Rough Paths and its Applications in Machine Learning

Lim Nengli

Singapore University of Technology and Design

July 20, 2017
History and motivation

- Originally formulated to study stochastic differential equations in a path-wise manner.
- Rough paths theory $\rightarrow$ theory of regularity structures
  - Solution to KPZ equation (Martin Hairer, Fields medal 2014).
  - Now begin applied to other hard problems in statistical physics.
- Recently, rough paths theory is finding applications in machine learning.
Controlled differential equations

Consider the following differential equation:

$$y'_t = V(y_t) x'_t$$

First-order approximation to the solution: For $s, t \in [0, T]$,

$$y_t - y_s = \int_s^t V(y_r) x'_r \, dr$$

$$\approx V(y_s) \int_s^t x'_r \, dr$$

$$= V(y_s) (x_t - x_s)$$
Second-order approximation

\[ y_t - y_s = \int_s^t V(y_{r1}) x'_{r1} \, dr_1 \]

\[ = \int_s^t \left( V(y_s) + \int_s^{r_1} V'(y_{r2}) y'_{r2} \, dr_2 \right) x'_{r1} \, dr_1 \]

\[ = V(y_s) \left[ \int_s^t x'_{r1} \, dr_1 \right] \]

\[ + \int_s^t \left( \int_s^{r_1} V'(y_{r2}) V(y_{r2}) x'_{r2} \, dr_2 \right) x'_{r1} \, dr_1 \]

\[ \simeq V(y_s) \left[ \int_s^t d x_{r1} \right] \]

\[ + V'(y_s) V(y_s) \left[ \int_s^t \left( \int_s^{r_1} d x_{r2} \right) d x_{r1} \right] \]
Separating signal from vector field

In one dimension:

\[ y_t - y_s \simeq V(y_s) [x_t - x_s] + V'(y_s) V(y_s) \left[ \frac{1}{2} (x_t - x_s)^2 \right] \]

In \( d \) dimensions:

\[
\begin{align*}
y_t - y_s & \simeq V(y_s) [x_t - x_s] \\
& + \nabla V(y_s) V(y_s) \left[ \sum_{i,j=1}^{d} \left( \int_{s}^{t} \int_{r_1}^{r_2} dx_{r_1}^{(i)} dx_{r_2}^{(j)} \right) e_i \otimes e_j \right]
\end{align*}
\]
2-dimensional case

Given

\[
y_t' = \begin{bmatrix} (y_t^{(1)})' \\ (y_t^{(2)})' \end{bmatrix} = \begin{bmatrix} V_1^{(1)}(y_t) & V_2^{(1)}(y_t) \\ V_1^{(2)}(y_t) & V_2^{(2)}(y_t) \end{bmatrix} \begin{bmatrix} (x_t^{(1)})' \\ (x_t^{(2)})' \end{bmatrix},
\]

we have

\[
y_t - y_s \simeq \begin{bmatrix} V_1^{(1)}(y_s) & V_2^{(1)}(y_s) \\ V_1^{(2)}(y_s) & V_2^{(2)}(y_s) \end{bmatrix} \begin{bmatrix} x_t^{(1)} - x_s^{(1)} \\ x_t^{(2)} - x_s^{(2)} \end{bmatrix}
+ \nabla V(y_s) V(y_s) \left( \begin{bmatrix} \int_s^t \int_s^{r_1} dx_{r_2}^{(1)} dx_{r_1}^{(1)} \\ \int_s^t \int_s^{r_1} dx_{r_2}^{(2)} dx_{r_1}^{(1)} \end{bmatrix} \begin{bmatrix} \int_s^t \int_s^{r_1} dx_{r_2}^{(1)} dx_{r_1}^{(2)} \\ \int_s^t \int_s^{r_1} dx_{r_2}^{(2)} dx_{r_1}^{(2)} \end{bmatrix} \right).
\]
2-dimensional case cont.

\[
\begin{bmatrix}
\int_s^t \int_{r_2}^{r_1} dx_{r_2}^{(1)} dx_{r_1}^{(1)} & \int_s^t \int_{r_2}^{r_1} dx_{r_2}^{(1)} dx_{r_1}^{(2)} \\
\int_s^t \int_{r_2}^{r_1} dx_{r_2}^{(2)} dx_{r_1}^{(1)} & \int_s^t \int_{r_2}^{r_1} dx_{r_2}^{(2)} dx_{r_1}^{(2)}
\end{bmatrix}
\]

\[= \frac{1}{2} \begin{bmatrix}
\left(x_{s,t}^{(1)}\right)^2 & x_{s,t}^{(1)}x_{s,t}^{(2)} \\
x_{s,t}^{(2)}x_{s,t}^{(1)} & \left(x_{s,t}^{(2)}\right)^2
\end{bmatrix} + \frac{1}{2} \begin{bmatrix}
0 & A_s^t \\
-A_s^t & 0
\end{bmatrix}
\]

\text{symmetric part}

\text{anti-symmetric part}

where

\[x_{s,t}^{(k)} := x_t^{(k)} - x_s^{(k)}, \quad k = 1, 2,\]

and the Lévy-area is given by

\[A_s^t = \int_s^t \int_{r_2}^{r_1} dx_{r_2}^{(1)} dx_{r_1}^{(2)} - \int_s^t \int_{r_2}^{r_1} dx_{r_2}^{(2)} dx_{r_1}^{(1)}.\]
Green’s theorem

\[ A = \frac{1}{2} \oint_C -x^{(2)} \, dx^{(1)} + x^{(1)} \, dx^{(2)} \]
Signature of a path

For all $0 \leq s \leq t \leq T$,

$$S_n(x)_{s,t} := (1, x_{s,t}, x_{s,t}^2, \ldots, x_{s,t}^n)$$

$$\in \mathbb{R} \oplus \mathbb{R}^d \oplus (\mathbb{R}^d \otimes \mathbb{R}^d) \oplus \cdots \oplus (\mathbb{R}^d)^{\otimes n},$$

where $x_{s,t}^k$ is the conventional $k$-th iterated integral of the path $x$ over the interval $[s, t]$:

$$x_{s,t}^k = \sum_{j_1, \ldots, j_k=1}^{d} \left( \int_{s<r_1<\cdots<r_k<t} \text{d}x_{r_1}^{(j_1)} \otimes \cdots \otimes \text{d}x_{r_k}^{(j_k)} \right) e_{j_1} \otimes \cdots \otimes e_{j_k}.$$
Chen’s Identity

- The signature is an element of a Lie group called the step-$n$ nilpotent group with $d$ generators.

- It satisfies

$$S_n(x)_{s,t} = S_n(x)_{s,u} \tilde{\otimes} S_n(x)_{u,t} \quad \forall \ s, u, t \in [0, T], \ s \leq u \leq t,$$

where given $a = (1, a^1, \ldots, a^n), b = (1, b^1, \ldots, b^n)$, group multiplication is performed by

$$a \tilde{\otimes} b := (1, c^1, \ldots, c^n), \quad c^k = \sum_{i=0}^{k} a^i \otimes b^{k-i}, \quad \forall \ 1 \leq k \leq n.$$ 

- E.g. $(1, a^1, a^2) \tilde{\otimes} (1, b^1, b^2) = (1, a^1 + b^1, a^2 + b^2 + a^1 \otimes b^1)$. 
Fractional Brownian motion: $W_t^H$

(i) Hurst parameter: $H \in (0, 1)$

(ii) Continuous paths, not differentiable a.e.

(iii) $W_t - W_s \sim \mathcal{N}(0, |t - s|^{2H})$

(iv) Covariance function: $R(s, t) = \frac{1}{2} (s^{2H} + t^{2H} - |t - s|^{2H})$

(v) $H = \frac{1}{2}$: Standard Brownian motion, $R(s, t) = s \land t$.

(vi) $H > \frac{1}{2}$: Increments along disjoint intervals are positively correlated.

(vii) $H < \frac{1}{2}$: Increments along disjoint intervals are negatively correlated.

(viii) Neither a Markov process nor a martingale (unless $H = \frac{1}{2}$)
Sample paths

\begin{figure}
\centering
\includegraphics[width=\textwidth]{sample_paths.png}
\caption{Sample paths for different Hurst exponents: $H=0.9$, $H=0.5$, and $H=0.1$.}
\end{figure}
Hölder continuity and rough paths

Definition
A function \( f \) is said to be \( \alpha \)-Hölder continuous on an interval \([0, T]\) if

\[
|f(t) - f(s)| \leq C |t - s|^\alpha, \quad \forall s, t \in [0, T].
\]

- Hölder continuity measures how "rough" a function is.
- Fact: Fractional Brownian motion with Hurst parameter \( H \) is almost surely \((H - \varepsilon)\)-Hölder continuous for any \( \varepsilon > 0 \).

Definition
Given \( \frac{1}{3} < \alpha \leq \frac{1}{2} \), \( X = (1, X_{s,t}, X_{s,t}^2) \) is an \( \alpha \)-Hölder rough path if it satisfies Chen’s identity, \( X \) is \( \alpha \)-Hölder continuous and \( X^2 \) is \( 2\alpha \)-Hölder continuous.
Itó integration as rough paths integration

- \[ \int_0^T Y_t \, dW_t = \lim_{\|\pi\| \to 0} \sum_i Y_{t_i} W_{t_i, t_{i+1}}, \]
  where the limit is taken in \( L^2(\Omega) \) and not almost surely path-wise because the paths are not regular enough.

- Even so, convergence in \( L^2(\Omega) \) relies on the fact \( W_t \) is a martingale, and that \( Y_t \) is adapted to the filtration of \( W_t \).

- Define

  \[
  W_{s,t}^{\text{Itô}} := (1, A_{s,t}^1, A_{s,t}^2) = \left(1, W_{s,t}, \int_s^t W_{s,r_1} \otimes dW_{r_1}\right)
  \]

  Then given a "Gubinelli derivative" \( Y' \),

  \[
  \int_0^T Y_t \, dW_t = \int_0^T Y_t \, dW_t^{\text{Itô}} = \lim_{\|\pi\| \to 0} Y_{t_i} A_{t_i, t_{i+1}}^1 + Y' A_{t_i, t_{i+1}}^2
  \]
  almost surely.
Properties of the signature

- The signature has more or less a one-to-one relation with its path (see T. Lyons and B. Hambly, *Uniqueness for the signature of a path of bounded variation and the reduced path group*, 2010).
- It is a graded summary of the data stream encoded in the path.
  - Iterated integrals capture non-linear aspects of the path.
  - Forms a natural “basis” for functionals on data streams.
- It provides a rich set of features that can be used in a machine learning pipeline.
Applications

- Finance:

- Sound compression:

- Identifying patterns in MEG scans etc.

Chinese handwriting recognition

- SCUT gPen:
  - Online Chinese handwriting recognition software
  - Began as a collaboration between Terry Lyons and Ben Graham (University of Warwick)
  - App developed by HCI-Lab in South China University of Technology

- State of the art: Won first place in ICDAR2013 competition with an error rate of 2.61% (Second place: 3.13%, Human error: 4.81%).

- Combines rough path theory and a deep convolutional neural network.

- Uses first 3 levels of the signature of the path.

Conclusion and ramblings

- Textbooks:
  - Peter Friz and Martin Hairer (2014), A Course on Rough Paths, Springer.

- Future direction:
  - Application to stochastic control and reinforcement learning:
    (i) Extend control theory to dynamical systems perturbed by coloured noise.
    (ii) Find efficient Monte-Carlo schemes to compute optimal path and control trajectories.