Inverting the signature of a path

Terry J. Lyons
University of Oxford

Weijun Xu
University of Warwick

July 1, 2014

Abstract

The main goal of this article is to develop methods to reconstruct a path from its signature. In particular, we give a procedure based on symmetrization that enables one to recover any $C^1$ path from its signature. It is natural to conjecture that this procedure also gives reconstruction to general bounded variation paths.

As a second theme, by using hyperbolic geometry, we also invert the signatures of piecewise linear paths.

1 Introduction

1.1 The signature of a path

Paths are natural objects to describe time evolving systems. The signature of a path (in $\mathbb{R}^d$) is a sequence of iterated integrals along the path, an object taking its value in the tensor algebra over $\mathbb{R}^d$. The graded structure of the tensor algebra allows the signature to capture the noncommutativity of the evolution along the path.

The study of relationships between paths and signatures dates back to Chen in 1950’s. In a series of papers ([3], [4], [5]), he showed that this sequence of iterated integrals determines piecewise smooth paths up to a reparametrization. The main motivation for him to study the signatures, however, is to study the cohomology of path spaces on manifolds ([6]).

Recently, Hambly and Lyons ([8], [9]) introduced the notion of tree-like paths, and proved that paths of bounded variation are completely determined by their signatures up to tree-like equivalence. In addition, within the class of bounded variation paths with the same signature, there is a unique one with minimal length, called the tree-reduced path. It is then natural to ask how one could reconstruct the reduced path from its signature, and this is the main goal of this paper. In fact, despite its independent theoretical interest, an effective reconstruction scheme would also be important in some practical cases (see for example [7] for using the signature in Chinese handwriting recognition and [10] for understanding the effect of the signature in a controlled system).

Before we give precise formulation of our results, we first introduce a few definitions and notations.

Let $\gamma : [0,1] \rightarrow \mathbb{R}^d$ be a path of bounded variation with length $|\gamma| = L$. We say $\gamma$ is parametrized at arc length or at uniform speed if for any $t \in [0,1]$, we have

$$|\gamma|_{[0,t]} = tL,$$
where $|\gamma|_{[0,t]}$ denotes the length of the segment of $\gamma$ in the interval $[0,t]$. Note that the exact value of length depends on the choice of norm on $\mathbb{R}^d$. In fact, we will use the standard Euclidean norm ($\ell^2$) and the somewhat less common $\ell^1$ norm in different contexts below. In these cases, we will have

$$|\gamma|_{\ell^2} = \int_0^1 \left( \sum_{j=1}^d (\dot{\gamma}^j(u))^2 \right)^{\frac{1}{2}} du \quad \text{and} \quad |\gamma|_{\ell^1} = \int_0^1 \sum_{j=1}^d |\dot{\gamma}^j(u)| du,$$

respectively. We will mention which exact norm we will be using in the relevant sections.

For any integer $n$ and sub-interval $[s,t] \subset [0,1]$, we let $X^n_{s,t}$ be the $n$-th level iterated tensor integral

$$X^n_{s,t}(\gamma) := \int_{s<u_1< \cdots <u_n<t} d\gamma_{u_1} \otimes \cdots \otimes d\gamma_{u_n}.$$

We then define the formal series $X_{s,t}(\gamma)$ in the tensor algebra $T(\mathbb{R}^d)$ by

$$X_{s,t}(\gamma) = 1 + \bigoplus_{n=1}^{+\infty} X^n_{s,t}(\gamma).$$

This formal series enjoys the following multiplicative identity

$$X_{s,u}(\gamma) \otimes X_{u,t}(\gamma) = X_{s,t}(\gamma), \quad \forall s < u < t, \quad (1)$$

first proved by K.T.Chen in [4]. The following definition of the signature was introduced in [9].

**Definition 1.1.** If $\gamma : [0,1] \to \mathbb{R}^d$ is a path of bounded variation, then the signature of $\gamma$ is defined by

$$X(\gamma) = X_{0,1}(\gamma).$$

Sometimes it is more convenient to express these tensor integrals in terms of their standard Euclidean coordinates, and we will do so in the rest of the article. If $(e_1, \cdots, e_d)$ is a standard basis of $\mathbb{R}^d$ and $w = e_{i_1} \cdots e_{i_n}$ is a word, then we let

$$C(w) = \int_{0<u_{i_1}< \cdots <u_{i_n}<1} d\gamma^i_{u_{i_1}} \cdots d\gamma^i_{u_{i_n}}$$

denote the coefficient of the word $w$ in the signature, and the signature of $\gamma$ is nothing but a monomial of all words. In fact, we have

$$X(\gamma) = \sum_{n=0}^{+\infty} \sum_{|w|=n} C(w)w,$$

where we have used the convention that $C(w) = 1$ if $w$ is an empty word.

Note that $X(\gamma)$ is a sequence of definite integrals along $\gamma$, and reparametrization of $\gamma$ does not change its signature. The signature $X(\gamma)$ characterizes properties of $\gamma$ as a control. In fact, Hambly and Lyons [9] proved the following theorem in the case of bounded variation paths.
Theorem 1.2 (Hambly, Lyons). Let $\alpha, \beta$ be two bounded variation paths on $\mathbb{R}^d$. Then, $X(\alpha) = X(\beta)$ if and only if $\alpha \ast \beta^{-1}$ is tree-like. Furthermore, within the class of all bounded variation paths with the same signature, there is a unique one with minimal length, called the tree reduced path.

The main purpose of this article is to address the inverse problem: how one could find this reduced path of bounded variation from its signature. Before we proceed to state our main results, we would like to mention some recent progress in proving uniqueness of signatures for some interesting random paths: particularly the works [11] and [1]. However, for the inversion problem, as these random paths are in general only Hölder-$\alpha$ for $\alpha < \frac{1}{2}$, it is not clear at this stage how the methods developed in this article could be modified to reconstruct those paths from their signatures.

1.2 Main results and outline of the paper

We now give a brief outline of the main results in this paper. In Section 2, we solve the inversion problem for axis paths. These are paths whose movements are parallel to Euclidean axes. We obtain the following reconstruction theorem.

Theorem 1.3. Let $X$ be the signature of an axis path $\gamma$. Then, there is a unique longest square free word $w$ with $C(w) \neq 0$. If $w = e_{i_1} \cdots e_{i_n}$, then $\gamma$ has the form

$$\gamma = (r_1 e_{i_1}) \ast \cdots \ast (r_n e_{i_n}),$$

where $r_k = \frac{2C(w_k)}{C(w)}$, and $w_k = e_{i_1} \cdots e_{i_k} e_{i_{k+1}} \cdots e_{i_n}$.

In Section 3, we recover the derivative at the end point of any $C^2$ path from its signature through a limiting process. A precise statement is the following.

Theorem 1.4. Let $\gamma : [0, 1] \to \mathbb{R}^d$ be a $C^2$ path of length $L$ parametrized at uniform speed with $\dot{\gamma}(t) = L\theta(t)$, where $\theta \in \mathbb{S}^{d-1}$ represents the direction of the travelling path. Let $(\eta_{\lambda}(t) \sinh \rho_{\lambda}(t), \cosh \rho_{\lambda}(t))^T$ denote the trajectory of the development of $\gamma_{\lambda}$ on $\mathbb{H}^d$, where $\eta_{\lambda} \in \mathbb{S}^{d-1}$ and $\rho_{\lambda} \in \mathbb{R}^+$. Then, the end point $(\eta_{\lambda}(1), \rho_{\lambda}(1))$ is determined by the signature of $\gamma$, and we have

$$\lim_{\lambda \to +\infty} \lambda L(\eta_{\lambda}(1) - \theta(1)) = -(I + \theta(1)\theta(1)^T)^{-1} \cdot \theta'(1),$$

where $I$ is the $d \times d$ identity matrix.

---

As defined in [9], a path $\gamma : [0, 1] \to \mathbb{R}^d$ is tree-like if there exists a positive continuous function $h$ defined on $[0, 1]$ such that $h(0) = h(1) = 0$, and that

$$|\gamma_t - \gamma_s| \leq h(s) + h(t) - 2 \inf_{u \in [s, t]} h(u)$$

for all $s < t$. Heuristically, one can think of tree-like paths as being a null path as a control; their trajectories are all cancelled out by themselves.
Loosely speaking, the theorem states that the end direction $\theta(1)$ can be obtained as the limit of a sequence of ‘observables’ $\eta_\lambda(1)$, which can be expressed in terms of the signature. This may at first seem surprising as $\theta(1)$ is a very local quantity, while each term in the signature represents some global effect of the path. In fact, it is a combination of many such global terms that gives asymptotically accurate information of this local quantity. As an application, we invert the signature of any piecewise linear path. This is the content of the following theorem.

**Theorem 1.5.** Let $\gamma : [0, 1] \to \mathbb{R}^d$ be a piecewise linear path with at least two pieces, and suppose the length of its last linear piece is $l$. Let $\eta_\lambda$ denote the direction of the end point on the hyperboloid as in the previous theorem. Then we have

$$\lim_{\lambda \to +\infty} \lim_{\tilde{\lambda} \to +\infty} \frac{1}{\lambda} \log |\eta_\lambda - \eta_{\tilde{\lambda}}| = -l.$$ 

This theorem tells how one could recover the length of the last linear piece of a piecewise linear path from its signature. Since we also know the direction of the last linear piece, we can walk back along the opposite direction for a distance of exactly $l$ to cancel this last piece. Thus, applying this procedure repeatedly gives a reconstruction for piecewise linear paths from their signatures.

One might wonder at this stage whether the same strategy could be employed for reconstructing more general smooth paths (say $C^2$). The main difficulty, however, lies in the fact that one needs to rescale the path with a very large factor $\lambda$ in order to obtain accuracy in the approximation of the derivative. As soon as the last part of the path is not linear, walking back towards the opposite derivative would inevitably create an error, which would be further enlarged by the multiplication of $\lambda$, and result in a damage of the accuracy. It turns out that the effect of the latter (enlargement of the error) tends to dominate the former (enhancing accuracy) in the long run, so it is not clear at this stage how one could implement this scheme beyond piecewise linear paths.

In Section 4, however, we will develop a completely different procedure, based on symmetrization of the signatures, that enables one to reconstruct any $C^1$ path from its signature. We give a loose statement of the main result in the following theorem.

**Theorem 1.6.** Let $X$ be the signature of a $C^1$ path $\gamma : [0, 1] \to \mathbb{R}^2$ when parametrized at uniform speed with respect to $\ell^1$ norm. We can construct a piecewise linear path

$$\tilde{\gamma} = \tilde{\gamma}_1 \ast \cdots \ast \tilde{\gamma}_k$$

from the signature $X$. Here, each $\tilde{\gamma}_j$ is a linear piece with the form

$$\tilde{\gamma}_j = \tilde{L}(a_x^{(j)} \rho_j x + a_y^{(j)} (1 - \rho_j) y),$$

and $a_x^{(j)}, a_y^{(j)} \in \{\pm 1\}$, $\rho_j \in [0, 1]$ and $\tilde{L} \in \mathbb{R}^+$ are all determined by $X(\gamma)$. Furthermore, the path $\tilde{\gamma}$ satisfies

$$\sup_{u \in \left[\frac{j-1}{k}, \frac{j}{k}\right]} \left| \tilde{L}(a_x^{(j)} \rho_j, a_y^{(j)} (1 - \rho_j)) - \dot{\gamma}_u \right| < C \epsilon_k$$

uniformly in $j$, where $\epsilon_k \to 0$ as $k \to +\infty$, and the speed of decay depends on the modulus of continuity of $\dot{\gamma}$. 

4
One advantage of this reconstruction scheme is that by performing a certain operation (symmetrization) on the signature, one could 'see' clearly how the path looks like. We will go into more details in Section 4. As a consequence, we also prove that tail signatures of \( C^1 \) paths already determine the path.

It turns out that in the case of monotone paths, the symmetrization procedure developed in Section 4 has a nice probabilistic interpretation, and we give a simple description in Section 5. This probabilistic interpretation also explains how the symmetrization procedure could be significantly simplified and strengthened for monotone paths.

As mentioned before, in this article, we will use two different norms on \( \mathbb{R}^d \) in different situations. In particular, we use the standard Euclidean norm (\( \ell^2 \)) in Section 3, while in Sections 4 and 5 we will be using the \( \ell^1 \) norm.

As usual, throughout the paper, \( c, C, C_k \), etc. will denote constants whose values may change from line to line.

### Acknowledgements

The research of Terry Lyons is supported by EPSRC grant EP/H000100/1 and the European Research Council under the European Union’s Seventh Framework Programme (FP7-IDEAS-ERC) / ERC grant agreement nr. 291244. Terry Lyons acknowledges the support of the Oxford-Man Institute.

## 2 Axis paths

A path \( \gamma \) is a (finite) axis path if it has the form

\[
\gamma = (r_1 e_{i_1}) \ast \cdots \ast (r_n e_{i_n}),
\]

where \( e_{i_k} \)'s are the standard Euclidean basis elements. In other words, it only moves parallel to Euclidean axes, and each piece has finite length. Here, the \( r_k \)'s can be arbitrary nonzero real numbers. If \( r_k \) is negative, it means that the path moves along the direction \( e_{i_k}^{-1} \) for distance \( |r_k| \). If two consecutive pieces have the same directions up to the sign (that is, \( i_k = i_{k+1} \)), then we can combine them together into one single piece and set \( r'_k = r_k + r_{k+1} \). Thus, we can always assume without loss of generality that \( i_k \neq i_{k+1} \) for all \( k = 1, \ldots, n-1 \) so that the path turns a right angle after each piece. Notice that integer lattice paths are special cases of axis paths.

The following notion of square free words has an important role in the characterization of axis paths.

**Definition 2.1.** A word \( w = e_{i_1} \cdots e_{i_n} \) is square free if for any \( k = 1, \cdots, n-1 \), we have \( i_k \neq i_{k+1} \).

If the path \( \gamma \) is of form (2), then its signature is

\[
X(\gamma) = \exp(r_1 e_{i_1}) \otimes \cdots \otimes \exp(r_n e_{i_n}).
\]

Let \( w = e_{i_1} \cdots e_{i_n} \), then \( w \) by assumption is square free with \( |w| = n \) and

\[
C(w) = r_1 \cdots r_n \neq 0.
\]
Moreover, if \( w' \) is any other square free word with \( |w'| \geq n \), then we must have \( C(w') = 0 \) as there is no such term in the expansion of the product of the above exponentials. In other words, for every finite axis path, there is a unique longest square free word \( w \) such that \( C(w) \neq 0 \). Now given the signature \( X \) of some axis path \( \gamma \), and suppose \( w = e_{i_1} \cdots e_{i_n} \) is this unique square free word, then \( \gamma \) has the form

\[
\gamma = (r_1 e_{i_1}) \ast \cdots \ast (r_n e_{i_n}),
\]

and all we need to do is to determine the coefficients \( r_k \)'s. Now for any \( k = 1, \cdots, n \), let

\[
w_k = e_{i_1} \cdots e_{i_k}^2 \cdots e_{i_n},
\]

then by direct computation, we have \( C(w_k) = \frac{1}{2} r_1 \cdots r_k^2 \cdots r_n \), and \( r_k = \frac{2C(w_k)}{C(w)} \).

Thus, we have proved the following theorem.

**Theorem 2.2.** Let \( \gamma \) be a (finite) axis path. Then, there is a unique longest square free word \( w \) with \( C(w) \neq 0 \). Suppose \( w = e_{i_1} \cdots e_{i_k}^2 \cdots e_{i_n} \), and let

\[
w_k = e_{i_1} \cdots e_{i_k}^2 \cdots e_{i_n}.
\]

Then,

\[
\gamma = (r_1 e_{i_1}) \ast \cdots \ast (r_n e_{i_n}),
\]

where \( r_k = \frac{2C(w_k)}{C(w)} \).

**Remark 2.3.** The notion of square free word was also used independently by LeJan and Qian in [11], where they have called them admissible words. The ideas are similar. There, the authors use this notion to identify the grids in \( \mathbb{R}^d \) together with their orders in which the Brownian path has visited, while in our case we use it to identify the directions together with their orders towards which the axis path travels.

As an immediate corollary, we have the following upper bound of the number of terms in the signature needed to reconstruct an axis path.

**Corollary 2.4.** For an axis path with \( n \) pieces, one needs at most \( n + 1 \) levels in the signature to reconstruct the path.

It is also natural to ask that given an integer \( n \), whether there exist two different paths that have the same signatures up to level \( n \). We now answer this question in affirmative by direct construction. Let \( \alpha^0 \) and \( \beta^0 \) be two one-step lattice paths, in \( x \) and \( y \) directions, respectively. Suppose we have now constructed \( \alpha^n \) and \( \beta^n \); we define \( \alpha^{n+1} \) and \( \beta^{n+1} \) by

\[
\alpha^{n+1} = \alpha^n \ast \beta^n, \quad \beta^{n+1} = \beta^n \ast \alpha^n,
\]

then for each \( n \), both \( \alpha^n \) and \( \beta^n \) have \( 2^n \) steps, and they are different in all the steps. We now claim that \( \alpha^n \) and \( \beta^n \) have the same signature up to level \( n \).
Proposition 2.5. For every $n$, we have

$$X^k(\alpha^n) = X^k(\beta^n), \quad \forall k \leq n$$

Proof. This is clearly true for $n = 0$. Suppose the proposition holds true for $m = 0, \ldots, n$, then for $m = n + 1$ and $k \leq n$, we have

$$X^k(\alpha^n * \beta^n) = \sum_{j=0}^{k} X^j(\alpha^n) \otimes X^{k-j}(\beta^n)$$

$$= \sum_{j=0}^{k} X^j(\beta^n) \otimes X^{k-j}(\alpha^n)$$

$$= X^k(\beta^n * \alpha^n).$$

Now, for $k = n + 1$, we have

$$X^{n+1}(\alpha^n * \beta^n) = X^{n+1}(\alpha^n) + X^{n+1}(\beta^n) + \sum_{j=1}^{n} X^j(\alpha^n) \otimes X^{n+1-j}(\beta^n)$$

$$= X^{n+1}(\beta^n * \alpha^n),$$

thus proving the proposition.

It is easy to see that the path $\alpha^n * (\beta^n)^{-1}$ as constructed above has trivial signature in the first $n$ levels, and has length $2^{n+1}$. An interesting question to ask is the following.

Question 2.6. Can one find a nontrivial lattice path with shorter length such that the first $n$ levels in its signature are all zero?

3 The derivative at the end point

In this section, we show how one can approximate the derivative at the end point of a relatively smooth path through a limiting process of its signature. The strategy is to rescale the path by a large factor $\lambda$ and develop it onto the hyperbolic space. As an application, we solve the inversion problem for piecewise linear paths.

3.1 Development onto the hyperbolic space

The idea of developing the rescaled path onto the hyperbolic space was first used in [13] and then further developed in [9]. The method works as follows: when the scale is large, the negative curvature of the hyperbolic space stretches out the path close to a geodesic, and hence the end point will give asymptotically accurate information of the length. Since the end point on the hyperbolic space is determined by the signature of the path, length can then be recovered from the signature.

The main result in this section is that, when using the hyperboloid model, the derivative at the end point of a relatively smooth path can also be recovered accurately from its signature. This result depends on the following further observation:
the rescaled path travels exponentially fast when it is high in the hyperboloid, and thus later directions tend to dominate earlier directions. Hence, when the scale tends to infinity, the direction one observes on the hyperboloid is very close to the real direction at the end point of the path. The main goal of this section is to give a rigorous quantitative characterization of this observation.

We first describe the how to develop the path into a hyperbolic space. The set up and notations mainly follow [9]. Consider the quadratic form on $\mathbb{R}^{d+1}$ defined by

$$I(x, y) = \sum_{j=1}^{d} x_j y_j - x_{d+1} y_{d+1},$$

and the surface

$$\mathbb{H}^d = \{ x \in \mathbb{R}^{d+1} : I(x, x) = -1, x_{d+1} > 0 \}.$$

For any $x \in \mathbb{H}^d$, $I$ is symmetric and positive definite on the tangent space $\{ y : I(y, x) = 0 \}$, so it gives a Riemannian structure on $\mathbb{H}^d$. In fact, $\mathbb{H}^d$ is the standard upper half $d$-dimensional hyperboloid with metric obtained by restricting $I$ to its tangent spaces.

If we let $SO(d)$ denote the group of orientation preserving isometries on $\mathbb{H}^d$, then its Lie algebra $so(d)$ is the set of $(d + 1) \times (d + 1)$ matrices that has the form

$$\begin{pmatrix} A & \beta \\ \beta^T & 0 \end{pmatrix},$$

where $A$ is $d \times d$ antisymmetric matrix and $\beta \in \mathbb{R}^d$. Now, let $\gamma : [0, 1] \rightarrow \mathbb{R}^d$ be a $C^1$ path with Euclidean ($\ell^2$) length $L$ parametrized at uniform speed. Then, we can write its derivative as

$$\dot{\gamma}(t) = L \theta(t),$$

where $\theta$ is a continuous path on $S^{d-1}$. We also define the linear operator $F : \mathbb{R}^d \rightarrow so(d)$ by

$$F : x \mapsto \begin{pmatrix} 0 & x \\ x^T & 0 \end{pmatrix}, \quad x \in \mathbb{R}^d.$$

Then, $F(\gamma(t))$ is a path in $so(d)$, and the linear differential equation of the Cartan development $\Gamma$ of $\gamma$ onto $SO(d)$ is

$$d\Gamma(t) = F(d\gamma(t))\Gamma(t), \quad \Gamma(0) = \text{id}.$$

Then, $\Gamma(t)$ is a path in $SO(d)$. Since the equation for $\Gamma$ is linear, the solution can be expressed by the signature as

$$\Gamma(t) = I + S^1_{0,t}(\gamma) + \cdots + S^n_{0,t}(\gamma) + \cdots,$$

where

$$S^n_{0,t}(\gamma) = \int_{0 < u_1 < \cdots < u_n < t} F(d\gamma(u_n)) \cdots F(d\gamma(u_1)).$$

If $o = (0, \cdots, 0, 1)^T \in \mathbb{R}^{d+1}$, then $\Gamma(t)o$ is a path on the hyperboloid $\mathbb{H}^d$, and we call it the development of $\gamma$ on $\mathbb{H}^d$. 

8
It is standard that every point on $\mathbb{H}^d$ can be uniquely expressed as
\[
\begin{pmatrix}
\eta 
\sinh \rho \\
\cosh \rho
\end{pmatrix}
\]
for some $\eta \in S^{d-1}$ and $\rho \in \mathbb{R}^+$ (see for example [2]), so we write
\[
\Gamma_\lambda(t) = \begin{pmatrix}
\eta_\lambda(t) \sinh \rho_\lambda(t) \\
\cosh \rho_\lambda(t)
\end{pmatrix}
\]
as the development of the rescaled path $\gamma_\lambda = \lambda \gamma$ on the hyperboloid. The following theorem was proved by Hambly and the first author in [9] (Proposition 3.8).

**Theorem 3.1.** Let $\gamma : [0, 1] \to \mathbb{R}^d$ be a $C^2$ path with length $L$ parametrized at uniform speed, and $\gamma_\lambda = \lambda \gamma$. If we represent the trajectory of its development on the hyperboloid by $(\eta_\lambda \sinh \rho_\lambda, \cosh \rho_\lambda)^T$, then the end point satisfies
\[
\lambda L - \frac{C}{\lambda} \leq \rho_\lambda(1) \leq \lambda L,
\]
where $C$ only depends on the path $\gamma$ but not $\lambda$.

This theorem tells that one can recover the length of a path from its signature by looking at the asymptotic behavior of $\rho_\lambda(1)$. In what follows, we will show that $\theta(1)$, the end direction of $\gamma$, can also be recovered from the asymptotic behavior of $\eta_\lambda(1)$. This can be achieved by deriving an explicit system of differential equations for the hyperbolic development $(\eta_\lambda, \rho_\lambda)$. Then, a careful asymptotic analysis of this system for large $\lambda$ will enable us to prove our main result (recovering $\theta(1)$) directly, as well as obtaining the above theorem (recovering $L$) as a consequence. Note that although the intrinsic working mechanism is the negative curvature of hyperbolic space, once the equation for $\Gamma(t)$ is given ((3)) and the trajectory on the hyperboloid is defined (by $\Gamma(t)o$), the rest will all be standard analysis in Euclidean spaces.

We first assume $\lambda = 1$, and give an explicit expression of the end point in terms of the signature. Note that this is exactly the last column of the matrix $\Gamma(1)$. For each $n \geq 0$, let
\[
E_{2n} = \{ w : w = e_1^2 \cdots e_n^2 \}.
\]
Also, for $k = 1, \cdots, d$, let
\[
E_{2n}^{(k)} = \{ w : w = \tilde{w} * e_k, \tilde{w} \in S_{2n} \}.
\]
Using the expression (4), we can easily obtain the following representation of $\Gamma(1)o$ in terms of the signature of $\gamma$.

**Proposition 3.2.** Suppose $\Gamma(1)o = (\eta \sinh \rho, \cosh \rho)^T$, where $\eta = (\eta^{(1)}, \cdots, \eta^{(d)})^T$. For each word $w$, let $C(w)$ denote the coefficient of $w$ in the signature of $\gamma$. Then, we have
\[
\cosh \rho = \sum_{n=0}^{\infty} \sum_{w \in E_{2n}} C(w),
\]
where by convention $C(w) = 0$ if $w$ is empty word. Also, for each $k = 1, \cdots, d$, we have
\[
\eta^{(k)} \sinh \rho = \sum_{n=0}^{\infty} \sum_{w \in E_{2n}^{(k)}} C(w).
\]
We now track the path $\Gamma(t)\gamma$ on $\mathbb{H}^d$. If $\theta(t) \equiv \theta$; that is, $\gamma$ is a straight line with length $L$, then $\Gamma(t)$ can be written explicitly as

$$
\Gamma(t) = \begin{pmatrix}
(cosh(Lt) - 1)\theta^T + I & \sinh(Lt)\theta
\end{pmatrix}
\begin{pmatrix}
\sinh(Lt)\theta^T \\
cosh(Lt)
\end{pmatrix},
$$

where $I$ now denotes the $d \times d$ identity matrix. Now, since the path $\gamma$ near time $t$ can be approximated by the line segment $L\theta(t)\Delta t$, the matrix for that infinitesimal development is

$$
\Gamma(t, \Delta t) = \begin{pmatrix}
(cosh(L\Delta t) - 1)\theta(t)\theta(T) + I & \sinh(L\Delta t)\theta(t)
\end{pmatrix}
\begin{pmatrix}
\sinh(L\Delta t)\theta(t)^T \\
cosh(L\Delta t)
\end{pmatrix}
= \begin{pmatrix}
I & L\theta(t)\Delta t
\end{pmatrix} + \mathcal{O}(\Delta t^2)
$$

Using the relation

$$
\begin{pmatrix}
(\eta(t + \Delta t)\sinh\rho(t + \Delta t) \\
cosh\rho(t + \Delta t)
\end{pmatrix} = \Gamma(t, \Delta t) \begin{pmatrix}
(\eta(t)\sinh\rho(t) \\
cosh\rho(t)
\end{pmatrix} + \mathcal{O}(\Delta^2),
$$

we see that the trajectory $(\eta(t), \rho(t))$ satisfies

$$
\begin{cases}
\eta(t + \Delta t)\sinh\rho(t + \Delta t) - \eta(t)\rho(t) = L\theta(t)\cosh\rho(t)\Delta t + \mathcal{O}(\Delta t^2) \\
cosh\rho(t + \Delta t) - \cosh\rho(t) = L\theta(t)^T\eta(t)\sinh\rho(t)\Delta t + \mathcal{O}(\Delta t^2)
\end{cases},
$$

Now for $\lambda > 0$, replacing $L$ by $\lambda L$ in [5], we deduce that the trajectory $(\eta_\lambda, \rho_\lambda)$ on $\mathbb{H}^d$ satisfies the following system of differential equations

$$
\begin{cases}
\eta_\lambda'(t) = \lambda L \cdot \frac{\cosh\rho_\lambda(t)}{\sinh\rho_\lambda(t)} \cdot (I + \eta_\lambda(t)\theta(t)^T)(\theta(t) - \eta_\lambda(t)) \\
\rho_\lambda'(t) = \lambda L\theta(t)^T\eta_\lambda(t) \\
\eta_\lambda(0) = \theta(0), \quad \rho_\lambda(0) = 0
\end{cases},
$$

where $I$ is the $d$-dimensional identity matrix.

**Remark 3.3.** To express $(\eta_\lambda(1), \rho_\lambda(1))$ in terms of the signature, one simply replaces $C(w)$ in Proposition 3.2 with $\lambda^n C(w)$ for each $|w| = n$.

**Example 3.4.** We now give an example to see how the hyperbolic development works. Let $\gamma = (L_1x) \ast (L_2y)$ be a two dimensional axis path, moving along $x$ direction for distance $L_1$ first, and then $y$ direction for distance $L_2$. Then, the development of the rescaled path $\gamma_\lambda$ can be characterized by the multiplication of the following two matrices

$$
\begin{pmatrix}
1 & 0 & 0 \\
0 & \cosh\lambda L_2 & \sinh\lambda L_2 \\
0 & \sinh\lambda L_2 & \cosh\lambda L_2
\end{pmatrix} \cdot \begin{pmatrix}
\cosh\lambda L_1 & 0 & \sinh\lambda L_1 \\
0 & 1 & 0 \\
\sinh\lambda L_1 & 0 & \cosh\lambda L_1
\end{pmatrix}
$$

Applying the above product to the base point $(0, 0, 1)^T \in \mathbb{R}^3$, we see the end point of $\gamma_\lambda$ on $\mathbb{H}^d$.

$$
\begin{pmatrix}
(\eta_\lambda^{(1)}(\sinh\rho_\lambda) \\
(\eta_\lambda^{(2)}(\sinh\rho_\lambda) \\
(\cosh\rho_\lambda)
\end{pmatrix} = \begin{pmatrix}
\sinh\lambda L_1 & \sinh\lambda L_2 \cosh\lambda L_1 \\
\sinh\lambda L_2 \cosh\lambda L_1 & \cosh\lambda L_1 \cosh\lambda L_2
\end{pmatrix} \cdot
$$
We have three observations relating to the previous example.

1. Since \( \cosh \rho_\lambda = \cosh \lambda L_1 \cosh \lambda L_2, \rho_\lambda \), the hyperbolic distance of the end point to the base point \( o \), satisfies

\[
\rho_\lambda \approx \lambda(L_1 + L_2),
\]

the right hand side being the length of the rescaled path. This corresponds to the Hambly-Lyons theorem on recovering length.

2. By the previous observation, we have \( \eta^{(1)}_\lambda = \frac{\sinh \lambda L_1}{\sinh \rho_\lambda} \approx 0 \), and hence \( \eta_\lambda \), the observed direction on \( \mathbb{H}^d \), satisfies

\[
(\eta^{(1)}_\lambda, \eta^{(2)}_\lambda) \approx (0,1).
\]

This reflects the fact that the direction of the second piece is vertical.

3. This observation is a quantitative version of the previous one. In fact, we have

\[
\eta^{(1)}_\lambda = \frac{\sinh \lambda L_1}{\sinh \rho_\lambda} \approx Ce^{-\lambda L_2}, \quad 1 - \eta^{(2)}_\lambda \approx Ce^{-2\lambda L_2},
\]

and thus

\[
\frac{1}{\lambda} \log |\eta_\lambda - (0,1)^T| \approx -L_2,
\]

the right hand side being the length of the last linear piece.

These observations are not coincidences, and they are consequences of the negative curvature of the hyperbolic space. In fact, when the scale \( \lambda \) is large, the negative curvature will stretch out the path close to a geodesic, and hence the end point will give asymptotically accurate information of the length. This result was first obtained in Hambly and Lyons ([9]). When the path is being 'stretched out', it travels exponentially fast when high on the hyperboloid, and hence later directions tend to dominate earlier directions in an overwhelming way. These explain the first two observations, and it turns out that they hold for general piecewise \( C^2 \) paths. Thus, when taking \( \lambda \) very large, one expects \( \eta_\lambda(1) \) to be close to \( \theta(1) \), and hence one can recover the tangent direction \( \theta(1) \) through the limiting behavior of \( \eta_\lambda(1) \).

The third observation is true for piecewise linear paths. As we shall see later, though the difference \( |\eta_\lambda - \theta| \) is of order \( O(\frac{1}{\lambda}) \) for general paths, it is exponentially small for piecewise linear paths, with \( L_2 \) in the above example replaced by the length of the last linear piece. This estimate, together with the fact that the direction at the end of the path can be approximated arbitrarily closely by its signature sequence, give an inversion theorem for signatures of piecewise linear paths.

In the rest of this section, we will prove the above observations for general paths with certain smoothness by analyzing the asymptotic behavior of \( (\eta_\lambda, \rho_\lambda) \) through the system ([6]). As an application, we also give an exact inversion procedure for piecewise linear paths.

### 3.2 Solving the differential equation

The goal of this section is to prove the following theorem, which gives a quantitative estimate of \( |\eta_\lambda(t) - \theta(t)| \) for large \( \lambda \).
Theorem 3.5. Let $\gamma$ be a $C^2$ path with length $L$ parametrized at unit speed. Then, for any $t > 0$, we have

$$\lim_{\lambda \to +\infty} \lambda L(\eta_\lambda(t) - \theta(t)) = -\left( I + \theta(t)\theta(t)^T \right)^{-1} : \theta'(t),$$

where $I$ is the $d$-dimensional identity matrix.

Proof. By replacing $\lambda L$ with $\lambda$, we can assume without loss of generality that $L = 1$. Let $f_\lambda(t) = \eta_\lambda(t) - \theta(t)$, then $(f_\lambda, \rho_\lambda)$ satisfies

$$\begin{cases}
    f'_\lambda(t) = -\lambda \cdot \frac{\cosh \rho_\lambda(t)}{\sinh \rho_\lambda(t)} \left( I + \theta(t)\theta(t)^T + f_\lambda(t)\theta(t)^T \right) f_\lambda(t) - \theta'(t) \\
    \rho'_\lambda(t) = \lambda (1 + \theta(t)^T f_\lambda(t)) \\
    f_\lambda(0) = 0, \quad \rho_\lambda(0) = 0
\end{cases}, \quad (7)$$

where $f_\lambda \in \mathbb{R}^d$, $\theta \in \mathbb{S}^{d-1}$ and $\rho_\lambda \in \mathbb{R}^+$. and in the equation for $\rho_\lambda$, we have used $\theta^T \theta = 1$. In what follows, we will show that when $\lambda$ is large, (7) is very close to a system of linear equations. The proof consists of four steps.

Step 1.

We claim that there exists $C > 0$ such that

$$\sup_t \| f_\lambda(t) \| < \frac{C}{\lambda}$$

for all large enough $\lambda$, and any $C > \| \theta' \|_{\infty}$ should suffice. In fact, whenever the quantity $| f_\lambda(t) |$ reaches the value $C/\lambda$, its magnitude will be forced to decrease. To see this, we compute

$$\frac{d}{dt} | f_\lambda(t) |^2 = 2 \langle f'_\lambda(t), f_\lambda(t) \rangle.$$

Substituting $f'_\lambda$ with (7), we have

$$\frac{1}{2} \frac{d}{dt} | f_\lambda(t) |^2 = -\lambda \cdot \frac{\cosh \rho_\lambda(t)}{\sinh \rho_\lambda(t)} \left( \langle (I + \theta(t)\theta(t)^T) f_\lambda(t), f_\lambda(t) \rangle \\ + \langle f_\lambda(t)\theta(t)^T f_\lambda(t), f_\lambda(t) \rangle \right) - \langle \theta'(t), f_\lambda(t) \rangle.$$

Since $\theta \theta^T$ is the projection matrix onto $\theta$, the first term in the parentheses on the right hand side above is always positive, and bounded from below by

$$\langle (I + \theta(t)\theta(t)^T) f_\lambda(t), f_\lambda(t) \rangle = | f_\lambda(t) |^2 + | \langle \theta(t), f_\lambda(t) \rangle |^2 \geq | f_\lambda(t) |^2.$$

Also, since $| \theta(t) | \equiv 1$, we estimate the second term by

$$| \langle f_\lambda(t)\theta(t)^T f_\lambda(t), f_\lambda(t) \rangle | \leq | f_\lambda(t) |^3.$$

Finally the last term satisfies $| \langle \theta'(t), f_\lambda(t) \rangle | \leq \| \theta' \|_{\infty} | f_\lambda(t) |$. Note that since $\frac{\cosh \rho_\lambda(t)}{\sinh \rho_\lambda(t)} \geq 1$, we have

$$\frac{1}{2 | f_\lambda(t) |} \cdot \frac{d}{dt} | f_\lambda(t) |^2 \leq -\lambda | f_\lambda(t) | (1 - | f_\lambda(t) |) + \| \theta' \|_{\infty}$$
Now let $C > \|\theta'\|_{\infty}$ be arbitrary. Since $f_\lambda(0) = 0$, it is then clear that if $\lambda$ is large enough and $\lambda |f_\lambda(t)|$ reaches $C$, the right hand side above will be negative and hence $|f_\lambda(t)|$ will be forced to decrease. Thus, we conclude that there exists a $C$ such that for all large $\lambda$, we have

$$\sup_{t \in [0,1]} |f_\lambda(t)| \leq \frac{C}{\lambda}. \quad (8)$$

**Step 2.**

We now show that the projection of $f_\lambda(t)$ onto the direction $\theta(t)$ is of order $O(\frac{1}{\lambda^2})$. Similar as before, we compute

$$\frac{d}{dt} |\langle \theta(t), f_\lambda(t) \rangle|^2 = 2\langle \theta(t), f_\lambda(t) \rangle \left( \langle \theta'(t), f_\lambda(t) \rangle + \langle \theta(t), f'_\lambda(t) \rangle \right).$$

Substituting $f'_\lambda(t)$ by (7), we have

$$\langle \theta(t), f'_\lambda(t) \rangle = -\langle \theta(t), \theta'(t) \rangle - \lambda \cdot \frac{\cosh \rho_\lambda(t)}{\sinh \rho_\lambda(t)} \left( 2\langle \theta(t), f_\lambda(t) \rangle + \langle \theta(t), f_\lambda(t) \rangle^2 \right).$$

Note that $|\theta| \equiv 1$, so $\langle \theta, \theta' \rangle \equiv 0$, and thus we get

$$\frac{1}{2} \frac{d}{dt} |\langle \theta(t), f_\lambda(t) \rangle|^2 = \langle \theta(t), f_\lambda(t) \rangle \left[ -\lambda \frac{\cosh \rho_\lambda(t)}{\sinh \rho_\lambda(t)} \left( 2\langle \theta(t), f_\lambda(t) \rangle + \langle \theta(t), f_\lambda(t) \rangle^2 \right) \right].$$

Since $\sup_t |f_\lambda(t)| < \frac{C}{\lambda}^\frac{1}{2}$ by the first step, we have

$$\sup_{t} |\langle \theta'(t), f_\lambda(t) \rangle| \leq \frac{\hat{C} \|\theta'\|_{\infty}}{\lambda^\frac{1}{2}}.$$  

Thus, in the sum $2\langle \theta(t), f_\lambda(t) \rangle + \langle \theta(t), f_\lambda(t) \rangle^2$, the second term is negligible. Also since $\frac{\cosh \rho_\lambda}{\sinh \rho_\lambda} \geq 1$, if $|\langle \theta(t), f_\lambda(t) \rangle| > \frac{\hat{C} \|\theta'\|_{\infty}}{\lambda^2}$, we will have

$$\lambda \frac{\cosh \rho_\lambda(t)}{\sinh \rho_\lambda(t)} \left( 2\langle \theta(t), f_\lambda(t) \rangle + \langle \theta(t), f_\lambda(t) \rangle^2 \right) > |\langle \theta'(t), f_\lambda(t) \rangle|$$

and the minus sign in front of $\lambda$ will make $\frac{d}{dt} |\langle \theta(t), f_\lambda(t) \rangle|^2$ negative, and hence the quantity $|\langle \theta(t), f_\lambda(t) \rangle|$ will be forced to decrease. Therefore, there exists a $C > 0$ such that

$$\sup_{t \in [0,1]} |\langle \theta(t), f_\lambda(t) \rangle| \leq \frac{C}{\lambda^2} \quad (9)$$

for all large $\lambda$.

**Step 3.**
We now show that for large $\lambda$, $\frac{\cosh \rho_{\lambda}(t)}{\sinh \rho_{\lambda}(t)}$ is close to 1 uniformly. To see this, note that

$$\rho'_{\lambda}(s) = \lambda (1 + \langle \theta(s), f_{\lambda}(s) \rangle).$$

Integrating both sides from 0 to $t$, and employing (9), we have

$$\lambda t - \frac{Ct}{\lambda} \leq \rho_{\lambda}(t) \leq \lambda t$$

for all $t$, where the second inequality follows from the geometry that the geodesic development gives the maximal length $\lambda t$. Now, since

$$\cosh \rho_{\lambda}(t) \sinh \rho_{\lambda}(t) - 1 = 2e^{-2\rho_{\lambda}(t)}1 - e^{-2\rho_{\lambda}(t)}.$$

The singularity at $t = 0$ has size $\frac{1}{\lambda t}$. But since $f_{\lambda}(0) = 0$ and $|f'_{\lambda}|$ is bounded, this singularity can be killed by a multiplication of $f_{\lambda}(t)$ for $t$ near 0. Also, since the numerator decays exponentially in $\lambda t$, we thus have

$$\left(\frac{\cosh \rho_{\lambda}(t)}{\sinh \rho_{\lambda}(t)} - 1\right)|f_{\lambda}(t)| \leq Ce^{-\lambda t},$$

where $C$ is independent of $\lambda$ and $t$.

**Step 4.**

We are now ready to prove the main claim. For convenience, write $P(s) = I + \theta(s)\theta(s)^T$, and we can rewrite (10) as

$$f'_{\lambda}(s) = -\lambda P(s)f_{\lambda}(s) - \theta'(s) + r_{\lambda}(s),$$

where

$$r_{\lambda}(s) = -\lambda \left(\frac{\cosh \rho_{\lambda}(s)}{\sinh \rho_{\lambda}(s)} - 1\right) \left(P(s) + f_{\lambda}(s)\theta(s)^T\right)f_{\lambda}(s) - \lambda f_{\lambda}(s)\theta(s)^T f_{\lambda}(s).$$

Since $|f_{\lambda}(s)\theta(s)^T f_{\lambda}(s)| < \frac{C}{\lambda^2}$ and $\left(\frac{\cosh \rho_{\lambda}(t)}{\sinh \rho_{\lambda}(t)} - 1\right)|f_{\lambda}(t)| \leq Ce^{-\lambda t}$ by the previous three steps, we have

$$\sup_s |r_{\lambda}(s)| < \frac{C}{\lambda^2}.$$

This shows that the differential equation defining $f_{\lambda}(s)$ is 'almost' linear when $\lambda$ is large. We will now show that it is also close to a constant-coefficient equation which we can write down the solution explicitly.

Now fix arbitrary $t > 0$. Since $|\theta'|$ is bounded, there exists a $\kappa > 0$ such that $\forall \epsilon > 0$ and all $s \in [t - \kappa \epsilon, t]$, we have

$$|P(s) - P(t)| < \epsilon.$$

Thus, we have

$$f'_{\lambda}(s) = -\lambda P(t)f_{\lambda}(s) - \theta'(s) + r_{\lambda}(s) + \tilde{r}_{\lambda}(s),$$

where

$$\tilde{r}_{\lambda}(s) = -\lambda \left(P(s) - P(t)\right)f_{\lambda}(s)$$
satisfies $|\tilde{r}_\lambda(s)| < C\epsilon$ for all $s \in [t - \kappa t, t]$. Now, if $g_\lambda$ solves
\[ g'_\lambda(s) = -\lambda P(t)g_\lambda(s) - \theta'(s), \quad s \in [t - \kappa t, t] \]
with initial condition $g_\lambda(t - \kappa t) = f_\lambda(t - \kappa t)$, then we can easily see that
\[ |f_\lambda(t) - g_\lambda(t)| < C\epsilon(\frac{1}{\lambda^2} + \epsilon). \tag{12} \]
Note that the equation defining $g_\lambda$ is linear with constant coefficient, so the terminal value $g_\lambda(t)$ can be expressed explicitly by
\[ g_\lambda(t) = e^{-\lambda \kappa t P(t)} f_\lambda(t - \kappa t) - e^{-\lambda t P(t)} \int_{t - \kappa t}^{t} e^{\lambda s P(s)} \theta'(s) ds. \]
If we set $\epsilon = \lambda^{-\frac{2}{3}}$ and send $\lambda \to +\infty$, then the first term on the right hand side above decays to 0 exponentially fast. For the second term, by replacing $\theta'(s)$ with the constant $\theta'(t)$, we have
\[ \int_{t - \kappa t}^{t} e^{\lambda s P(s)} \theta'(s) ds = \frac{1}{\lambda} P(t)^{-1} (e^{\lambda t P(t)} - e^{-\lambda (t - \kappa t) P(t)}) \theta'(t). \]
Since $P(t)$ is symmetric and positive definite with eigenvalues 2 and 1, so it has a uniformly bounded inverse, and the error of the above replacement is bounded by
\[ \left| e^{-\lambda t P(t)} \int_{t - \kappa t}^{t} e^{\lambda s P(s)} (\theta'(s) - \theta'(t)) ds \right| \leq \frac{C}{\lambda} \delta(\epsilon) \left( 1 - e^{-\lambda \kappa P(t)} \right), \]
where
\[ \delta(\epsilon) = \sup_{s \in [t - \kappa t, t]} \left| \theta'(s) - \theta'(t) \right| \to 0 \]
as $\epsilon \to 0$. Also recall that by setting $\epsilon = \lambda^{-\frac{2}{3}}$, we have
\[ |e^{-\lambda \kappa P(t)} f_\lambda(t - \kappa t)| < e^{-c\lambda^{1/3}}. \]
Thus, by sending $\lambda \to +\infty$ we have
\[ \lim_{\lambda \to +\infty} \lambda g_\lambda(t) = -P(t)^{-1} \theta'(t). \]
Finally, under the same assumption $\epsilon = \lambda^{-\frac{2}{3}}$, we have $|f_\lambda(t) - g_\lambda(t)| < C\lambda^{-\frac{4}{3}}$. Replacing $g_\lambda$ with $f_\lambda$ will give us
\[ \lim_{\lambda \to +\infty} \lambda f_\lambda(t) = -\left( I + \theta(t) \theta'(t)' \right)^{-1} \theta'(t). \]
Thus we have finished the proof.

**Remark 3.6.** We assume $C^2$ regularity in order to get the explicit error $|\eta_\lambda - \theta| = O(\frac{1}{\lambda})$, but the convergence $\eta_\lambda(t) \to \theta(t)$ itself does not require $C^2$. In fact, it holds as long as $\gamma$ is piecewise $C^1$, and the jump of the derivative is less than $\pi$. The same is true for Theorem 3.7.

**Remark 3.7.** We should note that Step 2 in the proof above is not necessary in establishing (11). Without that step, the term on the right hand side of (12) involving $\lambda$ will become $\frac{1}{\lambda^2}$, which would not affect the conclusion of the theorem. However, the estimate in Step 2 gives a sharp estimate (10), and this verifies in the case of $C^2$ paths Theorem 3.1 which recovers the length of a path from its signature in an accurate way.
3.3 Higher order derivatives

From now on, we assume our path $\gamma$ has unit length. The previous section shows that we can recover the derivative at the end point of $\gamma$ through a limiting process by

$$\theta(1) = \lim_{\lambda \to +\infty} \eta_\lambda(1),$$

where the right hand side can be 'observed' on the hyperboloid for each $\lambda$. By Theorem 3.5, we have

$$\theta'(1) = \lim_{\lambda \to +\infty} \lambda(I + \theta(1)\theta(1)^T)(\theta(1) - \eta_\lambda(1))).$$

Since $\theta(1)$ is now known, we can also get the value of $\theta'(1)$ through another limiting process. This suggests that we can actually recover higher order derivatives at time $t = 1$ provided $\gamma$ is sufficiently smooth.

We first give a heuristic argument to see how it works. Since for any $t > 0$ the difference $|\cosh \rho \lambda(t) - 1|$ is exponentially small in $\lambda t$, we can replace the term $\text{cosh} \rho \lambda(t)$ by 1 and rewrite the equation for $f_\lambda$ as

$$f'_\lambda(t) = -\lambda P(t)f_\lambda(t) - \lambda f_\lambda(t)\theta(t)\theta(t)^T f_\lambda(t) - \theta'(t), \quad (13)$$

where $P(t) = I + \theta(t)\theta(t)^T$. Let us also assume for a moment that for any $t > 0$, $f_\lambda(t)$ can be expanded around $\lambda = +\infty$ by

$$f_\lambda(t) = \frac{A_1(t)}{\lambda} + \cdots + \frac{A_n(t)}{\lambda^n} + \cdots.$$

Under suitable regularity conditions of $\gamma$, we can differentiate the above series termwise to get

$$f'_\lambda(t) = \frac{A'_1(t)}{\lambda} + \cdots + \frac{A'_n(t)}{\lambda^n} + \cdots.$$

Now, substituting the expansions of $f_\lambda$ and $f'_\lambda$ into (13), and comparing coefficients of $\frac{1}{\lambda^n}$ on both sides, we get

$$A_{n+1}(t) = -P(t)^{-1}A'_n(t) + \sum_{j=1}^{n} A_j(t)A_{n+1-j}(t) \quad (14)$$

for $n = 1, 2, \cdots$. It is clear that if $\theta \in C^k$ (or $\gamma \in C^{k+1}$), then $A_n$ can be defined as in the above recursive relation up to $n = k$ with $A_n \in C^{k-n}$. In fact, we have the following theorem.

**Theorem 3.8.** Let $\gamma$ be a $C^{k+1}$ path with length 1 parametrized at unit speed, and for $t > 0$, let $A_n(t)$ be defined as in the recursive relation (14) for $n = 1, \cdots, k$. Then, we have

$$\lim_{\lambda \to +\infty} \lambda^{n+1} \left(f_\lambda(t) - \sum_{j=1}^{n} \frac{A_j(t)}{\lambda^j}\right) = A_{n+1}(t)$$

for each $n = 0, 1, \cdots, k - 1$. 

16
Proof. We prove by induction. Since $\lambda f_\lambda(t) \to -P(t)^{-1}\theta'(t)$, the theorem clearly holds for $n = 0$. Suppose it holds up to $n - 1$. Let

$$g_\lambda(t) = \lambda^n \left( f_\lambda(t) - \sum_{j=1}^{n} \frac{A_j(t)}{\lambda^j} \right),$$

and we need to show that $\lambda g_\lambda(t) \to A_{n+1}(t)$. In fact, we have

$$f_\lambda(t) = \frac{g_\lambda(t)}{\lambda^n} + \sum_{j=1}^{n} \frac{A_j(t)}{\lambda^j}, \quad f'_\lambda(t) = \frac{g'_\lambda(t)}{\lambda^n} - \sum_{j=1}^{n} \frac{A'_j(t)}{\lambda^j}.$$

Substituting them into equation (13) gives a system of differential equations for $g_\lambda$. The recursive relation (14) for $A_j$’s up to $j = n$ suggests that the coefficients in that system up to level $\frac{1}{\lambda^n}$ are all cancelled. Then, multiplying both sides by $\lambda^n$ and rearranging terms, we simplify the equations to

$$g'_\lambda(t) = -\lambda P(t)g_\lambda(t) - \left( A'_n(t) + \sum_{j=1}^{n} A_j(t)A_{n+1-j}(t) \right)$$

$$- g_\lambda(t)\theta(t)t^TA_1(t) - A_1(t)\theta(t)t^Tg_\lambda(t) + r_\lambda(t),$$

where $\sup_t |r_\lambda(t)| < \frac{C}{\lambda^3}$ and we have used the fact that $|g_\lambda|$ is uniformly bounded, as implied by the induction hypothesis (in fact we have $g_\lambda \to 0$). Thus, using exactly the same analysis as in the proof of Theorem 3.5 one can show that

$$\lim_{\lambda \to +\infty} \lambda g_\lambda(t) = -P(t)^{-1}\left( A'_n(t) + \sum_{j=1}^{n} A_j(t)A_{n+1-j}(t) \right).$$

Note that

$$A_{n+1}(t) = -P(t)^{-1}\left( A'_n(t) + \sum_{j=1}^{n} A_j(t)A_{n+1-j}(t) \right)$$

has one less degree of regularity than $A_n$. We have thus proved the theorem. \hfill \Box

We have the following immediate consequence.

Corollary 3.9. Let $\gamma$ be a $C^{k+1}$ path parametrized at unit speed. Then, all the $(k+1)$ derivatives of $\gamma$ can be recovered from its signature sequence.

Proof. We can assume $k \geq 1$. For each $\lambda$, $\eta_\lambda(1)$ is an observable from the signature of $\gamma$. By Theorem 3.8, we have the expansion

$$\eta_\lambda(1) = A_0 + \frac{A_1}{\lambda} + \cdots + \frac{A_k}{\lambda^k} + o(\lambda^{-k}),$$

where $A_j = A_j(1)$ can be determined recursively by

$$A_0 = \lim_{\lambda \to +\infty} \eta_\lambda(1) = \theta(1)$$

We can replace $\frac{\cosh \rho}{\sinh \rho}$ with 1 for simplicity as their difference is exponentially small in $\lambda t$.

This is because all terms up to level $\frac{1}{\lambda^n}$ are completely cancelled except the major linear part $\lambda P(t)g_\lambda$. The remaining terms are of order at most $\frac{1}{\lambda^{n+1}}$, with $r_\lambda$ coming from the terms of order at most $\frac{1}{\lambda^{n+1}}$, and a multiplication of $\lambda^n$ implies it has order $\frac{1}{\lambda}$. 

17
and

\[ A_j = \lim_{\lambda \to +\infty} \lambda^j \left( \eta_\lambda(1) - \sum_{i=0}^{j-1} \frac{A_i}{\lambda^i} \right). \]

We have shown that \( \theta(1) = A_0 \). Suppose \( \theta^{(n)}(1) \) is a polynomial of \( A_0, \cdots, A_n, A_0'(1), \cdots, A_n'(1) \). But the recursive relation (14) implies that each \( A_j'(1) \) can be expressed as a polynomial of \( A_0, \cdots, A_j+1, \) so \( \theta^{(n+1)}(1) \) can be expressed as a polynomial of \( A_0, \cdots, A_n+1, \) and induction implies this holds up to \( \theta^{(k)} \). Thus, all \( k \) derivatives of \( \theta \) (and hence \( (k+1) \) derivatives of \( \gamma \)) can be expressed in terms of the signature sequence of \( \gamma \).

3.4 Inversion for piecewise linear paths

We now apply the results in the previous sections to invert the signatures for piecewise linear paths. Note that for such paths, the direction of the last piece can be recovered by the procedure in previous sections. If we also know the length of the last piece, then we can walk along the opposite direction for exactly the same distance so that this last piece is completely cancelled. We can thus remove the signature of the last linear piece from the whole signature, and repeat the same procedure for the remaining linear pieces so that the signature of the whole path can be inverted exactly. Thus, the key for inversion (for piecewise linear paths) is whether one can know the length of the last piece from the signature (or where one should stop when walking back). In fact, this is the third observation in Example 3.4. More generally, we will prove in this section that

\[ Ce^{-\lambda l} < |\eta_\lambda(1) - \theta(1)| < Ce^{-\lambda l} \]

if the length of the last piece is \( l \). Thus, one can recover that length from the asymptotic behavior of \( |\eta_\lambda(1) - \theta(1)| \). We have the following theorem.

**Theorem 3.10.** Let \( \gamma : [0, 1] \to \mathbb{R}^d \) be a piecewise linear path with at least two pieces, and suppose the length of its last linear piece is \( l \). For each \( \lambda > 0 \), let

\[ \left( \eta_\lambda \sinh \rho_\lambda \right. \left. \cosh \rho_\lambda \right) \]

denote the end point of the development of the rescaled path \( \gamma_\lambda = \lambda \gamma \) on the hyperboloid. Then, we have

\[ \lim_{\lambda \to +\infty} \lim_{\lambda \to +\infty} \frac{1}{\lambda} \log |\eta_{\lambda} - \eta_{\tilde{\lambda}}| = -l. \]

**Proof.** Since \( \eta_{\tilde{\lambda}} \to \theta(1) \), by sending \( \tilde{\lambda} \to +\infty \) first, it suffices to prove that

\[ \lim_{\lambda \to +\infty} \frac{1}{\lambda} \log |\eta_\lambda - \theta(1)| = -l. \]

Also, by symmetry of the hyperboloid, we can assume without loss of generality that \( \theta(1) = (1, 0, \cdots, 0)^T \), that is, the last piece is towards the direction \( e_1 \). Now we decompose \( \gamma \) into

\[ \gamma = \alpha * \beta, \]
where $\beta = le_1$ is the last linear piece of $\gamma$, and $\alpha$ is all the rest. Suppose the end point of the rescaled path $\alpha\lambda$ on the hyperboloid is

$$\begin{pmatrix} \zeta_\lambda \sinh \phi_\lambda \\ \cosh \phi_\lambda \end{pmatrix}.$$ 

Since $\beta = le_1$, the matrix of the development of $\beta_\lambda$ is given by

$$\begin{pmatrix} \cosh \lambda l & \sinh \lambda l \\ 1 & \cdots \\ \sinh \lambda l & \cosh \lambda l \end{pmatrix}.$$ 

Applying this matrix to the end point of $\alpha\lambda$, we then obtain the end point of $\gamma_\lambda$ on $\mathbb{H}^d$ to be

$$\begin{pmatrix} \eta_\lambda^{(1)} \sinh \rho_\lambda \\ \eta_\lambda^{(2)} \sinh \rho_\lambda \\ \vdots \\ \eta_\lambda^{(d)} \sinh \rho_\lambda \\ \cosh \rho_\lambda \end{pmatrix} = \begin{pmatrix} \zeta_\lambda^{(1)} \cosh \lambda l \sinh \phi_\lambda + \sinh \lambda l \cosh \phi_\lambda \\ \zeta_\lambda^{(2)} \sinh \phi_\lambda \\ \vdots \\ \zeta_\lambda^{(d)} \sinh \phi_\lambda \\ \zeta_\lambda^{(1)} \sinh \lambda l \sinh \phi_\lambda + \cosh \lambda l \cosh \phi_\lambda \end{pmatrix}.$$ 

We wish to estimate the distance $|\eta_\lambda - \theta(1)|$. Note that we can immediately have a bound for $\phi_\lambda$ and $\rho_\lambda$ in terms of the lengths of $\alpha$ and $\gamma$. In fact, if $\gamma$ has length $L$, then by Theorem 3.1 and Remark 3.6 we have

$$\lambda(L - l) - \frac{C}{\lambda} \leq \phi_\lambda \leq \lambda(L - l), \quad \lambda L - \frac{C}{\lambda} \leq \rho_\lambda \leq \lambda L.$$ 

Thus, there exist $c, C > 0$ such that for all large $\lambda$, we have

$$ce^{-\lambda l} \leq \frac{\sinh \phi_\lambda}{\sinh \rho_\lambda} \leq Ce^{-\lambda l}.$$ 

Applying the above bounds to $(\eta_\lambda^{(1)} \sinh \rho_\lambda, \ldots, \eta_\lambda^{(d)} \sinh \rho_\lambda)$, we get

$$c\left(\sum_{j=2}^{d} |\zeta_\lambda^{(j)}|^2\right)e^{-2\lambda l} \leq \sum_{j=2}^{d} |\eta_\lambda^{(j)}|^2 \leq C\left(\sum_{j=2}^{d} |\zeta_\lambda^{(j)}|^2\right)e^{-2\lambda l}. \quad (15)$$

Since $\zeta_\lambda$ converges to the direction of the last linear piece of $\alpha$, which is different from $e_1$, there exits $\delta > 0$ such that

$$\sum_{j=2}^{d} |\zeta_\lambda^{(j)}|^2 > \delta$$

for all large enough $\lambda$. This is where we have used the assumption that $\gamma$ has at least two pieces. Also, since $1 - |\eta_\lambda^{(1)}|^2 = |\eta_\lambda^{(2)}|^2 + \cdots + |\eta_\lambda^{(d)}|^2$, we have

$$ce^{-4\lambda l} \leq |1 - \eta_\lambda^{(1)}|^2 \leq Ce^{-4\lambda l}. \quad (16)$$

Combining (15) and (16), we have

$$c\left(\sum_{j=2}^{d} |\zeta_\lambda^{(j)}|^2\right)e^{-2\lambda l} \leq |\eta_\lambda - \theta(1)|^2 \leq C\left(\sum_{j=2}^{d} |\zeta_\lambda^{(j)}|^2\right)e^{-2\lambda l}.$$
Taking logarithm on both sides, sending $\lambda \to +\infty$, and using the fact that the quantity $\sum_{j=2}^{d} |\zeta_{\lambda}^{(j)}|^2$ is bounded away from 0 for all large $\lambda$, we get
\[
\lim_{\lambda \to +\infty} \frac{1}{\lambda} \log |\eta_{\lambda} - \theta(1)| = -l,
\]
thus proving the theorem.

Applying this theorem repeatedly to any piecewise linear path $\gamma$ will give an exact inversion procedure.

**Remark 3.11.** In view of the recursive relation (14), it is then not surprising that in the case of piecewise linear paths, the difference $|\eta_{\lambda}(1) - \theta(1)|$ is exponentially small in $\lambda$. This is because for piecewise linear paths, we have $\theta^{(j)}(1) = 0$ for all $j \geq 1$, and hence also $A_{j}(1) = 0$ for all $j \geq 1$.

**Remark 3.12.** The two limits in the above theorem are taken one after another rather than simultaneously. Sending $\tilde{\lambda} \to +\infty$ amounts to recover the direction $\theta(1)$, and after that, we can compute $|\eta_{\lambda} - \theta(1)|$ to identify the length of the last piece. However, since the error $|\eta_{\lambda}(1) - \theta(1)|$ is exponentially small, we also have
\[
\lim_{\lambda \to +\infty} \frac{1}{\lambda} \log |\eta_{\lambda} - \eta_{2\lambda}| = -l.
\]
In fact, there is nothing special about the number 2 here; any number larger than 1 will give the correct limit.

## 4 Symmetrization

In this section, we will develop a symmetrization procedure on the signatures with which one could reconstruct any $C^{1}$ path from its signature. Since there is essentially no difference between two or higher dimensions, for notational simplicity, we will only consider two dimensional paths here. Note that we will use $\ell^{1}$ norm in this and next section. If $z = (z_{1}, \ldots, z_{d})$, we set
\[
|z| = |z|_{\ell^{1}} = \sum_{j=1}^{d} |z_{j}|.
\]
Throughout, $\gamma : [0, 1] \to \mathbb{R}^{2}$ is parametrized at uniform speed with respect to $\ell^{1}$ length, and $\dot{\gamma}$ is continuous. We let $L = |\gamma|_{\ell^{1}}$ so we have
\[
|\dot{\gamma}| = |\dot{x}| + |\dot{y}| = L.
\]
We also set
\[
\delta_{k} = \sup_{|s-t| \leq \frac{1}{k}} |\dot{\gamma}_{s} - \dot{\gamma}_{t}|,
\]
which approximates the modulus of continuity of $\gamma$ for large $k$. In what follows, by merely using the information $X(\gamma)$, we will construct a piecewise linear path

$$\tilde{\gamma} = \tilde{\gamma}_1 \ast \cdots \ast \tilde{\gamma}_k,$$

where each $\tilde{\gamma}_j$ is a linear piece with the form

$$\tilde{\gamma}_j = \tilde{L} \left( a_x^{(j)} \rho_j x + a_y^{(j)} (1 - \rho_j) y \right).$$

Here, $a_x^{(j)}, a_y^{(j)} \in \{ \pm 1 \}$, $\rho_j \in [0, 1]$ and $\tilde{L} > 0$ are determined by $X(\gamma)$ and satisfy

$$\sup_{u \in \Delta_{k-1}} \left| \tilde{L} \left( a_x^{(j)} \rho_j, a_y^{(j)} (1 - \rho_j) \right) - \dot{\gamma}_u \right| < C \epsilon_k,$$

where $\epsilon_k \sim \sqrt{\delta_k}$. It is clear from the formulation that $\rho_j$ and $1 - \rho_j$ should represent the unsigned direction of the increment of each piece, $a_x^{(j)}, a_y^{(j)}$ represents the signs of these directions, and $\tilde{L}$ should approximate the $\ell^1$ length of $\gamma$.

In order to get a rough idea how the symmetrization procedure works, we first recall that in the case for axis paths, the inversion procedure can be roughly summarised as follows:

1. Identify a unique non-zero square free word in a sufficiently high level signature. That word gives the direction of each piece of the axis path.
2. Move one level up to recover the sign and length of each piece.

At first glance, this procedure seems to crucially depend on the very special structures of axis paths, and does not generalize directly to other situations. In particular, the vanishing/non-vanishing property of coefficients of square free words does not carry over to more general cases where the path can move along any direction in $\mathbb{R}^d$. However, fortunately, it turns out that similar results still hold if we replace the strict zero/non-zero criterion by a more robust notion of degeneracy/non-degeneracy. With this new notion of non-degeneracy, we are able to recover the unsigned directions $\rho_j$’s asymptotically.

On the other hand, the graded structure of the signature captures the non-commutative evolution of the path, but this local non-commutativity cannot be recorded in the unsigned directions $\rho_j$’s. More precisely, our theorem recovers the path at larger scales (bigger than $\frac{1}{k}$), while at smaller scales, we neglect all the non-commutative information and average them into a single line segment $\tilde{\gamma}_j$. Thus, in order to get the averaged local directions $\rho_j$’s, it is natural to introduce a quotient relation on the space of words based on the frequencies of different letters while neglecting their orders. The operation of the signatures under this new quotient space is called symmetrization. Before we give a precise description of the symmetrization procedure, we first introduce some notations. As usual, we let $\Delta_{k-1}$ denote the standard simplex

$$\Delta_{k-1} = \{ 0 < u_1 < \cdots < u_{k-1} < 1 \},$$

and use $u$ denote the point $u = (u_1, \ldots, u_{k-1}) \in \Delta_{k-1}$. For each $u \in \Delta_{k-1}$, we let

$$\Delta_{uj} x = x_{uj} - x_{uj-1}, \quad \Delta_{uj} y = y_{uj} - y_{uj-1},$$

and

$$|\Delta_{uj} \gamma| = |\Delta_{uj} x| + |\Delta_{uj} y|,$$
which reflects the $\ell^1$ norm we are working with. Similarly, we denote the standard increments by

$$\Delta_j x = x_{j/k} - x_{(j-1)/k}, \quad \Delta_j y = y_{j/k} - y_{(j-1)/k},$$

and the same for $|\Delta_j \gamma|$.

If $w$ is a word, we let $|w(x)|$ denote the number of letters $x$ in $w$, and $|w(y)|$ denote the number of letters $y$. For any word $w = e_{i_1} \cdots e_{i_{k-1}}$ and multi-index $\ell = \{\ell_1, \cdots, \ell_k\}$ with $0 \leq \ell_j \leq n$, we let $W_k^{2n}(w, \ell)$ be the set of words

$$W_k^{2n}(w, \ell) = \left\{ w' = w_1 \cdots e_{i_1} \cdots e_{i_{k-1}} \cdot w_k : |w_j(x)| = 2\ell_j, |w_j(y)| = 2n - 2\ell_j \right\}.$$

A typical word $w' \in W_k^{2n}(w, \ell)$ where $w = e_{i_1} \cdots e_{i_{k-1}}$ has the following form.

$$\begin{array}{cccccccc}
\ast\ast\ast\ast & e_{i_1} & \ast\ast\ast\ast & e_{i_2} & \cdots & \ast\ast\ast\ast & e_{i_{k-1}} & \ast\ast\ast\ast \\
w_1 & w_2 & \cdots & w_{k-1} & w_k \\
\end{array}$$

Here, each $w_j$ is a subword of length $2n$ with $2\ell_j$ letters $x$ and $2n - 2\ell_j$ letters $y$. The two consecutive subwords (blocks) $w_{j-1}$ and $w_j$ are separated by the letter $e_{i_j}$ from $w$. For example, for $n = 2$ and $k = 1$, we have

$$W_1^4(x, (1, 0)) = \{ xyxyy, yxxyy \}, \quad W_1^4(x, (1, 2)) = \{ xyxxx, yxxx \}$$

and

$$W_1^4(y, (1, 1)) = \{ xyyxy, xyyyx, xyxxy, xyxyx \}.$$  

With this definition, we set

$$S_k^{2n}(w, \ell) := \sum_{w' \in W_k^{2n}(w, \ell)} C(w'),$$

which is the symmetrized signature with $k$ blocks and block size $2n$. It is not hard to check that this quantity can also be expressed by

$$S_k^{2n}(w, \ell) = \int_{\Delta_k-1} P \prod_{j=1}^{k} \int_{\gamma_{i_j}} \int_{\gamma_{i_{j+1}}} P \left( \frac{2n}{2\ell_j} \right) (\Delta_{i_j} x)^{2\ell_j} (\Delta_{i_{j+1}} y)^{2n-2\ell_j} du. \quad (17)$$

In fact, these are the only quantities that we are going to use to recover the unsigned quantities only when it becomes necessary.

**Remark 4.1.** The reason why we have a letter $e_{i_j}$ between every two consecutive symmetrized blocks is to let $S_k^{2n}(w, \ell)$ have a closed form expression as in (17). This is mainly for technical convenience, and we expect results in this section still hold true when the symmetrization is taken without using these $e_{i_j}$’s to separate blocks.

**Remark 4.2.** Note that the symmetrization is taken only over even numbers of $x$’s and $y$’s in each block. This is to avoid cancellations of different signs inside the integration on the right hand side of (17), and is technically convenient.

Finally, for each $j$, we set $\ell_{-j}$ to be

$$\ell_j = (\ell_1, \cdots, \ell_{j-1}, \ell_{j+1}, \ell_k),$$

and we also occasionally write $\ell = (\ell_j, \ell_{-j})$.  

22
### 4.1 Concentration of the symmetrized signatures

We will show in this subsection that the symmetrized signatures $S^2_n(k)(w, \ell)$ of $C^1$ paths enjoy a very nice concentration property, which is the essential property that guarantees our reconstruction procedure to work. Loosely speaking, it states that although each $S^2_n(k)(w, \ell)$ is an iterated integral over the simplex $\Delta_{k-1}$ (as expressed in (17)), when the block size $n$ is large, most of these integrals are actually concentrated on a very small subset of $\Delta_{k-1}$. In fact, the domain of concentration is around the points $u \in \Delta_{k-1}$ such that the product $\prod_{j | \Delta u^j \gamma^j}$ is maximized, and these maximizers cannot be far away from the 'standard' locations $\{k \over k\}^{k-1}$. Before we give a precise statement of this property and show how it guarantees the effectiveness of the symmetrization procedure, we first state a few useful preliminary results.

**Lemma 4.3.** For all large enough $k$ and all $j = 1, \ldots, k$, we have

$$\frac{L - \delta_k}{k} \leq |\Delta_j \gamma| \leq \frac{L}{k}.$$

**Proof.** The inequality $|\Delta_j \gamma| \leq \frac{L}{k}$ follows immediately from the assumption that $\gamma$ is parametrized at uniform speed.

For the lower bound, we let $I_j = [j - 1, j]$. If both $\dot{x}_u$ and $\dot{y}_u$ keep their signs unchanged in the interval $I_j$, then we have

$$\left| \int_{I_j} \dot{x}_u du \right| = \int_{I_j} |\dot{x}_u| du, \quad \left| \int_{I_j} \dot{y}_u du \right| = \int_{I_j} |\dot{y}_u| du,$$

and it follows trivially that $|\Delta_j \gamma| = {\dot{L} \over k}$. If not, then either $\dot{x}$ or $\dot{y}$ is 0 at some point in the interval $I_j$. We suppose $\dot{y}_u = 0$ for some $u \in I_j$, then the continuity of $\dot{\gamma}$ implies that

$$\sup_{u \in I_j} |\dot{y}_u| \leq \delta_k,$$

and thus

$$|\dot{x}_u| \geq L - \delta_k$$

for all $u \in I_j$. In addition, the continuity of $\dot{x}$ also implies that $\dot{x}_u$ does not change its sign in $I_j$. Thus, we have

$$|\Delta_j \gamma| \geq \left| \int_{I_j} \dot{x}_u du \right| = \int_{I_j} |\dot{x}_u| du \geq \frac{L - \delta_k}{k}.$$

The case that $\dot{x}_u = 0$ for some $u \in I_j$ gives exactly the same bound. We have thus proved the lemma.

The previous lemma states that at the standard point $u = \{k \over k\}$, the product $\prod I_j |\Delta_j \gamma|$ is close to its possible maximum. On the other hand, if $u$ is far away from the standard location $\{k \over k\}$, then the product $\prod I_j |\Delta_j \gamma|$ must be small. This is the content of the next lemma.
Lemma 4.4. If \(|u_j - \frac{j}{k}| > \sqrt{\delta_k}\) for some \(i\), then we must have

\[
\prod_{j=1}^{k} \left( \frac{|\Delta_{u_j} \gamma|}{|\Delta_j \gamma|} \right) < (1 - 3\delta_k)^k. \tag{18}
\]

Proof. Since both the numerator and denominator of the left hand side of (18) scale exactly as \(L^{k-1}\), we can assume without loss of generality that \(L = 1\). Suppose

\[u_j - \frac{j}{k} = \epsilon\]

for some \(j\) and some \(\epsilon\), then \(u_j = \frac{j}{k} + \epsilon\), and the sum of all increments before and after the time \(t = u_j\) satisfy

\[
\sum_{i=0}^{j-1} |\Delta_{u_i} \gamma| \leq \frac{j}{k} + \epsilon, \quad \sum_{i=j+1}^{k} |\Delta_{u_i} \gamma| \leq \frac{k-j}{k} - \epsilon. \tag{19}
\]

Note that here we allow \(\epsilon\) to be either positive or negative. By the bound (19), the best possible maximum one can hope for \(\prod_{j} |\Delta_{u_j} \gamma|\) is the case when we have

\[|\Delta_{u_i} \gamma| = \frac{1}{k} + \frac{\epsilon}{j}, \quad \forall i \leq j-1 \quad \text{and} \quad |\Delta_{u_i} \gamma| = \frac{1}{k} - \frac{\epsilon}{k-j}, \quad \forall i \geq j+1.\]

Thus, we have

\[
\prod_{j=1}^{k} |\Delta_{u_j} \gamma| \leq \left( \frac{1}{k} + \frac{\epsilon}{j} \right)^j \cdot \left( \frac{1}{k} - \frac{\epsilon}{k-j} \right)^{k-j}.
\]

Using Lemma 4.3 we get

\[
\prod_{j=1}^{k} \left( \frac{|\Delta_{u_j} \gamma|}{|\Delta_j \gamma|} \right) \leq \left( \frac{(1+pe)^{\frac{1}{p}}(1-qe)^{\frac{1}{q}}}{1-\delta_k} \right)^{k},
\]

where \(p = \frac{k}{j}\) and \(q = \frac{k}{k-j}\) satisfy \(\frac{1}{p} + \frac{1}{q} = 1\). Now, let

\[f(x) = (1+px)^{\frac{1}{p}}(1-qx)^{\frac{1}{q}},\]

we have

\[f(0) = 1, \quad f'(0) = 0, \quad f''(0) = -(p+q) \leq -4.\]

It is then clear that if \(k\) is sufficiently large and \(\epsilon^2 > \delta_k\), we will have

\[f(\epsilon_k) < 1 - 4\delta_k,\]

which in turn implies

\[
\prod_{j=1}^{k} \left( \frac{|\Delta_{u_j} \gamma|}{|\Delta_j \gamma|} \right) < (1 - 3\delta_k)^k < 1.
\]

This finishes the proof of the proposition.
We now let 
\[ \epsilon_k := \sqrt{L \delta_k}, \]
and also let the set \( E_{k-1} \) be
\[ E_{k-1} = \{(u_1, \ldots, u_{k-1}) : |u_j - \frac{j}{k}| < \epsilon_k\}, \]
then it is very easy to show the following.

**Proposition 4.5.** Let \( k \) be a large fixed number. Then for all large \( n \) (depending on \( k \)), we have
\[ \int_{\Delta_{k-1} \cap E_{k-1}^c} \prod_{j=1}^{k} |\Delta u_j \gamma|^n du \geq (1 - C_k e^{-n}) \int_{\Delta_{k-1} \cap E_{k-1}^c} \prod_{j=1}^{k} |\Delta u_j \gamma|^n du. \]

**Proof.** We let \( E_{k-1}^c \) denote the set
\[ E_{k-1}^c = \{v : |v_j - \frac{j}{k}| < \frac{1}{k^2}, j = 1, \ldots, k\}. \]
By Lemma 4.4 and the continuity of \( \dot{\gamma} \), if \( k \) is large enough, we have
\[ \prod_{j=1}^{k} |\Delta u_j \gamma|^n \leq e^{-n} \prod_{j=1}^{k} |\Delta v_j \gamma|^n \]
for all \( u \in \Delta_{k-1} \cap E_{k-1}^c, v \in \Delta_{k-1} \cap E_{k-1}^c \), and all large \( n \). Averaging both sides of the above in their respective domains, we get
\[ \frac{1}{|\Delta_{k-1} \cap E_{k-1}^c|} \int_{\Delta_{k-1} \cap E_{k-1}^c} \prod_{j=1}^{k} |\Delta u_j \gamma|^n du \leq \frac{e^{-n}}{|\Delta_{k-1} \cap E_{k-1}^c|} \int_{\Delta_{k-1} \cap E_{k-1}^c} \prod_{j=1}^{k} |\Delta v_j \gamma|^n dv, \]
which in turn gives
\[ \int_{\Delta_{k-1} \cap E_{k-1}^c} \prod_{j=1}^{k} |\Delta u_j \gamma|^n du \leq C_k e^{-n} \int_{\Delta_{k-1} \cap E_{k-1}^c} \prod_{j=1}^{k} |\Delta u_j \gamma|^n du, \]
where we have enlarged the domain of the integration on the right hand side to \( \Delta_{k-1} \), and the constant \( C_k \) is given by
\[ C_k = \frac{|\Delta_{k-1} \cap E_{k-1}^c|}{|\Delta_{k-1} \cap E_{k-1}^c|}. \]
This immediately gives the proposition. \( \square \)

So far, we have proved the concentration property for the integrals
\[ \int_{\Delta_{k-1}} \prod_{j=1}^{k} |\Delta u_j \gamma|^n du, \]
but these quantities in general cannot be obtained from the signature \( X(\gamma) \). The next theorem shows that this concentration property actually carries to the real symmetrized signatures \( S_{2}^{kn}(w, \ell) \)'s as long as \( \dot{\gamma} \) is continuous.
Theorem 4.6. For every $k$, there exists $c_k > 0$ and $|w^*| = k-1$ such that

$$\sum_{\ell} |S_k^{2n}(w^*, \ell)| \geq c_k \int_{\Delta_{k-1}} \prod_{j=1}^{k} |\Delta_{u_j} \gamma|^{2n} du$$

for all large $n$, and the sum is taken over all multi-indices $\ell = (\ell_1, \cdots, \ell_k)$ whose $k$ components all run over $0, 1, \ldots, n$.

Proof. For any word $|w| = k - 1$, summing over the multi-indices $\ell$ gives

$$\sum_{\ell} |S_k^{2n}(w, \ell)| \geq \left| \int_{\Delta_{k-1}} \prod_{j=1}^{k-1} \gamma_{u_j}^{ij} \prod_{j=1}^{k} \left( \frac{2n}{2\ell_j} \right) (\Delta_{u_j} x)^{2\ell_j} (\Delta_{u_j} y)^{2n-2\ell_j} du \right|,$$

where we have interchanged the sum over $\ell$ and the product over $j$ since different components of $\ell$ are summed up independently. The integrand of the right hand side of (20) can be split into two products: the pointwise derivatives $\gamma_{u_j}^{ij}$ and the increments $\sum_{\ell} \left( \frac{2n}{2\ell_j} \right) (\Delta_{u_j} x)^{2\ell_j} (\Delta_{u_j} y)^{2n-2\ell_j}$. For the latter one, since

$$\sum_{\ell} \left( \frac{2n}{2\ell_j} \right) (\Delta_{u_j} x)^{2\ell_j} (\Delta_{u_j} y)^{2n-2\ell_j} = \frac{1}{2} \left( (\Delta_{u_j} x + \Delta_{u_j} y)^{2n} + (\Delta_{u_j} x - \Delta_{u_j} y)^{2n} \right),$$

which is bounded above by $|\Delta_{u_j} \gamma|^{2n}$ and bounded below by $\frac{1}{2}|\Delta_{u_j} \gamma|^{2n}$, we have

$$\frac{1}{2^k} \prod_{j=1}^{k} |\Delta_{u_j} \gamma|^{2n} \leq \prod_{j=1}^{k} \sum_{\ell} \left( \frac{2n}{2\ell_j} \right) (\Delta_{u_j} x)^{2\ell_j} (\Delta_{u_j} y)^{2n-2\ell_j} \leq \prod_{j=1}^{k} |\Delta_{u_j} \gamma|^{2n}. \quad (21)$$

Now we look at the first part, namely $\prod_{j=1}^{k} \gamma_{u_j}^{ij}$. Since we expect the whole integral on the right hand side of (21) to be concentrated in the domain $E_{k-1}$, we then choose a word $w^* = e_{i_1} * \cdots * e_{i_{k-1}}$ such that

$$|\gamma_{u_j}^{ij}| \geq \frac{L}{3} \quad (22)$$

for all $j$ and all $u \in \Delta_{k-1} \cap E_{k-1}$, and that none of the $\gamma_{u_j}^{ij}$ changes its sign in the domain. The continuity of $\gamma$ ensures that we can always find such a word as long as $k$ is large enough. The main purpose of choosing $w$ in this way is to ensure that the term $\prod_{j} \gamma_{u_j}^{ij}$ does not cause any degeneracy or cancellations in the integration in the domain of concentration $\Delta_{k-1} \cap E_{k-1}$.

We now decompose the right hand side of (20) into integrals over two disjoint domains: $\Delta_{k-1} \cap E_{k-1}$ and $\Delta_{k-1} \cap E_{k-1}^c$. For the first one, since the product $\prod_{j} \gamma_{u_j}^{ij}$ is bounded away from 0 by $(L/3)^{k-1}$ and does not change its sign in $E_{k-1}$, we can move the absolute value into the integral and combine (21) and (22) to get

$$\int_{\Delta_{k-1} \cap E_{k-1}} \prod_{j=1}^{k-1} \gamma_{u_j}^{ij} \prod_{j=1}^{k} \sum_{\ell} \left( \frac{2n}{2\ell_j} \right) (\Delta_{u_j} x)^{2\ell_j} (\Delta_{u_j} y)^{2n-2\ell_j} du \leq \frac{L^{k-1}}{6^{k-1}} \int_{\Delta_{k-1} \cap E_{k-1}} \prod_{j=1}^{k} |\Delta_{u_j} \gamma|^{2n} du.$$
For the second domain $\Delta_{k-1} \cap E_{k-1}^c$, by the second inequality in (21) and Proposition 4.5, the integration over that domain is bounded by

$$C_k e^{-2n} \int_{\Delta_{k-1}} \prod_{j=1}^k |\Delta_{u_j}\gamma|^{2n} du$$

for all large $n$. By Proposition (4.5), this part becomes negligible compared to $\int_{\Delta_{k-1} \cap E_{k-1}^c} \prod_{j=1}^k |\Delta_{u_j}\gamma|^{2n} du$ as $n \to +\infty$. Thus, we get

$$\sum_\ell |S_k^{2n}(w^*, \ell)| \geq c_k \int_{\Delta_{k-1} \cap E_{k-1}^c} \prod_{j=1}^k |\Delta_{u_j}\gamma|^{2n} du - C_k e^{-2n} \int_{\Delta_{k-1}} \prod_{j=1}^k |\Delta_{u_j}\gamma|^{2n} du. \quad (23)$$

Applying Proposition 4.5 for another time, we can enlarge the domain of integration on the right hand side to $\Delta_{k-1}$ at the cost of having a slightly smaller constant $c_k$. We have thus finished the proof of the theorem.

4.2 Reconstructing the path

We are now ready to reconstruct the path from the symmetrized signatures. Recall that the quantities we need to determine are $\rho_j \in [0, 1]$ representing the directions, $a^{(j)}_x, a^{(j)}_y \in \{\pm 1\}$ representing the signs of each piece with a given unsigned direction, and $\tilde{L}$ representing the length.

4.2.1 The unsigned directions

We first recover the unsigned directions $\rho_j$’s of each piece. At this stage, we are only using the quantities $S_k^{2n}(w, \ell)$ which can be obtained from the symmetrization and has an expression as in (17). Since we expect $\rho_j$ to be close to the increment of the $j$-th piece of $\gamma$, it is natural to introduce for each $j$ the unique real number $r_j \in [0, 1]$ such that

$$|\Delta_j x| : |\Delta_j y| = r_j : (1 - r_j).$$

We then have the following theorem.

**Theorem 4.7.** For $k$ large enough and each $j = 1, \cdots, k$, we have

$$\lim_{n \to +\infty} \left( \frac{\sum_w \sum_{|\ell| \geq c_k} \sum_{\ell-j} |S_k^{2n}(w, \ell)|}{\sum_w \sum_{\ell} |S_k^{2n}(w, \ell)|} \right) = 0, \quad (24)$$

where the sum is taken over all words $|w| = k - 1$ and all multi-indicies $\ell = (\ell_1, \cdots, \ell_k)$ within the appropriate range as indicated in the claim.

**Proof.** We assume without loss of generality that $L = 1$. By Theorem 4.6, the denominator $\sum_w \sum_\ell |S_k^{2n}(w, \ell)|$ has a positive proportion of $\int_{\Delta_{k-1} \prod_{j=1}^k |\Delta_{u_j}\gamma|^{2n} du}$ as $n \to +\infty$, so we only need to show the numerator has a vanishing proportion of that quantity.
In fact, since the length \( L = 1 \), we necessarily have \(|x_u|, |y_u| \leq 1\) uniformly in \( u \). Thus, for each word \( w \) and multi-index \( \ell \), we have

\[
|\mathcal{S}_k^{2n}(w, \ell)| \leq \int_{\Delta_{k-1}} \prod_{j=1}^k \frac{2n}{2\ell_j} (\Delta_{u_j}x)^{2\ell_j} (\Delta_{u_j}y)^{2n-2\ell_j} du.
\]

Now for each fixed \( j \), summing over all \( \ell_{-j} \) from 0 to \( n \), and \( \ell_j \) in the region \(|\ell_j/n - r_j| > \epsilon_k\), we have

\[
\sum |\mathcal{S}_k^{2n}(w, \ell)| < C_k \int_{\Delta_{k-1}} \prod_{j=1}^k \frac{2n}{2\ell_j} (\Delta_{u_j}x)^{2\ell_j} (\Delta_{u_j}y)^{2n-2\ell_j} \prod_{i \neq j} |\Delta_{u_i} \gamma|^{2n} du.
\]

Similar as before, we decompose the integral on the right hand side into two disjoint domains: \( \Delta_{k-1} \cap E_{k-1}^c \) and \( \Delta_{k-1} \cap E_{k-1} \), and we need to show that both of them have a vanishing proportion of \( \int_{\Delta_{k-1}} \prod_{j=1}^k |\Delta_{u_j} \gamma|^{2n} du \) as \( n \to +\infty \).

In fact, by Proposition 4.5, the above integral over the region \( \Delta_{k-1} \cap E_{k-1}^c \) is bounded by

\[
C_k e^{-2n} \int_{\Delta_{k-1}} \prod_{j=1}^k |\Delta_{u_j} \gamma|^{2n} du,
\]

which clearly vanishes compared with \( \int_{\Delta_{k-1}} \prod_j |\Delta_{u_j}|^{2n} du \) as \( n \to +\infty \). Then it suffices to bound the integration over the region \( \Delta_{k-1} \cap E_{k-1} \). We write down its expression as

\[
\int_{\Delta_{k-1} \cap E_{k-1}} \sum_{|\ell_j/n - r_j| > \epsilon_k} \frac{2n}{2\ell_j} (\Delta_{u_j}x)^{2\ell_j} (\Delta_{u_j}y)^{2n-2\ell_j} \prod_{i \neq j} |\Delta_{u_i} \gamma|^{2n} du.
\]

If the sum of \( \ell_j \) were over the whole range \( 0 \leq \ell_j \leq n \), then the first part of the integrand on the right hand side of (26) would just be

\[
\sum_{\ell_j=0}^n \frac{2n}{2\ell_j} (\Delta_{u_j}x)^{2\ell_j} (\Delta_{u_j}y)^{2n-2\ell_j} \sim |\Delta_{u_j} \gamma|^{2n},
\]

where we used \( f \sim g \) to denote that \( f \) is uniformly between two constant multiples \( g \). On the other hand, for fixed \( u \), the quantity \( \frac{2n}{2\ell_j} (\Delta_{u_j}x)^{2\ell_j} (\Delta_{u_j}y)^{2n-2\ell_j} \) is maximized near the values \( \ell_j \) such that

\[
\frac{\ell_j}{n-\ell_j} \frac{|\Delta_{u_j} x|}{|\Delta_{u_j} y|} < n^{-\frac{1}{2} + \eta},
\]

and it decays exponentially in \( n \) outside that region. Here \( \eta > 0 \) can be arbitrary. In fact, for fixed \( u \), these ‘probabilities’ approximate the \( \mathcal{N}(2np, 2np(1 - p)) \) distribution with

\[
\frac{p}{1 - p} = \frac{|\Delta_{u_j} x|}{|\Delta_{u_j} y|},
\]

28
and thus the sum of these 'probabilities' will be bounded by $n^\eta e^{-n^2}$ if it is taken over the region outside (28).

Since for any $u \in \Delta_{k-1} \cap E_{k-1}$, we have

$$\left| \frac{\Delta_{u,x}}{\Delta_{u,y}} - r_j \right| < \epsilon_k,$$

then if $|\ell_j - r_j| > \epsilon_k$, we will have

$$\left| \frac{\ell_j}{n} - \frac{\Delta_{u,x}}{\Delta_{u,y}} \right| \gg n^{-\frac{1}{2} + \eta},$$

which is outside the region (28). Thus, we deduce that (26) is bounded by

$$C_k n^\eta e^{-n^2} \int_{\Delta_{k-1}} \prod_{j=1}^k |\Delta_{u_j}|^{2n} du.$$

(29)

Since this holds true for every word $w$, by combining the bounds (25) and (29), we thus obtain the bound for the numerator in (24) as

$$\sum_w \sum_{|\ell_j - r_j| > \epsilon_k} \sum_{\ell_j - r_j} |S^n_k(w, \ell)| < C_k e^{-n} \cdot \int_{\Delta_{k-1}} \prod_{j=1}^k |\Delta_{u_j}|^{2n} du.$$

This is exactly the desired vanishing proportion, and we have thus finished the proof.

The following easy corollary enables one to select the directions $\rho_j$ (up to the sign) for each piece $\tilde{\gamma}_j$.

**Corollary 4.8.** For each $j$, there exists $\rho_j \in [0, 1]$ such that

$$\lim_{n \to +\infty} \left( \sum_w \sum_j \sum_{|\ell_j - r_j| < \epsilon_k} \frac{|S^n_k(w, \ell)|}{\left( \sum_w \sum_{\ell} |S^n_k(w, \ell)| \right)} \right) = 1.$$  

(30)

Moreover, if $\{\rho_j\}$ is any such set that satisfies (30), then we must have

$$|\rho_j - r_j| < 2\epsilon_k$$

for all $j = 1, \cdots, k$.

**Proof.** The existence of $\{\rho_j\}$ follows directly by setting $\rho_j = r_j$ and applying Theorem 4.7. On the other hand, if $|\rho_j - r_j| \geq 2\epsilon_k$ and $|\ell_j/n - \rho_j| < \epsilon$, then we necessarily have

$$|\ell_j/n - r_j| \geq \epsilon_k,$$

and Theorem 4.7 implies that this set of $\{\rho_j\}$ must violate (30). This completes the proof. 

29
Note that although Corollary 4.8 is formulated in the limit of \( n \to +\infty \), in practice, for large \( n \), one could just choose \( \{\rho_j\} \) by the criterion

\[
\left( \sum_w \sum_j \sum_{\ell \neq \rho_j < \epsilon_k} |S_{k}^{2n}(w, \ell)| \right) / \left( \sum_w \sum_{\ell} |S_{k}^{2n}(w, \ell)| \right) \geq \frac{2}{3},
\]

and the choice of the lower bound on the right hand side is of course flexible.

**Remark 4.9.** It is very important to note that Theorem 4.7 is only true for large enough \( k \). The reason is that the integral giving the quantity \( \sum_w \sum_{\ell} |S_{k}^{2n}(w, \ell)| \) is concentrated near the maximizers of \( \prod_j |\Delta_{u_j} \gamma| \). For any fixed \( k \), there might be more than one maximizers giving different unsigned directions. However, for fixed path \( \gamma \) and \( k \) large enough, all maximizers will be close to the standard location \{\( \frac{1}{k}\)\}, so the unsigned directions remain stable. However, in the case of monotone paths, there is only one unique maximizer for each \( k \), so one can actually recover any point on the path by symmetrizing only two blocks in the signature. We will go into more details about monotone paths in the Section 5.

### 4.2.2 The signs

We now turn to the recovery of the sign of the direction of each piece, which requires extra information than the \( S_{k}^{2n}(w, \ell) \)’s. Just as in the case for axis paths, we move one level up in the signatures to identify the signs. More precisely, for each \( i = 1, \cdots, k \), we introduce the quantities

\[
S_{k,i,x}^{2n}(w, \ell) = \int_{\Delta_{k-1}} \prod_{j=1}^{k-1} \hat{u}_{ij} \cdot \left( \frac{2n+1}{2\ell_i+1} \right) (\Delta_{u_i} x)^{2\ell_i+1} (\Delta_{u_i} y)^{2n-2\ell_i} \prod_{j \neq i} \left( \frac{2n}{2\ell_j} \right) (\Delta_{u_j} x)^{2\ell_j} (\Delta_{u_j} y)^{2n-2\ell_j} du,
\]

and

\[
S_{k,i,y}^{2n}(w, \ell) = \int_{\Delta_{k-1}} \prod_{j=1}^{k-1} \hat{u}_{ij} \cdot \left( \frac{2n+1}{2\ell_i} \right) (\Delta_{u_i} x)^{2\ell_i} (\Delta_{u_i} y)^{2n+1-2\ell_i} \prod_{j \neq i} \left( \frac{2n}{2\ell_j} \right) (\Delta_{u_j} x)^{2\ell_j} (\Delta_{u_j} y)^{2n-2\ell_j} du.
\]

In order to use these quantities to determine the signs, we first need to choose a word \( w^* \) so that \( \prod_j \hat{u}_{ij} \) is non-degenerate in the region \( \Delta_{k-1} \cap E_{k-1} \). Let \( \{\rho_j\} \) be the set of unsigned directions chosen by Corollary 4.8 then for each \( j = 1, \cdots, k-1 \) we let

\[
e_{ij} = x, \quad \text{if } \rho_j \geq \frac{1}{2},
\]

\[
e_{ij} = y, \quad \text{if } \rho_j < \frac{1}{2},
\]

and set the word \( w^* \) to be

\[
w^* = e_{i_1} \cdots e_{i_{k-1}}.
\]
In fact, this choice of \( w^* \) necessarily guarantees that
\[
\prod_{j=1}^{k-1} |z_{i,j}^u| \geq \left( \frac{L}{3} \right)^{k-1}
\]
for all \( u \in E_{k-1} \) and that the product does not change its sign in this domain. Note that Theorem 4.6 only gives the existence of such a word \( w^* \) while here we choose it explicitly based on the recovery of unsigned directions. We now determine the signs of the \( i \)-th piece (depending on \( n \)) as follows.

**Definition 4.10.** For the word \( w^* \) given above, we let
\[
a_{x}^{(i)} = 1, \quad \text{if} \quad \frac{\sum_{\ell} S_{k,i,x}^{2n}(w^*, \ell)}{\sum_{\ell} S_{k}^{2n}(w^*, \ell)} \geq 0,
\]
\[
a_{x}^{(i)} = -1, \quad \text{if} \quad \frac{\sum_{\ell} S_{k,i,x}^{2n}(w^*, \ell)}{\sum_{\ell} S_{k}^{2n}(w^*, \ell)} < 0.
\]

The choice for \( a_{y}^{(j)} \) is the same except replacing \( S_{k,i,x}^{2n}(w^*, \ell) \) by \( S_{k,i,y}^{2n}(w^*, \ell) \).

In the above definition, the choices of \( a_{x}^{(j)} \)’s and \( a_{y}^{(j)} \)’s depend on \( n \), the size of each block. It is possible that different values of \( n \) may yield different choices of signs. But it turns out that these choices of the signs remain stable as long as \( n \) is large enough, and they indeed give the correct signs as long as the directions are not close to degenerate. This is the content of the following theorem.

**Theorem 4.11.** If \( \rho_i \geq 5 \epsilon_k \), then
\[
\liminf_{n \to +\infty} \frac{\sum_{\ell} S_{k,i,x}^{2n}(w^*, \ell)}{\sum_{\ell} S_{k}^{2n}(w^*, \ell)} > 0 \quad \text{if} \quad \Delta_i x > 0,
\]
\[
\limsup_{n \to +\infty} \frac{\sum_{\ell} S_{k,i,x}^{2n}(w^*, \ell)}{\sum_{\ell} S_{k}^{2n}(w^*, \ell)} < 0 \quad \text{if} \quad \Delta_i x < 0.
\]

Similarly, if \( \rho_i \leq 1 - 5 \epsilon_k \), then
\[
\liminf_{n \to +\infty} \frac{\sum_{\ell} S_{k,i,y}^{2n}(w^*, \ell)}{\sum_{\ell} S_{k}^{2n}(w^*, \ell)} > 0 \quad \text{if} \quad \Delta_i y > 0,
\]
\[
\limsup_{n \to +\infty} \frac{\sum_{\ell} S_{k,i,y}^{2n}(w^*, \ell)}{\sum_{\ell} S_{k}^{2n}(w^*, \ell)} < 0 \quad \text{if} \quad \Delta_i y < 0.
\]

**Proof.** We only prove the first case when \( \rho_i \geq 5 \epsilon_k \) and \( \Delta_i x > 0 \), and the other three situations are similar. For every \( u \in \Delta_{k-1} \), we let
\[
\mathcal{N}(u) = \frac{1}{2^k} \prod_{j=1}^{k-1} z_{i,j}^{u_j} \cdot \left( (\Delta_{u,x} + \Delta_{u,y})^{2n+1} + (\Delta_{u,x} - \Delta_{u,y})^{2n+1} \right)
\]
\[
\prod_{j \neq i} \left( (\Delta_{u,x} + \Delta_{u,y})^{2n} + (\Delta_{u,x} - \Delta_{u,y})^{2n} \right),
\]
and
\[
\mathcal{D}(u) = \frac{1}{2^k} \prod_{j=1}^{k-1} z_{i,j}^{u_j} \prod_{j=1}^{k} \left( (\Delta_{u,x} + \Delta_{u,y})^{2n} + (\Delta_{u,x} - \Delta_{u,y})^{2n} \right).
\]
Then we can express the numerator and denominator as

\[
\sum_{\ell} S_{k,i,x}^{2n}(w^*, \ell) = \int_{\Delta_{k-1}} N(u)du, \quad \sum_{\ell} S_{k}^{2n}(w^*, \ell) = \int_{\Delta_{k-1}} D(u)du.
\]

It is easy to show that the numerator \( \sum_{\ell} S_{k,i,x}^{2n}(w^*, \ell) \) enjoys all the concentration properties described in the previous subsection, so we can again decompose its domain of integration into \( \Delta_{k-1} \cap E_{k-1} \) and \( \Delta_{k-1} \cap E_{k-1}^c \), where the integral over the second region is negligible to that over the first region for large \( n \). The case for the denominator is the same. So we have

\[
\liminf_{n \to +\infty} \frac{\sum_{\ell} S_{k,i,x}^{2n}(w^*, \ell)}{\sum_{\ell} S_{k}^{2n}(w^*, \ell)} = \liminf_{n \to +\infty} \frac{\int_{\Delta_{k-1} \cap E_{k-1}} N(u)du}{\int_{\Delta_{k-1} \cap E_{k-1}} D(u)du}
\] (31)

and it suffices to study the right hand side for large \( n \). By the assumption \( \rho_i \geq 5\epsilon_k \) and \( \Delta_i x > 0 \), we can deduce that for any \( u \in E_{k-1} \) and \( v_i \in [u_{i-1}, u_i] \), we have

\[ x_{v_i} \geq 2\epsilon_k > 0. \]

The constraint \( |\gamma| \equiv L \) also gives \( |y_{v_i}| \leq L - 2\epsilon_k \). Thus, the intermediate value theorem implies

\[ \Delta_{u,x} \geq 2\epsilon_k(u_i - u_{i-1}), \quad |\Delta_{u,y}| \leq (L - 2\epsilon_k)(u_i - u_{i-1}) \]

as long as \( u \in E_{k-1} \). On the other hand, the ratio of the integrand on the right hand side of (31) can be written as

\[
\frac{N(u)}{D(u)} = \frac{(\Delta_{u,x} + \Delta_{u,y})^{2n+1} + (\Delta_{u,x} - \Delta_{u,y})^{2n+1}}{(\Delta_{u,x} + \Delta_{u,y})^{2n} + (\Delta_{u,x} - \Delta_{u,y})^{2n}}.
\]

which is always positive as implied by the positivity of \( \Delta_{u,x} \). We can also easily deduce a pointwise bound on the ratio by

\[
\frac{|N(u)|}{|\Delta_{u,y}|} \leq 2|\Delta_{u,y}| \left( 1 + \left( \frac{|\Delta_{u,x}| + |\Delta_{u,y}|}{|\Delta_{u,x}| - |\Delta_{u,y}|} \right)^{2n} \right).
\] (32)

If \( |\Delta_{u,y}| \leq \frac{1}{2}\epsilon_k(u_i - u_{i-1}) \), we have

\[
\left| \frac{N(u)}{D(u)} - |\Delta_{u,y}| \right| \leq \epsilon_k(u_i - u_{i-1}).
\]

If \( |\Delta_{u,y}| \geq \frac{1}{2}\epsilon_k(u_i - u_{i-1}) \), then the right hand side of (32) decays exponentially. Using

\[ |\Delta_{u,y}| \geq |\Delta_{u,x}| > 2\epsilon_k(u_i), \]

we deduce that in both cases, we have

\[
\frac{N(u)}{D(u)} \geq \epsilon_k(u_i - u_{i-1})
\] (33)

for all \( u \in \Delta_{k-1} \cap E_{k-1} \) and large \( n \). If \( u_i - u_{i-1} \) is too small, then the product \( \prod |\Delta_{u,y}| \) will be strictly less than \( \left( \frac{L}{\delta_k} \right)^{k} \), and it becomes negligible when raised to the power \( 2n \). Thus, there exists \( \eta_k > 0 \) such that both integrals

\[
\int_{\Delta_{k-1} \cap E_{k-1}} N(u)du \quad \text{and} \quad \int_{\Delta_{k-1} \cap E_{k-1}} D(u)du
\]
are concentrated in the sub-domain where \( u_i - u_{i-1} \geq \eta_k > 0 \). Combining this with (33), we conclude that

\[
\liminf_{n \to +\infty} \frac{\int_{\Delta_{k-1} \cap E_{k-1}} N(u) du}{\int_{\Delta_{k-1} \cap E_{k-1}} D(u) du} \geq \eta_k > 0.
\]

We have thus finished the proof. \( \square \)

4.2.3 Length

We have now obtained for each \( j \) the signed direction \( (a_x^{(j)} \rho_j, a_y^{(j)}(1 - \rho_j)) \), and the only remaining quantity to be determined is \( \tilde{\gamma} \), which is expected to approximate the \( \ell^1 \) length of \( \gamma \). We can achieve this by a simple scaling argument. In fact, by the discussions above, if \( |\gamma|_\ell = L \), we will have

\[
\sup_{a \in \mathbb{R}} \left| \frac{1}{a} \right| L(a_x^{(j)} \rho_j, a_y^{(j)}(1 - \rho_j)) - \tilde{\gamma}_a < C \varepsilon_k
\]

for all \( j = 1, \ldots, k \). In particular, this implies

\[
|X^1(\gamma)| - C \varepsilon_k < L \left( \left| \sum_j a_x^{(j)} \rho_j \right| + \left| \sum_j a_y^{(j)}(1 - \rho_j) \right| \right) < |X^1(\gamma)| + C \varepsilon_k,
\]

(34)

where \( |X^1(\gamma)| = |x_1| + |y_1| \) is the \( \ell^1 \) norm of the increment.

If \( |X^1(\gamma)| > 0 \), then \( \varepsilon_k \) becomes negligible with respect to \( |X^1(\gamma)| \) for large \( k \), and \( \left| \sum_j a_x^{(j)} \rho_j \right| + \left| \sum_j a_y^{(j)}(1 - \rho_j) \right| \) is also guaranteed to be strictly positive and bounded away from 0 uniformly in \( k \). Then, (34) suggests that it is natural to set

\[
\tilde{L} := \frac{|X^1(\gamma)|}{\left| \sum_j a_x^{(j)} \rho_j \right| + \left| \sum_j a_y^{(j)}(1 - \rho_j) \right|},
\]

(35)

and clearly this choice of \( \tilde{L} \) satisfies

\[
\sup_{a \in \mathbb{R}} \left| \frac{1}{a} \right| \tilde{L}(a_x^{(j)} \rho_j, a_y^{(j)}(1 - \rho_j)) - \tilde{\gamma}_a < C \varepsilon_k
\]

In the case \( |X^1(\gamma)| = 0 \), the expression (35) determining \( \tilde{L} \) will have the form of \( \frac{0}{0} \), which causes a problem of the definition. The way we circumvent it is to attach a linear piece of positive length to the end of \( \gamma \). More precisely, we define

\[
Y := X(\gamma) \otimes \exp (a_x^{(k)} \rho_k x + a_y^{(k)}(1 - \rho_k)y).
\]

Then, \( Y \) is the signature of \( \gamma \star \beta \) where

\[
\beta = a_x^{(k)} \rho_k x + a_y^{(k)}(1 - \rho_k)y.
\]

The choices of \( a_x^{(k)}, a_y^{(k)} \) and \( \rho_k \) ensures that \( \beta \) concatenates almost smoothly to the end of \( \gamma \). In particular, it will not create any tree-like pieces. It is clear that

\[
|Y^1| = 1 \neq 0,
\]

and we can apply the previous procedure to the new signature \( Y \) to get a path asymptotically close to \( \gamma \star \beta \). Finally, removing \( \beta \) from that path gives the reconstruction of \( \gamma \). This finishes the choice of length as well as the whole reconstruction procedure.
4.2.4 Summary

We now end this section by summarising the symmetrization procedure in the following theorem.

**Theorem 4.12.** For $k$ large enough and each $j = 1, \ldots, k$, there exists $\rho_j \in [0, 1]$ such that

$$
\lim_{n \to +\infty} \left( \sum_w \sum_j \sum_{|\ell_j| < \epsilon_k} |S_{2n}^{2n}(w, \ell_j)| \right) \left( \sum_w \sum_{\ell} |S_{2n}^{2n}(w, \ell)| \right) = 1,
$$

and we choose the $\rho_j$'s that satisfy the above limit. We then choose the word $w^* = e_i^1 \cdots e_i^{k-1}$ by

$$
e_{i_j} = x, \quad \text{if } \rho_j \geq \frac{1}{2},
$$

$$
e_{i_j} = y, \quad \text{if } \rho_j < \frac{1}{2}.
$$

Also, we determine the signs $a^{(j)}_x, a^{(j)}_y \in \{\pm 1\}$ by

$$
a^{(j)}_x = 1, \quad \text{if } \frac{\sum_{\ell} S_{2n}^{2n}(w^*, \ell)}{\sum_{\ell} S_{2n}^{2n}(w^*, \ell)} \geq 0,
$$

$$
a^{(j)}_x = -1, \quad \text{if } \frac{\sum_{\ell} S_{2n}^{2n}(w^*, \ell)}{\sum_{\ell} S_{2n}^{2n}(w^*, \ell)} < 0.
$$

The choice for $a^{(j)}_y$ is the same except replacing $S_{2n}^{2n}(w^*, \ell)$ by $S_{2n}^{2n}(w^*, \ell)$.

Finally, if $|X^1(\gamma)| > 0$, we set $\tilde{L}$ by

$$
\tilde{L} = \frac{|X^1(\gamma)|}{|\sum_j a^{(j)}_x \rho_j| + |\sum_j a^{(j)}_y (1 - \rho_j)|},
$$

and this choice is guaranteed to make sense (the denominator would be uniformly bounded away from 0). If $|X^1(\gamma)| = 0$, then we right multiply $X(\gamma)$ by the signature

$$
\exp \left( a^{(k)}_x \rho_k x + a^{(k)}_y (1 - \rho_k) y \right),
$$

and determine $\tilde{L}$ by the procedure discussed just above.

In this way, we will obtain a piecewise linear path $\tilde{\gamma} = \tilde{\gamma}_1 \cdots \tilde{\gamma}_k$ such that each linear piece $\tilde{\gamma}_j$ has the form

$$
\tilde{\gamma}_j = \frac{\tilde{L}}{k} (a^{(j)}_x \rho_j x, a^{(j)}_y (1 - \rho_j) y),
$$

and satisfies

$$
\sup_{u \in [L^{1/2}, \epsilon]} \left| \tilde{L} (a^{(j)}_x \rho_j x, a^{(j)}_y (1 - \rho_j) y) - \dot{\gamma}_u \right| < C \epsilon_k
$$

for all large $k$ and all $j = 1, \ldots, k$.

We have the following easy consequence of the reconstruction theorem.
Corollary 4.13. Let \( \alpha, \beta : [0,1] \to \mathbb{R}^d \) be two \( C^1 \) paths parametrized at uniform speed. Then, \( \alpha = \beta \) if and only if there exists \( N > 0 \) such that
\[
X^n(\alpha) = X^n(\beta)
\]
for all \( n \geq N \).

Proof. Since the tail signatures of \( \alpha \) and \( \beta \) are the same, by Theorem 4.12, we can use the high level signatures (levels \( nk + k - 1 \) and \( nk + k \)) to produce a sequence of piecewise linear paths \( \gamma^{(k)} \) which converges in 1-variation to both \( \alpha \) and \( \beta \). Thus, we must have \( \alpha = \beta \) and hence \( X(\alpha) = X(\beta) \). \( \square \)

The above corollary shows that the tail signature of a \( C^1 \) path already determines the path. On the other hand, the information up to level \( nk + k \) already largely determines the path. Moreover, the larger \( n \) and \( k \) are, the finer structures of the path one could recover from its signature. Although this seems straightforward from the reconstruction theorem above, it will nevertheless be interesting to have a quantitative characterization of it.

Remark 4.14. One should also note that the path \( \tilde{\gamma} \) produced from this symmetrization procedure is not unique. In fact, what this procedure produces is not a single path, but instead a measure on piecewise linear paths which converges to the delta measure on the original path \( \gamma \) if \( \gamma \in C^1 \). In practice, one can just choose any piecewise linear path from the area where the measure is concentrated, and this path is guaranteed to be close to \( \gamma \) in 1-variation with explicit error bounds. In fact, this is what we have done in the formulation of Theorem 4.12.

5 Probabilistic interpretation

If the path \( \gamma \) is monotone, then the two bounds in Lemma 4.3 meet, and we obtain a strict equality there. In this case, the symmetrization procedure can be significantly simplified and strengthened: one can recover any point \( \gamma_t \) on the path by symmetrizing just two blocks with sizes \( nt \) and \( n(1-t) \) respectively (see Remark 4.9). Following the same line of argument as in the previous section, it is not hard to show that the error of the recovered point \( \gamma_t \) is of order \( O(n^{-\frac{3}{2}} + \eta) \) for any \( \eta > 0 \). The purpose of this section is to give a probabilistic interpretation of the symmetrization procedure for signatures of monotone paths, and show how the above result could be true with a such a simplified procedure.

Same as before, we only consider two dimensional paths here for simplicity, and the only difference for higher dimensions is notations. Throughout this section, \( \gamma : [0,1] \to \mathbb{R}^2 \) is a continuous path monotonically increasing in each of its directions. That is, we have
\[
\dot{x}_u \geq 0, \quad \dot{y}_u \geq 0, \quad \forall u \in [0,1].
\]
We also assume \( \gamma \) is parametrized at uniform speed with respect to \( \ell^1 \) length so we have
\[
\dot{x}_u + \dot{y}_u = L,
\]
where $L = |\gamma|_1$. We now consider two Poisson processes $X_t$ and $Y_t$, generating the letters $x$ and $y$, whose intensities at time $t$ are $\dot{x}_t$ and $\dot{y}_t$, respectively. We run $X$ and $Y$ on the time interval $[0, 1]$ independently and simultaneously, and let $W_t$ be the point process of the letters generated by $X$ and $Y$ in the order of their arrival time. We also use $W$ to denote the random variable of the outcome $\{W_t\}$ at time $t = 1$.

For example, if the letters $x, y, x, x, y$ are generated at times $0 < u_1 < \cdots < u_5 < 1$, then we have $W = xyxx$. It turns out that there is an interesting relationship between various probabilities relating to $W$ and the signature of $\gamma$. Formally, at an infinitesimal level, the probability that $W = w = e_i_1 \cdots e_i_n$ and that the occurrence times of these letters are $u_1 < \cdots < u_n$ is

$$e^{-L} \cdot \prod_{j=1}^{n} \hat{\gamma}_{i_j} u_j d_u,$$

where we have used $\dot{x}_t + \dot{y}_t = L$ to get the prefactor $e^{-L}$. Integrating the time vector $u$ over the simplex $\Delta_n$, we get

$$\mathcal{P}(W = w) = e^{-L} C(w).$$

One can also sum over all words with length $n$ to get

$$\mathcal{P}(|W| = n) = e^{-L} \cdot \frac{L^n}{n!}.$$

Thus, for $|w| = n$, we have the following conditional probability

$$\mathcal{P}(W = w| |W| = n) = \frac{n!}{L^n} C(w).$$

Using this probabilistic procedure, we now show that we can recover any point $\gamma_t$ on the path with high accuracy through a simplified symmetrization procedure. We give the argument for the middle point $t = \frac{1}{2}$, and other time points can be obtained similarly by symmetrizing blocks whose sizes are of respective proportions to each other.

For each $n$ and each $0 \leq k \leq n$, we define the set of words

$$E_{n,k} := \left\{ w = w_1 \ast w_2 : |w_1(x)| = k, |w_1(y)| = n - k, |w_2| = n \right\}.$$

That is, $E_{n,k}$ consists of words of length $2n$ such that its first $n$ letters include $k$ $x$’s and $n - k$ $y$’s. For words of length $2n$, we symmetrize the first half and adding all possibilities for the first half, we get for each $k$ the probability

$$\mathcal{P}(W \in E_{n,k}| |W| = 2n) = \sum_{w \in E_{n,k}} C(w),$$

and we denote this quantity by $S(n, k)$. Note that these are only quantities we are using to recover the middle point $\gamma_{1/2}$. In fact, we let $r \in [0, 1]$ be such that

$$r : (1 - r) = x_{1/2} : y_{1/2},$$

then we have the following theorem.
Theorem 5.1. Let \( \eta' > 0 \) be arbitrarily small number. Then for any \( \eta < \eta' \), we have

\[
\sum_{|\frac{n}{n} - r| > n^{-\frac{1}{2} + \eta'}} S(n, k) < Cn^\eta e^{-n^2\eta}
\]

for all large \( n \).

This theorem recovers the middle point \( \gamma_{1/2} \) on the path. In fact, by the above theorem, if \( \rho \in [0, 1] \) satisfies

\[
\lim_{n \to +\infty} \sum_{|\frac{n}{n} - \rho| < n^{-\frac{1}{2} + \eta}} S(n, k) = 1,
\]

then we necessarily have \( |\rho - r| < Cn^{-\frac{1}{2} + \eta} \). On the other hand, \( \rho = r \) satisfies the above limit. Thus, finding such a \( \rho \in [0, 1] \) gives the asymptotic increment of the first half of the path. Since length \( L \) is easily recovered by reading \( X^1(\gamma) \), we can thus recover the middle point \( \gamma_{1/2} \). Similar as before, we can fix a large \( n \) and choose \( \rho \) such that

\[
\sum_{|\frac{n}{n} - \rho| < n^{-\frac{1}{2} + \eta}} S(n, k) \geq \frac{2}{3},
\]

and the error \( |\rho - r| \) is then of order \( O(n^{-\frac{1}{2} + \eta}) \).

Before we prove Theorem 5.1, we first give a few useful lemmas.

Lemma 5.2. Let \( X_1, X_2 \) be two independent Poisson random variables with intensities \( \lambda_1 \) and \( \lambda_2 \). Then we have

\[
P(X = m|X + Y = 2n) = \binom{2n}{m} \left( \frac{\lambda_1}{\lambda_1 + \lambda_2} \right)^m \left( \frac{\lambda_2}{\lambda_1 + \lambda_2} \right)^{2n-m}
\]

for all \( n \) and \( m \leq 2n \).

We let \( \tau_n \) to be the random time at which the \( n \)-th letter in \( \mathcal{W} \) is generated. We then have the following.

Lemma 5.3. Let \( \eta > 0 \) be arbitrary. For all large \( n \), we have

\[
P(|\tau_n - \frac{1}{2}| > n^{-\frac{1}{2} + \eta}||\mathcal{W}| = 2n) < Cn^\eta e^{-n^2\eta}.
\]

Proof. We bound the upper half probability \( P(\tau_n - \frac{1}{2} > n^{-\frac{1}{2} + \eta}||\mathcal{W}| = 2n) \), and the bound for the other half is the same. We let \( t_n = n^{-\frac{1}{2} + \eta} \), and let \( X_1, X_2 \) be the number of letters generated in the time intervals \([0, \frac{1}{2} + t_n]\) and \([\frac{1}{2} + t_n, 1]\), respectively. Since \( \gamma \equiv 1 \), \( X_1, X_2 \) are independent Poisson random variables with intensities \( \frac{1}{2} + t_n \) and \( \frac{1}{2} - t_n \), respectively, and \( X_1 + X_2 = |\mathcal{W}| \). Thus, by Lemma 5.2 we have

\[
P(X_1 = m||\mathcal{W}| = 2n) = \binom{2n}{m} \left( \frac{1}{2} + t_n \right)^m \left( \frac{1}{2} - t_n \right)^{2n-m}.
\]
Since \( \tau_n > \frac{1}{2} + t_n \) implies \( X_1 < n \), we have
\[
\mathcal{P}(\tau_n > \frac{1}{2} + t_n | |W| = 2n) \leq \sum_{m<n} \binom{2n}{m} \left( \frac{1}{2} + t_n \right)^m \left( \frac{1}{2} - t_n \right)^{2n-m} 
\leq C n^\eta e^{-n^{2\eta}},
\]
where \( t_n = n^{-\frac{1}{2} + \eta} \).

\[\square\]

**Remark 5.4.** Roughly speaking, the above lemma states that on average, the patterns in the first \( n \) letters in words of length \( 2n \) are largely given by the first half of the path. It is this quantitative characterization that enables one to symmetrize only two rather than a large number of blocks for monotone paths. Also see Remark 4.9 for analytical explanations why this could be false for non-monotone paths.

**Proof of Theorem 5.1.**

We are now ready to prove Theorem 5.1. For simplicity, we use \( \mathcal{P}^{(2n)}(\cdot) \) to denote the conditional probability \( \mathcal{P}(\cdot | |W| = 2n) \). For fixed \( \epsilon \), we let
\[
K^{(\epsilon)}_n = \bigcup_{|\frac{1}{n} - r| > \epsilon} E_{n,k}.
\]
Thus, we have
\[
\sum_{|\frac{1}{n} - r| > \epsilon} S(n,k) = \mathcal{P}^{(2n)}(W \in K^{(\epsilon)}_n; |\tau_n - \frac{1}{2}| \leq t_n) + \mathcal{P}^{(2n)}(W \in K^{(\epsilon)}_n; |\tau_n - \frac{1}{2}| > t_n).
\]
By Lemma 5.3, the second term on the right hand side above is bounded by \( n^\eta e^{-n^{2\eta}} \) uniformly in \( \epsilon \), so we only need to bound the first one with \( \epsilon = n^{-\frac{1}{2} + \eta'} \) for \( \eta' > \eta \).

In fact, if we let \( f(n) \) denote the density of \( \tau_n \), then we have
\[
\mathcal{P}^{(2n)}(W \in K^{(\epsilon)}_n; |\tau_n - \frac{1}{2}| \leq t_n)
= \sum_{|\frac{1}{n} - r| > \epsilon} \int_{|\tau - \frac{1}{2}| \leq t_n} \mathcal{P}^{(2n)}(W \in E_{n,k}; |\tau_n = \tau| f(n)(\tau)d\tau
= \sum_{|\frac{1}{n} - r| > \epsilon} \int_{|\tau - \frac{1}{2}| \leq t_n} \binom{n}{k} \left( \frac{x_r}{x_r + y_r} \right)^k \left( \frac{y_r}{x_r + y_r} \right)^{n-k} f(n)(\tau)d\tau.
\]
Note that for \( |\tau - \frac{1}{2}| \leq t_n \), we have
\[
\left| \frac{x_r}{x_r + y_r} - r \right| < C t_n,
\]
and thus we get
\[
\mathcal{P}^{(2n)}(W \in K^{(\epsilon)}_n; |\tau_n - \frac{1}{2}| \leq t_n) \leq \sum_{|\frac{1}{n} - r| > \epsilon} \binom{n}{k} (r + \delta_n)^k (1 - r - \delta_n)^{n-k},
\]
where $|\delta_n| < C t_n$, and we have not specified its sign. Again, if $\epsilon = n^{-\frac{1}{2} + \eta'} \gg n^{-\frac{1}{2} + \eta}$, then by the binomial approximation to normal, we will have

$$P^{(2n)}(W \in K_n^{(\epsilon)} : |\tau_n - \frac{1}{2}| \leq t_n) < C n^\eta e^{-n^{2\eta}}.$$ 

This finishes the proof of Theorem 5.1.

A note on non-monotone paths.

In the case of general (non-monotone) bounded variation paths, one could also interpret the signatures as follows. In addition to the random word $W$ generated from the Poisson processes governed by $\gamma$, we also introduce a random variable $Z$ such that if

$$W = e_{i_1} \cdots e_{i_n},$$

and that the letters are generated at times $0 < u_1 < \cdots < u_n < 1$, then we define $Z$ to be

$$Z = \prod_{j=1}^{n} \text{sgn}(\dot{\gamma}_{i_j}^j),$$

where $\text{sgn}(x) \in \{\pm 1\}$ denotes the sign of $x$. Then, the signature of $\gamma$ can be related to the random pair $(W, Z)$ by

$$E(Z; W = w) = e^{-L} C(w).$$

It is clear that in the case of monotone paths, we have $Z \equiv 1$ and thus $E(Z; W = w) = P(W = w)$. But for non-monotone paths, $E(Z|W \in E)$ can be 0 (or very close to 0) even if the probability of $W \in E$ is almost 1. In this case, the degeneracy of expectation compensates the effect of dominating probabilities, so the latter is no longer visible from the signatures. This also explains why for non-monotone paths, one has to symmetrize a large number of blocks to get a piecewise linear path with many linear pieces rather than just finitely many blocks to get back individual points on the path.

References

[1] H. Boedihardjo, X. Geng, On the uniqueness of signature problem through a strengthened LeJan-Qian approximation scheme, Arxiv preprint: 1401.6165, 2014.

[2] J. Cannon, W. Floyd, R. Kenyon, W. Parry, Hyperbolic Geometry, Flavors of Geometry, MSRI Publications, Vol.31, 1997.

[3] K.-T. Chen, Iterated integrals and exponential homomorphisms, Proceedings of the London Mathematical Society, Vol.4, No.3 (1954), 502-512.

[4] K.-T. Chen, Integration of paths, geometric invariants and a generalized Baker-Hausdorff formula, Annals of Mathematics, Vol.65, No.1 (1957), 163-178.
[5] K.-T. Chen, Integration of paths - a faithful representation of paths by non-commutative formal power series, *Transactions of the A.M.S.*, Vol.89, No.2 (1958), 395-407.

[6] K.-T. Chen, Iterated path integrals, *Bulletin of the A.M.S.*, Vol.83, No.5 (1977), 831-879.

[7] B. Graham, Sparse arrays of signatures for online character recognition, Arxiv preprint, 2013.

[8] B.M.Hambly, T.J.Lyons, Some notes on trees and paths, unpublished manuscript, available at [http://arxiv.org/abs/0809.1365](http://arxiv.org/abs/0809.1365), 2008.

[9] B.M.Hambly, T.J.Lyons, Uniqueness for the signature of a path of bounded variation and the reduced path group, *Annals of Mathematics*, Vol.171, No.1 (2010), 109-167.

[10] M.Kawski, H.Sussmann, Noncommutative power series and formal Lie-Algebraic techniques in nonlinear control theory, *Operators, Systems and Linear Algebra: Three Decades of Algebraic Systems Theory* (1997), 111-129.

[11] Y.LeJan, Z.Qian, Stratonovich’s signatures of Brownian motion determine Brownian sample paths, preprint, to appear in *Probability Theory and Related Fields*.

[12] T.J.Lyons, Differential equations driven by rough signals, *Rev. Mat. Iberoamericana*, Vol.14, No.2 (1998), 215-310.

[13] T.J.Lyons, N.Sidorova, On the radius of convergence of the logarithmic signature, *Illinois Journal of Mathematics*, Vol.50 (2006), 763-790.

Mathematical and Oxford-Man Institutes, University of Oxford, Woodstock road, Oxford, OX2 6GG, UK.

Email: tlyons@maths.ox.ac.uk

Mathematics Institute, University of Warwick, Coventry, CV4 7AL, UK.

Email: weijun.xu@warwick.ac.uk