Signature features with the visibility transformation

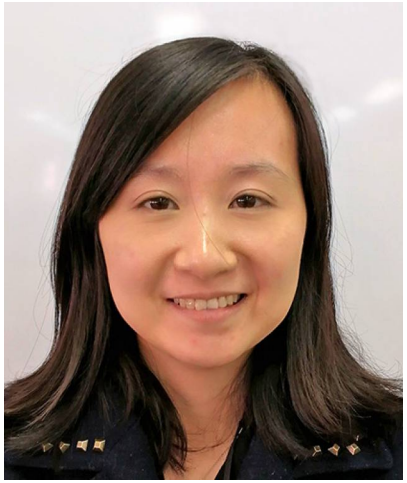
Poster presentation at ICPR2020, Milan Italy, 10-15 Jan 2021

Yue Wu (yue.wu@maths.ox.ac.uk)
DataSig and Oxford

DataSig

A rough path between
mathematics and data science

Co-authors and acknowledgements



Dr. Hao Ni
Department of Mathematics,
University College of London;
Alan Turning Institute

Prof. Terry Lyons FRS
Mathematical Institute,
University of Oxford;
Alan Turing Institute

Prof. Robin Hudson
Department of Mathematical Sciences,
Loughborough University

YW, HN, TJL were supported by the Alan Turing Institute under the EPSRC grant EP/N510129/1 and by EPSRC under EP/S026347/1.

Outline

1. Signatures as a natural feature set

2. Visibility transformation

Signatures as a natural feature set

A sequence of (input,output) pair, denoted by $\{(\mathbf{X}_i, y_i)\}_i$, we make inference f between input and output such that

$$f(\mathbf{X}_i) \approx y_i.$$

Signatures of path (data stream) as features:

- Determining solutions for controlled differential equations [Chen, 1958], [Hambly&Lyons, 2010], [Boedihardjo et al, 2016].
- Invariance under time reparameterisation.
- Unwrapping nonliearity.
- Fixed dimension under length variations, vectorisation.
- Faithful representation: tree-like equivalence.

Definition: signatures

Consider time-dependent path $X : [a, b] \to \mathbb{R}^d$. Now for any $t \in [a, b]$ and a fixed ordered multi-index collection $(i_1, \ldots, i_k)$, with $k \in \mathbb{N}$ and $i_j \in [d]$ for $j \in [k] := \{1, \ldots, k\}$, define the coordinate iterated integral by

$$S(X)_{a,t}^{i_1,\ldots,i_k} := \int_{a<t_k<t} \cdots \int_{a<t_1<t_2} \mathrm{d}X_{t_1}^{i_1} \ldots \mathrm{d}X_{t_k}^{i_k}. \tag{1}$$

Definition

The *signature* of a path X is the infinite collection of all iterated integrals of X:

$$S(X)_{a,b} := (1, S(X)_{a,b}^1, \ldots, S(X)_{a,b}^d, S(X)_{a,b}^{1,1}, S(X)_{a,b}^{1,2}, \ldots). \tag{2}$$

The finite collection of all terms $S(X)_{a,b}^{i_1,\ldots,i_k}$ with the multi-index of fixed length k is termed as the *kth level of the signature*. The truncated signature up to the pth level is denoted by $\lfloor S(X)_{a,b} \rfloor_p$.

Only capturing the effect of pattern change and not ones depending on the absolute position.

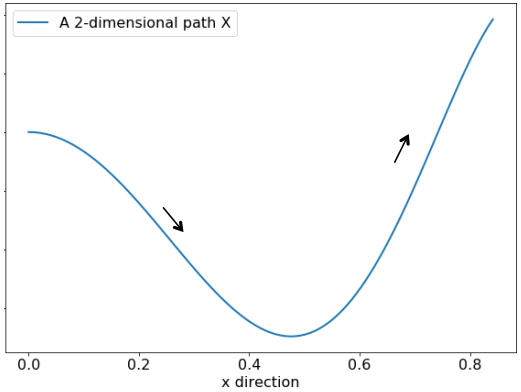
$$\text{Path X} \overset{\text{Signature}}{\Longrightarrow} \text{Incremental effects of path X.}$$

Outline

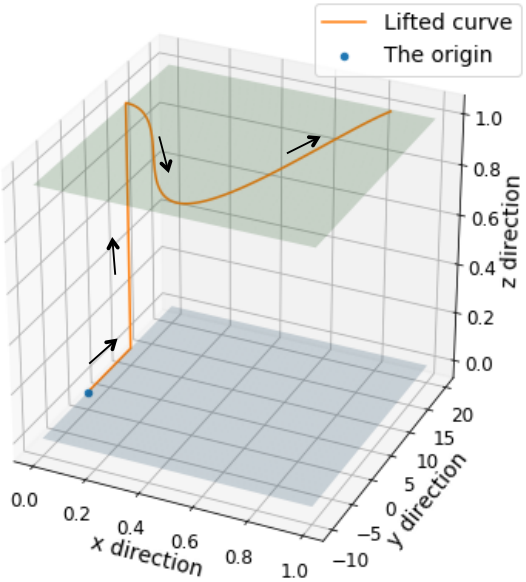
1. Signatures as a natural feature set

2. Visibility transformation

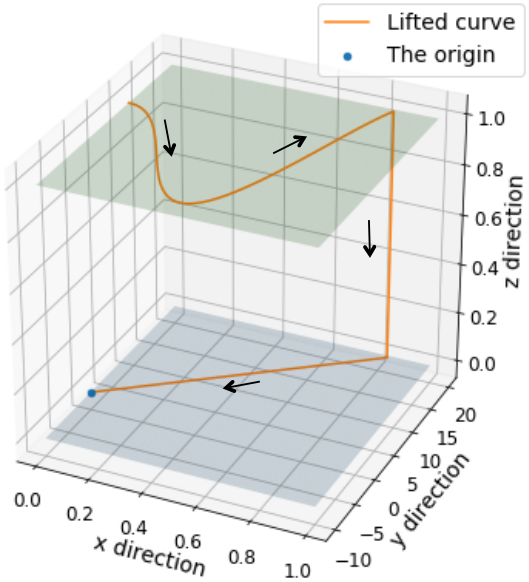
For continuous path of finite length



(a) Original 2-dimensional path X.



(b) $\mathbf{I}$-visibility transformation ($\gamma_{\mathbf{I}}$).



(c) $\mathbf{T}$-visibility transformation ($\gamma_{\mathbf{T}}$).

Main property

Theorem

Given X and a multi-index collection J with $d+1 \notin J$. Define $J^- := (d+1|J)$ and $J^+ := (J|d+1)$, where $d+1$ is prefixed (resp. postfixed) to J on the left(resp. right). Then

$$S(\gamma_{\mathbf{I}}(X))^{J^-} = S_X^J e_{d+1} e_{j_1} \dots e_{j_{|J|}},\ S(\gamma_{\mathbf{I}}(X))^{J^+} = \frac{1}{|J|!} \prod_{j \in J} X_0^j e_{j_1} \cdots e_{j_{|J|}} e_{d+1}, \tag{3}$$

where S_X^J is the corresponding coefficient of $S(X)$. Similarly, for the $\mathbf{T}$-visibility transformation, we have

$$S(\gamma_{\mathbf{T}}(X))^{J^+} = S_X^J e_{j_1} \dots e_{j_{|J|}} e_{d+1},\ S(\gamma_{\mathbf{T}}(X))^{J^-} = \frac{(-1)^{|J|+1}}{|J|!} \prod_{j \in J} X_1^j e_{d+1} e_{j_1} \cdots e_{j_{|J|}}. \tag{4}$$

Path X $\overset{\text{VT}}{\Longrightarrow}$ Path $\bar{\text{X}}$ $\overset{\text{Signature}}{\Longrightarrow}$ Positional and incremental effects of path X

discrete VT

Given streamed data $\mathbf{X} = (\mathbf{x}_1, \ldots, \mathbf{x}_n)$, $\mathbf{x}_i \in \mathbb{R}^d$.

$$\text{data } \mathbf{X} \xRightarrow[\text{linear interpolation}]{\text{piecewise}} \text{path } X \Rightarrow \text{signatures}$$

$$\text{data } \mathbf{X} \xRightarrow{\text{Python packages}} \text{signatures}$$

discrete VT

Given streamed data $\mathbf{X} = (\mathbf{x}_1, \ldots, \mathbf{x}_n)$, $\mathbf{x}_i \in \mathbb{R}^d$.

$$\text{data } \mathbf{X} \xRightarrow[\text{linear interpolation}]{\text{piecewise}} \text{path } X \Rightarrow \text{signatures}$$

$$\text{data } \mathbf{X} \xRightarrow{\text{Python packages}} \text{signatures}$$

One needs a discrete transform such that

$$\text{data } \mathbf{X} \xRightarrow{\text{discrete T}} \text{data } \bar{\mathbf{X}} \xRightarrow{\text{Python packages}} \text{signatures}$$

and

$$\text{data } \bar{\mathbf{X}} \xRightarrow[\text{linear interpolation}]{\text{piecewise}} \text{path } \bar{X}$$

Discrete VT

Take a discrete data with two 2-dimensional observations $[1,2]^T, [3,4]^T$ for example, the discrete $\mathbf{I}$-visibility transformation (IVT) would give

$$\begin{bmatrix}1\\2\end{bmatrix}, \begin{bmatrix}3\\4\end{bmatrix} \overset{\text{IVT}}{\Longrightarrow} \begin{bmatrix}0\\0\\0\end{bmatrix}, \begin{bmatrix}1\\2\\0\end{bmatrix}, \begin{bmatrix}1\\2\\1\end{bmatrix}, \begin{bmatrix}3\\4\\1\end{bmatrix},$$

and the discrete $\mathbf{T}$-visibility transformation (TVT) would give

$$\begin{bmatrix}1\\2\end{bmatrix}, \begin{bmatrix}3\\4\end{bmatrix} \overset{\text{TVT}}{\Longrightarrow} \begin{bmatrix}1\\2\\1\end{bmatrix}, \begin{bmatrix}3\\4\\1\end{bmatrix}, \begin{bmatrix}3\\4\\0\end{bmatrix}, \begin{bmatrix}0\\0\\0\end{bmatrix}.$$

Discrete VT

Take a discrete data with two 2-dimensional observations $[1,2]^T, [3,4]^T$ for example, the discrete $\mathbf{I}$-visibility transformation (IVT) would give

$$\begin{bmatrix}1\\2\end{bmatrix}, \begin{bmatrix}3\\4\end{bmatrix} \overset{\text{IVT}}{\Longrightarrow} \begin{bmatrix}0\\0\\0\end{bmatrix}, \begin{bmatrix}1\\2\\0\end{bmatrix}, \begin{bmatrix}1\\2\\1\end{bmatrix}, \begin{bmatrix}3\\4\\1\end{bmatrix},$$

and the discrete $\mathbf{T}$-visibility transformation (TVT) would give

$$\begin{bmatrix}1\\2\end{bmatrix}, \begin{bmatrix}3\\4\end{bmatrix} \overset{\text{TVT}}{\Longrightarrow} \begin{bmatrix}1\\2\\1\end{bmatrix}, \begin{bmatrix}3\\4\\1\end{bmatrix}, \begin{bmatrix}3\\4\\0\end{bmatrix}, \begin{bmatrix}0\\0\\0\end{bmatrix}.$$

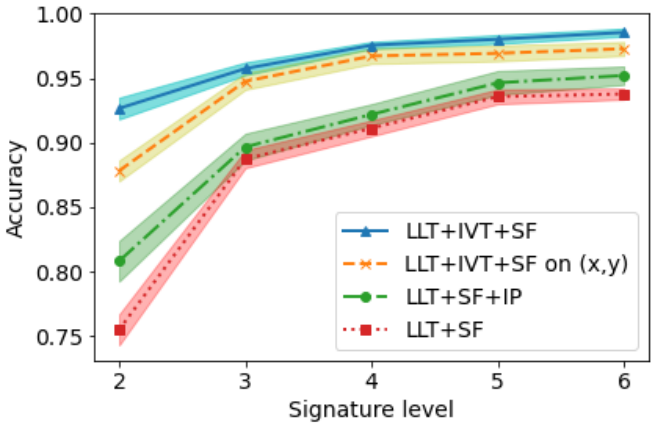
Application: character trajectories dataset

- The data [Williams et al, 2006] consists of 2858 instances for 20 different characters, and was captured using a WACOM tablet at 200Hz. Each character sample is a 3-dimensional pen tip velocity trajectory, namely (x, y, p). The original handwriting data contains training set (50%) and testing set (50%).

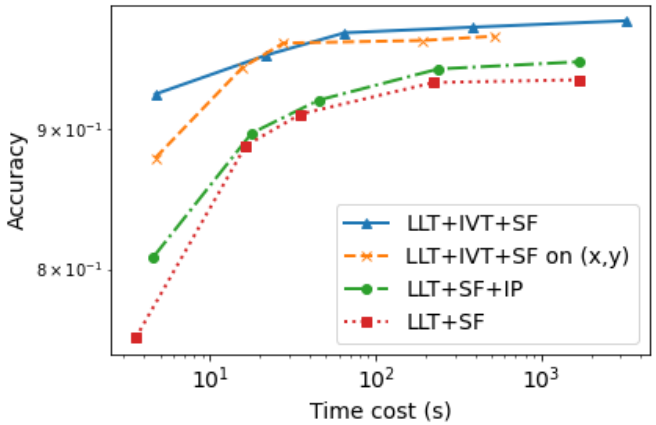
- 4 Different features: truncated signatures with the lead lag transform (LLT), truncated signatures with LLT and being prefixed by the explicit initial position (LLT+IP), and truncated signatures with LLT and the discrete $\mathbf{I}$-visibility transformation (LLT+IVT). We also extracted truncated signature features with LLT and IVT on the trajectory, namely the (x, y) path (LLT+IVT on (x, y)). In the experiment, the signature features were truncated to levels $\{2, 3, 4, 5, 6\}$.

- lightGBM Classifer: hyperparameter tuning implemented via grid search with cross validation.

Application: character trajectories dataset



(a) Signature level VS average accuracy with standard deviation.

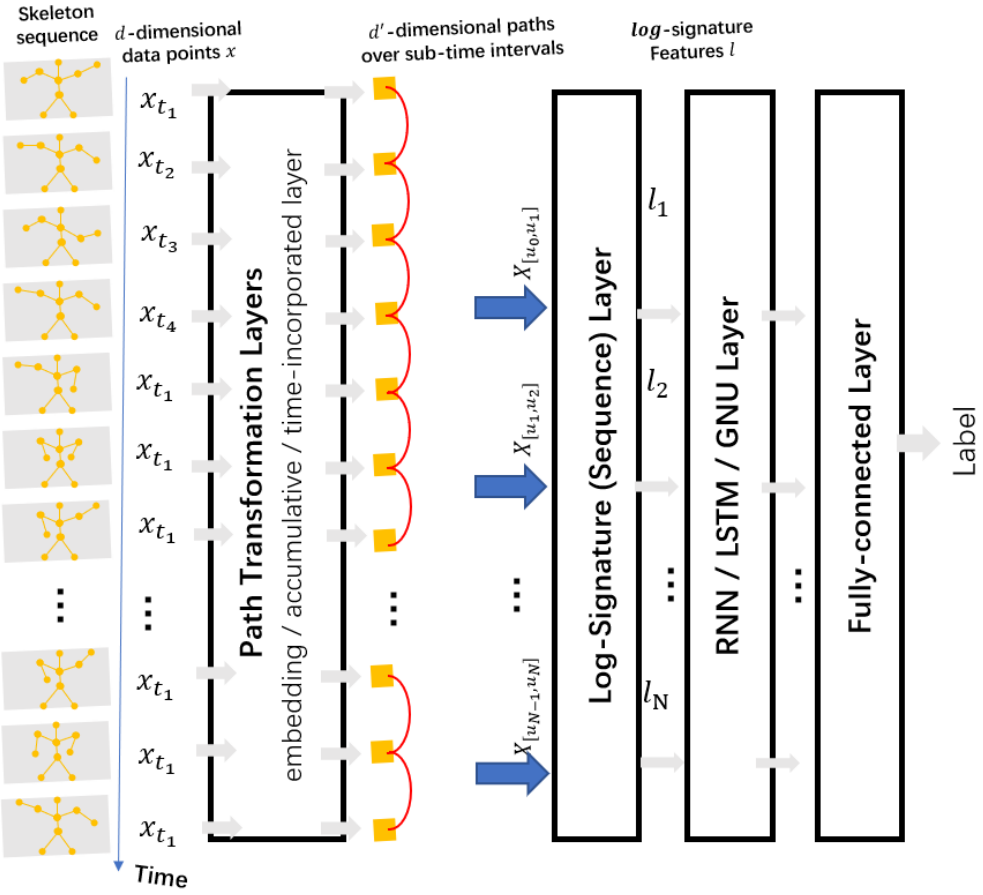


(b) CPU time VS average accuracy (log-log scale).

Method	Accuracy
ϕ(O,HMM)+SVM [Perina et al, 2009]	92.91%
TK [Tsuda et al, 2002]	93.67%
LLT+IVT+SF on (x, y)	**97.27%**
SDD [grabocka et al, 2016]	98.00%
MCDS [iosifidis et al, 2012]	98.25%
LLT+IVT+SF	**98.54%**

Gesture classification: Chalearn 2013 data

- The ChaLearn 2013 multi-modal gesture dataset [Escalera et al,2013] contains 23 hours of Kinect data of 27 subjects performing 20 Italian gestures.
- Liao et al [Liao et al, 2019] proposed a log-signature-based recurrent neural network model.



Gesture classification: Chalearn 2013 data

Modified PT-Logsig-RNN: adding a visibility transformation layer to architecture of PT-Logsig-RNN Model.

Table: Comparison of different methods on the Chalearn 2013 data.

Method	Accuracy
Deep LSTM [Shahroudy et al, 2016]	87.10%
Two-stream LSTM [Wang et al, 2017]	91.70%
ST-LSTM + Trust Gate [Liu et al, 2017]	92.00%
Three-stream net TTM [Li et al, 2019]	92.08%
PT-Logsig-RNN [Liao et al, 2019]	92.21%
Modified PT-Logsig-RNN	**92.89%**

Conclusion

To capture the informative information from streamed data for learning tasks, we explore a transformation that encodes the effects on the absolute position of streamed data into signature features with theoretical justifications. The enhanced feature is unified, theoretical-backed, and simple to implement with. It is superior to many benchmark methods that require handy data preparation and implementation of complicated algorithms in applications when absolute position of the data is intrusive.

References

[Boedihardjo et al, 2016] Boedihardjo, H., Geng, X., Lyons, T. and Yang, D., The signature of a rough path: uniqueness. *Advances in Mathematics*, 293 (2016): 720-737.

[Chen, 1958] Chen, K., Integration of paths|a faithful representation of paths by non-commutative formal power series. . *Amer. Math. Soc.*, 89 (1958): 395-407.

[Escalera et al,2013] Escalera, S., Gonzàlez, J., Baró, X., Reyes, M., Lopes, O., Guyon, I., Athitsos, V. and Escalante, H., *Multi-modal gesture recognition challenge 2013: Dataset and results*, *In Proceedings of the 15th ACM on International conference on multimodal interaction*, (2013): 445-452.

[grabocka et al, 2016] Grabocka, J., Wistuba, M. and Schmidt-Thieme, L., Fast classification of univariate and multivariate time series through shapelet discovery. *Knowledge and information systems*, 49.2 (2016): 429-454.

[Hambly & Lyons, 2010] Hambly, B. and Lyons, T., Uniqueness for the signature of a path of bounded variation and the reduced path group. *Annals of Mathematics*, 171.1 (2010):109-167.

[iosifidis et al, 2012] Iosifidis, A., Tefas, A. and Pitas, I., Multidimensional sequence classification based on fuzzy distances and discriminant analysis. *IEEE Transactions on Knowledge and Data Engineering*, 25.11 (2012): 2564-2575.

[Li et al, 2019] Li, C., Zhang, X., Liao, L., Jin, L. and Yang, W., Skeleton-based gesture recognition using several fully connected layers with path signature features and temporal transformer module. *In Proceedings of the AAAI Conference on Artificial Intelligence*, 33 (2019): 8585-8593.

[Liao et al, 2019] Liao, S., Lyons, T., Yang, W. and Ni, H., Learning stochastic differential equations using RNN with log signature features. *arXiv preprint*, arXiv:1908.08286.

References

[Liu et al, 2017] Liu, J., Shahroudy, A., Xu, D., Kot, A.C. and Wang, G., Skeleton-based action recognition using spatio-temporal lstm network with trust gates. *IEEE transactions on pattern analysis and machine intelligence*, 40.12(2017):3007-3021.

[Perina et al, 2009] Perina, A., Cristani, M., Castellani, U. and Murino, V., A new generative feature set based on entropy distance for discriminative classification. *International Conference on Image Analysis and Processing*, (2009): 199-208, Springer, Berlin, Heidelberg.

[Shahroudy et al, 2016] Shahroudy, A., Liu, J., Ng, T.T. and Wang, G., Ntu rgb+ d: A large scale dataset for 3d human activity analysis. *In Proceedings of the IEEE conference on computer vision and pattern recognition*, (2016):1010-1019.

[Tsuda et al, 2002] Tsuda, K., Kawanabe, M., Rätsch, G., Sonnenburg, S. and Müller, K.R., A new discriminative kernel from probabilistic models. *Advances in Neural Information Processing Systems*, (2002): 977-984.

[Wang et al, 2017] Wang, H. and Wang, L., Modeling temporal dynamics and spatial configurations of actions using two-stream recurrent neural networks. *In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2017: 499-508.

[Williams et al, 2006] Williams, B.H., Toussaint, M. and Storkey, A.J., Extracting motion primitives from natural handwriting data. *International Conference on Artificial Neural Networks*, (2006): 634-643, Springer, Berlin, Heidelberg.