Supplementary Material of "Signature-based validation of real-world economic scenarios"

Hervé Andrès[1,2], Alexandre Boumezoued[1], and Benjamin Jourdain[2]

[1]Milliman R&D, Paris, France
[2]CERMICS, École des Ponts, INRIA, Marne-la-Vallée, France.

February 27, 2024

The signature being already defined in the paper, the purpose of this supplementary material is to provide more insights on the signature thanks to examples and a brief overview of its main properties.

Appendix A. Some examples

First, we present several examples that allow to better understand the signature and the log-signature.

Example A.1. *If $X : [0,T] \to E$ is a linear path , i.e. $X_t = X_0 + (X_T - X_0)\frac{t}{T}$, then for any $n \geq 0$:*

$$\mathbf{X}^n = \frac{1}{n!}(X_T - X_0)^{\otimes n}. \tag{A.1}$$

Example A.2. *If E is a vector space of dimension 2, the second order term of the signature is given by:*

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

Note that the difference of the anti-diagonal coefficients of $\mathbf{X}^2$ corresponds, up to a factor $1/2$, to the Lévy area of the curve $t \mapsto (X_t^1, X_t^2)$ which is defined as:

$$\mathcal{A}^{Levy} = \frac{1}{2}\left(\int_0^T (X_t^1 - X_0^1)dX_t^2 - \int_0^T (X_t^2 - X_0^2)dX_t^1\right). \tag{A.3}$$

It is the signed area between the curve and the chord connecting the two endpoints (see Figure 1).

In Section 3.1.2., we mentioned that the lead-lag transformation allows to capture the quadratic variation of a path in the signature. More precisely, the Levy area of the lead-lag transformation is the quadratic variation up to a factor 1/2 as stated by the following proposition which is a direct consequence of the definition of the lead-lag transformation.

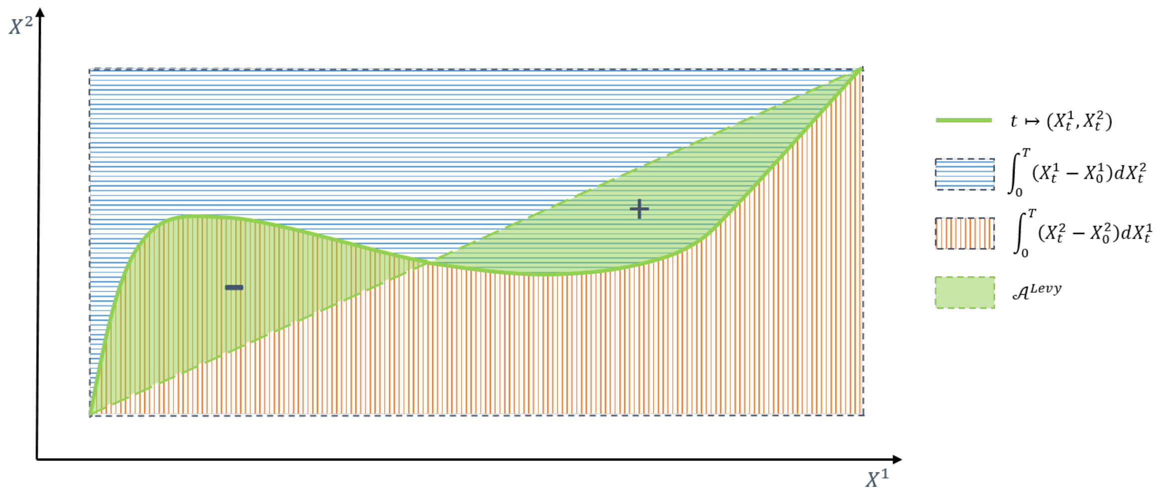

Figure 1: Illustration of the Lévy area. The blue dashed area corresponds to the integral $\int_0^T (X_t^1 - X_0^1) dX_t^2$ while the red dashed area corresponds to the integral $\int_0^T (X_t^2 - X_0^2) dX_t^1$. Taking the difference between these two areas yields the Levy area (represented in green transparent) up to a factor 2 because of a double counting. The '+' (resp. '−') sign indicates that the surrounding area is counted positively (resp. negatively).

Proposition A.1. *Let $t_0 = 0 < t_1 < \cdots < t_N = T$ be a partition of $[0, T]$ and $(X_{t_i})_{i=0,\ldots,N}$ be the vector of observations of a real-valued process X on this partition. The Levy area of the lead-lag transformation of $(X_{t_i})_{i=0,\ldots,N}$ is equal to the quadratic variation of X on the partition $(t_i)_{i=0,\ldots,N}$ up to a factor $1/2$, i.e.*

$$\frac{1}{2}\left(\int_0^T (X_t^{lead} - X_0^{lead}) dX_t^{lag} - \int_0^T (X_t^{lag} - X_0^{lag}) dX_t^{lead}\right) = \frac{1}{2}\sum_{i=0}^{N-1} (X_{t_{i+1}} - X_{t_i})^2. \tag{A.4}$$

Remark A.1. *We also mentioned in Section 3.1.2. that the cumulative lead-lag transformation $\tilde{X}$ of a sequence of observations $(X_{t_i})_{i=0,\ldots,N}$ on $[0, T]$ can be related to the statistical moments of X. Indeed, the term of order 1 of the signature of $\tilde{X}$ is given by:*

$$\tilde{\mathbf{X}}^1 = \begin{pmatrix} \tilde{X}_T - \tilde{X}_0 \\ \tilde{X}_T - \tilde{X}_0 \end{pmatrix} = \begin{pmatrix} \sum_{i=0}^N X_{t_i} \\ \sum_{i=0}^N X_{t_i} \end{pmatrix} \tag{A.5}$$

which is the empirical mean of X up to a factor $1/(N+1)$. From Proposition A.1, we also deduce that the Levy area of the cumulative lead-lag transformation is given by $\frac{1}{2}\sum_{i=0}^N (\tilde{X}_{t_{i+1}} - \tilde{X}_{t_i})^2 = \frac{1}{2}\sum_{i=0}^N X_{t_i}^2$ which is the empirical second order (non-central) moment of X up to a factor $1/(N+1)$. More generally, the n-th (non-central) moment of X can be obtained from the term of order n of the signature of the cumulative lead-lag transformation.

We have seen in our numerical experiments in Section 3.2.2. that the lead-lag transformation is not always sufficient to distinguish models that are too close from a statistical perspective. In the following example, we show that, as the time step converges to 0, the first two terms of the signature of the lead-lag transformation of a driftless Black-Scholes dynamics with constant volatility have the same distributions as the first two terms of the signature of the lead-lag transformation of a driftless Black-Scholes dynamics with a time-dependent deterministic volatility if the total variances at time T of both models are the same.

Example A.3. *Consider X and Y the solutions of the following SDE's*

$$dX_t = \sigma X_t dW_t \text{ and } dY_t = \gamma(t) Y_t dW_t \tag{A.6}$$

with $X_0 = Y_0 = 1$ and where $(W_t)_{t\geq 0}$ is a Brownian motion and γ is a deterministic function satisfying $\int_0^T \gamma(t)^2 dt = \sigma^2 T$. The explicit formulas of X and Y write:

$$X_t = \exp\left(\sigma W_t - \frac{1}{2}\sigma^2 t\right), \quad Y_t = \exp\left(\int_0^t \gamma(s) dW_s - \frac{1}{2}\int_0^t \gamma(s)^2 ds\right). \tag{A.7}$$

Let us denote by $\hat{X}_N$ (resp. $\hat{Y}_N$) the lead-lag transformation of X (resp. Y) on a partition $(t_i)_{i=0,\dots,N}$ of $[0,T]$ such that $t_i = iT/N$. The constraint $\int_0^T \gamma(t)^2 dt = \sigma^2 T$ implies that $X_T \overset{d}{=} Y_T$ so the first order terms of the signatures of $\hat{X}_N$ and $\hat{Y}_N$ (which reduce to the increments of X and Y over $[0,T]$) have the same distribution for all $N \geq 1$. The second order term of the signature of $\hat{Y}_N$ is given by (see the proof of the above proposition):

$$\hat{\mathbf{Y}}_N^2 = \begin{pmatrix} \frac{1}{2}(Y_T - Y_0)^2 & \sum_{i=0}^{N-1}\left[(Y_{t_{i+1}} - Y_{t_i})^2 + (Y_{t_i} - Y_0)(Y_{t_{i+1}} - Y_{t_i})\right] \\ \sum_{i=0}^{N-1}(Y_{t_i} - Y_0)(Y_{t_{i+1}} - Y_{t_i}) & \frac{1}{2}(Y_T - Y_0)^2 \end{pmatrix} \tag{A.8}$$

Now, given that Y is a square-integrable continuous martingale, the coefficient at position $(1,2)$ of $\hat{\mathbf{Y}}_N$ converges in probability as $N \to +\infty$ to $\langle Y\rangle_T + \int_0^T (Y_t - Y_0) dY_t = \frac{1}{2}\left[(Y_T - Y_0)^2 + \langle Y\rangle_T\right]$ where $\langle Y\rangle$ denotes the quadratic variation process of Y and the equality is obtained using the integration by parts formula. Similarly, the coefficient at position $(2,1)$ of $\hat{\mathbf{Y}}_N$ converges in probability as $N \to +\infty$ to $\int_0^T (Y_t - Y_0) dY_t = \frac{1}{2}\left[(Y_T - Y_0)^2 - \langle Y\rangle_T\right]$. The same convergences hold for $\hat{\mathbf{X}}_N$. Now remark that the processes $\left(\int_0^t \gamma(s) dW_s\right)_{t\geq 0}$ and $\left(W_{\int_0^t \gamma(s)^2 ds}\right)_{t\geq 0}$ are both Gaussian processes with the same mean and the same covariance function, we deduce that they have the same distribution. Analogously, $(\sigma W_t)_{t\geq 0}$ has the same distribution as $(W_{\sigma^2 t})$. We deduce that:

$$(X_t)_{t\geq 0} \overset{d}{=} \left(\exp\left(W_{\sigma^2 t} - \frac{1}{2}\sigma^2 t\right)\right)_{t\geq 0} \text{ and } (Y_t)_{t\geq 0} \overset{d}{=} \left(\exp\left(W_{\int_0^t \gamma(s)^2 ds} - \frac{1}{2}\int_0^t \gamma(s)^2 ds\right)\right)_{t\geq 0} \tag{A.9}$$

Setting $\varphi(t) = \frac{1}{\sigma^2}\int_0^t \gamma(s)^2 ds$, we deduce that $(Y_t)_{0\leq t\leq T} \overset{d}{=} (X_{\varphi(t)})_{0\leq t\leq T}$. As a consequence, $(Y_t, \langle Y\rangle_t)_{0\leq t\leq T}$ has the same distribution as $(X_{\varphi(t)}, \langle X\rangle_{\varphi(t)} - \langle X\rangle_{\varphi(0)})_{0\leq t\leq T}$. Since $\varphi(0) = 0$ and $\varphi(T) = T$, we conclude that the limit of $\hat{\mathbf{Y}}_N$ has the same distribution as the limit of $\hat{\mathbf{X}}_N$.

The log-signature

We now introduce more formally the log-signature. We recall that the space of formal series of tensors is defined as:

$$T(E) = \left\{(\mathbf{t}^n)_{n\geq 0} \mid \forall n \geq 0, \mathbf{t}^n \in E^{\otimes n}\right\} \tag{A.10}$$

with the convention $E^{\otimes 0} = \mathbb{R}$. This space can be equipped with the following operations: for $\mathbf{t}$, $\mathbf{u} \in T(E)$, $\lambda \in \mathbb{R}$,

$$\begin{aligned} \mathbf{t} + \mathbf{u} &= (\mathbf{t}^n + \mathbf{u}^n)_{n\geq 0} \\ \lambda\mathbf{t} &= (\lambda\mathbf{t}^n)_{n\geq 0} \\ \mathbf{t} \otimes \mathbf{u} &= \left(\mathbf{v}^n = \sum_{k=0}^n \mathbf{t}^k \otimes \mathbf{u}^{n-k}\right)_{n\geq 0}. \end{aligned} \tag{A.11}$$

Since by convention the term of order 0 of the signature is set to 1, the signature takes its values in the following affine subspace of $T(E)$:

$$T_1(E) = \left\{\mathbf{t} \in T(E) \mid \mathbf{t}^0 = 1\right\}. \tag{A.12}$$

A closely related subspace of $T(E)$ is the following:

$$T_0(E) = \left\{ \mathbf{t} \in T(E) \mid \mathbf{t}^0 = 0 \right\}. \tag{A.13}$$

In fact, there is a bijection between $T_1(E)$ and $T_0(E)$ (Lemma 2.21 in Lyons et al., 2007):

Proposition A.2. *Let us define respectively the exponential and logarithm mappings as:*

$$\begin{aligned} \exp: \quad T_0(E) &\to T_1(E) \\ \mathbf{t} &\mapsto \exp(\mathbf{t}) := \sum_{n \geq 0} \frac{\mathbf{t}^{\otimes n}}{n!} \end{aligned} \quad \textit{and} \quad \begin{aligned} \log: \quad T_1(E) &\to T_0(E) \\ \mathbf{t} &\mapsto \log(\mathbf{t}) := \sum_{n \geq 1} \frac{(-1)^{n-1}}{n} (\mathbf{t} - \mathbf{1})^{\otimes n} \end{aligned} \tag{A.14}$$

with the convention $t^{\otimes 0} = 1$ and where $\mathbf{1} = (1, 0, \dots, 0, \dots) \in T_1(E)$. The exponential mapping is bijective from $T_0(E)$ to $T_1(E)$ and its inverse is the logarithm mapping.

Example A.1 (continued)**.** *Using the exponential and the logarithm mappings, we can rewrite the signature in Example A.1 in the following way:*

$$S(X) = \exp(X_T - X_0) \tag{A.15}$$

where $X_T - X_0$ should be interpreted as the element $(0, X_T - X_0, 0, \dots, 0, \dots)$ of $T_0(E)$. Moreover,

$$\log(S(X)) = X_T - X_0. \tag{A.16}$$

Using the logarithm, it is therefore possible to define the log-signature of a path X as $\log(S(X))$. Although there is a one-to-one correspondence between the signature and the log-signature, the log-signature is a more parsimonious representation of the path than the signature in the sense that it removes the redundancies. This can be seen in Example A.1: the only non-zero term of the log-signature of a linear path is the term of order 1 which contains the increments of the path. In comparison to the signature, all the powers of the increments have disappeared. However, no information is lost. More generally, it can be shown (see for example Liao et al., 2019) that the log-signature has more zeros than the signature. As such, it represents a useful object for applications as it allows to avoid the exponential increase of the size of the truncated signature with the order. Indeed, if E is a vector space of dimension d, the term of order n of the signature has d^n elements.

Example A.4. *Let us consider $X \in \mathcal{C}^1([0,T], \mathbb{R}^2)$. The second order term of the log-signature writes:*

$$\mathbf{lX}^2 = \mathbf{X}^2 - \frac{1}{2} \mathbf{X}^1 \otimes \mathbf{X}^1 \tag{A.17}$$

where $\mathbf{X}^2$ comes from the first term ($n = 1$) of the log series in Equation (A.14) and $\mathbf{X}^1 \otimes \mathbf{X}^1$ comes from the second term ($n = 2$). We have:

$$\mathbf{X}^2 = \begin{pmatrix} \frac{(X_T^1 - X_0^1)^2}{2} & \int_0^T (X_t^1 - X_0^1) dX_t^2 \\ \int_0^T (X_t^2 - X_0^2) dX_t^1 & \frac{(X_T^2 - X_0^2)^2}{2} \end{pmatrix} \tag{A.18}$$

and

$$\mathbf{X}^1 \otimes \mathbf{X}^1 = \begin{pmatrix} (X_T^1 - X_0^1)^2 & (X_T^1 - X_0^1)(X_T^2 - X_0^2) \\ (X_T^1 - X_0^1)(X_T^2 - X_0^2) & (X_T^2 - X_0^2)^2 \end{pmatrix}. \tag{A.19}$$

Using the integration by part formula, we obtain:

$$\mathbf{lX}^2 = \underbrace{\frac{1}{2} \left(\int_0^T (X_t^1 - X_0^1) dX_t^2 - \int_0^T (X_t^2 - X_0^2) dX_t^1 \right)}_{\textit{Lévy area of } X} \begin{pmatrix} 0 & 1 \\ -1 & 0 \end{pmatrix} \tag{A.20}$$

Hence, the second order term of the log-signature reduces to the Lévy area.

Remark A.2. *Note that only the N first terms of the logarithm series (A.14) contribute to the N-th term of the log-signature. Indeed, for $n > N$, the contributions to $E^{\otimes N}$ of $(\mathbf{t}-\mathbf{1})^{\otimes n}$ always involve some product by $(\mathbf{t}-\mathbf{1})^0 = 0$.*

Appendix B. Properties

We have seen in the first subsection that the signature allows to capture some information about the path. A natural question at this stage is how much information about X does the signature of X contain. This subsection aims at answering this question.

Proposition B.1 (Invariance under time reparametrization). *Let $X \in \mathcal{C}^1([0,T],E)$ and consider $\varphi : [0,T] \to [0,T]$ a non-decreasing surjection. If we set $\tilde{X}_t = X_{\varphi(t)}$, then:*

$$S(\tilde{X}) = S(X). \tag{B.1}$$

This first property (see Proposition 7.10 in Friz and Victoir, 2010 for a proof) means that the speed at which the path is traversed is not captured by the signature. The signature is also invariant by translation. Indeed, if we define $\bar{X}_t = x + X_t$, then $d\bar{X}_t = dX_t$ and by definition of the signature we have $S(\bar{X}) = S(X)$. The next property we will outline is Chen's identity. Before introducing it, we need the following definition.

Definition B.1 (Concatenation). *Let $X \in \mathcal{C}^1([0,t],E)$ and $Y \in \mathcal{C}^1([t,T],E)$. The concatenation of X and Y is the path in $\mathcal{C}^1([0,T],E)$ defined as:*

$$(X * Y)_s = \begin{cases} X_s & \text{if } s \in [0,t] \\ X_t + Y_s - Y_t & \text{if } s \in [t,T]. \end{cases} \tag{B.2}$$

Theorem B.1 (Chen's identity). *Let $X \in \mathcal{C}^1([0,t],E)$ and $Y \in \mathcal{C}^1([t,T],E)$. Then,*

$$S_{[0,T]}(X * Y) = S_{[0,t]}(X) \otimes S_{[t,T]}(Y). \tag{B.3}$$

A proof can be found in Theorem 2.9 of Lyons et al. (2007). A useful application of Chen's identity is the computation of the signature of a piecewise linear path. Let $(t_i)_{0 \le i \le n}$ be a subdivision of $[0,T]$ and $X : [0,T] \to E$ be a path such that for $t \in [t_i, t_{i+1}]$ with $0 \le i \le n-1$,

$$X_t = X_{t_i} + \frac{X_{t_{i+1}} - X_{t_i}}{t_{i+1} - t_i}(t - t_i). \tag{B.4}$$

Then by Chen's identity and by using that $S_{[t_i,t_{i+1}]}(X) = \exp(X_{t_{i+1}} - X_{t_i})$ (since X is linear on each $[t_i, t_{i+1}]$),

$$S_{[0,T]}(X) = \bigotimes_{i=0}^{n-1} S_{[t_i,t_{i+1}]}(X) = \bigotimes_{i=0}^{n-1} \exp(X_{t_{i+1}} - X_{t_i}). \tag{B.5}$$

In general, the right hand side cannot be simplified to $\exp(X_T - X_0)$ because the tensor product $\otimes$ is not commutative. Another consequence of Chen's identity is the following proposition (Proposition 2.14 in Lyons et al., 2007).

Proposition B.2 (Time-reversal). *Let $X \in \mathcal{C}^1([0,T],E)$. Define $\overleftarrow{X}$ as $\overleftarrow{X}_t = X_{2T-t}$ for $t \in [T,2T]$. Then,*

$$S_{[0,2T]}(X * \overleftarrow{X}) = S_{[0,T]}(X) \otimes S_{[T,2T]}(\overleftarrow{X}) = \mathbf{1} \tag{B.6}$$

where we recall that $\mathbf{1} = (1, 0, \dots, 0, \dots) \in T_1(E)$.

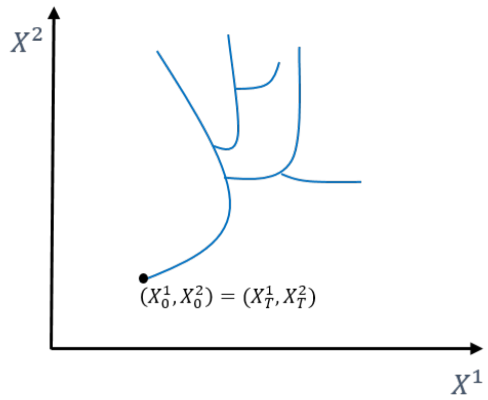

Figure 2: Example of a tree like path.

Because constant paths also have **1** as signature, the above proposition implies that $X * \overleftarrow{X}$ has the same signature as constant paths.

Due to the invariance by reparametrisation and by translation and the time-reversal property, it is clear that if two paths have the same signature, then they are not necessarily equal. In other words, the signature mapping is not injective. Fortunately, the presented invariances and the time-reversal property are essentially the only cases when paths can differ but have the same signature. To make this precise, we need the notion of tree-like paths.

Definition B.2 (Tree-like path)**.** *A path $X : [0, T] \to E$ is tree-like if there exists a continuous function $h : [0, T] \to [0, +\infty[$ such that $h(0) = h(T) = 0$ and for all $s, t \in [0, T]$ with $s \leq t$:*

$$\|X_t - X_s\|_E \leq h(s) + h(t) - 2 \inf_{u \in [s,t]} h(u). \tag{B.7}$$

This function is called a height function for the path X.

Remark B.1. *Note that a tree-like path necessarily satisfies $X_0 = X_T$. Indeed, by Definition B.2:*

$$\|X_T - X_0\|_E \leq h(0) + h(T) - 2 \inf_{u \in [0,T]} h(u) = 0 \tag{B.8}$$

because $h(0) = h(T) = 0$ and h is non-negative. Therefore, one way to turn a tree-like path into a path that is not tree-like is to consider the path $t \mapsto (t, X_t)$ obtained as the time transformation of X.

As suggested by their name, tree-like paths are paths whose graph looks like a tree (see Figure 2), i.e. an acyclic and connected graph in graph theory and the height function h corresponds to the depth of each node of the tree in a depth-first search. Another equivalent way to see tree-like paths is to see them as paths that can be reduced to a constant path by removing pieces of the form $W * \overleftarrow{W}$. For example, if X and Y are non-constant paths, $X * Y * \overleftarrow{Y} * \overleftarrow{X}$ is an example of tree-like path. This notion of tree-like paths is crucial to understand the information that is not captured by the signature as Hambly and Lyons (2010) showed that the signature determines the path up to tree-like equivalence, which we will now define.

Definition B.3 (Tree-like equivalence)**.** *For X and Y two paths, we say that X and Y are tree-like equivalent if $X * \overleftarrow{Y}$ is a tree-like path. This relation is denoted by $X \sim_t Y$.*

We can now state Hambly and Lyons's theorem.

Theorem B.2. *Let $X \in \mathcal{C}^1([0,T],E)$. Then $S(X) = \mathbf{1}$ if and only if X is a tree-like path. Moreover, if $Y \in \mathcal{C}^1([0,T],E)$ is another bounded variation path, then $S(X) = S(Y)$ if and only if $X \sim_t Y$.*

This theorem can be understood as follows: two paths will have the same signature if and only if one can be obtained from the second by using translations, by changing the traversal speeds and by removing parts of the form $W * \overleftarrow{W}$. This uniqueness result has then been extended to a more general class of paths (namely weakly geometric rough paths) by Boedihardjo et al. (2016).

Remark B.2. *The conclusion of Theorem B.2 still holds if the signature is replaced by the log-signature since the log mapping is a bijection. Note however that the first statement of the theorem should be modified as follows: $\log(S(X)) = \mathbf{0}$ if and only if X is a tree-like path where $\mathbf{0} = (0, \dots, 0, \dots) \in T_0(E)$.*

We have seen that in dimension 1, the signature only captures the path increment between 0 and T (see Example 2.1. in Section 2.2.) so that the signature will only allow to distinguish paths X and Y such that $X_T - X_0 \neq Y_T - Y_0$. This result is actually a consequence of the following proposition and of Theorem B.2.

Proposition B.3. *If E is a one-dimensional real vector space and X, Y are E-valued paths such that $X_T - X_0 = Y_T - Y_0$, then X and Y are tree-like equivalent.*

Proof. Since any one-dimensional real vector space is isometrically isomorph to $\mathbb{R}$, we can assume that $E = \mathbb{R}$. Let X and Y be two paths from $[0,T]$ to $\mathbb{R}$ such that $X_T - X_0 = Y_T - Y_0$. Let us set $Z = X * \overleftarrow{Y}$ and $h(t) = |Z_t - Z_0|$ for $t \in [0, 2T]$. Using the definition of concatenation operator and the fact that $X_T - X_0 = Y_T - Y_0$, we have $Z_0 = X_0$ and $Z_{2T} = X_T + Y_0 - Y_T = X_0$ so that $h(0) = h(2T) = 0$. The non-negativity of h results from the non-negativity of the absolute value. Moreover, the continuity of X and Y imply the continuity of Z by definition of the concatenation operator, so h is continuous as well. The only remaining property to show is inequality (B.7). Let $s, t \in [0, 2T]$ with $s \leq t$. Let us assume that $Z_s \leq Z_t$ (the proof in the case $Z_t \leq Z_s$ is similar) so that $|Z_t - Z_s| = Z_t - Z_s = Z_t - Z_0 - (Z_s - Z_0)$. We distinguish three cases:

- If $Z_0 \leq Z_s \leq Z_t$, then $h(t) = Z_t - Z_0$ and $h(s) = Z_s - Z_0$. Thus,

$$|Z_t - Z_s| = h(t) - h(s) \leq h(t) - \inf_{u \in [s,t]} h(u) \leq h(t) + h(s) - 2 \inf_{u \in [s,t]} h(u). \tag{B.9}$$

- If $Z_s \leq Z_0 \leq Z_t$, then $h(t) = Z_t - Z_0$ and $h(s) = Z_0 - Z_s$. Thus,

$$|Z_t - Z_s| = h(t) + h(s) = h(t) + h(s) - 2 \inf_{u \in [s,t]} h(u) \tag{B.10}$$

  because by the intermediate value theorem, there exists $v \in [s,t]$ such that $Z_v = Z_0$ which implies $\inf_{u \in [s,t]} h(u) = 0$.

- If $Z_s \leq Z_t \leq Z_0$, then $h(t) = Z_0 - Z_t$ and $h(s) = Z_0 - Z_s$. Thus,

$$|Z_t - Z_s| = h(s) - h(t) \leq h(s) - \inf_{u \in [s,t]} h(u) \leq h(t) + h(s) - 2 \inf_{u \in [s,t]} h(u). \tag{B.11}$$

Hence, h is a height function of Z and Z is tree-like. $\square$

Appendix C. Signature and stochastic processes

In the last two subsections, the signature has been presented in a deterministic setting. However, it is clear that the stated results in the previous subsection remain true for stochastic processes by defining the signature as a random variable. In view of the uniqueness theorem from Hambly and Lyons, a natural question at this stage is whether the signature allows to characterize the law of stochastic processes. A first positive answer has been provided by Chevyrev and Lyons (2016). They succeeded to construct a characteristic function for the signature of stochastic processes and they proved that it characterizes the law of stochastic processes in the same way as the traditional characteristic function does for random variables. However, this construction is quite abstract and as such is not suitable for applications so far. They also gave some technical conditions under which the expected signature (defined as $\mathbb{E}[S(X)]$ where X is a stochastic process) characterizes the law.

These results have then been extended by Chevyrev and Oberhauser (2022). They showed that by considering a normalization of the signature, the expected normalized signature characterizes the law of stochastic processes under mild regularity assumptions. This result is stronger than the one from Chevyrev and Lyons as it requires less assumptions. We now provide a brief description of their main result.

Let us denote by $T_1^*(E)$ the subset of $T^*(E)$ (see Equation (2.10) in Section 2.2.) defined by:

$$T_1^*(E) := \left\{\mathbf{t} \in T^*(E) \mid \mathbf{t}^0 = 1\right\}. \tag{C.1}$$

We define a tensor normalization as follows:

Definition C.1 (Tensor normalization)**.** *A tensor normalization is a continuous injective map of the form*

$$\begin{array}{rcl} \Lambda : T_1^*(E) & \to & \{\mathbf{t} \in T_1^*(E) \mid \|\mathbf{t}\| \leq K\} \\ \mathbf{t} & \mapsto & (\mathbf{t}^0, \lambda(\mathbf{t})\mathbf{t}^1, \lambda(\mathbf{t})^2\mathbf{t}^2, \dots, \lambda(\mathbf{t})^n\mathbf{t}^n, \dots). \end{array} \tag{C.2}$$

where $K > 0$ is a constant and $\lambda : T_1^*(E) \to (0, +\infty)$ is a positive function.

The existence of such object is discussed in Proposition 14 of Chevyrev and Oberhauser (2022). We can now state a simplified version of Chevyrev and Oberhauser's main theorem:

Theorem C.1. *Let $X = (X_t)_{t\in[0,T]}$ and $Y = (Y_t)_{t\in[0,T]}$ be two stochastic processes defined on a probability space $(\Omega, \mathcal{A}, \mathbb{P})$ such that X and Y are in $\mathcal{P}^1([0,T], E)$ almost surely where $\mathcal{P}^1([0,T], E) = \mathcal{C}^1([0,T], E)/\sim_t$ is the space of bounded variation paths quotiented by the tree-like equivalence relation. Let Λ be a tensor normalization and define the normalized signature as $\Phi = \Lambda \circ S$. Then,*

$$\mathbb{E}[\Phi(X)] = \mathbb{E}[\Phi(Y)] \text{ iff } X \overset{d}{=} Y. \tag{C.3}$$

Remark C.1. *This theorem can be extended to a more general space of processes, namely the space of geometric p-rough paths quotiented by the tree-like equivalence. This extension corresponds to Theorem 26 in Chevyrev and Oberhauser (2022).*

Remark C.2. *The proof of this theorem does not work anymore if we replace the signature by the log-signature. Indeed, one of the key ingredients of the proof is the shuffle product identity (stated and proved in Theorem 2.15 of Lyons et al., 2007) which holds for the signature but not for the log-signature.*

References

Boedihardjo, H., Geng, X., Lyons, T., and Yang, D. (2016). The signature of a rough path: uniqueness. In: *Adv. Math.* 293, pp. 720-737.

Chevyrev, I. and Lyons, T. (2016). Characteristic functions of measures on geometric rough paths. In: *Ann. Probab.* 44.6, pp. 4049-4082.

Chevyrev, I. and Oberhauser, H. (2022). Signature moments to characterize laws of stochastic processes. In: *Journal of Machine Learning Research* 23.176, pp. 1-42.

Friz, P. K. and Victoir, N. B. (2010). *Multidimensional stochastic processes as rough paths.* Vol. 120. Cambridge Studies in Advanced Mathematics. Theory and applications. Cambridge University Press, Cambridge, pp. xiv+656.

Hambly, B. and Lyons, T. (2010). Uniqueness for the signature of a path of bounded variation and the reduced path group. In: *Ann. of Math. (2)* 171.1, pp. 109-167.

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

Lyons, T., Caruana, M., and Lévy, T. (2007). *Differential equations driven by rough paths.* Vol. 1908. Lecture Notes in Mathematics. Lectures from the 34th Summer School on Probability Theory held in Saint-Flour, July 6-24, 2004, With an introduction concerning the Summer School by Jean Picard. Springer, Berlin, pp. xviii+109.