

Linear independence properties of the signature components of time-augmented stochastic processes.

Arthur Bourdon^{1,2}, Benjamin Jourdain², Hervé Andrès¹

¹Milliman R&D, Paris, France

²CERMICS, ENPC, Institut Polytechnique de Paris, CNRS, Marne-la-Vallée, France & MATHRISK team-project, Inria Paris, France.

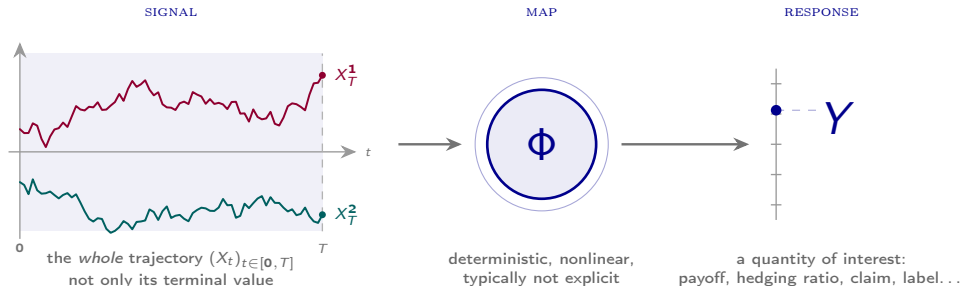
Contributed Talk at the *"XIII World Congress Bachelier"*, Bologna, Italia.

June 30, 2026



ALMA MATER STUDIORUM
UNIVERSITÀ DI BOLOGNA





$$Y = \Phi((X_t)_{t \in [0, T]})$$

Goal. Build an approximation $\tilde{\Phi}$ of Φ that is cheap to evaluate and tractable to fit.

The signature of a path

For a path of bounded variation $X := (X_t^1, \dots, X_t^d)_{t \in [0, T]}$, and for each word $i_1 \dots i_n$ with $i_k \in \{0, \dots, d\}$ and $n \in \mathbb{N}^*$ a positive integer, we define:

$$\forall t \in [0, T], \quad \widehat{X}_t^{i_1 \dots i_n} = \int_0^t \int_0^{t_1} \dots \int_0^{t_{n-1}} d\widehat{X}_{t_n}^{i_1} \dots d\widehat{X}_{t_1}^{i_n},$$

and

$$\forall t \in [0, T], \quad \widehat{X}_t^\emptyset = 1,$$

where $\emptyset$ denotes the empty word, $\widehat{X}_t^0 = t$ and $\widehat{X}_t^i = X_t^i$ for $i \in \{1, \dots, d\}$.

- The **time-augmented signature** of X is $\widehat{X} := (\widehat{X}^{i_1 \dots i_n})_{(i_1, \dots, i_n) \in \{0, \dots, d\}^n, n \in \mathbb{N}}$.
- For $N \in \mathbb{N}$, the **N -step truncated time-augmented signature** of X is $\widehat{X}^{\leq N} := (\widehat{X}^{i_1 \dots i_n})_{(i_1, \dots, i_n) \in \mathcal{A}^n, n \leq N}$
 $\rightsquigarrow \forall t \in [0, T], \quad \widehat{X}_t^{\leq N} \in \mathbb{R}^{\frac{(d+1)^{N+1} - 1}{d}} \rightsquigarrow$ **finite family of features**.
- Libraries for efficiently computing signatures include: *iisignature* [Reizenstein and Graham, 2020], *signatory*, *esig*.

When X is a **continuous semimartingale**: we define its time-augmented signature using **Stratonovich** integration theory.

The signature of a path: approximation property

Linear functionals on the time-augmented signature terminal value are universal approximators.

[Cuchiero et al., 2026, Bayer et al., 2025a, Bayer et al., 2025b].

$$\Phi(X) \simeq \underbrace{\langle \beta, \widehat{\mathbb{X}}_T^{\leq N} \rangle}_{\widehat{\Phi}(X)}, \quad \text{for large } N \text{ and some } \beta \in \mathbb{R}^{\frac{(d+1)N+1}{d}}$$

nonlinear path-dependent functional $\rightsquigarrow$ linear model in nonlinear path features

$$\beta_N^* := \arg \min_{\beta} \mathbb{E}[|\Phi(X) - \langle \beta, \widehat{\mathbb{X}}_T^{\leq N} \rangle|^2]$$

Finding a basis of the linear span of the components of $\widehat{\mathbb{X}}_T^{\leq N}$ is interesting for several reasons :

- β_N^* is unique $\iff$ components of $\widehat{\mathbb{X}}_T^{\leq N}$ are linearly independent $\rightsquigarrow$ **improves interpretability**
- reduction of **computational and storage cost of features are reduced**
- the **prediction performance is preserved**, while empirical improvements may arise in practice due to improved conditioning, reduced variance, or more stable cross-validation.

The signature of a path: the central role of words

Notations.

- $\mathcal{W}_0 = \{\emptyset\}$,
- $\mathcal{W}_N = \{i_1 i_2 \cdots i_N \mid (i_1, \dots, i_N) \in \mathcal{A}^N\}$ for $N \in \mathbb{N}^* = \mathbb{N} \setminus \{0\}$, $\rightsquigarrow$ set of words of length N
- $\mathcal{W}_{\leq N} = \bigsqcup_{k=0}^N \mathcal{W}_k$ for $N \in \mathbb{N}$, $\rightsquigarrow$ set of words of length not greater than N ,
- $\mathcal{W}_{<\infty} = \bigsqcup_{k=0}^{\infty} \mathcal{W}_k$, $\rightsquigarrow$ set of words of finite length

We denote the length of a word $w \in \mathcal{W}_{<\infty}$ by $\|w\|$.

Definition (Shuffle product)

The shuffle product between two words $\sqcup : \mathcal{W}_{<\infty} \times \mathcal{W}_{<\infty} \longrightarrow \text{Span}(\mathcal{W}_{<\infty})$ is defined inductively for all $w, v \in \mathcal{W}_{<\infty}$ and $i, j \in \mathcal{A}$ as follows :

$$\begin{cases} w \sqcup \emptyset = \emptyset \sqcup w = w \\ (wi) \sqcup (vj) = (w \sqcup (vj))i + ((wi) \sqcup v)j. \end{cases}$$

The shuffle product can be extended to $\text{Span}(\mathcal{W}_{<\infty}) \times \text{Span}(\mathcal{W}_{<\infty})$ by bilinearity.

Remark: We can easily observe that the words which appear in the shuffle of w and v all have length $\|w\| + \|v\|$.

Example: $12 \sqcup 01 = 1201 + 1021 + 0121 + 1012 + 0112 + 0112$.

Proposition (Shuffle product property)

Let $X : [0, T] \rightarrow \mathbb{R}^d$ be a path such that its time-augmented signature $\widehat{\mathbb{X}}$ is well defined. Then

$$\forall t \in [0, T], \forall \mathbf{w}, \mathbf{v} \in \mathcal{W}_{<\infty}, \quad \langle \mathbf{w}, \widehat{\mathbb{X}}_t \rangle \langle \mathbf{v}, \widehat{\mathbb{X}}_t \rangle = \langle \mathbf{w} \sqcup \mathbf{v}, \widehat{\mathbb{X}}_t \rangle.$$

Example: Let us consider a path $X := (X^1, X^2) : [0, T] \rightarrow \mathbb{R}^2$. Hence we have

$$\int_0^T dX_t^1 \int_0^T dX_t^2 = \widehat{\mathbb{X}}_T^1 \widehat{\mathbb{X}}_T^2 = \langle 1, \widehat{\mathbb{X}}_T \rangle \langle 2, \widehat{\mathbb{X}}_T \rangle = \langle 1 \sqcup 2, \widehat{\mathbb{X}}_T \rangle = \langle 12 + 21, \widehat{\mathbb{X}}_T \rangle = \int_0^T \int_0^t dX_s^1 dX_t^2 + \int_0^T \int_0^t dX_s^2 dX_t^1.$$

Generalization of the integration by parts formula.

The signature of a path: the central role of words

$(\Omega, \mathcal{F}, \mathbb{P})$ probability space and $X = (X_t)_{t \in [0, T]}$ be a continuous stochastic process of law $\mathbb{P}_X$ such that its time-augmented signature is well defined.

We denote the word $\underbrace{00 \cdots 0}_k$ by 0_k for $k \in \mathbb{N}$. Since $\langle 0_k, \hat{\mathbb{X}}_T \rangle = \frac{T^k}{k!} > 0$, the shuffle product property yields

$$\forall k \in \mathbb{N}, \quad \langle \mathbf{w}, \hat{\mathbb{X}}_T \rangle = \langle \frac{k!}{T^k} (\mathbf{w} \sqcup 0_k), \hat{\mathbb{X}}_T \rangle,$$

$$\frac{k!}{T^k} (\mathbf{w} \sqcup 0_k) \rightsquigarrow \text{linear combination of words of length } |||\mathbf{w}||| + k.$$

Consequence: For $N \in \mathbb{N}$ and $\mathbf{w} \in \mathcal{W}_{\leq N} \setminus \mathcal{W}_N$. Taking $k = N - |||\mathbf{w}||| > 0$, we deduce

$$\hat{\mathbb{X}}_T^{\mathbf{w}} = \text{linear combination of components associated with words of length } N.$$

The signature of a path : spanning results

Theorem

$\text{Span}(\hat{\mathbb{X}}_T^{\mathbf{w}} \mid \mathbf{w} \in \mathcal{W}_N) = \text{Span}(\hat{\mathbb{X}}_T^{\mathbf{w}} \mid \mathbf{w} \in \mathcal{W}_{\leq N}),$ whatever the law $\mathbb{P}_X$.

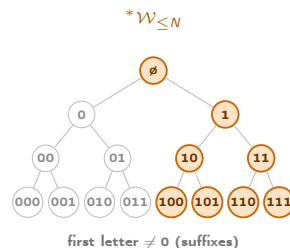
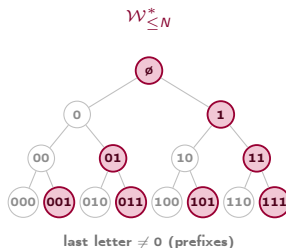
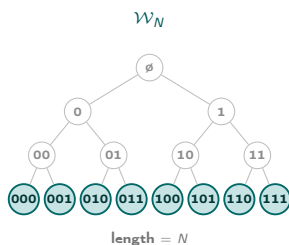
Other subsets of $\mathcal{W}_{\leq N}$ than $\mathcal{W}_N$ share the same spanning property, such as

■ $\mathcal{W}_{\leq N}^* := \{\emptyset\} \cup \{\mathbf{i}_1 \cdots \mathbf{i}_n \in \mathcal{W}_{\leq N} \mid \mathbf{i}_n \neq 0\}$

(prefixes)

■ ${}^*\mathcal{W}_{\leq N} := \{\emptyset\} \cup \{\mathbf{i}_1 \cdots \mathbf{i}_n \in \mathcal{W}_{\leq N} \mid \mathbf{i}_1 \neq 0\}$

(suffixes)



All these sets have the same cardinality $(d+1)^N$ (above: $d=1$, $N=3 \rightsquigarrow 8$ words out of 15).

Theorem

Let $m, d \in \mathbb{N}^*$, $(B_t)_{t \in [0, T]}$ be a m -dimensional Brownian motion, and Y be a random variable independent of $(B_t)_{t \in [0, T]}$. Let $(X_t)_{t \in [0, T]} := (X_t^1, \dots, X_t^d)_{t \in [0, T]}$ be the solution to:

$$dX_t = \sigma(t, X_t)dB_t + b(t, X_t)dt, \quad X_0 = Y, \quad (1)$$

where

- $b : [0, T] \times \mathbb{R}^d \rightarrow \mathbb{R}^d$ and $\sigma : [0, T] \times \mathbb{R}^d \rightarrow \mathbb{R}^{d \times m}$ are $C^{1,3}$ with linear growth in space uniformly in time,
- $\exists \lambda > 0, \forall t \in [0, T], \forall x \in \mathbb{R}^d, \forall \xi \in \mathbb{R}^d, \quad \xi \sigma(t, x) \sigma(t, x)^\top \xi^\top \geq \lambda \|\xi\|^2$.

Then for all $N \in \mathbb{N}$, the family of random variables $(\hat{X}_T^w)_{w \in \mathcal{W}_{\leq N}^*}$ is linearly independent.

The result still holds when considering [affine interpolation of the solution on a time grid](#).

First step:

For $(W_t)_{t \in [0, T]}$ a d -dimensional Brownian motion, $(\widehat{W}_T^{\mathbf{w}})_{\mathbf{w} \in \mathcal{W}_{\leq N}^*}$ is linearly independent.

- ① The family of stochastic processes $(\widehat{W}_t^{\leq N})_{t \in [0, T]}$ is linearly independent for the $dt \otimes d\mathbb{P}$ -a.e. equality.
- ② $\forall \mathbf{w} \in \mathcal{W}_{\leq N}^*, \quad \mathbb{E} \left[\widehat{W}_T^{\mathbf{w}} | \mathcal{F}_t \right] = \widehat{W}_t^{\mathbf{w}} + \sum_{k=1}^{M^{\mathbf{w}}} P_{k, \mathbf{w}}(T) \widehat{W}_t^{\mathbf{v}_{k, \mathbf{w}}},$ and $\deg(P_{k, \mathbf{w}}) + \|\mathbf{v}_{k, \mathbf{w}}\| \leq \|\mathbf{w}\| - 1.$

Second step:

The map $\mathbf{x} \mapsto \sum_{\mathbf{w} \in \mathcal{W}_{\leq N}^*} c_{\mathbf{w}} \langle \mathbf{w}, S_N(\widehat{\mathbf{x}})_T \rangle$ is continuous for the rough-path topology. Hence, for all $\mathbf{x} \in \text{Supp}(X), \quad \sum_{\mathbf{w} \in \mathcal{W}_{\leq N}^*} c_{\mathbf{w}} \langle \mathbf{w}, S_N(\widehat{\mathbf{x}})_T \rangle = 0.$ Here $S_N : X \mapsto \widehat{X}^{\leq N}.$

Third step:

By support theorem, we show that for any piecewise affine path $\gamma : [0, T] \rightarrow \mathbb{R}^d, S_2(\gamma) \in \text{Supp}(X).$ For W a d -dimensional Brownian motion, we construct $W^{(n)}$ the dyadic piecewise approximation of $W.$ Classical results gives $S_2(W^{(n)}) \rightarrow W$ for the rough-path topology.

Efficient computation of signature terms

For each word $\mathbf{w} \in \mathcal{W}_{<\infty}$ we denote by $S^{\mathbf{w}} : X \mapsto \widehat{\mathbb{X}}_T^{\mathbf{w}}$.

Proposition (Time-augmented signature of an affine path)

Let $T > 0$ and $X := (X_t^1, \dots, X_t^d)_{t \in [0, T]}$ be an affine path taking values in $\mathbb{R}^d$. Then for all $N \in \mathbb{N}^*$ and for all $\mathbf{w} = i_1 \cdots i_N \in \mathcal{W}_N$ we have the following equality

$$S^{\mathbf{w}}(X) = \frac{1}{N!} \prod_{k=1}^N \left(\widehat{X}_T^{i_k} - \widehat{X}_0^{i_k} \right),$$

where $\widehat{X}_t^0 = t$ and $\widehat{X}_t^i = X_t^i$ for $i \in \{1, \dots, d\}$.

Definition (Concatenation of paths)

Let $T_1, T_2 > 0$, and $X : [0, T_1] \rightarrow \mathbb{R}^d$, $Y : [0, T_2] \rightarrow \mathbb{R}^d$ be two paths. We denote by $X * Y : [0, T_1 + T_2] \rightarrow \mathbb{R}^d$ the concatenation of X and Y defined as

$$X * Y : t \mapsto \begin{cases} X_t & \text{if } t \in [0, T_1] \\ Y_{t-T_1} - Y_0 + X_{T_1} & \text{if } t \in [T_1, T_1 + T_2] \end{cases}$$

Proposition (Chen's relation)

Let $T_1, T_2 > 0$, and $X : [0, T_1] \rightarrow \mathbb{R}^d$, $Y : [0, T_2] \rightarrow \mathbb{R}^d$ be two paths of finite variation. For any word $i_1 \cdots i_m$ we have the following relation

$$S^{i_1 \cdots i_m}(X * Y) = \sum_{k=0}^m S^{i_1 \cdots i_k}(X) S^{i_{k+1} \cdots i_m}(Y),$$

with the convention $i_1 \cdots i_k = \emptyset$ if $k = 0$, and $i_{k+1} \cdots i_m = \emptyset$ if $k = m$.

Consider a time horizon $T > 0$, an integer $K \geq 2$, and $X = \Delta X_1 * \cdots * \Delta X_K$ defined as the concatenation of K affine paths $\Delta X_k : [0, \frac{T}{K}] \rightarrow \mathbb{R}^d$. We denote by

$$X_{|k} = \Delta X_1 * \cdots * \Delta X_k : \text{forward concatenation.}$$

For the **forward** approach, we iteratively compute the signature of $X_{|k+1} = X_{|k} * \Delta X_{k+1}$ by applying Chen's relation until we obtain the signature of the full path $X = X_{|K}$. The computation of $S^{i_1 \cdots i_m}(X_{|k+1})$ requires

- 1 computing $(S^{i_1 \cdots i_j}(X_{|k}))_{j \in \{0, \dots, m\}}$,
- 2 computing $(S^{i_{j+1} \cdots i_m}(\Delta X_{k+1}))_{j \in \{0, \dots, m\}}$,
- 3 performing $2m$ elementary operations. This corresponds to parsing the $m + 1$ elements in the summation, and performing $m - 1$ products (since there is no products for $j = 0$ and $j = m$).

We define

- $C(B, k)$ the required number of elementary operations to compute $(S^w(X_{|k}))_{w \in B}$,
- $\Phi_f(B) = \{w \in \mathcal{W}_{<\infty} \mid \exists v \in \mathcal{W}_{<\infty}, wv \in B\}$,
- $|||B||| = \sum_{w \in B} \|w\|$.

We obtain the following recursive formula

$$C(B, k) = C(\Phi_f(B), k - 1) + 2|||B|||.$$

Since

- 1 $\Phi_f \circ \Phi_f = \Phi_f$,
- 2 $C(B, 1) = \text{Card}(B)$,

we deduce

$$C(B, k) = \text{Card}(\Phi_f(B)) + 2(k - 2)|||\Phi_f(B)||| + 2|||B|||.$$

Total cost formula:

$$C_f(B, K) = \text{Card}(\Phi_f(B)) + 2(K - 2) |||\Phi_f(B)||| + 2 |||B|||.$$

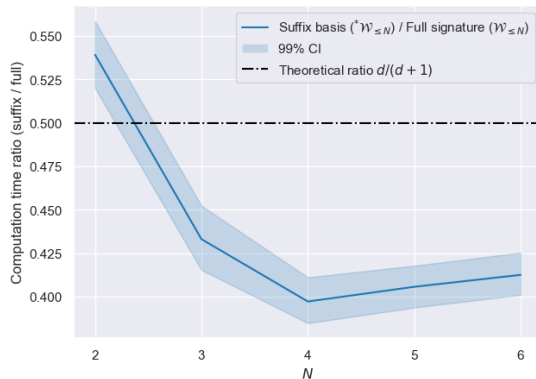
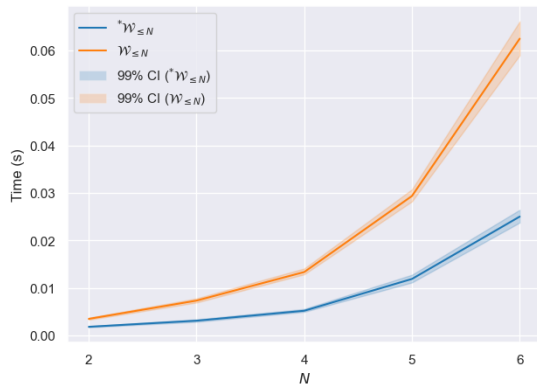
The minimum of $C_f(\cdot, K)$ on the set of subsets of $\mathcal{W}_N$ with cardinality $(d + 1)^N$ is attained by ${}^*\mathcal{W}_{\leq N}$. When applied to the set ${}^*\mathcal{W}_{\leq N}$, the gain in elementary operations behaves like

$$\frac{C_f({}^*\mathcal{W}_{\leq N}, K)}{C_f(\mathcal{W}_{\leq N}, K)} \xrightarrow{K \rightarrow \infty} \frac{d}{d + 1}.$$

Numerical results

We used the library `iisignature` and implemented an efficient iterative method to compute suffixes components.

Figure: Computation time of signature components and comparison with theoretical scaling. We sample 200 batches of size $n = 500$ of discretized Brownian motion sample paths and compute the empirical mean computation time across batches.



- We exhibit sets of words $\mathcal{B} \subset \mathcal{W}_{\leq N}$ s.t.
 - ① $\text{Card}(\mathcal{B}) = (d+1)^N$,
 - ② $\mathcal{B}$ does not depend on $\mathbb{P}_X$,
 - ③ $\text{Span}(\hat{\mathbb{X}}_T^{\mathbf{w}} \mid \mathbf{w} \in \mathcal{B}) = \text{Span}(\hat{\mathbb{X}}_T^{\leq N})$.
- $\Rightarrow \dim \left(\text{Span} \left(\hat{\mathbb{X}}_T^{\mathbf{w}} \mid \mathbf{w} \in \mathcal{W}_{\leq N} \right) \right) \leq (d+1)^N$, **whatever** $\mathbb{P}_X$.
- $^*\mathcal{W}_{\leq N}$ is such a subset and has the minimal computation cost when using Chen's relation.
- We show that if we consider a process solution to an SDE with non-degenerate diffusion coefficient, then $\dim \left(\text{Span} \left(\hat{\mathbb{X}}_T^{\mathbf{w}} \mid \mathbf{w} \in \mathcal{W}_{\leq N} \right) \right) = (d+1)^N$.



Bayer, C., Pelizzari, L., and Schoenmakers, J. (2025a).

Primal and dual optimal stopping with signatures.

Finance and Stochastics, 29(4):981–1014.



Bayer, C., Pelizzari, L., and Zhu, J.-J. (2025b).

Pricing american options under rough volatility using deep-signatures and signature-kernels.

arXiv preprint arXiv:2501.06758.



Cuchiero, C., Schmock, P., and Teichmann, J. (2026).

Global universal approximation of functional input maps on weighted spaces.

Constructive Approximation, pages 1–76.



Reizenstein, J. and Graham, B. (2020).

Algorithm 1004: The iisignature library: Efficient calculation of iterated-integral signatures and log signatures.

ACM Transactions on Mathematical Software (TOMS).

Thanks for your attention!

Questions?