Effective algorithms for inverting the signature of a path

Jiawei Chang

St Peter's College

University of Oxford

A thesis submitted for the degree of

Doctor of Philosophy

September 2018

Acknowledgements

I would like to express my deep gratitude to my supervisor Prof. Terry Lyons, who helped me whenever I encountered difficulties, without him this work would not have been possible.
I would like to thank my second supervisor Dr. Hao Ni, who gave me a chance to work with brilliant people and guided me through the collaboration.
Thanks to my collaborators Dr. Weijun Xu and Prof. Nick Duffield, who are wonderful mathematicians, and it was my pleasure working with them.
I would also like to thank Dr. Xi Geng, Jiawei Li, Christina Zou, Sina Nejad, Xingcheng Xu, Prof. David Blecher, and Prof. Coralia Cartis, who have given me valuable advice and guidance.
Finally I would like express my great gratitude to my parents, who have always been my greatest support and strength, and have always believed in me through my hard times. I dedicate this work to them.

Abstract

The signature of a path is an important concept in rough paths theory. It has been proved in the literature that the terms in the signature of a path are bounded above, and it would then be interesting to consider whether a lower bound exists for the signature of a bounded-variation path. It has also been proved that the signature of a bounded-variation path is unique up to some modifications, then a natural question is reconstructing the path from its signature, i.e. *inverting the signature of a path.*

We show the connection between the two questions above, and provide practical methods for signature inversion. First we prove a result about the super-multiplicativity and the decay of the signature of a path with bounded variation, then we describe the method of symmetrisation, which was first introduced by Lyons and Xu [25], and demonstrate explicitly how to invert the signature of a monotone path. Moreover we introduce the method of inverting the signature by insertion, and provide examples using the insertion method to invert the signature of a path. We compare these two methods of signature inversion, and illustrate the differences with computational results.

Motivation

The signature of a path was first studied by K.T. Chen ([11], [12]). It can be understood as a collection of non-commutative iterated integrals, and has always been an interesting and essential topic in rough paths theory.

The signature provides a characteristic description of a path. Chen [10] first showed that the non-commutative iterated integrals of a piecewise regular continuous path give a unqiue representation of the path up to some null modifications. Hambly and Lyons [19] furthered the result and showed that this non-commutative transform is faithful for paths of bounded variation up to tree-like pieces. Such a feature of the signature of a path lies at the heart of the theories related to applications of signatures, which is also a motivation of the development of the theory of rough paths.

The signature of a path can be applied to machine learning for sequential data mining. The signature of a path is invariant under re-parametrisation, which makes the signature a good candidate to record the features of data. For example one can refer to the very helpful primer of the connections of rough paths and data science by Chevyrev and Kormilitzin [13]. Moreover, Lyons, Ni and Levin [21] gave an example that when the underlying path is observed at discrete time steps while the data is sampled at finer scales, the signature describes the data better than simply recording the data stream. A particularly successful application of the signature method in machine learning is due to Graham [18] who won a worldwide competition of recognition of Chinese characters by using signatures as feature sets and state-of-the-art deep learning techniques.

Given the fact that the signature of a path is unique up to tree-like pieces [19], it is an important and natural topic to reconstruct the path from its signature for the completeness of the theory. Lyons and Xu ([24] and [25]) developed theories about inverting the signature of a C^1 path, and one particular idea they used is *sym-*

metrisation, which we will mention in one of the chapters. Geng investigated more complicated cases and developed a method of inverting the signature of a rough path [17]. Pfeffer, Seigal and Sturmfels [27] demonstrated a method of computing the shortest path with a given signature level.

Signature inversion also has practical motivations. Because in some cases the dataset is better summarised by the signature rather than discrete time-stamped data, sometimes signatures are stored instead of the raw data. Moreover, truncated signatures can lead to dimension reduction, as we may meet circumstances where it is cheaper to store the signature. As an example, we know that the signature of a d-dimensional path at level n is of size d^n, if the data is sampled at m time points such that

$$d + d^2 + ... + d^n \leq dm,$$

then it is more economical to store the signature of the path up to level n than to store the sequential data without losing much description about the data. In such cases, we need an algorithm to reconstruct the underlying data from its signature.

However, there has been no practical algorithms about inverting the signature of a path in the literature so far. The main aim of this thesis is therefore to provide effective algorithms for signature inversion for some classes of paths, and hopefully shed light on signature inversion in more complicated cases. Developing such algorithms requires knowledge about the behaviour of the signature of a path, and we will give an lower bound for the decay of the signature of a path of finite length. We will also extend the symmetrisation method developed by Lyons and Xu [25] and apply the method to monotone paths, and introduce a new method of inverting the signature of a path using insertion.

Contents

List of Symbols

$\mathbb{N}$: The set of all non-negative integers.
$\mathbb{R}$: The set of all real numbers.
$C^0(J, E)$: The set of continuous functions from J to E.
V': The dual space of the vector space V.
$\|X\|_{p,J}$: The p-variation of X over the interval J, Definition 1.1.1.
$S^n_J(X)$: The n-th term of the signature of a path X, Definition 1.2.3.
$\mathrm{Hom}(V, A)$: The set of bounded linear operators from V to A.
$\|x\|_{\to A}$: A tensor norm defined on the space where x lives, Definition 1.3.5.
$\mathrm{sym}(u)$: The symmetric part of the tensor u, Definition 1.4.1.
$\|\cdot\|_\pi$: The projective tensor norm, Definition 2.1.4.
$\|\cdot\|_\delta$: The injective tensor norm, Definition 2.1.4.
$\|\cdot\|_{HS}$: The Hilbert-Schmidt norm, Example 2.3.1.
$C_\gamma(\omega)$: The coefficient of the iterated integral of γ corresponding to the word ω, Definition 3.2.1.
$\mathcal{S}^{\mathbf{n}}_k(\boldsymbol{\ell})$: The symmetrised signature, Equation (3.4).
$\bar{S}^m_{s,t}(\gamma)$: the m-th term of the normalised signature of γ over $[s, t]$ where γ is parametrised at unit speed, Equation (4.3).
$\bar{S}_m$: The m-th term of the normalised signature of a path over the interval $[0, 1]$ where γ is parametrised at unit speed.
$I_{p,n}(x)$: Inserting element x into the p-th position of the n-th term of the normalised signature of a path which is parametrised at unit speed, Equation (4.4).
$R_{p,n+1}(x)$: Replacing the p-th element of the $(n + 1)$-th term in the normalised signature of a path which is parametrised at unit speed with x, Equation (4.5).
$\|\cdot\|_l$: The ℓ^l norm, Lemma 4.2.1.

List of Figures

Chapter 1

Introduction to the signature of a path

In this chapter we introduce some basic concepts used in rough paths theory, and some important definitions and theorems we will use in the following chapters. The rough paths theory was initiated by Lyons [22] in the 1990s, and the contents of this chapter are taken from the standard textbook about rough paths by Lyons et al. [23], the paper by Hambly and Lyons [19] and the paper by Boedihardjo and Geng [5].

1.1 Characterisation of paths

For a path X in a Banach space E, one way to characterise its roughness is to discuss the *p-variation* of the path X, which is defined below. In this section, we assume E is a Banach space and J is a compact interval.

Definition 1.1.1 (p-variation paths)**.** *Assume $p \geq 1$ is a real number. Let $X : J \to E$ be a continuous path. The p-*variation *of X on the interval J is defined by*

$$\|X\|_{p,J} = \left[\sup_{D \subset J} \sum_{j=0}^{r-1} \left\| X_{t_{j+1}} - X_{t_j} \right\|^p \right]^{\frac{1}{p}},$$

where $D = \{t_0 < t_1 < \cdots < t_r : t_i \in J \ \forall\, i = 0, \cdots, r\}$ is a partition of J.

In particular, a path with finite one-variation is also known as a path with *bounded variation*, or *finite variation.*
Lyons et al. described some basic and important properties about the paths of p-variation in Section 1.2 of [23], and we state some of the properties here for later use.

Lemma 1.1.1 (Lower semi-continuity)**.** *Let* $(X_n)_{n\geq 0}$ *be a sequence of elements of* $C^0(J,E)$ *which converges pointwise to a continuous path* X*. Then*

$$\|X\|_{p,J} \leq \liminf_{n\to\infty} \|X_n\|_{p,J}\,.$$

Lemma 1.1.2 (Lyons et al., Lemma 1.6 [23])**.** *Let* $X : J \to E$ *be a continuous path. Then the followings are true:*

1. *Let* $\varphi : J \to J$ *be a non-decreasing surjection. Then, for all* $p \geq 1$, $\|X\|_{p,J} = \|X \circ \varphi\|_{p,J}$.
2. *The function* $p \mapsto \|X\|_{p,J}$ *from* $[1,+\infty)$ *to* $[0,+\infty]$ *is non-increasing.*
3. *The function* $p \mapsto \log \|X\|_{p,J}^p$ *is convex, and continuous on any interval where it is finite.*
4. *For all* $p \geq 1$, $\|X\|_{p,J} \geq \sup_{s,t\in J} \|X_s - X_t\|$.

For each $p \geq 1$, let $\mathcal{V}^p(J,E)$ denote the set of continuous paths which have finite p-variation. For each $X \in \mathcal{V}^p(J,E)$, set

$$\|X\|_{\mathcal{V}^p(J,E)} = \|X\|_{p,J} + \sup_{t\in J} \|X_t\|\,.$$

In fact p-variation paths form a vector space, as proved in [23].

Proposition 1.1.1 (Lyons et al., Proposition 1.7 [23])**.** *For each* $p \geq 1$, *the set* $\mathcal{V}^p(J,E)$ *is a linear subspace of* $C^0(J,E)$ *on which* $\|\cdot\|_{\mathcal{V}^p(J,E)}$ *is a norm. Moreover,* $(\mathcal{V}^p(J,E), \|\cdot\|_{\mathcal{V}^p(J,E)})$ *is a Banach space.*

Another important definition which characterises the roughness of a path is called *Hölder continuity.*

Definition 1.1.2 (Hölder continuity)**.** $X : [0,T] \to E$ *is* Hölder continuous *if there exists a non-negative* α *such that*

$$\sup_{s,t\in[0,T]} \frac{\|X_t - X_s\|}{|t-s|^\alpha} < \infty.$$

Then X *is called a Hölder continuous path with exponent* α.

Definition 1.1.3 (Control functions)**.** *A* control function*, or* control*, on* $[0,T]$ *is a continuous non-negative function* ω *on* $\{(s,t) \in [0,T]^2 : 0 \leq s \leq t \leq T\}$ *which is super-additive in the sense that*

$$\omega(s,t) + \omega(t,u) \leq \omega(s,u) \quad \forall s \leq t \leq u \in J,$$

for which $\omega(t,t) = 0$ *for all* $t \in J$.

Lyons et al. [23] showed that there exists a relationship between p-variation and Hölder continuity: Note that for a path $X \in \mathcal{V}^p([0,T],E)$, the function $\omega_X(s,t) := \|X\|^p_{p,[s,t]}$ is a control function for $s \le t \in [0,T]$. The function $\omega_X(0,\cdot)$ provides a natural re-parametrisation for X. If we assume X is constant on no sub-interval of J, the function $t \to \omega_X(0,t)\frac{T}{\omega_X(0,T)}$ is an increasing bijection $J \to J$. Let $t \to \tau(t)$ be its inverse. Note that then, for all $s \le t$ in J,

$$\left\|X_{\tau(s)} - X_{\tau(s)}\right\|^p \le \omega_X(\tau(s),\tau(t)) \le \omega_X(0,\tau(t)) - \omega_X(0,\tau(s)) = \frac{\omega_X(0,T)}{T}(t-s).$$

Hence any path of finite p-variation is Hölder continuous with exponent $\frac{1}{p}$ up to re-parametrisation.
We also recall the following definitions of a path.

Definition 1.1.4 (Length of a path)**.** *Assume $X : I \to E$ is a continuous path. The* length of X, *L_X is defined such that for all $u \in I$,*

$$L_X(u) := \sup_D \sum_{u_{i-1},u_i \in D} \left\|X_{u_i} - X_{u_{i-1}}\right\|,$$

where D is a partition on $\{t \in I : t \le u\}$.

Definition 1.1.5 (Parametrisation at unit speed)**.** *Assume $X : I \to E$ is a continuous path of length $L < \infty$. X is* parametrised at unit speed *if*

$$L_X(t) = t \quad \forall\, t \in [0,L].$$

The next lemma guarantees that parametrisation will not significantly affect the properties of a path. This is a well-known fact and we include a proof in Appendix A just for completeness.

Lemma 1.1.3. *Assume $I = [0,T]$. If $X : I \to E$ is a continuous path of length $L < \infty$, then X can be re-parametrised at unit speed.*

Proof. See Appendix A. □

1.2 The signature of a path

Let X and Y be two real-valued functions on the interval $[0,T]$. It is known that if X has bounded variation, given that certain conditions are satisfied, the integral

$\int_0^t Y \mathrm{d}X$ can be defined for every $t \in [0,T]$. Moreover, as a function of t, this integral has bounded variation, and

$$\left\| \int_0^{\cdot} Y_s \mathrm{d}X_s \right\|_{1,[0,T]} \leq \|Y\|_{\infty,[0,T]} \|X\|_{1,[0,T]},$$

where $\|\cdot\|_{\infty,[0,T]}$ denotes the uniform norm on $[0,T]$. Young showed that if X has unbounded varation, the integral $\int Y \mathrm{d}X$ can still be defined.

Theorem 1.2.1 (Young, see e.g. Theorem 1.16 of [23]). *Let V and W be two Banach spaces. Let $p, q \geq 1$ be two real numbers such that $\frac{1}{p} + \frac{1}{q} > 1$. Let T be a positive real number. Assume $X \in \mathcal{V}^p([0,T], V)$ and $Y \in \mathcal{V}^q([0,T], \mathbf{L}(V,W))$, where $\mathbf{L}(V,W)$ denotes the set of continuous linear functions from V to W. Let $\mathcal{D} = \{0 = t_0 \leq t_1 \leq \cdots \leq t_{r-1} \leq t_r = T\}$ be a partition on $[0,T]$. Then for each $t \in [0,T]$, the integral*

$$\int_0^t Y_s \mathrm{d}X_s := \lim_{|\mathcal{D}| \to 0, \mathcal{D} \in [0,t]} \sum_{i=0}^{r-1} Y_{t_i}(X_{t_{i+1}} - X_{t_i})$$

is well-defined. As a function of t, this limit belongs to $\mathcal{V}^p([0,T], W)$ and there exists a constant $C_{p,q}$ which depends only on p and q such that the following inequality holds:

$$\left\| \int_0^{\cdot} (Y_s - Y_0) \mathrm{d}X_s \right\|_{p,[0,T]} \leq C_{p,q} \|Y\|_{q,[0,T]} \|X\|_{p,[0,T]}.$$

For many reasons, it is interesting to consider iterated integrals in tensor sense. In this section we describe the following setup as introduced by Lyons et al. in Section 2.2 of [23]. The convention $E^{\otimes 0} = \mathbb{R}$ is adopted.

Definition 1.2.1. *The space of* formal series of tensors of E, *denoted by $T((E))$, is defined to be the following space of sequences:*

$$T((E)) = \left\{ \mathbf{a} = (a_0, a_1, \cdots) : \forall n \geq 0, a_n \in E^{\otimes n} \right\}.$$

Let $\mathbf{a} = (a_0, a_1, \cdots)$ and $\mathbf{b} = (b_0, b_1, \cdots)$ be two elements of $T((E))$. We can equip $T((E))$ with two internal operations, addition and multiplication:

$$\mathbf{a} + \mathbf{b} = (a_0 + b_0, a_1 + b_1, \cdots),$$

and

$$\mathbf{a} \otimes \mathbf{b} = (c_0, c_1, \cdots),$$

where for each $n \geq 0$,

$$c_n = \sum_{k=0}^{n} a_k \otimes b_{n-k}.$$

We can also denote $\mathbf{a} \otimes \mathbf{b}$ by $\mathbf{ab}$.

The space $T((E))$ endowed with the operations described above and the natural action of $\mathbb{R}$ by $\lambda\mathbf{a} = (\lambda a_0, \lambda a_1, \cdots)$ is a real non-commutative unital algebra, with unit $\mathbf{1} = (1, 0, 0, \cdots)$.

An element $\mathbf{a} = (a_0, a_1, \cdots)$ of $T((E))$ is invertible if and only if $a_0 \neq 0$. Its inverse is then given by

$$\mathbf{a}^{-1} = \frac{1}{a_0} \sum_{n \geq 0} \left(\mathbf{1} - \frac{\mathbf{a}}{a_0} \right)^n,$$

which is well-defined because, for each given degree, only finitely many terms of the sum produce non-zero tensors of this degree. In particular, the subset

$$\tilde{T}((E)) = \{\mathbf{a} \in T((E)) : a_0 = 1\}$$

is a group.

It is often important to look only at finitely many terms of an element of $T((E))$. For each $n \geq 0$, the space $B_n = \{\mathbf{a} = (a_0, a_1, \cdots) : a_0 = \cdots = a_n = 0\}$ of formal series with no monomials of degree less than or equal to n is an ideal of $T((E))$.

Definition 1.2.2 (Truncated tensor algebra)**.** *Let $n \geq 1$ be an integer. The* truncated tensor algebra of order n of E *is defined as the quotient algebra*

$$T^{(n)}(E) = T((E))/B_n.$$

In fact $T^{(n)}(E)$ is canonically isomorphic to $\oplus_{k=0}^{n} E^{\otimes k}$ equipped with the product

$$(a_0, \cdots, a_n)(b_0, \cdots, b_n) = (c_0, \cdots, c_n),$$

where $c_k = \sum_{i=0}^{k} a_i \otimes b_{k-i}$ for all $k \in \{0, \cdots, n\}$.

We are now ready to define the signature of a path.

Definition 1.2.3 (Signature of a path)**.** *Let J denote a compact interval. Let $X : J \to E$ be a continuous path of finite p-variation for some $p < 2$. The* signature of *X is*

$$\mathbf{S}_J(X) = (1, S_J^1(X), S_J^2(X), \cdots),$$

where for each $n \geq 1$, $S_J^n(X) = \int_{u_1 < \cdots < u_n u_1, \cdots, u_n \in J} \mathrm{d}X_{u_1} \otimes \cdots \otimes \mathrm{d}X_{u_n}$.

Note that the signature of a path is an element of $\tilde{T}((E))$.
The range of the signature mapping is important to consider. Lyons et al. [23] discussed some important properties about the mapping, and we include some of them below.

Definition 1.2.4. *Let $X : [0, s] \to E$ and $Y : [s, t] \to E$ be two continuous paths. Their* concatenation *is the path $X * Y : [0, t] \to E$ defined by*

$$(X * Y) = \begin{cases} X_u & \text{if} \quad u \in [0, s] \\ X_s + Y_u - Y_s & \text{if} \quad u \in [s, t]. \end{cases}$$

It is well-known that the concatenated path preserves the variation. We include a proof in Appendix A just for completeness.

Lemma 1.2.1. *Assume $0 < s < t$. Let $X : [0, s] \to E$ and $Y : [s, t] \to E$ be two continuous paths both of finite p-variation for $p \geq 1$. Then $X * Y$ is of finite p-variation.*

Proof. See Appendix A. □

Chen [11] proved that the signature is a homomorphism, which is stated in the next theorem in the language of Lyons et al. [23].

Theorem 1.2.2 (Chen, see e.g. Theorem 2.9 of [23])**.** *Let $X : [0, s] \to E$ and $Y : [s, t] \to E$ be two continuous paths with finite one-variation. Then*

$$\mathbf{S}_{0,t}(X * Y) = \mathbf{S}_{0,s}(X) \otimes \mathbf{S}_{s,t}(Y).$$

Lyons et al. [23] extended Chen's theorem to p-variation paths for $p < 2$.

Theorem 1.2.3 (Lyons et al., Corollary 2.13 of [23])**.** *Theorem 1.2.2 holds for paths of finite p-variation for $p < 2$.*

By Lemma 1.2.1 and Theorem 1.2.3, we can see that the range of the signature mapping is closed under multiplication. If X is a path of finite p-variation, it was also showed in [23] that the inverse of $\mathbf{S}(X)$, denoted by $\mathbf{S}(X)^{-1}$, is the signature of a path.

Proposition 1.2.1 (Lyons et al., Proposition 2.14 of [23])**.** *Let $X : [0, T] \to E$ be a path of finite p-variation for $p < 2$. Let $\overleftarrow{X}$ be the path X run backwards, i.e. the path defined by $\overleftarrow{X}_t = X_{T-t}$, $t \in [0, T]$. Then*

$$\mathbf{S}(\overleftarrow{X}) = \mathbf{S}(X)^{-1}.$$

In particular, the range of $\mathbf{S} : \mathcal{V}^p([0, T], E) \to T((E))$ is a group.

It is also interesting to note that the signature of a path is a solution to a particular differential equation.

Lemma 1.2.2 (Lyons et al., Lemma 2.10 of [23])**.** *Let $X : [0,T] \to E$ be a path of finite p-variation for some $p < 2$. Let $\mathbf{L}(E, T^{(n)}(E))$ denote the set of continuous linear mappings from E to $T^{(n)}(E)$. Define $f : T^{(n)}(E) \to \mathbf{L}(E, T^{(n)}(E))$ by*

$$f(a_0, a_1, \cdots, a_n)x = (0, a_0 \otimes x, a_1 \otimes x, \cdots, a_{n-1} \otimes x).$$

Then the unique solution to the differential equation

$$\mathrm{d}S_t = f(S_t)\mathrm{d}X_t, \quad S_0 = (1, 0, \cdots, 0)$$

is the path $S : [0,T] \to T^{(n)}(E)$ defined for all $t \in [0,T]$ by

$$S_t = (1, S^1_{0,t}(X), \cdots, S^n_{0,t}(X)).$$

Remark 1.2.1. *We can also write the following differential equation for the full signature of X as a function from $[0,T]$ to $T((E))$:*

$$\mathrm{d}\mathbf{X}_{0,t} = \mathbf{X}_{0,t} \otimes \mathrm{d}X_t, \quad \mathbf{X}_{0,0} = \mathbf{1}.$$

If we adopt the terminology of [23], the signature of a path is then the solution to a rough differential equation. Since the solution to a rough differential equation is unique, we can conclude that the signature of a path is invariant under re-parametrisation.

We can also discuss functions on signatures. If we regard the tensor as a word over some alphabet, *shuffle product* is an important concept we should recall.

Definition 1.2.5 (Shuffle product)**.** *The shuffle product is defined inductively to be bilinear, such that for any words u, v, a and b,*

$$u \otimes a ⧢ v \otimes b := (u ⧢ v \otimes b) \otimes a + (u \otimes a ⧢ v) \otimes b.$$

Definition 1.2.6 (Group-like elements)**.** *Let V be a Banach space. Define*

$$\tilde{T}((V)) := \left\{(a_0, a_1, a_2, \cdots) : a_n \in V^{\otimes n}\, \forall n \geq 1, a_0 = 1\right\}.$$

An element $\mathbf{a} \in \tilde{T}((V))$ is called group-like *if for all $\phi, \psi : \tilde{T}((V)) \to \mathbb{R}$,*

$$\phi ⧢ \psi(\mathbf{a}) = \phi(\mathbf{a})\psi(\mathbf{a}).$$

Lyons et al. showed in Section 2.2 of [23] that the signature of a path is a group-like element.

1.3 Uniqueness of the signature of a path

We first state the properties of the norms on tensor products which we assume to be true.

Definition 1.3.1. *Let V be a Banach space. We say that its tensor powers are endowed with admissible norms if the following conditions hold:*

1. *For each $n \geq 1$, the symmetric group S_n acts by isometries on $V^{\otimes n}$, i.e.*

$$\|\sigma v\| = \|v\| \quad \forall v \in V^{\otimes n}, \forall \sigma \in S_n.$$

2. *The tensor product has norm 1, i.e. for all $n, m \geq 1$,*

$$\|v \otimes \omega\| \leq \|v\| \, \|\omega\| \quad \forall v \in V^{\otimes n}, \omega \in V^{\otimes m}.$$

For the rest of this chapter we consider paths of bounded-variation. Equipped with a tensor norm with properties described in Definition 1.3.1, the signature of a path is bounded above [23].

Proposition 1.3.1 (Lyons et al., Proposition 2.2 of [23])**.** *Let $X : [0, T] \to V$ be a path with finite variation. Then, for each $k \geq 1$, one has*

$$\left\| \int_{0<u_1<\cdots<u_k<T} \mathrm{d}X_{u_1} \otimes \cdots \otimes \mathrm{d}X_{u_k} \right\| \leq \frac{\|X\|_{1,[0,T]}^k}{k!}.$$

Remark 1.3.1. *If we recall the definition of the length of a path, and suppose $X : [0, T] \to V$ is a bounded-variation path of length $L > 0$, since the length of X is invariant under the re-parametrisation as described in Lemma 1.1.3, then by Proposition 1.3.1,*

$$\left\| \int_{0<u_1<\cdots<u_k<T} \mathrm{d}X_{u_1} \otimes \cdots \otimes \mathrm{d}X_{u_k} \right\| \leq \frac{L^k}{k!}.$$

In 2010, Hambly and Lyons [19] proved the uniqueness of the signature of a bounded-variation path. In this section we state some of their results for future references. More details can be found in [19].
Consider bounded-variation paths in $\mathbb{R}^d$ equipped with the Euclidean norm. An important tool used by Hambly and Lyons is the *hyperbolic development* of the path. Consider the quadratic form on $\mathbb{R}^{d+1}$ defined by

$$I_d(x, y) = \sum_{i=1}^{d} x_i y_i - x_{d+1} y_{d+1}, \quad \forall x, y \in \mathbb{R}^{d+1}.$$

Also, define

$$\mathbb{H} = \{x : I_d(x,x) = -1\}.$$

Then $\mathbb{H}$ is hyperbolic space. Let $SO(I_d)$ denote the group of matrices with positive determinant preserving the quadratic form I_d, i.e. $M \in SO(I_d)$ if $I_d((My^T)^T, (Mx^T)^T) = I_d(y,x)$.

Definition 1.3.2 (Cartan development)**.** *Define the mapping $F : \mathbb{R}^d \to Hom\left(\mathbb{R}^{d+1}, \mathbb{R}^{d+1}\right)$ such that for $x = (x_1, \cdots, x_d) \in \mathbb{R}^d$,*

$$F : x \mapsto \begin{pmatrix} 0 & \cdots & 0 & x_1 \\ \vdots & \ddots & \vdots & \vdots \\ 0 & \cdots & 0 & x_d \\ x_1 & \cdots & x_d & 0 \end{pmatrix}, \tag{1.1}$$

where $Hom(V, A)$ denotes the set of bounded linear operators from V to A. Then the Cartan development *of a continuous bounded-variation path $\gamma : [0,1] \to \mathbb{R}^d$ to $SO(I_d)$ is given by solving the following differential equation*

$$\begin{aligned} \mathrm{d}\Gamma_t &= \Gamma_t F(\mathrm{d}\gamma_t), \quad t \in [0,1], \\ \Gamma_0 &= I_{d+1}, \end{aligned}$$

where I_{d+1} is the $(d+1)\times(d+1)$ identity matrix, i.e. the development Γ satisfies

$$\mathrm{d}\Gamma_t = \Gamma_t \begin{pmatrix} 0 & \cdots & 0 & \mathrm{d}\gamma_t^1 \\ \vdots & \ddots & \vdots & \vdots \\ 0 & \cdots & 0 & \mathrm{d}\gamma_t^d \\ \mathrm{d}\gamma_t^1 & \cdots & \mathrm{d}\gamma_t^d & 0 \end{pmatrix}. \tag{1.2}$$

Note by Picard iteration, we have

$$\Gamma_t = \sum_{n=0}^{\infty} \int_{0<t_1<\cdots<t_n<t} F\left(\mathrm{d}\gamma_t\right) \cdots F\left(\mathrm{d}\gamma_t\right).$$

Definition 1.3.3 (Hyperbolic development)**.** *We define X to be the* hyperbolic development *of a continuous bounded-variation path $\gamma : [0,1] \to \mathbb{R}^d$ into $\mathbb{H}$ starting at $o = (0, 0, \cdots, 1)^T$ and given by*

$$X_t = \Gamma_t o,$$

where Γ is the Cartan development of γ.

In such a setting, we require $\mathbb{R}^d$ and $\mathbb{R}^{d+1}$ to be both equipped with the Euclidean norm. Note that under such assumptions, the hyperbolic development does not alter the length of the path, which was clearly stated by Boedihardjo and Geng [5] as described in the following lemma.

Lemma 1.3.1 (Boedihardjo and Geng, Fact 1 [5])**.** *The hyperbolic development described in Definition 1.3.3 is length preserving. Moreover, if γ_t is piecewise linear, then its hyperbolic development X_t is piecewise geodesic with the same intersection angles as those of γ_t.*

Definition 1.3.4. *If we develop path γ of fixed length l into path Γ in $SO(I_d)$, let $d(o, \Gamma_o)$ denote the length of the chord connecting the beginning and the end of the development of γ into hyperbolic space.*

Definition 1.3.5. *If V is a Banach space, A is a Banach algebra and $F_1, \cdots, F_k \in Hom(V, A)$, then write $F_1 \otimes \cdots \otimes F_k$ for the canonical linear extension of the multilinear map from $V^{\otimes k}$ to A such that*

$$(v_1, \cdots, v_k) \to F_1(v_1) \cdots F_k(v_k).$$

We can define the norm

$$\|x\|_{\to A} := \sup_{F_i \in Hom(V,A), \|F_i\|_{Hom(V,A)}=1} \|F_1 \otimes \cdots \otimes F_k(x)\|_A .$$

Note $A = \mathbb{R}$ yields the injective cross norm. In our case we consider the situation when $A = \mathrm{Hom}(\mathbb{R}^{d+1}, \mathbb{R}^{d+1})$ with the operator norm and $\mathbb{R}^{d+1}$ is given the Euclidean norm.

Lemma 1.3.2 (Hambly and Lyons, Lemma 3.1 [19])**.** *If F is as defined in Equation (1.1),*

$$\|F\|_{Hom(\mathbb{R}^{d+1}, \mathbb{R}^{d+1})} = 1.$$

We note the following proposition which gives an upper bound on the length of the chord connecting the beginning and the end of the hyperbolic development.

Proposition 1.3.2 (Hambly and Lyons, Proposition 3.13 [19])**.** *Let $G \in SO(I_d)$. Then $\|G\|_{Hom(\mathbb{R}^{d+1}, \mathbb{R}^{d+1})} \geq e^{d(o,Go)}$ where $\mathbb{R}^{d+1}$ is equipped with the Euclidean norm.*

The following lemma describes the relationships between distances in hyperbolic space.

Lemma 1.3.3 (Hambly and Lyons, Lemma 3.7 [19])**.** *Let $0 = T_0 < \cdots < T_i < \cdots < T_n = T$ be a partition of $[0,T]$. Let $(X_t)_{t\in[0,T]}$ be a continuous path, geodesic on the intervals $[T_i, T_{i+1}]|_{i=0,\cdots,n-1}$ in hyperbolic space with $n \geq 1$ where, at each T_i, the angle between the two geodesic segments $\angle X_{T_{i-1}} X_{T_i} X_{T_{i+1}}$ is in $[2\Omega, \pi]$, $\Omega > 0$. Suppose that each geodesic segment has length at least $K(\Omega) = \log\left(\frac{2}{1-\cos|\Omega|}\right)$. Let d denote the metric on the hyperbolic space. Then*

1. *$d(X_0, X_{T_i})$ is increasing in i and for each $i \leq n$,*
$$d(X_0, X_{T_i}) \geq d(X_0, X_{T_{i-1}}) + d(X_{T_{i-1}}, X_{T_i}) - K(\Omega) \geq K(\Omega),$$
and the angle between $\overrightarrow{X_{T_{i-1}} X_{T_i}}$ and $\overrightarrow{X_0 X_{T_i}}$ is at most Ω.

2. *We also have*
$$0 \leq \sum_{i=1}^{n} d(X_{T_{i-1}}, X_{T_i}) - d(X_0, X_{T_n}) \leq (n-1)K(\Omega).$$

Let $\gamma : [a,b] \to \mathbb{R}^d$ be a continuous path of finite length l, parametrised at unit speed. If $\alpha \in \mathbb{R}$, and without loss of generality we assume $\alpha > 0$, then the path $\gamma_\alpha : [\alpha a, \alpha b] \to \mathbb{R}$, $\gamma_\alpha := t \to \alpha\gamma(t/\alpha)$ is also parametrised at unit speed, and the length is αl:

$$\begin{aligned}\int_{a\alpha}^{b\alpha} |\dot{\gamma_\alpha}(t)| \, \mathrm{d}t &= \int_{a\alpha}^{b\alpha} \left|\dot{\gamma}(\frac{t}{\alpha})\right| \mathrm{d}t \\ &= \int_a^b \alpha \, |\dot{\gamma}(s)| \, \mathrm{d}s \\ &= \alpha l.\end{aligned}$$

If we consider the case where $u \to \gamma'(u)$ is continuous with modulus of continuity δ_γ, the derivative of γ_α has modulus of continuity $\delta_{\gamma_\alpha}(\alpha h) = \delta_\gamma(h)$. Its development from the identity matrix into $SO(I_d)$ is denoted by Γ_α. Hambly and Lyons [19] showed that the signature of a piecewise linear path has at least one term that is bounded below.

Theorem 1.3.1 (Hambly and Lyons, Theorem 13 [19])**.** *If γ is a non-degenerate piecewise linear path in $\mathbb{R}^d$, 2Ω is the smallest angle between adjacent edges, and $D > 0$ is the length of the shortest edge, then there is at least one n for which*

$$\left(\frac{2}{1-\cos|\Omega|}\right)^{1-\frac{1}{D}} \leq n! \left\| \int_{0<u_1<\cdots<u_n<T} \mathrm{d}\gamma(u_1) \otimes \cdots \otimes \mathrm{d}\gamma(u_n) \right\|_{\to Hom(\mathbb{R}^{d+1}, \mathbb{R}^{d+1})},$$

and in particular γ has non-trivial signature, i.e. $\mathbf{S}(\gamma) \neq (1, 0, 0, \cdots)$.

Another important definition mentioned by Hambly and Lyons [19] is the *tree-like paths.*

Definition 1.3.6 (Tree-like paths)**.** *$X : [0,T] \to E$ is a* tree-like path *in E if there exists a continuous function $h : [0,T] \to [0,+\infty)$ such that $h(0) = h(T) = 0$, and for all $s,t \in [0,T]$ with $s \le t$,*

$$\|X_t - X_s\| \le h(s) + h(t) - 2 \inf_{u \in [s,t]} h(u).$$

The function h is called a height function for X. We say X is a Lipschitz tree-like path if h can be chosen to be of bounded-variation.

Theorem 1.3.2 (Hambly and Lyons, Theorem 4 [19])**.** *Let X be a bounded-variation path in $\mathbb{R}^d$. The path is tree-like if and only if the signature of X is $(1,0,0,\cdots)$.*

Theorem 1.3.3 (Hambly and Lyons, Corollary 1.6 [19])**.** *For a bounded-variation path X in $\mathbb{R}^d$, there exists a unique path of minimal length $\tilde{X}$, called* the tree-reduced path*, with the same signature $\mathbf{S}(X) = \mathbf{S}(\tilde{X})$.*

Here we briefly explain the idea of the proof. Define the relation $\sim$ such that for bounded-variation paths X and Y in $\mathbb{R}^d$, $X \sim Y$ if $X * \overleftarrow{Y}$ is a Lipschitz tree-like path. $\sim$ is in fact an equivalence relation, and $\mathbf{S}(X) = \mathbf{S}(Y)$ if and only if $X * \overleftarrow{Y}$ is a Lipschitz tree-like path, that is, if $X \sim Y$, then $\mathbf{S}(X) = \mathbf{S}(Y)$. Within each equivalence class, there exists a unique path $\tilde{X}$ with minimal length, therefore $\mathbf{S}(X) = \mathbf{S}(\tilde{X})$.

1.4 Information about a path from its signature

We have seen that the signature of a tree-reduced bounded-variation path is unique, therefore the signature uniquely characterises the path up to tree-like pieces. We can then discuss the information about the path that is contained in the signature.

Definition 1.4.1 (Symmetric tensors)**.** *Let V be a vector space, and let $v \in V^{\otimes k}$. Then v is a* symmetric tensor *if for any permutation $\sigma \in S(k)$,*

$$\sigma(v) = v.$$

For $u \in V^{\otimes k}$, the symmetric part of u is the tensor defined by

$$sym(u) = \frac{1}{k!} \sum_{\sigma \in S(k)} \sigma(u).$$

As an example, for a tensor $x \otimes y \otimes z$, the *symmetric part* of the tensor is the sum of the tensors

$$\begin{aligned}
&\operatorname{sym}(x \otimes y \otimes z) \\
&= \frac{1}{3!}(x \otimes y \otimes z + y \otimes x \otimes z + y \otimes z \otimes x + x \otimes z \otimes y + z \otimes x \otimes y + z \otimes y \otimes x).
\end{aligned}$$

By recalling the fact that the signature of a path is a collection of tensors of different degrees, we conclude the following lemma. Note that the following lemma is in fact a special case of Exercise 3.15 stated in [23].

Lemma 1.4.1 (Lyons et al. [23])**.** *Let $X : [0,T] \to \mathbb{R}^d$ be a path of finite variation. Then for every $n \geq 1$, the symmetric part of the n-th level signature, $S^n_{0,T}(X)$, is equal to $\frac{(S^1_{0,T}(X))^{\otimes n}}{n!}$.*

Proof. See Appendix A. $\square$

Let $X : [0,T] \to \mathbb{R}^d$ be a bounded-variation path in $\mathbb{R}^d$. We can write the path coordinate-wise as $(X^i_t)_{i=1,\cdots,d}$. Define

$$X^i_{s,t} = X^i_t - X^i_s, \quad X^{ij}_{s,t} = \int_{s<u_1<u_2<t} \mathrm{d}X^i_{u_1}\mathrm{d}X^j_{u_2}, \quad i,j = 1,\cdots,d.$$

Then $(1, (X^i_{s,t})_{i=1,\cdots,d}, (X^{ij}_{s,t})_{i,j=1,\cdots,d})$ is the truncated signature of X up to level 2. We can see that $(X^i_{s,t})_{i=1,\cdots,d}$ gives the increment of X over $[s,t]$. For the second level signature, we can write it as sums of symmetric and anti-symmetric parts:

$$\begin{aligned}
X^{ij}_{s,t} &= \frac{1}{2}\int_{s<u_1,u_2<t} \mathrm{d}X^i_{u_1}\mathrm{d}X^j_{u_2} - \frac{1}{2}\int_{s<u_2<u_1<t} \mathrm{d}X^i_{u_1}\mathrm{d}X^j_{u_2} + \frac{1}{2}\int_{s<u_1<u_2<t} \mathrm{d}X^i_{u_1}\mathrm{d}X^j_{u_2} \\
&= \frac{1}{2}(X^i_t - X^i_s)(X^j_t - X^j_s) + \frac{1}{2}\int_{s<u_1<u_2<t} \mathrm{d}X^i_{u_1}\mathrm{d}X^j_{u_2} - \mathrm{d}X^j_{u_1}\mathrm{d}X^j_{u_2} \\
&= \frac{1}{2}(X^i_t - X^i_s)(X^j_t - X^j_s) + A^{ij}_{s,t},
\end{aligned}$$

where the anti-symmetric part $A^{ij}_{s,t} = \frac{1}{2}\int_{s<u_1<u_2<t} \mathrm{d}X^i_{u_1}\mathrm{d}X^j_{u_2} - \mathrm{d}X^j_{u_1}\mathrm{d}X^i_{u_2}$. Note we can write

$$A^{ij}_{s,t} = \frac{1}{2}\int_s^t (X^i_u - X^i_s)\mathrm{d}X^j_u - (X^j_u - X^j_s)\mathrm{d}X^i_u,$$

and by Green's Theorem, $A^{ij}_{s,t}$ is the area enclosed by the curve from (X^i_s, X^j_s) to (X^i_t, X^j_t) and the chord connecting the two points, provided orientation and multiplicity are taken into account.

From above we see that the signature of a path contains information about the increment and the area of the path. As the level goes higher, the truncated signature contains more information about the path. We have also seen that the signature of a bounded-variation path is unique up to tree-like pieces, therefore it is an interesting topic to develop algorithms to reconstruct the path with minimal length from the information available in the signature.

Chapter 2

Super-multiplicativity and a lower bound for the decay of the signature of a path of finite length

From Chapter 1 we have seen that there exists an upper bound for the n-th level of the signature of a path. It is then natural and interesting to study whether a lower bound of the signature of a path exists. For a path of length $L > 0$, for all $n \geq 1$, we multiply the n-th term of the signature by $n!L^{-n}$, and we say the resulting signature is '*normalised*'. It has been established [23] that the norm of the n-th term of the normalised signature of a bounded-variation path is bounded above by 1. In this chapter we discuss the super-multiplicativity of the norm of the signature of a path of finite length, and prove by Fekete's lemma the existence of a non-zero limit of the n-th root of the norm of the n-th term in the signature multiplied by $n!$ as n approaches infinity [9].

2.1 Reasonable tensor algebra norms

Definition 2.1.1 (Algebraic tensor product space). *Let $\{V_j\}_{j=1}^{N}$ be normed vector spaces over $\mathbb{F}=\mathbb{R}$ or $\mathbb{C}$. Their* algebraic tensor product space *is defined as the vector space*

$$V_1 \otimes \cdots \otimes V_N = \left\{ \sum_{i \in I} v_i^1 \otimes \cdots \otimes v_i^N : v_i^j \in V_j, \quad \forall i \in I, |I| < \infty, j = 1, \cdots, N. \right\},$$

where we identify $(u+v) \otimes w = u \otimes w + v \otimes w$ and $u \otimes (v+w) = u \otimes v + u \otimes w$.

Definition 2.1.2. *If $\phi_j \in V_j'$ are bounded linear functionals on V_j, $j = 1, \cdots, N$, then we define the dual action of $\phi_1 \otimes \cdots \otimes \phi_N$ on $V_1 \otimes \cdots \otimes V_N \to \mathbb{F}$ by*

$$(\phi_1 \otimes \cdots \otimes \phi_N)(v_1 \otimes \cdots \otimes v_N) := \prod_{i=1}^{N} \phi(v_i) \quad \forall v_i \in V_i, i = 1, \cdots, N,$$

and extending by linearity.

Definition 2.1.2 is in fact a generalisation of the mapping described on the tensor product of two Banach spaces in [29].
Recall that if $\phi \in V'$, then the norm of ϕ is defined as

$$\|\phi\| := \sup_{\|x\|=1} |\phi(x)| .$$

Definition 2.1.3 (Reasonable tensor algebra norms)**.** *Let $V, V \otimes V, \cdots, V^{\otimes n}$ be normed vector spaces. We assume that for all $v \in V^{\otimes n}$, $w \in V^{\otimes m}$,*

$$\|v \otimes w\| \leq \|v\| \, \|w\| \tag{2.1}$$

and the norm induced on the dual spaces satisfies that for all $\phi \in (V^{\otimes m})', \psi \in (V^{\otimes n})'$,

$$\|\phi \otimes \psi\| \leq \|\phi\| \, \|\psi\| . \tag{2.2}$$

Moreover, if $S(n)$ denotes the symmetric group over $\{1, 2, \cdots, n\}$, we assume that for all $n \geq 1$,

$$\|\sigma(v)\| = \|v\| \quad \forall \sigma \in S(n), \, v \in V^{\otimes n}.$$

If a norm satisfies the properties above, then we say it is a reasonable tensor algebra norm*.*

Proposition 2.1.1. *Let X and Y be normed vector spaces. If $\|\cdot\|$ is a tensor norm on $X \otimes Y$ which satisfies*

$$\|v \otimes w\| \leq \|v\| \, \|w\| \quad \forall v \in X, w \in Y,$$

and the norm induced on the dual spaces satisfies

$$\|\phi \otimes \psi\| \leq \|\phi\| \, \|\psi\| \quad \forall \phi \in X', \psi \in Y',$$

then $\|\cdot\|$ is called a reasonable cross norm*, and $\|x \otimes y\| = \|x\| \, \|y\|$ for every $x \in X$ and $y \in Y$; for every $\phi \in X'$ and $\psi \in Y'$, the norm of the linear functional $\phi \otimes \psi$ on $(X \otimes Y, \|\cdot\|)$ satisfies $\|\phi \otimes \psi\| = \|\psi\| \, \|\psi\|$.*

Proof. For a proof see for example Proposition 6.1 of [29] by Ryan. □

Proposition 2.1.1 implies that the inequalities in Equation (2.1) and (2.2) imply equality.

Remark 2.1.1. *Note that under the assumptions of Definition 2.1.3 for all* $a \in V^{\otimes m}$, $b \in V^{\otimes n}$, $c \in V^{\otimes l}$,

$$\|(a \otimes b) \otimes c\| = \|a \otimes (b \otimes c)\| = \|a\| \, \|b\| \, \|c\| \, .$$

Definition 2.1.4 (Projective tensor norm and injective tensor norm)**.** *Let* $\{V_j\}_{j=1}^N$ *be normed vector spaces over* $\mathbb{F}$*. The* projective tensor norm *on* $V_1 \otimes \cdots \otimes V_N$ *is defined such that for* $x \in V_1 \otimes \cdots \otimes V_N$,

$$\|x\|_\pi := \inf \left\{ \sum_{i \in I} \|v_i^1\| \cdots \|v_i^N\| : x = \sum_{i \in I} v_i^1 \otimes \cdots \otimes v_i^N, v_i^j \in V_j \, \forall i \in I, |I| < \infty \right\}.$$

The injective tensor norm *on* $V_1 \otimes \cdots \otimes V_N$ *is defined such that for* $x = \sum_{i \in I} v_i^1 \otimes \cdots \otimes v_i^N \in V_1 \otimes \cdots \otimes V_N$*,* $i \in I$*,* $|I| < \infty$*,*

$$\|x\|_\delta := \sup \left\{ \left| \sum_{i \in I} \prod_{j=1}^N \phi_j(v_i^j) \right| : \phi_j \in V_j', \|\phi_j\| \leq 1 \, \forall j = 1, \cdots, N \right\}$$

for any representation of x*.*

It is well-known that the projective tensor product and the injective tensor product are associative, i.e. if V_1, V_2 and V_3 are normed vector spaces, for $p = \delta$ or π, the natural mapping from $(V_1 \otimes_p V_2) \otimes_p$ to $V_1 \otimes_p (V_2 \otimes_p V_3)$ is an isometry, where $V_1 \otimes_p V_2$ denotes the completion of $V_1 \otimes V_2$ with norm p. One can refer to [3], [15], and [32] for related discussions.

The projective tensor norm and the injective tensor norm are reasonable tensor algebra norms, and this is in fact implied by the proofs given by Ryan [29]. We include the proofs in Appendix A just for completeness.

Lemma 2.1.1. *The projective tensor norm defined in Definition 2.1.4 satisfies the properties stated in Definition 2.1.3.*

Proof. See Appendix A. □

Lemma 2.1.2. *The injective tensor norm defined in Definition 2.1.4 satisfies the properties stated in Definition 2.1.3.*

Proof. See Appendix A. ☐

Lemma 2.1.3. *If α is a reasonable cross norm on $X \otimes Y$, and $u \in X \otimes Y$, then*

$$\|x\|_\delta \leq \alpha(x) \leq \|x\|_\pi .$$

As a result, any reasonable tensor algebra norm is sandwiched between the injective and projective tensor norms.

Proof. A proof of Lemma 2.1.3 can be found in Proposition 6.1 of [29]. ☐

Another example of a reasonable tensor algebra norm is the *Hilbert-Schmidt norm*, as we show in the following.
Suppose for now H is a Hilbert space equipped with inner-product $\langle \cdot, \cdot \rangle$. Then the natural norm induced by the inner-product is defined such that

$$\forall x \in H, \quad \|x\| = \sqrt{\langle x, x \rangle}.$$

Definition 2.1.5 (Hilbert-Schmidt inner-product)**.** *If H_1 and H_2 are Hilbert spaces with inner-products $\langle \cdot, \cdot \rangle_1$ and $\langle \cdot, \cdot \rangle_2$ respectively, define the function $\langle \cdot, \cdot \rangle$ on the algebraic tensor product space $H_1 \otimes H_2$ by*

$$\langle a \otimes u, b \otimes v \rangle = \langle a, b \rangle_1 \langle u, v \rangle_2 \quad \forall a, b \in H_1, u, v \in H_2 \tag{2.3}$$

and extending by linearity. Note this definition of $\langle \cdot, \cdot \rangle$ does not depend on the choice of basis.

It is easy to show that $\langle \cdot, \cdot \rangle$ defines an inner-product on $H_1 \otimes H_2$, and here we give a proof that for any $x \in H_1 \otimes H_2$, $\langle x, x \rangle \geq 0$ and $\langle x, x \rangle = 0$ if and only if $x = 0$:
Suppose $\{e_i : i \in I\}$ is an orthonormal basis of H_1 and $\{f_j : j \in J\}$ is an orthonormal basis of H_2. For any $x = \sum_{i \in I, j \in j} \lambda_{ij} e_i \otimes f_j \in H_1 \otimes H_2$,

$$\begin{aligned} \langle x, x \rangle &= \sum_{i \in I, j \in J} \langle \lambda_{ij} e_i \otimes f_j, \lambda_{ij} e_i \otimes f_j \rangle \\ &= \sum_{i \in I, j \in J} \lambda_{ij}^2 \geq 0, \end{aligned}$$

and clearly $\langle x, x \rangle = 0$ if and only if $x = 0$. Other properties of an inner-product can easily be checked.
Taking the completion of $H_1 \otimes H_2$ with $\langle \cdot, \cdot \rangle$ gives us a Hilbert space, and we denote it by $H_1 \hat{\otimes} H_2$.

Lemma 2.1.4. *The Hilbert-Schmidt tensor product is associative, i.e. if H_1, H_2 and H_3 are Hilbert spaces, then the natural mapping from $(H_1 \hat{\otimes} H_2) \hat{\otimes} H_3$ to $H_1 \hat{\otimes} (H_2 \hat{\otimes} H_3)$ is an isometry, where we equip the tensor product of two Hilbert spaces with the Hilbert-Schmidt inner-product.*

Proof. See Appendix A. □

We note the following important theorem which helps us to identify the dual space of a Hilbert space.

Theorem 2.1.1 (Riesz-Fréchet Theorem)**.** *Let H be a Hilbert space with inner-product $\langle \cdot, \cdot \rangle$, and let F be a continuous linear functional on H. There exists a unique $y \in H$ such that*

$$F(x) = \langle x, y \rangle \quad \forall x \in H.$$

Moreover, $\|F\| = \|y\|$.

Proof. For a proof see Theorem 6.8 of [33] by Young. □

Lemma 2.1.5. *Suppose H_1 and H_2 are Hilbert spaces with inner-product $\langle \cdot, \cdot \rangle_1$ and $\langle \cdot, \cdot \rangle_2$ respectively. Then for all $u \in H_1$, $v \in H_2$,*

$$\|u \otimes v\| = \|u\| \|v\|.$$

Moreover, for any $\phi \in H_1', \psi \in H_2'$,

$$\|\phi \otimes \psi\| = \|\phi\| \|\psi\|.$$

Proof. See Appendix A. □

It can then be deduced that the Hilbert-Schmidt norm is a reasonable tensor algebra norm.

Lemma 2.1.6. *The Hilbert-Schmidt norm is a reasonable tensor algebra norm.*

Proof. See Appendix A. □

2.2 Super-multiplicativity of the signature in reasonable tensor algebra norms

Let V be a Banach space, and $\gamma : J \to V$ be a continuous path of finite length over the compact interval J. The *signature* of γ is denoted by

$$S = (1, S_1, S_2, \cdots, S_n, \cdots), \tag{2.4}$$

where for each $n \geq 1$, $S_n = \int_{u_1 < \cdots < u_n, u_1, \cdots, u_n \in J} \mathrm{d}\gamma_{u_1} \otimes \cdots \otimes \mathrm{d}\gamma_{u_n}$.

Remark 2.2.1. *Note that the n-th term of S lives in the completed Banach space $V^{\otimes n}$ whenever the algebraic tensor product is completed with a reasonable tensor algebra norm.*

From now on we will fix a Banach space V, a reasonable tensor algebra norm, and we will take $V^{\otimes n}$ to be the completion of the algebraic tensor product with respect to that reasonable tensor algebra norm.

Theorem 2.2.1. *Suppose $\gamma : J \to V$ is a path of finite length. Then for $m, n \geq 0$, the signature of γ satisfies*

$$\|(m+n)! S_{m+n}\| \geq \|n! S_n\| \, \|m! S_m\| \quad \forall m, n \geq 0. \tag{2.5}$$

where $\|\cdot\|$ is any reasonable tensor algebra norm. $V^{\otimes 0}$ is defined to be $\mathbb{F}$, and $S_0 = 1$.

Proof. By Hahn-Banach Theorem, there exists $\phi_n \in (V^{\otimes n})'$, $\phi_m \in (V^{\otimes m})'$ such that $\|\phi_n\| = 1$, $\|\phi_m\| = 1$, and

$$\phi_n(S_n) = \|S_n\| \, , \ \phi_m(S_m) = \|S_m\| \, .$$

Equivalently, we can write

$$\phi_n(S) = \|S_n\| \, , \ \phi_m(S) = \|S_m\| \, ,$$

where we define $\phi_k(x) = 0$ for $x \notin V^{\otimes k}$ for all $k \geq 0$. From [23] we know that S is group-like, hence

$$\phi_m \sqcup\!\sqcup \, \phi_n(S) = \phi_m(S)\phi_n(S) = \|S_m\| \, \|S_n\|_.$$

Also,

$$\phi_m \sqcup\!\sqcup \, \phi_n(S_{m+n}) = \sum_{\sigma \in \text{shuffles}(m,n)} \sigma(\phi_m \otimes \phi_n)(S_{m+n})$$

$$= \sum_{\sigma \in \text{shuffles}(m,n)} (\phi_m \otimes \phi_n)(\sigma^{-1}(S_{m+n})),$$

so

$$|\phi_m \sqcup\!\sqcup\, \phi_n(S_{m+n})| \leq \#\text{shuffles}(m,n) \, \|\phi_m \otimes \phi_n\| \, \|S_{m+n}\| \, .$$

Note that $\#\text{shuffles}(m,n) = \frac{(m+n)!}{n!m!}$, and by Definition 2.1.3 we know that

$$\|\phi_m \otimes \phi_n\| \leq \|\phi_m\| \, \|\phi_n\| = 1.$$

Hence

$$\|(m+n)!S_{m+n}\| \geq \|n!S_n\| \, \|m!S_m\|$$

as expected. □

2.3 Limiting behaviour

Theorem 2.3.1 (Fekete's Lemma, see e.g. Lemma 1.2.1 of [30])**.** *If a sequence of real numbers* $\{a_n\}_{n\in\mathbb{N}}$ *satisfies the sub-additivity condition*

$$a_{m+n} \leq a_m + a_n \quad \forall m, n \in \mathbb{N},$$

Then

$$\lim_{n\to\infty} \frac{a_n}{n} = \inf_{n\in\mathbb{N}} \frac{a_n}{n}.$$

Definition 2.3.1 (Submonoids of $\mathbb{N}$)**.** *Let S be a subset of* $\mathbb{N}$*. Note we adopt the convention that* $0 \in \mathbb{N}$*. Then S is a* submonoid of $\mathbb{N}$ *if:*
(1). $0 \in S$*;*
(2). If $a, b \in S$*, then* $a + b \in S$*.*
Given a subset A of $\mathbb{N}$*, the* submonoid $\langle A \rangle$ generated by A *is defined as*

$$\langle A \rangle := \{\lambda_1 a_1 + \cdots + \lambda_r a_r : r \in \mathbb{N}, \lambda_i \in \mathbb{N}, a_i \in A \quad \forall i = 1, \cdots, r\} \, .$$

Definition 2.3.2 (Numerical semigroups)**.** *A submonoid S of* $\mathbb{N}$ *is called a* numerical semigroup *if* $\mathbb{N} \setminus S$ *is finite.*

Theorem 2.3.2. *Let M be a submonoid of* $\mathbb{N}$*. Then M has a unique minimal system of generators, which in addition is finite.*

Proof. For a proof see for example Corollary 2.8 of [28]. □

Lemma 2.3.1. *Assume A is a non-empty subset of $\mathbb{N}$. Let $d = gcd(A)$ be the greatest common divisor of the elements in A. Then $\langle d\rangle \setminus \langle A\rangle$ is finite.*

Proof. Note first that $\langle A\rangle \subset \langle d\rangle$. Define

$$S := \left\{\frac{n}{d} : n \in \langle A\rangle\right\}.$$

Then by [28], S is a numerical semigroup, so $\mathbb{N} \setminus S$ is finite, therefore $d\mathbb{N} \setminus \langle A\rangle$, i.e. $\langle d\rangle \setminus \langle A\rangle$, is finite. □

Lemma 2.3.2. *Assume $\gamma : J \to V$ is a continuous tree-reduced path of finite length. Define*

$$I := \{i \in \mathbb{N} : S_i \neq 0\}.$$

I is a submonoid of $\mathbb{N}$, therefore I is finitely generated.

Proof. Since $S_0 = 1$, $0 \in I$; Assume $m, n \in I$, then $S_m \neq 0$ and $S_n \neq 0$, so by super-multiplicativity,

$$\|(m+n)!S_{m+n}\| \geq \|m!S_m\| \, \|n!S_n\| > 0.$$

Hence $m + n \in I$. Therefore I is a submonoid of $\mathbb{N}$. Then by Theorem 2.3.2, I is finitely generated. □

For a tree-reduced continuous path $\gamma : J \to V$, by Lemma 2.3.2 there exist $n_1, \cdots, n_r \in \mathbb{N} \setminus \{0\}$ such that the set I of indices of the non-zero terms in the signature of γ is generated by the set

$$A := \{n_1, \cdots, n_r\}, \quad n_i \in \mathbb{N} \setminus \{0\} \quad \forall i = 1, \cdots, r.$$

Since γ is tree-reduced, by Hambly and Lyons [19] the signature of γ is non-trivial, and there must exist some $n \in \mathbb{N}\setminus\{0\}$ such that $S_n \neq 0$. Hence the set A is non-empty. We adopt these notations in the rest of this chapter unless otherwise specified.

Lemma 2.3.3. *Assume a tree-reduced continuous path $\gamma : J \to V$ is of finite length $L > 0$, and $A := \{n_1, \cdots, n_r\}$ is the generating set of the indices of all the non-zero terms of the signature of γ. Then for all $i = 1, \cdots, r$,*

$$\begin{aligned} &\lim_{n \in \langle n_i\rangle, n\to\infty} \|n!S_n\|^{1/n} \\ &= \sup_{n \in \langle n_i\rangle, n \geq 1} \|n!S_n\|^{1/n} \\ &=: \tilde{L} > 0. \end{aligned}$$

Proof. Define

$$f(n) := -\log(\|n!S_n\| / L^n).$$

Then by the super-multiplicativity of the signature,

$$f(m+n) \leq f(m) + f(n) \quad \forall\, m, n \in \langle n_i \rangle \quad i = 1, \cdots, r.$$

Define

$$\ell_i := \inf_{n \in \langle n_i \rangle, n \geq 1} \frac{f(n)}{n}.$$

Fix $\epsilon > 0$. Then there exists $K \in \langle n_i \rangle$ such that

$$\frac{f(K)}{K} \leq \ell_i + \epsilon.$$

Then for all $n \in < n_i >$ such that $n > K$, there exist $q \in \mathbb{N}$ and $s \in \langle n_i \rangle$, $s < K$ such that $n = qK + s$, hence

$$\begin{aligned}
\frac{f(n)}{n} &\leq \frac{qf(K) + f(s)}{qK + s} \\
&\leq \frac{qK}{qK + s}(\ell_i + \epsilon) + \frac{f(s)}{qK + s} \\
&\leq \ell_i + \epsilon + \frac{f(s)}{n}.
\end{aligned}$$

Note $f(s)$ is bounded by $\max_{j \in \langle n_i \rangle, j < K} f(j)$. Then as $n \to \infty$, we have

$$\limsup_{n \in \langle n_i \rangle, n \to \infty} \frac{f(n)}{n} \leq \ell_i + \epsilon.$$

Therefore for all $i = 1, \cdots, r$,

$$\lim_{n \in \langle n_i \rangle, n \to \infty} \frac{f(n)}{n} = \ell_i.$$

We also note that the submonoid of $\mathbb{N}$ generated by the least common multiple of A, $\langle \mathrm{lcm}(A) \rangle$, is a common infinite subsequence of $\langle n_i \rangle$ for all $i = 1, \cdots, r$. Hence $\ell_i = \ell_j := \ell$ for all $i, j = 1, \cdots, r$. Then for all $i = 1, \cdots, r$,

$$\begin{aligned}
&\lim_{n \in \langle n_i \rangle, n \to \infty} \|n!S_n\|^{1/n} \\
&= \sup_{n \in \langle n_i \rangle, n \geq 1} \|n!S_n\|^{1/n} \\
&= \tilde{L}.
\end{aligned}$$

Because $\langle n_i \rangle$ are indices of non-zero terms in the signature, $\tilde{L} > 0$. □

Remark 2.3.1. *Note that Lemma 2.3.3 in fact used the fact that Fekete's Lemma holds on $\langle n\rangle$ for all $n \in \mathbb{N}\setminus\{0\}$.*

Theorem 2.3.3 (Asymptotic behaviour of the signature)**.** *Assume a tree-reduced continuous path $\gamma : J \to V$ is of finite length $L > 0$, and $A := \{n_1, \cdots, n_r\}$ is the generating set of the indices of all the non-zero terms of the signature of γ. Let I denote the set of indices of all non-zero terms in the signature of γ. Then*

$$\begin{aligned}&\lim_{n\in I, n\to\infty} \|n!S_n\|^{1/n}\\ &= \sup_{n\in I, n\geq 1} \|n!S_n\|^{1/n} = \sup_{n\geq 1} \|n!S_n\|^{1/n}\\ &:= \tilde{L} > 0.\end{aligned}$$

Proof. Note $I = \langle A\rangle = \langle n_1, \cdots, n_r\rangle$. Define $f(n) := -\log(\|n!S_n\|/L^n)$. Then

$$f(m+n) \leq f(m) + f(n) \quad \forall\, m, n \in I.$$

By Lemma 2.3.3, we can define

$$\ell := \inf_{n\in\langle n_i\rangle, n\geq 1} \frac{f(n)}{n} \quad \forall\, i = 1, \cdots, r.$$

Fix $\epsilon > 0$. Again by Lemma 2.3.3 there exists $N \in \mathbb{N}$ such that for all $\lambda_i \geq N$,

$$\ell \leq \frac{f(\lambda_i n_i)}{\lambda_i n_i} \leq \ell + \epsilon \quad \forall i = 1, \cdots, r.$$

For all $n > N\sum_{i=1}^r n_i$ and $n \in I$, there exist $\lambda_1, \cdots \lambda_r \in \mathbb{N}$ such that $n = \sum_{i=1}^r \lambda_i n_i$, and

$$\begin{aligned} f(n) = f\left(\sum_{i=1}^r \lambda_i n_i\right) &= f\left(\sum_{\lambda_i\geq N} \lambda_i n_i + \sum_{\lambda_i<N} \lambda_i n_i\right)\\ &\leq f\left(\sum_{\lambda_i\geq N} \lambda_i n_i\right) + f\left(\sum_{\lambda_i<N} \lambda_i n_i\right)\\ &\leq (\ell+\epsilon)\left(\sum_{\lambda_i\geq N} \lambda_i n_i\right) + f\left(\sum_{\lambda_i<N} \lambda_i n_i\right).\end{aligned}$$

Then

$$\frac{f(n)}{n} \leq (\ell+\epsilon)\frac{\sum_{\lambda_i\geq N}\lambda_i n_i}{n} + \frac{f\left(\sum_{\lambda_i<N}\lambda_i n_i\right)}{n}.$$

Note that $f\left(\sum_{\lambda_i<N}\lambda_i n_i\right)$ is bounded above by $rN\max_{i=1,\cdots,r} f(n_i)$. Then as n goes to infinity,

$$\lim_{n\in I, n\to\infty}\frac{f(n)}{n}\le \ell+\epsilon.$$

By a similar argument as in the proof of Lemma 2.3.3, we have, for any $k\in I$ such that $k\ge 1$,

$$\lim_{n\in\langle k\rangle}\frac{f(n)}{n}=\inf_{n\in\langle k\rangle, n\ge 1}\frac{f(n)}{n}=\ell.$$

Therefore for all $k\in I\setminus\{0\}$,

$$\frac{f(k)}{k}\ge\ell.$$

Also note for any $\eta>0$, for any $k\in I\setminus\{0\}$, there exists $m\in\langle k\rangle\subset I$ such that

$$\frac{f(m)}{m}\le\ell+\eta,$$

hence

$$\ell=\inf_{n\in I, n\ge 1}\frac{f(n)}{n}.$$

Combining with the inequality we found for the limit of the sequence $(f(n)/n)$ for $n\in I$, we have

$$\ell\le\lim_{n\in I, n\to\infty}\frac{f(n)}{n}\le\ell+\epsilon,$$

which implies

$$\lim_{n\in I, n\to\infty}\frac{f(n)}{n}=\inf_{n\in I, n\ge 1}\frac{f(n)}{n},$$

and the result follows. $\square$

Remark 2.3.2. *By the result of Lemma 2.3.1, the signature of a tree-reduced path of finite length can only have finitely many zeros at levels which are multiples of the greatest common factor of $n_1,\cdots,n_r$. It is in fact a proof by Boedihardjo and Geng [4] that there can only be finitely many zero terms in the signature of a tree-reduced path of finite length, or in the language of Lemma 2.3.1, the greatest common divisor of $n_1,\cdots,n_r$ is 1, and we can apply the idea of Fekete's lemma and show that $\lim_{n\to\infty}\|n!S_n\|^{1/n}=\sup_{n\ge 1}\|n!S_n\|^{1/n}$, and the limit is non-zero.*

Corollary 2.3.1. *Let V be a Banach space. For any element*

$$\mathbf{a} = (a_0, a_1, a_2, \cdots) \in \left\{(b_0, b_1, b_2, \cdots) : b_0 = 1, b_n \in V^{\otimes n} \quad \forall n \geq 1\right\}$$

which is group-like, we have

$$\|(m+n)! a_{m+n}\| \geq \|m! a_m\| \, \|n! a_n\| \quad \forall m, n \geq 0.$$

Define $I := \{n \in \mathbb{N} : a_n \neq 0\}$. Then

$$\lim_{n \in I, n \to \infty} \|n! a_n\|^{1/n} = \sup_{k \in I, k \geq 1} \|k! a_k\|^{1/k}$$

under any reasonable tensor algebra norm $\|\cdot\|$.

Proof. Note that since $\mathbf{a}$ is group-like, the same arguments apply as in Theorem 2.2.1 and Theorem 2.3.3. $\square$

Remark 2.3.3. *It is an interesting question to ask whether there is a nice and simple form of the limit of $\|n! S_n\|^{1/n}$ mentioned in Theorem 2.3.3, and whether the limit is the same under any reasonable tensor algebra norm. Moreover, we know from [23] that for a path of finite length $L > 0$, an upper bound of $\|n! S_n\|$ is L^n. Furthermore, Hambly and Lyons (Theorem 9, [19]) proved that for a smooth enough path of finite length, the ratio $\|n! S_n\| / L^n$ converges to 1 under certain norms. Therefore we have the following conjecture.*

Conjecture 2.3.1. *Let V be a Banach space, and $\gamma : J \to V$ be a path of finite length $L > 0$. Then the signature of γ satisfies that*

$$\|n! S_n\|^{1/n} \to L \quad \text{as } n \to \infty,$$

under any reasonable tensor algebra norm .

Example 2.3.1. *Let us consider a monotone lattice path γ consisting of two linear pieces, and each piece is of length $\frac{1}{2}$. Under ℓ^1 norm, we have*

$$\|n! S_n\|_1 = \sum_{k=0}^{n} \binom{n}{k} \left(\frac{1}{2}\right)^k \left(\frac{1}{2}\right)^{n-k} = 1,$$

hence apparently $\lim_{n\to\infty} \|n! S_n\|_1^{1/n} = 1$, which is the length of the path.
If we consider the case under the Hilbert-Schmidt norm, we have

$$\|n! S_n\|_{HS}^{\frac{1}{n}} = \left(\sum_{k=0}^{n} \binom{n}{k}^2 \left(\frac{1}{2}\right)^{2k} \left(\frac{1}{2}\right)^{2(n-k)}\right)^{\frac{1}{2n}}.$$

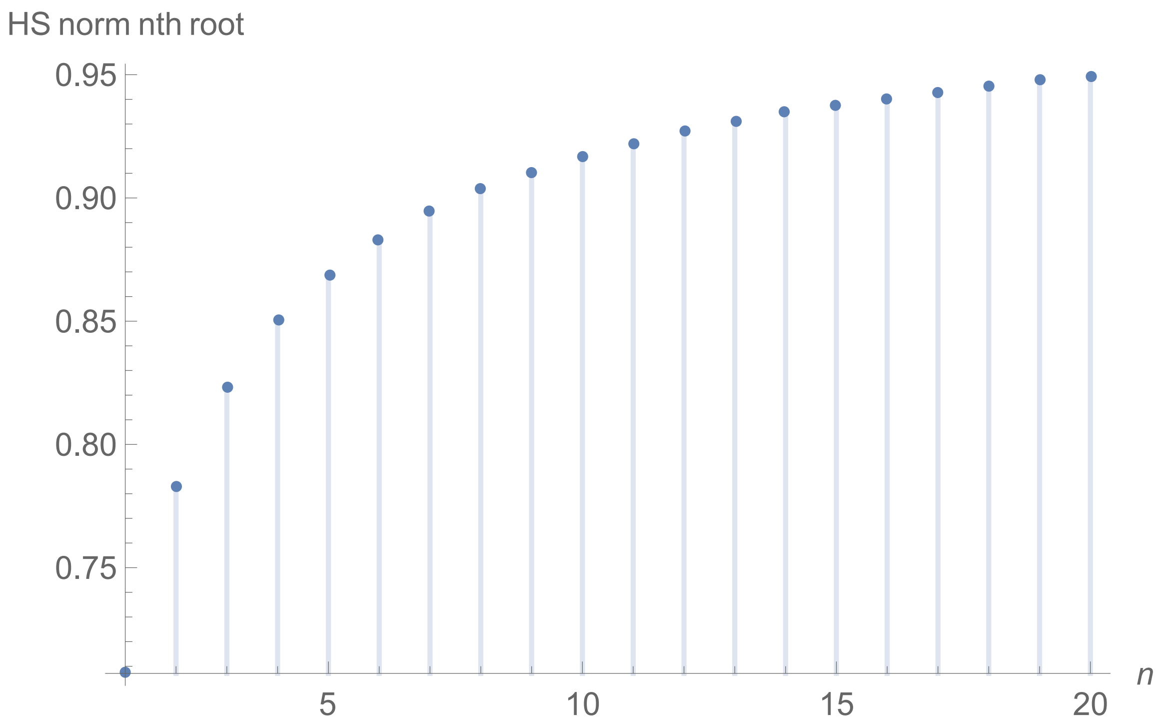


Figure 2.1: $\|n!S_n\|_{\mathrm{HS}}^{1/n}$ against the signature level n of a monotone lattice path

If we plot $\|n!S_n\|_{HS}^{1/n}$ against n, we obtain Figure 2.1, from which we can see that it is reasonable to assume $\|n!S_n\|_{HS}^{1/n}$ converges to the length of the path. Therefore both cases match our expectation as described in Conjecture 2.3.1.

Remark 2.3.4. *An interesting tensor norm to consider is the* Haagerup tensor norm *[3]. Clearly the Haagerup norm is not a reasonable tensor algebra norm, however under the Haagerup norm, for a path of finite length $L > 0$, we still have $n! \|S_n\| \leq L^n$. Therefore it is an interesting question to ask whether the signature will have the same behaviour as described in Theorem 2.3.3 under the Haagerup tensor norm, or the symmetrised forms of the Haagerup tensor norm.*

Remark 2.3.5. *Although it has been shown that $\|n!S_n\|$ eventually behaves like L^n under certain norms for paths which are smooth enough (see [19]), some simple examples show that in general for a path of finite length, $\|n!S_n\| / L^n$ does not necessarily converge to 1 as n increases. The result in Theorem 2.3.3 is the best description we have found so far about the decay of the signature of a path of finite length. Further discussions can be found in Section 4.5.*

Given the result for paths of finite length, it is then interesting to discuss whether a non-zero limit exists for p-variation paths for $p > 1$. However it is not the case as demonstrated in the following corollary.

Corollary 2.3.2. *Suppose $p \in \mathbb{R}$ and $p > 1$. Then for a path γ of finite p-variation, there does not exist a non-zero limit for $\|(n/p)!S_n\|^{1/n}$ under a reasonable tensor algebra norm.*

Proof. Consider a linear path γ of length 1. Note γ is a p-variation path for all $p > 1$. Then for all $p > 1$,

$$\left\| \left(\frac{n}{p}\right)! S_n \right\|^{\frac{1}{n}} = \left(\frac{\left(\frac{n}{p}\right)!}{n!} \right)^{\frac{1}{n}},$$

which converges to zero as n tends to infinity. Hence we cannot deduce a non-zero limit for a p-variation path in general. □

Chapter 3

Signature inversion using symmetrisation

From Lemma 1.4.1 we see that we can connect the symmetric part of a term in the signature with the total increment of the path. In this chapter, we introduce the method of *symmetrisation* by Lyons and Xu [25], which recovers the increments over the subintervals. We then show that the method can lead to practical algorithms for inverting monotone paths, which was jointly done with Duffield, Ni and Xu [8].

3.1 Symmetrisation

Lyons and Xu [25] first introduced the symmetrisation method of inverting the signature of a C^1 path. As an example, consider a path

$$\gamma : [0,1] \to \mathbb{R}^2, \quad \gamma_t = (x_t, y_t) \quad \forall t \in [0,1],$$

with continuous derivative $\dot{\gamma}_t = (\dot{x}_t, \dot{y}_t)$. Let Δ_n denote the simplex

$$\Delta_n := \{(t_1, t_2, \cdots, t_n) : 0 < t_1 < \cdots < t_n < 1\}.$$

Given $\mathbf{u} \in \Delta_n$, for every $j = 1, \cdots, k$, define

$$\Delta_{\mathbf{u},j} x = x_{u_j} - x_{u_{j-1}}, \quad \Delta_{\mathbf{u},j} y = y_{u_j} - y_{u_{j-1}}$$

to be the increment of γ over the time interval $[u_{j-1}, u_j] \subset [0,1]$. Suppose $\{e_1, e_2\}$ is a basis of $\mathbb{R}^2$, where e_1 is in the x-direction, and e_2 is in the y-direction. Then for $m \geq 1$, elements of $(\mathbb{R}^d)^{\otimes m}$ can be written as linear combinations of words $\omega = e_{i_1} \cdots e_{i_m}$. For each word ω, let $|\omega|_x$ denote the number of letter e_1's in ω, and $|\omega|_y$ denote the number of e_2's in ω. For integers $n \geq 0$ and $k \geq 0$, define

$$\mathbf{L}_k^n := \{(\ell_1, \cdots, \ell_k) : 0 \leq \ell_j \leq n, 1 \leq j \leq k\}.$$

For every $\boldsymbol{\ell} \in \mathbf{L}_k^n$ and every word $\omega = e_{i_1} \cdots e_{i_{k-1}}$, define

$$\mathbf{W}_k^{2n}(\omega, \boldsymbol{\ell}) = \left\{ \omega_1 * e_{i_1} * \cdots * e_{i_{k-1}} * \omega_k : |\omega_j|_x = 2\ell_j, |\omega_j|_y = 2n - 2\ell_j \right\},$$

where '$*$' denotes the concatenation of words. For $\omega' = e_{i_1} \cdots e_{i_m}$, define

$$C(\omega') = \int_{0<u_1<\cdots<u_m<1} \mathrm{d}\gamma_{u_1}^{i_1} \cdots \mathrm{d}\gamma_{u_m}^{i_m},$$

and

$$\mathcal{S}_k^{2n}(\omega, \boldsymbol{\ell}) = ((2n)!)^k \sum_{\omega' \in \mathbf{W}_k^{2n}(\omega, \boldsymbol{\ell})} C(\omega').$$

Proposition 3.1.1 (Lyons and Xu [25]). *Fix integer n and k. Let $\omega = e_{i_1} \cdots e_{i_{k-1}}$ and $\boldsymbol{\ell} = (\ell_1, \cdots, \ell_k) \in \mathbf{L}_k^n$. Then we can write*

$$\mathcal{S}_k^{2n}(\omega, \boldsymbol{\ell}) = \int_{\Delta_{k-1}} \prod_{j=1}^{k-1} \dot{\gamma}_{u_j} \prod_{j=1}^{k} \binom{2n}{2\ell_j} (\Delta_{\mathbf{u},j} x)^{2\ell_j} (\Delta_{\mathbf{u},j} y)^{2n-2\ell_j} \, \mathrm{d}\mathbf{u}.$$

Proof. See Appendix A. □

We can see that symmetrisation gives the piecewise increments over the subintervals. Lyons and Xu [25] used symmetrisation method to reconstruct C^1 paths, in the rest of this chapter we will show that the symmetrisation method can be used to recover monotone paths. The rest of this chapter was jointly done with Duffield, Ni and Xu [8], and the original article includes a very interesting description about the connection between the symmetrisation method and the large deviation principle, here we mainly describe the algorithm developed in the paper [8].

3.2 Probabilistic interpretation of the signature of a monotone path

Let $\{e_1, \cdots, e_d\}$ denote the standard basis of $\mathbb{R}^d$. For every integer $n \geq 0$, a word of length n is an ordered sequence of n letters from the set $\{e_1, \cdots, e_d\}$, where repetition is allowed. For a word ω, let $|\omega|$ denote the number of letters in ω. For two words $\omega_1 = e_{i_1} \cdots e_{i_n}$ and $\omega_2 = e_{j_1} \cdots e_{j_m}$, their concatenation $\omega_1 * \omega_2$ is a new word of length $n + m$ defined by

$$\omega_1 * \omega_2 = e_{i_1} \cdots e_{i_n} e_{j_1} \cdots e_{j_m}.$$

We use $\emptyset$ to denote the empty word, which is the unique word of length 0. Then we can write down the signature of a bounded-variation path coordinate-wise:

Definition 3.2.1. *Let $\gamma : [0,1] \to \mathbb{R}^d$ be a continuous path of bounded variation. For every integer $n \geq 0$ and every word $\omega = e_{i_1} \cdots e_{e_n}$, let*

$$C_\gamma(\omega) := \int_{0<u_1<\cdots<u_n<1} \mathrm{d}\gamma^{i_1}(u_1) \cdots \mathrm{d}\gamma^{i_n}(u_n),$$

where γ^i is the i-th component of γ. The signature of γ is the formal power series

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

where the second sum is taken over all words of length n, and we have set $C_\gamma(\emptyset) = 1$ by convention.

We use ℓ^1 norm from now on in this chapter. Note that the collection

$$\{C_\gamma(\omega) : |\omega| = n\}$$

is the n-th level coefficients in the signature.
In the rest of this chapter we assume $\gamma : [0,1] \to \mathbb{R}^d$ is a monotone path parametrised at unit speed, that is, $\|\dot{\gamma}(t)\|_1 = 1$ for all $t \in (0,1)$. Note in this case we can write $C_\gamma(\omega)$ as

$$C_\gamma(\omega) = \int_{0<u_1<\cdots<u_n<1} \dot{\gamma}^{i_1}(u_1) \cdots \dot{\gamma}^{i_n}(u_n)\mathrm{d}u_1 \cdots \mathrm{d}u_n.$$

Note if γ is decreasing in any of its components, we can reflect γ^i to $-\gamma^i$ in that component and the corresponding change in the signature is immediate. Therefore without loss of generality we can assume that γ is monotonically increasing, that is

$$\dot{\gamma}^i(t) \geq 0 \quad \forall i = 1, \cdots, d.$$

Under ℓ^1 norm we have

$$\sum_{i=1}^{d} \dot{\gamma}^i(u) = 1 \quad \forall u \in (0,1)$$

since γ is parametrised at unit speed.
Since γ is monotonically increasing, we have $C_\gamma(\omega) \geq 0$ for every word ω. Note that for any integer $N \geq 0$,

$$\sum_{|\omega|=N} C_\gamma(\omega) = \sum_{i_1 \cdots, i_N \in \{1, \cdots, d\}} \int_{0<u_1<\cdots<u_N<1} \mathrm{d}\gamma^{i_1}(u_1) \cdots \mathrm{d}\gamma^{i_N}(u_N)$$

$$
\begin{aligned}
&= \sum_{i_1,\cdots,i_N\in\{1,\cdots,d\}} \int_{0<u_1<\cdots<u_N<1} \dot\gamma^{i_1}(u_1)\cdots\dot\gamma^{i_N}(u_N)\mathrm{d}u_1\cdots\mathrm{d}u_N \\
&= \int_{0<u_1<\cdots<u_N<1} \prod_{k=1}^{N}\sum_{i=1}^{d}\dot\gamma^{i}(u_k)\mathrm{d}u_1\cdots\mathrm{d}u_N \\
&= \int_{0<u_1<\cdots<u_N<1} \mathrm{d}u_1\cdots\mathrm{d}u_N \\
&= \frac{1}{N!}.
\end{aligned}
$$

This suggests that for every N, the quantities $\{N!C_\gamma(\omega) : |\omega| = N\}$ constitute a probability measure on the words of length N, giving each word ω with $|\omega| = N$ the 'probability' $N!C_\gamma(\omega)$. Therefore we can try to give the probabilistic interpretation of the signature of a monotone path. For such a purpose, we introduce the definition of *Poisson processes.*

Definition 3.2.2 (Poisson processes). *A counting process $\{N(t), t \geq 0\}$ is called a* Poisson process with rate $\lambda(t) > 0$ *if*

1. *$N(0) = 0$;*
2. *$N(t)$ has independent increments;*
3. *$N(t) - N(s) \sim$ Poisson($\int_{s<u<t} \lambda(u)\mathrm{d}u$) for $t > s$.*

Let $\gamma : [0, 1] \to \mathbb{R}^d$ be a monotone path parametrised at unit speed with respect to ℓ^1 norm. Consider d independent Poisson processes running simultaneously on the time interval $[0, 1]$ respectively. The intensity for the i-th coordinate component of this Poisson process is $\dot\gamma^i(t)$. Let $W(t)$ be the word of ordered letters that arrive up to time t. For example, if at times $0 < u_1 < \cdots < u_5 < t$, the letters e_3, e_2, e_2, e_1, e_3 arrive, then

$$
W(t) = e_3e_2e_2e_1e_3, \quad W(v) = e_3e_2,\ v \in [u_2, u_3).
$$

Suppose the arrival times are τ_j for $j = 1, 2, \cdots$ with the convention that $\tau_0 = 0$, then W can be defined as a lattice path by setting $W(\tau_0) = 0$, and for $t \in (\tau_j, \tau_{j+1}]$,

$$
W(t) = W(\tau_j) + \frac{t-\tau_j}{\tau_{j+1}-\tau_j}e_{i_j}, \tag{3.1}
$$

where e_{i_j} is the arriving letter at time τ_j. Therefore W is a lattice path we create that is associated with γ.

Now we want to consider the law of W conditional on the total number of arrivals up to time 1. Let $\mathcal{N}(t)$ be the process counting the total number of arrivals up to time t. Since γ is parametrised at unit speed, $\mathcal{N}(t)$ is a homogeneous Poisson process on $[0,1]$ with intensity 1.

Now we condition on the event $\mathcal{N}(1)=N$, that is, there are totally N arrivals up to times 1. Let

$$\mathbb{P}^N(\cdot) := \mathbb{P}(\cdot|\mathcal{N}(1)=N)$$

denote the conditional probability. Thus, for every word ω with $|\omega|=N$, with the abuse of notation that W denotes the word generated by the processes, we have

$$C_\gamma(\omega) = \frac{1}{N!}\mathbb{P}^N(W=\omega) = \frac{1}{N!}\mathbb{P}(W=\omega|\mathcal{N}(1)=N).$$

Define $W_N(t)=\frac{1}{N}W(t)$, where $W(t)$ is as defined in Equation (3.1). We now would like to parametrise the path W_N at unit speed. Recall that $\mathcal{N}(t)$ records the number of letters that have arrived until t for $t\in[0,1]$. We still condition on $\mathcal{N}(1)=N$. We now try to normalise the interval $[0,N]$ to $[0,1]$.

For $j=1,\cdots,N$, let $\tau_j\in[0,1]$ denote the arrival time of the j-th letter in the process W, hence

$$\mathcal{N}(t)=j \quad \text{if } t\in[\tau_j,\tau_{j+1}).$$

Define $T_N:[0,1]\to[0,\tau_N]$ such that

$$T_N(q) := \tau_j + (Nq-j)\left(\tau_{j+1}-\tau_j\right),\ q\in(\frac{j}{N},\frac{j+1}{N}].$$

We can see that $T_N(\cdot)$ maps $\frac{j}{N}$ to the arrival time τ_j of the j-th letter, and linear interpolates in between. Hence T_N is a strictly increasing function, therefore we can find the inverse $Q_N:[0,\tau_N]\to[0,1]$ defined by

$$Q_N(t) = \frac{j}{N} + \frac{t-\tau_j}{N(\tau_{j+1}-\tau_j)},\quad t\in(\tau_j,\tau_{j+1}]. \tag{3.2}$$

Thus we can re-parametrise the path W_N by considering

$$X_N = W_N\circ T_N : [0,1]\to\left(\frac{\mathbb{Z}}{N}\right)^d,\quad q\mapsto W_N\left(T_N(q)\right). \tag{3.3}$$

Hence X_N is the lattice path W_N re-parametrised at unit speed: for $\tau_j<t<\tau_{j+1}$,

$$\begin{aligned}\dot{X}_N(t) &= \dot{W}_N\left(T_N(t)\right)\dot{T}_N(t)\\ &= N\left(\tau_{j+1}-\tau_j\right)\frac{1}{N}\frac{1}{\tau_{j+1}-\tau_j}\end{aligned}$$

$$= 1.$$

For every integer $N \geq 0$ and $k \geq 0$, let $\mathcal{P}_{N,k}$ denote the set of k-partitions of N, i.e.

$$\mathcal{P}_{N,k} = \left\{ \mathbf{n} = (n_1, \cdots, n_k) : n_j > 0, \sum_{j=1}^{k} n_j = N \right\}.$$

For $\mathbf{n} \in \mathcal{P}_{N,k}$, let

$$L_k^{\mathbf{n}} = \left\{ \boldsymbol{\ell} = (\boldsymbol{\ell}_1, \cdots, \boldsymbol{\ell}_k) : \boldsymbol{\ell}_j = (\ell_j^1, \cdots, \ell_j^d), \sum_{i=1}^{d} \ell_j^i = n_j, \forall j = 1, \cdots, k \right\}.$$

Now for $\mathbf{n} \in \mathcal{P}_{N,k}$ and $\boldsymbol{\ell} \in \mathcal{L}_k^{\mathbf{n}}$, define the set of words $\mathcal{W}_k^{\mathbf{n}}(\boldsymbol{\ell})$ by

$$\mathcal{W}_k^{\mathbf{n}}(\boldsymbol{\ell}) = \left\{ \omega = \omega_1 * \cdots * \omega_k : |\omega_j|_{e_i} = \ell_j^i, \forall i = 1, \cdots, d, j = 1, \cdots, k \right\},$$

where $|\omega_j|_{e_i}$ denotes the number of the letter e_i in the word ω_j. Now for a given length of the words, if we consider all the words in which the number of each letter is fixed, we are 'symmetrising' the signature. We can define symmetrised signatures by

$$\mathcal{S}_k^{\mathbf{n}}(\boldsymbol{\ell}) := N! \sum_{\omega \in \mathcal{W}_k^{\mathbf{n}}(\boldsymbol{\ell})} C(\omega). \tag{3.4}$$

That is, $\mathcal{S}_k^{\mathbf{n}}(\boldsymbol{\ell})$ is the sum of the coefficients of all words ω such that $\omega = \omega_1 * \cdots * \omega_k$, and the number of letters in ω_j is ℓ_j^i. By recalling the random word W generated by the Poisson process associated to the monotone path γ, we have

$$\mathbb{P}^N(W \in \mathcal{W}_k^{\mathbf{n}}(\boldsymbol{\ell})) = \mathcal{S}_k^{\mathbf{n}}(\boldsymbol{\ell}).$$

So each $\mathcal{W}_k^{\mathbf{n}}$ corresponds to a random piecewise linear path composed of k pieces, each piece is a d-dimensional vector ω_j for $j = 1, .., k$.

3.3 A numerical example

Now we consider a monotone path $\gamma : [0,1] \to \mathbb{R}^2$. From above we know that for a random word W,

$$\mathbb{P}^N(W \in \mathcal{W}_k^{\mathbf{n}}(\boldsymbol{\ell})) = \mathcal{S}_k^{\mathbf{n}}(\boldsymbol{\ell}).$$

Let $\{x, y\}$ be the alphabet in the 2-dimensional case. We know from before that each $\mathcal{W}_k^{\mathbf{n}}$ corresponds to a random piecewise linear path, which has k pieces, and the number of letters in each piece is $|\omega_j|_x + |\omega_j| = n_j$ for $j = 1, \cdots, k$. Since the path γ

is parametrised at unit speed, each piece is of length $\frac{1}{k}$. If we denote the increments in the two directions over the j-th piece by Δx_j and Δy_j respectively for $j = 1, \cdots, k$, then

$$\mathbb{P}^N\left(\Delta x_j = \frac{m}{n_j}\frac{1}{k}, \Delta y_j = \frac{n_j - m}{n_j}\frac{1}{k}\right) = \frac{\sum_{\mathbf{n}=(n_1,\cdots,n_k),|\omega_j|_x=m} \mathcal{S}_k^{\mathbf{n}}(\boldsymbol{\ell})}{\sum_{\mathbf{n}=(n_1,\cdots,n_k)} \mathcal{S}_k^{\mathbf{n}}(\boldsymbol{\ell})} \tag{3.5}$$

for $m = 0, \cdots, n_j$. Therefore we can compute the probabilities of all possible combinations of Δx_j and Δy_j over the j-th piece for $j = 1, .., k$ and construct the piecewise linear path using the averaged increments.

We now consider the path $\gamma : [0, 1] \to \mathbb{R}^2$ such that

$$\gamma_t \mapsto (x_t, y_t), y_t = x_t^3 \quad \forall t \in [0, 1],$$

and we parametrise γ at unit speed.

We compute the probabilities of different combinations of increments according to Equation (3.5). We illustrate the results in two ways.

First we fix the length of each piece and vary the number of partitions: Let $n := n_j = 2$ for $j = 1, \cdots, k$, and vary $k = 4, 5, 6$. Note in this case we require information from the (nk)-th level of the signature. If we plot the probabilities as matrices, as shown in Figure 3.1, 3.2 and 3.3, we can see that the probability matrices characterise the underlying path in the expected way, and as we take finer partitions, the approximation becomes more accurate.

Instead we can fix the total number of partitions and vary the length of each piece: Fix $k = 2$ and vary $n = n_j = 1, \cdots, 5$ for $j = 1, \cdots, k$. Again we see from Figure 3.4, 3.5, 3.6, 3.7 and 3.8 that using higher levels of the signature gives us better characterisation.

If we plot the approximation paths obtained by taking the expectations of the increments over each piece, we again see from Figure 3.9 and 3.10 that we get better approximation results if we use higher levels of the signature. Note the computation of the signature in this example is done via the C++ package *Libalgebra* [6] which was developed through a project led by Lyons.

The key idea of the code for generating the approximation paths is that, given the alphabet A and a fixed proportion m of increments of δx_j, we define a function `AllPermutations`(A, j, m) to generate all the words which have m letters in the j-th block in the direction of x_j. Then we define a function `SigCoeff` which takes the coefficients in front of the words included in `AllPermutations`(A, j, m), as follows

```
def SigCoeff(A,j,m):
    sigcoeff=0
```

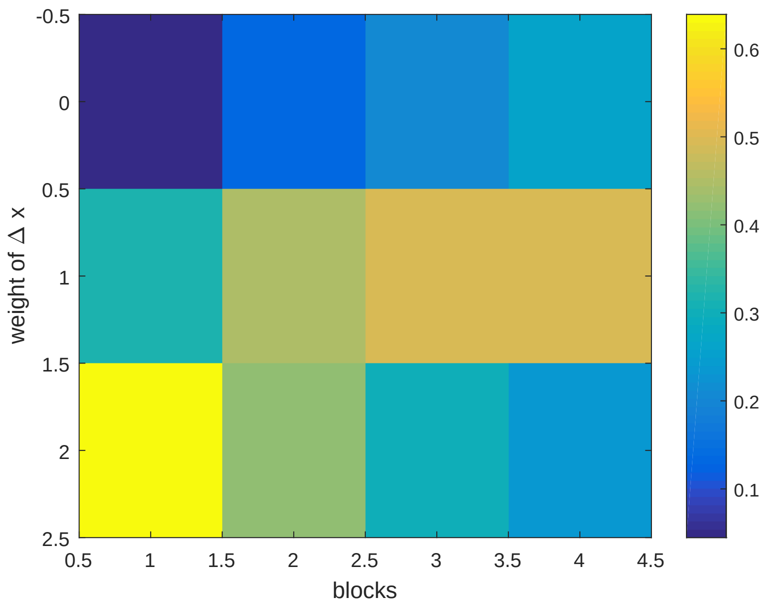


Figure 3.1: Probability matrix where $n = 2, k = 4$

```
for w in AllPermutations(A,j,m)
    sigcoeff+=Signature(w)
return sigcoeff.
```

With the function above it is easy to compute the probability of a given increment over a block, therefore the expectation can be computed.

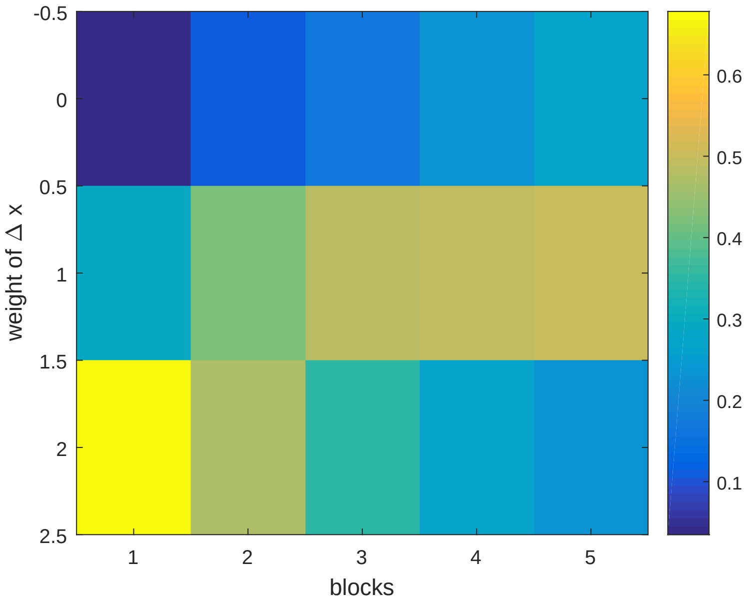


Figure 3.2: Probability matrix where $n = 2, k = 5$

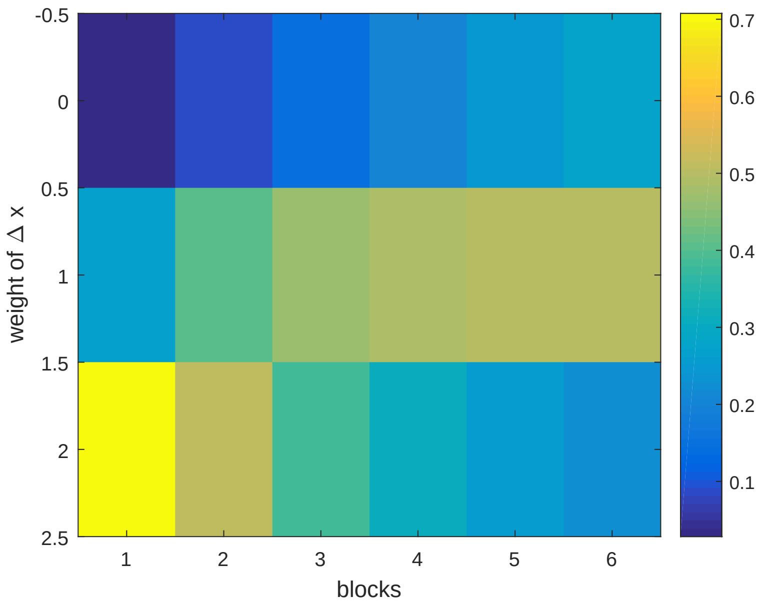


Figure 3.3: Probability matrix where $n = 2, k = 6$

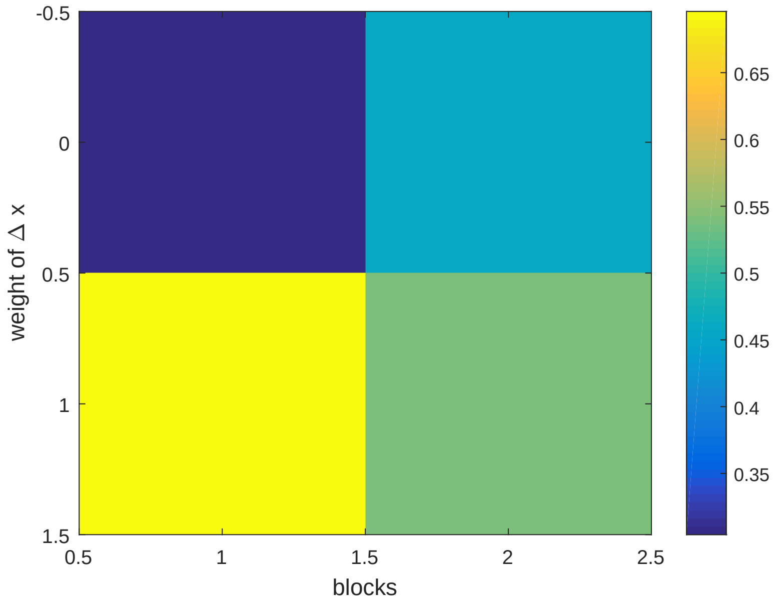


Figure 3.4: Probability matrix where $n = 1, k = 2$

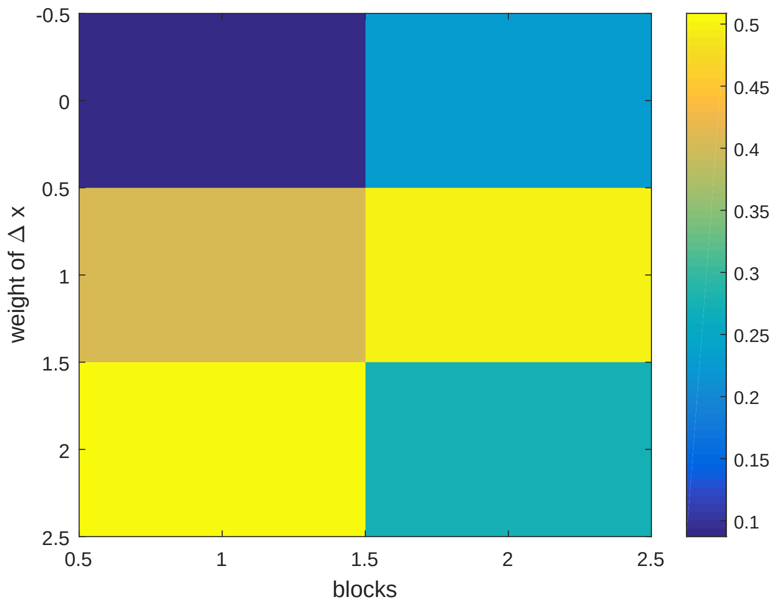


Figure 3.5: Probability matrix where $n = 2, k = 2$

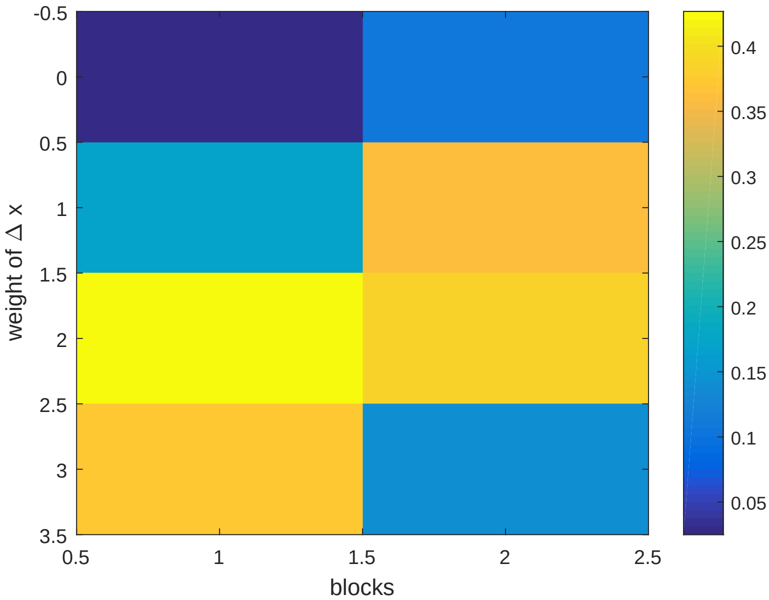


Figure 3.6: Probability matrix where $n = 3, k = 2$

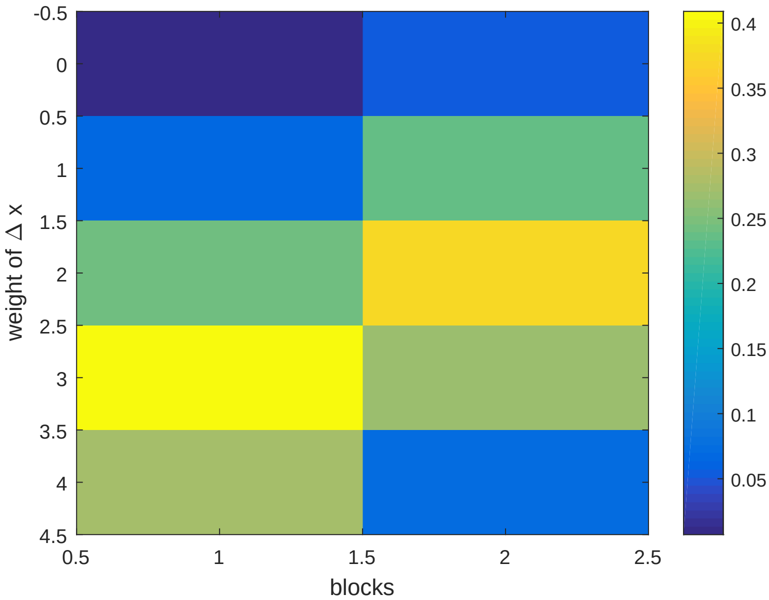


Figure 3.7: Probability matrix where $n = 4, k = 2$

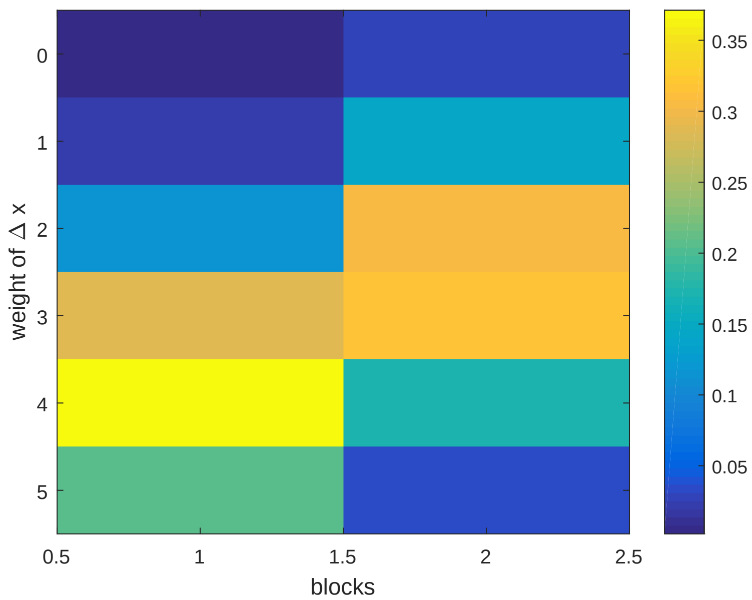


Figure 3.8: Probability matrix where $n = 5, k = 2$

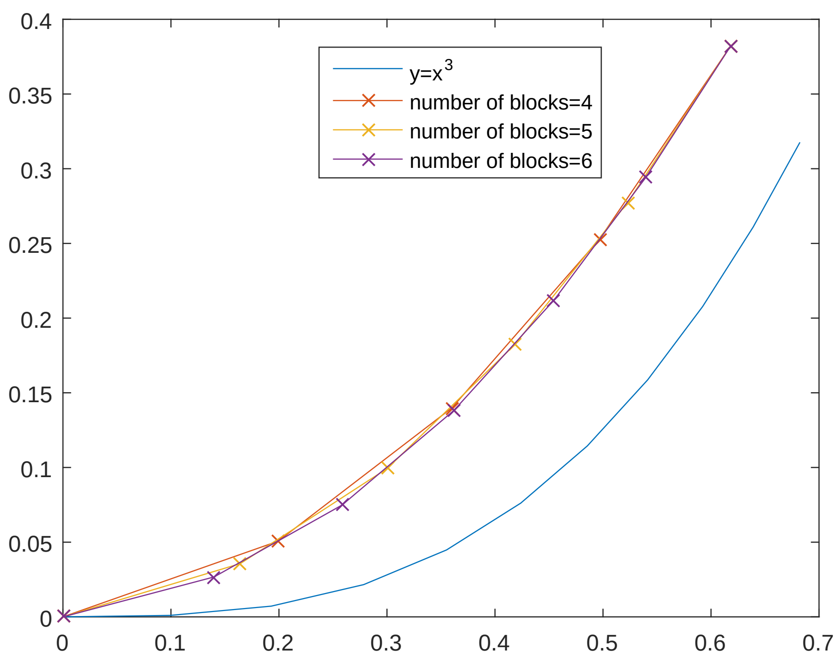


Figure 3.9: Reconstructed paths with $n = 2$ and $k = 4, 5, 6$

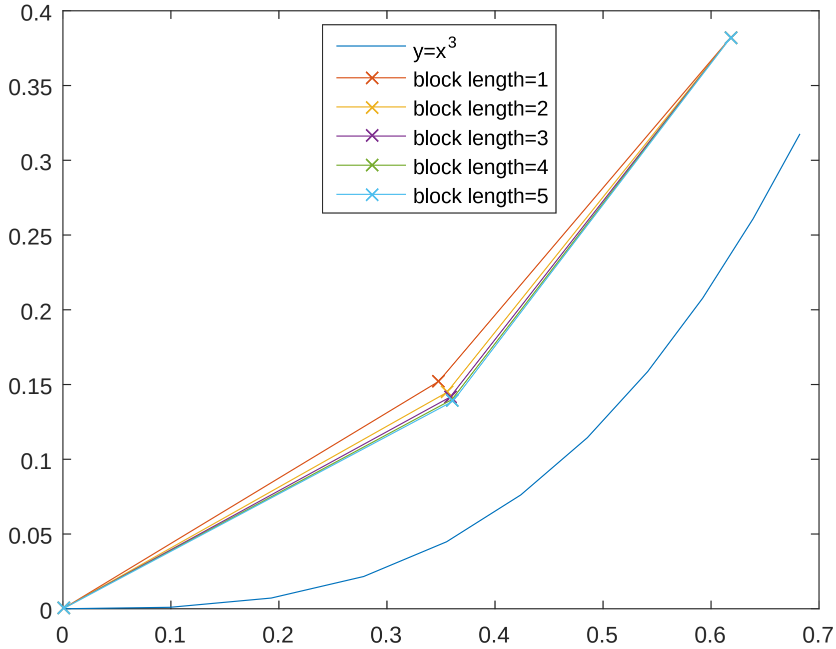


Figure 3.10: Reconstructed paths with $k = 2$ and $n = 1, 2, 3, 4, 5$

Chapter 4

Signature inversion using insertion

In this chapter, we develop a new method of inverting the signature of the path by trying to approximate a level of the signature by a lower level of signature. We justify the motivation of the method by considering the signatures of simple paths, and then we illustrate how we can approximate a path by solving an optimisation problem and demonstrate the method for a particular set of paths. In this chapter we mainly consider paths of unit length parametrised at unit speed.

4.1 Introductory examples

Consider a path $\gamma : [0,T] \to \mathbb{R}^d$. If γ is linear, the signature of γ at level n is $S^n_{0,T}(\gamma) = \frac{(\gamma_T - \gamma_0)^{\otimes n}}{n!}$ for $n \in \mathbb{N}$, so we have the linear relation

$$S^n_{0,T}(\gamma) \otimes (\gamma_T - \gamma_0) = (n+1) S^{n+1}_{0,T}(\gamma). \tag{4.1}$$

Assume instead that γ is a piecewise linear path, and is linear on $[0,u]$ and $[u,T]$ respectively for $u \in (0,T)$. Then by Chen's identity, $S^n_{0,T}(\gamma) = \sum_{k=0}^{n} \frac{(\gamma_u - \gamma_0)^{\otimes k}}{k!} \otimes \frac{(\gamma_T - \gamma_u)^{\otimes (n-k)}}{(n-k)!}$ for $n \in \mathbb{N}$. We note the following lemma for such a path.

Lemma 4.1.1. *Consider a piecewise path $\gamma : [0,T] \to \mathbb{R}^d$ that is linear on $[0,u]$ and $[u,T]$ respectively for $u \in (0,T)$. Then*

$$(\gamma_u - \gamma_0) \otimes S^n_{0,T}(\gamma) + S^n_{0,T}(\gamma) \otimes (\gamma_T - \gamma_u) = (n+1) S^{n+1}_{0,T}(\gamma). \tag{4.2}$$

Proof.

$$\begin{aligned}
&(\gamma_u - \gamma_0) \otimes S^n_{0,T}(\gamma) + S^n_{0,T} \otimes (\gamma_T - \gamma_u) \\
&= \sum_{k=0}^{n} \frac{(\gamma_u - \gamma_0)^{\otimes (k+1)}}{k!} \otimes \frac{(\gamma_T - \gamma_u)^{\otimes (n-k)}}{(n-k)!} + \sum_{k=0}^{n} \frac{(\gamma_u - \gamma_0)^{\otimes k}}{k!} \otimes \frac{(\gamma_T - \gamma_u)^{\otimes (n-k+1)}}{(n-k)!}.
\end{aligned}$$

Note for $k = 1, \cdots, n$,

$$\begin{aligned}
&\frac{(\gamma_u - \gamma_0)^{\otimes k}}{(k-1)!} \otimes \frac{(\gamma_T - \gamma_u)^{\otimes(n-k+1)}}{(n-k+1)!} + \frac{(\gamma_u - \gamma_0)^{\otimes k}}{k!} \otimes \frac{(\gamma_T - \gamma_u)^{\otimes(n-k+1)}}{(n-k)!} \\
=&\frac{k+n-k+1}{k!(n-k+1)!}(\gamma_u - \gamma_0)^{\otimes k} \otimes (\gamma_T - \gamma_u)^{\otimes(n-k+1)} \\
=&(n+1)\frac{(\gamma_u - \gamma_0)^{\otimes k} \otimes (\gamma_T - \gamma_u)^{\otimes(n-k+1)}}{k!(n-k+1)!},
\end{aligned}$$

and

$$\begin{aligned}
\frac{(\gamma_u - \gamma_0)^{\otimes(n+1)}}{n!} &= (n+1)\frac{(\gamma_u - \gamma_0)^{\otimes(n+1)}}{(n+1)!}, \\
\frac{(\gamma_T - \gamma_u)^{\otimes(n+1)}}{n!} &= (n+1)\frac{(\gamma_T - \gamma_u)^{\otimes(n+1)}}{(n+1)!}.
\end{aligned}$$

Then

$$\begin{aligned}
&(\gamma_u - \gamma_0) \otimes S^n_{0,T}(\gamma) + S^n_{0,T}(\gamma) \otimes (\gamma_T - \gamma_u) \\
=&(n+1)\sum_{k=0}^{n+1} \frac{(\gamma_u - \gamma_0)^{\otimes k}}{k!} \otimes \frac{(\gamma_T - \gamma_u)^{\otimes(n+1-k)}}{(n+1-k)!} \\
=&(n+1)S^{n+1}_{0,T}(\gamma).
\end{aligned}$$

□

We note that by solving the linear equation (4.2) for $\gamma_u - \gamma_0$ and $\gamma_T - \gamma_u$, we are able to reconstruct exactly the underlying path. From a computational perspective, it is worth recalling the definition of the *singular value decomposition.*

Theorem 4.1.1 (Singular value decomposition)**.** *Suppose A is a matrix of dimension $m \times n$ with entries from $\mathbb{R}$ or $\mathbb{C}$. Then there exists a factorisation, called a* singular value decomposition (SVD) *of A, of the form*

$$A = U\Sigma V^T,$$

where
U is an $m \times m$ unitary matrix;
V is an $n \times n$ unitary matrix and V^T is the conjugate transpose of V;
Σ is an $m \times n$ diagonal matrix with non-negative real numbers on the diagonal.
The diagonal entries of Σ are called the singular values *of A.*

We can now computationally reconstruct a path consisting of two linear pieces. If $\gamma : [0,t] \to \mathbb{R}^d$ is a path consisting of linear pieces, Let the $2d \times d^{n+1}$ matrix A represent the linear mapping $\cdot \otimes S^n_{0,T}(\gamma) + S^n_{0,T}(\gamma) \otimes \cdot : (\mathbb{R}^d, \mathbb{R}^d) \to \mathbb{R}^{d^{n+1}}$, and $b \in \mathbb{R}^{d^{n+1}}$ represent $S^{n+1}_{0,T}(\gamma)$. Then Equation (4.2) can be written as, for a vector $X \in (\mathbb{R}^d, \mathbb{R}^d)$,

$$AX = (n+1)b.$$

By using SVD on A, we can obtain a simple computational algorithm that recovers γ, as shown in Example 4.1.1. Note the computation of the signature in the example is by the C++ package *Libalgebra* [6], and the matrix computation is done via *LAPACK* [1], the version of LAPACK used is provided by *Intel Math Kernel Library.*

Example 4.1.1. *For a two-dimensional path $\gamma : [0,T] \to \mathbb{R}^2$, $\gamma_t = (x_t, y_t)$ where*

$$y = \begin{cases} 2x & \forall x \in [0,1) \\ -\frac{2}{3}x + \frac{8}{3} & \forall x \in [1,4]. \end{cases}$$

By using two adjacent levels of the signature of γ, for instance, the third and fourth levels, we can fully reconstruct the underlying path γ by solving Equation (4.2), as shown in Figure 4.1.

Note from above we have shown that for a linear path or a piecewise linear path composed of two linear pieces, we are able to recover the path exactly by comparing two adjacent levels of the signature. This leads to the idea that we may as well get some information about the underlying path if we compare two adjacent levels of the signature of a more complicated path.
Another observation we can make is that if a path $\gamma : [0,1] \to \mathbb{R}^d$ is differentiable with derivative f, then by 'normalising' the signature, i.e. multiplying the n-th level of the signature of a path by $n!$, we have the integral

$$\int_{0<u_1<\cdots<u_n<1} f(u_1) \otimes \cdots \otimes f(u_n) n! \mathrm{d}u_1 \cdots \mathrm{d}u_n,$$

which is the expectation of a function under the distribution of n uniform order statistics over $[0,1]$. Note that the marginal distributions of uniform order statistics are beta distributions.

4.2 Comparing two adjacent levels of the signature of a C^1 path

From now on we may omit the symbol $\otimes$ in tensor multiplication for simplicity. First we define the properties of admissible norms we assume to be true for the rest of this chapter.

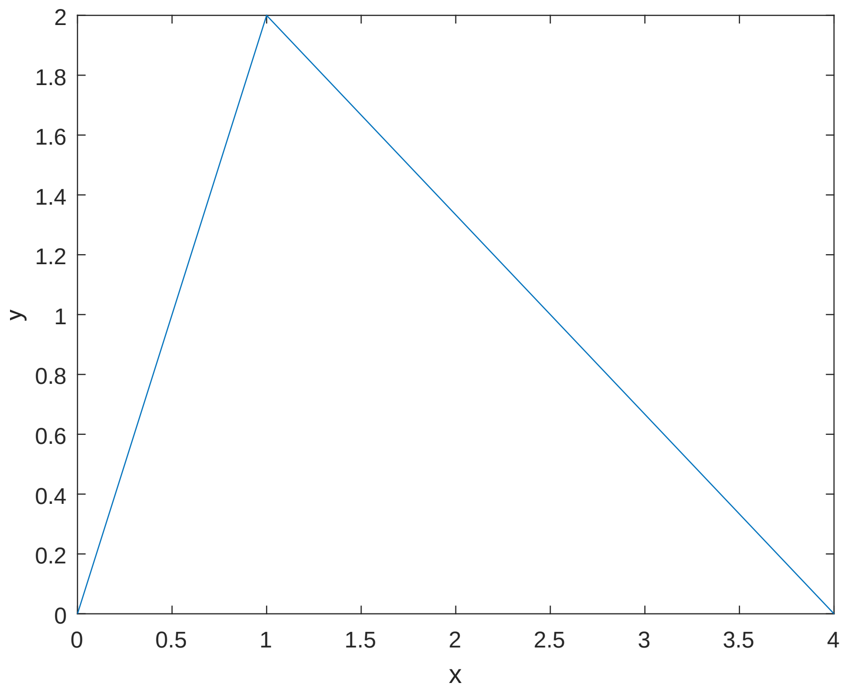


Figure 4.1: Reconstruction of y as a function of x for the path in Example 4.1.1

Definition 4.2.1. *Let V be a Banach space. Suppose the tensor powers are endowed with a tensor norm such that*

1. *For all $n \geq 1$, the norm of a tensor is invariant under permutation, i.e.*

$$\|\sigma(v)\| = \|v\| \quad \forall v \in V^{\otimes n}, \forall \sigma \in S(n),$$

where $S(n)$ is the symmetric group on n letters;

2. *For all $n, m \geq 1$,*

$$\|v \otimes \omega\| = \|v\| \, \|\omega\| \quad \forall v \in V^{\otimes n}, \omega \in V^{\otimes m}.$$

In the following lemma, we give a collection of norms which satisfy the properties stated in Definition 4.2.1.

Lemma 4.2.1. *Let $V = \mathbb{R}^d$ with a basis $\{e_1, \cdots, e_d\}$. Then for any element $u \in V^{\otimes n}$, we can recognise u as a vector in $\mathbb{R}^{d^n}$, and in this case, for any $l > 0$, ℓ^l norm satisfies the properties in Definition 4.2.1.*

Proof. Assume m, n are non-negative integers. Let $v = \sum_{i_1,..,i_n \in \{1,\cdots,d\}} \lambda_{i_1\cdots i_n} e_{i_1} \cdots e_{i_n} \in V^{\otimes n}$. Then for any permutation $\sigma \in S(n)$,

$$\begin{aligned}
\|\sigma(v)\|_l &= \left\| \sum_{i_1,\cdots,i_n \in \{1,\cdots,d\}} \lambda_{i_1\cdots i_n} e_{\sigma(i_1)} \cdots e_{\sigma(i_n)} \right\|_l \\
&= \left(\sum_{i_1,\cdots,i_n \in \{1,\cdots,d\}} \lambda^l_{i_1\cdots i_n} \right)^{\frac{1}{l}} \\
&= \|v\|_l \,.
\end{aligned}$$

Moreover, take $\omega = \sum_{j_1,\cdots,j_m \in \{1,\cdots,d\}} \mu_{j_1\cdots j_m} e_{j_1} \cdots e_{j_m} \in V^{\otimes m}$. Then

$$\begin{aligned}
\|v \otimes \omega\|_l &= \left\| \sum_{i_1,\cdots,i_n \in \{1,\cdots,d\}} \lambda_{i_1\cdots i_n} \left(\sum_{j_1,..,j_m \in \{1,\cdots,d\}} \mu_{j_1\cdots j_m} \right) e_{i_1} \cdots e_{i_n} e_{j_1} \cdots e_{j_m} \right\|_l \\
&= \left(\sum_{i_1,\cdots,i_n,j_1,\cdots,j_m \in \{1,\cdots,d\}} (\lambda_{i_1\cdots i_n} \mu_{j_1\cdots j_m})^l \right)^{\frac{1}{l}} \\
&= \left(\sum_{i_1,\cdots,i_n \in \{1,\cdots,d\}} \lambda^l_{i_1\cdots i_n} \sum_{j_1,\cdots,j_m \in \{1,\cdots,d\}} \mu^l_{j_1\cdots j_m} \right)^{\frac{1}{l}} \\
&= \left(\sum_{i_1,\cdots,i_n \in \{1,\cdots,d\}} \lambda^l_{i_1\cdots i_n} \right)^{\frac{1}{l}} \left(\sum_{j_1,\cdots,j_m \in \{1,\cdots,d\}} \mu^l_{j_1\cdots j_m} \right)^{\frac{1}{l}} \\
&= \|v\|_l \|\omega\|_l \,.
\end{aligned}$$

□

In this section, unless otherwise specified, let $\gamma : [0,1] \to \mathbb{R}^d$ be a path with continuous derivative $f : (0,1) \to \mathbb{R}^d$ such that $||f(t)|| = 1$ for all $t \in (0,1)$. Define the modulus of continuity of f by

$$\delta(h) := \sup_{|u-v|<h} \|f(u) - f(v)\| \,.$$

Definition 4.2.2 (Normalised signature)**.** *Assume γ is a continuous path with bounded variation over the interval $[s,t]$ parametrised at unit speed. Define the* normalised signature *of γ over $[s,t]$ as*

$$\left(1, \bar{S}^1_{s,t}(\gamma), \bar{S}^2_{s,t}(\gamma), \cdots, \bar{S}^m_{s,t}(\gamma), \cdots\right),$$

where for all $m \geq 1$,

$$\bar{S}^m_{s,t}(\gamma) := \frac{m! \int_{s<u_1<\cdots<u_m<t} f(u_1)\cdots f(u_m) \mathrm{d}u_1 \cdots \mathrm{d}u_m}{(t-s)^m}. \tag{4.3}$$

For simplicity, we write $\bar{S}_m := \bar{S}^m_{0,1}(\gamma)$.

Definition 4.2.3 (Insertion map)**.** *Assume $\gamma : [0,1] \to \mathbb{R}^d$ is a continuous path of bounded variation parametrised at unit speed. For $p = 1, \cdots, n+1$, define the mapping function $I_{p,n} : \mathbb{R}^d \to (\mathbb{R}^d)^{\otimes(n+1)}$ by*

$$I_{p,n}(x) := \int_{0<u_1<\cdots<u_n<1} f(u_1)\cdots f(u_{p-1}) x f(u_p) \cdots f(u_n) n! \mathrm{d}u_1 \cdots \mathrm{d}u_n, \tag{4.4}$$

i.e. $I_{p,n}(x)$ is the function that inserts x into the p-th position of the n-th normalised signature. Note that the operation of inserting $x \in \mathbb{R}^d$ into a homogeneous tensor $t \in \mathbb{R}^{d^{\otimes n}}$ at p-th position is well-defined.

Similarly, for $p = 1, \cdots, n+1$, define the mapping $R_{p,n+1} : \mathbb{R}^d \to (\mathbb{R}^d)^{\otimes(n+1)}$ by

$$R_{p,n+1}(x) := \int_{0<u_1<\cdots<u_{n+1}<1} f(u_1)\cdots f(u_{p-1}) x f(u_{p+1}) \cdots f(u_{n+1}) (n+1)! \mathrm{d}u_1 \cdots \mathrm{d}u_{n+1}, \tag{4.5}$$

i.e. $R_{p,n+1}$ replaces the p-th element of the $(n+1)$-th normalised signature by x.
A simple observation is that the function $I_{p,n}$ is linear, as stated in the following lemma.

Lemma 4.2.2. *Assume $\gamma : [0,1] \to \mathbb{R}^d$ is differentiable almost surely, and the derivative $f : (0,1) \to \mathbb{R}^d$ satisfies $\|f(t)\| = 1$ if defined. For all $a, b \in \mathbb{R}$, $x, y \in \mathbb{R}^d$, $n \geq 1$, $p \in \{1, \cdots, n+1\}$, $I_{p,n}(ax+by) = aI_{p,n}(x) + bI_{p,n}(y)$.*

Proof.

$$\begin{aligned}
&I_{p,n}(ax+by) \\
=&\int_{0<u_1<\cdots<u_n<1} f(u_1)\cdots f(u_{p-1})(ax+by) f(u_p)\cdots f(u_n) n! \mathrm{d}u_1\cdots \mathrm{d}u_n \\
=&a\int_{0<u_1<\cdots<u_n<1} f(u_1)\cdots f(u_{p-1}) x f(u_p)\cdots f(u_n) n! \mathrm{d}u_1\cdots \mathrm{d}u_n \\
&+ b\int_{0<u_1<\cdots<u_n<1} f(u_1)\cdots f(u_{p-1}) y f(u_p)\cdots f(u_n) n! \mathrm{d}u_1\cdots \mathrm{d}u_n \\
=&aI_{p,n}(x) + bI_{p,n}(y).
\end{aligned}$$

□

Because of the properties of the norm stated in Definition 4.2.1, we are able to state the following property of the distances between images of the map $I_{p,n}$ which we will use later.

Lemma 4.2.3. *Assume $\gamma : [0,1] \to \mathbb{R}^d$ is differentiable almost everywhere with derivative $f : (0,1) \to \mathbb{R}^d$ such that $\|f(t)\| = 1$ if defined. Then following the properties of the norm defined in Definition 4.2.1, for any $x, y \in \mathbb{R}^d$, $n \geq 1$, $p \in \{1, \cdots, n+1\}$,*

$$\|I_{p,n}(x) - I_{p,n}(y)\| = \left\|\bar{S}_n\right\| \|x - y\| . \tag{4.6}$$

Proof.

$$\begin{aligned}
&\|I_{p,n}(x) - I_{p,n}(y)\| \\
&= \left\| \int_{0<u_1<\cdots<u_n<1} f(u_1) \cdots (x-y) \cdots f(u_n) n! \mathrm{d}u_1 \cdots \mathrm{d}u_n \right\| \\
&= \left\| \int_{0<u_1<\cdots<u_n<1} f(u_1) \cdots f(u_n) n! \mathrm{d}u_1 \cdots \mathrm{d}u_n (x-y) \right\| \\
&= \left\| \int_{0<u_1<\cdots<u_n<1} f(u_1) \cdots f(u_n) n! \mathrm{d}u_1 \cdots \mathrm{d}u_n \right\| \|x-y\| \\
&= \left\|\bar{S}_n\right\| \|x - y\| .
\end{aligned}$$

□

Corollary 4.2.1. *The function $I_{p,n}$ is Lipschitz continuous.*

Proof. It is a direct consequence of Equation (4.6) that $I_{p,n}$ is Lipschitz. □

With a suitable choice of x, we hope to approximate $\bar{S}_{n+1}$ by $I_{p,n}(x)$. For such a purpose, we first introduce some definitions that we will use.

Definition 4.2.4 (Sub-Gaussian variables)**.** *A random variable X with finite mean $\mathbb{E}[X] = \mu$ is* sub-Gaussian *if there exists $\sigma \in \mathbb{R}$, $\sigma > 0$ such that*

$$\mathbb{E}\left[\exp\left(\lambda\left(X - \mu\right)\right)\right] \leq \exp\left(\frac{\lambda^2\sigma^2}{2}\right) \quad \forall\, \lambda \in \mathbb{R}.$$

Such a constant σ^2 is called a proxy variance, *and X is then called σ^2-sub-Gaussian. If X is sub-Gaussian, the* optimal proxy variance *is defined as*

$$\sigma^2_{opt} := \min\left\{\sigma^2 \geq 0 \, \textit{such that} \, X \textit{ is sub-Gaussian}\right\}.$$

It is useful to recall the definitions of the beta distribution and the Dirichlet distribution for later use.

Definition 4.2.5 (Beta distributions)**.** *A beta variable $X \sim Beta(\alpha, \beta)$ has a probability density function $f : (0,1) \to [0,1]$ such that*

$$f(x) = \frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)} x^{\alpha-1}(1-x)^{\beta-1},$$

where $\Gamma(z)$ is the gamma function*. Note if n is a positive integer, $\Gamma(n) = (n-1)!$.*

Definition 4.2.6 (Dirichlet distributions)**.** *The Dirichlet distribution of order $K \geq 2$ with parameters $\alpha_1, \cdots, \alpha_K > 0$ has a probability density function given by*

$$f(x_1, \cdots, x_K; \alpha_1, \cdots, \alpha_K) = \frac{1}{B(\boldsymbol{\alpha})} \prod_{i=1}^{K} x_i^{\alpha_i - 1}$$

where $x_i \geq 0$ and $\sum_{i=1}^{K} x_i = 1$ for all $i = 1, \cdots, K$. Note

$$B(\boldsymbol{\alpha}) := \frac{\prod_{i=1}^{K} \Gamma(\alpha_i)}{\Gamma(\sum_{i=1}^{K} \alpha_i)}, \quad \boldsymbol{\alpha} = (\alpha_1, \cdots, \alpha_K)$$

where $\Gamma(z)$ is the gamma function.

In particular, beta variables are sub-Gaussian, and an upper bound can be found for the optimal proxy variance, as proved by Marchal and Arbel [26].

Theorem 4.2.1 (Marchal and Arbel [26])**.** *For any $\alpha, \beta > 0$, the beta distribution $Beta(\alpha, \beta)$ is $\sigma^2_{opt}(\alpha, \beta)$-sub-Gaussian. An upper bound for the optimal proxy variance of $Beta(\alpha, \beta)$ is given by $\sigma^2(\alpha, \beta) = \frac{1}{4(\alpha+\beta+1)}$.*

Because of the sub-Gaussian property of the beta distribution, an upper bound exists for the tails of the beta distribution.

Lemma 4.2.4. *Suppose $X \sim Beta(\alpha, \beta)$ with mean $\mathbb{E}[X] = \mu$. Assume σ_0^2 is an upper bound for the optimal proxy variance of X. Then for any $a \in \mathbb{R}$, $t \in (0,1)$,*

$$\mathbb{P}(|X - a| \geq t) \leq \exp\left(-\frac{t^2}{2\sigma_0^2}\right)\left(\exp\left(\frac{t(\mu - a)}{\sigma_0^2}\right) + \exp\left(-\frac{t(\mu - a)}{\sigma_0^2}\right)\right). \quad (4.7)$$

Proof. See Appendix A. □

We are now ready to compare $\bar{S}_{n+1}$ and $R_{p,n+1}(x)$.

Proposition 4.2.1. *Assume $\gamma : [0,1] \to \mathbb{R}^d$ is a path with continuous derivative $f : (0,1) \to \mathbb{R}^d$ such that $\|f(t)\| = 1$ for all $t \in (0,1)$, and $\delta(h) := \sup_{|u-v|<h} \|f(u) - f(v)\|$*

for $h \in [0,1]$. *Fix* $\theta \in (0,1)$. *For* $n \geq 1$, *choose* $p \in \{1, \cdots, n+1\}$ *such that* $p = \lfloor \theta(n+2) \rfloor$. *Then*

$$\left\| R_{p,n+1}(f(\theta)) - \bar{S}_{n+1} \right\| \leq \delta \left(\sqrt{\frac{\log(n+3)}{n+3}} \right) + \frac{C}{(n+3)^2} \to 0$$

as $n \to \infty$, *where* C *is some constant.*

Proof. Note for a fixed $\theta \in (0,1)$ and a given $n \in \mathbb{N}$, it might happen that $\lfloor \theta(n+2) \rfloor = 0$. We can avoid this by considering large enough n such that $p = \lfloor \theta(n+2) \rfloor$ is at least 1. Note for any $x \in \mathbb{R}^d$,

$$\begin{aligned}
&\left\| R_{p,n+1}(x) - \bar{S}_{n+1} \right\| \\
=&\left\| \int_{0<u_1<\cdots<u_{n+1}<1} f(u_1) \cdots f(u_{p-1})(x - f(u_p)) \cdots f(u_{n+1})(n+1)! \mathrm{d}u_1 \cdots \mathrm{d}u_{n+1} \right\| \\
\leq& \int_{0<u<1} \|x - f(u)\| (n+1)! \frac{u^{p-1}}{(p-1)!} \frac{(1-u)^{n+1-p}}{(n+1-p)!} \mathrm{d}u,
\end{aligned}$$

which give rise to a variable $U \sim \mathrm{Beta}(p, n-p+2)$, and $\mathbb{E}[U] = \frac{p}{n+2}$. Note that a beta variable has its majority of the density concentrated near the mean. In order to minimise $\left\| R_{p,n+1}(x) - \bar{S}_{n+1} \right\|$, it is interesting to consider the case when $x = f(\theta)$. Note by the result from Theorem 4.2.1, U is sub-Gaussian, and an upper bound for the optimal proxy variance of U is $1/(4(n+3))$. Note

$$0 \leq \theta - \frac{p}{n+2} \leq \frac{1}{n+2}.$$

Then Lemma 4.2.4 implies that for all $h \in (0,1)$,

$$\begin{aligned}
&\mathbb{P}\left(|U - \theta| \geq h\right) \\
\leq& \exp\left(-2(n+3)h^2\right) \\
&\left(\exp\left(4(n+3)h\left(\theta - \frac{p}{n+2}\right)\right) + \exp\left(-4(n+3)h\left(\theta - \frac{p}{n+2}\right)\right)\right) \\
\leq& 2\exp\left(-2(n+3)h^2\right)\exp(8h) \\
\leq& 2e^8 \exp\left(-2(n+3)h^2\right).
\end{aligned}$$

Then for any $h \in (0,1)$,

$$\begin{aligned}
&\left\| R_{p,n+1}(f(\theta)) - \bar{S}_{n+1} \right\| \\
\leq& \int_{0<u<1} \|f(\theta) - f(u)\| (n+1)! \frac{u^{p-1}}{(p-1)!} \frac{(1-u)^{n+1-p}}{(n+1-p)!} \mathrm{d}u
\end{aligned}$$

$$
\begin{aligned}
&= \int_{|u-\theta|<h} \|f(\theta) - f(u)\| \,(n+1)! \frac{u^{p-1}}{(p-1)!} \frac{(1-u)^{n+1-p}}{(n+1-p)!} \mathrm{d}u \\
&\quad + \int_{|u-\theta|>h} \|f(\theta) - f(u)\| \,(n+1)! \frac{u^{p-1}}{(p-1)!} \frac{(1-u)^{n+1-p}}{(n+1-p)!} \mathrm{d}u \\
&\le \delta(h) + 2\mathbb{P}\left(|U-\theta| \ge h\right) \\
&\le \delta(h) + 4e^8 \exp\left(-2(n+3)h^2\right).
\end{aligned}
$$

Hence

$$
\left\|R_{p,n+1}(f(\theta)) - \bar{S}_{n+1}\right\| \le \delta(h) + 4e^8 \exp\left(-2(n+3)h^2\right).
$$

In particular, if we choose $h = \sqrt{\frac{\log(n+3)}{n+3}}$, we have

$$
\left\|R_{p,n+1}(f(\theta)) - \bar{S}_{n+1}\right\| \le \delta\left(\sqrt{\frac{\log(n+3)}{n+3}}\right) + \frac{4e^8}{(n+3)^2},
$$

which converges to 0 as n goes to infinity. □

Example 4.2.1. *Following the assumptions in Proposition 4.2.1, consider the case when $\delta(h) = h^\rho$ for some $\rho \in (0,1]$. By taking $h = \sqrt{\log(n+3)/(n+3)}$, we have*

$$
\left\|R_{p,n+1}(f(\theta)) - \bar{S}_{n+1}\right\| \le \left(\frac{\log(n+3)}{n+3}\right)^{\frac{\rho}{2}} + \frac{4e^8}{(n+3)^2},
$$

so it has rate of convergence $O((\log(n+3)/(n+3))^{\rho/2})$.

We introduce another quantity: for $p = \{1, \cdots, n+1\}$, define

$$
M_{p,n+1} := \int_{0<u_1<\cdots<u_{n+1}<1} f(u_1)\cdots f(u_{p-1}) f(u_{p+1}) \cdots f(u_{n+1}) (n+1)! \mathrm{d}u_1 \cdots \mathrm{d}u_{n+1}, \tag{4.8}
$$

which is essentially the $(n+1)$-th normalised signature with the p-th element removed. We first observe the equivalence of $\|I_{p,n}(x) - R_{p,n+1}(x)\|$ and $\left\|M_{p,n+1} - \bar{S}_n\right\|$.

Lemma 4.2.5. *Assume $\gamma : [0,1] \to \mathbb{R}^d$ is differentiable almost everywhere, and has derivative f such that $\|f(t)\| = 1$ for all $t \in (0,1)$ where it is defined. Then following the properties of the tensor norm defined in Definition 4.2.1, we have for all $n \ge 1$, $p \in \{1, \cdots, n+1\}$, $x \in \mathbb{R}^d$ such that $\|x\| = 1$,*

$$
\|I_{p,n}(x) - R_{p,n+1}(x)\| = \left\|M_{p,n+1} - \bar{S}_n\right\|.
$$

Proof. Note

$$
\begin{aligned}
&\|I_{p,n}(x) - R_{p,n+1}(x)\| \\
=&\Big\| \int_{0<u_1<\cdots<u_n<1} f(u_1)\cdots f(u_{p-1}) x f(u_p)\cdots f(u_n) n! \mathrm{d}u_1\cdots \mathrm{d}u_n \\
&- \int_{0<u_1<\cdots<u_{n+1}<1} f(u_1)\cdots f(u_{p-1}) x f(u_{p+1})\cdots f(u_{n+1})(n+1)! \mathrm{d}u_1\cdots \mathrm{d}u_{n+1}\Big\| \\
=&\Big\| \Big(\int_{0<u_1<\cdots<u_n<1} f(u_1)\cdots f(u_{p-1}) f(u_p)\cdots f(u_n) n! \mathrm{d}u_1\cdots \mathrm{d}u_n \\
&- \int_{0<u_1<\cdots<u_{n+1}<1} f(u_1)\cdots f(u_{p-1}) f(u_{p+1})\cdots f(u_{n+1})(n+1)! \mathrm{d}u_1\cdots \mathrm{d}u_{n+1}\Big) x\Big\| \\
=&\Big\| \int_{0<u_1<\cdots<u_n<1} f(u_1)\cdots f(u_{p-1}) f(u_p)\cdots f(u_n) n! \mathrm{d}u_1\cdots \mathrm{d}u_n \\
&- \int_{0<u_1<\cdots<u_{n+1}<1} f(u_1)\cdots f(u_{p-1}) f(u_{p+1})\cdots f(u_{n+1})(n+1)! \mathrm{d}u_1\cdots \mathrm{d}u_{n+1}\Big\| \|x\| \\
=&\big\|M_{p,n+1} - \bar{S}_n\big\|
\end{aligned}
$$

by the properties of the norm described in Definition 4.2.1. □

Because of Lemma 4.2.5, we would then like to compare $M_{p,n+1}$ and $\bar{S}_n$. Note that we can rewrite $M_{p,n+1}$ as

$$
\begin{aligned}
M_{p,n+1} = \int_{0<u_{p-1}<u_{p+1}<1} & \bar{S}^{p-2}_{0,u_{p-1}}(\gamma) f(u_{p-1}) f(u_{p+1}) \bar{S}^{n-p}_{u_{p+1},1}(\gamma) \\
& (n+1)!(u_{p+1}-u_{p-1}) \frac{u_{p-1}^{p-2}}{(p-2)!} \frac{(1-u_{p+1})^{n-p}}{(n-p)!} \mathrm{d}u_{p-1}\mathrm{d}u_{p+1}.
\end{aligned}
$$

Similarly,

$$
\begin{aligned}
\bar{S}_n = \int_{0<u_{p-1}<u_p<1} & \bar{S}^{p-2}_{0,u_{p-1}}(\gamma) f(u_{p-1}) f(u_p) \bar{S}^{n-p}_{u_p,1}(\gamma) \\
& n! \frac{u_{p-1}^{p-2}}{(p-2)!} \frac{(1-u_p)^{n-p}}{(n-p)!} \mathrm{d}u_{p-1}\mathrm{d}u_p.
\end{aligned}
$$

From above we can see that if we want to learn about the difference between $M_{p,n+1}$ and $\bar{S}_n$, it is interesting to discuss the continuity of $\bar{S}^m_{s,t}(\gamma)$ for $m \geq 1$. Chen [10] studied the distance between two integrals of different functions over the same region.

Lemma 4.2.6 (Chen [10]). *Let $f_i(t), g_i(t)$, $i = 1, \cdots, q$, be piecewise continuous with $\|f_i(t)\| \leq M$, $\|g_i(t)\| \leq M$, $\|g_i(t) - f_i(t)\| \leq m$ for $a \leq t \leq b$. Then for $q \geq 1$,*

$$
\left\| \int_{a<t_1<\cdots<t_q<b} g_1(t_1)\cdots g_q(t_q)\mathrm{d}t_1\cdots \mathrm{d}t_q - \int_{a<t_1<\cdots<t_q<b} f_1(t_1)\cdots f_q(t_q)\mathrm{d}t_1\cdots \mathrm{d}t_q \right\|
$$

$$\leq \frac{mM^{q-1}(b-a)^q}{(q-1)!}.$$

The lemma above gives us an idea that we may re-parametrise the path in order to find the modulus of continuity of $\bar{S}^m_{s,t}(\gamma)$.

Lemma 4.2.7. *Assume $\gamma : [0,1] \to \mathbb{R}^d$ is a path with continuous derivative $f : (0,1) \to \mathbb{R}^d$ such that $\|f(t)\| = 1$ for all $t \in (0,1)$, and $\delta(h) := \sup_{|u-v|<h}\|f(u)-f(v)\|$ for $h \in [0,1]$. Then for $m \geq 1$, for $s, u \in (0,t)$,*

$$\left\|\bar{S}^m_{u,t}(\gamma) - \bar{S}^m_{s,t}(\gamma)\right\| \leq m\delta(|u-s|).$$

Proof. Define $\lambda = \frac{t-u}{t-s}$.

$$\begin{aligned}
&\left\|\bar{S}^m_{u,t}(\gamma) - \bar{S}^m_{s,t}(\gamma)\right\| \\
=&\left\|\frac{m!\int_{s<v_1<\cdots<v_m<t} f(\lambda v_1 + t(1-\lambda))\cdots f(\lambda v_m + t(1-\lambda))\mathrm{d}v_1\cdots\mathrm{d}v_m}{(t-s)^m}\right. \\
&\left. - \frac{m!\int_{s<u_1<\cdots<u_m<t} f(u_1)\cdots f(u_m)\mathrm{d}u_1\cdots\mathrm{d}u_m}{(t-s)^m}\right\| \text{ by re-parametrisation} \\
\leq&\frac{m!\int_{s<u_1<\cdots<u_m<t}\sum_{i=1}^m\|f(\lambda u_i + t(1-\lambda)) - f(u_i)\|\mathrm{d}u_1\cdots\mathrm{d}u_m}{(t-s)^m} \\
\leq&\frac{m!\int_{s<u_1<\cdots<u_m<t}\delta(|1-\lambda|(t-u_1)) + \cdots + \delta(|1-\lambda|(t-u_m))\mathrm{d}u_1\cdots\mathrm{d}u_m}{(t-s)^m} \\
\leq&\frac{m!\int_{s<u_1<\cdots<u_m<t}\delta(|1-\lambda|(t-s)) + \cdots + \delta(|1-\lambda|(t-s))\mathrm{d}u_1\cdots\mathrm{d}u_m}{(t-s)^m} \\
=&m\delta(|u-s|).
\end{aligned}$$

□

Similarly we can prove the same modulus of continuity of $\bar{S}^m_{s,t}(\gamma)$ in t.

Lemma 4.2.8. *Assume $\gamma : [0,1] \to \mathbb{R}^d$ is a path with continuous derivative $f : (0,1) \to \mathbb{R}^d$ such that $\|f(t)\| = 1$ for all $t \in (0,1)$, and $\delta(h) := \sup_{|u-v|<h}\|f(u)-f(v)\|$ for $h \in [0,1]$. Then for $m \geq 1$, for $u, t \in (s,1)$,*

$$\left\|\bar{S}^m_{s,t}(\gamma) - \bar{S}^m_{s,u}(\gamma)\right\| \leq m\delta(|u-t|).$$

Proof. Define $\lambda = \frac{t-s}{u-s}$. Note

$$\begin{aligned}
&\left\|\bar{S}^m_{s,t}(\gamma) - \bar{S}^m_{s,u}(\gamma)\right\| \\
=&\left\|\frac{m!\int_{s<v_1<\cdots<v_m<u} f(\lambda v_1 + s(1-\lambda))\cdots f(\lambda v_m + s(1-\lambda))\mathrm{d}v_1\cdots\mathrm{d}v_m}{(u-s)^m}\right.
\end{aligned}$$

$$
\begin{aligned}
&-\frac{m!\int_{s<u_1<\cdots<u_m<u} f(u_1)\cdots f(u_m)\mathrm{d}u_1\cdots\mathrm{d}u_m}{(u-s)^m}\Bigg\| \\
\leq&\frac{m!\int_{s<u_1<\cdots<u_m<u}\sum_{i=1}^m\|f(\lambda u_i+s(1-\lambda))-f(u_i)\|\mathrm{d}u_1\cdots\mathrm{d}u_m}{(u-s)^m} \\
\leq&\frac{m!\int_{s<u_1<\cdots<u_m<u}\delta(|1-\lambda|(u_1-s))+\cdots+\delta(|1-\lambda|(u_m-s))\mathrm{d}u_1\cdots\mathrm{d}u_m}{(u-s)^m} \\
\leq&\frac{m!\int_{s<u_1<\cdots<u_m<u}\delta(|1-\lambda|(u-s))+\cdots+\delta(|1-\lambda|(u-s))\mathrm{d}u_1\cdots\mathrm{d}u_m}{(u-s)^m} \\
=&m\delta(|u-t|).
\end{aligned}
$$

□

Now we are ready to find an upper bound for the distance between $M_{p,n+1}$ and $\bar{S}_n$.

Proposition 4.2.2. *Assume $\gamma : [0,1] \to \mathbb{R}^d$ is a path with continuous derivative $f : (0,1) \to \mathbb{R}^d$ such that $\|f(t)\| = 1$ for all $t \in (0,1)$, and define $\delta(h) := \sup_{|u-v|<h}\|f(u)-f(v)\|$ for $h \in [0,1]$. Fix $\theta \in (0,1)$. For $n \geq 1$, choose $p \in \{1,\cdots,n+1\}$ such that $p = \lfloor\theta(n+2)\rfloor$. Then*

$$
\left\|M_{p,n+1}-\bar{S}_n\right\| \leq O\left(n\delta\left(\sqrt{\frac{\log n}{n}}\right)+\frac{1}{n^2}\right).
$$

Proof. Recall that we can write

$$
\begin{aligned}
M_{p,n+1}=&\int_{0<u_{p-1}<u_{p+1}<1}\bar{S}^{p-2}_{0,u_{p-1}}(\gamma)f(u_{p-1})f(u_{p+1})\bar{S}^{n-p}_{u_{p+1},1}(\gamma) \\
&(n+1)!(u_{p+1}-u_{p-1})\frac{u_{p-1}^{p-2}}{(p-2)!}\frac{(1-u_{p+1})^{n-p}}{(n-p)!}\mathrm{d}u_{p-1}\mathrm{d}u_{p+1},
\end{aligned}
\tag{4.9}
$$

and

$$
\begin{aligned}
\bar{S}_n=&\int_{0<u_{p-1}<u_p<1}\bar{S}^{p-2}_{0,u_{p-1}}(\gamma)f(u_{p-1})f(u_p)\bar{S}^{n-p}_{u_p,1}(\gamma) \\
&n!\frac{u_{p-1}^{p-2}}{(p-2)!}\frac{(1-u_p)^{n-p}}{(n-p)!}\mathrm{d}u_{p-1}\mathrm{d}u_p.
\end{aligned}
\tag{4.10}
$$

Note that Equation (4.9) gives rise to a Dirichlet distribution with mean $(\frac{p-1}{n+2},\frac{p+1}{n+2})$. We know that the density of a Dirichlet distribution is concentrated near the mean, therefore we can consider

$$
\left\|M_{p,n+1}-\bar{S}^{p-2}_{0,\frac{p-1}{n+2}}(\gamma)f(\frac{p-1}{n+2})f(\frac{p+1}{n+2})\bar{S}^{n-p}_{\frac{p+1}{n+2},1}(\gamma)\right\|
$$

and

$$\left\|\bar{S}_n - \bar{S}^{p-2}_{0,\frac{p-1}{n+2}}(\gamma) f(\frac{p-1}{n+2}) f(\frac{p+1}{n+2}) \bar{S}^{n-p}_{\frac{p+1}{n+2},1}(\gamma)\right\|$$

respectively to approach $\|M_{p,n+1} - \bar{S}_n\|$.

$$\begin{aligned}
&\left\|M_{p,n+1} - \bar{S}^{p-2}_{0,\frac{p-1}{n+2}}(\gamma) f(\frac{p-1}{n+2}) f(\frac{p+1}{n+2}) \bar{S}^{n-p}_{\frac{p+1}{n+2},1}(\gamma)\right\| \\
=&\Big\| \int_{0<u_1<u_2<1} (\bar{S}^{p-2}_{0,u_1}(\gamma) f(u_1) f(u_2) \bar{S}^{n-p}_{u_2,1}(\gamma) - \bar{S}^{p-2}_{0,\frac{p-1}{n+2}}(\gamma) f(\frac{p-1}{n+2}) f(\frac{p+1}{n+2}) \bar{S}^{n-p}_{\frac{p+1}{n+2},1}(\gamma)) \\
&(n+1)!(u_2-u_1) \frac{u_1^{p-2}}{(p-2)!} \frac{(1-u_2)^{n-p}}{(n-p)!} \mathrm{d}u_1 \mathrm{d}u_2 \Big\| \\
\leq &\int_{0<u_1<u_2<1} \left\|\bar{S}^{p-2}_{0,u_1}(\gamma) - \bar{S}^{p-2}_{0,\frac{p-1}{n+2}}(\gamma)\right\| (n+1)!(u_2-u_1) \frac{u_1^{p-2}}{(p-2)!} \frac{(1-u_2)^{n-p}}{(n-p)!} \mathrm{d}u_1 \mathrm{d}u_2 \\
&+ \int_{0<u_1<u_2<1} \left\|f(u_1) - f(\frac{p-1}{n+2})\right\| (n+1)!(u_2-u_1) \frac{u_1^{p-2}}{(p-2)!} \frac{(1-u_2)^{n-p}}{(n-p)!} \mathrm{d}u_1 \mathrm{d}u_2 \\
&+ \int_{0<u_1<u_2<1} \left\|f(u_2) - f(\frac{p+1}{n+2})\right\| (n+1)!(u_2-u_1) \frac{u_1^{p-2}}{(p-2)!} \frac{(1-u_2)^{n-p}}{(n-p)!} \mathrm{d}u_1 \mathrm{d}u_2 \\
&+ \int_{0<u_1<u_2<1} \left\|\bar{S}^{n-p}_{u_2,1}(\gamma) - \bar{S}^{n-p}_{\frac{p+1}{n+2},1}(\gamma)\right\| (n+1)!(u_2-u_1) \frac{u_1^{p-2}}{(p-2)!} \frac{(1-u_2)^{n-p}}{(n-p)!} \mathrm{d}u_1 \mathrm{d}u_2 \\
= &\int_{0<u<1} \left\|\bar{S}^{p-2}_{0,u}(\gamma) - \bar{S}^{p-2}_{0,\frac{p-1}{n+2}}(\gamma)\right\| (n+1)! \frac{u^{p-2}}{(p-2)!} \frac{(1-u)^{n-p+2}}{(n-p+2)!} \mathrm{d}u \\
&+ \int_{0<u<1} \left\|f(u) - f(\frac{p-1}{n+2})\right\| (n+1)! \frac{u^{p-2}}{(p-2)!} \frac{(1-u)^{n-p+2}}{(n-p+2)!} \mathrm{d}u \\
&+ \int_{0<u<1} \left\|f(u) - f(\frac{p+1}{n+2})\right\| (n+1)! \frac{u^{p}}{p!} \frac{(1-u)^{n-p}}{(n-p)!} \mathrm{d}u \\
&+ \int_{0<u<1} \left\|\bar{S}^{n-p}_{u,1}(\gamma) - \bar{S}^{n-p}_{\frac{p+1}{n+2},1}(\gamma)\right\| (n+1)! \frac{u^{p}}{p!} \frac{(1-u)^{n-p}}{(n-p)!} \mathrm{d}u.
\end{aligned}$$

Note that now we have densities of distributions Beta$(p-1, n-p+3)$ and Beta$(p+1, n-p+1)$ in the integrals. We can split each integral into two parts: the part near the mean and the tail part. Using the modulus of continuity we have stated for $\bar{S}^{p-2}_{0,u}(\gamma)$ and $\bar{S}^{n-p}_{v,1}(\gamma)$ in Lemma 4.2.7 and Lemma 4.2.8, we have for $h \in (0,1)$,

$$\begin{aligned}
&\left\|M_{p,n+1} - \bar{S}^{p-2}_{0,\frac{p-1}{n+2}}(\gamma) f(\frac{p-1}{n+2}) f(\frac{p+1}{n+2}) \bar{S}^{n-p}_{\frac{p+1}{n+2},1}(\gamma)\right\| \\
\leq &\int_{|u-\frac{p-1}{n+2}|<h} \left\|\bar{S}^{p-2}_{0,u}(\gamma) - \bar{S}^{p-2}_{0,\frac{p-1}{n+2}}(\gamma)\right\| (n+1)! \frac{u^{p-2}}{(p-2)!} \frac{(1-u)^{n-p+2}}{(n-p+2)!} \mathrm{d}u \\
&+ \int_{|u-\frac{p-1}{n+2}|>h} \left\|\bar{S}^{p-2}_{0,u}(\gamma) - \bar{S}^{p-2}_{0,\frac{p-1}{n+2}}(\gamma)\right\| (n+1)! \frac{u^{p-2}}{(p-2)!} \frac{(1-u)^{n-p+2}}{(n-p+2)!} \mathrm{d}u
\end{aligned}$$

$$
\begin{aligned}
&+\int_{|u-\frac{p-1}{n+2}|<h}\left\|f(u)-f(\frac{p-1}{n+2})\right\|(n+1)!\frac{u^{p-2}}{(p-2)!}\frac{(1-u)^{n-p+2}}{(n-p+2)!}\mathrm{d}u\\
&+\int_{|u-\frac{p-1}{n+2}|>h}\left\|f(u)-f(\frac{p-1}{n+2})\right\|(n+1)!\frac{u^{p-2}}{(p-2)!}\frac{(1-u)^{n-p+2}}{(n-p+2)!}\mathrm{d}u\\
&+\int_{|u-\frac{p+1}{n+2}|<h}\left\|f(u)-f(\frac{p+1}{n+2})\right\|(n+1)!\frac{u^{p}}{p!}\frac{(1-u)^{n-p}}{(n-p)!}\mathrm{d}u\\
&+\int_{|u-\frac{p+1}{n+2}|>h}\left\|f(u)-f(\frac{p+1}{n+2})\right\|(n+1)!\frac{u^{p}}{p!}\frac{(1-u)^{n-p}}{(n-p)!}\mathrm{d}u\\
&+\int_{|u-\frac{p+1}{n+2}|<h}\left\|\bar{S}^{n-p}_{u,1}(\gamma)-\bar{S}^{n-p}_{\frac{p+1}{n+2},1}(\gamma)\right\|(n+1)!\frac{u^{p}}{p!}\frac{(1-u)^{n-p}}{(n-p)!}\mathrm{d}u\\
&+\int_{|u-\frac{p+1}{n+2}|>h}\left\|\bar{S}^{n-p}_{u,1}(\gamma)-\bar{S}^{n-p}_{\frac{p+1}{n+2},1}(\gamma)\right\|(n+1)!\frac{u^{p}}{p!}\frac{(1-u)^{n-p}}{(n-p)!}\mathrm{d}u\\
\leq&(p-2)\delta(h)+2\mathbb{P}\left(\left|U_1-\frac{p-1}{n+2}\right|\geq h\right)+\delta(h)+2\mathbb{P}\left(\left|U_1-\frac{p-1}{n+2}\right|\geq h\right)\\
&+\delta(h)+2\mathbb{P}\left(\left|U_2-\frac{p+1}{n+2}\right|\geq h\right)+(n-p)\delta(h)+2\mathbb{P}\left(\left|U_2-\frac{p+1}{n+2}\right|\geq h\right)\\
=&n\delta(h)+4\mathbb{P}\left(\left|U_1-\frac{p-1}{n+2}\right|\geq h\right)+4\mathbb{P}\left(\left|U_2-\frac{p+1}{n+2}\right|\geq h\right)
\end{aligned}
$$

where $U_1\sim \mathrm{Beta}(p-1,n-p+3)$, and $U_2\sim\mathrm{Beta}(p+1,n-p+1)$. Note by Theorem 4.2.1, we know that U_1 and U_2 are sub-Gaussian, and we have $1/(4(n+3))$ as an upper bound for the optimal proxy variances of U_1 and U_2. Then by Lemma 4.2.4 we can find an upper found for the tail of the beta distributions:

$$
\mathbb{P}\left(\left|U_1-\frac{p-1}{n+2}\right|\geq h\right)\leq 2\exp\left(-2h^2(n+3)\right),
$$

$$
\mathbb{P}\left(\left|U_2-\frac{p+1}{n+2}\right|\geq h\right)\leq 2\exp\left(-2h^2(n+3)\right).
$$

So

$$
\begin{aligned}
&\left\|M_{p,n+1}-\bar{S}^{p-2}_{0,\frac{p-1}{n+2}}(\gamma)f(\frac{p-1}{n+2})f(\frac{p+1}{n+2})\bar{S}^{n-p}_{\frac{p+1}{n+2},1}(\gamma)\right\|\\
\leq&n\delta(h)+16\exp\left(-2h^2(n+3)\right)
\end{aligned}
$$

for $h\in(0,1)$.

Using similar arguments, we can find an upper bound for $\left\|\bar{S}_n-\bar{S}^{p-2}_{0,\frac{p-1}{n+2}}(\gamma)f(\frac{p-1}{n+2})f(\frac{p+1}{n+2})\bar{S}^{n-p}_{\frac{p+1}{n+2},1}(\gamma)\right\|$:

$$
\left\|\bar{S}_n-\bar{S}^{p-2}_{0,\frac{p-1}{n+2}}(\gamma)f(\frac{p-1}{n+2})f(\frac{p+1}{n+2})\bar{S}^{n-p}_{\frac{p+1}{n+2},1}(\gamma)\right\|
$$

$$
\begin{aligned}
\leq & \int_{0<u_1<u_2<1} \left\| \bar{S}^{p-2}_{0,u_1}(\gamma) - \bar{S}^{p-2}_{0,\frac{p-1}{n+2}}(\gamma) \right\| n! \frac{u_1^{p-2}}{(p-2)!} \frac{(1-u_2)^{n-p}}{(n-p)!} \mathrm{d}u_1 \mathrm{d}u_2 \\
& + \int_{0<u_1<u_2<1} \left\| f(u_1) - f(\frac{p-1}{n+2}) \right\| n! \frac{u_1^{p-2}}{(p-2)!} \frac{(1-u_2)^{n-p}}{(n-p)!} \mathrm{d}u_1 \mathrm{d}u_2 \\
& + \int_{0<u_1<u_2<1} \left\| f(u_2) - f(\frac{p+1}{n+2}) \right\| n! \frac{u_1^{p-2}}{(p-2)!} \frac{(1-u_2)^{n-p}}{(n-p)!} \mathrm{d}u_1 \mathrm{d}u_2 \\
& + \int_{0<u_1<u_2<1} \left\| \bar{S}^{n-p}_{u_2,1}(\gamma) - \bar{S}^{n-p}_{\frac{p+1}{n+2},1}(\gamma) \right\| n! \frac{u_1^{p-2}}{(p-2)!} \frac{(1-u_2)^{n-p}}{(n-p)!} \mathrm{d}u_1 \mathrm{d}u_2 \\
= & \int_{0<u<1} \left\| \bar{S}^{p-2}_{0,u}(\gamma) - \bar{S}^{p-2}_{0,\frac{p-1}{n+2}}(\gamma) \right\| n! \frac{u^{p-2}}{(p-2)!} \frac{(1-u)^{n-p+1}}{(n-p+1)!} \mathrm{d}u \\
& + \int_{0<u<1} \left\| f(u) - f(\frac{p-1}{n+2}) \right\| n! \frac{u^{p-2}}{(p-2)!} \frac{(1-u)^{n-p+1}}{(n-p+1)!} \mathrm{d}u \\
& + \int_{0<u<1} \left\| f(u) - f(\frac{p+1}{n+2}) \right\| n! \frac{u^{p-1}}{(p-1)!} \frac{(1-u)^{n-p}}{(n-p)!} \mathrm{d}u \\
& + \int_{0<u<1} \left\| \bar{S}^{n-p}_{u,1}(\gamma) - \bar{S}^{n-p}_{\frac{p+1}{n+2},1}(\gamma) \right\| n! \frac{u^{p-1}}{(p-1)!} \frac{(1-u)^{n-p}}{(n-p)!} \mathrm{d}u.
\end{aligned}
$$

Now in the integrals we see the densities of beta variables $U_3 \sim \mathrm{Beta}(p-1, n-p+2)$ and $U_4 \sim \mathrm{Beta}(p, n-p+1)$. Again by Theorem 4.2.1, $1/(4(n+2))$ is an upper bound for the optimal proxy variance of U_3 and the optimal proxy variance of U_4. Then by Lemma 4.2.4, we can find upper bounds for the tails of U_3 and U_4: For $h \in (0,1)$,

$$
\begin{aligned}
& \mathbb{P}\left(\left| U_3 - \frac{p-1}{n+2} \right| \geq h \right) \\
\leq & \exp\left(-2h^2(n+2)\right) \left(\exp(4\frac{p-1}{n+1}h) + \exp(-4\frac{p-1}{n+1}h) \right) \\
\leq & 2\exp\left(-2h^2(n+2) + 4h\right),
\end{aligned}
$$

and

$$
\begin{aligned}
& \mathbb{P}\left(\left| U_4 - \frac{p+1}{n+2} \right| \geq h \right) \\
\leq & \exp\left(-2h^2(n+2)\right) \left(\exp(4\frac{p-n-1}{n+1}h) + \exp(-4\frac{p-n-1}{n+1}h) \right) \\
\leq & 2\exp\left(-2h^2(n+2) + 4h\right).
\end{aligned}
$$

So for $h \in (0,1)$, again by considering the concentration parts and tail parts of the beta distributions, we have

$$
\left\| \bar{S}_n - \bar{S}^{p-2}_{0,\frac{p-1}{n+2}}(\gamma) f(\frac{p-1}{n+2}) f(\frac{p+1}{n+2}) \bar{S}^{n-p}_{\frac{p+1}{n+2},1}(\gamma) \right\|
$$

$$
\begin{aligned}
&\leq n\delta(h) + 4\mathbb{P}\left(\left|U_3 - \frac{p-1}{n+2}\right| \geq h\right) + 4\mathbb{P}\left(\left|U_4 - \frac{p+1}{n+2}\right| \geq h\right) \\
&\leq n\delta(h) + 8\exp\left(-2h^2(n+2) + 4h\right) + 8\exp\left(-2h^2(n+2) + 4h\right) \\
&= n\delta(h) + 16\exp\left(-2h^2(n+2) + 4h\right).
\end{aligned}
$$

Then by triangle inequality, for $h_1, h_2 \in (0,1)$,

$$
\begin{aligned}
&\left\|M_{p,n+1} - \bar{S}_n\right\| \\
\leq &n\delta(h_1) + 16\exp\left(-2h_1^2(n+3)\right) \\
+ &n\delta(h_2) + 16\exp\left(-2h_2^2(n+2) + 4h_2\right).
\end{aligned}
$$

If we choose $h_1 = \sqrt{\frac{\log(n+3)}{n+3}}$, $h_2 = \sqrt{\frac{\log(n+2)}{n+2}}$, then

$$
\|M_{p,n+1} - \bar{S}_n\| \leq n\delta\left(\sqrt{\frac{\log(n+3)}{n+3}}\right) + n\delta\left(\sqrt{\frac{\log(n+2)}{n+2}}\right) + \frac{C_1}{(n+3)^2} + \frac{C_2}{(n+2)^2}
$$

for some constants C_1 and C_2. $\square$

Example 4.2.2. *Following the assumptions in Proposition 4.2.2, in the case when $\delta(h) = h^\rho$ for $\rho \in (0,1]$, we have*

$$
\left\|M_{p,n+1} - \bar{S}_n\right\| \leq n\left(\frac{\log(n+3)}{n+3}\right)^{\frac{\rho}{2}} + n\left(\frac{\log(n+2)}{n+2}\right)^{\frac{\rho}{2}} + \frac{C_1}{(n+3)^2} + \frac{C_2}{(n+2)^2}.
$$

If we use a triangle inequality argument, by the results of Proposition 4.2.1 and Proposition 4.2.2, if a differentiable path $\gamma : [0,1] \to \mathbb{R}^d$ is parametrised at unit speed, and the derivative $f : (0,1) \to \mathbb{R}^d$ has modulus of continuity δ, for $\theta \in (0,1)$, if we choose $p = \lfloor \theta(n+2) \rfloor$, then

$$
\begin{aligned}
&\left\|I_{p,n}(f(\theta)) - \bar{S}_{n+1}\right\| \\
\leq &\left\|R_{p,n+1}(f(\theta)) - \bar{S}_{n+1}\right\| + \left\|M_{p,n+1} - \bar{S}_n\right\| \\
\leq &O\left(n\delta\left(\sqrt{\frac{\log n}{n}}\right) + \frac{1}{n^2}\right).
\end{aligned}
$$

Therefore we need to have a δ function that decays faster than $1/n$ if we want $\left\|I_{p,n}(f(\theta)) - \bar{S}_{n+1}\right\|$ converges to 0 as n increases. Because piecewise linear paths are locally linear, it is interesting to consider such a path, which leads to the topic of our next section.

4.3 Comparing two adjacent levels of the signature of a piecewise linear path

In this section we consider a non-degenerate piecewise linear path $\gamma : [0,1] \to \mathbb{R}^d$ with continuous derivative $f : (0,1) \to \mathbb{R}^d$ such that $\|f(t)\| = 1$ for all $t \in (0,1)$ where it is defined. For $0 \le s < t \le 1$, $n \ge 1$, and $x \in \mathbb{R}^d$, the definitions of $\bar{S}^n_{s,t}(\gamma)$, $I_{p,n}(x)$, $R_{p,n+1}(x)$ and $M_{p,n+1}$ are the same as in Equation (4.3), (4.4), (4.5) and (4.8). We want to show that for a piecewise linear path, $I_{p,n}(x)$ is eventually a good approximation to $\bar{S}_{n+1}$ as n increases.
First we state the modulus of continuity of the normalised signature for a piecewise linear path.

Lemma 4.3.1. *Assume $\gamma : [0,1] \to \mathbb{R}^d$ is a piecewise linear path with derivative $f : (0,1) \to \mathbb{R}^d$ such that $\|f(t)\| = 1$ for all $t \in (0,1)$ if defined. Then for $m \ge 1$, for $s, u \in (0,t)$,*

$$\left\|\bar{S}^m_{u,t}(\gamma) - \bar{S}^m_{s,t}(\gamma)\right\| \le m \sup_{\substack{r_1,r_2 \in (\min(s,u),t) \\ |r_1-r_2| \le |u-s|}} \|f(r_1) - f(r_2)\|.$$

Proof. Define $\lambda = \frac{t-u}{t-s}$. Then we can write

$$\begin{aligned}
&\left\|\bar{S}^m_{u,t}(\gamma) - \bar{S}^m_{s,t}(\gamma)\right\| \\
=&\left\|\frac{m! \int_{s<u_1<\cdots<u_m<t} f(\lambda u_1 + t(1-\lambda)) \cdots f(\lambda u_m + t(1-\lambda)) \mathrm{d}u_1 \cdots \mathrm{d}u_m}{(t-s)^m}\right. \\
&\left. - \frac{m! \int_{s<u_1<\cdots<u_m<t} f(u_1) \cdots f(u_m) \mathrm{d}u_1 \cdots \mathrm{d}u_m}{(t-s)^m}\right\| \\
\le&\frac{m! \int_{s<u_1<\cdots<u_m<t} \sum_{i=1}^m \|f(\lambda u_i + t(1-\lambda)) - f(u_i)\| \mathrm{d}u_1 \cdots \mathrm{d}u_m}{(t-s)^m}.
\end{aligned}$$

Note that for all $u_i \in (s,t)$,

$$0 \le |\lambda u_i + t(1-\lambda) - u_i| \le |u-s|.$$

Hence

$$\begin{aligned}
&\left\|\bar{S}^m_{u,t}(\gamma) - \bar{S}^m_{s,t}(\gamma)\right\| \\
\le&\frac{m! \int_{s<u_1<\cdots<u_m<t} \sum_{i=1}^m \sup_{|r_1-r_2| \le |u-s|} \|f(r_1) - f(r_2)\| \mathrm{d}u_1 \cdots \mathrm{d}u_m}{(t-s)^m} \\
=&m \sup_{\substack{r_1,r_2 \in (\min(s,u),t) \\ |r_1-r_2| \le |u-s|}} \|f(r_1) - f(r_2)\|.
\end{aligned}$$

∎

Lemma 4.3.2. *Assume $\gamma : [0,1] \to \mathbb{R}^d$ is a piecewise linear path with derivative $f : (0,1) \to \mathbb{R}^d$ such that $\|f(t)\| = 1$ for all $t \in (0,1)$ if defined. Then for $u, t \in (s,1)$,*

$$\left\|\bar{S}^m_{s,t}(\gamma) - \bar{S}^m_{s,u}(\gamma)\right\| \leq m \sup_{\substack{r_1, r_2 \in (s, \max(t,u)) \\ |r_1 - r_2| \leq |t-u|}} \|f(r_1) - f(r_2)\|.$$

Proof. Define $\lambda = \frac{t-s}{u-s}$. Then

$$\begin{aligned}
&\left\|\bar{S}^m_{s,t}(\gamma) - \bar{S}^m_{s,u}(\gamma)\right\| \\
=&\left\|\frac{m! \int_{s<u_1<\cdots<u_m<u} f(\lambda u_1 + s(1-\lambda)) \cdots f(\lambda u_m + s(1-\lambda)) \mathrm{d}u_1 \cdots \mathrm{d}u_m}{(u-s)^m}\right. \\
&- \left.\frac{m! \int_{s<u_1<\cdots<u_m<u} f(u_1) \cdots f(u_m) \mathrm{d}u_1 \cdots \mathrm{d}u_m}{(t-s)^M}\right\| \\
\leq& \frac{m! \int_{s<u_1<\cdots<u_m<u} \sum_{i=1}^m \|f(\lambda u_i + s(1-\lambda)) - f(u_i)\| \mathrm{d}u_1 \cdots \mathrm{d}u_m}{(u-s)^m}.
\end{aligned}$$

Note for all $s < u_i < u$,

$$0 \leq |\lambda u_i + s(1-\lambda) - u_i| \leq |t-u|,$$

then

$$\begin{aligned}
&\left\|\bar{S}^m_{s,t}(\gamma) - \bar{S}^m_{s,u}(\gamma)\right\| \\
\leq& \frac{m! \int_{s<u_1<\cdots<u_m<u} \sum_{i=1}^m \sup_{|r_1 - r_2| \leq |t-u|} \|f(r_1) - f(r_2)\| \mathrm{d}u_1 \cdots \mathrm{d}u_m}{(u-s)^m} \\
=& m \sup_{\substack{r_1, r_2 \in (s, \max(t,u)) \\ |r_1 - r_2| \leq |t-u|}} \|f(r_1) - f(r_2)\|.
\end{aligned}$$

□

We now discuss the behaviour of $\left\|I_{p,n}(f(\theta)) - \bar{S}_{n+1}\right\|$ of a piecewise linear path.

Theorem 4.3.1. *Assume $\gamma : [0,1] \to \mathbb{R}^d$ is a piecewise linear path with derivative $f : (0,1) \to \mathbb{R}^d$ such that $\|f(t)\| = 1$ for all $t \in (0,1)$ where it is defined. Assume γ is differentiable at $\theta \in (0,1)$. Choose $p \in \{1, \cdots, n+1\}$ such that $p = \lfloor \theta(n+2) \rfloor$. For all $h \in (0,1)$, $\left\|I_{p,n}(f(\theta)) - \bar{S}_{n+1}\right\| \leq \epsilon_{\theta,n,h}$, where*

$$\begin{aligned}
\epsilon_{\theta,n,h} = O\Bigg(&\sup_{|u-\theta|<h} \|f(\theta) - f(u)\| + \sup_{|u-\theta|<h} \left\|\bar{S}^{p-2}_{0,u}(\gamma) - \bar{S}^{p-2}_{0,\theta}(\gamma)\right\| \\
&+ \sup_{|u-\theta|<h} \left\|\bar{S}^{n-p}_{u,1}(\gamma) - \bar{S}^{n-p}_{\theta,1}(\gamma)\right\| + \exp\left(-2nh^2\right) \Bigg).
\end{aligned}$$

Proof. Using similar ideas discussed in Proposition 4.2.1, for all $h \in (0,1)$,

$$
\begin{aligned}
&\left\|R_{p,n+1}(f(\theta)) - \bar{S}_{n+1}\right\| \\
\leq &\int_{0<u_1<\cdots<u_{n+1}<1} \|f(u_1)\cdots f(u_{p-1})(f(\theta)-f(u_p))\cdots f(u_{n+1})\|(n+1)!\mathrm{d}u_1\cdots \mathrm{d}u_{n+1} \\
\leq &\int_{0<u<1} \|f(\theta)-f(u)\|(n+1)!\frac{u^{p-1}}{(p-1)!}\frac{(1-u)^{n-p+1}}{(n-p+1)!}\mathrm{d}u \\
= &\int_{|u-\theta|<h} \|f(\theta)-f(u)\|(n+1)!\frac{u^{p-1}}{(p-1)!}\frac{(1-u)^{n-p+1}}{(n-p+1)!}\mathrm{d}u \\
&+ \int_{|u-\theta|>h} \|f(\theta)-f(u)\|(n+1)!\frac{u^{p-1}}{(p-1)!}\frac{(1-u)^{n-p+1}}{(n-p+1)!}\mathrm{d}u \\
\leq &\sup_{|u-\theta|<h} \|f(\theta)-f(u)\| + 2\mathbb{P}\left(|U-\theta| \geq h\right),
\end{aligned}
$$

and $U \sim \mathrm{Beta}(p, n-p+2)$, $\mathbb{E}[U] = \frac{p}{n+2}$. By Theorem 4.2.1, an upper bound for the optimal proxy variance of U is $1/(4(n+3))$. Then by Lemma 4.2.4, we can find an upper bound for the tail density of U:

$$
\begin{aligned}
&\mathbb{P}\left(|U-\theta| \geq h\right) \\
\leq &\exp\left(-2(n+3)h^2\right) \\
&\left(\exp\left(4(n+3)h\left(\theta - \frac{p}{n+2}\right)\right) + \exp\left(-4(n+3)h\left(\theta - \frac{p}{n+2}\right)\right)\right) \\
\leq &2\exp\left(-2(n+3)h^2\right)\exp\left(4(n+3)h\frac{1}{n+2}\right) \\
\leq &2\exp\left(-2(n+3)h^2\right)\exp(8h) \\
\leq &2e^8\exp\left(-2(n+3)h^2\right)
\end{aligned}
$$

for $h \in (0,1)$. Then

$$
\left\|R_{p,n+1}(f(\theta)) - \bar{S}_{n+1}\right\| \leq \sup_{|u-\theta|<h} \|f(\theta)-f(u)\| + 2e^8\exp\left(-2(n+3)h^2\right).
$$

Taking $h = \sqrt{\frac{\log(n+3)}{n+3}}$, we have

$$
\left\|R_{p,n+1}(f(\theta)) - \bar{S}_{n+1}\right\| \leq \sup_{|u-\theta|<\sqrt{\frac{\log(n+3)}{n+3}}} \|f(\theta)-f(u)\| + \frac{2e^8}{(n+3)^2}. \tag{4.11}
$$

Note that since γ is differentiable at θ, f is continuous in some neighbourhood near θ. Therefore as n increases, the right-hand side of (4.11) tends to zero.

We now look for an upper bound on $\left\|M_{p,n+1}(x) - \bar{S}_n\right\|$. Similarly as stated in Proposition 4.2.2, we first consider

$$
\left\|M_{p,n+1} - \bar{S}_{0,\theta}^{p-2}(\gamma)f(\theta)f(\theta)\bar{S}_{\theta,1}^{n-p}(\gamma)\right\|
$$

and

$$\left\|\bar{S}_n - \bar{S}_{0,\theta}^{p-2}(\gamma)f(\theta)f(\theta)\bar{S}_{\theta,1}^{n-p}(\gamma)\right\|$$

respectively.

Note for all $h \in (0,1)$,

$$\begin{aligned}
&\left\|M_{p,n+1} - \bar{S}_{0,\theta}^{p-2}(\gamma)f(\theta)f(\theta)\bar{S}_{\theta,1}^{n-p}(\gamma)\right\| \\
=&\Bigg\|\int_{0<u_1<u_2<1}(\bar{S}_{0,u_1}^{p-2}(\gamma)f(u_1)f(u_2)\bar{S}_{u_2,1}^{n-p}(\gamma) - \bar{S}_{0,\theta}^{p-2}(\gamma)f(\theta)f(\theta)\bar{S}_{\theta,1}^{n-p}(\gamma)) \\
&(n+1)!(u_2-u_1)\frac{u_1^{p-2}}{(p-2)!}\frac{(1-u_2)^{n-p}}{(n-p)!}\mathrm{d}u_1\mathrm{d}u_2\Bigg\| \\
&\le \int_{0<u<1}\left\|\bar{S}_{0,u}^{p-2}(\gamma) - \bar{S}_{0,\theta}^{p-2}(\gamma)\right\|(n+1)!\frac{u^{p-2}}{(p-2)!}\frac{(1-u)^{n-p+2}}{(n-p+2)!}\mathrm{d}u \\
&+\int_{0<u<1}\|f(u)-f(\theta)\|(n+1)!\frac{u^{p-2}}{(p-2)!}\frac{(1-u)^{n-p+2}}{(n-p+2)!}\mathrm{d}u \\
&+\int_{0<u<1}\|f(u)-f(\theta)\|(n+1)!\frac{u^{p}}{p!}\frac{(1-u)^{n-p}}{(n-p)!}\mathrm{d}u \\
&+\int_{0<u<1}\left\|\bar{S}_{u,1}^{n-p}(\gamma) - \bar{S}_{\theta,1}^{n-p}(\gamma)\right\|(n+1)!\frac{u^{p}}{p!}\frac{(1-u)^{n-p}}{(n-p)!}\mathrm{d}u \\
=&\int_{|u-\theta|<h}\left\|\bar{S}_{0,u}^{p-2}(\gamma) - \bar{S}_{0,\theta}^{p-2}(\gamma)\right\|(n+1)!\frac{u^{p-2}}{(p-2)!}\frac{(1-u)^{n-p+2}}{(n-p+2)!}\mathrm{d}u \\
&+\int_{|u-\theta|>h}\left\|\bar{S}_{0,u}^{p-2}(\gamma) - \bar{S}_{0,\theta}^{p-2}(\gamma)\right\|(n+1)!\frac{u^{p-2}}{(p-2)!}\frac{(1-u)^{n-p+2}}{(n-p+2)!}\mathrm{d}u \\
&+\int_{|u-\theta|<h}\|f(u)-f(\theta)\|(n+1)!\frac{u^{p-2}}{(p-2)!}\frac{(1-u)^{n-p+2}}{(n-p+2)!}\mathrm{d}u \\
&+\int_{|u-\theta|>h}\|f(u)-f(\theta)\|(n+1)!\frac{u^{p-2}}{(p-2)!}\frac{(1-u)^{n-p+2}}{(n-p+2)!}\mathrm{d}u \\
&+\int_{|u-\theta|<h}\|f(u)-f(\theta)\|(n+1)!\frac{u^{p}}{p!}\frac{(1-u)^{n-p}}{(n-p)!}\mathrm{d}u \\
&+\int_{|u-\theta|>h}\|f(u)-f(\theta)\|(n+1)!\frac{u^{p}}{p!}\frac{(1-u)^{n-p}}{(n-p)!}\mathrm{d}u \\
&+\int_{|u-\theta|<h}\left\|\bar{S}_{u,1}^{n-p}(\gamma) - \bar{S}_{\theta,1}^{n-p}(\gamma)\right\|(n+1)!\frac{u^{p}}{p!}\frac{(1-u)^{n-p}}{(n-p)!}\mathrm{d}u \\
&+\int_{|u-\theta|>h}\left\|\bar{S}_{u,1}^{n-p}(\gamma) - \bar{S}_{\theta,1}^{n-p}(\gamma)\right\|(n+1)!\frac{u^{p}}{p!}\frac{(1-u)^{n-p}}{(n-p)!}\mathrm{d}u \\
\le&\sup_{|u-\theta|<h}\left\|\bar{S}_{0,u}^{p-2}(\gamma) - \bar{S}_{0,\theta}^{p-2}(\gamma)\right\| + 2\mathbb{P}\left(|U_1-\theta|\ge h\right) \\
&+\sup_{|u-\theta|<h}\|f(u)-f(\theta)\| + 2\mathbb{P}\left(|U_1-\theta|\ge h\right)
\end{aligned}$$

$$
\begin{aligned}
&+\sup_{|u-\theta|<h}\|f(u)-f(\theta)\|+2\mathbb{P}\left(|U_2-\theta|\geq h\right)\\
&+\sup_{|u-\theta|<h}\left\|\bar{S}^{n-p}_{u,1}(\gamma)-\bar{S}^{n-p}_{\theta,1}(\gamma)\right\|+2\mathbb{P}\left(|U_2-\theta|\geq h\right),
\end{aligned}
$$

where $U_1\sim\mathrm{Beta}(p-1,n-p+3)$, and $U_2\sim\mathrm{Beta}(p+1,n-p+1)$. By Theorem 4.2.1, an upper bound for both the optimal proxy variance of U_1 and the optimal proxy variance of U_2 is $1/(4(n+3))$, then by Lemma 4.2.4,

$$
\begin{aligned}
&\mathbb{P}\left(|U_1-\theta|\geq h\right)\\
\leq&\exp\left(-2h^2(n+3)\right)\\
&\left(\exp\left(4(n+3)h\left(\theta-\frac{p-1}{n+2}\right)\right)+\exp\left(-4(n+3)h\left(\theta-\frac{p-1}{n+2}\right)\right)\right)\\
\leq&\exp\left(-2h^2(n+3)\right)\left(\exp\left(4(n+3)h\frac{2}{n+2}\right)+\exp\left(4(n+3)h\frac{2}{n+2}\right)\right)\\
\leq&2\exp\left(-2h^2(n+3)\right)\exp(16h)\\
\leq&2e^{16}\exp\left(-2h^2(n+3)\right)
\end{aligned}
$$

for $h\in(0,1)$. Similarly, we can show that

$$
\mathbb{P}\left(|U_2-\theta|\geq h\right)\leq 2e^{16}\exp\left(-2h^2(n+3)\right).
$$

Then for all $h\in(0,1)$,

$$
\begin{aligned}
&\left\|M_{p,n+1}-\bar{S}^{p-2}_{0,\theta}f(\theta)f(\theta)\bar{S}^{n-p}_{\theta,1}(\gamma)\right\|\\
\leq&\sup_{|u-\theta|<h}\left\|\bar{S}^{p-2}_{0,u}(\gamma)-\bar{S}^{p-2}_{0,\theta}(\gamma)\right\|+\sup_{|u-\theta|<h}\left\|\bar{S}^{n-p}_{u,1}(\gamma)-\bar{S}^{n-p}_{\theta,1}(\gamma)\right\|\\
&+2\sup_{|u-\theta|<h}\|f(u)-f(\theta)\|+C_1\exp\left(-2h^2(n+3)\right)
\end{aligned}
$$

where C_1 is some constant. Similarly,

$$
\begin{aligned}
&\left\|\bar{S}_n-\bar{S}^{p-2}_{0,\theta}(\gamma)f(\theta)f(\theta)\bar{S}^{n-p}_{\theta,1}(\gamma)\right\|\\
=&\Bigg\|\int_{0<u_1<u_2<1}(\bar{S}^{p-2}_{0,u_1}(\gamma)f(u_1)f(u_2)\bar{S}^{n-p}_{u_2,1}(\gamma)-\bar{S}^{p-2}_{0,\theta}(\gamma)f(\theta)f(\theta)\bar{S}^{n-p}_{\theta,1}(\gamma))\\
&n!\frac{u_1^{p-2}}{(p-2)!}\frac{(1-u_2)^{n-p}}{(n-p)!}\mathrm{d}u_1\mathrm{d}u_2\Bigg\|\\
\leq&\int_{0<u<1}\left\|\bar{S}^{n-p}_{0,u}(\gamma)-\bar{S}^{p-2}_{0,\theta}(\gamma)\right\|n!\frac{u^{p-2}}{(p-2)!}\frac{(1-u)^{n-p+1}}{(n-p+1)!}\mathrm{d}u\\
&+\int_{0<u<1}\|f(u)-f(\theta)\|n!\frac{u^{p-2}}{(p-2)!}\frac{(1-u)^{n-p+1}}{(n-p+1)!}\mathrm{d}u\\
&+\int_{0<u<1}\|f(u)-f(\theta)\|n!\frac{u^{p-1}}{(p-1)!}\frac{(1-u)^{n-p}}{(n-p)!}\mathrm{d}u
\end{aligned}
$$

$$
\begin{aligned}
&+\int_{0<u<1}\left\|\bar{S}^{n-p}_{u,1}(\gamma)-\bar{S}^{n-p}_{\theta,1}(\gamma)\right\| n!\frac{u^{p-1}}{(p-1)!}\frac{(1-u)^{n-p}}{(n-p)!}\mathrm{d}u\\
\leq &\sup_{|u-\theta|<h'}\left\|\bar{S}^{p-2}_{0,u}(\gamma)-\bar{S}^{p-2}_{0,\theta}(\gamma)\right\|+2\mathbb{P}\left(|U_3-\theta|\geq h'\right)\\
&+\sup_{|u-\theta|<h'}\|f(u)-f(\theta)\|+2\mathbb{P}\left(|U_3-\theta|\geq h'\right)\\
&+\sup_{|u-\theta|<h'}\|f(u)-f(\theta)\|+2\mathbb{P}\left(|U_4-\theta|\geq h'\right)\\
&+\sup_{|u-\theta|<h'}\left\|\bar{S}^{n-p}_{u,1}(\gamma)-\bar{S}^{n-p}_{\theta,1}(\gamma)\right\|+2\mathbb{P}\left(|U_4-\theta|\geq h'\right),
\end{aligned}
$$

for all $h'\in(0,1)$. Here $U_3\sim\mathrm{Beta}(p-1,n-p+2)$, and $U_4\sim\mathrm{Beta}(p,n-p+1)$. Note $\mathbb{E}[U_3]=\frac{p-1}{n+1}$, $\mathbb{E}[U_4]=\frac{p}{n+1}$. By Theorem 4.2.1, an upper bound for both the optimal proxy variance of U_3 and the optimal proxy variance of U_4 is $1/(4(n+2))$. Note that

$$
\begin{aligned}
\left|\theta-\frac{p-1}{n+1}\right| &\leq\left|\theta-\frac{p-1}{n+2}\right|+\left|\frac{p-1}{n+2}-\frac{p-1}{n+1}\right|\\
&\leq\frac{2}{n+2}+\frac{p-1}{(n+1)(n+2)}.
\end{aligned}
$$

Then by Lemma 4.2.4,

$$
\begin{aligned}
&\mathbb{P}\left(|U_3-\theta|\geq h'\right)\\
\leq &\exp\left(-2(n+2)h'^2\right)\\
&\left(\exp\left(4(n+2)h'\left(\theta-\frac{p-1}{n+1}\right)\right)+\exp\left(-4(n+2)h'\left(\theta-\frac{p-1}{n+1}\right)\right)\right)\\
\leq &2\exp\left(-2(n+2)h'^2\right)\exp\left(4(n+2)h'\left(\frac{2}{n+2}+\frac{p-1}{(n+1)(n+2)}\right)\right)\\
\leq &2\exp\left(-2(n+2)h'^2\right)\exp(12h')\\
\leq &2e^{12}\exp\left(-2(n+2)h'^2\right),
\end{aligned}
$$

and similarly,

$$
\begin{aligned}
&\mathbb{P}\left(|U_4-\theta|\geq h'\right)\\
\leq &\exp\left(-2(n+2)h'^2\right)\\
&\left(\exp\left(4(n+2)h'\left(\theta-\frac{p}{n+1}\right)\right)+\exp\left(-4(n+2)h'\left(\theta-\frac{p}{n+1}\right)\right)\right)\\
\leq &\exp\left(-2(n+2)h'^2\right)\left(\exp\left(4(n+2)h'\left(\frac{1}{n+2}+\frac{p}{(n+1)(n+2)}\right)\right)\right.\\
&\left.+\exp\left(-4(n+2)h'\left(\frac{1}{n+2}+\frac{p}{(n+1)(n+2)}\right)\right)\right)
\end{aligned}
$$

$$
\begin{aligned}
&\leq 2\exp\left(-2(n+2)h'^2\right)\exp(8h')\\
&\leq 2e^8\exp\left(-2(n+2)h'^2\right).
\end{aligned}
$$

Then for all $h' \in (0,1)$, we have

$$
\begin{aligned}
&\left\|\bar{S}_n - \bar{S}_{0,\theta}^{p-2}(\gamma)f(\theta)f(\theta)\bar{S}_{\theta,1}^{n-p}(\gamma)\right\|\\
\leq &\sup_{|u-\theta|<h'}\left\|\bar{S}_{0,u}^{p-2}(\gamma)-\bar{S}_{0,\theta}^{p-2}(\gamma)\right\| + \sup_{|u-\theta|<h'}\left\|\bar{S}_{u,1}^{n-p}(\gamma)-\bar{S}_{\theta,1}^{n-p}(\gamma)\right\|\\
&+2\sup_{|u-\theta|<h'}\|f(u)-f(\theta)\| + C_2\exp\left(-2(n+2)h'^2\right),
\end{aligned}
$$

where C_2 is some constant.
Hence by triangle inequality, for all $h, h' \in (0,1)$,

$$
\begin{aligned}
&\left\|M_{p,n+1}-\bar{S}_n\right\|\\
\leq &\sup_{|u-\theta|<h}\left\|\bar{S}_{0,u}^{p-2}(\gamma)-\bar{S}_{0,\theta}^{p-2}(\gamma)\right\| + \sup_{|u-\theta|<h}\left\|\bar{S}_{u,1}^{n-p}(\gamma)-\bar{S}_{\theta,1}^{n-p}(\gamma)\right\|\\
&+2\sup_{|u-\theta|<h}\|f(u)-f(\theta)\| + C_1\exp\left(-2h^2(n+3)\right)\\
&+\sup_{|u-\theta|<h'}\left\|\bar{S}_{0,u}^{p-2}(\gamma)-\bar{S}_{0,\theta}^{p-2}(\gamma)\right\| + \sup_{|u-\theta|<h'}\left\|\bar{S}_{u,1}^{n-p}(\gamma)-\bar{S}_{\theta,1}^{n-p}(\gamma)\right\|\\
&+2\sup_{|u-\theta|<h'}\|f(u)-f(\theta)\| + C_2\exp\left(-2(n+2)h'^2\right).
\end{aligned}
$$

Note the result of Lemma 4.2.5 also holds for piecewise linear paths, so

$$
\|I_{p,n}(f(\theta))-R_{p,n+1}(f(\theta))\| = \left\|M_{p,n+1}-\bar{S}_n\right\|,
$$

then for all $h, h', t \in (0,1)$,

$$
\begin{aligned}
&\left\|I_{p,n+1}(f(\theta))-\bar{S}_{n+1}\right\|\\
\leq &\|I_{p,n}(f(\theta))-R_{p,n+1}(f(\theta))\| + \left\|R_{p,n+1}(f(\theta))-\bar{S}_{n+1}\right\|\\
\leq &\sup_{|u-\theta|<h}\left\|\bar{S}_{0,u}^{p-2}(\gamma)-\bar{S}_{0,\theta}^{p-2}(\gamma)\right\| + \sup_{|u-\theta|<h}\left\|\bar{S}_{u,1}^{n-p}(\gamma)-\bar{S}_{\theta,1}^{n-p}(\gamma)\right\|\\
&+2\sup_{|u-\theta|<h}\|f(u)-f(\theta)\| + C_1\exp\left(-2(n+3)h^2\right)\\
&+\sup_{|u-\theta|<h'}\left\|\bar{S}_{0,u}^{p-2}(\gamma)-\bar{S}_{0,\theta}^{p-2}(\gamma)\right\| + \sup_{|u-\theta|<h'}\left\|\bar{S}_{u,1}^{n-p}(\gamma)-\bar{S}_{\theta,1}^{n-p}(\gamma)\right\|\\
&+2\sup_{|u-\theta|<h'}\|f(u)-f(\theta)\| + C_2\exp\left(-2(n+2)h'^2\right)\\
&+\sup_{|u-\theta|<t}\|f(\theta)-f(u)\| + 2e^8\exp\left(-2(n+3)t^2\right).
\end{aligned}
$$

Define for all $h \in (0,1)$,

$$\begin{aligned}\epsilon_{\theta,n,h} :=& 2 \sup_{|u-\theta|<h} \left\| \bar{S}^{p-2}_{0,u}(\gamma) - \bar{S}^{p-2}_{0,\theta}(\gamma) \right\| + 2 \sup_{|u-\theta|<h} \left\| \bar{S}^{n-p}_{u,1}(\gamma) - \bar{S}^{n-p}_{\theta,1}(\gamma) \right\| \\ &+ 5 \sup_{|u-\theta|<h} \|f(u) - f(\theta)\| + C_1 \exp\left(-2(n+3)h^2\right) \\ &+ C_2 \exp\left(-2(n+2)h^2\right) + 2e^8 \exp\left(-2(n+3)h^2\right),\end{aligned}$$

therefore $\epsilon_{\theta,n,h}$ is as required. $\square$

Remark 4.3.1. *Let us continue the setting in Theorem 4.3.1.*
Note that if $S_n = 0$, then

$$\left\| I_{p,n}(x) - \bar{S}_{n+1} \right\| = \left\| \bar{S}_{n+1} \right\| \quad \forall\, x \in \mathbb{R}^d,$$

therefore the difference $\left\| I_{p,n}(x) - \bar{S}_{n+1} \right\|$ only depends on $\bar{S}_{n+1}$ if $S_n = 0$.
If $S_{n+1} = 0$, then

$$\left\| I_{p,n}(x) - \bar{S}_{n+1} \right\| = \left\| \bar{S}_n \right\| \|x\| = \left\| \bar{S}_n \right\| \quad \forall\, x \in \mathbb{R}^d,\ \|x\| = 1,$$

hence $\left\| I_{p,n}(x) - \bar{S}_{n+1} \right\|$ only depends on $\bar{S}_n$ if $S_{n+1} = 0$.
In both cases above, $\left\| I_{p,n}(x) - \bar{S}_{n+1} \right\|$ does not depend on x, and we cannot extract any useful information about the underlying path by varying x. Therefore it is more interesting to consider the case when S_n and S_{n+1} are non-zero.
Now suppose the underlying path γ is such that there exists $N \in \mathbb{N}$ such that $S_n \neq 0$ for all $n \geq N$. Let us try to understand the upper bound $\epsilon_{\theta,n,h}$ obtained in Theorem 4.3.1. Essentially, we expect that given a suitable choice of $h \in (0,1)$ depending on n, $\epsilon_{\theta,n,h}$ converges to zero as n goes to infinity. Note that since γ is piecewise linear and differentiable at θ, there exists $t \in (0,1)$ such that for $u \in (\theta - t, \theta + t)$, $\|f(u) - f(\theta)\| = 0$. Therefore the key is to understand the behaviour of other terms in $\epsilon_{\theta,n,h}$.
Let us first look at the term $\exp(-2nh^2)$. In order for this term to go to zero, we shall choose h such that

$$h\sqrt{n} \to \infty \quad \textit{as} \quad n \to \infty,$$

in other words, if the spread of the concentration area is too small, the the tail of a sub-Gaussian distribution would not be neglectable.
On the other hand, $\sup_{|u-\theta|<h} \|\bar{S}^{p-2}_{0,u}(\gamma) - \bar{S}^{p-2}_{0,\theta}(\gamma)\|$ and $\sup_{|u-\theta|<h} \|\bar{S}^{n-p}_{u,1}(\gamma) - \bar{S}^{n-p}_{\theta,1}(\gamma)\|$ clearly would be closer to zero when h is smaller. Therefore there is a trade-off between

the concentration part and the tail part, and the question is whether there exists an $h^* \in (0,1)$ such that $h^*\sqrt{n} \to \infty$ as $n \to \infty$, and

$$\sup_{|u-\theta|<h^*} \left\| \bar{S}^{p-2}_{0,u}(\gamma) - \bar{S}^{p-2}_{0,\theta}(\gamma) \right\| \to 0 \quad as \quad n \to \infty, \tag{4.12}$$

as well as

$$\sup_{|u-\theta|<h^*} \left\| \bar{S}^{n-p}_{u,1}(\gamma) - \bar{S}^{n-p}_{\theta,1}(\gamma) \right\| \to 0 \quad as \quad n \to \infty. \tag{4.13}$$

Note that in this case Lemma 4.3.1 and Lemma 4.3.2 cannot provide converging upper bounds on the modulus of continuity of the normalised signature as the degree increases. Another way is to look at the derivative of the normalised signature. Note that

$$\sup_{|u-\theta|<h} \left\| \bar{S}^{p-2}_{0,u}(\gamma) - \bar{S}^{p-2}_{0,\theta}(\gamma) \right\| \leq \sup_{|t-\theta|<h} \left\| \frac{\partial \bar{S}^{p-2}_{0,t}(\gamma)}{\partial t} \right\| h,$$

and

$$\sup_{|u-\theta|<h} \left\| \bar{S}^{n-p}_{u,1}(\gamma) - \bar{S}^{m}_{\theta,1}(\gamma) \right\| \leq \sup_{|t-\theta|<h} \left\| \frac{\partial \bar{S}^{n-p}_{t,1}(\gamma)}{\partial t} \right\| h,$$

where

$$\frac{\partial \bar{S}^{p-2}_{0,u}(\gamma)}{\partial u} = \frac{p-2}{u} \left(\bar{S}^{p-3}_{0,u}(\gamma) \otimes f(u) - \bar{S}^{p-2}_{0,u}(\gamma) \right),$$

and

$$\frac{\partial \bar{S}^{n-p}_{u,1}(\gamma)}{\partial u} = \frac{n-p}{1-u} \left(f(u) \otimes \bar{S}^{n-p-1}_{u,1}(\gamma) + \bar{S}^{n-p}_{u,1}(\gamma) \right).$$

Recall that $p = \lfloor \theta(n+2) \rfloor$. In order to satisfy Equation (4.12) and (4.13), we need to have

$$n^{1/2} \left\| \bar{S}^{p-3}_{0,u}(\gamma) \otimes f(u) - \bar{S}^{p-2}_{0,u}(\gamma) \right\| \to 0 \quad as \quad n \to \infty, \tag{4.14}$$

and

$$n^{1/2} \left\| f(u) \otimes \bar{S}^{n-p-1}_{u,1}(\gamma) + \bar{S}^{n-p}_{u,1}(\gamma) \right\| \to 0 \quad as \quad n \to \infty. \tag{4.15}$$

We have not been able to prove (4.14) and (4.15), and it is still an open question.

4.4 A converging upper bound for $\left\|I_{p,n}\left(f\left(\theta\right)\right)-\bar{S}_{n+1}\right\|$

Theorem 4.3.1 was our attempt to prove that $\left\|I_{p,n}(f(\theta))-\bar{S}_{n+1}\right\|$ converging to zero using the sub-Gaussian property of the beta distributions, and as stated in Remark 4.3.1, there are some unsolved questions which need to be answered in order for this method to work. Intuitively it is reasonable to expect that if the derivative of the underlying path is inserted at the 'correct' position into the n-th term in the normalised signature, the resulting tensor shall be a well-behaved approximation of the $(n+1)$-th term in the normalised signature, and as we have a finer partition of the interval, the approximation should be more accurate. We first note the following theorem by Hoeffding [20].

Theorem 4.4.1 (Hoeffding's inequality [20])**.** *Let $X_1,\cdots,X_n$ be independent random variables strictly bounded by the intervals $[a_i,b_i]$ respectively, define $S_n=\sum_{i=1}^{n}X_i$. Then for any $t>0$,*

$$\mathbb{P}\left(S_n-\mathbb{E}[S_n]\geq t\right)\leq\exp\left(-\frac{2t^2}{\sum_{i=1}^{n}(b_i-a_i)^2}\right),$$

$$\mathbb{P}\left(|S_n-\mathbb{E}[S_n]|\geq t\right)\leq 2\exp\left(-\frac{2t^2}{\sum_{i=1}^{n}(b_i-a_i)^2}\right).$$

Notice since a binomial variable is a sum of independent Bernoulli variables, Hoeffding's inequality applies to binomial variables. We note the following example.

Example 4.4.1. *Assume $\tilde{\gamma}:[0,1]\to\mathbb{R}^d$ is a linear path with derivative $g:(0,1)\to\mathbb{R}^d$ such that $\|g(t)\|=1$ for all $t\in(0,1)$. Then for any $\theta\in(0,1)$ and $q=\lfloor\theta(n+2)\rfloor$,*

$$\left\|I_{q,n}(g(\theta))-\bar{S}_{n+1}\right\|=\left\|n!\frac{(g(\theta))^{\otimes(n+1)}}{n!}-(n+1)!\frac{(g(\theta))^{\otimes(n+1)}}{(n+1)!}\right\|=0,$$

therefore

$$\left\|I_{q,n}(g(\theta))-\bar{S}_{n+1}\right\|\to 0\quad as\quad n\to\infty.$$

Now let us consider a slightly more complicated case. Assume $\{e_1,e_2\}$ is a basis of $\mathbb{R}^2$ and $\gamma:[0,1]\to\mathbb{R}^2$ is a piecewise linear path such that

$$\gamma(t)=\begin{cases}te_2 & t\in[0,\frac{2}{3}]\\(t-\frac{2}{3})e_1+\frac{2}{3}e_2 & t\in(\frac{2}{3},1].\end{cases}$$

Note that the derivative $f : (0,1) \to \mathbb{R}^2$ of γ satisfies $\|f(t)\| = 1$ for all $t \in (0,1)$ where f is defined. Note in this case,

$$\bar{S}_n = n! \sum_{k=0}^{n} \left(\frac{2}{3}\right)^k \frac{e_2^{\otimes k}}{k!} \otimes \left(\frac{1}{3}\right)^{n-k} \frac{e_1^{\otimes(n-k)}}{(n-k)!}.$$

Note that if we choose $\theta = 1/2$ and $p = \lfloor \theta(n+2) \rfloor$, then $f(\frac{1}{2}) = e_2$, and we can write

$$\begin{aligned} I_{p,n}\left(f\left(\frac{1}{2}\right)\right) =& n! \sum_{k=p-1}^{n} \left(\frac{2}{3}\right)^k \frac{e_2^{\otimes(k+1)}}{k!} \left(\frac{1}{3}\right)^{n-k} \frac{e_1^{\otimes(n-k)}}{(n-k)!} \\ &+ n! \sum_{k=0}^{p-2} \left(\frac{2}{3}\right)^k \frac{e_2^{\otimes k}}{k!} \left(\frac{1}{3}\right)^{n-k} \frac{e_1^{\otimes(p-1-k)} \otimes e_2 \otimes e_1^{\otimes(n-p+1)}}{(n-k)!}, \end{aligned}$$

then

$$\begin{aligned} &I_{p,n}\left(f\left(\frac{1}{2}\right)\right) - \bar{S}_{n+1} \\ =& n! \sum_{k=p-1}^{n} \left(\frac{2}{3}\right)^k \frac{e_2^{\otimes(k+1)}}{k!} \left(\frac{1}{3}\right)^{n-k} \frac{e_1^{\otimes(n-k)}}{(n-k)!} \\ &+ n! \sum_{k=0}^{p-2} \left(\frac{2}{3}\right)^k \frac{e_2^{\otimes k}}{k!} \left(\frac{1}{3}\right)^{n-k} \frac{e_1^{\otimes(p-1-k)} \otimes e_2 \otimes e_1^{\otimes(n-p+1)}}{(n-k)!} \\ &- (n+1)! \sum_{k=p}^{n+1} \left(\frac{2}{3}\right)^k \frac{e_2^{\otimes k}}{k!} \left(\frac{1}{3}\right)^{n+1-k} \frac{e_1^{\otimes(n+1-k)}}{(n+1-k)!} \\ &- (n+1)! \sum_{k=0}^{p-1} \left(\frac{2}{3}\right)^k \frac{e_2^{\otimes k}}{k!} \left(\frac{1}{3}\right)^{n+1-k} \frac{e_1^{\otimes(n+1-k)}}{(n+1-k)!} \\ =& \sum_{k=p-1}^{n} \left(\frac{2}{3}\right)^{k+1} \left(\frac{1}{3}\right)^{n-k} \frac{(n+1)!}{(k+1)!(n-k)!} \left(\frac{3}{2}\frac{k+1}{n+1} - 1\right) e_2^{\otimes(k+1)} e_1^{\otimes(n-k)} \\ &+ \sum_{k=0}^{p-2} \left(\frac{2}{3}\right)^k \left(\frac{1}{3}\right)^{n-k} \frac{n!}{k!(n-k)!} e_2^{\otimes k} e_1^{\otimes(p-1-k)} \otimes e_2 \otimes e_1^{\otimes(n-p+1)} \\ &- \sum_{k=0}^{p-1} \left(\frac{2}{3}\right)^k \left(\frac{1}{3}\right)^{n+1-k} \frac{(n+1)!}{k!(n+1-k)!} e_2^{\otimes k} e_1^{\otimes(n+1-k)}. \end{aligned}$$

Hence

$$\begin{aligned} \left\| I_{p,n}\left(f\left(\frac{1}{2}\right)\right) - \bar{S}_{n+1} \right\| \leq & \sum_{k=p-1}^{n} \left(\frac{2}{3}\right)^{k+1} \left(\frac{1}{3}\right)^{n-k} \frac{(n+1)!}{(k+1)!(n-k)!} \left|\frac{3}{2}\frac{k+1}{n+1} - 1\right| \\ &+ \sum_{k=0}^{p-2} \left(\frac{2}{3}\right)^k \left(\frac{1}{3}\right)^{n-k} \frac{n!}{k!(n-k)!} \end{aligned}$$

$$
\begin{aligned}
&+\sum_{k=0}^{p-1}\left(\frac{2}{3}\right)^k\left(\frac{1}{3}\right)^{n+1-k}\frac{(n+1)!}{k!(n+1-k)!}\\
=&\sum_{k=p}^{n+1}\left(\frac{2}{3}\right)^k\left(\frac{1}{3}\right)^{n+1-k}\frac{(n+1)!}{k!(n+1-k)!}\left|\frac{3}{2}\frac{k}{n+1}-1\right|\\
&+\sum_{k=0}^{p-2}\left(\frac{2}{3}\right)^k\left(\frac{1}{3}\right)^{n-k}\frac{n!}{k!(n-k)!}\\
&+\sum_{k=0}^{p-1}\left(\frac{2}{3}\right)^k\left(\frac{1}{3}\right)^{n+1-k}\frac{(n+1)!}{k!(n+1-k)!}. \qquad (4.16)
\end{aligned}
$$

Let us investigate the binomial sums on the right-hand side of (4.16) respectively. Note

$$
\begin{aligned}
&\sum_{k=p}^{n+1}\left(\frac{2}{3}\right)^k\left(\frac{1}{3}\right)^{n+1-k}\frac{(n+1)!}{k!(n+1-k)!}\left|\frac{3}{2}\frac{k}{n+1}-1\right|\\
\leq&\sum_{k=0}^{n+1}\left(\frac{2}{3}\right)^k\left(\frac{1}{3}\right)^{n+1-k}\frac{(n+1)!}{k!(n+1-k)!}\left|\frac{3}{2}\frac{k}{n+1}-1\right|\\
=&\sum_{\substack{t=k/(n+1)\\ k=0,\cdots,n+1}}\left(\frac{2}{3}\right)^{(n+1)t}\left(\frac{1}{3}\right)^{(n+1)(1-t)}\frac{(n+1)!}{((n+1)t)!\,((n+1)(1-t))!}\left|\frac{3}{2}t-1\right|. \qquad (4.17)
\end{aligned}
$$

Assume $X\sim$Binomial$(n+1,2/3)$. Note that (4.17) is the expectation of a function of the random variable $Y:=X/(n+1)$, and

$$
\mathbb{E}[Y]=\frac{1}{n+1}\mathbb{E}[X]=\frac{2}{3},\quad Var[Y]=\frac{1}{(n+1)^2}Var[X]=\frac{2}{9}\frac{1}{n+1},
$$

hence we can see that as n increases, the distribution of Y will be mostly concentrated around $\mathbb{E}[Y]$, and since $\left|\frac{3}{2}Y-1\right|=0$ at $Y=\mathbb{E}[Y]$, the value of the sum of (4.17) would converge to zero as n increases. For a more formal argument, we have for any $\lambda>0$,

$$
\begin{aligned}
&\mathbb{P}\left(\left|Y-\frac{2}{3}\right|\geq\frac{\lambda}{3}\sqrt{\frac{2}{n+1}}\right)\\
=&\mathbb{P}\left(\left|X-\frac{2}{3}(n+1)\right|\geq\frac{\lambda}{3}\sqrt{2(n+1)}\right)\\
\leq&2\exp\left(-\frac{4}{9}\lambda^2\right),
\end{aligned}
$$

where the last inequality comes from Hoeffding's inequality.

Then for any $\lambda>0$,

$$
\sum_{k=p}^{n+1}\left(\frac{2}{3}\right)^k\left(\frac{1}{3}\right)^{n+1-k}\frac{(n+1)!}{k!(n+1-k)!}\left|\frac{3}{2}\frac{k}{n+1}-1\right|
$$

$$\leq \sum_{\substack{t=k/(n+1)\\ k=0,\cdots,n+1\\ \left|t-\frac{2}{3}\right|<\frac{\lambda}{3}\sqrt{\frac{2}{n+1}}}} \left(\frac{2}{3}\right)^{(n+1)t}\left(\frac{1}{3}\right)^{(n+1)(1-t)} \frac{(n+1)!}{((n+1)t)!((n+1)(1-t))!}\left|\frac{3}{2}t-1\right|$$

$$+ \sum_{\substack{t=k/(n+1)\\ k=0,\cdots,n+1\\ \left|t-\frac{2}{3}\right|\geq\frac{\lambda}{3}\sqrt{\frac{2}{n+1}}}} \left(\frac{2}{3}\right)^{(n+1)t}\left(\frac{1}{3}\right)^{(n+1)(1-t)} \frac{(n+1)!}{((n+1)t)!((n+1)(1-t))!}\left|\frac{3}{2}t-1\right|$$

$$\leq \frac{\lambda}{\sqrt{2(n+1)}} + 2\mathbb{P}\left(\left|Y-\frac{2}{3}\right| \geq \frac{\lambda}{3}\sqrt{\frac{2}{n+1}}\right)$$

$$\leq \frac{\lambda}{\sqrt{2(n+1)}} + 4\exp\left(-\frac{4}{9}\lambda^2\right).$$

Note that for a given $n>0$, $\lambda/\sqrt{2(n+1)}$ *is a strictly increasing linear function of* λ *from* $(0,\infty)$ *to* $(0,\infty)$, *and* $4\exp\left(-\frac{4}{9}\lambda^2\right)$ *is a strictly decreasing function of* λ *from* $(0,\infty)$ *to* $(0,4)$. *So for each* n, *there exists* $\lambda_n>0$ *such that*

$$\frac{\lambda_n}{\sqrt{2(n+1)}} = 4\exp\left(-\frac{4}{9}\lambda_n^2\right). \tag{4.18}$$

Differentiating (4.18) with respect to n *gives*

$$\frac{\partial\lambda_n}{\partial n}(2(n+1))^{\frac{1}{2}} - \lambda_n(2(n+1))^{-\frac{3}{2}} = -\frac{32}{9}\frac{\partial\lambda_n}{\partial n}\lambda_n\exp\left(-\frac{4}{9}\lambda_n^2\right),$$

$$\frac{\partial\lambda_n}{\partial n}\left((2(n+1))^{\frac{1}{2}} + \frac{32}{9}\lambda_n\exp\left(-\frac{4}{9}\lambda_n^2\right)\right) = \lambda_n\left(2(n+1)\right)^{-\frac{3}{2}},$$

therefore $\frac{\partial\lambda_n}{\partial n}>0$, *and* λ_n *is a strictly increasing function in* n *which tends to infinity, so* $4\exp\left(-4/9\lambda_n^2\right)$ *is a strictly decreasing function in* n, *and*

$$\begin{aligned}\sum_{k=p}^{n+1}\left(\frac{2}{3}\right)^k\left(\frac{1}{3}\right)^{n+1-k}\frac{(n+1)!}{k!(n+1-k)!}\left|\frac{3}{2}\frac{k}{n+1}-1\right| \leq & \frac{2\lambda_n}{\sqrt{2(n+1)}}\\ =& 8\exp\left(-\frac{4}{9}\lambda_n^2\right)\\ \to& 0\end{aligned}$$

as n *goes to infinity.*

For the other two binomial sums in (4.16), by Hoeffding's inequality,

$$\sum_{k=0}^{p-2}\left(\frac{2}{3}\right)^k\left(\frac{1}{3}\right)^{n-k}\frac{n!}{k!(n+1-k)!} + \sum_{k=0}^{p-1}\left(\frac{2}{3}\right)^k\left(\frac{1}{3}\right)^{n+1-k}\frac{(n+1)!}{k!(n+1-k)!}$$

$$\leq \exp\left(-C_1 n + B_1\right) + \exp\left(-C_2 n + B_2\right),$$

where $C_1 > 0$, $C_2 > 0$, B_1 and B_2 are some constants. Therefore

$$\left\| I_{p,n}\left(\frac{1}{2}\right) - \bar{S}_{n+1} \right\| \leq \frac{2\lambda_n}{\sqrt{2(n+1)}} + \exp\left(-C_1 n + B_1\right) + \exp\left(-C_2 n + B_2\right) \to 0$$

as $n \to \infty$. Note since $2\lambda_n / \sqrt{2(n+1)}$ is decreasing in n, and λ_n is increasing in n, the rate of increasing of λ_n must be smaller than the increasing rate of $\sqrt{2(n+1)}$, therefore the upper bound obtained for $\left\| I_{p,n}\left(\frac{1}{2}\right) - \bar{S}_{n+1} \right\|$ decreases at a rate slower than $O(1/\sqrt{n})$.

Example 4.4.1 has inspired us that we can use the tail behaviour of the binomial distribution to prove that $\left\| I_{p,n}\left(f(\theta)\right) - \bar{S}_{n+1} \right\|$ converges to 0.

Theorem 4.4.2. *Suppose $\gamma : [0,1] \to \mathbb{R}^d$ is a tree-reduced continuous bounded-variation path, and the derivative of γ $f : (0,1) \to \mathbb{R}^d$ is defined almost everywhere, and $\|f(t)\| = 1$ for all $t \in (0,1)$ if defined. Assume γ is linear on $[s,t]$ for $0 \leq s < t \leq 1$, and let $\theta \in (s,t)$. If we choose $p = \lfloor \theta(n+2) \rfloor$, then*

$$\left\| I_{p,n}\left(f(\theta)\right) - \bar{S}_{n+1} \right\| \to 0 \quad \text{as} \quad n \to \infty,$$

and the rate of convergence of the upper bound obtained for $\left\| I_{p,n}\left(f(\theta)\right) - \bar{S}_{n+1} \right\|$ is slower than $O\left(1/\sqrt{n+1}\right)$.

Proof. Since γ is tree-reduced, by Boedihardjo and Geng [4], there exists $N \in \mathbb{N}$ such that for all $n \geq N$, $S^n_{0,1}(\gamma) \neq 0$. We now only consider the case when $n \geq N$.
By Chen's identity,

$$\begin{aligned}
\bar{S}_n &= n! \sum_{k_1+k_2+k_3=n} S^{k_1}_{0,s}(\gamma) \otimes S^{k_2}_{s,t}(\gamma) \otimes S^{k_3}_{t,1}(\gamma) \\
&= n! \sum_{k_1+k_2+k_3=n} S^{k_1}_{0,s}(\gamma) \otimes \frac{(t-s)^{k_2} f(\theta)^{\otimes k_2}}{k_2!} \otimes S^{k_3}_{t,1}(\gamma).
\end{aligned}$$

Define the sets

$$I := \left\{(k_1,k_2,k_3) : k_1+k_2+k_3 = n, k_i = 0, \cdots, n, \ i = 1,2,3\right\}$$

and

$$J := \left\{(k_1,k_2,k_3) : (k_1,k_2,k_3) \in I, k_1 \leq p-1, k_3 \leq n+1-p\right\}.$$

For $x \in \mathbb{R}^d$, let $x \uparrow S_{u,v}^k(\gamma)$ denote the resulting tensor of inserting x into any position of $S_{u,v}^k(\gamma)$ for any $0 \leq u < v \leq 1$. Then

$$\begin{aligned}
I_{p,n}\left(f(\theta)\right) = & \sum_{(k_1,k_2,k_3)\in J} n! S_{0,s}^{k_1}(\gamma) \otimes \frac{(t-s)^{k_2} f(\theta)^{\otimes k_2+1}}{k_2!} \otimes S_{t,1}^{k_3}(\gamma) \\
& + \sum_{(k_1,k_2,k_3)\in I, k_1\geq p} n! \left(f(\theta) \uparrow S_{0,s}^{k_1}(\gamma)\right) \otimes \frac{(t-s)^{k_2} f(\theta)^{\otimes k_2}}{k_2!} \otimes S_{t,1}^{k_3}(\gamma) \\
& + \sum_{(k_1,k_2,k_3)\in I, k_3\geq n-p+2} n! S_{s,t}^{k_1}(\gamma) \otimes \frac{(t-s)^{k_2} f(\theta)^{\otimes k_2}}{k_2!} \otimes \left(f(\theta) \uparrow S_{t,1}^{k_3}(\gamma)\right).
\end{aligned}$$

Note also

$$\begin{aligned}
\bar{S}_{n+1} = & \sum_{k_1+k_2+k_3=n+1} (n+1)! S_{0,s}^{k_1}(\gamma) \otimes \frac{(t-s)^{k_2} f(\theta)^{\otimes k_2}}{k_2!} \otimes S_{t,1}^{k_3}(\gamma) \\
= & \sum_{(k_1,k_2,k_3)\in J} (n+1)! S_{0,s}^{k_1}(\gamma) \frac{(t-s)^{k_2+1} f(\theta)^{\otimes k_2+1}}{(k_2+1)!} \otimes S_{t,1}^{k_3}(\gamma) \\
& + \sum_{\substack{k_1+k_2+k_3=n+1 \\ k_1\geq p}} (n+1)! S_{0,s}^{k_1}(\gamma) \otimes \frac{(t-s)^{k_2} f(\theta)^{\otimes k_2}}{k_2!} \otimes S_{t,1}^{k_3}(\gamma) \\
& + \sum_{\substack{k_1+k_2+k_3=n+1 \\ k_3\geq n-p+2}} (n+1)! S_{0,s}^{k_1}(\gamma) \otimes \frac{(t-s)^{k_2} f(\theta)^{\otimes k_2}}{k_2!} \otimes S_{t,1}^{k_3}(\gamma).
\end{aligned}$$

Then since for any $0 \leq u < v \leq 1$ and any $k \geq 1$, $\left\|S_{u,v}^k(\gamma)\right\| \leq (v-u)^k/k!$, we have

$$\begin{aligned}
& \left\|I_{p,n}\left(f(\theta)\right) - \bar{S}_{n+1}\right\| \\
\leq & \sum_{(k_1,k_2,k_3)\in J} \frac{(n+1)!}{k_1!(k_2+1)!k_3!} s^{k_1}(1-t)^{k_3}(t-s)^{k_2+1} \left|\frac{k_2+1}{n+1}\frac{1}{t-s} - 1\right| \\
& + \sum_{(k_1,k_2,k_3)\in I, k_1\geq p} \frac{n!}{k_1!k_2!k_3!} s^{k_1}(t-s)^{k_2}(1-t)^{k_3} \\
& + \sum_{(k_1,k_2,k_3)\in I, k_3\geq n-p+2} \frac{n!}{k_1!k_2!k_3!} s^{k_1}(t-s)^{k_2}(1-t)^{k_3} \\
& + \sum_{\substack{k_1+k_2+k_3=n+1 \\ k_1\geq p}} \frac{(n+1)!}{k_1!k_2!k_3!} s^{k_1}(t-s)^{k_2}(1-t)^{k_3} \\
& + \sum_{\substack{k_1+k_2+k_3=n+1 \\ k_3\geq n-p+2}} \frac{(n+1)!}{k_1!k_2!k_3!} s^{k_1}(t-s)^{k_2}(1-t)^{k_3} \\
\leq & \sum_{0\leq k\leq n+1} \frac{(n+1)!}{k!(n+1-k)!}(t-s)^k(1-t+s)^{n+1-k} \left|\frac{k}{n+1}\frac{1}{t-s} - 1\right|
\end{aligned}$$

$$
\begin{aligned}
&+\sum_{p\le k\le n}\frac{n!}{k!(n-k)!}s^k(1-s)^{n-k}\\
&+\sum_{n-p+2\le k\le n}\frac{n!}{k!(n-k)!}(1-t)^k t^{n-k}\\
&+\sum_{p\le k\le n+1}\frac{(n+1)!}{k!(n+1-k)!}s^k(1-s)^{n+1-k}\\
&+\sum_{n-p+2\le k\le n+1}\frac{(n+1)!}{k!(n+1-k)!}(1-t)^k t^{n+1-k}.
\end{aligned}
\tag{4.19}
$$

Note

$$
\begin{aligned}
&\sum_{0\le k\le n+1}\frac{(n+1)!}{k!(n+1-k)!}(t-s)^k(1-t+s)^{n+1-k}\left|\frac{k}{n+1}\frac{1}{t-s}-1\right|\\
=&\sum_{\substack{r=k/(n+1)\\ k=0,\cdots,n+1}}\frac{(n+1)!}{((n+1)r)!\,((n+1)(1-r))!}(t-s)^{(n+1)r}(1-t+s)^{(n+1)(1-r)}\left|\frac{1}{t-s}r-1\right|,
\end{aligned}
\tag{4.20}
$$

then if we assume X ~Binomial$(n+1,t-s)$ and define $Y:=X/(n+1)$, notice that (4.20) is the expectation of the function $|Y/(t-s)-1|$, and

$$
\mathbb{E}[Y]=t-s,\quad \mathrm{Var}[Y]=\frac{(t-s)(1-t+s)}{n+1}.
$$

By Hoeffding's inequality, for any $\lambda>0$,

$$
\begin{aligned}
&\mathbb{P}\left(|Y-(t-s)|\ge\lambda\sqrt{\frac{(t-s)(1-t+s)}{n+1}}\right)\\
=&\mathbb{P}\left(|X-(t-s)(n+1)|\ge\lambda\sqrt{(t-s)(1-t+s)(n+1)}\right)\\
\le&2\exp\left(-2(t-s)(1-t+s)\lambda^2\right),
\end{aligned}
$$

therefore by considering the cases when $|Y-(t-s)|<\lambda\sqrt{(t-s)(1-t+s)/(n+1)}$ and $|Y-(t-s)|\ge\lambda\sqrt{(t-s)(1-t+s)/(n+1)}$ respectively, we have

$$
\begin{aligned}
&\sum_{0\le k\le n+1}\frac{(n+1)!}{k!(n+1-k)!}(t-s)^k(1-t+s)^{n+1-k}\left|\frac{k}{n+1}\frac{1}{t-s}-1\right|\\
\le&\lambda\sqrt{\frac{1-t+s}{(t-s)(n+1)}}\\
&+\max\left(1,\left(\frac{1}{t-s}-1\right)\right)\mathbb{P}\left(|Y-(t-s)|\ge\lambda\sqrt{\frac{(t-s)(1-t+s)}{n+1}}\right)
\end{aligned}
$$

$$\leq \lambda\sqrt{\frac{1-t+s}{(t-s)(n+1)}} + 2\max\left(1, \left(\frac{1}{t-s}-1\right)\right)\exp\left(-2(t-s)(1-t+s)\lambda^2\right).$$

By similar arguments as in Example 4.4.1, there exists a strictly increasing sequence $(\lambda_n)_n$ such that for each $n \in \mathbb{N}$, $\lambda_n > 0$, and

$$\lambda_n\sqrt{\frac{1-t+s}{(t-s)(n+1)}} = 2\max\left(1, \left(\frac{1}{t-s}-1\right)\right)\exp\left(-2(t-s)(1-t+s)\lambda_n^2\right). \tag{4.21}$$

Suppose $\lambda_n \to \lambda^* < \infty$ as $n \to \infty$. Then taking limits on both sides of Equation (4.21) gives

$$0 = 2\max\left(1, \left(\frac{1}{t-s}-1\right)\right)\exp\left(-2(t-s)(1-t+s)\lambda^{*2}\right),$$

which does not hold if λ^* is finite. Hence we must have $\lambda^* = \infty$. Therefore

$$\begin{aligned}
&\sum_{0\leq k\leq n+1} \frac{(n+1)!}{k!(n+1-k)!}(t-s)^k(1-t+s)^{n+1-k}\left|\frac{k}{n+1}\frac{1}{t-s}-1\right| \\
\leq &2\lambda_n\sqrt{\frac{1-t+s}{(t-s)(n+1)}} \\
= &4\max\left(1, \left(\frac{1}{t-s}-1\right)\right)\exp\left(-2(t-s)(1-t+s)\lambda_n^2\right) \to 0
\end{aligned}$$

as $n \to \infty$. If we define $X_1 \sim$Binomial(n, s), $X_2 \sim$Binomial$(n, 1-t)$, $X_3 \sim$Binomial$(n+1, s)$, and $X_4 \sim$Binomial$(n+1, 1-t)$, by applying Hoeffding's inequality to the other binomial sums on the right-hand side of (4.19), there exists $M \in \mathbb{N}$ such that for all $n \geq M$,

$$\begin{aligned}
&\left\|I_{p,n}\left(f(\theta)\right) - \bar{S}_{n+1}\right\| \\
\leq &2\lambda_n\sqrt{\frac{1-t+s}{(t-s)(n+1)}} + \mathbb{P}\left(X_1 \geq p\right) \\
&+ \mathbb{P}\left(X_2 \geq n-p+2\right) + \mathbb{P}\left(X_3 \geq p\right) + \mathbb{P}\left(X_4 \geq n-p+2\right) \\
\leq &2\lambda_n\sqrt{\frac{1-t+s}{(t-s)(n+1)}} + \exp\left(-C_1 n + B_1\right) \\
&+ \exp\left(-C_2 n + B_2\right) + \exp\left(-C_3 n + B_3\right) + \exp\left(-C_4 n + B_4\right),
\end{aligned}$$

where we have used Hoeffding's inequality and $C_1 > 0, C_2 > 0, C_3 > 0, C_4 > 0$, B_1, B_2, B_3 and B_4 are some constants. Therefore

$$\left\|I_{p,n}\left(f(\theta)\right) - \bar{S}_{n+1}\right\| \to 0 \quad \text{as} \quad n \to \infty.$$

Moreover, the rate of convergence of the upper bound obtained for $\left\|I_{p,n}\left(f(\theta)\right)-\bar{S}_{n+1}\right\|$ is slower than $O\left(1/\sqrt{(t-s)(n+1)}\right)$. $\square$

We also have the following example as a numerical demonstration of our claim.

Example 4.4.2. *Assume $\gamma \in \mathbb{R}^2$ is a piecewise linear path which is an approximation to the quadratic path over the unit interval $[0,1]$ parametrised at unit speed, i.e. $\gamma^{(2)}(t) = (\gamma^{(1)}(t))^2$ for all $t \in [0,1]$. Fixing $\theta = 0.3$, if we compute the difference $\left\|I_{p,n}(f(\theta)) - \bar{S}_{n+1}\right\|$ under the ℓ^1 norm and ℓ^2 norm, we obtain Figure 4.2 and 4.3. From the figures we can see that under both the ℓ^1 norm and ℓ^2 norm, the difference between the n-th term in the normalised signature with $f(\theta)$ inserted at the p-th position and the $(n+1)$-th term in the normalised signature decreases as n increases, although not monotonically. The reason for the non-monotonicity is that given $p = \lfloor\theta(n+2)\rfloor$, $I_{p,n}(f(p/(n+2)))$ is a good approximation of $\bar{S}_{n+1}$, while $I_{p,n}(f(\theta))$ may not be as a good approximation as $I_{p,n}(f(p/(n+2)))$ if $\theta(n+2)$ is not an integer, hence we may observe small increases when $p/(n+2)$ is a bit far from θ.*
As a justification, we now choose $\theta = 0.5$ and $p = \lfloor 0.5(n+2)\rfloor$ for $n = 4, 6, 8, 10, 12$, and plot $\left\|I_{p,n}(f(\theta)) - \bar{S}_{n+1}\right\|_1$ in Figure 4.4. In this case since the signature level n used is even, $0.5(n+2)$ is an integer, and from the figure we can see that we get monotone convergence as n increases in this case.

4.5 A lower bound for the signature of a path

We have so far discussed finding an upper bound on $\left\|I_{p,n}(f(\theta)) - \bar{S}_{n+1}\right\|$ for a path which is differentiable at θ. In the light of Lemma 4.2.3, we know that for $x, y \in \mathbb{R}^d$,

$$\left\|\bar{S}_n\right\| \|x-y\| = \|I_{p,n}(x) - I_{p,n}(y)\| \leq \left\|I_{p,n}(x) - \bar{S}_{n+1}\right\| + \left\|I_{p,n}(y) - \bar{S}_{n+1}\right\|.$$

Given an upper bound on the right-hand side of the inequality, if we can obtain a lower bound on $\left\|\bar{S}_n\right\|$, we can get an upper bound on $\|x-y\|$. In fact, finding a lower bound for the signature is itself an interesting topic. We will see in the following example that the rate of decay of the signature depends on the path as well as the norm we choose.

Example 4.5.1. *This is an extension of Example 2.3.1. If we consider a monotone lattice path γ of consisting of two pieces, and each piece is of length $\frac{1}{2}$, the ℓ^1 norm of the signature at level n is*

$$\left\|n! S^n_{0,T}(\gamma)\right\|_1 = \sum_{k=0}^{n} \binom{n}{k} \left(\frac{1}{2}\right)^k \left(\frac{1}{2}\right)^{n-k} = 1,$$

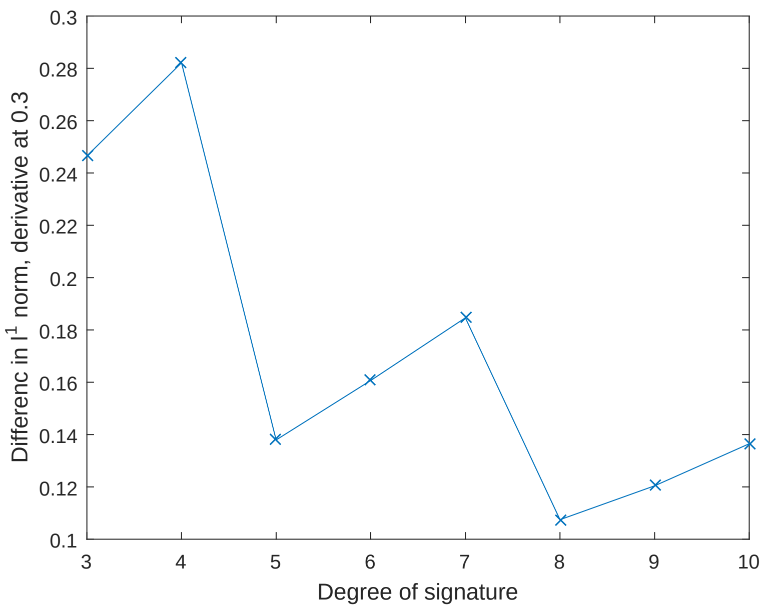


Figure 4.2: $\left\|I_{p,n}(f(\theta)) - \bar{S}_{n+1}\right\|_1$ for $p = \lfloor 0.3(n+2) \rfloor$, $n = 3, \cdots, 10$.

hence the signature is clearly bounded below. If we consider the norm of the signature under the Hilbert-Schmidt norm, then

$$\left\|n! S^n_{0,T}(\gamma)\right\|_{HS} = \sqrt{\sum_{k=0}^{n} \binom{n}{k}^2 \left(\frac{1}{2}\right)^{2k} \left(\frac{1}{2}\right)^{2(n-k)}}.$$

As we can see from Figure 4.5, the $\left\|S^n_{0,T}(\gamma)\right\|_{HS}$ decreases in such a way that there is no obvious constant non-zero lower bound for the signature.

Therefore it is important to take into account the effects of the norm when we look for a lower bound for the signature.
We first recall the norm in Definition 1.3.5 by Hambly and Lyons [19]: If V is a Banach space, A is a Banach algebra, and $F_1, \cdots, F_k \in \mathrm{Hom}(V, A)$, then the canonical linear extension $F_1 \otimes \cdots \otimes F_k$ from $V^{\otimes k}$ to A is defined as

$$(v_1, \cdots, v_k) \to F_1(v_1) \cdots F_k(v_k).$$

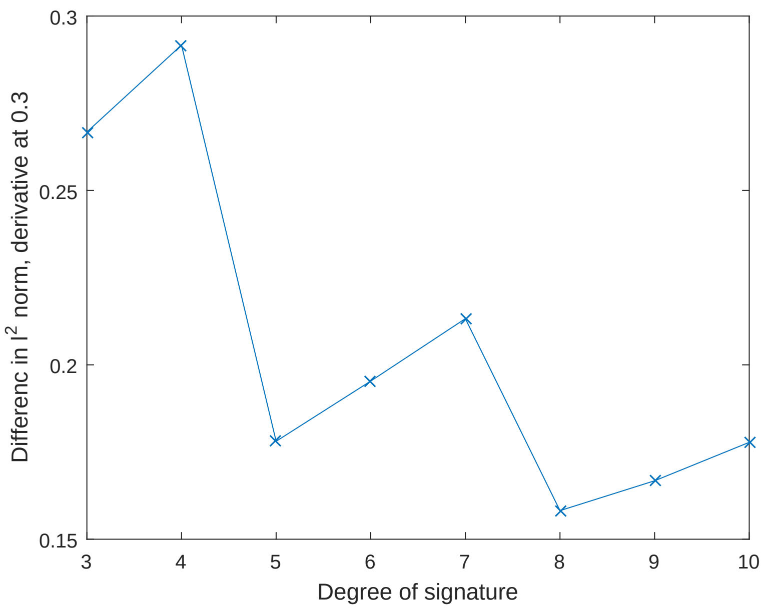


Figure 4.3: $\left\|I_{p,n}(f(\theta)) - \bar{S}_{n+1}\right\|_2$ for $p = \lfloor 0.3(n+2) \rfloor$, $n = 3, \cdots, 10$.

The norm

$$\|x\|_{\to A} := \sup_{F_i \in \mathrm{Hom}(V,A), \|F_i\|_{\mathrm{Hom}(V,A)}=1} \|F_1 \otimes \cdots \otimes F_k(x)\|_A .$$

As stated by Hambly and Lyons [19], the norm $\|\cdot\|_{\to A}$ is smaller than the projective tensor norm. We give a proof of this claim in the following lemma.

Lemma 4.5.1. *For all $x \in V^{\otimes k}$, $\|x\|_{\to A} \leq \|x\|_\pi$, where $\|\cdot\|_\pi$ is the projective tensor norm as defined in Definition 2.1.4.*

Proof. For all $x \in V^{\otimes k}$, if $x = \sum_{i \in I} v_{i_1} \otimes \cdots \otimes v_{i_k}$ is an representation of x for some indexing set I, then for any $F_i \in \mathrm{Hom}(V, A)$ such that $\|F_i\|_{\mathrm{Hom}(V,A)} = 1$, $i = 1, \cdots, k$, we have

$$\|(F_1 \otimes \cdots \otimes F_k)(x)\|_A = \left\| \sum_{i \in I} F_1(v_1) \cdots F_k(v_k) \right\|_A$$

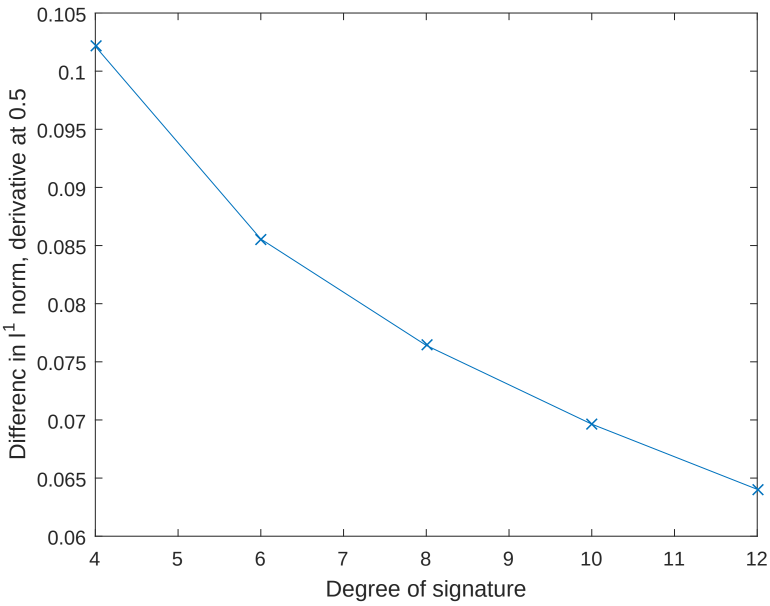


Figure 4.4: $\left\|I_{p,n}(f(\theta)) - \bar{S}_{n+1}\right\|_1$ for $p = \lfloor 0.5(n+2) \rfloor$, $n = 4, 6, 8, 10, 12$.

$$\leq \sum_{i \in I} \|F(v_{i_1})\|_A \cdots \| F(v_{i_k})\|_A$$

$$\leq \sum_{i \in I} \|v_{i_1}\| \cdots \|v_{i_k}\|,$$

for an abitrary representation of x. Then by the definition of the projective tensor norm, for any $F_i \in \mathrm{Hom}(V, A)$ such that $\|F_i\|_{\mathrm{Hom}(V,A)} = 1$, $i = 1, \cdots, k$,

$$\|(F_1 \otimes \cdots \otimes F_k)(x)\|_A \leq \|x\|_\pi,$$

hence

$$\|x\|_{\to A} \leq \|x\|_\pi.$$

$\square$

We also note the following useful lemma which gives a bound on the tail behaviour of the Poisson distribution by Canonne [7].

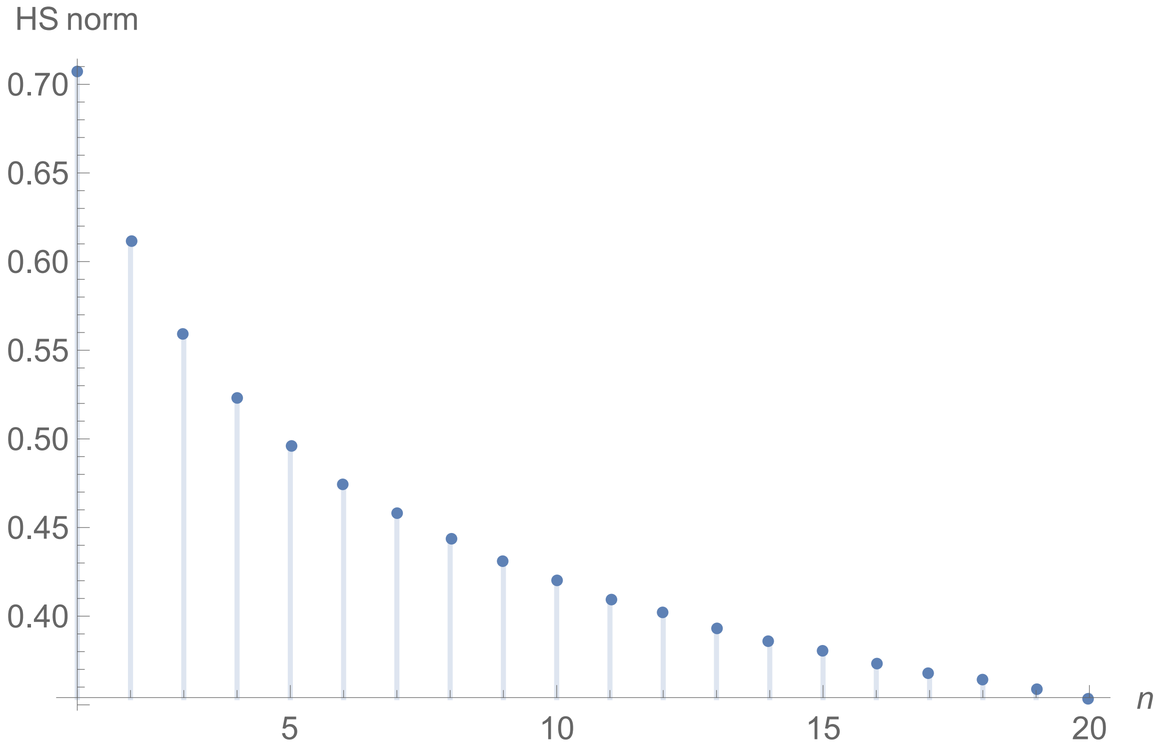


Figure 4.5: The Hilbert-Schmidt norm of the normalised signature of a monotone lattice path at level n

Lemma 4.5.2 (Canonne [7]). *Let $X \sim Poisson(\lambda)$ for some parameter $\lambda > 0$. Then for any $h > 0$, we have*

$$\mathbb{P}\left(|X-\lambda| \geq h\right) \leq 2\exp\left(-\frac{h^2}{2(\lambda+h)}\right).$$

We extend the argument by Hambly and Lyons (Theorem 13, [19]) and prove in the following theorem that a non-zero lower bound exists for more than one level of the signature of a piecewise linear path.

Theorem 4.5.1. *Let $\gamma : [0,1] \to \mathbb{R}^d$ be a non-degenerate piecewise linear path consisting of $M > 0$ linear pieces. Suppose $2\Omega > 0$ is the smallest angle between two adjacent edges. Equip $\mathbb{R}^d$ and $\mathbb{R}^{d+1}$ with the Euclidean norm. Then for any $c \in (0,1)$, there exists at least an increasing subsequence $(n_k)_{k\geq 1} \in \mathbb{N}$ such that*

$$\left\|\bar{S}_{n_k}\right\|_{\to Hom(\mathbb{R}^{d+1},\mathbb{R}^{d+1})} \geq c\exp\left(-(M-1)K(\Omega)\right) \quad \forall k \geq 1,$$

where $K(\Omega) := \log\left(\frac{2}{1-\cos|\Omega|}\right)$.

Proof. Without loss of generality, we can assume γ is of length 1. Suppose $D > 0$ is the length of the shortest edge of γ. For $\alpha > 0$, write the path $\alpha\gamma$ as γ_α. Then for all α such that $\alpha \geq \frac{K(\Omega)}{D}$, the shortest path of γ_α is at least of length $K(\Omega)$. Then by Lemma 1.3.3, the Cartan development Γ_α of γ_α, as defined by Equation (1.2), satisfies

$$d(o, \Gamma_\alpha o) \geq \alpha - (M-1)K(\Omega).$$

Also by Proposition 1.3.2, we know that

$$\|\Gamma_\alpha\|_{\mathrm{Hom}(\mathbb{R}^{d+1},\mathbb{R}^{d+1})} \geq \exp(d(o, \Gamma_\alpha o)).$$

Then if we recall the definition of the map F as in (1.1), for all α such that $\alpha > \frac{K(\Omega)}{D}$, we have

$$\begin{aligned}
&\exp\left(\alpha - (M-1)K(\Omega)\right) && (4.22)\\
\leq & \|\Gamma_\alpha\|_{\mathrm{Hom}(\mathbb{R}^{d+1},\mathbb{R}^{d+1})}\\
\leq & \sum_{n=0}^{\infty} \alpha^n \left\| \int_{0<u_1<\cdots<u_n<1} F(\mathrm{d}\gamma_{u_1}) \cdots F(\mathrm{d}\gamma_{u_n}) \right\|_{\mathrm{Hom}(\mathbb{R}^{d+1},\mathbb{R}^{d+1})}\\
= & \sum_{n=0}^{\infty} \alpha^n \left\| (F \otimes \cdots \otimes F) \left(\int_{0<u_1<\cdots<u_n<1} \mathrm{d}\gamma_{u_1} \otimes \cdots \otimes \mathrm{d}\gamma_{u_n} \right) \right\|_{\mathrm{Hom}(\mathbb{R}^{d+1},\mathbb{R}^{d+1})}\\
\leq & \sum_{n=0}^{\infty} \alpha^n \left\| \int_{0<u_1<\cdots<u_n<1} \mathrm{d}\gamma_{u_1} \otimes \cdots \otimes \mathrm{d}\gamma_{u_n} \right\|_{\to\mathrm{Hom}(\mathbb{R}^{d+1},\mathbb{R}^{d+1})}\\
\leq & \sum_{n=0}^{\infty} \frac{\alpha^n}{n!} \left\| \bar{S}_n \right\|_{\to\mathrm{Hom}(\mathbb{R}^{d+1},\mathbb{R}^{d+1})}, && (4.23)
\end{aligned}$$

where the third inequality follows from the definition of the norm $\|\cdot\|_{\to\mathrm{Hom}(\mathbb{R}^{d+1},\mathbb{R}^{d+1})}$. Multiplying both sides of (4.22) by $\exp(-\alpha)$ gives

$$\exp\left(-(M-1)K(\Omega)\right) \leq \exp(-\alpha) \sum_{n=0}^{\infty} \frac{\alpha^n}{n!} \left\| \bar{S}_n \right\|_{\to\mathrm{Hom}(\mathbb{R}^{d+1},\mathbb{R}^{d+1})}. \tag{4.24}$$

Note the right hand side of (4.24) is the expectation of the function $\left\| \bar{S}_n \right\|_{\to\mathrm{Hom}(\mathbb{R}^{d+1},\mathbb{R}^{d+1})}$ under the Poisson distribution with parameter α. Note the distribution has mean α, variance α. We have the following claim:
For all $c \in (0,1)$,

$$\begin{aligned}
&\mathbb{P}\left(n \,\text{such that}\, \left\| \bar{S}_n \right\|_{\to\mathrm{Hom}(\mathbb{R}^{d+1},\mathbb{R}^{d+1})} \geq c \exp\left(-(M-1)K(\Omega)\right) \right)\\
\geq & (1-c)\exp\left(-(M-1)K(\Omega)\right).
\end{aligned}$$

We prove the above claim by contradiction. Suppose that

$$\mathbb{P}\left(n \,\text{such that}\, \left\| \bar{S}_n \right\|_{\to\mathrm{Hom}(\mathbb{R}^{d+1},\mathbb{R}^{d+1})} \geq c \exp\left(-(M-1)K(\Omega)\right) \right)$$

$$<(1-c)\exp\left(-(M-1)K(\Omega)\right).$$

We know from Remark 1.3.1 that $\left\|\bar{S}_n\right\|_{\to\mathrm{Hom}(\mathbb{R}^{d+1},\mathbb{R}^{d+1})} \leq 1$. Then if we think about how large the expectation can be, we have

$$\begin{aligned}
&\exp(-\alpha)\sum_{n=0}^{\infty}\frac{\alpha^n}{n!}\left\|\bar{S}_n\right\|_{\to\mathrm{Hom}(\mathbb{R}^{d+1},\mathbb{R}^{d+1})} \\
<&(1-c)\exp\left(-(M-1)K(\Omega)\right)+c\exp(-(M-1)K(\Omega)) \\
=&\exp(-(M-1)K(\Omega))
\end{aligned}$$

which contradicts (4.24). So we must have

$$\begin{aligned}
&\mathbb{P}\left(n\,\text{such that}\,\left\|\bar{S}_n\right\|_{\to\mathrm{Hom}(\mathbb{R}^{d+1},\mathbb{R}^{d+1})} \geq c\exp\left(-(M-1)K(\Omega)\right)\right) \\
\geq&(1-c)\exp\left(-(M-1)K(\Omega)\right).
\end{aligned}$$

Then by Lemma 4.5.2, we have an estimate for $X\sim$Poisson(α) such that for all $h>0$,

$$\mathbb{P}\left(|X-\alpha|\geq h\right)\leq 2\exp\left(-\frac{h^2}{2(\alpha+h)}\right).$$

In particular,

$$\mathbb{P}\left(|X-\alpha|\geq \alpha^{3/4}\right)\leq 2\exp\left(-\frac{\alpha^{3/2}}{2(\alpha+\alpha^{3/4})}\right). \tag{4.25}$$

Note the right-hand side of (4.25) is a decreasing function in α, hence there exists α^* such that for all $\alpha>\alpha^*$,

$$\begin{aligned}
\mathbb{P}(|X-\alpha|\geq\alpha^{3/4}) &\leq 2\exp\left(-\frac{\alpha^{3/2}}{2(\alpha+\alpha^{3/4})}\right) \\
&< (1-c)\exp\left(-(M-1)K(\Omega)\right) \\
&\leq \mathbb{P}\left(n\,\text{such that}\,\left\|\bar{S}_n\right\|_{\to\mathrm{Hom}(\mathbb{R}^{d+1},\mathbb{R}^{d+1})} \geq c\exp(-(M-1)K(\Omega))\right).
\end{aligned}$$

Then there must be some n near the mean α such that

$$\left\|\bar{S}_n\right\|_{\to\mathrm{Hom}(\mathbb{R}^{d+1},\mathbb{R}^{d+1})} \geq c\exp(-(M-1)K(\Omega)),$$

i.e. for $\alpha>\alpha^*$,

$$\begin{aligned}
&\mathbb{P}\left(X=n\,\text{where}\,|n-\alpha|<\alpha^{3/4}\,\text{and}\,\|\bar{S}_n\|_{\to\mathrm{Hom}(\mathbb{R}^{d+1},\mathbb{R}^{d+1})}\geq c\exp(-(M-1)K(\Omega))\right) \\
\geq&(1-c)\exp\left(-(M-1)K(\Omega)\right)-2\exp\left(-\frac{\alpha^{3/2}}{2(\alpha+\alpha^{3/4})}\right) \\
>&0.
\end{aligned}$$

Hence for large enough α, there exists at least one $n \in (\alpha - \alpha^{3/4}, \alpha + \alpha^{3/4})$ such that $\left\|\bar{S}_n\right\|_{\to \mathrm{Hom}(\mathbb{R}^{d+1}, \mathbb{R}^{d+1})} \geq c \exp(-(M-1)K(\Omega))$. Note that α grows faster than $\alpha^{3/4}$, so as α increases, the interval $(\alpha - \alpha^{3/4}, \alpha + \alpha^{3/4})$ moves rightwards. Hence there exists a strictly increasing subsequence $(n_k)_{k \geq 1} \in \mathbb{N}$ such that

$$\left\|\bar{S}_{n_k}\right\|_{\to \mathrm{Hom}(\mathbb{R}^{d+1}, \mathbb{R}^{d+1})} \geq c \exp(-(M-1)K(\Omega)) \quad \forall\, k \geq 1.$$

□

Corollary 4.5.1. *Let $\gamma : [0,1] \to \mathbb{R}^d$ be a non-degenerate piecewise linear path consisting of $M > 0$ linear pieces. Suppose $2\Omega > 0$ is the smallest angle between two adjacent edges. Equip $\mathbb{R}^d$ and $\mathbb{R}^{d+1}$ with the Euclidean norm. Then for any $c \in (0,1)$, there exists at least a subsequence $(n_k)_{k \geq 1} \in \mathbb{N}$ such that*

$$\left\|\bar{S}_{n_k}\right\|_{\pi} \geq c \exp(-(M-1)K(\Omega)) \quad \forall k \geq 1,$$

where $K(\Omega) := \log\left(\frac{2}{1-\cos|\Omega|}\right)$ and $\|\cdot\|_\pi$ is the projective tensor norm induced from the Euclidean space $\mathbb{R}^d$.

Proof. Without loss of generality, we assume γ is of length 1. There are two ways to justify this result.
By Lemma 4.5.1, we know that $\|\cdot\|_\pi$ is a bigger norm than $\|\cdot\|_{\to \mathrm{Hom}(\mathbb{R}^{d+1}, \mathbb{R}^{d+1})}$, hence by Theorem 4.5.1, for any $c \in (0,1)$, there exists a subsequence $(n_k)_{k \geq 1} \in \mathbb{N}$ such that

$$c \exp(-(M-1)K(\Omega)) \leq \left\|\bar{S}_{n_k}\right\|_{\to \mathrm{Hom}(\mathbb{R}^{d+1}, \mathbb{R}^{d+1})} \leq \left\|\bar{S}_{n_k}\right\|_{\pi}.$$

An alternative proof directly applies the argument in the proof of Theorem 4.5.1 to the norm $\|\cdot\|_\pi$. Note that for any $x \in \mathbb{R}^{d^{\otimes k}}$, and $F : \mathbb{R}^d \to \mathrm{Hom}(\mathbb{R}^{d+1}, \mathbb{R}^{d+1})$ as defined in (1.1), if $x = \sum_{i \in I} v_{i_1} \otimes \cdots \otimes v_{i_k}$ for an indexing set I, we have

$$\begin{aligned}
&\|(F \otimes \cdots \otimes F)(x)\|_{\mathrm{Hom}(\mathbb{R}^{d+1}, \mathbb{R}^{d+1})} \\
&= \left\|\sum_{i \in I} F(v_{i_1}) \cdots F(v_{i_k})\right\|_{\mathrm{Hom}(\mathbb{R}^{d+1}, \mathbb{R}^{d+1})} \\
&\leq \sum_{i \in I} \|F(v_{i_1})\|_{\mathrm{Hom}(\mathbb{R}^{d+1}, \mathbb{R}^{d+1})} \cdots \|F(v_{i_k})\|_{\mathrm{Hom}(\mathbb{R}^{d+1}, \mathbb{R}^{d+1})} \\
&\leq \|F\|^k_{\mathrm{Hom}(\mathbb{R}^d, \mathrm{Hom}(\mathbb{R}^{d+1}, \mathbb{R}^{d+1}))} \sum_{i \in I} \|v_{i_1}\| \cdots \|v_{i_k}\|,
\end{aligned}$$

for an abritrary representation of x. Hence

$$\|(F\otimes\cdots\otimes F)(x)\|_{\mathrm{Hom}(\mathbb{R}^{d+1},\mathbb{R}^{d+1})}\leq\|F\|^k_{\mathrm{Hom}(\mathbb{R}^d,\mathrm{Hom}(\mathbb{R}^{d+1},\mathbb{R}^{d+1}))}\|x\|_\pi\leq\|x\|_\pi.$$

Therefore when we equip $(\mathbb{R}^d)^{\otimes k}$ with $\|\cdot\|_\pi$, we have

$$\|(F\otimes\cdots\otimes F)\|_{\mathrm{Hom}\left(\left(\mathbb{R}^d\right)^{\otimes k},\mathrm{Hom}(\mathbb{R}^{d+1},\mathbb{R}^{d+1})\right)}\leq 1.$$

As in the proof of Theorem 4.5.1, and if we use $\|\cdot\|_\pi$ instead of $\|\cdot\|_{\to\mathrm{Hom}(\mathbb{R}^{d+1},\mathbb{R}^{d+1})}$, (4.22) becomes

$$\begin{aligned}
&\exp(\alpha-(M-1)K(\Omega))\\
\leq&\|\Gamma_\alpha\|_{\mathrm{Hom}(\mathbb{R}^{d+1},\mathbb{R}^{d+1})}\\
\leq&\sum_{n=0}^{\infty}\alpha^n\left\|\int_{0<u_1<\cdots<u_k<1}F(\mathrm{d}\gamma_{u_1})\cdots F(\mathrm{d}\gamma_{u_n})\right\|_{\mathrm{Hom}(\mathbb{R}^{d+1},\mathbb{R}^{d+1})}\\
=&\sum_{n=0}^{\infty}\alpha^n\left\|(F\otimes\cdots\otimes F)\int_{0<u_1<\cdots<u_n<1}\mathrm{d}\gamma_{u_1}\otimes\cdots\otimes\mathrm{d}\gamma_{u_n}\right\|_{\mathrm{Hom}(\mathbb{R}^{d+1},\mathbb{R}^{d+1})}\\
\leq&\sum_{n=0}^{\infty}\alpha^n\left\|\int_{0<u_1<\cdots<u_n<1}\mathrm{d}\gamma_{u_1}\otimes\cdots\otimes\mathrm{d}\gamma_{u_n}\right\|_\pi\\
\leq&\sum_{n=0}^{\infty}\frac{\alpha^n}{n!}\left\|\bar{S}\right\|_\pi
\end{aligned}$$

and then the rest of the proof of Theorem 4.5.1 applies. □

Remark 4.5.1. *We have seen from Conjecture 2.3.1 that we expect the n-th root of the n-th term in the signature of a path of finite length multiplied by $n!$ to converge to the length of the path under a reasonable tensor algebra norm. Hambly and Lyons [19] showed that (Theorem 4.6.2) a stronger decay result holds in special cases: If γ is a path of finite length $L>0$ with the modulus of continuity of its derivative $\delta(\epsilon)=o(\epsilon^{3/4})$, then*

$$L^{-k}k!\left\|\int_{0<u_1<\cdots<u_k<1}\mathrm{d}\gamma_{u_1}\otimes\cdots\otimes\mathrm{d}\gamma_{u_k}\right\|_{\to Hom(\mathbb{R}^{d+1},\mathbb{R}^{d+1})}\to 1$$

as $n\to\infty$. However we have seen from Example 4.5.1 such a strong result does not hold for piecewise linear paths at least under the Hilbert-Schmidt norm. The significance of Theorem 4.5.1 and Corollary 4.5.1 is that we have a stronger result for a piecewise linear path than we conjectured in Conjecture 2.3.1.

4.6 Inverting the signature of a path

Assume $\gamma : [0,1] \to \mathbb{R}^d$ is a continuous bounded-variation path with derivative $f : (0,1) \to \mathbb{R}^d$ such that $\|f(t)\|_2 = 1$ for all $t \in (0,1)$ almost everywhere. Assume γ is linear on $[s,t] \subset [0,1]$, and $\theta \in (s,t)$. For $n \geq 1$, choose $p \in \{1, \cdots, n+1\}$ such that $p = \lfloor \theta(n+2) \rfloor$. In this section we use the result of Theorem 4.4.2, i.e. there exists $\epsilon^{\gamma}_{\theta,n}$ such that $\|I_{p,n}(f(\theta)) - \bar{S}_{n+1}\|_\pi \leq \epsilon^{\gamma}_{\theta,n}$ and $\epsilon^{\gamma}_{\theta,n} \to 0$ as $n \to \infty$.
Define the set

$$A^{\gamma}_{\theta,n} := \left\{ x \in \mathbb{R}^d : \|x\|_2 = 1, \left\| I_{p,n}(x) - \bar{S}_{n+1} \right\|_\pi \leq \epsilon^{\gamma}_{\theta,n} \right\}.$$

Note $f(\theta) \in A^{\gamma}_{\theta,n}$. We adopt these notations in this section. We first note the following lemma.

Lemma 4.6.1. *The projective tensor norm $\|\cdot\|_\pi$ satisfies Definition 4.2.1.*

Proof. We know from Lemma 2.1.1 that the projective tensor norm is a reasonable tensor algebra norm, it then follows directly from Proposition 2.1.1 that the projective tensor norm satisfies the properties stated in Definition 4.2.1. □

We now give a strategy to invert the signature of a non-degenerate piecewise linear path.

Theorem 4.6.1. *Assume $\gamma : [0,1] \to \mathbb{R}^d$ is a non-degenerate piecewise linear path with derivative $f : (0,1) \to \mathbb{R}^d$ such that $\|f(t)\|_2 = 1$ for all $t \in (0,1)$ if defined. Assume γ is differentiable at $\theta \in (0,1)$. For $n \geq 1$, choose $p \in \{1, \cdots, n+1\}$ such that $p = \lfloor \theta(n+2) \rfloor$. Then there exists a subsequence $(n_k)_{k \geq 1} \in \mathbb{N}$ such that for all $k \geq 1$, for all $x_{\theta,n_k}, y_{\theta,n_k} \in A^{\gamma}_{\theta,n_k}$,*

$$\left\| x_{\theta,n_k} - y_{\theta,n_k} \right\|_2 \to 0 \quad as \quad k \to \infty.$$

Proof. Note for all $n \geq 1$, $x_{\theta,n}, y_{\theta,n} \in A^{\gamma}_{\theta,n}$,

$$\begin{aligned} &\left\| I_{p,n}(x_{\theta,n}) - I_{p,n}(y_{\theta,n}) \right\|_\pi \\ \leq &\left\| I_{p,n}(x_{\theta,n}) - \bar{S}_{n+1} \right\|_\pi + \left\| I_{p,n}(y_{\theta,n}) - \bar{S}_{n+1} \right\|_\pi \\ \leq &2\epsilon^{\gamma}_{\theta,n}. \end{aligned}$$

By Corollary 4.5.1, there exists a subsequence $(n_k)_{k \geq 1} \in \mathbb{N}$ such that for all $k \geq 1$, $\|\bar{S}_{n_k}\|_\pi \geq \frac{1}{2} \exp(-(M-1)K(\Omega))$, where $K(\Omega) = \log(\frac{2}{1-\cos|\Omega|})$, 2Ω is the smallest angle between two adjacent edges, and $M > 0$ is the number of linear pieces of γ. Then by Lemma 4.2.3,

$$\begin{aligned}
&\|x_{\theta,n_k} - y_{\theta,n_k}\|_2 \\
=&\frac{\|I_{p,n_k}(x_{\theta,n_k}) - I_{p,n_k}(y_{\theta,n_k})\|_\pi}{\left\|\bar{S}_{n_k}\right\|_\pi} \\
\leq&\frac{4\epsilon^\gamma_{\theta,n_k}}{\exp(-(M-1)K(\Omega))}.
\end{aligned}$$

Since $\epsilon_{\theta,n_k} \to 0$ as $k \to \infty$, we have $\|x_{\theta,n_k} - y_{\theta,n_k}\|_2 \to 0$ as $k \to \infty$. □

We are then able to derive a corollary which is more useful for computation.

Corollary 4.6.1. *Assume $\gamma : [0,1] \to \mathbb{R}^d$ is a non-degenerate piecewise linear path with derivative $f : (0,1) \to \mathbb{R}^d$ such that $\|f(t)\|_2 = 1$ for all $t \in (0,1)$ if defined. Assume γ is differentiable at $\theta \in (0,1)$. For $n \geq 1$, choose $p \in \{1, \cdots, n+1\}$ such that $p = \lfloor \theta(n+2) \rfloor$. Define*

$$x^*_{\theta,n} := argmin_{x \in \mathbb{R}^d, \|x\|_2 = 1} \left\|I_{p,n}(x) - \bar{S}_{n+1}\right\|_\pi . \tag{4.26}$$

Then there exists at least a subsequence $(n_k)_{k\geq 1} \in \mathbb{N}$ such that $x^*_{n_k}$ converges to $f(\theta)$ as k increases.

Proof. By Theorem 4.6.1, we know that there exists a subsequence $(n_k)_{k\geq 1} \in \mathbb{N}$ such that for all $k \geq 1$, for all $x_{\theta,n_k}, y_{\theta,n_k} \in A^\gamma_{\theta,n_k}$, $\|x_{\theta,n_k} - y_{\theta,n_k}\|_2 \to 0$ as $k \to \infty$. We know that $f(\theta) \in A^\gamma_{\theta,n_k}$, and $\left\|I_{p,n_k}(x^*_{\theta,n_k}) - \bar{S}_{n_k+1}\right\|_\pi \leq \epsilon^\gamma_{\theta,n_k}$ due to the fact that x^*_{θ,n_k} gives the shortest distance between $I_{p,n_k}(x)$ and $\bar{S}_{n_k+1}$ among all $x \in \mathbb{R}^d$ such that $\|x\|_2 = 1$. Therefore $x^*_{\theta,n_k} \in A^\gamma_{\theta,n_k}$, hence $\left\|x^*_{\theta,n_k} - f(\theta)\right\|_2 \to 0$ as $k \to \infty$. □

We can also develop such an algorithm for another set of paths. First we recall the following theorem by Hambly and Lyons [19].

Theorem 4.6.2 (Hambly and Lyons, Theorem 9 [19])**.** *Let J be a closed and bounded interval. Let $\gamma : J \to \mathbb{R}^d$ be a continuous path of finite length $\ell > 0$. Recall that the modulus of continuity of the derivative is defined as $\delta(h) := \sup_{|u-v|\leq h} \|\dot{\gamma}(u) - \dot{\gamma}(v)\|_2$. If $\delta(h) = o\left(h^{3/4}\right)$, then*

$$\ell^k k! \left\| \int_{0<u_1<\cdots<u_k<1} \mathrm{d}\gamma_{u_1} \otimes \cdots \otimes \mathrm{d}\gamma_{u_k} \right\|_{\to Hom\left(\mathbb{R}^{d+1}, \mathbb{R}^{d+1}\right)} \to 1$$

as $k \to \infty$.

Theorem 4.6.3. *Let $\gamma : [0,1] \to \mathbb{R}^d$ be a continuous path with derivative $f : (0,1) \to \mathbb{R}^d$ such that $\|f(t)\|_2 = 1$ for all $t \in (0,1)$. Suppose further that the modulus of continuity of f is $\delta(h) = o\left(h^{3/4}\right)$. Assume γ is linear over the interval $[s,t] \subset [0,1]$ and $\theta \in (s,t)$. Then for $n \geq 1$, choose $p \in \{1, ..., n+1\}$ such that $p = \lfloor \theta(n+2) \rfloor$. Define*

$$x^*_{\theta,n} := argmin_{x \in \mathbb{R}^d, \|x\|_2 = 1} \left\| I_{p,n}(x) - \bar{S}_{n+1} \right\|_\pi .$$

Then $x^*_{\theta,n}$ converges to $f(\theta)$ as n increases.

Proof. By Theorem 4.6.2, for any $c \in (0,1)$, there exists $N \in \mathbb{N}$ such that for all $n \geq N$,

$$\left\| \bar{S}_n \right\|_{\to \mathrm{Hom}\left(\mathbb{R}^{d+1}, \mathbb{R}^{d+1}\right)} \geq 1 - c.$$

By Lemma 4.5.1, the projective norm is bigger than the norm $\|\cdot\|_{\to \mathrm{Hom}\left(\mathbb{R}^{d+1}, \mathbb{R}^{d+1}\right)}$, hence for all $n \geq N$,

$$\left\| \bar{S}_n \right\|_\pi \geq 1 - c.$$

Then for all $n \geq N$, for all $x_{\theta,n}, y_{\theta,n} \in A^\gamma_{\theta,n}$, we have

$$\begin{aligned} & \|x_{\theta,n} - y_{\theta,n}\|_2 \\ = & \frac{\|I_{\theta,n}(x_{\theta,n}) - I_{\theta,n}(y_{\theta,n})\|_\pi}{\left\| \bar{S}_n \right\|_\pi} \\ \leq & \frac{2\epsilon^\gamma_{\theta,n}}{1 - c}. \end{aligned}$$

Since $\epsilon^\gamma_{\theta,n} \to 0$ as $n \to \infty$, we have $\|x_{\theta,n} - y_{\theta,n}\|_2 \to 0$ as $n \to \infty$. Since $x^*_{\theta,n}, f(\theta) \in A^\gamma_{\theta,n}$, we have

$$\left\| x^*_{\theta,n} - f(\theta) \right\|_2 \to 0 \quad \text{as} \quad n \to \infty.$$

□

Remark 4.6.1. *Note that if we take $p = \lfloor \theta(n+2) \rfloor$, we may get $p = 0$ if n is small. But we can always take higher orders of the signature, and this will not affect our result.*

Note that so far in this chapter we have assumed that the underlying path is parametrised at unit speed. However, in practice when we only have the information from the signature, it is impossible to know whether the path is parametrised at unit speed. We prove in the following lemma that our algorithm still works with a slight alteration.

Lemma 4.6.2. *For a non-degenerate piecewise linear path $\gamma : [a, b] \to \mathbb{R}^d$ of length $L > 0$ and differentiable at $\theta \in (a, b)$, we can slightly change (4.26) and obtain an approximation to the derivative of γ when it is parametrised at unit speed when we choose the position of insertion p appropriately, even if the original speed of parametrisation is unknown. Moreover the same changes apply to the result of Theorem 4.6.3.*

Proof. By Lemma 1.1.3 we know that γ can be re-parametrised at unit speed. Let the function $\phi : [a, b] \to [u, v]$ be such that the path $\tilde{\gamma} := \gamma \circ \phi$ is parametrised at unit speed.

We first try to determine what value p should be, i.e. the position at which the element shall be inserted into the n-th level of the normalised signature of $\tilde{\gamma}$. As in the proof of Proposition 4.2.1, for any norm which satisfies properties stated in Definition 4.2.1, if we insert $x \in \mathbb{R}^d$ into the n-th level of the signature of $\tilde{\gamma}$, then

$$
\begin{aligned}
&L^{-(n+1)}\Big\| \int_{u<t_1<\cdots<t_{n+1}<v} (n+1)!\dot{\tilde{\gamma}}_{t_1} \otimes \cdots \otimes \dot{\tilde{\gamma}}_{t_{p-1}} \otimes x \otimes \dot{\tilde{\gamma}}_{t_{p+1}} \otimes \cdots \otimes \dot{\tilde{\gamma}}_{t_{n+1}} \mathrm{d}t_1 \cdots \mathrm{d}t_{n+1} \\
&- (n+1)! \int_{u<t_1<\cdots<t_{n+1}<v} \dot{\tilde{\gamma}}_{t_1} \otimes \cdots \otimes \dot{\tilde{\gamma}}_{t_{n+1}} \mathrm{d}t_1 \cdots \mathrm{d}t_{n+1} \Big\| \\
=&L^{-(n+1)}(n+1)! \Big\| \int_{u<t_1<\cdots<t_{n+1}<v} \dot{\tilde{\gamma}}_{t_1} \otimes \cdots \otimes \dot{\tilde{\gamma}}_{t_{p-1}} \\
&\otimes (x - \dot{\tilde{\gamma}}_{t_p}) \otimes \dot{\tilde{\gamma}}_{t_{p+1}} \otimes \cdots \otimes \dot{\tilde{\gamma}}_{t_{n+1}} \mathrm{d}t_1 \cdots \mathrm{d}t_{n+1} \Big\| \\
\leq& \int_{u<t<v} \|x - \dot{\tilde{\gamma}}_t\| L^{-(n+1)}(n+1)! \frac{(t-u)^{p-1}}{(p-1)!} \frac{(v-t)^{n+1-p}}{(n+1-p)!} \mathrm{d}t,
\end{aligned}
$$

which gives rise to the expectation of a function about a *non-standard beta variable* $U \sim \mathrm{Beta}(p, n-p+2)$ over the interval (u, v). We can change the variable in the integral to obtain a standard beta variable $U' := (U-u)/(v-u)$, which is over the interval $(0, 1)$. Therefore we can see that the expectation of U is $p(v-u)/(n+2)+u$. With the same argument we had in the proof of Proposition 4.2.1, we know that we shall choose

$$
p = \left\lfloor \frac{\phi(\theta) - u}{v - u}(n+2) \right\rfloor
$$

in order to approximate the derivative of $\tilde{\gamma}$ at $\phi(\theta)$.

With a slight extension of the analysis in this chapter, we see that for $n \geq 1$, the solution to

$$
\min_{\|x\|_2=1} \left\| L^{-n} I^{\tilde{\gamma}}_{p,n}(x) - L^{-(n+1)}(n+1)! S^{n+1}_{u,v}(\tilde{\gamma}) \right\|_\pi \tag{4.27}
$$

gives an approximation to the derivative of $\tilde{\gamma}$ at $\phi(\theta)$, where

$$I^{\tilde{\gamma}}_{p,n}(x) := n! \int_{u<t_1<\cdots<t_n<v} \mathrm{d}\tilde{\gamma}_{t_1} \otimes \cdots \otimes \mathrm{d}\tilde{\gamma}_{t_{p-1}} \otimes x \otimes \mathrm{d}\tilde{\gamma}_{t_p} \otimes \cdots \otimes \mathrm{d}\tilde{\gamma}_{t_n}.$$

Note

$$\begin{aligned} S^n_{u,v}(\tilde{\gamma}) &= \int_{u<t_1<\cdots<t_n<v} \mathrm{d}\tilde{\gamma}_{t_1} \otimes \cdots \otimes \mathrm{d}\tilde{\gamma}_{t_n} \\ &= \int_{a<t_1<\cdots<t_n<b} (\phi'(t_1)\mathrm{d}\gamma_{\phi(t_1)}) \otimes \cdots \otimes (\phi'(t_n)\mathrm{d}\gamma_{\phi(t_n)}) \\ &= \int_{a<t_1<\cdots<t_n<b} \mathrm{d}\gamma_{t_1} \otimes \cdots \otimes \mathrm{d}\gamma_{t_n} \\ &= S^n_{a,b}(\gamma). \end{aligned}$$

Hence (4.27) can be written as a problem about γ:

$$\min_{\|x\|_2=1} \left\| L^{-n} I^{\gamma}_{p,n}(x) - L^{-(n+1)}(n+1)! S^{n+1}_{a,b}(\gamma) \right\|_\pi ,$$

which is equivalent to solving the following optimisation problem

$$\min_{\|x\|_2=1} \left\| L I^{\gamma}_{p,n}(x) - (n+1)! S^{n+1}_{a,b}(\gamma) \right\|_\pi . \tag{4.28}$$

Hence if we solve problem (4.28), we will obtain an approximation to the derivative of γ at θ when it is parametrised at unit speed. Therefore we can still recover the path, but maybe at a different speed of parametrisation from the underlying. The same argument clearly applies to the result of Theorem 4.6.3. □

Remark 4.6.2. *The significance of Lemma 4.6.2 is that it provides us with a generalised version of the insertion algorithm we have developed in this chapter, and we will then be able to reconstruct a path even if it is parametrised at an unknown speed. In fact it shows that the insertion algorithm developed in this chapter for inverting the signature of a path requires the knowledge of the length of the path. A particular example can be found in the next chapter in Example 5.2.3.*

Chapter 5

Computational reconstruction of a path from its signature

We have seen in Chapter 4 that a path can be reconstructed by solving an optimisation problem after inserting an element into a level of the signature of the path. In this chapter we demonstrate computationally how to use this method to recover a path, and we also include comparisons between this insertion method and the symmetrisation method discussed in Chapter 3.

5.1 Setting of the optimisation problem

Suppose $\gamma : [0, 1] \to \mathbb{R}^d$ is a tree-reduced continuous bounded-variation path with derivative $f : (0, 1) \to \mathbb{R}^d$ such that $\|f(t)\|_2 = 1$ for all $t \in (0, 1)$ almost everywhere, and γ is differentiable at $\theta \in (0, 1)$. We have seen from Chapter 4 that given certain assumptions are satisfied, the key to reconstruct the path from the signature is to solve the optimisation problem

$$\min_{\|x\|_2=1} \left\| I_{p,n}(x) - \bar{S}_{n+1} \right\|_\pi , \tag{5.1}$$

where $p = \lfloor \theta(n + 2) \rfloor$. If we want to computationally reconstruct the path from its signature, it is necessary to consider programmes which solve the non-linear optimisation problem (5.1). Note in practice the projective tensor norm $\|\cdot\|_\pi$ is difficult to compute, we can generalise the problem to a wider set of tensor norms:

Problem. For a norm function $\|\cdot\|$ which satisfies Definition 4.2.1, assume a tree-reduced continuous bounded-variation path $\gamma : [0, 1] \to \mathbb{R}^d$ with derivative $f : (0, 1) \to \mathbb{R}^d$ such that $\|f(t)\| = 1$ for all $t \in (0, 1)$ almost everywhere, and γ is differentiable at $\theta \in (0, 1)$. For all $n \geq 1$, define $g : \mathbb{R}^d \to \mathbb{R}$ such that

$$g(x) := \left\| I_{p,n}(x) - \bar{S}_{n+1} \right\|$$

for $p \in \{1, \cdots, n+1\}$. We are interested in the following optimisation problem

$$\min_{\|x\|=1} g(x). \tag{5.2}$$

Lemma 5.1.1. *There exists at least one solution to (5.2).*

Proof. We first show that g is a continuous function: for $x, y \in \mathbb{R}^d$, we have

$$\begin{aligned} |g(x) - g(y)| &\leq \|I_{p,n}(x) - I_{p,n}(y)\| \\ &= \|I_{p,n}(x-y)\| \\ &= \|x-y\| \left\|\bar{S}_n\right\|, \end{aligned}$$

so g is Lipschitz hence continuous. The set $\{x \in \mathbb{R}^d : \|x\| = 1\}$ is closed and bounded in $\mathbb{R}^d$, so there exists $x^* \in \{x \in \mathbb{R}^d : \|x\| = 1\}$ such that $g(x^*) = \min_{\|x\|=1} g(x)$. $\square$

After proving the existence, a natural question to ask is whether the solution is unique. If we have a *convex optimisation problem*, we would know the answer to this question.

Definition 5.1.1 (Convex optimisation problem). *If $F : \mathbb{R}^d \to \mathbb{R}$ is a convex function and $\mathcal{X}$ is a convex set in $\mathbb{R}^d$, then*

$$\min_{x \in \mathcal{X}} F(x)$$

is called a convex optimisation problem.

Proposition 5.1.1. *If X is a convex subset of $\mathbb{R}^d$ and $f : \mathbb{R}^d \to (-\infty, \infty]$ is a proper convex function, then a local minimum of f over X is also a global minimum of f over X. If in addition f is strictly convex, then there exists at most one global minimum of f over X. For an example one can refer to [2] for details.*

Lemma 5.1.2. *g is a convex function.*

Proof. For any $x, y \in \mathbb{R}^d$ and $0 \leq \lambda \leq 1$,

$$\begin{aligned} g(\lambda x + (1-\lambda)y) &= \left\|I_{p,n}(\lambda x + (1-\lambda)y) - \bar{S}_{n+1}\right\| \\ &= \left\|I_{p,n}(\lambda x) + I_{p,n}((1-\lambda)y) - \bar{S}_{n+1}\right\| \\ &= \left\|I_{p,n}(\lambda x) + I_{p,n}((1-\lambda)y) - \lambda\bar{S}_{n+1} - (1-\lambda)\bar{S}_{n+1}\right\| \\ &\leq \left\|\lambda(I_{p,n}(x) - \bar{S}_{n+1})\right\| + \left\|(1-\lambda)(I_{p,n}(y) - \bar{S}_{n+1})\right\| \\ &= \lambda g(x) + (1-\lambda) g(y), \end{aligned}$$

therefore g is a convex function. $\square$

However, (5.2) is not quite a convex optimisation problem, since the set

$$\left\{\|x\| = 1 : x \in \mathbb{R}^d\right\}$$

is not convex. There is no general theory about the uniqueness of the solution to optimisation over the sphere, therefore we have to explore the uniqueness of the solution depending on the tensor norm used. In the next section we will prove that if we identify $(\mathbb{R}^d)^{\otimes n}$ with $(\mathbb{R}^d)^n$, then under ℓ^2 norm the minimiser is unique using the method of Lagrange multipliers.

5.2 Application of the method of Lagrange multipliers

If $\left\|I_{p,n}(x) - \bar{S}_{n+1}\right\|$ is a smooth function under the norm we choose, then a practical method to find a minimum to the problem is using *Lagrange multipliers*. As an example, let us consider a tree-reduced d-dimensional path $\gamma : [0,1] \to \mathbb{R}^d$ parametrised at unit speed, i.e. $\|\dot{\gamma}_t\|_2 = 1$. For any $n \geq 1$, $p \in \{1, \cdots, n+1\}$, let $A \in \mathbb{R}^{d^{n+1} \times d}$ denote the matrix representing the linear mapping $I_{p,n}$, and $b \in \mathbb{R}^{d^{n+1}}$ be the normalised signature of γ at level $n+1$. We now try to find a solution to

$$\min_{x \in \mathbb{R}^d, \|x\|_2 = 1} \|Ax - b\|_2 . \tag{5.3}$$

We first note the following property of A.

Lemma 5.2.1. *Assume that in $\mathbb{R}^d$, A is same the matrix as in (5.3). The singular values of A are the same and equal to $\left\|\bar{S}_n\right\|_2$.*

Proof. Let $\{e_1, e_2, \cdots, e_d\}$ be a basis of $\mathbb{R}$, and $\bar{S}_n = \sum_{i \in I(n)} a_{i_1 i_2 \cdots i_n} e_{i_1} \otimes \cdots \otimes e_{i_n}$ for the set $I(n)$ of all words of length n over the alphabet $\{1, \cdots, d\}$. Note the matrix can be obtained by applying the map $I_{p,n}$ on the basis $\{e_1, \cdots, e_n\}$ of $\mathbb{R}$, which gives elements in $\mathbb{R}^{\otimes(n+1)}$. Therefore we can identify the entries in A by the bases of $\mathbb{R}$ and $\mathbb{R}^{\otimes(n+1)}$ simultaneously, and write entries of A as $A_{i_1 \cdots i_{n+1}, j}$ for all $i \in I(n+1)$ and $j \in \{1, \cdots, d\}$. Then by the definition of $I_{p,n}$, we have

$$A_{i_1 \cdots i_{p-1} i_p i_{p+1} \cdots i_{n+1}, j} = \begin{cases} a_{i_1 \cdots i_{p-1} i_{p+1} \cdots i_{n+1}} & \text{if } j = i_p, \\ 0 & \text{otherwise.} \end{cases}$$

Hence

$$A^T A = \begin{pmatrix} \left\|\bar{S}_n\right\|_2^2 & 0 & \cdots & 0 \\ 0 & \left\|\bar{S}_n\right\|_2^2 & 0 & \vdots \\ \vdots & 0 & \ddots & \vdots \\ 0 & \cdots & 0 & \|\bar{S}_n\|_2^2 \end{pmatrix},$$

which is a diagonal matrix with all diagonal entries equal to $\left\|\bar{S}_n\right\|_2^2$. Then by the definition of singular values, the singular values of A are equal to $\left\|\bar{S}_n\right\|_2$. $\square$

Because the objective function in (5.3) is differentiable, we can use the classical *method of Lagrange multipliers.*
Consider the non-linear programming problem

$$\begin{array}{rrl} & \text{minimise} & f(x) \\ (P) & \text{subject to} & h(x) = 0, \\ & & r(x) \leq 0, \end{array}$$

where the functions $f : \mathbb{R}^n \to \mathbb{R}$, $r : \mathbb{R}^n \to \mathbb{R}^p$ and $h : \mathbb{R}^n \to \mathbb{R}^m$ are continuously differentiable. Define $\Omega := \{x \in \mathbb{R}^n : h(x) = 0, r(x) \leq 0\}$.

Theorem 5.2.1 (Karush-Kuhn-Tucker theorem (KKT))**.** *Let $x^* \in \Omega$ be a local minimiser of the problem (P). If x^* satisfies some regularity conditions, then there exists $\lambda^* \in \mathbb{R}^m$ and $\mu^* \in \mathbb{R}^p$ such that*

$$\begin{aligned} -\nabla f(x^*) = &\sum_{i=1}^{m} \lambda_i^* \nabla h_i(x^*) + \sum_{j=1}^{p} \mu_j^* \nabla r_j(x^*), \\ &\mu_j^* \geq 0, \quad j = 1, \cdots, p, \\ &\mu_j^* r_j(x^*) = 0, \quad j = 1, \cdots, p. \end{aligned}$$

Note that when $p = 0$, i.e. when there is no inequality constraints, these multipliers are called Lagrange multipliers.

Definition 5.2.1 (Linear independence constraint qualification (LICQ))**.** *For the optimisation problem (P), the* linear independence constraint qualification *holds at $\bar{x} \in \Omega$ if the equality constraint gradients $\nabla h_i(\bar{x})$, $i = 1, \cdots, m$ and the active inequality constraint gradients $\nabla r_j(\bar{x})$, for those indices $j \in \{1, \cdots, p\}$ such that $r_j(\bar{x}) = 0$, are linearly independent.*

In fact it is a well-known fact that LICQ is a regularity condition which guarantees that the local minimisers must satisfy KKT conditions. For an example one can see from [16] that the following corollary holds.

Corollary 5.2.1. *If a local minimizer $x^* \in \Omega$ satisfies LICQ, then x^* satisfies the KKT conditions stated in Theorem 5.2.1.*

Now we can show that problem (5.3) admits a unique solution on the sphere.

Proposition 5.2.1. *There exists a unique solution to problem (5.3), and we can develop an explicit formula for the minimum using the method of Lagrange multipliers.*

Proof. Applying *singular value decomposition* on A, we can write

$$A = U\Sigma V^T,$$

where $U \in \mathbb{R}^{d^{n+1}\times d^{n+1}}$ is an orthogonal matrix, $\Sigma \in \mathbb{R}^{d^{n+1}\times d}$ is a diagonal matrix, and $V \in \mathbb{R}^{d\times d}$ is an orthogonal matrix. Note

$$\begin{aligned} \|Ax - b\|_2 &= \left\|U\Sigma V^T x - b\right\|_2 \\ &= \left\|\Sigma V^T x - U^T b\right\|_2 . \end{aligned}$$

Recall from Lemma 5.2.1 that the singular values of A are equal to $\left\|\bar{S}_n\right\|_2$. Define $\lambda := \left\|\bar{S}_n\right\|_2$, and write

$$\Sigma = \begin{pmatrix} \lambda & 0 & \cdots & 0 \\ 0 & \ddots & & \vdots \\ \vdots & & \ddots & 0 \\ 0 & \cdots & 0 & \lambda \\ 0 & \cdots & \cdots & 0 \\ \vdots & \vdots & \vdots & \vdots \end{pmatrix}, \quad V^T x = \begin{pmatrix} q_1 \\ q_2 \\ \vdots \\ q_d \end{pmatrix}, \quad U^T b = \begin{pmatrix} y_1 \\ y_2 \\ \vdots \\ y_{d^{n+1}} \end{pmatrix}.$$

Then

$$\left\|\Sigma V^T x - U^T b\right\|_2 = \left\| \begin{pmatrix} \lambda & 0 & \cdots & 0 \\ 0 & \ddots & & \vdots \\ \vdots & & \ddots & 0 \\ 0 & \cdots & 0 & \lambda \\ 0 & \cdots & \cdots & 0 \\ \vdots & \vdots & \vdots & \vdots \end{pmatrix} \begin{pmatrix} q_1 \\ q_2 \\ \vdots \\ q_d \end{pmatrix} - \begin{pmatrix} y_1 \\ y_2 \\ \vdots \\ y_{d^{n+1}} \end{pmatrix} \right\|_2$$

$$= \left\| \begin{pmatrix} \lambda q_1 - y_1 \\ \lambda q_2 - y_2 \\ \vdots \\ \lambda q_d - y_d \\ -y_{d+1} \\ \vdots \\ -y_{d^{n+1}} \end{pmatrix} \right\|_2 .$$

Note that $\|V^T x\|_2 = \|x\|_2$ since V is orthogonal. Note also that q_i for $i = 1, \cdots, d$ only appear in the first d entries, and therefore (5.3) is equivalent to

$$\min \sum_{i=1}^{d} (\lambda q_i - y_i)^2 \quad \text{subject to} \quad \sum_{i=1}^{d} q_i^2 = 1.$$

Define η to be the Lagrange multiplier. Then the extreme values of $\{q_i\}$ satisfy

$$\begin{aligned} 2\lambda(\lambda q_i - y_i) &= 2\eta q_i \quad \forall i = 1, \cdots, d, \\ \sum_{i=1}^{d} q_i^2 &= 1. \end{aligned}$$

Therefore we can get an equation about η:

$$\sum_{i=1}^{d} (\lambda y_i)^2 = (\lambda^2 - \eta)^2,$$

after solving which we can get $\eta = \lambda^2 \pm \lambda \sqrt{\sum_{i=1}^{d} y_i^2}$. Using the fact that $q_i = \frac{\lambda y_i}{\lambda^2 - \eta}$, we have the following choices of solutions:

$$\eta = \lambda^2 \pm \lambda \sqrt{\sum_{i=1}^{d} y_i^2}.$$

Therefore we have

$$q_i = \frac{y_i}{\sqrt{\sum_{j=1}^{d} y_j^2}}, \quad \forall i = 1, \cdots, d, \quad \text{or} \quad q_i = -\frac{y_i}{\sqrt{\sum_{j=1}^{d} y_j^2}}, \quad \forall i = 1, \cdots, d.$$

By Lemma 5.1.1 we know that there must exist at least one feasible global minimum to (5.3). Since we only have one equality constraint, LICQ is trivially satisfied. Hence the solutions from using the Lagrange multipliers must contain at least one local minimum by Corollary 5.2.1. Since a global minimum is also a local minimum, the global minimum(s) must be among the solutions we have obtained from using the Lagrange multipliers. Hence we can substitute the solutions into the objective function and compare.

When $q_i = \frac{y_i}{\sqrt{\sum_{j=1}^d y_j^2}}$, the objective function is equal to

$$\sum_{i=1}^{d}\left(\lambda\frac{y_i}{\sqrt{\sum_{j=1}^d y_j^2}} - y_i\right)^2 = \left(\frac{\lambda}{\sqrt{\sum_{j=1}^d y_j^2}} - 1\right)^2 \sum_{i=1}^{d} y_i^2;$$

When $q_i = -\frac{y_i}{\sqrt{\sum_{j=1}^d y_j^2}}$, the objective function is equal to

$$\sum_{i=1}^{d}\left(\lambda\frac{y_i}{\sqrt{\sum_{j=1}^d y_j^2}} + y_i\right)^2 = \left(\frac{\lambda}{\sqrt{\sum_{j=1}^d y_j^2}} + 1\right)^2 \sum_{i=1}^{d} y_i^2.$$

By assuming $\|\bar{S}_n\|_2 > 0$, we have $\lambda > 0$, then from above we can see that $q_i = \frac{y_i}{\sqrt{\sum_{j=1}^d y_j^2}}$ for $i = 1, \cdots, d$ give smaller value, therefore $\left(\frac{y_i}{\sqrt{\sum_{j=1}^d y_j^2}}\right)_{i=1,\cdots,d}$ is the global minimum. Hence the solution to (5.3) is unique. Finally we get the minimum x^* by $x = VV^T x$. □

Corollary 5.2.2. *Assume a tree-reduced continuous bounded-variation path $\gamma : [a, b] \to \mathbb{R}^d$ is of length $L > 0$ and differentiable at $\theta \in (a, b)$, and suppose γ is parametrised at an unknown speed. Then there exists a unique solution to problem*

$$\min_{x\in\mathbb{R}^d, \|x\|_2=1} \|LAx - b\|_2\,, \tag{5.4}$$

where A and b are as described in (5.3).

Proof. We have seen from Lemma 4.6.2 that if we do not know the speed of parametrisation of the path, we can solve the optimisation problem (4.28) to get a approximation of the derivative of the path. If we use ℓ^2 norm, then (4.28) becomes (5.4). Note the only difference between (5.4) and (5.3) is the constant L in front of the matrix A, therefore a similar analysis as in Proposition 5.2.1 applies, and (5.4) admits a unique solution. □

We now demonstrate some examples of inverting the signature of a path by solving (5.3). All of the following computation is done in C++, and the graphs are plotted in MATLAB. The computation of signatures used is done via the C++ library *Libalgebra* [6]. The matrix computation algorithms used are from *LAPACK* [1], and the version used is provided by Intel Math Kernel Library.

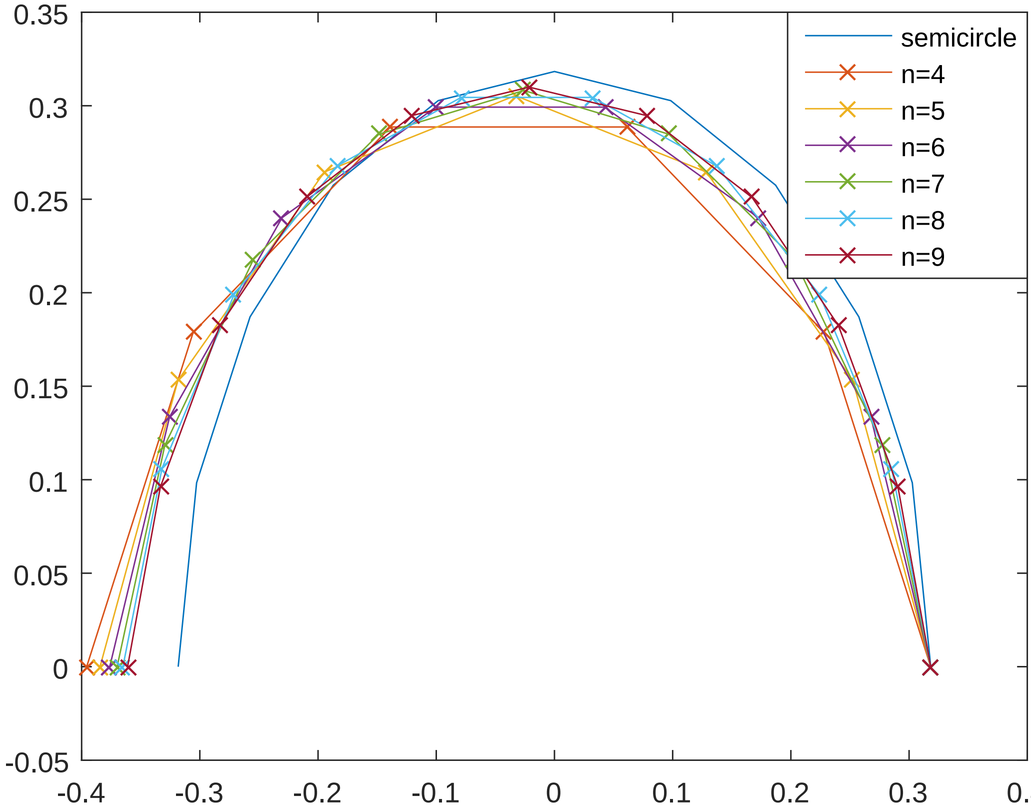


Figure 5.1: Reconstruction of a semicircle under ℓ^2 norm, where n is the level of signature used

Example 5.2.1 (Semicircle)**.** *Let* $\gamma : [0,1] \to \mathbb{R}^2$ *be the path of a semicircle, i.e.* $\gamma_t^{(1)} = \frac{1}{\pi}\cos(\pi t)$, $\gamma_t^{(2)} = \frac{1}{\pi}\sin(\pi t)$ *for* $t \in [0,1]$*. If we use* ℓ^2 *norm, we can use the formulae obtained in Proposition 5.2.1 to get an approximation to the derivative of the path at different time points. Thus we are able to approximate the increments over subintervals by Mean Value Theorem, as shown in Figure 5.1. We can see that using higher levels of signature gives better approximations to the true path.*

Example 5.2.2 (Circle)**.** *Assume in this example* $\gamma : [0,1] \to \mathbb{R}^2$ *is the path of a circle such that* $\gamma_t = (\frac{1}{2\pi}\cos(2\pi t), \frac{1}{2\pi}\sin(2\pi t))$ *for* $t \in [0,1]$*. Again we used the formulae obtained in Proposition 5.2.1 to get an approximation to the derivative at different times in* ℓ^2 *norm, therefore an approximation of the increments over the sub-intervals, as shown in Figure 5.2.*

Example 5.2.3 (Digit '8')**.** *One interesting case to consider is when the path is self-crossing. A good example of this kind is digits.*
The dataset we use is Pen-Based Recognition of Handwritten Digits Data Set *from*

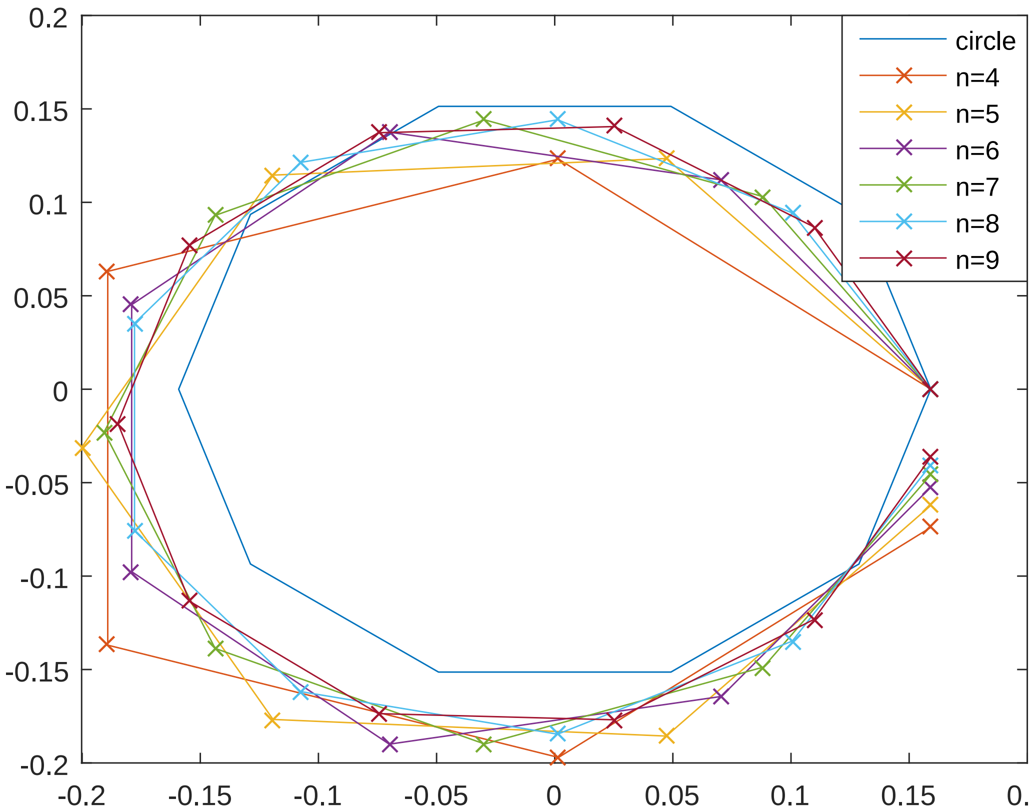


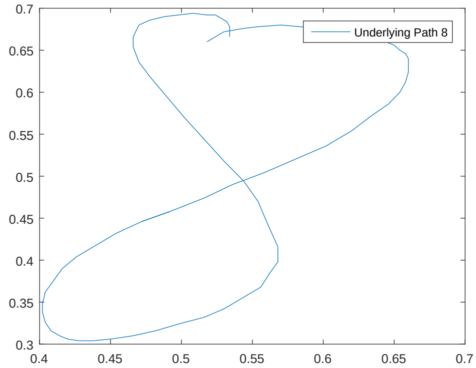
Figure 5.2: Reconstruction of a circle under ℓ^2 norm, where n is the level of signature used

UC Irvine Machine Learning Repository [14], which is a digit database by collecting 250 handwritten digit samples from 44 writers. The dataset records the (x, y) *coordinates on the* 2*-dimensional plane as the participants write. The raw data captured consists of integer values between* 0 *and* 500*, and then a resampling algorithm is applied so that the points are regularly spaced in arc length.*
We have taken one sample of the digit '8' from the training data, and normalise the input vectors so that they consist of values in $[0, 1]$*. Note that the path now is not necessarily parametrised at unit speed. In this case, we can solve a slightly altered optimisation problem by the result of Lemma 4.6.2, therefore we need an approximation of the length of the path. Due to Conjecture 2.3.1, we can approximate the length of the path by taking the n-th root of the n-th level of the signature multiplied by* $n!$.
We then reconstruct the underlying path using the method of Lagrange multipliers to approximate the derivative of the path at different points by the results of Proposition 5.2.1 and Corollary 5.2.2, and use splines to smooth the derivatives, and then integrate over $[1/(n+2), (n+1)/(n+2)]$ *in MATLAB to approximate the underlying path, where n is the level of the lower level signature used. Compared to the underlying path in Figure 5.3a, we can see from Figure 5.3b, 5.3c, 5.3d, 5.3e, 5.3f, 5.3g, and 5.3h that overall we get better approximations when we use higher levels of the signature of the path. Note that the paths reconstructed are at different scales from the underlying path. This is because we have reconstructed the path parametrised at unit speed, as shown in Lemma 4.6.2. Also note that if* $\gamma : [u, u + L] \to \mathbb{R}^d$ *is a path parametrised at unit speed and of length* L*, then*

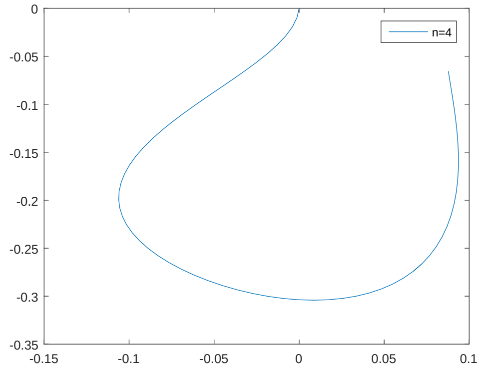
$$\begin{aligned}&\int_u^{u+L} \dot{\gamma}(t)\mathrm{d}t \\ =&L\int_0^1 \dot{\gamma}(Ls+u)\mathrm{d}s,\end{aligned}$$

therefore the path obtained is the underlying path parametrised at unit speed and scaled by $1/L$*. The shapes of the reconstructed paths are not affected even though we have a different speed of parametrisation.*

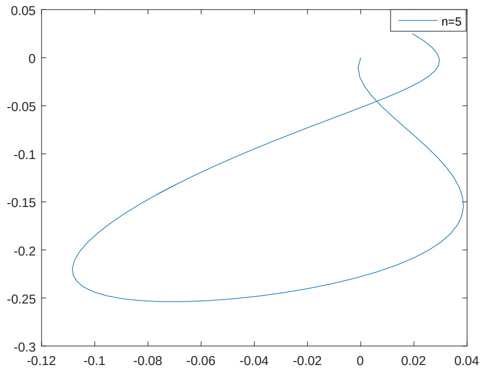
Example 5.2.4 (Robustness of the insertion method)**.** *In this example, we show that we can build a pipeline to invert the signatures of paths by the insertion method. We arbitrarily choose 20 samples from the training set (consisting of handwritten digits by 30 writers) of the Pen-Based Recognition of Handwritten Digits Data Set [14] and normalise the data as described in Example 5.2.3. Then we reconstruct the underlying path using signature level* 9 *and* 10 *using the method of Lagrange multipliers*



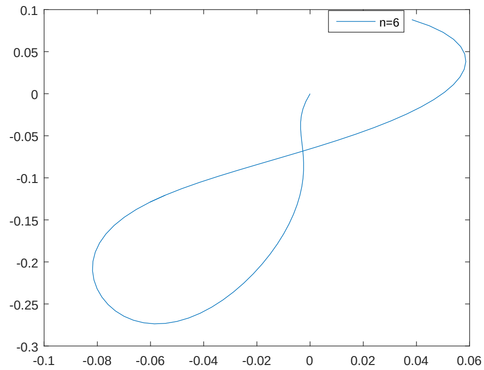
(a) Underlying path of the digit ‘8’



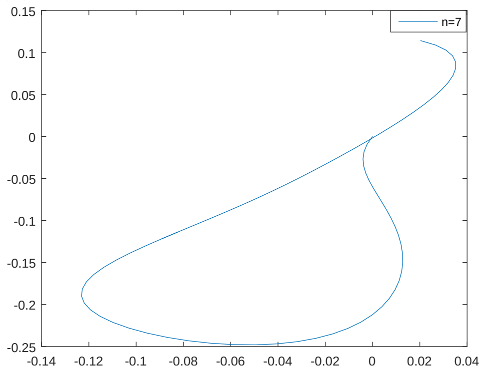
(b) Reconstruction using signature level 4



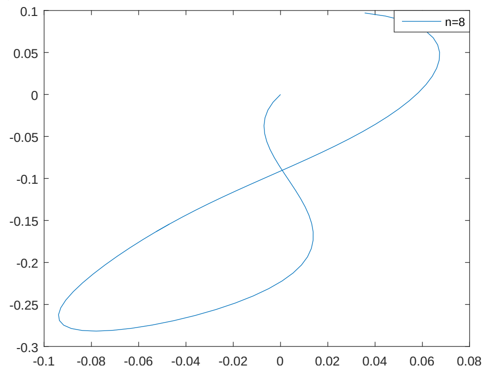
(c) Reconstruction using signature level 5



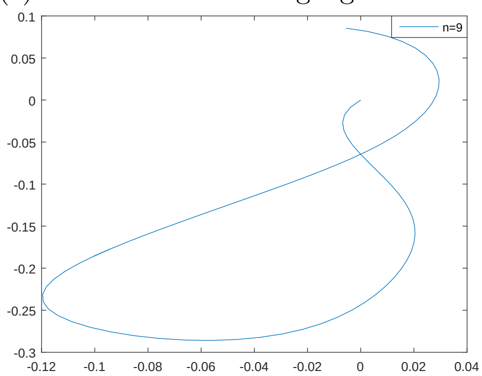
(d) Reconstruction using signature level 6



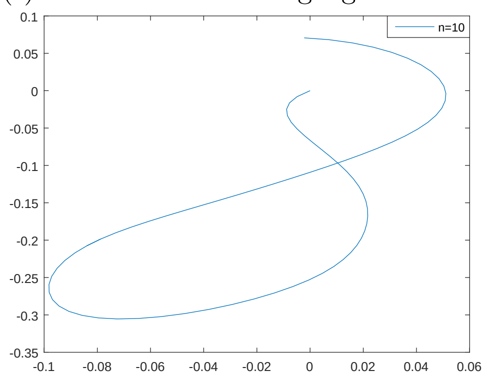
(e) Reconstruction using signature level 7



(f) Reconstruction using signature level 8



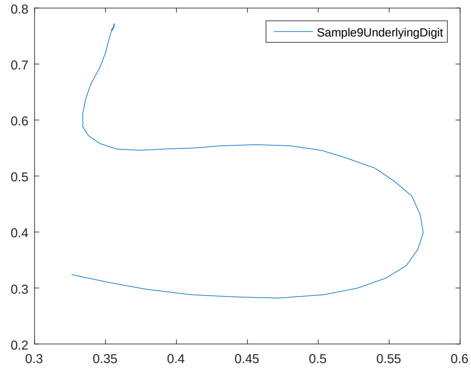
(g) Reconstruction using signature level 9



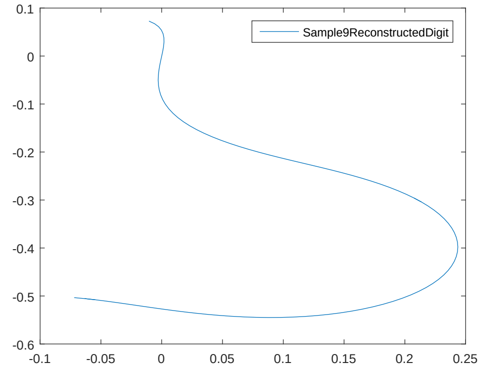
(h) Reconstruction using signature level 10

Figure 5.3: Reconstruction of the digit ‘8’ using the insertion method

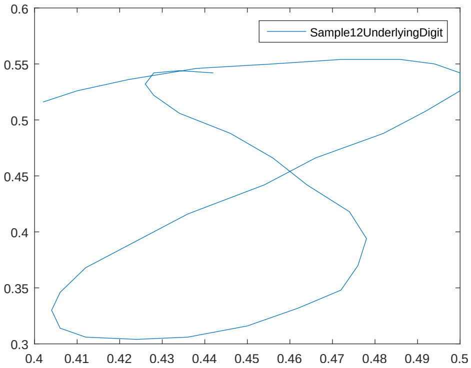
as described in Proposition 5.2.1 and Corollary 5.2.2, and obtain Figure 5.4, 5.5, 5.6, 5.7 and 5.8. Note we export the derivatives computed in C++ into MATLAB, and use the splines to approximate the derivatives, and then unlike Example 5.2.3, we integrate the splines over $[0, 1]$. *This is because signature level* 9 *is relatively higher than most of the signature levels used in Example 5.2.3, so the splines are supposed to behave better at extrapolation. We can see that the insertion method is in general quite robust, however it may not be able to give an accurate approximation at the corner of the path.*



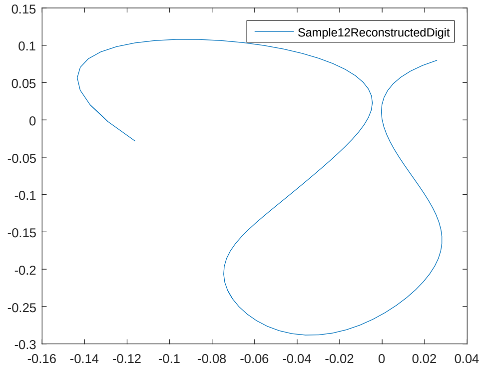
(a) Sample 9, underlying digit



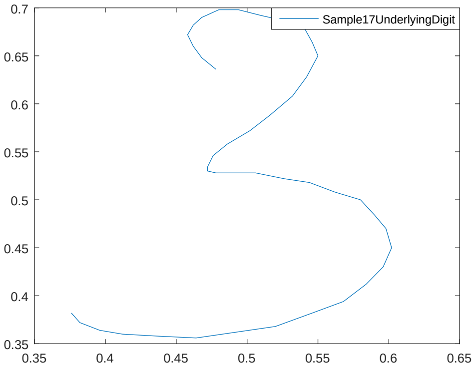
(b) Sample 9, reconstructed digit



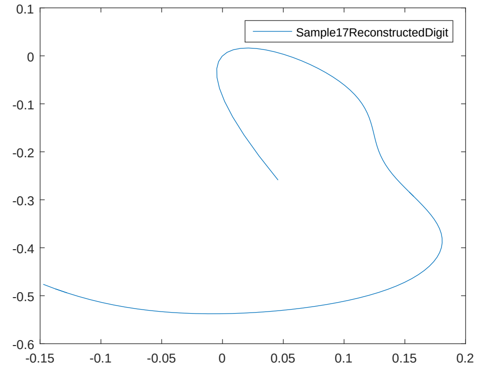
(c) Sample 12, underlying digit



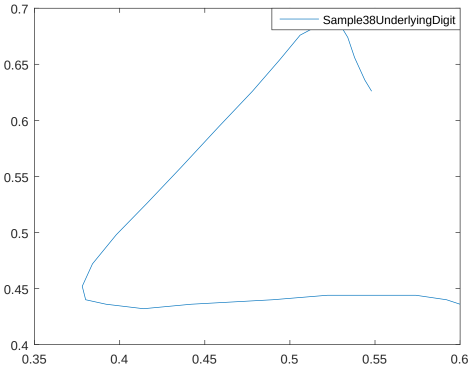
(d) Sample 12, reconstructed digit



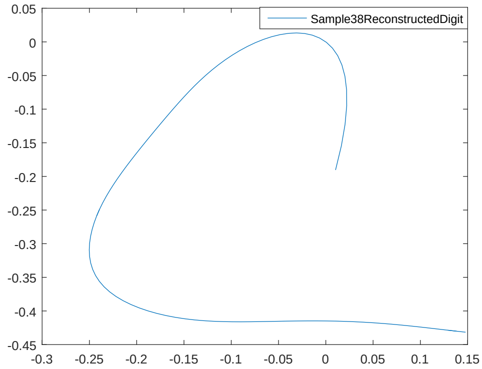
(e) Sample 17, underlying digit



(f) Sample 17, reconstructed digit

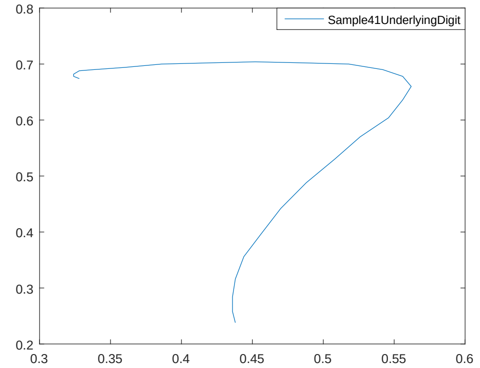


(g) Sample 38, underlying digit

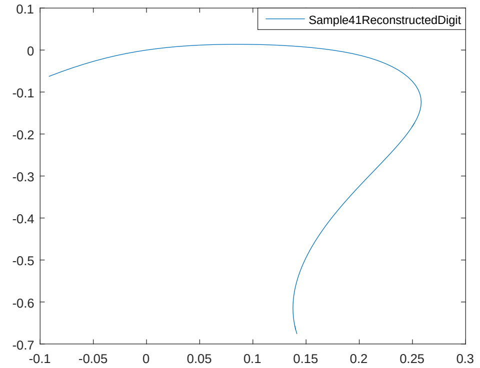


(h) Sample 38, reconstructed digit

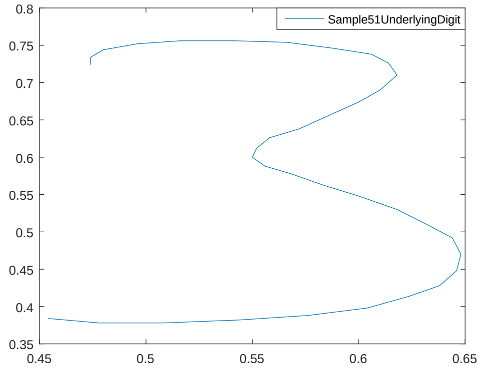
Figure 5.4: Reconstruction of digits from the data set [14] using signature level 9 and 10



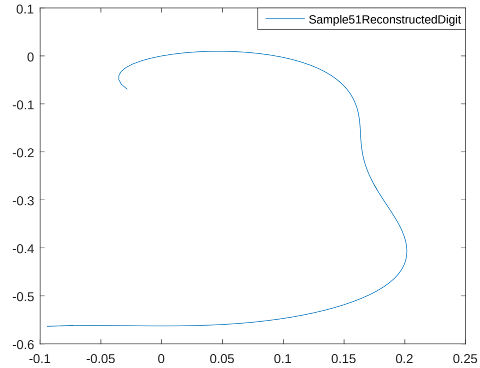
(a) Sample 41, underlying digit



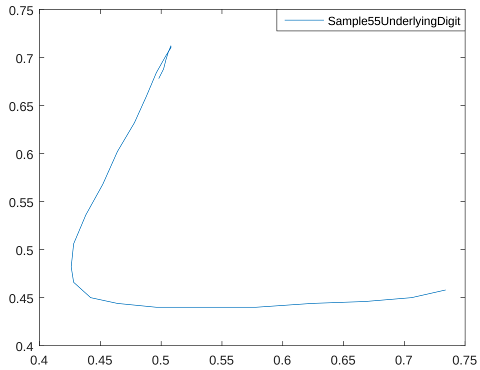
(b) Sample 41, reconstructed digit



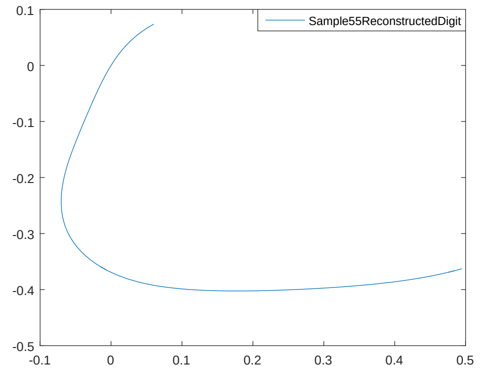
(c) Sample 51, underlying digit



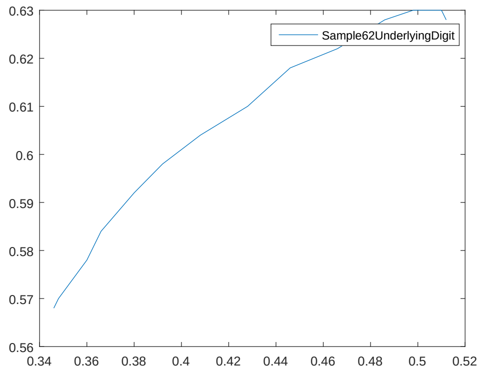
(d) Sample 51, reconstructed digit



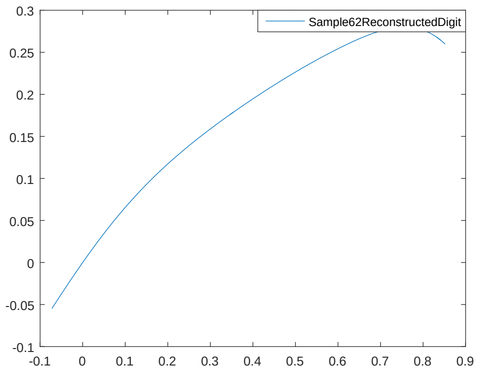
(e) Sample 55, underlying digit



(f) Sample 55, reconstructed digit

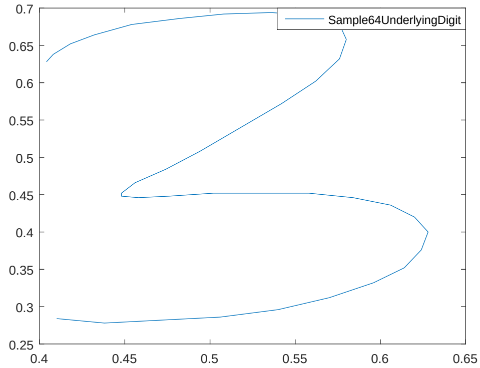


(g) Sample 62, underlying digit

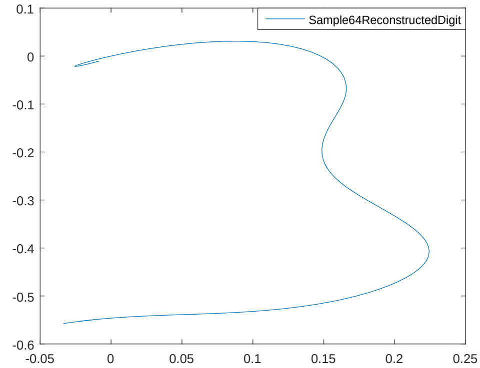


(h) Sample 62, reconstructed digit

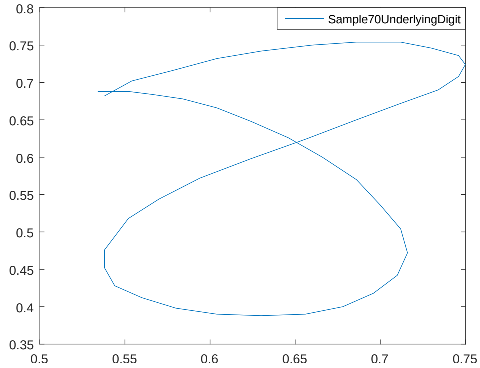
Figure 5.5: Reconstruction of digits from the data set [14] using signature level 9 and 10



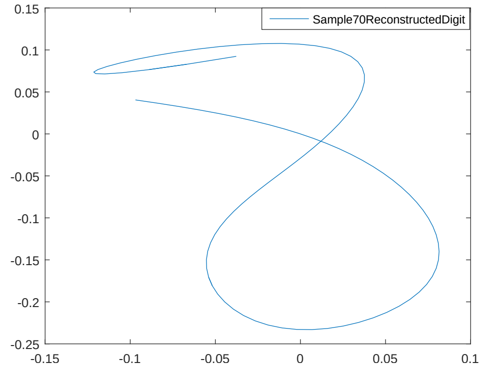
(a) Sample 64, underlying digit



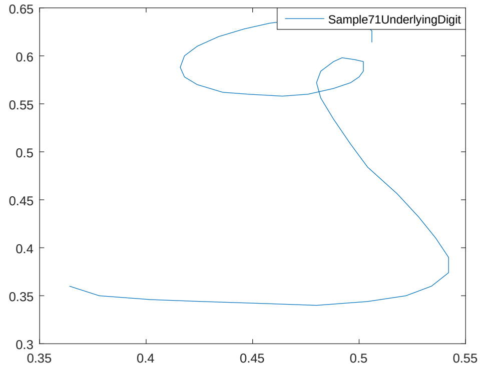
(b) Sample 64, reconstructed digit



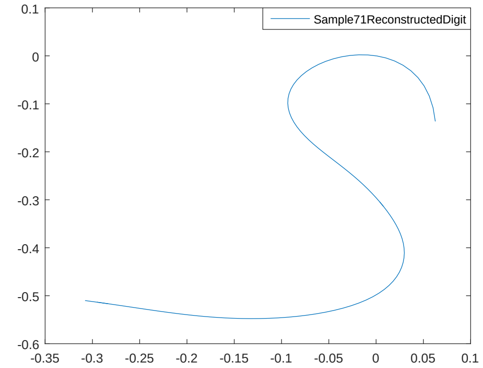
(c) Sample 70, underlying digit



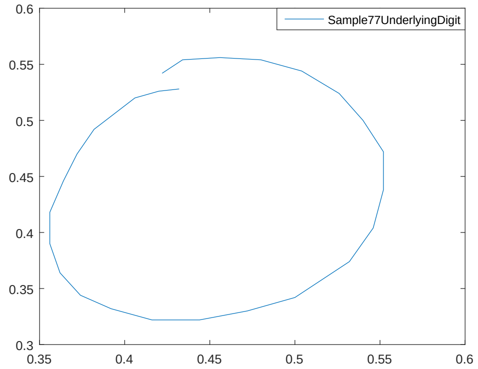
(d) Sample 70, reconstructed digit



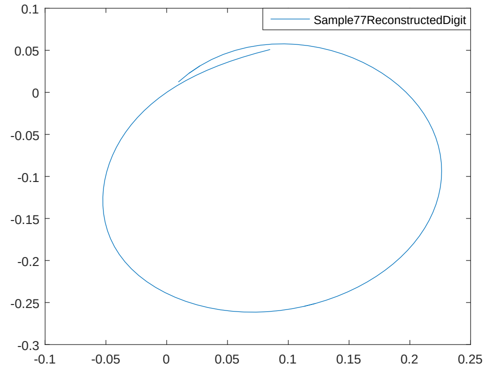
(e) Sample 71, underlying digit



(f) Sample 71, reconstructed digit

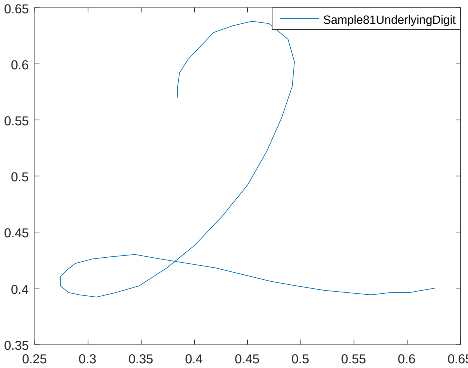


(g) Sample 77, underlying digit

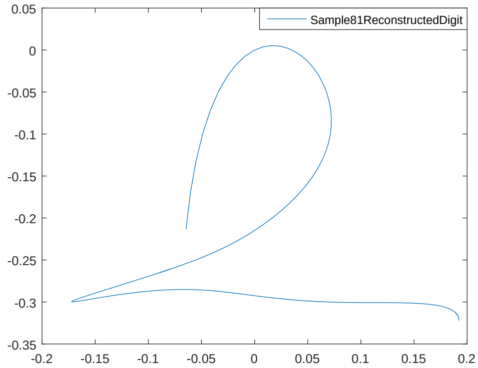


(h) Sample 77, reconstructed digit

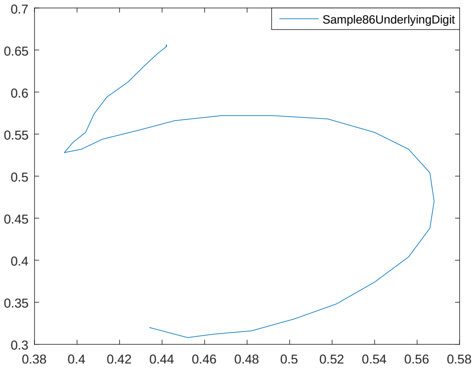
Figure 5.6: Reconstruction of digits from the data set [14] using signature level 9 and 10



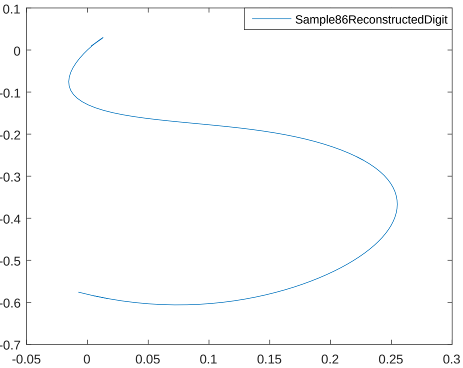
(a) Sample 81, underlying digit



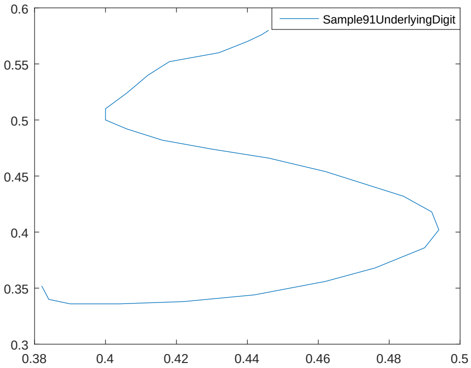
(b) Sample 81, reconstructed digit



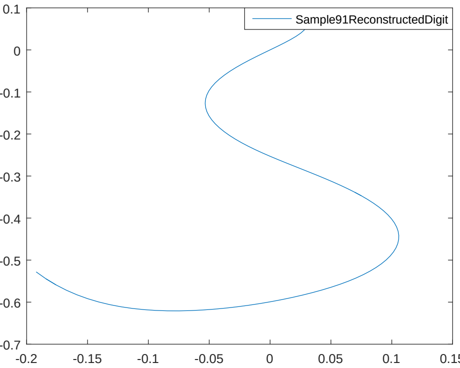
(c) Sample 86, underlying digit



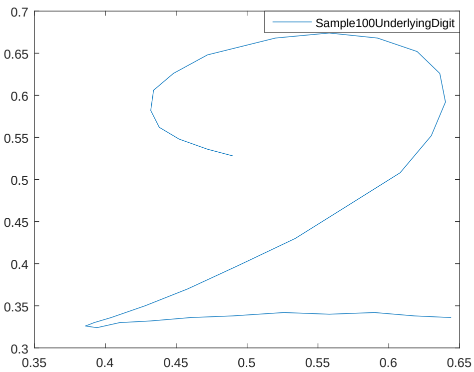
(d) Sample 86, reconstructed digit



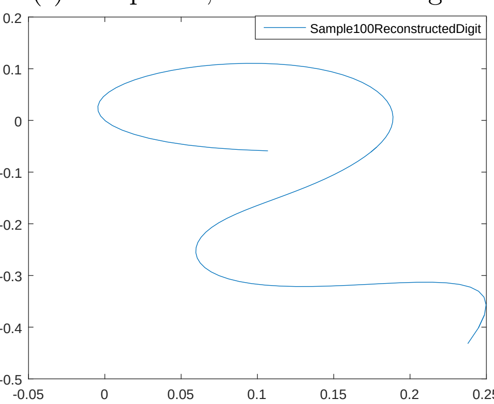
(e) Sample 91, underlying digit



(f) Sample 91, reconstructed digit



(g) Sample 100, underlying digit



(h) Sample 100, reconstructed digit

Figure 5.7: Reconstruction of digits from the data set [14] using signature level 9 and 10

Remark 5.2.1. *From a computational point of view, in general if we want to use the insertion method described in Chapter 4 to invert the signature, we need a nonlinear optimisation solver. However most of such solvers require a good initial guess. Hence when doing computation, we need to keep in mind that such factors may affect the results.*

5.3 Comparison with symmetrisation

So far we have discussed two methods of inverting the signature of a path: the symmetrisation method and the insertion method. It is interesting to compare the two methods.

Complexity. Let us first consider the complexity of the symmetrisation method. If we recall the setting of constructing the probability of increments over the sub-intervals, for $m = 0, \cdots, n_j$,

$$\mathbb{P}^N\left(\Delta x_j = \frac{m}{n_j}\frac{1}{k}, \Delta y_j = \frac{n_j - m}{n_j}\frac{1}{k}\right) = \frac{\sum_{\mathbf{n}=(n_1,\cdots,n_k),|\omega_j|_x=m} \mathcal{S}_k^{\mathbf{n}}(\boldsymbol{\ell})}{\sum_{\mathbf{n}=(n_1,\cdots,n_k)} \mathcal{S}_k^{\mathbf{n}}(\boldsymbol{\ell})},$$

where $\boldsymbol{\ell} \in L_k^{\mathbf{n}}$, and

$$L_k^{\mathbf{n}} = \left\{\boldsymbol{\ell} = (\boldsymbol{\ell}_1, \cdots, \boldsymbol{\ell}_k) : \boldsymbol{\ell}_j = (\ell_j^1, \cdots, \ell_j^d), \sum_{i=1}^{d} \ell_j^i = n_j, \forall j = 1, \cdots, k\right\},$$

for $\mathbf{n} \in \mathcal{P}_{N,k}$, and

$$\mathcal{P}_{N,k} = \left\{\mathbf{n} = (n_1, \cdots, n_k) : n_j > 0, \sum_{j=1}^{k} n_j = N\right\}.$$

Note for a given $\mathbf{n} = (n_1, \cdots, n_k)$, the number possible combinations of $\ell_j^1, \cdots, \ell_j^d$ satisfying $\sum_{i=1}^d \ell_j^i = n_j$ is

$$\sum_{i=0}^{d-1} \binom{d}{d-i}\binom{n_j - 1}{d-i-1}$$

for all $j = 1, \cdots, k$. Here we define $\binom{m}{n} = 0$ if $m < n$. Then the size of $L_k^{\mathbf{n}}$ is $\prod_{j=1}^k(\sum_{i=0}^{d-1}\binom{d}{d-i}\binom{n_j-1}{d-i-1})$. If we choose $n_1 = \cdots = n_k$, then the size of $L_k^{\mathbf{n}}$ is $[\sum_{i=0}^{d-1}\binom{d}{d-i}\binom{\frac{N}{k}-1}{d-i-1}]^k$, where N is the level of signature used, and also note the complexity comes from calculating the number of combinations of letters when fixing $\Delta x_j = \frac{m}{n_j}\frac{1}{k}$ for every $m = 0, \cdots, n_j$ and $j = 1, \cdots, k$, hence the order of complexity when all the k blocks are of equal length is $O(k(\frac{N}{k}+1)[\sum_{i=0}^{d-1}\binom{d}{d-i}\binom{\frac{N}{k}-1}{d-i-1}]^{k-1})$.

As for the insertion method, the complexity depends on the optimisation algorithm we use. If we consider the method of Lagrange multipliers, from the previous section we can see that the main complexity comes from singular value decomposition and matrix multiplication. If we use the `DGESVD` function from LAPACK [1] to compute SVD, LAPACK Users' Guide [1] suggests that `DGESVD` requires *Householder reflectors*, and from Trefethen and Bau [31] we can see that the cost of such an operation is $O(d^{N+3})$. The cost of matrix multiplication in our case is $O(d^{2N+2})$. Hence the overall complexity of the insertion method when using the method of Lagrange multipliers is $O(d^{2N+2})$.
For the ease of comparison, for the symmetrisation method we fix the length of each of block to be 1, i.e. $n_1 = \cdots = n_k = 1$. Then when the dimension $d = 2$, the complexity of the symmetrisation method is $O(N2^N)$, while the complexity of the insertion method is $O(2^{2N+2})$. Hence the symmetrisation method has an advantage over the insertion method in terms of complexity.
Accuracy. In general, the rate of convergence of the both methods shall depend on the modulus of continuity of the underlying path. We have not been able to determine the rate of convergence for both methods, therefore we need to compare the accuracy of the methods by implementation.

When using the symmetrisation method, we consider the iterated integrals coordinatewise, and recover the magnitude of the increments over the subintervals. With the symmetrisation method we are not able to recover the direction of the path directly, and this is also one of the reasons we have focused mainly on the symmetrisation algorithm for inverting monotone paths. For C^1 paths, a method of determining the direction of the path is given by Lyons and Xu [25].
The insertion method directly deals with terms in the signature of a path, which include the information about the direction of the path. Therefore the insertion method is able to reconstruct the derivative of a path with the direction.
We use the following example to demonstrate the differences between the symmetrisation and the insertion method.

Example 5.3.1. *Consider the path $\gamma : [0,1] \to \mathbb{R}^2$ parametrised at unit speed such that $\gamma^{(2)} = (\gamma^{(1)})^2$.*
If we use the symmetrisation method, we use ℓ^1 norm. When we fix the length of each block equal to 1 *and increase the number of blocks, we obtain Figure 5.10.*
If we use the insertion method and choose to use ℓ^2 norm, we can reconstruct the quadratic curve using the method of Lagrange multipliers as described in Proposition

	Time consumed(secs)
Symmetrisation in C++, using signature level $n = 5, \cdots, 11$	2.885
Insertion in C++, using signature level $n = 4, \cdots, 10$	0.702

Table 5.1: Computational time when applying the methods using different platforms

5.2.1 and obtain Figure 5.9, and here again Libalgebra [6] and LAPACK [1] are used in the computation. When reconstructing the path, for a given level of signature n, we can get an approximation of the derivative of the path at $\frac{p}{n+2}$ *for* $p = 1, \cdots, n+1$. *Then we approximate the increment of the path over the subinterval* $[\frac{p-1}{n+1}, \frac{p}{n+1}]$ *by* $\frac{1}{n+1}x^*_{\frac{p}{n+2},n}$, *where* $x^*_{\frac{p}{n+2},n}$ *is the solution to the optimisation problem defined in (5.3). Again we can see that as we take higher order of signatures, the approximation path obtained becomes closer to the underlying path.*

*We can also export the signature from C++ to use the built-in optimisation programme **`FMINCON`** in MATLAB. Here we demonstrate the result when we solve the problem under* ℓ^1 *norm in Figure 5.11.*

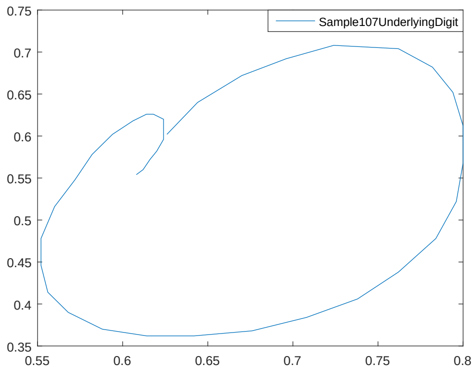
If we compare the time consumed when using each of the method, as shown in Table 5.1, we can see that the insertion method takes less time than the symmetrisation method in C++. This is different from the complexity analysis we have, however we have to bear in mind that in implementation there are other factors that affect the performance, such as the complexity generated from using Libalgebra [6]. Moreover, LAPACK [1] is highly optimised in terms of memory usage.

In terms of accuracy, we can see from Figure 5.9, 5.10 and 5.11 that both methods give results with similar accuracy. One way to measure the accuracy is to compare the levels of the signature reconstructed with the levels of the original signature used. Denote the ℓ^2 *error between the original and the reconstructed signatures at the n-th level by* e_n. *Note that the insertion method takes two levels of the signature as inputs, i.e. the insertion method requires information from two levels of the signature to construct one approximation path, while the symmetrisation method only requires one level for one reconstruction. Hence for comparison, it is reasonable to compare* e_n *from the insertion method with* e_{n+1} *from the symmetrisation method. Table 5.2 shows that the symmetrisation method gives better approximation results than the insertion method. We also note that the accuracy of the insertion method may also depend on the optimisation package chosen.*

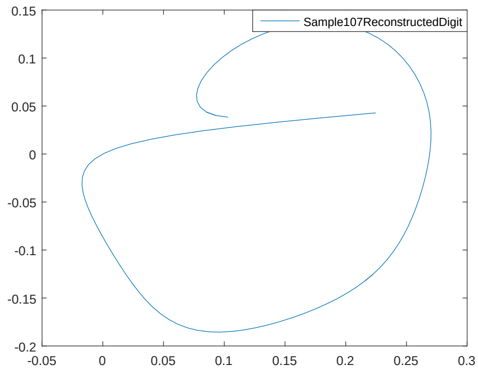
Note: The code of the computational results in Chapter 4 and Chapter 5 is available on *https://sourceforge.net/projects/signatureinversioninsertion/*.

Error	Insertion method using Lagrange multipliers	Symmetrisation method
e_4^2	1.2334e-05	-
e_5^2	4.833595e-07	3.57113e-08
e_6^2	1.28592e-08	5.28732e-10
e_7^2	2.49103e-10	5.77203e-12
e_8^2	3.68427e-12	4.84631e-14
e_9^2	4.30637e-14	3.22984e-16
e_{10}^2	4.08225e-16	1.751e-18
e_{11}^2	-	7.87613e-21

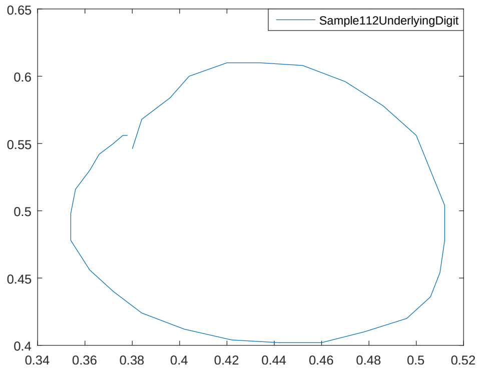
Table 5.2: Error between the reconstructed and the underlying signature at different levels



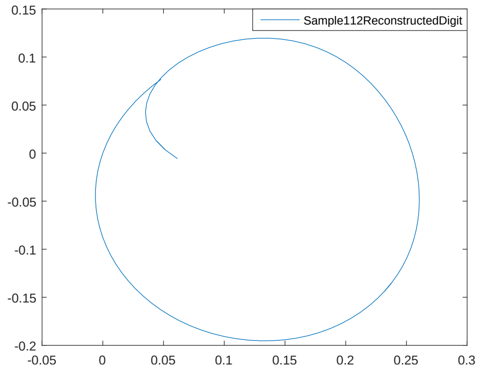
(a) Sample 107, underlying digit



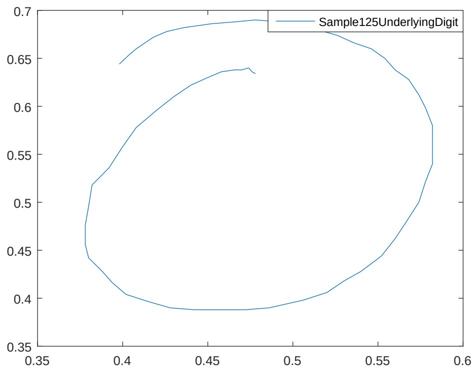
(b) Sample 107, reconstructed digit



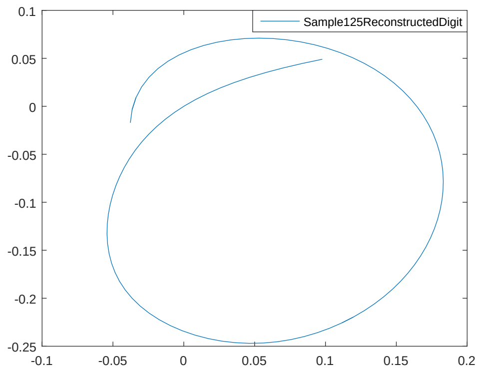
(c) Sample 112, underlying digit



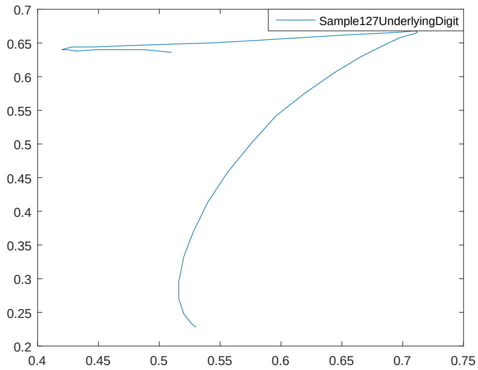
(d) Sample 112, reconstructed digit



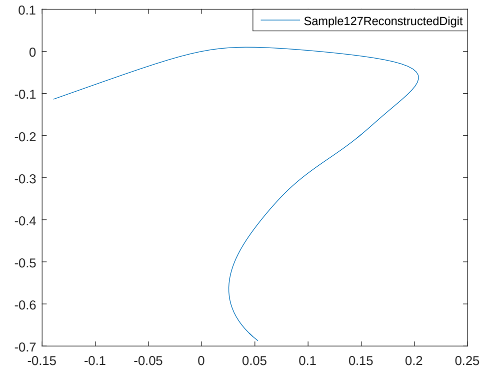
(e) Sample 125, underlying digit



(f) Sample 125, reconstructed digit



(g) Sample 127, underlying digit



(h) Sample 127, reconstructed digit

Figure 5.8: Reconstruction of digits from the data set [14] using signature level 9 and 10

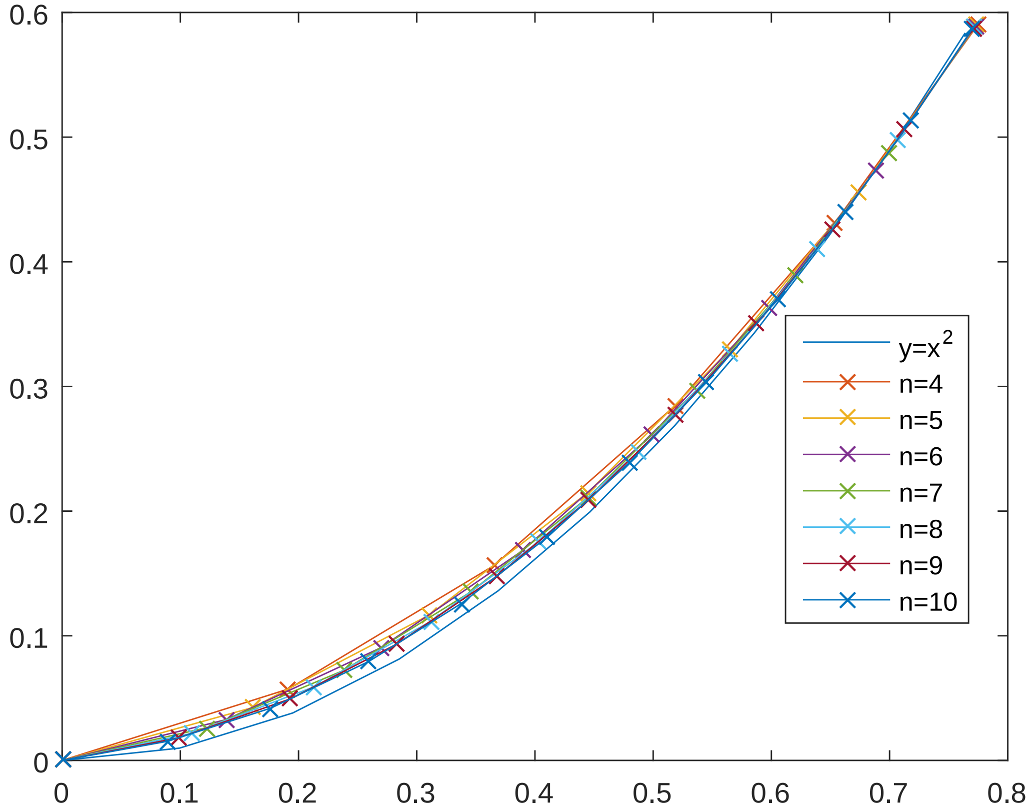


Figure 5.9: Reconstruction of $y = x^2$ under ℓ^2 norm using the method of Lagrange multipliers, where n is the level of signature

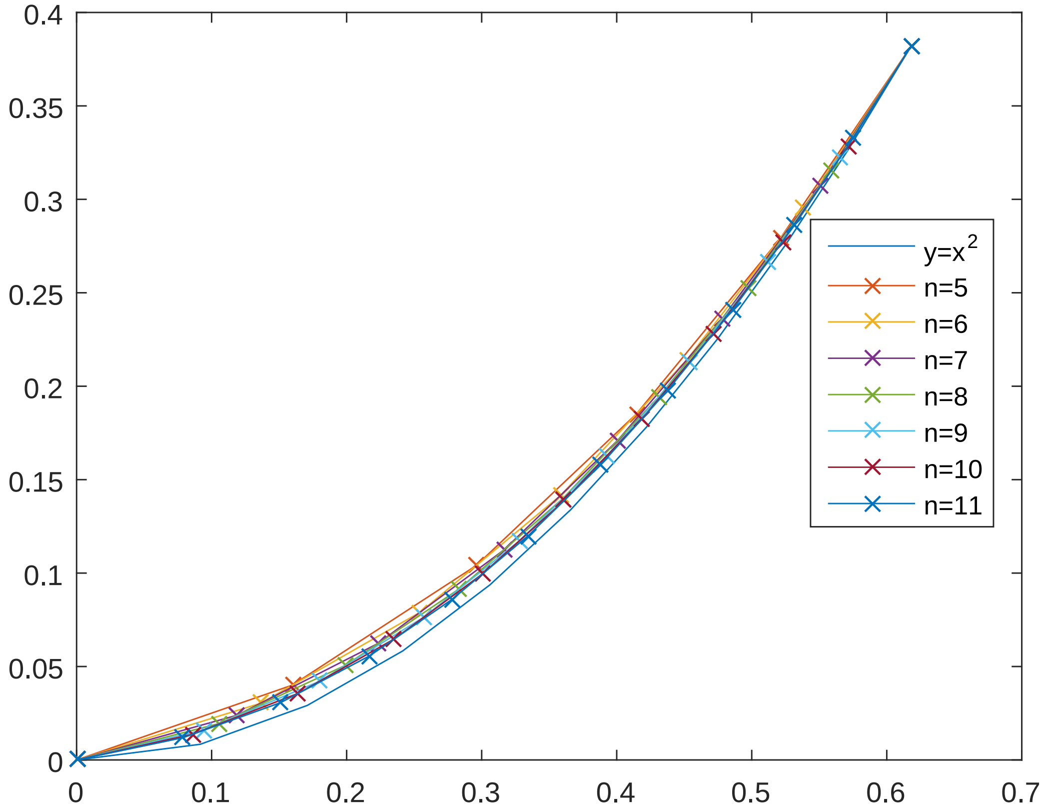


Figure 5.10: Reconstruction of $y = x^2$ using the symmetrisation method, block length=1

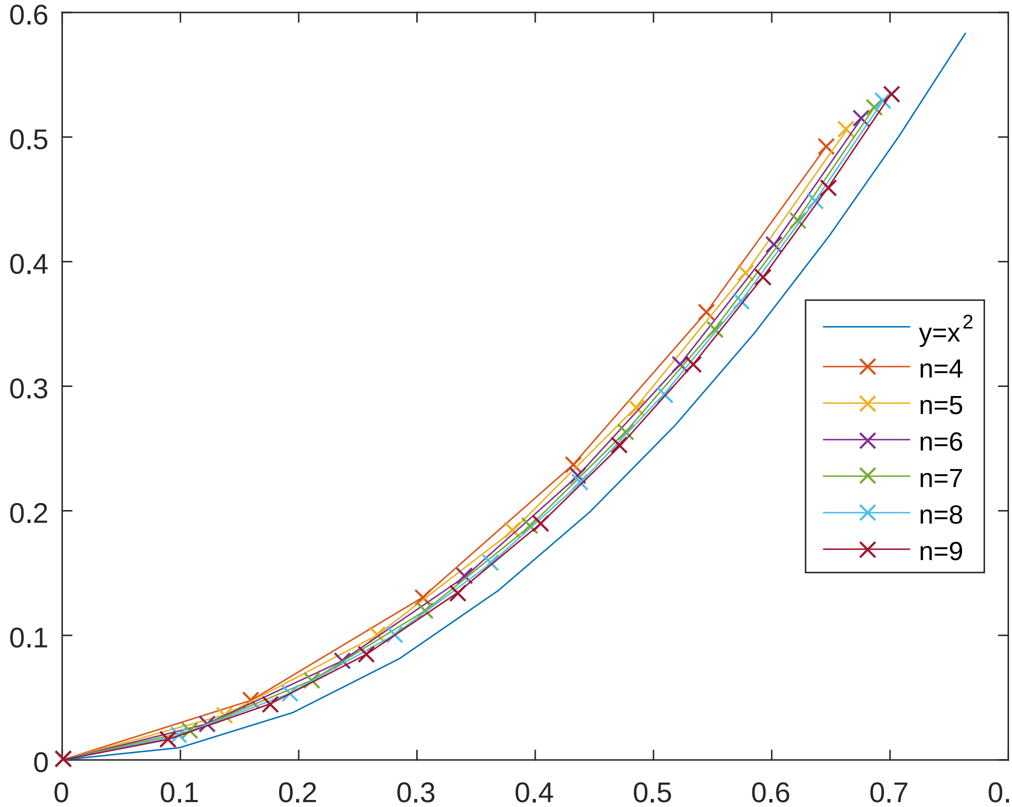


Figure 5.11: Reconstruction of $y = x^2$ under ℓ^1 norm using the optimisation programme in MATLAB, where n is the level of signature

Chapter 6

Conclusions and future work

The signature of a path has always been an important concept in rough paths theory. We have showed various properties regarding the signature of a path.
It has always been an interesting topic that whether the normalised signature is bounded below. We have proved in Chapter 2 that the n-th root of the n-th level of the signature of a path multiplied by $n!$ with finite length converges to a non-zero limit, and we have conjectured that the limit is the length of the path. In fact, Hambly and Lyons [19] showed that under the Hilbert–Schmidt norm, the n-th root of the n-th level of the signature multiplied by $n!$ of a continuously differentiable path converges to the length of the path. It is therefore an interesting problem to prove the conjecture for any path with finite length under any reasonable tensor algebra norm.
A stronger convergence was proved by Hambly and Lyons [19], which states that for path which satisfy certain smooth conditions, the n-th term of the normalised signature tends to 1 as n tends to infinity under a particular choice of norm. However, we gave an example in Chapter 4 that this strong convergence does not hold for the piecewise linear path under the Hilbert-Schmidt norm. Hence it is worth considering the conditions required on paths and norm functions to observe the strong convergence.

We have also included two methods to invert the signature of a path. The symmetrisation method, which was first introduced by Lyons and Xu [25], recovers the increments of a path over sub-intervals. We demonstrated an algorithm using the symmetrisation method to reconstruct a monotone path.
We have also introduced the insertion method, which recovers the derivative of a path at different times. The motivation is that we can approximate the $(n+1)$-th level of the normalised signature of a path by inserting the derivative of the path

into the n-th level of the normalised signature. We used the idea of concentration of measures and showed an upper bound exists for the difference between the inserted n-th level and the $(n+1)$-th level of the normalised signature of a path. Combining with the lower bound we found for a subsequence of terms in the normalised signature of a piecewise linear path in Theorem 4.5.1, we have showed that the solution to $\min_{\|x\|_2=1} \left\|I_{p,n}(x) - \bar{S}_{n+1}\right\|_\pi$ converges to the derivative of the piecewise linear path if certain constraints are satisfied. Moreover, we have also showed that the insertion method can also be used to reconstruct smooth enough paths from their signatures under some conditions in Theorem 4.6.3.
There are some interesting questions we can ask about the insertion method:

1. Theorem 4.3.1 requires knowledge about the modulus of continuity of the normalised signature of a non-degenerate piecewise linear path to develop an upper bound for the difference between $I_{p,n}$ and $\bar{S}_n$. However, it is an open question that whether the upper bounds obtained in Lemma 4.3.1 and Lemma 4.3.2 are optimal;

2. We proved in Corollary 4.5.1 that a subsequence of terms of the normalised signature of a non-degenerate piecewise linear path is bounded below by a constant under the projective tensor norm. It is then a natural question to ask whether there exists a lower bound for the normalised signature of a piecewise linear path under other tensor norms, and in general, the rate of decay of the terms in the normalised signature of a path;

3. We have seen from Chapter 5 that we can reconstruct a tree-reduced smooth path with the insertion method by using norms other than the projective tensor norm. It is then interesting to explore the feasibility of the insertion method under other tensor norms induced by a norm function on $\mathbb{R}^d$ other than the ℓ^2 norm.

If the above questions can be answered, then it is very possible that the insertion method can be used to reconstruct a more general set of paths, such as C^1 paths or even, bounded-variation paths, under some easily computable norm.

To conclude, inverting the signature of a path is an interesting topic which involves knowledge from different subjects such as rough paths, probability, algebra and analysis, and there are still some open questions in this topic awaiting to be discussed.

Appendix A

1. Proof of Lemma 1.1.3

Proof. Let $L_X(u)$ be the length of X up to time $u \in I$. Because $X(\cdot)$ is continuous, $L_X(\cdot)$ is a continuous increasing function on I. Define

$$\tau(t) := \sup\{u \in I : L_X(u) \leq t\}.$$

Note that $L_X(\tau(t)) = t$ for $t \in [0, L]$. Define $Y_t = X_{\tau(t)}$ for $t \in [0, L]$. Let $L_Y(t)$ denote the length of Y up to time t. Recall that the length of a path is defined as

$$L_X(u) := \sup_{D_X} \sum_{u_{i-1}, u_i \in D_X} \left\| X_{u_i} - X_{u_{i-1}} \right\| \quad \forall\, u \in I,$$

where D_X is a partition on $\{u' \in I : u' \leq u\}$. We can show that the length of the path is invariant under this re-parametrisation:
For a partition D_Y on $\{t' \in [0, L] : t' < t\}$, define $D_X = \{\tau(t') : t' \in D_Y\}$. Note because $\tau(\cdot)$ is not onto, more than one points in D_Y can be mapped to the same value in D_X. However, if $L_X(u) = L_X(v)$ for $u, v \in I$, then $\|X_u - X_v\| = 0$. Hence

$$\sum_{t_{i-1}, t_i \in D_Y} \left\| Y_{t_i} - Y_{t_{i-1}} \right\| = \sum_{u_{j-1}, u_j \in D_X} \left\| X_{u_j} - X_{u_{j-1}} \right\|,$$

and

$$L_Y(t) = \sup_{D_Y} \sum_{t_{i-1}, t_i \in D_Y} \left\| Y_{t_i} - Y_{t_{i-1}} \right\| = \sup_{D_X} \sum_{u_{j-1}, u_j \in D_X} \left\| X_{u_j} - X_{u_{j-1}} \right\| \leq L_X(\tau(t)).$$

Also for a partition P_X on $\{u' \in I : u' \leq \tau(t)\}$, we can define

$$P_Y = \{L_X(u') : u' \in P_X\},$$

and

$$\sum_{u_{j-1},u_j \in P_X} \left\| X_{u_j} - X_{u_{j-1}} \right\| = \sum_{t_{i-1},t_i \in P_Y} \left\| Y_{t_i} - Y_{t_{i-1}} \right\|,$$

so

$$L_X(\tau(t)) = \sup_{P_X} \sum_{u_{j-1},u_j \in P_X} \left\| X_{u_j} - X_{u_{j-1}} \right\| = \sup_{P_Y} \sum_{t_{i-1},t_{j-1} \in P_Y} \left\| Y_{t_i} - Y_{t_{i-1}} \right\| \leq L_Y(t).$$

Therefore $L_X(\tau(t)) = L_Y(t)$ for $t \in [0, L]$. Hence

$$L_Y(t) - L_Y(s) = L_X(\tau(t)) - L_X(\tau(s)) = t - s.$$

Therefore we have re-parametrised the path at a unit speed. □

2. Proof of Lemma 1.2.1

Proof. Let $D = \{0 = t_0 < ... < t_r = t\}$ be a partition of $[0, t]$. Assume $t_j \leq s < t_{j+1}$ for some $j \in \{0, ..., r-1\}$. Define $D_1 = \{t_0 < ... < t_j\}$, $D_2 = \{t_j < ... < t_r\}$. So $D_1 \cup D_2 = D$. Then

$$\begin{aligned}
&\sum_{t_i,t_{i+1} \in D} \left\| (X * Y)_{t_{i+1}} - (X * Y)_{t_i} \right\|^p \\
&= \sum_{t_i,t_{i+1} \in D_1} \left\| (X * Y)_{t_{i+1}} - (X * Y)_{t_i} \right\|^p + \sum_{t_i t_{i+1} \in D_2} \left\| (X * Y)_{t_{i+1}} - (X * Y)_{t_i} \right\|^p \\
&= \sum_{t_i,t_{i+1} \in D_1} \left\| X_{t_{i+1}} - X_{t_i} \right\|^p + \sum_{t_i,t_{i+1} \in D_2} \left\| (Y_{t_{i+1}} - Y_{t_i} \right\|^p \\
&< \infty.
\end{aligned}$$

Therefore

$$\|X * Y\|_{p,[0,t]} = \left(\sup_D \sum_{t_i,t_{i+1} \in D} \left\| (X * Y)_{t_{i+1}} - (X * Y)_{t_i} \right\|^p \right)^{\frac{1}{p}} < \infty.$$

□

3. Proof of Lemma 1.4.1

Proof. This is in fact a simplified statement of Exercise 3.15 in [23], and we provide a proof here:

$$\mathrm{sym}(S^n_{0,T}(X)) = \frac{1}{n!} \sum_{\sigma \in S(n)} \int_{0<u_1<...<u_n<T} \mathrm{d}X_{\sigma(u_1)} \otimes ... \otimes \mathrm{d}X_{\sigma(u_n)}$$

$$
\begin{aligned}
&= \frac{1}{n!} \sum_{\sigma \in S(n)} \int_{0<\sigma^{-1}(u_1)<\ldots<\sigma^{-1}(u_n)<T} \mathrm{d}X_{u_1} \otimes \ldots \otimes \mathrm{d}X_{u_n} \\
&= \frac{1}{n!} \int_{0<u_1,\ldots,u_n<T} \mathrm{d}X_{u_1} \otimes \ldots \otimes \mathrm{d}X_{u_n} \\
&= \frac{1}{n!} (X_T - X_0)^{\otimes n} \\
&= \frac{1}{n!} \left(S^1_{0,T}(X)\right)^{\otimes n}.
\end{aligned}
$$

□

4. Proof of Lemma 2.1.1

Proof. The idea used is the same as in Proposition 2.1 of [29]. Suppose V is a normed vector space. Let $\phi \in (V^{\otimes m})'$ and $\psi \in (V^{\otimes n})'$. For any $x = \sum_{i\in I} u_i \otimes v_i \in V^{\otimes(m+n)}$, where $u_i \in V^{\otimes m}$ and $v_i \in V^{\otimes n}$ for all $i \in I$,

$$
\begin{aligned}
|(\phi \otimes \psi)(x)| &= \left| \sum_{i\in I} \phi(u_i)\psi(v_i) \right| \\
&\leq \sum_{i\in I} |\phi(u_i)| \, |\psi(v_i)| \\
&\leq \sum_{i\in I} \|\phi\| \, \|u_i\| \, \|\psi\| \, \|v_i\|
\end{aligned}
$$

for any arbitrary representation of x. Hence for all $x \in V^{\otimes(m+n)}$,

$$
|(\phi \otimes \psi)(x)| \leq \|\phi\| \, \|\psi\| \, \|x\|_\pi .
$$

Therefore $\|\phi \otimes \psi\| \leq \|\phi\| \, \|\psi\|$.
Now let $u \in V^{\otimes m}$ and $v \in V^{\otimes n}$. Note by the definition of projective tensor norm, we have

$$
\|u \otimes v\|_\pi \leq \|u\| \, \|v\| .
$$

For any permutation $\sigma \in S(n)$, we have, for any representation of $x = \sum_{i\in I} v_{i_1} \otimes \cdots \otimes v_{i_n} \in V^{\otimes n}$,

$$
\begin{aligned}
\|\sigma(x)\|_\pi &= \inf \sum_{i\in I} \left\|v_{\sigma(i_1)}\right\| \cdots \left\|v_{\sigma(i_n)}\right\| \\
&= \inf \sum_{i\in I} \|v_{i_1}\| \cdots \|v_{i_n}\| \\
&= \|x\|_\pi .
\end{aligned}
$$

Hence the projective tensor norm satisfies the properties specified in Definition 2.1.3. □

5. Proof of Lemma 2.1.2.

Proof. This proof is also a generalisation of the argument in [29]. Suppose V is a normed vector space. Let $\phi \in (V^{\otimes m})'$ and $\psi \in (V^{\otimes n})'$. If $\|\phi\| = 0$ or $\|\psi\| = 0$, then for all $x \in V^{\otimes(m+n)}$, $(\phi \otimes \psi)(x) = 0$, which implies trivially that $\|\phi\| \|\psi\| = \|\phi \otimes \psi\|$. If we assume $\|\phi\| > 0$ and $\|\psi\| > 0$. Then for $x = \sum_{i \in I} u_i \otimes v_i \in V^{\otimes(m+n)}$, where $u_i \in V^{\otimes m}$ and $v_i \in V^{\otimes n}$ for all $i \in I$,

$$\begin{aligned} |(\phi \otimes \psi)(x)| &= \left| \sum_{i \in I} \phi(u_i)\psi(v_i) \right| \\ &= \|\phi\| \|\psi\| \sum_{i \in I} \frac{\phi(u_i)}{\|\phi\|} \frac{\psi(v_i)}{\|\psi\|}. \end{aligned}$$

Note that $\frac{\phi}{\|\phi\|}$ and $\frac{\psi}{\|\psi\|}$ are linear functionals of norm 1, then by the definition of the injective tensor norm,

$$|(\phi \otimes \psi)(x)| \leq \|\phi\| \|\psi\| \|x\|_\delta$$

for any arbitrary representation of x. Hence

$$\|\phi \otimes \psi\| \leq \|\phi\| \|\psi\|.$$

Let $u \in V^{\otimes m}$, and $v \in V^{\otimes n}$. By Hahn-Banach Theorem, there exist $\phi_1 \in (V^{\otimes m})'$, and $\phi_2 \in (V^{\otimes n})'$ such that $\|\phi_1\| = 1$, $\|\phi_2\| = 1$, and $\phi_1(u) = \|u\|$, $\phi_2(v) = \|v\|$. Then

$$\begin{aligned} \|u\| \|v\| &= |\phi_1 \otimes \phi_2(u \otimes v)| \\ &\leq \|\phi_1 \otimes \phi_2\| \|u \otimes v\| \\ &\leq \|u \otimes v\|_\delta. \end{aligned}$$

Let $x = \sum_{i \in I} v_{i_1} \otimes \cdots \otimes v_{i_n} \in V^{\otimes n}$ for some indexing set I where $|I| < \infty$. If $\sigma \in S(n)$, then

$$\begin{aligned} \|\sigma(x)\|_\delta &= \left| \sup_{\|\phi_j\| \leq 1} \sum_{i \in I} \phi_1(v_{i_{\sigma(1)}}) \cdots \phi_n(v_{i_{\sigma(n)}}) \right| \\ &= \left| \sup_{\|\phi_j\| \leq 1} \sum_{i \in I} \phi_1(v_{i_1}) \cdots \phi_n(v_{i_n}) \right| \\ &= \|x\|_\delta. \end{aligned}$$

Hence the injective tensor norm satisfies the properties described in Definition 2.1.3. □

6. Proof of Lemma 2.1.4

Proof. We need to prove that for $x \in H_1 \otimes H_2 \otimes H_3$, $\|x\|_{(H_1 \hat{\otimes} H_2)\hat{\otimes} H_3} = \|x\|_{H_1 \hat{\otimes}(H_2 \hat{\otimes} H_3)}$. Assume $\{e_i : i \in I\}$, $\{f_j : j \in J\}$ and $\{w_k : k \in K\}$ are orthonormal bases of H_1, H_2 and H_3 respectively. Then for $x = \sum_{i\in I, j\in J, k\in K} \lambda_{ijk} e_i \otimes f_j \otimes w_k$,

$$\|x\|^2_{(H_1 \hat{\otimes} H_2)\hat{\otimes} H_3} = \sum_{i\in I, j\in J, k\in K} \lambda^2_{ijk} = \|x\|^2_{H_1 \hat{\otimes}(H_2 \hat{\otimes} H_3)},$$

as expected. □

7. Proof of Lemma 2.1.5

Proof. By the definition of the Hilbert-Schmidt inner-product $\langle \cdot, \cdot \rangle$ on $H_1 \otimes H_2$, for all $u \in H_1$ and $v \in H_2$,

$$\begin{aligned} \langle u \otimes v, u \otimes v \rangle &= \langle u, u \rangle_1 \langle v, v \rangle_2 \\ &= \|u\|^2 \|v\|^2, \end{aligned}$$

hence

$$\|u \otimes v\| = \|u\| \|v\|.$$

By Riesz-Fréchet Theorem, H_1' and H_2' are Hilbert spaces, and for any $\phi \in H_1'$ and $\psi \in H_2'$,

$$\|\phi \otimes \psi\| = \|\phi\| \|\psi\|.$$

We can show that the tensor product of duals here match with Definition 2.1.2: If $\{e_i : i \in I\}$ is an orthonormal basis of H_1 and $\{e_j' : j \in J\}$ is an orthonormal basis of H_2, we denote the corresponding dual bases as $\{f_i : i \in I\}$ and $\{f_j' : j \in J\}$ respectively. Then $\{e_i \otimes e_j' : i \in I, j \in J\}$ is an orthonormal basis of $H_1 \otimes H_2$, and $\{f_i \otimes f_j' : i \in I, j \in J\}$ is the corresponding dual basis of $H_1' \otimes H_2'$, and

$$f_i \otimes f_j'(e_r \otimes e_s') = \begin{cases} 1 & \text{if } i = r, j = s \\ 0 & \text{otherwise.} \end{cases}$$

Note this coincides with Definition 2.1.2. □

8. Proof of Lemma 2.1.6

Proof. We are left to check that the Hilbert-Schmidt norm is invariant under permutations: if $\{e_i : i \in I\}$ is an orthonormal basis of H, for any $v = \sum_{j_k \in I} \beta_i e_{j_1} \otimes \cdots \otimes e_{j_n} \in H^{\otimes n}$, for any $\sigma \in S(n)$,

$$\begin{aligned}\|\sigma(v)\|^2 &= \sum_{j_k \in I} \left\langle \beta_j e_{j_{\sigma(1)}} \otimes \cdots \otimes e_{j_{\sigma(n)}}, \beta_j e_{j_{\sigma(1)}} \otimes \cdots \otimes e_{j_{\sigma(n)}} \right\rangle \\ &= \sum_{j_k \in I} \beta_j^2 \\ &= \|v\|^2 .\end{aligned}$$

Hence the Hilbert-Schmidt norm is a reasonable tensor algebra norm. □

9. Proof of Proposition 3.1.1

Proof. For all words ω of length m and $|\omega|_x = \ell$, $|\omega|_y = m - \ell$, we have

$$\begin{aligned}\sum_{|\omega|=m, |\omega|_x=\ell,, |\omega|_y=m-\ell} C(\omega) &= \frac{1}{\ell!(m-\ell)!} \int_{0<u_1,\dots,u_m<1} \mathrm{d}\gamma_{u_1}^{i_1} \dots \mathrm{d}\gamma_{u_m}^{i_m} \\ &= \frac{1}{\ell!(m-\ell)!} (\Delta x)^{\ell} (\Delta y)^{m-\ell},\end{aligned}$$

where $|e_{j_1} \dots e_{j_m}|_x = \ell$ and $|e_{j_1} \dots e_{j_m}|_y = m - \ell$, and Δx, Δy are the increments of γ over $[0, 1]$. Note the factorials come from the fact that the repeated integrals are not counted. Hence

$$\begin{aligned}\mathcal{S}_k^{2n}(\omega, \boldsymbol{\ell}) &= \int_{\Delta_{k-1}} \prod_{j=1}^{k} \binom{2n}{2\ell_j} (\Delta x_j)^{2\ell_j} (\Delta y_j)^{2n-2\ell_j} \mathrm{d}\gamma_{u_1} \dots \mathrm{d}\gamma_{u_{k-1}} \\ &= \int_{\Delta_{k-1}} \prod_{j=1}^{k-1} \dot{\gamma}_{u_j} \prod_{j=1}^{k} \binom{2n}{2\ell_j} (\Delta x_j)^{2\ell_j} (\Delta y_j)^{2n-2\ell_j} \mathrm{d}u_1 \dots \mathrm{d}u_{k-1}.\end{aligned}$$

□

10. Proof of Lemma 4.2.4

Proof.

$$\begin{aligned}\mathbb{P}(X - a \geq t) &= \mathbb{P}(X - \mu \geq a - \mu + t) \\ &= \mathbb{P}(\exp(\lambda(X - \mu)) \geq \exp(\lambda(a - \mu + t))) \quad \forall \lambda > 0 \\ &\leq \frac{\mathbb{E}[\exp(\lambda(X - \mu))]}{\exp(\lambda(a - \mu + t))} \quad \text{by Markov's inequality}\end{aligned}$$

$$
\begin{aligned}
&\leq \exp\left(\frac{\lambda^2\sigma_{opt}^2}{2} - \lambda(a-\mu+t)\right) \quad \text{by sub-Gaussian property} \\
&\leq \exp\left(\frac{\lambda^2\sigma_0^2}{2} - \lambda(a-\mu+t)\right).
\end{aligned}
$$

Similarly,

$$
\begin{aligned}
\mathbb{P}(X-a \leq -t) &= \mathbb{P}(-X+a \geq t) \\
&= \mathbb{P}(-X+\mu \geq \mu - a + t) \\
&= \mathbb{P}\left(\exp(-\lambda(X-\mu)) \geq \exp\left(\lambda(\mu-a+t)\right)\right) \quad \forall \lambda > 0 \\
&\leq \frac{\mathbb{E}[\exp(-\lambda(X-\mu))]}{\exp(\lambda(\mu-a+t))} \\
&\leq \exp\left(\frac{\lambda^2\sigma_0^2}{2} - \lambda(\mu-a+t)\right).
\end{aligned}
$$

Then

$$
\mathbb{P}\left(|X-a| \geq t\right) \leq \exp\left(\frac{\lambda^2\sigma_0^2}{2} - \lambda t\right)\left(\exp(\lambda(\mu-a)) + \exp(-\lambda(\mu-a))\right),
$$

and choosing $\lambda = \frac{t}{\sigma_0^2}$ gives

$$
\mathbb{P}\left(|X-a| \geq t\right) \leq \exp\left(-\frac{t^2}{2\sigma_0^2}\right)\left(\exp\left(\frac{t(\mu-a)}{\sigma_0^2}\right) + \exp\left(\frac{-t(\mu-a)}{\sigma_0^2}\right)\right).
$$

□

Bibliography

[1] E. Anderson, Z. Bai, C. Bischof, L. S. Blackford, J. Demmel, Jack J. Dongarra, J. Du Croz, S. Hammarling, A. Greenbaum, A. McKenney, and D. Sorensen. *LAPACK Users' Guide (Third Ed.)*. Society for Industrial and Applied Mathematics, Philadelphia, PA, USA, 1999.

[2] Dimitri P Bertsekas, Angelia Nedi, Asuman E Ozdaglar, et al. *Convex Analysis and Optimization*. Athena Scientific, 2003.

[3] David P Blecher and Vern I Paulsen. Tensor products of operator spaces. *Journal of Functional Analysis*, 99(2):262–292, 1991.

[4] Horatio Boedihardjo and Xi Geng. A non-vanishing property for the signature of a path. *arXiv preprint arXiv:1808.05903*, 2018.

[5] Horatio Boedihardjo and Xi Geng. The tail asymptotics of the brownian signature. *to appear on Transactions of the American Mathematical Society*, 2018. `DOI:https://doi.org/10.1090/tran/7683`.

[6] Stephen Buckley, Djalil Chafai, Lajos Gyurko, Arend Janssen, and Terry Lyons. Libalgebra C++ Package, Computational Rough Paths. `https://sourceforge.net/projects/coropa/`.

[7] Clément Canonne. A short note on Poisson tail bounds. `http://www.cs.columbia.edu/~ccanonne/files/misc/2017-poissonconcentration.pdf`, 2017.

[8] Jiawei Chang, Nick Duffield, Hao Ni, Weijun Xu, et al. Signature inversion for monotone paths. *Electronic Communications in Probability*, 22, 2017.

[9] Jiawei Chang, Terry Lyons, and Hao Ni. Super-multiplicativity and a lower bound for the decay of the signature of a path of finite length. *Comptes Rendus Mathematique*, 2018.

[10] Kuo-sai Chen. Integration of paths–a faithful representation of paths by non-commutative formal power series. *Transactions of the American Mathematical Society*, 89(2):395–407, 1958.

[11] Kuo-Tsai Chen. Integration of paths, geometric invariants and a generalized Baker-Hausdorff formula. *Annals of Mathematics*, pages 163–178, 1957.

[12] Kuo-Tsai Chen. Iterated path integrals. *Bulletin of the American Mathematical Society*, 83(5):831–879, 1977.

[13] Ilya Chevyrev and Andrey Kormilitzin. A primer on the signature method in machine learning. *arXiv preprint arXiv:1603.03788*, 2016.

[14] Dua Dheeru and Efi Karra Taniskidou. UCI Machine Learning Repository. University of California, Irvine, School of Information and Computer Sciences, 2017. `http://archive.ics.uci.edu/ml`.

[15] Edward G Effros and Zhong-Jin Ruan. Self-duality for the Haagerup tensor product and Hilbert space factorizations. *Journal of Functional Analysis*, 100(2):257–284, 1991.

[16] Rodrigo G Eustaquio, Elizabeth W Karas, and Ademir A Ribeiro. Constraint qualifications for nonlinear programming. *Federal University of Parana*, 2008.

[17] Xi Geng. Reconstruction for the signature of a rough path. *Proceedings of the London Mathematical Society*, 114(3):495–526, 2017.

[18] Benjamin Graham. Sparse arrays of signatures for online character recognition. *arXiv preprint arXiv:1308.0371*, 2013.

[19] Ben Hambly and Terry Lyons. Uniqueness for the signature of a path of bounded variation and the reduced path group. *Annals of Mathematics*, pages 109–167, 2010.

[20] Wassily Hoeffding. Probability inequalities for sums of bounded random variables. *Journal of the American statistical association*, 58(301):13–30, 1963.

[21] Daniel Levin, Terry Lyons, and Hao Ni. Learning from the past, predicting the statistics for the future, learning an evolving system. *arXiv preprint arXiv:1309.0260*, 2013.

[22] Terry J Lyons. Differential equations driven by rough signals. *Revista Matemática Iberoamericana*, 14(2):215–310, 1998.

[23] Terry J Lyons, Michael Caruana, and Thierry Lévy. *Differential equations driven by rough paths*. Springer, 2007.

[24] Terry J Lyons and Weijun Xu. Hyperbolic development and inversion of signature. *Journal of Functional Analysis*, 272(7):2933–2955, 2017.

[25] Terry J Lyons and Weijun Xu. Inverting the signature of a path. *Journal of the European Mathematical Society*, 20(7):1655–1687, 2018.

[26] Olivier Marchal, Julyan Arbel, et al. On the sub-Gaussianity of the beta and Dirichlet distributions. *Electronic Communications in Probability*, 22, 2017.

[27] Max Pfeffer, Anna Seigal, and Bernd Sturmfels. Learning paths from signature tensors. *arXiv preprint arXiv:1809.01588*, 2018.

[28] José Carlos Rosales and Pedro A García-Sánchez. *Numerical semigroups*, volume 20. Springer Science & Business Media, 2009.

[29] Raymond A Ryan. *Introduction to tensor products of Banach spaces*. Springer Science & Business Media, 2013.

[30] J Michael Steele. *Probability theory and combinatorial optimization*, volume 69. Siam, 1997.

[31] Lloyd N Trefethen and David Bau III. *Numerical linear algebra*, volume 50. Siam, 1997.

[32] Gerd Wittstock, B Betz, HJ Fischer, A Lambert, K Louis, M Neufang, and I Zimmermann. What are operator spaces. *An online dictionary. URL:http://www.math.uni-sb.de/~ag-wittstock/projekt2001.html*, 2001.

[33] Nicholas Young. *An introduction to Hilbert space*. Cambridge University Press, 1988.