Appendix

Shujian Liao[1]
shujian.liao.18@ucl.ac.uk
Terry Lyons[2,3]
terry.lyons@maths.ox.ac.uk
Weixin Yang[2,3]
yangw2@maths.ox.ac.uk
Kevin Schlegel[1,3]
k.schlegel@ucl.ac.uk
Hao Ni[1,3]
h.ni@ucl.ac.uk

[1] University College London
London, UK
[2] University of Oxford
Oxford, UK
[3] The Alan Turing Institute
London, UK

The (log)-signature of a path is a core mathematical object in rough path theory, which is a branch of stochastic analysis. In this appendix, we give a brief overview of the signature and log signature of a path, and their properties which are useful in the context of machine learning. Besides we provide illustrative examples via the python notebook in the supplementary material. In the following, for concreteness, we focus on paths of bounded variation, but the definition of the signature can be generalized to paths of finite p-variation. Interested readers can refer to [6] for a rigorous introduction of the (log)-signature of a path.

1 The signature of a path

Let us recall the definition of the signature of a path. Let $J := [S,T]$ denote a compact interval and $E := \mathbb{R}^d$. $BV(J,E)$ denotes the space of all continuous paths of finite length from J to E.

Definition 1.1 (The Signature of a Path). *Let $X \in BV(J,E)$. Let $I = (i_1, i_2, \cdots, i_n)$ be a multi-index of length n where $i_j \in \{1, \cdots, d\}, \forall j \in \{1, \cdots, n\}$. Define the coordinate signature of the path X_J associated with the index I as follows:*

$$X_J^I = \underset{\substack{u_1<\cdots<u_k \\ u_1,\ldots,u_k \in J}}{\int\cdots\int} dX_{u_1}^{(i_1)} \ldots dX_{u_n}^{(i_2)}.$$

The signature of X is defined as follows:

$$S(X)_J = (1, \mathbf{X}_J^1, \ldots, \mathbf{X}_J^k, \ldots) \tag{1}$$

where $\mathbf{X}_J^k = \underset{\substack{u_1<\cdots<u_k \\ u_1,\ldots,u_k \in J}}{\int\cdots\int} dX_{u_1} \otimes \cdots \otimes dX_{u_k} = (X_J^I)_{I=(i_1,\cdots,i_k)}, \forall k \geq 1.$

Let $S_k(X)_J$ denote the truncated signature of X of degree k, i.e.

$$S_k(X)_J = (1, \mathbf{X}_J^1, \ldots, \mathbf{X}_J^k). \tag{2}$$

The signature of a path has a geometric interpretation. The first level signature $\mathbf{X}_J^1$ is the increment of the path X, i.e. $X_T - X_S$, while the second level signature represents the signed area enclosed by the curve X and the cord connecting the start and end points of the path X.

The signature of X arises naturally as the basis function to represent the solution to linear controlled differential equation based on the Picard's iteration [5]. It plays the role of non-commutative monomials on the path space. In particular, if X is a one-dimensional path, the k^{th} level of the signature of X can be computed explicitly by induction for every $k \in \mathbb{N}$ as follows

$$\mathbf{X}_J^k = \frac{(X_T - X_S)^k}{k!}. \tag{3}$$

The signature of a d-dimensional linear path is given explicitly in the below lemma.

Lemma 1.1. *Let $X : [S,T] \to E$ be a linear path. Then*

$$S^n(X) = \frac{(X_T - X_S)^{\otimes n}}{n!}. \tag{4}$$

Equivalently speaking, for any multi-index $I = (i_1, \cdots, i_n)$,

$$S^I = \frac{\prod_{j=1}^n (X_T^{(i_j)})}{n!}. \tag{5}$$

The signature of a path can be interpreted as a solution to the controlled differential equation driven by a path in the tensor algebra space.

Theorem 1.1. *[6] Let $X \in BV([0,T],E)$. Define $f : T^{(n)} \to L(E, T^n(E))$ by*

$$f(a_0, a_1, \cdots, a_n)x = (0, a_0 \otimes x, a_1 \otimes x, \cdots, a_{n-1} \otimes x). \tag{6}$$

Then the unique solution to the differential equation

$$dS_t = f(S_t)dX_t, S_0 = (1, 0, \cdots, 0), \tag{7}$$

is the path $S : [0,T] \to T^{(n)}(E)$ defined for all $t \in [0,T]$ by

$$S_t = S_n(X_{[0,t]}) = (1, \mathbf{X}_{[0,t]}^1, \ldots, \mathbf{X}_{[0,t]}^n)).$$

Formally the signature of the path $X_{0,t}$ as a function from $[0,T]$ to $T((E))$ is the solution to the following equation

$$dS(X)_{[0,t]} = S(X_{[0,t]}) \otimes dX_t, S(X_{[0,0]}) = 1.$$

1.1 Multiplicative Property

The signature of paths of finite length (bounded variation) has the multiplicative property, also called Chen's identity.

Definition 1.2. *Let $X \in BV([0,s],E)$ and $Y \in BV([s,t],E)$ be two continuous paths. Their concatenation is the path denoted by $X * Y \in BV([0,t],E)$ defined by*

$$(X * Y)_u = \begin{cases} X_u, & u \in [0,s], \\ Y_u - Y_s + X_s, & u \in [s,t]. \end{cases}$$

Theorem 1.2 (Chen's identity)**.** . *Let $X \in BV([0,s],E)$ and $Y \in BV([s,t],E)$. Then*

$$S(X * Y) = S(X) \otimes S(Y).$$

Chen's identity asserts that the signature is a homomorphism between the path space and the signature space.

The multiplicative properties of the signature allows us to compute the truncated signature of a piecewise linear path.

Lemma 1.2. *Let X be a E-valued piecewise linear path, i.e. X is the concatenation of a finite number of linear paths, and in other words there exists a positive integer l and linear paths $X_1, X_2, \cdots, X_l$ such that $X = X_1 * X_2 * \cdots * X_l$. Then*

$$S(X) = \otimes_{i=1}^{l} \exp(X_i). \tag{8}$$

1.2 Uniqueness of the signature

Let us start with introducing the definition of the tree-like path.

Definition 1.3 (Tree-like Path)**.** *A path $X \in BV(J,E)$ is tree-like if there exists a continuous function $h : J \to [0,+\infty)$ such that $h(S) = h(T) = 0$ and such that, for all $s,t \in J$ with $s \leq t$,*

$$||X_t - X_s|| \leq h(s) + h(t) - 2 \inf_{u \in [s,t]} h(u).$$

Intuitively a tree-like path is a trajectory in which there is a section where the path exactly retraces itself. The tree-like equivalence is defined as follows: we say that two paths X and Y are the same up to the tree-like equivalence if and only if the concatenation of X and the inverse of Y is tree-like. Now we are ready to characterize the kernel of the signature transformation.

Theorem 1.3 (Uniqueness of the signature)**.** *Let $X \in BV(J,E)$. Then $S(X)$ determines X up to the tree-like equivalence defined in Definition 1.3.[1]*

Theorem 1.3 shows that the signature of the path can recover the path trajectory under a mild condition. The uniqueness of the signature is important, as it ensures it to be a discriminative feature set of un-parameterized streamed data.

Remark 1.1. *A simple sufficient condition for the uniqueness of the signature of a path of finite length is that one component of X is monotone. Thus the signature of the time-joint path determines its trajectory (see [4]).*

1.3 Invariance under time parameterization

Lemma 1.3 (Invariance under time parameterization)**.** *[6] Let $X \in V_1(J,E)$ and a path $\tilde{X} : J \to E$ be a time re-parameterization of X. Then*

$$S(X_J) = S(\tilde{X}_J). \tag{9}$$

Re-parameterizing a path inside the interval does not change its signature. As can be seen in Figure 1, speed changes result in different time series representation but the same signature feature. It means that signature feature can reduce dimension massively by removing the redundancy caused by the speed of traversing the path. It is very useful for applications where the output is invariant w.r.t. the speed of an input path, e.g. online handwritten character recognition and video classification.

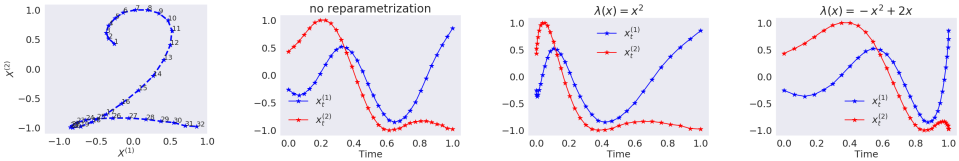

Figure 1: The first figure represents the trajectory of the digit 2, and the rest of figures plot the coordinates of the pen location via different speed respectively, which share the same signature and log signature given in the first subplot.

1.4 Shuffle Product Property

We introduce a special class of linear forms on $T((E))$; Suppose $(e^*_{n_1}, \cdots, e^*_{n_d}, \cdots)$ are elements of E^*. We can introduce coordinate iterated integrals by setting $X_u^{(i)} := \langle e_i^*, X_u \rangle$, and rewriting $\langle e^*_{i_1} \otimes \cdots \otimes e^*_{i_n}, S(X) \rangle$ as the scalar iterated integral of coordinate projection. In this way, we realize the n^{th} degree coordinate iterated integrals as the restrictions of linear functionals in $E^{\otimes n}$ to the space of signatures of paths. If $(e_1, \cdots, e_d)$ is a basis for a finite dimensional space E, and $(e^*_1, \cdots, e^*_d)$ is a basis for the dual E^* it therefore follows that

$$\mathbf{X}_J = \sum_{\substack{k \geq 0 \\ i_1, \cdots, i_k \\ \in 1,2,\cdots,d\}}} \int_{\substack{u_1 < \cdots < u_k \\ u_1, \ldots, u_k \in J}} \cdots \int dX_{u_1}^{(i_1)} \otimes \cdots \otimes dX_{u_k}^{(i_k)} e_{n_1} \otimes \cdots \otimes e_{n_k}.$$

Theorem 1.4 (Shuffle Algebra)**.** *The linear forms on $T((E))$ induced by $T(E^*)$, when restricted to the range $S(BV([0,T],E)$ of the signature, form an algebra of real valued functions of bounded variation.*

The proof can be found in page 35 in [6]. The proof is based on the Fubini theorem, and it is to show that for any $e^*, f^* \in T(E^*)$, such that for all $\mathbf{a} \in S(\mathcal{V}^p([0,T],E)$,

$$e^*(\mathbf{a}) f^*(\mathbf{a}) = (e^* \sqcup\!\sqcup f^*)(\mathbf{a}) \tag{10}$$

1.5 Universality of the signature

Any functional on the path can be rewritten as a function on the signature based on the uniqueness of the signature (Theorem 1.3). The signature of the path has the universality property, i.e. that any continuous functional on the signature can be well approximated by linear functionals on the signature (Theorem 1.5)[4].

Theorem 1.5 (Signature Approximation Theorem)**.** *Suppose $f : S_1 \to \mathbb{R}$ is a continuous function, where S_1 is a compact subset of $S(BV(J,E))$*[1]*. Then $\forall \varepsilon > 0$, there exists a linear functional $L \in T((E))^*$ such that*

$$\sup_{a \in S_1} ||f(a) - L(a)|| \leq \varepsilon. \tag{11}$$

Proof. It can be proved by the shuffle product property of the signature and the Stone-Weierstrass Theorem. □

[1] $S(BV(J,E))$ denotes the range of the signature for $x \in BV(J,E)$.

2 The log-signature of a path

Before introducing the log-signature, we define the Lie algebra, which the log-signature takes value in.

2.1 Lie algebra and Lie series

If F_1 and F_2 are two linear subspaces of $T((E))$, let us denote by $[F_1, F_2]$ the linear span of all the elements of the form $[a,b]$, where $a \in F_1$ and $b \in F_2$. Consider the sequence $(L_n)_{n\geq 0}$ of subspaces of $T((E))$ defined recursively as follows:

$$L_0 = 0; \forall n \geq 1, L_n = [E, L_{n-1}]. \tag{12}$$

Definition 2.1. *The space of Lie formal series over E, denoted as $\mathcal{L}((E))$ is defined as the following subspace of $T((E))$:*

$$\mathcal{L}((E)) = \{l = (l_0, \cdots, l_n, \cdots) | \forall n \geq 0, l_n \in L_n\}. \tag{13}$$

Theorem 2.1 (Theorem 2.23 [6])**.** *Let $X \in BV(J,E)$. Then the log-signature of X is a Lie series in $\mathcal{L}((E))$.*

2.2 The bijection between the signature and log-signature

Similar to the way of defining the logarithm of a tensor series, we have the exponential mapping of the element in $T((E))$ defined in a power series form.

Definition 2.2 (Exponential map)**.** *Let $a = (a_0, a_1, \cdots) \in T((E))$. Define the exponential map denoted by* exp *as follows:*

$$\exp(a) = \sum_{n=0}^{\infty} \frac{a^{\otimes n}}{n!}. \tag{14}$$

Lemma 2.1. *The inverse of the logarithm on the domain $\{a \in T((E)) | a_0 \neq 0\}$ is the exponential map.*

Theorem 2.2. *The dimension of the space of the truncated log signature of d-dimensional path up to degree n over d letters is given by:*

$$\mathcal{D}\mathcal{L}_n = \frac{1}{n} \sum_{d|n} \mu(d) q^{n|d}$$

where μ is the Mobius function, which maps n to

$$\begin{cases} 0, & \text{if } n \text{ has one or more repeated prime factors} \\ 1, & \text{if } n = 1 \\ (-1)^k & \text{if } n \text{ is the product of } k \text{ distinct prime numbers} \end{cases}$$

The proof can be found in Corollary 4.14 p. 96 of [8].

2.3 Calculation of the log-signature

Let's start with a linear path. The log signature of a linear path X_J is nothing else, but the increment of the path $X_T - X_S$.

The Baker-Cambpell-Hausdorff (BCH) formula gives a general method to compute the log-signature of the concatenation of two paths, which uses the multiplicativity of the signature and the free Lie algebra. It provides a way to compute the log-signature of a piecewise linear path by induction.

Theorem 2.3. *For any $S_1, S_2 \in \mathcal{L}((E))$*

$$Z = log(e^{S_1} e^{S_2}) = \sum_{\substack{n \geq 1 \\ p_1,\ldots,p_n \geq 0 \\ q_1,\ldots,q_n \geq 0 \\ p_i + q_i > 0}} \frac{(-1)^{n+1}}{n} \frac{1}{p_1! q_1! \ldots p_n! q_n!} r(S_1^{p_1} S_2^{q_1} \ldots S_1^{p_n} S_2^{q_n}) \tag{15}$$

where $r : A^* \to A^*$ is the right-Lie-bracketing operator, such that for any word $w = a_1 \ldots a_n$

$$r(w) = [a_1, \ldots, [a_{n-1}, a_n] \ldots].$$

This version of BCH is sometimes called the Dynkin's formula.

Proof. See remark of appendix 3.5.4 p. 81 in [?]. □

2.4 Uniqueness of the log-signature

Like the signature, the log-signature has the uniqueness stated in the following theorem.

Theorem 2.4 (Uniqueness of the log-signature)**.** *Let $X \in BV(J, E)$. Then $lS(X)$ determines X up to the tree-like equivalence defined in Definition 1.3.*

Theorem 2.4 shows that the signature of the path can recover the path trajectory under a mild condition.

Lemma 2.2. *A simple sufficient condition for the uniqueness of the log-signature of a path of finite length is that one component of X is monotone.*

3 Comparison of the Signature and Log-signature

Both the signature and log-signature take the functional view on discrete time series data, which allows a unified way to treat time series of variable length and missing data. For example, we chose one pen-digit data of length 53 and simulate 1000 samples of modified pen trajectories by dropping at most 16 points from it, to mimic the missing data of variable length case (See one example in Figure 2). Figure 3 shows that the mean absolute relative error of the signature and log-signature of the missing data is no more than 6%. Besides, the log-signature feature is more robust against missing data, and it is of lower dimension compared with the signature feature.

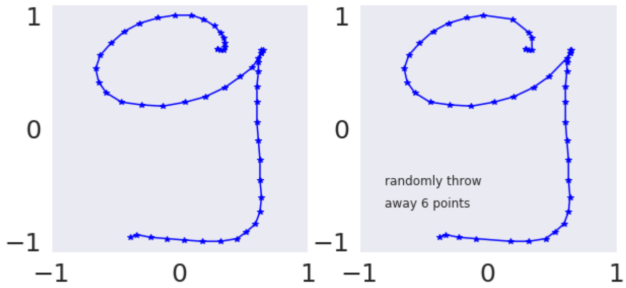

Figure 2: (Left) The chosen pen trajectory of digit 9. (Right) The simulated path by randomly dropping at most 16 points of the pen trajectory on the left.

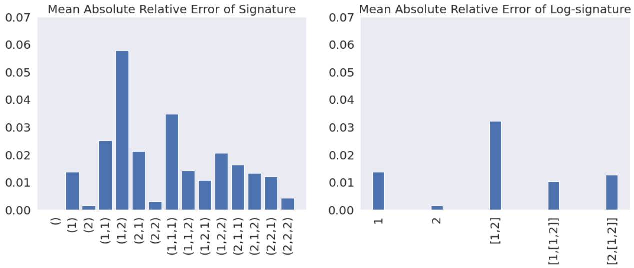

Figure 3: Signature and Log-Signature Comparison for the missing data case.

Figure 4: Signature and Log-Signature Comparison for the missing data case.

4 Backpropagation of the Log-Signature Layer

In this section, we provide the detailed derivation of backpropagation through time in our paper. Following the notations in Section 4 of our paper, $l_k := l_k^M$ denotes the log-signatures of a path $x^{\hat{\mathcal{D}}}$ over time partition $[u_k, u_{k+1}]$ of degree M. Let us consider the derivative of the scalar function F on $(l_k)_{k=1}^N$ with respect to path $x^{\hat{\mathcal{D}}}$, given the derivatives of F with respect to $(l_k)_{k=1}^N$. W.l.o.g assume that $\hat{\mathcal{D}} \subset \mathcal{D}$. By the Chain rule, it holds that

$$\frac{\partial F((l_1, \cdots, l_N))}{\partial x_{t_i}} = \sum_{k=1}^{N} \frac{\partial F(l_1, \cdots, l_N)}{\partial l_k} \frac{\partial l_k}{\partial x_{t_i}}. \tag{16}$$

where $k \in \{1, \cdots, N\}$ and $i \in \{0, 1, \cdots, n\}$.

If $t_i \notin [u_{k-1}, u_k]$, $\frac{\partial l_k}{\partial x_{t_i}} = 0$; otherwise $\frac{\partial l_k}{\partial x_{t_i}}$ is the derivative of the single log-signature l_k with respect to path x_{u_{k-1}, u_k} where $t_i \in \mathcal{D} \cap [u_{k-1}, u_k]$. The log signature $lS(x^{\hat{\mathcal{D}}})$ with respect to x_{t_i} is proved differentiable and the algorithm of computing the derivatives is given in [2], denoted by $\nabla_{x_{t_i}} LS(x^{\hat{\mathcal{D}}})$. This is a special case for our log-signature layer when $N = 1$. In general, for any $N \in \mathbb{Z}^+$, it holds that $\forall i \in \{0, 1, \cdots, n\}$ and $k \in \{1, \cdots, N\}$,[2]

$$\frac{\partial l_k}{\partial x_{t_i}} = \mathbf{1}_{t_i \in [u_{k-1}, u_k]} \nabla_{x_{t_i}} LS(x_{u_{k-1}, u_k}). \tag{17}$$

The backpropagation of the Log-Signature Layer can be implemented using Equation (16) and (17).

5 Other Path Transformation Layers

In this section, we introduce two other useful Path Transformation Layers which are often accompanied with the Embedding Layer in the PT-Logsig-RNN to improve the performance.

Accumulative Layer. Accumulative Layer (AL) maps the input sequence $(X_{t_i})_{i=1}^n$ to its partial sum sequence Y_{t_i}, where $Y_{t_i} = \sum_{j=1}^{i} X_{t_j}$, and $i = 1, \cdots, n$. One advantage of using the Accumulative Layer along with Log-Signature Layer is to extract the quadratic variation and other higher order statistics of an input path X effectively [7].

Time-incorporated Layer. Time-incorporated Layer (TL) is to add the time dimension to the input $(X_{t_i})_{i=1}^n$; in formula, the output is $(t_i, X_{t_i})_{i=1}^n$. As speed information is informative in most of SHAR tasks, e.g. distinguish running and walking, the log-signature of the time-incorporated transformation of a path can preserve such information.

6 Numerical Experiments

Network Architecture of PT-Logsig-RNN on Chalearn2013 data can be found in Table A.1.

Layer		Output shape	Description
Input		$(n, 20, 3)$	Input with n time steps
Embedding Layer	Conv2D	$(n, 20, 32)$	Kernel size=$1 \times 1 \times 32$
	Conv2D	$(n, 20, 16)$	Kernel size=$3 \times 1 \times 16$
	Flatten	$(n, 320)$	Flatten spatial dimensions
	Conv1D	$(n, 30)$	Kernel size=$1 \times 304 \times 30$
Accumulative Layer (AL)		$(n, 30)$	Partial sum on the series
Time-incorporated Layer (TL)		$(n, 31)$	Add temporal dimension
Logsig Layer		$(N, 496)$	$M = 2, N = 4$ after tuning
Add Starting Points		$(N, 527)$	Starting points of TL output
LSTM		$(N, 128)$	Return sequential output
Output		20	

Table A.1: Architecture of the PT-Logsig-RNN on Chalearn 2013

[2] Our implementation of Log-Signature layer in Tensorflow is implemented using the iisignature python package [3] and the customized layer in Keras. We implement the pytorch version of the Log-Signature layer using signatory package with some modification to handle time series of variable length.

References

[1] B.M. 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, 2010.

[2] Reizenstein Jeremy. *Iterated-Integral Signatures in Machine Learning*. PhD thesis, 2019.

[3] Reizenstein Jeremy and Graham Benjamin. The iisignature library: efficient calculation of iterated-integral signatures and log signatures. *arXiv preprint arXiv:1802.08252.*, 2018.

[4] Daniel Levin, Terry Lyons, and Hao Ni. Learning from the past, predicting the statistics for the future, learning an evolving system. *arXiv preprint arXiv:1309.0260*, 2013.

[5] Terry J Lyons. Differential equations driven by rough signals. *Revista Matemática Iberoamericana,*, 14(2):215–310, 1998.

[6] Terry J Lyons, Michael Caruana, and Thierry Lévy. *Differential equations driven by rough paths*. Springer, 2007.

[7] Hao Ni. A multi-dimensional stream and its signature representation. *arXiv preprint arXiv:1509.03346*, 2015.

[8] Christophe Reutenauer. Free lie algebras. In *Handbook of algebra*, volume 3, pages 887–903. Elsevier, 2003.