A Quasi-sure Non-degeneracy Property for the Brownian Rough Path

H. Boedihardjo[*], X. Geng[†], X. Liu[‡] and Z. Qian[§]

March 2, 2016

Abstract

In the present paper, we are going to show that outside a slim set in the sense of Malliavin (or quasi-surely), the signature path (which consists of iterated path integrals in every degree) of Brownian motion is non-self-intersecting. This property relates closely to a non-degeneracy property for the Brownian rough path arising naturally from the uniqueness of signature problem in rough path theory. As an important consequence we conclude that quasi-surely, the Brownian rough path does not have any tree-like pieces and every sample path of Brownian motion is uniquely determined by its signature up to reparametrization.

1 Introduction

In 1954, motivated from the study of homotopy theory and loop space homology, Chen [4] proposed a way of representing a vector-valued path x by a fully non-commutative tensor series

$$S(x) = \sum_{n=0}^{\infty} \int_{0 < t_1 < \cdots < t_n < T} dx_{t_1} \otimes \cdots \otimes dx_{t_n} \tag{1.1}$$

of iterated path integrals. In recent literature, this representation is known as the *signature* of a path. Intuitively, the first degree of $S(x)$ is the increment of x, and the second degree of $S(x)$ encodes the geometric signed area enclosed by x and the chord connecting its end points. In general, the signature is a global quantity which captures the total "area" in each degree produced by the underlying path.

---

[*]Department of Mathematics and Statistics , University of Reading, Reading RG6 6AX, United Kingdom. Email: h.s.boedihardjo@reading.ac.uk.

[†]Mathematical Institute, University of Oxford, Oxford OX2 6GG, United Kingdom. Email: xi.geng@maths.ox.ac.uk.

[‡]Mathematical Institute, University of Oxford, Oxford OX2 6GG, United Kingdom. Email: xuan.liu@maths.ox.ac.uk.

[§]Mathematical Institute, University of Oxford, Oxford OX2 6GG, United Kingdom. Email: zhongmin.qian@maths.ox.ac.uk.

The fundamental importance of the signature representation lies in the fact that it is essentially faithful: the signature uniquely determines the underlying path in a certain sense. This is a deep point as it reveals the relationship between local and global properties of a path. The first result along this direction was contained in Chen's original work [5] in 1958, in which he proved that an irreducible and piecewise regular path is uniquely determined by its signature up to translation and reparametrization.

However, the class of paths Chen studied is very special as it does not reveal a crucial invariance property of the signature map: a piece along which the path x goes out and traces back does not contribute to the signature of x. The characterization of this invariance property in a precise mathematical form is the key point of understanding in what sense a generic path x is uniquely determined by its signature. It was after five decades that Hambly and Lyons [10] first gave a complete characterization in the case of continuous paths with bounded variation. In particular, they showed that a continuous path with bounded variation is uniquely determined by its signature up to tree-like equivalence in their sense defined in terms of a height function.

Since the work of Hambly and Lyons, many efforts have been made to explore beyond the bounded variation setting. For applications in probability theory, a natural class of paths to be considered is the space of rough paths, as it is well known that a large amount of interesting stochastic processes can be regarded as rough paths in a canonical way. However, in the rough path setting, Hambly and Lyons' characterization does not apply any more as their tree-like characterization forces the underlying path to have bounded variation. It was in a recent work of Boedihardjo, Geng, Lyons and Yang [2] that the right characterization for the above invariance property was identified in terms of a real tree structure and the corresponding uniqueness result for signature was established.

On the other hand, if we consider the uniqueness problem for sample paths of a stochastic process, we might expect stronger results since a stochastic process usually has non-degenerate sample paths and the above invariance phenomenon will not appear at all. A series of probabilistic works have been done along this direction, originally for Brownian motion by Le Jan and Qian [14], which was later extended to hypoelliptic diffusions by Geng and Qian [9] and Gaussian processes by Boedihardjo and Geng [1]. Formally the result can be stated as the fact that with probability one, every sample path of the underlying stochastic process is uniquely determined by its signature up to translation and reparametrization.

The techniques involved in studying the uniqueness problem for signature in the deterministic and probabilistic settings are very different. Moreover, the deterministic result is weaker but it treats all possible rough paths in one goal, while the probabilistic result is stronger but we have to work in the support of the law of the underlying process on path space. The link between the deterministic and probabilistic approaches seems to be missing, and the main goal of the present paper is to fill in this gap in a relatively robust way.

To be more precise, we will be interested in the following non-degeneracy

property for the Brownian rough path: *it is not possible for a path having a piece along which the path goes out and traces back* (the precise mathematical statement will be made in the next section). As discussed before, this non-degeneracy property arises naturally from the deterministic uniqueness of signature problem for rough paths. In particular, we are going to prove this non-degeneracy property in Malliavin's capacity setting, which is much stronger than the probability measure case and it reveals finer analytic structure over the Wiener space than the underlying probability measure. According to the deterministic uniqueness result for signature in [2], a direct consequence is that outside a slim set in the sense of Malliavin or quasi-surely (see definition in the next Section), every sample path of Brownian motion is uniquely determined by its signature up to reparametrization.

The main motivation of investigating quasi-sure analysis for the Brownian rough path lies in the fundamental work of Sugita [20] in 1988 that capacity is a universal object with respect to a large class of positive generalized Wiener functionals. Therefore, the quasi-sure analysis provides a powerful universal tool in studying degenerate functionals (for instance the Brownian bridge or pinned diffusions) on the Brownian rough path.

According to [2], the aforementioned quasi-sure non-degeneracy property for the Brownian rough path is *equivalent* to the quasi-sure non-self-intersection for the Brownian signature path. Indeed, we are going to obtain a quantitative constraint on the degree n of signature, the dimension d of Brownian motion and the capacity index (r, q) (see (2.1) in the next Section for definition), under which the truncated Brownian signature path up to degree n is non-self-intersecting outside a set of zero (r, q)-capacity.

Intersection properties for random walks and stochastic processes is a classical topic in probability theory, and it has important applications in statistical field theory. The non-self-intersection of sample paths of Brownian motion was studied extensively in the literature. The first result dates back to 1944, in which Kakutani [13] proved that almost every sample path of Brownian motion is non-self-intersecting if the dimension $d \geqslant 5$. Later on, it was known by Dvoretzky, Erdős and Kakutani [6] that the optimal dimension is $d = 4$. The technique of Kakutani was extended to the capacity setting on Wiener space by Fukushima [8]. In particular, he showed that outside a set of zero $(1, 2)$-capacity, every sample path of Brownian motion is non-self-intersecting if $d \geqslant 7$. This result was further extended by Takeda [21] for general (r, q)-capacities under the constraint $d > rq + 4$. It is remarkable that in Fukushima's setting, Lyons [15] proved that the optimal dimension is $d = 6$. However, it is not known (and we expect that it is not true) that outside a slim set every sample path of Brownian motion is non-self-intersecting regardless of the dimension d.

Our technique of proving the quasi-sure non-self-intersection of the Brownian signature path is inspired by the general ideas contained in the aforementioned series of works. In particular, the key ingredient is to establish a maximal type capacity estimate and a small ball capacity estimate for the signature path. However, it will be clear that our technique is robust enough to be extended to more general Gaussian processes as it does not rely on the explicit distribution

3

of Brownian motion and any martingale properties, which is indeed the case for the aforementioned works. In contrast to the Brownian motion, as the Brownian signature path is an infinite dimensional process taking values in the algebra of tensor series, it is not entirely surprising that a quasi-sure non-self-intersection result can be expected.

According to Sugita's work in [20], our result implies the corresponding almost-sure non-degeneracy property and uniqueness of signature result for any probability measure on W associated with a positive generalized Wiener functional.

The present paper is organized in the following way. In Section 2 we formulate our main result, in Section 3 we develop the proofs and in Section 4 we give a few remarks as conclusion.

2 Formulation of Main Result

In this section, we present the basic notions in quasi-sure analysis and formulate our main result. We refer the reader to [18] and [19] for a systematic introduction to the Malliavin calculus and quasi-sure analysis.

Let $(W, \mathcal{B}(W), \mathbb{P})$ be the canonical Wiener space over $\mathbb{R}^d$. In other words, W is the space of continuous paths $w : [0, 1] \to \mathbb{R}^d$ starting at the origin equipped with the uniform topology, $\mathcal{B}(W)$ is the Borel σ-algebra and $\mathbb{P}$ is the canonical Wiener measure. Let $\mathcal{H}$ be the space of absolutely continuous paths in W with square integrable derivative with respect to the Lebesgue measure. It is well known that the canonical embedding $\iota : \mathcal{H} \to W$ give rise to the structure of an abstract Wiener space in the sense of Gross. Let $\iota^* : W^* \to \mathcal{H}^* \cong \mathcal{H}$ be the corresponding dual embedding.

Consider the space $\mathcal{P}$ of polynomial functionals over W, which consists of functionals of the form $F = f(\varphi_1, \cdots, \varphi_n)$, where f is a polynomial over $\mathbb{R}^n$ and $\varphi_1, \cdots, \varphi_n \in W^*$. The *Malliavin derivative* of F is the $\mathcal{H}$-valued functional

$$DF = \sum_{i=1}^{n} \frac{\partial f}{\partial x^i}(\varphi_1, \cdots, \varphi_n) \iota^* \varphi_i.$$

This definition extends to Hilbert space valued polynomial functionals in a natural way. In particular, the r-th derivative of $F \in \mathcal{P}$ can be defined inductively as an $\mathcal{H}^{\otimes r}$-valued functional. For $r \in \mathbb{N}$ and $q \geqslant 1$, the (r, q)-Sobolev norm of F is defined to be

$$\|DF\|_{r,q} = \left(\sum_{i=0}^{r} \mathbb{E}[\|D^i F\|_{\mathcal{H}^{\otimes i}}^q] \right)^{\frac{1}{q}}.$$

The (r, q)-*Sobolev space* $\mathbb{D}_{r,q}$ is the completion of $\mathcal{P}$ under the (r, q)-Sobolev norm.

Throughout the rest we always assume that $r \in \mathbb{N}$ and $q > 1$.

Let O be an open subset of W. The (r,q)-*capacity* of O is defined to be

$$\mathrm{Cap}_{r,q}(O) = \inf\{\|F\|_{r,q} : \ F \in \mathbb{D}_{r,q}, \ F \geqslant 1 \text{ on } O, \ F \geqslant 0 \text{ on } W \text{ for } \mathbb{P}-\text{a.s.}\}.$$

For a general subset $A \subset W$, its (r,q)-capacity is defined to be

$$\mathrm{Cap}_{r,q}(A) = \inf\{\mathrm{Cap}_{r,q}(O) : \ O \text{ open}, \ A \subset O\}. \tag{2.1}$$

It is not hard to see that the (r,q)-capacity is non-negative, increasing and sub-additive. Moreover, the following inequality holds:

$$\mathbb{P}(A)^{\frac{1}{q}} = \mathrm{Cap}_{0,q}(A) \leqslant \mathrm{Cap}_{r,q}(A), \ \forall A \in \mathcal{B}(W).$$

Therefore, capacities are finer scales in measuring the size of a set from an analytic view point than the underlying probability measure.

According to Malliavin, a *slim* set is a subset having zero (r,q)-capacity for every (r,q). A property on paths is said to hold *quasi-surely* if it holds outside a slim set. We are interested in properties which hold quasi-surely.

We will also be working with functions that are defined quasi-surely. A function f on W is said to be (r,q)-*quasi-continuous* if for every $\varepsilon > 0$, there exists an open subset $O \subset W$, such that $\mathrm{Cap}_{r,q}(O) < \varepsilon$ and $f|_{O^c}$ is continuous. A main property for quasi-continuous functions that we will be using is a version of Chebyshev's inequality (see [18], Theorem 2.2):

$$\mathrm{Cap}_{r,q}(|f| > R) \leqslant \frac{M_{r,q}\|f\|_{r,q}}{R}, \ R > 0, \tag{2.2}$$

for any (r,q)-quasi-continuous function $f \in \mathbb{D}_{r,q}$, where $M_{r,q}$ is a constant depending only on r and q.

Now we are in a position to formulate our main result.

The basic object we are interested in is the *Brownian rough path*

$$\mathbf{w} = (1, w^1, w^2) : \ [0,1] \to G^2(\mathbb{R}^d) = \exp\left(\mathbb{R}^d \oplus [\mathbb{R}^d, \mathbb{R}^d]\right)$$

in dimension $d \geqslant 2$, which is the canonical lifting of Brownian motion to the free nilpotent Lie group of order 2 over $\mathbb{R}^d$. Heuristically, through the logarithmic diffeomorphism onto the Lie algebra, the Brownian rough path $\mathbf{w}$ is equivalent to the process

$$\sum_{j=1}^{d} w_t^j e_j + \frac{1}{2} \sum_{1 \leqslant j < k \leqslant d} \left(\int_0^t w_s^j dw_s^k - w_s^k dw_s^j \right) [e_j, e_k],$$

which is the original Brownian motion coupled with its Lévy area process.

In rough path theory, it is an important result of Lyons [16] that any rough path $\mathbf{X}$ with roughness p (or a p-rough path) extends uniquely to a continuous path $\mathbb{X}$ taking values in the algebra

$$T((\mathbb{R}^d)) = \mathbb{R} \oplus \mathbb{R}^d \oplus (\mathbb{R}^d)^{\otimes 2} \oplus \cdots$$

of tensor series, such that the projection of $\mathbb{X}$ onto the truncated tensor algebra up to every degree $n \geqslant \lfloor p \rfloor$ has finite p-variation. Here the truncated tensor algebra $T^{(n)}(\mathbb{R}^d)$ is equipped with the Hilbert-Schmidt norm. Lyons' lifting $\mathbb{X}$ of a rough path $\mathbf{X}$ is also known as the *signature path* of $\mathbf{X}$. This is a generalized notion of taking iterated path integrals in each degree, and the signature is just the end point of the signature path (see (1.1)).

According to Inahama [12] (see also [3]), the Brownian rough path is quasi-surely well defined as the limit of the lifting of dyadic piecewise linear interpolation of Brownian motion under p-variation metric for $2 < p < 3$. Therefore, from Lyons' extension theorem, the Brownian signature path and the signature are well defined quasi-surely.

Remark 2.1. Due to the multiplicative structure in $T((\mathbb{R}^d))$, a rough path $\mathbf{X}$ can either mean an actual path $\mathbf{X}_t$ indexed by a single parameter t or a multiplicative functional $\mathbf{X}_{s,t}$ indexed by a pair $s \leqslant t$. These two notions are interchangeable with each other by setting $\mathbf{X}_{s,t} = \mathbf{X}_s^{-1} \otimes \mathbf{X}_t$ and $\mathbf{X}_t = \mathbf{X}_{0,t}$. In this paper, except for the situation in which we write down the notation explicitly with a double subscript, when referring to a rough path or a signature path, we always mean the actual path with one single parameter.

Our main result can be stated as follows.

Theorem 2.1. *For $n \in \mathbb{N}$, define*

$$\mathcal{O}_n = \{w \in W : \ S_n(\mathbf{w})_{0,s} = S_n(\mathbf{w})_{0,t} \text{ for some } 0 \leqslant s < t \leqslant 1\},$$

where $S_n(\mathbf{w})_{0,t}$ is the truncated Brownian signature path up to degree n. Then $\mathcal{O}_n$ has zero (r,q)-capacity provided

$$\binom{n+d-1}{n} > rq + 4.$$

In particular, the Brownian signature path is non-self-intersecting quasi-surely.

Remark 2.2. Takeda [21] proved that if $d > rq + 4$, outside a set of zero (r,q)-capacity every sample path of Brownian motion is non-self-intersecting. This corresponds to the case of $n = 1$ in Theorem 2.1. Therefore, our main result extends Takeda's result to the higher degree situation.

The non-self-intersection of the Brownian signature path has an important geometric interpretation on the Brownian rough path, which corresponds to the non-degeneracy property mentioned in the introduction and arises naturally from the uniqueness of signature problem in rough path theory. According to [2], this non-degeneracy property can be made precise by using the language of a real tree. Recall that a *real tree* is a metric space τ in which every two distinct points can be joined by a unique non-self-intersecting path (up to reparametrization), and this path is a geodesic.

A continuous path $x : [0,1] \to X$ in some topological space X is called *tree-like*, if there exists a real tree τ, and two continuous maps $\alpha : [0,1] \to \tau$ and $\beta : \tau \to X$ such that $\alpha(0) = \alpha(1)$ and $x = \beta \circ \alpha$. In other words, a tree-like

path is a path which can be realized as a loop in some real tree. A *tree-like piece* of a continuous path x is a pair $s < t$ such that $x|_{[s,t]}$ is tree-like.

From the feature of a real tree, it is clear that the aforementioned non-degeneracy property means the fact that a path does not have any tree-like pieces.

Definition 2.1. A continuous path is called *tree-reduced* if it does not have any tree-like pieces.

It is clear that if a path is non-self-intersecting, then it is tree-reduced.

According to the deterministic uniqueness result for signature in [2], we know that a weakly geometric rough path (a continuous path in the free nilpotent group of order $\lfloor p \rfloor$ with finite p-variation for some $p \geqslant 1$) is tree-like if and only if it has trivial signature. Therefore, a tree-like piece in a rough path corresponds to a loop in its signature path and vice versa. It follows immediately that Theorem 2.1 is *equivalent* to the following, which is already interesting on its own.

Theorem 2.2. *The Brownian rough path is tree-reduced quasi-surely.*

Another important consequence of Theorem 2.1 is a quasi-sure uniqueness result for the signature of Brownian motion. In the uniqueness of signature aspect, this partially extends the work of Le Jan and Qian [14] to the capacity setting. Note that their original work is stronger than uniqueness as it gives an explicit way to reconstruct a sample path of Brownian motion from its signature .

Theorem 2.3. *Outside a slim set $\mathcal{N} \subset W$, two sample paths w and w' of Brownian motion have the same signature if and only if they differ from each other by a reparametrization. In other words, quasi-surely every sample path of Brownian motion is uniquely determined by its signature up to reparametrization.*

3   Proof of the Main Theorem

In this section, we are going to develop the proof of Theorem 2.1 and point out how Theorem 2.3 follows easily from this and the deterministic uniqueness result for signature.

Along the general ideas in the aforementioned works of Kakutani, Fukushima and Takeda, our proof of Theorem 2.1 contains three main steps: a large deviation type capacity estimate for the maximal functional on the signature path, a small ball capacity estimate for the signature path, and a subdivision argument.

A crucial point in our proof is a general and useful technique in rough path theory on controlling higher degree signature components. It consists of a quantitative statement of Lyons' lifting theorem and a technique used by Hambly and Lyons [11] in the construction of stochastic area for Brownian motion on the Sierpinski gasket. We state the result as follows and refer the reader to the monograph by Lyons and Qian [17] for its proof.

Theorem 3.1. *Let* $\mathbf{X} = (1, X^1, \cdots, X^{\lfloor p \rfloor})$ *be a p-rough path.*

(1) Suppose that $\mathbb{X} = (1, X^1, X^2, \cdots)$ *is the signature path of* $\mathbf{X}$. *If there exists a control function* $\omega(s, t)$ *such that*

$$\left| X_{s,t}^i \right| \leqslant \frac{\omega(s,t)^{\frac{i}{p}}}{\beta(i/p)!} \tag{3.1}$$

for $1 \leqslant i \leqslant \lfloor p \rfloor$ *and* $0 \leqslant s < t \leqslant 1$, *where* β *is a constant such that*

$$\beta \geqslant p^2 \left(1 + \sum_{l=1}^{\infty} \left(\frac{2}{l} \right)^{(\lfloor p \rfloor + 1)/p} \right), \tag{3.2}$$

then the inequality (3.1) holds for all $i > \lfloor p \rfloor$ *as well.*

(2) Given a constant $\gamma > p - 1$, *for* $0 \leqslant s \leqslant t \leqslant 1$ *and* $1 \leqslant i \leqslant \lfloor p \rfloor$, *define*

$$\rho_i(\mathbf{X}; s, t) = \sum_{m=1}^{\infty} m^\gamma \sum_{k=1}^{2^m} \left| X_{t_m^{k-1}, t_m^k}^i \right|^{\frac{p}{i}}, \tag{3.3}$$

where $(t_m^k)_{0 \leqslant k \leqslant 2^m}$ *is the dyadic partition of* $[s, t]$. *Then there exists a constant* $C = C(p, \gamma)$, *such that*

$$\sup_{\mathcal{P}([s,t])} \sum_l \left| X_{t_{l-1}, t_l}^i \right|^{\frac{p}{i}} \leqslant C(p, \gamma) \sum_{j=1}^i \rho_j(\mathbf{X}; s, t)$$

for all $1 \leqslant i \leqslant \lfloor p \rfloor$ *and* $0 \leqslant s \leqslant t \leqslant 1$, *where the supremum is taken over all finite partitions of* $[s, t]$.

As we are interested in the Brownian rough path, throughout the rest, we always fix $2 < p < 3$ and the two constants β, γ arising from Theorem 3.1. For simplicity, we always omit the dependence on p, β and γ for a constant, and the value of a constant may change from line to line even the same notation is used.

3.1 A Maximal Type Capacity Estimate

As the first step, we are going to estimate the (r, q)-capacity of the event $\{w : \max_{t \in [t_0, t_1]} \|S_n(\mathbf{w})_{0,t} - S_n(\mathbf{w})_{0,t_0}\| > \eta\}$, where $[t_0, t_1]$ is a dyadic sub-interval of $[0, 1]$ (i.e. $[t_0, t_1] = [(k-1)/2^m, k/2^m]$ for some k, m). Our main idea is to control the maximal function by the series defined by (3.3), and to observe that the increments $w_{s,t}^i$ $(i = 1, 2)$ are "evenly distributed" over a dyadic partition.

For $m \geqslant 1$, let $\mathbf{w}^{(m)} = (1, w^{(m),1}, w^{(m),2})$ be the lifting of the dyadic piecewise linear interpolation of $w|_{[t_0,t_1]}$ over the dyadic partition of $[t_0, t_1]$ into small intervals of length $1/2^m$. In other words, $w^{(m),1}$ is just the increment process of the interpolation, while $w^{(m),2}$ is defined by second order iterated integrals of the interpolation.

Lemma 3.1. *We have the following estimates:*

$$\sup_{m \geqslant 1} \|w_{t_0,t_1}^{(m),i}\|_{L^2} \leqslant C_d |t_1 - t_0|^{\frac{i}{2}}, \text{ for } i = 1, 2,$$

and

$$\|w_{t_0,t_1}^{(m+1),2} - w_{t_0,t_1}^{(m),2}\|_{L^2} \leqslant C_d \frac{|t_1 - t_0|}{2^{m/2}}, \text{ for } m \geqslant 1,$$

where C_d is a constant depending only on the dimension d.

Proof. See [17], Proposition 4.4.1 (and also [3], Lemma 2.3). □

Lemma 3.2. *Suppose $N \in \mathbb{N}$. Then for $i = 1, 2$, we have $|w_{t_0,t_1}^i|^{2N} \in \oplus_{j=0}^{2iN} \mathcal{H}_j$ and*

$$\||w_{t_0,t_1}^i|^{2N}\|_{4N,q} \leqslant C_{N,q,d} |t_1 - t_0|^{iN},$$

where $\mathcal{H}_j$ is the j-th Wiener-Itô chaos and $C_{N,q,d}$ is a constant depending only on N, q and d.

Proof. We only need to consider the case when $i = 2$, as w_{t_0,t_1}^1 is just the increment of Brownian motion in which case the assertion is obvious.

First of all, we have

$$\left| \left| w_{t_0,t_1}^{(m+1),2} \right|^{2N} - \left| w_{t_0,t_1}^{(m),2} \right|^{2N} \right| \leqslant \left| w_{t_0,t_1}^{(m+1),2} - w_{t_0,t_1}^{(m),2} \right|$$

$$\cdot \sum_{k=0}^{2N-1} \left| w_{t_0,t_1}^{(m+1),2} \right|^k \left| w_{t_0,t_1}^{(m),2} \right|^{2N-1-k}.$$

From the hypercontractivity of the Ornstein-Uhlenbeck semigroup, it is well known that the L^q ($q > 1$) and L^2-norms are comparable over a given Wiener-Itô chaos. In particular, we have (see [19], Proposition 2.14):

$$\|F\|_{L^q} \leqslant C_{N,q} \|F\|_{L^2} \tag{3.4}$$

for any $F \in \oplus_{j=0}^N \mathcal{H}_j$. Since $|w_{t_0,t_1}^{(m),2}|^{2k} \in \oplus_{j=0}^{4k} \mathcal{H}_j$, by using Lemma 3.1 and (3.4), it is straight forward to see that

$$\left\| \left| w_{t_0,t_1}^{(m+1),2} \right|^{2N} - \left| w_{t_0,t_1}^{(m),2} \right|^{2N} \right\|_{L^2} \leqslant C_{N,d} \frac{|t_1 - t_0|^{2N}}{2^{m/2}}. \tag{3.5}$$

Therefore, $\left| w_{t_0,t_1}^{(m),2} \right|^{2N}$ converges in L^2 as $m \to \infty$. This shows that $|w_{t_0,t_1}^2|^{2N} \in \oplus_{j=0}^{4N} \mathcal{H}_j$ as it is the almost-sure limit of $\left| w_{t_0,t_1}^{(m),2} \right|^{2N}$.

Moreover, the Sobolev norm of $\left| w_{t_0,t_1}^{(m+1),2} \right|^{2N} - \left| w_{t_0,t_1}^{(m),2} \right|^{2N}$ can by controlled by the L^2-norm uniformly as they are polynomials of a fixed degree (see [3], Lemma 2.2) . In particular, we obtain from (3.5) that

$$\left\| \left| w_{t_0,t_1}^{(m+1),2} \right|^{2N} - \left| w_{t_0,t_1}^{(m),2} \right|^{2N} \right\|_{4N,q} \leqslant C_{N,q,d} \frac{|t_1 - t_0|^{2N}}{2^{m/2}}.$$

This implies that $w_{t_0,t_1}^{(m),2} \to w_{t_0,t_1}^2$ in $\mathbb{D}_{4N,q}$ as $m \to \infty$.

Therefore, by Lemma 3.1 we have

$$
\begin{aligned}
\||w_{t_0,t_1}^2|^{2N}\|_{4N,q} &= \lim_{m\to\infty} \||w_{t_0,t_1}^{(m),2}|^{2N}\|_{4N,q} \\
&\leqslant C_{N,q} \lim_{m\to\infty} \|w_{t_0,t_1}^{(m),2}\|_{L^2}^{2N} \\
&\leqslant C_{N,q,d}|t_1 - t_0|^{2N}.
\end{aligned}
$$

$\square$

Now we are able to establish the required maximal capacity estimate.

Proposition 3.1. *(1) Let $\alpha_p = \sum_{j=1}^\infty 1/(\beta(j/p)!) < 1$. Then for any $n \in \mathbb{N}$, we have:*

$$
\mathrm{Cap}_{r,q} \left(\max_{t_0 \leqslant t \leqslant t_1} \|S_n(\mathbf{w})_{t_0,t} - \mathbf{1}\| > \alpha \right)
$$

$$
\leqslant \begin{cases} C_{N,q,d} \dfrac{|t_1-t_0|^N}{\alpha^{2N}} \left(1 + \dfrac{|t_1-t_0|^N}{\alpha^{2N}}\right), & \text{if } 0 < \alpha \leqslant \alpha_p; \\[2ex] C_{N,q,d} \dfrac{|t_1-t_0|^N}{\alpha^{2N/n}} \left(1 + \dfrac{|t_1-t_0|^N}{\alpha^{2N/n}}\right), & \text{if } \alpha > \alpha_p, \end{cases} \tag{3.6}
$$

where $N > r$ and $C_{N,q,d}$ is a constant depending only on N, q, d.

(2) Suppose $N > r$ and $\delta > 0$. Then for any $n \in \mathbb{N}$ and $0 < \eta < 2^{-1/\delta} \wedge \alpha_p$, we have

$$
\mathrm{Cap}_{r,q} \left(\max_{t_0 \leqslant t \leqslant t_1} \|S_n(\mathbf{w})_{0,t} - S_n(\mathbf{w})_{0,t_0}\| > \eta \right)
$$

$$
\leqslant C_{N,q,d} \left(\frac{|t_1-t_0|^N}{\eta^{2N(1+\delta)}} \cdot \left(1 + \frac{|t_1-t_0|^N}{\eta^{2N(1+\delta)}}\right) + \eta^{2N\delta/n} \right), \tag{3.7}
$$

where $C_{N,q,d}$ is a constant depending only on N, q, d.

Proof. (1) Let $C_p = (1/p)! + (2/p)!$ and β be given by (3.2). Define a control function $\omega(s,t)$ by

$$
\omega(s,t) = \beta C_p \sum_{i=1}^2 \sup_{\mathcal{P}([s,t])} \sum_l |w_{t_{l-1},t_l}^i|^{\frac{p}{i}}, \quad 0 \leqslant s \leqslant t \leqslant 1. \tag{3.8}
$$

Since $\mathbf{w}$ is a quasi-surely well defined p-rough path, according to Theorem 3.1 (1), we have

$$
\max_{t_0 \leqslant t \leqslant t_1} |w_{t_0,t}^i| \leqslant \frac{\omega(t_0,t_1)^{i/p}}{\beta(i/p)!}
$$

for all $i \geqslant 1$, where w^i denotes the i-th degree component of the Brownian signature path. For the moment let $\lambda > 0$ be such that

$$
\sum_{i=1}^n \frac{\lambda^{i/p}}{\beta(i/p)!} \leqslant \alpha
$$

10

for given $\alpha > 0$. It follows that

$$\left\{ w : \max_{t_0 \leqslant t \leqslant t_1} \|S_n(\mathbf{w})_{t_0,t} - \mathbf{1}\| > \alpha \right\}$$

$$= \left\{ w : \max_{t_0 \leqslant t \leqslant t_1} \left(\sum_{i=1}^{n} |w_{t_0,t}^i|^2 \right) > \alpha^2 \right\}$$

$$\subseteq \left\{ w : \sum_{i=1}^{n} \max_{t_0 \leqslant t \leqslant t_1} |w_{t_0,t}^i|^2 > \sum_{i=1}^{n} \frac{\lambda^{2i/p}}{(\beta(i/p)!)^2} \right\}$$

$$\subseteq \bigcup_{i=1}^{n} \left\{ w : \max_{t_0 \leqslant t \leqslant t_1} |w_{t_0,t}^i| > \frac{\lambda^{i/p}}{\beta(i/p)!} \right\}$$

$$\subseteq \{ w : \omega(t_0, t_1) > \lambda \}. \tag{3.9}$$

Let $\rho_i(\mathbf{w}; t_0, t_1)$ $(i = 1, 2)$ be given by (3.3). According to Theorem 3.1 (2), we obtain that

$$\{ w : \omega(t_0, t_1) > \lambda \} \quad \subseteq \quad \{ w : \rho_1(\mathbf{w}; t_0, t_1) > C\lambda \}$$
$$\bigcup \{ w : \rho_2(\mathbf{w}; t_0, t_1) > C\lambda \}, \tag{3.10}$$

where $C > 0$ is some constant depending only on p and γ in that theorem.

For $\theta > 0$, let $C_\theta > 0$ be a constant such that

$$C_\theta \sum_{m=1}^{\infty} m^\gamma 2^{-m\theta} \leqslant C.$$

It follows that

$$\{ w : \rho_i(\mathbf{w}; t_0, t_1) > C\lambda \} \quad \subseteq \quad \bigcup_{m=1}^{\infty} \left\{ w : \sum_{k=1}^{2^m} \left| w_{t_m^{k-1}, t_m^k}^i \right|^{\frac{p}{i}} > C_\theta \lambda 2^{-m\theta} \right\}$$

$$\subseteq \quad \bigcup_{m=1}^{\infty} \bigcup_{k=1}^{2^m} \left\{ w : \left| w_{t_m^{k-1}, t_m^k}^i \right|^{\frac{p}{i}} > C_\theta \lambda 2^{-m(\theta+1)} \right\}.$$

Therefore, for any $N > r$, we have

$$\mathrm{Cap}_{r,q} \left(\rho_i(\mathbf{w}; t_0, t_1) > C\lambda \right)$$

$$\leqslant \quad \sum_{m=1}^{\infty} \sum_{k=1}^{2^m} \mathrm{Cap}_{r,q} \left(\left| w_{t_m^{k-1}, t_m^k}^i \right|^{\frac{p}{i}} > C_\theta \lambda 2^{-m(\theta+1)} \right)$$

$$\leqslant \quad \sum_{m=1}^{\infty} \sum_{k=1}^{2^m} \mathrm{Cap}_{4N,q} \left(\left| w_{t_m^{k-1}, t_m^k}^i \right|^{2N} > (C_\theta \lambda)^{\frac{2iN}{p}} 2^{-\frac{2im(1+\theta)N}{p}} \right).$$

On the other hand, from the proof of Lemma 3.2, we know that $|w_{t_m^{k-1}, t_m^k}^{(l),i}|^{2N} \to$
$|w_{t_m^{k-1}, t_m^k}^i|^{2N}$ in $\mathbb{D}_{4N,q}$ as well as quasi-surely when $l \to \infty$. Since $|w_{t_m^{k-1}, t_m^k}^{(l),i}|^{2N}$

11

is continuous on W, according to [18], Theorem 2.3.5, we see that $\left|w^i_{t^{k-1}_m, t^k_m}\right|^{2N}$ is $(4N, q)$-quasi-continuous. Therefore, by using the Chebyshev inequality for capacity (2.2) and Lemma 3.2, we have

$$
\begin{aligned}
\mathrm{Cap}_{r,q}(\rho_i(\mathbf{w}; t_0, t_1) > C\lambda) \quad &\leqslant \quad C_{N,q}(C_\theta \lambda)^{-\frac{2iN}{p}} \sum_{m=1}^{\infty} 2^{\frac{2im(1+\theta)N}{p}} \\
&\qquad \cdot \sum_{k=1}^{2^m} \left\| \left|w^i_{t^{k-1}_m, t^k_m}\right|^{2N} \right\|_{4N,q} \\
&\leqslant \quad C_{N,q,d}(C_\theta \lambda)^{-\frac{2iN}{p}} |t_1 - t_0|^{iN} \\
&\qquad \cdot \sum_{m=1}^{\infty} 2^{m\left(iN\left(\frac{2(1+\theta)}{p}-1\right)+1\right)}.
\end{aligned}
$$

Now we choose θ to be small enough such that

$$
\left(\frac{2(1+\theta)}{p} - 1\right) N + 1 < 0.
$$

This is possible since $2 < p < 3$. Then we arrive at

$$
\mathrm{Cap}_{r,q}\left(\rho_i(\mathbf{w}; t_0, t_1) > C\lambda\right) \leqslant C_{N,q,d} \frac{|t_1 - t_0|^{iN}}{\lambda^{2iN/p}}.
$$

Combining with (3.9) and (3.10), we obtain that

$$
\begin{aligned}
\mathrm{Cap}_{r,q}\left(\max_{t_0 \leqslant t \leqslant t_1} \|S_n(\mathbf{w})_{t_0,t} - \mathbf{1}\| > \alpha\right) \quad &\leqslant \quad C_{N,q,d} \frac{|t_1 - t_0|^N}{\lambda^{2N/p}} \\
&\qquad \cdot \left(1 + \frac{|t_1 - t_0|^N}{\lambda^{2N/p}}\right).
\end{aligned}
$$

Now (3.6) follows by setting

$$
\lambda = \begin{cases} (\alpha/\alpha_p)^p, & \text{if } 0 < \alpha \leqslant \alpha_p; \\ (\alpha/\alpha_p)^{p/n}, & \text{if } \alpha > \alpha_p. \end{cases}
$$

(2) From the multiplicative property of a rough path, we know that

$$
S_n(\mathbf{w})_{0,t} - S_n(\mathbf{w})_{0,t_0} = S_n(\mathbf{w})_{0,t_0} \otimes (S_n(\mathbf{w})_{t_0,t} - \mathbf{1}).
$$

Therefore, for any $\delta > 0$, we have

$$
\begin{aligned}
&\{w : \max_{t_0 \leqslant t \leqslant t_1} \|S_n(\mathbf{w})_{0,t} - S_n(\mathbf{w})_{0,t_0}\| > \eta\} \\
\subseteq \quad &\{w : \|S_n(\mathbf{w})_{0,t_0}\| > \eta^{-\delta}\} \bigcup \{w : \max_{t_0 \leqslant t \leqslant t_1} \|S_n(\mathbf{w})_{t_0,t} - \mathbf{1}\| > \eta^{1+\delta}\}.
\end{aligned}
$$

If $0 < \eta < \alpha_p$, then $\eta^{1+\delta} < \alpha_p$, and from (3.6) we conclude that

$$
\begin{aligned}
\mathrm{Cap}_{r,q}\left(\max_{t_0 \leqslant t \leqslant t_1} \|S_n(\mathbf{w})_{t_0,t} - \mathbf{1}\| > \eta^{1+\delta}\right) \quad &\leqslant \quad C_{N,q,d}\frac{|t_1 - t_0|^N}{\eta^{2N(1+\delta)}} \\
&\qquad \cdot \left(1 + \frac{|t_1 - t_0|^N}{\eta^{2N(1+\delta)}}\right).
\end{aligned}
$$

On the other hand, as $\eta^{-\delta}/2 > 1 > \alpha_p$, by applying (3.6) for the case $[t_0, t_1] = [0,1]$, we obtain that

$$
\begin{aligned}
\mathrm{Cap}_{r,q}\left(\|S_n(\mathbf{w})_{0,t_0}\| > \eta^{-\delta}\right) \quad &\leqslant \quad \mathrm{Cap}_{r,q}\left(\max_{0 \leqslant t \leqslant 1}\|S_n(\mathbf{w})_{0,t} - \mathbf{1}\| > \frac{1}{2}\eta^{-\delta}\right) \\
&\leqslant \quad C_{N,q,d}\eta^{\frac{2N\delta}{n}}\left(1 + \eta^{\frac{2N\delta}{n}}\right) \\
&\leqslant \quad C_{N,q,d}\eta^{\frac{2N\delta}{n}}.
\end{aligned}
$$

Now (3.7) follows immediately. $\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\square$

Remark 3.1. In the probability measure case (i.e. $r = 0$), it is possible to strengthen the maximal inequalities in Proposition 3.1 to an exponential type by using a Fernique type estimate for the p-variation of the Brownian rough path. However, this approach cannot be applied to the capacity case as the Chebyshev inequality for capacity involves the Sobolev norm instead of the L^q-norm. Indeed, it is even not clear whether the p-variation is differentiable in the sense of Malliavin.

3.2 A Small Ball Capacity Estimate

The second step is to establish an estimate for the (r, q)-capacity of the event $\{w : |S_n(\mathbf{w})_{0,t_1} - S_n(\mathbf{w})_{0,t_0}| \leqslant \eta\}$, where t_0, t_1 are two dyadic points in $[0,1]$. The key ingredient here is to observe the hypoellipticity of a collection of signature components regarded as a stochastic differential equation (short for SDE), so that the required estimate will follow from the Malliavin calculus for hypoelliptic SDEs. It should be pointed out that $S_n(\mathbf{w})_{0,t}$, as a path in the truncated tensor algebra, is not hypoelliptic as it lives on the free nilpotent Lie group.

Recall from rough path theory (see [7], Proposition 7.8) that the truncated signature path $S_n(\mathbf{w})_{t_0,t}$ satisfies the linear differential equation

$$
\begin{cases}
dS_n(\mathbf{w})_{t_0,t} = S_n(\mathbf{w})_{t_0,t} \otimes dw_t, \\
S_n(\mathbf{w})_{t_0,t_0} = \mathbf{1}.
\end{cases}
\tag{3.11}
$$

In our case we can either interpret (3.11) as a rough differential equation or a Stratonovich type SDE driven by Brownian motion. Under the canonical basis of $\mathbb{R}^d$, we can write it as

$$
dw_{t_0,t}^I = w_{t_0,t}^{I'} \circ dw^i
\tag{3.12}
$$

starting at zero, where I runs over all words over $\{1, \cdots, d\}$ with length at most n, and I' is the word obtained by dropping the last letter i. We are interested in a *consistent* collection $\mathcal{I}$ of words in the sense that $I \in \mathcal{I} \implies I' \in \mathcal{I}$.

Definition 3.1. A word I over $\{1, \cdots, d\}$ is said to be *non-degenerate* if it has the form

$$I = \left(i_0, \underbrace{i_1, \cdots, i_1}_{l_1 \text{ copies}}, \cdots, \underbrace{i_k, \cdots, i_k}_{l_k \text{ copies}} \right),$$

where $1 \leqslant i_0 < i_1 < \cdots < i_k \leqslant d$ and $l_1, \cdots, l_k \geqslant 0$.

The collection $\mathcal{I}_{d,n}$ of non-degenerate words with length at most n is clearly consistent.

Lemma 3.3. *The cardinality of $\mathcal{I}_{d,n}$ is given by*

$$|\mathcal{I}_{d,n}| = \left(\begin{array}{c} n + d - 1 \\ n \end{array} \right).$$

Proof. For $1 \leqslant k \leqslant d$, let $\mathcal{I}_{d,n}(k)$ be the set of words $I \in \mathcal{I}_{d,n}$ whose first letter is k. It is not hard to see that there is a bijection between $\mathcal{I}_{d,n}(k)$ and the set of non-negative integer solutions to the equation

$$x_{k+1} + \cdots + x_d + y = n - 1.$$

It follows that

$$|\mathcal{I}_{d,n}(k)| = \left(\begin{array}{c} n - 1 + d - k \\ d - k \end{array} \right).$$

Therefore,

$$|\mathcal{I}_{d,n}| = \sum_{k=1}^{d} \left(\begin{array}{c} n - 1 + d - k \\ d - k \end{array} \right) = \sum_{l=n}^{n+d-1} \left(\begin{array}{c} l - 1 \\ n - 1 \end{array} \right).$$

The last expression is easily seen to be $\left(\begin{array}{c} n + d - 1 \\ n \end{array} \right)$ as it can be modeled by choosing subsets of $\{1, \cdots, n + d - 1\}$ with n elements and with l being the largest one. $\qquad\square$

Now we have the following result.

Lemma 3.4. *The restriction of (3.12) to the collection $\mathcal{I}_{d,n}$ of non-degenerate words defines an $|\mathcal{I}_{d,n}|$-dimensional linear SDE satisfying Hörmander's condition at the origin in the sense that the linear span of its generating vector fields and their Lie brackets of any order at the origin is $\mathbb{R}^{|\mathcal{I}_{d,n}|}$.*

Proof. First of all, the consistency of $\mathcal{I}_{d,n}$ implies that the restriction of (3.12) to $\mathcal{I}_{d,n}$ is itself a linear SDE of dimension $|\mathcal{I}_{d,n}|$. It suffices to verify Hörmander's condition at the origin.

We use the notation $X^{k;I}$ for a component to keep track of the length k of the word $I \in \mathcal{I}_{d,n}$. In geometric notation, the generating vector fields of the SDE (3.12) restricted to $\mathcal{I}_{d,n}$ are given by

$$V_i = \sum_{I_i} x^{k-1;I_i'} \partial_{k;I_i}, \ 1 \leqslant i \leqslant d,$$

where we set $x^{0;I_i'} = 1$. Here the sum is taken over all words $I_i \in \mathcal{I}_{d,n}$ whose last letter is i, and I_i' is the word obtained by dropping the last letter from I_i.

We write

$$V_i = \partial_{1;(i)} + P_i,$$

where

$$P_i = \sum_{|I_i| \geqslant 2} x^{k-1;I_i'} \partial_{k;I_i}$$

is a vector field with homogeneous linear coefficients. For $1 \leqslant i < j \leqslant d$, we then have

$$
\begin{aligned}
[V_i, V_j] &= [\partial_{1;(i)} + P_i, \partial_{1;(j)} + P_j] \\
&= \partial_{1;(i)} P_j - \partial_{1;(j)} P_i + [P_i, P_j] \\
&= \partial_{2;(i,j)} - \partial_{1;(j)} P_i + [P_i, P_j],
\end{aligned}
$$

where the ∂P denotes the vector field obtained by differentiating the coefficients of P.

Now the key observation is that if $i \leqslant j$, then P_i does not depend on $x^{1;(j)}$. Indeed, if this is not the case, then $I_i = (j, i)$ has to be a word appearing in the summation, contradicting the construction of $\mathcal{I}_{d,n}$. Therefore,

$$\partial_{1;(j)} P_i = 0$$

and we have

$$[V_i, V_j] = \partial_{2;(i,j)} + [P_i, P_j].$$

Note that $[P_i, P_j]$ is a vector field with homogeneous linear coefficients of the form $x^{k-1;I_i'}$ or $x^{l-1;I_j'}$ $(k, l \geqslant 2)$.

If $1 \leqslant i < j_1 \leqslant j_2 \leqslant d$, then

$$
\begin{aligned}
[[V_i, V_{j_1}], V_{j_2}] &= [\partial_{2;(i,j_1)} + [P_i, P_{j_1}], \partial_{1;(j_2)} + P_{j_2}] \\
&= \partial_{3;(i,j_1,j_2)} - \partial_{1;j_2}[P_i, P_{j_1}] + [[P_i, P_{j_1}], P_{j_2}].
\end{aligned}
$$

Again we know that the second term on the right hand side vanishes as $[P_i, P_{j_1}]$ does not depend on $x^{1;(j_2)}$.

By an induction argument, we obtain the fact that

$$[\cdots[[V_i, V_{j_1}], V_{j_2}], \cdots, V_{j_m}] = \partial_{m+1;(i,j_1,\cdots,j_m)} + P_{i,j_1,\cdots,j_m},$$

15

for all $1 \leqslant i < j_1 \leqslant \cdots \leqslant j_m \leqslant d$ and $1 \leqslant m \leqslant n-1$, where

$$P_{i,j_1,\cdots,j_m} = [\cdots[[P_i, P_{j_1}], P_{j_2}], \cdots, P_{j_m}]$$

is a vector field with homogeneous linear coefficients not depending on any $x^{1;(j)}$ with $j \geqslant j_m$. In particular,

$$P_{i,j_1,\cdots,j_m}(0) = 0$$

and we conclude that the linear span of V_i and their Lie brackets at the origin coincides with $\mathrm{Span}\{\partial_{k;I} : I \in \mathcal{I}_{d,n}\}$, which is $\mathbb{R}^{|\mathcal{I}_{d,n}|}$. $\qquad\square$

According to Lemma 3.4 and Hörmander's theorem from the Malliavin calculus, we know that the law of $(w^I_{t_0,t})_{I \in \mathcal{I}_{d,n}}$ has a smooth density with respect to the Lebesgue measure. A small ball probability estimate follows immediately from this fact. To obtain a corresponding capacity estimate, we need to the following lemma.

Lemma 3.5. *For any $i \geqslant 1$, we have $w^i_{t_0,t_1} \in \oplus^i_{j=0}\mathcal{H}_j$ and*

$$\|w^i_{t_0,t_1}\|_{r,q} \leqslant C_{r,q,i,d}|t_1 - t_0|^{\frac{i}{2}}, \tag{3.13}$$

where $C_{r,q,i,d}$ is a constant depending only on r, q, i and d.

Proof. Let $\widetilde{\mathbf{w}}^{(m)}$ be the lifting of the m-th dyadic piecewise linear interpolation of w over $[0,1]$. Define the control function $\omega_m(s,t)$ in the same way as in (3.8) by replacing $\mathbf{w}$ by $\widetilde{\mathbf{w}}^{(m)}$. According to Theorem 3.1, we have

$$|\widetilde{w}^{(m),i}_{s,t}| \leqslant \frac{\omega_m(s,t)^{i/p}}{\beta(i/p)!}, \ \forall i \geqslant 1,$$

and

$$\omega_m(s,t) \leqslant C\left(\rho_1(\widetilde{\mathbf{w}}^{(m)}; s, t) + \rho_2(\widetilde{\mathbf{w}}^{(m)}; s, t)\right),$$

where C is a constant depending only on p and γ. Therefore,

$$\|w^{(m),i}_{t_0,t_1}\|_{L^2} \leqslant C_i\left(\|\rho_1(\widetilde{\mathbf{w}}^{(m)}; t_0, t_1)^{\frac{i}{p}}\|_{L^2} + \|\rho_2(\widetilde{\mathbf{w}}^{(m)}; t_0, t_1)^{\frac{i}{p}}\|_{L^2}\right).$$

It follows that

$$\begin{aligned}
\|\rho_j(\widetilde{\mathbf{w}}^{(m)}; t_0, t_1)\|_{L^{\frac{2i}{p}}} &\leqslant \sum_{l=1}^{\infty} l^\gamma \sum_{k=1}^{2^l} \||\widetilde{w}^{(m),j}_{t_l^{k-1}, t_l^k}|^{\frac{p}{j}}\|_{L^{\frac{2i}{p}}} \\
&\leqslant C_i \sum_{l=1}^{\infty} l^\gamma \sum_{k=1}^{2^l} \|\widetilde{w}^{(m),j}_{t_l^{k-1}, t_l^k}\|_{L^2}^{\frac{p}{j}} \\
&\leqslant C_{i,d}|t_1 - t_0|^{\frac{p}{2}},
\end{aligned}$$

where the last inequality follows from the fact that

$$\|\widetilde{w}^{(m),j}_{s,t}\|_{L^2} \leqslant C_d|t - s|^{\frac{j}{2}}$$

16

for all $0 \leqslant s \leqslant t \leqslant 1$ and $m \geqslant 1$, even in the case when $[s, t]$ is not a dyadic sub-interval of $[0, 1]$. This can be seen easily based on the computation in [17], pp. 68–70. Therefore, we obtain that

$$\|\widetilde{w}_{t_0,t_1}^{(m),i}\|_{L^2} \leqslant C_{i,d}|t_1 - t_0|^{\frac{i}{2}}. \tag{3.14}$$

On the other hand, since $\widetilde{w}_{t_0,t_1}^{(m),i} \in \oplus_{j=0}^{i} \mathcal{H}_j$, from (3.14) and (3.4) we know that the L^q-norm of $\widetilde{w}_{t_0,t_1}^{(m),i}$ is uniformly bounded for any $q > 2$. As $\widetilde{w}_{t_0,t_1}^{(m),i} \to w_{t_0,t_1}^{i}$ $\mathbb{P}$-almost-surely, it follows that the convergence holds in L^2 as well. Therefore, $w_{t_0,t_1}^{i} \in \oplus_{j=0}^{i} \mathcal{H}_j$ and it also satisfies (3.14). Finally, (3.13) follows from [3], Lemma 2.2. $\qquad\square$

Now we are able to establish the required small ball capacity estimate.

Proposition 3.2. *Given any $\tau > 1$, we have the following estimate:*

$$\mathrm{Cap}_{r,q}\left(\|S_n(\mathbf{w})_{0,t_1} - S_n(\mathbf{w})_{0,t_0}\| \leqslant \eta\right) \leqslant \frac{C_{r,q,n,d,\tau}}{|t_1 - t_0|^{\lambda_{n,d}/\tau q}} \eta^{\frac{|\mathcal{I}_{d,n}|}{\tau q} - r}$$

for every $0 < \eta < 1$ and $n \in \mathbb{N}$, where $C_{r,q,n,d,\tau}$ is a constant depending only on r, q, n, d and τ.

Proof. Write $X_{t_0,t} = (w_{t_0,t}^{I})_{I \in \mathcal{I}_{d,n}}$ as a diffusion in $\mathbb{R}^{|\mathcal{I}_{d,n}|}$. By the multiplicative structure of the signature path and the fact that $\|S_n(\mathbf{w})_{0,t_0}\| \geqslant 1$, we have

$$\begin{aligned}
\mathrm{Cap}_{r,q}\left(\|S_n(\mathbf{w})_{0,t_1} - S_n(\mathbf{w})_{0,t_0}\| \leqslant \eta\right) &\leqslant \mathrm{Cap}_{r,q}\left(\|S_n(\mathbf{w})_{t_0,t_1} - \mathbf{1}\| \leqslant \eta\right) \\
&\leqslant \mathrm{Cap}_{r,q}\left(\|X_{t_0,t_1}\| \leqslant \eta\right).
\end{aligned}$$

Now consider a function $f \in C^\infty(\mathbb{R}^{|\mathcal{I}_{d,n}|})$ such that

$$\begin{cases}
0 \leqslant f \leqslant 1, \\
f = 1 \text{ on } |x| \leqslant \eta \text{ and } f = 0 \text{ on } |x| \geqslant 2\eta, \\
|\nabla^k f| \leqslant \frac{C_r}{\eta^k}, \text{ for } k \leqslant r,
\end{cases}$$

where C_r is a constant depending only on r. Let $F = f(X_{t_0,t_1})$. It follows that is smooth in the sense of Malliavin. By using Lemma 3.5 and the chain rule, we obtain that

$$\|F\|_{r,q'} \leqslant \frac{C_{r,q',n,d}}{\eta^r}, \ \forall q' > 1. \tag{3.15}$$

Moreover, the same reason as in the proof of Proposition 3.1 shows that F is (r, q)-quasi-continuous. Therefore, according to the Chebyshev inequality (2.2),

for any $\tau > 1$, we have

$$
\begin{aligned}
\mathrm{Cap}_{r,q}(\|X_{t_0,t_1}\| \leqslant \eta) \;&\leqslant\; \mathrm{Cap}_{r,q}(F \geqslant 1) \\
&\leqslant\; C_{r,q}\|F\|_{r,q} \\
&\leqslant\; C_{r,q} \sum_{i=0}^{r} \left(\mathbb{E}[\|D^i F\|^q \mathbf{1}_{\{|X_{t_0,t_1}| \leqslant 2\eta\}}] \right)^{\frac{1}{q}} \\
&\leqslant\; C_{r,q}\|F\|_{r,q_1} \mathbb{P}(\|X_{t_0,t_1}\| \leqslant 2\eta)^{\frac{1}{\tau q}} \\
&\leqslant\; \frac{C_{r,q,n,d,\tau}}{\eta^r} \mathbb{P}(\|X_{t_0,t_1}\| \leqslant 2\eta)^{\frac{1}{\tau q}},
\end{aligned}
$$

where $q_1 = \tau q/(\tau - 1)$.

Finally, according to Lemma 3.4 and Hörmander's theorem from the Malliavin calculus (here we use a quantitative version in [19], Theorem 6.16), X_{t_0,t_1} has a smooth density $p_{t_0,t_1}(x)$ with respect to the Lebesgue measure on $\mathbb{R}^{|\mathcal{I}_{d,n}|}$. In particular, $p_{t_0,t_1}(x)$ satisfies the following estimate:

$$
\sup_{x \in \mathbb{R}^{|\mathcal{I}_{d,n}|}} p_{t_0,t_1}(x) \leqslant \frac{C_{n,d}}{(t_1 - t_0)^{\lambda_{n,d}}}.
$$

Therefore,

$$
\begin{aligned}
\mathrm{Cap}_{r,q}(\|X_{t_0,t_1}\| \leqslant \eta) \;&\leqslant\; \frac{C_{r,q,n,d,\tau}}{\eta^r} \left(\int_{\{x : |x| \leqslant 2\eta\}} p_{t_0,t_1}(x)dx \right)^{\frac{1}{\tau q}} \\
&\leqslant\; \frac{C_{r,q,n,d,\tau}}{|t_1 - t_0|^{\lambda_{n,d}/\tau q}} \eta^{\frac{|\mathcal{I}_{d,n}|}{\tau q} - r}.
\end{aligned}
$$

$\square$

3.3 Kakutani's Sub-division Argument

Following the original sub-division argument of Kakutani [13], we are now in a position to complete the proof of Theorem 2.1.

Here a remarkable fact that we are going to use is the sub-additivity for the q-th power of the (r,q)-capacity (up to a constant depending only on r, q) instead of the original sub-additivity. This sub-additivity can be easily seen by using the renowned Meyer's inequalities and the integral representation formula (see [19], Chapter 4)

$$
(I - \mathcal{L})^{-\frac{r}{2}} = \frac{1}{\Gamma(r/2)} \int_0^\infty e^{-t} t^{\frac{r}{2} - 1} T_t dt,
$$

where $\mathcal{L}$ is the generator of the Ornstein-Uhlenbeck semigroup T_t. If we use the sub-additivity for the capacity itself, we will end up with the quantitative constraint $|\mathcal{I}_{d,n}| > rq + 4q$ for the non-self-intersection property, which is not as sharp as the version we are going to obtain.

Proof of Theorem 2.1. Fix two dyadic sub-intervals $[s_0, s_1]$, $[t_0, t_1]$ with equal length Δ and $s_1 < t_0$. Let $\mathcal{A}_n$ be the event that $S_n(\mathbf{w})_{0,s} = S_n(\mathbf{w})_{0,t}$ for some $s \in [s_0, s_1]$ and $t \in [t_0, t_1]$. It follows that

$$
\begin{aligned}
\mathcal{A}_n \quad \subseteq \quad & \{w : \|S_n(\mathbf{w})_{0,s_0} - S_n(\mathbf{w})_{0,t_0}\| \leqslant 2\eta\} \\
& \bigcup \{w : \max_{s_0 \leqslant s \leqslant s_1} \|S_n(\mathbf{w})_{0,s} - S_n(\mathbf{w})_{0,s_0}\| > \eta\} \\
& \bigcup \{w : \max_{t_0 \leqslant t \leqslant t_1} \|S_n(\mathbf{w})_{0,t} - S_n(\mathbf{w})_{0,t_0}\| > \eta\}
\end{aligned}
$$

for every $\eta > 0$. Combining with Proposition 3.1 and Proposition 3.2, we have

$$
\begin{aligned}
\mathrm{Cap}_{r,q}(\mathcal{A}_n)^q \quad \leqslant \quad & \frac{C_{r,q,n,d,\tau}}{|t_0 - s_0|^{\lambda_{n,d}/\tau}} \eta^{\frac{|\mathcal{I}_{d,n}|}{\tau} - rq} \\
& + C_{N,q,d}\left(\frac{\Delta^{qN}}{\eta^{2Nq(1+\delta)}}\left(1 + \frac{\Delta^{qN}}{\eta^{2Nq(1+\delta)}}\right)\right. \\
& \left. + \eta^{\frac{2Nq\delta}{n}}\right),
\end{aligned} \tag{3.16}
$$

for any $\tau > 1, N > r, \delta > 0$ and small η.

Now we divide the intervals $[s_0, s_1]$ and $[t_0, t_1]$ into dyadic sub-intervals with length $\Delta/2^l$. Note that any $I \subseteq [s_0, s_1]$ and $J \subseteq [t_0, t_1]$ are separated from each other by distance at least $t_0 - s_1$. Therefore, by applying (3.16) to the dyadic sub-intervals, we obtain that

$$
\begin{aligned}
\mathrm{Cap}_{r,q}(\mathcal{A}_n)^q \quad \leqslant \quad & \frac{C_{r,q,n,d,\tau}}{|t_0 - s_1|^{\lambda_{n,d}/\tau q}} 2^{2l} \eta^{\frac{|\mathcal{I}_{d,n}|}{\tau} - rq} \\
& + C_{N,q,d} 2^{2l}\left(\frac{2^{-Nql}\Delta^{Nq}}{\eta^{2Nq(1+\delta)}}\left(1 + \frac{2^{-Nql}\Delta^{Nq}}{\eta^{2Nq(1+\delta)}}\right)\right. \\
& \left. + \eta^{\frac{2Nq\delta}{n}}\right).
\end{aligned}
$$

Setting $\eta = 2^{-\sigma l}$ with $\sigma > 0$, we arrive at

$$
\begin{aligned}
\mathrm{Cap}_{r,q}(\mathcal{A}_n)^q \quad \leqslant \quad & \frac{C_{r,q,n,d,\tau}}{|t_0 - s_1|^{\lambda_{n,d}/\tau}} 2^{-l\left(\sigma\left(\frac{|\mathcal{I}_{d,n}|}{\tau} - rq\right) - 2\right)} \\
& + C_{N,q,d} 2^{-l(Nq(1-2\sigma(1+\delta))-2)} \\
& \cdot \left(1 + 2^{-Nql(1-2\sigma(1+\delta))}\right).
\end{aligned} \tag{3.17}
$$

To expect that the right hand side goes to zero as $l \to \infty$, we choose the parameters τ, δ and σ such that

$$
\sigma\left(\frac{|\mathcal{I}_{d,n}|}{\tau} - rq\right) > 2
$$

and

$$
1 - 2\sigma(1+\delta) > 0.
$$

19

This is equivalent to

$$\frac{2}{|\mathcal{I}_{d,n}|/\tau - rq} < \sigma < \frac{1}{2(1+\delta)}, \qquad (3.18)$$

provided that the left hand side is positive. As $\tau > 1$ and $\delta > 0$ is arbitrary, when

$$|\mathcal{I}_{d,n}| > rq + 4,$$

a choice of parameters satisfying (3.18) is certainly possible. Therefore, after choosing τ, δ and σ, when N is large we conclude that the right hand side of (3.17) converges to zero as $l \to \infty$. In other words, we have

$$\mathrm{Cap}_{r,q}(\mathcal{A}_n) = 0.$$

On the other hand, if $S_n(\mathbf{w})_{0,s} = S_n(\mathbf{w})_{0,t}$ for some $s < t$, apparently there exist two disjoint dyadic sub-intervals $[s_0, s_1]$ and $[t_0, t_1]$ containing s and t respectively. Therefore,

$$\mathrm{Cap}_{r,q}(\mathcal{O}_n) = 0,$$

which gives the first assertion of Theorem 2.1.

To prove the second assertion, let $\mathcal{O}$ be the event that the Brownian signature path has self-intersection at some $0 \leqslant s < t \leqslant 1$. Then we have

$$\mathcal{O} \subseteq \bigcap_{n \geqslant 1} \mathcal{O}_n.$$

Therefore, $\mathcal{O}$ has zero (r, q)-capacity for every r and q. In other words, $\mathcal{O}$ is a slim set. $\qquad\square$

Based on the deterministic uniqueness result for signature in [2], it is not hard to see that the quasi-sure uniqueness for the signature of Brownian motion (Theorem 2.3) is a direct consequence of Theorem 2.1.

Proof of Theorem 2.3. From [2], Theorem 4.1, we know that the space of signatures for weakly geometric p-rough paths has a canonical real tree structure. In particular, if g is the signature of some weakly geometric p-rough path $\mathbf{X}$, then there exists a unique weakly geometric p-rough path $\widetilde{\mathbf{X}}$ (up to reparametrization) such that its signature is g and its signature path is non-self-intersecting.

In our case, let $\mathcal{N}$ be the slim set outside which every Brownian signature path is non-self-intersecting. Suppose $w, w' \in \mathcal{N}^c$ are two sample paths of Brownian motion with the same signature. It follows that the corresponding rough paths $\mathbf{w}$ and $\mathbf{w}'$ differ by a reparametrization, and hence w are w' differ by a reparametrization. Therefore, quasi-surely every sample path of Brownian motion is uniquely determined by its signature up to reparametrization.

$\qquad\square$

4   Final Remarks

We give a few remarks to conclude our paper.

First of all, from the details of the proof, it is not hard to see that our technique is robust as it only involves the Gaussian nature of Brownian motion and the structure of its covariance function. In particular, its explicit distribution, martingale property and Markov property are not used at all. Therefore, our work extends to any Gaussian rough path under the intrinsic capacities induced by the underlying Gaussian measure over the associated abstract Wiener space, for the cases where the Gaussian rough path $\mathbf{X}$ is well defined quasi-surely and Hörmander's theorem for rough differential equations driven by $\mathbf{X}$ is applicable. A fundamental example where everything works is the fractional Brownian motion B^H with Hurst parameter $H > 1/4$.

On the other hand, one might ask if we could strengthen Theorem 2.1 to the intrinsic dimension of the truncated signature path $S_n(\mathbf{w})_{0,t}$ instead of restricting it to the collection $\mathcal{I}_{d,n}$ of components. Indeed, it is known that (see [7]) $S_n(\mathbf{w})_{0,t}$ satisfies an intrinsic hypoelliptic differential equation

$$dS_n(\mathbf{w})_{0,t} = \sum_{i=1}^{d} U_i(S_n(\mathbf{w})_{0,t}) \circ dw_t^i$$

on the free nilpotent Lie group $G^n(\mathbb{R}^d)$ of order n over $\mathbb{R}^d$. Therefore, it is reasonable to expect that $S_n(\mathbf{w})_{0,t}$ is non-self-intersecting outside a set of zero (r, q)-capacity provided

$$\dim G^n(\mathbb{R}^d) > rq + 4.$$

This will be sharper as $|\mathcal{I}_{d,n}|$ grows with rate n^d while $\dim G^n(\mathbb{R}^d)$ grows with rate d^n/n as $n \to \infty$. However, what is missing is the analysis on the vector fields U_i in order to guarantee a priori estimates on the density which is needed in our proof. This is non-trivial as the vector fields are in fact polynomial of degree n when pulled back to the free nilpotent Lie algebra. It is not clear how to develop a localization method which is consistent with our argument. We do not pursue this direction because unlike the full signature path, the truncated signature path up to a given degree does not have a natural interpretation on the geometric behavior of the Brownian rough path.

However, the case when $n = 2$ is particularly interesting because it is just the Brownian rough path. In this case,

$$\dim G^2(\mathbb{R}^d) = |\mathcal{I}_{d,2}| = \frac{d^2 + d}{2}.$$

Based on the works of Dvoretzky, Erdős and Kakutani [6] and Lyons [15] as we mentioned in the introduction, it is natural to expect that the Brownian rough path has self-intersection with positive probability when $d = 2$ while it is non-self-intersecting outside a set of zero $(1, 2)$-capacity when $d = 3$. Moreover, it is even not unreasonable to expect that the Brownian rough path is non-self-intersecting outside a set of zero $(r, 2)$-capacity if and only if $d^2 + d \geqslant 4r + 8$.

References

[1] H. Boedihardjo and X. Geng, The uniqueness of signature problem in the non-Markov setting, *Stochastic Process. Appl.* 125 (12): 4674–4701, 2015.

[2] H. Boedihardjo, X. Geng, T. Lyons and D. Yang, The Signature of a rough path: uniqueness, arXiv preprint, 2015, to appear in *Adv. Math.*.

[3] H. Boedihardjo, X. Geng and Z. Qian, Quasi-sure existence of Gaussian rough paths and large deviation principles for capacities, arXiv preprint, 2014, to appear in *Osaka J. Math.*.

[4] K. Chen, Iterated integrals and exponential homomorphisms, *Proc. London Math. Soc.* 4 (3): 502–512, 1954.

[5] K.T. Chen, Integration of paths-a faithful representation of paths by non-commutative formal power series, *Trans. Amer. Math. Soc.* 89: 395–407, 1958.

[6] A. Dvoretzky, P. Erdős, S. Kakutani, Double points of paths of Brownian motion in n-space, *Acta Sci. Math. Szeged* 12: 75–81, 1950.

[7] P. Friz and N. Victoir, *Multidimensional stochastic processes as rough paths*, Cambridge University Press, 2010.

[8] M. Fukushima, Basic properties of Brownian motion and a capacity on the Wiener space, *J. Math. Soc. Japan* 36 (1): 161–176, 1984.

[9] X. Geng and Z. Qian, On an inversion theorem for Stratonovich's signatures of multidimensional diffusion paths, *Ann. Inst. Henri Poincaré Probab. Stat.* 52 (1): 429–447, 2016.

[10] 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.

[11] B.M. Hambly and T. Lyons, Stochastic area for Brownian motion on the Sierpinski gasket, Ann. Probab. 26 (1): 132–148, 1998.

[12] Y. Inahama, Quasi-sure existence of Brownian rough paths and a construction of Brownian Pants, *Infin. Dimens. Anal. Quantum Probab. Relat. Top.* 9 (4): 513–528, 2006.

[13] S. Kakutani, On Brownian Motions in n-Space, *Proceedings of the Imperial Academy* 20 (9): 648–652, 1944.

[14] Y. Le Jan and Z. Qian, Stratonovich's signatures of Brownian motion determine Brownian sample paths, *Probab. Theory Relat. Fields* 157: 209–223, 2013.

[15] T. Lyons, The critical dimension at which quasi-every Brownian path is self-avoiding, Adv. in Appl. Probab. 18: 87–99, 1986.

[16] T. Lyons, Differential equations driven by rough signals, *Rev. Mat. Iberoamericana* 14 (2): 215–310, 1998.

[17] T. Lyons and Z. Qian, *System control and rough path*s, Oxford University Press, 2002.

[18] P. Malliavin, *Stochastic analysis*, Springer-Verlag, Berlin, 1997.

[19] I. Shigekawa, *Stochastic analysis,* Translations of Mathematical Monographs, American Mathematical Society, 1998.

[20] H. Sugita, Positive generalized Wiener functions and potential theory over abstract Wiener spaces, *Osaka J. Math.* 25: 665–696, 1988.

[21] M. Takeda, (r, p)-Capacity on the Wiener space and properties of Brwonian motion, *Z. Wahrscheinlichkeitstheorie verw. Gebiete* 68: 149–162, 1984.