Université de Rennes 2025-2026

SÉMINAIRE DE MASTER 2

On the expected signature of random paths to characterize their laws.

Arthur Maritch-Roy, supervisé par Clément Levrard

Outline

The signature of a path has been introduced by K. Chen in 1954 [5] while studying the cohomology of the loop space. He showed that it is an injection from an appropriate set of paths onto the space of formal series. Later on, Hambly and Lyons found it could be a useful tool for rough paths theory. They found the kernel of the signature on paths of bounded variation [10], that is tree-like paths, and their result have been extended to weakly geometry rough paths by Boedihardjo et al. [4].
In the 2010s, the signature has begun to be used for machine learning [12] and Chevyrev and Oberhauser [7] used it to build a practical tool to characterize the law of a random path up to tree-like equivalence.

In this document, we define the signature of a path of bounded variation and show some of its useful algebraic properties. Then we show why it can be useful to characterize the law of a random path in a particular space of paths, as is has been done in [6]. We also give some elements of the proof of Hambly and Lyons [10], to show in the last section how to build a normalized signature like in [7], using the formalism of reproducing kernel Hilbert spaces.

Contents

1 Signature of a path

1.1 Paths of bounded variation

In all this paragraph, $V = \mathbb{R}^d$ will denote a finite-dimensional real vector space, equipped with the topology induced by the Euclidean norm.

Definition 1. A *path* (on V) is a continuous function $\gamma : [0,1] \to V$ which maps a non-empty interval of $\mathbb{R}$ into V. We define its *length* (or *total variation*) by

$$|\gamma| = |\gamma|_{[0,1]} := \sup_{\substack{\sigma=\{t_i\}_i\subset[0,1] \\ \sigma \text{ subdivision of } [0,1]}} \sum_i \|\gamma(t_{i+1}) - \gamma(t_i)\|.$$

This is a general definition that can be extended to any path that takes values in a metric space. We say that γ is of *bounded variation* if $|\gamma| < +\infty$, and we write $\gamma \in \mathrm{BV}(V)$.

For a question of simplicity, all paths will be defined on $[0,1]$, but it can be done on any interval $[a,b] \subset \mathbb{R}$.

Theorem 2. (Lebesgue, [16]) If $\gamma \in \mathrm{BV}(V)$ then γ is differentiable almost everywhere (for Lebesgue measure).

This theorem allows us to define $\mathrm{d}\gamma(s) = \gamma'(s)\,\mathrm{d}s$ in order to give sense to $\int_0^1 f(s)\,\mathrm{d}\gamma(s)$ if $f : [0,1] \to V$ is a measurable function. For a question of notation, we will prefer the notations γ_s and $\mathrm{d}\gamma_s$ in the following In the general case where V is a metric space, we need it to be a Banach space to define these integrals, through Riemann-Stieltjes integration.

Corollary 3. Let $\gamma \in \mathrm{BV}(V)$. Then $|\gamma| = \int_0^1 \|\gamma'(t)\|\,\mathrm{d}t$. In particular, $|\cdot|$ defines a semi-norm on $\mathrm{BV}(V)$, and it defines a norm if we add the first value: $\gamma \mapsto |\gamma_0| + \int_0^1 |\gamma'(t)|\,\mathrm{d}t$ defines a norm on BV, which we will note $\|\cdot\|_{\mathrm{BV}}$.

Definition 4. In order to have a metric on $\mathrm{BV}(V)$ that does not depend on time parametrization, we define

$$\rho(x,y) = \inf_{\tau \text{ parametrization}} |x - y \circ \tau|,$$

which is a pseudo-metric on $\mathrm{BV}(V)$. It then defines a non Hausdorff topology on $\mathrm{BV}(V)$ (if x and y are two constant paths, then $\rho(x,y)$ is always zero).

From now, we will always take $V = \mathbb{R}^d$.

1.2 Signature of a path of bounded variation

Definition 5. For $i \in \{1,\dots,d\}$, and $\gamma = (\gamma^1,\dots,\gamma^d) \in \mathrm{BV}(\mathbb{R}^d)$, we define (if γ is defined on $[a,b]$):

$$\mathrm{S}(\gamma)^i_{a,b} := \int_{a<s<b} \mathrm{d}\gamma^i_s = \gamma^i_b - \gamma^i_a.$$

Recursively, for $I = (i_1,\dots,i_k) \in \{1,\dots,d\}^k$, we define

$$\mathrm{S}(\gamma)^{i_1,\dots,i_k}_{a,b} = \int_{a<s<b} \mathrm{S}(\gamma)^{i_1,\dots,i_{k-1}}_{a,s}\,\mathrm{d}\gamma_s. \tag{1}$$

One can check (by induction on k) this corresponds to

$$\mathrm{S}(\gamma)^I = \int_{a<t_1<\dots<t_k<b} \mathrm{d}\gamma^{i_1}_{t_1}\cdots\mathrm{d}\gamma^{i_k}_{t_k}. \tag{2}$$

The *signature* of the path γ on $[a,b]$ is the collection

$$\mathrm{S}(\gamma)_{a,b} = (\mathrm{S}(\gamma)^I_{a,b})_{I \text{ multi index}}, \tag{3}$$

where we take $\mathrm{S}(\gamma)^{\emptyset}_{a,b} = 1$ as a convention.

Because we will consider paths on $[0,1]$, $\mathrm{S}(\gamma)$ (resp. $\mathrm{S}(\gamma)^I$) will be a shorthand for $\mathrm{S}(\gamma)_{0,1}$ (resp. $\mathrm{S}(\gamma)^I_{0,1}$). Indexing on all multi-indices like in (3) is not well-defined because we do not have any order on this multi-indices. Let us then introduce an algebra in which the signature will live.

Definition 6. (tensor series algebra)
We define $\mathcal{T}(\mathbb{R}^d)$ the *tensor series algebra* on $\mathbb{R}^d$ the algebra of elements of the form

$$(\lambda_0, \lambda_1, \dots, \lambda_d, \lambda_{1,1}, \dots, \lambda_{1,d}, \dots, \lambda_{d,d}, \lambda_{1,1,1}, \lambda_{1,1,2}, \dots) =: \sum_{k=0}^{\infty} \sum_{1 \le i_1, \dots, i_k \le d} \lambda_{i_1,\dots,i_k} e_{i_1} \otimes \cdots \otimes e_{i_k},$$

where the sum and the scalar multiplication are define coordinate by coordinate, and the product (also denoted $\otimes$) is the one that extends:

$$\mathbb{R}^{\otimes k} \otimes \mathbb{R}^{\otimes l} \ni (e_{i_1} \otimes \cdots \otimes e_{i_k}) \otimes (e_{j_1} \otimes \cdots \otimes e_{j_l}) = e_{i_1} \otimes \cdots \otimes e_{i_k} \otimes e_{j_1} \otimes \cdots \otimes e_{j_l} \in \mathbb{R}^{\otimes k+l}.$$

If we denote by $\wedge$ the concatenation between two multi-indexes, we can write $e_{I \wedge J} = e_I \otimes e_J$ if $I = (i_1, \dots, i_k)$ and $J = (j_1, \dots, j_m)$.

Then we can see the signature $\mathrm{S}(\gamma)$, $\gamma \in \mathrm{BV}(\mathbb{R}^d)$, as an element of $\mathcal{T}(\mathbb{R}^d)$

$$\mathrm{S}(\gamma) = 1 + \sum_{k=1}^{+\infty} \sum_{I=(i_1,\dots,i_k)} \mathrm{S}(\gamma)^I \, e_{i_1} \otimes \cdots \otimes e_{i_k}. \tag{4}$$

Another notation that can be used is seeing $\mathrm{S}(\gamma) = \sum_{k=0}^{\infty} \mathrm{S}_k(\gamma)$, where $\mathrm{S}_k(\gamma) = \sum_{I=(i_1,\dots,i_k)} \mathrm{S}(\gamma)^I \, e_I$, as an element of $\prod_{k \ge 0} (\mathbb{R}^d)^{\otimes k}$:

$$\mathrm{S}_k(\gamma) = \int_{0<t_1<\cdots<t_k<1} \mathrm{d}\gamma_{t_1} \otimes \cdots \otimes \mathrm{d}\gamma_{t_k}.$$

Here are some properties of the signature that are easy to check.

Proposition 7. Let $\gamma \in \mathrm{BV}(\mathbb{R}^d)$.

- For $a \in \mathbb{R}^d$, $\mathrm{S}(\gamma + a) = \mathrm{S}(\gamma)$
- Let $\varphi : [0,1] \to [0,1]$ a strictly increasing function, $\mathrm{S}(\gamma \circ \varphi) = \mathrm{S}(\gamma)$.
- Let $\lambda \in \mathbb{R}$, $I = (i_1, \dots, i_k)$, $\mathrm{S}(\lambda\gamma)^I = \lambda^k \, \mathrm{S}(\gamma)^I$.
- For all I, $\gamma \mapsto \mathrm{S}(\gamma)^I$ is continuous for the metric $|\cdot|$.

Example 8. Let $\gamma \in \mathrm{BV}(\mathbb{R})$. Then

$$\mathrm{S}(\gamma)^{1,\dots,1} = \frac{(\gamma_1 - \gamma_0)^k}{k!}.$$

In particular, it only depends on the extreme values of γ.

Example 9. Let $x = (x_1, \dots, x_g) \in \mathbb{R}^d$ and $\gamma : t \mapsto tx$. Then we have

$$S(\gamma)_{0,1}^{I=i_1,\dots,i_k} = \frac{x_1^{\beta_1} \cdots x_d^{\beta_d}}{k!},$$

where β_j is the number of j in I. Therefore:

$$\begin{aligned}
\mathrm{S}(\gamma) &= 1 + \sum_{k=1}^{\infty} \frac{1}{k!} \sum_{I=(i_1,\dots,i_k)} x_{i_1} \cdots x_{i_d} e_{i_1} \otimes \cdots \otimes e_{i_k} \\
&= 1 + \sum_{k=1}^{\infty} \frac{(\sum_{i=1}^{d} x_i e_i)^{\otimes k}}{k!} \\
&= \sum_{k=0}^{\infty} \frac{x^{\otimes k}}{k!}.
\end{aligned}$$

Remark 10. Here we defined the signature for paths of bounded variation, but all we did and all we will do in the sequel can be generalized to rough paths (by using rough paths integrals), see [4]. For example, one can define the signature of a brownian path with Stratonovitch integrals, see [8].

1.3 Algebraic properties of the signature

This paragraph is dedicated to showing the shuffle product identity and Chen's identity which will be useful later on.

Definition 11 (Shuffle product, [6, Def 1.12])**.** Let $I = (i_1, \dots, i_k)$ and $J = (j_1, \dots, j_m)$ be two multi-indexes. The *shuffle product* $I \sqcup\!\sqcup J$ is the *multi-set* defined by

$$I \sqcup\!\sqcup J = \{(i_1, \dots, i_k, j_1, \dots, j_m), (i_1, \dots, i_{k-1}, j_1, i_k, \dots, j_m), \dots, (j_1, \dots, j_m, i_1, \dots, i_k)\}.$$

In order to make it clear, let us give an example (for a question of readability, I removed parentheses and commas):

$$(1, 2, 3) \sqcup\!\sqcup (4, 5) = \{12345, 12435, 14235, 41235, 41253, 41523, 45123, 12453, 14523, 14253\}.$$

The elements of $I \sqcup\!\sqcup J$ are the multi indices of length $k + m$ made exactly of the elements of $I \cup J$, but where the order in I and J is preserved.
One can check that the shuffle product is commutative, and that $|I \sqcup\!\sqcup J| = \binom{k+m}{k}$.

Theorem 12 (Shuffle product identity, [6, Thm 1.14])**.** Let $\gamma \in \mathrm{BV}(\mathbb{R}^d)$. For all I, J:

$$\mathrm{S}(\gamma)^I\, \mathrm{S}(\gamma)^J = \sum_{K \in I \sqcup\!\sqcup J} \mathrm{S}(\gamma)^K \tag{5}$$

Proof. Let k, m the number of elements in I, J. We proceed by induction on $k + m$. The result is true when k or $m = 0$. For $k + m \geq 2$

$$\begin{aligned}
\mathrm{S}(\gamma)^I\, \mathrm{S}(\gamma)^J &= \int_{s,t \in [0,1]} \mathrm{S}(\gamma)_{0,s}^{i_1,\dots,i_{k-1}}\, \mathrm{S}(\gamma)_{0,t}^{j_1,\dots,j_{m-1}}\, \mathrm{d}\gamma_s^{i_k} \mathrm{d}\gamma_t^{j_m} \\
&= \int_0^1 \mathrm{S}(\gamma)_{0,t}^{I}\, \mathrm{S}(\gamma)_{0,t}^{j_1,\dots,j_{m-1}}\, \mathrm{d}\gamma_t^{j_m} + \int_0^1 \mathrm{S}(\gamma)_{0,s}^{J}\, \mathrm{S}(\gamma)_{0,s}^{i_1,\dots,i_{k-1}}\, \mathrm{d}\gamma_s^{i_k}.
\end{aligned}$$

By induction hypothesis, we thus have

$$\begin{aligned}
\mathrm{S}(\gamma)^I\, \mathrm{S}(\gamma)^J &= \int_0^1 \sum_{K \in I \sqcup\!\sqcup (J \setminus \{j_m\})} \mathrm{S}(\gamma)_{0,t}^{K}\, \mathrm{d}\gamma_s^{j_m} + \int_0^1 \sum_{K \in (I \setminus \{i_k\} \sqcup\!\sqcup J)} \mathrm{S}(\gamma)_{0,s}^{K}\, \mathrm{d}\gamma_s^{i_k} \\
&= \sum_{\substack{K' = K \wedge (j_m) \\ K \in I \sqcup\!\sqcup (J \setminus \{j_m\})}} \mathrm{S}(\gamma)_{0,1}^{K'} + \sum_{\substack{K' = K \wedge (i_k) \\ K \in (I \setminus \{i_k\} \sqcup\!\sqcup J)}} \mathrm{S}(\gamma)_{0,1}^{K'} \\
&= \sum_{K \in I \sqcup\!\sqcup J} \mathrm{S}(\gamma)_{0,1}^{K}.
\end{aligned}$$

Where $\wedge$ denotes the concatenation between two multi-indexes. Note that the induction hypothesis must cover the paths on all the intervals. □

Definition 13. Let $\gamma_1, \gamma_2 \in \mathrm{BV}(\mathbb{R}^d)$, we define the *concatenation* of γ_1 and γ_2, denoted by $\gamma_1 * \gamma_2$ by

$$(\gamma_1 * \gamma_2)(t) = \gamma_1(2t) \text{ if } t < 1/2, \gamma_2(2t-1) - (\gamma_1(1) - \gamma_2(0)) \text{ if } t \geq 1/2. \tag{6}$$

Theorem 14 (Chen's identity, [6, Thm 1.17])**.** Let $\gamma : [a,c] \to \mathbb{R}^d \in \mathrm{BV}(\mathbb{R}^d)$, and $b \in]a,c[$. Then for $i_1, \dots, i_k \in \{1, \dots, d\}$:

$$\mathrm{S}(\gamma)^{i_1,\dots,i_k}_{a,c} = \sum_{m=0}^{k} \mathrm{S}(\gamma)^{i_1,\dots,i_m}_{a,b}\, \mathrm{S}(\gamma)^{i_{m+1},\dots,i_k}_{b,c}. \tag{7}$$

Equivalently, if $\gamma_1, \gamma_2 \in \mathrm{BV}(\mathbb{R}^d)$, we can write in the free tensor algebra:

$$\mathrm{S}(\gamma_1 * \gamma_2) = \mathrm{S}(\gamma_1) \otimes \mathrm{S}(\gamma_2). \tag{8}$$

Proof. We prove the result by induction on k. The result holds for $k = 0$, let us assume it holds for a $k \geq 0$.

$$\begin{aligned}
\mathrm{S}(\gamma)_{a,c} &= \int_a^b \mathrm{S}(\gamma)^{i_1,\dots,i_{k-1}}_{a,t}\, \mathrm{d}\gamma^{i_k}_t + \int_b^c \mathrm{S}(\gamma)^{i_1,\dots,i_{k-1}}_{a,t}\, \mathrm{d}\gamma^{i_k}_t \\
&= \mathrm{S}(\gamma)^{i_1,\dots,i_k}_{a,b} + \int_b^c \sum_{m=0}^{k-1} \mathrm{S}(\gamma)^{i_1,\dots,i_m}_{a,b}\, \mathrm{S}(\gamma)^{i_{m+1},\dots,i_{k-1}}_{b,t}\, \mathrm{d}\gamma^{i_k}_t \\
&= \mathrm{S}(\gamma)^{i_1,\dots,i_k}_{a,b} + \sum_{m=0}^{k-1} \mathrm{S}(\gamma)^{i_1,\dots,i_k}_{a,b} \int_b^c \mathrm{S}(\gamma)^{i_{m+1},\dots,i_{k-1}}_{b,t}\, \mathrm{d}\gamma^{i_k}_t \\
&= \sum_{m=0}^{k} \mathrm{S}(\gamma)^{i_1,\dots,i_m}_{a,b}\, \mathrm{S}(\gamma)^{i_{m+1},\dots,i_k}_{b,c}.
\end{aligned}$$

Rewriting this for concatenation, we get, for all $i_1, \dots, i_k$:

$$\mathrm{S}(\gamma_1 * \gamma_2)^{i_1,\dots,i_k} = \sum_{m=0}^{k} \mathrm{S}(\gamma_1)^{i_1,\dots,i_m}\, \mathrm{S}(\gamma_2)^{i_{m+1},\dots,i_k},$$

so that in the tensor algebra we get

$$\begin{aligned}
\mathrm{S}(\gamma_1 * \gamma_2) &= 1 + \sum_{k=1}^{\infty} \sum_{1 \leq i_1,\dots,i_k \leq d} \left(\sum_{m=0}^{k} \mathrm{S}(\gamma_1)^{i_1,\dots,i_m}\, \mathrm{S}(\gamma_2)^{i_{m+1},\dots,i_k} \right) e_{i_1} \otimes \cdots \otimes e_{i_k} \\
&= 1 + \sum_{k=1}^{\infty} \sum_{1 \leq i_1,\dots,i_k \leq d} \left(\sum_{m=0}^{k} (\mathrm{S}(\gamma_1)^{i_1,\dots,i_m}\, e_{i_1} \otimes \cdots \otimes e_{i_m}) \otimes (\mathrm{S}(\gamma_2)^{i_{m+1},\dots,i_k}\, e_{i_{m+1}} \otimes \cdots \otimes e_{i_k}) \right) \\
&= 1 + \sum_{I \neq \emptyset} \sum_{J \wedge K = I} (\mathrm{S}(\gamma_1)^J\, e_J) \otimes (\mathrm{S}(\gamma_2)^K\, e_K) \\
&= \sum_{I} \sum_{J \wedge K = I} (\mathrm{S}(\gamma_1)^J\, e_J) \otimes (\mathrm{S}(\gamma_2)^K\, e_K) \\
&= \mathrm{S}(\gamma_1) \otimes \mathrm{S}(\gamma_2).
\end{aligned}$$

□

Definition 15. Let $\gamma : [a,b] \to \mathbb{R}^d \in \mathrm{BV}(\mathbb{R}^d)$, we define $\overleftarrow{\gamma} \in \mathrm{BV}(\mathbb{R}^d)$ by $\overleftarrow{\gamma}_t := \gamma_{a+b-t}$.

Proposition 16 ([6, Thm 1.25])**.** For $1 \leq i_1, \dots, i_k \leq d$, we have $\mathrm{S}(\overleftarrow{\gamma})^{i_1,\dots,i_k} = (-1)^k\, \mathrm{S}(\gamma)^{i_k,\dots,i_1}$. Also we have $\mathrm{S}(\gamma * \overleftarrow{\gamma}) = 1$ so by Chen's identity:

$$\mathrm{S}(\gamma) \otimes \mathrm{S}(\overleftarrow{\gamma}) = 1. \tag{9}$$

Proof. These two properties can be shown using the same method as before, *i.e.* by induction. □

The set $\mathrm{BV}([0,1], \mathbb{R}^d)$ forms a monoid with the binary operation $*$, and S is a homomorphism from $(\mathrm{BV}(\mathbb{R}^d), *)$ into $(\mathcal{T}(\mathbb{R}^d), \otimes)$. If we are able to determine the kernel of S, then we would have an isomorphism

$$\mathrm{BV}(\mathbb{R}^d) / \ker(\mathrm{S}) \xrightarrow{\sim} \mathcal{T}(\mathbb{R}^d).$$

The aim of the next section is to find the kernel of the signature.

2 The use of signature for characterising the law

In this section, we show why and how the expected signature can extend the moments to path valued random variables, as it is done in [6].

2.1 The moment problem for vector-valued data

For real valued random variables, the moment problem consists in answering the following question: if I know the moments of my law, *i.e.* the sequence $(\int x^k \, \mathrm{d}\mu(x))_k$, can I know the law μ. There is no positive answer for any measure: the log normal distribution is not characterized by its moments. But for example if the measure is compactly supported, one can show using Stone-Weierstrass theorem that it is determined by its moments. It is not a necessary condition to be characterized by its moments, as the existence of an exponential moment for the law suffices for example (because it implies that the characteristic function admits a power series expansion in a neighbourhood of zero).

We can adapt the proof using Stone-Weierstrass theorem to vector-valued data. What will play the role of the moments is the sequence

$$\left(\mathbb{E}\left(\frac{X^{\otimes m}}{m!}\right)\right)_{m\geq 0},$$

which can be written as a formal power series

$$\mathbb{E}\left(\exp(X)\right) = \sum_{m=0}^{\infty} \frac{\mathbb{E}(X^m)T^m}{m!} \in \mathbb{R}[\![T]\!].$$

Using this tool, one can show the following theorem

Theorem 17 ([7])**.** Let $X, Y \in K \subset \mathbb{R}^d$, where K is compact, be two random vectors such that for all $m \geq 0$:

$$\mathbb{E}(X^{\otimes m}) = \mathbb{E}(Y^{\otimes m}).$$

Then X and Y have the same law.

2.2 Why the signature generalizes the idea of moments

We motivate the use of signature for characterizing a path as it is done in [6].

We saw in example 9 that the signature of $\gamma : [0,1] \to \mathbb{R}^d, t \mapsto tx$ is equal to

$$\sum_{k=0}^{\infty} \frac{x^{\otimes k}}{k!} =: \exp(x),$$

so if Γ is a random path which is almost surely of the form $t \mapsto tX$, where $X \in \mathbb{R}^d$ is a random vector, the expected signature of Γ corresponds to the moments of X. Starting from this, we show why integrating this when Γ is not almost surely linear could be interesting. If $\gamma \in \mathrm{BV}(\mathbb{R}^d)$, we could see $\mathrm{S}(\gamma)$ as "the exponential of γ". Let us give another reason that explains this analogy.

Let y be a function that satisfies the following differential equation:

$$y'(t) = a(t)y(t), y(0) = 1 \tag{10}$$

where $a \in \mathrm{C}^1(\mathbb{R}, \mathbb{R})$. We know that the solution is of the form $y(t) = \mathrm{e}^{\int_0^t a}$. One way to compute the solution of (10) is to calculate Picard iterations:

$$y_0(t) = 1, y_{n+1}(t) = 1 + \int_0^t a(s) y_n(s) \, \mathrm{d}s.$$

In our case, we can calculate them and we find

$$y_n(t) = \int_{0<t_1<\ldots<t_n<1} a(t_1) \cdots a(t_n) \, \mathrm{d}t_1 \cdots \mathrm{d}t_n = \sum_{k=0}^{n} \frac{A(t)^k}{k!},$$

which goes to the solution when n goes to infinity (A is such that $A' = a$ and $A(0) = 0$). Knowing the formal series $\sum \frac{X^k}{k!}$ and evaluating it in $a(t)$ gives access to the state of the system at time t.

Now let us consider a differential equation driven by a path $\gamma \in \mathrm{BV}(\mathbb{R}^d)$:

$$\mathrm{d}Y_t = V(Y_t) \, \mathrm{d}\gamma_t, Y_0 = y \in \mathbb{R}^p,$$

where $Y \in \mathrm{BV}(\mathbb{R}^p)$ and $V : \mathbb{R}^p \to \mathcal{L}(\mathbb{R}^d, \mathbb{R}^p)$. As before, we define Picard iterations by

$$Y_t^{n+1} = y + \sum_{i=1}^{d} \int_0^t V_i(Y_s^n) \, \mathrm{d}\gamma_s^i.$$

We can calculate them and find that

$$Y_t^n = y + \sum_{k=1}^{n} \sum_{i_1, \ldots, i_k=1}^{d} V_{i_k}(\cdots (V_{i_1}(y)) \cdots) \, \mathrm{S}(\gamma)_{0,t}^{i_1, \ldots, i_k},$$

so knowing the signature $\mathrm{S}(\gamma)_{0,t}$ gives access to the state of the system at time t. Indeed, one can check that the previous sum converges as n goes to infinity (we will do it later on in a simpler case, see Lemma 33), and that the limit is the solution of the differential equation. To see this as a generalization of what we did before, one can see that knowing the formal tensor power series $\mathrm{S}(\gamma)_{0,t}$ allows one to recover the state of the system at time t.

Let us see how we can use this generalized exponential to get generalized moments that can characterize the law.

2.3 A first result

In this paragraph, we show a simple case where the signature characterizes the path, and we show how we can use it to characterize the law with the expected signature, using the same type of arguments as in the previous paragraph. We will want to prove a density result, so we will use Stone-Weierstrass theorem, using the properties we show before to check its hypotheses.

Let $y_0 \in \mathbb{R}^{d-1}$ and let $\mathcal{Y}$ be the following subset of paths:

$$\mathcal{Y} := \{\gamma \in \mathrm{BV}(\mathbb{R}^d), \gamma_t^1 = t \, \forall t \in [0, 1], \gamma_0 = (0, y)\}.$$

On this set of paths, the signature characterizes the path.

Proposition 18 ([6, Prop 1.34])**.** Let X, Y be two paths of $\mathcal{Y}$ such that $\mathrm{S}(X) = \mathrm{S}(Y)$. Then $X = Y$.

Proof. This comes from the fact that for $k \geq 1$ ones and any $i \in \{2, \dots, d\}$:

$$\mathrm{S}(X)^{1,\dots,1,i} = \int_0^1 \frac{t^k}{k!} (X^i)'_t \, \mathrm{d}t,$$

and so because X and Y have the same signature, the measures $(X^i)'_t \, \mathrm{d}t$ and $(Y^i)'_t \, \mathrm{d}t$ on $[0,1]$ have the same moments and then are equal (see [3]). Finally, for all $t \in [0,1]$:

$$X_t = (0,y) + \int_0^t X'_u \, \mathrm{d}u = (0,y) + \int_0^t Y'_u \, \mathrm{d}u = Y_t.$$

□

Definition 19. A *linear form of the signature* is a mapping from $\mathrm{BV}(\mathbb{R}^d)$ into $\mathbb{R}$ that can be written

$$\gamma \mapsto \sum_{I \in \mathcal{I}} \lambda_I \, \mathrm{S}(\gamma)^I \tag{11}$$

where $\mathcal{I}$ is a finite set of multi-indexes and $\lambda_I \in \mathbb{R}$. We will denote by $\mathcal{F}(\mathcal{X})$ the set of all linear forms of the signature of paths in $\mathcal{X}$.

Proposition 20 ([6], page 29)**.** Let $\mathcal{X}$ be a compact subspace of $\mathrm{BV}(\mathbb{R}^d)$ (for the genuine bounded variation norm $\| \cdot \|_{\mathrm{BV}}$). The set $\mathcal{F}(\mathcal{X})$ is a unitary sub algebra of $\mathrm{C}((\mathcal{X}, \| \cdot \|_{\mathrm{BV}}), \mathbb{R})$.

Proof. The continuity comes from Proposition 7. It is clear that $\mathcal{F}(\mathcal{X})$ is a vector space, and it is an algebra thanks to shuffle-product identity (Theorem 12). It is unitary because 1 is the signature of any constant path. □

In order to apply Stone-Weierstrass theorem, the algebra has to separate the points. If we consider paths on $\mathcal{Y}$, it will be the case.

Proposition 21 ([6, Prop 1.35])**.** If $\mathcal{K}$ is a compact subset of $(\mathcal{Y}, \| \cdot \|_{\mathrm{BV}})$, then $\mathcal{F}(\mathcal{K})$ is a dense subspace of $\mathrm{C}(\mathcal{K}, \mathbb{R})$.

Proof. It follows from Propositions 18 and 20, together with Stone-Weierstrass theorem. □

Finally, we can mimic the argument we use for polynomials and measures on a compact of $\mathbb{R}$, but for the signature and measures on a compact subset of $\mathcal{Y} \subset \mathrm{BV}(\mathbb{R}^d)$.

Proposition 22 ([6, Cor 1.36])**.** Let $\mathbb{P}$ and $\mathbb{Q}$ be two probability measures on a compact space $\mathcal{K} \subset \mathcal{Y}$ (still with the bounded variation norm) such that

$$\int_{\mathcal{K}} \mathrm{S}(X) \; \mathbb{P}(\mathrm{d}X) = \int_{\mathcal{K}} \mathrm{S}(X) \, \mathbb{Q}(\mathrm{d}X),$$

which has to be understood the following way:

$$\text{if for all } I, \int_{\mathcal{K}} \mathrm{S}(X)^I \; \mathbb{P}(\mathrm{d}X) = \int_{\mathcal{K}} \mathrm{S}(X)^I \, \mathbb{Q}(\mathrm{d}X)$$

then $\mathbb{P} = \mathbb{Q}$.

Proof. We proceed the same way as for measures on $\mathbb{R}$. Thanks to the hypothesis we have that for all $\ell \in \mathcal{F}(\mathcal{K})$,

$$\int_{\mathcal{K}} \ell(X) \; \mathbb{P}(\mathrm{d}X) = \int_{\mathcal{K}} \ell(X) \, \mathbb{Q}(\mathrm{d}X),$$

and thanks to Proposition 21, it holds for any continuous mapping $f \in \mathrm{C}(\mathcal{K}, \mathbb{R})$. It is known that indicators of open subspaces of $\mathrm{C}(\mathcal{K}, \mathbb{R})$ can be monotonously approached by continuous functions, so that $\mathbb{P}(U) = \mathbb{Q}(U)$ for each U open subset of $\mathrm{C}(\mathcal{K}, \mathbb{R})$, and then $\mathbb{P} = \mathbb{Q}$. □

Why we want to obtain the kernel of the signature is clearer now: it could allow us to get density results that can get us to characterize the law of random paths (up to the equivalence relation induced by the kernel, which will be studied in the sequel).

3 Characterizing a path with the signature, tree-like equivalence

We saw in the previous section that if a path can be written $\gamma * \overleftarrow{\gamma}$, its signature is trivial. Unfortunately, being of the previous form is not a necessary condition to have trivial signature. The necessary and sufficient condition to have trivial signature is to be *tree-like*, and this is the notion we will study in this section. We will give some ideas of the proof of this theorem of Hambly and Lyons (see [10]).

3.1 Tree-like paths and tree-like equivalence

Definition 23. A path $\gamma \in \mathrm{C}([0,1],\mathbb{R}^d)$ (or any $[a,b]$) is a *tree-like* path is there exists a non-negative function, called *height* function, such that

- $h(0) = h(T) = 0$
- for all t, s, $\|\gamma_s - \gamma_t\| \leq h(s) + h(t) - 2\inf_{u\in[s,t]} h(u)$.

If h is of bounded variation, the path is called a *Lipschitz tree-like path.*

In order to see why it is called a tree-like path, we will explain the following theorem.

Theorem 24 ([10, Thm 7])**.** If $\gamma \in \mathrm{BV}(\mathbb{R}^d)$ is a tree-like path with height function h and $\gamma(0) = 0$, then there exists $\widetilde{h}$ a height function for γ such that it is also a tree-like path, and so γ is a Lipschitz tree-like path. Moreover, $\|\widetilde{h}\|_{\mathrm{BV}(\mathbb{R})} \leq \|\gamma\|_{\mathrm{BV}(\mathbb{R}^d)}$.

Proof. We will not prove the entire result, but only give the general idea. For $t \in [0,T]$, let

$$g_t : v \in [0,t] \mapsto \inf_{v<u<t} h(u).$$

Because this is an infimum on a decreasing set and because h is continuous, g_t is continuous and non-decreasing and we have

- $g_t(0) = 0$
- $g_t(t) = h(t)$.

The intermediate value theorem guarantees that $g_t([0,t]) = [0,h(t)]$ so we can define, for $x < h(t)$:

$$\tau_t(x) = \sup\{u \in [0,t], g_t(u) = x\} = \inf\{u \in [0,t], g_t(u) > x\}$$

which has to be understood as the last time when we have height x. We can then define

$$s \preccurlyeq t \text{ if and only if } \exists x \in [0,h(t)], \tau_t(x) = s.$$

Let us see γ as a path that as a shape of a tree, and see h as the "depth" of a point on the tree. The set $\{t, t \preccurlyeq t_0\}$, for a given t_0 has to be seen as the "shortest" path in the tree that leads to γ_{t_0}.
Using this intuition, it has been shown that $\preccurlyeq$ is a partial order, and that $\{t \preccurlyeq t_0\}$ is totally ordered for $\preccurlyeq$, in one to one correspondence with $[0,h(t_0)]$. For a fixed t and $x,y \in [0,h(t)]$, we have:

$$\begin{aligned}\|\gamma_{\tau_t(x)} - \gamma_{\tau_t(y)}\| &\leq h(\tau_t(x)) + h(\tau_t(y)) - 2\inf_{[\tau_t(x),\tau_t(y)]} h \\ &\leq x + y - 2\inf_{[x,y]} h\circ\tau_t = x + y - 2x = y - x,\end{aligned}$$

so that $\gamma\circ\tau_t$ is continuous and of bounded variation. In Figure 3.1, the path $\gamma\circ\tau_{t_0} : [0,h(t_0)] \to \mathbb{R}^d$ is the path whose trajectory is in green, going from the root to the point γ_{t_0} (we assume that $w' < w$). In our example, the paths $\gamma\circ\tau_w$ and $\gamma\circ\tau_{w'}$ correspond until the point $h(w_0)$. We can

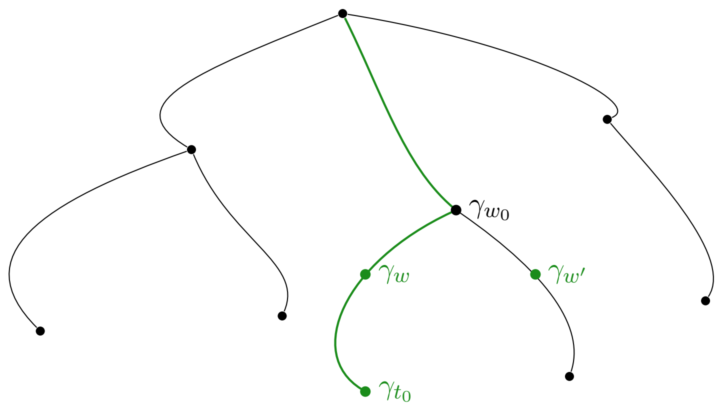

Figure 1: Example of tree-like path. The part of the tree in green is the subset $\{t \preccurlyeq t_0\}$. The points w and w' have same height on γ but $w' \not\preccurlyeq t_0$ because the last time that has height $h(w')$ until we reach γ_{t_0} is w'.

show (but we will not do it here), that we can always find such a w_0. If $\widetilde{h}(t)$ denotes the total variation of $\gamma \circ \tau_t$, we have, for w, w':

$$\begin{aligned}\|\gamma_w - \gamma_{w'}\| = \|\gamma_{\tau_w(w)} - \gamma_{\tau_{w'}(w')}\| &\leq \widetilde{h}(w) - \widetilde{h}(w_0) + \widetilde{h}(w') - \widetilde{h}(w_0) \\ &= \widetilde{h}(w) + \widetilde{h}(w') - 2\widetilde{h}(w_0).\end{aligned}$$

We can show (it is clear on the figure) that $\widetilde{h}(w_0) = \inf_{[w',w]} \widetilde{h}$, so $\widetilde{h}$ is a height function for γ, and we can prove that its total variation is bounded by the one of the path. $\square$

Remark 25.

- All the proof can be done without seeing the path γ as a tree in the figure. We have done this here because a tree-like path can be seen through the lens of $\mathbb{R}$-trees, which is done in [10].
- Conversely, any tree like path has less bounded variation than any height function for it.

3.2 The kernel of the signature is formed of tree-like paths

The goal of this paragraph is to give elements of the proof of the main theorem.

Theorem 26 (Hambly and Lyons, [10, Thm 1] (2010))**.** *Let $\gamma \in \mathrm{BV}(\mathbb{R}^d)$. Then γ has trivial signature if and only if it is a Lipschitz tree-like path.*

We will focus on giving ideas of why any tree-like path has trivial signature, and then only say a few words on the converse.

Definition 27. A *piecewise linear path* $\gamma \in \mathrm{BV}(\mathbb{R}^d)$ is a path such that there exists a partition $\{t_i\}$ such that γ is linear on each segment $[t_i, t_{i+1}]$. We say it is *non-degenerate* if each angle formed by $[\gamma_{t_{i-1}}, \gamma_{t_i}], [\gamma_{t_i}, \gamma_{t_{i+1}}]$ is non zero ($\mod \pi$).

We will admit the following important result.

Proposition 28 ([10, Thm 9])**.** *If γ is a non-degenerate piecewise linear path, it has non trivial signature.*

In the reference we give, it is done by quantitative estimates of the signature, with geometrical arguments.

Proposition 29 ([10, Cor 6.2])**.** *Any piecewise linear path γ that has trivial signature is tree-like, with a height function that has the same total variation as γ.*

Proof of Proposition 29. Let us proceed by induction on the minimum number r of edges of γ. We have $\mathrm{S}(\gamma) = 1$ so by Proposition 28 γ is degenerate and because r is minimal there is a segment in which γ goes back and forth. We can write:

$$\gamma = \gamma_- * \tau * \gamma_+,$$

where τ is a path that goes from a point a to a point b on a straight line then back from b to a following the same line. We already know that $\mathrm{S}(\tau) = 1$, so by Chen's identity:

$$1 = \mathrm{S}(\gamma) = \mathrm{S}(\gamma_-) \otimes 1 \otimes \mathrm{S}(\gamma_+) = \mathrm{S}(\gamma_- * \gamma_+)$$

By induction hypothesis, we conclude that $\gamma_- * \gamma_+$ is tree-like. Now it is not difficult to find a height function for the whole path thanks to the one we just obtained for $\gamma_- * \gamma_+$. □

To finish this side of the proof, we have to use the fact that any Lipschitz tree-like path can be approximated in L^1 by piecewise linear Lipschitz tree-like paths. Using this and the fact that iterated integrals are continuous in L^1, we obtain that any Lipschitz tree-like path has trivial signature.

Now let us say a few words on the proof of the converse of this result. It can be shown that any path $\gamma \in \mathrm{BV}(\mathbb{R}^d)$ could be approximated in $\mathrm{BV}(\mathbb{R}^d)$ by weakly piecewise linear paths (*i.e.* paths that can be written $\gamma_1 * \cdots * \gamma_r$, where each γ_i lives in a line) that can be written as integrals of γ under rank one forms. Using this, one can show that these approximations have trivial signature. Now, one can show that any weakly piecewise linear path that has trivial signature is tree-like, so that γ will be tree-like as well.

Theorem 30. Two paths $\gamma_1, \gamma_2 \in \mathrm{BV}(\mathbb{R}^d)$ have the same signature if and only if they are *tree-like equivalent*, noted $\gamma_1 \sim_t \gamma_2$, *i.e.* if and only if $\gamma_1 * \overleftarrow{\gamma_2}$ is tree-like. We denote by $\mathrm{P}(\mathbb{R}^d)$ the set of equivalence classes of $\mathrm{BV}(\mathbb{R}^d)$ under $\sim_t$, called the *reduced paths space.*

Corollary 31. Let $\mathbb{P}, \mathbb{Q}$ be two measures on a compact subset $\mathcal{K}$ of $\mathrm{BV}(\mathbb{R}^d)$. If $\mathbb{P}$ and $\mathbb{Q}$ have the same expected signatures, then $\mathbb{P} \circ \pi = \mathbb{Q} \circ \pi$, where $\pi : \mathrm{BV}(\mathbb{R}^d) \to \mathrm{P}(\mathbb{R}^d)$ is the canonical projection.

Proof. Now, the unitary algebra that separates the points is the set of linear forms of the signature from $\pi(\mathcal{K})$ to $\mathbb{R}$, on which we can apply the Stone-Weierstrass theorem. By the hypothesis, we have:

$$\begin{aligned} \forall \ell \in \mathcal{F}(\mathcal{K}), \int_{\pi(\mathcal{K})} \overline{\ell}(\pi(\gamma)) \; \mathbb{P} \circ \pi(\mathrm{d}X) &= \int_{\mathcal{K}} \ell(X) \; \mathbb{P}(\mathrm{d}X) \\ &= \int_{\mathcal{K}} \ell(X) \, \mathbb{Q}(\mathrm{d}X) = \int_{\pi(\mathcal{K})} \overline{\ell}(\pi(\gamma)) \, \mathbb{Q} \circ \pi(\mathrm{d}X), \end{aligned}$$

where $\overline{\ell} : \pi(\gamma) \mapsto \ell(\gamma)$. By Stone Weierstrass theorem, $\mathcal{F}(\pi(\mathcal{K}))$ is a dense subset of $\mathrm{C}(\pi(\mathcal{K}), \mathbb{R})$ for the quotient topology so using the same argument as before $\mathbb{P} \circ \pi = \mathbb{Q} \circ \pi$. □

4 Embedding the paths into a Reproducing Kernel Hilbert Space through signature

If we want to have a tool that characterize the law of a stochastic process of bounded variation, we must have to find more general than the signature. Previously, the expected signature was defined thanks to the compactness of the set of paths we were considering. For vector-valued data, we use the characteristic function rather than moments to characterize the law in the general case. Our goal in this section is to build a tool that will characterize the law of a stochastic process.

We will base our construction upon the signature (one can see the characteristic function as a "normalisation" of the moments by developing the exponential), and try to get density results (weaker than Stone Weierstrass) that will ensure the *characteristicness* of our tool. This will be done in the formalism of Reproducing Kernel Hilbert Spaces. The normalized signatures we will study now have been introduced by Chevyrev and Oberhauser in 2022 [7].

4.1 A Hilbert space for the signature

We will begin by introducing a Hilbert Space in which the signature will live. Until now, we saw elements of $\mathcal{T}(\mathbb{R}^d)$ as elements on a formal power series where the inderterminates $(e_i)_i$ are non commutative. These indeterminates must be seen as the canonical basis of $\mathbb{R}^d$, so that an element of $\mathcal{T}(\mathbb{R}^d)$ can be seen as an element of $\prod_{k\geq 0}(\mathbb{R}^d)^{\otimes k}$.

Definition 32 ([7, Def 10])**.** Let $\mathbf{T}(\mathbb{R}^d)$ be the Banach space

$$\mathbf{T}(\mathbb{R}^d) := \left\{ \mathbf{t} = (\mathbf{t}_k)_k \in \prod_{k\geq 0}(\mathbb{R}^d)^{\otimes k}, \|\mathbf{t}\|^2 := \sum_{k\geq 0} \|\mathbf{t}_k\|^2_{(\mathbb{R}^d)^{\otimes k}} < +\infty \right\}$$

If we equip it with the inner product $\langle \mathbf{u}, \mathbf{v}\rangle := \sum_{k=0}^{+\infty}\langle \mathbf{u}_k, \mathbf{v}_k\rangle$, it is a Hilbert Space.

We must check that this Hilbert space is compatible with the signature.

Lemma 33. Let $\gamma, x, y \in \mathrm{BV}(\mathbb{R}^d)$ and $k \geq 1$. Then we have

- $\|\,\mathrm{S}_k(x)\|_{(\mathbb{R}^d)^{\otimes k}} \leq \frac{|x|^k}{k!}$
- $\|\,\mathrm{S}_k(x) - \mathrm{S}_k(y)\|_{(\mathbb{R}^d)^{\otimes k}} \leq \frac{1}{(k-1)!}(|x| + |y|)^{k-1}|x-y|$

Corollary 34. The map $\mathrm{S} : \mathrm{BV}(\mathbb{R}^d) \to \mathbf{T}(\mathbb{R}^d)$ is well-defined and continuous with respect to the pseudo-metric defined on paths of bounded variation. Moreover, by definition of the quotient topology, $\mathrm{S} : \mathrm{BV}(\mathbb{R}^d)/\sim_t \to \mathbf{T}(\mathbb{R}^d)$ is continuous.

Proof. For the first point, we write:

$$\begin{aligned}
\|\,\mathrm{S}_k(\gamma)\|_{(\mathbb{R}^d)^{\otimes k}} &= \left\| \int_{0<t_1<\dots<t_k<1} \gamma'_{t_1}\otimes\dots\otimes\gamma'_{t_k}\,\mathrm{d}t_1\cdots\mathrm{d}t_k \right\|_{(\mathbb{R}^d)^{\otimes k}} \\
&\leq \int_{0<t_1<\dots<t_k<1} \|\gamma'_{t_1}\otimes\dots\otimes\gamma'_{t_k}\|_{(\mathbb{R}^d)^{\otimes k}}\,\mathrm{d}t_1\cdots\mathrm{d}t_k \\
&= \int_{0<t_1<\dots<t_k<1} \|\gamma'_{t_1}\|\cdots\|\gamma'_{t_k}\|\,\mathrm{d}t_1\cdots\mathrm{d}t_k \\
&= \frac{1}{k!}\left(\int_0^1 \|\gamma'_t\|\mathrm{d}t\right)^k \\
&= \frac{|\gamma|^k}{k!}.
\end{aligned}$$

And for the second one, we will use the following identity (for all $a_j, b_j \in \mathbb{R}^d$):

$$a_1\otimes\dots\otimes a_k - b_1\otimes\dots\otimes b_k = \sum_{j=1}^{k} b_1\otimes\dots\otimes b_{j-1}\otimes(a_j - b_j)\otimes a_{j+1}\cdots\otimes a_k,$$

in order to write

$$\begin{aligned}
\| \mathrm{S}_k(x) - \mathrm{S}_k(y)\|_{(\mathbb{R}^d)^{\otimes k}} &\leq \sum_{j=1}^{k} \int_{0<t_1<\cdots<t_k<1} \|y'_{t_1}\| \cdots \|y'_{t_{j-1}}\| \|x'_{t_j} - y'_{t_j}\| \|x'_{t_{j+1}}\| \cdots \|x'_{t_k}\| \,\mathrm{d}t_1 \cdots \mathrm{d}t_k \\
&= \sum_{j=1}^{k} \frac{1}{(j-1)!(k-j)!} \int_0^1 \|x'_{t_j} - y'_{t_j}\| \left(\int_0^{t_j} \|y'_s\| \,\mathrm{d}s\right)^{j-1} \left(\int_{t_j}^1 \|x'_s\| \,\mathrm{d}s\right)^{k-j} \mathrm{d}t_j \\
&\leq \frac{|x-y|}{(k-1)!} \sum_{j=1}^{k} \binom{k-1}{j-1} \|y\|^{j-1} \|x\|^{k-j} \\
&= \frac{(|x|+|y|)^{k-1}}{(k-1)!} |x-y|.
\end{aligned}$$

□

Because the map S is injective and continuous from the space of reduced paths into $\mathbf{T}(\mathbb{R}^d)$, we could be able to see a random path as its image by S, *i.e.* as a random element in $\mathbf{T}(\mathbb{R}^d)$.

4.2 Random variables in a Hilbert space and RKHSs

In this paragraph, H will denote a separable Hilbert space.

Proposition 35. A mapping $X : (\Omega, \mathcal{F}, \mathbb{P}) \to (H, \mathcal{B}(H))$ is a random variable if and only if $\langle x, X\rangle$ is a (real valued) random variable for any $x \in H$.

Proof. If X is a random variable, so are the $\langle x, X\rangle$ because $\langle x, \cdot\rangle$ is a continuous hence measurable mapping for all x. Conversely, because H is separable, one can write any ball as a countable intersection of hyperplanes, so that $X^{-1}(B) = \bigcap_{n\in\mathbb{N}} \langle x_n, \cdot\rangle^{-1}(\mathbb{R}_+)$ is measurable. As it is true for each ball, and as the balls span the topology of H, we get the result. □

Using this and Riesz representation theorem, we can define the expectation of a random variable defined in H by duality.

Definition 36. Let $X : \Omega \to H$ be a random variable. If $\mathbb{E}(\|X\|)$ is finite, there exists a unique $m_X \in H$ such that

$$\forall x \in H \,,\, \mathbb{E}(\langle x, X\rangle) = \langle x, m_X\rangle.$$

We define $\mathbb{E}(X) := m_X$, and we have $\| \mathbb{E}(X)\| \leq \mathbb{E}(\|X\|)$.

Definition 37. Let $(V, \langle \cdot, \cdot\rangle)$ be an inner product space, with $V \subset \mathbb{R}^{\mathcal{X}}$, where $\mathcal{X}$ is a set. It is a *reproducing kernel Hilbert Space* (RKHS) if it is a Hilbert space for which every linear form

$$\ell_x : f \mapsto f(x) \qquad x \in \mathcal{X}$$

is bounded.

As a consequence, in $(f_n) \in V^{\mathbb{N}}$ converges to f for the norm, we have for all x:

$$|f_n(x) - f(x)| \leq \|\ell_x\| \|f_n - f\|$$

so the convergence for the norm implies pointwise convergence.

Definition 38 (kernel of a RKHS,[1])**.** Let $V \subset \mathbb{R}^{\mathcal{X}}$ be a RKHS. There exists a unique

$$K : \mathcal{X} \times \mathcal{X} \to \mathbb{R}$$

such that $\langle f, K(\cdot, y)\rangle = f(y)$ for all $f \in V, y \in \mathcal{X}$. The application K is called the *kernel* of V.

Note that the symmetry of the inner product implies the one of K, but in general K is not bilinear (there is not even a law on $\mathcal{X}$). The existence and uniqueness of the kernel come from Riesz representation theorem. We say that K is *positive* if for all $r \geq 1$, $(c_i)_{1\leq i\leq r} \in \mathbb{R}^r$, $(x_i)_{1\leq i\leq r} \in \mathcal{X}^r$, one has

$$\sum_{i=1}^{r}\sum_{j=1}^{r} c_i c_j K(x_i, x_j) \geq 0,$$

One can check that the kernel of any RKHS is positive.

Theorem 39 (Moore-Aronszajn theorem, [1])**.** If K is a symmetric positive kernel on $\mathcal{X} \times \mathcal{X}$, there exists a unique RKHS for which K is a reproductive kernel.

This theorem is proven by considering a completion of the vector space V_0 spanned by the functions $\{K(x,\cdot), x \in \mathcal{X}\}$. As a consequence, V_0 is always dense in V and any Cauchy sequence in V_0 converges pointwise to an element of V. Conversely, if a subspace $V_0 \subset \mathbb{R}^{\mathcal{X}}$ is such that the ℓ_x are bounded and each Cauchy sequence converging pointwise to zero converges in norm to zero, then it has a RKHS completion.

Let us go back to our first example. We define a symmetric positive kernel on $\mathbb{R}^{\mathcal{K}}$, where $\mathcal{K}$ is a compact subset of $\mathcal{Y}$ (we take $\mathcal{Y}$ for a question of simplicity, but everything can be done in the reduced path space) by

$$K(X,Y) = \langle \mathrm{S}(X), \mathrm{S}(Y)\rangle.$$

Thanks to Moore-Aronszajn theorem, K defines a RKHS $H \subset \mathbb{R}^{\mathcal{K}}$. In this context, the embedding of a path $X \in \mathcal{K}$ into the RKHS H is $K(X,\cdot) = (Y \mapsto \langle \mathrm{S}(X), \mathrm{S}(Y)\rangle)$. What makes the embedding possible is that the mapping $Y \mapsto \mathrm{S}(Y)$ is injective. In a more general case, we would have to consider the embedding of an element of $\pi(\mathcal{K})$ into a RKHS defined by the signature.

Proposition 40 (continuity of functions in a RKHS, [13, Thm 2.3])**.** If $\mathcal{X}$ is a topological space and (V,K) is a RKHS ($V \subset \mathbb{R}^{\mathcal{X}}$). Then $V \subset \mathrm{C}(\mathcal{X},\mathbb{R})$ if and only if:

1. $K(x,\cdot) \in \mathrm{C}(\mathcal{X},\mathbb{R})$ for all $x \in \mathcal{X}$
2. $y \mapsto K(y,y)$ is bounded on a neighbourhood of every point $x \in \mathcal{X}$.

Proof. If V is composed of continuous functions, then so is V_0, so 1. holds. Therefore, $y \mapsto K(y,y) = \langle K(y,\cdot), K(y,\cdot)\rangle$ is continuous because the inner product is continuous.
Conversely, 1. implies that V_0 is a subspace of continuous functions. For $x \in \mathcal{X}$, $f \in V$, one has

$$|f(x)| \leq \sqrt{K(x,x)}\|f\|,$$

so that the norm convergence implies uniform convergence on $\{x, K(x,x) \leq M\}$ for all $M > 0$. But for $f \in V$, on can find $(f_n) \in V_0^{\mathbb{N}}$ such that $f_n \xrightarrow[n\to\infty]{\|\cdot\|} f$. Therefore, f_n converges uniformly to f in each $\{x, K(x,x) \leq M\}$ so thanks to 2. f is continuous. □

As a consequence of Lemma 33 and Proposition 40, we see that the RKHS defined previously by the signature is made of continuous functions.

Now let us introduce the notions that will be useful for our law characterization problem.

Definition 41. Let $X : \Omega \to \mathcal{X}$ be a random variable, and $V \subset \mathbb{R}^{\mathcal{X}}$ a RKHS defined by its kernel K. We assume it is a measurable kernel. The embedding of X into the RKHS V is

$$\check{X} := K(X,\cdot).$$

It is a random variable in the Hilbert space V.

Remark 42. If $\mathbb{E}(\|\check{X}\|) < \infty$, we know $\langle \mathbb{E}(\check{X}), f \rangle$ for all $f \in V$ so we know $\mathbb{E}(f(X))$ for all f in V. If V contains enough functions, we will be able to recover the law of X thanks to the expectation of the embedding $\check{X}$.

Definition 43. A RKHS (V, K) of $\mathbb{R}^{\mathcal{X}}$ is *characteristic* to $\mathcal{P}$ if the map $P \mapsto \mathbb{E}_P(\check{X})$ is injective, where X is a random variable distributed under P, and P belongs to the set of all probability measures of $\mathcal{P}$.

Being characteristic is not easy to check through this definition, so we use density results (as we did before).

Definition 44. If $\mathcal{X}$ is a topological space and $V \subset \mathrm{C}(\mathcal{X}, \mathbb{R})$ a RKHS of kernel K. It is *universal* to $\mathcal{F} \subset \mathrm{C}(\mathcal{X}, \mathbb{R})$ if it is dense in $\mathcal{F}$ with respect to the topology of $\mathcal{F}$ (which is expected to be weaker than the uniform topology).

There are many different types of universality and characteristicness, their link are studied in [15]. The topology we will consider is the strict topology, defined as follows.

Definition 45. Let $\mathcal{X}$ be a topological space. A function $\psi : \mathcal{X} \to \mathbb{R}$ *vanishes at infinity* if for all $\varepsilon > 0$, there exists a compact set $\mathcal{K} \subset \mathcal{X}$ such that $\sup_{\mathcal{X} \setminus \mathcal{K}} |\psi| < \varepsilon$. If $\mathrm{C}_0(\mathcal{X}, \mathbb{R})$ is the set of functions that vanish at infinity, we define the *strict topology* on $\mathrm{C}_\mathrm{b}(\mathcal{X}, \mathbb{R})$ as the topology generated by the seminorms

$$f \mapsto \sup_{x \in \mathcal{X}} |f(x)\psi(x)| \qquad \psi \in \mathrm{C}_0(\mathcal{X}, \mathbb{R}).$$

The argument that generalizes the use of Stone-Weierstrass theorem is the following.

Theorem 46. Let $\mathcal{X}$ be a topological space, and V a RKHS defined by its kernel K. If V_0 is a subalgebra of $\mathrm{C}_\mathrm{b}(\mathcal{X}, \mathbb{R})$ which separates the points and such that for all x, there exists f such that $f(x) \neq 0$, then K is universal to $\mathrm{C}_\mathrm{b}(\mathcal{X}, \mathbb{R})$ equipped with the strict topology.
Moreover, if elements of V_0 have finite expectancies in V under every positive measure on $\mathcal{X}$, this implies that K is characteristic to finite regular positive measures on $\mathcal{X}$.

We admit this theorem here, which uses a generalization of Stone-Weierstrass theorem (see [9] Thms 3.1 & 4.6), both with a duality method using Hahn-Banach theorem (see [14]).

Now, the goal is to create a kernel in which we can embed every path of $\mathrm{P}(\mathbb{R}^d)$, and in which we will be able to apply Theorem 46.

4.3 Tensor normalization

In order to get a tool that generalizes the characteristic function, we have to find a mapping built over the signature for which the expectation over any probability measure on $\mathrm{BV}(\mathbb{R}^d)$ will be finite. Let $\mathbf{T}_1(\mathbb{R}^d)$ denote the subset of $\mathbf{T}(\mathbb{R}^d)$ for which the first component is equal to one.

Definition 47. A *tensor normalization* is a continuous injective map of the form

$$\Lambda : \mathbf{T}_1(\mathbb{R}^d) \to \{\mathbf{t} \in \mathbf{T}_1(\mathbb{R}^d), \|\mathbf{t}\| \leq R\}, \qquad \mathbf{t} \mapsto \delta_{\lambda(\mathbf{t})}(\mathbf{t}),$$

where $R > 0$ is a constant, $\lambda : \mathbf{T}_1(\mathbb{R}^d) \to \mathbb{R}_+^*$ is a function, and $\delta_a(\mathbf{t})$ denotes $(a^k \mathbf{t}_k) \in \mathbf{T}_1(\mathbb{R}^d)$.

It has been shown in [7] that such maps actually exist. We will not show it here.

Theorem 48 ([7, Thm 21])**.** Let Λ be a tensor normalization, and define the *normalized signature* $\Phi := \Lambda \circ \mathrm{S}$. Then

1. Φ is a continuous injection from $\mathrm{P}(\mathbb{R}^d)$ into a bounded subset of $\mathbf{T}_1(\mathbb{R}^d)$.

2. $\langle \Phi(\cdot), \Phi(\cdot) \rangle$ defines a characteristic kernel on probability measures on $\mathrm{P}(\mathbb{R}^d)$

Proof. The first point comes from the definition of a tensor normalization. As a consequence of 1., $\mathbb{E}(\check{X})$ is defined for every random variable X on $\mathrm{P}(\mathbb{R}^d)$. Then, thanks to the definition of Λ and thanks to the shuffle product identity, one can show that V_0 satisfies the required conditions to apply Theorem 46, which gives the desired result. □

As a consequence of Theorem 46, one can also show that the normalized signature defines a continuous injection and a characteristic kernel on the whole $\mathcal{Y}$ space, equipped with the genuine bounded variation topology.

We conclude this document by saying a few words on why RKHSs are useful in statistics, and why it can be useful in our case.

The RKHS theory allows to linearise problems that are not linear in the first place. For example, kernel SVM generalizes linear classification: a non-linear separation problem becomes a linear problem, but in a RKHS which is larger than the original space (see [2], 5.3.2).

Also, we are able to see probability measures as vectors of a Hilbert space. As we said before, we can embed a probability measure or a random variable into a RKHS $V \subset \mathbb{R}^{\mathcal{X}}$ (with kernel K) by $X \mapsto \mathbb{E}(\check{X})$, and if the RKHS is large enough, this embedding is an injection. We can define a metric on the space of probability measures on $\mathcal{X}$ thanks to the RKHS norm. Using the definition of a norm and the reproducing property of the kernel, one can show this metric is

$$\| \mathbb{E}_{X\sim\mu, Y\sim\nu}(K(X-Y,\cdot))\|_V^2 =: d_K^2(\mu,\nu) = \mathbb{E}(K(X,X')) - 2\,\mathbb{E}(K(X,Y)) + \mathbb{E}(K(Y,Y')),$$

where X' and Y' are independent copies of $X \sim \mu$ and $Y \sim \nu$ respectively.

In our case, the embedding is an injection if we consider measures on $\mathrm{P}(\mathbb{R}^d)$ or on $\mathcal{Y}$, with the kernel defined by the normalized signature. It has been shown in [7] that the underlying metric is weaker than the convergence in law, and they give conditions so that they are equivalent (see Thm 30). As it is easy to calculate (there exist algorithms to compute inner products of signature, see [11]), it can be a useful tool to build a statistic test. For instance, Chevyrev and Oberhauser [7] used it to build a two-sample test for stochastic processes.

References

[1] N. Aronszajn. Theory of reproducing kernels. *Transactions of the American Mathematical Society*, 68(3):337–404, 1950.

[2] A. Berlinet and C. Thomas-Agnan. *Reproducing Kernel Hilbert Spaces in Probability and Statistics*. Springer US, 2003.

[3] P. Billingsley. *Measure and Probability*. John Wiley & Sons: New York, 1995.

[4] H. Boedihardjo, X. Geng, T. Lyons, and D. Yang. The signature of a rough path: Uniqueness, 2015.

[5] K.-T. Chen. Iterated integrals and exponential homomorphisms. *Proceedings of The London Mathematical Society*, pages 502–512, 1954.

[6] I. Chevyrev and A. Kormilitzin. A primer on the signature method in machine learning, 2025.

[7] I. Chevyrev and H. Oberhauser. Signature moments to characterize laws of stochastic processes, 2022.

[8] T. Fawcett. *Problems in Stochastic Analysis: Connections Between Rough Paths and Non-commutative Harmonic Analysis*. University of Oxford, 2003.

[9] R. Giles. A generalization of strict topology. *Transactions of the American Mathematical Society*, 1971.

[10] B. Hambly and T. Lyons. Uniqueness for the signature of a path of bounded variation and the reduced path group. *Annals of Mathematics*, 171(1):109–167, Mar. 2010.

[11] F. J. Király and H. Oberhauser. Kernels for sequentially ordered data, 2016.

[12] T. Lyons, H. Ni, and H. Oberhauser. A feature set for streams and an application to high-frequency financial tick data. In *Proceedings of the 2014 International Conference on Big Data Science and Computing*, BigDataScience '14, New York, NY, USA, 2014. Association for Computing Machinery.

[13] S. Saitoh. *Theory of Reproducing Kernels and Applications*. Springer, Berlin, Heidelberg, 2016.

[14] C.-J. Simon-Gabriel and B. Schölkopf. Kernel distribution embeddings: Universal kernels, characteristic kernels and kernel metrics on distributions. *Journal of Machine Learning Research*, 19(44):1–29, 2018.

[15] B. K. Sriperumbudur, K. Fukumizu, and G. R. Lanckriet. Universality, characteristic kernels and rkhs embedding of measures. *Journal of Machine Learning Research*, 12(70):2389–2410, 2011.

[16] W. P. Ziemer. *Weakly differentiable functions*. Springer-Verlag, Berlin, Heidelberg, 1989.