An Elementary Introduction to Signature Methods and Rough Path Theory for ML

Thomas Li and Damian B
Santa Rosa Junior College

Motivation of the Signature

- **Picard for ODEs.** Given $\frac{dy}{dx} = f(x,y)$, $y(x_0) = y_0$, Picard iterations

$$y_0(x) = y_0,$$
$$y_{n+1}(x) = y_0 + \int_{x_0}^{x} f\big(t, y_n(t)\big)\, dt$$

converge to the power-series solution. E.g. $\frac{dy}{dx} = y,\ y_0 = 1$: $y_1 = 1 + x,\ y_2 = 1 + x + \frac{x^2}{2}, \cdots \to e^x$.

- **From ODE to CDE.** Let

$$X : [a,b] \to \mathbb{R}^d, \quad Y : [a,b] \to \mathbb{R}^e, \quad F : \mathbb{R}^e \to V(\mathbb{R}^d, \mathbb{R}^e)$$ (e by d matrices)

with $F = [F_1 \mid \cdots \mid F_d]$ and $dX_s = (dX_s^1, \ldots, dX_s^d)^T$. Then

$$dY_t = F(Y_t)\, dX_t, \quad Y_t = Y_a + \sum_{i=1}^{d} \int_a^t F_i(Y_s)\, dX_s^i.$$

$$Y_t^0 = Y_a,$$
$$Y_t^1 = Y_a + \sum_{i=1}^{d} F_i(Y_a) \int_a^t dX_s^i,$$
$$Y_t^2 = Y_t^1 + \sum_{i,j=1}^{d} F_i(F_j(Y_a)) \int_a^t \int_a^s dX_u^j\, dX_s^i$$

...

In the limit,

$$Y_t = Y_a + \sum_{k=1}^{\infty} \sum_{i_1,\ldots,i_k} F_{i_1\cdots i_k}(Y_a) \int_{a<t_1<\cdots<t_k<t} dX_{t_1}^{i_1} \cdots dX_{t_k}^{i_k}.$$

- **Path Signature.** For a continuous path $X = (X^1, \ldots, X^d)$, define

$$S(X)_{a,t} = \left(1,\ \int_a^t dX,\ \int_{a<s_1<s_2<t} dX_{s_1} \otimes dX_{s_2}, \ldots\right) \in T(\mathbb{R}^d).$$

Its level-k entries are $\boxed{S(X)_{a,t}^{i_1\ldots i_k} = \int_{a<t_1<\cdots<t_k<t} dX_{t_1}^{i_1} \cdots dX_{t_k}^{i_k}}$. For level 2: $dX_{s_1} \otimes dX_{s_2}$, we have a matrix whose (i,j) entries are: $(\int_{a<s_1<s_2<t} dX_{s_1}^i dX_{s_2}^j)_{i,j}$

- **Uniqueness (Hambly–Lyons, 2010).** For continuous bounded-variation (or rough) paths $X, Y : [0,T] \to^d$,

$$S(X) = S(Y) \quad \Longleftrightarrow \quad X, Y \text{ are tree-like equivalent.}$$

Thus Signatures form a universal feature map: any reasonable path-functional can be approximated by a linear functional on truncated signatures. Furthermore, Signature terms decay factorially, meaning that truncating at low orders will capture most of the path's information.

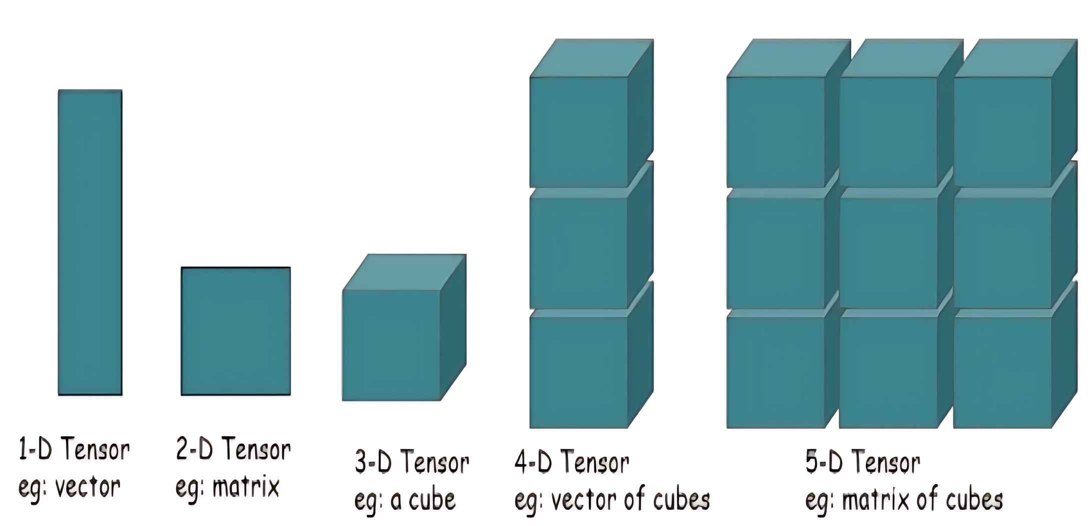


Figure: Visualization of the Tensor Algebra where Signature terms are located in.

Signature-based Methods in ML

Data often arrives as *streamed* observations, often indexed by time. Machine Learning seeks to systematically understand the behavior of this streamed data, despite the fact that it is often extremely irregular, nonstationary, and high-dimensional. **Signature Methods** and the broader framework of **Rough Path Theory** serve as a unique way to represent high dimensional data, yet preserve information about the data completely and uniquely (analogous to that of power series). It extracts a hierarchy of coordinate-invariant, mathematically complete features via *iterated integrals*. We present an *overview* of the theory of signatures and applications to ML: including the **log-signature** and how it may be applied to tasks like *time series classification*. We additionally hope to present this mathematical framework in an intuitive, elementary fashion to improve comprehension.

Log-Signature

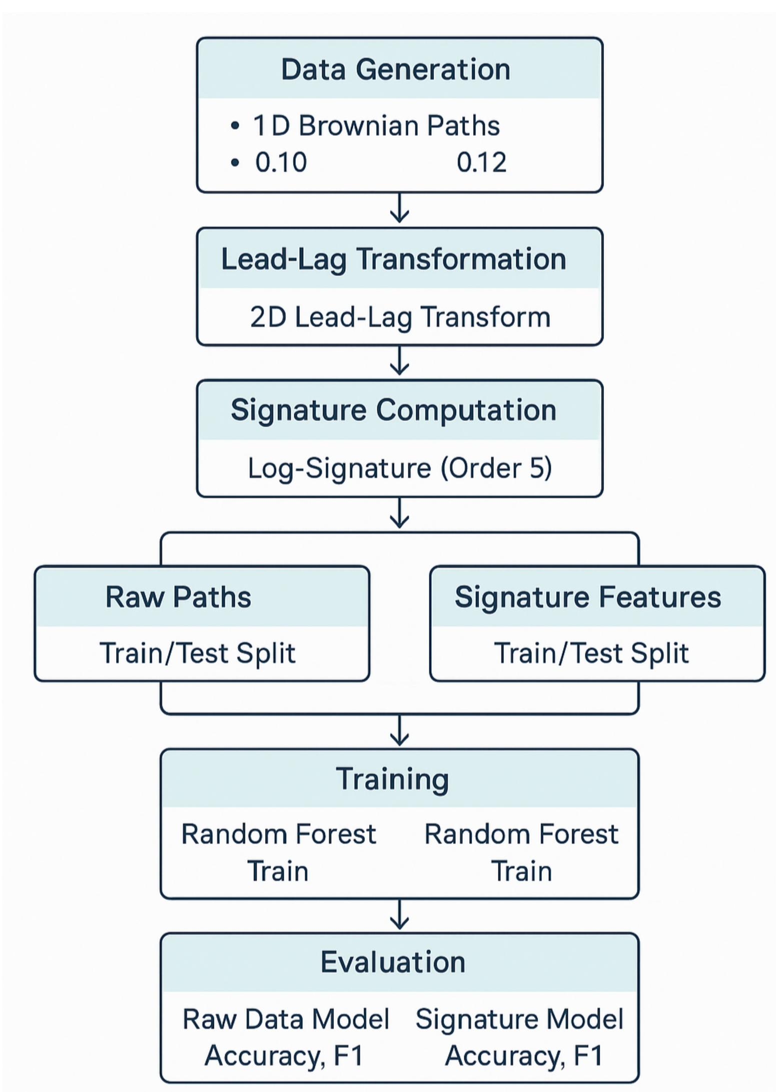
$$\ln \mathbf{S}(X)_{a,b} = \sum_{n=1}^{\infty} (-1)^{n-1} \frac{(\mathbf{S}(X)_{a,b} - 1)^{\otimes n}}{n} = (\mathbf{S} - 1) - \tfrac{1}{2}(\mathbf{S} - 1)^{\otimes 2} + \cdots$$

$$= \sum_{i=1}^{d} S(X)_{a,b}^{i} e_i + \sum_{i,k=1}^{d} \left(S(X)_{a,b}^{i,k} - \tfrac{1}{2} S(X)_{a,b}^{i} S(X)_{a,b}^{k}\right)(e_i \otimes e_k) + \cdots$$

$$= \sum_{i=1}^{d} S^i e_i + \tfrac{1}{2} \sum_{i<k} (S^{i,k} - S^{k,i})\, [e_i, e_k] + \ \ldots \text{ where } [e_i, e_k] = e_i \otimes e_k - e_k \otimes e_i.$$

- Ensures linear independence of signature features.
- In practice, we save the coefficients of each lie algebra term (the bracket terms) in a vector. This feature vector mathematically lies in the lie algebra space.

Applications & Results



Case Study: Brownian Motion Classification.

- Simulated $N = 1000$ paths, $T = 70$ time steps: $X_t = \sum_{i=1}^{t} \epsilon_i,\ \epsilon_i \sim \mathcal{N}(0, \Delta^2),\ \Delta \in \{0.10, 0.12\}$
- Goal: distinguish between low and high volatility paths based on discrete increments.
- *Raw features (save path data):* $\mathbf{X}_{\text{raw}} \in \mathbb{R}^{N \times T}$
- **Lead–Lag Transformation**: Each 1D path $X_i = \{X_i(t_j)\}_{j=1}^{T}$ is transformed into a 2D path by pairing each point with its previous value:

$$\widetilde{X}_i = \left\{\big(X_i(t_j),\ X_i(t_{j-1})\big)\right\}_{j=1}^{T} \in \mathbb{R}^{T \times 2}.$$

- **Log-Signature Computation**: We compute truncated log-sigs up to order 5 and save to a feature matrix.

$$\mathbf{X}_{\text{sig}} = \begin{bmatrix} \ln S(\widetilde{X}_1)_{0,T} \\ \ln S(\widetilde{X}_2)_{0,T} \\ \vdots \\ \ln S(\widetilde{X}_N)_{0,T} \end{bmatrix} \in \mathbb{R}^{N \times D}.$$

Applications Results

Classification Method: Random Forest

- **200 Trees:** Each model aggregates predictions from 200 decision trees ("mini-experts"), with the majority vote determining the output. This stabilizes performance and reduces variance.
- **Max Depth = 15:** Trees can make up to 15 binary splits (e.g., "Is feature 37 $>$ 0.42?"). This limits overfitting by preventing any tree from memorizing noise.
- **Feature Sampling = $\sqrt{p}$:** At each split, each tree considers a random subset of $\sqrt{p}$ features (where p is the total number of features), injecting randomness and reducing correlation between trees.

Evaluation Metrics.

- **Accuracy** $= \frac{\text{correct predictions}}{\text{total predictions}}$
- **Precision** $= \frac{\text{TP}}{\text{TP+FP}}$: correct "positive" guesses
- **Recall** $= \frac{\text{TP}}{\text{TP+FN}}$: How many actual positives were correctly predicted.

Single-Run Results:

Feature	Accuracy	Precision	Recall
Raw	57.3%	0.56	0.68
Signature	85.0%	0.89	0.80

Repeated Trials (5 runs). $\overline{\text{Accuracy}}_{\text{raw}} = 58.1\% \pm 2.8\%$, $\overline{\text{Accuracy}}_{\text{sig}} = 84.9\% \pm 1.2\%$

Takeaway: Signature features yield a $\sim 27\%$ absolute and $\sim 46\%$ relative accuracy improvement over classifying and recognizing Brownian Motion!

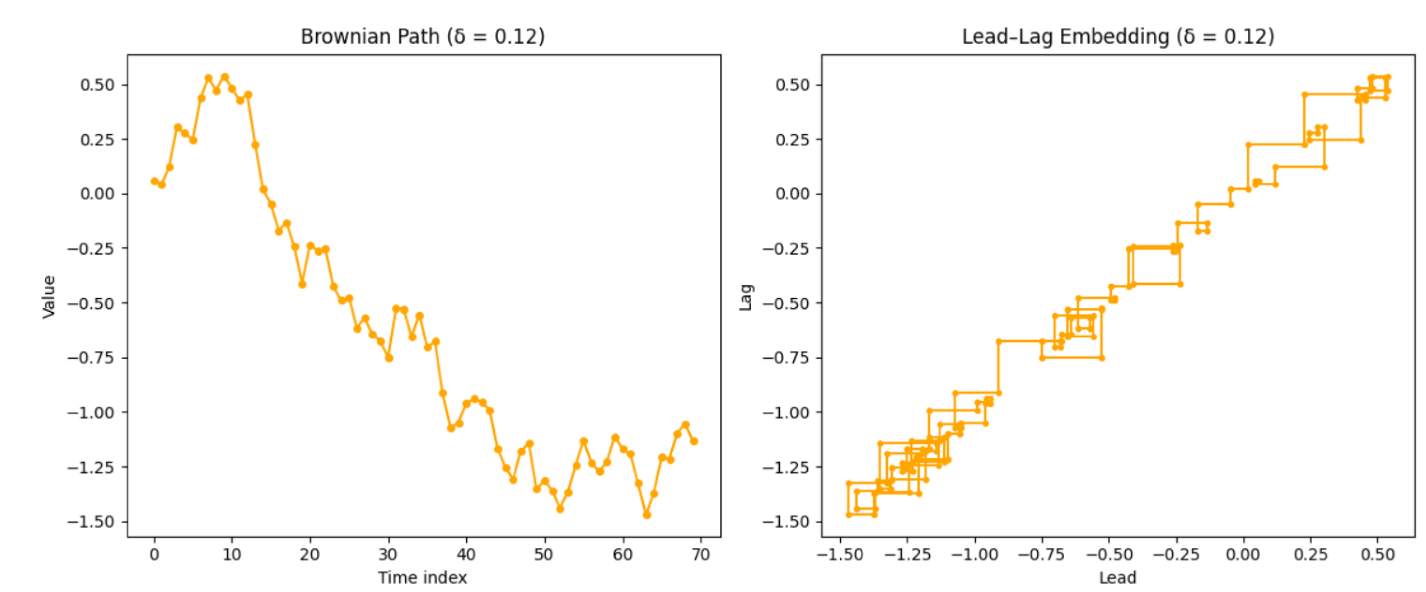


Figure: Sample Brownian motion path and its lead-lag transformation with $\Delta = 0.12$

Conclusion

Path signatures offer superior classification performance for distinguishing stochastic processes with subtle parameter differences. Brownian Motion is a classic example of highly chaotic data, and signature methods demonstrate a clear advantage in capturing structure.
Signatures aren't just theoretical—they've been applied in diverse, real-world domains:

- **Financial Time Series:** Used for anomaly detection—even identifying market manipulation.
- **Brain–Computer Interfaces:** Enable enhanced EEG signal classification to support assistive technologies for the disabled.
- **Chinese Handwriting Recognition:** Signatures capture dynamic pen stroke information, improving accuracy in recognizing complex characters.