The Signature of a Rough Path: Uniqueness

Xi Geng

Lincoln College

University of Oxford

A thesis submitted for the degree of

Doctor of Philosophy

Trinity 2015

Dedicated to my beloved wife Jiena, and my parents Gaofei and Yuezhen
for their enduring love and support.

Acknowledgements

First of all, I would like to express my greatest gratitude to my supervisor Professor Zhongmin Qian for his continuous encouragement, enormous inspiration and precious suggestions during my DPhil study. In the past three years, his deep insight and broad knowledge in mathematics gradually shaped my own perspective and my invaluable enthusiasm in the subject. In particular, his inspiring teaching and discussions led me into the fields of stochastic analysis, differential geometry and theoretical physics. It is the inner elegance, profound significance of the theories and deep interactions among them that make me be so obsessed with mathematics. It is also his lasting influence that formed my understanding on how to think and work in mathematics independently and persistently. He is my great mentor in mathematics and my life.

Moreover, I would like to thank Professor Jiagang Ren (Sun Yat-sen University) so much for his wonderful teaching in lectures and seminars when I was an undergraduate at Sun Yat-sen University. He is the one that brought me a first look at frontier mathematics, and his admirable spirit in doing mathematics will always be with me. Without him, I would not be studying at Oxford and probably even not choose to be a research mathematician.

Also, I am extremely grateful to my wife Jiena Chen, my parents Gaofei Geng, Yuezhen Gao, my friends and my colleagues Horatio Boedihardjo, Wei Pan, Lihao Liang, Weijun Xu, Zichen Zhang, Danyu Yang, Hao Ni etc. for their tremendous support during my DPhil study. They brought me great happiness for studying and living at Oxford.

Last but not the least, I would like to express my special thanks to the Oxford-Man Institute which gave me full financial support and provided me with excellent resources for my DPhil study, and in particular to Professor Terry Lyons who gave me plenty of precious advices in mathematics and also the opportunity to work with him as a postdoctoral research fellow.

Abstract

The main contribution of the present thesis is in two aspects.

The first one, which is the heart of the thesis, is to explore the fundamental relation between rough paths and their signatures. Our main goal is to give a geometric characterization of the kernel of the signature map in different situations. In Chapter Two, we start by establishing a general fact that a continuous Jordan curve on a Riemannian manifold can be arbitrarily well approximated by piecewise minimizing geodesic interpolations which are again Jordan. This result enables us to prove a generalized version of Green's theorem for planar Jordan curves with finite p-variation for $1 \leqslant p < 2$, and to prove that two such Jordan curves have the same signature if and only if they are equal up to reparametrization. In Chapter Three, we investigate the problem for general weakly geometric rough paths. In particular, we show that a weakly geometric rough path has trivial signature if and only if it is tree-like in the sense we will define later on. In Chapter Four, we study the problem in the probabilistic setting. In particular, we show that for a class of stochastic processes, with probability one the sample paths are determined by their signatures up to reparametrization. A fundamental example is Gaussian processes including fractional Brownian motion with Hurst parameter $H > 1/4$, the Ornstein-Uhlenbeck process and the Brownian bridge.

The second one is an application of rough path theory to the study of nonlinear diffusions on manifolds under the framework of nonlinear expectations. In Chapter Five, we begin by studying the geometric rough path nature of G-Brownian motion. This enables us to introduce rough differential equations driven by G-Brownian motion from a pathwise point of view. Next we establish the fundamental relation between rough (pathwise theory) and stochastic (L^2-theory) differential equations driven by G-Brownian motion. This is a crucial point of understanding nonlinear diffusions and their generating heat flows on manifolds from an intrinsic point of view. Finally, from the pathwise point of view we construct G-Brownian motion on a compact Riemannian manifold and establish its generating heat flow for a class of G-functions under orthogonal invariance. As an independent interest, we also develop the Euler-Maruyama scheme for stochastic differential equations driven by G-Brownian motion.

Contents

List of Figures

Chapter 1

Introduction

1.1 Motivation and Main Results of the Thesis

The set of continuous paths in $\mathbb{R}^d$ forms a semigroup with involution, with the group operation and involution given by concatenation and reversal of paths. As early as 1954, K.T. Chen [8] observed that the map sending a continuous path $x: [0,1] \to \mathbb{R}^d$ with bounded total variation to the formal series

$$1 + \int_0^1 dx_s^i E_i + \int_0^1 \int_0^{s_2} dx_{s_1}^i dx_{s_2}^j E_i E_j + \cdots , \tag{1.1.1}$$

where $E_1, \cdots, E_d$ are indeterminates and x^i denotes the i-th coordinate component of x, is a homomorphism from the semigroup of continuous paths to the algebra of non-commutative formal power series. In general, this map is not injective; it is apparent that any path concatenated with its reversal is mapped to the trivial formal series. It seems however that the map is essentially injective if we restrict our attention to paths that "do not track back along themselvs". Indeed, in 1958 K.T. Chen himself [9] already proved that the map is injective on the space of regular, irreducible paths. In 2010, B.M. Hambly and T. Lyons [35] extended K.T. Chen's result to the space of continuous paths with bounded total variation and first introduced the notion of tree-like paths to describe paths that track back along themselves. In particular, they proved that the formal series corresponding to a path, which they called the signature of the path, is trivial if and only if the path is tree-like.

Aside from its interesting algebraic properties, the map also gained attention through the fundamental role it plays in the theory of path integration. In 1936, L.C. Young [69] defined the Stieltjes type integral $\int_0^1 y_t dx_t$ in terms of a Riemann sum when x and y have finite p- and q-variation respectively, where $1/p+1/q > 1$. In particular, this allows us to define, for a Lipschitz one form ϕ, the path integral $\int_0^1 \phi(x_t)\, dx_t$ when x is a multidimensional path with finite p-variation for $1 \leqslant p < 2$. In the same paper, L.C. Young gave an example where the integral $\int_0^1 \phi(x_t)\, dx_t$ defined using a Riemann sum diverges if x has only finite 2-variation. In other words, the Stieltjes integration map $x \to \int_0^1 \phi(x_t)\, dx_t$ does not have a closable graph under the p-variation metric if $p \geqslant 2$. The seemingly insurmountable $p = 2$ barrier, at least in the deterministic setting, was to remain for another sixty years. In 1998, T. Lyons [45] showed that the Stieltjes integration map has a closable graph under the p-variation metric if the path x takes values in the step-$\lfloor p \rfloor$ free nilpotent Lie group. He called these paths weakly geometric p-rough paths. The first step in the construction of such integrals is to define the signature for weakly geometric rough paths, which can be viewed as a generalization of the formal series (1.1.1) of iterated path integrals. The integration of one forms against such paths is then defined via Taylor's expansion and by using the multiplicative structure of the signature in an essential way. Later on, there have been extensions of T. Lyons' integration theory to more general settings, see for example M. Gubinelli [30] for controlling rough paths, T. Lyons and D. Yang [49] for integrating time-varying cocyclic one forms against rough paths.

From a theoretical point of view, it is a fundamental question about whether we could further extend B.M. Hambly and T. Lyons' result to the case of weakly geometric rough paths; namely whether the signature of a weakly geometric rough path determines the path uniquely up to tree-like equivalence. From a practical point of view, there has also been work done, for example by D. Levin, T. Lyons and H. Ni [44], on analyzing time series data using the signature map. The justification of their method implicitly uses the fact that the map from a path to its signature is injective in a certain sense.

Before any answer to the question in the deterministic setting, there has already been exciting progress on the problem in the probabilistic setting. In 2012, Y. Le Jan and Z. Qian [43] proved that with probability one, the Stratonovich signatures of Brownian motion determine Brownian sample paths. Later on, their result was extended to hypoelliptic diffusions by X. Geng and Z. Qian [28], and to Chordal SLE$_\kappa$ curves with $\kappa \leqslant 4$ by H. Boedihardjo, H. Ni and Z. Qian [6]. It should be pointed out that in [6], the authors already gave a complete answer to the question for planar simple (i.e. non-self-intersecting) curves with finite p-variation for $1 \leqslant p < 2$. In this case tree-like equivalence reduces to the equivalence of reparametrization. This is the first result in the deterministic setting beyond the bounded total variation case.

The main contribution of the present thesis is to investigate this problem in general in both deterministic and probabilistic settings. From now on, we call this problem the *uniqueness of signature* problem.

In Chapter 2, we first consider the special case of planar Jordan curves. In fact, we begin by establishing a general result regarding simple piecewise geodesic approximation of simple and Jordan curves on an arbitrary Riemannian manifold. As two important consequences, we prove a generalized Green's theorem for planar Jordan curves with finite p-variation for $1 \leqslant p < 2$ and solve the uniqueness of signature problem for this case (more precisely, we show that the curve is uniquely determined by its signature up to reparametrization). The contents of this chapter are based on joint work with H. Boedihardjo in the paper [3] in 2013.

In Chapter 3, we solve the general uniqueness of signature problem for weakly geometric rough paths. More precisely, we prove that the signature of a weakly geometric rough path is trivial if and only if it is tree-like. The contents of this chapter are based on joint work with H. Boedihardjo, T. Lyons and D. Yang in the paper [5] in 2014.

In Chapter 4, we study the problem in the probabilistic setting by further extending the results and techniques in [28],[43]. In particular, we prove that for a certain class of non-Markov processes, with probability one the signatures of the process determine the sample paths uniquely up to reparametrization. As a fundamental example, we show that our method applies to a class of Gaussian processes including fractional Brownian motion with Hurst parameter $H > 1/4$, the Ornstein-Uhlenbeck process and the Brownian bridge. The contents of this chapter are based on joint work with H. Boedihardjo in the paper [4] in 2014.

From T. Lyons' point of view, the path integration theory is essential for the study of differential equations driven by rough paths. In fact, in the same paper [45], T. Lyons proved the existence and uniqueness of solutions to differential equations driven by rough paths by regarding the equation as a rough integral equation and then by using Picard iteration. Moreover, he established the continuity of the solution map with respect to the driving path under the p-variation metric, which is usually known as the universal limit theorem. Later on, reformulations and extensions of T. Lyons' theory of rough differential equations appeared, for example in A.M. Davie [16], P. Friz and N. Victoir [25] via discrete approximation.

The theory of rough paths leads to an enormous number of applications among different fields, in particular in probability theory. As the sample paths of many interesting stochastic processes are irregular, pathwise solutions to stochastic differential equations driven by such processes were not well understood before the appearance of rough path theory. Once we have established the rough path nature for sample paths of stochastic processes, the pathwise theory of stochastic differential equations is a direct consequence of the deterministic results. A fundamental example is Brownian motion. It was shown by E.M. Sipiläinen [60], T. Lyons and Z. Qian [47] that with probability one, sample paths of Brownian motion can be lifted as geometric $(2+\varepsilon)$-rough paths in a canonical way. Moreover, pathwise solutions to stochastic differential equations driven by Brownian motion coincide with Stratonovich's solutions, which relates to Itô's solutions in terms of the famous result of E. Wong and M. Zakai [66]. The rough path regularity of other important stochastic processes and related applications are well summarized in the monograph by P. Friz and N. Victoir [26]. Other important applications of rough path theory in probability include, for example to the Malliavin calculus for Gaussian rough differential equations by T. Cass and P. Friz [7], to support theorem and large deviations by M. Ledoux, Z. Qian and T. Zhang [42], and to stochastic partial differential equations by M. Hairer [31], [32].

In the last chapter of the present thesis, we explore another application of rough path theory in studying nonlinear diffusions on manifolds under the framework of nonlinear expectations, originally introduced by S. Peng [56] in 2007. The theory of nonlinear expectations, or more precisely, of G-expections, is motivated from the study of probability model uncertainty. In contrast to classical stochastic analysis, the fundamental feature of G-diffusions is that the generating heat flows are nonlinear. Starting with a G-function which captures the underlying nonlinearity, it is an interesting question to ask what the associated intrinsic nonlinear heat flow looks like on a Riemannian manifold.

The last chapter of the thesis is devoted to an answer to this question based on the theory of rough paths. As a crucial point, we first show that quasi-surely, sample paths of G-Brownian motion can be lifted canonically to geometric p-rough paths for $2 < p < 3$. This enables us to introduce the notion of rough differential equations driven by G-Brownian motion in the pathwise sense. Next we establish the fundamental relation between stochastic (in the L^2-sense of S. Peng) and rough differential equations driven by G-Brownian motion. It follows that we are able to construct G-diffusions on a differentiable manifold easily from a pathwise point of view. This is the starting point of constructing G-Brownian motion on a compact Riemannian manifold via J. Eells, K.D. Elworthy and P. Malliavin's approach. The last part of the chapter is devoted to such construction for a wide and interesting class of G-functions under orthogonal invariance. In particular, we establish the generating nonlinear heat flow for such G-Brownian motion and construct the canonical G-expectation on the path space over the manifold. As a result of independent interest, we also develop the Euler-Maruyama scheme for stochastic differential equations driven by G-Brownian motion. The contents of this chapter are based on joint work with Z. Qian and D. Yang in the paper [29] in 2013.

1.2 Background on Rough Path Theory

In this section, we recall the basic notions of rough path theory. The contents of this section are based on the monographs by T. Lyons and Z. Qian [47], T. Lyons, M. Caruana and T. Lévy [46], and P. Friz and N. Victoir [26]. These are excellent references for a systematic introduction to rough path theory and its applications.

In the present thesis, we only consider finite dimensional paths and hence we restrict ourselves to the finite dimensional setting. However, it should be pointed out that the original rough path theory of T. Lyons was developed for general Banach space-valued paths.

Let $T\left(\mathbb{R}^d\right)$ denote the infinite dimensional tensor algebra over $\mathbb{R}^d$. Let π_N denote the projection map from $T\left(\mathbb{R}^d\right)$ to $\left(\mathbb{R}^d\right)^{\otimes N}$ and $\pi^{(N)}$ denote the projection map from $T\left(\mathbb{R}^d\right)$ to the truncated N-th tensor algebra

$$T^N\left(\mathbb{R}^d\right) := \oplus_{i=0}^N \left(\mathbb{R}^d\right)^{\otimes i}.$$

Here we equip $\left(\mathbb{R}^d\right)^{\otimes N}$ with the Euclidean norm by identifying it with $\mathbb{R}^{d^N}$. Let $\triangle = \{(s,t) : 0 \leqslant s \leqslant t \leqslant 1\}$ be the standard 2-simplex. Throughout the rest of this section, $p \geqslant 1$ is a fixed constant.

Definition 1.2.1. A *multiplicative functional of degree* $N \in \mathbb{N}$ is a continuous map $\mathbf{X} = \left(1, X^1_{\cdot,\cdot}, \cdots, X^N_{\cdot,\cdot}\right) : \triangle \to T^N\left(\mathbb{R}^d\right)$ satisfying the following so-called Chen's identity:

$$\mathbf{X}_{s,u} \otimes \mathbf{X}_{u,t} = \mathbf{X}_{s,t},\ \forall 0 \leqslant s \leqslant u \leqslant t \leqslant 1.$$

Let $\mathbf{X}, \mathbf{Y}$ be two multiplicative functionals of degree N. Define

$$d_p\left(\mathbf{X}, \mathbf{Y}\right) = \max_{1 \leqslant i \leqslant N} \sup_{\mathcal{P}_{[0,1]}} \left(\sum_l \left| X^i_{t_{l-1},t_l} - Y^i_{t_{l-1},t_l} \right|^{\frac{p}{i}} \right)^{\frac{i}{p}},$$

where $\mathcal{P}_{[0,1]}$ denotes all finite partitions of $[0,1]$. d_p is called the *p-variation metric.* If $d_p\left(\mathbf{X}, \mathbf{1}\right) < \infty$ where $\mathbf{1} = (1, 0, \cdots, 0)$, we say that $\mathbf{X}$ has *finite total p-variati*on. A multiplicative functional of degree $\lfloor p \rfloor$ with finite total p-variation is called a *p-rough path*. The space of p-rough paths is denoted by $\Omega_p(\mathbb{R}^d)$.

It can be proved that $\left(\Omega_p(\mathbb{R}^d), d_p\right)$ is a complete metric space.

The p-variation metric d_p is in general hard to use. Equivalently, when describing convergence and continuity for rough paths, we usually use the notion of control.

Definition 1.2.2. A *control* over $[0,1]$ is a continuous function $\omega: \ \Delta \to [0,\infty)$ such that

$$\omega(s,t) + \omega(t,u) \leqslant \omega(s,u)$$

for any $0 \leqslant s \leqslant t \leqslant u \leqslant 1$.

Definition 1.2.3. A multiplicative functional $\mathbf{X}$ of degree N $(1 \leqslant N \leqslant \infty)$ is said to have *finite p-variation* if there exists a control ω such that

$$\left|X^i_{s,t}\right| \leqslant \omega(s,t)^{\frac{i}{p}}, \text{ for all } 1 \leqslant i \leqslant N.$$

It can be shown (see [47]) that a multiplicative functional $\mathbf{X}$ of degree N $(N \in \mathbb{N})$ has finite total p-variation if and only if it has finite p-variation. The key point lies in the fact that according to the multiplicative structure, if $\mathbf{X}$ has finite total p-variation, then

$$\omega(s,t) := \sum_{i=1}^{N} \sup_{\mathcal{P}_{[s,t]}} \sum_{l} \left|X^i_{t_{l-1},t_l}\right|^{\frac{p}{i}}, \ (s,t) \in \Delta,$$

is a control. In many situations, to emphasize the control of a p-rough path $\mathbf{X}$, we usually say that $\mathbf{X}$ is controlled by ω.

The following so-called Lyons' extension theorem asserts that the signature of a p-rough path is well defined and is locally Lipschitz continuous with respect to the p-rough path in some sense. We refer the reader to [47] for the proof.

Theorem 1.2.1. *(1) Let $\mathbf{X}$ be a p-rough path. Then for any $i \geqslant \lfloor p \rfloor + 1$, there exists a unique continuous map $X^i: \ \Delta \to \left(\mathbb{R}^d\right)^{\otimes i}$ such that*

$$S(\mathbf{X}) := \left(1, X^1, \cdots, X^{\lfloor p \rfloor}, \cdots X^i, \cdots\right)$$

is a multiplicative functional in $T\left(\mathbb{R}^d\right)$ with finite p-variation. Moreover, if ω is a control such that

$$\left|X^i_{s,t}\right| \leqslant \frac{\omega(s,t)^{\frac{i}{p}}}{\beta\left(\frac{i}{p}\right)!}, \ \forall 1 \leqslant i \leqslant \lfloor p \rfloor \text{ and } \forall (s,t) \in \Delta, \tag{1.2.1}$$

where β is some constant satisfying

$$\beta \geqslant p^2 \left(1 + \sum_{r=3}^{\infty} \left(\frac{2}{r-2}\right)^{\frac{\lfloor p \rfloor + 1}{p}}\right),$$

then (1.2.1) holds for all $i \geqslant \lfloor p \rfloor + 1$.

(2) Let $\mathbf{X}, \mathbf{Y}$ be two p-rough paths, and let β be a constant satisfying

$$\beta \geqslant 2p^2 \left(1 + \sum_{r=3}^{\infty} \left(\frac{2}{r-2}\right)^{\frac{\lfloor p \rfloor + 1}{p}}\right).$$

Suppose that there exists a control ω such that

$$\left|X_{s,t}^i\right|, \ \left|Y_{s,t}^i\right| \leqslant \frac{\omega(s,t)^{\frac{i}{p}}}{\beta \left(\frac{i}{p}\right)!}, \ \forall 1 \leqslant i \leqslant \lfloor p \rfloor \text{ and } \forall (s,t) \in \Delta,$$

and

$$\left|X_{s,t}^i - Y_{s,t}^i\right| \leqslant \varepsilon \frac{\omega(s,t)^{\frac{i}{p}}}{\beta \left(\frac{i}{p}\right)!}, \ \forall 1 \leqslant i \leqslant \lfloor p \rfloor \text{ and } \forall (s,t) \in \Delta, \tag{1.2.2}$$

then (1.2.2) holds for all $i \geqslant \lfloor p \rfloor + 1$.

Definition 1.2.4. $S(\mathbf{X})_{0,1} \in T\left(\left(\mathbb{R}^d\right)\right)$ defined in Theorem 1.2.1 is called the *signature* of the p-rough path $\mathbf{X}$.

If $x : [0,1] \to \mathbb{R}^d$ is a path with finite p-variation for some $1 \leqslant p < 2$, then as a p-rough path no higher levels of x are needed and we can express the signature of x explicitly as (see (1.1.1))

$$S(x)_{0,1} = \left(1, \int_{0<s_1<1} dx_{s_1}, \ldots, \int_{0<s_1<\cdots<s_n<1} dx_{s_1} \otimes \ldots \otimes dx_{s_n}, \ldots\right),$$

where the iterated integrals are defined in the sense of L.C. Young.

There is a special class of rough paths called geometric rough paths. They play a fundamental role in rough path theory and its applications.

Definition 1.2.5. Given $p \geqslant 1$. Let $G\Omega_p\left(\mathbb{R}^d\right)$ denote the completion of the set

$$\Omega_p^{\infty}\left(\mathbb{R}^d\right) := \left\{S_{\lfloor p \rfloor}(x) := \pi^{(\lfloor p \rfloor)}\left(S(x)\right) : \ x \text{ has bounded total variation}\right\}$$

under the p-variation metric d_p. $G\Omega_p(\mathbb{R}^d)$ is called the space of *geometric p-rough paths.*

The importance of geometric rough paths lies in the fact that it is canonical from an analytic point of view, as the construction of the signature naturally arises as iterated path integrals for paths with bounded total variation. Moreover, from a probabilistic point of view, almost surely the sample paths of many interesting stochastic processes (e.g. Brownian motion, Markov processes, martingales, Gaussian processes, under certain conditions), can be regarded as geometric rough paths in a canonical way, or more precisely via piecewise linear approximation. See [26] for a detailed discussion.

The following so-called shuffle product formula implies that polynomial functionals of the signature are essentially linear, which is a crucial feature of the signature. Therefore, by an approximation argument we can see that the structure of certain regular functionals of the signature is rather simple. We refer the reader to [46] for the proof.

Proposition 1.2.1. *Let* $\mathbf{X}$ *be a geometric* p*-rough path for some* $p \geqslant 1$*. For multi-indices* $I = (i_1, \cdots, i_r)$ *and* $J = (j_1, \cdots, j_s)$*, set*

$$(k_1, \cdots, k_{r+s}) = (i_1, \cdots, i_r, j_1, \cdots, j_s).$$

Then we have

$$\mathbf{X}^I \mathbf{X}^J = \sum_{\sigma \in \text{Shuffle}(r,s)} \mathbf{X}^{\left(k_{\sigma^{-1}(1)}, \cdots, k_{\sigma^{-1}(r+s)}\right)},$$

where Shuffle(r, s) *denotes the set of permutations* σ *of order* $r + s$ *such that* $\sigma(1) < \cdots < \sigma(r)$ *and* $\sigma(r+1) < \cdots < \sigma(r+s)$*.*

The fundamental results in rough path theory are continuity theorems for rough path integrals and rough differential equations (written RDEs hereafter) under the p-variation metric. As we pointed out in the last section, from T. Lyons' original point of view the study of RDEs is based on the theory of path integration; an RDE is equivalently regarded as a rough integral equation (a fixed point problem) and the solution is constructed via Picard iteration.

Instead of presenting T. Lyons' original approach, here we follow [26] to adopt a relatively simpler and equivalent formulation which does not rely on the theory of path integration . In fact, path integration is a direct consequence of the theory of RDEs in this respect. The key idea is using the Wong-Zakai type approximation.

The following result, known as the universal limit theorem, asserts the existence, uniqueness and local Lipschitz continuity of solutions to RDEs driven by geometric rough paths in some sense. The definition of solutions to RDEs is contained in the statement of the theorem itself. We refer the reader to [26] for the proof.

Theorem 1.2.2. *(1) Let* $V = (V_1, \cdots, V_d)$ *be a family of* Lip^γ*-vector fields on* $\mathbb{R}^e$ *for some* $\gamma > p$ *(a* Lip^γ*-vector field is a vector field with bounded continuous derivatives of orders up to* $\lfloor \gamma \rfloor$ *and its* $\lfloor \gamma \rfloor$*-th derivative is* $(\gamma - \lfloor \gamma \rfloor)$*-Hölder continuous). For any given* $y_0 \in \mathbb{R}^e$*, define the map*

$$F(y_0, \cdot): \ \Omega_p^\infty\left(\mathbb{R}^d\right) \to G\Omega_p\left(\mathbb{R}^e\right)$$

in the following way. For any $\mathbf{X} \in \Omega_p^\infty\left(\mathbb{R}^d\right)$ *which is the lifting of some path* x *with bounded total variation, let* y *be the unique path in* $\mathbb{R}^e$ *with bounded total variation which is the solution to the ordinary differential equation*

$$dy_t = \sum_{i=1}^d V_i(y_t) dx_t^i, \ t \in [0, 1],$$

with initial value y_0*.* $F(y_0, \mathbf{X})$ *is defined to be the lifting of* y *in* $\Omega_p^\infty\left(\mathbb{R}^e\right) \subset G\Omega_p\left(\mathbb{R}^e\right)$*. Then the map* $F(y_0, \cdot)$ *is uniformly continuous on bounded subsets under the* p*-variation metric. Therefore, it extends uniquely to a continuous map on* $G\Omega_p(\mathbb{R}^d)$*. For given* $\mathbf{X} \in G\Omega_p\left(\mathbb{R}^d\right)$*, the corresponding* $\mathbf{Y} = F(y_0, \mathbf{X})$ *is defined to be the unique solution to the RDE*

$$d\mathbf{Y} = V(\mathbf{Y}) d\mathbf{X}$$

with initial condition y_0*.*

(2) Let $V^{(1)}, V^{(2)}$ *be two families of* Lip^γ*-vector fields on* $\mathbb{R}^e$ *for some* $\gamma > p$*, let* $\mathbf{X}^{(1)}, \mathbf{X}^{(2)}$ *be two geometric* p*-rough paths over* $\mathbb{R}^d$*, and let* $y_0^{(1)}, y_0^{(2)} \in \mathbb{R}^e$*. According to (1), define* $\mathbf{Y}^{(1)} = F\left(y_0^{(1)}, \mathbf{X}^{(1)}\right)$ *and* $\mathbf{Y}^{(2)} = F\left(y_0^{(2)}, \mathbf{X}^{(2)}\right)$ *respectively. Suppose that for some control* ω *we have*

$$\left|X_{s,t}^{(1),i}\right|, \ \left|X_{s,t}^{(2),i}\right| \leqslant \omega(s,t)^{\frac{i}{p}}, \ \forall 1 \leqslant i \leqslant \lfloor p \rfloor \text{ and } (s,t) \in \Delta,$$

and

$$\left|X_{s,t}^{(1),i} - X_{s,t}^{(2),i}\right| \leqslant \varepsilon \omega(s,t)^{\frac{i}{p}}, \ \forall 1 \leqslant i \leqslant \lfloor p \rfloor \text{ and } (s,t) \in \Delta,$$

then

$$\left|Y_{s,t}^{(1),i} - Y_{s,t}^{(2),i}\right| \leqslant C(\left|y_0^{(1)} - y_0^{(2)}\right| + \left\|V^{(1)} - V^{(2)}\right\|_{\mathrm{Lip}^{\gamma-1}} + \varepsilon)\omega(s,t)^{\frac{i}{p}},$$

for all $1 \leqslant i \leqslant \lfloor p \rfloor$ *and* $(s,t) \in \Delta$*, where* C *is some positive constant depending only on* $p, \gamma, \omega(0,T), \left\|V^{(1)}\right\|_{\mathrm{Lip}^\gamma}, \left\|V^{(2)}\right\|_{\mathrm{Lip}^\gamma}$*.*

Remark 1.2.1. In general, the definition of solutions itself does not require such regularity assumptions on the generating vector fields as in Theorem 1.2.2, in which case the first result of Theorem 1.2.2 does not hold any more. However, the underlying key idea of the general formulation is the same as in Theorem 1.2.2, namely the Wong-Zakai type approximation. Here we are not going to present the most general formulation. A detailed presentation can be found in [26].

As mentioned before, rough path integrals can be regarded as a special case of solutions to RDEs.

Let ϕ be an $\mathbb{R}^e$-valued one form on $\mathbb{R}^d$, which can be formally written as

$$\phi = \sum_{i=1}^{d} \phi_i(x) dx^i,$$

where ϕ_i are $\mathbb{R}^e$-valued functions on $\mathbb{R}^d$ $(i = 1, \cdots, d)$. Given a geometric p-rough path $\mathbf{X}$ over $\mathbb{R}^d$, we want to define the rough path integral $\mathbf{Y} = \int \phi(d\mathbf{X})$ as a geometric p-rough path in $\mathbb{R}^e$. The idea is to consider the following RDE:

$$d(\mathbf{Z}, \mathbf{Y}) = V(\mathbf{Z}, \mathbf{Y}) d\mathbf{X} \tag{1.2.3}$$

with initial condition $(z_0, y_0) = (x_0, 0)$, where x_0 is some given point in $\mathbb{R}^d$ being understood as the starting point of $\mathbf{X}$. Here the generating vector fields $V = (V_1, \cdots, V_d)$ are given by

$$V_i(z, y) = \begin{pmatrix} e_i \\ \phi_i(z) \end{pmatrix}, \ i = 1, \cdots, d,$$

where $\{e_1, \cdots, e_d\}$ is the standard basis of $\mathbb{R}^d$. The RDE (1.2.3) is defined for the coupled geometric rough path $(\mathbf{Z}, \mathbf{Y})$ in $\mathbb{R}^d \oplus \mathbb{R}^e$, and its projection onto the $\mathbb{R}^e$-component is then defined to be the rough path integral $\int \phi(d\mathbf{X})$. Under Lip^γ-regularity on the one form ϕ, the existence, uniqueness and local Lipschitz continuity of the integral is a direct consequence of Theorem 1.2.2. We refer the reader to [26] for the details.

Unlike general rough paths, geometric rough paths in fact take values in a much smaller subspace which is a Lie group with very nice analytic structure. It is this special structure that provides powerful tools in the study of geometric rough paths.

We first introduce some algebraic notions.

Given $N \in \mathbb{N}$, let

$$\begin{aligned} \mathfrak{t}^N\left(\mathbb{R}^d\right) &= \left\{g \in T^N\left(\mathbb{R}^d\right) : \pi_0(g) = 0\right\}; \\ 1 + \mathfrak{t}^N\left(\mathbb{R}^d\right) &= \left\{g \in T^N\left(\mathbb{R}^d\right) : \pi_0\left(g\right) = 1\right\}. \end{aligned}$$

We equip $\mathfrak{t}^N\left(\mathbb{R}^d\right)$ with the multiplication defined by

$$[g,h] = g \otimes h - h \otimes g, \ \ g,h \in \mathfrak{t}^N\left(\mathbb{R}^d\right).$$

Respectively, the multiplication on $1 + \mathfrak{t}^N\left(\mathbb{R}^d\right)$ is induced by the tensor product $\otimes$ on $T^N\left(\mathbb{R}^d\right)$. Then we can prove the following result. We refer the reader to [26] for the proof.

Proposition 1.2.2. *The space* $\left(1 + \mathfrak{t}^N\left(\mathbb{R}^d\right), \otimes\right)$ *is a Lie group with manifold topology induced by the Euclidean topology, and the space* $\left(\mathfrak{t}^N\left(\mathbb{R}^d\right), +, [\cdot,\cdot]\right)$ *is a Lie algebra. Moreover,* $\mathfrak{t}^N\left(\mathbb{R}^d\right)$ *is identified with the Lie algebra of the Lie group* $1+\mathfrak{t}^N\left(\mathbb{R}^d\right)$ *under the exponential map* $\exp : \ \mathfrak{t}^N\left(\mathbb{R}^d\right) \to 1 + \mathfrak{t}^N\left(\mathbb{R}^d\right)$ *given by*

$$a \mapsto 1 + \sum_{k=1}^{N} \frac{a^{\otimes k}}{k!}, \ \ a \in \mathfrak{t}^N\left(\mathbb{R}^d\right).$$

Now let $\mathfrak{g}^N\left(\mathbb{R}^d\right)$ be the Lie subalgebra of $\mathfrak{t}^N\left(\mathbb{R}^d\right)$ generated by $\pi_1\left(\mathfrak{t}^N\left(\mathbb{R}^d\right)\right) \cong \mathbb{R}^d$. It is an important result that the Lie group with Lie algebra $\mathfrak{g}^N\left(\mathbb{R}^d\right)$ is exactly the group of truncated signatures. More precisely, let

$$G^N\left(\mathbb{R}^d\right) = \left\{S_N(x)_{0,1} : \ x \text{ has bounded variation}\right\}.$$

Then we have the following result. We refer the reader to [26] for the proof.

Proposition 1.2.3. $G^N\left(\mathbb{R}^d\right) = \exp\left(\mathfrak{g}^N\left(\mathbb{R}^d\right)\right)$, *and* $G^N\left(\mathbb{R}^d\right)$ *is a closed subgroup of* $1 + \mathfrak{t}^N\left(\mathbb{R}^d\right)$.

It follows from the theory of Lie groups (see for example the monograph by F.W. Warner [63]) that there exists a unique manifold structure on $G^N\left(\mathbb{R}^d\right)$ under which it is a Lie subgroup of $1+\mathfrak{t}^N\left(\mathbb{R}^d\right)$. The corresponding manifold topology is the relative topology.

Definition 1.2.6. $G^N\left(\mathbb{R}^d\right)$ is called the *free nilpotent group of step* N over $\mathbb{R}^d$.

It is a remarkable feature of $G^N\left(\mathbb{R}^d\right)$ that it carries a natural norm structure under which it becomes a geodesic space. Recall that a geodesic space is a metric space (E, d) such that for any two points $a, b \in E$, there exists a continuous path $\gamma: [0,1] \to X$ such that $\gamma(0) = a,\ \gamma(1) = b$ and

$$d(\gamma_s, \gamma_t) = |t - s| \cdot d(a, b),\ \forall 0 \leqslant s < t \leqslant 1. \tag{1.2.4}$$

By an obvious rescaling, the geodesic γ can actually be defined on any closed interval $[s, t]$.

For $g \in G^N\left(\mathbb{R}^d\right)$, define

$$\|g\| = \inf\left\{\int_0^1 |dx| :\ x \text{ has bounded total variation and } S_N(x) = g\right\}.$$

Then we have the following result. We refer the reader to [26] for the proof.

Proposition 1.2.4. *(1) For any $g \in G^N\left(\mathbb{R}^d\right)$, there exists some minimizing path x^* with bounded total variation such that $S_N(x^*) = g$ and $\|g\| = \int_0^1 |dx^*|$. Moreover, x^* can be (and will be from now on) parametrized to be Lipschitz and of constant velocity.*

(2) The norm $\|\cdot\|$ induces a left-invariant metric d on $G^N\left(\mathbb{R}^d\right)$ by letting

$$d(g, h) = \|g^{-1} \otimes h\|,\ g, h \in G^N\left(\mathbb{R}^d\right)$$

Under this metric, $G^N\left(\mathbb{R}^d\right)$ becomes a geodesic space, and for $g, h \in G^N\left(\mathbb{R}^d\right)$, a geodesic $\mathbf{X}$ joining g and h is given by

$$\mathbf{X}_t = g \otimes S_N(x^*)_{0,t},\ t \in [0,1],$$

where x^* is a minimizing path associated with $g^{-1} \otimes h$ given by (1).

(3) Let ρ be the induced Euclidean metric on $G^N\left(\mathbb{R}^d\right)$. Then

$$\mathrm{Id}:\ \left(G^N\left(\mathbb{R}^d\right), d\right) \rightleftarrows \left(G^N\left(\mathbb{R}^d\right), \rho\right)$$

is Lipschitz continuous on bounded sets in the "$\rightarrow$" direction and is $1/N$-Hölder continuous on bounded sets in the "$\leftarrow$" direction. In particular, the topology induced by d coincides with the manifold topology.

Definition 1.2.7. The metric d on $G^N\left(\mathbb{R}^d\right)$ is called the *Carnot–Carathéodory metric.*

Given $p \geqslant 1$, it is nature to regard the group $G^{\lfloor p \rfloor}\left(\mathbb{R}^d\right)$ as the state space of geometric rough paths which captures nonlinear higher level increments. This leads to the study of $G^{\lfloor p \rfloor}(\mathbb{R}^d)$-valued continuous paths.

Definition 1.2.8. A continuous path $\mathbf{X}: [0,1] \to G^{\lfloor p \rfloor}\left(\mathbb{R}^d\right)$ starting at the unit $\mathbf{1}$ with finite p-variation under the Carnot–Carathéodory metric, i.e.

$$\|\mathbf{X}\|_p := \sup_{\mathcal{P}_{[0,1]}} \left(\sum_l d\left(\mathbf{X}_{t_{l-1}}, \mathbf{X}_{t_l}\right)^p \right)^{\frac{1}{p}} < \infty,$$

is called a *weakly geometric p-rough path.* The space of weakly geometric p-rough paths is denoted by $WG\Omega_p\left(\mathbb{R}^d\right)$.

Similarly we can introduce the p-variation metric on $WG\Omega_p\left(\mathbb{R}^d\right)$ based on the Carnot–Carathéodory metric. For $\mathbf{X}, \mathbf{Y} \in WG\Omega_p\left(\mathbb{R}^d\right)$, define

$$\overline{d}_p(\mathbf{X}, \mathbf{Y}) = \left(\sup_{\mathcal{P}_{[0,1]}} \sum_l d\left(\mathbf{X}_{t_{l-1},t_l}, \mathbf{Y}_{t_{l-1},t_l}\right)^p \right)^{\frac{1}{p}},$$

where $\mathbf{X}_{t_{l-1},t_l} := \mathbf{X}_{t_{l-1}}^{-1} \otimes \mathbf{X}_{t_l}$ and similarly for $\mathbf{Y}$. It can also be proved (see [26]) that $\left(WG\Omega_p\left(\mathbb{R}^d\right), \overline{d}_p\right)$ is a complete metric space.

Given $\mathbf{X} \in WG\Omega_p\left(\mathbb{R}^d\right)$, by setting $\mathbf{X}_{s,t} = \mathbf{X}_s^{-1} \otimes \mathbf{X}_t$ for $(s,t) \in \Delta$, we can regard $\mathbf{X}$ as a multiplicative functional of degree $\lfloor p \rfloor$. Therefore, the p-variation metric d_p defined at the beginning can be applied to the space $WG\Omega_p\left(\mathbb{R}^d\right)$. It is an important consequence of Proposition 1.2.4 (3) that d_p and $\overline{d}_p$ are comparable on bounded sets. More precisely, we have the following result. We refer the reader to [26] for the proof.

Proposition 1.2.5. *A continuous path* $\mathbf{X}: [0,1] \to G^{\lfloor p \rfloor}\left(\mathbb{R}^d\right)$ *starting at the unit* $\mathbf{1}$ *is a weakly geometric p-rough path if and only if it is a p-rough path. Moreover, the identity map*

$$\mathrm{Id}: \left(WG\Omega_p\left(\mathbb{R}^d\right), \overline{d}_p\right) \rightleftarrows \left(WG\Omega_p\left(\mathbb{R}^d\right), d_p\right)$$

is Lipschitz continuous on bounded sets in the "$\to$" direction and is $1/\lfloor p \rfloor$*-Hölder continuous on bounded sets in the "$\leftarrow$" direction. In particular, the notion of convergence in* $WG\Omega_p\left(\mathbb{R}^d\right)$ *under* d_p *and* $\overline{d}_p$ *are equivalent.*

From Proposition 1.2.5, it is obvious that a geometric p-rough path $\mathbf{X}$ actually takes values in $G^{\lfloor p \rfloor}\left(\mathbb{R}^d\right)$ if we set $\mathbf{X}_t = \mathbf{X}_{0,t}$ for $t \in [0,1]$. Therefore, $G\Omega_p\left(\mathbb{R}^d\right)$ can be equivalently regarded as a subspace of $WG\Omega_p\left(\mathbb{R}^d\right)$. A nontrivial and important fact is that weakly geometric rough paths are almost indistinguishable from geometric rough paths. In fact, we have the following result. We refer the reader to [26] for the proof.

Proposition 1.2.6. *Let* $\mathbf{X}$ *be a weakly geometric* p*-rough path.*

(1) When regarded as a multiplicative functional of degree $\lfloor p \rfloor$, $\mathbf{X}$ *is a geometric* q*-rough path for any* $p < q < \lfloor p \rfloor + 1$.

(2) $\mathbf{X}$ *is a geometric* p*-rough path if and only if*

$$\lim_{\|\mathcal{P}\| \to 0} \overline{d}_p \left(\mathbf{X}^{\mathcal{P}}, \mathbf{X} \right) = 0,$$

where $\mathbf{X}^{\mathcal{P}}$ *denotes a piecewise geodesic interpolation of* $\mathbf{X}$ *over the partition points in* $\mathcal{P}$.

Let us point out that the finite dimensional rough path theory can be equivalently formulated under the framework of weakly geometric rough paths, and many important results such as Lyons' extension theorem and the universal limit theorem can be proved in an equivalent way in this setting. See [26] for a systematic presentation using this approach.

To end this section, we remark that in the previous formulations, it is of no importance whether we require all paths to be defined on the unit interval $[0, 1]$ or on any arbitrary interval $[s, t]$. In fact, the whole theory is invariant under reparametrization in the general sense (for example, the signature is invariant). As we will use the concept of reparametrization several times, we give its formal definition in the following.

Definition 1.2.9. A *reparametrization* in the strict sense from $[a, b]$ to $[c, d]$ is a continuous, strictly increasing map $\sigma : [a, b] \to [c, d]$ with $\sigma(a) = c$ and $\sigma(b) = d$. If σ is just increasing (i.e. non-decreasing), we call σ a reparametrization in the general sense.

For the study of the uniqueness of signature problem, we restrict ourselves to reparametrizations in the strict sense. In this thesis, unless otherwise stated, a reparametrization is understood in this sense.

Chapter 2

Simple Piecewise Geodesic Interpolation of Simple and Jordan Curves with Applications

2.1 Introduction

The classical proofs of many properties of Jordan curves (e.g. the Jordan curve theorem) or functions on Jordan curves (e.g. Cauchy's theorem) begin with consideration of the case of polygonal Jordan curves. As part of the proof of the Jordan curve theorem in H. Tverberg [62], it was shown that for every planar Jordan curve, there is a polygonal Jordan curve that approximates the original Jordan curve arbitrarily well. We begin this chapter by proving a stronger and more general fact that given a Jordan curve on a connected Riemannian manifold M and n points on the curve, there exists a simple, piecewise minimizing geodesic, arbitrarily fine interpolation which includes these n points as interpolation points. Its proof relies on another main result of this chapter for non-closed simple curves. This case was first treated by Werness [64], in which the author used an inductive proof which is not constructive. Here we provide another proof of this result which has the advantage of being explicit and constructive.

We would like to emphasize that our approximation, unlike that of [62] (which is then a direct consequence of our result), does not rely on the flatness of the Euclidean metric or the regularity of the curve. Moreover it respects the parametrization of the curve, i.e. it is an interpolation rather than merely an approximation in the uniform norm. This is particularly important for applications in the context of rough path theory, where we approximate continuous paths by bounded total variation ones in the p-variation metric. Such an idea is fundamental to study the roughness of continuous paths, and particularly of sample paths of continuous stochastic processes. We will come back to this point in the last two chapters.

We then give two applications of our main result for Jordan curves.

By taking advantage of the result that the p-variation of the piecewise linear interpolation of a path is bounded by the p-variation of the path itself, our approximation theorem yields immediately a generalized Green's theorem for planar Jordan curves with finite p-variation, where $1 \leqslant p < 2$. To the best of our knowledge, in the rough path literature, the only other attempt so far in extending Green's theorem to non-rectifiable curves appeared in P. Yam [67], where Green's theorem was proved for the boundaries of α-Hölder domains for $\frac{1}{3} < \alpha < 1$. Our result is a partial generalization of P. Yam's because P. Yam's result requires the curve to be α-Hölder under the conformal parametrization whereas our result only requires the curve to be α-Hölder under some parametrization. This difference lies in the fact that P. Yam used the conformal map from the unit disk to the interior of the Jordan curve to construct the approximation.

Another application of our main result is a solution to the uniqueness of signature problem for planar Jordan curves with finite p-variation for $1 \leqslant p < 2$. Since Jordan curves are highly non-degenerate, it is natural to expect that the curve is determined by its signature up to reparametrization in this case. In fact, we will see that with the aid of our main result, this can be proved by using the same technique as in [6] for planar simple curves (with the same regularity). It should be pointed out that under such regularity, the path is actually "smooth" from the view of rough path theory as no higher level increments are needed in this case and the signature is uniquely defined in the sense of L.C. Young as iterated path integrals. Of course the uniqueness of signature problem is nontrivial, even for the case when the path is really smooth.

Throughout the rest of this chapter, all curves are assumed to be continuous.

2.2 Simple Piecewise Geodesic Interpolation of Simple and Jordan Curves

In this section, we prove our main results about simple piecewise geodesic approximation of simple and Jordan curves in Riemannian manifolds. Although the most interesting and nontrivial case lies in the Euclidean plane, we formulate the problems in a Riemannian geometric setting of arbitrary dimension since our proofs do not rely on Euclidean geometry (i.e. the "flatness" of the Euclidean metric) at all.

Throughout this section, let M be a d-dimensional connected Riemannian manifold ($d \geqslant 2$).

The following lemma, which is an easy fact from Riemannian geometry, is essential for us to formulate our main results.

Lemma 2.2.1. *For any compact set $K \subset M$, there exists some $\varepsilon = \varepsilon_K > 0$, such that for any $x, y \in K$ with*

$$d(x, y) < \varepsilon,$$

there exists a unique minimizing geodesic in M joining x and y, where $d(\cdot, \cdot)$ denotes the Riemannian distance function.

Proof. For any $x \in K$, choose δ_x small enough such that $B(x, \delta_x)$ is a geodesically convex normal ball (see the monograph by M.P. do Carmo [21], Chapter 3, Proposition 4.2). By compactness, we have a finite covering of K :

$$K \subset \bigcup_{i=1}^{k} B\left(x_i, \frac{\delta_{x_i}}{2}\right),$$

where $x_1, \cdots, x_k \in K$. Let $\varepsilon = \frac{1}{2}\min\{\delta_{x_1}, \cdots, \delta_{x_k}\}$. It follows that for any $x, y \in K$ with $d(x, y) < \varepsilon$, there exists some $1 \leqslant i \leqslant k$, such that $x, y \in B(x_i, \delta_{x_i})$. Therefore, by geodesic convexity we know that x and y can be joined by a unique minimizing geodesic in M which lies in $B(x_i, \delta_{x_i})$. □

Now we are in a position to state our main results.

The first main result is a simple piecewise geodesic approximation theorem for non-closed simple curves in M.

Theorem 2.2.1. *Let γ be a non-closed simple curve in M. Then for all $\varepsilon > 0$, there exists a finite partition*

$$\mathcal{P}_\varepsilon : \ 0 = t_0 < t_1 < \cdots < t_{n-1} < t_n = 1$$

of $[0, 1]$, such that

(1) the mesh size of the partition $\|\mathcal{P}_\varepsilon\| = \max_{i=1,\cdots,n}(t_i - t_{i-1}) < \varepsilon$;

(2) for any $i = 1, \cdots, n$, $\gamma_{t_{i-1}}$ and γ_{t_i} can be joined by a unique minimizing geodesic in M, and the piecewise geodesic interpolation (more precisely, piecewise minimizing

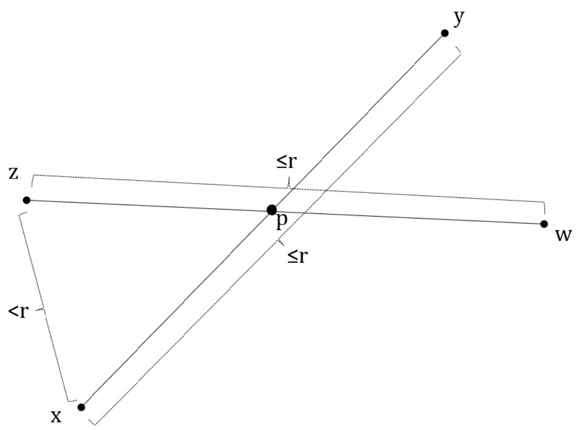

Figure 2.2.1: This figure illustrates the relative positions of the points in Lemma 2.2.2 in the Euclidean case. The lengths of the line segments $\overline{xy}$ and $\overline{zw}$ are less than or equal to r. Here the length of $\overline{zx}$ is strictly less than r.

geodesic interpolation, and the same thereafter) $\gamma^{\mathcal{P}_\varepsilon}$ *of* γ *over the partition points in* $\mathcal{P}_\varepsilon$ *is a simple curve.*

The proof of Theorem 2.2.1 relies on the following crucial lemma, which depends heavily on properties of minimizing geodesics. In the Euclidean case, we illustrate the lemma in Figure 2.2.1, which says that if the length of the straight line segments $\overline{xy}$ and $\overline{zw}$ are both less than or equal to r, then at least one of the four line segments $\overline{xv}, \overline{xw}, \overline{yv}, \overline{yw}$ has length strictly less than r.

Lemma 2.2.2. *Let* $x, y, z, w \in M$ *and* $\alpha : [0,1] \to M$ *(respectively,* $\beta : [0,1] \to M$*) be a minimizing geodesic joining* x *and* y *(respectively,* z *and* w*). Assume that* $\alpha([0,1]) \bigcap \beta([0,1]) \neq \emptyset$ *and for some* $r > 0$, $d(x,y) \leqslant r$, $d(z,w) \leqslant r$*. Then at least one of* $d(x,z), d(y,z), d(x,w), d(y,w)$ *is strictly less than* r*.*

Proof. Let $\alpha(u) = \beta(v) = p$ for some $u, v \in [0,1]$. Since α and β are minimizing geodesics, we know that

$$\begin{aligned} d(x,y) = d(x,p) + d(p,y) &\leqslant r, \\ d(z,w) = d(z,p) + d(p,w) &\leqslant r. \end{aligned}$$

Therefore, at least one of the following four cases happens:

(1) $d(x,p) \leqslant \frac{r}{2},\ d(z,p) \leqslant \frac{r}{2}$;
(2) $d(x,p) \leqslant \frac{r}{2},\ d(p,w) \leqslant \frac{r}{2}$;
(3) $d(p,y) \leqslant \frac{r}{2},\ d(z,p) \leqslant \frac{r}{2}$;

(4) $d\left(p,y\right) \leqslant \frac{r}{2}$; $d\left(p,w\right) \leqslant \frac{r}{2}$.

First assume that Case (1) holds. It follows that

$$d\left(x,z\right) \leqslant d\left(x,p\right) + d\left(z,p\right) \leqslant r.$$

If $d\left(x,z\right) = r$, then

$$d(x,p) = d(z,p) = \frac{r}{2},$$

and hence Case (4) holds, which implies

$$d\left(y,w\right) \leqslant d\left(p,y\right) + d\left(p,w\right) \leqslant r.$$

If $d\left(y,w\right) = r$, then

$$d\left(p,y\right) = d\left(p,w\right) = \frac{r}{2}.$$

Consequently, we have $u = v = \frac{1}{2}$.

Now define

$$\widetilde{\alpha}\left(t\right) = \begin{cases} \alpha\left(t\right), & t \in \left[0, \frac{1}{2}\right]; \\ \beta\left(1-t\right), & t \in \left[\frac{1}{2}, 1\right]. \end{cases}$$

Since

$$\text{Length}\left(\widetilde{\alpha}\right) = r = d\left(x,z\right),$$

$\widetilde{\alpha}$ is minimizing. Moreover, since any geodesic has constant speed, by definition we know that $\widetilde{\alpha}$ is parametrized proportionally to arc length. It follows from the first variation formula (see [21], Chapter 9, Proposition 2.4) that $\widetilde{\alpha}$ must be a geodesic. However, since $\widetilde{\alpha}\Big|_{[0,\frac{1}{2}]} = \alpha\Big|_{[0,\frac{1}{2}]}$, by the uniqueness of geodesics we have $\widetilde{\alpha} = \alpha$ and hence $y = z$. Similarly we have $x = w$.

The other cases can be treated in the same way, which completes the proof of the lemma. □

With the help of Lemma 2.2.2, we can now prove Theorem 2.2.1. The key idea is to construct a sequence of times $t_1, t_2, \ldots$ so that t_{i+1} is the *last* exit time of γ for a small geodesic ball around γ_{t_i} after time t_i. The uniform continuity of the inverse of the map $t \to \gamma_t$ guarantees that t_i and t_{i+1} are close. We then need to argue that adjacent geodesic segments as well as non-adjacent geodesic segments in the approximation curve do not intersect. The latter uses Lemma 2.2.2. We illustrate the first step of the construction in Figure 2.2.2.

Proof of Theorem 2.2.1. Fix $\varepsilon > 0$. Since γ is a continuous and injective map from

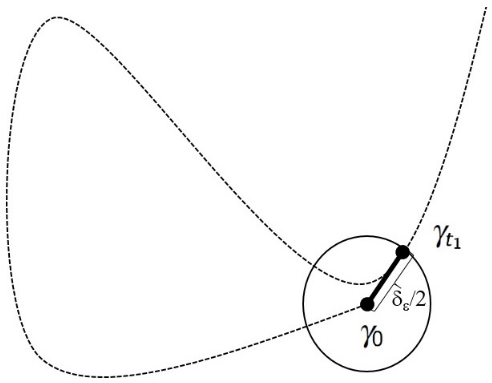

Figure 2.2.2: This figure illustrates the first step in the construction given in the proof of Theorem 2.2.1. The dotted line represents the simple curve γ. The solid geodesic segment joining the point γ_0 and γ_{t_1} represents the first step in the construction of the piecewise geodesic interpolation of γ. Note that we take t_1 to be the last exit time of γ in a $\frac{\delta_\varepsilon}{2}$-geodesic ball centered at γ_0.

the compact space $[0,1]$ to the Hausdorff space M, it is a homeomorphism from $[0,1]$ to its image. By compactness and hence uniform continuity of γ^{-1} we know that there exists $\delta_\varepsilon > 0$ such that for any $s, t \in [0,1]$,

$$d(\gamma_s, \gamma_t) < \delta_\varepsilon \implies |t - s| < \varepsilon.$$

We further assume that $\delta_\varepsilon < \varepsilon_{\gamma([0,1])}$, where $\varepsilon_{\gamma([0,1])}$ is the positive number in Lemma 2.2.1 depending on the compact set $\gamma([0,1]) \subset M$. It follows from Lemma 2.2.1 that for any $s, t \in [0,1]$ with $d(\gamma_s, \gamma_t) < \delta_\varepsilon$, γ_s and γ_t can be joined by a unique minimizing geodesic in M. Now define an increasing sequence of points $\{t_i\}_{i=0}^{\infty}$ in $[0,1]$ inductively by setting $t_0 = 0$ and

$$t_i = \sup \left\{ t \in [t_{i-1}, 1] : \ \gamma_t \in \overline{B}\left(\gamma_{t_{i-1}}, \frac{\delta_\varepsilon}{2}\right) \right\}, \ i \geqslant 1.$$

We claim that there exists some $l \geqslant 1$, such that for all $i \geqslant l$, $t_i = 1$. In fact, if it is not the case, then for any $i \geqslant 1$, we have

$$t_{i-1} < t_i < 1 \text{ and } d\left(\gamma_{t_{i-1}}, \gamma_{t_i}\right) = \frac{\delta_\varepsilon}{2}.$$

On the other hand, by the uniform continuity of γ, there exists some $\eta_\varepsilon > 0$, such that for any $s, t \in [0,1]$,

$$|t-s| < \eta_\varepsilon \implies d(\gamma_s, \gamma_t) < \frac{\delta_\varepsilon}{2}.$$

Therefore, for any $i \geqslant 1$, $|t_i - t_{i-1}| \geqslant \eta_\varepsilon$, which is an obvious contradiction. Now set

$$l = \min\{i \geqslant 1 : \ t_i = 1\},$$

and define

$$\mathcal{P}_\varepsilon : \ 0 = t_0 < t_1 < \cdots < t_{l-1} < t_l = 1$$

to be a finite partition of $[0,1]$. Then it is easy to see that $\|\mathcal{P}_\varepsilon\| < \varepsilon$, where $\|\mathcal{P}_\varepsilon\|$ denote the mesh size of the partition $\mathcal{P}_\varepsilon$.

It remains to show that the piecewise geodesic interpolation $\gamma^{\mathcal{P}_\varepsilon}$ of γ over the points of $\mathcal{P}_\varepsilon$ is a simple curve.

To see this, first notice that for adjacent intervals $[t_{i-1}, t_i]$, $[t_i, t_{i+1}]$, we have

$$\gamma^{\mathcal{P}_\varepsilon}|_{[t_{i-1},t_i]} \bigcap \gamma^{\mathcal{P}_\varepsilon}|_{[t_i,t_{i+1}]} = \{\gamma_{t_i}\}.$$

In fact, if it is not the case, then there exist $s_1 \in [t_{i-1}, t_i)$ and $s_2 \in (t_i, t_{i+1}]$ such that

$$\gamma^{\mathcal{P}_\varepsilon}_{s_1} = \gamma^{\mathcal{P}_\varepsilon}_{s_2} \neq \gamma_{t_i}.$$

If $i < l-1$, then by applying Lemma 2.2.1 with $x = \gamma_{t_i}$ and $y = \gamma^{\mathcal{P}_\varepsilon}_{s_1} = \gamma^{\mathcal{P}_\varepsilon}_{s_2}$, $\gamma^{\mathcal{P}_\varepsilon}|_{[s_1,t_i]}$ is a reparametrization of the reversal of $\gamma^{\mathcal{P}_\varepsilon}|_{[t_i,s_2]}$, which we denote as $\overleftarrow{\gamma^{\mathcal{P}_\varepsilon}}|_{[t_i,s_2]}$. In particular, $\gamma^{\mathcal{P}_\varepsilon}|_{[t_i,t_{i+1}]}$ and $\overleftarrow{\gamma^{\mathcal{P}_\varepsilon}}|_{[t_{i-1},t_i]}$ are geodesics that start at the same position with the same initial velocity. By the uniqueness of geodesics, either $\gamma^{\mathcal{P}_\varepsilon}([t_i, t_{i+1}]) \subseteq \overleftarrow{\gamma^{\mathcal{P}_\varepsilon}}([t_{i-1}, t_i])$ or $\overleftarrow{\gamma^{\mathcal{P}_\varepsilon}}([t_{i-1}, t_i]) \subseteq \gamma^{\mathcal{P}_\varepsilon}([t_i, t_{i+1}])$. In particular we have either $\gamma^{\mathcal{P}_\varepsilon}|_{[t_{i-1},t_i]}$ passes through $\gamma_{t_{i+1}}$ or $\gamma^{\mathcal{P}_\varepsilon}|_{[t_i,t_{i+1}]}$ passes through $\gamma_{t_{i-1}}$. As $\overleftarrow{\gamma^{\mathcal{P}_\varepsilon}}|_{[t_{i-1},t_i]}$ and $\gamma^{\mathcal{P}_\varepsilon}|_{[t_i,t_{i+1}]}$ are minimizing geodesics and we have $d\left(\gamma_{t_i}, \gamma_{t_{i-1}}\right) = d\left(\gamma_{t_i}, \gamma_{t_{i+1}}\right)$, we conclude that $\gamma_{t_{i-1}} = \gamma_{t_{i+1}}$ which contradicts that γ is simple. Figure 2.2.3 illustrates this argument.

If $i = l-1$, then arguing as in the case $i < l-1$, we have either $\gamma^{\mathcal{P}_\varepsilon}([t_i, t_{i+1}]) \subseteq \overleftarrow{\gamma^{\mathcal{P}_\varepsilon}}([t_{i-1}, t_i])$ or $\overleftarrow{\gamma^{\mathcal{P}_\varepsilon}}([t_{i-1}, t_i]) \subseteq \gamma^{\mathcal{P}_\varepsilon}([t_i, t_{i+1}])$. However, as $i = l-1$, we have $d\left(\gamma_{t_i}, \gamma_{t_{i+1}}\right) \leqslant \frac{\delta_\varepsilon}{2} = d\left(\gamma_{t_{i-1}}, \gamma_{t_i}\right)$ and hence $\gamma^{\mathcal{P}_\varepsilon}([t_i, t_{i+1}]) \subseteq \overleftarrow{\gamma^{\mathcal{P}_\varepsilon}}([t_{i-1}, t_i])$. In particular, $\gamma^{\mathcal{P}_\varepsilon}|_{[t_{i-1},t_i]}$ passes through $\gamma_{t_{i+1}}$. Therefore, $d\left(\gamma_{t_{i-1}}, \gamma_{t_{i+1}}\right) \leqslant d\left(\gamma_{t_{i-1}}, \gamma_{t_i}\right)$ which contradicts the construction of $\{t_i\}_{i=0}^{l}$.

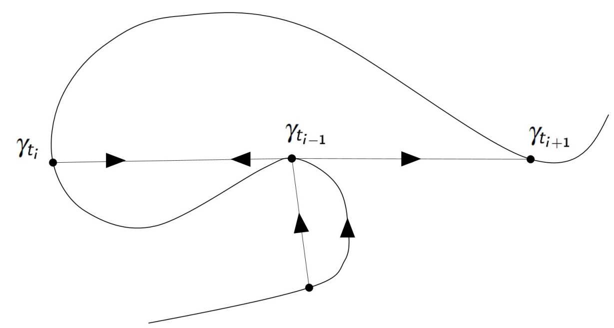

Figure 2.2.3: This figure illustrates the argument in the proof of Theorem 2.2.1 that two adjacent line segment of the approximation curve we constructed cannot intersect. The straight line represents the geodesic segments in the piecewise geodesic interpolation of γ. $\gamma_{t_{i-1}}, \gamma_{t_i}, \gamma_{t_{i+1}}$ are subdivision points of the curve. If the two adjacent line segments do intersect as in the figure, then $\gamma_{t_{i+1}}$ is closer to $\gamma_{t_{i-1}}$ than to γ_{t_i} which contradicts our construction.

On the other hand, if $[t_{i-1}, t_i]$ and $[t_{j-1}, t_j]$ $(i < j)$ are non-adjacent intervals and

$$\gamma^{\mathcal{P}_\varepsilon}|_{[t_{i-1},t_i]} \bigcap \gamma^{\mathcal{P}_\varepsilon}|_{[t_{j-1},t_j]} \neq \emptyset,$$

then by Lemma 2.2.2 we know that at least one of

$$d\left(\gamma_{t_{i-1}}, \gamma_{t_{j-1}}\right), d\left(\gamma_{t_i}, \gamma_{t_{j-1}}\right), d\left(\gamma_{t_{i-1}}, \gamma_{t_j}\right), d\left(\gamma_{t_i}, \gamma_{t_j}\right)$$

is strictly less than $\frac{\delta_\varepsilon}{2}$. However, this again contradicts the construction of $\{t_i\}_{i=0}^{l}$.

Now the proof is complete. □

The previous technique is crucial for us to prove our second main result, which is concerned with simple piecewise geodesic approximation of Jordan curves. This result significantly strengthens Theorem 2.2.1.

Theorem 2.2.2. *Let* $\gamma: [0,1] \to M$ *be a Jordan curve. Assume that* $0 < \tau_1 < \cdots < \tau_k < 1$ *are* k *fixed points in* $[0,1]$. *Then for any* $\varepsilon > 0$, *there exists a finite partition*

$$\mathcal{P}_\varepsilon: \ 0 = t_0 < t_1 < \cdots < t_{n-1} < t_n = 1$$

of $[0,1]$, *such that*

(1) $\tau_1, \cdots, \tau_k$ *are partition points of* $\mathcal{P}_\varepsilon$*;*

(2) $\|\mathcal{P}_\varepsilon\| < \varepsilon$*;*

(3) for $i = 1, \cdots, n$, $\gamma_{t_{i-1}}$ *and* γ_{t_i} *can be joined by a unique minimizing geodesic in* M, *and the piecewise geodesic interpolation* $\gamma^{\mathcal{P}_\varepsilon}$ *of* γ *over the partition points in*

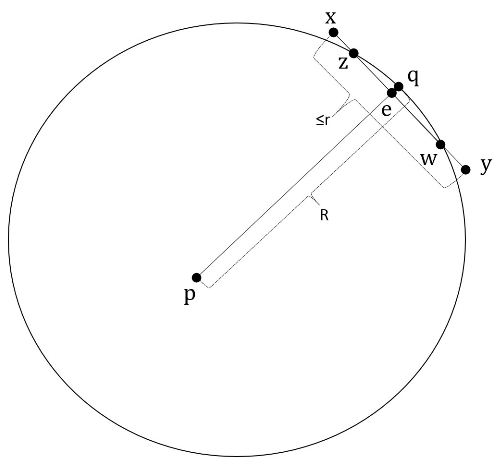

Figure 2.2.4: This figure illustrates the relative positions of the points involved in Lemma 2.2.3.

$\mathcal{P}_\varepsilon$ is a Jordan curve.

The proof of Theorem 2.2.2 relies on the following geometric fact. It is illustrated by Figure 2.2.4.

Lemma 2.2.3. *Let $B(p,R)$ be a geodesically convex normal ball centered at $p \in M$, and let $q \in \partial B(p,R)$. Assume that $x, y \in \overline{B}(p,R)^c$ and there exists a minimizing geodesic $\alpha:\ [0,1] \to M$ joining x and y. If $\alpha([0,1]) \bigcap \overline{pq} \neq \emptyset$ and $d(x,y) \leqslant r$ for some $0 < r < R$, where $\overline{pq}$ denotes the image of the unique minimizing geodesic in M joining p and q, then*

$$d(x,q) < r,\ d(y,q) < r.$$

Proof. The conclusion is obvious if $q \in \alpha([0,1])$. Otherwise, let $t \in (0,1)$ be the unique time such that $e := \alpha(t) \in B(p,R)$ is the intersection point of $\alpha([0,1])$ and $\overline{pq}$. By using the fact that $B(p,R)$ is a geodesically convex normal ball, it is easy to show that there exists a unique $u \in (0,t)$ and a unique $v \in (t,1)$, such that $z := \alpha(u)$ and $w := \alpha(v)$ lie on $\partial B(p,R)$. Observe that e and p are distinct, since their equality contradicts the fact that $r < R$. Now it follows from properties of

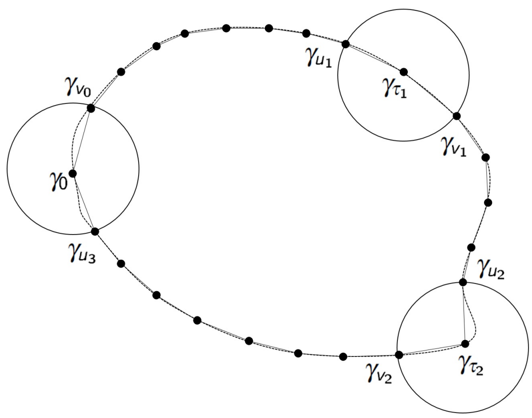

Figure 2.2.5: This figure illustrates the idea of proving of Theorem 2.2 when $k = 2$. The dotted line represents the curve γ. The solid line represents the piecewise geodesic interpolation of γ.

minimizing geodesics that

$$\begin{aligned} d\,(x,q) &\leqslant d\,(x,e) + d\,(e,q) \\ &= d\,(x,e) + d\,(p,q) - d\,(p,e) \\ &= d\,(x,e) + d\,(p,w) - d\,(p,e) \\ &\leqslant d\,(x,e) + d\,(e,w) \\ &= d\,(x,w) \\ &< d\,(x,y) \\ &\leqslant r. \end{aligned}$$

Similarly, we have $d\,(y,q) < r$. □

Now we can prove Theorem 2.2.2. Our proof is constructive and the idea is as follows. Recall that the times $\tau_1, \ldots, \tau_k$ should to included in our partition. Firstly, We find small disjoint geodesic balls around the points $\gamma_{\tau_i}, \ldots, \gamma_{\tau_k}, \gamma_1$. Secondly, we connect each point γ_{τ_i} by two radial minimizing geodesics to the point where γ *first* enters the geodesic ball around γ_{τ_i} before time τ_i and to the point where γ *last* exists the geodesic ball. Finally, we construct a simple piecewise geodesic interpolation for each piece of simple curve outside those geodesic balls inductively, by using the algorithm in Theorem 2.2.1. To make sure that those approximation curves do not intersect the geodesic segments inside those geodesic balls, we need to use Lemma 2.2.3. Figure 2.2.5 illustrates the idea when $k = 2$.

Proof of Theorem 2.2.2. Take an arbitrary $\tau \in (0, \tau_1)$. Since γ is a Jordan curve, we know that $\gamma_\tau, \gamma_{\tau_1}, \cdots, \gamma_{\tau_k}, \gamma_{\tau_{k+1}} \in M$ are all distinct, where we set $\tau_{k+1} = 1$. By the Hausdorff property, there exists some $\delta > 0$ such that the closed metric balls $\overline{B}(\gamma_{\tau_1}, \delta), \cdots, \overline{B}(\gamma_{\tau_{k+1}}, \delta)$ are all disjoint and $\gamma_\tau \notin \bigcup_{i=1}^{k+1} \overline{B}(\gamma_{\tau_i}, \delta)$.

For the moment, by periodic extension and restriction we regard γ as defined on $[\tau, \tau + 1]$ with starting and end points being $\gamma(\tau)$.

Now fix $\varepsilon > 0$. Without loss of generality we assume that

$$\varepsilon < \min \left\{ \tau, \tau_1 - \tau, \frac{\tau_2 - \tau_1}{2}, \cdots, \frac{\tau_{k+1} - \tau_k}{2} \right\}.$$

First of all, by the uniform continuity of $\gamma|_{[\tau,\tau_i]}^{-1}$ and $\gamma|_{[\tau_i,\tau+1]}^{-1}$, there exists some $\delta_\varepsilon > 0$, such that for all $i = 1, \cdots, k+1$, any $s, t \in [\tau, \tau_i]$ or $s, t \in [\tau_i, \tau + 1]$,

$$d(\gamma_s, \gamma_t) < \delta_\varepsilon \implies |t - s| < \varepsilon.$$

Now set $U_i = B(\gamma_{\tau_i}, \delta_\varepsilon)$. Here we assume that δ_ε is small enough so that each U_i is a geodesically convex normal ball and Lemma 2.2.1 holds for those γ_s, γ_t with $d(\gamma_s, \gamma_t) < 2\delta_\varepsilon$. Define

$$\begin{aligned} u_i &= \inf \left\{ t \in [\tau, \tau_i] : \ \gamma_t \in \overline{U_i} \right\}, \\ v_i &= \sup \left\{ t \in [\tau_i, \tau + 1] : \ \gamma_t \in \overline{U_i} \right\}. \end{aligned}$$

To return to the original time interval $[0, 1]$, let $v_0 = v_{k+1} - 1$. We have $|v_0| < \varepsilon$, $|\tau_i - u_i| < \varepsilon$, $|v_i - \tau_i| < \varepsilon$ and

$$0 < v_0 < u_1 < \tau_1 < v_1 < \cdots < u_k < \tau_k < v_k < u_{k+1} < 1,$$

and

$$\gamma_{u_i} \neq \gamma_{v_i}, \ d(\gamma_{\tau_i}, \gamma_{u_i}) = d(\gamma_{\tau_i}, \gamma_{v_i}) = \delta_\varepsilon.$$

Moreover, we have

$$\gamma|_{(v_0,u_1) \bigcup (v_1,u_2) \bigcup \cdots \bigcup (v_{k-1},u_k) \bigcup (v_k,u_{k+1})} \bigcap \left(\bigcup_{i=1}^{k+1} \overline{U_i} \right) = \emptyset.$$

We take $v_0, u_1, \tau_1, v_1, \cdots, u_k, \tau_k, v_k, u_{k+1}$ as part of the partition points in $\mathcal{P}_\varepsilon$. In particular, v_0 is the first point, u_{k+1} is the last point (except 0 and 1), and u_i, τ_i, v_i are successive points in $\mathcal{P}_\varepsilon$, so the piecewise geodesic interpolation of γ over those small intervals is a finite sequence of radial geodesics of the balls centered at γ_{τ_i} with radius δ_ε for $i = 1, \cdots, k+1$.

For the next step, notice that $\gamma|_{[v_0,u_1]}, \gamma|_{[v_1,u_2]}, \cdots, \gamma|_{[v_k,u_{k+1}]}$ are $k+1$ non-closed simple curves with disjoint images. We use the constructive procedure in the proof of Theorem 2.2.1 to define a simple piecewise geodesic approximation of each $\gamma|_{[v_{i-1},u_i]}$ ($i = 1, \cdots, k+1$) with partition size smaller than ε inductively, such that the resulting piecewise geodesic closed curve over $[0,1]$ is Jordan.

Let $\gamma^{(0)}$ be the Jordan curve such that

$$\gamma^{(0)} = \gamma, \text{ on } [v_0, u_1] \bigcup [v_1, u_2] \bigcup \cdots \bigcup [v_{k-1}, u_k] \bigcup [v_k, u_{k+1}],$$

and it is the minimizing geodesic (radial segment of the corresponding normal ball) on each small interval of

$$[0, v_0], [u_1, \tau_1], [\tau_1, v_1], \cdots, [u_k, \tau_k], [\tau_k, v_k], [u_{k+1}, 1].$$

By the construction in the proof of Theorem 2.2.1, we may find a partition

$$\mathcal{P}^{(1)}_{[v_0,u_1]}: \; v_0 = w^{(1)}_0 < w^{(1)}_1 < \cdots < w^{(1)}_{l_1-1} < w^{(1)}_{l_1} = u_1$$

so that $\left\|\mathcal{P}^{(1)}_{[v_0,u_1]}\right\| < \varepsilon$, the geodesic interpolation $\gamma^{\mathcal{P}^{(1)}_{[v_0,u_1]}}$ of $\gamma|_{[v_0,u_1]}$ over the partition points in $\mathcal{P}^{(1)}_{[v_0,u_1]}$ is simple and

$$\begin{aligned} d\left(\gamma_{w^{(1)}_{i-1}}, \gamma_{w^{(1)}_i}\right) &= \delta^{(1)}_\varepsilon, \; i = 1, \cdots l_1 - 1, \\ d\left(\gamma_{w^{(1)}_{l_1-1}}, \gamma_{u_1}\right) &\leqslant \delta^{(1)}_\varepsilon, \end{aligned}$$

for some $\delta^{(1)}_\varepsilon > 0$.

Moreover, we may choose $\delta^{(1)}_\varepsilon$small enough so that $\operatorname{dist}\left(\gamma^{\mathcal{P}^{(1)}_{[v_0 u_1]}}, \gamma^{(0)}|_{[\tau_1,1]}\right) > 0$ and $\delta^{(1)}_\varepsilon < \delta_\varepsilon$.

Now we show that

$$\gamma^{\mathcal{P}^{(1)}_{[v_0,u_1]}} \bigcap \gamma^{(0)}|_{[0,v_0) \bigcup (u_1,\tau_1]} = \emptyset.$$

In fact, if $\gamma^{\mathcal{P}^{(1)}_{[v_0,u_1]}} \bigcap \gamma^{(0)}|_{[0,v_0)} \neq \emptyset$, then from the construction of $\left\{w^{(1)}_i\right\}$, there exists some $i \geqslant 2$, such that $\gamma_{w_{i-1}^{(1)}}, \gamma_{w_i^{(1)}} \in \overline{U_{k+1}}^c$ and

$$\overline{\gamma_{w_{i-1}^{(1)}} \gamma_{w_i^{(1)}}} \bigcap \gamma^{(0)}|_{[0,v_0)} \neq \emptyset,$$

where $\overline{\gamma_{w_{i-1}^{(1)}} \gamma_{w_i^{(1)}}}$ denotes the image of the unique minimizing geodesic joining $\gamma_{w_{i-1}^{(1)}}$ and $\gamma_{w_i^{(1)}}$. However, since $d\left(\gamma_{w_{i-1}^{(1)}}, \gamma_{w_i^{(1)}}\right) \leqslant \delta_\varepsilon^{(1)} < \delta_\varepsilon$, we know from Lemma 2.2.3 that

$$d\left(\gamma_{v_0}, \gamma_{w_{i-1}^{(1)}}\right) < \delta_\varepsilon^{(1)},\ d\left(\gamma_{v_0}, \gamma_{w_i^{(1)}}\right) < \delta_\varepsilon^{(1)},$$

which is an obvious contradiction to the construction of $\left\{w_i^{(1)}\right\}_{i=0}^{l_1}$.

On the other hand, if $\gamma^{\mathcal{P}^{(1)}_{[v_0,u_1]}} \bigcap \gamma^{(0)}|_{(u_1,\tau_1]} \neq \emptyset$, then there exists some $i \leqslant l_1 - 1$, such that $\overline{\gamma_{w_{i-1}^{(1)}} \gamma_{w_i^{(1)}}} \bigcap \gamma^{(0)}|_{(u_1,\tau_1]} \neq \emptyset$. Since $\gamma_{w_{i-1}^{(1)}}, \gamma_{w_i^{(1)}} \in \overline{U_1}^c$ and $d\left(\gamma_{w_{i-1}^{(1)}}, \gamma_{w_i^{(1)}}\right) = \delta_\varepsilon^{(1)} < \delta_\varepsilon$, we know again from Lemma 2.2.3 that

$$d\left(\gamma_{u_1}, \gamma_{w_{i-1}^{(1)}}\right) < \delta_\varepsilon^{(1)},\ d\left(\gamma_{u_1}, \gamma_{w_i^{(1)}}\right) < \delta_\varepsilon^{(1)}.$$

But this is also a contradiction to the construction of $\left\{w_i^{(1)}\right\}_{i=0}^{l_1}$.

Therefore, the closed curve $\gamma^{(1)}$ over $[0,1]$ defined by

$$\gamma_t^{(1)} = \begin{cases} \gamma_t^{\mathcal{P}^{(1)}_{[v_0,u_1]}}, & t \in [v_0, u_1]\,; \\ \gamma_t^{(0)}, & t \in [0,1] \setminus [v_0, u_1]\,, \end{cases}$$

is a Jordan curve.

Now consider $\gamma|_{[v_1,u_2]}$. The previous argument can be carried through easily with respect to the Jordan curve $\gamma^{(1)}$, and we obtain a finite partition

$$\mathcal{P}^{(2)}_{[v_1,u_2]}:\ v_1 = w_0^{(2)} < w_1^{(2)} < \cdots < w_{l_2-1}^{(2)} < w_{l_2}^{(2)} = u_2,$$

such that $\left\|\mathcal{P}^{(2)}_{[v_1,u_2]}\right\| < \varepsilon$, and the closed curve $\gamma^{(2)}$ over $[0,1]$ defined by

$$\gamma_t^{(2)} = \begin{cases} \gamma_t^{\mathcal{P}^{(2)}_{[v_1,u_2]}}, & t \in [v_1, u_2]\,; \\ \gamma_t^{(1)}, & t \in [0,1] \setminus [v_1, u_2]\,, \end{cases}$$

is a Jordan curve, where $\gamma^{\mathcal{P}^{(2)}_{[v_1,u_2]}}$ is the geodesic interpolation of $\gamma|_{[v_1,u_2]}$ over the partition points in $\mathcal{P}^{(2)}_{[v_1,u_2]}$. By induction, we are able to construct simple piecewise geodesic approximation of each piece of γ outside $\bigcup_{i=1}^{k+1} \overline{U_i}$ and finally obtain a finite partition $\mathcal{P}_\varepsilon$ of $[0,1]$ with partition points

$$\{0\} \bigcup \left(\bigcup_{i=1}^{k+1} \left\{ v_{i-1}, w_1^{(i)}, \cdots, w_{l_i-1}^{(i)}, u_i, \tau_i \right\} \right),$$

such that $\|\mathcal{P}_\varepsilon\| < \varepsilon$, and the geodesic interpolation $\gamma^{\mathcal{P}_\varepsilon}$ (which is $\gamma^{(k+1)}$ by induction) of γ over the points of $\mathcal{P}_\varepsilon$ is a Jordan curve.

Now the proof is complete. □

Remark 2.2.1. By slight modifying the proof, it is not hard to see that Theorem 2.2.2 also holds for non-closed simple curves. In this respect, it strengthens the result of Theorem 2.2.1.

Remark 2.2.2. It is possible to generalize our main results to infinite dimensional spaces with suitable geodesic properties. For technical simplicity we are not going to present the details.

2.3 Applications

In this section, we demonstrate two applications of Theorem 2.2.2. Here we assume that $M = \mathbb{R}^2$.

2.3.1 Green's Theorem for Jordan Curves with Finite p-variation for $1 \leqslant p < 2$

We first prove a generalized version of Green's theorem for planar Jordan curves with finite p-variation, where $1 \leqslant p < 2$.

The following continuity result on Young's integrals is crucial for us. We refer the reader to [46] for the proof.

Theorem 2.3.1. *Let $p, q \geqslant 1$ be such that $\frac{1}{p} + \frac{1}{q} > 1$. Let $x, y : [0,1] \to \mathbb{R}^d$ be two continuous paths with finite p- and q-variation respectively. Then the following limit exists:*

$$\int_0^1 x \otimes dy := \lim_{\|\mathcal{P}\| \to 0} \sum_{\mathcal{P}:\ 0=t_0<\ldots<t_n=1} x_{t_i} \otimes \left(y_{t_{i+1}} - y_{t_i} \right).$$

Moreover, the path $\int_0^{\cdot} x \otimes dy$ *has finite q-variation and*

$$\left\| \int_0^{\cdot} x \otimes dy \right\|_q \leqslant 2\zeta\left(\frac{1}{p}+\frac{1}{q}\right)\|x\|_p\|y\|_q,$$

where $\zeta(\cdot)$ *is the classical Riemann zeta function.*

The following lemma demonstrates the importance of piecewise linear approximation under the p-variation metric. We refer the reader to [46] for the proof.

Lemma 2.3.1. *Let* $x:[0,1]\to\mathbb{R}^d$ *be a path with finite p-variation, where* $p\geqslant 1$*. Let* $f:\mathbb{R}^d\to\mathbb{R}^e$ *be a Lipschitz function with Lipschitz constant* C*. Then*

(1) $\|f(x)\|_p \leqslant C\|x\|_p$.

(2) For any $q>p$,

$$\lim_{\|\mathcal{P}\|\to 0}\left\|f(x)-f\left(x^{\mathcal{P}}\right)\right\|_q \to 0,$$

where $x^{\mathcal{P}}$ *denotes the piecewise linear interpolation of* x *over the partition points in* $\mathcal{P}$.

We now prove a generalized version of Green's theorem for non-rectifiable Jordan curves.

Theorem 2.3.2. *Let* $f,g:\mathbb{R}^2\to\mathbb{R}$ *be functions with continuous first order derivatives, and let* $\gamma:[0,1]\to\mathbb{R}^2$ *be a positively oriented Jordan curve with finite p-variation, where* $1\leqslant p<2$*. Let* x,y *denote the first and second coordinate components of* γ *respectively. Then we have*

$$\int_0^1\left(f(\gamma_s)\,dy_s-g(\gamma_s)\,dx_s\right)=\int_{\mathrm{Int}(\gamma)}\left(\frac{\partial f}{\partial x}+\frac{\partial g}{\partial y}\right)dxdy,$$

where the integral on the L.H.S. is understood as Young's integral, and $\mathrm{Int}(\gamma)$ *denotes the interior of* γ.

Proof. Fix $\varepsilon>0$. According to Theorem 2.2.2, let $\mathcal{P}_\varepsilon$ be a finite partition of $[0,1]$ such that $\|\mathcal{P}_\varepsilon\|<\varepsilon$, and the piecewise linear interpolation $\gamma^{\mathcal{P}_\varepsilon}$ of γ over the partition points in $\mathcal{P}_\varepsilon$ is a Jordan curve. Let $x^{\mathcal{P}_\varepsilon}, y^{\mathcal{P}_\varepsilon}$ be the first and second components of $\gamma^{\mathcal{P}_\varepsilon}$ respectively. It follows from the classical Green's theorem for piecewise smooth Jordan curves that

$$\int_0^1\left(f\left(\gamma_s^{\mathcal{P}_\varepsilon}\right)dy_s^{\mathcal{P}_\varepsilon}-g\left(\gamma_s^{\mathcal{P}_\varepsilon}\right)dx_s^{\mathcal{P}_\varepsilon}\right)=\int_{\mathrm{Int}(\gamma^{\mathcal{P}_\varepsilon})}\left(\frac{\partial f}{\partial x}+\frac{\partial g}{\partial y}\right)dxdy.$$

For any $q \in (p, 2)$, we know that

$$\begin{aligned}
&\left| \int_0^1 \left(f\left(\gamma_s^{\mathcal{P}_\varepsilon}\right) dy_s^{\mathcal{P}_\varepsilon} - \int_0^1 f\left(\gamma_s\right) dy_s \right) \right| \\
= &\left| \int_0^1 \left(f\left(\gamma_s^{\mathcal{P}_\varepsilon}\right) - f\left(\gamma_s\right)\right) dy_s^{\mathcal{P}_\varepsilon} + \int_0^1 f\left(\gamma_s\right) d\left(y_s^{\mathcal{P}_\varepsilon} - y_s\right) \right| \\
\leqslant &\left| \int_0^1 \left(f\left(\gamma_s^{\mathcal{P}_\varepsilon}\right) - f\left(\gamma_s\right)\right) dy_s^{\mathcal{P}_\varepsilon} \right| + \left| \int_0^1 f\left(\gamma_s\right) d\left(y_s^{\mathcal{P}_\varepsilon} - y_s\right) \right| \\
\leqslant &\, 2\zeta\left(\frac{2}{q}\right) \left(\left\| f\left(\gamma_\cdot^{\mathcal{P}_\varepsilon}\right) - f\left(\gamma_\cdot\right)\right\|_q \left\|\gamma\right\|_q + \left\| f\left(\gamma_\cdot\right)\right\|_q \left\|\gamma_\cdot^{\mathcal{P}_\varepsilon} - \gamma_\cdot\right\|_q \right),
\end{aligned}$$

where the final inequality follows from Theorem 2.3.1 and Lemma 2.3.1. Therefore, by Lemma 2.3.1,

$$\int_0^1 f\left(\gamma_s^{\mathcal{P}_\varepsilon}\right) dy_s^{\mathcal{P}_\varepsilon} \to \int_0^1 f\left(\gamma_s\right) dy_s$$

as $\varepsilon \to 0$. Similarly,

$$\int_0^1 g\left(\gamma_s^{\mathcal{P}_\varepsilon}\right) dy_s^{\mathcal{P}_\varepsilon} \to \int_0^1 g\left(\gamma_s\right) dy_s$$

as $\varepsilon \to 0$.

On the other hand, as γ has finite p variation, it has a $1/p$-Hölder parametrization. Therefore, γ has Hausdorff dimension less than 2. In particular, this means that $\gamma([0,1])$ has zero Lebesgue measure. By applying the bounded convergence theorem to the integrand $\left(\frac{\partial f}{\partial x} + \frac{\partial g}{\partial y}\right) \mathbf{1}_{\mathrm{Int}(\gamma^{\mathcal{P}_\varepsilon})}$, we have

$$\int_{\mathrm{Int}(\gamma^{\mathcal{P}_\varepsilon})} \left(\frac{\partial f}{\partial x} + \frac{\partial g}{\partial y}\right) dxdy \to \int_{\mathrm{Int}(\gamma)} \left(\frac{\partial f}{\partial x} + \frac{\partial g}{\partial y}\right) dxdy$$

as $\varepsilon \to 0$.

Now the proof is complete. □

Remark 2.3.1. For paths with bounded total variation, there is a version of Green's theorem which works for non-simple closed curves, involving the winding number of a path. An interesting inequality in this respect is the Banchoff-Pohl inequality, which generalizes the isoperimetric inequality and asserts that the winding number of a rectifiable curve is square-integrable. The reason for the "simple closed" condition in our version of Green's theorem is that in general the winding number of a non-simple non-rectifiable curve is not integrable. The fact that we can approximate the rough Jordan curves by piecewise linear interpolations which are still Jordan means that the winding number of each approximation is an indicator function, which is bounded by the indicator function of a neighborhood of $\mathrm{Int}(\gamma)$. This point is crucial in our situation.

Remark 2.3.2. A direct consequence of Theorem 2.3.2 is Cauchy's theorem for Jordan curves with finite p-variation where $1 \leqslant p < 2$, according to the Cauchy-Riemann equation for holomorphic functions.

2.3.2 The Uniqueness of Signature Problem for Planar Jordan Curves with Finite p-variation for $1 \leqslant p < 2$

As the second application of Theorem 2.2.2, we prove that up to reparametrization, a planar Jordan curve with finite p-variation for $1 \leqslant p < 2$ is uniquely determined by its signature.

For notational simplicity, assume that $\{\mathbf{e}_1, \ldots, \mathbf{e}_d\}$ is the standard basis of $\mathbb{R}^d$, and $\{\mathbf{e}_1^*, \ldots, \mathbf{e}_d^*\}$ is the corresponding dual basis of $\mathbb{R}^{d*}$. We embed $T\left(\left(\mathbb{R}^d\right)^*\right)$ into $T\left(\mathbb{R}^d\right)^*$ by extending the relation

$$\mathbf{e}_{i_1}^* \otimes \ldots \otimes \mathbf{e}_{i_n}^* \left(\mathbf{e}_{j_1} \otimes \ldots \otimes \mathbf{e}_{j_k}\right) = \begin{cases} 1 & \text{if } n = k \text{ and } i_1 = j_1, \cdots, i_k = j_k, \\ 0 & \text{otherwise,} \end{cases}$$

linearly. Throughout the rest of this section $1 \leqslant p < 2$ is some fixed constant.

The following basic properties of the signature follow easily from the definition and Lyons' extension theorem.

Proposition 2.3.1. *Let $x : [0,1] \to \mathbb{R}^d$ be a continuous path with finite p-variation. Then the following holds.*

(1) Let $\sigma : [0,1] \to [0,1]$ be a continuous increasing surjection, then

$$S\left(x_\cdot\right)_{0,1} = S\left(x_{\sigma(\cdot)}\right)_{0,1}.$$

(2) For all $v \in \mathbb{R}^d$,

$$S\left(v + x_\cdot\right)_{0,1} = S\left(x_\cdot\right)_{0,1}.$$

(3) Let x_n be a sequence of continuous paths with finite p-variation and

$$\lim_{n\to\infty} \left\|x_n - x\right\|_p \to 0,$$

Then for each $i \in \mathbb{N}$,

$$\lim_{n\to\infty}\left|\pi_i\left(S(x_n)_{0,1}\right)-\pi_i(S(x))\right|\to 0.$$

It turns out that some terms in the signature of a curve can be reduced to single line integrals. This is the key idea for proving our uniqueness of signature result.

Proposition 2.3.2. *Let $\gamma:\ [0,1]\to\mathbb{R}^2$ be a positively oriented Jordan curve with finite p-variation. Let $x., y.$ be the first and second coordinate components of γ respectively. Then for $k,n\geqslant 0$, we have*

$$\begin{aligned}
&\mathbf{e}_1^{*\otimes k+1}\otimes\mathbf{e}_2^{*\otimes n+1}\left(S(\gamma)_{0,1}\right)\\
=\;&\int_0^1\int_0^{s_{n+k+2}}\cdots\int_0^{s_2}dx_{s_1}\cdots dx_{s_{k+1}}dy_{s_{k+2}}\cdots dy_{s_{n+k+2}}\\
=\;&\frac{1}{k!n!}\int_{\operatorname{Int}(\gamma)}(x-x_0)^k(y_1-y)^n\,dxdy.
\end{aligned}$$

Proof. Note that

$$\begin{aligned}
&\int_0^1\int_0^{s_{n+k+2}}\cdots\int_0^{s_2}dx^{\mathcal{P}_\varepsilon}_{s_1}\cdots dx^{\mathcal{P}_\varepsilon}_{s_{k+1}}dy^{\mathcal{P}_\varepsilon}_{s_{k+2}}\cdots dy^{\mathcal{P}_\varepsilon}_{s_{n+k+2}}\\
=\;&\frac{1}{(k+1)!n!}\int_0^1\left(x^{\mathcal{P}_\varepsilon}_{s_{k+2}}-x^{\mathcal{P}_\varepsilon}_0\right)^{k+1}\left(y^{\mathcal{P}_\varepsilon}_1-y^{\mathcal{P}_\varepsilon}_{s_{k+2}}\right)^n dy^{\mathcal{P}_\varepsilon}_{s_{k+2}}\\
=\;&\frac{1}{k!n!}\int_{\operatorname{Int}(\gamma^{\mathcal{P}_\varepsilon})}(x-x_0)^k(y_1-y)^n\,dxdy,
\end{aligned}$$

where $\mathcal{P}_\varepsilon$ is the partition given by Theorem 2.2.2, so that $\gamma^{\mathcal{P}_\varepsilon}$ is a Jordan curve.

By Proposition 2.3.1,

$$\begin{aligned}
&\int_0^1\int_0^{s_{n+k+2}}\cdots\int_0^{s_2}dx^{\mathcal{P}_\varepsilon}_{s_1}\cdots dx^{\mathcal{P}_\varepsilon}_{s_{k+1}}dy^{\mathcal{P}_\varepsilon}_{s_{k+2}}\cdots dy^{\mathcal{P}_\varepsilon}_{s_{n+k+2}}\\
\to\;&\int_0^1\int_0^{s_{n+k+2}}\cdots\int_0^{s_2}dx_{s_1}\cdots dx_{s_{k+1}}dy_{s_{k+2}}\cdots dy_{s_{n+k+2}}
\end{aligned}$$

as $\varepsilon\to 0$. On the other hand, as in the proof of Theorem 2.3.2 it is not hard to see that

$$\int_{\operatorname{Int}(\gamma^{\mathcal{P}_\varepsilon})}(x-x_0)^k(y_1-y)^n\,dxdy\to\int_{\operatorname{Int}(\gamma)}(x-x_0)^k(y_1-y)^n\,dxdy$$

as $\varepsilon\to 0$.

Therefore, the result follows. □

Remark 2.3.3. The case of $n = 1, k = 0$ for Proposition 2.3.2 has already been proved by Werness [64]. The main difficulty in extending to the general case involves the interchange of iterated path integrals.

The following lemma is the main reason why our result only works for Jordan curves.

Lemma 2.3.2. *Let $\gamma, \widetilde{\gamma} : [0,1] \to \mathbb{R}^2$ be two positively oriented Jordan curves such that $\mathrm{Im}(\gamma) = \mathrm{Im}(\widetilde{\gamma})$ and $\gamma_0 = \widetilde{\gamma}_0$. There exists a continuous and strictly increasing map $\sigma : [0,1] \to [0,1]$ with $\sigma(0) = 0$, $\sigma(1) = 1$, such that $\gamma_{\sigma(t)} = \widetilde{\gamma}$. In other words, γ and $\widetilde{\gamma}$ are equal up to reparametrization.*

Proof. As γ and $\widetilde{\gamma}$ are Jordan curves, $\mathrm{Im}(\gamma)\backslash\{\gamma_0\}$ and $\mathrm{Im}(\widetilde{\gamma})\backslash\{\widetilde{\gamma}_0\}$ are both homeomorphic to $(0,1)$. Therefore, the function $\sigma : (0,1) \to (0,1)$ defined by $\sigma(t) = \gamma^{-1} \circ \widetilde{\gamma}(t)$ is a homeomorphism $(0,1) \to (0,1)$. Hence, it is strictly monotone. This implies that $\lim_{t\to 0} \sigma(t)$ exists. Moreover, it is easy to see that $\lim_{t\to 0} \sigma(t) \in \{0,1\}$.

If $\lim_{t\to 0} \sigma(t) = 0$, then σ can be extended to the desired map on $[0,1]$. As $\gamma_{\sigma(t)} = \widetilde{\gamma}_t$, we know that γ and $\widetilde{\gamma}$ are equal up to reparametrization. If $\lim_{t\to 0} \sigma(t) = 1$, then $\lim_{t\to 1} \sigma(t) = 0$ and $\sigma(t)$ is decreasing. This implies that $\sigma(1-t)$ is a continuous increasing function. Therefore, γ and $\widetilde{\gamma}$ have opposite orientations, which is a contradiction. □

Now we are in a position to provide a solution to the uniqueness of signature problem for planar Jordan curves.

Theorem 2.3.3. *Let $\gamma, \tilde{\gamma} : [0,1] \to \mathbb{R}^2$ be two Jordan curves with finite p-variation starting at the origin. Then $S(\gamma)_{0,1} = S(\tilde{\gamma})_{0,1}$ if and only if γ and $\tilde{\gamma}$ are equal up to reparametrization.*

Proof. Sufficiency follows from Proposition 2.3.1. We now consider the necessity part.

As $\mathbf{e}_1^* \otimes \mathbf{e}_2^* \left(S(\gamma)_{0,1}\right) = \mathbf{e}_1^* \otimes \mathbf{e}_2^* \left(S(\tilde{\gamma})_{0,1}\right)$, by Proposition 2.3.2 we have

$$(-1)^{\varepsilon(\gamma)} \int_{\mathrm{Int}(\gamma)} dxdy = (-1)^{\varepsilon(\tilde{\gamma})} \int_{\mathrm{Int}(\tilde{\gamma})} dxdy,$$

where $\varepsilon(\gamma)$ is 0 if γ is positively oriented and 1 otherwise. As $\int_{\mathrm{Int}(\gamma)} dxdy$ and $\int_{\mathrm{Int}(\tilde{\gamma})} dxdy$ are both positive, we must have γ and $\tilde{\gamma}$ oriented in the same direction.

Without loss of generality, assume both γ and $\tilde{\gamma}$ are positively oriented. By Proposition 2.3.2 and that $S(\gamma)_{0,1} = S(\tilde{\gamma})_{0,1}$, we have

$$\int_{\mathrm{Int}(\gamma)} (x - x_0)^k (y_1 - y)^n \, dxdy = \int_{\mathrm{Int}(\tilde{\gamma})} (x - \tilde{x}_0)^k (\tilde{y}_1 - y)^n \, dxdy$$

for all $n, k \geqslant 0$. Therefore,

$$\int_{\mathrm{Int}(\gamma)} e^{i(\lambda_1 x+\lambda_2 y)} dxdy = \int_{\mathrm{Int}(\tilde{\gamma})} e^{i(\lambda_1 x+\lambda_2 y)} dxdy$$

for all $\lambda_1, \lambda_2 \in \mathbb{R}$.

Both $\mathbf{1}_{\mathrm{Int}(\gamma)}$ and $\mathbf{1}_{\mathrm{Int}(\tilde{\gamma})}$ are in L^1 and by the injectivity of the Fourier transform on L^1, we have

$$\mathbf{1}_{\mathrm{Int}(\gamma)}(x, y) = \mathbf{1}_{\mathrm{Int}(\tilde{\gamma})}(x, y)$$

for almost every $(x, y) \in \mathbb{R}^2$. In particular, this implies that both $\mathrm{Int}(\gamma) \backslash \overline{\mathrm{Int}(\tilde{\gamma})} \subset \mathrm{Int}(\gamma) \backslash \mathrm{Int}(\tilde{\gamma})$ and $\mathrm{Int}(\tilde{\gamma}) \backslash \overline{\mathrm{Int}(\gamma)} \subset \mathrm{Int}(\gamma) \backslash \mathrm{Int}(\tilde{\gamma})$ are null sets in $\mathbb{R}^2$. However, since both $\mathrm{Int}(\tilde{\gamma}) \backslash \overline{\mathrm{Int}(\gamma)}$ and $\mathrm{Int}(\gamma) \backslash \overline{\mathrm{Int}(\tilde{\gamma})}$ are open, they must be empty. Therefore,

$$\overline{\mathrm{Int}(\tilde{\gamma})} = \overline{\mathrm{Int}(\gamma)}.$$

By the Jordan curve theorem, we have

$$\overline{\mathbb{R}^2 \backslash \overline{\mathrm{Int}(\tilde{\gamma})}} = \mathbb{R}^2 \backslash \mathrm{Int}(\tilde{\gamma}).$$

Therefore, $\mathrm{Int}(\tilde{\gamma}) = \mathrm{Int}(\gamma)$ and so $\mathrm{Im}(\gamma) = \mathrm{Im}(\tilde{\gamma})$ and by Lemma 2.3.2, γ and $\tilde{\gamma}$ are equal up to reparametrization. □

Remark 2.3.4. The proof of Theorem 2.3.3 gives an explicit way of computing the moments of the finite measure $\mathbf{1}_{\mathrm{Int}(\gamma)}(x, y)\, dxdy$ from the signature of γ. In particular, by applying Fourier inversion it gives us a way of reconstructing the path from its signature. However, this is not explicit due to the possible difficulty of inverting the Fourier transform.

Chapter 3

The Uniqueness of Signature Problem for Weakly Geometric Rough Paths

3.1 Introduction

In this chapter, we give a complete solution to the uniqueness of signature problem for weakly geometric rough paths.

In [35], B.M. Hambly and T. Lyons introduced the notion of tree-like paths and proved that the signature of a continuous path with bounded total variation is trivial if and only if the path is tree-like. According to their formulation, a continuous path $x: [0,1] \to \mathbb{R}^d$ with bounded total variation is called tree-like if there exists a continuous function $h: [0,1] \to [0,\infty)$ (called a *height function*) such that $h(0) = h(1) = 0$ and

$$|x_t - x_s| \leqslant h(t) + h(s) - 2 \inf_{u \in [s,t]} h(u) \tag{3.1.1}$$

for all $0 \leqslant s \leqslant t \leqslant 1$. This notion is to describe paths that track back along themselves completely. Since the signature of x represents the aggregation of its global higher order increments, it is heuristically not obvious that the trajectory of X could be determined from its signature in any sense. In this respect, the result of B.M. Hambly and T. Lyons is profound and it reveals the fundamental relationship between the geometry of the path and its signature, which is a deep exploration of the connection between geometric and algebraic features of the path.

The fundamental idea of their proof is via approximation. More precisely, for the "if" part, they first showed that if a piecewise linear path is tree-like, then it has trivial signature. The proof is based on induction on the number of edges. Then they showed that a tree-like path with bounded total variation can be approximated by a sequence of piecewise linear tree-like paths under the 1-variation metric. The conclusion of the "if" part is then an easy consequence of the continuity of the signature map under the 1-variation metric. The logic of proving the "only if" part is similar, but the development is much more involved. First of all, again by induction they showed that if a weakly piecewise linear path (i.e. a path with bounded total variation whose image lies in a polygonal curve) has trivial signature, then it must be tree-like. The next step is to represent the original path by the indefinite path integral of a rank 1 one form along the path. It then follows from standard compactness results in measure theory that this one form can be approximated by a sequence of locally constant rank 1 one forms in L^1-norm (this is sufficient in the case of paths with bounded total variation). It is both the local constancy and rank 1 features which guarantee that the indefinite path integrals of this sequence of approximating one forms yield a sequence of weakly piecewise linear paths, which converges to the original path under the 1-variation metric. Moreover, according to the shuffle product formula the signatures of this integral sequence are functionals of the signature of the original path and hence they have trivial signatures. It follows that this sequence of approximating paths must be tree-like. Finally the conclusion of the "only if" part follows from a compactness result for tree-like paths.

The main contribution of this chapter is the extension of B.M. Hambly and T. Lyons' result to the case of weakly geometric rough paths. Before developing the complete mathematical proof, we first illustrate the underlying idea informally.

Let's look at B.M. Hambly and T. Lyons' proof of the "if" part more closely. Although their proof relies on the assumption of bounded total variation in a crucial way, the underlying idea is robust. From a more geometric point of view, a tree-like path is nothing but just a continuous loop in some real tree. If we start with a finite set of points on the path, the key to producing a piecewise linear tree-like path with this set as end points (but not necessarily all the end points) is simply to identify all nodes associated with this set. The piecewise linear tree-like path is then constructed almost immediately from the basic structure of a real tree. This argument certainly does not rely on the regularity of the path and can be developed rigorously in the language of real trees (or equivalently in terms of "height functions" as in [35]). Our proof of the "if" part for weakly geometric rough paths is based on this idea.

On the other hand, if we look into B.M. Hambly and T. Lyons' proof of the "only if" part more carefully, it is not hard to capture the fundamental difficulties in extending the technique to weakly geometric rough paths. First of all, the representation of the path by the indefinite path integral of a rank 1 one form along the path depends crucially on the fact that the path has bounded total variation, since the rank 1 feature comes from projections onto the tangent lines of the path. In general, such a representation is possible if we relax the rank 1 condition. However, this condition is necessary to guarantee that the approximating sequence is weakly piecewise linear, a point which is again crucial in the development. Even if such a one form can be constructed, the corresponding approximation is also highly nontrivial beyond the bounded total variation case since the L^1-norm on the space of one forms is not strong enough to yield the corresponding convergence for the approximating paths under the p-variation metric.

Therefore, in developing the "only if" part of the assertion for weakly geometric rough paths, we do need a substantially new idea to get around the difficulties we mentioned before. The fundamental point is to prove that for any weakly geometric p-rough path $\mathbf{X}$, there exists a unique weakly geometric p-rough path $\widetilde{\mathbf{X}}$ up to reparametrization such that $S\left(\widetilde{\mathbf{X}}\right)_{0,1} = S\left(\mathbf{X}\right)_{0,1}$ and the signature path $S\left(\widetilde{\mathbf{X}}\right)_{0,\cdot}$ of $\widetilde{\mathbf{X}}$ is a simple curve. This almost indicates, at least in a very intuitive way (it would be clear in the context of Definition 3.2.1 below), that the signature group $\mathcal{S}_p$ over the space of weakly geometric p-rough paths is a real tree under some tree metric. This tree metric is actually not hard to construct in terms of the p-variation of the "reduced" path $\widetilde{\mathbf{X}}$. With the observation that the signature path $S(\mathbf{X})_{0,\cdot}$ of a weakly geometric p-rough path $\mathbf{X}$ with trivial signature is just a continuous loop in $\mathcal{S}_p$ under the tree metric, it follows easily that the path $\mathbf{X}$ is tree-like.

3.2 Preliminaries on Real Trees and Formulation of the Main Result

In the formulation and proof of our main result, we use the basic language of real trees instead of height functions, which is technically simpler and geometrically more intuitive. It can be seen from the following discussion that these two settings are equivalent. In this section, we recall the basic notions of real trees. The contents of this section are based on the monographs by I. Chiswell [11], and C. Favre and M. Jonsson [23].

Definition 3.2.1. A metric space (τ, d) is called a *real tree* if for any two distinct points $g, h \in \tau$, there exists a unique continuous simple curve $\gamma : [0,1] \to \tau$ up to reparametrization such that $\gamma(0) = g, \gamma(1) = h$. Moreover, this simple curve is a geodesic (i.e. it satisfies (1.2.4)).

For two points g, h in a real tree τ, we use $[g, h]$ to denote the image of the unique continuous simple curve joining g to h (it degenerates to a single point when $g = h$). $[g, h]$ is usually called the *segment* with end points g, h.

We list some simple geometric properties of segments in the following proposition. We refer the reader to [11] for the proof.

Proposition 3.2.1. *(1) For $g, h \in \tau$ and any continuous curve $\alpha : [0,1] \to \tau$ with $\alpha(0) = g, \alpha(1) = h$, we have $[g, h] \subset \alpha([0,1])$.*

(2) For $r, g, h \in \tau$, there exists some unique $w \in \tau$ such that $[r, g] \bigcap [r, h] = [r, w]$. Moreover, $[w, g] \bigcap [w, h] = \{w\}$ and $[g, h] = [g, w] \bigcup [w, h]$.

The notion of partial order is important in the study of real trees. Let $r \in \tau$ be a fixed point, which is called a *root* (the choice of roots is of no importance). Define the relation "$\lesssim$" on τ by

$$g \lesssim h \text{ iff } [r, g] \subset [r, h].$$

It is easy to show that "$\lesssim$" defines a partial order on τ. Moreover, for $g, h \in \tau$, the unique element $w \in \tau$ given by Proposition 3.2.1 (2) is the infimum of g and h under the partial order "$\lesssim$". We denote w by $g \wedge h$.

The following result gives a geometric characterization of general real trees. We refer the reader to [11] for the proof.

Proposition 3.2.2. *A metric space (τ, d) is a real tree if and only if it is a geodesic space and contains no subspace which is homeomorphic to the unit circle S^1.*

For compact real trees, we have a more explicit and constructive characterization based on height functions.

Let $h : [0,1] \to [0, \infty)$ be a continuous function. Introduce the equivalence relation "$\sim$" on $[0,1]$ by

$$s \sim t \text{ iff } h(s) = h(t) = \inf_{u \in [s,t]} h(u), \tag{3.2.1}$$

and define the functional d on the quotient space $[0,1]/_\sim$ by

$$d([s], [t]) = h(s) + h(t) - 2 \inf_{u \in [s,t] \text{ or } [t,s]} h(u). \tag{3.2.2}$$

It is shown by B.M. Hambly and T. Lyons [34] that d is well-defined and it defines a metric on $[0,1]/_\sim$ which makes it into a real tree. Moreover, the canonical projection $\pi : [0,1] \to [0,1]/_\sim$, where $[0,1]$ is equipped with the Euclidean topology, is a continuous map. Therefore, $([0,1]/_{\sim}, d)$ is a compact real tree. This tree is called the *contour tree* associated with the height function h.

Conversely, we have the following characterization. We refer the reader to [34] for the proof.

Proposition 3.2.3. *Every compact real tree τ is isometric to the contour tree associated with some height function.*

It is worth sketching the proof of Proposition 3.2.3 in a few words as it is then clear how the tree is realized as a contour tree. Firstly, by a topological argument we can always find a continuous loop $\alpha : [0,1] \to \tau$ onto the whole tree. Let h be the height function given by

$$h(t) = d\left(\alpha(t), \alpha(0)\right), \ t \in [0,1].$$

Then the contour tree associated with h is isometric to τ, and the isometry is given by $i([t]) = \alpha(t)$.

For a general partially ordered set, it is a natural question to ask if it can be equipped with a real tree metric. The following result gives an affirmative answer to this question for a class of partially ordered sets. This is important for us since our signature group turns out to be a special example.

Let $(\mathcal{P}, \lesssim)$ be a partially ordered set. A subset $S \subset \mathcal{P}$ is said to be *full* if $s_1, s_2 \in S$, $t \in \mathcal{P}$ with $s_1 \lesssim t \lesssim s_2$ implies that $t \in S$.

Now we have the following result. We refer the reader to [23] for the proof.

Proposition 3.2.4. *Let $(\mathcal{P}, \lesssim)$ be a partially ordered set satisfying the following conditions:*

(1) $\mathcal{P}$ has a unique minimal element;

(2) any two elements $s, t \in \mathcal{P}$ have an infimum $s \wedge t$ (and hence unique by the definition of infimum);

(3) for any $t \in \mathcal{P}$, the set $\{s \in \mathcal{P} : \ s \lesssim t\}$ is totally ordered;

(4) There exists an increasing function $L : \ \mathcal{P} \to [0, \infty)$, such that the restriction of L on any full, totally ordered subset is a bijection onto a real interval (i.e. a connected subset of $\mathbb{R}$).

Then the functional d defined by

$$d(s,t) = L(s) + L(t) - 2L(s \wedge t), \ s, t \in \mathcal{P},$$

is a metric which makes $\mathcal{P}$ into a real tree.

Now we are in a position to formulate our main result. We first define tree-like paths in the setting of real trees.

Definition 3.2.2. Let V be a topological space. A continuous path $\beta : [0,1] \to V$ is called *tree-like* if there exists a real tree τ, a continuous map $\alpha : [0,1] \to \tau$ with $\alpha(0) = \alpha(1)$ and a map $\psi : \tau \to V$ such that $\beta = \psi \circ \alpha$. We also say that β is a tree-like path *realized* on the tree τ.

It is easy to see that being tree-like is invariant under reparametrization.

The following is the main result of this chapter.

Theorem 3.2.1. *Let $\mathbf{X} : [0,1] \to G^{\lfloor p \rfloor}\left(\mathbb{R}^d\right)$ be a weakly geometric p-rough path. Then $S(\mathbf{X})_{0,1} = \mathbf{1}$ if and only if $\mathbf{X}$ is tree-like.*

3.3 Proof of the Sufficiency Part

In this section, we prove the sufficiency of Theorem 3.2.1. The proof essentially consists of two parts: showing that a "piecewise linear" tree-like path has trivial signature, and constructing a "piecewise linear" tree-like approximation of the original tree-like path.

First of all, we have the following result.

Proposition 3.3.1. *Let $\mathbf{X} = \psi \circ \alpha$ be a tree-like weakly geometric p-rough path realized on some real tree τ. If there exists a finite partition*

$$\mathcal{P} : \ 0 = t_0 < t_2 < \cdots < t_{k-1} < t_k = 1$$

of $[0,1]$ such that α is monotone with respect to the root $r := \alpha(0)$ on each sub-interval $[t_{i-1}, t_i]$, then the signature of $\mathbf{X}$ is trivial.

Proof. We prove this by induction on the number $|\mathcal{P}|$ of partition points in $\mathcal{P}$.

If $|\mathcal{P}| = 2$, then $\alpha \equiv r$ since $\alpha(0) = \alpha(1) = r$, and the claim is trivial.

Suppose that the claim is true for the case when $|\mathcal{P}| < n$, and that $\mathbf{X}$ satisfies the assumptions with a partition $\mathcal{P}$ consisting of n points. Since $\alpha(\mathcal{P})$ is a finite set in τ, it has a maximal element in itself, say, $g = \alpha(t_i)$ for some $t_i \in \mathcal{P}$. Since $\alpha(t_{i-1}), \alpha(t_{i+1}) \leqslant g$, from the definition of the partial order it is easy to see that either $\alpha(t_{i-1}) \leqslant \alpha(t_{i+1})$ or $\alpha(t_{i+1}) \leqslant \alpha(t_{i-1})$.

Let us assume the first case. Set

$$t' = \inf\left\{t \in [t_{i-1}, t_i] : \ \alpha(t) = \alpha(t_{i+1})\right\}.$$

It follows that $\mathbf{X}\big|_{[t_i,t_{i+1}]}$ is a reparametrization of $\mathbf{X}\big|_{[t',t_i]}$ in the general sense and hence

$$S\left(\mathbf{X}\big|_{[t',t_{i+1}]}\right) = S\left(\mathbf{X}\big|_{[t',t_i]}\right) \otimes S\left(\mathbf{X}\big|_{[t_i,t_{i+1}]}\right) = \mathbf{1}.$$

Now let α' be the continuous loop in τ such that

$$\alpha'(t) = \alpha(t),\ t \in [0, t_{i-1}] \bigcup [t_{i+1}, 1],$$

and $\alpha'\big|_{[t_{i-1},t_{i+1}]}$ is the unique continuous simple curve (choose some parametrization) joining $\alpha(t_{i-1})$ to $\alpha(t_{i+1})$. It is then easy to see that $\mathbf{X}' := \psi \circ \alpha'$ has finite p-variation (controlled by the p-variation of $\mathbf{X}$), and we have

$$S\left(\mathbf{X}'\big|_{[t_{i-1},t_{i+1}]}\right) = S\left(\mathbf{X}\big|_{[t_{i-1},t']}\right)$$

since $\mathbf{X}'\big|_{[t_{i-1},t_{i+1}]}$ is a reparametrization of $\mathbf{X}\big|_{[t_{i-1},t']}$. Therefore,

$$\begin{aligned} S(\mathbf{X})_{0,1} &= S\left(\mathbf{X}\big|_{[0,t_{i-1}]}\right) \otimes S\left(\mathbf{X}\big|_{[t_{i-1},t']}\right) \otimes S\left(\mathbf{X}\big|_{[t',t_{i+1}]}\right) \otimes S\left(\mathbf{X}\big|_{[t_{i+1},1]}\right) \\ &= S\left(\mathbf{X}'\big|_{[0,t_{i-1}]}\right) \otimes S\left(\mathbf{X}'\big|_{[t_{i-1},t_{i+1}]}\right) \otimes S\left(\mathbf{X}'\big|_{[t_{i+1},1]}\right) \\ &= S(\mathbf{X}')_{0,1}. \end{aligned}$$

On the other hand, it is obvious that $\mathbf{X}'$ satisfies the assumptions with the partition $\mathcal{P}\backslash\{t_i\}$. Therefore, by the induction hypothesis we know that $\mathbf{X}'$ has trivial signature and so does $\mathbf{X}$. The second case can be treated in the same way.

Now the proof is complete. □

The second part of the proof is to show that a tree-like weakly geometric p-rough path can be approximated by the ones in Proposition 3.3.1 in a sense that the continuity of the signature map should follow.

Let $\mathbf{X} = \psi \circ \alpha$ be a tree-like weakly geometric p-rough path realized on some real tree τ, and choose $\alpha(0) = \alpha(1)$ to be the root of τ. Given a finite partition $\mathcal{P}$ of $[0,1]$, define

$$B = \{\alpha(t_{i_1}) \wedge \cdots \wedge \alpha(t_{i_l}) :\ l \geqslant 1,\ t_{i_1}, \cdots, t_{i_l} \in \mathcal{P}\}.$$

Heuristically, B contains all nodes associated with the finite point set $\alpha(\mathcal{P})$ in τ. Note that if $g, h \in B$, then $g \wedge h \in B$. Moreover, by the definition of the partial order and Proposition 3.2.1 (1), every point in the set B is reachable by α.

Set $s_0 = 0$, and define inductively

$$s_i = \inf\{s \in [s_{i-1}, 1] :\ \alpha(s) \in B\backslash\{\alpha(s_{i-1})\}\}$$

for $i \geqslant 1$.Since B is a finite set in τ, by the continuity of α it is easy to see that there exists some $k \geqslant 1$ such that

$$0 = s_0 < s_1 < \cdots < s_k \leqslant 1,$$

$\alpha(s_k) = \alpha(0)$ and $\alpha([s_k, 1]) \bigcap (B \backslash \{\alpha(0)\}) = \emptyset$. We denote this partition of $[0, 1]$ (including the end points $\{0, 1\}$) by $\mathcal{P}'$. Note that $\mathcal{P}$ may not be a subset of $\mathcal{P}'$, but apparently we have $\alpha(\mathcal{P}') = B$. The next key observation is that for each s_i, either $\alpha(s_{i-1}) \leqslant \alpha(s_i)$ or $\alpha(s_i) \leqslant \alpha(s_{i-1})$. In fact, if $\alpha(s_{i-1}) \wedge \alpha(s_i) \notin \{\alpha(s_{i-1}), \alpha(s_i)\}$, then by Proposition 3.2.1 we know that $\alpha(s) = \alpha(s_{i-1}) \wedge \alpha(s_i)$ for some $s \in (s_{i-1}, s_i)$, which is a contradiction to the definition of s_i. Moreover, again by the definition we can see that for any $s_i, s_j \in \mathcal{P}'$, there are only three possibilities: $[\alpha(s_{i-1}), \alpha(s_i)] = [\alpha(s_{j-1}), \alpha(s_j)]$ or $[\alpha(s_{i-1}), \alpha(s_i)] \bigcap [\alpha(s_{j-1}), \alpha(s_j)] = \emptyset$ or their intersection is a single point which is an end point of these two segments.

Now define a continuous loop α' such that on each sub-interval $[s_{i-1}, s_i]$ in $\mathcal{P}'$, α' is the unique geodesic (parametrized on $[s_{i-1}, s_i]$) joining $\alpha(s_{i-1})$ to $\alpha(s_i)$. It follows that $\alpha'(0) = \alpha'(1) = \alpha(0)$ and α' is monotone on each sub-interval $[s_{i-1}, s_i]$. Moreover, by Proposition 3.2.1 it is easy to see that $\tau' := \alpha'([0, 1]) \subset \tau$ is a real tree under the induced tree metric. Define a map $\psi' : \tau' \to G^{\lfloor p \rfloor}(\mathbb{R}^d)$ in the following way. For each s_i, choose a geodesic γ in $G^{\lfloor p \rfloor}(\mathbb{R}^d)$ joining $\mathbf{X}_{s_{i-1}}$ to $\mathbf{X}_{s_i}$ in such a way that if $[\alpha(s_{i-1}), \alpha(s_i)] = [\alpha(s_{j-1}), \alpha(s_j)]$ then the corresponding geodesics are either the same or are reversals of each other. For $g \in [\alpha(s_{i-1}), \alpha(s_i)]$, define $\psi'(g)$ to be the unique point on the geodesic γ such that

$$\frac{d(\alpha(s_{i-1}), g)}{d(\alpha(s_{i-1}), \alpha(s_i))} = \frac{d\left(\mathbf{X}_{s_{i-1}}, \psi'(g)\right)}{d\left(\mathbf{X}_{s_{i-1}}, \mathbf{X}_{s_i}\right)}.$$

Then ψ' is a well defined map from τ' to $G^{\lfloor p \rfloor}(\mathbb{R}^d)$. Let $\mathbf{X}' = \psi' \circ \alpha'$. From the construction of α' and ψ', we know that $\mathbf{X}'$ is a piecewise geodesic interpolation of $\mathbf{X}$ over the partition points in $\mathcal{P}'$. It follows from [26], Proposition 5.20 that

$$\|\mathbf{X}'\|_p \leqslant 3^{1-\frac{1}{p}} \|\mathbf{X}\|_p. \tag{3.3.1}$$

In particular, $\mathbf{X}'$ is a weakly geometric p-rough path. Moreover, it is obvious that $\mathbf{X}'$ is tree-like and satisfies the assumption in Proposition 3.3.1. Therefore, $\mathbf{X}'$ has trivial signature.

To complete the proof of the sufficiency of Theorem 3.2.1, it suffices to show that $\mathbf{X}'$ converges uniformly to the original path $\mathbf{X}$ as $\|\mathcal{P}\| \to 0$. In fact, by [26], Proposition 8.15 and Lemma 8.16, we have

$$\begin{aligned} &\overline{d}_q(\mathbf{X}', \mathbf{X}) \\ &\leqslant C \cdot (\|\mathbf{X}'\|_p + \|\mathbf{X}\|_p)^{\frac{p}{q}} \cdot \max\{d_\infty(\mathbf{X}', \mathbf{X}), \\ &\quad d_\infty(\mathbf{X}', \mathbf{X})^{\frac{1}{\lfloor p \rfloor}} \cdot (\|\mathbf{X}'\|_\infty + \|\mathbf{X}\|_\infty)^{1-\frac{1}{\lfloor p \rfloor}}\Big\} \end{aligned}$$

for any $q \in (p, \lfloor p \rfloor + 1)$, where the subscript "$\infty$" denotes the uniform norm or uniform metric under the Carnot–Carathéodory metric. It follows from (3.3.1) and the uniform convergence that $\mathbf{X}'$ converges to $\mathbf{X}$ under the q-variation metric as $\|\mathcal{P}\| \to 0$. This argument is based on the generalization of Lemma 2.3.1 to weakly geometric rough paths. By the continuity of the signature map under the q-variation metric (which follows immediately from Lyons' extension theorem), we conclude that $S(\mathbf{X})_{0,1} = \mathbf{1}$.

Now it remains to establish the following result.

Lemma 3.3.1. $\mathbf{X}'$ *converges to* $\mathbf{X}$ *uniformly as* $\|\mathcal{P}\| \to 0$.

Proof. By the continuity of $\mathbf{X}$, for any $\varepsilon > 0$, there exists some $\delta > 0$ such that

$$|t - s| < \delta \implies d(\mathbf{X}_s, \mathbf{X}_t) < \frac{\varepsilon}{2}.$$

Given any finite partition $\mathcal{P}$ of $[0, 1]$ such that $\|\mathcal{P}\| < \frac{\delta}{2}$, construct $\mathcal{P}'$ as before.

We claim that for any $s_i \in \mathcal{P}'$ and $s \in [s_{i-1}, s_i]$, $d(\mathbf{X}_s, \mathbf{X}_{s_{i-1}}) < \varepsilon$. In fact, if this is not the case, then we have $s - s_{i-1} \geqslant \delta$. On the other hand, let

$$t^* = \sup\{t \in [0, s) : \ t \in \mathcal{P}\}.$$

It follows that $t^* \in (s_{i-1}, s)$ and $s - t^* < \delta/2$. Therefore, we have $d(\mathbf{X}_{t^*}, \mathbf{X}_s) < \frac{\varepsilon}{2}$ which implies that $\alpha(t^*) \in B\backslash\{\alpha(s_{i-1})\}$. However, this contradicts the definition of s_i.

Consequently, for any $s \in [s_{i-1}, s_i]$ with $s_i \in \mathcal{P}$, we have

$$\begin{aligned} d(\mathbf{X}'_s, \mathbf{X}_s) &\leqslant d(\mathbf{X}'_s, \mathbf{X}_{s_{i-1}}) + d(\mathbf{X}_{s_{i-1}}, \mathbf{X}_s) \\ &= d(\mathbf{X}'_s, \mathbf{X}'_{s_{i-1}}) + d(\mathbf{X}_{s_{i-1}}, \mathbf{X}_s) \\ &\leqslant d(\mathbf{X}'_{s_i}, \mathbf{X}'_{s_{i-1}}) + d(\mathbf{X}_{s_{i-1}}, \mathbf{X}_s) \\ &= d(\mathbf{X}_{s_i}, \mathbf{X}_{s_{i-1}}) + d(\mathbf{X}_{s_{i-1}}, \mathbf{X}_s) \\ &< 2\varepsilon. \end{aligned}$$

Now the proof is complete. □

Remark 3.3.1. The use of piecewise geodesic interpolation for $\mathbf{X}$ is not necessary; the whole point is that the piecewise geodesic interpolation is realized on the tree τ by α'. In fact, we can simply take $\mathbf{X}'' = \psi \circ \alpha'$, and argue along a similar way as before. However, the continuity of $\mathbf{X}''$ is not obvious and requires more careful analysis.

3.4 Proof of the Necessity Part

Now we prove the necessity of Theorem 3.2.1. Let

$$\mathcal{S}_p = \left\{ S(\mathbf{X})_{0,1} : \ \mathbf{X} \in WG\Omega_p\left(\mathbb{R}^d\right) \right\}$$

be the signature group over the space of weakly geometric p-rough paths. As we have pointed out before, the key point is to show that for each $g \in \mathcal{S}_p$, there exists a unique $\mathbf{X} \in WG\Omega_p\left(\mathbb{R}^d\right)$ up to reparametrization such that $S(\mathbf{X})_{0,1} = g$ and $S(\mathbf{X})_{0,\cdot}$ is a simple curve. Such an $\mathbf{X}$ is called the *reduced path* associated with g.

First of all, the existence of a reduced path is purely topological and requires only the Hausdorff property. Such a path is obtained by a topological procedure of erasing all possible loops of the original path in a maximal way. More precisely, we have the following result.

Proposition 3.4.1. *Let T be a Hausdorff topological space, and $\alpha : [0,1] \to T$ be a continuous path. Then there exist disjoint open intervals $\{I_i : 1 \leqslant i \leqslant \infty\}$ in $(0,1)$ such that the continuous path $\widetilde{\alpha}$ defined by*

$$\widetilde{\alpha}_t = \begin{cases} \alpha_t, & t \in \left(\bigcup_{i=1}^{\infty} I_i\right)^c; \\ \alpha_{\inf I_i}, & t \in I_i, \end{cases} \tag{3.4.1}$$

satisfies the property that if $s \neq t$ and $\widetilde{\alpha}_s = \widetilde{\alpha}_t$, then $s, t \in \overline{I_i}$ for some i.

Proof. The construction relies on Zorn's lemma.

Let

$$\mathcal{P} = \left\{ \bigcup_{i=1}^{\infty} I_i : \ I_i \text{ are disjoint open intervals in } (0,1) \text{ with } \alpha_{\inf I_i} = \alpha_{\sup I_i} \right\}.$$

Define a partial order "$\lesssim$" on $\mathcal{P}$ by inclusion. We claim that $(\mathcal{P}, \lesssim)$ is inductively ordered, i.e. every totally ordered subset of $\mathcal{P}$ has an upper bound.

Let $\mathcal{J} \subset \mathcal{P}$ be a totally ordered subset, and define $\overline{J} = \bigcup_{J \in \mathcal{J}} J$. Then $\overline{J}$ is an open subset of $(0,1)$. It suffices to show that $\overline{J} \in \mathcal{P}$. According to the structure of open sets in $\mathbb{R}^1$, $\overline{J}$ can be written as the countable disjoint union of open intervals $I_i \subset (0,1)$ uniquely up to permutation. Fix any such I_i and $\varepsilon > 0$. Since $[\inf I_i + \varepsilon, \sup I_i - \varepsilon]$ is covered by $J \in \mathcal{J}$, by compactness there exists $J_1, \cdots, J_k \in \mathcal{J}$ such that

$$[\inf I_i + \varepsilon, \sup I_i - \varepsilon] \subset \bigcup_{i=1}^{k} J_i.$$

Since $\mathcal{J}$ is totally ordered, there exists a greatest element among $\{J_1, \cdots, J_k\}$ and assume it to be J_1. It follows that

$$[\inf I_i + \varepsilon, \sup I_i - \varepsilon] \subset J_1.$$

By connectedness, $[\inf I_i + \varepsilon, \sup I_i - \varepsilon]$ is contained in some connected component C of J_1. On the other hand, we know that $C \subset \overline{J}$ and $C \bigcap I_i \neq \emptyset$. Therefore, $C \subset I_i$. Since $J_1 \in \mathcal{P}$ we have $\alpha_{\inf C} = \alpha_{\sup C}$. By letting $\varepsilon \to 0$, it follows from the continuity of α and the Hausdorff property that $\alpha_{\inf I_i} = \alpha_{\sup I_i}$. Therefore, $\overline{J} \in \mathcal{P}$.

According to Zorn's lemma, $\mathcal{P}$ contains a maximal element, say $I = \bigcup_{i=1}^{\infty} I_i$. Define $\widetilde{\alpha}$ by (3.4.1). From this construction and the continuity of α, it is easy to see that for any $t \in [0,1]$ and any open neighborhood U of $\widetilde{\alpha}_t$, there exists some $\delta > 0$ such that

$$s \in (t-\delta, t+\delta) \bigcap [0,1] \implies \widetilde{\alpha}_s \in U.$$

The continuity of $\widetilde{\alpha}$ then follows easily. Now it remains to show that if $s \neq t$ and $\widetilde{\alpha}_s = \widetilde{\alpha}_t$, then $s, t \in \overline{I_i}$ for some i.

In fact, assume that $\widetilde{\alpha}_s = \widetilde{\alpha}_t$ for some $s < t$, and suppose on the contrary that s, t do not belong to the same $\overline{I_i}$ for all i. There are four cases corresponding to whether $\widetilde{\alpha}_s, \widetilde{\alpha}_t$ belong to I. If $s \in I_i, t \in I_j$ for some $i \neq j$, then

$$\alpha_{\inf I_i} = \widetilde{\alpha}_s = \widetilde{\alpha}_t = \alpha_{\sup I_j}.$$

Therefore, $I' := I \bigcup (\inf I_i, \sup I_j)$ is an element in $\mathcal{P}$ strictly containing I, which contradicts the maximality of I. If $s \in I_i$ for some i and $t \notin I$, then

$$\alpha_{\inf I_i} = \widetilde{\alpha}_s = \widetilde{\alpha}_t = \alpha_t$$

Therefore, $I' := I \bigcup (\inf I_i, t)$ is an element in $\mathcal{P}$ strictly containing I (note that

t cannot be equal to $\sup I_i$, otherwise $s, t \in \overline{I_i}$). The remaining two cases can be treated in a similar way.

Now the proof is complete. $\square$

Although the path $\widetilde{\alpha}$ does not contain loops, it is not simple as it stays constant on each I_i. It is another standard topological procedure of obtaining a simple curve from $\widetilde{\alpha}$.

Let $\alpha : [0,1] \to T$ be a continuous path in some Hausdorff space T, and let $\widetilde{\alpha}$ be the path constructed in Proposition 3.4.1. It follows that $\widetilde{\alpha}([0,1])$ is a compact connected Hausdorff subspace of T. From a general fact in topology (see the monograph by S. Willard [65]) that a compact connected Hausdorff space is arcwise-connected (i.e. any two distinct points can be joined by a continuous simple curve), there exists a continuous simple curve $\widehat{\alpha}$ joining $\widetilde{\alpha}_0$ to $\widetilde{\alpha}_1$ with image lying in $\mathrm{Im}\left(\widetilde{\alpha}\right)$. Moreover, we have the following result.

Lemma 3.4.1. *$\widehat{\alpha}([0,1]) = \widetilde{\alpha}([0,1])$, and $\sigma := \widehat{\alpha}^{-1} \circ \widetilde{\alpha} : [0,1] \to [0,1]$ is increasing.*

Proof. We first show that if $\widehat{\alpha}\left([0,1]\right) = \widetilde{\alpha}\left([0,1]\right)$, then σ is increasing. Since $\widehat{\alpha} : [0,1] \to \widehat{\alpha}\left([0,1]\right)$ is a homeomorphism, we know that $\sigma\left(s\right) = \sigma\left(t\right)$ if and only if $\widetilde{\alpha}_s = \widetilde{\alpha}_t$, and by the construction of $\widetilde{\alpha}$ this is equivalent to $s, t \in \overline{I}_i$ for some i. If σ is not increasing, since $\sigma\left(0\right) = 0$ and $\sigma\left(1\right) = 1$, by continuity there exists some $s < u < t$ such that

$$\sigma\left(u\right) < \sigma\left(s\right) = \sigma\left(t\right). \tag{3.4.2}$$

Therefore $s, t \in \overline{I}_i$ for some i and by the construction of $\widetilde{\alpha}$ we know that $\widetilde{\alpha}_u = \widetilde{\alpha}_s = \widetilde{\alpha}_t$, contradicting (3.4.2).

Now we show that $\widehat{\alpha}\left([0,1]\right) = \widetilde{\alpha}\left([0,1]\right)$.

We first prove that for any $0 < t < 1$, $\widetilde{\alpha}\left([0,1]\right) \setminus \{\widetilde{\alpha}_t\}$ is disconnected. In fact, if $t \notin \bigcup_i \overline{I}_i$, then $\widetilde{\alpha}\left([0,1]\right) \setminus \{\widetilde{\alpha}_t\}$ can be written as the disjoint union of $\widetilde{\alpha}\left([0,t)\right)$ and $\widetilde{\alpha}\left((t,1]\right)$ which are both non-empty. By continuity and Hausdorff property, $\widetilde{\alpha}\left([t,1]\right)^c$ is open in T. Since

$$\widetilde{\alpha}\left([0,t)\right) = \widetilde{\alpha}\left([t,1]\right)^c \bigcap \widetilde{\alpha}\left([0,1]\right),$$

we know that $\widetilde{\alpha}\left([0,t)\right)$ is open in $\widetilde{\alpha}\left([0,1]\right)$. Similarly, $\widetilde{\alpha}\left((t,1]\right)$ is open in $\widetilde{\alpha}\left([0,1]\right)$. If $t \in \overline{I}_i$ for some i, then $\widetilde{\alpha}\left([0,1]\right) \setminus \{\widetilde{\alpha}_t\}$ can be written as the disjoint union $\widetilde{\alpha}\left([0, \inf I_i)\right)$ and $\widetilde{\alpha}\left((\sup I_i, 1]\right)$ which are both non-empty. Since

$$\widetilde{\alpha}\left([0, \inf I_i)\right) = \widetilde{\alpha}\left([\inf I_i, 1]\right)^c \bigcap \widetilde{\alpha}\left([0,1]\right),$$

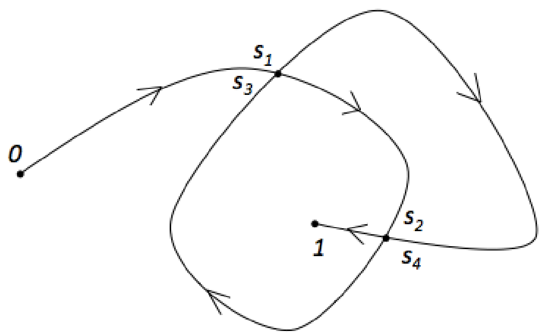

Figure 3.4.1: This figure illustrates the non-uniqueness of the simplified path $\widehat{\alpha}$.

we know that $\widetilde{\alpha}([0,\inf I_i))$ is open in $\widetilde{\alpha}([0,1])$, and similarly for $\widetilde{\alpha}((\sup I_i,1])$. Therefore, $\widetilde{\alpha}([0,1])\setminus\{\widetilde{\alpha}_t\}$ is disconnected.

Suppose on the contrary that there exists some $0<t<1$ such that $\widetilde{\alpha}_t\notin\widehat{\alpha}([0,1])$, then $\widehat{\alpha}([0,1])\subset\widetilde{\alpha}([0,1])\setminus\{\widetilde{\alpha}_t\}$. But this contradicts connectedness since $\widehat{\alpha}(0)$ and $\widehat{\alpha}(1)$ lie in different components of $\widetilde{\alpha}([0,1])\setminus\{\widetilde{\alpha}_t\}$. Therefore $\widehat{\alpha}([0,1])=\widetilde{\alpha}([0,1])$.

Now the proof is complete. □

Remark 3.4.1. In the case when $\alpha_0=\alpha_1$, $\widetilde{\alpha}$ degenerates to a constant point. In this case, the simplified path $\widehat{\alpha}$ should be understood as the constant path.

Remark 3.4.2. The simplified path $\widehat{\alpha}$ associated with α is in general not unique. For example, consider α as in Figure 3.4.1. Apparently there are two ways of erasing the loops of α : either erasing the loop from time s_1 to s_3 or the loop from s_2 to s_4; the resulting simple curves have different images.

Now we apply the previous discussion to signature paths.

Recall that $\mathcal{S}_p$ is the space of signatures for weakly geometric p-rough paths, which can be regarded as a subspace of the infinite tensor algebra $T\left(\mathbb{R}^d\right)$. For any element $\mathbb{X}\in T\left(\mathbb{R}^d\right)$ and multi-index $I=(i_1,\cdots,i_n)\in\{1,\cdots,d\}^n$, we use X^I to denote the I-th component of $\mathbb{X}$ (we have omitted the constant term of $\mathbb{X}$). We also use $X^{(N)}$ to denote the truncation of $\mathbb{X}$ up to degree N. Now define

$$L_d^2=\left\{\mathbb{X}=\left(X^I\right)_{I:\ \mathrm{multi-index}}\in T\left(\mathbb{R}^d\right):\sum_{I:\ \text{multi-index}}\left(X^I\right)^2<\infty\right\}.$$

From Lyons' extension theorem (the factorial decay of signature), it is easy to see that $\mathcal{S}_p$ is a subspace of L_d^2. Moreover, for any $\mathbf{X}\in WG\Omega_p\left(\mathbb{R}^d\right)$, the signature path

$S(\mathbf{X})_{0,\cdot}$ is continuous under the L^2-metric. From now on, we always equip $\mathcal{S}_p$ with the L^2-metric.

Given $g = S(\mathbf{X})_{0,1} \in \mathcal{S}_p$ for some $\mathbf{X} \in WG\Omega_p\left(\mathbb{R}^d\right)$, define a simplified continuous path $\widehat{\mathbb{X}}_\cdot$ associated with the signature path $S(\mathbf{X})_{0,\cdot}$ as before. From the construction it is not hard to see that

$$\left\|\pi^{(N)}\left(\widehat{\mathbb{X}}\right)\right\|_p \leqslant \left\|S_N\left(\mathbf{X}\right)_{0,\cdot}\right\|_p$$

for every $N \in \mathbb{N}$. Therefore, $\widehat{\mathbf{X}} := \pi^{(\lfloor p \rfloor)}\left(\widehat{\mathbb{X}}\right)$ is a weakly geometric p-rough path, and by Lyons' extension theorem $\widehat{\mathbb{X}}$ is the signature path of $\widehat{\mathbf{X}}$.

In contrast to Remark 3.4.2, the uniqueness of the simplified signature path for weakly geometric rough paths is a special and crucial feature of the signature. It is a consequence of the shuffle product formula. Such uniqueness is the key to proving that the signature group can be equipped with a real tree metric and to conclude the necessity of Theorem 3.2.1.

Now our main goal is to establish the following result.

Proposition 3.4.2. *Let $\mathbf{X}, \mathbf{Y}$ be two weakly geometric p-rough paths such that*

$$S(\mathbf{X})_{0,1} = S(\mathbf{Y})_{0,1}.$$

If their corresponding signature paths $S(\mathbf{X})_{0,\cdot}, S(\mathbf{Y})_{0,\cdot}$ are both simple, then they differ by a reparametrization. In particular, $\mathbf{X}$ and $\mathbf{Y}$ differ by a reparametrization.

The main idea of proving Proposition 3.4.2 is to show that $S(\mathbf{X})_{0,\cdot}$ and $S(\mathbf{Y})_{0,\cdot}$ have the same image, which can be proved by contradiction. If not, we are then able to construct a finite dimensional one form such that the path integrals of this one form along the truncated signatures of $\mathbf{X}, \mathbf{Y}$ are different. This certainly leads to a contradiction since the two integrals are functionals of the signature, according to polynomial approximation and the shuffle product formula, provided the one form is smooth enough, and hence they should be identical.

To start with, we first establish the following approximation result.

Lemma 3.4.2. *Let $\mathbf{X}, \mathbf{Y} \in WG\Omega_p\left(\mathbb{R}^d\right)$ with the same signature. Then for any $N \in \mathbb{N}$ and any C^m-one form (continuously differentiable up to order m) ϕ on $\mathbb{R}^{(N)}$ with $m > p$, we have*

$$\int_0^1 \phi\left(dX_u^{(N)}\right) = \int_0^1 \phi\left(dY_u^{(N)}\right). \tag{3.4.3}$$

Proof. From Proposition 1.2.6 (1), we know that $\mathbf{X}, \mathbf{Y}$ are geometric q-rough paths for all $q \in (p, \lfloor p \rfloor + 1)$, and hence the shuffle product formula (Proposition 1.2.1) applies to them. Fix any $q \in (p, \lfloor p \rfloor + 1)$.

Write $\phi = \sum_{|I| \leqslant N} \phi_I dX^I$, where ϕ_I are C_c^m-functions on $\mathbb{R}^{(N)}$. Let K be a compact neighborhood of $X^{(N)}([0,1]) \bigcup Y^{(N)}([0,1])$. According to T. Bagby, L. Bos and N. Levenberg [2], Theorem 1, for each multi-index I, there exists a polynomial sequence $\phi_I^{(n)}$ such that

$$\left\| \phi_I^{(n)} - \phi_I \right\|_{C_K^m} := \sup_{|\alpha| \leqslant m} \sup_K \left| D^\alpha \left(\phi_I^{(n)} - \phi_I \right) \right| \to 0$$

as $n \to \infty$. Let $\phi^{(n)} = \sum_{|I \leqslant N|} \phi_I^{(n)} dX^I$. From the shuffle product formula it is easy to see that (3.4.3) holds for the polynomial one forms $\phi^{(n)}$.

By regarding $\mathbf{X}, \mathbf{Y}$ as geometric q-rough paths, the result follows from the continuity of the integration maps $\phi \mapsto \int_0^1 \phi \left(dX_u^{(N)} \right), \int_0^1 \phi \left(dY_u^{(N)} \right)$ under the Lip^{m-1}-norm (on K) when $m > p$ (see [26], Theorem 10.47) and the fact that the Lip^{m-1}-norm on K is dominated by the C_K^m-norm. □

The crucial point of proving Proposition 3.4.2 is to find a way to reduce the problem to finite dimensions via truncating the signature. For $N \in \mathbb{N}$, we use $\mathbb{R}^{(N)}$ to denote the Euclidean space of truncated tensor elements up to degree N, and such a truncation of $\mathbb{X} \in T\left(\left(\mathbb{R}^d\right)\right)$ is denoted by $X^{(N)}$. We also use $B(\mathbb{X}, R)$ ($B\left(X^{(N)}, R\right)$, respectively) to denote the open ball in L_d^2 (in $\mathbb{R}^{(N)}$,respectively) with radius R.

Firstly, we need the following lemma.

Lemma 3.4.3. *Let $\mathbf{X} \in WG\Omega_p\left(\mathbb{R}^d\right)$ and $\mathbb{X}_\cdot := S(\mathbf{X})_{0,\cdot}$ be a simple curve. Then for any $\varepsilon > 0$, there exists $N(\varepsilon) \in \mathbb{N}$, such that $X_s^{(N)} \neq X_t^{(N)}$ for every $N \geqslant N(\varepsilon)$ and $(s,t) \in \Delta$ with $|t-s| \geqslant \varepsilon$.*

Proof. Let $\Delta_\varepsilon = \{(s,t) \in \Delta : \ t - s \geqslant \varepsilon\}$. For each $(s,t) \in \Delta_\varepsilon$, since $\mathbb{X}_s \neq \mathbb{X}_t$, there exists some $N_{s,t} \in \mathbb{N}$ such that

$$X_s^{(N_{s,t})} \neq X_t^{(N_{s,t})}. \tag{3.4.4}$$

By continuity, (3.4.4) holds in a neighborhood of (s,t). The result then follows easily from a compactness argument on Δ_ε. □

The following result is the key to constructing the finite dimensional one form mentioned before.

Lemma 3.4.4. *Let* $\mathbf{X} \in WG\Omega_p\left(\mathbb{R}^d\right)$ *and* $\mathbb{X}$ *be the corresponding signature path. Suppose for some* $s < \tau < t$ *that* $\mathbb{X}|_{[s,t]}$ *is simple,* $\|\mathbb{X}_s - \mathbb{X}_\tau\| = \|\mathbb{X}_t - \mathbb{X}_\tau\| = R$, and $\mathbb{X}|_{(s,t)} \subset B(\mathbb{X}_\tau, R)$ where $0 < R < 1$. *Then there exists some* $N_0 \in \mathbb{N}$, *such that for any* $N \geqslant N_0$, *there exists some* $h \in C_c^m\left(\mathbb{R}^{(N)}\right)$ *($m > p$) supported on the closed ball* $\overline{B^{(N)}}\left(X_\tau^{(N)}, R\right)$ *with*

$$\int_s^t \phi\left(dX_u^{(N)}\right) \neq 0,$$

where the C_c^m*-one form* ϕ *on* $\mathbb{R}^{(N)}$ *is defined by*

$$\phi\left(X^{(N)}\right) = \sum_{|I| \leqslant N} h\left(X^{(N)}\right) X^I dX^I.$$

The construction of the one form ϕ in Lemma 3.4.4 is local. The underlying idea is that ϕ is "almost" supported near $X_s^{(N)}$ and $X_t^{(N)}$, and more importantly it is locally radial near these two points with respect to $X_\tau^{(N)}$ but globally not. This breaks the symmetry of ϕ and prevents the exact cancellation of the path integral, so that it should be nonzero. The global construction of ϕ relies on the following general result from differential geometry, which is usually known as the partition of unity. We refer the reader to the monograph by S.S. Chern, W. Chen and K.S. Lam [10] for the proof.

Lemma 3.4.5. *(Partition of Unity) Let* $\{U_i\}_{i=1}^k$ *be an open cover of a differentiable manifold* M. *Then there exists* C^∞*-functions* $\{\varphi_i\}_{i=1}^k$ *on* M *such that*

(1) $0 \leqslant \varphi_i \leqslant 1$;

(2) $\mathrm{supp}\varphi_i \subset U_i$;

(3) $\sum_{i=1}^k \varphi_i = 1$.

Proof of Lemma 3.4.4. Let $\varepsilon > 0$ be such that $\overline{B}(\mathbb{X}_s, \varepsilon) \bigcap \overline{B}(\mathbb{X}_t, \varepsilon) = \emptyset$, and define

$$\begin{aligned} s_1 &= \inf\left\{u \in [s, \tau] : \ \mathbb{X}_u \in B(\mathbb{X}_s, \varepsilon)^c\right\}, \\ t_1 &= \sup\left\{u \in [\tau, t] : \ \mathbb{X}_u \in B(\mathbb{X}_t, \varepsilon)^c\right\}. \end{aligned}$$

It follows that $\mathbb{X}([s, s_1)) \subset B(\mathbb{X}_s, \varepsilon)$, $\mathbb{X}((t_1, t]) \subset B(\mathbb{X}_t, \varepsilon)$. Moreover, by assumption and continuity, there exists some $0 < \rho < R$, such that $\mathbb{X}([s_1, t_1]) \subset B(\mathbb{X}_\tau, \rho)$. Let

$$A_{\rho,R} = \left\{\mathbb{X} \in L_d^2 : \ \rho < \|\mathbb{X} - \mathbb{X}_\tau\| < R\right\}$$

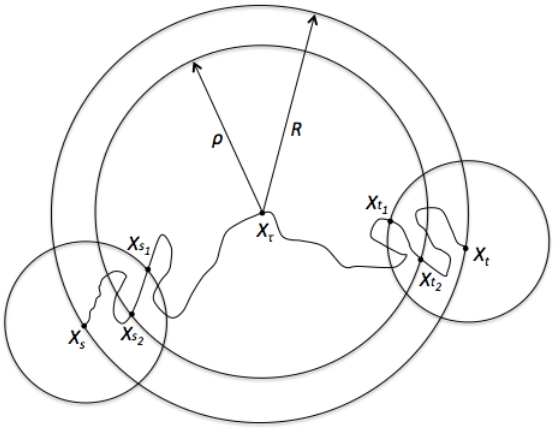

Figure 3.4.2: This figure illustrates the corresponding infinite dimensional configuration in the proof of Lemma 3.4.4.

be the open annulus, and define

$$\begin{aligned} s_2 &= \sup\left\{u \in [s,\tau] : \ \mathbb{X}_u \in \overline{A}_{\rho,R}\right\}, \\ t_2 &= \inf\left\{u \in [\tau,t] : \ \mathbb{X}_u \in \overline{A}_{\rho,R}\right\}. \end{aligned} \tag{3.4.5}$$

It follows that $s < s_2 < s_1$, $t_1 < t_2 < t$ and $\mathbb{X}((s_2,t_2)) \subset B(\mathbb{X}_\tau,\rho)$. Figure 3.4.2 illustrates the corresponding infinite dimensional configuration.

Let

$$\varepsilon_1 = \max\left\{ \sup_{u\in[s,s_2]} \|\mathbb{X}_u - \mathbb{X}_s\|, \sup_{u\in[t_2,t]} \|\mathbb{X}_u - \mathbb{X}_t\| \right\} < \varepsilon. \tag{3.4.6}$$

Then there exists some $N_1 \in \mathbb{N}$, such that for each $N \geqslant N_1$,

$$\sqrt{\sum_{|I|>N} |X_s^I|^2} < \frac{\varepsilon-\varepsilon_1}{2}, \ \sqrt{\sum_{|I|>N} |X_t^I|^2} < \frac{\varepsilon-\varepsilon_1}{2}.$$

It follows that for each $N \geqslant N_1$,

$$\overline{B^{(N)}}\left(X_s^{(N)}, \frac{\varepsilon+\varepsilon_1}{2}\right) \bigcap \overline{B^{(N)}}\left(X_t^{(N)}, \frac{\varepsilon+\varepsilon_1}{2}\right) = \emptyset \tag{3.4.7}$$

and

$$X^{(N)}([s,s_2]) \subset B^{(N)}\left(X_s^{(N)}, \frac{\varepsilon+\varepsilon_1}{2}\right), \ X^{(N)}([t_2,t]) \subset B^{(N)}\left(X_t^{(N)}, \frac{\varepsilon+\varepsilon_1}{2}\right).$$

Of course we also have $X^{(N)}([s_2,t_2]) \subset \overline{B^{(N)}}\left(X_\tau^{(N)},\rho\right)$. On the other hand, since

$$\|X_s - X_\tau\| = \|X_t - X_\tau\| = R,$$

for any fixed $R_1 \in (\rho, R)$, there exists $N_2 \in \mathbb{N}$, such that for each $N \geqslant N_2$,

$$R_1 < \left\|X_s^{(N)} - X_\tau^{(N)}\right\|, \left\|X_t^{(N)} - X_\tau^{(N)}\right\| \leqslant R.$$

We take $N_0 = N_1 \vee N_2$ and fix any $N \geqslant N_0$.

Let

$$A_{\rho,R}^{(N)} = \left\{X^{(N)} \in \mathbb{R}^{(N)} :\ \rho < \left\|X^{(N)} - X_\tau^{(N)}\right\| < R\right\}$$

be the finite dimensional open annulus, and define

$$\begin{aligned} s_3 &= \sup\left\{u \in [s,\tau] :\ X_u^{(N)} \in \overline{A^{(N)}}_{\rho,R}\right\}, \\ t_3 &= \inf\left\{u \in [\tau,t] :\ X_u^{(N)} \in \overline{A^{(N)}}_{\rho,R}\right\}. \end{aligned}$$

Then $s < s_3 \leqslant s_2$, $t_2 \leqslant t_3 < t$,

$$\left\|X_{s_3}^{(N)} - X_\tau^{(N)}\right\| = \left\|X_{t_3}^{(N)} - X_\tau^{(N)}\right\| = \rho,$$

and

$$X^{(N)}((s_3,t_3)) \subset B^{(N)}\left(X_\tau^{(N)},\rho\right). \tag{3.4.8}$$

Consider a small open neighborhood $A_{\rho-\eta,R+\eta}^{(N)}$ of $\overline{A^{(N)}}_{\rho,R}$, and let U, V be two open subsets of $A_{\rho-\eta,R+\eta}^{(N)}$ such that $A_{\rho-\eta,R+\eta}^{(N)} = U \bigcup V$, and

$$\begin{aligned} \overline{B^{(N)}}\left(X_s^{(N)}, \frac{\varepsilon+\varepsilon_1}{2}\right) \bigcap A_{\rho-\eta,R+\eta}^{(N)} &\subset U \bigcap V^c, \\ \overline{B^{(N)}}\left(X_t^{(N)}, \frac{\varepsilon+\varepsilon_1}{2}\right) \bigcap A_{\rho-\eta,R+\eta}^{(N)} &\subset V \bigcap U^c. \end{aligned} \tag{3.4.9}$$

This is possible because of (3.4.7). Figure 3.4.3 illustrates the corresponding finite dimensional configuration .

Let $h_i(r) \in C_c^m(\mathbb{R}^1)$ $(i = 1,2)$ be such that h_i are supported on $[\rho, R]$. Take $\{\varphi_1, \varphi_2\}$ to be a partition of unity on $A_{\rho-\eta,R+\eta}^{(N)}$ subordinate to the open cover $\{U, V\}$

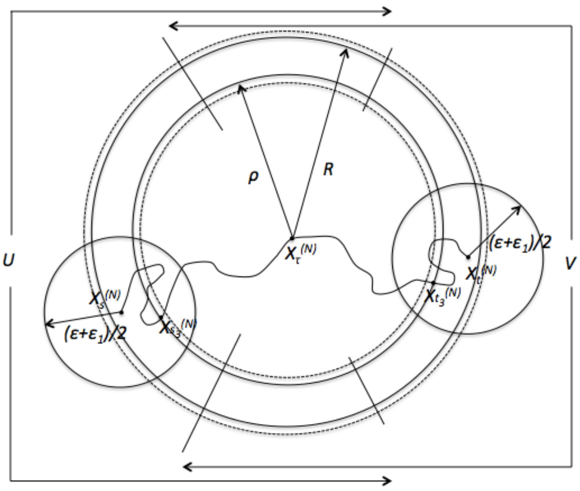

Figure 3.4.3: This figure illustrates the corresponding finite dimensional configuration in the proof of Lemma 3.4.4.

according to Lemma 3.4.5. Define $h \in C_c^m\left(\mathbb{R}^{(N)}\right)$ by

$$h\left(X^{(N)}\right) = \begin{cases} 0, & X^{(N)} \in \overline{B^{(N)}}\left(X_\tau^{(N)}, \rho\right); \\ \sum_{i=1}^{2} \varphi_i\left(X^{(N)}\right) h_i\left(\left\|X^{(N)} - X_\tau^{(N)}\right\|\right), & X^{(N)} \in A_{\rho,R}^{(N)}; \\ 0, & X^{(N)} \in B^{(N)}\left(X_\tau^{(N)}, R\right)^c. \end{cases}$$

It is not hard to see that h is well-defined, of class C^m and compactly supported on $\overline{B^{(N)}}\left(X_\tau^{(N)}, R\right)$. Consider the C_c^m-one form φ on $\mathbb{R}^{(N)}$ defined by

$$\phi = \sum_{|I| \leqslant N} h\left(X^{(N)}\right) X^I dX^I.$$

It is crucial point to notice that ϕ is radial in both of $B^{(N)}\left(X^{(N)}, (\varepsilon + \varepsilon_1)/2\right)$ and $B^{(N)}\left(X^{(N)}, (\varepsilon + \varepsilon_1)/2\right)$, although globally it is not. It follows from (3.4.8), (3.4.9) and the definition of h that

$$\begin{aligned} \int_s^t \phi\left(dX_u^{(N)}\right) &= \int_s^{s_3} \phi\left(dX_u^{(N)}\right) + \int_{s_3}^{t_3} \phi\left(dX_u^{(N)}\right) + \int_{t_3}^{t} \phi\left(dX_u^{(N)}\right) \\ &= \frac{1}{2}\left(\int_s^{s_3} h_1\left(r_u^{(N)}\right) d\left(r_u^{(N)}\right)^2 + \int_{t_3}^{t} h_2\left(r_u^{(N)}\right) d\left(r_u^{(N)}\right)^2\right) \\ &= \frac{1}{2}\left(\int_{r_s^{(N)}}^{\rho} h_1(r) dr^2 + \int_{\rho}^{r_t^{(N)}} h_2(r) dr^2\right). \end{aligned} \tag{3.4.10}$$

where $r_u^{(N)} := \left\| X_u^{(N)} - X_\tau^{(N)} \right\|$ denotes the radial vector.

Obviously we can find h_1, h_2 such that (3.4.10) is nonzero. For instance, we can take

$$h_i(r) = \begin{cases} \left(r^2-\rho^2\right)^{k_i}\left(R^2-r^2\right)^{k_i}, & \rho \leqslant r \leqslant R; \\ 0, & \text{otherwise}, \end{cases} \tag{3.4.11}$$

where $k_1, k_2 > K$ are large positive integers to be taken later on. It follows that

$$\begin{aligned} & \int_s^t \phi\left(dX_u^{(N)}\right) \\ = \; & \frac{1}{2}\left(\int_\rho^{r_t^{(N)}} h_2(r)dr^2 - \int_\rho^{r_s^{(N)}} h_1(r)dr^2\right) \\ \geqslant \; & \frac{1}{2}\left(\int_\rho^{R_1} \left(r^2-\rho^2\right)^{k_2}\left(R^2-r^2\right)^{k_2} dr^2 - \int_\rho^{R} \left(r^2-\rho^2\right)^{k_1}\left(R^2-r^2\right)^{k_1} dr^2\right). \end{aligned}$$

Since $0 < R < 1$, when fixing k_2 we can choose k_1 large enough so that the R.H.S. is strictly positive.

Now the proof is complete.

□

Now we are in a position to prove Proposition 3.4.2.

Proof of Proposition 3.4.2. It suffices to show that $\mathbb{X}_\cdot := S(\mathbf{X})_{0,\cdot}$ and $\mathbb{Y} := S(\mathbf{Y})_{0,\cdot}$ have the same image.

Assume on the contrary that $\mathbb{X}_\tau \notin \mathbb{Y}([0,1])$ for some $\tau \in (0,1)$. Then there exists some $0 < R_0 < 1$ and some $N_3 \in \mathbb{N}$, such that for each $N \geqslant N_3$ and $u \in [0,1]$, we have

$$\|Y_u^{(N)} - X_\tau^{(N)}\| > R_0. \tag{3.4.12}$$

Define

$$\begin{aligned} s_0 &= \sup\{u \in [0,\tau] : \; \mathbb{X}_u \in B(\mathbb{X}_\tau, R_0)^c\}, \\ t_0 &= \inf\{u \in [\tau,1] : \; \mathbb{X}_u \in B(\mathbb{X}_\tau, R_0)^c\}. \end{aligned}$$

Choose $\varepsilon > 0$ so that $\overline{B}(\mathbb{X}_{s_0}, 2\varepsilon) \bigcap \overline{B}(\mathbb{X}_{t_0}, 2\varepsilon) = \emptyset$. For such ε, there exists some $\delta > 0$, such that

$$|u - v| < \delta \implies \|\mathbb{X}_u - \mathbb{X}_v\| < \varepsilon.$$

By the uniform continuity of $\mathbb{X}^{-1}$, there exists $\varepsilon_0 << \varepsilon$, such that

$$\|\mathbb{X}_u - \mathbb{X}_v\| < \varepsilon_0 \implies |u - v| < \delta.$$

Consider the balls $B(\mathbb{X}_{s_0}, \varepsilon_0)$ and $B(\mathbb{X}_{t_0}, \varepsilon_0)$. By arguing in the same way as in the beginning of the proof of Lemma 3.4.4, we can find some $s \in (s_0, \tau)$, $t \in (\tau, t_0)$ and some $0 < R < R_0$, such that

$$\mathbb{X}([s_0, s]) \subset B(\mathbb{X}_{s_0}, \varepsilon_0),\ \mathbb{X}([t, t_0]) \subset B(\mathbb{X}_{t_0}, \varepsilon_0),$$

and

$$\|\mathbb{X}_s - \mathbb{X}_\tau\| = \|\mathbb{X}_t - \mathbb{X}_\tau\| = R,\ \mathbb{X}((s,t)) \subset B(\mathbb{X}_\tau, R).$$

Also note that

$$\mathbb{X}([s_0, s]) \subset B(\mathbb{X}_s, \varepsilon),\ \mathbb{X}([t, t_0]) \subset B(\mathbb{X}_t, \varepsilon).$$

By applying Lemma 3.4.4 to $\mathbb{X}|_{[s,t]}$ and $B(\mathbb{X}_\tau, R)$, we know that there exists some $N_0 \in \mathbb{N}$, such that for each $N \geqslant N_0$, there exists some $h \in C_c^m\left(\mathbb{R}^{(N)}\right)$ $(m > p)$ supported on $\overline{B^{(N)}}\left(X_\tau^{(N)}, R\right)$ with

$$\int_s^t \phi\left(dX_u^{(N)}\right) \neq 0,$$

where

$$\phi = \sum_{|I| \leqslant N} h\left(X^{(N)}\right) X^I dX^I.$$

Here when applying the proof of Lemma 3.4.4 we should start with the disjoint balls $\overline{B}(\mathbb{X}_s, \varepsilon)$ and $\overline{B}(\mathbb{X}_t, \varepsilon)$. Moreover, in order to take the pieces $\mathbb{X}|_{[s_0,s]}$ and $\mathbb{X}|_{[t,t_0]}$ into account, the "ε_1" defined by (3.4.6) should be

$$\varepsilon_1' := \max\left\{\sup_{u \in [s_0, s_2]} \|\mathbb{X}_u - \mathbb{X}_s\|, \sup_{u \in [t_2, t_0]} \|\mathbb{X}_u - \mathbb{X}_t\|\right\} < \varepsilon$$

here, where s_2, t_2 are the last exit and first entry times for the closed annulus $\overline{A}_{\rho,R}$ defined by (3.4.5). It then follows from our explicit construction of h that ϕ is radial along $X^{(N)}|_{[s_0,s]}$ and $X^{(N)}|_{[t,t_0]}$, since they lie inside the balls $B^{(N)}\left(X_s^{(N)}, (\varepsilon + \varepsilon_1')/2\right)$ and $B^{(N)}\left(X_t^{(N)}, (\varepsilon + \varepsilon_1')/2\right)$ respectively.

Fix $R_2 \in (R, R_0)$. Choose $N_4 \in \mathbb{N}$ such that for each $N \geqslant N_4$,

$$R_2 < \left\| X_{s_0}^{(N)} - X_{\tau}^{(N)} \right\|, \left\| X_{t_0}^{(N)} - X_{\tau}^{(N)} \right\| \leqslant R_0.$$

Moreover, for the positive number $R_2 - R$, there exists some $\eta > 0$, such that

$$|u - v| < \eta \implies \|\mathbb{X}_u - \mathbb{X}_v\| < R_2 - R.$$

We then apply Lemma 3.4.3 for η to get some $N_5 \in \mathbb{N}$, such that $X_s^{(N)} \neq X_t^{(N)}$ for every $N \geqslant N_5$ and $(s,t) \in \Delta$ with $t - s \geqslant \eta$.

Now take $N \geqslant \max\{N_0, N_3, N_4, N_5\}$, and define

$$\begin{aligned} s_4 &= \inf\left\{u \in [s_0, \tau] : \; X_u^{(N)} \in \overline{B^{(N)}}\left(X_\tau^{(N)}, R\right)\right\}, \\ t_4 &= \sup\left\{u \in [\tau, t_0] : \; X_u^{(N)} \in \overline{B^{(N)}}\left(X_\tau^{(N)}, R\right)\right\}. \end{aligned}$$

It follows that

$$\|\mathbb{X}_{s_0} - \mathbb{X}_{s_4}\| \geqslant \left\| X_{s_0}^{(N)} - X_{s_4}^{(N)} \right\| \geqslant R_2 - R,$$

and hence $s_4 - s_0 \geqslant \eta$. Similarly we have $t_0 - t_4 \geqslant \eta$. Therefore, we know that

$$X^{(N)}([s_4, t_4]) \bigcap X^{(N)}([0, s_0] \bigcup [t_0, 1]) = \emptyset.$$

By continuity, there exists open sets $U \subset\subset V$ in $\mathbb{R}^{(N)}$ such that

$$X^{(N)}([s_4, t_4]) \subset U, \;\; X^{(N)}([0, s_0] \bigcup [t_0, 1]) \bigcup \mathrm{Im}\left(Y^{(N)}\right) \subset V^c.$$

Figure 3.4.4 illustrates the corresponding finite dimensional configuration.

Let $\zeta \in C_c^\infty\left(\mathbb{R}^{(N)}\right)$ be a bump function with respect to $\{U, V\}$, i.e. $0 \leqslant \zeta \leqslant 1$, $\zeta = 1$ on U and $\zeta = 0$ on V^c (see [10] for the construction of bump functions). Define the C_c^m-one form Φ on $\mathbb{R}^{(N)}$ by $\Phi = \zeta \cdot \phi$. It follows that

$$\begin{aligned} \int_0^1 \Phi\left(dX_u^{(N)}\right) &= \int_0^{s_0} \Phi\left(dX_u^{(N)}\right) + \int_{s_0}^{t_0} \Phi\left(dX_u^{(N)}\right) + \int_{t_0}^1 \Phi\left(dX_u^{(N)}\right) \\ &= \int_{s_0}^{t_0} \zeta\left(X_u^{(N)}\right) \phi\left(dX_u^{(N)}\right). \end{aligned}$$

Since ϕ is supported on $\overline{B^{(N)}}\left(X_\tau^{(N)}, R\right)$, we know from the definition of s_4, t_4 and ζ

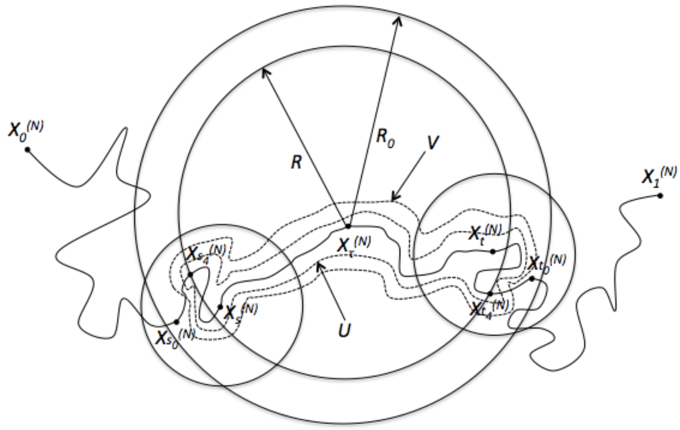

Figure 3.4.4: This figure illustrates the corresponding finite dimensional configuration in the proof of Proposition 3.4.2.

that

$$\int_0^1 \Phi\left(dX_u^{(N)}\right) = \int_{s_4}^{t_4} \phi\left(dX_u^{(N)}\right).$$

Based on the explicit construction of ϕ, it follows from the local radial property of ϕ and (3.4.10) that

$$\begin{aligned}\int_{s_4}^{t_4} \phi\left(dX_u^{(N)}\right) &= \frac{1}{2}\left(\left(\int_R^{r_s^{(N)}} + \int_{r_s^{(N)}}^{\rho}\right) h_1(r)dr^2 + \left(\int_\rho^{r_t^{(N)}} + \int_{r_t^{(N)}}^{R}\right) h_2(r)dr^2\right) \\ &= \frac{1}{2}\left(\int_\rho^R (h_2(r) - h_1(r))dr^2\right).\end{aligned}$$

This is nonzero if for instance we take h_1 and h_2 as given by (3.4.11) with $k_1 >> k_2 > K$. Therefore we arrive at

$$\int_0^1 \Phi\left(dX_u^{(N)}\right) \neq 0.$$

On the other hand, since Φ is supported on $\overline{B^{(N)}}\left(X_\tau^{(N)}, R\right) \subset \overline{B^{(N)}}\left(X_\tau^{(N)}, R_0\right)$, we know from (3.4.12) that

$$\int_0^1 \Phi\left(dY_u^{(N)}\right) = 0.$$

This is a contradiction to Lemma 3.4.2.

Now the proof is complete.

□

Remark 3.4.3. In the proofs of Lemma 3.4.4 and Proposition 3.4.2, we have implicitly used the following fact: the truncated signature path $X^{(N)}$ of $\mathbf{X}$, regarded as a continuous path taking values in the truncated tensor algebra $W = T^{(N)}(\mathbb{R}^d)$, admits a canonical lifting as a weakly geometric p-rough path over W whose signature is uniquely determined (and can be explicitly computed) by the signature of $\mathbf{X}$. The proof of this fact can be obtained by first looking at the case of $p = 1$ and then using approximation based on the geometric rough path nature of $\mathbf{X}$.

Remark 3.4.4. Proposition 3.4.2 itself already gives the uniqueness of signature for simple weakly geometric rough paths: if $\mathbf{X}, \mathbf{Y} \in WG\Omega_p(\mathbb{R}^d)$ are both simple and have the same signature, then they differ by a reparametrization.

To conclude the necessity of Theorem 3.2.1, we now show that the signature group $\mathcal{S}_p$ can be equipped with a tree metric d, and if $\mathbf{X} \in WG\Omega_p(\mathbb{R}^d)$ has trivial signature, then it is a tree-like path realized on the real tree $(\mathcal{S}_p, d)$.

For $g \in \mathcal{S}_p$, let $\mathbb{X}^g$ and $\mathbf{X}^g$ denote the unique simplified signature path and reduced weakly geometric path (choose some parametrization) associated with g respectively.

Introduce the relation "$\lesssim$" on $\mathcal{S}_p$ by

$$g \lesssim h \text{ iff } \mathrm{Im}(\mathbb{X}^g) \subset \mathrm{Im}(\mathbb{X}^h).$$

It is easy to see that "$\lesssim$" defines a partial order on $\mathcal{S}_p$. Now define a functional $L: \mathcal{S}_p \to [0, \infty)$ by

$$L(g) = \|\mathbf{X}^g\|_p^p, \ g \in \mathcal{S}_p.$$

Then we have the following result.

Proposition 3.4.3. *The partially ordered set $(\mathcal{S}_p, \lesssim)$ together with the functional L satisfies the four conditions in Proposition 3.2.4. In particular, the metric given by*

$$d(g, h) = L(g) + L(h) - 2L(g \wedge h), \ g, h \in \mathcal{S}_p,$$

defines a real tree metric on $\mathcal{S}_p$.

Proof. (1) Obviously, the unit element "$\mathbf{1}$" is the unique minimal element under the partial order "$\lesssim$".

(2) Let $g, h \in \mathcal{S}_p$. Set

$$t^* = \inf\left\{t \in [0,1] : \ \mathbb{X}_t^h \notin \mathrm{Im}(\mathbb{X}^g)\right\},$$

and define $g \wedge h = \mathbb{X}_{t^*}^h$. We now show that $g \wedge h$ is an infimum of $\{g, h\}$. In fact, it is obvious that $g \wedge h \lesssim h$ since $\mathbb{X}^{g \wedge h} = \mathbb{X}^h|_{[0,t^*]}$. Moreover, by continuity we know that

$\mathbb{X}^h_{t^*} = \mathbb{X}^g_{s^*}$ for some $s^* \in [0,1]$, and by the uniqueness of simplified signature path (Proposition 3.4.2) we have

$$\mathbb{X}^{g\wedge h} = \mathbb{X}^h|_{[0,t^*]} = \mathbb{X}^g|_{[0,s^*]},$$

which implies that $g \wedge h \lesssim g$. Now assume that $w \in \mathcal{S}_p$ with $w \lesssim g, h$. It follows that

$$\mathbb{X}^w([0,1]) = \mathbb{X}^h([0,t']) \subset \mathrm{Im}\,(\mathbb{X}^g)$$

for some $t' \in [0,1]$. Apparently $t^* \geqslant t'$ and hence $w \lesssim g \wedge h$. Therefore, $g \wedge h$ is an infimum of $\{g,h\}$. By the uniqueness of infimum, this already implies that $g \wedge h = h \wedge g$ is the unique infimum of $\{g,h\}$.

(3) Let $g \in \mathcal{S}_p$ and $h, w \lesssim g$. Then there exists unique s^*, t^* such that $h = \mathbb{X}^g_{s^*}$ and $w = \mathbb{X}^g_{t^*}$ (unless $g = \mathbf{1}$ which is a trivial case). Apparently $h \lesssim w$ or $w \lesssim h$ according to whether $s^* \leqslant t^*$ or $t^* \leqslant s^*$.

(4) It is obvious that L is increasing. Moreover, let $g, h \in \mathcal{S}_p$ with g strictly less than h. It follows that there exists some $0 \leqslant t^* < 1$ such that $\mathbb{X}^g = \mathbb{X}^h|_{[0,t^*]}$. Apparently $\mathbf{X}^h|_{[t^*,1]}$ cannot be a constant path otherwise $g = h$. It follows that

$$L(g) = \|\mathbf{X}^g\|_p^p = \left\|\mathbf{X}^h|_{[0,t^*]}\right\|_p^p < \left\|\mathbf{X}^h\right\|_p^p = L(h),$$

and hence L is strictly increasing. Now let A be a full, totally ordered subset of $\mathcal{S}_p$. It follows that $L: \ A \to L(A)$ is a bijection. To complete the proof, it remains to show that $L(A)$ is a real interval. Indeed, let

$$\theta = \inf L(A),\ \ \Theta = \sup L(A),$$

and $c \in (\theta, \Theta)$. It follows that for some $g, h \in A$, we have

$$L(g) < c < L(h).$$

Since A is totally ordered, g must be strictly less than h, and there exists a unique $0 \leqslant t^* < 1$ such that $\mathbb{X}^g = \mathbb{X}^h|_{[0,t^*]}$. If we define

$$\varphi(t) = \left\|\mathbf{X}^h|_{[0,t]}\right\|_p^p,\ \ t \in [t^*,1],$$

by [26], Proposition 5.8 we know that φ is continuous. Therefore, there exists some (unique) $t^* < t' < 1$ such that

$$\varphi(t') = \left\|\mathbf{X}^h|_{[0,t']}\right\|_p^p = c.$$

Let $w = \mathbb{X}^h_{t'} \in \mathcal{S}_p$. It follows that $g \leqslant w \leqslant h$ and hence $w \in A$. In particular, $c \in L(A)$. Now we have

$$(\theta, \Theta) \subset L(A) \subset [\theta, \Theta],$$

which of course implies that $L(A)$ is a real interval.

Now the proof is complete. □

Corollary 3.4.1. *For $g \in \mathcal{S}_p$, the reduced path $\mathbf{X}^g$ is p-variation minimizing among all weakly geometric p-rough paths with signature g.*

Proof. For any $\mathbf{X} \in WG\Omega_p\left(\mathbb{R}^d\right)$ with $S(\mathbf{X})_{0,1} = g$, we apply Proposition 3.4.1 and Lemma 3.4.1 to the signature path $S(\mathbf{X})_{0,\cdot}$ to obtain a continuous simple curve $\widehat{\mathbb{X}}$ joining $\mathbf{1}$ to g. From the previous discussion we know that

$$\left\|\pi^{(\lfloor p\rfloor)}\left(\widehat{\mathbb{X}}\right)\right\|_p \leqslant \|\mathbf{X}\|_p.$$

On the other hand, by the uniqueness of simplified signature path and reduced path, $\mathbf{X}^g$ and $\pi^{(\lfloor p\rfloor)}\left(\widehat{\mathbb{X}}\right)$ differ by a reparametrization, and hence have the same p-variation. Therefore,

$$\|\mathbf{X}^g\|_p \leqslant \|\mathbf{X}\|_p,$$

and $\mathbf{X}^g$ is a p-variation minimizer. □

Remark 3.4.5. In general, $\mathbf{X}^g$ is not the unique p-variation minimizer (up to reparametrization in the general sense).

To see this, consider the following example for $d = 2$ and $1 < p < 2$. Let $\widehat{AB}$ be an arc of the unit circle centered at $O \in \mathbb{R}^2$ with central angle θ_0, and let C be the midpoint of $\widehat{AB}$. Let D be a point on the extension of the radius vector $\overrightarrow{OC}$ and let $|CD| = \varepsilon$. Consider the paths $x, y:\ [0,1] \to \mathbb{R}^2$ defined by the trajectories

$$x = \widehat{AC} \sqcup \overrightarrow{CD} \sqcup \overrightarrow{DC} \sqcup \widehat{CB},\ y = \widehat{AB},$$

respectively, where "$\sqcup$" means concatenation. It is easy to see that x, y have the same signature g and y is the reduced path of x (obviously they are not equal up to reparametrization in the general sense). Now we show that $\|x\|_p = \|y\|_p = \left|\overrightarrow{AB}\right|$ provided θ_0 and ε are small enough.

In fact, let $E \in \widehat{AB}$ and denote the central angle $\angle EOB$ by θ. Consider the function

$$f(\theta) = \left|\overrightarrow{AE}\right|^p + \left|\overrightarrow{EB}\right|^p, \ \theta \in [0, \theta_0],$$

which can be written as

$$f(\theta) = 2^p \left(\sin^p \frac{\theta}{2} + \sin^p \frac{\theta_0 - \theta}{2} \right)$$

according to Euclidean geometry. Computing the second derivative of f, we obtain that

$$f''(\theta) = \frac{p}{2^{2-p}} \left((p-1) \left(\frac{\cos^2 \frac{\theta}{2}}{\sin^{2-p} \frac{\theta}{2}} + \frac{\cos^2 \frac{\theta_0 - \theta}{2}}{\sin^{2-p} \frac{\theta_0 - \theta}{2}} \right) - \left(\sin^p \frac{\theta}{2} + \sin^p \frac{\theta_0 - \theta}{2} \right) \right).$$

Since $1 < p < 2$, we konw that when θ_0 is small, $f''(\theta)$ is uniformly positive and hence f is convex on $[0, \theta]$. Also note that $f(0) = f(\theta_0) = \left|\overrightarrow{AB}\right|^p$. Therefore, for θ_0 small enough f obtains its maximum on the end points and we have

$$\left|\overrightarrow{AE}\right|^p + \left|\overrightarrow{EB}\right|^p \leqslant \left|\overrightarrow{AB}\right|^p, \ \forall E \in \widehat{AB}.$$

Now we fix such a θ_0. This already implies that $\|y\|_p = \left|\overrightarrow{AB}\right|$. Moreover, by considering the symmetry of $f(\theta)$ it is easy to see that f obtains its minimum at $\theta = \theta_0/2$. Set

$$\lambda = \left|\overrightarrow{AB}\right|^p - \left|\overrightarrow{AC}\right|^p - \left|\overrightarrow{CB}\right|^p > 0.$$

It remains to show that when ε is small enough, $\|x\|_p = \left|\overrightarrow{AB}\right|$. To this end, let

$$\mathcal{P}: \ 0 = t_0 < t_1 < \cdots < t_n = 1$$

be a finite partition of $[0, 1]$, and let t_k, t_l be the first and last partition points at which x is in $\overline{CD}$ respectively. If such points don't exist, then obviously we have

$$\sum_{i=1}^{n} \left|x_{t_i} - x_{t_{i-1}}\right|^p \leqslant \left|\overrightarrow{AB}\right|^p.$$

Otherwise, we have

$$\begin{aligned}
\sum_{i=1}^{n}\left|x_{t_i}-x_{t_{i-1}}\right|^p &= \sum_{i=1}^{k-1}\left|x_{t_i}-x_{t_{i-1}}\right|^p+\left|x_{t_k}-x_{t_{k-1}}\right|^p \\
&\quad +\sum_{i=k+1}^{l}\left|x_{t_i}-x_{t_{i-1}}\right|^p+\left|x_{t_{l+1}}-x_{t_l}\right|^p+\sum_{i=l+2}^{n}\left|x_{t_i}-x_{t_{i-1}}\right|^p \\
&\leqslant \left|x_{t_{k-1}}-A\right|^p+\left|B-x_{t_{l+1}}\right|^p+\left|x_{t_k}-x_{t_{k-1}}\right|^p \\
&\quad +\left|x_{t_{l+1}}-x_{t_l}\right|^p+2\varepsilon^p,
\end{aligned}$$

where we have used the previous discussion and the fact that $\overrightarrow{CD}$ is a geodesic. It follows that

$$\begin{aligned}
&\sum_{i=1}^{n}\left|x_{t_i}-x_{t_{i-1}}\right|^p \\
&\leqslant\left|\overrightarrow{AC}\right|^p+\left|\overrightarrow{CB}\right|^p+\left(\left|x_{t_k}-x_{t_{k-1}}\right|^p-\left|C-x_{t_{k-1}}\right|^p\right) \\
&\quad+\left(\left|x_{t_{l+1}}-x_{t_l}\right|^p-\left|x_{t_{l+1}}-C\right|^p\right)+2\varepsilon^p \\
&=\left|\overrightarrow{AB}\right|^p-\left(\left|\overrightarrow{AB}\right|^p-\left|\overrightarrow{AC}\right|^p-\left|\overrightarrow{CB}\right|^p-\left(\left|x_{t_k}-x_{t_{k-1}}\right|^p-\left|C-x_{t_{k-1}}\right|^p\right)\right. \\
&\quad\left.-\left(\left|x_{t_{l+1}}-x_{t_l}\right|^p-\left|x_{t_{l+1}}-C\right|^p\right)-2\varepsilon^p\right).
\end{aligned}$$

On the other hand, we have

$$\begin{aligned}
\left|x_{t_k}-x_{t_{k-1}}\right|^p-\left|C-x_{t_{k-1}}\right|^p &\leqslant \left(\left|C-x_{t_{k-1}}\right|+\varepsilon\right)^p-\left|C-x_{t_{k-1}}\right|^p \\
&\leqslant \max\left\{\left(\sqrt{\varepsilon}+\varepsilon\right)^p, \theta_0^p\left(\left(1+\sqrt{\varepsilon}\right)^p-1\right)\right\} \\
&=: \mu(\varepsilon),
\end{aligned}$$

where the “max” comes from considering the cases whether $\left|C-x_{t_{k-1}}\right|\leqslant\sqrt{\varepsilon}$ or not. The same inequality holds for $\left|x_{t_{l+1}}-x_{t_l}\right|^p-\left|x_{t_{l+1}}-C\right|^p$. Therefore, we have

$$\sum_{i=1}^{n}\left|x_{t_i}-x_{t_{i-1}}\right|^p \leqslant \left|\overrightarrow{AB}\right|^p-\left(\lambda-2\mu(\varepsilon)-2\varepsilon^p\right)\leqslant\left|\overrightarrow{AB}\right|^p,$$

provided ε is small enough so that

$$2\mu(\varepsilon)+2\varepsilon^p<\lambda.$$

Now by taking supremum over all finite partitions of $[0,1]$, we conclude that

$$\|x\|_p \leqslant \left|\overrightarrow{AB}\right| = \|y\|_p \leqslant \|x\|_p.$$

Therefore, x is also a p-variation minimizer with signature g.

However, it should be pointed out that when $p = 1$, $\mathbf{X}^g$ is always the unique 1-variation (length) minimizer (up to reparametrization in the general sense) with signature $g \in \mathcal{S}_1$. This follows easily from the triangle inequality. See [35] for a discussion as well.

Finally, the necessity of Theorem 3.2.1 is a consequence of Proposition 3.4.3. In fact, given $\mathbf{X} \in WG\Omega_p\left(\mathbb{R}^d\right)$ with trivial signature, define a loop $\alpha : [0,1] \rightarrow \mathcal{S}_p$ by

$$\alpha(t) = S(\mathbf{X})_{0,t},\ t \in [0,1],$$

and define $\psi : \mathcal{S}_p \rightarrow G^{\lfloor p \rfloor}\left(\mathbb{R}^d\right)$ by projection $\pi^{(\lfloor p \rfloor)}$ (ψ is well-defined according to the definition of $\mathcal{S}_p$). It is obvious that $\mathbf{X} = \psi \circ \alpha$. Now it remains to establish the following nontrivial fact.

Lemma 3.4.6. *α is continuous under the tree metric d.*

Proof. Define

$$h(t) = \left\|\mathbf{X}^t\right\|_p^p,\ t \in [0,1],$$

where $\mathbf{X}^t$ is the reduced path associated with $S(\mathbf{X})_{0,t}$. We also use $\mathbb{X}^t$ to denote the corresponding simplified signature path.We first show that h is a continuous function.

Fix $t \in [0,1]$. By the continuity of the p-variation norm, for any $\varepsilon > 0$, there exists some $\delta > 0$ such that

$$s \in (t-\delta, t) \implies \|\mathbf{X}\|_{p;[s,t]}^p < \varepsilon^{2p},$$

where $\|\cdot\|_{p;[s,t]}$ denotes the p-variation of $\mathbf{X}|_{[s,t]}$. Let $\mathbf{X}'$ be the concatenation of $\mathbf{X}^s$ and $\mathbf{X}|_{[s,t]}$, where we parametrize $\mathbf{X}^s$ on $[0,s]$ so that $\mathbf{X}'$ is defined on $[0,t]$. For any finite partition

$$\mathcal{P} :\ 0 = u_0 < u_1 < \cdots < u_k \leqslant s \leqslant u_{k+1} < \cdots < u_{k+l-1} < u_{k+l} = t$$

of $[0,t]$, we have

$$\begin{aligned}
&\sum_{i=1}^{k+l} d\left(\mathbf{X}'_{u_{i-1}},\mathbf{X}'_{u_i}\right)^p \\
= &\left(\sum_{i=1}^{k} d\left(\mathbf{X}'_{u_{i-1}},\mathbf{X}'_{u_i}\right)^p + d\left(\mathbf{X}'_{u_k},\mathbf{X}'_s\right)^p\right) \\
&+\left(d\left(\mathbf{X}'_s,\mathbf{X}'_{u_{k+1}}\right)^p + \sum_{i=k+2}^{k+l} d\left(\mathbf{X}'_{u_{i-1}},\mathbf{X}'_{u_i}\right)^p\right) \\
&+d\left(\mathbf{X}'_{u_k},\mathbf{X}'_{u_{k+1}}\right)^p - d\left(\mathbf{X}'_{u_k},\mathbf{X}'_s\right)^p - d\left(\mathbf{X}'_s,\mathbf{X}'_{u_{k+1}}\right)^p \\
\leqslant\ & \|\mathbf{X}^s\|_p^p + \|\mathbf{X}\|_{p;[s,t]}^p + d\left(\mathbf{X}^s_{u_k},\mathbf{X}_{u_{k+1}}\right)^p - d\left(\mathbf{X}^s_{u_k},\mathbf{X}_s\right)^p \\
\leqslant\ & \|\mathbf{X}^s\|_p^p + \varepsilon^{2p} + \left(d\left(\mathbf{X}^s_{u_k},\mathbf{X}_s\right) + d\left(\mathbf{X}_s,\mathbf{X}_{u_{k+1}}\right)\right)^p - d\left(\mathbf{X}^s_{u_k},\mathbf{X}_s\right)^p.
\end{aligned}$$

It follows that

$$\begin{aligned}
&\sum_{i=1}^{k+l} d\left(\mathbf{X}'_{u_{i-1}},\mathbf{X}'_{u_i}\right)^p \\
&\leqslant \|\mathbf{X}^s\|_p^p + \max\left\{(2^p+1)\,\varepsilon^p, \varepsilon^{2p} + \|\mathbf{X}\|_p^p\left((1+\varepsilon)^p - 1\right)\right\},
\end{aligned}$$

where the "max" comes from considering the cases whether $d\left(\mathbf{X}^s_{u_k},\mathbf{X}_s\right) \leqslant \varepsilon$ or not, and we have also used the fact that $\|\mathbf{X}^s\|_p \leqslant \|\mathbf{X}\|_p$ (see Corollary 3.4.1). Therefore, by taking supremum over all possible partitions of $[0,t]$ and by the definition of $\mathbf{X}^t$, we have

$$\begin{aligned}
&h(t) - h(s) \\
\leqslant\ & \|\mathbf{X}'\|_p^p - \|\mathbf{X}^s\|_p^p \\
\leqslant\ & \max\left\{(2^p+1)\,\varepsilon^p, \varepsilon^{2p} + \|\mathbf{X}\|_p^p\left((1+\varepsilon)^p - 1\right)\right\}.
\end{aligned}$$

By taking $s \uparrow t$ and then $\varepsilon \to 0$, we have

$$\limsup_{s\uparrow t}\left(h(t) - h(s)\right) \leqslant 0.$$

On the other hand, let

$$\mathbf{X}''_u = \begin{cases} \mathbf{X}^t_u, & u \in [0,t]; \\ \mathbf{X}_{2t-u}, & u \in [t,2t]. \end{cases}$$

It follows that

$$h(t) - h(s) \geqslant \left\|\mathbf{X}^t\right\|_p^p - \|\mathbf{X}''\|_{p;[0,2t-s]}^p.$$

Again by the continuity of the p-variation norm, we have

$$\liminf_{s\uparrow t} \left(h(t) - h(s)\right) \geqslant 0.$$

Therefore, we conclude that h is left continuous. By a similar argument we can show that h is also right continuous.

Now we establish the continuity of $\alpha(t) = S(\mathbf{X})_{0,t}$ under the tree metric. Again fix t. From the continuity of height function, apparently we only need to show that

$$\lim_{s\to t} \left\|\mathbf{X}^{\alpha(s)\wedge\alpha(t)}\right\|_p = \left\|\mathbf{X}^t\right\|_p.$$

Note that for each s, there exists a unique $\sigma(s) \in [0,1]$, such that $\mathbb{X}^t_{\sigma(s)} = \alpha(s)\wedge\alpha(t) \in \mathbb{X}^t([0,1])$. We now show that $\sigma(s) \to 1$ as $s \to t$.

In fact, from the construction we know that the concatenation of the reversal of $\mathbb{X}^s$ from $\alpha(s)$ to $\alpha(s)\wedge\alpha(t)$ with $\mathbb{X}^t$ from $\alpha(s)\wedge\alpha(t)$ to $\alpha(t)$ is a continuous simple curve under the L^2-metric, unless $\alpha(s) = \alpha(t)$ which is a trivial case. We denote this path by $\mathbb{Y}^s$ and its projection onto $G^{\lfloor p\rfloor}\left(\mathbb{R}^d\right)$ by $\mathbf{Y}^s$. Then $\mathbf{Y}^s$ has finite p-variation. By uniqueness $\alpha(s)^{-1}\otimes\mathbf{Y}^s$ must be the reduced path associated with $\alpha(s)^{-1}\otimes\alpha(t)$ and $\|\mathbf{Y}^s\|_p \leqslant \|\mathbf{X}\|_{p;[s,t]}$. Therefore, $\|\mathbf{Y}^s\|_p \to 0$ as $s \to t$, and thus $\|\mathbf{X}^t\|_{p;[\sigma(s),1]} \to 0$ as $s \to t$.

For any subsequence $s_n \to t$ such that $\sigma(s_n)$ converges to some $u \in [0,1]$, by the continuity of the p-variation norm we have $\|\mathbf{X}^t\|^p_{p;[\sigma(s_n),1]} \to \|\mathbf{X}^t\|_{p;[u,1]}$. Therefore, $\|\mathbf{X}^t\|^p_{p;[u,1]} = 0$, which implies $u = 1$ since $\mathbb{X}^t$ is the simplified signature path. Consequently, $\sigma(s) \to 1$ as $s \to t$.

Now the result follows from the fact that $\mathbf{X}^{\alpha(s)\wedge\alpha(t)} = \mathbf{X}^t|_{[0,\sigma(s)]}$ and the continuity of p-variation. □

Combining all the previous results, now the proof of Theorem 3.2.1 is complete.

3.5 Final Remarks

As we have seen before, our proof of Theorem 3.2.1 is developed under the setting of real trees. Alternatively, it is also possible to develop the proof by using height functions as in the original work by B.M. Hambly and T. Lyons for continuous paths with bounded total variation. Namely, it is possible to prove the following result without realizing the path on any real tree.

Theorem 3.5.1. *A weakly geometric p-rough path* $\mathbf{X}$ *has trivial signature if and only if there exists some continuous function* $h: [0,1] \to [0,\infty)$ *with* $h(0) = h(1) = 0$, *such that for any* $s, t \in [0,1]$, *if*

$$h(s) = h(t) = \inf_{u \in [s,t]} h(u),$$

then $\mathbf{X}_s = \mathbf{X}_t$.

The proof of this result is of course essentially the same as the proof of Theorem 3.2.1 under the setting of real trees. Instead of developing the technical details again, let us just give a proof of the following result.

Proposition 3.5.1. *A continuous path* $\beta: [0,1] \to V$ *in some topological space* V *is tree-like if and only if there exists some continuous function* $h: [0,1] \to [0,\infty)$ *with* $h(0) = h(1) = 0$, *such that for any* $0 \leqslant s \leqslant t \leqslant 1$, *if*

$$h(s) = h(t) = \inf_{u \in [s,t]} h(u), \tag{3.5.1}$$

then $\beta(s) = \beta(t)$.

Proof. Necessity. Assume that β is a tree-like path in V realized on some real tree (τ, d), so we have $\beta = \psi \circ \alpha$ for some continuous loop α and some map ψ. Choose $\alpha(0)$ to be the root of τ and define the partial order accordingly. Set

$$h(t) = d\left(\alpha(t), \alpha(0)\right), \ t \in [0,1].$$

It is obvious that h is non-negative, continuous and $h(0) = h(1) = 0$. Now assume that $s, t \in [0,1]$ satisfies (3.5.1). If $\alpha(s) \neq \alpha(t)$, then either $\alpha(s), \alpha(t)$ are comparable which contradicts the fact that $h(s) = h(t)$, or $\alpha(s), \alpha(t)$ are not comparable which contradicts the fact that h attains its minimum at s, t since in this case $\alpha(u) = \alpha(s) \wedge \alpha(t)$ at some $u \in [s,t]$ and

$$h(u) = d\left(\alpha(s) \wedge \alpha(t), \alpha(0)\right) < h(s) = h(t).$$

Therefore, $\alpha(s) = \alpha(t)$ and hence $\beta(s) = \beta(t)$.

Sufficiency. Assume that there exists a continuous function h satisfying the conditions in the Proposition. From the discussion in Section 3.2, we know that $([0,1]/_{\sim}, d)$

is a real tree where the equivalence relation is defined by (3.2.1) and the tree metric is defined by (3.2.2), and the canonical projection $\alpha: [0,1] \to [0,1]/_{\sim}$ is continuous. Now define a map $\psi: [0,1]/_{\sim} \to V$ by $\psi([t]) = \beta(t)$, which is well defined since $s \sim t$ if and only if (3.5.1) holds which implies that $\beta(s) = \beta(t)$ by the assumption. From the construction it is obvious that $\beta = \psi \circ \alpha$. Therefore, β is tree-like. □

Remark 3.5.1. It should be pointed out that the notion of tree-like paths in the sense of B.M. Hambly and T. Lyons (see 3.1.1) is equivalent to our formulation. Let $x: [0,1] \to \mathbb{R}^d$ be a continuous path with bounded total variation. We only need to show that if it is tree-like in our sense (we use the height function formulation in Proposition 3.5.1), then it is tree-like in the sense of B.M. Hambly and T. Lyons. In fact, since x is tree-like, it has trivial signature and from the previous discussion it can be realized on the real tree $\mathcal{S}_1$ via the signature path and projection. From the proof of Proposition 3.5.1, we know that the height function of x is defined by

$$\begin{aligned} h(t) &= d\left(S(x)_{0,t}, \mathbf{1}\right) \\ &= L\left(S(x)_{0,t}\right) \\ &= \left\|x^t\right\|_1, \end{aligned}$$

where x^t denotes the reduced path associated with $S(x)_{0,t}$. For given $0 \leqslant s \leqslant t \leqslant 1$, it follows that

$$\begin{aligned} |x_t - x_s| &\leqslant \left\|x^{s,t}\right\|_1 \\ &= L(S(x)_{0,s}) + L(S(x)_{0,t}) - 2L(S(x)_{0,s} \wedge S(x)_{0,t}), \end{aligned}$$

where $x^{s,t}$ denotes the reduced path from x_s to x_t, and we have also used the additivity of the 1-variation norm over adjacent intervals. Moreover, by Proposition 3.2.1 we have

$$S(x)_{0,s} \wedge S(x)_{0,t} \in S(x)([s,t]),$$

and hence

$$\inf_{u \in [s,t]} h(u) \leqslant L(S(x)_{0,s} \wedge S(x)_{0,t}).$$

This implies that

$$|x_t - x_s| \leqslant h(s) + h(t) - 2 \inf_{u \in [s,t]} h(u),$$

and x is tree-like in the sense of B.M. Hambly and T. Lyons.

A immediate consequence of Theorem 3.2.1 is the fact that being a tree-like deformation of each other is an equivalence relation, which is nontrivial from the definition. More precisely, if we introduce the relation "$\sim$" on $WG\Omega_p\left(\mathbb{R}^d\right)$ by

$$\mathbf{X}\sim\mathbf{Y} \text{ iff } \overrightarrow{\mathbf{X}}\sqcup\overleftarrow{\mathbf{Y}} \text{ is tree} - \text{like},$$

where $\overleftarrow{\mathbf{Y}}$ means the reversal of $\mathbf{Y}$, then Theorem 3.2.1 implies that "$\sim$" is an equivalence relation on $WG\Omega_p\left(\mathbb{R}^d\right)$.

On the other hand, consider the RDE

$$d\mathbf{Y}=V(\mathbf{Y})d\mathbf{X}$$

on $\mathbb{R}^e$ with initial condition y_0. When $\mathbf{X}$ has bounded total variation, we have the following Taylor expansion:

$$\mathbf{Y}_t\simeq\sum_{\substack{N\geqslant 0\\ 1\leqslant i_1,\cdots,i_N\leqslant d}}\left(V_{i_1}\cdots V_{i_N}I\right)(y_0)\cdot X_t^{(i_1,\cdots,i_N)},\ t\in[0,1], \tag{3.5.2}$$

where $V=\{V_1,\cdots,V_d\}$ is a given family of vector fields on $\mathbb{R}^e$, I is the identity map on $\mathbb{R}^e$, and

$$X_t^{(i_1,\cdots,i_N)}:=\int_{0<t_1<\cdots<t_N<t}dX_{t_1}^{i_1}\cdots dX_{t_N}^{i_N}$$

are the signature terms associated with $\mathbf{X}$. From the formula 3.5.2 and the shuffle product formula, it is not hard to see that the signature of the solution path is uniquely determined by the signature of $\mathbf{X}$, and by a limiting argument it holds when $\mathbf{X}$ is a weakly geometric rough path. This point can be made mathematically rigorous, at least in the case when the generating vector fields are regular enough. Therefore, Theorem 3.2.1 implies that the tree-like equivalence class of the solution path $\mathbf{Y}$ is determined by the tree-like equivalence class of the driving path $\mathbf{X}$; or equivalently the signature of $\mathbf{Y}$ is determined by the signature of $\mathbf{X}$.

Finally, so far we have seen that the signature map $\mathbf{X}\in WG\Omega_p\left(\mathbb{R}^d\right)\mapsto S(\mathbf{X})_{0,1}$ defines an isomorphism from the space of weakly geometric p-rough paths modulo tree-like equivalence onto its image $\mathcal{S}_p$. To have a fundamental understanding of the signature map, which is an interesting and important mathematical problem on its own, we should at least be able to answer two more questions.

(1) Given a tensor element $g\in T\left(\mathbb{R}^d\right)$, when can it be identified as the signature of some weakly geometric rough path? Or equivalently, can we characterize the signature group $\mathcal{S}_p$ intrinsically?

(2) Given a tensor element g in the signature group $\mathcal{S}_p$, how can we reconstruct the reduced weakly geometric p-rough path whose signature is g? In other words, can we describe the inverse of the signature map (modulo tree-like equivalence) in an analytic way? Another related interesting question is: how does the tensor element g reveal the geometry of the reduced path?

These questions are the main motivation of my ongoing research, and at this stage they are still far from being well understood. We will come back to the second one partially in the probabilistic setting in the next chapter.

Chapter 4

The Uniqueness of Signature Problem in the Probabilistic Setting: Non-Markov Processes

4.1 Introduction

In this chapter, we deviate from the deterministic setting and investigate the probabilistic situation for sample paths of stochastic processes.

As mentioned in the first chapter, in the probabilistic setting Y. Le Jan and Z. Qian [43] proved that with probability one, the Stratonovich signatures of Brownian motion determine the Brownian sample paths. Their strategy, in particular the approximation scheme constructed in the proof, came from the study of cyclic cohomology in algebraic topology. Moreover, their proof relies heavily on the explicit distribution of Brownian motion, the strong Markov property and the potential theory for the Laplace operator. Later on this result was extended to hypoelliptic diffusions by X. Geng and Z. Qian [28], in which the technique of Y. Le Jan and Z. Qian is strengthened but the proof still relies on the strong Markov property and potential theory in a crucial way. A similar result for Chordal SLE_κ curves with $\kappa \leqslant 4$ was obtained by H. Boedihardjo, H. Ni and Z. Qian [6], as a direct application of the deterministic result for planar simple curves together with the sample path properties for SLE curves.

It should be pointed out that the result of Le Jan and Qian is stronger than the general deterministic results we have seen so far, as it not only gives the injectivity but also gives an explicit way of how a sample path can be constructed from its signature outside a null set in the path space. In the deterministic setting, such reconstruction was studied by T. Lyons and W. Xu [48] for C^1-paths via symmetrization, and also implicitly contained in the work in Chapter 2 (see Remark [3]) via Fourier transform. A general inversion scheme for the signature of weakly geometric rough paths remains a significant open problem in rough path theory.

The main purpose of this chapter is to further simplify and strengthen the method of Y. Le Jan and Z. Qian to include a class of non-Markov processes. In particular, we establish the almost-sure uniqueness of signature (up to reparametrization) for a class of Gaussian processes including fractional Brownian motion with Hurst parameter $H > 1/4$, the Ornstein-Uhlenbeck process and the Brownian bridge. More importantly, our technique also yields an explicit inversion scheme for the signature of sample paths.The fundamental difficulty in exploiting the idea of Y. Le Jan and Z. Qian of course lies in the unavailability of those probabilistic and analytic tools arising from the strong Markov property and potential theory which are both crucial in their proof. The key of getting around this difficulty is to find methods which enable us to analyze pathwisely.

The well-definedness of the signature when the sample paths of the process have finite p-variation for $p \geqslant 1$ are well studied in the probability literature (see for example [26] for a detailed presentation). For instance, it was shown by L. Coutin and Z. Qian [14] that with probability one, the sample paths of fractional Brownian motion with Hurst parameter $H > \frac{1}{4}$ can be lifted canonically as geometric rough paths, while it is believed that no such canonical lifting exists for $H \leqslant \frac{1}{4}$. More generally, L. Coutin and Z. Qian [13] showed that under certain conditions on the decorrelation of the increments of a Gaussian process, with probability one the lifting of the dyadic piecewise linear interpolation of the Gaussian sample paths in $G\Omega_p\left(\mathbb{R}^d\right)$ is a Cauchy sequence under the p-variation metric. In [26], P. Friz and N. Victoir extended this result to a larger class of Gaussian processes under certain regularity condition on the covariance function. Moreover, they showed that the lifting of any sequence of piecewise linear interpolations of the Gaussian sample paths in $G\Omega_p(\mathbb{R}^d)$ converges to the same limit. This limit is usually known as the canonical lifting of the Gaussian process in $G\Omega_p\left(\mathbb{R}^d\right)$.

In establishing our main result, we state explicitly under what conditions on the law of the process the almost-sure uniqueness of signature holds. We hope that this provides a general framework for solving the almost-sure uniqueness of signature problem for other interesting processes. Note that our result is not a direct corollary of the result in Chapter 3, since it is highly nontrivial to prove the existence of a null set outside which no two paths can be tree-like deformations of each other.

4.2 Main Results

In this section we state the main results of this chapter and illustrate the idea of the proofs.

Let $X = \{X_t : t \in [0,1]\}$ be a d-dimensional continuous stochastic process starting at the origin, where $d \geqslant 2$. We always assume that X is realized on the path space $(W, \mathcal{B}(W), \mathbb{P})$,where W is the space of $\mathbb{R}^d$-valued continuous paths over $[0,1]$ starting at the origin, $\mathcal{B}(W)$ is the Borel σ-algebra over W, and $\mathbb{P}$ is the law of X.

In the rest of this chapter, we make the following assumptions on the law $\mathbb{P}$.

Assumption (A): There exists a $\mathbb{P}$-null set $\mathcal{N}_0$ and a map

$$S : W \backslash \mathcal{N}_0 \to C\left(\triangle; T\left(\mathbb{R}^d\right)\right),$$

such that for each $x \in W \backslash \mathcal{N}_0$ and $(s,t) \in \Delta$, $\pi_1\left(S(x)_{s,t}\right) = x_t - x_s$ and $S(x)$ is the multiplicative extension of some geometric rough path $\mathbf{X}$ in terms of Lyons' extension theorem. We call such a map S a $\mathbb{P}$-*almost sure lifting.* The integral against x is then defined as integrating against the geometric rough path $\mathbf{X}$.

Assumption (B): For any $0 < t < 1$, the law of x_t is absolutely continuous with respect to the Lebesgue measure.

Assumption (C): For any open cube $H \subset \mathbb{R}^d$, there exists a differential one form (i.e. a C^∞-one form) $\phi = \sum_{i=1}^d \phi_i dx^i$ supported on the closure of H, such that for any $0 \leqslant s < t \leqslant 1$, if we let

$$A^H_{s,t} = \{x \in W : \text{ there exists some } u \in (s,t) \text{ such that } x_u \in H\}, \tag{4.2.1}$$

then

$$\mathbb{P}\left(\left\{x \in W : \int_s^t \phi(dx_u) = 0\right\} \bigcap A^H_{s,t}\right) = 0.$$

Here $\int_s^t \phi(dx_u) = \sum_{i=1}^d \int_s^t \phi_i(x_u) dx^i_u$ is defined in the sense of rough paths according to Assumption (A).

Remark 4.2.1. As we have mentioned before, Assumption (A) is quite natural for a large class of stochastic processes. Assumption (B) is also verified for most of these processes, e.g. hypoelliptic diffusions, Gaussian processes, solutions to hypoelliptic rough differential equations driven by Gaussian processes. These examples are well studied in [26]. Assumption (C) suggests a certain kind of non-degeneracy for sample paths of the process, which is essential for the recovery of a path from its signature in our setting. By a closer look at Assumption (C), it actually excludes the possibility of the sample paths having tree-like pieces. Therefore, with probability one the sample paths are already "reduced" paths in the tree-like equivalence classes and it is natural to expect an inversion scheme for the signature in our setting. This is the main goal of this chapter.

In the last section of this chapter, as a fundamental example we show that these assumptions are all verified for a class of Gaussian processes including fractional Brownian motion with Hurst parameter $H > 1/4$, the Ornstein-Uhlenbeck process and the Brownian bridge.

Now we are in a position to state our main results. Let $\mathcal{R}$ be the group of reparametrizations from $[0,1]$ to itself.

Theorem 4.2.1. *Assume that the law $\mathbb{P}$ of the stochastic process satisfies Assumption (A), (B) and (C). Let S be the $\mathbb{P}$-almost sure lifting as in Assumption (A). Then there exists a $\mathbb{P}$-null set $\mathcal{N}$, such that for any $x, x' \in \mathcal{N}^c$, if $S(x)_{0,1} = S(x')_{0,1}$, then there exists some $\sigma \in \mathcal{R}$, such that*

$$x_t = x'_{\sigma(t)}, \ \forall t \in [0,1].$$

As a fundamental example, we prove the following result for a class of Gaussian processes satisfying conditions to be specified later on in the final section.

Theorem 4.2.2. *Let $\mathbb{P}$ be the law of a Gaussian process satisfying conditions specified in Section 4.5. Then $\mathbb{P}$ satisfies Assumption (A), (B), (C). In particular, the result holds for fractional Brownian motion with Hurst parameter $H > 1/4$, the Ornstein-Uhlenbeck process and the Brownian bridge.*

Before going into the mathematical proofs, we first describe the strategy informally. The approximation scheme we develop is an adaptation from the work of Y. Le Jan and Z. Qian [43]. However, the main difficulties are in the development of each step in the non-Markov setting, which will be clear in the detailed proofs.

Step One. Prove that if two paths have the same signature, then the iterated integrals of the paths along any finite sequence of differential one forms are the same. Following [43], these iterated integrals along one forms are called *extended signatures.*

Step Two. Decompose the Euclidean space $\mathbb{R}^d$ into disjoint identical open cubes with small tunnels between them. For each such cube, we define a differential one form supported on the closure of the cube according to Assumption (C).

Step Three. Show that, for each path x outside a $\mathbb{P}$-null set, the ordered sequence of cubes visited by x corresponds to the unique maximal sequence of differential one forms along which the extended signature of x is nonzero. This together with step one allows us to recover the ordered sequence of cubes visited by x from its signature.

Step Four. Construct a polygonal approximation of x by joining the centers of cubes visited by x in order. This polygonal path is parametrized so that it is at the center of the cube at the time when the cube is first visited by x. Show that with probability one, as the size of cubes tends to zero, the polygonal path converges to the original path x under the uniform topology.

Step Five. Since the signature is invariant under reparametrizations of the path, it is not possible to recover the exact visit times of the cubes. If two paths have the same signature, then the corresponding polygonal paths constructed in (3) are only equal up to reparametrization. Therefore, we need to introduce a variant of the Fréchet distance on W measuring the distance of two paths modulo parametrization. We should also prove that outside a $\mathbb{P}$-null set this is indeed a metric. It then follow from step four that if two paths x and x' have the same signature, their corresponding approximation paths converge to the same limit under this metric, which implies that x and x' are equal up to reparametrization.

For the Gaussian case, Assumption (A) is verified from [26] and Assumption (B) is trivial by definition. By using the Malliavin calculus, for each open cube H we explicitly construct a differential one form ϕ supported on $\overline{H}$ such that the functional $x \to \int_s^t \phi(dx_u)$ has a density conditioned on the set $A_{s,t}^H$. This certainly verifies Assumption (C).

4.3 Signature Determines Extended Signatures

Starting from this section, we develop the detailed proofs of our main results.

As the first step, we prove that if two sample paths as geometric rough paths have the same signature, then they have the same extended signature. Note that the signatures and extended signatures are well-defined $\mathbb{P}$-almost surely according to Assumption (A).

From now on, for a geometric rough path $\mathbf{X}$ and a finite sequence $(\phi^1, \cdots, \phi^n)$ of differential one forms $\phi^1, \ldots, \phi^n$, we use $[\phi^1, \cdots, \phi^n]_{0,1}(x)$ to denote the iterated path integral $\int_0^1 \cdots \int_0^{s_2} \phi^1(d\mathbf{X}_{s_1}) \cdots \phi^n(d\mathbf{X}_{s_n})$, where $x_\cdot := \pi_1(S(\mathbf{X})_{0,\cdot})$ is the first level path of $\mathbf{X}$. A simple way of understanding this integral is by

$$\int_0^1 \cdots \int_0^{s_2} \phi^1 (d\mathbf{X}_{s_1}) \dots \phi^n (d\mathbf{X}_{s_n}) = \lim_{k\to\infty} \int_0^1 \cdots \int_0^{s_2} \phi^1 \left(dx_{s_1}^{(k)}\right) \cdots \phi^n \left(dx_{s_n}^{(k)}\right),$$

where by the definition of geometric rough paths $x^{(k)}$ is a sequence of paths with bounded total variation whose lifting converges to $\mathbf{X}$ under the p-variation metric. Sometimes we also use the notation $\int_0^1 \dots \int_0^{s_2} \phi^1 (dx_{s_1}) \dots \phi^n (dx_{s_n})$ to denote the path integral. Note that the ordering of $(\phi^1, \cdots, \phi^n)$ is non-commutative in this notation.

Now we have the following result. The proof is almost identical to the one of Lemma 3.4.2.

Proposition 4.3.1. *Given $p \geqslant 1$, let $\mathbf{X}, \mathbf{X}'$ be two geometric p-rough paths. Suppose that $\phi^1, \dots, \phi^n$ are C^α-one forms for $\alpha > p$. If $S(\mathbf{X})_{0,1} = S(\mathbf{X}')_{0,1}$, then*

$$\left[\phi^1, \dots, \phi^n\right]_{0,1}(x) = \left[\phi^1, \dots, \phi^n\right]_{0,1}(x'), \tag{4.3.1}$$

where x and x' are the first level paths of $\mathbf{X}$ and $\mathbf{X}'$ respectively.

Proof. We write ϕ^i as $\phi^i = \sum_{j=1}^d \phi_j^i(x)\, dx^j$. Let K be a compact neighborhood of $x([0,1]) \bigcup x'([0,1])$. As in the proof of Lemma 3.4.2, according to [2], Theorem 1, for each $\alpha > 0$ and each j, there exists a polynomial sequence $\phi_j^{i(m)}$ such that

$$\left\| \phi_j^{i(m)} - \phi_j^i \right\|_{C_K^\alpha} \to 0$$

as $m \to \infty$. Let $\phi^{i(m)}(x) = \sum_{j=1}^d \phi_j^{i(m)}(x)\, dx^j$. We know from the shuffle product formula (see also [43], p. 4) that (4.3.1) holds for the finite sequence $\left(\phi^{1(m)}, \cdots, \phi^{n(m)}\right)$ for all m. Now the result follows from the continuity of the integration map

$$(\phi^1, \cdots, \phi^n) \mapsto \left[\phi^1, \cdots, \phi^n\right]_{0,1}(x), \left[\phi^1, \cdots, \phi^n\right]_{0,1}(x')$$

under the Lip^α-norm when $\alpha > p$ together with the fact that the Lip^α-norm is controlled by the C_K^α-norm. $\square$

4.4 The Strengthened Le Jan-Qian Approximation Scheme and the Uniqueness of Signature

Now fix $\varepsilon, \delta > 0$ with $\delta << \varepsilon$.

For any integer point $z = (z^1, z^2, \ldots, z^d) \in \mathbb{Z}^d$, let $H_z^{\varepsilon,\delta}$ be the open cube centered at εz with edges of length $\varepsilon - \delta$. In other words,

$$H_z^{\varepsilon,\delta} = \left\{ x \in \mathbb{R}^d : \left| x^i - \varepsilon z^i \right| < \frac{\varepsilon - \delta}{2}, \ \forall i = 1, \cdots, d \right\}.$$

Geometrically, the space $\mathbb{R}^d$ is divided into disjoint identical open cubes and small closed tunnels.

For any $x \in W$ and $k \geqslant 1$, define recursively

$$\tau_k^{\varepsilon,\delta} = \inf \left\{ t \in \left[\tau_{k-1}^{\varepsilon,\delta}, 1 \right] : \ x_t \in \bigcup_{z \neq \boldsymbol{m}_{k-1}^{\varepsilon,\delta}} H_z^{\varepsilon,\delta} \right\},$$

and $\boldsymbol{m}_k^{\varepsilon,\delta}$ to be the integer point $z \in \mathbb{Z}^d$ such that

$$x_{\tau_k^{\varepsilon,\delta}} \in H_z^{\varepsilon,\delta},$$

where $\tau_0^{\varepsilon,\delta} = 0, \ \boldsymbol{m}_0^{\varepsilon,\delta} = 0 \in \mathbb{Z}^d$. Let

$$N^{\varepsilon,\delta} = \sup \left\{ k \geqslant 1 : \ \tau_k^{\varepsilon,\delta} < 1 \right\},$$

where $\sup \emptyset := 0$. The sequence $\left\{ \tau_k^{\varepsilon,\delta} \right\}$ records the successive visit times of the open cubes by the path, the sequence $\left\{ \boldsymbol{m}_k^{\varepsilon,\delta} \right\}$ records the cubes visited in order, and $N^{\varepsilon,\delta}$ records the total number of cubes visited. Note that revisiting the same cube after visiting some other cubes counts, but revisiting before visiting any other cubes does not count. By continuity and compactness, it is easy to see that for any $x \in W$, $0 \leqslant N^{\varepsilon,\delta} < \infty$.

Here and thereafter, for notational simplicity we drop the dependence on x for these random variables on W.

Remark 4.4.1. It is important to use the open cubes instead of the closed ones, as we are only interested in the case when a path x travels through the interior of a cube. Note that these $\tau_k^{\varepsilon,\delta}$ are not stopping times with respect to the natural filtration.

For each cube $H_z^{\varepsilon,\delta}$, let $\phi_z^{\varepsilon,\delta}$ be the differential one form given in Assumption (C). In particular, $\phi_z^{\varepsilon,\delta}$ is supported on the closure of $H_z^{\varepsilon,\delta}$, and $\phi_z^{\varepsilon,\delta} = 0$ on $\partial H_z^{\varepsilon,\delta}$.

4.4.1 Recovery of Cubes Visited in Order by Using the Extended Signature

Let $\mathcal{W}_m$ ($m \geqslant 0$) be the set of words $(z_0 = 0, z_1, \cdots, z_m)$ with $z_i \neq z_{i+1}$, $z_i \in \mathbb{Z}^d$, and let $\mathcal{W} = \bigcup_{m \geqslant 0} \mathcal{W}_m$. Elements of $\mathcal{W}$ are called *admissible words.*

For $w = (z_0, z_1, \cdots, z_m) \in \mathcal{W}$, define

$$E_w^{\varepsilon,\delta} = \left\{ x \in W : \ N^{\varepsilon,\delta} = m, \ \boldsymbol{m}_k^{\varepsilon,\delta} = z_k, \ k = 0, \cdots, m \right\}.$$

It follows that W can be written as the disjoint union $W = \bigcup_{w \in \mathcal{W}} E_w^{\varepsilon,\delta}$.

Now we have the following result.

Lemma 4.4.1. *For any $m \geqslant 0$, if $w = (z_0 = 0, \cdots, z_m) \in \mathcal{W}_m$ and $x \in E_w^{\varepsilon,\delta}$, then*

(1)

$$\left[\phi_{z_0}^{\varepsilon,\delta}, \cdots, \phi_{z_m}^{\varepsilon,\delta}\right]_{0,1}(x) = \prod_{i=1}^{m+1} \int_{\tau_{i-1}^{\varepsilon,\delta}}^{\tau_i^{\varepsilon,\delta}} \phi_{z_{i-1}}^{\varepsilon,\delta}(dx_t), \tag{4.4.1}$$

where $\tau_{m+1}^{\varepsilon,\delta} = 1$ by definition since $x \in E_w^{\varepsilon,\delta}$.

(2) For any $w' = (z_0, z_1', \cdots, z_n') \in \mathcal{W}_n$ with $n > m$,

$$\left[\phi_{z_0}^{\varepsilon,\delta}, \cdots, \phi_{z_n'}^{\varepsilon,\delta}\right]_{0,1}(x) = 0.$$

(3) For any $w' = (z_0, z_1', \cdots, z_m')$ with $w' \neq w$,

$$\left[\phi_{z_0}^{\varepsilon,\delta}, \cdots, \phi_{z_m'}^{\varepsilon,\delta}\right]_{0,1}(x) = 0.$$

Proof. We prove this result by induction on m.

If $m = 0$, assume that $x \in E_{(z_0)}^{\varepsilon,\delta}$. Then (1) and (3) are trivial. To see (2), let $w' = (z_0, z_1', \cdots, z_n') \in \mathcal{W}_n$ with $n > 0$. Since w' is an admissible word, there is some $0 < k \leqslant n$ such that x does not visit the open cube $H_{z_k'}^{\varepsilon,\delta}$ and the corresponding extended signature is zero by definition (here we have implicitly used the definition of extended signatures of geometric rough paths and the joint continuity of the integration map with respect to the one forms and the driving path). If $m = 1$, assume that $w = (z_0, z_1) \in \mathcal{W}_1$ and $x \in E_w^{\varepsilon,\delta}$. Then (3) follows by the same argument as before. To see

(1), first we have

$$\begin{aligned}\left[\phi_{z_0}^{\varepsilon,\delta},\phi_{z_1}^{\varepsilon,\delta}\right]_{0,1}(x) &= \int_0^1 \left[\phi_{z_0}^{\varepsilon,\delta}\right]_{0,t}(x)\phi_{z_1}^{\varepsilon,\delta}(dx_t)\\ &= \int_{\tau_1^{\varepsilon,\delta}}^1 \left[\phi_{z_0}^{\varepsilon,\delta}\right]_{0,t}(x)\phi_{z_1}^{\varepsilon,\delta}(dx_t),\end{aligned}$$

since $\phi_{z_1}^{\varepsilon,\delta}$ is supported in $H_{z_1}^{\varepsilon,\delta}$. Moreover, if $\tau_1^{\varepsilon,\delta}\leqslant t\leqslant 1$, then

$$\left[\phi_{z_0}^{\varepsilon,\delta}\right]_{0,t}(x)=\left[\phi_{z_0}^{\varepsilon,\delta}\right]_{0,\tau_1^{\varepsilon,\delta}}(x),$$

since $\phi_{z_0}^{\varepsilon,\delta}$ is supported in $H_{z_0}^{\varepsilon,\delta}$. Therefore,

$$\left[\phi_{z_0}^{\varepsilon,\delta},\phi_{z_1}^{\varepsilon,\delta}\right]_{0,1}(x)=\left(\int_0^{\tau_1^{\varepsilon,\delta}}\phi_{z_0}^{\varepsilon,\delta}(dx_t)\right)\left(\int_{\tau_1^{\varepsilon,\delta}}^1\phi_{z_1}^{\varepsilon,\delta}(dx_t)\right)$$

and (1) follows. If $w'=(z_0,z_1',\cdots,z_n')\in\mathcal{W}_n$ with $n>1$, there are two case. The first case is that there is some $0<k\leqslant n$ such that z_k' is different from z_0 and z_1. In this case (2) follows by the same argument as before. The second case is

$$w'=(z_0,z_1,z_0,z_1,\cdots,z_n'),$$

where $n>1$ and z_n' is either z_0 or z_1. If $z_n'=z_0$, then

$$\left[\phi_{z_0}^{\varepsilon,\delta},\cdots,\phi_{z_n'}^{\varepsilon,\delta}\right]_{0,1}(x)=\int_0^{\tau_1^{\varepsilon,\delta}}\left[\phi_{z_0}^{\varepsilon,\delta},\cdots,\phi_{z_{n-1}'=z_1}^{\varepsilon,\delta}\right]_{0,t}(x)\phi_{z_0}(dx_t).$$

But during $\left[0,\tau_1^{\varepsilon,\delta}\right]$ the path x never visits the interior of $H_{z_1}^{\varepsilon,\delta}$, so the integral on the R.H.S. is zero and hence the extended signature corresponding to w' is zero. If $z_n'=z_1$,

$$\left[\phi_{z_0}^{\varepsilon,\delta},\cdots,\phi_{z_n'}^{\varepsilon,\delta}\right]_{0,1}(x)=\int_{\tau_1^{\varepsilon,\delta}}^1\left[\phi_{z_0}^{\varepsilon,\delta},\cdots,\phi_{z_{n-1}'=z_0}^{\varepsilon,\delta}\right]_{0,t}(x)\phi_{z_1}^{\varepsilon,\delta}(dx_t).$$

For $\tau_1^{\varepsilon,\delta} \leqslant t \leqslant 1$, we have

$$\begin{aligned}
&\left[\phi_{z_0}^{\varepsilon,\delta}, \cdots, \phi_{z'_{n-1}=z_0}^{\varepsilon,\delta}\right]_{0,t}(x) \\
&= \left[\phi_{z_0}^{\varepsilon,\delta}, \cdots, \phi_{z'_{n-1}=z_0}^{\varepsilon,\delta}\right]_{0,\tau_1^{\varepsilon,\delta}}(x) + \int_{\tau_1^{\varepsilon,\delta}}^1 \left[\phi_{z_0}^{\varepsilon,\delta}, \cdots, \phi_{z'_{n-2}=z_1}^{\varepsilon,\delta}\right]_{0,t}(x)\phi_{z_0}^{\varepsilon,\delta}(dx_t) \\
&= \left[\phi_{z_0}^{\varepsilon,\delta}, \cdots, \phi_{z'_{n-1}=z_0}^{\varepsilon,\delta}\right]_{0,\tau_1^{\varepsilon,\delta}}(x).
\end{aligned}$$

But during $\left[0, \tau_1^{\varepsilon,\delta}\right]$ the path x does not visit the interior of $H_{z_1}^{\varepsilon,\delta}$ and the last term contains the differential one form $\phi_{z_1}^{\varepsilon,\delta}$, thus it is zero and $\left[\phi_{z_0}^{\varepsilon,\delta}, \cdots, \phi_{z'_n}^{\varepsilon,\delta}\right]_{0,1}(x) = 0$. Therefore (2) again follows.

Now assume that the claim is true for all non-negative integer less than m, and we show that it is true for m. Let $w = (z_0, \cdots, z_m) \in \mathcal{W}_m$ and $x \in E_w^{\varepsilon,\delta}$.

We first show (1). In fact,

$$\begin{aligned}
\left[\phi_{z_0}^{\varepsilon,\delta}, \cdots, \phi_{z_m}^{\varepsilon,\delta}\right]_{0,1}(x) &= \int_0^{\tau_m^{\varepsilon,\delta}} \left[\phi_{z_0}^{\varepsilon,\delta}, \cdots, \phi_{z_{m-1}}^{\varepsilon,\delta}\right]_{0,t}(x)\phi_{z_m}^{\varepsilon,\delta}(dx_t) \\
&\quad + \int_{\tau_m^{\varepsilon,\delta}}^1 \left[\phi_{z_0}^{\varepsilon,\delta}, \cdots, \phi_{z_{m-1}}^{\varepsilon,\delta}\right]_{0,t}(x)\phi_{z_m}^{\varepsilon,\delta}(dx_t) \\
&= \int_0^{\tau_m^{\varepsilon,\delta}} \left[\phi_{z_0}^{\varepsilon,\delta}, \cdots, \phi_{z_{m-1}}^{\varepsilon,\delta}\right]_{0,t}(x)\phi_{z_m}^{\varepsilon,\delta}(dx_t) \\
&\quad + \left[\phi_{z_0}^{\varepsilon,\delta}, \cdots, \phi_{z_{m-1}}^{\varepsilon,\delta}\right]_{0,\tau_m^{\varepsilon,\delta}}(x) \int_{\tau_m^{\varepsilon,\delta}}^1 \phi_{z_m}^{\varepsilon,\delta}(dx_t),
\end{aligned}$$

where the last equality comes from the fact that

$$\left[\phi_{z_0}^{\varepsilon,\delta}, \cdots, \phi_{z_{m-1}}^{\varepsilon,\delta}\right]_{0,t}(x) = \left[\phi_{z_0}^{\varepsilon,\delta}, \cdots, \phi_{z_{m-1}}^{\varepsilon,\delta}\right]_{0,\tau_m^{\varepsilon,\delta}}(x), \ \forall t \in \left[\tau_m^{\varepsilon,\delta}, 1\right],$$

since $z_{m-1} \neq z_m$ and hence during $\left[\tau_m^{\varepsilon,\delta}, 1\right]$ the path does not visit the interior of $H_{z_{m-1}}^{\varepsilon,\delta}$. Now we want to use the induction hypothesis (1) on the term $\left[\phi_{z_0}^{\varepsilon,\delta}, \cdots, \phi_{z_{m-1}}^{\varepsilon,\delta}\right]_{0,\tau_m^{\varepsilon,\delta}}(x)$. To this end, let $\widetilde{x}$ be a path in W such that $\widetilde{x} = x$ on $\left[0, \tau_m^{\varepsilon,\delta}\right]$ and $\widetilde{x}$ stays inside the tunnel on $\left[\tau_m^{\varepsilon,\delta}, 1\right]$. It follows that $\widetilde{x} \in E_{\widetilde{w}}^{\varepsilon,\delta}$ where $\widetilde{w} = (z_0, \cdots, z_{m-1}) \in \mathcal{W}_{m-1}$, and

$$\left[\phi_{z_0}^{\varepsilon,\delta}, \cdots, \phi_{z_{m-1}}^{\varepsilon,\delta}\right]_{0,\tau_m^{\varepsilon,\delta}}(x) = \left[\phi_{z_0}^{\varepsilon,\delta}, \cdots, \phi_{z_{m-1}}^{\varepsilon,\delta}\right]_{0,1}(\widetilde{x}).$$

Therefore, by the induction hypothesis (1) and the definition of $\widetilde{x}$ we have

$$\begin{aligned}\left[\phi_{z_0}^{\varepsilon,\delta},\cdots,\phi_{z_{m-1}}^{\varepsilon,\delta}\right]_{0,1}(\widetilde{x}) &= (\prod_{i=1}^{m-1}\int_{\tau_{i-1}^{\varepsilon,\delta}}^{\tau_i^{\varepsilon,\delta}}\phi_{z_{i-1}}^{\varepsilon,\delta}(d\widetilde{x}_t))(\int_{\tau_{m-1}^{\varepsilon,\delta}}^{1}\phi_{z_{m-1}}^{\varepsilon,\delta}(d\widetilde{x}_t))\\ &= (\prod_{i=1}^{m-1}\int_{\tau_{i-1}^{\varepsilon,\delta}}^{\tau_i^{\varepsilon,\delta}}\phi_{z_{i-1}}^{\varepsilon,\delta}(dx_t))(\int_{\tau_{m-1}^{\varepsilon,\delta}}^{\tau_m^{\varepsilon,\delta}}\phi_{z_{m-1}}^{\varepsilon,\delta}(dx_t)).\end{aligned}$$

Consequently (1) follows once we show that

$$\int_0^{\tau_m^{\varepsilon,\delta}}\left[\phi_{z_0}^{\varepsilon,\delta},\cdots,\phi_{z_{m-1}}^{\varepsilon,\delta}\right]_{0,t}(x)\phi_{z_m}^{\varepsilon,\delta}(dx_t)=0.$$

But this is an easy consequence of the fact that

$$\int_0^{\tau_m^{\varepsilon,\delta}}\left[\phi_{z_0}^{\varepsilon,\delta},\cdots,\phi_{z_{m-1}}^{\varepsilon,\delta}\right]_{0,t}(x)\phi_{z_m}^{\varepsilon,\delta}(dx_t)=\left[\phi_{z_0}^{\varepsilon,\delta},\cdots,\phi_{z_m}^{\varepsilon,\delta}\right]_{0,1}(\widetilde{x})$$

and the induction hypothesis (2).

Now we show (2). Let $w'=(z_0,z_1',\cdots,z_n')\in\mathcal{W}_n$ with $n>m$. As before, the case when there exists some $0<k\leqslant n$ such that $z_k'\notin\{z_0,\cdots,z_m\}$ is trivial. Otherwise, write

$$\left[\phi_{z_0}^{\varepsilon,\delta},\cdots,\phi_{z_n'}^{\varepsilon,\delta}\right]_{0,1}(x)=\sum_i\int_{\tau_{i-1}^{\varepsilon,\delta}}^{\tau_i^{\varepsilon,\delta}}\left[\phi_{z_0}^{\varepsilon,\delta},\cdots,\phi_{z_{n-1}'}^{\varepsilon,\delta}\right]_{0,t}(x)\phi_{z_n'}^{\varepsilon,\delta}(dx_t),\tag{4.4.2}$$

where the sum is over those $i\leqslant m+1$ such that $z_{i-1}=z_n'$. Since $z_{n-1}'\neq z_n'$, for each such i we have

$$\begin{aligned}&\int_{\tau_{i-1}^{\varepsilon,\delta}}^{\tau_i^{\varepsilon,\delta}}\left[\phi_{z_0}^{\varepsilon,\delta},\cdots,\phi_{z_{n-1}'}^{\varepsilon,\delta}\right]_{0,t}(x)\phi_{z_n'}^{\varepsilon,\delta}(dx_t)\\ =&\left[\phi_{z_0}^{\varepsilon,\delta},\cdots,\phi_{z_{n-1}'}^{\varepsilon,\delta}\right]_{0,\tau_{i-1}^{\varepsilon,\delta}}(x)\int_{\tau_{i-1}^{\varepsilon,\delta}}^{\tau_i^{\varepsilon,\delta}}\phi_{z_n'}^{\varepsilon,\delta}(dx_t).\end{aligned}$$

Define a new path $\widetilde{x}\in W$ such that $\widetilde{x}=x$ on $\left[0,\tau_{i-1}^{\varepsilon,\delta}\right]$ and $\widetilde{x}$ stays inside the tunnel on $\left[\tau_{i-1}^{\varepsilon,\delta},1\right]$. Then $\widetilde{x}\in E_{\widetilde{w}}^{\varepsilon,\delta}$ with $\widetilde{w}=(z_0,\cdots,z_{i-2})$. Since during $\left[\tau_{i-1}^{\varepsilon,\delta},1\right]$ the path $\widetilde{x}$ does not visit the interior of $H_{z_n'}^{\varepsilon,\delta}$, we have

$$\left[\phi_{z_0}^{\varepsilon,\delta},\cdots,\phi_{z_{n-1}'}^{\varepsilon,\delta}\right]_{0,\tau_{i-1}^{\varepsilon,\delta}}(x)=\left[\phi_{z_0}^{\varepsilon,\delta},\cdots,\phi_{z_{n-1}'}^{\varepsilon,\delta}\right]_{0,1}(\widetilde{x}).$$

Now observe that $i-2 < m \leqslant n-1$, and so by the induction hypothesis (2) we know that

$$\left[\phi_{z_0}^{\varepsilon,\delta}, \cdots, \phi_{z'_{n-1}}^{\varepsilon,\delta}\right]_{0,1}(\widetilde{x}) = 0.$$

Therefore, each term in the R.H.S. is zero and (2) follows.

Finally we show (3). Let $w' = (z_0, z'_1, \cdots, z'_m) \in \mathcal{W}_m$ with $w' \neq w$. If $z'_m = z_m$, then

$$\begin{aligned}
&\left[\phi_{z_0}^{\varepsilon,\delta}, \cdots, \phi_{z'_m}^{\varepsilon,\delta}\right]_{0,1}(x) \\
&= \left[\phi_{z_0}^{\varepsilon,\delta}, \cdots, \phi_{z'_m}^{\varepsilon,\delta}\right]_{0,\tau_m^{\varepsilon,\delta}}(x) + \left[\phi_{z_0}^{\varepsilon,\delta}, \cdots, \phi_{z'_{m-1}}^{\varepsilon,\delta}\right]_{0,\tau_m^{\varepsilon,\delta}}(x) \int_{\tau_m^{\varepsilon,\delta}}^1 \phi_{z_m}^{\varepsilon,\delta}(dx_t).
\end{aligned}$$

Define $\widetilde{x} \in W$ by $\widetilde{x} = x$ on $\left[0, \tau_m^{\varepsilon,\delta}\right]$ and staying inside the tunnel on $\left[\tau_m^{\varepsilon,\delta}, 1\right]$. It follows from the induction hypothesis (2) that

$$\left[\phi_{z_0}^{\varepsilon,\delta}, \cdots, \phi_{z'_m}^{\varepsilon,\delta}\right]_{0,\tau_m^{\varepsilon,\delta}}(x) = \left[\phi_{z_0}^{\varepsilon,\delta}, \cdots, \phi_{z'_m}^{\varepsilon,\delta}\right]_{0,1}(\widetilde{x}) = 0.$$

Moreover, in this case we know that $(z_0, \cdots, z'_{m-1}) \neq (z_0, \cdots, z_{m-1})$. Therefore, by induction hypothesis (3) we have

$$\left[\phi_{z_0}^{\varepsilon,\delta}, \cdots, \phi_{z'_{m-1}}^{\varepsilon,\delta}\right]_{0,\tau_m^{\varepsilon,\delta}}(x) = \left[\phi_{z_0}^{\varepsilon,\delta}, \cdots, \phi_{z'_{m-1}}^{\varepsilon,\delta}\right]_{0,1}(\widetilde{x}) = 0.$$

Consequently (3) follows. For the case $z'_m \neq z_m$ and there exists some $i \leqslant m+1$ with $z_{i-1} = z'_m$ (otherwise it is trivial), we know that i must be strictly less than $m-1$. By writing $\left[\phi_{z_0}^{\varepsilon,\delta}, \cdots, \phi_{z'_m}^{\varepsilon,\delta}\right]_{0,1}(x)$ as a sum of the form (4.4.2), the result (3) follows easily from the induction hypothesis (2) by a similar argument.

Now the proof is complete. □

Define a map $M^{\varepsilon,\delta}: W \to \mathbb{Z}_+$ by sending a path $x \in W$ to

$$\sup\left\{m \geqslant 0 : \exists w = (z_0, z_1, \cdots, z_m) \in \mathcal{W}_m \text{ s.t. } \left[\phi_{z_0}^{\varepsilon,\delta}, \phi_{z_1}^{\varepsilon,\delta}, \cdots, \phi_{z_m}^{\varepsilon,\delta}\right]_{0,1}(x) \neq 0\right\}.$$

Note that by Lemma 4.4.1, $M^{\varepsilon,\delta} \leqslant N^{\varepsilon,\delta}$ for $\mathbb{P}$-almost surely. Moreover, we are able to prove the following recovery result.

Proposition 4.4.1. *For each $x \in W$ outside a $\mathbb{P}$-null set, there exists a unique word $w = (z_0, \cdots, z_{M^{\varepsilon,\delta}(x)}) \in \mathcal{W}_{M^{\varepsilon,\delta}(x)}$ such that*

$$\left[\phi_{z_0}^{\varepsilon,\delta}, \cdots, \phi_{z_{M^{\varepsilon,\delta}(x)}}^{\varepsilon,\delta}\right]_{0,1}(x) \neq 0.$$

This word is exactly given by $M^{\varepsilon,\delta}(x) = N^{\varepsilon,\delta}(x)$*, and*

$$z_i = \boldsymbol{m}_i^{\varepsilon,\delta}(x),\ i = 0, \cdots, M^{\varepsilon,\delta}(x).$$

Proof. Let $\mathcal{N}^{\varepsilon,\delta}$ be the set

$$\bigcup_{m=0}^{\infty} \bigcup_{w=(z_0,\cdots,z_m)\in\mathcal{W}_m} \bigcup_{i=0}^{m} \bigcup_{\substack{0\leqslant r_1<r_2\leqslant 1\\ r_1,r_2\in\mathbb{Q}}} \left(\left\{ x \in W : \int_{r_1}^{r_2} \phi_{z_i}^{\varepsilon,\delta}(dx_u) = 0 \right\} \bigcap A_{r_1,r_2}^{z_i,\varepsilon,\delta} \right),$$

where $A_{r_1,r_2}^{z_i,\varepsilon,\delta}$ is the set defined in (4.2.1) associated with the cube $H_{z_i}^{\varepsilon,\delta}$ and the differential one form $\phi_{z_i}^{\varepsilon,\delta}$. It follows from Assumption (C) that $\mathcal{N}^{\varepsilon,\delta}$ is a $\mathbb{P}$-null set.

For any $x \in (\mathcal{N}^{\varepsilon,\delta})^c$, let $w = (z_0, \cdots, z_m)$ be the word in $\mathcal{W}_m$ with $m = N^{\varepsilon,\delta}$ and $z_i = \boldsymbol{m}_i^{\varepsilon,\delta}$, for $i = 0, \ldots, m$, so $x \in E_w^{\varepsilon,\delta}$.

By (4.4.1) in Lemma 4.4.1, if $\left[\phi_{z_0}^{\varepsilon,\delta}, \cdots, \phi_{z_m}^{\varepsilon,\delta}\right]_{0,1}(x) = 0$, then there exists some $i = 1, \cdots, m+1$ such that $\int_{\tau_{i-1}^{\varepsilon,\delta}}^{\tau_i^{\varepsilon,\delta}} \phi_{z_{i-1}}^{\varepsilon,\delta}(dx_t) = 0$. By the definition of $\tau_k^{\varepsilon,\delta}$ and continuity, we can find some rational numbers $r_1 < \tau_{i-1}^{\varepsilon,\delta}$ and $r_2 < \tau_i^{\varepsilon,\delta}$ (if $m = 0$ take $r_1 = 0$ and $r_2 = 1$; otherwise if $i = 1$, take $r_1 = 0$ and if $i = m + 1$, take $r_2 = 1$) such that there exists some $u \in (r_1, r_2)$ with $x_u \in H_{z_{i-1}}^{\varepsilon,\delta}$ and

$$\int_{r_1}^{r_2} \phi_{z_{i-1}}^{\varepsilon,\delta}(dx_t) = \int_{\tau_{i-1}^{\varepsilon,\delta}}^{\tau_i^{\varepsilon,\delta}} \phi_{z_{i-1}}^{\varepsilon,\delta}(dx_t) = 0.$$

This implies that $x \in \mathcal{N}^{\varepsilon,\delta}$, which is a contradiction. Therefore, we have

$$\left[\phi_{z_0}^{\varepsilon,\delta}, \cdots, \phi_{z_m}^{\varepsilon,\delta}\right]_{0,1}(x) \neq 0.$$

By the second and third part of Lemma 4.4.1, we know that $M^{\varepsilon,\delta}(x) = m$ and w is the unique word in $\mathcal{W}_m$ such that the corresponding extended signature of x is nonzero. □

Together with the result in Section 4.3, proposition 4.4.1 tells us that outside a $\mathbb{P}$-null set, given the signature of a path x we can recover the sequence of open cubes $H_z^{\varepsilon,\delta}$ which x has visited in order.

4.4.2 An Approximation Result

Now we construct a polygonal approximation of a path based on the ordered sequence of open cubes visited by the path and the corresponding visit times. With

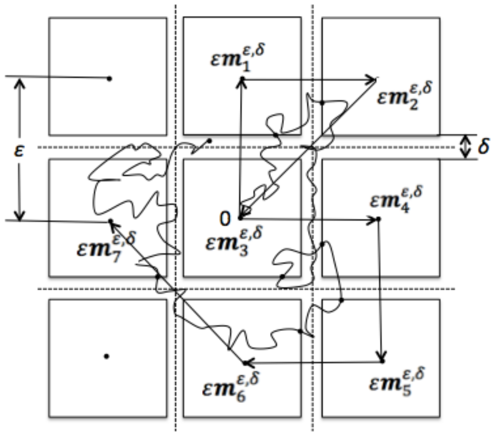

Figure 4.4.1: This figure illustrates the corresponding approximation scheme. The dotted lines represent the degenerate tunnels. According to Assumption (B) on the process, the probability that a path stays in these tunnels for a positive time period is zero, a crucial fact used in the proof of Proposition 4.4.2.

probability one, such polygonal approximations converge to the original path under the uniform topology. This result is crucial for the recovery of a path up to reparametrization from its signature.

Let $x \in W$ and define the word $w = (z_0, \cdots, z_m) \in \mathcal{W}_m$ by $m = N^{\varepsilon,\delta}$ and $z_i = \boldsymbol{m}_i^{\varepsilon,\delta}$ for $i = 0, \cdots, m$. Construct a polygonal path $x^{\varepsilon,\delta}$ as follows. If $m = 0$, let $x_t^{\varepsilon,\delta} = 0$ for $t \in [0,1]$; otherwise for $1 \leqslant k \leqslant m$, define

$$x_t^{\varepsilon,\delta} = \frac{\tau_k^{\varepsilon,\delta} - t}{\tau_k^{\varepsilon,\delta} - \tau_{k-1}^{\varepsilon,\delta}} \varepsilon z_{k-1} + \frac{t - \tau_{k-1}^{\varepsilon,\delta}}{\tau_k^{\varepsilon,\delta} - \tau_{k-1}^{\varepsilon,\delta}} \varepsilon z_k, \ t \in \left[\tau_{k-1}^{\varepsilon,\delta}, \tau_k^{\varepsilon,\delta}\right],$$

and

$$x_t^{\varepsilon,\delta} = \varepsilon z_m, \ t \in \left[\tau_m^{\varepsilon,\delta}, 1\right].$$

The approximation scheme is illustrated by Figure 4.4.1.

Now we have the following approximation result.

Proposition 4.4.2. *For each $n \geqslant 1$ and $\varepsilon_n = 1/n$, there exists $\delta_n > 0$, such that for $\mathbb{P}$-almost surely,*

$$\lim_{n\to\infty} \sup_{0\leqslant t\leqslant 1} \left| x_t^{\varepsilon_n,\delta_n} - x_t \right| = 0. \tag{4.4.3}$$

Proof. For each ε, δ, let

$$T^{\varepsilon,\delta} = \mathbb{R}^d \backslash \bigcup_{z\in\mathbb{Z}^d} H_z^{\varepsilon,\delta}$$

be the set of closed tunnels, and define

$$A^{\varepsilon,\delta} = \left\{x \in W : \exists [s,t] \subset x^{-1}\left(T^{\varepsilon,\delta}\right), |x_t - x_s| \geqslant \varepsilon\right\}.$$

We first show that for any fixed $\varepsilon > 0$,

$$\bigcap_{\delta>0} A^{\varepsilon,\delta} \subset \left\{x \in W : \exists 1 \leqslant i \leqslant d, k \in \mathbb{Z}, q \in \mathbb{Q}\bigcap(0,1) \text{ s.t. } x_q^i = \frac{2k-1}{2}\varepsilon\right\}. \quad (4.4.4)$$

Let $x \in \bigcap_{\delta>0} A^{\varepsilon,\delta}$, and δ_n be a sequence such that $\delta_n \downarrow 0$. Then for each $n \geqslant 1$, there exists $0 \leqslant s_n < t_n \leqslant 1$ such that $[s_n, t_n] \subset x^{-1}\left(T^{\varepsilon,\delta_n}\right)$ and $|x_{t_n} - x_{s_n}| \geqslant \varepsilon$. By compactness we can find a subsequence (s_{n_l}, t_{n_l}) of (s_n, t_n) such that (s_{n_l}, t_{n_l}) converges to some (s,t). The condition $\left|x_{t_{n_l}} - x_{s_{n_l}}\right| \geqslant \varepsilon$ then implies that $s < t$. Therefore, for fixed u, v with $s < u < v < t$, there exists some $N \in \mathbb{N}$ such that $[u,v] \subset \bigcap_{l\geqslant N}[s_{n_l}, t_{n_l}]$, and hence

$$\begin{aligned} [u,v] &\subset \bigcap_{l\geqslant N} x^{-1}\left(T^{\varepsilon,\delta_{n_l}}\right) \\ &= x^{-1}\left(\bigcup_{k\in\mathbb{Z}} \bigcup_{1\leqslant i\leqslant d} \mathbb{R}^{i-1} \times \left\{\frac{2k-1}{2}\varepsilon\right\} \times \mathbb{R}^{d-i}\right). \end{aligned}$$

In particular, this implies (4.4.4) and by Assumption (B) we have $\mathbb{P}\left(\bigcap_{\delta>0} A^{\varepsilon,\delta}\right) = 0$.

Now we show that for each ε, δ,

$$\left\{x \in W : \sup_{0\leqslant u\leqslant 1} \left|x_u^{\varepsilon,\delta} - x_u\right| \geqslant 11\sqrt{d}\varepsilon\right\} \subset A^{\varepsilon,\delta}. \quad (4.4.5)$$

To see this, first notice that if x belongs to the left hand side of (4.4.5), then either

(1) there exists some $u \in \left[\tau_{k-1}^{\varepsilon,\delta}, \tau_k^{\varepsilon,\delta}\right]$ for some $1 \leqslant k \leqslant N^{\varepsilon,\delta}$, such that $\left|x_u^{\varepsilon,\delta} - x_u\right| \geqslant 11\sqrt{d}\varepsilon$; or

(2) there exists some $u \in \left[\tau_{N^{\varepsilon,\delta}}^{\varepsilon,\delta}, 1\right]$, such that $\left|x_u - \varepsilon \boldsymbol{m}_{N^{\varepsilon,\delta}}^{\varepsilon,\delta}\right| \geqslant 11\sqrt{d}\varepsilon$.

In the first case, we know that x does not visit any cube other than $H_{\boldsymbol{m}_{k-1}^{\varepsilon,\delta}}^{\varepsilon,\delta}$ during $(\tau_{k-1}^{\varepsilon,\delta}, \tau_k^{\varepsilon,\delta})$. If the distance between the cubes $H_{\boldsymbol{m}_k^{\varepsilon,\delta}}^{\varepsilon,\delta}$ and $H_{\boldsymbol{m}_{k-1}^{\varepsilon,\delta}}^{\varepsilon,\delta}$ is at least $3\sqrt{d}\varepsilon$, by continuity there exist $\tau_{k-1}^{\varepsilon,\delta} < s < t < \tau_k^{\varepsilon,\delta}$, such that

$$\left|x_s - x_{\tau_{k-1}^{\varepsilon,\delta}}\right| = \sqrt{d}\varepsilon, \ \left|x_t - x_{\tau_{k-1}^{\varepsilon,\delta}}\right| = 2\sqrt{d}\varepsilon,$$

and $[s,t] \subset x^{-1}\left(T^{\varepsilon,\delta}\right)$. Moreover, by the triangle inequality we have $|x_t - x_s| \geqslant$

ε. Therefore, $x \in A^{\varepsilon,\delta}$. If the distance between $H^{\varepsilon,\delta}_{\boldsymbol{m}_k^{\varepsilon,\delta}}$ and $H^{\varepsilon,\delta}_{\boldsymbol{m}_{k-1}^{\varepsilon,\delta}}$ is strictly less than $3\sqrt{d}\varepsilon$, we know that $\left|x_u^{\varepsilon,\delta} - \varepsilon \boldsymbol{m}_{k-1}^{\varepsilon,\delta}\right| \leqslant 4\sqrt{d}\varepsilon$ for all $u \in \left(\tau_{k-1}^{\varepsilon,\delta}, \tau_k^{\varepsilon,\delta}\right)$. Since $\sup_{0\leqslant u\leqslant 1}\left|x_u^{\varepsilon,\delta} - x_u\right| \geqslant 11\sqrt{d}\varepsilon$, there exists $u \in \left(\tau_{k-1}^{\varepsilon,\delta}, \tau_k^{\varepsilon,\delta}\right)$ such that

$$\left|x_u - \varepsilon \boldsymbol{m}_{k-1}^{\varepsilon,\delta}\right|, \ \left|x_u - \varepsilon \boldsymbol{m}_{k}^{\varepsilon,\delta}\right| \geqslant 7\sqrt{d}\varepsilon.$$

It follows again from continuity that there exist $\tau_{k-1}^{\varepsilon,\delta} < s < t < \tau_k^{\varepsilon,\delta}$ such that

$$\left|x_s - \varepsilon \boldsymbol{m}_{k-1}^{\varepsilon,\delta}\right| = 5\sqrt{d}\varepsilon, \ \left|x_t - \varepsilon \boldsymbol{m}_{k-1}^{\varepsilon,\delta}\right| = 6\sqrt{d}\varepsilon,$$

and $[s,t] \subset x^{-1}\left(T^{\varepsilon,\delta}\right)$. Therefore, $|x_s - x_t| \geqslant \varepsilon$ and we have $x \in A^{\varepsilon,\delta}$.

In the second case, there exist $\tau_{N^{\varepsilon,\delta}}^{\varepsilon,\delta} < s < t \leqslant 1$ such that

$$\left|x_s - \varepsilon \boldsymbol{m}_{N^{\varepsilon,\delta}}^{\varepsilon,\delta}\right| = \sqrt{d}\varepsilon, \ \left|x_t - \varepsilon \boldsymbol{m}_{N^{\varepsilon,\delta}}^{\varepsilon,\delta}\right| = 2\sqrt{d}\varepsilon,$$

and $[s,t] \subset x^{-1}(T^{\varepsilon,\delta})$. Again we have $|x_t - x_s| \geqslant \varepsilon$ and hence $x \in A^{\varepsilon,\delta}$.

Now for $\varepsilon_n = 1/n$, if we choose δ_n small enough such that $\mathbb{P}\left(A^{\varepsilon_n,\delta_n}\right) \leqslant \varepsilon_n^2$, we have

$$\begin{aligned} \sum_{n=1}^{\infty} \mathbb{P}\left(\left\{x \in W : \sup_{0\leqslant u\leqslant 1} |x_u^{\varepsilon_n,\delta_n} - x_u| \geqslant 11\sqrt{d}\varepsilon_n\right\}\right) &\leqslant \sum_{n=1}^{\infty} \mathbb{P}\left(A^{\varepsilon_n,\delta(\varepsilon_n)}\right) \\ &< \infty, \end{aligned}$$

It follows from the Borel-Cantelli lemma that

$$\mathbb{P}\left(\limsup_{n\to\infty}\left\{x \in W : \sup_{0\leqslant u\leqslant 1} |x_u^{\varepsilon_n,\delta_n} - x_u| \geqslant 11\sqrt{d}\varepsilon_n\right\}\right) = 0,$$

and hence the uniform convergence (4.4.3) holds $\mathbb{P}$-almost surely. □

Remark 4.4.2. From the previous proof, it is not hard to see that the result of Proposition 4.4.2 holds for all continuous stochastic processes starting at the origin whose law satisfies Assumption (B).

From now on, we always assume that $\varepsilon_n = 1/n$, and take δ_n as in the previous proof.

4.4.3 A Variant of the Fréchet Distance on Path Space

Now we are coming to the last step of the proof of Theorem 4.2.1.

Under Assumption (A), (B), (C), what we have obtained so far is that there exists some $\mathbb{P}$-null set $\mathcal{N}$, such that for any path $x \in \mathcal{N}^c$, the signature $S(x)_{0,1}$ is well-defined, and for each $n \geqslant 1$, we can recover the ordered sequence of open cubes $H_z^{\varepsilon_n,\delta_n}$ visited by x from its signature. Moreover, the polygonal approximation $x^{\varepsilon_n,\delta_n}$ constructed before converges to x uniformly.

By possibly enlarging the $\mathbb{P}$-null set $\mathcal{N}$ (still a $\mathbb{P}$-null set), we show that for any two paths $x, x' \in \mathcal{N}^c$, if $S(x)_{0,1} = S(x')_{0,1}$, then x and x' differ by a reparametrization $\sigma \in \mathcal{R}$ in the sense of Definition 1.2.9.

Now we introduce an equivalence relation "$\sim$" on W by

$$x \sim x' \iff (x_t)_{0\leqslant t\leqslant 1} = \left(x'_{\sigma(t)}\right)_{0\leqslant t\leqslant 1}, \text{ for some } \sigma \in \mathcal{R}.$$

Let $W/_\sim$ be the quotient space consisting of $\sim$-equivalence classes. For any $[x], [x'] \in W/_\sim$, define

$$d\left([x],[x']\right) = \inf_{\sigma\in\mathcal{R}} \sup_{t\in[0,1]} \left|x_t - x'_{\sigma(t)}\right|. \tag{4.4.6}$$

If we only assume that σ is non-decreasing, the function $d(\cdot,\cdot)$ is usually know as the Fréchet distance. It was originally introduced by Fréchet to study the shape of geometric spaces. Here we emphasize that σ is strictly increasing.

It is easy to see that $d(\cdot,\cdot)$ does not depend on the choice of representatives in the corresponding equivalence classes, and $d(\cdot,\cdot)$ is non-negative and symmetric. Moreover, $d(\cdot,\cdot)$ satisfies the triangle inequality. In fact, for any $x, x', x'' \in W$ and $\sigma, \theta \in \mathcal{R}$, we have

$$\sup_{t\in[0,1]} \left|x_t - x''_{\sigma(t)}\right| \leqslant \sup_{t\in[0,1]} \left|x_t - x'_{\theta(t)}\right| + \sup_{t\in[0,1]} \left|x'_{\theta(t)} - x''_{\sigma(t)}\right|.$$

It follows that

$$\begin{aligned} d\left([x],[x'']\right) &= \inf_{\sigma\in\mathcal{R}} \sup_{t\in[0,1]} \left|x_t - x''_{\sigma(t)}\right| \\ &\leqslant \sup_{t\in[0,1]} \left|x_t - x'_{\theta(t)}\right| + \inf_{\sigma\in\mathcal{R}} \sup_{t\in[0,1]} \left|x'_{\theta(t)} - x''_{\sigma(t)}\right| \\ &= \sup_{t\in[0,1]} \left|x_t - x'_{\theta(t)}\right| + d\left([x'],[x'']\right). \end{aligned}$$

By taking infimum over $\theta \in \mathcal{R}$, we obtain the triangle inequality.

It should be pointed out that unlike the Fréchet distance, $d(\cdot,\cdot)$ is not a metric on

$W/_\sim$. For example, consider the one dimensional case. Let $x_t = t,\ t \in [0,1]$, and

$$x'_t = \begin{cases} 2t, & t \in [0, \frac{1}{2}]; \\ 1, & t \in [\frac{1}{2}, 1]. \end{cases}$$

Then it is easy to see that $d([x],[x']) = 0$, but obviously x' is not a reparametrization of x in the sense of Definition 1.2.9. However, we are going to show that, if we exclude paths with certain degeneracy, then on the corresponding quotient space $d(\cdot,\cdot)$ is indeed a metric.

Let D be the set of paths $x \in W$ such that there exist some $0 \leqslant s < t \leqslant 1$ with

$$x_u = x_s,\ \forall u \in [s,t].$$

We first make an important remark that under Assumption (C), D is a $\mathbb{P}$-null set. To see this, let $\{H_n\}_{n\geqslant 1}$ be a covering of $\mathbb{R}^d$ consisting of open cubes, and for each n let ϕ_n be the differential one form associated with H_n according to Assumption (C). It follows that

$$D \subset \bigcup_{r_1,r_2\in\mathbb{Q}\bigcap[0,1]} \bigcup_{n\geqslant 1} \left(\left\{x \in W : \int_{r_1}^{r_2} \phi_n(dx_u) = 0\right\} \bigcap A_{r_1,r_2}^{H_n}\right\}.$$

Therefore, by Assumption (C) we know that $\mathbb{P}(D) = 0$.

Now we have the following result.

Proposition 4.4.3. *Define the equivalence relation "$\sim$" on $W_0 = D^c \subset W$ as before, and let $W_0/_\sim$ be the corresponding quotient space. Then $d(\cdot,\cdot)$, defined in the same way as in (4.4.6), is a metric on $W_0/_\sim$.*

Proof. It suffices to show that, for any $x, x' \in W_0$, if

$$\inf_{\sigma\in\mathcal{R}} \sup_{t\in[0,1]} \left|x_t - x'_{\sigma(t)}\right| = 0, \tag{4.4.7}$$

then

$$x_t = x'_{\sigma(t)},\ \forall t \in [0,1], \tag{4.4.8}$$

for some $\sigma \in \mathcal{R}$.

In fact, by (4.4.7), for any $n \geqslant 1$, there exists $\sigma_n \in \mathcal{R}$, such that

$$\left|x_t - x'_{\sigma_n(t)}\right| \leqslant \frac{1}{n},\ \forall t \in [0,1]. \tag{4.4.9}$$

It follows from compactness, denseness, and a standard diagonal selection argument that we can find a subsequence $\{\sigma_{n_k}\}$ such that for any $r \in \mathbb{Q}\bigcap[0,1]$,

$$\lim_{k\to\infty} \sigma_{n_k}(r) =: \widetilde{\sigma}(r)$$

exists.

Now define $\sigma:\ [0,1] \to [0,1]$ by

$$\sigma(t) = \begin{cases} \inf\left\{\widetilde{\sigma}(r):\ r > t, r \in \mathbb{Q}\bigcap[0,1]\right\}, & 0 \leqslant t < 1; \\ 1, & t = 1. \end{cases}$$

We want to show that $\sigma \in \mathcal{R}$, and it satisfies (4.4.8).

(1) It is easy to see that σ is increasing. Let $0 \leqslant t < 1$. For any $\varepsilon > 0$, there exists some $r > t, r \in \mathbb{Q}\bigcap[0,1]$, such that

$$\sigma(t) \leqslant \widetilde{\sigma}(r) < \sigma(t) + \varepsilon.$$

Therefore, for any $t' \in (t,r)$, if we take some $r' \in \mathbb{Q}\bigcap[0,1]$ with $t' < r' < r$, then

$$\sigma(t) \leqslant \sigma(t') \leqslant \widetilde{\sigma}(r') \leqslant \widetilde{\sigma}(r) < \sigma(t) + \varepsilon.$$

It follows that σ is right continuous.

(2) σ is also left continuous.

In fact, assume on the contrary that for some $0 < t \leqslant 1$, $\sigma(t-) \neq \sigma(t)$. Fix any $\sigma(t-) < s < \sigma(t)$, and define for $k \geqslant 1$, $t_{n_k} = \sigma_{n_k}^{-1}(s)$. It follows that for any $r > t, r \in \mathbb{Q}\bigcap[0,1]$,

$$s < \sigma(t) \leqslant \widetilde{\sigma}(r).$$

Since $\lim_{k\to\infty} \sigma_{n_k}(r) = \widetilde{\sigma}(r)$, we know that when k is large enough, $s < \sigma_{n_k}(r)$, which is equivalent to $t_{n_k} < r$ for k large enough. Therefore, we have $\limsup_{k\to\infty} t_{n_k} \leqslant r$. But this is true for all $r > t, r \in \mathbb{Q}\bigcap[0,1]$, which implies that $\limsup_{k\to\infty} t_{n_k} \leqslant t$. On the other hand, for any $r < t, r \in \mathbb{Q}\bigcap[0,1]$, we have

$$\widetilde{\sigma}(r) \leqslant \sigma(r) \leqslant \sigma(t-) < s,$$

A similar argument yields that $\liminf_{k\to\infty} t_{n_k} \geqslant t$. Therefore, $\lim_{k\to\infty} t_{n_k}$ exists and is equal to t. Now from (4.4.9) we know that

$$\left|x_{t_{n_k}} - x'_s\right| \leqslant \frac{1}{n_k},\ \forall k \geqslant 1,$$

and hence $x_t = x'_s$. But this is true for all $\sigma(t-) < s < \sigma(t)$, which contradicts the fact that $x' \in W_0$. Therefore, σ is left continuous. A similar argument also shows that $\sigma(0) = 0$.

(3) For any $r \in \mathbb{Q}\bigcap[0,1]$, $\sigma(r) = \widetilde{\sigma}(r)$.

In fact, it is obvious that $\sigma(r) \geqslant \widetilde{\sigma}(r)$. On the other hand, for any $t < r$ we have $\sigma(t) \leqslant \widetilde{\sigma}(r)$, and by the left continuity of σ we have $\sigma(r) \leqslant \widetilde{\sigma}(r)$.

(4) σ is strictly increasing.

In fact, if for some $0 \leqslant s < t \leqslant 1$, $\sigma(s) = \sigma(t)$, then σ remains constant over $[s,t]$. In particular, for any $r \in \mathbb{Q}\bigcap[s,t]$, from (4.4.9) and the previous step we have

$$x_r = x'_{\widetilde{\sigma}(r)} = x'_{\sigma(r)} = x'_{\sigma(s)},$$

which implies that x is constant over $[s,t]$, contradicting the fact that $x \in W_0$.

Now it is obvious that $\sigma \in \mathcal{R}$, and (4.4.8) follows. □

From now on, we include D in the $\mathbb{P}$-null set $\mathcal{N}$.

Now we are in a position to complete the proof of Theorem 4.2.1.

Assume that $x, x' \in \mathcal{N}^c$ and $S(x)_{0,1} = S(x')_{0,1}$. For $n \geqslant 1$, let $\left(\phi_{z_0}^{\varepsilon_n,\delta_n}, \cdots, \phi_{z_m}^{\varepsilon_n,\delta_n}\right)$ ($\left(\phi_{z_0}^{\varepsilon_n,\delta_n}, \cdots, \phi_{z'_{m'}}^{\varepsilon_n,\delta_n}\right)$, respectively) be the unique maximal sequence of differential one forms along which the extended signature of x (x', respectively) is nonzero. It follows from Theorem 4.3.1 that $m = m'$ and $z_i = z'_i$ for $i = 0, \ldots, m$. Moreover, by Proposition 4.4.1 we know that

$$N^{\varepsilon_n,\delta_n}(x) = N^{\varepsilon_n,\delta_n}(x') = m,$$

and

$$\boldsymbol{m}_i^{\varepsilon_n,\delta_n}(x) = \boldsymbol{m}_i^{\varepsilon_n,\delta_n}(x') = z_i,\ \forall i = 0, \cdots, m.$$

It follows that in the quotient space $W/_\sim$, $\left[x^{\varepsilon_n,\delta_n}\right] = \left[(x')^{\varepsilon_n,\delta_n}\right]$, where $x^{\varepsilon_n,\delta_n}$ and $(x')^{\varepsilon_n,\delta_n}$ are the polygonal approximation of x and x' respectively. On the other hand, by Proposition 4.4.2 we know that

$$x^{\varepsilon_n,\delta_n} \to x,\ (x')^{\varepsilon_n,\delta_n} \to x',$$

under the uniform topology as $n \to \infty$. Therefore, by the triangle inequality of the distance function $d(\cdot,\cdot)$ we have $d([x],[x']) = 0$. Since $D \subset \mathcal{N}$, it follows from Proposition 4.4.3 that there exists $\sigma \in \mathcal{R}$, such that (4.4.8) holds.

Now the proof of Theorem 4.2.1 is complete.

4.5 A Fundamental Example: Gaussian Processes

As we have remarked before, Assumption (A) and (B) are natural for a large class of stochastic processes. However, Assumption (C) is in general difficult to verify. In this section, as a fundamental example of Theorem 4.2.1, we show that Assumption (A), (B), (C) hold for a class of Gaussian processes including fractional Brownian motion with Hurst parameter $H > 1/4$, the Ornstein-Uhlenbeck process and the Brownian bridge. The main idea of verifying Assumption (C) for Gaussian processes is to apply local regularity results for Gaussian functionals from the Malliavin calculus, based on pathwise integration by parts which is possible due to the regularity of sample paths and Cameron-Martin paths.

The class of Gaussian processes we study in this section is specified in the following.

Let $\mathbb{P}$ be the law of a centered, non-degenerate, continuous Gaussian process over $[0,1]$ starting at the origin with i.i.d components. We assume that $\mathbb{P}$ satisfies the following conditions: there exists $H \in \left(\frac{1}{4}, 1\right)$ such that

(G1) for all $\rho \in \left(\frac{1}{2H} \vee 1, 2\right]$, the ρ-variation of the covariance function (see [26], Definition 5.50) of each component of X is controlled by a 2-dimensional Hölder-dominated control (see [26], Definition 5.51);

(G2) there exists $\delta > 0$ and $c_\delta > 0$, such that for all $0 \leqslant s < t \leqslant 1$ with $|t-s| \leqslant \delta$, we have

$$\mathbb{E}\left[(X_t - X_s)^2\right] \geqslant c_\delta (t-s)^{2H};$$

(G3) the Cameron-Martin space $\mathcal{H}$ associated with $\mathbb{P}$ satisfies the property that

$$C_0^{1+H-}([0,1];\mathbb{R}^d) \subset \mathcal{H} \subset C_0^{q-var}([0,1];\mathbb{R}^d),\ \forall q > \left(H + \frac{1}{2}\right)^{-1},$$

where $C_0^{1+H^-}([0,1];\mathbb{R}^d)$ is the space of differentiable paths in W with Hölder continuous derivatives of any order smaller than H, and $C_0^{q-var}([0,1];\mathbb{R}^d)$ is the space of paths in W with finite total q-variation.

Now we prove our second main result, namely Theorem 4.2.2. Note that in this case the verification of Assumption (A) is a standard result for Gaussian rough paths according to (G1) (see [26], Theorem 15.33), and Assumption (B) is trivial. The main difficulty is the verification of Assumption (C).

For any open cube $H_{x_0,\eta}$ with center $x_0 = (x_0^1, \cdots, x_0^d) \in \mathbb{R}^d$ and edges of length 2η, we construct a differential one form ϕ supported on the closure of $H_{x_0,\eta}$, such that for any $0 \leqslant s < t \leqslant 1$,

$$\mathbb{P}\left(\left\{x \in W : \int_s^t \phi(dx_u) = 0\right\} \bigcap A_{s,t}^{H_{x_0,\eta}}\right) = 0, \tag{4.5.1}$$

where $A_{s,t}^{H_{x_0,\eta}}$ is the set defined by (4.2.1). In other words, Assumption (C) holds.

Let $h(t) \in C_c^\infty(\mathbb{R}^1)$ be a function such that

$$\begin{cases} h(t) > 0, & t \in (-1,1); \\ h(t) = 0, & t \notin (-1,1), \end{cases}$$

and $h'(t)$ is everywhere nonzero in $(-1,1)$ except at $t = 0$. For example, the standard mollifier function

$$h(t) = \begin{cases} e^{\frac{-1}{1-|t|^2}}, & t \in (-1,1); \\ 0, & t \notin (-1,1), \end{cases}$$

satisfies the properties.

Define a differential one form $\phi(x) = \sum_{i=1}^d \phi_i(x)dx^i$ on $\mathbb{R}^d$ by

$$\begin{aligned} \phi_1(x) &= h\left(\frac{x^1 - x_0^1}{\eta}\right) \cdots h\left(\frac{x^d - x_0^d}{\eta}\right) \exp\left(h^2\left(\frac{x^2 - x_0^2}{\eta}\right)\right), \ x \in \mathbb{R}^d, \\ \phi_i &= 0, \ \text{for all } i = 2, \cdots, d. \end{aligned} \tag{4.5.2}$$

It is easy to see that the support of ϕ is exactly the boundary of the $H_{x_0,\eta}$. Moreover, we have

$$\begin{aligned} \frac{\partial \phi_1}{\partial x^2}(x) &= \frac{1}{\eta}\left(\prod_{i \neq 2} h\left(\frac{x^i - x_0^i}{\eta}\right)\right) h'\left(\frac{x^2 - x_0^2}{\eta}\right) \\ &\quad \cdot \exp\left(h^2\left(\frac{x^2 - x_0^2}{\eta}\right)\right)\left(1 + 2h^2\left(\frac{x^2 - x_0^2}{\eta}\right)\right), \end{aligned}$$

which is everywhere nonzero in $H_{x_0,\eta}$ except on the slice $\{x \in H_{x_0,\eta} : \ x^2 = x_0^2\}$.

To verify Assumption (C) for such a differential one form ϕ, we need the following Lemma.

Lemma 4.5.1. *Fix $0 \leqslant s < t \leqslant 1$. Let f be a smooth function on $\mathbb{R}^d$ with com-*

pact support. Then there exists a $\mathbb{P}$-null set $\mathcal{N}_1$ such that for any $x \in (\mathcal{N}_1)^c$, if $\int_u^v f(x_r)dx_r^1 = 0$ for all u, v with $[u,v] \subset [s,t]$, then $f(x_u) = 0$ for all $u \in [s,t]$.

Proof. Fix $\frac{1}{2H} < \rho < \frac{1}{H}$. According to (G1) and [26], Theorem 15.33, outside some $\mathbb{P}$-null set $\mathcal{N}_0'$, a sample path x admits a canonical lifting to a geometric 2ρ-rough path $\mathbf{X}$ as well as a $G^{\lfloor 2\rho \rfloor}(\mathbb{R}^d)$-valued $\frac{1}{2\rho}$-Hölder continuous path ($G^N(\mathbb{R}^d)$ is the free nilpotent group of step N over $\mathbb{R}^d$, see [26], Theorem 7. 30). Since the path integral $\int_u^v f(x_r)dx_r^1$ can be regarded as the projection of the solution to the rough differential equation

$$\begin{cases} dx_r^1 = dx_r^1, \\ \cdots, \\ dx_r^d = dx_r^d, \\ dx_r^{d+1} = f(x_r^1, \cdots, x_r^d)dx_r^1 \end{cases}$$

over $[u,v]$ with initial condition $(x_u^1, \cdots, x_u^d, x_u^{d+1}) = (x_u^1, \cdots, x_u^d, 0)$, according to [26], Corollary 10.15, we know that pathwisely

$$\left| \int_u^v f(x_r)dx_r^1 - f(x_u)\mathbf{X}_{u,v}^{1;1} - \sum_{i=1}^d \frac{\partial f}{\partial x^i}(x_u)\mathbf{X}_{u,v}^{2;i,1} - \sum_{i,j=1}^d \frac{\partial^2 f}{\partial x^i \partial x^j}(x_u)\mathbf{X}_{u,v}^{3;i,j,1} \right|$$
$$\leqslant C_1 \left\| \mathbf{X} \right\|_{\frac{1}{2\rho}-\text{Höl};[u,v]}^{2\rho\theta} |u-v|^\theta,$$

where $\theta > 1$ and C_1 is some positive constant depending only on ρ, θ and the uniform bounds on the derivatives of f. If $\int_u^v f(x_r)dx_r^1 = 0$, then we have

$$\begin{aligned} &\left| f(x_u) \left(x_v^1 - x_u^1 \right) \right| \\ &\leqslant C_1 \left\| \mathbf{X} \right\|_{\frac{1}{2\rho}-\text{Höl};[u,v]}^{2\rho\theta} |u-v|^\theta + \|Df\|_\infty |\pi_2(\mathbf{X}_{u,v})| + \|D^2 f\|_\infty |\pi_3(\mathbf{X}_{u,v})| \end{aligned} \tag{4.5.3}$$

On the other hand, according to (G1) and [26], Proposition 15.19, Corollary 15.21 and Theorem 15.33, we know that

$$\mathbb{E} \left| \pi_j \left(\mathbf{X}_{u,v} \right) \right|^2 \leqslant C_2 |u-v|^{j/\rho} \tag{4.5.4}$$

for each level j, where C_2 is some positive constant depending only on ρ. Now we choose α, γ such that $H < \alpha < \gamma < \frac{1}{\rho}$. According to (G2) and (4.5.4), it follows from the Borel-Cantelli lemma that

$$\begin{aligned}\mathcal{N}(u) \quad := \quad & \left\{x \in W: \left|x^1_{u+\frac{1}{2^n}} - x^1_u\right| \leqslant \frac{1}{2^{\alpha n}}, \text{ for infinitely many } n\right\} \\ & \bigcup \left\{x \in W: \left|\pi_2\left(\mathbf{X}_{u,u+\frac{1}{2^n}}\right)\right| \geqslant \frac{1}{2^{\gamma n}}, \text{ for infinitely many } n\right\} \\ & \bigcup \left\{x \in W: \left|\pi_3\left(\mathbf{X}_{u,u+\frac{1}{2^n}}\right)\right| \geqslant \frac{1}{2^{\gamma n}}, \text{ for infinitely many } n\right\}\end{aligned}$$

is a $\mathbb{P}$-null set.

Let $x \in (\mathcal{N}_0' \bigcup \mathcal{N}(u))^c$. Then there exists some $N \geqslant 1$, such that

$$\left|x^1_{u+\frac{1}{2^n}} - x^1_u\right| > \frac{1}{2^{\alpha n}}, \ \left|\pi_2\left(\mathbf{X}_{u,u+\frac{1}{2^n}}\right)\right| < \frac{1}{2^{\gamma n}}, \ \left|\pi_3\left(\mathbf{X}_{u,u+\frac{1}{2^n k}}\right)\right| < \frac{1}{2^{\gamma n}},$$

for all $n > N$. Therefore, by (4.5.3) with $v = u + 1/2^n$, for any $n > N$ we have

$$\begin{aligned}& |f(x_u)| \\ \leqslant\ & \frac{1}{2^{n(\theta-\alpha)}} C_1 \left\|\mathbf{X}\right\|^{2\rho\theta}_{\frac{1}{2\rho}-\text{Höl};[0,1]} + \frac{1}{2^{n(\gamma-\alpha)}} \left(\|Df\|_\infty + \|D^2 f\|_\infty\right).\end{aligned}$$

By taking $n \to \infty$, we have $f(x_u) = 0$.

Now the result follows easily if we take

$$\mathcal{N}_1 = \mathcal{N}_0' \bigcup \bigcup_{u \in \mathbb{Q} \bigcap [s,t]} \mathcal{N}(u).$$

□

Remark 4.5.1. By the denseness argument, it is easy to see that the $\mathbb{P}$-null set $\mathcal{N}_1$ can be taken uniformly in s, t.

Now we complete the proof of Theorem 4.2.2.

In what follows, for simplicity we use Einstein's summation convention: repeated indices of superscript and subscript are automatically summed over from 1 to d.

Let $F(x) = \int_s^t \phi(dx_u) = \int_s^t \phi_i(x_u) dx^i_u$. It follows that F is smooth in the sense of Malliavin (see for example [7], and Y. Inahama [38]). Since F is a random variable on the abstract Wiener space $(W, \mathcal{H}, \mathbb{P})$, it suffices to show that outside a $\mathbb{P}$-null set, for any $x \in A^{H_{x_0,\eta}}_{s,t}$ the Malliavin derivative $DF(x)$ is a nonzero element in the Cameron-Martin space $\mathcal{H}$. It then follows from standard local regularity results from the Malliavin calculus (see for example the monograph by D. Nualart [51], Theorem

2.1.1 and the following remark on p. 93) that the measure

$$\lambda(B) = \mathbb{P}\left(\{F \in B\} \bigcap A_{s,t}^{H_{x_0,\eta}}\right),\ B \in \mathcal{B}\left(\mathbb{R}^1\right),$$

is absolutely continuous with respect to the Lebesgue measure in $\mathbb{R}^1$. In particular, (4.5.1) holds.

Let $\mathcal{N}_1$ be the null set in Lemma 4.5.1. We know that $\mathbb{P}$-almost surely sample paths can be lifted as geometric p-rough paths for $1 < p < 4$ with $Hp > 1$, and according to (G3) we have $\mathcal{H} \subset C_0^{q-var}([0,1];\mathbb{R}^d)$ for any $q > (H+1/2)^{-1}$. Obviously we can choose such p, q so that $1/p + 1/q > 1$. Therefore, in the sense of Young's integrals we know that for any $x \in A_{s,t}^{H_{x_0,\eta}} \bigcap \mathcal{N}_1^c$ and $h \in \mathcal{H}$,

$$\begin{aligned}
\langle DF(x), h\rangle_{\mathcal{H}} &= \frac{d}{d\varepsilon}|_{\varepsilon=0} F(x+\varepsilon h) \\
&= \frac{d}{d\varepsilon}|_{\varepsilon=0} \int_s^t \phi_i(x_u+\varepsilon h_u)d(x_u^i+\varepsilon h_u^i) \\
&= \int_s^t \frac{\partial \phi_i}{\partial x^j}(x_u)h_u^j dx_u^i + \int_s^t \phi_i(x_u)dh_u^i,
\end{aligned}$$

where the interchange of differentiation and integration can be verified easily by the geometric rough path nature of x and the continuity of the integration map.

Integration by parts shows that

$$\int_s^t \phi_i(x_u)dh_u^i = \phi_i(x_t)h_t^i - \phi_i(x_s)h_s^i - \int_s^t h_u^i \frac{\partial \phi_i}{\partial x^j}(x_u)dx_u^j.$$

Therefore,

$$\langle DF(x), h\rangle_{\mathcal{H}} = \left(\phi_i(x_t)h_t^i - \phi_i(x_s)h_s^i\right) + \int_s^t \left(\frac{\partial \phi_i}{\partial x^j} - \frac{\partial \phi_j}{\partial x^i}\right)(x_u)h_u^j dx_u^i.$$

Let

$$Y_{u,j} = \int_s^u \left(\frac{\partial \phi_i}{\partial x^j} - \frac{\partial \phi_j}{\partial x^i}\right)(x_v)dx_v^i,\ u \in [0,1],\ j = 1, \cdots, d. \tag{4.5.5}$$

It follows from integration by parts again that

$$\begin{aligned}
\langle DF(x), h\rangle_{\mathcal{H}} &= (\phi_i(x_t)h_t^i - \phi_i(x_s)h_s^i) + \int_s^t h_u^i dY_{u,i} \\
&= (\phi_i(x_t) + Y_{t,i})h_t^i - (\phi_i(x_s) + Y_{s,i})h_s^i - \int_s^t Y_{u,i}dh_u^i.
\end{aligned}$$

Now we define $h = (h^1, \cdots, h^d)$ by

$$h_u^i = \int_s^u (\phi_i(x_t) + Y_{t,i} - Y_{v,i})dv, \ u \in [0,1], \ i = 1, \cdots, d, \tag{4.5.6}$$

then $h_s^i = 0$ for $i = 1, \ldots, d$. Technically if $s > 0$ we modify h^i smoothly on $\left[0, \frac{s}{2}\right)$ so that $h_0^i = 0$ for all i. Note that the modification does not change the value of $\langle DF(x), h\rangle_{\mathcal{H}}$ as it depends only on the value of h on $[s,t]$. By the regularity of sample paths, it is easy to see that $h \in C_0^{1+H^-}([0,1];\mathbb{R}^d)$, which is also in $\mathcal{H}$ according to (G3). Therefore,

$$\langle DF(x), h\rangle_{\mathcal{H}} = \sum_{i=1}^d \int_s^t (\phi_i(x_t) + Y_{t,i} - Y_{u,i})^2 du.$$

If $DF(x) = 0$, then $\langle DF(x), h\rangle_{\mathcal{H}} = 0$, which implies that for all $i = 1, \cdots, d$, and $u \in [s,t]$, $\phi_i(x_t) + Y_{t,i} - Y_{u,i} = 0$. It follows from taking $i = 2$ and our construction of ϕ that

$$\int_u^v \frac{\partial \phi_1}{\partial x^2}(x_r)dx_r^1 = 0, \ \forall [u,v] \subset [s,t].$$

Therefore, by Lemma 4.5.1 we have for all $u \in [s,t]$, $\frac{\partial \phi_1}{\partial x^2}(x_u) = 0$.

On the other hand, since $x \in A_{s,t}^{H_{x_0,\eta}}$, there exists some $u \in (s,t)$ such that $x_u \in H_{x_0,\eta}$. From the construction of ϕ we have already seen that $\frac{\partial \phi_1}{\partial x^2}$ is everywhere nonzero in $H_{x_0,\eta}$ except on the "slice"

$$L_{x_0,\eta} = \left\{x \in H_{x_0,\eta} : \ x^2 = x_0^2\right\}.$$

Therefore, by continuity there exists some open interval $(u,v) \subset [s,t]$, such that $x_r \in L_{x_0,\eta}$ for all $r \in (u,v)$. But this implies that there exists some $r \in \mathbb{Q}\bigcap(s,t)$ such that $x_r^2 = x_0^2$. Since for any $r \in (0,1)$, the law of x_r is absolutely continuous with respect to the Lebesgue measure, we know that

$$\mathcal{N}_2 := \bigcup_{r\in\mathbb{Q}\bigcap(0,1)} \{x_r^2 = x_0^2\}$$

is a $\mathbb{P}$-null set. By further removing $\mathcal{N}_2$, we arrive at a contradiction. Therefore, for any $x \in A_{s,t}^H \bigcap \mathcal{N}_1^c \bigcap \mathcal{N}_2^c$, $DF(x)$ is a nonzero element in $\mathcal{H}$.

Now the proof of Theorem 4.2.2 is complete.

In the rest of this chapter we consider three specific examples of Gaussian processes which all verify conditions (G1), (G2) and (G3): fractional Brownian motion with

Hurst parameter $H > 1/4$, the Ornstein-Uhlenbeck process and the Brownian bridge.

4.5.1 Fractional Brownian Motion with Hurst Parameter $H > 1/4$

Let X be the d-dimensional fractional Brownian motion with Hurst parameter H for $H > 1/4$. In other words, X is a Gaussian process starting at the origin with i.i.d. components, and the covariance function of X^i is given by

$$R^H(s,t) = \frac{1}{2}\left(s^{2H} + t^{2H} - |t-s|^{2H}\right), \ s,t \in [0,1].$$

In this case the parameter H in the conditions (G1), (G2) and (G3) is just the Hurst parameter. The verification of Condition (G1) is the content of [26], Proposition 15.5 if $H \in (1/4, 1/2]$ (the case when $H > 1/2$ is trivial in the rough path setting), and (G2) follows from direct calculation. The verification of (G3) is contained in the following two lemmas.

Let $\mathcal{H}^H$ be the Cameron-Martin space associated with X.

Lemma 4.5.2. *$\mathcal{H}^H$ contains $C_0^\alpha([0,1];\mathbb{R}^d)$ for all $\alpha > H + 1/2$.*

Proof. We assume $H \neq 1/2$, as the result is well-known for Brownian motion. According to L. Decreusefond and A. Ustunel [18], Theorem 2.1, we have

$$\mathcal{H}^H = \mathcal{I}_{0+}^{H+\frac{1}{2}}\left(L^2[0,1]\right),$$

where

$$\mathcal{I}_{0+}^{\alpha}(f)(x) = \int_0^x f(t)(x-t)^{\alpha-1}\,dt$$

is the fractional integral operator.

If $0 < H < 1/2$, from fractional calculus (see the monograph by S. Samko, A. Kilbas and O. Marichev [59], p. 233) we know that $\mathcal{I}_{0+}^{H+\frac{1}{2}}(L^2[0,1])$ contains all α-Hölder continuous functions whenever $\alpha > H + 1/2$. If $H > 1/2$, by the fundamental theorem of calculus we know that $h \in \mathcal{I}_{0+}^{H+\frac{1}{2}}(L^2[0,1])$ if and only if h is differentiable with derivative in $\mathcal{I}_{0+}^{H-\frac{1}{2}}(L^2[0,1])$. Therefore, in both cases we have $\mathcal{H}^H$ containing $C_0^\alpha\left([0,1];\mathbb{R}^d\right)$ for all $\alpha > H + 1/2$. □

Lemma 4.5.3. *(1) (see [18], Theorem 2.1, Theorem 3.3 and [59], Theorem 3.6) If $H > 1/2$, we have*

$$\mathcal{H}^H \subset C_0^H\left([0,1];\mathbb{R}^d\right). \tag{4.5.7}$$

(2) (see [24], Corollary 1) If $0 < H \leqslant 1/2$*, then for any* $q > (H+1/2)^{-1}$*, we have*

$$\mathcal{H}^H \subset C_0^{q-var}\left([0,1];\mathbb{R}^d\right).$$

Remark 4.5.2. From the proof of Theorem 4.2.2 we can see that the embedding $\mathcal{H}^H \subset C_0^{q-var}([0,1];\mathbb{R}^d)$ is only used for making sense of path integrals in the sense of L.C. Young. Therefore, when $H > 1/2$, (4.5.7) is obviously sufficient for us to carry out all the calculations before as we are also in the setting of Young's integrals.

4.5.2 The Ornstein-Uhlenbeck Process

Let

$$X_t = \int_0^t e^{-(t-s)}dB_s, \ t \in [0,1],$$

be the standard Ornstein-Uhlenbeck process in $\mathbb{R}^d$ starting at the origin, where B is the standard d-dimensional Brownian motion.

We take $H = 1/2$. The verification of Condition (G1) is contained in [26], p. 405 and (G2) follows by direct calculation. (G3) is a consequence of the fact that the Cameron-Martin space $\mathcal{H}^{\mathrm{OU}}$ associated with X is the same as the one for Brownian motion with a different but equivalent inner product (see the monograph by D. Stroock [61], Theorem 8.5.4).

Remark 4.5.3. The uniqueness of signature for the Ornstein-Uhlenbeck process is the direct consequence of the general result in [28], as it is the solution to a (hypo)elliptic stochastic differential equation (which we write as SDE hereafter).

4.5.3 The Brownian Bridge

Finally we consider the d-dimensional Brownian bridge

$$X_t = B_t - tB_1, \ t \in [0,1].$$

In this case we also take $H = 1/2$. Similar to the case of the Ornstein-Uhlenbeck process, (G1) and (G2) follows quite easily by direct calculations. However, (G3) is not satisfied as the Cameron-Martin space $\mathcal{H}^{\mathrm{Bridge}}$ associated with X is the one for Brownian motion with vanishing terminal condition: $h_1 = 0$ (see [61], pp. 334–335). Of course the embedding $\mathcal{H}^{\mathrm{Bridge}} \subset C^{q-var}([0,1];\mathbb{R}^d)$ still holds for any $q > 1$.

The main problem in the verification of Assumption (C) is that in the explicit construction of our Cameron-Martin path, the h given by (4.5.6) may not satisfy

$h_1 = 0$. However, it is just a technical issue to overcome such a difficulty.

Recall that we want to show $DF(x) \neq 0$ for $x \in A_{s,t}^{H_{x_0,\eta}}$, where $F = \int_s^t \phi(dx_u)$ and ϕ is the differential one form given by (4.5.2). From our proof before it is easy to see that everything follows in the same way if $t < 1$, since we can always modify h^i on $((t+1)/2, 1]$ so that $h_1^i = 0$ and the value of $\langle DF(x), h\rangle$ does not change as it depends only on the value of h on $[s,t]$. Therefore, we only need to consider the case when $t = 1$.

On the path space W let $x \in A_{s,t}^{H_z^{\varepsilon,\delta}}$ and take $\varepsilon > 0$ such that $x|_{[1-\varepsilon,1]} \subset H_0^{\varepsilon,\delta}$ (this is possible since $x_1 = 0$). Define ϕ by (4.5.2) for the open cube $H_z^{\varepsilon,\delta}$, and define $Y_{u,j}$ by (4.5.5). Now we need to consider two cases.

(1) If $z \neq 0$, then

$$\phi_i(x_1) + Y_{1,i} - Y_{v,i} = 0, \ \forall v \in [1-\varepsilon, 1],$$

since ϕ is supported on the closure of $H_z^{\varepsilon,\delta}$. Therefore, for any $h \in \mathcal{H}$,

$$\langle DF(x), h\rangle = \int_s^{1-\varepsilon} (\phi_i(x_1) + Y_{1,i} - Y_{v,i})\, dh_u^i.$$

To apply our previous argument, we just define h by (4.5.6) but modified on $(1-\varepsilon/2, 1]$ so that $h_1^i = 0$, and the resulting h is an element in $\mathcal{H}^{\text{Bridge}}$. By making use of Remark 4.5.1, the proof follows easily in the same way.

(2) If $z = 0$, based on our argument before, for any $\psi^i \in C^1([1-\varepsilon, 1])$ ($i = 1, \cdots, d$) with

$$\psi_{1-\varepsilon}^i = C_i := \int_s^{1-\varepsilon} (\phi_i(x_1) + Y_{1,i} - Y_{v,i})\, dv$$

and $\psi_1^i = 0$, the function

$$h_u^i = \begin{cases} \int_s^u (\phi_i(x_1) + Y_{1,i} - Y_{v,i})\, dv, & u \in [0, 1-\varepsilon]; \\ \psi_u^i, & u \in [1-\varepsilon, 1], \end{cases} \tag{4.5.8}$$

defines an element $h \in \mathcal{H}^{\text{Bridge}}$. It follows that

$$\begin{aligned} \langle DF(x), h\rangle &= \sum_{i=1}^d \int_s^{1-\varepsilon} (\phi_i(x_1) + Y_{1,i} - Y_{v,i})^2\, dv \\ &\quad + \sum_{i=1}^d \int_{1-\varepsilon}^1 (\phi_i(x_1) + Y_{1,i} - Y_{v,i})\, d\psi_v^i. \end{aligned}$$

Now we take ψ^i of the form

$$\psi_u^i = C_i - \int_{1-\varepsilon}^{u} \xi_v^i dv, \ u \in [1-\varepsilon, 1],$$

where $\xi^i \in C([1-\varepsilon, 1])$ with $\int_{1-\varepsilon}^{1} \xi_v^i dv = C_i$. If $\langle DF(x), h \rangle = 0$, then we have

$$\sum_{i=1}^{d} \int_{s}^{1-\varepsilon} (\phi_i(x_1) + Y_{1,i} - Y_{v,i})^2 \, dv - \sum_{i=1}^{d} \int_{1-\varepsilon}^{1} (\phi_i(x_1) + Y_{1,i} - Y_{v,i}) \, \xi_v^i dv = 0.$$

It follows that for any $\zeta^i \in C([1-\varepsilon, 1])$ with $\int_{1-\varepsilon}^{1} \zeta_v^i dv = 0$, we have

$$\sum_{i=1}^{d} \int_{1-\varepsilon}^{1} (\phi_i(x_1) + Y_{1,i} - Y_{v,i}) \, \zeta_v^i dv = 0,$$

which by an elementary argument implies that

$$\phi_i(x_1) + Y_{1,i} - Y_{v,i} = \text{const.}, \ \forall v \in [1-\varepsilon, 1] \text{ and } 1 \leqslant i \leqslant d.$$

It follows from taking $i = 2$ that

$$\int_{u}^{v} \frac{\partial \phi_1}{\partial x^2}(x_r) dx_r^1 = 0, \ \forall [u, v] \subset [1-\varepsilon, 1].$$

Now the proof follows again by making use of Remark 4.5.1 and the fact that $x|_{[1-\varepsilon,1]} \subset H_0^{\varepsilon,\delta}$.

Remark 4.5.4. By the same argument with a technical modification of ψ so that the h defined by (4.5.8) is regular enough to lie in the Cameron-Martin space, the result holds for general Gaussian bridge processes

$$X_t = G_t - tG_1, \ t \in [0, 1],$$

as long as the underlying Gaussian process G itself satisfies conditions (G1), (G2) and (G3).

Chapter 5

G-Brownian Motion as Rough Paths and Differential Equations Driven by G-Brownian Motion

5.1 Introduction

The classical Feynman-Kac formula (see for example M. Kac [39] or the monograph by I.A. Karatzas and S.E. Shreve [40]) provides us with a way to represent the solution to a linear parabolic partial differential equation (PDE hereafter) in terms of the conditional expectation of a certain functional of a diffusion process. However, it works only for the linear case, mainly due to the linear nature of diffusion processes. To understand nonlinear parabolic PDEs from a probabilistic point of view, S. Peng and E. Pardoux initiated the study of backward stochastic differential equations (BSDEs hereafter) in a series of important works [52], [53], [54]. In particular, they showed that the solutions to a certain type of quasilinear parabolic PDE can be expressed in terms of the solutions to BSDEs. This result suggests that BSDEs reveal a certain type of nonlinear dynamics, and this was made explicit by S. Peng [55]. More precisely, S. Peng introduced a notion of nonlinear expectation called the g-expectation in terms of the solutions to BSDEs which is filtration consistent. However, it was developed under the framework of the classical Itô calculus and did not investigate the fully nonlinear situation.

Motivated by the study of fully nonlinear dynamics, S. Peng [56] introduced the notion of G-expectation in an intrinsic way which does not rely on any particular probability space. It reveals the probability distribution uncertainty in a fundamental way which is useful in many practical situations. The underlying mechanism corresponding to this type of uncertainty is a fully nonlinear parabolic PDE. In [56], [57], he also introduced the concept of G-Brownian motion which is generated by the so-called nonlinear G-heat equation, and related stochastic calculus such as G-Itô's integrals, G-Itô's formula, SDEs driven by G-Brownian motion. One of the major contributions of this theory is the corresponding nonlinear Feynman-Kac formula proved by S. Peng [58], which gives us a way to represent the solution to a fully nonlinear parabolic PDE in terms of the solution to a forward-backward SDE under the framework of G-expectation.

The case of classical Brownian motion is special, since we have a complete SDE theory in the L^2-sense, as well as the notion of Stratonovich type integrals and differential equations. The fundamental relation between the two types of stochastic differentials (one-dimensional case) can be expressed by

$$X \circ dY = XdY + \frac{1}{2}dX \cdot dY.$$

It is proved in rough path theory (see [26], [47], and also [37], [66] from the view of the Wong-Zakai type approximation) that the Stratonovich type integrals and differential equations are equivalent to path integrals and RDEs in the sense of rough paths. In other words, the following two types of differential equation driven by Brownian motion

$$\begin{aligned} dX_t &= \sum_{\alpha=1}^{d} V_\alpha(X_t)\, dW_t^\alpha + b(X_t)\, dt, \quad \text{(Itô type SDE)} \\ dY_t &= \sum_{\alpha=1}^{d} V_\alpha(Y_t)\, dW_t^\alpha + \left(b(Y_t) - \sum_{\alpha=1}^{d} \frac{1}{2} DV_\alpha(Y_t) \cdot V_\alpha(Y_t) \right) dt, \quad \text{(RDE)} \end{aligned}$$

which are both well-defined under some regularity assumptions on the generating vector fields, are equivalent in the sense that if their solutions X_t and Y_t satisfy $X_0 = Y_0$, then $X = Y$ almost surely.

Under the framework of G-expectation, SDEs driven by G-Brownian motion introduced by S. Peng, can be regarded as nonlinear diffusion processes in Euclidean space. The idea of constructing G-Itô's integrals and SDEs driven by G-Brownian motion is similar to the classical Itô calculus, which is also an L^2-theory but under the G-expectation instead of a probability measure. What is missing is the notion of the Stratonovich type integral, mainly due to the reason that the theory of G-martingales is still not well understood. In particular, we don't have the corresponding nonlinear

Doob-Meyer type decomposition theorem and the notion of quadratic variation processes for G-martingales. However, by the key observation in the classical case that the Stratonovich type integrals and rough path integrals are essentially equivalent in the sense of rough paths, we can start with the study of sample path regularity of G-Brownian motion from the view of rough path theory. In particular, it may lead us to a complete pathwise theory of SDEs driven by G-Brownian motion in the sense of rough paths (namely RDEs). The basic language for describing path structure under G-expectation is quasi sure analysis and related capacity theory, which was developed by L. Denis, M. Hu and S. Peng [20]. In particular, they generalized the Kolmogorov continuity theorem and studied sample path properties of G-Brownian motion. They also studied the relation between G-expectation and upper expectation associated to a family of probability measures which defines a Choquet capacity and the relation between the corresponding two types of L^p-spaces. The pathwise properties and homeomorphic flows for SDEs driven by G-Brownian motion in the quasi sure setting were then studied by F. Gao [27].

The main motivation of this chapter is to study nonlinear diffusions and their associated intrinsic "distributions" on manifolds from the view of rough paths. In particular, as the ultimate goal we are interested in constructing G-Brownian motion on a Riemannian manifold and establishing its generating nonlinear heat flow.

This chapter is organized in the following way. Section 5.2 is a brief review on G-expectation and related stochastic calculus. In Section 5.3 we develop the Euler-Maruyama scheme for SDEs driven by G-Brownian motion. In Section 5.4, by using techniques in rough path theory, we show that quasi surely the sample paths of G-Brownian motion can be lifted to the second level in a canonical way so that they become geometric p-rough paths for $2 < p < 3$. In Section 5.5 we establish the fundamental relation between SDEs and RDEs driven by G-Brownian motion by using the rough Taylor expansion. In Section 5.6 we introduce the notion of SDEs on a differentiable manifold driven by G-Brownian motion from the RDE point of view by using a pathwise localization technique. In Section 5.7, we study the infinitesimal diffusive nature and the generating PDE for nonlinear diffusion processes in a (Riemannian) geometric setting. In particular, from the view of J. Eells, K.D. Elworthy and P. Malliavin, it leads us to the construction of G-Brownian motion on a compact Riemannian manifold and the generating nonlinear heat flow for a wide and interesting class of G-functions whose invariant group (to be introduced in that section) is the orthogonal group. As a consequence we also construct the canonical G-expectation on the path space over the manifold.

Throughout the rest of this chapter, for simplicity we again use the Einstein convention of summation.

5.2 Preliminaries on G-expectation and Related Stochastic Calculus

In this section we recall the basic notions of G-expectation and related stochastic calculus. The contents of this section are based on the monograph by S. Peng [58].

Let Ω be a non-empty set, and $\mathcal{H}$ be a vector space of functionals on Ω such that $\mathcal{H}$ contains all constant functionals and for any $X_1, \cdots, X_n \in \mathcal{H}$ and any $\varphi \in C_{l,Lip}(\mathbb{R}^n)$,

$$\varphi(X_1, \cdots, X_n) \in \mathcal{H},$$

where $C_{l,Lip}(\mathbb{R}^n)$ denotes the space of functions φ on $\mathbb{R}^n$ satisfying

$$|\varphi(x) - \varphi(y)| \leqslant C(1 + |x|^m + |y|^m)(|x - y|), \ \forall x, y \in \mathbb{R}^n,$$

for some constant $C > 0$ and $m \in \mathbb{N}$ depending on φ. $\mathcal{H}$ can be regarded as the space of random variables.

Definition 5.2.1. A *sublinear expectation* $\mathbb{E}$ on $(\Omega, \mathcal{H})$ is a functional $\mathbb{E}: \ \mathcal{H} \to \mathbb{R}$ such that

(1) if $X \leqslant Y$, then $\mathbb{E}[X] \leqslant \mathbb{E}[Y]$;
(2) for any constant c, $\mathbb{E}[c] = c$;
(3) for any $X, Y \in \mathcal{H}$, $\mathbb{E}[X + Y] \leqslant \mathbb{E}[X] + \mathbb{E}[Y]$;
(4) for any $\lambda \geqslant 0$ and $X \in \mathcal{H}$, $\mathbb{E}[\lambda X] = \lambda \mathbb{E}[X]$.

The triple $(\Omega, \mathcal{H}, \mathbb{E})$ is called a *sublinear expectation space.*

The relation between sublinear expectations and linear expectations, which was proved by S. Peng [58], is contained in the following representation theorem.

Theorem 5.2.1. *Let $(\Omega, \mathcal{H}, \mathbb{E})$ be a sublinear expectation space. Then there exists a family of linear expectations (linear functionals) $\{\mathbb{E}_\theta : \theta \in \Theta\}$ on $\mathcal{H}$, such that*

$$\mathbb{E}[X] = \sup_{\theta \in \Theta} \mathbb{E}_\theta[X], \ \forall X \in \mathcal{H}.$$

Under the framework of a sublinear expectation space, we also have the notion of independence and distribution (law).

Definition 5.2.2. (1) A random vector $Y \in \mathcal{H}^n$ is said to be *independent from* another random vector $X \in \mathcal{H}^m$ under the sublinear expectation $\mathbb{E}$, if for any $\varphi \in C_{l,Lip}(\mathbb{R}^m \times \mathbb{R}^n)$,

$$\mathbb{E}[\varphi(X,Y)] = \mathbb{E}\left[\mathbb{E}[\varphi(x,Y)]_{x=X}\right].$$

(2) Given a random vector $X \in \mathcal{H}^n$, the *distribution* (or the law) of X is defined as the functional $\mathbb{F}_X : C_{l,Lip}(\mathbb{R}^n) \to \mathbb{R}$ given by

$$\mathbb{F}_X[\varphi] = \mathbb{E}[\varphi(X)], \ \varphi \in C_{l,Lip}(\mathbb{R}^n).$$

This is a sublinear expectation on $(\mathbb{R}^n, C_{l,Lip}(\mathbb{R}^n))$. By saying that two random vectors X, Y (possibly defined on different sublinear expectation spaces) are *identically distributed*, we mean that their distributions are the same.

Now we introduce the notion of G-distribution, which is the generalization of degenerate distributions and normal distributions. It captures the uncertainty of probability distributions and plays a fundamental role in the theory of sublinear expectation.

Let $S(d)$ be the space of $d \times d$ symmetric matrices, and let $G : \mathbb{R}^d \times S(d) \to \mathbb{R}$ be a continuous and sublinear function monotonic in $S(d)$ in the sense that:

(1) $G(p+\bar{p}, A+\bar{A}) \leqslant G(p,A) + G(\bar{p},\bar{A}), \ \forall p,\bar{p} \in \mathbb{R}^d, \ A,\bar{A} \in S(d)$;

(2) $G(\lambda p, \lambda A) = \lambda G(p,A), \ \forall \lambda \geqslant 0$;

(3) $G(p,A) \leqslant G(p,\bar{A}), \ \forall A \leqslant \bar{A}$.

Definition 5.2.3. Let $X, \eta \in \mathcal{H}^d$ be two random vectors. (X,η) is called G-*distributed* if for any $\varphi \in C_{l,Lip}(\mathbb{R}^d \times \mathbb{R}^d)$, the function

$$u(t,x,y) := \mathbb{E}[\varphi(x+\sqrt{t}X, y+t\eta)], \ (t,x,y) \in [0,\infty) \times \mathbb{R}^d \times \mathbb{R}^d,$$

is a viscosity solution to the following parabolic PDE (called a G-heat equation):

$$\partial_t u - G\left(D_y u, D_x^2 u\right) = 0, \tag{5.2.1}$$

with Cauchy condition $u|_{t=0} = \varphi$.

Remark 5.2.1. From the general theory of viscosity solutions (see for example M.G. Crandall, H. Ishii and P.L. Lions [15] or [58]), the G-heat equation (5.2.1) has a unique viscosity solution. By solving the G-heat equation (5.2.1) (in some special cases, it is explicitly solvable), we can compute the sublinear expectation of some functionals of a G-distributed random vector. The case of convex functionals, for instance, the power function $|x|^k$, is quite interesting.

It can be proved (see [58]) that for such a function G, there exists a bounded, closed and convex subset $\Gamma \subset \mathbb{R}^d \times \mathbb{R}^{d\times d}$, such that G has the following representation:

$$G(p,A) = \sup_{(q,Q)\in\Gamma} \left\{ \frac{1}{2}\mathrm{tr}\big(AQQ^T\big) + \langle p,q\rangle \right\}, \ \forall (p,A) \in \mathbb{R}^d \times S(d).$$

The set Γ captures the uncertainty of probability distribution (mean uncertainty and variance uncertainty) of a G-distributed random vector.

In particular, if G only depends on $p \in \mathbb{R}^d$, then there exists some bounded, closed and convex subset $\Lambda \subset \mathbb{R}^d$, such that

$$G(p) = \sup_{q\in\Lambda} \langle p,q\rangle.$$

In this case a G-distributed random vector η is called *maximal distributed* and is denoted by $\eta \sim N(\Lambda, \{0\})$. Similarly, if G only depends on $A \in S(d)$, then there exists some bounded, closed and convex subset $\Sigma \subset S_+(d)$ (the space of symmetric and non-negative definite matrices) such that

$$G(A) = \frac{1}{2}\sup_{B\in\Sigma} \mathrm{tr}(AB), \ \forall A \in S(d). \tag{5.2.2}$$

A G-distributed random vector X for such G is called *G-normal distributed* and is denoted by $X \sim N(\{0\}, \Sigma)$.

Now we introduce the concept of G-Brownian motion and related stochastic calculus.

From now on, let $G: \ S(d) \to \mathbb{R}$ be a function given by (5.2.2).

Definition 5.2.4. A d-dimensional process B_t is called a *G-Brownian motion* if

(1) $B_0(\omega) = 0, \ \forall \omega \in \Omega$;

(2) for each $s, t \geqslant 0$, $B_{t+s} - B_t \sim N(\{0\}, s\Sigma)$ and is independent from $(B_{t_1}, \cdots, B_{t_n})$ for any $n \geqslant 1$ and $0 \leqslant t_1 < \cdots < t_n \leqslant t$.

Similar to the classical situation, a G-Brownian motion can be constructed explicitly on the canonical path space by using independent G-normal random vectors. Here we omit the details and refer the reader to [58] for the construction.

In summary, let $\Omega = C_0([0,\infty);\mathbb{R}^d)$ be the space of $\mathbb{R}^d$-valued continuous paths starting at the origin, and let $B_t(\omega) := \omega_t$ be the coordinate process. For any $T \geqslant 0$, define

$$L_{ip}(\Omega_T) = \left\{\varphi(B_{t_1}, \cdots, B_{t_n}) : \ n \geqslant 1, t_1, \cdots, t_n \in [0, T], \varphi \in C_{l,Lip}(\mathbb{R}^{d\times n})\right\},$$

and

$$L_{ip}(\Omega) = \bigcup_{n=1}^{\infty} L_{ip}(\Omega_n).$$

Then on $(\Omega, L_{ip}(\Omega))$ we can define the canonical sublinear expectation $\mathbb{E}$ such that the coordinate process B_t becomes a G-Brownian motion, which is usually called the *G-expectation* and denoted by $\mathbb{E}^G$. $(\Omega, L_{ip}(\Omega), \mathbb{E}^G)$ is also called the *canonical G-expectation space*. Throughout the rest of this chapter, we restrict ourselves to the canonical G-expectation space and its completion (to be defined later on).

Now we introduce the notion of conditional G-expectation and related properties.

For

$$X = \varphi(B_{t_1}, B_{t_2} - B_{t_1}, \cdots, B_{t_n} - B_{t_{n-1}}) \in L_{ip}(\Omega),$$

where $0 \leqslant t_1 < t_2 < \cdots < t_n$, the *$G$-conditional expectation* of X given Ω_{t_j} is defined by

$$\mathbb{E}^G[X|\Omega_{t_j}] = \psi(B_{t_1}, B_{t_2} - B_{t_1}, \cdots, B_{t_j} - B_{t_{j-1}}),$$

where

$$\psi(x_1, \cdots, x_j) := \mathbb{E}^G[\varphi(x_1, \cdots, x_j, B_{t_{j+1}} - B_{t_j}, \cdots, B_{t_n} - B_{t_{n-1}})], \ x_1, \cdots, x_j \in \mathbb{R}^d.$$

The conditional G-expectation $\mathbb{E}^G[\cdot|\Omega_t]$ has the following basic properties: for any $X, Y \in L_{ip}(\Omega)$,

(1) if $X \leqslant Y$, then $\mathbb{E}^G[X|\Omega_t] \leqslant \mathbb{E}^G[Y|\Omega_t]$;

(2) $\mathbb{E}^G[X + Y|\Omega_t] \leqslant \mathbb{E}^G[X|\Omega_t] + \mathbb{E}^G[Y|\Omega_t]$;

(3) for any $\eta \in L_{ip}(\Omega_t)$,

$$\begin{aligned}\mathbb{E}^G[\eta|\Omega_t] &= \eta,\\ \mathbb{E}^G[\eta X|\Omega_t] &= \eta^+\mathbb{E}^G[X|\Omega_t] + \eta^-\mathbb{E}[-X|\Omega_t],\end{aligned}$$

where η^+, η^- denote the positive and negative parts of η respectively;

(4) $\mathbb{E}^G[\mathbb{E}^G[X|\Omega_t]|\Omega_s] = \mathbb{E}^G[X|\Omega_{t\wedge s}]$. In particular, $\mathbb{E}^G[\mathbb{E}^G[X|\Omega_t]] = \mathbb{E}^G[X]$.

For any $p \geqslant 1$, let L^p_G (respectively, $L^p_G(\Omega_t)$)) be the completion of $L_{ip}(\Omega)$ (respectively, $L_{ip}(\Omega_t)$) under the semi-norm $\|X\|_p := \left(\mathbb{E}^G\left[|X|^p\right]\right)^{\frac{1}{p}}$. Then $\mathbb{E}^G$ can be continuously extended to a sublinear expectation on $L^p_G(\Omega)$ (respectively, $L^p_G(\Omega_t)$), which is still denoted by $\mathbb{E}^G$.

For $t < T \leqslant \infty$, the conditional G-expectation $\mathbb{E}^G[\cdot|\Omega_t]: \ L_{ip}(\Omega_T) \to L_{ip}(\Omega_t)$ is a continuous map under $\|\cdot\|_1$ and can be uniquely extended to a continuous map

$$\mathbb{E}^G[\cdot|\Omega_t]: \ L_G^1(\Omega_T) \to L_G^1(\Omega_t),$$

which can still be interpreted as the conditional G-expectation. It is easy to show that the properties (1) to (4) for the conditional G-expectation still hold true on $L_G^1(\Omega_T)$ as long as it is well-defined.

Now we introduce the related stochastic calculus for G-Brownian motion.

First of all, similar to the idea in the classical case, we still have the notion of Itô's integrals against a 1-dimensional G-Brownian motion. More precisely, in the one dimensional case we can first define Itô's integrals of simple processes and then pass to the limit under the G-expectation $\mathbb{E}^G$ in some suitable functional spaces. Let $M_G^{p,0}(0,T)$ be the space of simple processes $\eta_t(\omega)$ on $[0,T]$ of the form

$$\eta_t(\omega) = \sum_{k=1}^N \xi_{k-1}(\omega)\mathbf{1}_{[t_{k-1},t_k)}(t),$$

where $\pi_T^N := \{t_0, t_1, \cdots, t_N\}$ is a partition of $[0,T]$ and $\xi_k \in L_G^p(\Omega_{t_k})$, and introduce the semi-norm

$$\|\eta\|_{M_G^p(0,T)} = \left(\mathbb{E}^G\left[\int_0^T |\eta_t|^p \, dt\right]\right)^{\frac{1}{p}}$$

on $M_G^{p,0}(0,T)$. Let $M_G^p(0,T)$ be the completion of $M_G^{p,0}(0,T)$ under $\|\cdot\|_{M_G^p(0,T)}$. It is straight forward to define Itô's integrals $\int_0^T \eta_t dB_t$ of simple processes. Moreover, such an integral operator is linear and continuous under $\|\cdot\|_{M_G^p(0,T)}$ and hence can be extended to a bounded linear operator

$$\mathcal{I}: \ M_G^2(0,T) \to L_G^2(0,T).$$

The operator $\mathcal{I}$ is called the *Itô's integral operator* against a G-Brownian motion. For $0 \leqslant s < t \leqslant T$, define

$$\int_s^t \eta_u dB_u = \int_0^T \mathbf{1}_{[s,t]}(u)\eta_u dB_u.$$

We list some important properties of G-Itô's integrals in the following.

Proposition 5.2.1. *Let $\eta, \theta \in M_G^2(0,T)$ and let $0 \leqslant s \leqslant r \leqslant t \leqslant T$. Then*

(1)

$$\int_s^t \eta_u dB_u = \int_s^r \eta_u dB_u + \int_r^t \eta_u dB_u;$$

(2) if α is bounded in $L_G^1(\Omega_s)$, then

$$\int_s^t (\alpha\eta_u + \theta_u) dB_u = \alpha \int_s^t \eta_u dB_u + \int_s^t \theta_u dB_u;$$

(3) for any $X \in L_G^1(\Omega)$,

$$\mathbb{E}^G \left[X + \int_r^T \eta_u dB_u \middle| \Omega_s \right] = \mathbb{E}^G \left[X | \Omega_s \right];$$

(4)

$$\underline{\sigma}^2 \mathbb{E}^G \left[\int_0^T \eta_t^2 dt \right] \leqslant \mathbb{E}^G \left[\left(\int_0^T \eta_t dB_t \right)^2 \right] \leqslant \overline{\sigma}^2 \mathbb{E}^G \left[\int_0^T \eta_t^2 dt \right],$$

where $\overline{\sigma}^2 := \mathbb{E}^G[B_1^2]$ and $\underline{\sigma}^2 := -\mathbb{E}^G[-B_1^2]$.

Secondly, we have the notion of quadratic variation process of G-Brownian motion. In the case of 1-dimensional G-Brownian motion, the *quadratic variation process* $\langle B \rangle_t$ is defined as

$$\langle B \rangle_t = B_t^2 - 2 \int_0^t B_s dB_s,$$

which can be regarded as the L_G^2-limit of the sum $\sum_{j=1}^{k_N} \left(B_{t_j^N} - B_{t_{j-1}^N} \right)^2$ as $\mu\left(\pi_t^N\right) \to 0$, where $\pi_t^N := \{t_j^N\}_{j=0}^{k_N}$ is any finite partition of $[0,t]$ and

$$\mu\left(\pi_t^N\right) := \max\left\{ t_j^N - t_{j-1}^N : \ j = 1, 2, \cdots, k_N \right\}.$$

It follows that $\langle B \rangle_t$ is an increasing process with $\langle B \rangle_0 = 0$.

Similar to the definition of G-Itô's integrals, we can define the integration against $\langle B \rangle_t$ where B_t is a 1-dimensional G-Brownian motion. We refer the reader to [58] for a precise definition but we remark that the integral operator against $\langle B \rangle_t$ is a continuous linear map

$$Q_{0,T} : \ M_G^1(0,T) \to L_G^1(\Omega_T).$$

The following identity can be regarded as the G-Itô isometry.

Proposition 5.2.2. *Let $\eta \in M_G^2(0,T)$, then*

$$\mathbb{E}^G\left[\left(\int_0^T \eta_t dB_t\right)^2\right] = \mathbb{E}^G\left[\int_0^T \eta_t^2 d\langle B\rangle_t\right].$$

Now we consider the multi-dimensional case. Let B_t be a d-dimensional G-Brownian motion, and for any $v \in \mathbb{R}^d$, denote

$$B_t^v = \langle v, B_t\rangle,$$

where $\langle\cdot,\cdot\rangle$ is the Euclidean inner product. Then for $a, \overline{a} \in \mathbb{R}^d$, the *cross variation process* $\langle B^a, B^{\overline{a}}\rangle_t$ is defined as

$$\langle B^a, B^{\overline{a}}\rangle_t = \frac{1}{4}(\langle B^{a+\overline{a}}, B^{a+\overline{a}}\rangle_t - \langle B^{a-\overline{a}}, B^{a-\overline{a}}\rangle_t).$$

In the same way as the case of the quadratic variation process, we have

$$\begin{aligned}\langle B^a, B^{\overline{a}}\rangle_t &= (L_G^2-)\lim_{\mu(\pi_t^N)\to 0}\sum_{j=1}^{k_N}\left(B^a_{t_j^N} - B^a_{t_{j-1}^N}\right)\left(B^{\overline{a}}_{t_j^N} - B^{\overline{a}}_{t_{j-1}^N}\right)\\ &= B_t^a B_t^{\overline{a}} - \int_0^t B_s^a dB_s^{\overline{a}} - \int_0^t B_s^{\overline{a}} dB_s^a.\end{aligned}$$

Note that unlike the classical case, the cross variation process is not deterministic. The following result characterizes the distribution of the matrix-valued process $\langle B\rangle_t := (\langle B^\alpha, B^\beta\rangle_t)_{\alpha,\beta=1}^d$, where B_t is a d-dimensional G-Brownian motion and B_t^α is the α-th component of B_t. We refer the reader to [58] for the proof. Recall that the function G has the representation (5.2.2).

Proposition 5.2.3. $\langle B\rangle_t \sim N(t\Sigma, \{0\})$.

As in the classical case, we also have the important G-Itô formula under G-expectation, which takes a similar form to the classical one. The main difference is that $dB_t^\alpha \cdot dB_t^\beta$ should be $d\langle B^\alpha, B^\beta\rangle_t$ instead of $\delta_{\alpha\beta}dt$. We are not going to state the full result of G-Itô's formula here. See [58] for a detailed discussion.

Now we introduce the notion of SDEs driven by G-Brownian motion.

For $p \geqslant 1$, let $\overline{M}_G^p(0,T;\mathbb{R}^n)$ be the completion of $M_G^{p,0}(0,T;\mathbb{R}^n)$ under the norm

$$\|\eta\|_{\overline{M}_G^p(0,T;\mathbb{R}^n)} := \left(\int_0^T \mathbb{E}^G\left[|\eta_t|^p\right]dt\right)^{\frac{1}{p}}.$$

It is easy to see that $\overline{M}_G^p(0,T;\mathbb{R}^n) \subset M_G^p(0,T;\mathbb{R}^n)$.

Consider the following N-dimensional SDE driven by G-Brownian motion over $[0,T]$:

$$dX_t = b(t,X_t)dt + h_{\alpha\beta}(t,X_t)d\langle B^\alpha, B^\beta\rangle_t + V_\alpha(t,X_t)dB_t^\alpha \tag{5.2.3}$$

with initial condition $\xi \in \mathbb{R}^N$. Here we assume that the coefficients $b^i, h^i_{\alpha\beta}, V^i_\alpha$ are Lipschitz functions in the space variable, uniformly in time. A solution to (5.2.3) is a process in $\overline{M}_G^2(0,T;\mathbb{R}^N)$ satisfying the equation (5.2.3) in its integral form.

The existence and uniqueness of solutions to the SDE (5.2.3) is contained in the following result. We refer the reader to [58] for the proof.

Theorem 5.2.2. *There exists a unique solution $X \in \overline{M}_G^2(0,T;\mathbb{R}^N)$ to the SDE (5.2.3).*

Finally, we introduce the notion of quasi sure analysis for G-expectation. It plays an important role in studying pathwise properties of stochastic processes under the framework of G-expectation.

First of all, on the canonical sublinear expectation space $\left(\Omega, L_{ip}(\Omega), \mathbb{E}^G\right)$, we can prove a refinement of Theorem 5.2.1: there exists a weakly compact family $\mathcal{P}$ of probability measures on $(\Omega, \mathcal{B}(\Omega))$, such that for any $X \in L_{ip}(\Omega)$ and $P \in \mathcal{P}$, $\mathbb{E}_P[X]$ is well-defined and

$$\mathbb{E}^G[X] = \max_{P\in\mathcal{P}} \mathbb{E}_P[X], \quad \forall X \in L_{ip}(\Omega),$$

where "max" means that the supremum is attainable (for each X). Moreover, there is an explicit characterization of the family $\mathcal{P}$. Let G be represented in the following way:

$$G(A) = \frac{1}{2}\sup_{Q\in\Gamma} \mathrm{tr}\left(AQQ^T\right),$$

for some bounded, closed and convex subset $\Gamma \subset \mathbb{R}^{d\times d}$, and let $\mathcal{A}_\Gamma$ be the collection of all Γ-valued and $\{\mathcal{F}_t^W : t \geqslant 0\}$-adapted processes on $[0,\infty)$, where $\{\mathcal{F}_t^W : t \geqslant 0\}$ is the natural filtration of the coordinate process on Ω. Let $\mathcal{P}_0$ be the collection of probability laws of the following classical Itô integral processes with respect to the standard Wiener measure:

$$B_t^\gamma = \int_0^t \gamma_s dW_s,\ t \geqslant 0,\ \gamma \in A_\Gamma.$$

Then $\mathcal{P} = \overline{\mathcal{P}_0}$. We refer the reader to [20] for the proof of this result.

For any $\mathcal{B}(\Omega)$-measurable random variable X such that $\mathbb{E}_P[X]$ is well-defined for all $P \in \mathcal{P}$, define the upper expectation

$$\hat{\mathbb{E}}[X] = \sup_{P \in \mathcal{P}} \mathbb{E}_P[X].$$

Then we can prove that for any $0 \leqslant T \leqslant \infty$ and $X \in L^1_G(\Omega_T)$,

$$\mathbb{E}^G[X] = \hat{\mathbb{E}}[X].$$

See again [20] for a detailed discussion and other related properties.

For this particular family $\mathcal{P}$, define the set function c by

$$c(A) = \sup_{P \in \mathcal{P}} P(A), \;\; A \in \mathcal{B}(\Omega).$$

Then we have the following result.

Proposition 5.2.4. *The set function c is a Choquet capacity (for an introduction of capacity theory, see G. Choquet [12], C. Dellacherie [19]). In other words,*

(1) for any $A \in \mathcal{B}(\Omega)$, $0 \leqslant c(A) \leqslant 1$;

(2) if $A \subset B$, then $c(A) \leqslant c(B)$;

(3) if A_n is a sequence in $\mathcal{B}(\Omega)$, then $c\left(\bigcup_n A_n\right) \leqslant \sum_n c(A_n)$;

(4) if A_n is increasing in $\mathcal{B}(\Omega)$, then $c\left(\bigcup_n A_n\right) = \lim_{n\to\infty} c(A_n)$.

Definition 5.2.5. A property depending on $\omega \in \Omega$ is said to hold *quasi surely*, if it holds outside a $\mathcal{B}(\Omega)$-measurable subset of zero capacity.

We end this section by stating the following Markov inequality and Borel-Cantelli lemma under the capacity c, which are both crucial for our study. We refer the reader to [58] for the proof.

Theorem 5.2.3. *(1) For any $X \in L^p_G(\Omega)$ and $\lambda > 0$, we have*

$$c(|X| > \lambda) \leqslant \frac{\mathbb{E}^G[|X|^p]}{\lambda^p}.$$

(2) Let A_n be a sequence in $\mathcal{B}(\Omega)$ such that $\sum_{n=1}^{\infty} c(A_n) < \infty$. Then

$$c(\limsup A_n) = 0.$$

5.3 The Euler-Maruyama Approximation for SDEs Driven by G-Brownian Motion

In this section, we develop the Euler-Maruyama approximation scheme for SDEs driven by G-Brownian motion.

This result can be used to establish the Wong-Zakai type approximation which reveals the relationship between SDEs (in the sense of $L_G^2(\Omega;\mathbb{R}^N)$ by S. Peng) and RDEs driven by G-Brownian motion. In Section 5.5, the study of this relationship is our main focus. However, based on the result in the next section which reveals the rough path nature of G-Brownian motion, we use the rough Taylor expansion in the theory of RDEs instead of developing the Wong-Zakai type approximation to show that the solution to an SDE solves some associated RDE with a correction term in terms of the cross variation process of multidimensional G-Brownian motion. Such an approach reveals the natural of G-Brownian motion and differential equations in the sense of rough paths in a more essential way.

We also believe that there are other interesting applications of the Euler-Maruyama approximation, such as in numerical analysis under G-expectation, and in practical models under probability distribution uncertainty.

Consider the following N-dimensional SDE driven by the canonical d-dimensional G-Brownian motion over $[0,1]$ on the sublinear expectation space $(\Omega, L_G^2(\Omega), \mathbb{E}^G)$ which is the L_G^2-completion of the canonical path space $(\Omega, L_{ip}(\Omega), \mathbb{E}^G)$:

$$dX_t \quad = \quad b(X_t)dt + h_{\alpha\beta}(X_t)d\left\langle B^\alpha, B^\beta\right\rangle_t + V_\alpha(X_t)dB_t^\alpha, \tag{5.3.1}$$

with initial condition $X_0 = \xi \in \mathbb{R}^N$, where the coefficients $b^i, h^i_{\alpha\beta}, V^i_\alpha$ are bounded and uniformly Lipschitz. From Theorem 5.2.2 we know that the SDE has a unique solution.

The Euler-Maruyama approximation of the solution X_t to (5.3.1) is defined as follows. The underlying idea is similar to the classical situation.

For $n \geqslant 1$, consider the dyadic partition of the time interval $[0,1]$, i.e.

$$t_k^n = \frac{k}{2^n},\ k = 0, 1, \cdots, 2^n.$$

Define X_t^n to be the approximation of X_t in the following inductive way:

$$X_0^n \quad = \quad \xi,$$

and for $t \in [t^n_{k-1}, t^n_k]$,

$$(X^n_t)^i = (X^n_{k-1})^i + V^i_\alpha(X^n_{k-1})\Delta^n_k B^\alpha + b^i(X^n_{k-1})\Delta t^n + h^i_{\alpha\beta}(X^n_{k-1})\Delta^n_k \langle B^\alpha, B^\beta\rangle,$$

where

$$X^n_{k-1} := X^n_{t^n_{k-1}},\ \Delta^n_k B^\alpha := B^\alpha_{t^n_k} - B^\alpha_{t^n_{k-1}},$$

$$\Delta t^n := \frac{1}{2^n},\ \Delta^n_k\langle B^\alpha, B^\beta\rangle := \langle B^\alpha, B^\beta\rangle_{t^n_k} - \langle B^\alpha, B^\beta\rangle_{t^n_{k-1}}.$$

In this section, we prove that X^n_t converges to the solution X_t to (5.3.1) in $L^2_G(\Omega; \mathbb{R}^N)$ with convergence rate 1/2, which coincides with the classical case when B_t reduces to the classical Brownian motion.

First of all, we need the following lemma.

Lemma 5.3.1. *Let η_t be a bounded process in $M^2_G(0,1)$. Then for any $v \in \mathbb{R}^d$, $0 \leqslant s < t \leqslant 1$,*

$$\mathbb{E}^G\left[\left(\int_s^t \eta_u d\langle B^v\rangle_u\right)^2\right] \leqslant \overline{\sigma}^2_v (t-s)\mathbb{E}^G\left[\int_s^t \eta^2_u d\langle B^v\rangle_u\right],$$

where $\overline{\sigma}^2_v := 2G(v \cdot v^T)$ and $B^v := \langle v, B\rangle$, in which $\langle\cdot,\cdot\rangle$ denotes the Euclidean inner product of $\mathbb{R}^d$.

Proof. By approximation, it suffices to consider

$$\eta_u = \sum_{j=1}^k \zeta_{j-1} 1_{[u_{j-1}, u_j)},$$

where $s = u_0 < u_1 < \cdots < u_k = t$ and $\zeta_j \in L_{ip}(\Omega_{u_j})$ are bounded. In this case, by definition

$$\int_s^t \eta_u d\langle B^v\rangle_u = \sum_{j=1}^k \zeta_{j-1}(\langle B^v\rangle_{u_j} - \langle B^v\rangle_{u_{j-1}}),$$

and

$$\int_s^t \eta^2_u d\langle B^v\rangle_u = \sum_{j=1}^k \zeta^2_{j-1}\left(\langle B^v\rangle_{u_j} - \langle B^v\rangle_{u_{j-1}}\right),$$

which are both defined in the pathwise sense for step functions. Since $\langle B^v\rangle$ is increasing, the Cauchy-Schwarz inequality yields that

$$\left(\int_s^t \eta_u d\langle B^v\rangle_u\right)^2 \leqslant (\langle B^v\rangle_t - \langle B^v\rangle_s)\cdot \int_s^t \eta^2_u d\langle B^v\rangle_u.$$

Since the ζ_j are bounded, if we use M to denote an upper bound of η_u^2, it follows that for any $c \geqslant \overline{\sigma}_v^2$,

$$\begin{aligned}&\left(\int_s^t \eta_u d\langle B^v\rangle_u\right)^2\\ &\leqslant M\left(\langle B^v\rangle_t - \langle B^v\rangle_s - c(t-s)\right)^+\left(\langle B^v\rangle_t - \langle B^v\rangle_s\right) + c(t-s)\int_s^t \eta_u^2 d\langle B^v\rangle_u.\end{aligned}$$

Let $\varphi(x) = (x - c(t-s))^+ x$. Since $\langle B^v\rangle_t - \langle B^v\rangle_s$ is $N((t-s)[\underline{\sigma}_v^2, \overline{\sigma}_v^2] \times \{0\})$-distributed, it follows that

$$\begin{aligned}\mathbb{E}^G\left[\varphi(\langle B^v\rangle_t - \langle B^v\rangle_s)\right] &= \sup_{\underline{\sigma}_v^2 \leqslant x \leqslant \overline{\sigma}_v^2} \varphi(x(t-s))\\ &= (t-s)^2 \sup_{\underline{\sigma}_v^2 \leqslant x \leqslant \overline{\sigma}_v^2} (x-c)^+ x\\ &= 0.\end{aligned}$$

Therefore, by the sub-linearity of G, we have

$$\mathbb{E}^G\left[\left(\int_s^t \eta_u d\langle B^v\rangle_u\right)^2\right] \leqslant c(t-s)\mathbb{E}^G\left[\int_s^t \eta_u^2 d\langle B^v\rangle_u\right], \quad c \geqslant \overline{\sigma}_v^2.$$

Now the proof is complete. □

Now we are in position to state and prove our main result of this section.

Theorem 5.3.1. *We have the following error estimate for the Euler-Maruyama approximation:*

$$\sup_{t\in[0,1]} \mathbb{E}^G\left[|X_t^n - X_t|^2\right] \leqslant C\Delta t^n,$$

where C is some positive constant only depending on d, N, G and the coefficients of (5.3.1). In particular,

$$\lim_{n\to\infty} \sup_{t\in[0,1]} \mathbb{E}^G\left[|X_t^n - X_t|^2\right] = 0.$$

Proof. For $t \in [t_{k-1}^n, t_k^n]$, by construction we have

$$X_t^i - (X_t^n)^i = I_1^i + J_1^i + K_1^i + I_2^i + J_2^i + K_2^i,$$

where

$$\begin{aligned}
I_1^i &= \sum_{l=1}^{k-1}\int_{t_{l-1}^n}^{t_l^n}\left(V_\alpha^i(X_s)-V_\alpha^i(X_s^n)\right)dB_s^\alpha+\int_{t_{k-1}^n}^{t}\left(V_\alpha^i(X_s)-V_\alpha^i(X_s^n)\right)dB_s^\alpha,\\
J_1^i &= \sum_{l=1}^{k-1}\int_{t_{l-1}^n}^{t_l^n}\left(b^i(X_s)-b^i(X_s^n)\right)ds+\int_{t_{k-1}^n}^{t}\left(b^i(X_s)-b^i(X_s^n)\right)ds,\\
K_1^i &= \sum_{l=1}^{k-1}\int_{t_{l-1}^n}^{t_l^n}\left(h_{\alpha\beta}^i(X_s)-h_{\alpha\beta}^i(X_s^n)\right)d\left\langle B^\alpha,B^\beta\right\rangle_s\\
&\quad+\int_{t_{k-1}^n}^{t}\left(h_{\alpha\beta}^i(X_s)-h_{\alpha\beta}^i(X_s^n)\right)d\left\langle B^\alpha,B^\beta\right\rangle_s,\\
I_2^i &= \sum_{l=1}^{k-1}\int_{t_{l-1}^n}^{t_l^n}\left(V_\alpha^i(X_s^n)-V_\alpha^i(X_{l-1}^n)\right)dB_s^\alpha+\int_{t_{k-1}^n}^{t}\left(V_\alpha^i(X_s)-V_\alpha^i(X_{l-1}^n)\right)dB_s^\alpha,\\
J_2^i &= \sum_{l=1}^{k-1}\int_{t_{l-1}^n}^{t_l^n}\left(b^i(X_s^n)-b^i(X_{l-1}^n)\right)ds+\int_{t_{k-1}^n}^{t}\left(b^i(X_s)-b^i(X_{l-1}^n)\right)ds,\\
K_2^i &= \sum_{l=1}^{k-1}\int_{t_{l-1}^n}^{t_l^n}\left(h_{\alpha\beta}^i(X_s)-h_{\alpha\beta}^i(X_s^n)\right)d\left\langle B^\alpha,B^\beta\right\rangle_s\\
&\quad+\int_{t_{k-1}^n}^{t}\left(h_{\alpha\beta}^i(X_s)-h_{\alpha\beta}^i(X_s^n)\right)d\left\langle B^\alpha,B^\beta\right\rangle_s.
\end{aligned}$$

It follows that

$$\left(X_t^i-(X_t^n)^i\right)^2\leqslant 6\left((I_1^i)^2+(J_1^i)^2+(K_1^i)^2+(I_2^i)^2+(J_2^i)^2+(K_2^i)^2\right). \tag{5.3.2}$$

Throughout the rest of this section, we always use the same notation C to denote constants only depending on d, N, G and the coefficients of (5.3.1), although they may be different from line to line.

The following estimates are important for further development.

(1) From the G-Itô isometry, the distribution of $\langle B^\alpha\rangle$ and the Lipschitz property, we have,

$$\mathbb{E}^G\left[\left(\int_{t_{l-1}^n}^{u}(V_\alpha^i(X_s)-V_\alpha^i(X_s^n))dB_s^\alpha\right)^2\right]\leqslant C\int_{t_{l-1}^n}^{u}\mathbb{E}^G\left[|X_s-X_s^n|^2\right]ds,\ \forall u\in[t_{l-1}^n,t_l^n].$$

(2) Similarly, by the Cauchy-Schwarz inequality, we have

$$\begin{aligned}&\mathbb{E}^G\left[\left(\int_{t_{l-1}^n}^u \left(b^i(X_s)-b^i(X_s^n)\right)ds\right)^2\right]\\ &\leqslant C(u-t_{l-1}^n)\int_{t_{l-1}^n}^u \mathbb{E}^G\left[|X_s-X_s^n|^2\right]ds,\end{aligned}$$

for all $u\in[t_{l-1}^n,t_l^n]$. By the definition of $\langle B^\alpha,B^\beta\rangle$ and Lemma 5.3.1, we also have

$$\begin{aligned}&\mathbb{E}^G\left[\left(\int_{t_{l-1}^n}^u \left(h_{\alpha\beta}^i(X_s)-h_{\alpha\beta}^i(X_s^n)\right)d\langle B^\alpha,B^\beta\rangle_s\right)^2\right]\\ &\leqslant C(u-t_{l-1}^n)\int_{t_{l-1}^n}^u \mathbb{E}^G\left[|X_s-X_s^n|^2\right]ds,\end{aligned}$$

for all $u\in[t_{l-1}^n,t_l^n]$.

(3) By construction and similar arguments to (1), (2), we have

$$\begin{aligned}\mathbb{E}^G\left[\left(\int_{t_{l-1}^n}^u (V_\alpha^i(X_s^n)-V_\alpha^i(X_{l-1}^n))dB_s^\alpha\right)^2\right] &\leqslant C(u-t_{l-1}^n)^2,\\ \mathbb{E}^G\left[\left(\int_{t_{l-1}^n}^u \left(b^i(X_s^n)-b^i(X_{l-1}^n)\right)ds\right)^2\right] &\leqslant C(u-t_{l-1}^n)^3,\\ \mathbb{E}^G\left[\left(\int_{t_{l-1}^n}^u \left(h_{\alpha\beta}^i(X_s^n)-h_{\alpha\beta}^i(X_s^n)\right)d\langle B^\alpha,B^\beta\rangle_s\right)^2\right] &\leqslant C(u-t_{l-1}^n)^3,\end{aligned}$$

for all $u\in[t_{l-1}^n,t_l^n]$.

(4) By conditioning and from the properties of Itô's integrals against G-Brownian motion, we know that the G-expectation of each "cross term" in $(I_1^i)^2$ and in $(I_2^i)^2$ is zero.

Combining (1) to (4) and applying the following elementary inequality to $(J_1^i)^2$, $(J_2^i)^2$, $(K_1^i)^2$ and $(K_2^i)^2$:

$$(a_1+\cdots+a_m)^2\leqslant m(a_1^2+\cdots+a_m^2),$$

it is not hard to obtain that

$$\mathbb{E}^G\left[\|X_t - X_t^n\|^2\right] \leqslant C\int_0^t \mathbb{E}^G\left[\|X_s - X_s^n\|^2\right] ds + C(\Delta t^n),\ \forall t \in [0,1].$$

By using Gronwall's inequality, we arrive at

$$\mathbb{E}^G\left[\|X_t - X_t^n\|^2\right] \leqslant C(\Delta t^n),$$

which completes the proof of the theorem. □

5.4 G-Brownian Motion as Rough Paths and RDEs Driven by G-Brownian Motion

In this section, we study the geometric rough path nature for sample paths of G-Brownian motion. More precisely, we show that: on the canonical path space, outside a Borel-measurable set of capacity zero, the sample paths of G-Brownian motion can be lifted canonically as geometric p-rough paths for $2 < p < 3$. As pointed out before, such a result enables us to make sense of RDEs driven by G-Brownian motion in the pathwise sense.

Recall that $(\Omega, L_{ip}(\Omega), \mathbb{E}^G)$ is the canonical path space associated with the function G, on which the coordinate process

$$B_t(\omega) := \omega_t,\ t \in [0,1],$$

is a d-dimensional G-Brownian motion with continuous sample paths.

By the following moment inequality for B_t:

$$\mathbb{E}^G\left[|B_t - B_s|^{2q}\right] \leqslant C_q(t-s)^q,\ \forall 0 \leqslant s < t \leqslant 1,\ q > 1, \tag{5.4.1}$$

and the generalized Kolmogorov criterion (see [58] for details), we know that quasi-surely, the sample paths of B_t are α-Hölder continuous for any $\alpha \in (0, \frac{1}{2})$. Therefore, if the sample paths of B_t can be regarded as geometric rough paths, the exact roughness should be $2 < p < 3$. The situation here is the same as the classical Brownian motion, and the fundamental reason lies in the distribution of B_t (or more precisely, the moment inequality (5.4.1)), which yields the same kind of Hölder continuity for sample paths of B_t as the classical case.

From now on, we assume that $p \in (2,3)$ is some fixed constant.

As in the last section, for $n \geqslant 1$, $k = 0, 1, \cdots, 2^n$, let $t_k^n = \frac{k}{2^n}$ be the dyadic partition of $[0,1]$, and let B_t^n be the piecewise linear interpolation of B_t over the partition points $\{t_0^n, t_1^n, \cdots, t_{2^n}^n\}$. Since the sample paths of B_t^n have bounded total variation, B_t^n has a unique lifting

$$\boldsymbol{B}_{s,t}^n = \left(1, B_{s,t}^{n,1}, B_{s,t}^{n,2}\right), \ 0 \leqslant s < t \leqslant 1,$$

to the space $G\Omega_p(\mathbb{R}^d)$ of geometric p-rough paths (in fact, for any $p \geqslant 1$) determined by iterated path integrals.

Our goal is to show that quasi-surely, $\boldsymbol{B}^n$ is a Cauchy sequence under the p-variation metric d_p. It follows that quasi-surely, the sample paths of B_t can be lifted as geometric p-rough paths, which are defined as limits of $\boldsymbol{B}^n$ under d_p. Such a lifting via dyadic piecewise linear interpolation can be regarded as a canonical lifting.

Throughout the rest of this section, we use $\|\cdot\|_q$ to denote the L^q-norm under the G-expectation $\mathbb{E}^G$. Moreover, we use the same notation C to denote constants only depending on d, G, p, although they may be different from line to line.

The following estimates are crucial for the proof of the main result of this section.

Lemma 5.4.1. *Let $m, n \geqslant 1$, and $k = 1, 2, \cdots, 2^n$. Then*

(1)

$$\left\| B_{t_{k-1}^n, t_k^n}^{m,j} \right\|_{\frac{p}{j}} \leqslant \begin{cases} C\left(\frac{1}{2^{\frac{n}{2}}}\right)^j, & n \leqslant m; \\ C\left(\frac{2^{\frac{m}{2}}}{2^n}\right)^j, & n > m, \end{cases}$$

where $j = 1, 2$.

(2)

$$\left\| B_{t_{k-1}^n, t_k^n}^{m+1,1} - B_{t_{k-1}^n, t_k^n}^{m,1} \right\|_p \leqslant \begin{cases} 0, & n \leqslant m; \\ C\frac{2^{\frac{m}{2}}}{2^n}, & n > m, \end{cases}$$

$$\left\| B_{t_{k-1}^n, t_k^n}^{m+1,2} - B_{t_{k-1}^n, t_k^n}^{m,2} \right\|_{\frac{p}{2}} \leqslant \begin{cases} C\frac{1}{2^{\frac{m}{2}}2^{\frac{n}{2}}}, & n \leqslant m; \\ C\frac{2^m}{2^{2n}}, & n > m. \end{cases}$$

Here $\|\cdot\|_q$ denotes the L^q-norm under the G-expectation $\mathbb{E}^G$, and C is some positive constant not depending on m, n, k.

Proof. (1) The first level.

If $n \leqslant m$, then

$$B_{t_{k-1}^n, t_k^n}^{m,1} = B_{t_k^n} - B_{t_{k-1}^n}.$$

It follows from the moment inequality (5.4.1) that

$$\mathbb{E}^G\left[\left|B^{m,1}_{t^n_{k-1},t^n_k}\right|^p\right] \leqslant C\frac{1}{2^{\frac{np}{2}}},$$

and thus

$$\left\|B^{m,1}_{t^n_{k-1},t^n_k}\right\|_p \leqslant C\frac{1}{2^{\frac{n}{2}}}.$$

Also it is trivial to see that

$$B^{m+1,1}_{t^n_{k-1},t^n_k} - B^{m,1}_{t^n_{k-1},t^n_k} = \left(B_{t^n_k} - B_{t^n_{k-1}}\right) - \left(B_{t^n_k} - B_{t^n_{k-1}}\right) = 0.$$

If $n > m$, then by construction we know that

$$B^{m,1}_{t^n_{k-1},t^n_k} = \frac{2^m}{2^n}\left(B_{t^m_l} - B_{t^m_{l-1}}\right),$$

where l is the unique integer such that $[t^n_{k-1}, t^n_k] \subset [t^m_{l-1}, t^m_l]$. Therefore,

$$\left\|B^{m,1}_{t^n_{k-1},t^n_k}\right\|_p = \frac{2^m}{2^n}\left\|B_{t^m_l} - B_{t^m_{l-1}}\right\|_p \leqslant C\frac{2^{\frac{m}{2}}}{2^n}.$$

On the other hand, if $[t^n_{k-1}, t^n_k] \subset [t^{m+1}_{2l-2}, t^{m+1}_{2l-1}]$, then

$$\begin{aligned} B^{m+1,1}_{t^n_{k-1},t^n_k} - B^{m,1}_{t^n_{k-1},t^n_k} &= \frac{2^{m+1}}{2^n}\left(B_{t^{m+1}_{2l-1}} - B_{t^{m+1}_{2l-2}}\right) - \frac{2^m}{2^n}\left(B_{t^m_l} - B_{t^m_{l-1}}\right) \\ &= \frac{2^m}{2^n}\left(\left(B_{\frac{2l-1}{2^{m+1}}} - B_{\frac{2l-2}{2^{m+1}}}\right) - \left(B_{\frac{2l}{2^{m+1}}} - B_{\frac{2l-1}{2^{m+1}}}\right)\right). \end{aligned}$$

It follows that

$$\left\|B^{m+1,1}_{t^n_{k-1},t^n_k} - B^{m,1}_{t^n_{k-1},t^n_k}\right\|_p \leqslant C\frac{2^{\frac{m}{2}}}{2^n}.$$

Similarly, if $[t^n_{k-1}, t^n_k] \subset [t^{m+1}_{2l-1}, t^{m+1}_{2l}]$, we obtain the same estimate.

(2) The second level.

Since $\frac{p}{2} < 2$, by monotonicity it suffices to establish the desired estimates under the L^2-norm.

First consider the term $B^{m+1,2}_{t^n_{k-1},t^n_k} - B^{m,2}_{t^n_{k-1},t^n_k}$.

If $n \leqslant m$, by the construction of $B^{m,2}_{s,t}$, we have

$$
\begin{aligned}
B^{m,2;\alpha,\beta}_{t^n_{k-1},t^n_k} &= \int_{t^n_{k-1}}^{t^n_k} B^{m,1;\alpha}_{t^n_{k-1},v} dB^{m,1;\beta}_v \\
&= \sum_{l=2^{(m-n)}(k-1)+1}^{2^{m-n}k} \frac{\Delta^m_l B^\beta}{\Delta t^m} \int_{t^m_{l-1}}^{t^m_l} \left(\frac{v - t^m_{l-1}}{\Delta t^m} B^\alpha_{t^m_l} + \frac{t^m_l - v}{\Delta t^m} B^\alpha_{t^m_{l-1}} - B^\alpha_{t^n_{k-1}} \right) dv \\
&= \sum_{l=2^{(m-n)}(k-1)+1}^{2^{m-n}k} \left(\frac{B^\alpha_{t^m_{l-1}} + B^\alpha_{t^m_l}}{2} - B^\alpha_{t^n_{k-1}} \right) \Delta^m_l B^\beta.
\end{aligned}
$$

Therefore,

$$
\begin{aligned}
& B^{m+1,2;\alpha,\beta}_{t^n_{k-1},t^n_k} - B^{m,2;\alpha,\beta}_{t^n_{k-1},t^n_k} \\
=& \sum_{l=2^{(m+1-n)}(k-1)+1}^{2^{m+1-n}k} \left(\frac{B^\alpha_{t^{m+1}_{l-1}} + B^\alpha_{t^{m+1}_l}}{2} - B^\alpha_{t^n_{k-1}} \right) \Delta^m_l B^\beta \\
& - \sum_{l=2^{(m-n)}(k-1)+1}^{2^{m-n}k} \left(\frac{B^\alpha_{t^m_{l-1}} + B^\alpha_{t^m_l}}{2} - B^\alpha_{t^n_{k-1}} \right) \Delta^m_l B^\beta \\
=& \sum_{l=2^{(m-n)}(k-1)+1}^{2^{(m-n)}k} \left(\left(\frac{B^\alpha_{t^{m+1}_{2l-2}} + B^\alpha_{t^{m+1}_{2l-1}}}{2} - B^\alpha_{t^n_{k-1}} \right) \Delta^{m+1}_{2l-1} B^\beta \right. \\
& + \left(\frac{B^\alpha_{t^{m+1}_{2l-1}} + B^\alpha_{t^{m+1}_{2l}}}{2} - B^\alpha_{t^n_{k-1}} \right) \Delta^{m+1}_{2l} B^\beta \\
& \left. - \left(\frac{B^\alpha_{t^{m+1}_{2l-2}} + B^\alpha_{t^{m+1}_{2l}}}{2} - B^\alpha_{t^n_{k-1}} \right) \left(\Delta^{m+1}_{2l-1} B^\beta + \Delta^{m+1}_{2l} B^\beta \right) \right) \\
=& \frac{1}{2} \sum_{l=2^{m-n}(k-1)+1}^{2^{m-n}k} \left(\Delta^{m+1}_{2l-1} B^\alpha \Delta^{m+1}_{2l} B^\beta - \Delta^{m+1}_{2l} B^\alpha \Delta^{m+1}_{2l-1} B^\beta \right).
\end{aligned}
$$

By using the notation of tensor products, we have

$$
B^{m+1,2}_{t^n_{k-1},t^n_k} - B^{m,2}_{t^n_{k-1},t^n_k} = \frac{1}{2} \sum_{l=2^{m-n}(k-1)+1}^{2^{m-n}k} \left(\Delta^{m+1}_{2l-1} B \otimes \Delta^{m+1}_{2l} B - \Delta^{m+1}_{2l} B \otimes \Delta^{m+1}_{2l-1} B \right).
$$

It follows that

$$
\begin{aligned}
& \mathbb{E}^G\left[\left|B^{m+1,2}_{t^n_{k-1},t^n_k}-B^{m,2}_{t^n_{k-1},t^n_k}\right|^2\right] \\
= & \frac{1}{4}\mathbb{E}^G\left[\left|\sum_{l=2^{m-n}(k-1)+1}^{2^{m-n}k}\left(\Delta^{m+1}_{2l-1}B\otimes\Delta^{m+1}_{2l}B-\Delta^{m+1}_{2l}B\otimes\Delta^{m+1}_{2l-1}B\right)\right|^2\right] \\
\leqslant & C\sum_{\substack{\alpha\neq\beta \\ \alpha,\beta=1,\cdots,d}}\mathbb{E}^G\left[\left|\sum_{l}\left(\Delta^{m+1}_{2l-1}B^\alpha\Delta^{m+1}_{2l}B^\beta-\Delta^{m+1}_{2l}B^\alpha\Delta^{m+1}_{2l-1}B^\beta\right)\right|^2\right] \\
\leqslant & C\sum_{\alpha\neq\beta}\sum_{l,r}\mathbb{E}^G\left[\left(\Delta^{m+1}_{2l-1}B^\alpha\Delta^{m+1}_{2l}B^\beta-\Delta^{m+1}_{2l}B^\alpha\Delta^{m+1}_{2l-1}B^\beta\right)\right. \\
& \left.\cdot\left(\Delta^{m+1}_{2r-1}B^\alpha\Delta^{m+1}_{2r}B^\beta-\Delta^{m+1}_{2r}B^\alpha\Delta^{m+1}_{2r-1}B^\beta\right)\right] \\
\leqslant & C\sum_{\alpha\neq\beta}\sum_{l,r}\left(\mathbb{E}^G\left[\Delta^{m+1}_{2l-1}B^\alpha\Delta^{m+1}_{2r-1}B^\alpha\Delta^{m+1}_{2l}B^\beta\Delta^{m+1}_{2r}B^\beta\right]\right. \\
& +\mathbb{E}^G\left[\Delta^{m+1}_{2l}B^\alpha\Delta^{m+1}_{2r}B^\alpha\Delta^{m+1}_{2l-1}B^\beta\Delta^{m+1}_{2r-1}B^\beta\right] \\
& +\mathbb{E}^G\left[-\Delta^{m+1}_{2l-1}B^\alpha\Delta^{m+1}_{2r}B^\alpha\Delta^{m+1}_{2r-1}B^\beta\Delta^{m+1}_{2l}B^\beta\right] \\
& \left.+\mathbb{E}^G\left[-\Delta^{m+1}_{2r-1}B^\alpha\Delta^{m+1}_{2l}B^\alpha\Delta^{m+1}_{2l-1}B^\beta\Delta^{m+1}_{2r}B^\beta\right]\right),
\end{aligned}
$$

where the summation over l and r is taken from $2^{m-n}(k-1)+1$ to $2^{m-n}k$. Here we have used the sub-linearity of $\mathbb{E}$. Now we study every term separately. If $l<r$, by the properties of conditional G-expectation and the distribution of B_t, we have

$$
\begin{aligned}
& \mathbb{E}^G\left[\Delta^{m+1}_{2l-1}B^\alpha\Delta^{m+1}_{2l}B^\beta\Delta^{m+1}_{2r-1}B^\alpha\Delta^{m+1}_{2r}B^\beta\right] \\
= & \mathbb{E}^G\left[\mathbb{E}^G\left[\Delta^{m+1}_{2l-1}B^\alpha\Delta^{m+1}_{2l}B^\beta\Delta^{m+1}_{2r-1}B^\alpha\Delta^{m+1}_{2r}B^\beta\middle|\,\Omega_{t^{m+1}_{2r-1}}\right]\right] \\
= & \mathbb{E}^G\left[\eta^+\mathbb{E}^G\left[\Delta^{m+1}_{2r}B^\beta\middle|\,\Omega_{t^{m+1}_{2r-1}}\right]+\eta^-\mathbb{E}\left[-\Delta^{m+1}_{2r}B^\beta\middle|\,\Omega_{t^{m+1}_{2r-1}}\right]\right] \\
= & 0,
\end{aligned}
$$

where $\eta=\Delta^{m+1}_{2l-1}B^\alpha\Delta^{m+1}_{2l}B^\beta\Delta^{m+1}_{2r-1}B^\alpha$. Similarly, we can prove that for any $l\neq r$,

$$
\begin{aligned}
& \mathbb{E}^G\left[\Delta^{m+1}_{2l-1}B^\alpha\Delta^{m+1}_{2r-1}B^\beta\Delta^{m+1}_{2l}B^\beta\Delta^{m+1}_{2r}B^\beta\right] \\
= & \mathbb{E}^G\left[\left(\Delta^{m+1}_{2l}B^\alpha\Delta^{m+1}_{2r}B^\alpha\Delta^{m+1}_{2l-1}B^\beta\Delta^{m+1}_{2r-1}B^\beta\right)\right] \\
= & \mathbb{E}^G\left[\left(-\Delta^{m+1}_{2l-1}B^\alpha\Delta^{m+1}_{2r}B^\alpha\Delta^{m+1}_{2r-1}B^\beta\Delta^{m+1}_{2l}B^\beta\right)\right] \\
= & \mathbb{E}^G\left[\left(-\Delta^{m+1}_{2r-1}B^\alpha\Delta^{m+1}_{2l}B^\alpha\Delta^{m+1}_{2l-1}B^\beta\Delta^{m+1}_{2r}B^\beta\right)\right] \\
= & 0.
\end{aligned}
$$

On the other hand, if $l = r$, it is straight forward that

$$\begin{aligned}\mathbb{E}^G\left[\left(\Delta_{2l-1}^{m+1}B^\alpha\right)^2\left(\Delta_{2l}^{m+1}B^\beta\right)^2\right] &\leqslant \frac{1}{2}\left(\mathbb{E}^G\left[\left(\Delta_{2l-1}^{m+1}B^\alpha\right)^4\right]+\mathbb{E}^G\left[\left(\Delta_{2l}^{m+1}B^\beta\right)^4\right]\right)\\ &\leqslant C\frac{1}{2^{2m}},\end{aligned}$$

and similarly,

$$\begin{aligned}&\mathbb{E}^G\left(-\Delta_{2l-1}^{m+1}B^\alpha\Delta_{2l-1}^{m+1}B^\beta\Delta_{2l}^{m+1}B^\alpha\Delta_{2l}^{m+1}B^\beta\right)\\ \leqslant\ &\frac{1}{4}\left(\mathbb{E}^G\left[\left(\Delta_{2l-1}^{m+1}B^\alpha\right)^4\right]+\mathbb{E}^G\left[\left(\Delta_{2l-1}^{m+1}B^\beta\right)^4\right]\right.\\ &\left.+\mathbb{E}^G\left[\left(\Delta_{2l}^{m+1}B^\alpha\right)^4\right]+\mathbb{E}^G\left[\left(\Delta_{2l}^{m+1}B^\beta\right)^4\right]\right)\\ \leqslant\ &C\frac{1}{2^{2m}}.\end{aligned}$$

Combining all the estimates above, we arrive at

$$\mathbb{E}^G\left[\left|B_{t_{k-1}^n,t_k^n}^{m+1,2}-B_{t_{k-1}^n,t_k^n}^{m,2}\right|^2\right] \leqslant C\sum_{\alpha\neq\beta}\sum_{l=2^{m-n}(k-1)+1}^{2^{m-n}k}\frac{1}{2^{2m}}\leqslant C\frac{1}{2^m2^n},$$

and hence

$$\left\|B_{t_{k-1}^n,t_k^n}^{m+1,2}-B_{t_{k-1}^n,t_k^n}^{m,2}\right\|_2\leqslant C\frac{1}{2^{\frac{m}{2}}2^{\frac{n}{2}}}.$$

If $n > m$, by construction we have

$$\begin{aligned}B_{t_{k-1}^n,t_k^n}^{m,2;\alpha,\beta} &= \int_{t_{k-1}^n<u<v<t_k^n} d(B^m)_u^\alpha d(B^m)_v^\beta\\ &= \int_{t_{k-1}^n}^{t_k^n} B_{t_{k-1}^n,v}^{m,1;\alpha}d(B^m)_v^\beta\\ &= \frac{\Delta_l^mB^\alpha\Delta_l^mB^\beta}{(\Delta t^m)^2}\int_{t_{k-1}^n}^{t_k^n}(v-t_{k-1}^n)dv\\ &= \frac{1}{2}2^{2(m-n)}\Delta_l^mB^\alpha\Delta_l^mB^\beta,\end{aligned}$$

where l is the unique integer such that $[t_{k-1}^n, t_k^n] \subset [t_{l-1}^m, t_l^m]$. In other words, we have

$$B_{t_{k-1}^n,t_k^n}^{m,2}=\frac{1}{2}2^{2(m-n)}(\Delta_l^mB)^{\otimes 2},$$

It follows that

$$B_{t_{k-1}^n,t_k^n}^{m+1,2}-B_{t_{k-1}^n,t_k^n}^{m,2}=2^{2(m-n)+1}(\Delta_{2l-1}^{m+1}B)^{\otimes 2}-2^{2(m-n)-1}(\Delta_l^mB)^{\otimes 2}$$

if $[t_{k-1}^n, t_k^n] \subset [t_{2l-2}^{m+1}, t_{2l-1}^{m+1}]$, and

$$B_{t_{k-1}^n, t_k^n}^{m+1,2} - B_{t_{k-1}^n, t_k^n}^{m,2} = 2^{2(m-n)+1}(\Delta_{2l}^{m+1} B)^{\otimes 2} - 2^{2(m-n)-1}(\Delta_l^m B)^{\otimes 2}$$

if $[t_{k-1}^n, t_k^n] \subset [t_{2l-1}^{m+1}, t_{2l}^{m+1}]$. By using the Minkowski inequality, the Cauchy-Schwarz inequality and the sub-linearity of $\mathbb{E}$, it is easy to obtain that

$$\left\| B_{t_{k-1}^n, t_k^n}^{m+1,2} - B_{t_{k-1}^n, t_k^n}^{m,2} \right\|_2 \leqslant C \frac{2^m}{2^{2n}}.$$

Now consider the term $B_{t_{k-1}^n, t_k^n}^{m,2}$.

If $n \geqslant m$, by using

$$B_{t_{k-1}^n, t_k^n}^{m,2} = 2^{2(m-n)-1}(\Delta_l^m B)^{\otimes 2},$$

we can proceed in the same way as before to obtain that

$$\left\| B_{t_{k-1}^n, t_k^n}^{m,2} \right\|_2 \leqslant C \frac{2^m}{2^{2n}}.$$

If $n < m$, then

$$B_{t_{k-1}^n, t_k^n}^{m,2} = \sum_{l=n+1}^{m} \left(B_{t_{k-1}^n, t_k^n}^{l,2} - B_{t_{k-1}^n, t_k^n}^{l-1,2} \right) + B_{t_{k-1}^n, t_k^n}^{n,2}.$$

It follows that

$$\begin{aligned} \left\| B_{t_{k-1}^n, t_k^n}^{m,2} \right\|_2 &\leqslant \sum_{l=n+1}^{m} \left\| B_{t_{k-1}^n, t_k^n}^{l,2} - B_{t_{k-1}^n, t_k^n}^{l-1,2} \right\|_2 + \left\| B_{t_{k-1}^n, t_k^n}^{n,2} \right\|_2 \\ &\leqslant C \left(\frac{1}{2^{\frac{n}{2}}} \sum_{l=n+1}^{\infty} \frac{1}{2^{\frac{l}{2}}} + \frac{1}{2^n} \right) \\ &\leqslant C \frac{1}{2^n}. \end{aligned}$$

Now the proof is complete. □

In order to study the convergence of $\boldsymbol{B}^m$ in the space $G\Omega_p(\mathbb{R}^d)$, we need to control the p-variation metric d_p in a proper way. For $\mathbf{X}, \widetilde{\mathbf{X}} \in G\Omega(\mathbb{R}^d)$, define

$$\rho_j(\mathbf{X}, \widetilde{\mathbf{X}}) = \left(\sum_{n=1}^{\infty} n^{\gamma} \sum_{k=1}^{2^n} \left| X_{t_{k-1}^n, t_k^n}^j - \widetilde{X}_{t_{k-1}^n, t_k^n}^j \right|^{\frac{p}{j}} \right)^{\frac{j}{p}}, \ j = 1, 2, \tag{5.4.2}$$

where $\gamma > p - 1$ is some fixed universal constant. The functional ρ_j was initially introduced by B.M. Hambly and T. Lyons [33] to construct the stochastic area process associated with the Brownian motion on the Sierpinski gasket. We use $\rho_j(\mathbf{X})$ to denote $\rho_j(\mathbf{X}, \widetilde{\mathbf{X}})$ with $\widetilde{\mathbf{X}} = (1, 0, 0)$.

The following estimate for the p-variation metric is crucial for us. We refer the reader to [47] for the proof.

Proposition 5.4.1. *There exists some positive constant $R = R(p, \gamma)$, such that for any $\mathbf{X}, \widetilde{\mathbf{X}} \in G\Omega(\mathbb{R}^d)$,*

$$d_p(\mathbf{X}, \widetilde{\mathbf{X}}) \leqslant R \cdot \max\{\rho_1(\mathbf{X}, \widetilde{\mathbf{X}}), \rho_1(\mathbf{X}, \widetilde{\mathbf{X}})(\rho_1(\mathbf{X}) + \rho_1(\widetilde{\mathbf{X}})), \rho_2(\mathbf{X}, \widetilde{\mathbf{X}})\}.$$

Now let

$$I(\mathbf{X}, \widetilde{\mathbf{X}}) := \max\{\rho_1(\mathbf{X}, \widetilde{\mathbf{X}}), \rho_1(\mathbf{X}, \widetilde{\mathbf{X}})(\rho_1(\mathbf{X}) + \rho_1(\widetilde{\mathbf{X}})), \rho_2(\mathbf{X}, \widetilde{\mathbf{X}})\}, \tag{5.4.3}$$

and observe that

$$\begin{aligned}
&\{\omega : \ \mathbf{B}^m \text{ is not Cauchy under } d_p\} \\
&\subset \left\{\omega : \ \sum_{m=1}^{\infty} d_p\left(\mathbf{B}^m, \mathbf{B}^{m+1}\right) = \infty\right\} \\
&\subset \limsup_{m\to\infty} \left\{\omega : \ d_p\left(\mathbf{B}^m, \mathbf{B}^{m+1}\right) > \frac{R}{2^{m\beta}}\right\} \\
&\subset \limsup_{m\to\infty} \left\{\omega : \ I(\mathbf{B}^m, \mathbf{B}^{m+1}) > \frac{1}{2^{m\beta}}\right\}.
\end{aligned} \tag{5.4.4}$$

where β is some positive constant to be chosen. Notice that the R.H.S. of (5.4.4) is $\mathcal{B}(\Omega)$-measurable so its capacity is well-defined. Therefore, in order to prove that quasi-surely, $\mathbf{B}^m$ is a Cauchy sequence under d_p, it suffices to show that the R.H.S. of (5.4.4) has capacity zero. This can be shown by using the Borel-Cantelli lemma.

According to (5.4.3), we may first need to establish estimates for

$$c\left(\rho_j\left(\mathbf{B}^m, \mathbf{B}^{m+1}\right) > \lambda\right), \ j = 1, 2,$$

and

$$c\left(\rho_1\left(\mathbf{B}^m\right) > \lambda\right),$$

where $m \geqslant 1$ and $\lambda > 0$. They are contained in the following lemma.

Lemma 5.4.2. *For $m \geqslant 1$, $\lambda > 0$, we have the following estimates.*

(1)

$$c(\rho_1(\mathbf{B}^m) > \lambda) \leqslant C\lambda^{-p}.$$

(2) Let $\theta \in (0, \frac{p}{2} - 1)$ be some constant such that

$$n^{\gamma+1} \leqslant C \frac{2^{n(p-1)}}{2^{n(p-\theta-1)}}, \ \forall n \geqslant 1.$$

Then we have

$$c\left(\rho_j\left(\mathbf{B}^m, \mathbf{B}^{m+1}\right) > \lambda\right) \leqslant C\lambda^{-\frac{p}{j}} \frac{1}{2^{m\left(\frac{p}{2}-\theta-1\right)}}, \ j = 1, 2.$$

Proof. First consider

$$c\left(\rho_1\left(\mathbf{B}^m\right) > \lambda\right) = c\left(\sum_{n=1}^{\infty} n^{\gamma} \sum_{k=1}^{2^n} \left|B^{m,1}_{t^n_{k-1},t^n_k}\right|^p > \lambda^p\right).$$

Define

$$A_N = \left\{\omega : \ \sum_{n=1}^{N} n^{\gamma} \sum_{k=1}^{2^n} \left|B^{m,1}_{t^n_{k-1},t^n_k}\right|^p > \lambda^p\right\} \in \mathcal{B}(\Omega),$$

and

$$A = \left\{\omega : \ \sum_{n=1}^{\infty} n^{\gamma} \sum_{k=1}^{2^n} \left|B^{m,1}_{t^n_{k-1},t^n_k}\right|^p > \lambda^p\right\} \in \mathcal{B}(\Omega).$$

It is obvious that $A_N \uparrow A$. By the properties of the capacity c, we have

$$c(A) = \lim_{N\to\infty} c(A_N).$$

On the other hand, by the sub-linearity of $\mathbb{E}^G$, the Chebyshev inequality for the capacity c and Lemma 5.4.1, we have

$$\begin{aligned}
c(A_N) &\leqslant \lambda^{-p} \sum_{n=1}^{N} n^{\gamma} \sum_{k=1}^{2^n} \mathbb{E}\left[\left|B^{m,1}_{t^n_{k-1},t^n_k}\right|^p\right] \\
&\leqslant C\lambda^{-p}\left[\sum_{n=1}^{m} n^{\gamma} 2^n \frac{1}{2^{\frac{np}{2}}} + \sum_{n=m+1}^{\infty} n^{\gamma} 2^n \frac{2^{\frac{mp}{2}}}{2^{np}}\right] \\
&= C\lambda^{-p}\left[\sum_{n=1}^{m} n^{\gamma} \frac{1}{2^{n\left(\frac{p}{2}-1\right)}} + 2^{\frac{mp}{2}} \sum_{n=m+1}^{\infty} n^{\gamma} \frac{1}{2^{n(p-1)}}\right] \\
&\leqslant C\lambda^{-p}.
\end{aligned}$$

It follows that

$$c\left(\rho_1\left(\mathbf{B}^m\right)>\lambda\right)=c(A)\leqslant C\lambda^{-p}.$$

Now consider

$$c\left(\rho_1\left(\mathbf{B}^m,\mathbf{B}^{m+1}\right)>\lambda\right)=c\left(\sum_{n=1}^{\infty}n^{\gamma}\sum_{k=1}^{2^n}\left|B^{m+1,1}_{\frac{k-1}{2^n},\frac{k}{2^n}}-B^{m,1}_{\frac{k-1}{2^n},\frac{k}{2^n}}\right|^p>\lambda^p\right).$$

For similar reasons we have

$$\begin{aligned}c\left(\rho_1\left(\mathbf{B}^m,\mathbf{B}^{m+1}\right)>\lambda\right) &\leqslant \lambda^{-p}\sum_{n=1}^{\infty}n^{\gamma}\sum_{k=1}^{2^n}\mathbb{E}\left[\left|B^{m+1,1}_{t^n_{k-1},t^n_k}-B^{m,1}_{t^n_{k-1},t^n_k}\right|^p\right]\\ &\leqslant C\lambda^{-p}\left(\sum_{n=m+1}^{\infty}n^{\gamma}2^n\frac{2^{\frac{mp}{2}}}{2^{np}}\right)\\ &= C\lambda^{-p}2^{\frac{mp}{2}}\sum_{n=m+1}^{\infty}n^{\gamma}\frac{1}{2^{n(p-1)}}.\end{aligned}$$

Since $\theta\in(0,\frac{p}{2}-1)$ satisfies

$$n^{\gamma+1}\leqslant C\frac{2^{n(p-1)}}{2^{n(p-\theta-1)}},\ \ \forall n\geqslant 1,$$

we arrive at

$$c\left(\rho_1\left(\mathbf{B}^m,\mathbf{B}^{m+1}\right)>\lambda\right)\leqslant C\lambda^{-p}\frac{1}{2^{m\left(\frac{p}{2}-\theta-1\right)}}.$$

Finally, consider the second level part. For similar reasons, we have

$$\begin{aligned}&c\left(\rho_2\left(\mathbf{B}^m,\mathbf{B}^{m+1}\right)>\lambda\right)\\ \leqslant\ & C\lambda^{-\frac{p}{2}}\left[\sum_{n=1}^{m}n^{\gamma}2^n\frac{1}{2^{\frac{mp}{4}}2^{\frac{np}{4}}}+2^{\frac{mp}{2}}\sum_{n=m+1}^{\infty}n^{\gamma}2^n\frac{1}{2^{np}}\right]\\ =\ & C\lambda^{-\frac{p}{2}}\left[\frac{1}{2^{\frac{mp}{4}}}\sum_{n=1}^{m}n^{\gamma}2^{n\left(1-\frac{p}{4}\right)}+2^{\frac{mp}{2}}\sum_{n=m+1}^{\infty}n^{\gamma}\frac{1}{2^{n(p-1)}}\right]\\ \leqslant\ & C\lambda^{-\frac{p}{2}}\left[\frac{1}{2^{\frac{mp}{4}}}m^{\gamma+1}2^{m\left(1-\frac{p}{4}\right)}+2^{\frac{mp}{2}}\frac{1}{2^{m(p-\theta-1)}}\right]\\ \leqslant\ & C\lambda^{-\frac{p}{2}}\frac{1}{2^{m\left(\frac{p}{2}-\theta-1\right)}}.\end{aligned}$$

□

Now we are in position to prove the main result of this section.

Theorem 5.4.1. *Outside a $\mathcal{B}(\Omega)$-measurable set of capacity zero, $\mathbf{B}^m$ is a Cauchy sequence under the p-variation metric d_p. In particular, quasi-surely, the sample paths of B_t can be lifted as geometric p-rough paths*

$$\mathbf{B}_{s,t} = \left(1, B^1_{s,t}, B^2_{s,t}\right), \ 0 \leqslant s < t \leqslant 1,$$

which are defined as the limit of sample (geometric p-rough) paths of $\boldsymbol{B}^m$ in $G\Omega_p\left(\mathbb{R}^d\right)$ under the p-variation metric d_p.

Proof. By Lemma 5.4.2, we have

$$\begin{aligned} & c\left(I\left(\mathbf{B}^m, \mathbf{B}^{m+1}\right) > \frac{1}{2^{m\beta}}\right) \\ \leqslant & \sum_{j=1}^{2} c\left(\rho_j\left(\mathbf{B}^m, \mathbf{B}^{m+1}\right) > \frac{1}{2^{m\beta}}\right) \\ & + c\left(\rho_1\left(\mathbf{B}^m, \mathbf{B}^{m+1}\right)\left(\rho_1\left(\mathbf{B}^m\right) + \rho_1\left(\mathbf{B}^{m+1}\right)\right) > \frac{1}{2^{m\beta}}\right) \\ \leqslant & 2c\left(\rho_1\left(\mathbf{B}^m, \mathbf{B}^{m+1}\right) > \frac{1}{2^{2m\beta}}\right) + c\left(\rho_2\left(\mathbf{B}^m, \mathbf{B}^{m+1}\right) > \frac{1}{2^{m\beta}}\right) \\ & + c\left(\rho_1\left(\mathbf{B}^m\right) > \frac{2^{m\beta}}{2}\right) + c\left(\rho_1\left(\mathbf{B}^{m+1}\right) > \frac{2^{m\beta}}{2}\right) \\ \leqslant & C\left[\frac{1}{2^{m\beta p}} + \frac{1}{2^{m\left(\frac{p}{2}-\theta-2\beta p-1\right)}} + \frac{1}{2^{m\left(\frac{p}{2}-\theta-\frac{\beta p}{2}-1\right)}}\right], \end{aligned}$$

where $\theta \in \left(0, \frac{p}{2}-1\right)$ is some fixed constant.

If we choose β such that

$$0 < \beta < \frac{p-2\theta-2}{4p},$$

then

$$\sum_{m=1}^{\infty} c\left(I\left(\mathbf{B}^m, \mathbf{B}^{m+1}\right) > \frac{1}{2^{m\beta}}\right) < \infty.$$

By the Borel-Cantelli lemma, we have

$$c\left(\limsup_{m\to\infty}\left\{\omega: \ I\left(\mathbf{B}^m, \mathbf{B}^{m+1}\right) > \frac{1}{2^{m\beta}}\right\}\right) = 0,$$

and the result follows from the inclusion (5.4.4). □

With the help of Theorem 5.4.1 and the regularity of $\left\langle B^\alpha, B^\beta\right\rangle_t$ (by definition the sample paths of $\left\langle B^{\alpha}, B^{\beta}\right\rangle_t$ have bounded total variation), we are able to apply the universal limit theorem in rough path theory to define RDEs driven by G-Brownian motion in the pathwise sense. More precisely, consider the following N-dimensional RDE in the sense of rough paths:

$$dY_t = \widetilde{b}(Y_t)dt + \widetilde{h}_{\alpha\beta}(Y_t)d\left\langle B^{\alpha}, B^{\beta}\right\rangle_t + V_{\alpha}(Y_t)dB_t^{\alpha}, \tag{5.4.5}$$

with initial condition $Y_0 = x$, where $\widetilde{b}, \widetilde{h}_{\alpha\beta}, V_{\alpha}$ are C_b^3-vector fields on $\mathbb{R}^N$. Then outside a $\mathcal{B}(\Omega)$-measurable set of capacity zero, (5.4.5) has a unique solution $\mathbf{Y}$ in $G\Omega_p(\mathbb{R}^N)$. $\mathbf{Y}$ is constructed as the limit of the lifting of Y_t^n in $G\Omega_p\left(\mathbb{R}^N\right)$ under the p-variation metric, where Y_t^n is the unique classical solution to the following ordinary differential equation:

$$dY_t^n \;=\; \widetilde{b}(Y_t^n)dt + \widetilde{h}_{\alpha\beta}(Y_t^n)d\left\langle B^{\alpha}, B^{\beta}\right\rangle_t + V_{\alpha}(Y_t^n)d(B^n)_t^{\alpha}, \tag{5.4.6}$$

with $Y_0^n = x$, in which B_t^n is the dyadic piecewise linear interpolation of B_t.

In practice, we usually only consider the first level $Y := x + \pi_1(\mathbf{Y})$ of the solution instead of the full geometric rough path $\mathbf{Y}$. Therefore, without ambiguity we simply regard Y as the solution to the RDE (5.4.5). It is easy to see that quasi-surely, Y is the uniform limit of the solution to (5.4.6) with initial condition $Y_0^n = x$.

To end this section, we give an explicit description of the second level $B_{s,t}^2$ of B_t defined in Theorem 5.4.1 which reveals the nature of $B_{s,t}^2$. Such a result is crucial to understand the relationship between SDEs and RDEs driven by G-Brownian motion.

Lemma 5.4.3. *Assume that X_n converges to X in $L_G^2(\Omega)$ and converges to Y quasi-surely. Then $X = Y$ quasi-surely.*

Proof. By the Chebyshev inequality for the capacity, we have

$$c\left(|X_n - X| > \varepsilon\right) \leqslant \frac{1}{\varepsilon^2}\mathbb{E}^G\left[|X_n - X|^2\right],\ \forall \varepsilon > 0.$$

Since

$$X_n \to X \quad \text{in } L_G^2(\Omega),$$

we can extract a subsequence X_{n_k}, such that for any $k \geqslant 1$,

$$\mathbb{E}^G\left[|X_{n_k} - X|^2\right] \leqslant \frac{1}{k^4}.$$

It follows that

$$c\left(|X_{n_k}-X|>\frac{1}{k}\right)\leqslant\frac{1}{k^2},\quad\forall k\geqslant 1,$$

and

$$\sum_{k=1}^{\infty}c\left(|X_{n_k}-X|>\frac{1}{k}\right)<\infty.$$

By the Borel-Cantelli lemma for the capacity, we conclude that X_{n_k} converges to X quasi-surely. By assumption it follows that $X=Y$ quasi-surely. □

The following result shows the nature of the second level of B_t. In the case when B_t reduces to the classical Brownian motion, it is essentially the relation between Stratonovich's and Itô's integrals.

Proposition 5.4.2. *Let $\boldsymbol{B}_{s,t}=\left(1,B^1_{s,t},B^2_{s,t}\right)$ be the quasi-surely defined lifting of B_t in Theorem 5.4.1. Then for any $0\leqslant s<t\leqslant 1$, we have*

$$B^{2;\alpha,\beta}_{s,t}=\int_s^t B^\alpha_{s,u}dB^\beta_u+\frac{1}{2}\left\langle B^\alpha,B^\beta\right\rangle_{s,t}\qquad(5.4.7)$$

quasi-surely, where the integral on the R.H.S. of (5.4.7) is Itô's integral.

Proof. We know from Theorem 5.4.1 that

$$\lim_{n\to\infty}d_p(\mathbf{B}^n,\mathbf{B})=0$$

quasi-surely. From the definition of d_p, it is straight forward that $B^{n,2}_{s,t}$ converges uniformly to $B^2_{s,t}$ quasi-surely.

Without loss of generality, we assume that s,t are both dyadic points in $[0,1]$. It follows that when n is large enough,

$$\begin{aligned}
B^{n,2;\alpha,\beta}_{s,t} &= \int_{s<u<v<t}d(B^n)^\alpha_u d(B^n)^\beta_v\\
&= \int_s^t (B^n)^\alpha_{s,v}d(B^n)^\beta_v\\
&= \sum_{k:[t^n_{k-1},t^n_k]\subset[s,t]}\frac{\Delta^n_k B^\beta}{\Delta t^n}\int_{t^n_{k-1}}^{t^n_k}\left(\frac{v-t^n_{k-1}}{\Delta t^n}B^\alpha_k+\frac{t^n_k-v}{\Delta t^n}B^\alpha_{k-1}-B^\alpha_s\right)dv\\
&= \sum_{k:[t^n_{k-1},t^n_k]\subset[s,t]}\left(\frac{B^\alpha_{k-1}+B^\alpha_k}{2}-B^\alpha_s\right)\Delta^n_k B^\beta\\
&= \sum_{k:[t^n_{k-1},t^n_k]\subset[s,t]}(B^\alpha_{k-1}-B^\alpha_s)\Delta^n_k B^\beta+\frac{1}{2}\sum_k\Delta^n_k B^\alpha\Delta^n_k B^\beta.
\end{aligned}$$

From properties of Itô integrals and the cross-variation $\left\langle B^{\alpha}, B^{\beta}\right\rangle_t$, we know that the R.H.S. of the above equality converges to $\int_s^t B_{s,u}^{\alpha} dB_u^{\beta} + \frac{1}{2}\left\langle B^{\alpha}, B^{\beta}\right\rangle_{s,t}$ in $L_G^2(\Omega)$.

Consequently, by Lemma 5.4.3 $B_{s,t}^2$ must coincide with $\int_s^t B_{s,u}^{\alpha} dB_u^{\beta} + \frac{1}{2}\left\langle B^{\alpha}, B^{\beta}\right\rangle_{s,t}$ quasi surely. □

5.5 The Relationship between SDEs and RDEs Driven by G-Brownian Motion

So far we have seen that there are two types of well-defined differential equations driven by G-Brownian motion: SDEs which are defined in the L_G^2-sense under the G-expectation $\mathbb{E}^G$, and RDEs which are quasi surely defined in the pathwise sense. This section is devoted to the study of the fundamental relationship between these two types of differential equation.

Consider the following N-dimensional SDE driven by G-Brownian motion on $(\Omega, L_G^2(\Omega), \mathbb{E})$:

$$dX_t = b(X_t)dt + h_{\alpha\beta}(X_t)d\left\langle B^{\alpha}, B^{\beta}\right\rangle_t + V_{\alpha}(X_t)dB_t^{\alpha}, \tag{5.5.1}$$

with initial condition $X_0 = x \in \mathbb{R}^N$. Here we assume that $b, h_{\alpha\beta}, V_{\alpha}$ are C_b^3-vector fields on $\mathbb{R}^N$.

Our aim is to find the correct RDE of the form (5.4.5) whose solution coincides with X_t quasi surely in the pathwise sense.

Let us first illustrate the idea informally. We use the rough Taylor expansion for RDEs and Proposition 5.4.2 to seek the correct form of the RDE for X_t.

Consider the following general RDE:

$$dY_t = \widetilde{b}(Y_t)dt + \widetilde{h}_{\alpha\beta}(Y_t)d\left\langle B^{\alpha}, B^{\beta}\right\rangle_t + \widetilde{V}_{\alpha}(Y_t)dB_t^{\alpha}, \tag{5.5.2}$$

with initial condition $Y_0 = x$, where $\widetilde{b}, \widetilde{h}_{\alpha\beta}, \widetilde{V}_{\alpha}$ are C_b^3-vector fields on $\mathbb{R}^N$. By the regularity of the cross variation process $\langle B^{\alpha}, B^{\beta}\rangle$ and the roughness of B_t studied in the last section, we know from the rough Taylor expansion theorem that quasi-surely, for some control $\omega(s,t)$, the solution Y_t to (5.5.2) satisfies, when $\omega(s,t) \leqslant 1$, that

$$\begin{aligned} &\left| Y_{s,t} - \widetilde{b}(Y_s)(t-s) - \widetilde{h}_{\alpha\beta}(Y_s)\left\langle B^{\alpha}, B^{\beta}\right\rangle_{s,t} - \widetilde{V}_{\alpha}(Y_s)B_{s,t}^{1;\alpha} - D\widetilde{V}_{\beta}(Y_s)\cdot\widetilde{V}_{\alpha}(Y_s)B_{s,t}^{2;\alpha,\beta}\right| \\ &\leqslant C\omega(s,t)^{\theta}, \end{aligned} \tag{5.5.3}$$

where C and $\theta > 1$ are two constants not depending on s, t. Note that inequality (5.5.3) reveals the local behavior of the solution Y_t. It follows from Proposition 5.4.2 that

$$\left|Y_{s,t} - \widetilde{I}_{s,t}\right| \leqslant C\omega(s,t)^{\theta}$$

quasi-surely, where

$$\begin{aligned} \widetilde{I}_{s,t} : &= \widetilde{b}(Y_s)(t-s) + (\widetilde{h}_{\alpha\beta}(Y_s) + \frac{1}{2}D\widetilde{V}_\beta(Y_s) \cdot \widetilde{V}_\alpha(Y_s))d\left\langle B^\alpha, B^\beta \right\rangle_t \\ &\quad + \widetilde{V}_\alpha(Y_s)B^{1;\alpha}_{s,t} + D\widetilde{V}_\beta(Y_s) \cdot \widetilde{V}_\alpha(Y_s)\int_s^t B^\alpha_{s,u}dB^\beta_u. \end{aligned} \tag{5.5.4}$$

Now if we consider the global behavior of Y_t, we may sum up inequality (5.5.4) over dyadic intervals $[t^n_{k-1}, t^n_k]$ and then take the limit (in $L^2_G(\Omega; \mathbb{R}^N)$) to obtain that

$$\begin{aligned} & Y_{s,t} \\ =& \int_s^t \widetilde{b}\left(Y_u\right)du + \int_s^t \left(\widetilde{h}_{\alpha\beta}\left(Y_u\right) + \frac{1}{2}D\widetilde{V}_\beta\left(Y_u\right) \cdot \widetilde{V}_\alpha\left(Y_u\right)\right)d\left\langle B^\alpha, B^\beta\right\rangle_u + \int_s^t \widetilde{V}_\alpha\left(Y_u\right)dB^\alpha_u \\ &+ \left(L^2_G -\right)\lim_{n\to\infty}\sum_{k:[t^n_{k-1},t^n_k]\subset[s,t]} D\widetilde{V}_\alpha\left(Y_{t^n_{k-1}}\right)\cdot\widetilde{V}_\beta\left(Y_{t^n_{k-1}}\right)\int_{t^n_{k-1}}^{t^n_k} B^\alpha_{t^n_{k-1},u}dB^\beta_u, \end{aligned} \tag{5.5.5}$$

quasi-surely, where the integrals against B_t are interpreted as Itô's integrals. On the other hand, by the distribution of B_t and properties of G-Itô's integrals, it is not hard to prove that the L^2_G-limit in the last term of the above identity is zero. Therefore, we know that Y_t solves the SDE

$$dX_t = \widetilde{b}(X_t)dt + \left(\widetilde{h}_{\alpha\beta}(X_t) + \frac{1}{2}D\widetilde{V}_\beta(X_t)\cdot\widetilde{V}_\alpha(X_t)\right)d\left\langle B^\alpha, B^\beta\right\rangle_t + \widetilde{V}_\alpha(X_t)dB^\alpha_t.$$

In other words, if X_t is the solution to the SDE (5.5.1), it is natural to expect that quasi-surely, X_t is the solution to the following RDE:

$$dY_t = b(Y_t)dt + \left(h_{\alpha\beta}(Y_t) - \frac{1}{2}DV_\beta(Y_t)\cdot V_\alpha(Y_t)\right)d\left\langle B^\alpha, B^\beta\right\rangle + V_\alpha(Y_t)dB^\alpha_t, \tag{5.5.6}$$

with the same initial condition.

Now we prove this assertion rigorously.

From now on, assume that X_t is the solution to the SDE (5.5.1) and Y_t is the solution to the RDE (5.5.6) with the same initial condition $x \in \mathbb{R}^N$, where the coefficients $b, h_{\alpha\beta}, V_\alpha$ are C^3_b-vector fields on $\mathbb{R}^N$. For simplicity we also use the same notation to denote constants only depending on d, N, G, p and the coefficients of (5.5.1), although they may be different from line to line.

The following lemma enables us to show that the L_G^2-limit in the last term of the identity (5.5.5) is zero.

Lemma 5.5.1. *Let $f \in C_b(\mathbb{R}^N)$, and $s < t$ be two dyadic points in $[0,1]$ (i.e. $s = t_k^m$ and $t = t_l^m$ for some m and $k < l$). Then for any $\alpha, \beta = 1, 2, \cdots, d$,*

$$\lim_{n\to\infty} \mathbb{E}^G\left[\left(\sum_{k:[t_{k-1}^n, t_k^n]\subset[s,t]} f\left(Y_{t_{k-1}^n}\right)\int_{t_{k-1}^n}^{t_k^n} B_{t_{k-1}^n,u}^{\alpha} dB_u^{\beta}\right)^2\right] = 0.$$

Proof. From direct calculation, we have

$$\begin{aligned}
&\mathbb{E}^G\left[\left(\sum_{k:[t_{k-1}^n, t_k^n]\subset[s,t]} f\left(Y_{t_{k-1}^n}\right)\int_{t_{k-1}^n}^{t_k^n} B_{t_{k-1}^n,u}^{\alpha} dB_u^{\beta}\right)^2\right] \\
\leqslant\ & \|f\|_\infty^2 \sum_{k:[t_{k-1}^n, t_k^n]\subset[s,t]} \mathbb{E}^G\left[\left(\int_{t_{k-1}^n}^{t_k^n} B_{t_{k-1}^n,u}^{\alpha} dB_u^{\beta}\right)^2\right] \\
&+2 \sum_{\substack{k<l \\ [t_{k-1}^n, t_k^n],[t_{l-1}^n, t_l^n]\subset[s,t]}} \left[\mathbb{E}^G f\left(Y_{t_{k-1}^n}\right)\left(\int_{t_{k-1}^n}^{t_k^n} B_{t_{k-1}^n,u}^{\alpha} dB_u^{\beta}\right)\right. \\
&\left.\cdot f\left(Y_{t_{l-1}^n}\right)\left(\int_{t_{l-1}^n}^{t_l^n} B_{t_{l-1}^n,u}^{\alpha} dB_u^{\beta}\right)\right] \\
\leqslant\ & C\|f\|_\infty^2 \sum_{k:[t_{k-1}^n, t_k^n]\subset[s,t]} (\Delta t^n)^2 \\
&+2 \sum_{\substack{k<l \\ [t_{k-1}^n, t_k^n],[t_{l-1}^n, t_l^n]\subset[s,t]}} \left(\mathbb{E}^G\left[\left(f\left(Y_{t_{k-1}^n}\right)\left(\int_{t_{k-1}^n}^{t_k^n} B_{t_{k-1}^n,u}^{\alpha} dB_u^{\beta}\right) f\left(Y_{t_{l-1}^n}\right)\right)^+\right.\right. \\
&\left.\cdot\mathbb{E}^G\left[\int_{t_{l-1}^n}^{t_l^n} B_{t_{l-1}^n,u}^{\alpha} dB_u^{\beta}\,\middle|\,\Omega_{t_{l-1}^n}\right]\right] \\
&+\mathbb{E}^G\left[\left(f(Y_{t_{k-1}^n})\left(\int_{t_{k-1}^n}^{t_k^n} B_{t_{k-1}^n,u}^{\alpha} dB_u^{\beta}\right) f\left(Y_{t_{l-1}^n}\right)\right)^-\right. \\
&\left.\left.\cdot\mathbb{E}^G\left[-\int_{t_{l-1}^n}^{t_l^n} B_{t_{l-1}^n,u}^{\alpha} dB_u^{\beta}\,\middle|\,\Omega_{t_{l-1}^n}\right]\right]\right) \\
\leqslant\ & C\|f\|_\infty^2 \Delta t^n,
\end{aligned}$$

and the result follows easily. □

Now we are in a position to prove the main result of this section.

Theorem 5.5.1. *Quasi-surely, we have*

$$X_t = Y_t, \ \forall t \in [0,1].$$

Proof. Since the coefficients of the RDE (5.5.6) are in $C_b^3(\mathbb{R}^N)$, quasi-surely we can define the following pathwise control

$$\begin{aligned} \omega(s,t) &= \left(\|V\|_{2,\infty}\|\boldsymbol{B}\|_{p;\,[s,\,t]}\right)^p + \|b\|_{1,\infty}(t-s) \\ &\quad + \left\|h - \frac{1}{2}DV\cdot V\right\|_{1,\infty}\|\langle B,B\rangle\|_{1;[s,t]} \end{aligned}$$

for $(s,t)\in\Delta$, where $\|\cdot\|_{m,\infty}$ denotes the maximum of uniform norms of derivatives up to order m. It follows from the rough Taylor expansion (see [26], Corollary 12.8) that quasi-surely, there exists some positive constant $\theta > 1$, such that for $0 \leqslant s < t \leqslant 1$, when $\omega(s,t)\leqslant 1$, we have

$$|Y_{s,t} - I_{s,t}| \leqslant C\omega(s,t)^\theta,$$

where

$$\begin{aligned} I_{s,t} &= b(Y_s)(t-s) + \left(h_{\alpha\beta}(Y_s) - \frac{1}{2}DV_\beta(Y_s)\cdot V_\alpha(Y_s)\right)\left\langle B^\alpha, B^\beta\right\rangle_{s,t} + V_\alpha(Y_s)B_{s,t}^{1;\alpha} \\ &\quad + DV_\beta(Y_s)\cdot V_\alpha(Y_s)B_{s,t}^{2;\alpha,\beta} \end{aligned}$$

By Proposition 5.4.2, we have quasi-surely,

$$\begin{aligned} &\Big| Y_{s,t} - b(Y_s)(t-s) - h_{\alpha\beta}(Y_s)\left\langle B^\alpha, B^\beta\right\rangle_{s,t} - V_\alpha(Y_s)B_{s,t}^{1;\alpha} \\ &\qquad - DV_\beta(Y_s)\cdot V_\alpha(Y_s)\int_s^t B_{s,u}^\alpha dB_u^\beta\Big| \leqslant C\omega(s,t)^\theta. \quad (5.5.7) \end{aligned}$$

Now consider fixed $s<t$ being two dyadic points in $[0,1]$. When n is large enough, by applying inequality (5.5.7) on each small dyadic interval $\left[t_{k-1}^n, t_k^n\right]\subset[s,t]$ and summing up through the triangle inequality, we obtain that

$$\begin{aligned} \left|Y_{s,t} - I_{s,t}^n\right| &\leqslant C\sum \omega\left(t_{k-1}^n, t_k^n\right)^\theta \\ &\leqslant C\omega(s,t)\max\left\{\omega\left(t_{k-1}^n, t_k^n\right)^{\theta-1} :\ \left[t_{k-1}^n, t_k^n\right]\subset[s,t]\right\}, \end{aligned}$$

quasi-surely, where

$$\begin{aligned} I_{s,t}^{n} &= \sum b\left(Y_{t_{k-1}^{n}}\right)\Delta t^{n}+\sum h_{\alpha\beta}\left(Y_{t_{k-1}^{n}}\right)\Delta_{k}^{n}\left\langle B^{\alpha},B^{\beta}\right\rangle+\sum V_{\alpha}\left(Y_{t_{k-1}^{n}}\right)\Delta_{k}^{n}B^{\alpha} \\ &\quad+\sum DV_{\beta}\left(Y_{t_{k-1}^{n}}\right)\cdot V_{\alpha}\left(Y_{t_{k-1}^{n}}\right)\int_{t_{k-1}^{n}}^{t_{k}^{n}}B_{t_{k-1}^{n},u}^{\alpha}dB_{u}^{\beta}, \end{aligned}$$

and each sum is over all k such that $\left[t_{k-1}^{n},t_{k}^{n}\right]\subset[s,t]$. It follows that quasi-surely,

$$I_{s,t}^{n}\to Y_{s,t},\ n\to\infty.$$

On the other hand, the following convergence in $L_{G}^{2}\left(\Omega;\mathbb{R}^{N}\right)$ holds:

$$\begin{aligned} \sum b\left(Y_{t_{k-1}^{n}}\right)\Delta t^{n} &\to \int_{s}^{t}b\left(Y_{u}\right)du, \\ \sum h_{\alpha\beta}\left(Y_{t_{k-1}^{n}}\right)\Delta_{k}^{n}\left\langle B^{\alpha},B^{\beta}\right\rangle &\to \int_{s}^{t}h_{\alpha\beta}\left(Y_{u}\right)d\left\langle B^{\alpha},B^{\beta}\right\rangle_{u}, \\ \sum V_{\alpha}\left(Y_{t_{k-1}^{n}}\right)\Delta_{k}^{n}B^{\alpha} &\to \int_{s}^{t}V_{\alpha}\left(Y_{u}\right)dB_{u}^{\alpha}, \end{aligned}$$

as $n\to\infty$.

The reason is the following. For simplicity we only consider the third one, as the first two are similar (and in fact easier). It is straight forward that

$$\begin{aligned} &\int_{0}^{1}\left|V_{\alpha}\left(Y_{t}\right)-\sum_{k=1}^{2^{n}}V_{\alpha}\left(Y_{t_{k-1}^{n}}\right)\mathbf{1}_{[t_{k-1}^{n},t_{k}^{n})}(t)\right|^{2}dt \\ &=\sum_{k=1}^{2^{n}}\int_{t_{k-1}^{n}}^{t_{k}^{n}}\left|V_{\alpha}\left(Y_{t}\right)-V_{\alpha}\left(Y_{t_{k-1}^{n}}\right)\right|^{2}dt \\ &\leqslant C\sum_{k=1}^{2^{n}}\int_{t_{k-1}^{n}}^{t_{k}^{n}}\left|Y_{t}-Y_{t_{k-1}^{n}}\right|^{2}dt \\ &\leqslant C\sum_{k=1}^{2^{n}}\|Y\|_{p;\left[t_{k-1}^{n},t_{k}^{n}\right]}^{2}\Delta t^{n} \\ &\leqslant C\left(\sum_{k=1}^{2^{n}}\|Y\|_{p;\left[t_{k-1}^{n},t_{k}^{n}\right]}^{p}\Delta t^{n}\right)^{\frac{2}{p}} \\ &\leqslant C\left(\Delta t^{n}\right)^{\frac{2}{p}}\|Y\|_{p;[0,1]}^{2}, \end{aligned}$$

where C depends only on V_{α}. Therefore, it suffices to show that $\|Y\|_{p-var;[0,1]}\in L_{G}^{2}(\Omega)$, as it implies the G-Itô integrability of $V_{\alpha}\left(Y_{t}\right)$ and the desired convergence in

$L_G^2\left(\Omega;\mathbb{R}^N\right)$ holds. For simplicity we assume that Y_t is the solution to the following RDE

$$dY_t = V_\alpha\left(Y_t\right) dB_t^\alpha$$

with $Y_0=\xi$ (there is no substantial difference because dt and $d\left\langle B^\alpha, B^\beta\right\rangle_t$ are more regular than dB_t), then by [26], Theorem 10. 14, we know that

$$\|Y\|_{p;[0,1]} \leqslant C\,\|\mathbf{B}\|_{p;[0,1]} \vee \|\mathbf{B}\|_{p;[0,1]}^p .$$

Therefore, we only need to show that $\|\mathbf{B}\|_{p;[0,1]}^p \in L_G^2(\Omega)$. For this purpose, we use Proposition 5.4.1 to control the p-variation norm by the functions ρ_1,ρ_2 defined in (5.4.2). It follows that

$$\|\mathbf{B}\|_p \leqslant C(1+\rho_1(\mathbf{B})^2+\rho_2(\mathbf{B})).$$

Therefore, it remains to show that $\rho_1(\mathbf{B})^{2p}, \rho_2(\mathbf{B})^p \in L_G^1(\Omega)$. First consider level one. By the distribution of B_t, we have

$$\begin{aligned}
\left\|\sum_{n=1}^{\infty} n^\gamma \sum_{k=1}^{2^n}\left|B^1_{t^n_{k-1},t^n_k}\right|^p\right\|_2 &\leqslant \sum_{n=1}^{\infty} n^\gamma \sum_{k=1}^{2^n}\left\|\left|B^1_{t^n_{k-1},t^n_k}\right|^p\right\|_2 \\
&\leqslant \sum_{n=1}^{\infty} n^\gamma\left(\Delta t^n\right)^{\frac{p}{2}-1} \\
&< \infty,
\end{aligned}$$

and we know that $\rho_1(\mathbf{B})^{2p} \in L_G^1(\Omega)$. Now consider level two. By Proposition 5.4.2 and the distribution of B_t and $\langle B,B\rangle_t$, we have

$$\begin{aligned}
\left\|\sum_{n=1}^{\infty} n^\gamma \sum_{k=1}^{2^n}\left|B^2_{t^n_{k-1},t^n_k}\right|^{\frac{p}{2}}\right\|_2 &= \left\|\sum_{n=1}^{\infty} n^\gamma \sum_{k=1}^{2^n}\left|\int_{t^n_{k-1}}^{t^n_k} B_{t^n_{k-1},u}\otimes dB_u + \frac{1}{2}\langle B,B\rangle_{t^n_{k-1},t^n_k}\right|^{\frac{p}{2}}\right\|_2 \\
&\leqslant \sum_{n=1}^{\infty} n^\gamma \sum_{k=1}^{2^n}\left\|\left|\int_{t^n_{k-1}}^{t^n_k} B_{t^n_{k-1},u}\otimes dB_u + \frac{1}{2}\langle B,B\rangle_{t^n_{k-1},t^n_k}\right|^{\frac{p}{2}}\right\|_2 \\
&\leqslant C\sum_{n=1}^{\infty} n^\gamma\left(\Delta t^n\right)^{\frac{p}{2}-1} \\
&< \infty.
\end{aligned}$$

It follows that $\rho_2(\mathbf{B})^p \in L_G^1(\Omega)$. Therefore, the desired L_G^2-convergence holds.

In addition, by Lemma 5.5.1 we also have the following L_G^2-convergence:

$$\sum DV_\beta\left(Y_{t_{k-1}^n}\right)\cdot V_\alpha\left(Y_{t_{k-1}^n}\right)\int_{t_{k-1}^n}^{t_k^n}B_{t_{k-1}^n,u}^\alpha dB_u^\beta\to 0,\ n\to\infty.$$

Consequently, in $L_G^2\left(\Omega;\mathbb{R}^N\right)$,

$$I_{s,t}^n\to\int_s^t b\left(Y_u\right)du+\int_s^t h_{\alpha\beta}\left(Y_u\right)d\left\langle B^\alpha,B^\beta\right\rangle_u+\int_s^t V_\alpha\left(Y_u\right)dB_u^\alpha,$$

as $n\to\infty$.

From Lemma 5.4.3, we conclude that

$$Y_{s,t}=\int_s^t b\left(Y_u\right)du+\int_s^t h_{\alpha\beta}\left(Y_u\right)d\left\langle B^\alpha,B^\beta\right\rangle_u+\int_s^t V_\alpha\left(Y_u\right)dB_u^\alpha$$

quasi-surely. Since X_t and Y_t are both quasi-surely continuous, it follows that X coincides with Y quasi-surely. □

Remark 5.5.1. As we have mentioned at the beginning of Section 5.3, it is possible to prove Theorem 5.5.1 by establishing the Wong-Zakai type approximation. More precisely, if we let X_t^n to be the Euler-Maruyama approximation of the SDE (5.5.1) and let Y_t^n to be the unique classical solution to the following ODE:

$$dY_t^n=b\left(Y_t^n\right)dt+\left(h_{\alpha\beta}\left(Y_t^n\right)-\frac{1}{2}DV_\beta\left(Y_t^n\right)\cdot V_\alpha\left(Y_t^n\right)\right)d\left\langle B^\alpha,B^\beta\right\rangle_t+V_\alpha\left(Y_t^n\right)d\left(B^n\right)_t^\alpha$$

with $X_0^n=Y_0^n=\xi$, where B_t^n is the dyadic piecewise linear interpolation of B_t, then by using our main result in Section 5.3 and establishing related L_G^2-estimates, we can prove that

$$\sup_{t\in[0,1]}\mathbb{E}^G\left[\left|X_t^n-Y_t^n\right|^2\right]\leqslant C\sqrt{1+\xi^2}\left(\Delta t^n\right)^{\frac{1}{2}}.$$

In other words, Y_t^n converges to the solution X_t to the SDE (5.5.1) in the L_G^2-sense. However, we know that Y_t^n converges uniformly to the solution Y_t to the RDE (5.5.6) quasi-surely. Again by Lemma 5.4.3 and continuity, we conclude that X coincides with Y quasi-surely.

From the above discussion, if we forget the RDE (5.5.6) and only consider the L_G^2-limit of Y_t^n, it seems that there is nothing to do with rough paths at all as everything is well-defined in the classical sense. However, the crucial point of understanding the convergence of Y_t^n in the pathwise sense lies in the fact that B_t can be regarded as a geometric rough path (i.e. the lifting defined in Section 5.4) with approximating sequence in $G\Omega_p\left(\mathbb{R}^d\right)$ being the lifting of the natural dyadic piecewise linear interpolation B_t^n. This is exactly what the universal limit theorem tells us.

Remark 5.5.2. From the RDE point of view, it is possible to reduce the regularity assumptions on the coefficients. In particular, since the regularity of t and $\langle B^\alpha, B^\beta\rangle_t$ are both "better" than B_t, the regularity assumptions on the coefficients of dt and $d\left\langle B^\alpha, B^\beta\right\rangle_t$ can be weaker than the one imposed on the coefficient of dB_t. However, we are not going to present the results in such generality. We refer the reader to [26] for general existence and uniqueness results for RDEs.

5.6 SDEs on a Differentiable Manifold Driven by G-Brownian Motion

Our main result in Section 5.5 can be used to construct SDEs on a differentiable manifold driven by G-Brownian motion, which is the main focus of this section. Our development is based on the idea in the classical case, for which the reader may refer to the monographs by K.D. Elworthy [22], E.P. Hsu [36], N. Ikeda and S.Watanabe [37]. This part is the foundation of constructing G-Brownian motion on a Riemannian manifold in the next section.

In classical stochastic analysis, SDEs on a manifold are constructed using the Stratonovich type formulation, which can be regarded as a pathwise approach. The reason for using the Stratonovich type formulation instead of the Itô type one is the following. First of all, the notion of SDE can be introduced by using test functions on the manifold from an intrinsic point of view, which is consistent with ordinary differential calculus and invariant under diffeomorphisms. Moreover, when we construct solutions extrinsically, we can prove that almost-surely, the solutions to the extended SDEs which start on the manifold always live on it. This reveals the intrinsic nature of ordinary differential equations.

In the setting of G-expectations, we adopt the same idea for the development. However, there is a major difficulty in this situation. The method of constructing solutions in the classical case from the extrinsic point of view depends heavily on the localization technique, which is not available in the setting of G-expectations, mainly due to the reason that concepts of information flow and stopping times are not well understood. To get around this difficulty, we use our main result in Section 5.5 to obtain a pathwise construction. The advantage of such an approach is that we can still use localization arguments but do not need to care about measurability and integrability issues under G-expectation.

Now assume that M is a differentiable manifold. For technical reasons we further assume that M is compact (it is not necessary if we impose restrictive geometric conditions on the manifold and the generating vector fields). Let $\{b, h_{\alpha\beta}, V_\alpha : \alpha, \beta = 1, \cdots, d\}$ be a family of C^3-vector fields on M, and let B_t be the canonical d-dimensional G-Brownian motion on the path space $(\Omega, L^2_G(\Omega), \mathbb{E}^G)$, where G is a function given by (5.2.2).

Consider the following symbolic Stratonovich type SDE over $[0,1]$:

$$\begin{cases} dX_t &= b\left(X_t\right)dt + h_{\alpha\beta}\left(X_t\right)d\left\langle B^\alpha, B^\beta\right\rangle_t + V_\alpha(X_t)\circ dB^\alpha_t, \\ X_0 &= \xi \in M, \end{cases} \tag{5.6.1}$$

on M.

Definition 5.6.1. A solution X_t to the SDE (5.6.1) is an M-valued continuous stochastic process such that for any $f \in C^\infty(M)$ and $\alpha, \beta = 1, \cdots, d$,

$$\{h_{\alpha\beta}f(X_t) : t \in [0,1]\} \in M^1_G(0,1),\ \{V_\alpha f(X_t) : t \in [0,1]\} \in M^2_G(0,1),$$

and the following equality holds on $[0,1]$:

$$f(X_t) = f(\xi) + \int_0^t bf(X_s)ds + \int_0^t h_{\alpha\beta}f(X_s)d\left\langle B^\alpha, B^\beta\right\rangle_s + \int_0^t V_\alpha f(X_s)\circ dB^\alpha_s, \tag{5.6.2}$$

where the last term is defined as

$$\int_0^t V_\alpha f(X_s)\circ dB^\alpha_s := \int_0^t V_\alpha f(X_s)dB^\alpha_s + \frac{1}{2}\int_0^t V_\beta V_\alpha f(X_s)d\left\langle B^\alpha, B^\beta\right\rangle_s.$$

Remark 5.6.1. Definition 5.6.1 is intrinsic. It is easy to see that Definition 5.6.1 is consistent with the Euclidean case.

Now we construct the solution to (5.6.1) from the extrinsic point of view.

According to the Whitney embedding theorem (see the monograph by G. de Rham [17]), M can be embedded into some ambient Euclidean space $\mathbb{R}^N$ as a submanifold such that the image $i(M)$ of M is closed in $\mathbb{R}^N$. We simply regard M as a subset of $\mathbb{R}^N$.

Let $F^1, \cdots, F^N \in C^\infty(M)$ be the coordinate functions on M. The following result is easy to prove. It is similar to the classical case.

Proposition 5.6.1. *X_t is a solution to (5.6.1) if and only if for any $i = 1, \cdots, N$, $\alpha, \beta = 1, \cdots, d$,*

$$\{h_{\alpha\beta}F^i(X_t) : t \in [0,1]\} \in M_G^1(0,1), \ \{V_\alpha F^i(X_t) : t \in [0,1]\} \in M_G^2(0,1),$$

and for any $t \in [0,1]$,

$$F^i(X_t) = F^i(\xi) + \int_0^t bF^i(X_s)ds + \int_0^t h_{\alpha\beta}F^i(X_s)d\left\langle B^\alpha, B^\beta\right\rangle_s + \int_0^t V_\alpha F^i(X_s) \circ dB_s^\alpha. \tag{5.6.3}$$

Proof. Necessity is obvious since $F^i \in C^\infty(M)$ for any $i = 1, 2, \cdots, N$.

Now consider sufficiency. Let $f \in C^\infty(M)$, and choose a C^∞-extension $\widetilde{f}$ of f with compact support in $\mathbb{R}^N$ (it is possible since M is compact). Then for any $x \in M$,

$$f(x) = \widetilde{f}\left(F^1(x), \cdots, F^N(x)\right),$$

and thus

$$f(X_t) = \widetilde{f}\left(F^1(X_t), \cdots, F^N(X_t)\right), \ \forall t \in [0,1].$$

Since M is compact and $\widetilde{f}$ is smooth with compact support, it follows from the G-Itô formula that for $t \in [0,1]$,

$$\begin{aligned} & \widetilde{f}\left(F^1(X_t), \cdots, F^N(X_t)\right) \\ = \ & f(\xi) + \int_0^t \frac{\partial \widetilde{f}}{\partial y^i}\left(bF^i(X_s)ds + h_{\alpha\beta}F^i(X_s)d\left\langle B^\alpha, B^\beta\right\rangle_s \right. \\ & \left. + V_\alpha F^i(X_s) \circ dB_s^\alpha\right) \\ = \ & f(\xi) + \int_0^t \left(bf(X_s)ds + h_{\alpha\beta}f(X_s)d\left\langle B^\alpha, B^\beta\right\rangle_s + V_\alpha f(X_s) \circ dB_s^\alpha\right), \end{aligned}$$

where we have used the simple fact that for any C^1-vector field V on M,

$$Vf = \sum_{i=1}^N \frac{\partial \widetilde{f}}{\partial y^i} VF^i.$$

By Definition 5.6.1, we know that X_t is a solution to the SDE (5.6.1). □

Now we prove the existence and uniqueness of (5.6.1) by using the main result of Section 5.5, namely, a pathwise approach based on the associated RDE.

Let $\widetilde{b}, \widetilde{h}_{\alpha\beta}, \widetilde{V}_\alpha$ be C_b^3-extensions (not unique) of the vector fields $b, h_{\alpha\beta}, V_\alpha$. Consider the following Stratonovich type SDE in the ambient space $\mathbb{R}^N$:

$$dX_t = \widetilde{b}(X_t)dt + \widetilde{h}_{\alpha\beta}(X_t)d\left\langle B^\alpha, B^\beta\right\rangle_t + \widetilde{V}_\alpha(X_t) \circ dB_t^\alpha \tag{5.6.4}$$

with $X_0 = x \in \mathbb{R}^N$, which is interpreted as the following Itô type SDE:

$$dX_t = \widetilde{b}(X_t)dt + \left(\widetilde{h}_{\alpha\beta}(X_t) + \frac{1}{2}D\widetilde{V}_\alpha(X_t) \cdot \widetilde{V}_\beta(X_t)\right) d\left\langle B^\alpha, B^\beta\right\rangle_t + \widetilde{V}_\alpha(X_t)dB_t^\alpha.$$

According to Section 5.5, we can alternatively interpret (5.6.4) as an RDE which is defined pathwisely. Both the SDE and the RDE have unique solutions, and according to Theorem 5.5.1 they coincide quasi-surely. Our aim is to show that quasi-surely, the solution X_t to (5.6.4) never leaves M and it is the unique solution to (5.6.1).

The following result is important to prove the existence and uniqueness of the SDE (5.6.1) on the manifold M.

Proposition 5.6.2. *Let x_t be a continuous path with bounded total variation in $\mathbb{R}^d$, and let $W_1, \cdots, W_d$ be a family of C^1-vector fields on M and $\widetilde{W}_1, \cdots, \widetilde{W}_d$ be their C_b^1-extensions to $\mathbb{R}^N$. Consider the following ODE in the ambient space $\mathbb{R}^N$ over $[0,1]$:*

$$dy_t = \widetilde{W}_\alpha(y_t)dx_t^\alpha \tag{5.6.5}$$

with $y_0 = x \in M$. Then the solution $y_t \in M$ for all $t \in [0,1]$. Moreover, y_t does not depend on extensions of the vector fields.

Proof. Let $F(x) := d(x, M)^2$ be the squared distance function to the submanifold M. It follows that F is smooth in an open neighborhood of M. By using the cut-off function we may assume that $F \in C_b^\infty(M)$. Now we are able to choose an open neighborhood U of M, such that for any $x \in U$, $F(x) = 0$ if and only if $x \in M$. Moreover, since $\widetilde{W}_\alpha$ $(\alpha = 1, 2, \cdots, d)$ are tangent vector fields of M when restricted to M, U can be chosen such that for any $x \in U$ and $\alpha = 1, 2, \cdots, d$,

$$\left|\widetilde{W}_\alpha F(x)\right| \leqslant CF(x), \tag{5.6.6}$$

for some positive constant C depending on U. The function $F(x)$ was used in [36] to construct SDEs on M driven by classical Brownian motion.

Since x_t is a path with bounded total variation and $y_0 = \xi \in M$, by the change of variables formula in ordinary calculus, we have

$$F(y_t) = \int_0^t \widetilde{W}_\alpha F(y_s) dx_s^\alpha, \ \forall t \in [0,1].$$

Define $\tau := \inf\{t \in [0,1] : \ y_t \notin U\}$. It follows from (5.6.6) that

$$F(y_t) \leqslant C \int_0^t F(y_s) d|x|_s, \ \forall t \in [0,\tau],$$

where $|x|_t$ is the total variation of the path x_t.

By iteration and Fubini's theorem, on $[0,\tau]$ we have

$$\begin{aligned} F(y_t) &\leqslant C^2 \int_0^t \left(\int_0^s F(y_u) d|x|_u \right) d|x|_s \\ &= C^2 \int_0^t \left(|x|_t - |x|_s \right) F(y_s) d|x|_s. \end{aligned}$$

By induction, it is easy to see that for any $k \geqslant 1$,

$$F(y_t) \leqslant C^k \int_0^t \frac{\left(|x|_t - |x|_s \right)^{k-1}}{(k-1)!} F(y_s) d|x|_s, \ \forall t \in [0,\tau].$$

Since F is bounded, we obtain further that for any $k \geqslant 1$,

$$F(y_t) \leqslant \|F\|_\infty \frac{C^k (|x|_t - |x|_0)^k}{k!}, \ \forall t \in [0,\tau].$$

By letting $k \to \infty$, it follows that $F(y_t) \equiv 0$ on $[0,\tau]$, which implies that $y_t \in M$ for any $t \in [0,\tau]$. Since y_t is continuous, the only possibility is that y_t never leaves M on $[0,1]$.

If we rewrite the ODE (5.6.5) in its integral form:

$$y_t = \xi + \int_0^t \widetilde{W}_\alpha(y_s) dx_s^\alpha, \ t \in [0,1], \tag{5.6.7}$$

we know from the previous discussion that equation (5.6.7) depends only on the values of $\widetilde{W}_\alpha$ on M, that is, of W_α $(\alpha = 1, 2, \cdots, d)$. In other words, if $\hat{W}_\alpha$ is another extension of W_α and $\widehat{y}_t$ is the solution to the corresponding ODE with the same initial condition, $\widehat{y}_t$ is also a solution to (5.6.5). By uniqueness, we have $y = \widehat{y}$. Therefore, y_t does not depend on the extensions of the vector fields. □

With the help of Proposition 5.6.2, we can prove the following existence and uniqueness result.

Theorem 5.6.1. *Let $b, h_{\alpha\beta}, V_\alpha$ be C^3-vector fields on M. Then the Stratonovich type SDE (5.6.1) has a solution X_t which is unique quasi surely.*

Proof. Fix C_b^3-extensions $\widetilde{b}, \widetilde{h}_{\alpha\beta}, \widetilde{V}_\alpha$ of $b, h_{\alpha\beta}, V_\alpha$, and let X_t be the solution to the Stratonovich type SDE (5.6.4) in $\mathbb{R}^N$ over $[0,1]$. By Theorem 5.5.1, quasi-surely X_t coincides with the solution to (5.6.4) when it is interpreted as an RDE. Since M is closed in $\mathbb{R}^N$, it follows from Proposition 5.6.2 and Theorem 1.2.2 (the universal limit theorem) that quasi-surely, X_t never leaves M over $[0,1]$. In this case, (5.6.4) is equivalent to (5.6.3), which implies from Proposition 5.6.1 that X_t is a solution to (5.6.1). On the other hand, if Y_t is another solution to (5.6.1), then it is a solution to (5.6.4) (interpreted as an SDE or an RDE). By the uniqueness of RDEs, we know that $X = Y$ quasi-surely. □

Remark 5.6.2. It is possible to formulate uniqueness in the L_G^2-sense when M is regarded as a closed submanifold of $\mathbb{R}^N$. However, we use the quasi sure formulation because the notion itself is intrinsic although the proof is developed from the extrinsic point of view.

5.7 G-Brownian Motion on a Compact Riemannian Manifold and the Generating Nonlinear Heat Flow

In this section, we construct G-Brownian motion on a compact Riemannian manifold for a wide and interesting class of G-functions, based on J. Eells, K.D. Elworthy and P. Malliavin's horizontal lifting construction (see [22], [36], [37] for the construction of Brownian motion on a Riemannian manifold and related topics). Roughly speaking, we "roll" an Euclidean G-Brownian motion up to the Riemannian manifold "without slipping" via a proper frame bundle (for the class of G-functions we are interested in, such a bundle is the orthonormal frame bundle).

It should be pointed out that, unlike the classical case, the non-compact situation becomes much more complicated as we may encounter issues of integrability and localization under G-expectation when explosion is taken into account. In particular, the notion of localization and random times is not well understood under G-expectation. Here we only consider the compact case and leave the study of explosion in the non-compact case for future research.

In the classical case, we know that the law of a d-dimensional Brownian motion B_t is invariant under orthogonal transformations on $\mathbb{R}^d$. This is a crucial point to obtain a linear parabolic PDE (the standard heat equation associated with the Bochner horizontal Laplacian $\Delta_{\mathcal{O}(M)}$) on the orthonormal frame bundle $\mathcal{O}(M)$ over a Riemannian manifold M governing the law of the horizontal lifting ξ_t of B_t to $\mathcal{O}(M)$, which is invariant under orthogonal transformations along fibers. It is such an invariance that enables us to "project" the PDE down to the base manifold M and obtain the standard heat equation associated with the Laplace-Beltrami operator Δ_M on M. This heat equation governs the law of the development $X_t = \pi(\xi_t)$ of B_t to the Riemannian manifold M via the horizontal lifting ξ_t. As a stochastic process on M, although X_t depends on the initial orthonormal frame ξ at x as well as the initial position $x \in M$, the law of X_t depends only on the initial position x, and it is characterized by the Laplace-Beltrami operator Δ_M via the heat equation. Equivalently, it can be shown that the law of X_t is the unique solution to the martingale problem on M associated with Δ_M starting at x. X_t is called the Brownian motion on M starting at x in the sense of Eells, Elworthy and Malliavin.

It is quite natural to expect that the Brownian sample paths X_t on M depend on the initial orthonormal frame ξ at x if we look back into the Euclidean case, in which we actually fix the standard orthonormal basis in advance and define Brownian motion in the corresponding coordinate system. If we use another orthonormal basis, we obtain a process (still a Brownian motion) which is an orthogonal transformation of the original Brownian motion. Therefore, it is the law, which is characterized by the Laplace operator on $\mathbb{R}^d$, rather than the sample paths that captures the intrinsic nature of the Brownian motion, and this idea can be developed in a Riemannian geometric setting.

It should be remarked that in a pathwise manner, we can lift B_t horizontally to the total frame bundle $\mathcal{F}(M)$ instead of $\mathcal{O}(M)$ by solving the same SDE generated by the horizontal vector fields but using a general frame instead of an orthonormal one as initial condition. Moreover, we can write down the generating heat equation on $\mathcal{F}(M)$ which takes the same form as the one on $\mathcal{O}(M)$. The key difference here is that although the horizontal lifting of B_t can be projected onto M, the heat equation on $\mathcal{F}(M)$ cannot. In other words, the heat equation is not invariant under non-degenerate linear transformations along fibers. This becomes less interesting to us, as we are not able to obtain an intrinsic law of the development of B_t on M which is independent of initial frames. The fundamental reason for using the orthonormal frame bundle is that the Laplace operator on $\mathbb{R}^d$ is invariant exactly under orthogonal transformations.

The case of G-Brownian motion can be studied in a similar manner. From the last section we are able to solve SDEs on a differentiable manifold (in particular, on $\mathcal{F}(M)$) driven by an Euclidean G-Brownian motion B_t. By projection we obtain the development X_t of B_t to M. As we have pointed out before, such a development is not interesting unless we are able to prove that the law of X_t depends only on the initial position x rather than the initial frame. In fact, if the law of X_t depends on the initial frame, we might not be able to write down the generating PDE of X_t intrinsically on M although it is possible on $\mathcal{F}(M)$. Therefore, for a given G-function, it is crucial to identify a proper frame bundle over M with a specific structure group such that parallel transport preserves fibers and the generating PDE of the horizontal lifting ξ_t of B_t to such a frame bundle is invariant under actions by the structure group along fibers. It follows that the law of X_t is independent of initial frames in the fiber over x (x is the starting point of X_t) and we should be able to obtain the generating PDE of X_t, which is associated with G and intrinsically defined on M.

As we will see, such an idea depends on a crucial algebraic quantity associated with the G-function called the invariant group $I(G)$ of G, which will be defined later on. In this section, we are mainly interested in the case when $I(G)$ is the orthogonal group. We will see that it includes a wide class of G-functions. In particular, one example is the generalization of the one-dimensional Barenblatt equation to higher dimensions.

The concept of the invariant group of G is motivated from the study of the infinitesimal diffusive nature of SDEs driven by G-Brownian motion and their generating PDEs, which is discussed below.

We first consider the Euclidean case.

From now on, we always assume that $G: S(d) \to \mathbb{R}$ is a given continuous, sublinear and monotonic function. Equivalently, from Section 5.2 we know that G is represented by

$$G(A) = \frac{1}{2} \sup_{B \in \Sigma} \mathrm{tr}(AB), \ \forall A \in S(d), \tag{5.7.1}$$

where Σ is some bounded, closed and convex subset of $S_+(d)$. Let B_t be the standard d-dimensional G-Brownian motion on the path space.

Assume that $V_1, \cdots, V_d$ are C_b^3 -vector fields on $\mathbb{R}^N$. Consider the following N-

dimensional Stratonovich type SDE over $[0,1]$:

$$\begin{cases} dX_{t,x} = V_\alpha(X_{t,x}) \circ dB_t^\alpha, \\ X_{0,x} = x, \end{cases} \tag{5.7.2}$$

which is either interpreted as an RDE or the associated Itô type SDE

$$\begin{cases} dX_{t,x} = V_\alpha(X_{t,x}) dB_t^\alpha + \frac{1}{2} DV_\alpha(X_{t,x}) V_\beta(X_{t,x}) \left\langle B^\alpha, B^\beta \right\rangle_t, \\ X_{t,x} = x, \end{cases}$$

according to the main result of Section 5.5.

The following result characterizes the generator of the SDE (5.7.2) in terms of G. It describes the infinitesimal diffusive nature of (5.7.2). One might compare it with the case of linear diffusion processes.

Proposition 5.7.1. *For any* $p \in \mathbb{R}^N,\ A \in S(N)$,

$$\begin{aligned} &\lim_{\delta \to 0^+} \frac{1}{\delta} \mathbb{E}^G \left[\langle p, X_{\delta,x} - x \rangle + \frac{1}{2} \langle A(X_{\delta,x} - x), X_{\delta,x} - x \rangle \right] \\ =&G\left(\left(\frac{1}{2} \langle p, DV_\alpha(x) V_\beta(x) + DV_\beta(x) V_\alpha(x) \rangle + \langle AV_\alpha(x), V_\beta(x) \rangle \right)_{1 \leqslant \alpha, \beta \leqslant d} \right). \end{aligned} \tag{5.7.3}$$

Proof. From the distribution of B_t we know that

$$\begin{aligned} G(A) &= \frac{1}{2} \mathbb{E}^G \left[\langle AB_1, B_1 \rangle \right] \\ &= \frac{1}{2t} \mathbb{E}^G \left[\langle AB_t, B_t \rangle \right],\ \forall t > 0. \end{aligned}$$

Therefore, the R.H.S. of (5.7.3) is equal to

$$I_\delta = \frac{1}{2\delta} \mathbb{E}^G \left[\left(\langle p, DV_\alpha(x) V_\beta(x) \rangle + \langle AV_\alpha(x), V_\beta(x) \rangle \right) B_\delta^\alpha B_\delta^\beta \right],$$

for any $\delta > 0$.

Since

$$X_{\delta,x} - x = \int_0^\delta V_\alpha(X_{s,x}) dB_s^\alpha + \frac{1}{2} \int_0^\delta DV_\alpha(X_{s,x}) V_\beta(X_{s,x}) d\left\langle B^\alpha, B^\beta \right\rangle_s,$$

by the properties of $\mathbb{E}^G$ and the distribution of B_t, we have

$$\begin{aligned}
&\left|\frac{1}{\delta}\mathbb{E}^G\left[\langle p, X_{\delta,x}-x\rangle+\frac{1}{2}\langle A(X_{\delta,x}-x), X_{\delta,x}-x\rangle\right]-I_\delta\right| \\
\leqslant&\left|\frac{1}{2\delta}\mathbb{E}^G\left[\int_0^\delta \langle p, DV_\alpha(X_{s,x})\cdot V_\beta(X_{s,x})\rangle\, d\left\langle B^\alpha, B^\beta\right\rangle_s\right.\right. \\
&\left.+\left\langle A\int_0^\delta V_\alpha(X_{s,x})dB_s^\alpha, \int_0^\delta V_\beta(X_{s,x})dB_s^\beta\right\rangle\right] \\
&\left.-\frac{1}{2\delta}\mathbb{E}^G\left[\langle p, DV_\alpha(x)\cdot V_\beta(x)\rangle\left\langle B^\alpha, B^\beta\right\rangle_\delta+\left\langle AV_\alpha(x)B_\delta^\alpha, V_\beta(x)B_\delta^\beta\right\rangle\right]\right|+C\delta^{\frac{1}{2}}+C\delta \\
\leqslant&\frac{1}{2\delta}\left(C\int_0^\delta\sqrt{\mathbb{E}^G\left[|X_{s,x}-x|^2\right]}ds+C\int_0^\delta \mathbb{E}^G\left[|X_{s,x}-x|^2\right]ds\right. \\
&\left.+C\delta^{\frac{1}{2}}\sqrt{\int_0^\delta \mathbb{E}^G\left[|X_{s,x}-x|^2\right]ds}\right)+C\delta^{\frac{1}{2}}+C\delta,
\end{aligned}$$

where we have also used the fact that G-Itô integrals and $B_\delta^\alpha B_\delta^\beta-\langle B^\alpha, B^\beta\rangle_\delta$ do not have mean uncertainty. Here C always denotes positive constants independent of δ.

Now the result follows easily from the fact that

$$\mathbb{E}^G[|X_{t,x}-x|^2]\leqslant Ct,\ \forall t\in[0,1].$$

□

The infinitesimal diffusive nature of (5.7.2) characterized by Proposition 5.7.1 enables us to establish the generating PDE of (5.7.2) in terms of viscosity solutions. The understanding of this PDE, especially its intrinsic nature, is essential for the development in a geometric setting.

Theorem 5.7.1. *Let $\varphi\in C_b^\infty\left(\mathbb{R}^N\right)$, and define*

$$u(t,x)=\mathbb{E}^G[\varphi(X_{t,x})],\ (t,x)\in[0,1]\times\mathbb{R}^N.$$

Then $u(t,x)$ is the unique viscosity solution to the following nonlinear parabolic PDE:

$$\begin{cases}\frac{\partial u}{\partial t}-G\left(\left(\widehat{V_\alpha V_\beta}u\right)_{1\leqslant\alpha,\beta\leqslant d}\right)=0, \\ u(0,x)=\varphi(x),\end{cases}\tag{5.7.4}$$

where $\widehat{V_\alpha V_\beta}$ denotes the symmetrization of the second order differential operator $V_\alpha V_\beta$, *that is,*

$$\widehat{V_\alpha V_\beta} = \frac{1}{2}(V_\alpha V_\beta + V_\beta V_\alpha).$$

Proof. The continuity of u in t and x can be shown in a standard way by using the Lipschitz continuity of φ (in fact, u is Lipchitz in x and 1/2-Hölder continuous in t). Here the proof is omitted.

Fix $(t_0, x_0) \in (0,1) \times \mathbb{R}^N$. Let $v(t,x) \in C_b^{2,3}\left([0,1] \times \mathbb{R}^N\right)$ be a test function such that

$$u(t_0, x_0) = v(t_0, x_0)$$

and

$$u(t,x) \leqslant v(t,x), \ \forall (t,x) \in [0,1] \times \mathbb{R}^N.$$

For $0 < \delta < t_0$, by the uniqueness of the SDE (5.7.2) and the fact that B_t and $\langle B^\alpha, B^\beta \rangle_t$ have independent and identically distributed increments, we know that

$$\begin{aligned} \mathbb{E}^G\left[\varphi(X_{t_0,x_0})|\Omega_\delta\right] &= \mathbb{E}^G\left[\varphi(X_{\delta,x_0} + \int_\delta^{t_0} V_\alpha\left(X_{s,x_0}\right) dB_s^\alpha \right. \\ &\quad \left. \left. + \frac{1}{2}\int_\delta^{t_0} DV_\alpha\left(X_{s,x_0}\right) \cdot V_\beta\left(X_{s,x_0}\right) d\left\langle B^\alpha, B^\beta\right\rangle_s\right) \right| \Omega_\delta\right] \\ &= \mathbb{E}^G\left[\varphi\left(X_{t_0-\delta,y}\right)\right]\Big|_{y=X_{\delta,x_0}}. \end{aligned}$$

Therefore,

$$\begin{aligned} v(t_0,x_0) &= \mathbb{E}^G\left[\varphi\left(X_{t_0,x_0}\right)\right] \\ &= \mathbb{E}^G\left[\mathbb{E}^G\left[\varphi\left(X_{t_0,x_0}\right)\right|\Omega_\delta\right]\right] \\ &= \mathbb{E}^G\left[u\left(t_0-\delta, X_{\delta,x_0}\right)\right] \\ &\leqslant \mathbb{E}^G\left[v\left(t_0-\delta, X_{\delta,x_0}\right)\right]. \end{aligned}$$

It follows that

$$\begin{aligned} 0 &\leqslant \mathbb{E}^G\left[v\left(t_0-\delta, X_{\delta,x_0}\right) - v(t_0,x_0)\right] \\ &= \mathbb{E}^G\left[v\left(t_0-\delta, X_{\delta,x_0}\right) - v\left(t_0, X_{\delta,x_0}\right) + v\left(t_0, X_{\delta,x_0}\right) - v(t_0,x_0)\right] \\ &= \mathbb{E}^G\left[-\delta\int_0^1 \frac{\partial v}{\partial t}\left(t_0-(1-\alpha)\delta, X_{\delta,x_0}\right) d\alpha + \langle \nabla v(t_0,x_0), X_{\delta,x_0} - x_0\rangle \right. \\ &\quad \left. + \int_0^1\int_0^1 \left\langle \nabla^2 v(t_0, x_0 + \alpha\beta\left(X_{\delta,x_0} - x_0\right))\left(X_{\delta,x_0} - x_0\right), X_{\delta,x_0} - x_0\right\rangle \alpha d\alpha d\beta \right] \end{aligned}$$

$$\begin{aligned}&\leqslant -\delta\frac{\partial v}{\partial t}(t_0,x_0)+\mathbb{E}^G\Big[\langle\nabla v(t_0,x_0),X_{\delta,x_0}-x_0\rangle\\&+\frac{1}{2}\left\langle\nabla^2 v(t_0,x_0)\left(X_{\delta,x_0}-x_0\right),X_{\delta,x_0}-x_0\right\rangle\Big]+\mathbb{E}^G\left[|I_\delta|\right]+\mathbb{E}^G\left[|J_\delta|\right],\end{aligned}$$

where

$$\begin{aligned}I_\delta &= -\delta\int_0^1\left(\frac{\partial v}{\partial t}(t_0-(1-\alpha)\delta,X_{\delta,x_0})-\frac{\partial v}{\partial t}(t_0,x_0)\right)d\alpha,\\ J_\delta &= \int_0^1\int_0^1\big\langle(\nabla^2 v(t_0,x_0+\alpha\beta(X_{\delta,x_0}-x_0))\\&\quad-\nabla^2 v(t_0,x_0))(X_{\delta,x_0}-x_0),X_{\delta,x_0}-x_0\big\rangle\,\alpha d\alpha d\beta.\end{aligned}$$

By a standard argument one can easily show that

$$\mathbb{E}^G[|I_\delta|]+\mathbb{E}^G[|J_\delta|]\leqslant C\delta^{\frac{3}{2}},$$

where C is a positive constant independent of δ. On the other hand, the R.H.S. of (5.7.3) applying to

$$p=\nabla v(t_0,x_0),\ \ A=\nabla^2 v(t_0,x_0),$$

is exactly the same as $G\left(\left(\widehat{V_\alpha V_\beta}v(t_0,x_0)\right)_{1\leqslant\alpha,\beta\leqslant d}\right)$. Therefore, by Proposition 5.7.1, we arrive at

$$\frac{\partial v}{\partial t}(t_0,x_0)-G\left(\left(\widehat{V_\alpha V_\beta}v(t_0,x_0)\right)_{1\leqslant\alpha,\beta\leqslant d}\right)\leqslant 0.$$

Consequently, $u(t,x)$ is a viscosity sub-solution to (5.7.4).

Similarly, one can show that $u(t,x)$ is a viscosity supersolution to (5.7.4). Therefore, $u(t,x)$ is a viscosity solution to (5.7.4).

The reason for uniqueness is the following. Define a function $F:\ \mathbb{R}^N\times\mathbb{R}^N\times S(N)\to\mathbb{R}$ by the R.H.S. of (5.7.3), that is,

$$\begin{aligned}&F(x,p,A)\\=&G\left(\left(\frac{1}{2}\langle p,DV_\alpha(x)\cdot V_\beta(x)+DV_\beta(x)\cdot V_\alpha(x)\rangle+\langle AV_\alpha(x),V_\beta(x)\rangle\right)_{1\leqslant\alpha,\beta\leqslant d}\right),\end{aligned}$$

for $(x,p,A)\in\mathbb{R}^N\times\mathbb{R}^N\times S(N)$. It is easy to prove that F is sublinear in (p,A) and monotonically increasing in $S(N)$, due to the same properties held by G. Moreover, F satisfies the continuity condition (Assumption (G) in [58], Appendix C) for the uniqueness of the associated nonlinear PDE, due to the regularity of the given vector fields V_α. In other words, all properties of G to ensure uniqueness are preserved in F, and the space dependence of F are uniformly controlled. Therefore, according to the uniqueness results (see [15], [58]), the parabolic PDE has a unique viscosity solution, which is given by $u(t,x)$. □

Example 5.7.1. An example which motivates the study of G-Brownian motion on a Riemannian manifold is the following.

Let $Q \in GL(d,\mathbb{R})$, where $GL(d,\mathbb{R})$ is the group of $d \times d$ real invertible matrices. Define $B_t^Q = QB_t$, and for $\varphi \in C_b^\infty(\mathbb{R}^d)$, define

$$u(t,x) = \mathbb{E}^G\left[\varphi\left(x + B_t^Q\right)\right],\ (t,x) \in [0,1] \times \mathbb{R}^d.$$

Then $u(t,x)$ is the unique viscosity solution to the PDE:

$$\begin{cases} \frac{\partial u}{\partial t} - G\left(Q^T \cdot \nabla^2 u \cdot Q\right) = 0, \\ u(0,x) = \varphi(x). \end{cases}$$

In fact, it follows directly from Theorem 5.7.1 if we regard $x + B_t^Q$ as the solution to the SDE over $[0,1]$:

$$\begin{cases} dX_{t,x} = Q_\alpha \circ dB_t^\alpha, \\ X_{0,x} = x, \end{cases} \tag{5.7.5}$$

where $Q = (Q_1, \cdots, Q_d)$, and each Q_α is a constant vector field on $\mathbb{R}^d$ (so the SDE (5.7.5) coincides exactly with the Itô type one).

The result of Theorem 5.7.4 is similar to the discussion of the nonlinear Feynman-Kac formula in [58], in which the solution to a forward-backward SDE is used to represent the viscosity solution to an associated nonlinear backward parabolic PDE. Here the intrinsic nature of (5.7.4) is essential and should be emphasized below.

It is not hard to see that the nonlinear second order differential operator

$$G\left(\left(\widehat{V_\alpha V_\beta}\cdot\right)_{1\leqslant\alpha,\beta\leqslant d}\right)$$

is intrinsically defined on $\mathbb{R}^N$, since $V_1, \cdots, V_d$ are vector fields independent of coordinates. Moreover, in local coordinates it preserves the same properties of the G-function which is defined under the standard coordinate system of $\mathbb{R}^d$. In particular, it shares the same ellipticity as G. Therefore, from our results in Section 5.6, we should be able to establish the generating PDE of a nonlinear diffusion process on a differentiable manifold.

Assume that M is a compact manifold, and $V_1, \cdots, V_d$ are C^3-vector fields on M. According to Section 5.6, the Stratonovich type SDE over $[0,1]$

$$\begin{cases} dX_{t,x} = V_\alpha(X_{t,x}) \circ dB_t^\alpha, \\ X_{0,x} = x \in M, \end{cases} \tag{5.7.6}$$

has a unique solution. The following result is immediate from Theorem 5.7.1.

Theorem 5.7.2. *Let $\varphi \in C^\infty(M)$, and define*

$$u(t,x) = \mathbb{E}^G[\varphi(X_{t,x})], \ (t,x) \in [0,1] \times M,$$

then $u(t,x)$ is the unique viscosity solution to the following nonlinear parabolic PDE on M:

$$\begin{cases} \frac{\partial u}{\partial t} - G\left(\left(\widehat{V_\alpha V_\beta} u\right)_{1 \leqslant \alpha,\beta \leqslant d}\right) = 0, \\ u(0,x) = \varphi(x), \end{cases} \tag{5.7.7}$$

where $\widehat{V_\alpha V_\beta}$ is the symmetrization of $V_\alpha V_\beta$, defined in the same way as in Theorem 5.7.1. Here the notion of viscosity solutions to the PDE (5.7.7) can be defined in the same way as in the Euclidean case by using test functions (see D. Azagra, J. Ferrera and B. Sanz [1]).

Proof. The result follows easily from an extrinsic point of view.

In fact, assume that M is embedded into an ambient Euclidean space $\mathbb{R}^N$ as a closed submanifold, and take a C^3-extension $\widetilde{V}_\alpha$ of V_α with compact support. Consider the following Stratonovich type SDE over $[0,1]$:

$$\begin{cases} dX_{t,x} = \widetilde{V}_\alpha(X_{t,x}) \circ dB_t^\alpha, \\ X_{0,x} = x \in \mathbb{R}^N. \end{cases}$$

Let $\widetilde{\varphi}$ be a C^∞-extension of φ with compact support, and define

$$\widetilde{u}(t,x) = \mathbb{E}^G[\widetilde{\varphi}(X_{t,x})], \ (t,x) \in [0,1] \times \mathbb{R}^N.$$

It follows from Theorem 5.7.1 that $\widetilde{u}(t,x)$ is the unique viscosity solution to the nonlinear parabolic PDE generated by the vector fields $\widetilde{V}_\alpha$.

According to Section 5.6, if $x \in M$, $X_{t,x}$ never leaves M quasi-surely. Therefore, when restricted to M, $\widetilde{u} = u$. In particular, we know that u is continuous. To see that u is a viscosity sub-solution to (5.7.7), let $(t_0, x_0) \in (0,1) \times M$, and $v(t,x) \in C^{2,3}([0,1] \times M)$ be a test function such that

$$v(t_0, x_0) = u(t_0, x_0)$$

and

$$u(t,x) \leqslant v(t,x), \ \forall (t,x) \in [0,1] \times M.$$

Take an $C_b^{2,3}$-extension $\widetilde{v}$ of v such that

$$\widetilde{u}(t,x) \leqslant \widetilde{v}(t,x), \ \forall (t,x) \in [0,1] \times \mathbb{R}^N.$$

It follows from previous discussion that

$$\frac{\partial \widetilde{v}}{\partial t}(t_0, x_0) - G\left(\left(\widehat{\widetilde{V}_\alpha \widetilde{V}_\beta} \widetilde{v}(t_0, x_0)\right)_{1 \leqslant \alpha, \beta \leqslant d}\right) \leqslant 0.$$

Since

$$\widetilde{V}_\alpha|_M = V_\alpha, \ \widetilde{v}|_M = v,$$

from the intrinsic nature of the generating PDE, we know that

$$\frac{\partial \widetilde{v}}{\partial t}(t_0, x_0) = \frac{\partial v}{\partial t}(t_0, x_0)$$

and

$$G\left(\left(\widehat{\widetilde{V}_\alpha \widetilde{V}_\beta} \widetilde{v}(t_0, x_0)\right)_{1 \leqslant \alpha, \beta \leqslant d}\right) = G\left(\left(\widehat{V_\alpha V_\beta} v(t_0, x_0)\right)_{1 \leqslant \alpha, \beta \leqslant d}\right).$$

It follows that

$$\frac{\partial v}{\partial t}(t_0, x_0) - G\left(\left(\widehat{V_\alpha V_\beta} v(t_0, x_0)\right)_{1 \leqslant \alpha, \beta \leqslant d}\right) \leqslant 0.$$

Therefore, $u(t,x)$ is a viscosity sub-solution to (5.7.7). Similarly we can show that it is a viscosity supersolution as well, and thus a viscosity solution.

The uniqueness of (5.7.7) follows from the same reason as in the proof of Theorem 5.7.1 once we notice that the second order differential operator $G\left(\left(\widehat{V_\alpha V_\beta} \cdot\right)_{1 \leqslant \alpha, \beta \leqslant d}\right)$ on M shares exactly the same properties as G (in particular, the same ellipticity), which can be seen either in an extrinsic way or via local computation. Another way to see the uniqueness is to use the results in [1] as long as we assign a complete

Riemannian metric on M, which is always possible according to K. Nomizu and H. Ozeki [50] even in the non-compact situation. In this case

$$G\left(\left(\widehat{V_\alpha V_\beta}u\right)_{1\leqslant\alpha,\beta\leqslant d}\right)=G\left(\left(\frac{1}{2}\left\langle\nabla u,\nabla_{V_\alpha}V_\beta+\nabla_{V_\beta}V_\alpha\right\rangle+\mathrm{Hess}u(V_\alpha,V_\beta)\right)_{1\leqslant\alpha,\beta\leqslant d}\right),$$

where ∇ is the Levi-Civita connection corresponding to the Riemannian metric. The uniqueness of (5.7.7) follows from [1], Theorem 5.1 directly, as the assumptions in the theorem are verified by the properties of G. Note that we do not need the Ricci curvature condition in [1] due to the compactness of M and uniform continuity of $G\left(\left(\widehat{V_\alpha V_\beta}\cdot\right)_{1\leqslant\alpha,\beta\leqslant d}\right)$. □

Remark 5.7.1. The study of the SDE (5.7.6) as a nonlinear diffusion process on M is independent of the geometry of M. The fundamental reason is that (5.7.6) is defined in the pathwise sense as an RDE generated by the vector fields V_α on M. Such an RDE depends only on the differential structure of M. The infinitesimal diffusive nature of (5.7.6) can be studied by local computation.

Now we turn to the study of G-Brownian motion on a compact Riemannian manifold. The Riemannian structure (the Levi-Civita connection) is used to "roll" the Euclidean G-Brownian motion up to the manifold "without slipping" by solving an SDE generated by the fundamental horizontal vector fields on a proper frame bundle (usually known as horizontal lifting). This is the fundamental idea of Eells, Elworthy and Malliavin on the construction of Brownian motion on a Riemannian manifold.

As is pointed out at the beginning of this section, the essential point of this development is the invariance of the generating PDE on the frame bundle under actions by the structure group along fibers. The key to capturing such invariance is Theorem 5.7.4 and Example 5.7.1, which leads to the following important concept.

Definition 5.7.1. The *invariant group* $I(G)$ of G is defined by

$$I(G)=\left\{Q\in GL(d,\mathbb{R}):\ \forall A\in S(d),\ G\left(Q^TAQ\right)=G(A)\right\}.$$

It is easy to check the $I(G)$ is a group, and hence a subgroup of $GL(d,\mathbb{R})$.

By using the representation (5.7.1) of G, we have the following equivalent characterization of the invariant group $I(G)$.

Proposition 5.7.2. *Let G be represented by*

$$G(A)=\frac{1}{2}\sup_{B\in\Sigma}\operatorname{tr}(AB),\ \forall A\in S(d),$$

where Σ is some bounded, closed and convex subset of $S_+(d)$. Then Σ is uniquely determined by G and the invariant group $I(G)$ of G is given by

$$I(G)=\left\{Q\in GL(d,\mathbb{R}):\ Q\Sigma Q^T=\Sigma\right\}. \tag{5.7.8}$$

Proof. It suffices to show the uniqueness of Σ, and (5.7.8) follows immediately from the commutativity of the trace operator and the uniqueness of Σ. Note that for any $Q\in GL(d,\mathbb{R})$, $Q\Sigma Q^T$ is also a bounded, closed and convex subset of $S_+(d)$.

Introduce a symmetric bilinear form $\langle\cdot,\cdot\rangle_{\text{tr}}$ on the finite dimensional vector space $S(d)$ by

$$\langle A_1,A_2\rangle_{\text{tr}}=\operatorname{tr}(A_1A_2),\ A_1,A_2\in S(d).$$

It is easy to check that $\langle\cdot,\cdot\rangle_{\text{tr}}$ is indeed an inner product, thus $(S(d),\langle\cdot,\cdot\rangle_{\text{tr}})$ is a finite dimensional Hilbert space. The norm $\|\cdot\|_{\text{tr}}$ induced by $\langle\cdot,\cdot\rangle_{\text{tr}}$ is equivalent to any other matrix norm on $S(d)$ since $S(d)$ is finite dimensional.

Let Σ_1,Σ_2 be two bounded, closed and convex subsets of $S_+(d)$, such that

$$\sup_{B\in\Sigma_1}\operatorname{tr}(AB)=\sup_{B\in\Sigma_2}\operatorname{tr}(AB),\ \forall A\in S(d).$$

If $\Sigma_1\neq\Sigma_2$, without loss of generality assume that $B_0\in\Sigma_2\backslash\Sigma_1$. According to the Mazur separation theorem in functional analysis (see the monograph by K. Yosida [68]), there exists a bounded linear functional $f\in S(d)^*$ and some $\alpha\in\mathbb{R}$, such that

$$f(B)<\alpha<f(B_0),\ \forall B\in\Sigma_1.$$

By the Riesz representation theorem, there exists a unique $A^*\in S(d)$, such that

$$f(B)=\langle A^*,B\rangle_{\text{tr}}=\operatorname{tr}(A^*B),\ \forall B\in S(d).$$

It follows that

$$\sup_{B\in\Sigma_1}\operatorname{tr}(A^*B)\leqslant\alpha<\operatorname{tr}(A^*B_0)\leqslant\sup_{B\in\Sigma_2}\operatorname{tr}(A^*B),$$

which is a contradiction. Therefore, $\Sigma_1=\Sigma_2$. □

We now give some examples for the invariant groups $I(G)$ of different G-functions.

Example 5.7.2. If $\Sigma = \{0\}$, then it is obvious that $I(G) = GL(d,\mathbb{R})$, which is a non-compact group.

Example 5.7.3. It is possible that $I(G)$ is a finite group.

Consider Σ to be the set of diagonal matrices

$$\Lambda = \mathrm{diag}(\lambda_1, \cdots, \lambda_d)$$

such that each $\lambda_\alpha \in [0,1]$, then Σ is a bounded, closed and convex subset of $S_+(d)$. We claim that

$$I(G) = \left\{(\pm e_{\sigma(1)}, \cdots, \pm e_{\sigma(d)}) : \ \sigma \text{ is a permutation of order } d\right\}, \tag{5.7.9}$$

where $\{e_1, \cdots, e_d\}$ is the standard orthonormal basis of $\mathbb{R}^d$, each e_i being regarded as a column vector.

In fact, if $Q \in GL(d,\mathbb{R})$ has the form (5.7.9), by direct computation one can show easily that

$$Q\Sigma Q^T = \Sigma. \tag{5.7.10}$$

Conversely, if Q satisfies (5.7.10), by choosing

$$\Lambda = \mathrm{diag}(1, 0, \cdots, 0),$$

we know that

$$\left(Q\Lambda Q^T\right)^\alpha_\beta = Q^\alpha_1 Q^\beta_1.$$

Therefore, if $Q\Lambda Q^T \in \Sigma$, the first column of Q must contain at most one nonzero element q_1 such that $q_1^2 \leqslant 1$. Similarly for other columns of Q. Moreover, the corresponding nonzero elements in any two different columns of Q must be in different rows, otherwise Q is degenerate. Consequently, Q has the form

$$Q = \left(q_1 e_{\sigma(1)}, \cdots, q_d e_{\sigma(d)}\right)$$

with $q_i^2 \leqslant 1$ $(i = 1, 2, \cdots, d)$. On the other hand, for the identity matrix I_d, there exists $\Lambda \in \Sigma$, such that

$$Q\Lambda Q^T = I_d.$$

By taking determinants on both sides, we have

$$q_1^2 \cdots q_d^2 \det(\Lambda) = 1,$$

which implies that $q_\alpha = \pm 1$ $(\alpha = 1, 2, \cdots, d)$. Therefore, Q has the form of (5.7.9).

Note that in this case $I(G)$ is a finite subgroup of the orthogonal group $O(d)$ with order $2^d d!$. Moreover, G is given by

$$G(A) = \frac{1}{2}\sum_{\alpha=1}^{d}(A_\alpha^\alpha)^+, \ \forall A \in S(d).$$

Example 5.7.4. Now we give some examples of G such that $I(G) = O(d)$. This case is our main interest in this section.

(1) $\Sigma = \{I_d\}$.

Obviously (5.7.10) is equivalent to $Q \in O(d)$.

This corresponds to the case of classical Brownian motion, in which

$$G(A) = \frac{1}{2}\mathrm{tr}(A)$$

and the generator is $\frac{1}{2}\Delta$.

(2) Σ is given by the segment joining λI_d and μI_d, where $0 \leqslant \lambda < \mu$.

If $Q \in GL(d, \mathbb{R})$ such that (5.7.10) holds, then

$$\mu QQ^T = tI_d,$$

for some $t \in [\lambda, \mu]$. On the other hand, there exists some $t' \in [\lambda, \mu]$ such that

$$t'QQ^T = \mu I_d.$$

The only possibility is that $QQ^T = I_d$, which means $Q \in O(d)$. The converse is trivial.

In this case, G is given by

$$G(A) = \frac{1}{2}\left(\mu(\mathrm{tr}A)^+ - \lambda(\mathrm{tr}A)^-\right).$$

The corresponding G-heat equation can be regarded as the generalization of the one-dimensional Barenblatt equation to higher dimensions.

(3) Σ is given by the subset of matrices $B \in S_+(d)$ such that the eigenvalues of B lie in the bounded interval $[\lambda, \mu]$, where $0 \leqslant \lambda < \mu$. Equivalently,

$$\Sigma = \left\{B \in S_+(d) : \ \lambda \leqslant x^T Bx \leqslant \mu, \ \forall x \in \mathbb{R}^d \text{ with } |x| = 1\right\}.$$

It follows that Σ is a bounded, closed and convex subset of $S_+(d)$.

Since Σ is characterized by eigenvalues, and the eigenvalues of a symmetric matrix are preserved under change of orthonormal basis, it follows that for any $Q \in O(d)$, (5.7.10) holds. Conversely, let $Q \in GL(d, \mathbb{R})$ with (5.7.10). Then there exists $B_1, B_2 \in \Sigma$, such that

$$\mu QQ^T = B_1,\ QB_2Q^T = \mu I_d.$$

It follows that all eigenvalues of QQ^T lie in $\left[\frac{\lambda}{\mu}, 1\right]$, and

$$\det\left(QQ^T\right)\det(B_2) = \mu^d.$$

Therefore, the only possibility is that all eigenvalues of QQ^T are equal to 1, which implies that Q is an orthogonal matrix.

In this case G can be expressed by

$$\begin{aligned} G(A) &= \frac{1}{2}\sup_{B\in\Sigma}\operatorname{tr}(AB) \\ &= \frac{1}{2}\sup_{P\in O(d)}\sup_{\lambda\leqslant c_1,\cdots,c_d\leqslant\mu}\operatorname{tr}\left(AP^T\operatorname{diag}(c_1,\cdots,c_d)P\right) \\ &= \frac{1}{2}\sup_{P\in O(d)}\sup_{\lambda\leqslant c_1,\cdots,c_d\leqslant\mu}\operatorname{tr}\left(PAP^T\operatorname{diag}(c_1,\cdots,c_d)\right) \\ &= \frac{1}{2}\sup_{P\in O(d)}\sup_{\lambda\leqslant c_1,\cdots,c_d\leqslant\mu}\sum_{\alpha=1}^{d} c_\alpha\left(PAP^T\right)_\alpha^\alpha \\ &= \frac{1}{2}\sup_{P\in O(d)}\sum_{\alpha=1}^{d}\left(\mu\left(\left(PAP^T\right)_\alpha^\alpha\right)^+ - \lambda\left(\left(PAP^T\right)_\alpha^\alpha\right)^-\right). \end{aligned}$$

Similar to Example 5.7.4, for those Σ's characterized by eigenvalues, we can construct a large class of G with $I(G) = O(d)$.

Remark 5.7.2. If Σ has at least one non-degenerate element, that is, if there exists some positive definite matrix $B_0 \in \Sigma$, then $I(G)$ is a compact group. In fact, if we introduce a matrix norm $\|\cdot\|_{B_0}$ on the space $\operatorname{Mat}(d, \mathbb{R})$ of real $d\times d$ matrices by

$$\|A\|_{B_0} = \sqrt{\operatorname{tr}\left(AB_0A^T\right)},\ A \in \operatorname{Mat}(d, \mathbb{R}),$$

it follows that

$$\sup_{Q\in I(G)}\|Q\|_{B_0} = \sup_{Q\in I(G)}\sqrt{\operatorname{tr}\left(QB_0Q^T\right)} \leqslant \sup_{B\in\Sigma}\sqrt{\operatorname{tr}(B)} < \infty,$$

since Σ is bounded. It is obvious that $I(G)$ is closed. Therefore, it is compact.

Now suppose that (M, g) is a d-dimensional compact Riemannian manifold.

We first recall some basic notions on frame bundles, which is the central concept in the horizontal lifting construction. We refer the reader to the monograph [10], and also the one by S. Kobayashi and K. Nomizu [41] for a systematic introduction.

Let $\mathcal{F}(M)$ be the total frame bundle over M defined by

$$\mathcal{F}(M) = \bigcup_{x \in M} \mathcal{F}_x(M),$$

where the fibre $\mathcal{F}_x(M)$ is the set of all frames (bases of the tangent space $T_x(M)$) at x. A frame $\xi = (\xi_1, \cdots, \xi_d) \in \mathcal{F}_x(M)$ can be equivalently regarded as a linear isomorphism from $\mathbb{R}^d$ to T_xM (also denoted by ξ) if we let

$$\xi(e_\alpha) = \xi_\alpha, \ \alpha = 1, 2, \cdots, d,$$

and extend linearly to $\mathbb{R}^d$, where we always fix $\{e_1, \cdots, e_d\}$ to be the standard orthonormal basis of $\mathbb{R}^d$. $\mathcal{F}(M)$ is a principal bundle with structure group $GL(d, \mathbb{R})$ acting along fibers from the right.

Fix a frame $\xi \in \mathcal{F}_x(M)$. A vector $X \in T_\xi\mathcal{F}(M)$ is called *vertical* if it is tangent to the fibre $\mathcal{F}_x(M)$. The space of vertical vectors at ξ is called the *vertical subspace*, and it is denoted by $V_\xi\mathcal{F}(M)$. $V_\xi\mathcal{F}(M)$ is a d^2-dimensional vector space, which is independent of the Riemannian structure.

A smooth curve $\xi_t = (\xi_{1,t} \cdots, \xi_{d,t}) \in \mathcal{F}(M)$ is called *horizontal* if $\xi_{\alpha,t}$ is a parallel vector field along the projection curve $x_t = \pi(\xi_t)$ for each $\alpha = 1, 2, \cdots, d$. Given a smooth curve $x_t \in M$ and a frame $\xi_0 = (\xi_1, \cdots, \xi_d) \in \mathcal{F}_{x_0}(M)$, by solving a first order linear ODE, we can determine a unique parallel vector field $\xi_{\alpha,t}$ along x_t with $\xi_{\alpha,0} = \xi_\alpha$ for each $\alpha = 1, 2, \cdots, d$. The smooth curve

$$\xi_t = (\xi_{1,t}, \cdots, \xi_{d,t}) \in \mathcal{F}(M)$$

is then the unique horizontal curve with $x_t = \pi(\xi_t)$ and initial position ξ_0. ξ_t is called the *horizontal lifting* of x_t from ξ_0. A vector $X \in T_\xi\mathcal{F}(M)$ is called *horizontal* if it is tangent to a horizontal curve through ξ. The space of horizontal vectors at ξ is called the *horizontal subspace*, and it is denoted by $H_\xi\mathcal{F}(M)$. It is a d-dimensional vector space characterized by the Levi-Civita connection ∇.

As ξ varies, $V_\xi\mathcal{F}(M)$ (respectively, $H_\xi\mathcal{F}(M)$) determines a vertical (respectively, horizontal) subspace field on M. The following result reveals the fundamental structure of $\mathcal{F}(M)$. We refer the reader to [41] for the proof.

Theorem 5.7.3. *The horizontal subspace field $H\mathcal{F}(M)$, which is determined by ∇, has the following properties.*

(1) For each $\xi \in \mathcal{F}_x(M)$, the tangent space $T_\xi \mathcal{F}(M)$ has the decomposition

$$T_\xi \mathcal{F}(M) = H_\xi \mathcal{F}(M) \oplus V_\xi \mathcal{F}(M).$$

Moreover, $H_\xi \mathcal{F}(M)$ is isomorphic to $T_x M$ under the canonical projection $\pi : \mathcal{F}(M) \to M$.

(2) $H\mathcal{F}(M)$ is invariant under actions by the structure group $GL(d, \mathbb{R})$. More precisely, for any $\xi \in \mathcal{F}(M)$, $Q \in GL(d, \mathbb{R})$,

$$Q_*(H_\xi \mathcal{F}(M)) = H_{\xi Q} \mathcal{F}(M).$$

It should be pointed out the Riemannian structure is not essential for the existence of the above horizontal-vertical decomposition; it is the affine connection (the Levi-Civita connection) ∇ that plays the key role. Moreover, on the contrary it can be proved (see[41]) that given any horizontal subspace field $H\mathcal{F}(M)$ satisfying the two properties in Theorem 5.7.3, there exists an affine connection ∇^H such that $H\mathcal{F}(M)$ is the horizontal subspace field determined by ∇^H.

On $\mathcal{F}(M)$ there is a canonical way to define a frame field globally, which is not always possible on a general Riemannian manifold. This makes $\mathcal{F}(M)$ simpler than the base space M to some extent. Fix $w \in \mathbb{R}^d$. For any $\xi \in \mathcal{F}_x(M)$ regarded as a linear isomorphism $\xi : \mathbb{R}^d \to T_x M$, we know that $\xi(w)$ is a tangent vector in $T_x M$. By Theorem 5.7.3 (1), $\xi(w)$ corresponds to a unique vector $H_w(\xi) \in H_\xi \mathcal{F}(M)$. It follows that H_w is a globally defined horizontal vector field on $\mathcal{F}(M)$. If we take $w = e_\alpha$ ($\alpha = 1, 2, \cdots, d$), then we obtain a family of horizontal vector fields $\{H_{e_1}, \cdots, H_{e_d}\}$ as a basis of the horizontal subspace $H_\xi \mathcal{F}(M)$ at each frame $\xi \in \mathcal{F}(M)$. $\{H_{e_1}, \cdots, H_{e_d}\}$ are called the *fundamental horizontal fields* of $\mathcal{F}(M)$, simply denoted by $\{H_1, \cdots, H_d\}$.

Now we introduce the concept of development and anti-development based on [36], which is crucial in the construction of G-Brownian motion on M. Assume that $x_t \in M$ is a smooth curve and ξ_t is the horizontal lifting of x_t from ξ_0. Then we can determine a smooth curve

$$w_t = \int_0^t \xi_s^{-1} \dot{x}_s ds \in \mathbb{R}^d$$

starting at 0 (w_t is regarded as a column vector in $\mathbb{R}^d$). w_t is called the *anti-development* of x_t in $\mathbb{R}^d$ with respect to ξ_0. If ξ_t and η_t are two horizontal liftings of x_t with $\xi_0 = \eta_0 Q$ for some $Q \in GL(d, \mathbb{R})$, then the two corresponding anti-developments are related by

$$w_t^\eta = Q w_t^\xi.$$

The key relation between the anti-development w_t of x_t and the horizontal lifting ξ_t is the following ODE on $\mathcal{F}(M)$:

$$d\xi_t = H_\alpha(\xi_t) dw_t^\alpha. \tag{5.7.11}$$

Conversely, given a smooth curve $w_t \in \mathbb{R}^d$ starting at 0, by solving the ODE (5.7.11) on $\mathcal{F}(M)$ with initial frame ξ_0, we obtain a horizontal curve $\xi_t \in \mathcal{F}(M)$. The projection $x_t = \pi(\xi_t)$ is called the *development* of w_t in M with respect to ξ_0. If we use another initial frame $\eta_0 = \xi_0 Q^{-1}$ and the driven process $v_t = Q w_t \in \mathbb{R}^d$, by solving (5.7.11) from η_0 and projection onto M we obtain the same curve x_t. In this way, we obtain a one-to-one correspondence between the Euclidean curve w_t and the manifold curve x_t via the horizontal curve ξ_t in $\mathcal{F}(M)$, which depends on the initial frame ξ_0. The procedure of getting x_t from w_t is usually known as "rolling without slipping".

A crucial point here is that such a procedure is carried out by solving the ODE (5.7.11) in the pathwise sense, which fits well in the context of rough paths if the Euclidean curve w_t is interpreted as a rough path. In this case, (5.7.11) should be interpreted as an RDE. This is an important reason why we need to develop the notion of Stratonovich type SDEs on a differentiable manifold.

For a general Euclidean G-Brownian motion B_t, from Section 5.6 we are able to solve (5.7.11) pathwisely if the driving curve dw_t is replaced by dB_t in the Stratonovich sense (or in the RDE sense). By projecting the solution $\xi_t \in \mathcal{F}(M)$ to the manifold M, we obtain a process $X_t \in M$ pathwisely which depends on the initial position x_0 and the initial frame $\xi_0 \in \mathcal{F}_{x_0}(M)$. A disadvantage of using the total frame bundle $\mathcal{F}(M)$ is that in this way it is not possible to write down the generating PDE governing the law of X_t intrinsically on M, which does not depend on the initial frame ξ_0. Note that the generating PDE of ξ_t is well-defined on $\mathcal{F}(M)$ according to Theorem 5.7.2, which takes the form

$$\frac{\partial u}{\partial t} - G\left(\left(\widehat{H_\alpha H_\beta} u\right)_{1\leqslant \alpha,\beta\leqslant d}\right) = 0. \tag{5.7.12}$$

The main reason for this disadvantage is that the PDE (5.7.12) is not invariant under actions by $GL(d,\mathbb{R})$ along fibers, since the G-function does not have this kind of invariance.

To fix this issue, a possible way is to use the invariant group $I(G)$ of G as the structure group, so that the generating PDE is invariant under actions by $I(G)$ along fibers due to the form (5.7.12) takes. Therefore, we need to use a proper frame bundle (a submanifold of $\mathcal{F}(M)$ which is a principal bundle over M with structure group $I(G)$ and fibers being a suitable class of frames) instead of $\mathcal{F}(M)$. The fibers of such frame bundle should be preserved by parallel transport so the fundamental horizontal fields can be restricted to it and we are able to solve the RDE

$$d\xi_t = H_\alpha(\xi_t) \circ dB_t^\alpha$$

on the frame bundle. It then turns out that we are able to establish the generating PDE of the projection process $X_t = \pi(\xi_t)$ intrinsically on M, which does not depend on the initial frame. Therefore, although as a process the sample paths of X_t depend on the initial frame (this is not surprising since in the Euclidean case we also don't have a canonical Brownian motion if we do not fix the frame $\{e_1, \cdots, e_d\}$ in advance), the law of X_t does not. In this way we obtain a canonical PDE on M associated with the original G-function, which can be regarded as the generating PDE governing the law of X_t. The process X_t can be defined as a G-Brownian motion on M and the solution to the generating PDE plays the role of the canonical Wiener measure (the solution of the martingale problem for the operator $\frac{1}{2}\Delta_M$) on M in the nonlinear setting.

The construction of such a frame bundle for a G-function with an arbitrary invariant group $I(G)$ is not clear to us at the moment. However, in the case when $I(G)$ is the orthogonal group $O(d)$, which contains a wide and interesting class of G-functions, there is a very natural frame bundle serving us well for this purpose: the orthonormal frame bundle $\mathcal{O}(M)$.

From now on, let G be given by (5.7.1) with $I(G) = O(d)$.

The orthonormal frame bundle $\mathcal{O}(M)$ over M is defined by

$$\mathcal{O}(M) = \bigcup_{x \in M} \mathcal{O}_x(M),$$

where the fibre $\mathcal{O}_x(M)$ is the set of orthonormal bases of T_xM. Since M is compact, $\mathcal{O}(M)$ is a compact submanifold of $\mathcal{F}(M)$. Moreover, since the Levi-Civita connection is compatible with the Riemannian metric g, parallel transport preserves the fibers of $\mathcal{O}(M)$. Therefore, statements about $\mathcal{F}(M)$ before on the horizontal aspect can be carried through in the case of $\mathcal{O}(M)$ directly. In particular, the fundamental horizontal fields H_α can be restricted to $\mathcal{O}(M)$. The only difference is in the vertical direction: the fibre becomes orthonormal frames, and the structure group which acts on fibers becomes the orthogonal group; the dimension in the vertical direction is reduced to $d(d-1)/2$.

For $\xi \in \mathcal{O}_x(M)$, according to Section 5.6, let $U_{t,\xi} \in \mathcal{O}(M)$ be the unique solution to the following RDE over $[0,1]$:

$$\begin{cases} dU_{t,\xi} = H_\alpha(U_{t,\xi}) \circ dB_t^\alpha, \\ U_{0,\xi} = \xi. \end{cases} \tag{5.7.13}$$

Let $X_{t,\xi} = \pi(U_{t,\xi})$ be the projection of $U_{t,\xi}$ onto M.

Definition 5.7.2. $X_{t,\xi}$ is called a *G-Brownian motion on M* with respect to the initial orthonormal frame $\xi \in \mathcal{O}_x(M)$, and $U_{t,\xi}$ is called a horizontal G-Brownian motion in $\mathcal{O}(M)$ starting at ξ.

For any $\varphi \in C_{Lip}(M)$ (under the Riemannian distance), define

$$u(t,\xi) = \mathbb{E}^G\left[\varphi(X_{t,\xi})\right], \ (t,\xi) \in [0,1] \times \mathcal{O}(M).$$

Let $\hat{\varphi} = \varphi \circ \pi$ be the lifting of φ to $\mathcal{O}(M)$. It is obvious that

$$u(t,\xi) = \mathbb{E}^G\left[\hat{\varphi}(U_{t,\xi})\right].$$

By Theorem 5.7.2, we know that $u(t,\xi)$ is the unique viscosity solution to the following nonlinear parabolic PDE:

$$\begin{cases} \frac{\partial u}{\partial t} - G\left(\left(\widehat{H_\alpha H_\beta} u\right)_{1 \leqslant \alpha,\beta \leqslant d}\right) = 0, \\ u(0,\xi) = \hat{\varphi}(\xi), \end{cases} \tag{5.7.14}$$

on $\mathcal{O}(M)$.

The following result tells us that the law of $X_{t,\xi}$ depends only on the initial position x.

Proposition 5.7.3. *If $\xi, \eta \in \mathcal{O}_x(M)$, then*

$$u(t,\xi) = u(t,\eta).$$

Proof. For any fixed orthogonal matrix $Q \in O(d)$, let $\widetilde{B}_t = QB_t$, which is an orthogonal transformation of the original G-Brownian motion B_t, and let $W_{t,\zeta}$ be the pathwise solution to the following RDE over $[0,1]$:

$$\begin{cases} dW_{t,\zeta} = H_\alpha(W_{t,\zeta}) \circ d\widetilde{B}_t^\alpha, \\ W_{0,\zeta} = \zeta \in \mathcal{O}(M), \end{cases} \tag{5.7.15}$$

on $\mathcal{O}(M)$. If we regard $\widetilde{B}_t$ as the solution to the SDE

$$d\widetilde{B}_t = Q_\alpha dB_t^\alpha$$

starting at 0 with constant coefficients, then the RDE (5.7.15) is equivalent to

$$\begin{cases} dW_{t,\zeta} = H_\beta(W_{t,\zeta})Q_\alpha^\beta \circ dB_t^\alpha, \\ W_{0,\zeta} = \zeta, \end{cases}$$

in which the generating vector fields are $H_\beta Q_\alpha^\beta$. Since the invariant group $I(G)$ of G is the orthogonal group, by Theorem 5.7.2 we know that the function

$$v(t,\zeta) = \mathbb{E}^G\left[\hat{\varphi}(W_{t,\zeta})\right], \ (t,\zeta) \in [0,1] \times \mathcal{O}(M)$$

is the unique viscosity solution to the same PDE (5.7.14) on $\mathcal{O}(M)$. Therefore,

$$u(t,\zeta) = v(t,\zeta), \ \forall (t,\zeta) \in [0,1] \times \mathcal{O}(M).$$

Now since $\xi, \eta \in \mathcal{O}_x(M)$, there exists some $Q \in O(d)$ such that $\xi = \eta Q$. Define $W_{t,\zeta}$ as before. By the previous discussion on the relation between different anti-developments, we know that

$$X_{t,\xi} = \pi(U_{t,\xi}) = \pi(W_{t,\eta}), \ \forall t \in [0,1].$$

Therefore,

$$\begin{aligned} u(t,\xi) &= \mathbb{E}^G[\varphi \circ \pi(U_{t,\xi})] \\ &= \mathbb{E}^G[\varphi \circ \pi(W_{t,\eta})] \\ &= v(t,\eta) \\ &= u(t,\eta). \end{aligned}$$

□

From Proposition 5.7.3, we know that $u(t,\xi)$ is invariant along each fibre. Therefore, the law of $X_{t,\xi}$ depends only on the initial position $x \in M$ and not on the initial frame ξ. We use $u(t,x)$ to denote $u(t,\xi)$, where x is the base point of ξ. In this situation it is possible to establish the PDE for $u(t,x)$ intrinsically on M by "projecting down" (5.7.14), which should become the generating PDE governing the law of $X_{t,\xi}$.

For any $u \in C^\infty(M)$, take an orthonormal frame $\xi = (\xi_1, \cdots, \xi_d) \in \mathcal{O}_x(M)$, and consider the quantity

$$G((\mathrm{Hess}u(\xi_\alpha, \xi_\beta))_{1\leqslant\alpha,\beta\leqslant d}).$$

Since $I(G) = O(d)$, it is easy to see that the above quantity is independent of the orthonormal frame $\xi \in \mathcal{O}_x(M)$. In other words, G can be regarded as a functional of the Hessian, and the nonlinear second order differential operator $G(\mathrm{Hess}(\cdot))$ is globally well-defined on M.

Now we have the following result.

Theorem 5.7.4. *$u(t,x)$ is the unique viscosity solution to the following nonlinear heat equation on M :*

$$\begin{cases} \frac{\partial u}{\partial t} - G(\mathrm{Hess}u) = 0, \\ u(0,x) = \varphi(x). \end{cases} \tag{5.7.16}$$

Proof. It suffices to show that: if $f \in C^\infty(M)$, and $\hat{f} = f \circ \pi$ is the lifting of f to $\mathcal{O}(M)$, then for any $\xi = (\xi_1, \cdots, \xi_d) \in \mathcal{O}_x(M)$,

$$\mathrm{Hess}f(\xi_\alpha, \xi_\beta)(x) = H_\alpha H_\beta \hat{f}(\xi).$$

Note that uniqueness follows for the same reason as pointed out in the proof of Theorem 5.7.2 by using results in [1].

In fact, for any $\xi = (\xi_1, \cdots, \xi_d) \in \mathcal{O}_x(M)$, let ξ_t be a horizontal curve through ξ such that $H_\beta(\xi)$ is tangent to ξ_t at $t = 0$, and let x_t be its projection onto M. It follows that the tangent vector of x_t at $t = 0$ is ξ_β, and

$$\begin{aligned} H_\beta \hat{f}(\xi) &= \frac{d\hat{f}(\xi_t)}{dt}\Big|_{t=0} \\ &= \frac{df(x_t)}{dt}\Big|_{t=0} \\ &= \langle \xi_\beta, \nabla f(x) \rangle_g. \end{aligned}$$

Therefore, if we now assume that ξ_t is a horizontal curve through ξ with tangent vector $H_\alpha(\xi)$ at ξ and still $x_t = \pi(\xi_t)$, then

$$\begin{aligned} H_\alpha H_\beta \hat{f}(\xi) &= H_\alpha \langle \xi_\beta, \nabla f(\pi(\xi))\rangle_g \\ &= \frac{d}{dt}|_{t=0} \langle \xi_{\beta,t}, \nabla f(x_t)\rangle_g \\ &= \left\langle \frac{D\xi_{\beta,t}}{dt}|_{t=0}, \nabla f(x) \right\rangle_g + \langle \xi_\beta, \nabla_{\xi_\alpha} \nabla f(x)\rangle_g \\ &= \mathrm{Hess} f(\xi_\alpha, \xi_\beta)(x), \end{aligned}$$

where we have used the fact that $\xi_{\beta,t}$ is parallel along x_t. □

Since $X_{t,\xi}$ is the projection of $U_{t,\xi}$ and $U_{t,\xi}$ is the solution to the RDE (5.7.13) which is equivalent to an Itô type SDE from an extrinsic point of view, by Theorem 5.7.4 we can see that as a process on M the law of the G-Brownian motion $X_{t,\xi}$ is characterized by the nonlinear parabolic PDE (5.7.16).

Example 5.7.5. When G is given by a functional of the trace, as in Example 5.7.4 (1), (2), the generating PDE (5.7.16) takes a more explicit form in terms of the Laplace-Beltrami operator Δ_M on M. This is due to the fact that

$$\Delta_M = \mathrm{tr}(\mathrm{Hess}).$$

For instance, if $G(A) = \frac{1}{2}\mathrm{tr}(A)$, then (5.7.16) becomes the classical heat equation on M:

$$\frac{\partial u}{\partial t} - \frac{1}{2}\Delta_M u = 0,$$

which governs the law of classical Brownian motion on M (see [36], [37]). If G is given by

$$G(A) = \frac{1}{2}\left(\mu(\mathrm{tr}A)^+ - \lambda(\mathrm{tr}A)^-\right),$$

where $0 \leqslant \lambda < \mu$, then (5.7.16) becomes

$$\frac{\partial u}{\partial t} - \frac{1}{2}\left(\mu(\Delta_M u)^+ - \lambda(\Delta_M u)^-\right) = 0.$$

It is a generalization of the one-dimensional Barenblatt equation to higher dimensions in a Riemannian geometric setting.

As pointed out before, as a process the G-Brownian motion $X_{t,\xi}$ on M depends on the initial orthonormal frame ξ and hence there is not a canonical choice of a particular one. However, if we consider the path space $W(M) = C([0,1]; M)$, then for each $x \in M$, it is possible to define a canonical sublinear expectation $\mathbb{E}_x$ on the space $\mathcal{H}(M)$ of functionals on $W(M)$ of the form

$$f(x_{t_1}, \cdots, x_{t_n}),$$

where $0 \leqslant t_1 < \cdots < t_n \leqslant 1$ and $f \in C_{Lip}(M)$, such that under $\mathbb{E}_x$ the law of the coordinate process is characterized by the PDE (5.7.16) with $\mathbb{E}_x[\varphi(x_0)] = \varphi(x)$ for any $\varphi \in C_{Lip}(M)$.

To see this, we define $\mathbb{E}_x$ explicitly. We use $u_\varphi(t,x)$ to denote the solution to (5.7.16), emphasizing the dependence on φ. For a functional of the form $f(x_t)$, we simply define

$$\mathbb{E}_x[f(x_t)] = u_f(t,x).$$

For a functional of the form $f(x_s, x_t)$, $\mathbb{E}_x f(x_s, x_t)$ should be defined by $\mathbb{E}^G[f(X_{s,\xi}, X_{t,\xi})]$, where $X_{t,\xi}$ is a G-Brownian motion on M with respect to an initial orthonormal frame $\xi \in \mathcal{O}_x(M)$. Similar to the proof of Theorem 5.7.1 we know that

$$\begin{aligned} \mathbb{E}^G[f(X_{s,\xi}, X_{t,\xi})] &= \mathbb{E}^G\left[\mathbb{E}^G\left[f(X_{s,\xi}, X_{t,\xi}) | \, \Omega_s\right]\right] \\ &= \mathbb{E}^G\left[\mathbb{E}^G\left[f(\pi(U_{s,\xi}), \pi(U_{t,\xi})) | \, \Omega_s\right]\right] \\ &= \mathbb{E}^G\left[\mathbb{E}^G\left[f(\pi(\eta), X_{t-s,\eta})\right] \big|_{\eta = U_{s,\xi}}\right]. \end{aligned}$$

But since the law of $X_{t-s,\eta}$ does not depend on the initial orthonormal frame η, we obtain that

$$\mathbb{E}^G\left[f(\pi(\eta), X_{t-s,\eta})\right] \big|_{\eta = U_{s,\xi}} = u_{f(X_{s,\xi},\cdot)}(t-s, X_{s,\xi}).$$

Therefore, we define

$$\mathbb{E}_x\left[f(x_s, x_t)\right] = \mathbb{E}^G\left[f(X_{s,\xi}, X_{t,\xi})\right] = u_g(s,x),$$

where

$$g(y) := u_{f(y,\cdot)}(t-s, y), \; y \in M.$$

Inductively, assume that

$$u_f^{(n)}(t_1, \cdots, t_n, x) = \mathbb{E}_x[f(x_{t_1}, \cdots, x_{t_n})]$$

is already defined. For a functional of the form $f(x_{t_1}, \cdots, x_{t_{n+1}})$, define

$$\mathbb{E}_x\left[f(x_{t_1}, \cdots, x_{t_{n+1}})\right] = u_g(t_1, x),$$

where

$$g(y) := u^{(n)}_{f(y,\cdot,\cdots,\cdot)}(t_2 - t_1, \cdots, t_{n+1} - t_1, y), \ y \in M.$$

Then $\mathbb{E}_x$ is the desired sublinear expectation on $\mathcal{H}(M)$.

Remark 5.7.3. As we have pointed out before, for non-compact Riemannian manifolds, the RDE (5.7.13) may possibly explode at some finite time and so may the corresponding G-Brownian motion as well. An interesting question is the study of geometric conditions for explosion from both the deterministic and PDE point of view. It possibly relates to the curvature and topology of the Riemannian manifold.

On the other hand, for those G-functions with the same invariant group, they may have some important features in common; while for those with different invariant groups, their structures should be very different. The study of classification of G-functions in terms of the invariant group is interesting, and it might give us some hints on generalizing our results to the case when $I(G) \neq O(d)$. We believe that in some cases it is still possible to construct a proper frame bundle (or more generally a principal bundle) with structure group $I(G)$ on which we can apply similar technique as before. But in some extreme cases, for instance when $I(G)$ is a finite group as in Example 5.7.3, it seems difficult to proceed along this approach unless we have a globally defined frame field over the Riemannian manifold M, which is usually not true. We should explore different ideas for those extreme cases.

Bibliography

[1] D. Azagra, J. Ferrera and B. Sanz, Viscosity solutions to second order partial differential equations on Riemannian manifolds, *J. Differential Equations* 245 (2): 307–336, 2009.

[2] T. Bagby, L. Bos and N. Levenberg, Multivariate simultaneous approximation, *Constr. Approx.* 18 (3): 569–577, 2002.

[3] H. Boedihardjo and X. Geng, Simple piecewise geodesic interpolation of simple and Jordan curves with applications, to appear in *Constr. Approx.*., 2014.

[4] H. Boedihardjo and X. Geng, The uniqueness of signature problem in the non-Markov setting, *arXiv*: 1406.7871, 2014.

[5] H. Boedihardjo, X. Geng, T. Lyons and D. Yang, The signature of a rough path: uniqueness, *arXiv*: 1406.7871, 2014.

[6] H. Boedihardjo, H. Ni and Z. Qian, Uniqueness of signature for simple curves, *J. Funct. Anal.* 267 (6): 1778–1806, 2014.

[7] T. Cass and P. Friz, Densities for rough differential equations under Hörmander's condition, *Ann. of Math.* 171 (3): 2115–2141, 2010.

[8] K. Chen, Iterated integrals and exponential homomorphisms, *Proc. London Math. Soc.* 4 (3): 502–512, 1954.

[9] K. Chen, Integration of paths-a faithful representation of paths by non-commutative formal power series, *Trans. Amer. Math. Soc.* 89: 395–407, 1958.

[10] S.S. Chern, W. Chen and K.S. Lam, *Lectures on differential geometry*, World Scientific, 1999.

[11] I. Chiswell, *Introduction to trees*, World Scientific, 2001.

[12] G. Choquet, Theory of capacities, *Ann. Inst. Fourier* 87 (5): 131–295, 1953.

[13] L. Coutin and Z. Qian, Stochastic differential equations for fractional Brownian motions, *C. R.Acad. Sci. Paris Ser. I Math.* 331: 75–80, 2000.

[14] L. Coutin and Z. Qian, Stochastic analysis, rough path analysis and fractional Brownian motions, *Probab. Theory Relat. Fields* 122: 108–140, 2002.

[15] M.G. Crandall, H. Ishii and P.L. Lions, User's guide to viscosity solutions of second order partial differential equations, *Bull. Amer. Math. Soc.* 27 (1): 1–67, 1992.

[16] A.M. Davie, Differential equations driven by rough paths: an approach via discrete approximation, *Appl. Math. Res. Express.* AMRX (2): Art. ID abm009, 40, 2007.

[17] G. de Rham, *Riemannian Manifolds,* Springer-Verlag, 1984.

[18] L. Decreusefond and A. Ustunel, Stochastic Analysis of the Fractional Brownian Motion, *Potential Anal.* 10: 177–214, 1997.

[19] C. Dellacherie, *Capacités et Processus Stochastiques,* Springer-Verlag, 1972.

[20] L. Denis, M. Hu, and S. Peng, Function spaces and capacity related to a sublinear expectation: application to G-Brownian motion paths, *Potential Anal.* 34 (2): 139–161, 2011.

[21] M.P. do Carmo, *Riemannian geometry*, Springer, 1992.

[22] K.D. Elworthy, *Stochastic Differential Equations on Manifolds,* Springer, 1998.

[23] C. Favre and M. Jonsson, *The valuative tree*, Springer, 2004.

[24] P. Friz and N. Victoir, A variation embedding theorem and applications, *J. Funct. Anal.* 239: 631–637, 2006.

[25] P. Friz and N. Victoir, Euler estimates for rough differential equations. *J. Differential Equations* 244 (2): 388–412, 2008.

[26] P. Friz and N. Victoir, *Multidimensional stochastic processes as rough paths*, Cambridge University Press, 2010.

[27] F. Gao, Pathwise properties and homeomorphic flows for stochastic differential equations driven by G-Brownian motion, *Stochastic Process. Appl.* 119 (10): 3356–3382, 2009.

[28] X. Geng and Z. Qian, On an inversion theorem for Stratonovich's signatures of multidimensional diffusion paths, to appear in *Ann. Inst. Poincaré (B).*

[29] X. Geng, Z. Qian and D. Yang, G-Brownian motion as rough paths and differential equations driven by G-Brownian motion, to appear in *Séminaire de Probabilités.*

[30] M. Gubinelli, Controlling rough paths, *J. Funct. Anal.* 216 (1): 86–140, 2004.

[31] M Hairer, Solving the KPZ equation, *Ann. of Math.* 178 (2): 559–664, 2013.

[32] M. Hairer, A theory of regularity structures, *arXiv*: 1303.5113, 2013.

[33] B.M. Hambly and T. Lyons, Stochastic area for Brownian motion on the sierpinski gasket, *Ann. Probab.* 26 (1): 132–148, 1998.

[34] B.M. Hambly and T. Lyons, Some notes on trees and paths, *arXiv:* 0809.1365, 2008.

[35] B.M. Hambly and T. Lyons, Uniqueness for the signature of a path of bounded variation and the reduced path group, *Ann. of Math.* 171 (1), 109–167, 2010.

[36] E.P. Hsu, *Stochastic Analysis on Manifolds*, American Mathematical Society, 2002.

[37] N. Ikeda and S.Watanabe, *Stochastic Differential Equations and Diffusion Processes,* North-Holland, 1989.

[38] Y. Inahama, Malliavin differentiability of solutions of rough differential equations, *arXiv:* 1312.7621, 2014.

[39] M. Kac, On distributions of certain wiener functionals, *Trans. Amer. Math. Soc.* 65 (1): 1–13, 1949.

[40] I.A. Karatzas and S.E. Shreve, *Brownian Motion and Stochastic Calculus,* Springer-Verlag, 1991.

[41] S. Kobayashi and K. Nomizu, *Foundations of Differential Geometry I, II,* Interscience, 1963 and 1969.

[42] M. Ledoux, Z. Qian and T. Zhang, Large deviations and support theorem for diffusion processes via rough paths, *Stochastic Process. Appl.* 102 (2): 265–283, 2002.

[43] Y. Le Jan and Z. Qian, Stratonovich's signatures of Brownian motion determine Brownian sample paths, *Probab. Theory Relat. Fields* 157, 440–454, 2012.

[44] D. Levin and T. Lyons, H. Ni, Learning from the past, predicting the statistics for the future, learning an evolving system, *arXiv:* 1309.0260, 2013.

[45] T. Lyons, Differential equations driven by rough signals, *Rev. Mat. Iberoamericana* 14 (2), 215–310, 1998.

[46] T. Lyons, M. Caruana and T. Lévy, *Differential equations driven by rough paths*, Springer, 2007.

[47] T. Lyons and Z. Qian, *System control and rough path*s, Oxford University Press, 2002.

[48] T. Lyons and W. Xu, Inverting the signature of a path, *arXiv*:1406.7833, 2014.

[49] T. Lyons and D. Yang, Integration of time-varying cocyclic one-forms against rough paths, *arXiv*: 1408.2785, 2014.

[50] K. Nomizu and H. Ozeki, The existence of complete riemannian metrics, *Proc. Amer. Math. Soc.* 12 (6): 889–891, 1961.

[51] D. Nualart, *The Malliavin calculus and related topics,* Springer-Verlag, 2006.

[52] E. Pardoux and S. Peng, Adapted solution of a backward stochastic differential equation, *Syst. Control Lett.* 14 (1): 55–61, 1990.

[53] E. Pardoux and S. Peng, Backward stochastic differential equations and quasilinear parabolic partial differential equations, *Stochastic Partial Differential Equations and Their Applications*, 200–217, 1992.

[54] E. Pardoux and S. Peng, Backward doubly stochastic differential equations and systems of quasilinear SPDEs, *Probab. Theory Relat. Fields* 98 (2): 209–227, 1994.

[55] S. Peng, Backward SDE and related g-expectation, *Pitman Research Notes in Mathematics series*, 141–160, 1997.

[56] S. Peng, G-expectation, G-Brownian motion and related stochastic calculus of Itô type, *Stoch. Anal. Appl.* Abel Symposia (2): 541–567, 2007.

[57] S. Peng, Multi-dimensional G-Brownian motion and related stochastic calculus under G-expectation, *Stochastic Process. Appl.* 118 (12): 2223–2253, 2008.

[58] S. Peng, Nonlinear expectations and stochastic calculus under uncertainty, *arXiv*: 1002.4546, 2010.

[59] S. Samko and A. Kilbas, O Marichev, *Fractional integrals and derivatives: theory and applications*, Gordon and Breach, 1993.

[60] E.M. Sipiläinen, A pathwise view of solutions of stochastic differential equations, PhD thesis, University of Edinburgh, 1993.

[61] D. Stroock, *Probability theory, an analytic view*, Cambridge University Press, 2nd Edition, 1993.

[62] H. Tverberg. A proof of the Jordan curve theorem, *Bull. Lond. Math. Soc.* 12: 34–38, 1980.

[63] F.W. Warner, *Foundations of differentiable manifolds and Lie groups*, Springer, 1983.

[64] B. Werness, Regularity of Schramm-Loewner evolutions, annular crossings, and rough path theory. *Electron. J. Probab.* 17: 1–21, 2012.

[65] S. Willard, *General topology*, Addison-Wesley, 1970.

[66] E. Wong and M. Zakai, On the relation between ordinary and stochastic differential equations, *Internat. J. Engrg. Sci.* 3 (2): 213–229, 1965.

[67] P. Yam, Analytical and topological aspects of signatures, PhD thesis, University of Oxford, 2008.

[68] K. Yosida, *Functional Analysis*, Springer-Verlag, 1980.

[69] L. C. Young, An inequality of Hölder type connected with Stieltjes integration. *Acta Math.* 67: 251–282, 1936.