The Signature Method

Nikolas Tapia

NTNU Trondheim

Apr. 16th, 2019 @ Santiago, Chile

Goals

1. Signatures
    1. For paths on $\mathbb{R}^d$
    2. Some applications
    3. Geometric Rough Paths
2. Shape recognition
    1. Signatures on Lie groups

Signatures

Consider a d-dimensional vector space V and define

$$T(V) := \mathbb{R}1 \oplus V \oplus (V \otimes V) \oplus (V \otimes V \otimes V) \oplus \cdots .$$

For $p \geq 1$, the *degree p component* $T(V)_p = V^{\otimes p}$ is spanned by the set

$$\{e_{i_1 \cdots i_p} := e_{i_1} \otimes \cdots \otimes e_{i_p} : i_1, \ldots, i_p = 1, \ldots, d\}$$

In particular $\dim T(V) = \infty$.

For a given $\psi \in T(V)^* := T(\!(V)\!)$ we write

$$\psi = \sum_{p \geq 0} \sum_{i_1, \ldots, i_p = 1}^{d} \langle \psi, e_{i_1 \cdots i_p} \rangle e_{i_1 \cdots i_p}.$$

There are two products on $T(V)$:

1. the tensor product: $e_{i_1\cdots i_p}\otimes e_{i_{p+1}\cdots i_{p+q}}=e_{i_1\cdots i_{p+q}}\in T(V)_{p+q}$ and,
2. the shuffle product:

$$e_{i_1\cdots i_p}\sqcup\!\sqcup\, e_{i_{p+1}\cdots i_{p+q}}=\sum_{\sigma\in\mathrm{Sh}(p,q)}e_{i_{\sigma(1)}i_{\sigma(2)}\cdots i_{\sigma(p+q)}}\in T(V)_{p+q}.$$

Examples:

$$e_i\sqcup\!\sqcup\, e_j=e_{ij}+e_{ji},\quad e_i\sqcup\!\sqcup\, e_{jk}=e_{ijk}+e_{jik}+e_{jki}.$$

On both cases $1\in T(V)_0$ acts as the unit.

The shuffle algebra carries a coalgebra structure: define $\Delta\colon T(V) \to T(V) \dot{\otimes} T(V)$ by

$$\Delta e_{i_1 \cdots i_p} := e_{i_1 \cdots i_p} \dot{\otimes} 1 + 1 \dot{\otimes} e_{i_1 \cdots i_p} + \sum_{j=1}^{p-1} e_{i_1 \cdots i_j} \dot{\otimes} e_{i_{j+1} \cdots i_p}.$$

This structure is dual to the tensor product in the sense that if $\varphi, \psi \in T(\!(V)\!)$ then

$$\varphi \otimes \psi = \sum_{p \geq 0} \sum_{i_1, \ldots, i_p = 1}^{d} \langle \varphi \dot{\otimes} \psi, \Delta e_{i_1 \cdots i_p} \rangle e_{i_1 \cdots i_p}.$$

Let $x\colon [0,1] \to \mathbb{R}^d$ be a curve of bounded variation.

Definition

Its signature over the interval $[s,t] \subset [0,1]$ *is the tensor series with coefficients*

$$\langle S(x)_{s,u}, 1\rangle := 1, \quad \langle S(x)_{s,t}, e_{i_1\cdots i_p}\rangle := \int_s^t \langle S(x)_{s,u}, e_{i_1\cdots i_{p-1}}\rangle \, \mathrm{d}x_u^{i_p}.$$

Example:

$$\langle S(x)_{s,t}, e_i\rangle = \int_s^t \mathrm{d}x_u^i = x_t^i - x_s^i, \quad \langle S(x)_{s,t}, e_{ij}\rangle = \int_s^t \int_s^u \mathrm{d}x_v^i \mathrm{d}x_u^j.$$

In total:

$$S(x)_{s,t} = 1 + \int_s^t \mathrm{d}x_u^i \, e_i + \int_s^t \int_s^{u_2} \mathrm{d}x_{u_1}^i \mathrm{d}x_{u_2}^j \, e_{ij} + \iiint\limits_{s<u_1<u_2<u_3<t} \mathrm{d}x_{u_1}^i \mathrm{d}x_{u_2}^j \mathrm{d}x_{u_3}^k \, e_{ijk} + \cdots$$

Chen (1954) shows that $S(x)$ satsifies:

1. the shuffle relation:

$$\langle S(x)_{s,t}, e_{i_1\dots i_p} \sqcup\!\sqcup\, e_{i_{p+1}\cdots i_{p+q}}\rangle = \langle S(x)_{s,t}, e_{i_1\dots i_p}\rangle\langle S(x)_{s,t}, e_{i_{p+1}\cdots i_{p+q}}\rangle.$$

2. Chen's rule: for any $s < u < t$, we have

$$S(x)_{s,t} = S(x)_{s,u} \otimes S(x)_{u,t}.$$

3. If y is another path and $x \cdot y$ is their concatenation then

$$S(x \cdot y)_{s,t} = S(x)_{s,t} \otimes S(y)_{s,t}.$$

The shuffle identity generalizes integration by parts.

$$
\begin{aligned}
\langle S(x)_{s,t}, e_{ij} + e_{ji} \rangle &= \int_s^t \int_s^u \mathrm{d}x_v^i \mathrm{d}x_u^j + \int_s^t \int_s^u \mathrm{d}x_v^j \mathrm{d}x_u^i \\
&= \int_s^t (x_u^i - x_s^i)\, \mathrm{d}x_u^j + \int_s^t (x_u^j - x_s^j)\, \mathrm{d}x_u^i \\
&= (x_t^i - x_s^i)(x_t^j - x_s^j) \\
&= \langle S(x)_{s,t}, e_i \rangle \langle S(x)_{s,t}, e_j \rangle .
\end{aligned}
$$

Chen's rule generalizes the splitting of integrals.

$$
\begin{aligned}
\langle S(x)_{s,t}, e_i \rangle &= \int_s^t \mathrm{d}x_v^i \\
&= \int_s^u \mathrm{d}x_v^i + \int_u^t \mathrm{d}x_v^i \\
&= \langle S(x)_{s,u} \otimes S(x)_{u,t}, e_i \rangle.
\end{aligned}
$$

$$
\begin{aligned}
\langle S(x)_{s,t}, e_{ij}\rangle &= \int_s^t (x_v^i - x_s^i)\,\mathrm{d}x_v^j \\
&= \int_s^u (x_v^i - x_s^i)\,\mathrm{d}x_v^j + \int_u^t (x_v^i - x_s^i)\,\mathrm{d}x_v^j \\
&= \int_s^u (x_v^i - x_s^i)\,\mathrm{d}x_v^j + \int_u^t (x_v^i - x_u^i)\,\mathrm{d}x_v^j + (x_u^i - x_s^i)(x_t^j - x_u^j) \\
&= \langle S(x)_{s,u} \otimes S(x)_{u,t}, e_{ij}\rangle
\end{aligned}
$$

Signatures can be easily computed for certain paths.

If x is a straight line, i.e. $x_t = a + bt$ with $a, b \in \mathbb{R}^d$ then

$$\langle S(x)_{s,t}, e_{i_1\cdots i_p}\rangle = \frac{(t-s)^p}{p!}\prod_{j=1}^{p} b_{i_j}.$$

Indeed

$$\begin{aligned}\langle S(x)_{s,t}, e_{i_1\cdots i_p}\rangle &= \int_s^t \frac{(u-s)^{p-1}}{(p-1)!}\prod_{j=1}^{p-1} b_{i_j} b_p \,\mathrm{d}u \\ &= \frac{(t-s)^p}{p!}\prod_{j=1}^{p} b_{i_j}.\end{aligned}$$

Therefore

$$S(x)_{s,t} = 1 + (t-s)b + \frac{(t-s)^2}{2} b \otimes b + \frac{(t-s)^3}{6} b \otimes b \otimes b + \cdots = \exp_\otimes((t-s)b).$$

By Chen's rule, if x is a general piecewise linear path with slopes $b_1, \ldots, b_m \in \mathbb{R}^d$ between times $s < t_1 < \cdots < t_{m-1} < t$ then

$$S(x)_{s,t} = \exp_\otimes((t_1 - s)b_1) \otimes \cdots \otimes \exp_\otimes((t - t_{m-1})b_m).$$

Some further properties:

1. Invariant under reparametrization: if φ is an increasing diffeomorphism on $[0, 1]$ then
$$S(x \circ \varphi)_{s,t} = S(x)_{s,t}.$$
2. Characterizes the path up-to *irreducibility*. If $S(x) = S(y)$ for two irreducible paths then y is a translation of x.

The signature takes values on a group G with $\otimes$ as composition.

In practice, a suitable truncation of $S(x)$ is considered, and this also belongs to a group G_m, $m > 1$.

The basic workflow for signatures in applications is to convert data streams into paths.

This can be done in several ways: linear interpolation, axis paths, lead-lag transforms, cumulative sums, etc...

One also has to choose the truncation level.

There is some redundancy in the signature due to the shuffle relations. A more efficient approach is to work with the so-called *log-signature*.

Applying some transformations one can read off some information from the signature:

1. Mean
2. Quadratic variation, i.e. variance

For Machine Learning applications, levels of the signature are selected as explanatory variables for the features of a path. An example of objective function (taken from Gyurkó, Lyons, Kontkowski & Field; 2014)

$$\min_{\beta}\left[\sum_{k=1}^{L}\left(\sum_{|w|\leq M}\beta_w\langle S(x_k)_{0,1},w\rangle - y_k\right)^2 + \alpha\sum_{|w|\leq M}|\beta_w|\right]$$

This framework has been applied to

1. finanical data streams,
2. sound compression (Lyons & Sidorova, 2005),
3. chinese character recognition (Graham, 2013; Lianwen, Weixin & Manfei, 2015),
4. pattern recognition in MEG scans (Gyurkó, Lyons & Oberhauser, 2014) and
5. behavioural patterns of patients with bipolar disorder.

Rough Paths

Definition (Lyons (1998))

A rough path of roughness $\theta > 1$ is a map $X\colon [0,1]^2 \to T(\!(V)\!)_{\leq m}$ such that $X_{s,t} = X_{s,u} \otimes X_{u,t}$ and

$$|\langle X_{s,t}, e_{i_1\cdots i_p}\rangle| \leq C_p|t-s|^{p/\theta}, \quad p < m$$

where $m := \lceil \theta \rceil$.

The (trucated) signature is the "canonical lift" of a path of bounded variation to a rough path of roughness θ.

Theorem (Lyons (1998))

Any path $X\colon [0,1]^2 \to T(\!(V)\!)_{\leq m}$ satisying Chen's rule and the analytic bound admits a unique extension $\hat{X}\colon [0,1] \to T(\!(V)\!)$ with the same properties.

Definition (Lyons (1998))

A geometric rough path of roughness θ is the limit of canonical lifts of bounded variation paths in a certain θ-variation metric.

Geometric rough paths are G_m valued, where again $m := \lceil \theta \rceil$.

Definition (Friz–Victoir (2006))

A weakly-geometric rough path of roughness θ is a G_m-valued path of finite θ-variation.

Geometric rough paths provide a "universal" description of flows controlled by x.

For a 1-dimensional smooth path x, consider the controlled differential equation

$$\dot{y}_t = V(y_t)\dot{x}_t.$$

To first order we have

$$y_t - y_s = V(y_s)\int_s^t \dot{x}_u \,\mathrm{d}u + o(|t-s|)$$

To second order

$$y_t - y_s = V(y_s)\int_s^t \dot{x}_u \,\mathrm{d}u + V'(y_s)V(y_s)\int_s^t\int_s^u \dot{x}_v\dot{x}_u \,\mathrm{d}v\mathrm{d}u + o(|t-s|^2).$$

We know that Brownian motion a.s. has finite θ-variation for any $\theta > 2$.

For $2 < \theta < 3$ and any *fixed* realization

$$X_{s,t} = 1 + (B^i_t - B^i_s)\ e_i + \int_s^t (B^i_u - B^i_s) \circ \mathrm{d}B^j_u\ e_{ij}$$

is a weakly-geometric rough path over B.

Shape analysis

Shapes are modeled as unparametrized curves, i.e. equivalence classes of elements under the action of the orientation-preserving diffeomorphism group of a fixed interval.

The similarity between two shapes $[c]$ and $[c']$ is defined by a distance d_S on shape space, defined as

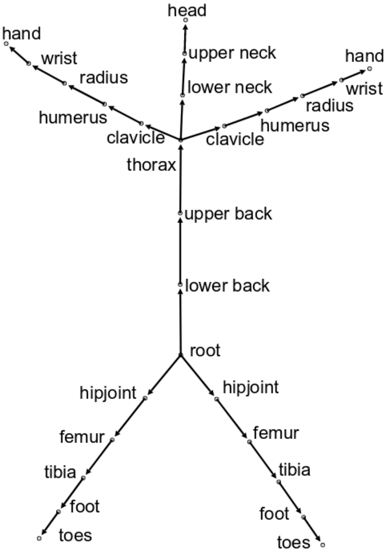
$$d_S([c],[c']) := \inf_{\varphi} d_P(c, c' \circ \varphi).$$

A possible choice is the *elastic metric*

$$d_P(c,c') := \sqrt{\int_I \|R(c)_t - R(c')_t\|^2 \,\mathrm{d}t}$$

where $R(c)_t := \frac{(L_{c_t}^{-1})_*(\dot{c}_t)}{\sqrt{\|\dot{c}_t\|}}$ is the Square-Root Velocity Transform (SRVT).

Motion capture records a set of 3 Euler angles for some of the actor's joints.



For each joint recording we have a path in $SO(3)$.

The goal is to cluster these recordings according to some labels, e.g. "walk", "run", "jump", etc. . .

Current methods use the SRVT and the elastic metric to do this.

These are computationally demanding (dynamic programming) and in the end we throw away the solution to the optimization problem.

Let G be a d-dimensional Lie group with Lie algebra $\mathfrak{g}$.

The Maurer–Cartan form on G is the pushforward of left translation:

$$\omega_g(v) = (L_{g^{-1}})_* v, \quad v \in T_g G.$$

It is a $\mathfrak{g}$-valued 1-form, i.e. a smooth section of $(M \times \mathfrak{g}) \otimes T^*G$. In other words, ω_g maps $T_g G$ into $\mathfrak{g}$. In particular, it can be written as

$$\omega = X_1 \otimes \omega^1 + \cdots + X_d \otimes \omega^d$$

where $\omega^1, \ldots, \omega^d$ are suitable 1-forms on G and $X_1, \ldots, X_d$ is a basis of $\mathfrak{g}$.

Chen defines the signature over the interval $[s,t]$ of a smooth curve $\alpha\colon[0,1]\to G$ as the tensor series $S(\alpha)_{s,t}$ with coefficients

$$\langle S(\alpha)_{s,t},1\rangle := 1,\quad \langle S(\alpha)_{s,t},e_{i_1\dots i_p}\otimes e_j\rangle := \int_s^t \langle S(\alpha)_{s,u},e_{i_1\cdots i_p}\rangle\omega^j_{\alpha_u}(\dot\alpha_u)\,\mathrm{d}u.$$

When $G=\mathbb{R}^d$ this definition coincides with the previous one by observing that $\omega^i=\mathrm{d}x^i$, i.e.

$$\omega_{\alpha_t}(\dot\alpha_t)=\dot\alpha_t^1 e_1+\cdots+\dot\alpha_t^d e_d$$

with $e_1,\dots,e_d$ the canonical basis of $\mathbb{R}^d$.

An example: let $G = H_3$ be the Heisenberg group, that is,

$$H_3 := \left\{ \begin{pmatrix} 1 & x & z \\ 0 & 1 & y \\ 0 & 0 & 1 \end{pmatrix} : x, y, z \in \mathbb{R} \right\}.$$

Its Lie algebra $\mathfrak{h}_3$ is spanned by the matrices

$$X := \begin{pmatrix} 0 & 1 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix}, \quad Y := \begin{pmatrix} 0 & 0 & 0 \\ 0 & 0 & 1 \\ 0 & 0 & 0 \end{pmatrix}, \quad Z := \begin{pmatrix} 0 & 0 & 1 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix}$$

with $[X, Y] = Z$, $[X, Z] = [Y, Z] = 0$.

In this group, the Maurer–Cartan form is given by

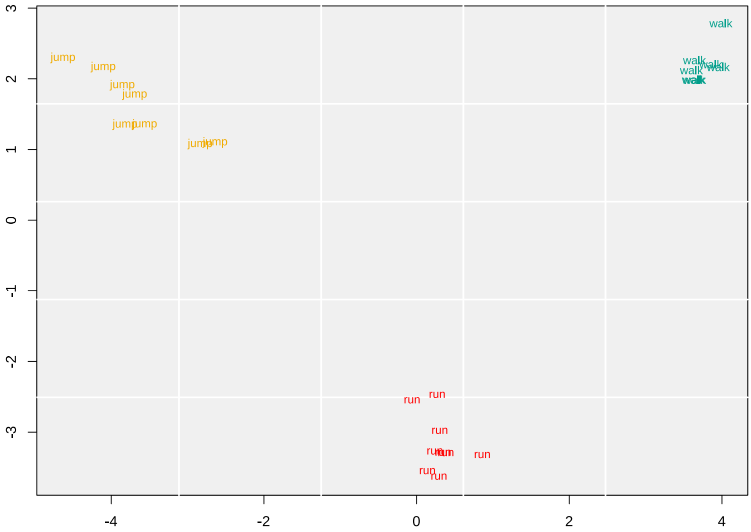
$$\omega_g = \begin{pmatrix} 0 & \mathrm{d}x & \mathrm{d}z - x\mathrm{d}y \\ 0 & 0 & \mathrm{d}y \\ 0 & 0 & 0 \end{pmatrix} \text{ when } g = \begin{pmatrix} 1 & x & z \\ 0 & 1 & y \\ 0 & 0 & 1 \end{pmatrix}.$$

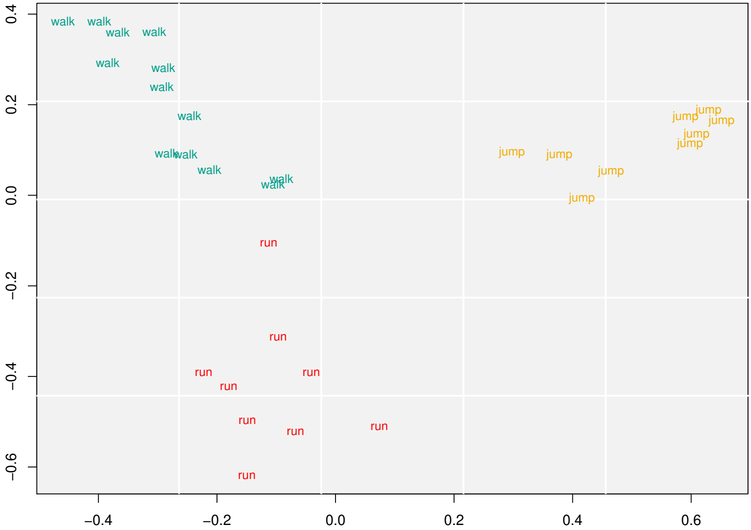
In particular

$$S(\alpha)_{s,t} = 1 + \int_s^t \dot\alpha_u^x \,\mathrm{d}u \; e_1 + \int_s^t \dot\alpha_u^y \,\mathrm{d}u \; e_2 + \int_s^t (\dot\alpha_u^z - \alpha_u^x \dot\alpha_u^y) \,\mathrm{d}u \; e_3 + \cdots \in T(\!(\mathbb{R}^3)\!)$$

where

$$\alpha_t = \begin{pmatrix} 1 & \alpha_t^x & \alpha_t^z \\ 0 & 1 & \alpha_t^y \\ 0 & 0 & 1 \end{pmatrix}.$$


jump
walk
run
-4
-2
0
2
4
-3
-2
-1
0
1
2
3



walk
run
jump
0.4
0.2
0.0
−0.2
−0.4
−0.6
−0.4
−0.2
0.0
0.2
0.4
0.6

Gracias!