A High Order Solver for Signature Kernels Multiscale Homogenization of Goursat PDEs

Maud Lemercier and Terry Lyons

Mathematical Institute, University of Oxford.

Abstract

Signature kernels are at the core of several machine learning algorithms for analysing multivariate time series. The kernel of two bounded variation paths (such as piecewise linear interpolations of time series data) is typically computed by solving a Goursat problem for a hyperbolic partial differential equation (PDE) in two independent time variables. However, this approach becomes considerably less practical for highly oscillatory input paths, as they have to be resolved at a fine enough scale to accurately recover their signature kernel, resulting in significant time and memory complexities. To mitigate this issue, we first show that the signature kernel of a broader class of paths, known as *smooth rough paths*, also satisfies a PDE, albeit in the form of a system of coupled equations. We then use this result to introduce new algorithms for the numerical approximation of signature kernels. As bounded variation paths (and more generally geometric $\boldsymbol{p}$-rough paths) can be approximated by piecewise smooth rough paths, one can replace the PDE with rapidly varying coefficients in the original Goursat problem by an explicit system of coupled equations with piecewise constant coefficients derived from the first few iterated integrals of the original input paths. While this approach requires solving more equations, they do not require looking back at the complex and fine structure of the initial paths, which significantly reduces the computational complexity associated with the analysis of highly oscillatory time series.

Keywords: rough analysis, kernel, hyperbolic linear PDEs

1 Introduction

Kernels are at the core of several well-established methods for classification [1], regression [2, 3], novelty detection [4] and statistical hypothesis testing [5, 6]. Real-valued kernels, defined as symmetric positive definite functions $\kappa : \mathcal{X} \times \mathcal{X} \to \mathbb{R}$, arise in Bayesian statistics as covariance functions for Gaussian process priors [7] over real-valued functions $f : \mathcal{X} \to \mathbb{R}$. In deep learning, they have played a pivotal role in understanding the large-scale limits of neural networks [8, 9]. They also underpin the construction of statistical scoring rules and discrepancies for fitting the parameters of deep generative models [10–12]. They are useful in mesh-free methods for solving partial differential equations and inverse problems [13, 14]. All these techniques are transferable to different input spaces in the sense that they can be tailored to different data types by choosing a suitable kernel. A real-valued kernel can always be represented as an inner product $\kappa(x, y) = \langle \varphi(x), \varphi(y) \rangle$ between some representations of the inputs x and y in a Hilbert space H with inner product $\langle \cdot, \cdot \rangle$ via a feature map $\varphi : \mathcal{X} \to H$. Depending on the problem at hand, better results might be achieved by mapping the inputs to a high-dimensional or even infinite-dimensional feature space. Numerically, this is tractable if the inner product can be obtained without computing every individual coordinate of $\varphi(x)$ and $\varphi(y)$, a strategy which is commonly referred to as a *kernel trick*.

In this article, we consider so-called *signature kernels*, a class of symmetric positive definite functions defined on some spaces of paths, known for their effectiveness in time series data analysis. These kernel functions have an explicit representation in terms of the signature, a central map in stochastic analysis [15]. The latter maps any smooth path $x : [0,T] \to E$ taking its values in a vector space E to the solution of a system of linear ordinary differential equation driven by this path

$$\dot{\mathbf{Z}}_t = \mathbf{Z}_t \otimes \dot{x}_t \tag{1}$$

and started at $\mathbf{Z}_0 = \mathbf{1}$. At any time, the output $\boldsymbol{Z}_t$ leaves in the space of tensor series over E,

$$T((E)) = \left\{ \mathbf{A} = (a^0, a^1, a^2, \dots) \mid a^0 \in \mathbb{R}, \forall k > 0,\ a^k \in E^{\otimes k} \right\}$$

which is an algebra endowed with the addition and multiplication operations $+$ and $\otimes$ (defined in Section 2) and unitary element $\mathbf{1} := (1, 0, 0, \dots)$. The solution $\mathbf{Z}_t$ can be expressed explicitly in terms of the k-fold iterated integrals of the path x, that is, $\mathbf{Z}_t = S(x)_{0,t}$ where

$$S(x)_{s,t} = \left(1,\ \int_{s<\tau<t} \dot{x}(\tau) d\tau,\ \ \dots,\ \int \dots \int_{s<\tau_1<\dots<\tau_k<t} \left\{ \dot{x}(\tau_1) \otimes \dots \otimes \dot{x}(\tau_k) \right\} d\tau_1 \dots d\tau_k,\ \ \dots \right).$$

While the signature $S(x) := S(x)_{0,T}$ of a path is an infinite collection of summary statistics, in machine learning applications, it is common practice to use the initial terms to embed time series data into feature vectors that can then be integrated with various methods from classical multivariate statistics [16–18]. In other words, each path x in a dataset is mapped to the collection of features $(1, S^1(x), \dots, S^n(x))$, defined as the projection of $S(x)$ on $T^n(E) := \bigoplus_{k=0}^n E^{\otimes k}$ with the convention $E^{\otimes 0} = \mathbb{R}$.

If the state space E of the paths is endowed with an inner product $\langle \cdot, \cdot \rangle_1$, and we denote by $\langle \cdot, \cdot \rangle_k$ the canonical (Hilbert-Schmidt) inner product on $E^{\otimes k}$ derived from $\langle \cdot, \cdot \rangle_1$, then an inner product on $T(E)$—the subalgebra of $T((E))$ in which all but finitely many projections are zero—can be defined for any $\mathbf{A}, \mathbf{B} \in T(E)$ by

$$\langle \mathbf{A}, \mathbf{B} \rangle := \sum_{k=0}^{\infty} \langle a^k, b^k \rangle_k.$$

We note that other choices are possible [19] but will not be considered in this article. The completion of $T(E)$ with respect to $\langle \cdot, \cdot \rangle$ is a Hilbert space $(\mathcal{H}, \langle \cdot, \cdot \rangle_{\mathcal{H}})$, and the factorial decay of the signature terms [20, Prop. 2.2] ensures that the signature of a path segment $S(x)_{s,t}$ actually takes its values in $\mathcal{H} \subset T((E))$. With this, a kernel can then be defined as

$$\kappa(x, y) := \langle S(x), S(y) \rangle_{\mathcal{H}} = \sum_{k=0}^{\infty} \langle S^k(x), S^k(y) \rangle_k. \tag{2}$$

These so-called *signature kernels*, originally introduced in [21], have found several applications in Bayesian modelling [22, 23], distribution regression [24–26], generative modelling [27, 28], theoretical deep learning [29, 30] and numerical analysis [31].

The function in eq. (2) can be approximated by the inner product in the truncated tensor algebra $T^n(E)$ of the truncated signatures of the paths x and y. For smooth paths, the approximation error decays factorially with the truncation level n

$$\left| \kappa(x, y) - \sum_{k=0}^{n} \langle S^k(x), S^k(y) \rangle_k \right| \leq \frac{2e^c c^{n+1}}{(n+1)!},$$

where the constant $c = \max\{\|x\|_1, \|y\|_1\}$ depends on the 1-variation (see Definition 1) of the input paths. The k-fold iterated integrals up to $k = n$ of piecewise linear paths taking their values in a vector space E of dimension $d \in \mathbb{N}$ can be efficiently computed using highly optimized Python libraries [32–35]. However, the paths need to be of finite dimension and the computational time and memory both increase exponentially with d. Alternatively, the inner product in $T^n(E)$ can be computed using a Horner scheme [21] which does not require the explicit computation of iterated integrals; consequently it remains tractable for infinite dimensional Hilbert-space valued paths (e.g. certain classes of spatiotemporal functions). However, a high truncation level n might be necessary to obtain a sufficiently accurate approximation if c is large.

Leveraging the definition of the signature as the solution of the differential equation (1), the signature kernel $\kappa(x, y) = u(T, T)$ of the pair of paths (x, y) has been shown in [36] to be the solution at final times $(s, t) = (T, T)$ of the linear hyperbolic second-order PDE

$$\frac{\partial^2 u}{\partial s \partial t} = u \cdot \langle \dot{x}_s, \dot{y}_t \rangle_1, \tag{3}$$

with boundary conditions $u(0, t) = 1$ and $u(s, 0) = 1$. This result provides a second approach for approximating the signature kernel of two paths x and y without computing their higher order iterated integrals. For piecewise linear paths, the error now depends on the step size $\Delta t \Delta s$ of the numerical solver used to compute a numerical solution $\hat{u}$ of the PDE in eq. (3)

$$|\kappa(x, y) - \hat{u}(T, T)| \leq c' \Delta t \Delta s,$$

where c' is a function of $\sup_{s,t \in [0,T]} |\langle \dot{x}_s, \dot{y}_t \rangle|$. However, highly oscillatory input paths force the use of small step sizes and the memory and time complexities become prohibitive. Optimizing the cost of a single kernel evaluation is critical, as kernel methods typically involve evaluations on multiple pairs of inputs to construct Gram matrices. Moreover, it is worth noting that the broader class of weighted signature kernels introduced in [19], necessitate solving multiple PDEs within a single kernel evaluation. One straightforward approach to overcome this challenge is to subsample the input paths. However, this strategy fails when the observations originate from the discretization of underlying continuous paths of large (possibly unbounded) 1-variation. In such situations, it is crucial to ensure that the paths are resolved at a fine enough scale to accurately recover their signature kernel.

Improving the scalability of kernel methods has been a long-standing challenge [37]. Recently, random approximation techniques for signature kernels have been proposed in [38] and [30]. However, when the paths are highly oscillatory, as is often the case with real-world sequential data, they may simply be better described as so-called *rough paths* [15]. These objects $\mathbf{X}$ can be seen as a path x augmented with a finite number of higher order objects $x^2, \ldots, x^m$, which locally, remove the need to look at the fine structure of the path. Rough path theory guarantees that the signature kernel is still well-defined in this setting. However, although Horner schemes converging to the kernels of geometric rough paths have been designed [21], and these kernels have also been shown to satisfy an equation akin to the integral form of eq. (3) in [25], effective algorithms for computing signature kernels of rough paths, have been, to our knowledge, limited so far. The main difficulty in adapting the aforementioned Horner schemes and the PDE (3) to rough paths lies in the fact that their derivations both rely on the following property of the inner product

$$\langle \mathbf{A} \otimes a^k, \mathbf{B} \otimes b^k \rangle_{\mathcal{H}} = \langle \mathbf{A}, \mathbf{B} \rangle_{\mathcal{H}} \langle a^k, b^k \rangle_k, \tag{4}$$

where a^k and b^k are two elements of $E^{\otimes k}$. For example, the PDE in eq. (3) is derived by taking $k = 1$ and replacing $(\mathbf{A}, \mathbf{B})$ with $(S(x)_{0,s}, S(y)_{0,t})$ and (a^1, b^1) with $(\dot{x}_s, \dot{y}_t) \in E \times E$. As p-rough paths $\mathbf{X} = (1, x^1, \ldots, x^{\lfloor p \rfloor})$ take their values in $\bigoplus_{k=0}^{\lfloor p \rfloor} E^{\otimes k}$, generalisations of the previous ideas would involve terms of the form $\langle \mathbf{A} \otimes a^k, \mathbf{B} \otimes b^q \rangle_{\mathcal{H}}$ with $k \neq q$ which do not factor into two inner products as it is the case above.

1.1 Main results

In this article, we find that the trick to solve the aforementioned algebraic problem is to rewrite these terms as $\langle \ell(\mathbf{B})(\mathbf{A}), r(a^k)(b^q)\rangle_{\mathcal{H}}$ if $k \leq q$ or $\langle \ell(\mathbf{A})(\mathbf{B}), r(b^q)(a^k)\rangle_{\mathcal{H}}$ otherwise. We will give the precise definition of the maps $\ell(\mathbf{A})(\cdot)$ and $r(\mathbf{A})(\cdot)$ in Section 3. Furthermore, we choose to study the kernels of *smooth rough paths* [39]. This choice is instrumental in solving our computational problem: the numerical intractability of eq. (3).

First, it allows us to use classical calculus to show that the kernel of any pair of smooth rough paths $\mathbf{X} = (1, x^1, \ldots, x^m)$ and $\mathbf{Y} = (1, y^1, \ldots, y^n)$, also solves a linear PDE. However, in general, it is not sufficient to solve a *single* PDE: as one adds more terms to describe the input paths, one also needs to solve a larger system of coupled equations (Section 3.1). We show that the additional variables to consider are the initial terms of $\ell(S(\mathbf{X}))(S(\mathbf{Y}))$ and $\ell(S(\mathbf{Y}))(S(\mathbf{X}))$ both taking their values in $\mathcal{H} \subset T((E))$. Given that their scalar component is the signature kernel $\langle S(\mathbf{X}), S(\mathbf{Y})\rangle_{\mathcal{H}}$, it is not entirely surprising that the new system is formulated in terms of these variables. Importantly, for any two smooth rough paths of finite degrees m and n, these equations can be decomposed into two sets. The first set includes the signature kernel and the lower-order terms of $\ell(S(\mathbf{X}))(S(\mathbf{Y}))$ and $\ell(S(\mathbf{Y}))(S(\mathbf{X}))$. Its size is determined by m and n and it forms a closed system which can be solved independently from the other. Once the values of the unknown variables within this set are obtained, the complementary state can be subsequently determined. In other words, finitely many extra state is needed to compute the signature kernel, and this extra state is enough to determine all the higher order terms of $\ell(S(\mathbf{X}))(S(\mathbf{Y}))$ and $\ell(S(\mathbf{Y}))(S(\mathbf{X}))$.

Second, we use this new result to derive high order schemes for the numerical approximation of signature kernels of arbitrary rough paths. Although the inner product of the signatures of any pair of smooth paths solves a Goursat problem for a *single* PDE of type eq. (3), obtaining an accurate numerical solution might require unacceptable amounts of computational time and memory if the paths, hence the coefficients of the PDE, are too oscillatory. This problem can be alleviated by approximating the inputs by suitable piecewise smooth rough paths, termed *piecewise log-linear paths* [39] (a.k.a. piecewse abelian paths [40] or pure rough paths [41]), whose kernel solves an augmented system of equations and provides a good approximation of the kernel of the initial paths. We note that substituting the path driving a differential equation by a piecewise log-linear approximation is classical in the field of numerical analysis of rough differential equations [42–44], where it is known as the log-ODE method, an approach which has recently found applications in deep learning [45, 46]. A piecewise log-linear approximation is determined by two parameters: a partition $D = \{0 = t_0 \leq t_1 \leq \ldots \leq t_n = T\}$ of the time interval $[0, T]$ and a degree, i.e. an integer $m \geq 1$. In our case, we apply the approximation to two inputs. Degree-1 approximations correspond to classical piecewise linear approximations, in which case the system collapses to a scalar PDE of type eq. (3). By choosing higher degrees, the original PDE is replaced with a bigger system of coupled equations. The new coefficients, which are piecewise constant, are derived from the *log-signatures* of the original input paths, taken over each interval of the partition. Ultimately, the discretization of these PDEs provides a numerical scheme for computing the kernels of arbitrary rough paths without transitions back to the huge but irrelevant fine structure.

1.2 Outline

In Section 2, we recall the concepts from rough path theory, including the definition of smooth rough paths as introduced by [39], essential for the rest of the paper. In Section 3 we present our main results. First, we show that the signature kernel of two smooth rough paths solves an augmented system of PDEs (Section 3.1) which generalizes the Goursat problem of [36] to a broader class of inputs. Second, we leverage this result to derive new numerical schemes for the efficient computation of signature kernels. The key idea is to replace the inputs by piecewise log-linear paths, a natural extension of classical piecewise linear approximations (Section 3.2). In Section 4, these findings are illustrated on simulated data. Finally, in Section 5, we conclude and outline potential future work directions.

2 Preliminaries

In this section, we recall the notions from rough path theory necessary for Section 3.

Definition 1 (p-variation)**.** *Let $x : [0,T] \to E$ be a continuous path valued in a normed vector space $(E, \|\cdot\|)$. Denoting $x_{t_i,t_{i+1}} := x(t_i) - x(t_{i+1})$, the p-variation of x on any interval $[s,t] \subseteq [0,T]$ is defined by*

$$\|x\|^p_{p\text{-var},[s,t]} = \sup\Big\{ \sum_{t_i \in D} \|x_{t_i,t_{i+1}}\|^p \;\Big|\; D \text{ finite partition of } [s,t] \Big\}.$$

Any continuous path $x : [0,T] \to E$ of finite 1-variation can be canonically lifted to a path $\mathbf{Z} : t \mapsto S(x)_{0,t}$ with values in $T((E))$, the space of tensor series over E, simply by considering all its iterated (Riemann-Stieltjes) integrals. This space is endowed with two internal operations: an addition and a product. For any two elements $\mathbf{A} = (a^0, a^1, a^2, \ldots)$ and $\mathbf{B} = (b^0, b^1, b^2, \ldots)$ in $T((E))$ and any scalar $\lambda \in \mathbb{R}$,

$$\lambda\mathbf{A} + \mathbf{B} = (\lambda a^0 + b^0,\, \lambda a^1 + b^1,\, \ldots)$$
$$\mathbf{A} \otimes \mathbf{B} = (c^0,\, c^1,\, \ldots) \quad \text{with } c^n = \sum_{k=0}^{n} a^k \otimes b^{n-k}.$$

The space $T((E))$ endowed with these operations is a (non-commutative) algebra with unitary element $\mathbf{1} = (1, 0, 0, \ldots)$. It is often important to look only at finitely many terms of an element of $T((E))$. To this aim, we define $T^n(E)$ the truncated tensor algebra over E of order $n \in \mathbb{N}$. More precisely, $T^n(E)$ is the quotient of $T((E))$ by the ideal $T^{>n}(E)$ defined by

$$T^{>n}(E) = \left\{ \mathbf{A} = (0, 0, \ldots, a^{n+1}, \ldots) \mid \mathbf{A} \in T((E)) \right\}.$$

We denote by π_n the quotient map from $T((V))$ to $T^n(V)$. We identify $T^n(E)$ with $\bigoplus_{k=0}^n E^{\otimes k}$ equipped with the product $(a^0, a^1, \ldots, a^n) \otimes_n (b^0, b^1, \ldots, b^n) = (c^0, c^1, \ldots, c^n)$ with $c^i = \sum_{k=0}^i a^k \otimes b^{i-k}$. With this identification π_n becomes a projection.

The unital associative algebra $(T((E)), +, \otimes)$ carries a Lie bracket $[\cdot,\cdot]$ defined by $[\mathbf{A}, \mathbf{B}] = \mathbf{A} \otimes \mathbf{B} - \mathbf{B} \otimes \mathbf{A}$, and there are several canonical Lie algebras associated to $T((E))$.

Definition 2 (Lie series)**.** *The space of Lie formal series over E, denoted as $\mathcal{L}((E))$ is defined as the following subspace of $T((E))$*

$$\mathcal{L}((E)) = \left\{ \mathbf{L} = (l^0, l^1, \ldots) \mid \forall k \geq 0,\ l^k \in L_k \right\}$$

where $L_0 = 0$, $L_1 = E$, and $L_{k+1} = [E, L_k]$, with $[E, F]$ denoting the linear span of all elements of the form $[e, f]$ where $(e, f) \in E \times F$ for any two linear subspaces E, F of $T((E))$.

If x is a path segment, then $\log S(x) \in \mathcal{L}((E))$ [47] where the logarithm map is defined for any $\mathbf{A} \in T^{>0}(E)$ by

$$\log(\mathbf{1} + \mathbf{A}) = \sum_{k \geq 1} \frac{(-1)^{k-1}}{k} \mathbf{A}^{\otimes k}.$$

We denote the space of Lie polynomials by $\mathcal{L}(E)$. For any $n \geq 1$, the *step-n free Lie algebra* is defined by $\mathcal{L}^n(E) := \pi_n(\mathcal{L}((E)))$ with elements called Lie polynomials of degree n.
The map $\log_n$ associates to each $\mathbf{A} \in \pi_n(T^{>0}(E))$ the element of $T^n(E)$ defined as

$$\log_n(\mathbf{1} + \mathbf{A}) = \sum_{k=1}^{n} \frac{(-1)^{k-1}}{k} \mathbf{A}^{\otimes k}.$$

Note that is satisfies $\pi_n(\log(\mathbf{1}+\mathbf{A})) = \log_n(\pi_n(\mathbf{1}+\mathbf{A}))$. Finally, we note that the path signature $\mathbf{Z}$ actually takes its values in a curved subspace $G(E) \subset T((E))$ with a group structure. It is given by $G(E) = \exp\{\mathcal{L}((E))\}$ where

$$\exp(\mathbf{A}) = \sum_{k=0}^{\infty} \frac{\mathbf{A}^{\otimes k}}{k!}.$$

For each $n \geq 1$, we denote $G^n(E) = \pi_n(G(E))$ and $\exp_n(\mathbf{A}) = \sum_{k=0}^{n} \frac{\mathbf{A}^{\otimes k}}{k!}$.

2.1 Rough paths

More generally, a path of bounded p-variation with values in $G^{\lfloor p \rfloor}(E)$ has a canonical lift to a path with values in $G(E)$. In the sequel, denote by Δ_T the simplex $\Delta_T := \{(s,t) \in [0,T]^2 \mid 0 \leq s \leq t \leq T\}$ and we say that a continuous map $\omega : \Delta_T \to [0,+\infty)$ is a *control function* if it is super-additive, that is, $\omega(s,t) + \omega(t,u) \leq \omega(s,u)$ for all $0 \leq s \leq t \leq u \leq T$.

Definition 3 (Multiplicative functional)**.** *Let $n \geq 1$ be an integer. Let $\mathbb{X} : \Delta_T \to T^n(E)$ be a continuous map. For each $(s,t) \in \Delta_T$, denote by $\mathbb{X}_{s,t}$ the image by $\mathbb{X}$ of (s,t) and write*

$$\mathbb{X}_{s,t} = (x^0_{s,t}, x^1_{s,t}, \ldots, x^n_{s,t}) \in T^n(E).$$

The function $\mathbb{X}$ is called a multiplicative functional of degree n in E if
(i) $x^0_{s,t} = 1$ for all $(s,t) \in \Delta_T$ and
(ii) Chen's identity holds, that is,

$$\mathbb{X}_{s,u} \otimes \mathbb{X}_{u,t} = \mathbb{X}_{s,t}, \quad \forall s,t,u \in [0,T], \quad s \leq u \leq t.$$

Given a path $\mathbf{X}_t = (1, x^1_t, \ldots, x^n_t)$ in $T^n(E)$ we say that $\mathbb{X}_{s,t} = \mathbf{X}_s^{-1} \otimes \mathbf{X}_t$ is the multiplicative functional determined by $\mathbf{X}$. Conversely, given a multiplicative functional $\mathbb{X}_{s,t}$ and a point $\mathbf{X}_0 \in T^n(E)$, we say that $\mathbf{X}_t = \mathbf{X}_0 \otimes \mathbb{X}_{0,t}$ is the path starting at $\mathbf{X}_0$ determined by $\mathbb{X}_{s,t}$. The p-variation over $[s,t]$ of a multiplicative functional $\mathbb{X}$ of degree $\lfloor p \rfloor$ is defined by

$$\|\mathbb{X}\|_{p-\mathrm{var},[s,t]} := \sup_{D \subset [0,T]} \max_{1 \leq k \leq \lfloor p \rfloor} \left(\sum_{t_i \in D} \|x^k_{s,t}\|^{p/k} \right)^{1/p} \tag{5}$$

where the supremum is taken over all finite partitions D of $[s,t]$.

Definition 4 (p-rough path)**.** *Let $p \geq 1$ be a real number. A p-rough path $\mathbb{X}$ in E is a multiplicative functional of degree $\lfloor p \rfloor$ in E with finite p-variation.*

Theorem 1 (Thm. 2.2.1 in [15])**.** *Let $p \geq 1$ be a real number and $n \geq 1$ an integer. Let $\mathbb{X} : \Delta_T \to T^n(E)$ be a multiplicative functional with finite p-variation controlled by a control w and assume that $n \geq \lfloor p \rfloor$. Then there exists a unique extension of $\mathbb{X}$ to a multiplicative functional $\Delta_T \to T((E))$ which possesses finite p-variation. More precisely, for every $m \geq p+1$, there exists a unique continuous function $x^m : \Delta_T \to E^{\otimes m}$ such that*

$$(s,t) \mapsto (1, x^1_{s,t}, \ldots, x^{\lfloor p \rfloor}_{s,t}, \ldots, x^m_{s,t}, \ldots) \in T((E))$$

is a multiplicative functional with finite p-variation controlled by ω. By this we mean that $\|x^i_{s,t}\| < \frac{\omega(s,t)^{i/p}}{\beta_p (i/p)!}$ $\forall i \geq 1$ and $\forall (s,t) \in \Delta_T$ where $\beta_p = p^2 \left(1 + \sum_{r=3}^{\infty} \frac{2}{r-2}^{(\lfloor p \rfloor + 1)/p}\right)$.

If a p-rough path $\mathbb{X}$ controlled by w takes its values in $G^n(E)$, we say it is *weakly geometric*, and we call its unique extension its signature $S(\mathbb{X}) : (s,t) \mapsto (1, x^1_{s,t}, \ldots, x^{\lfloor p \rfloor}_{s,t}, \ldots, x^m_{s,t}, \ldots)$.

Definition 5 (Signature kernel). *Let $p \geq 1$ and $q \geq 1$ be two real numbers. The signature kernel is the map defined for any two weakly geometric p- and q-rough paths $\mathbb{X}, \mathbb{Y}$ by*

$$\kappa(\mathbb{X}, \mathbb{Y}) = \langle S(\mathbb{X}), S(\mathbb{Y}) \rangle_{\mathcal{H}} \tag{6}$$

This kernel is well-defined [36] since $\langle S(\mathbb{X})_{s_1,s_2}, S(\mathbb{Y})_{t_1,t_2} \rangle_{\mathcal{H}} = \sum_{k=0}^{\infty} \langle x^k_{s_1,s_2}, y^k_{t_1,t_2} \rangle_k$ which is bounded by $\sum_{k=0}^{\infty} \|x^k_{s_1,s_2}\| \|y^k_{t_1,t_2}\| \leq \sum_{k=0}^{\infty} \frac{\omega_{\mathbb{X}}(s_1,s_2)^{k/p} \cdot \omega_{\mathbb{Y}}(t_1,t_2)^{k/q}}{\beta_p(k/p)! \beta_q(k/q)!} < +\infty$.

2.2 Smooth rough paths

We now recall the definition of smooth rough paths. This section is based on [39].

Definition 6 (m-smooth geometric rough path). *A level-m smooth geometric rough path (m-sgrp) over E is any path $\mathbf{X} : [0,T] \to G^m(E)$ such that for any word w of length $|w| \leq m$, the map $t \mapsto \langle \mathbf{X}_t, w \rangle$ is smooth. A smooth geometric rough path (sgrp) over E is any path $\mathbf{X} : [0,T] \to G(E)$ such that for any word w, the map $t \mapsto \langle \mathbf{X}_t, w \rangle$ is smooth.*

Definition 7 (m-smooth geometric rough model). *A level-m smooth geometric rough model (m-sgrm) over E is any non-zero map $\mathbb{X} : \Delta_T \to G^m(E)$ such that*
(i) Chen's relation holds, that is, for any $s, t, u \in [0,T]$ such that $s \leq u \leq t$

$$\mathbb{X}_{s,u} \otimes_m \mathbb{X}_{u,t} = \mathbb{X}_{s,t}$$

(ii) For any word w of length $|w| \leq m$, the map $t \mapsto \langle \mathbb{X}_{s,t}, w \rangle$ is smooth for one $s \in [0,T]$.
A smooth geometric rough model (in short: sgrm) is a map with values in $G(E)$ such that (i) and (ii) hold with all restrictions on the word's length omitted and $\otimes_m$ replaced with $\otimes$.

Any m-sgrp $\mathbf{X} : [0,T] \to G^m(E)$ induces an m-sgrm $\mathbb{X} : \Delta_T \to G^m(E)$ defined by $\mathbb{X}_{s,t} = \mathbf{X}_s^{-1} \otimes_m \mathbf{X}_t$. Conversely, an m-sgrm $\mathbb{X}$ induces an m-sgrp $\mathbf{X} : [0,T] \to G^m(E)$ with $\mathbf{X}_t = \mathbb{X}_{0,t}$.

Definition 8 (Extension of m-sgrp). *A sgrp $\mathbf{Z}$ is called extension of some m-sgrp $\mathbf{X}$ if*

$$\langle \mathbf{Z}_t, w \rangle = \langle \mathbf{X}_t, w \rangle, \quad \textit{for all } t \in [0,T] \textit{ and word } w \textit{ of length } |w| \leq m$$

if this holds for a m'-sgrp $\mathbf{Z}$ with $m < m' < \infty$, we call is m'-extension of $\mathbf{X}$.

Definition 9 (Extension of m-sgrm). *A sgrm $\mathbb{Z}$ is called extension of some m-sgrm $\mathbb{X}$ if*

$$\langle \mathbb{Z}_{s,t}, w \rangle = \langle \mathbb{X}_{s,t}, w \rangle, \quad \textit{for all } s,t \in [0,T] \textit{ and word } w \textit{ of length } |w| \leq m$$

if this holds for a m'-sgrm $\mathbb{Z}$ with $m < m' < \infty$, we call is m'-extension of $\mathbb{X}$.

Whenever $\dim(E) > 1$, extensions are non unique. However, analogously to [15, Thm. 2.2.1], one can enforce a condition to guarantee the uniqueness of the extension [39].

Definition 10 (Diagonal derivative and signature of an m-smooth rough path). *Given an m-sgrm $\mathbb{X}$ for some $m \in \mathbb{N}$, there exists a unique sgrm extension $\mathbb{Z}$ of $\mathbb{X}$ which is minimal in the sense that for all $s \in [0,T]$ one has*

$$\dot{\boldsymbol{z}}_s := \partial_h|_{h=0} \mathbb{Z}_{s,s+h} \in \mathcal{L}^m(\mathbb{E}) \subset \mathcal{L}((E)).$$

We call $\dot{\boldsymbol{z}}$ the diagonal derivative of $\mathbb{Z}$. It satisfies $\dot{\boldsymbol{z}}_s = \dot{\boldsymbol{x}}_s$. For a fixed interval $[s,t] \subset [0,T]$, $\mathbb{Z}_{s,t}$ only depends on $\{\mathbb{X}_{u,v} : s \leq u \leq v \leq t\}$ and the signature of $\mathbb{X}$ on $[s,t]$ is defined by

$$S(\mathbb{X})_{s,t} = \mathbb{Z}_{s,t} \in G(E). \tag{7}$$

To compute the signature of an m-sgrp $\mathbb{X}$, it suffices to solve

$$\dot{\mathbf{Z}}_t = \mathbf{Z}_t \otimes \dot{\boldsymbol{x}}_t, \quad \text{started at } \mathbf{Z}_0 = \mathbf{1} \in G(E). \tag{8}$$

3 Main results

Before stating the first result of the article, we recall the definitions of the adjoints of the linear maps of left and right tensor multiplication by an element of $\mathcal{H}$.

Definition 11 (Adjoint of left tensor multiplication)**.** *Let* $\mathbf{A}, \mathbf{B}, \mathbf{C}$ *be three elements of* $\mathcal{H}$*. Denote by* $\ell(\mathbf{A}) : \mathcal{H} \to \mathcal{H}$ *the adjoint of left multiplication by* $\mathbf{A}$*, defined by*

$$\langle \mathbf{C}, \mathbf{A} \otimes \mathbf{B} \rangle = \langle \ell(\mathbf{A})(\mathbf{C}), \mathbf{B} \rangle, \tag{9}$$

which can be written as $\ell(\mathbf{A})(\mathbf{C}) = \sum_u \langle \mathbf{A}, e_u \rangle \sum_v \langle \mathbf{C}, e_{uv} \rangle e_v$*.*

Definition 12 (Adjoint of right tensor multiplication)**.** *Let* $\mathbf{A}, \mathbf{B}, \mathbf{C}$ *be three elements of* $\mathcal{H}$*. Denote by* $r(\mathbf{B}) : \mathcal{H} \to \mathcal{H}$ *the adjoint of right multiplication by* $\mathbf{B}$*, defined by*

$$\langle \mathbf{C}, \mathbf{A} \otimes \mathbf{B} \rangle = \langle r(\mathbf{B})(\mathbf{C}), \mathbf{A} \rangle, \tag{10}$$

which can be written as $r(\mathbf{B})(\mathbf{C}) = \sum_v \langle \mathbf{B}, e_v \rangle \sum_u \langle \mathbf{C}, e_{uv} \rangle e_u$*.*

Proposition 2. *Let* $\mathbf{A}$ *be a tensor in* $\mathcal{H}$*. Let* $b \in E^{\otimes p}$ *and* $c \in E^{\otimes q}$ *with* $p \leq q$*. One has*

$$r(\mathbf{B})(\mathbf{A} \otimes \mathbf{C}) = \mathbf{A} \otimes r(\mathbf{B})(\mathbf{C}) \tag{11}$$

where $\mathbf{B}$ *and* $\mathbf{C}$ *are the embeddings of* b *and* c *into* $\mathcal{H}$*.*

In the following section, we will establish that the signature kernel $\langle S(\mathbb{X}), S(\mathbb{Y}) \rangle_{\mathcal{H}}$ of two m- amd n-smooth rough paths $\mathbb{X}$ and $\mathbb{Y}$ solves an augmented system of PDEs which generalises the original Goursat problem (retrieved when $m = n = 1$). In the general case, the dynamics of the kernel might be influenced by the first degrees of extra state defined in terms of $\ell(S(\mathbb{Y}))(S(\mathbb{X}))$ and $\ell(S(\mathbb{X}))(S(\mathbb{Y}))$. This will be the case when at least one of $\mathbb{X}$ or $\mathbb{Y}$ is not given by the minimal extension of a 1-smooth rough path. We note that the first scalar components of $\ell(S(\mathbb{Y}))(S(\mathbb{X}))$ and $\ell(S(\mathbb{X}))(S(\mathbb{Y}))$ are equal to the kernel $\langle S(\mathbb{X}), S(\mathbb{Y}) \rangle_{\mathcal{H}}$.

3.1 The signature kernel of smooth rough paths solves a PDE

We choose to work in the smooth setting of [39], as this allows us to use classical calculus and focus on linear algebra aspects to extend the results from [36].

Theorem 3. *Let* $\mathbb{X}$ *and* $\mathbb{Y}$ *be two* m*- and* n*-smooth geometric rough paths on the intervals* $I = [a, b]$ *and* $J = [c, d]$ *respectively. Let* $\dot{\boldsymbol{x}}_s$ *and* $\dot{\boldsymbol{y}}_t$ *be their diagonal derivatives,*

$$\dot{\boldsymbol{x}}_s := \partial_v|_{v=s} \mathbb{X}_{s,v} \in \mathcal{L}^m(E)$$
$$\dot{\boldsymbol{y}}_t := \partial_v|_{v=t} \mathbb{Y}_{t,v} \in \mathcal{L}^n(E)$$

which are valued in the Lie polynomials of degree less equal m *and* n *respectively.*
The real-valued functions indexed on the plane $u : I \times J \to \mathbb{R}$ *and the two tensor-valued functions indexed on the plane* $\phi^n : I \times J \to \pi_n(T^{>0}(E))$ *and* $\psi^m : I \times J \to \pi_m(T^{>0}(E))$ *defined for all* $(s, t) \in I \times J$ *by*

$$u(s,t) := \langle S(\mathbb{X})_{a,s}, S(\mathbb{Y})_{c,t} \rangle_{\mathcal{H}}$$
$$\phi^n(s,t) := \pi_n \left(\ell(S(\mathbb{Y})_{c,t})(S(\mathbb{X})_{a,s}) - u(s,t) \cdot \mathbf{1} \right)$$
$$\psi^m(s,t) := \pi_m \left(\ell(S(\mathbb{X})_{a,s})(S(\mathbb{Y})_{c,t}) - u(s,t) \cdot \mathbf{1} \right)$$

solve the following system of linear PDE

$$\frac{\partial^2 u}{\partial s \partial t} = u \cdot \langle \dot{\boldsymbol{x}}_s, \dot{\boldsymbol{y}}_t \rangle + \langle \phi^n, r(\dot{\boldsymbol{x}}_s)(\dot{\boldsymbol{y}}_t) \rangle + \langle \psi^m, r(\dot{\boldsymbol{y}}_t)(\dot{\boldsymbol{x}}_s) \rangle \tag{12}$$

$$\frac{\partial \phi^n}{\partial s} = u \cdot \pi_n(\dot{\boldsymbol{x}}_s) + \phi^n \otimes_n \pi_n(\dot{\boldsymbol{x}}_s) + \ell(\psi^m)(\dot{\boldsymbol{x}}_s) - \langle \psi^m, \pi_m(\dot{\boldsymbol{x}}_s)\rangle \cdot \mathbf{1}_n \tag{13}$$

$$\frac{\partial \psi^m}{\partial t} = u \cdot \pi_m(\dot{\boldsymbol{y}}_t) + \psi^m \otimes_m \pi_m(\dot{\boldsymbol{y}}_t) + \ell(\phi^n)(\dot{\boldsymbol{y}}_t) - \langle \phi^n, \pi_n(\dot{\boldsymbol{y}}_t)\rangle \cdot \mathbf{1}_m \tag{14}$$

with boundary conditions

$$u(0,t) = 1,\ \psi(0,t) = \pi_m(\mathbb{Y}_{c,t} - \mathbf{1}),\ \phi(0,t) = \mathbf{0}_n,$$
$$u(s,0) = 1,\ \phi(s,0) = \pi_n(\mathbb{X}_{a,s} - \mathbf{1}),\ \psi(s,0) = \mathbf{0}_m.$$

For notational convenience we used $\langle \dot{\boldsymbol{x}}_s, \dot{\boldsymbol{y}}_t\rangle := \sum_{k=1}^{m\wedge n} \langle \dot{x}_s^k, \dot{y}_t^k\rangle_k$, the arguments of the adjoints are canonically embedded into $\mathcal{H}$, and we don't write that $r(\dot{\boldsymbol{x}}_s)(\dot{\boldsymbol{y}}_t)$ and $r(\dot{\boldsymbol{y}}_t)(\dot{\boldsymbol{x}}_s)$ are projected onto $T^n(E)$ and $T^m(E)$ respectively in eq. (20).

Remark 1. This theorem states that the signature kernel of two m- and n-smooth geometric rough paths is the first component of a system of linear PDEs. Furthermore, the additional variables ϕ^n and ψ^m take their values in truncated tensor algebras. Note that in eq. (20), the PDE for the kernel, the n^{th} degree of $r(\dot{\boldsymbol{x}}_s)(\dot{\boldsymbol{y}}_t)$ and the m^{th} degree of $r(\dot{\boldsymbol{y}}_t)(\dot{\boldsymbol{x}}_s)$ are both zero. Therefore, the inner products involving them can be rewritten as

$$\langle \phi^n, r(\dot{\boldsymbol{x}}_s)(\dot{\boldsymbol{y}}_t)\rangle = \langle \phi^{n-1}, \pi_{n-1}(r(\dot{\boldsymbol{x}}_s)(\dot{\boldsymbol{y}}_t))\rangle$$
$$\langle \psi^m, r(\dot{\boldsymbol{y}}_t)(\dot{\boldsymbol{x}}_s)\rangle = \langle \psi^{m-1}, \pi_{m-1}(r(\dot{\boldsymbol{y}}_t)(\dot{\boldsymbol{x}}_s))\rangle.$$

This means that the dynamics of u only depend on degrees of the adjoint variables lower than $n-1$ and $m-1$ respectively. This is in line with the previous results on signature kernels, since for two 1-smooth rough paths, with diagonal derivatives $\dot{\boldsymbol{x}}_s = (0, \dot{x}_s)$ and $\dot{\boldsymbol{y}}_t = (0, \dot{y}_t)$ the reader can check that the system, written component-wise, would read as

$$\frac{\partial^2 u}{\partial s \partial t} = u \cdot \langle \dot{x}_s, \dot{y}_t\rangle_1 \tag{15}$$

$$\frac{\partial \phi^0}{\partial s} = 0 \tag{16}$$

$$\frac{\partial \psi^0}{\partial t} = 0 \tag{17}$$

$$\frac{\partial \phi^{(k)}}{\partial s} = u \cdot \dot{x}_s^{(k)} \qquad \text{for all } k \in \{1, \dots, d\} \tag{18}$$

$$\frac{\partial \psi^{(k)}}{\partial t} = u \cdot \dot{y}_t^{(k)} \qquad \text{for all } k \in \{1, \dots, d\}. \tag{19}$$

Since the states $\phi^{(k)}$ and $\psi^{(k)}$ do not influence the dynamics of u, to compute the signature kernel, we can discard the equations for the adjoint variables of degree greater than 1 and solve eq. (15) which corresponds to solving the original Goursat PDE problem eq. (3).

Remark 2. If we consider the minimal extensions of $\mathbb{X}$ and $\mathbb{Y}$ to two η-smooth geometric rough paths for some $\eta > \max\{m, n\}$, then their diagonal derivatives would both take their values in $\mathcal{L}^\eta(E)$. For notational convenience we keep denoting them by $\dot{\boldsymbol{x}}_s$ and $\dot{\boldsymbol{y}}_t$. Then, their signature kernel and signature adjoints solve the following system

$$\frac{\partial^2 u}{\partial s \partial t} = u \cdot \langle \dot{\boldsymbol{x}}_s, \dot{\boldsymbol{y}}_t\rangle + \langle \phi^\eta, r(\dot{\boldsymbol{x}}_s)(\dot{\boldsymbol{y}}_t)\rangle + \langle \psi^\eta, r(\dot{\boldsymbol{y}}_t)(\dot{\boldsymbol{x}}_s)\rangle \tag{20}$$

$$\frac{\partial \phi^\eta}{\partial s} = u \cdot \pi_\eta(\dot{\boldsymbol{x}}_s) + \phi^\eta \otimes_\eta \pi_\eta(\dot{\boldsymbol{x}}_s) + \ell(\psi^\eta)(\dot{\boldsymbol{x}}_s) - \langle \psi^\eta, \pi_\eta(\dot{\boldsymbol{x}}_s)\rangle \cdot \mathbf{1}_\eta \tag{21}$$

$$\frac{\partial \psi^\eta}{\partial t} = u \cdot \pi_\eta(\dot{\boldsymbol{y}}_t) + \psi^\eta \otimes_\eta \pi_\eta(\dot{\boldsymbol{y}}_t) + \ell(\phi^\eta)(\dot{\boldsymbol{y}}_t) - \langle \phi^\eta, \pi_\eta(\dot{\boldsymbol{y}}_t)\rangle \cdot \mathbf{1}_\eta \tag{22}$$

complemented by appropriate boundary conditions. As $\eta > \max\{m, n\}$, we have more equations than in theorem 3. However, for $\dot{\boldsymbol{x}}_s$ all degrees greater than $m+1$ are zero, and similarly for $\dot{\boldsymbol{y}}_t$, all degrees greater than $n+1$ are zero. If we project the solution of this system $U^\eta = (u, \phi^\eta, \psi^\eta)$ and consider $U^{\text{proj}} = (u, \pi_n(\phi^\eta), \pi_m(\psi^\eta))$ we have $U^{\text{proj}} = U^{n,m}$, where $U^{n,m} := (u, \phi^n, \psi^m)$ denotes the solution of the system in theorem 3. In other words, when minimally extending the input smooth rough paths to higher degrees, we write a bigger set of equations, but if we solve them and project the solution, we get the same state as if we had solved the reduced set of equations, which shows consistency. To fix ideas, let's continue the example in the previous remark. If we extend the 1-sgrp to 2-sgrp, then we get

$$\frac{\partial^2 u}{\partial s \partial t} = u \cdot \langle \dot{x}_s, \dot{y}_t \rangle_1 \tag{23}$$

$$\frac{\partial \phi^{(k)}}{\partial s} = u \cdot \dot{x}_s^{(k)} \qquad \text{for all } k \in \{1, \ldots, d\} \tag{24}$$

$$\frac{\partial \psi^{(k)}}{\partial t} = u \cdot \dot{y}_t^{(k)} \qquad \text{for all } k \in \{1, \ldots, d\} \tag{25}$$

$$\frac{\partial \phi^{(k,p)}}{\partial s} = \phi^{(k)} \cdot \dot{x}_s^{(p)} \qquad \text{for all } k, p \in \{1, \ldots, d\} \tag{26}$$

$$\frac{\partial \psi^{(k,p)}}{\partial t} = \psi^{(k)} \cdot \dot{y}_t^{(p)} \qquad \text{for all } k, p \in \{1, \ldots, d\} \tag{27}$$

For 1-sgrps or any minimal extensions of 1-sgrps, the dynamics of u are independent from any other state. Otherwise, we need more equations to get u.

Remark 3. Let $\mathbb{X}$ and $\mathbb{Y}$ be two good sgrps with diagonal derivatives

$$\dot{\boldsymbol{x}}_s := \partial_v|_{v=s} \mathbb{X}_{s,v} \in \mathcal{L}(E)$$
$$\dot{\boldsymbol{y}}_t := \partial_v|_{v=t} \mathbb{Y}_{t,v} \in \mathcal{L}(E)$$

valued in the space of Lie polynomials $\mathcal{L}(E)$. The real-valued function indexed on the plane $u : I \times J \to \mathbb{R}$ and the two tensor-valued functions $\phi, \psi : I \times J \to T^{>0}(E)$ defined for all $(s,t) \in I \times J$ by

$$u(s,t) := \langle S(\mathbb{X})_{a,s}, S(\mathbb{Y})_{c,t} \rangle_{\mathcal{H}}$$
$$\phi(s,t) := \ell(S(\mathbb{Y})_{c,t})(S(\mathbb{X})_{a,s}) - u(s,t) \cdot \mathbf{1}$$
$$\psi(s,t) := \ell(S(\mathbb{X})_{a,s})(S(\mathbb{Y})_{c,t}) - u(s,t) \cdot \mathbf{1}$$

solve the following system of linear PDE

$$\frac{\partial^2 u}{\partial s \partial t} = u \cdot \langle \dot{\boldsymbol{x}}_s, \dot{\boldsymbol{y}}_t \rangle_{\mathcal{H}} + \langle \phi, r(\dot{\boldsymbol{x}}_s)(\dot{\boldsymbol{y}}_t) \rangle_{\mathcal{H}} + \langle \psi, r(\dot{\boldsymbol{y}}_t)(\dot{\boldsymbol{x}}_s) \rangle_{\mathcal{H}} \tag{28}$$

$$\frac{\partial \phi}{\partial s} = u \cdot \dot{\boldsymbol{x}}_s + \phi \otimes \dot{\boldsymbol{x}}_s + \ell(\psi)(\dot{\boldsymbol{x}}_s) - \langle \psi, \dot{\boldsymbol{x}}_s \rangle_{\mathcal{H}} \cdot \mathbf{1} \tag{29}$$

$$\frac{\partial \psi}{\partial t} = u \cdot \dot{\boldsymbol{y}}_t + \psi \otimes \dot{\boldsymbol{y}}_t + \ell(\phi)(\dot{\boldsymbol{y}}_t) - \langle \phi, \dot{\boldsymbol{y}}_t \rangle_{\mathcal{H}} \cdot \mathbf{1} \tag{30}$$

with boundary conditions

$$u(0,t) = 1, \ \psi(0,t) = \mathbb{Y}_{c,t} - \mathbf{1}, \ \phi(0,t) = \mathbf{0},$$
$$u(s,0) = 1, \ \phi(s,0) = \mathbb{X}_{a,s} - \mathbf{1}, \ \psi(s,0) = \mathbf{0}.$$

As we consider *good* smooth rough paths, there is some $\eta \in \mathbb{N}$ so that $\dot{x}^k \equiv \dot{y}^k \equiv 0$ for all $k > \eta$, and there is a particular structure in the coupling of the equations. As aforementioned, the dynamics for u only depend on the degrees $k = 1, \ldots, \eta$ of ϕ and ψ because the action of the adjoint $r(\dot{\boldsymbol{x}}_s)$ and $r(\dot{\boldsymbol{y}}_t)$ can only decrease the maximum degree (in fact there are also independent from the degree η of ϕ and ψ, because the degree η of $r(\dot{\boldsymbol{x}}_s)(\dot{\boldsymbol{y}}_t)$ is given by $r(\dot{x}_s^0)(\dot{y}_t^\eta) = r(0)(\dot{y}_t^\eta) = 0$). Futhermore, the system $(u, \phi^\eta, \psi^\eta)$ is closed because the dynamics of ϕ^η and ψ^η do not depend on degrees strictly higher than η of ϕ and ψ.

3.2 Numerical Method: Application to Piecewise Log-Linear Paths

Having shown that the signature kernel of two smooth rough paths $\mathbb{X}$ and $\mathbb{Y}$, augmented with additional bilinear functions in $S(\mathbb{X})$ and $S(\mathbb{Y})$ solves a system of linear PDEs, we now explain how this result provides a high order numerical method for computing signature kernels of p-rough paths, including those which solve the Goursat problem eq. (3). The idea is to approximate each input path by piecewise log-linear paths, which are piecewise smooth rough paths, and generalize the classical piecewise linear approximation.

By reparametrising the paths by $\tau(t) = Cw(0,t)$ with $C \in (0,+\infty)$ we can assume that they are controlled by $C^{-1}(t-s)$.

Definition 13 (piecewise log-linear approximation)**.** *Let $p \geq 1$ be a real number and let $\mathbb{X}$ be a p-rough path. Let $D = \{0 = t_1 \leq t_2 \leq \ldots \leq t_n = T\}$ be a partition of $[0,T]$, and $m \geq \lfloor p \rfloor$ an integer. We call $\mathbb{X}^{m,D}$ the piecewise log-linear approximation of $\mathbb{X}$ of Lie degree m on D, the piecewise log-linear path defined for any $t_i \leq s < t \leq t_{i+1}$ for any $t_i, t_{i+1} \in D$ by*

$$\mathbb{X}^{m,D}_{s,t} = \exp_m\left(\frac{t-s}{t_{i+1}-t_i}\log_m(\mathbb{X}_{t_i,t_{i+1}})\right) \tag{31}$$

and extended to be multiplicative on $s,t \in [0,T]$.

Note that the paths $t \mapsto \pi_m(S(\mathbb{X})_{0,t})$ and $t \mapsto \mathbb{X}^{m,D}_{0,t}$ agree on D when $m = \lfloor p \rfloor$. Furthermore, for all $t \in [t_i, t_{i+1}]$, the diagonal derivative of $\mathbb{X}^{m,D}_{s,t}$ is expressed in terms of the log-signature

$$\dot{\boldsymbol{x}}_t := \partial_v|_{v=t}\mathbb{X}^{m,D}_{t,v} = \frac{1}{t_{i+1}-t_i}\cdot\log_m(\mathbb{X}_{t_i,t_{i+1}}) \tag{32}$$

In Lemma 3, we show that $\mathbb{X}^{\lfloor p \rfloor,D}_{s,t}$ is controlled by $t-s$. Furthermore, its signature is given by

$$S(\mathbb{X}^{m,D})_{s,t} = \exp\left(\frac{t-s}{t_{i+1}-t_i}\cdot\log_m(\mathbb{X}_{t_i,t_{i+1}})\right)$$

for any $t_i \leq s < t \leq t_{i+1}$ for any $t_i, t_{i+1} \in D$. Importantly, the results from Section 3.1 apply to these piecewise log-linear paths, which are special types of piecewise smooth rough paths. The explicit form of the diagonal derivatives gives us the coefficients of the PDE. We also have the following bound (proved in Appendix B) on the approximation of the signature of a p-rough path by the signature of a piecewise log-linear approximation.

Lemma 1. *Let $\mathbb{X}$ be a weakly geometric p-rough path defined on $[0,T]$ with values in E. Let $D = \{0 = t_0 < \ldots < t_n = T\}$ be a regular partition of $[0,T]$ and $m = \lfloor p \rfloor$. Denote by $\mathbb{X}^{m,D}$ the piecewise log-linear approximation of $\mathbb{X}$ of degree m on D. We have*

$$\|S(\mathbb{X})_{0,T} - S(\mathbb{X}^{m,D})_{0,T}\| \leq C_p w(0,T)\exp(w(0,T);p)^3\left(\frac{w(0,T)}{n}\right)^{\frac{m+1}{p}-1}$$

where C_p only depends on p and $\exp(a;p) := \sum_{k=0}^{\infty} a^{k/p}/(k/p)!$.

A similar type of estimate can be obtained for the inner product of signatures, that is, for the signature kernel. Let $\mathbb{X}$ be a weakly geometric p-rough path on $[0,S]$ and $\mathbb{Y}$ a weakly geometric q-rough path on $[0,T]$. Let D and D' be a partition of $[0,S]$ and $[0,T]$ respectively. Assume $m = \lfloor p \rfloor$ and $n = \lfloor q \rfloor$. It suffices to use the identity

$$\langle S(\mathbb{X}), S(\mathbb{Y})\rangle - \langle S(\mathbb{X}^{m,D}), S(\mathbb{Y}^{n,D'})\rangle = \langle S(\mathbb{X}) - S(\mathbb{X}^{m,D}), S(\mathbb{Y}^{n,D'})\rangle + \langle S(\mathbb{Y}) - S(\mathbb{Y}^{n,D'}), S(\mathbb{X})\rangle,$$

and then apply Cauchy-Schwarz

$$\begin{aligned}\left|\kappa(\mathbb{X},\mathbb{Y}) - \langle S(\mathbb{X}^{m,D}), S(\mathbb{Y}^{n,D'})\rangle\right| &\leq \left|\langle S(\mathbb{X}) - S(\mathbb{X}^{m,D}), S(\mathbb{Y}^{n,D'})\rangle\right| + \left|\langle S(\mathbb{Y}) - S(\mathbb{Y}^{n,D'}), S(\mathbb{X})\rangle\right| \\ &\leq \|S(\mathbb{X}) - S(\mathbb{X}^{m,D})\| \cdot \|S(\mathbb{Y}^{n,D'})\| + \|S(\mathbb{Y}) - S(\mathbb{Y}^{n,D'})\| \cdot \|S(\mathbb{X})\|\end{aligned}$$

The final step to produce a numerical scheme for approximating the signature kernel is to discretize the PDEs. To maintain the rate of convergence, this discretization must yield an accurate enough approximation of the solution. A discretization method is given in Algorithm 1, where for simplifity, we take $m = n$ and a single step for each PDE. We leave as future work the verification that the chosen discretization ensures that the convergence rate remains optimal.

Algorithm 1 PDE discretization method

1: **Input:** Truncated log-signatures over a partition $\dot{\boldsymbol{x}}_i \in \mathcal{L}^m(\mathbb{R}^d)$ for $i = 1, \dots, N_x$ and $\dot{\boldsymbol{y}}_j \in \mathcal{L}^m(\mathbb{R}^d)$ for $j = 1, \dots, N_y$. Initial conditions $u_{0,\cdot}, u_{\cdot,0}, \phi_{0,\cdot}, \phi_{\cdot,0}, \psi_{0,\cdot}, \psi_{\cdot,0}$.
2: **for** i in $0, \dots, N_x - 1$ **do**
3:     **for** j in $0, \dots, N_y - 1$ **do**
4:         $\dot{\boldsymbol{x}}_i^{\text{proj}}, \dot{\boldsymbol{y}}_j^{\text{proj}} \leftarrow \pi_{m-1}(\dot{\boldsymbol{x}}_i), \pi_{m-1}(\dot{\boldsymbol{y}}_j)$
5:         *// Update the adjoint states*
6:         $\phi_{i+1,j+1} \leftarrow \phi_{i,j+1} + u_{i,j} \cdot \dot{\boldsymbol{x}}_i^{\text{proj}} + \phi_{i,j+1} \otimes \dot{\boldsymbol{x}}_i^{\text{proj}} + \ell(\psi_{i,j+1})(\dot{\boldsymbol{x}}_i) - \langle \psi_{i,j+1}, \dot{\boldsymbol{x}}_i^{\text{proj}} \rangle$
7:         $\psi_{i+1,j+1} \leftarrow \psi_{i+1,j} + u_{i,j} \cdot \dot{\boldsymbol{y}}_j^{\text{proj}} + \psi_{i+1,j} \otimes \dot{\boldsymbol{y}}_j^{\text{proj}} + \ell(\phi_{i+1,j})(\dot{\boldsymbol{y}}_j) - \langle \phi_{i+1,j}, \dot{\boldsymbol{y}}_j^{\text{proj}} \rangle$
8:         *// Intermediate states*
9:         $f_1 \leftarrow u_{i,j} \cdot \langle \dot{\boldsymbol{x}}_i, \dot{\boldsymbol{y}}_j \rangle + \langle \phi_{i,j}, r(\dot{\boldsymbol{x}}_i)(\dot{\boldsymbol{y}}_j) \rangle + \langle \psi_{i,j}, r(\dot{\boldsymbol{y}}_j)(\dot{\boldsymbol{x}}_i) \rangle$
10:         $f_2 \leftarrow u_{i,j+1} \cdot \langle \dot{\boldsymbol{x}}_i, \dot{\boldsymbol{y}}_j \rangle + \langle \phi_{i,j+1}, r(\dot{\boldsymbol{x}}_i)(\dot{\boldsymbol{y}}_j) \rangle + \langle \psi_{i,j+1}, r(\dot{\boldsymbol{y}}_j)(\dot{\boldsymbol{x}}_i) \rangle$
11:         $f_3 \leftarrow u_{i+1,j} \cdot \langle \dot{\boldsymbol{x}}_i, \dot{\boldsymbol{y}}_j \rangle + \langle \phi_{i+1,j}, r(\dot{\boldsymbol{x}}_i)(\dot{\boldsymbol{y}}_j) \rangle + \langle \psi_{i+1,j}, r(\dot{\boldsymbol{y}}_j)(\dot{\boldsymbol{x}}_i) \rangle$
12:         $u^p \leftarrow u_{i+1,j} + u_{i,j+1} - u_{i,j} + f_1$
13:         $f_4 \leftarrow u^p \cdot \langle \dot{\boldsymbol{x}}_i, \dot{\boldsymbol{y}}_j \rangle + \langle \phi_{i+1,j+1}, r(\dot{\boldsymbol{x}}_i)(\dot{\boldsymbol{y}}_j) \rangle + \langle \psi_{i+1,j+1}, r(\dot{\boldsymbol{y}}_j)(\dot{\boldsymbol{x}}_i) \rangle$
14:         *// Update the kernel state*
15:         $u_{i+1,j+1} = u_{i+1,j} + u_{i,j+1} - u_{i,j} + (1./4) * (f_1 + f_2 + f_3 + f_4)$
16:     **end for**
17: **end for**
18: **return** u_{N_x,N_y}.

If the data takes the form of time series $x_1, \dots, x_n$ with $x_i \in \mathbb{R}^d$, one first needs to construct the higher order description. In this case, the input of Algorithm 1 might be obtained by embedding the data into a continuous path $x : [0,T] \to \mathbb{R}^d$ and then computing a sequence of log-signatures $\log_m S(x)_{s_i,s_{i+1}}$ over a partition $\{0 = s_0 \leq \ldots \leq s_{N_x} = T\}$. Such constructions are straightforward using Python packages such as `esig`, `iisignature`, `signatory`, `signax` [32–34, 48] or `RoughPy` [35]. It is worth noting that that for small degrees $m = 1$ and $m = 2$, the equations can still be described in terms of matrix-vector products and dot products, and the numerical schemes can be implemented using native Python and `NumPy` functions. However, packages designed for working with fundamental objects from free non-commutative algebra (see lines $6-14$ in Algorithm 1) such as `RoughPy`, significantly streamline the implementation and offer a seamless transition between different schemes. An implementation of the newly proposed methods for computation of signature kernels is made accessible at https://github.com/maudl3116/high-order-sigkernel.

4 Numerical Illustration

To fix ideas, let's consider an example where we have two multivariate Brownian motion sample paths W and V on $[0,T]$. Almost surely W and V have infinite two-variation and finite p-variation for every $p>2$. Therefore, it is a priori not possible to define their signature kernel as the solution of the PDE in eq. (3). However, the PDE is satisfied by the kernel of any two piecewise linear approximations $\boldsymbol{w}$ and $\boldsymbol{v}$ of W and V associated with any two partitions D and D' of $[0,T]$. The solution at final times $u(T,T)=u^{D,D'}(T,T)$ of

$$\frac{\partial^2 u}{\partial s \partial t}(s,t) = u(s,t)\cdot(\dot{\boldsymbol{w}}_s)^\top \dot{\boldsymbol{v}}_t \tag{33}$$

with boundary conditions $u(0,t)=1$ and $u(s,0)=1$, converges to the signature kernel of W and V as the mesh sizes of D and D' tend to 0, since the signatures of $\boldsymbol{w}$ and $\boldsymbol{v}$ converge to the Stratonovich signatures of W and V [20, Sec. 3.3.2]. Since $\boldsymbol{w}$ and $\boldsymbol{v}$ are piecewise linear, we solve a sequence of PDEs with constant coefficient of the form

$$\frac{\partial^2 u}{\partial s \partial t}(s,t) = u(s,t)\cdot c_{i,j} \tag{34}$$

for $s_i \leq s \leq s_{i+1}$ and $t_j \leq t \leq t_{j+1}$ where $c_{i,j} = W^\top_{s_i,s_{i+1}} V_{t_j,t_{j+1}}$. This procedure we have just described, corresponds to the case where we use a piecewise log-linear paths approximation of degree 1 for both input paths. For the first path (and similarly for the second) this approximation is given on each $[s_i, s_{i+1}]$ by

$$\dot{\boldsymbol{w}}(s) = (0, W_{s_i,s_{i+1}}) \in \mathbb{R}\oplus\mathbb{R}^d$$

Increasing the degree of the approximation consists in adding terms on top of the increment. For example, when $m=2$ (and similarly $n=2$) the approximation is given by

$$\dot{\boldsymbol{w}}(s) = (0, W_{s_i,s_{i+1}}, A_{s_i,s_{i+1}}) \in \mathbb{R}\oplus\mathbb{R}^d\oplus[\mathbb{R}^d,\mathbb{R}^d],$$

where the additional term is the Lévy area given for all $p,q=1,\ldots,d$ by

$$A^{(p,q)}_{s,s'} = \frac{1}{2}\int_{\sigma'=s}^{s'}\int_{\sigma=s}^{\sigma'} \circ dW^{(p)}(\sigma)\circ dW^{(q)}(\sigma') - \circ dW^{(q)}(\sigma)\circ dW^{(p)}(\sigma').$$

The signature kernel of these piecewise differentiable paths is the first component of the solution of the linear system of $2d+1$ partial differential equations

$$\begin{aligned}
\frac{\partial^2 u}{\partial s\partial t} &= u\cdot c_{i,j} + \sum_{k,p=1}^{d} \phi^{(k)}\cdot B^{(k,p)}_{t_j,t_{j+1}}\cdot W^{(p)}_{s_i,s_{i+1}} + \psi^{(k)}\cdot A^{(k,p)}_{s_i,s_{i+1}}\cdot V^{(p)}_{t_j,t_{j+1}}\\
\frac{\partial \phi^{(k)}}{\partial s} &= u\cdot W^{(k)}_{s_i,s_{i+1}} + \sum_{p=1}^{d}\psi^{(p)}\cdot A^{(p,k)}_{s_i,s_{i+1}}, \qquad \text{for } k=1,\ldots,d\\
\frac{\partial \psi^{(k)}}{\partial t} &= u\cdot V^{(k)}_{t_j,t_{j+1}} + \sum_{p=1}^{d}\phi^{(p)}\cdot B^{(p,k)}_{t_j,t_{j+1}} \qquad \text{for } k=1,\ldots,d
\end{aligned}$$

where the coefficient $c_{i,j}$ is now given by the inner product in $T^2(\mathbb{R}^d)$ of $\dot{\boldsymbol{w}}(s)$ and $\dot{\boldsymbol{v}}(t)$, i.e.

$$c_{i,j} := \sum_{k=1}^{d} W^{(p)}_{s_i,s_{i+1}} V^{(p)}_{t_j,t_{j+1}} + \sum_{k,p=1}^{d} A^{(p,k)}_{s_i,s_{i+1}} B^{(p,k)}_{t_j,t_{j+1}}.$$

The dynamics of $u(s,t)$ are now forced by the additional state variables $\phi^{(k)}(s,t)$ and $\psi^{(k)}(s,t)$ for $k=1,\ldots,d$.

Using the notion of piecewise log-linear approximation of degree m on a partition D we have unified the original PDE method for computing signature kernels [36] with the newly derived schemes in this article. Now, we want to compare these different schemes on this Brownian motion example. In each case, we will need an estimator of the error

$$\hat{E} = |\hat{u}^{\text{fine}}(1,1) - \hat{U}_0(1,1)| \tag{35}$$

where $\hat{u}^{\text{fine}}$ denotes the numerical solution (obtained with Algorithm 1) of the one-dimensional Goursat problem obtained by approximating the inputs by piecewise log-linear paths of degree 1 on a regular partition of small mesh size $\Delta t = 1/n$; and $\hat{U}_0$ is the first coordinate of numerical solution of the PDE system obtained by approximating the inputs by piecewise log-linear paths of degree m on a regular partition of mesh size $\Delta t = k/n$. For the experiment, we consider 2-dimensional Brownian motion, and construct the fine grid with $n = 1\,024$. Figure 1 shows the error as a function of the mesh size coarsening factor k for different degrees $m \in \{1, 2, 3, 4\}$. The experiment is repeated on 100 pairs of sample paths from $\mathbb{R}^2$-valued Brownian motion on the unit time interval $[0, 1]$, and the dots on Figure 1 correspond to the mean of the errors calculated according to eq. (35). We see that, for any fixed partitioning of the interval $[0, 1]$, the error decreases with the degree m. As expected, for any degree m, the error also decreases with the mesh size.

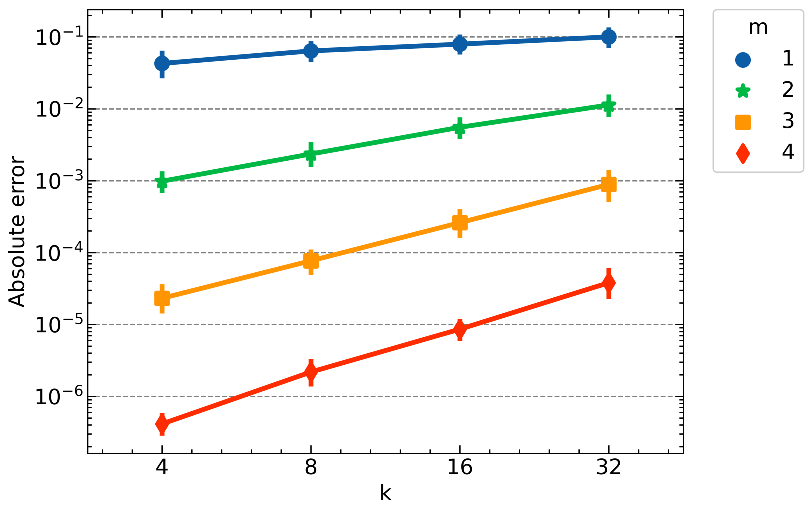

Fig. 1 Approximation of the signature kernel of two BM sample paths.

5 Conclusion

In this paper we show that the signature kernel of smooth rough paths solves a system of PDEs, thereby extending the results of [36] to a broader class of paths. Highly oscillatory inputs may now be described by piecewise log-linear paths which can be thought of a generalization of piecewise linear interpolations. In light of our result, their similarity measure can be computed by solving the associated PDEs using state-of-the-art software packages such as `RoughPy`.

Possible extensions of this work include developing similar schemes for so-called *weighted signature kernels* [19] and developing adaptive versions, where the partition size and the degree of the piecewise log-linear paths are adjusted instead of being fixed at the beginning, possibly building on [44].

Acknowledgements. This work was supported in part by EPSRC (NSFC) under Grant EP/S026347/1, in part by The Alan Turing Institute under the EPSRC grant EP/N510129/1, the Data Centric Engineering Programme (under the Lloyd's Register Foundation grant G0095), the Defence and Security Programme (funded by the UK Government) and the Office for National Statistics & The Alan Turing Institute (strategic partnership).

Declarations

References

[1] I. Steinwart, A. Christmann, *Support vector machines* (Springer Science & Business Media, 2008)

[2] H. Drucker, C.J. Burges, L. Kaufman, A. Smola, V. Vapnik, Support vector regression machines. Advances in neural information processing systems **9** (1996)

[3] C. Brouard, M. Szafranski, F. d'Alché Buc, Input output kernel regression: Supervised and semi-supervised structured output prediction with operator-valued kernels. Journal of Machine Learning Research **17**, np (2016)

[4] B. Schölkopf, J.C. Platt, J. Shawe-Taylor, A.J. Smola, R.C. Williamson, Estimating the support of a high-dimensional distribution. Neural computation **13**(7), 1443–1471 (2001)

[5] A. Gretton, K.M. Borgwardt, M.J. Rasch, B. Schölkopf, A. Smola, A kernel two-sample test. The Journal of Machine Learning Research **13**(1), 723–773 (2012)

[6] G. Wynne, A.B. Duncan, A kernel two-sample test for functional data. The Journal of Machine Learning Research **23**(1), 3159–3209 (2022)

[7] C.K. Williams, C.E. Rasmussen, *Gaussian processes for machine learning*, vol. 2 (MIT press Cambridge, MA, 2006)

[8] J. Lee, Y. Bahri, R. Novak, S.S. Schoenholz, J. Pennington, J. Sohl-Dickstein, Deep neural networks as gaussian processes. arXiv preprint arXiv:1711.00165 (2017)

[9] A. Jacot, F. Gabriel, C. Hongler, Neural tangent kernel: Convergence and generalization in neural networks. Advances in neural information processing systems **31** (2018)

[10] C.L. Li, W.C. Chang, Y. Cheng, Y. Yang, B. Póczos, Mmd gan: Towards deeper understanding of moment matching network. arXiv preprint arXiv:1705.08584 (2017)

[11] L. Pacchiardi, Statistical inference in generative models using scoring rules. Ph.D. thesis, University of Oxford (2022)

[12] T. Matsubara, J. Knoblauch, F.X. Briol, C.J. Oates, et al., Robust generalised bayesian inference for intractable likelihoods. Journal of the Royal Statistical Society Series B **84**(3), 997–1022 (2022)

[13] Y. Chen, B. Hosseini, H. Owhadi, A.M. Stuart, Solving and learning nonlinear pdes with gaussian processes. Journal of Computational Physics **447**, 110668 (2021)

[14] P. Batlle, M. Darcy, B. Hosseini, H. Owhadi, Kernel methods are competitive for operator learning. Journal of Computational Physics **496**, 112549 (2024)

[15] T.J. Lyons, Differential equations driven by rough signals. Revista Matemática Iberoamericana **14**(2), 215–310 (1998)

[16] I. Chevyrev, A. Kormilitzin, A primer on the signature method in machine learning. arXiv preprint arXiv:1603.03788 (2016)

[17] I. Perez Arribas, G.M. Goodwin, J.R. Geddes, T. Lyons, K.E. Saunders, A signature-based machine learning model for distinguishing bipolar disorder and borderline personality disorder. Translational psychiatry **8**(1), 274 (2018)

[18] A. Fermanian, Embedding and learning with signatures. Computational Statistics & Data Analysis **157**, 107148 (2021)

[19] T. Cass, T. Lyons, X. Xu, Weighted signature kernels. The Annals of Applied Probability **34**(1A), 585–626 (2024)

[20] T. Lyons, M. Caruana, T. Lévy, Differential equations driven by rough paths. Ecole d'été de Probabilités de Saint-Flour XXXIV pp. 1–93 (2004)

[21] F.J. Király, H. Oberhauser, Kernels for sequentially ordered data. Journal of Machine Learning Research **20** (2019)

[22] C. Toth, H. Oberhauser, *Bayesian learning from sequential data using gaussian processes with signature covariances*, in *International Conference on Machine Learning* (PMLR, 2020), pp. 9548–9560

[23] M. Lemercier, C. Salvi, T. Cass, E.V. Bonilla, T. Damoulas, T.J. Lyons, *SigGPDE: Scaling Sparse Gaussian Processes on Sequential Data*, in *International Conference on Machine Learning* (PMLR, 2021), pp. 6233–6242

[24] M. Lemercier, C. Salvi, T. Damoulas, E. Bonilla, T. Lyons, *Distribution regression for sequential data*, in *International Conference on Artificial Intelligence and Statistics* (PMLR, 2021), pp. 3754–3762

[25] C. Salvi, M. Lemercier, C. Liu, B. Horvath, T. Damoulas, T. Lyons, Higher order kernel mean embeddings to capture filtrations of stochastic processes. Advances in Neural Information Processing Systems **34**, 16635–16647 (2021)

[26] T. Cochrane, P. Foster, V. Chhabra, M. Lemercier, T. Lyons, C. Salvi, *Sk-tree: a systematic malware detection algorithm on streaming trees via the signature kernel*, in *2021 IEEE international conference on cyber security and resilience (CSR)* (IEEE, 2021), pp. 35–40

[27] J. Dyer, J. Fitzgerald, B. Rieck, S.M. Schmon, *Approximate bayesian computation for panel data with signature maximum mean discrepancies*, in *NeurIPS 2022 Temporal Graph Learning Workshop* (2022)

[28] Z. Issa, B. Horvath, M. Lemercier, C. Salvi, Non-adversarial training of neural sdes with signature kernel scores. Advances in Neural Information Processing Systems **36** (2024)

[29] A. Fermanian, P. Marion, J.P. Vert, G. Biau, Framing rnn as a kernel method: A neural ode approach. Advances in Neural Information Processing Systems **34**, 3121–3134 (2021)

[30] N. Muca Cirone, M. Lemercier, C. Salvi, Neural signature kernels as infinite-width-depth-limits of controlled resnets. arXiv e-prints pp. arXiv–2303 (2023)

[31] A. Pannier, C. Salvi, A path-dependent pde solver based on signature kernels. arXiv preprint arXiv:2403.11738 (2024)

[32] T.L. et al, Coropa computational rough paths (software library) (2010). URL http://coropa.sourceforge.net/

[33] J. Reizenstein, B. Graham, The iisignature library: efficient calculation of iterated-integral signatures and log signatures. arXiv preprint arXiv:1802.08252 (2018)

[34] P. Kidger, T. Lyons, *Signatory: differentiable computations of the signature and logsignature transforms, on both CPU and GPU*, in *International Conference on Learning Representations* (2020)

[35] P.R. Sam Morley, T. Lyons. Roughpy 0.1.0 - pypi (2023). URL https://pypi.org/project/RoughPy/

[36] C. Salvi, T. Cass, J. Foster, T. Lyons, W. Yang, The signature kernel is the solution of a goursat pde. SIAM Journal on Mathematics of Data Science **3**(3), 873–899 (2021)

[37] A. Rahimi, B. Recht, Random features for large-scale kernel machines. Advances in neural information processing systems **20** (2007)

[38] C. Toth, H. Oberhauser, Z. Szabo, Random fourier signature features. arXiv preprint arXiv:2311.12214 (2023)

[39] C. Bellingeri, P.K. Friz, S. Paycha, R. Preiß, Smooth rough paths, their geometry and algebraic renormalization. Vietnam Journal of Mathematics **50**(3), 719–761 (2022)

[40] G. Flint, T. Lyons, Pathwise approximation of sdes by coupling piecewise abelian rough paths. arXiv preprint arXiv:1505.01298 (2015)

[41] H. Boedihardjo, X. Geng, N.P. Souris, Path developments and tail asymptotics of signature for pure rough paths. Advances in Mathematics **364**, 107043 (2020)

[42] T. Lyons, Rough paths, signatures and the modelling of functions on streams. Proceedings of the International Congress of Mathematicians, Korea (2014)

[43] Y. Boutaib, L.G. Gyurkó, T. Lyons, D. Yang, Dimension-free euler estimates of rough differential equations. Revue Roumaine de Mathmatiques Pures et Appliques, 59 (2014)

[44] C. Bayer, S. Breneis, T. Lyons, *An adaptive algorithm for rough differential equations* (Weierstraß-Institut für Angewandte Analysis und Stochastik Leibniz-Institut ..., 2023)

[45] J. Morrill, C. Salvi, P. Kidger, J. Foster, *Neural rough differential equations for long time series*, in *International Conference on Machine Learning* (PMLR, 2021), pp. 7829–7838

[46] B. Walker, A.D. McLeod, T. Qin, Y. Cheng, H. Li, T. Lyons, Log neural controlled differential equations: The lie brackets make a difference. arXiv preprint arXiv:2402.18512 (2024)

[47] K.T. Chen, Integration of paths, geometric invariants and a generalized baker-hausdorff formula. Annals of Mathematics **65**(1), 163–178 (1957)

[48] A.T. et al. signax 0.2.1 - pypi (2024). URL https://pypi.org/project/RoughPy/

Appendix A Proof of Theorem 3

A.1 Equation for the kernel

Let $\mathbb{X}$ and $\mathbb{Y}$ be two smooth rough paths and denote by $\dot{\boldsymbol{x}}$ and $\dot{\boldsymbol{y}}$ their diagonal derivatives in $\mathcal{L}(E)$. By bilinearity of the inner product and the definition of the signature, we have

$$\begin{aligned}u(s,t) - u(\sigma,t) - u(s,\tau) + u(\sigma,\tau) &= \langle S(\mathbb{X})_{0,s} - S(\mathbb{X})_{0,\sigma}, S(\mathbb{Y})_{0,t} - S(\mathbb{Y})_{0,\tau}\rangle \\ &= \langle \int_\sigma^s S(\mathbb{X})_{0,s'} \otimes \dot{\boldsymbol{x}}_{s'} ds', \int_\tau^t S(\mathbb{Y})_{0,t'} \otimes \dot{\boldsymbol{y}}_{t'} dt'\rangle\end{aligned}$$

and

$$u(s,t) = u(\sigma,t) + u(s,\tau) - u(\sigma,\tau) + \langle \int_\sigma^s (S(\mathbb{X})_{0,s'} \otimes \dot{\boldsymbol{x}}_{s'})\, ds', \int_\tau^t (S(\mathbb{Y})_{0,t'} \otimes \dot{\boldsymbol{y}}_{t'})\, dt' \rangle.$$

Expanding the inner product term, we obtain

$$\begin{aligned} u(s,t) = u(\sigma,t) + u(s,\tau) - u(\sigma,\tau) &+ \sum_{k=0}^{\infty} \int_\sigma^s \int_\tau^t u(s',t') \cdot \langle \dot{\boldsymbol{x}}_{s'}^k, \dot{\boldsymbol{y}}_{t'}^k \rangle_k ds' dt' \\ &+ \sum_{k=1}^{\infty} \sum_{q=0}^{k-1} \int_\sigma^s \int_\tau^t \left\langle S(\mathbb{X})_{0,s'} \otimes \dot{\boldsymbol{x}}_{s'}^k, S(\mathbb{Y})_{0,t'} \otimes \dot{\boldsymbol{y}}_{t'}^q \right\rangle_{\mathcal{H}} ds' dt' \\ &+ \sum_{k=0}^{\infty} \sum_{q=k+1}^{\infty} \int_{s_i}^s \int_{t_j}^t \left\langle S(\mathbb{X})_{0,s'} \otimes \dot{\boldsymbol{x}}_{s'}^k, S(\mathbb{Y})_{0,t'} \otimes \dot{\boldsymbol{y}}_{t'}^q \right\rangle_{\mathcal{H}} ds' dt' \end{aligned}$$

This is not an integral equation yet. The first step towards obtaining an integral equation is to rewrite the integrands of the last two integrals. By Theorem 2,

$$\begin{aligned} \left\langle S(\mathbb{X})_{0,s'} \otimes \dot{\boldsymbol{x}}_{s'}^k, S(\mathbb{Y})_{0,t'} \otimes \dot{\boldsymbol{y}}_{t'}^q \right\rangle_{\mathcal{H}} &= \left\langle \ell(S(\mathbb{Y})_{0,t'})(S(\mathbb{X})_{0,s'}), r(\dot{\boldsymbol{x}}_{s'}^k)(\dot{\boldsymbol{y}}_{t'}^q) \right\rangle_{\mathcal{H}}, \quad \text{if } k \leq q \\ \left\langle S(\mathbb{X})_{0,s'} \otimes \dot{\boldsymbol{x}}_{s'}^k, S(\mathbb{Y})_{0,t'} \otimes \dot{\boldsymbol{y}}_{t'}^q \right\rangle_{\mathcal{H}} &= \left\langle \ell(S(\mathbb{X})_{0,s'})(S(\mathbb{Y})_{0,t'}), r(\dot{\boldsymbol{y}}_{t'}^q)(\dot{\boldsymbol{x}}_{s'}^k) \right\rangle_{\mathcal{H}}, \quad \text{if } k \geq q. \end{aligned}$$

And introduce the following $T^{>0}(E)$-valued state variables

$$\begin{aligned} \phi(s,t) &= \ell(S(\mathbb{Y})_{0,t})(S(\mathbb{X})_{0,s}) - u(s,t) \\ \psi(s,t) &= \ell(S(\mathbb{X})_{0,s})(S(\mathbb{Y})_{0,t})) - u(s,t) \end{aligned}$$

With these, we have

$$\begin{aligned} u(s,t) = u(\sigma,t) + u(s,\tau) - u(\sigma,\tau) &+ \int_\sigma^s \int_\tau^t u(s',t') \cdot \langle \dot{\boldsymbol{x}}_{s'}, \dot{\boldsymbol{y}}_{t'} \rangle_{\mathcal{H}} ds' dt' \\ &+ \sum_{k=1}^{\infty} \sum_{q=0}^{k-1} \int_\sigma^s \int_\tau^t \left\langle \psi(s',t'), r(\dot{\boldsymbol{y}}_{t'}^q)(\dot{\boldsymbol{x}}_{s'}^k) \right\rangle ds' dt' \\ &+ \sum_{k=0}^{\infty} \sum_{q=k+1}^{\infty} \int_{s_i}^s \int_{t_j}^t \left\langle \phi(s',t'), r(\dot{\boldsymbol{x}}_{s'}^k)(\dot{\boldsymbol{y}}_{t'}^q) \right\rangle ds' dt' \end{aligned}$$

We have $\langle \psi(s',t), r(\dot{\boldsymbol{y}}_{t'}^q)(\dot{\boldsymbol{x}}_{s'}^k) \rangle = 0$ for $q \geq k$ and $\forall q$ when $k = 0$. This comes from $\langle \psi(s',t), e_\varnothing \rangle = 0$ (case $k = q$), $r(\dot{\boldsymbol{y}}_{t'}^q)(\dot{\boldsymbol{x}}_{s'}^k) = 0$ (case $q > k$), and $\dot{\boldsymbol{x}}^0 \equiv 0$ (case $k = 0$). Therefore,

$$\begin{aligned} \sum_{k=1}^{\infty} \sum_{q=0}^{k-1} \int_\sigma^s \int_\tau^t \left\langle \psi(s',t'), r(\dot{\boldsymbol{y}}_{t'}^q)(\dot{\boldsymbol{x}}_{s'}^k) \right\rangle ds' dt' &= \sum_{k=0}^{\infty} \sum_{q=0}^{\infty} \int_\sigma^s \int_\tau^t \left\langle \psi(s',t), r(\dot{\boldsymbol{y}}_{t'}^q)(\dot{\boldsymbol{x}}_{s'}^k) \right\rangle ds' dt' \\ &= \int_\sigma^s \int_\tau^t \langle \psi(s',t'), r(\dot{\boldsymbol{y}}_{t'})(\dot{\boldsymbol{x}}_{s'}) \rangle ds' dt' \end{aligned}$$

Similarly, using the fact that $r(\dot{\boldsymbol{x}}_{s'}^k)(\dot{\boldsymbol{y}}_{t'}^q) = 0$ for $k > q$ and $\langle \phi(s',t), e_\varnothing \rangle = 0$ and $\dot{\boldsymbol{y}}^0 \equiv 0$,

$$\sum_{k=0}^{\infty} \sum_{q=k+1}^{\infty} \int_{s_i}^s \int_{t_j}^t \left\langle \phi(s',t'), r(\dot{\boldsymbol{x}}_{s'}^k)(\dot{\boldsymbol{y}}_{t'}^q) \right\rangle ds' dt' = \sum_{k=0}^{\infty} \sum_{q=0}^{\infty} \int_{s_i}^s \int_{t_j}^t \left\langle \phi(s',t'), r(\dot{\boldsymbol{x}}_{s'}^k)(\dot{\boldsymbol{y}}_{t'}^q) \right\rangle ds' dt'$$

$$= \int_{s_i}^{s} \int_{t_j}^{t} \langle \phi(s',t'), r(\dot{\boldsymbol{x}}_{s'})(\dot{\boldsymbol{y}}_{t'}) \rangle \, ds' dt'$$

With this, we get the first equation (in integral form)

$$\begin{aligned} u(s,t) = u(\sigma,t) + u(s,\tau) - u(\sigma,\tau) + \int_{\sigma}^{s} \int_{\tau}^{t} u(s',t') \cdot \langle \dot{\boldsymbol{x}}_{s'}, \dot{\boldsymbol{y}}_{t'} \rangle_{\mathcal{H}} ds' dt' \\ + \int_{\sigma}^{s} \int_{\tau}^{t} \langle \psi(s',t), r(\dot{\boldsymbol{y}}_{t'})(\dot{\boldsymbol{x}}_{s'}) \rangle \, ds' dt' \\ + \int_{\sigma}^{s} \int_{\tau}^{t} \langle \phi(s',t), r(\dot{\boldsymbol{x}}_{s'})(\dot{\boldsymbol{y}}_{t'}) \rangle \, ds' dt' \end{aligned}$$

A.2 Equation for the adjoints

Now, we derive the equation for ϕ_w for any non empty word w.

$$\begin{aligned} \langle \phi(s,t) - \phi(\sigma,t), e_w \rangle &= \langle S(\mathbb{X})_{0,s} - S(\mathbb{X})_{0,\sigma}, S(\mathbb{Y})_{0,t} \otimes e_w \rangle \\ &= \Big\langle \int_{\sigma}^{s} S(\mathbb{X})_{0,s'} \otimes \dot{\boldsymbol{x}}_{s'} ds', S(\mathbb{Y})_{0,t} \otimes e_w \Big\rangle \\ &= \int_{\sigma}^{s} \langle S(\mathbb{X})_{0,s'} \otimes \dot{\boldsymbol{x}}_{s'}, S(\mathbb{Y})_{0,t} \otimes e_w \rangle ds' \end{aligned}$$

Writing $k = |w|$ and rewriting the integrand

$$\begin{aligned} \langle S(\mathbb{X})_{0,s'} \otimes \dot{\boldsymbol{x}}_{s'}, S(\mathbb{Y})_{0,t} \otimes e_w \rangle = u(s',t) \langle \dot{\boldsymbol{x}}_{s'}^{k}, e_w \rangle \\ + \sum_{q=0}^{k-1} \langle \ell(S(\mathbb{Y})_{0,t})(S(\mathbb{X})_{0,s'}), r(\dot{\boldsymbol{x}}_{s'}^{q})(e_w) \rangle \\ + \sum_{q=k+1}^{\infty} \langle \ell(S(\mathbb{X})_{0,s'})(S(\mathbb{Y})_{0,t}), r(e_w)(\dot{\boldsymbol{x}}_{s'}^{q}) \rangle \end{aligned}$$

Putting everything together, we get

$$\begin{aligned} \langle \phi(s,t), e_w \rangle = \langle \phi(\sigma,t), e_w \rangle + \int_{\sigma}^{s} u(s',t) \langle \dot{\boldsymbol{x}}_{s'}^{k}, e_w \rangle ds' + \sum_{q=0}^{k-1} \int_{\sigma}^{s} \langle \phi(s',t), r(\dot{\boldsymbol{x}}_{s'}^{q})(e_w) \rangle ds' \\ + \sum_{q=k+1}^{\infty} \int_{\sigma}^{s} \langle \psi(s',t), r(e_w)(\dot{\boldsymbol{x}}_{s'}^{q}) \rangle ds' \end{aligned}$$

And using $\langle \phi, e_{\varnothing} \rangle = \langle \psi, e_{\varnothing} \rangle = 0$ and $r(\dot{\boldsymbol{x}}_{s'}^{q})(e_w) = 0$ for $q > k$ and $r(e_w)(\dot{\boldsymbol{x}}_{s'}^{q}) = 0$ for $q < k$

$$\begin{aligned} \langle \phi(s,t), e_w \rangle = \langle \phi(\sigma,t), e_w \rangle + \int_{\sigma}^{s} u(s',t) \langle \dot{\boldsymbol{x}}_{s'}, e_w \rangle ds' + \int_{\sigma}^{s} \langle \phi(s',t), r(\dot{\boldsymbol{x}}_{s'})(e_w) \rangle ds' \\ + \int_{\sigma}^{s} \langle \psi(s',t), r(e_w)(\dot{\boldsymbol{x}}_{s'}) \rangle ds' \end{aligned}$$

Now, our goal is to find an equation in $T(E)$ for the tensor ϕ

$$\phi(s,t) = 0 \cdot e_{\varnothing} + \sum_{k=1}^{\infty} \sum_{w:|w|=k} \langle \phi(s,t), e_w \rangle e_w$$

Plugging in the above,

$$\phi(s,t) = 0 \cdot e_\varnothing + \sum_{k=1}^{\infty} \sum_{w:|w|=k} \langle \phi(\sigma,t), e_w \rangle e_w + \sum_{k=1}^{\infty} \sum_{w:|w|=k} \int_\sigma^s \langle \phi(s',t), r(\dot{\boldsymbol{x}}_{s'})(e_w) \rangle ds' e_w$$
$$+ \sum_{k=1}^{\infty} \sum_{w:|w|=k} \int_\sigma^s \langle \psi(s',t), r(e_w)(\dot{\boldsymbol{x}}_{s'}) \rangle ds' e_w$$

The penultimate term can be rewritten as

$$\begin{aligned}
\sum_{k=1}^{\infty} \sum_{w:|w|=k} \int_\sigma^s \langle \phi(s',t), r(\dot{\boldsymbol{x}}_{s'})(e_w) \rangle ds' e_w &= \sum_{k=0}^{\infty} \sum_{w:|w|=k} \int_\sigma^s \langle \phi(s',t), r(\dot{\boldsymbol{x}}_{s'})(e_w) \rangle ds' e_w \\
&= \sum_{k=0}^{\infty} \sum_{w:|w|=k} \int_\sigma^s \langle \phi(s',t) \otimes \dot{\boldsymbol{x}}_{s'}, e_w \rangle ds' e_w \\
&= \int_\sigma^s \phi(s',t) \otimes \dot{\boldsymbol{x}}_{s'} ds'
\end{aligned}$$

where the first equality comes from the fact that $\langle \phi(s',t), e_\varnothing \rangle = 0$ and the second equality from the definition of the right adjoint. For the last term, using $\langle \psi(s',t), r(e_\varnothing)(\dot{\boldsymbol{x}}_s) \rangle = \langle \psi(s',t), \dot{\boldsymbol{x}}_s \rangle$, and adjoint operations, we obtain

$$\begin{aligned}
\sum_{k=1}^{\infty} \sum_{w:|w|=k} \int_\sigma^s \langle \psi(s',t), r(e_w)(\dot{\boldsymbol{x}}_{s'}) \rangle ds' e_w &= \sum_{k=0}^{\infty} \sum_{w:|w|=k} \int_\sigma^s \langle \psi(s',t), r(e_w)(\dot{\boldsymbol{x}}_{s'}) \rangle ds' e_w \\
&\quad - \int_\sigma^s \langle \psi(s',t), \dot{\boldsymbol{x}}_{s'} \rangle ds' e_\varnothing \\
&= \sum_{k=0}^{\infty} \sum_{w:|w|=k} \int_\sigma^s \langle \psi(s',t) \otimes e_w, \dot{\boldsymbol{x}}_{s'} \rangle ds' e_w \\
&\quad - \int_\sigma^s \langle \psi(s',t), \dot{\boldsymbol{x}}_{s'} \rangle ds' e_\varnothing \\
&= \sum_{k=0}^{\infty} \sum_{w:|w|=k} \int_\sigma^s \langle e_w, \ell(\psi(s',t))(\dot{\boldsymbol{x}}_{s'}) \rangle ds' e_w \\
&\quad - \int_\sigma^s \langle \psi(s',t), \dot{\boldsymbol{x}}_{s'} \rangle ds' e_\varnothing \\
&= \int_\sigma^s \ell(\psi(s',t))(\dot{\boldsymbol{x}}_{s'}) ds' - \int_\sigma^s \langle \psi(s',t), \dot{\boldsymbol{x}}_{s'} \rangle ds' e_\varnothing
\end{aligned}$$

Therefore,

$$\phi(s,t) = \phi(\sigma,t) + \int_\sigma^s \phi(s',t) \otimes \dot{\boldsymbol{x}}_{s'} ds' + \int_\sigma^s \ell(\psi(s',t))(\dot{\boldsymbol{x}}_{s'}) ds' - \int_\sigma^s \langle \psi(s',t), \dot{\boldsymbol{x}}_{s'} \rangle ds' e_\varnothing$$

Appendix B Numerical Analysis

Lemma 2. *Suppose $\mathbb{X}$ and $\mathbb{Y}$ are p-rough paths then, for any $\beta > 0$ there exists a common control w so that for all $s, t \in [0, T]$ and for any degree $k \geq 1$ we have*

$$\|x^k_{s,t}\| \leq \frac{w(s,t)^{k/p}}{\beta(k/p)!} \qquad \text{and} \qquad \|y^k_{s,t}\| \leq \frac{w(s,t)^{k/p}}{\beta(k/p)!}.$$

we note that by reparametrizing $\mathbb{X}$ and $\mathbb{Y}$ by a common reparametrization $\tau(t) = w(0,t)$ of $\mathbb{X}$ and $\mathbb{Y}$, we can take $w(s,t) = t - s$.

Lemma 3. *Suppose that $\mathbb{X}$ is a weakly geometric p-rough path defined on $[0, T]$ with values in E and that $m = \lfloor p \rfloor$. Suppose that $\mathbb{X}$ is controlled by ω/C. That is to say, for all s, t and all $k \leq m$ we have*

$$\left\|x^k_{s,t}\right\| \leq \frac{\omega(s,t)^{k/p}}{C^{k/p}\beta_p\left(k/p\right)!}.$$

Let $\mathbf{X}_r$ be the tensor $\exp_m\left(r \log \mathbb{X}_{s,t}\right)$ where $0 \leq r \leq 1$ then for all $k \geq 1$

$$\left\|x^k_r\right\| \leq \frac{\left(r\omega(s,t)\right)^{k/p}}{\beta_p\left(k/p\right)!}.$$

Proof. Working in the tensor algebra truncated at level m,

$$\begin{aligned}\mathbf{X}_r &:= \left(1 + \sum_{i=1\ldots m} x^k_{s,t}\right)^r \\ &= 1 + r\left(\sum_{i=1\ldots m} x^k_{s,t}\right) + \frac{r(r-1)}{2!}\left(\sum_{i=1\ldots m} x^k_{s,t}\right)^2 + \ldots + \frac{r(r-1)\ldots(r-(m-1))}{m!}\left(\sum_{i=1\ldots m} x^k_{s,t}\right)^m\end{aligned}$$

For any $k \geq 1$

$$x^k_r = r x^k_{s,t} + \frac{r(r-1)}{2!}\sum_{j_1+j_2=k} x^{j_1}_{s,t}x^{j_2}_{s,t} + \ldots + \frac{r(r-1)\ldots(r-(m-1))}{m!}\sum_{j_1+\ldots+j_2=k} x^{j_1}_{s,t}\ldots x^{j_m}_{s,t}$$

where all $j_i \geq 1$. Estimating this with the neoclassical inequality we get

$$\left\|x^k_r\right\| \leq rC_{k,p}\frac{\omega(s,t)^{k/p}}{C^{k/p}}.$$

Recall that $r < 1$ and $k/p \leq 1$ so that

$$\begin{aligned}\left\|x^k_r\right\| &\leq rC_{k,p}\frac{\omega(s,t)^{k/p}}{C^{k/p}} \\ &\leq rC_{k,p}\frac{(r\omega(s,t))^{k/p}}{C^{k/p}}\end{aligned}$$

Now set $C = \max\left\{\left(\frac{C_{k,p}}{\beta_p(k/p)!}\right)^{p/k} \mid k \leq m\right\}$ then we have

$$\left\|x^k_r\right\| \leq \frac{\left(r\omega\left(s,t\right)\right)^{k/p}}{\beta_p\left(k/p\right)!}.$$

□

Lemma 4. *Assume $p \in [1, +\infty)$ and $a > 0$, then*

$$\sum_{k=m}^{\infty} \frac{a^{k/p}}{(k/p)!} \leq \exp(a;p) \frac{a^{m/p}}{(m/p)!}$$

where $\exp(a,p)$ *is the Mittag-Leffler function* $\exp(a;p) := \sum_{k \geq 0} \frac{a^{k/p}}{(k/p)!}$.

Proof. Using $\Gamma(x+1)\Gamma(y+1) < \Gamma(x+y+1)$ for any $x, y > 0$, we have

$$\begin{aligned}
\sum_{k=m}^{\infty} \frac{a^{k/p}}{(k/p)!} &= \sum_{k=0}^{\infty} \frac{a^{(k+m)/p}}{((k+m)/p)!} \\
&= a^{m/p} \sum_{k=0}^{\infty} \frac{a^{k/p}}{((k+m)/p)!} \\
&= \frac{a^{m/p}}{(m/p)!} \sum_{k=0}^{\infty} \frac{(k/p)!(m/p)!}{((k+m)/p)!} \frac{a^{k/p}}{(k/p)!} \\
&\leq \frac{a^{m/p}}{(m/p)!} \sum_{k=0}^{\infty} \frac{a^{k/p}}{(k/p)!}
\end{aligned}$$

□

B.1 Proof of Lemma 1

Let $\mathbb{X} = (1, x^1, \ldots, x^{\lfloor p \rfloor})$ be a weakly geometric p-rough path defined on $[0,T]$ with values in E and let $m = \lfloor p \rfloor$. Let $\mathbb{X}^{m,D}$ be the piecewise log-linear approximation of $\mathbb{X}$ of degree m on a partition $D = \{0 = t_0 < \ldots < t_n = T\}$ of $[0,T]$. For any $i = 0, \ldots, n-1$, by Lemma 3 and Lemma 2, we have

$$\begin{aligned}
\|S(\mathbb{X})_{t_i,t_{i+1}} - S(\mathbb{X}^{m,D})_{t_i,t_{i+1}}\| &\leq \sum_{k=0}^{\infty} \|x^k_{t_i,t_{i+1}} - (\mathbb{X}^{m,D}_{t_i,t_{i+1}})^k\| \\
&\leq \sum_{k=m+1}^{\infty} \|x^k_{t_i,t_{i+1}} - (\mathbb{X}^{m,D}_{t_i,t_{i+1}})^k\| \\
&\leq 2 \sum_{k=m+1}^{\infty} \frac{w(t_i,t_{i+1})^{k/p}}{\beta_p (k/p)!}
\end{aligned}$$

where we have used the fact that $x^k_{t_i,t_{i+1}} = (\mathbb{X}^{m,D}_{t_i,t_{i+1}})^k$ for all $k \leq m$. Then, we have

$$S(\mathbb{X})_{0,T} - S(\mathbb{X}^{m,D})_{0,T} = \sum_{i=0}^{n-1} S(\mathbb{X}^{m,D})_{0,t_i} \otimes \big(S(\mathbb{X})_{t_i,t_{i+1}} - S(\mathbb{X}^{m,D})_{t_i,t_{i+1}}\big) \otimes S(\mathbb{X})_{t_{i+1},t_n}$$

and taking the norm and applying Lemma 4, we get

$$\begin{aligned}
\|S(\mathbb{X})_{0,T} - S(\mathbb{X}^{m,D})_{0,T}\| &\leq \sum_{i=0}^{n-1} \|S(\mathbb{X}^{m,D})_{0,t_i}\| \|S(\mathbb{X})_{t_i,t_{i+1}} - S(\mathbb{X}^{m,D})_{t_i,t_{i+1}}\| \|S(\mathbb{X})_{t_{i+1},t_n}\| \\
&\leq \left(\sum_{k=0}^{\infty} \frac{w(0,T)^{k/p}}{\beta_p (k/p)!}\right) \left(2n \sum_{k=m+1}^{\infty} \frac{(w(0,T)/n)^{k/p}}{\beta_p (k/p)!}\right) \left(\sum_{k=0}^{\infty} \frac{w(0,T)^{k/p}}{\beta_p (k/p)!}\right) \\
&\leq \frac{2n}{\beta_p^3} \frac{(w(0,T)/n)^{(m+1)/p}}{((m+1)/p)!} \exp\left(\frac{w(0,T)}{n}; p\right) (\exp(w(0,T);p))^2 \\
&\leq \frac{2}{\beta_p^3} \frac{w(0,T)}{((m+1)/p)!} \exp(w(0,T);p)^2 \exp\left(\frac{w(0,T)}{n}; p\right) \left(\frac{w(0,T)}{n}\right)^{\frac{m+1}{p}-1}
\end{aligned}$$