repeat calculations of similar kind in the future.

We directly prove the continuity estimate (1.19) by induction on i. The result of Part (i) is a special case by taking $\mathbf{X} = \tilde{\mathbf{X}}$ and $\tilde{\mathcal{Y}} = 0$. When $i = 2$, one has

$$\begin{aligned} Y_{s,t}^{N-2} - \tilde{Y}_{s,t}^{N-2} &= \big(Y_s^{N-1} X_{s,t}^1 - \tilde{Y}_s^{N-1} \tilde{X}_{s,t}^1\big) + \big(\mathcal{R}\mathcal{Y}_{s,t}^{N-2} - \mathcal{R}\tilde{\mathcal{Y}}_{s,t}^{N-2}\big) \\ &= Y_s^{N-1}(X_{s,t}^1 - \tilde{X}_{s,t}^1) + (Y_s^{N-1} - \tilde{Y}_s^{N-1})\tilde{X}_{s,t}^1 + \big(\mathcal{R}\mathcal{Y}_{s,t}^{N-2} - \mathcal{R}\tilde{\mathcal{Y}}_{s,t}^{N-2}\big). \end{aligned}$$

Since $\mathcal{R}\mathcal{Y}^{N-1} = Y^{N-1}$, it follows that

$$\begin{aligned} \|\delta Y^{N-2}\|_\alpha &\leqslant (1 + T^\alpha)\,\big(|Y_0^{N-1}| + \|Y^{N-1}\|_\alpha\big)\|\delta X^1\|_\alpha \\ &\quad + (1 + T^\alpha)\,\big(\big(|\delta Y_0^{N-1}| + \|\delta Y^{N-1}\|_\alpha\big)\|\tilde{X}^1\|_\alpha + T^\alpha \|\delta\mathcal{R}^{N-2}\|_{2\alpha}\big) \\ &\leqslant (1 + T^\alpha)\,\big(T^\alpha + \|\tilde{\mathbf{X}}\|_\alpha + |Y_0^{N-1}| + \|\mathcal{R}\mathcal{Y}^{N-1}\|_\alpha\big) \\ &\quad \times \big(\|\delta X^1\|_\alpha + |\delta Y_0^{N-1}| + \|\delta\mathcal{R}^{N-1}\|_\alpha + \|\delta\mathcal{R}^{N-2}\|_{2\alpha}\big). \end{aligned}$$

Therefore, the estimate (1.19) holds in this case.

Suppose that (1.19) holds for $\delta Y^{N-1}, \cdots, \delta Y^{N-i}$. Using that

$$\delta Y_{s,t}^{N-(i+1)} = \sum_{j=1}^{i} Y_s^{N-j} X_{s,t}^{i+1-j} - \sum_{j=1}^{i} \tilde{Y}_s^{N-j} \tilde{X}_{s,t}^{i+1-j} + \delta\mathcal{R}^{N-(i+1)},$$

one has

$$\begin{aligned}\|\delta Y^{N-(i+1)}\|_\alpha &\leqslant \big(1+T^{(i+1)\alpha}\big)\Big(\sum_{j=1}^{i}\big(|Y_0^{N-j}|+\|Y^{N-j}\|_\alpha\big)\|\delta X^{i+1-j}\|_{(i+1-j)\alpha} \\ &\qquad + \sum_{j=1}^{i}\big(|\delta Y_0^{N-j}|+\|\delta Y^{N-j}\|_\alpha\big)\|\tilde{X}^{i+j}\|_{(i+j)\alpha} + \|\delta\mathcal{R}^{N-(i+1)}\|_{(i+1)\alpha}\Big) \\ &\leqslant \big(1+T^{(i+1)\alpha}\big)\big(1+\max_{1\leqslant j\leqslant i}\big(|Y_0^{N-j}|+\|Y^{N-j}\|_\alpha\big)+\|\tilde{\mathbf{X}}\|_\alpha\big) \\ &\qquad \times\Big(\rho_\alpha(\mathbf{X},\tilde{\mathbf{X}})+\sum_{j=1}^{i}\big(|\delta Y_0^{N-j}|+\|\delta Y^{N-j}\|_\alpha\big)+\|\delta\mathcal{R}^{N-(i+1)}\|_{(i+1)\alpha}\Big).\end{aligned} \tag{1.20}$$

By the induction hypothesis with $\mathbf{X}=\tilde{\mathbf{X}}$ and taking $\tilde{Y}=0$, one has

$$\|Y^{N-j}\|_\alpha \leqslant M\big(T,\|\mathbf{X}\|_\alpha,\max_{1\leqslant l\leqslant j-1}|Y_0^{N-l}|,\max_{1\leqslant l\leqslant j}\|\mathcal{R}\mathcal{Y}^{N-l}\|_\alpha\big). \tag{1.21}$$

Similarly, for each $1\leqslant j\leqslant i$ one has

$$\begin{aligned}\|\delta Y^{N-j}\|_\alpha &\leqslant M\big(T,\|\mathbf{X}\|_\alpha,\|\tilde{\mathbf{X}}\|_\alpha,\max_{1\leqslant l\leqslant j-1}\big|Y_0^{N-l}\big|,\max_{1\leqslant l\leqslant j-1}\|\mathcal{R}\mathcal{Y}^{N-l}\|_\alpha\big) \\ &\qquad \times\Big(\rho_\alpha(\mathbf{X},\tilde{\mathbf{X}})+\|\delta\mathcal{R}^{N-j}\|_{j\alpha}+\sum_{l=1}^{j-1}\big(\big|\delta Y_0^{N-l}\big|+\|\delta\mathcal{R}^{N-l}\|_{l\alpha}\big)\Big).\end{aligned} \tag{1.22}$$

The induction step follows by substituting (1.21) and (1.22) into (1.20). □

2 Rough differential equations and the universal limit theorem

The goal of this chapter is to establish existence, uniqueness and continuity of the rough differential equation (RDE)

$$d\mathcal{Y} = F(\mathcal{Y})d\mathbf{X} \tag{2.1}$$

with given initial condition Y_0. Here $\mathbf{X}$ is a weakly geometric rough path over V, $F : U \to \mathcal{L}(V;U)$ is a suitably regular function, and $\mathcal{Y}$ is a controlled rough path over U with respect to $\mathbf{X}$. The major steps towards establishing the well-posedness of (2.1) are summarised as follows.

Step 1. Show that controlled rough paths are stable under regular transformations:

$$\mathcal{Y} \text{ controlled by } \mathbf{X},\ F \text{ suitably regular } \implies F(\mathcal{Y}) \text{ controlled by } \mathbf{X}.$$

Step 2. Define the notion of rough integration $\int \mathcal{Z} d\mathbf{X}$ where $\mathbf{X}$ is a rough path and $\mathcal{Z}$ is controlled by $\mathbf{X}$.
Step 3. Interpret the RDE (2.1) in its integral form and construct solutions based on fixed point arguments.

In the following sections, we develop these steps mathematically. For the sake of simplicity, we will only work under the regularity regime of $\alpha \in (1/3, 1/2]$ ($N = 2$). In most places, the extension to the general case is routine and only requires more technicalities. However, for the stability property in Step 1 the extension is not trivial at all. We will give a brief discussion on this point in the last part of Section 2.1. Summarised concisely, the moral is that

$$\left.\begin{array}{r}\alpha > 1/3,\ \mathbf{X} : \text{arbitrary rough path} \\ \alpha \leqslant 1/3,\ \mathbf{X} : \text{weakly geometric}\end{array}\right\} \implies \text{Steps 1 \& 3;}$$
$$\alpha \in (0,1],\ \mathbf{X} : \text{arbitrary rough path} \implies \text{Step 2.}$$

Notation. In the rest of this chapter, we will always use $M(\cdots)$ to denote a universal function that is continuous and increasing in all variables. The shape of M may differ from line to line and is of no importance in the discussion.

2.1 Stability of controlled rough paths

As the first step towards solving RDEs, in this section we show that controlled rough paths are stable under regular transformations. Let $\mathbf{X}$ be an α-Hölder rough path over V and let $\mathcal{Y} \in \mathcal{D}_{\mathbf{X};\alpha}(U)$, where $\alpha \in (1/3, 1/2]$ is given fixed. Let $F : U \to W$ be a C_b^2-function (twice continuously differentiable with bounded derivatives).

We want to define $\mathcal{Z} = F(\mathcal{Y})$ as a W-valued rough path controlled by $\mathbf{X}$. By definition, it must come with a zeroth level path Z^0 and a derivative path Z^1. The definition of Z^0 is obvious: one simply sets $Z_t^0 \triangleq F(Y_t^0)$. To motivate the construction of Z^1, we recall that the remainder

$$\mathcal{R}\mathcal{Z}_{s,t} \triangleq Z_{s,t}^0 - Z_s^1 X_{s,t}^1$$

is required to have regularity $|t-s|^{2\alpha}$. By the second order Taylor expansion of F (with integral form remainder), one has

$$F(Y_t^0) = F(Y_s^0) + DF(Y_s^0)(Y_{s,t}^0) + \int_0^1 (1-\theta) D^2 F(Y_s^0 + \theta Y_{s,t}^0)(Y_{s,t}^0)^{\otimes 2} d\theta, \quad (2.2)$$

where

$$DF : U \to \mathcal{L}(U;W),\ DF(x)(y) \triangleq \frac{d}{d\varepsilon}\Big|_{\varepsilon=0} F(x+\varepsilon y)$$

and

$$D^2F : U \to \mathcal{L}(U \otimes U; W),\ D^2F(x)(y \otimes z) \triangleq \frac{\partial^2}{\partial\varepsilon\partial\eta}\Big|_{(\varepsilon,\eta)=(0,0)} F(x+\varepsilon y+\eta z) \quad (2.3)$$

are the derivatives of F.

Note that the last term on the right hand side of (2.2) has regularity $|t-s|^{2\alpha}$ due to the $(Y_{s,t}^0)^{\otimes 2}$ term. Since $\mathcal{Y} \in \mathcal{D}_{\mathbf{X};\alpha}(U)$, by writing $Y_{s,t}^0 = Y_s^1 X_{s,t}^1 + \mathcal{R}\mathcal{Y}_{s,t}$ it is clear from (2.2) that

$$F(Y_t^0) = F(Y_s^0) + DF(Y_s^0) Y_s^1 X_{s,t}^1 + \text{“} * \text{”},$$

where $*$ denotes an expression that has regularity $|t-s|^{2\alpha}$. In view of Definition 1.23, the derivative path of $F(\mathcal{Y})$ should be defined as

$$Z_t^1 \triangleq DF(Y_t^0) Y_t^1. \quad (2.4)$$

Now checking all required conditions for being a controlled rough path is routine.

Theorem 2.1. *Let $F \in C_b^2$ and $\mathcal{Y} \in \mathcal{D}_{\mathbf{X};\alpha}$. The path*

$$\mathcal{Z} \triangleq F(\mathcal{Y}) \triangleq (Z^0, Z^1) :\ Z_t^0 \triangleq F(Y_t^0), Z_t^1 \triangleq DF(Y_t^0) Y_t^1$$

is an α-Hölder rough path controlled by $\mathbf{X}$. In addition, the following estimate holds:

$$\|\mathcal{Z}\|_{\mathbf{X};\alpha} \leqslant \|F\|_{C_b^2} \cdot M(T, |Y_0^1|, \|\mathcal{Y}\|_{\mathbf{X};\alpha}, \|\mathbf{X}\|_\alpha). \quad (2.5)$$

Proof. From the previous discussion, it is already clear that $\mathcal{R}\mathcal{Z}_{s,t}$ has regularity $|t-s|^{2\alpha}$. It is also obvious that Z^0, Z^1 are both α-Hölder continuous. Therefore, $\mathcal{Z} \in \mathcal{D}_{\mathbf{X};\alpha}(W)$. It remains to establish the uniform estimate (2.5). According to the definition of $\mathcal{Z}$, the Taylor expansion (2.2) and the Hölder estimate for Y^0 given by (1.18), one sees that

$$\begin{aligned} |\mathcal{R}\mathcal{Z}_{s,t}| &= \big|DF(Y_s^0)\mathcal{R}\mathcal{Y}_{s,t} + \int_0^1 (1-\theta)D^2F(Y_s^0 + \theta Y_{s,t}^0)(Y_{s,t}^0)^{\otimes 2} d\theta\big| \\ &\leqslant \|F\|_{C_b^2} \cdot M(T, |Y_0^1|, \|\mathcal{Y}\|_{\mathbf{X};\alpha}, \|\mathbf{X}\|_\alpha) \cdot |t-s|^{2\alpha} \end{aligned} \tag{2.6}$$

Similarly,

$$\begin{aligned} |Z_{s,t}^1| &= \big|DF(Y_t^0)Y_t^1 - DF(Y_s^0)Y_s^1\big| \\ &= \big|\big(DF(Y_t^0) - DF(Y_s^0)\big)Y_t^1 + DF(Y_s^0)Y_{s,t}^1\big| \\ &\leqslant \|D^2F\|_\infty |Y_{s,t}^0| \cdot \big(|Y_1^0| + T^\alpha \|Y^1\|_\alpha\big) + \|DF\|_\infty |Y_{s,t}^1| \\ &\leqslant \|F\|_{C_b^2} \cdot M(T, |Y_0^1|, \|\mathcal{Y}\|_{\mathbf{X};\alpha}, \|\mathbf{X}\|_\alpha) \cdot |t-s|^\alpha. \end{aligned}$$

The desired estimate (2.5) thus follows. □

Remark 2.2. The main useful information from (2.5) is that the right hand side does not depend on Y_0^0. This point is important in the construction of global RDE solutions.

A simple adaptation of the above calculation yields the following continuity estimate.

Proposition 2.3. *Suppose that $F \in C_b^3$. Let $\mathbf{X}, \tilde{\mathbf{X}}$ be α-Hölder rough paths and let $\mathcal{Y}, \tilde{\mathcal{Y}}$ be controlled by $\mathbf{X}, \tilde{\mathbf{X}}$ respectively. Then*

$$\begin{aligned} d_{\mathbf{X},\tilde{\mathbf{X}};\alpha}(F(\mathcal{Y}), F(\tilde{\mathcal{Y}})) \leqslant{}& \|F\|_{C_b^3} M(T, |Y_0^1|, |\tilde{Y}_0^1|, \|\mathcal{Y}\|_{\mathbf{X};\alpha}, \|\tilde{\mathcal{Y}}\|_{\tilde{\mathbf{X}};\alpha}, \|\mathbf{X}\|_\alpha, \|\tilde{\mathbf{X}}\|_\alpha) \\ &\times \big(\rho_\alpha(\mathbf{X}, \tilde{\mathbf{X}}) + d_{\mathbf{X},\tilde{\mathbf{X}};\alpha}(\mathcal{Y}, \tilde{\mathcal{Y}}) + |Y_0^0 - \tilde{Y}_0^0| + |Y_0^1 - \tilde{Y}_0^1|\big). \end{aligned} \tag{2.7}$$

Proof. The calculation is lengthy but routine: one uses the second order Taylor expansion to represent $\mathcal{R}\mathcal{Z}_{s,t}$ as in (2.6) and keeps using the elementary inequality

$$|ab - \tilde{a}\tilde{b}| \leqslant |a - \tilde{a}| \cdot |b| + |\tilde{a}| \cdot |b - \tilde{b}|$$

when making the comparison between $\mathcal{Z}$ and $\tilde{\mathcal{Z}}$. Since there is no essential difference from the calculation performed in the proof of Lemma 1.26, we do not provide the tedious details. □

The case when $\alpha \leqslant 1/3$

In the regime of $\alpha \in (1/3, 1/2]$, one does not need to assume that $\mathbf{X}$ is weakly geometric (the above argument makes no use of such an assumption). We must point out that, however, this algebraic feature becomes crucial when $\alpha \leqslant 1/3$. We illustrate this point when $\alpha \in (1/4, 1/3]$, in which case the calculation is simple and explicit. The general case of $\alpha \in (0, 1/2]$ requires deeper algebraic considerations and we refer the reader to Boedihardjo-Geng [BG20] for a detailed discussion.

Let $\mathbf{X}$ be an α-Hölder weakly geometric rough path with $\alpha \in (1/4, 1/3]$. Let $F \in C_b^3$ and $\mathcal{Y} = (Y^0, Y^1, Y^2) \in \mathcal{D}_{\mathbf{X};\alpha}(U)$. The construction of $\mathcal{Z} = F(\mathcal{Y}) = (Z^0, Z^1, Z^2)$ is similar to the previous case. One first uses a third order Taylor expansion of F to see that

$$Z^0_{s,t} = DF(Y^0_s)Y^0_{s,t} + \frac{1}{2}D^2F(Y^0_s)(Y^0_{s,t})^{\otimes 2} + \text{``}*\text{''},$$

where $*$ records an expression of regularity $|t-s|^{3\alpha}$. Next, by using the relation

$$Y^0_{s,t} = Y^1_s X^1_{s,t} + Y^2_s X^2_{s,t} + \mathcal{R}\mathcal{Y}^0_{s,t},$$

one can further write

$$\begin{aligned}
Z^0_{s,t} &= DF(Y^0_s)Y^1_s X^1_{s,t} + DF(Y^0_s)Y^2_s X^2_{s,t} + \frac{1}{2}D^2F(Y^0_s)(Y^1_s X^1_{s,t})^{\otimes 2} + \text{``}*\text{''} \\
&= DF(Y^0_s)Y^1_s X^1_{s,t} + DF(Y^0_s)Y^2_s X^2_{s,t} + \frac{1}{2}D^2F(Y^0_s)(Y^1_s \otimes Y^1_s)(X^1_{s,t} \otimes X^1_{s,t}) + \text{``}*\text{''}
\end{aligned}$$

Here $Y^1_s \otimes Y^1_s \in \mathcal{L}(V^{\otimes 2}; U^{\otimes 2})$ is the unique linear operator induced by

$$(Y^1_s \otimes Y^1_s)(v_1 \otimes v_2) \triangleq (Y^1_s v_1) \otimes (Y^1_s v_2), \quad v_1, v_2 \in V.$$

Here comes the key algebraic ingredient. Since $\mathbf{X}$ is weakly geometric, according to the shuffle product formula (1.15), one has

$$X^1_{s,t} \otimes X^1_{s,t} = (\mathrm{id} + P)(X^2_{s,t}),$$

where $P : V^{\otimes 2} \to V^{\otimes 2}$ is the permutation operator induced by

$$P(v_1 \otimes v_2) = v_2 \otimes v_1, \quad v_1, v_2 \in V.$$

It follows that

$$Z^0_{s,t} = DF(Y^0_s)Y^1_s X^1_{s,t} + \big(DF(Y^0_s)Y^2_s + \frac{1}{2}D^2F(Y^0_s)(Y^1_s \otimes Y^1_s)(\mathrm{id} + P)\big)X^2_{s,t} + \text{``}*\text{''}. \tag{2.8}$$

This identity forces us to define

$$Z_t^1 \triangleq DF(Y_t^0)Y_t^1$$

and

$$Z_t^2 \triangleq DF(Y_s^0)Y_s^2 + \frac{1}{2}D^2F(Y_s^0)(Y_s^1 \otimes Y_s^1)(\mathrm{id}+P).$$

In this way, it is already ensured by (2.8) that the zeroth level remainder

$$\mathcal{R}\mathcal{Z}_{s,t}^0 \triangleq Z_{s,t}^0 - Z_s^1 X_{s,t}^1 - Z_s^2 X_{s,t}^2$$

has regularity $|t-s|^{3\alpha}$. In addition, it is clear that second level remainder $\mathcal{R}\mathcal{Z}_{s,t}^2 \triangleq Z_{s,t}^2$ has regularity $|t-s|^\alpha$ since Z^2 is an α-Hölder continuous path. What is less obvious is the 2α-regularity of

$$\mathcal{R}\mathcal{Z}_{s,t}^1 \triangleq Z_{s,t}^1 - Z_s^2 X_{s,t}^1. \tag{2.9}$$

We prove this in the following lemma.

Lemma 2.4. *The first level remainder satisfies*

$$\sup_{0\leqslant s<t\leqslant T} \frac{|\mathcal{R}\mathcal{Z}_{s,t}^1|}{|t-s|^{2\alpha}} < \infty.$$

Proof. Since (2.9) takes values in $\mathcal{L}(V;W)$, in order to see things better we feed the equation an arbitrary test vector $v \in V$. By the definition of Z^1, one has

$$\begin{aligned} Z_{s,t}^1 v &= DF(Y_t^0)Y_t^1 v - DF(Y_s^0)Y_s^1 v \\ &= \big(DF(Y_t^0) - DF(Y_s^0)\big)Y_t^1 v + DF(Y_s^0)Y_{s,t}^1 v \\ &= D^2F(Y_s^0)(Y_{s,t}^0 \otimes Y_t^1 v) + DF(Y_s^0)Y_{s,t}^1 v + \text{``}*\text{''} \\ &= D^2F(Y_s^0)\big((Y_s^1 X_{s,t}^1) \otimes Y_s^1 v\big) + DF(Y_s^0)\big(Y_s^2(X_{s,t}^1 \otimes v)\big) + \text{``}*\text{''} \end{aligned}$$

where $*$ denotes an expression of regularity $|t-s|^{2\alpha}$ which may differ from line to line. On the other hand, by the definition of Z^2, one has

$$\begin{aligned} Z_s^2(X_{s,t}^1 \otimes v) &= DF(Y_s^0)Y_s^2(X_{s,t}^1 \otimes v) + \frac{1}{2}D^2F(Y_s^0)(Y_s^1 \otimes Y_s^1)(\mathrm{id}+P)(X_{s,t}^1 \otimes v) \\ &= DF(Y_s^0)Y_s^2(X_{s,t}^1 \otimes v) + \frac{1}{2}D^2F(Y_s^0)\big((Y_s^1 X_{s,t}^1) \otimes Y_s^1 v + Y_s^1 v \otimes Y_s^1 X_{s,t}^1\big). \end{aligned}$$

Note that D^2F is a symmetric operator, i.e.

$$D^2F(x)(y \otimes z) = D^2F(x)(z \otimes y)$$

which is clear from its definition (2.3). Therefore,

$$D^2F(Y_s^0)\big((Y_s^1X_{s,t}^1)\otimes Y_s^1v\big)=\frac{1}{2}D^2F(Y_s^0)\big((Y_s^1X_{s,t}^1)\otimes Y_s^1v+Y_s^1v\otimes Y_s^1X_{s,t}^1\big)$$

and thus

$$Z_{s,t}^1v-Z_s^2(X_{s,t}^1\otimes v)=\text{“}*\text{”},$$

which has the desired regularity $|t-s|^{2\alpha}$. □

As a consequence, one concludes that $\mathcal{Z}\in\mathcal{D}_{\mathbf{X};\alpha}(W)$. The extensions of Theorem 2.1 and Proposition 2.3 to this case are now straight forward. For the general case of $\alpha\in(0,1/2]$, the construction of $\mathcal{Z}=(Z^0,Z^1,\cdots,Z^{N-1})$ ($N=\lfloor\alpha\rfloor$) is done in a similar way as before based on the Taylor expansion of F and the shuffle product formula for $\mathbf{X}$. The main challenge is to develop the algebraic formalism that allows one to show that $\mathcal{R}\mathcal{Z}_{s,t}^i$ has regularity $|t-s|^{(N-i)\alpha}$ for each i. This point has an algebraic nature that does not follow from standard Hölder estimates and requires deeper tools from free Lie algebras.

2.2 Rough integration

Let V,U be given Banach spaces. Let $\mathbf{X}=(X^1,X^2)$ be an α-Hölder rough path over V and let $\mathcal{Z}=(Z^0,Z^1)$ be an $\mathcal{L}(V;U)$-valued path controlled by $\mathbf{X}$. Here the underlying roughness is again fixed to be $\alpha\in(1/3,1/2]$. Note that

$$Z_t^0\in\mathcal{L}(V;U),\ Z_t^1\in\mathcal{L}(V;\mathcal{L}(V;U))\cong\mathcal{L}(V\otimes V;U),$$

where the latter identification is induced by

$$f\mapsto\tilde{f}:\ \tilde{f}(v_1\otimes v_2)=f(v_1)(v_2),\ f\in\mathcal{L}(V;\mathcal{L}(V;U)).$$

All paths are assumed to be defined on $[0,T]$. The aim of this section is to construct the rough integral

$$\int_s^t ZdX\in U.$$

As we will see, this integral can be realised as a U-valued path controlled by $\mathbf{X}$. In addition, one can establish a continuity estimate which will be needed in the study of differential equations in the next section.

A natural idea of defining the integral $\int_s^t ZdX$ is to write down a suitable Riemann sum approximation and look for convergence. As we pointed out in Section 1.1, under the current regularity regime ($\alpha\in(1/3,1/2]$) the crucial point is to include an extra second order term in the approximation. To be precise, given any finite partition

$$\mathcal{P}: s = t_0 < t_1 < \cdots < t_{n-1} < t_n = t$$

of $[s,t]$, one defines the enhanced Riemann sum approximation

$$\int_{\mathcal{P}} ZdX \triangleq \sum_{t_i \in \mathcal{P}} \big(Z^0_{t_{i-1}} X^1_{t_{i-1},t_i} + Z^1_{t_{i-1}} X^2_{t_{i-1},t_i}\big).$$

The following theorem ensures the convergence of $\int_{\mathcal{P}} ZdX$ as $|\mathcal{P}| \to 0$, yielding the precise definition of the rough integral $\int_s^t ZdX$.

Theorem 2.5 (Construction of the rough integral)**.** *For each pair of $s < t$, the limit*

$$\int_s^t ZdX \triangleq \lim_{|\mathcal{P}| \to 0} \int_{\mathcal{P}} ZdX$$

exists in U. In addition, if one defines the integral path $\mathcal{I} = (I^0, I^1)$ by

$$I^0_t \triangleq \int_0^t ZdX, \ \ I^1_t \triangleq Z^0_t, \quad 0 \leqslant t \leqslant T, \tag{2.10}$$

then $\mathcal{I}$ is a U-valued α-Hölder rough path controlled by $\mathbf{X}$ and the following estimate holds:

$$\|\mathcal{I}\|_{\mathbf{X};\alpha} \leqslant C_\alpha \big(T^\alpha (1 + \|\mathbf{X}\|_\alpha) \|\mathcal{Z}\|_{\mathbf{X};\alpha} + |Z^1_0| \cdot \|\mathbf{X}\|_\alpha\big), \tag{2.11}$$

where C_α is a universal constant depending only on α.

Proof. The proof we give here relies on an elegant idea of "point removal", which was used explicitly by Lyons [Lyo98] and essentially went back to Young [You36] in the first place.

Let

$$\mathcal{P}: s = t_0 < t_1 < \cdots < t_{n-1} < t_n = t$$

be a finite partition of $[s,t]$. Given arbitrary $t_j \in \mathcal{P}$, denote $\mathcal{P} \backslash \{t_j\}$ as the partition obtained by removing the point t_j from $\mathcal{P}$. From the definition, one has

$$\begin{aligned}\int_{\mathcal{P}} ZdX - \int_{\mathcal{P}\backslash\{t_j\}} ZdX = \big(&Z^0_{t_{j-1}} X^1_{t_{j-1},t_j} + Z^1_{t_{j-1}} X^2_{t_{j-1},t_j} + Z^0_{t_j} X^1_{t_j,t_{j+1}} + Z^1_{t_j} X^2_{t_j,t_{j+1}}\big) \\ &- \big(Z^0_{t_{j-1}} X^1_{t_{j-1},t_{j+1}} + Z^1_{t_{j-1}} X^2_{t_{j-1},t_{j+1}}\big).\end{aligned} \tag{2.12}$$

By using Chen's identity

$$\begin{cases} X^1_{t_{j-1},t_{j+1}} = X^1_{t_{j-1},t_j} + X^1_{t_j,t_{j+1}}, \\ X^2_{t_{j-1},t_{j+1}} = X^2_{t_{j-1},t_j} + X^2_{t_j,t_{j+1}} + X^1_{t_{j-1},t_j} \otimes X^1_{t_j,t_{j+1}} \end{cases}$$

and the definition of $\mathcal{R}\mathcal{Z}$, the expression (2.12) simplifies to

$$\int_{\mathcal{P}} ZdX - \int_{\mathcal{P}\backslash\{t_j\}} ZdX = \mathcal{R}\mathcal{Z}_{t_{j-1},t_j} X^1_{t_j,t_{j+1}} + Z^1_{t_{j-1},t_j} X^2_{t_j,t_{j+1}}.$$

As a result, one has the following estimate:

$$\begin{aligned} \Big|\int_{\mathcal{P}} ZdX - \int_{\mathcal{P}\backslash\{t_j\}} ZdX\Big| &\leqslant \|\mathcal{R}\mathcal{Z}\|_{2\alpha}\|X^1\|_\alpha |t_j - t_{j-1}|^{2\alpha}|t_{j+1}-t_j|^\alpha \\ &\quad + \|Z^1\|_\alpha \|X^2\|_{2\alpha}|t_j - t_{j-1}|^\alpha |t_{j+1}-t_j|^{2\alpha} \\ &\leqslant 2\|\mathcal{Z}\|_{\mathbf{X};\alpha}\|\mathbf{X}\|_\alpha |t_{j+1}-t_{j-1}|^{3\alpha}. \end{aligned} \tag{2.13}$$

Since

$$\sum_{j=1}^{n-1}(t_{j+1}-t_{j-1}) = (t_n - t_1) + (t_{n-1}-t_0) \leqslant 2(t-s),$$

there must exist some j such that

$$t_{j+1} - t_{j-1} \leqslant \frac{2(t-s)}{n-1}.$$

We choose this particular t_j in the estimate (2.13). It follows that

$$\Big|\int_{\mathcal{P}} ZdX - \int_{\mathcal{P}\backslash\{t_j\}} ZdX\Big| \leqslant 2\|\mathcal{Z}\|_{\mathbf{X};\alpha}\|\mathbf{X}\|_\alpha \cdot \big(\frac{2(t-s)}{n-1}\big)^{3\alpha}.$$

Now the new partition $\mathcal{P}\backslash\{t_j\}$ has one point less than the original one and we continue this procedure until all points in $\mathcal{P}$ (except for the endpoints s, t) are removed. A simple application of the triangle inequality yields that

$$\begin{aligned} \Big|\int_{\mathcal{P}} ZdX - \int_{\{s,t\}} ZdX\Big| &\leqslant 2\|\mathcal{Z}\|_{\mathbf{X};\alpha}\|\mathbf{X}\|_\alpha \big(2(t-s)\big)^{3\alpha} \\ &\quad \times \big(\frac{1}{(n-1)^{3\alpha}} + \frac{1}{(n-2)^{3\alpha}} + \cdots + 1\big) \\ &\leqslant C_\alpha \|\mathcal{Z}\|_{\mathbf{X};\alpha}\|\mathbf{X}\|_\alpha |t-s|^{3\alpha}, \end{aligned} \tag{2.14}$$

where

$$C_\alpha \triangleq 2^{3\alpha+1}\cdot\sum_{n=1}^{\infty}\frac{1}{n^{3\alpha}}<\infty. \quad (\text{recall that } 3\alpha>1)$$

The inequality (2.14) is uniform with respect to all $s<t$ and all partitions of $[s,t]$.

To prove the convergence of Riemann sum approximation, let $\mathcal{P},\tilde{\mathcal{P}}$ be two given partitions of $[s,t]$. Let $\mathcal{P}'$ be the partition formed by all points in $\mathcal{P}$ and $\tilde{\mathcal{P}}$. For each sub-interval $[s_l,s_{l+1}]$ in $\mathcal{P}$, the estimate (2.14) implies that

$$\Big|\int_{\mathcal{P}'\cap[s_l,s_{l+1}]}ZdX-\int_{\{s_l,s_{l+1}\}}ZdX\Big|\leqslant C_\alpha|s_{l+1}-s_l|^{3\alpha}\|\mathcal{Z}\|_{\mathbf{X};\alpha}\|\mathbf{X}\|_\alpha.$$

By summing over l, one obtains that

$$\begin{aligned}\Big|\int_{\mathcal{P}'}ZdX-\int_{\mathcal{P}}ZdX\Big|&\leqslant C_\alpha\|\mathcal{Z}\|_{\mathbf{X};\alpha}\|\mathbf{X}\|_\alpha\Big(\sum_{s_l\in\mathcal{P}}(s_{l+1}-s_l)^{3\alpha}\Big)\\&\leqslant C_\alpha\|\mathcal{Z}\|_{\mathbf{X};\alpha}\|\mathbf{X}\|_\alpha|\mathcal{P}|^{3\alpha-1}|t-s|.\end{aligned}$$

A similar estimate holds with $\mathcal{P}$ replaced by $\tilde{\mathcal{P}}$. Therefore, one has

$$\Big|\int_{\tilde{\mathcal{P}}}ZdX-\int_{\mathcal{P}}ZdX\Big|\leqslant C_\alpha\|\mathcal{Z}\|_{\mathbf{X};\alpha}\|\mathbf{X}\|_\alpha(|\mathcal{P}|^{3\alpha-1}+|\tilde{\mathcal{P}}|^{3\alpha-1})|t-s|.$$

The left hand side can be made arbitrarily small when the mesh sizes of $\mathcal{P},\tilde{\mathcal{P}}$ are small enough. The existence of $\lim_{|\mathcal{P}|\to0}\int_{\mathcal{P}}ZdX$ thus follows from the Cauchy criterion.

Let $\mathcal{I}=(I^0,I^1)$ be defined by (2.10). By taking $|\mathcal{P}|\to0$ in the estimate (2.14), one obtains that

$$\big|I^0_{s,t}-(Z^0_sX^1_{s,t}+Z^1_sX^2_{s,t})\big|\leqslant C_\alpha\|\mathcal{Z}\|_{\mathbf{X};\alpha}\|\mathbf{X}\|_\alpha|t-s|^{3\alpha}.$$

Therefore,

$$\begin{aligned}|\mathcal{R}\mathcal{I}_{s,t}|&\leqslant C_\alpha\|\mathcal{Z}\|_{\mathbf{X};\alpha}\|\mathbf{X}\|_\alpha|t-s|^{3\alpha}+|Z^1_s|\cdot|X^2_{s,t}|\\&\leqslant C_\alpha T^\alpha\|\mathcal{Z}\|_{\mathbf{X};\alpha}\|\mathbf{X}\|_\alpha|t-s|^{2\alpha}+(|Z^1_s-Z^1_0|+|Z^1_0|)|X^2_{s,t}|\\&\leqslant C_\alpha T^\alpha\|\mathcal{Z}\|_{\mathbf{X};\alpha}\|\mathbf{X}\|_\alpha|t-s|^{2\alpha}+\big(T^\alpha\|Z^1\|_\alpha+|Z^1_0|\big)\|X^2\|_{2\alpha}|t-s|^{2\alpha}\end{aligned}$$

and thus

$$\|\mathcal{R}\mathcal{I}\|_{2\alpha}\leqslant(1+C_\alpha)T^\alpha\|\mathbf{X}\|_\alpha\|\mathcal{Z}\|_{\mathbf{X};\alpha}+|Z^1_0|\cdot\|\mathbf{X}\|_\alpha. \tag{2.15}$$

In a similar way, one also has

$$\begin{aligned}|I^1_{s,t}| &= |Z^0_{s,t}| = |Z^1_s X^1_{s,t} + \mathcal{R}\mathcal{Z}_{s,t}| \\ &\leqslant \big(T^\alpha \|Z^1\|_\alpha + |Z^1_0|\big)\|X^1\|_\alpha |t-s|^\alpha + T^\alpha \|\mathcal{R}\mathcal{Z}\|_{2\alpha}|t-s|^\alpha\end{aligned}$$

and thus

$$\|I^1\|_\alpha \leqslant T^\alpha(1+\|\mathbf{X}\|_\alpha)\|\mathcal{Z}\|_{\mathbf{X};\alpha} + |Z^1_0| \cdot \|\mathbf{X}\|_\alpha. \tag{2.16}$$

The estimate (2.15) shows that the remainder $\mathcal{R}\mathcal{I}$ has the required 2α-Hölder regularity and hence $\mathcal{I}$ is a rough path controlled by $\mathbf{X}$. The desired estimate (2.11) clearly follows from (2.15) and (2.16). $\square$

We sometimes use $\int \mathcal{Z} d\mathbf{X}$ to denote the controlled rough path $\mathcal{I}$ defined by (2.10). A simple adaptation of the above proof yields the following continuity estimate.

Theorem 2.6 (Continuity of rough integrals)**.** *Let $\mathbf{X}, \tilde{\mathbf{X}}$ be geometric rough paths over V and let $\mathcal{Z}, \tilde{\mathcal{Z}}$ be paths controlled by $\mathbf{X}, \tilde{\mathbf{X}}$ respectively. Define the rough integral paths $\mathcal{I}, \tilde{\mathcal{I}}$ as in Theorem 2.5. Then*

$$\begin{aligned}d_{\mathbf{X},\tilde{\mathbf{X}};\alpha}(\mathcal{I},\tilde{\mathcal{I}}) \leqslant\ & C_\alpha M(T, \|\mathbf{X}\|_\alpha, \|\tilde{\mathcal{Z}}\|_{\tilde{\mathbf{X}};\alpha}, |\tilde{Z}^1_0|) \\ & \times \big(T^\alpha d_{\mathbf{X},\tilde{\mathbf{X}};\alpha}(\mathcal{Z},\tilde{\mathcal{Z}}) + \rho_\alpha(\mathbf{X},\tilde{\mathbf{X}}) + |Z^1_0 - \tilde{Z}^1_0|\big).\end{aligned}$$

Exercise 2.7. Adapt the argument for Theorem 2.5 to prove Theorem 2.6. [*Hint: As in the previous proof, the key point is to obtain a uniform estimate for*

$$\Big(\int_{\mathcal{P}} Z dX - \int_{\{s,t\}} Z dX\Big) - \Big(\int_{\mathcal{P}} \tilde{Z} d\tilde{X} - \int_{\{s,t\}} \tilde{Z} d\tilde{X}\Big).]$$

Exercise 2.8. Let $\mathbf{X}$ be an α-Hölder weakly geometric rough path over V with $\alpha \in (1/4, 1/3]$. Let $F : V \to \mathcal{L}(V;U)$ be a C^3_b-function (in geometric terms F is called a U-valued *one-form*). Define

$$Z^0_t \triangleq F(X^1_{0,t}),\ Z^1_t \triangleq DF(X^1_{0,t}),\ Z^2_t \triangleq D^2F(X^1_{0,t})$$

respectively. Show that $\mathcal{Z} \triangleq (Z^0, Z^1, Z^2) \in \mathcal{D}_{\mathbf{X};\alpha}(\mathcal{L}(V;U))$. As a result of Theorem 2.5, one can define the integral

$$\begin{aligned}\int_0^t F(\mathbf{X}) d\mathbf{X} \triangleq \lim_{|\mathcal{P}|\to 0} \sum_{t_i \in \mathcal{P}} \big(& F(X^1_{0,t_{i-1}}) X^1_{t_{i-1},t_i} + DF(X^1_{0,t_{i-1}}) X^2_{t_{i-1},t_i} \\ & + D^2F(X^1_{0,t_{i-1}}) X^3_{t_{i-1},t_i}\big).\end{aligned}$$

Compare this situation with the case of $\alpha \in (1/3, 1/2]$.

Exercise 2.9. Let $\mathbf{X}$ be an α-Hölder weakly geometric rough path ($\alpha \in (1/3, 1/2]$) and let $\mathcal{Z} = (Z^0, Z^1)$ be controlled by $\mathbf{X}$.

(i) Suppose that $\mathbf{X}$ is the canonical lifting of a smooth path. Show that the value of the rough integral $\int_0^t Z dX$ does not depend on the derivative path Z^1.
(ii) Construct an example of $\mathbf{X}$ and $\mathcal{Z}, \tilde{\mathcal{Z}}$ controlled by $\mathbf{X}$, such that $Z^0 = \tilde{Z}^0$ but

$$\int_0^\cdot Z dX \neq \int_0^\cdot \tilde{Z} dX.$$

2.3 Rough differential equations

We now come to the fundamental problem that drives the development of rough path theory: solving differential equations driven by rough paths (RDEs). To be more specific, in this section we shall address the following two questions mathematically:

Question 1. How can one make sense of the RDE

$$\begin{cases} dY_t = F(Y_t) d\mathbf{X}_t, \\ Y_0 : \text{given initial condition}, \end{cases} \tag{2.17}$$

where $\mathbf{X}$ is a weakly geometric rough path and F is a suitably regular function?

Question 2. How can one prove the continuity of the solution map

$$(Y_0, \mathbf{X}) \mapsto Y$$

with respect to suitable path topologies?

Having all the necessary rough path tools at hand, the solutions to these questions become a standard matter of analysis and the essential idea can be summarised as follows.

(i) Interpret the RDE (2.17) as an integral equation:

$$\mathcal{Y}_t = Y_0 + \int_0^t F(\mathcal{Y}_s) d\mathbf{X}_s$$

in the space of rough paths controlled by $\mathbf{X}$.
(ii) Consider the transformation

$$\mathcal{M} : \mathcal{Y} \mapsto Y_0 + \int_0^\cdot F(\mathcal{Y}) d\mathbf{X}$$

over the control rough path space. Show that $\mathcal{M}$ is a contraction mapping when restricted on a metric ball over a *small time* period.
(iii) Patch the small-time solutions in Step (ii) to obtain a global solution.
(iv) Uniqueness and continuity of the solution map follow from the continuity estimates for regular transforms and rough integrals derived in Sections 2.1 and 2.2 respectively.

We now give the precise mathematical definition of solutions. Let V, U be Banach spaces. Let $\mathbf{X}$ be an α-Hölder weakly geometric rough path over V with $\alpha \in (0, 1/2]$. Let $F : U \to \mathcal{L}(V; U)$ be a continuously differentiable function with bounded derivatives up to order $N+1$ ($N \triangleq \lfloor 1/\alpha \rfloor$). All paths are assumed to be defined on $[0, T]$.

Definition 2.10. Given $Y_0 \in U$, we say that $\mathcal{Y} = (Y^0, Y^1, \cdots, Y^{N-1}) \in \mathcal{D}_{\mathbf{X};\alpha}(U)$ is a *solution to the RDE (2.17) with initial condition* Y_0, if

$$Y_t^0 = Y_0 + \Big(\int_0^{\cdot} F(\mathcal{Y}) d\mathbf{X}\Big)_t^0,\ Y_t^i = \Big(\int_0^{\cdot} F(\mathcal{Y}) d\mathbf{X}\Big)_t^i \quad (1 \leqslant i \leqslant N-1)$$

for all $t \in [0, T]$.

Remark 2.11. If $\mathcal{Y} = (Y^0, Y^1, \cdots, Y^{N-1})$ is a solution to the RDE, the derivative paths $Y^1, \cdots, Y^N$ are all uniquely determined by the zeroth level path Y^0. Indeed, for each fixed t, the value of Y_t^i is explicitly determined by the value of Y_t^0. One can easily see this in the case of $\alpha \in (1/3, 1/2]$ ($Y_t^1 = F(Y_t^0)$).

The main theorem in this section is stated as follows.

Theorem 2.12. *(i) [Existence and uniqueness] For each $Y_0 \in U$, there exists a unique solution $\mathcal{Y} \in \mathcal{D}_{\mathbf{X};\alpha}(U)$ to the RDE (2.17) in the sense of Definition 2.10.*
(ii) [Continuity estimate] Let $\mathbf{X}, \tilde{\mathbf{X}}$ be α-Hölder weakly geometric rough paths over V and let $Y_0, \tilde{Y}_0 \in U$. Let $\mathcal{Y}, \tilde{\mathcal{Y}}$ be the solutions to (2.17) driven by $\mathbf{X}, \tilde{\mathbf{X}}$ with initial conditions $Y_0, \tilde{Y}_0$ respectively. Then the following estimate holds true:

$$d_{\mathbf{X},\tilde{\mathbf{X}};\alpha}(\mathcal{Y}, \tilde{\mathcal{Y}}) \leqslant M(\alpha, T, \|F\|_{C_b^{N+1}}, \|\mathbf{X}\|_\alpha, \|\tilde{\mathbf{X}}\|_\alpha)\big(\rho_\alpha(\mathbf{X}, \tilde{\mathbf{X}}) + |Y_0 - \tilde{Y}_0|\big). \quad (2.18)$$

Remark 2.13. The continuity estimate (2.18) is nowadays commonly known as the *universal limit theorem.*

In the following subsections, we develop the major steps for proving Theorem 2.12. To ease notation, we continue to assume that $\alpha \in (1/3, 1/2]$ so that $N = 2$. Let $F : U \to \mathcal{L}(V; U)$ be a given fixed C_b^3-function throughout the rest.

2.3.1 Composition of regular transformation and rough integration

We first state a lemma that contains the core estimates needed for the proof of Theorem 2.12. All paths below are assumed to be defined on $[0,\tau]$ with $\tau > 0$.

Lemma 2.14. *(i) Given weakly geometric rough path* $\mathbf{X}$ *and* $\mathcal{Y} \in \mathcal{D}_{\mathbf{X};\alpha}(U)$*, we define the* U*-valued controlled rough path* $\mathcal{J} \triangleq Y_0^0 + \int_0^{\cdot} F(\mathcal{Y})d\mathbf{X}$*. More specifically,*

$$J_t^0 = Y_0^0 + \int_0^t F(Y)dX, \ \ J_t^1 = F(Y_t^0).$$

Then

$$\|\mathcal{J}\|_{\mathbf{X};\alpha} \leqslant C_\alpha\big(\tau^\alpha M(\tau, \|F\|_{C_b^2}, \|\mathbf{X}\|_\alpha, |Y_0^1|, \|\mathcal{Y}\|_{\mathbf{X};\alpha}) + |F(\mathcal{Y})_0^1| \cdot \|\mathbf{X}\|_\alpha\big). \tag{2.19}$$

(ii) Given another weakly geometric rough path $\tilde{\mathbf{X}}$ *and* $\tilde{\mathcal{Y}} \in \mathcal{D}_{\tilde{\mathbf{X}};\alpha}(U)$*, we define* $\tilde{\mathcal{J}}$ *similarly. Then*

$$\begin{aligned} d_{\mathbf{X},\tilde{\mathbf{X}};\alpha}(\mathcal{J},\tilde{\mathcal{J}}) \leqslant C_\alpha M(\tau, \|F\|_{C_b^3}, \|\mathbf{X}\|_\alpha, \|\tilde{\mathbf{X}}\|_\alpha, |Y_0^1|, |\tilde{Y}_0^1|, \|\mathcal{Y}\|_{\mathbf{X};\alpha}, \|\tilde{\mathcal{Y}}\|_{\tilde{\mathbf{X}};\alpha}) \\ \times \big(\tau^\alpha d_{\mathbf{X},\tilde{\mathbf{X}};\alpha}(\mathcal{Y},\tilde{\mathcal{Y}}) + |F(\mathcal{Y})_0^1 - F(\tilde{\mathcal{Y}})_0^1| + \rho_\alpha(\mathbf{X},\tilde{\mathbf{X}})\big). \end{aligned} \tag{2.20}$$

Proof. This is an immediate consequence of Proposition 2.3 and Theorem 2.6 . □

Remark 2.15. The appearance of the factor τ^α and the fact that the function M does not depend on $Y_0^0, \tilde{Y}_0^0$ are both crucial for the proof of Theorem 2.12.

2.3.2 Local existence

We prove existence and uniqueness by using the Banach fixed point theorem. Recall that solutions to the RDE (2.17) are defined as fixed points of the transformation

$$\mathcal{M} : \mathcal{D}_{\mathbf{X};\alpha}(U) \to \mathcal{D}_{\mathbf{X};\alpha}(U), \ \mathcal{M}(\mathcal{Y}) \triangleq Y_0 + \int_0^{\cdot} F(\mathcal{Y})d\mathbf{X}.$$

Note that the "constant" (i.e. the function M) appearing in the continuity estimate (2.20) depends on $\|\mathcal{Y}\|_{\mathbf{X};\alpha}, \|\tilde{\mathcal{Y}}\|_{\tilde{\mathbf{X}};\alpha}$. As a result, one can only hope for $\mathcal{M}$ being a contraction when it is restricted on a bounded subset, say a metric ball. The contraction factor comes from a further restriction on a small time interval so that the coefficient $C_\alpha \tau^\alpha M$ in the relevant estimate can be made smaller than one.

We assume for now that all paths below are defined on $[0,\tau]$. Let $Y_0 \in U$ be a given fixed initial condition. To determine a metric ball for the restriction of $\mathcal{M}$, it is natural to fix a center $\mathcal{W} = (W^0, W^1) \in \mathcal{D}_{\mathbf{X};\alpha}(U)$ that satisfies $W_0^0 = Y_0$. A canonical choice is that

$$W_t^0 \triangleq Y_0 + F(Y_0)X_{0,t}^1, \ W_t^1 = F(Y_0), \quad 0 \leqslant t \leqslant \tau. \tag{2.21}$$

Lemma 2.16. *The path $\mathcal{W}$ defined by (2.21) is an α-Hölder rough path controlled by $\mathbf{X}$.*

Proof. It is clear that W^0 and W^1 are both α-Hölder continuous. In addition, one has

$$\mathcal{R}\mathcal{W}_{s,t} \triangleq W_{s,t}^0 - W_s^1 X_{s,t}^1 = F(Y_0)X_{s,t}^1 - F(Y_0)X_{s,t}^1 = 0.$$

□

Now we define the closed subset

$$B_\tau(\mathcal{W}, R) \triangleq \{\mathcal{Y} \in \mathcal{D}_{\mathbf{X};\alpha}(U) : \|\mathcal{Y} - \mathcal{W}\|_{\mathbf{X};\alpha} \leqslant R, \mathcal{Y}_0 = \mathcal{W}_0\},$$

where the radius R is to be chosen later on and the subscript τ means that all relevant paths are restricted to $[0, \tau]$. We want $\mathcal{M}$ to map $B_\tau(\mathcal{W}, R)$ into $B_\tau(\mathcal{W}, R)$. Given $\mathcal{Y} \in B_\tau(\mathcal{W}, R)$, it is clear that $(\mathcal{M}\mathcal{Y})_0 = \mathcal{W}_0$. In addition, from (2.19) one knows that

$$\|\mathcal{M}\mathcal{Y} - \mathcal{W}\|_{\mathbf{X};\alpha} \leqslant C_\alpha \tau^\alpha M(\tau, \|F\|_{C_b^2}, \|\mathbf{X}\|_\alpha, R) + C_\alpha |F(\mathcal{Y})_0^1| \cdot \|\mathbf{X}\|_\alpha.$$

Since

$$F(\mathcal{Y})_0^1 = F'(Y_0^0)(F(Y_0^0) \circ \cdot) \in \mathcal{L}(V; \mathcal{L}(V; U)), \tag{2.22}$$

one has $|F(\mathcal{Y})_0^1| \leqslant M(\|F\|_{C_b^1})$ and thus

$$C_\alpha |F(\mathcal{Y})_0^1| \cdot \|\mathbf{X}\|_\alpha \leqslant C_\alpha M(\|F\|_{C_b^1}, \|\mathbf{X}\|_\alpha).$$

By choosing

$$R \triangleq 2C_\alpha M(\|F\|_{C_b^1}, \|\mathbf{X}\|_\alpha), \tag{2.23}$$

it follows that

$$\|\mathcal{M}\mathcal{Y} - \mathcal{W}\|_{\mathbf{X};\alpha} \leqslant C_\alpha \tau^\alpha M(\tau, \|F\|_{C_b^2}, \|\mathbf{X}\|_\alpha, R) + \frac{R}{2}, \quad \forall \mathcal{Y} \in B_\tau(\mathcal{W}, R).$$

By taking τ to be small enough such that

$$C_\alpha \tau^\alpha M(\tau, \|F\|_{C_b^2}, \|\mathbf{X}\|_\alpha, R) < \frac{R}{2},$$

one can now ensure that

$$\mathcal{M}(B_\tau(\mathcal{W},R)) \subseteq B_\tau(\mathcal{W},R).$$

To make $\mathcal{M}$ into a contraction, we recall from (2.20) that

$$\|\mathcal{M}\mathcal{Y} - \mathcal{M}\tilde{\mathcal{Y}}\|_{\mathbf{X};\alpha} \leqslant C_\alpha \tau^\alpha M(\tau, \|F\|_{C_b^3}, \|\mathbf{X}\|_\alpha, R)\|\mathcal{Y} - \tilde{\mathcal{Y}}\|_{\mathbf{X};\alpha}$$

for all $\mathcal{Y}, \tilde{\mathcal{Y}} \in B_\tau(\mathcal{W},R)$. By further reducing τ so that

$$C_\alpha \tau^\alpha M(\tau, \|F\|_{C_b^3}, \|\mathbf{X}\|_\alpha, R) < \frac{1}{2},$$

one obtains the contraction property

$$\|\mathcal{M}\mathcal{Y} - \mathcal{M}\tilde{\mathcal{Y}}\|_{\mathbf{X};\alpha} \leqslant \frac{1}{2}\|\mathcal{Y} - \tilde{\mathcal{Y}}\|_{\mathbf{X};\alpha}.$$

According to the Banach fixed point theorem, under the above choices of R, τ there is a unique $\mathcal{Y} \in B_\tau(\mathcal{W},R)$ such that $\mathcal{M}\mathcal{Y} = \mathcal{Y}$, namely a solution to the RDE (2.17) on the small time interval $[0,\tau]$.

To summarise, we have proven the following local existence result.

Lemma 2.17. *There exists $\tau > 0$, which is independent of the initial condition Y_0 and depends only on α, $\|F\|_{C_b^3}$ and $\|\mathbf{X}\|_\alpha$, such that the RDE (2.17) admits a solution $\mathcal{Y}$ on $[0,\tau]$ that satisfies*

$$\|\mathcal{Y}\|_{\mathbf{X};\alpha} \leqslant C_\alpha M(\|F\|_{C_b^1}, \|\mathbf{X}\|_\alpha).$$

2.3.3 Global existence, uniqueness and continuity

The global existence of solutions is an easy consequence of Lemma 2.17 via a standard patching argument. Let τ be as in that lemma. The crucial point is that τ does not depend on the initial condition. As a result, after obtaining a solution on $[0,\tau]$, one can then treat $\mathcal{Y}_\tau$ as a new initial condition to obtain a solution to the RDE driven by $\mathbf{X}$ on $[\tau, 2\tau]$. This procedure continues inductively and the resulting path $\mathcal{Y}$ is clearly a global solution on $[0,T]$ (or indeed on $[0,\infty)$ if the underlying time horizon is infinite).

To prove uniqueness, let $\tilde{\mathcal{Y}}$ be another solution to the RDE with the same initial condition Y_0. Define

$$\sigma \triangleq \sup\{t \geqslant 0 : \tilde{\mathcal{Y}}_s = \mathcal{Y}_s \text{ on } [0,t]\}.$$

Suppose on the contrary that $\sigma < \infty$. Note that $\tilde{\mathcal{Y}}_\sigma = \mathcal{Y}_\sigma$. According to (2.20), for each $\tau > 0$ one has

$$\|\tilde{\mathcal{Y}}|_{[\sigma,\sigma+\tau]} - \mathcal{Y}|_{[\sigma,\sigma+\tau]}\|_{\mathbf{X};\alpha} \leqslant C_\alpha \tau^\alpha M(\tau, \|F\|_{C_b^3}, \|\mathcal{Y}\|_{\mathbf{X};\alpha}, \|\tilde{\mathcal{Y}}\|_{\mathbf{X};\alpha}, \|\mathbf{X}\|_\alpha) \\ \times \|\tilde{\mathcal{Y}}|_{[\sigma,\sigma+\tau]} - \mathcal{Y}|_{[\sigma,\sigma+\tau]}\|_{\mathbf{X};\alpha}.$$

By taking τ to be small, one finds that

$$\|\tilde{\mathcal{Y}}|_{[\sigma,\sigma+\tau]} - \mathcal{Y}|_{[\sigma,\sigma+\tau]}\|_{\mathbf{X};\alpha} \leqslant \frac{1}{2}\|\tilde{\mathcal{Y}}|_{[\sigma,\sigma+\tau]} - \mathcal{Y}|_{[\sigma,\sigma+\tau]}\|_{\mathbf{X};\alpha}.$$

This implies $\tilde{\mathcal{Y}} = \mathcal{Y}$ on $[\sigma, \sigma + \tau]$, contradicting the definition of σ. Therefore, $\tilde{\mathcal{Y}} = \mathcal{Y}$ for all time.

Finally, we prove the local Lipschitz-continuity estimate (2.18). Set $B \triangleq \max\{\|\mathbf{X}\|_\alpha, \|\tilde{\mathbf{X}}\|_\alpha\}$. The following lemma provides the key ingredient of the proof.

Lemma 2.18. *There exists $\tau > 0$ depending only on $\alpha, B, \|F\|_{C_b^3}$, such that for any interval $[\sigma, \sigma + \tau]$ of length τ, one has*

$$d_{\mathbf{X},\tilde{\mathbf{X}};\alpha}(\mathcal{Y}, \tilde{\mathcal{Y}})|_{[\sigma,\sigma+\tau]} \leqslant M(\alpha, \|F\|_{C_b^3}, B)\big(|Y_\sigma^0 - \tilde{Y}_\sigma^0| + \rho_\alpha(\mathbf{X}, \tilde{\mathbf{X}})\big). \tag{2.24}$$

and

$$|Y_{\sigma+\tau}^0 - \tilde{Y}_{\sigma+\tau}^0| \leqslant M(\alpha, \|F\|_{C_b^3}, B)\big(|Y_\sigma^0 - \tilde{Y}_\sigma^0| + \rho_\alpha(\mathbf{X}, \tilde{\mathbf{X}})\big). \tag{2.25}$$

Proof. Let R be defined by (2.23) with $\|\mathbf{X}\|_\alpha$ replaced by B and let τ be as in Lemma 2.17 which now depends on B and $\|F\|_{C_b^3}$. Given any interval $[\sigma, \sigma + \tau]$ of length τ, the continuity estimate (2.20) implies that

$$d_{\mathbf{X},\tilde{\mathbf{X}};\alpha}(\mathcal{Y}, \tilde{\mathcal{Y}})|_{[\sigma,\sigma+\tau]} \leqslant C_\alpha M(\|F\|_{C_b^3}, B)\big(\tau^\alpha d_{\mathbf{X},\tilde{\mathbf{X}};\alpha}(\mathcal{Y}, \tilde{\mathcal{Y}})|_{[\sigma,\sigma+\tau]} \\ + |F(\mathcal{Y})_\sigma^1 - F(\tilde{\mathcal{Y}})_\sigma^1| + \rho_\alpha(\mathbf{X}, \tilde{\mathbf{X}})\big). \tag{2.26}$$

In addition, from the definition of $F(\mathcal{Y})^1$ (cf. (2.4) and Remark 2.11) it is easily seen that

$$\|F(\mathcal{Y})_\sigma^1 - F(\tilde{\mathcal{Y}})_\sigma^1\| \leqslant M(\|F\|_{C_b^2})|Y_\sigma^0 - \tilde{Y}_\sigma^0|. \tag{2.27}$$

By substituting (2.27) into (2.26) and further reducing τ to be such that

$$C_\alpha M(\|F\|_{C_b^3}, B)\tau^\alpha < \frac{1}{2},$$

one arrives at (2.24). The estimate (2.25) is obtained as follows:

$$\begin{aligned}|Y^0_{\sigma+\tau}-\tilde{Y}^0_{\sigma+\tau}| &\leqslant |Y^0_\sigma-\tilde{Y}^0_\sigma|+\tau^\alpha\|(Y^0-\tilde{Y}^0)|_{[\sigma,\sigma+\tau]}\|_\alpha \\ &\leqslant |Y^0_\sigma-\tilde{Y}^0_\sigma|+\tau^\alpha M(\tau,\|F\|_{C^1_b},B)\big(|Y^0_\sigma-\tilde{Y}^0_\sigma| \\ &\quad + d_{\mathbf{X},\tilde{\mathbf{X}};\alpha}(\mathcal{Y},\tilde{\mathcal{Y}})|_{[\sigma,\sigma+\tau]}+\rho_\alpha(\mathbf{X},\tilde{\mathbf{X}})\big) \quad \text{(by (1.19))} \\ &\leqslant M(\alpha,\|F\|_{C^3_b},B)\big(|Y^0_\sigma-\tilde{Y}^0_\sigma|+\rho_\alpha(\mathbf{X},\tilde{\mathbf{X}})\big). \quad \text{(by (2.24))}.\end{aligned}$$

□

To complete the proof of (2.18), let τ be as in Lemma 2.18. We divide $[0,T]$ evenly into $K \triangleq \lfloor T/\tau \rfloor + 1$ sub-intervals. By applying (2.25) inductively to each sub-interval, one sees that

$$|Y^0_\sigma-\tilde{Y}^0_\sigma| \leqslant M(\alpha,T,\|F\|_{C^3_b},B)\big(|Y_0-\tilde{Y}_0|+\rho_\alpha(\mathbf{X},\tilde{\mathbf{X}})\big)$$

at every partition point σ. This inequality, together with the inequality (2.24) applied to each sub-interval, yields the desired continuity estimate (2.18) on $[0,T]$.

The proof of Theorem 2.12 is now complete.

We give a few remarks to conclude our discussion.

(i) All results and quantitative estimates in this chapter extend to the case of arbitrary Hölder regularity. Except for Theorem 2.1 (stability of controlled rough paths) that requires deeper algebraic considerations, the extension of all other results is only a technical matter. We refer the reader to [BG20] for a detailed discussion on this general case.
(ii) If the coefficient function F and/or its derivatives are not uniformly bounded, one can still prove existence and uniqueness in the same way, however, the solution to the RDE (2.17) may only be defined up to its intrinsic explosion time.
(iii) In the continuity estimates (2.7) and (2.18), one can also take into account a perturbation of F, say $\tilde{F}$. In this case, an extra term of $\|F-\tilde{F}\|_{C^{N+1}_b}$ will appear on the right hand side of the relevant estimate.
(iv) In Theorem 2.12, the regularity assumption on F is not optimal. There is a general notion of γ-Lipschitz functions on Banach spaces (cf. Stein [Ste70]), which formally requires that $F \in C^{\lfloor\gamma\rfloor}$ and $D^{\lfloor\gamma\rfloor}F$ is $(\gamma-\lfloor\gamma\rfloor)$-Hölder continuous. The optimal results can be concisely stated as follows:

$$F \text{ is } \gamma\text{-Lipschitz with } \gamma>\alpha^{-1} \implies \text{existence and uniqueness} \tag{2.28}$$

and in finite dimensions

$$F \text{ is } \gamma\text{-Lipschitz with } \gamma>\alpha^{-1}-1 \implies \text{existence.} \tag{2.29}$$

The result of (2.28) can be obtained by sacrificing the Hölder regularity of the remainders of the controlled rough path $F(\mathcal{Y})$. More specifically, $\mathcal{R}F(\mathcal{Y})^i_{s,t}$ should have regularity $|t-s|^{(\gamma-1-i)\alpha}$ instead of $|t-s|^{(N-i)\alpha}$. Correspondingly, the definition of controlled rough paths needs to be relaxed to allow more flexible Hölder exponents for the remainders (cf. [Gub04] for the case of $\alpha > 1/3$ whose extension to the general case does not involve essential difficulties). The proof of (2.29) relies on the Leray-Schauder fixed point theorem (cf. [Gub04]), which requires the compactness of the transformation $\mathcal{M}$ and it is only true in finite dimensions. To the best of my knowledge, I am not aware of the correctness of (2.29) in infinite dimensions.
(v) The controlled rough path approach we presented here is essentially equivalent to Lyons' original approach (cf. [Lyo98]) and Davie's approach (cf. Davie [Dav08] as well as [FV10]). In other words, for the same formal RDE (2.17), all these approaches lead to exactly the same solution path Y (more precisely, the same base level path of the full rough path solution as these approaches formulate solutions in different rough path spaces). Rough path theory can also be regarded as a special example of the more general theory of regularity structures developed by Hairer [Hai14].

Exercise 2.19. Consider the RDE (2.17) where $\mathbf{X}$ is a weakly geometric rough path over V and $F : U \to \mathcal{L}(V;U)$ is a continuous linear function.

(i) Show that there exists a unique global solution for any given initial condition $Y_0 \in U$.
(ii) Suppose that $\mathbf{X}$ is the canonical lifting of a path x with bounded total variation. Show that the zeroth level path of the solution is explicitly given by

$$Y_t^0 = Y_0 + \sum_{n=1}^{\infty} F^{(n)}(Y_0) \int_{0<t_1<\cdots<t_n<t} dx_{t_1} \otimes \cdots \otimes dx_{t_n},$$

where $F^{(n)} : U \to \mathcal{L}(V^{\otimes n};U)$ is the linear operator induced by

$$F^{(n)}(y)(v_1 \otimes \cdots \otimes v_n) \triangleq F(\cdots(F(F(y)v_1)v_2)\cdots)v_n, \quad y \in U, v_1, \cdots, v_n \in V.$$

(iii) Suppose that $U = V = \mathbb{R}^2$. Construct an example of an α-Hölder rough path $\mathbf{X}$ with $\alpha \in (1/3, 1/2]$ and a smooth function F with linear growth, i.e.

$$|F(u)| \leqslant C(1+|u|) \quad \forall u \in \mathbb{R}^2,$$

such that the zeroth level path of the RDE solution explodes to infinity in finite time.

Exercise 2.20. Consider the RDE (2.17) where $U = \mathbb{R}^n, V = \mathbb{R}^d$. We write $F = (V_1, \cdots, V_d)$ where each $V_i : \mathbb{R}^n \to \mathbb{R}^n$ is now a C_b^∞ vector field on $\mathbb{R}^n$.

(i) Suppose that the vector fields $V_1, \cdots, V_d$ are *commutative* in the sense that $[V_i, V_j] = 0$ for all i, j (cf. (4.3) for the definition of Lie brackets). Show that the zeroth level path of the RDE solution with initial condition Y_0 is explicitly given by

$$Y_t^0 = \exp\big(\sum_{i=1}^d X_t^{1;i} V_i\big)(Y_0).$$

Here $X^{1;i}$ denotes the i-th coordinate component of X^1. Given any C_b^∞ vector field W on $\mathbb{R}^n$, the notation $\exp W : U \to U$ denotes the time one mapping of the flow associated with W, i.e.

$$\exp(W)(y) \triangleq z_1,$$

where $(z_t)_{0\leqslant t\leqslant 1}$ is the unique solution to the ODE

$$\begin{cases} \dot{z}_t = W(z_t), & 0 \leqslant t \leqslant 1; \\ z_0 = y. \end{cases}$$

This part shows that RDE theory reduces to a trivial situation if the driving path $\mathbf{X}$ is one dimensional (i.e. $V = \mathbb{R}^1$) or if the vector fields are commutative when $\dim V > 1$.

(ii) Suppose that the vector fields $V_1, \cdots, V_d$ are *step-2 nilpotent* in the sense that

$$[V_i, [V_j, V_k]] = 0 \quad \forall i, j, k = 1, \cdots, d.$$

Show that

$$Y_t^0 = \exp\big(\sum_{i=1}^d X_t^{1;i} V_i + \frac{1}{2} \sum_{1\leqslant i<j\leqslant d} \big(X_{0,t}^{2;i,j} - X_{0,t}^{2;j,i}\big)[V_i, V_j]\big)(Y_0)$$

in this case, where $X^{2;i,j}$ denotes the (i, j)-coordinate component of X^2 under the canonical tensor basis of $(\mathbb{R}^d)^{\otimes 2}$ (cf. Example 1.3). Note that this part does not trivialise the rough path perspective, as the second level path X^2 is not canonically defined and it comes with the definition of $\mathbf{X}$.

Exercise 2.21. Let $\gamma : [0, L] \to \mathbb{R}^2$ be a continuously differentiable path with unit speed parametrisation (i.e. $|\dot{\gamma}_t| = 1$ for all t). For each $n \geqslant 1$, we equip

$(\mathbb{R}^2)^{\otimes n}$ with the Hilbert-Schmidt tensor norm $\|\cdot\|_{\mathrm{HS}}$ (cf. the first part of Example 1.5) and define

$$g_n \triangleq \int_{0<t_1<\cdots<t_n<L} \dot{\gamma}_{t_1} \otimes \cdots \otimes \dot{\gamma}_{t_n} dt_1 \cdots dt_n \in (\mathbb{R}^2)^{\otimes n}.$$

The aim of this problem is to show that

$$\lim_{n\to\infty} (n!\|g_n\|_{\mathrm{HS}})^{1/n} = \text{Length of } \gamma \tag{2.30}$$

by using the method of differential equations.

(i) Show that

$$\|g_n\|_{\mathrm{HS}} \leqslant \frac{L^n}{n!} \quad \forall n \geqslant 1.$$

(ii) Let $\{a_n : n \geqslant 1\}$ be a sub-additive real sequence, i.e. $a_{m+n} \leqslant a_m + a_n$ for all $m, n \geqslant 1$. Show that $\lim_{n\to\infty} \frac{a_n}{n}$ exists and is equal to $\inf \frac{a_n}{n}$.

(iii) Use the shuffle product formula (1.15) and the result of Part (ii) to show that the limit on the left hand side of (2.30) exists. Denote this limit as $\tilde{L}$.

(iv) Define the linear mapping

$$\Phi : \mathbb{R}^2 \to M_2,\ (x, y) \mapsto \begin{pmatrix} x & y \\ y & -x \end{pmatrix},$$

where M_2 denotes the space of 2×2 real matrices. For each $\lambda > 0$, let $\Gamma_t^\lambda \in M_2$ $(0 \leqslant t \leqslant L)$ be the solution to the differential equation

$$\begin{cases} d\Gamma_t^\lambda = \Gamma_t^\lambda \Phi(\lambda d\gamma_t), & 0 \leqslant t \leqslant L; \\ \Gamma_0^\lambda = \mathrm{Id}. \end{cases}$$

Show that Γ_t^λ is explicitly given by

$$\Gamma_t^\lambda = \sum_{n=0}^{\infty} \lambda^n \Phi^{\otimes n}(g_n), \quad 0 \leqslant t \leqslant L,$$

where $\Phi^{\otimes n} : (\mathbb{R}^2)^{\otimes n} \to M_2$ is the linear operator induced by

$$\Phi^{\otimes n}(v_1 \otimes \cdots \otimes v_n) \triangleq \Phi(v_1) \cdots \Phi(v_n), \quad v_1, \cdots, v_n \in V.$$

(v) Use the result of Part (vi) to show that

$$\tilde{L} \geqslant \overline{\lim_{\lambda\to\infty}} \frac{\|\Gamma_L^\lambda\|_{\mathbb{R}^2\to\mathbb{R}^2}}{\lambda}.$$

(vi) Write $\gamma_t = (\cos\beta_t, \sin\beta_t)$ where β_t is the continuous angular path of γ_t. Define $w_t^\lambda \triangleq \Gamma_t^\lambda \xi$ where $\xi \in \mathbb{R}^2$ is a given fixed unit (column) vector. Show that w_t^λ satisfies the ODE

$$\begin{cases} \dot{w}_t^\lambda = \lambda \begin{pmatrix} \cos\alpha_t & \sin\alpha_t \\ \sin\alpha_t & -\cos\alpha_t \end{pmatrix} w_t^\lambda, \quad 0 \leqslant t \leqslant L; \\ w_0^\lambda = \xi, \end{cases}$$

where $\alpha_t \triangleq \beta_{L-t}$ is the reversal of β.
(vii) Let $(\rho_t^\lambda, \phi_t^\lambda)$ denote the polar coordinates of w_t^λ (i.e. $w_t^\lambda = \rho_t^\lambda e^{i\phi_t^\lambda}$). Show that

$$\begin{cases} \dot{\rho}_t^\lambda = \lambda \rho_t^\lambda \cos(\alpha_t - 2\phi_t^\lambda), \\ \dot{\phi}_t^\lambda = \lambda \sin(\alpha_t - 2\phi_t^\lambda). \end{cases}$$

(viii) Use the result of Part (v) to conclude that

$$\tilde{L} \geqslant \varlimsup_{\lambda\to\infty} \int_0^L \cos(\alpha_t - 2\phi_t^\lambda) dt.$$

(ix) Show that $2\phi_t^\lambda$ converges to α_t as $\lambda \to \infty$ uniformly in $t \in [0, L]$.
(x) Use the above steps to conclude that $\tilde{L} = L$.

3 The Lie algebraic foundation of rough path theory[1]

When the roughness $\alpha \leqslant 1/3$, the algebraic structure encoded in the shuffle product formula (1.15) enters the theory in an essential way. In Section 1.2.2, we derived this formula from the analytic perspective of iterated integrals. In this chapter, we give an algebraic characterisation of the shuffle product structure from the perspective of free Lie algebras. Stated in a concise way, *a tensor series satisfies the shuffle product formula if and only if its logarithm is a Lie series*. This elegant result, which was originally due to Chen [Che57], lays the algebraic foundation of rough path theory and has far-reaching implications.

After introducing the basic definitions, we develop the proof of Chen's theorem in Section 3.2. The core of this part are various equivalent characterisations of Lie series. In Section 3.3, we give a derivation of the classical Baker-Campbell-Hausdorff formula for independent interest. This section is self-contained and does not rely on other parts of the notes. In Section 3.4, we discuss an important application of Chen's theorem to the study of rough differential equations (the Chen-Strichartz formula). Our discussion follows the main lines of [Reu93] (Chen's theorem) and [Bau04] (the Chen-Strichartz formula). This is a purely algebraic chapter and essentially no analysis is involved.

3.1 Basic definitions

In this section, we introduce the basic concepts that are needed for stating Chen's theorem.

Recall that in Section 1.2.1, we have conceptually defined the (algebraic) tensor product $V \otimes W$ of two (real) vector spaces V, W. An important property of $V \otimes W$ is that any bilinear mapping

$$f : V \times W \to Z$$

into an arbitrary vector space Z can be uniquely lifted as a linear mapping $g : V \otimes W \to Z$ such that $f = g \circ \varphi$, where

$$\varphi : V \times W \to V \otimes W, \ \varphi(v, w) \triangleq v \otimes w$$

denotes the canonical embedding. This property uniquely characterises the tensor product space $V \otimes W$ up to isomorphism. As a result, in order to specify a linear

[1] This chapter may be skipped on first reading.

mapping on $V \otimes W$, it is enough to specify a bilinear mapping on $V \times W$, or equivalently specifying the values on monomials $v \otimes w$. Similar remark applies to higher order tensor products.

Throughout the rest, V is a given fixed vector space. We define the *infinite tensor algebra*

$$T((V)) \triangleq \prod_{n=0}^{\infty} V^{\otimes n} = \{\xi = (\xi_0, \xi_1, \xi_2, \cdots) : \xi_n \in V^{\otimes n} \ \forall n\},$$

where $V^{\otimes n}$ is the n-th algebraic tensor product of V and by convention we set $V^{\otimes 0} \triangleq \mathbb{R}$. Elements in $T((V))$ are referred to as (formal) *tensor series*. Given $\xi, \eta \in T((V))$, we define $\xi + \eta$ and $\xi \otimes \eta$ by

$$(\xi + \eta)_n \triangleq \xi_n + \eta_n, \ (\xi \otimes \eta)_n \triangleq \sum_{k=0}^{n} \xi_k \otimes \eta_{n-k}, \quad n = 0, 1, 2, \cdots .$$

One easily checks that $(T((V)), +, \otimes)$ is an algebra with unit $\mathbf{1} = (1, 0, 0, \cdots)$. These definitions are essentially the same as in Section 1.2.1. Denote $\pi_n : T((V)) \to V^{\otimes n}$ as the natural projection.

Let $T_1((V))$ (respectively, $T_0((V))$) be the subspace of tensor series ξ such that $\xi_0 = 1$ (respectively, $\xi_0 = 0$). Define the following two functions

$$\exp : T_0((V)) \to T_1((V)), \ \exp(\xi) \triangleq \sum_{n=0}^{\infty} \frac{\xi^{\otimes n}}{n!} \tag{3.1}$$

and

$$\log : T_1((V)) \to T_0((V)), \ \log(\xi) \triangleq \sum_{n=1}^{\infty} (-1)^{n-1} \frac{(\xi - \mathbf{1})^{\otimes n}}{n}. \tag{3.2}$$

Note that the above series are well-defined since they are *locally finite*, in the sense that the computations of $\pi_n(\exp(\xi))$ and $\pi_n(\log(\xi))$ only involve finite summations. It can be shown that exp and log are inverse to each other:

$$\exp \circ \log = \mathrm{Id}_{T_1((V))}, \ \log \circ \exp = \mathrm{Id}_{T_0((V))}.$$

Given $n \geqslant 1$, let $\mathcal{S}_n$ denote the group of permutations of order n. Recall from Section 1.2.1 that for each $\sigma \in \mathcal{S}_n$, there is an associated permutation operator $P_\sigma : V^{\otimes n} \to V^{\otimes n}$ induced by (1.7), which permutes the slots of an n-tensor under σ. The following definition extracts the algebraic structure encoded in the shuffle product formula (1.15). Recall that $\mathcal{S}(m, n)$ denotes the set of (m, n)-shuffles (cf. Lemma 1.15).

Definition 3.1. Let $\xi \in T_1((V))$ be a tensor series. We say that ξ is a *group-like* element, if it satisfies the following property:

$$\xi_m \otimes \xi_n = \sum_{\sigma \in \mathcal{S}(m,n)} P_\sigma(\xi_{m+n}), \quad \forall m, n \geqslant 0.$$

The space of group-like elements is denoted as $G((V))$.

Example 3.2. Let $x : [0,T] \to V = \mathbb{R}^d$ be a smooth path. According to Example 1.11 and Lemma 1.15, for each fixed $s < t$, the path x lifts to a group-like element

$$\mathbb{X}_{s,t} \triangleq \big(1, x_t - x_s, \int_{s<u<v<t} dx_u \otimes dx_v, \cdots, \int_{s<t_1<\cdots<t_n<t} dx_{t_1} \otimes \cdots \otimes dx_{t_n}, \cdots\big) \quad (3.3)$$

in a canonical way. This result has an important extension to the rough path case. Let $\mathbf{X} = (1, X^1, \cdots, X^N)$ be an α-Hölder rough path over V ($N \triangleq \lfloor \alpha \rfloor$). It was a basic theorem of Lyons [Lyo98] that there exists a unique extension of $\mathbf{X}$ to a multiplicative functional

$$\mathbb{X} = (1, X^1, \cdots, X^N, X^{N+1}, \cdots) : \Delta_T \to T_1((V))$$

such that $\mathbb{X}$ is α-Hölder continuous in the sense of (1.12) for all $n \geqslant 1$. In addition, if $\mathbf{X}$ is weakly geometric, then $\mathbb{X}_{s,t}$ is group-like for all $s < t$.

A key concept in Chen's theorem is the notion of Lie series. Given two tensor series $\xi, \eta \in T((V))$, we define the *Lie bracket* between ξ and η as

$$[\xi, \eta] \triangleq \xi \otimes \eta - \eta \otimes \xi.$$

It is obvious that $[\xi, \eta] = -[\eta, \xi]$. In addition, the Lie bracket satisfies the following so-called *Jacobi identity*:

$$[\xi, [\eta, \gamma]] + [\eta, [\gamma, \xi]] + [\gamma, [\xi, \eta]] = 0 \quad \forall \xi, \eta, \gamma \in T((V)).$$

We define the subspaces $\mathcal{L}_n(V)$ ($n \geqslant 1$) inductively in the following manner: $\mathcal{L}^1(V) \triangleq V$ and

$$\mathcal{L}_{n+1}(V) \triangleq [V, \mathcal{L}_n(V)] \triangleq \mathrm{Span}\{[v, \xi] : v \in V, \xi \in \mathcal{L}_n(V)\}, \quad n \geqslant 1.$$

Elements in $\mathcal{L}_n(V)$ are called *homogeneous Lie polynomials* of degree n. They can be written as finite linear combinations of elements of the form

$$[v_1, [v_2, \cdots, [v_{n-1}, v_n]]] \quad (v_1, \cdots, v_n \in V).$$

Definition 3.3. A *Lie series* is a tensor series $\xi = (\xi_0, \xi_1, \xi_2, \cdots)$ such that $\xi_0 = 0$ and $\xi_n \in \mathcal{L}_n(V)$ for all $n \geqslant 1$. The space of Lie series is denoted as $\mathcal{L}((V))$.

Example 3.4. Let $v_1, v_2 \in V$. One may surprisingly find that $\pi_n \log(e^{v_1} \otimes e^{v_2}) \in \mathcal{L}_n(V)$ at least for the first few n's where explicit calculation is still manageable. It is indeed true that $\log(e^{v_1} \otimes e^{v_2})$ is a Lie series. As we will see, this is essentially the content of the Baker-Campbell-Hausdorff formula or can be seen as an immediate consequence of the more general Chen's theorem.

3.2 Chen's theorem and Dynkin's formula

Under the notation in the previous section, we can now state Chen's remarkable theorem.

Theorem 3.5. *Let $\xi \in T_1((V))$ be a tensor series. Then ξ is a group-like element if and only if $\log \xi$ is a Lie series.*

This theorem has the following important application to rough paths. Indeed, Chen proved his theorem in the context of iterated path integrals (for paths with bounded total variation).

Corollary 3.6. *Let $V = \mathbb{R}^d$ and let $\mathbb{X} : \Delta_T \to T_1((V))$ be the Lyons extension of a weakly geometric rough path over V (cf. Example 3.2). Then $\log \mathbb{X}_{s,t}$ is a Lie series for all $s < t$.*

As the following corollary suggests, Chen's theorem can be viewed as a generalisation of the classical Baker-Campbell-Hausdorff formula. We give an independent proof of the BCH formula in Section 3.3 below.

Corollary 3.7. *Let $v, w \in V$. Then $\log(e^v \otimes e^w)$ is a Lie series.*

Proof. According to Chen's theorem, it is equivalent to showing that $e^v \otimes e^w$ is group-like. This is easy from the perspective of path integrals: one simply observes that $e^v \otimes e^w$ is the canonical lifting of the piecewise linear path

$$x_t \triangleq \begin{cases} tv, & 0 \leqslant t \leqslant 1; \\ v + (t-1)w, & 1 \leqslant t \leqslant 2 \end{cases}$$

defined by the global iterated integrals over $[0, 2]$ (cf. (3.3)). $\square$

Remark 3.8. The shuffle product formula indicates that there are delicate algebraic dependencies among different components of a group-like element g. Chen's theorem suggests that all such depedencies are eliminated by looking at $\log g$. Note that there are no algebraic relations among components of $\log g$, as $\mathcal{L}((V))$ is freely generated by the space V.

In what follows, we develop the proof of Chen's theorem. We adopt a more modern perspective of free Lie algebras [Reu93] instead of the original argument of Chen [Che57]. An advantage is that the free Lie algebra approach reveals the algebraic essence of the theorem in a more fundamental way. A price to pay is that it contains a few algebraic considerations that may not be obvious at first glance.

3.2.1 The coproduct and several basic operators

A key tool for proving Chen's theorem is the use of a coproduct operator. To this end, we first introduce the following *doubly infinite tensor algebra*

$$\mathcal{T}_2((V)) \triangleq \prod_{m,n=0}^{\infty} V^{\otimes m} \boxtimes V^{\otimes n}.$$

Here the component $V^{\otimes m} \boxtimes V^{\otimes n}$ is understood as the tensor product between the two vector spaces $V^{\otimes m}$ and $V^{\otimes n}$ in the algebraic sense of Section 1.2.1. The notation $\boxtimes$ is used to distinguish the new tensor product from the one $\otimes$ used for $V^{\otimes n}$. A generic element in $\mathcal{T}_2((V))$ is a formal infinite tensor series

$$\Xi = (\Xi_{m,n})_{m,n\geqslant 0}, \quad \Xi_{m,n} \in V^{\otimes m} \boxtimes V^{\otimes n}. \tag{3.4}$$

Given $\xi, \eta \in T((V))$, one can form their $\boxtimes$-tensor product

$$\xi \boxtimes \eta \triangleq (\xi_m \boxtimes \eta_n)_{m,n\geqslant 0} \in \mathcal{T}_2((V)) \tag{3.5}$$

in the obvious way. There is a natural grading structure on $\mathcal{T}_2((V))$ defined by grouping components in (3.4) based on homogeneity. More specifically, one has

$$\mathcal{T}_2((V)) \cong \prod_{n=0}^{\infty} \mathcal{V}^n \text{ where } \mathcal{V}^n \triangleq \bigoplus_{k=0}^{n} V^{\otimes k} \boxtimes V^{\otimes (n-k)}$$

and we take the convention that $\mathcal{V}^0 = \langle \mathbf{1} \boxtimes \mathbf{1} \rangle \cong \mathbb{R}$. Given $\xi, \eta \in T((V))$, under the above grading structure the n-th component of $\xi \boxtimes \eta$ is

$$(\xi \boxtimes \eta)_n \triangleq \sum_{k=0}^{n} \xi_k \boxtimes \eta_{n-k} \in \mathcal{V}^n.$$

The space $\mathcal{T}_2((V))$ has a natural addition $+$ and a multiplication $*$ induced by

$$(\xi_1 \boxtimes \eta_1) * (\xi_2 \boxtimes \eta_2) \triangleq (\xi_1 \otimes \xi_2) \boxtimes (\eta_1 \otimes \eta_2) \in V^{\otimes(m+p)} \boxtimes V^{\otimes(n+q)}$$

for $\xi_1 \in V^{\otimes m}, \eta_1 \in V^{\otimes n}, \xi_2 \in V^{\otimes p}, \eta_2 \in V^{\otimes q}$. It follows that $(\mathcal{T}_2((V)), +, *)$ is an algebra with unit $\mathbf{1} \boxtimes \mathbf{1}$.

Remark 3.9. There is a natural embedding of the actual tensor product $T((V)) \boxtimes T((V))$ into $\mathcal{T}_2((V))$ induced by (3.5). However, the latter space is in general larger than the former one.

We can now define the coproduct operator precisely.

Definition 3.10. The *coproduct* δ is the unique algebra homomorphism

$$\delta : T((V)) \to \mathcal{T}_2((V))$$

such that $\delta(\mathbf{1}) = \mathbf{1} \boxtimes \mathbf{1}$ and

$$\delta(v) = v \boxtimes 1 + 1 \boxtimes v \quad \forall v \in V.$$

The requirement that δ is an algebra homomorphism forces the unique way of defining it. It is useful to know how it acts on a generic monomial explicitly. Let $v_1 \otimes \cdots \otimes v_n \in V^{\otimes n}$. Given $I = \{i_1 < \cdots < i_k\} \subseteq \{1, \cdots, n\}$, we denote

$$v|_I \triangleq v_{i_1} \otimes \cdots \otimes v_{i_k} \in V^{\otimes k}$$

and by convention we set $v|_\emptyset \triangleq \mathbf{1}$. Since δ is a homomorphism, by the definition of the $*$-multiplication, one finds that

$$\begin{aligned} \delta(v_1 \otimes \cdots \otimes v_n) &= \delta(v_1) * \cdots * \delta(v_n) \\ &= (v_1 \boxtimes \mathbf{1} + \mathbf{1} \boxtimes v_1) * \cdots * (v_n \boxtimes \mathbf{1} + \mathbf{1} \boxtimes v_n) \\ &= \sum_I v|_I \boxtimes v|_{I^c} \in \mathcal{V}^n, \end{aligned} \tag{3.6}$$

where the summation is over all subsets $I \subseteq \{1, \cdots, n\}$ (including $\emptyset$).

In addition to the coproduct δ, we need to introduce a few more operators on these tensor algebras. As we will see, they are all naturally related with each other in the characterisations of Lie series.

(i) Let $\alpha : T((V)) \to T((V))$ be the linear operator induced by $\alpha(\mathbf{1}) \triangleq \mathbf{1}$ and

$$\alpha(v_1 \otimes \cdots \otimes v_n) \triangleq (-1)^n v_n \otimes \cdots \otimes v_1, \quad n \geqslant 1, v_i \in V.$$

Note that α is an *anti-automorphism*, i.e.

$$\alpha(\xi \otimes \eta) = \alpha(\eta) \otimes \alpha(\xi).$$

(ii) Let $D : T((V)) \to T((V))$ be the linear operator induced by $D(\mathbf{1}) \triangleq 0$ and

$$D(v_1 \otimes \cdots \otimes v_n) \triangleq n v_1 \otimes \cdots \otimes v_n.$$

Note that D is a *derivation* of the algebra $T((V))$, in the sense that

$$D(\xi \otimes \eta) = D(\xi) \otimes \eta + \xi \otimes D(\eta).$$

(iii) The *right normed bracketing* is the linear operator $R : T((V)) \to T((V))$ induced by $R(\mathbf{1}) \triangleq 0$ and

$$R(v_1 \otimes \cdots \otimes v_n) \triangleq [v_1, [v_2, [\cdots [v_{n-1}, v_n] \cdots]]]. \tag{3.7}$$

It is clear from the definition that the image of a tensor series under R is a Lie series.
(iv) We define two adjoint operators $\mathrm{ad}, \mathrm{Ad} : T((V)) \to \mathrm{End}(T((V)))$ in the following way:

$$\mathrm{ad}(\xi)(\eta) \triangleq [\xi, \eta], \quad \xi, \eta \in T((V))$$

and Ad is the unique algebra homomorphism such that $\mathrm{Ad}(v) = \mathrm{ad}(v)$ for $v \in V$. Note that Ad is well-defined: given $\xi, \eta \in T((V))$, the definition of $\pi_n(\mathrm{Ad}(\xi)(\eta))$ only involves the first $n-1$ components of ξ. In general, $\mathrm{ad} \neq \mathrm{Ad}$. For instance, given $v, w \in V$ and $\eta \in T((V))$, one has

$$\mathrm{ad}(v \otimes w)(\eta) = [v \otimes w, \eta] = v \otimes w \otimes \eta - \eta \otimes v \otimes w$$

while

$$\begin{aligned} \mathrm{Ad}(v \otimes w)(\eta) &= \mathrm{Ad}(v)\mathrm{Ad}(w)\eta = \mathrm{ad}(v)\mathrm{ad}(w)\eta = [v, [w, \eta]] \\ &= v \otimes w \otimes \eta - v \otimes \eta \otimes w - w \otimes \eta \otimes v + \eta \otimes w \otimes v. \end{aligned}$$

(v) We set $\bar{\delta} \triangleq (\mathrm{Id} \boxtimes \alpha) \circ \delta : T((V)) \to \mathcal{T}_2((V))$, where $\mathrm{Id} \boxtimes \alpha$ is the linear operator induced by

$$(\mathrm{Id} \boxtimes \alpha)(\xi \boxtimes \eta) \triangleq \xi \boxtimes \alpha(\eta).$$

(vi) Define two linear operators $\mathrm{conc}, \lambda : \mathcal{T}_2((V)) \to T((V))$ induced by

$$\mathrm{conc}(\xi \boxtimes \eta) \triangleq \xi \otimes \eta, \ \lambda(\xi \boxtimes \eta) \triangleq D(\xi) \otimes \eta$$

respectively.
(vii) Finally, we introduce one more linear operator $\mu : \mathcal{T}_2((V)) \to \mathrm{End}(T((V)))$ by

$$\mu(\xi_1 \boxtimes \xi_2)(\eta) \triangleq \xi_1 \otimes \eta \otimes \xi_2, \quad \xi_i, \eta \in T((V)).$$

Chen's theorem relates the shuffle product structure with the Lie structure, and the key ingredient for connecting them is the coproduct operator. The following lemma provides the first (easier) part of the proof. It characterises the shuffle product structure in terms of δ.

Lemma 3.11. *Let $\xi \in T_1((V))$ be a tensor series. Then ξ is group-like if and only if*

$$\delta(\xi) = \xi \boxtimes \xi. \tag{3.8}$$

Proof. Since δ is degree-preserving, the relation (3.8) is equivalent to saying that

$$\delta(\xi_n) = (\xi \boxtimes \xi)_n = \sum_{k=0}^{n} \xi_k \boxtimes \xi_{n-k}, \quad \forall n \geqslant 1.$$

In addition, one knows from (3.6) that

$$\delta(\xi_n) = \sum_{I \subseteq \{1,\cdots,n\}} \mathcal{P}_I(\xi_n),$$

where $\mathcal{P}_I : V^{\otimes n} \to V^{\otimes |I|} \boxtimes V^{\otimes (n-|I|)}$ is the linear operator induced by

$$\mathcal{P}_I(v_1 \otimes \cdots \otimes v_n) \triangleq v|_I \boxtimes v|_{I^c}.$$

As a consequence, the relation (3.8) is equivalent to saying that

$$\xi_k \boxtimes \xi_{n-k} = \sum_{\substack{I \subseteq \{1,\cdots,n\} \\ |I|=k}} \mathcal{P}_I(\xi_n) \quad \forall n \geqslant 1, k = 0, \cdots, n. \tag{3.9}$$

On the other hand, for each fixed $k \leqslant n$ there is a one-to-one correspondence between subsets $I \subseteq \{1, 2, \cdots, n\}$ of k elements and $(k, n-k)$-shuffles. Indeed, given $I = \{i_1 < \cdots < i_k\}$, one defines $\sigma \in \mathcal{S}(k, n-k)$ by mapping $\{1, \cdots, k\}$ onto $\{i_1, \cdots, i_k\}$ and mapping $\{k+1, \cdots, n\}$ onto I^c in the obvious order-preserving manner. Under this correspondence, by applying the operator conc on both sides of (3.9), one precisely obtains the shuffle product formula:

$$\xi_k \otimes \xi_{n-k} = \sum_{\substack{I \subseteq \{1,\cdots,n\} \\ |I|=k}} \mathrm{conc}\big(\mathcal{P}_I(\xi_n)\big) = \sum_{\sigma \in \mathcal{S}(k,n-k)} P_\sigma(\xi_n). \tag{3.10}$$

The equivalence between (3.10) and (3.9) follows from the simple observation that

$$\mathrm{conc}: V^{\otimes k} \boxtimes V^{\otimes(n-k)} \to V^{\otimes n}$$

is an isomorphism. $\square$

Remark 3.12. A deeper way of understanding all the above notions is to put them into the more general context of Hopf algebras. Since we only consider (weakly geometric) rough path applications, we stick to the more explicit shuffle/Lie algebraic approach rather than delving into the general theory of Hopf algebras. The reader is referred to [Gub10] for a natural generalisation of rough path theory to a non-geometric setting (branched rough paths) as well as related algebraic structures.

3.2.2 Characterisations of Lie series

In view of Lemma 3.11, the harder part of proving Chen's theorem is to show that

$$\log \xi \text{ is a Lie series} \iff \delta(\xi) = \xi \boxtimes \xi.$$

To this end, we need to develop several equivalent characterisations of Lie series. The main theorem is stated as follows. We always assume that $\dim V \geqslant 2$ and remark that the one-dimensional situation is trivial.

Theorem 3.13. *Let $\xi \in T((V))$ be a tensor series. Then the following statements are equivalent:*

(i) ξ is a Lie series;
(ii) $\mathrm{ad}(\xi) = \mathrm{Ad}(\xi)$ and $\xi_0 = 0$;
(iii) $\delta(\xi) = \xi \boxtimes \mathbf{1} + \mathbf{1} \boxtimes \xi$;
(iv) $\bar{\delta}(\xi) = \xi \boxtimes \mathbf{1} - \mathbf{1} \boxtimes \xi$;
(v) $\xi_0 = 0$ and $R(\xi) = D(\xi)$.

We prepare a few basic lemmas towards the proof of Theorem 3.13. They reveal how the various operators defined before are related with each other.

Lemma 3.14. *One has*

$$\lambda \circ \bar{\delta} = R, \ \mathrm{conc} \circ \bar{\delta} = \pi_0.$$

Proof. From the definition, one has

$$\bar{\delta}(\mathbf{1}) = (\mathrm{Id} \boxtimes \alpha) \circ \delta(\mathbf{1}) = \mathbf{1} \boxtimes \mathbf{1}.$$

As a result,

$$\lambda \circ \bar{\delta}(\mathbf{1}) = D(\mathbf{1}) \otimes \mathbf{1} = 0 = R(\mathbf{1}), \ \mathrm{conc} \circ \bar{\delta}(\mathbf{1}) = \mathbf{1} = \pi_0(\mathbf{1}).$$

We now show by induction that

$$\lambda \circ \bar{\delta}(v_1 \otimes \cdots \otimes v_n) = R(v_1 \otimes \cdots \otimes v_n), \ \mathrm{conc} \circ \bar{\delta}(v_1 \otimes \cdots \otimes v_n) = 0 \tag{3.11}$$

for all $n \geqslant 1$. The case of $n = 1$ is obvious by definition. Suppose that (3.11) is true for n. Let $v \in V$ and $\xi \in V^{\otimes n}$. We write

$$\delta(\xi) = \sum_i \xi_i \boxtimes \eta_i$$

with some tensors ξ_i, η_i. It follows that

$$\bar{\delta}(\xi) = (\mathrm{Id} \boxtimes \alpha)\big(\sum_i \xi_i \boxtimes \eta_i\big) = \sum_i \xi_i \boxtimes \alpha(\eta_i),$$

and by the induction hypothesis one has

$$\begin{aligned}
\lambda \circ \bar{\delta}(\xi) &= \sum_i D(\xi_i) \otimes \alpha(\eta_i) = R(\xi), \\
\mathrm{conc} \circ \bar{\delta}(\xi) &= \sum_i \xi_i \otimes \alpha(\eta_i) = 0.
\end{aligned}$$

Therefore,

$$\begin{aligned}
\bar{\delta}(v \otimes \xi) &= (\mathrm{Id} \boxtimes \alpha)(\delta(v) * \delta(\xi)) \\
&= (\mathrm{Id} \boxtimes \alpha)\big((v \boxtimes \mathbf{1} + \mathbf{1} \boxtimes v) * \big(\sum_i \xi_i \boxtimes \eta_i\big)\big) \\
&= \sum_i (\mathrm{Id} \boxtimes \alpha)\big((v \otimes \xi_i) \boxtimes \eta_i + \xi_i \boxtimes (v \otimes \eta_i)\big) \\
&= \sum_i \big((v \otimes \xi_i) \boxtimes \alpha(\eta_i) - \xi_i \boxtimes (\alpha(\eta_i) \otimes v)\big).
\end{aligned}$$

Recall that the operator D is a derivation. As a consequence, one has

$$
\begin{aligned}
\lambda\circ\bar{\delta}(v\otimes\xi) &= \sum_i \big(D(v\otimes\xi_i)\otimes\alpha(\eta_i) - D(\xi_i)\otimes(\alpha(\eta_i)\otimes v). \\
&= D(v)\otimes\big(\sum_i \xi_i\otimes\alpha(\eta_i)\big) + v\otimes\big(\sum_i D(\xi_i)\otimes\alpha(\eta_i)\big) \\
&\quad - \big(\sum_i D(\xi_i)\otimes\alpha(\eta_i)\big)\otimes v \\
&= v\otimes 0 + v\otimes R(\xi) - R(\xi)\otimes v \\
&= R(v\otimes\xi).
\end{aligned}
$$

Similarly,

$$
\mathrm{conc}\circ\bar{\delta}(v\otimes\xi) = \sum_i \big((v\otimes\xi_i)\otimes\alpha(\eta_i) - \xi_i\otimes(\alpha(\eta_i)\otimes v)\big) = 0.
$$

This completes the induction step for $n+1$. $\square$

Lemma 3.15. *One has*

$$
\mathrm{ad}(\xi) = \mu(\xi\boxtimes\mathbf{1} - \mathbf{1}\boxtimes\xi),\ \mathrm{Ad}(\xi) = \mu(\bar{\delta}(\xi)), \quad \forall\xi\in T((V)). \tag{3.12}
$$

In addition, μ is injective if $\dim V \geqslant 2$.

Proof. The first part of (3.12) is immediate from definition. The second part is apparently true when $\xi\in V$. Since Ad is an algebra homomorphism, it is enough to show that $\mu\circ\bar{\delta}$ is also an algebra homomorphism. By definition, it is easy to see that $\mu\circ(\mathrm{Id}\boxtimes\alpha)$ is an algebra homomorphism. Since δ is a homomorphism, so is $\mu\circ(\mathrm{Id}\boxtimes\alpha)\circ\delta = \mu\circ\bar{\delta}$.

To show the injectivity of μ, let $\{e_i : i\in\mathcal{I}\}$ be a basis of V which contains at least two linearly independent elements e_1, e_2. Let $\Xi = \sum c e_I\boxtimes e_J \in \mathcal{T}_2((V))$ be a nonzero element, where $e_I \triangleq e_{i_1}\otimes\cdots\otimes e_{i_r}$ for $I = \{i_1,\cdots,i_r\}\subseteq\mathcal{I}$. Denote $\mathcal{W}$ the set of terms $ce_I\boxtimes e_J$ in Ξ with nonzero coefficient c and minimal total degree $|I|+|J|$. Let $c^* e_{I^*}\boxtimes e_{J^*}$ be a term in $\mathcal{W}$ where the degree $|I|$ with respect to the first component is minimal. Let $N > \max\{|I| : ce_I\boxtimes e_J\in\mathcal{W}\}$. We claim that

$$
\eta \triangleq \mu(\Xi)\big(e_1^{\otimes N}\otimes e_2\big) \neq 0.
$$

Indeed, by the definition of μ, one has

$$
\eta = \sum c e_I\otimes e_1^{\otimes N}\otimes e_2\otimes e_J. \tag{3.13}
$$

The monomial $c^* e_{I^*} \otimes e_1^{\otimes N} \otimes e_2 \otimes e_{J^*}$ is a nonzero term in the expansion (3.13). Moreover, all other terms in (3.13) are different from this term. To see this, suppose that

$$e_{I^*} \otimes e_1^{\otimes N} \otimes e_2 \otimes e_{J^*} = e_I \otimes e_1^{\otimes N} \otimes e_2 \otimes e_J$$

for another pair of (I, J) in the expansion of Ξ. By comparing degrees, one knows that $e_I \otimes e_J$ is a monomial in $\mathcal{W}$. By the minimality of I^*, e_I is an extension of e_{I^*}, say $e_I = e_{I^*} \otimes e_{I'}$. It follows that

$$e_1^{\otimes N} \otimes e_2 \otimes e_{J^*} = e_{I'} \otimes e_1^N \otimes e_2 \otimes e_J.$$

By the choice of N, one concludes that $e_1^{\otimes N}$ is an extension of $e_{I'}$. Therefore, $I' = \emptyset$ and $e_I = e_{I^*}$, $e_{J^*} = e_J$. The claim (3.13) thus follows. This implies that $\mu(\Xi) \neq 0$. □

The following lemma gives a useful necessary condition for being a Lie series.

Lemma 3.16. *Let ξ be a Lie series. Then $\alpha(\xi) = -\xi$.*

Proof. It is clear that $\alpha(v) = -v$ for $v \in V$. Suppose that the claim is true for Lie polynomials ξ and η. Then

$$\begin{aligned} \alpha([\xi, \eta]) &= \alpha(\xi \otimes \eta - \eta \otimes \xi) = \alpha(\eta) \otimes \alpha(\xi) - \alpha(\xi) \otimes \alpha(\eta) \\ &= (-\eta) \otimes (-\xi) - (-\xi) \otimes (-\eta) = -[\xi, \eta]. \end{aligned}$$

By induction on the degree, the claim is true for all Lie polynomials. Since α is degree-preserving, it is true for Lie series as well. □

We are now in a position to prove Theorem 3.13.

Proof of Theorem 3.13. (i) $\Longrightarrow$ (ii). One has $\mathrm{ad} = \mathrm{Ad}$ on V. Suppose that the claim is true for Lie polynomials ξ_1, ξ_2. Then for any $\eta \in T((V))$, one has

$$\begin{aligned} \mathrm{Ad}([\xi_1, \xi_2])(\eta) &= \mathrm{Ad}(\xi_1 \otimes \xi_2 - \xi_2 \otimes \xi_1)(\eta) = \big(\mathrm{Ad}(\xi_1)\mathrm{Ad}(\xi_2) - \mathrm{Ad}(\xi_2)\mathrm{Ad}(\xi_1)\big)(\eta) \\ &= \big(\mathrm{ad}(\xi_1)\mathrm{ad}(\xi_2) - \mathrm{ad}(\xi_2)\mathrm{ad}(\xi_1)\big)(\eta) = [\xi_1, [\xi_2, \eta]] - [\xi_2, [\xi_1, \eta]] \\ &= [[\xi_1, \xi_2], \eta] = \mathrm{ad}([\xi_1, \xi_2])(\eta), \end{aligned}$$

where the second last equality follows from the Jacobi identity. As a consequence, the claim is true for all Lie polynomials. It is also true for Lie series since the operators ad, Ad are locally finite, i.e. the computation of $\pi_n(\mathrm{ad}(\xi)(\eta))$ and $\pi_n(\mathrm{Ad}(\xi)(\eta))$ only involves finitely many components in ξ, η.

(ii) $\Longrightarrow$ (iv). Let $\xi \in T_0((V))$ be such that $\mathrm{Ad}(\xi) = \mathrm{ad}(\xi)$. According to Lemma 3.15, one concludes that $\bar{\delta}(\xi) = \xi \boxtimes \mathbf{1} - \mathbf{1} \boxtimes \xi$.

(iv) $\Longrightarrow$ (v). Let $\xi \in T((V))$ be such that $\bar{\delta}(\xi) = \xi \boxtimes \mathbf{1} - \mathbf{1} \boxtimes \xi$. According to Lemma 3.14, one has

$$\pi_0(\xi) = \mathrm{conc} \circ \bar{\delta}(\xi) = \mathrm{conc}(\xi \boxtimes \mathbf{1} - \mathbf{1} \boxtimes \xi) = \xi \otimes \mathbf{1} - \mathbf{1} \otimes \xi = 0$$

and

$$R(\xi) = \lambda \circ \bar{\delta}(\xi) = \lambda(\xi \boxtimes \mathbf{1} - \mathbf{1} \boxtimes \xi) = D(\xi) \otimes \mathbf{1} - D(\mathbf{1}) \otimes \xi = D(\xi).$$

(v) $\Longrightarrow$ (i). Let $\xi = (0, \xi_1, \xi_2, \cdots) \in T_0((V))$ be such that $R(\xi) = D(\xi)$. Then

$$\xi_n = \frac{1}{n} D(\xi_n) = \frac{1}{n} R(\xi_n).$$

It is clear from the definition of R that $R(\xi_n)$ is a homogeneous Lie polynomial of degree n. So is ξ_n.

(i) $\Longrightarrow$ (iii). Observe that $\delta = (\mathrm{Id} \boxtimes \alpha) \circ \bar{\delta}$. Given a Lie series ξ, one has

$$\begin{aligned}
\delta(\xi) &= (\mathrm{Id} \boxtimes \alpha) \circ \bar{\delta}(\xi) \\
&= (\mathrm{Id} \boxtimes \alpha)(\xi \boxtimes \mathbf{1} - \mathbf{1} \boxtimes \xi) \quad \text{(by (iv))} \\
&= \xi \boxtimes \mathbf{1} + \mathbf{1} \boxtimes \xi. \quad \text{(by Lemma 3.16)}
\end{aligned}$$

(iii) $\Longrightarrow$ (v). Suppose that $\delta(\xi) = \xi \boxtimes \mathbf{1} + \mathbf{1} \boxtimes \xi$. Then one has

$$\bar{\delta}(\xi) = \xi \boxtimes \mathbf{1} + \mathbf{1} \boxtimes \alpha(\xi),$$

which implies by Lemma 3.14 that

$$\pi_0(\xi) = \xi + \alpha(\xi), \;\; R(\xi) = D(\xi).$$

The first identity further implies that $\pi_0(\xi) = 0$.

□

3.2.3 Proof of Chen's theorem

We now complete the proof of Chen's theorem. Let $\xi \in T_1((V))$. It remains to show that

$$\log \xi \text{ is a Lie series} \Longleftrightarrow \delta(\xi) = \xi \boxtimes \xi. \tag{3.14}$$

One first observes that

$$\delta(\xi) = \delta(\exp(\log \xi)) = \exp(\delta(\log \xi)),$$

since δ is an algebra homomorphism. According to Theorem 3.13, $\log \xi$ is a Lie series iff

$$\delta(\log \xi) = \log \xi \boxtimes \mathbf{1} + \mathbf{1} \boxtimes \log \xi,$$

which is equivalent to the relation that

$$\delta(\xi) = \exp(\log \xi \boxtimes \mathbf{1} + \mathbf{1} \boxtimes \log \xi). \tag{3.15}$$

Since $\log \xi \boxtimes \mathbf{1}$ and $\mathbf{1} \boxtimes \log \xi$ are commutative, (3.15) is further equivalent to

$$\delta(\xi) = \exp(\log \xi \boxtimes \mathbf{1}) * \exp(\mathbf{1} \boxtimes \log \xi) = (\xi \boxtimes \mathbf{1}) * (\mathbf{1} \boxtimes \xi) = \xi \boxtimes \xi.$$

This completes the proof of (3.14) and thus of Chen's theorem.

Corollary 3.17. *The space $G((V))$ of group-like elements is a multiplicative subgroup of $T_1((V))$.*

Proof. Let $\xi, \eta \in G((V))$. Then ξ, η satisfy the equation (3.8). Since δ is a homomorphism, one has

$$\delta(\xi \otimes \eta) = \delta(\xi) * \delta(\eta) = (\xi \boxtimes \xi) * (\eta \boxtimes \eta) = (\xi \otimes \eta) \boxtimes (\xi \otimes \eta).$$

It follows from Lemma 3.11 that $\xi \otimes \eta \in G((V))$. In addition, since

$$\xi^{-1} = e^{-\log \xi}$$

and $\log \xi$ is a Lie series, one knows that $-\log \xi$ is also a Lie series and thus $\xi^{-1} \in G((V))$ by Chen's theorem. □

3.2.4 An application: Dynkin's formula

A simple application of Chen's theorem is an explicit formula of $\log(e^v \otimes e^w)$. Such a formula was originally due to Dynkin.

Theorem 3.18 (Dynkin's formula)**.** *Let $v, w \in V$. Then*

$$\log(e^v \otimes e^w) = \sum_{n=1}^{\infty} \frac{(-1)^{n-1}}{n} \cdot \sum \frac{R(v^{\otimes p_1} \otimes w^{\otimes q_1} \otimes \cdots \otimes v^{\otimes p_n} \otimes w^{\otimes q_n})}{(p_1 + q_1 + \cdots + p_n + q_n) p_1! q_1! \cdots p_n! q_n!}, \tag{3.16}$$

where the inner summation is taken over all $p_i, q_j \geqslant 0$ such that $p_i + q_i > 0$.

Proof. By the definition (3.1) of the exponential function, one has

$$g \triangleq e^v \otimes e^w = \Big(\sum_{p=0}^{\infty} \frac{v^{\otimes p}}{p!}\Big) \otimes \Big(\sum_{q=0}^{\infty} \frac{w^{\otimes q}}{q!}\Big) \implies g_k = \sum_{p+q=k} \frac{v^{\otimes p} \otimes w^{\otimes q}}{p!q!}$$

where $g_k \triangleq \pi_k(g)$. Its logarithm is given by (cf. (3.2))

$$\log g = \sum_{n=1}^{\infty} \frac{(-1)^{n-1}}{n} (g - \mathbf{1})^{\otimes n}.$$

In particular,

$$\begin{aligned}
\pi_r(\log g) &= \sum_{n=1}^{\infty} \frac{(-1)^{n-1}}{n} \pi_r\big((g-\mathbf{1})^{\otimes n}\big) \\
&= \sum_{n=1}^{\infty} \frac{(-1)^{n-1}}{n} \sum_{\substack{k_1,\cdots,k_n \geqslant 1 \\ k_1+\cdots+k_n = r}} g_{k_1} \otimes \cdots \otimes g_{k_n} \\
&= \sum_{n=1}^{\infty} \frac{(-1)^{n-1}}{n} \sum \frac{v^{\otimes p_1} \otimes w^{\otimes q_1} \otimes \cdots \otimes v^{\otimes p_n} \otimes w^{\otimes q_n}}{p_1! q_1! \cdots p_n! q_n!},
\end{aligned}$$

where the inner summation is taken over all $p_i, q_j \geqslant 0$ such that

$$p_i + q_i > 0, \ \sum_{i=1}^{n} (p_i + q_i) = r.$$

On the other hand, one knows from Corollary 3.7 that $\log g$ is a Lie series. As a result, by applying Theorem 3.13 (v) one sees that

$$\begin{aligned}
\pi_r(\log g) &= \frac{1}{r} D(\pi_r(\log g) = \frac{1}{r} R(\pi_r(\log g)) \\
&= \frac{1}{r} \sum_{n=1}^{\infty} \frac{(-1)^{n-1}}{n} \sum \frac{R(v^{\otimes p_1} \otimes w^{\otimes q_1} \otimes \cdots \otimes v^{\otimes p_n} \otimes w^{\otimes q_n})}{p_1! q_1! \cdots p_n! q_n!}.
\end{aligned}$$

The desired formula (3.16) thus follows. □

3.3 The Baker-Campbell-Hausdorff formula

In this section, we give an independent derivation of another explicit formula of $\log(e^v \otimes e^w)$ as a Lie series without using Chen's theorem. Such a formula was originally due to Baker, Campbell and Hausdorff in a series of papers around 1900s, which appeared much earlier than Dynkin's formula (1940s). The BCH formula has a nice combinatorial nature that relates to the Bernoulli numbers in a natural way.

We continue to use the previous notation. In what follows, let v, w be two fixed, linearly independent elements in V. Without loss of generality, we assume that V is the two-dimensional space generated by v, w. Every element ξ in $V^{\otimes n}$ has a unique representation

$$\xi = \sum_{i_1,\cdots,i_n=1}^{2} c_{i_1\cdots i_n} e_{i_1} \otimes \cdots \otimes e_{i_n}$$

where $\{e_1, e_2\} \triangleq \{v, w\}$.

Our starting point is a crucial formula about derivations. Let $\mathcal{A}$ be an algebra and let $\varphi : \mathcal{A} \to \mathcal{A}$ be a homomorphism. A linear operator $\mathfrak{D} : \mathcal{A} \to \mathcal{A}$ is called a *φ-derivation*, if

$$\mathfrak{D}(xy) = \mathfrak{D}(x)\varphi(y) + \varphi(x)\mathfrak{D}(y).$$

It is simply called a *derivation* if it is a φ-derivation with $\varphi = \mathrm{Id}$.

Lemma 3.19. *Suppose that $\mathfrak{D}$ is a φ-derivation on $\mathcal{A}$. Then one has*

$$\mathfrak{D}(x^n) = \sum_{k=1}^{n} \binom{n}{k} (\mathrm{ad}\varphi(x))^{k-1}(\mathfrak{D}(x))\varphi(x)^{n-k} \quad \forall x \in \mathcal{A}, n \geqslant 1, \tag{3.17}$$

where $\mathrm{ad}(x)(y) \triangleq [x, y] \triangleq xy - yx$.

Proof. Let us denote $a \triangleq \varphi(x), b \triangleq \mathfrak{D}(x)$. We prove (3.17) by induction. The case of $n = 1$ reduces to $\mathfrak{D}(x) = \mathfrak{D}(x)$ which is trivial. Suppose that (3.17) is true for n. Since $\mathfrak{D}$ is a φ-derivation, one has

$$\begin{aligned}\mathfrak{D}(x^{n+1}) &= \mathfrak{D}(x \cdot x^n) = \mathfrak{D}(x)\varphi(x^n) + \varphi(x)\mathfrak{D}(x^n) = ba^n + a\mathfrak{D}(x^n) \\ &= ba^n + a\Big(\sum_{k=1}^{n} \binom{n}{k} (\mathrm{ad}a)^{k-1}(b)a^{n-k}\Big). \quad \text{(induction hypothesis)}\end{aligned}$$

By using the relation $au = [a, u] + ua$, one can rewrite the above equation as

$$\mathfrak{D}(x^{n+1}) = ba^n + \sum_{k=1}^{n} \binom{n}{k} (\mathrm{ad}a)^k(b)a^{n-k} + \sum_{k=1}^{n} \binom{n}{k} (\mathrm{ad}a)^{k-1}(b)a^{n-k+1}. \tag{3.18}$$

By applying a change of indices and using the fact that

$$\binom{n}{l-1}+\binom{n}{l}=\binom{n+1}{l},$$

one simplifies (3.18) to the desired equation

$$\mathfrak{D}(x^{n+1})=\sum_{l=1}^{n+1}\binom{n+1}{l}(\mathrm{ad}a)^{l-1}(b)a^{n+1-l}.$$

This completes the induction step. $\square$

Returning to the tensor algebra $T((V))$, one has the following important corollary of Lemma 3.19.

Proposition 3.20. *Let $f(t)=\sum_{n\geqslant 0}a_nt^n$ ($a_n\in\mathbb{R}$) be a formal power series. Let $\mathfrak{D}:T((V))\to T((V))$ be a φ-derivation for some homomorphism φ. Suppose that $\xi,\varphi(\xi)\in T_0((V))$. Then*

$$\mathfrak{D}(f(\xi))=\sum_{k=1}^{\infty}\frac{1}{k!}(\mathrm{ad}\varphi(\xi))^{k-1}(\mathfrak{D}(\xi))\otimes f^{(k)}(\varphi(\xi)) \tag{3.19}$$

for all $\xi\in T_0((V))$.

Remark 3.21. Since $\xi,\varphi(\xi)\in T_0((V))$, the equation (3.19) is locally finite and thus well-defined.

Proof. Since the equation (3.19) is linear in f, it suffices to consider the case when $f(t)=t^n$. But this is precisely Lemma 3.19. $\square$

Now let us write $e^v\otimes e^w=e^H$ where $H\in T_0((V))$. We rearrange the tensor series H according to the degree *relative to* v. In other words, we write

$$H=\sum_{n=1}^{\infty}H_n,$$

where $H_n\in T_0((V))$ is a tensor series in which every monomial has precisely n of v's. It is clear that $H_0=w$. The BCH formula gives an explicit way of computing H_n and in particular shows that H_n is a Lie series. We first derive the formula for H_1.

Proposition 3.22. *The tensor series H_1 is given by*

$$H_1 = v + \frac{1}{2}[v,w] + \sum_{n=1}^{\infty} \frac{B_{2n}}{(2n)!}(\mathrm{ad}w)^{2n}(v),$$

where the B_{2n}'s are the Bernoulli numbers, i.e. the Taylor coefficients of

$$\frac{x}{e^x - 1} = \sum_{m=0}^{\infty} \frac{B_m}{m!} x^m.$$

In particular, H_1 is a Lie series.

Remark 3.23. By explicit calculation, one finds that

$$B_0 = 1, B_1 = -\frac{1}{2}, B_2 = \frac{1}{6}, B_4 = -\frac{1}{30}, B_6 = \frac{1}{42}$$

and $B_{2n+1} = 0$ for all $n \geqslant 1$.

Proof. Let $\varphi : T((V)) \to T((V))$ be the algebra homomorphism induced by $\varphi(v) \triangleq 0$, $\varphi(w) \triangleq w$. Let $\mathfrak{D} : T((V)) \to T((V))$ be the linear operator such that $D = \mathrm{Id}$ on each of those basis monomials $e_{i_1} \otimes \cdots \otimes e_{i_n}$ ($e_i = v$ or w) that contains precisely one v and $D = 0$ otherwise. In other words, $\mathfrak{D}$ annihilates all terms that have no or more than one v's in them. It follows that

$$\mathfrak{D}(e^H) = \mathfrak{D}(e^v \otimes e^w) = v \otimes e^w. \tag{3.20}$$

On the other hand, it is readily checked that $\mathfrak{D}$ is a φ-derivation. According to Proposition 3.20 with $f(\xi) = e^{\xi}$,

$$\mathfrak{D}(e^H) = \sum_{k=1}^{\infty} \frac{1}{k!}(\mathrm{ad}\varphi(H))^{k-1}(\mathfrak{D}(H)) \otimes e^{\varphi(H)} =: \Big(\frac{e^{\mathrm{ad}\varphi(H)} - \mathrm{Id}}{\mathrm{ad}\varphi(H)}\Big)(\mathfrak{D}(H)) \otimes e^{\varphi(H)}.$$

By the definitions of φ and $\mathfrak{D}$, one has

$$\varphi(H) = H_0 = w,\ \mathfrak{D}(H) = H_1.$$

As a result,

$$\mathfrak{D}(e^H) = \Big(\frac{e^{\mathrm{ad}w} - \mathrm{Id}}{\mathrm{ad}w}\Big)(H_1) \otimes e^w. \tag{3.21}$$

By comparing (3.20) and (3.21), one sees that

$$v \otimes e^w = \big(\frac{e^{\mathrm{ad}w} - \mathrm{Id}}{\mathrm{ad}w}\big)(H_1) \otimes e^w \iff v = \big(\frac{e^{\mathrm{ad}w} - \mathrm{Id}}{\mathrm{ad}w}\big)(H_1). \tag{3.22}$$

Consequently,

$$H_1 = \big(\frac{\mathrm{ad}w}{e^{\mathrm{ad}w} - \mathrm{Id}}\big)(v) = v + \frac{1}{2}[v, w] + \sum_{n=1}^{\infty} \frac{B_{2n}}{(2n)!}(\mathrm{ad}w)^{2n}(v).$$

$\square$

We now derive the BCH formula (the computation of H_n) by using the Lie series H_1. Let $H_1\frac{\partial}{\partial w}$ denote the derivation induced by

$$\big(H_1\frac{\partial}{\partial w}\big)(v) \triangleq 0, \ \big(H_1\frac{\partial}{\partial w}\big)(w) \triangleq H_1.$$

Theorem 3.24. *For each $n \geqslant 1$, the tensor series H_n is given by*

$$H_n = \frac{1}{n!}\big(H_1\frac{\partial}{\partial w}\big)^n(w).$$

In addition, H_n is a Lie series.

Proof. Denote $\mathfrak{D} \triangleq H_1\frac{\partial}{\partial w}$. By applying Proposition 3.20 to the case when $\varphi = \mathrm{Id}$ and $f(\xi) = e^{\xi}$, one finds that

$$\mathfrak{D}(e^w) = \big(\frac{e^{\mathrm{ad}w} - \mathrm{Id}}{\mathrm{ad}w}\big)(\mathfrak{D}(w)) \otimes e^w = \big(\frac{e^{\mathrm{ad}w} - \mathrm{Id}}{\mathrm{ad}w}\big)(H_1) \otimes e^w = v \otimes e^w,$$

where the last equality comes from (3.22). Since $\mathfrak{D}$ is a derivation and $\mathfrak{D}(v) = 0$, by applying $\mathfrak{D}$ again, one has

$$\mathfrak{D}^2(e^w) = \mathfrak{D}(v) \otimes e^w + v \otimes \mathfrak{D}(e^w) = v \otimes v \otimes e^w = v^{\otimes 2} \otimes e^w,$$

and inductively,

$$e^{\mathfrak{D}}(e^w) = e^v \otimes e^w.$$

The next observation is that $e^{\mathfrak{D}}$ is an algebra homomorphism. Indeed, since $\mathfrak{D}$ is a derivation, one proves by induction that

$$\mathfrak{D}^n(\xi \otimes \eta) = \sum_{k=0}^{n} \binom{n}{k} \mathfrak{D}^k(\xi) \otimes \mathfrak{D}^{n-k}(\eta).$$

As a result,

$$e^{\mathfrak{D}}(\xi\otimes\eta)=\sum_{n=0}^{\infty}\frac{1}{n!}\mathfrak{D}^n(\xi\otimes\eta)=\sum_{n=0}^{\infty}\frac{1}{n!}\sum_{k=0}^{n}\binom{n}{k}\mathfrak{D}^k(\xi)\otimes\mathfrak{D}^{n-k}(\eta)$$
$$=\sum_{n=0}^{\infty}\sum_{k=0}^{n}\frac{1}{k!(n-k)!}\mathfrak{D}^k(\xi)\otimes\mathfrak{D}^{n-k}(\eta)=e^{\mathfrak{D}}(\xi)\otimes e^{\mathfrak{D}}(\eta).$$

Therefore, $e^{\mathfrak{D}}$ is a homomorphism. It follows that

$$e^{\mathfrak{D}}(e^w)=\exp(e^{\mathfrak{D}}(w))=e^v\otimes e^w=e^H \implies e^{\mathfrak{D}}(w)=H.$$

By the definition of H_1 and $\mathfrak{D}$, every application of $\mathfrak{D}$ on w increases the degree of v by one. In particular, the v-degree of the term $\frac{1}{n!}\mathfrak{D}^n(w)$ in the expansion of $e^{\mathfrak{D}}$ is precisely n. As a consequence, one concludes that

$$H_n=\frac{1}{n!}\mathfrak{D}^n(w).$$

To see that H_n is a Lie series, one observes that

$$H_1 \text{ is a Lie series, } H_{n+1}=\frac{1}{n+1}\mathfrak{D}(H_n),$$

and $\mathfrak{D}$ maps a Lie series into a Lie series. The last claim is a simple consequence of the relation

$$\mathfrak{D}([\xi,\eta])=[\mathfrak{D}(\xi),\eta]+[\xi,\mathfrak{D}(\eta)],$$

which follows from the fact that $\mathfrak{D}$ is a derivation. $\square$

Remark 3.25. The BCH formula can be concisely expressed as

$$\log(e^v\otimes e^w)=\exp\big(\big(\frac{\mathrm{ad}w}{e^{\mathrm{ad}w}-\mathrm{Id}}\big)(v)\frac{\partial}{\partial w}\big)(w).$$

Remark 3.26. One can also obtain a dual formula by expressing $\log(e^v\otimes e^w)$ relative to the degree of w. Letting H_1' be the series in H whose w-degree is one, it can be shown that

$$H_1'=\big(\frac{\mathrm{ad}v}{1-e^{-\mathrm{ad}v}}\big)(w)=w+\frac{1}{2}[v,w]+\sum_{n=1}^{\infty}\frac{B_{2n}}{(2n)!}(\mathrm{ad}v)^{2n}(w),$$

and

$$H=\exp\big(H_1'\frac{\partial}{\partial v}\big)(v).$$

The proof is left as an exercise.

3.4 The Chen-Strichartz formula

Chen's theorem asserts that the logarithm of a group-like element g is a Lie series. However, it does not provide any explicit formula of $\log g$ in terms of commutators. The BCH formula and Dynkin's formula do give explicit logarithms as Lie series, but for the special case of $e^v \otimes e^w$, or a bit more generally, for $e^{v_1} \otimes \cdots \otimes e^{v_n}$ with $v_1, \cdots, v_n \in V$. In the context of paths, these cases correspond to piecewise linear paths. It was until 1987 that Strichartz derived an explicit (and rather elegant) formula of $\log g$ as a Lie series for any group-like element g. In particular, it covers the case of weakly geometric rough paths. Strichartz's formula has several deep applications to the theory of differential equations and sub-Riemannian geometry. Due to its fundamental relation with Chen's theorem, this formula (as well as its development in the context of differential equations) is often referred to as the *Chen-Strichartz formula*.

3.4.1 The logarithm of a group-like element

We first derive the formula in the intrinsic context of group-like elements. Given a permutation $\sigma \in \mathcal{S}_n$, we set

$$e(\sigma) \triangleq \#\{j = 1, \cdots, n-1 : \sigma(j) > \sigma(j+1)\}. \tag{3.23}$$

Recall that P_σ is the tensor permutation and R is the right normed bracketing operators on $T((V))$ (cf. (1.7) and (3.7) respectively).

Theorem 3.27 (The Chen-Strichartz formula for group-like elements)**.** *Let $g = (1, g_1, g_2, \cdots)$ be a group-like element. Then*

$$\log g = \sum_{n=1}^{\infty} \sum_{\sigma \in \mathcal{S}_n} \frac{(-1)^{e(\sigma)}}{n^2 \binom{n-1}{e(\sigma)}} R(P_\sigma(g_n)). \tag{3.24}$$

Proof. The following equation is a simple generalisation of the shuffle product formula:

$$g_{p_1} \otimes \cdots \otimes g_{p_m} = \sum_{\sigma \in \mathcal{S}(p_1, \cdots, p_m)} P_\sigma(g_{p_1 + \cdots + p_m}) \quad \forall m \geqslant 2, p_1, \cdots, p_m \in \mathbb{N}, \tag{3.25}$$

where the summation is taken over all $(p_1, \cdots, p_m)$-shuffles, i.e. those $\sigma \in \mathcal{S}_{p_1 + \cdots + p_m}$ such that

$$\sigma(1) < \cdots < \sigma(p_1),\ \sigma(p_1+1) < \cdots < \sigma(p_1+p_2),\ \cdots,$$
$$\sigma(p_1 + \cdots + p_{m-1} + 1) < \cdots < \sigma(p_1 + \cdots + p_m).$$

It follows from (3.25) that

$$\log g=\sum_{k=1}^{\infty}\frac{(-1)^{k-1}}{k}(g-\mathbf{1})^{\otimes k}=\sum_{n=1}^{\infty}\sum_{m=1}^{n}\frac{(-1)^{m-1}}{m}\sum_{\substack{p_1,\cdots,p_m\geqslant 1\\ p_1+\cdots+p_m=n}}g_{p_1}\otimes\cdots\otimes g_{p_m}$$
$$=\sum_{n=1}^{\infty}\sum_{m=1}^{n}\frac{(-1)^{m-1}}{m}\sum_{p_i\geqslant 1,p_1+\cdots+p_m=n}\ \sum_{\sigma\in\mathcal{S}(p_1,\cdots,p_m)}P_\sigma(g_n)$$
$$=\sum_{n=1}^{\infty}\sum_{m=1}^{n}\sum_{\sigma\in\mathcal{S}_n}\frac{(-1)^{m-1}}{m}d(n,m,\sigma)P_\sigma(g_n),$$

where for given $n\geqslant m\geqslant 1$ and $\sigma\in\mathcal{S}_n$, we define

$$d(n,m,\sigma)\triangleq\#\{(p_1,\cdots,p_m):p_1,\cdots,p_m\geqslant 1,p_1+\cdots+p_m=n,\sigma\in\mathcal{S}(p_1,\cdots,p_m)\}.$$

Since $\log g$ is a Lie series (Chen's theorem), one knows from Theorem 3.13 (v) that

$$\pi_n(\log g)=\frac{1}{n}D(\pi_n(\log g))=\frac{1}{n}R(\pi_n(\log g)).$$

Therefore,

$$\log g=\sum_{n=1}^{\infty}\sum_{\sigma\in\mathcal{S}_n}\sum_{m=1}^{n}\frac{(-1)^{m-1}}{mn}d(n,m,\sigma)R(P_\sigma(g_n)).\tag{3.26}$$

We claim that

$$\sum_{m=1}^{n}\frac{(-1)^{m-1}}{mn}d(n,m,\sigma)=\frac{(-1)^{e(\sigma)}}{n^2\binom{n-1}{e(\sigma)}},\tag{3.27}$$

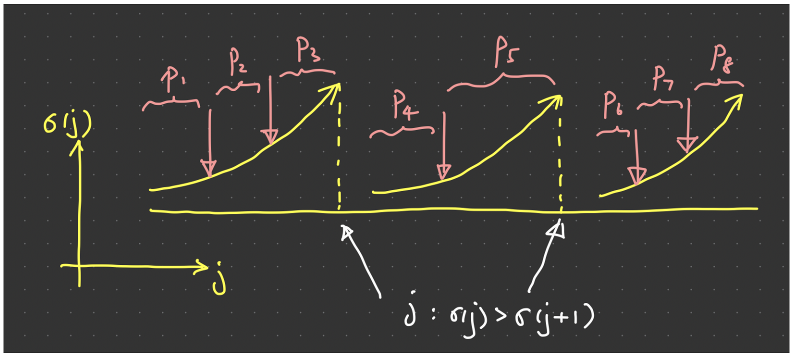
where $e(\sigma)$ is the number defined by (3.23). To prove this, one first observes that

$$d(n,m,\sigma)=\binom{n-e(\sigma)-1}{m-e(\sigma)-1},$$

which follows from a simple combinatorial argument of counting partitions: inserting $m-e(\sigma)-1$ sticks into $n-e(\sigma)-1$ positions, so that the inserted sticks together with the fixed ones

$$\{j=1,\cdots,n-1:\sigma(j)>\sigma(j+1)\}$$

give arise to a partition $(p_1,\cdots,p_m)$.



In particular, the summation on the left hand side of (3.27) indeed starts from $m = e(\sigma) + 1$. On the other hand, for any $k \geqslant 1$, $r \geqslant 0$ one has

$$\begin{aligned}\sum_{j=0}^{r} \frac{(-1)^j}{k+j}\binom{r}{j} &= \sum_{j=0}^{r}(-1)^j\binom{r}{j}\int_0^1 x^{k+j-1}dx = \int_0^1 (1-x)^r x^{k-1}dx \\ &= \frac{(k-1)!r!}{(k+r)!}. \quad \text{(Beta function)}\end{aligned}$$

By taking $r = n - e(\sigma) - 1$ and $k = e(\sigma) + 1$, one finds that the left hand side of (3.27) is equal to

$$\sum_{m=e(\sigma)+1}^{n} \frac{(-1)^{m-1}}{mn}\binom{n-e(\sigma)-1}{m-e(\sigma)-1} = \sum_{j=0}^{r}\frac{(-1)^{j+k-1}}{(k+j)n}\binom{r}{j} = \frac{(-1)^{e(\sigma)}}{n^2\binom{n-1}{e(\sigma)}}.$$

Therefore, the relation (3.27) holds.

By substituting (3.27) into (3.26), the desired formula (3.24) follows. □

Remark 3.28. If $V = \mathbb{R}^d$ with basis $\{e_1, \cdots, e_d\}$, the formula (3.26) can be rewritten as

$$\log g = \sum_{n=1}^{\infty} \sum_{i_1,\cdots,i_n=1}^{d} \Lambda_I(g) e_I, \tag{3.28}$$

where $I \triangleq (i_1, \cdots, i_n)$, the coefficient $\Lambda_I(g)$ is defined by

$$\Lambda_I(g) \triangleq \sum_{\sigma\in\mathcal{S}_n} \frac{(-1)^{e(\sigma)}}{n^2\binom{n-1}{e(\sigma)}} g_n^{i_{\sigma^{-1}(1)},\cdots,i_{\sigma^{-1}(n)}}, \tag{3.29}$$

and

$$e_I \triangleq R(e_{i_1}\otimes\cdots\otimes e_{i_n}) = [e_{i_1}, [e_{i_2}, \cdots, [e_{i_{n-1}}, e_{i_n}]\cdots]].$$

3.4.2 Solutions to nilpotent differential equations

To conclude this chapter, we give an important application of Theorem 3.27 to rough differential equations. This is a generalisation of Exercise 2.20 to the higher order case.

Consider the following RDE:

$$\begin{cases} dY_t = \sum_{i=1}^d V_i(Y_t)dX_t^i, \quad 0 \leqslant t \leqslant T; \\ Y_0 = y_0 \in \mathbb{R}^n, \end{cases} \tag{3.30}$$

where $\mathbf{X}$ is an α-Hölder weakly geometric rough path over $\mathbb{R}^d$, Y_t takes values in $\mathbb{R}^n$, and $V_1, \cdots, V_d$ are C_b^∞-vector fields. According to [Lyo98], $\mathbf{X}$ has a unique extension to an α-Hölder continuous, multiplicative functional $\mathbb{X} : \Delta_T \to T_1((\mathbb{R}^d))$. In addition, $\mathbb{X}_{s,t}$ is group-like for all $s < t$ (cf. Example 3.2).

To simplify technicalities, we assume that the vector fields $V_1, \cdots, V_d$ are *nilpotent*. More precisely, there exists $N \geqslant 1$ such that the Lie brackets (cf. (4.3))

$$V_I \triangleq [V_{i_1}, [V_{i_2}, \cdots, [V_{i_{n-1}}, V_{i_n}]]] \tag{3.31}$$

vanish identically for all $n > N$ and $1 \leqslant i_1, \cdots, i_n \leqslant d$. This avoids the consideration of convergence issues and all related series become finite.

Before stating the Chen-Strichartz formula for the solution Y_t, we need to introduce one more notation. Given a C_b^∞-vector field W on $\mathbb{R}^n$, we denote $\exp W : \mathbb{R}^n \to \mathbb{R}^n$ as the time-one map of the flow of diffeomorphisms associated with W. In other words,

$$\exp(W)(y) \triangleq z_1, \tag{3.32}$$

where $(z_t)_{0 \leqslant t \leqslant 1}$ is the unique solution to the ODE

$$\begin{cases} \dot{z}_t = W(z_t), \quad 0 \leqslant t \leqslant 1; \\ z_0 = y. \end{cases}$$

This map $\exp W$ defines an element in the group of diffeomorphisms over $\mathbb{R}^n$, which is denoted as $\mathcal{D}(\mathbb{R}^n)$.

Theorem 3.29 (The Chen-Strichartz formula for nilpotent RDEs)**.** *Suppose that the vector fields $V_1, \cdots, V_d$ are nilpotent. Then the solution Y_t (first level path) to the RDE (3.30) is given by*

$$Y_t = \exp\big(\sum_{n=1}^\infty \sum_{i_1, \cdots, i_n = 1}^d \Lambda_I(\mathbb{X}_{0,t})V_I\big)(y_0), \tag{3.33}$$

where $\Lambda_I(\mathbb{X}_{0,t})$ is defined by (3.29) with $g = \mathbb{X}_{0,t}$ and V_I is defined by (3.31).

Remark 3.30. Due to the nilpotency assumption, the series in the expression (3.33) is indeed a finite sum.

Proof. The essential point is the following algebraic consideration. Let $\mathcal{L}(V_1, \cdots, V_d)$ denote the (nilpotent) Lie algebra generated by the vector fields $\{V_1, \cdots, V_d\}$. Consider the linear map

$$F : \mathbb{R}^d \to \mathcal{L}(V_1, \cdots, V_d), \ e_i \mapsto V_i.$$

This map extends naturally to a Lie algebra homomorphism $F : \mathcal{L}((\mathbb{R}^d)) \to \mathcal{L}(V_1, \cdots, V_d)$, which is well-defined due to the nilpotency assumption. According to the Chen-Strichartz formula (3.24) for group-like elements, one has

$$\sum_{n=1}^{\infty} \sum_{i_1, \cdots, i_n=1}^{d} \Lambda_I(\mathbb{X}_{0,t}) V_I = F(\log \mathbb{X}_{0,t}).$$

Next, let us consider the compostion

$$\hat{F} : G((\mathbb{R}^d)) \to \mathcal{D}(\mathbb{R}^n), \ \hat{F} \triangleq \exp \circ F \circ \log.$$

We claim that $\hat{F}$ is a group anti-homomorphism, i.e. $\hat{F}(g_1 \otimes g_2) = \hat{F}(g_2)\hat{F}(g_1)$. To see this, let $g_1, g_2 \in G((\mathbb{R}^d))$ with logarithms l_1, l_2 respectively. One can write

$$g_1 \otimes g_2 = e^{l_1} \otimes e^{l_2} = e^{H(l_1, l_2)},$$

where $(l_1, l_2) \mapsto H(l_1, l_2)$ is the BCH functional. It is important to note that the shape of H is universal, in the sense that it does not rely on the specific Lie algebra under consideration. By the definition of $\hat{F}$, one thus has

$$\begin{aligned}\hat{F}(g_1 \otimes g_2) &= \hat{F}(e^{H(l_1,l_2)}) = \exp\big(F(H(l_1, l_2))\big) \\ &= \exp\big(H(F(l_1), F(l_2))\big) = \exp\big(F(l_2)\big) \exp\big(F(l_1)\big) = \hat{F}(g_2)\hat{F}(g_1).\end{aligned}$$

where the third identity holds since F is a Lie homomorphism and the fourth identity follows from the universality of H. Note that the order reversal is a consequence of the action convention in $\mathcal{D}(\mathbb{R}^n)$; the product $e^V e^W$ means acting by e^W first and then followed by e^V. In terms of the map $\hat{F}$, the desired formula (3.33) is equivalent to that (we only consider $t = T$)

$$Y_T = \hat{F}(\mathbb{X}_{0,T})(y_0). \tag{3.34}$$

We begin with the simplest case where $X_t = tw$ $(0 \leqslant t \leqslant T)$ is a line segment with $w = (w^1, \cdots, w^d) \in \mathbb{R}^d$. In this case, the RDE (3.33) becomes the ODE

$$\dot{Y}_t = \sum_{i=1}^d w^i V_i(Y_t),\ Y_0 = y_0$$

whose solution is given by

$$Y_T = \exp\big(\sum_{i=1}^d T w^i V_i\big)(y_0).$$

It is clear in this case that $\log(\mathbb{X}_{0,T}) = Tw$ and thus (3.34) holds. Next, suppose that X is a piecewise linear path given by the concatenation of the vectors $w_1, \cdots, w_k \in \mathbb{R}^d$. To be precise, we assume that $\dot{X}_t = w_i$ when $t \in [t_{i-1}, t_i]$ $(t_0 = 0,\ t_k = T)$. Essentially, this is still the ODE case and the solution is given by

$$Y_T = \exp\big(\sum_{i=1}^d (t_k - t_{k-1}) w_k^i V_i\big) \circ \cdots \circ \exp\big(\sum_{i=1}^d t_1 w_1^i V_i\big)(y_0).$$

It follows from the first case and the anti-homomorphism property of $\hat{F}$ that

$$Y_T = \hat{F}(\mathbb{X}_{t_{k-1},t_k}) \cdots \hat{F}(\mathbb{X}_{0,t_1})(y_0) = \hat{F}(\mathbb{X}_{0,T})(y_0).$$

Therefore, the formula (3.34) holds in this case.

To treat the general case, the key observation is that the formula (3.33) is universal. More precisely, one can write $Y_T = \Psi(\mathbb{X}_{0,T}, y_0)$, where

$$\Psi : G((\mathbb{R}^d)) \times \mathbb{R}^n \to \mathbb{R}^n,\ \Psi(g, y) \triangleq \exp\big(\sum_{m=1}^\infty \sum_{i_1,\cdots,i_m=1}^d \Lambda_I(g) V_I\big)(y).$$

Note that Ψ is well-defined since $V_1, \cdots, V_d$ are nilpotent. Suppose that X is a continuous path with bounded total variation. Since X can be approximated by piecewise linear paths (e.g. interpolation over finite partitions) under the total variation distance, by passing to the limit one obtains the same formula for X. If $\mathbf{X}$ is weakly geometric, one knows from [FV10] that $\mathbf{X}$ is geometric (with slightly sacrificed Hölder regularity), in the sense that it can be approximated by (the lifting of) paths with bounded total variation under the rough path metric. By passing to the limit again, one obtains the formula for $\mathbf{X}$. To reduce technicalities, we will not discuss the details about how the limiting procedures are justified. □

Remark 3.31. The map $\hat{F}$ constructed above is just the "Taylor expansion" map. Indeed, a routine application of Picard's iteration to the RDE (3.30) leads to a formal Taylor expansion for Y_T :

$$Y_t \simeq y_0 + \sum_{n=1}^{\infty} (V_{i_1} \cdots V_{i_n} I)(y_0) \cdot \mathbb{X}_{0,t}^{n;i_1,\cdots,i_n}.$$

The right hand side formally coincides with $\hat{F}(\mathbb{X}_{0,t})(y_0)$.

A special situation of Theorem 3.29 is when the vector fields $V_1, \cdots, V_d$ are *commutative*, i.e. $[V_i, V_j] = 0$ for all i, j. In this case, the Chen-Strichartz formula reads

$$Y_t = \exp\big(\sum_{i=1}^{d} X_t^i V_i\big)(y_0). \tag{3.35}$$

The formula (3.35) can be proved in a much simpler way in this case (cf. Exercise 2.20). This special formula indicates that RDE theory becomes trivial in the commutative case. The solution given by (3.35) is a priori well-defined without imposing any rough path structure on the driving path X. This is particularly true when dimension $d = 1$, in which case there is only one vector field V_1 in the equation (thus always commutative). In the general nilpotent case, one sees from the Chen-Strichartz formula that the solution Y_T at the end time depends on the driving path $\mathbf{X}$ through the global value $\mathbb{X}_{0,T}$ of the Lyons extension. This group-like element $\mathbb{X}_{0,T}$, known as the *signature* of $\mathbf{X}$, encodes essentially all information about the driving path $\mathbf{X}$ (cf. Section 4.3 below). As a result, the determinism of $\mathbb{X}_{0,T} \mapsto Y_T$ is not too surprising.

Example 3.32. Let $\mathbf{X}$ be an α-Hölder weakly geometric rough path over $\mathbb{R}^d$ and let $N \geqslant \lfloor 1/\alpha \rfloor$. Let $S_N(\mathbf{X})_t$ denote the Lyons extension of $\mathbf{X}$ up to level N (the first N components of $\mathbb{X}_{0,t}$). As a path in the truncated tensor algebra $T^N(\mathbb{R}^d)$, it satisfies the differential equation

$$dS_N(\mathbf{X})_t = S_N(\mathbf{X})_{0,t} \otimes d\mathbf{X}_t,\ S_N(\mathbf{X})_0 = \mathbf{1}, \tag{3.36}$$

where the tensor product is now taken in $T_1^N(\mathbb{R}^d)$. The point is that $S_N(\mathbf{X})_t$ lives on the subgroup $G^N(\mathbb{R}^d)$ as it satisfies the shuffle product formula. As a result, the RDE (3.36) should be intrinsically defined on $G^N(\mathbb{R}^d)$.

To this end, let

$$\mathfrak{g}^N(\mathbb{R}^d) \triangleq \bigoplus_{n=0}^{N} \mathcal{L}_n(\mathbb{R}^d)$$

be the space of Lie polynomials of degree at most N. According to Chen's theorem, under the exponential map (3.1) (truncated at level N), the space $\mathfrak{g}^N(\mathbb{R}^d)$ is precisely the Lie algebra of $G^N(\mathbb{R}^d)$. Given a fixed basis $\{e_1, \cdots, e_d\}$ of $\mathbb{R}^d = \mathcal{L}_1(\mathbb{R}^d) \subseteq \mathfrak{g}^N(\mathbb{R}^d)$, let U_i $(i = 1, \cdots, d)$ denote the left invariant vector field on $G^N(\mathbb{R}^d)$ associated with the e_i. Then the intrinsic differential equation for $S_N(\mathbf{X})_t$ is given by

$$\mathrm{d}S_N(\mathbf{X})_t = \sum_{i=1}^{d} U_i(S_N(\mathbf{X})_t)\mathrm{d}X_t^i,$$

or written in a more geometric form as

$$d\Gamma_t = (L_{\Gamma_t})_*(dX_t),$$

where $\Gamma_t \triangleq S_N(\mathbf{X})_t$ and $(L_g)_* : T_{\mathbf{1}}G^N(\mathbb{R}^d) \to T_g G^N(\mathbb{R}^d)$ denotes the differential map of the left translation L_g. Since $\mathfrak{g}^N(\mathbb{R}^d)$ is the Lie algebra generated by $\{e_1, \cdots, e_d\}$, it is clear that

$$\mathrm{Span}\{U_i, [U_j, U_k], \cdots [U_{i_1}, [U_{i_2}, \cdots, [U_{i_{N-1}}, U_{i_N}]]\}|_g = T_g G^N(\mathbb{R}^d)$$

for all $g \in G^N(\mathbb{R}^d)$. In other words, the vector fields satisfy Hörmander's condition at every point on $G^N(\mathbb{R}^d)$ (cf. Definition 4.6 in the below).

Remark 3.33. The formula (3.33) in the non-nilpotent case is usually applied in a truncated setting in small time (thus an approximation scheme). This is particularly useful when the underlying vector fields are assumed to be hypoelliptic and the driving path is a suitable random process. We refer the reader to [Bau04] for a more general discussion along this direction.

4 Some applications of rough path theory

In this chapter, we discuss some applications of rough path theory. We focus on conveying essential ideas and explaining strategies that are not reflected in classical ordinary/stochastic calculus. As a result, this is a highly non-technical chapter and most of the technicalities (e.g. justifications of differentiability, integrability, convergence etc.) will not be presented. Proper references are given for the fine details.

4.1 Differential equations driven by Gaussian rough paths

One of the most successful applications of rough path theory is the extension of Itô's calculus to the non-semimartingale setting. As a typical situation, we consider an SDE of the form

$$dY_t = \sum_{i=1}^{d} V_i(Y_t)dX_t^i, \quad 0 \leqslant t \leqslant T, \tag{4.1}$$

where $X = (X^1, \cdots, X^d)$ is a suitable d-dimensional Gaussian process with continuous sample paths, $V_1, \cdots, V_d$ are C_b^∞-vector fields on $\mathbb{R}^n$ and Y takes values in $\mathbb{R}^n$. A basic example to have in mind is that X is a d-dimensional *fractional Brownian motion* (fBM) with Hurst parameter $H \in (1/4, 1)$. Namely, $X^1, \cdots, X^d$ are i.i.d. Gaussian processes with mean zero and covariance function

$$R(s,t) \triangleq \mathbb{E}[X_s^1 X_t^1] = \frac{t^{2H} + s^{2H} - |t-s|^{2H}}{2}, \quad s, t \in [0, T].$$

The threshold of $H = 1/4$ will be explained later on. Note that the Brownian motion corresponds to $H = 1/2$.

Convention. To ease notation, throughout the rest we take the standard convention that repeated indices appearing in an expression are automatically summed over its range, e.g. $V_i dX^i \triangleq \sum_{i=1}^d V_i dX^i$.

4.1.1 Gaussian processes as rough paths

To make sense of the equation (4.1) from the rough path perspective, one needs to regard the driving process X as a (random) rough path in a natural way, i.e. a canonical rough path lifting $\mathbf{X}$ of the first level path X.

In what follows, we only consider the fBM case. If X is an fBM with Hurst parameter $H > 1/2$, the equation (4.1) is defined in the sense of Young (cf. Lyons

[Lyo94]) and rough path theory is not needed in this case. We now assume that $H \in (1/3, 1/2]$.

For each $m \geqslant 1$, we define a random functional

$$\Delta_T \ni (s,t) \mapsto \mathbf{X}_{s,t}^{(m)} \in G^2(\mathbb{R}^d)$$

in the following way. Let $X^{(m)}$ be the piecewise linear interpolation of X over the m-th order dyadic partition of $[0, T]$, namely

$$X_{kT/2^m}^{(m)} = X_{kT/2^m},\ k = 0, 1, \cdots, 2^m$$

and $X_t^{(m)}$ is linear on each dyadic interval $[kT/2^m, (k+1)T/2^m]$. Since $X^{(m)}$ is piecewise linear, one can define

$$\mathbf{X}_{s,t}^{(m)} \triangleq (1, X_t^{(m)} - X_s^{(m)}, \int_{s<u<v<t} dX_u^{(m)} \otimes dX_v^{(m)})$$

in the classical sense. The following result, which was originally due to Coutin-Qian [CQ02], defines an a.s. rough path lifting of X. We will not discuss its technical proof here.

Theorem 4.1. *Let $\alpha \in (1/3, H)$ be fixed. With probability one, the sequence $\{\mathbf{X}^{(m)} : m \geqslant 1\}$ of random α-Hölder rough paths is Cauchy with respect to the rough path metric ρ_α defined by (1.14). As a result, it has an a.s. limit $\mathbf{X}$ in the space of α-Hölder rough paths.*

If $H \in (1/4, 1/3]$, one needs to take $\alpha \in (1/4, H)$. In this case, $X^{(m)}$ should be lifted up to the third level and the above theorem remains valid. However, if $H \leqslant 1/4$, it was shown by Coutin-Qian in the same paper that $\mathbf{X}^{(m)}$ no longer converges in a reasonable sense. How one can define the SDE (4.1) in a natural way in this case remains an open problem.

By taking the lifting of X given by Theorem 4.1, one can now interpret (4.1) as an RDE driven by $\mathbf{X}$ in the pathwise sense under the framework of Chapter 2.

Example 4.2. In the Brownian motion case ($H = 1/2$), the second level lifting $X_{s,t}^{2;i,j}$ coincides with the Stratonovich integral $\int_s^t X_{s,v}^i \circ dX_v^j$ (the (i,j)-superscript means extracting coordinate components). When $i \neq j$, this integral is the same as the Itô integral due to the independence of X^i, X^j. In general, the rough integral $\int F(\mathbf{X})d\mathbf{X}$ constructed in the sense of Exercise 2.8 for this case coincides with the Stratonovich integral. The RDE (4.1) coincides with the Stratonovich equation, which is equivalent to the following Itô type SDE:

$$dY_t = V_i(Y_t)dX_t^i + \frac{1}{2}\frac{\partial V_i}{\partial y^j}V_i^j(Y_t)dt.$$

This is a direct consequence of the Wong-Zakai approximation theorem (cf. Wong-Zakai [WZ65]) and the universal limit theorem in Section 2.3.

Remark 4.3. There is a more robust way of constructing canonical liftings of Gaussian processes based on regularity properties of the covariance function. We refer the reader to [FV10] for a discussion which also reveals the fundamental barrier of $H = 1/4$ in a deeper way.

After making proper sense of the differential equation, many interesting questions related to properties of the solution can be raised (e.g. ergodicity, small time asymptotics, tail behaviour, density properties, large deviations etc.). We discuss one of the most extensively studied problems: existence of density for the solution.

4.1.2 Existence of density for hypoelliptic differential equations

In the 1970s, Malliavin discovered an elegant probabilistic proof of a renowned theorem of Hörmander in PDE theory. The core of Malliavin's work lies in showing that, when the vector fields of the SDE (4.1) satisfy a hypoellipticity condition (cf. Definition 4.6 below), the solution Y_t admits a smooth density function with respect to the Lebesgue measure on $\mathbb{R}^n$. In Malliavin's setting, the driving process X is a Brownian motion. For proving his theorem, Malliavin developed a power machinery of differential calculus on path space, known as *the Malliavin calculus*, which has led to far-reaching applications to a wide range of problems in stochastic analysis.

The extension of Malliavin's theorem to the situation when the driving process is a Gaussian rough path has only been extensively studied over the last decade with the aid of rough path theory. Before that, the main challenge was the lack of a proper stochastic integration theory for general Gaussian processes. Rough path theory provides robust analytic tools for the differential calculus of such equations. In particular, it produces deeper insight even back into the original theorem of Malliavin in the diffusion case.

The Hörmander condition and the main theorem

We begin by introducing the key assumption: the Hörmander condition. We fix the SDE (4.1) with initial condition $Y_0 = y_0 \in \mathbb{R}^n$. For simplicity, we assume that X is a d-dimensional fBM with Hurst parameter $H > 1/4$. With the canonical lifting $\mathbf{X}$ of X given by Theorem 4.1, the SDE is defined and solved pathwisely under the framework of Section 2.3.

The definition of the Hörmander condition requires the notion of Lie brackets which we now recall. A *smooth vector field* on $\mathbb{R}^n$ is a smooth function $V : \mathbb{R}^n \to \mathbb{R}^n$ (assigning a vector to each point in space in a smooth manner). It can be equivalently viewed as a differential operator:

$$V = V^i \frac{\partial}{\partial y^i} : C^\infty(\mathbb{R}^n) \to C^\infty(\mathbb{R}^n). \tag{4.2}$$

where V^i is the i-th component of V. We often write $V = (V^1, \cdots, V^n)^T$ as a column vector and denote DV as the $n \times n$ matrix whose (i, j)-entry is $\frac{\partial V^i}{\partial y^j}$.

Definition 4.4. Let V, W be two smooth vector fields on $\mathbb{R}^n$. The *Lie bracket* of V and W is the vector field defined by

$$[V, W] \triangleq DW \cdot V - DV \cdot W, \tag{4.3}$$

where $\cdot$ denotes matrix multiplication.

Remark 4.5. It is easily seen that

$$[V, W]f = V(Wf) - W(Vf), \quad \forall f \in C^\infty(\mathbb{R}^n)$$

when vector fields are viewed as differential operators by (4.2).

We are now able to define the Hörmander condition and state the main theorem.

Definition 4.6. Let $\mathcal{V} = \{V_1, \cdots, V_d\}$ be a family of smooth vector fields on $\mathbb{R}^n$. We say that $\mathcal{V}$ satisfies *the Hörmander condition* at a given point $y_0 \in \mathbb{R}^n$ (or $\mathcal{V}$ is *hypoelliptic* at y_0), if the following family of vectors

$$V_i(y_0),\ [V_i, V_j](y_0),\ [V_i, [V_j, V_k]](y_0),\ [V_i, [V_j, [V_k, V_l]]](y_0),\ \cdots$$

linearly span the whole space $\mathbb{R}^n$.

Theorem 4.7. *Consider the RDE (4.1) starting at $y_0 \in \mathbb{R}^n$, where X is a d-dimensional fBM with Hurst parameter $H > 1/4$, the vector fields are C_b^∞ and satisfy the Hörmander condition at y_0. Then for each $t > 0$, the solution Y_t admits a smooth density function with respect to the Lebesgue measure on $\mathbb{R}^n$.*

Example 4.8. Let $B_t = (X_t, Y_t)$ be a two-dimensional fBM and define

$$Z_t \triangleq \frac{1}{2}\int_0^t X_s dY_t - Y_s dX_s.$$

Note that Z_t is well-defined as it comes from the second level lifting of B_t. The triple (X_t, Y_t, Z_t) satisfies the following RDE on $\mathbb{R}^3$:

$$dX_t = dX_t, dY_t = dY_t, dZ_t = -\frac{1}{2}Y_t dX_t + \frac{1}{2}X_t dY_t.$$

The associated two vector fields are given by

$$V_1(x,y,z) = \begin{pmatrix} 1 \\ 0 \\ -y/2 \end{pmatrix}, \ V_2(x,y,z) = \begin{pmatrix} 0 \\ 1 \\ x/2 \end{pmatrix}, \quad (x,y,z) \in \mathbb{R}^3.$$

One finds that $[V_1, V_2] = \begin{pmatrix} 0 \\ 0 \\ 1 \end{pmatrix}$. Therefore, $\{V_1, V_2\}$ satisfies the Hörmander condition at every point in $\mathbb{R}^3$. On the other hand, let $S(B)_{0,t} \in T^2(\mathbb{R}^2)$ denote the canonical lifting of B as a rough path ($H \in (1/3, 1/2]$). Then $S(B)_{0,t}$ satisfies the following (linear) RDE in $T^2(\mathbb{R}^2)$:

$$\begin{cases} dS(B)_{0,t} = S(B)_{0,t} \otimes dB_t, & t \in [0,T]; \\ S(B)_{0,0} = \mathbf{1}. \end{cases}$$

By putting it into a more standard form of (4.1) with state space $\mathbb{R}^2 \otimes (\mathbb{R}^2)^{\otimes 2}$, the resulting equation does *not* satisfy the Hörmander condition at any point. The process (X_t, Y_t, Z_t) is essentially equal to $\log S(B)_{0,t}$ (cf. Exercise 1.22 (iii) or Chen's theorem). More generally, given an α-Hölder weakly geometric rough path $\mathbf{X}$ over $\mathbb{R}^d$ ($\alpha \in (0,1]$, $d \geqslant 2$), the differential equation

$$d\mathbf{X}_{0,t} = \mathbf{X}_{0,t} \otimes d\mathbf{X}_t$$

does not satisfy the Hörmander condition over $T_1^N(\mathbb{R}^d)$ ($N = \lfloor 1/\alpha \rfloor$) while the associated differential equation for $\log \mathbf{X}_{0,t}$ over $\log G^N(\mathbb{R}^d)$ does (cf. Example 3.32). The point is that a weakly geometric rough path intrinsically lives on the lower dimensional submanifold $G^N(\mathbb{R}^d)$.

Theorem 4.7 holds for a wide class of driving Gaussian rough paths under suitable conditions. The existence of density was first established by Cass-Friz [CF10] and smoothness was later obtained by Cass-Litterer-Hairer-Tindel [CHLT15]. In what follows, we explain the main strategy for the existence of density part in a formal way. The monograph [FH14] contains the complete details for the proof.

A geometric interpretation

Before discussing the actual proof, we take a detour to describe the heuristics behind the Hörmander condition. It ensures that the solution Y_t is "diffusive", in the sense that it is able to travel along any arbitrary direction starting at y_0.

To elaborate this point, it is helpful to first re-interpret the relevant concepts in a geometric setting. Let M be a differentiable manifold. For one who has not seen this notion before, just think of a surface in $\mathbb{R}^3$ (e.g. the unit sphere). A vector field V on M is a mapping $y \mapsto V(y)$ assigning to every point $y \in M$ a vector $V(y)$ that is tangential to M at y. As before, a vector field V can be equivalently viewed as a differential operator acting on smooth functions on M:

$$(Vf)(y) \triangleq \text{directional derivative of } f \text{ along the direction } V(y) \text{ at } y.$$

The Lie bracket of two smooth vector fields V, W is defined by

$$[V, W] \triangleq VW - WV : C^\infty(M) \to C^\infty(M)$$

from the perspective of differential operators. It is known that $[V, W]$ is indeed a vector field.

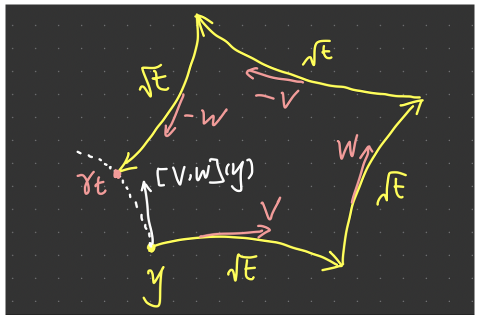
The Lie bracket $[V, W]$ has the following geometric interpretation. First of all, a vector field V determines a flow of (local) diffeomorphisms $\Phi_t^V : M \to M$ defined by the ODE

$$\frac{d\Phi_t^V(y)}{dt} = V(\Phi_t^V(y)), \ \Phi_0^V(y) = y.$$

Heuristically, $t \mapsto \Phi_t^V(y)$ carries the initial point $y \in M$ to flow on the manifold M along the vector field V. The direction of $[V, W]$ at any given location y is determined by

$$[V, W](y) = \frac{d}{dt}\Big|_{t=0} \Phi_{\sqrt{t}}^{-W} \circ \Phi_{\sqrt{t}}^{-V} \circ \Phi_{\sqrt{t}}^{W} \circ \Phi_{\sqrt{t}}^{V}(y).$$

In other words, to determine the direction of $[V, W](y)$, starting from y one first flow along V for $\sqrt{t}$ amount of time, then flow along W, and then along $-V$ and finally along $-W$ all for the same amount of time. Denoting γ_t as the endpoint of this dynamics, the direction $[V,W](y)$ is represented by the tangent vector of this curve γ_t at $t=0$.



Suppose that $V_1,\cdots,V_d$ are d smooth vector fields on M. Consider the RDE (4.1) where X is an fBM in $\mathbb{R}^d$ (canonically lifted as rough paths by Theorem 4.1). Since rough path theory is consistent with ordinary calculus, the RDE (4.1) is intrinsically well-defined on the manifold (i.e. the solution Y_t lives on the manifold M for all time). The Hörmander condition at y_0 asserts that the family of vectors

$$V_i(y_0),\ [V_i,V_j](y_0),\ [V_i,[V_j,V_k]](y_0),\ [V_i,[V_j,[V_k,V_l]]](y_0),\ \cdots$$

linearly span the tangent space of M at y_0.

We now explain in a heuristic way how the Hörmander condition enables the solution Y_t to explore all tangential directions starting at y_0. First of all, in an infinitely small amount of time one has the approximation

$$\delta Y\approx V_i(y_0)\cdot\delta X^i.$$

Since X is non-degenerate (i.e. $(\delta X^1,\cdots,\delta X^d)$ achieves all possible values in $\mathbb{R}^d$), the solution is able to explore all directions in $\mathcal{L}^1\triangleq\mathrm{Span}\{V_1,\cdots,V_d\}$. Given $W_1,W_2\in\mathcal{L}^1$, for the same reason, infinitesimally the solution can also explore the square flow

$$W_1\rightarrow W_2\rightarrow -W_1\rightarrow -W_2,$$

thus travelling along the direction $[W_1,W_2](y_0)$. By continuity, this is true in a small neighbourhood of y_0 (i.e. starting at any location y near y_0, the solution is able to travel along $[W_1,W_2](y)$). For the same reason again, the solution is able to explore the direction

$$[W_1,[W_2,W_3]](y_0)\triangleq\big[W_1\rightarrow[W_2,W_3]\rightarrow -W_1\rightarrow -[W_2,W_3]\big]$$

for any $W_1, W_2, W_3 \in \mathcal{L}^1$. One can now argue inductively to see that, starting at y_0, in an infinitely small amount of time the solution is able to explore any arbitrary direction from

$$\mathrm{Span}\{V_i(y_0),\ [V_i, V_j](y_0),\ [V_i, [V_j, V_k]](y_0),\ [V_i, [V_j, [V_k, V_l]]](y_0),\ \cdots\}.$$

As a consequence of the Hörmander condition, the solution is thus able to travel along all directions tangential to M at y_0.

The above discussion is general but rather vague. It is only based on two fundamental properties: the Hörmander condition and a suitable notion of non-degeneracy for the driving process X. From this perspective, the Gaussian nature of X (and thus the machinery of the Malliavin calculus) does not seem to play a significant role. However, I am not sure if it is possible to produce a purely analytic proof of Theorem 4.7 in a general setting that takes the spirit of the above heuristics as a starting point.

The Malliavin calculus

Returning to the main course of proving Theorem 4.7, we first explain the fundamental ideas behind the Malliavin calculus for establishing existence and smoothness of density in general. The reader is referred to Nualart [Nua06] or Shigekawa [Shi04] for an excellent introduction to the subject.

Let X be an fBM realised on the canonical path space $(\mathcal{W}, \mathcal{B}(\mathcal{W}), \mu)$. In other words, $\mathcal{W}$ is the space of continuous paths $w : [0, T] \to \mathbb{R}^d$ with $w_0 = 0$. $\mathcal{B}(\mathcal{W})$ is the Borel σ-algebra with respect to the uniform topology. μ is the unique probability measure over $\mathcal{W}$ under which the coordinate process $X_t(w) \triangleq w_t$ becomes an fBM.

A core structure in the Malliavin calculus is an embedded Hilbert space defined as follows. Let $\mathcal{C}_1$ denote the closed subspace of $L^2(\mathcal{W}, \mu)$ generated by $\{X_t^i : 0 \leqslant t \leqslant T, 1 \leqslant i \leqslant d\}$.

Definition 4.9. The *Cameron-Martin subspace* of $(\mathcal{W}, \mathcal{B}(\mathcal{W}), \mu)$, denoted as $\mathcal{H}$, is the subspace of elements $h \in \mathcal{W}$ that can be expressed as

$$h_t = \mathbb{E}[ZX_t], \quad 0 \leqslant t \leqslant T,$$

where $Z \in \mathcal{C}_1$. This is a Hilbert space with respect to the inner product

$$\langle h_1, h_2 \rangle \triangleq \mathbb{E}[Z_1 Z_2], \quad h_1, h_2 \in \mathcal{H},$$

where $Z_i \in \mathcal{C}_1$ is the (unique) random variable associated with the path h_i ($i = 1, 2$).

Given a random variable $F : \mathcal{W} \to \mathbb{R}$ that satisfies a proper notion of differentiability, one can define its derivative $DF : \mathcal{W} \to \mathcal{H}^* \cong \mathcal{H}$ by

$$D_h F(w) \triangleq \langle DF(w), h \rangle_{\mathcal{H}} \triangleq \frac{d}{d\varepsilon}\Big|_{\varepsilon=0} F(w + \varepsilon h), \quad \forall h \in \mathcal{H}$$

In other words, $D_h F(w)$ is the directional derivative of F at w along the direction h and $DF(w)$ is viewed as an element in $\mathcal{H}$ through the Riesz representation theorem.

Definition 4.10. Let $F = (F^1, \cdots, F^n) : \mathcal{W} \to \mathbb{R}^n$ be a smooth random vector. The *Malliavin covariance matrix* of F is the $n \times n$ random matrix defined by

$$\gamma_F = (\gamma_F^{ij})_{1 \leqslant i,j \leqslant n} : \; \gamma_F^{ij} \triangleq \langle DF^i, DF^j \rangle_{\mathcal{H}}.$$

The following result is a fundamental theorem in the Malliavin calculus. It provides a neat sufficient condition for the existence and smoothness of density.

Theorem 4.11. *Let $F = (F^1, \cdots, F^n) : \mathcal{W} \to \mathbb{R}^n$ be a smooth random vector. If the Malliavin covariance matrix γ_F is a.s. invertible, then F admits a density function f with respect to the Lebesgue measure $\mathbb{R}^n$. If γ_F further satisfies*

$$\det \gamma_F^{-1} \in L^p(\mathcal{W}, \mu) \quad \forall p > 1,$$

then the density function f is smooth in $\mathbb{R}^n$.

Sketch of proof. Let $\varphi : \mathbb{R}^n \to \mathbb{R}$ be an arbitrary function. By the chain rule of differentiation, one has

$$D\varphi(F) = \partial_i \varphi(F) DF^i,$$

where $\partial_i \varphi \triangleq \frac{\partial \varphi}{\partial x^i}$. It follows that

$$\langle D\varphi(F), DF^j \rangle_{\mathcal{H}} = \partial_i \varphi(F) \gamma_F^{ij} \quad \forall j = 1, \cdots, n. \tag{4.4}$$

The invertibility of γ_F allows one to deduce from (4.4) that

$$\partial_i \varphi(F) = \langle D\varphi(F), (\gamma_F^{-1})^{ij} DF^j \rangle_{\mathcal{H}}.$$

As a consequence, for any random variable G one has

$$\begin{aligned} \mathbb{E}[\partial_i \varphi(F) \cdot G] &= \mathbb{E}\big[\langle D\varphi(F), G(\gamma_F^{-1})^{ij} DF^j \rangle_{\mathcal{H}}\big] \\ &= \mathbb{E}\big[\varphi(F) H_i(G)\big], \end{aligned} \tag{4.5}$$

where

$$H_i(G) \triangleq D^*\big(G(\gamma_F^{-1})^{ij} DF^j\big)$$

and D^* denotes the adjoint of the differential operator D, i.e.

$$\mathbb{E}[\langle D\Phi, \Psi\rangle_{\mathcal{H}}] = \mathbb{E}[\Phi D^* \Psi].$$

for any smooth real valued Φ and $\mathcal{H}$-valued Ψ.

One can iterate the integration by parts formula (4.5) to see that

$$\mathbb{E}[\partial^2_{ij}\varphi(F)\cdot G] = \mathbb{E}[\partial_i\varphi(F)\cdot H_j(G)] = \mathbb{E}[\varphi(F)\cdot H_i(H_j(G))] \quad \forall \text{smooth } G.$$

In particular, by iterating this for n times with $G = 1$, one arrives at

$$\mathbb{E}[\partial^n_{12\cdots n}\varphi(F)] = \mathbb{E}[\varphi(F)\cdot R_n], \tag{4.6}$$

where $R_n \triangleq H_1(H_2(\cdots H_n(1)))$. Note that (4.6) is true for all φ.

Now let $\psi \in C_c^\infty(\mathbb{R}^n)$ be an arbitrary function and define

$$\varphi(x^1, \cdots, x^n) \triangleq \int_{-\infty}^{x^1} \cdots \int_{-\infty}^{x^n} \psi(y)dy.$$

By applying (4.6) to this φ, one obtains that

$$\mathbb{E}[\psi(F)] = \mathbb{E}\big[\big(\int_{-\infty}^{F^1} \cdots \int_{-\infty}^{F^n} \psi(y)dy\big)R_n = \int_{\mathbb{R}^n} \mathbb{E}[\mathbf{1}_{\{F>y\}}R_n]\psi(y)dy,$$

where $\{F > y\} \triangleq \{F^i > y^i \ \forall i = 1, \cdots, n\}$. This formula not only suggests that F has a density function f, but also that

$$f(x) = \mathbb{E}[\mathbf{1}_{\{F>x\}}R_n].$$

In a similar way, by iterating (4.5) to an arbitrary order one can write down a formula for any partial derivative of f, which in particular yields the smoothness of f.

$\square$

Remark 4.12. The justification of the integration by parts formula (4.5) requires suitable L^p-integrability of γ_F^{-1} as well as certain derivatives of G, F (to ensure that $G(\gamma_F^{-1})^{ij}DF^j \in \mathrm{Dom}(D^*)$ and the expectation is finite). It is true that

$$F \text{ differentiable and } \gamma_F \text{ a.s. invertible} \implies \text{existence of density.}$$

However, the above argument does not directly reflect this (cf. Nualart [Nua06] for a precise proof).

Calculus of variations for RDEs

In order to apply Theorem 4.11 to the RDE solution Y_t, we need to perform a series of (formal) differential calculus for the equation. There are two essential ingredients for this part:

(i) a relation between the Malliavin covariance matrix of Y_t and the Jacobian of the RDE (4.1) (cf. Lemma 4.13 below);
(ii) the appearance of Lie brackets in the equation of pulling back vector fields by the Jacobian (cf. Lemma 4.14 below).

We first introduce some basic definitions. Given $s \leqslant t$, we use $U^X_{t\leftarrow s} : \mathbb{R}^n \to \mathbb{R}^n$ to denote the (random) *flow of diffeomorphisms* determined by the RDE. In other words, $U^{X(w)}_{t\leftarrow s}(y_0)$ is the solution at time t of the RDE (4.1) driven by the sample path $X(w)$ (lifted as a rough path) with initial condition y_0 at starting time s. If $s > t$, we set $U^X_{t\leftarrow s} \triangleq (U^X_{s\leftarrow t})^{-1}$. The *Jacobian* of the RDE at y_0, denoted as $J^{X,y_0}_{t\leftarrow s}$, is the (random) linear transformation on $\mathbb{R}^n$ defined by

$$J^{X,y_0}_{t\leftarrow s}\xi \triangleq \frac{d}{d\varepsilon}\Big|_{\varepsilon=0} U^X_{t\leftarrow s}(y_0 + \varepsilon\xi), \quad \xi \in \mathbb{R}^n.$$

We also regard $J^{X,y_0}_{t\leftarrow s}$ as an $n \times n$ matrix. Geometrically, $J^{X,y_0}_{t\leftarrow s}$ is the push-forward mapping which carries a tangent vector at y_0 to a tangent vector at $U^{X(w)}_{t\leftarrow s}(y_0)$ along the solution path by the flow of diffeomorphisms. By differentiating the RDE (4.1) with respect to the initial condition y_0, it is easily seen that the Jacobian $J^{X,y_0}_{t\leftarrow s}$ satisfies the following (homogeneous) linear RDE:

$$\begin{cases} dJ^{X,y_0}_{t\leftarrow s} = DV_i(Y^s_t)J^{X,y_0}_{t\leftarrow s}dX^i_t, & t > s; \\ J^{X,y_0}_{s\leftarrow s} = \mathrm{Id}, \end{cases} \tag{4.7}$$

where $Y^s_t \triangleq U^X_{t\leftarrow s}(y_0)$. In addition, the uniqueness of solutions for the RDE (4.1) implies that

$$J^{X,y_0}_{u\leftarrow s} = J^{X,Y^s_t}_{u\leftarrow t} \cdot J^{X,y_0}_{t\leftarrow s} \quad \forall s, t, u \in [0, T]. \tag{4.8}$$

To simply notation we will omit the superscript letters on the Jacobian.

We are now ready to establish two key lemmas related to the aforementioned Points (i) and (ii). The first lemma provides a neat relation between the Malliavin covariance matrix γ_{Y_t} and the Jacobian $J_{t\leftarrow s}$, where Y_t is the solution to the RDE (4.1).

Lemma 4.13. *For any $\xi \in \mathbb{R}^n$ (viewed as a column vector), one has*

$$\xi^T \gamma_{Y_t} \xi = \sup_{h \in \mathcal{H}: \|h\|_{\mathcal{H}}=1} \left| \xi^T \int_0^t J_{t\leftarrow s} V_i(Y_s) dh_s^i \right|^2. \tag{4.9}$$

Proof. Let $h \in \mathcal{H}$ be a given Cameron-Martin path (cf. Definition 4.9). For each $\varepsilon > 0$, one has

$$dU_{t\leftarrow 0}^{X(w+\varepsilon h)} = V_i(U_{t\leftarrow 0}^{X(w+\varepsilon h)}) d(X_t^i + \varepsilon h_t^i).$$

By differentiation with respect to ε, one finds that the Malliavin derivative $D_h Y_t$ satisfies the following (inhomogeneous) RDE:

$$\begin{cases} dD_h Y_t = DV_i(Y_t) D_h Y_t dX_t^i + V_i(Y_t) dh_t^i, \\ D_h Y_0 = 0. \end{cases}$$

In view of the Jacobian equation (4.7), a simple variation of constants argument yields that

$$D_h Y_t = \int_0^t J_{t\leftarrow s} V_i(Y_s) dh_s^i. \tag{4.10}$$

The relation (4.9) is a consequence of (4.10) as well as the following duality relation:

$$\xi^T \gamma_{Y_t} \xi = \|\xi^T DY_t\|_{\mathcal{H}}^2 = \sup_{h \in \mathcal{H}: \|h\|_{\mathcal{H}}=1} \left| \xi^T D_h Y_t \right|^2.$$

□

Let W an arbitrary smooth vector field on $\mathbb{R}^n$. Recall that $J_{0\leftarrow t} W(Y_t)$ is the pull-back of W to the initial location y_0 by the flow of diffeomorphisms. The second lemma reveals how the Lie brackets arise naturally when one considers the differential equation for $J_{0\leftarrow t} W(Y_t)$. Recall that $Y_0 = y_0$.

Lemma 4.14. *The pull-back process $t \mapsto J_{0\leftarrow t} W(Y_t)$ satisfies the following equation:*

$$J_{0\leftarrow t} W(Y_t) - W(y_0) = \int_0^t J_{0\leftarrow s} [V_i, W](Y_s) dX_s^i.$$

Proof. Note that $J_{0\leftarrow t} = (J_{t\leftarrow 0})^{-1}$. According to (4.7), one sees that $J_{0\leftarrow t}$ satisfies the following equation:

$$dJ_{0\leftarrow t} = -J_{t\leftarrow 0} DV_i(Y_t) dX_t^i.$$

On the other hand, by the RDE of Y_t and the chain rule, one sees that

$$dW(Y_t) = DW(Y_t)V_i(Y_t)dX_t^i.$$

Therefore,

$$\begin{aligned} d\big(J_{0\leftarrow t}W(Y_t)\big) &= dJ_{0\leftarrow t}\cdot W(Y_t) + J_{0\leftarrow t}\cdot dW(Y_t) \\ &= -J_{0\leftarrow t}DV_i(Y_t)W(Y_t)dX_t^i + J_{0\leftarrow t}DW(Y_t)V_i(Y_t)dX_t^i \\ &= J_{0\leftarrow t}[V_i, W](Y_t)dX_t^i. \end{aligned}$$

□

The sketched proof of Theorem 4.7

Let $t>0$ be given fixed. We now sketch the proof of the existence of density for the solution Y_t under the Hörmander condition at y_0 (cf. Definition 4.6). In view of Theorem 4.11, it boils down to showing that γ_{Y_t} is invertible a.s. Note that γ_{Y_t} is symmetric and non-negative definite. As a result,

$$\gamma_{Y_t} \text{ invertible} \iff \xi^T\gamma_{Y_t}\xi > 0\ \forall \xi\in\mathbb{R}^n\backslash\{0\}.$$

Suppose on the contrary that, at a given fixed sample path $w\in\mathcal{W}$,

$$\xi^T\gamma_{Y_t}\xi = 0 \quad \text{for some non-zero vector } \xi.$$

According to Lemma 4.13, one has

$$\xi^T\int_0^t J_{t\leftarrow s}V_i(Y_s)dh_s^i = 0 \quad \forall h\in\mathcal{H}.$$

By using the relation (4.8), the above property can be rewritten as

$$\int_0^t \eta^T J_{0\leftarrow s}V_i(Y_s)dh_s^i = 0 \quad \forall h\in\mathcal{H}. \tag{4.11}$$

where $\eta^T \triangleq \xi^T J_{t\leftarrow 0}$. Note that $\eta\neq 0$ as $J_{t\leftarrow 0}$ is invertible. In the fBM case, it is known that (cf. Boedihardjo-Geng [BG15]) the Cameron-Martin subspace $\mathcal{H}$ contains all smooth paths in $\mathcal{W}$. In particular, the property (4.11) is sufficient to conclude that the integrand

$$\eta^T J_{0\leftarrow s}V_i(Y_s) = 0 \quad \forall s\in[0,t],\ i=1,\cdots,d. \tag{4.12}$$

As a consequence of Lemma 4.14, the property (4.12) implies that

$$\int_0^s \eta^T J_{0\leftarrow u}[V_j, V_i](Y_u)dX_u^j = 0 \quad \forall s \in [0,t],\ i = 1, \cdots, d. \tag{4.13}$$

We claim that the integrand

$$\eta^T J_{0\leftarrow u}[V_j, V_i](Y_u) = 0 \quad \forall u \in [0,t],\ i,j = 1, \cdots, d. \tag{4.14}$$

Indeed, by viewing the integral

$$I_s^0 \triangleq \int_0^s \eta^T J_{0\leftarrow u}[V_j, V_i](Y_u)dX_u^j$$

as a rough integral in the sense of Section 2.2, one has

$$0 = I_{u,v}^0 = I_u^1 X_{u,v} + \mathcal{R}\mathcal{I}_{u,v} \implies |I_u^1 X_{u,v}| = O(|v-u|^{2\alpha}), \tag{4.15}$$

where

$$I_u^1 \triangleq \big(\eta^T J_{0\leftarrow u}[V_j, V_i](Y_u)\big)_{1\leqslant i\leqslant d}.$$

On the other hand, it is known that (cf. Lifshits [Lif95]) the modulus of continuity of fBM sample paths is $\delta^H|\log\delta|^{1/2}$. In particular, if one assumes that α is close to H, then

$$\overline{\lim_{|u-v|\to 0}} \frac{|X_{u,v}|}{|v-u|^{2\alpha}} = +\infty.$$

As a result, the regularity property (4.15) can only be true when I_u^1 is identically zero. Therefore, the claim (4.14) holds.

Now one can iterate the use of Lemma 4.14 to conclude in a similar way that

$$\eta^T J_{0\leftarrow u} W(Y_u) \equiv 0 \quad \forall u \in [0,t]$$

for all vector fields

$$W \in \bigoplus_{m=1}^{\infty} \mathcal{L}_{\mathcal{V}}^m,$$

where $\mathcal{L}_{\mathcal{V}}^m$ is the vector space of smooth vector fields defined inductively by

$$\mathcal{L}_{\mathcal{V}}^1 \triangleq \mathrm{Span}\{V_1, \cdots, V_d\},\ \mathcal{L}_{\mathcal{V}}^m \triangleq \mathrm{Span}\{[W,Z] : W \in \mathcal{L}_V^1, Z \in \mathcal{L}_{\mathcal{V}}^m\}.$$

In particular, at $u = 0$ one has

$$\eta^T W(y_0) = 0 \quad \forall W \in \bigoplus_{m=1}^{\infty} \mathcal{L}_{\mathcal{V}}^m.$$

Since $\eta \neq 0$, this is a contradiction to the Hörmander condition which asserts that

$$\mathrm{Span}\left\{W(y_0) : W \in \bigoplus_{m=1}^{\infty} \mathcal{L}_{\mathcal{V}}^m\right\} = \mathbb{R}^n.$$

Consequently, γ_{Y_t} is invertible a.s. The a.s. property comes from the fact that the above rough path analysis can be performed precisely at every sample path $w \in \mathcal{W}$ outside a μ-null set.

Remark 4.15. In the spirit of the above argument, the extension of Theorem 4.7 to a more general Gaussian process X requires two basic assumptions:

$$\int_0^t f_s^T dh_s = 0 \ \forall h \in \mathcal{H} \implies f_s = 0 \ \forall s \in [0,t]$$

and

$$\int_0^s Y_u dX_u = 0 \ \forall s \in [0,t] \implies Y_u = 0 \ \forall u \in [0,t]. \tag{4.16}$$

Once these conditions are justified, the previous analysis can be adapted to treat the case of more general Gaussian processes.

Remark 4.16. To prove the smoothness of density, one needs to establish the integrability condition

$$\det \gamma_{Y_t}^{-1} \in L^p(\mathcal{W}, \mu) \quad \forall p > 1.$$

This requires a much more quantitative version of the condition (4.16), known as *the Norris lemma*, which quantifies the property that

$$\int_0^{\cdot} Y dX \text{ is small} \implies Y \text{ is small}.$$

Such a lemma can be established pathwisely by using rough path analysis (cf. [CHLT15]).

4.2 The Brox diffusion

Another remarkable application of rough path theory is the mathematical solution to the KPZ equation by Hairer [Hai13]. The essential ideas and techniques developed in this work further lead to Hairer's groundbreaking theory of regularity structures for singular stochastic partial differential equations (cf. Hairer [Hai14]). To keep things relatively elementary, we use a simple example to illustrate how rough path theory can be applied to solve certain singular PDEs. The work we discuss here was due to Delarue-Diel [DD16].

Let $U : \mathbb{R} \to \mathbb{R}$ be a smooth function. Recall that the weak solution to the SDE

$$dX_t = \dot{U}(X_t)dt + dB_t \quad (B_t : \text{1D Brownian motion})$$

can be constructed in terms of the martingale problem associated with the generator

$$\mathcal{A} \triangleq \frac{1}{2}\frac{d^2}{dx^2} + \dot{U}(x)\frac{d}{dx}.$$

Here $\dot{U}(x)$ denotes the derivative of $U(x)$. From diffusion theory, the construction of a diffusion process $X = \{X_t\}$ with generator $\mathcal{A}$ is essentially equivalent to solving the PDE

$$\begin{cases} \dfrac{\partial u}{\partial t} = \mathcal{A}u + g, \quad (t,x) \in (0,\infty) \times \mathbb{R}; \\ u(0,\cdot) = f \end{cases} \tag{4.17}$$

for arbitrary (suitably regular) input functions $g(t,x)$ and $f(x)$. It is classical that the PDE (4.17) is well-posed when U is continuously differentiable.

Let us now consider the situation when $U = W : \mathbb{R} \to \mathbb{R}$ is itself a two-sided Brownian motion on $\mathbb{R}$ that is independent of B. The SDE

$$dX_t = \dot{W}(X_t)dt + dB_t, \tag{4.18}$$

known as the *Brox diffusion*, defines a diffusion process evolving in a Brownian random environment W. The corresponding PDE takes the form

$$\begin{cases} \dfrac{\partial u}{\partial t} = \dfrac{1}{2}\dfrac{\partial^2 u}{\partial x^2} + \dot{W}(x)\dfrac{\partial u}{\partial x} + g(t,x), \quad (t,x) \in (0,\infty) \times \mathbb{R}; \\ u(0,x) = f(x). \end{cases} \tag{4.19}$$

The weak and strong existence of X was established by Flandoli-Russo-Wolf [FRW03] and Bass-Chen [BC01] respectively. It was shown by Seignourel [Sei00] that the Brox diffusion is the weak scaling limit of discrete random walks in Bernoulli random environments (Sinan's random walks).

Since $\dot{W}$ makes no sense, neither the SDE (4.18) nor (4.19) could be defined in the classical sense. Nonetheless, rough path theory can be applied to give meaning to the PDE (4.19). Once the PDE is solved properly, the Brox diffusion can also constructed by using the standard theory of martingale problems.

The rough path method of Delarue-Diel is motivated from the following formal calculation. We begin by considering the following heat equation:

$$\begin{cases} \dfrac{\partial w}{\partial t} = \dfrac{1}{2}\dfrac{\partial^2 w}{\partial t^2} + g(t,x), \\ w(0,x) = f(x). \end{cases} \tag{4.20}$$

A simple application of Duhamel's principle yields that

$$w(t,x) = (P_t f)(x) + \int_0^t \int_{\mathbb{R}} p_{t-s}(x-y) g(s,y) dy ds,$$

where $p_t(x) \triangleq \frac{1}{\sqrt{2\pi t}} e^{-\frac{x^2}{2t}}$ denotes the heat kernel on $\mathbb{R}$ and

$$(P_t f)(x) \triangleq \int_{\mathbb{R}} f(y) p_t(x-y) dy$$

is the associated heat semigroup. Now let $u(t,x)$ be the (formal) solution to the PDE (4.19) and set $J(t,x) \triangleq w(t,x) - u(t,x)$. Then J satisfies $J(0,\cdot) = 0$ and

$$\begin{aligned} \frac{\partial J}{\partial t} &= \frac{\partial w}{\partial t} - \frac{\partial u}{\partial t} \\ &= \frac{1}{2}\frac{\partial^2 w}{\partial t^2} + g - \frac{1}{2}\frac{\partial^2 u}{\partial t^2} - \dot{W} \cdot \frac{\partial u}{\partial x} - g \\ &= \frac{1}{2}\frac{\partial^2 J}{\partial t^2} - \dot{W} \cdot \frac{\partial u}{\partial x}. \end{aligned}$$

Another application of Duhamel's principle shows that

$$J(t,x) = -\int_0^t \int_{\mathbb{R}} p_{t-s}(x-y) \dot{W}(y) \frac{\partial u}{\partial x}(s,y) dy ds.$$

Therefore, one has

$$\begin{aligned} u(t,x) &= w(t,x) - J(t,x) \\ &= (P_t f)(x) + \int_0^t \int_{\mathbb{R}} p_{t-s}(x-y) g(s,y) dy ds \\ &\quad + \int_0^t \int_{\mathbb{R}} p_{t-s}(x-y) \dot{W}(y) \frac{\partial u}{\partial x}(s,y) dy ds. \end{aligned} \tag{4.21}$$

Note that (4.21) is only formal as the last integral is ill-defined.

We now apply an integration by parts to see that

$$\int_{\mathbb{R}} p_{t-s}(x-y)\dot{W}(y)\frac{\partial u}{\partial x}(s,y)dy = \int_{\mathbb{R}} p_{t-s}(x-y)d\big(\int_x^y \frac{\partial u}{\partial x}(s,z)dW(z)\big)$$
$$= -\int_{\mathbb{R}} \frac{\partial}{\partial x}p_{t-s}(x-y)\big(\int_x^y \frac{\partial u}{\partial x}(s,z)dW(z)\big)dy.$$

By substituting this into (4.21), one obtains that

$$u(t,x) = (P_t f)(x) + \int_0^t \int_{\mathbb{R}} p_{t-s}(x-y)g(s,y)dyds$$
$$- \int_0^t \int_{\mathbb{R}} \frac{\partial}{\partial x}p_{t-s}(x-y)\big(\int_x^y \frac{\partial u}{\partial x}(s,z)dW(z)\big)dyds.$$

As a consequence, the function $v(t,x) \triangleq \frac{\partial u}{\partial x}(t,x)$ satisfies the following integral equation:

$$v(t,x) = \frac{\partial P_t f}{\partial x} + \int_0^t \int_{\mathbb{R}} \frac{\partial}{\partial x}p_{t-s}(x-y)g(s,y)dyds$$
$$- \int_0^t \int_{\mathbb{R}} \frac{\partial^2}{\partial x^2}p_{t-s}(x-y)\big(\int_x^y v(s,z)dW(z)\big)dyds. \qquad (4.22)$$

Here comes the point where rough path theory enters the picture. The dW-integral in the last term of (4.22) can be regarded as a rough integral. This requires viewing W as a rough path and $z \mapsto v(s,z)$ (for each fixed s) as a rough path controlled by W. Note that the "rough path variable" here is the space variable z (the time variable s is a parameter). One can then use the results from Section 2.2 to make sense of the integral $\int_x^y v(s,z)dW(z)$ rigorously. The moral is that, one can define a Banach space $\mathcal{B}$ with a suitably designed norm Θ, which consists of continuous functions $v(t,x)$ such that for each fixed t, the path $\mathbb{R} \ni x \mapsto v(t,x)$ is controlled by W on every compact interval. In the spirit of RDE theory, the solution to the integral equation (4.22) can be formulated as the fixed point of the transformation $\mathcal{M} : \mathcal{B} \to \mathcal{B}$ defined by

$$(\mathcal{M}v)(t,x) \triangleq \frac{\partial P_t f}{\partial x} + \int_0^t \int_{\mathbb{R}} \frac{\partial}{\partial x}p_{t-s}(x-y)g(s,y)dyds$$
$$- \int_0^t \int_{\mathbb{R}} \frac{\partial^2}{\partial x^2}p_{t-s}(x-y)\big(\int_x^y v(s,z)dW(z)\big)dyds.$$

By constructing the norm Θ carefully and making use of delicate heat kernel estimates, it is possible to make $\mathcal{M}$ into a contraction mapping under Θ. The existence and uniqueness of the solution v is then a direct consequence of the Banach fixed point theorem as in the RDE case.

Remark 4.17. The above perspective is analytic and pathwise (with respect to the Brownian motion W). The assumption that W is a Brownian motion plays no essential role (one can assume that $x \mapsto W(x)$ is an arbitrary α-Hölder continuous path with at most polynomial growth at infinity). On the other hand, the above argument does not fully capture the remarkable analysis developed in the Delarue-Diel paper. There are two significant challenges in their work. The first challenge is that the potential W under their consideration is time-dependent. This raises substantial difficulty when lifting $W(t,x)$ to a time-dependent rough path, as the spatial rough path structures at different times will interfere with each other in this case. If W is time-independent, the lifting of W is essentially trivial – one simply takes

$$\mathbf{W} = (W^1, W^2) : \ W^1_{x,y} \triangleq W(y) - W(x), W^2_{x,y} \triangleq \frac{1}{2}(W(y) - W(x))^2.$$

The second challenge is that the rough path metric Θ needs to be tuned in a delicate way so that the transformation $\mathcal{M}$ is a contraction mapping under Θ. The analysis for this part is indeed highly non-trivial.

Remark 4.18. The Delarue-Diel construction of the Brox diffusion has a natural extension to multidimensions. In this case, $x \mapsto W(x)$ is a spatially rough field. One needs to apply either Hairer's theory of regularity structures or Gubinelli's theory of para-controlled calculus to formulate a multidimensional analogue of the equation (4.22) properly. We refer the reader to Cannizzaro-Chouk [CC18] for an approach based on the para-controlled calculus.

4.3 The signature uniqueness theorem

In recent years, there has been an emerging approach of using the signature transform in rough path theory combined with neural network techniques to study problems in data sciences (cf. e.g. [LZL19, WlN19, XSJ17]). In a vague form, the underlying problem has the following nature: learning the (nonlinear) functional relationship between an input data stream (an ordered sequence of points in a vector space) and its output effect. For instance, use the atmosphere data observed over the last 24 hours to predict local weather in the next 24 hours. The essential idea behind the signature-based approach, in its earlier stage but not state-of-the-art, can be summarised as follows.

An input data stream has a natural ordering so that it can be easily turned into a continuous path in a vector space V (e.g. joining the sequential points by line segments). The signature-based approach takes the so-called *signature transform* of x as a feature set and feed it into the machinery of deep learning to learn the underlying target function. More precisely, the signature transform maps a path $x : [0, T] \to V$ to the sequence

$$S(x) \triangleq \Big(x_T - x_0, \int_{0<s<t<T} dx_s \otimes dx_t, \cdots, \int_{0<t_1<\cdots<t_n<T} dx_{t_1} \otimes \cdots \otimes dx_{t_n}, \cdots\Big) \quad (4.23)$$

of global iterated integrals along x. In practice, one takes a truncation of $S(x)$ up to a certain level N, denoted as $S_N(x)$, and use it as a feature set representing the input path x. The original problem is thus transformed into the problem of learning the functional relationship between $S_N(x)$ and the output effect y. The latter can be solved by using traditional statistical methods (e.g. linear regression) or modern neural network techniques from deep learning.

A theoretical basis of the aforementioned approach is a fundamental theorem in rough path theory, known as the *signature uniqueness theorem*, which asserts that every rough path is uniquely determined by its signature up to tree-like pieces. As a result, the signature transform encodes all essential information about the original data stream. The effectiveness of this approach is reflected in the following two properties of the signature transform:

Property 1. The truncated signature $S_N(x)$ converges to the actual signature $S(x)$ factorially fast as $N \to \infty$. This property suggests that the truncation does not lose much information even when N is not large.

Property 2. According to the shuffle product formula (1.15), polynomial functions on the space of signatures are always linear. As a consequence, the original non-linear problem on the data stream space is linearised when lifted up to the signature space. This perspective is robust and model-free.

Remark 4.19. The linearisation property becomes less relevant in recent works. Indeed, more efficient methods from deep learning have been introduced to replace the traditional linear regression on the signature space.

To conclude the present notes, we explain the signature uniqueness theorem and the essential idea behind its proof. We begin with the definition of the signature transform. Let V be a Banach space. Recall that $T^N(V)$ is the truncated tensor algebra of order N defined by (1.8) and $T((V))$ is the infinite tensor algebra defined by

$$T((V)) \triangleq \prod_{n=0}^{\infty} V^{\otimes n} \triangleq \{\xi = (\xi_0, \xi_1, \xi_2, \cdots) : \xi_n \in V^{\otimes n} \ \forall n\}.$$

Let $\mathbf{X}$ be a rough path over V and let

$$\mathbb{X} = (1, X^1, X^2, \cdots) : \Delta_T \to T((V))$$

be the Lyons extension of $\mathbf{X}$ (cf. Example 3.2).

Definition 4.20. The quantity $\mathbb{X}_{0,T} \in T((V))$, denoted as $S(\mathbf{X})$, is called the *signature* of the rough path $\mathbf{X}$.

Example 4.21. Let $x : [0, T] \to V$ be a smooth path. Then the signature of x is given by the global iterated integrals (4.23) defined in the classical sense.

Tree-like paths

To state the signature uniqueness theorem, we first need to introduce the notion of tree-like paths. Heuristically, a tree-like path is a path that travels out and reverses back to cancel itself. A mathematical way of capturing this property is through the notion of real trees.

Definition 4.22. A *real tree* is a metric space $(\mathcal{T}, \rho)$ in which any two distinct points can be joined by a unique continuous, non-self-intersecting path. In addition, this path is a geodesic in the metric space.

From the definition, it is clear that a real tree does not contain non-degenerate loops (i.e. subspaces homeomorphic to the circle). Tree-like paths are defined to be paths that can be realised as loops on a real tree (hence they must possess the aforementioned self-cancelling property).

Definition 4.23. Let $x : [0, T] \to V$ be a continuous path in some topological space V. We say that x is *tree-like*, if there exist a real tree $\mathcal{T}$ together with continuous mappings

$$\phi : [0, T] \to \mathcal{T}, \ \psi : \mathcal{T} \to V$$

such that $\phi(0) = \phi(T)$ and $x = \psi \circ \phi$. A *tree-like piece* of a continuous path x is a portion $[s, t]$ such that $x|_{[s,t]}$ is tree-like. We say that x is *tree-reduced* if it does not contain any tree-like pieces.

The following figure gives examples of a real tree, a tree-reduced path and a path containing a tree-like piece.

The signature uniqueness theorem

The signature uniqueness theorem can now be stated as follows. It was first proved by Hambly-Lyons [HL10] for paths with bounded total variation and then extended to the rough path case by Beodihardjo et al. [BGLY16]

Theorem 4.24. *Let $\mathbf{X}$ be a weakly geometric rough path in a Banach space V. Then $\mathbf{X}$ has trivial signature if and only if $t \mapsto \mathbf{X}_{0,t}$ tree-like. As a result, two weakly geometric rough paths $\mathbf{X}, \mathbf{Y}$ have the same signature if and only if they are equal up to tree-like pieces.*

We only explain the heuristic idea behind the proof of Theorem 4.24. The sufficiency part is convincing. It is an elementary exercise to see that $S(x \sqcup \overleftarrow{x}) = \mathbf{1}$ if x is a piecewise linear or smooth path in $\mathbb{R}^d$. The general case is treated by approximating a tree-like rough path by tree-like piecewise geodesic paths on the real tree.

The main challenge lies in the necessity part. Let $\mathcal{G}_\alpha$ denote the space of signatures for weakly geometric α-Hölder rough paths. The key point is to show that $\mathcal{G}_\alpha$ is a real tree with respect to a suitable metric. Presuming this is true, given any $\mathbf{X} : [0,T] \to G^N(V)$, one can write $\mathbf{X} = \pi^N \circ \mathbb{X}$ where $\mathbb{X} : [0,T] \to \mathcal{G}_\alpha$ is the signature path of $\mathbf{X}$. Saying that $\mathbf{X}$ has trivial signature precisely means that $\mathbb{X}$ is a loop on the real tree $\mathcal{G}_\alpha$. It follows that $\mathbf{X}$ is tree-like in the sense of Definition 4.23.

The core of the argument is thus to show that $\mathcal{G}_\alpha$ is a real tree. For this purpose, an essential property is that for any $g \in \mathcal{G}_\alpha$, there exists a unique path $\mathbf{X}$ (up to reparametrisation) such that:

(i) $S(\mathbf{X}) = g$;
(ii) the signature path $\mathbb{X}$ of $\mathbf{X}$ is non-self-intersecting.

Since we already know by definition that g comes from a rough path, the existence part follows from a topological procedure of loop removal. The uniqueness part is proved by contradiction. Suppose on the contrary that $\mathbf{X}, \mathbf{Y}$ are two paths satisfying Properties (i), (ii) but their signature paths $\mathbb{X}, \mathbb{Y}$ have non-identical images. Let $\mathcal{U}$ be a small neighbourhood around some point $\mathbb{X}_t$ that is disjoint from the entire image of $\mathbb{Y}$. One can then construct a smooth one-form Φ that is compactly supported in $\mathcal{U}$, such that

$$\int_0^T \Phi(\mathbb{X})d\mathbb{X} = 1, \ \int_0^T \Phi(\mathbb{Y})d\mathbb{Y} = 0. \tag{4.24}$$

On the other hand, since $\mathbf{X}, \mathbf{Y}$ have the same signature g, as a consequence of the shuffle product formula one sees that

$$\int_0^T \Psi(\mathbb{X})d\mathbb{X} = \int_0^T \Psi(\mathbb{Y})d\mathbb{Y} \quad \forall \text{polynomial one-form } \Psi. \tag{4.25}$$

By a standard approximation procedure, (4.25) remains valid for all smooth one-forms Ψ. This is a contradiction to (4.24) and thus the uniqueness of $\mathbf{X}$ holds. The above property shows that for any element $g \in \mathcal{G}_\alpha$, there is a unique non-self-intersecting path joining g and the identity $\mathbf{1}$ on $\mathcal{G}_\alpha$. With some extra effort, one can further show that the same is true for any pair of distinct elements $g, h \in \mathcal{G}_\alpha$. This essentially yields the tree property (i.e being loop-free). The construction of a real tree metric requires deeper tools from real tree theory and will not be discussed here.

References

[BC01] R. Bass and Z. Chen. Stochastic differential equations for Dirichlet processes. *Probab. Theory Relat. Fields* 121 (2001): 422–446.

[Bau04] F. Baudoin, An introduction to the geometry of stochastic flows, Imperial College Press, 2004.

[BG15] H. Boedihardjo and X. Geng. The uniqueness of signature problem in the non-Markov setting. *Stochastic Process. Appl.* 125 (12) (2015): 4674-4701.

[BG20] H. Boedihardjo and X. Geng. Lipschitz stability of controlled rough paths and rough differential equations. *arXiv:2009.13084*, 2020.

[BGLY16] H. Boedihardjo, X. Geng, T. Lyons and D. Yang. The signature of a rough path: uniqueness. *Adv. Math.* 293 (2016): 720-737.

[CC18] G. Cannizzaro and K. Chouk. Multidimensional SDEs with singular drift and universal construction of the polymer measure with white noise potential. *Ann. Probab.* 46 (2018): 1710– 1763.

[CF10] T. Cass and P. Friz. Densities for rough differential equations under Hörmander's condition. *Ann. of Math.* (2) 171 (3) (2010): 2115–2141.

[CHLT15] T. Cass, M. Hairer, C. Litterer and S. Tindel. Smoothness of the density for solutions to Gaussian rough differential equations. *Ann. Probab.* 43 (1) (2015): 188–239.

[CQ02] L. Coutin and Z. Qian. Stochastic analysis, rough path analysis and fractional Brownian motions. *Probab. Theory Relat. Fields* 122 (2002): 108–140.

[Che57] K.T. Chen. Integration of paths, geometric invariants and a generalized Baker-Hausdorff formula. Ann. of Math. 65 (1957): 163–178.

[Dav08] A.M. Davie. Differential equations driven by rough paths: an approach via discrete approximation. *Appl. Math. Res. Express.* AMRX 2008 (1): 1–40.

[DD16] F. Delarue and R. Diel. Rough paths and 1d SDE with a time dependent distributional drift: application to polymers. *Probab. Theory Relat. Fields* 165 (2016):1–63.

[FRW03] F. Flandoli, F. Russo and J. Wolf. Some SDEs with distributional drift. I. General calculus. *Osaka J. Math.* 40 (2003): 493–542.

[FH14] P.K. Friz and M. Hairer. *A course on rough paths.* Springer, 2014.

[FV10] P.K. Friz and N. Victoir. *Multidimensional stochastic processes as rough paths.* Cambridge University Press, 2010.

[Gub04] M. Gubinelli. Controlling rough paths. *J. Funct. Anal.* 216 (1) (2004): 86–140.

[Gub10] M. Gubinelli. Ramification of rough paths. *J. Differential Equations* 248 (4) (2010): 693–721.

[Hai13] M. Hairer. Solving the KPZ equation. *Ann. of Math.* (2) 178 (2) (2013): 559–664.

[Hai14] M. Hairer. A theory of regularity structures. *Invent. Math.* 198 (2) (2014): 269–504.

[HL10] B. Hambly and T. Lyons. Uniqueness for the signature of a path of bounded variation and the reduced path group. *Ann. of Math.* 171 (1) (2010): 109–167.

[LZL19] C. Li, X. Zhang, L. Liao et al. Skeleton-based gesture recognition using several fully connected layers with path signature features and temporal transformer module. The Thirty-third AAAI Conference on Artificial Intelligence, 2019.

[Lif95] M.A. Lifshits. *Gaussian random functions.* Springer, 1995.

[Lyo94] T.J. Lyons. Differential equations driven by rough signals. I. An extension of an inequality of L. C. Young. *Math. Res. Lett.* 1 (4) (1994): 451–464.

[Lyo98] T.J. Lyons. Differential equations driven by rough signals. *Rev. Mat. Iberoamericana* 14 (2) (1998): 215–310.

[Nua06] D. Nualart. The Malliavin calculus and related topics, 2nd Edition. Springer-Verlag, 2006.

[Reu93] C. Reutenauer, *Free Lie algebras*, Clarendon Press, Oxford, 1993.

[Sei00] P. Seignourel. Discrete schemes for processes in random media. *Probab. Theory Relat. Fields* 118 (2000): 293–322.

[Shi04] I. Shigekawa. *Stochastic analysis.* Americal Mathematical Society, 2004.

[Ste70] E.M. Stein. *Singular integrals and differentiability properties of functions.* Princeton University Press, 1970.

[Str87] R.S. Strichartz. The Campbell-Baker-Hausdorff-Dynkin formula and solutions of differential equations. *J. Funct. Anal.* 72 (1987): 320–345.

[WlN19] B. Wang, M. Liakata, H. Ni et al. A path signature approach for speech emotion recognition. *Interspeech* (2019): 1661–1665.

[WZ65] E. Wong and M. Zakai. On the relation between ordinary and stochastic differential equations. *Internat. J. Engrg. Sci.* 3 (2) (1965): 213–229.

[XSJ17] Z. Xie, Z. Sun, L. Jin et al. Learning spatial-semantic context with fully convolutional recurrent network for online handwritten Chinese text recognition. *IEEE transactions on pattern analysis and machine intelligence* 40 (8) (2017): 1903–1917.

[You36] L.C. Young. An inequality of Hölder type connected with Stieltjes integration. *Acta Math.* 67 (1936): 251–282.