Signatures methods in finance

Christa Cuchiero
partly based on a course given jointly with Sara Svaluto-Ferro

University of Vienna

Mini course

Soesterberg, January 2024

Part I

Introduction to the theory of signature

- partly based on Chapter 7 of "Multidimensional stochastic processes as rough paths - Theory and Applications" by Friz & Victoir (2010)
- We refer to the slides from January 22, 2024.

Part II

Signature methods in Stochastic Portfolio Theory

- based on joint work with Janka Möller; https://arxiv.org/abs/2310.02322
- We refer to the slides from January 23, 2024.

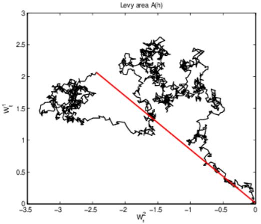
Part III

An affine and polynomial perspective to signature based models

- An overview of affine and polynomial processes by means of Lévy's stochastic area formula
- Signature Stochastic Differential Equations (SDEs) from an affine and polynomial perspective
  - based on joint work with S. Svaluto-Ferro and J. Teichmann ('23); https://arxiv.org/abs/2302.01362
  - Related works (in particular with applications in finance):
    - Sig-SDE models for finance, e.g. I. Arribas, C. Salvi & L. Szpruch ('20)
    - Signature-based models, see e.g. C.C., G. Gazzani, J. Möller & S. Svaluto-Ferro ('23)
    - Neural SDEs, e.g. P. Gierjatovicz et al.
    - Signature stochastic volatility models: pricing and hedging with Fourier by E. Abi-Jaber and L.Gérard

Lévy's stochastic area

- In the article "Le mouvement Brownien plan" (1940), P. Lévy began studying what he called the "stochastic area", i.e., the signed area enclosed by the trajectory of a 2-dimensional Brownian motion W and its chord.



Source: S. J. Malham, Anke Wiese

- In formulas, up to a factor of $\frac{1}{2}$ Lévy's stochastic area is thus given by

$$L_t := \int_0^t W_s^1 dW_s^2 - W_s^2 dW_s^1, \quad t \geq 0.$$

Lévy's stochastic area formula

In subsequent papers, P. Lévy then answered the following ...

Lévy's stochastic area formula

In subsequent papers, P. Lévy then answered the following ...

Questions:

- What is the characteristic function of L_t, i.e., $\mathbb{E}[e^{\mathrm{i}\lambda L_t}]$ for $\lambda \in \mathbb{R}$.
- What is the conditional characteristic function of L_t given W_t, i.e., $\mathbb{E}[e^{\mathrm{i}\lambda L_t}|W_t = y]$ for $y \in \mathbb{R}^2$ and $\lambda \in \mathbb{R}$?

Lévy's stochastic area formula

In subsequent papers, P. Lévy then answered the following ...

Questions:

- What is the characteristic function of L_t, i.e., $\mathbb{E}[e^{i\lambda L_t}]$ for $\lambda \in \mathbb{R}$.
- What is the conditional characteristic function of L_t given W_t, i.e., $\mathbb{E}[e^{i\lambda L_t} | W_t = y]$ for $y \in \mathbb{R}^2$ and $\lambda \in \mathbb{R}$?

- Lévy's formula was generalized and studied by many authors, see e.g., M. Yor (1980), Helmes and Schwane (1983) and the references therein.
- Alternatively to these proofs the above formulas can be derived via the theory of affine processes, in spirit of C.C., S. Svaluto-Ferro & J. Teichmann, "Signature SDEs from an affine and polynomial perspective" ('23).

Definition of affine diffusion processes

Simplest setting (for illustrative purposes): Itô diffusion with state space $S \subseteq \mathbb{R}^d$.

$$dX_t = B(X_t)dt + \sqrt{A(X_t)}dW_t, \quad X_0 = x, \qquad (*)$$

where the characteristics $A : \mathbb{R}^d \to S^+(\mathbb{R}^d)$ and $B : \mathbb{R}^d \to \mathbb{R}^d$ are continuous functions and W a Brownian motion on $\mathbb{R}^d$.

Definition of affine diffusion processes

Simplest setting (for illustrative purposes): Itô diffusion with state space $S \subseteq \mathbb{R}^d$.

$$dX_t = B(X_t)dt + \sqrt{A(X_t)}dW_t, \quad X_0 = x, \qquad (*)$$

where the characteristics $A : \mathbb{R}^d \to S^+(\mathbb{R}^d)$ and $B : \mathbb{R}^d \to \mathbb{R}^d$ are continuous functions and W a Brownian motion on $\mathbb{R}^d$.

Definition

A weak solution X of $(*)$ is called affine process if B and A are affine functions, i.e.,

$$B(x) = b + \sum_{i=1}^{d} \beta_i x_i, \quad A(x) = a + \sum_{i=1}^{d} \alpha_i x_i,$$

for characteristics $b, \beta_i \in \mathbb{R}^d$ and $a, \alpha_i \in \mathbb{R}^{d\times d}$.

Definition of affine diffusion processes

Simplest setting (for illustrative purposes): Itô diffusion with state space $S \subseteq \mathbb{R}^d$.

$$dX_t = B(X_t)dt + \sqrt{A(X_t)}dW_t, \quad X_0 = x, \qquad (*)$$

where the characteristics $A : \mathbb{R}^d \to S^+(\mathbb{R}^d)$ and $B : \mathbb{R}^d \to \mathbb{R}^d$ are continuous functions and W a Brownian motion on $\mathbb{R}^d$.

Definition

A weak solution X of $(*)$ is called affine process if B and A are affine functions, i.e.,

$$B(x) = b + \sum_{i=1}^{d} \beta_i x_i, \quad A(x) = a + \sum_{i=1}^{d} \alpha_i x_i,$$

for characteristics b, $\beta_i \in \mathbb{R}^d$ and $a, \alpha_i \in \mathbb{R}^{d \times d}$.

From this definition affine processes appear as a narrow class, whose universal character announced in the title of this talk is at this stage by no means visible.

Affine transform formula - Riccati ODEs

The remarkable implication is that exponential moments, i.e., $\mathbb{E}[\exp(\langle u, X_t \rangle)]$ for $u \in \mathbb{C}^d$, in particular the characteristic function, can be expressed as solutions of Riccati ordinary differential equations (ODEs).

Theorem (D. Duffie, D. Filipovic & W. Schachermayer ('03), C.C. & J. Teichmann ('13))

Let $(X_t)_{t \geq 0}$ be an affine process and let $u \in \mathbb{C}^d$ such that $\mathbb{E}[\exp(|\langle u, X_t \rangle|)] < \infty$. Then,

$$\mathbb{E}_x \left[\exp(\langle u, X_t \rangle)\right] = \Phi(t) \exp(\langle \psi(t), x \rangle),$$

where Φ and ψ solve the Riccati ODEs given by

$$\partial_t \Phi(t) = \Phi(t) F(\psi_t), \quad \Phi(0) = 1, \qquad \partial_t \psi(t) = R(\psi(t)), \quad \psi(0) = u,$$

where

$$F(u) = b^\top u + u^\top a u, \qquad R_i(u) = \beta_i^\top u + u^\top \alpha_i u, \quad i = 1, \dots, d.$$

Back to Lévy's stochastic area formula

The affine transform formula is thus tailor-made to compute the characteristic function of the Lévy stochastic area L, if we can embed it within an affine process.

Lemma

Let W be a 2-dimensional Brownian motion and consider the 4-dimensional process $(X_t)_{t\geq 0} = (x_1 + W_t^1, x_2 + W_t^2, x_3 + L_t, \|(x_1,x_2)^\top + W_t\|^2)_{t\geq 0}$.

Then X is an affine process with initial value $x = (x_1, x_2, x_3, \|(x_1,x_2)^\top\|^2) \in \mathbb{R}^4$ *and characteristics*

$$a = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{pmatrix}, \quad b = \begin{pmatrix} 0 \\ 0 \\ 0 \\ 2 \end{pmatrix},$$

$$\alpha_1 = \begin{pmatrix} 0 & 0 & 0 & 2 \\ 0 & 0 & 1 & 0 \\ 0 & 1 & 0 & 0 \\ 2 & 0 & 0 & 0 \end{pmatrix}, \quad \alpha_2 = \begin{pmatrix} 0 & 0 & -1 & 0 \\ 0 & 0 & 0 & 2 \\ -1 & 0 & 0 & 0 \\ 0 & 2 & 0 & 0 \end{pmatrix}, \quad \alpha_4 = \begin{pmatrix} 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 4 \end{pmatrix},$$

and all others are 0.

Back to Lévy's stochastic area formula

The affine transform formula is thus tailor-made to compute the characteristic function of the Lévy stochastic area L, if we can embed it within an affine process.

Lemma

Let W be a 2-dimensional Brownian motion and consider the 4-dimensional process $(X_t)_{t\geq 0} = (x_1 + W_t^1, x_2 + W_t^2, x_3 + L_t, \|(x_1, x_2)^\top + W_t\|^2)_{t\geq 0}$.

Then X is an affine process with initial value $x = (x_1, x_2, x_3, \|(x_1, x_2)^\top\|^2) \in \mathbb{R}^4$ *and characteristics*

$$a = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{pmatrix}, \quad b = \begin{pmatrix} 0 \\ 0 \\ 0 \\ 2 \end{pmatrix},$$

$$\alpha_1 = \begin{pmatrix} 0 & 0 & 0 & 2 \\ 0 & 0 & 1 & 0 \\ 0 & 1 & 0 & 0 \\ 2 & 0 & 0 & 0 \end{pmatrix}, \quad \alpha_2 = \begin{pmatrix} 0 & 0 & -1 & 0 \\ 0 & 0 & 0 & 2 \\ -1 & 0 & 0 & 0 \\ 0 & 2 & 0 & 0 \end{pmatrix}, \quad \alpha_4 = \begin{pmatrix} 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 4 \end{pmatrix},$$

and all others are 0.

Key idea: lift the process to a higher-dimensional state space to make it affine

Lévy's stochastic area formula via affine processes

To compute $\mathbb{E}[e^{\mathrm{i}\lambda L_t}]$, set $\mathbb{R}^4 \ni u = (0, 0, \mathrm{i}\lambda, 0)^\top$. Then the Riccati ODEs reduce to $\psi_1 = \psi_2 = 0$, $\psi_3 = \mathrm{i}\lambda$ and

$$\partial_t \psi_4(t) = \frac{1}{2}(4(\psi_4(t))^2 - \lambda^2),\ \psi_4 = 0, \quad \partial_t \Phi(t) = 2\Phi(t)\psi_4(t),\ \Phi(t) = 1,$$

whose solutions are given by

$$\psi_4(t) = -\frac{\lambda \tanh(\lambda t)}{2}, \quad \Phi(t) = \frac{1}{\cosh(\lambda t)}.$$

Theorem

The characteristic function of $x_3 + L_t$ is given by

$$\mathbb{E}_x[e^{i\lambda(x_3 + L_t)}] = \frac{1}{\cosh(\lambda t)} \exp(x_3 i\lambda + \|(x_1, x_2)^\top\|^2 \psi_4(t)), \quad \lambda \in \mathbb{R}.$$

By setting $x = 0$, we then get the first Lévy stochastic area formula

$$\mathbb{E}_0[e^{i\lambda L_t}] = \frac{1}{\cosh(\lambda t)}, \quad \lambda \in \mathbb{R}.$$

Lévy's stochastic area formula via affine processes

- For the second one, we can compute the joint characteristic function $\mathbb{E}_0[e^{\mathrm{i}\lambda L_t + \mathrm{i}\langle v, W_t\rangle}]$ by solving additionally to ψ_4 from above the "Riccati" ODEs

$$\begin{pmatrix}\partial_t\psi_1(t)\\ \partial_t\psi_2(t)\end{pmatrix} = \begin{pmatrix}4\psi_4(t) & 2\mathrm{i}\lambda\\ -2\mathrm{i}\lambda & 4\psi_4(t)\end{pmatrix}\begin{pmatrix}\psi_1(t)\\ \psi_2(t)\end{pmatrix}, \qquad \begin{pmatrix}\psi_1(0)\\ \psi_2(0)\end{pmatrix} = \mathrm{i}v,$$
$$\partial_t\Phi(t) = \Phi(t)(\psi_1^2(t) + \psi_2^2(t) + 2\psi_4(t)), \qquad \Phi(0) = 1,$$

  which yields

$$\mathbb{E}_0[e^{\mathrm{i}\lambda L_t + \mathrm{i}\langle v, W_t\rangle}] = \frac{1}{\cosh(\lambda t)}\exp\Big(-\frac{\|v\|^2}{2\lambda\coth(\lambda t)}\Big)$$

- Since $\mathbb{E}_0[e^{\mathrm{i}\lambda L_t + \mathrm{i}\langle v, W_t\rangle}] = \int_{\mathbb{R}^2} e^{\mathrm{i}\langle v, y\rangle}\mathbb{E}[e^{\mathrm{i}\lambda L_t} | W_t = y]\frac{1}{2\pi t}e^{-\frac{1}{2t}\|y\|^2}dy$ holds, Fourier inversion yields...

Lévy's stochastic area formula via affine processes

- For the second one, we can compute the joint characteristic function $\mathbb{E}_0[e^{i\lambda L_t + i\langle v, W_t\rangle}]$ by solving additionally to ψ_4 from above the "Riccati" ODEs

$$\begin{pmatrix}\partial_t\psi_1(t)\\ \partial_t\psi_2(t)\end{pmatrix} = \begin{pmatrix}4\psi_4(t) & 2i\lambda\\ -2i\lambda & 4\psi_4(t)\end{pmatrix}\begin{pmatrix}\psi_1(t)\\ \psi_2(t)\end{pmatrix}, \qquad \begin{pmatrix}\psi_1(0)\\ \psi_2(0)\end{pmatrix} = iv,$$
$$\partial_t\Phi(t) = \Phi(t)(\psi_1^2(t) + \psi_2^2(t) + 2\psi_4(t)), \qquad \Phi(0) = 1,$$

  which yields

$$\mathbb{E}_0[e^{i\lambda L_t + i\langle v, W_t\rangle}] = \frac{1}{\cosh(\lambda t)}\exp\Big(-\frac{\|v\|^2}{2\lambda\coth(\lambda t)}\Big)$$

- Since $\mathbb{E}_0[e^{i\lambda L_t + i\langle v, W_t\rangle}] = \int_{\mathbb{R}^2} e^{i\langle v, y\rangle}\mathbb{E}[e^{i\lambda L_t}|W_t = y]\frac{1}{2\pi t}e^{-\frac{1}{2t}\|y\|^2}dy$ holds, Fourier inversion yields...

Theorem

$$\mathbb{E}[e^{i\lambda L_t}|W_t = y] = \frac{\lambda t}{\sinh\lambda t}\exp\left(\|y\|^2\frac{1-\lambda t\coth(\lambda t)}{2t}\right), \quad \lambda\in\mathbb{R}.$$

Definition of polynomial diffusion processes

Any affine diffusion process is also a polynomial process. Therefore moments of the Lévy area L_t can also be computed by so-called polynomial technology.

Definition

A weak solution X of $(*)$ is called polynomial process if B is affine and A quadratic.

- In this finite dimensional diffusion framework polynomial processes are always more general than affine ones. This does not necessarily hold true in the presence of jumps.
- As we shall see, in certain infinite dimensional setups the notions of affine and polynomial diffusion processes coincide.

Definition of polynomial diffusion processes

Any affine diffusion process is also a polynomial process. Therefore moments of the Lévy area L_t can also be computed by so-called polynomial technology.

Definition

A weak solution X of $(*)$ is called polynomial process if B is affine and A quadratic.

- In this finite dimensional diffusion framework polynomial processes are always more general than affine ones. This does not necessarily hold true in the presence of jumps.
- As we shall see, in certain infinite dimensional setups the notions of affine and polynomial diffusion processes coincide.
- Denote by $\mathcal{P}_k$ polynomials on $S \subseteq \mathbb{R}^d$ up to degree $k \in \mathbb{N}$, i.e. $\mathcal{P}_k = \{x \mapsto \sum_{|\mathbf{i}|=0}^{k} u_{\mathbf{i}} x^{\mathbf{i}} \mid u_{\mathbf{i}} \in \mathbb{R}\}$, where we use multi-index notation $\mathbf{i} = (i_1, \ldots, i_d) \in \mathbb{N}^d$, $|\mathbf{i}| = i_1 + \cdots + i_d$ and $x^{\mathbf{i}} = x^{i_1} \cdots x^{i_d}$. The dimension of $\mathcal{P}_k$ is denoted by N.

Moment formula

- We write $u \in \mathbb{R}^N$ for the coefficients vector and define $p(x,u) := \sum_{|\mathbf{i}|=0}^{k} u_{\mathbf{i}} x^{\mathbf{i}}$.
- Note that for a polynomial process, the generator $\mathcal{A}$ maps $\mathcal{P}_k$ to $\mathcal{P}_k$, i.e. polynomials to polynomials of same or smaller degree.
- Hence there is a linear map L_N from $\mathbb{R}^N$ to $\mathbb{R}^N$ such that

$$\mathcal{A}(p(\cdot,u))(x) = p(x, L_N u).$$

Theorem (C.C., M. Keller-Ressel & J. Teichmann ('12), D. Filipovic & M. Larsson ('16))

Let $(X_t)_{t\in[0,T]}$ be a polynomial process. Denote by $c(t)$ the solution of the linear ODE given by

$$\partial_t c(t) = L_N c(t), \quad c(0) = u \in \mathbb{R}^N.$$

Then,

$$\mathbb{E}_x\Big[\sum_{|\mathbf{i}|=0}^{k} u_{\mathbf{i}} X_t^{\mathbf{i}}\Big] = \sum_{|\mathbf{i}|=0}^{k} c_{\mathbf{i}}(t) x^{\mathbf{i}}.$$

Affine and polynomial processes as universal model class?

- Despite the rather narrow definition, already the finite dimensional affine and polynomial processes contains many well-known models, e.g.,
  - Ornstein-Uhlenbeck, Feller-type and Wishart processes, the Black-Scholes and the Heston model, the Fisher-Snedecor process, the Wright-Fisher diffusion as well as all possible combinations thereof.

Affine and polynomial processes as universal model class?

- Despite the rather narrow definition, already the finite dimensional affine and polynomial processes contains many well-known models, e.g.,
  - Ornstein-Uhlenbeck, Feller-type and Wishart processes, the Black-Scholes and the Heston model, the Fisher-Snedecor process, the Wright-Fisher diffusion as well as all possible combinations thereof.
- As we shall see, the true universal character becomes visible through infinite dimensional affine and polynomial
- They appear either as infinite dimensional analogs of the finite dimensional ones, usually with a much more intricate structure, or as lifts, in spirit of the lift of the Lévy area.

Affine and polynomial processes as universal model class?

- Despite the rather narrow definition, already the finite dimensional affine and polynomial processes contains many well-known models, e.g.,
  - Ornstein-Uhlenbeck, Feller-type and Wishart processes, the Black-Scholes and the Heston model, the Fisher-Snedecor process, the Wright-Fisher diffusion as well as all possible combinations thereof.
- As we shall see, the true universal character becomes visible through infinite dimensional affine and polynomial
- They appear either as infinite dimensional analogs of the finite dimensional ones, usually with a much more intricate structure, or as lifts, in spirit of the lift of the Lévy area.
  - Infinite dimensional analogs can often be realized as measure valued or Hilbert space valued processes.
  - Markovian lifts appear for instance as lifts of stochastic Volterra processes or signature lifts.

Infinite dimensional examples and their applications

Measure-valued affine and polynomial processes:

- Most prominent examples: Dawson-Watanabe and Flemming-Viot processes
- Measure-valued branching processes in the sense of Z. Li ('11)
- Characterization of measure-valued affine and polynomial diffusions: C.C., M. Larsson & S. Svaluto-Ferro ('19) and C.C., L. Di Persio, F. Guida & S. Svaluto-Ferro ('21).
- Applications: population genetics, chemistry, energy market modeling

Infinite dimensional examples and their applications

Measure-valued affine and polynomial processes:

- Most prominent examples: Dawson-Watanabe and Flemming-Viot processes
- Measure-valued branching processes in the sense of Z. Li ('11)
- Characterization of measure-valued affine and polynomial diffusions: C.C., M. Larsson & S. Svaluto-Ferro ('19) and C.C., L. Di Persio, F. Guida & S. Svaluto-Ferro ('21).
- Applications: population genetics, chemistry, energy market modeling

Hilbert space valued affine and polynomial processes:

- e.g., F. Benth and I. Simonsen ('18), T. Schmidt, S. Tappe and W. Yu ('20), S. Cox, S.Karbach & A. Khedher ('22), S. Cox, C.C. & A. Khedher ('23)
- Applications: Forward curve modeling, infinite dimensional covariance modeling and stochastic volatility models, infinite dimensional Wishart processes

Infinite dimensional examples and their applications

Measure-valued affine and polynomial processes:

- Most prominent examples: Dawson-Watanabe and Flemming-Viot processes
- Measure-valued branching processes in the sense of Z. Li ('11)
- Characterization of measure-valued affine and polynomial diffusions: C.C., M. Larsson & S. Svaluto-Ferro ('19) and C.C., L. Di Persio, F. Guida & S. Svaluto-Ferro ('21).
- Applications: population genetics, chemistry, energy market modeling

Hilbert space valued affine and polynomial processes:

- e.g., F. Benth and I. Simonsen ('18), T. Schmidt, S. Tappe and W. Yu ('20), S. Cox, S.Karbach & A. Khedher ('22), S. Cox, C.C. & A. Khedher ('23)
- Applications: Forward curve modeling, infinite dimensional covariance modeling and stochastic volatility models, infinite dimensional Wishart processes

Markovian lifts of Volterra processes and affine Volterra processes:

- e.g. E. Abi Jaber, M. Larsson & S.Pulido ('19), E. Abi Jaber & O. El Euch ('19), C.C. & J. Teichmann ('20), C.C. & S. Svaluto-Ferro ('21), A. Bondi, S. Pulido & S. Scotti ('22)
- Applications: Rough volatility modeling and forward variance curve modeling, delay equations, etc.

Universal approximation property of affine and polynomial processes?

Classical universal approximation theorems (UATs) state that families of certain functions (e.g. neural networks or polynomials) are dense, say, in $C(K)$ with $K \subseteq \mathbb{R}^d$ some compact set.

Universal approximation property of affine and polynomial processes?

Classical universal approximation theorems (UATs) state that families of certain functions (e.g. neural networks or polynomials) are dense, say, in $C(K)$ with $K \subseteq \mathbb{R}^d$ some compact set.

Research question:

- What is an analog formulation for stochastic processes?

Universal approximation property of affine and polynomial processes?

Classical universal approximation theorems (UATs) state that families of certain functions (e.g. neural networks or polynomials) are dense, say, in $C(K)$ with $K \subseteq \mathbb{R}^d$ some compact set.

Research question:

- What is an analog formulation for stochastic processes?
- Can we prove a universal approximation property of affine and polynomial processes say in the space of all Itô-processes, i.e. stochastic processes with path-dependent characteristics driven by Brownian motion?

Universal approximation property of affine and polynomial processes?

Classical universal approximation theorems (UATs) state that families of certain functions (e.g. neural networks or polynomials) are dense, say, in $C(K)$ with $K \subseteq \mathbb{R}^d$ some compact set.

Research question:

- What is an analog formulation for stochastic processes?
- Can we prove a universal approximation property of affine and polynomial processes say in the space of all Itô-processes, i.e. stochastic processes with path-dependent characteristics driven by Brownian motion?
- What is the right topology, e.g. comparable to polynomials being dense in continuous functions on compacts?

Signature SDEs from an affine and polynomial perspective

One first step to solve these questions is achieved by

- "linearizing" generic classes of Itô-processes (with path-dependent characteristics) called signature SDEs via their signature prolongations;
- showing that these signature prolongations are (infinite dimensional) affine and polynomial processes;
- characterizing their full law on path space via the explicitly computable characteristic function of the signature.

Signature SDEs from an affine and polynomial perspective

One first step to solve these questions is achieved by

- "linearizing" generic classes of Itô-processes (with path-dependent characteristics) called signature SDEs via their signature prolongations;
- showing that these signature prolongations are (infinite dimensional) affine and polynomial processes;
- characterizing their full law on path space via the explicitly computable characteristic function of the signature.

Open question: What is the precise universal approximation property of these signature SDEs?

Partial answer: A global universal approximation property of signature SDEs on the level of characteristics holds, since linear functions of the signature serve as linear regression basis for paths functionals.

Entire functions of the signature

- For an Itô-diffusion X with state space $S \subseteq \mathbb{R}^d$, we denote by $\mathcal{S}(S)$ the set of group like elements of $T((\mathbb{R}^d))$ whose first level lies in S.
- For $\mathbf{x} \in \mathcal{S}(S)$ and $\mathbf{u} \in T((\mathbb{R}^d)) + \mathrm{i}\, T((\mathbb{R}^d))$ set $|\mathbf{u}|_{\mathbf{x}} = \sum_{n=0}^{\infty} |\langle \pi_n(\mathbf{u}), \pi_n(\mathbf{x}) \rangle|$, where π_n denotes the projection on $(\mathbb{R}^d)^{\otimes n}$.
- Dual elements:
  $\mathcal{S}(S)^* := \{\mathbf{u} \in T((\mathbb{R}^d)) + \mathrm{i}\, T((\mathbb{R}^d)) \colon |\mathbf{u}|_{\mathbf{x}} < \infty \text{ for all } \mathbf{x} \in \mathcal{S}(S)\}$.
- For $\mathbf{u} \in \mathcal{S}(S)^*$ entire maps of group like elements are defined as

$$\mathcal{S}(S) \ni \mathbf{x} \mapsto \langle \mathbf{u}, \mathbf{x} \rangle := \lim_{N \to \infty} \sum_{n=0}^{N} \langle \pi_n(\mathbf{u}), \pi_n(\mathbf{x}) \rangle.$$

Entire functions of the signature

- For an Itô-diffusion X with state space $S \subseteq \mathbb{R}^d$, we denote by $\mathcal{S}(S)$ the set of group like elements of $T((\mathbb{R}^d))$ whose first level lies in S.
- For $\mathbf{x} \in \mathcal{S}(S)$ and $\mathbf{u} \in T((\mathbb{R}^d)) + \mathrm{i}\,T((\mathbb{R}^d))$ set $|\mathbf{u}|_{\mathbf{x}} = \sum_{n=0}^{\infty} |\langle \pi_n(\mathbf{u}), \pi_n(\mathbf{x})\rangle|$, where π_n denotes the projection on $(\mathbb{R}^d)^{\otimes n}$.
- Dual elements:
  $\mathcal{S}(S)^* := \{\mathbf{u} \in T((\mathbb{R}^d)) + \mathrm{i}\,T((\mathbb{R}^d)) \colon |\mathbf{u}|_{\mathbf{x}} < \infty \text{ for all } \mathbf{x} \in \mathcal{S}(S)\}$.
- For $\mathbf{u} \in \mathcal{S}(S)^*$ entire maps of group like elements are defined as

$$\mathcal{S}(S) \ni \mathbf{x} \mapsto \langle \mathbf{u}, \mathbf{x}\rangle := \lim_{N\to\infty} \sum_{n=0}^{N} \langle \pi_n(\mathbf{u}), \pi_n(\mathbf{x})\rangle.$$

Products of entire functions are again entire functions.

Proposition (Shuffle property)

Let $(X_t)_{t\in[0,T]}$ be an Itô-process and $\mathbf{u}, \mathbf{v} \in \mathcal{S}(S)^*$. Then $\mathbf{u} \sqcup\!\sqcup \mathbf{v} \in \mathcal{S}(S)^*$ and

$$\langle \mathbf{u}, \mathbb{X}\rangle \langle \mathbf{v}, \mathbb{X}\rangle = \langle \mathbf{u} \sqcup\!\sqcup \mathbf{v}, \mathbb{X}\rangle.$$

Signature SDEs

- We introduce signature SDEs with state space $S \subseteq \mathbb{R}^d$ driven by some d-dimensional Brownian motion W via

$$dX_t = b(\mathbb{X}_t)dt + \sqrt{a(\mathbb{X}_t)}dW_t, \quad X_0 = x. \tag{Sig-SDE}$$

  The coefficients $b : \mathcal{S}(S) \to \mathbb{R}^d$ and $a : \mathcal{S}(S) \to \mathbb{S}^d_+$ are componentwise entire maps of group-like elements, i.e.

$$b_i(\mathbf{x}) = \langle \mathbf{b}_i, \mathbf{x} \rangle \text{ and } a_{ij}(\mathbf{x}) = \langle \mathbf{a}^{ij}, \mathbf{x} \rangle,$$

  where $\mathbf{b}_i, \mathbf{a}_{ij} \in \mathcal{S}(S)^*$.
- In the one-dimensional case this corresponds to an SDE with real-analytic coefficients. Neural SDEs with real-analytic activation functions are also included.

Signature SDEs

- We introduce signature SDEs with state space $S \subseteq \mathbb{R}^d$ driven by some d-dimensional Brownian motion W via

$$dX_t = b(\mathbb{X}_t)dt + \sqrt{a(\mathbb{X}_t)}dW_t, \quad X_0 = x. \tag{Sig-SDE}$$

  The coefficients $b : \mathcal{S}(S) \to \mathbb{R}^d$ and $a : \mathcal{S}(S) \to \mathbb{S}^d_+$ are componentwise entire maps of group-like elements, i.e.

$$b_i(\mathbf{x}) = \langle \mathbf{b}_i, \mathbf{x} \rangle \text{ and } a_{ij}(\mathbf{x}) = \langle \mathbf{a}^{ij}, \mathbf{x} \rangle,$$

  where $\mathbf{b}_i, \mathbf{a}_{ij} \in \mathcal{S}(S)^*$.

- In the one-dimensional case this corresponds to an SDE with real-analytic coefficients. Neural SDEs with real-analytic activation functions are also included.
- Universality on the level of characteristics as all continuous path-functionals can be approximated by linear and thus entire functions of the signature.
- By the shuffle property the characteristics of $\mathbb{X}$ are again entire functions. $\Rightarrow$ $\mathbb{X}$ is an $\mathcal{S}(S)$-valued affine and polynomial process

Main result

Theorem (C.C., S. Svaluto-Ferro, J.Teichmann ('23))

Let X be given by (Sig-SDE) *and fix* $\mathcal{U} \subseteq \mathcal{S}(S)^*$. *Consider the maps* $R : \mathcal{U} \to T((\mathbb{R}^d))$ *and* $L : \mathcal{U} \to T((\mathbb{R}^d))$ *given by*

$$R(\mathbf{u}) = \mathbf{b}^\top \sqcup\!\sqcup \mathbf{u}^{(1)} + \frac{1}{2} \operatorname{Tr}\left(\mathbf{a} \sqcup\!\sqcup \left(\mathbf{u}^{(2)} + \mathbf{u}^{(1)} \sqcup\!\sqcup (\mathbf{u}^{(1)})^\top\right)\right),$$

$$L(\mathbf{u}) = \mathbf{b}^\top \sqcup\!\sqcup \mathbf{u}^{(1)} + \frac{1}{2} \operatorname{Tr}\left(\mathbf{a} \sqcup\!\sqcup \mathbf{u}^{(2)}\right),$$

where $\mathbf{u}^{(1)}, \mathbf{u}^{(2)}$ *denotes certain shifts of* $\mathbf{u}$. *Under some technical conditions,* $\mathbb{X}$ *is an* $\mathcal{S}(S)$*-valued affine and polynomial process satisfying*

$$\mathbb{E}[\exp(\langle \mathbf{u}, \mathbb{X}_t \rangle)] = \exp(\langle \psi(t), \mathbb{X}_0 \rangle), \quad \mathbb{E}[\langle \mathbf{u}, \mathbb{X}_t \rangle] = \langle \mathbf{c}(t), \mathbb{X}_0 \rangle,$$

where ψ *and* $\mathbf{c}$ *are* $\mathcal{U}$*-valued solutions of the extended tensor algebra valued Riccati and linear ODEs, i.e.*

$$\psi(t) = \mathbf{u} + \int_0^t R(\psi(s)) ds, \quad \mathbf{c}(t) = \mathbf{u} + \int_0^t L(\mathbf{c}(s)) ds.$$

Proof ingredients - dynamics of the signature

The following lemma is important to derive the form of L and R.

Lemma

Let X be given by (Sig-SDE) *and fix some multi-index I. Then it holds that*

$$d\langle e_I, \mathbb{X}_t\rangle = \langle e_{I'} \sqcup\!\sqcup \mathbf{b}_{i_{|I|}} + \frac{1}{2} e_{I''} \sqcup\!\sqcup \mathbf{a}_{i_{|I|-1} i_{|I|}}, \mathbb{X}_t\rangle dt + \sum_{j=1}^{d} \langle e_{I'}, \mathbb{X}_t\rangle \sigma(\mathbb{X}_t)_{i_{|I|} j} dB_t^j,$$

where $\sigma(\mathrm{x}) = \sqrt{a(\mathrm{x})}$.

Proof of the lemma

Proof.

By definition we know that $d\langle e_I, \mathbb{X}_t\rangle = \langle e_{I'}, \mathbb{X}_t\rangle \circ dX_t^{i_{|I|}}$ holds. By definition of the Stratonovich integral we have

$$
\begin{aligned}
d\langle e_I, \mathbb{X}_t\rangle &= \langle e_{I'}, \mathbb{X}_t\rangle \circ dX_t^{i_{|I|}} \\
&= \frac{1}{2}\langle e_{I''}, \mathbb{X}_t\rangle d[X^{i_{|I|-1}}, X^{i_{|I|}}]_t + \langle e_{I'}, \mathbb{X}_t\rangle dX_t^{i_{|I|}} \\
&= \Big(\frac{1}{2}\langle e_{I''}, \mathbb{X}_t\rangle\langle \mathbf{a} i_{|I|-1} i_{|I|}, \mathbb{X}_t\rangle + \langle e_{I'}, \mathbb{X}_t\rangle\langle \mathbf{b} i_{|I|}, \mathbb{X}_t\rangle\Big) dt \\
&\quad + \sum_{j=1}^{d} \langle e_{I'}, \mathbb{X}_t\rangle \sigma(\mathbb{X}_t)_{i_{|I|}j} dB_t^j \\
&= \langle \frac{1}{2} e_{I''} \sqcup\!\sqcup\, \mathbf{a} i_{|I|-1} i_{|I|} + e_{I'} \sqcup\!\sqcup\, \mathbf{b} i_{|I|}, \mathbb{X}_t\rangle dt + \sum_{j=1}^{d} \langle e_{I'}, \mathbb{X}_t\rangle \sigma(\mathbb{X}_t)_{i_{|I|}j} dB_t^j,
\end{aligned}
$$

This proves the assertion. □

Proof ingredients and remarks

- This proves that

$$\langle e_I, \mathbb{X}_t \rangle - \int_0^t \langle L(e_I), \mathbb{X}_s \rangle ds$$

  and also

$$\langle \mathbf{u}, \mathbb{X}_t \rangle - \int_0^t \langle L(\mathbf{u}), \mathbb{X}_s \rangle ds$$

  are local martingales and thus that $\mathbb{X}$ is a polynomial process (defined via the martingale problem).

- Similary we get that

$$\exp(\langle \mathbf{u}, \mathbb{X}_t \rangle) - \int_0^t \exp(\langle \mathbf{u}, \mathbb{X}_s \rangle) \langle R(\mathbf{u}), \mathbb{X}_s \rangle ds$$

  is a local martingale, and thus that $\mathbb{X}$ is an affine process (defined via the martingale problem).

- For the moment and affine transform formula to hold it is essential that these are true martingales and the we have solutions to the ODEs.

Consequences

Polynomial point of view

- For generic (universal) classes of stochastic processes, expected signature can be computed by solving infinite dimensional linear ODEs.
- If X is a classical polynomial process, the truncated expected signature can be computed by solving a finite dimensional linear ODE.

Consequences

Polynomial point of view

- For generic (universal) classes of stochastic processes, expected signature can be computed by solving infinite dimensional linear ODEs.
- If X is a classical polynomial process, the truncated expected signature can be computed by solving a finite dimensional linear ODE.

Affine point of view

- Knowing $\mathbf{u} \mapsto \mathbb{E}[\exp(i\langle \mathbf{u}, \mathbb{X}_t \rangle)]$ which is now computable via infinite dimensional Riccati ODEs means to characterize the law of the path-valued random variable $X_{[0,t]}$ without any exponential moment conditions.
$\Rightarrow$ Path characteristic function
- One can now define maximum mean discrepancy distances via

$$\sup_{\mathrm{u} \in \mathcal{S}(S)^*} |\mathbb{E}[\exp(i\langle \mathbf{u}, \mathbb{X}_t] - \mathbb{E}[\exp(i\langle \mathbf{u}, \mathbb{Y}_t \rangle)]|$$

for time-series data generation in an adversarial manner.

Numerical illustration

- Computation of the Laplace transform of a geometric Brownian motion, i.e., $\mathbb{E}[\exp(-\lambda \exp(X_t))]$ where X is a one-dimensional Brownian motion and $\lambda \in \mathbb{R}$.
- In the one dimensional setup $\mathbb{X}_t := (1, X_t, \frac{X_t^2}{2!}, \ldots)$, and the function R is sequence-valued and here of the form

$$R(\mathbf{u})_k = \frac{1}{2}\Big(\mathbf{u}_{k+2} + \sum_{i+j=k} \binom{k}{i} \mathbf{u}_{i+1}\mathbf{u}_{j+1}\Big), \quad k \in \mathbb{N},$$

such that $\mathbb{E}[\exp(-\lambda \exp(X_t))] = \mathbb{E}[\exp(\langle \mathbf{u}, \mathbb{X}_t \rangle)] = \exp(\langle \boldsymbol{\psi}(t), \mathbb{X}_0 \rangle)$, where $\mathbf{u} = -\lambda(1, 1, 1, \ldots)$ and $\partial_t \boldsymbol{\psi}(t) = R(\boldsymbol{\psi}(t))$.

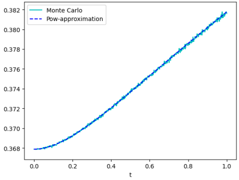
Numerical illustration

- Computation of the Laplace transform of a geometric Brownian motion, i.e., $\mathbb{E}[\exp(-\lambda \exp(X_t))]$ where X is a one-dimensional Brownian motion and $\lambda \in \mathbb{R}$.
- In the one dimensional setup $\mathbb{X}_t := (1, X_t, \frac{X_t^2}{2!}, \ldots)$, and the function R is sequence-valued and here of the form

$$R(\mathbf{u})_k = \frac{1}{2}\Big(\mathbf{u}_{k+2} + \sum_{i+j=k} \binom{k}{i} \mathbf{u}_{i+1}\mathbf{u}_{j+1}\Big), \quad k \in \mathbb{N},$$

such that $\mathbb{E}[\exp(-\lambda \exp(X_t))] = \mathbb{E}[\exp(\langle \mathbf{u}, \mathbb{X}_t \rangle)] = \exp(\langle \psi(t), \mathbb{X}_0 \rangle)$, where $\mathbf{u} = -\lambda(1, 1, 1, \ldots)$ and $\partial_t \psi(t) = R(\psi(t))$.

Numerical implementation for $\lambda = 1$ via a standard ODE solver for the truncated Riccati ODEs with truncation level 20.



$t \mapsto \mathbb{E}[\exp(-\exp(X_t))]$

Conclusion

- Lévy's stochastic area formula as an early example from the literature showing the powerfulness of the unifying affine framework. It can actually be also embedded in the signature SDE framework.

Conclusion

- Lévy's stochastic area formula as an early example from the literature showing the powerfulness of the unifying affine framework. It can actually be also embedded in the signature SDE framework.
- Signature SDEs as generic class of Itô-processes that are affine and polynomial when lifted to the state space of group-like elements
  - ⇒ One step in the direction of universality of affine and polynomial processes
  - ⇒ The characteristic function and the expected value of entire functions of the signature can be computed via Riccati or linear ODEs.
  - ⇒ Tractability properties for neural SDEs and Sig-SDE models
    - ⋆ a systematic polynomial way to compute expected signature,
    - ⋆ Fourier pricing,
    - ⋆ adversarial times series generation.

Outlook

- Existence and uniqueness theory for signature SDEs (joint work with M. Larsson and S. Svaluto-Ferro)
- Universality on the level of stochastic processes
- Theory of entire and real-analytic processes, including jumps, extending the theory of polynomial processes to semigroups mapping real-analytic functions to real-analytic functions (joint work with F. Primavera and S. Svaluto-Ferro)
- Numerical procedures, in particular for solving tensor-algebra valued Riccati and linear ODEs