Signatures of BV Paths and its Application to Online Handwriting Recognition

Yangxi Ou
Department of Mathematical Sciences
Carnegie Mellon University

MATH 820 Advanced Topics in Analysis: Variational and PDE Techniques in Data Analysis
Center for ~~Non~~linear Analysis Seminar Room, December 14th, 2016

This report is produced per the course requirement of MATH 820 in Fall 2016 offered by Professo Dejan Slepcev. Its main goal is to introduce to fellow students the signature (sometimes called tensorial exponential) of a continuous path of bounded variations, and its application to analysis of data streams (i.e. one-parameter representations in a state space), with writer-independent online handwriting digit recognition as a prototype of concepts. With an emphasis on the big picture, the report is not intended to be rigorous or comprehensive, and in particular, theorems and propositions are not proved. Interested readers can refer to the citations for more details. The author claims no expertise in either signatures of paths (the abstract algebraic leg of rough path theory) or data analysis (especially handwriting recognition), but is responsible for any mistake not due to the references.

Contents

1 The Signature of a Path, Tensor Algebra, and Linearity

Description and prediction of dynamical systems such as weather forecasting, speech recognition, and financial investment are fundamental, intriguing, and difficult problems in both daily life and research, which have been studied by the most brilliant minds of human beings since as late as the advents of calculus and classical mechanics. A classical and currently dominant mathematical tool for such problems is ordinary differential equations (ODE), in which the infinitesimal change of the system with respect to time is described as a time-dependent function of the current state of the system and external forcing. However, ODE theory often requires suitable regularities of the system and forcing to produce meaningful outputs, and hence is limited in dealing with highly oscillatory and complex systems, for example systems of merely bounded variations, or signals corrupted by noises. A more general framework, called rough path theory, has recently been advocated by Lyons to address the lack of (classical) regularities in modeling evolving systems. According to Lyons and Geng, rough path theory has proven successful ranging from commercial applications such as Chinese handwriting recognition by Graham to theoretical developments such as regularity structures and stochastic partial differential equations (SPDE) by the Fields medalist Hairer. In this section, we shall sample taste the abstract algebraic leg of rough path theory, namely the signature of a path, or sometimes called the (tensorial) exponential map. Later in the next section, we shall experiment with handwriting data and apply the theory of signature to a mini-task of online handwriting digit recognition, in contrast to Graham. We conclude this report with general remarks of signatures as extracted features in analyzing data streams.

Let's start with the mathematical formulation of evolving systems, paths, and data streams, as they appear in the introductory paragraph above.

Definition 1.1 (Path)**.** Let E be a set and $I \subseteq \mathbb{R}$ be an interval. A path in E is a function $X : I \to E$. If E has additional mathematical structures such as σ-algebra, topology, or metric, it is usual to restrict X to be a homomorphism in some corresponding category. In natural science and engineering, it is common to refer X as an evolving system with state space E; in social science and data analysis, X is often called an E-valued data stream.

In this report, we consider $E = \mathbb{R}^d$ for some $d \in \mathbb{N}_+$, $I = [a, b]$ a compact interval, and X being continuous and of bounded variations, without losing much generality in applications yet keeping the introduction accessible.

Next, we give the definition of the signature $S(X)$ of the path $X : I \to E$ along with its basic properties, and demonstrate its significance and advantages in data science thanks to abstract algebra and category theory.

1.1 Tensor Algebra, Definition and Motivation of Signatures

We need some basics of tensor algebra in order to define signature (as suggested by the alternate name tensorial exponential map).

Definition 1.2. For $n \in \mathbb{N}$, let $T^{(n)}(E) := \bigoplus_{k=0}^{n} E^{\otimes k} := \left\{a = (a_0, a_1, \ldots, a_n) \,\middle|\, \forall k \leq n, a_k \in E^{\otimes k}\right\}$, i.e. the space of non-commutative polynomials of E-tensors of degree at most n, where $E^{\otimes 0} := \mathbb{R}$, the underlying scalar field by convention. Similarly, let $T^{(<\omega)}(E) := \bigoplus_{n=0}^{<\omega} E^{\otimes n} := \bigoplus_{n=0}^{\infty} E^{\otimes n} := \bigcup_{n\in\mathbb{N}} \bigoplus_{k=0}^{n} E^{\otimes k}$, i.e. the space of non-commutative polynomials of E-tensors. Finally, let $T^{(\omega)}(E) := \bigoplus_{n=0}^{\omega} E^{\otimes n} := \left\{a = (a_0, a_1, \ldots, a_n, \ldots) \,\middle|\, \forall n \in \mathbb{N}, a_n \in E^{\otimes n}\right\}$, i.e. the space of non-commutative formal power series of E-tensors. All of $T^{(n)}(E)$, $T^{(<\omega)}(E)$, and $T^{(\omega)}(E)$ are non-commutative associative unital algebra over $\mathbb{R}$, with addition and scalar multiplication given componentwise, multiplication given by polynomial multiplication or sequence convolution (truncated at degree n for $T^{(n)}(E)$), the zero being sequence of graded zero tensors, and the unit being $1 = (1, 0, 0, \ldots)$, scalar 1 followed by graded zero tensors. Denote the truncation at degree n by $\pi_n : T^{(m)}(E) \to T^{(n)}(E)$, which is a homomorphism, where $n \leq m \in (\omega+1) = \mathbb{N} \cup \{\omega\}$ or m is $< \omega$. Note that the inclusion $T^{(<\omega)}(E) \hookrightarrow T^{(\omega)}(E)$ is also homomorphic, but it is not true if $< \omega$ is replaced by $n \in \mathbb{N}$.

For $n \in (\omega + 1)$, $a \in T^{(n)}(E)$ is invertible if and only if $a_0 \neq 0$. Let $\tilde{T}^{(n)}(E) := \left\{a \in T^{(n)}(E) \,\middle|\, a_0 = 1\right\}$, the (special) group of elements with scalar term 1. Analogously, $\tilde{T}^{(<\omega)}(E)$ can be defined but it is not a group.

Now, we are ready to give the definition of signature and its approximation in applications, truncated signature.

Definition 1.3. Let $X : I = [a, b] \to E$ be a continuous path of bounded variations. Define the signature or the tensorial exponential $S(X) \in T^{(\omega)}(E)$ of X by $S(X)_n := \int_{a<t_1<\cdots<t_n<b} \bigotimes_{i=1}^{n} \mathrm{d}X_{t_i} \in E^{\otimes n}, \forall n \in \mathbb{N}$, where $S(X)_0 := 1$ by convention. For $n \in \mathbb{N}$, the truncated signature is then given by $S^{(n)} := \pi_n \circ S$.

One might wonder where the definition of signature via iterated integrals comes from. Indeed, the idea arises naturally in solving a linear ODE theoretically and numerically, as demonstrated below. The example of linear ODE somehow justifies the name exponential map as well (a more sophisticated connection is to Lie theory). Also, it is helpful to compare signature with Taylor series where iterated partial derivatives are taken locally around an expansion center a to obtain a (formal) power series (converging to the original function f in some open neighbourhood $U \ni a$ if and only if $f \in C^\omega$, i.e. analytic).

Consider a linear differential equations driven by $X : I \to E$

1.2 Signature as Group Homomorphism, Tree-like Path and Faithful Representation

1.3 Shuffle Product Formula: Polynomial is Linear

2 Application of Signature to Online Handwriting Recognition

2.1 Handwriting Recognition: Online VS Offline

2.2 Signature as Extracted Features: Apriori Analysis

2.3 Experiment: Recognizing Digits using Signature and Linear Regression

2.3.1 Data Source and Description

2.3.2 Program Design and Work Flow

2.3.3 Results

3 Summary

References

I will include references including data sources later.