Weighted Signature Kernels and Applications

Xingcheng Xu

Shanghai AI Lab

Seminar on Rough Paths, SPDEs and Related Topics
TU Berlin

December 7, 2023

Joint work with Thomas Cass and Terry Lyons.

Motivation

- High-dimensional sequential learning with signature and kernel methods.
- The selection of an effective kernel is challenging and somewhat task-dependent. When the training data consist of sequential data such as time series, these challenges are magnified.
- The path signature transform has decisive advantages in capturing complex interactions between multivariate data streams.

This Work

- Introduce general signature kernels indexed by a weight function ϕ which generalise the ordinary signature kernel.
- Interpret in many examples as an average of PDE solutions, and show how it can be estimated computationally using suitable quadrature formulae.
- Articulate a novel connection between signature kernels and the notion of the hyperbolic development of a path.
- Extend the analysis to derive closed-form formulae for expected signature kernels involving Brownian motion. Evaluate the use of different general signature kernels as a basis for non-parametric goodness-of-fit tests to Wiener measure on path space.

Weighted Signature Kernels

What is Signature?

- Suppose $x : [a, b] \to V$ is a stream of information/high dimensional (continuous) times series defined on $[a, b]$ taking values in V (e.g. $V = \mathbb{R}^d$).

- The signature $S_{a,t}(x) = \oplus_{n=0}^{\infty} S_{a,t}^{(n)}(x) \in T((V))$ with

$$S_{a,t}^{(n)}(x) = \int \cdots \int_{a<t_1<\cdots<t_n<t} dx_{t_1} \otimes \cdots \otimes dx_{t_n}, \text{ for n} \geq 1.$$

  is the response of the exponential nonlinear system to the stream x:

$$dS_{a,t}(x) = S_{a,t}(x) dx_t, \; S_{a,a}(x) = 1.$$

- The signature provides the infinite graded sequence of statistics which identifies a path in an essentially unique manner.

Signature as Feature

The signature uniquely determines paths.

Theorem (Uniqueness of signature; HL10, Annals of Math.)

Assume $x : [a, b] \to V$ is a continuous path of bounded variation. Then the signature $S_{a,b}(x)$ determines x up to tree-like equivalence. Specially, the path $\bar{x} : t \to (t, x_t)$ is always uniquely determined by its signature $S_{a,b}(\bar{x})$.

The terms of the signature factorially decay.

Theorem (Factorial Decay; LCL04.)

Assume $x : [a, b] \to V$ is a continuous path of bounded variation. Then for every $n \in \mathbb{N}$ and every $a \leq s < t \leq b$, we have $\|S_{s,t}^{(n)}(x)\|_{V^{\otimes n}} \leq \frac{\omega_x(s,t)}{n!}$.

Regression on Signatures

Continuous functions of paths are approximately linear on signatures.

Theorem (Universal Nonlinearity; BDLLS20.)

Let $\mathcal{V}_1([a,b];V)$ denote the space of continuous paths $[a,b] \to V$ of bounded variation. Suppose $\mathcal{K} \subset \mathcal{V}_1([a,b];V)$ is compact and $F : \mathcal{K} \to \mathbb{R}$ is continuous. Then for any $\epsilon > 0$ there exists a truncation level $N \in \mathbb{N}$ such that for every $x \in \mathcal{K}$ we have

$$\left| F(x) - \sum_{i=0}^{N} \sum_{\mathcal{I} \in \{1,\cdots,d\}^i} \alpha_{\mathcal{I}} S_{a,b}^{\mathcal{I}}(x) \right| < \epsilon.$$

Signature Kernel Methods

- Overcomes bottlenecks of the signature features:
  - computational complexity
  - expressiveness

- Kernelization allows to simultaneously consider a rich set of non-linearities while avoiding the combinatorial explosion in the computation of signatures.

- Theoretical guarantee of universality (the ability to approximate non-linear functions) and characteristicness (the ability to characterize probability measures) from stochastic analysis and the properties of the classical signature.

- Leverages modular tools from kernel learning: Kernel methods are well-established tools in machine learning which are fundamental for classification, nonlinear regression and outlier detection involving small or moderate-sized data sets.

Weighted Signature Kernels

- Define the general/weighted signature kernel of two paths x and y by

$$K^{\phi}_{s,t}(x,y) := \langle S_{a,s}(x), S_{a,t}(y)\rangle_{\phi} := \sum_{k=0}^{\infty} \phi(k) \langle S^{(k)}_{a,s}(x), S^{(k)}_{a,t}(y)\rangle_{V^{\otimes k}}$$

- The signature kernel with $\phi \equiv 1$ solves the partial differential equation ([SCFLY20])

$$\frac{\partial^2 K_{s,t}(x,y)}{\partial s \partial t} = K_{s,t}(x,y)\langle \dot{x}_s, \dot{y}_t\rangle$$

  with boundary conditions $K_{a,\cdot}(x,y) \equiv K_{\cdot,a}(x,y) \equiv 1$.

- Methods that allow for the efficient computation of general signature kernels with a broad class of different weightings ϕ are introduced by Cass, Lyons, and Xu (2023+). WSKs can be interpreted as an average of PDE solutions, and could be computed via suitable quadrature rules.

Kernel Learning: Signature Kernel Classification/Regression

- Kernelization in a nutshell: Given a set $\mathcal{K}$, we want a feature map $\psi : \mathcal{K} \to \mathcal{H}$ that injects elements of $\mathcal{K}$ into a Hilbert space $\mathcal{H}$ such that ψ provides sufficient non-linearities so that linear functionals of ψ, $x \mapsto \langle w, \psi(x)\rangle_{\mathcal{H}}$ are expressive enough to approximate a rich set of functions.

- Signature feature map $\psi = S$: continuous paths space (times series data space) $x \in \mathcal{K} \subset \mathcal{V}_1([a,b];V)$ to the signature space as a subspace of a Hilbert space $\mathcal{H} = T_\phi(V)$.

- Gram Matrix $G = \Big(K^\phi(x_i, x_j)\Big)_{x_i \in \mathcal{D}}$ where $\mathcal{D}$ is a finite set of data given. Computation cost $|\mathcal{D}|^2$.

- Low-rank matrix approximations $\tilde{G} = CW^{-1}C^T$ where $C = \Big(K^\phi(x_i, x_j)\Big)_{x_i \in \mathcal{D}, x_j \in \mathcal{J}}$, $W = \Big(K^\phi(x_i, x_j)\Big)_{x_i, x_j \in \mathcal{J}}$, and $\mathcal{J}$ is of size r, sampled uniformly at random from $\mathcal{D}$ with $r \ll |\mathcal{D}|$. Computation cost $O(|\mathcal{D}|)$.

Weighted Signature Kernels by Randomisation

- Condition on ϕ: The function $\phi : \mathbb{N} \cup \{0\} \to \mathbb{R}_+$ is such that the series $\sum_{k \in \mathbb{N}} C^k \phi(k) (k!)^{-2}$ is summable for every $C > 0$.

- The Hamburger moment problem: find a probability measure μ on $\mathbb{R}$ such that

$$\phi(k) = \int \lambda^k d\mu(\lambda) \text{ for all } k \in \mathbb{N} \cup \{0\}.$$

Then, under some conditions on μ, able to justify the following identity

$$K^{\phi}_{s,t}(x, y) = \sum_{k=0}^{\infty} \int \lambda^k \left\langle S^k_{a,s}(x), S^k_{a,t}(y) \right\rangle_k d\mu(\lambda) = \int K_{s,t}(\lambda x, y) \, d\mu(\lambda).$$

Weighted Signature Kernels by Randomisation: Examples

- Factorial-WSK: The ϕ-signature kernel under $\phi(k) = \left(\frac{k}{2}\right)!$ satisfies

$$K^{\phi}_{s,t}(x,y) = \mathbb{E}_{\pi}\left[K_{s,t}\left(\pi^{1/2}x, y\right)\right] = \mathbb{E}_{\pi}\left[K_{s,t}\left(x, \pi^{1/2}y\right)\right],$$

where $\pi \sim \mathrm{Exp}(1)$ is an exponentially distributed random variable with intensity 1.

- Beta-WSK: The ϕ-signature kernel under $\phi(k) = \frac{\Gamma(m+1)\Gamma(k+1)}{\Gamma(k+m+1)}$ satisfies

$$K^{\phi}_{s,t}(x,y) = \mathbb{E}_{\pi}\left[K_{s,t}\left(\pi x, y\right)\right] = \mathbb{E}_{\pi}\left[K_{s,t}\left(x, \pi y\right)\right],$$

where $\pi \sim B(1,m)$ is a Beta-distributed random variable.

Weighted Signature Kernels by Integral Transforms

- The trigonometric moment problem: find a measure μ on $[0, 2\pi]$ such that

$$\phi(k) = \int_0^{2\pi} e^{ik\theta} d\mu(\theta) \text{ for } k \in \mathbb{Z}.$$

- The Integral Transforms: A class of integral transforms having the form

$$\phi(u) = \int_C r(u, z)\, d\mu(z), \text{ with } r(u, z) = g(z)^{\alpha u} \in \mathbb{C}$$

where $\alpha \in \mathbb{R}$, μ a finite signed Borel measure. This class includes the Fourier-, Laplace- and Mellin-Stieltjes transforms. Justify the following identity:

$$K_{s,t}^{\phi}(x, y) = \int_C K_{s,t}\left(g(z)^{\alpha} x, y\right) \mu(dz).$$

Weighted Signature Kernels by Integral Transforms: Examples

- Fourier-Stieltjes transform: $C=\mathbb{R},\ g(z)=e^{-2\pi iz},\ \alpha=1$, i.e.

$$\phi(u)=\hat{\mu}(u):=\int_{\mathbb{R}} e^{-2\pi iuz}\mu(dz);$$

- Laplace-Stieltjes transform: $C=(0,\infty),\ g(z)=e^{-z},\ \alpha=1$, i.e.

$$\phi(u)=\tilde{\mu}(u):=\int_0^\infty e^{-uz}\mu(dz);$$

- Mellin-Stieltjes transform: $C=(0,\infty),\ g(z)=z,\ \alpha=1$, i.e.

$$\phi(u)=\mu_{\mathrm{Mel}}(u+1)=\int_0^\infty z^u\mu(dz),\ \operatorname{Re} u>-1.$$

Computing Weighted Signature Kernels

- Gaussian Quadrature Rule (e.g. [SM03]):

$$\int K_{s,t}(zx, y)\, w(z)\, dz \approx \sum_{k=0}^{n} w_k K_{s,t}(z_k x, y)\,.$$

- For a general weight function w, suppose that $\mathcal{P} = \{p_n : n \in \mathbb{N} \cup \{0\}\}$ is a system of orthogonal polynomials w.r.t. the weight function w, that is $\deg(p_n) = n$ and $\langle p_n, p_m \rangle_w = \int p_m p_n w dz = 0$ for $n \neq m$.

- The quadrature points z_k, $k = 0, 1, ..., n$ are the zeros of the polynomial p_{n+1}, the corresponding quadrature weights are

$$w_k := \int w(z) \prod_{i=0, i \neq k}^{n} \left(\frac{z - z_i}{z_k - z_i} \right)^2 dz.$$

Expected Weighted Signature Kernels

Expected Weighted Signature Kernels

- Consider how ϕ-signature kernels can be combined with the notion of expected signatures to compare the laws of two stochastic processes.

- Study the expected weighted signature kernels

$$K^{\phi}_{s,t}(\mathcal{W},\mu) := \langle \mathbb{E}_{B\sim\mathcal{W}}[S_{0,s}(\circ B)], \mathbb{E}_{X\sim\mu}[S_{0,t}(X)]\rangle_{\phi},$$

  where $\mathcal{W}$ is Wiener measure, and the measure μ will typically discrete and supported on bounded variation paths. $\mathbb{E}[S_{0,s}(\circ B)]$ denotes the expected Stratonovich signature for Brownian motion B.

- An initial step: x is a fixed deterministic continuous path of bounded variation.

$$K^{\phi}_{s,t}(\mathcal{W},x) := \langle \mathbb{E}_{B\sim\mathcal{W}}[S_{0,s}(\circ B)], S_{0,t}(x)\rangle_{\phi},$$

Expected Weighted Signature Kernels: Closed-form Solution

Theorem (Cass, Lyons, X., 2023+)

Let $\phi(k) = \left(\frac{k}{2}\right)!$ for $k \in \mathbb{N} \cup \{0\}$. Suppose $x : [0,1] \to V$ is any continuous path of bounded variation, it holds that

$$K^{\phi}_{s,t}(\mathcal{W}, x) := \langle \mathbb{E}_{B\sim\mathcal{W}}\left[S_{0,s}(\circ B)\right], S_{0,t}(x)\rangle_{\phi} = \cosh\left(\rho_{\sqrt{s/2}x}(t)\right),$$

for all $t \in [0,1]$. In this notation, $\rho_{\lambda x}(t) := \mathrm{d}_{\mathbb{H}^{\mathrm{d}}}(o, \sigma_{\lambda x}(t))$ is the distance between the hyperbolic development $\sigma_{\lambda x}(t)$ of the path $\lambda x(\cdot)$ from $T_o\mathbb{H}^d$ onto the d-dimensional hyperbolic space. $\mathbb{H}^d$ started at the base point $o = (0, 0, ..., 1) \in \mathbb{H}^d$, and $\mathrm{d}_{\mathbb{H}^{\mathrm{d}}} : \mathbb{H}^d \times \mathbb{H}^d \to [0, \infty)$ is the Riemannian distance on $\mathbb{H}^d$.

Expected Weighted Signature Kernels

- The representation:

$$\cosh \rho_{\lambda x}(t) = 1 + \sum_{n=1}^{\infty} \lambda^{2n} \int_{0<t_1<\cdots<t_{2n}<t} \langle dx_{t_1}, dx_{t_2} \rangle \cdots \langle dx_{t_{2n-1}}, dx_{t_{2n}} \rangle.$$

- For $x_t \in \mathbb{R}^d$, let x^λ be the rescaling of x by λ. The ODE

$$d\Gamma^{\lambda x}(t) = F(dx_t^\lambda)\Gamma^{\lambda x}(t), \;\; t \in [0,1], \;\; \text{with } \Gamma^{\lambda x}(0) = I_{d+1} \tag{1}$$

has a unique solution, and furthermore the last entry $\Gamma^{\lambda x}_{d+1,d+1}(t) = \cosh \rho_{\lambda x}(t)$.

- If x is a piecewise linear path defined by the concatenation $x_{v_1} * x_{v_2} * \cdots * x_{v_n}$, i.e. x is such that $x'_{v_i}(t) = v_i \in \mathbb{R}^d$ for $t \in (t_{i-1}, t_i)$. Then the solution to (1) is given explicitly by the matrix product

$$\Gamma^{\lambda x}(1) = A(v_n, \Delta_n, \lambda)\, A(v_{n-1}, \Delta_{n-1}, \lambda) \cdots A(v_1, \Delta_1, \lambda), \tag{2}$$

where $A(v, \Delta, \lambda) := I_{d+1} + \sinh(\lambda |v| \Delta) M + (\cosh(\lambda |v| \Delta) - 1) M^2$, $\Delta_i = t_i - t_{i-1}$, and $M = F(\tilde{v})$ with $\tilde{v} = v/|v|$.

Computing Expected Weighted Signature Kernels: Factorial

- In contrast to the earlier case on the ϕ-signature kernel of two paths, we need only solve an ODE to calculate $\langle \mathbb{E}\left[S\left(\circ B\right)\right], S\left(x\right)\rangle_{\phi}$ rather than a PDE.

- For general path x, the ODE is known, and is determined by given linear vector fields. Any ODE solver such as Runge-Kutta could in principle be used to obtain numerical solutions.

- For piecewise linear path x, the exact solution is given explicitly as a product of matrices.

The Original Kernel of Expected Signatures

Theorem (Cass, Lyons and X., 2023+)

Let $\phi \equiv 1$. Then

$$\left|\left|\mathbb{E}\left[S\left(\circ B\right)_{0,s}\right]\right|\right|_{\phi}^{2} = \frac{1}{2\pi i}\oint_{C} z^{-1}e^{z+s^2 d/(4z)}dz$$

where the contour C is the unit circle in $\mathbb{C}$ traversed anticlockwise. Furthermore,

$$K_{s,t}^{\phi}\left(\mathcal{W}, x\right) = \frac{1}{2\pi i}\oint_{C} z^{-1}e^{z}\Gamma_{d+1,d+1}^{c_s(z)x}(t)dz$$

where $c_s(z) = \sqrt{s/2z} \in \mathbb{C}$ and $\Gamma_{d+1,d+1}^{c_s(z)x}(t)$ is the last entry of the solution to ODE (1).

Computing Expected Weighted Signature Kernels: Original

The efficient approximation of the Hankel-type contour integrals of the form

$$I = \frac{1}{2\pi i} \oint_H e^z f(z) dz = \frac{1}{2\pi i} \int_{-\infty}^{+\infty} e^{\varphi(\theta)} f(\varphi(\theta)) \varphi'(\theta) d\theta.$$

The approach is to approximate by

$$I_N = -iN^{-1} \sum_{k=1}^{N} e^{z_k} f(z_k) w_k = -\sum_{k=1}^{N} c_k f(z_k) \tag{3}$$

on the finite interval $[-\pi, \pi]$ with N points which are regularly spaced on the interval and $z_k = \varphi(\theta_k)$, $w_k = \varphi'(\theta_k)$ and $c_k = iN^{-1} e^{z_k} w_k$. Three classes of contours (see e.g. [TWS06]):

- Parabolic contours $\varphi(\theta) = N(0.1309 - 0.1194\theta^2 + 0.2500i\theta)$
- Hyperbolic contours $\varphi(\theta) = 2.246N(1 - \sin(1.1721 - 0.3443i\theta))$
- Cotangent contours $\varphi(\theta) = N(0.5017\theta \cot(0.6407\theta) - 0.6122 + 0.2645i\theta)$

The Beta-Weighted Kernel of Expected Signatures and Computation

Let $\phi(k) = \frac{\Gamma(m+1)\Gamma(k+1)}{\Gamma(k+m+1)}$. Then

$$K^{\phi}_{s,t}(\mathcal{W}, x) = \frac{\Gamma(m+1)}{2\pi i} \underbrace{\oint_C z^{-(m+1)} e^z}_{(3)\ \textit{Contour approximation}} \left[\underbrace{\frac{1}{\sqrt{2\pi}} \int_{-\infty}^{+\infty} \overbrace{\Gamma^{c_s(\rho,z)x}_{d+1,d+1}(t)}^{(1)\ \textit{explicit solution}} \ e^{-\frac{\rho^2}{2}} d\rho}_{(2)\ \textit{Gaussian quadrature}} \right] dz,$$

where $c_s(\rho, z) = z^{-1}\rho\sqrt{s} \in \mathbb{C}$.

More general weighted expected signature kernels, see Theorem 6.5 of [CLX23+].

Characterise Laws of Stochastic Processes

- a measure of similarity of two laws on path space through the quantity

$$d_{\phi}(\mathcal{W},\mu) := \left|\left|\mathbb{E}^{\mathcal{W}}[S(X)] - \mathbb{E}^{\mu}[S(X)]\right|\right|_{\phi},$$

  which is seen to be a maximum mean discrepancy (MMD) distance between $\mathcal{W}$ and μ.

- a measure of alignment of the two expected signatures of $\mathcal{W}$ and μ given by

$$\cos\angle_{\phi}(\mathcal{W},\mu) := \frac{\left\langle \mathbb{E}^{\mathcal{W}}[S(X)], \mathbb{E}^{\mu}[S(X)]\right\rangle_{\phi}}{||\mathbb{E}^{\mathcal{W}}[S(X)]||_{\phi}\,||\mathbb{E}^{\mu}[S(X)]||_{\phi}},$$

  which can be interpreted as an analogue of the Pearson correlation coefficient for measures on path space.

- Design goodness-of-fit tests: The ratio $\frac{d_{\phi}(\mathcal{W},\mu^{*})}{d_{\phi}(\mathcal{W},\mu)} < \alpha < 1$. By an appropriate selection of the threshold α, one might decide whether μ resembles Wiener measure or not.

Applications of Weighted Signature Kernels

Multivariate Time Series Classification

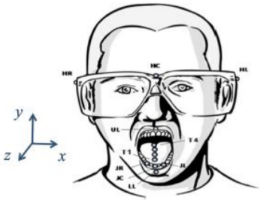
- We employ our weighted signature kernels to several multivariate time series classification problems on the UEA datasets available at `https://timeseriesclassification.com/`.

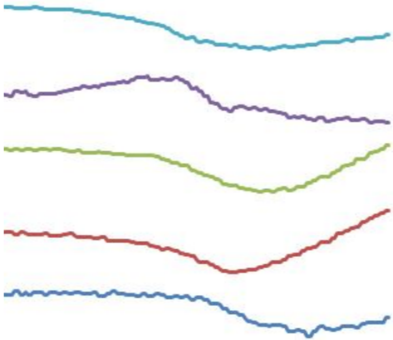
- We use the support vector classifier (SVC), and compare the performance under the same SVC settings (time series pre-processing, hyperparameter selection, etc.) with the original signature kernel, factorially-weighted signature kernel and Beta-weighted signature kernel (here, we use $m = 1$ for the Beta weights). The only difference is the kernel we use in the SVC models.

- In Table 1, we show the performance of SVC with different kernels. As the results show, the test accuracy of SVC with the factorially- or Beta-weighted signature kernel are better than that of the original signature kernel for most of the datasets.

Multivariate Time Series: ArticularyWordRecognition

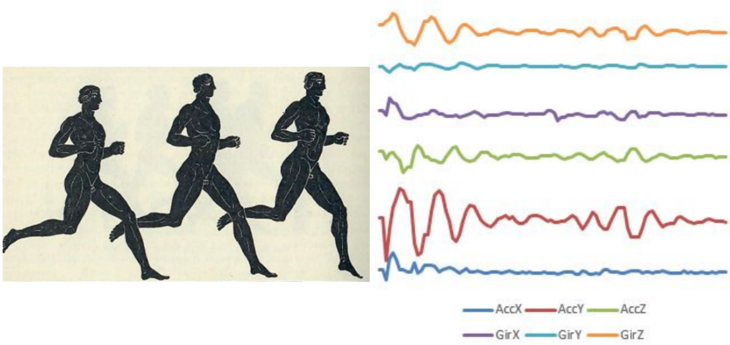
An Electromagnetic Articulograph (EMA) is an apparatus used to measure the movement of the tongue and lips during speech. The motion tracking using EMA is registered by attaching small sensors on the surface of the articulators. Collected from multiple native English speakers producing 25 words.





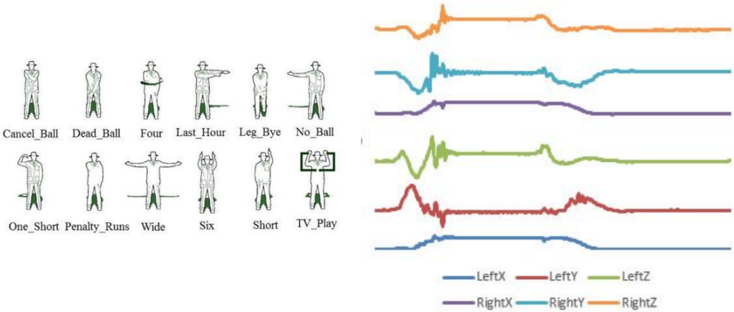
Multivariate Time Series: BasicMotions

Performed four activities whilst wearing a smart watch. The watch collects 3D accelerometer and a 3D gyroscope data. Four classes: standing, walking, running and playing badminton.



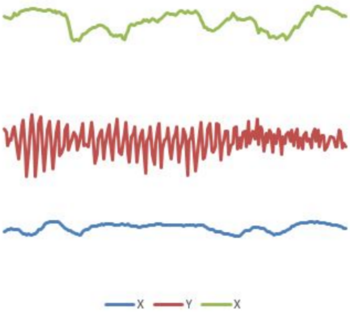
Multivariate Time Series: Cricket

Cricket requires an umpire to signal different events in the game to a distant scorer. The signals are communicated with motions of the hands.



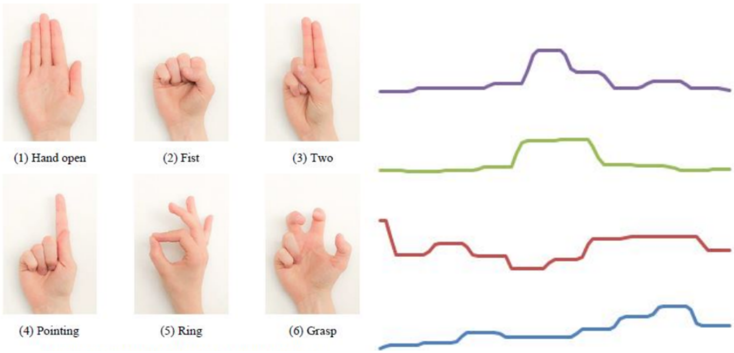
Multivariate Time Series: Epilepsy

Data was generated with healthy participants simulating four different activities: WALKING, RUNNING, SAWING, SEIZURE MIMICKING.



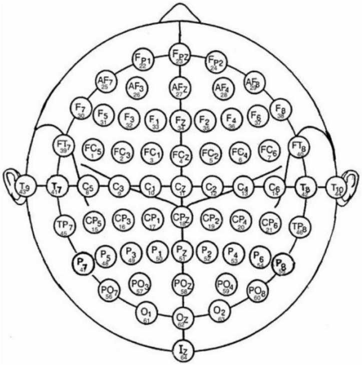
Multivariate Time Series: ERing

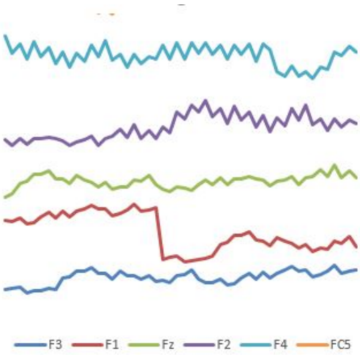
Hand and finger gestures:



Multivariate Time Series: FingerMovements

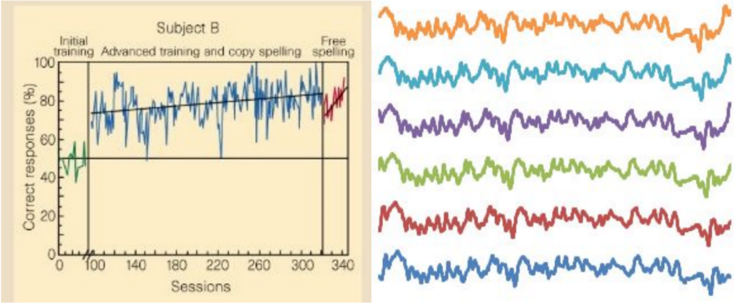
This dataset was recorded from a normal subject during a no-feedback session. The subject sat in a normal chair, fingers in the standard typing position at the computer keyboard. Brain-Computer Interface (BCI) detects upcoming finger movements and predicts their laterality.





Multivariate Time Series: SelfRegulationSCP1

Self-regulation of Slow Cortical Potentials. The data were taken from a healthy subject. The subject was asked to move a cursor up and down on a computer screen, while his cortical potentials were taken.



Multivariate Time Series Classification

Table: Test set classification accuracy (in %) on UEA multivariate time series datasets

Datasets	**Without add-time operation**			**With add-time operation**		
	Original	*Factorial*	*Beta(1)*	*Original*	*Factorial*	*Beta(1)*
ArticularyWordRecognition	81.3	80	79.7	94.3	92.3	93.3
BasicMotions	87.5	**90**	100	97.5	97.5	95
Cricket	62.5	58.3	**75**	84.7	81.9	83.3
Epilepsy	90.6	88.4	90.6	92	92	93.5
ERing	75.6	**78.1**	74.4	80	87.4	**86.3**
FingerMovements	44	**47**	**45**	49	**51**	57
Libras	48.9	**50**	**57.2**	66.1	65	68.3
NATOPS	73.9	73.3	**78.9**	90.6	88.3	91.7
RacketSports	69.1	68.4	67.1	78.9	78.3	79.6
SelfRegulationSCP1	50.5	50.5	**51.5**	70.3	71	68.6
UWaveGestureLibrary	74.1	73.4	76.9	71.9	70.3	70.6

Discrete Measures on Brownian Paths

- We proved the existence of a unique optimal probability measure μ^* supported on $X = \{x_1, \cdots, x_n\}$ such that

$$\mu^* = \arg\min_{\mu \in C_n} \left|\left|\mathbb{E}\left[S(\circ B)_{0,1}\right] - \mathbb{E}^{\mu}\left[S(X)_{0,1}\right]\right|\right|_{\phi}^{2}.$$

- Brownian Paths: We randomly sample n i.i.d. Brownian motion paths in $\mathbb{R}^d$. Each path sampled over the time interval $[0, 1]$, on an equally-spaced partition $0 = t_1 < t_2 < \cdots < t_m = 1$. Denote the resulting finite set piecewise linearly interpolated Brownian sample paths as

$$\mathcal{S}(n, m, d) = \{B_i\}_{i=1}^{n} \text{ with } B_i = \{B_i(t_j) \in \mathbb{R}^d\}_{j=1}^{m}.$$

Discrete Measures on Brownian Paths

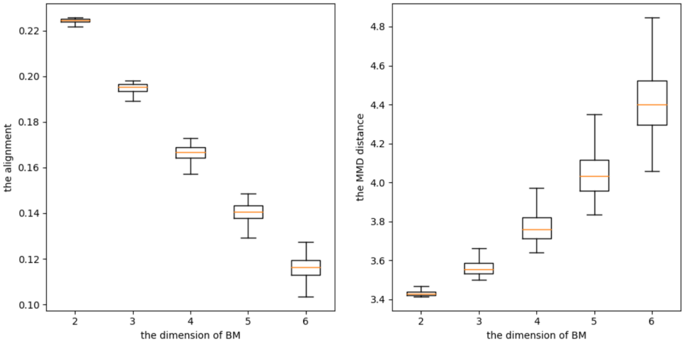


Figure: Boxplots of the factorially-weighted signature kernel. (a) The left panel shows the distribution of the values of the alignment $\cos \angle_\phi(\mu^*, \mathcal{W})$ of the optimal measure and the Wiener measure across 400 independent experiments. (b) The right panel shows the same for the MMD distance $d_\phi(\mu^*, \mathcal{W})$.

Discrete Measures on Brownian Paths

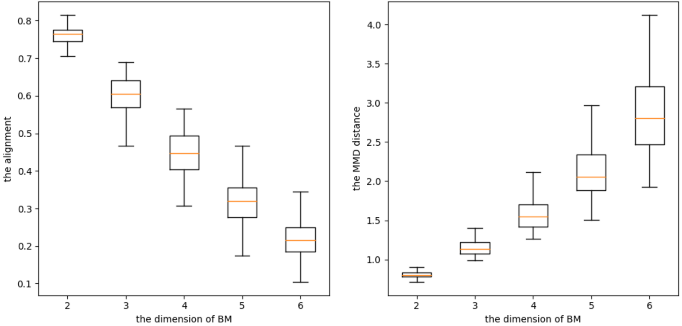


Figure: The optimal measure under the original signature kernel

Applications in Signal Processing

- The alignment and the similarity defined by the ϕ-signature kernel give us a way of determining how large a given discrete measure is different to the Wiener measure. A natural application of these methods in signal processing is to mitigate/detect the (additive) contamination of white noise under different types of perturbation.

- The observed visibility = the astrophysical sky signals + thermal noise + RFI.

- Narrow-band RFI measure across antennas: The received signals are linear superpositions of independent Brownian motions with a single-frequency sinusoidal wave of a fixed amplitude.

$$X_i^{(j)}(t) = B_i^{(j)}(t) + \epsilon \sin(2\pi\nu t - \phi_i^{(j)}),\ j = 1, 2, \cdots, d$$

- Short duration high energy bursts: Brownian signal undergoes a perturbation at a uniformly distributed random time.

$$X_i^{(j)}(t) = B_i^{(j)}(t) + \epsilon\sqrt{(t - U_i)^+},\ j = 1, 2, \cdots, d$$

Narrow-band RFI Contamination

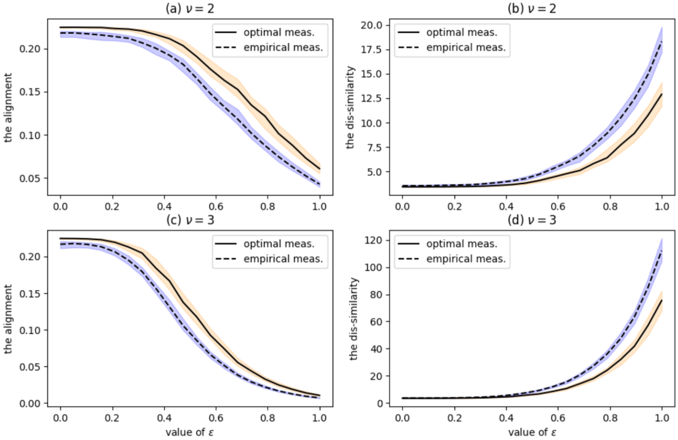


Figure: The case for the factorially-weighted signature kernel.

Narrow-band RFI Contamination

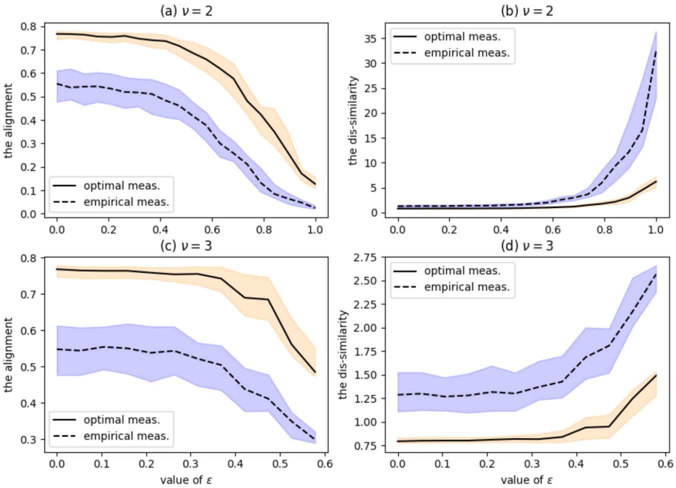


Figure: The case for the original signature kernel.

Short Duration High Energy Bursts

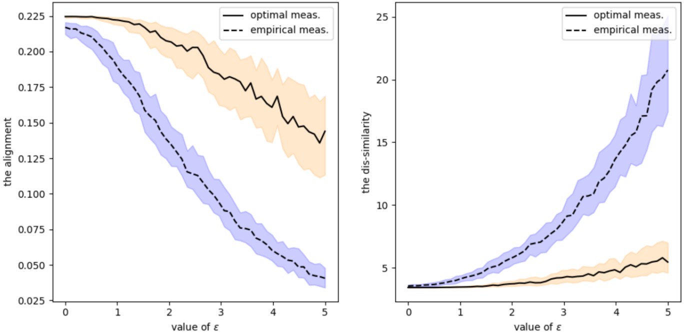


Figure: The case for the factorially-weighted signature kernel.

Short Duration High Energy Bursts

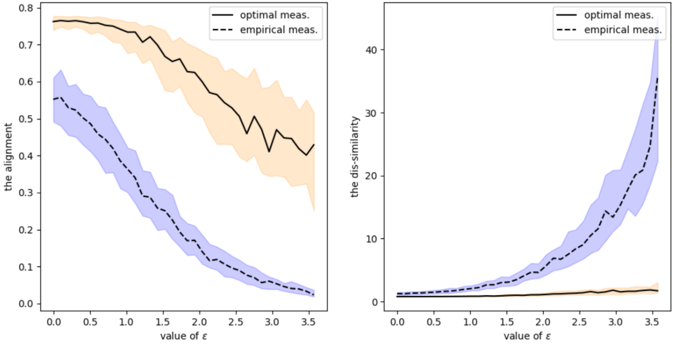
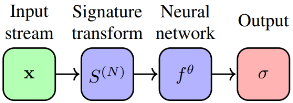


Figure: The case for the original signature kernel.

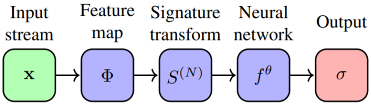
Coupling with Deep Neural Networks

Coupling with Deep Neural Networks

- Deep signature models ([LM2023])



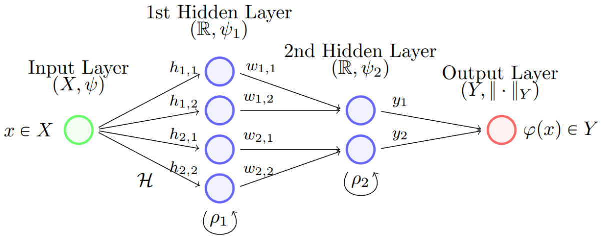
(a) Neural-signature model.
Trainable parameters: θ.



(b) Neural-signature-augment model.
Trainable parameters: θ.

- Neural signature kernel (as limits of controlled ResNets, [CLS2023]).

- Functional input neural networks ([CST2023])



Thanks!