The insertion method to invert the signature of a path

Adeline Fermanian, Jiawei Chang, Terry Lyons and Gérard Biau

Abstract The signature is a representation of a path as an infinite sequence of its iterated integrals. Under certain assumptions, the signature characterizes the path, up to translation and reparameterization. Therefore, a crucial question of interest is the development of efficient algorithms to invert the signature, i.e., to reconstruct the path from the information of its (truncated) signature. In this article, we study the insertion procedure, originally introduced by Chang and Lyons (2019), from both a theoretical and a practical point of view. After describing our version of the method, we give its rate of convergence for piecewise linear paths, accompanied by an implementation in Pytorch. The algorithm is parallelized, meaning that it is very efficient at inverting a batch of signatures simultaneously. Its performance is illustrated with both real-world and simulated examples.

1 Introduction

To any multivariate path, that is, a smooth function $X : [0, 1] \to \mathbb{R}^d$, $d \geqslant 1$, one can associate its signature, denoted by $S(X)$, which is a representation of X as an

Adeline Fermanian

LOPF, Califrais' Machine Learning Lab, Paris, France, e-mail: adeline.fermanian@califrais.fr

Jiawei Chang

Department of Mathematics, University of Oxford, United Kingdom, e-mail: jiawei.chang@hotmail.com

Terry Lyons

Department of Mathematics, University of Oxford, The Alan Turing Institute, United Kingdom, e-mail: terry.lyons@maths.ox.ac.uk

Gérard Biau

Sorbonne Université, CNRS, Laboratoire de Probabilités, Statistique et Modélisation, F-75005 Paris, France, e-mail: gerard.biau@sorbonne-universite.fr

infinite sequence of tensors of iterated integrals. The signature was first introduced by Chen (1957), then was at the heart of rough path theory (Lyons et al., 2007; Friz and Victoir, 2010), and is now widely used in machine learning (Levin et al., 2013; Chevyrev and Kormilitzin, 2016; Kidger et al., 2019; Király and Oberhauser, 2019; Liao et al., 2019; Toth and Oberhauser, 2020; Fermanian, 2021). The combination of signatures and machine learning algorithms has been successfully applied in various domains such as character recognition (Yang et al., 2016; Wilson-Nunn et al., 2018), finance (Lyons et al., 2014; Pérez Arribas, 2018), human action recognition (Li et al., 2017; Yang et al., 2022), medicine (Morrill et al., 2019, 2020), and emotion recognition (Wang et al., 2019).

The signature is a faithful representation of a path as a geometric object, in the sense that if X has at least one monotone coordinate, then $S(X)$ uniquely determines X, up to translation and reparameterization (Hambly and Lyons, 2010). A natural question, therefore, is to design methods and fast algorithms that can invert the signature, i.e., reconstruct X from the information in $S(X)$. This question is of particular interest in the area of time series generation, where time series over a fixed horizon can be seen as paths. The goal is to generate (fake) time series that follow the same distribution as those of a dataset. Similar to the moment-generating function for random variables, the expected signature characterizes distributions of paths (Chevyrev and Oberhauser, 2022), making it a very attractive tool for generative modeling (Ni et al., 2020; Lou et al., 2023). Thus, instead of generating time series directly, a possible strategy is to generate signatures and then invert them to get back to the time domain (Bühler et al., 2020). Moreover, beyond the theoretical interest, we can also think of several other applications of signature inversion such as trend estimation or data smoothing.

Inverting signatures has been an active area of research in recent years, with methods often being theoretically oriented, but practically intractable or slow in their algorithmic execution. For example, Lyons and Xu (2017) provide a procedure for piecewise linear paths, Lyons and Xu (2018) for $\mathscr{C}_1$ paths, and Chang et al. (2017) for monotone paths. Despite their mathematical interest, none of these methods comes with a package that allows them to be implemented and used in a data processing pipeline. For completeness, we also mention that some authors (e.g., Kidger et al., 2019; Bühler et al., 2020) have proposed neural-network-based inversion strategies, but without any theoretical guarantees and at a high computational cost.

In this paper, inspired by the preliminary results of Chang and Lyons (2019), we propose a new, fast, and computationally tractable method, which we call the insertion method. This method exploits the property that signatures of order n and $n + 1$ are connected by inserting the derivatives of the path into a tensor product, which is then integrated. Retrieving the derivatives of a path can then be written as a linear minimization problem, which can be solved very efficiently. We show through a sound theoretical analysis that this approach allows to reconstruct any piecewise linear path with a given rate of convergence, asymptotically in n. Working with piecewise linear paths is by no means a restriction since in practice the signature is always computed by linear interpolation of the measurement points of X. Besides the theoretical guarantees, our method is accompanied by a fast algorithm that, given

the signature of a piecewise linear path of order n, outputs a piecewise linear path of n pieces. This algorithm is implemented in the PyTorch framework (Paszke et al., 2019), and is now part of the Signatory (Kidger and Lyons, 2021) package, freely available to the statistics and machine learning community.

The rest of the paper is organized as follows. In Section 2 we recall some basic definitions and elementary results on signatures and tensors, which are essential for a good understanding of our approach. Section 3 describes the insertion method, gives the main theoretical results, and presents the algorithm. Section 4 shows our method in action through some examples of signature inversion, together with a brief study of the computational cost of the algorithm. For the sake of clarity, most of the technical proofs have been moved to the Supplementary Material.

2 Preliminaries

In this section, we introduce the mathematical framework along with some of the notation used throughout the paper. In all the following, we assume that $\mathbb{R}^d$ is equipped with the Euclidean norm, denoted by $\|\cdot\|$. We refer to Friz and Victoir (2010, Chapter 1) for further details on paths of bounded variation, and to Lyons et al. (2007) for an introduction to signatures and tensors.

2.1 Paths of bounded variation and admissible norms

Definition 1 Let $X : [0, 1] \to \mathbb{R}^d$ be a continuous path. For any $[u, v] \subset [0, 1]$, the total variation (or length) of X on $[u, v]$ is defined by

$$\|X\|_{TV;[u,v]} = \sup_{(u_0,\ldots,u_n)\in D_{u,v}} \sum_{i=1}^{n} \|X_{u_i} - X_{u_{i-1}}\|,$$

where $D_{u,v}$ denotes the set of all finite partitions of $[u, v]$, i.e.,

$$D_{u,v} = \big\{(u_0, \ldots, u_n) \mid n \geqslant 1, \, u = u_0 \leqslant u_1 \leqslant \cdots \leqslant u_{n-1} \leqslant u_n = v\big\}.$$

The path X is said to be of bounded variation on $[u, v]$ if its total variation is finite.

Throughout the document, we denote by $BV(\mathbb{R}^d)$ the set of continuous paths of bounded variation on $[0, 1]$ with values in $\mathbb{R}^d$. Moreover, when $[u, v] = [0, 1]$, we often write $\|X\|_{TV}$ instead of $\|X\|_{TV;[0,1]}$.

Example

Let $X : [0, 1] \to \mathbb{R}^d$ be a continuous piecewise linear path, and let $0 = t_0 < t_1 < \cdots < t_{M-1} < t_M = 1$ be the minimal partition such that X is linear on each $[t_{i-1}, t_i]$.

So, there exist $\alpha_1, \ldots, \alpha_M, \beta_1, \ldots, \beta_M \in \mathbb{R}^d$ such that

$$X_t = \alpha_i + \beta_i t, \quad \text{for } t \in [t_{i-1}, t_i], \quad i \in \{1, \ldots, M\}.$$

One has

$$\|X\|_{TV} = \sum_{i=1}^{M} \|\beta_i\|(t_{i-1} - t_i).$$

For E and F two finite-dimensional real vector spaces, their tensor product is denoted by $E \otimes F$ (see Purbhoo, 2012). It is also a vector space, and, if $(e_i)_{i \in I}$, $(f_j)_{j \in J}$, are bases of E and F respectively, then $(e_i \otimes f_j)_{i \in I, j \in J}$ is a basis of $E \otimes F$. If $\dim(E) = d$, $\dim(F) = e$, then $\dim(E \otimes F) = d \times e$.

For $n \geqslant 0$, the nth tensor power of E is defined as the order n tensor product of E with itself, i.e., $E^{\otimes n} = E \otimes \cdots \otimes E$, with the convention $E^{\otimes 0} = \mathbb{R}$. If $(e_1, \ldots, e_d)$ is a basis of E, then any element a of $E^{\otimes n}$ can be written as

$$a = \sum_{(i_1, \ldots, i_n) \in \{1, \ldots, d\}^n} a_{(i_1, \ldots, i_n)} e_{i_1} \otimes \cdots \otimes e_{i_n}, \qquad a_{(i_1, \ldots, i_n)} \in \mathbb{R}.$$

The space $E^{\otimes n}$ is of dimension d^n, which means that we can identify $E^{\otimes n}$ with $\mathbb{R}^{d^n}$. In particular, $E^{\otimes 2}$ can be identified with the space of $(d \times d)$ matrices.

From now on, it is assumed that $E = \mathbb{R}^d$, where $\mathbb{R}^d$ is equipped with its canonical basis, denoted in this article by $(e_1, \ldots, e_d)$. We want to equip the tensor spaces $(\mathbb{R}^d)^{\otimes n}$, $n \geqslant 1$, with norms inherited from the Euclidean norm on $\mathbb{R}^d$. An essential condition is that these norms should behave "well" with respect to the tensor product, a property summarized by the notion of admissible norms. This important notion is based on the notion of permutation on $(\mathbb{R}^d)^{\otimes n}$, recalled in the definition below.

Definition 2 Let σ be a permutation of $\{1, \ldots, n\}$, i.e., a bijective function of $\{1, \ldots, n\}$ over itself. For any $a \in (\mathbb{R}^d)^{\otimes n}$ of the form

$$a = \sum_{(i_1, \ldots, i_n) \in \{1, \ldots, d\}^n} a_{(i_1, \ldots, i_n)} e_{i_1} \otimes \cdots \otimes e_{i_n},$$

we let

$$\sigma(a) = \sum_{(i_1, \ldots, i_n) \in \{1, \ldots, d\}^n} a_{(i_1, \ldots, i_n)} e_{i_{\sigma(1)}} \otimes \cdots \otimes e_{i_{\sigma(n)}}.$$

Example

Let $n = 2$, $\sigma(1) = 2$, and $\sigma(2) = 1$. Then, for, any $a \in (\mathbb{R}^2)^{\otimes 2}$, one has

$$a = a_{(1,1)} e_1 \otimes e_1 + a_{(1,2)} e_1 \otimes e_2 + a_{(2,1)} e_2 \otimes e_1 + a_{(2,2)} e_2 \otimes e_2,$$

and thus

$$\sigma(a) = a_{(1,1)}e_2 \otimes e_2 + a_{(1,2)}e_2 \otimes e_1 + a_{(2,1)}e_1 \otimes e_2 + a_{(2,2)}e_1 \otimes e_1.$$

Definition 3 (Admissible norms) Assume that, for each $n \geqslant 1$, $(\mathbb{R}^d)^{\otimes n}$ is endowed with a norm $\|\cdot\|_n$. The norms $(\|\cdot\|_n)_{n \geqslant 1}$ are said to be admissible if

(i) For any $n \geqslant 1$, any permutation σ of $\{1, \ldots, n\}$, any $a \in (\mathbb{R}^d)^{\otimes n}$,

$$\|\sigma(a)\|_n = \|a\|_n.$$

(ii) For any $n, m \geqslant 1$, any $a \in (\mathbb{R}^d)^{\otimes n}$, any $b \in (\mathbb{R}^d)^{\otimes m}$,

$$\|a \otimes b\|_{n+m} = \|a\|_n \|b\|_m.$$

Throughout the paper, to keep the vocabulary simple, we refer to "an admissible norm" instead of "a collection of admissible tensor norms" and, since no confusion is possible, we write $\|\cdot\|$ instead of $\|\cdot\|_n$. There are several admissible norms. The most natural one is the Euclidean tensor norm, defined for $a \in (\mathbb{R}^d)^{\otimes n}$ by

$$\|a\| = \Bigg(\sum_{(i_1, \ldots, i_n) \in \{1, \ldots, d\}^n} a^2_{(i_1, \ldots, i_n)} \Bigg)^{1/2}.$$

Another important admissible norm is the projective norm, which is the largest possible admissible norm. Denote by $\|\cdot\|$ any admissible norm, and write any element $a \in (\mathbb{R}^d)^{\otimes n}$ as

$$a = \sum_{i=1}^k a_{1,i} \otimes \cdots \otimes a_{n,i}, \qquad a_{1,i}, \ldots, a_{n,i} \in \mathbb{R}^d. \tag{1}$$

Note that such a representation always exists but is clearly not unique. Then, for statement (ii) in Definition 3 to be satisfied, by the triangle inequality, we must have

$$\|a\| \leqslant \sum_{i=1}^k \|a_{1,i}\| \cdots \|a_{n,i}\|.$$

Since this inequality is valid for any representation of the form (1), we must therefore have

$$\|a\| \leqslant \inf \Bigg\{ \sum_{i=1}^k \|a_{1,i}\| \cdots \|a_{n,i}\| \ \Big| \ a = \sum_{i=1}^k a_{1,i} \otimes \cdots \otimes a_{n,i}, \ k \geqslant 1 \Bigg\}.$$

It is easily shown that

$$\|a\|_\pi = \inf \Bigg\{ \sum_{i=1}^k \|a_{1,i}\| \cdots \|a_{n,i}\| \ \Big| \ a = \sum_{i=1}^k a_{1,i} \otimes \cdots \otimes a_{n,i}, \ k \geqslant 1 \Bigg\}$$

is an admissible norm, called the projective norm. We refer to Ryan (2002) for more details on tensor norms.

Throughout the article, it will be assumed that the tensor powers of $\mathbb{R}^d$ have been equipped with an admissible norm. We close the subsection with the definition of the tensor algebra.

Definition 4 The tensor algebra $T(\mathbb{R}^d)$ is defined by

$$T(\mathbb{R}^d) = \big\{ (a_0, \ldots, a_k, \ldots) \mid \forall k \geqslant 0, a_k \in (\mathbb{R}^d)^{\otimes k} \big\}.$$

For $n \geqslant 0$, the truncated tensor algebra is

$$T^n(\mathbb{R}^d) = \big\{ (a_0, \ldots, a_n) \mid \forall k \in \{0, \ldots, n\}, a_k \in (\mathbb{R}^d)^{\otimes k} \big\}.$$

2.2 The signature of a path

We are now in a position to define the signature of a path of bounded variation and to give some elementary examples.

Definition 5 (Signature of a path) Let $X \in BV(\mathbb{R}^d)$. The signature of X is defined by

$$S(X) = (1, \mathbf{X}^1, \mathbf{X}^2, \ldots, \mathbf{X}^n, \ldots) \in T(\mathbb{R}^d),$$

where, for each $n \geqslant 1$,

$$\mathbf{X}^n = \int \cdots \int_{0 \leqslant u_1 \leqslant \cdots \leqslant u_n \leqslant 1} dX_{u_1} \otimes \cdots \otimes dX_{u_n} \in (\mathbb{R}^d)^{\otimes n}.$$

Recall that when X is a continuously differentiable path, then "$dX_t = X_t' dt$", where X' is the derivative of X with respect to t. However, the integrals in Definition 5 are still well defined for continuous paths of bounded variation using the notion of Riemann-Stieljes integrals (e.g., Lyons et al., 2007).

Example

Assume that $d = 2$ and let $X_t = (X_t^1, X_t^2)$. Since $(\mathbb{R}^2)^{\otimes n}$ can be identified with $\mathbb{R}^{2^n}$, the signatures of order 1 and 2 are equal to

$$\mathbf{X}^1 = \int_0^1 dX_t = \begin{pmatrix} \int_0^1 dX_t^1 \\ \int_0^1 dX_t^2 \end{pmatrix}$$

$$\mathbf{X}^2 = \int_0^1 \int_0^t dX_s \otimes dX_t = \begin{pmatrix} \int_0^1 \int_0^t dX_s^1 dX_t^1 & \int_0^1 \int_0^t dX_s^1 dX_t^2 \\ \int_0^1 \int_0^t dX_s^2 dX_t^1 & \int_0^1 \int_0^t dX_s^2 dX_t^2 \end{pmatrix}.$$

For any $n \geqslant 1$, the truncated signature of order n is denoted by

$$S^n(X) = (1, \mathbf{X}^1, \ldots, \mathbf{X}^n).$$

For a multi-index $I = (i_1, \ldots, i_n) \in \{1, \ldots, d\}^n$, the coefficient of $\mathbf{X}^n$ corresponding to I is

$$S^I(X) = \int_{(u_1,\ldots,u_n)\in\Delta_n} dX_{u_1}^{i_1} \ldots dX_{u_n}^{i_n},$$

where $X_t = (X_t^1, \ldots, X_t^d)$, $t \in [0, 1]$, and

$$\Delta_n = \big\{(u_1, \ldots, u_n) \in [0, 1]^n \mid 0 \leqslant u_1 \leqslant \cdots \leqslant u_n \leqslant 1\big\}.$$

With this notation, the signature of order n takes the form

$$\mathbf{X}^n = \sum_{I \in \{1,\ldots,d\}^n} S^I(X) e_{i_1} \otimes \cdots \otimes e_{i_n}.$$

We will sometimes consider the signature of a path restricted to a certain interval $[u, v] \subset [0, 1]$. Its signature is then denoted by $S_{[u,v]}(X)$ (respectively $\mathbf{X}_{[u,v]}^n$, and so on). Similarly, the simplex in $[u, v]^n$ is denoted by

$$\Delta_{n;[u,v]} = \big\{(u_1, \ldots, u_n) \in [u, v]^n \mid u \leqslant u_1 \leqslant \cdots \leqslant u_n \leqslant v\big\}.$$

Note that the signature $S(X)$ is an element of $T(\mathbb{R}^d)$ and that the truncated signature $S^n(X)$ is an element of $T^n(\mathbb{R}^d)$.

Example

Let $X : [0, 1] \to \mathbb{R}^d$ be a linear path, i.e., $X_t = \alpha + \beta t$, $\alpha, \beta \in \mathbb{R}^d$. Then, for any $n \geqslant 1$, $[u, v] \subset [0, 1]$,

$$\begin{aligned}
\mathbf{X}_{[u,v]}^n &= \int_{\Delta_{n;[u,v]}} dX_{u_1} \otimes \cdots \otimes dX_{u_n} \\
&= \int_{\Delta_{n;[u,v]}} (\beta du_1) \otimes \cdots \otimes (\beta du_n) \\
&= \beta^{\otimes n} \int_{\Delta_{n;[u,v]}} du_1 \ldots du_n = \beta^{\otimes n} \frac{(v-u)^n}{n!}.
\end{aligned} \qquad (2)$$

We now recall without proofs the properties of signatures that will be essential to our method of inversion, and refer to Lyons et al. (2007, Chapter 2) for more material.

Proposition 1

(i) *Let $X \in BV(\mathbb{R}^d)$, $u, v, w \in [0, 1]$ such that $u \leqslant v \leqslant w$. Then*

$$\mathbf{X}^n_{[u,w]} = \sum_{k=0}^{n} \mathbf{X}^k_{[u,v]} \otimes \mathbf{X}^{n-k}_{[v,w]}.$$

(*ii*) *Let* $X \in BV(\mathbb{R}^d)$, $\psi : [0,1] \to [0,1]$ *a reparameterization (i.e., a continuous non-decreasing surjection), and* $\widetilde{X}_t = X_{\psi(t)}$ *for any* $t \in [0,1]$. *Then* $S(\widetilde{X}) = S(X)$.

The first statement of Proposition 1 is known as Chen's identity (Lyons et al., 2007, Theorem 2.9). It provides a procedure for computing signatures of piecewise linear paths. Indeed, starting from the explicit formula for the signature of a linear path given by (2), this identity allows to compute the signature of a concatenation of two linear sections. Iterating this scheme allows to compute the signature of any piecewise linear path. This is the basis of the tools available in Python such as esig, iisignature (Reizenstein and Graham, 2020), and Signatory (Kidger and Lyons, 2021).

Since we are interested in reconstructing a path from its signature, we need to discuss what it means for two paths to have the same signature. It is clear from their definition that signatures are invariant by translation, and according to Proposition 1, (*ii*) they are also invariant by reparameterization. The appropriate notion to encapsulate how signatures characterize paths is the so-called property of tree-like equivalence, introduced by Hambly and Lyons (2010) and defined as follows. The notation $\overleftarrow{X}$ denotes the time-reversal of a path X, defined, for any $t \in [0,1]$, by $\overleftarrow{X}_t = X_{1-t}$.

Definition 6 (Tree-like equivalence)

(*i*) A path $X \in BV(\mathbb{R}^d)$ is tree-like if there exists a continuous function $h : [0,1] \to [0,+\infty)$ such that $h(0) = h(1) = 0$ and, for any $[s,t] \subset [0,1]$,

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

(*ii*) Two paths $X, Y \in BV(\mathbb{R}^d)$ are tree-like equivalent if $X * \overleftarrow{Y}$ is a tree-like path. This relation is denoted by $X \sim Y$.

Informally, a tree-like path is a path that retraces itself backward. This concept is important insofar as it is involved in the following uniqueness theorem, due to Hambly and Lyons (2010, Theorem 4, Corollary 1.6).

Theorem 1 (Uniqueness) *Let* $X, Y \in BV(\mathbb{R}^d)$. *Then* $S(X) = S(Y)$ *if and only if* $X \sim Y$. *Moreover, for any* $X \in BV(\mathbb{R}^d)$, *there exists a unique path (up to translation and reparameterization) of minimal length in its equivalence class, denoted by* $\overline{X}$ *and called the reduced path.*

Theorem 1 can be summarized as follows: the signature loses the information about the initial position of the path, of the time parameterization, and of any tree-like piece. We conclude this subsection with Proposition 2, which gives an upper bound on the size of signature coefficients.

Proposition 2 *Let $X \in BV(\mathbb{R}^d)$, $[u, v] \subset [0, 1]$. Then, for any admissible tensor norm and $n \geqslant 1$, one has*

$$\|\mathbf{X}^n_{[u,v]}\| \leqslant \frac{\|X\|^n_{TV;[u,v]}}{n!}.$$

3 The insertion method

3.1 Description

We are now ready to describe the insertion method and its associated algorithm. Suppose that we are given a piecewise linear path $X : [0, 1] \to \mathbb{R}^d$ as in Subsection 2.1. Recall that we denote by

$$0 = t_0 < t_1 < \cdots < t_{M-1} < t_M = 1$$

the minimal partition such that X is linear on each $[t_{i-1}, t_i]$. Thus, there exist intercepts $\alpha_1, \ldots, \alpha_M \in \mathbb{R}^d$ and non-zero slopes $\beta_1, \ldots, \beta_M \in \mathbb{R}^d$ such that

$$X_t = \alpha_i + \beta_i t, \quad \text{for } t \in [t_{i-1}, t_i], \quad i \in \{1, \ldots, M\}. \tag{3}$$

For each $i \in \{1, \ldots, M - 1\}$, we let ω_i be the angle between the linear segments $[t_{i-1}, t_i]$ and $[t_i, t_{i+1}]$, i.e.,

$$\omega_i = \text{Arccos}\left(\frac{\langle \beta_i, \beta_{i+1} \rangle}{\|\beta_i\| \|\beta_{i+1}\|}\right) \in [0, \pi],$$

where $\langle \cdot, \cdot \rangle$ is the Euclidean scalar product and Arccos is the arccosine function. Assuming that the partition is minimal means that $\omega_i \neq 0$. In addition, we assume that X is the reduced path in its equivalence class, for the tree-like equivalence (see Theorem 1), which means that $\omega_i \neq \pi$. (If $\omega_i = \pi$, then segment $i + 1$ retraces backwards segment i and is therefore a tree-like part of the path, which is excluded if X is reduced.) In summary, we have the following assumption:

$$(A_1) \quad \forall i \in \{1, \ldots, M\}, \quad \omega_i \in]0, \pi[.$$

We denote by $\omega = \min_i \omega_i$ the smallest angle between two linear segments. Invariance by translation means that we cannot recover α_0 from $S(X)$. Due to the continuity of X, the α_i for $i \in \{1, \ldots, M\}$ are completely determined by $\alpha_0, \beta_1, \ldots, \beta_M$. Therefore, the only information we hope to recover is the information on the slopes β_i, up to the reparameterization of the path. So we choose a specific parameterization of X. Let $\ell = \|X\|_{TV}$ be the length of X, assumed to be strictly positive, and let ϕ be the reparameterization defined as

$$\phi : [0, 1] \to [0, 1]$$

$$t \mapsto \frac{\|X\|_{TV;[0,t]}}{\ell}.$$

The function ϕ is strictly increasing and continuous piecewise linear. In fact, if $t \in [t_{i-1}, t_i]$, then

$$\phi(t) = \ell^{-1} \sum_{k=1}^{i-1} \|\beta_k\| (t_k - t_{k-1}) + \ell^{-1} \|\beta_i\| (t - t_{i-1}).$$

The path $\overline{X}_t = X_{\phi^{-1}(t)}$ is therefore a reparameterization of X, and $S(X) = S(\overline{X})$. More precisely, $\overline{X}$ is piecewise linear on the partition $0 = u_0 < u_1 < \cdots < u_M = 1$, where

$$u_i = \ell^{-1} \sum_{k=1}^{i} \|\beta_k\| (t_k - t_{k-1}).$$

To see this, just note that for each $i \in \{1, \ldots, M\}$ and $u \in [u_{i-1}, u_i]$, one has

$$\phi^{-1}(u) = \frac{\ell}{\|\beta_i\|} (u - u_{i-1}) + t_{i-1},$$

and

$$\overline{X}_u = \alpha_i + \beta_i \phi^{-1}(u) = \alpha_i + \frac{\ell}{\|\beta_i\|} \beta_i (u - u_{i-1}) + \beta_i t_{i-1}.$$

The path $\overline{X}$ is thus piecewise linear and its slopes have a constant norm, equal to ℓ. Note that recovering X up to translation and reparameterizations amounts to recovering the direction of each linear segment, i.e., $\beta_1/\|\beta_1\|, \ldots, \beta_M/\|\beta_M\|$. So, from now on, we will assume that X is piecewise linear on a partition $t_0 = 0 < t_1 < \cdots < t_M = 1$ and that $\|\beta_i\| = \ell$ for each $i \in \{1, \ldots, M\}$:

$$(A_2) \quad \forall i \in \{1, \ldots, M\}, \quad \|\beta_i\| = \ell.$$

Throughout, to avoid degenerate situations, it is assumed that for each $n \geqslant 1$, $\|\mathbf{X}^n\| \neq 0$. There is no harm in doing so since according to Boedihardjo and Geng (2019) this assumption is satisfied for n large enough when X is not tree-like, which is our case under Assumption (A_1). The general idea of the insertion method is as follows. Signatures of order n and $n + 1$ are connected by inserting the derivatives of the path into a tensor product, which is then integrated. Therefore, retrieving the derivative of the path boils down to finding a vector in $\mathbb{R}^d$ such that, once inserted into the tensor product of the signature of order n, minimizes the distance between that modified signature and the signature of order $n + 1$. This can be written as a linear minimization problem, which can be solved very efficiently. It allows to retrieve the slopes β_i of the path at different intervals and thus to reconstruct the path by integration.

The first ingredient of the method is the insertion map. It is a linear map defined as a tensor product between a vector $y \in \mathbb{R}^d$ and the signature of X of order n along the dimension p.

Definition 7 (Insertion map) Let $X \in BV(\mathbb{R}^d)$, $n \geqslant 1$, and $p \in \{1, \ldots, n+1\}$. The insertion map $\mathscr{L}_{p,X}^n : \mathbb{R}^d \to (\mathbb{R}^d)^{\otimes(n+1)}$ is defined, for any $y \in \mathbb{R}^d$, by

$$[p = 1] \quad \mathscr{L}_{1,X}^n(y) = \int_{\Delta_n} y \otimes dX_{u_1} \otimes \cdots \otimes dX_{u_n},$$

$$[2 \leqslant p < n] \quad \mathscr{L}_{p,X}^n(y) = \int_{\Delta_n} dX_{u_1} \otimes \cdots \otimes dX_{u_{p-1}} \otimes y \otimes dX_{u_p} \otimes \cdots \otimes dX_{u_n},$$

$$[p = n + 1] \quad \mathscr{L}_{n+1,X}^n(y) = \int_{\Delta_n} dX_{u_1} \otimes \cdots \otimes dX_{u_n} \otimes y.$$

A first observation is that the map $\mathscr{L}_{p,X}^n$ is linear. Moreover, it depends only on the signature of X of order n and is Lipschitz continuous for any admissible tensor norm (see Definition 3). These properties are summarized in the next proposition.

Proposition 3 *Let $X \in BV(\mathbb{R}^d)$. For any $n \geqslant 1$, $p \in \{1, \ldots, n+1\}$, $\mathscr{L}_{p,X}^n$ is linear, Lipschitz-continuous, and its Lipschitz constant is $\|\mathbf{X}^n\|$.*

Proof The linearity is an immediate consequence of the linearity of the tensor product. Moreover, for $y, z \in \mathbb{R}^d$, we have

$$\|\mathscr{L}_{p,X}^n(y) - \mathscr{L}_{p,X}^n(z)\|$$

$$= \Big\| \int_{\Delta_n} dX_{u_1} \otimes \cdots \otimes dX_{u_{p-1}} \otimes (y - z) \otimes dX_{u_p} \otimes \cdots \otimes dX_{u_n} \Big\|$$

$$= \Big\| \int_{\Delta_n} dX_{u_1} \otimes \cdots \otimes dX_{u_n} \otimes (y - z) \Big\| \qquad \text{(by Definition 3, } (i))$$

$$= \Big\| \int_{\Delta_n} dX_{u_1} \otimes \cdots \otimes dX_{u_n} \Big\| \cdot \|y - z\| \qquad \text{(by Definition 3, } (ii))$$

$$= \|\mathbf{X}^n\| \cdot \|y - z\|.$$

$$\square$$

The general idea of the insertion method is to solve the following optimization problem:

$$y_{p,n}^* \in \arg\min_{y \in \mathbb{R}^d} \|\mathscr{L}_{p,X}^n(y) - (n+1)\mathbf{X}^{n+1}\|. \tag{4}$$

The main result of the section is that the minimizer of (4) is close to β_i for the projective norm, provided the parameter $p/(n+1)$ is well-chosen inside $]t_{i-1}, t_i[$.

Theorem 2 *Let X be a piecewise linear path as defined by (3), following assumptions (A_1) and (A_2). For any $n \geqslant 1$, $p \in \{1, \ldots, n+1\}$, define*

$$y^*_{p,n} \in \arg\min_{y \in \mathbb{R}^d} \|\mathscr{L}^n_{p,X}(y) - (n+1)\mathbf{X}^{n+1}\|_\pi.$$

Let

$$K(\omega) = \log\left(\frac{2}{1 - \cos(\omega/2)}\right) \quad and \quad n_1 = \lfloor 4e^{2(M-1)K(\omega)}\rfloor.$$

Then, for any $i \in \{1, \ldots, M\}$, $n > \max(n_1, 2/(t_i - t_{i-1}))$, and $p = \lfloor (3t_i + t_{i-1})(n+1)/4\rfloor$, there exists an integer $k_n \in [n - n^{3/4}, n + n^{3/4}]$ such that

$$\|y^*_{p,k_n} - \beta_i\| \leqslant 4\ell e^{(M-1)K(\omega)}\left(\frac{1}{\sqrt{k_n + 1}}\sqrt{\frac{1 - (t_i - t_{i-1})}{t_i - t_{i-1}}} + 4\exp\left(-\frac{k_n}{16}(t_i - t_{i-1})^2\right)\right).$$

Since the intervals $[n - n^{3/4}, n + n^{3/4}]$ shift towards infinity with a width that grows slower than n, we have the following corollary.

Corollary 1 *For any $i \in \{1, \ldots, M\}$, there exists a strictly increasing sequence of integers (k_n) and a sequence of positive integers (p_n) such that $y^*_{p_n,k_n}$ converges to β_i as n increases.*

Let us comment on Theorem 2. To simplify the interpretation, consider the case of a regular partition, where $t_i = i/M$. Then, Theorem 2 becomes: for any $n > \max(n_1, 2M)$ and $p = (2M-1)(n+1)/4M$,

$$\|y^*_{p,k_n} - \beta_i\| \leqslant 4\ell e^{(M-1)K(\omega)}\left(\sqrt{\frac{M-1}{k_n + 1}} + 4\exp\left(-\frac{k_n}{16M^2}\right)\right). \tag{5}$$

As the number M of piecewise linear pieces increases (or, equivalently, the length of the interval $t_i - t_{i-1}$ decreases), it becomes harder to recover β_i—both the right-hand-side of (5) and the truncation order n necessary for the result to hold increase.

Theorem 2 relies heavily on the following two propositions, the proofs of which are given in the Supplementary Material.

Proposition 4 *Let X be a piecewise linear path as defined by (3), following assumptions (A_1) and (A_2). For any $i \in \{1, \ldots, M\}$, $n \geqslant 2/(t_i - t_{i-1})$, and $p = \lfloor (3t_i + t_{i-1})(n+1)/4\rfloor$, one has*

$$\|\mathscr{L}^n_{p,X}(\beta_i) - (n+1)\mathbf{X}^{n+1}\|$$
$$\leqslant \frac{\ell^{n+1}}{n!}\left(\frac{1}{\sqrt{n+1}}\sqrt{\frac{1 - (t_i - t_{i-1})}{t_i - t_{i-1}}} + 4\exp\left(-\frac{n}{16}(t_i - t_{i-1})^2\right)\right).$$

Proposition 5 *Let X be a piecewise linear path as defined by (3), following assumptions (A_1) and (A_2). For any $n > n_1$, there exists an integer $k_n \in [n - n^{3/4}, n + n^{3/4}]$ such that*

$$\|\mathbf{X}^{k_n}\|_\pi \geqslant \ell^{k_n}\frac{e^{-(M-1)K(\omega)}}{2k_n!},$$

where $\|\cdot\|_\pi$ is the projective norm.

Proof (Theorem 2) The identity of Proposition 3 is true for every admissible norm so it is especially true for the projective norm $\| \cdot \|_\pi$. Thus, for any $n \geqslant 1$,

$$\|\mathscr{L}^n_{p,X}(y^*_{p,n}) - \mathscr{L}^n_{p,X}(\beta_i)\|_\pi = \|\mathbf{X}^n\|_\pi \, \|y^*_{p,n} - \beta_i\|.$$

Proposition 5 guarantees the existence of an integer $k_n \in [n - n^{3/4}, n + n^{3/4}]$ such that

$$\|\mathbf{X}^{k_n}\|_\pi \geqslant \ell^{k_n} \frac{e^{-(M-1)K(\omega)}}{2k_n!}.$$

Moreover, Proposition 4 (which is valid for any tensor norm) yields

$$\|\mathscr{L}^{k_n}_{p,X}(\beta_i) - (k_n + 1)\,\mathbf{X}^{k_n+1}\|_\pi$$
$$\leqslant \frac{\ell^{k_n+1}}{k_n!}\left(\frac{1}{\sqrt{k_n + 1}}\sqrt{\frac{1 - (t_i - t_{i-1})}{t_i - t_{i-1}}} + 4\exp\left(-\frac{k_n}{16}(t_i - t_{i-1})^2\right)\right).$$

By the very definition of $y^*_{p,n}$, we have

$$\|\mathscr{L}^{k_n}_{p,X}(y^*_{p,k_n}) - (k_n + 1)\,\mathbf{X}^{k_n+1}\|_\pi \leqslant \|\mathscr{L}^{k_n}_{p,X}(\beta_i) - (k_n + 1)\,\mathbf{X}^{k_n+1}\|_\pi.$$

Combining these inequalities, we obtain

$$\|y^*_{p,k_n} - \beta_i\| = \frac{1}{\|\mathbf{X}^{k_n}\|_\pi}\|\mathscr{L}^{k_n}_{p,X}(y^*_{p,k_n}) - \mathscr{L}^{k_n}_{p,X}(\beta_i)\|_\pi$$
$$\leqslant \frac{1}{\|\mathbf{X}^{k_n}\|_\pi}\big(\|\mathscr{L}^{k_n}_{p,X}(y^*_{p,k_n}) - (k_n + 1)\,\mathbf{X}^{k_n+1}\|_\pi$$
$$+ \|(k_n + 1)\,\mathbf{X}^{k_n+1} - \mathscr{L}^{k_n}_{p,X}(\beta_i)\|_\pi\big)$$
$$\leqslant \frac{2}{\|\mathbf{X}^{k_n}\|_\pi}\|\mathscr{L}^{k_n}_{p,X}(\beta_i) - (k_n + 1)\,\mathbf{X}^{k_n+1}\|_\pi$$
$$\leqslant 4\ell e^{(M-1)K(\omega)}\left(\frac{1}{\sqrt{k_n + 1}}\sqrt{\frac{1 - (t_i - t_{i-1})}{t_i - t_{i-1}}}\right.$$
$$\left. + 4\exp\left(-\frac{k_n}{16}(t_i - t_{i-1})^2\right)\right).$$

$\square$

Remark 1 Note that Theorem 2 holds for the projective norm. An extension of this theorem to other admissible norms is possible, provided a lower bound similar to Proposition 5 is available. However, the existence of such a bound is still a conjecture (see Chang et al., 2018). On the other hand, computing the projective norm is computationally expensive, so in practice it is much better to solve (4) with the Euclidean tensor norm, as we will do in the next subsection.

3.2 Algorithm

The last step necessary to have a complete algorithm is to solve (4) with the choice of the Euclidean tensor norm. Let $A_p \in \mathbb{R}^{d^{n+1} \times d}$ be the matrix representing the linear map $\mathscr{L}_{p,X}^n(\cdot)$ in the canonical basis of $\mathbb{R}^d$, where $(\mathbb{R}^d)^{\otimes n+1}$ is identified with $\mathbb{R}^{d^{n+1}}$. Thus, for any $y \in \mathbb{R}^d$, $\mathscr{L}_{p,X}^n(y) = A_p y$. The following lemma, which is essential for finding an explicit solution to (4), gives information about the form of the matrix $A_p^\top A_p$.

Lemma 1 *For any $p \in \{1, \ldots, n+1\}$, one has $A_p^\top A_p = \|\mathbf{X}^n\|^2 I_d$, where I_d is the identity matrix.*

Lemma 1, proved in the Supplementary Material, allows us to obtain an explicit solution to Problem (4).

Proposition 6 *For any $p \in \{1, \ldots, n+1\}$, Problem (4) with the Euclidean tensor norm has a unique solution, equal to*

$$y_{p,n}^* = (n+1) \frac{A_p^\top \mathbf{X}^{n+1}}{\|\mathbf{X}^n\|^2}.$$

Proof Problem (4) takes the form

$$y_{p,n}^* \in \arg\min_{y \in \mathbb{R}^d} \|A_p y - (n+1)\mathbf{X}^{n+1}\|^2.$$

This is exactly the same minimization problem than in least-squares regression. Therefore,

$$y_{p,n}^* = (n+1)(A_p^\top A_p)^{-1} A_p^\top \mathbf{X}^{n+1}.$$

By Lemma 1, $A_p^\top A_p = \|\mathbf{X}^n\|^2 I_d$, and thus

$$y_{p,n}^* = (n+1) \frac{A_p^\top \mathbf{X}^{n+1}}{\|\mathbf{X}^n\|^2}.$$

$\square$

The insertion algorithm is presented in Algorithm 1. Given a signature of order $n \geqslant 2$, the finest piecewise approximation we can get is to let p vary in $\{1, \ldots, n\}$. For each such p, we solve (4), which gives an approximation of the slope $y_{p,n}^*$ on $[(p-1)/n, p/n]$. We then integrate to obtain a piecewise-linear path $\tilde{X}$. Note that since we cannot recover the initial position of the path, it is provided by the user.

It is important to note that Theorem 2 applies only to piecewise linear paths, but this is not a restriction per se, and Algorithm 1 is in fact applicable to any continuous path of bounded variation. To understand this remark, one only has to remember that in practice the path X is never observed continuously, but only at

---

Algorithm 1: Insertion algorithm

Data: d: dimension of the underlying path; n: truncation order of the signature; $\mathbf{X}^n$ and $\mathbf{X}^{n+1}$: signature of order n; X_0: starting point of the path.

Result: $\widetilde{X} = (X_0, \widetilde{X}_{1/n}, \ldots, \widetilde{X}_1) \in \mathbb{R}^{(n+1)\times d}$: array of the $(n+1)$ positions in $\mathbb{R}^d$ of the reconstructed path.

1 Extract $\mathbf{X}^n$ and $\mathbf{X}^{n-1}$, the terms of order n and $n-1$ of $S^n(X)$.

2 $\widetilde{X}_0 = X_0$

3 **for** $p \in \{1, \ldots, n\}$ **do**

4 $\quad$ Compute A_p, the matrix representing the linear map $\mathscr{L}^{n-1}_{p,X}(\cdot)$, function of $\mathbf{X}^{n-1}$.

5 $\quad y^*_{p,n} = n \frac{A_p^\top \mathbf{X}^n}{\|\mathbf{X}^{n-1}\|^2}$

6 $\quad \widetilde{X}_{p/n} = \widetilde{X}_{p-1/n} + \frac{y^*_{p,n}}{n}$

---

discrete-time points $t_1, \ldots, t_M$, $0 \leqslant t_1 < \cdots < t_M \leqslant 1$, with values $X_{t_1}, \ldots, X_{t_M}$. The signature computed by the softwares is not the signature of X, but the signature of the piecewise linear path obtained by linear interpolation of $X_{t_1}, \ldots, X_{t_M}$. This operation is performed sequentially using Chen's identity, as explained in the remark after Proposition 1, with a complexity $O(Md^n)$. Thus, Algorithm 1 should rather be understood as a method to approximate any continuous path by a linear path composed of n pieces, with two signatures close to each other. Note that it can be shown that the signatures of a given path and of its piecewise linear approximation converge as the discretization gets thinner (see, e.g., Bleistein et al., 2023, Lemma 3.3), which theoretically justifies working with piecewise linear approximations.

A natural question is at what order n the path should be reconstructed. If n is chosen too small, the quality of the approximation is of course likely to be poor, but if n is too large, we run into memory problems, since the size of the signature is of the order of d^n. In practice, n is chosen to be as large as possible, given the memory constraints. Finally, Algorithm 1 is computationally cheap, since the most expensive operation inside the loop is the matrix multiplication $A_p^\top \mathbf{X}^n$ (of the order $O(d^{n+1})$). Moreover, the algorithm can be easily parallelized to allow it to take as input a batch of N signatures, each of order n, and thus of length $(d^{n+1}-1)/(d-1)$ for $d \geqslant 1$. In this parallel context, the data has the form of an array of size $\mathbb{R}^{N\times(d^{n+1}-1)/(d-1)}$ and the algorithm outputs a tensor of multiple paths of size $\mathbb{R}^{N\times(n+1)\times d}$. We refer the reader to the code that we have included in the Signatory package (Kidger and Lyons, 2021) for more details.

4 Experimental results

Examples of inversion

We show in Figures 1, 2, 3, and 4 some examples of reconstruction for paths in dimension 2 and 3. Figures 1 and 2 are simulations of a half circle and a spiral,
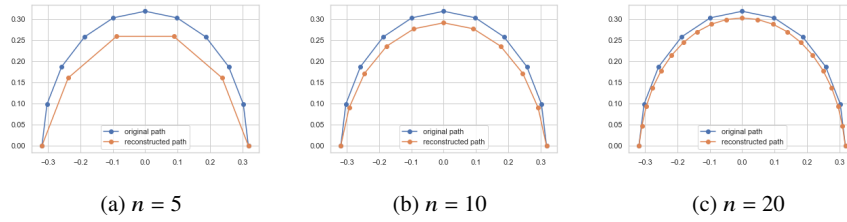
(a) $n = 5$ (b) $n = 10$ (c) $n = 20$

Fig. 1: Signature inversion of a half circle.

(a) $n = 5$ (b) $n = 10$ (c) $n = 15$
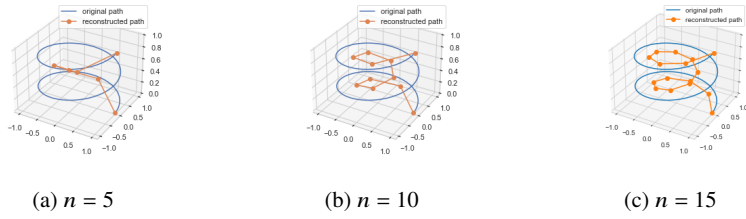
Fig. 2: Signature inversion of a spiral in 3d.

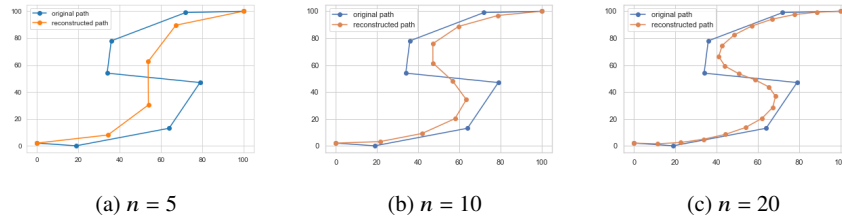(a) $n = 5$ (b) $n = 10$ (c) $n = 20$

Fig. 3: Signature inversion of an example of the class "5" from the Pendigits dataset.

while Figures 3 and 4 are two examples of the Pendigits dataset from the UCI Machine Learning Repository (Dua and Graff, 2019). In these four figures, we have knowledge of the "true" path, in blue, from which we compute its truncated signature with three different values of n. We then invert this signature using Algorithm 1 and plot the reconstructed path in orange. This way we can compare the quality of the reconstructed path with the original one. We observe that in all these experiments, the reconstructed paths are smooth and close to the original ones for $n = 20$.

Running times

Figure 5 shows the running time of Algorithm 1 when several hyperparameters vary. The paths are randomly generated as piecewise linear paths with $M = 10$ pieces,
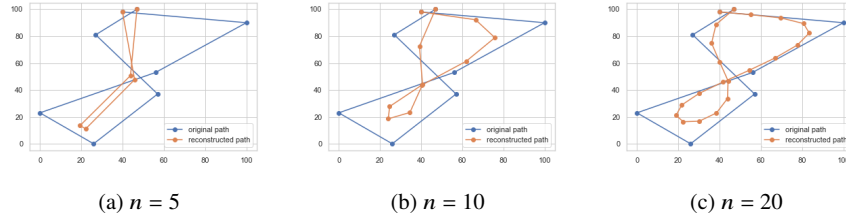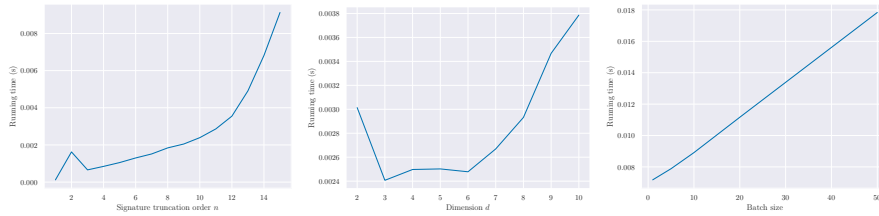
(a) $n = 5$ (b) $n = 10$ (c) $n = 20$

Fig. 4: Signature inversion of an example of the class "8" from the Pendigits dataset.

(a) Running time when n varies. (b) Running time when d varies. (c) Running time when N varies.

Fig. 5: Running time in seconds to invert the signature of several paths, when we let some hyperparameters vary: the truncation order of the signature n, the dimension of the path d, and the number of signatures that are inverted.

starting at 0. The endpoint of each linear piece is generated uniformly at random in $[0, 1]^d$. The experiments were performed on a standard laptop computer.

In Figure 5a, we invert one path in dimension $d = 2$ for different values of n. In Figure 5b, we set $n = 4$ and let d vary. In Figure 5c, we set $n = 10$ and $d = 2$, and let the number of paths N vary. We see that as n increases, the running time increases exponentially, while the dependence on the dimension d is polynomial, and the dependence on the number of paths N is linear. In total, we can invert 50 signatures truncated at order 10 in about 0.02s.

Trend estimation

One interesting application of Algorithm 1 is to perform nonparametric trend estimation. In fact, if we compute the signature of a highly-sampled time series and invert it, we get an approximation of the original path sampled at a much lower frequency. In Figure 6, we show some results on both simulated and real-world financial data of this approach. We can see that the signatures are able to capture the overall trend of the curve. In this case, the choice of truncation order n corresponds to the degree of smoothness desired for the trend, which can then be adapted to the application at hand. An interesting extension of our work would be to study both theoretically and empirically the resulting estimator of the trend, and compare it with traditional
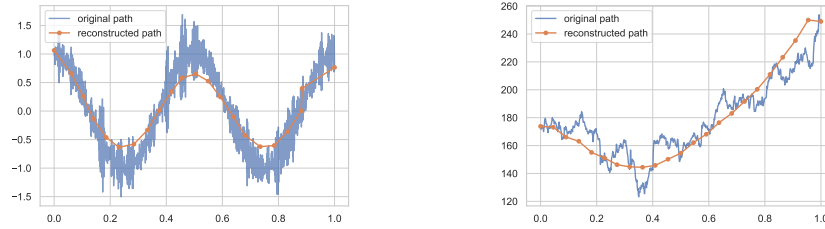
(a) Simulation of a path as a cosine function plus an AR(1) noise.

(b) Daily open value of the FDX between 2015 and 2017.

Fig. 6: Examples of trend estimation: the signature of the original path in blue is computed with a truncation order of $n = 21$, then it is inverted, which gives the orange reconstructed path.

approaches such as running average. We can expect the signature-based estimator to be less sensitive to outliers and noise.

References

L. Bleistein, A. Fermanian, A.-S. Jannot, and A. Guilloux. Learning the dynamics of sparsely observed interacting systems. In A. Krause, E. Brunskill, K. Cho, B. Engelhardt, S. Sabato, and J. Scarlett, editors, *Proceedings of the 40th International Conference on Machine Learning*, volume 202, pages 2603–2640. PMLR, 2023.

H. Boedihardjo and X. Geng. A non-vanishing property for the signature of a path. *Comptes Rendus Mathématique*, 357:120–129, 2019.

H. Bühler, B. Horvath, T. Lyons, I. Pérez Arribas, and B. Wood. A data-driven market simulator for small data environments. *arXiv:2006.14498*, 2020.

J.W. Cannon, W.J. Floyd, R. Kenyon, and W.R. Parry. Hyperbolic geometry. In S. Levy, editor, *Flavors of Geometry*, volume 31, pages 59–115. MSRI Publications, 1997.

J. Chang and T. Lyons. Insertion algorithm for inverting the signature of a path. *arXiv:1907.08423*, 2019.

J. Chang, N. Duffield, H. Ni, and W. Xu. Signature inversion for monotone paths. *Electronic Communications in Probability*, 22:1–11, 2017.

J. Chang, T. Lyons, and H. Ni. Super-multiplicativity and a lower bound for the decay of the signature of a path of finite length. *Comptes Rendus Mathématique*, 356:720–724, 2018.

K.-T. Chen. Integration of paths, geometric invariants and a generalized Baker-Hausdorff formula. *Annals of Mathematics*, 65:163–178, 1957.

I. Chevyrev and A. Kormilitzin. A primer on the signature method in machine learning. *arXiv:1603.03788*, 2016.

I. Chevyrev and H. Oberhauser. Signature moments to characterize laws of stochastic processes. *Journal of Machine Learning Research*, 23(176):1–42, 2022.

D. Dua and C. Graff. UCI Machine Learning Repository. University of California, Irvine, School of Information and Computer Sciences, 2019. URL http://archive.ics.uci.edu/ml.

A. Fermanian. Embedding and learning with signatures. *Computational Statistics & Data Analysis*, 157:107148, 2021.

P.K. Friz and N.B. Victoir. *Multidimensional Stochastic Processes as Rough Paths: Theory and Applications*, volume 120 of *Cambridge Studies in Advanced Mathematics*. Cambridge University Press, Cambridge, 2010.

B. Hambly and T. Lyons. Uniqueness for the signature of a path of bounded variation and the reduced path group. *Annals of Mathematics*, 171:109–167, 2010.

P. Kidger and T. Lyons. Signatory: Differentiable computations of the signature and logsignature transforms, on both CPU and GPU. In *International Conference on Learning Representations*, 2021.

P. Kidger, P. Bonnier, I. Pérez Arribas, C. Salvi, and T. Lyons. Deep signature transforms. In H. Wallach, H. Larochelle, A. Beygelzimer, F. d'Alché-Buc, E. Fox, and R. Garnett, editors, *Advances in Neural Information Processing Systems*, volume 32, pages 3099–3109. Curran Associates, Inc., 2019.

F.J. Király and H. Oberhauser. Kernels for sequentially ordered data. *Journal of Machine Learning Research*, 20(31):1–45, 2019.

D. Levin, T. Lyons, and H. Ni. Learning from the past, predicting the statistics for the future, learning an evolving system. *arXiv:1309.0260*, 2013.

C. Li, X. Zhang, and L. Jin. LPSNet: A novel log path signature feature based hand gesture recognition framework. In *2017 IEEE International Conference on Computer Vision Workshop*, pages 631–639. IEEE Computer Society, 2017.

S. Liao, T. Lyons, W. Yang, and H. Ni. Learning stochastic differential equations using RNN with log signature features. *arXiv:1908.08286*, 2019.

H. Lou, S. Li, and H. Ni. PCF-GAN: Generating sequential data via the characteristic function of measures on the path space. *arXiv:2305.12511*, 2023.

B. Loustau. Hyperbolic Geometry. *arXiv:2003.11180*, 2020.

T. Lyons and W. Xu. Inverting the signature of a path. *Journal of the European Mathematical Society*, 20:1655–1687, 2018.

T. Lyons, H. Ni, and H. Oberhauser. A feature set for streams and an application to high-frequency financial tick data. In *BigDataScience '14: Proceedings of the*

2014 International Conference on Big Data Science and Computing, pages 1–8. Association for Computing Machinery, 2014.

T.J. Lyons and W. Xu. Hyperbolic development and inversion of signature. *Journal of Functional Analysis*, 272:2933–2955, 2017.

T.J. Lyons, M.J. Caruana, and T. Lévy. *Differential Equations Driven by Rough Paths*, volume 1908 of *Lecture Notes in Mathematics*. Springer, Berlin, 2007.

J. Morrill, A. Kormilitzin, A.J. Nevado-Holgado, S. Swaminathan, S. Howison, and T. Lyons. The signature-based model for early detection of sepsis from electronic health records in the intensive care unit. In *2019 Computing in Cardiology*, pages 1–4, 2019.

J.H. Morrill, A. Kormilitzin, A.J. Nevado-Holgado, S. Swaminathan, S.D. Howison, and T.J. Lyons. Utilization of the signature method to identify the early onset of sepsis from multivariate physiological time series in critical care monitoring. *Critical Care Medicine*, 48:e976–e981, 2020.

H. Ni, L. Szpruch, M. Wiese, S. Liao, and B. Xiao. Conditional Sig-Wasserstein GANs for time series generation. *arXiv:2006.05421*, 2020.

A. Paszke, S. Gross, F. Massa, A. Lerer, J. Bradbury, G. Chanan, T. Killeen, Z. Lin, N. Gimelshein, L. Antiga, A. Desmaison, A. Kopf, E. Yang, Z. DeVito, M. Raison, A. Tejani, S. Chilamkurthy, B. Steiner, L. Fang, J. Bai, and S. Chintala. PyTorch: An imperative style, high-performance deep learning library. In H. Wallach, H. Larochelle, A. Beygelzimer, F. d'Alché-Buc, E. Fox, and R. Garnett, editors, *Advances in Neural Information Processing Systems*, volume 32, pages 8026–8037. Curran Associates, Inc., 2019.

J. Paupert. Introduction to Hyperbolic Geometry. *Arizona State University, School of Mathematical and Statistical Sciences*, 2016.

K. Purbhoo. Notes on Tensor Products and the Exterior Algebra. 2012.

I. Pérez Arribas. Derivatives pricing using signature payoffs. *arXiv:1809.09466*, 2018.

J.F. Reizenstein and B. Graham. Algorithm 1004: The iisignature library: Efficient calculation of iterated-integral signatures and log signatures. *ACM Transactions on Mathematical Software*, 46:1–21, 2020.

R.A. Ryan. *Introduction to Tensor Products of Banach Spaces*. Springer, London, 2002.

C. Toth and H. Oberhauser. Bayesian learning from sequential data using Gaussian processes with signature covariances. In H. Daumé III and A. Singh, editors, *Proceedings of the 37th International Conference on Machine Learning*, volume 119, pages 9548–9560. PMLR, 2020.

B. Wang, M. Liakata, H. Ni, T. Lyons, A.J. Nevado-Holgado, and K. Saunders. A path signature approach for speech emotion recognition. In *Interspeech 2019*, pages 1661–1665, 2019.

D. Wilson-Nunn, T. Lyons, A. Papavasiliou, and H. Ni. A path signature approach to online Arabic handwriting recognition. In *2018 IEEE 2nd International Workshop on Arabic and Derived Script Analysis and Recognition*, pages 135–139. IEEE, 2018.

W. Yang, L. Jin, and M. Liu. DeepWriterID: An end-to-end online text-independent writer identification system. *IEEE Intelligent Systems*, 31:45–53, 2016.

W. Yang, T. Lyons, H. Ni, C. Schmid, and L. Jin. Developing the path signature methodology and its application to landmark-based human action recognition. In G. Yin and T. Zariphopoulou, editors, *Stochastic Analysis, Filtering, and Stochastic Optimization: A Commemorative Volume to Honor Mark H.A. Davis's Contributions*, pages 431–464. Springer, 2022.

Supplementary Material

5 Proof of Proposition 4

We start by proving a relation similar to Chen's identity (Proposition 1, (i)) for the insertion operator. For any $X \in BV(\mathbb{R}^d)$, $n \geqslant 1$, and $p \in \{1, \dots, n+1\}$, the insertion operator restricted to an interval $[s, t] \subset [0, 1]$ is defined by

$$\mathscr{L}^n_{p, X; [s,t]}(y) = \int_{(u_1, \dots, u_n) \in \Delta_{n; [s,t]}} dX_{u_1} \otimes \cdots \otimes dX_{u_{p-1}} \otimes y \otimes dX_{u_p} \otimes \cdots \otimes dX_{u_n}.$$

Lemma 2 *Let $X \in BV(\mathbb{R}^d)$, $n \geqslant 1$, $p \in \{1, \dots, n+1\}$, and $u, v, w \in [0, 1]$ such that $u < v < w$. Then*

$$\mathscr{L}^n_{p, X; [u,w]}(y) = \sum_{k=0}^{p-1} \mathbf{X}^k_{[u,v]} \otimes \mathscr{L}^{n-k}_{p-k, X; [v,w]}(y) + \sum_{k=p}^{n} \mathscr{L}^k_{p, X; [u,v]}(y) \otimes \mathbf{X}^{n-k}_{[v,w]}.$$

Proof This formula is obtained by splitting the integration domain in two parts, with a running index k corresponding to the number of integration variables in each part of the interval. We have

$$\mathscr{L}^n_{p, X; [u,w]}(y)$$

$$= \int_{u \leqslant u_1 \leqslant \cdots \leqslant u_n \leqslant w} \cdots \int dX_{u_1} \otimes \cdots \otimes dX_{u_{p-1}} \otimes y \otimes \cdots \otimes dX_{u_n}$$

$$= \sum_{k=0}^{n} \int_{u \leqslant u_1 \leqslant \cdots \leqslant u_k \leqslant v \leqslant u_{k+1} \leqslant \cdots \leqslant u_n \leqslant w} \cdots \int dX_{u_1} \otimes \cdots \otimes dX_{u_{p-1}} \otimes y \otimes \cdots \otimes dX_{u_n}$$

$$= \sum_{k=0}^{n} \int_{(u_1, \dots, u_k) \in \Delta_{k; [u,v]}} \int_{(u_{k+1}, \dots, u_n) \in \Delta_{n-k; [v,w]}} dX_{u_1} \otimes \cdots \otimes dX_{u_{p-1}} \otimes y \otimes \cdots \otimes dX_{u_n},$$

where we use the convention $u_0 = u$ and $u_{n+1} = w$. To simplify this expression, we analyze separately the case $k \leqslant p - 1$ and the case $k > p$. So,

$$\mathscr{L}^n_{p,X;[u,w]}(y)$$

$$= \sum_{k=0}^{p-1} \int_{\Delta_{k;[u,v]}} \int_{\Delta_{n-k;[v,w]}} dX_{u_1} \otimes \cdots \otimes dX_{u_{p-1}} \otimes y \otimes \cdots \otimes dX_{u_n}$$

$$+ \sum_{k=p}^{n} \int_{\Delta_{k;[u,v]}} \int_{\Delta_{n-k;[v,w]}} dX_{u_1} \otimes \cdots \otimes dX_{u_{p-1}} \otimes y \otimes \cdots \otimes dX_{u_n}$$

$$= \sum_{k=0}^{p-1} \left(\int_{(u_1,\ldots,u_k)\in\Delta_{k;[u,v]}} dX_{u_1} \otimes \cdots \otimes dX_{u_k} \right)$$

$$\otimes \left(\int_{(u_{k+1},\ldots,u_n)\in\Delta_{n-k;[v,w]}} dX_{u_{k+1}} \otimes \cdots \otimes dX_{u_{p-1}} \otimes y \otimes \cdots \otimes dX_{u_n} \right)$$

$$+ \sum_{k=p}^{n} \left(\int_{(u_1,\ldots,u_k)\in\Delta_{k;[u,v]}} dX_{u_1} \otimes \cdots \otimes dX_{u_{p-1}} \otimes y \otimes \cdots \otimes dX_{u_k} \right)$$

$$\otimes \left(\int_{(u_{k+1},\ldots,u_n)\in\Delta_{n-k;[v,w]}} dX_{u_{k+1}} \otimes \cdots \otimes dX_{u_n} \right)$$

$$= \sum_{k=0}^{p-1} \mathbf{X}^k_{[u,v]} \otimes \mathscr{L}^{n-k}_{p-k,X;[v,w]}(y) + \sum_{k=p}^{n} \mathscr{L}^k_{p,X;[u,v]}(y) \otimes \mathbf{X}^{n-k}_{[v,w]}.$$

$$\square$$

We are now ready to prove Proposition 2. The first step uses Proposition 1, (i) and Lemma 2 to split both the signature $\mathbf{X}^{n+1}$ and the insertion operator $\mathscr{L}^n_{p,X}(y)$ on the intervals $[0, t_{i-1}]$, $[t_{i-1}, t_i]$, and $[t_i, 1]$. With respect to the insertion operator, we split the sum depending on the location of p, similar to the proof of Lemma 2. Let $\mathscr{S}$ denote the set

$$\mathscr{S} = \left\{ (n_1, n_2, n_3) \in \{0, \ldots, n+1\} \mid n_1 + n_2 + n_3 = n+1 \right\}.$$

For a fixed $p \in \{1, \ldots, n+1\}$, we use the following partition, illustrated in Figure 7:

$$\mathscr{S} = \left\{ (n_1, n_2, n_3) \in \mathscr{S} \mid n_1 \geqslant p \right\} \cup \left\{ (n_1, n_2, n_3) \in \mathscr{S} \mid n_1 < p, \, n_1 + n_2 \geqslant p \right\}$$

$$\cup \left\{ (n_1, n_2, n_3) \in \mathscr{S} \mid n_1 + n_2 < p \right\}.$$

With this notation, we have
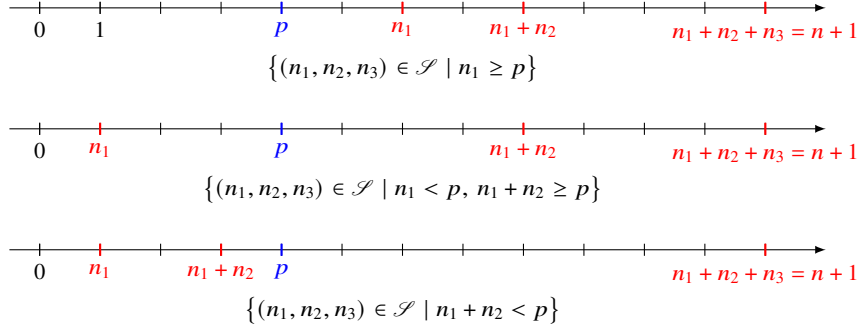
Fig. 7: Partition of $\mathscr{S}$.

$$
\begin{aligned}
\mathscr{L}_{p,X}^n(y) = & \sum_{\substack{(n_1,n_2,n_3)\in\mathscr{S} \\ p\leqslant n_1}} \mathscr{L}_{p,X;[0,t_{i-1}]}^{n_1-1}(y) \otimes \mathbf{X}_{[t_{i-1},t_i]}^{n_2} \otimes \mathbf{X}_{[t_i,1]}^{n_3} \\
& + \sum_{\substack{(n_1,n_2,n_3)\in\mathscr{S} \\ n_1<p\leqslant n_1+n_2}} \mathbf{X}_{[0,t_{i-1}]}^{n_1} \otimes \mathscr{L}_{p-n_1,X;[t_{i-1},t_i]}^{n_2-1}(y) \otimes \mathbf{X}_{[t_i,1]}^{n_3} \\
& + \sum_{\substack{(n_1,n_2,n_3)\in\mathscr{S} \\ n_1+n_2<p}} \mathbf{X}_{[0,t_{i-1}]}^{n_1} \otimes \mathbf{X}_{[t_{i-1},t_i]}^{n_2} \otimes \mathscr{L}_{p-n_1-n_2,X;[t_i,1]}^{n_3-1}(y) \\
= & \sum_{k=p-1}^n \mathscr{L}_{p,X;[0,t_{i-1}]}^{k}(y) \otimes \mathbf{X}_{[t_{i-1},1]}^{n-k} \\
& + \sum_{\substack{(n_1,n_2,n_3)\in\mathscr{S} \\ n_1<p\leqslant n_1+n_2}} \mathbf{X}_{[0,t_{i-1}]}^{n_1} \otimes \mathscr{L}_{p-n_1,X;[t_{i-1},t_i]}^{n_2-1}(y) \otimes \mathbf{X}_{[t_i,1]}^{n_3} \\
& + \sum_{k=0}^{p-1} \mathbf{X}_{[0,t_i]}^{k} \otimes \mathscr{L}_{p-k,X;[t_i,1]}^{n-k}(y).
\end{aligned}
$$

With respect to $\mathbf{X}^{n+1}$, a straightforward extension of Proposition 1 yields

$$
\mathbf{X}^{n+1} = \sum_{(n_1,n_2,n_3)\in\mathscr{S}} \mathbf{X}_{[0,t_{i-1}]}^{n_1} \otimes \mathbf{X}_{[t_{i-1},t_i]}^{n_2} \otimes \mathbf{X}_{[t_i,1]}^{n_3},
$$

and thus

$$
\begin{aligned}
\mathbf{X}^{n+1} = & \sum_{k=p-1}^n \mathbf{X}_{[0,t_{i-1}]}^{k+1} \otimes \mathbf{X}_{[t_{i-1},1]}^{n+1-k} + \sum_{\substack{(n_1,n_2,n_3)\in\mathscr{S} \\ n_1<p\leqslant n_1+n_2}} \mathbf{X}_{[0,t_{i-1}]}^{n_1} \otimes \mathbf{X}_{[t_{i-1},t_i]}^{n_2} \otimes \mathbf{X}_{[t_i,1]}^{n_3} \\
& + \sum_{k=0}^{p-1} \mathbf{X}_{[0,t_i]}^{k} \otimes \mathbf{X}_{[t_i,1]}^{n+1-k}.
\end{aligned}
$$

Putting everything together, we are led to

$$
n!(\mathscr{L}^n_{p,X}(\beta_i) - (n+1)\mathbf{X}^{n+1})
$$

$$
= n! \sum_{k=p-1}^{n} (\mathscr{L}^k_{p,X;[0,t_{i-1}]}(\beta_i) - (n+1)\mathbf{X}^{k+1}_{[0,t_{i-1}]}) \otimes \mathbf{X}^{n-k}_{[t_{i-1},1]}
$$

$$
+ n! \sum_{\substack{(n_1,n_2,n_3)\in\mathscr{S} \\ n_1 < p \leqslant n_1+n_2}} \mathbf{X}^{n_1}_{[0,t_{i-1}]} \otimes (\mathscr{L}^{n_2-1}_{p-n_1,X;[t_{i-1},t_i]}(\beta_i) - (n+1)\mathbf{X}^{n_2}_{[t_{i-1},t_i]}) \otimes \mathbf{X}^{n_3}_{[t_i,1]}
$$

$$
+ n! \sum_{k=0}^{p-1} \mathbf{X}^k_{[0,t_i]} \otimes (\mathscr{L}^{n-k}_{p-k,X;[t_i,1]}(\beta_i) - (n+1)\mathbf{X}^{n+1-k}_{[t_i,1]})
$$

$$
:= A_1 + A_2 + A_3. \tag{6}
$$

We now bound A_1, A_2, and A_3 separately. Note that the same arguments as in the proof of Proposition 3 give for any admissible norm $\|\mathscr{L}^n_{p,X;[s,t]}(y)\| = \|\mathbf{X}^n\| \cdot \|y\|$ and thus, by Proposition 2,

$$
\|\mathscr{L}^n_{p,X;[s,t]}(y)\| \leqslant \frac{\|X\|^n_{TV;[s,t]}}{n!} \|y\|.
$$

This yields, together with the triangle inequality,

$$
\|A_1\| = \left\| n! \sum_{k=p-1}^{n} (\mathscr{L}^k_{p,X;[0,t_{i-1}]}(\beta_i) - (n+1)\mathbf{X}^{k+1}_{[0,t_{i-1}]}) \otimes \mathbf{X}^{n-k}_{[t_{i-1},1]} \right\|
$$

$$
\leqslant n! \sum_{k=p-1}^{n} \left\| \mathscr{L}^k_{p,X;[0,t_{i-1}]}(\beta_i) \right\| \cdot \left\| \mathbf{X}^{n-k}_{[t_{i-1},1]} \right\|
$$

$$
+ (n+1)! \sum_{k=p-1}^{n} \left\| \mathbf{X}^{k+1}_{[0,t_{i-1}]} \right\| \cdot \left\| \mathbf{X}^{n-k}_{[t_{i-1},1]} \right\|
$$

$$
\leqslant n! \sum_{k=p-1}^{n} \|\beta_i\| \frac{\|X\|^k_{TV;[0,t_{i-1}]}}{k!} \frac{\|X\|^{n-k}_{TV;[t_{i-1},1]}}{(n-k)!}
$$

$$
+ (n+1)! \sum_{k=p-1}^{n} \frac{\|X\|^{k+1}_{TV;[0,t_{i-1}]}}{(k+1)!} \frac{\|X\|^{n-k}_{TV;[t_{i-1},1]}}{(n-k)!}.
$$

Recall that our choice of parameterization of X ensures that for any $j \in \{1,\ldots,M\}$, $\|\beta_j\| = \ell$. Therefore,

$$
\|X\|_{TV;[0,t_{i-1}]} = \sum_{j=1}^{i-1} \|\beta_j\|(t_j - t_{j-1}) = \ell \sum_{j=1}^{i-1} (t_j - t_{j-1}) = \ell t_{i-1}.
$$

Similarly, $\|X\|_{TV;[t_{i-1},1]} = \ell(1 - t_{i-1})$. It follows that

$$\|A_1\| \leqslant \ell^{n+1}\left(\sum_{k=p-1}^{n}\binom{n}{k}t_{i-1}^k(1-t_{i-1})^{n-k} + \sum_{k=p}^{n+1}\binom{n+1}{k}t_{i-1}^k(1-t_{i-1})^{n+1-k}\right). \quad (7)$$

Bounding A_3 in a similar way, we obtain

$$\|A_3\| \leqslant \ell^{n+1}\left(\sum_{k=0}^{p-1}\binom{n}{k}t_i^k(1-t_i)^{n-k} + \sum_{k=0}^{p-1}\binom{n+1}{k}t_i^k(1-t_i)^{n+1-k}\right). \quad (8)$$

We now turn to the term A_2. We have

$$\|A_2\| \leqslant n! \sum_{\substack{(n_1,n_2,n_3)\in\mathscr{S} \\ n_1<p\leqslant n_1+n_2}} \left\|\mathbf{X}_{[0,t_{i-1}]}^{n_1}\right\| \left\|\mathscr{L}_{p-n_1,X;[t_{i-1},t_i]}^{n_2-1}(\beta_i) - (n+1)\mathbf{X}_{[t_{i-1},t_i]}^{n_2}\right\| \left\|\mathbf{X}_{[t_i,1]}^{n_3}\right\|$$

$$\leqslant n! \sum_{(n_1,n_2,n_3)\in\mathscr{S}} \frac{t_{i-1}^{n_1}(1-t_i)^{n_3}\ell^{n_1}\ell^{n_3}}{n_1!n_3!} \left\|\mathscr{L}_{p-n_1,X;[t_{i-1},t_i]}^{n_2-1}(\beta_i) - (n+1)\mathbf{X}_{[t_{i-1},t_i]}^{n_2}\right\|.$$

Since X is linear on $[t_{i-1},t_i]$,

$$\mathbf{X}_{[t_{i-1},t_i]}^{n_2} = \frac{(t_i-t_{i-1})^{n_2}}{n_2!}\beta_i^{\otimes n_2}, \quad \mathscr{L}_{p-n_1,X;[t_{i-1},t_i]}^{n_2-1}(\beta_i) = \frac{(t_i-t_{i-1})^{n_2-1}}{(n_2-1)!}\beta_i^{\otimes n_2}.$$

So,

$$\left\|\mathscr{L}_{p-n_1,X;[t_{i-1},t_i]}^{n_2-1}(\beta_i) - (n+1)\mathbf{X}_{[t_{i-1},t_i]}^{n_2}\right\| = \frac{(t_i-t_{i-1})^{n_2}}{n_2!}\left|\frac{n_2}{t_i-t_{i-1}} - (n+1)\right| \cdot \|\beta_i^{\otimes n_2}\|$$

$$= \frac{(t_i-t_{i-1})^{n_2}\ell^{n_2}}{n_2!}\left|\frac{n_2}{t_i-t_{i-1}} - (n+1)\right|,$$

and

$\|A_2\|$

$$\leqslant n! \sum_{(n_1,n_2,n_3)\in\mathscr{S}} \frac{t_{i-1}^{n_1}\ell^{n_1}}{n_1!}\frac{(t_i-t_{i-1})^{n_2}\ell^{n_2}}{n_2!}\left|\frac{n_2}{t_i-t_{i-1}}-(n+1)\right|\frac{(1-t_i)^{n_3}\ell^{n_3}}{n_3!}$$

$$= \ell^{n+1} \sum_{(n_1,n_2,n_3)\in\mathscr{S}} \frac{(n+1)!}{n_1!n_2!n_3!}t_{i-1}^{n_1}(t_i-t_{i-1})^{n_2}(1-t_i)^{n_3}\left|\frac{n_2}{(n+1)(t_i-t_{i-1})}-1\right|$$

$$= \ell^{n+1} \sum_{k=0}^{n+1}\left(\frac{1}{k!}\left(\sum_{n_1=0}^{k}\frac{k!}{n_1!(k-n_1)!}t_{i-1}^{n_1}(1-t_i)^{k-n_1}\right)\right.$$

$$\left.\times \frac{(n+1)!}{(n+1-k)!}(t_i-t_{i-1})^{n+1-k}\left|\frac{n+1-k}{(n+1)(t_i-t_{i-1})}-1\right|\right)$$

$$= \ell^{n+1} \sum_{k=0}^{n+1}\frac{(n+1)!}{k!(n+1-k)!}(1-(t_i-t_{i-1}))^k(t_i-t_{i-1})^{n+1-k}\left|\frac{n+1-k}{(n+1)(t_i-t_{i-1})}-1\right|$$

$$= \ell^{n+1} \sum_{k=0}^{n+1}\binom{n+1}{k}(1-(t_i-t_{i-1}))^{n+1-k}(t_i-t_{i-1})^k\left|\frac{k}{(n+1)(t_i-t_{i-1})}-1\right|. \qquad (9)$$

The right-hand sides of (7), (8), and (9) correspond to probability mass functions of binomial random variables. Indeed, let

$$Y_{1,n}\sim\mathrm{Binom}(n,t_{i-1}),\quad Y_{2,n}\sim\mathrm{Binom}(n,t_i-t_{i-1}),\quad\text{and}\quad Y_{3,n}\sim\mathrm{Binom}(n,t_i).$$

Then

$$\|A_1\|\leqslant\ell^{n+1}\big(\mathbb{P}(Y_{1,n}\geqslant p-1)+\mathbb{P}(Y_{1,n+1}\geqslant p)\big),$$

$$\|A_3\|\leqslant\ell^{n+1}\big(\mathbb{P}(Y_{3,n}\leqslant p-1)+\mathbb{P}(Y_{3,n+1}\leqslant p-1)\big),$$

$$\|A_2\|\leqslant\ell^{n+1}\mathbb{E}\left[\left|\frac{Y_{2,n+1}}{(n+1)(t_i-t_{i-1})}-1\right|\right].$$

First, since $\mathbb{E}Y_{2,n+1}=(n+1)(t_i-t_{i-1})$ and $\mathrm{Var}(Y_{2,n+1})=(n+1)(t_i-t_{i-1})(1-(t_i-t_{i-1}))$, by Hölder's inequality,

$$\mathbb{E}\left[\left|\frac{Y_{2,n+1}}{(n+1)(t_i-t_{i-1})}-1\right|\right] = \frac{1}{(n+1)(t_i-t_{i-1})}\mathbb{E}\big[|Y_{2,n+1}-\mathbb{E}Y_{2,n+1}|\big]$$

$$\leqslant \frac{1}{(n+1)(t_i-t_{i-1})}\mathbb{E}\big[|Y_{2,n+1}-\mathbb{E}Y_{2,n+1}|^2\big]^{1/2}$$

$$\leqslant \frac{1}{(n+1)(t_i-t_{i-1})}\sqrt{(n+1)(t_i-t_{i-1})(1-(t_i-t_{i-1}))}$$

$$\leqslant \frac{1}{\sqrt{n+1}}\sqrt{\frac{1-(t_i-t_{i-1})}{t_i-t_{i-1}}}.$$

The other terms decay exponentially fast if p is well chosen. We give the details for the bound on $\|A_1\|$ below but the one on $\|A_3\|$ is treated in the same way. First, we

have

$$\|A_1\| \leqslant \ell^{n+1} \big(\mathbb{P}(Y_{1,n} \geqslant p - 1) + \mathbb{P}(Y_{1,n+1} \geqslant p) \big),$$
$$\leqslant \ell^{n+1} \big(\mathbb{P}(Y_{1,n} - \mathbb{E}Y_{1,n} \geqslant p - 1 - nt_{i-1}) $$
$$+ \mathbb{P}(Y_{1,n+1} - \mathbb{E}Y_{1,n+1} \geqslant p - (n+1)t_{i-1}) \big).$$

Recall that $p = \lfloor (n+1)(3t_i + t_{i-1})/4 \rfloor$. Thus, for $n \geqslant 2/(t_i - t_{i-1})$, we have

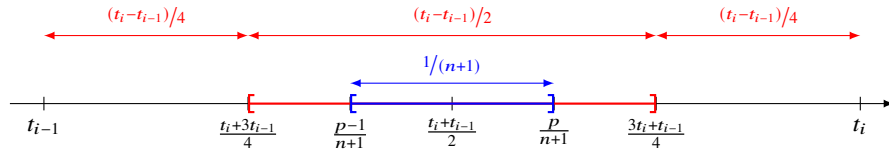$$\frac{t_i + 3t_{i-1}}{4} \leqslant \frac{p}{n+1} < \frac{3t_i + t_{i-1}}{4}.$$

Fig. 8: Illustration on the choice of p.

This is summarized in Figure 8. In particular,

$$p - (n+1)t_{i-1} = (n+1)\Big(\frac{p}{n+1} - t_{i-1} \Big)$$
$$\geqslant (n+1)\Big(\frac{t_i + 3t_{i-1}}{4} - t_{i-1} \Big) = (n+1)\frac{t_i - t_{i-1}}{4} > 0,$$

and

$$p - 1 - nt_{i-1} = n\Big(\frac{p}{n+1} - t_{i-1} + \frac{p}{n+1} - \frac{p-1}{n} \Big) \geqslant n\Big(\frac{t_i - t_{i-1}}{4} + \frac{n+1-p}{n(n+1)} \Big)$$
$$> n\frac{t_i - t_{i-1}}{4} > 0.$$

By Hoeffding's inequality, we obtain

$$\|A_1\| \leqslant \ell^{n+1} \Big(\exp\Big(-n\Big(\frac{p-1}{n} - t_{i-1}\Big)^2 \Big) + \exp\Big(-(n+1)\Big(\frac{p}{n+1} - t_{i-1}\Big)^2 \Big) \Big)$$
$$\leqslant \ell^{n+1} \Big(\exp\Big(-\frac{n(t_i - t_{i-1})^2}{16} \Big) + \exp\Big(-\frac{(n+1)(t_i - t_{i-1})^2}{16} \Big) \Big)$$
$$\leqslant 2\ell^{n+1} \exp\Big(-\frac{n(t_i - t_{i-1})^2}{16} \Big).$$

With similar arguments, we obtain

$$\|A_3\| \leqslant 2\ell^{n+1} \exp\left(-\frac{n(t_i - t_{i-1})^2}{16}\right).$$

Finally, combining (6) with the bounds above on $\|A_1\|$, $\|A_2\|$, and $\|A_3\|$, we conclude that

$$\|\mathcal{L}_{p,X}^n(\beta_i) - (n+1)\mathbf{X}^{n+1}\|$$

$$\leqslant \frac{\ell^{n+1}}{n!}\left(\frac{1}{\sqrt{n+1}}\sqrt{\frac{1-(t_i-t_{i-1})}{t_i-t_{i-1}}} + 4\exp\left(-\frac{n(t_i-t_{i-1})^2}{16}\right)\right).$$

6 Proof of Proposition 5

The proof is based on several ingredients of Hambly and Lyons (2010). Their core idea is to move the path to a hyperbolic space, which is called the "development" of the path. First, note that if $Y \in BV(\mathbb{R}^d)$ is defined by $Y_t = X_t/\ell$, where $\ell = \|X\|_{TV}$ is the length of X, then

$$\mathbf{Y}^n = \int_{(u_1,\ldots,u_n)\in\Delta_n} dY_{u_1}\otimes\cdots\otimes dY_{u_n} = \int_{(u_1,\ldots,u_n)\in\Delta_n} \frac{1}{\ell}dX_{u_1}\otimes\cdots\otimes\frac{1}{\ell}dX_{u_n} = \frac{1}{\ell^n}\mathbf{X}^n.$$

Therefore, we will assume without loss of generality that $\ell = 1$, and the lower bound obtained must be multiplied by ℓ^n to return to the general case.

We start the proof with a series of definitions.

Definition 8 The hyperboloid model is the subspace of $\mathbb{R}^{d+1}$ defined by

$$\mathbb{H} = \{y \in \mathbb{R}^{d+1} \mid B(y,y) = -1\},$$

where, for any $x, y \in \mathbb{R}^{d+1}$,

$$B(x,y) = \sum_{i=1}^d x_i y_i - x_{d+1}y_{d+1}.$$

This hyperbolic space has several well-known good properties (Cannon et al., 1997; Paupert, 2016; Loustau, 2020). The main one is that the hyperbolic distance between two points $x, y \in \mathbb{H}$, denoted by d, can be easily computed as

$$d(x,y) = \operatorname{arcosh}(-B(x,y)).$$

For E_1 and E_2 two finite-dimensional normed vector spaces, we let $L(E_1, E_2)$ be the vector space of linear operators from E_1 to E_2. Equipped with the operator norm, $L(E_1, E_2)$ is itself a normed vector space. Recall that for $f \in L(E_1, E_2)$,

$$\|f\|_{L(E_1,E_2)} = \sup_{x\in E_1, \|x\|_{E_1}=1} \|f(x)\|_{E_2}.$$

We will use linear maps on $\mathbb{R}^d$ equipped with the Euclidean norm, and consider linear maps as matrix multiplications. Therefore, we write fx instead of $f(x)$ for function evaluation. In particular, we let $F : \mathbb{R}^d \to L(\mathbb{R}^{d+1}, \mathbb{R}^{d+1})$ be defined for any $y \in \mathbb{R}^d$ by

$$Fy = \begin{pmatrix} 0 & \cdots & 0 & y_1 \\ \vdots & & \vdots & \vdots \\ 0 & \cdots & 0 & y_d \\ y_1 & \ldots & y_d & 0 \end{pmatrix}.$$

Lemma 3 *F is a bounded linear map and its operator norm is*

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

Proof For any $y \in \mathbb{R}^d$, $z \in \mathbb{R}^{d+1}$, we have

$$Fyz = \begin{pmatrix} 0 & \cdots & 0 & y_1 \\ \vdots & & \vdots & \vdots \\ 0 & \cdots & 0 & y_d \\ y_1 & \ldots & y_d & 0 \end{pmatrix} \begin{pmatrix} z_1 \\ \vdots \\ z_{d+1} \end{pmatrix} = \begin{pmatrix} y_1 z_{d+1} \\ \vdots \\ y_d z_{d+1} \\ \sum_{i=1}^d y_i z_i \end{pmatrix}.$$

Thus,

$$\|Fyz\|^2 = z_{d+1}^2 \sum_{i=1}^d y_i^2 + \Big(\sum_{i=1}^d y_i z_i \Big)^2$$

$$\leqslant z_{d+1}^2 \|y\|^2 + \|y\|^2 \sum_{i=1}^d z_i^2$$

(by the Cauchy-Schwartz inequality)

$$\leqslant \|y\|^2 \|z\|^2.$$

This yields

$$\|Fy\|_{L(\mathbb{R}^{d+1}, \mathbb{R}^{d+1})} = \sup_{z \in \mathbb{R}^{d+1}, \|z\|=1} \|Fyz\| \leqslant \sup_{z \in \mathbb{R}^{d+1}, \|z\|=1} \|y\| \, \|z\| = \|y\|.$$

This inequality becomes an equality when $z = e_{d+1}$ (where $(e_1, \ldots, e_{d+1})$ is the canonical basis of $\mathbb{R}^{d+1}$). Therefore,

$$\|Fy\|_{L(\mathbb{R}^{d+1}, \mathbb{R}^{d+1})} = \|y\|,$$

and

$$\|F\|_{L(\mathbb{R}^d, L(\mathbb{R}^{d+1}, \mathbb{R}^{d+1}))} = \sup_{y \in \mathbb{R}^d, \|y\|=1} \|Fy\|_{L(\mathbb{R}^{d+1}, \mathbb{R}^{d+1})} = \sup_{y \in \mathbb{R}^d, \|y\|=1} \|y\| = 1.$$

$\square$

We are now ready to define the development of a path $X \in BV(\mathbb{R}^d)$ to the hyperbolic space. For $t \in [0, 1]$, we let $\Gamma_t : \mathbb{R}^{d+1} \to \mathbb{R}^{d+1}$ be such that, for any $y \in \mathbb{R}^{d+1}$, $\Gamma_t y$ is the solution at time t of the controlled differential equation

$$dY_t = F(dX_t)Y_t, \quad Y_0 = y. \tag{10}$$

Lemma 4 *For any $t \in [0, 1]$, Γ_t is a (well-defined) linear map preserving the function B, i.e., for any $y, \widetilde{y} \in \mathbb{R}^{d+1}$,*

$$B(\Gamma_t y, \Gamma_t \widetilde{y}) = B(y, \widetilde{y}).$$

Proof Equation (10) is a linear controlled differential equation and, by Picard's theorem (see, e.g., Lyons et al., 2007, Theorem 1.3), for any $y \in \mathbb{R}^d$, it has a unique solution. Therefore Γ_t is well-defined. Moreover, it is clearly a linear operator: if $y, \widetilde{y} \in \mathbb{R}^d$, and $Y_t = \Gamma_t y$, $\widetilde{Y}_t = \Gamma_t \widetilde{y}$, then $Y_t + \widetilde{Y}_t$ follows equation (10) with initial point $y + \widetilde{y}$, so by uniqueness of the solution, $\Gamma_t(y + \widetilde{y}) = \Gamma_t y + \Gamma_t \widetilde{y}$.

The fact that Γ preserves B is proved in Hambly and Lyons (2010), Section 3. □

The differential equation (10) may be rewritten as a differential equation on Γ_t, which itself can be thought of as a $(d + 1) \times (d + 1)$ matrix:

$$d\Gamma_t = F(dX_t)\Gamma_t, \quad \Gamma_0 = I_{d+1}, \tag{11}$$

where I_{d+1} is the identity matrix. Now, let $y_0 = (0, \ldots, 0, 1)^\top \in \mathbb{R}^{d+1}$. The development of X to the hyperboloid space $\mathbb{H}$ is defined as the path $Y : [0, 1] \to \mathbb{H}$, $Y_t = \Gamma_t y_0$. Observe, since $B(y_0, y_0) = -1$, that $y_0 \in \mathbb{H}$ and thus, by Lemma 4, that $Y_t \in \mathbb{H}$ for any $t \in [0, 1]$. To summarize, equation (11) maps a path X in $\mathbb{R}^d$ into a new path Y in $\mathbb{H}$. It turns out that this embedding preserves a number of properties of the path X. In particular (Hambly and Lyons, 2010):

- Any linear piece of X is mapped to a geodesic in $\mathbb{H}$. So, if X is piecewise linear, then Y is geodesic on each $[t_{i-1}, t_i]$, and $d(Y_{t_i}, Y_{t_{i-1}}) = \|X_{t_i} - X_{t_{i-1}}\|$. Therefore, since X is linear on $[t_{i-1}, t_i]$, $d(Y_{t_i}, Y_{t_{i-1}}) = \|\beta_i\|(t_i - t_{i-1})$.
- The function $t \mapsto \Gamma_t$ preserves the angles between linear segments. If 2ω is the angle between the linear pieces $[X_{t_{i-1}}, X_{t_i}]$ and $[X_{t_i}, X_{t_{i+1}}]$, then the angle between the geodesics $[Y_{t_{i-1}}, Y_{t_i}]$ and $[Y_{t_i}, Y_{t_{i+1}}]$ is also equal to 2ω.

The last ingredient to prove Proposition 5 is the following lemma, due to Hambly and Lyons (2010, Lemma 3.7 and Proposition 3.13).

Lemma 5

(*i*) *Let $Y : [0, 1] \to \mathbb{H}$ be a continuous path, geodesic on the intervals $[t_{i-1}, t_i]$, where $0 = t_0 < t_1 < \cdots < t_M = 1$ is a partition of $[0, 1]$. If ω is the smallest angle between two geodesic segments and each geodesic segment has length at least $K(\omega)$, then*

$$0 \leqslant \sum_{i=1}^{M} d(Y_{t_{i-1}}, Y_{t_i}) - d(y_0, Y_1) \leqslant (M - 1)K(\omega).$$

(*ii*) *Let* $SO(B)$ *denote the group of* $(d+1) \times (d+1)$ *matrices with positive determinant preserving the isometry* B: *if* $G \in SO(B)$, *for any* $x, y \in \mathbb{R}^{d+1}$, $B(Gx, Gy) = B(x, y)$. *Then, for any* $G \in SO(B)$,

$$\|G\|_{L(\mathbb{R}^{d+1}, \mathbb{R}^{d+1})} \geqslant e^{d(y_0, Gy_0)},$$

where $y_0 = (0, \ldots, 0, 1)^\top \in \mathbb{R}^{d+1}$.

We now have all the components necessary to prove Proposition 5. Recall that it is assumed, without loss of generality, that X has total variation 1. For $\alpha > 0$, we let $X_t^\alpha = \alpha X_t$, and denote by Γ_t^α and Y_t^α the corresponding hyperbolic developments. Let D be the length of the shortest linear segment of X, and take α such that $\alpha > K(\omega)/D$. Clearly, the path $t \mapsto Y_t^\alpha$ satisfies the assumptions of Lemma 5, (*i*). Therefore

$$0 \leqslant \sum_{i=1}^{M} d(Y_{t_{i-1}}^\alpha, Y_{t_i}^\alpha) - d(y_0, Y_1^\alpha) \leqslant (M-1)K(\omega).$$

Since X is linear on each segment $[t_{i-1}, t_i]$, $d(Y_{t_{i-1}}^\alpha, Y_{t_i}^\alpha) = \|X_{t_i}^\alpha - X_{t_{i-1}}^\alpha\|$, and thus

$$\sum_{i=1}^{M} d(Y_{t_{i-1}}^\alpha, Y_{t_i}^\alpha) = \sum_{i=1}^{M} \|X_{t_i}^\alpha - X_{t_{i-1}}^\alpha\| = \alpha.$$

Hence,

$$d(y_0, Y_1^\alpha) \geqslant \alpha - (M-1)K(\omega).$$

Therefore, upon noting that $d(y_0, \Gamma_1^\alpha y_0) = d(y_0, Y_1^\alpha)$, according to Lemma 5,

$$\|\Gamma_1^\alpha\|_{L(\mathbb{R}^{d+1}, \mathbb{R}^{d+1})} \geqslant e^{d(y_0, \Gamma_1^\alpha y_0)} \geqslant e^{\alpha - (M-1)K(\omega)}. \tag{12}$$

Let us now prove that Γ_1 is in fact a linear function of the signature of X. To see this, observe, since Γ_1 is the solution at time 1 to the linear controlled differential equation (11), that

$$\begin{aligned}
\Gamma_1^\alpha &= I_{d+1} + \int_0^1 F(dX_t^\alpha)\Gamma_t \\
&= I_{d+1} + \int_0^1 F(dX_t^\alpha)\left(I_{d+1} + \int_0^t F(dX_s^\alpha)\Gamma_s\right) \\
&= I_{d+1} + \int_0^1 F(dX_t^\alpha) + \int_0^1 \int_0^t F(dX_t^\alpha)F(dX_s^\alpha)\Gamma_s \\
&= \ldots
\end{aligned}$$

Iterating this procedure yields the identity

$$\Gamma_1^\alpha = I_{d+1} + \sum_{k=1}^\infty \int_{\Delta_k} F(dX_{t_1}^\alpha) \cdots F(dX_{t_k}^\alpha). \tag{13}$$

In order to show that the series above is well-defined, denote by $F^{\otimes k}$ the linear map $F^{\otimes k} : (\mathbb{R}^d)^{\otimes k} \to L(\mathbb{R}^{d+1}, \mathbb{R}^{d+1})$ such that, for any tensor $u = \sum_{j=1}^\ell u_{1,j} \otimes \cdots \otimes u_{k,j}$,

$$F^{\otimes k}(u) = \sum_{j=1}^\ell F(u_{1,j}) \cdots F(u_{k,j}).$$

Thus,

$$\|F^{\otimes k}(u)\|_{L(\mathbb{R}^{d+1}, \mathbb{R}^{d+1})} \leqslant \sum_{j=1}^p \|F(u_{1,j}) \cdots F(u_{k,j})\|_{L(\mathbb{R}^{d+1}, \mathbb{R}^{d+1})}$$

$$\leqslant \sum_{j=1}^p \|F(u_{1,j})\|_{L(\mathbb{R}^{d+1}, \mathbb{R}^{d+1})} \cdots \|F(u_{k,j})\|_{L(\mathbb{R}^{d+1}, \mathbb{R}^{d+1})}$$

(because the operator norm is sub-multiplicative)

$$\leqslant \sum_{j=1}^p \|u_{1,j}\| \cdots \|u_{k,j}\|$$

(by Lemma 3).

Taking the infimum over any representation of u yields $\|F^{\otimes k}(u)\|_{L(\mathbb{R}^{d+1}, \mathbb{R}^{d+1})} \leqslant \|u\|_\pi$. (Note that this is not true for any tensor norm.) Endowing $(\mathbb{R}^d)^{\otimes k}$ with the projective norm, the kth term in the sum (13) becomes

$$F(dX_{t_1}^\alpha) \cdots F(dX_{t_k}^\alpha) = F^{\otimes k}\left(\int_{\Delta_k} dX_{t_1}^\alpha \otimes \cdots \otimes dX_{t_k}^\alpha\right)$$

$$= \alpha^k F^{\otimes k}\left(\int_{\Delta_k} dX_{t_1} \otimes \cdots \otimes dX_{t_k}\right)$$

$$= \alpha^k F^{\otimes k}(\mathbf{X}^k),$$

and

$$\left\|\alpha^k F^{\otimes k}(\mathbf{X}^k)\right\|_{L(\mathbb{R}^{d+1}, \mathbb{R}^{d+1})} = \alpha^k \left\|F^{\otimes k}(\mathbf{X}^k)\right\|_{L(\mathbb{R}^{d+1}, \mathbb{R}^{d+1})} \leqslant \alpha^k \|\mathbf{X}^k\|_\pi \leqslant \alpha^k/k!$$

by Proposition 2. This shows that the right-hand side of (13) is a convergent series. In addition, we have just proved that

$$\Gamma_1^\alpha = I_{d+1} + \sum_{k=1}^\infty \alpha^k F^{\otimes k}(\mathbf{X}^k),$$

i.e., Γ_1^α is a linear function of the signature, and, moreover, that

$$\|\Gamma_1^\alpha\|_{L(\mathbb{R}^{d+1}, \mathbb{R}^{d+1})} \leqslant 1 + \sum_{k=1}^\infty \alpha^k \|\mathbf{X}^k\|_\pi. \tag{14}$$

Next, let, for any $k \geqslant 1$, $b_k = k! \|\mathbf{X}^k\|_\pi$, and $b_0 = 1$. Note that $b_k \leqslant 1$ since we have taken X of total variation 1. Thus, combining (12) with (14) yields

$$e^{-(M-1)K(\omega)} \leqslant e^{-\alpha} \sum_{k=0}^\infty \frac{\alpha^k}{k!} b_k. \tag{15}$$

The last step of the proof is to use inequality (15) to prove that for each $n > n_1$ there exists an integer $k_n \in [n - n^{3/4}, n + n^{3/4}]$ such that $b_{k_n} \geqslant e^{-(M-1)K(\omega)}/2$. Indeed, if such a sequence (k_n) exists, then

$$\|\mathbf{X}^{k_n}\|_\pi = \frac{b_{k_n}}{k_n!} \geqslant \frac{e^{-(M-1)K(\omega)}}{2k_n!},$$

which is the desired conclusion.

Let $Z \sim \mathrm{Poisson}(\alpha)$ be a random variable following a Poisson distribution with parameter α. The right-hand-side of (15) is then exactly $\mathbb{E}[b_Z]$. In the next few lines, we show that

$$\mathbb{P}\left(b_Z \geqslant \frac{1}{2} e^{-(M-1)K(\omega)} \text{ and } Z \in J_n\right) > 0, \tag{16}$$

where J_n denotes the interval $[n - n^{3/4}, n + n^{3/4}]$. It follows that there exists $k_n \in J_n$ such that $b_{k_n} \geqslant \frac{1}{2} e^{-(M-1)K(\omega)}$ (otherwise the probability above would be zero). In order to prove (16), observe, since $b_k \leqslant 1$ for all $k \geqslant 0$, that

$$e^{-(M-1)K(\omega)} \leqslant \mathbb{E}[b_Z] = \mathbb{E}[b_Z \mathbf{1}_{b_Z \geqslant \frac{1}{2} e^{-(M-1)K(\omega)}}] + \mathbb{E}[b_Z \mathbf{1}_{b_Z < \frac{1}{2} e^{-(M-1)K(\omega)}}]$$

$$\leqslant \mathbb{P}\left(b_Z \geqslant \frac{1}{2} e^{-(M-1)K(\omega)}\right) + \frac{1}{2} e^{-(M-1)K(\omega)}.$$

This yields the inequality

$$\mathbb{P}\left(b_Z \geqslant \frac{1}{2} e^{-(M-1)K(\omega)}\right) \geqslant \frac{1}{2} e^{-(M-1)K(\omega)} > 0.$$

Moreover, by Chebyshev's inequality,

$$\mathbb{P}(b_Z \notin J_n) = \mathbb{P}(|Z - n| > n^{3/4}) \leqslant \frac{1}{\sqrt{n}}.$$

Let $n_1 = \lfloor 4e^{2(M-1)K(\omega)} \rfloor$. Then, for any $n > n_1$,

$$\mathbb{P}\Big(b_Z \geqslant \frac{1}{2}e^{-(M-1)K(\omega)} \text{ and } Z \in J_n\Big) \geqslant \mathbb{P}\Big(b_Z \geqslant \frac{1}{2}e^{-(M-1)K(\omega)}\Big) - \mathbb{P}\big(b_Z \notin J_n\big)$$

$$\geqslant \frac{1}{2}e^{-(M-1)K(\omega)} - \frac{1}{\sqrt{n}}$$

$$\geqslant \frac{1}{2}e^{-(M-1)K(\omega)} - \frac{1}{\sqrt{n_1}} > 0.$$

This shows inequality (16).

7 Proof of Lemma 1

Recall that $(e_1, \ldots, e_d)$ denotes the canonical basis of $\mathbb{R}^d$. Each column of A_p corresponds to $\mathscr{L}^n_{p,X}(e_j)$, and, for any $(n+1)$-tuple $(i_1, \ldots, i_{n+1}) \in \{1, \ldots, d\}^{n+1}$, one has

$$(\mathscr{L}^n_{p,X}(e_j))_{i_1,\ldots,i_{n+1}} = \begin{cases} S^{(i_1,\ldots,i_{p-1},i_{p+1},\ldots,i_{n+1})} & \text{if } i_p = j, \\ 0 & \text{otherwise.} \end{cases}$$

Therefore, each row of A_p, indexed by an $(n+1)$-tuple $(i_1, \ldots, i_{n+1})$, contains only one non-zero element, namely the one in the column $j = i_p$. Let $a_{i_1,\ldots,i_{n+1},j}$ denote the element in column j corresponding to the row $(i_1, \ldots, i_{n+1})$. Then, for any $q, r \in \{1, \ldots, d\}$,

$$(A_p^\top A_p)_{q,r} = \sum_{(i_1,\ldots,i_{n+1}) \in \{1,\ldots,d\}^{n+1}} a_{i_1,\ldots,i_{n+1},q}\, a_{i_1,\ldots,i_{n+1},r}$$

$$= \sum_{i_p=1}^{d} \sum_{(i_1,\ldots,i_{p-1},i_{p+1},\ldots,i_{n+1}) \in \{1,\ldots,d\}^n} a_{i_1,\ldots,i_{n+1},q}\, a_{i_1,\ldots,i_{n+1},r}$$

For both $a_{i_1,\ldots,i_{n+1},q}$ and $a_{i_1,\ldots,i_{n+1},r}$ to be non-zero, q and r must be equal to i_p. Therefore, $A_p^\top A_p$ is diagonal. The diagonal terms, corresponding to $i_p = q = r$, are equal to

$$(A_p^\top A_p)_{q,q} = \sum_{(i_1,\ldots,i_{p-1},i_{p+1},\ldots i_{n+1}) \in \{1,\ldots,d\}^n} (S^{(i_1,\ldots,i_{p-1},i_{p+1},\ldots,i_{n+1})})^2 = \|\mathbf{X}^n\|^2.$$