The signature of a path, and inversion

Weijun Xu

June 24, 2020

The signature of a path

Euclidan coordinates for $\mathbf{R}^d$: $(e_1, \ldots, e_d)$.

A path $\gamma = (\gamma^1, \ldots, \gamma^d) : [0,1] \to \mathbf{R}^d$ continuously differentiable.

For word $w = e_{i_1} \cdots e_{i_n}$, define

$$C_\gamma(w) = \int_{0<u_1<\cdots<u_n<1} \mathrm{d}\gamma^{i_1}(u_1) \cdots \mathrm{d}\gamma^{i_n}(u_n).$$

The signature of γ is the collection of all $C(w)$'s, denoted by $\mathrm{Sig}(\gamma)$:

$$\mathrm{Sig}(\gamma) := \sum_{n \geq 0} \underbrace{\sum_{w:|w|=n} C_\gamma(w) \cdot w}_{\mathrm{Sig}^{(n)}(\gamma)} \in \bigoplus_{n=0}^{+\infty} (\mathbf{R}^d)^{\otimes n}.$$

Independent of parametrisation. It captures the ordered evolution along the path through the order of the letters.

First two levels

Let $\gamma : [0, 1] \to \mathbf{R}^d$. Consider the e_i-e_j plane.

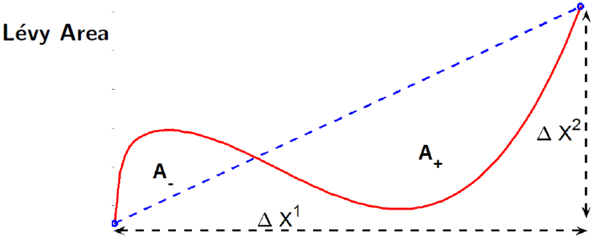
First level: $C(e_i) = \gamma^i(1) - \gamma^i(0)$ is the increment.

Second level signature:

$$C(e_i^2) = \frac{1}{2}\left(\gamma^i(1) - \gamma^i(0)\right)^2 = \frac{1}{2}\left(C(e_i)\right)^2$$

$$C(e_i e_j) + C(e_j e_i) = \left(\gamma^1(1) - \gamma^i(0)\right)\left(\gamma^j(1) - \gamma^j(0)\right) = C(e_i)C(e_j)$$

$$C(e_i e_j) - C(e_j e_i) = A_+ - A_- \; .$$



Examples

Consider $\mathbf{R}^2$ with standard basis $(e_1, e_2) = (x, y)$.

1. Path movement: $(0,0) \to (a,0) \to (a,b)$.

   $\mathrm{Sig}(ax * by) = \exp(ax)\exp(by)$. $C(w) \neq 0$ only if $w = x^k y^\ell$.

2. Path movement: $(0,0) \to (0,b) \to (a,b)$.

   $\mathrm{Sig}(by * ax) = \exp(by)\exp(ax)$. $C(w) \neq 0$ only if $w = y^k x^\ell$.

3. Straightline segment from $(0,0)$ to (a,b).

   $\mathrm{Sig}(ax + by) = \exp(ax + by)$.

They have the same $\mathrm{Sig}^{(1)}$, but different $\mathrm{Sig}^{(2)}$.

Exponential homomorphism

In the tensor language:

$$\mathrm{Sig}^{(n)}(\gamma) = \int_{0<u_1<\cdots<u_n<1} \mathrm{d}\gamma(u_1) \otimes \cdots \otimes \mathrm{d}\gamma(u_n) \in (\mathbf{R}^d)^{\otimes n} .$$

Path group with concatenation $(*)$ and inverse: $\alpha, \beta : [0,1] \to \mathbf{R}^d$, then

$$\text{Concatenation:} \quad (\alpha * \beta)(t) = \begin{cases} \alpha(2t) , & t \in [0, \frac{1}{2}] , \\ \alpha(1) + \beta(2t-1) , & t \in [\frac{1}{2}, 1]. \end{cases}$$

$$\text{Inverse:} \qquad \alpha^{-1}(t) = \alpha(1-t), \quad t \in [0,1].$$

Signature is a group homomorphism:

$$\mathrm{Sig}(\alpha * \beta) = \mathrm{Sig}(\alpha) \otimes \mathrm{Sig}(\beta) , \qquad \mathrm{Sig}(\alpha * \alpha^{-1}) = 1 .$$

Three questions

1. What elements in the tensor algebra are in the image of the signature?
2. Given an element in the image of the signature map, how many paths are there in the pre-image?
3. Given a signature, how can we reconstruct (all) paths with that signature?

Signature image

$\mathrm{Sig}(\gamma) = \{C_\gamma(w)\}_w$ satisfies the following shuffle relation:

$$C(w')C(w'') = \sum_{w \in w' \sqcup\!\sqcup w''} C(w) , \qquad \forall w', w'' . \tag{1}$$

Here, $\sqcup\!\sqcup$ denotes the shuffle product, giving all possible ways to putting two words together while preserving their own orders. For example:

$$ab \sqcup\!\sqcup c = abc + acb + cab .$$

Remark: this relation forces $C(\emptyset) = 1$.

Kuo-Tsai Chen: $\mathrm{Sig}(\gamma)$ is the exponential of a Lie element.

Wei-Liang Chow: If $\mathcal{S} \in \bigoplus_{n=0}^{+\infty} (\mathbf{R}^d)^{\otimes n}$ satisfies the shuffle relation (1), then for every N, there exists a piecewise linear path γ such that

$$\mathrm{Sig}(\gamma) = \mathcal{S} \quad \text{in} \quad \bigoplus_{n=0}^{N} (\mathbf{R}^d)^{\otimes n} .$$

Signature image

Recall

$$\mathrm{Sig}^{(n)}(\gamma) = \int_{0<u_1<\cdots<u_n<1} \mathrm{d}\gamma(u_1) \otimes \cdots \otimes \mathrm{d}\gamma(u_n) \ .$$

The domain of integration has size $\frac{1}{n!}$, and there are d^n terms in $\mathrm{Sig}^{(n)}$:

$$\|\mathrm{Sig}^{(n)}(\gamma)\| \leq \frac{L^n}{n!} \ .$$

Remark: $n!$ replaced by $(\alpha n)!$ if γ is α-Hölder.

The necessary conditions for $\mathcal{S} \in \bigoplus_{n=0}^{\infty} (\mathbf{R}^d)^{\otimes n}$ to be in the image of the signature are to satisfy both the shuffle property and the factorial decay.

Whether they are sufficient or not is unclear.

An example

For every n, we give two paths α and β such that $\text{Sig}^{(k)}(\alpha) = \text{Sig}^{(k)}(\beta)$ for every $k \leq n$.

Consider dimension two. Let $\alpha_0 = x$, $\beta_0 = y$. Define

$$\alpha_{k+1} = \alpha_k * \beta_k \ , \qquad \beta_{k+1} = \beta_k * \alpha_k \ .$$

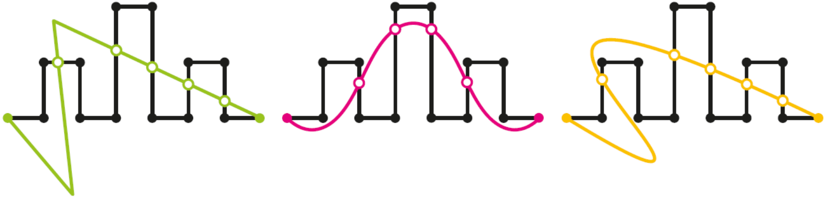
Then α_n and β_n have the same signature up to level n.

e.g.: $\alpha_3 = xyyxyxxy$ and $\beta_3 = yxxyxyyx$ are two lattice paths with 8 steps, and have the same signatures up to level three.

$\text{Sig}^{(n)}$ describes finer (local) information of the path when n becomes bigger.

More examples

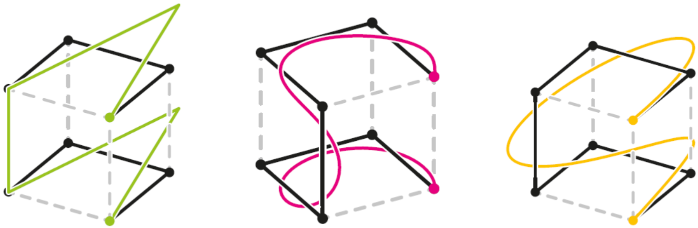
Pfeffer-Seigal-Sturmfels: the (shortest) path with $m = 3$ steps (left), a piecewise linear path with $m = 100$ steps (middle), and a polynomial path of degree 3 (right).



All having the same signature with the skyline path up to level three.

More examples

Pfeffer-Seigal-Sturmfels: the (shortest) path with $m = 5$ steps (left), a piecewise linear path with $m = 100$ steps (middle), and a polynomial path of degree 5 (right).



The first two have the same signature with the Klee-Minty path up to level three. The third one is close but not equal.

Uniqueness

Chen (1950's): continuously differentiable curves (when parametrised at unit speed) are determined by their signatures.

Hambly-Lyons: paths of finite lengths are uniquely determined by their signatures up to tree-like equivalence.

If α, β are two paths of finite lengths, then $\mathrm{Sig}(\alpha) = \mathrm{Sig}(\beta)$ if and only if $\alpha * \beta^{-1}$ is equivalent to a null path.

Finite length paths can have very subtle tree-cancellations, while there is no such nontrivial equivalence if the curve is continuously differentiable (when parametrised at unit speed).

Boedihardjo-Geng-Lyons-Yang: uniqueness for rough paths.

Question: how to reconstruct the reduced path from its signature?

Uniqueness and reconstruction

Uniqueness with semi-constructive proofs:

- Le Jan-Qian: Brownian motion sample paths.
- Boedihardjo-Geng: more general Gaussian processes.
- Geng: deterministic rough paths.

Inversion for axis paths

These are paths that move parallel to Euclidean axes. They have the form

$$\gamma = r_1 e_{i_1} * r_2 e_{i_2} * \cdots * r_N e_{i_N} \ .$$

Information to recover: ordered directions $(e_{i_1}, \dots, e_{i_N})$ and length of each step $(r_1, \dots, r_N)$.

Observation:

- $w = e_{i_1} e_{i_2} \cdots e_{i_N}$ is square-free (no two adjacent letters are the same), and $C(w) = r_1 r_2 \cdots r_{i_N} \neq 0$.
- If w' is any other square-free word with length $|w'| \geq N$, then $C(w') = 0$.

Conclusion: there exists a unique longest square free word w such that $C(w) \neq 0$, then this word tells the ordered directions of the path movement.

Inversion for axis paths

These are paths that move parallel to Euclidean axes. They have the form

$$\gamma = r_1 e_{i_1} * r_2 e_{i_2} * \cdots * r_N e_{i_N} \ .$$

To recover the lengths, let $w = e_{i_1} \cdots e_{i_N}$ be the unique longest square-free word as above, and define

$$w_k := e_{i_1} \cdots e_{i_k}^2 \cdots e_{i_N} \ .$$

Then

$$C(w_k) = \frac{1}{2} r_1 \cdots r_k^2 \cdots r_N \quad \Rightarrow \quad r_k = \frac{2C(w_k)}{C(w)} \ .$$

Inversion for axis paths

These are paths that move parallel to Euclidean axes. They have the form

$$\gamma = r_1 e_{i_1} * r_2 e_{i_2} * \cdots * r_N e_{i_N} \; .$$

1. Find the unique longest square-free word with non-zero coefficient. This word tells the ordered directions of the path movement.
2. Move one level up and compare the coefficients to recover the length of each step.

Rely on special structures of the lattice.

Pfeffer-Seigal-Sturmfels: reconstruct paths that arise from a *fixed dictionary*.

Main reconstruction theorem

Theorem (Lyons, X.)

For every k, by using $Sig(\gamma)$ up to level $N = \mathcal{O}(k^3 \log k)$, we explicitly construct a piecewise linear path $\widetilde{\gamma}$ with k pieces such that

$$\sup_{u \in [0,1]} \left| \widetilde{\gamma}'(u) - \gamma'(u) \right| < \varepsilon_k$$

when both are parametrised at unit speed (with respect to ℓ^1 norm), and $\varepsilon_k \to 0$ as $k \to +\infty$, the speed depending on modulus of continuity of γ'.

The error $\varepsilon_k = \mathcal{O}(k^{-\frac{\alpha^2}{2}})$ if $\gamma \in \mathcal{C}^{1,\alpha}$.

Commutative in smaller scales; noncommutative in larger scales.

Key: how to rule out noncommutativity in small scales?

Main reconstruction theorem

We work with $\mathbf{R}^2$ for notational simplicity. The piecewise linear path $\widetilde{\gamma}$ has the form

$$\widetilde{\gamma} = \widetilde{\gamma}_1 * \cdots * \widetilde{\gamma}_k,$$

where

$$\widetilde{\gamma}_j = \frac{\widetilde{L}}{k}\Big(a_x^{(j)}\rho_j x + a_y^{(j)}(1-\rho_j)y\Big).$$

Hope: each $\widetilde{\gamma}_j$ approximates $\gamma_{[\frac{j-1}{k},\frac{j}{k}]}$ in the ℓ^1 sense.

- $\rho_j, 1-\rho_j \in [0,1]$ represents the unsigned direction;
- $a_x^{(j)}, a_y^{(j)} \in \{\pm 1\}$ represents the sign;
- $\widetilde{L} > 0$ approximates the ℓ^1 length.

Recovering the increment

Symmetrisation averages out the order.

Summing over all words of length n with k x's and $n-k$ y's:

$$\mathcal{S}(k, n-k) = n! \sum_{w \in \mathcal{W}_{k,n-k}} C(w) = \binom{n}{k} (\Delta x)^k (\Delta y)^{n-k}.$$

Maximum: $\frac{k^*}{n-k^*} \approx \frac{|\Delta x|}{|\Delta y|}$ $\Rightarrow$ recovers unsigned direction.

More robust way of doing it: find k^* such that

$$\sum_{k:\, |\frac{k}{n} - \frac{k^*}{n}| < \varepsilon} |\mathcal{S}(k, n-k)| \approx \sum_{k} |\mathcal{S}(k, n-k)|$$

There are more than one such k^*, but all of them are close to each other.

Move one level up: comparing $\mathcal{S}(k^*+1, n-k^*)$ and $\mathcal{S}(k^*, n-k^*)$ gives the sign of the x direction.

Symmetrisation

Symmetrising k blocks with block size $2n$:

$$\underbrace{*\,*\,*\,*\,*}_{2n}\, e_{i_1} \underbrace{*\,*\,*\,*\,*}_{2n}\, e_{i_2} \cdots\cdots e_{i_{k-1}} \underbrace{*\,*\,*\,*\,*}_{2n}.$$

Key: pattern in block j are roughly determined by $\gamma_{[\frac{j-1}{k}, \frac{j}{k}]}$.

Steps:

1. Recovering the unsigned directions by checking non-degeneracy.
2. Recovering the signs by moving one level up.
3. Recovering the length by a scaling argument.

Remark: only uses level $2nk + k - 1$ and $2nk + k$.

Summary

Consequences of the reconstruction:

1. Tail signatures already determine $\mathcal{C}^1$ paths.
2. 'Verification' that higher level signatures describe finer structures of the path.

   Quantitative description? Relevant lower bounds (for large n): Hambly-Lyons, Boedihardjo-Geng.

What have we learned:

1. Symmetrisation counts the frequency but neglects the order; so it gives local increments.
2. A certain non-degeneracy criterion is often needed in recovering the directions (Le Jan-Qian, Boedihardjo-Geng, Geng).

Some questions

1. Improve efficiency?

   Insertion algorithm by Chang-Lyons; optimisation scheme by Yang.

   Algebraic structures explored in Améndola-Friz-Sturmfels, Pfeffer-Seigal-Sturmfels.

2. Inversion for rough paths? (Geng)

3. Identify the image of the signatures in the tensor algebra.

   Expect to involve highly nontrival interplays between algebraic structure (group-like) and analytic properties (decay) $\rightsquigarrow$ no clue at this moment.

   More reasonable to start with monotone paths first.