Rough paths - streamed data

Terry Lyons

Mathematical Institute, Oxford
Alan Turing Institute

Hao Ni, Thomas Cass, Wiexin Yang, Imanol Perez, Paul More, Jiawei Yang, ….

Oxford
Mathematics

Rough paths - streamed data

Period of rapid change
- in our understanding and modelling of data
- in the role of optimisation
- In how we use information

A data scientist might ask what is left for mathematics to say?

The mathematics of information still has a role
- In this meeting we will see serious data challenges
- Where the mathematics of rough path theory can add value; it
  - removes an infinite dimensional invariance (gauge invariance)
  - and captures data more robustly
- Adds significant value to existing data science techniques

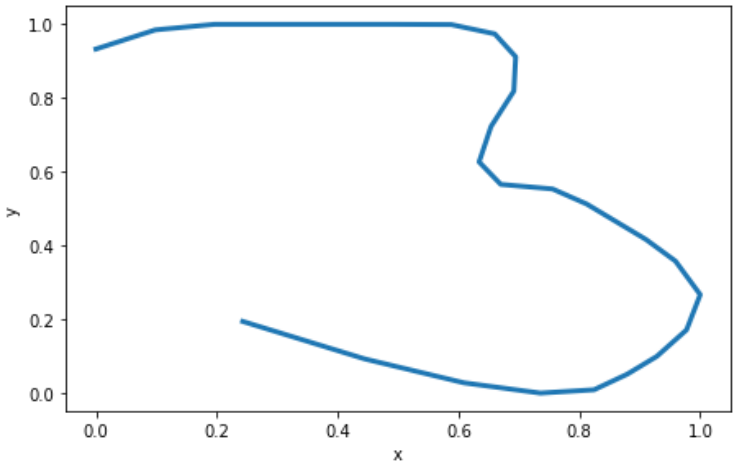
Rough paths - streamed data

- a character drawn on the screen of an iphone
- an order book
- a piece of text
- progression through hospital record
- astronomical data

- An evolving stream of emotions

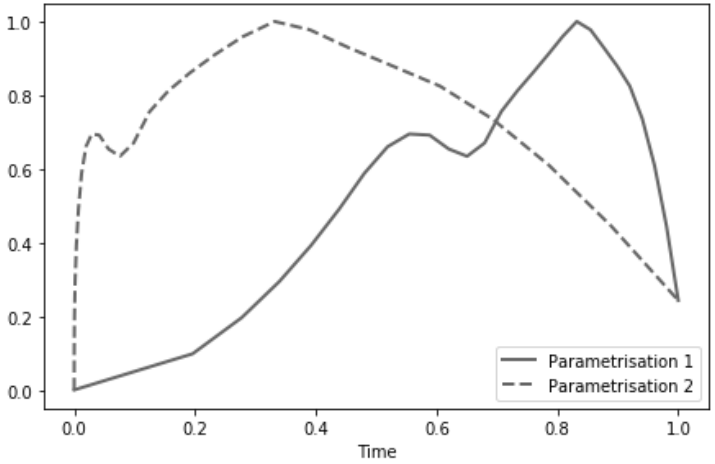
Rough paths: streamed data

– The letter “3” is drawn from top to bottom



Rough paths: streamed data

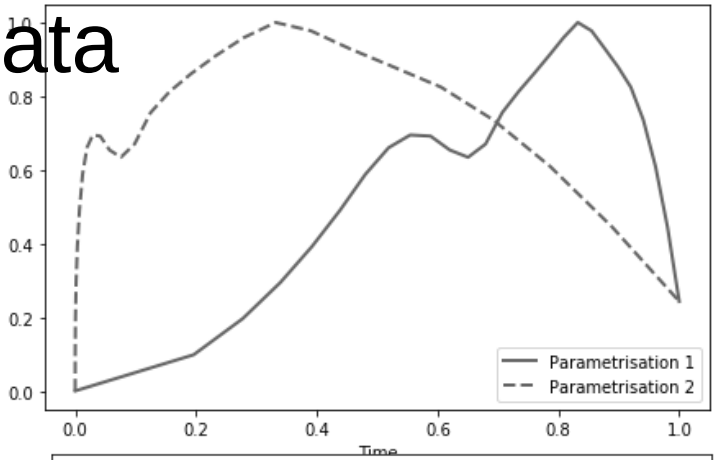
- The letter “3” is drawn from top to bottom
- The x coordinate of the evolving symbol plotted against time at different speeds
- This is the classic Kolmogorov approach. But it is very unstable.

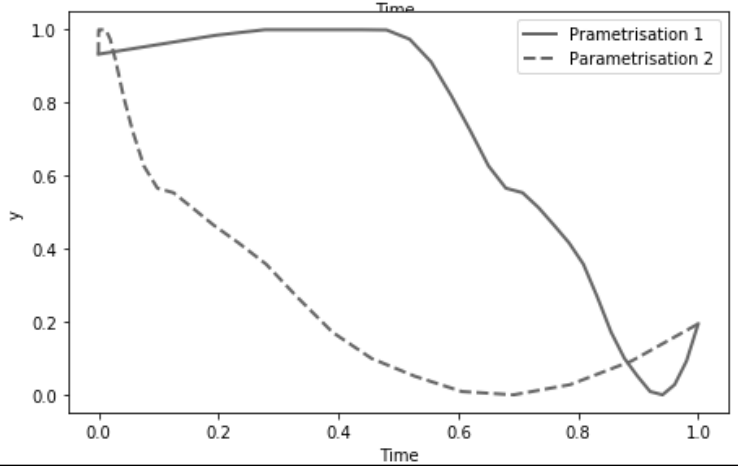


Rough paths: streamed data

- The number “3” x, y coordinates – same picture drawn at two different speeds

- This is the classic Kolmogorov approach. But it is very unstable.
- No consistent wavelets
- Reparameterizations do not form a linear space!

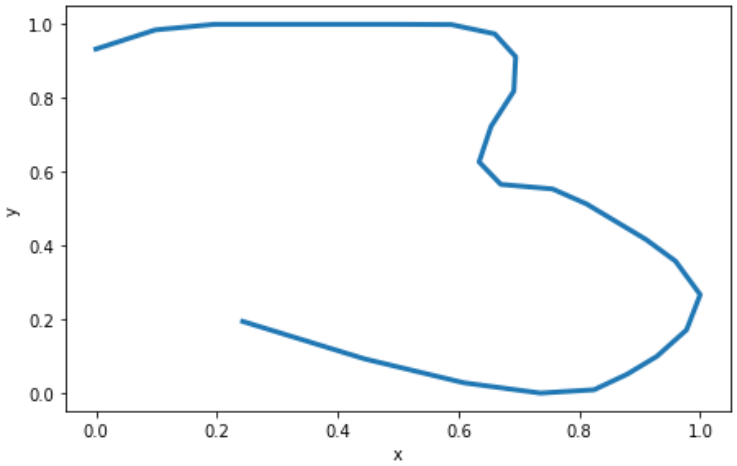




Rough paths: streamed data

- The letter “3” is drawn from top to bottom
- How does one describe the three or any path modulo the symmetry of parametrisation?

Paths and streams are of a varied type! No canonical parametrisation.



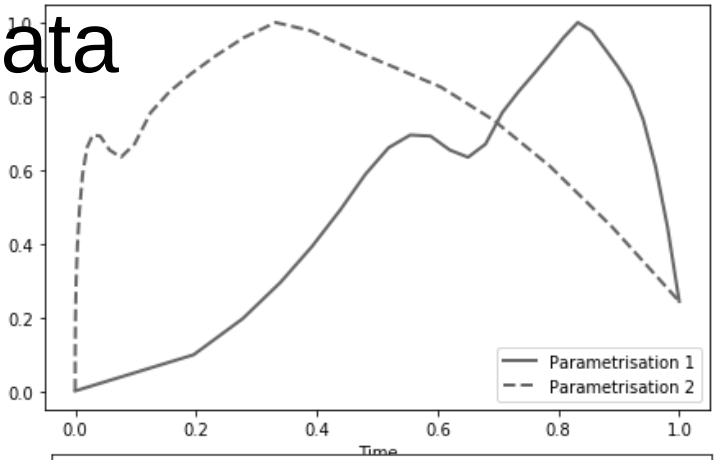
Rough paths: streamed data

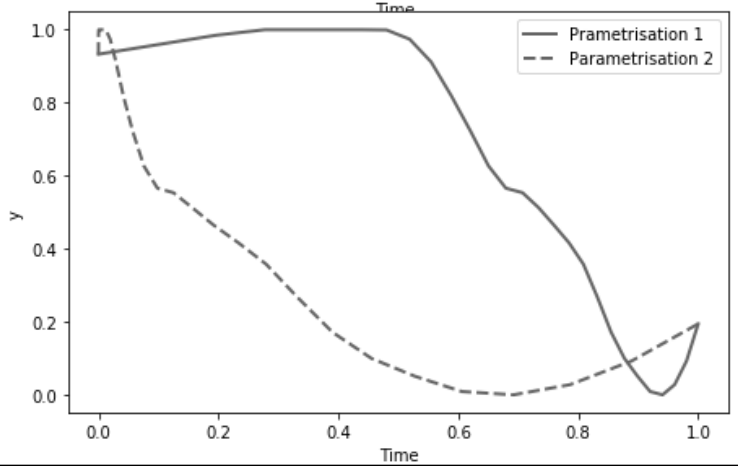
- Many typed of stream
- Reducing video to landmarks and pose is now feasible!
- The matchstick men and women are data streams in 30-75 dimensions
  - Still need to make sense of the data!
  - Starting times shift
  - Speed varies

http://mvig.sjtu.edu.cn/research/alphapose.html

Rough paths: streamed data

- How do we make a difference?
- Describe this multidimensional path in terms of its effects on certain nonlinear systems
- The first few terms in solution of $dS = S \otimes d\gamma$.





The signature of a path describes an unparameterized stream γ

Signature is a *top down* description for unparameterized paths that describes a path segment through its effects of stylised nonlinear systems

$$dS = S \otimes d\gamma$$

removing an infinite dimensional invariance allowing prediction and classification with *much* smaller learning sets.

gives fixed dimensional feature sets regardless of the sample points (missing data/common parameterisation not issues).

The signature – faithful and universal features describing an unparameterised stream

The signature of a stream γ over $I = [s, t]$ defined by

$\sum_{k=0}^{\infty} S_k$ where $S_0 = 1$ and

$$S_k(\gamma, I) := \iint_{s<u_1<\cdots<u_k<t} d\gamma_{u_1} d\gamma_{u_2} \dots d\gamma_{u_k}$$

These “Fourier-like” coefficients exactly describe the *un-parameterised* stream. (Hambly Lyons Annals Math 2010)

The signature of a stream – a faithful and universal feature set for a stream

Suppose γ is a stream or path $I = [s,t] \to E$ with values in $E := \mathbb{R}\langle e_1, \dots, \rangle$ a vector space, whose basis we refer to as the alphabet, and I is an interval of information. Then any word $\boldsymbol{e} = (\boldsymbol{e_{i_1}}, \dots, \boldsymbol{e_{i_n}})$ defines a real valued feature of the stream γ over I:

$$\begin{aligned} \boldsymbol{\phi_e}(\boldsymbol{\gamma}, \boldsymbol{I}) &:= \iint_{\boldsymbol{s<u_1<\cdots<u_k<t}} \langle \boldsymbol{e_{i_1}} | \boldsymbol{d\gamma_{u_1}} \rangle \langle \boldsymbol{e_{i_2}} | \boldsymbol{d\gamma_{u_2}} \rangle \dots \langle \boldsymbol{e_{i_k}} | \boldsymbol{d\gamma_{u_k}} \rangle \\ &= \langle \boldsymbol{e} | \boldsymbol{S_I} \rangle. \end{aligned}$$

Words are independent and !span! numerical functions on streams.

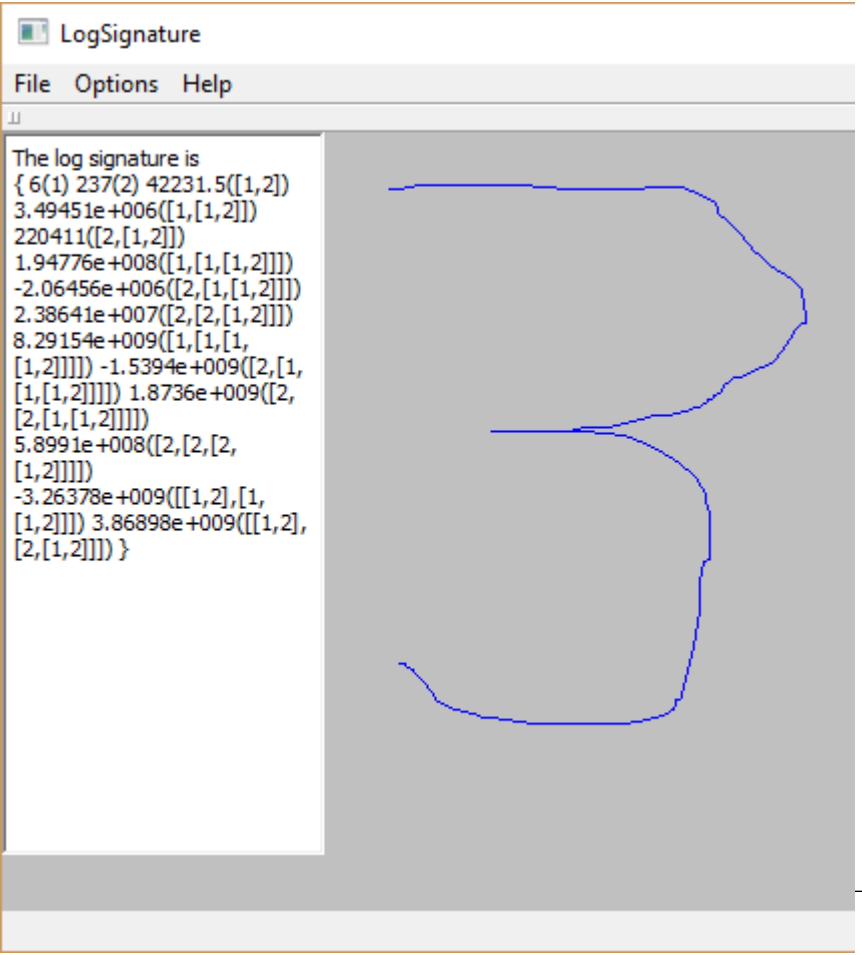
Analysis, Geometry, Combinatorial Hopf \ Dendriform \ Sensor Algebras

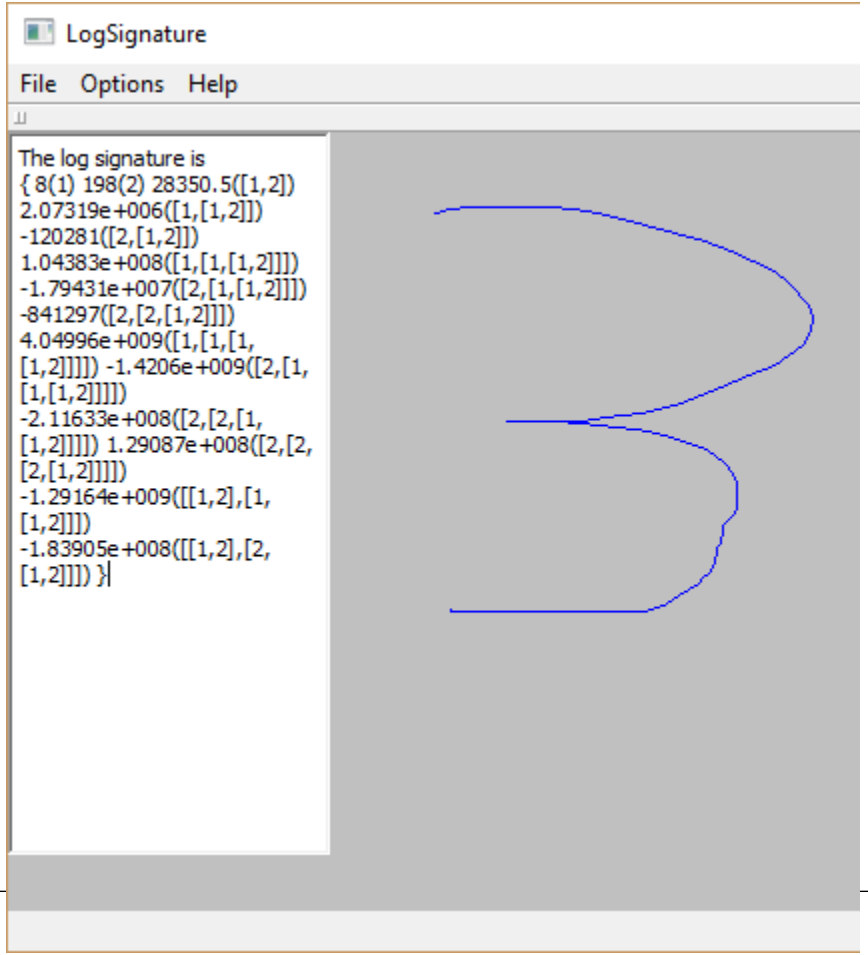
Signature leads to linear space of real valued functionals on streams

Pointwise multiplication and integration of these functionals

$\langle \alpha | \gamma \rangle \langle \beta | \gamma \rangle = \langle \alpha \uplus \beta | \gamma \rangle \qquad \int \langle \alpha | \gamma \rangle d \langle \beta | \gamma \rangle = \langle \alpha \prec \beta | \gamma \rangle$

can usefully be described in purely algebraic language

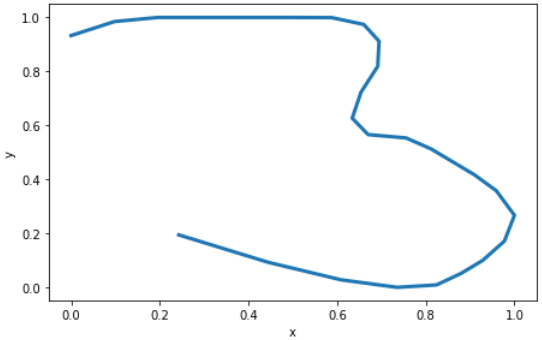

LogSignature
File Options Help
The log signature is
{ 6(1) 237(2) 42231.5([1,2])
3.49451e+006([1,[1,2]])
220411([2,[1,2]])
1.94776e+008([1,[1,[1,2]]])
-2.06456e+006([2,[1,[1,2]]])
2.38641e+007([2,[2,[1,2]]])
8.29154e+009([1,[1,[1,
[1,2]]]]) -1.5394e+009([2,[1,
[1,[1,2]]]]) 1.8736e+009([2,
[2,[1,[1,2]]]])
5.8991e+008([2,[2,[2,
[1,2]]]])
-3.26378e+009([[1,2],[1,
[1,2]]]) 3.86898e+009([[1,2],
[2,[1,2]]]) }

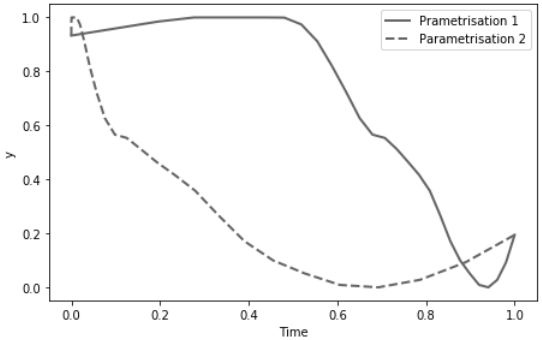

LogSignature
File Options Help
The log signature is
{ 8(1) 198(2) 28350.5([1,2])
2.07319e+006([1,[1,2]])
-120281([2,[1,2]])
1.04383e+008([1,[1,[1,2]]])
-1.79431e+007([2,[1,[1,2]]])
-841297([2,[2,[1,2]]])
4.04996e+009([1,[1,[1,
[1,2]]]]) -1.4206e+009([2,[1,
[1,[1,2]]]])
-2.11633e+008([2,[2,[1,
[1,2]]]]) 1.29087e+008([2,[2,
[2,[1,2]]]])
-1.29164e+009([[1,2],[1,
[1,2]]])
-1.83905e+008([[1,2],[2,
[1,2]]]) }

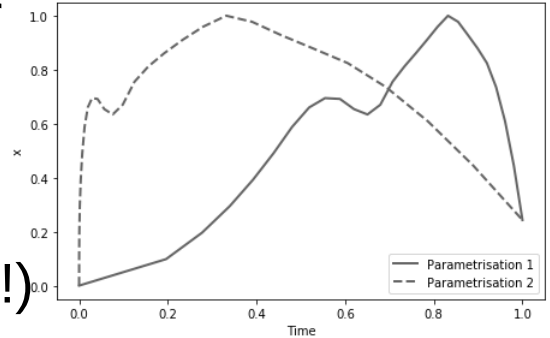
Rough paths: streamed data

- Fourier Series unstable
  - Cameron Clark
  - Dickinson
- n-dimensional linear summaries predict effects O(1/n^1/2)
- signatures O(1/(n/p)!)

Parametrisations of the handwritten 3







Log-signature:
[0.2429 -0.7386 -0.6237 -0.2292 0.0323]

Fourier series (parametrisation 1):
[32.8005 -4.3984 -4.5158 -0.7769 -0.7814]

Fourier series (parametrisation 2):
[24.0718 14.0634 0.708 -4.8552 -0.0539]

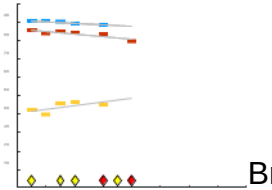
Understanding *evolving* human action samples small and noisy – dimension high

Yang, W., Lyons, T., Ni, H., Schmid, C., Jin, L. and Chang, J., 2017.
Leveraging the Path Signature for Skeleton-based Human Action Recognition. *arXiv preprint arXiv:1707.03993*.

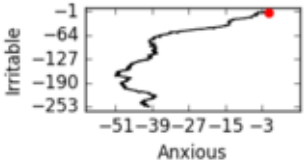
Leveraging the mathematics

Finger drawn characters

Moving matchstick man



Brain Weights



Evolving mood

Chinese handwriting

Action Recognition

Alzheimer’s Disease

Complex social data