Derivative Pricing with Signatures

by

Ludovica Pasquini

To obtain the degree of Master of Science in Applied Mathematics at TU Delft,
to be defended publicly on November 21, 2025.

Student number:	5974577	
Project duration:	February 24, 2025 – November 21, 2025	
Thesis committee:	F. Fang,	TU Delft, Daily Supervisor
	K. Vuik,	TU Delft, Responsible Supervisor
	A. Combi,	Rabobank, External Supervisor
	L. Souto,	Rabobank, External Supervisor
	A. Derumigny,	TU Delft

TU Delft Rabobank

Acknowledgements

I share this milestone with many people, who have contributed to this work and to the person I am, more than they realize.

I would like to begin by thanking Alessandro Combi and Luis Souto, who have offered me constant support, believing in my work and in me, even when I did not. Their enthusiasm, ideas, and understanding have truly shaped this thesis and my growth with it. I also would like to thank all the people I met at Rabobank, for making my time there so enjoyable.

I am also grateful to Professor Fang Fang, for her insightful guidance and constructive feedback throughout this project. I further thank Professor Kees Vuik and Professor Alexis Derumigny for being part of the committee.

I owe a great deal to my friends, whose presence and encouragement have carried me until here.

I cannot express in words the gratitude I feel towards my family, for their unwavering and unconditional support.

A special place in these acknowledgments goes to Carla and Paola, le mie nonne.

Ludovica Pasquini
Delft, November 2025

Abstract

This thesis investigates the use of path signatures as feature representations for derivative pricing within a regression-based framework. Motivated by the universal approximation theorem for signatures, which ensures that any continuous path functional can be approximated by a linear functional of its signature, the study evaluates the practical performance of this approach for a range of path-dependent derivatives.

The analysis focuses on two research questions: how does signature-based regression compare with standard Monte Carlo simulation in terms of accuracy and stability, and whether more compact signature representations can alleviate the exponential growth in the signature dimension with respect to the number of assets. To address these questions, the work introduces an alternative representation of the signature features, the filtered signature, which uses the signatures of individual time-augmented asset paths rather than the full joint multi-asset signature, with the aim of reducing dimensionality while retaining essential path information. The effectiveness of this representation, along with the overall performance of the regression framework, is then evaluated through a series of numerical experiments.

For the financial products analyzed, numerical results indicate that, when combined with analytically derived expected signatures, the price estimator obtained through signature-based regression exhibits markedly lower variance than standard Monte Carlo. As a result, the required number of simulated paths to achieve the same level of accuracy is reduced. The results further show that the filtered signature representation attains pricing accuracy and stability comparable to the full signature, while relying on significantly fewer terms. These findings suggest that essential and enough information about the path is encoded within the individual signatures of the time-augmented underlying assets, supporting the adoption of compact signature representations for multi-asset derivative pricing.

Contents

List of Figures

List of Tables

1

Introduction

This thesis investigates the use of path signatures as feature representations for derivative pricing within a linear regression framework. Given a path $X : [0, T] \to \mathbb{R}^d$, the signature collects all iterated integrals of the form

$$\int_{0<u_1<\ldots<u_n<T} dX_{u_1}^{(i_1)} \ldots dX_{u_n}^{(i_n)},$$

encoding the order-by-order interactions between its components. There are two key properties that make signatures particularly valuable for approximation and learning tasks. First, under mild conditions, the signature of a path uniquely characterizes the path up to so-called tree-like equivalences [1]. Second, linear combinations of the signature satisfy a universal approximation property: they can approximate any continuous function defined on a compact subset of the path space. This result, established in [2], positions signatures as a natural generalization of polynomial approximations. Just as polynomials are dense in the space of continuous real-valued functions on a closed interval, linear functionals on signatures can approximate any continuous functional defined on a compact set of paths.

The concept of path signatures originates in the seminal work of Chen [3, 4], who introduced iterated integrals and established their algebraic structure through Chen's identity. In modern stochastic analysis, this theory has been extended through Lyons' rough path framework [5, 6, 7], which generalizes classical integration theory to irregular signals, including diffusions and other processes lacking traditional differentiability. Within this framework, signatures can be defined for stochastic processes with non-smooth trajectories, making them a central object in probability theory and a natural tool for studying functionals of random paths. From the viewpoint of differential equations, signatures admit a compelling interpretation. We consider a system driven by a signal $X \in \mathbb{R}^d$, evolving according to a controlled differential equation $dY_t = V(Y_t)\, dX_t$, with initial condition $Y_0 = y_0$, where $V : \mathbb{R}^e \to \mathbb{R}^{e\times d}$ encodes the dynamics. Then the solution Y is entirely determined by the vector field V and by the signature of the driving signal X; in other words, the signature completely summarizes how an input path influences an output. This observation illustrates the generality of the signature representation and motivates its increasingly broad range of applications in diverse fields.

In recent years, signatures have been successfully employed as feature representations within machine learning algorithms, achieving state-of-the-art performance across a range of tasks. Applications include character and handwriting recognition [8], medical predictive models [9], and human action recognition [10]. Many of these applications share a common principle: they aim to learn or approximate a functional that depends on the evolution of a path, leveraging the descriptive power of the signature to capture the relevant dynamics. This perspective naturally extends to quantitative finance, where derivative prices, optimal trading strategies or risk measures are themselves functionals of underlying asset paths. Consequently, signatures have recently found growing use in financial modeling. Bayer et al. [11] reformulated the optimal stopping problem by introducing randomized stopping times and continuous-time policies, which can then be uniformly approximated by linear functionals of the signature, allowing for efficient numerical solutions to problems such as American option pricing. Similarly, Abi Jaber and Gérard [12] introduced signature volatility models, where the volatility process is represented as a

linear combination of the elements of the signature of a Brownian motion. This construction unifies several classical stochastic volatility models and allows for efficient Fourier-based pricing and hedging within a common algebraic framework. With a similar goal in mind, Pérez-Arribas et al. [13] develop an alternative to neural Stochastic Differential Equation (SDE) by introducing the so-called Signature SDE model, which is calibrated to exotic financial products whose payoffs depend non-linearly on the entire price trajectory. By replacing backpropagation-based optimization with linear regression on signature terms, they achieve consistent calibration under both the pricing and real-world measures. Together, these developments illustrate how signatures provide a unifying language for modeling, pricing, and hedging in finance.

Among the various applications of signatures in finance, [14] provides a clear example of how the properties of signatures can be used directly for derivative pricing. The authors introduce a family of payoffs, the *signature payoffs*, that approximate continuous functionals of the underlying path through a linear functional of its signature. This class of payoffs is universal: it can approximate any continuous payoff functional of the underlying process. With the aim of pricing derivatives, this result is particularly useful because, once the signature of the underlying process is computed, pricing reduces to estimating a finite set of coefficients via linear regression and applying them to the expected signature, exploiting the linearity of both the coefficients and the expectation operator. The regression framework based on signatures has been analyzed in [15, 16], showcasing that the linear regressions built on signature features can successfully learn complex functionals of multivariate stochastic processes, confirming the potential of this approach for financial applications.

Building on these findings and on the universal approximation theorem for signatures, we aim to extend the signature-based regression framework for derivative pricing to payoffs depending on multiple underlying assets. While existing studies have mainly focused on single-asset settings, we investigate whether the regression approach remains effective when the payoff depends jointly on several correlated processes. Although the universality of signature payoffs and existing multivariate studies suggest that the framework should extend to multi-asset derivatives, its practical feasibility is less clear. In applied settings, performance is limited by two factors: computational cost of estimating high-dimensional signatures, which grows exponentially with the number of assets and truncation level, and the statistical efficiency of signature-regression based pricing relative to standard Monte Carlo simulation. These considerations motivate the following research questions.

RQ1: *How does signature-based regression compare with standard Monte Carlo simulation for pricing derivatives?*

RQ2: *Since the size of the signature grows exponentially with the number of assets, can more compact representations retain comparable pricing accuracy while alleviating this curse of dimensionality?*

By addressing these questions, this thesis makes two main contributions. First, it provides an empirical comparison between signature-based regression and standard Monte Carlo simulation for derivative pricing, illustrating how the two methods differ in terms of accuracy and stability across a range of different payoffs. Second, it investigates whether more compact signature representations can mitigate the computational challenges arising from the curse of dimensionality while still producing accurate price estimates.

The remainder of this thesis is organized as follows. Chapter 2 reviews the mathematical theory needed to introduce signatures in the context of financial mathematics, together with the definition of the expected signature and the financial models considered throughout the work. Chapter 3 develops the regression framework underlying the pricing methodology and derives the analytical formulas for the expected signature used throughout the numerical experiments. Chapter 4 reports numerical results, comparing the performance of the signature-based approach under different signature representations and benchmarking it against standard Monte Carlo simulation across a range of representative derivative payoffs. Finally, Chapter 5 summarizes the main findings and discusses strengths, limitations and future directions.

2

Mathematical Framework

In the first part of this thesis, we aim to offer a clear theoretical foundation for the tools that will be used throughout this work. The chapter is structured as follows. We begin in Section 2.1 by reviewing the necessary algebraic background: tensor products and tensor algebras, which provide the natural space for signatures and their operations. In Section 2.2, we introduce the analytical framework for paths of varying regularity, together with the corresponding notions of integration, including stochastic integration for semimartingales. These ingredients allow us, in Section 2.3, to rigorously define the signature of a path and discuss its main properties. In Section 2.4, we extend these ideas to stochastic processes by introducing the expected signature of a path and deriving the partial differential equation that characterizes its evolution. Finally, in Section 2.5 we present several stochastic diffusion models commonly used in finance that will serve as test cases throughout this thesis.

2.1. Preliminaries

This section recalls the main algebraic tools, tensor products and tensor algebras, that provide the natural space in which signatures are defined[1].

2.1.1. Tensor products.

When studying path signatures we often encounter iterated integrals, which naturally lead to objects that are multilinear in their increments. To accommodate such objects, we need a space that records all possible multilinear interactions between vectors of a given space. This is precisely what the *tensor product* provides. Let E be a real vector space, generally assumed to be $E = \mathbb{R}^d$ throughout this work. Most of the following theory extends to infinite-dimensional spaces [18].

Definition 1 (Tensor product)**.** *Let E and F be finite-dimensional vector spaces with respective bases B_E and B_F. Their* tensor product*, denoted by $E \otimes F$, is another vector space together with a bilinear map*

$$\varphi : E \times F \to E \otimes F, \qquad (e, f) \mapsto e \otimes f,$$

with the property that any bilinear map out of $E \times F$ factors uniquely through φ. Concretely, if $(e_i)_{i \in B_E}$ and $(f_j)_{j \in B_F}$ are bases of E and F, then the family

$$\{\, e_i \otimes f_j \;\mid\; i \in B_E,\, j \in B_F \,\}$$

is a basis of $E \otimes F$. For a complete discussion concerning the existence and uniqueness of tensor products, we refer to [17].

[1] We assume that the reader is familiar with basic notions of linear and multilinear algebra, such as vector spaces, linear maps, and bilinear forms. The definitions and conventions adopted here follow standard presentations in the literature on tensor analysis and rough path theory (see, e.g., [6, 7, 17]), with minor notational adaptations for consistency.

The elements of tensor product spaces are called *tensors*, and each of them can be written as a finite linear combination of expressions of the form $e_i \otimes f_j$. Tensors generalize familiar objects: scalars are tensors of order 0, vectors of order 1, and matrices of order 2. More generally, the *order* (or rank) of a tensor corresponds to the number of vector spaces being tensored together. For instance, if E has dimension d with canonical basis $(e_1, \dots, e_d)$, then each index $j \in \{1, \dots, d\}$ designates one of the coordinate directions of E. A basis element of $E \otimes F$ is then specified by choosing a direction from E and a direction from F, e.g. $e_i \otimes f_j$. A key feature of tensor products is that they are generally *non-commutative*: the order in which vectors appear matters. The construction of a vector space $E \otimes F$ can be extended by tensoring one vector space with itself several times. This allows us to describe multilinear interactions of arbitrary order.

Definition 2 (n-fold tensor products.)**.** *Let $E = \mathbb{R}^d$ be a vector space of dimension $d \in \mathbb{N}$. For each $n \in \mathbb{N}$, the* n-fold tensor product *of E with itself is defined as*

$$E^{\otimes n} = \underbrace{E \otimes \ldots \otimes E}_{n \text{ times}},$$

with the convention $E^{\otimes 0} := \mathbb{R}$.

A basis of $E^{\otimes n}$ is obtained by tensoring together n basis vectors of E. If $(e_1, \dots, e_d)$ is the canonical basis of E, then a basis element of $E^{\otimes n}$ has the form

$$e_{i_1} \otimes e_{i_2} \otimes \cdots \otimes e_{i_n}, \qquad i_k \in \{1, \dots, d\}.$$

Thus $E^{\otimes n}$ has dimension d^n, and its elements are linear combinations of such elementary tensors. The spaces $E^{\otimes n}$, for $n \geq 0$, play a central role in the algebraic description of signatures, as we will see in Section 2.3.

2.1.2. Word notation

To work effectively with tensor product spaces and their elements, it is convenient to have a simple system for indexing their basis vectors. The natural language for this purpose is that of *alphabets* and *words*.

Definition 3 (Alphabet and Words)**.** *An* alphabet *is a finite set of symbols that we denote by $\Sigma = \{1, \dots, d\}$, where each symbol corresponds to one coordinate direction of $E = \mathbb{R}^d$. A* word I of length n *over Σ is a sequence*

$$I = (i_1, \dots, i_n), \qquad i_1, \dots, i_n \in \Sigma.$$

The length of a word I is denoted by $|I|$ and corresponds to the order of the associated tensor; the empty word, denoted by $\varnothing$, is the unique word of length 0. For each $n \geq 0$, we denote by Σ^n the set of all words of length n, and by $\Sigma^+ := \bigcup_{n \geq 0} \Sigma^n$ the set of all finite words.

Words provide a convenient way to label basis elements of tensor product spaces. Let $(e_1, \dots, e_d)$ denote the canonical basis of E. Each word $I = (i_1, \dots, i_n) \in \Sigma^n$ determines a tensor basis element

$$e_I := e_{i_1} \otimes \cdots \otimes e_{i_n} \in E^{\otimes n}.$$

The collection $\{e_I : I \in \Sigma^n\}$ forms the canonical basis of $E^{\otimes n}$, and by convention $e_\varnothing := 1 \in \mathbb{R}$ corresponds to the empty word. For two words $I, J \in \Sigma^+$, the tensor product of their corresponding basis elements satisfies

$$e_I \otimes e_J = e_{IJ} \in E^{\otimes(|I|+|J|)},$$

meaning that tensor multiplication is encoded algebraically as concatenation of the associated words.

2.1.3. Tensor algebras

Collecting all n-fold tensor products together leads us naturally to the construction of the tensor algebra. The tensor algebra $T(E)$ is the standard algebraic construction used in multilinear algebra that gathers tensors of all orders into a single graded vector space:

$$T(E) := \bigoplus_{n \geq 0} E^{\otimes n}.$$

It consists of finite linear combinations of tensors of arbitrary order. In the context of path signatures, it is customary to work instead with its formal completion, the *extended tensor algebra* $T((E))$, which allows infinite series and provides a natural algebraic setting for their representation.

Definition 4 (Extended Tensor Algebra)**.** *Let E be a real vector space. The extended tensor algebra of E, denoted by $T((E))$, is the algebra obtained by endowing the space of formal series of tensors*

$$T((E)) := \Big\{ a = (a_0, a_1, a_2, \dots) \;\Big|\; a_n = \sum_{|I|=n} a_I e_I \in E^{\otimes n},\ a_I \in \mathbb{R},\ n \geq 0 \Big\}$$

with the following operations. Let $a = (a_0, a_1, \dots)$ and $b = (b_0, b_1, \dots)$ be two elements of $T((E))$:

- *Addition:* $a + b := (a_0 + b_0,\ a_1 + b_1,\ \dots)$,
- *Product:* $a \otimes b := (c_0, c_1, c_2, \dots)$, *where* $c_n := \sum_{k=0}^{n} a_k \otimes b_{n-k}$,
- *Scalar multiplication:* $\lambda a := (\lambda a_0,\ \lambda a_1,\ \dots), \quad \lambda \in \mathbb{R}$.

With these operations, $(T((E)), +, \otimes)$ forms a real, associative, non-commutative algebra. The element

$$\mathbf{1} := (1, 0, 0, \dots)$$

is the multiplication identity: for any element $A \in T((\mathbb{R}^d))$,

$$A \otimes \mathbf{1} = \mathbf{1} \otimes A = A.$$

Each element of $T((E))$ can be written as a formal and possibly infinite linear combination

$$a = \sum_{I \in \Sigma^+} a_I \, e_I,$$

where $a_I \in \mathbb{R}$ and $e_I = e_{i_1} \otimes \cdots \otimes e_{i_n}$ for $I = (i_1, \dots, i_n)$, and multiplication corresponds to concatenation of words: $e_I \otimes e_J = e_{IJ}$.

In practical computations, it is, of course, unfeasible to retain all (infinitely many) tensor levels of an element in $T((E))$: the algebra is thus truncated at a level m of choice.

Definition 5 (Truncated Tensor Algebra)**.** *Let $k \in \mathbb{N}$. The* truncated tensor algebra *of order m is defined as the space of finite sequences of length $m + 1$:*

$$T^k(E) = \big\{ a := (a_0, a_1, \dots, a_k) \,\big|\, a_m \in E^{\otimes m},\ 0 \leq m \leq k \big\},$$

where all higher-order terms (i.e. $k > m$) vanish identically. In relation to the extended algebra $T((E))$, it is defined as the quotient algebra

$$T^k(E) = T((E))/B_k,$$

where $B_k = \{a = (a_0, a_1, \dots) \mid a_0 = \dots = a_k = 0\}$ is the formal series with no monomials of degree less than or equal to k.

This is a finite-dimensional algebra, which corresponds to the space of non-commutative polynomials of degree at most k.

To evaluate or manipulate elements of $T((E))$, it is often useful to consider their dual space of linear functionals. For a finite-dimensional vector space $E = \mathbb{R}^d$ with canonical basis $(e_1, \dots, e_d)$, denote by E^* the dual space of linear maps $E \to \mathbb{R}$ with dual basis $(e_1^*, \dots, e_d^*)$. This construction extends naturally to tensor powers: for each $n \geq 0$ there is a canonical isomorphism

$$(E^{\otimes n})^* \;\cong\; (E^*)^{\otimes n}$$

that identifies each tensor $e_I^* := e_{i_1}^* \otimes \cdots \otimes e_{i_n}^*$ with the corresponding basis element $e_I := e_{i_1} \otimes \cdots \otimes e_{i_n}$, of $E^{\otimes n}$ so that $\langle e_I^*, e_J \rangle = \delta_{I,J}$.

Definition 6 (Dual Tensor Algebra)**.** *The* dual tensor algebra *associated with E is defined as*

$$T((E))^* := \bigoplus_{n=0}^{\infty} (E^{\otimes n})^* \;\cong\; \bigoplus_{n=0}^{\infty} (E^*)^{\otimes n}.$$

Elements of $T((E))^*$ have finite support (i.e., only finitely many nonzero components), and each e_I^* acts on $a = (a_0, a_1, \dots) \in T((E))$ by

$$e_I^*(a) := e_I^*(a_n), \qquad |I| = n,$$

that is, it extracts the coefficient of the tensor basis element e_I in the component of degree $|\, I \,|$. Thus, the family $\{e_I^*\}_{I \in \Sigma^+}$ forms the canonical basis of $T((E))^*$, dual to the tensor basis $\{e_I\}_{I \in \Sigma^+}$ of $T((E))$.

Any linear functional $\ell \in T((E))^*$ can therefore be expressed as a finite linear combination

$$\ell = \sum_{I \in \Sigma^+} \ell_I \, e_I^*, \qquad \ell_I \in \mathbb{R},$$

and its action on $a = \sum_{J \in \Sigma^+} a_J e_J \in T((E))$ is given by

$$\ell(a) = \sum_{I \in \Sigma^+} \ell_I \, a_I.$$

Among all linear functionals, a particularly important class consists of those that extract a single tensor coefficient.

Definition 7 (Word Projection)**.** *Fix the canonical basis $(e_1, \dots, e_d)$ of $E = \mathbb{R}^d$ and its dual basis $(e_1^*, \dots, e_d^*)$ of $(E)^*$. For each word $I = (i_1, \dots, i_n) \in \Sigma^n$, we define $e_I := e_{i_1} \otimes \cdots \otimes e_{i_n}$ and its dual $e_I^* := e_{i_1}^* \otimes \cdots \otimes e_{i_n}^*$. The* word projection *associated with I is the linear functional*

$$\pi^{(I)} : T((E)) \to \mathbb{R}, \qquad \pi^{(I)}(a) := \langle e_I^*, a \rangle,$$

which extracts the coefficient of the basis tensor e_I from a. Equivalently, if $a = \sum_{J \in \Sigma^+} a_J e_J$, then $\pi^{(I)}(a) = a_I$.

With abuse of notation, for $a_m \in E^{\otimes m}$, $m \in \mathbb{N}$, $\pi^{(I)}(a_m)$ extracts the coefficient from a_m corresponding to the word I, if $|I| = m$.

2.2. Paths, variation and integration

Paths and variation A path is a continuous mapping

$$X : [0, T] \to \mathbb{R}^d, \qquad t \mapsto X_t = \big(X_t^{(1)}, \dots, X_t^{(d)}\big),$$

where each component $X^{(i)} : [0, T] \to \mathbb{R}$ represents the evolution of the i-th coordinate of the path, $i \in \{1, \dots d\}$. To introduce the signature of a path, we first need to clarify which paths we can integrate against. Classical Riemann integration requires smoothness or bounded variation of the integrator, but many paths of interest in finance are highly irregular. A natural way to quantify irregularity, or *roughness*, of a path X is through its p-variation

$$\|X\|_{p\text{-var};[0,T]} := \left(\sup_{\{t_i\} \in \Pi} \sum_i |X_{t_i} - X_{t_{i-1}}|^p \right)^{1/p}, \tag{2.1}$$

where the supremum runs over all finite partitions $\Pi = \{0 = t_0 < \ldots < t_n = T\}$ of $[0, T]$. If (2.1) is finite, we say that X has *finite p-variation* and denote the space of paths of finite p-variation as:

$$\mathcal{V}^p([0, T]; \mathbb{R}^d) := \{X : [0, T] \to \mathbb{R}^d : X \text{ continuous}, \, \|X\|_{p\text{-var};[0,T]} < \infty\}. \tag{2.2}$$

Throughout this work, we will mainly consider the time-augmented path $\widehat{X} := (t, X_t) \in \mathbb{R}^{1+d}$, whose introduction will be motivated later in Subsection 2.3.3.6. We define the space of time-augmented paths of finite p-variation as

$$\mathcal{A}^p([0,T];\mathbb{R}^d) := \left\{\widehat{X} = (t, X_t) : X_t \in \mathcal{V}^p([0,T];\mathbb{R}^d)\right\}. \tag{2.3}$$

Since $t \mapsto t$ has bounded variation, $\widehat{X}$ meets the same integration requirements as X, and $\widehat{X} \in \mathcal{A}^p$ whenever $X \in \mathcal{V}^p$.

The p-variation of a path quantifies its roughness and also determines which integration framework applies. Since signatures are defined through iterated integrals, the admissible integration theory depends on the value of p. We distinguish three cases.

$p = 1$**: Paths of bounded variation.** In this case increments are summable, and the integral $\int f\, dX$ can be defined in the classical Riemann-Stieltjes sense.

$p \in (1,2)$**: Intermediate roughness.** Here the integrator is not of bounded variation, but Young's theorem ensures that $\int Y\, dX$ exists whenever the integrand has complementary variation, i.e. $X \in \mathcal{V}^p$ and $Y \in \mathcal{V}^q$ with $1/p + 1/q > 1$.

$p \geq 2$**: Rough paths.** These fail both the bounded variation and Young's conditions. To recover a meaningful integration theory, one must *enhance* the path X with its iterated integrals, obtaining the so-called *enhanced path* $(X, \mathbb{X})$. These iterated integrals are defined as the limit of their corresponding piece-wise smooth approximations, ensuring algebraic consistency (through Chen's relation, see Theorem 2.4) and stability of the resulting integrals.

The last regime, first developed by Lyons in [5], forms the basis of *rough path theory*, which provides a deterministic framework for integration and differential equations driven by highly irregular paths. This theory underlies the modern definition of the path signature, since signatures are constructed from the iterated integrals appearing in a rough path enhancement.

The rough path framework can be viewed as a deterministic counterpart of stochastic integration: it provides a pathwise construction of integrals that coincides with the *Stratonovich integral* when the driving path is a semimartingale. In this sense, rough path theory unifies deterministic and stochastic integration under a common geometric perspective.

In this work, we will not require the full formalism of rough path theory. The driving processes we consider are Brownian motions, hence semimartingales, for which integration is already well-defined in the stochastic sense (Itô or Stratonovich). Nevertheless, the rough path viewpoint remains fundamental: it explains why signatures form a natural language for describing irregular paths.

Stochastic integration We now recall the stochastic integration framework relevant for this work. We consider a filtered probability space $(\Omega, \mathcal{F}, (\mathcal{F}_t)_{t\in[0,T]}, \mathbb{P})$ satisfying the standard conditions[2]. Let X be a continuous semimartingale and Y a predictable process with suitable integrability[3]. There are two main ways to define stochastic integrals: Itô and Stratonovich, which we summarize below together with their relationship.

Definition 8 (Itô integral)**.** *The Itô integral of Y with respect to X is defined as the limit in probability of left-point Riemann sums along partitions $\Pi = \{0 = t_0 < \cdots < t_n = t\}$ with $|\Pi| \to 0$:*

$$\int_0^t Y_s \circ dX_s := \lim_{|\Pi|\to 0} \sum_{[t_i,t_{i+1}]\in\Pi} Y_{t_i}\left(X_{t_{i+1}} - X_{t_i}\right).$$

This integral is adapted to the filtration and defines a martingale under mild conditions; Itô's lemma then modifies the chain rule by adding quadratic-variation terms.

[2]The filtration is assumed to be complete, i.e. $\mathcal{F}_0$ contains all $\mathbb{Q}$-null sets of $\mathcal{F}$, and right-continuous, i.e. $\mathcal{F}_t = \bigcap_{s>t} \mathcal{F}_s$ for all $t \geq 0$. These are known as the *usual conditions*.

[3]For basic definitions and properties of semimartingales, we refer the reader to standard references such as [19, 20].

Definition 9 (Stratonovich integral)**.** *The Stratonovich integral of Y with respect to X is defined as the limit in probability of symmetric Riemann sums:*

$$\int_0^t Y_s \, dX_s := \lim_{|\Pi|\to 0} \sum_{[t_i,t_{i+1}]\in\Pi} \frac{Y_{t_i} + Y_{t_{i+1}}}{2} \left(X_{t_{i+1}} - X_{t_i}\right).$$

It preserves the classical chain rule and coincides with limits of smooth approximations.

Proposition 2.1 (Stratonovich-Itô relation)**.** *If X and Y are continuous semimartingales, then for all $t \in [0,T]$,*

$$\int_0^t Y_s \, dX_s = \int_0^t Y_s \circ dX_s + \tfrac{1}{2}\,[Y,X]_t,$$

where $[Y,X]_t$ denotes the quadratic covariation between X and Y.

With a suitable notion of integration at hand, we can now define *iterated integrals*, which play a central role in the construction of the signature, and describe how a path interacts with itself at higher orders.

Definition 10 (Iterated integrals)**.** *Given a continuous path $X : [0,T] \to \mathbb{R}^d$, its n-th iterated integral up to time $t \le T$ is defined as*

$$\int_{0<u_1<\cdots<u_n<t} dX_{u_1} \otimes \cdots \otimes dX_{u_n} \;\in\; (\mathbb{R}^d)^{\otimes n}.$$

Writing $dX_u = \sum_{i=1}^d dX_u^{(i)}\, e_i$, where $\{e_1,\dots,e_d\}$ is the canonical basis of $\mathbb{R}^d$, this integral expands in coordinates as

$$\int_{0<u_1<\cdots<u_n<t} dX_{u_1} \otimes \cdots \otimes dX_{u_n} = \sum_{i_1,\dots,i_n=1}^{d} \left(\int_{0<u_1<\cdots<u_n<t} dX_{u_1}^{(i_1)} \cdots dX_{u_n}^{(i_n)} \right) e_{i_1} \otimes \cdots \otimes e_{i_n}.$$

At level $n=1$, an iterated integral corresponds to the path increment $X_t - X_0$, while for $n=2$, the coefficients encode signed areas traversed by the path. The tensor notation provides a compact way to represent all the coordinate-wise iterated integrals simultaneously and distinguishes the order of integration: for instance, $e_i \otimes e_j$ corresponds to $\int dX^{(i)} dX^{(j)}$, which generally differs from $e_j \otimes e_i$.

2.3. Signatures

The signature of a path collects all of its iterated integrals into a single algebraic object taking values in the extended tensor algebra. It provides a hierarchical description of the path through all orders of interaction.

Definition 11 (Signature)**.** *Let $X \in \mathcal{V}^p([0,T];\mathbb{R}^d)$ be a continuous path of finite p-variation with components $X = (X^{(1)},\dots,X^{(d)})$. For $0 \le s < t \le T$, the signature of X over $[s,t]$ is the formal series*

$$S(X)_{s,t} := \big(1,\, S_1(X)_{s,t},\, S_2(X)_{s,t},\, \dots, S_n(X)_{s,t}, \dots\big) \in T((\mathbb{R}^d)),$$

where each level is defined as

$$S_n(X)_{s,t} := \int_{s<u_1<\cdots<u_n<t} dX_{u_1} \otimes \cdots \otimes dX_{u_n} \in (\mathbb{R}^d)^{\otimes n},$$

with $n \ge 1$, $S_0(X)_{s,t} := 1$, and where $T((\mathbb{R}^d))$ is the extended tensor algebra.

Remark 2.2. *In the setting of this thesis, when X is a stochastic process, the definition of the iterated integrals is understood path-wise and in the Stratonovich sense. Although this is not a conventional choice in financial mathematics, it is required to ensure that the signature retains its multiplicative property under path concatenation (see Chen's identity, Theorem 2.4).*

Applying the word projection $\pi^{(I)}(\cdot)$ associated with a word $I = (i_1, \dots, i_n)$ recovers the corresponding iterated integral:

$$\pi^{(I)}\big(S(X)_{s,t}\big) = \int_{s<u_1<\cdots<u_n<t} dX_{u_1}^{(i_1)} \cdots dX_{u_n}^{(i_n)} \in \mathbb{R}. \tag{2.4}$$

Using this notation, each level of the signature decomposes as

$$S_n(X)_{s,t} = \sum_{|I|=n} \pi^{(I)}\big(S(X)_{s,t}\big)\, e_I,$$

where the length $|I| = n$ of the word I corresponds to the order of the integral.

In practice, it is impossible to use the full (infinite-dimensional) signature. To make computations tractable we will use its truncated version, retaining only iterated tensor levels up to a fixed order $k \in \mathbb{N}$:

$$S(X)_{s,t}^k := \big(1,\, S_1(X)_{s,t},\, S_2(X)_{s,t}, \dots, S_k(X)_{s,t}\big) \in T^k(\mathbb{R}^d),$$

where $T^k(\mathbb{R}^d)$ is the truncated tensor algebra.

Remark 2.3. *Unless otherwise specified, the notation $S(X)$ denotes the signature of X over the interval $[0, T]$, that is, $S(X)_{0,T}$. Similarly, the truncated signature $S^k(X)$ denotes $S^k(X)_{0,T}$.*

Figure 2.1 offers a visual and intuitive representation of the first three levels of the signature, where each block represents tensors of increasing order: a point for the scalar term, a line (vector) for first-level increments, a square (matrix) for second-level interactions, a cube (third-order tensor) for third-level iterated integrals.

Figure 2.1: Visualization of the signature represented as a collection of tensors of increasing dimensions. Reproduced from [21].

The link between the algebraic structure of signatures and their analytical meaning emerges clearly in the context of differential equations, as illustrated in the following section.

2.3.1. Motivation Picard iterations

Iterated integrals, and consequently signature terms, naturally emerge in the study of ordinary differential equations (ODEs) driven by a signal [22, Section 1.2.3]. We consider the controlled ODE

$$dY_t = \sum_{i=1}^{d} V_i(Y_t)\, dX_t^{(i)}, \qquad Y_0 = y_0 \in \mathbb{R}^e, \tag{2.5}$$

where $X : [0, T] \to \mathbb{R}^d$ denotes the driving path and $Y : [0, T] \to \mathbb{R}^e$ the state variable. The map

$$V = (V_1, \dots, V_d) : \mathbb{R}^e \to \mathcal{L}(\mathbb{R}^d, \mathbb{R}^e)$$

consists of d smooth vector fields V_i which can be written in coordinates as $V_i = (V_i^{(1)}, \dots, V_i^{(e)})$, $i = 1, \dots, d$. Here, $\mathcal{L}(\mathbb{R}^d, \mathbb{R}^e)$ denotes the space of linear maps from $\mathbb{R}^d$ to $\mathbb{R}^e$, which can be identified with the set of real $e \times d$ matrices. If the driving path X has bounded variation and Y is continuous, then the integral $\int_0^T V(Y_t)\, dX_t$ is well defined in the Riemann-Stieltjes sense. Under standard assumptions, namely, that the vector fields V_i are globally Lipschitz, the Picard-Lindelöf theorem[4] guarantees the existence and uniqueness of a global solution Y_t in $[0, T]$. If the derivatives of V_i are not bounded on all $\mathbb{R}^e$, the result still holds locally, up to a possible explosion time. Each vector field V_i acts on the state

[4] See [23] for the proof.

variable Y through the ODE (2.5). To make this action explicit, it is useful to view every V_i as a first order differential operator acting on smooth functions $f : \mathbb{R}^e \to \mathbb{R}$:

$$(V_i\, f)(y) = \sum_{m=1}^{e} V_i^m(y) \frac{\partial f}{\partial y^m}(y)\,, \quad i = 1, \dots, d.$$

This operator viewpoint allows us to compose vector fields and obtain, for instance, expressions such as $V_{i_1} V_{i_2}\, I(y)$, where $I : \mathbb{R}^e \to \mathbb{R}^e$, $I(y) = y$, denotes the identity function; such compositions will later appear in the expansion of the solution of the ODE.

The integral form of (2.5) reads:

$$Y_t = y_0 + \sum_{i=1}^{d} \int_0^t V_i(Y_s)\, dX_s^{(i)},$$

which expresses the evolution of Y_t in terms of the increments of the driving paths X. A convenient way to construct its solution is through Picard iterations, the method of successive approximations on which the Picard-Lindelöf theorem is based. Starting from $Y_t^{[0]} = y_0$, we define recursively the k-th Picard iteration as

$$Y_t^{[k]} = y_0 + \sum_{i=1}^{d} \int_0^t V_i(Y_s^{[k-1]})\, dX_s^{(i)}, \quad k \geq 1.$$

Under the assumptions stated earlier, the Picard-Lindelöf theorem guarantees that the sequence $(Y_t^{[k]})_{k\geq 1}$ converges uniformly on $[0, T]$ to the unique solution Y_t of (2.5), that is, $Y_t = \lim_{k\to\infty} Y_t^{[k]}$. Expanding the first iterations illustrates how iterated integrals of the driving signal naturally arise. At the first step,

$$Y_t^{[1]} = y_0 + \sum_{i=1}^{d} V_i\, I(y_0) \int_0^t dX_s^{(i)},$$

and after a second iteration,

$$Y_t^{[2]} = y_0 + \sum_{i=1}^{d} V_i\, I(y_0) \int_0^t dX_s^{(i)} + \sum_{i,j=1}^{d} V_j V_i\, I(y_0) \int_{0<u_1<u_2<t} dX_{u_1}^{(i)} dX_{u_2}^{(j)}.$$

Proceeding inductively, one sees that every term of the Picard expansion involves an iterated integral of the driving signal X:

$$Y_t^{[n]} = y_0 + \sum_{i=1}^{d} \int_0^t V_i(Y_s^{[n-1]}) dX_s \tag{2.6}$$

$$= y_0 + \sum_{k=1}^{n} \sum_{i_1,\dots i_k=1}^{d} (V_{i_k} \dots V_{i_1}\, I)\,(y_0) \underbrace{\int_{0<u_1<\dots<u_k<t} dX_{u_1}^{(i_1)} \dots dX_{u_k}^{(i_k)}}_{\text{Iterated integrals of the driving signal } X}. \tag{2.7}$$

Hence, the entire solution Y_t is determined by a linear combination of the collection of iterated integrals of the driving signal, i.e. by the signature of X. This means that the signature summarizes the effect of a path X on an output path Y, when Y is driven by X. Extrapolating this observation beyond the mathematical setting, it follows that whenever an output depends on the evolution of an input path, its behaviour can be expressed in terms of the signature of the driving signal.

2.3.2. Geometric interpretation of the signature

Although the definition of the signature is algebraic in nature, its components admit an informative geometric interpretation that reflects how the path evolves through its coordinate directions. Each iterated integral captures how the coordinates of the path interact over time, and its meaning depends on the order of integration n, the dimension of the path d, and the pattern of indices appearing in the corresponding word $(i_1, \dots, i_n)$. We can distinguish three main cases, following the intuitive exposition in [24, Section 4].

2.3.2.1. Case 1: All indices identical

When all integrators correspond to the same component, that is, $i_1 = \cdots = i_n = i$, the corresponding iterated integral reduces to

$$\pi^{(i,\dots,i)}(S(X)) = \int_{a<t_1<\cdots<t_n<b} dX_{t_1}^{(i)} \cdots dX_{t_n}^{(i)} = \frac{1}{n!}(X_b^{(i)} - X_a^{(i)})^n.$$

These terms depend only on the total displacement of the i-th coordinate and can be viewed as the signed volume of an n-dimensional simplex whose side length is equal to $|X_b^{(i)} - X_a^{(i)}|$.

2.3.2.2. Case 2: All indices distinct ($n \leq d$)

When all indices $i_1, \dots, i_n$ are distinct, the iterated integrals encode the relative evolution of several coordinates. For $n = 2$, they correspond to oriented areas in the planes $(X^{(i_1)}, X^{(i_2)})$; for $n = 3$, to oriented volumes in three-dimensional space; and, more generally, for $n \leq d$, to oriented hyper-volumes spanned by the trajectory across n coordinate directions. The sign of each term reflects the temporal order in which the path traverses these coordinates. These components capture the geometric shape of the path and how it fills space.

2.3.2.3. Case 3: Repeated indices ($n > d$, partial repetitions).

For higher levels, repetitions of indices become unavoidable. In this regime, iterated integrals no longer correspond to new geometric dimensions but rather describe higher-order interactions between coordinates, including repetitions of the same component. Repeated indices such as (i, i, j) or (i, j, j) measure how motion in one coordinate repeatedly intertwines with that of another, encoding the sequential structure of the path over time. Thus, higher-order terms enrich the description of the path by capturing the temporal ordering and combinatorial structure of the trajectory.

2.3.3. Properties of the Signature

The signature enjoys a number of fundamental properties that reflect its algebraic and geometric structure. We begin with Chen's identity [4], which provides the algebraic rule governing concatenations of paths.

2.3.3.1. Chen's identity

We first recall the notion of concatenation of paths. Given two paths $X : [a, b] \to \mathbb{R}^d$ and $Y : [b, c] \to \mathbb{R}^d$, their concatenation is defined as the path $X * Y : [a, c] \to \mathbb{R}^d$ for which $(X * Y)_t = X_t$ for $t \in [a, b]$ and $(X * Y)_t = X_b + (Y_t - Y_b)$ for $t \in [b, c]$. This operation geometrically corresponds to joining the end of a path with the beginning of another one.

Theorem 2.4 (Chen's identity)**.** *For two paths $X : [a, b] \to \mathbb{R}^d$ and $Y : [b, c] \to \mathbb{R}^d$,*

$$S(X * Y)_{a,c} = S(X)_{a,b} \otimes S(Y)_{b,c}. \tag{2.8}$$

This identity shows that the signature behaves multiplicatively under concatenation of paths, which is an essential feature that supports many of its algebraic and practical applications. For stochastic processes, the use of the Stratonovich integral is essential to preserve (2.8). This choice ensures that the signature retains its multiplicative property under concatenation of paths, consistent with the classical chain rule. In contrast, Itô integrals introduce additional correction terms arising from quadratic variation, which break the additivity of iterated integrals across intervals and hence destroy the algebraic structure encoded by Chen's identity.

One of the applications of Chen's identity is in the computational approximation of path signatures. For a d-dimensional linear path

$$X_t = \begin{pmatrix} a_1 + b_1 t \\ \vdots \\ a_d + b_d t \end{pmatrix}, \qquad t \in [0, T],$$

and for a word $I = (i_1, \ldots, i_k) \in \Sigma^+$, the corresponding signature coefficient can be computed explicitly as

$$S^{(i_1,\ldots,i_k)}(X) = \int_{0 \le u_1 < \cdots < u_k \le t} dX_{u_1}^{(i_1)} \cdots dX_{u_k}^{(i_k)} = \frac{b_{i_1} \cdots b_{i_k}}{k!}.$$

If we approximate a general path by a sufficiently fine piecewise-linear approximation, the signature of the entire path can be reconstructed by concatenating the signatures of its linear segments through Chen's identity. This principle underlies the algorithmic design of signature computation libraries such as `iisignature` [25] and `signatory` [26].

2.3.3.2. Invariances of the signature

A key geometric feature of the signature is its invariance under several transformations of the underlying path. These invariances reflect the fact that the signature depends only on the shape or geometry of the path, rather than on superficial aspects such as its parametrization or starting point. While these properties are advantageous in some contexts, in many others it is useful to recover the information they suppress, which can be achieved through suitable augmentations. We now describe the main invariances of the signature.

2.3.3.3. Invariance under reparametrization

The signature is invariant under time reparametrizations: it depends only on the geometric trace of the path and not on the speed at which it is traversed. For clarity, we illustrate this property in the smooth case. Formally, a function $\psi : [a, b] \to [a, b]$ is called a *reparametrization* if it is surjective, continuous, and non-decreasing. Given two continuous paths $X, Y : [a, b] \to \mathbb{R}^d$, the reparametrized paths $\tilde{X}_t = X_{\psi(t)}$, $\tilde{Y}_t = Y_{\psi(t)}$ trace the same geometric curves respectively, possibly at a different speed. By the chain rule,

$$\dot{\tilde{X}}_t = \dot{X}_{\psi(t)}\, \dot{\psi}(t),$$

and thus, for any component-wise path integral,

$$\int_a^b \tilde{Y}_t \, \mathrm{d}\tilde{X}_t = \int_a^b Y_{\psi(t)}\, \dot{X}_{\psi(t)}\, \dot{\psi}(t)\, \mathrm{d}t = \int_a^b Y_u \, \mathrm{d}X_u,$$

where the last equality follows from the change of variable $u = \psi(t)$. This shows that path integrals are invariant under time reparametrizations. Consequently, since each term of the signature is defined as an iterated integral of the components of paths, we have $S(X) = S(\tilde{X})$.

More generally, the invariance of the signature under time reparametrization holds for rough or non-smooth paths as well. This extension is non-trivial and is established in the general setting in [7].

2.3.3.4. Tree-like equivalence

Another type of invariance arises from the notion of tree-like equivalence, first introduced in [1]. This equivalence class contains all the paths that differ by tree-like segments, whose iterated integrals vanish, and therefore do not alter the value of the signature. Formally, a segment Y is said to be tree-like if there exist functions $h : [0, T] \to \mathbb{R}_+$ and $\varphi : \mathbb{R}+ \to \mathbb{R}^d$ such that $Y_t = \varphi(h_t)$, where h increases and decreases as if it traversed the branches of a tree, with $h(0) = h(T) = 0$. Consequently, the signature of a tree-like path is trivial:

$$S(Y) = \mathbf{1}.$$

If we consider a path X and concatenate it with a tree-like segment Y, we obtain the path $X^\circ = X * Y$, which is tree-like invariant to X. Indeed, according to Chen's identity 2.4, the signature of the concatenated path X° satisfies

$$S(X^\circ) = S(X * Y) = S(X) \otimes \mathbf{1} = S(X).$$

2.3.3.5. Characterization

The signature is a powerful characterization tool, widely used for its ability to encode the essential geometric information of a path. Because it represents the path in an algebraic form, a natural question arises: to what extent does this representation retain all the information about the original path? The characterization capability of the signature is formalized in the following theorem [1] for paths of finite p-variation for $p = 1$ and extended to the case of $p > 2$ in [27].

Theorem 2.5 (Characterization)**.** *Let $X : [a, b] \to \mathbb{R}^d$ be a d-dimensional p-finite variation path. Then the signature $S(X)$ uniquely characterizes X up to tree-like equivalences.*

This means that, modulo tree-like segments, the signature provides a complete and lossless representation of a path. In practical applications, however, it is often desirable to obtain strict injectivity without equivalence classes. This can be achieved through time augmentation.

2.3.3.6. Completing the characterization: time augmentation

Although the presence of tree-like equivalences prevents the signature from being strictly injective, this limitation can be overcome by augmenting the path with an additional monotone coordinate, typically representing time. More precisely, we consider the time-augmented path

$$\widehat{X} := (t, X_t) \in \mathbb{R}^{d+1}.$$

Because the time component is strictly increasing, any tree-like retracing in the original spatial coordinates would necessarily alter the time coordinate, thereby preventing the cancellation responsible for the tree-like invariance. For the same reason, this augmentation also removes invariance under time reparameterization. As a result, the signature of $\widehat{X}$, $S(\widehat{X})$, uniquely characterizes the path $\widehat{X}$ (and thus X) without equivalence classes. It is important to remember that having a monotone coordinate is a sufficient condition to ensure complete characterization, while a necessary condition can be found in [1].

From a practical perspective, this augmentation plays a central role in our later experiments, where all stochastic processes are represented in their time-augmented form to guarantee full correspondence between a path and its signature.

2.3.3.7. Other types of augmentations

Just as the time augmentation restores injectivity by breaking tree-like equivalence, other augmentations have been proposed to selectively remove or restore specific invariance properties of the signature. For instance, the invisibility-reset transformation [10, 28] and base point augmentations [26], restore sensitivity to translation by adding, respectively, an additional channel or a fixed starting point. In high-dimensional settings, projection-based and learned augmentations [29, 30] are used to modify the path representation before computing the signature, either to reduce dimensionality or to adapt to data-driven tasks. Beyond correction of invariances, some augmentations serve to enrich the path information. The lead-lag transformation, introduced in [22] and further studied in [31], duplicates each coordinate into a leading and a lagging version, generating cross-iterated integrals that encode quadratic variation and temporal ordering information.

2.3.3.8. Universal approximation

Having established that the (time-augmented) signature uniquely characterizes paths, we now turn to its approximation capabilities. The collection of linear functionals acting on the signature forms an algebra that is dense in the space of continuous functionals on compact subsets of the path space. This result, often referred to as the *universal approximation theorem for signatures*, is particularly relevant in the context of derivative pricing, as it implies that any continuous path-dependent payoff, if considered as a functional of the underlying price trajectory, can be approximated using linear functionals of the signature. This is formalized in the following theorem.

Theorem 2.6 (Universal Approximation Theorem (UAT))**.** *Let $p \geq 1$ and let $f : \mathcal{A}^p([0, T]; \mathbb{R}^d) \to \mathbb{R}$ be a continuous functional. Let $K \subset \mathcal{A}^p([0, T]; \mathbb{R}^d)$ be a compact set. Assume that for each $\widehat{X} \in K$, the signature $S(\widehat{X})$ is well defined. Then, for any $\varepsilon > 0$, there exists $\ell \in T((\mathbb{R}^{d+1}))^*$ such that*

$$| f(\widehat{X}) - \ell \left(S(\widehat{X}) \right) | < \varepsilon.$$

Sketch of proof. Since paths in $\mathcal{A}^p([0,T];\mathbb{R}^d)$ are time-augmented and hence uniquely represented by their signatures, the function f induces a continuous function $\mathcal{F}$ on the image $S(K)$ such that $f(\widehat{X}) = \mathcal{F}(S(\widehat{X}))$. The set of linear functionals on signatures, $\{\ell(S(\widehat{X})) : \ell \in T((\mathbb{R}^{d+1})^*)\}$, forms an algebra that separates points and contains constants (details can be found in [6, Corollary 2.16]). By the Stone-Weierstrass theorem, such an algebra is dense in the space of continuous functions on $S(K)$ with respect to the uniform norm. Hence, for any $\varepsilon > 0$, there exists a linear functional ℓ such that $\left|f(X) - \ell(S(\widehat{X}))\right| < \varepsilon$ for all $\widehat{X} \in K$. □

This results highlights that the signature unravels the non-linear structure of functions defined on the space of paths, revealing a latent linear structure within their geometry.

2.4. Expected signatures

In stochastic settings, the driving signal X is random, and so is its signature $S(X)$. In many applications, we are not interested in a single realization of the path but rather in its average effect. This motivates the notion of the *expected signature*, which encodes the distributional information of a stochastic process in a compact algebraic form. The formal definition follows.

Definition 12 (Expected Signature [32])**.** *Given a probability space $(\Omega, \mathcal{F}, \mathbb{P})$, X is a stochastic process which takes value in $\mathbb{R}^d$. Suppose that for almost every $\omega \in \Omega$, the signature of $X(\omega)$, denoted by $S(X(\omega))$, is well defined and, under the probability measure $\mathbb{P}$, its expectation, denoted by $\mathbb{E}[S(X)]$, is finite. For $0 \le s < t \le T$ we call $\mathbb{E}[S(X)]$ the expected signature of X over $[s,t]$ the formal series*

$$\mathbb{E}[S(X)]_{s,t} := \big(1,\, \mathbb{E}[S_1(X)],\, \mathbb{E}[S_2(X)],\, \dots \mathbb{E}[S_n(X)], \dots\big) \in T((\mathbb{R}^d)), \tag{2.9}$$

$$\mathbb{E}[S_n(X)] = \mathbb{E}\left[\int_{0<u_1<\cdots<u_n<T} dX_{u_1} \otimes \cdots \otimes dX_{u_n}\right], \tag{2.10}$$

with $n \ge 1$ and $\mathbb{E}[S_0(X)] := 1$.

Each component $\mathbb{E}[S_n(X)]$ belongs to $(\mathbb{R}^d)^{\otimes n}$ and can be interpreted as a generalised moment of order n.

2.4.1. Interpretation and connection with moment generating functions

The expected signature admits an interpretation that parallels the role of Moment Generating Functions in classical probability. Let $(\Omega, \mathcal{F}, \mathbb{P})$ be a probability space and let $X : [0,T] \to \mathbb{R}^d$ be a stochastic process. Chevyrev and Lyons [2] proved that, under suitable integrability assumptions, the expected signature $\mathbb{E}[S(X)]$ uniquely determines the law of X: if two processes X and Y satisfy $\mathbb{E}[S(X)] = \mathbb{E}[S(Y)]$, then they have the same distribution. This characterization result was later extended by Chevyrev and Oberhauser [33], who relaxed the assumptions by introducing an appropriate renormalization of the signature. This property can be viewed as a non-commutative analogue of a well-known fact for random variables: if Z is a real-valued random variable, its moment generating function $t \mapsto \mathbb{E}[e^{tZ}]$ characterizes its distribution. In a similar way, for a stochastic process X, the expected signature plays the role of a *moment generating function of paths*, encoding all joint moments of its increments in tensor form (see [34] for an intuitive discussion of this analogy). This is evident in the one-dimensional case for a path $X = (X_t)_{t\in[0,T]}$, where the expected signature over $[0,T]$ coincides with the sequence of (scaled) moments of the increment $X_T - X_0$:

$$\mathbb{E}[S(X)]_{0,T} = \Big(1,\, \mathbb{E}[X_T - X_0],\, \frac{1}{2!}\mathbb{E}[(X_T - X_0)^2],\, \dots,\, \frac{1}{k!}\mathbb{E}[(X_T - X_0)^k],\, \dots\Big).$$

Having established the probabilistic interpretation of the expected signature, we now investigate its analytical behaviour when the underlying process X evolves according to an Itô stochastic differential equation.

2.4.1.1. Expected signature of an Itô diffusion process

Let $(\Omega, \mathcal{F}, \mathbb{P})$ be a probability space endowed with a filtration satisfying the usual conditions. Let $W = \{W_t\}_{t\geq 0}$ be a k-dimensional Brownian motion in $\mathbb{R}^k$.

Consider the time-homogeneous Itô diffusion $X = \{X_t\}_{t\geq 0}$ taking values in $\mathbb{R}^d$, defined as the strong solution to the stochastic differential equation

$$dX_t = \mu(X_t)\, dt + V(X_t)\, dW_t, \tag{2.11}$$

where $\mu : \mathbb{R}^d \to \mathbb{R}^d$ is the *drift vector* and $V : \mathbb{R}^d \to \mathbb{R}^{d\times k}$ is the *volatility matrix*.

Explicitly, writing $\mu(x) = (\mu^{(1)}(x), \dots, \mu^{(d)}(x))^\top$ and $V(x) = \big(V_i^{(j)}(x)\big)_{1\leq j\leq d,\, 1\leq i\leq k}$, Equation (2.11) reads component-wise as

$$dX_t^{(j)} = \mu^{(j)}(X_t)\, dt + \sum_{i=1}^{k} V_i^{(j)}(X_t)\, dW_t^{(i)}, \qquad j = 1, \dots, d.$$

To ensure that Equation 2.11 admits a unique strong solution with continuous sample paths, we impose the following standard assumption.

Condition 2.7 (Global Lipschitz continuity and linear growth)**.** *The drift $\mu : \mathbb{R}^d \to \mathbb{R}^d$ and the volatility matrix $V : \mathbb{R}^d \to \mathbb{R}^{d\times k}$ are globally Lipschitz continuous; that is, there exists a constant $C > 0$ such that*

$$|\mu(x) - \mu(y)| + \|V(x) - V(y)\| \leq C\,|x - y|, \qquad \forall\, x, y \in \mathbb{R}^d.$$

Condition 2.7 implies a linear growth bound, $|\mu(x)| + \|V(x)\| \leq C(1 + |x|)$, which guarantees that the SDE (2.11) admits a unique strong solution with continuous sample paths, a result first due to Itô [35].

For such a process, each sample path $t \mapsto X_t(\omega)$ admits a well-defined (Stratonovich) signature $S(X)_{0,t}(\omega)$, as discussed in Section 2.3. The expected signature of X is therefore the function

$$\Phi(t, x) := \mathbb{E}^x[S(X)]_{0,t}, \qquad (t, x) \in \mathbb{R}_+ \times \mathbb{R}^d,$$

where $\mathbb{E}^x[\cdot]$ denotes expectation under the law of the diffusion starting at $X_0 = x$. We write $\Phi_n(t, x) \in (\mathbb{R}^d)^{\otimes n}$ for its tensor components, with $\Phi_0(t, x) = 1$ and $\Phi_n(0, x) = \mathbf{1}$ for $n \geq 1$.

In order to derive the partial differential equation satisfied by Φ, we require a mild smoothness assumption ensuring that Itô's formula can be applied componentwise.

Condition 2.8 (Regularity of the expected signature)**.** *Each coordinate projection $\pi^{(I)}(\Phi)$ belongs to $C^{1,2}([0, T] \times \mathbb{R}^d)$, i.e. it has continuous first-order derivatives in time and continuous second-order derivatives in space.*

The evolution of the expected signature is governed by the infinitesimal generator of the process X. This second-order differential operator, denoted by A, describes the expected rate of change of a function of the process. For a sufficiently smooth function, its action is

$$(Af)(x) = \sum_{i=1}^{d} \mu^{(i)}(x)\partial_{x_i} f(x) + \frac{1}{2}\sum_{i,j=1}^{d} b_{ij}(x)\partial^2_{x_i,x_j} f(x),$$

where $b(x) = V(x)V(x)^\top$ is a $d\times d$ diffusion covariance matrix, with components $b_{ij}(x) = \sum_{l=1}^{k} V_{il}(x)V_{jl}(x)$. The operator A acts component-wise on each component of the tensor-valued function Φ_n .

2.4.2. The PDE of the expected signature

A fundamental result, developed in [32], is that the expected signature $\Phi(t, x)$ of an Itô diffusion process satisfies a parabolic Partial Differential Equation (PDE). This concept is adapted to the notation of this thesis in the following theorem.

Theorem 2.9. *Let X be the d-dimensional Itô diffusion defined by*

$$dX_t = \mu(X_t)\,dt + V(X_t)\,dW_t,$$

with $\mu : \mathbb{R}^d \to \mathbb{R}^d$ and $V : \mathbb{R}^d \to \mathbb{R}^{d\times k}$ satisfying Condition 2.7. Let $\Phi(t,x) = \mathbb{E}^x[S(X)_{0,t}]$ be the expected signature, and assume each coordinate projection $\pi^{(I)}(\Phi) \in C^{1,2}([0,T]\times\mathbb{R}^d)$.

For each $n \geq 2$, define the source term

$$f_n(t,x) := \left(\sum_{j=1}^{d} \mu^{(j)}(x)\, e_j\right) \otimes \Phi_{n-1}(t,x) + \sum_{j=1}^{d}\left(\sum_{i=1}^{d} b_{i,j}(x)\, e_i\right) \otimes \partial_{x_j}\Phi_{n-1}(t,x) + \frac{1}{2}\left(\sum_{i,j=1}^{d} b_{i,j}(x)\, e_i \otimes e_j\right) \otimes \Phi_{n-2}(t,x), \tag{2.12}$$

and for $n = 1$, $f_1(t,x) := -\sum_{j=1}^{d} \mu^{(j)}(x)\, e_j$.

Then the components of the expected signature satisfy the parabolic system

$$(-\partial_t + A)\Phi_n(t,x) = -f_n(t,x), \qquad n \geq 1, \tag{2.13}$$

with initial conditions $\Phi_0(t,x) \equiv \mathbf{1} = (1,0,\ldots)$ and $\Phi_n(0,x) = 0$ for $n \geq 1$.

The core of the proof involves constructing a process $M_u = S(X)_{0,u} \otimes \Phi(t-u, X_u)$ and showing that it is a martingale. By applying Itô's product rule and setting the resulting drift term to zero, the parabolic system (2.13) for Φ follows. The complete proof can be found in [32, Chapter 4].

Remark 2.10. *The signature evolution is naturally written in Stratonovich form, $dS(X)_{0,u} = S(X)_{0,u} \otimes dX_u$, which preserves Chen's multiplicativity. In order to apply Itô's formula, it is necessary to move to Itô form, hence introducing the correction term defined in Proposition 2.1, $\frac{1}{2}\, S(X)_{0,u} \otimes \sum_{i,j} b_{i,j}(X_u)\, e_i \otimes e_j \circ du$. When applying Itô's product rule to $M_u = S(X)_{0,u} \otimes \Phi(t-u, X_u)$, this drift contributes the term related to Φ_{n-2} in Equation (2.12).*

Solving the PDE system (2.13) directly is often intractable. However, in certain canonical models, such as Geometric Brownian Motion or Ornstein-Uhlenbeck processes, the structure of the coefficients allows explicit recursive formulas for Φ. These explicit solutions will be the focus of Section 3.5 in Chapter 3 and subsequently used in the derivative pricing framework.

2.5. Financial models

Let $(\Omega, \mathcal{F}, \mathbb{Q})$ be a probability space endowed with a complete, right-continuous filtration $\{\mathcal{F}_t\}_{t\geq 0}$ satisfying the usual conditions. Let $W = \{W_t\}_{t\geq 0}$ be a k-dimensional Brownian motion in $\mathbb{R}^k$ under the risk-neutral measure $\mathbb{Q}$, with correlation matrix defined component-wise as $\mathrm{Corr}(W_t^{(i)}, W_t^{(j)}) = \rho_{i,j}$. In this work, we focus on a class of diffusion models commonly employed in finance to describe the dynamics of asset prices and interest rates. Each model is formulated as a time-homogeneous Itô diffusion process of the general form

$$dX_t = \mu(X_t)\,dt + V(X_t)\,dW_t, \tag{2.14}$$

where $X = \{X_t\}_{t\geq 0}$ taking values in $\mathbb{R}^d$ is the state vector, $\mu : \mathbb{R}^d \to \mathbb{R}^d$ is the drift function, and $V : \mathbb{R}^d \to \mathbb{R}^{d\times k}$ is the volatility matrix. The choice of μ and V determines the stochastic structure of X_t and governs the analytical tractability of its expected signature.

Throughout this work, we focus on two families of diffusions, the *Geometric Brownian Motion* (GBM) and the *Ornstein-Uhlenbeck* (OU) process, both of which admit explicit moment structures and compact representations of their expected signatures.

2.5.1. Geometric Brownian Motion (GBM)

The multidimensional GBM describes log-normal asset dynamics under constant volatility. For $X_t = (X_t^{(1)}, \dots, X_t^{(d)})^\top$, the dynamics are given by

$$dX_t^{(i)} = \left(r - \frac{1}{2}(VV^\top)_{ii}\right) X_t^{(i)}\, dt + X_t^{(i)} \sum_{j=1}^{d} \sigma_{ij}\, dW_t^{(j)}, \quad i = 1, \dots, d, \tag{2.15}$$

where r is the constant risk-free rate, $V = (\sigma_{ij})_{i,j=1}^d$ is the volatility matrix, and and $\langle dW_t^{(j)}, dW_t^{(k)} \rangle = \rho_{jk}\, dt$ denotes the instantaneous correlation between the Brownian components $W^{(j)}$ and $W^{(k)}$.

This process admits the explicit solution

$$X_t = X_0 \exp\left(\left(r\mathbf{1} - \tfrac{1}{2}\mathrm{diag}(VV^\top)\right) t + VW_t\right),$$

from which its joint moments can be computed as

$$\mathbb{E}\left[\prod_{i=1}^{d} (X_t^{(i)})^{n_i}\right] = \prod_{i=1}^{d} (X_0^{(i)})^{n_i} \exp\left(\left(\sum_{i=1}^{d} n_i r + \tfrac{1}{2} \sum_{i,k=1}^{d} n_i n_k (VV^\top)_{ik}\right) t\right).$$

As explained in Section 2.3, in order to ensure uniqueness of the signature, we consider the time-augmented process

$$\widehat{X} = (t, X_t) \in \mathbb{R}^{d+1},$$

whose drift and diffusion coefficients are given by

$$\widehat{\mu}(x_1, \dots, x_{d+1}) = (1,\, rx_2,\, \dots,\, rx_{d+1}), \qquad \widehat{V}(x_1, \dots, x_{d+1}) = \begin{bmatrix} 0_{1\times d} \\ \mathrm{diag}(x_2, \dots, x_{d+1})\, V \end{bmatrix}.$$

By construction, the augmented path $\widehat{X}$ belongs to the space $\mathcal{A}^p([0,T]; \mathbb{R}^d)$ of time-augmented paths of finite p-variation. This multidimensional time-augmented GBM serves as the foundation for the explicit computation of expected signatures presented in 3, Section 3.5.

2.5.2. Ornstein-Uhlenbeck (OU) Process

The Ornstein-Uhlenbeck process [36] provides mean-reverting dynamics, often used to model short rates or volatilities. In d dimensions, it satisfies the linear SDE

$$dX_t = \Theta(\mu - X_t)\, dt + \Sigma\, dW_t, \tag{2.16}$$

where $\Theta \in \mathbb{R}^{d\times d}$ is the mean-reversion matrix, $\mu \in \mathbb{R}^d$ the long-term mean, and $V \in \mathbb{R}^{d\times k}$ the volatility matrix. The explicit solution is

$$X_t = e^{-\Theta t} X_0 + \left(I - e^{-\Theta t}\right)\mu + \int_0^t e^{-\Theta(t-s)} V\, dW_s,$$

implying $\mathbb{E}[X_t] = e^{-\Theta t} X_0 + (I - e^{-\Theta t})\mu$.

As before, we work with its *time-augmented* version $\widehat{X} = (t, X_t) \in \mathbb{R}^{d+1}$, characterized by

$$\widehat{\mu}(x_1, \dots, x_{d+1}) = (1,\, [\Theta(\mu - (x_2, \dots, x_{d+1})^\top)]^\top), \qquad \widehat{V}(x_1, \dots, x_{d+1}) = \begin{bmatrix} 0_{1\times d} \\ V \end{bmatrix}.$$

Again, $\widehat{X}$ belongs to $\mathcal{A}^p([0,T]; \mathbb{R}^d)$. The affine structure of the OU process implies that each component of its expected signature is a polynomial function of the initial condition x, a property exploited in the computations developed later in Chapter 3, Section 3.5.

Both models, when expressed in the unified form (2.14), admit time-augmented representations that satisfy the hypotheses of Theorem 2.9. This allows the expected signature to be characterized by a parabolic PDE and computed recursively via the Feynman-Kac formulation, as detailed in Section 3.5.

2.5.3. FX Dynamics

Foreign exchange (FX) derivatives involve two currencies, typically a domestic and a foreign one, with corresponding short rates $\{r_t^d\}$ and $\{r_t^f\}$. Let S_t denote the spot FX rate, expressed as the price of one unit of the foreign currency in domestic currency units. Under the domestic risk-neutral measure $\mathbb{Q}^d$, its dynamics satisfy

$$dS_t = (r_t^d - r_t^f)\, S_t\, dt + \sigma_S\, S_t\, dW_t^{(S)}, \tag{2.17}$$

where $\sigma_S > 0$ is the instantaneous volatility and $W_t^{(S)}$ is a Brownian motion under $\mathbb{Q}^d$. The drift term reflects the interest rate differential between the two currencies. To account for the stochastic behavior of interest rates, we adopt the Ornstein-Uhlenbeck specification for both the domestic and foreign short rates:

$$\begin{aligned} dr_t^d &= (\theta_d \ - a_d\, r_t^d)\, dt + \sigma_d\, dW_t^{(d)}, \\ dr_t^f &= (\tilde{\theta}_f \ - a_f\, r_t^f)\, dt + \sigma_f\, dW_t^{(f)}, \end{aligned} \tag{2.18}$$

with $a_i > 0$, $\sigma_i > 0$, and where $\tilde{\theta}_f$ denotes the (measure-adjusted) mean level of the foreign short rate under $\mathbb{Q}^d$[5].

The Brownian motions $(W_t^{(S)}, W_t^{(d)}, W_t^{(f)})$ are jointly correlated with correlation matrix

$$\mathrm{Corr}\big(W_t^{(i)}, W_t^{(j)}\big) = \rho_{ij}, \qquad i, j \in \{S, d, f\}.$$

The resulting three-dimensional state process $X_t = (S_t, r_t^d, r_t^f)$ follows the diffusion system obtained combining the (2.17) and (2.18). This structure is useful in the context of this thesis, as it demonstrates that the signature-based regression can accommodate different types of underlying dynamics within a unified representation.

[5]If the foreign short rate is specified under its own risk-neutral measure with mean level $\theta_f^{(f)}$, then changing to the domestic measure $\mathbb{Q}^d$ introduces the standard quanto adjustment $\tilde{\theta}_f = \theta_f^{(f)} - \rho_{Sf}\, \sigma_S\, \sigma_f$, consistent with S_t being quoted as domestic price of one unit of foreign currency.

3

Signature-based Regression

The idea of representing functions as linear combinations of suitable basis is classical in analysis. For instance, Taylor's theorem shows how any smooth function can be expressed locally as a combination of monomials centered at a given point. Similarly, Fourier series provide another celebrated example, where functions are approximated by linear combinations of trigonometric basis elements. In an analogous way, functions on paths can be represented through linear functionals of their signatures.

This functional perspective provides the foundation for using path signatures as regression features in a pricing context. Building on the theoretical foundations presented in Chapter 2, this chapter develops a practical framework for derivative pricing based on path signatures: the signature-based regression. We begin by formally defining *signature payoffs* in Section 3.1 and demonstrating how their pricing reduces to evaluating a linear functional on the expected signature in Section 3.2. Section 3.3 then introduces the methodology used to estimate these functionals, while Section 3.4 focuses on the approximation of payoffs written on d underlyings and discusses different choices of regression bases: the *full signature* of the joint path, the *filtered signature* formed from individual asset signatures, and an intermediate *hybrid* representation that groups selected assets together. Finally, Section 3.5 derives analytical expressions for the expected signatures of the dynamics introduced in Section 2.5, thereby completing the formulation of the signature-based pricing framework.

3.1. Signature payoffs

A payoff function P assigns a monetary value to each price path of the underlying asset. Formally, it is any function of the path that is measurable with respect to the information available at maturity T. In financial terms, once the price path X is realized, the holder of the derivative receives the amount $P(X)$ at time T. Viewing payoffs as functionals on a path space suggests representing them through suitable features of the path itself. Among many available representations [37, 38], the signature provides a particularly powerful and structured representation of the path. Building on this idea, payoffs are represented as continuous functionals of the signature of the underlying price path [14, 39]. This motivates defining a specific class of payoffs: *signature payoffs*.

Definition 13 (Signature payoff [14])**.** *Let $\ell \in T((\mathbb{R}^d))^*$. For a given maturity $T > 0$, the ℓ-signature payoff $\mathcal{S}^\ell$ at maturity $T > 0$ is defined for any path $\widehat{X}$ belonging to the space of time-augmented paths of finite p-variation $\mathcal{A}^p([0,T];\mathbb{R}^d)$, as the mapping*

$$\begin{aligned} \mathcal{S}^\ell : \mathcal{A}^p([0,T];\mathbb{R}^d) &\to \mathbb{R}, \\ \widehat{X} &\mapsto \ell(S(\widehat{X})), \end{aligned}$$

where ℓ acts linearly on the signature $S(\widehat{X})$.

The theoretical justification for this definition follows from the universal approximation property of signatures (Theorem 2.6 in Chapter 2). Such theorem ensures that any continuous functional of a path can be uniformly approximated by a linear functional acting on its signature. Since payoffs are precisely (real-valued) functionals of the underlying price path, it follows that they can be approximated *arbitrarily well* by signature payoffs.

Corollary 3.1. *Let $P : \mathcal{A}^p([0,T];\mathbb{R}^d) \to \mathbb{R}$ be a continuous payoff function, and let $E \subset \mathcal{A}^p([0,T];\mathbb{R}^d)$ be compact. Then, for every $\varepsilon > 0$, there exists $\ell \in T((\mathbb{R}^d)^*)$ such that*

$$| P(X) - \mathcal{S}^\ell(\widehat{X}) | < \varepsilon \quad \forall\, \widehat{X} \in E,$$

where $\mathcal{S}^\ell(\widehat{X}) \coloneqq \ell(S(\widehat{X}))$ denotes the associated signature payoff.

Proof. This follows directly from Theorem 2.6, applied to $f = P$. Since signature payoffs are linear functionals of the signature, the existence of such an ℓ is guaranteed. □

Corollary 3.1 formally establishes that signature payoffs form a dense family in the space of continuous payoffs. Consequently, the fair value of any continuous payoff can be approximated arbitrarily well by pricing an appropriate signature payoff.

In practice, arbitrarily well means that the quality of the approximation depends on the truncation order k of the signature. Truncating the signature at level k corresponds to retaining iterated integrals up to order k, analogously to truncating a Taylor expansion after k terms. As established in the rough path framework of [7, Chapter 10], the error originating from replacing the full signature with its truncated version decreases as higher-order terms are included. In particular, for stochastic processes with sample paths of finite p-variation (such as Itô diffusion processes), the remainder term of the step-k expansion admits a higher-order bound in the path increments. This result follows from Davie's estimate [7, Theorem 10.29], which provides a quantitative stability bound for solutions of rough differential equations with respect to perturbations of their driving rough paths. Here, a p-geometric rough path refers to a path enhanced with its iterated integrals up to level p, obtained as the limit of the corresponding iterated integrals of smooth approximations and satisfying Chen's relation. Specifically, if X_t and $\tilde{X}_t$ are p-geometric rough paths driving the ODEs

$$dY_t = V(Y_t)\, dX_t, \qquad \text{and} \qquad d\tilde{Y}_t = V(\tilde{Y}_t)\, d\tilde{X}_t,$$

where the vector fields V are Lip^γ, $\gamma > p$, then the corresponding solutions Y_t and $\tilde{Y}_t$ satisfy

$$|Y_t - \tilde{Y}_t| \le C\, \|V\|_{\mathrm{Lip}^\gamma}\, \|X - \tilde{X}\|_{p\text{-var}}\, \exp\big(C\, \|V\|^p_{\mathrm{Lip}^\gamma}\big), \qquad \gamma > p,$$

for some constant C depending only on (p, γ). When $\tilde{X}$ represents the truncated signature $X^{(m)}$, the two drivers coincide up to order m, and their difference is of order $\mathcal{O}(\|X\|^{k+1}_{p\text{-var}})$. In the context of signature truncation and its approximation power, this means that neglecting higher-order iterated integrals introduces an error that decreases with the truncation order k.

In summary, the universal approximation property and the controlled truncation error of signatures establish the theoretical link between path functionals and their signature representation.

3.2. Pricing signature payoffs

As discussed in Section 3.1, signatures can approximate continuous payoffs with arbitrary accuracy. This suggests a natural approach to derivative pricing: given a payoff function P, one can construct a signature payoff that closely approximates it, and then use the fair value of the signature payoff as an approximation for the fair value of P.

Let $\mathcal{S}^\ell$ denote a signature payoff. The crucial property enabling efficient pricing is the *linearity* of both $\mathcal{S}^\ell$ in ℓ and of the expectation operator. In particular, according to Fundamental Theorem of Asset pricing [40], the arbitrage-free price of an option at maturity T is given by its discounted expected value

under the risk-neutral measure $\mathbb{Q}$. If Z_T denotes the (deterministic) discount factor from T to 0, then the fair value of a signature payoff can be written as

$$\text{Price}\big(\mathcal{S}^\ell\big) = Z_T\,\mathbb{E}^{\mathbb{Q}}\Big[\mathcal{S}^\ell(\widehat{X})\Big] = Z_T\,\mathbb{E}^{\mathbb{Q}}\Big[\ell(S(\widehat{X}))\Big] = Z_T\,\ell\Big(\mathbb{E}^{\mathbb{Q}}[S(\widehat{X})]\Big)\,, \tag{3.1}$$

where $\mathbb{E}^{\mathbb{Q}}[S(\widehat{X})]$ is the expected signature of the process X under $\mathbb{Q}$. When interest rates are stochastic, the discount factor must be taken inside the expectation. In that case, the price is $\mathbb{E}^{\mathbb{Q}}\left[Z_T\ell(S(\widehat{X}))\right]$, and the linearity argument remains unchanged; it simply requires augmenting the state with one additional dimension for the short-rate process. The linearity of ℓ allows the expectation to be interchanged with the linear functional, so that all model-specific information is contained in $\mathbb{E}^{\mathbb{Q}}[S(\widehat{X})]$, while ℓ depends only on the particular payoff being approximated. Once $\mathbb{E}^{\mathbb{Q}}[S(\widehat{X})]$ has been computed, pricing any signature payoff reduces to applying the corresponding linear functional to this expected value. Therefore, the same expected signature can be reused across many derivatives if we assume the same model, reducing the computational cost for large portfolios. To illustrate the pricing procedure, consider the payoffs of N derivatives $\{P^{\{1\}},\dots,P^{\{N\}}\}$. Using Corollary 3.1, each payoff $P^{\{i\}}$ can be approximated by a signature payoff $\mathcal{S}_i^\ell$, so that the corresponding fair values can be obtained through the following steps:

1. Replace each payoff $P^{\langle i\rangle}$ with its signature approximation $\mathcal{S}_i^\ell$.
2. Compute the expected signature $\mathbb{E}^{\mathbb{Q}}[S(\widehat{X})]$ under the chosen market model.
3. Apply each linear functional ℓ_i to the expected signature to obtain the approximate prices

$$\text{Price}(P^{\{i\}}) \approx Z_T\,\ell_i\big(\mathbb{E}^{\mathbb{Q}}[S(\widehat{X})]\big), \quad i=1,\dots,N.$$

This procedure decouples model-specific and payoff-specific components, which in theory allows for efficient pricing. In practice, however, its effectiveness depends critically on how accurately and stably the coefficients ℓ can be determined.

3.3. Estimation of the linear functional through signature-based regression

As discussed in the previous section, the fair value of a signature payoff depends partly on the linear functional ℓ acting on the expected signature of the underlying process, as guaranteed by the Universal Approximation Theorem (Theorem 2.6). In practice, ℓ is not known a priori and must be estimated from data. This motivates the introduction of the signature-based regression methodology, which formulates the estimation of ℓ as a regularized linear regression problem in the space of truncated signature features.

3.3.1. Formulation of the regression problem

Let $\{(X^{\{i\}},P^{\{i\}})\}_{i=1}^N$ denote a collection of N sample paths of the underlying process, together with their corresponding discounted payoffs $P^{\{i\}} = Z_T P(X^{\{i\}})$ computed under the risk-neutral measure. For each path $\widehat{X}^{(i)}$, let $S(\widehat{X}^{\{i\}}) \in \mathbb{R}^{N(d,k)}$ denote its truncated signature up to order k and $N(d,k)$ its number of elements. The goal is to determine the coefficients $\ell \in \mathbb{R}^{N(d,k)}$ such that

$$P^{\{i\}} \approx \ell^\top S(X^{\{i\}}), \qquad i=1,\dots,N. \tag{3.2}$$

Putting all observations into a matrix, let

$$\bar{S} = \begin{bmatrix} S(\widehat{X}^{\{1\}})^\top \\ \vdots \\ S(\widehat{X}^{\{N\}})^\top \end{bmatrix} \in \mathbb{R}^{N\times N(d,k)}, \qquad P = \begin{bmatrix} P^{\{1\}} \\ \vdots \\ P^{\{N\}} \end{bmatrix} \in \mathbb{R}^N.$$

Then, the estimation of ℓ reduces to solving the ordinary least squares (OLS) problem

$$\ell_{\text{OLS}} = \arg\min_{\ell\in\mathbb{R}^{N(d,k)}} \|P - \bar{S}\,\ell\|_2^2. \tag{3.3}$$

This regression formulation connects the theoretical pricing framework introduced in Section 3.2 to a practical learning procedure: ℓ represents the linear functional that best maps the truncated signature of the path to its observed payoff.

3.3.2. Dimensionality and truncation effect

The dimensionality $N(d,k)$ of the feature space, as well as the computational burden of estimating the linear functional ℓ, grow rapidly with both the truncation order of the signature and the dimensionality of the underlying process. Since all models considered in this work rely on the time-augmented path

$$\widehat{X} = (t, X_t^{(1)}, \dots, X_t^{(d)}) \in \mathbb{R}^{d+1},$$

the relevant alphabet has size $d+1$. Indeed, for the signature of a time-augmented d-dimensional path truncated at order k, the number of terms (excluding the first term corresponding to the empty word) is

$$N(d,k) = \sum_{m=1}^{k} (d+1)^m = \frac{(d+1)^{k+1} - (d+1)}{d}. \tag{3.4}$$

Table 3.1: Number of signature terms (excluding the constant term) as a function of the dimension $d+1$ of the time-augmented path and truncation order k. The total number of regression features is $N(d,k) = \frac{(d+1)^{k+1}-(d+1)}{d}$.

Dimension $d+1$	$k=2$	$k=3$	$k=4$	$k=5$
2	6	14	30	62
3	12	39	120	363
5	30	155	780	3,905
7	56	399	2,800	19,607
10	110	1,110	11,110	111,110

Table 3.1 highlights the rapid exponential growth of the signature's dimensionality with respect to both d and k. This exponential scaling not only increases memory and computational requirements, but also raises risk of numerical instability in the results and risk of overfitting when estimating ℓ. Consequently, while the signature-based regression framework is theoretically elegant, its practical implementation requires balancing expressive power and numerical stability. Higher truncation orders provide richer representations of path-dependent structures but dramatically increase the cost and instability of estimating ℓ.

3.3.3. Regularization and stability

The dimensionality of ℓ increases rapidly with both d and k (see Table 3.1), which can lead to ill-conditioned or over-fitted solutions, especially when the number of features is comparable to or larger than the number of paths N. To address this issue, regularization terms are introduced to stabilize the estimation, introducing different types of regression.

- Ridge regression [41]
$$\ell_{\text{ridge}} = \arg\min_{\ell} \|P - \bar{S}\,\ell\|_2^2 + \lambda\|\ell\|_2^2.$$
This penalizes large coefficients and improves numerical stability.
- Lasso regression [42]:
$$\ell_{\text{lasso}} = \arg\min_{\ell} \|P - \bar{S}\,\ell\|_2^2 + \lambda\|\ell\|_1,$$
which encourages sparsity in ℓ by setting many coefficients to zero, thus performing an implicit feature selection.
- Elastic Net [43]: a convex combination of the previous two penalties, balancing shrinkage and sparsity:
$$\ell_{\text{EN}} = \arg\min_{\ell} \|P - \bar{S}\,\ell\|_2^2 + \lambda_1\|\ell\|_1 + \lambda_2\|\ell\|_2^2.$$

Regularization mitigates instability caused by high-dimensional feature spaces, but it does not address the source of the dimensional explosion itself. This motivates the exploration of alternative representations of the path that reduce the number of signature features.

3.4. Variants of the signature representation

In a multi-dimensional setting, the exponential growth of the signature coefficients in both the number of dimensions d and order of truncation k motivates the exploration of techniques that reduce the number of terms retained in the signature representation of the path, thereby alleviating the curse of dimensionality. The goal is to achieve this reduction while still preserving as much relevant information as possible about the individual asset paths and the dependence structure encoded in the underlying multi-asset model. In this work, two main approaches are introduced and assessed:

1. the *full signature* approach: baseline, uses the full signature of the joint time-augmented process;
2. the *filtered signature* approach: discards cross terms involving different assets;

Before introducing the two approaches, it is useful to clarify the notion of *cross terms* in the signature. For a multi-asset time-augmented path

$$\widehat{X} = (t, X_t^{(1)}, \dots, X_t^{(d)}) \in \mathbb{R}^{d+1},$$

each coordinate corresponds to a letter in the alphabet $\Sigma = \{1, 2, \dots, d+1\}$, where 1 denotes the time component, and the remaining letters correspond to the assets. The signature $S(\widehat{X})$ consists of iterated integrals indexed by words $I \in \Sigma^+$, such as $(1,2)$ or $(3,2,1)$. A *cross term* is any component of the signature whose word involves more than one asset index, for example $(2,3)$ or $(1,2,3)$ when $d \geq 2$. Such terms encode interactions between different coordinates of the path. It is important to note that these cross-coordinate interactions are not the same as statistical dependence or correlation. Since signatures of time-augmented paths uniquely determine the path, any dependence between assets is automatically reflected in the sample paths themselvesas long as the driving Brownian motions are correlated, regardless of whether cross terms are explicitly included. Based on these observations, we present the two variants.

3.4.1. Full signature approach

Let $\widehat{X}$ be the time-augmented path representing the underlying multi-dimensional process. In this baseline case, the path is encoded through its full signature $S(\widehat{X})$, which includes all possible iterated integrals of the components of $\widehat{X}$, as seen in Section 2.3. Due to the properties of the signature, this representation characterizes completely the path and captures both the individual dynamics of each asset and their mutual interactions through the cross terms.

In this approach, the number of signature terms increases exponentially with the dimension d and the truncation order k (see Table 3.1). As a consequence, the full signature approach suffers from a pronounced curse of dimensionality.

3.4.2. Filtered signature approach

The key observation underlying the filtered-signature approach lies in the characterization capabilities of the signature. For a d-dimensional path $X = (X_t^{(1)}, \dots, X_t^{(d)})$, we may regard each component as a time-augmented path $\widehat{X}^{\langle i \rangle} = (t,\, X_t^{(i)})$ and recall that the signature $S(\widehat{X}^{\langle i \rangle})$ uniquely characterizes its evolution. This motivates the question of whether the collection of single-asset signatures $\{S(\widehat{X}^{\langle i \rangle})\}_{i=1}^{d}$ could retain relevant and sufficient information contained in the full signature $S(\widehat{X})$. Notice, however, that this would not be the case for the signature without time augmentation, as that object only encodes information at the final time t when $d = 1$.

If the d assets are simulated jointly under a model that already accounts for their correlation or dependence, this dependence is implicitly reflected in the sample paths themselves and therefore in the corresponding single-asset signatures. Although the cross terms of the full signature explicitly encode interactions between different components, the information regarding co-movements may, in principle,

be recovered indirectly from the individual time-augmented paths. The filtered signature approach is based on this hypothesis: since the information required to characterize each path is already contained in the single-asset time-augmented signatures, the resulting price estimates obtained from the filtered signature should remain close to those based on the full signature, while substantially reducing the dimensionality of the feature space and the associated computational cost.

Formally, this is achieved by restricting the alphabet used to construct the signature. Define the filtered alphabet $\Sigma_{\text{fil}} = \{w \in \Sigma^+ \text{ contains at most one of } \{2, \dots, d+1\}\}$, and let the corresponding filtered signature be defined as

$$S_{\text{fil}} = \left(\pi^{(I)}(S(X))\right)_{I \in \Sigma_{fil},\, |I| \leq k}$$

In this way, all signature components corresponding to mixed words such as $(1, i, j)$ or (j, i), where $i, j \in \{2, \dots, d+1\}$ are discarded.

The total number of terms in the truncated filtered signature can be obtained as follows. Each asset i contributes all words of length up to k formed from the two-letter alphabet $\{1, i\}$ for $i = 2, \dots, d+1$ (corresponding to time and asset i), yielding 2^m words at each level m. Summing over all assets gives $d \sum_{m=1}^{k} 2^m = d(2^{k+1} - 2)$ terms. However k pure-time words appear once per asset and must be counted only once overall, which requires subtracting $(d-1)k$. Hence the number of filtered-signature terms is

$$N_{\text{fil}}(d, k) = d\,(2^{k+1} - 2) - (d-1)k,$$

which is linear in the number of assets d. This linear complexity means the filtered signature does not suffer from the curse of dimensionality, making it computationally tractable for high-dimensional problems.

It is useful to distinguish between two quantities: the number of coefficients that must be computed, which corresponds to the signatures of the d time-augmented paths $\widehat{X}^{\langle 1 \rangle} = (t, X_t^{(i)})$, and the number of coefficients that are retained, given by N_{fil}. While the former determines the computational cost of generating the signatures, the latter represents the dimensionality of the feature matrix used in the regression and, therefore, the storage cost.

The comparison of the two approaches is summarized in Table 3.2 and the reduction in the number of terms can be visualized in Figure 3.1.

Table 3.2: Comparison of signature representations for multi-dimensional paths.

Approach	**Cross Terms**	**Dimensionality of the feature space**
Full Signature	Included	$\frac{(d+1)^{k+1}-(d+1)}{d}$
Filtered Signature	Excluded	$d\,(2^{k+1} - 2) - (d-1)k$

3.4.2.1. Hybrid Variant of the Filtered Signature

A natural extension of the filtered representation consists in computing the signature of groups of assets rather than of each asset individually. Let $G_1, ..., G_m$ be a partition of the asset indices. For each group G_j, we construct the grouped time-augmented path

$$\widehat{X}^{\#G_j} = (t, X_t^{(i)} : i \in G_j)$$

and compute its signature $S(\hat{X}^{\#G_j})$. The hybrid signature is then obtained by concatenating the signatures for every partition of the assets. This representation of the underlyings is useful when the multi-asset model is composed of lower-dimensional factors whose expected signatures are available analytically, while the expected signature of the full $(d+1)$-dimensional process is not.

The choice to represent a path through its full or filtered signature gives rise to a trade-off between dimensionality reduction and information retention. While the full signature captures the complete joint structure of the multi-asset path, its size rapidly becomes computationally prohibitive. The

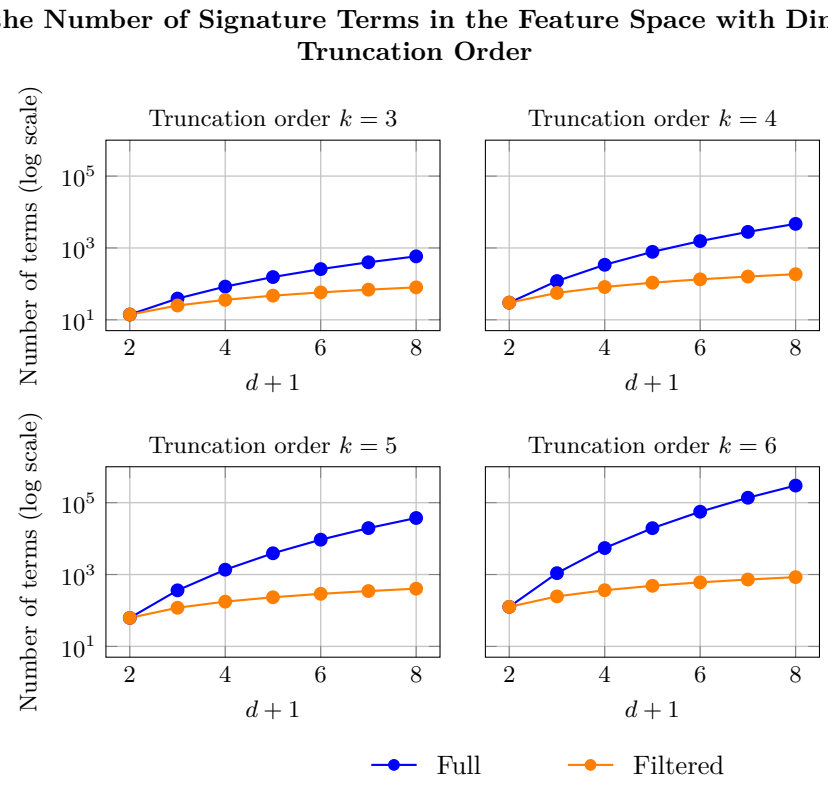


Figure 3.1: Comparison between the growth of the number of signature terms for the full and filtered representations across increasing input dimensions $d + 1$ and truncation orders $k = 3$ to 6. The y-axis is on a logarithmic scale.

filtered variant mitigates this issue at the cost of neglecting cross terms, under the assumption that sufficient information is contained in the individual paths $(t, X_t^{(i)})$, for $i = 2, \dots, d + 1$. This trade-off will be empirically examined in Section 4, where the performance of both approaches is compared.

3.4.3. Final estimation procedure

Having introduced the different signature representations, we now outline the practical steps involved in the signature-based pricing method. In practice, the estimation procedure consists of the following steps:

1. Simulate a set of sample paths $\{X^{\{i\}}\}_{i=1}^N$ under the risk-neutral measure $\mathbb{Q}$ and some chosen dynamics.
2. Compute their truncated signatures $S^k(X^{\{i\}})$ up to level k, choosing between the approaches introduced in Section 3.4.
3. Evaluate the discounted payoff $P^{\{i\}} = Z_T P(X^{\{i\}})$ for each path.
4. Fit ℓ by solving a regularized regression problem of the form (3.3).

Once the coefficients ℓ have been learned, they can be combined with the expected signature of the considered model to compute the approximations of the prices, as showed in (3.1). The following section addresses the computation of the expected signature $\mathbb{E}^{\mathbb{Q}}[S^k(X)]$, which represents the second core component of the signature-based pricing framework.

3.5. Expected Signature Computations

The price approximation introduced in Section 3.2 relies on two complementary ingredients. The first is the linear functional ℓ, estimated through the regression procedure of Section 3.3, which captures the dependence of the payoff on the path. The second is the expected signature $\mathbb{E}^{\mathbb{Q}}[S^k(X)]$, which captures

the behavior of the underlying process under the risk-neutral measure $\mathbb{Q}$. This section is dedicated to the second ingredient of the price, the expected signature of the underlying model, where we derive explicit expressions for the models introduced in 2.5.

3.5.1. Expected Signature: One Dimensional GBM

Under the Black-Scholes model, the $\mathbb{Q}$-dynamics of the price path are given by the Geometric Brownian Motion (GBM):

$$dX_t = rX_t\,dt + \sigma X_t\,dW_t,$$

with r and σ the constant risk-free interest rate and the constant volatility, respectively, and W a standard Brownian motion. The n-th moment of the process is given by

$$\mathbb{E}[X_t^n] = X_0^n E_n(t), \quad \text{where } E_n(t) := \exp\left(\left(nr + \frac{1}{2}n(n-1)\sigma^2\right)t\right), \quad \text{for } 0 \le t \le T. \tag{3.5}$$

We consider the time-augmented path

$$\widehat{X} = (t, X_t), \qquad \widehat{X}(0) = (0, X_0),$$

which fits the Itô diffusion setting of Theorem 2.9 with drift and diffusion vectors

$$\mu(x_1, x_2) = (1,\, rx_2), \qquad V(x_1, x_2) = (0,\, \sigma x_2).$$

The corresponding alphabet is $\Sigma = \{1, 2\}$, representing time and asset, respectively.

The expected signature $\Phi(t, x) = \mathbb{E}^x[S(\widehat{X})_{0,t}]$ satisfies the general parabolic system:

$$(-\partial_t + A)\Phi_n(t, x) = -f_n(t, x), \qquad n \ge 1, \tag{3.6}$$

with A the infinitesimal generator of $\widehat{X}$ and

$$f_n(t,x) = \begin{cases} (e_1 + rx\,e_2) \otimes \Phi_{n-1} + \sigma^2 x^2 e_2 \otimes \partial_x \Phi_{n-1} + \frac{1}{2}\sigma^2 x^2 e_2 \otimes e_2 \otimes \Phi_{n-2} & \text{if } n \ge 2, \\ (e_1 + rx\,e_2) & \text{if } n = 1. \end{cases} \tag{3.7}$$

The initial conditions are $\Phi_0(t, x) = \mathbf{1}$ and $\Phi_n(0, x) = 0$ for $n \ge 1$.

A key simplification in the computation of $\Phi(t, x)$ arises from the multiplicative structure of the Black-Scholes model. Since X_t scales linearly with its initial condition, the dependence of $\Phi(t, x)$ on x separates from its dependence on t. As established and proved[1] in [14], for the time-augmented process $\widehat{X} = (t, X_t)$ there exists a deterministic tensor-valued function $F : [0, T] \to T((\mathbb{R}^2))$ such that,

$$\pi^{(I)}\big(\Phi(t, x)\big) = x^{\alpha(I)}\,\pi^{(I)}\big(F(t)\big) \quad \forall\, I \in \{1, 2\}, \tag{3.8}$$

where $\alpha(I)$ denotes the number of occurrences of the letter 2 in I. This separation of variables reduces the PDE hierarchy for Φ to a deterministic recursion in t, allowing the components of $F(t)$ to be explicitly computed level by level, as shown in the following derivations.

Derivation. Substituting (3.8) into the PDE system reveals that each level component of F evolves independently of the initial condition x, and therefore can be computed recursively in time for each $n \ge 0$. Moreover, applying the Feynman-Kac theorem[2], Φ_n is given by

$$\Phi_n(t, x) = \int_0^t \mathbb{E}^x[f_n(t - s, X_s)]\,ds. \tag{3.9}$$

Before proceeding, we introduce a convenient notation.

Definition 14 (Truncation of a word)**.** *For a word $I = (i_1, \ldots, i_n) \in \Sigma^n$, its* first *and* second truncation *are*

$$I' = (i_2, \ldots, i_n), \qquad I'' = (i_3, \ldots, i_n)$$

whenever these are defined.

[1] In [14], a three-dimensional augmented process $(t, X_t, X_0 t/T)$ is considered. In practice, the third component $X_0 t/T$ plays no essential role for models with constant coefficients: it merely enforces linear dependence on the initial value. In the present setting, where X_t already scales multiplicatively with X_0, the two-dimensional augmentation (t, X_t) suffices, and the argument of the proof extends naturally.

[2] see, for example, Section 3.2 [44]

Base Cases ($n = 0, 1$) From the definition of the expected signature, the zeroth-level term is $\pi^{(\varnothing)}(\Phi(t,x)) = 1$. For the first level, from the Feymann-Kac representation (3.9),

$$\Phi_1(t, X_0) = \int_0^t \mathbb{E}[f_1(t-s),\, X_s]\, ds = \int_0^t \mathbb{E}\left[(e_1 + rX_s\, e_2)\right]\, ds = t\, e_1 + X_0\, (e^{rt} - 1)\, e_2.$$

Comparing with (3.8), the first-order term of F is

$$F_1(t) = t\, e_1 + \left(e^{rt} - 1\right) e_2.$$

Thus, the first-level components of the expected signature $\Phi(t, X_0)$ are:

$$\pi^{(I)}(\Phi(t, X_0)) = X_0^{\alpha(I)} \pi^{(I)}(F(t)) = \begin{cases} t & \text{if } I = (1) \\ (F(t)) = X_0(e^{rt} - 1) & \text{if } I = (2) \end{cases}.$$

Recursive Step ($n \geq 2$) For higher levels, each component Φ_n can be obtained by projecting the Feynman-Kac representation (3.9) onto a given word I of length n:

$$\pi^{(I)}(\Phi(t, X_0)) = \int_0^t \mathbb{E}^x \Big[\pi^{(I)}(f(t-s, X_s))\Big]\, ds.$$

The projection $\pi^{(I)}(f_n)$ can be computed directly from (3.7) as a function of the already known lower-level components Φ_{n-1}, Φ_{n-2}, and the moment $X_s^{\alpha(I)}$. Using the moment formula $\mathbb{E}[X_s^n] = X_0^n E_n(s)$ (Equation 3.5), we can compute the expectation and obtain a purely deterministic recursion in t.

As an illustration, consider a word starting with the time component, $I = (1, i_2, \ldots, i_n)$. Applying the projection $\pi^{(1,I')}$ to (3.9) gives

$$\pi^{(1,I')}\big(\Phi_n(t, X_0)\big) = \int_0^t \mathbb{E}\Big[\pi^{(1,I')}\big(f_n(t-s, X_s)\big)\Big]\, ds.$$

From the definition of f_n,

$$\begin{aligned} \pi^{(1,I')}\big(f_n(t,x)\big) &= \pi^{(1,I')}\big[(e_1 + rxe_2) \otimes \Phi_{n-1}(t,x) + \sigma^2 x^2 e_2 \otimes \partial_x \Phi_{n-1}(t,x) \\ &\qquad + \tfrac{1}{2}\sigma^2 x^2 (e_2 \otimes e_2) \otimes \Phi_{n-2}(t,x)\big] \\ &= \pi^{(I')}\big(\Phi_{n-1}(t,x)\big) \\ &\overset{(3.8)}{=} x^{\alpha(I')} \pi^{(I')}\big(F(t)\big). \end{aligned}$$

Thus, after integrating in time and taking expectations by exploiting the moment function $E_{\alpha(I')}(s)$, we obtain the recursive relation for $\pi^{(1,I')}(F(t))$.

The structure of the PDE implies that each term generates new words by prefixing one or two letters to those of lower order:

- the drift term $e_1 \otimes \Phi_{n-1}$ produces words beginning with 1;
- the diffusion term $e_2 \otimes \Phi_{n-1}$ produces words beginning with 2;
- the Itô correction $e_2 \otimes e_2 \otimes \Phi_{n-2}$ contributes words with a double prefix $(2, 2)$.

Accordingly, the recursion divides into three cases:

$$I = (1, I'), \qquad I = (2, 1, I''), \qquad \text{and} \quad I = (2, 2, I''),$$

where I' and I'' denote the truncation of I.

Carrying out the same reasoning as above for each case leads to the following recursive relations:

$$\pi^{(1,I')}(F(t)) = \int_0^t E_{\alpha(I')}\pi^{(I')}(F(t-s))\,ds,$$
$$\pi^{(2,1,I'')}(F(t)) = \big(r+\sigma^2\alpha((1,I''))\big)\int_0^t E_{\alpha((2,1,I''))}(s)\pi^{(1,I'')}(F(t-s))\,ds,$$
$$\pi^{(2,2,I'')}(F(t)) = \Big(\int_0^t \big(r+\sigma^2\alpha((2,I''))\big)E_{\alpha((2,2,I''))}(s)\pi^{(2,I'')}(F(t-s))\,ds, + \int_0^t \frac{1}{2}\sigma^2 E_{\alpha((2,2,I''))}(s)\pi^{(I'')}(F(t-s))\,ds\Big).$$

The detailed derivations for each case are reported in Appendix A.1.

3.5.2. Expected Signature: d-Dimensional GBM

The derivations carried out in Section 3.5.1 extend to the time-augmented d-dimensional GBM

$$\widehat{X}_t = (t, X_t^{(1)}, \dots, X_t^{(d)}) \in \mathbb{R}^{d+1},$$

where the d assets follow

$$dX_t^{(i)} = r\,X_t^{(i)}\,dt + \sigma_i X_t^{(i)}\,dW_t^{(i)}, \qquad i = 1,\dots,d, \tag{3.10}$$

and the Brownian motions are correlated according to

$$\mathbb{E}\big[dW_t^{(i)}dW_t^{(j)}\big] = \rho_{ij}\,dt, \qquad \rho_{ii} = 1.$$

Let $D = \mathrm{diag}(\sigma_1,\dots,\sigma_d)$ denote the diagonal volatility matrix and $R = (\rho_{ij})_{i,j=1}^d$ the correlation matrix. The covariance matrix of the Brownian component is then

$$C := DRD, \qquad C_{ij} = \sigma_i\sigma_j\rho_{ij}.$$

The alphabet is $\Sigma = \{1,\dots,d+1\}$, corresponding to time and the d stochastic components, respectively. For a word I and $w \in \{2,\dots,d+1\}$, let $\alpha_w(I)$ denote the number of occurrences of letter w in I, and set

$$n_i := \alpha_{i+1}(I), \qquad i = 1,\dots,d.$$

Then the mixed-moment factor associated with $(n_1,\dots,n_d) \in \mathbb{N}^d$ is given by

$$E_n(t) = \exp\left(\Big(\sum_{i=1}^d n_i r + \tfrac{1}{2}\sum_{i=1}^d n_i(n_i-1)C_{ii} + \sum_{1\le i<j\le d} n_i n_j\,C_{ij}\Big)t\right). \tag{3.11}$$

The computation of the expected signature in higher dimensions follows the same recursive structure as in the one-dimensional case. In particular, Φ satisfies the parabolic system defined in Theorem 2.9. Each component of $\Phi(t,x)$ therefore depends on lower-order ones through tensor products and spatial derivatives, leading to a hierarchy of PDEs that can be solved iteratively in the truncation level n.

The multiplicative structure of the geometric Brownian motion allows again the dependence on the initial values $x = (x_1,\dots,x_d)$ to separate from the time dependence. Specifically, the solution admits a factorized form

$$\pi^{(I)}\big(\Phi(t,x)\big) = \prod_{m=1}^d \big(x_{m+1}\big)^{\alpha_{m+1}(I)}\,\pi^{(I)}\big(F(t)\big), \tag{3.12}$$

so it suffices to compute the scalar functions $\pi^{(I)}(F(t))$.

For a word I and its decompositions

$$I = (1, I'), \qquad I = (i, 1, I''), \qquad I = (i, i, I''), \qquad I = (i, j, I''),\ i \neq j,$$

define the asset-count vector $n(I) := \big(\alpha_2(I), \dots, \alpha_{d+1}(I)\big) \in \mathbb{N}^d$.

The first-level components are

$$\pi^{(1)}(F(t)) = t, \qquad \pi^{(i)}\left(F(t)\right) = e^{rt} - 1, \quad i \in \{2, \dots, d+1\}.$$

Then, for $|I| \geq 2$,

$$\pi^{(1,I')}\big(F(t)\big) = \int_0^t E_{n(I')}(s)\ \pi^{I'}\big(F(t-s)\big)\, ds, \tag{3.13}$$

$$\pi^{(i,1,I'')}\big(F(t)\big) = \int_0^t E_{n(I)}(s)\ \pi^{(1,I'')}\big(F(t-s)\big)\, ds, \tag{3.14}$$

$$\begin{aligned}\pi^{(i,i,I'')}\big(F(t)\big) = A_{i,i}(I') \int_0^t E_{n(I)}(s)\ \pi^{(i,I'')}\big(F(t-s)\big)\, ds\ + \\ +\ B_{i,i} \int_0^t E_{n(I)}(s)\ \pi^{I''}\big(F(t-s)\big)\, ds,\end{aligned} \tag{3.15}$$

$$\begin{aligned}\pi^{(i,j,I'')}\big(F(t)\big) = A_{i,j}(I') \int_0^t E_{n(I)}(s)\ \pi^{(j,I'')}\big(F(t-s)\big)\, ds\ + \\ +\ B_{i,j} \int_0^t E_{n(I)}(s)\ \pi^{I''}\big(F(t-s)\big)\, ds,\ \ i \neq j.\end{aligned} \tag{3.16}$$

with coefficients $A_{\cdot,\cdot}$ and $B_{\cdot,\cdot}$ determined as follows. Let $C = VV^\top$; for any I' and letters $i, j \in \{2, \dots, d+1\}$,

$$A_{i,j}(I') := r + \sum_{m=2}^{d+1} C_{i-1,\,m-1}\, \alpha_m(I'), \tag{3.17}$$

$$B_{i,j} := \tfrac{1}{2}\, C_{i-1,\,j-1}, \tag{3.18}$$

and, by convention for the time prefix, $A_{1,\bullet} \equiv 1$. Equations (3.13) to (3.16) together with (3.17) and the first level initial conditions determine all components $\pi^I(F(t))$ recursively. The symmetry under exchanging the assets is immediate from (3.17)-(3.18).

3.5.3. Expected Signature: One-dimensional OU

We now turn to the Ornstein-Uhlenbeck process, whose dynamics differ qualitatively from the Geometric Brownian Motion case due to the lack of a multiplicative structure. The process is defined under the risk-neutral measure as

$$dX_t = \theta(\mu - X_t)\, dt + \sigma\, dW_t,$$

where $\theta > 0$ denotes the mean-reversion rate, μ the long-run mean level, and σ the volatility.

As in the GBM setting, we consider the time-augmented process $\widehat{X} = (t, X_t)$ with drift and diffusion functions

$$\mu(x_1, x_2) = (1,\, \theta(\mu - x_2)), \qquad V(x_1, x_2) = (0,\, \sigma).$$

The alphabet is $\Sigma = \{1, 2\}$, corresponding to the time and stochastic components, respectively.

Substituting these coefficients into the PDE of Theorem 2.9, we obtain for each level $n \geq 1$

$$(-\partial_t + A)\Phi_n(t, x) = -f_n(t, x), \tag{3.19}$$

where with A is the infinitesimal generator of $\widehat{X}$ and the source function is defined as:

$$f_n(t,x) = \begin{cases} (e_1 + \theta(\mu - x)e_2) \otimes \Phi_{n-1} + \sigma^2 e_2 \otimes \partial_x \Phi_{n-1} + \frac{1}{2}\sigma^2 e_2 \otimes e_2 \otimes \Phi_{n-2} & \text{if } n \geq 2, \\ e_1 + \theta(\mu - x)e_2 & \text{if } n = 1. \end{cases} \tag{3.20}$$

The initial conditions are $\Phi_0(t,x) = \mathbf{1}$ and $\Phi_n(0,x) = 0$ for $n \geq 1$.

Applying the Feynman-Kac theorem gives the integral representation

$$\Phi_n(t,x) = \int_0^t \mathbb{E}^x[f_n(t-s, X_s)]\, ds,$$

and, by projection on each basis element e_I, we obtain

$$\pi^{(I)}(\Phi_n(t,x)) = \int_0^t \mathbb{E}^x\left[\pi^{(I)}(f_n(t-s, X_s))\right] ds. \tag{3.21}$$

3.5.3.1. Affine dependence and polynomial structure

As in the GBM case, evaluating the expectations in the Feynman-Kac formula requires understanding how X_t depends on its initial value. While the GBM exhibits a multiplicative dependence on the initial condition, i.e. $X_t = X_0 \cdot Z_t$, the OU dynamics are affine in the initial condition. Specifically, denoting a process starting at X_0 as $X_t^{X_0}$, the explicit solution of the SDE is given by:

$$X_t^{X_0} = X_0 e^{-\theta t} + \mu(1 - e^{-\theta t}) + \int_0^t \sigma e^{\theta(s-t)} dW_s.$$

If the process starts at 0, we analogously have:

$$X_t^0 = \mu(1 - e^{-\theta t}) + \int_0^t \sigma e^{\theta(s-t)}\, dW_s.$$

Hence, the dependence on the initial condition is affine:

$$X_t^{X_0} = X_0 e^{-\theta t} + X_t^0. \tag{3.22}$$

According to [32, Theorem 4.12], for diffusion processes with affine dynamics (such as the OU model), each component of the expected signature can be written as a polynomial function of the initial value, and the coefficients are smooth functions of time. That is,

$$\pi^{(I)}(\Phi_n(t,x)) = \sum_{k=0}^{\alpha(I)} c_k^{(I)}(t)\, x^k.$$

The degree of the polynomial is at most $\alpha(I)$. This follows from the recursive structure of the source term f_n in (3.20), where multiplication by $(\mu - x)$ and differentiation with respect to x can increase or decrease the polynomial degree by at most one at each step.

Building on this, we introduce three polynomial functions that will facilitate the recursion. The first one, $R_I(x,\tau)$, is the *source polynomial*:

$$R_I(x,\tau) = \pi^{(I)}(f_n(\tau,x)) = \sum_{j=0}^{\alpha(I)} a_j^{(I)}(\tau) \cdot x^j \tag{3.23}$$

whose coefficients $a_j^{(I)}(\tau)$ evolve deterministically in time and are known from the recursive definition of f_n (3.20). The second is $Q_I(x,s)$, the *expected source polynomial*:

$$
\begin{aligned}
Q_I(x,s) &= \mathbb{E}[\pi^{(I)}(f_n(t-s, X_s))] \\
&= \mathbb{E}[\pi^{(I)}(f_n(t-s, X_s^0 + xe^{-\theta s}))] \\
&= \mathbb{E}\left[R_I(X_s^0 + xe^{-\theta s}, t-s)\right] \qquad (3.24)\\
&= \sum_{k=0}^{\alpha(I)} \beta_k^{(I)}(s) \cdot x^k. \qquad (3.25)
\end{aligned}
$$

where the coefficients $\beta_k^{(I)}(s)$ result from expanding the expectation in powers of x and depend on the coefficients $a_j^{(I)}$, the binomial coefficient, and the OU moments.

Finally, we present $P_I(x,t)$, the *expected signature polynomial*:

$$
P_I(x,t) := \pi^{(I)}(\Phi_n(t,x)) = \int_0^t Q_I(x,s)\, ds = \sum_{k=0}^{\alpha(I)} c_k^{(I)}(t)\, x^k, \qquad (3.26)
$$

where $c_k^{(I)}(t)$ are the final time-dependent coefficients obtained after integrating $\beta_k^{(I)}(s)$ with respect to time.

Our goal is to identify the coefficients $c_k^{(I)}(t)$, since they characterize the components of the expected signature Φ. In order to do so, we exploit the relationships between R_I, Q_I and P_I. The explicit derivation of these relationships is presented next.

From the relation between R_I and Q_I (3.24), we can derive an explicit expression for the coefficients $\beta_k^{(I)}(s)$ of the expected source polynomial $Q_I(X_0, s)$. Substituting the affine decomposition (3.22) into the expectation and expanding via the binomial theorem yields

$$
\begin{aligned}
Q_I(X_0,s) &= \mathbb{E}^{X_0}\left[\pi^{(I)}(f_n(t-s,\, X_s))\right] \\
&= \mathbb{E}\left[\pi^{(I)}\left(f_n\left(t-s,\, X_s^0 + X_0 e^{-\theta s}\right)\right)\right] \\
&= \mathbb{E}\left[R_I\left(X_s^0 + X_0 e^{-\theta s},\, t-s\right)\right] \\
&= \mathbb{E}\left[\sum_{j=0}^{\alpha(I)} a_j^{(I)}(t-s)\left(X_s^0 + X_0 e^{-\theta s}\right)^j\right] \\
&= \mathbb{E}\left[\sum_{j=0}^{\alpha(I)} a_j^{(I)}(t-s) \sum_{k=0}^{j} \binom{j}{k} (X_s^0)^{j-k} (X_0 e^{-\theta s})^k\right]. \qquad (3.27)
\end{aligned}
$$

The expression in (3.27) is a polynomial in X_0. To identify its coefficients, we reorganize the double summation by collecting all terms associated with the same power k of X_0:

$$
Q_I(X_0,s) = \sum_{k=0}^{\alpha(I)} \underbrace{\left(\sum_{j=k}^{\alpha(I)} a_j^{(I)}(t-s) \binom{j}{k} \mathbb{E}\left[(X_s^0)^{j-k}\right] e^{-\theta s k}\right)}_{\beta_k^{(I)}(s)} X_0^k. \qquad (3.28)
$$

Hence, the coefficients of the polynomial Q_I are given explicitly by

$$
\beta_k^{(I)}(s) = e^{-\theta s k} \sum_{j=k}^{\alpha(I)} a_j^{(I)}(t-s) \binom{j}{k} \mathbb{E}\left[(X_s^0)^{j-k}\right]. \qquad (3.29)
$$

We can now express the expected signature polynomial $P_I(X_0,t) = \pi^{(I)}(\Phi_n(t,X_0))$ in terms of the coefficients $\beta_k^{(I)}(s)$. Substituting the definition of Q_I and integrating over s yields

$$\begin{aligned} P_I(X_0,t) &= \int_0^t Q_I(X_0,s)\,ds \\ &= \int_0^t \sum_{k=0}^{\alpha(I)} \beta_k^{(I)}(s)\, X_0^k\, ds \\ &= \sum_{k=0}^{\alpha(I)} \left(\int_0^t \beta_k^{(I)}(s)\, ds \right) X_0^k \\ &= \sum_{k=0}^{\alpha(I)} \underbrace{\left(\int_0^t (e^{-\theta s})^k \sum_{j=k}^{\alpha(I)} a_j^{(I)}(t-s) \binom{j}{k} \mathbb{E}\big[(X_s^0)^{j-k}\big]\, ds \right)}_{c_k^{(I)}(t)} X_0^k. \end{aligned} \tag{3.30}$$

Hence, the coefficients of the polynomial P_I that characterize the levels of Φ are defined as

$$c_k^{(I)}(t) := \int_0^t (e^{-\theta s})^k \sum_{j=k}^{\alpha(I)} a_j^{(I)}(t-s) \binom{j}{k} \mathbb{E}\big[(X_s^0)^{j-k}\big]\, ds.$$

3.5.3.2. Recursive structure

The computation of each component of the expected signature proceeds recursively in the word length $|I| = n$. Starting from the base levels

$$\Phi_0 = \mathbf{1}, \qquad \Phi_1(t,x) = t\, e_1 + (\mu - x)(1 - e^{-\theta t})\, e_2,$$

all higher-order terms Φ_n are obtained by evaluating the source function f_n in (3.20), which depends only on Φ_{n-1} and Φ_{n-2} and their spatial derivatives. For each word I of length n, projecting f_n with $\pi^{(I)}$ yields

$$\pi^{(I)}(f_n) = \text{combination of } \pi^{(I')}(\Phi_{n-1})\,,\ \partial_x \pi^{(I')}(\Phi_{n-1})\,,\ \text{and } \pi^{(I'')}(\Phi_{n-2})\,,$$

where I' and I'' denote shorter subwords of I. This recursive dependence determines the coefficients $a_j^{(I)}$ of the source polynomial R_I. From these, the coefficients $\beta_k^{(I)}$ of the expected source polynomial Q_I are obtained using the analytical expression in (3.29), which involves the OU moments $\mathbb{E}[(X_s^0)^{j-k}]$. Finally the coefficients $c_k^{(I)}$ of the expected signature polynomial P_I follow from time integration as in (3.30). This structure allows all components of the expected signature to be computed recursively using only the coefficients of the previous levels.

Example 3.2. *To illustrate the recursive structure of the computation, let us work out explicitly the case of the word $I = (2,1)$.*

Recall that the recursion starts from $n = 0$ and $n = 1$, where the components of the expected signature are explicitly known:

$$\pi^{\varnothing}(\Phi_0(t,x)) = 1, \qquad \pi^{(1)}(\Phi_1(t,x)) = t, \qquad \pi^{(2)}(\Phi_1(t,x)) = (\mu - x)\,(1 - e^{-\theta t}).$$

These serve as the base cases for the recursion.

For $I = (2,1)$ we have $n = 2$ and $\alpha(I) = 1$. The recursive definition of f_n gives

$$\begin{aligned}\pi^{(2,1)}(f_2(\tau,x)) &= \theta(\mu-x)\,\pi^{(1)}(\Phi_1(\tau,x)) + \sigma^2\,\partial_x\pi^{(1)}(\Phi_1(t,x)) \\ &= \theta(\mu-x)\,\tau + \sigma^2\partial_x(\tau) \\ &= -\theta\tau\,x + \theta\mu\tau. \end{aligned} \tag{3.31}$$

We thus identify the polynomial dependence on x as

$$\pi^{(2,1)}(f_2(\tau,x)) = \underbrace{-\theta\tau}_{a_1^{(2,1)}(\tau)}\; x + \underbrace{\theta\mu\tau}_{a_0^{(2,1)}(\tau)} = \sum_{j=0}^{1} a_j^{(2,1)}(\tau)\,x^j.$$

Clearly the coefficients $a_j^{(2,1)}(\tau)$ are known functions of time.

Substituting this into the representation of Φ_n, we obtain

$$\pi^{(2,1)}(\Phi_2(t,X_0)) = \int_0^t \mathbb{E}^{X_0}\left[\pi^{(2,1)}(f_2(t-s,\,X_s))\right] ds \tag{3.32}$$

$$= \int_0^t \mathbb{E}^{X_0}\left[\sum_{j=0}^{1} a_j^{(2,1)}(t-s)\cdot X_s^j\right] ds. \tag{3.33}$$

Using the relation between Q_I and R_I in (3.26), the expectation can be rewritten as a polynomial in X_0:

$$\int_0^t \sum_{k=0}^{1} \underbrace{\left(\sum_{j=k}^{1} a_j^{(2,1)}(t-s)\binom{j}{k}\,\mathbb{E}\big[(X_s^0)^{\,j-k}\big]\,(e^{-\theta s})^k\right)}_{\beta_k^{(2,1)}(s)} X_0^k\,ds. \tag{3.34}$$

Finally, exchanging the sum and the integral and integrating in s gives

$$\pi^{(2,1)}(\Phi_2(t,X_0)) = \sum_{k=0}^{1} \underbrace{\left(\int_0^t \sum_{j=k}^{1} a_j^{(2,1)}(t-s)\binom{j}{k}\,\mathbb{E}\big[(X_s^0)^{\,j-k}\big]\,(e^{-\theta s})^k\,ds\right)}_{c_k^{(2,1)}(t)} X_0^k. \tag{3.35}$$

We notice that $\mathbb{E}\big[(X_s^0)^{\,j-k}\big]$ is also explicitly known as a function of s, hence the coefficients $c_k^{(2,1)}(t)$ can be computed. In this way, we have determined the expected signature component corresponding to the word $(2,1)$. The same reasoning applies to any other word of length $n \geq 2$.

3.5.4. Expected Signature: d-dimensional OU

We now extend the previous computations to the time-augmented d-dimensional Ornstein-Uhlenbeck process

$$\widehat{X} = (t, X_t^{(1)}, \dots X_t^{(d)}),$$

where the vector process $X_t = (X_t^{(1)}, \dots, X_t^{(d)})^\top$ satisfies the dynamics

$$dX_t = \Theta(\mu - X_t)\,dt + \Sigma\,dW_t, \tag{3.36}$$

where the parameters are defined as in Section 2.5

The alphabet is defined by $\Sigma = \{1, 2, \dots, d+1\}$ corresponding to time and $X^{(1)}, \dots, X^{(d)}$ respectively. The infinitesimal generator of X_t is

$$A = \sum_{i=1}^{d} \big(\Theta(\mu - x)\big)_i \frac{\partial}{\partial x_i} + \frac{1}{2}\sum_{i,j=1}^{d} (\Sigma\Sigma^\top)_{ij}\,\frac{\partial^2}{\partial x_i \partial x_j}.$$

Substituting the coefficients into the general PDE of Theorem 2.9 yields, for each level $n \geq 1$,

$$(-\partial_t + A)\Phi_n(t,x) = -f_n(t,x), \tag{3.37}$$

where the source function is

$$f_n(t,x) = \begin{cases} \Big(e_1 + \sum_{i=1}^{d} (\Theta(\mu - x))_i e_{i+1}\Big) \otimes \Phi_{n-1}(t,x) & \text{if } n \geq 2, \\ \quad + \sum_{i,j=1}^{d} (\Sigma\Sigma^\top)_{ij}\, e_{i+1} \otimes \partial_{x_j} \Phi_{n-1}(t,x) & \\ \quad + \frac{1}{2} \sum_{i,j=1}^{2} (\Sigma\Sigma^\top)_{ij}\, e_{i+1} \otimes e_{j+1} \otimes \Phi_{n-2}(t,x), & \\ \Big(e_1 + \sum_{i=1}^{d} (\Theta(\mu - x))_i e_{i+1}\Big), & \text{if } n = 1. \end{cases} \tag{3.38}$$

with the convention that $\Phi_0 = \mathbf{1}$ and $\Phi_n(0,x) = 0$ for $n \geq 1$. The integral representation follows from the Feynman-Kac theorem:

$$\Phi_n(t,x) = \int_0^t \mathbb{E}^x[f_n(t-s, X_s)]\, ds,$$

and its projection onto the word I is

$$\pi^{(I)}(\Phi_n(t,x)) = \int_0^t \mathbb{E}^x\Big[\pi^{(I)}(f_n(t-s, X_s))\Big]\, ds.$$

3.5.4.1. Affine dependence and polynomial structure

As in the d-dimensional case, the two dimensional OU process admits an explicit affine solution:

$$X_t^x = e^{-\Theta t}x + X_t^0, \qquad X_t^0 = (\mathrm{I} - e^{-\Theta t})\mu + \int_0^t e^{-\Theta(t-s)}\Sigma\, dW_s.$$

The random part X_t^0 is Gaussian and independent of x, so that each component $\pi^{(I)}(\Phi_n(t,x))$ can be expressed as a polynomial in the initial state $x = (x^{(1)}, \dots, x^{(d)})$ of total degree at most $\alpha(I)$, where $\alpha(I)$ counts the number of indices that refer to stochastic components $\{2, \dots, d\}$ in the word I. That is,

$$\pi^{(I)}(\Phi_n(t,x)) = \sum_{|k| \leq \alpha(I)} c_k^{(I)}(t)\, x^k, \qquad x^k = (x^{(1)})^{k_1}(x^{(2)})^{k_2}.$$

The coefficients $c_k^{(I)}(t)$ satisfy the same recursive construction as in the one-dimensional setting: they are obtained by expanding the source polynomial $\pi^{(I)}(f_n)$, taking expectations exploiting the the affine relation above, and integrating over s.

3.5.4.2. Recursive structure

The recursive computation proceeds as in the one-dimensional case, replacing scalar parameters θ and σ with matrices Θ and V. All terms of Φ_n remain polynomials in x, and the affine dependence $X_t^x = e^{-\Theta t}x + X_t^0$ ensures that expectations can be evaluated analytically using the Gaussian moment formula of X_t^0. The base cases of the recursion are $\Phi_0 = 1$ and, using $\mathbb{E}^x[X_s] = e^{-\Theta s}x + (\mathrm{I} - e^{-\Theta s})\mu$,

$$\Phi_1(t,x) = t\, e_1 + \big((\mathrm{I} - e^{-\Theta t})(\mu - x)\big)_1 e_2 + \big((\mathrm{I} - e^{-\Theta t})(\mu - x)\big)_2 e_3.$$

3.5.5. Hybrid Expected Signature for FX

The pricing of FX derivatives in this work is based on a hybrid model that combines the analytical signatures of the Black-Scholes (GBM) and Ornstein-Uhlenbeck processes introduced in the previous sections. Rather than deriving the expected signature of the full Black-Scholes-Ornstein-Uhlenbeck

system, corresponding to the time-augmented path $(t, X_t^{(1)}, X_t^{(2)}, X_t^{(3)})$, we approximate it by composing the individual expected signatures of its two components: a one-dimensional Black-Scholes preocess $(t, X^{(1)})$ for the exchange rate, and a two-dimensional Ornstein-Uhlenbeck process $(t, X^{(2)}, X^{(3)})$ for the domestic and foreign short rates. This aligns with the hybrid representation of the regression base introduced in 3.4.2.1.

Each expected signature coefficient is determined by the letters appearing in the corresponding word:

- Words involving only $\{1, 2\}$ (time and FX rate) are evaluated using the analytical expected signature of the BS model.
- Words involving only $\{1, 3, 4\}$ (time and the two short rates) are evaluated using the OU expected signature.
- Words containing terms referred to cross terms between the FX and rates components, such as $(2, 3)$ or $(2, 4)$, are excluded.

The resulting expected signature can therefore be interpreted as a composition of two independent signatures, each capturing the temporal evolution of its own subdynamics.

4

Numerical Results

This chapter presents the empirical evaluation of the signature-based regression framework introduced in Chapter 3, with the aim of answering the two research questions stated in Chapter 1:

- how signature-based regression compares with standard Monte Carlo simulation for derivative pricing, and
- whether more compact signature representations can retain comparable pricing accuracy while mitigating the curse of dimensionality.

To address these questions, the chapter is organized as follows. In Section 4.1, we validate the analytical formulas of the expected signature derived in Section 3.5, and compare them with their Monte Carlo estimators. This analysis also serves to identify practical choices for key parameters, such as the number of training paths, that will be used in the subsequent experiments. Section 4.2 examines the regression setup in more detail, with particular attention to the choice of truncation order. We analyze the trade-off between representational richness and pricing accuracy. The main pricing results are presented in Section 4.3. Here, we evaluate the performance of the signature-based regression framework on a range of benchmark derivatives: one-dimensional European and Asian calls, a two-dimensional arithmetic basket option, and a three-dimensional FX-type derivative, under the models introduced in Section 2.5. For each case, we study the behavior of the price estimates as the number of training paths increases, assess the stability across repeated runs, and investigate the effect of moneyness. This section provides the empirical evidence needed to address both research questions, showing how the method scales with dimension and how full versus filtered signature representations impact pricing performance. Finally, Section 4.4 complements the accuracy results with a timing analysis. By breaking down the computational cost of each stage of the pricing workflow, we provide a direct comparison with classical Monte Carlo simulation as well as between the different signature representations.

Parameters Before presenting the numerical results, we summarize the common parameter choices used across the different pricing experiments. Unless otherwise specified, each pricing experiment is repeated over $n_{\text{runs}} = 20$ independent runs to assess the variability due to random sampling. In the experiments, Monte Carlo plays three distinct roles: (i) it provides the simulated training paths of size N_{train} used to train the regression; (ii) the same paths are used when $\mathbb{E}[\text{Sig}]$ is estimated empirically; and (iii) it serves as a pricing method for comparison under the same computational budget. In addition, a separate Monte Carlo simulation with $N_{\text{bench}} = 100{,}000$ paths is used to obtain a high-accuracy benchmark price. These distinctions are summarized in Table 4.1. All simulations employ antithetic variates to reduce sampling variance, and the Brownian increments are standardized at each time step to have mean zero and unit variance. Unless otherwise specified, simulations are performed over a time horizon of $T = 1$ under a risk-free rate $r = 5\%$ and volatility $\sigma = 20\%$, with the underlying spot price initialized at $S_0 = 100$. The reference strike is set to $K = 100$ (at-the-money).

For the one-dimensional experiments, both the European and Asian call options are priced under the Black-Scholes model. The two-dimensional case considers an arithmetic basket call option driven by

Table 4.1: Summary of Monte Carlo usages in the pricing experiments and corresponding number of paths employed.

Purpose	Number of paths	Description
Training regression	N_{train}	Paths used to fit the linear model
MC estimate of $\mathbb{E}$[Sig]	N_{train}	Same paths used to estimate expected signature
Direct MC comparison	N_{train}	Pricing with same computational budget as regression
Benchmark MC price	$N_{\text{bench}} = 100{,}000$	High-accuracy reference price

two correlated GBM, with correlation coefficient $\rho = 0.4$, and volatilities $\sigma_1 = 20\%$ and $\sigma_2 = 40\%$. The three-dimensional case corresponds to a FX derivative priced under the Black-Scholes-Ornstein-Uhlenbeck (BSOU) model. The specific parameters used in this last setting are reported in Table 4.2. Across all experiments, prices are reported for both regression variants: the one based on the theoretical

Table 4.2: Model parameters for the FX derivative under the Black-Scholes-Ornstein-Uhlenbeck (BSOU) model.

Parameter	Description	Value
S_0	Spot FX rate	1.10
K	Strike	1.05
σ_X	FX volatility	0.10
r_0^d	Initial domestic short rate	0.03
r_0^f	Initial foreign short rate	0.015
a_d, θ_d, σ_d	Domestic OU parameters	0.5, 0.02, 0.01
a_f, θ_f, σ_f	Foreign OU parameters	0.7, 0.01, 0.016
ρ_{Xd}	Corr. (FX, domestic rate)	0.2
ρ_{Xf}	Corr. (FX, foreign rate)	0.1
ρ_{df}	Corr. (domestic, foreign rates)	0.3

expected signature and the one based on its Monte Carlo estimate. In the two-dimensional case, we compare the performance of the full signature with the filtered representation. In the three-dimensional FX setting, we consider the hybrid representation constructed as the combination of the signatures of the individual components. This allows us to directly assess how different signature representations behave across increasing model dimensions.

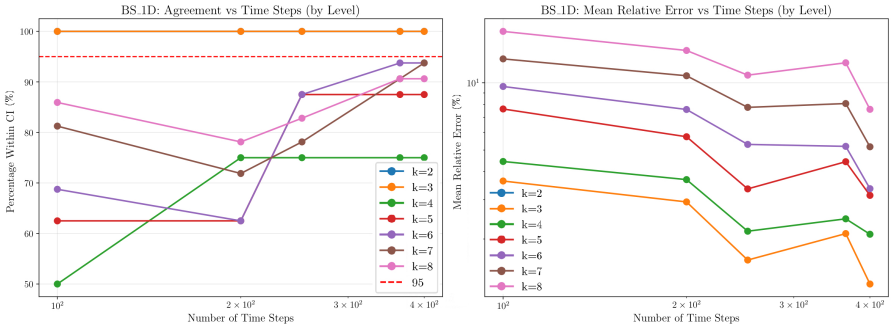
In what follows, Theoretical $\mathbb{E}$[Sig] and MC $\mathbb{E}$[Sig] denote, respectively, the use of the analytically derived expected signature and the Monte Carlo-estimated expected signature in the regression-based pricing procedure.

4.1. Comparison between Analytical and Monte Carlo Expected Signature

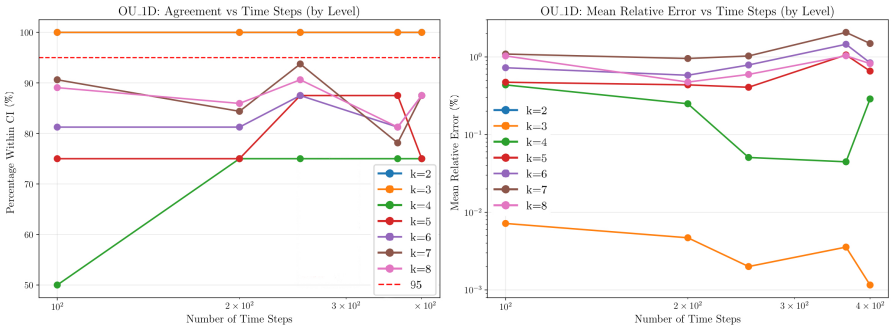
This section serves two purposes. First, it validates the analytical expressions derived in Section 3.5 for the expected signature under the Black-Scholes and Ornstein-Uhlenbeck dynamics by comparing them with Monte Carlo estimates obtained from simulated sample paths. Second, the analysis provides a practical guideline for selecting an adequate time discretization over a fixed time horizon $T = 1$ to be used in subsequent experiments.

To validate the analytical expressions derived in Section 3.5 for the expected signature under the GBM and OU dynamics, we compare them against Monte Carlo estimates obtained from simulated sample paths. Once validated, the analytical coefficients are treated as the ground truth in subsequent experiments throughout this chapter. As a measure of accuracy, we report the mean relative errors (in %) for each level of truncation k as a function of the number of simulated paths, N_{train}, and the size of the time discretization, n_{steps}.

Figures 4.1a and 4.1b compare the theoretical and Monte Carlo estimates of the expected signature for $n_{steps} \in \{100, 200, 252, 365, 400\}$. Each figure shows, on the left panel, the percentage of terms of the theoretical expected signature that lies within the 95% confidence interval of the corresponding Monte Carlo estimate, for each truncation order $k \in \{2, \ldots, 8\}$. The comparison is based on $5,000$ paths, a deliberately moderate sample size, ensuring that any subsequent experiments using equal or more paths will produce estimates of equal or higher accuracy. The right panel reports the corresponding mean relative error for each order of truncation k as a function of the number of time steps.



(a) Black-Scholes model: proportion (%) of coefficients within the 95% confidence interval (left) and corresponding mean relative error (%) (right) per level of truncation k vs time steps. The red dashed line corresponds to the 95% threshold.



(b) Ornstein–Uhlenbeck model: proportion (%) of coefficients the 95% confidence interval (left) and corresponding mean relative error (%) (right) per level of truncation k vs time steps. The red dashed line corresponds to the 95% threshold.

Figure 4.1: Comparison between the Monte Carlo and theoretical expected signatures across increasing number of time-steps ($n_{steps} \in \{100, 200, 252, 365, 400\}$) for the one-dimensional Black-Scholes (top) Ornstein-Uhlenbeck (bottom) models. Each left panel reports the percentage of coefficients whose theoretical values lie within the 95% confidence interval of the Monte Carlo estimate for different truncation orders k; each right panel shows the corresponding mean relative error as a function of the number of time steps for different truncation orders k.

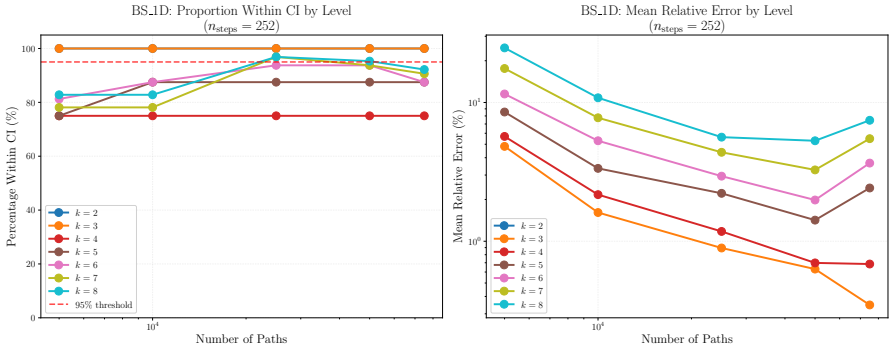
For the Black-Scholes model, the agreement improves steadily with finer discretizations, with over 75% of the coefficients falling within the confidence level for $n_{steps} > 200$. Moreover, the right panel shows that the mean relative error decreases systematically across truncation levels as the discretization becomes finer, with a value of $n_{steps} > 200$ producing reasonable results.

For the OU model, the overall agreement appears higher and more stable across discretization levels compared to the Black-Scholes case. As shown in the left panel of Figure 4.1b, most truncation levels achieve over 80% of coefficients within the confidence interval for 100 times steps, with some fluctuations at the finer resolutions. The right panel shows that the mean relative error remains small throughout, especially for lower truncation orders. This behavior reflects the smoother dynamics of the mean-reverting process.

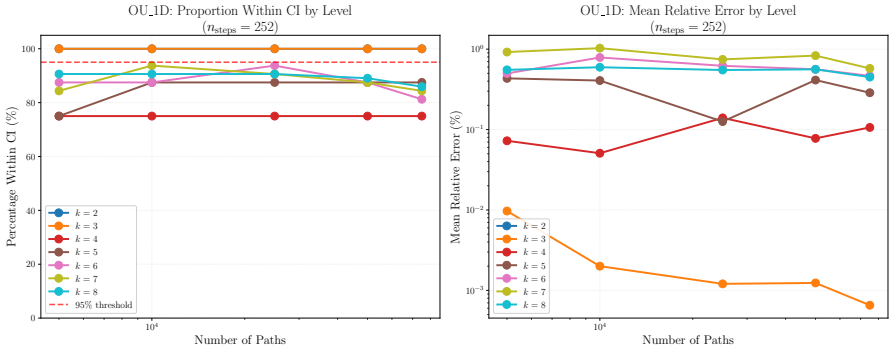
Based on these observations, and to ensure consistency across models, we fix a unified discretization of $n_{\text{steps}} = 252$, corresponding to a typical number of trading days in a year, for all the subsequent experiments.

Next, we study the effect of the number of paths on the accuracy of the MC $\mathbb{E}$[Sig] estimate. Figures 4.2a and 4.2b show the effect of increasing N_{train} on the accuracy of the MC $\mathbb{E}$[Sig], with the time discretization fixed to $n_{\text{steps}} = 252$. In both the GBM and OU models, the left panels indicate that the

proportion of coefficients whose theoretical values lie within the Monte Carlo 95% confidence interval is already high even for a moderate number of paths. This confirms that, once the time discretization is sufficiently fine, the MC $\mathbb{E}$[Sig] estimate quickly becomes comparable to its analytical counterpart.



(a) Black-Scholes model: proportion (%) of coefficients the 95% confidence interval (left) and corresponding mean relative error (%) (right) per level of truncation k vs time steps.



(b) Ornstein–Uhlenbeck model: proportion (%) of coefficients the 95% confidence interval (left) and corresponding mean relative error (%) (right) per level of truncation k vs time steps.

Figure 4.2: Comparison between the Monte Carlo and theoretical expected signatures across number of paths $N_{paths} \in \{5,000,\ 10,000,\ 25,000,\ 50,000,\ 75,000\}$ for the one-dimensional Black-Scholes (top) and Ornstein-Uhlenbeck (bottom) models. Each left panel reports the percentage of coefficients whose theoretical values lie within the 95% confidence interval of the Monte Carlo estimate for different truncation orders k; each right panel shows the corresponding mean relative error (%) as a function of the number of paths for different truncation orders k.

The right panels, however, reveal a clear difference between the two models. For the GBM (Figure 4.2a), the mean relative error is noticeably larger and decreases steadily as more paths are added, reflecting the higher variance inherent in the GBM dynamics. In contrast, for the OU model (Figure 4.2b), the mean relative error is already very small for 5,000 paths and shows no substantial improvement with additional samples. This behaviour is consistent with the smoother, mean-reverting dynamics of the OU process, which yield more stable Monte Carlo estimates. To maintain a unified framework across models and truncation levels, we continue to use $N_{\text{train}} = 5{,}000$ paths for the remainder of the numerical experiments.

4.2. Regression Settings

This section specifies the regression settings used to estimate the linear functional ℓ in the pricing framework. Two main aspects are analyzed: the choice of the regularization method and the truncation order of the signature, which together determine the stability of the approximation.

4.2.1. Regularization

Different regularization strategies can be employed for the estimation of the coefficient vector ℓ, such as Ridge, Lasso, and Elastic-Net regression. In this work, all three approaches were tested, and no consistent improvement was observed when using Lasso or Elastic-Net over Ridge in terms of either

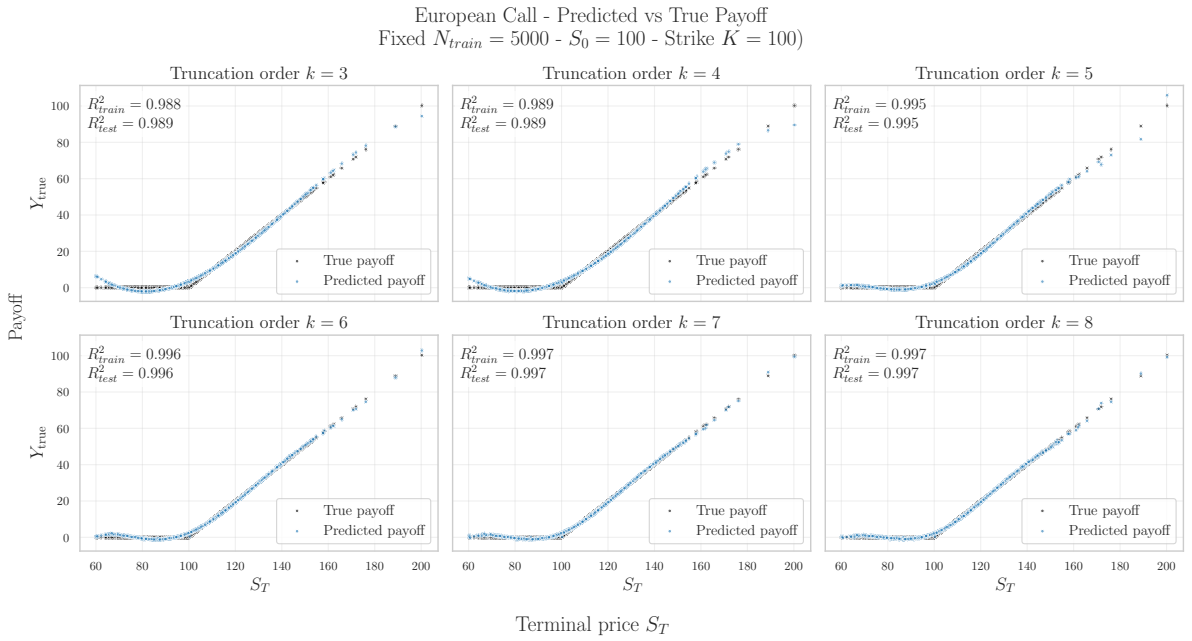


Figure 4.3: Predicted vs. true payoff for a European call at fixed $N_{\text{train}} = 5{,}000$ and varying truncation k

payoff approximation accuracy or the resulting option prices. Therefore, following common practice in the literature on regression frameworks for signatures [16], Ridge regression with a fixed regularization parameter $\alpha = 1$ is adopted throughout the experiments. This choice ensures a stable and well-conditioned estimator while avoiding sensitivity to hyperparameter tuning, especially since no noticeable improvement is observed in the resulting price estimates (beyond small differences in the payoff approximation).

4.2.2. Truncation Order

In this section we turn to the approximation power of the signature and determine what can be considered an acceptable level of accuracy for the payoff representation. This allows us to identify a range of truncation orders that provide a good trade-off between expressiveness of the signature representation and numerical stability, which we can then keep using for the remainder of the experiments.

We perform this preliminary analysis on an European call option, which serves as a canonical benchmark in option pricing. Being one of the most liquid and analytically tractable derivatives, it provides a clean test case for studying the approximation capabilities of the signature representation before extending the analysis to more complex payoffs.

Figures 4.3 shows the predicted versus the true payoff for an European call option, obtained using truncation orders ranging from $k = 3$ to $k = 8$, and training size $N_{\text{train}} = 5{,}000$. For each truncation order, an independent test set equal to 20% of the training set is generated for out-of-sample evaluation. The number of training paths are chosen large enough to provide a reliable regression estimate but still moderate to observe the difference between the truncation orders. As k increases, the regression progressively captures the non-linear structure of the payoff, with the coefficient of determination improving from $R^2 \approx 0.988$ at $k = 3$ to $R^2 \approx 0.997$ at $k = 7 - 8$. This indicates that, beyond $k = 6$, the signature features already provide a sufficiently expressive basis to approximate the call payoff accurately.

Our goal is not solely to fit the payoff itself but to obtain accurate and stable price estimates. For this reason, it is essential to assess whether the improved in-sample fit observed for larger truncation orders actually translates into better pricing performance, or whether it merely reflects overfitting to the training data. To investigate this, we compute for the same range of truncation orders the corresponding price estimates as the number of training paths N_{train} increases, and compare them with the benchmark Black-Scholes price. This allows us to quantify the extent to which increasing k genuinely improves the approximation of the pricing functional, and to identify the point beyond which additional signature

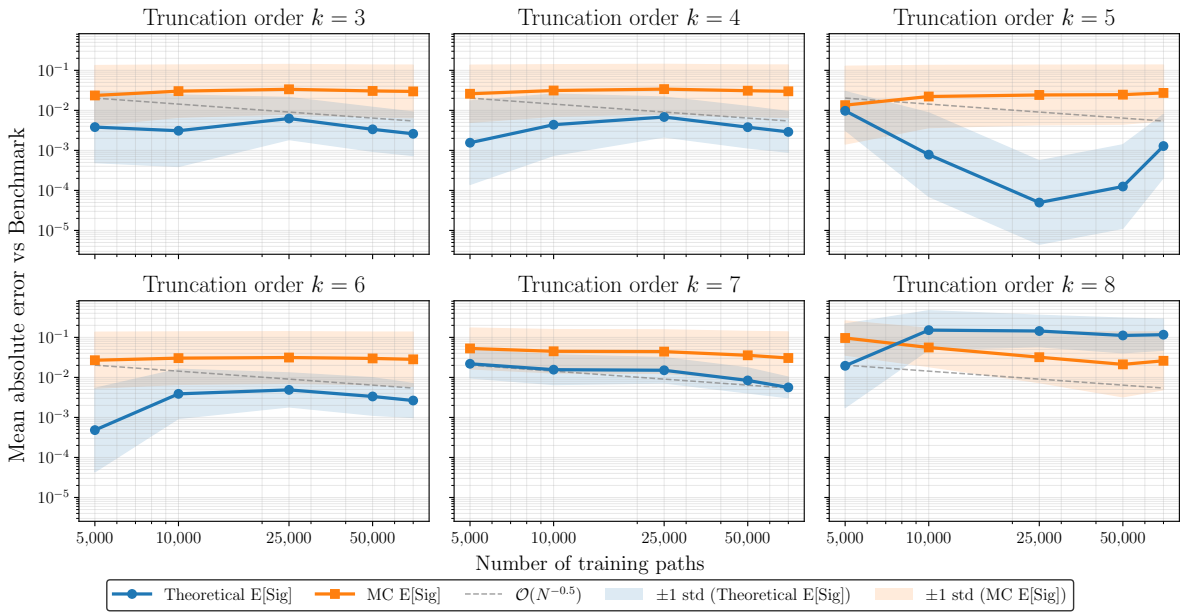


Figure 4.4: Mean absolute pricing error with respect to the benchmark price (Black-Scholes formula) as a function of the number of training paths $N_{\text{train}} \in \{5{,}000\,,\ 10{,}000\,,\ 25{,}000\,,50{,}000\,,\ 75{,}000\}$, for truncation orders $k \in \{3, \ldots, 8\}$

terms no longer contribute meaningfully to pricing accuracy.

Although the regression achieves very high coefficients of determination ($R^2 \approx 0.99$) for $k = 4 - 5$, (as seen in Figure 4.3), the associated price estimates in Figure 4.4 do not improve at the same rate. The figure shows the mean absolute pricing error with respect to the Black-Scholes price as a function of the number of training paths, for each truncation order $k \in \{3, ..., 8\}$. The prices reported in the picture are computed using the Theoretical $\mathbb{E}$[Sig] (blue) and the MC $\mathbb{E}$[Sig] (orange). For low and intermediate truncation orders ($k = 3 - 6$), both pricing estimates remain stable across the train sizes N_{train} considered. For higher truncation orders ($k \geq 7$), however, the behavior slightly changes. While the MC $\mathbb{E}$[Sig]-based prices remain relatively stable, the prices computed using the Theoretical $\mathbb{E}$[Sig] shift upward for all training sizes as k increases. Overall, truncation levels around $k = 4 - 6$ provide the most reliable and stable prices, whereas higher orders introduce a persistent upward bias in the Theoretical $\mathbb{E}$[Sig] price estimates without delivering noticeable improvements in accuracy. This behavior indicates that the linear regression introduces a consistent bias in the price estimation, which persists even as the number of training paths grows. Truncation levels around $k = 4 - 6$ provide the most reliable balance between payoff approximation, quality and pricing stability, thus we use them in the following experiments to see how other payoffs are approximated.

4.3. Pricing results

This section presents the pricing results obtained using the signature-based regression approach introduced in Chapter **??**. Our main objective is to evaluate how signature-based regression compares with standard Monte Carlo pricing. We consider several payoffs across multiple dimensions: starting with one-dimensional European and Asian call options, we move to a two-dimensional arithmetic basket option and finally address a three-dimensional FX model that incorporates stochastic interest rates. For each payoff, we report the convergence of the estimated prices as the number of training paths N_{train} increases, the variability across independent runs, and the sensitivity of the regression-based prices to the moneyness of the option. The regression settings are those fixed in 4.2.1.

4.3.1. One Dimensional Case: European and Asian Call

European Call We start with the variability of the option price across independent runs. For a fixed configuration of $N_{\text{train}} = 5000$, truncation order $k = 4$, and strike $K = 100$, we perform $n_{\text{runs}} = 20$

independent runs, which we present in Figure 4.5. This representation highlights the variability due to random sampling and provides an indication of the stability of the regression-based prices. Although the three approaches oscillate around the benchmark price, the plot illustrates that the variance across independent runs is remarkably smaller with the theoretical expected signature compared to the MC$\mathbb{E}$[Sig] and direct Monte Carlo approaches.

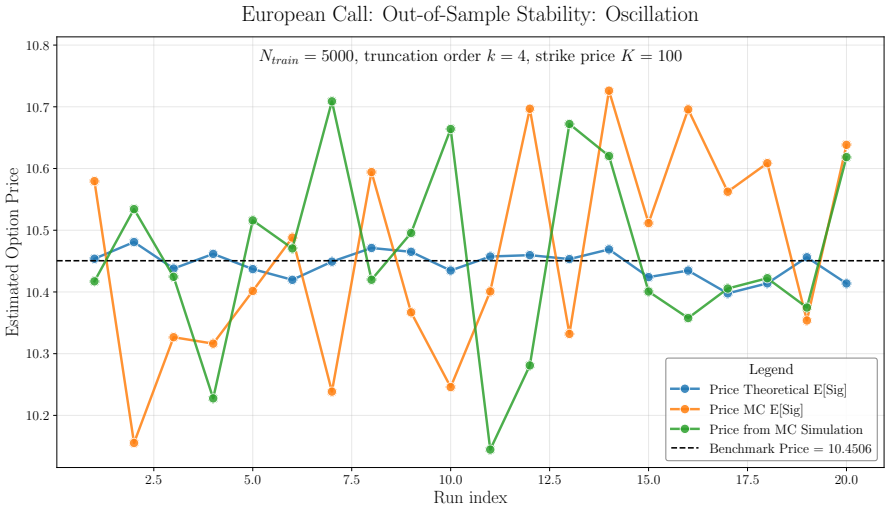


Figure 4.5: Oscillation of the signature-based estimator for the at-the-money European call ($K = 100$, $S_0 = 100$, $r = 5\%$, $\sigma = 20\%$, $T = 1$, $k = 4$, 20 independent runs of 5,000 training paths each). Each line reports the estimated option price obtained with a different pricing approach, as listed in the legend. The dashed line marks the analytical Black-Scholes benchmark (10.4506).

The variance reduction effect is further shown in Table 4.3, which also analyzes the effect of moneyness on the estimated prices across all strike levels.

The signature-based regression using the theoretical expected signature achieves prices that are very close to the analytical Black-Scholes benchmarks, with absolute errors below 0.003 and relative errors consistently below 0.02%. The standard deviations across runs remain small, confirming the high stability already observed in Figure 4.5. When the expected signature is estimated via Monte Carlo, the mean prices remain accurate, but the variance increases by roughly one order of magnitude, reflecting the additional noise introduced by sampling the expected signature empirically. The dispersion of the Monte Carlo expected signature results is comparable to that of direct Monte Carlo pricing, suggesting that both approaches are similarly affected by sampling variability.

Asian Call We next consider an Asian call (introduced in Section) option to evaluate the signature-based regression framework on a more path-dependent payoff. The parameters used are the same as those defined for the European Call.

Figure 4.6 reports the relative pricing error of the Asian call option as a function of the number of training paths, for truncation orders $k = 4, 5, 6$. For each truncation level, both pricing approaches, using Theoretical $\mathbb{E}$[Sig] (blue) and the MC $\mathbb{E}$[Sig] (orange), display a stabilization trend as the training size increases, showcasing the bias in the price estimation introduced by the linear regression. The theoretical $\mathbb{E}$[Sig] price estimator achieves lower variance compared to the MC $\mathbb{E}$[Sig].

To assess the variability of the price, we report in Figure 4.7 the oscillation of the estimated prices across 20 independent runs with $N_{\text{train}} = 5000$ and truncation order $k = 4$. The benchmark is computed using Monte Carlo with $100,000$ simulated paths and the corresponding confidence interval is of level 99%. Theoretical $\mathbb{E}$[Sig] prices (blue) exhibit minimal oscillation and remain always within the confidence interval. In contrast, the MC $\mathbb{E}$[Sig] (orange) and direct Monte Carlo (green) estimates fluctuate more noticeably, reflecting again their dependence on the sampling noise in the simulated paths and on the expected signature estimator. Nevertheless, all methods oscillate around the benchmark level. These results are similar to the ones obtained for the European call, with less pronounced oscillations due to the nature of the Asian payoff.

Table 4.3: Signature-based pricing of one-dimensional European call options under the Black–Scholes model ($r = 5\%$, $\sigma = 20\%$, $T = 1$, 20 runs of 5,000 training paths each). Benchmark prices are computed using the analytical Black–Scholes formula.

Strike K	Approach	Mean Price	Std. Dev.	Abs. Error	Rel. Error (%)
Benchmark: Reference Exact = 16.6994					
90 (ITM)	Theoretical $\mathbb{E}$[Sig]	16.7010	0.0200	0.0016	0.010
	MC $\mathbb{E}$[Sig]	16.7180	0.1020	0.0186	0.111
	Direct MC Simulation	16.7090	0.1380	0.0096	0.057
Benchmark: Reference Exact = 10.4506					
100 (ATM)	Theoretical $\mathbb{E}$[Sig]	10.4521	0.0247	0.0015	0.0149
	MC $\mathbb{E}$[Sig]	10.4246	0.1081	0.0259	0.249
	Direct MC Simulation	10.4588	0.1496	0.0082	0.078
Benchmark: Reference Exact = 6.0401					
110 (OTM)	Theoretical $\mathbb{E}$[Sig]	6.0415	0.0230	0.0014	0.023
	MC $\mathbb{E}$[Sig]	6.0580	0.1000	0.0179	0.296
	Direct MC Simulation	6.0480	0.1430	0.0079	0.131

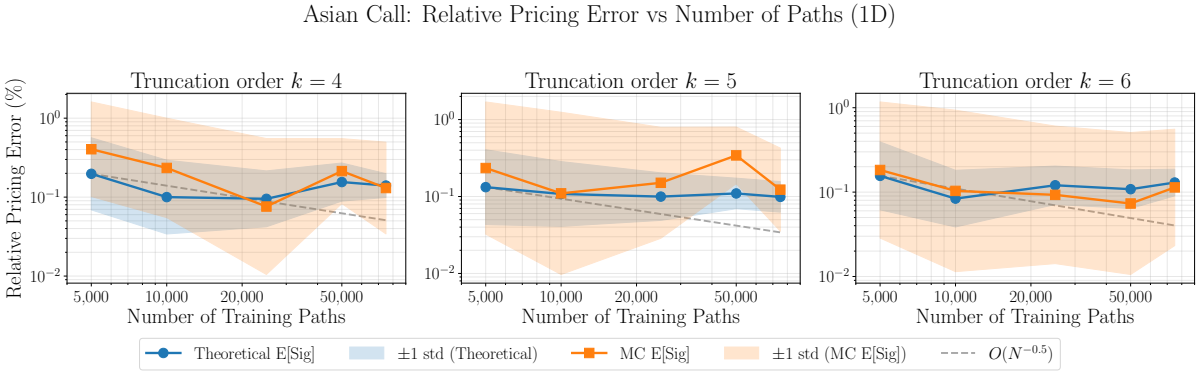


Figure 4.6: Relative pricing error for the one-dimensional Asian call option ($S_0 = 100$, $K = 100$, $r = 5\%$, $\sigma = 20\%$, $T = 1$) as a function of the number of training paths. Results are shown for truncation orders $k \in \{4, 5, 6\}$, comparing the regression based on the theoretical expected signature (blue) and on its Monte Carlo estimate (orange). Shaded areas represent ±1 standard deviation across 20 independent runs, shown symmetrically in log scale (i.e., multiplicatively around the mean). The dashed gray line illustrates the reference Monte Carlo convergence rate $\mathcal{O}(N^{-1/2})$

To further assess the robustness of the pricing framework, we analyze the performance across different moneyness levels (ITM, ATM, OTM). Again, the mean prices are computed over 20 independent runs with $N_{\text{train}} = 5{,}000$ and truncation order $k = 4$, as reported in Table 4.4. The overall behavior closely mirrors that of the European case in Table 4.3: both the theoretical and Monte Carlo versions of $\mathbb{E}$[Sig] yield accurate and stable prices across all strikes, with the theoretical estimator providing the most consistent results. A clear trend is observed in the standard deviations, where the Theoretical $\mathbb{E}$[Sig] approach achieves the smallest variability among all methods.

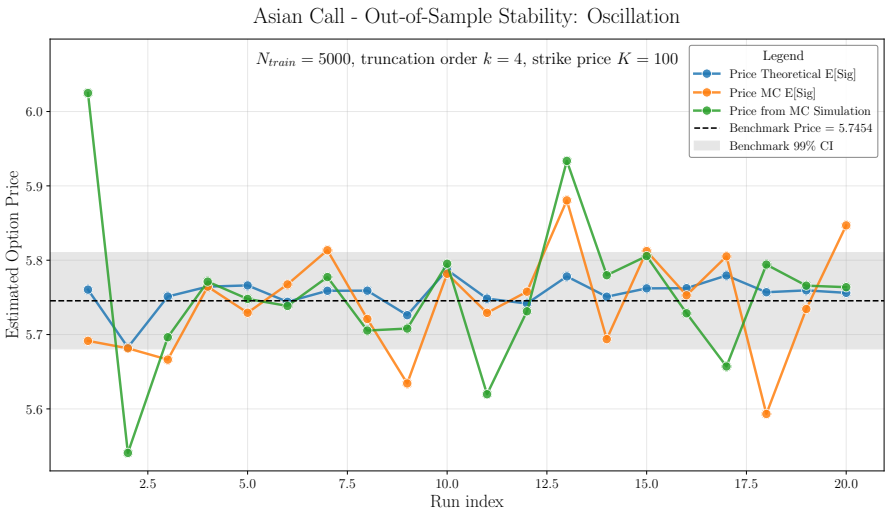


Figure 4.7: Oscillation of the signature-based estimator for the at-the-money Asian call ($K = 100$, $S_0 = 100$, $r = 5\%$, $\sigma = 20\%$, $T = 1$, $k = 4$, 20 independent runs of 5,000 training paths each). Each line reports the estimated option price obtained with a different pricing approach, as listed in the legend. The dashed line marks the Monte Carlo estimate of the price computed with 100,000 paths used as benchmark (5.7454), and the shaded gray areas represent the 99% confidence interval ($5.68723 - 5.81453$)

Table 4.4: Signature-based pricing of one-dimensional Asian call options under the Black-Scholes model ($r = 5\%$, $\sigma = 20\%$, $T = 1$, 20 runs of 5,000 training paths each). Benchmark prices are computed from 100,000 Monte Carlo simulations.

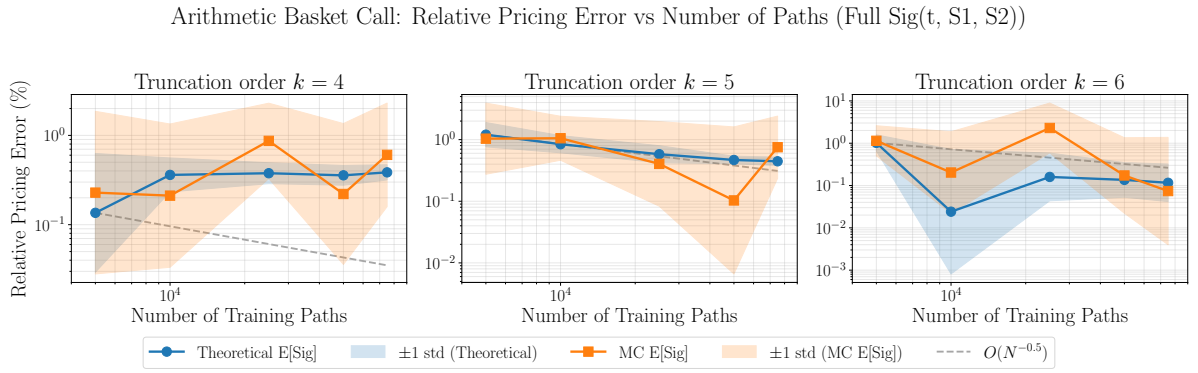
Strike K	Approach	Mean Price	Std. Dev.	Abs. Error	Rel. Error (%)
Benchmark: Reference MC Price = 12.5903 CI: (12.5055 – 12.6750)					
90 (ITM)	Theoretical $\mathbb{E}$[Sig]	12.5947	0.0117	0.0044	0.0349
	MC $\mathbb{E}$[Sig]	12.5894	0.0422	0.0009	0.0071
	Direct MC Simulation	12.5946	0.0596	0.0044	0.0349
Benchmark: Reference MC Price = 5.7677 CI: (5.7028 – 5.8326)					
100 (ATM)	Theoretical $\mathbb{E}$[Sig]	5.7548	0.0218	0.0129	0.2241
	MC $\mathbb{E}$[Sig]	5.7428	0.0711	0.0249	0.4309
	Direct MC Simulation	5.7543	0.1014	0.0134	0.2330
Benchmark: Reference MC Price = 1.9824 CI: (1.9425 – 2.0222)					
110 (OTM)	Theoretical $\mathbb{E}$[Sig]	1.9842	0.0113	0.0018	0.0914
	MC $\mathbb{E}$[Sig]	1.9749	0.0569	0.0075	0.3769
	Direct MC Simulation	1.9861	0.0783	0.0038	0.1900

4.3.2. Two Dimensional Case: Arithmetic Basket

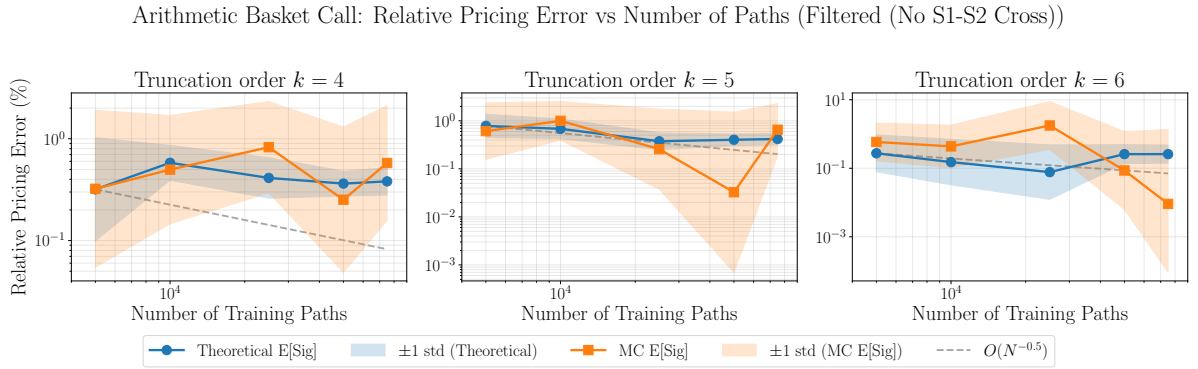
We now turn to the two-dimensional setting, where we consider an arithmetic basket option. We test the performance of the two signature representations introduced in Section 3.4. This experiment extends the pricing framework to multiple underlying assets, allowing us to assess how the inclusion of cross-asset interactions affects the quality of the pricing estimate.

Figure 4.8 reports the relative pricing error of the arithmetic basket call option for the full (panel 4.8a) and filtered (panel 4.8b) signature representations. As observed for the one dimensional case, there is no clear sign of convergence as the number of training paths increases, with the error oscillating around 0.3% for $k = 4$ and 1% for $k = 5$. The full and filtered representations yield remarkably similar results overall, with overlapping error values across all settings, except for $k = 6$, where the filtered signature

displays a lower pricing error. This suggests that, at least in the two-dimensional arithmetic basket setting, removing the cross-terms does not compromise the accuracy of the model.



(a) Full signature representation.



(b) Filtered signature representation.

Figure 4.8: Convergence of the arithmetic basket call option prices for the two signature-based regression approaches. Each panel reports the relative pricing error as a function of the number of training paths ($S_0^{(1)} = 95$, $S_0^{(2)} = 105$, $K = 100$, $r = 5\%$, $\sigma_1 = 20\%$, $\sigma_2 = 30\%$, $\rho = 0.5$, $T = 1$). Results are shown for truncation orders $k \in \{4, 5, 6\}$, comparing the regression based on the theoretical expected signature (blue) and on its Monte Carlo estimate (orange). Shaded areas represent ±1 standard deviation across 20 independent runs, shown symmetrically in log scale (i.e., multiplicatively around the mean). The dashed gray line illustrates the reference $\mathcal{O}(N^{-1/2})$ rate.

To complement the analysis, Figure 4.9 examines the stability of the pricing estimates, showing on the left the results for the full signature approach and on the right those for the filtered one. The experiment follows the same setup as for the European and Asian options. In the left panel, the price computed using the Theoretical $\mathbb{E}$[Sig] (blue) remains consistently within the confidence interval of the benchmark (shaded gray area). It exhibits mild oscillations, comparable to those observed in Figure 4.7, though the overall amplitude remains limited. The right panel shows the results of the price computed without considering the cross terms in the linear regression. Here, the theoretical expected signature oscillates slightly more (stdev= 0.0550) compared MC $\mathbb{E}$[Sig]MC expected signature (orange) exhibits little change in variability. Overall, both approaches produce prices that oscillate around the benchmark level across all runs. As in the previous experiments, a clear reduction in variance is observed for the theoretical expected signature, confirming its superior stability relative to the Monte Carlo-based estimators.

Table 4.5 reports the pricing results across different moneyness levels, comparing the full and filtered signature representations. both representations yield accurate price estimates close to the Monte Carlo benchmark. The full signature exhibits slightly smaller relative errors and lower variability than the filtered one, although their overall performance remains comparable. This suggests that the cross-asset interaction terms retained in the full signature contribute only marginally to improving numerical stability and precision, and that, at least for the arithmetic basket case, both representations perform similarly across all moneyness levels, confirming the hypothesis that relevant and enough information is

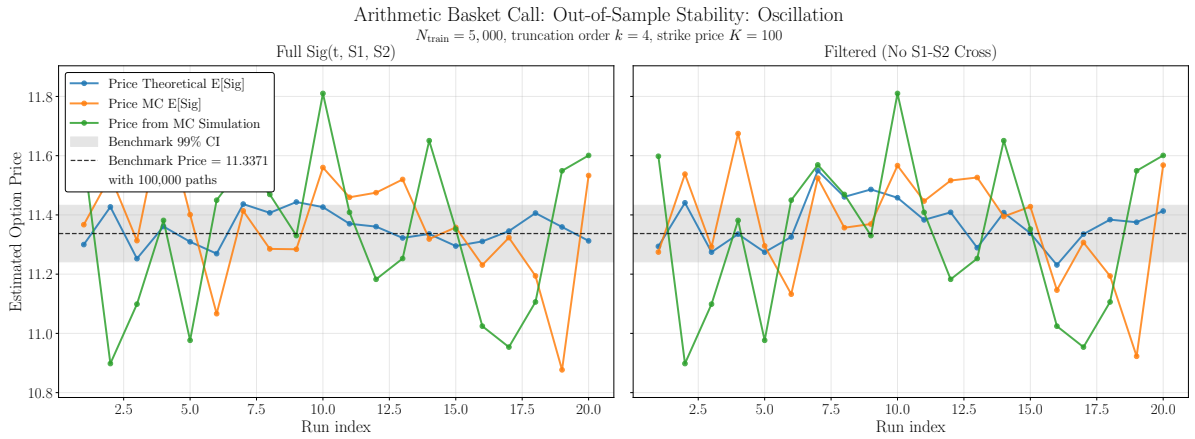


Figure 4.9: Oscillation of the signature-based estimator for the at-the-money arithmetic basket call ($K = 100$, $S_0^{(1)} = 95$, $S_0^{(2)} = 105$, $r = 5\%$, $\sigma_1 = 20\%$, $\sigma_2 = 40\%$, $\rho = 0.5$, $T = 1$, $k = 4$, 20 independent runs of 5,000 training paths each). The left panel shows the full signature approach, while the right panel shows the filtered variant. Each line reports the estimated option price obtained with a different pricing method, as listed in the legend. The dashed lines corresponds to the benchmark price (11.3371) computed from 100,000 Monte Carlo simulations with 99% confidence intervals (11.2417 – 11.4325).

Table 4.5: Comparison of pricing results of a two-dimensional arithmetic basket call option across moneyness levels for the full and filtered signature representations. Each result is averaged over 20 independent runs of 5,000 training paths. Benchmark prices are computed from 100,000 Monte Carlo simulations with 99% confidence intervals reported in parentheses.

Strike K	Approach	Mean Price	Std. Dev.	Abs. Error	Rel. Error (%)
Benchmark: Reference MC Price = 16.6594 CI: (16.5668 – 16.7519)					
90 (ITM)	Full Theoretical $\mathbb{E}$[Sig]	16.6485	0.0251	0.0109	0.0653
	Full MC $\mathbb{E}$[Sig]	16.6680	0.2016	0.0086	0.0518
	Filtered Theoretical $\mathbb{E}$[Sig]	16.6704	0.0586	0.0110	0.0660
	Filtered MC $\mathbb{E}$[Sig]	16.6942	0.1883	0.0348	0.2088
	Direct MC Simulation	16.6598	0.1928	0.0004	0.0022
Benchmark: Reference MC Price = 11.3848 CI: (11.3543 – 11.4152)					
100 (ATM)	Full Theoretical $\mathbb{E}$[Sig]	11.3679	0.0287	0.0169	0.1486
	Full MC $\mathbb{E}$[Sig]	11.4007	0.2005	0.0159	0.1397
	Filtered Theoretical $\mathbb{E}$[Sig]	11.4041	0.0550	0.0193	0.1695
	Filtered MC $\mathbb{E}$[Sig]	11.4406	0.1926	0.0558	0.4901
	Direct MC Simulation	11.4105	0.1979	0.0257	0.2257
Benchmark: Reference MC Price = 7.5535 CI: (7.4638 – 7.6431)					
110 (OTM)	Full Theoretical $\mathbb{E}$[Sig]	7.5220	0.0382	0.0315	0.4168
	Full MC $\mathbb{E}$[Sig]	7.5677	0.1865	0.0142	0.1882
	Filtered Theoretical $\mathbb{E}$[Sig]	7.5633	0.0392	0.0099	0.1306
	Filtered MC $\mathbb{E}$[Sig]	7.6124	0.1716	0.0590	0.7807
	Direct MC Simulation	7.5890	0.1822	0.0355	0.4705

retained in the single asset time-augmented signatures. Within each representation, the prices computed with the theoretical expected signature clearly outperform those computed with their Monte Carlo counterpart in terms of stability.

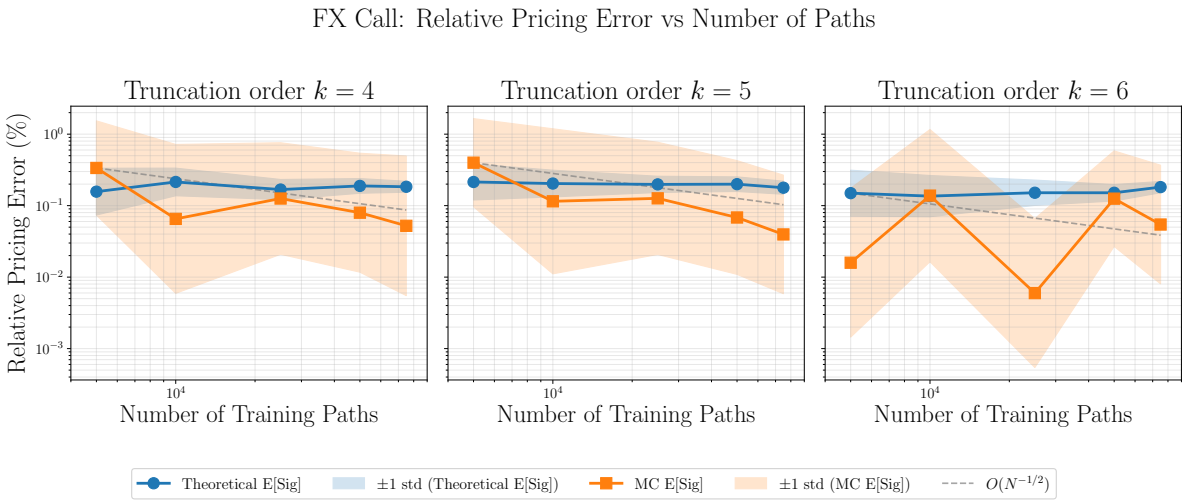


Figure 4.10: Relative pricing error for the one-dimensional FX derivative (parameters as defined in Table 4.2, $K = 1.10$) as a function of the number of training paths. Results are shown for truncation orders $k \in \{4, 5, 6\}$, comparing the regression based on the Theoretical $\mathbb{E}$[Sig] (blue) and on its Monte Carlo estimate (orange). Shaded areas represent ± 1 standard deviation across 20 independent runs, shown symmetrically in log scale (i.e., multiplicatively around the mean). The dashed gray line illustrates the convergence rate $\mathcal{O}(N^{-1/2})$

Overall, the two-dimensional arithmetic basket experiment confirms the findings observed in the one-dimensional setting: the Theoretical $\mathbb{E}$[Sig] yields the most stable and accurate results, exhibiting substantially reduced variance compared to the MC $\mathbb{E}$[Sig]e. Regarding the comparison between the full and filtered representations, the two approaches deliver broadly similar accuracy: the filtered signature shows slightly larger oscillations around the benchmark, but the degradation is mild and does not compromise the overall quality of the estimates. Building on these insights, we next consider the three-dimensional setting corresponding to the foreign exchange (FX) model with stochastic domestic and foreign interest rates, to examine whether a variant of the filtered representation, the hybrid representation, continues to perform competitively in a higher-dimensional and more complex setting.

4.3.3. Three Dimensional Case: Pricing FX derivative

This experiment extends the signature-based regression framework to a higher-dimensional setting, where the payoff depends jointly on the exchange rate and the short-rate dynamics of both the domestic and foreign currencies. This hybrid signature is constructed by combining the signatures of the time-augmented GBM and the time-augmented two-dimensional OU processes, as detailed in Section 3.4.2.1 and the corresponding expected signature as in Section 3.5.5. The goal is to assess whether this reduced representation still provides a sufficiently accurate and stable approximation of the joint path to yield reliable price estimates in this more challenging setting.

Figure 4.10 shows that the regression based on the Theoretical $\mathbb{E}$[Sig] remains clearly more stable than the Monte Carlo design. The MC $\mathbb{E}$[Sig] exhibits substantially wider standard deviations, although the width of the variability bands decreases slightly as the number of paths used to estimate it increases. This indicates that additional samples improve stability to some extent, but the Monte Carlo approach remains considerably more variable than the Theoretical $\mathbb{E}$[Sig] across all training sizes.

The oscillation behavior is observed in Figure 4.11, which reports the price obtained from the different pricing methods over 20 independent runs, for a fixed number of training paths equal to $5,000$. Here, using the Theoretical $\mathbb{E}$[Sig] produces values tightly clustered around the benchmark Monte Carlo price (computed using $100,000$ paths). The MC $\mathbb{E}$[Sig], on the other hand, yields more unstable estimation of the price across the runs, but with a standard deviation reduced compared to direct Monte Carlo simulation price estimate. This suggests that the hybrid structure successfully captures the essential interactions among the three processes.

Table 4.6: Out-of-sample stability analysis for FX call options (20 runs of 5,000 training paths each). Benchmark prices are computed via Monte Carlo simulations (CI at 99% confidence level).

Strike K	Approach	Mean Price	Std. Dev.	Abs. Error	Rel. Error (%)
Benchmark:	Reference MC Price = 0.128064	CI: (0.127443 – 0.128685)			
	Theoretical $\mathbb{E}$[Sig]	0.128102	0.000105	0.000038	0.030
0.99 (ITM)	MC $\mathbb{E}$[Sig]	0.128417	0.001195	0.000353	0.276
	Direct MC Simulation	0.128105	0.001488	0.000041	0.032
Benchmark:	Reference MC Price = 0.081599	CI: (0.081057 – 0.082141)			
	Theoretical $\mathbb{E}$[Sig]	0.081464	0.000146	0.000135	0.166
1.05 (ATM)	MC $\mathbb{E}$[Sig]	0.081220	0.001046	0.000379	0.464
	Direct MC Simulation	0.081870	0.001561	0.000271	0.332
Benchmark:	Reference MC Price = 0.046312	CI: (0.045881 - 0.046742)			
	Theoretical $\mathbb{E}$[Sig]	0.045979	0.000129	0.000332	0.717
1.11 (OTM)	MC $\mathbb{E}$[Sig]	0.046290	0.000818	0.000021	0.046
	Direct MC Simulation	0.046329	0.000821	0.000018	0.038

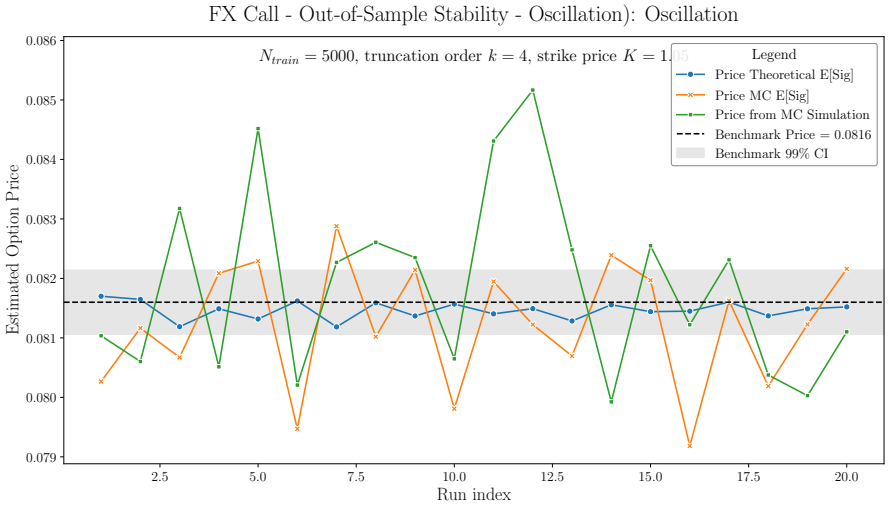


Figure 4.11: Oscillation of the signature-based estimator for the at-the-money FX derivative (parameters as defined in Table 4.2, $K = 1.10$). Each line reports the estimated option price obtained with a different pricing approach, as listed in the legend. The dashed line marks the Monte Carlo estimate of the price computed with 100,000 paths used as benchmark (0.0816). The shaded gray area represents the Monte Carlo 99% CI (0.081057 – 0.082141)

The quantitative results in Table 4.6 complete the pricing analysis and provide a detailed view of the estimator performance across different levels of moneyness. Across ITM, ATM and OTM strike values K, the Theoretical $\mathbb{E}$[Sig] consistently yields the smallest standard deviation, with a variability of order 10^{-4}. The MC $\mathbb{E}$[Sig] and the direct Monte Carlo simulation both perform reasonably well, but their variability is an order of magnitude larger. In terms of accuracy, measured by absolute and relative errors, the behavior depends on moneyness. For the ITM and ATM strikes, the Theoretical $\mathbb{E}$[Sig] produces pricing errors that are extremely small and comparable to those of direct Monte Carlo. However, as the option moves out of the money, the accuracy degrades: the Theoretical $\mathbb{E}$[Sig] displays a noticeably larger error (relative error of 0.717%), whereas both the MC $\mathbb{E}$[Sig] and the direct Monte Carlo method remain much closer to the benchmark, with relative error below 0.05%. This suggests that the Theoretical $\mathbb{E}$[Sig] price, while highly stable, becomes slightly biased in the tail region where the payoff is influenced by rare behavior.

Taken together, these experiments indicate that the hybrid signature representation provides a robust and effective approach for pricing FX derivatives. Moreover, the Theoretical $\mathbb{E}[\text{Sig}]$, even in its hybrid form, continues to act as a strong variance-reduction mechanism compared with standard Monte Carlo simulation.

Having concluded the analysis of pricing accuracy and stability, we now examine the computational requirements of the signature-based framework.

4.4. Time Analysis and Computational Aspects

To quantify the practical computational burden of the signature-based pricing framework, we perform a timing analysis of the full pricing pipeline for a fixed truncation order $k = 4$, which has been identified in Section 4.3 as an appropriate level of truncation for pricing.

The total runtime is decomposed into the following components:

- path simulation: generation of N_{train} training paths under the model dynamics;
- payoff evaluation: computation of the target values for regression;
- model fitting: estimation of the regression coefficients ℓ;
- signature computation: evaluation of the truncated signature serving as the regression basis;
- expected signature computation: evaluation of the theoretical or Monte Carlo expected signature used for pricing.

The time required to obtain the final price, which corresponds to evaluating the expected payoff in the case of direct Monte Carlo, and to applying the estimated linear coefficients to the expected signature vector in the signature-based approaches, is negligible compared to the other components of the pipeline. Each component is measured separately to identify the main computational bottlenecks.

The following analysis is divided into two parts. First, we examine the computational cost of each stage of the pricing procedures. Second, we investigate the relation between computation time and pricing accuracy by measuring how long it takes to achieve a given level of absolute relative error and standard deviation. This complementary analysis allows us to assess whether the higher computational cost of the signature-based approach is offset by its improved statistical efficiency. In particular, we aim to quantify the behavior observed in Section 4.3, where the pricing standard deviation obtained with the Theoretical $\mathbb{E}[\text{Sig}]$ was consistently smaller than that based on its Monte Carlo estimate. The timing experiments are performed for the following three representative settings: the one-dimensional Asian call option; the two-dimensional arithmetic basket option, where we highlight the computational differences between the full and filtered representations; and the three-dimensional FX model, where we analyze the hybrid approach.

4.4.1. Timing Breakdown of the Pricing Pipeline

Figure 4.12 reports the average runtime decomposition of the Theoretical $\mathbb{E}[\text{Sig}]$ pricing pipeline for the Asian call option at truncation order $k = 4$. The total computational cost scales approximately linearly with the number of training paths, with the signature computation step representing the dominant contribution across all settings. Path simulation also accounts for a substantial portion of the total time, whereas the payoff evaluation and regression fitting steps are comparatively negligible. The computation of the theoretical expected signature adds an almost imperceptible overhead relative to the data generation phase. The red line shows the total runtime of the standard Monte Carlo estimator for the same number of paths. Both methods scale similarly with sample size, but the signature-based pipeline remains more expensive when compared on a per-path basis.

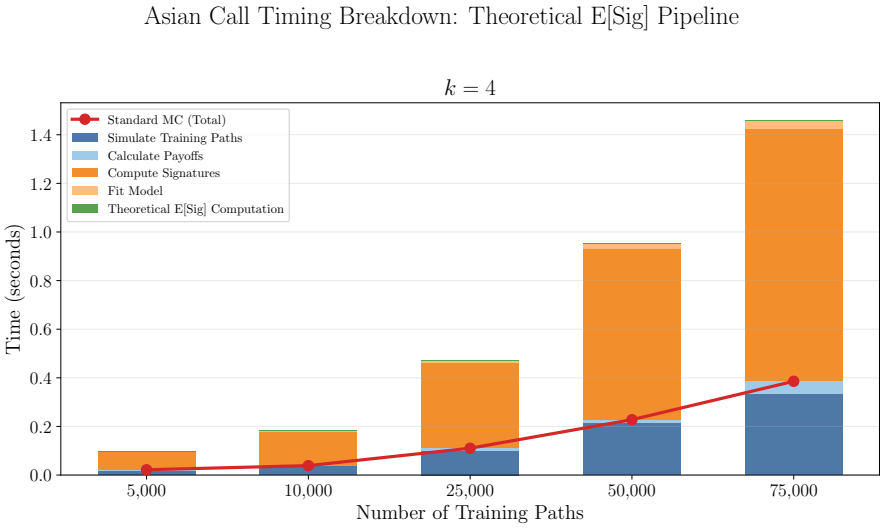


Figure 4.12: Timing breakdown of the theoretical expected signature pricing pipeline for the one-dimensional Asian call option at truncation order $k = 4$. Each stacked bar reports the average runtime across 20 independent runs, decomposed into the main computational components indicated in the legend. The red line indicates the total runtime of the standard Monte Carlo estimator for the same number of simulated paths.

Figure 4.13 compares the computational cost of the full (left) and filtered (right) signature representations at truncation order $k = 4$ for the two-dimensional arithmetic basket option. Both pipelines exhibit roughly linear growth in runtime with the number of training paths, and the overall cost is significantly higher compared to the one-dimensional case (Figure 4.12). The total runtime increases from approximately 0.35 seconds for 5,000 paths to about 5.8 seconds for 75,000 paths in the full approach, and to around 5.1 seconds for the filtered one. The simulation of paths and the computation of payoffs are common to both methods and account for a comparable share of the total runtime. The main difference arises in the signature computation step, which remains the dominant cost for both representations. For the full case, it represents about 3.8 seconds, over 65% of the total runtime at 75,000 paths, whereas for the filtered case the share is smaller, consistent with the lower number of iterated integrals to compute (120 versus 60). However, the reduction in runtime is not directly proportional to this halving in the number of terms, possibly due to additional overhead in reassembling the filtered components or to the non-linear scaling of the `iisignature` implementation with dimensionality.

We report runtime measurements for the pipeline using the *theoretical* expected signature; the Monte Carlo-based version yields essentially the same timings, since it replaces the Theoretical $\mathbb{E}[\mathrm{Sig}]$ computation with the empirical averaging of the same signature values already computed for the regression, without altering the cost of the remaining components.

Figure 4.14 reports the computational cost decomposition of the theoretical expected signature pipeline for the FX derivative at truncation order $k = 4$. Although no direct comparison with a full signature representation is available in this case, the results remain consistent with the two-dimensional arithmetic basket experiment. Adding one extra dimension to the underlying process increases the runtime, but not excessively, confirming the scalability of the approach with respect to dimensionality. Again, the most time is spent computing the signature of course. As in previous cases, the computation of path signatures constitutes the dominant contribution to the overall cost. It is worth noting that, in this experiment, the time required for the evaluation of the theoretical expected signature of the two-dimensional time-augmented Ornstein-Uhlenbeck process is non-negligible, mainly due to limited code optimization rather than intrinsic model complexity.

4.4.2. Runtime Analysis: Time to Accuracy and Stability

To complement the previous analysis, we examine how the mean relative error and the relative standard deviation evolve with total runtime, providing a direct measure of how efficiently each pricing approach attains a desired level of accuracy. As before, we consider the three option types analyzed in the

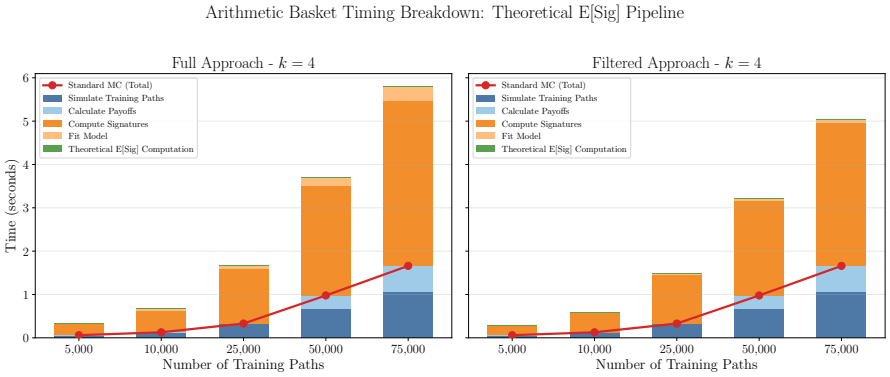


Figure 4.13: Timing breakdown of the theoretical expected signature pricing pipeline for the two-dimensional arithmetic basket call option at truncation order $k = 4$. The left panel shows the full signature representation, while the right panel reports the filtered variant. The red line indicates the total runtime of the standard Monte Carlo estimator for the same number of simulated paths.

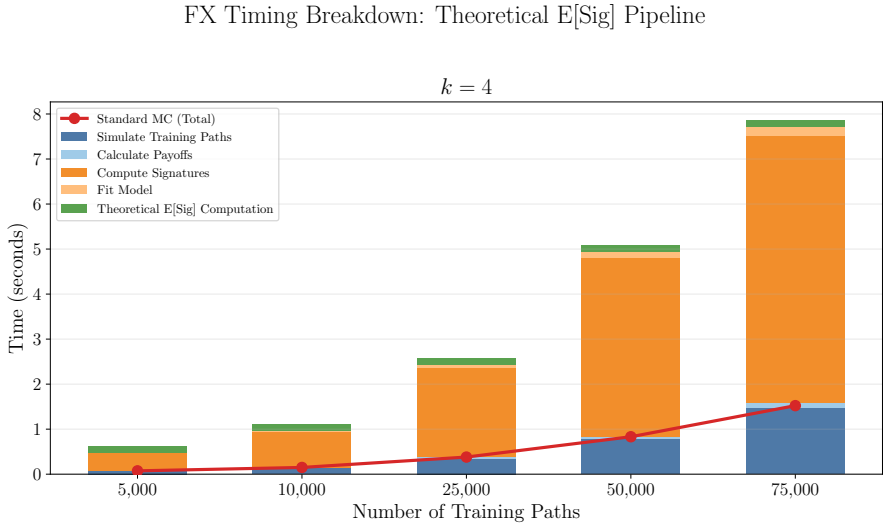
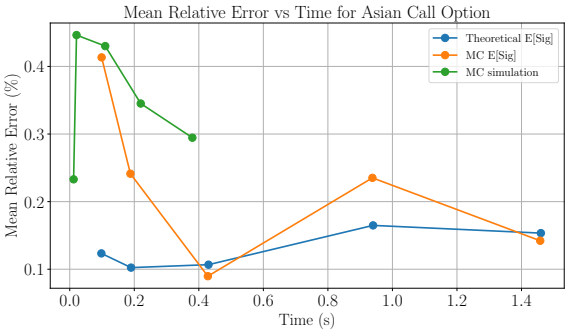
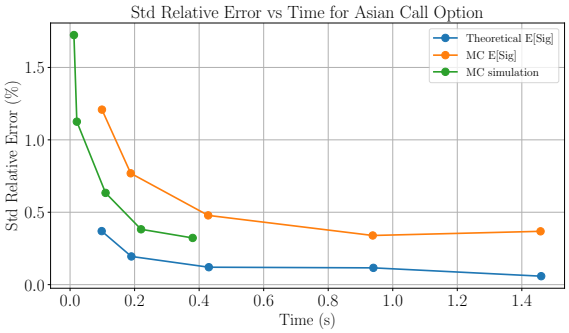


Figure 4.14: Timing breakdown of the Theoretical $\mathbb{E}$[Sig] pricing pipeline for the FX derivative model at truncation order $k = 4$. The stacked bars show the total average runtime decomposition into the main computational components indicated in the legend. The red line represents the total runtime of the standard Monte Carlo estimator for the same number of simulated paths.

previous sections: the one-dimensional Asian option, the two-dimensional arithmetic basket option, and the three-dimensional FX derivative. Again, we fix the order of truncation to 4.

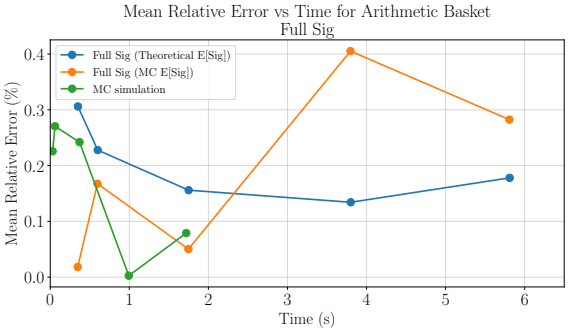


(a) Mean relative error versus computation time

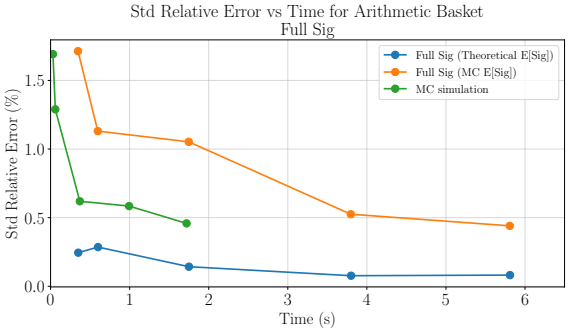


(b) Relative standard deviation versus computation time

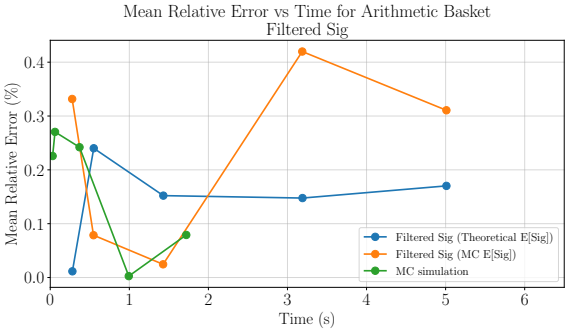
Figure 4.15: Time-accuracy analysis for the one-dimensional Asian call option. The left panel reports the mean relative error as a function of total runtime, while the right panel shows the evolution of the relative standard deviation.



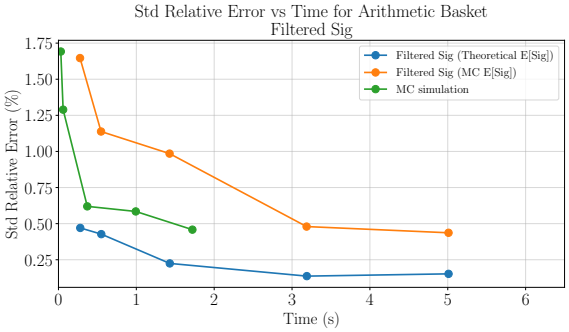
(a) Full signature: mean relative error



(b) Full signature: relative standard deviation

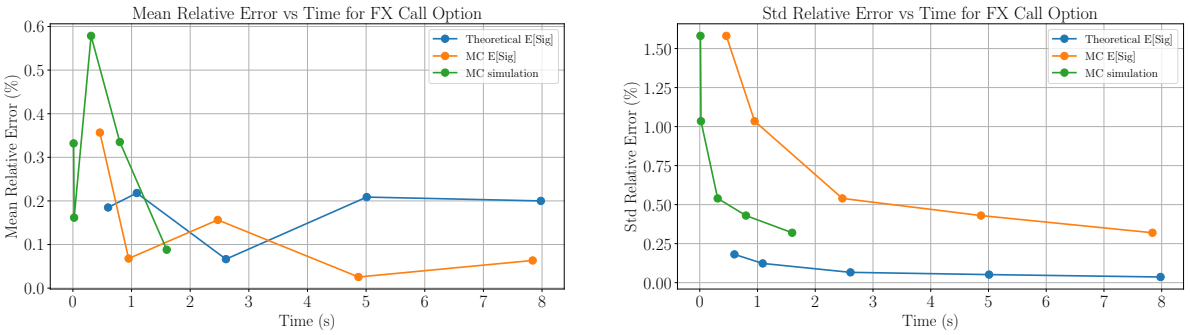


(c) Filtered signature: mean relative error



(d) Filtered signature: relative standard deviation

Figure 4.16: Comparison of computation time versus pricing accuracy for the two-dimensional arithmetic basket option. Top panels show the mean relative error; bottom panels show the relative standard deviation, for both full and filtered signature representations.



(a) Mean relative error versus computation time

(b) Relative standard deviation versus computation time

Figure 4.17: Time-accuracy analysis for the FX derivative pricing experiment. The left panel shows the evolution of the mean relative error with total runtime, while the right panel reports the corresponding relative standard deviation.

Figures 4.15-4.17b summarize the relation between computational time and pricing accuracy for all considered options. Across all cases, the same qualitative behavior is observed. Theoretical $\mathbb{E}$[Sig] prices (blue) consistently yield lower relative variance and reach stable accuracy levels faster than their Monte Carlo counterparts (orange), confirming the efficiency of the analytical formulation. The reduction in variability appears already for small training sizes and persists across payoffs and model dimensions, yielding to rather accurate prices with low variance fast. As observed in the convergence plots presented earlier (see Figures 4.6, 4.8a, 4.8b, 4.10), the relative error of the theoretical signature-based estimators does not decrease with the number of training paths and therefore does not improve with computational time. In contrast, the MC $\mathbb{E}$[Sig] benefits from larger sample sizes, since the empirical estimation of the expected signature becomes more accurate as the number of simulated paths increases.

For the two-dimensional arithmetic basket, the difference between the full and filtered signatures becomes visible mainly on the time side of the time-accuracy trade-off. Both representations achieve comparable levels of accuracy and relative standard deviation, and these stability levels are already reached for relatively small training sizes, regime in which the computational times of the two methods do not differ dramatically in this two-dimensional setting. Naturally, the reduction in computational time would become more pronounced in higher-dimensional We also observe that, for both the full and filtered representations, the Theoretical $\mathbb{E}$[Sig] prices lie consistently below the direct Monte Carlo estimates across all runtimes. This behavior mirrors the plots discussed earlier (Figures 4.8a, 4.8b). In contrast, the prices based on the MC $\mathbb{E}$[Sig] clearly exhibit larger fluctuations: their curves are more irregular, and the associated variance remains noticeably higher throughout the runtime range, compared to both Monte Carlo simulation and Theoretical $\mathbb{E}$[Sig] pipeline. For a fixed computational budget, the Monte Carlo-based estimates remain noticeably less stable.

Analogous considerations hold for the three dimensional case, presented in Figures 4.17a and 4.17b. The Theoretical $\mathbb{E}$[Sig] achieves stable accuracy levels rapidly and maintains the lowest relative variance across all computational budgets. Its mean relative error curve is relatively flat and remains consistently close to the benchmark, indicating that additional computational time does not substantially alter the price estimator once the regression coefficients have stabilized. In contrast, the price based on the MC $\mathbb{E}$[Sig] signature exhibits noticeably larger fluctuations, particularly at smaller runtimes, and only gradually settles as more computational time (that directly correspond to a larger number of paths to approximate the expected signature) is allocated to estimate the expected signature more accurately. Although direct Monte Carlo simulation becomes reasonably stable for larger budgets, its variance remains higher than that of the theoretical expected signature. This confirms that the hybrid signature representation performs well when combined with the theoretical expected signature, in full agreement with the behavior observed in both the one and two-dimensional settings. Across all dimensions, the main distinction between stable and unstable price estimates is not the choice of signature representation itself, but rather the method used to compute the expected signature.counterpart.

5

Conclusion

This thesis investigated the use of path signatures as a feature representation for derivative pricing through a regression-based framework. The universal approximation theorem for signatures provides the theoretical motivation behind this approach: it guarantees that any continuous functional of a path can be approximated arbitrarily well by a linear functional of its signature. This property makes the signature a powerful and expressive object for representing path-dependent quantities such as derivative payoffs. The study was structured around two research questions.

RQ1 asked how signature-based regression compares with standard Monte Carlo simulation for pricing derivatives. Across all experiments, the results show a substantial reduction in variance when using signature-based regression combined with the analytical expected signature. This stability improvement is achieved without loss in pricing accuracy: the method produces errors comparable to standard Monte Carlo. Because accurate and stable estimates are obtained with fewer simulated paths, the approach can yield savings in computational time, demonstrating its practical advantage for the derivatives considered.

RQ2 examined whether more compact signature representations of paths could mitigate the exponential growth in signature dimensionality, while retaining pricing accuracy and stability. To address this, two alternative path representations were defined and compared: the *full* signature, which includes all cross-asset iterated integrals, and the *filtered* signature, which contains only terms with individual assets or pre-defined asset groups (in its hybrid variation). The empirical results from the experiments conducted in this thesis showed that the filtered signature achieves pricing accuracy and standard deviation levels comparable to the full signature, while reducing the number of terms, and therefore computational cost of the pricing procedure. These findings also indicate that relevant and enough information for pricing is already encoded in the signatures of the individual time-augmented paths, including the dependence structure between assets, and that explicit cross-asset iterated integrals are not always necessary for price estimation purposes.

5.1. Limitations

Several important limitations should be acknowledged regarding this piece of research. First of all, the methodology is highly dependent on the underlying model assumptions. The expected signature is a fundamental quantity in the pricing pipeline, and its accurate evaluation is crucial for obtaining stable and reliable results. When an analytical expression for the expected signature is available, as for the Geometric Brownian Motion and Ornstein-Uhlenbeck models, the approach produces prices with low variance. However, for many stochastic models of practical interest, deriving a closed-form expression is not feasible. In such cases, one must resort to a Monte Carlo approximation, and the method becomes far less convenient. As seen in the experiments, achieving a sufficiently accurate estimate of the expected signature requires a large number of paths, making the approach computationally expensive, thus limiting its applicability in more complex models. Consequently, the framework is not particularly flexible.

A second limitation concerns the computational cost of the overall pipeline. While the number of simulated paths required to obtain accurate and stable price estimates is relatively small, the dominant bottleneck remains the cost of computing signatures themselves. Signature computation scales rapidly with both the path dimension and the truncation order, and the total runtime is largely determined by this step. The filtered representation alleviates the growth in the number of signature terms, bringing the dependence on the dimension down to linear. However, this structural improvement is not sufficient to compensate for the fact that currently available signature libraries are still relatively costly, especially in high dimensions or for high truncation orders. Consequently, scaling the method to larger-dimensional models or deeper truncations would require either more efficient signature engines or dedicated approximation strategies.

Finally, the regression step itself introduces practical limitations. Although regularization techniques mitigate overfitting, they also introduce hyperparameters that must be tuned, which increases computational complexity and sensitivity. In this thesis, a fixed Ridge parameter of $\alpha = 1$ was used across all experiments, an assumption that simplifies the framework, but may not be optimal. Moreover, the linear regression model introduces a systematic bias that limits the attainable accuracy of the method: the estimator cannot converge to the true price simply by increasing the number of training paths or by raising the truncation order.

5.2. Future Work

This research opens several avenues for future work, primarily focused on theoretical justification, computational efficiency, and model extension. A first direction is to develop a rigorous mathematical explanation for one of the key empirical findings of this thesis:the observation that the filtered signature appears to constitute a sufficiently rich representation for pricing purposes. Although the numerical experiments performed support this hypothesis, its theoretical justification remains open. Establishing conditions under which the filtered signature retains all the relevant path information would provide valuable insight into the structure and expressiveness of signature features.

From a computational point of view, future work could focus on improving the efficiency of the signature computation. The currently available libraries are not optimized for large-scale applications, and signature evaluation represents the main computational bottleneck. Developing dedicated algorithms or surrogate models to approximate signature values could substantially reduce runtime and make the approach more practical.

Finally, replacing the linear regression step with more flexible models, such as kernel methods or neural networks, could further enhance approximation accuracy while maintaining interpretability. However, this would require careful consideration, as one of the key advantages of the current linear framework is that the pricing formula remains analytically tractable once the expected signature is known. Preserving the balance between flexibility and analytical transparency constitutes a promising and challenging direction for future exploration.

References

[1] Ben Hambly and Terry Lyons. "Uniqueness for the signature of a path of bounded variation and the reduced path group". In: *Annuals of Mathematics* 171.1 (2010), pp. 109–167. URL: https://doi.org/http://doi.org/10.4007/annals.2010.171.109.

[2] Ilya Chevyrev and Terry Lyons. "Characteristic functions of measures on geometric rough paths". In: *The Annals of Probability* 44.6 (Nov. 2016). ISSN: 0091-1798. DOI: 10.1214/15-aop1068. URL: http://dx.doi.org/10.1214/15-AOP1068.

[3] Kuo-Tsai Chen. "Integration of Paths, Geometric Invariants and a Generalized Baker- Hausdorff Formula". In: *Annals of Mathematics* 65.1 (1957), pp. 163–178. ISSN: 0003486X, 19398980. URL: http://www.jstor.org/stable/1969671.

[4] Kuo-Tsai Chen. "Integration of Paths–A Faithful Representation of Paths by Noncommutative Formal Power Series". In: *Transactions of the American Mathematical Society* 89.2 (1958), pp. 395–407. ISSN: 00029947, 10886850. URL: http://www.jstor.org/stable/1993193.

[5] Terry J. Lyons. "Differential equations driven by rough signals." In: *Revista Matemática Iberoamericana* 14.2 (1998), pp. 215–310. URL: http://eudml.org/doc/39555.

[6] Terry Lyons Thierry Lévy and Michea Caruana. *Differential Equations Driven by Rough Paths*. Vol. 1908. 2007.

[7] Peter K. Friz and Nicolas B. Victoir. "Multidimensional Stochastic Processes as Rough Paths: Theory and Applications". In: Cambridge Studies in Advanced Mathematics. Cambridge University Press, 2010.

[8] Weixin Yang, Lianwen Jin, and Manfei Liu. *DeepWriterID: An End-to-end Online Text-independent Writer Identification System*. 2015. arXiv: 1508.04945 [cs.CV]. URL: https://arxiv.org/abs/1508.04945.

[9] James Morrill et al. "The signature-based model for early detection of sepsis from electronic health records in the intensive care unit". In: *2019 Computing in Cardiology (CinC)*. Vol. 46. 2020, pp. 1–4. DOI: 10.23919/CinC49843.2019.9005739.

[10] Weixin Yang et al. *Developing the Path Signature Methodology and its Application to Landmark-based Human Action Recognition*. 2019. arXiv: 1707.03993 [cs.CV]. URL: https://arxiv.org/abs/1707.03993.

[11] Christian Bayer et al. *Optimal stopping with signatures*. 2021. arXiv: 2105.00778 [math.PR]. URL: https://arxiv.org/abs/2105.00778.

[12] Eduardo Abi Jaber and Louis-Amand Gérard. "Signature Volatility Models: Pricing and Hedging with Fourier". In: *SIAM Journal on Financial Mathematics* 16.2 (May 2025), pp. 606–642. ISSN: 1945-497X. DOI: 10.1137/24m1636952. URL: http://dx.doi.org/10.1137/24M1636952.

[13] Imanol Pérez Arribas, Cristopher Salvi, and Lukasz Szpruch. *Sig-SDEs model for quantitative finance*. 2020. arXiv: 2006.00218 [q-fin.CP]. URL: https://arxiv.org/abs/2006.00218.

[14] Imanol Pérez Arribas. *Derivatives pricing using signature payoffs*. 2018. arXiv: 1809.09466 [q-fin.CP]. URL: https://arxiv.org/abs/1809.09466.

[15] James Morrill et al. *A Generalised Signature Method for Multivariate Time Series Feature Extraction*. 2021. arXiv: 2006.00873 [cs.LG]. URL: https://arxiv.org/abs/2006.00873.

[16] Adeline Fermanian. *Functional linear regression with truncated signatures*. 2022. arXiv: 2006.08442 [stat.ME]. URL: https://arxiv.org/abs/2006.08442.

[17] Wolfgang Hackbusch. *Tensor Spaces and Numerical Tensor Calculus*. Vol. 42. Springer Series in Computational Mathematics. Springer, 2012. ISBN: 978-3-642-28026-9.

[18] J. N. Bernstein and S. I. Gelfand. “Tensor products of finite and infinite dimensional representations of semisimple Lie algebras”. In: *Compositio Mathematica* 41.2 (1980), pp. 245–285. URL: http://eudml.org/doc/89458.

[19] Philip E. Protter. *Stochastic Integration and Differential Equations*. Springer Berlin, Heidelberg, 2005. ISBN: 978-3-662-10061-5.

[20] Daniel Revuz and Marc Yor. *Continuous Martingales and Brownian Motion*. Springer Berlin, Heidelberg, 1999.

[21] Anna Scampicchio and Melanie N. Zeilinger. *On the role of the signature transform in nonlinear systems and data-driven control*. 2025. arXiv: 2409.05685 [eess.SY]. URL: https://arxiv.org/abs/2409.05685.

[22] Ilya Chevyrev and Andrey Kormilitzin. *A Primer on the Signature Method in Machine Learning*. 2025. arXiv: 1603.03788 [stat.ML]. URL: https://arxiv.org/abs/1603.03788.

[23] R. Kent Nagle and Edward B. Saff. *Fundamentals of Differential Equations and Boundary Value Problems*. 8th ed. Greg Tobin, 2012. Chap. 13.

[24] Stephan Sturm. *Path Signatures for Feature Extraction. An Introduction to the Mathematics Underpinning an Efficient Machine Learning Technique*. 2025. arXiv: 2506.01815 [cs.LG]. URL: https://arxiv.org/abs/2506.01815.

[25] Jeremy Reizenstein and Benjamin Graham. *The iisignature library: efficient calculation of iterated-integral signatures and log signatures*. 2018. arXiv: 1802.08252 [cs.DS]. URL: https://arxiv.org/abs/1802.08252.

[26] Patrick Kidger and Terry Lyons. *Signatory: differentiable computations of the signature and logsignature transforms, on both CPU and GPU*. 2021. arXiv: 2001.00706 [cs.LG]. URL: https://arxiv.org/abs/2001.00706.

[27] Horatio Boedihardjo et al. “The signature of a rough path: Uniqueness”. In: *Advances in Mathematics* 293 (2016), pp. 720–737. ISSN: 0001-8708. DOI: https://doi.org/10.1016/j.aim.2016.02.011. URL: https://www.sciencedirect.com/science/article/pii/S0001870816301104.

[28] Yue Wu et al. *Signature features with the visibility transformation*. 2020. arXiv: 2004.04006 [cs.LG]. URL: https://arxiv.org/abs/2004.04006.

[29] Terry Lyons and Harald Oberhauser. *Sketching the order of events*. 2017. arXiv: 1708.09708 [stat.ML]. URL: https://arxiv.org/abs/1708.09708.

[30] Shujian Liao et al. *Learning stochastic differential equations using RNN with log signature features*. 2019. arXiv: 1908.08286 [cs.LG]. URL: https://arxiv.org/abs/1908.08286.

[31] Guy Flint, Ben Hambly, and Terry Lyons. “Discretely sampled signals and the rough Hoff process”. In: *Stochastic Processes and their Applications* 126.9 (Sept. 2016), pp. 2593–2614. ISSN: 0304-4149. DOI: 10.1016/j.spa.2016.02.011. URL: http://dx.doi.org/10.1016/j.spa.2016.02.011.

[32] Hao Ni. “Expected Signature of a Stochastic Process”. PhD thesis. University of Oxford, 2012.

[33] Ilya Chevyrev and Harald Oberhauser. *Signature moments to characterize laws of stochastic processes*. 2022. arXiv: 1810.10971 [math.ST]. URL: https://arxiv.org/abs/1810.10971.

[34] Adeline Fermanian. “Learning Time-Dependent Data with the Signature Transform”. PhD thesis. École Polytechnique, 2021.

[35] Kiyoshi Itô. “On a Stochastic Integral Equation”. In: *Proceedings of the Japan Academy* 22.2 (1946), pp. 32–35. DOI: 10.3792/pja/1195572371. URL: https://doi.org/10.3792/pja/1195572371.

[36] G. E. Uhlenbeck and L. S. Ornstein. “On the Theory of the Brownian Motion”. In: *Phys. Rev.* 36 (5 Sept. 1930), pp. 823–841. DOI: 10.1103/PhysRev.36.823. URL: https://link.aps.org/doi/10.1103/PhysRev.36.823.

[37] Mahir Lokvancic. *Pricing Path-Dependent Payoff Using Quadrature on Wiener Space*. https://ssrn.com/abstract=3503410. SSRN working paper, DOI: http://dx.doi.org/10.2139/ssrn.3503410. 2019.

[38] Kiseop Lee, Seongje Lim, and Hyungbin Park. *Option pricing under path-dependent stock models*. 2023. arXiv: 2211.10953 [math.PR]. URL: https://arxiv.org/abs/2211.10953.

[39] Terry Lyons, Sina Nejad, and Imanol Perez Arribas. *Non-parametric pricing and hedging of exotic derivatives*. 2019. arXiv: 1905.00711 [q-fin.MF]. URL: https://arxiv.org/abs/1905.00711.

[40] Darrell Duffie. *Dynamic Asset Pricing Theory*. Princeton University Press, 2001.

[41] Arthur E. Hoerl and Robert W. Kennard. "Ridge Regression: Biased Estimation for Nonorthogonal Problems". In: *Technometrics* 12.1 (1970), pp. 55–67. ISSN: 00401706.

[42] Robert Tibshirani. "Regression Shrinkage and Selection via the Lasso". In: *Journal of the Royal Statistical Society. Series B (Methodological)* 58.1 (1996), pp. 267–288. ISSN: 00359246.

[43] Hui Zou and Trevor Hastie. "Regularization and Variable Selection Via the Elastic Net". In: *Journal of the Royal Statistical Society Series B: Statistical Methodology* 67.2 (Mar. 2005), pp. 301–320. ISSN: 1369-7412. DOI: 10.1111/j.1467-9868.2005.00503.x. eprint: https://academic.oup.com/jrsssb/article-pdf/67/2/301/49795094/jrsssb_67_2_301.pdf. URL: https://doi.org/10.1111/j.1467-9868.2005.00503.x.

[44] Cornelius W. Osterlee and Lech A. Grzelak. *Mathematical Modeling And Computation In Finance: With Exercises And Python And Matlab Computer Codes*. World Scientific Publishing Europe Ltd., 2020. ISBN: 978-1-78634-794-7.

A

Appendix

A.1. Computation of the Expected Signature of the Time-Augmented One-Dimensional GBM

This appendix provides the step-by-step computations of the components of the expected signature associated with the time-augmented one-dimensional Geometric Brownian Motion discussed in Section 3.5.1.

Case 1: $I = (1, I')$.

$$\begin{aligned}\pi^{(1,I')}\left(f_n(t,x)\right) &= \pi^{(1,I')}(e_1 + rxe_2) \otimes \Phi_{n-1}(t,x) + \\ &\quad + \sigma^2 x^2 e_2 \otimes \partial_x \Phi_{n-1}(t,x) + \tfrac{1}{2}\sigma^2 x^2 (e_2 \otimes e_2) \otimes \Phi_{n-2}(t,x) \\ &= \pi^{(I')}\left(\Phi_{n-1}(t, X_0)\right) \\ &\overset{(3.8)}{=} x^{\alpha(I')} \pi^{I'}\left(F(t)\right)\end{aligned} \tag{A.1}$$

Applying $\pi^{(I)}$ to both sides of the Feyman-Kac relation: (**??**):

$$\begin{aligned}\pi^{(1,I')}\left(\Phi_n(t, X_0)\right) &= \pi^{(1,I')}\left(\int_0^t \mathbb{E}\left[f_n(t-s, X_s)\right] ds\right) \\ &\overset{\text{lin.}}{=} \int_0^t \mathbb{E}\left[\pi^{(1,I')}\left(f_n(t-s, X_s)\right)\right] ds \\ &\overset{\text{A.1}}{=} \int_0^t \mathbb{E}\left[X_s^{\alpha(I')} \pi^{(I')}\left(F(t-s)\right)\right] ds \\ &= X_0^{\alpha(I')} \int_0^t E_{\alpha(I')} \pi^{(I')}\left(F(t-s)\right) ds \\ &= X_0^{\alpha(I)} \int_0^t E_{\alpha(I')} \pi^{(I')}\left(F(t-s)\right) ds,\end{aligned}$$

where the last equivalence holds because $\alpha(I') = \alpha((1, I'))$, according to the definition of $\alpha(\cdot)$.

Case 2: $I = (2, 1, I'')$

$$\begin{aligned}
\pi^{(2,1,I'')}\left(f_n(t,x)\right) &= \pi^{(2,1,I'')}(e_1 + rxe_2) \otimes \Phi_{n-1}(t,x) + \\
&\quad + \sigma^2 x^2 e_2 \otimes \partial_x \Phi_{n-1}(t,x) + \tfrac{1}{2}\sigma^2 x^2 (e_2 \otimes e_2) \otimes \Phi_{n-2}(t,x)) \\
&= rx\pi^{(1,I'')}\left(\Phi_{n-1}(t,x)\right) + \sigma^2 x^2 \left(\partial_x \pi^{(1,I'')}\left(\Phi_{n-1}(t,x)\right)\right) \\
&\overset{(3.8)}{=} \left(rx + \sigma^2 x^2 \partial_x\right) \left(x^{\alpha((1,I''))}\pi^{(1,I'')}\left(F(t)\right)\right) \\
&= \left(r + \sigma^2 \alpha((1,I''))\right) x^{\alpha((2,1,I''))}\pi^{(1,I'')}\left(F(t)\right).
\end{aligned} \tag{A.2}$$

As done for Case 1, using Feyman-Kac:

$$\begin{aligned}
\pi^{(2,1,I'')}\left(\Phi_n(t,X_0)\right) &= \pi^{(2,1,I'')}\left(\int_0^t \mathbb{E}\left[f_n(t-s,X_s)\right] ds\right) \\
&\overset{\text{lin.}}{=} \int_0^t \mathbb{E}\left[\pi^{(2,1,I'')}\left(f_n(t-s,X_s)\right)\right] ds \\
&\overset{A.2}{=} \int_0^t \mathbb{E}\left[\left(r + \sigma^2\alpha((1,I''))\right) X_s^{\alpha((2,1,I''))}\pi^{(1,I'')}\left(F(t-s)\right)\right] ds \\
&= X_0^{\alpha(2,1,I'')}\left(r + \sigma^2\alpha((1,I''))\right) \int_0^t E_{\alpha((2,1,I''))}(s)\, \pi^{(1,I'')}\left(F(t-s)\right)\, ds.
\end{aligned}$$

Case 3 : $I = (2, 2, I'')$

$$\begin{aligned}
\pi^{(2,2,I'')}\left(f_n(t,x)\right) &= rx\pi^{(2,I'')}\left(\Phi_{n-1}(t,x)\right) + \sigma^2 x^2 \left(\partial_x \pi^{(2,I'')}\left(\Phi_{n-1}(t,x)\right)\right) + \\
&\quad + \tfrac{1}{2}\sigma^2 x^2 \pi^{I''}\left(\Phi_{n-2}(t,x)\right) \\
&\overset{(3.8)}{=} \left(rx + \sigma^2 x^2 \partial_x\right) \cdot \left(x^{\alpha((2,I''))}\pi^{(2,I'')}\left(F(t)\right)\right) + \tfrac{1}{2}\sigma^2 x^2 x^{\alpha(I'')}\pi^{(I'')}\left(F(t)\right) \\
&= x^{\alpha((2,2,I''))}\left(\left(r + \sigma^2\alpha((2,I''))\right)\, \pi^{(2,I'')}\left(F(t)\right) + \tfrac{1}{2}\sigma^2\pi^{(I'')}(F(t))\right).
\end{aligned} \tag{A.3}$$

Then:

$$\begin{aligned}
\pi^{(2,2,I'')}\left(\Phi_n(t,X_0)\right) &= \pi^{(2,2,I'')}\left(\int_0^t \mathbb{E}\left[f_n(t-s,X_s)\right]\, ds\right) \\
&\overset{\text{lin.}}{=} \int_0^t \mathbb{E}\left[\pi^{(2,2,I'')}\left(f_n(t-s,X_s)\right)\right]\, ds \\
&\overset{A.3}{=} \int_0^t \mathbb{E}\left[X_s^{\alpha((2,2,I''))}\left(\left(r + \sigma^2\alpha((2,I''))\right)\pi^{(2,I'')}(F(t-s)) + \right.\right. \\
&\quad \left.\left. + \tfrac{1}{2}\sigma^2\pi^{(I'')}(F(t-s))\right)\right]\, ds \\
&= X_0^{\alpha((2,2,I''))}\,(\int_0^t \left(r + \sigma^2\alpha((2,I''))\right) E_{\alpha((2,2,I''))}(s)\, \pi^{(2,I'')}(F(t-s))\, ds + \\
&\quad + \int_0^t \tfrac{1}{2}\sigma^2 E_{\alpha((2,2,I''))}(s)\, \pi^{(I'')}(F(t-s))\, ds\ )\,.
\end{aligned}$$