Deep Signature Transforms

A tool for the modelling of functions on streams

Patric Bonnier*, **Patrick Kidger***, **Imanol Perez Arribas***, **Cristopher Salvi***, **Terry Lyons** (* *Equal contribution.*)

`kidger@maths.ox.ac.uk` `arxiv.org/abs/1905.08494` `github.com/patrick-kidger/Deep-Signature-Transforms`

What is the signature transform?

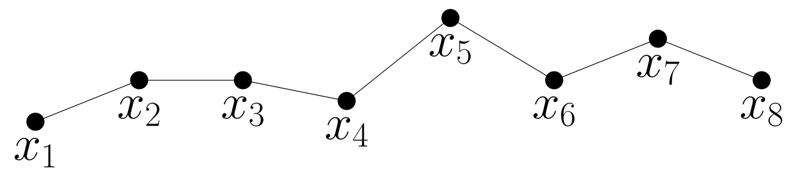

Figure 1: Embed a stream of data into path-space as a piecewise linear path.

Consider $\mathbf{x} = (x_1, \ldots, x_n)$ with $x_i \in \mathbb{R}^d$.
Linearly interpolate $\mathbf{x}$ to get a path f. Then its *signature transform of depth N* [CK16] is

$$\mathrm{Sig}^N(\mathbf{x}) = \left(\left(\int \cdots \int_{0<t_1<\cdots<t_k<1} \prod_{j=1}^{k} \frac{\mathrm{d} f_{i_j}}{\mathrm{d}t}(t_j) \mathrm{d}t_1 \cdots \mathrm{d}t_k \right)_{1 \leqslant i_1, \ldots, i_k \leqslant d} \right)_{1 \leqslant k \leqslant N}.$$

The signature transform is roughly analogous to the Fourier transform. Instead of frequency, it extracts information about *order* and *area*.

In fact, order and area precisely capture all possible nonlinear effects; the signature transform is a *universal nonlinearity*, and the following model exhibits universal approximation.

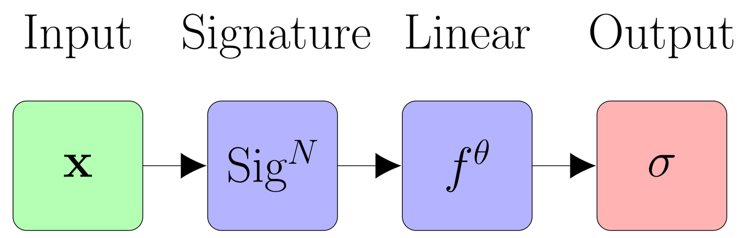

Figure 2: N is a hyperparameter. θ is learnt.

The signature transform has many other nice properties: robustness to missing or irregularly sampled data; optional translation invariance; optional sampling invariance; encoding of certain physical quantities; efficient data compression.

Context

The signature transform has previously been used for:

- Feature transformations [CK16]
- Kernels [KO19]
- Gaussian processes [TO19]
- Handwriting recognition [YJNL16]
- Action recognition [LZJ17]
- Pricing of financial derivatives [PA18]

But in neural networks, it has previously only been used as a fixed feature transformation.

Signature transforms in neural networks

The signature transform is actually differentiable, so it may be embedded in a neural network.

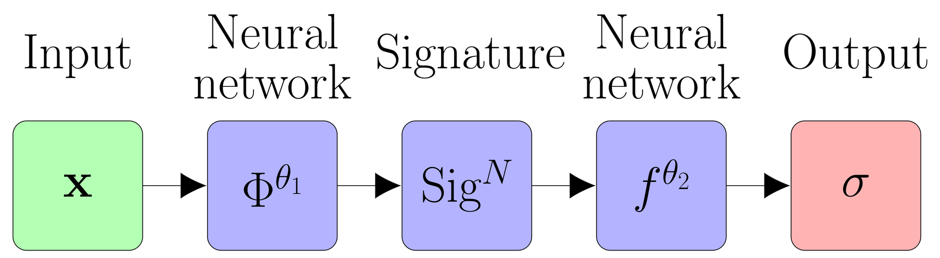

Figure 3

The only requirement is that Φ^{θ_1} must preserve the stream-like nature of the data (as the signature needs a stream-like input). This will be true if Φ^{θ_1} is a CNN or RNN, for example.

We then go on to introduce *lifts*, which lift a stream of data into a *stream of streams*. (These are analogous to windowing functions in signal processing.)
The signature transform is then applied to each sub-stream, to create a stream of signatures.

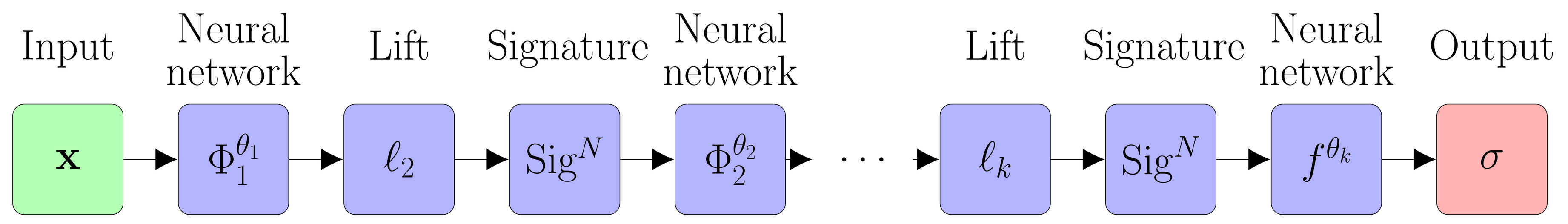

Figure 4: Deep signature model.

What this means in neural network terms:
The lift and signature transform together act as a mathematically motivated choice of pooling operation, known to extract meaningful information.

What this means in signal processing terms:
With lifts, the signature transform is analogous to the short time Fourier transform.

Example: fractional Brownian motion

Fractional Brownian motion (fBM) is a Gaussian process $B^H \colon [0, \infty) \to \mathbb{R}$. It depends upon its *Hurst parameter* $H \in (0, 1)$, which describes how rough it is.

We train several models to estimate the Hurst parameter from a realisation of fBM.

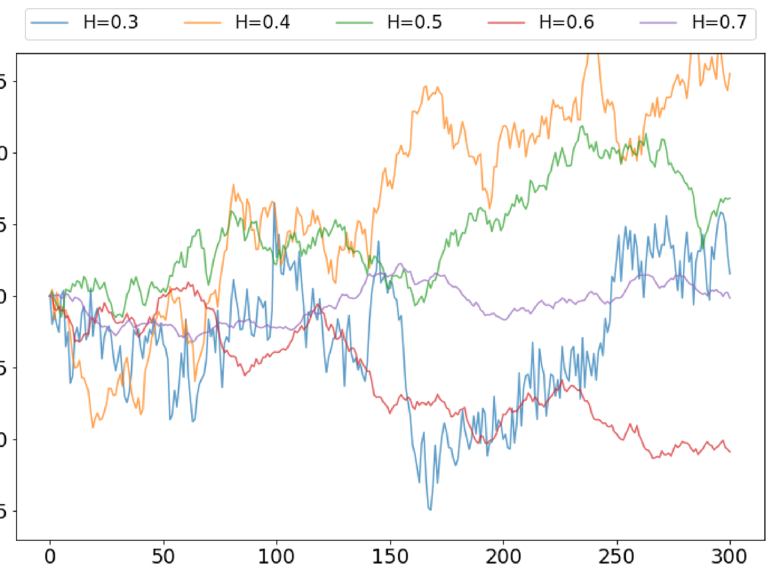

Figure 5: Examples of fBM

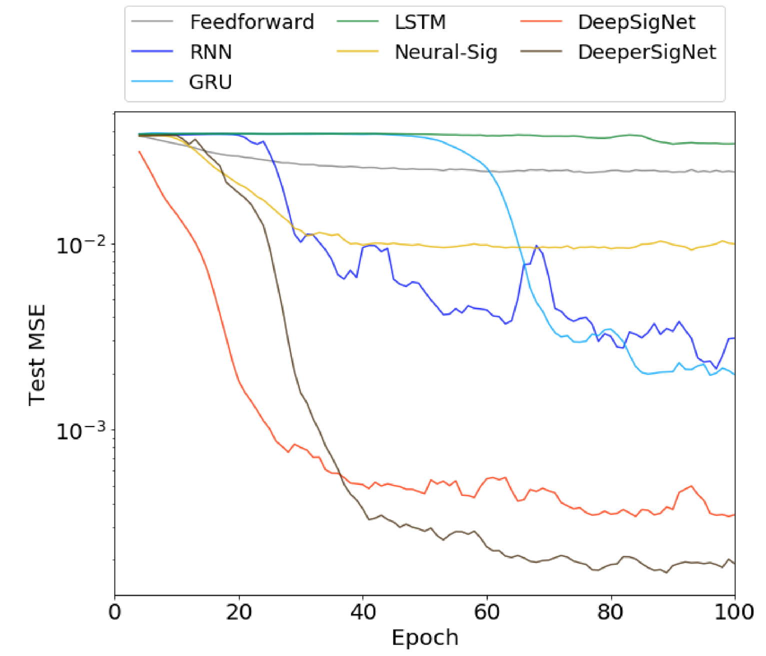

Figure 6: Typical validation loss during training

References

[CK16] I. Chevyrev and A. Kormilitzin. A primer on the signature method in machine learning. *arXiv:1603.03788*, 2016.

[KO19] Franz J Király and Harald Oberhauser. Kernels for sequentially ordered data. *Journal of Machine Learning Research*, 2019.

[LZJ17] Chenyang Li, Xin Zhang, and Lianwen Jin. LPSNet: a novel log path signature feature based hand gesture recognition framework. In *Proceedings of the IEEE International Conference on Computer Vision*, pages 631–639, 2017.

[PA18] Imanol Perez Arribas. Derivatives pricing using signature payoffs. *arXiv:1809.09466*, 2018.

[TO19] Csaba Toth and Harald Oberhauser. Variational gaussian processes with signature covariances. *arXiv:1906.08215*, 2019.

[YJNL16] Weixin Yang, Lianwen Jin, Hao Ni, and Terry Lyons. Rotation-free online handwritten character recognition using dyadic path signature features, hanging normalization, and deep neural network. In *2016 23rd International Conference on Pattern Recognition (ICPR)*, pages 4083–4088. IEEE, 2016.

Work supported by EPSRC grants EP/R513295/1, EP/L015811/1, EP/N510129/1.