EAI Math Reading Group

Signature Transform

Outline

- Basics of Signatures

  [A Primer on the Signature Method in ML – Chevyrev, Kormilitzin (2016)]

- Application : Expressive power of SSMs

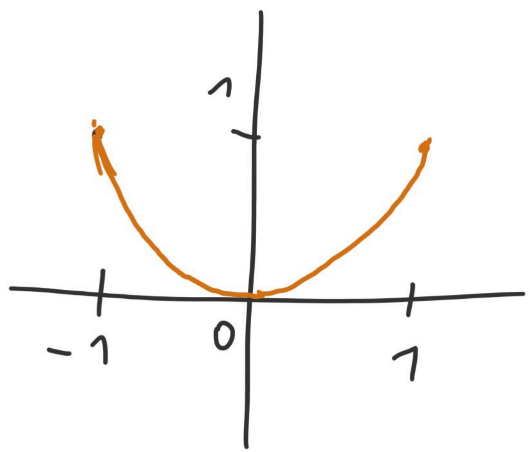
  [Theoretical Foundations of Deep Selective SSMs – Cirone et al. (2024)]

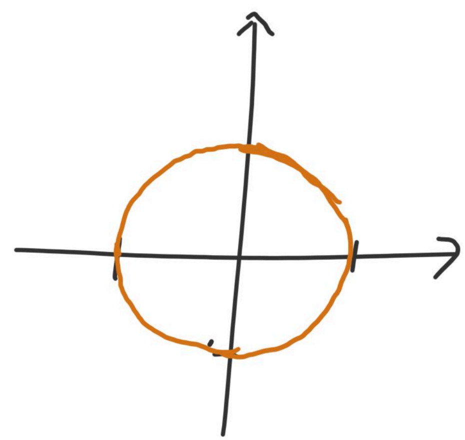
Paths

$$X : [a, b] \longrightarrow \mathbb{R}^d$$

$$t \longmapsto X_t = \{X_t^1, \dots, X_t^d\}$$

e.g. $X_t = \{t, t^2\}$ $t \in [-1, 1]$ $\qquad$ $X_t = \{\cos t, \sin t\}$ $t \in [0, 2\pi]$





Path integrals

integral of $f: \mathbb{R} \to \mathbb{R}$ against $X: [a,b] \to \mathbb{R}$

$$\int_a^b f(X_t)\, dX_t = \int_a^b f(X_t) \underbrace{\dot{X}_t}_{\hookrightarrow \frac{dX_t}{dt}} \, dt$$

integral of $Y: [a,b] \to \mathbb{R}$ against $X: [a,b] \to \mathbb{R}$

$$\int_a^b Y_t \, dX_t = \int_a^b Y_t \dot{X}_t \, dt$$

Examples

$$Y_t = \{t^2\} \qquad X_t = t \quad \Rightarrow \quad dX_t = 1 . dt$$

$$\int_a^b Y_t \, dX_t = \int_a^b t^2 \, dt = \left[\frac{t^3}{3}\right]_a^b = \frac{1}{3}(b^3 - a^3)$$

$$Y_t = \{t^2\} \qquad X_t = \{t^3\} \quad \Rightarrow \quad dX_t = 3t^2 dt$$

$$\int_a^b Y_t \, dX_t = \int_a^b t^2 \cdot 3t^2 \, dt = \left[\frac{3t^5}{5}\right]_a^b = \frac{3}{5}(b^5 - a^5)$$

Signature of a path

Given $X : [a,b] \to \mathbb{R}^d \sim (X^1, \dots, X^d)$

$$S(X)^i_{a,t} = \int_a^t dX^i_s \triangleq \int_{a<s<t} dX^i_s = X^i_t - X^i_a$$

NB. $S(X)^i_a : [a,b] \to \mathbb{R}$ is a path $\to$ iterate!

$$S(X)^{i,j}_{a,t} = \int_{a<s<t} S(X)^i_{a,s}\, dX^j_s = \int_{a<r<s<t} dX^i_r\, dX^j_s$$

$$= \int_a^t \int_a^s dX^i_r\, dX^j_s$$

Iterated integrals

For k-tuple of indices $(i_1, \dots, i_k) \in [d]^k$

$$S(X)^{i_1,\dots,i_k}_{a,t} = \int_{a<s<t} S(X)^{i_1,\dots,i_{k-1}}_{a,s} \, dX^{i_k}_s$$

The signature $S(X)_{a,b}$ of $X : [a,b] \to \mathbb{R}^d$

is the collection of all iterated integrals of X

$$S(X)_{a,b} = \left(1, S(X)^1_{a,b}, \dots, S(X)^d_{a,b}, S(X)^{1,1}_{a,b}, S(X)^{1,2}_{a,b}, \dots\right)$$

NB. multi-indices live in $W_d = \{(i_1, \dots, i_k) \mid k \geq 1, i_1, \dots, i_k \in \{1, \dots, d\}\}$

Small Recap

The signature of a path $X : [a,b] \to \mathbb{R}^d$ is the collection of iterated integrals over all multi-indices

$$S(X) = \Big(\underbrace{1}_{0 \text{ order}} \, , \underbrace{S(X)^1_{a,b}, \dots, S(X)^d_{a,b}}_{\substack{1^{st} \text{ order} \\ (d \text{ terms})}} \, , \underbrace{S(X)^{1,1}_{a,b}, \dots, S(X)^{d,d}_{a,b}}_{\substack{2^{d} \text{ order} \\ (d^2 \text{ terms})}} \, , \dots$$

→ Can be thought as an element of $T(\mathbb{R}^d)$ ↪ tensor algebra on $\mathbb{R}^d$

Application: Picard iteration

Given a 1^{st} order ODE

$$\frac{dy}{dt} = f(y,t) \qquad y(t_0) = y_0$$

Integrating w/ respect to dt yields a fixed point equation

$$y(t) = \underbrace{y(t_0) + \int_{t_0}^{t} f(y(s),s)\,ds}_{}$$

$$y = F(y)$$

① Define

$$y_k(t) = y(t_0) + \int_{t_0}^{t} f(y_{k-1}(s), s)\,ds$$

for $k \in \mathbb{N}$, with $y_0(t) = y_0(t_0) = y_0$ $\quad \forall\, t \in [t_0, t_{max}]$

<u>Picard-Lindelöf theorem</u>:

f continuous in t and Lipschitz-continuous in y: $|f(y_1) - f(y_2)| < K|y_1 - y_2|$

$$\Rightarrow \quad y_k \xrightarrow[k \to \infty]{} y$$

<u>Example</u>: $\frac{dy}{dt} = y(t)$; $y(0) = 1$

$$y_k(t) = \sum_{n=0}^{k} \frac{1}{n!} t^n \longrightarrow e^t$$

Controlled Differential Equations

For $V : \mathbb{R}^e \longrightarrow \underbrace{L(\mathbb{R}^d, \mathbb{R}^e)}_{\text{linear maps } \mathbb{R}^{e \times d}}$, $Y : [a,b] \to \mathbb{R}^e$, $X : [a,b] \to \mathbb{R}^d$

$$dY_t = V(Y_t)\, dX_t \quad , \quad Y_a = y$$

$$\Leftrightarrow \qquad Y_t = y + \int_a^t V(Y_s)\, dX_s \qquad \rightsquigarrow \text{ Use Picard iteration}$$

$$Y_t^0 = y$$

$$Y_t^1 = y + \int_a^t V(Y_s^0)\, dX_s = \left(\int_a^t dV(X_s) + I_e \right)(y)$$

$$Y_t^2 = y + \int_a^t V(Y_s^1)\, dX_s = \left(\int_a^t \int_a^s dV(X_u)\, dV(X_s) + \int_a^t dV(X_s) + I_e \right)(y)$$

The n-th iterate is

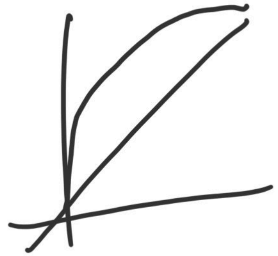
$$y_t^n = y + \int_a^t V(y_s^{n-1})\, dX_s = \left(\sum_{k=1}^{n} \int_{a<t_1<\dots<t_k<t} dV(X_{t_1}) \dots dV(X_{t_k}) \right)(y)$$

$\rightarrow$ terms $\displaystyle\int_{a<t_1<\dots<t_k<t} dV(X_{t_1}) \dots dV(X_{t_k}) \in L(\mathbb{R}^d, \mathbb{R}^e)$

$\rightsquigarrow$ depend <u>linearly</u> on terms up to level k of $S(X)_{a,t}$

- $y_t^k \rightarrow y_t \quad \rightsquigarrow \quad y_t$ depends linearly on $S(x)_{a,t}$

Properties of the Signature



- Invariance under time reparameterization

For $X, Y : [a,b] \to \mathbb{R}$, $\psi : [a,b] \to [a,b]$ monotonous, continuous, nondecreasing

$$\tilde{X}_t = X_{\psi(t)}, \quad \tilde{Y}_t = Y_{\psi(t)} :$$

$$d\tilde{X}_t = dX_{\psi(t)} \dot{\psi}(t)$$

and

$$\int_a^b \tilde{Y}_t \, d\tilde{X}_t = \int_a^b Y_{\psi(t)} \, dX_{\psi(t)} \dot{\psi}(t) \, dt = \int_a^b Y_s \, dX_s$$

$\longrightarrow$ Path integrals (hence signature) <u>don't</u> depend on time parametrization

Practical use

Given discrete time series data $\{X_1, X_2, \ldots, X_T\}$

1) Construct continuous path (e.g. linear interpolation, rectilinear, splines, lead-lag, ...)

   $\rightarrow X_t$

2) Compute (truncated) signature $S(X)$

3) Use signature terms as features

4) Profit?

Linear CDEs and SSMs

let $X : [0,1] \to \mathbb{R}^{d_X}$, $U : [0,1] \to \mathbb{R}^{d_U}$
(input)

be continous paths

$$dZ_t = \sum_{i=1}^{d_X} A_i Z_t \, dX_t^i + B \, dU_t \qquad Z_0 \in \mathbb{R}^N$$

is a linear CDE with solution

$$Z_t = \sum_{I \in W_{d_X}} A_I Z_0 \, S(X)_{0,t}^I + \sum_{i=1}^{d_U} \sum_{I \in W_{d_X}} A_I B_i \int_0^t S(x)_{s,t}^I \, dU_s^i$$

$I = (i_1, \dots, i_k) \in W_{d_X}$

SSMs are linear CDEs

- S4: $z^i_l = \bar{A}_i z^i_{l-1} + \bar{B}_i x^i_l$ $\rightsquigarrow$ $X_t = t$; $U_t = \int_0^t y_s \, ds$

input embedding

- Mamba: $z^i_l = \bar{A}(x^i_l) z^i_{l-1} + \bar{B}(x^i_l) x^i_l$ $\rightsquigarrow$ $dZ_t = A Z_t \, dX_t + B \, dU_t$

$\sigma(\tilde{\alpha} y_t + \tilde{\beta}) dt$ $\quad$ $\sigma(\tilde{\alpha} y_t + \tilde{\beta}) y_t dt$

Expressivity of linear CDEs

Given paths X_t, U_t s.t. $X_t^1 = t$, $X_t^2 = t^2$, then linear CDEs can approximate over $\mathbb{K} \times [0,1]$ any functional of the form

$$\Psi(X_{[0,t]}) + \int_0^t \phi(X_{[s,t]}) \cdot dU_s$$

with

$$\left.\begin{aligned} &\Psi : C_{1,0}([0,1], \mathbb{R}^{d_x}) \to \mathbb{R} \\ &\phi : \mathbb{R}^{d_x} \to \mathbb{R}^{d_u} \end{aligned}\right\} \text{continuous}$$

If the matrices $A_1, \dots, A_{d_x}$ are diagonal, we get functionals of the form

$$\psi(X_t) + \int_0^t \varphi(X_t - X_s) \cdot dU_s$$

with $\psi : \mathbb{R}^{d_x} \to \mathbb{R}$, $\varphi : \mathbb{R}^{d_x} \to \mathbb{R}^{d_u}$ continuous

Randomized Linear CDEs

For any compact $\mathbb{K} \subset \mathbb{X}$ and any functional $F : \mathbb{K} \times [0,1] \to \mathbb{R}$ in the CDE closure,

$$\lim_{N \to \infty} \mathbb{P}\left[\left\{ \exists\, v \in \mathbb{R}^N \text{ s.t. } \sup_{(x,t)} \left| F(x,t) - \langle v, Z_t \rangle \right| \leq \varepsilon \right\}\right] = 1$$

where Z_t is generated by a linear CDE with

$$[A_i]_{n,n'} \overset{iid}{\sim} N\left(0, \tfrac{1}{N}\right) \qquad \underbrace{[Z_0]_n}_{\text{initial state}}, [B]_{n,j} \overset{iid}{\sim} N(0,1)$$

TLDR: with high enough state dim N, only need to generate Z_t randomly and train v

Random Linear CDEs

- random LCDEs Z_t can approximate any F with high probability as $\langle Z_t, v \rangle$ $\rightsquigarrow$ only train v

  $\hookrightarrow$ [Reservoir Computing]

- Z_t depends linearly on $S(X)$

  $\hookrightarrow$ $S(X)$ are the "right features" for path to path transforms

Path to Path learning

For a map $G : \mathcal{C}_{1,0}([0,1], \mathbb{R}^{d_x}) \times [0,1] \to \mathbb{R}$

$\sim \mathcal{C}_{1,0}([0,1], \mathbb{R}^{d_x}) \longrightarrow \mathcal{C}_{1,0}([0,1], \mathbb{R})$

G can be approximated arbitrarily well by some neural network $F : \mathbb{R}^N \to \mathbb{R}$ applied to a LCDE Z_t

$$\sup_{(X,t)} \left| F(Z_t) - G(X_t, t) \right| < \varepsilon \qquad (\varepsilon > 0 \text{ arbitrary})$$

Wrapping up

- Rigorous theory of SSM expressivity

- Diagonal SSMs : weaker, but can be chained to increase expressive power

  $\rightsquigarrow$ K chained diagonal SSMs $\approx$ K levels of Signature

- (Randomized) LCDEs $\rightsquigarrow$ Kernel version of SSMs $\rightarrow$ possible NTK style results for SSMs ?

- Feature / Layer in deep models $\rightarrow$ Deep Signature Transforms
  Kidger