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 $\otimes$ 16th, 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 Tensor Algebra, the Signature of a Path, 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 [8] and Geng [1], 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 3. 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 [3]. We conclude this report with general remarks of signatures as extracted features in analyzing data streams.

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

Definition 1.1 (Path)**.** Let E be a set and I be a linearly ordered set. A path in E is a function $X : I \to E$. If E and I have 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 all these cases, I is interpreted as time.

In this report, we consider $E = \mathbb{R}^d$ for some $d \in \mathbb{N}_+$, $I = [a, b] \subset \mathbb{R}$ a compact interval, and X being continuous and of bounded variations, i.e. $X \in C \cap BV(I; E)$[1], 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 (Tensor Algebra, Formal Series)**.** 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, formally $\sum_{k=0}^{n} a_k x^{\otimes k}$ ($x^{\otimes k}$ is a placeholder for the k-th component), where $E^{\otimes 0} := \mathbb{R}$, the underlying scalar field by convention. Similarly, let $T^{(<\omega)}(E) := \bigoplus_{n=0}^{<\omega} E^{\otimes n} := \bigoplus_{n=0}^{\infty} E^{\otimes n} :=$

[1]The technical reason why we work in this space will be clear after seeing the definition of signature, Definition 1.3.

$\bigcup_{n\in\mathbb{N}} \bigoplus_{k=0}^{n} E^{\otimes k}$, i.e. the space of non-commutative polynomials of E-tensors, formally the same as above, with varying $n \in \mathbb{N}$. Finally, let $T^{(\omega)}(E) := \bigoplus_{n=0}^{\omega} E^{\otimes n} := \left\{a = (a_0, a_1, \dots, a_n, \dots) \,\middle|\, \forall n \in \mathbb{N}, a_n \in E^{\otimes n}\right\}$, i.e. the space of non-commutative formal power series of E-tensors, formally $\sum_{n=0}^{\infty} a_n x^{\otimes n}$. 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 $1 = (1, 0, 0, \dots)$, 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 $n \le m \in (\omega + 1) = \mathbb{N} \cup \{\omega\}$ 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.

The algebraic properties of various spaces of tensors defined above will be further discussed in the next subsection. They will play a crucial role in studying paths, and in our case, justifying why signature is a reasonable candidate feature for handwriting recognition. Now, we are ready to give the definition of signature and truncated signature, the finite dimensional approximation in applications.

Definition 1.3 (Signature or Tensorial Exponential)**.** 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. In the notation of formal series, we may write $S(X) = 1 + \sum_{n=1}^{\infty} \int_{a<t_1<\cdots<t_n<b} \prod_{k=1}^{n} \mathrm{d}\langle X_{t_i}, x\rangle$. For $n \in \mathbb{N}$, the n-th truncated signature is given by $S^{(n)} := \pi_n \circ S$. The transform S ($S^{(n)}$) is called the (truncated) exponential map.

We remark that the iterated integral is componentwise well-defined, since $X \in C \cap BV(I; E)$, where $BV(I; E)$ is the topological dual of $C(I; E)$. Although a more general setup in rough path theory via Young's integral is possible, we stick with the better-understood space $C \cap BV$, in which more results are available[2].

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). We refer interested readers to Appendix A for a more elaborate discussion. After we explain the shuffle product formula at the end of this section, we shall provide another heuristic motivation to the definition of signature, with close connection to statistics, in particular linear regression.

Consider a linear differential equation driven by $X : I = [0, T] \to E$, that is $\mathrm{d}Y_t = A(\mathrm{d}X_t, Y_t), Y_0 = y_0 \in E$, where $A \in \mathcal{L}(E \times E; E)$, or a $(1, 2)$-E-tensor. It is well-known that the solution to this linear ODE uniquely exists: one applies the Picard-Lindelof iteration and shows its (eventual) contractivity. The iteration is actually a practical numerical scheme to compute the solution, starting with

[2] Young's integral is defined for continuous integrators of finite p-variations and continuous integrands of fintie q-variations, where $p, q \in [1, \infty)$ and $1/p + 1/q > 1$. One undesirable consequence of Young's integral is that Fubini's theorem does not hold when $p > 1$.

$Y_\cdot^{(0)} :\equiv y_0$, and inductively integrating $Y_0^{(n)} + \int_0^\cdot A(\mathrm{d}X_t, Y_t^{(n)}) =: Y_\cdot^{(n+1)}$. The closed form expression of the n-th approximation is

$$Y_\cdot^{(n)} = y_0 + \sum_{k=1}^{n} \int_0^\cdot A\left(\mathrm{d}X_{t_k}, \int_0^{t_k} A\left(\mathrm{d}X_{t_{k-1}}, \cdots, \int_0^{t_2} A(\mathrm{d}X_{t_1}, y_0)\right)\right) = y_0 + \sum_{k=1}^{n}\left(A\int_0^\cdot \mathrm{d}X_{t_k}\right)\left(A\int_0^{t_k} \mathrm{d}X_{t_{k-1}}\right)\cdots\left(A\int_0^{t_2} \mathrm{d}X_{t_1}\right) y_0$$
$$= y_0 + \sum_{k=1}^{n} A^{\otimes k}\left(\left(\bigotimes_{i=1}^{k-1}\int_0^{t_{i+1}} \mathrm{d}X_{t_i}\right)\otimes\int_0^\cdot \mathrm{d}X_{t_k}\right) y_0 = \left(I + \sum_{k=1}^{n} A^{\otimes k}\int_{0<t_1<\cdots<t_k<\cdot}\bigotimes_{i=1}^{k} \mathrm{d}X_{t_i}\right) y_0.$$

Iterated integrals appear inductively in the approximate solutions to the linear ODE, and observe that the operator $y_0 \mapsto Y_t^{(n)}$ is nothing but evaluating the formal tensorial polynomial $S^{(n)}(X|_{[0,t]})$ at the tensor A. As the iteration scheme converges to the solution, we have that $Y_t = \left(I + \sum_{k=1}^\infty A^{\otimes k}\int_{0<t_1<\cdots<t_k<t}\bigotimes_{i=1}^k \mathrm{d}X_{t_i}\right) y_0 = S(X|_{[0,t]})(A)(y_0)$, more than just an artificial formal construct in algebra. Moreover, as in the proof of Picard-Lindelof, one can show factorial decay of the summands, i.e. $\forall k \in \mathbb{N}_+$, $\left|\int_{0<t_1<\cdots<t_k<T}\bigotimes_{i=1}^k \mathrm{d}X_{t_i}\right| \le \frac{1}{k!}\|X\|_{BV([0,T];E)}^k$. The rate of convergence is exactly that in the MacLaurin series of exponential function. In the special case $d = 1$ (i.e.$E = \mathbb{R}$), the series of iterated integrals reduces to $Y_\cdot = \exp\left(A\int_0^\cdot \mathrm{d}X_t\right) y_0 = \exp(A(X_\cdot - X_0))y_0$ as usual. Note that it is not required that Y_t lives in the same space as X_t does, as long as A acts properly. If $X_t \in \mathbb{R}$ and $Y_t \in E$, then A can be represented as a square matrix and the solution to the ODE is still as above with matrix exponential.

One might find the tensorial algebraic exercise an overkill, but it turns out useful and (sometimes surprisingly) efficient in numerical computations. Suppose that we have to solve the ODE for many different infinitesimal generators[3] A and initial conditions y_0 (the initial conditions don't have to be in the same space as aforementioned!) but a fixed driving path X. Rather than performing Picard-Lindelof as above for every pair (A, y_0), we first compute and store the (truncated) signature of X independently, then evaluate the polynomial (the degree of the polynomial depends on the norm of A and tolerance for error) at different tensor A, which could be implemented efficiently. Essentially we are treating the path X as a constant but the pair of generator and initial condition (A, y_0) as the only variable here, a technique known as currying in computer science. Tensor algebra simply realizes this idea mathematically, separating (A, y_0) from X in the seemingly clumsy iteration. In this sense, the signature $S(X)$ summarizes all effects of linear operators on the path X, a fact comparable to Taylor series representation of an analytic function, and Hausdorff Moment Problem, whose solution confirms that the moments of a compactly supported random variable uniquely determines its probability distribution. It is now clear from the ODE example why iterated integrals are worth studying and why signature is also called the exponential map.

In the following subsection, we shall see that signature indeed contains "more" information: it is a faithful representation of the path up to a "small" set of "null" paths.

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

The goal of this subsection is to understand nice properties of signatures, culminating at the theorem in Hambly and Lyons [4] that the signature $S(X) \in T^{(\omega)}(E)$ is essentially a faithful representation of the path $X \in C \cap BV(I; E)$. Let's start from the basics.

[3]The solution to this ODE is obviously a time-homogeneous flow, and Lie theory kicks in its study. Signature then coincides with the exponential map in Lie group/algebra. In differential geometry, the ODE represents development of one manifold over another, in other words, rolling one surface onto another. Indeed, geometry is the primary origination of signature, in K.T. Chen's thesis around 60 years ago, per Hambly and Lyons [4].

Proposition 1.4 (Invariance under Translation and Time Change)**.** *Let $X \in C \cap BV(I;E)$ be a continuous path of bounded variations, $y \in E$ a vector, and $\tau : [a,b] =: I \to J := [c,d]$ a continuous and increasing function, then $X + y$ and $X_{\tau(\cdot)}$ are also continuous paths in E of bounded variations with the same signature as X, that is $S(X+y) = S(X) = S\left(X_{\tau(\cdot)}\right)$.*

The vector y is interpreted as translation while the function τ can be be viewed as time change. The invariance properties follow directly from those of (iterated) integrals. They will prove desirable in our mini-task of online handwriting digit recognition. Due to invariance, we sometimes abuse notation and identify paths differed only by a translation or a reparametrization[4]. Without loss of generality, we may always assume $I = [0,1]$, the unit compact interval, whenever convenient. We denote the path space $C \cap BV(I;E)$ under such identification by $Path(E)$.

The next thing to consider naturally is algebraic structure on $Path(E)$ and $T^{(\omega)}(E)$ that signature preserves. In contrast to Taylor series, the exponential map is not a linear operator, and hence vector space (and therefore algebra) is not the structure we are hunting for. Recall that $S(X)_0 \equiv 1$, $\forall X \in Path(E)$, which always lives in the group $\tilde{T}^{(\omega)}(E)$. So one should expect S being a (non-commutative) group homomorphism to $\tilde{T}^{(\omega)}(E)$. But what should be the corresponding group structure on $Path(E)$? Of course, we can always abstractly pullback the group structure $\left(\tilde{T}^{(\omega)}(E), \otimes, 1\right)$ to the set $Path(E)$, but it is more straightforward to guess (part of) the group structure: the multiplication is given by concatenation of paths, as in fundamental groups.

Definition 1.5 (Concatenation and Identity Path)**.** Let $X, Y \in Path(E)$ with respective representatives $X, Y \in C \cap BV(I;E)$ be two continuous paths of bounded variations, where $I := [0,1]$. Define the concatenated path $X * Y \in Path(E)$ via the representative in $C \cap BV(I;E)$ given by $(X * Y)_t := (X_{2t} - X_0)\mathbb{I}_{[0,1/2]}(t) + (Y_{2t-1} - Y_0)\mathbb{I}_{[1/2,1]}(t)$. Concatenation is well-defined, that is independent of the choice of representatives. It is associative (but not commutative) and hence makes the path space $Path(E)$ a semigroup, which is futhermore a non-commutative monoid, with the identity being the constant path $\mathrm{id} :\equiv 0$.

Proposition 1.6 (Signature as Monoid Homomorphism)**.** $S : (Path(E), *, \mathrm{id}) \to \left(\tilde{T}^{(\omega)}(E), \otimes, 1\right)$ *is a monoid homomorphism, that is* $\forall X, Y \in Path(E)$*,* $S(X * Y) = S(X) \otimes S(Y)$ *and* $S(\mathrm{id}) = 1$*.*

Preservation of identities is immediate as nontrivial integrals are zeros, while preservation of multiplications can be proved by partitioning the unit interval I into two halves and applying Fubini's theorem[5]. It seems that we are stating the obvious, but let's remark that this property allows us to compute the signature of a long path by "appending" signatures of many short ones, opening the door to optimize algorithms to large scale calculations or to perform real-time tasks as data streams feed in around the clock.

Now that $Path(E)$ is a monoid, can we push it to be a well-defined group by specifying the inversion? A first answer is no: the identity path is characterized by a singleton image, i.e. $X = \mathrm{id} \Leftrightarrow \#X(I) = 1$, and concatenation only enlarges images of paths. So other than the trivial case, X admits no inverse in $(Path(E), *, \mathrm{id})$, and we must impose extra structure on $Path(E)$ to make it a group. Mimicking the approach in fundamental groups, in which path reversal is the inversion, we shall identify $X * \overleftarrow{X}$ with id in $Path(E)$ and consider the equivalence relation "generated" by this identification, where $\overleftarrow{X}$ is the reversed path of X.

[4] Note that we do not identify paths with the same signature, at least have not yet. The "identification" is the content of the theorem on faithful representation by Hambly and Lyons.

[5] This is another technical reason why we restrict the paths to be continuous and of bounded variations.2

Definition 1.7 (Reversal and Tree-like Path). Let $X \in Path(E)$ with representative $X \in C \cap BV(I = [0,1]; E)$ be a continuous path of bounded variations. Define $\overleftarrow{X} \in Path(E)$, the reversed path of X, via the representative $\overleftarrow{X} := X_{1-\cdot} \in C \cap BV(I; E)$. Reversal is well-defined, independent of representatives. Observe that double inversion is the identity function, i.e. $\overleftarrow{\overleftarrow{X}} = X$, and that inversion (reversely) respects concatenation, i.e. $\overleftarrow{Y} * \overleftarrow{X} = \overleftarrow{X * Y}$, where $Y \in Path(E)$.

X is called a tree-like path, if there exists a height function $h : [0,1] \to [0,\infty)$ such that $h(0) = 0 = h(1)$ and $\forall 0 \le s \le t \le 1$, $\|X_t - X_s\| \le h(s) + h(t) - 2 \inf h([s,t])$. Note that $X_0 = X_1$ is enforced in the definition of a tree-like path, and the height function is like a "group" version of the arc length function[6] of X. Indeed, the path $Y * X * \overleftarrow{Y}$ is tree-like provided X is, so is id trivially. Tree-like paths also form a sub-monoid of $Path(E)$ stable under reversal. Thus, the relation $X \sim Y$ defined by $X * \overleftarrow{Y}$ being tree-like is an equivalence relation preserved by concatenation and reversal, and $\left(RPath(E) := Path(E)/\sim, *, \mathrm{id}, \overleftarrow{\cdot}\right)$ is therefore a group.

Proposition 1.8 (Signature as Group Homomorphism). $S : \left(RPath(E), *, \mathrm{id}, \overleftarrow{\cdot}\right) \to \left(\tilde{T}^{(\omega)}(E), \otimes, 1, \cdot^{-1}\right)$ *is a group homomorphism, that is* $\forall X, Y \in RPath(E)$, $S(X * Y) = S(X) \otimes S(Y)$, $S(\mathrm{id}) = 1$, *and* $S\left(\overleftarrow{X}\right) = S(X)^{-1}$.

A direct computation with some ODE theory shows that $S(X) \otimes S\left(\overleftarrow{X}\right) = 1$. One completes the proof by showing that signature respects tree-like paths, which is more technical. Interested readers could consult Hambly and Lyons [4] for the proofs of this fact and the following celebrated theorem, which says that exponential map is a lossless compression of all essential information of paths.

Theorem 1.9 (Signature as Faithful Representation). $S : \left(RPath(E), *, \mathrm{id}, \overleftarrow{\cdot}\right) \to \left(\tilde{T}^{(\omega)}(E), \otimes, 1, \cdot^{-1}\right)$ *is injective, i.e. the kernel of* $S : Path(E) \to \tilde{T}^{(\omega)}$ *is the set of tree-like paths. Moreover, among all paths in* $Path(E)$ *sharing the same signature, there is a unique one with the minimal arc length.*

We do warn readers that S is far from surjective, and the image of S can be further restricted to a proper Lie subgroup following the results in the next subsection and some classic Lie theory, though we are not going to pursue in this direction[7]. According to Geng [1], a characterization of the image of S is still unknown, even though the truncated counterpart is well understood.

At the end of this subsection, we emphasize again that the discrete sequence of signature encodes almost all information of a path in a graded fashion with increasing granularity, just like the (coefficients of) Taylor series to an analytic function, the moment sequence to a compactly supported random variable, or the spectra to a periodic function. Whence, studying paths with signatures is theoretically justified, and this is one of the two bedrocks in our mini-task in Section 2.

1.3 Shuffle Product Formula: Polynomial is Linear

So far we have been considering a path or its signature itself. In this section we shift gears to study

[6] One more technical reason why we restrict the paths to be continuous and of bounded variations.2

[7] In this case, one will dance among multiple algebras, including a Lie algebra of polynomials and series in which tensor product $\otimes$ is replaced with Lie bracket $[,]$. Exponential map on the Lie algebra will be distinguished from S and utilized to bound the set of signatures.

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 Functionals

2.3.1 Data Source and Description

2.3.2 Program Design and Work Flow

2.3.3 Results

3 Summary

Appendix A Signature and Taylor Series in Tensor Algebra

appendix means to be appended (later): "dual" relation between Taylor series and signature

Appendix B MATLAB Implementation of Signature Computation

appendix means to be appended (later): *cumsum* and *kron*

Acknowledgments

The author thanks Dr. Geng for introducing the fascinating and rapidly growing field of rough path theory in his presentation [1]. This is the first time that the author is aware of the theory of rough path, and in particular signatures. Dr. Geng also points the author to the compact references Lyons et al. [10, 9].

The author thanks Professor Slepcev for assigning the final project in his Fall 2016 offering of the course MATH 820 Advanced Topics in Analysis: Variational and PDE Techniques in Data Analysis, and teaching the author geometry in his Spring 2016 offering of the course MATH 759 Differential Geometry. It is the pressure from the course to actually "do something" that drives the author to adventure some extra miles into his unknown territory, but if not equipped with basic tools from geometry, the author might not survive in this foreign wonderland filled with fancy abstract nonsense. (Clive Newstead also helps indirectly by advertising the course PHIL 713 Category Theory.)

The author would also like to thank Antoine Remond-Tiedrez for informing of the powerful public computing resources on campus of Carnegie Mellon University, which make animated visualization and intensive model validations (and future to-do optimizations) much easier for the author (with a shitty 7-year-old 32-bit machine running MATLAB R2009a). The comment of “good space, bad objects” versus “bad space, good objects” comes from Antoine as well.

Bibliographic Notes

Section 1 is adapted from Chapter 2 in Lyons et al. [10], originating from a lecture note given at the (annual) Saint-Flour Probability Summer School in July 2004. The majority of the chapter is summarized in this report, with the exception of materials relating to Lie theory (and proofs!). Geng's presentation [1], mainly on the work of Hambly and Lyons [4], is also blended into this section. The lecture note by Lyons et al. [10] is rigorous enough whenever proofs are included, while the paper Hambly and Lyons [4] is a more serious treatment, and itself the proof of Theorem 1.9 (Signature as Faithful Representation). For more intuitions and a broader view, one can refer to the survey article Lyons [8] written for the International Congress of Mathematicians 2014, Korea, in which Hairer received his Fields medal .

Section 2 reports the design and findings of the mini-task of online handwriting digit recognition that the author conducts. The mini-task is inspired by a comment in Geng's presentation [1] that Chinese handwriting recognition based on signatures has been or is being commercialized. The paper by Graham [3] is one of the rare English literature that documents signature-based handwriting recognition. The author also casually learns the jargon and some practices of handwriting recognition from researchers' webpages, such as LeCun et al. [5], MATLAB [11, 12]. The raw data is downloaded from University of California Irvine (UCI) Machine Learning Repository [6, 7]. Although MNIST [5] is considered a benchmark dataset in handwriting recognition, it is offline and does not suit the situations here. Online data are fewer and harder to find than offline data. The author is responsible for the majority of contents in Section 2 and the MATLAB codes used, although MATLAB [11, 12] help files and support webpages are heavily referenced. For other analyses of data streams, references include the papers Lyons [8], Lyons et al. [9], Gergely Gyurkó et al. [2], where the latter two concentrate on applications to financial time series (which the author has better knowledge of and more experience with).

The Taylor series part of Appendix A is standard material in any undergraduate mathematical curriculum. The author rewrites it in the (alien) tensorial notations to make a point in formality. The author has not yet encountered or been aware of literature in such notations to the best of his knowledge, and finds it overkill in the statement and usual applications of Taylor's theorem.

Appendix B is included for completeness and to support the claim that signature can be computed efficiently, albeit a seemingly scary monster in mathematics and notations.

References

[1] Xi Geng. The exponential transform of a path: a faithful representation. Center for Computational Finance (CCF) Seminar, 10 2016. URL http://math.cmu.edu/CCF/ccfseminar.php?SeminarSelect=1331.

[2] L. Gergely Gyurkó, T. Lyons, M. Kontkowski, and J. Field. Extracting information from the signature of a financial data stream. *ArXiv e-prints*, July 2013.

[3] Benjamin Graham. Sparse arrays of signatures for online character recognition. *CoRR*, abs/1308.0371, 2013. URL http://arxiv.org/abs/1308.0371.

[4] B. Hambly and T. Lyons. Uniqueness for the signature of a path of bounded variation and the reduced path group. *ArXiv Mathematics e-prints*, July 2005.

[5] Yann LeCun, Corinna Cortes, and Christopher J.C. Burges. The MNIST database of handwritten digits. URL http://yann.lecun.com/exdb/mnist/index.html.

[6] M. Lichman. UCI machine learning repository uji pen characters data set, 2013. URL http://archive.ics.uci.edu/ml/datasets/UJI+Pen+Characters.

[7] M. Lichman. UCI machine learning repository pen-based recognition of handwritten digits data set, 2013. URL http://archive.ics.uci.edu/ml/datasets/Pen-Based+Recognition+of+Handwritten+Digits.

[8] T. Lyons. Rough paths, Signatures and the modelling of functions on streams. *ArXiv e-prints*, May 2014.

[9] Terry Lyons, Hao Ni, and Harald Oberhauser. A feature set for streams and an application to high-frequency financial tick data. In *Proceedings of the 2014 International Conference on Big Data Science and Computing*, BigDataScience '14, pages 5:1–5:8, New York, NY, USA, 2014. ACM. ISBN 978-1-4503-2891-3. doi: 10.1145/2640087.2644157. URL http://doi.acm.org/10.1145/2640087.2644157.

[10] T.J. Lyons, M.J. Caruana, and T. Lévy. *Differential Equations Driven by Rough Paths: Ecole dEté de Probabilités de Saint-Flour XXXIV-2004*. Lecture Notes in Mathematics. Springer Berlin Heidelberg, 2007. ISBN 9783540712855. URL https://books.google.com/books?id=hOm5BQAAQBAJ.

[11] MATLAB. *R2009a*. The MathWorks Inc., Natick, Massachusetts, 2009. URL http://www.mathworks.com/help/matlab/.

[12] MATLAB. *R2016a*. The MathWorks Inc., Natick, Massachusetts, 2016. URL http://www.mathworks.com/help/matlab/.