# Inverting the signature of a path

Weijun Xu

Joint works with Terry Lyons

also with Jiawei Chang, Nick Duffield and Hao Ni

April 30, 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

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

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

$$\mathrm{Sig}(\gamma) := \sum_{n\geq 0} \underbrace{\sum_{w:|w|=n} C_\gamma(w) \cdot w}_{\mathrm{Sig}^{(n)}(\gamma)} .$$

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

## Examples

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

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

$$ax * by \mapsto \exp(ax)\exp(by) \,.$$

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

$$by * ax \mapsto \exp(by)\exp(ax) \,.$$

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

$$ax + by \mapsto \exp(ax + by) \,.$$

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

## Another example

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

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 $\alpha_n$ and $\beta_n$ 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.

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

# Uniqueness

Chen (1950's): continuously differentiable curves are determined by their signatures.

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

If $\alpha, \beta$ 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*

$$\sup_{u \in [0,1]} |\widetilde{\gamma}'_u - \gamma'_u| < \varepsilon_k$$

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

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 $\ell^1$ 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 $\ell^1$ 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 $2nk + k$.

## Probabilisitc interpretation

Terry told me the following probabilistic interpretation of the signature during my PhD.

Suppose $\gamma : [0,1] \to \mathbf{R}^2$ is monotone in the sense that $x_t' \geq 0$ and $y_t' \geq 0$ for all $t \in [0,1]$.

Think of the following Poisson process $(\mathcal{X}_t, \mathcal{Y}_t)_{t \in [0,1]}$:

- $\mathcal{X}_t$ generates letter $x$ with intensity $x_t'$; $\mathcal{Y}_t$ generates letter $y$ with intensity $y_t'$; simultaneously and independently.
- We arrange the letters in the order of their arrival time (up to time 1), getting a (random) word $\mathcal{W}$.
- For example, if there are 5 arrivals in total in $[0,1]$, say $x, y, y, x, y$ at times $0 \leq u_1 < u_2 < u_3 < u_4 < u_5 \leq 1$, then $\mathcal{W} = xyyxy$.

Probabilistic interpretation of signature:

$$C_\gamma(w) = e^L \, \mathrm{Pr}(\mathcal{W} = w) \;, \qquad L = \ell^1 \text{ length of } \gamma \;.$$

# Probabilistic interpretation

For monotone paths:

$$C_\gamma(w) = e^L \,\mathrm{Pr}(\mathcal{W} = w)$$

Chang-Duffield-Ni-X.: inversion for monotone paths.

General non-monotone paths? $x_t'$ and $y_t'$ can change signs.

- Poisson process $(\mathcal{X}_t, \mathcal{Y}_t)$ with intensities $|x_t'|$ and $|y_t'|$.
- Each letter of arrival also carries a sign: if $x$ arrives at time $u$, then $+1$ if $x_u' > 0$, and $-1$ if $x_u' < 0$. Same for $y$.
- Same random word $\mathcal{W}$ as before, but $\mathcal{W}$ also have a sign — the product of the signs of its letters.

Now, we have

$$C_\gamma(w) = e^L \mathbf{E}\big[\mathrm{sign}(\mathcal{W}) \cdot \mathbf{1}_{\mathcal{W}=w}\big]\;, \quad L = \ell^1 \text{ length of } \gamma\;.$$

## 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

   A reverse question: does a version of Bernstein's theorem hold?

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;

   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.