Algebraic aspects of signatures

Nikolas Tapia based on joint work with J. Diehl & K. Ebrahimi-Fard
Introduction

Classification of time series is a very active field of research.

Most methods rely on extraction of features.

Signatures\(^a,b\) provide features that are interesting for a number of applications.

Also useful for other tasks such as analysing control systems, pathwise solutions to Stochastic Differential Equations, among others.


Continuous-time signatures

Let \( X : [0, 1] \to \mathbb{R} \) \textbf{continuous} path.

**Definition**

Given \( p \geq 1 \), define the \( p \)-variation of \( X \) over the interval \([s, t] \subseteq [0, 1]\) by

\[
\| X \|_{p;[s,t]} := \left( \sup_{\pi \in \mathcal{P}[s,t]} \sum_{[u,v] \in \pi} |X_v - X_u|^p \right)^{1/p}.
\]

The space of all paths such that \( \| X \|_{p;[s,t]} < \infty \) is denoted by \( V^p([s, t]) \).

Can be generalized to functions \( \Xi : [0, 1]^2 \to \mathbb{R} \) by replacing the increment \( X_v - X_u \) by \( \Xi_{u,v} \).

This generalization is an essential piece in T. Lyon's theory of \textit{rough paths}.\(^a\)

Young integration

Theorem (Young\textsuperscript{a})

Let $X \in V^p$, $Y \in V^q$ with $\frac{1}{p} + \frac{1}{q} > 1$. The integral

$$\int_s^t Y_u \, dX_u := \lim_{|\pi| \to 0} \sum_{i=0}^{n(\pi)} Y_{t_i}(X_{t_{i+1}} - X_{t_i})$$

is well defined and

$$\left| \int_s^t Y_u \, dX_u - Y_s(X_t - X_u) \right| \leq C_{p,q} \|X\|_{p;[s,t]} \|Y\|_{q;[s,t]}.$$

In particular, the iterated integral $\int X \, dX$ is well defined as long as $X \in V^p$ for $1 \leq p < 2$.

More generally, if $X = (X^1, \ldots, X^d)$ takes values in $\mathbb{R}^d$ then the integrals $\int X^i \, dX^j$ are defined.

\textsuperscript{a}L. C. Young. „An inequality of the Hölder type, connected with Stieltjes integration“. In: Acta Mathematica 67.1 (1936), p. 251.
Signatures

Definition

The signature of the path $X : [0, 1] \to \mathbb{R}^d$ is the collection of iterated integrals

$$S(X)_{s,t} := 1 + \int_s^t dX_u + \int_s^t \int_s^u dX_v \otimes dX_u + \cdots + \int \cdots \int dX_{u_1} \otimes \cdots \otimes dX_{u_n} + \cdots$$

Theorem

The signature satisfies

1. Chen’s identity: $S(X)_{s,u} \otimes S(X)_{u,t} = S(X)_{s,t}$.
2. Reparametrization invariance: $S(X \circ \varphi)_{s,t} = S(X)_{s,t}$.
3. It is the unique solution to the fixed-point equation

$$S(X)_{s,t} = 1 + \int_s^t S(X)_{s,u} \otimes dX_u.$$
Additionally, the shuffle relations are satisfied:

\[ S(X)^I_{s,t} S(X)^J_{s,t} = \sum_{K \in \text{Sh}(I,J)} S(X)^K_{s,t}. \]

This introduces some redundancy, e.g.

\[ S(X)^{ij}_{s,t} = S(X)^i_{s,t} S(X)^j_{s,t} - S(X)^{ij}_{s,t}. \]

A way to compress the available information is to work with the so-called log-signature

\[ \Omega(X)_{s,t} := \log \otimes S(X)_{s,t} \in \mathcal{L}(\mathbb{R}^d). \]

\( \Omega(X) \) corresponds to a pre-Lie Magnus expansion w.r.t. the pre-Lie product

\[ X \triangleright Y := \int_s^t \int_s^u [dX_v, dY_u] \]
The map $I \mapsto S(X)^I_{s,t}$ defines a linear map from the tensor algebra to the reals. The shuffle relations then mean that this map is a character, i.e.

$$S(X)^I_{s,t} S(X)^J_{s,t} = S(X)^{I \shuffle J}_{s,t}$$

where the shuffle product is recursively defined, for $I = (i_1, \ldots, i_n)$, $J = (j_1, \ldots, j_m)$, by

$$I \shuffle J = (I' \shuffle J)i_n + (I \shuffle J')j_m$$

where $I' = (i_1, \ldots, i_{n-1})$, $J' = (j_1, \ldots, j_{m-1})$ and

$$S(X)^I_{s,t} = \int \cdots \int_{s<u_1<\cdots<u_n<t} dX_{u_1}^i \cdots dX_{u_n}^i.$$ 

Remark

We can think of $S(X)$ as a formal series

$$S(X)_{s,t} = \sum_I S(X)^I_{s,t} I.$$ 

Signatures

Why should we care?

1. Useful for the description of the solutions of controlled systems: if $\dot{Y}_t = V(Y_t) \dot{X}_t$ then
   $$Y_t - Y_s = \sum_I V_I(Y_s) S(X)_s^I + R_{s,t}.$$

2. Captures features of $X$, useful for applications to Machine Learning, pattern recognition, time series analysis, etc.

In principle hard to compute. However, if $X$ is piecewise linear then
$$S(X)_{s,t} = \exp\otimes(v_1) \otimes \cdots \otimes \exp\otimes(v_k)$$
and we can use the Baker–Campbell–Hausdorff formula.$^a$

---

Signatures

However:

1. For a one-dimensional signal:

\[
\int \cdots \int_{s < u_1 < \cdots < u_n < t} dX_{u_1} \cdots dX_{u_n} = \frac{(X_t - X_s)^n}{n!}.
\]

This can be cured to some extent by introducing more dimensions\(^a\) and other tricks\(^b\).

2. In practice we are confronted with discrete data.

   This can also be avoided by interpolation.

3. A more severe problem is tree-like equivalence\(^c\).

We propose instead a new framework operating directly at the discrete level.


Discrete signatures

A composition of an integer $n$ is a sequence $(i_1, \ldots, i_k)$ with $i_1 + \cdots + i_k = n$.

**Definition (Gessel\textsuperscript{a})**

Given a composition $I = (i_1, \ldots, i_k)$ define

$$M_I(z) := \sum_{j_1 < j_2 < \cdots < j_k} z_{j_1}^i \cdots z_{j_k}^i.$$  

For example

$$M_{(1)}(z) = \sum_j z_j, \quad M_{(1,1)} = \sum_{j_1 < j_2} z_{j_1} z_{j_2}, \quad M_{(2)}(z) = \sum_j z_j^2.$$  

Note that

$$M_{(1)}(z)^2 = M_{(2)}(z) + 2M_{(1,1)}.$$  

The map $I \mapsto M_I(z)$ defines a linear map from compositions to the reals. The product rule above can be expressed as

$$M_I(z)M_J(z) = M_{I \star J}(z)$$

where the quasi-shuffle product\footnote{M. E. Hoffman. „Quasi-shuffle products“. In: *J. Algebraic Combin.* 11.1 (2000), pp. 49–68.} is recursively defined, for $I = (i_1, \ldots, i_n)$, $J = (j_1, \ldots, j_m)$, by

$$I \star J = (I' \star J)i_n + (I \star J')j_m + c(I, J)$$

where $I'$ and $J'$ are defined as before, and

$$c(I, J) := (i_1, \ldots, i_{n-1}, j_1, \ldots, j_{m-1}, i_n + j_m).$$

Definition

*Given a discrete time series $x = (x_0, x_1, \ldots, x_N)$, its discrete signature is*

$$DS(x)_{n,m} = \sum_I M_I(\Delta^m_n x)I$$

where $\Delta^m_n x = (x_{n+1} - x_n, \ldots, x_m - x_{m-1})$. 
Discrete signatures

Why should we care?

1. Can be used to analyse solutions of controlled recurrence equations of the form

\[ y_{k+1} = y_k + V(y_k)(x_{k+1} - x_k) \]

relevant e.g. for applications to Residual Neural Networks.\(^a\),\(^b\),\(^c\)

2. Invariant under time warping, useful for applications to time series classification.\(^d\)

3. No need to transform the data in any way. Even if \(x\) is one-dimensional we get more information, e.g.

\[ DS(x)_{0,N}^{(2)} = \sum_{j=1}^{N} (x_j - x_{j-1})^2 \neq (x_N - x_0)^2. \]

4. No need for BCH formula.

\(^a\)Current project with P. Friz (TU) and C. Bayer (WIAS)
We can actually count the number of invariants.

**Theorem (Diehl, Ebrahimi-Fard, T.; Novelli, Thibon[^1])**

The number of time-warping invariants of a $d$-dimensional time series has generating function

$$
G(t) := \sum_{n=0}^{\infty} c_n(d) t^n = \frac{(1 - t)^d}{2(1 - t)^d - 1} = 1 + dt + \frac{d(3d + 1)}{2} t^2 + \frac{d(13d^2 + 9d + 2)}{6} t^3 + \cdots.
$$

Compare with the corresponding generating function for the shuffle algebra:

$$
H(t) = \frac{1}{1 - dt} = 1 + dt + d^2 t^2 + d^3 t^3 + \cdots.
$$

Discrete signatures

Theorem (Diehl, Ebrahimi-Fard, T.)

Let \( x \) be a time series and define an infinite-dimensional path \( X = (X^I) \) where for a composition \( I \), \( X^I \) is the linear interpolation of the sequence

\[
n \mapsto M_I(\Delta_0^n x).
\]

Then

\[
S(X)^I_{0,N} = DS(x)^{\Phi(I)}_{0,N}
\]

where \( \Phi \) is Hoffman's isomorphism\(^a\).

In the one-dimensional case, the \textit{elementary symmetric functions}

\[
M_{(1,1,\ldots,1)}(\Delta x) = \sum_{j_1 < \cdots < j_n} \Delta x_{j_1} \cdots \Delta x_{j_n}
\]

arise as a left-point Riemann sum associated to a \textit{piecewise constant} interpolation of \( x \).

A few possible extensions:

1. Multi-parameter data, e.g. one-dimensional time series depending on two parameters.
2. General functions on increments, e.g.

\[ \sum_{j_1 < j_2} f_{i_1}(\Delta x_{j_1}) f_{i_2}(\Delta x_{j_2}). \]

And some questions and ongoing projects:

1. Understanding \( \log DS(x) \). Chow’s theorem.
2. Numerical experiments and use in time warping.\(^a\)
3. Robustness of Residual Neural Networks.
4. Learning dynamics of Stochastic Differential Equations.\(^b,c\)


\(^b\)with C. Bayer and M. Eigel (WIAS)

Thanks!