Weierstrass Institute for
Applied Analysis and Stochastics

Leibniz
Association

Signatures and applications in finance

Christian Bayer

Memory can determine the dynamics of a stochastic process in different ways, e.g.,

Hidden Markov process: X is a component or function of an underlying Markov process Z. E.g., the price process in a stochastic volatility model

$$dS_t = \sqrt{v_t} S_t dB_t, \quad dv_t = \alpha(v_t)dt + \beta(v_t)dW_t, \quad Z = (S, v).$$

Delay equations: The dynamics of X at time t depends explicitly on $(X_s)_{t-h \le s \le t}$.

Memory kernel: The dynamics of X at time t depends on

$$\int_{-\infty}^{t} K(t,s) X_s ds, \quad \int_{-\infty}^{t} K(t,s) dX_s, \dots$$

Special case: $K(t,s) = K(t-s)$ (Volterra equation).

Processes with memory are the rule, not the exception!

Claim

The path signature is a universal tool for approximating functions of paths, comparable to polynomials in finite dimensions.

1. Introduction to signatures and rough paths (time permitting).
2. Universality of signatures and signature kernels: model-free statistics for stochastic processes.
3. Optimal stopping as an example of using signatures for stochastic optimal control of non-Markov processes.

Paths

- A (d-dimensional) path is a continuous function $x : I \to \mathbb{R}^d$, $I \subset \mathbb{R}$ being an interval.

Paths

- A (d-dimensional) path is a continuous function $x : I \to \mathbb{R}^d$, $I \subset \mathbb{R}$ being an interval.
- A path x is smooth if it is C^1 – more precisely, bounded variation would suffice.

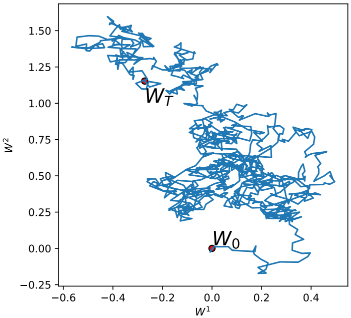


Figure: Sample of a $2d$ Brownian motion W.

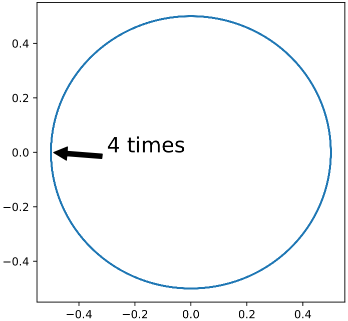


Figure: Path $[0, 1] \ni t \mapsto \frac{1}{4}(\sin(8\pi t), \cos(8\pi t))$.

Controlled differential equation

Let $x : [0, T] \to \mathbb{R}^d$ be a smooth path, $V : \mathbb{R}^e \to \mathbb{R}^{e\times d}$ smooth, $y_0 \in \mathbb{R}^e$, and consider

$$dy(t) = V(y(t))\,dx(t), \quad t \in [0, T], \quad y(0) = y_0.$$

Controlled differential equation

Let $x : [0, T] \to \mathbb{R}^d$ be a smooth path, $V : \mathbb{R}^e \to \mathbb{R}^{e \times d}$ smooth, $y_0 \in \mathbb{R}^e$, and consider

$$\mathrm{d}y(t) = V(y(t))\,\mathrm{d}x(t), \quad t \in [0, T], \quad y(0) = y_0.$$

- y solves an ODE: $\dot{y}(t) = V(y(t))\dot{x}(t)$, but difficult to generalize to rough paths.

Controlled differential equation

Let $x : [0, T] \to \mathbb{R}^d$ be a smooth path, $V : \mathbb{R}^e \to \mathbb{R}^{e\times d}$ smooth, $y_0 \in \mathbb{R}^e$, and consider

$$\mathrm{d}y(t) = V(y(t))\,\mathrm{d}x(t), \quad t \in [0, T], \quad y(0) = y_0.$$

- y solves an ODE: $\dot{y}(t) = V(y(t))\dot{x}(t)$, but difficult to generalize to rough paths.
- First order expansion: For $s < u < t$, $y(u) = y(s) + \text{H.O.T.}$, implying that

  $V(y(u)) = V(y(s)) + \text{H.O.T.}$, and hence $y(t) = y(s) + V(y(s))x_{s,t} + \text{H.O.T.}$, $x_{s,t} := x(t) - x(s)$.

Controlled differential equation

Let $x : [0, T] \to \mathbb{R}^d$ be a smooth path, $V : \mathbb{R}^e \to \mathbb{R}^{e\times d}$ smooth, $y_0 \in \mathbb{R}^e$, and consider

$$\mathrm{d}y(t) = V(y(t))\,\mathrm{d}x(t), \quad t \in [0, T], \quad y(0) = y_0.$$

- y solves an ODE: $\dot{y}(t) = V(y(t))\dot{x}(t)$, but difficult to generalize to rough paths.
- First order expansion: For $s < u < t$, $y(u) = y(s) +$ H.O.T., implying that

$$V(y(u)) = V(y(s)) + \text{H.O.T., and hence } y(t) = y(s) + V(y(s))x_{s,t} + \text{H.O.T.}, \quad x_{s,t} := x(t) - x(s).$$

- Second order expansion: $y(u) = y(s) + V(y(s))x_{s,u} +$ H.O.T., implying that

$$V(y(u)) = V(y(s)) + DV(y(s))V(y(s))x_{s,u}, \quad y(t) = y(s) + V(y(s))x_{s,t} + DV(y(s))V(y(s))\mathbb{x}_{s,t} + \text{H.O.T.}$$

$$\mathbb{x}^{(i,j)}_{s,t} := \int_s^t x^i_{s,u}\mathrm{d}x^j(u) = \int_{s<t_1<t_2<t} \mathrm{d}x^i(t_1)\,\mathrm{d}x^j(t_2), \quad i, j = 1, \ldots, d.$$

Controlled differential equation

Let $x : [0, T] \to \mathbb{R}^d$ be a smooth path, $V : \mathbb{R}^e \to \mathbb{R}^{e\times d}$ smooth, $y_0 \in \mathbb{R}^e$, and consider

$$dy(t) = V(y(t))\, dx(t), \quad t \in [0, T], \quad y(0) = y_0.$$

- y solves an ODE: $\dot{y}(t) = V(y(t))\dot{x}(t)$, but difficult to generalize to rough paths.
- First order expansion: For $s < u < t$, $y(u) = y(s) +$ H.O.T., implying that

$V(y(u)) = V(y(s))+$H.O.T., and hence $y(t) = y(s) + V(y(s))x_{s,t} +$ H.O.T., $x_{s,t} := x(t) - x(s)$.

- Second order expansion: $y(u) = y(s) + V(y(s))x_{s,u} +$ H.O.T., implying that

$V(y(u)) = V(y(s)) + DV(y(s))V(y(s))x_{s,u}$, $y(t) = y(s) + V(y(s))x_{s,t} + DV(y(s))V(y(s))\mathbb{x}_{s,t} +$ H.O.T.

$$\mathbb{x}^{(i,j)}_{s,t} := \int_s^t x^i_{s,u} dx^j(u) = \int_{s<t_1<t_2<t} dx^i(t_1)\, dx^j(t_2), \quad i, j = 1, \dots, d.$$

- Third order expansion: involves iterated integrals of order three...

Path signature

Given a (smooth) path $x : [s,t] \to \mathbb{R}^d$, the associated signature $\mathbb{x}^{<\infty}_{s,t}$ is the collection of all iterated integrals, i.e., $\mathbb{x}^{<\infty}_{s,t} := \left(\mathbb{x}^{=n}_{s,t}\right)_{n=0}^{\infty}$, where

$$\mathbb{x}^{=0}_{s,t} := 1,\ \mathbb{x}^{=n}_{s,t} := \left(\mathbb{x}^{(i_1,\dots,i_n)}_{s,t}\right)_{(i_1,\dots,i_n)\in\{1,\dots,d\}^n},\ \mathbb{x}^{(i_1,\dots,i_n)}_{s,t} := \int_{s<t_1<\cdots<t_n<t} \mathrm{d}x^{i_1}(t_1)\cdots \mathrm{d}x^{i_n}(t_n).$$

Path signature

Given a (smooth) path $x : [s,t] \to \mathbb{R}^d$, the associated signature $\mathbb{x}^{<\infty}_{s,t}$ is the collection of all iterated integrals, i.e., $\mathbb{x}^{<\infty}_{s,t} := \left(\mathbb{x}^{=n}_{s,t}\right)_{n=0}^{\infty}$, where

$$\mathbb{x}^{=0}_{s,t} := 1, \; \mathbb{x}^{=n}_{s,t} := \left(\mathbb{x}^{(i_1,\dots,i_n)}_{s,t}\right)_{(i_1,\dots,i_n)\in\{1,\dots,d\}^n}, \; \mathbb{x}^{(i_1,\dots,i_n)}_{s,t} := \int_{s<t_1<\dots<t_n<t} \mathrm{d}x^{i_1}(t_1)\cdots \mathrm{d}x^{i_n}(t_n).$$

The signature is parameterization-invariant: i.e., for $\gamma : [u,v] \to [s,t]$ increasing and C^1, the change of variables formula – with $r = \gamma(\bar{r})$ – implies that

$$\int_u^v f(\gamma(\bar{r}))\mathrm{d}x(\gamma(\bar{r})) = \int_u^v f(\gamma(\bar{r}))\dot{x}(\gamma(\bar{r}))\dot{\gamma}(\bar{r})\mathrm{d}\bar{r} = \int_s^t f(r)\dot{x}(r)\mathrm{d}r = \int_s^t f(r)\mathrm{d}x(r).$$

Hence, denoting $z \circ \gamma = x$, we have $\mathbb{z}^{<\infty}_{u,v} = \mathbb{x}^{<\infty}_{s,t}$.

Theorem (Chen 1958, Hambly and Lyons 2010)

A (smooth) path x is uniquely determined by its initial value and its signature – up to re-parameterization and tree-like excursions.

Theorem (Chen 1958, Hambly and Lyons 2010)

A (smooth) path x is uniquely determined by its initial value and its signature – up to re-parameterization and tree-like excursions.

- The theorem was proved by Chen for C^1-paths in 1958 and extended to bounded-variation paths by Hambly and Lyons in 2010.
- Extended to (weakly geometric) rough paths.
- Tree-like paths are essentially paths, which start and end in the same point and "completely re-trace their history". These paths have trivial signatures.

Theorem (Chen 1958, Hambly and Lyons 2010)

A (smooth) path x is uniquely determined by its initial value and its signature – up to re-parameterization and tree-like excursions.

- The theorem was proved by Chen for C^1-paths in 1958 and extended to bounded-variation paths by Hambly and Lyons in 2010.
- Extended to (weakly geometric) rough paths.
- Tree-like paths are essentially paths, which start and end in the same point and "completely re-trace their history". These paths have trivial signatures.

Open problem

How can we computationally and efficiently recover the path (with unit speed) from its signature?

Tensor algebra

Given a (finite-dimensional) vector space V, let $V^{\otimes 0} := \mathbb{R}$, $V^{\otimes(n+1)} := V^{\otimes n} \otimes V$, and denote

$$T(V) := \bigoplus_{n=0}^{\infty} V^{\otimes n}, \quad T((V)) := \prod_{n=0}^{\infty} V^{\otimes n}, \quad T^N(V) := \bigoplus_{n=0}^{N} V^{\otimes n}$$

Both $T(V)$ and $T((V))$ (and, with obvious modifications, the truncated tensor algebra $T^N(V)$) are algebras with usual addition and the product

$$\mathbf{a} \otimes \mathbf{b} := \left(\sum_{i+j=n} a_i \otimes b_j \right)_{n=0}^{\infty}, \text{ where } \mathbf{a} = (a_n)_{n=0}^{\infty},\ \mathbf{b} = (b_n)_{n=0}^{\infty}.$$

Recall that $\mathbf{a} = (a_n)_{n=0}^{\infty} \in T((V))$ is contained in $T(V)$ iff $a_n = 0 \in V^{\otimes n}$ for all but finitely many n.

- Let $e_1, \ldots, e_d$ denote a basis of $\mathbb{R}^d$, and $x : [s,t] \to \mathbb{R}^d$ a smooth path with $x(u) = \sum_{i=1}^d x^i(u) e_i =: x^i(u) e_i$.
- Recall that $\{ e_{i_1} \otimes \cdots \otimes e_{i_n} \mid (i_1, \ldots, i_n) \in \{1, \ldots, d\}^n \}$ is a basis of $(\mathbb{R}^d)^{\otimes n}$.
- We denote the basis of $(\mathbb{R}^d)^{\otimes 0} \simeq \mathbb{R}$ by $\mathbf{1}$ – which we identify with $(\mathbf{1}, 0, \ldots) \in T((\mathbb{R}^d))$. Note that $\mathbf{1}$ is the neutral element of the algebra $T((\mathbb{R}^d))$ w.r.t. $\otimes$.

- Let $e_1, \dots, e_d$ denote a basis of $\mathbb{R}^d$, and $x : [s,t] \to \mathbb{R}^d$ a smooth path with $x(u) = \sum_{i=1}^d x^i(u) e_i =: x^i(u) e_i$.
- Recall that $\{ e_{i_1} \otimes \cdots \otimes e_{i_n} \mid (i_1, \dots, i_n) \in \{1, \dots, d\}^n \}$ is a basis of $(\mathbb{R}^d)^{\otimes n}$.
- We denote the basis of $(\mathbb{R}^d)^{\otimes 0} \simeq \mathbb{R}$ by $\mathbf{1}$ – which we identify with $(\mathbf{1}, 0, \dots) \in T((\mathbb{R}^d))$. Note that $\mathbf{1}$ is the neutral element of the algebra $T((\mathbb{R}^d))$ w.r.t. $\otimes$.

Definition (Path signature)

We define the signature $\mathbb{x}_{s,t}^{<\infty} \in T((\mathbb{R}^d))$ by setting

$$\mathbb{x}_{s,t}^{<\infty} := \mathbf{1} + \sum_{n=1}^{\infty} \sum_{(i_1,\dots,i_n)\in\{1,\dots,d\}^n} \mathbb{x}_{s,t}^{(i_1,\dots,i_n)} e_{i_1} \otimes \cdots \otimes e_{i_n} =: \mathbf{1} + \sum_{n=1}^{\infty} \int_{s<t_1<\cdots<t_n<t} \mathrm{d}x(t_1) \otimes \cdots \otimes \mathrm{d}x(t_n),$$

as well as its truncated version $\mathbb{x}_{s,t}^{\leq N} \in T^N(\mathbb{R}^d)$ by truncation at level N.

Theorem (Chen's identity)

Given a (smooth) path $x : [r,t] \to \mathbb{R}^d$, *then for any* $r < s < t$ *we have*

$$\mathbb{X}^{<\infty}_{r,t} = \mathbb{X}^{<\infty}_{r,s} \otimes \mathbb{X}^{<\infty}_{s,t}.$$

Theorem (Chen's identity)

Given a (smooth) path $x : [r,t] \to \mathbb{R}^d$, then for any $r < s < t$ we have

$$\mathbb{x}^{<\infty}_{r,t} = \mathbb{x}^{<\infty}_{r,s} \otimes \mathbb{x}^{<\infty}_{s,t}.$$

- Formally, Chen's identity follows easily from the differential equation satisfied by the signature:

$$\mathrm{d}\mathbb{x}^{<\infty}_{s,t} = \mathbb{x}^{<\infty}_{s,t} \otimes \mathrm{d}x(t), \quad \mathbb{x}^{<\infty}_{s,s} = \mathbf{1} \in T((\mathbb{R}^d)).$$

- Chen's identity is a consequence of linearity of the integral. Hence, it is a fundamental property valid for all notions of signatures, including for rough paths.

Theorem (Chen's identity)

Given a (smooth) path $x : [r,t] \to \mathbb{R}^d$, then for any $r < s < t$ we have

$$\mathbb{x}^{<\infty}_{r,t} = \mathbb{x}^{<\infty}_{r,s} \otimes \mathbb{x}^{<\infty}_{s,t}.$$

- Given two paths $x : [a,b] \to \mathbb{R}^d$ and $y : [c,e] \to \mathbb{R}^d$, define their concatenation product $z := x \circ y : [a, b + (e-c)] \to \mathbb{R}^d$ by

$$z(u) := \begin{cases} x(u), & a \le u \le b, \\ y(u - b + c) - y(c) + x(b), & b < u \le b + (e-c). \end{cases}$$

  By Chen's identity (and re-parameterization invariance), $\mathbb{z}^{<\infty}_{a,b+(e-c)} = \mathbb{x}^{<\infty}_{a,b} \otimes \mathbb{y}^{<\infty}_{c,e}$.
- Let $\overleftarrow{x}$ the time-reversal of x, so $z := x \circ \overleftarrow{x}$ is tree-like, $\mathbb{z}^{<\infty}_{r,t} = \mathbf{1}$. Hence, $\overleftarrow{\mathbb{x}}^{<\infty}_{s,t} = (\mathbb{x}^{<\infty}_{r,s})^{-1}$.

- Consider all words $\mathtt{w}$ in the letters $\{\mathtt{1}, \dots, \mathtt{d}\}$, endowed with the concatenation product.

- Consider all words $\mathtt{w}$ in the letters $\{\mathtt{1},\dots,\mathtt{d}\}$, endowed with the concatenation product.
- Let $\mathcal{W}_d$ denote the linear span of all such words: For words $\mathtt{w}_1,\dots,\mathtt{w}_k$, a typical element $\ell \in \mathcal{W}_d$ is of the form $\ell = \lambda_1 \mathtt{w}_1 + \cdots + \lambda_k \mathtt{w}_k, \quad \lambda_1,\dots,\lambda_k \in \mathbb{R}$.
- Extending the concatenation product in a distributive way to $\mathcal{W}_d$, we obtain an algebra, including the empty word $\varnothing$ as neutral element w.r.t. multiplication (i.e., concatenation).

- Consider all words $\mathtt{w}$ in the letters $\{\mathtt{1}, \dots, \mathtt{d}\}$, endowed with the concatenation product.
- Let $\mathcal{W}_d$ denote the linear span of all such words: For words $\mathtt{w}_1, \dots, \mathtt{w}_k$, a typical element $\ell \in \mathcal{W}_d$ is of the form $\ell = \lambda_1 \mathtt{w}_1 + \cdots + \lambda_k \mathtt{w}_k, \quad \lambda_1, \dots, \lambda_k \in \mathbb{R}$.
- Extending the concatenation product in a distributive way to $\mathcal{W}_d$, we obtain an algebra, including the empty word $\varnothing$ as neutral element w.r.t. multiplication (i.e., concatenation).
- Note that $\mathcal{W}_d$ is isomorphic to the algebra $T(\mathbb{R}^d)$, and, hence, (trivially) $T((\mathbb{R}^d)^*)$.

- Consider all words $\mathtt{w}$ in the letters $\{\,\mathtt{1},\dots,\mathtt{d}\,\}$, endowed with the concatenation product.
- Let $\mathcal{W}_d$ denote the linear span of all such words: For words $\mathtt{w}_1,\dots,\mathtt{w}_k$, a typical element $\ell \in \mathcal{W}_d$ is of the form $\ell = \lambda_1 \mathtt{w}_1 + \cdots + \lambda_k \mathtt{w}_k, \quad \lambda_1,\dots,\lambda_k \in \mathbb{R}$.
- Extending the concatenation product in a distributive way to $\mathcal{W}_d$, we obtain an algebra, including the empty word $\varnothing$ as neutral element w.r.t. multiplication (i.e., concatenation).
- Note that $\mathcal{W}_d$ is isomorphic to the algebra $T(\mathbb{R}^d)$, and, hence, (trivially) $T((\mathbb{R}^d)^*)$.

Definition (Duality pairing)

Define a bi-linear map $\langle \cdot, \cdot \rangle : \mathcal{W}_d \times T((\mathbb{R}^d)) \to \mathbb{R}$: For a word $\ell = \mathtt{i}_1 \cdots \mathtt{i}_k \in \mathcal{W}_d$, and for

$$T((\mathbb{R}^d)) \ni \mathbf{a} = a^{\varnothing}\mathbf{1} + \sum_{n=1}^{\infty} \sum_{(i_1,\dots,i_n)\in\{\,1,\dots,d\,\}^n} a^{(i_1,\dots,i_n)} e_{i_1} \otimes \cdots \otimes e_{i_n},$$

set $\langle \mathtt{i}_1 \cdots \mathtt{i}_k, \mathbf{a} \rangle := a^{(i_1,\dots,i_k)}$, and extend bi-linearly to $\mathcal{W}_d$ in the first argument.

Definition (Shuffle product)

Define a commutative product $\sqcup\!\sqcup$ on $\mathcal{W}_d$ as follows: For words $\mathtt{w}$, $\mathtt{v}$ and letters $\mathtt{i}$, $\mathtt{j}$ define

$$\mathtt{w} \sqcup\!\sqcup \varnothing := \varnothing \sqcup\!\sqcup \mathtt{w} := \mathtt{w}, \quad \mathtt{wi} \sqcup\!\sqcup \mathtt{vj} := (\mathtt{w} \sqcup\!\sqcup \mathtt{vj})\mathtt{i} + (\mathtt{wi} \sqcup\!\sqcup \mathtt{v})\mathtt{j},$$

and extend to $\mathcal{W}_d$ by bi-linearity.

Definition (Shuffle product)

Define a commutative product $\sqcup\!\sqcup$ on $\mathcal{W}_d$ as follows: For words $\mathtt{w}, \mathtt{v}$ and letters $\mathtt{i}, \mathtt{j}$ define

$$\mathtt{w} \sqcup\!\sqcup \varnothing := \varnothing \sqcup\!\sqcup \mathtt{w} := \mathtt{w}, \quad \mathtt{wi} \sqcup\!\sqcup \mathtt{vj} := (\mathtt{w} \sqcup\!\sqcup \mathtt{vj})\mathtt{i} + (\mathtt{wi} \sqcup\!\sqcup \mathtt{v})\mathtt{j},$$

and extend to $\mathcal{W}_d$ by bi-linearity.

Example: $\mathtt{12} \sqcup\!\sqcup \mathtt{34} = \mathtt{1234} + \mathtt{1324} + \mathtt{1342} + \mathtt{3124} + \mathtt{3142} + \mathtt{3412}$.

Definition (Shuffle product)

Define a commutative product $\sqcup\!\sqcup$ on $\mathcal{W}_d$ as follows: For words $\mathtt{w}, \mathtt{v}$ and letters $\mathtt{i}, \mathtt{j}$ define

$$\mathtt{w} \sqcup\!\sqcup \varnothing := \varnothing \sqcup\!\sqcup \mathtt{w} := \mathtt{w}, \quad \mathtt{wi} \sqcup\!\sqcup \mathtt{vj} := (\mathtt{w} \sqcup\!\sqcup \mathtt{vj})\mathtt{i} + (\mathtt{wi} \sqcup\!\sqcup \mathtt{v})\mathtt{j},$$

and extend to $\mathcal{W}_d$ by bi-linearity.

Example: $\mathtt{12} \sqcup\!\sqcup \mathtt{34} = \mathtt{1234} + \mathtt{1324} + \mathtt{1342} + \mathtt{3124} + \mathtt{3142} + \mathtt{3412}$.

Theorem (Shuffle identity)

Given a smooth path $x : [s,t] \to \mathbb{R}^d$ and $\ell_1, \ell_2 \in \mathcal{W}_d$, we have

$$\left\langle \ell_1, \mathbb{x}^{<\infty}_{s,t} \right\rangle \left\langle \ell_2, \mathbb{x}^{<\infty}_{s,t} \right\rangle = \left\langle \ell_1 \sqcup\!\sqcup \ell_2, \mathbb{x}^{<\infty}_{s,t} \right\rangle.$$

Remarks on the shuffle identity

- Follows from the chain rule, hence relies on smoothness of paths.

Remarks on the shuffle identity

- Follows from the chain rule, hence relies on smoothness of paths.
- Example: Let $\ell_1 = \ell_2 = \mathtt{i}$. Then, by definition, $\mathtt{i} \sqcup\!\sqcup\, \mathtt{i} = 2\mathtt{ii}$. Hence,

$$\left\langle \ell_1 \sqcup\!\sqcup\, \ell_2,\, \mathbb{x}^{<\infty}_{s,t} \right\rangle = 2\left\langle \mathtt{ii},\, \mathbb{x}^{<\infty}_{s,t} \right\rangle = 2\int_s^t (x^i(u) - x^i(s))\mathrm{d}x^i(u) = 2\int_s^t \underbrace{x^i(u)\dot{x}^i(u)}_{=\frac{1}{2}\frac{\mathrm{d}}{\mathrm{d}u}(x^i(u))^2}\mathrm{d}u - 2x^i(s)x^i_{s,t}$$

$$= (x^i(t))^2 - (x^i(s))^2 - 2x^i(s)x^i(t) + 2(x^i(s))^2 = (x^i_{s,t})^2 = \left\langle \mathtt{i},\, \mathbb{x}^{<\infty}_{s,t} \right\rangle^2 = \left\langle \ell_1,\, \mathbb{x}^{<\infty}_{s,t} \right\rangle \left\langle \ell_2,\, \mathbb{x}^{<\infty}_{s,t} \right\rangle.$$

Note the redundancies in the signature!

- Follows from the chain rule, hence relies on smoothness of paths.
- Example: Let $\ell_1 = \ell_2 = \mathtt{i}$. Then, by definition, $\mathtt{i} \sqcup\!\sqcup \mathtt{i} = 2\mathtt{ii}$. Hence,

$$\left\langle \ell_1 \sqcup\!\sqcup \ell_2, \mathbb{X}^{<\infty}_{s,t} \right\rangle = 2 \left\langle \mathtt{ii}, \mathbb{X}^{<\infty}_{s,t} \right\rangle = 2 \int_s^t (x^i(u) - x^i(s)) \mathrm{d}x^i(u) = 2 \int_s^t \underbrace{x^i(u)\dot{x}^i(u)}_{=\frac{1}{2}\frac{\mathrm{d}}{\mathrm{d}u}(x^i(u))^2} \mathrm{d}u - 2x^i(s)x^i_{s,t}$$

$$= (x^i(t))^2 - (x^i(s))^2 - 2x^i(s)x^i(t) + 2(x^i(s))^2 = (x^i_{s,t})^2 = \left\langle \mathtt{i}, \mathbb{X}^{<\infty}_{s,t} \right\rangle^2 = \left\langle \ell_1, \mathbb{X}^{<\infty}_{s,t} \right\rangle \left\langle \ell_2, \mathbb{X}^{<\infty}_{s,t} \right\rangle.$$

  Note the redundancies in the signature!
- Given $p \in \mathbb{R}[x]$ (e.g., $p(x) = \lambda_0 + \lambda_1 x + \cdots + \lambda_n x^n$) and $\ell \in \mathcal{W}_d$, there is $p^{\sqcup\!\sqcup}(\ell) \in \mathcal{W}_d$, s.t.,

$$p\left(\left\langle \ell, \mathbb{X}^{<\infty}_{s,t} \right\rangle\right) = \left\langle p^{\sqcup\!\sqcup}(\ell), \mathbb{X}^{<\infty}_{s,t} \right\rangle, \quad p^{\sqcup\!\sqcup}(\ell) \coloneqq \lambda_0 \varnothing + \lambda_1 \ell + \cdots + \lambda_n \ell^{\sqcup\!\sqcup n} \in \mathcal{W}_d.$$

Polynomials in the signature are linear functionals in the signature.

Recall that signatures are invertible w.r.t. the tensor multiplication. Do they form a group?

Definition (Group-like elements)

$$G(\mathbb{R}^d) := \left\{ \mathbf{a} \in T((\mathbb{R}^d)) \,\middle|\, \forall \ell_1, \ell_2 \in \mathcal{W}_d : \langle \ell_1, \mathbf{a} \rangle \langle \ell_2, \mathbf{a} \rangle = \langle \ell_1 \sqcup\!\sqcup \ell_2, \mathbf{a} \rangle \right\}$$

Recall that signatures are invertible w.r.t. the tensor multiplication. Do they form a group?

Definition (Group-like elements)

$$G(\mathbb{R}^d) := \left\{ \mathbf{a} \in T((\mathbb{R}^d)) \,\middle|\, \forall \ell_1, \ell_2 \in \mathcal{W}_d : \langle \ell_1, \mathbf{a} \rangle \langle \ell_2, \mathbf{a} \rangle = \langle \ell_1 \sqcup\!\sqcup \ell_2, \mathbf{a} \rangle \right\}$$

- From the shuffle-identity, for any smooth path $x : [s,t] \to \mathbb{R}^d$, $\mathbb{x}^{<\infty}_{s,t} \in G(\mathbb{R}^d)$.
- If $\mathbf{a} \in G(\mathbb{R}^d)$, then $\mathbf{a} = \mathbf{1} + \tilde{\mathbf{a}}$ (with $\langle \varnothing, \tilde{\mathbf{a}} \rangle = 0$), and $\mathbf{a}^{-1} = \sum_{k=0}^{\infty} (-1)^k \tilde{\mathbf{a}}^{\otimes k}$.

Recall that signatures are invertible w.r.t. the tensor multiplication. Do they form a group?

Definition (Group-like elements)

$$G(\mathbb{R}^d) := \left\{ \mathbf{a} \in T((\mathbb{R}^d)) \,\middle|\, \forall \ell_1, \ell_2 \in \mathcal{W}_d : \langle \ell_1, \mathbf{a} \rangle \langle \ell_2, \mathbf{a} \rangle = \langle \ell_1 \sqcup\!\sqcup \ell_2, \mathbf{a} \rangle \right\}$$

- From the shuffle-identity, for any smooth path $x : [s,t] \to \mathbb{R}^d$, $\mathbb{x}_{s,t}^{<\infty} \in G(\mathbb{R}^d)$.
- If $\mathbf{a} \in G(\mathbb{R}^d)$, then $\mathbf{a} = \mathbf{1} + \tilde{\mathbf{a}}$ (with $\langle \emptyset, \tilde{\mathbf{a}} \rangle = 0$), and $\mathbf{a}^{-1} = \sum_{k=0}^{\infty} (-1)^k \tilde{\mathbf{a}}^{\otimes k}$.
- We can also define a group $G^N(\mathbb{R}^d) \subset T^N(\mathbb{R}^d)$ by truncation. $G^N(\mathbb{R}^d)$ is a Lie group.

Define $\exp : T((\mathbb{R}^d)) \to T((\mathbb{R}^d))$ and $\log : \left\{ \mathbf{a} \in T((\mathbb{R}^d)) \,\middle|\, \langle \varnothing, \mathbf{a} \rangle = 1 \right\} \to T((\mathbb{R}^d))$ by

$$\exp(\mathbf{a}) := \mathbf{1} + \sum_{k=1}^{\infty} \frac{1}{k!} \mathbf{a}^{\otimes k}, \quad \log(\mathbf{a}) := \sum_{k=1}^{\infty} \frac{(-1)^{k+1}}{k} \tilde{\mathbf{a}}^{\otimes k}, \qquad \text{with } \mathbf{a} = \mathbf{1} + \tilde{\mathbf{a}}.$$

Define $\exp : T((\mathbb{R}^d)) \to T((\mathbb{R}^d))$ and $\log : \left\{ \mathbf{a} \in T((\mathbb{R}^d)) \,\middle|\, \langle \varnothing, \mathbf{a} \rangle = 1 \right\} \to T((\mathbb{R}^d))$ by

$$\exp(\mathbf{a}) := \mathbf{1} + \sum_{k=1}^{\infty} \frac{1}{k!} \mathbf{a}^{\otimes k}, \quad \log(\mathbf{a}) := \sum_{k=1}^{\infty} \frac{(-1)^{k+1}}{k} \tilde{\mathbf{a}}^{\otimes k}, \quad \text{with } \mathbf{a} = \mathbf{1} + \tilde{\mathbf{a}}.$$

Lie algebra

$\mathfrak{g}(\mathbb{R}^d) := \log(G(\mathbb{R}^d))$ is a Lie algebra under the commutator $[\mathbf{a}, \mathbf{b}] := \mathbf{a} \otimes \mathbf{b} - \mathbf{b} \otimes \mathbf{a}$. In fact, it is the free Lie algebra generated by $e_1, \ldots, e_d$. Similarly, define $\mathfrak{g}^N(\mathbb{R}^d)$.

Define $\exp : T((\mathbb{R}^d)) \to T((\mathbb{R}^d))$ and $\log : \left\{ \mathbf{a} \in T((\mathbb{R}^d)) \,\middle|\, \langle \varnothing, \mathbf{a} \rangle = 1 \right\} \to T((\mathbb{R}^d))$ by

$$\exp(\mathbf{a}) := \mathbf{1} + \sum_{k=1}^{\infty} \frac{1}{k!} \mathbf{a}^{\otimes k}, \quad \log(\mathbf{a}) := \sum_{k=1}^{\infty} \frac{(-1)^{k+1}}{k} \tilde{\mathbf{a}}^{\otimes k}, \quad \text{with } \mathbf{a} = \mathbf{1} + \tilde{\mathbf{a}}.$$

Lie algebra

$\mathfrak{g}(\mathbb{R}^d) := \log(G(\mathbb{R}^d))$ is a Lie algebra under the commutator $[\mathbf{a}, \mathbf{b}] := \mathbf{a} \otimes \mathbf{b} - \mathbf{b} \otimes \mathbf{a}$. In fact, it is the free Lie algebra generated by $e_1, \ldots, e_d$. Similarly, define $\mathfrak{g}^N(\mathbb{R}^d)$.

- Note that $\exp : \mathfrak{g}(\mathbb{R}^d) \to G(\mathbb{R}^d)$ and $\log : G(\mathbb{R}^d) \to \mathfrak{g}(\mathbb{R}^d)$ are both bijective, and the same holds, mutatis mutandis, for the truncated versions $G^N(\mathbb{R}^d)$, $\mathfrak{g}^N(\mathbb{R}^d)$. Hence, $\mathfrak{g}^N(\mathbb{R}^d)$ is a global chart of the Lie group $G^N(\mathbb{R}^d)$.

Define $\exp : T((\mathbb{R}^d)) \to T((\mathbb{R}^d))$ and $\log : \left\{ \mathbf{a} \in T((\mathbb{R}^d)) \,\middle|\, \langle \varnothing, \mathbf{a} \rangle = 1 \right\} \to T((\mathbb{R}^d))$ by

$$\exp(\mathbf{a}) := \mathbf{1} + \sum_{k=1}^{\infty} \frac{1}{k!} \mathbf{a}^{\otimes k}, \quad \log(\mathbf{a}) := \sum_{k=1}^{\infty} \frac{(-1)^{k+1}}{k} \tilde{\mathbf{a}}^{\otimes k}, \quad \text{with } \mathbf{a} = \mathbf{1} + \tilde{\mathbf{a}}.$$

Lie algebra

$\mathfrak{g}(\mathbb{R}^d) := \log(G(\mathbb{R}^d))$ is a Lie algebra under the commutator $[\mathbf{a}, \mathbf{b}] := \mathbf{a} \otimes \mathbf{b} - \mathbf{b} \otimes \mathbf{a}$. In fact, it is the free Lie algebra generated by $e_1, \ldots, e_d$. Similarly, define $\mathfrak{g}^N(\mathbb{R}^d)$.

- Note that $\exp : \mathfrak{g}(\mathbb{R}^d) \to G(\mathbb{R}^d)$ and $\log : G(\mathbb{R}^d) \to \mathfrak{g}(\mathbb{R}^d)$ are both bijective, and the same holds, mutatis mutandis, for the truncated versions $G^N(\mathbb{R}^d)$, $\mathfrak{g}^N(\mathbb{R}^d)$. Hence, $\mathfrak{g}^N(\mathbb{R}^d)$ is a global chart of the Lie group $G^N(\mathbb{R}^d)$.
- $\dim \mathfrak{g}^N(\mathbb{R}^d)$ grows much slower than $\dim T^N(\mathbb{R}^d)$. E.g., for $d = 3$ and $N = 4$: $\dim T^N(\mathbb{R}^d) = 120$, $\dim \mathfrak{g}^N(\mathbb{R}^d) = 32$. Hence, the Lie algebra removes many redundancies (at the cost of the shuffle identity).

Definition (Log-signature)

Given a smooth path $x : [s, t] \to \mathbb{R}^d$, define the (truncated) log-signature by $\mathbb{l}^{<\infty}_{s,t} := \log(\mathbb{x}^{<\infty}_{s,t}) \in \mathfrak{g}(\mathbb{R}^d)$ – and similarly its truncated version $\mathbb{l}^{\leq N}_{s,t} \in \mathfrak{g}^N(\mathbb{R}^d)$.

Definition (Log-signature)

Given a smooth path $x : [s,t] \to \mathbb{R}^d$, define the (truncated) log-signature by $\mathbb{l}^{<\infty}_{s,t} := \log(\mathbb{x}^{<\infty}_{s,t}) \in \mathfrak{g}(\mathbb{R}^d)$ – and similarly its truncated version $\mathbb{l}^{\leq N}_{s,t} \in \mathfrak{g}^N(\mathbb{R}^d)$.

Example: $N = 2$

- A basis of $\mathfrak{g}^2(\mathbb{R}^d)$ is given by e_i, $i = 1, \ldots, d$, together with $[e_i, e_j]$, $1 \leq i < j \leq d$.

Definition (Log-signature)

Given a smooth path $x : [s,t] \to \mathbb{R}^d$, define the (truncated) log-signature by $\mathbb{l}^{<\infty}_{s,t} := \log(\mathbb{x}^{<\infty}_{s,t}) \in \mathfrak{g}(\mathbb{R}^d)$ – and similarly its truncated version $\mathbb{l}^{\leq N}_{s,t} \in \mathfrak{g}^N(\mathbb{R}^d)$.

Example: $N = 2$

- A basis of $\mathfrak{g}^2(\mathbb{R}^d)$ is given by e_i, $i = 1, \dots, d$, together with $[e_i, e_j]$, $1 \leq i < j \leq d$.
- By the definition of $\log$ applied to $\mathbb{x}^{\leq 2}_{s,t} = \mathbf{1} + x^i_{s,t} e_i + \mathbb{x}^{(i,j)}_{s,t} e_i \otimes e_j$, we get $\log \mathbb{x}^{\leq 2}_{s,t} = (\mathbb{x}^{\leq 2}_{s,t} - \mathbf{1}) - \frac{1}{2}(\mathbb{x}^{\leq 1}_{s,t} - \mathbf{1})^{\otimes 2} = x^i_{s,t} e_i + \left(\mathbb{x}^{(i,j)}_{s,t} - \frac{1}{2} x^i_{s,t} x^j_{s,t}\right) e_i \otimes e_j.$
- Note that $\mathbb{x}^{(i,j)}_{s,t} + \mathbb{x}^{(j,i)}_{s,t} = \int_{s<t_1<t_2<t} \mathrm{d}x^i(t_1)\mathrm{d}x^j(t_2) + \int_{s<t_2<t_1<t} \mathrm{d}x^i(t_1)\mathrm{d}x^j(t_2) = \int_s^t \int_s^t \mathrm{d}x^i(t_1)\mathrm{d}x^j(t_2) = x^i_{s,t} x^j_{s,t}$. Hence, $\mathbb{x}^{(i,i)}_{s,t} - \frac{1}{2}(x^i_{s,t})^2 = 0$, $\mathbb{x}^{(i,j)}_{s,t} - \frac{1}{2} x^i_{s,t} x^j_{s,t} = \frac{1}{2}(\mathbb{x}^{(i,j)}_{s,t} - \mathbb{x}^{(j,i)}_{s,t})$.
- In total: $\log \mathbb{x}^{\leq 2}_{s,t} = \sum_{i=1}^d x^i_{s,t} e_i + \sum_{1 \leq i < j \leq d} \frac{1}{2}(\mathbb{x}^{(i,j)}_{s,t} - \mathbb{x}^{(j,i)}_{s,t})[e_i, e_j] =: \sum_{i=1}^d x^i_{s,t} e_i + \sum_{1 \leq i < j \leq d} \mathbb{a}^{(i,j)}_{s,t}$.

Example: Signatures and areas of $x(t) = (\alpha\cosh(\theta_1 t) - \alpha,\ \cos(\theta_2 t))$, $d = 2$

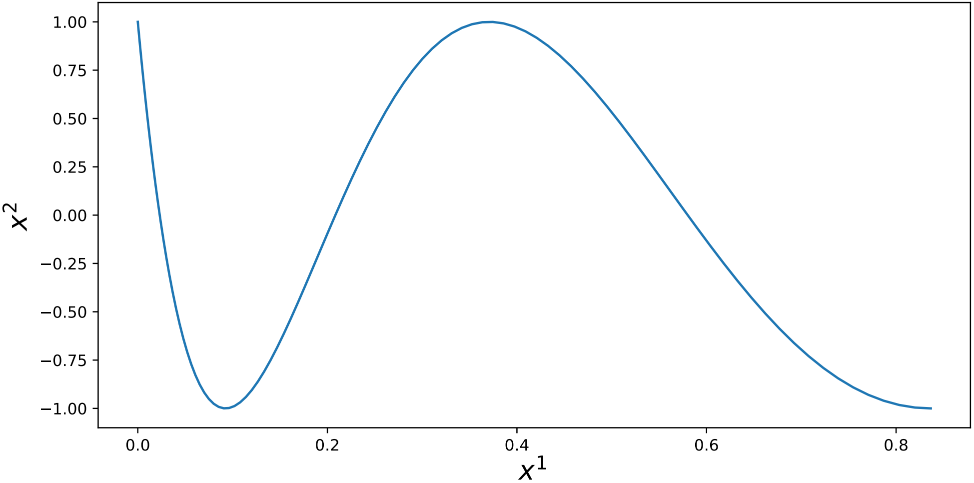


Figure: The path – up to re-parameterization.

Example: Signatures and areas of $x(t) = (\alpha \cosh(\theta_1 t) - \alpha,\ \cos(\theta_2 t))$, $d = 2$

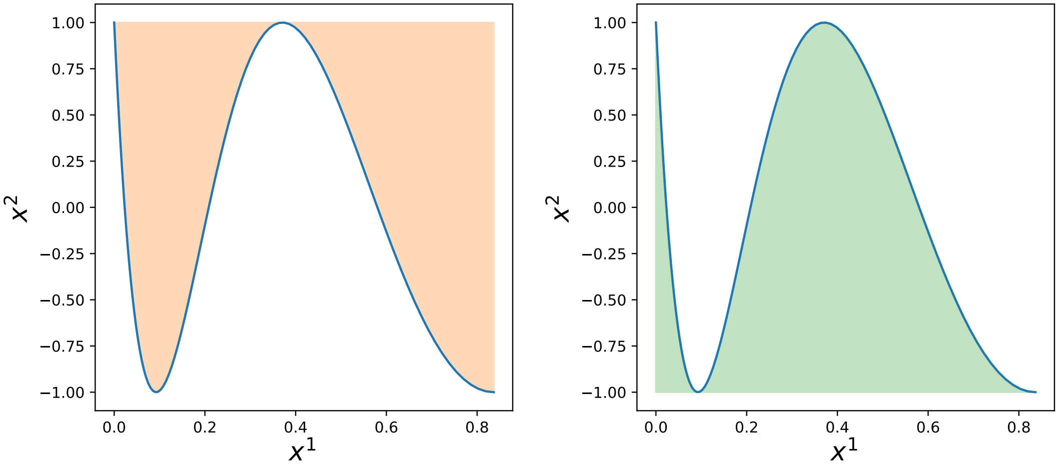


Figure: The shuffle identity $\mathbb{x}_{s,t}^{(1,2)} + \mathbb{x}_{s,t}^{(2,1)} = x_{s,t}^1 x_{s,t}^2$

Example: Signatures and areas of $x(t) = (\alpha \cosh(\theta_1 t) - \alpha,\ \cos(\theta_2 t))$, $d = 2$

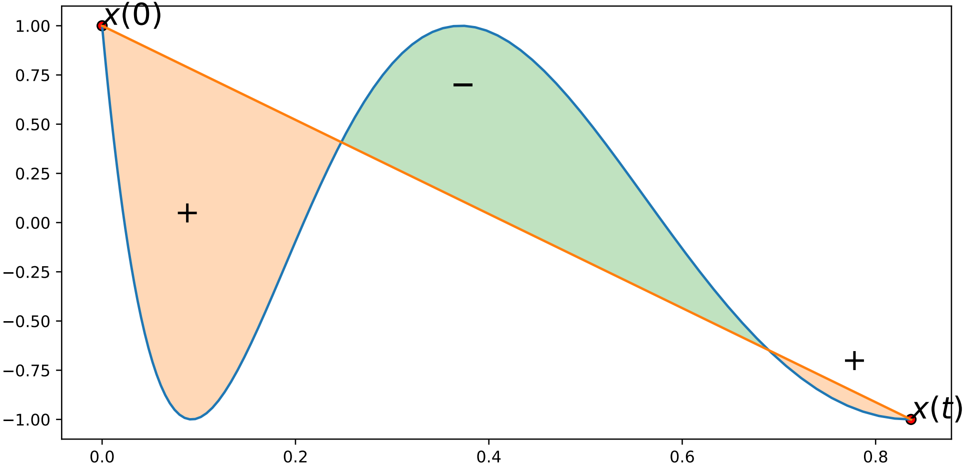


Figure: Interpretation of Lévy's area

Example: Signatures and areas of $x(t) = (\alpha\cosh(\theta_1 t) - \alpha,\ \cos(\theta_2 t))$, $d = 2$

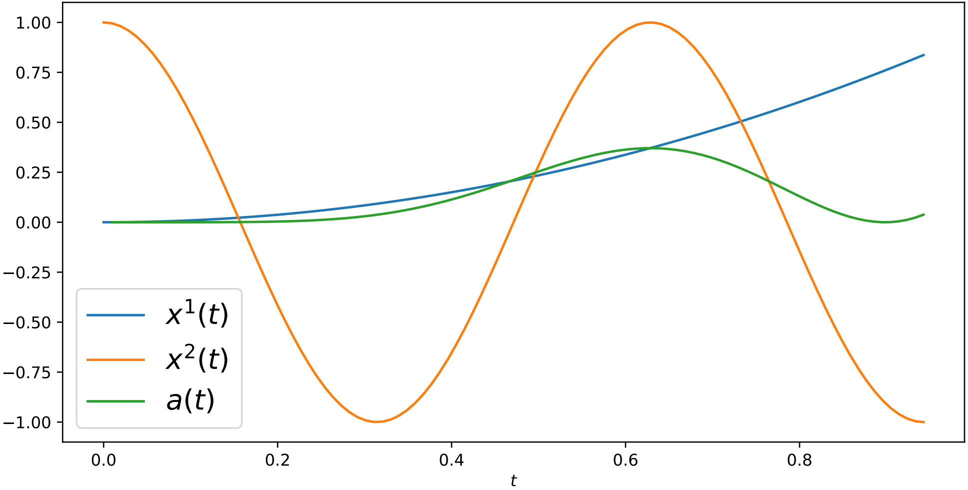


Figure: The path and the induced area path $t \mapsto \mathbb{a}_{0,t}$.

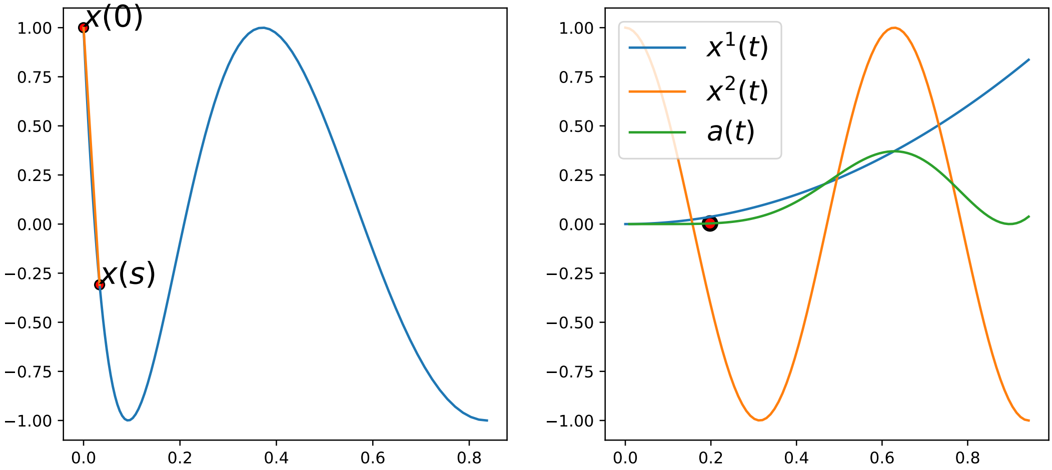


Figure: Construction of the induced area path $t \mapsto \mathbb{a}_{0,t}$.

Example: Signatures and areas of $x(t) = (\alpha\cosh(\theta_1 t) - \alpha,\ \cos(\theta_2 t))$, $d = 2$

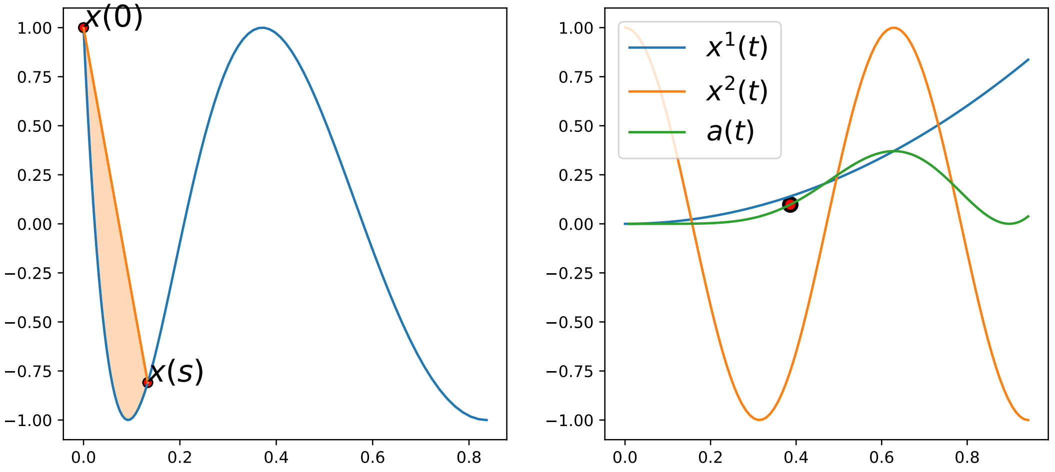


Figure: Construction of the induced area path $t \mapsto \mathbb{a}_{0,t}$.

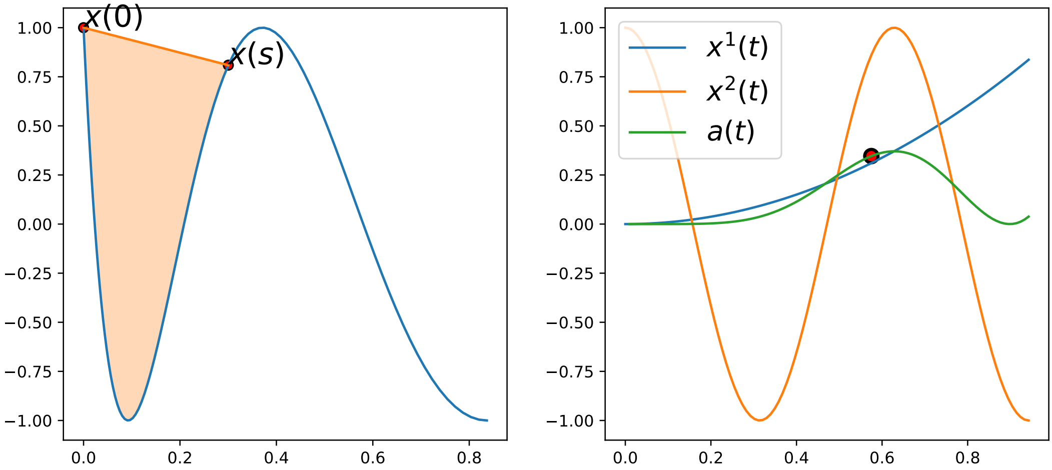
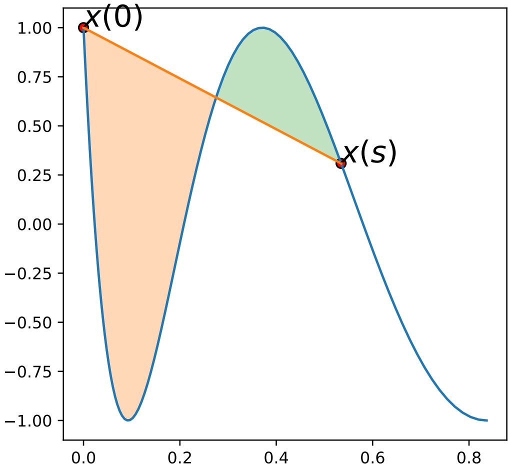


Figure: Construction of the induced area path $t \mapsto \mathbb{a}_{0,t}$.

Example: Signatures and areas of $x(t) = (\alpha \cosh(\theta_1 t) - \alpha, \cos(\theta_2 t))$, $d = 2$



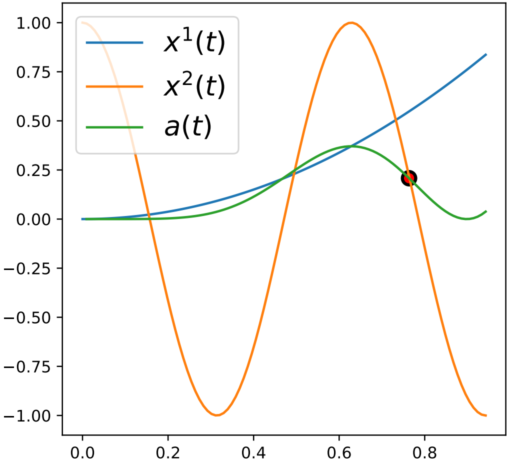


Figure: Construction of the induced area path $t \mapsto \mathbb{a}_{0,t}$.

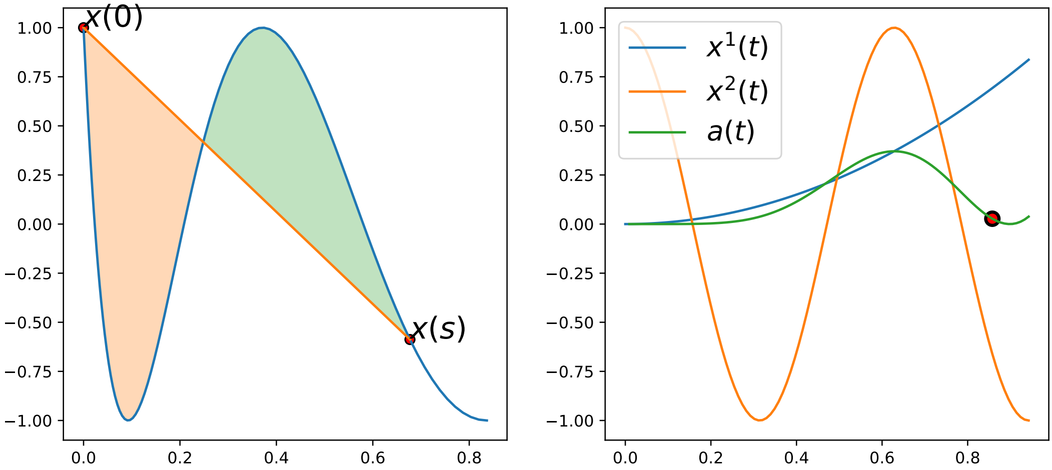
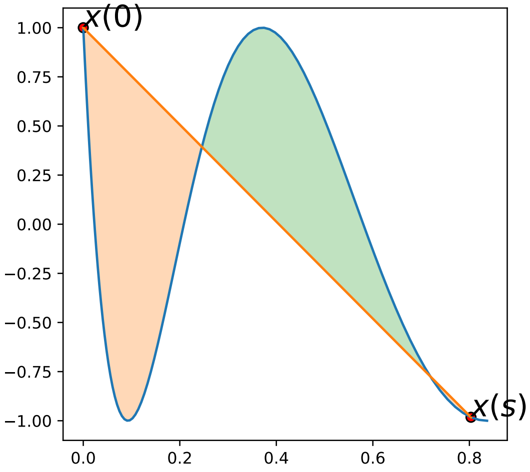


Figure: Construction of the induced area path $t \mapsto \mathbb{a}_{0,t}$.

Example: Signatures and areas of $x(t) = (\alpha \cosh(\theta_1 t) - \alpha,\ \cos(\theta_2 t))$, $d = 2$



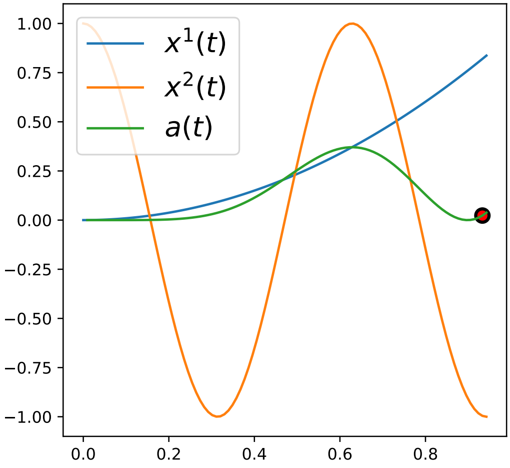


Figure: Construction of the induced area path $t \mapsto \mathbb{a}_{0,t}$.

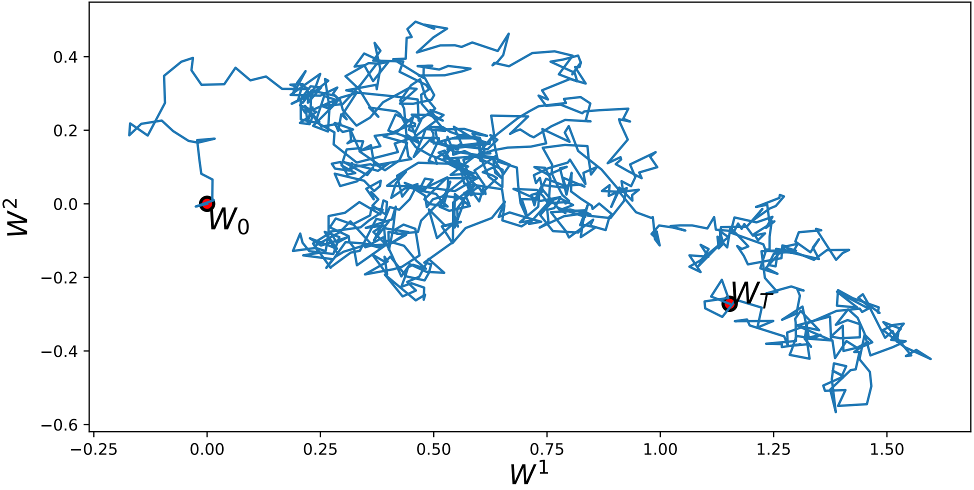


Figure: Path of a two-dimensional Brownian motion

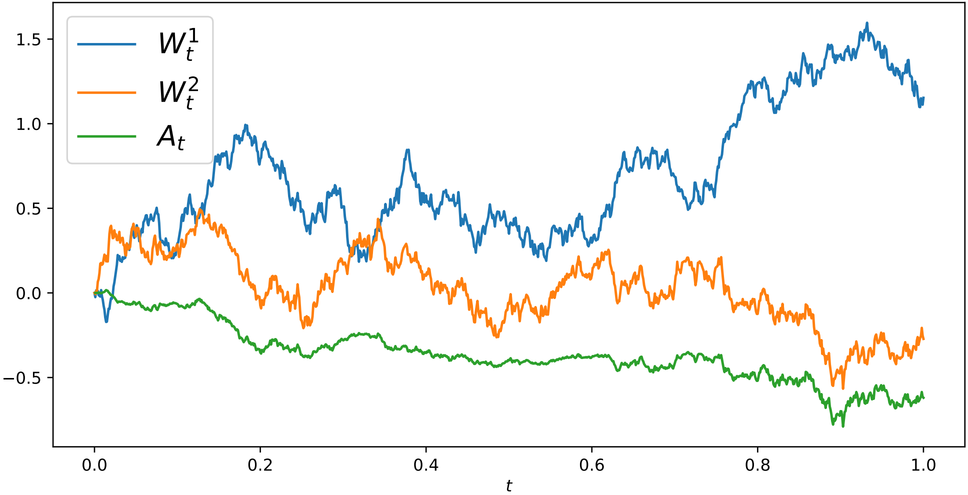


Figure: Path and area of a two-dimensional Brownian motion

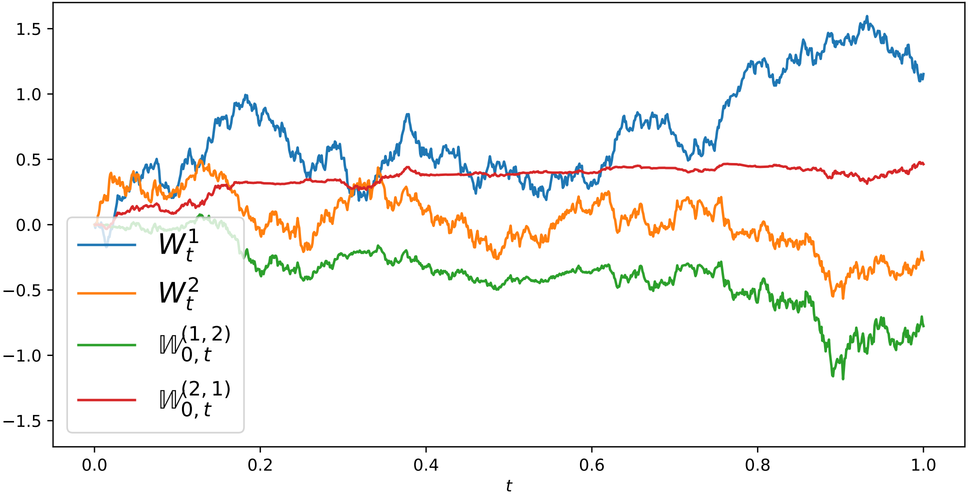


Figure: Path of W and non-trivial entries of $\mathbb{W}_{0,t}^{\leq 2}$ – note that $\mathbb{W}_{0,t}^{(i,i)} = \frac{1}{2}(W_{0,t}^i)^2$.

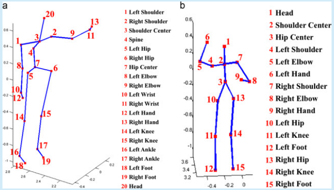
- Input data: a path or, more realistically, a time series in d dimensions.

- Input data: a path or, more realistically, a time series in d dimensions.
- Feature transformation: extract a finite dimensional projection of the path-signature.
- ML framework: plug the features into a standard ML framework, e.g., random forest or deep neural network.

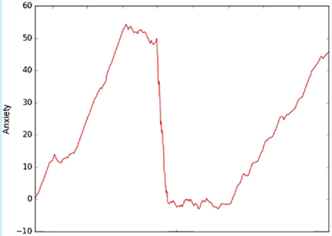
- Input data: a path or, more realistically, a time series in d dimensions.
- Feature transformation: extract a finite dimensional projection of the path-signature.
- ML framework: plug the features into a standard ML framework, e.g., random forest or deep neural network.

Examples [Terry Lyons and co-authors]

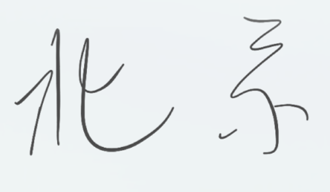
Human action recognotion



Psychiatric diagnosis



Chinese handwriting



- Time-extended path: Recall that the signature $\mathbb{x}_{s,t}^{<\infty}$ is invariant under re-parameterization. If this is not appropriate, extend x to $\overline{x}(u) := (u, x(u)) \in \mathbb{R}^{d+1}$. Its signature $\overline{\mathbb{x}}_{s,t}^{<\infty}$ effectively respects the given parameterization.

- Time-extended path: Recall that the signature $\mathbb{x}^{<\infty}_{s,t}$ is invariant under re-parameterization. If this is not appropriate, extend x to $\overline{x}(u) := (u, x(u)) \in \mathbb{R}^{d+1}$. Its signature $\overline{\mathbb{x}}^{<\infty}_{s,t}$ effectively respects the given parameterization.
- Interpolation in time: Given a time series $(x_1, x_2, \ldots)$, choose the appropriate interpolation to construct a path. Popular choices: piece-wise linear or piece-wise axis-parallel.
- Discrete time signature: Alternatively, choose discrete time signatures.

- Time-extended path: Recall that the signature $\mathbb{x}_{s,t}^{<\infty}$ is invariant under re-parameterization. If this is not appropriate, extend x to $\overline{x}(u) := (u, x(u)) \in \mathbb{R}^{d+1}$. Its signature $\overline{\mathbb{x}}_{s,t}^{<\infty}$ effectively respects the given parameterization.
- Interpolation in time: Given a time series $(x_1, x_2, \ldots)$, choose the appropriate interpolation to construct a path. Popular choices: piece-wise linear or piece-wise axis-parallel.
- Discrete time signature: Alternatively, choose discrete time signatures.
- Lead-lag-transform: Especially for financial time series, extend a time series $(x_1, x_2, x_3, \ldots)$ to $((x_1, x_1), (x_2, x_1), (x_3, x_2), (x_4, x_3), \ldots)$. (Related to quadratic variation.)

- Time-extended path: Recall that the signature $\mathbb{x}_{s,t}^{<\infty}$ is invariant under re-parameterization. If this is not appropriate, extend x to $\overline{x}(u) := (u, x(u)) \in \mathbb{R}^{d+1}$. Its signature $\overline{\mathbb{x}}_{s,t}^{<\infty}$ effectively respects the given parameterization.
- Interpolation in time: Given a time series $(x_1, x_2, \ldots)$, choose the appropriate interpolation to construct a path. Popular choices: piece-wise linear or piece-wise axis-parallel.
- Discrete time signature: Alternatively, choose discrete time signatures.
- Lead-lag-transform: Especially for financial time series, extend a time series $(x_1, x_2, x_3, \ldots)$ to $((x_1, x_1), (x_2, x_1), (x_3, x_2), (x_4, x_3), \ldots)$. (Related to quadratic variation.)

Modern trends

- Neural (rough) DEs.
- Signature kernel methods

- K.-T. Chen. Iterated integrals and exponential homomorphisms, Proceedings of the London Mathematical Society 3(1):502–512, 1954.
- I. Chevyrev, A. Kormilitzin. A primer on the signature method in machine learning, arXiv:1603.03788, 2016.
- M. Fliess. Fonctionnelles causales non linéaires et indéterminées non commutatives, Bulletin de la société mathématique de France 109:3–40, 1981.
- P.K. Friz, B. Nicolas. Multidimensional stochastic processes as rough paths: theory and applications, Vol. 120. Cambridge University Press, 2010.

Figure: Kuo-Tsai Chen (1923–1987)

Can we solve $\mathrm{d}y(t) = V(y(t))\mathrm{d}x(t)$ for a non-smooth path $x : [0, T] \to \mathbb{R}^d$ – e.g., α-Hölder?

Can we solve $\mathrm{d}y(t) = V(y(t))\mathrm{d}x(t)$ for a non-smooth path $x : [0, T] \to \mathbb{R}^d$ – e.g., α-Hölder?

- Standard recipe: Let x_n be smooth paths such that $\|x_n - x\|_? \xrightarrow{n\to\infty} 0$. Define y as limit of solutions y_n to $\mathrm{d}y_n(t) = V(y_n(t))\mathrm{d}x_n(t)$.

Can we solve $\mathrm{d}y(t) = V(y(t))\mathrm{d}x(t)$ for a non-smooth path $x : [0, T] \to \mathbb{R}^d$ – e.g., α-Hölder?

- Standard recipe: Let x_n be smooth paths such that $\|x_n - x\|_? \xrightarrow{n\to\infty} 0$. Define y as limit of solutions y_n to $\mathrm{d}y_n(t) = V(y_n(t))\mathrm{d}x_n(t)$.

Example

- Let $x_n(t) := \left(\sin(n^2 t)/n, \cos(n^2 t)/n\right)$, $t \in [0, 2\pi]$, with limit $x(t) \equiv 0$, and the area

$$z_n(t) := \frac{1}{2}\int_0^t x_n^1(s)\mathrm{d}x_n^2(s) - \frac{1}{2}\int_0^t x_n^2(s)\mathrm{d}x_n^1(s)$$

Can we solve $\mathrm{d}y(t) = V(y(t))\mathrm{d}x(t)$ for a non-smooth path $x : [0,T] \to \mathbb{R}^d$ – e.g., α-Hölder?

- Standard recipe: Let x_n be smooth paths such that $\|x_n - x\|_? \xrightarrow{n\to\infty} 0$. Define y as limit of solutions y_n to $\mathrm{d}y_n(t) = V(y_n(t))\mathrm{d}x_n(t)$.

Example

- Let $x_n(t) := \left(\sin(n^2 t)/n, \cos(n^2 t)/n\right)$, $t \in [0, 2\pi]$, with limit $x(t) \equiv 0$, and the area

$$z_n(t) := \frac{1}{2}\int_0^t x_n^1(s)\mathrm{d}x_n^2(s) - \frac{1}{2}\int_0^t x_n^2(s)\mathrm{d}x_n^1(s)$$
$$= -\frac{1}{2}\left(\int_0^t \sin\left(n^2 s\right)^2 \mathrm{d}s + \int_0^t \cos\left(n^2 s\right)^2 \mathrm{d}s\right)$$

Can we solve $\mathrm{d}y(t) = V(y(t))\mathrm{d}x(t)$ for a non-smooth path $x : [0, T] \to \mathbb{R}^d$ – e.g., α-Hölder?

- Standard recipe: Let x_n be smooth paths such that $\|x_n - x\|_? \xrightarrow{n\to\infty} 0$. Define y as limit of solutions y_n to $\mathrm{d}y_n(t) = V(y_n(t))\mathrm{d}x_n(t)$.

Example

- Let $x_n(t) := \left(\sin(n^2 t)/n,\ \cos(n^2 t)/n\right)$, $t \in [0, 2\pi]$, with limit $x(t) \equiv 0$, and the area

$$z_n(t) := \frac{1}{2}\int_0^t x_n^1(s)\mathrm{d}x_n^2(s) - \frac{1}{2}\int_0^t x_n^2(s)\mathrm{d}x_n^1(s)$$
$$= -\frac{1}{2}\int_0^t 1\mathrm{d}s = -\frac{1}{2}t \not\to 0 = z(t) \text{ as } n \to \infty.$$

Can we solve $\mathrm{d}y(t) = V(y(t))\mathrm{d}x(t)$ for a non-smooth path $x : [0,T] \to \mathbb{R}^d$ – e.g., α-Hölder?

- Standard recipe: Let x_n be smooth paths such that $\|x_n - x\|_? \xrightarrow{n\to\infty} 0$. Define y as limit of solutions y_n to $\mathrm{d}y_n(t) = V(y_n(t))\mathrm{d}x_n(t)$.

Example

- Let $x_n(t) := \big(\sin(n^2 t)/n,\ \cos(n^2 t)/n\big)$, $t \in [0, 2\pi]$, with limit $x(t) \equiv 0$, and the area

$$z_n(t) := \frac{1}{2}\int_0^t x_n^1(s)\mathrm{d}x_n^2(s) - \frac{1}{2}\int_0^t x_n^2(s)\mathrm{d}x_n^1(s)$$
$$= -\frac{1}{2}\int_0^t 1\mathrm{d}s = -\frac{1}{2}t \not\to 0 = z(t) \text{ as } n \to \infty.$$

- Note that $y_n(t) := (x_n^1(t), x_n^2(t), z_n(t))$ solves controlled DE with $V(y) := \begin{pmatrix} 1 & 0 \\ 0 & 1 \\ \frac{1}{2}y^2 & -\frac{1}{2}y^1 \end{pmatrix}$.

$$dy(t) = V(y(t))dx(t), \ t \in [0, T], \ y(0) = y_0, \ x : [0, T] \to \mathbb{R}^d, \ 0 = t_0 < \cdots < t_n = T.$$

$$\mathrm{d}y(t) = V(y(t))\mathrm{d}x(t),\ t \in [0,T],\ y(0) = y_0,\ x : [0,T] \to \mathbb{R}^d,\ 0 = t_0 < \cdots < t_n = T.$$

Case: smooth path x. If x is smooth, we have $\left|x_{t_i,t_{i+1}}\right| = \mathcal{O}(|t_{i+1} - t_i|)$. By Taylor,

$$y(t_{i+1}) = y(t_i) + V(y(t_i))x_{t_i,t_{i+1}} + \text{H.O.T.}_i, \quad |\text{H.O.T.}_i| = \mathcal{O}(|t_{i+1} - t_i|^2) = o(|t_{i+1} - t_i|).$$

Ignoring error propagation, the Euler scheme converges as $\sum_{i=0}^{n-1} |\text{H.O.T.}_i| = o(1)$, $n \to \infty$.

$$\mathrm{d}y(t) = V(y(t))\mathrm{d}x(t),\ t \in [0,T],\ y(0) = y_0,\ x : [0,T] \to \mathbb{R}^d,\ 0 = t_0 < \cdots < t_n = T.$$

Case: smooth path x. If x is smooth, we have $\left|x_{t_i,t_{i+1}}\right| = \mathcal{O}(|t_{i+1} - t_i|)$. By Taylor,

$$y(t_{i+1}) = y(t_i) + V(y(t_i))x_{t_i,t_{i+1}} + \text{H.O.T.}_i, \quad |\text{H.O.T.}_i| = \mathcal{O}(|t_{i+1} - t_i|^2) = o(|t_{i+1} - t_i|).$$

Ignoring error propagation, the Euler scheme converges as $\sum_{i=0}^{n-1} |\text{H.O.T.}_i| = o(1)$, $n \to \infty$.

Case: α-Hölder path x, $\alpha > \frac{1}{2}$. We have $\left|x_{t_i,t_{i+1}}\right| = \mathcal{O}(|t_{i+1} - t_i|^\alpha)$. By Taylor,

$$y(t_{i+1}) = y(t_i) + V(y(t_i))x_{t_i,t_{i+1}} + \text{H.O.T.}_i, \quad |\text{H.O.T.}_i| = \mathcal{O}(|t_{i+1} - t_i|^{2\alpha}) = o(|t_{i+1} - t_i|).$$

Ignoring error propagation, the Euler scheme converges as $\sum_{i=0}^{n-1} |\text{H.O.T.}_i| = o(1)$, $n \to \infty$.

$$\mathrm{d}y(t) = V(y(t))\mathrm{d}x(t),\; t \in [0,T],\; y(0) = y_0,\; x : [0,T] \to \mathbb{R}^d,\; 0 = t_0 < \cdots < t_n = T.$$

Case: smooth path x. If x is smooth, we have $\left|x_{t_i,t_{i+1}}\right| = \mathcal{O}(|t_{i+1} - t_i|)$. By Taylor,

$$y(t_{i+1}) = y(t_i) + V(y(t_i))x_{t_i,t_{i+1}} + \text{H.O.T.}_i, \quad |\text{H.O.T.}_i| = \mathcal{O}(|t_{i+1} - t_i|^2) = o(|t_{i+1} - t_i|).$$

Ignoring error propagation, the Euler scheme converges as $\sum_{i=0}^{n-1} |\text{H.O.T.}_i| = o(1)$, $n \to \infty$.

Case: α-Hölder path x, $\alpha > \frac{1}{2}$. We have $\left|x_{t_i,t_{i+1}}\right| = \mathcal{O}(|t_{i+1} - t_i|^\alpha)$. By Taylor,

$$y(t_{i+1}) = y(t_i) + V(y(t_i))x_{t_i,t_{i+1}} + \text{H.O.T.}_i, \quad |\text{H.O.T.}_i| = \mathcal{O}(|t_{i+1} - t_i|^{2\alpha}) = o(|t_{i+1} - t_i|).$$

Ignoring error propagation, the Euler scheme converges as $\sum_{i=0}^{n-1} |\text{H.O.T.}_i| = o(1)$, $n \to \infty$.

Remark: (Young '30s) $\int_0^T f(s)\mathrm{d}g(s)$ well-defined for f α-Hölder, g β-Hölder iff $\alpha + \beta > 1$.

$$\mathrm{d}y(t) = V(y(t))\mathrm{d}x(t),\ t \in [0,T],\ y(0) = y_0,\ x : [0,T] \to \mathbb{R}^d,\ 0 = t_0 < \cdots < t_n = T.$$

Now consider x to be α-Hölder with $\frac{1}{3} < \alpha \leq \frac{1}{2}$. By the previous calculation, the Euler scheme diverges. Recall the formal second order expansion:

$$y(t_{i+1}) + V(y(t_i))x_{t_i,t_{i+1}} + DV(y(t_i))V(y(t_i))\mathbb{x}^{=2}_{t_i,t_{i+1}} + \text{H.O.T.}_i.$$

$$\mathrm{d}y(t) = V(y(t))\mathrm{d}x(t),\ t \in [0,T],\ y(0) = y_0,\ x : [0,T] \to \mathbb{R}^d,\ 0 = t_0 < \cdots < t_n = T.$$

Now consider x to be α-Hölder with $\frac{1}{3} < \alpha \leq \frac{1}{2}$. By the previous calculation, the Euler scheme diverges. Recall the formal second order expansion:

$$y(t_{i+1}) + V(y(t_i))x_{t_i,t_{i+1}} + DV(y(t_i))V(y(t_i))\mathbb{x}^{=2}_{t_i,t_{i+1}} + \text{H.O.T.}_i.$$

Key observation

Assume that we could define $\mathbb{x}^{=2}_{t_i,t_{i+1}} = \left(\int_{t_i}^{t_{i+1}} x^j_{t_i,s}\mathrm{d}x^k(s)\right)_{j,k=1,\dots,d}$. Then we would expect

$$\left|x_{t_i,t_{i+1}}\right| = O(|t_{i+1}-t_i|^{\alpha}), \quad \left|\mathbb{x}^{=2}_{t_i,t_{i+1}}\right| = O(|t_{i+1}-t_i|^{2\alpha}), \quad |\text{H.O.T.}_i| = O(|t_{i+1}-t_i|^{3\alpha}) = o(|t_{i+1}-t_i|).$$

Hence, we expect convergence of the extended Euler scheme

$$\bar{y}_{i+1} = \bar{y}_i + V(\bar{y}_i)x_{t_i,t_{i+1}} + DV(\bar{y}_i)V(\bar{y}_i)\mathbb{x}^{=2}_{t_i,t_{i+1}}.$$

Definition (α-Hölder rough paths)

Let $\frac{1}{3} < \alpha \leq \frac{1}{2}$. An α-Hölder rough path on $\mathbb{R}^d$ is a pair $\mathbf{x} = (x, \mathbb{x})$, $x : [0, T] \to \mathbb{R}^d$, $\mathbb{x} : [0, T]^2 \to \mathbb{R}^d \otimes \mathbb{R}^d$, continuous, such that Chen's identity (truncated to $N = 2$) holds and

$$\sup_{s \neq t} \frac{\left| x_{s,t} \right|}{|t - s|^{\alpha}} < \infty, \quad \sup_{s \neq t} \frac{\left| \mathbb{x}_{s,t} \right|}{|t - s|^{2\alpha}} < \infty.$$

Definition (α-Hölder rough paths)

Let $\frac{1}{3} < \alpha \le \frac{1}{2}$. An α-Hölder rough path on $\mathbb{R}^d$ is a pair $\mathbf{x} = (x, \mathbb{x})$, $x : [0,T] \to \mathbb{R}^d$, $\mathbb{x} : [0,T]^2 \to \mathbb{R}^d \otimes \mathbb{R}^d$, continuous, such that Chen's identity (truncated to $N = 2$) holds and

$$\sup_{s \neq t} \frac{|x_{s,t}|}{|t-s|^{\alpha}} < \infty, \qquad \sup_{s \neq t} \frac{|\mathbb{x}_{s,t}|}{|t-s|^{2\alpha}} < \infty.$$

- The definition can be extended to general $\alpha > 0$, by providing $\lfloor 1/\alpha \rfloor$ iterated integrals.

Definition (α-Hölder rough paths)

Let $\frac{1}{3} < \alpha \leq \frac{1}{2}$. An α-Hölder rough path on $\mathbb{R}^d$ is a pair $\mathbf{x} = (x, \mathbb{x})$, $x : [0, T] \to \mathbb{R}^d$, $\mathbb{x} : [0, T]^2 \to \mathbb{R}^d \otimes \mathbb{R}^d$, continuous, such that Chen's identity (truncated to $N = 2$) holds and

$$\sup_{s \neq t} \frac{|x_{s,t}|}{|t - s|^{\alpha}} < \infty, \qquad \sup_{s \neq t} \frac{|\mathbb{x}_{s,t}|}{|t - s|^{2\alpha}} < \infty.$$

- The definition can be extended to general $\alpha > 0$, by providing $\lfloor 1/\alpha \rfloor$ iterated integrals.
- Every α-Hölder path can be extended to an α-Hölder rough path, but the extension is generally not unique. (N.b.: If x is smooth, there is a canonical choice.)

Definition (α-Hölder rough paths)

Let $\frac{1}{3} < \alpha \leq \frac{1}{2}$. An α-Hölder rough path on $\mathbb{R}^d$ is a pair $\mathbf{x} = (x, \mathbb{x})$, $x : [0,T] \to \mathbb{R}^d$, $\mathbb{x} : [0,T]^2 \to \mathbb{R}^d \otimes \mathbb{R}^d$, continuous, such that Chen's identity (truncated to $N = 2$) holds and

$$\sup_{s \neq t} \frac{\left| x_{s,t} \right|}{|t-s|^{\alpha}} < \infty, \quad \sup_{s \neq t} \frac{\left| \mathbb{x}_{s,t} \right|}{|t-s|^{2\alpha}} < \infty.$$

- The definition can be extended to general $\alpha > 0$, by providing $\lfloor 1/\alpha \rfloor$ iterated integrals.
- Every α-Hölder path can be extended to an α-Hölder rough path, but the extension is generally not unique. (N.b.: If x is smooth, there is a canonical choice.)
- The theory of rough paths was developed by Terry Lyons starting from 1994. Important re-formulations and generalizations were due to Massimiliano Gubinelli (controlled rough paths) and Martin Hairer (regularity structures).

Universal limit theorem

Given an α-Hölder rough path $\mathbf{x}$, and $V \in C^\gamma$ for $\gamma \geq 1/\alpha$. Then there is a unique solution of the rough differential equation

$$\mathrm{d}y(t) = V(y(t))\mathrm{d}\mathbf{x}(t), \quad y(0) = y_0.$$

The map $(y_0, V, \mathbf{x}) \to y$ is locally Lipschitz continuous – w.r.t. appropriate topologies.

Universal limit theorem

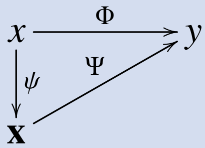
Given an α-Hölder rough path $\mathbf{x}$, and $V \in C^\gamma$ for $\gamma \geq 1/\alpha$. Then there is a unique solution of the rough differential equation

$$\mathrm{d}y(t) = V(y(t))\mathrm{d}\mathbf{x}(t), \quad y(0) = y_0.$$

The map $(y_0, V, \mathbf{x}) \to y$ is locally Lipschitz continuous – w.r.t. appropriate topologies.

- As the signature solves the RDE $\mathrm{d}\mathbb{x}^{<\infty}_{s,t} = \mathbb{x}^{<\infty}_{s,t} \otimes \mathrm{d}\mathbf{x}(t)$, $\mathbb{x}^{<\infty}_{s,s} = \mathbf{1}$, this implies that every rough path has a uniquely defined signature.
- The solution y depends on the rough path $\mathbf{x}$, i.e., the choice of extension of x.

Rough path principle



Given a d-dimensional Brownian motion W. We can construct iterated integrals (in an L^2 or almost sure sense) as follows

- $$\mathbb{W}^{(i,j),\text{Ito}}_{s,t} := \int_s^t W^i_{s,u} \mathrm{d}W^j_u, \quad \mathbb{W}^{=2,\text{Ito}}_{s,t} := \sum_{1\le i,j\le d} \mathbb{W}^{(i,j),\text{Ito}}_{s,t} e_i \otimes e_j,$$
- $$\mathbb{W}^{(i,j),\text{Strat}}_{s,t} := \int_s^t W^i_{s,u} \circ \mathrm{d}W^j_u, \quad \mathbb{W}^{=2,\text{Strat}}_{s,t} := \sum_{1\le i,j\le d} \mathbb{W}^{(i,j),\text{Strat}}_{s,t} e_i \otimes e_j.$$

Given a d-dimensional Brownian motion W. We can construct iterated integrals (in an L^2 or almost sure sense) as follows

- $\mathbb{W}_{s,t}^{(i,j),\text{Ito}} := \int_s^t W_{s,u}^i \mathrm{d}W_u^j, \quad \mathbb{W}_{s,t}^{=2,\text{Ito}} := \sum_{1\le i,j\le d} \mathbb{W}_{s,t}^{(i,j),\text{Ito}} e_i \otimes e_j,$
- $\mathbb{W}_{s,t}^{(i,j),\text{Strat}} := \int_s^t W_{s,u}^i \circ \mathrm{d}W_u^j, \quad \mathbb{W}_{s,t}^{=2,\text{Strat}} := \sum_{1\le i,j\le d} \mathbb{W}_{s,t}^{(i,j),\text{Strat}} e_i \otimes e_j.$
- Both $\mathbf{W}^{\text{Ito}}(\omega)$ and $\mathbf{W}^{\text{Strat}}(\omega)$ are a.s. α-Hölder rough paths, for any $\alpha < \frac{1}{2}$.

Given a d-dimensional Brownian motion W. We can construct iterated integrals (in an L^2 or almost sure sense) as follows

- $\mathbb{W}^{(i,j),\text{Ito}}_{s,t} := \int_s^t W^i_{s,u} \mathrm{d}W^j_u, \quad \mathbb{W}^{=2,\text{Ito}}_{s,t} := \sum_{1 \le i,j \le d} \mathbb{W}^{(i,j),\text{Ito}}_{s,t} e_i \otimes e_j,$
- $\mathbb{W}^{(i,j),\text{Strat}}_{s,t} := \int_s^t W^i_{s,u} \circ \mathrm{d}W^j_u, \quad \mathbb{W}^{=2,\text{Strat}}_{s,t} := \sum_{1 \le i,j \le d} \mathbb{W}^{(i,j),\text{Strat}}_{s,t} e_i \otimes e_j.$
- Both $\mathbf{W}^{\text{Ito}}(\omega)$ and $\mathbf{W}^{\text{Strat}}(\omega)$ are a.s. α-Hölder rough paths, for any $\alpha < \frac{1}{2}$.
- Solutions of RDEs driven by $\mathbf{W}^{\text{Ito}}$ coincide (a.s.) with the corresp. Ito-SDE solutions.
- Solutions of RDEs driven by $\mathbf{W}^{\text{Strat}}$ coincide (a.s.) with the corresp. Stratonovich-SDE solutions.

Given a d-dimensional Brownian motion W. We can construct iterated integrals (in an L^2 or almost sure sense) as follows

- $\mathbb{W}^{(i,j),\text{Ito}}_{s,t} := \int_s^t W^i_{s,u} \mathrm{d}W^j_u, \quad \mathbb{W}^{=2,\text{Ito}}_{s,t} := \sum_{1 \le i,j \le d} \mathbb{W}^{(i,j),\text{Ito}}_{s,t} e_i \otimes e_j,$
- $\mathbb{W}^{(i,j),\text{Strat}}_{s,t} := \int_s^t W^i_{s,u} \circ \mathrm{d}W^j_u, \quad \mathbb{W}^{=2,\text{Strat}}_{s,t} := \sum_{1 \le i,j \le d} \mathbb{W}^{(i,j),\text{Strat}}_{s,t} e_i \otimes e_j.$
- Both $\mathbf{W}^{\text{Ito}}(\omega)$ and $\mathbf{W}^{\text{Strat}}(\omega)$ are a.s. α-Hölder rough paths, for any $\alpha < \frac{1}{2}$.
- Solutions of RDEs driven by $\mathbf{W}^{\text{Ito}}$ coincide (a.s.) with the corresp. Ito-SDE solutions.
- Solutions of RDEs driven by $\mathbf{W}^{\text{Strat}}$ coincide (a.s.) with the corresp. Stratonovich-SDE solutions.
- $\omega \mapsto \mathbf{W}^{\text{Ito/Strat}}(\omega)$ is discontinuous, the solution map in $\mathbf{W}^{\text{Ito/Strat}}(\omega)$ is continuous.

Given a d-dimensional Brownian motion W. We can construct iterated integrals (in an L^2 or almost sure sense) as follows

- $\mathbb{W}^{(i,j),\text{Ito}}_{s,t} := \int_s^t W^i_{s,u} \mathrm{d}W^j_u, \quad \mathbb{W}^{=2,\text{Ito}}_{s,t} := \sum_{1 \le i,j \le d} \mathbb{W}^{(i,j),\text{Ito}}_{s,t} e_i \otimes e_j,$
- $\mathbb{W}^{(i,j),\text{Strat}}_{s,t} := \int_s^t W^i_{s,u} \circ \mathrm{d}W^j_u, \quad \mathbb{W}^{=2,\text{Strat}}_{s,t} := \sum_{1 \le i,j \le d} \mathbb{W}^{(i,j),\text{Strat}}_{s,t} e_i \otimes e_j.$
- Both $\mathbf{W}^{\text{Ito}}(\omega)$ and $\mathbf{W}^{\text{Strat}}(\omega)$ are a.s. α-Hölder rough paths, for any $\alpha < \frac{1}{2}$.
- Solutions of RDEs driven by $\mathbf{W}^{\text{Ito}}$ coincide (a.s.) with the corresp. Ito-SDE solutions.
- Solutions of RDEs driven by $\mathbf{W}^{\text{Strat}}$ coincide (a.s.) with the corresp. Stratonovich-SDE solutions.
- $\omega \mapsto \mathbf{W}^{\text{Ito/Strat}}(\omega)$ is discontinuous, the solution map in $\mathbf{W}^{\text{Ito/Strat}}(\omega)$ is continuous.
- Note that $\mathbb{W}^{<\infty,\text{Strat}}$ satisfies the shuffle identity, but $\mathbb{W}^{<\infty,\text{Ito}}$ does not.

Let $\mathscr{C}^\alpha([0,T];\mathbb{R}^d)$ denote the space of α-Hölder rough paths.

Let $\mathscr{C}^\alpha([0,T];\mathbb{R}^d)$ denote the space of α-Hölder rough paths.

- While $\mathscr{C}^\alpha([0,T];\mathbb{R}^d)$ is not a linear space, it is a complete metric space with the appropriate Hölder-distance.

Let $\mathscr{C}^\alpha([0,T];\mathbb{R}^d)$ denote the space of α-Hölder rough paths.

- While $\mathscr{C}^\alpha([0,T];\mathbb{R}^d)$ is not a linear space, it is a complete metric space with the appropriate Hölder-distance.

Given a smooth path $x:[0,T]\to\mathbb{R}^d$, construct a corresponding α-Hölder rough path $\mathbf{x}$ by

$$\mathbf{x}=(x,\mathbb{x}),\quad x_{s,t}:=x(t)-x(s),\quad \mathbb{x}^{(i,j)}_{s,t}:=\int_s^t x^i(u)\mathrm{d}x^j(u).$$

Let $\mathscr{C}^\alpha([0,T];\mathbb{R}^d)$ denote the space of α-Hölder rough paths.

- While $\mathscr{C}^\alpha([0,T];\mathbb{R}^d)$ is not a linear space, it is a complete metric space with the appropriate Hölder-distance.

Given a smooth path $x:[0,T]\to\mathbb{R}^d$, construct a corresponding α-Hölder rough path $\mathbf{x}$ by

$$\mathbf{x} = (x, \mathbb{x}), \quad x_{s,t} := x(t) - x(s), \quad \mathbb{x}^{(i,j)}_{s,t} := \int_s^t x^i(u)\mathrm{d}x^j(u).$$

Let $\mathscr{C}_g^\alpha([0,T];\mathbb{R}^d) \subset \mathscr{C}^\alpha([0,T];\mathbb{R}^d)$ denote the closure of smooth rough paths in $\mathscr{C}^\alpha([0,T];\mathbb{R}^d)$. $\mathbf{x} \in \mathscr{C}_g^\alpha([0,T];\mathbb{R}^d)$ is called geometric.

Let $\mathscr{C}^\alpha([0,T];\mathbb{R}^d)$ denote the space of α-Hölder rough paths.

- While $\mathscr{C}^\alpha([0,T];\mathbb{R}^d)$ is not a linear space, it is a complete metric space with the appropriate Hölder-distance.

Given a smooth path $x : [0,T] \to \mathbb{R}^d$, construct a corresponding α-Hölder rough path $\mathbf{x}$ by

$$\mathbf{x} = (x, \mathbb{x}), \quad x_{s,t} := x(t) - x(s), \quad \mathbb{x}_{s,t}^{(i,j)} := \int_s^t x^i(u) \mathrm{d}x^j(u).$$

Let $\mathscr{C}_g^\alpha([0,T];\mathbb{R}^d) \subset \mathscr{C}^\alpha([0,T];\mathbb{R}^d)$ denote the closure of smooth rough paths in $\mathscr{C}^\alpha([0,T];\mathbb{R}^d)$. $\mathbf{x} \in \mathscr{C}_g^\alpha([0,T];\mathbb{R}^d)$ is called geometric.

- The signature $\mathbb{x}_{s,t}^{<\infty}$ of a geometric rough path $\mathbf{x} \in \mathscr{C}_g^\alpha$ satisfies the shuffle identity. Symbolically,

$$\forall \mathbf{x} \in \mathscr{C}_g^\alpha([0,T];\mathbb{R}^d),\ \forall 0 \le s \le t \le T : \ \mathbb{x}_{s,t}^{<\infty} \in G(\mathbb{R}^d).$$

- P.K. Friz, M. Hairer. A course on rough paths, Springer International Publishing, 2020.
- P.K. Friz, N. Victoir. Multidimensional stochastic processes as rough paths: theory and applications, Vol. 120. Cambridge University Press, 2010.
- T. Lyons. Differential equations driven by rough signals, Revista Matemática Iberoamericana 14(2):215–310, 1998.

Figure: Terry Lyons

W.l.o.g., all paths start at 0, i.e., $x(0) = 0$.

- Let $\Omega_1 := \mathscr{C}^{1-\text{var}}([0,T];V)$ denote the space of bounded variation functions taking values in a (finite-dimensional) Banach space V with the norm $\|x\|_{1-\text{var}} := |x(0)| + |x|_{1-\text{var}}$, where

$$|x|_{1-\text{var}} := \sup_{N\in\mathbb{N}} \sup_{0\le t_0<t_1<\cdots<t_N\le T} \sum_{i=1}^{N} |x(t_{i+1}) - x(t_i)| .$$

W.l.o.g., all paths start at 0, i.e., $x(0) = 0$.

- Let $\Omega_1 := \mathscr{C}^{1-\mathrm{var}}([0,T];V)$ denote the space of bounded variation functions taking values in a (finite-dimensional) Banach space V with the norm $\|x\|_{1-\mathrm{var}} := |x(0)| + |x|_{1-\mathrm{var}}$, where

$$|x|_{1-\mathrm{var}} := \sup_{N\in\mathbb{N}} \sup_{0\le t_0<t_1<\cdots<t_N\le T} \sum_{i=1}^{N} |x(t_{i+1}) - x(t_i)|\,.$$

- Given $x \in \mathscr{C}^{1-\mathrm{var}}([0,T];\mathbb{R}^d)$, we obtain $t \mapsto \mathbb{x}_{0,t}^{\le N} \in C^{1-\mathrm{var}}\left([0,T];T^N(\mathbb{R}^d)\right)$ and the lift $x \mapsto \mathbb{x}_{0,\cdot}^{\le N}$ is continuous: $\left\|\mathbb{x}_{0,\cdot}^{\le N}\right\|_{1-\mathrm{var}} \le |x|_{1-\mathrm{var}}$ – provided that $V := G^N(\mathbb{R}^d)$ is equipped with the Carnot-Caratheodory metric.

W.l.o.g., all paths start at 0, i.e., $x(0) = 0$.

- Let $\Omega_1 := \mathscr{C}^{1-\mathrm{var}}([0,T];V)$ denote the space of bounded variation functions taking values in a (finite-dimensional) Banach space V with the norm $\|x\|_{1-\mathrm{var}} := |x(0)| + |x|_{1-\mathrm{var}}$, where

$$|x|_{1-\mathrm{var}} := \sup_{N\in\mathbb{N}} \sup_{0\le t_0<t_1<\cdots<t_N\le T} \sum_{i=1}^{N} |x(t_{i+1}) - x(t_i)| .$$

- Given $x \in \mathscr{C}^{1-\mathrm{var}}([0,T];\mathbb{R}^d)$, we obtain $t \mapsto \mathbb{x}^{\le N}_{0,t} \in C^{1-\mathrm{var}}\left([0,T]; T^N(\mathbb{R}^d)\right)$ and the lift $x \mapsto \mathbb{x}^{\le N}_{0,\cdot}$ is continuous: $\left\|\mathbb{x}^{\le N}_{0,\cdot}\right\|_{1-\mathrm{var}} \le |x|_{1-\mathrm{var}}$ – provided that $V := G^N(\mathbb{R}^d)$ is equipped with the Carnot-Caratheodory metric.
- Given $x \in \mathscr{C}^{1-\mathrm{var}}([0,T];\mathbb{R}^d)$, we define $\widehat{x}(t) := (t, x(t)) \in \mathbb{R}^{1+d}$ and denote $\widehat{\Omega}_1 := \{\, \widehat{x} \mid x \in \Omega_1 \,\}$. Note that $\widehat{x}$ is uniquely determined by its signature $\widehat{\mathbb{x}}^{<\infty}_{0,T}$ and $\widehat{x}(0)$!

Theorem

Let $A := \{ f_\ell \mid \ell \in \mathcal{W}_{1+d} \}$ *where for any* $\ell \in \mathcal{W}_{1+d}$ *we set*

$$f_\ell : \widehat{\Omega}_1 \to \mathbb{R}, \quad \widehat{x} \mapsto \left\langle \ell, \widehat{\mathbb{x}^{<\infty}_{0,T}} \right\rangle.$$

Then $A \subset C(\widehat{\Omega}_1; \mathbb{R})$ *is dense w.r.t. uniform convergence on compacts.*

Theorem

Let $A := \{ f_\ell \mid \ell \in \mathcal{W}_{1+d} \}$ *where for any* $\ell \in \mathcal{W}_{1+d}$ *we set*

$$f_\ell : \widehat{\Omega}_1 \to \mathbb{R}, \quad \widehat{x} \mapsto \left\langle \ell, \widehat{\mathbb{x}^{<\infty}_{0,T}} \right\rangle.$$

Then $A \subset C(\widehat{\Omega}_1; \mathbb{R})$ *is dense w.r.t. uniform convergence on compacts.*

The proof is based on the classical Stone – Weierstrass theorem. We give a sufficient version below:

Theorem (Stone – Weierstrass)

Let X *be a compact metric space and consider a subalgebra* $A \subset C(X; \mathbb{R})$ *that is point-separating and vanishes nowhere. Then* $A \subset C(X; \mathbb{R})$ *is dense w.r.t. uniform convergence.*

- We can replace $\widehat{\Omega}_1$ by $\mathscr{P}_1$, the set of bounded variation paths modulo re-parameterization and tree-like excursion.

- We can replace $\widehat{\Omega}_1$ by $\mathscr{P}_1$, the set of bounded variation paths modulo re-parameterization and tree-like excursion.
- We can immediately generalize the theorem to the rough setting, i.e., by replacing Ω_1 and $\widehat{\Omega}_1$ by their rough analogues for $p > 1$:

$$\Omega_p := \mathscr{C}_g^{1/p}([0,T];\mathbb{R}^d), \quad \widehat{\Omega}_p := \left\{ \mathbf{x} = (x, \mathbb{x}) \in \mathscr{C}_g^{1/p}([0,T];\mathbb{R}^{1+d}) \,\middle|\, \forall t \in [0,T] : \; x^0(t) = t \right\}.$$

- We can replace $\widehat{\Omega}_1$ by $\mathscr{P}_1$, the set of bounded variation paths modulo re-parameterization and tree-like excursion.
- We can immediately generalize the theorem to the rough setting, i.e., by replacing Ω_1 and $\widehat{\Omega}_1$ by their rough analogues for $p > 1$:

$$\Omega_p := \mathscr{C}_g^{1/p}([0,T];\mathbb{R}^d), \quad \widehat{\Omega}_p := \left\{ \mathbf{x} = (x, \mathbb{x}) \in \mathscr{C}_g^{1/p}([0,T];\mathbb{R}^{1+d}) \,\middle|\, \forall t \in [0,T] : \ x^0(t) = t \right\}.$$

- Unlike $\mathscr{C}^{1/p}([0,T];\mathbb{R}^d)$, $\mathscr{C}_g^{1/p}([0,T];\mathbb{R}^d)$ is separable, hence a Polish space. Any rough process defined as a random variable taking values in Ω_p or $\widehat{\Omega}_p$, respectively, is tight.

Corollary

Given a rough process $\widehat{\mathbf{X}}$ taking values in $\widehat{\Omega}_p$, $p > 1$. Then for any $f \in C\left(\widehat{\Omega}_p;\mathbb{R}\right)$ and $\epsilon > 0$ there is $\ell \in \mathcal{W}_{1+d}$ s.t.

$$P\left(\left|f(\widehat{\mathbf{X}}) - \left\langle \ell, \widehat{\mathbb{X}}^{<\infty}_{0,T} \right\rangle\right| > \epsilon\right) < \epsilon.$$

Theorem (Stone – Weierstrass theorem; Giles'71)

Let X be a ~~compact~~ metric space and consider a subalgebra $A \subset C_b(X;\mathbb{R})$ that is point-separating and vanishes nowhere. Then $A \subset C_b(X;\mathbb{R})$ is dense w.r.t. the strict topology.

- The strict topology on $C_b(X;\mathbb{R})$ is the topology generated by the seminorms $p_\psi(f) := \sup_{x \in X} |f(x)\psi(x)|$, $f \in C_b(X;\mathbb{R})$, indexed by the functions $\psi : X \to \mathbb{R}$ vanishing at infinity.

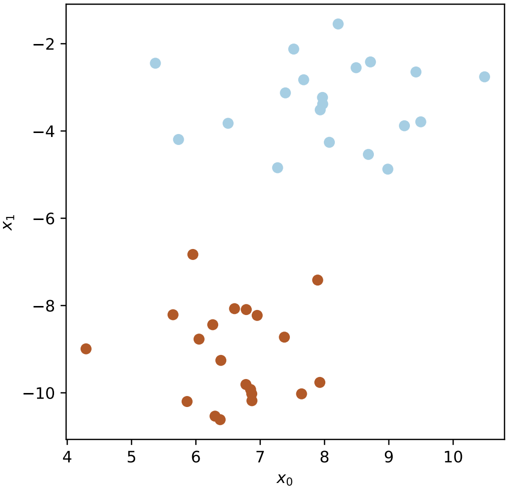
Theorem (Stone – Weierstrass theorem; Giles'71)

Let X be a ~~compact~~ metric space and consider a subalgebra $A \subset C_b(X;\mathbb{R})$ that is point-separating and vanishes nowhere. Then $A \subset C_b(X;\mathbb{R})$ is dense w.r.t. the strict topology.

- The strict topology on $C_b(X;\mathbb{R})$ is the topology generated by the seminorms $p_\psi(f) := \sup_{x\in X} |f(x)\psi(x)|$, $f \in C_b(X;\mathbb{R})$, indexed by the functions $\psi : X \to \mathbb{R}$ vanishing at infinity.
- Replace the (unbounded) functions $\widehat{x} \mapsto \left\langle \ell, \widehat{\mathbb{x}}^{<\infty}_{0,T} \right\rangle$ by the bounded functions $\widehat{x} \mapsto \left\langle \ell, \Lambda(\widehat{\mathbb{x}}^{<\infty}_{0,T}) \right\rangle$ for a tensor normalization $\Lambda : T((\mathbb{R}^d)) \to T((\mathbb{R}^d))$.
- Tensor normalizations are continuous, injective maps Λ s.t. $\Lambda(\mathbf{a})$ is in a bounded ball in $T((\mathbb{R}^d))$ and $\Lambda(\mathbf{a}) = \delta_{\lambda(\mathbf{a})}\mathbf{a}$ for some $\lambda : T((\mathbb{R}^d)) \to \mathbb{R}$.

- Consider data $x_i \in E$ for a (finite-dimensional) space E, with labels $y_i \in \{-1, +1\}$, $i = 1, \dots, M$.
- Classify data points by a separating hyperplane, i.e., find $w \in E$ and $b \in \mathbb{R}$ s.t. for all $i = 1, \dots, M$:

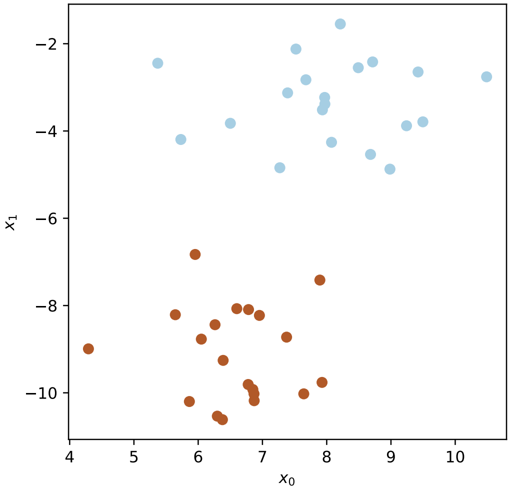
$$y_i = +1 \iff \langle w, x_i \rangle_E - b > 0,$$
$$y_i = -1 \iff \langle w, x_i \rangle_E - b < 0.$$



- Consider data $x_i \in E$ for a (finite-dimensional) space E, with labels $y_i \in \{-1, +1\}$, $i = 1, \ldots, M$.
- Classify data points by a separating hyperplane, i.e., find $w \in E$ and $b \in \mathbb{R}$ s.t. for all $i = 1, \ldots, M$:

$$y_i = +1 \iff \langle w, x_i \rangle_E - b > 0,$$
$$y_i = -1 \iff \langle w, x_i \rangle_E - b < 0.$$

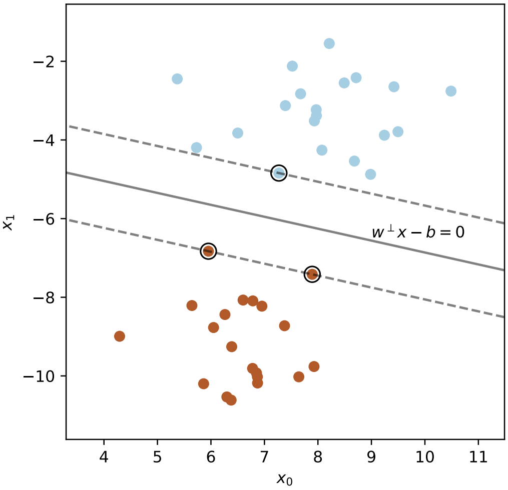
- If at all possible, there will be infinitely many solutions. Hence, we try to find the best solution.



Solution

$$\min_{w \in E,\, b \in \mathbb{R}} \frac{1}{2} \|w\|_E^2 \quad \text{subject to}$$

$$\forall i \in \{1, \ldots, M\}: \; y_i (\langle w, x_i \rangle_E - b) \geq 1.$$

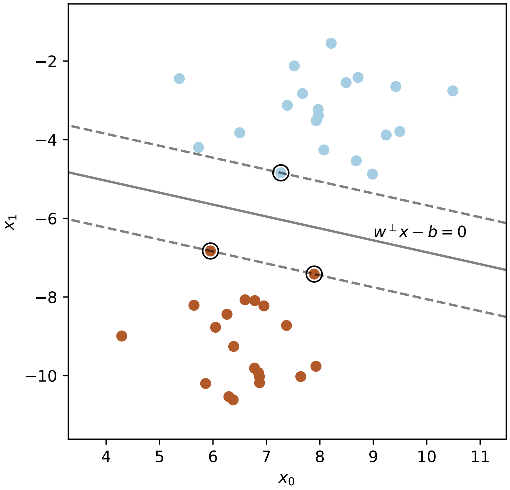


Solution

$$\min_{w\in E,\, b\in\mathbb{R}} \frac{1}{2}\|w\|_E^2 \text{ subject to}$$

$$\forall i \in \{1,\dots,M\}: \; y_i(\langle w, x_i\rangle_E - b) \geq 1.$$

- What if separation by hyperplanes is not possible, or data lives in a non-linear space $\mathcal{X}$?
- Lift data $x_i \mapsto \Phi(x_i)$ using a non-linear feature map $\Phi : \mathcal{X} \to \mathcal{H}$ for some (infinite-dimensional) Hilbert space $\mathcal{H}$.
- Which Φ? Evaluation very expensive!?



Definition

A reproducing kernel Hilbert space (RKHS) is a Hilbert space $\mathcal{H}$ of functions $f : \mathcal{X} \to \mathbb{R}$ s.t. for all $x \in \mathcal{X}$, the evaluation functional $\mathrm{ev}_x : \mathcal{H} \to \mathbb{R}$, $f \mapsto f(x)$ is continuous.

Definition

A reproducing kernel Hilbert space (RKHS) is a Hilbert space $\mathcal{H}$ of functions $f : \mathcal{X} \to \mathbb{R}$ s.t. for all $x \in \mathcal{X}$, the evaluation functional $\mathrm{ev}_x : \mathcal{H} \to \mathbb{R}, \ f \mapsto f(x)$ is continuous.

- By Riesz representation, for every $x \in \mathcal{X}$ we can find $k_x \in \mathcal{H}$ such that

$$\forall f \in \mathcal{H} : \ \mathrm{ev}_x(f) = \langle k_x, f \rangle_{\mathcal{H}} \, .$$

- Define $k : \mathcal{X} \times \mathcal{X} \to \mathbb{R}$, $k(x, y) := \left\langle k_x, k_y \right\rangle_{\mathcal{H}}$ called the kernel.

Definition

A reproducing kernel Hilbert space (RKHS) is a Hilbert space $\mathcal{H}$ of functions $f : \mathcal{X} \to \mathbb{R}$ s.t. for all $x \in \mathcal{X}$, the evaluation functional $\mathrm{ev}_x : \mathcal{H} \to \mathbb{R},\ f \mapsto f(x)$ is continuous.

- By Riesz representation, for every $x \in \mathcal{X}$ we can find $k_x \in \mathcal{H}$ such that

$$\forall f \in \mathcal{H} :\ \mathrm{ev}_x(f) = \langle k_x, f \rangle_{\mathcal{H}} .$$

- Define $k : \mathcal{X} \times \mathcal{X} \to \mathbb{R}$, $k(x, y) := \left\langle k_x, k_y \right\rangle_{\mathcal{H}}$ called the kernel.

1. By the analogue properties of $\langle \cdot, \cdot \rangle_{\mathcal{H}}$, k is symmetric and positive definite, i.e., $\forall x_1, \dots, x_k \in \mathcal{X}$, the matrix $(k(x_i, x_j)) \in \mathbb{R}^{k \times k}$ is positive definite.
2. $k_x(y) = \mathrm{ev}_y(k_x) = \left\langle k_y, k_x \right\rangle_{\mathcal{H}} = k(x, y)$, i.e., for any $x \in \mathcal{X}$, $k_x = k(x, \cdot)$.

Definition

A reproducing kernel Hilbert space (RKHS) is a Hilbert space $\mathcal{H}$ of functions $f : \mathcal{X} \to \mathbb{R}$ s.t. for all $x \in \mathcal{X}$, the evaluation functional $\mathrm{ev}_x : \mathcal{H} \to \mathbb{R}$, $f \mapsto f(x)$ is continuous.

- By Riesz representation, for every $x \in \mathcal{X}$ we can find $k_x \in \mathcal{H}$ such that

$$\forall f \in \mathcal{H} : \ \mathrm{ev}_x(f) = \langle k_x, f \rangle_{\mathcal{H}} .$$

- Define $k : \mathcal{X} \times \mathcal{X} \to \mathbb{R}$, $k(x,y) := \left\langle k_x, k_y \right\rangle_{\mathcal{H}}$ called the kernel.

1. By the analogue properties of $\langle \cdot, \cdot \rangle_{\mathcal{H}}$, k is symmetric and positive definite, i.e., $\forall x_1, \dots, x_k \in \mathcal{X}$, the matrix $(k(x_i, x_j)) \in \mathbb{R}^{k \times k}$ is positive definite.
2. $k_x(y) = \mathrm{ev}_y(k_x) = \left\langle k_y, k_x \right\rangle_{\mathcal{H}} = k(x,y)$, i.e., for any $x \in \mathcal{X}$, $k_x = k(x, \cdot)$.
3. Conversely, given a symmetric, positive definite kernel $k : \mathcal{X} \times \mathcal{X} \to \mathbb{R}$, we obtain a RKHS as completion of $\widetilde{H} := \langle \{ k(x, \cdot) \mid x \in \mathcal{X} \} \rangle$ with $\langle k(x, \cdot), k(y, \cdot) \rangle_{\widetilde{\mathcal{H}}} := k(x,y)$.

Given data $x_i \in \mathcal{X}$, choose a RKHS $\mathcal{H}$ on $\mathcal{X}$ and features $\Phi(x) := k(x, \cdot) \in \mathcal{H}$.

Given data $x_i \in \mathcal{X}$, choose a RKHS $\mathcal{H}$ on $\mathcal{X}$ and features $\Phi(x) := k(x, \cdot) \in \mathcal{H}$.

$$\min_{w \in \mathcal{H},\, b \in \mathbb{R}} \frac{1}{2} \|w\|_{\mathcal{H}}^2 \text{ subject to } \forall i \in \{1, \dots, M\}: \; y_i \left(\langle w, \Phi(x_i) \rangle_{\mathcal{H}} - b \right) \geq 1.$$

- By the representer theorem, $w \in \langle \{ k(x_i, \cdot) \mid i = 1, \dots, M \} \rangle$, i.e.,

$$\exists \alpha \in \mathbb{R}^M : \; w = \sum_{i=1}^{M} \alpha_i k(x_i, \cdot), \text{ hence } \|w\|_{\mathcal{H}}^2 = \sum_{i=1}^{M} \alpha^\top K \alpha, K := (k(x_i, x_j))_{i,j=1}^{M} \in \mathbb{R}^{M \times M}.$$

- Similarly, $\langle w, \Phi(x_i) \rangle_{\mathcal{H}} = \sum_{j=1}^{M} \alpha_j \left\langle k(x_j, \cdot), k(x_i, \cdot) \right\rangle_{\mathcal{H}} = (\alpha^\top K)_i$.

Given data $x_i \in \mathcal{X}$, choose a RKHS $\mathcal{H}$ on $\mathcal{X}$ and features $\Phi(x) := k(x, \cdot) \in \mathcal{H}$.

$$\min_{w \in \mathcal{H},\ b \in \mathbb{R}} \frac{1}{2} \|w\|_{\mathcal{H}}^2 \text{ subject to } \forall i \in \{1, \dots, M\} : \ y_i \left(\langle w, \Phi(x_i) \rangle_{\mathcal{H}} - b \right) \geq 1.$$

- By the representer theorem, $w \in \langle \{ k(x_i, \cdot) \mid i = 1, \dots, M \} \rangle$, i.e.,

$$\exists \alpha \in \mathbb{R}^M : \ w = \sum_{i=1}^{M} \alpha_i k(x_i, \cdot), \text{ hence } \|w\|_{\mathcal{H}}^2 = \sum_{i=1}^{M} \alpha^\top K \alpha, K := (k(x_i, x_j))_{i,j=1}^{M} \in \mathbb{R}^{M \times M}.$$

- Similarly, $\langle w, \Phi(x_i) \rangle_{\mathcal{H}} = \sum_{j=1}^{M} \alpha_j \left\langle k(x_j, \cdot), k(x_i, \cdot) \right\rangle_{\mathcal{H}} = (\alpha^\top K)_i$.
- Need evaluations of the kernel k (for the Gram matrix K), but not of Φ – kernel trick.

Let $\mathcal{X}_1 := \left\{ x \in \mathscr{C}^{1-\mathrm{var}}([0,T];\mathbb{R}^d) \,\middle|\, T > 0,\ x(0) = 0 \right\}$ – and similarly $\widehat{\mathcal{X}}_1$.

Goal: Define an appropriate kernel for paths / time series.

Definition

Given $x, y \in \mathcal{X}_1$ defined on $[0,t], [0,s]$, respectively. We define

$$\mathrm{k}_{\mathrm{sig}}(x,y) := \left\langle \mathbb{x}^{<\infty}_{0,t}, \mathbb{y}^{<\infty}_{0,s} \right\rangle := \sum_{n=0}^{\infty} \sum_{\alpha \in \{1,\dots,d\}^n} \mathbb{x}^{\alpha}_{0,t} \mathbb{y}^{\alpha}_{0,s}.$$

Let $\mathcal{X}_1 := \left\{ x \in \mathscr{C}^{1-\mathrm{var}}([0,T];\mathbb{R}^d) \;\middle|\; T > 0,\ x(0) = 0 \right\}$ – and similarly $\widehat{\mathcal{X}}_1$.

Goal: Define an appropriate kernel for paths / time series.

Definition

Given $x, y \in \mathcal{X}_1$ defined on $[0,t], [0,s]$, respectively. We define

$$\mathrm{k}_{\mathrm{sig}}(x,y) := \left\langle \mathbb{x}_{0,t}^{<\infty}, \mathbb{y}_{0,s}^{<\infty} \right\rangle := \sum_{n=0}^{\infty} \sum_{\alpha \in \{1,\dots,d\}^n} \mathbb{x}_{0,t}^{\alpha} \mathbb{y}_{0,s}^{\alpha}.$$

- It is easy to see that $\left| \mathbb{x}_{[0,t]}^{=n} \right| \le \frac{\|x\|_{1-\mathrm{var}}^n}{n!}$, therefore the sum is finite.
- The definition can easily be extended to rough paths or time series – e.g., by piecewise-linear interpolation.
- Extension: For a kernel $\kappa : \mathbb{R}^d \times \mathbb{R}^d \to \mathbb{R}$, first lift $t \mapsto x(t) \to \kappa_x := t \mapsto \kappa(x(t), \cdot) \in \mathcal{H}$, then compute the signature kernel of the lifted path $\mathrm{k}_{\mathrm{sig}}(\kappa_x, \kappa_y)$.

Direct computation is impossible, due to the exponential growth of the signature – recall that $\mathbb{x}^{=n} \in (\mathbb{R}^d)^{\otimes n}$, i.e., has d^n terms. However, a recursive construction exists – comparable to the Horner scheme for polynomials. Even more powerful:

Theorem [Salvi et al., '21]

Assume that $x, y \in C^1$, and let $K_{x,y}(u, v) := \mathrm{k}_{\mathrm{sig}}\left(x|_{[0,u]}, y|_{[0,v]}\right)$ for $u \in [0, t]$, $v \in [0, s]$. Then $K_{x,y}$ solves the PDE

$$\frac{\partial^2}{\partial u \partial v} K_{x,y}(u, v) = \langle \dot{x}(u), \dot{y}(v) \rangle K_{x,y}(u, v), \quad K_{x,y}(0, \cdot) = K_{x,y}(\cdot, 0) = 1.$$

$$\mathrm{MMD}_{\mathsf{sig}}(\mu,\nu) := \Big[\int_{\mathcal{X}_1\times\mathcal{X}_1} \mathsf{k}_{\mathsf{sig}}(x,x')\mu(\mathrm{d}x)\mu(\mathrm{d}x') + \int_{\mathcal{X}_1\times\mathcal{X}_1} \mathsf{k}_{\mathsf{sig}}(y,y')\nu(\mathrm{d}y)\nu(\mathrm{d}y') - 2\int_{\mathcal{X}_1\times\mathcal{X}_1} \mathsf{k}_{\mathsf{sig}}(x,y)\mu(\mathrm{d}x)\nu(\mathrm{d}y)\Big]^{1/2}$$

- Given $\mathcal{K} \subset \widehat{\mathcal{X}_1}$ compact, then $\mathrm{MMD}_{\mathsf{sig}}$ is characteristic for $\mathcal{P}_1(\mathcal{K})$, the probability measures supported on $\mathcal{K}$, i.e., $\mathrm{MMD}_{\mathsf{sig}}(\mu,\nu) = 0 \iff \mu = \nu$.

$$\mathrm{MMD}_{\mathsf{sig}}(\mu,\nu) := \Big[\int_{\mathcal{X}_1\times\mathcal{X}_1} \mathsf{k}_{\mathsf{sig}}(x,x')\mu(\mathrm{d}x)\mu(\mathrm{d}x') + \int_{\mathcal{X}_1\times\mathcal{X}_1} \mathsf{k}_{\mathsf{sig}}(y,y')\nu(\mathrm{d}y)\nu(\mathrm{d}y') - 2\int_{\mathcal{X}_1\times\mathcal{X}_1} \mathsf{k}_{\mathsf{sig}}(x,y)\mu(\mathrm{d}x)\nu(\mathrm{d}y)\Big]^{1/2}$$

- Given $\mathcal{K}\subset\widehat{\mathcal{X}_1}$ compact, then $\mathrm{MMD}_{\mathsf{sig}}$ is characteristic for $\mathcal{P}_1(\mathcal{K})$, the probability measures supported on $\mathcal{K}$, i.e., $\mathrm{MMD}_{\mathsf{sig}}(\mu,\nu)=0 \iff \mu=\nu$.
- In the compact case, $\mathrm{MMD}_{\mathsf{sig}}$ is a metric for weak convergence.

$$\mathrm{MMD}_{\mathsf{sig}}(\mu,\nu) := \Big[\int_{\mathcal{X}_1\times\mathcal{X}_1} \mathsf{k}_{\mathsf{sig}}(x,x')\mu(\mathrm{d}x)\mu(\mathrm{d}x') + \int_{\mathcal{X}_1\times\mathcal{X}_1} \mathsf{k}_{\mathsf{sig}}(y,y')\nu(\mathrm{d}y)\nu(\mathrm{d}y') - 2\int_{\mathcal{X}_1\times\mathcal{X}_1} \mathsf{k}_{\mathsf{sig}}(x,y)\mu(\mathrm{d}x)\nu(\mathrm{d}y)\Big]^{1/2}$$

- Given $\mathcal{K} \subset \widehat{\mathcal{X}_1}$ compact, then $\mathrm{MMD}_{\mathsf{sig}}$ is characteristic for $\mathcal{P}_1(\mathcal{K})$, the probability measures supported on $\mathcal{K}$, i.e., $\mathrm{MMD}_{\mathsf{sig}}(\mu,\nu) = 0 \iff \mu = \nu$.
- In the compact case, $\mathrm{MMD}_{\mathsf{sig}}$ is a metric for weak convergence.
- For $\mathcal{P}_1(\widehat{\mathcal{X}_1})$ we obtain a metric by switching to normalized signatures, as discussed earlier. However, convergence under $\mathrm{MMD}_{\mathsf{sig}}$ does not imply weak convergence.

- B. E. Boser, I. M. Guyon, V. N. Vapnik. A training algorithm for optimal margin classifiers, Proceedings of the fifth annual workshop on Computational learning theory, 144–152. 1992.
- B. Hambly, T. Lyons. Uniqueness for the signature of a path of bounded variation and the reduced path group, Annals of Mathematics 71(1): 109–167, 2010.
- D. Lee, H. Oberhauser. The Signature Kernel, arXiv preprint arXiv:2305.04625, 2023.

Figure: Vladimir Vapnik

- K. Muandet, K. Fukumizu, B. Sriperumbudur, B. Schölkopf. Kernel mean embedding of distributions: A review and beyond, Foundations and Trends® in Machine Learning 10(1-2):1–141, 2017.

Outline

1 Path signatures

2 Rough Paths

3 Universality and the signature kernel

4 **Signature based representations for optimal stopping**

Setting

Given a d-dimensional stochastic process $(X_t)_{t\in[0,T]}$ controlled by α. Goal: maximize some reward function.

Markovian case: If X is a Markov process, the optimal control satisfies $\alpha_t^* = \alpha^*(t, X_t)$.

Popular methods include:

- Solving the (deterministic) Hamilton–Jacobi–Belman PDE for the value function.
- Approximate α^* in some parametric class of functions on $\mathbb{R}^d$ and optimize the reward.
- Least squares Monte Carlo, involving computations of conditional expectations $E[V_{t+\Delta t} \mid X_t]$.

Setting

Given a d-dimensional stochastic process $(X_t)_{t\in[0,T]}$ controlled by α. Goal: maximize some reward function.

Markovian case: If X is a Markov process, the optimal control satisfies $\alpha_t^* = \alpha^*(t, X_t)$. Popular methods include:

- Solving the (deterministic) Hamilton–Jacobi–Belman PDE for the value function.
- Approximate α^* in some parametric class of functions on $\mathbb{R}^d$ and optimize the reward.
- Least squares Monte Carlo, involving computations of conditional expectations $E[V_{t+\Delta t} \mid X_t]$.

Non-Markovian case: Now we can only expect α_t^* to be $\mathcal{F}_t$-measurable, i.e., $\alpha_t^* = \alpha^*(t, (X_s)_{s\le t})$. For all methods above, we are left with approximations in spaces of functions of paths.

Following [Kalsi, Lyons, Perez Arribas '20], a general recipe for solving stochastic optimal control problems using path signatures can be described as follows:

1. Assume that controls α_t are continuous functions $\phi(\widehat{X}|_{[0,t]})$ of the path and, hence, of the signature $\theta(\widehat{\mathbb{X}}^{<\infty}_{0,t})$ – and similarly for the loss function $L_\theta(\widehat{\mathbb{X}}^{<\infty}_{0,T})$.

Following [Kalsi, Lyons, Perez Arribas '20], a general recipe for solving stochastic optimal control problems using path signatures can be described as follows:

1. Assume that controls α_t are continuous functions $\phi(\widehat{X}|_{[0,t]})$ of the path and, hence, of the signature $\theta(\widehat{\mathbb{X}}^{<\infty}_{0,t})$ – and similarly for the loss function $L_\theta(\widehat{\mathbb{X}}^{<\infty}_{0,T})$.
2. As continuous functions, $\alpha_t = \theta(\widehat{\mathbb{X}}^{<\infty}_{0,t}) \approx \left\langle \ell_\theta, \widehat{\mathbb{X}}^{<\infty}_{0,t} \right\rangle$, $L_\theta(\widehat{\mathbb{X}}^{<\infty}_{0,T}) \approx \left\langle f_L(\ell_\theta), \widehat{\mathbb{X}}^{<\infty}_{0,T} \right\rangle$ for some $\ell_\theta, f_L(\ell_\theta) \in \mathcal{W}_d$ – by universality.

Following [Kalsi, Lyons, Perez Arribas '20], a general recipe for solving stochastic optimal control problems using path signatures can be described as follows:

1. Assume that controls α_t are continuous functions $\phi(\widehat{X}|_{[0,t]})$ of the path and, hence, of the signature $\theta(\widehat{\mathbb{X}}_{0,t}^{<\infty})$ – and similarly for the loss function $L_\theta(\widehat{\mathbb{X}}_{0,T}^{<\infty})$.
2. As continuous functions, $\alpha_t = \theta(\widehat{\mathbb{X}}_{0,t}^{<\infty}) \approx \left\langle \ell_\theta, \widehat{\mathbb{X}}_{0,t}^{<\infty} \right\rangle$, $L_\theta(\widehat{\mathbb{X}}_{0,T}^{<\infty}) \approx \left\langle f_L(\ell_\theta), \widehat{\mathbb{X}}_{0,T}^{<\infty} \right\rangle$ for some $\ell_\theta, f_L(\ell_\theta) \in \mathcal{W}_d$ – by universality.
3. Interchange expectation and truncate the signature at level N:
$E\left[L_\theta(\widehat{\mathbb{X}}_{0,T}^{<\infty})\right] \approx E\left[\left\langle f_L(\ell_\theta), \widehat{\mathbb{X}}_{0,T}^{<\infty} \right\rangle\right] = \left\langle f_L(\ell_\theta), \mathbb{E}\left[\widehat{\mathbb{X}}_{0,t}^{<\infty}\right] \right\rangle \approx \left\langle f_L(\ell_\theta), \mathbb{E}\left[\widehat{\mathbb{X}}_{0,t}^{\leq N}\right] \right\rangle$.

Following [Kalsi, Lyons, Perez Arribas '20], a general recipe for solving stochastic optimal control problems using path signatures can be described as follows:

1. Assume that controls α_t are continuous functions $\phi(\widehat{X}|_{[0,t]})$ of the path and, hence, of the signature $\theta(\widehat{\mathbb{X}}_{0,t}^{<\infty})$ – and similarly for the loss function $L_\theta(\widehat{\mathbb{X}}_{0,T}^{<\infty})$.
2. As continuous functions, $\alpha_t = \theta(\widehat{\mathbb{X}}_{0,t}^{<\infty}) \approx \left\langle \ell_\theta, \widehat{\mathbb{X}}_{0,t}^{<\infty} \right\rangle$, $L_\theta(\widehat{\mathbb{X}}_{0,T}^{<\infty}) \approx \left\langle f_L(\ell_\theta), \widehat{\mathbb{X}}_{0,T}^{<\infty} \right\rangle$ for some $\ell_\theta, f_L(\ell_\theta) \in \mathcal{W}_d$ – by universality.
3. Interchange expectation and truncate the signature at level N:
$E\left[L_\theta(\widehat{\mathbb{X}}_{0,T}^{<\infty})\right] \approx E\left[\left\langle f_L(\ell_\theta), \widehat{\mathbb{X}}_{0,T}^{<\infty} \right\rangle\right] = \left\langle f_L(\ell_\theta), \mathbb{E}\left[\widehat{\mathbb{X}}_{0,t}^{<\infty}\right]\right\rangle \approx \left\langle f_L(\ell_\theta), \mathbb{E}\left[\widehat{\mathbb{X}}_{0,t}^{\leq N}\right]\right\rangle$.
4. Optimize $\ell_\theta \mapsto \left\langle f_L(\ell_\theta), \mathbb{E}\left[\widehat{\mathbb{X}}_{0,t}^{\leq N}\right]\right\rangle$ – a fully deterministic optimization problem.

Following [Kalsi, Lyons, Perez Arribas '20], a general recipe for solving stochastic optimal control problems using path signatures can be described as follows:

1. Assume that controls α_t are continuous functions $\phi(\widehat{X}|_{[0,t]})$ of the path and, hence, of the signature $\theta(\widehat{\mathbb{X}}^{<\infty}_{0,t})$ – and similarly for the loss function $L_\theta(\widehat{\mathbb{X}}^{<\infty}_{0,T})$.
2. As continuous functions, $\alpha_t = \theta(\widehat{\mathbb{X}}^{<\infty}_{0,t}) \approx \left\langle \ell_\theta, \widehat{\mathbb{X}}^{<\infty}_{0,t} \right\rangle$, $L_\theta(\widehat{\mathbb{X}}^{<\infty}_{0,T}) \approx \left\langle f_L(\ell_\theta), \widehat{\mathbb{X}}^{<\infty}_{0,T} \right\rangle$ for some $\ell_\theta, f_L(\ell_\theta) \in \mathcal{W}_d$ – by universality.
3. Interchange expectation and truncate the signature at level N:
$E\left[L_\theta(\widehat{\mathbb{X}}^{<\infty}_{0,T})\right] \approx E\left[\left\langle f_L(\ell_\theta), \widehat{\mathbb{X}}^{<\infty}_{0,T} \right\rangle\right] = \left\langle f_L(\ell_\theta), \mathbb{E}\left[\widehat{\mathbb{X}}^{<\infty}_{0,t}\right] \right\rangle \approx \left\langle f_L(\ell_\theta), \mathbb{E}\left[\widehat{\mathbb{X}}^{\leq N}_{0,t}\right] \right\rangle$.
4. Optimize $\ell_\theta \mapsto \left\langle f_L(\ell_\theta), \mathbb{E}\left[\widehat{\mathbb{X}}^{\leq N}_{0,t}\right] \right\rangle$ – a fully deterministic optimization problem.

No convergence result known so far, but pathwise density for steps 1 + 2 with high probability is proved in [Kalsi, Lyons, Perez Arribas '20]. Problem: discontinuity of (optimal) controls.

Optimal stopping problem

Given a stochastic reward process $(Y_t)_{t\in[0,T]}$ adapted to a filtration $(\mathcal{F}_t)_{t\in[0,T]}$ generated by a d-dimensional stochastic process $(X_t)_{t\in[0,T]}$. Let $\mathcal{S}$ denote the set of $(\mathcal{F}_t)_{t\in[0,T]}$–stopping times. Compute $\sup_{\tau\in\mathcal{S}} \mathbb{E}\left[Y_\tau\right]$.

Optimal stopping problem

Given a stochastic reward process $(Y_t)_{t\in[0,T]}$ adapted to a filtration $(\mathcal{F}_t)_{t\in[0,T]}$ generated by a d-dimensional stochastic process $(X_t)_{t\in[0,T]}$. Let $\mathcal{S}$ denote the set of $(\mathcal{F}_t)_{t\in[0,T]}$–stopping times. Compute $\sup_{\tau\in\mathcal{S}} \mathbb{E}[Y_\tau]$.

- Optimal stopping times are generally hitting times of sets, hence discontinuous functions on path-space.

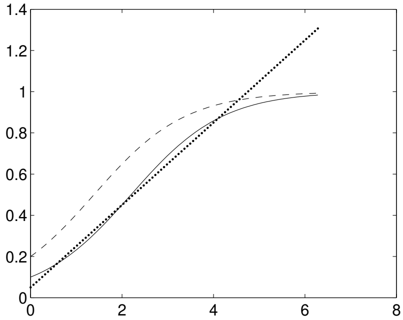


Figure: Discontinuity of hitting times

Optimal stopping problem

Given a stochastic reward process $(Y_t)_{t\in[0,T]}$ adapted to a filtration $(\mathcal{F}_t)_{t\in[0,T]}$ generated by a d-dimensional stochastic process $(X_t)_{t\in[0,T]}$. Let $\mathcal{S}$ denote the set of $(\mathcal{F}_t)_{t\in[0,T]}$–stopping times. Compute $\sup_{\tau\in\mathcal{S}} \mathbb{E}\left[Y_\tau\right]$.

- Optimal stopping times are generally hitting times of sets, hence discontinuous functions on path-space.
- **Example:** X models a stock price – possibly with additional factors such as stochastic volatilities – and $Y_t = h(X_t)$ for some payoff function h.
- **Example:** $X = Y = W^H$ . . . fractional Brownian motion

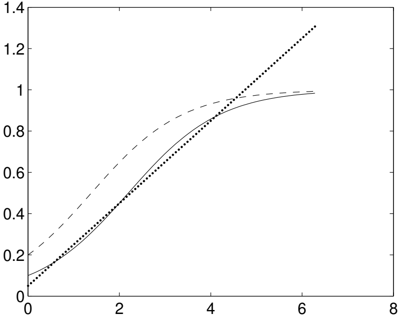


Figure: Discontinuity of hitting times

[Becker, Cheredito, Jentzen '19] consider the optimal stopping problem for fractional Brownian motion. In the general setting, their strategy is as follows:

1. Fix a time-grid $0 = t_0 < \cdots < t_J = T$ and define a (discrete time) $(J+1)d$-dimensional Markov process $(Z_j)_{j=0}^J$ by

$$\begin{aligned} Z_0 &:= (X_{t_0}, 0, \ldots, 0), \\ Z_1 &:= (X_{t_0}, X_{t_1}, 0, \ldots, 0), \\ Z_2 &:= (X_{t_0}, X_{t_1}, X_{t_2}, 0, \ldots, 0), \\ &\vdots \end{aligned}$$

[Becker, Cheredito, Jentzen '19] consider the optimal stopping problem for fractional Brownian motion. In the general setting, their strategy is as follows:

1. Fix a time-grid $0 = t_0 < \cdots < t_J = T$ and define a (discrete time) $(J+1)d$-dimensional Markov process $(Z_j)_{j=0}^J$ by
$$\begin{aligned} Z_0 &:= (X_{t_0}, 0, \dots, 0), \\ Z_1 &:= (X_{t_0}, X_{t_1}, 0, \dots, 0), \\ Z_2 &:= (X_{t_0}, X_{t_1}, X_{t_2}, 0, \dots, 0), \\ &\vdots \end{aligned}$$
2. Solve the discrete-time Markovian optimal stopping problem. [Becker, Cheredito, Jentzen '19] use deep neural networks to approximate stopping decisions $f_j(Z_j) \approx \mathrm{DNN}_j(Z_j; \theta)$ – "stop at time t_j unless stopped earlier".

How can we construct stopping times and adapted processes using rough paths?

Stopped rough paths

Let $\widehat{\Omega}_t^p := \left\{ \mathbf{x} \in \mathscr{C}_g^{1/p}([0,t];\mathbb{R}^{1+d}) \;\middle|\; x^1(s) = s \right\}$. The space of stopped rough paths is defined as $\Lambda_T := \bigcup_{t\in[0,T]} \widehat{\Omega}_t^p$.

How can we construct stopping times and adapted processes using rough paths?

Stopped rough paths

Let $\widehat{\Omega}^p_t := \left\{ \mathbf{x} \in \mathscr{C}_g^{1/p}([0,t];\mathbb{R}^{1+d}) \,\middle|\, x^1(s) = s \right\}$. The space of stopped rough paths is defined as $\Lambda_T := \bigcup_{t\in[0,T]} \widehat{\Omega}^p_t$.

- Λ_T is a Polish space with a Dupire type metric.
- We can now define adapted processes or stopping times as functionals on Λ_T.

How can we construct stopping times and adapted processes using rough paths?

Stopped rough paths

Let $\widehat{\Omega}_t^p := \left\{ \mathbf{x} \in \mathscr{C}_g^{1/p}([0,t];\mathbb{R}^{1+d}) \,\middle|\, x^1(s) = s \right\}$. The space of stopped rough paths is defined as $\Lambda_T := \bigcup_{t\in[0,T]} \widehat{\Omega}_t^p$.

- Λ_T is a Polish space with a Dupire type metric.
- We can now define adapted processes or stopping times as functionals on Λ_T.

Rough stochastic processes

Given a probability space $(\Omega, \mathcal{F}, \mathbb{P})$, a rough stochastic process is a random variable $\widehat{\mathbf{X}}$ taking values in $\widehat{\Omega}_T^p$. We further define the natural filtration generated by $\widehat{\mathbf{X}}$, i.e., $\mathcal{F}_t := \sigma\left(\mathbf{X}_{0,s} : 0 \le s \le t\right)$.

Given $\ell \in \mathcal{W}_{1+d}$, define a signature stopping rule $\tau_\ell \in \mathcal{S}$ as

$$\tau_\ell := \inf\left\{ t \in [0,T] \,\middle|\, \left\langle \ell, \widehat{\mathbb{X}}^{<\infty}_{0,t} \right\rangle \geq 1 \right\}.$$

Note that τ_ℓ is the first hitting time of a hyperplane in $T((\mathbb{R}^{1+d}))$.

Given $\ell \in \mathcal{W}_{1+d}$, define a signature stopping rule $\tau_\ell \in \mathcal{S}$ as

$$\tau_\ell := \inf\left\{ t \in [0,T] \,\middle|\, \left\langle \ell, \widehat{\mathbb{X}}^{<\infty}_{0,t} \right\rangle \geq 1 \right\}.$$

Note that τ_ℓ is the first hitting time of a hyperplane in $T((\mathbb{R}^{1+d}))$.

Theorem

Given an $(\mathcal{F}_t)_{t\in[0,T]}$–adapted continuous reward process $(Y_t)_{t\in[0,T]}$ with $\mathbb{E}\,\|Y\|_\infty < \infty$, then

$$\sup_{\tau\in\mathcal{S}} \mathbb{E}\left[Y_{\tau\wedge T}\right] = \sup_{\ell\in\mathcal{W}_{1+d}} \mathbb{E}\left[Y_{\tau_\ell\wedge T}\right].$$

Given $\ell \in \mathcal{W}_{1+d}$, define a signature stopping rule $\tau_\ell \in \mathcal{S}$ as

$$\tau_\ell := \inf \left\{ t \in [0, T] \,\middle|\, \left\langle \ell, \widehat{\mathbb{X}}^{<\infty}_{0,t} \right\rangle \geq 1 \right\}.$$

Note that τ_ℓ is the first hitting time of a hyperplane in $T((\mathbb{R}^{1+d}))$.

Theorem

Given an $(\mathcal{F}_t)_{t\in[0,T]}$–adapted continuous reward process $(Y_t)_{t\in[0,T]}$ with $\mathbb{E}\,\|Y\|_\infty < \infty$, then

$$\sup_{\tau\in\mathcal{S}} \mathbb{E}\left[Y_{\tau\wedge T}\right] = \sup_{\ell\in\mathcal{W}_{1+d}} \mathbb{E}\left[Y_{\tau_\ell\wedge T}\right].$$

- While optimal stopping times $\tau^* \in \mathcal{S}$ typically exist, we do not expect an optimizer $\ell^* \in \mathcal{W}_{1+d}$ to exist.

Given $\theta \in C(\Lambda_T, \mathbb{R})$ define a continuous stopping rule by

$$\tau_\theta := \inf\left\{ t \in [0,T] \,\middle|\, \int_0^t \theta(\widehat{\mathbf{X}}|_{[0,s]})^2 \mathrm{d}s \geq 1 \right\}.$$

Lemma

$$\sup_{\theta \in C(\Lambda_T, \mathbb{R})} \mathbb{E}\left[Y_{\tau_\theta \wedge T}\right] = \sup_{\tau \in \mathcal{S}} \mathbb{E}\left[Y_{\tau \wedge T}\right]$$

Proof of the Lemma is based on approximation of measurable by continuous functions.

- If a continuous stopping rule τ_θ was continuous as a function of the signature, we could approximate it by signature stopping rules:

$$\inf\left\{ t \in [0,T] \,\middle|\, \int_0^t \theta(\widehat{\mathbf{X}}|_{[0,s]})^2 \mathrm{d}s \geq 1 \right\} \approx \inf\left\{ t \in [0,T] \,\middle|\, \left\langle \ell, \widehat{\mathbb{X}}^{<\infty}_{0,t} \right\rangle \geq 1 \right\}.$$

- Unfortunately, this is just not the case.

- If a continuous stopping rule τ_θ was continuous as a function of the signature, we could approximate it by signature stopping rules:

$$\inf\left\{ t \in [0,T] \,\middle|\, \int_0^t \theta(\widehat{\mathbf{X}}|_{[0,s]})^2 \mathrm{d}s \geq 1 \right\} \approx \inf\left\{ t \in [0,T] \,\middle|\, \left\langle \ell, \widehat{\mathbb{X}}_{0,t}^{<\infty} \right\rangle \geq 1 \right\}.$$

- Unfortunately, this is just not the case.
- **Randomization:** Replace the fixed level 1 above by an (independent) random level Z.
- Interpretation: If $Z \sim \mathrm{Exp}(1)$, stop at the first jump time of a pure jump process with intensity $\theta(\widehat{\mathbf{X}}|_{[0,s]})^2$.

- If a continuous stopping rule τ_θ was continuous as a function of the signature, we could approximate it by signature stopping rules:

$$\inf\left\{ t \in [0,T] \,\middle|\, \int_0^t \theta(\widehat{\mathbf{X}}|_{[0,s]})^2 \mathrm{d}s \geq 1 \right\} \approx \inf\left\{ t \in [0,T] \,\middle|\, \left\langle \ell, \widehat{\mathbb{X}}^{<\infty}_{0,t} \right\rangle \geq 1 \right\}.$$

- Unfortunately, this is just not the case.
- **Randomization:** Replace the fixed level 1 above by an (independent) random level Z.
- Interpretation: If $Z \sim \mathrm{Exp}(1)$, stop at the first jump time of a pure jump process with intensity $\theta(\widehat{\mathbf{X}}|_{[0,s]})^2$.

Let $Z \geq 0$ be a r.v. independent of $\widehat{\mathbb{X}}$ with (smooth) c.d.f. F_Z.

$$\tau^r_\theta := \inf\left\{ t \in [0,T] \,\middle|\, \int_0^t \theta\left(\widehat{\mathbb{X}}|_{[0,s]}\right)^2 \mathrm{d}s \geq Z \right\}, \quad \tau^r_\ell := \inf\left\{ t \in [0,T] \,\middle|\, \int_0^t \left\langle \ell, \widehat{\mathbb{X}}^{<\infty}_{0,t} \right\rangle^2 \mathrm{d}s \geq Z \right\}.$$

Lemma

$$\sup_{\theta \in C(\Lambda_T, \mathbb{R})} \mathbb{E}\left[Y_{\tau^r_\theta \wedge T}\right] = \sup_{\theta \in C(\Lambda_T, \mathbb{R})} \mathbb{E}\left[Y_{\tau_\theta \wedge T}\right], \quad \sup_{\ell \in \mathcal{W}_{1+d}} \mathbb{E}\left[Y_{\tau^r_\ell \wedge T}\right] = \sup_{\ell \in \mathcal{W}_{1+d}} \mathbb{E}\left[Y_{\tau_\ell \wedge T}\right].$$

Proof: Formal proof by dominated convergence. Informally: The buyer of an American option may very well randomize her exercise decision.

Lemma

$$\sup_{\theta \in C(\Lambda_T, \mathbb{R})} \mathbb{E}\left[Y_{\tau_\theta^r \wedge T}\right] = \sup_{\theta \in C(\Lambda_T, \mathbb{R})} \mathbb{E}\left[Y_{\tau_\theta \wedge T}\right], \quad \sup_{\ell \in \mathcal{W}_{1+d}} \mathbb{E}\left[Y_{\tau_\ell^r \wedge T}\right] = \sup_{\ell \in \mathcal{W}_{1+d}} \mathbb{E}\left[Y_{\tau_\ell \wedge T}\right].$$

Proof: Formal proof by dominated convergence. Informally: The buyer of an American option may very well randomize her exercise decision.

Lemma (Regularization by randomization)

Let $\widetilde{F}(t) := F_Z\left(\int_0^t \theta\left(\widehat{\mathbf{X}}|_{[0,s]}\right) \mathrm{d}s\right)$, then $\mathbb{E}\left[Y_{\tau_\theta^r \wedge T} \mid \widehat{\mathbf{X}}\right] = \int_0^T Y_t \mathrm{d}\widetilde{F}(t) + Y_T(1 - \widetilde{F}(T))$.

- Note that the R.H.S. is a smooth function of $\widehat{\mathbf{X}}$.

Lemma

For every $\varepsilon > 0$ there is a compact set $\mathcal{K} \subset \widehat{\Omega}_T^p$ s.t. $\mathbb{P}(\mathbf{X} \in \mathcal{K}) > 1 - \varepsilon$ and for every $\theta \in C(\Lambda_T, \mathbb{R})$ there is a sequence $\ell_n \in \mathcal{W}_{1+d}$ s.t.

$$\sup_{\mathbf{x} \in \mathcal{K};\ t \in [0,T]} \left| \theta(\widehat{\mathbf{x}}|_{[0,t]}) - \left\langle \ell_n, \mathbb{x}_{0,t}^{<\infty} \right\rangle \right| \xrightarrow{n \to \infty} 0.$$

The above Stone–Weierstrass theorem implies that (randomized) continuous stopping rules can be approximated by (randomized) signature stopping rules, given that

$$\mathbb{E}[Y_\tau] \leq \mathbb{E}\left[\|Y\|_\infty \right] < \infty.$$

Let, for simplicity, $Z \sim \mathrm{Exp}(1)$. Then we end up with

$$\sup_{\tau \in \mathcal{S}} \mathbb{E}\left[Y_{\tau \wedge T}\right] = Y_0 + \sup_{\ell \in \mathcal{W}_{d+1}} \mathbb{E}\left[\int_0^T \exp\left(-\int_0^t \left\langle \ell, \widehat{\mathbb{X}}^{<\infty}_{0,s} \right\rangle^2 \mathrm{d}s\right) \mathrm{d}Y_t\right].$$

Let, for simplicity, $Z \sim \mathrm{Exp}(1)$. Then we end up with

$$\sup_{\tau \in \mathcal{S}} \mathbb{E}\left[Y_{\tau \wedge T}\right] = Y_0 + \sup_{\ell \in \mathcal{W}_{d+1}} \mathbb{E}\left[\int_0^T \exp\left(-\int_0^t \left\langle \ell, \widehat{\mathbb{X}}^{<\infty}_{0,s} \right\rangle^2 \mathrm{d}s\right) \mathrm{d}Y_t\right].$$

- Recalling that $\widehat{X}_s = (s, X_s)$, we have

$$\int_0^t \left\langle \ell, \widehat{\mathbb{X}}^{<\infty}_{0,s} \right\rangle^2 \mathrm{d}s = \int_0^t \left\langle \ell \sqcup\!\sqcup\, \ell, \widehat{\mathbb{X}}^{<\infty}_{0,s} \right\rangle \mathrm{d}s = \left\langle (\ell \sqcup\!\sqcup\, \ell)\mathbf{1}, \widehat{\mathbb{X}}^{<\infty}_{0,t} \right\rangle$$

Let, for simplicity, $Z \sim \mathrm{Exp}(1)$. Then we end up with

$$\sup_{\tau \in \mathcal{S}} \mathbb{E}\left[Y_{\tau \wedge T}\right] = Y_0 + \sup_{\ell \in \mathcal{W}_{d+1}} \mathbb{E}\left[\int_0^T \exp\left(-\int_0^t \left\langle \ell, \widehat{\mathbb{X}}^{<\infty}_{0,s} \right\rangle^2 \mathrm{d}s\right) \mathrm{d}Y_t\right].$$

- Recalling that $\widehat{X}_s = (s, X_s)$, we have

$$\int_0^t \left\langle \ell, \widehat{\mathbb{X}}^{<\infty}_{0,s} \right\rangle^2 \mathrm{d}s = \int_0^t \left\langle \ell \sqcup\!\sqcup \ell, \widehat{\mathbb{X}}^{<\infty}_{0,s} \right\rangle \mathrm{d}s = \left\langle (\ell \sqcup\!\sqcup \ell)\mathbf{1}, \widehat{\mathbb{X}}^{<\infty}_{0,t} \right\rangle$$

- Approximate exp by polynomials, giving the exponential shuffle $\exp^{\sqcup\!\sqcup}(\ell) := \sum_{n=0}^{\infty} \frac{1}{n!} \ell^{\sqcup\!\sqcup n}$.
- Often, Y can also be approximated by a linear functional on $\widehat{\mathbb{X}}^{<\infty}$. Otherwise, consider a RP extending $t \mapsto (t, X_t, Y_t)$. E.g., in the case $d = 1$, $Y \equiv X$, we obtain

$$\mathbb{E}\left[Y_{\tau_\ell \wedge T}\right] = \left\langle \exp^{\sqcup\!\sqcup}\left(-(\ell \sqcup\!\sqcup \ell)\mathbf{1}\right)\mathbf{2}, \mathbb{E}\left[\widehat{\mathbb{X}}^{<\infty}_{0,T}\right]\right\rangle \approx \left\langle \exp^{\sqcup\!\sqcup}\left(-(\ell \sqcup\!\sqcup \ell)\mathbf{1}\right)\mathbf{2}, \mathbb{E}\left[\widehat{\mathbb{X}}^{\leq N}_{0,T}\right]\right\rangle.$$

Theorem

Let $\mathbb{E}\left[\|Y\|_\infty\right] < \infty$. *Given* $\kappa > 0$, *define the stopping time* $\sigma = \sigma_\kappa$ *by* $\sigma := \inf\left\{ t \geq 0 \,\middle|\, \|\widehat{\mathbb{X}}\|_{p-\text{var};[0,t]} \geq \kappa \right\} \wedge T$. *Then,*

$$\sup_{\tau \in \mathcal{S}} \mathbb{E}\left[Y_{\tau \wedge T}\right] = \mathbb{E}\left[Y_0\right] + \lim_{\kappa \to \infty} \lim_{K \to \infty} \lim_{N \to \infty} \sup_{|\ell| + \deg(\ell) \leq K} \mathbb{E}\left[\int_0^{\sigma_\kappa} \left\langle \exp^{\sqcup\!\sqcup}(-(\ell \sqcup\!\sqcup \ell)\mathbf{1}), \widehat{\mathbb{X}}^{\leq N}_{0,t} \right\rangle \mathrm{d}Y_t\right].$$

Theorem

Let $\mathbb{E}\left[\|Y\|_\infty\right] < \infty$. *Given* $\kappa > 0$, *define the stopping time* $\sigma = \sigma_\kappa$ *by*
$\sigma := \inf\left\{ t \geq 0 \,\middle|\, \|\widehat{\mathbb{X}}\|_{p-\mathrm{var};[0,t]} \geq \kappa \right\} \wedge T$. *Then,*

$$\sup_{\tau \in \mathcal{S}} \mathbb{E}\left[Y_{\tau\wedge T}\right] = \mathbb{E}\left[Y_0\right] + \lim_{\kappa\to\infty} \lim_{K\to\infty} \lim_{N\to\infty} \sup_{|\ell|+\deg(\ell)\leq K} \mathbb{E}\left[\int_0^{\sigma_\kappa} \left\langle \exp^{\sqcup\!\sqcup}(-(\ell \sqcup\!\sqcup \ell)1), \widehat{\mathbb{X}}^{\leq N}_{0,t} \right\rangle \mathrm{d}Y_t\right].$$

If Y *is a linear functional of* $\widehat{\mathbb{X}}^{<\infty}$, *this formula can be further simplified. E.g., if* $d = 1$ *and* $Y = X$, *then*

$$\sup_{\tau \in \mathcal{S}} \mathbb{E}\left[Y_{\tau\wedge T}\right] = \mathbb{E}\left[Y_0\right] + \lim_{\kappa\to\infty} \lim_{K\to\infty} \lim_{N\to\infty} \sup_{|\ell|+\deg(\ell)\leq K} \left\langle \exp^{\sqcup\!\sqcup}(-(\ell \sqcup\!\sqcup \ell)1)2, \mathbb{E}\left[\widehat{\mathbb{X}}^{\leq N}_{0,\sigma_\kappa}\right] \right\rangle.$$

1. Optimal stopping of Brownian motion X: By Fawcett's formula,

$$\mathbb{E}\left[\widehat{\mathbb{X}}_{0,T}^{<\infty}\right] = \exp\left(T\left(e_1 + \frac{1}{2}e_2 \otimes e_2\right)\right).$$

We immediately see that $\left\langle \exp^{\sqcup\!\sqcup}(-(\ell \sqcup\!\sqcup \ell)\mathbf{1})\mathbf{2}, \mathbb{E}\left[\widehat{\mathbb{X}}_{0,T}^{\leq N}\right]\right\rangle = 0.$

1. Optimal stopping of Brownian motion X: By Fawcett's formula,

$$\mathbb{E}\left[\widehat{\mathbb{X}}^{<\infty}_{0,T}\right] = \exp\left(T\left(e_1 + \frac{1}{2}e_2 \otimes e_2\right)\right).$$

We immediately see that $\left\langle \exp^{\sqcup\!\sqcup}(-(\ell \sqcup\!\sqcup \ell)1)2,\ \mathbb{E}\left[\widehat{\mathbb{X}}^{\leq N}_{0,T}\right]\right\rangle = 0.$

2. Obtain approximately optimal strategy, not just approximation to value function. Let $\ell^* = \ell^*_{\kappa,K,N}$ an optimizer in the theorem. Construct

$$\tau^r_{\ell^*} := \inf\left\{ t \in [0,T] \,\middle|\, \left\langle (\ell^* \sqcup\!\sqcup \ell^*)1,\ \widehat{\mathbb{X}}^{\leq N}_{0,t}\right\rangle \geq Z \right\}.$$

- $\mathbb{E}\left[Y_{\tau^r_{\ell^*}}\right] \approx \mathbb{E}[Y_0] + \left\langle \exp^{\sqcup\!\sqcup}(-(\ell^* \sqcup\!\sqcup \ell^*)1)2,\ \mathbb{E}\left[\widehat{\mathbb{X}}^{\leq N}_{0,\sigma_\kappa}\right]\right\rangle \approx \sup_{\tau\in\mathcal{S}} \mathbb{E}\left[Y_{\tau\wedge T}\right]$
- Obviously, $\mathbb{E}\left[Y_{\tau^r_{\ell^*}}\right] \leq \sup_{\tau\in\mathcal{S}} \mathbb{E}\left[Y_{\tau\wedge T}\right]$.

1. Optimal stopping of Brownian motion X: By Fawcett's formula,

$$\mathbb{E}\left[\widehat{\mathbb{X}}^{<\infty}_{0,T}\right] = \exp\left(T\left(e_1 + \frac{1}{2}e_2 \otimes e_2\right)\right).$$

We immediately see that $\left\langle \exp^{\sqcup\!\sqcup}(-(\ell \sqcup\!\sqcup \ell)1)2,\ \mathbb{E}\left[\widehat{\mathbb{X}}^{\leq N}_{0,T}\right]\right\rangle = 0$.

2. Obtain approximately optimal strategy, not just approximation to value function. Let $\ell^* = \ell^*_{\kappa,K,N}$ an optimizer in the theorem. Construct

$$\tau^r_{\ell^*} := \inf\left\{ t \in [0,T] \,\middle|\, \left\langle (\ell^* \sqcup\!\sqcup \ell^*)1,\ \widehat{\mathbb{X}}^{\leq N}_{0,t}\right\rangle \geq Z \right\}.$$

- $\mathbb{E}\left[Y_{\tau^r_{\ell^*}}\right] \approx \mathbb{E}[Y_0] + \left\langle \exp^{\sqcup\!\sqcup}(-(\ell^* \sqcup\!\sqcup \ell^*)1)2,\ \mathbb{E}\left[\widehat{\mathbb{X}}^{\leq N}_{0,\sigma_\kappa}\right]\right\rangle \approx \sup_{\tau\in\mathcal{S}} \mathbb{E}\left[Y_{\tau\wedge T}\right]$
- Obviously, $\mathbb{E}\left[Y_{\tau^r_{\ell^*}}\right] \leq \sup_{\tau\in\mathcal{S}} \mathbb{E}\left[Y_{\tau\wedge T}\right]$.

3. Dual method based on minimization of martingales.

Recall that $\mathbb{L}^{<\infty}_{s,t} := \log \mathbb{X}^{<\infty}_{s,t} \in \mathfrak{g}(\mathbb{R}^d)$ and $\mathbb{L}^{\leq N}_{s,t} := \log \mathbb{X}^{\leq N}_{s,t} \in \mathfrak{g}^N(\mathbb{R}^d)$.

Recall that $\mathbb{L}^{<\infty}_{s,t} := \log \mathbb{X}^{<\infty}_{s,t} \in \mathfrak{g}(\mathbb{R}^d)$ and $\mathbb{L}^{\leq N}_{s,t} := \log \mathbb{X}^{\leq N}_{s,t} \in \mathfrak{g}^N(\mathbb{R}^d)$.

- The log-signature $\mathbb{L}^{\leq N}_{s,t}$ contains the same information as $\mathbb{X}^{\leq N}_{s,t}$, but removes algebraic redundancies.
- No shuffle identity holds for (truncated) log-signatures, but $\dim \mathfrak{g}^N(\mathbb{R}^d) \ll \dim T^N(\mathbb{R}^d)$. E.g., for $d = 3$, $N = 6$: 196 vs 1092.

Recall that $\mathbb{L}^{<\infty}_{s,t} := \log \mathbb{X}^{<\infty}_{s,t} \in \mathfrak{g}(\mathbb{R}^d)$ and $\mathbb{L}^{\leq N}_{s,t} := \log \mathbb{X}^{\leq N}_{s,t} \in \mathfrak{g}^N(\mathbb{R}^d)$.

- The log-signature $\mathbb{L}^{\leq N}_{s,t}$ contains the same information as $\mathbb{X}^{\leq N}_{s,t}$, but removes algebraic redundancies.
- No shuffle identity holds for (truncated) log-signatures, but $\dim \mathfrak{g}^N(\mathbb{R}^d) \ll \dim T^N(\mathbb{R}^d)$. E.g., for $d = 3$, $N = 6$: 196 vs 1092.
- Use of the shuffle identity is not free, but often translated into very high degrees of truncation. E.g., suppose that $\deg = 3$ contains enough information, but a polynomial of degree 3 is to be linearized. Hence, the truncation degree $N = 9$ is required. (For $d = 3$, this leads to a dimension $\dim T^9(\mathbb{R}^3) = 29524$ – compare with $\dim T^3(\mathbb{R}^3) = 39$, $\dim \mathfrak{g}^3(\mathbb{R}^3) = 14$.)

Signatures are useful as features when their algebraic properties are efficiently used. Otherwise, log-signatures are probably preferable.

A class of fully connected Artificial Neural Networks

Given $K, q, I \in \mathbb{N}$ and an activation function φ (i.e., continuous, non-polynomial), let $\mathrm{DNN}(K, q, I; \varphi)$ denote the set of fully connected artificial neural networks with I hidden layers of dimension q, input dimension K and output dimension 1, i.e., for $\vartheta \in \mathrm{DNN}(K, q, I; \varphi)$ there are affine maps $A_0 : \mathbb{R}^K \to \mathbb{R}^q$, $A_1, \dots, A_{I-1} : \mathbb{R}^q \to \mathbb{R}^q$, $A_I : \mathbb{R}^q \to \mathbb{R}$ s.t.,

$$\vartheta = A_I \circ \varphi \circ A_{I-1} \circ \varphi \circ \cdots \circ \varphi \circ A_0.$$

A class of fully connected Artificial Neural Networks

Given $K, q, I \in \mathbb{N}$ and an activation function φ (i.e., continuous, non-polynomial), let $\mathrm{DNN}(K, q, I; \varphi)$ denote the set of fully connected artificial neural networks with I hidden layers of dimension q, input dimension K and output dimension 1, i.e., for $\vartheta \in \mathrm{DNN}(K, q, I; \varphi)$ there are affine maps $A_0 : \mathbb{R}^K \to \mathbb{R}^q$, $A_1, \ldots, A_{I-1} : \mathbb{R}^q \to \mathbb{R}^q$, $A_I : \mathbb{R}^q \to \mathbb{R}$ s.t.,

$$\vartheta = A_I \circ \varphi \circ A_{I-1} \circ \varphi \circ \cdots \circ \varphi \circ A_0.$$

Deep signature stopping rule

Given $\vartheta \in \mathrm{DNN}(K, q, I; \varphi)$ with $K = \dim \mathfrak{g}^N(\mathbb{R}^d)$ for some N, we define a deep signature stopping rule by

$$\tau_\vartheta := \inf\left\{ t \in [0, T] \,\middle|\, \int_0^t \vartheta\left(\mathbb{L}^{\leq N}_{0,s}\right)^2 \mathrm{d}s \geq 1 \right\}.$$

Let $\mathcal{T}_{\text{log}} := \bigcup_{N,q,I \in \mathbb{N}} \mathrm{DNN}\big(\dim \mathfrak{g}^N(\mathbb{R}^d), q, I; \varphi\big)$.

Theorem

If $\mathbb{E}\left[||Y||_\infty\right] < \infty$, we have

$$\sup_{\tau \in \mathcal{S}} \mathbb{E}\left[Y_{\tau \wedge T}\right] = \sup_{\vartheta \in \mathcal{T}_{\text{log}}} \mathbb{E}\left[Y_{\tau_\vartheta^r \wedge T}\right].$$

Let $\mathcal{T}_{\text{log}} := \bigcup_{N,q,I \in \mathbb{N}} \mathrm{DNN}\big(\dim \mathfrak{g}^N(\mathbb{R}^d), q, I; \varphi\big)$.

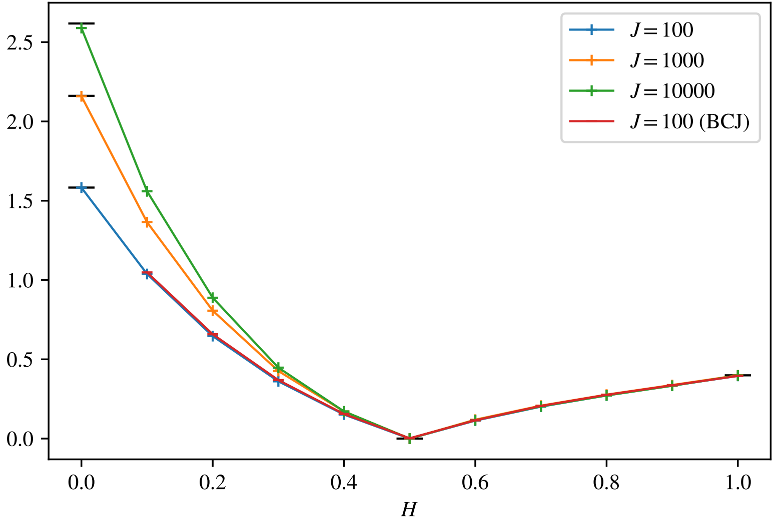
Theorem

If $\mathbb{E}\left[\|Y\|_\infty\right] < \infty$, we have

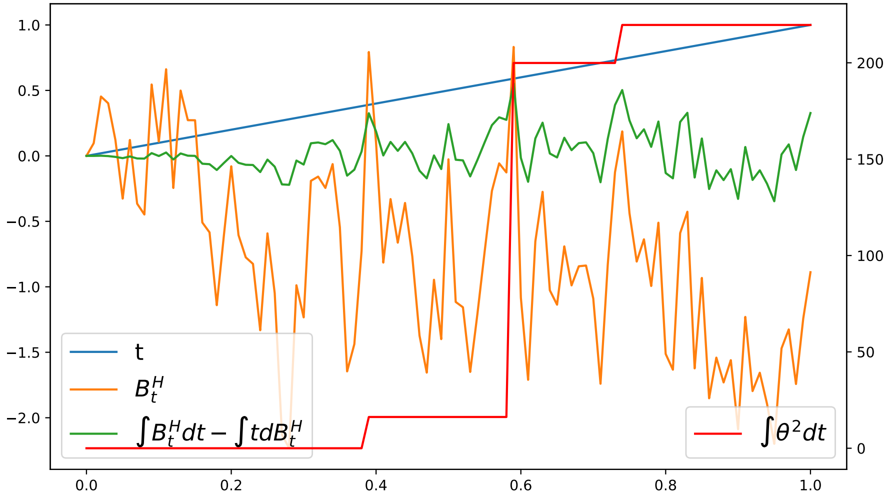
$$\sup_{\tau \in \mathcal{S}} \mathbb{E}\left[Y_{\tau \wedge T}\right] = \sup_{\vartheta \in \mathcal{T}_{\text{log}}} \mathbb{E}\left[Y_{\tau^r_\vartheta \wedge T}\right].$$

- Proof: Combination of the classical universal approximations theorem for neural networks and our earlier arguments.

Example: Optimal stopping of fractional Brownian motion $(W_t^H)_{t\in[0,1]}$ – approximate values



Example: Optimal stopping of fractional Brownian motion $(W_t^H)_{t\in[0,1]}$ – sample strategy



- C. Bayer, P. Hager, S. Riedel, J. Schoenmakers. Optimal stopping with signatures. Annals of Appl. Prob. 33(1):238–273, 2023.
- S. Becker, P. Cheridito, and A. Jentzen. Deep optimal stopping. J. Machine Learning Research, 2019.
- J. Kalsi, T. Lyons, and I. Perez Arribas. Optimal execution with rough path signatures. SIAM J. Financial Math., 2020.