Weierstrass Institute for
Applied Analysis and Stochastics

Leibniz
Association

Optimal stopping with signatures

Christian Bayer

Joint work with: P. Hager, S. Riedel, J. Schoenmakers

Quantitative Finance

Conference in honour of Michael Dempster's 85th birthday

Recent trend for using processes with memory in finance and beyond:

- Rough volatility: Model stochastic volatility by fractional Brownian motion, e.g., the rough Bergomi model:

$$dS_t = \sqrt{v_t} S_t dZ_t,$$
$$v_t = \xi(t) \exp\left(\eta \widehat{W}_t - \frac{1}{2}\eta^2 t^{2H}\right), \ \widehat{W}_t := \int_0^t K(t-s) dW_s, \ K(r) := \sqrt{2H} r^{H-\frac{1}{2}}.$$

- Order flow models by self-exciting jump processes, e.g., Hawkes processes.
- Statistical mechanics models based on Generalized Langevin Equations.

Recent trend for using processes with memory in finance and beyond:

- Rough volatility: Model stochastic volatility by fractional Brownian motion, e.g., the rough Bergomi model:

$$\begin{aligned} dS_t &= \sqrt{v_t} S_t dZ_t, \\ v_t &= \xi(t) \exp\left(\eta \widehat{W}_t - \frac{1}{2}\eta^2 t^{2H}\right), \ \widehat{W}_t := \int_0^t K(t-s) dW_s, \ K(r) := \sqrt{2H} r^{H-\frac{1}{2}}. \end{aligned}$$

- Order flow models by self-exciting jump processes, e.g., Hawkes processes.
- Statistical mechanics models based on Generalized Langevin Equations.

Many numerical methods rely on the Markov property: (pricing) PDEs, polynomial regression methods, dynamic programming, ...

- Given a smooth path $X : [0,T] \to \mathbb{R}^d$, i.e., a continuous path of bounded variation. W.l.o.g., $X(0) = 0$.
- For a word $\alpha = \mathtt{i}_1 \cdots \mathtt{i}_n$, $\mathtt{i}_j \in \{ \mathtt{1}, \dots, \mathtt{d} \}$, set the iterated integral

$$X_{s,t}^{\mathtt{i}_1 \cdots \mathtt{i}_n} := \int_{s<t_1<\cdots<t_n<t} \mathrm{d}X^{\mathtt{i}_1}(t_1) \cdots \mathrm{d}X^{\mathtt{i}_n}(t_n), \quad X_{s,t}^{\varnothing} := 1.$$

- Given a smooth path $X : [0,T] \to \mathbb{R}^d$, i.e., a continuous path of bounded variation. W.l.o.g., $X(0) = 0$.
- For a word $\alpha = \mathtt{i}_1 \cdots \mathtt{i}_n$, $\mathtt{i}_j \in \{\mathtt{1}, \dots, \mathtt{d}\}$, set the iterated integral

$$X_{s,t}^{\mathtt{i}_1\cdots\mathtt{i}_n} := \int_{s<t_1<\cdots<t_n<t} \mathrm{d}X^{\mathtt{i}_1}(t_1)\cdots \mathrm{d}X^{\mathtt{i}_n}(t_n), \quad X_{s,t}^{\varnothing} := 1.$$

- The signature is the collection of all iterated integrals

$$\mathbb{X}_{s,t}^{<\infty} := \sum_{n=0}^{\infty} \sum_{\mathtt{i}_1,\dots,\mathtt{i}_n \in \{\mathtt{1},\dots,\mathtt{d}\}} X_{s,t}^{\mathtt{i}_1\cdots\mathtt{i}_n} e_{\mathtt{i}_1} \otimes \cdots \otimes e_{\mathtt{i}_n} \in T((\mathbb{R}^d)) := \prod_{n=0}^{\infty} (\mathbb{R}^d)^{\otimes n}.$$

- Also define the truncated signature

$$\mathbb{X}_{s,t}^{\leq N} := \sum_{n=0}^{N} \sum_{\mathtt{i}_1,\dots,\mathtt{i}_n \in \{\mathtt{1},\dots,\mathtt{d}\}} X_{s,t}^{\mathtt{i}_1\cdots\mathtt{i}_n} e_{\mathtt{i}_1} \otimes \cdots \otimes e_{\mathtt{i}_n} \in T^N(\mathbb{R}^d) := \prod_{n=0}^{N} (\mathbb{R}^d)^{\otimes n}$$

- $T((\mathbb{R}^d))$ is an algebra: with $\mathbf{a} = (a_n)_{n=0}^{\infty}$, $\mathbf{b} = (b_n)_{n=0}^{\infty}$ set

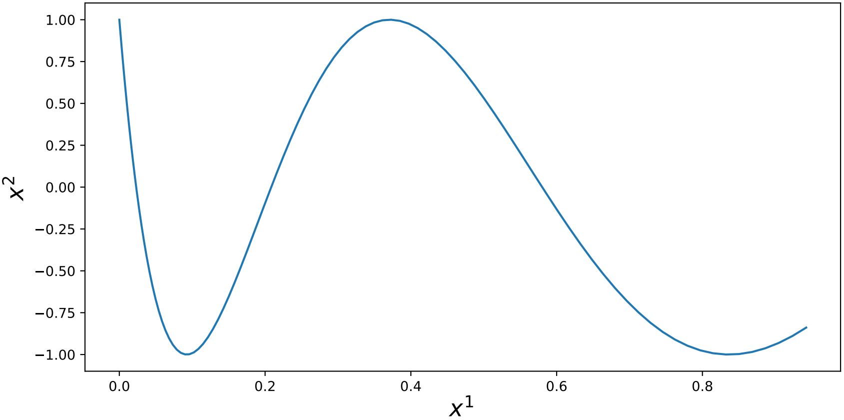
$$\mathbf{a} \otimes \mathbf{b} := \Big(\sum_{i+j=n} a_i \otimes b_j \Big)_{n=0}^{\infty}.$$

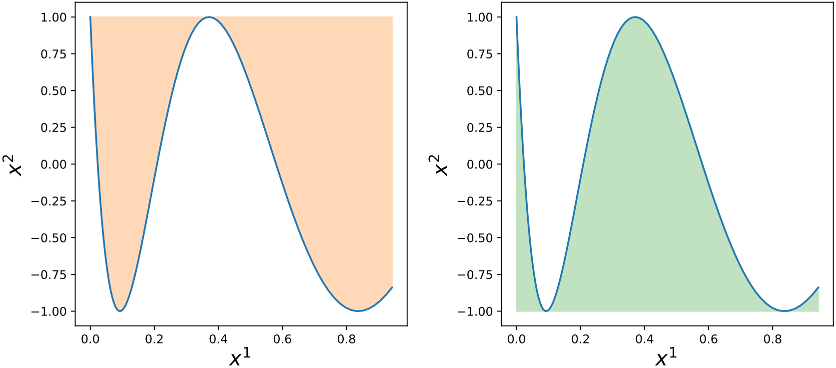
Chen's theorem

$$\mathbb{X}_{s,u}^{<\infty} \otimes \mathbb{X}_{u,t}^{<\infty} = \mathbb{X}_{s,t}^{<\infty}, \quad 0 \le s \le u \le t \le T.$$

- Different topologies have been suggested, leading to (Banach- or Hilbert-) subspaces of $T((\mathbb{R}^d))$. Here, we consider the full space $T((\mathbb{R}^d))$. In contrast, $T^N(\mathbb{R}^d)$ is finite dimensional and endowed with the usual Euclidean topology.

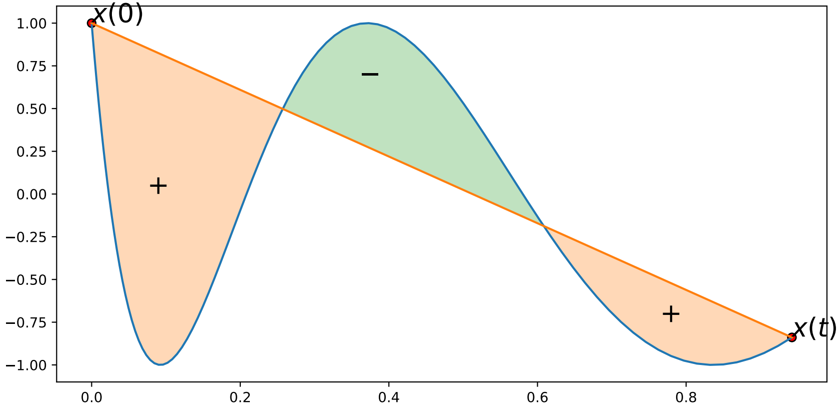
A smooth example path: $X(t) = (\alpha \cosh(\theta_1 t) - \alpha,\ \cos(\theta_2 t))$, $d = 2$





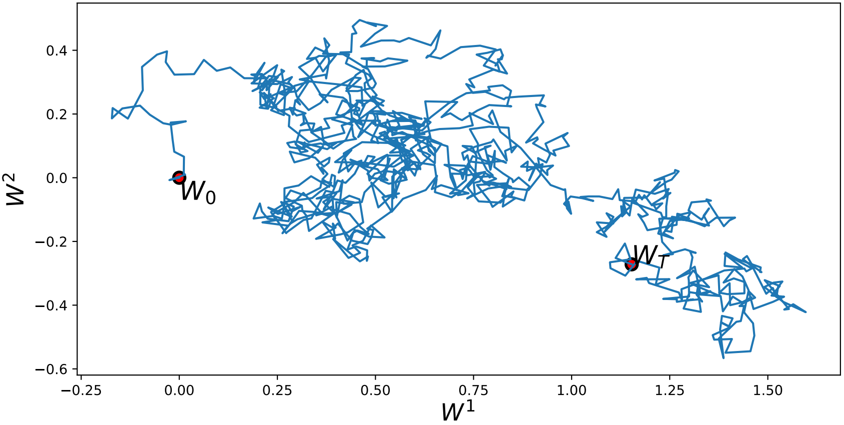
$$\mathbb{X}_{s,t}^{(1,2)} + \mathbb{X}_{s,t}^{(2,1)} = X_{s,t}^1 X_{s,t}^2$$

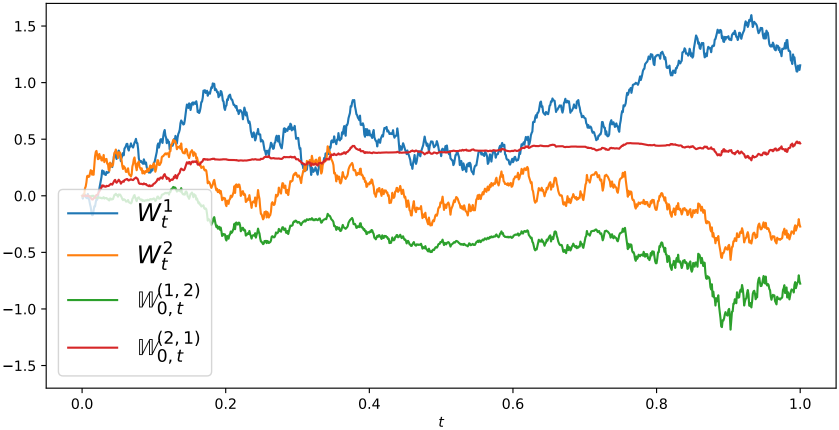
Calculation of the area



$$A_t^{(1,2)} := \frac{1}{2}\mathbb{X}_{s,t}^{(1,2)} - \frac{1}{2}\mathbb{X}_{s,t}^{(2,1)}$$

A two-dimensional Brownian path





Note that $\mathbb{W}_{s,t}^{(i,i)} = \frac{1}{2}(W_{s,t}^i)^2$.

- Let $\mathcal{W}_d$ denote the linear span of words $\mathtt{w}$ in the letters $\{\mathtt{1}, \dots, \mathtt{d}\}$.
- $\mathcal{W}_d$ is an algebra with the concatenation product – equivalent to the tensor product.

- Let $\mathcal{W}_d$ denote the linear span of words $\mathtt{w}$ in the letters $\{\mathtt{1}, \dots, \mathtt{d}\}$.
- $\mathcal{W}_d$ is an algebra with the concatenation product – equivalent to the tensor product.
- Duality with the pairing $\langle \cdot, \cdot \rangle$ defined for $\ell = \lambda_1 \mathtt{w}_1 + \cdots + \lambda_k \mathtt{w}_k \in \mathcal{W}_d$ and $\mathbf{a} \in T((\mathbb{R}^d))$ by

$$\langle \ell, \mathbf{a} \rangle := \lambda_1 a^{\mathtt{w}_1} + \cdots + \lambda_k a^{\mathtt{w}_k},$$

where $a^{\mathtt{i}_1 \cdots \mathtt{i}_m}$ is the coefficient of $\mathbf{a}$ w.r.t. $e_{\mathtt{i}_1} \otimes \cdots \otimes e_{\mathtt{i}_m}$.

- Let $\mathcal{W}_d$ denote the linear span of words $\mathtt{w}$ in the letters $\{\mathtt{1}, \dots, \mathtt{d}\}$.
- $\mathcal{W}_d$ is an algebra with the concatenation product – equivalent to the tensor product.
- Duality with the pairing $\langle \cdot, \cdot \rangle$ defined for $\ell = \lambda_1 \mathtt{w}_1 + \cdots + \lambda_k \mathtt{w}_k \in \mathcal{W}_d$ and $\mathbf{a} \in T((\mathbb{R}^d))$ by

$$\langle \ell, \mathbf{a} \rangle := \lambda_1 a^{\mathtt{w}_1} + \cdots + \lambda_k a^{\mathtt{w}_k},$$

  where $a^{\mathtt{i}_1 \cdots \mathtt{i}_m}$ is the coefficient of $\mathbf{a}$ w.r.t. $e_{\mathtt{i}_1} \otimes \cdots \otimes e_{\mathtt{i}_m}$.
- Shuffle product on $\mathcal{W}_d$: For words $\mathtt{w}, \mathtt{v}$ and letters $\mathtt{i}, \mathtt{j}$ defined by

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

- Example: $\mathtt{12} \sqcup\!\sqcup \mathtt{34} = \mathtt{1234} + \mathtt{1324} + \mathtt{1342} + \mathtt{3124} + \mathtt{3142} + \mathtt{3412}$
- The shuffle product is a commutative product on $\mathcal{W}_d$.

Shuffle identity for signatures

$$\forall \ell_1, \ell_2 \in \mathcal{W}_d : \left\langle \ell_1, \mathbb{X}^{<\infty}_{s,t} \right\rangle \left\langle \ell_2, \mathbb{X}^{<\infty}_{s,t} \right\rangle = \left\langle \ell_1 ⧢ \ell_2, \mathbb{X}^{<\infty}_{s,t} \right\rangle$$

Shuffle identity for signatures

$$\forall \ell_1, \ell_2 \in \mathcal{W}_d : \left\langle \ell_1, \mathbb{X}^{<\infty}_{s,t} \right\rangle \left\langle \ell_2, \mathbb{X}^{<\infty}_{s,t} \right\rangle = \left\langle \ell_1 \sqcup\!\sqcup \ell_2, \mathbb{X}^{<\infty}_{s,t} \right\rangle$$

- Define a group $G(\mathbb{R}^d) \subset T((\mathbb{R}^d))$ w.r.t. $\otimes$ by

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

- Note that $\mathbb{X}^{<\infty}_{s,t} \in G(\mathbb{R}^d)$ for any $s \leq t$ and any smooth path X.

Shuffle identity for signatures

$$\forall \ell_1, \ell_2 \in \mathcal{W}_d : \left\langle \ell_1, \mathbb{X}_{s,t}^{<\infty} \right\rangle \left\langle \ell_2, \mathbb{X}_{s,t}^{<\infty} \right\rangle = \left\langle \ell_1 ⧢ \ell_2, \mathbb{X}_{s,t}^{<\infty} \right\rangle$$

- Define a group $G(\mathbb{R}^d) \subset T((\mathbb{R}^d))$ w.r.t. $\otimes$ by

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

- Note that $\mathbb{X}_{s,t}^{<\infty} \in G(\mathbb{R}^d)$ for any $s \le t$ and any smooth path X.

- For $p \in \mathbb{R}[x]$, i.e., $p(x) = \lambda_0 + \lambda_1 x + \cdots + \lambda_n x^n$, and $\ell \in \mathcal{W}_d$, we have

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

- For $\mathbb{X} : \Delta_T \to T^{\lfloor p \rfloor}(\mathbb{R}^d)$, $\Delta_T := \{ (s,t) \mid 0 \le s \le t \le T \}$, $p \ge 1$, let

$$\|\mathbb{X}\|_{p-\mathrm{var}} := \max_{k=1,\dots,\lfloor p \rfloor} \sup_{\mathcal{D} \text{ partition of } [0,T]} \left[\sum_{t_i \in \mathcal{D}} \left| \pi_k(\mathbb{X}_{t_i,t_{i+1}}) \right|^{\frac{p}{k}} \right]^{\frac{k}{p}}$$

Rough paths

Given $p > 1$, the set Ω_T^p of (geometric) p-rough paths is the closure of $\left\{ \mathbb{X}^{\le \lfloor p \rfloor}_{\cdot,\cdot} \;\middle|\; X \text{ smooth} \right\}$ under $\|\cdot\|_{p-\mathrm{var}}$.

- For $\mathbb{X} : \Delta_T \to T^{\lfloor p \rfloor}(\mathbb{R}^d)$, $\Delta_T := \{ (s,t) \mid 0 \le s \le t \le T \}$, $p \ge 1$, let

$$\|\mathbb{X}\|_{p-\text{var}} := \max_{k=1,\dots,\lfloor p \rfloor} \sup_{\mathcal{D} \text{ partition of } [0,T]} \left[\sum_{t_i \in \mathcal{D}} \left| \pi_k(\mathbb{X}_{t_i,t_{i+1}}) \right|^{\frac{p}{k}} \right]^{\frac{k}{p}}$$

Rough paths

Given $p > 1$, the set Ω_T^p of (geometric) p-rough paths is the closure of $\left\{ \mathbb{X}_{\cdot,\cdot}^{\le \lfloor p \rfloor} \;\middle|\; X \text{ smooth} \right\}$ under $\|\cdot\|_{p-\text{var}}$.

- Given a rough path $\mathbb{X}$, we can construct $\mathbb{X}^{<\infty}$ in a unique, pathwise, continuous way – as well as solving differential equations.
- Example: Let W be a Brownian motion, set $\mathbb{W}(\omega) : \Delta_T \to T^2(\mathbb{R}^d)$ by

$$W_{s,t}^i := W_t^i - W_s^i, \quad W_{s,t}^{i,j} := \int_s^t (W_u^i - W_s^i) \circ \mathrm{d}W_u^j, \quad 1 \le i, j \le d.$$

  This a.s. defines a rough path for $2 < p < 3$, i.e., $\mathbb{W} \in \Omega_T^p$ a.s.

Continuous functionals $f : \Omega_T^p \to \mathbb{R}$ can be approximated by linear functionals $\mathbb{X} \mapsto \left\langle \ell, \mathbb{X}_{0,T}^{<\infty} \right\rangle$, $\ell \in \mathcal{W}_d$.

- This is a consequence of Stone–Weierstrass and the shuffle identity (and holds on compact subsets of Ω_T^p).

Continuous functionals $f : \Omega_T^p \to \mathbb{R}$ can be approximated by linear functionals $\mathbb{X} \mapsto \left\langle \ell, \, \mathbb{X}_{0,T}^{<\infty} \right\rangle$, $\ell \in \mathcal{W}_d$.

- This is a consequence of Stone–Weierstrass and the shuffle identity (and holds on compact subsets of Ω_T^p).

For every rough stochastic process $\widehat{\mathbb{X}}$, the process $t \mapsto \widehat{\mathbb{X}}_{0,t}^{<\infty}$ is a Markov process.

- Every rough path $\mathbb{X}$ *with one strictly monotone component* is uniquely determined by its signature.
- Consider the process $\widehat{X}_t := (t, X_t)$, and its rough path lift to $\widehat{\mathbb{X}} : \Delta_T \to T^{\lfloor p \rfloor}(\mathbb{R}^{d+1})$. Then $\widehat{\mathbb{X}}|_{\Delta_t}$ is uniquely determined by $\widehat{\mathbb{X}}_{0,t}^{<\infty}$, for any $0 \le t \le T$.
- Assuming that X_0 is trivial, the above result follows.

- 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]

- Action recognition
- Medical diagnosis
- Chinese handwriting

[Becker, Cheredito, Jentzen '19] consider the problem $\sup_{0 \le \tau \le 1} \mathbb{E}\left[W_\tau^H\right]$,

where W^H is fractional Brownian motion with Hurst index H – connection to rough stochastic volatility models.

[Becker, Cheredito, Jentzen '19] consider the problem $\sup_{0 \le \tau \le 1} \mathbb{E}\left[W^H_\tau\right]$, where W^H is fractional Brownian motion with Hurst index H – connection to rough stochastic volatility models.

- Fix a time-grid $0 = t_0 < t_1 < \cdots < t_J = 1$, and define a Markov process $X_j \in \mathbb{R}^J$ by

$$\begin{aligned} X_0 &= (0, 0, \dots, 0) \\ X_1 &= (W^H_{t_1}, 0, \dots, 0) \\ X_2 &= (W^H_{t_1}, W^H_{t_2}, 0, \dots, 0) \\ &\vdots \end{aligned}$$

- Use deep neural networks to parameterize stopping decisions $f_j(X_j) \approx \mathrm{DNN}_j(X_j; \theta)$ – "stop at time j unless stopped earlier".

Motivation: Optimal stopping of fractional Brownian motion

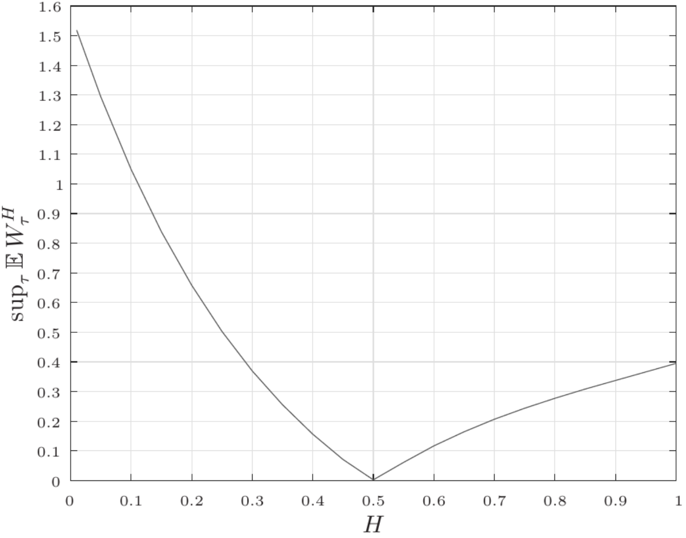


Figure: Plot from [Becker, Cheridito, Jentzen '19], licensed under CC BY 4.0.

On a probability space $(\Omega, \mathcal{F}, \mathbb{P})$ we are given:

- A stochastic process $(X_t)_{t\in[0,T]}$ such that $\widehat{X}_t := (t, X_t)$ extends to a p-rough path $\widehat{\mathbb{X}}$.
- A continuous reward-process $(Y_t)_{t\in[0,T]}$ adapted to the filtration $(\mathcal{F}_t)_{t\in[0,T]}$ generated by $\widehat{\mathbb{X}}$ such that $\mathbb{E}\,\|Y\|_\infty < \infty$.

Optimal stopping problem

Let $\mathcal{S}$ be the set of $(\mathcal{F}_t)_{t\in[0,T]}$-stopping times taking values in $[0, T]$. Solve

$$\sup_{\tau\in\mathcal{S}} \mathbb{E} Y_\tau.$$

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

1. Controls u_t are continuous functions of the path $\phi(\widehat{X}|_{[0,t]})$ and, hence, of the signature $\theta(\widehat{\mathbb{X}}^{<\infty}_{0,t})$ – and similarly for the loss function.
2. We may approximate $\theta(\widehat{\mathbb{X}}^{<\infty}_{0,T})$ by linear functionals $\left\langle \ell, \widehat{\mathbb{X}}^{<\infty}_{0,T} \right\rangle$.
3. Interchange expectation and truncate the signature at level N.
4. Optimize $\ell \mapsto \left\langle \ell, \mathbb{E}\left[\widehat{\mathbb{X}}^{\leq N}_{0,T}\right]\right\rangle$.

No convergence result known so far, but *pathwise* density for steps **1.** + **2.** with high probability is proved in [Kalsi, Lyons, Perez Arribas '20].

Given $\ell \in \mathcal{W}_{d+1}$, set the signature stopping time

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

i.e., a hitting time of a hyperplane in $T((\mathbb{R}^d))$.

Theorem (B., Hager, Riedel, Schoenmakers '23)

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

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

Given $\ell \in \mathcal{W}_{d+1}$, set the signature stopping time

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

i.e., a hitting time of a hyperplane in $T((\mathbb{R}^d))$.

Theorem (B., Hager, Riedel, Schoenmakers '23)

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

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

- While an optimizer $\tau^* \in \mathcal{S}$ of the R.H.S. generally exists, we do not know if there also is an optimizer $\ell^* \in \mathcal{W}_{d+1}$ of the L.H.S.

Step 1: Controls as continuous functionals of paths

- Let $\widehat{\Omega}_t^p$ the set of p-RPs on $[0,t]$ with values in $\mathbb{R}^{1+d}$, the *first component being* $s \mapsto s$
- Let $\Lambda_T := \bigcup_{t\in[0,T]} \widehat{\Omega}_t^p$ be the space of stopped rough paths.
- Λ_T is Polish with Dupire's functional metric based on the p-variation distance.

Step 1: Controls as continuous functionals of paths

- Let $\widehat{\Omega}_t^p$ the set of p-RPs on $[0,t]$ with values in $\mathbb{R}^{1+d}$, the *first component being* $s \mapsto s$
- Let $\Lambda_T := \bigcup_{t\in[0,T]} \widehat{\Omega}_t^p$ be the space of stopped rough paths.
- Λ_T is Polish with Dupire's functional metric based on the p-variation distance.
- Given $\theta \in C(\Lambda_T, \mathbb{R})$, we define a continuous stopping rule by

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

Lemma

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

Step 2: Approximation by linear functionals of the signature

- Problem: Candidate stopping times τ or τ_θ are typically discontinuous functions of the path.

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

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

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

Lemma

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

and similarly for signature stopping rules.

Randomization regularizes the optimal stopping problem

$$\mathbb{E}\left[Y_{\tau^r_\theta \wedge T} \mid \widehat{\mathbb{X}}\right] = Y_0 + \int_0^T \left[1 - F_Z\left(\int_0^t \theta\left(\widehat{\mathbb{X}}|_{[0,s]}\right)^2 \mathrm{d}s\right)\right] \mathrm{d}Y_t$$

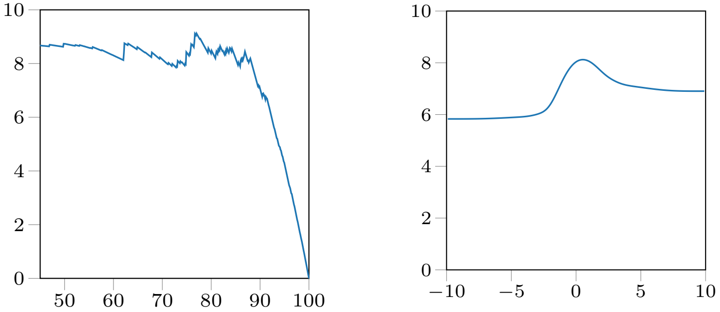


Figure: Example loss function based on 100 samples from [B, Tempone, Wolfers '20]. **L** No randomization. **R** With randomization.

Randomization regularizes the optimal stopping problem

$$\mathbb{E}\left[Y_{\tau_\theta^r \wedge T} \mid \widehat{\mathbb{X}}\right] = Y_0 + \int_0^T \left[1 - F_Z\left(\int_0^t \theta\left(\widehat{\mathbb{X}}|_{[0,s]}\right)^2 \mathrm{d}s\right)\right] \mathrm{d}Y_t$$

- The approximation by linear functionals now follows by Stone-Weierstrass together with dominated convergence, noting that for any stopping time τ (randomized or not, signature based or not):

$$\mathbb{E}\left[Y_\tau\right] \leq \mathbb{E}\left[\|Y\|_{\infty;[0,T]}\right] < \infty.$$

- Randomization can be used to substantially improve the accuracy of numerical approximations of optimal stopping problems.

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

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

- Recalling that $\widehat{X}_t = (t, X_t)$, we have

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

- $\exp$ can be approximated by polynomials, leading to the exponential shuffle.
- Often, Y can also be approximated by a linear functional on $\widehat{\mathbb{X}}^{<\infty}$. Otherwise, consider a RP extending $t \mapsto (t, X_t, Y_t)$
- Need to truncate the signature.

- For $\ell \in \mathcal{W}_{d+1}$, define

$$\exp^{\sqcup\!\sqcup}(\ell) := \sum_{n=0}^{\infty} \frac{1}{n!} \ell^{\sqcup\!\sqcup n}.$$

- As a formal series, $\exp^{\sqcup\!\sqcup}(l)$ does not define a linear map on $T((\mathbb{R}^{1+d}))$, but it does define
  - a linear map on $T^N(\mathbb{R}^{1+d})$;
  - a map on the group-like elements $G(\mathbb{R}^{1+d})$, i.e., on signatures.

Lemma

Let $\mathbf{g} \in G(\mathbb{R}^{1+d})$, $\ell \in \mathcal{W}_{d+1}$. Then

$$\left|\exp\left(\langle \ell, \mathbf{g}\rangle\right) - \langle \exp^{\sqcup\!\sqcup}(\ell), \pi_{\leq N}(\mathbf{g})\rangle\right| \leq 4 \exp(\langle \ell, \mathbf{1}\rangle) \frac{\left(|\ell| \left|\pi_{\leq \deg(\ell)}(\mathbf{g})\right|\right)^{\left\lfloor \frac{N}{\deg(\ell)} \right\rfloor + 1}}{(\lfloor N/\deg(\ell) \rfloor + 1)!}.$$

Theorem (B., Hager, Riedel, Schoenmakers '21)

Let $\mathbb{E}\left[\|Y\|_\infty\right] < \infty$. *Given* $\kappa > 0$, *define the stopping time* $\sigma = \sigma_\kappa$ *by*

$$\sigma := \inf\left\{ t \geq 0 \;\middle|\; \left\|\widehat{\mathbb{X}}\right\|_{p-\mathrm{var};[0,t]} \geq \kappa \right\} \wedge T. \text{ Then,}$$

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

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:

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

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

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

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

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

- The linear signature stopping rules have nice theoretical properties, but do not seem to work in practice, due to the exponential.

- The linear signature stopping rules have nice theoretical properties, but do not seem to work in practice, due to the exponential.

Log-signatures

- $\mathbb{L}_{0,t}^{<\infty} := \log^{\otimes}(\mathbb{X}_{0,t}^{<\infty}) \in \mathfrak{g}(\mathbb{R}^{1+d}) := \log^{\otimes}(G(\mathbb{R}^{1+d}))$, a free Lie algebra.
- Reduces redundancies in the signature, dimension reduction.

- Deep signature stopping rule: $\theta\left(\widehat{\mathbb{X}}_{0,t}^{\leq N}\right) := \vartheta\left(\log^{\otimes}\left(\widehat{\mathbb{X}}_{0,t}^{\leq N}\right)\right)$, ϑ being a standard (deep) neural network.

- The linear signature stopping rules have nice theoretical properties, but do not seem to work in practice, due to the exponential.

Log-signatures

- $\mathbb{L}^{<\infty}_{0,t} := \log^{\otimes}(\mathbb{X}^{<\infty}_{0,t}) \in \mathfrak{g}(\mathbb{R}^{1+d}) := \log^{\otimes}(G(\mathbb{R}^{1+d}))$, a free Lie algebra.
- Reduces redundancies in the signature, dimension reduction.

- Deep signature stopping rule: $\theta\left(\widehat{\mathbb{X}}^{\leq N}_{0,t}\right) := \vartheta\left(\log^{\otimes}\left(\widehat{\mathbb{X}}^{\leq N}_{0,t}\right)\right)$, ϑ being a standard (deep) neural network.
- Obtain similar theoretical convergence result, but also works well in numerical examples.

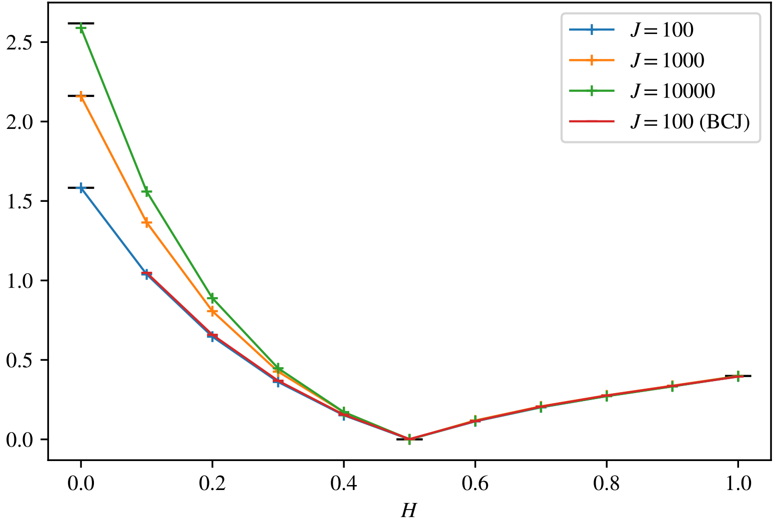


Figure: Approximation based on J time steps, log-signature truncated at $N = 3$ ($\dim \mathfrak{g}^{\leq N} = 5$), NN with 2 hidden layers.

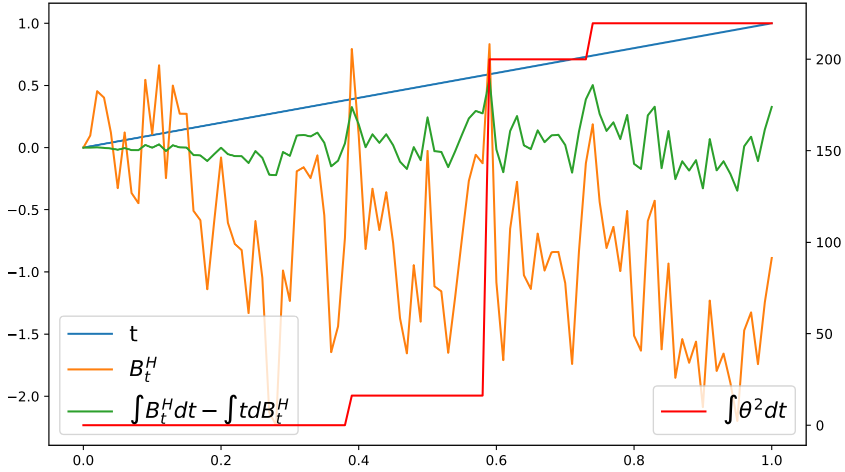


Figure: Approximate randomized stopping rule and select log-signature entries for one trajectory of a fractional Brownian motion with $H = 0.1$

Thank you for your attention!

- C. Bayer, M. Eigel, L. Sallandt, P. Trunschke. *Pricing high-dimensional Bermudan options with hierarchical tensor formats*. SIAM J. Financial Math., to appear, 2023.
- C. Bayer, P. Hager, S. Riedel, J. Schoenmakers. *Optimal stopping with signatures*. Annals of Appl. Prob., 2023.
- C. Bayer, R. Tempone, and S. Wolfers. *Pricing American options by exercise rate optimization*. Quant. Fin., 2020.
- 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.