Path Signatures

Zhiyu Wang

Mathematical Institute
University of Oxford

Presentation for Department of Mathematics
March 2025

Time-ordered data

- Financial times-series.
- Text: "The quick brown fox jumped over the lazy dog."

- Time-evolving network.

Time-ordered data as paths

We can think of time-ordered data as a path, i.e. a continuous function

$$X : [a, b] \to E$$

for some set E (target space).

- Financial time series: $X : [a, b] \to \mathbb{R}^d$, $[a, b] =$ time horizon, $d =$ number of assets tracked;
- Text: $X : [0, T] \to \mathcal{A}$, $T =$ length of text, $\mathcal{A} = \{a, A, b, B, \dots\} =$ alphabet;
- Time-evolving network: $X : [0, T] \to \mathcal{G}, \mathcal{G} =$ a set of graphs.

Features of paths

We would like to summarise important features of paths.
Goal: introduce a special feature, the **signature** of a path, which have desired properties listed below.

1. Sometimes only the order $X_{t_1}, \dots, X_{t_N}$ matters and the time-paramterisation $(t_1, \dots, t_N)$ is irrelevant (Time-representation invariance);
2. One-to-one correspondence between paths and their features (Uniqueness);
3. We observe only finitely many points $\{X_{t_i}\}_{i=1}^N$, not the whole path $\{X_t\}_{t \in [0,T]}$ (Factorial decay).

Applications

The signature method has been

- combined with convolutional neural nets (CNNs) to win first prize in ICDAR 2013 Online Isolated Chinese Character recognition competition;
- combined with gradient boosting regression to win first prize in PhysioNet 2019 Computing in Cardiology Challenge;
- combined with graph convolutional networks (GCNs) for Skeleton-Based Action Recognition;
- combined with generative adversarial networks (GANs) under the conditional Sig-Wasserstein framework for time series generation...

- Rough Path Theory and Signatures
- Kernel Learning
- Signatures Meet Machine Learning
- Deep Signature Transforms

Rough Path Theory and Signatures

- Paths
- Tensor Algebras
- Path Signatures
- Properties of Path Signatures
  - Geometric Properties
  - Algebraic Properties
  - Upper Estimates
  - Uniqueness
  - Universal Nonlinearity
  - Log Signatures and Lie Algebras
- Rough Paths

Paths

University of Oxford
Mathematical Institute

Paths form one of the basic elements of this chapter. Let us first introduce the concepts of continuous E-valued paths and finite p-variation paths. Let $(E, \|\cdot\|_E)$ be a finite dimensional real Banach space with $\dim(E) = d$.

> **Definition (Path)**
>
> A *path* is a continuous function $X : [a, b] \to E$.

We use the subscript notation $X_t = X(t)$ to denote dependence on $t \in [a, b]$ and superscript notation to denote its coordinates $X_t = (X_t^1, X_t^2, \ldots, X_t^d)$ We denote the space of all continuous paths by $\mathrm{C}^0([a, b], E)$.

p-variation

Definition (p-variation of a path)

Let $p \geq 1$ be a real number. Let $X : [a, b] \to E$ be a continuous path. The *p-variation of X* on the interval $[a, b]$ is defined by

$$\|X\|_{p,[a,b]} = \left[\sup_{a \leq t_0 \leq \ldots \leq t_r \leq b} \sum_{j=0}^{r-1} \|X_{t_j} - X_{t_{j+1}}\|_E^p \right]^{\frac{1}{p}}.$$

Remark

A Brownian motion $B : [a, b] \to \mathbb{R}$ has finite p-variation for $p > 2$. Yet, the 2-variation of B is almost surely infinite as proven in [Fre71].

Topology on path space

Let $\mathcal{V}^p([a,b],E)$ denote the space of continuous paths $X:[a,b]\to E$ of finite p-variation, i.e.

$$\mathcal{V}^p([a,b],E)=\{X\in C^0([a,b],E)\mid \|X\|_{p,[a,b]}<\infty\}.$$

Paths of finite 1-variation are also called bounded variation paths.
It is easy to check that the p-variation $\|\cdot\|_{p,[a,b]}$ is a seminorm and not a norm as all constant paths have zero p-variation. To rectify this, we can define the p-variation norm on $\mathcal{V}^p([a,b],E)$ by

$$\|X\|_{\mathcal{V}^p}=\|X\|_{p,[a,b]}+\sup_{t\in[a,b]}\|X_t\|.$$

Proposition

$(\mathcal{V}^p([a,b],E),\|\cdot\|_{\mathcal{V}^p})$ is a Banach space.

Tensor algebra

It turns out that the path signature which we will define later possesses a rich algebraic structure encapsulated by the tensor algebra.

Definition (Extended tensor algebra/Space of formal series; [LCL07, Definition 2.4])

The *extended tensor algebra* of E, denoted by $T((E))$, is the direct (Cartesian) product

$$T((E)) := \prod_{i=0}^{\infty} E^{\otimes i} = \left\{ \boldsymbol{a} = (a_0, a_1, \ldots, a_n, \ldots) \mid a_n \in E^{\otimes n} \right\}.$$

Here, $E^{\otimes n} = \underbrace{E \otimes \cdots \otimes E}_{n \text{ times}}$ denotes the n-fold tensor product of E with itself. By convention, $E^{\otimes 0} := \mathbb{R}$.

Extended tensor algebra as a vector space and beyond

$T((E))$ is naturally equipped with a vector space structure. Let $\boldsymbol{a} = (a_0, a_1, \ldots)$ and $\boldsymbol{b} = (b_0, b_1, \ldots)$ be two elements of $T((E))$. Let $\lambda \in \mathbb{R}$. Define the addition and scalar multiplication as follows:

$$\boldsymbol{a} + \boldsymbol{b} = (a_0 + b_0, a_1 + b_1, \ldots) \quad \text{and} \quad \lambda \boldsymbol{a} = (\lambda a_0, \lambda a_1, \ldots).$$

Moreover, this vector space becomes a real unital algebra if we define the product of $\boldsymbol{a}$ and $\boldsymbol{b}$ as

$$\boldsymbol{a} \otimes \boldsymbol{b} = (c_0, c_1, \ldots), \quad c_n = \sum_{k=0}^{n} a_k \otimes b_{n-k} \quad \forall\, n \in \mathbb{N}. \tag{1}$$

In fact, we can define norms and inner products on $T((E))$. We will discuss this later.

Subspaces of extended tensor algebra

Definition

Let $T_1((E))$ be the subset of $T((E))$ where

$$T_1((E)) := \{(a_0, a_1, a_2, \ldots) \in T((E)) \mid a_0 = 1\}.$$

Proposition

$(T_1((E)), \otimes)$ *is a group.*

Tensor algebra

Definition (Tensor algebra; [LCL07, Definition 2.5])

The *tensor algebra* over E is the direct sum of n-fold tensor products

$$T((E)) \supset T(E) := \bigoplus_{i=0}^{\infty} E^{\otimes i} = \left\{ \boldsymbol{a} = (a_k)_{k=0}^{\infty} \mid a_k \in E^{\otimes k}, \exists N \in \mathbb{N} \text{ s.t. } a_k = 0 \, \forall k \geq N \right\}.$$

Remark

Note that an element in $T((E))$ is an infinite sequence of possibly infinitely many non-zero tensors, while an element in $T(E)$ is an infinite sequence of finitely many non-zero tensors. In other words, $T(E)$ is a proper subset of $T((E))$.

Truncated tensor algebra

Definition (Truncated tensor algebra)

The *truncated tensor algebra* at order $n \in \mathbb{N}$ is

$$T(E) \supseteq T^{(n)}(E) := \bigoplus_{i=0}^{n} E^{\otimes i} = \left\{ \boldsymbol{a} = (a_k)_{k=0}^{\infty} \mid a_k \in E^{\otimes k}, a_k = 0 \, \forall k \geq n \right\}.$$

The product $\otimes$ given by (1) is also defined on $T^{(n)}(E)$ for $0 \leq k \leq n$. Moreover, note that the dimension of the truncated tensor algebra at order n is given by

$$\dim \left(T^{(n)}(E) \right) = \sum_{i=0}^{n} d^i = \frac{d^{n+1} - 1}{d - 1}.$$

Intuitive understanding of $E^{\otimes n}$

Remark

Informally, we view the 0-fold tensor product $E^{\otimes 0}$ as the set of all scalars $a_0 \in \mathbb{R}$ with dimension $d^0 = 1$. Similarly, $E^{\otimes 1}$ is the set of real vectors $v \in E$ with dimension $d^1 = d$ and $E^{\otimes 2}$ is the set of real matrices $A \in \mathbb{R}^{d \times d}$ with dimension d^2. Thus, a zero tensor at level 2 can be understood as the zero matrix and more formally, let $\{e_{i_1} \otimes e_{i_2} \otimes \cdots \otimes e_{i_n} \mid 0 \leq i_1 \leq m_1, \ldots, 0 \leq i_n \leq m_n\}$ be a basis of $E^{\otimes n}$, then the zero tensor is

$$0 = \sum_{0 \leq i_1 \leq m_1, \ldots, 0 \leq i_n \leq m_n} 0 \cdot \left(e_{i_1} \otimes e_{i_2} \otimes \cdots \otimes e_{i_n}\right).$$

Definition (Canonical homomorphism)

Let $n \in \mathbb{N}$. The *canonical homomorphism* is

$$\pi_n : T((E)) \to T^{(n)}(E), \quad (a_0, a_1, \dots) \mapsto (a_0, a_1, \dots, a_n, 0, \dots)\,.$$

Word

Before formally defining path signatures, let us first introduce the concepts of word and Young integral.

Definition (Words)

A *word* is a finite tuple

$$\mathcal{W}(\mathcal{A}_d) \ni \mathrm{I} = (i_1, \ldots, i_k) =: i_1 \ldots i_k$$

with $i_1, \ldots, i_k \in \{1, 2, \ldots, d\}$. The length of word I is $|\mathrm{I}| := k$. The set $\mathcal{A}_d := \{1, 2, \ldots, d\}$ is the *alphabet*. Moreover, denote by $\mathcal{W}(\mathcal{A}_d)$ the countably infinite set of all words on the alphabet $\mathcal{A}_d$

$$\mathcal{W}(\mathcal{A}_d) := \{\emptyset, 1, \ldots, d, 11, \ldots, dd, \ldots\}.$$

An example of words

Example ($d = 3$)

Let $\mathcal{A}_3 = \{1, 2, 3\}$ be the alphabet. Then

$$\mathcal{W}(\mathcal{A}_3) = \{\emptyset, 1, 2, 3, 11, 12, 13, 21, 22, 23, 31, 32, 33, 111, 112, 113, 121, \ldots\}.$$

Young integrals

Since the paths in our discussion are all E-valued, we need a suitable framework for integration along paths in Banach spaces. This is provided by the Young integral. Given Banach spaces V and W, the Young integral determines a way for a path $[a, b] \to \mathcal{L}(V, W)$ to be integrated along a path $[a, b] \to V$. The notation $\mathcal{L}(V, W)$ denotes the space of bounded linear operators $V \to W$. To be more precise, let $X : [a, b] \to V$ and $Y : [a, b] \to \mathcal{L}(V, W)$. Given a partition $\mathcal{D} = (a = t_0, t_1, \ldots, t_r = b)$ of $[a, b]$ we let $|\mathcal{D}| := \min\{t_{i+1} - t_i \mid i \in \{0, \ldots, r-1\}\}$. We first define

$$\int_{\mathcal{D}} YdX := \sum_{i=0}^{r-1} Y_{t_r}\left(X_{t_{r+1}} - X_{t_r}\right)$$

Then, given $t \in [a, b]$, the Young integral of Y along X over the interval $[a, t]$ is defined as

Young integrals (continued)

$$\int_a^t Y_s dX_s := \lim_{j\to\infty} \int_{\mathcal{D}_j} YdX$$

where $\{\mathcal{D}_j\}_{j=1}^{\infty}$ is a sequence of partitions of $[a,t]$ with $|\mathcal{D}_j| \to 0$ as $j \to \infty$. Formally, we have the following theorem.

Theorem (Young; [LCL07, Theorem 1.16])

Let V and W be two Banach spaces. Let $p, q \in \mathbb{R}_{\geq 1}$ with $\frac{1}{p} + \frac{1}{q} > 1$, $X \in \mathcal{V}^p([a,b], V)$, and $Y \in \mathcal{V}^q([a,b], \mathcal{L}(V,W))$. Then, for each $t \in [a,b]$, the limit

$$\int_0^t Y_s dX_s = \lim_{|\mathcal{D}|\to 0,\, \mathcal{D}\subset[a,t]} \int_{\mathcal{D}} YdX$$

exists. Moreover, the mapping $t \mapsto \int_a^t Y_s dX_s$ belongs to $\mathcal{V}^p([a,b], W)$.

Coordinate iterated integrals

Definition (Iterated integral)

For $p \in [1,2)$, $X \in \mathcal{V}^p([a,b],E)$, word $\mathrm{I} = i_1 \dots i_k$, and domain $[s,t] \subseteq [a,b]$, the *iterated integral* $S(X)^{\mathrm{I}}_{s,\cdot}$ is defined inductively by

$$\mathcal{V}^p([a,b],E) \ni S(X)^{\mathrm{I}}_{s,\cdot} = \int_s^{\cdot} S(X)^{i_1 \dots i_{k-1}}_{s,u} dX^{i_k}_u, \quad S(X)^{\emptyset}_{s,t} = 1.$$

An example of iterated integrals

Mathematical Institute

Example

We can explicitly write out the iterated integral as follows:

$$S(X)^{\emptyset}_{s,t} = 1, \quad S(X)^{i}_{s,t} = \int_s^t dX^i_u = X^i_t - X^i_s,$$
$$S(X)^{ij}_{s,t} = \int_s^t \int_s^u dX^i_v dX^j_u, \quad S(X)^{ijk}_{s,t} = \int_s^t \int_s^{t_3} \int_s^{t_2} dX^i_{t_1} dX^j_{t_2} dX^k_{t_3}.$$

Generally, one has

$$S(X)^{i_1 \dots i_k}_{s,t} = \int_s^t \int_s^{t_k} \cdots \int_s^{t_2} dX^{i_1}_{t_1} \dots dX^{i_{k-1}}_{t_{k-1}} dX^{i_k}_{t_k}.$$

Signature coefficients

To make our notation compact, we use the signature coefficient.

Definition (Signature coefficient)

With the notation as above, define the *signature coefficient* $S(X)_{s,t}^{i_1 \dots i_k}$ indexed by word $i_1 \dots i_k \in \mathcal{W}(\mathcal{A}_d)$ as

$$E \ni S(X)_{s,t}^{i_1 \dots i_k} := \int_{s<t_1<\cdots<t_k<t} dX_{t_1}^{i_1} \cdots dX_{t_k}^{i_k} = \int_s^t \int_s^{t_k} \cdots \int_s^{t_2} dX_{t_1}^{i_1} dX_{t_1}^{i_1} \dots dX_{t_k}^{i_k}.$$

Signature of finite p-variation paths with $p \in [1, 2)$

Now we are ready to give the definition of the signature of finite p-variation paths with $p \in [1, 2)$. The path signature is indexed by words.

Definition (Signature I)

Let $p \in [1, 2)$ and $X \in \mathcal{V}^p([a, b], E)$. The *signature* $S(X)_{s,t}$ of X over domain $[s, t] \subseteq [a, b]$ is the infinite set of real numbers indexed by words $\mathrm{I} \in \mathcal{W}(\mathcal{A}_d)$

$$\left\{S(X)^{\mathrm{I}}_{s,t}\right\}_{I \in \mathcal{W}} = \{\underbrace{S^{\emptyset}_{s,t}}_{1}, \underbrace{S(X)^{1}_{s,t}}_{X^1_b - X^1_a}, \ldots, \underbrace{S(X)^{d}_{s,t}}_{X^d_b - X^d_a}, S(X)^{11}_{s,t}, S(X)^{12}_{s,t}, \ldots, S(X)^{dd}_{s,t}, S(X)^{111}_{s,t}, \ldots\}.$$

Signature of finite p-variation paths with $p \in [1,2)$

Observe that there is a natural grading on the signature $S(X)_{s,t}$ based on the length of indexed words. With a slight abuse of notation, we have

$$S(X)_{s,t} = \left\{1, S^{[1]}_{s,t}(X), S^{[2]}_{s,t}(X), \dots\right\}, \quad S(X)^{[k]}_{s,t} = \left\{S(X)^{\mathrm{I}}_{s,t}\right\}_{|\mathrm{I}|=k}.$$

Thus, the signature can be elegantly encoded as an element in the extended tensor algebra $T((E))$. We present the following alternative definition of the signature.

Signature of finite p-variation paths with $p \in [1, 2)$

Oxford Mathematical Institute

Definition (Signature II; [LCL07, Definition 2.6])

Let $p \in [1, 2)$ and $X \in \mathcal{V}^p([a, b], E)$. The *signature* $S(X)_{s,t}$ of X over domain $[s, t] \subseteq [a, b]$ is the infinite sequence of tensors

$$
\begin{aligned}
T_1((E)) \ni S(X)_{s,t} &:= \left(1, S(X)^{[1]}_{s,t}, S(X)^{[2]}_{s,t}, \ldots, S(X)^{[k]}_{s,t}, \ldots\right) \\
&:= \left(1, \int_{s<t_1<t} dX_t, \ldots, \int_{s<t_1<\cdots<t_k<t} dX_{t_1} \otimes \cdots \otimes dX_{t_k}, \ldots\right),
\end{aligned}
$$

where the *k-fold iterated tensor integral of X_t* is a tensor in $E^{\otimes k}$ defined as

$$
\int_{s<t_1<\cdots<t_k<t} dX_{t_1} \otimes \ldots \otimes dX_{t_k} := \sum_{I \in \mathcal{W}(\mathcal{A}_d), |I| = k} \underbrace{S(X)^{i_1 \ldots i_k}_{s,t}}_{\text{signature coefficient}} e_{i_1} \otimes \cdots \otimes e_{i_k}.
$$

Examples of path signatures

Example (Linear path)

Let $x \in E$ and define $X_t = tx$ for $t \in [0,1]$. We have

$$S(X)_{0,1}^{i_1 \dots i_k} = \frac{x^{i_1} \dots x^{i_k}}{k!}.$$

Example (One-dimensional path)

Consider $X \in \mathcal{V}^1([a,b],E)$. We have

$$S(X)_{a,b}^{1\dots 1} = \frac{(X_b - X_a)^k}{k!}.$$

Hence $S(X)_{a,b}$ only captures the increment for one-dimensional paths.

Truncated signatures

Definition (Truncated signature)

Let $n \in \mathbb{N}_+$. The level-n *truncated signature* is defined as

$$T^{(n)}(E) \ni S^{(n)}(X)_{s,t} = \pi_n\left(S(X)_{s,t}\right).$$

Remark

Note that $S^{[n]}(X)_{s,t} \in E^{\otimes n}$ while $S^{(n)}(X)_{s,t} \in T^{(n)}(E)$.

Geometric properties

Proposition (Time-parametrisation invariance)

Consider $X \in \mathcal{V}^1([a,b],E)$ *and a non-decreasing bijection* $\psi : [x,y] \to [a,b]$*. Define* $Y \in \mathcal{V}^1([x,y],E)$ *by* $Y_t = X_{\psi(t)}$*. Then*

$$S(X)_{a,b} = S(Y)_{x,y}.$$

Proposition (Base-point invariance)

Consider $X \in \mathcal{V}^1([a,b],E)$ *and* $h \in E$*. Define* $Y \in \mathcal{V}^1([a,b],E)$ *as* $Y_t = X_t + h$*. Then*

$$S(X)_{a,b} = S(Y)_{a,b}.$$

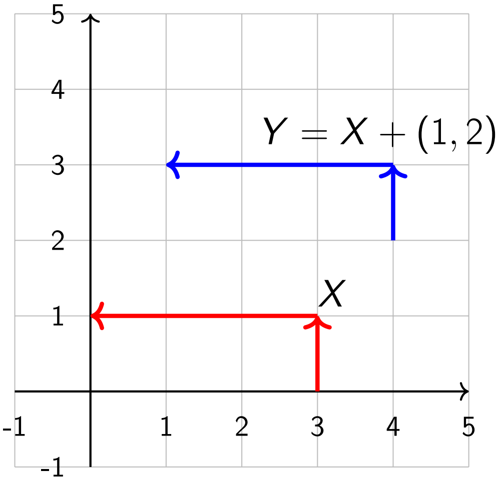


Figure: Illustration of time-parametrisation invariance of signatures

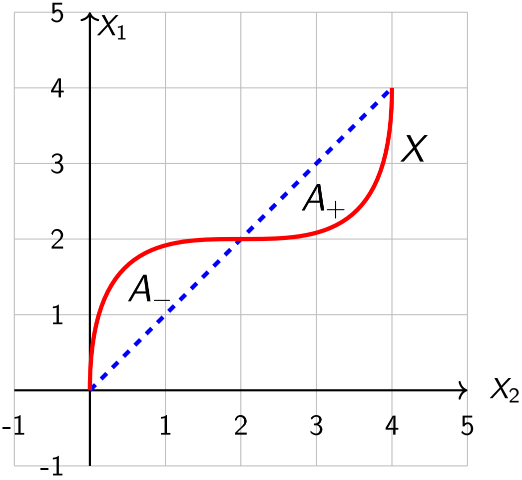


Figure: Lévy areas

Geometric properties (continued)

Definition (Lévy area)

The *Lévy area* of a 2-dimensional path $X = (X^1, X^2) \in \mathcal{V}^1\left([a,b], \mathbb{R}^2\right)$ is the real number $A_+ - A_-$.

Proposition (2nd order signatures compute Lévy area)

One has

$$S(X)^{12}_{a,b} - S(X)^{21}_{a,b} = 2\left(A_+ - A_-\right).$$

Algebraic properties

Proposition (Chen's identity)

Let $X \in \mathcal{V}^1([a,b], E)$ and $c \in [a,b]$. One has

$$S(X)_{a,b} = S(X)_{a,c} \otimes S(X)_{c,b}.$$

Remark

If we were to use Definition (Signature I) to formalise Chen's identity, then we would have had a much more cumbersome form

$$S(X)_{a,b}^{i_1 \dots i_k} = \sum_{m=0}^{k} S(X)_{a,c}^{i_1 \dots i_m} S(X)_{c,b}^{i_{m+1} \dots i_k}, \quad \forall\, i_1 \dots i_k \in \mathcal{W}(\mathcal{A}_d).$$

Concatenation of paths

Definition (Concatenation of paths)

For $X, Y \in \mathcal{V}^1([0,1], E)$, define their concatenation $X * Y \in \mathcal{V}^1([0,1], E)$

$$(X * Y)_t := \begin{cases} X_{2t} & \text{if } t \in [0, 1/2], \\ X_1 + (Y_{2t-1} - Y_0) & \text{if } t \in [1/2, 1]. \end{cases}$$

Proposition (Chen's identity for concatenation of paths)

We have $S(X)_{0,1} \otimes S(Y)_{0,1} = S(X * Y)_{0,1}$.

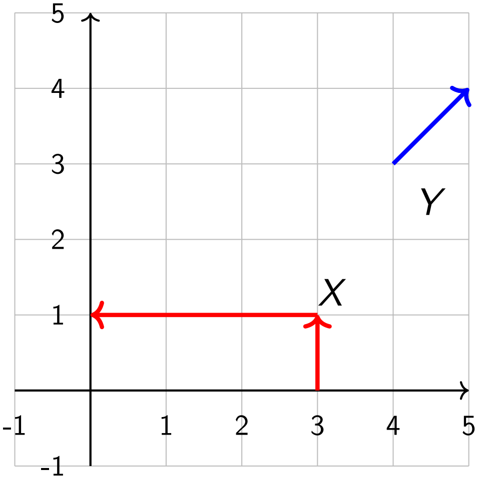


Figure: Paths X and Y

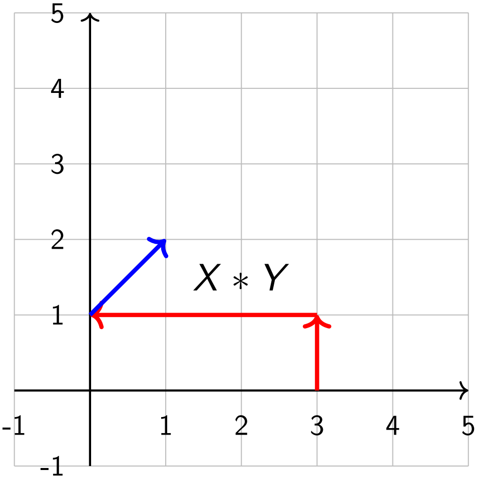


Figure: Concatenation $X * Y$

Time reversal of a path

Definition (Time reversal)

For a path $X : [a,b] \to E$ in $\mathcal{V}^1([0,1], E)$, define its time-reversal $\overleftarrow{X} : [a,b] \to E$ by $\overleftarrow{X}_t = X_{a+b-t}$ for all $t \in [a,b]$.

Proposition

We have

$$S(X)_{a,b} \otimes S(\overleftarrow{X})_{a,b} = 1.$$

Equivalently,

$$S(X)^{i_1 \dots i_k}_{a,b} = (-1)^k S(\overleftarrow{X})^{i_k \dots i_1}_{a,b}.$$

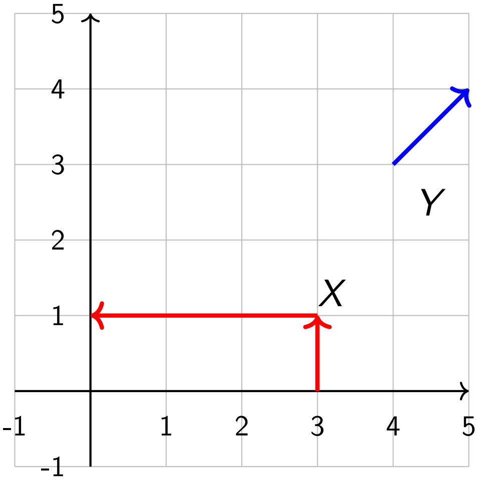


Figure: Paths X and Y

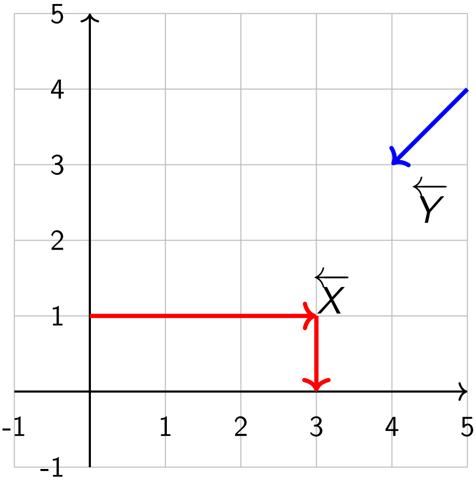


Figure: Paths $\overleftarrow{X}$ and $\overleftarrow{Y}$

A detour to shuffle products

Definition ((k, m)-shuffle)

For $k, m \in \mathbb{N}$, a permutation σ of $\{1, \ldots, k+m\}$ is a (k, m)-*shuffle* if

$$\sigma^{-1}(1) < \ldots < \sigma^{-1}(k) \quad \text{and} \quad \sigma^{-1}(k+1) < \ldots < \sigma^{-1}(k+m).$$

Let $\mathrm{Shuffles}(k, m) \subseteq \mathcal{S}_{k+m}$ denote the set of all (k, m)-shuffles.

Definition (Shuffle product)

Consider two words $\mathrm{I}, \mathrm{J} \in \mathcal{W}(\mathcal{A}_d)$ where $\mathrm{I} = i_1 \ldots i_k$ and $\mathrm{J} = j_1 \ldots j_m$. Let word R be given by

$$\mathrm{R} = r_1 \ldots r_k r_{k+1} \ldots r_{k+m} = i_1 \ldots i_k j_1 \ldots j_m.$$

The *shuffle product* of I and J is the finite multiset of words

$$\mathrm{I} \sqcup\!\sqcup \mathrm{J} = \left\{ r_{\sigma(1)} \ldots r_{\sigma(k+m)} \mid \sigma \in \mathrm{Shuffles}(k, m) \right\}.$$

Algebraic properties (continued)

UNIVERSITY OF OXFORD
Mathematical Institute

Example

$121 \sqcup\!\sqcup 23 = \{\!\{12123, 12213, 12213, 21213, 12231, 12231, 21231, 12321, 21321, 23121\}\!\}$

Proposition (Shuffle product identity)

For $X \in \mathcal{V}^1([a,b],E)$ and words $\mathsf{I}, \mathsf{J} \in \mathcal{W}(\mathcal{A}_d)$, one has

$$S(X)^{\mathsf{I}}_{a,b} S(X)^{\mathsf{J}}_{a,b} = \sum_{\mathsf{K} \in \mathsf{I} \sqcup\!\sqcup \mathsf{J}} S(X)^{\mathsf{K}}_{a,b}.$$

Example

For $X \in \mathcal{V}^1([a,b],E)$, we have

$$S(X)^{12}_{a,b} S(X)^{23}_{a,b} = S(X)^{1223}_{a,b} + S(X)^{1223}_{a,b} + S(X)^{2123}_{a,b} + S(X)^{1232}_{a,b} + S(X)^{2132}_{a,b} + S(X)^{2312}_{a,b}.$$

Extended tensor algebra as a normed space

Recall that $(E, \|\cdot\|_E)$ is a Banach space. In order to equip the extended tensor algebra $T((E))$ with a norm, we define admissible tensor norms.

Definition (Admissible tensor norm; [LCL07, Definition 1.25])

A family of norms on $E^{\otimes n}$ with $n \in \mathbb{N}_+$ is *admissible* if

1. For $n \in \mathbb{N}_+$, the norm $\|\cdot\|_{E^{\otimes n}}$ on $E^{\otimes n}$ is invariant under the action of the symmetric group S_n on $E^{\otimes n}$, i.e.
$$\forall n \in \mathbb{N}_+ \, \forall x \in E^{\otimes n} \, \forall \rho \in S_n, \|\rho(x)\|_{E^{\otimes n}} = \|x\|_{E^{\otimes n}}.$$
Here, for $\rho \in S_n$ and $a_1 \otimes \ldots \otimes a_n \in E^{\otimes n}$, then $\rho\left(a_1 \otimes \ldots \otimes a_n\right) := a_{\rho(1)} \otimes \ldots \otimes a_{\rho(n)}$ and is extended to the entirety of $E^{\otimes n}$ by linearity.
2. For $n, m \in \mathbb{N}_+$, the norm $\|\cdot\|_{E^{\otimes n}}$ is submultiplicative, i.e.
$$\forall n, m \in \mathbb{N}_+ \, x, y \in E^{\otimes n}, \|x \otimes y\|_{E^{\otimes (n+m)}} \leq \|x\|_{E^{\otimes n}} \|y\|_{E^{\otimes m}}. \tag{2}$$

Extended tensor algebra as a normed space (continued)

Now we can define a norm on $T((E))$ by

$$\|v\|_{T((E))} := \sum_{n=0}^{\infty} \|\pi_n(v)\|_{E^{\otimes n}} .$$

With the properties of the admissible tensor norm $\|\cdot\|_{E^{\otimes n}}$, one can check that $\|\cdot\|_{T((E))}$ is indeed a norm on $T((E))$.

Controls

Let $\Delta_{[a,b]} := \{(s,t) \in \mathbb{R}^2 \mid a \leq s \leq t \leq b\}$ be a simplex.

Definition (Control; [LCL07, Definition 1.9])

A *control* on $[a,b]$ is a continuous function $\omega : \Delta_{[a,b]} \to [0,\infty)$ such that

1. ω is super-additive, i.e.

$$\forall s,t,u \in [a,b] \text{ with } s \leq u \leq t, \omega(s,u) + \omega(u,t) \leq \omega(s,t);$$

2. ω vanishes on the diagonal, i.e.

$$\forall t \in [a,b], \omega(t,t) = 0.$$

Examples of controls

Mathematical Institute

Example

Let $p \in \mathbb{R}_{\geq 1}$ and $X \in \mathcal{V}^p([a,b],E)$. Define $\omega_X : \Delta_{[a,b]} \to [0,\infty)$ as

$$\omega_X(s,t) := \|X\|^p_{p,[s,t]} = \sup_{\mathcal{D}\subset[s,t]} \sum_{i=0}^{r-1} \left\|X_{t_i} - X_{t_{i+1}}\right\|^p_E .$$

Then ω_X is a control.

Proposition

Let $p \in [1,2)$ and $X \in \mathcal{V}^p([a,b],E)$. Then there exists a control $\omega : \Delta_{[a,b]} \to [0,\infty)$ such that for every $n \in \mathbb{N}_+$ and every $(s,t) \in \Delta_{[a,b]}$, we have

$$\left\| S^{[n]}(X)_{s,t} \right\|_{E^{\otimes n}} \leq \frac{\omega(s,t)^n}{\Gamma\left(1+\frac{n}{p}\right)},$$

where $\Gamma(\cdot)$ is the Gamma function.

Factorial decay

Corollary (Factorial decay; [LCL07, Proposition 2.2])

For $p = 1$*, we have for every* $n \in \mathbb{N}_+$ *and every* $(s, t) \in \Delta_{[a,b]}$

$$\left\| S^{[n]}(X)_{s,t} \right\|_{E^{\otimes n}} \leq \frac{\|X\|_{1,[s,t]}^{n}}{n!}.$$

Consequently, truncating the signature to a finite depth effectively captures the most significant terms in a norm-based sense. We will see later that this factorial decay property not only reduces computational complexity but also makes it feasible to use signatures as features in machine learning models.

Height functions

Definition (Height function; [HL10, Definition 1.3])

Given a path $X : [a, b] \to E$, a *height function* for X is a non-negative continuous function $h : [a, b] \to \mathbb{R}$ such that $h(a) = h(b) = 0$ and

$$\|X_t - X_s\| \leq h(s) + h(t) - 2 \inf_{u \in [s,t]} h(u) \text{ for all } a \leq s < t \leq b.$$

Tree-like equivalence

Definition (Tree-like path; [HL10, Definition 1.3])

A bounded variation path is called *tree-like* if it has a height function.

Definition (Tree-like equivalence; [HL10, Definition 1.4])

Let $X, Y \in \mathcal{V}^1([a,b], E)$. X and Y are *tree-like equivalent* if $X * \overleftarrow{Y}$, the concatenation of X with the reversed path of Y, is a tree-like path. We denote it by $X \sim Y$.

Proposition

Tree-like equivalence $\sim$ is an equivalence relation.

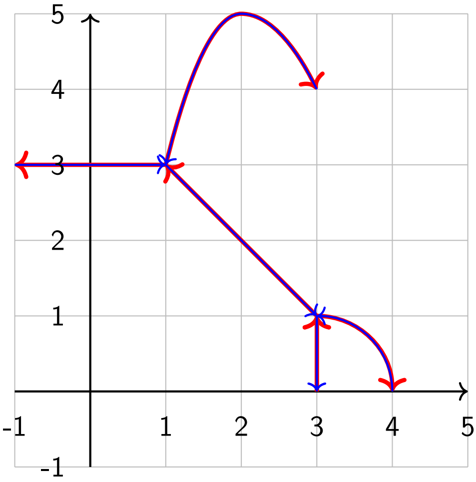


Figure: A tree-like path

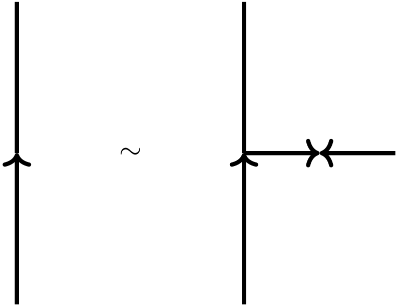


Figure: Tree-like equivalent paths

Informally, a path is called tree-like if it completely back-tracks itself.

Signature determines paths up to tree-like equivalence

Theorem (Uniqueness of the signature)
Let $X, Y \in \mathcal{V}^1([a,b], E)$. Then X is tree-like equivalent to Y if and only if $S(X) = S(Y)$.

The following proposition is a useful sufficient condition for the uniqueness of the signature.

Proposition
Let $X \in \mathcal{V}^1([a,b], E)$ with a fixed starting point and at least one coordinate of X is a monotone function. Then $S(X)$ determines X uniquely.

Quotient space of bounded variation paths $\mathcal{P}^1([a,b],E)$

One of the most fundamental results in the study of the signature is its **universal nonlinearity**, which states that continuous functions on path space can be well approximated by linear functionals on the signature. We first define a quotient space of bounded variation paths modulo the tree-like equivalence relation.

Definition

Define $\mathcal{P}^1([a,b],E) := \mathcal{V}^1([a,b],E)/\sim$. Any element in $\mathcal{P}^1([a,b],E)$ is an equivalence class $[X]$ for some $X \in \mathcal{V}^1([a,b],E)$. By Theorem 49, we define the signature of $[X] \in \mathcal{P}^1([a,b],E)$ as the signature of any representative $\tilde{X} \in \mathcal{V}^1([a,b],E)$ of the equivalence class $[X]$.

Topology on $\mathcal{P}^1([a,b],E)$

Definition ([LQ02, Equation (3.70)])

Given $[X],[Y] \in \mathcal{P}^1([a,b],E)$, define

$$d_{\mathcal{P}^1}([X],[Y]) := \sup_{a \le t_0 \le \ldots \le t_r \le b} \left(\sum_{j=0}^{r-1} \left\| S(X)^1_{t_j,t_{j+1}} - S(Y)^1_{t_j,t_{j+1}} \right\|_E \right), \tag{3}$$

where $S(X)^1_{t_j,t_{j+1}}$ and $S(Y)^1_{t_j,t_{j+1}}$ are signature coefficients defined as in Definition 18.

Remark

Theorem (Uniqueness of the signature) ensures that (3) is well-defined.

Signature is a universal approximator of continuous functions

Theorem (Universal Nonlinearity)

Let $\mathcal{K} \subset \mathcal{P}^1([a,b],E)$ be compact and $f : \mathcal{K} \to \mathbb{R}$ a continuous function with respect to the topology generated by (3). Then for any $\varepsilon > 0$ there exists a truncation level $k \in \mathbb{N}$ and $\alpha_{i,\mathrm{I}} \in \mathbb{R}$ such that for all $X \in \mathcal{K}$

$$\left| f(x) - \sum_{i=0}^{k} \sum_{\mathrm{I} \in \mathcal{W}(\mathcal{A}_d)} \alpha_{i,\mathrm{I}} S(X)^{\mathrm{I}}_{a,b} \right| \leq \varepsilon.$$

Neural networks are universal approximators

Theorem ([PHGDGN20, Theorem 1])

σ is not polynomial iff for every continuous function $f : K \subset \mathbb{R}^d \to \mathbb{R}$ defined on a compact set K and $\varepsilon > 0$, there exists a two-layer neural network with m neurons and weights $\mathrm{W}, \mathrm{b}, \mathrm{v}$ s.t.

$$\max_{x \in K} \left| f(x) - \sum_{j \leq m} v_j \sigma \left(w_j^{\mathrm{T}} x + b_j \right) \right| < \varepsilon$$

- Fixed number of layers ("bounded depth")
- Does not tell how many neurons m are needed ("arbitrary width")
- There are stronger results, including bounded depth and width

Exponential function on extended tensor algebra

Definition (Tensor exponential)

The *tensor exponential* $\exp : T((E)) \to T((E))$ is defined as

$$T((E)) \ni \exp(\boldsymbol{a}) := \sum_{n=0}^{\infty} \frac{\boldsymbol{a}^{\otimes n}}{n!} = 1 + \boldsymbol{a} + \frac{\boldsymbol{a}^{\otimes 2}}{2!} + \frac{\boldsymbol{a}^{\otimes 3}}{3!} + \cdots, \tag{4}$$

where $1 = (1, 0, 0, \ldots)$.

Remark

The exponential series defined in (4) is convergent as shown in [LCL07, Lemma 2.19] in the sense that the level-k tensor of $\exp(\boldsymbol{a})$ converges in $E^{\otimes k}$ under the admissible tensor norm $\| \cdot \|_{E^{\otimes k}}$.

Logarithm function on extended tensor algebra

Definition (Tensor logarithm)

The *tensor logarithm* $\log : T_{>0}((E)) \to T((E))$ is defined as

$$T((E)) \ni \log(\boldsymbol{a}) := \log(\pi_0(\boldsymbol{a})) + \sum_{n=1}^{\infty} \frac{(-1)^n}{n} \left(1 - \frac{\boldsymbol{a}}{\pi_0(\boldsymbol{a})}\right)^{\otimes n}, \tag{5}$$

where $1 = (1, 0, 0, \dots)$ and $\pi_0 : T((E)) \to \mathbb{R}$ is the canonical homomorphism.

Remark

On the other hand, the tensor logarithm series is well-defined in (5) as it only contains finitely many terms at each level. This can be seen by noting that

$$1 - \frac{\boldsymbol{a}}{\pi_0(\boldsymbol{a})} = \left(0, -\frac{\pi_1(\boldsymbol{a})}{\pi_0(\boldsymbol{a})} \dots\right)$$

and hence $\left(1 - \frac{\boldsymbol{a}}{\pi_0(\boldsymbol{a})}\right)_k^{\otimes n} = 0$ for all $k < n$.

Exp and Log are inverses of each other

Moreover, if we restrict the tensor exponential and logarithm on certain subsets, then they are inverses of each other.
Recall that $T_1((E)) := \{(a_0, a_1, a_2, \ldots) \in T((E)) \mid a_0 = 1\}$. Similarly, let $T_{>0}((E)) := \{(a_0, a_1, a_2, \ldots) \in T((E)) \mid a_0 > 0\}$ and $T_0((E)) := \{(a_0, a_1, a_2, \ldots) \in T((E)) \mid a_0 = 0\}$.

Proposition

The restricted exponential and logarithm function $\exp : T_0((E)) \to T_1((E))$ *and* $\log : T_1((E)) \to T_0((E))$ *are bijective and inverses of each other.*

Signatures have redundancy

Shuffle product identity $\Rightarrow$ signature terms not algebraically independent. Recall the example we have seen earlier:

Example

For $X \in \mathcal{V}^1([a,b],E)$, we have

$$S(X)^{12}_{a,b}S(X)^{23}_{a,b} = S(X)^{1223}_{a,b}+S(X)^{1223}_{a,b}+S(X)^{2123}_{a,b}+S(X)^{1232}_{a,b}+S(X)^{2132}_{a,b}+S(X)^{2312}_{a,b}.$$

To remove this redundancy, we introduce the Log signature.

Log signatures and truncated log signature

Definition (Log signature)

Let $p \in [1, 2)$ and $X \in \mathcal{V}^p([a, b], E)$. The *log signature* $\log S(X)_{s,t}$ of X over domain $[s, t] \subseteq [a, b]$ is

$$T_0((E)) \ni \log S(X)_{s,t} \overset{(5)}{:=} \sum_{n=1}^{\infty} \frac{(-1)^n}{n} (1 - S(X)_{s,t})^{\otimes n}.$$

Definition (Truncated log signature)

Let $n \in \mathbb{N}_+$. The level-n *truncated log signature* is defined as

$$T^{(n)}(E) \ni \log^{(n)} S(X)_{s,t} = \pi_n (\log S(X)_{s,t}).$$

Lie algebra

Definition (Lie algebra)

A *Lie algebra* $\mathfrak{g}$ over $\mathbb{R}$ is a real vector space with a map $[\cdot,\cdot] : \mathfrak{g} \times \mathfrak{g} \to \mathfrak{g}$ called the *Lie bracket* such that

1. the Lie bracket $[\cdot,\cdot]$ is bilinear;
2. $[x, y] = -[y, x]$;
3. Jacobi's identity holds: $[x,[y,z]] + [y,[z,x]] + [z,[x,y]] = 0$.

Proposition

Let $[\cdot,\cdot] : T((E)) \times T((E)) \to T((E))$ *be defined as*

$$[\boldsymbol{a}, \boldsymbol{b}] = \boldsymbol{a} \otimes \boldsymbol{b} - \boldsymbol{b} \otimes \boldsymbol{a}. \tag{6}$$

Then $(T((E)), [\cdot,\cdot])$ *forms a Lie algebra.*

Subspaces generated by lie brackets

Oxford Mathematical Institute

Definition

Let F_1 and F_2 be subspaces of $T((E))$. Define

$$[F_1, F_2] := \operatorname{span}\left(\{[\boldsymbol{a}, \boldsymbol{b}] \mid \boldsymbol{a} \in F_1 \text{ and } \boldsymbol{b} \in F_2\}\right).$$

Now a sequence of subspaces $(L_i)_{i=0}^{\infty}$ of $T((E))$ can be recursively defined as

$$L_0 = \{0\}, \quad L_1 = E, \quad L_2 = [E, L_1] = [E, E], \quad L_3 = [E, L_2] = [E, [E, E]], \quad \ldots,$$

and generally, $L_{n+1} := [E, L_n] \subset E^{\otimes(n+1)}$.

Log signatures live in the space of lie formal series

Definition (Lie formal series; [LCL07, Definition 2.22])

The space of *Lie formal series* over E is the subspace $\mathcal{L}((E)) \subset T((E))$ given by

$$\mathcal{L}((E)) := \{ \boldsymbol{l} \in T((E)) \mid \forall n \in \mathbb{Z}_{\geq 0},\, \pi_n(\boldsymbol{l}) \in L_n \} \,.$$

Moreover, for $n \in \mathbb{N}$ we define the *Lie polynomials of degree* n to be the subset

$$\mathcal{L}^{(n)}(E) := \pi_n(\mathcal{L}((E))) = \bigoplus_{i=0}^{n} L_i .$$

Proposition

We have $\log S(X)_{s,t} \in \mathcal{L}((E))$ *and* $\log^{(n)} S(X)_{s,t} \in \mathcal{L}^{(n)}(E)$.

Rough paths

Note that we have only defined the signature of finite p-variation paths with $p \in [1, 2)$.
In the case where $p \geq 2$, we need a new framework beyond the Young integral.
For instance, for a Brownian motion, on the other hand, the integrals can be understood in the sense of Itô or Stratonovich.

- $G\Omega_p(E) \subset WG\Omega_p(E) \subset \Omega_p(E)$
- We can define the signature of a *p-rough paths*

Kernel Learning

Mathematical Institute

- Reproducing Kernel Hilbert Spaces
- From Kernels to RKHSs
- Operations with Kernels
- Signature Kernels

Evaluation functionals

Definition (Evaluation functional)

Let $\mathcal{H} \subseteq \mathbb{R}^{\mathcal{X}}$ be a Hilbert space of functions. The *evaluation functional* at a fixed $x \in \mathcal{X}$ is the map

$$\delta_x : \mathcal{H} \to \mathbb{R}, \quad f \mapsto f(x).$$

Remark

Evaluation functionals are always linear. For $f, g \in \mathcal{H}$ and $\lambda \in \mathbb{R}$, we have $\delta_x(f + \lambda g) = (f + \lambda g)(x) = f(x) + \lambda g(x) = \delta_x(f) + \lambda \delta_x(g)$. However, evaluation functionals are not always continuous.

An example of discontinuous evaluation functional

Example (Discontinuous evaluation functional)

Let μ be the Lebesgue measure on $\mathbb{R}$. Consider the Hilbert space $(L^2([0,1];\mu), \|\cdot\|_{L^2})$ of square integrable equivalence classes of functions where $\forall f_1, f_2 \in L^2([0,1])$,

$$\|f_1 - f_2\|_{L^2} = \left(\int_0^1 |f_1(x) - f_2(x)|^2 \, d\mu(x) \right)^{\frac{1}{2}}.$$

Now, observe that the sequence of monomials $(q_n)_{n=1}^{\infty}$ with $q_n = x^n$ converges to the zero function in L^2 norm since

$$\lim_{n\to\infty} \|q_n - 0\|_{L^2} = \lim_{n\to\infty} \left(\int_0^1 x^{2n} dx \right)^{\frac{1}{2}} = \lim_{n\to\infty} \frac{1}{\sqrt{2n+1}} = 0$$

but does not converge pointwise at $x = 1$ since

$$1 = \lim_{n\to\infty} \delta_1(q_n) \neq \delta_1\left(\lim_{n\to\infty} q_n \right) = 0.$$

Mathematical Institute

Definition (RKHS)

A Hilbert space of functions $\mathcal{H} \subseteq \mathbb{R}^{\mathcal{X}}$ is a *Reproducing Kernel Hilbert Space (RKHS)* if the evaluation functional δ_x is continuous for all $x \in \mathcal{X}$.

Remark

It may seem evident from Example (Discontinuous evaluation functional) that $L^2([0,1];\mu)$ is not an RKHS. However, it fails to be a valid counterexample since it consists of equivalence classes of functions rather than individual functions, and thus is not strictly a space of functions. Consequently, it does not even fall within the scope of the discussion. Only by invoking the *Axiom of Choice* can one provide a non-constructive proof of the existence of a Hilbert space of functions that is not an RKHS.

Functions are well-behaved in RKHSs

A notable consequence is that RKHSs exhibit particularly well-behaved properties compared to general Hilbert spaces of functions. Specifically, we have the property that two functions close in RKHS norm implies that they are also close posintwise.

Proposition ([BTA04, Corollary 1])

Let $\mathcal{H} \subseteq \mathbb{R}^{\mathcal{X}}$ be an RKHS. If $\lim_{n\to\infty} \|f_n - f\|_{\mathcal{H}} = 0$, then $\lim_{n\to\infty} f_n(x) = f(x)$ for all $x \in \mathcal{X}$.

Reproducing kernels

The definition of an RKHS in does not explicitly mention a kernel. We now introduce reproducing kernels, kernels, and positive semi-definite functions to illustrate how they relate to RKHSs.

Definition (Reproducing kernel; [BTA04, Definition 1])

Let $\mathcal{H} \subseteq \mathbb{R}^{\mathcal{X}}$ be a Hilbert space of functions. A bivariate function $k : \mathcal{X} \times \mathcal{X} \to \mathbb{R}$ is a *reproducing kernel* of $\mathcal{H}$ if

1. $\forall x \in \mathcal{X}, k(\cdot, x) \in \mathcal{H}$;
2. reproducing property holds: $\forall x \in \mathcal{X}\, \forall f \in \mathcal{H}, \langle f, k(\cdot, x) \rangle_{\mathcal{H}} = f(x)$.

Existence and uniqueness of reproducing kernels

Proposition (Existence of reproducing kernel; [BTA04, Theorem 1])
Let $\mathcal{H} \subseteq \mathbb{R}^{\mathcal{X}}$ be a Hilbert space of functions. Then $\mathcal{H}$ is an RKHS if and only if there exists a reproducing kernel of $\mathcal{H}$.

Proposition (Uniqueness of reproducing kernel)
Let $\mathcal{H} \subseteq \mathbb{R}^{\mathcal{X}}$ be an RKHS. The reproducing kernel of $\mathcal{H}$ is unique.

Positive semi-definite functions

Let us now characterise reproducing kernels with positive semi-definite functions.

Definition (Positive semi-definite function; [BTA04, Definition 2])

A bivariate function $k : \mathcal{X} \times \mathcal{X} \to \mathbb{R}$ is *positive semi-definite* if k is symmetric and for any $(x_i)_i^n \subseteq \mathcal{X}, (\alpha_i)_{i=1}^n \subseteq \mathbb{R}$

$$\sum_{i=1}^{n} \sum_{j=1}^{n} \alpha_i \alpha_j k\left(x_i, x_j\right) \geq 0.$$

Relation to positive semi-definite matrices

Remark

One can immediately see that a bivariate function $k : \mathcal{X} \times \mathcal{X} \to \mathbb{R}$ is positive semi-definite if and only if the *Gram matrix* of k

$$K := \begin{bmatrix} k(x_1, x_1) & \cdots & k(x_1, x_n) \\ \vdots & \ddots & \vdots \\ k(x_n, x_1) & \cdots & k(x_n, x_n) \end{bmatrix}$$

is positive semi-definite for any $(x_i)_i^n \subseteq \mathcal{X}$.

Inner products are positive semi-definite

In fact, every inner product is a positive semi-definite function. More generally, we have this following useful proposition.

Proposition

Let $\mathcal{H}$ be a Hilbert space and $\phi : \mathcal{X} \to \mathcal{H}$. Then the bivariate function $h : \mathcal{X} \times \mathcal{X} \to \mathbb{R}$

$$h(x, y) = \langle \phi(x), \phi(y) \rangle_{\mathcal{H}}$$

is positive semi-definite.

Corollary ([BTA04, Lemma 2])

Let $\mathcal{H} \subseteq \mathbb{R}^{\mathcal{X}}$ be an RKHS. The reproducing kernel $k : \mathcal{X} \times \mathcal{X} \to \mathbb{R}$ of $\mathcal{H}$ is a positive semi-definite function.

Kernels

Kernel methods are a versatile algorithmic framework which allows construction of nonlinear machine learning algorithms by employing linear tools in a nonlinearly transformed feature space. Let us now introduce the definition of a kernel.

Definition (Kernel; [SC08, Definition 4.1])

A bivariate function $k : \mathcal{X} \times \mathcal{X} \to \mathbb{R}$ is a *kernel* if there exists a Hilbert space $\mathcal{H}$ called *feature space* and a *feature map* $\varphi : \mathcal{X} \to \mathcal{H}$ such that for any $x, x' \in \mathcal{X}$

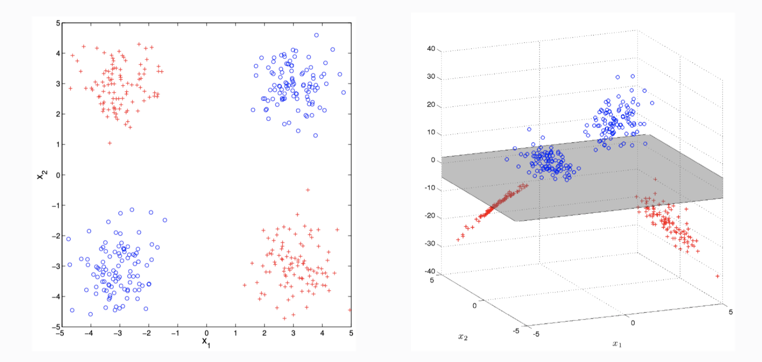
$$k\left(x, x'\right) = \langle \varphi(x), \varphi\left(x'\right) \rangle_{\mathcal{H}}.$$

Proposition

Kernels are positive semi-definite functions.

Proposition (Reproducing kerenls are kernels; [SC08, Lemma 4.19])

Let $\mathcal{H} \subseteq \mathbb{R}^{\mathcal{X}}$ be an RKHS. The reproducing kernel of $\mathcal{H}$ is a kernel.



- No linear classifier separates red from blue.
- Linear separation after mapping to a higher dimensional feature space:

$$x \mapsto \varphi(x) = \begin{pmatrix} x^{(1)} & x^{(2)} & x^{(1)}x^{(2)} \end{pmatrix}^\top \in \mathbb{R}^3$$

Feature maps and feature spaces are not unique for a given kernel

Example

Consider $\mathcal{X} = \mathbb{R}^p$ with $p \in \mathbb{N}_+$ and the kernel $k : \mathcal{X} \times \mathcal{X} \to \mathbb{R}$ given by

$$k(x, y) = xy = \begin{bmatrix} \frac{x}{\sqrt{2}} & \frac{x}{\sqrt{2}} \cdots \frac{x}{\sqrt{2}} \end{bmatrix} \begin{bmatrix} \frac{y}{\sqrt{2}} \\ \frac{y}{\sqrt{2}} \\ \vdots \\ \frac{y}{\sqrt{2}} \end{bmatrix}.$$

We can define two feature maps

$$\phi_1(x) = x, \quad \phi_2(x) = \begin{bmatrix} \frac{x}{\sqrt{2}} & \frac{x}{\sqrt{2}} \cdots \frac{x}{\sqrt{2}} \end{bmatrix}$$

and their corresponding feature spaces

$$\mathcal{H}_1 = \mathbb{R}^p, \quad \mathcal{H}_2 = \mathbb{R}^{2p}.$$

From Kernels to RKHSs

Previously, we established that for any RKHS $\mathcal{H}$, there exists a unique reproducing kernel associated with $\mathcal{H}$, which is a positive semi-definite function. We then explored kernels—functions expressible as inner products in a feature space—observing that all reproducing kernels are valid kernels. In example above (feature maps and feature spaces are not unique for a given kernel), we saw that the representation of a kernel as an inner product in a feature space may not be unique. However, neither of the feature spaces in that example is an RKHS, as they are not spaces of functions on $\mathcal{X} = \mathbb{R}^p$.
In this section, we essentially aim to show that the converse of Proposition (kernels are positive semi-definite functions) holds. **Namely, for every positive semi-definite function $k : \mathcal{X} \times \mathcal{X} \to \mathbb{R}$, there exists a unique RKHS $\mathcal{H}$ for which k is a reproducing kernel of $\mathcal{H}$.** Let us first give a preview of how such an RKHS is constructed and then give detailed proofs.

Outline of construction of RKHSs

1. Define $\mathcal{H}_0 := \operatorname{span}\{k(\cdot, x)\}_{x\in\mathcal{X}} \subseteq \mathbb{R}^{\mathcal{X}}$;
2. Define a bivariate function on $\mathcal{H}_0$ by $\langle f, g\rangle_{\mathcal{H}_0} = \sum_{i=1}^{n}\sum_{j=1}^{m} \alpha_i\beta_j k\left(y_j, x_i\right)$;
3. Prove that $(\mathcal{H}_0, \langle\cdot,\cdot\rangle_{\mathcal{H}_0})$ is an inner product space;
4. Prove that $\mathcal{H}_0$ is a pre-RKHS;
5. Define $\mathcal{H} := \{f : \mathcal{X} \to \mathbb{R} \mid \exists$ Cauchy $(f_n)_{n=1}^{\infty} \subseteq \mathcal{H}_0$ s.t. $f_n \to f$ pointwise$\}$ so that $\mathcal{H}_0 \subseteq \mathcal{H}$;
6. Prove that an inner product on $\mathcal{H}$ can be defined by $\langle f, g\rangle_{\mathcal{H}} := \lim_{n\to\infty} \langle f_n, g_n\rangle_{\mathcal{H}_0}$;
7. Prove that $(\mathcal{H}, \langle\cdot,\cdot\rangle_{\mathcal{H}})$ is a Hilbert space;
8. Prove that evaluation functionals are continuous on $\mathcal{H}$ and hence $\mathcal{H}$ is an RKHS with the reproducing kernel k.

Pre-RKHSs

Definition (Pre-RKHS)

Let $\mathcal{H}_0 \subseteq \mathbb{R}^{\mathcal{X}}$ be an inner product space of functions with the inner product $\langle \cdot, \cdot \rangle_{\mathcal{H}_0}$. $\mathcal{H}_0$ is a *pre-RKHS* if

1. the evaluation functional δ_x is continuous for all $f \in \mathcal{H}_0$;
2. any Cauchy sequence $(f_n)_{n=1}^{\infty} \subseteq \mathcal{H}_0$ converging pointwise to 0 also converges in the induced $\mathcal{H}_0$-norm to 0.

Construction of a pre-RKHS

Proposition (Construction of a pre-RKHS)

Let $k : \mathcal{X} \times \mathcal{X} \to \mathbb{R}$ be a positive semi-definite function. Define the space of functions $\mathcal{H}_0$ as

$$\mathcal{H}_0 := \left\{ \sum_{i=1}^{n} \alpha_i k(\cdot, x_i) : \mathcal{X} \to \mathbb{R} \mid n \in \mathbb{N}_+, \alpha_1, \ldots, \alpha_n \in \mathbb{R}, x_1, \ldots, x_n \in \mathcal{X} \right\}. \tag{7}$$

Define a bivariate function on $\mathcal{H}_0$ by

$$\langle f, g \rangle_{\mathcal{H}_0} = \sum_{i=1}^{n} \sum_{j=1}^{m} \alpha_i \beta_j k(y_j, x_i) \tag{8}$$

where $f = \sum_{i=1}^{n} \alpha_i k(\cdot, x_i)$ and $g = \sum_{j=1}^{m} \beta_j k(\cdot, y_j)$. Then $(\mathcal{H}_0, \langle \cdot, \cdot \rangle_{\mathcal{H}_0})$ is an inner product space. Moreover, $\mathcal{H}_0$ is a pre-RKHS.

From $\mathcal{H}_0$ to $\mathcal{H}$

Definition

Let $\mathcal{H} \subseteq \mathbb{R}^{\mathcal{X}}$ be a Hilbert space of functions such that there exists a Cauchy sequence $(f_n)_{n=1}^{\infty} \subseteq \mathcal{H}_0$ converging pointwise to f for all $f \in \mathcal{H}$, i.e.

$$\mathcal{H} := \{f : \mathcal{X} \to \mathbb{R} \mid \exists \text{ Cauchy } (f_n)_{n=1}^{\infty} \subseteq \mathcal{H}_0 \text{ s.t. } f_n \to f \text{ pointwise } \}. \tag{9}$$

Remark

Clearly, we have $\mathcal{H}_0 \subseteq \mathcal{H} \subseteq \mathbb{R}^{\mathcal{X}}$.

Inner product on $\mathcal{H}$

Proposition ([BTA04, Lemma 5])

For $f, g \in \mathcal{H}$ and Cauchy sequences $(f_n)_{n=1}^{\infty}$, $(g_n)_{n=1}^{\infty} \subseteq \mathcal{H}_0$ converging pointwise to f and g, we define $a_n = \langle f_n, g_n \rangle_{\mathcal{H}_0}$. Then the sequence of real numbers $(a_n)_{n=1}^{\infty}$ is convergent and its limit depends only on f and g.

Proposition ([BTA04, Lemma 6])

Let $(f_n)_{n=1}^{\infty} \subseteq \mathcal{H}_0$ be a Cauchy sequence converging pointwise to $f \in \mathcal{H}$. If $\lim_{n\to\infty} \langle f_n, f_n \rangle_{\mathcal{H}_0} = 0$, then $f = 0$.

We can thus define an inner product on $\mathcal{H}$ by

$$\langle f, g \rangle_{\mathcal{H}} := \lim_{n\to\infty} \langle f_n, g_n \rangle_{\mathcal{H}_0}.$$

$\mathcal{H}$ is an RKHS

Mathematical Institute

Proposition ([BTA04, Corollary 2])

$\mathcal{H}_0$ is dense in $\mathcal{H}$.

Proposition ([BTA04, Lemma 8])

The evaluation functionals are continuous on $\mathcal{H}$.

Proposition ([BTA04, Lemma 8])

$\mathcal{H}$ is an RKHS.

Now we are ready to present the converse of Proposition (kernels are positive semi-definite functions).

Moore-Aronszajn

Theorem (Moore-Aronszajn; [BTA04, Theorem 3])

For every positive semi-definite function $k : \mathcal{X} \times \mathcal{X} \to \mathbb{R}$, *there exists a unique RKHS* $\mathcal{H}$ *for which* k *is a reproducing kernel of* $\mathcal{H}$.

Operations with kernels

New kernels can be constructed by applying operations to existing ones. We now establish key algebraic properties of the set of kernels on $\mathcal{X}$.

> **Proposition (Sums of kernels; [SC08, Lemma 4.5])**
> *Let $k_1 : \mathcal{X} \times \mathcal{X} \to \mathbb{R}$ and $k_2 : \mathcal{X} \times \mathcal{X} \to \mathbb{R}$ be kernels. Then $k_1 + k_2 : \mathcal{X} \times \mathcal{X} \to \mathbb{R}$ is also a kernel.*

Extended tensor algebra as an inner product space

Before formally defining the signature kernel, we need to define an inner product on $T((E))$. Let $\mathcal{B} = \{e_1, \ldots, e_d\}$ be a basis for E. Let $\mathcal{B}^* = \{e_1^*, \ldots, e_d^*\}$ be the correspondingly induced dual basis on E^*. For each $n \in \mathbb{Z}_{\geq 2}$, $\mathcal{B}$ determines the basis

$$\mathcal{B}^{\otimes n} := \{e_{\mathrm{K}} = e_{k_1} \otimes \ldots \otimes e_{k_n} \mid \mathrm{K} = k_1 \cdots k_n \in \mathcal{W}(\mathcal{A}_d)\}$$

for $E^{\otimes n}$ and the corresponding dual basis

$$(\mathcal{B}^*)^{\otimes n} := \left\{e_{\mathrm{K}}^* = e_{k_1}^* \otimes \ldots \otimes e_{k_n}^* \mid \mathrm{K} = k_1 \cdots k_n \in \mathcal{W}(\mathcal{A}_d)\right\}$$

for $(E^*)^{\otimes n}$. We first use the basis $\mathcal{B}$ to define an inner product on E. To do so we set

$$\langle e_i, e_j \rangle_E := \delta_{ij} = \begin{cases} 1 & \text{if} \quad i = j \\ 0 & \text{if} \quad i \neq j \end{cases}$$

and extending bilinearly to the whole of E.

Extended tensor algebra as an inner product space (continued)

Given any $n \in \mathbb{Z}_{\geq 2}$ we use the basis $\mathcal{B}^{\otimes n}$ to define an inner product on $E^{\otimes n}$. We set

$$\langle e_{i_1} \otimes \ldots \otimes e_{i_n}, e_{j_1} \otimes \ldots \otimes e_{j_n} \rangle_{E^{\otimes n}} := \prod_{k=1}^{n} \langle e_{i_k}, e_{j_k} \rangle_E = \delta_{i_1 j_1} \ldots \delta_{i_n j_n}$$

and again extending bilinearly to the whole of $E^{\otimes n}$. Finally, we define an inner product on the tensor algebra $T((E))$ by

$$\langle \boldsymbol{a}, \boldsymbol{b} \rangle_{T((E))} = \sum_{n=0}^{\infty} \langle \pi_n(\boldsymbol{a}), \pi_n(\boldsymbol{b}) \rangle_{E^{\otimes n}}$$

for $\boldsymbol{a}, \boldsymbol{b} \in T((E))$ where we choose the inner product on $E^{\otimes 0} = \mathbb{R}$ to be given by usual multiplication. We assume now that the tensor algebra $T((E))$ has been equipped with an inner product in this manner.

Signature kernels

Definition (Signature kernel)

Let $X \in \mathcal{V}^1([a,b],E)$ and $Y \in \mathcal{V}^1([c,d],E)$. The *signature kernel* $k_{X,Y} : [a,b] \times [c,d] \to \mathbb{R}$ is the bivariate function

$$k_{X,Y}(s,t) = \langle S(X)_{a,s}, S(Y)_{c,t} \rangle_{T((E))}$$

for $s \in [a,b]$ and $t \in [c,d]$.

Proposition

Signature kernel is a kernel.

Truncated signature kernel

Definition (Truncated signature kernel)

Let $X \in \mathcal{V}^1([a,b],E)$ and $Y \in \mathcal{V}^1([c,d],E)$. The *truncated signature kernel* $k^{(n)}_{X,Y} : [a,b] \times [c,d] \to \mathbb{R}$ is the bivariate function

$$k^{(n)}_{X,Y}(s,t) = \sum_{k=0}^{n} \langle S^{(k)}(X)_{a,s}, S^{(k)}(Y)_{c,t} \rangle_{E^{\otimes k}}$$

for $s \in [a,b]$ and $t \in [c,d]$.

Proposition

Truncated signature kernel is a kernel.

Computing signature kernels

The signature kernel is impractical without any computation method. Fortunately, as shown in [SCF+21], it satisfies a hyperbolic partial differential equation (PDE) of the Goursat problem class.

Theorem (Signature kernel solves Goursat PDE; [SCF+21, Theorem 2.5])
Let $X \in \mathcal{V}^1([a,b], E)$ and $Y \in \mathcal{V}^1([c,d], E)$. Then the signature kernel $k_{X,Y}$ is the solution of the Goursat PDE

$$\frac{\partial^2 k_{X,Y}}{\partial s \partial t} = \langle \dot{X}_s, \dot{Y}_t \rangle_E k_{X,Y}, \quad k_{X,Y}(a, \cdot) = k_{X,Y}(\cdot, c) = 1$$

where $\dot{X}_s = \frac{dX_p}{dp}\Big|_{p=s}$, $\dot{Y}_t = \frac{dY_q}{dq}\Big|_{q=t}$ are the derivatives of X and Y at time s and t respectively.

Remark

Note that paths of bounded variations are almost everywhere differentiable [SS09, Theorem 3.4]. The existence and uniqueness of the solution to the Goursat PDE follow from [Lee60, Theorem 2, 4].

General Pipeline

Mathematical Institute

$$\text{discrete data} \xrightarrow{(1)} \text{continuous path} \xrightarrow{(2)} \text{signature of path} \xrightarrow{(3)} \text{features of data}$$

- We have already seen how to do the signature transform for (1);
- Now we discuss how to do (2) and (3).

From discrete data to continuous paths

Consider E-valued time series $\{x_{t_i}\}_{i=1}^n$. We create $\mathbb{R}^d$-valued path in two steps.
Step 1: Apply feature map $\varphi : E \to \mathbb{R}^d$ to obtain $\mathbb{R}^d$-valued time series

$$\{X_{t_i}\}_{i=1}^n = \{\varphi(x_{t_i})\}_{i=1}^n .$$

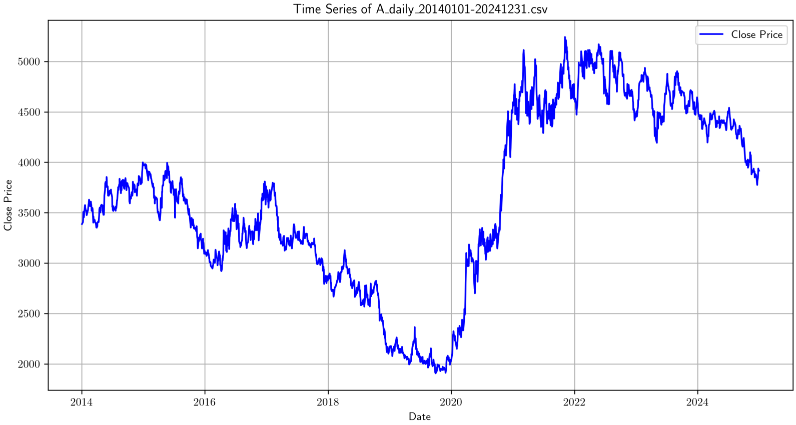
Example

- Canonical embedding: $\mathcal{X} = \mathbb{R}^d, \varphi = \mathrm{id}$;
- Kernel embeddings: $\mathbb{R}^d \rightsquigarrow \mathcal{H}$ where $\mathcal{H}$ is an RKHS.

Sometimes even if $E = \mathbb{R}^d$, applying non-linear φ can be very helpful.

From discrete data to continuous paths (continued)

Step 2: connect points to form continuous path.



Path transformations

Time parametrisation. Recall that signature is independent of time parametrisation. For much data in finance, time parametrisation matters! We need to encode parametrisation before computing signatures.

1D time series. Recall that if $X : [a, b] \to \mathbb{R}$, then $S(X)^{1\ldots1}_{a,b} = \frac{(X_b - X_a)^k}{k!}$. $\Rightarrow$ only the increment is encoded by signature $\Rightarrow$ signature can't be applied naively to 1D time series.

We describe two transformations:

- Time augmentation;
- Lags.

Time augmentation

Definition (Time augmentation)

For $X \in \mathcal{V}^1\left([a,b],\mathbb{R}^d\right)$, define its time-augmented path $\overline{X} \in \mathcal{V}^1\left([a,b],\mathbb{R}^{1+d}\right)$

$$\overline{X}_t = \left(\overline{X}_t^0, \ldots, \overline{X}_t^d\right) = \left(t, X_t^1, \ldots, X_t^d\right).$$

And recall that

Proposition

Let $X \in \mathcal{V}^1([a,b],E)$ with a fixed starting point and at least one coordinate of X is a monotone function. Then $S(X)$ determines X uniquely.

Lags

Omitted for now...

General Pipeline

$$\text{discrete data} \xrightarrow{(1)} \text{continuous path} \xrightarrow{(2)} \text{signature of path} \xrightarrow{(3)} \text{features of data}$$

- We have already seen how to do (1) and (2).
- Now we discuss how to do (3).

Features and learning algorithms

Typically, the first $m \geq 1$ levels $\left\{S^{(k)}(X)_{a,b}\right\}_{k=0}^{m}$ are used (truncated signature), where

$$S^{(k)}(X)_{a,b} = \left\{S(X)_{a,b}^{i_1,\ldots,i_k}\right\}_{1\leq i_1,\ldots,i_k\leq d}.$$

Type of learning algorithm used (linear regression, support vector machines, random forest, etc.) is almost arbitrary.
And with the signature kernel and truncated signature kernel, we can perform a kernelised version of learning algorithms.

Deep signature transforms

By [BKA+19], we can even integrate beyond the standard pipeline with learning algorithms to neural networks.

Technical analysis of financial time series

See `signature.ipynb`

Patric Bonnier, Patrick Kidger, Imanol Pérez Arribas, Cristopher Salvi, and Terry J. Lyons.
Deep signatures.
CoRR, abs/1905.08494, 2019.

Alain Berlinet and Christine Thomas-Agnan.
Reproducing Kernel Hilbert Space in Probability and Statistics.
01 2004.

D. Freedman.
Brownian Motion and Diffusion.
Holden-Day series in probability and statistics. Holden-Day, 1971.

Ben Hambly and Terry Lyons.
Uniqueness for the signature of a path of bounded variation and the reduced path group.
Annals of Mathematics, 171(1):109–167, March 2010.

T.J. Lyons, M.J. Caruana, and T. Lévy.
Differential Equations Driven by Rough Paths: Ecole d'Eté de Probabilités de Saint-Flour XXXIV-2004.
Lecture Notes in Mathematics. Springer Berlin Heidelberg, 2007.

Milton Lees.
The goursat problem.
Journal of the Society for Industrial and Applied Mathematics, 8(3):518–530, 1960.

Terry Lyons and Zhongmin Qian.
System Control and Rough Paths.
Oxford University Press, 12 2002.

Eduardo Paluzo-Hidalgo, Rocio Gonzalez-Diaz, and Miguel A. Gutiérrez-Naranjo.
Two-hidden-layer feed-forward networks are universal approximators: A constructive approach.
Neural Networks, 131:29–36, 2020.

I. Steinwart and A. Christmann.
Support Vector Machines.
Information Science and Statistics. Springer New York, 2008.

Cristopher Salvi, Thomas Cass, James Foster, Terry Lyons, and Weixin Yang.
The signature kernel is the solution of a goursat pde.
SIAM Journal on Mathematics of Data Science, 3(3):873–899, 2021.

E.M. Stein and R. Shakarchi.
Real Analysis: Measure Theory, Integration, and Hilbert Spaces.
Princeton University Press, 2009.