Signature Stochastic Differential Equations (SDEs) from an affine and polynomial perspective

Josef Teichmann (joint work with
Christa Cuchiero and Sara Svaluto-Ferro)

ETH Zürich

April 2022

Section 1

Some thoughts on regularity

Regularity

- Analysis: from analytic to measurable functions.
- Partial differential equations: from (weakly) differentiable functions to distributions and viscosity solutions.
- Stochastic Analysis: from bounded variation to semi-martingales.
- Rough Analysis reveals the inner structure of low regularity objects from, e.g., stochastic analysis.

Regularity

- Analysis: from analytic to measurable functions.
- Partial differential equations: from (weakly) differentiable functions to distributions and viscosity solutions.
- Stochastic Analysis: from bounded variation to semi-martingales.
- Rough Analysis reveals the inner structure of low regularity objects from, e.g., stochastic analysis.

Regularity

- Analysis: from analytic to measurable functions.
- Partial differential equations: from (weakly) differentiable functions to distributions and viscosity solutions.
- Stochastic Analysis: from bounded variation to semi-martingales.
- Rough Analysis reveals the inner structure of low regularity objects from, e.g., stochastic analysis.

Regularity

- Analysis: from analytic to measurable functions.
- Partial differential equations: from (weakly) differentiable functions to distributions and viscosity solutions.
- Stochastic Analysis: from bounded variation to semi-martingales.
- Rough Analysis reveals the inner structure of low regularity objects from, e.g., stochastic analysis.

Ito-Malliavin-Lyons

- Solutions of stochastic differential equations driven by Brownian motion are in general *only* measurable on Wiener space (Ito) with weak differentiability properties under mild regularity assumptions on the vector fields (Malliavin).
- Solutions of stochastic differential equations can be split in a measurable map (signature) and a differentiable map under mild regularity assumptions on the vector fields (Lyons).

Ito-Malliavin-Lyons

- Solutions of stochastic differential equations driven by Brownian motion are in general *only* measurable on Wiener space (Ito) with weak differentiability properties under mild regularity assumptions on the vector fields (Malliavin).
- Solutions of stochastic differential equations can be split in a measurable map (signature) and a differentiable map under mild regularity assumptions on the vector fields (Lyons).

Signature

Signature is the collection of all iterated integrals of a multi-variate path $u : \mathbb{R} \to \mathbb{R}^d$ with respect to some integration theory. Let us assume first order calculus for the moment (finite variation paths, Stratonovich Brownian motion).

Signature has many remarkable properties, most importantly linear functionals on signature of a (continuous) finite variation or rough path u (extended by time)

$$\Big\{ \sum_{k,i_1,\ldots,i_k} \ell_{i_1 \cdots i_k} \int_{0 \le t_1 \le \ldots t_k \le t} du^{i_1}(t_1) \cdots du^{i_k}(t_k) \,|\, (i_1, \ldots, i_k) \in \{0, \ldots, d\}^k \Big\}$$

form a point separating algebra of path space functionals on paths starting at 0.

Therefore signature is a *universal linearizer* on path space and often analyzed from an algebraic point of view, since polynomials of signature can be expressed as linear combinations of signature.

Characteristics of semi-martingales

Since signature appears as universal linearizer, one can linearize integrands or characteristics of general stochastic processes, which relates those processes to polynomial or affine theory.

Signature as a probabilistic object

It is the goal of this work to contribute to the probabilistic theory of signatures of stochastic processes in a twofold way:

- develop a dynamic theory of expectations of polynomials of signatures.
- develop an affine theory for expectations of Fourier functionals of signatures.

Notice that signature of, e.g., Brownian motion is a highly non-trivial probabilistic object: we have hypo-ellipticity phenomena, moment indeterminate laws, co-monotonicities, etc.

Signature as a probabilistic object

It is the goal of this work to contribute to the probabilistic theory of signatures of stochastic processes in a twofold way:

- develop a dynamic theory of expectations of polynomials of signatures.
- develop an affine theory for expectations of Fourier functionals of signatures.

Notice that signature of, e.g., Brownian motion is a highly non-trivial probabilistic object: we have hypo-ellipticity phenomena, moment indeterminate laws, co-monotonicities, etc.

Section 2

Signatures

Controlled ordinary differential equations (CODE)

We consider differential equations of the form

$$dY_t = \sum_i V_i(Y_t) du_t^i \,, \; Y_0 = y \in E$$

to construction evolutions in state space E (could be a manifold of finite or infinite dimension) depending on local characteristics, initial value $y \in E$ and the control u.

If the map $y \to Y_T$ is considered CODEs are a model for feedforward neural networks, residual networks, etc (see joint work with Christa Cuchiero and Martin Larsson).

CODEs: control as input

For this talk we fix $y \in E$ and consider

$$u \mapsto W \,\mathrm{Evol}_{s,t}(y)$$

and train the readout and/or the vector fields.

Does this also correspond to classes of networks? Yes: these are continuous time versions of rNNs, LSTMs, etc.

It can be used for time series, predictions, etc.

Reservoir Computing (RC)

... We aim to learn an input-output map on a high- or infinite dimensional input state space. Consider the input as well as the output dynamic, e.g. a time series. An example: learn a given evolution on state space E:

Paradigm of Reservoir computing (Herbert Jäger, Lyudmila, Grigoryeva, Wolfgang Maas, Juan-Pablo Ortega, et al.)

Split the input-output map into a generic part of generalized rNN-type (the *reservoir*), which is *not* trained and a readout part, which is trained.

Often the readout is chosen linear and the reservoir has random features. The reservoir is usually a numerically very tractable dynamical system.

Applications of RC

- Often reservoirs can be realized physically, whence ultrafast evaluations are possible. Only the readout map W has to be trained.
- One can learn dynamic phenomena *without* knowing the specific characteristics.
- It works unreasonably well with generalization tasks.

Applications of RC

- Often reservoirs can be realized physically, whence ultrafast evaluations are possible. Only the readout map W has to be trained.
- One can learn dynamic phenomena *without* knowing the specific characteristics.
- It works unreasonably well with generalization tasks.

Applications of RC

- Often reservoirs can be realized physically, whence ultrafast evaluations are possible. Only the readout map W has to be trained.
- One can learn dynamic phenomena *without* knowing the specific characteristics.
- It works unreasonably well with generalization tasks.

An instance of RC are CODEs/RDEs/SDEs

Consider a controlled differential equation

$$dY_t = \sum_{i=1}^{d} V_i(Y_t) du_t^i \,, \; Y_0 = y \in E$$

for some smooth vector fields $V_i : E \to TE$, $i = 1, \ldots, d$ and d independent (Stratonovich) Brownian motions u^i, or finite variation continuous controls, or a rough path, or a semi-martingale. This describes a controlled dynamics on E.

We want to learn the dynamics, i.e. the map

(input control u) $\mapsto$ (solution Y).

Obviously a complicated, non-linear map, ...

Transport operators

We introduce some notation for this purpose:

Definition

Let $V : E \to E$ be a smooth vector field, and let $f : E \to \mathbb{R}$ be a smooth function, then we call

$$Vf(x) = df(x) \bullet V(x)$$

the transport operator associated to V, which maps smooth functions to smooth functions and determines V uniquely.

Taylor expansion

Theorem

Let Evol *be a smooth evolution operator on a convenient manifold E which satisfies (again the time derivative is taken with respect to the forward variable t) a controlled ordinary differential equation*

$$d\,\mathrm{Evol}_{s,t}(x) = \sum_{i=1}^{d} V_i(\mathrm{Evol}_{s,t}(x))du^i(t)$$

then for any smooth function $f : E \to \mathbb{R}$, and every $x \in E$

$$f\left(\mathrm{Evol}_{s,t}(x)\right) =$$
$$= \sum_{k=0}^{M} \sum_{i_1,\dots,i_k=1}^{d} V_{i_1} \cdots V_{i_k} f(x) \int_{s \le t_1 \le \cdots \le t_k \le t} du^{i_1}(t_1) \cdots du^{i_k}(t_k) +$$
$$+ R_M(s,t,f)$$

Taylor expansion

with remainder term

$$R_M(s,t,f) =$$
$$= \sum_{i_0,\dots,i_M=1}^{d} \int_{s \le t_0 \le \cdots \le t_M \le t} V_{i_0} \cdots V_{i_M} f\big(\operatorname{Evol}_{s,t_0}(x)\big)\, du^{i_0}(t_0) \cdots du^{i_M}(t_M)$$

holds true for all times $s \le t$ and every natural number $M \ge 0$.

A lot of work has been done to understand the analysis, algebra and geometry of this expansion (Kua-Tsai Chen, Gerard Ben-Arous, Terry Lyons). It is a starting point of *rough path analysis* (Terry Lyons, Peter Friz, etc).

Hopf algebraic interpretation

Definition

Consider the free algebra $\mathbb{A}_d$ of formal series generated by d non-commutative indeterminates $e_1, \ldots, e_d$ (actually a Hopf Algebra). A typical element $a \in \mathbb{A}_d$ is written as

$$a = \sum_{k=0}^{\infty} \sum_{i_1,\ldots,i_k=1}^{d} a_{i_1 \ldots i_k} e_{i_1} \cdots e_{i_k} \ ,$$

sums and products are defined in the natural way. We consider the complete locally convex topology making all projections $a \mapsto a_{i_1 \ldots i_k}$ continuous on $\mathbb{A}_d$, hence a convenient vector space.

Vector fields in $\mathbb{A}_d$

Definition

We define on $\mathbb{A}_d$ smooth vector fields

$$a \mapsto ae_i$$

for $i = 1, \ldots, d$.

Signature

Theorem

Let u be a smooth control, then the controlled differential equation

$$d\,\mathrm{Sig}_{s,t}(a) = \sum_{i=1}^{d} \mathrm{Sig}_{s,t}(a) e_i du^i(t)\,,\ \mathrm{Sig}_{s,s}(a) = a \qquad (1)$$

has a unique smooth evolution operator, called signature of u and denoted by Sig, *given by*

$$\mathrm{Sig}_{s,t}(a) = a \sum_{k=0}^{\infty} \sum_{i_1,\dots,u_k=1}^{d} \int_{s\le t_1\le\cdots\le t_k\le t} du^{i_1}(t_1)\cdots du^{i_k}(t_k)\, e_{i_1}\cdots e_{i_k}\,. \qquad (2)$$

Actually $\mathrm{Sig}(e)$ *takes values in a Lie group G and any element of G can be reached up to arbitrary order of accuracy by such evolutions starting at e. Additionally the restriction of linear maps on G is an algebra.*

Signature as an abstract reservoir

Theorem (Signature is a reservoir)

Let Evol *be a smooth evolution operator on a convenient vector space E which satisfies (again the time derivative is taken with respect to the forward variable t) a controlled ordinary differential equation*

$$d\,\mathrm{Evol}_{s,t}(x) = \sum_{i=1}^{d} V_i(\mathrm{Evol}_{s,t}(x))du^i(t)\,.$$

Then for any smooth (test) function $f : E \to \mathbb{R}$ and for every $M \geq 0$ there is a time-homogenous linear $W = W(V_1, \ldots, V_d, f, M, x)$ from $\mathbb{A}_d^M$ to the real numbers $\mathbb{R}$ such that

$$f\big(\mathrm{Evol}_{s,t}(x)\big) = W\big(\pi_M(\mathrm{Sig}_{s,t}(1))\big) + \mathcal{O}\big((t-s)^{M+1}\big)$$

for $s \leq t$.

Signature as reservoir

- This explains that any solution can be represented – up to a linear readout – by a universal reservoir, namely signature. Similar constructions can be done in regularity structures, too (branched rough paths, etc).
- This is used in many instances of provable machine learning by, e.g., groups in Oxford (Harald Oberhauser, Terry Lyons, etc), and also ...
- ... at JP Morgan, in particular great recent work on 'Nonparametric pricing and hedging of exotic derivatives' by Terry Lyons, Sina Nejad and Imanol Perez Arribas.
- in contrast to reservoir computing: signature is high dimensional (i.e. infinite dimensional) and a precisely defined, non-random object.
- Can we approximate signature by a lower dimensional random object with similar properties?

Signature as reservoir

- This explains that any solution can be represented – up to a linear readout – by a universal reservoir, namely signature. Similar constructions can be done in regularity structures, too (branched rough paths, etc).
- This is used in many instances of provable machine learning by, e.g., groups in Oxford (Harald Oberhauser, Terry Lyons, etc), and also ...
- ... at JP Morgan, in particular great recent work on 'Nonparametric pricing and hedging of exotic derivatives' by Terry Lyons, Sina Nejad and Imanol Perez Arribas.
- in contrast to reservoir computing: signature is high dimensional (i.e. infinite dimensional) and a precisely defined, non-random object.
- Can we approximate signature by a lower dimensional random object with similar properties?

Signature as reservoir

- This explains that any solution can be represented – up to a linear readout – by a universal reservoir, namely signature. Similar constructions can be done in regularity structures, too (branched rough paths, etc).
- This is used in many instances of provable machine learning by, e.g., groups in Oxford (Harald Oberhauser, Terry Lyons, etc), and also ...
- ... at JP Morgan, in particular great recent work on 'Nonparametric pricing and hedging of exotic derivatives' by Terry Lyons, Sina Nejad and Imanol Perez Arribas.
- in contrast to reservoir computing: signature is high dimensional (i.e. infinite dimensional) and a precisely defined, non-random object.
- Can we approximate signature by a lower dimensional random object with similar properties?

Signature as reservoir

- This explains that any solution can be represented – up to a linear readout – by a universal reservoir, namely signature. Similar constructions can be done in regularity structures, too (branched rough paths, etc).
- This is used in many instances of provable machine learning by, e.g., groups in Oxford (Harald Oberhauser, Terry Lyons, etc), and also ...
- ... at JP Morgan, in particular great recent work on 'Nonparametric pricing and hedging of exotic derivatives' by Terry Lyons, Sina Nejad and Imanol Perez Arribas.
- in contrast to reservoir computing: signature is high dimensional (i.e. infinite dimensional) and a precisely defined, non-random object.
- Can we approximate signature by a lower dimensional random object with similar properties?

Signature as reservoir

- This explains that any solution can be represented – up to a linear readout – by a universal reservoir, namely signature. Similar constructions can be done in regularity structures, too (branched rough paths, etc).
- This is used in many instances of provable machine learning by, e.g., groups in Oxford (Harald Oberhauser, Terry Lyons, etc), and also ...
- ... at JP Morgan, in particular great recent work on 'Nonparametric pricing and hedging of exotic derivatives' by Terry Lyons, Sina Nejad and Imanol Perez Arribas.
- in contrast to reservoir computing: signature is high dimensional (i.e. infinite dimensional) and a precisely defined, non-random object.
- Can we approximate signature by a lower dimensional random object with similar properties?

Signature for semi-martingales

We shall consider now $\mathbb{R}_{\geq 0}$ as time interval except otherwise mentioned. The stochastic basis satisfies usual conditions.

Let us introduce some notation: we denote by $\mathbb{S}$ the set of simple predictable processes, i.e. for $\omega \in \Omega$, $s \in T$

$$H_s(\omega) = H_0(\omega) 1_{\{0\}}(s) + \sum_{i=1}^{n} H_i(\omega) 1_{]T_i(\omega), T_{i+1}(\omega)]}(s)$$

for an increasing, finite sequence of stopping times $0 = T_0 \leq T_1 \leq \ldots T_{n+1} < \infty$ and H_i being $\mathcal{F}_{T_i}$ measurable, by $\mathbb{L}$ the set of adapted, caglad processes and by $\mathbb{D}$ the set of adapted, cadlag processes on $\mathbb{R}_{\geq 0}$.

These vector spaces are endowed with the metric

$$d(X, Y) := \sum_{n \geq 0} \frac{1}{2^n} E\big[|(X - Y)|_n^* \wedge 1\big],$$

which makes $\mathbb{L}$ and $\mathbb{D}$ complete topological vector spaces. We call this topology the ucp-topology ("uniform convergence on compacts in probability"). Notice that predictable strategies as well as integrators are considered $\mathbb{R}$ valued here, which, however, *contains* the $\mathbb{R}^n$ case.

Good integrators

Definition

An adapted, cadlag process X is called good integrator if the map

$$J_X : \mathbb{S} \to \mathbb{D}$$

with

$$(H \bullet X)_t := J_X(H)_t := H_0 X_0 + \sum_{i=1}^{n} H_i (X_{T_{i+1} \wedge t} - X_{T_i \wedge t}),$$

for $H \in \mathbb{S}$, is continuous with respect to the ucp-topologies on the respective spaces (this can even be weakened).

Bichteler-Dellacherie Theorem

X is a good integrator if and only if $X = M + A$, where M is a local martingale and A is a process of finite total variation, i.e. X is a semimartingale.

The Emery topology

The Emery topology on the set of semimartingales $\mathbb{SEM}$ is defined by the metric

$$d_E(S_1, S_2) := \sum_{n \geq 0} \frac{1}{2^n} \sup_{K \in \mathbb{S},\, \|K\|_\infty \leq 1} E\left[|(K \bullet (S_1 - S_2))|_n^* \wedge 1 \right].$$

We can by means of the Bichteler-Dellacherie theorem easily prove the following important theorem.

Theorem

The set of semi-martingales $\mathbb{SEM}$ is a topological vector space and complete with respect to the Emery topology.

Theorem

For every semi-martingale X the map J_X from the space $\mathbb{L}$ of càglàd processes to $\mathbb{SEM}$ of semi-martingales is continuous.

Ito's formula

We are now already able to formulate and prove Ito's formula in all generality:

Theorem

Let $X^1, \ldots, X^n$ be good integrators and $f : \mathbb{R}^n \to \mathbb{R}$ a C^2 function, then for $t \geq 0$

$$f(X_t) = \sum_{i=1}^{n} (\partial_i f(X_-) \bullet X^i)_t + \frac{1}{2} \sum_{i,j=1}^{n} (\partial^2_{ij} f(X_-) \bullet [X^i, X^j])_t +$$

$$+ \sum_{0 \leq s \leq t} \left\{ f(X_s) - f(X_s)_- - \sum_{i=1}^{n} \partial_i f(X_s)_- \Delta X^i_s - \frac{1}{2} \sum_{i,j=1}^{n} \partial^2_{ij} f(X_s)_- \Delta X^i_s \Delta X^j_s \right\}.$$

(we apply $X_{0-} = 0$ here.)

Semimartingale Signature (existence)

Theorem

Let $X^1, \dots, X^n$ be good integrators. Consider a free algebra $\mathbb{A}^d$ of power series generated by (non-commutative) generators $e_0, e_i, e_{ij}, e_{ijk}, \dots$, for $i \leq j \leq k \leq \dots \in \{1, \dots, d\}$, then semimartingale signature

$$\text{sem-Sig} = 1 + \int_0^{\cdot} (\text{sem-Sig}_s\, ds) e_0 + \sum_{i=1}^{d} (\text{sem-Sig}_- \bullet X^i) e_i +$$

$$+ \sum_{i \leq j = 1}^{d} (\text{sem-Sig}_- \bullet [X^i, X^j]) e_{ij} +$$

$$\sum_{i \leq j \leq k} (\sum_{s \leq \cdot} \text{sem-Sig}_{s-}\, \Delta X^i_s \Delta X^j_s \Delta X^k_s) e_{ijk} + \dots$$

is a well defined $\mathbb{A}^d$ valued process.

Semi-martingale Signature (density)

The set of all $\langle \ell, \text{sem-Sig} \rangle$ for $\ell \in (\mathbb{A}^d)^*$ is an algebra of semimartingales.

Section 3

Signature Stochastic Differential Equations (SDEs)

Sig-SDEs

Indeed, if X satisfies the generic equation

$$dX_t = b(\widehat{\mathbb{X}}_t)dt + \sqrt{a(\widehat{\mathbb{X}}_t)}dB_t, \quad X_0 \in S \subseteq \mathbb{R}^d\,, \tag{3}$$

where $(\widehat{\mathbb{X}}_t)_{t\geq 0}$ is signature of $t \mapsto (X_t, t)$ and b and a are linear maps, then

- Ito's formula yields that the characteristics of $(\widehat{\mathbb{X}}_t)_{t\geq 0}$ are linear in $(\widehat{\mathbb{X}}_t)_{t\geq 0}$,
- $\mathcal{A}\exp(\langle \mathrm{u}, \mathrm{x}\rangle) = \exp(\langle \mathrm{u}, \mathrm{x}\rangle)\langle R(\mathrm{u}), \mathrm{x}\rangle$ and $\mathcal{A}(\langle u, x\rangle) = \langle L(u), x\rangle$, for x in the state space with R and L being quadratic and linear operators expressible by natural operations on $\mathbb{A}^d$.
- $(\widehat{\mathbb{X}}_t)_{t\geq 0}$ is a $\mathbb{A}_d$-valued linear, hence affine and polynomial process.

Sig-SDEs as affine and polynomial processes

- This means that (under appropriate conditions)
  - ... $\mathbb{E}[\widehat{\mathbb{X}}_T]$ can be computed via polynomial technology, i.e. by solving an infinite dimensional linear ODE.
  - ... $\log \mathbb{E}[\exp(\langle u, \widehat{\mathbb{X}}_T \rangle)]$ can be computed via affine technology, i.e. by solving an infinite dimensional Riccati ODE.
- Sig-SDEs go beyond Markovian settings due to possibly path-dependent coefficients. Signature itself remains Markovian with linear characteristics, which is the essential feature.
- Special cases are Markovian SDEs with b and a analytic in X.
- If b and a only depend on the signature up to order 1 and 2 respectively, then $(\widehat{\mathbb{X}}_t^{\leq N})_{t \geq 0}$ is a finite dimensional polynomial process.