An overview of 2-d signatures methods

Samy Tindel

Purdue University

Mini-Courses: Rough Paths, Signatures,
and their applications in Machine Learning
BI – Oslo – November 2023

Joint works with

- Guang Lin and Sheng Zhang
- Joscha Diehl, Kurusch Ebrahimi-Fard and Fabian Harang

Signatures for images (ongoing CAS project)

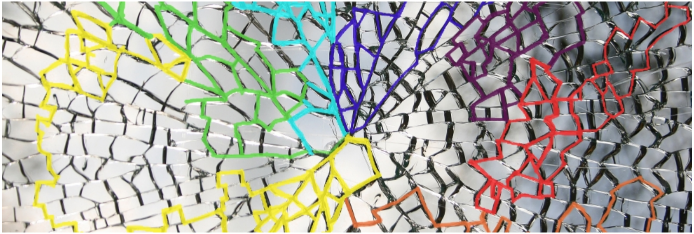


Figure: Fragmented glass

Outline

Outline

1. 1-d signatures as features
   - A motivation for feature extraction
   - Basic properties of 1-d signatures
2. Introducing 2d-signatures
   - Basic properties of 2d-signatures
   - Numerical experiment
3. Extended signature in the plane
   - A very simple signature
   - The extended signature
4. A PDE perspective on 2d-signatures

Outline

A basic classification task

Data:

- Points $\mathbf{X} = \{\mathbf{x}_i; i = 1, \ldots, n\}$ with $\mathbf{x}_i \in \mathbb{R}^d$
- Labels $\{y_i; i = 1, \ldots, n\}$ with $y_i \in \{0, 1\}$
- When labels are known, the learning is supervised

Aim:

- Find a proper separation between labels 0 and labels 1

Linear separation

Separation using hyperplanes:

- We use a classification $\hat{y} = \text{sign}(\mathbf{v} \cdot \mathbf{x})$
- $\mathbf{v}$ optimized
  $\hookrightarrow$ According to our data:

$$\mathbf{v} = \min_{\mathbf{w} \in \mathbb{R}^d} \sum_{i=1}^{n} \left\| \text{sign}\left(\mathbf{w} \cdot \mathbf{x}_i\right) - y_i \right\|^2$$

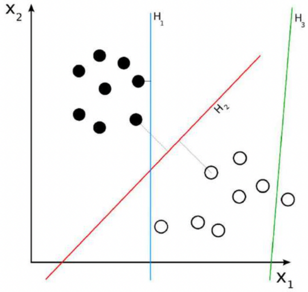


Figure: Separation of 2 subgroups according to H_1, H_2, H_3

Separation using neural networks

Definition of the multilayer neural network:

- Recursion $\mathbf{x}^{k+1} = S(\mathbf{w}^k \mathbf{x}^k + \mathbf{d}^k)$ for $k = 0, \ldots, n_{\text{layer}}$
- $\mathbf{w}^k$ matrix-valued, $\mathbf{d}^k$ vector-valued
- S defined componentwise by σ below
- $\mathbf{w}^k$ and $\mathbf{d}^k$ to be optimized

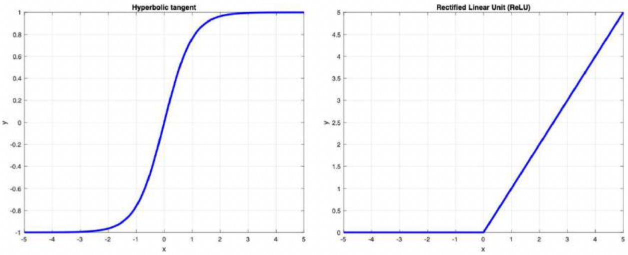
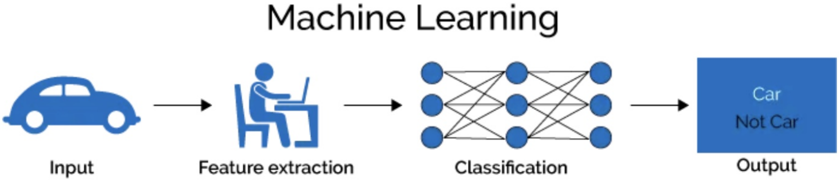


Figure: Sigmoid $\sigma(x) = \frac{2}{\pi} \tanh(x)$ and ReLU $\sigma(x) = \max\{x, 0\}$

Feature extraction problem

Objection to previous situation:

- In classification problem, $\mathbf{X}$ was supposed to be fixed
- If $\mathbf{X}$ is high-dimensional, this might be a problem
  $\hookrightarrow$ feature extraction needed



Feature extraction (2)

Main goal of feature extraction: Given input $\mathbf{X}$

- **extract information** to be fed as additional input to the machine learning algorithm
- Be sparse in the additional input:
  - more information given
    $\Longrightarrow$ more computationally expensive learning task
  - Unnecessary noise added with more information

Wishlist for good features:

- Computationally efficient
- Accurate description of the data distribution
- If possible: interpretable.

References

Bensoussan, A., Zhou, X. et al (2020).
Machine learning and control theory.
arXiv preprint arXiv:2006.05604.

Outline

Notation for 1-d parameter signature

Signal: We first consider a $\mathbb{R}^d$-valued

$$x = \left\{x_s^i;\, s \in [0, T],\, i = 1 \ldots, d\right\}$$

Second order 1-d signature: For our signal x

$$\begin{aligned} \mathbf{x}_{st}^{\mathbf{1},i_1} &= x_t^{i_1} - x_s^{i_1} = \int_{s<r_1<t} \mathrm{d}x_{r_1}^{i_1} \\ \mathbf{x}_{st}^{\mathbf{2},i_1,i_2} &= \int_{s<r_1<r_2<t} \mathrm{d}x_{r_1}^{i_1}\, \mathrm{d}x_{r_2}^{i_2}. \end{aligned}$$

Signature as a feature

Claim:

- **Signature is a good feature for the signal** x

Some properties to review:

1. Natural feature
2. Algebraic properties
3. Analytic properties
4. Accurate description of the path
5. Computationally efficient
6. Possibility of interpretation

The founding fathers

Kuo Tsai Chen

Kyosi Itô

Terry Lyons

Brief history survey:

- K. T. Chen, 50s: Structure of iterated-integrals signatures
- K. Itô, 50s: Itô stochastic calculus
- T. Lyons, 90s: Theory of rough paths.

Signatures are natural features (1)

Notation: For a function $z : [0, T] \to \mathbb{R}^d$,

$$\delta z_{st} = z_t - z_s$$

Change of variable formula: Consider smooth functions

- $x : [0, T] \to \mathbb{R}^d$
- $f : \mathbb{R}^d \to \mathbb{R}$

Then we have

$$\delta f(x)_{st} = \int_{s \leq r_1 \leq t} \partial_{i_1} f(x_{r_1}) \, \mathrm{d}x_{r_1}^{i_1}$$

Signatures are natural features (2)

Iterating change of variable: Write

$$
\begin{aligned}
\delta f(x)_{st} &= \int_{s\le r_1\le t} \partial_{i_1} f(x_{r_1})\,\mathrm{d}x_{r_1}^{i_1} \\
&= \partial_{i_1} f(x_s) \int_{s\le r_1\le t} \mathrm{d}x_{r_1}^{i_1} + \int_{s\le r_1\le t} \delta\left[\partial_{i_1} f(x)\right]_{sr_1}\, \mathrm{d}x_{r_1}^{i_1} \\
&= \partial_{i_1} f(x_s) \int_{s\le r_1\le t} \mathrm{d}x_{r_1}^{i_1} + \int_{s\le r_2\le r_1\le t} \partial_{i_1 i_2} f(x_{r_2})\,\mathrm{d}x_{r_2}^{i_2}\mathrm{d}x_{r_1}^{i_1} \\
&= \partial_{i_1} f(x_s) \int_{s\le r_1\le t} \mathrm{d}x_{r_1}^{i_1} + \partial_{i_1 i_2} f(x_s) \int_{s\le r_2\le r_1\le t} \mathrm{d}x_{r_2}^{i_2}\mathrm{d}x_{r_1}^{i_1} \\
&\quad + \int_{s\le r_2\le r_1\le t} \delta\left[\partial_{i_1 i_2} f(x)\right]_{sr_2}\, \mathrm{d}x_{r_2}^{i_2}\mathrm{d}x_{r_1}^{i_1}
\end{aligned}
$$

Signatures are natural features (3)

Approximations in change of variable: For a smooth enough f,

$$f(x_t) - f(x_s) \simeq \sum_{i_1} \partial_{i_1} f(x_s)\, \mathbf{x}_{st}^{\mathbf{1},i_1} + \sum_{i_1,i_2} \partial^2_{i_1,i_2} f(x_s)\, \mathbf{x}_{st}^{\mathbf{2},i_1,i_2}$$

This is one of the reasons why signatures are natural features

Full signature: For higher order approximations one can recur to

$$[S(x)]_{st} = 1 + \sum_{n=1}^{\infty} \int_{s<r_1<r_2<\cdots<r_n<t} \mathrm{d}x_{r_1} \otimes \mathrm{d}x_{r_2} \otimes \cdots \otimes \mathrm{d}x_{r_n} \qquad (1)$$

Next aim: Give a proper meaning to (1)

Where do signatures live (1)?

Words: Define a set of words $\mathcal{W} = \cup_{n\geq 0}\mathcal{W}_n$ with

$$\mathcal{W}_n = \Big\{ w = (i_1, \ldots, i_n) \; ; \; n \geq 0 \text{ and } i_j \in \{1, \ldots, d\} \text{ for all } j = 1, \ldots, n \Big\}$$

Notation for simplexes: For $a < b$ we set

$$\Delta^n_{a,b} = \big\{ r \in [0, T]^n \, ; \, a \leq r^1 \leq \cdots \leq r^n \leq b \big\}.$$

Evaluation on words: For $w = (i_1, \ldots, i_n) \in \mathcal{W}$ we set

$$\langle S_{st}(x), \, w \rangle = \int_{\Delta^n_{[s,t]}} \mathrm{d}x^{i_1}_{r^1} \cdots \mathrm{d}x^{i_n}_{r^n} \, .$$

Where do signatures live (2)?

Tensor algebra: We set

$$\mathcal{T}(\mathbb{R}^d) = \bigoplus_{n=0}^{\infty} (\mathbb{R}^d)^{\otimes n},$$

Canonical basis for $\mathbb{R}^d$:

$$(e_1, \ldots, e_d)$$

Canonical basis for $\mathcal{T}(\mathbb{R}^d)$:

$$\{e_w = e_{i_1} \otimes \cdots \otimes e_{i_n} \; ; \; w = (i_1, \ldots, i_n) \in \mathcal{W}\}.$$

Where do signatures live (3)?

Signature as evaluation: We see $S_{st}(x)$ as a linear map,

$$S_{st}(x) : \mathcal{T}(\mathbb{R}^d) \longrightarrow \mathbb{R}, \qquad e_w \longmapsto \langle S(x),\, w \rangle_{st}$$

Tensor series: We also write

$$S(x) = \sum_{w \in \mathcal{W}} \langle S(x), w \rangle \; e_w, \quad \text{and} \quad S_{st}(x) \in \mathcal{T}((\mathbb{R}^d))$$

Chen's algebraic relation (1)

Notation for products: For $g, h \in \mathcal{T}((\mathbb{R}^d))$,

$$[g \otimes h]^n = \sum_{k=0}^{n} g^{n-k} \otimes h^k$$

Notation for simplexes: For $a < b$ we set

$$\Delta_{a,b}^n = \left\{ r \in [0, T]^n ; \, a \leq r^1 \leq \cdots \leq r^n \leq b \right\}.$$

Chen's algebraic relation (2)

Theorem 1.

Consider

- $x : [0, T] \to \mathbb{R}^d$ differentiable path
- $S(x)$ its signature seen as an element of $\mathcal{T}((\mathbb{R}^d))$
- $(s, u, t) \in \Delta^3$

Then

$$S_{su}(x) \otimes S_{ut}(x) = S_{st}(x)$$

Shuffle algebraic identity (1)

Shuffle of permutations: Consider

$$\sigma \in \Sigma_{\{1,\ldots,n\}}, \quad \text{and} \quad \tau \in \Sigma_{\{n+1,\ldots,n+k\}} .$$

Then we set

$$\mathrm{Sh}(\sigma, \tau) = \left\{\rho \in \Sigma_{\{1,\ldots,n+k\}} \text{ ; } \rho \text{ does not change the order of } \sigma \text{ and } \tau\right\}$$

Example: Take

$$\sigma = \{1, 3, 2\} \in \Sigma_{\{1,2,3\}}, \qquad \tau = \{5, 4\} \in \Sigma_{\{4,5\}}$$

Then

$$\rho = \{1, 5, 3, 2, 4\} \in \mathrm{Sh}(\sigma, \tau)$$

Shuffle algebraic identity (2)

Theorem 2.

Consider

- $x : [0, T] \to \mathbb{R}^d$ differentiable path
- $S(x)$ its signature seen as an element of $\mathcal{T}((\mathbb{R}^d))$
- $(s, t) \in \Delta^2$
- w, w' two words in $\mathcal{W}$

Then

$$\langle S_{st}(x), w \rangle \langle S_{st}(x), w' \rangle = \sum_{\phi \in \mathrm{Sh}(w, w')} \int_{\Delta_{st}^{n+n'}} \prod_{i=1}^{n+n'} \mathrm{d}x_{r_i}^{\hat{w}_{\phi(i)}}$$

Two basic analytic relations

Theorem 3.

Consider

- $x : [0, T] \to \mathbb{R}^d$ differentiable path
- $S(x)$ its signature seen as an element of $\mathcal{T}((\mathbb{R}^d))$
- $(s, t) \in \Delta^2$ and w, w' two words in $\mathcal{W}$

Then

1. Denoting $x^\phi = x \circ \phi$, we have an invariance,

$$[S(x)]_{\phi(s)\phi(t)} = S(x^\phi)_{st}$$

2. The following analytic estimate holds true,

$$||S_n(x)|| \leq \frac{(C_{\sigma,x})^n}{n!}$$

Computational efficiency

Example of discretization of : Consider

- The element $\langle S_{0,T}(x),\, (1,2)\rangle$ in the signature
- $\{t_i = t_i^n; 0 \le i \le n\}$ uniform partition of $[0,\, T]$

Then

1. We have

$$\langle S_{0,T}(x),\, (1,2)\rangle \simeq \sum_{i=0}^{n-1} \left(\delta x^1_{0 t_{i-1}} + \delta x^1_{t_{i-1} t_i} \right) \delta x^2_{t_i t_{i+1}}$$

2. This requires $O(n)$ operations

Interpretation of double iterated integral

Illustration:

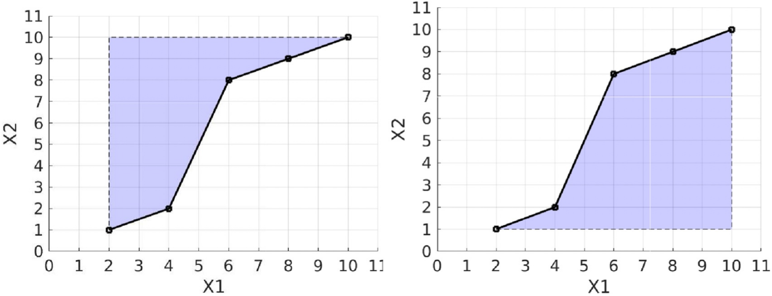


Figure: Left: $\langle S(x), (1,2) \rangle$ and Right: $\langle S(x), (2,1) \rangle$

Interpretation:

- If $\langle S_{0,T}(x), (1,2) \rangle$ is large, then x^2 goes faster than x^1
- If $\langle S_{0,T}(x), (2,1) \rangle$ is large, then x^1 goes faster than x^2

Characterization of paths

Basic characterization: For two bounded variation paths,

$$S(x)_{01} = S(y)_{01} \qquad \text{iff} \qquad x \sim y,$$

where $x \sim y$ means that x, y only differs by a tree-like path

Some references:

- Characterization: Lyons-Hambly '10
- Characterization, rough paths setting:
  Boedihardjo, Geng, Lyons, Yang '16
- Reconstruction in the $\mathcal{C}^1$ case: Lyons-Xu '18
- Reconstruction in the Hölder case: Xi Geng '17

1-d signatures and data analysis

Bottomline: 1-d signatures are successful features
$\hookrightarrow$ for numerous data analysis procedures

Classical examples (Lyons and collaborators):

- Chinese character recognition
- Finance time series
- Topological data analysis
- Diagnosis prediction

A study on Alzheimer disease

Illustration:

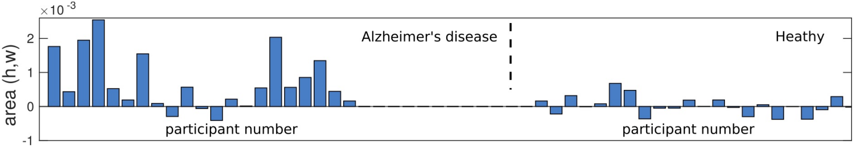


Figure: Comparison of $\langle S(x), (1,2)\rangle$, where 1 = Hippocampus, 2 = Whole brain

References

Friz, Peter K. and Victoir, Nicolas B.
Multidimensional stochastic processes as rough paths
Cambridge University Press, 2010

Geng, Xi
Reconstruction for the signature of a rough path.
Proceedings of the London Mathematical Society, 2017

Hambly, Ben and Lyons, Terry
Uniqueness for the signature of a path of bounded variation and the reduced path group.
Annals of Mathematics, 2010

References

Lyons, Terry; Ni, Hao and Oberhauser, Harald
A feature set for streams and an application to high-frequency financial tick data
Proceedings of the 2014 BigDataScience, 2014

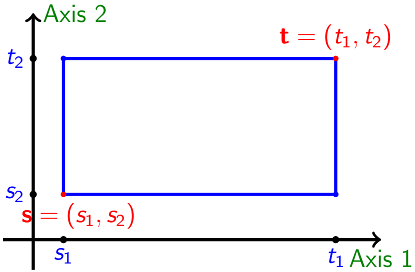
Moore, P. J.; Lyons, T. J. and Gallacher, J.
Using path signatures to predict a diagnosis of Alzheimer's disease
PLOS ONE, 2019

Outline

Outline

Notation for calculus in the plane

Basic notation for points in the plane:



Rectangles: We set

$$R = [\mathbf{s}, \mathbf{t}] := [s_1, t_1] \times [s_2, t_2], \quad \text{and} \quad [\mathbf{0}, \mathbf{T}] = [0, T]^2$$

Notation for 2d-signatures

Field: We first consider a $\mathbb{R}^d$-valued

$$X = \left\{ X^i_{\mathbf{s}} = X^i_{s_1;s_2};\ \mathbf{s} \in [\mathbf{0}, \mathbf{T}],\ i = 1\dots, d \right\}\ ,$$

where $i \equiv$ color (rgb) for an image

Differentials: We set

$$\begin{aligned} \mathrm{d}^i X_{\mathbf{r}} = \mathrm{d}_{12} X^i_{s;t} &= \partial_{12} X^i_{\mathbf{r}}\, \mathrm{d}r_1 \mathrm{d}r_2 \\ \hat{\mathrm{d}}^{ij} X_{\mathbf{r}} = \mathrm{d}_{\hat{1}\hat{2}} X^{ij}_{s;t} &= \partial_1 X^i_{\mathbf{r}}\, \partial_2 X^j_{\mathbf{r}}\, \mathrm{d}r_1 \mathrm{d}r_2. \end{aligned}$$

How do 2d-signatures show up?

Rectangular increment: For a field X we set

$$\square_{\mathbf{s},\mathbf{t}} X := X_{t_1;t_2} - X_{s_1;t_2} - X_{t_1;s_2} + X_{s_1;s_2}$$

Change of variable in the plane:

$$\square_{\mathbf{s},\mathbf{t}} f(X) = \int_{[\mathbf{s},\mathbf{t}]} \partial_i f(X_{\mathbf{r}}) \, \mathrm{d}^i X_{\mathbf{r}} + \int_{[\mathbf{s},\mathbf{t}]} \partial_{ij} f(X_{\mathbf{r}}) \, \hat{\mathrm{d}}^{ij} X_{\mathbf{r}}.$$

Problem:

Proper iteration of this formula for approximations

Part of a second order signature

Some elements of the signature:

Increment	*Interpretation*	*Regularity*	*Increment*	*Interpretation*
$\mathbf{x}^{1;2}$	$\int_1 \int_2 d_{12}x$	(γ_1, γ_2)	$\mathbf{x}^{\hat{1};\hat{2}}$	$\int_1 \int_2 d_{\hat{1}\hat{2}}x$
$\mathbf{x}^{11;02}$	$\int_1 d_1x \int_2 d_{12}x$	$(2\gamma_1, \gamma_2)$	$\mathbf{x}^{1\hat{1};0\hat{2}}$	$\int_1 d_1x \int_2 d_{\hat{1}\hat{2}}x$
$\mathbf{x}^{01;22}$	$\int_2 d_2x \int_1 d_{12}x$	$(\gamma_1, 2\gamma_2)$	$\mathbf{x}^{0\hat{1};2\hat{2}}$	$\int_2 d_2x \int_1 d_{\hat{1}\hat{2}}x$
$\mathbf{x}^{11;22}$	$\int_1 \int_2 d_{12}xd_{12}x$	$(2\gamma_1, 2\gamma_2)$	$\mathbf{x}^{1\hat{1};2\hat{2}}$	$\int_1 \int_2 d_{12}xd_{\hat{1}\hat{2}}x$
$\mathbf{x}^{\hat{1}1;\hat{2}2}$	$\int_1 \int_2 d_{\hat{1}\hat{2}}xd_{12}x$	$(2\gamma_1, 2\gamma_2)$	$\mathbf{x}^{\hat{1}\hat{1};\hat{2}\hat{2}}$	$\int_1 \int_2 d_{\hat{1}\hat{2}}xd_{\hat{1}\hat{2}}x$

References on rough sheets:

- Chouk-Gubinelli, unpublished
- Chouk-T, EJP '15, Skorohod-Stratonovich corrections

Other properties of 2-d signatures

Algebraic and analytic properties:

- Not clear, since the notion of signature is not clear
- Coordinate-wise reparametrization invariance
- Signatures generated by Jacobian minor operators
  $\hookrightarrow$ in Giusti, Lee, Nanda, Oberhauser
- Signatures generated by line integrals
  $\hookrightarrow$ in Diehl, Ebrahimi-Farad, Tapia
- Non-commutative Stokes point of view
  $\hookrightarrow$ in Lee-Oberhauser
- Overall, still a lot to be done

A modest goal

Our aim:

- Explore data analysis properties of 2d-signatures
- Simple numerical experiment on texture classification
  $\hookrightarrow$ in order to see if this makes sense empirically
- Try to find a signature for 2d-objects which has
  1. Simple enough structure
  2. Good algebraic-analytic properties
  3. Good discriminating properties

References

Chouk, Khalil and Gubinelli, Massimiliano
Rough sheets
Arxiv preprint, 2013

Chouk, Khalil and Tindel, Samy
Skorohod and Stratonovich integration in the plane
Electronic Journal of Probability, 2015

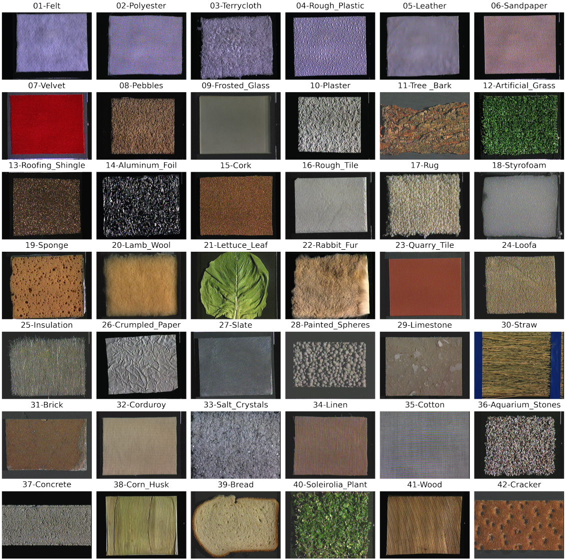
Giusti, Chad ; Lee, Darrick; Nanda, Vidit and Oberhauser, Harald
A topological approach to mapping space signatures
arXiv preprint, 2022

Diehl, Joscha; Ebrahimi-Fard, Kurusch; Tapia, Nikolas
Generalized iterated-sums signatures
J. Algebra 632 (2023), 801–824.

Outline

Texture dataset

- 42 textures
- Dataset: CuRRET
- Supervised class. procedure



Supervised learning

Procedure:

- We randomly sample (100×100)-sized images from each texture
- 10 samples from every texture used for training
- 100 images from every texture sampled for testing

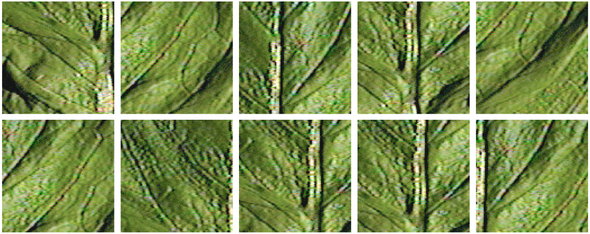


Figure: Ten samples from the texture "21-Lettuce Leaf"

2d-simplexes

Points in the plane: Consider

- $\mathbf{s} = (s_1, s_2)$ in $[0, T]^2$
- $\mathbf{t} = (t_1, t_2)$ in $[0, T]^2$
- $s_1 \le t_1$ and $s_2 \le t_2$

2d-simplexes:

$$\Delta^n_{[\mathbf{s},\mathbf{t}]} := \Delta^n_{s_1,t_1} \times \Delta^n_{s_2,t_2} = \Big\{ (\mathbf{r}^1, \ldots, \mathbf{r}^n) \in \big([0, T]^2\big)^n ; \\ s_1 \le r_1^1 \le \cdots \le r_1^n \le t_1 \text{ and } s_2 \le r_2^1 \le \cdots \le r_2^n \le t_2 \Big\}. \quad (2)$$

Features

A list of features: We include discretized versions of

$$\begin{aligned}
\mathbf{X}_{\mathbf{s},\mathbf{t}}^{(1,2);i} &= \int_{[\mathbf{s},\mathbf{t}]} \mathrm{d}^{i} x_{\mathbf{r}_1}^{i} \\
\mathbf{X}_{\mathbf{s},\mathbf{t}}^{(\hat{1},\hat{2});i} &= \int_{[\mathbf{s},\mathbf{t}]} \hat{\mathrm{d}}^{ii} x_{\mathbf{r}_1}^{i} \\
\mathbf{X}_{\mathbf{s},\mathbf{t}}^{(11,22);ii} &= \int_{\Delta_{\mathbf{s},\mathbf{t}}^{2}} \mathrm{d}^{i} x_{\mathbf{r}_1}^{i} \mathrm{d}^{i} x_{\mathbf{r}_2}^{i} \\
\mathbf{X}_{\mathbf{s},\mathbf{t}}^{(\hat{1}\hat{1},\hat{2}\hat{2});ii} &= \int_{\Delta_{\mathbf{s},\mathbf{t}}^{2}} \hat{\mathrm{d}}^{ii} x_{\mathbf{r}_1}^{i} \hat{\mathrm{d}}^{ii} x_{\mathbf{r}_2}^{i} \\
\mathbf{X}_{\mathbf{s},\mathbf{t}}^{(1\hat{1},2\hat{2});ii} &= \int_{\Delta_{\mathbf{s},\mathbf{t}}^{2}} \mathrm{d}^{i} x_{\mathbf{r}_1}^{i} \hat{\mathrm{d}}^{ii} x_{\mathbf{r}_2}^{i} \\
\mathbf{X}_{\mathbf{s},\mathbf{t}}^{(\hat{1}1,\hat{2}2);ii} &= \int_{\Delta_{\mathbf{s},\mathbf{t}}^{2}} \hat{\mathrm{d}}^{ii} x_{\mathbf{r}_1}^{i} \mathrm{d}^{i} x_{\mathbf{r}_2}^{i}
\end{aligned}$$

More about the procedure

Rotations:

- We average our features (See Mallat-Sifre)
  $\hookrightarrow$ over $\frac{\pi}{2}$ rotations

Dimension of feature space:

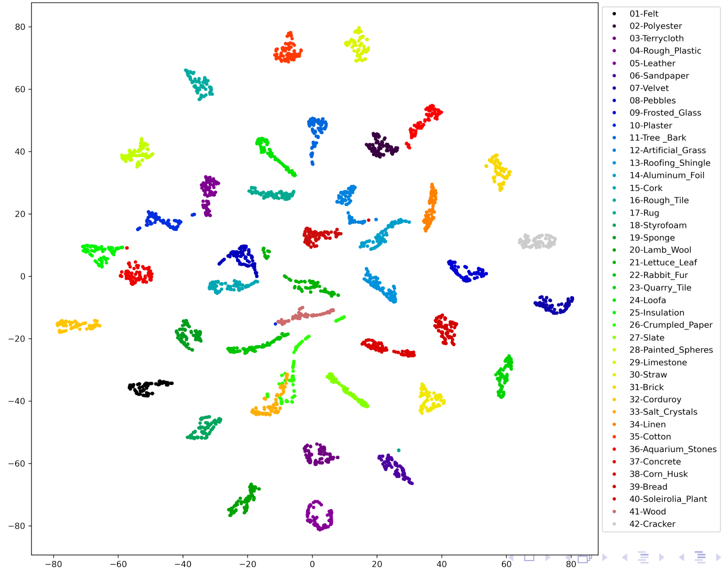
- For $\mathbf{X}_{\mathbf{s},\mathbf{t}}^{(1,2);i}$, i.e rectangular increments
  $\hookrightarrow$ PCA on all small increments
- Number of PCA components ≤ 40
- Overall, feature dimension ≤ 52
  $\hookrightarrow$ Considered as small

Classification method:

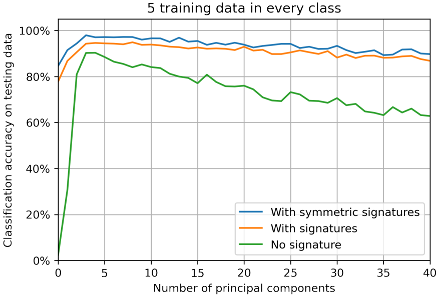
- Random forests

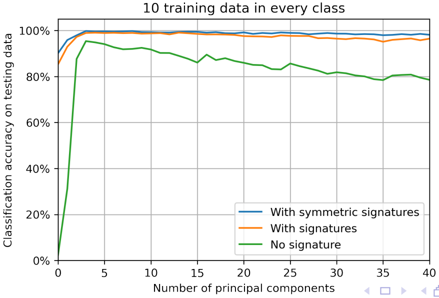
Outcome 1: visualization

Projection using t-distributed stochastic neighbor embedding:



Outcome 2: accuracy





Outline

Brief summary

Conclusion 1:

- Signatures based on 2-d increments are worth exploring

Conclusion 2:

- We should look for simple enough structures
- At least simpler than structure from calculus in the plane

References

Zhang, Sheng and Lin, Guang and Tindel, Samy
Two-dimensional signature of images and texture classification
Proceedings of the Royal Society A, 2022

Outline

2d-simplexes (repeated)

Points in the plane: Consider

- $\mathbf{s} = (s_1, s_2)$ in $[0, T]^2$
- $\mathbf{t} = (t_1, t_2)$ in $[0, T]^2$
- $s_1 \leq t_1$ and $s_2 \leq t_2$

2d-simplexes:

$$\Delta^n_{[\mathbf{s},\mathbf{t}]} := \Delta^n_{s_1,t_1} \times \Delta^n_{s_2,t_2} = \Big\{ (\mathbf{r}^1, \ldots, \mathbf{r}^n) \in \left([0, T]^2\right)^n ; \; s_1 \leq r_1^1 \leq \cdots \leq r_1^n \leq t_1 \text{ and } s_2 \leq r_2^1 \leq \cdots \leq r_2^n \leq t_2 \Big\}. \quad (3)$$

Definition of the simple signature

Definition 4.

Consider

- $\mathbf{s}, \mathbf{t}$ in $[0, T]^2$
- $w = (i_1, \ldots, i_n) \in \mathcal{W}$

Then we set

$$\left\langle \mathbf{S}^{\mathrm{Id}}_{\mathbf{s},\mathbf{t}}(X),\, w \right\rangle = \int_{\Delta^n_{[\mathbf{s},\mathbf{t}]}} \mathrm{d}^{i_1} X_{\mathbf{r}^1} \cdots \mathrm{d}^{i_n} X_{\mathbf{r}^n}$$

Recursive definition: We also have

$$\left\langle \mathbf{S}^{\mathrm{Id}}_{\mathbf{s},\mathbf{t}}(X),\, w \right\rangle = \int_{[\mathbf{s},\mathbf{t}]} \left\langle \mathbf{S}^{\mathrm{Id}}_{\mathbf{s},\mathbf{r}}(X),\, (i_1, \ldots, i_{n-1}) \right\rangle \mathrm{d}^{i_n} X_{\mathbf{r}}$$

How does the simple signature show up?

Equation: Let

- $X : [0, T]^2 \to \mathbb{R}^d$
- $v \in \mathbb{R}^d$ and $\{A^i \,;\, i = 1, \ldots, d\}$ matrices in $\mathbb{R}^{d,d}$

Then let Y be the solution to

$$Y_{\mathbf{t}} = v + \sum_{i=1}^{d} \int_{[\mathbf{s},\mathbf{t}]} A^i Y_{\mathbf{r}} \, \mathrm{d}^i X_{\mathbf{r}}$$

Expansion: Y can be formally expanded as

$$\Box_{\mathbf{s},\mathbf{t}} Y = \sum_{w \in \mathcal{W}} A^{\circ w} v \left\langle \mathbf{S}^{\mathrm{Id}}_{\mathbf{s},\mathbf{t}}(X), w \right\rangle$$

Lack of shuffle property (1)

Desirable property: Take

- $\langle \mathbf{S}^{\mathrm{Id}}_{\mathbf{s},\mathbf{t}}(X),\, w_1 \rangle$ and $\langle \mathbf{S}^{\mathrm{Id}}_{\mathbf{s},\mathbf{t}}(X),\, w_2 \rangle$ in the signature

We wish to have

$$\langle \mathbf{S}^{\mathrm{Id}}_{\mathbf{s},\mathbf{t}}(X),\, w_1 \rangle \langle \mathbf{S}^{\mathrm{Id}}_{\mathbf{s},\mathbf{t}}(X),\, w_2 \rangle = \sum \text{Elements of the signature}$$

Simple example: Consider $X : [0, T]^2 \to \mathbb{R}$ and

- $\langle \mathbf{S}^{\mathrm{Id}}_{\mathbf{s},\mathbf{t}}(X),\, w_1 \rangle = \int_{[\mathbf{s},\mathbf{t}]} \mathrm{d}X_{\mathbf{r}}$
- $\langle \mathbf{S}^{\mathrm{Id}}_{\mathbf{s},\mathbf{t}}(X),\, w_2 \rangle = \int_{[\mathbf{s},\mathbf{t}]} \mathrm{d}X_{\mathbf{v}}$

Then define

$$\Pi_{\mathbf{s},\mathbf{t}} = \int_{[\mathbf{s},\mathbf{t}]} \mathrm{d}X_{\mathbf{r}} \int_{[\mathbf{s},\mathbf{t}]} \mathrm{d}X_{\mathbf{v}}$$

Lack of shuffle property (2)

Relation for Π: Recall that

$$\Pi_{\mathbf{s},\mathbf{t}} = \int_{[\mathbf{s},\mathbf{t}]} \mathrm{d}X_{\mathbf{r}} \int_{[\mathbf{s},\mathbf{t}]} \mathrm{d}X_{\mathbf{v}}$$

Then

$$\Pi_{\mathbf{s},\mathbf{t}} = \underbrace{\int_{\Delta^2_{[\mathbf{s},\mathbf{t}]}} \mathrm{d}X_{r_1^1;r_2^1} \mathrm{d}X_{r_1^2;r_2^2}}_{\Pi^1_{\mathbf{s},\mathbf{t}}} + \underbrace{\int_{\Delta^2_{[\mathbf{s},\mathbf{t}]}} \mathrm{d}X_{r_1^1;r_2^2} \mathrm{d}X_{r_1^2;r_2^1}}_{\Pi^2_{\mathbf{s},\mathbf{t}}}$$

Identifying Π^1: One can easily see that

$$\Pi^1_{\mathbf{s},\mathbf{t}} = \langle \mathbf{S}(X), (1,1) \rangle_{\mathbf{s},\mathbf{t}}$$

Lack of shuffle property (3)

Problem with Π^2: Recall that

$$\Pi^2_{\mathbf{s},\mathbf{t}} = \int_{\Delta^2_{[\mathbf{s},\mathbf{t}]}} \mathrm{d}X_{r_1^1;r_2^2}\mathrm{d}X_{r_1^2;r_2^1}$$

Then

- This object is not in the signature
- This is due to the permutation $r_2^2 \longleftrightarrow r_2^1$

Remark:

- This problem with permutations pops up at many places
- We thus introduce a new signature involving permutations

Outline

Definition of extended words

Definition 5.

For $n \geq 1$ we set

$$\hat{\mathcal{W}}_n = \{(w, \nu) | \, w \in \mathcal{W}_n, \text{ and } \nu \in \Sigma_{\{1,\ldots,n\}}\}$$

Then the set of extended words is given by

$$\hat{\mathcal{W}} = \bigcup_{n=0}^{\infty} \hat{\mathcal{W}}_n$$

2d-simplexes (repeated)

Points in the plane: Consider

- $\mathbf{s} = (s_1, s_2)$ in $[0, T]^2$
- $\mathbf{t} = (t_1, t_2)$ in $[0, T]^2$
- $s_1 \leq t_1$ and $s_2 \leq t_2$

2d-simplexes:

$$\Delta^n_{[\mathbf{s},\mathbf{t}]} := \Delta^n_{s_1,t_1} \times \Delta^n_{s_2,t_2} = \Big\{ (\mathbf{r}^1, \ldots, \mathbf{r}^n) \in \left([0, T]^2\right)^n ; \\ s_1 \leq r^1_1 \leq \cdots \leq r^n_1 \leq t_1 \text{ and } s_2 \leq r^1_2 \leq \cdots \leq r^n_2 \leq t_2 \Big\}. \quad (4)$$

Definition of the extended signature

Definition 6.

Consider

- $\mathbf{s}, \mathbf{t}$ in $[0, T]^2$
- $(w, \nu) \in \hat{\mathcal{W}}_n$ with
  - $w = (i_1, \ldots, i_n)$
  - $\nu \in \Sigma_{\{1,\ldots,n\}}$

Then we set

$$\langle \mathbf{S}_{\mathbf{s},\mathbf{t}}(X), (w, \nu) \rangle = \int_{\Delta^n_{[\mathbf{s},\mathbf{t}]}} \prod_{i=1}^{n} \mathrm{d}X^{w_i}(r_1^i, r_2^{\nu_i})$$

Claim:

This extended signature has good algebraic properties

Shuffle of words

Definition 7.

Let

- $n \geq 1$, $k \geq 1$
- Word $w = (i_1, \ldots, i_n)$
- Word $v = (j_1, \ldots, j_k)$
- $[w, v] = (i_1, \ldots, i_n, j_1, \ldots, j_k)$

Then the shuffle of v and w is given by

$$\mathrm{Sh}(w, v) = \Big\{ \text{Permutations of } [w, v]; \\ \text{orders of } w \text{ and } v \text{ are not changed} \Big\}$$

Shuffle of permutations

Definition 8.

Let

- $n \geq 1, \; k \geq 1$
- Permutation $\sigma \in \Sigma_{\{1,\ldots,n\}}$
- Permutation $\tau \in \Sigma_{\{n+1,\ldots,n+k\}}$
- $[w, v] = (i_1, \ldots, i_n, j_1, \ldots, j_k)$

Then the shuffle of σ and τ is given by

$$\mathrm{Sh}(\sigma, \tau) = \Big\{\rho \in \Sigma_{\{1,\ldots,n+k\}} \; ; \; \rho \text{ does not change the order of } \sigma \text{ and } \tau\Big\}$$

Shuffle relation

Theorem 9.

Let

- $X : [0, T]^2 \to \mathbb{R}^d$ smooth path
- $\mathbf{s}, \mathbf{t}$ in $[0, T]^2$
- (w, ν) and (w', ν') elements of $\hat{\mathcal{W}}$

Then we have

$$\langle \mathbf{S}_{\mathbf{s},\mathbf{t}}(X), (w, \nu)\rangle \langle \mathbf{S}_{\mathbf{s},\mathbf{t}}(X), (w', \nu')\rangle$$
$$= \sum_{\phi \in \mathrm{Sh}(w,w')} \sum_{\rho \in \mathrm{Sh}(\nu,\nu')} \langle \mathbf{S}_{\mathbf{s},\mathbf{t}}(X), (\phi([w, w']), \rho \circ \phi([\nu, \nu']))\rangle$$

Other results and perspectives

Other results:

1. Partial versions of Chen's relations
   - Splits in direction 1 and 2
   - Symmetrized signature
2. Invariances by change of variables

Perspectives:

1. Full algebraic setting for Chen
2. Relation with stochastic calculus in the plane
3. Relation with non-commutative Stokes theorem

Outline

Brief summary

Examples of natural notions of 2d-signatures: Based on

- Calculus in the plane
- Jacobian minors
- Noncommutative Stokes

Another natural notion: Based on

- PDEs for image processing

Smoothing an image

Setting: We consider

- $u_0 : \Omega \equiv [0, T]^2 \to \mathbb{R}$ (original noisy image)
- G_σ Gaussian kernel

Smoothed version: For σ to be calibrated,

$$u_\sigma = G_\sigma * u_0$$

PDE version: u_σ can also be computed through

$$\begin{cases} \partial_t u = \operatorname{div}(\nabla u), & \text{in } \Omega \\ \frac{\partial u}{\partial \mathbf{n}} = 0, & \text{on } \partial\Omega, \end{cases}$$

Modulating the diffusion

Problem with diffusion equation:

- Images become very blurry
- Main problem: respect the corners and edges
- Solution: smaller diffusion when gradient is large

New equation: For g decaying at ∞,

$$\begin{cases} \partial_t u = \operatorname{div}\left(g(|\nabla u|)\,\nabla u\right), & \text{in } \Omega \\ \frac{\partial u}{\partial \mathbf{n}} = 0, & \text{on } \partial\Omega\,, \end{cases}$$

A class of PDEs for image processing

Basic model: By Rudin-Osher-Fatemi, $> 18,000$ citations

$$\partial_t u = \operatorname{div}\left(\frac{\nabla u}{|\nabla u|}\right) - \lambda\,(u - u_0)$$

Remarks about the model:

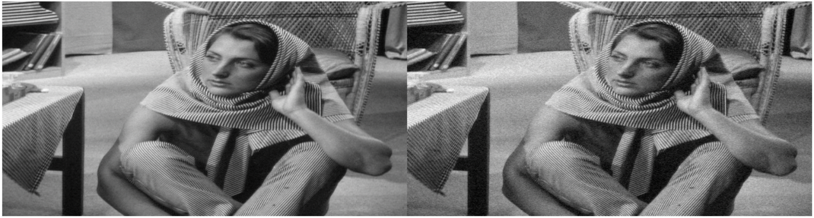
- Numerous extensions (4th order, anisotropic)
- Model justified by optimization considerations

Generic smoothed model: With λ regularization parameter

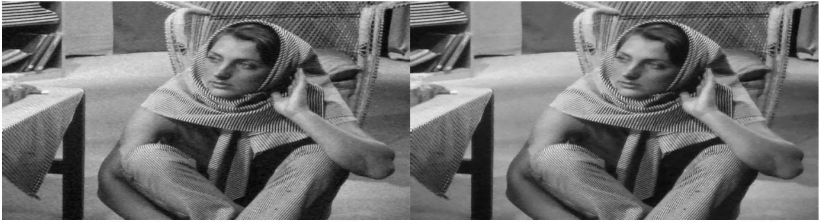
$$\partial_t u = \underbrace{\operatorname{div}\left(\varphi(\nabla u)\,\nabla u\right)}_{\text{smoothing + edges}} - \underbrace{\lambda\,(u - u_0)}_{\text{stay close to original } u_0}$$

An implementation from Osher-Solé-Vese (2003)

Corrupted image:



Restored image:



Features from PDEs

Basic idea:

- Use regularity structures methods to expand the PDE
  $\hookrightarrow$ Produces a hierarchy of linear PDEs
- Use the solutions to this family of PDEs as features

Justification:

- Smoothing methods already been used for representation
- Regularity structures $\longrightarrow$ algebraic/analytic machinery
- Approach already used by Chevyrev-Gerasimovics-Weber

A generic coefficient

Method implemented:

- Taken from Otto-Sauer-Smith-Weber, using multiindex notation
- Below $\mathcal{E}_m$ is an awful index set

Basic operator: We set

$$\mathcal{A}_0 = \varphi_0 \Delta u - \lambda\, u$$

Hierarchy of PDEs: We get

$$(\partial_t - \mathcal{A}_0)\, \Pi_{\mathbf{x}m} = \sum_{n,k,p,m_p^k,m^{k+1} \in \mathcal{E}_m} \operatorname{div}\left(\left(\prod_{l=1}^{d} \prod_{j=1}^{n(l)} \nabla^{(k)} \Pi_{\mathbf{x}m_j^l} \right) \nabla \Pi_{\mathbf{x}m^{k+1}} \right)$$

References

Osher, Stanley; Solé, Andrés; Vese, Luminita.
Image decomposition and restoration using total variation minimization and the H^1 norm.
Multiscale Model. Simul.1, 2003

Rudin, Leonid I.; Osher, Stanley; Fatemi, Emad.
Nonlinear total variation based noise removal algorithms.
Phys. D, 1992

Felix Otto, Jonas Sauer, Scott Smith, Hendrik Weber.
A priori bounds for quasi-linear SPDEs in the full sub-critical regime.
Arxiv preprint, 2021.