The discrete signature of a time series

Joscha Diehl (Universität Greifswald)

joint with Kurusch Ebrahimi-Fard (NTNU), Nikolas Tapia (TU Berlin), Max Pfeffer (MPI Leizpig)

10. May

Time-stretch invariants

Assume there is a discrete deterministic time-series

$$(X_1, X_2, .., X_M) \in \mathbb{R}^M,$$

that we want to know about.
But! We only get noisy observations

$$Y_n^{(\ell)} = X_n + W_n^{(\ell)}, \quad \ell = 1, \dots L,$$

where $W^{(\ell)}$ are iid samples of a random walk.

Of course

$$\frac{1}{L} \sum_{\ell=1}^{L} Y_m^{(\ell)} \to_{L\to\infty} X_m, \qquad m = 1, \dots, M.$$

So: if we observe often enough, we can recover X.

Time-stretch invariants

Now still assume $X \in \mathbb{R}^M$ unknown, but additionally we do not know the speed at which it is run.

To be specific,

$$Y_n^{(\ell)} = X_{\tau^{(\ell)}(n)} + W_n^{(\ell)}, \quad \ell = 1, \dots L, n = 1, \dots, N,$$

where

$$\tau^{(\ell)} : \{1, .., N\} \to \{1, .., M\},$$

are non-decreasing, surjective and unknown.

Time-stretch invariants

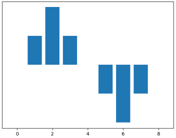


Fig: Original

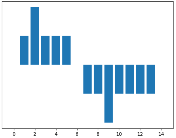


Fig: Time-stretched

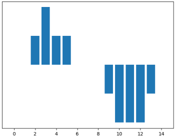


Fig: Time-stretched

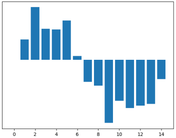


Fig: Time-stretched + noise

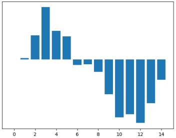


Fig: Time-stretched + noise

Time-stretch invariants

$$Y_n^{(\ell)} = X_{\tau^{(\ell)}(n)} + W_n^{(\ell)}, \quad \ell = 1, \ldots L, n = 1, \ldots, N.$$

How to recover X now ?

Current available method.

1. Align the different samples.
2. Average.

This works for large signal-to-noise ratio .

Time-stretch invariants

It was no chance of working for small signal-to-noise ratio.

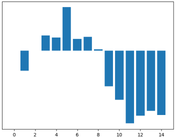


Fig: Time-stretched + noise

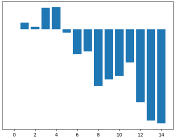


Fig: Time-stretched + noise

Our strategy

1. Calculate time-strecth invariant features of the time-series.
2. Average them. Law of large numbers $\rightsquigarrow$ noise disappears.
3. Invert the first step: find time-series that matches the averaged features.

Time-stretch invariants

First idea Use iterated-integrals signature on the linearly interpolated path,

$$\mathbf{Sig}(Y)_{0,N} = \left(1, \int_0^N dY, \int_0^N dY \otimes dY, \int_0^N dY \otimes dY \otimes dY, ..\right)$$
$$= \left(1, Y_{0,N}, \frac{1}{2!}(Y_{0,N})^2, \frac{1}{3!}(Y_{0,N})^3, \ldots\right)$$

For $d = 1$ one only gets one feature: the total displacement.

Time-stretch invariants

First idea: Use iterated-integrals signature on the linearly interpolated path,

$$\mathbf{Sig}(Y)_{0,N} = \left(1, \int_0^N dY, \int_0^N dY \otimes dY, \int_0^N dY \otimes dY \otimes dY, ..\right)$$
$$= \left(1, Y_{0,N}, \frac{1}{2!}(Y_{0,N})^2, \frac{1}{3!}(Y_{0,N})^3, \ldots\right)$$

For $d = 1$ one only gets one feature: the total displacement.

(There are ways to turn a one-dim time series into a multi-dim one though; more on this later.)

Instead: we look for all polynomials on time-series that are invariant in the desired sense.

Time-stretch invariants

Example

$$f\left(Y^{(\ell)}\right) := \sum_{n=2}^{N} \left(Y_n^{(\ell)} - Y_{n-1}^{(\ell)}\right)^2.$$

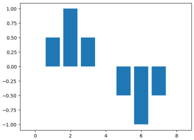
Then

$$\begin{aligned}
\mathbb{E}\left[f\left(Y^{(\ell)}\right)\right] &= \mathbb{E}\Big[\sum_{n=2}^{N}\left(X^{(\ell)}_{\tau^{(\ell)}(n)} - X^{(\ell)}_{\tau^{(\ell)}(n-1)}\right)^2 + 2\sum_{n=2}^{N}\left(X^{(\ell)}_{\tau^{(\ell)}(n)} - X^{(\ell)}_{\tau^{(\ell)}(n-1)}\right)\left(W_n^{(\ell)} - W_{n-1}^{(\ell)}\right) \\
&\quad + \sum_{n=2}^{N}\left(W_n^{(\ell)} - W_{n-1}^{(\ell)}\right)^2\Big] \\
&= \sum_{n=2}^{N}\left(X^{(\ell)}_{\tau^{(\ell)}(n)} - X^{(\ell)}_{\tau^{(\ell)}(n-1)}\right)^2 + (N-1)\cdot\sigma^2 \\
&= \sum_{m=2}^{M}\left(X_m - X_{m-1}\right)^2 + (N-1)\cdot\sigma^2.
\end{aligned}$$

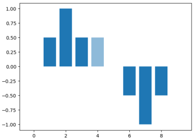
Time-stretch invariants

We now work on sequences of real numbers that eventually are zero, $Y \in \mathbb{R}_0^{\mathbb{N}}$. Define, $\mathrm{Still}_n : \mathbb{R}_0^N \to \mathbb{R}_0^N$ for example as

Still_4:



$\mapsto$



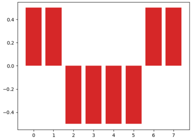
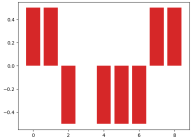
Relation to previous consideration: embed X in $\mathbb{R}_0^N$ and then $X_{\tau(\cdot)}$ can be realized as standing still a couple of times.

We call $F : \mathbb{R}_0^{\mathbb{N}} \to \mathbb{R}$ **invariant to standing still** if for all $Y \in \mathbb{R}_0^{\mathbb{N}}$, all $n \geq 1$

$$F(\mathrm{Still}_n(Y)) = F(Y).$$

Time-stretch invariants

It simplifies matters to think in terms of increments $y_i := Y_i - Y_{i-1}$. "Standing still" then becomes "inserting zeros".

Zero_4:  $\mapsto$ 

Definition

We call $G : \mathbb{R}_0^{\mathbb{N}} \to \mathbb{R}$ **invariant to inserting zeros** if for all $y \in \mathbb{R}_0^{\mathbb{N}}$, all $n \geq 1$

$$G(\mathrm{Zero}_n(y)) = G(y).$$

Time-stretch invariants

Lemma
All polynomial invariants to inserting zeros are given by the quasisymmetric functions

$$\sum_{i_1 < \cdots < i_p} y_{i_1}^{\alpha_1} \cdot \cdots \cdot y_{i_p}^{\alpha_p}, \quad p \geq 1, \alpha \in \mathbb{N}_{\geq 1}^{p}.$$

Example We have already seen $\alpha = (2)$.

$\alpha = (5, 7, 2)$ gives

$$\sum_{i_1 < i_2 < i_3} y_{i_1}^5 y_{i_2}^7 y_{i_3}^2.$$

Time-stretch invariants

Goal *Store these features in a* ***Hopf algebra structure****, as is done for the classical iterated-integrals signature.*

Let H be the space of formal inifinite linear combinations of integer compositions. Define

$$\textbf{DiscreteSig}(y) := \sum_{\alpha} \sum_{i_1 < \cdots < i_p} y_{i_1}^{\alpha_1} \cdot \cdots \cdot y_{i_p}^{\alpha_p} \cdot \alpha \in H$$

$$= () + \sum_{i_1} y_{i_1} \cdot (1) + \sum_{i_1} y_{i_1}^2 \cdot (2) + \sum_{i_1 < i_2} y_{i_1} y_{i_2} \cdot (1,1)$$

$$+ \sum_{i_1 < i_2} y_{i_1} y_{i_2}^2 \cdot (1,2) + \ldots$$

(*The Hopf algebra of quasisymmetric function was studied by Malvenuto/Reutenauer 1994.*)

Time-stretch invariants

Lemma (Chen's identity)

For $y, y' \in \mathbb{R}_0^{\mathbb{N}}$ let denote $y \sqcup y' \in \mathbb{R}_0^{\mathbb{N}}$ their concatenation. Then:

$$\mathbf{DiscreteSig}(y \sqcup y') = \mathbf{DiscreteSig}(y) \bullet \mathbf{DiscreteSig}(y').$$

Here $\bullet$ is the concatenation product on H. For example

$$(2,3,1) \bullet (7,4) = (2,3,1,7,4).$$

Time-stretch invariants

Lemma (Shuffle identity)

$$\Big\langle \alpha, \textbf{DiscreteSig}(y) \Big\rangle \cdot \Big\langle \beta, S(y) \Big\rangle = \Big\langle \alpha \overset{q}{\sqcup\!\sqcup} \beta, \textbf{DiscreteSig}(y) \Big\rangle,$$

Here $\overset{q}{\sqcup\!\sqcup}$ is the quasi-shuffle on H^*. For example

$$(1,2) \overset{q}{\sqcup\!\sqcup} (3) = (1,2,3) + (1,3,2) + (3,1,2) + (1,5) + (4,2).$$

So: just as for classical signature, the discrete signature is a character on some Hopf algebra.

What about Chow's theorem?

Recall:

Theorem (Chow's theorem for classical signature)

For every $L \in \mathfrak{g}$*, the free Lie algebra, for every* $n \geq 1$*, there exists a piecewise linear path* X *such that*

$$\mathrm{proj}_{\leq n} \mathrm{Signature}(X)_{0,1} = \mathrm{proj}_{\leq n} \exp(L).$$

Time-stretch invariants

This is not true here anymore!

To wit: up to degree 2 the Lie algebra of H is spanned by two vectors $L_{(1)}$ and $L_{(2)}$. The logarithm of the discrete signature up to degree 2 is given by

$$\log \mathbf{DiscreteSig}(y) = \sum_{i_1} y_{i_1} \cdot L_{(1)} + \sum_{i_1} y_{i_1}^2 \cdot L_{(2)}.$$

Since the coefficient of $L_{(2)}$ is non-negative, not every element of the Lie algebra can be reached!

(The problem seems to evaporate over $\mathbb{C}$ ...)

Multidimensional / Relation to other signatures

Time-stretch invariants

Multidimensional / Relation to other signatures

Open questions / Observations

Multidimensional / Relation to other signatures

For a times-series in $y \in (\mathbb{R}^d)^{\mathbb{N}_0}$ something similar works.
Let us look at the first few terms of the discrete signature for $d = 2$. Introduce commuting variables a_1, a_2, then

$$\begin{aligned}
\textbf{DiscreteSig}(y) = {} & () + \sum_{i_1} y^{(1)}_{i_1} (a_1) + \sum_{i_1} y^{(2)}_{i_1} (a_2) \\
& + \sum_{i_1} (y^{(1)}_{i_1})^2 (a_1^2) + \sum_{i_1} y^{(1)}_{i_1} y^{(2)}_{i_1} (a_1 a_2) + \sum_{i_1} (y^{(2)}_{i_1})^2 (a_2^2) \\
& + \sum_{i_1 < i_2} y^{(1)}_{i_1} y^{(1)}_{i_2} (a_1, a_1) + \sum_{i_1 < i_2} y^{(1)}_{i_1} y^{(2)}_{i_2} (a_1, a_2) \\
& + \sum_{i_1 < i_2} y^{(2)}_{i_1} y^{(1)}_{i_2} (a_2, a_1) + \sum_{i_1 < i_2} y^{(2)}_{i_1} y^{(2)}_{i_2} (a_2, a_2) + \dots
\end{aligned}$$

Then: Chen's lemma ✓, shuffle identity ✓.
(It fits nicely into the algebraic framework of quasi-shuffle algebras of *Hoffman 2000*.)

Multidimensional / Relation to other signatures

The discrete signature contains all (polynomial) time-stretch invariants, so it must contain the classical signature.

Denote $\mathbf{Sig}(X)$ the classical signature of the linearly inerpolated path.

Lemma

There exists a map

$$\Phi : H \to T((\mathbb{R}^d))$$

such that

$$\mathbf{Sig}(X) = \Phi(\mathbf{DiscreteSig}(\Delta X))$$

Remark 1. It is (the dual of) the isomorphism of Hoffman.

2. The other direction is not possible (**DiscreteSig** contains strictly more information).

Multidimensional / Relation to other signatures

For one-dimensional signals, the classical signature is not very interesting. There exist several ways to enhance a $1d$ curve to a multidim curve though:

1. Add time. (Destroys time-stretch invariance.)
2. Add 1-variation. (Not poynomial.)
3. Lead-lag procedure of Flint/Hambly/Lyons 2016.

Lemma

1. There is a map from our discrete signature to the lead-lag signature.
2. For $d \geq 2$ (the logarithm of) the lead-lag signature contains redundant terms.

Multidimensional / Relation to other signatures

For one-dimensional signals, the classical signature is not very interesting. There exist several ways to enhance a $1d$ curve to a multidim curve though:

1. Add time. (Destroys time-stretch invariance.)
2. Add 1-variation. (Not poynomial.)
3. Lead-lag procedure of Flint/Hambly/Lyons 2016.

Lemma

1. There is a map from our discrete signature to the lead-lag signature.
2. For $d \geq 2$ (the logarithm of) the lead-lag signature contains redundant terms.

Open questions / Observations

Time-stretch invariants

Multidimensional / Relation to other signatures

Open questions / Observations

Open questions / Observations

Open questions

- Weaker statement of Chow type (image is Zariski dense, image has positive measure, ..).
- Inverting **DiscreteSig** numerically.
- Multi-parameter case.

Open questions / Observations

Observation

Sample a continuous path X discretely $\rightsquigarrow X^n$. Then, if X is smooth,

$$\mathbf{DiscreteSig}(X^n) \quad \text{“}\to\text{”} \quad \mathbf{Sig}(X).$$

In particular: for a one-dim signal in the limit there is <u>no</u> information (apart from the increment).

On the other hand: for a martginale X

$$\mathbf{DiscreteSig}(X^n) \quad \text{“}\to\text{”} \quad \mathbf{Sig}(X, \langle X \rangle),$$

which gives a lot more information.

What is going on here?

Open questions / Observations

Thank you!