Imperial College London

Imperial College London

Department of Mathematics

Signature Inversion via Orthogonal Polynomials

Author:
Zhuoyue Huang

Supervisor:
Dr Cristopher Salvi

Github Link: `https://github.com/Zhuoyue-Huang/M3R-signature-inversion`

June 13, 2023

Declaration

The work contained in this thesis is my own work unless otherwise stated.

Signed: Zhuoyue Huang Date: June 13, 2023

Acknowledgement

I wish to first express my profound gratitude to my supervisor, Dr Cristopher Salvi, for his invaluable guidance, expertise, and support throughout the progression of this thesis. Through our weekly meetings, Cris consistently nurtured my ideas, fostering an environment that encouraged curiosity and innovation. I am incredibly fortunate to have had the privilege of his mentorship, and the lessons imparted will indubitably shape my future endeavours. Moreover, I would like to express my deepest appreciation to my parents, Mrs Jing Li and Mr Bo Huang, for their love, support, and encouragement throughout my academic journey. Their unwavering support, both emotionally and financially, has allowed me to pursue my dreams and complete this thesis. I would like to extend my sincere appreciation to Fengzhe Zhang, whose heartfelt support has made this challenging journey more enjoyable. Thank you all for being a part of this milestone in my life.

Abstract

In this thesis, we propose a method to reconstruct a path from its truncated signature. This is achieved through explicit recursive formulae for the coefficients of an orthogonal polynomial expansion of the path, represented as linear functionals of the signature. A key example of our approach is the application of 'shift-and-scale' Hermite polynomials, which facilitates the accurate point-wise recovery of the path with relatively small errors. Consequently, this leads to the successful implementation of point-wise signature inversion utilizing Hermite polynomials. The novel techniques proposed in this study offer an innovative insight into the field of signature inversion and potentially pave the way for more accurate and efficient computational methods.

Contents

List of Figures

Introduction

The study of path-dependent problems arises in various scientific disciplines, such as finance, fluid dynamics, and machine learning. Traditional methods like Fourier analysis often struggle to tackle these problems due to their inherent complexity and nonlinearity (e.g. [1, 2]). This difficulty motivates the search for new mathematical tools capable of capturing and summarizing the essential information contained in paths, ultimately leading us to the concept of signatures. Because of rich geometric and deterministic properties encoded in signatures, under regularity conditions, a path can be reconstructed from its signatures, leading to the study of signature inversion.

This thesis, divided into three chapters, focuses on a novel method of signature inversion using orthogonal polynomials. In Chapter 1, we introduce the basic definition and properties of signatures, alongside the concept of universality, which characterises the ability to approximate non-linear functions by linear functionals in feature space [3]. Chapter 2 introduces two methods of path decomposition via orthogonal polynomials and Fourier series. Armed with these tools, the coefficients for the bases can be found by the universality concept of signatures in Chapter 3, where all contents are original and novel. Moreover, we elucidate the orthogonal polynomial inversion method and its results on randomly generated paths. The thesis concludes by comparing the performance of the analytical inversion method via orthogonal polynomials and a learning-based method via Fourier series across a range of paths.

Chapter 1

Signatures

1.1 Preliminaries and definitions

1.1.1 Paths

Before elaborating on the signature, we shall first introduce the definition of a path and its integral. A path $x : [a, b] \to V$ is defined as a continuous and differentiable mapping from interval $[a, b]$ to V, where V denotes a d-dimensional Banach space. Normally, the interval $[a, b]$ is treated as time to indicate the time dependency of a path. In this case, we could write a high-dimensional path by its coordinates

$$x = x_t = (x_t^1, x_t^2, \dots, x_t^d),$$

where $t \in [a, b]$ and each coordinate $x^i : [a, b] \to \mathbb{R}$ can be viewed as a 1-dimensional path. For example, $x = (t, t^2)$ for $t \in [-10, 10]$ can be roughly seen as a 'U' shape in $\mathbb{R}^2$. It is also a time-augmented path that encodes the time information in the path.

Now we will introduce the path integral for 1-dimensional paths. For a fixed function $f : \mathbb{R} \to \mathbb{R}$, let $y_t = f(x_t)$, we can define the path integral of y against x as

$$\int_a^b y_t dx_t = \int_a^b f(x_t) dx_t = \int_a^b f(x_t)\dot{x}_t dt = \int_a^b y_t \dot{x}_t dt,$$

where $\dot{x}_t$ is the derivative of x_t with respect to t, i.e. dx_t/dt. Consider $f(x) = -x$ and $x_t = t^2$ for $t \in [0, 1]$ as an example, path integral of x against f can be written as

$$\int_0^1 -x_t dx_t = \int_0^1 -t^2 \cdot 2t dt = -\frac{1}{2}.$$

The requirement that paths are differentiable can be relaxed further. Therefore, we introduce *p-variation* for less restrictive paths.

Definition 1.1.1 (p-variation)**.** Let $1 \le p \le \infty$. The *p-variation* of a path $x : [a, b] \to V$ on a subinterval $[c, d] \subset [a, b]$ is

$$\|x\|_{p,[c,d]} = \left(\sup_{\mathcal{D} \subset [c,d]} \sum_{t_i \in \mathcal{D}} \left\| x_{t_{i+1}} - x_{t_i} \right\|^p \right)^{1/p},$$

where $\|\cdot\|$ is any norm on V and $\mathcal{D}$ consists all increasing and finite sequences on $[c, d]$. We denote $C_p([a, b], V)$ the space of paths x such that it has finite p-variation on $[a, b]$, i.e. $C_p([a, b], V) = \left\{ x : [a, b] \to V \text{ s.t. } \|x\|_{p,[a,b]} < \infty \right\}$. The notation of $C_p([a, b], V)$ will be simplified to $C_p(V)$ as a short-hand notation. We say x has *bounded variation* if $x \in C_1(V)$.

In this thesis, to guarantee the existence of iterated integrals defined later [4], we will assume paths are in $C_p(V)$ for $1 \le p < 2$.

1.1.2 Signatures of a path

With basic knowledge of paths, now we define signatures of a path based on [5] by iterated integrals. Let a path $x \in C_p(V)$ for $1 \leq p < 2$ with the i-th coordinate path x_t^i. We define

$$S(x)_{a,t}^i = \int_{a<s<t} dx_s^i = x_t^i - x_a^i, \tag{1.1.1}$$

which can be interpreted as the increment of the i-th coordinate path from a to t. In this case, one can observe that $S(x)_{a,t}^i$ is similar to x^i as they are both 1-dimensional real-valued paths. By collecting all possible terms of $S(x)_{a,t}^i$ in order, we define

$$S(x)_{a,t}^{(1)} = (S(x)_{a,t}^i : i \in \{1, \dots, d\}) \in V.$$

Now we can define the *double-iterated* integral for any pair of indices $i, j \in \{1, \dots, d\}$ as

$$S(x)_{a,t}^{i,j} = \int_{a<s<t} S(x)_{a,s}^i dx_s^j = \int_{a<s<t} \int_{a<r<s} dx_r^i dx_s^j := \int_{a<r<s<t} dx_r^i dx_s^j, \tag{1.1.2}$$

where $S(x)_{a,s}^i$ is defined previously in (1.1.1) and $a < r < s < t$ is a short-hand notation of two intervals with respect to r and s. Given $S(x)_{a,s}^i$ and x_s^j as 1-dimensional path, $S(x)_{a,t}^{i,j}$ is also a 1-dimensional path. Likewise, the collection of all double-iterated integrals can be defined as

$$S(x)_{a,t}^{(2)} = (S(x)_{a,t}^{i,j} : i, j \in \{1, \dots, d\}) \in V^{\otimes 2},$$

where $V^{\otimes 2} := V \times V$. Similarly, we can define *triple-iterated* integral for any combination of $i, j, k \in \{1, \dots d\}$ by double-iterated integral (1.1.2)

$$S(x)_{a,t}^{i,j,k} = \int_{a<s<t} S(x)_{a,s}^{i,j} dx_s^k = \int_{a<w<r<s<t} dx_w^i dx_r^j dx_s^k,$$

which can also be viewed as a 1-dimensional path integral. We also collect all triple-iterated integrals

$$S(x)_{a,t}^{(3)} = (S(x)_{a,t}^{i,j,k} : i, j, k \in \{1, \dots, d\}) \in V^{\otimes 3}.$$

In this case, we can generalise it to *n-fold iterated integral* as below. For any combination of $i_1, \dots, i_n \in \{1, \dots d\}$,

$$S(x)_{a,t}^{i_1,\dots,i_n} = \int_{a<t_n<t} S(x)_{a,t_n}^{i_1,\dots,i_{n-1}} dx_{t_n}^{i_n} = \int_{a<t_1<\cdots<t_n<t} dx_{t_1}^{i_1} \dots dx_{t_n}^{i_n}.$$

Inductively, one can see that the n-fold iterated integral is also a 1-dimensional real-valued path. Collection of all n-fold iterated integrals gives

$$S(x)_{a,t}^{(n)} = (S(x)_{a,t}^{i_1,\dots,i_n} : i_1, \dots, i_n \in \{1, \dots, d\}) \in V^{\otimes n}. \tag{1.1.3}$$

Definition 1.1.2 (signature)**.** Let a path $x \in C_p(V)$ for $1 \leq p < 2$. The *signature* $S(x)_{a,b}$ of x is the infinite collection of all iterated integrals of x on $[a, b]$

$$S(x)_{a,b} = (1, S(x)_{a,b}^{(1)}, S(x)_{a,b}^{(2)}, S(x)_{a,b}^{(3)}, \dots, S(x)_{a,b}^{(n)}, \dots) \in \prod_{n=0}^{\infty} V^{\otimes n}, \tag{1.1.4}$$

where $S(x)_{a,b}^{(0)} = 1 \in V^{\otimes 0}$ and $V^{\otimes 0} := \mathbb{R}$ by convention and $S(x)_{a,b}^{(n)}$ defined in (1.1.3) is usually called as the *n-th level* signature.

In other words, the signature is an infinite and real-valued sequence containing information about all folds of iterated integrals. In the real world, we will only get a finite resolution of signatures. Then we say the signature up to the n-th level is *n-step truncated signature* $S_n(x)_{a,b}$, denoted by

$$S_n(x)_{a,b} = (1, S(x)_{a,b}^{(1)}, S(x)_{a,b}^{(2)}, S(x)_{a,b}^{(3)}, \dots, S(x)_{a,b}^{(n)}, 0, 0, \cdots) \in \prod_{k=0}^{\infty} V^{\otimes k}.$$

Now we will demonstrate the definition of the signature by following two examples.

Example 1.1.1 (Signature for 1-dimensional paths[5]). *Consider an arbitrary 1-dimensional path $x \in C_p(\mathbb{R})$ for $1 \leq p < 2$ and let $x = x_t$, the signature of the path is*

$$S(x)_{a,b} = (1, S(x)^1_{a,b}, S(x)^{1,1}_{a,b}, S(x)^{1,1,1}_{a,b}, \dots),$$

where

$$S(x)^1_{a,b} = \int_{a<s<b} dx_s = x_b - x_a$$

$$S(x)^{1,1}_{a,b} = \int_{a<r<s<b} dx_r dx_s = \int_{a<s<b} (x_s - x_a) dx_s = \frac{1}{2}x_s^2 - x_a x_s \Big|_{x_a}^{x_b} = \frac{(x_b - x_a)^2}{2!}$$

$$S(x)^{1,1,1}_{a,b} = \int_{a<w<r<s<b} dx_w dx_r dx_s = \int_{a<s<b} (\frac{1}{2}x_s^2 - x_a x_s) dx_s = \frac{1}{6}x_s^3 - \frac{1}{2}x_a x_s^2 \Big|_{x_a}^{x_b} = \frac{(x_b - x_a)^3}{3!}.$$

With a bit more algebraic work, one can show that

$$S(x)^{\overbrace{1,1,\dots,1}^{n}} = \frac{(x_b - x_a)^n}{n!}.$$

It is an interesting result, as we can conclude that the information stored in the signature of 1-dimensional paths is no more than the increment of paths $x_b - x_a$.

Example 1.1.2 (Geometric interpretation of a 2-dimensional path). *Consider a path $\hat{x} \in C_1([0,9], \mathbb{R}^2)$, where $\hat{x} = (x_t^1, x_t^2) = (t, x(t))$. Here $x(t)$ is defined as*

$$x_t^2 = x(t) = \begin{cases} \sqrt{3}t & t \in [0,2] \\ 2\sqrt{3} & t \in [2,8] \\ \sqrt{3}t - 6\sqrt{3} & t \in [8,9] \end{cases},$$

which is continuous and piece-wise differentiable. In this case, $\dot{x}_t^1 = 1$ and $\dot{x}_t^2$ can be expressed as

$$\dot{x}_t^2 = \dot{x}(t) = \begin{cases} \sqrt{3} & t \in (0,2) \\ 0 & t \in (2,8) \\ \sqrt{3} & t \in (8,9) \end{cases}.$$

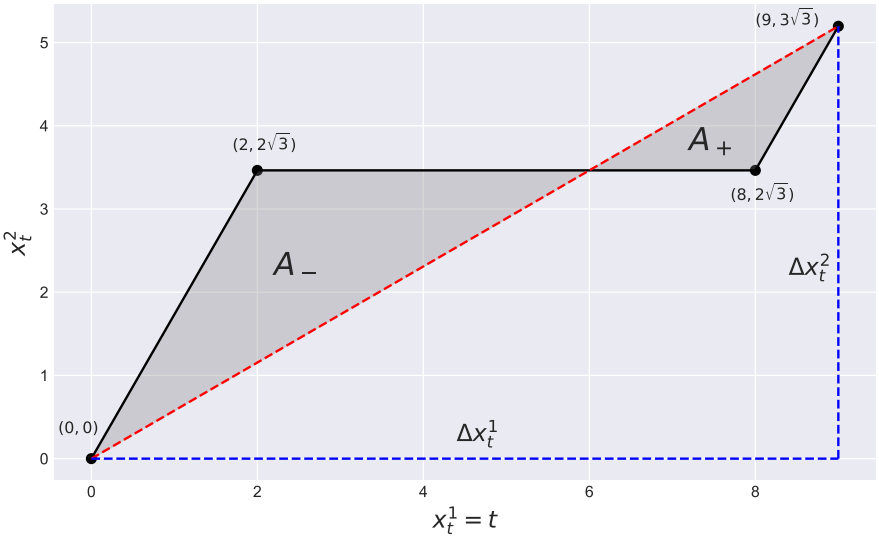

Figure 1.1: Path in Example 1.1.2 shaded by signed Lévy area

Then one can compute the signature of the path

$$\begin{aligned} S(x)_{0,9} &= (1, S(x)^{(1)}_{0,9}, S(x)^{(2)}_{0,9}, S(x)^{(3)}_{0,9}, \dots, S(x)^{(n)}_{0,9}, \dots) \\ &= (1, S(x)^1_{0,9}, S(x)^2_{0,9}, S(x)^{1,2}_{0,9}, S(x)^{2,1}_{0,9}, S(x)^{1,1,1}_{0,9}, \dots), \end{aligned}$$

where

$$
\begin{aligned}
S(x)^1_{0,9} &= \int_{0<s<9} dx^1_s = x^1_9 - x^1_0 = 9 \\
S(x)^2_{0,9} &= \int_{0<s<9} dx^2_s = x^2_9 - x^2_0 = 3\sqrt{3} \\
S(x)^{1,1}_{0,9} &= \int_{0<r<s<9} dx^1_r dx^1_s = \int_{0<s<9} x^1_s dx^1_s = \frac{1}{2}\left(x^1_s\right)^2\Big|_0^9 = \frac{81}{2} \\
S(x)^{1,2}_{0,9} &= \int_{0<r<s<9} dx^1_r dx^2_s = \int_{0<s<9} s dx^2_s = \int_{0<s<9} s\dot{x}^2_s ds = \frac{\sqrt{3}}{2}s^2\Big|_0^2 + \frac{\sqrt{3}}{2}s^2\Big|_8^9 = \frac{21}{2}\sqrt{3} \\
S(x)^{2,1}_{0,9} &= \int_{0<r<s<9} dx^2_r dx^1_s = \int_{0<s<9} x^2_s ds = \frac{\sqrt{3}}{2}s^2\Big|_0^2 + 2\sqrt{3}s\Big|_2^8 + \frac{\sqrt{3}}{2}s^2 - 6\sqrt{3}s\Big|_8^9 = \frac{33}{2}\sqrt{3} \\
S(x)^{2,2}_{0,9} &= \int_{0<r<s<9} dx^2_r dx^2_s = \int_{0<s<9} x^2_s dx^2_s = \frac{1}{2}\left(x^2_s\right)^2\Big|_0^9 = \frac{27}{2}.
\end{aligned}
$$

From Figure 1.1, let A_- and A_+ represent the signed value of the shaded region, signed Lévy area of the path is defined as $A_- + A_+$. In this case, the signed Lévy area is $-3\sqrt{3}$. Surprisingly,

$$
\frac{1}{2}\left(S(x)^{1,2}_{0,9} - S(x)^{2,1}_{0,9}\right) = \frac{1}{2}\left(\frac{21}{2}\sqrt{3} - \frac{33}{2}\sqrt{3}\right) = -3\sqrt{3} = A_- + A_+,
$$

which is exactly the signed Lévy area [5]! From the example, we could then taste a bit of the richness of underlying geometric properties encoded in the signature of a path.

1.1.3 Tensor algebra

Recall that signature (1.1.4) is defined in space $\prod_{n=0}^{\infty} V^{\otimes n}$ based on an infinite collection of iterated integrals. To have the structure of an algebra on the space, we will introduce the algebra of formal power series.

Definition 1.1.3 (Formal power series [5])**.** Let $e_1, \cdots, e_d$ be d formal indeterminates. The algebra of non-commuting formal power series in d indeterminates is the vector space of all series of the form

$$
\sum_{n=0}^{\infty} \sum_{i_1,\ldots,i_n \in \{1,\ldots d\}} \lambda_{i_1,\ldots,i_n} e_{i_1}, \cdots, e_{i_n},
$$

where $\lambda_{i_1,\ldots,i_n} \in \mathbb{R}$ and an element $e_{i_1}, \cdots, e_{i_n}$ is called monomials.

Note that the non-commutativity of power series emphasises that different orders of indeterminates would be distinct monomials; for example, $e_1e_2e_3$, $e_2e_1e_3$ and $e_3e_2e_1$ are all different. To define the space of formal power series as a vector space, one can define addition and scalar multiplication of formal power series by changing the corresponding coefficient of elements. Moreover, by defining the tensor product, the space of formal power series equipped with the product is an algebra [5], denoted by $T((V))$, where tensor algebra is its subalgebra.

Definition 1.1.4 (tensor product)**.** Given monomials $e_{i_1}, \cdots, e_{i_n}$ and $e_{j_1}, \cdots, e_{j_k}$, we define tensor product $\otimes$ between monomials by concatenating monomials with current order

$$
e_{i_1}, \cdots, e_{i_n} \otimes e_{j_1}, \cdots, e_{j_k} = e_{i_1}, \cdots, e_{i_n} e_{j_1}, \cdots, e_{j_k}.
$$

More generally, if we have two formal power series w and v, then

$$
\begin{aligned}
w \otimes v &= \left(\sum_{n=0}^{\infty} \sum_{i_1,\ldots,i_n \in \{1,\ldots d\}} w_{i_1,\ldots,i_n} e_{i_1}, \cdots, e_{i_n}\right) \otimes \left(\sum_{n=0}^{\infty} \sum_{i_1,\ldots,i_n \in \{1,\ldots d\}} v_{i_1,\ldots,i_n} e_{i_1}, \cdots, e_{i_n}\right) \\
&= w_0 v_0 + \sum_{i=1}^{d} (w_0 v_i + w_i v_0) e_i + \sum_{i,j=1}^{d} (w_0 v_{i,j} + w_i v_j + w_{i,j} v_0) e_i e_j + \cdots
\end{aligned}
$$

Interestingly, the indexing strategy of formal power series resembles that of signatures. One can see that the formal power series provides a useful way to express signatures of a path $x \in C_p(V)$ where $1 \leq p < 2$

$$S(x)_{a,b} = \sum_{n=0}^{\infty} \sum_{i_1,\ldots,i_n \in \{1,\ldots d\}} S(x)_{a,b}^{i_1,\ldots,i_n} e_{i_1}, \cdots, e_{i_n}.$$

The representation of the signature as a formal power series is a powerful and convenient tool in mathematical analysis, as we can apply the property of formal power series and simplify the computation of signatures.

Example 1.1.3 (Signatures of linear paths [4, 6])**.** *Suppose there is a linear path $x \in C_1(V)$. Then the path x is linear in terms of t, i.e.*

$$x_t = x_a + \frac{t-a}{b-a}\left(x_b - x_a\right).$$

It follows that its derivative can be written as

$$dx_t = \frac{(x_b - x_a)}{b-a} dt.$$

Then recall the definition of signatures, it holds that

$$\begin{aligned} S(x)_{a,b}^{i_1,\ldots,i_n} &= \int_{a<t_1<\cdots<t_n<b} dx_{t_1}^{i_1} \ldots dx_{t_n}^{i_n} \\ &= \frac{\prod_{k=1}^{n}\left(x_b^{i_k} - x_a^{i_k}\right)}{(b-a)^n} \int_{a<t_1<\cdots<t_n<b} dt_1 \ldots dt_n \\ &= \frac{\prod_{k=1}^{n}\left(x_b^{i_k} - x_a^{i_k}\right)}{(b-a)^n} \frac{(b-a)^n}{n!} \\ &= \frac{\prod_{k=1}^{n}\left(x_b^{i_k} - x_a^{i_k}\right)}{n!}. \end{aligned}$$

Therefore, the n-fold iterated integral thus the whole signatures can be expressed in terms of formal power series as

$$\begin{aligned} S(x)_{a,b}^{(n)} &= \frac{\left(x_b - x_a\right)^{\otimes n}}{n!} \\ S(x)_{a,b} &= \sum_{n=0}^{\infty} \frac{\left(x_b - x_a\right)^{\otimes n}}{n!} := \exp_{\otimes}\left(x_b - x_a\right). \end{aligned}$$

After introducing Chen's identity in Theorem 1.3.2 later, we will demonstrate how it is powerful in computing signatures.

1.2 Analytical properties

Now we will review several important properties of signatures. From those analytical properties, one should see why the signature is a good candidate for representing a path.

1.2.1 Injectivity

The first property shows that signature is characteristic for determining a path. KT Chen [7] first showed that non-commutative iterated integrals are faithful to their irreducible piece-wise regular continuous path. Then, the condition was extended to continuous paths of bounded variation by Hambly and Lyons [8], where the signature determines a path x up to tree-like equivalence. A further extension to less irregular paths can be found in [9], which generalised the tree-like notion to weakly geometric rough paths.

Definition 1.2.1 (Tree-like equivalence [3])**.** For paths $x, y \in C_p(V)$ and $1 \leq p < 2$, $S(x)_{a,b} = S(y)_{a,b}$ if and only if $x \sim_\tau y$, where $\sim_\tau$ denotes tree-like equivalence.

Theorem 1.2.2 (Injectivity of signatures [9]). *For $1 \leq p < 2$, the signature map $S : C_p(V) \to T((V))$ is injective up to tree-like equivalence.*

Informally, tree-like equivalence on the space of paths with continuous bounded variation can be treated as a dimension reduction of paths with the same parameterisation, where the space with reduction is called *unparameterised paths*.

Definition 1.2.3 (Unparameterised paths [10]). For $x \in C_p(V)$ and $1 \leq p < 2$, the $\sim_\tau$ equivalence class of x is denoted by $[x]$. The space of *p-unparameterised paths* $\mathcal{C}_p(V)$ is defined as the quotient space $C_p(V)/\sim_\tau = \{[x] : x \in C_p(V)\}$.

Remark 1.2.4. If a path x has at least one monotone coordinate, $[x]$ collapses to a single point, i.e. $[x] = \{x\}$. Therefore, the signature uniquely determines the path x. One can further conclude that time-augmented paths have unique signatures.

1.2.2 Invariance under time reparameterisations

If a function $\phi : [c, d] \to [a, b]$ is surjective, non-decreasing and continuously differentiable, we define ϕ as a *reparameterisation*. For example, $\phi(x) = x^2$ for $x \in [0, 1]$ is a reparameterisation that is slower at first and then goes quicker than the original parameterisation.

To demonstrate the idea of invariance, Let us restrict paths $x, y : [a, b] \to \mathbb{R}$ to be continuous and differentiable, and define the reparameterisation as $\tilde{x}_t = x_{\phi(t)}, \tilde{y}_t = y_{\phi(t)}$. By the chain rule,

$$\dot{\tilde{x}}_t = \dot{\phi}(t)\dot{x}_{\phi(t)}.$$

The path integral on $\tilde{x}_t$ can be expressed as

$$\int_c^d \tilde{y}_t d\tilde{x}_t = \int_a^b y_{\phi(t)}\dot{\phi}(t)\dot{x}_{\phi(t)}dt = \int_a^b y_u dx_u,$$

which shows that the path integral is invariant under reparameterisation. Since the signature of a d-dimensional path $\hat{x}$ is defined by its coordinate iterated integrals, inductively, one can see that the signature of a path and its reparameterisation is the same, i.e. $S(\hat{x}_t)_{a,b} = S(\hat{x}_{\phi(t)})_{a,b}$. This means that if one traverses the same path but changes the speed or pacing, the signature remains the same. More generally, one can show that the restriction of differentiability is unnecessary. Reparameterisation invariance also holds for $x \in C_p(V)$, where $1 \leq p < 2$ [4]. From the property of invariance under reparameterisation, the signature is not determined by the recording time of each sample but concentrates on the structure and geometry of the path itself.

1.2.3 Factorial Decay

Lastly, we will present a quantitative result of the magnitude of the signature at each level with respect to the length of paths.

Theorem 1.2.5 (Factorial decay [4, 11]). *Consider $x \in C_1(V)$ is a continuous path with signature $S(x)_{a,b}$ and 1-variation $\|x\|_{1,[a,b]}$. Then for every $n \in \mathbb{N}$, we have*

$$\|S^n(x)_{a,b}\|_{V^{\otimes n}} \leq \frac{\|x\|^n_{1,[a,b]}}{n!}. \tag{1.2.1}$$

Theorem 1.2.5 shows terms of signature decay in a factorial way, illustrating the most significant information is contained in the first few levels of signature. The property of factorial decay justifies we can characterise a path approximately by a truncated level of signature, which is widely used in applications.

1.3 Algebraic properties

Algebraic properties make the signature a powerful tool for understanding the geometry and dynamics of paths. In this section, we will briefly introduce the signature's ability to encode and manipulate the composition, order, and complex interactions of paths.

1.3.1 Chen's identity

Chen's identity is one of the most fundamental algebraic properties of signature as it describes the behaviour of signature under the concatenation of paths.

Definition 1.3.1 (Concatenation)**.** Consider paths $x \in C_p([a,b],V)$ and $y \in C_p([b,c],V)$ for $1 \le p < 2$, we define the *concatenation* of x and y, denoted by $x * y$ as a path $[a,c] \to V$

$$(x*y)_t := \begin{cases} x_t & \text{if } a \le t \le b \\ x_b - y_b + y_t & \text{if } b \le t \le c \end{cases}.$$

Now we will present Chen's identity, which intuitively set an equivalence relation between the concatenation of paths and the tensor product of signatures.

Theorem 1.3.2 (Chen's identity)**.** *Consider paths $x \in C_p([a,b],V)$ and $y \in C_p([b,c],V)$ for $1 \le p < 2$. Then*

$$S(x*y)_{a,c} = S(x)_{a,b} \otimes S(y)_{b,c}.$$

The theorem provides a method to simplify the analysis of longer paths by converting them into manageable shorter ones. If we have a path $x : [t_0, t_n] \to \mathbb{R}^d$ under regularity conditions, then inductively, we can decompose the signature of x to

$$S(x)_{t_0,t_n} = S(x)_{t_0,t_1} \otimes S(x)_{t_1,t_2} \otimes \cdots \otimes S(x)_{t_{n-1},t_n}.$$

Moreover, if we have a sequence of data $(t_i, x_i) \in \mathbb{R}^{d+1}$, we can treat x as a piecewise linear path interpolating the data [4, 6]. Based on Example 1.1.3, one can observe that

$$S(x)_{t_0,t_n} = \exp_\otimes (x_{t_1} - x_{t_0}) \otimes \exp_\otimes (x_{t_2} - x_{t_1}) \cdots \exp_\otimes \left(x_{t_n} - x_{t_{n-1}}\right),$$

which is widely used in Python packages such as `ESig` [12]. In this case, n-step truncated signatures can be computed by the tensor product up to monomials with order n.

1.3.2 Shuffle product

Apart from the tensor product, we can also equip space of formal power series with shuffle product to form Lie algebra [5, 10]. Given a word $I = (i_1, \cdots, i_n) \in \{1, \cdots, d\}^n$, informally, it can be viewed as the index of an element of signatures.

$$S_{a,b}(x)^I := S_{a,b}(x)^{i_1,\cdots,i_n}.$$

Recall signatures of a path are defined by iterated integrals. We can treat I as a linear functional on $T((V))^*$, where $T((V))^*$ is the dual space of $T((V))$.

$$S_{a,b}(x)^I = (I, S_{a,b}(x)). \tag{1.3.1}$$

With the basic setup above, the definition of the shuffle product will be introduced below.

Definition 1.3.3 (shuffle product [6])**.** Consider two words $I = (i_1, \cdots, i_n)$ and $J = (j_1, \cdots j_p)$. the *shuffle product* $\sqcup\!\sqcup$ is defined by

$$I \sqcup\!\sqcup J := \sum_{\sigma \in \mathrm{Shuff}(n,p)} \left(a_{\sigma^{-1}(1)}, \cdots, a_{\sigma^{-1}(n+p)}\right),$$

where $(a_1, \cdots, a_{n+p}) := (i_1, \cdots, i_n, j_1, \cdots j_p)$, and S_n contains all permutation of n elements,

$$\mathrm{Shuff}(n,p) := \{\sigma \in S_{n+p} : \sigma(1) < \cdots < \sigma(n) \text{ and } \sigma(n+1) < \cdots < \sigma(n+p)\}.$$

Intuitively, one can think of it as shuffling two decks of cards together. Each deck represents a word. When shuffling them together, we effectively combine the information from each word in every possible order - capturing the higher-level details. From the definition, one can observe that the shuffle product is commutative. Equipped with the shuffle product, the shuffle combination of higher-level terms of signatures can be seen expressed by their products of lower-level terms, which is called the shuffle identity.

Theorem 1.3.4 (Shuffle identity [11]). *Consider a path $x \in C_p(V)$ for $1 \le p < 2$. For any word $I = (i_1, \cdots, i_n) \in \{1, \cdots, d\}^n$ and $J = (j_1, \cdots j_p) \in \{1, \cdots, d\}^p$,*

$$(I, S(x)_{a,b})\,(J, S(x)_{a,b}) = (I \sqcup\!\sqcup J, S(x)_{a,b})\,.$$

Example 1.3.1. *Suppose a differentiable path $x = (x_t^1, x_t^2) \in C_p(\mathbb{R}^2)$ and words $I = \mathbf{1}, J = \mathbf{2}$.*

$$\begin{aligned} S(x)^1_{a,b} S(x)^2_{a,b} &= \int_{a<t<b} dx_t^1 \int_{a<t<b} dx_t^2 \\ &= \int_{a<t<b} \dot{x}_t^1 dt \int_{a<t<b} \dot{x}_t^2 dt \\ &\overset{\text{by parts}}{=} \int_{a<t<b} S(x)^2_{a,b} \dot{x}_t^1 dt + \int_{a<t<b} S(x)^1_{a,b} \dot{x}_t^2 dt \\ &= S(x)^{2,1}_{a,b} + S(x)^{1,2}_{a,b}. \end{aligned}$$

By the shuffle identity, we have

$$\begin{aligned} S(x)^1_{a,b} S(x)^2_{a,b} &= (\mathbf{1}, S(x)_{a,b})\,(\mathbf{2}, S(x)_{a,b}) \\ &= (\mathbf{1} \sqcup\!\sqcup \mathbf{2}, S(x)_{a,b}) \\ &= (\mathbf{12} + \mathbf{21}, S(x)_{a,b}) \\ &= S(x)^{1,2}_{a,b} + S(x)^{2,1}_{a,b}, \end{aligned}$$

which is exactly the same as what we derived via integration by parts! Indeed, shuffle products over iterated integrals can be proved by this useful calculus tool [4].

Example 1.3.2. *Consider a path $x = (x_t^1, x_t^2, x_t^3) \in C_1(\mathbb{R}^3)$ and words $I = \mathbf{321}, J = \mathbf{1}$. By the shuffle identity, we have*

$$\begin{aligned} S(x)^{3,2,1}_{a,b} S(x)^1_{a,b} &= (\mathbf{321}, S(x)_{a,b})\,(\mathbf{1}, S(x)_{a,b}) \\ &= (\mathbf{321} \sqcup\!\sqcup \mathbf{1}, S(x)_{a,b}) \\ &= (\mathbf{1321} + \mathbf{3121} + 2\mathbf{3211}, S(x)_{a,b}) \\ &= S(x)^{1,3,2,1}_{a,b} + S(x)^{3,1,2,1}_{a,b} + 2S(x)^{3,2,1,1}_{a,b}. \end{aligned}$$

1.3.3 Time-reversal

Time-reversal property of signatures explains the inverse relation of signatures under tensor product $\otimes$ between a path and its time-reversal path.

Definition 1.3.5 (time-reversal). Consider a path $x \in C_p([a,b], V)$, the *time-reversal* path is defined by $\overleftarrow{x} \in C_p([a,b], V)$ such that $\overleftarrow{x}_t = x_{a+b-t}$ for all $t \in [a,b]$.

Intuitively, one can think of a time-reversal path as the path flipping the start and end points.

Theorem 1.3.6 (Time-reversed signatures). *Consider a path $x \in C_p(V)$ for $1 \le p < 2$ and its time-reversal path $\overleftarrow{x}$, then*

$$S(x)_{a,b} \otimes S(\overleftarrow{x})_{a,b} = 1,$$

which is the identity of formal power series, where all monomial terms have zero coefficients.

Recall Chen's identity in Theorem 1.3.2, one can intuitively treat the tensor product between signatures as a concatenation of path and its reverse. In this case, the concatenation result could be viewed as a path that is not moving. The time-reversal property is important as it provides insight for studying the reverse pattern and its relationship with the original path.

1.3.4 Universality

The final important property emphasises that signature linearises continuous functions of an unparameterised path x ([3, 10, 13, 14]).

Theorem 1.3.7 (Universality)**.** *Let $\mathcal{C}_p(V)$ denotes the space of unparameterised paths and $1 \leq p < 2$. Consider a compact set $\mathcal{X} \subset \mathcal{C}_p(V)$ and an arbitrary continuous function $f : \mathcal{X} \to \mathbb{R}$. Then for any $\epsilon > 0$, there exists a truncation level $n \in \mathbb{N}$ and $\alpha_J \in \mathbb{R}$,*

$$\sup_{x \in \mathcal{X}} \left| f(x) - \sum_{k=0}^{n} \sum_{J \in \{1,\cdots,d\}^k} \alpha_J S(x)^J \right| < \epsilon.$$

Sketch proof [4]. Let $C(\mathcal{X})$ be the space of continuous functions over $\mathcal{X}$ with sup-norm topology. Consider the subset of $C(\mathcal{X})$

$$\begin{aligned} \mathcal{A}_{\mathcal{X}} &:= \{[x] \to (I, S(x)) : x \in \mathcal{X}, I \in T((V))^*\} \\ &:= \{\Phi_I : I \in T((V))^*\}. \end{aligned}$$

The strategy of the sketch proof is first showing that $\mathcal{A}_{\mathcal{X}}$ is a subalgebra of $C(\mathcal{X})$ with certain properties, and then proving that $\mathcal{A}_{\mathcal{X}}$ is dense by Stone-Weierstrass theorem.

For $I, J \in T((V))^*$, $\Phi_I, \Phi_J \in \mathcal{A}_{\mathcal{X}}$. Let $\lambda \in \mathbb{R}$ be a scalar, by definition of $\mathcal{A}_{\mathcal{X}}$, we have that $\Phi_I + \Phi_J = \Phi_{I+J}$ and $\Phi_{\lambda I} = \lambda \Phi_I$. From shuffle identity in 1.3.4, $\Phi_I \Phi_J = \Phi_{I \sqcup\!\sqcup J} \in \mathcal{A}_{\mathcal{X}}$, thus $\mathcal{A}_{\mathcal{X}}$ is closed under the pointwise product. Let i be the linear functional such that $(i, S(x)) = 1$ for extracting the 0-th level of signature. Then $\Phi_{\lambda i} \equiv \lambda$ and $\mathcal{A}_{\mathcal{X}}$ contains all constant functions. As $(I, S(x)) = (I, S(y))$ if and only if $S(x) = S(y)$, followed by injectivity of S up to tree-like equivalence in 1.2.2, $\mathcal{A}_{\mathcal{X}}$ separates points in $\mathcal{X} \subset \mathcal{C}_p(V)$.

Now we have shown that $\mathcal{A}_{\mathcal{X}}$ is a subalgebra of $C(\mathcal{X})$ that contains constant functions and separates points. Applying the Stone-Weierstrass theorem, $\mathcal{A}_{\mathcal{X}}$ is dense, which implies the universality property. □

Theorem 1.3.7 shows that the linear combination of elements in the signature of paths can approximate continuous mappings of the paths in arbitrary precision uniformly. This is a fundamental result for inversion, as we shall see later that universality guarantees the existence of linear functionals in the dual space $T((V))^*$.

1.4 Applications of signatures

The properties summarised above make signatures effective and efficient as features of the path. In recent works, signatures are widely used as a tool for feature extraction, where the routine [5] is normally

$$\text{data} \to \text{path} \to \text{signatures} \to \text{features}$$

In this section, we aim to provide a very brief overview of its application for interested readers.

Machine learning is one of the main areas of applications, where features found from signatures can be learned to summarise and generalise into new cases. In particular, the reparametrization invariance of the signature is beneficial in machine learning tasks such as time series classification and recognition tasks involving sequences, as it means features from signatures do not depend on the specific timing of the observations in the time series. Take character recognition as an example, combined with a convolutional neural network, the signatures method improved the accuracy of Chinese character recognition [15], where valuable information was found in the first three levels of signatures. A similar pattern is also shown in financial data [16], which demonstrates the sufficiency of low levels of truncated signatures to make useful predictions. To model continuous-time time series, one can compute neural controlled differential equations and rough differential equations by learning from log signatures [17, 18]. Moreover, signatures can be also applied to the field of bioinformatics for identifying disorders [5, 19, 20] and diseases in early stages [6, 21, 22].

In summary, the signature-based learning method has permeated numerous fields with its unique and powerful approach to understanding sequential data. Despite the breadth of its current applications, the full potential of the signature method remains to be fully explored and understood.

Chapter 2

Compositions of paths in bases

The universality of signatures in the last chapter of Theorem 1.3.7 provides a rich playground for finding linear functionals that can direct to other characteristics of a path. For this purpose, we will introduce orthogonal polynomials and the Fourier series.

2.1 Orthogonal polynomials

2.1.1 Inner product and orthogonality

Consider a dot product $\langle \boldsymbol{x}, \boldsymbol{y} \rangle := \sum_{i=1}^{n} x_i y_i$, where $\boldsymbol{x}, \boldsymbol{y} \in \mathbb{R}^n$. If weights $w_1, \cdots, w_n \in \mathbb{R}_+$ are defined, $\langle \boldsymbol{x}, \boldsymbol{y} \rangle_w := \sum_{i=1}^{n} w_i x_i y_i$ is measured as a weighted dot product, where $\langle \cdot, \cdot \rangle_w$ can be written as $\langle \cdot, \cdot \rangle$ for simplicity. In general, we can extend the product for vector space $\mathbb{V}$ over F.

Definition 2.1.1 (inner product and inner product space)**.** Let $F = \mathbb{R}$ or $\mathbb{C}$ and let $\mathbb{V}$ be a vector space over F. An *inner product* $\langle u, v \rangle$ on $\mathbb{V}$ is a map $\mathbb{V} \times \mathbb{V} \to F$ satisfying

(1) $\langle \lambda_1 v_1 + \lambda_2 v_2, u \rangle = \lambda_1 \langle v_1, u \rangle + \lambda_2 \langle v_2, u \rangle$,

(2) $\langle u, v \rangle = \overline{\langle v, u \rangle}$,

(3) $\langle v, v \rangle > 0$ if $v \neq 0$,

for all $v_i, v, u \in \mathbb{V}$ and $\lambda_i \in F$. Then the *inner product space* is defined as the vector space $\mathbb{V}$ with an inner product $\langle \cdot, \cdot \rangle$.

For example, let $d\alpha$ be a non-negative Borel measure supported in the interval $[a, b]$ and $\mathbb{V} = L^2_w(a, b)$. For $p \in [1, \infty)$, $L^p_w(\Omega)$ is the linear space of measurable functions from Ω to $\mathbb{R}$ such that their weighted p-norms are bounded, i.e.

$$L^2_w(\Omega) = \left\{ v \text{ is measurable in } \Omega \middle| \int_\Omega |v(t)|^2 w(t) dt < \infty \right\}.$$

One can define $\langle f, g \rangle = \int_a^b f(t) g(t) d\alpha(t)$ as a Stieltjes integral for all $f, g \in \mathbb{V}$, which can be proved as an inner product by the three properties in the Definition 2.1.1. Note that if $\alpha(t)$ is absolutely continuous, which will be the setting throughout the report, then one can find $d\alpha(t) = w(t)dt$ [23]. In this case, the definition of inner product over function space reduces to an integral with respect to a weight function, i.e.

$$\langle f, g \rangle = \int_a^b f(t) g(t) w(t) dt.$$

We can then refer an orthogonal polynomial system to orthogonal with respect to the *weight* function w. From now on, we will stick to $F = \mathbb{R}$ and $\mathbb{V} = L^2_w(\Omega)$. $\mathbb{P}[x] \subset L^2_w(\Omega)$ denoted by the space of all polynomials. For a polynomial $p \in \mathbb{P}_n[t]$ with degree n, it is *monic* if the coefficient of the n-th degree is one.

Definition 2.1.2 (Orthogonality and orthogonal polynomials)**.** For an arbitrary vector space $\mathbb{V}$, u and v are *orthogonal* if $\langle u, v \rangle = 0$ with all $u, v \in \mathbb{V}$. When $\mathbb{V} = \mathbb{P}[t]$, a sequence of polynomials

(or *orthogonal polynomial system*) $(p_n)_{n\in\mathbb{N}} \in \mathbb{P}[t]$ is called orthogonal polynomials with respect to a weight w if for all $m \neq n$,

$$\langle p_n, p_m \rangle = \int p_n(t) p_m(t) w(t) dt = 0,$$

where $\deg(p_n) = n$ is the degree of polynomial. Furthermore, we say the sequence of orthogonal polynomials is *orthonormal* if $\langle p_n, p_n \rangle = 1$ for all $n \in \mathbb{N}$.

For simplification, inner product notation $\langle \cdot, \cdot \rangle$ will be used without specifying the integral formulation for orthogonal polynomial. To construct a sequence of orthogonal polynomials in Definition 2.1.2, one can follow the Gram-Schmidt orthogonalisation process, which is stated and proved below.

Theorem 2.1.3 (Gram-Schmidt orthogonalisation process (Thm. 2.1 in [24]))**.** *The polynomial system $(p_n)_{n\in\mathbb{N}}$ can be constructed recursively by*

$$p_0 = 1, \qquad p_n = t^n - \sum_{i=1}^{n-1} \frac{\langle t^n, p_i \rangle}{\langle p_i, p_i \rangle} p_i \quad \text{for } n \geq 1 \tag{2.1.1}$$

is a monic orthogonal polynomial system with respect to the inner product $\langle \cdot, \cdot \rangle$.

Proof. We will prove it by induction. Given $F = \mathbb{R}$, the inner product is symmetric, then it is sufficient to prove

$$\langle p_n, p_m \rangle = 0, \quad \forall n \geq 1, 0 \leq m \leq n-1. \tag{2.1.2}$$

Now consider the base case $n = 1$, the relation from (2.1.2) gives

$$\langle p_1, p_0 \rangle = \left\langle t - \frac{\langle t, p_0 \rangle}{\langle p_0, p_0 \rangle} p_0, p_0 \right\rangle = \langle t, p_0 \rangle - \frac{\langle t, p_0 \rangle}{\langle p_0, p_0 \rangle} \langle p_0, p_0 \rangle = 0,$$

which shows p_0 and p_1 are orthogonal. In the inductive step, let us assume the relation from (2.1.2) holds true up to $n \geq 1$. In the case of p_{n+1}, for $0 \leq m \leq n$,

$$\begin{aligned}
\langle p_{n+1}, p_m \rangle &= \left\langle t^{n+1} - \sum_{i=1}^{n} \frac{\langle t^{n+1}, p_i \rangle}{\langle p_i, p_i \rangle} p_i, p_m \right\rangle \\
&= \langle t^{n+1}, p_m \rangle - \sum_{i=1}^{n} \frac{\langle t^{n+1}, p_i \rangle}{\langle p_i, p_i \rangle} \langle p_i, p_m \rangle \\
&= \langle t^{n+1}, p_m \rangle - \frac{\langle t^{n+1}, p_m \rangle}{\langle p_m, p_m \rangle} \langle p_m, p_m \rangle \\
&= 0,
\end{aligned}$$

where we have used the properties of the inner product and orthogonality of p_m and p_i for $0 \leq i \leq n$ and $i \neq m$. □

From the orthogonalisation process in Theorem 2.1.3, we can see that the n-th polynomial p_n has degree n exactly, which means $(p_n)_{n\in\mathbb{N}}$ is a basis spanning $\mathbb{P}[t]$. Furthermore, the orthogonal construction makes the orthogonal polynomial system an orthogonal basis with respect to the corresponding inner product. The following proposition forms an explicit expression of coefficients of $(p_k)_{k\in\{0,\cdots,n\}}$ for an arbitrary n-th degree polynomials.

Proposition 2.1.4 (Orthogonal polynomials expansion)**.** *Consider an arbitrary polynomial $x(t) \in \mathbb{P}_n[t]$, then one can express $x(t)$ by a sequence of orthogonal polynomials $(p_k)_{k\in\{0,\cdots,n\}}$, i.e.*

$$x(t) = \sum_{k=0}^{n} \frac{\langle p_k, x \rangle}{\langle p_k, p_k \rangle} p_k(t).$$

Proof. Since $(p_k)_{k\in\{0,\cdots,n\}}$ is a basis of $\mathbb{P}_n[t]$, $x(t)$ can be written as a linear combination of the sequence of polynomial

$$x(t) = \sum_{k=0}^{n} \alpha_k^x p_k(t).$$

By applying the inner product with p_i on both sides, we get

$$\langle p_i, x\rangle = \left\langle p_i, \sum_{k=0}^{n} \alpha_k^x p_k \right\rangle = \sum_{k=0}^{n} \alpha_k^x \langle p_i, p_k\rangle = \alpha_i^x \langle p_i, p_i\rangle,$$

where we used orthogonality of p_i and p_k for $i \neq k$. Therefore, the k-th coefficient is

$$\alpha_k^x = \frac{\langle p_k, x\rangle}{\langle p_k, p_k\rangle} = \frac{1}{\langle p_k, p_k\rangle} \int x(t) p_k(t) w(t) dt. \tag{2.1.3}$$

□

Remark 2.1.5. We have proved the orthogonal polynomial expansion for $x \in \mathbb{P}_n[t]$. In general, by the closure of orthogonal polynomials system in $L_w^2(a,b)$ [23], arbitrary $f \in L_w^2(a,b)$ can be written as an infinite sequence of orthogonal polynomials.

$$f(t) \sim \sum_{k=0}^{\infty} \frac{\langle p_k, f\rangle}{\langle p_k, p_k\rangle} p_k(t).$$

The N-th degree approximation of f is the best approximating polynomial [25] with degree less or equal to N, denoted by

$$P_N f(t) = \sum_{k=0}^{N} \frac{\langle p_k, f\rangle}{\langle p_k, p_k\rangle} p_k(t). \tag{2.1.4}$$

Linear functional on orthogonal polynomials

Recall *linear functional* on a vector space $\mathbb{V}$ is a linear map $\mathcal{L} : \mathbb{V} \to F$ such that

$$\mathcal{L}(\alpha_1 v_1 + \alpha_2 v_2) = \alpha_1 \mathcal{L}(v_1) + \alpha_2 \mathcal{L}(v_2).$$

In the setting of orthogonal polynomials, the process of extracting the k-th coefficient can be defined as $\mathcal{L}_k(x) = \alpha_k^x$ for $x \in L_w^2$.

Proposition 2.1.6. *$\mathcal{L}_k(x) = \alpha_k^x$ is a linear functional with respect to a given orthogonal polynomials system $(p_n)_{n\in\mathbb{N}}$.*

Proof. Choose $x_1(t), x_2(t) \in L_w^2$ and $c_1, c_2 \in \mathbb{R}$ in arbitrary. By the coefficient of orthogonal polynomials in (2.1.3),

$$\mathcal{L}_k(c_1 x_1 + c_2 x_2) = \frac{\langle p_k, c_1 x_1 + c_2 x_2\rangle}{\langle p_k, p_k\rangle} = \frac{c_1\langle p_k, x_1\rangle + c_2\langle p_k, x_2\rangle}{\langle p_k, p_k\rangle} = c_1 \mathcal{L}(x_1) + c_2 \mathcal{L}(x_2),$$

which is followed by the Definition 2.1.1 of the inner product. □

2.1.2 Basic properties

There are many important properties of orthogonal polynomials, such as three terms recurrence relation, Christoffel–Darboux formula and zeros, which are widely used in fields like approximation and differential equations [23]. Here we will list two main properties that are significant for our application purpose.

The three-term recurrence relation

Theorem 2.1.7 (Three-term recurrence relation)**.** *An orthogonal polynomials system $(p_n)_{n\in\mathbb{N}}$ with respect to a measure $d\alpha$ satisfies the three-term recurrence relation.*

$$p_0(t) = 1, \quad p_1(t) = A_1 t + B_1, \quad p_{n+1}(t) = (A_{n+1}t + B_{n+1}) p_n(t) + C_{n+1} p_{n-1}(t),$$

for all $n \in \mathbb{N}$, and $A_i > 0$ for all $i \in \mathbb{N}_0$.

Before proving the recurrence relation, we will first show that an orthogonal polynomial is orthogonal to all polynomials with a degree lower than that of itself.

Lemma 2.1.8. *A polynomial $q(t) \in \mathbb{P}_n[t]$ satisfying $\langle q, r\rangle = 0$ for all $r(t) \in \mathbb{P}_m[t]$ with $m < n$ if and only if $q(t) = p_n(t)$ up to some constant coefficient, where $p_n(t)$ denotes the orthogonal polynomial with degree n.*

Proof. $\Longrightarrow$: Consider $q(t) = \alpha_n t^n + O(t^{n-1})$ and $p_n(t) = \tilde{\alpha}_n t^n + O(t^{n-1})$. Then we define

$$s(t) = q(t) - \frac{\alpha_n}{\tilde{\alpha}_n} p_n(t) = O(t^{n-1}),$$

which has a degree at most $n-1$. Therefore, for all $m < n$,

$$\langle s, p_m\rangle = \langle q, p_m\rangle - \frac{\alpha_n}{\tilde{\alpha}_n}\langle p_n, p_m\rangle = 0.$$

The former inner product $\langle q, p_m\rangle = 0$ by assumption, while the latter inner product $\langle p_n, p_m\rangle = 0$ by orthogonality. By Proposition 2.1.4,

$$s(t) = \sum_{m=0}^{n-1} \frac{\langle p_m, s\rangle}{\langle p_m, p_m\rangle} p_m(t) = 0 \quad \Longrightarrow \quad q(t) = \frac{\tilde{\alpha}_n}{\alpha_n} p_n(t).$$

$\Longleftarrow$: Consider $r(t) = \sum_{k=0}^m r_k p_k(t)$. Let $q(t) = cp_n(t)$ by assumption, then for all $m < n$,

$$\langle q, r\rangle = \left\langle cp_n(t), \sum_{k=0}^m r_k p_k(t)\right\rangle = c\sum_{k=0}^m r_k \langle p_n(t), p_k(t)\rangle = 0,$$

where we used the linearity of the inner product and orthogonality of $(p_n)_{n\in\mathbb{N}}$. □

Now we have enough tools to prove the famous three-term recurrence relation.

Proof of Theorem 2.1.7. Consider a sequence of orthogonal polynomials $(p_n)_{n\in\mathbb{N}}$. When $n = 1$, p_1 can be expressed as $A_1 t + B_1$ for $A_1, B_1 \in \mathbb{R}$. This is because p_1 is an element in an orthogonal basis with degree 1. Based on the inner product of orthogonal polynomials,

$$\langle p_k, tp_n\rangle = \int tp_k(t)p_n(t)w(t)dt = \langle tp_k, p_n\rangle.$$

Therefore, for $0 \le k < n-1$, we have $\langle p_k, tp_n\rangle = 0$ by Lemma 2.1.8. Since $tp_n(t)$ has degree $n+1$, by Proposition 2.1.4,

$$tp_n(t) = \sum_{k=0}^{n+1} \frac{\langle p_k, tp_n\rangle}{\langle p_k, p_k\rangle} p_k(t) = \sum_{k=n-1}^{n+1} \frac{\langle p_k, tp_n\rangle}{\langle p_k, p_k\rangle} p_k(t) = \alpha_{n-1}p_{n-1}(t) + \alpha_n p_n(t) + \alpha_{n+1}p_{n+1}(t)$$

$$\Longrightarrow\ p_{n+1} = \left(\frac{1}{\alpha_{n+1}}t - \frac{\alpha_n}{\alpha_{n+1}}\right)p_n(t) - \frac{\alpha_{n-1}}{\alpha_{n+1}}p_{n-1}(t),$$

which completes the proof. □

Remark 2.1.9. Among all other properties, recurrence is the core property of orthogonal polynomials in our setting, as one can find higher-order coefficients based on lower-order coefficients given the analytical form of the orthogonal polynomials. This idea coincides with the shuffle identity of signatures in 1.3.4. As we shall see later, one can construct an explicit recurrence relation for coefficients of orthogonal polynomials by linear functionals acting on signatures.

Approximation results for functions in L^2_w

Without loss of generality, consider $f \in L^2_w(-1, 1)$, as we can always transform an arbitrary interval $[a, b]$ linearly into the interval $[-1, 1]$. Recall the N-th degree approximation $P_N f(t)$ defined in (2.1.4). The uniform convergence of the N-th degree approximation $P_N f(t)$ to f can be found in [25], where we obtain

$$\frac{1}{\sqrt{2\pi}}\|f - P_N f\|_2 \le \|f - P_N f\|_\infty \le (1 + \|P_N\|)\|f - q\|_\infty, \qquad q \in \mathbb{P}_N,$$

where $\|P_N\|$ relates to the system of orthogonal polynomials, and $\|f-q\|_\infty$ depends on the smoothness k of f. In the case of Chebyshev polynomials, where the weight function is $w(t) = 1/\sqrt{1-t^2}$, $\|P_N\| = \frac{4}{\pi}\log n + \mathcal{O}(1)$ [25]. For some $\alpha \in (0, 1]$,

$$\|f - P_N f\|_2 \leq c_k \frac{\log N}{N^{k+\alpha}} \qquad \text{for } N \geq 2.$$

The bound result is shown numerically in Figure 2.1. The orthogonal polynomials closely resemble the original paths x_1 and x_2, which demonstrates good convergence results. On the other hand, the orthogonal polynomials generalise the general trend of x_3 but fail to approximate the non-differentiable points.

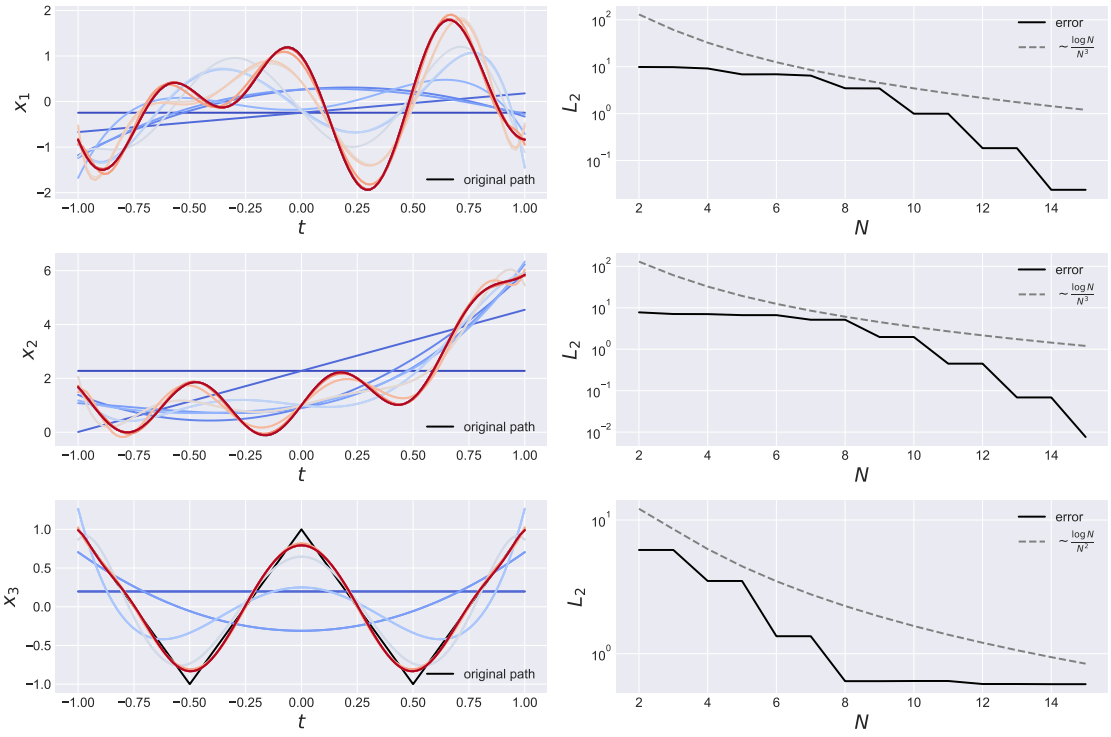

Figure 2.1: Approximation (left) and convergence of L_2 error (right) results for Chebyshev polynomials by increasing degree N, where their colours change from blue to red in the left plots; paths are given by (from top to bottom): $x_1(t) = \cos(10t) - \sin(2\pi t)$, $x_2(t) = \sin(10t) + e^{2t} - t$, $x_3(t) = 2|2t-1| - 1$

2.1.3 Examples

In this subsection, we will provide two general orthogonal polynomials, namely Jacobi polynomials and Hermite polynomials, which will be set as a comparison in the next chapter. Figure 2.2 visualises one case for each of the first few polynomials of these two kinds, where one can see that different weight functions would lead to different behaviours of orthogonal polynomials.

Jacobi polynomials

Jacobi polynomials $p_n^{(\alpha,\beta)}$ are a system of orthogonal polynomials with respect to the weight function $w : (-1, 1) \to \mathbb{R}$ such that

$$w(t; \alpha, \beta) := (1-t)^\alpha (1+t)^\beta.$$

There are many well-known special cases of Jacobi polynomials, such as Legendre polynomials $p_n^{(0,0)}$ and Chebyshev polynomials $p_n^{(-1/2,-1/2)}$. In general, the analytical expression of Jacobi

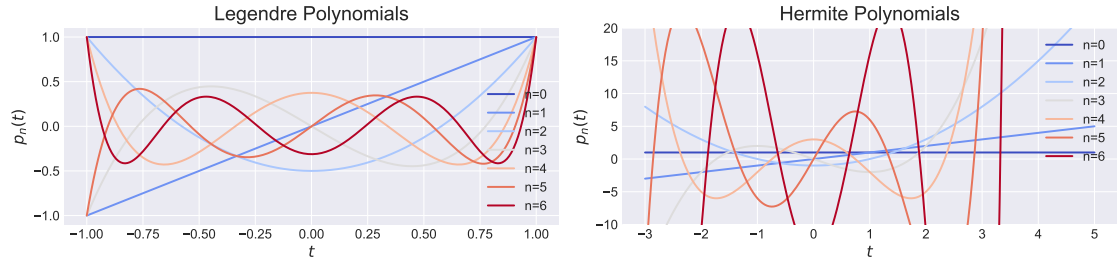

Figure 2.2: Visualisation of first 7 Legendre and Hermite polynomials

polynomials [26] is defined by the hypergeometric function ${}_2F_1$:

$$p_n^{(\alpha,\beta)}(t) = \frac{(\alpha+1)_n}{n!}{}_2F_1(-n, 1+\alpha+\beta+n; \alpha+1; \frac{1}{2}(1-t)),$$

where $(\alpha+1)_n$ is Pochhammer's symbol. For orthogonality, Jacobi polynomials satisfy

$$\int_{-1}^{1}(1-t)^\alpha(1+t)^\beta p_m^{(\alpha,\beta)}(t)p_n^{(\alpha,\beta)}(t) = \frac{2^{\alpha+\beta+1}\Gamma(n+\alpha+1)\Gamma(n+\beta+1)}{(2n+\alpha+\beta+1)\Gamma(n+\alpha+\beta+1)n!}\delta_{nm}, \qquad \alpha,\beta > -1,$$

where δ_{mn} is the Kronecker delta. For fixed α, β, the recurrence relation of Jacobi polynomial is

$$\begin{aligned} p_n^{(\alpha,\beta)}(t) =& \frac{2n+\alpha+\beta-1}{2n(n+\alpha+\beta)(2n+\alpha+\beta-2)}\left((2n+\alpha+\beta)(2n+\alpha+\beta-2)t+\alpha^2-\beta^2\right)p_{n-1}^{(\alpha,\beta)}(t) \\ &- \frac{(n+\alpha-1)(n+\beta-1)(2n+\alpha+\beta)}{n(n+\alpha+\beta)(2n+\alpha+\beta-2)}p_{n-2}^{(\alpha,\beta)}(t). \end{aligned}$$

Hermite polynomials

Hermite polynomials H are a system of orthogonal polynomials with respect to the weight function $w : (-\infty, \infty) \to \mathbb{R}$ such that $w(t) = \exp(-t^2/2)$. This is called probabilist's Hermite polynomials, which we will use throughout the thesis. There is another form called physicist's Hermite polynomials with respect to weight $w(t) = \exp(-t^2)$. The explicit expression of probabilist's Hermite polynomials can be written as

$$H_n(t) = n!\sum_{m=0}^{\lfloor \frac{n}{2} \rfloor}\left(-\frac{1}{2}\right)^m \frac{t^{n-2m}}{m!(n-2m)!},$$

with orthogonality property

$$\int_{-\infty}^{\infty} H_m(t)H_n(t)e^{-\frac{t^2}{2}}dt = \sqrt{2\pi}n!\delta_{mn}. \tag{2.1.5}$$

Lastly, we state the recurrence relation of Hermite polynomials as $H_{n+1}(t) = xH_n(t) - nH_{n-1}(t)$. Note that the weight of Hermite polynomials can be viewed as an unnormalised normal distribution. If we are more interested in a particular region far away from the origin, we can define a 'shift-and-scale' version of Hermite polynomials with respect to the weight

$$w^{t_0,\epsilon}(t) = \exp((t-t_0)^2/2\epsilon^2),$$

where t_0 denotes the new centre and ϵ measures standard deviation. Let $(H_n^{t_0,\epsilon})_{n\in\mathbb{N}}$ denotes the shift-and-scale Hermite polynomials. Then the orthogonality property is

$$\int_{-\infty}^{\infty} H_m^{t_0,\epsilon}(t)H_n^{t_0,\epsilon}(t)e^{-\frac{(t-t_0)^2}{2\epsilon^2}}dt = \epsilon\int_{-\infty}^{\infty} H_m^{t_0,\epsilon}(t_0+\epsilon y)H_n^{t_0,\epsilon}(t_0+\epsilon y)e^{-\frac{y^2}{2}}dy,$$

by substitution $y = (t-t_0)/\epsilon$. Now we have the weight of Hermite polynomials. If

$$H_n^{t_0,\epsilon}(t_0+\epsilon y) = H_n(y), \qquad n \in \mathbb{N}, \tag{2.1.6}$$

then $(H_n^{t_0,\epsilon})_{n\in\mathbb{N}}$ is an orthogonal polynomial system with orthogonality

$$\int_{-\infty}^{\infty} H_m^{t_0,\epsilon}(t)H_n^{t_0,\epsilon}(t)e^{-\frac{(t-t_0)^2}{2\epsilon^2}}dt = \epsilon\int_{-\infty}^{\infty} H_m(y)H_n(y)e^{-\frac{y^2}{2}}dy = \epsilon\sqrt{2\pi}n!\delta_{mn},$$

which is followed by the orthogonality of Hermite polynomials in (2.1.5). Similarly, the connection between Hermite and shift-and-scaled Hermite in (2.1.6) provides a way to find the explicit form and recurrence relation of $(H_n^{t_0,\epsilon})_{n\in\mathbb{N}}$, which are

$$H_n^{t_0,\epsilon}(t) = n!\sum_{m=0}^{\lfloor\frac{n}{2}\rfloor}\left(-\frac{1}{2}\right)^m\frac{1}{m!(n-2m)!}\left(\frac{t-t_0}{\epsilon}\right)^{n-2m} \tag{2.1.7}$$

$$H_{n+1}^{t_0,\epsilon}(t) = \frac{1}{\epsilon}(t-t_0)H_n^{t_0,\epsilon}(t) - nH_{n-1}^{t_0,\epsilon}(t). \tag{2.1.8}$$

Remark 2.1.10. Note that there are a simple expression of $(H_n^{t_0,\epsilon})_{n\in\mathbb{N}}$ at $t = t_0$. One can easily observe that

$$H_n^{t_0,\epsilon}(t_0) = \begin{cases}\left(-\frac{1}{2}\right)^{\frac{n}{2}}\frac{n!}{\frac{n}{2}!} & \text{for even } n\\ 0 & \text{for odd } n\end{cases}.$$

2.2 Fourier series

Trigonometric series are also possible ways to represent a function. Here we only present a brief introduction to the Fourier series, which highlights the main results that will be used later.

2.2.1 Trigonometric series

Let $f \in L^1(-\pi,\pi)$, The Fourier series of f is defined by

$$F(t) = \frac{a_0}{2} + \sum_{k=1}^{\infty}\left(a_k\cos(kt) + b_k\sin(kt)\right),$$

where

$$a_k = \frac{1}{\pi}\int_{-\pi}^{\pi} f(t)\cos(kt)dx, \qquad k \geq 0$$
$$b_k = \frac{1}{\pi}\int_{-\pi}^{\pi} f(t)\sin(kt)dx, \qquad k \geq 1,$$

which can be derived from the orthogonal bases $\{\cos kt\}_k$ and $\{\sin kt\}_k$. More generally, we can extend the period to $2l \in \mathbb{R}$ and form exponential series [27]. For $f \in L^1(-l,l)$ and $k \in \mathbb{Z}$,

$$F(t) = \sum_{n=-\infty}^{\infty} c_k e^{i\frac{2\pi}{l}kt}, \qquad c_k = \frac{1}{l}\int_0^l f(t)e^{-i\frac{2\pi}{l}kt}dt. \tag{2.2.1}$$

Linear functional on Fourier series

In the setting of the Fourier series, the process of extracting the k-th coefficient c_k in the exponential form can be defined as $\mathcal{L}_k(x) = c_k^x$ for $x \in L^1(-l,l)$.

Proposition 2.2.1. *$\mathcal{L}_k(x) = c_k^x$ is a linear functional on the space of Fourier series.*

Proof. Choose $x_1(t), x_2(t) \in L^1(-l,l)$ and $\alpha_1,\alpha_2 \in \mathbb{R}$ in arbitrary. By the coefficient of Fourier series in (2.2.1),

$$\begin{aligned}\mathcal{L}_k(\alpha_1x_1+\alpha_2x_2) &= \frac{1}{l}\int_0^l(\alpha_1x_1+\alpha_2x_2)(t)e^{-i\frac{2\pi}{l}kt}dt\\ &= \frac{\alpha_1}{l}\int_0^l x_1(t)e^{-i\frac{2\pi}{l}kt}dt + \frac{\alpha_2}{l}\int_0^l x_2(t)e^{-i\frac{2\pi}{l}kt}dt\\ &= \alpha_1\mathcal{L}_k(x_1) + \alpha_2\mathcal{L}_k(x_2).\end{aligned}$$

which is followed by the linearity of integration. □

2.2.2 Convergence

Note that $F(t)$ and $f(t)$ are closely related. Under some regularity conditions, $F(t)$ converges to $f(t)$. But in other cases, $F(t)$ may not converge to $f(t)$ or even a limit [25]. To examine the convergence of the Fourier series, we define the partial sum of the Fourier series as

$$S_N f(t) = \frac{a_0}{2} + \sum_{k=1}^{N} \left(a_k \cos(kt) + b_k \sin(kt)\right).$$

Now we present pointwise convergence and uniform convergence results [25, 28] of the Fourier series for various functions.

Theorem 2.2.2 (Pointwise convergence for bounded variation)**.** *For 2π-periodic functions of bounded variation on $[-\pi, \pi]$, its Fourier series at arbitrary t converges to*

$$\frac{1}{2}\left(f(t^-) + f(t^+)\right).$$

Theorem 2.2.3 (Uniform convergence for piecewise smooth functions)**.** *If f is a 2π-periodic piecewise smooth function,*

(a) if f is also continuous, then the Fourier series converges uniformly and continuously to f;

(b) if f is not continuous, then the Fourier series converges uniformly to f on every closed interval without discontinuous points.

Theorem 2.2.4 (Uniform error bounds)**.** *Let $f \in C_p^{k,\alpha}(2\pi)$ be a 2π-periodic k-th continuously differentiable function that is Hölder continuous with the exponent $\alpha \in (0, 1]$. Then the 2-norm and infinity-norm bound of a partial sum can be expressed as*

$$\frac{1}{\sqrt{2\pi}}\|f - S_N f\|_2 \le \|f - S_N f\|_\infty \le c_k \frac{\log N}{N^{k+\alpha}}, \qquad \text{for } N \ge 2.$$

For functions only defined in an interval $[a, b]$, we can always shift and extend them to be 2π-periodic functions. These theorems guarantee the convergence of common functions we will use in later experiments. To illustrate, Figure 2.3 shows how convergence theories match with numerical results. Compared with the path $x_2(t) = \sin(10t) + e^{2t} - t$, the other 2 paths have better convergence generally. The main reason is that the Fourier series of x_2 at $t = \pm 1$ does not pointwise converge to $x_2(\pm 1)$. Since the Fourier series treats the interval $[-1, 1]$ as one period over $\mathbb{R}$, by Theorem 2.2.2, the series will converge to $(x_2(-1) + x_2(1))/2$ at $t = \pm 1$, leading to incorrect convergence at boundaries. The same result can also be found in the approximation figure of $x_2(t)$ on the left, where the Fourier series tends to get better and better estimates of points around $t = \pm 1$ and jump back to $(x_2(-1) + x_2(1))/2$ suddenly at $t = \pm 1$. On the other hand, the boundary values match the other two paths, resulting in pointwise convergence for all points.

Compare Figure 2.1 and 2.3, one can observe orthogonal polynomials are better at approximating continuously differentiable path, while Fourier series estimate path with spikes better and their computation is more stable in the long run. Figure 2.4 in the later section provides a summary of convergence results for different types of orthogonal polynomials and Fourier series, which also matches the current results.

2.2.3 Discrete Fourier transform

In reality, it is not feasible to calculate coefficients of the Fourier series in (2.2.1) explicitly for some functions. To calculate these integrals, one way is to discretise them and solve them as a system of linear equations. This is the main idea of the discrete Fourier transform, which will be introduced for the computation of coefficients in the later chapter.

Consider a function $f \in L^1(0, 2\pi)$ such that $f(0) = f(2\pi)$, let $h = 2\pi/N$ and $t_j = jh$. By the trapezium rule, we can express coefficients c_k as

$$c_k = \frac{1}{2\pi}\int_0^{2\pi} f(t)e^{-ikt}dt = \frac{h}{2\pi}\left(\frac{f(t_0)}{2} + \sum_{j=1}^{N-1} f(t_j)e^{-ikt_j} + \frac{f(t_N)}{2}\right) := \frac{1}{N}\sum_{j=0}^{N-1} f(t_j)e^{-i\frac{2\pi}{N}kj}.$$

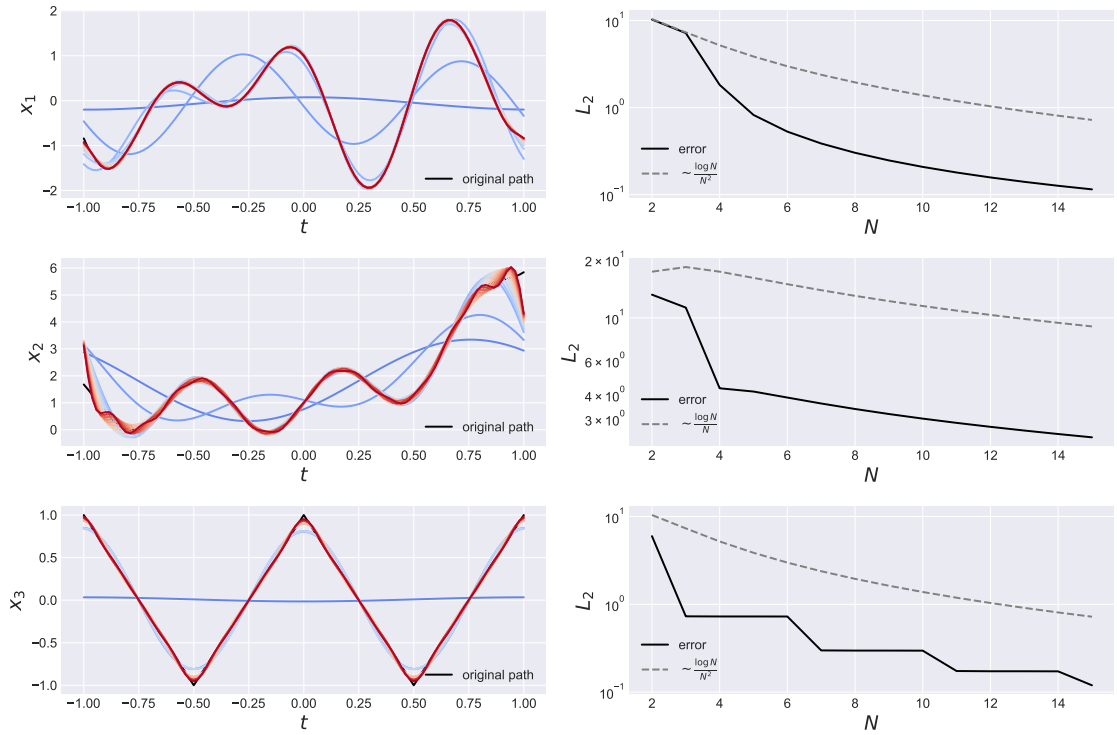

Figure 2.3: Approximation (left) and L_2 convergence (right) results for Fourier series by increasing order N, with the same experimental setting as Figure 2.1. Results with higher order N can be found in Figure A.1

Then the discrete Fourier transform is the sequence of $(c_k)_{k=0}^{N-1}$. To express discrete Fourier transform in the vector form, let $\omega_N = e^{-i\frac{2\pi}{N}}$ and

$$\boldsymbol{f} = \begin{bmatrix} f(t_0) \\ f(t_1) \\ \vdots \\ f(t_N - 1) \end{bmatrix}, \qquad \boldsymbol{F} = \begin{bmatrix} c_0 \\ c_1 \\ \vdots \\ c_{N-1} \end{bmatrix}, \qquad Q_N = \frac{1}{\sqrt{N}} \begin{bmatrix} 1 & 1 & \cdots & 1 \\ 1 & \omega^{-1} & \cdots & \omega^{-(N-1)} \\ \vdots & \vdots & \ddots & \vdots \\ 1 & \omega^{-(N-1)} & \cdots & \omega^{(-(N-1))^2} \end{bmatrix}.$$

Then the discrete Fourier transform can be formulated as

$$\boldsymbol{F} = \frac{1}{\sqrt{N}} Q_N \boldsymbol{f},$$

while its inverse transform can be derived based on unitary Q_N [28] as

$$\boldsymbol{f} = \sqrt{N}(Q_N)^{-1}\boldsymbol{F} = \sqrt{N}\overline{Q}_N^T \boldsymbol{F},$$

where $\overline{Q}_N^T$ is the complex conjugate of Q_N. From the transform and inverse transform, we can see that they are computationally efficient as one does not need to calculate the inverse of a large matrix. The idea of Fast Fourier transform (FFT) takes one step further and expresses Q_{2N} as a simple linear transformation of two copies of Q_N. Fast Fourier transform is powerful since it shrinks the computation of matrix multiplication to one-half of the original size iteratively, resulting $\mathcal{O}(n \log n)$ number of multiplication for Q_n in FFT [25].

2.3 Approximation quality of orthogonal polynomials and Fourier series

In the final section of this chapter, we compare the approximation results of the methods introduced above numerically.

2.3.1 Experiment setup

The experiment is set to compare

- Legendre polynomials: $w(t) = 1$
- two types of Jacobi polynomials: $w(t) = \sqrt{1+t}$, $w(t) = \sqrt{1-t}$
- three types of shift-and-scale Hermite polynomials with different variance for *pointwise approximation*: $\epsilon = 0.1, 0.05, 0.01$
- Fourier series

Regarding pointwise approximation for Hermite polynomials, each sample point t_i of the function will be approximated by a system of Hermite polynomial centred at the point t_i, i.e., $(H_n^{t_i,\epsilon})_{n\in\mathbb{N}}$. To test approximation quality, we simulate random polynomial functions and random trigonometric functions, where sample functions can be found in B.1 and B.2. The L_2 error is then obtained by an average of L_2 error from approximations of 10 functions in each type.

2.3.2 Approximation results

Figure 2.4 illustrates the L_2 error reduction corresponding to an increase in the order of orthogonal polynomials and Fourier series. Among all bases considered, the Fourier series delivers the least desirable approximation result for both path types, attributable to the non-guarantee of pointwise convergence at ± 1 given boundary inconsistencies. Three Jacobi polynomials, including Legendre polynomials, exhibit comparable approximation outcomes, with a slight convergence advantage noted for Legendre polynomials. Conversely, Hermite polynomials demonstrate a significantly reduced approximation error, potentially due to their shifting focus on the point of interest. However, reducing ϵ to achieve greater concentration on sample points can quickly inflate the coefficients of Hermite polynomials, particularly when ϵ is exceedingly small. This behaviour is corroborated by the analytic form and recurrence relation of shift-and-scale Hermite polynomials as shown in (2.1.7) and (2.1.8). Accordingly, Hermite polynomials with $\epsilon = 0.01$ do not outperform those with $\epsilon = 0.05$. It is also worth noting the step pattern of decrease observable in Hermite polynomials, a phenomenon that can be traced back to Remark 2.1.10.

The approximation results elucidate that the Fourier series, in requiring additional assumptions about function values at boundaries, fail to achieve pointwise convergence across all points as effectively as orthogonal polynomials, which generally excel in approximating smooth paths. Among the orthogonal polynomials, Hermite polynomials, even of low degrees, can approximate functions with remarkable precision. However, this precision comes at the cost of extended computation times for each sample point. To mitigate computational expense, we henceforth use Hermite polynomials with $\epsilon = 0.05$ as the representative of the Hermite family. The findings presented in Figure 2.4 play a crucial role in our signature inversion method discussed in the next chapter. They establish a benchmark for the best possible performance attainable in path reconstruction from signatures.

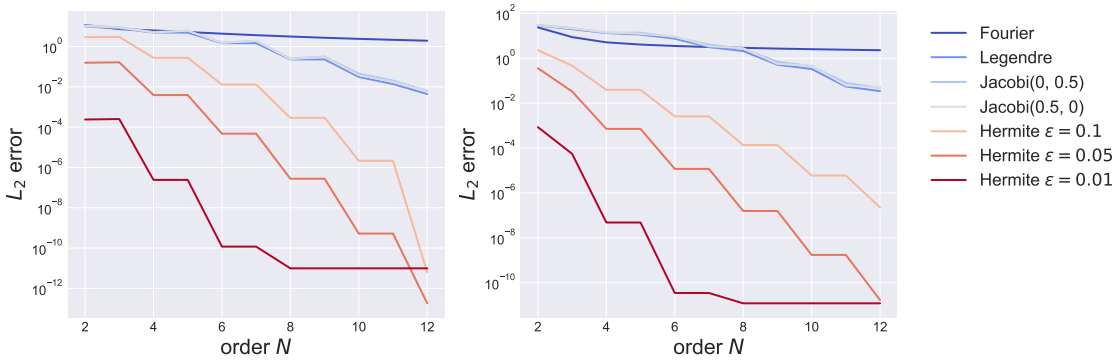

Figure 2.4: L_2 approximation error by different bases. The figures (from left to right) are the corresponding error over the average of 10 random polynomial and trigonometric functions

Chapter 3

Signature inversion

As we described in Chapter 1, signatures encode rich geometric information of a path. Given the uniqueness of paths up to tree-like pieces in [8], it is common to wonder if one could reconstruct a path from its signature. The inversion of signatures is still an active and challenging research area. Inspired by a probabilistic inverse problem [29] with prior knowledge about Brownian motion, X. Geng [30] developed a constructive insight about signature inversion for rough paths. Recently, methods like hyperbolic construction of a path [31] and symmetrization [32] focus on the reconstruction of C^1 path, which are paths that have 1st order continuous derivatives. For implementation, many algorithms [33, 34] are designed for piecewise-linear paths for practical purposes.

In this chapter, we introduce a novel and analytic method to invert the signatures of a path by orthogonal polynomials, with a corresponding learning-based method for the Fourier series. Figure 3.1 shows the proposed method. Compared with inverting signature directly, we will first find the connection between signatures and coefficients for different bases. After deriving the corresponding coefficients, a path can be reconstructed from the basis. Step 2 in Figure 3.1 is quite straightforward. Therefore, in the following sections, we will focus on explaining how the connection in Step 1 is constructed for orthogonal polynomials and Fourier series.

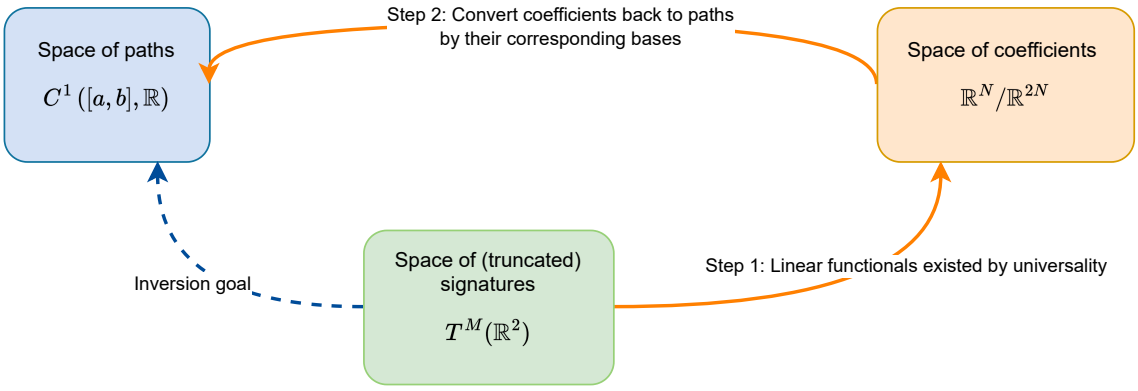

Figure 3.1: Proposed idea of signature inversion

3.1 An analytical method by orthogonal polynomials

3.1.1 Shuffle operations on paths

Recall that in (1.3.1), we can treat a word I as linear function indexing from the signatures $S(x)_{a,b} \in T((V))$ as

$$S_{a,b}(x)^I = (I, S_{a,b}(x)) .$$

In the theorem of shuffle identity 1.3.4, the multiplication of lower level terms of signatures expresses higher order terms based on the rule of shuffle product $\sqcup\!\sqcup$ as

$$(I, S(x)_{a,b})\,(J, S(x)_{a,b}) = (I \sqcup\!\sqcup J, S(x)_{a,b})\,.$$

To simplify the notation in later proofs, we briefly introduce the right half shuffle product $\succ$ [35, 36]. Suppose I, J denote words and $\mathbf{a}$ denotes a letter, which constitutes words. Then the right half shuffle product can be defined recursively by words

$$I \succ \mathbf{a} := I\mathbf{a}, \qquad I \succ J\mathbf{a} := (I \succ J + J \succ I) \cdot \mathbf{a},$$

where $\cdot$ is a concatenation product. Moreover, the right half shuffle product can express iterated integrals by

$$\int_a^b S_{a,t}(x)^I d\left(S_{a,t}(x)^J\right) = (I \succ J, S(x)_{a,b})\,.$$

Example 3.1.1. *Consider a two-dimensional real-valued path $\hat{x} = (t, x(t)) : [a, b] \to \mathbb{R}^2$ with $x(a) = 0$. For the first-level signature, one can compute as follow:*

$$\begin{aligned} S(\hat{x})^1_{a,t} &= \int_a^t ds = t - a = (\boldsymbol{1}, S_{a,t}(\hat{x})) \\ S(\hat{x})^2_{a,t} &= \int_a^t d\left(x(s)\right) = x(t) - x(a) = x(t) = (\boldsymbol{2}, S_{a,t}(\hat{x}))\,. \end{aligned}$$

Then one can express all integrals in terms of powers of $t - a$ and $x(t)$ by signatures of $\hat{x}$. For example, let $n, m \in \mathbb{N}_0$,

$$\begin{aligned} \int_a^b (t-a)^n x(t)^m dt &= \int_a^b (t-a)^n x(t)^m d(t-a) \\ &= \int_a^b \left(S(\hat{x})^1_{a,t}\right)^n \left(S(\hat{x})^2_{a,t}\right)^m d\left(S(\hat{x})^1_{a,t}\right) \\ &= ((\boldsymbol{1}^{\sqcup\!\sqcup n} \sqcup\!\sqcup \boldsymbol{2}^{\sqcup\!\sqcup m}) \succ \boldsymbol{1}, S_{a,b}(\hat{x}))\,, \end{aligned}$$

which is followed by the shuffle identity and right shuffle product in integrals.

3.1.2 Connecting signatures and coefficients of polynomials

Recall Theorem 2.1.7 that all orthogonal polynomials have 3-term recurrence relation. For a sequence of orthogonal polynomials $(p_n)_{n\in\mathbb{N}}$ with respect to the inner product

$$\langle f, g\rangle := \int_a^b f(t)g(t)\omega(t)dt, \tag{3.1.1}$$

where $\omega : [a, b] \to \mathbb{R}$. The polynomial p_n satisfies the recurrence form

$$p_n(t) = (A_n t + B_n)p_{n-1}(t) + C_n p_{n-2}(t), \qquad n \geq 2, \tag{3.1.2}$$

with $p_0(t) = 1$ and $p_1(t) = A_1 t + B_1$. Let $x \in L^2_\omega$ such that $x(a) = 0$. The following function converges pointwise to

$$x(t) \sim \sum_{n=0}^{\infty} \alpha_n^x p_n(t),$$

where

$$\alpha_n^x = \frac{1}{\langle p_n, p_n\rangle} \int_a^b x(t)p_n(t)\omega(t)dt.$$

If $\omega(t)x(t)$ is well defined over $t \in [a, b]$, let $\hat{x} : [a, b] \to \mathbb{R}^2$ be the time-augmented weighted path such that $\hat{x} = (t, \omega(t)x(t))$ with respect to the inner product (3.1.1). In this chapter, We will use the shorter notation of signatures $S(\hat{x}) := S(\hat{x})_{a,b}$.

Proposition 3.1.1. *Let $\hat{x} = (t, \omega(t)x(t))$, where $\omega(t)$ corresponds to the weight function of a system of orthogonal polynomials. For any $n \in \mathbb{N} \cup \{0\}$, there exists a unique $\ell_n \in T((\mathbb{R}^2))^*$ such that for any path x,*

$$\alpha_n^x = (\ell_n, S(\hat{x})).$$

Furthermore, the linear functional $(\ell_n)_{n\geq 0}$ satisfy the following recurrence relation

$$\ell_n = A_n \frac{\langle p_{n-1}, p_{n-1}\rangle}{\langle p_n, p_n\rangle} \mathbf{1} \succ \ell_{n-1} + (A_n a + B_n)\frac{\langle p_{n-1}, p_{n-1}\rangle}{\langle p_n, p_n\rangle}\ell_{n-1} + C_n \frac{\langle p_{n-2}, p_{n-2}\rangle}{\langle p_n, p_n\rangle}\ell_{n-2},$$

with

$$\ell_0 = \frac{A_0}{\langle p_0, p_0\rangle}\mathbf{21} \quad \text{and} \quad \ell_1 = \frac{A_1}{\langle p_1, p_1\rangle}(\mathbf{121} + \mathbf{211}) + \frac{A_1 a + B_1}{\langle p_1, p_1\rangle}(\mathbf{21}).$$

Proof. By the universality of signature in Theorem 1.3.7, there exists such linear functionals. One can express the first two coefficients in an orthogonal polynomial expansion of x by its signature:

$$\begin{aligned}
\alpha_0^x &= \frac{1}{\langle p_0, p_0\rangle}\int_a^b A_0 x(t)\omega(t)dt \\
&= \left(\frac{A_0}{\langle p_0, p_0\rangle}\mathbf{2} \succ \mathbf{1}, S(\hat{x})\right) \\
&= \left(\frac{A_0}{\langle p_0, p_0\rangle}\mathbf{21}, S(\hat{x})\right) \\
&:= (\ell_0, S(\hat{x})) \\
\alpha_1^x &= \frac{1}{\langle p_1, p_1\rangle}\int_a^b (A_1 t + B_1)x(t)\omega(t)dt \\
&= \frac{A_1}{\langle p_1, p_1\rangle}\int_a^b (t-a)x(t)\omega(t)dt + \frac{A_1 a + B_1}{\langle p_1, p_1\rangle}\int_a^b x(t)\omega(t)dt \\
&= \frac{A_1}{\langle p_1, p_1\rangle}((\mathbf{1} \sqcup\!\sqcup \mathbf{2}) \succ \mathbf{1}, S(\hat{x})) + \frac{A_1 a + B_1}{\langle p_1, p_1\rangle}(\mathbf{21}, S(\hat{x})) \\
&= \left(\frac{A_1}{\langle p_1, p_1\rangle}(\mathbf{121} + \mathbf{211}) + \frac{A_1 a + B_1}{\langle p_1, p_1\rangle}(\mathbf{21}), S(\hat{x})\right) \\
&:= (\ell_1, S(\hat{x})).
\end{aligned}$$

Then one can find ℓ_n recursively by multiplying both sides of the equation (3.1.2) by $x(t)\omega(t)$ and integrating on $[a, b]$:

$$\begin{aligned}
\int_a^b p_n(t)x(t)\omega(t)dt = \int_a^b (A_n t + B_n)p_{n-1}(t)x(t)\omega(t)dt + \int_a^b C_n p_{n-2}(x)x(t)\omega(t)dt \\
= A_n \int_a^b (t-a)d\left(\int_a^t p_{n-1}(s)x(s)\omega(s)ds\right) \\
+ (A_n a + B_n)\int_a^b p_{n-1}(t)x(t)\omega(t)dt \\
+ C_n \int_a^b p_{n-2}(x)x(t)\omega(t)dt.
\end{aligned}$$

By definition of α_i^x,

$$\begin{aligned}
\int_a^b p_n(t)x(t)\omega(t)d = \langle p_n, p_n\rangle\alpha_n^x = \langle p_n, p_n\rangle(\ell_n, S(\hat{x})) \\
A_n \int_a^b (t-a)d\left(\int_a^t p_{n-1}(s)x(s)\omega(s)ds\right) = A_n\langle p_{n-1}, p_{n-1}\rangle(\mathbf{1} \succ \ell_{n-1}, S(\hat{x})) \\
(A_n a + B_n)\int_a^b p_{n-1}(t)x(t)\omega(t)dt = (A_n a + B_n)\langle p_{n-1}, p_{n-1}\rangle(\ell_{n-1}, S(\hat{x})) \\
C_n \int_a^b p_{n-2}(x)x(t)\omega(t)dt = C_n\langle p_{n-2}, p_{n-2}\rangle(\ell_{n-2}, S(\hat{x})).
\end{aligned}$$

Therefore, the recurrence relation of signature with respect to orthogonal polynomials is

$$\begin{aligned}(\ell_n, S(\hat{x})) =& A_n \frac{\langle p_{n-1}, p_{n-1}\rangle}{\langle p_n, p_n\rangle}(\mathbf{1} \succ \ell_{n-1}, S(\hat{x})) \\ &+ (A_n a + B_n)\frac{\langle p_{n-1}, p_{n-1}\rangle}{\langle p_n, p_n\rangle}(\ell_{n-1}, S(\hat{x})) \\ &+ C_n \frac{\langle p_{n-2}, p_{n-2}\rangle}{\langle p_n, p_n\rangle}(\ell_{n-2}, S(\hat{x})).\end{aligned}$$

□

From the proof, there are several assumptions about orthogonal polynomials made to derive the recurrence relation. Firstly, the interval defined on the inner space is compact. Secondly, $w(t)$ is well defined on the closed interval. This may lead to a limited range of orthogonal polynomials. For example, since the range of Hermite polynomials is not bounded, the polynomial is not applicable based on our theorem. However, if we use a scale-and-shift version of the polynomials with most of the weight density centred at a point, their weight can be truncated to a compact interval numerically. In Section 2.1.3, the relation between the original Hermite polynomials and the scale-and-shift Hermite polynomials is stated in (2.1.6). One can centre the weight density by a small ϵ and shift the density to an interested point t_i. Since the non-zero density is centred as a small interval, Lemma 3.1.1 can be used on the truncated density over the interval.

3.1.3 Taylor approximation of the weight function

The results in 3.1.1 require signatures of $\hat{x} = (t, w(t)x(t))$. However, sometimes one may only have signatures of $\tilde{x} = (t, x(t))$. Here we propose a theoretically applicable method by approximating the weight function in the Taylor series.

Consider the Taylor approximation of ω around $t = a$, i.e.,

$$\omega(t) \approx \sum_{i=0}^{M} \frac{d^i\omega}{dt^i}\Big|_{t=a}(t-a)^i := \sum_{i=0}^{M} \omega_i (t-a)^i.$$

Letting $\tilde{x}_t = (t, x(t))$ and

$$c_i := (\mathbf{2} \sqcup\!\sqcup \mathbf{1}^{\sqcup\!\sqcup i}) \succ \mathbf{1} = i!(\mathbf{21\ldots1} + \mathbf{121\ldots1} + \ldots + \mathbf{1\ldots121}),$$

we have

$$\begin{aligned}\alpha_0^x &= \frac{1}{\langle p_0, p_0\rangle}\int_a^b A_0 x(t)\omega(t)dt \\ &= \frac{A_0}{\langle p_0, p_0\rangle}\sum_{i=0}^{M}\omega_i \int_a^b (t-a)^i x(t)\mathrm{d}t \\ &= \left(\frac{A_0}{\langle p_0, p_0\rangle}\sum_{i=0}^{M}\omega_i(\mathbf{2} \sqcup\!\sqcup \mathbf{1}^{\sqcup\!\sqcup i}) \succ \mathbf{1}, S(\tilde{x})\right) \\ &= \left(\frac{A_0}{\langle p_0, p_0\rangle}\sum_{i=0}^{M}\omega_i c_i, S(\tilde{x})\right) \\ &:= (\ell_0, S(\tilde{x})), \\ \alpha_1^x &= \frac{1}{\langle p_1, p_1\rangle}\int_a^b (A_1 t + B_1)x(t)\omega(t)\mathrm{d}t \\ &= \frac{1}{\langle p_1, p_1\rangle}\int_a^b \left(A_1(t-a) + A_1 a + B_1\right)x(t)\omega(t)\mathrm{d}t \\ &= \frac{1}{\langle p_1, p_1\rangle}\sum_{i=0}^{M}\omega_i \int_a^b \left(A_1(t-a)^{i+1} + (A_1 a + B_1)(t-a)^i\right)x(t)\mathrm{d}t \\ &= \left(\frac{1}{\langle p_1, p_1\rangle}\sum_{i=0}^{M}\omega_i\left(A_1(\mathbf{2} \sqcup\!\sqcup \mathbf{1}^{\sqcup\!\sqcup i+1}) + (A_1 a + B_1)(\mathbf{2} \sqcup\!\sqcup \mathbf{1}^{\sqcup\!\sqcup i})\right) \succ \mathbf{1}, S(\tilde{x})\right)\end{aligned}$$

$$= \left(\frac{1}{\langle p_1, p_1 \rangle} \sum_{i=0}^{M} \omega_i (A_1 c_{i+1} + (A_1 a + B_1) c_i), S(\tilde{x}) \right)$$
$$:= (\ell_1, S(\tilde{x}))$$

By induction the same relation of 3.1.1 holds, and because it holds for any path x and signatures span the tensor algebra we must have

$$\ell_n = A_n \frac{\langle p_{n-1}, p_{n-1} \rangle}{\langle p_n, p_n \rangle} \mathbf{1} \succ \ell_{n-1} + (A_n a + B_n) \frac{\langle p_{n-1}, p_{n-1} \rangle}{\langle p_n, p_n \rangle} \ell_{n-1} + C_n \frac{\langle p_{n-2}, p_{n-2} \rangle}{\langle p_n, p_n \rangle} \ell_{n-2}.$$

Remark 3.1.2. There are several reasons why we would say the Taylor approximation method is 'theoretically applicable'.

1. Firstly and most importantly, the expansion of weight function around the starting point a is hard to find analytically; even if one manages to find the series, it may diverge.

2. Secondly, if the series converge, one still needs to determine how many orders of approximation lead to an error within a certain tolerance. Moreover, if the convergence rate is slow, more orders of series are needed, resulting in higher levels of truncation of signature. In this case, the computation of signature would increase exponentially, which costs computational inefficiency and potential underflow/overflow errors.

For these reasons, signature inversion with Taylor approximation will not be demonstrated in the result section later.

3.2 A learning-based method by Fourier series

Since a linear map within arbitrary precision is guaranteed by the universality of signature in 1.3.7, there exist linear functionals $\ell_\cdot$ such that

$$a_k^x = (\ell_{a_k}, S(\hat{x})), \qquad b_k^x = (\ell_{b_k}, S(\hat{x}))$$

Unfortunately, there is no recurrence relation in the Fourier series that could connect with shuffle identity in 1.3.4. Currently, the coefficients of the Fourier series cannot be found recursively. However, the coefficient can still be learned by linear regression, which leads to our second proposal.

In this case, one can first construct a dataset of paths, and then extract coefficients of Fourier series from training paths. After choosing suitable truncated levels of signatures and order of Fourier series, the linear regression model can learn to find coefficients from signatures, i.e., treating signatures as features and coefficients as targets. For a new path, its coefficients of Fourier series can be predicted based on the signatures of the path. Therefore, the path will then be reconstructed by the coefficients with trigonometric bases.

3.3 Results outline

In this section, we will summarise those two proposed methods by visualisation of sample paths and error estimation for a set of paths.

3.3.1 Visualisation of inversion by different bases

To evaluate inversion results, both low-frequency and high-frequency trigonometric paths are generated as a comparison. Figure 3.2 presents the outcomes of inversions via five different bases; the first four, employing orthogonal polynomial bases, explicitly invert the signature, while the fifth utilises the Fourier series to learn reconstruction through linear regression. In each plot, the path reconstruction from signature is depicted in red, whereas the reconstruction derived solely from the bases is shown in blue. The latter serves as a benchmark, representing the optimal outcome achievable through inversion. The reconstruction results rely on 4 main factors, which are

1. degree/order of bases (N) and levels of truncated signatures;

2. 'complexity' of paths, such as frequency and smoothness;

3. the inversion method, whether it is explicit or learning-based;

4. the weight function of orthogonal polynomials.

Factor 1 significantly influences the path reconstruction by bases, consequently leading to a varied performance in signature inversion. The orders of the bases employed here are depicted in Table 3.1, establishing a relationship with the levels of truncated signatures as supported by Proposition 3.1.1. While higher levels of signatures could potentially be utilised, the coefficients of orthogonal polynomials depend only on the signature level, not exceeding 2 of the degree of the orthogonal polynomials. For the Fourier series, the base order is chosen to align with that of the orthogonal polynomials, setting the signature level at 6, given that the number of features grows exponentially by $\sum_{i=0}^{M} 2^i = 2^{M+1} - 1$. Figure 2.4 indicates that as the order increases, the approximation of the original paths by various bases improves, thereby it is expected that reconstruction from signatures can increasingly resemble the original paths.

Factor 2 also crucially contributes to the approximation by bases. A comparison between the first and second columns of Figure 3.2 reveals that all bases can approximate the simple path featured in the first column more accurately. However, the Jacobi polynomials are less effective in approximating the high-frequency path with the current degree of polynomials, as demonstrated in the second column. Consequently, more complex paths might yield less satisfactory inversion results due to the limitations of bases.

Factor 3 has a direct and clear influence on signature inversion. Despite the Fourier series successfully approximating both paths, the learning-based inversion method is unable to establish a definitive linear map between coefficients and signatures. This difficulty explains the observed match of general patterns, rather than exact matches, between the original paths and their reconstruction from signatures. In contrast, signature inversion via orthogonal polynomials is explicit and consequently, one can reasonably expect the signature inversion to closely approximate the reconstruction from bases.

Relative to the above factors, factor 4 plays a minor role in the reconstruction process. As observed in Figure 3.2, the left tail of Jacobi$(0, 0.5)$ and the right tail of Jacobi$(0.5, 0)$ approximations tend towards divergence, likely due to overflow errors as their weight functions approach zero at $t \to \mp 1$. Meanwhile, the signature inversion of Hermite polynomials, conducted on a pointwise basis like the base approximation, yields precise results even when lower degrees of polynomials are utilised, due to each sample point being estimated at the centre of the weight function.

Approximation methods	Orders (N)	Levels of truncated signatures
Legendre	10	N+2=12
Jacobi$(0, 0.5)$	10	N+2=12
Jacobi$(0.5, 0)$	10	N+2=12
Hermite ($\epsilon = 0.05$)	2	N+2=4
Fourier	10	6

Table 3.1: Approximation methods with corresponding orders used in Figure 3.2

3.3.2 Error comparison of explicit signature inversion

As the learning-based inversion method via the Fourier series only reveals general patterns of the original path, the error between the reconstructed and ground-truth paths is expected to be large. Therefore, we only compare the L_2 error of signature inversion via orthogonal polynomials.

Figure 3.3 depicts the L_2 error and time consumption per path as the polynomial degree accumulates. Since both polynomial and trigonometric path approximations via orthogonal polynomials exhibit a similar trend of decreasing error with increasing polynomial degrees, trigonometric paths were chosen for illustration. A comparison between Figure 2.4 and the left plot of Figure 3.3 reveals that the inversion results fail to reach their lower bounds, defined by the orthogonal polynomial approximations. The L_2 error for signature inversion via orthogonal polynomials is approximately

Figure 3.2: Inversion results for low-frequency and high-frequency examples, with approximation bases (from top to bottom) Legendre (Jacobi$(0,0)$), Jacobi$(0,0.5)$, Jacobi$(0.5,0)$, Hermite ($\epsilon = 0.05$) and Fourier

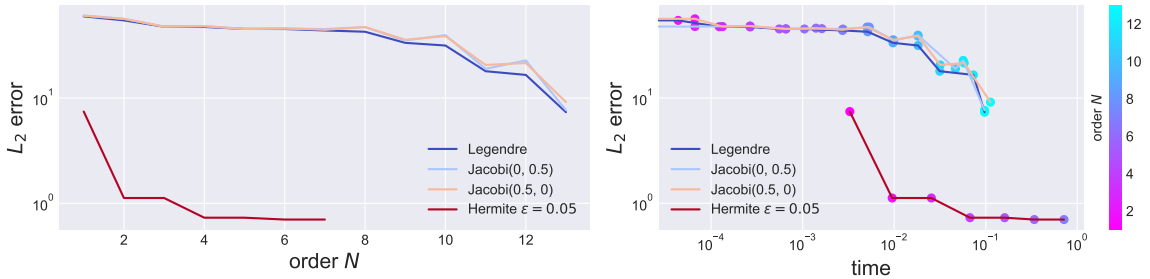

Figure 3.3: L_2 error of signature inversion via orthogonal polynomials with respect to order and time; error and time are calculated by an average of 15 paths with 200 sample points

tenfold larger than the error associated with polynomial approximation. This discrepancy arises due to underflow errors incurred during the computation of linear functionals. However, these errors appear visually negligible in the figures, largely attributable to a large number of sample points and the wide vertical span of the paths.

Drawing parallels with the pattern exhibited by polynomial approximation, pointwise signature inversion via Hermite polynomials outperforms inversion via Jacobi polynomials. For pointwise inversion with Hermite polynomials, a smaller polynomial degree typically yields desirable results. Conversely, increasing the degree of Hermite polynomials induces numerical instability due to the computation of terms with ϵ, resulting in exponential time consumption. Supporting this observation, the right-hand plot of Figure 3.3 presents the time consumption against the L_2 error, with an increasing degree of polynomials. The plotted points, differentiated by colour, correspond to the degrees of the polynomials used for inversion. Drawing from the plot, we can deduce the following:

- The inversion time is roughly equivalent across different Jacobi polynomials.
- For a given running time, signature inversion using lower-order Hermite polynomials yields superior path reconstruction compared to inversion via higher-order Jacobi polynomials.
- When comparing inversions in equal precision, the use of Hermite polynomials accelerates the process by approximately a factor of ten compared to Jacobi polynomials.

3.3.3 More on rough paths

Finally, we provide a brief demonstration of signature inversion on rough paths. Paths are generated from fractional Brownian motion (FBM) [37] with Hurst index 0.5 and 0.9, which ranges from 0 to 1. Sample paths of these two kinds can be found in Figures B.3 and B.4, which supports that the higher the Hurst is, the smoother the realisation is.

Figure 3.5 shows the inversion results via different bases on FBM paths. All inversions via polynomials reveal the general trend of the paths. Notably, pointwise inversion via Hermite polynomials captures more subtle changes in the paths. On the other hand, learning-based inversion via the Fourier series generalised less well for FBM with Hurst 0.5, which is due to its varying and unpredictable characteristics. Figure 3.4 illustrates the L_2 error and time consumption per FBM path as the polynomial degree accumulates, which shares similar patterns to Figure 3.3.

In summary, this work has derived an analytical signature inversion method using orthogonal polynomials, which is also numerically validated on both smooth paths (trigonometric waves) and rough paths (FBM). Among the polynomial choices, Hermite polynomials excel in capturing precise path information efficiently, thereby laying the groundwork for future research in this domain.

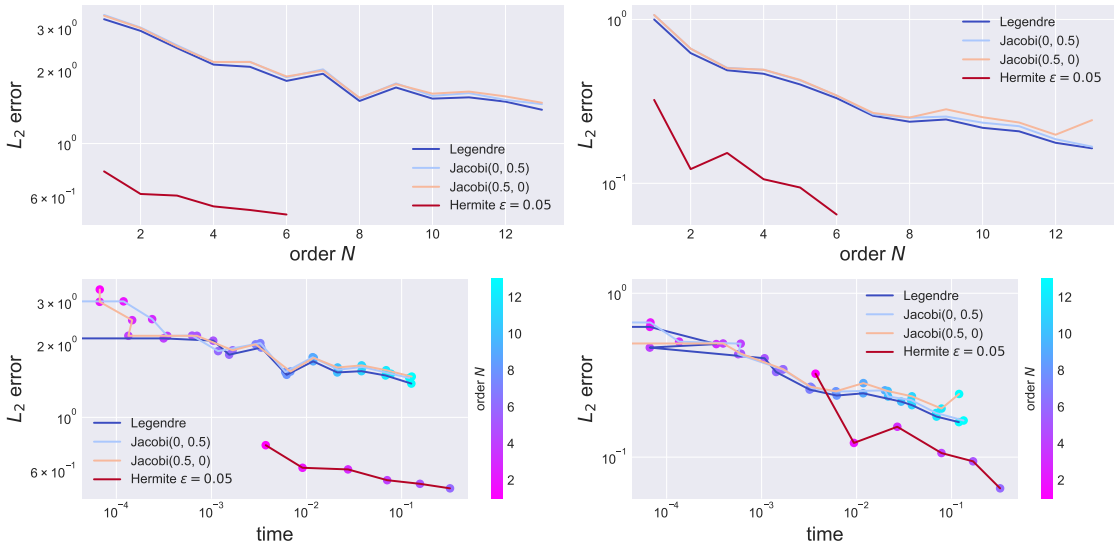

Figure 3.4: L_2 error of signature inversion via orthogonal polynomials with respect to order and time; settings are the same as Figure 3.3, paths are generated from fractional Brownian motion with the Hurst index 0.5 and 0.9 (from left to right)

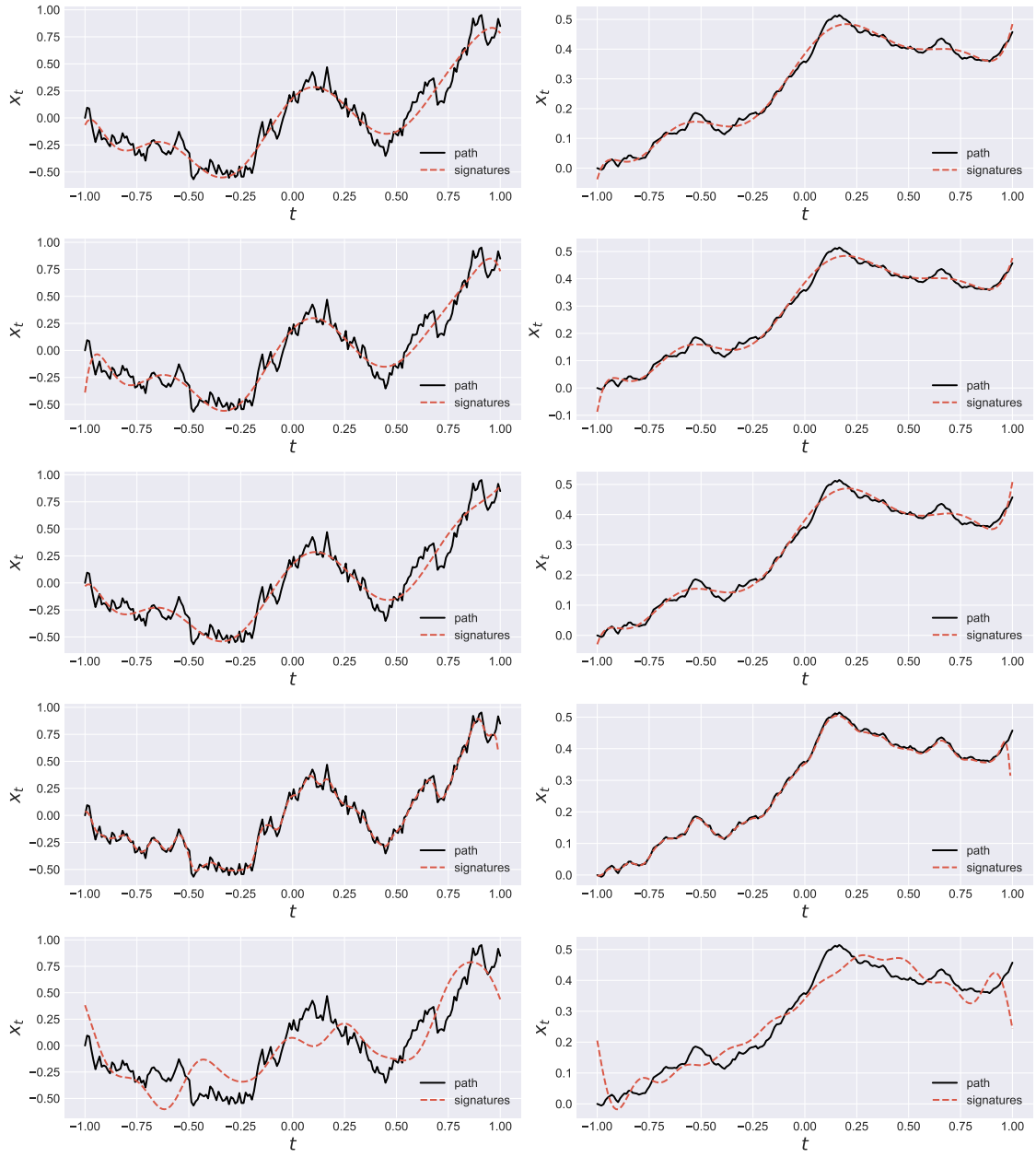

Figure 3.5: Inversion results on fractional Brownian motion with Hurst 0.5 and 0.9, with approximation bases (from top to bottom) Legendre (Jacobi$(0,0)$), Jacobi$(0,0.5)$, Jacobi$(0.5,0)$, Hermite ($\epsilon = 0.05$) and Fourier

Chapter 4

Summary and future work

In this thesis, we explored an innovative method of signature inversion via orthogonal polynomials. We began in Chapter 1 by establishing the foundational definitions and properties of signatures, focusing on the concept of universality. This concept is pivotal in linearising non-linear functions by signatures, which was later extended to coefficients of orthogonal polynomials. Building on path decomposition methods in Chapter 2, Chapter 3 provided an in-depth examination of the analytical inversion method using orthogonal polynomials, evidenced by results from randomly generated paths. With a match between theoretical and numerical results, we can conclude that the inversion is exact and its performance depends on the capability of orthogonal polynomials approximating paths.

As we reflect on the accomplishments and insights gained through the research presented in this thesis, we also recognise numerous potentials for future exploration.

1. To begin with, there is potential to delve deeper into the convergence results for orthogonal polynomials. By determining the analytical error bound of orthogonal polynomial approximation in relation to path smoothness, the type of orthogonal polynomials, and the degree of the polynomials, we could establish an equivalent theoretical error bound for signature inversion. This would enhance our understanding of the reconstruction performance.

2. Additionally, the concept of Taylor's expansion of the weight function offers an insightful basis for the reconstruction of paths from 'non-weighting' time-augmented signatures, which are frequently encountered in inversion problems. Developing a well-behaved Taylor expansion around certain points and aligning it with the starting point could lead to the derivation of an analytical inversion method from time-augmented signatures.

3. Moreover, the sample paths in this report are only 1-dimensional. However, in practical applications such as quantitative finance and bioinformatics, paths are likely to be high-dimensional and rough. To address this discrepancy, future work could extend the method to accommodate multi-dimensional paths. The refined methodology could subsequently be applied to real-world data such as DNA and protein sequences, opening up more possibilities for practical applications.

In conclusion, the research presented in this thesis serves as a foundation for further exploration and advancement in the field. By pursuing the avenues outlined above, we can continue to push the boundaries of knowledge and contribute to practical applications in various domains.

Bibliography

[1] Norden E. Huang, Zheng Shen, Steven R. Long, Manli C. Wu, Hsing H. Shih, Quanan Zheng, Nai-Chyuan Yen, Chi Chao Tung, and Henry H. Liu. The empirical mode decomposition and the hilbert spectrum for nonlinear and non-stationary time series analysis. *Proceedings of the Royal Society. A, Mathematical, physical, and engineering sciences*, 454(1971):903–995, 1998.

[2] Danilo P. Mandic, Naveed Ur Rehman, Zhaohua Wu, and Norden E. Huang. Empirical mode decomposition-based time-frequency analysis of multivariate signals: The power of adaptive data analysis. *IEEE signal processing magazine*, 30(6):74–86, 2013.

[3] Ilya Chevyrev and Harald Oberhauser. Signature moments to characterize laws of stochastic processes. *Journal of Machine Learning Research*, June 2022.

[4] Thomas Cass and Cristopher Salvi. Rough paths in machine learning. Class notes from Imperial College London, 2023.

[5] Ilya Chevyrev and Andrey Kormilitzin. A primer on the signature method in machine learning. 1603.03788, arXiv, March 2016.

[6] Terry Lyons and Andrew D. McLeod. Signature methods in machine learning. 2206.14674, arXiv, 2023.

[7] 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.

[8] Ben Hambly and Terry Lyons. Uniqueness for the signature of a path of bounded variation and the reduced path group. *Annals of mathematics*, 171(1):109–167, 2010.

[9] Horatio Boedihardjo, Xi Geng, Terry Lyons, and Danyu Yang. The signature of a rough path: Uniqueness. *Advances in mathematics (New York. 1965)*, 293:720–737, 2016.

[10] Thomas Cass and William F. Turner. Topologies on unparameterised path space. 2206.11153, arXiv, 2022.

[11] Terry Lyons, M. Caruana, and T. Lévy. *Differential Equations Driven by Rough Paths*, volume 1908 of *École d'Été de Probabilités de Saint-Flour, 1908*. Springer Berlin Heidelberg, Berlin, Heidelberg, 1st ed. edition, 01 2004.

[12] Terry Lyons. `ESig`. `https://esig.readthedocs.io/en/latest/`, 2017.

[13] Franz J. Király and Harald Oberhauser. Kernels for sequentially ordered data. *J. Mach. Learn. Res.*, 20(1):1041–1085, jan 2019.

[14] Maud Lemercier, Cristopher Salvi, Theodoros Damoulas, Edwin V. Bonilla, and Terry Lyons. Distribution regression for sequential data. 2006.05805, arXiv, 2021.

[15] Benjamin Graham. Sparse arrays of signatures for online character recognition. 1308.0371, arXiv, 2013.

[16] Lajos Gergely Gyurkó, Terry Lyons, Mark Kontkowski, and Jonathan Field. Extracting information from the signature of a financial data stream. 1307.7244, arXiv, 2014.

[17] Shujian Liao, Terry Lyons, Weixin Yang, and Hao Ni. Learning stochastic differential equations using rnn with log signature features. 2008.03408, arXiv, 2019.

[18] James Morrill, Cristopher Salvi, Patrick Kidger, James Foster, and Terry Lyons. Neural rough differential equations for long time series. 2008.03408, arXiv, 2021.

[19] Imanol Perez Arribas, Guy M. Goodwin, John R. Geddes, Terry Lyons, and Kate E. A. Saunders. A signature-based machine learning model for distinguishing bipolar disorder and borderline personality disorder. *Translational Psychiatry*, 8, 2017.

[20] Bo Wang, Yue Wu, Niall Taylor, Terry Lyons, Maria Liakata, Alejo J Nevado-Holgado, and Kate E A Saunders. Learning to detect bipolar disorder and borderline personality disorder with language and speech in non-clinical interviews. 2008.03408, arXiv, 2021.

[21] Paul Moore, Terry Lyons, and J. Gallacher. Using path signatures to predict a diagnosis of alzheimer's disease. *PLOS ONE*, 14:e0222212, 09 2019.

[22] Andrey Kormilitzin, Kate E. A. Saunders, Paul J. Harrison, John R. Geddes, and Terry Lyons. Detecting early signs of depressive and manic episodes in patients with bipolar disorder using the signature-based model. 2008.03408, arXiv, 2017.

[23] Gabor. Szego. *Orthogonal polynomials*. Colloquium publications, American Mathematical Society ; v.23. American Mathematical Society, Providence, R.I, 3rd ed. edition, 1967.

[24] Mama Foupouagnigni. An introduction to orthogonal polynomials. In Mama Foupouagnigni and Wolfram Koepf, editors, *Orthogonal Polynomials*, pages 3–24, Cham, 2020. Springer International Publishing.

[25] Kendall. Atkinson. *Theoretical Numerical Analysis A Functional Analysis Framework*. Texts in Applied Mathematics, 39. Springer New York, New York, NY, 3rd ed. 2009. edition, 2009.

[26] Mourad Ismail. *Classical and quantum orthogonal polynomials in one variable /*. Encyclopedia of mathematics and its applications ; v. 98. Cambridge University Press, Cambridge, 2005.

[27] David Morin. Chapter 3 fourier analysis. Notes, Nov 2009. `https://scholar.harvard.edu/files/david-morin/files/waves_fourier.pdf`.

[28] Edward Beckenstein George Bachman, Lawrence Narici. *Fourier and wavelet analysis*. Universitext. Springer, New York, 1 edition, 2000.

[29] Yves Le Jan and Zhongmin Qian. Stratonovich's signatures of brownian motion determine brownian sample paths. *Probability theory and related fields*, 157(1-2):209–223, 2013.

[30] Xi Geng. Reconstruction for the signature of a rough path. *Proceedings of the London Mathematical Society*, 114(3):495–526, 2017.

[31] Terry J. Lyons and Weijun Xu. Hyperbolic development and inversion of signature. *Journal of functional analysis*, 272(7):2933–2955, 2017.

[32] Terry Lyons and Weijun Xu. Inverting the signature of a path. *Journal of the European Mathematical Society : JEMS*, 20(7):1655–1687, 2018.

[33] Jeremy Reizenstein and Benjamin Graham. Algorithm 1004: The iisignature library: Efficient calculation of iterated-integral signatures and log signatures. *ACM Transactions on Mathematical Software (TOMS)*, 2020.

[34] Jiawei Chang and Terry Lyons. Insertion algorithm for inverting the signature of a path. 1907.08423, arXiv, 2019.

[35] Loic Foissy and Frédéric Patras. Natural endomorphisms of shuffle algebras. *Int. J. Algebra Comput.*, 23:989–, 2012.

[36] Eric Gehrig and Matthias Kawski. A hopf-algebraic formula for compositions of noncommuting flows. In *Proceedings of the 47th IEEE Conference on Decision and Control, CDC 2008*, Proceedings of the IEEE Conference on Decision and Control, pages 1569–1574. Institute of Electrical and Electronics Engineers Inc., 2008. 47th IEEE Conference on Decision and Control, CDC 2008 ; Conference date: 09-12-2008 Through 11-12-2008.

[37] Benoit B. Mandelbrot and John W. Van Ness. Fractional brownian motions, fractional noises and applications. *Siam Review*, 10:422–437, 1968.

Appendix A

More convergence results for Fourier series

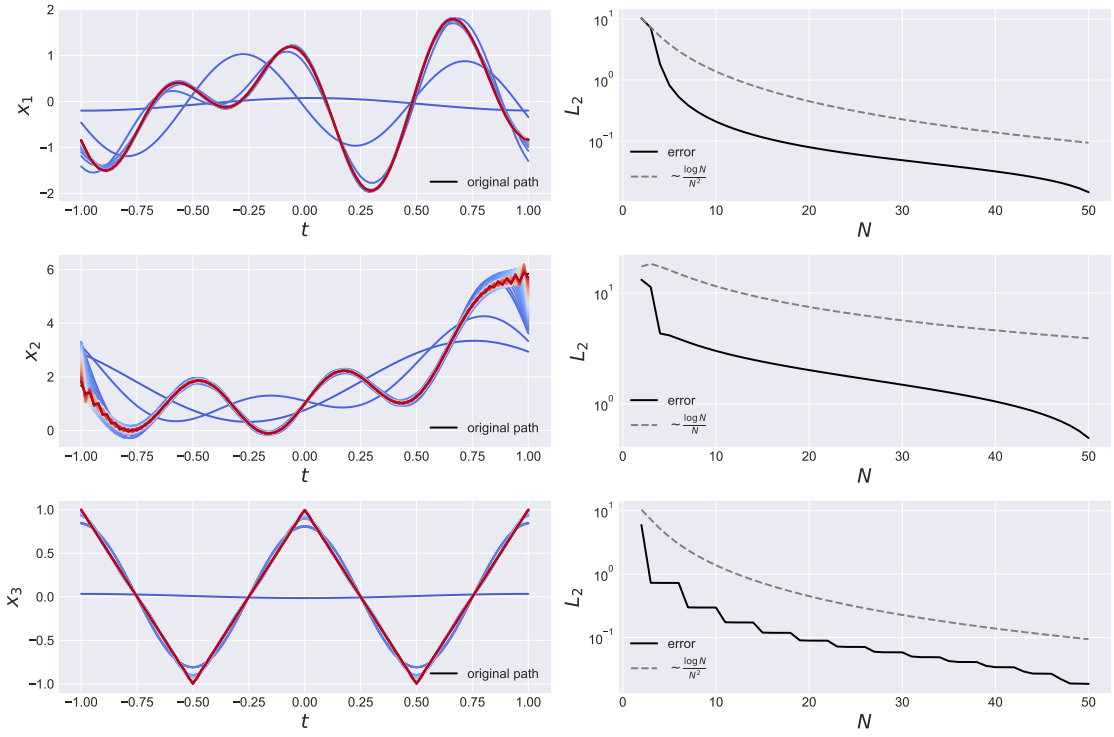

Figure A.1: Approximation (left) and L_∞ convergence (right) results for Fourier series by more order N, with the same experimental setting as figure 2.1

Appendix B

Samples of random generated path

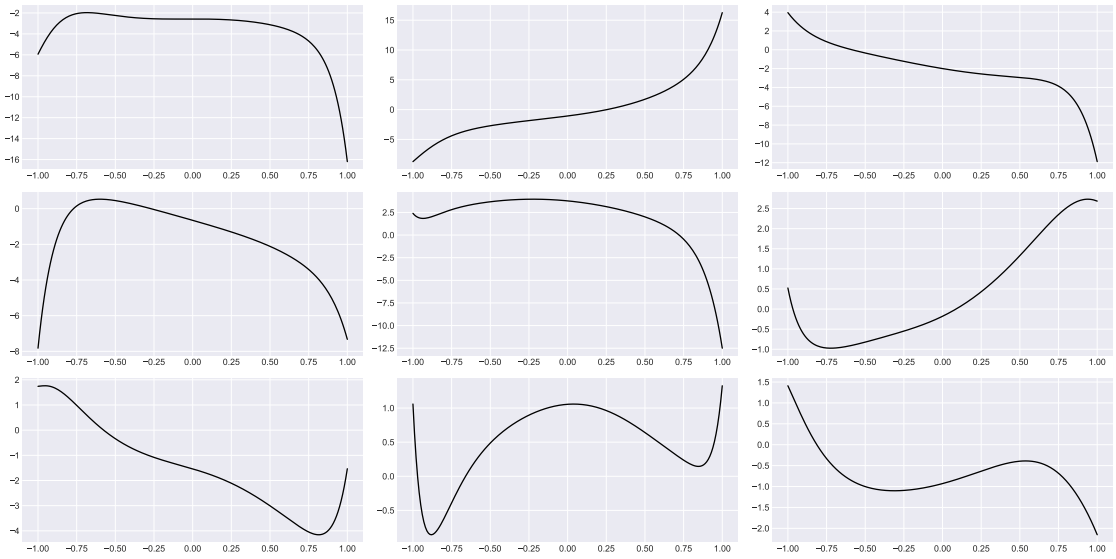

Figure B.1: Samples of generated polynomial paths with the possible degree ranging from 10 to 20

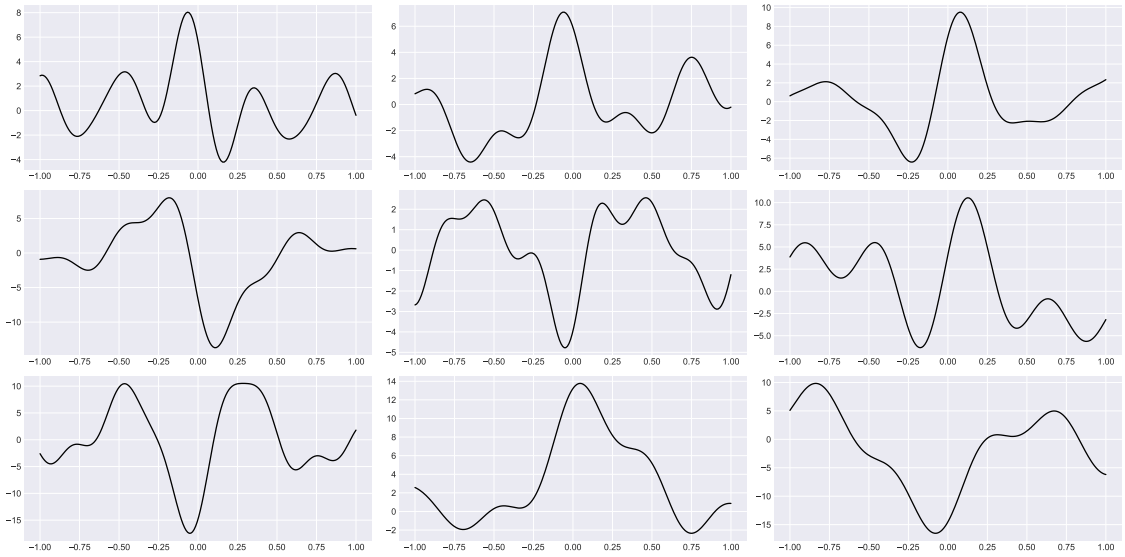

Figure B.2: Samples of generated trigonometric paths with highest possible frequency 15

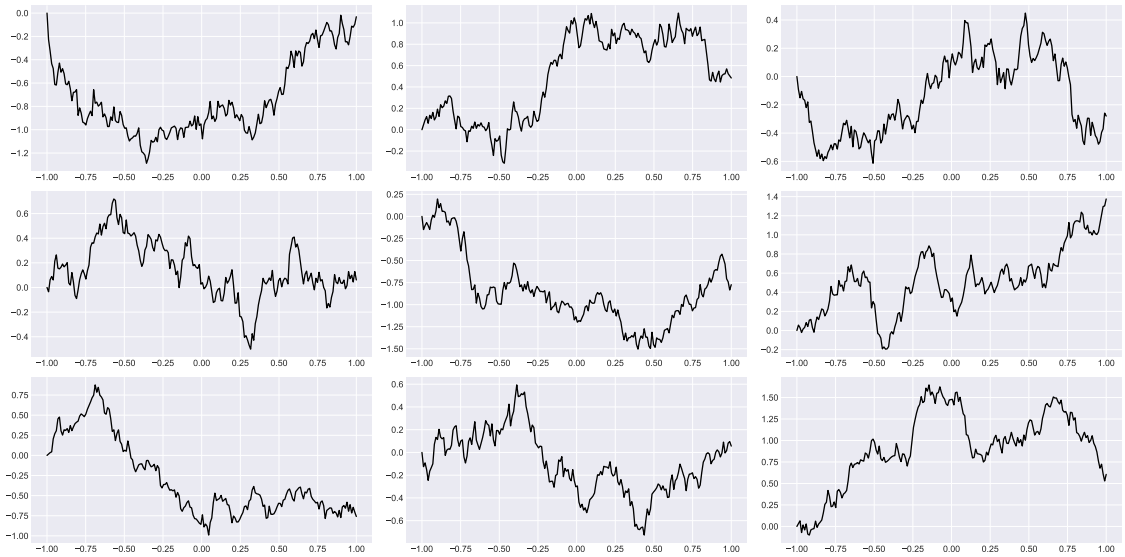

Figure B.3: Samples of paths generated from fractional Brownian motion with Hurst index 0.5

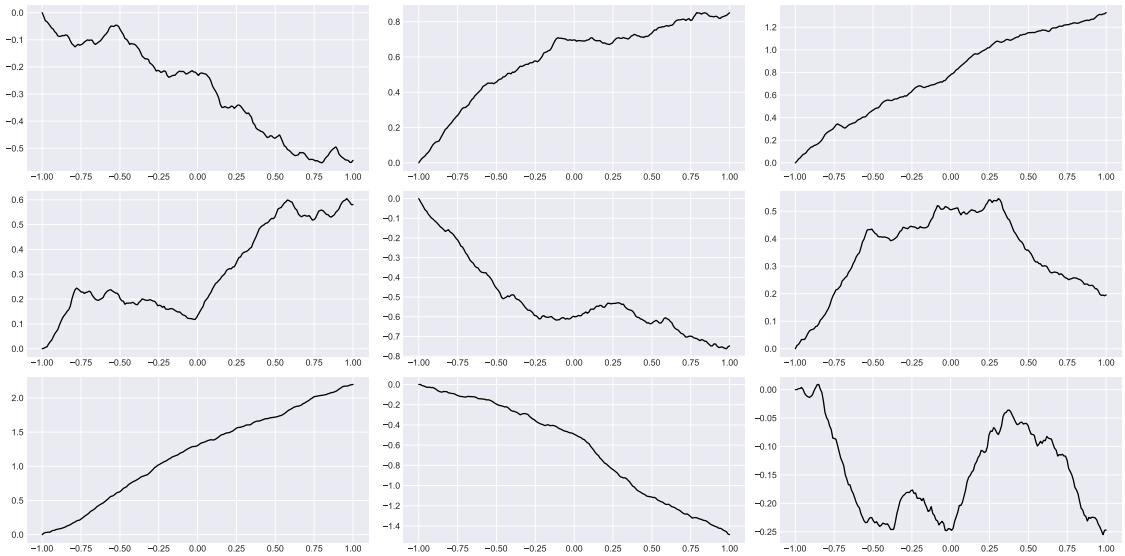

Figure B.4: Samples of paths generated from fractional Brownian motion with Hurst index 0.9