Rough Paths, Hopf Algebras and Chinese handwriting

Greg Gyurko
Hao Ni
Terry Lyons
Andrey Kormlitzn
Harald Oberhauser
Danyu Yang

... my students ...

Mathematical Institute, University of Oxford

Kings College, London, 7th October 2016
extending a function

defining \text{Sqrt}(x)

- There are many important functions initially defined on special restricted domains
- It can be a mathematical challenge to find the correct home for these functions
- For $x = \frac{p^2}{q^2}$ one has $\sqrt{x} := \frac{p}{q}$
- For $x \in \mathbb{Q}$
extending a function

- If \( y(\gamma) := \sin \frac{1}{\gamma^2} \) then \( y(0) =? \)
  - Close the graph \( G = \{(\gamma, y(\gamma)), \gamma \in (0, 1]\} \) in \([0, 1] \times [-1, 1]\).
  - The closure of \( G \) is not the graph of a function!
**closed graphs**

**extension needs a correct topology/metric on** $\gamma$

**Definition**

A map $y : \Lambda \subset \Gamma \to Y$ is *closable* if there exist two convergent sequences $\gamma_n, \tau_n \in \Lambda$ with the same limit $\gamma$ so that $y(\gamma_n)$ and $y(\tau_n)$ both converge then their limits agree. (The closure of the graph of the function is the graph of a function).

If the graph is not closable one needs to change the topology and the completion on which the extension lives!

**a more sophisticated example**

a ball $y$ is rolled along a smooth path $\gamma$. What is the correct topology on paths $\gamma$?
**Is built on:**

- The analysis of LC Young
- The geometry of KT Chen
- The non-commutative algebra of Bourbaki
- The Lipchitz function theory of EM Stein
- The set theory of Luzitania

**It contributes to:**

- Calculus: extending the DEs of Newton and Ito to RDEs
- Stochastic analysis
- Martin Hairer’s work on SPDEs
- Data Science
Rough path theory is about the effective description of sequential data

I aim to give a few hints and present experimental validation.

Tools and goals:

- Paths, streams and controlled systems
- The Signature of a stream
- Rough path theory
  - a transform
  - an effective local description of oscillation in data
- Machine learning
  - feature sets specialized for streamed data
  - linearise polynomials
some sequential data

We have a stereotyped view of paths!

This is a piece of text where each character is coded as a byte and each byte is represented as four steps with each in one of four directions.
more sequential data

A market:

Offers to buy, offers to sell, and occasional transactions. Rich data not adequately summarized by a simple stochastic differential equation.

Source: QuantHouse, 2012 (www.quanthouse.com)
a model for a stream with effects

A controlled differential equation:

- The model

\[ dy_t = f(y_t) \, d\gamma_t, \, y_0 = a \]
An expansion in iterated integrals

Consider the case where the target space $F$ is linear, the vector fields $f(\cdot)d\gamma$ are linear, and the path $\gamma \in E$ is smooth enough. Let $A$ be the bi-linear map taking $F \otimes E$ into $F$ that $f$ induces. Picard iteration gives:

\[
\begin{align*}
\frac{dy_t}{dt} &= Ay_t d\gamma_t \\
y_T &= y_S + A \int_{S \leq u \leq T} d\gamma_u \ y_S + \ldots \\
\quad &= \left( \sum_{0}^{\infty} A^n \int \ldots \int d\gamma_{u_1} \otimes \ldots \otimes d\gamma_{u_n} \right) y_S
\end{align*}
\]
A fundamental object (Chen, Feynman)

- The signature solves a linear differential equation:
  Let $S_T := S(\gamma_{[S,T]})$ then
  \[
  dS_t = S_t \otimes d\gamma_t \\
  S_T = 1
  \]

- It is the universal non-commutative exponential.
- It is a transformation of the path to a sequence of coefficients.
the abstract framework

the basic controlling object

▶ is a path in V lifted up to the group-like elements in the tensor algebra over V
▶ integrated against a slowly varying co-cyclic one form (Yang 2015)
  ▶ classical integrands become (nearly) exact and the polynomial approximations vary slowly
  ▶ Young’s methods produce estimates don’t depend on V or on the tensor algebra
the abstract framework

estimates don’t depend on $V$ or on the tensor algebra

- a group $G$ of paths,
  - embedded in a combinatorial hopf algebra $A$ over some Banach space $V$ of labels
  - with dual of $A$ the “functions” on the group
  - and $A \ast \otimes A$ the operators on $G$

- because the linear maps from the degree one $V \subset A$ to $A$ extend to algebra maps one can define these paths on manifolds with a connection. This abstract property allows the theory without $V$

- permutation algebras (Yang 2016)
remarkable estimates

factorial decay

Let $\gamma$ be a path of finite length $L$ and parametrization $\gamma$ to have unit speed.

$$S^n_J = \int \cdots \int \dot{\gamma}_{u_1} \otimes \cdots \otimes \dot{\gamma}_{u_n} du_1 \cdots du_n$$

$$\leq \frac{|L|^n}{n!}$$

$$dy_t = Ay_t d\gamma_t$$

$$y_T = \left( \sum_{n=0}^{\infty} A^n \int \cdots \int d\gamma_{u_1} \otimes \cdots \otimes d\gamma_{u_n} \right) y_S$$
Predicting a controlled equation

Suppose $Y_t$ satisfies the following SDE:

$$dY_t = a(1 - Y_t)dX_t^{(1)} + bY_t^2dX_t^{(2)}, Y_0 = 0.$$  

where $X_t = (X_t^{(1)}, X_t^{(2)}) = (t, W_t)$, and the integral is in the Stratonovich sense, and $(a, b)$ is chosen to $(1, 2)$. (Tessy Papavasiliou and Christophe Ladroue)

Data to learn from

We generate 800 independent samples of pairs $(\{X_t\}_{0 \leq t \leq T}, Y_T)$ using Milstein’s method with discretization step 0.001. Half of the samples are used for the training set, and the rest is for the backtesting set.
 Skipping class

1. \((\text{Input, Output}) = (\{X_t\}_{0 \leq t \leq T}, Y_T)\);
2. Try to find a functional \(\hat{f}\) to fit the learning set data:

\[
\hat{Y}_T = \hat{f}(\{X_t\}_{0 \leq t \leq T})
\]

3. Measure of goodness of fitting \((R^2)\).
Our procedure

- We compute the truncated signature of each sample path of \( \{X_t\}_{t \in [0,T]} \) of order \( d \), denoted by \( S_d(X_{0,T}) \);
- In the learning set, we run a least squares based linear regression of \( Y_T \) against \( S_d(X_{0,T}) \), and the computed linear functional is denoted by \( \hat{f} \);
- Plot, for both the training set and the backtesting set, the collection of value pairs \( (Y_T, \hat{f}(S_d(X_{0,T}))) \).
Figure: $T = 0.25$, Number of simulations = 400
machine learning
learning a function $f$ from observations

$$y_n = f(\gamma_n), \ n = 1, \ldots, N$$

- Introduce features $\phi_i$ of $\gamma$; and look for $\lambda_i \in \mathbb{R}$ so that

$$y_n \approx \sum_i \lambda_i \phi_i(\gamma_n) \quad \forall n$$

$$f \approx \sum_i \lambda_i \phi_i$$

- Many methods for doing this
  - Least Squares and Singular Value Decomposition
  - Support Vector Machines and deep learning

- If the features span a separating algebra then linear combinations have good density properties and the approach is reasonable. Cf the idea of smooth functions, monomials, and Taylor’s theorem.
Coordinate iterated integrals span a graded algebra of real valued functions on paths

\[ e = e_1 \otimes \ldots \otimes e_n \in (E^*)^n \subset T(E^*) \]

\[ \phi_e(\gamma) := \langle e, S(\gamma) \rangle = \int \ldots \int \langle e_1, d\gamma_{u_1} \rangle \ldots \langle e_n, d\gamma_{u_n} \rangle \]

\[ u_1 \leq \ldots \leq u_n \in J^n \]

The shuffle product \( \shuffle \) on \( T(E^*) \) corresponds to pointwise product of coordinate integrals

\[ \phi_e(\gamma) \phi_f(\gamma) = \phi_{e \shuffle f}(\gamma) \]

The co-tensors are the pre-dual of an enveloping algebra so have a product structure that linearises polynomials on the group!!
coordinate iterated integrals separate streams

A locally finite basis for polynomials on stream space

- Signatures parametrize and separate streams (Hambly Lyons Annals of Math 2010, Boedihardjo Geng Lyons Yang ArXiv 2014)
- Co-ordinate iterated integrals separate signatures
- Linearisation of smooth functions
  - Log signature parametrizes Signatures
  - Polynomials in the log signature of a stream are linear functionals of the signature
- See signatures in practise
Experiments

- Simple classification problem
  - 30 minutes of normalised financial market data
  - learning and backTesting sets
- Objective: learn a simple classification problem

\[ f \left( \text{time series} \right) = 1 \quad \text{time slot}=10.30-11.00 \]
\[ f \left( \text{time series} \right) = 0 \quad \text{time slot}=14.00-14.30 \]

- Use the co-ordinates of the signature of the normalised financial market data \( \gamma \) as features \( \phi_i \left( \gamma \right) \)

\[ f \left( \gamma \right) \approx \sum_i \lambda_i \phi_i \left( \gamma \right) \]
classification of time-buckets from standardised data

Methodology

- linear regression based pair-wise separation
- LASSO (least absolute shrinkage and selection operator) shrinkage with cross-validation
- apply statistical indicators (ROC etc)
classification of time-buckets from standardised data

commodity future, front montY: 10:30-11:00 vs 14:00-14:30

Learning set: K-S distance: 0.8, correct classification ration 90%
Out-of-sample set: K-S distance: 0.84, classification ratio 89%

Rough Paths, Hopf Algebras and the classification of streams - Kings College, London, 7th October 2016 - p25/29
classification of time-buckets from standardised data

commodity future, front montY: 10:30-11:00 vs 14:00-14:30

ROC curve
area under curve: learning set: 0.954, out of sample: 0.958
classification of time-buckets from standardised data

commodity future, front montY: 10:30-11:00 vs 14:00-14:30

Rough Paths, Hopf Algebras and the classification of streams - Kings College, London, 7th October 2016 - p27/29
classification of time-buckets from standardised data

commodity future, front montY: 12:00-12:30 vs 12:30-13:00
Thanks

- Funding
  - Oxford Man Institute
  - ERC grant ESig (agreement no. 291244)
- People: So many...

References

- Citations etc. arXiv:1405.4537
- Books: Lyons and Qian; Friz and Victoir; Lyons, Caruana and Levy; Friz and Hairer.