UNIVERSITAT DE BARCELONA

ADVANCED MATHEMATICS
MASTER'S FINAL PROJECT

Rough Volatility Models using the Signature Transform: Theory and Calibration

---

Author:
Pere Díaz Lozano

Supervisor:
Dr. Josep Vives Santa-Eulàlia

Facultat de Matemàtiques i Informàtica

June 28, 2023

UNIVERSITAT DE BARCELONA

Abstract

Facultat de Matemàtiques i Informàtica

MSc

Rough Volatility Models using the Signature Transform: Theory and Calibration

by Pere DÍAZ LOZANO

In this thesis we study a general stochastic volatility model where the dynamics of the volatility process are described by using the signature transform, a key object in rough path theory which is also very popular in the machine learning community due to its fundamental properties in approximation theory.

More specifically, we will present a general model for the evolution of the price of the underlying asset where the dynamics of the volatility are described by linear functions of the (time extended) signature of a primary underlying process. We will finally use this model in practice, showing how it can be efficiently calibrated to market prices of options by a Monte Carlo simulation.

List of symbols and notation

$(\Omega, \mathcal{F}, \mathbb{P})$	a probability space
$\mathbb{T}$	an index set representing time, usually taken to be $[0, T]$ or $[0, \infty)$
$\mathbb{F} = (\mathcal{F}_t)_{t \geq 0}$	a filtration in Ω
$s \wedge t$	minimum between s and t
$S = (S_t)_{t \geq 0}$	price process of the underlying
$\mathbb{P}^*$	risk neutral probability measure
V	a d-dimensional real vector space
$T((V))$	extended tensor algebra over V
$T(V)$	tensor algebra over V
$T^N(V)$	truncated tensor algebra of order N over V
π_n	canonical projection $T((V)) \to V^{\otimes n}$
$\pi_{\leq N}$	canonical projection $T((V)) \to T^N(V)$
$\sqcup\!\sqcup$	shuffle product
$S(\mathbf{X})$	signature of $\mathbf{X}$
$S(\mathbf{X})^n$	n-th term of the signature of $\mathbf{X}$
$S(\mathbf{X})^{\leq N}$	truncated signature of order N of $\mathbf{X}$
$G(V)$	space of group-like elements
$G^N(V)$	projection of $G(V)$ to $T^N(V)$
$\lfloor p \rfloor$	integer part of p
$\Omega_T^p(V)$	space of p-rough paths
$G\Omega_T^p(V)$	space of geometric p-rough paths
$\mathcal{W}G\Omega_T^p(V)$	space of weakly geometric p-rough paths
$\mathcal{W}G\widehat{\Omega}_T^p(V)$	space of time augmented weakly geometric p-rough paths
$\Lambda_T^p(V)$	space of weakly geometric stopped p-rough paths

Contents

Chapter 1

Introduction

Mathematical finance could be defined as the area of mathematics focused on the quantitative behavior of the financial markets. One of its central problems is concerned with the valuation of a type of financial instruments called *option contracts*. In a nutshell, these are legal agreements that promise some payment in the future, depending on the behavior of the price of another asset, which is usually called the *underlying*.

The main assumption when pricing options is that there is no arbitrage in the market, which essentially means that there are no opportunities for investors to make risk-free profits without using their own money or taking any risks.

One of the most celebrated results in mathematical finance is the *First Fundamental Theorem of Asset Pricing*, which has the following important consequence: if there exists a probability measure that is equivalent to the real one under which the tradable assets are local martingales, then the market is free of arbitrage. It can then be proved that the price of an option at the initial time is given by the expectation of the final payoff under this new measure.

For example, it is very common to assume that, under this equivalent probability measure, the price of the underlying $S = (S_t)_{t \geq 0}$ satisfies

$$dS_t = S_t \sigma_t dB_t,$$

where $(B_t)_{t \geq 0}$ is a Brownian motion. The integration is in the Itô sense, which results in S being a local martingale. The process $\sigma = (\sigma_t)_{t \geq 0}$ is called the volatility process, and it is the only unknown in the previous equation. Depending on the choice of σ, we get a different market model for S.

In practice, one usually approximates the dynamics of the volatility process by a parameterized family of models. For example, the Black-Scholes model assumes σ to be constant. Other models assume that it satisfies some parameterized SDE, like the Heston model.

In classical quantitative finance, all these parameters have an economic meaning. However, in the past few years Machine Learning has successfully entered the world of stochastic modelling and mathematical finance. The paradigm of calibrating a few well interpretable parameters has changed to learning the model's characteristics as a whole, providing more robust and data-driven models that match the observed market prices better.

1.1 Contributions

In this thesis we follow the work of [Cuc+23], where the volatility process is modeled by a linear functional of the signature of a multidimensional continuous semimartingale. In our case, we extend these models by only requiring the underlying process to have sample paths that are almost surely weakly geometric rough paths. This way we are able to incorporate models where the driving noise is much rougher than semimartingales, like for example a one dimensional fractional Brownian motion.

We will see how the universal approximation properties of the signature transform makes the model mathematically well-founded. Moreover, the linearity of the volatility model allows for a very fast and computationally cheap calibration of the parameters to match market option prices.

Finally, we will test how these models perform in practice, by calibrating them to option prices generated from classical stochastic volatility models and also from real market data, concluding that in all cases we get a very good fit.

1.2 Thesis outline

Chapter 2 briefly introduces the basics of stochastic processes, focusing on Brownian motion and fractional Brownian motion.

Chapter 3 is an introduction to the field of mathematical finance, and more specifically, option pricing theory. We will derive the Black-Scholes formula, and also discuss the related concept of implied volatility. We will end with a brief introduction to several kinds of models that are used in classical mathematical finance, like stochastic volatility models and rough volatility models.

Chapter 4 rigorously defines the signature of a path and shows its most important properties.

In Chapter 5 we present the signature based model for the volatility and show how it can be used to price options in a fast and efficient way. We end with the results of the calibration to three different datasets of observed option prices.

Finally, Chapter A in the appendix gives an overview of the fundamental results in stochastic analysis that are used throughout the thesis.

Chapter 2

Stochastic processes

In this first chapter we introduce the basic concepts of the theory of stochastic processes, focusing on Brownian motion and fractional Brownian motion.

2.1 Stochastic process

Along this chapter, $(\Omega, \mathcal{F}, \mathbb{P})$ denotes a probability space, where Ω is a set denoting the space of outcomes of a random experience, $\mathcal{F}$ is a σ-algebra of parts of Ω and $\mathbb{P}$ is a probability on $\mathcal{F}$.

Let us start by giving the formal definition of a stochastic process, which in a nutshell gathers the evolution of random phenomena whose values belong to a set S, called the state space.

Definition 2.1 (Stochastic process). *A stochastic process with state space S is a family $X = (X_t)_{t\in\mathbb{T}}$ of random variables $X_t : \Omega \to S$ indexed by a set $\mathbb{T}$.*

In this chapter we will always choose S to be $\mathbb{R}$, and we will equip it with the Borel σ-algebra $\mathcal{B}(\mathbb{R})$. However, we remark that everything that we will introduce can be easily extended to the multidimensional case.

The set $\mathbb{T}$ is thought of as a set of times, which in our case will usually be $\mathbb{T} = [0, T]$ with $T < \infty$, or $\mathbb{T} = [0, \infty)$.

In the context of modelling, the observed values of a given process are the relevant data. As with samples in statistics, these observations correspond to fixed values of $\omega \in \Omega$. This motivates the following definition.

Definition 2.2 (Sample paths). *The sample paths (or trajectories) of a stochastic process $X = (X_t)_{t\in\mathbb{T}}$ are the family of functions indexed by $\omega \in \Omega$, $X(\omega) : \mathbb{T} \to \mathbb{R}$, defined by $X(\omega)(t) = X_t(\omega)$.*

Definition 2.3 (Continuous process). *We say that a stochastic process is continuous if its sample paths are $\mathbb{P}$-a.s. continuous.*

Along this section, unless stated otherwise, we will write $X = (X_t)_{t\in\mathbb{T}}$ and $Y = (Y_t)_{t\in\mathbb{T}}$ to denote two stochastic processes. Let us proceed with two basic definitions.

Definition 2.4 (Modifications). *If $\mathbb{P}(X_t = Y_t) = 1$ for all $t \in \mathbb{T}$, we say that X and Y are modifications of each other.*

Definition 2.5 (Indistinguishable). *If $\mathbb{P}(\omega \in \Omega | X_t(\omega) = Y_t(\omega), \forall t \in \mathbb{T}) = 1$, we say that X and Y are indistinguishable. In other words, they have the same sample paths with probability one.*

We clearly have that, if X and Y are indistinguishable, then they are also modifications of each other. However, the converse statement is not true.

For example, take $\mathbb{T} = \mathbb{R}$, $X_t = 0$ and $Y_t = \mathbf{1}_{\{Z=t\}}$, with $Z \sim \mathcal{N}(0,1)$ being defined on the same probability space. Then, we have that $\mathbb{P}(X_t = Y_t) = \mathbb{P}(Z \neq t) = 1$ for all $t \in \mathbb{R}$, but since all the sample paths of Y have exactly one point which is different from zero, we clearly have that

$$\mathbb{P}(\omega \in \Omega | X_t(\omega) = Y_t(\omega), \forall t \in \mathbb{T}) = \mathbb{P}(\omega \in \Omega | Y_t(\omega) = 0, \forall t \in \mathbb{T}) = 0.$$

The problem here is the uncountability of the real numbers used for the time index. In fact, by the countable additivity of measures, if X and Y are modifications of each other, we then have that the sample paths of X and Y are almost surely identical on any given countable subset of times. This in particular means that

$$\mathbb{P}(\omega \in \Omega | X_t(\omega) = Y_t(\omega), \forall t \in \mathbb{T} \cap \mathbb{Q}) = 1.$$

By the density of $\mathbb{Q} \cap \mathbb{T}$ in $\mathbb{T}$, we deduce that the converse is true whenever the sample paths of X and Y are almost surely continuous. In other words, if two modifications of the same process are continuous, then they must be indistinguishable.

The probabilistic features of stochastic process are gathered by the joint distributions of their variables, as given in the next definition.

Definition 2.6 (Law of a stochastic process). *We say that X and Y have the same law, and we write $X \overset{law}{=} Y$, to indicate that $(X_{t_1}, \ldots, X_{t_m})$ and $(Y_{t_1}, \ldots, Y_{t_m})$ have the same law for all $m \geq 1$ and all $t_1, \ldots, t_m \in \mathbb{T}$, i.e.*

$$\mathbb{P}((X_{t_1}, \ldots, X_{t_m}) \in A) = \mathbb{P}((Y_{t_1}, \ldots, Y_{t_m}) \in A)$$

for all $A \in \mathcal{B}(\mathbb{R}^m)$.

Remark 2.1. *It is worthwhile noting that if X is a modification of Y, then they necessarily have the same law. Indeed, from Definition 2.4 we deduce that $\mathbb{P}(X_1 = Y_{t_1}, \ldots, X_{t_m} = Y_{t_m}) = 1$ for all $m \geq 1$ and all $t_1, \ldots, t_m \in \mathbb{T}$.*

2.2 Filtrations and martingales

Next we define what a filtration is, as well as the notion of a stochastic process being adapted to it.

Definition 2.7 (Filtration, adapted process). *A filtration $\mathbb{F} = (\mathcal{F}_t)_{t \in \mathbb{T}}$ is an increasing family of σ-algebras which are included in the σ-algebra $\mathcal{F}$. A process $X = (X_t)_{t \in \mathbb{T}}$ is said to be adapted to the filtration $\mathbb{F}$ if X_t is $\mathcal{F}_t$-measurable for each $t \in \mathbb{T}$. We will also write that X is $\mathbb{F}$-adapted.*

Intuitively, one might think of $\mathcal{F}_t$ as the collection of all events that are observable up to time t. Then, the fact that a stochastic process $(X_t)_{t \in \mathbb{T}}$ is adapted to $\mathbb{F}$ can be thought of as X_t only depending on what it is observable up to time t.

Remark 2.2. *We will always work with filtrations where if $A \in \mathcal{F}$ with $\mathbb{P}(A) = 0$, then $A \in \mathcal{F}_t$ for all $t \in \mathbb{T}$. That is, $\mathcal{F}_0$ contains all the $\mathbb{P}$-null sets of $\mathcal{F}$.*

Let us now define the natural filtration associated to a stochastic process. Intuitively, this records its past behaviour at each time. It is in a sense the simplest filtration such that the given process is adapted.

Definition 2.8 (Natural filtration)**.** *Given a process $X = (X_t)_{t\in\mathbb{T}}$, we can define its natural filtration $\mathbb{F} = (\mathcal{F}_t)_{t\in\mathbb{T}}$ as $\mathcal{F}_t := \sigma(\{X_s, s \leq t\})$. In other words, $\mathcal{F}_t$ is the smallest σ-algebra such that $\mathbb{F}$ is a filtration and X is adapted to it.*

In general, the natural filtration does not satisfy Remark 2.2. For that reason, whenever we talk about the natural filtration, we will always consider its completion, that is, the filtration in which all the $\mathcal{F}$-negligible sets are $\mathcal{F}_0$-measurable. We will refer to this as the natural filtration associated to the process X, without making explicit that it is the completion.

Definition 2.9 ((Sub, Super)Martingale)**.** *Let $X = (X_t)_{t\in\mathbb{T}}$ be a process adapted to a filtration $\mathbb{F} = (\mathcal{F}_t)_{t\in\mathbb{T}}$ and such that $\mathbb{E}[|X_t|] < \infty$ for all $t \in \mathbb{T}$. Then, we say that*

- *X is an martingale w.r.t $\mathbb{F}$ if $\mathbb{E}[X_t|\mathcal{F}_s] = X_s$ for all $s < t$,*
- *X is a submartingale w.r.t $\mathbb{F}$ if $\mathbb{E}[X_t|\mathcal{F}_s] \geq X_s$ for all $s < t$,*
- *X is a supermartingale w.r.t $\mathbb{F}$ if $\mathbb{E}[X_t|\mathcal{F}_s] \leq X_s$ for all $s < t$.*

In order to indicate the filtration we are referring to, it is also common to write that X is an $\mathbb{F}$-(sub, super)martingale. Whenever it is clear to what filtration we are referring to, we will simply say that X is a (sub, super)martingale.

Let us try to explain in an intuitive way what the martingale condition is telling us. Assume that $X = (X_t)_{t\in\mathbb{T}}$ is a stochastic process where X_t indicates the amount of capital that the player of some random game has at time t. Then, saying that X is a martingale tells us that, having the information of the game up to time s (that is, $\mathcal{F}_s$), in average, we expect the player to keep the capital unchanged. Hence, the term martingale refers in some sense to a *fair game*. For a supermartingale, in average, the capital is expected to decrease (this favours the Casino), while for a submartingale, it is expected to increase (favoring the player).

The following result is useful whenever we need to work with filtrations that are larger than the natural ones. The proof is a simple consequence of the tower property of conditional expectations.

Proposition 2.1. *Let $X = (X_t)_{t\in\mathbb{T}}$ be a martingale with respect to a filtration $\mathbb{F} = (\mathcal{F}_t)_{t\in\mathbb{T}}$. Let $\mathbb{G} = (\mathcal{G}_t)_{t\in\mathbb{T}}$ be another filtration such that $\mathcal{G}_t \subset \mathcal{F}_t$ for all $t \in \mathbb{T}$, and assume that X is also $\mathbb{G}$-adapted. Then X is also a martingale with respect to $\mathbb{G}$.*

We now proceed to define a generalization of the notion of a martingale, namely a local martingale, which will be crucial in the following chapters. Before doing so, we need to define what a stopping time is.

Definition 2.10 (Stopping time)**.** *Let $\mathbb{F} = (\mathcal{F}_t)_{t\in\mathbb{T}}$ be a filtration. An $\mathbb{F}$-stopping time is a random variable $\tau : \Omega \to \mathbb{T}$ such that $\{\tau \leq t\} \in \mathcal{F}_t$ for all $t \in \mathbb{T}$.*

We will establish slightly different definitions of a local martingale depending on the set $\mathbb{T}$. For convenience, let us write $T = \infty$ if $\mathbb{T} = [0, \infty)$.

Definition 2.11 (Local martingale)**.** *Let $X = (X_t)_{t\in\mathbb{T}}$ be a process adapted to a filtration $\mathbb{F}$. Then, we say that X is a local martingale w.r.t. $\mathbb{F}$ if there exists a sequence of $\mathbb{F}$-stopping times $\{\tau\}_{n\geq 0}$ such that*

1. *The sequence* $\{\tau\}_{n\geq 0}$ *is almost surely increasing:* $\mathbb{P}(\tau_n < \tau_{n+1}) = 1$ *for all* $n \geq 0$.
2. $\mathbb{P}(\lim_{n\to\infty} \tau_n = T) = 1$.
3. *The stopped process* $(X_{t\wedge\tau_n})_{t\in\mathbb{T}}$ *is an* $\mathbb{F}$*-martingale for each* $n \geq 0$.

2.3 Gaussian processes

The Gaussian distribution is arguably the most important distribution in probability theory. Similarly, Gaussian processes also play a central role in the theory of stochastic processes.

Definition 2.12 (Gaussian process). *A stochastic process* $X = (X_t)_{t\in\mathbb{T}}$ *is said to be Gaussian if, for any* $m \geq 1$ *and any* $t_1, \ldots, t_m \in \mathbb{T}$, $(X_1, \ldots, X_{t_m})$ *is a Gaussian random vector.*

The mean of X *is then defined as the function* $m : \mathbb{T} \to \mathbb{R}$ *given by* $m(t) := \mathbb{E}[X_t]$, *while the covariance of* X *is the function* $\Gamma : \mathbb{T}^2 \to \mathbb{R}$ *given by* $\Gamma(s,t) := Cov(X_s, X_t)$. *When* $m \equiv 0$, *we say that* X *is centered.*

In terms of uniqueness, just like for Gaussian random vectors we have the following result.

Proposition 2.2. *Two Gaussian processes have the same law if and only if they have the same mean and the same covariance.*

Proof. See, e.g., [Kal97, Lemma 11.1]. □

The canonical way of proving the existence of a centered Gaussian stochastic process with covariance function Γ is to use Kolmogorov's theorem.

Theorem 2.1 (Kolmogorov). *Consider a symmetric function* $\Gamma : \mathbb{T}^2 \to \mathbb{R}$. *Then, there exists a centered Gaussian process having* Γ *as a covariance function if and only if*

$$\sum_{k,l=1}^{m} a_k a_l \Gamma(t_k, t_l) \geq 0$$

for all $m \geq 1$, $t_1, \ldots, t_m \in \mathbb{T}$ *and* $a_1, \ldots, a_m \in \mathbb{R}$. *In this case, it is said that* Γ *is of positive type.*

Proof. See, e.g., [Dud02, Theorem 12.1.3]. □

Let us end this subsection by stating the Kolmogorov-Chentsov continuity criterion, that gives us information about the continuity and the Hölder regularity of the sample paths of a stochastic process.

Lemma 2.1 (Kolmogorov-Chentsov). *Let* $X = (X_t)_{t\in\mathbb{T}}$ *be a stochastic process satisfying the following property: for any compact interval* $I \subset \mathbb{T}$, *there exists positive real numbers* $\alpha \geq 1$, β *and* C *such that*

$$\mathbb{E}[|X_t - X_s|^\alpha] \leq C|t-s|^{1+\beta}$$

for any $s, t \in I$. *Then, the process* X *has a continuous modification* $\tilde{X}$. *Moreover, for any* $\gamma \in (0, \beta/\alpha)$, *the process* $\tilde{X}$ *has sample paths that are almost surely locally* γ*-Hölder* *continuous, i.e. for any compact interval $I \subset \mathbb{T}$*

$$\sup_{s,t \in I} \frac{|\tilde{X}_t(\omega) - \tilde{X}_s(\omega)|}{|t-s|^\gamma} < \infty$$

for almost all $\omega \in \Omega$.

Proof. See, e.g., [Kal97, Theorem 2.23]. □

2.4 Brownian motion

The most well-known and used process in stochastic analysis is the standard Brownian motion.

Definition 2.13 (Standard Brownian motion)**.** *A one-dimensional standard Brownian motion $B = (B_t)_{t\in\mathbb{T}}$ is a centered Gaussian process with covariance function given by $\mathbb{E}[B_t B_s] = s \wedge t$.*

The existence of such process is ensured by Kolmogorov's Theorem 2.1. Indeed, all we need to do is check that the symmetric function given by $\Gamma(s,t) = s \wedge t$ is of positive type.

By using the identity given by

$$s \wedge t = \int_0^\infty \mathbf{1}_{[0,s]}(r)\mathbf{1}_{[0,t]}(r)dr,$$

we have that, for all $m \geq 1$, $t_1, \dots, t_m \in \mathbb{T}$ and $a_1, \dots, a_m \in \mathbb{R}$,

$$\begin{aligned}\sum_{i,j=1}^m a_i a_j (t_i \wedge t_j) &= \sum_{i,j=1}^m a_i a_j \int_0^\infty \mathbf{1}_{[0,t_i]}(r)\mathbf{1}_{[0,t_j]}(r)dr \\ &= \int_0^\infty \Big(\sum_{i=1}^m a_i \mathbf{1}_{[0,t_i]}(r)\Big)^2 dr \geq 0.\end{aligned}$$

As for the continuity and regularity of the sample paths, we have the following result.

Proposition 2.3. *There is a continuous modification of a standard Brownian motion such that its sample paths are almost surely locally γ-Hölder continuous for any $\gamma \in (0, \frac{1}{2})$.*

Proof. The result follows immediately from the Kolmogorov-Chentsov Lemma 2.1. Indeed, since the process is centered and Gaussian with covariance given by $\mathbb{E}[B_t B_s] = s \wedge t$, it is easy to check that $B_t - B_s \sim \mathcal{N}(0, |t-s|)$ for any $s, t \in \mathbb{T}$.

Therefore, the $2n$-order moments of $B_t - B_s$ can be computed explicitly, and are in fact given by

$$\mathbb{E}[|B_t - B_s|^{2n}] = C_n |t-s|^n, \quad C_n = \frac{(2n)!}{2^n n!}.$$

For any $n \in \mathbb{N}$, we can then apply the Kolmogorov-Chentsov Lemma 2.1 with $\alpha = 2n$, $C = C_n$ and $\beta = n-1$, obtaining that the sample paths are almost surely Hölder continuous for any $\gamma \in (0, \frac{n-1}{2n})$. By passing n to the limit, we get that this holds true for any $\gamma \in (0, \frac{1}{2})$. □

In order to avoid speaking about a continuous modification each time, in the sequel we will assume continuity of the Brownian motion itself.

The next proposition summarizes three important properties of Brownian motion.

Proposition 2.4. *A Brownian motion $(B_t)_{t\geq 0}$ has independent increments. That is, for any $m \geq 4$ and $0 \leq t_1 \leq \cdots \leq t_m$, the random variables*

$$B_{t_2} - B_{t_1}, \ldots, B_{t_m} - B_{t_{m-1}}$$

are mutually independent.

Proof. Since for a Gaussian random vector we have that pairwise independence implies mutual independence, we only need to prove that $B_{t_2} - B_{t_1}$ and $B_{t_4} - B_{t_3}$ are independent for any $0 \leq t_1 \leq t_2 \leq t_3 \leq t_4$. Since these random variables are normally distributed, all we need to show is that they are uncorrelated, i.e. $\mathbb{E}[(B_{t_2} - B_{t_1})(B_{t_4} - B_{t_3})] = 0$. This is straightforward to check. □

Let us end this section by showing that a standard Brownian motion is a martingale.

Proposition 2.5. *Let $B = (B_t)_{t\in\mathbb{T}}$ be a standard Brownian motion adapted to some filtration $\mathbb{F} = (\mathcal{F}_t)_{t\in\mathbb{T}}$. Then, we have that B is an $\mathbb{F}$-martingale.*

Proof. First of all, since $B_t \sim \mathcal{N}(0,t)$, we clearly have that $\mathbb{E}[|B_t|] < \infty$. Second, for any $s,t \in \mathbb{T}$ with $s < t$, we have that

$$\mathbb{E}[B_t|\mathcal{F}_s] - B_s = \mathbb{E}[B_t - B_s|\mathcal{F}_s] = \mathbb{E}[B_t - B_s] = 0,$$

where in the first equality we used the linearity of the conditional expectation and the fact that B_s is $\mathcal{F}_s$-measurable, and in the second equality we used the fact that $B_t - B_s$ is independent of $\mathcal{F}_s$, due to the independence of the increments. □

2.5 Fractional Brownian motion

In this section we will introduce a Gaussian process called fractional Brownian motion, which will be used in Chapter 5. This was first introduced in [MN68], and it is defined as follows.

Definition 2.14 (Fractional Brownian motion)**.** *A one-dimensional fractional Brownian motion (fBm) with Hurst parameter $H \in (0,1)$ is a centered Gaussian process $B^H = (B^H_t)_{t\in\mathbb{T}}$ with covariance function*

$$\mathbb{E}[B^H_t B^H_s] = \frac{1}{2}\Big(t^{2H} + s^{2H} - |t-s|^{2H}\Big). \tag{2.1}$$

Notice that by setting $H = \frac{1}{2}$ we recover the definition of the classical Brownian motion.

The most straightforward way to prove the existence of fBm is to show that the symmetric function $\Gamma(s,t) = \frac{1}{2}\Big(t^{2H} + s^{2H} - |t-s|^{2H}\Big)$ is of positive type.

Proposition 2.6. *The function given by (2.1) is symmetric and of positive type. Therefore, there exists a centered Gaussian process $B^H = (B^H_t)_{t\in\mathbb{T}}$ with covariance function given by (2.1).*

Proof. See, e.g., [Nou, Proposition 1.6]. □

The following proposition tells us that there is a continuous modification of B^H. Moreover, the samples paths of a fBm have a regularity that decreases with H.

Proposition 2.7. *There is a continuous modification of a fBm with Hurst parameter H such that its sample paths are almost surely locally γ-Hölder continuous for any $\gamma \in (0, H)$.*

Proof. Just like in the Brownian motion case, the result follows immediately from the Kolmogorov-Chentsov Lemma 2.1. Indeed, one can easily deduce from the covariance function of $B^H = (B_t^H)_{t\in\mathbb{T}}$ that $B_t^H - B_s^H \sim \mathcal{N}(0, |t-s|^{2H})$ for any $0 \leq s < t$.

Therefore, the $2n$-order moments of $B_t^H - B_s^H$ are given by

$$\mathbb{E}[|B_t^H - B_s^H|^{2n}] = C_n|t-s|^{2Hn}, \quad C_n = \frac{(2n)!}{2^n n!}.$$

For any $n \in \mathbb{N}$, we can then apply the Kolmogorov-Chentsov Lemma 2.1 with $\alpha = 2n$, $C = C_n$ and $\beta = 2Hn - 1$, obtaining that the sample paths are almost surely Hölder continuous for any $\gamma \in (0, \frac{2Hn-1}{2n})$. By passing n to the limit, we get that this holds true for any $\gamma \in (0, H)$. □

In order to avoid speaking about a continuous modification each time, in the sequel we will assume continuity of the fractional Brownian motion itself.

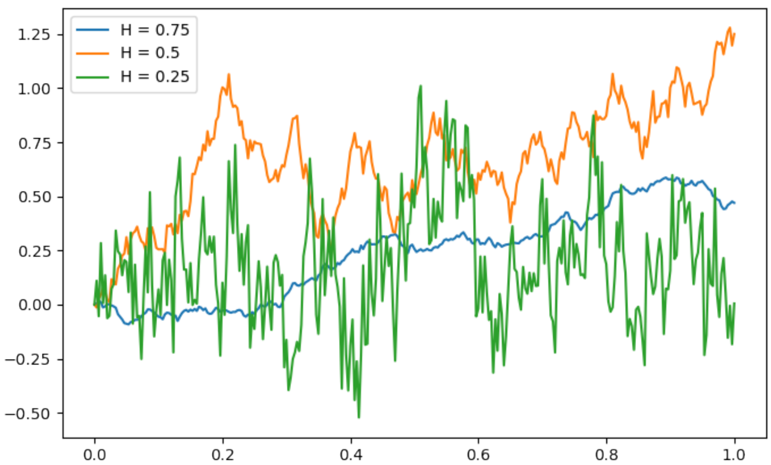

FIGURE 2.1: Sample paths of fBm with different Hurst parameters H.

2.5.1 Properties of the fractional Brownian motion

Next we will give some interesting properties of fractional Brownian motion.

Proposition 2.8. *Let $B^H = (B_t^H)_{t\geq 0}$ be a fractional Brownian motion with Hurst parameter $H \in (0,1)$. Then:*

1. *(Stationarity of increments) For any $h > 0$, $(B_{t+h}^H - B_h^H)_{t\geq 0} \overset{law}{=} (B_t^H)_{t\geq 0}$.*

2. *(H-self-similarity) For any $\lambda > 0$, we have that $(\lambda^{-H} B_{\lambda t}^H)_{t\geq 0} \overset{law}{=} (B_t^H)_{t\geq 0}$.*

Conversely, any continuous Gaussian process $B^H = (B_t^H)_{t\geq 0}$ *with* $B_0^H = 0$, $Var(B_1^H) = 1$ *and such that (1) and (2) hold is a fractional Brownian motion with Hurst parameter* H.

Proof. The two properties are proven in the same way. One first shows that the processes on the left hand side are centered, Gaussian and have (2.1) as covariance function. Then, Proposition 2.2 yields the result.

The converse is proven in a similar manner. See [Nou, Proposition 2.2] for the explicit proof. □

Proposition 2.9. *Let* $B^H = (B_t^H)_{t\in\mathbb{T}}$ *be a fBm with Hurst parameter* $H \in (0,1)$. *Let* $0 \leq s_1 < t_1 < s_2 < t_2$ *with* $s_i, t_i \in \mathbb{T}$, *so that the intervals* $[s_1, t_1]$ *and* $[s_2, t_2]$ *are disjoint.*

We then have that:

1. *If* $H \in (0, 1/2)$, *the increments of* B^H *are negatively correlated:*

$$\mathbb{E}[(B_{t_1}^H - B_{s_1}^H)(B_{t_2}^H - B_{s_2}^H)] < 0.$$

2. *If* $H \in (1/2, 1)$, *the increments of* B^H *are positively correlated:*

$$\mathbb{E}[(B_{t_1}^H - B_{s_1}^H)(B_{t_2}^H - B_{s_2}^H)] > 0.$$

Proof. From the covariance function of $B^H = (B_t^H)_{t\in\mathbb{T}}$, one obtains that

$$\mathbb{E}[(B_{t_1}^H - B_{s_1}^H)(B_{t_2}^H - B_{s_2}^H)] = \frac{1}{2}\Big(|t_2 - s_1|^{2H} + |s_2 - t_1|^{2H} - |t_2 - t_1|^{2H} - |s_2 - s_1|^{2H}\Big).$$

Set $a_1 = t_2 - s_1$, $a_2 = s_2 - t_1$, $b_1 = t_2 - t_1$, $b_2 = s_2 - s_1$, which trivially yields $a_1 + a_2 = b_1 + b_2$. Then, we can write the right hand side of the previous expression as

$$\frac{f(a_1) + f(a_2)}{2} - \frac{f(b_1) + f(b_2)}{2},$$

where $f(x) = x^{2H}$. If $H \in (0, 1/2)$, we can use the concavity of f to get

$$\frac{f(a_1) - f(a_2)}{2} - \frac{f(b_1) - f(b_2)}{2} < f\Big(\frac{a_1 + a_2}{2}\Big) - f\Big(\frac{b_1 + b_2}{2}\Big) = 0.$$

If $H \in (1/2, 1)$, we can use the convexity of f to get

$$\frac{f(a_1) - f(a_2)}{2} - \frac{f(b_1) - f(b_2)}{2} > f\Big(\frac{a_1 + a_2}{2}\Big) - f\Big(\frac{b_1 + b_2}{2}\Big) = 0.$$

□

Therefore, for $H \in (0, 1/2)$, the fBm has the property of counter-persistence: if it was increasing in the past, it is more likely to decrease in the future, and vice versa. In contrast, if $H \in (1/2, 1)$, the fBm is persistent, it is more likely to trend than to break it. Notice that if $H = 1/2$, i.e. if we are in the Brownian motion case, then the increments are uncorrelated.

Lemma 2.2. *If $H \in (1/2, 1)$, a fBm exhibits long-range dependence, in the sense that*

$$\sum_{k=0}^{\infty} Cov[B_1^H, B_k^H - B_{k-1}^H] = \infty.$$

If $H \in (0, 1/2)$, a fBm exhibits short-range dependence, in the sense that

$$\sum_{k=0}^{\infty} Cov[B_1^H, B_k^H - B_{k-1}^H] < \infty.$$

Proof. We have that

$$Cov[B_1^H, B_k^H - B_{k-1}^H] = \frac{1}{2}\left(k^{2H} - 2|k-1|^{2H} + |k-2|^{2H}\right). \tag{2.2}$$

The proof follows by showing that (2.2) is of order k^{2H-2} as $k \to \infty$. □

2.5.2 Semimartingale property

We will now see that a fBm with Hurst parameter $H \in (0, 1)$ is not a semimartingale unless $H = \frac{1}{2}$. The most important consequence of this result is the fact that one is unable to use the Itô integral to define integration against a fBm.

Theorem 2.2. *Let B^H be a fBm with Hurst parameter $H \neq \frac{1}{2}$. Then B^H is not a semimartingale.*

The proof essentially relies on the following result.

Proposition 2.10. *Let $t \in \mathbb{T}$ and $p \in [1, \infty)$. Then, in $L^2(\Omega)$ and as $n \to \infty$, one has*

$$t^{-pH} \sum_{k=1}^{n} |B_{tk/n}^H - B_{t(k-1)/n}^H|^p \to \begin{cases} 0 & \text{if } p > 1/H \\ \mathbb{E}[|G|^p] & \text{if } p = 1/H, \text{ with } G \sim \mathcal{N}(0,1) \\ +\infty & \text{if } p < 1/H \end{cases}$$

Proof. The proof is a consequence of [Nou, Theorem 2.1, Corollary 2.1], together with the H-self-similarity of fBm. □

Proof for Theorem 2.2. Let us only focus on the case where $H \in (0, 1/2)$, which is the one we will use in the subsequent chapters.

The proof relies on the fact that, if $S = (S_t)_{t\in\mathbb{T}}$ denotes a semimartingale, then for any $t \in \mathbb{T}$,

$$\sum_{k=1}^{n} (S_{tk/n} - S_{t(k-1)/n})^2 \to \langle S \rangle_t < \infty$$

in probability as $n \to \infty$, where $\langle S \rangle_t$ denotes the quadratic variation of S on $[0, t]$. Then, if $H < \frac{1}{2}$, Proposition 2.10 yields that $\sum_{k=1}^{n} (B_{tk/n}^H - B_{t(k-1)/n}^H)^2 \to \infty$ in probability, and so B^H is not a semimartingale.

See [Nou, Theorem 2.2] for the proof in the case where $H \in (1/2, 1)$. □

Another interesting consequence of Proposition 2.10 is given by the following property about the irregularity of the sample paths of a fBm.

Proposition 2.11. *For any $\gamma > H$, the sample paths of a fBm are almost surely nowhere locally Hölder continuous of order γ.*

Proof. Let $[0,T] \subset \mathbb{T}$. Suppose that for some $\gamma > H$,

$$|B_t^H - B_s^H| \le C|t-s|^\gamma \quad \mathbb{P}\text{-a.s.}$$

for all $s,t \in [0,T]$. Choose $p > 0$ such that $H < 1/p < \gamma$. Then,

$$\sum_{k=1}^{n} |B_{Tk/n}^H - B_{T(k-1)/n}^H|^p \le C^p T^{\gamma p} \left|\frac{1}{n}\right|^{p\gamma - 1} \quad \mathbb{P}\text{-a.s.}$$

where the left hand side diverges as a consequence of Proposition 2.10 and the right hand side converges to zero, arriving at a contradiction. This completes the proof. □

2.5.3 Stochastic representation

In this section we will show that a fractional Brownian motion can be represented as an integral against a classical Brownian motion. This will be useful later, as it will allow us to decompose a fBm into the sum of two separate processes, from which we will only keep the one with a simple expression that contains all the irregularity of a fBm, which is what we are interested in.

This representation of fBm that we will focus on is called *moving average representation*, and was first obtained in [MN68].

Before showing that, we will need some results on the Wiener integral of a deterministic function against a two-sided Brownian motion $B = (B_t)_{t\in\mathbb{R}}$, defined as

$$B_t := \begin{cases} B_t^1 & \text{if } t \ge 0 \\ B_{-t}^2 & \text{if t < 0,} \end{cases}$$

where B^1 and B^2 are two independent standard Brownian motions defined on $[0,\infty)$.

Proposition 2.12. *Let $B = (B_t)_{t\in\mathbb{R}}$ be a two-sided Brownian motion. Then, the family*

$$\left\{ \int_{\mathbb{R}} f(u)dB_u \ : \ f \in L^2(\mathbb{R}) \right\}$$

is a collection of Gaussian random variables characterized by the following:

$$\mathbb{E}\left[\int_{\mathbb{R}} f(u)dB_u \right] = 0,$$

$$\mathbb{E}\left[\int_{\mathbb{R}} f(u)dB_u \int_{\mathbb{R}} g(u)dB_u \right] = \int_{\mathbb{R}} f(u)g(u)du. \tag{2.3}$$

Proof. See [Nou, Section 5.1]. □

Proposition 2.13. *Let $H \in (0,1)$. Set*

$$c_H = \sqrt{\frac{1}{2H} + \int_0^\infty \left((1+u)^{H-\frac{1}{2}} - u^{H-\frac{1}{2}}\right)^2 du}$$

and let $B = (B_t)_{t\in\mathbb{R}}$ be a two-sided Brownian motion. Then, the process $B^H = (B_t^H)_{t\in\mathbb{T}}$ given by

$$B_t^H = \frac{1}{c_H}\left(\int_{-\infty}^0 \left((t-u)^{H-\frac{1}{2}} - (-u)^{H-\frac{1}{2}}\right) dB_u + \int_0^t (t-u)^{H-\frac{1}{2}} dB_u\right) \tag{2.4}$$

is a fractional Brownian motion with Hurst parameter H.

Proof. Let us first check that both Itô integrals in (2.4) are well-defined. In order to do so, we need to show that the integrands are square integrable in their corresponding integration regions (see Section A.1). For any $t \in \mathbb{T}$, let

$$f_t(u) := (t-u)^{H-\frac{1}{2}} - (-u)^{H-\frac{1}{2}}.$$

Since $2H - 1 > -1$, we have that $f_t(u)$ is square integrable at $u \to 0^-$. When $u \to -\infty$, one has that,

$$f_t(u) \sim \left(H - \frac{1}{2}\right)(-u)^{H-\frac{3}{2}}.$$

Since $2H - 3 < -1$, we have that $(-u)^{H-\frac{3}{2}}$ is square integrable at $-\infty$, and so is $f_t(u)$. Therefore, the first Itô integral is well defined. The same proof also applies to show that the constant c_H is finite. As for the second integral, since $2H - 1 > -1$, we clearly have that $(t-u)^{H-\frac{1}{2}}$ is square integrable in $[0,t]$.

Consequently, we have that B_t^H is well-defined for any $t \in \mathbb{T}$.

Let us now show that B^H is a fractional Brownian motion with Hurst parameter H. By Proposition 2.12, we clearly have that B^H is centered and Gaussian. Therefore, by Proposition 2.2, all we need to prove is that the covariance function of B^H is the same as in (2.1).

Let us rewrite (2.4) compactly as

$$B_t^H = \frac{1}{c_H}\int_{\mathbb{R}} \left((t-u)^{H-\frac{1}{2}}\mathbf{1}_{\{u<t\}} - (-u)^{H-\frac{1}{2}}\mathbf{1}_{\{u<0\}}\right) dB_u.$$

Fix $s, t \in \mathbb{T}$ with $s < t$. By using (2.3), we get that

$$\begin{aligned}
\mathbb{E}[(B_t^H - B_s^H)^2] &= \frac{1}{c_H^2}\int_{\mathbb{R}} \left((t-u)^{H-\frac{1}{2}}\mathbf{1}_{\{u<t\}} - (s-u)^{H-\frac{1}{2}}\mathbf{1}_{\{u<s\}}\right)^2 du \\
&= \frac{1}{c_H^2}\int_{\mathbb{R}} \left((t-s-v)^{H-\frac{1}{2}}\mathbf{1}_{\{v<t-s\}} - (-v)^{H-\frac{1}{2}}\mathbf{1}_{\{v<0\}}\right)^2 dv \\
&= \frac{(t-s)^{2H}}{c_H^2}\int_{\mathbb{R}} \left((1-w)^{H-\frac{1}{2}}\mathbf{1}_{\{w<1\}} - (-w)^{H-\frac{1}{2}}\mathbf{1}_{\{w<0\}}\right)^2 dw \\
&= (t-s)^{2H},
\end{aligned}$$

where in the second equality we set $u = s + v$ and in the third equality we set $w = \frac{v}{t-s}$. Using that $B_0^H = 0$ $\mathbb{P}$-a.s., we get that

$$\begin{aligned}\mathbb{E}[B_s^H B_t^H] &= \frac{1}{2}\big(\mathbb{E}[(B_s^H - B_0^H)^2] + \mathbb{E}[(B_t^H - B_0^H)^2] - \mathbb{E}[(B_t^H - B_s^H)^2]\big)\\ &= \frac{1}{2}\big(s^{2H} + t^{2H} - (t-s)^{2H}\big),\end{aligned}$$

concluding the proof. □

It turns out that all the irregularity of B^H comes from the second term in the previous decomposition, since the first term has absolutely continuous sample paths.

Theorem 2.3. *The process $Z^H = (Z_t^H)_{t\in\mathbb{T}}$ given by*

$$Z_t^H = \int_0^\infty \big((t+u)^{H-\frac{1}{2}} - u^{H-\frac{1}{2}}\big) dB_u$$

has absolutely continuous trajectories.

Proof. First of all, we notice that $Z_t^H = 0$ almost surely. Consider now the stochastic process given by

$$Y_t = \int_0^\infty \Big(H - \frac{1}{2}\Big)(t+u)^{H-\frac{3}{2}} dB_u$$

for any $t \in \mathbb{T}$, $t > 0$. Since $f_t(u) = (t+u)^{H-\frac{3}{2}}$ is square integrable in $[0,\infty)$ for any $t > 0$, we have that the stochastic integral is well-defined. Then, by using the Stochastic Theorem of Fubini (see, e.g., [Ver12]), we have that

$$\begin{aligned}\int_0^t Y_s ds &= \int_0^t \Big(\int_0^\infty \Big(H - \frac{1}{2}\Big)(t+u)^{H-\frac{3}{2}} dB_u\Big) ds\\ &= \int_0^\infty \Big(\int_0^t \Big(H - \frac{1}{2}\Big)(t+u)^{H-\frac{3}{2}} ds\Big) dB_u\\ &= \int_0^\infty \big((t+u)^{H-\frac{1}{2}} - u^{H-\frac{1}{2}}\big) dB_u\\ &= Z_t^H,\end{aligned}$$

concluding the proof. □

As we already mentioned, an immediate consequence of the previous result is that the process $X^H = (X_t^H)_{t\in\mathbb{T}}$ given by

$$X_t^H = \frac{1}{\Gamma(H+\frac{1}{2})}\int_0^t (t-u)^{H-\frac{1}{2}} dB_u \tag{2.5}$$

has sample paths with the same irregularity as a conventional fBm with Hurst parameter $H \in (0,1)$. The process given by (2.5) is called a fractional Brownian motion of Riemann-Liouville type, see [ML01].

Because of its easier representation, this is the usual process that is used when one wants to consider processes with sample paths as irregular as the ones provided by a fractional Brownian motion, see for example [AMN01] or [ER19].

Remark 2.3. *Notice that in the fBm of Riemann-Liouville type, the constant that appears multiplying the integral is $1/\Gamma(H+\frac{1}{2})$, instead of $1/c_H$. The reason for this is that, when the fBm was first introduced in [MN68], the original expression used to define it was*

$$B_t^H = \frac{1}{\Gamma(H+\frac{1}{2})}\left(\int_{-\infty}^0 \left((t-u)^{H-\frac{1}{2}} - (-u)^{H-\frac{1}{2}}\right) dB_u + \int_0^t (t-u)^{H-\frac{1}{2}} dB_u\right),$$

where Γ denotes the gamma function. After that, it was proved that it was a centered Gaussian process with covariance function given by

$$\mathbb{E}[B_s^H B_t^H] = \frac{V_H}{2}(s^{2H} + t^{2H} - |t-s|^{2H}),$$

where V_H is the constant given by

$$V_H = \frac{\Gamma(2-2H)\cos(\pi H)}{\pi H(1-2H)}.$$

We opted for the more modern approach of defining fBm as a Gaussian process with covariance function given by (2.1), which has an easier representation. However, for simplicity we will keep the constant $1/\Gamma(H+\frac{1}{2})$ instead of $1/c_H$ in the definition of the fBm of Riemann-Liouville type.

Chapter 3

Mathematical finance

Stochastic processes provide useful models of random phenomena, among which a particularly interesting instance is given by the evolution of the values of financial assets listed on a Stock Exchange.

In this chapter we give a brief introduction to the basic concepts of mathematical finance, that will be needed through the rest of the thesis.

3.1 Financial derivatives

Often, people need to get into financial arrangements to exchange a set of different cash-flows at different times. These arrangements are financial contracts, and each person that enters into the agreement is called a counterparty. For example, when someone needs to buy a house, they get a mortgage loan where they receive a large sum of money in exchange for future monthly payments.

When a financial contract depends on the evolution of another asset, it is called a *derivative contract*. This is because the price of the financial contract is 'derived' from the evolution of the asset, which is named the *underlying*. This is usually assumed to be a real stochastic process $S = (S_t)_{t\geq 0}$, defined on some filtered probability space $(\Omega, \mathcal{F}, \mathbb{F}, \mathbb{P})$. The process S is also assumed to be adapted to the filtration $\mathbb{F} := (\mathcal{F}_t)_{t\geq 0}$, which represents the flow of public information.

For simplicity, we will always work with European derivatives, defined as follows.

Definition 3.1 (European derivative). *A European derivative is a pair (Z, T), where T is a strictly positive real number and Z is an $\mathcal{F}_T$-measurable random variable. We call T the maturity or expiration date, and Z the payoff.*

The maturity T indicates the time at which the contract expires and the rights and obligations of the parties under the contract are settled. The payoff Z indicates the amount of money that the buyer of the contract (also called the holder) will receive from the seller (also called the issuer) at the expiration date. Notice that, if Z turns out to be negative, then it is the issuer who will receive money from the holder.

Let us give an example of a very simple European derivative contract.

Definition 3.2 (Forward contract). *A forward contract with maturity $T > 0$ and delivery price $K > 0$ is a European derivative given by payoff $Z = S_T - K$.*

This can be understood as an agreement where the holder promises to buy an asset from the issuer at some specified time T and at some specified price K.

Notice that in a forward contract, the payoff only depends on the value of the underlying at maturity. However, we remark that a European derivative can also be a functional of the whole price process up to time T. For example, an Asian forward with delivery price K is a European derivative whose payoff is given by

$$Z = \frac{1}{T}\int_0^T S_t dt - K.$$

These last two examples are known as *lock products*, which obligate the counterparties to fulfill the terms of the contract at maturity time.

Another kind of derivatives are *option products*. They offer the holder of the contract the right, but not the obligation, to exercise the contract at maturity. As a result, the payoff of an option is a non-negative random variable, since the buyer of the product will not exercise the option if this gives a negative profit (that is, a loss). Just like with derivatives, we will only focus on European options, defined as follows.

Definition 3.3 (European option). *A European option is a European derivative where the holder of the contract has the right, but not the obligation, to exercise the contract at maturity. Equivalently, one can define it as a European derivative with non-negative payoff.*

The simplest and most popular European options are calls and puts, which are also called plain vanilla options. These kind of contracts have an old history, as they became increasingly practised at the end of the 60s with the institution of the Chicago Board Options Exchange (CBOE).

Mathematically, they are defined as follows.

Definition 3.4 (Call option). *A call option with maturity T and strike price K is a European option with payoff given by*

$$Z = (S_T - K)_+ = \begin{cases} S_T - K & \text{if } S_T > K \\ 0 & \text{if } S_T \leq K. \end{cases}$$

Definition 3.5 (Put option). *A put option with maturity T and strike price K is a European option with payoff given by*

$$Z = (K - S_T)_+ = \begin{cases} K - S_T & \text{if } K > S_T \\ 0 & \text{if } K \leq S_T. \end{cases}$$

A call option can be understood as a contract which guarantees to its holder the right (but not the obligation) to buy the underlying at a given time T and at a price K fixed in advance, which is called the strike price. In case $S_T > K$, we have that the holder exercises the option, buying the underlying at the previously settled price K. Then he can immediately sell it to the market at price S_T, therefore making a profit of $S_T - K$. On the contrary, in case $S_T \leq K$, then the holder of the contract does not exercise the option, and so the resulting payoff is simply zero.

One can argue in a similar manner that a put option can be understood as a contract which guarantees to its holder the right (but not the obligation) to sell the underlying at a given time T and at a price K fixed in advance, which is also called the strike price.

Since the appearance of these derivatives in the market, a problem of interest has been to evaluate the right price of an option. For example, in a call option, the issuer of the contract faces a risk: if at maturity time the value of the underlying is greater than the strike price, he would be compelled to hand to the holder of the contract the amount $S_T - K$. The question one is interested in is: how much should the issuer be paid in order to compensate the risk that he is going to face?

In this chapter we will present the key arguments that lead to the determination of the fair price of an option.

3.2 General framework

Through this chapter we consider a filtered probability space $(\Omega, \mathcal{F}, \mathbb{F}, \mathbb{P})$ with finite horizon $T > 0$, where $\mathbb{F} = (\mathcal{F}_t)_{t\in[0,T]}$ and $\mathcal{F}_T = \mathcal{F}$. Whenever not specified, stochastic processes are always supposed to be defined there.

We will consider a financial market model with $d+1$ assets. The first one, which we denote by $S^0 = (S_t^0)_{t\in[0,T]}$, is assumed to evolve as

$$dS_t^0 = rS_t^0 dt, \quad t \in [0,T], \quad S_0^0 = 1,$$

where r is a non-negative constant called the interest rate. The previous equation can be easily solved, yielding

$$S_t^0 = e^{rt}, \qquad t \in [0,T].$$

This asset is usually called the bank account, since it can be viewed as a secure place where one can store money. Indeed, the fact that this asset evolves in a deterministic fashion means that it does not have any risk: if one invests (or borrows) any quantity at a particular time, then one knows exactly how this amount will evolve through time.

Moreover, S_t^0 represents the value of a unit of money (like a dollar) at time t. The fact that r is positive accounts for the concept that a dollar received in the future is worth less than a dollar received today.

The rest of the assets are denoted by $S^i = (S_t^i)_{t\in[0,T]}$, with $i \in \{1,\dots,d\}$. As it is usual, we will also assume that each S^i is a one-dimensional Itô process. In particular, since S^i needs to be strictly positive, we will assume that

$$S_t^i = S_0^i + \int_0^t S_s^i \mu_s^i ds + \int_0^t S_s^i \sigma_s^i dB_s^i,$$

where $(B_t^i)_{t\in[0,T]}$ is a one dimensional Brownian motion, and where $(\mu_t^i)_{t\in[0,T]}$ and $(\sigma_t^i)_{t\in[0,T]}$ are adapted processes with

$$\int_0^T (|\mu_t^i| + |\sigma_t^i|^2) dt < \infty \quad \mathbb{P}\text{-a.s.},$$

so that the integrals are well-defined. Notice that, by applying Itô's formula (see Section A.2) with $f(t,x) = \log(x)$, we easily arrive at

$$S_t^i = S_0^i \exp\Big(\int_0^t (\mu_s^i - \frac{1}{2}(\sigma_s^i)^2) ds + \int_0^t \sigma_s^i dB_s^i \Big),$$

which is strictly positive for any $t \in [0, T]$. Finally, let us also consider the discounted value of S^i, defined as

$$\tilde{S}^i = (\tilde{S}^i_t)_{t\in[0,T]}, \quad \tilde{S}^i_t := e^{-rt} S^i_t.$$

We remark that this is still an Itô process, since

$$\begin{aligned} d\tilde{S}^i_t &= -re^{-rt} S^i_t dt + e^{-rt} dS^i_t \\ &= -re^{-rt} S^i_t dt + e^{-rt} S^i_t (\mu^i_t dt + \sigma^i_t dB^i_t) \\ &= \tilde{S}^i_t ((\mu^i_t - r) dt + \sigma^i_t dB^i_t). \end{aligned} \tag{3.1}$$

Notice that the path given by e^{-rt} is of bounded variation, so the cross-variation between the two processes is zero and the integration by parts rule that we used is the classical one.

The reason why one typically considers the discounted value of an asset rather than its absolute value is because the discounted value takes into account the time value of money, and provides a more accurate representation of its worth in today's terms.

3.2.1 Trading strategies and the self-financing condition

Next we define what a trading strategy in continuous time is. We need to start by defining predictable processes.

Definition 3.6 (Predictable process). *A one-dimensional process $\psi = (\psi_t)_{t\in[0,T]}$ is said to be an elementary predictable process w.r.t. $\mathbb{F} = (\mathcal{F}_t)_{t\in[0,T]}$ if it can be written as*

$$\psi_t = \psi_0 \mathbb{1}_{\{t=0\}}(t) + \sum_{k=1}^{n} H_k \mathbb{1}_{\{t_{k-1}<t\le t_k\}}(t),$$

where H_k are $\mathcal{F}_{t_{k-1}}$-measurable and bounded random variables.

A process is said to be predictable if it belongs to the set of processes generated by elementary predictable processes. In other words, if it is measurable with respect to the σ-field generated by limits of elementary predictable processes.

Definition 3.7 (Trading strategy). *We call $\phi = \big((\phi^0_t, \phi^1_t, \ldots, \phi^d_t)\big)_{t\in[0,T]}$ a trading strategy if each $(\phi^i_t)_{t\in[0,T]}$ is a predictable process. The i-th component of ϕ indicates the number of units of S^i owned at each time $t \in [0, T]$.*

Therefore, the value of the trading strategy ϕ at time $t \in [0, T]$ is given by

$$V_t(\phi) := \phi_t \cdot S_t = \sum_{i=0}^{d} \phi^i_t S^i_t.$$

Note that we are not making any assumptions about the sign of the quantities. If for any $i \in \{1, \ldots, d\}$ we have that $\phi^i_t < 0$, then this amounts to borrowing this number of units of the i-th risky asset and converting them into cash (this is called *short-selling*). If $\phi^0_t < 0$, then this means that the trader is borrowing this amount of monetary units and converting them into the risky assets (that is, a loan to buy risky assets). In fact, the value of ϕ^i_t can be any real number, so that divisibility and total liquidity of the market are assumed. We will also assume that there are no transaction costs.

Consider now a strategy where the trades happen discretely rather than continuously in time. That is, we let ϕ be given by an elementary predictable process

$$\phi_t^i = \phi_0^i \mathbb{1}_{\{t=0\}}(t) + \sum_{k=1}^{n} H_k^i \mathbb{1}_{\{t_{k-1}<t\leq t_k\}}(t), \quad i \in \{0,\dots,d\},$$

where H_k^i are $\mathcal{F}_{t_{k-1}}$-measurable random variables. This means that, at each time t_k, the trader makes the decision of changing the amount of the i-th asset by $\phi_{t_k}^i - \phi_{t_{k-1}}^i$.

We are interested in trading strategies that do this without adding or taking out any wealth. The cost of all these operations at time t_k is given by

$$\sum_{i=0}^{d}(\phi_{t_k}^i - \phi_{t_{k-1}}^i)S_{t_k}^i = (\phi_{t_k} - \phi_{t_{k-1}}) \cdot S_{t_k}.$$

Therefore, if we impose this to be zero, we arrive at

$$\phi_{t_{k-1}} \cdot S_{t_k} = \phi_{t_k} \cdot S_{t_k} = V_{t_k}(\phi).$$

If this holds for all $1 \leq k \leq n$, we call the discrete strategy ϕ self-financing.

What if ϕ is not discrete? Notice that the previous condition implies that

$$V_{t_{k+1}}(\phi) - V_{t_k}(\phi) = \phi_{t_k} \cdot (S_{t_{k+1}} - S_{t_k}).$$

Since in the continuous time case we would be passing to the limit, the self-financing condition can be defined as follows.

Definition 3.8 (Self-financing strategy). *A strategy $\phi = ((\phi_t^0, \phi_t^1, \dots, \phi_t^d))_{t\in[0,T]}$ is self-financing if:*

- $\int_0^T (|\phi_t^0| + \sum_{i=1}^d |\phi_t^i|^2)dt < \infty \quad \mathbb{P}$-*a.s.*
- $V_t(\phi) = V_0 + \int_0^t \phi_s \cdot dS_s$.

Notice that the first condition is needed in order for the integrals in the second condition to be well-defined.

We finish with a necessary and sufficient condition for a trading strategy to be self-financing in terms of discounted values. Just like with a risky stock, we define the discounted value of a strategy ϕ as $\tilde{V}_t(\phi) := e^{-rt}V_t(\phi)$.

Proposition 3.1. *Let $\phi = ((\phi_t^0, \phi_t^1, \dots, \phi_t^d))_{t\in[0,T]}$ be a strategy such that the first condition of the previous definition is satisfied. Then, ϕ is self-financing if and only if*

$$\tilde{V}_t(\phi) = V_0(\phi) + \sum_{i=1}^{d} \int_0^t \phi_s^i d\tilde{S}_s^i. \tag{3.2}$$

Proof. For ease of notation, let us write $\tilde{V}_t$ instead of $\tilde{V}_t(\phi)$. Assume first that ϕ is self-financing. Then, by applying the integration by parts rule to compute the differential of $\tilde{V}_t = e^{-rt}V_t$, we get

$$d\tilde{V}_t = -re^{-rt}V_t dt + e^{-rt}dV_t.$$

From the previous equality and the fact that $S_t^0 = e^{rt}$, we get

$$\begin{aligned} d\tilde{V}_t &= -re^{-rt}(\phi_t^0 e^{rt} + \sum_{i=1}^d \phi_t^i S_t^i)dt + e^{-rt}(\phi_t^0 re^{rt}dt + \sum_{i=1}^d \phi_t^i dS_t^i) \\ &= \sum_{i=1}^d \phi_t^i(-re^{-rt}S_t^i dt + e^{-rt}dS_t^i) \\ &= \sum_{i=1}^d \phi_t^i d\tilde{S}_t^i. \end{aligned}$$

By noticing that $V_0 = \tilde{V}_0$, we arrive at (3.2). The converse is justified similarly. □

3.2.2 Arbitrage and the first fundamental theorem of asset pricing

The two main goals of this section are to define what an arbitrage opportunity is and to see a sufficient condition for the market to be free of it. Before doing that, we need to define the concept of an admissible strategy.

Definition 3.9 (Admissible strategy). *A strategy ϕ is admissible if it is self-financing and its discounted value is bounded from below, i.e. there exists a positive number K such that $\tilde{V}_t(\phi) \geq -K$ for all $t \in [0, T]$.*

Definition 3.10 (Arbitrage opportunity). *We say that there is an arbitrage opportunity if there exists an admissible strategy ϕ with zero initial value, non-negative final value and a non-zero probability of having a strictly positive final value, that is*

1. $V_0(\phi) = 0$,
2. $\mathbb{P}(V_T(\phi) \geq 0) = 1$,
3. $\mathbb{P}(V_T(\phi) > 0) > 0$.

Note that if there is an arbitrage opportunity we can get a strictly positive return with a null initial investment with no risk at all. The crucial assumption when determining the fair price of a derivative is that there are no arbitrage opportunities, in which case the market model is said to be *free of arbitrage*.

Our next goal will be to show a sufficient condition for the market model to be free of arbitrage.

Definition 3.11 (Equivalent probability measures). *Let $\mathbb{P}$ and $\mathbb{P}^*$ be two probability measures defined on a measurable space $(\Omega, \mathcal{F})$. Then, they are said to be equivalent if their null sets are the same, i.e.*

$$\mathbb{P}(A) = 0 \iff \mathbb{P}^*(A) = 0$$

for any $A \in \mathcal{F}$. In this case, we write $\mathbb{P} \sim \mathbb{P}^*$.

Definition 3.12 (Risk-neutral probability measure). *We say that $\mathbb{P}^*$ is a risk-neutral probability measure if it is equivalent to $\mathbb{P}$ and the discounted price processes $\tilde{S}^i = (\tilde{S}_t^i)_{t\in[0,T]}$ are local martingales under $\mathbb{P}^*$. The original probability $\mathbb{P}$ is then called the physical probability measure.*

We are now ready to state a version of the First Fundamental Theorem of Asset Pricing, which says that the existence of a risk-neutral probability measure is enough for the market model to be free of arbitrage.

Theorem 3.1 (First Fundamental Theorem of Asset Pricing, simple version)**.** *Assume that there exists a risk-neutral probability measure $\mathbb{P}^*$. Then, the market model is free of arbitrage.*

Before going through the proof, let us state and prove a technical lemma that we will use.

Lemma 3.1. *Let $X = (X_t)_{t\in[0,T]}$ be a local martingale with $\mathbb{E}[|X_0|] < \infty$. Assume that it is bounded from below. Then, it is a supermartingale, i.e.*

1. $\mathbb{E}[|X_t|] < \infty$ *for all* $t \in [0, T]$,
2. $\mathbb{E}[X_t|\mathcal{F}_s] \leq X_s$ *for all* $0 \leq s \leq t \leq T$.

Proof. Let C be a real number such that $X_t \geq C$ for all $t \geq 0$, and consider the local martingale given by $Y = (Y_t)_{t\geq 0}$, where $Y_t = X_t - C$. By definition of local martingale, there exists a sequence of stopping times $\{\tau_n\}_{n\geq 0}$ with $\mathbb{P}(\lim_{n\to\infty} \tau_n = T) = 1$ such that $Y^n = (Y^n_t)_{t\in[0,T]}$ with $Y^n_t = Y_{t\wedge\tau_n}$ is a martingale for each $n \geq 0$, that is,

$$\mathbb{E}[Y^n_t|\mathcal{F}_s] = Y^n_s$$

for any $0 \leq s \leq t \leq T$. The first statement follows by Fatou's Lemma,

$$\mathbb{E}[|Y_t|] = \mathbb{E}[Y_t] = \mathbb{E}[\liminf_{n\to\infty} Y^n_t] \leq \liminf_{n\to\infty} \mathbb{E}[Y^n_t] = \mathbb{E}[Y_0] < \infty.$$

The second statement follows by the conditional form of Fatou's Lemma (see, e.g., [MR05, Lemma A.0.2]),

$$\mathbb{E}[Y_t|\mathcal{F}_s] = \mathbb{E}[\liminf_{n\to\infty} Y^n_t|\mathcal{F}_s] \leq \liminf_{n\to\infty} \mathbb{E}[Y^n_t|\mathcal{F}_s] = \liminf_{n\to\infty} Y^n_s = Y_s.$$

Finally, since Y is a supermartingale, we also have that X is a supermartingale. □

Before going through the proof of Theorem 3.1, let us mention that there is a more general version of this result, in which the notion of arbitrage is replaced by a stronger concept called No Free Lunch with Vanishing Risk (NFLVR). In such theorem, the existence of a risk-neutral probability measure is not only a sufficient but also a necessary condition for the market model to satisfy the NFLVR condition. See [DS94] for further details.

Proof of Theorem 3.1. Consider an admissible strategy ϕ with zero initial value. Recall that, by Proposition 3.1,

$$\tilde{V}_t(\phi) = \sum_{i=1}^{d} \int_0^t \phi^i_s d\tilde{S}^i_s.$$

Each S^i is a local martingale under $\mathbb{P}^*$, and so $\tilde{V}_t(\phi)$ is also a local martingale under $\mathbb{P}^*$ (see Proposition A.4). Since it is bounded from below, by Proposition 3.1 we have that $\tilde{V}_t(\phi)$ is a supermartingale under $\mathbb{P}^*$. Consequently,

$$\mathbb{E}_{\mathbb{P}^*}[\tilde{V}_T(\phi)] = \mathbb{E}_{\mathbb{P}^*}[\tilde{V}_T(\phi)|\mathcal{F}_0] \leq \tilde{V}_0(\phi) = 0 \implies \mathbb{E}_{\mathbb{P}^*}[V_T(\phi)] \leq 0.$$

If we assumed that $\mathbb{P}(V_T(\phi) \geq 0) = 1$, by using $\mathbb{P} \sim \mathbb{P}^*$ we would have that $\mathbb{P}^*(V_T(\phi) \geq 0) = 1$, concluding that $\mathbb{P}^*(V_T(\phi) = 0) = 1$. Since $\mathbb{P} \sim \mathbb{P}^*$, we conclude that $\mathbb{P}(V_T(\phi) = 0) = 1$, ruling out that ϕ is an arbitrage strategy. □

3.2.3 Pricing options

In the previous section we have seen a sufficient condition for the market model to be free of arbitrage. However, we still don't know how to give a fair price to any option. This will be the goal of the following pages. Unless stated otherwise, from now on we will always assume the existence of a risk-neutral probability measure $\mathbb{P}^*$.

The first case that is usually studied is when one can replicate the final payoff of an option by using an admissible strategy.

Definition 3.13 (Replicable option)**.** *We say that an option is replicable if its payoff Z is equal to the final value of an admissible strategy. That is, there exists an admissible strategy ϕ such that $V_T(\phi) = Z$.*

Definition 3.14 (Complete market model)**.** *We say that the market model is complete if any option with a square integrable payoff with respect to $\mathbb{P}^*$ is replicable.*

If an option with payoff Z is replicable and we denote by ϕ the replicating strategy, by a simple no arbitrage argument one can easily see that its price at time t is given by $V_t(\phi)$.

However, to impose the market model to be complete is a very strong restriction. Indeed, only in the simplest models, like Black-Scholes, one is able to replicate any square integrable option. In the more sophisticated ones, where the volatility is stochastic and evolves as a separate Itô process, the market model is not complete.

Remark 3.1. *How should one price options in such incomplete market models? The answer is fairly simple: if we denote by $X = (X_t)_{t\in[0,T]}$ the price process of an option with payoff Z, we can introduce X as a new risky asset in the market. By doing so, one is (trivially) able to replicate Z, since we can simply define the strategy given by buying one unit of X at time $t = 0$ and hold it until maturity time. Notice that, in this case, the value of this strategy at time t is precisely given by X_t.*

The following theorem gives us an expression for the process X such that the market model is free of arbitrage.

Theorem 3.2. *Consider an option with payoff Z that is integrable with respect to $\mathbb{P}^*$, and denote by $X = (X_t)_{t\in[0,T]}$ its price process. Then, if*

$$X_t = \mathbb{E}_{\mathbb{P}^*}[e^{-r(T-t)}Z|\mathcal{F}_t],$$

we have that the market model is free of arbitrage.

Proof. Let us denote by $\tilde{X} = (\tilde{X}_t)_{t\in[0,T]}$ the discounted value of X, where $\tilde{X}_t = e^{-rt}X_t$. By Theorem 4.8 and Remark 3.1, it is enough to prove that $\tilde{X}$ is a martingale under $\mathbb{P}^*$. Indeed,

$$\begin{aligned}\mathbb{E}_{\mathbb{P}^*}[\tilde{X}_t|\mathcal{F}_s] &= \mathbb{E}_{\mathbb{P}^*}[e^{-rt}X_t|\mathcal{F}_s] = \mathbb{E}_{\mathbb{P}^*}[e^{-rt}\mathbb{E}_{\mathbb{P}^*}[e^{-r(T-t)}Z|\mathcal{F}_t]|\mathcal{F}_s] \\ &= e^{-rs}\mathbb{E}_{\mathbb{P}^*}[e^{-r(T-s)}Z|\mathcal{F}_s] = \tilde{X}_s,\end{aligned}$$

where in the last equality we have used the tower property of conditional expectations. $\square$

In our practical applications, we will be only interested in pricing options at the initial time $t = 0$, and so we will be using the expression given by

$$X_0 = \mathbb{E}_{\mathbb{P}^*}[e^{-rT}Z].$$

Moreover, we will only study options with payoffs that depend on one underlying, and so we will only consider a market model with one risky asset.

3.2.4 Call-Put parity condition

Let us next introduce a formula that relates the prices of European call options and European put options with the same underlying asset, strike price, and expiration date. This is achieved in a model-free manner, only assuming the existence of a risk-neutral probability. This is the reason why in the upcoming sections we will only focus on pricing call options.

Theorem 3.3 (Call-put parity). *Let $\mathbb{P}^*$ be a risk-neutral probability measure, and let C_t and P_t denote the corresponding price of a call option and a put option, respectively, at time $t \in [0, T]$, both with strike price K and maturity T. Then, we have that*

$$C_t - P_t = S_t - Ke^{-r(T-t)}.$$

Proof. First of all, note that

$$S_T - (S_T - K)_+ + (K - S_T)_+ = K,$$

which means that

$$S_T - C_T + P_T = K.$$

Multiplying both sides by e^{-rT}, we arrive at

$$\tilde{S}_T - \tilde{C}_T + \tilde{P}_T = Ke^{-rT}.$$

Now, $(\tilde{S}_t)_{t\in[0,T]}$, $(\tilde{C}_t)_{t\in[0,T]}$ and $(\tilde{P}_t)_{t\in[0,T]}$ are local martingales under $\mathbb{P}^*$, and so is the process $X = (X_t)_{t\in[0,T]}$ defined by

$$X_t := \tilde{S}_t - \tilde{C}_t + \tilde{P}_t.$$

Therefore, we have that for any $t \in [0, T]$

$$Ke^{-rT} = \mathbb{E}[Ke^{-rT}|\mathcal{F}_t] = \mathbb{E}[X_T|\mathcal{F}_t] = \mathbb{E}[\lim_{n\to\infty} X_{T\wedge\tau_n}|\mathcal{F}_t],$$

where $\{\tau_n\}_{n\geq 0}$ is the sequence of stopping times from the definition of a local martingale. By the Dominated Convergence Theorem, we can interchange the limit and the expectation, arriving at

$$\mathbb{E}[\lim_{n\to\infty} X_{T\wedge\tau_n}|\mathcal{F}_t] = \lim_{n\to\infty} \mathbb{E}[X_{T\wedge\tau_n}|\mathcal{F}_t] = X_t.$$

We therefore obtain that

$$Ke^{-rT} = \tilde{S}_t - \tilde{C}_t + \tilde{P}_t,$$

and the result follows by multiplying both sides by e^{rt}. □

3.2.5 Change of measure

For the simplest market models, the usual approach is to present the dynamics of S under the physical measure $\mathbb{P}$, and then find an equivalent probability measure $\mathbb{P}^*$ under which $\tilde{S}$ is a local martingale.

In order to do so, Girsanov's theorem is the standard tool used to explicitly find that change of measure.

Theorem 3.4 (Girsanov)**.** *Let $B = (B_t)_{t\in[0,T]}$ be a Brownian motion defined in a filtered probability space $(\Omega, \mathcal{F}, \mathbb{F}, \mathbb{P})$, and $(\theta_t)_{t\in[0,T]}$ an $\mathbb{F}$-adapted process satisfying $\int_0^T \theta_s^2 ds < \infty$ $\mathbb{P}$-a.s. and such that the process $L = (L_t)_{t\in[0,T]}$ defined by*

$$L_t = \exp\Big(-\int_0^t \theta_s dB_s - \frac{1}{2}\int_0^t \theta_s^2 ds\Big) \tag{3.3}$$

is a martingale under $\mathbb{P}$. Let $\mathbb{P}^*$ be the probability defined on $(\Omega, \mathcal{F})$ having density L_T with respect to $\mathbb{P}$. Then, the process $(B_t^*)_{t\in[0,T]}$ defined by $B_t^* = B_t + \int_0^t \theta_s ds$ is an $\mathbb{F}$-standard Brownian motion under $\mathbb{P}^*$.

Proof. See [Bal17, Theorem 12.1]. □

Recall that under the physical probability measure $\mathbb{P}$, the discounted stock price $\tilde{S} = (\tilde{S}_t)_{t\in[0,T]}$ could be written as

$$\tilde{S}_t = \tilde{S}_0 + \int_0^t \tilde{S}_s(\mu_s - r)dt + \int_0^t \tilde{S}_s\sigma_s dB_s, \quad \forall t \in [0,T].$$

Assume now that $\sigma_t > 0$ for all $t \in [0,T]$, and write $\theta_t = \frac{\mu_t - r}{\sigma_t}$. Then, if we assume that all the assumptions in Girsanov's theorem are satisfied, we would have that $B_t^* = B_t + \int_0^t \theta_s ds$ would be an $\mathbb{F}$-standard Brownian motion under $\mathbb{P}^*$.

Therefore, we could write

$$\begin{aligned}
\tilde{S}_t &= \tilde{S}_0 + \int_0^t \tilde{S}_s(\mu_s - r)dt + \int_0^t \tilde{S}_s\sigma_s dB_s \\
&= \tilde{S}_0 + \int_0^t \tilde{S}_s(\mu_s - r)dt + \int_0^t \tilde{S}_s\sigma_s d(B_s^* - \int_0^t \frac{\mu_s - r}{\sigma_s} ds) \\
&= \tilde{S}_0 + \int_0^t \tilde{S}_s\sigma_s dB_s^*,
\end{aligned}$$

which makes $\tilde{S}$ a local martingale under $\mathbb{P}^*$.

As we already explained, the previous approach is the one that is usually taken when introducing the most popular market models, like Black-Scholes or Heston's. However, some more sophisticated models are presented directly under the risk-neutral probability measure $\mathbb{P}^*$, and one simply assumes that the physical probability measure $\mathbb{P}$ is equivalent to $\mathbb{P}^*$.

Moreover, the assumption that $\sigma_t > 0$ might not be true in some recently proposed market models that rely on machine learning. As we will see in Chapter 5, this will also be the case for our model.

For all this reasons, in the next sections we will present different models directly under the risk-neutral probability measure $\mathbb{P}^*$, without worrying about the physical probability measure $\mathbb{P}$.

Through the rest of the chapter we will always assume that we are working in the filtered probability space $(\Omega, \mathcal{F}, \mathbb{F}, \mathbb{P}^*)$, where $\mathbb{P}^*$ is a risk-neutral probability measure.

3.3 The Black-Scholes model

In 1973, Black and Scholes [BS73] and Merton [Mer73] proposed the most well-known model for the dynamics of S. Under the risk-neutral probability measure $\mathbb{P}^*$, this reads as

$$dS_t = S_t(rdt + \sigma dB_t),$$

or equivalently,

$$d\tilde{S}_t = \sigma \tilde{S}_t dB_t,$$

where $\sigma > 0$ is a constant and $(B_t)_{t\in[0,T]}$ is a Brownian motion. Under this dynamics, by a direct application of Itô's formula with $f(x) = \log x$ we have that

$$S_t = S_0 \exp\Big((r - \frac{1}{2}\sigma^2)t + \sigma B_t\Big). \tag{3.4}$$

3.3.1 Pricing in the Black-Scholes model

The reason why this model has become so important is because of its simplicity, and more importantly, due to the fact that there exists a closed form formula for the price of a call option.

Theorem 3.5. *Consider S given by (3.4). Then, the price of a call option at time $t = 0$ with strike price K, maturity time T and spot price S_0, is given by*

$$C_{BS}(T, S_0, K, r, \sigma) = S_0\Phi(d_+) - Ke^{-rT}\Phi(d_-)$$

with

$$d_\pm = \frac{\log(S_0/K) + (r \pm \frac{1}{2}\sigma^2)T}{\sigma\sqrt{T}},$$

and where Φ stands for the cumulative probability function of the standard normal law. Notice that we have included σ and r as parameters of the pricing function.

Proof. Recall that by Theorem 3.2, we have that $C_{BS}(T,S_0,K,r,\sigma) = \mathbb{E}_{\mathbb{P}^*}(e^{-rT}(S_T - K)_+)$. Then,

$$\begin{aligned}\mathbb{E}_{\mathbb{P}^*}(e^{-rT}(S_T-K)_+) &= e^{-rT}\mathbb{E}_{\mathbb{P}^*}(S_T\mathbf{1}_{\{S_T>K\}}) - Ke^{-rT}\mathbb{E}_{\mathbb{P}^*}(\mathbf{1}_{\{S_T>K\}})\\ &= e^{-rT}S_0\mathbb{E}_{\mathbb{P}^*}\left(\frac{S_T}{S_0}\mathbf{1}_{\{\frac{S_T}{S_0}>\frac{K}{S_0}\}}\right) - Ke^{-rT}\mathbb{E}_{\mathbb{P}^*}\left(\mathbf{1}_{\{\frac{S_T}{S_0}>\frac{K}{S_0}\}}\right).\end{aligned}$$

Notice that $\frac{S_T}{S_0} = \exp\left((r-\frac{1}{2}\sigma^2)T + \sigma B_T\right)$. Now, we have that

$$\begin{aligned}\mathbb{E}_{\mathbb{P}^*}\left(\mathbf{1}_{\{\frac{S_T}{S_0}>\frac{K}{S_0}\}}\right) &= \mathbb{P}^*\left(\frac{S_T}{S_0} > \frac{K}{S_0}\right)\\ &= \mathbb{P}^*\left(\log\frac{S_T}{S_0} > \log\frac{K}{S_0}\right)\\ &= \mathbb{P}^*\left(\frac{B_T}{\sqrt{T}} < \frac{\log\frac{S_0}{K} + (r-\frac{1}{2}\sigma^2)T}{\sigma\sqrt{T}}\right)\\ &= \Phi\left(\frac{\log\frac{S_0}{K} + (r-\frac{1}{2}\sigma^2)T}{\sigma\sqrt{T}}\right)\\ &= \Phi(d_-).\end{aligned}$$

On the other hand, we have that

$$\begin{aligned}e^{-rT}\mathbb{E}_{\mathbb{P}^*}\left(\frac{S_T}{S_0}\mathbf{1}_{\{\frac{S_T}{S_0}>\frac{K}{S_0}\}}\right) &= e^{-rT}\mathbb{E}_{\mathbb{P}^*}\left(\exp\left\{(r-\frac{1}{2}\sigma^2)T + \sigma B_T\right\}\mathbf{1}_{\left\{\frac{B_T}{\sqrt{T}}<\frac{\log S_0/K+(r-\frac{1}{2}\sigma^2)T}{\sigma\sqrt{T}}\right\}}\right)\\ &= \mathbb{E}_{\mathbb{P}^*}\left(\exp\left\{-\frac{1}{2}\sigma^2T + \sigma B_T\right\}\mathbf{1}_{\left\{\frac{B_T}{\sqrt{T}}<\frac{\log S_0/K+(r-\frac{1}{2}\sigma^2)T}{\sigma\sqrt{T}}\right\}}\right)\\ &= \frac{1}{\sqrt{2\pi}}\int_{-\infty}^{\frac{\log S_0/K+(r-\frac{1}{2}\sigma^2)T}{\sigma\sqrt{T}}} \exp\left\{-\frac{1}{2}\sigma^2T - \sigma\sqrt{T}y - \frac{1}{2}y^2\right\}dy\\ &= \frac{1}{\sqrt{2\pi}}\int_{-\infty}^{\frac{\log S_0/K+(r-\frac{1}{2}\sigma^2)T}{\sigma\sqrt{T}}} \exp\left\{-\frac{1}{2}\left(\sigma\sqrt{T}+y\right)^2\right\}dy\\ &= \frac{1}{\sqrt{2\pi}}\int_{-\infty}^{\frac{\log S_0/K+(r+\frac{1}{2}\sigma^2)T}{\sigma\sqrt{T}}} \exp\left\{-\frac{1}{2}u^2\right\}du\\ &= \Phi(d_+).\end{aligned}$$

□

3.3.2 The implied volatility

Notice that the price of a call option depends on just one unknown parameter, the volatility σ, which in the Black-Scholes model is assumed to be constant. Given the observed market price of a call option, one can determine the value of σ such that the Black-Scholes price exactly matches it. This is called the implied volatility.

Definition 3.15 (Implied volatility)**.** *Let $C^{market}(T,K)$ be the observed market price of a call option with strike K and maturity time T. We define its implied volatility, denoted by $\sigma_{BS}(T,K)$, as the volatility we use in the Black-Scholes formula to obtain the market price of* *the option. That is,*

$$C^{market}(T,K) = C_{BS}(T,S_0,K,r,\sigma_{BS}(T,K)).$$

One can easily check that there is a unique value for the implied volatility. Indeed, we have that

$$\frac{\partial C_{BS}(T,S_0,K,r,\sigma)}{\partial \sigma} = S_0\sqrt{T}\Phi(d_+) > 0,$$

and so the Black-Scholes price of a call is strictly increasing with respect to the volatility: that is, a higher volatility means a higher price on the call. In addition, the upper and lower limits are, respectively,

$$\lim_{\sigma\to+\infty} C_{BS}(T,S_0,K,r,\sigma) = S_0,$$
$$\lim_{\sigma\to 0^+} C_{BS}(T,S_0,K,r,\sigma) = (S_0 - Ke^{-rT})_+,$$

which are two non-arbitrage bounds on the price of a call option. Therefore, by the inverse function theorem we have that for every non-arbitrage price there is a unique implied volatility.

The volatility surface

In this section we will talk about some empirical facts about the volatility surface, focusing on equity markets, that is, markets where stocks and shares of companies are traded.

The surface generated by the function $\sigma_{\mathrm{BS}}(K,T)$ is called the volatility surface. By observing a set of prices for calls with different strikes and maturities in the market, we can get an approximation of it. Here is an example with real market data.

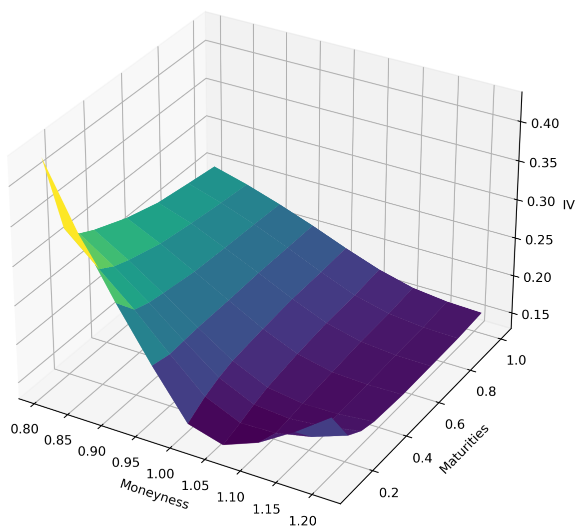

FIGURE 3.1: Implied volatility surface, S&P500 index as of March 17, 2021. The moneyness is defined as K/S_0.

Notice that in this example the empirical implied volatility is not a constant function. Indeed, this is not a particularity for that trading day, and instead one can observe it in every trading day of the year: we say that it is a stylized fact of the volatility. This is the main reason why we know that the Black-Scholes model is very inaccurate in representing the real dynamics of S.

In some sense, the volatility surface describes the deviation of market option prices from the theoretical values obtained within the Black-Scholes model. It can be then summarized as the wrong parameter to put in the wrong formula to get the right price [Leo14].

Besides giving us empirical proof that the Black-Scholes model generates unrealistic dynamics, the implied volatility is widely used as a way of quoting prices of calls more than the prices themselves, since it gives a better idea of their relative value.

The implied volatility curve

Another interesting way of looking at the implied volatility surface is by fixing the maturity date T and consider the mapping $K \mapsto \sigma_{\text{BS}}(T, K)$, which is called the implied volatility curve.

We say that we have a *skew* if implied volatilities are higher for low strikes than for high strikes, and a *smile* if the curve has a minimum, which usually occurs for K lying not far away from the forward value given by $S_0 e^{rT}$.

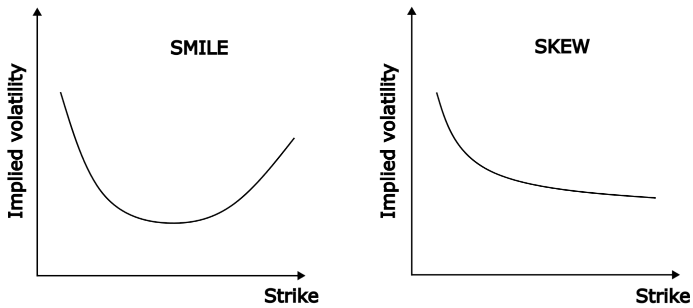

FIGURE 3.2: Implied volatility smile and skew, for T fixed.

For short maturities, the smile (or skew) effect is very pronounced, but it almost completely disappears for longer maturities. Another important characteristic of the volatility surface is the term structure of at-the-money (ATM) volatility skew, which is defined at time t as

$$\psi(\tau) := \left| \frac{\partial \sigma_{\text{BS}}(\tau, K)}{\partial K} \right|_{K=S_0}.$$

Empirically, it turns out that this can be well-approximated by a power-law function $\Phi(\tau) = A\tau^{-\alpha}$, with $0 < \alpha < 1/2$, see [GJR18].

3.4 Modelling volatility

Besides not being constant, the level and orientation of the volatility surface tend to move according to changes in market prices and market conditions. Then, in order to reflect all the characteristics that we have seen, it is necessary to introduce models of volatility. By introducing uncertainty in the behavior of volatility, the evolution of financial assets can be estimated more realistically. Therefore, one usually looks for models of the stock price S of the form

$$dS_t = S_t(rdt + \sigma_t dB_t),$$

and the goal is to determine the process $\sigma = (\sigma_t)_{t\in[0,T]}$.

Let us make a small overview of the different families of models that exist in the literature. As we have seen in the previous section, in the Black-Scholes framework the volatility is constant. In Dupire's local volatility model [Dup94], the volatility σ_t is a deterministic function of the spot price S_t and time t, chosen to exactly match observed plain vanilla option prices exactly. Such a model is by definition time-inhomogeneous; its dynamics are highly unrealistic, typically generating future volatility surfaces completely unlike those we observe.

On the other hand, in classical stochastic volatility models, the volatility is modeled as a continuous Brownian semimartingale, usually by making them satisfy an SDE like

$$d\sigma_t = \alpha(\sigma_t)dt + \beta(\sigma_t)dW_t, \tag{3.5}$$

where $(W_t)_{t\in[0,T]}$ is a Brownian motion possibly correlated with $(B_t)_{t\in[0,T]}$. That is,

$$dW_t = \rho dB_t + \sqrt{1-\rho^2}dZ_t,$$

where $(Z_t)_{t\in[0,T]}$ is a standard Brownian motion independent of $(B_t)_{t\in[0,T]}$.

3.4.1 The Heston model

Let us now introduce the Heston model [Hes93], which is the most well-known classical stochastic volatility model. It became very popular for being able to capture the volatility smile and having a semi-analytical closed form for the price of a plain vanilla option.

In the Heston model the price process S satisfies the following system of stochastic differential equations

$$\begin{aligned} dS_t &= S_t(rdt + \sqrt{v_t}dB_t) \\ dv_t &= \kappa(\theta - v_t)dt + \nu\sqrt{v_t}dW_t \\ dW_t &= \rho dB_t + \sqrt{1-\rho^2}dZ_t, \end{aligned}$$

where $v := (v_t)_{t\in[0,T]}$ is the variance process, that is, $v_t = \sigma_t^2$. Notice that by using the Itô formula with $f(x) = \sqrt{x}$ we could instead directly model the volatility process, but with a more complicated equation.

In the previous SDE system, we have that:

- $\theta > 0$ is the long-run mean level of the variance,
- $\kappa > 0$ is the rate at which the variance process reverts to the mean θ,
- $\nu > 0$ is the volatility of volatility parameter.

One usually chooses the parameters such that they satisfy the Feller condition $2\kappa\theta \geq \nu^2$, which ensures positivity of the variance process.

The biggest difference between the Heston model and other stochastic volatility models is the existence of a fast and easily implemented quasi-closed form solution for European options, which as we will see in Section 5.2 becomes critical when calibrating any model to known option prices. See [GT06, Chapter 2] for the Heston solution of a call option.

Although the Heston model is more realistic than the Black-Scholes model, generated option prices are not consistent with the ones observed in the market, especially for options with short maturities. This problem also holds in any stochastic volatility model where the volatility is modeled as a continuous Brownian semimartingale. In particular, the at-the-money volatility skew tends to a constant value as $\tau \to 0$ for any model of the volatility of the form (3.5). Indeed, as it is proved for example in [GT06, Chapter 7],

$$\psi(\tau) := \left|\frac{\partial \sigma_{\mathrm{BS}}(\tau, K)}{\partial K}\right|_{K=S_0} \xrightarrow{\tau \to 0} \frac{\rho \beta(\sigma_0)}{2\sigma_0}$$

where σ_0 is the initial value for the volatility process. This is in contrast with the empirical observations made on the volatility surface of real market data, where the ATM volatility skew tends to infinity as $\tau \to 0$.

3.4.2 Rough volatility models

In the celebrated paper [GJR18], Gatheral, Jaisson and Rosenbaum showed empirically that, from an analysis of the time series of realized volatility[1] using recent high frequency data, the logarithm of realized volatility behaves essentially as a fractional Brownian motion with Hurst parameter of order 0.1 at any reasonable time scale.

The previous explanation justifies the use of fractional Brownian motion to model the volatility process in the physical measure. But what about in the risk-neutral probability measure? Is there any advantage in using fBm over Brownian motion whenever we are interested in pricing options?

It turns out that the answer is positive. Indeed, in [Fuk11, Section 3.3] it is shown that in stochastic volatility models where the volatility is driven by a (Riemann-Liouville) fractional Brownian motion with Hurst parameter H, the generated ATM volatility skew is of the form $\psi(\tau) \sim \tau^{H-\frac{1}{2}}$, at least for small τ. By choosing $H \approx 0.1$, we get dynamics that generate volatility surfaces that are more realistic than classical stochastic volatility models, especially for shorter maturities. This new class of models are called rough volatility models, and they were first proposed in [ALV07].

Probably the most important representative is the Rough Heston model, see [ER19]. Its dynamics are given by

$$dS_t = S_t(rdt + \sqrt{v_t}dB_t)$$
$$v_t = v_0 + \frac{1}{\Gamma(\alpha)}\int_0^t (t-s)^{\alpha-1}\kappa(\theta - v_s)ds + \frac{1}{\Gamma(\alpha)}\int_0^t (t-s)^{\alpha-1}\kappa\nu\sqrt{v_s}dW_s$$

where $\alpha = H + \frac{1}{2}$ and $dW_t = \rho dB_t + \sqrt{1-\rho^2}dZ_t$. All the other parameters have a similar meaning to the classical Heston model.

[1]Realized volatility is an estimator of the underlying volatility process given a set of discretely sampled data from the underlying price process.

Chapter 4

The signature transform

In this chapter we introduce the fundamental object in this thesis, called the *signature of a path* or *signature transform*. As we will see, whenever the given path is regular enough, its signature is given by the infinite sequence of its iterated integrals in the Young-Stieltjes sense.

For more irregular paths, like trajectories of a Brownian motion, things are not that simple. Indeed, one is not able to define the signature from a given path, and is in fact forced to lift this path to a higher dimensional object, called a rough path, where one is able to define its signature. Since this lift is not unique, we get that signatures are defined on rough paths rather than on paths.

In our case, we will only consider lifts that are convenient to us from a theoretical and practical point of view, such that the signature will capture the most important analytic and geometric qualities of the given (rough) path, making it a very convenient transformation in modelling problems.

4.1 Paths with finite p-variation

Definition 4.1 (Partition of interval)**.** *A partition or a subdivision of* $[0, T]$ *is an increasing sequence of real numbers* $\mathcal{D} = (t_0, t_1, \ldots, t_n)$ *such that* $0 = t_0 < t_1 < \cdots < t_n = T$*. We will use the notation* $\mathcal{D} \subset [0, T]$ *to indicate that* $\mathcal{D}$ *is a partition of* $[0, T]$*.*

In the sequel, V will be a d-dimensional real vector space whose norm is denoted by $|\cdot|$. Let us first define a notion of irregularity of a V-valued continuous path on a compact interval $[0, T]$, namely its p-variation.

Definition 4.2 (p-variation)**.** *Let* $p \geq 1$ *be a real number and* $x : [0, T] \to V$ *be a continuous path. The p-variation of x on the interval* $[0, T]$ *is defined as*

$$\|x\|_{p;[0,T]} := \left[\sup_{\mathcal{D}\subset[0,T]} \sum_{t_i \in \mathcal{D}} |x_{t_{i+1}} - x_{t_i}|^p \right]^{1/p}.$$

We denote by $\mathcal{V}^p([0, T]; V)$ *the linear space of paths which have finite p-variation.*

Let us emphasize that we are considering the supremum over all subdivisions of $[0, T]$, and not a limit as the mesh of the subdivision goes to zero. Unless $p = 1$, this is quite different.

As we already mentioned, the p-variation is a measure of the regularity or smoothness of a path. It is easy to check that the p-variation decreases with p, which means that for any $1 < p < q$ we have the set of inclusions

$$\mathcal{V}^1([0,T];V) \subset \mathcal{V}^p([0,T];V) \subset \mathcal{V}^q([0,T];V) \subset C([0,T];V).$$

Let us now show an important relation between α-Hölder continuous paths and paths of finite p-variation.

Proposition 4.1. *Let $\alpha \in (0,1]$, and let $x : [0,T] \to V$ be a α-Hölder continuous path, i.e. such that*

$$|x_t - x_s| \leq C|t-s|^\alpha \quad \text{for all } 0 \leq s \leq t \leq T$$

for some $C > 0$. We then have that the $\frac{1}{\alpha}$-variation of x is finite. Conversely, if x is continuous and has finite p-variation, then there exists a reparameterization τ such that $x \circ \tau$ is $\frac{1}{p}$-Hölder continuous.

This proposition provides us with many examples of paths belonging to $\mathcal{V}^p([0,T];V)$ for different $p \geq 1$. For example, the sample paths of a Brownian motion are almost surely α-Hölder continuous for any $\alpha < \frac{1}{2}$, which means that they are almost surely of finite p-variation for any $p > 2$. More generally, we can make the same statement for continuous semimartingales.

In the case of a fractional Brownian motion with Hurst parameter H, its sample paths are almost surely α-Hölder continuous for any $\alpha < H$, which means that they are almost surely of finite p-variation for any $p > \frac{1}{H}$.

4.2 The Young-Stieltjes integral

In this section we give a brief exposition of the Young-Stieltjes integral, also known as the Riemann-Stieltjes integral, along with some well-known results.

Definition 4.3 (Young-Stieltjes integral)**.** *Let $x, y : [0,T] \to \mathbb{R}$ be two continuous paths. Let $\{\mathcal{D}_n\}_{n\geq 1}$ with $\mathcal{D}_n = (t_0^n, t_1^n, \ldots, t_n^n)$ be a sequence of partitions of $[0,T]$ with $|\mathcal{D}_n| \to 0$, and ξ_i^n some points in $[t_i^n, t_{i+1}^n]$. Assume that*

$$\sum_{i=0}^{n-1} y_{\xi_i^n}(x_{t_{i+1}^n} - x_{t_i^n})$$

converges when n tends to ∞ to a limit I independent of the choice of ξ_i^n and the sequence $\{\mathcal{D}_n\}_{n\geq 1}$. Then, we say that the Young-Stieltjes integral of y against x on $[0,T]$ exists and write

$$\int_0^T y_u dx_u := I.$$

We call y the integrand and x the integrator.

It is well-known that if the integrator x has finite 1-variation (also called of bounded variation), then the integral $\int_0^t y_u dx_u$ exists for every $t \in [0,T]$, and as a function of t, this integral itself has bounded variation, with the bound

$$\left\| \int_0^\bullet y_u dx_u \right\|_{1;[0,T]} \leq \|y\|_{\infty;[0,T]} \|x\|_{1;[0,T]}, \tag{4.1}$$

where $\|y\|_{\infty;[0,T]}$ is the uniform norm on $[0,T]$.

In 1936, L.C. Young proved that in certain situations where x has unbounded variation, $\int_0^T y_u dx_u$ still exists in the sense explained above, see [You36]. Of course, this requires more regularity on y than just continuity. To be precise, we have the following result.

Theorem 4.1 (Young). *Let $p, q \geq 1$ be two real numbers such that $p^{-1} + q^{-1} > 1$. Consider $x \in \mathcal{V}^p([0,T];\mathbb{R})$ and $y \in \mathcal{V}^q([0,T];\mathbb{R})$. Then, for each $t \in [0,T]$, the Young-Stieltjes integral exists.*

Moreover, as a function of t, the integral belongs to $\mathcal{V}^p([0,T];\mathbb{R})$, and there exists a constant $C_{p,q}$, which depends only on p and q, such that the following inequality holds:

$$\left\| \int_0^{\bullet} (y_u - y_0) dx_u \right\|_{p;[0,T]} \leq C_{p,q} \|y\|_{q;[0,T]} \|x\|_{p;[0,T]} .$$

Proof. See, e.g., [LCL04, Theorem 1.16]. □

Finally, let us mention that the Young-Stieltjes integral satisfies the usual integration by parts rule. As we will see, this will be useful later.

Theorem 4.2 (Integration by parts, Young-Stieltjes integral). *Let $x \in \mathcal{V}^p([0,T];\mathbb{R})$ and $y \in \mathcal{V}^q([0,T];\mathbb{R})$, for $p, q \geq 1$ such that $p^{-1} + q^{-1} > 1$. Then we have that*

$$\int_s^t y_u dx_u + \int_s^t x_u dy_u = x_t y_t - y_s x_s$$

for any $0 \leq s \leq t \leq T$.

Proof. See, e.g., [Joh81, Theorem 53.3]. □

The main limitation of the Young-Stieltjes integral lies in the constrain $p^{-1} + q^{-1} > 1$, which must hold on the regularities of x and y for the integral $\int_0^t y_u dx_u$ to exist. With probabilistic applications in mind, this is a drastic restriction, since for example trajectories of a Brownian motion have finite p-variation only for $p > 2$. This means that, for example, simple integrals like $\int_0^t B_u dB_u$ are almost surely not well-defined in the Young-Stieltjes sense.

4.3 Formal series of tensors

This section is devoted to the definition and study of tensor algebras over V, which are formal series of tensors equipped with certain algebraic operations. These will be crucial in the following sections, as signatures of paths take their values in this kind of spaces.

Let us start by defining the n-th tensor power over V.

Definition 4.4 (n-th tensor power). *Let n be a positive integer. The n-th tensor power over V is defined as*

$$V^{\otimes n} := V \otimes \overset{n)}{\cdots} \otimes V,$$

where $\otimes$ denotes the tensor product. By convention, $V^{\otimes 0} := \mathbb{R}$.

Let $\{e_1, \ldots, e_d\} \subset V$ be a basis of V. Then, the set given by

$$\{e_{i_1} \otimes \cdots \otimes e_{i_n} \mid i_1, \ldots, i_n \in \{1, \ldots, d\}\} \tag{4.2}$$

is a basis of $V^{\otimes n}$. Therefore, any element of $V^{\otimes n}$ can be written as

$$\sum_{i_1,\ldots,i_n \in \{1,\ldots,d\}} \lambda_{i_1,\ldots,i_n} e_{i_1} \otimes \cdots \otimes e_{i_n} \quad \text{with } \lambda_{i_1,\ldots,i_n} \in \mathbb{R},$$

which means that $V^{\otimes n}$ can be identified as the space of formal non-commuting homogeneous polynomials of degree n in d indeterminates $e_1, ..., e_d$ and real coefficients. Notice that, in particular, this means that $V^{\otimes n}$ is a real vector space of dimension d^n.

Let us set some notation on the multi-index $I := (i_1, \ldots, i_n)$. If $n \geq 1$ or $n \geq 2$, we write $I' := (i_1, \ldots, i_{n-1})$ and $I'' := (i_1, \ldots, i_{n-2})$. We will denote its length by writing $|I| := n$. We consider the empty index $I := \emptyset$ and set $|I| := 0$. We will also use the notation

$$\{I : |I| = n\} := \{1, \ldots, d\}^n,$$

omitting the parameter d whenever this does not introduce ambiguity. Finally, for each $|I| \geq 1$, we set

$$e_I := e_{i_1} \otimes \cdots \otimes e_{i_n}.$$

We are now ready to define the extended tensor algebra over V.

Definition 4.5 (Extended tensor algebra)**.** *We define the extended tensor algebra over V, which sometimes is also called its formal series of tensors, as*

$$T((V)) := \{\boldsymbol{a} = (a_0, a_1, ...) \mid a_n \in V^{\otimes n} \textit{ for all } n \geq 0\}.$$

For any $n \geq 0$, we also define the canonical projection

$$\begin{array}{rccc} \pi_n : & T((V)) & \to & V^{\otimes n} \\ & (a_i)_{i=0}^{\infty} & \mapsto & a_n. \end{array}$$

The set given by e_I for all $|I| = n$, $n \geq 0$, is basis of $T((V))$. Therefore, any element $\mathbf{a} \in T((V))$ can be written as

$$\mathbf{a} = \sum_{|I| \geq 0} a_I e_I \quad \text{with } a_I \in \mathbb{R}.$$

Notice that $T((V))$ can be identified as the space of formal non-commuting power series in d indeterminates and real coefficients.

The extended tensor algebra can be endowed with a natural vector space structure, defining an addition and a product by scalars. For any $\mathbf{a} = \sum_{|I| \geq 0} a_I e_I, \mathbf{b} = \sum_{|I| \geq 0} b_I e_I \in T((V))$ and $\alpha, \beta \in \mathbb{R}$, we set

$$\alpha \mathbf{a} + \beta \mathbf{b} := \sum_{|I| \geq 0} (\alpha a_I + \beta b_I) e_I.$$

When we equip $T((V))$ with the tensor product $\otimes$, it becomes an algebra. This is defined as

$$\mathbf{a} \otimes \mathbf{b} := \sum_{n=0}^{\infty} \sum_{|I|+|J|=n} a_I b_J e_I \otimes e_J.$$

Observe that $(T((V)), +, \cdot, \otimes)$ is a real non-commutative algebra with neutral element $\mathbf{1} = (1, 0, \ldots, 0, \ldots)$. The invertible elements $\mathbf{a} = (a_i)_{i=0}^{\infty} \in T((V))$ are those that $a_0 \neq 0$, and their inverse is given by

$$\mathbf{a}^{-1} = \frac{1}{a_0} \sum_{n \geq 0} \left(\mathbf{1} - \frac{\mathbf{a}}{a_0}\right)^{\otimes n}.$$

Let us now define the subalgebra of $T((V))$ consisting of all sequences of tensors with finite length.

Definition 4.6 (Tensor algebra). *We define the tensor algebra over V as*

$$T(V) := \{\boldsymbol{a} = (a_0, a_1, \ldots) \in T((V)) \mid \exists N \in \mathbb{N} \text{ s.t. } a_n = 0 \text{ for all } n \geq N\}.$$

Finally, let us define the truncated tensor algebra over V, which is the linear subspace of $T((V))$ of all sequences of a given length.

Definition 4.7 (Truncated tensor algebra). *Let $N \geq 0$. We define the truncated tensor algebra of order N over V as*

$$T^N(V) := \{\boldsymbol{a} = (a_0, a_1, \ldots) \in T((V)) \mid a_n = 0 \text{ for all } n > N\}.$$

We also define the canonical projection

$$\begin{array}{rccc} \pi_{\leq N}\colon & T((V)) & \to & T^N(V) \\ & (a_i)_{i=0}^{\infty} & \mapsto & (a_i)_{i=0}^{N}. \end{array}$$

We equip $T^N(V)$ with the truncated tensor product $\otimes_{T^N(V)}$, defined as

$$\mathbf{a} \otimes_{T^N(V)} \mathbf{b} := \pi_{\leq N}(\mathbf{a} \otimes \mathbf{b})$$

for any $\mathbf{a}, \mathbf{b} \in T^N(V) \subset T((V))$. Notice that $(T^N(V), +, \cdot, \otimes_{T^N(V)})$ is a real non-commutative algebra with neutral element $\mathbf{1} = (1, 0, \ldots, 0) \in T^N(V)$. With some abuse of notation, whenever there is no confusion we will write $\otimes$ to denote the truncated tensor product between elements of $T^N(V)$.

Let us also remark that, since $V^{\otimes n}$ is a real vector space of dimension d^n, we have that $T^N(V)$ is a real vector space of dimension $\sum_{n=0}^{N} d^n$.

In the sequel, we will assume that the tensor powers over V are endowed with a norm that satisfies the following condition.

Definition 4.8. *We say that the tensor powers over V are endowed with admissible norms if the following conditions hold:*

1. *For each $n \geq 1$, the symmetric group S_n acts by isometries on $V^{\otimes n}$, i.e.*

$$\|\sigma v\|_{V^{\otimes n}} = \|v\|_{V^{\otimes n}} \quad \forall v \in V^{\otimes n}, \forall \sigma \in S_n.$$

2. *The tensor product has norm* 1*, i.e. for all* $n, m \geq 1$*,*

$$\|v \otimes w\|_{V^{\otimes(n+m)}} \leq \|v\|_{V^{\otimes n}} \|w\|_{V^{\otimes m}} \quad \forall v \in V^{\otimes n}, w \in V^{\otimes m}$$

An example of such a norm is given by

$$\|a\|_{V^{\otimes n}} := \sup_{|I|=n} |a_I|$$

for any $n \geq 1$ and $a = \sum_{|I|=n} a_I e_I \in V^{\otimes n}$. We will then equip $T(V)$ and $T^N(V)$ with the norms given by

$$\|\mathbf{a}\|_{T(V)} := \max_{n \geq 0} \|\pi_n(\mathbf{a})\|_{V^{\otimes n}}, \quad \|\mathbf{a}\|_{T^N(V)} := \max_{0 \leq n \leq N} \|\pi_n(\mathbf{a})\|_{V^{\otimes n}}.$$

4.3.1 Linear maps on the extended tensor algebra

Whenever one does analysis, one works hard to find a core of real functions on the space one is working on, and ideally this core should be an algebra. As we will see later, in our case this core will be given by the vector space generated by a set of linear maps on $T((V))$.

For any $\ell = \sum_{|I|\geq 0} \ell_I e_I \in T(V)$ and any $\mathbf{a} = \sum_{|I|\geq 0} a_I e_I \in T((V))$, we define the bilinear map $\langle \cdot, \cdot \rangle : T(V) \times T((V)) \to \mathbb{R}$ given by

$$\langle \ell, \mathbf{a} \rangle := \sum_{|I|\geq 0} \ell_I a_I. \tag{4.3}$$

Notice that every $\ell \in T(V)$ induces a real valued linear map on $T((V))$. Since ℓ only has a finite number of non-zero elements, we have that (4.3) is a finite sum, and therefore it is well-defined.

Observe also that in particular we have $a_I = \langle e_I, \mathbf{a} \rangle$, and so we can write

$$\mathbf{a} = \sum_{|I|\geq 0} \langle e_I, \mathbf{a} \rangle e_I.$$

Besides the sum and the tensor product, there is a third operation on $T(V)$ that will be very important in the following pages: the shuffle product $\sqcup\!\sqcup$.

Definition 4.9 (Shuffle product)**.** *For any two multi-indices* $I = (i_1, \ldots, i_n)$ *and* $J = (j_1, \ldots, j_m)$*, the shuffle product between* e_I *and* e_J *is defined recursively as*

$$e_I \sqcup\!\sqcup\, e_J := (e_{I'} \sqcup\!\sqcup\, e_J) \otimes e_{i_n} + (e_I \sqcup\!\sqcup\, e_{J'}) \otimes e_{j_m},$$

with $e_I \sqcup\!\sqcup\, e_\emptyset := e_\emptyset \sqcup\!\sqcup\, e_I = e_I$*. This extends bi-linearly to any* $\ell^1, \ell^2 \in T(V)$ *as*

$$\ell^1 \sqcup\!\sqcup\, \ell^2 = \sum_{|I|,|J|\geq 0} \ell^1_I \ell^2_J (e_I \sqcup\!\sqcup\, e_J).$$

It is easy to show that the shuffle product is associative and commutative, which means that $(T(V), +, \cdot, \sqcup\!\sqcup)$ is a commutative algebra.

4.4 The signature of a path of bounded variation

We are now ready to define the signature of a path of bounded variation.

Definition 4.10 (Signature of a path of bounded variation). *Let $0 \leq s \leq t \leq T$. Let $x : [0, T] \to V$ be a continuous path of bounded variation. We define the signature of x over $[s, t]$ by*

$$S(x)_{s,t} := (1, S(x)^1_{s,t}, \dots, S(x)^n_{s,t}, \dots) \in T((V))$$

with

$$S(x)^n_{s,t} := \int \cdots \int_{s<u_1<\cdots<u_n<t} dx_{u_1} \otimes \cdots \otimes dx_{u_n} = \int_s^t S(x)^{n-1}_{s,u} \otimes dx_u \in V^{\otimes n},$$

where the integration is in the Young-Stieltjes sense. For any $n \in \mathbb{N}$, $S(X)^n_{s,t}$ is called the n-th level of the signature of x over $[s, t]$. Similarly, we define the truncated signature of order N as

$$S(x)^{\leq N}_{s,t} := \pi_{\leq N}(S(x)_{s,t}) = (1, S(x)^1_{s,t}, \dots, S(x)^N_{s,t}) \in T^N(V).$$

An alternative way of defining the signature is by defining its components recursively, as

$$\langle e_\emptyset, S(x)_{s,t} \rangle := 1, \quad \langle e_I, S(x)_{s,t} \rangle := \int_s^t \langle e_{I'}, S(x)_{s,u} \rangle dx^{i_n}_u$$

where $I = (i_1, \dots, i_n)$, $I' = (i_1, \dots, i_{n-1})$ and $x^{i_n}_u$ indicates the i_n-th component of x evaluated at u.

Let us now state and prove a very important property of signatures of paths of bounded variation.

Theorem 4.3 (Chen's theorem). *Let $x : [0, T] \to V$ be a continuous path of bounded variation. Then, for any $0 \leq s \leq u \leq t \leq T$, we have that*

$$S(x)^{\leq N}_{s,t} = S(x)^{\leq N}_{s,u} \otimes S(x)^{\leq N}_{u,t} \tag{4.4}$$

holds for any $N \geq 0$. We remark that $\otimes$ is the tensor product truncated at level N. Notice that, since this holds for all N, we also have that

$$S(x)_{s,t} = S(x)_{s,u} \otimes S(x)_{u,t},$$

where now the tensor product $\otimes$ is in $T((V))$.

Proof. Let us prove it by induction on N. For $N = 0$, the theorem simply states that $1 = 1 \otimes 1$. Assume now that (4.4) holds for $N - 1$. Let's now prove that it holds for N.

First observe that in $T^N(V)$, due to the truncation up to level N, we have that

$$S(x)^{N-1}_{s,u} \otimes \int_u^t S(x)^{N-1}_{u,r} \otimes dx_r = S(x)^N_{s,u} \otimes \int_u^t S(x)^{N-1}_{u,r} \otimes dx_r.$$

Hence, using the induction hypothesis, splitting $S(x)^N_{s,t}$ at u, we get

$$\begin{aligned} S(x)^N_{s,t} &= \mathbf{1} + \int_s^t S(x)^{N-1}_{s,r} \otimes dx_r = \mathbf{1} + \int_s^u S(x)^{N-1}_{s,r} \otimes dx_r + \int_u^t S(x)^{N-1}_{s,r} \otimes dx_r \\ &= \mathbf{1} + \int_s^u S(x)^{N-1}_{s,r} \otimes dx_r + \int_u^t S(x)^{N-1}_{s,u} \otimes S(x)^{N-1}_{u,r} \otimes dx_r \\ &= S(x)^N_{s,u} + S(x)^N_{s,u} \otimes \int_u^t S(x)^{N-1}_{u,r} \otimes dx_r = S(x)^N_{s,u} \otimes (\mathbf{1} + (S(x)^N_{u,t} - \mathbf{1})) \\ &= S(x)^N_{s,u} \otimes S(x)^N_{u,t}. \end{aligned}$$

□

By definition, the signature of a path of bounded variation is an element of $T((V))$. However, at first glance it is not clear in what subspace of $T((V))$ it lives in. It turns out that signatures take values in a certain group called the space of *group-like* elements.

Definition 4.11 (Group-like element). *An element $\boldsymbol{a} \in T((V)) \setminus \{0\}$ is said to be group-like if, given any $\ell_1, \ell_2 \in T(V)$, the following equality holds:*

$$\langle \ell_1, \boldsymbol{a} \rangle \langle \ell_2, \boldsymbol{a} \rangle = \langle \ell_1 \sqcup\!\sqcup \ell_2, \boldsymbol{a} \rangle.$$

The space of all group-like elements $\boldsymbol{a} \in T((V)) \setminus \{0\}$ is denoted by $G(V)$. We also define, for any $N \in \mathbb{N}$, the projection of $G(V)$ into $T^N(V)$,

$$G^N(V) := \pi_{\leq N}(G(V)) \subset T^N(V).$$

Proposition 4.2 (Shuffle property). *Let $x : [0, T] \to V$ be a continuous path of bounded variation. Then, for any $0 \leq s \leq t \leq T$, we have that $S(x)_{s,t}$ is group-like. That is, for any $\ell_1, \ell_2 \in T(V)$, we have that*

$$\langle \ell_1, S(x)_{s,t} \rangle \langle \ell_2, S(x)_{s,t} \rangle = \langle \ell_1 \sqcup\!\sqcup \ell_2, S(x)_{s,t} \rangle.$$

Proof. By the bi-linearity of the shuffle product, it is enough to prove it for any e_I, e_J, with I and J multi-indices.

We set $n := |I| + |J|$, and proceed by induction on this variable. For $n = 0$, we necessarily have $I = J = \emptyset$, and the result clearly holds.

Assume now that the result holds for some fixed n. Let's show that this implies that it also holds for $n + 1$. Let us write $I = (i_1, \ldots, i_{n+1-m})$ and $J = (j_1, \ldots, j_m)$.

Notice that for a fixed $s \in [0, T]$ we have

$$d\langle e_I, S(x)_{s,t} \rangle = \langle e_{I'}, S(x)_{s,t} \rangle dx_t^{i_{n+1-m}}, \quad d\langle e_J, S(x)_{s,t} \rangle = \langle e_{J'}, S(x)_{s,t} \rangle dx_t^{j_m}.$$

By using integration by parts, we arrive at

$$\begin{aligned} &\langle e_I, S(x)_{s,t} \rangle \langle e_J, S(x)_{s,t} \rangle \\ &\qquad = \int_s^t \langle e_{I'}, S(x)_{s,u} \rangle \langle e_J, S(x)_{s,u} \rangle dx_u^{i_{n+1-m}} + \int_s^t \langle e_I, S(x)_{s,u} \rangle \langle e_{J'}, S(x)_{s,u} \rangle dx_u^{j_m} \\ &\qquad = \langle (e_{I'} \sqcup\!\sqcup e_J) \otimes e_{i_{n+1-m}}, S(x)_{s,t} \rangle + \langle (e_I \sqcup\!\sqcup e_{J'}) \otimes e_{j_m}, S(x)_{s,t} \rangle \\ &\qquad = \langle e_I \sqcup\!\sqcup e_J, S(x)_{s,t} \rangle, \end{aligned}$$

where in the second equality we used the inductive hypothesis. □

Notice that an immediate consequence of the shuffle product property is that polynomials on signatures turn out to be linear functions on signatures. This will be exploited in Section 4.7.

4.5 Rough paths and their signatures

In the last section, we have defined the signature of a path of bounded variation and have also presented some of its most important properties. By using Theorem 4.1, we could easily extend the definition of the signature to continuous paths of finite p-variation, for $p < 2$. Moreover, in this case all the results that we stated hold true, with the proofs still being valid.

However, for paths with infinite p-variation for any $p < 2$, we would not be able to define its signature, since the iterated integrals would lack a meaning. With stochastic calculus in mind, we are mainly interested in this case. The goal of this section is to give a rigorous account of what it takes to define the signature of such irregular paths.

4.5.1 Multiplicative functionals

Set $\Delta_T := \{(s,t) \in [0,T]^2 \ : \ s \leq t\}$, and let $x : [0,T] \to V$ be a continuous path of bounded variation. For any $N \geq 0$, let us consider the map $S(x)^{\leq N} : \Delta_T \to T^N(V)$ given by taking the truncated signature of order N of x over each $(s,t) \in \Delta_T$.

Chen's theorem asserts that the map $S(x)^{\leq N}_{s,t}$ is *multiplicative*. That is, for all $0 \leq s \leq u \leq t \leq T$, one has $S(x)^{\leq N}_{s,u} \otimes S(x)^{\leq N}_{u,t} = S(x)^{\leq N}_{s,t}$.

In what follows, we will take as our basic object of study a general function from Δ_T into $T^N(V)$ which satisfies this multiplicative property.

Definition 4.12 (Multiplicative functional)**.** *Let $N \in \mathbb{N}$. Let $\boldsymbol{X} : \Delta_T \to T^N(V)$ be a continuous map. For each $(s,t) \in \Delta_T$, denote by $\boldsymbol{X}_{s,t}$ the image by $\boldsymbol{X}$ of (s,t) and write*

$$\boldsymbol{X}_{s,t} = (\boldsymbol{X}^0_{s,t}, \boldsymbol{X}^1_{s,t}, \ldots, \boldsymbol{X}^N_{s,t}) \in \mathbb{R} \oplus V \oplus \cdots \oplus V^{\otimes N}.$$

The function $\boldsymbol{X}$ is called a multiplicative functional of degree N in V if $\boldsymbol{X}^0_{s,t} = 1$ for all $(s,t) \in \Delta_T$ and

$$\boldsymbol{X}_{s,u} \otimes \boldsymbol{X}_{u,t} = \boldsymbol{X}_{s,t} \quad \forall s,u,t \in [0,T], \quad s \leq u \leq t, \tag{4.5}$$

where we remark that the tensor product $\otimes$ is truncated at level N.

In reference to Chen's theorem, the multiplicative property (4.5) is called Chen's identity. In order to get a better understanding of what this property means, let us first take a look at the important examples given by $N = 1$ and $N = 2$.

Example 4.1. *Consider a functional $\boldsymbol{X} : \Delta_T \to T^1(V) = \mathbb{R} \oplus V$ which is multiplicative. In this case, this translates to*

$$(1, \boldsymbol{X}^1_{s,t}) = (1, \boldsymbol{X}^1_{s,u}) \otimes (1, \boldsymbol{X}^1_{u,t}) = (1, \boldsymbol{X}^1_{s,u} + \boldsymbol{X}^1_{u,t})$$

for all $0 \le s \le u \le t \le T$. *Therefore, the multiplicative condition in* $T^1(V)$ *reduces to the additivity in* V *of the mapping* $(s,t) \mapsto \boldsymbol{X}^1_{s,t}$, *that is, to the identity* $\boldsymbol{X}^1_{s,t} = \boldsymbol{X}^1_{s,u} + \boldsymbol{X}^1_{u,t}$.

This identity is equivalent to the existence of a path $x : [0,T] \to V$ *such that* $\boldsymbol{X}^1_{s,t} = x_t - x_s$ *for all* $(s,t) \in \Delta_T$, *where the path* x *is uniquely determined by* $\boldsymbol{X}$ *up to the addition of a constant element of* V. *We therefore have that a multiplicative functional of degree* 1 *in* V *is just a path in* V *modulo translation, described through its increments.*

Example 4.2. *In the case* $N = 2$, $\boldsymbol{X} : \Delta_T \to T^2(V)$ *is a multiplicative functional of degree* 2 *if*

$$\begin{aligned} &\textit{(level 1)} \quad \boldsymbol{X}^1_{s,t} = \boldsymbol{X}^1_{s,u} + \boldsymbol{X}^1_{u,t}, \\ &\textit{(level 2)} \quad \boldsymbol{X}^2_{s,t} = \boldsymbol{X}^2_{s,u} + \boldsymbol{X}^2_{u,t} + \boldsymbol{X}^1_{s,u} \otimes \boldsymbol{X}^1_{u,t}. \end{aligned}$$

If $\mathbf{X}$ is a multiplicative functional of degree N in V, then $\pi_1 \circ \mathbf{X} : \Delta_T \to T^1(V)$ is a multiplicative functional of degree 1. Therefore, by the remark made earlier, there exists a classical path $x : [0,T] \to V$ which underlies $\mathbf{X}$ in the sense that $\mathbf{X}^1_{s,t} = x_t - x_s$ for all $(s,t) \in \Delta_T$.

However, we remark that $\mathbf{X}$ is not at all the signature of x in general. Even if x is of finite p-variation for some $p < 2$ (otherwise its signature is not well defined), $\mathbf{X}$ may be different from the signature of x. For instance, if w is any non-zero element in $V^{\otimes 2}$, then $\mathbf{X}_{s,t} := (1, 0, (t-s)w) \in T^2(V)$ is a multiplicative functional of degree 2. However, the path x determined by $\pi_1 \circ \mathbf{X}$ is a constant path and the signature of a constant path truncated at level 2 is simply $(1,0,0)$.

More generally, if $\psi : [0,T] \to V^{\otimes N}$ is any continuous function, then it is not hard to check that the functional $\mathbf{X}_{s,t} := (1, 0, \ldots, 0, \psi(t) - \psi(s)) \in T^N(V)$ is multiplicative. This is a particular instance of the following result.

Lemma 4.1. *Let* $N \ge 0$ *be an integer. Let* $\boldsymbol{X}, \boldsymbol{Y} : \Delta_T \to T^N(V)$ *be two multiplicative functionals which agree on the first* $N-1$ *levels. Then the difference function* $\Psi : \Delta_T \to V^{\otimes N}$ *defined by*

$$\Psi_{s,t} := \boldsymbol{X}^N_{s,t} - \boldsymbol{Y}^N_{s,t} \in V^{\otimes N}$$

is additive, which means that for all $0 \le s \le u \le t \le T$,

$$\Psi_{s,t} = \Psi_{s,u} + \Psi_{u,t}.$$

Conversely, if $\boldsymbol{X} : \Delta_T \to T^N(V)$ *is a multiplicative functional and if* $\Psi : \Delta_T \to V^{\otimes N}$ *is an additive function, then* $(s,t) \mapsto \boldsymbol{X}_{s,t} + \Psi_{s,t}$ *is still a multiplicative functional.*

So far, the only condition imposed on a multiplicative functional has been Chen's identity, which is a purely algebraic condition. Let us now introduce the notion of p-variation for multiplicative functionals.

Definition 4.13. *Given a continuous functional* $\boldsymbol{X} : \Delta_T \to T^N(V)$, *we define its* p*-variation as*

$$\|\boldsymbol{X}\|_{p\text{-}var} := \max_{1 \le k \le N} \sup_{\mathcal{D} \subset [0,T]} \left(\sum_{t_i \in \mathcal{D}} \|\pi_k(\boldsymbol{X}_{t_i, t_{i+1}})\|^{p/k}_{V^{\otimes k}} \right)^{k/p},$$

where the supremum is taken over all partitions $\mathcal{D}$ of $[0, T]$. We say that $\mathbf{X}$ has finite p-variation if $\|\mathbf{X}\|_{p\text{-}var} < \infty$.

Moreover, given two multiplicative functionals $\mathbf{X}, \mathbf{Y} : \Delta_T \to T^N(V)$ with finite p-variation, we define their p-variation distance as

$$d_{p\text{-}var}(\mathbf{X}, \mathbf{Y}) := \|\mathbf{X} - \mathbf{Y}\|_{p\text{-}var}.$$

Lemma 4.1 tells us that, whenever two multiplicative functionals $\mathbf{X}, \mathbf{Y} : \Delta_T \to T^N(V)$ agree on the first $N-1$ levels, their difference is an additive function $\Psi_{s,t} : \Delta_T \to V^{\otimes N}$, which can be written as $\Psi_{s,t} = \psi(t) - \psi(s)$ for some $\psi : [0, T] \to V^{\otimes N}$.

If we impose $\mathbf{X}$ and $\mathbf{Y}$ to be of finite p-variation, this would imply that ψ is of finite p/N in the sense of Definition 4.2. By Proposition 4.1 we would have that there is a reparameterization τ such that $\psi \circ \tau$ is N/p-Hölder continuous. If we select N such that $N/p > 1$, this would necessarily mean that ψ is constant, concluding that a multiplicative functional with finite p-variation is completely determined by its first $\lfloor p \rfloor$ levels, where $\lfloor p \rfloor$ is the integer part of p.

This short argument leads us to the following fundamental theorem.

Theorem 4.4 (Extension theorem)**.** *Let $p \geq 1$ be a real number and $N \geq 1$ an integer. Let $\boldsymbol{X} : \Delta_T \to T^N(V)$ be a multiplicative functional with finite p-variation. Assume that $N \geq \lfloor p \rfloor$. Then, for every $n > N$, there exists a unique continuous multiplicative functional of degree N, which we denote by $\boldsymbol{Y} : \Delta_T \to T^N(V)$, such that:*

1. *$\boldsymbol{Y}$ is an extension of $\boldsymbol{X}$, i.e. $\pi_{\leq N}(\boldsymbol{Y}) = \boldsymbol{X}$.*
2. *$\boldsymbol{Y}$ has finite p-variation.*

Moreover, the extension map thus defined is continuous under the p-variation distance.

Proof. See [LCL04], Theorems 3.7 and 3.10. □

Note that a particular example of this theorem is given by the truncated signature of order N of a path of bounded variation x, which is a multiplicative functional of degree N completely determined by its first term, $x_t - x_s$. Again, the same holds true for a path of finite p-variation, with $p < 2$.

Another very interesting example is given by the signature of a trajectory of a Brownian motion $B : [0, T] \to V$. Consider for example its Stratonovich signature truncated at level 2, defined as the map $S^{\text{Strat}}(B)^{\leq 2} : \Delta_T \to T^2(V)$ given by

$$S^{\text{Strat}}(B)^{\leq 2}_{s,t} := \left(1, B_t - B_s, \iint_{s<u_1<u_2<t} \circ dB_{u_1} \otimes \circ dB_{u_2}\right),$$

where $\circ$ indicates that the integrals are understood in the Stratonovich sense.

One can easily check that $S^{\text{Strat}}(B)^{\leq 2}$ is a multiplicative functional of degree 2. However, we could also consider the Itô signature, also truncated at level 2, defined as

$$S^{\text{Itô}}(B)^{\leq 2}_{s,t} := \left(1, B_t - B_s, \iint_{s<u_1<u_2<t} dB_{u_1} \otimes dB_{u_2}\right),$$

where now the integrals are understood in the Itô sense. Again, one can easily check that $S^{\text{Itô}}(B)^{\leq 2}_{s,t}$ is a multiplicative functional of degree 2. Since both multiplicative functionals have finite p-variation if and only if $p > 2$, and their projection onto V is given by $B_t - B_s$, we conclude that the first term does not completely determine the higher order terms.

However, once we give a meaning to the second order iterated integrals (i.e., we choose Itô or Stratonovich), by the previous theorem we get uniqueness of the higher order levels, which are given by the iterated integrals in the sense that we chose.

As such, for a Brownian motion B we should think of these lifts to continuous maps from Δ_T to $T^2(V)$ as the fundamental objects, and reverse what is defined by what, so that $B_t - B_s$ is defined as the projection of the Stratonovich/Itô signature by π_1. The same can be said about a continuous path of finite p-variation and a multiplicative functional $\mathbf{X} : \Delta_T \to T^{\lfloor p \rfloor}(V)$ of finite p-variation such that $\pi_1(\mathbf{X}_{s,t}) = x_t - x_s$.

4.5.2 Spaces of rough paths

In the last section we explained that a multiplicative functional with finite p-variation is completely determined by its first $\lfloor p \rfloor$ terms. Therefore, the following definition is completely natural.

Definition 4.14 (Rough path). *Let $p \geq 1$. A p-rough path is a multiplicative functional of degree $\lfloor p \rfloor$ with finite p-variation. We denote the space of p-rough paths defined over Δ_T by $\Omega_T^p(V)$.*

We therefore have that a p-rough path is a continuous mapping from Δ_T to $T^{\lfloor p \rfloor}(V)$ which satisfies an algebraic condition, Chen's identity, and an analytic condition, the finiteness of its p-variation.

We remark that, due to the non-linearity of Chen's relation, we have that $(\Omega_T^p(V), +, \cdot)$ is not a vector space, since the sum of two multiplicative functionals is (in general) not a multiplicative functional. Nevertheless, the space of rough paths is complete under the p-variation metric (see [Lyo02, Section 3.3.1]).

Next, we define the space of geometric p-rough paths, essentially defined as the closure in $(\Omega_T^p(V), d_p)$ of truncated signatures of order $\lfloor p \rfloor$ of paths of bounded variation.

Definition 4.15 (Geometric p-rough path). *A geometric p-rough path is a p-rough path $\boldsymbol{X} \in \Omega_T^p(V)$ for which there exists a sequence of paths of bounded variation $\{x_n\}_{n\geq 0}$ such that*

$$\lim_{n\to\infty} d_{p\text{-}var}(\boldsymbol{X}, S(x_n)^{\leq \lfloor p \rfloor}) = 0,$$

where $S(x_n)^{\leq \lfloor p \rfloor}$ denotes the truncated signature of order $\lfloor p \rfloor$ of x_n. The space of geometric p-rough paths defined over Δ_T is denoted by $G\Omega_T^p(V)$.

Recall that the signature of a path of bounded variation, truncated at level N, takes its values in the space of group-like elements, which is denoted by $G^N(V)$ (see Proposition 4.2). Since $G^N(V)$ is defined by algebraic identities, which are related to the shuffle product, it still holds true that a geometric p-rough path takes its values in $G^{\lfloor p \rfloor}(V)$.

Unfortunately, this property does not completely characterise geometric rough paths among all rough paths, see [FV06]. This makes the following definition necessary.

Definition 4.16 (Weakly geometric p-rough path)**.** *A weakly geometric p-rough path is a p-rough path which takes its values in $G^{\lfloor p \rfloor}(V)$. The space of weakly geometric p-rough paths defined over Δ_T is denoted by $\mathcal{W}G\Omega_T^p(V)$.*

We then have the following set of strict inclusions:

$$G\Omega_T^p(V) \subset \mathcal{W}G\Omega_T^p(V) \subset \Omega_T^p(V).$$

As it is explained in [LCL04, Section 3.2.2] and [FV06], the difference between geometric and weakly geometric rough paths is generally insignificant, and it could be compared to the difference between C^1 and Lipschitz functions. In fact, we have the following interesting result.

Proposition 4.3. *Let $p > 1$. Then, for any $1 < p < q$, we have that*

$$\mathcal{W}G\Omega_T^p(V) \subset G\Omega_T^q(V).$$

Proof. See [FV10, Corollary 8.24]. □

4.5.3 Examples of weakly geometric p-rough paths

In our case, we shall be interested in stochastic processes whose trajectories can be lifted to weakly geometric p-rough paths almost surely. In this section we will go over a few of them that will be useful in the sequel.

The most widely used stochastic integral with respect to a Brownian motion is the Itô integral. Let us start this section by showing that the resulting signature is not a weakly geometric p-rough path for any $p > 2$, as it doesn't take its values in the space of group-like elements $G^2(V)$.

Consider a d-dimensional Brownian motion $B = (B_t)_{t\in[0,T]}$, and denote its Itô signature truncated at level 2 as $S^{\text{Itô}}(B)^{\leq 2}$. Let us show that

$$\langle e_1, S^{\text{Itô}}(B)_{s,t}^{\leq 2}\rangle \langle e_1, S^{\text{Itô}}(B)_{s,t}^{\leq 2}\rangle \neq \langle e_1 \sqcup\!\sqcup\, e_1, S^{\text{Itô}}(B)_{s,t}^{\leq 2}\rangle.$$

First, we have that

$$\langle e_1, S^{\text{Itô}}(B)_{s,t}^{\leq 2}\rangle \langle e_1, S^{\text{Itô}}(B)_{s,t}^{\leq 2}\rangle = (B_t - B_s)^2.$$

Since $e_1 \sqcup\!\sqcup\, e_1 = 2e_{11}$, using Itô's formula we have that

$$\langle e_1 \sqcup\!\sqcup\, e_1, S^{\text{Itô}}(B)_{s,t}^{\leq 2}\rangle = 2\langle e_{11}, S^{\text{Itô}}(B)_{s,t}^{\leq 2}\rangle = 2\int_s^t (B_u - B_s)dB_u = (B_t - B_s)^2 - (t-s).$$

Thus, we conclude that the Itô signature of a Brownian motion does not take its values in $G^2(V)$, and therefore is not a weakly geometric p-rough path.

The situation is different if we consider the Stratonovich signature of a Brownian motion. Indeed, since the Stratonovich integration by parts rule is the same as the Young-Stieltjes one, we can use the same proof as in Proposition 4.2 to prove that the truncated Stratonovich signature of a Brownian motion takes its values in $G^2(V)$.

Since Stratonovich integration is well defined for any continuous semimartingale X, we arrive at the following result.

Theorem 4.5. *The truncated Stratonovich signature of order* 2 *of a multidimensional continuous semimartingale $X = (X_t)_{t\in[0,T]}$, defined like*

$$S(X)^{\leq 2}_{s,t} := \left(1, X_t - X_s, \iint_{s<u_1<u_2<t} \circ dX_{u_1} \otimes \circ dX_{u_2}\right), \tag{4.6}$$

is almost surely a weakly geometric p-rough path for any $p \in (2,3)$.

Another example of a stochastic process whose trajectories can almost surely be lifted to a weakly geometric rough path is given by a multidimensional fractional Brownian motion with Hurst parameter $H > 1/4$.

Theorem 4.6. *Let $B^H = (B^H_t)_{t\in[0,T]}$ be a d-dimensional fractional Brownian motion with Hurst parameter $H \in (1/4, 1)$. Then, for any $p \in (1/H, 4)$, there exists an almost surely weakly geometric p-rough path $\boldsymbol{B}^H : \Delta_T \to T^{\lfloor p\rfloor}(V)$ such that $\pi_1(\boldsymbol{B}^H) = B^H_t - B^H_s$.*

Proof. For $H > 1/2$, we have that the trajectories are almost surely of finite p-variation for $p < 1/H < 2$, which means that the weakly geometric p-rough path is simply given by the increments of the path.

For $H = 1/2$, we recover the Brownian motion case.

For $H \in (1/4, 1/2)$, we refer to [CQ02, Theorem 2], where they actually prove that B^H can be lifted almost surely to a geometric p-rough path, which in particular is a weakly geometric p-rough path. □

With rough volatility models in mind, the restriction $H > 1/4$ on the previous proposition is a great limitation for our purposes in the next chapter, where we will want to be able to compute the signature of a weakly geometric p-rough path whose projection into the first component is a fBm with $H < 1/4$.

Luckily for us, the next proposition tells us that if the process is one-dimensional, there is a trivial lift to a weakly geometric p-rough path.

Proposition 4.4. *Let $p \geq 1$. Let $x : [0,T] \to \mathbb{R}$ be a continuous path of finite p-variation. Then, there is a trivial lift to a weakly geometric p-rough path given by*

$$\mathbf{X}_{s,t} = \left(1, (x_t - x_s), \frac{1}{2}(x_t - x_s)^2, \ldots, \frac{1}{\lfloor p\rfloor!}(x_t - x_s)^{\lfloor p\rfloor}\right).$$

The following result will be very useful in the upcoming sections.

Proposition 4.5. *Let $x : [0,T] \to V$ be a continuous path that can be lifted to a weakly geometric p-rough path $\mathbf{X} \in \mathcal{WG}\Omega_p(V)$. Then, the path $\widehat{x} : [0,T] \to \mathbb{R} \oplus V$ given by $\widehat{x}_t = (t, x_t)$ can also be lifted to a weakly geometric p-rough path, which we denote by $\widehat{\mathbf{X}}$.*

Proof. This result is an immediate consequence of [FV10, Theorem 9.26]. Intuitively speaking, we can make use of the abundant regularity of the first component of $\widehat{x}$ in order to define the new terms that we are introducing in $\widehat{\mathbf{X}}$ (that is, the ones that are not already in $\mathbf{X}$), by imposing that they satisfy the shuffle property.

Denote by $0, \ldots, d$ the indices of $\widehat{x}$, where 0 corresponds to the time component.

Let us define the new terms inductively on the length of the multi-index I. For $|I| = 1$, we trivially have that the first level is already well-defined.

Assume now that we have defined all the levels up until the $(n-1)$-th, and that they satisfy the shuffle property. We can then define the n-th level as follows.

Let $I = (i_1, \dots, i_n)$ be a multi-index with $i_k \in \{0, \dots, d\}$. Assume that there is some $i_k = 0$ (otherwise, $\langle e_I, \widehat{\mathbf{X}} \rangle$ is already in $\mathbf{X}$ and it is therefore well-defined).

Let $I_1 := (i_1, \dots, i_k)$ and $I_2 := (i_{k+1}, \dots, i_n)$ be such that $i_k = 0$ and there is no other 0 in I_2. Notice that, by definition of shuffle product, we have

$$e_I = e_{I'} \sqcup\!\sqcup\, e_{i_n} - (e_{I''} \sqcup\!\sqcup\, e_{i_n}) \otimes e_{i_{n-1}}.$$

Now, we can impose the shuffle property and define

$$\langle e_I, \widehat{\mathbf{X}}_{s,t} \rangle := \langle e_{I'}, \widehat{\mathbf{X}}_{s,t} \rangle \langle e_{i_n}, \widehat{\mathbf{X}} \rangle - \langle (e_{I''} \sqcup\!\sqcup\, e_{i_n}) \otimes e_{i_{n-1}}, \widehat{\mathbf{X}}_{s,t} \rangle.$$

The first term is well-defined by induction. If $n-1 = k$ and thus $i_{n-1} = 0$, then the second term is also well-defined as

$$\langle (e_{I''} \sqcup\!\sqcup\, e_{i_n}) \otimes e_{i_{n-1}}, \widehat{\mathbf{X}}_{s,t} \rangle := \int_s^t \langle e_{I''} \sqcup\!\sqcup\, e_{i_n}, \widehat{\mathbf{X}}_{s,u} \rangle du. \tag{4.7}$$

If $i_{n-1} \neq 0$, we can make use of the following identity,

$$(e_{I''} \sqcup\!\sqcup\, e_{i_n}) \otimes e_{i_{n-1}} = e_{I''} \sqcup\!\sqcup\, (e_{i_n} \otimes e_{i_{n-1}}) - (e_{I'''} \sqcup\!\sqcup\, (e_{i_n} \otimes e_{i_{n-1}})) \otimes e_{i_{n-2}},$$

obtained also by using the definition of the shuffle product. Again, we can define

$$\langle (e_{I''} \sqcup\!\sqcup\, e_{i_n}) \otimes e_{i_{n-1}}, \widehat{\mathbf{X}}_{s,t} \rangle := \langle e_{I''}, \widehat{\mathbf{X}}_{s,t} \rangle \langle e_{i_n} \otimes e_{i_{n-1}}, \widehat{\mathbf{X}} \rangle - \langle (e_{I'''} \sqcup\!\sqcup\, (e_{i_n} \otimes e_{i_{n-1}})) \otimes e_{i_{n-2}}, \widehat{\mathbf{X}}_{s,t} \rangle.$$

The first term is well-defined by induction. If $n-2 = k$ and thus $i_{n-2} = 0$, then the second term is also well-defined, in a similar way as (4.7). Otherwise, we can iterate the process until we arrive at i_k.

Using this reasoning, we can express any component of the signature of $\widehat{\mathbf{X}}$ as the sum of products of already well-defined terms and a finite-variation integral. □

Since these time augmented weakly geometric p-rough paths will be very important later, let us formally define them and set some notation.

Definition 4.17. *A time augmented weakly geometric p-rough path is a weakly geometric p-rough path $\mathbf{X} \in \mathcal{W}G\Omega_T^p(\mathbb{R} \oplus V)$ such that the projection into the first level is given by the increments of a time augmented path, i.e.*

$$\pi_1(\mathbf{X}_{s,t}) = \widehat{x}_t - \widehat{x}_s, \quad \textit{where } \widehat{x}_t := (t, x_t).$$

The space of time augmented weakly geometric p-rough paths in V is denoted by $\mathcal{W}G\widehat{\Omega}_T^p(V)$.

Notice that, by the previous proposition, any weakly geometric p-rough path can be identified with a time augmented weakly geometric p-rough path.

4.6 The signature of a p-rough path

Recall that, by the Extension Theorem 4.4, every p-rough path $\mathbf{X} \in \Omega_T^p(V)$ can be extended to a multiplicative functional of arbitrarily high degree with finite p-variation. We make use of such extensions to define the full signature of a p-rough path.

Definition 4.18 (Signature of a p-rough path)**.** *Let $\boldsymbol{X} : \Delta_T \to T^{\lfloor p \rfloor}(V)$ be a p-rough path. We define the truncated signature of order $N \geq \lfloor p \rfloor$ of $\boldsymbol{X}$ as the unique extension*

$$S(\boldsymbol{X})^{\leq N} := (1, S(\boldsymbol{X})^1, \dots, S(\boldsymbol{X})^N) : \Delta_T \to T^N(V)$$

given by the Extension Theorem 4.4. Likewise, we refer to

$$S(\boldsymbol{X}) := (1, S(\boldsymbol{X})^1, \dots, S(\boldsymbol{X})^n, \dots) : \Delta_T \to T((V))$$

as the signature of $\boldsymbol{X}$.

Remark 4.1. *Consider a weakly geometric p-rough path $\boldsymbol{X}$, and let $q > p$. By Proposition 4.3, there exists a sequence of paths of bounded variation such that their truncated signatures of order $\lfloor q \rfloor$ converge to $\boldsymbol{X}$ in the q-variation metric.*

Since the Extension Theorem 4.4 tells us that the signature map is continuous in the q-variation metric, we can conclude that the signature of a weakly geometric p-rough path is group-like. That is,

$$\langle \ell_1, S(\boldsymbol{X})_{s,t} \rangle \langle \ell_2, S(\boldsymbol{X})_{s,t} \rangle = \langle \ell_1 \sqcup\!\sqcup \ell_2, S(\boldsymbol{X})_{s,t} \rangle.$$

for any $\ell_1, \ell_2 \in T(V)$.

4.7 Universal approximation theorems

From now on it will be useful to think of a p-rough path as a path with domain $[0, T]$, rather than Δ_T. The reason we are allowed to do this is because, by Chen's relation, any multiplicative functional $\mathbf{X} : \Delta_T \to T^N(V)$ is completely determined by the path from $[0, T]$ to $T^N(V)$ given by $\mathbf{X}_{0,t}$, since

$$\mathbf{X}_{s,t} = (\mathbf{X}_{0,s})^{-1} \otimes \mathbf{X}_{0,t}.$$

We will therefore write $\mathbf{X}_t := \mathbf{X}_{0,t}$.

The goal of this final section is to present a couple of universal approximation theorems that will justify why signatures of weakly geometric p-rough paths are so useful and mathematically well-founded to use in modelling problems. In a nutshell, these state that one can use signatures to approximate arbitrarily well non-linear continuous maps by linear ones.

4.7.1 Uniqueness of the signature

Let us start by proving that the signature of a time augmented weakly geometric p-rough path $\mathbf{X}$, evaluated at the final time T, completely determines $\mathbf{X}$.

Theorem 4.7 (Uniqueness of the signature)**.** *Let $\boldsymbol{X}, \boldsymbol{Y} \in \mathcal{W}G\widehat{\Omega}^p_T(V)$. We then have that*

$$S(\boldsymbol{X})_T = S(\boldsymbol{Y})_T \iff \boldsymbol{X}_t = \boldsymbol{Y}_t \quad \forall t \in [0, T].$$

Proof. ($\Leftarrow$) This implication is a consequence of the uniqueness of the extension in Theorem 4.4.

($\Rightarrow$) We need to prove that

$$S(\mathbf{X})_T = S(\mathbf{Y})_T \implies \langle \mathbf{X}_t, e_I \rangle = \langle \mathbf{Y}_t, e_I \rangle \quad \forall |I| \leq \lfloor p \rfloor, \forall t \in [0, T].$$

By the shuffle property, we have that

$$\langle \big(e_I \sqcup\!\sqcup (e_0^{\otimes k})\big) \otimes e_0, S(\mathbf{X})_T \rangle = \int_0^T \langle e_I, \mathbf{X}_t \rangle \frac{t^k}{k!} dt.$$

The proof follows from the fact that for a continuous function $f : [0, T] \to \mathbb{R}$ with $f(0) = 0$ and $\int_0^T f(t) t^k ds = 0$ for all $n \in \mathbb{N}$, it holds that $f \equiv 0$. □

4.7.2 First universal approximation theorem

We are now ready to state and prove the first universal approximation theorem (First UAT).

Theorem 4.8 (First UAT)**.** *Let $K \subset \mathcal{W}G\widehat{\Omega}^p_T(V)$ be a compact subset, and let us consider a continuous*[1] *map $f : \mathcal{W}G\widehat{\Omega}^p_T(V) \to \mathbb{R}$. Then, for any $\epsilon > 0$ there exists some $\ell \in T(\mathbb{R} \oplus V)$ such that*

$$\sup_{\boldsymbol{X} \in K} |f(\boldsymbol{X}) - \langle \ell, S(\boldsymbol{X})_T \rangle| < \epsilon.$$

Proof. The result follows by applying the Stone-Weierstrass theorem. To check the needed properties, observe that by the shuffle property, the set

$$A := \operatorname{span}\{\mathbf{X} \mapsto \langle e_I, S(\mathbf{X})_T \rangle : I \in \{0, \dots, d\}^{|I|}\}$$

is an algebra of maps from K to $\mathbb{R}$, which by the Extension Theorem 4.4 are continuous with respect to the p-variation distance. Choosing $I = \emptyset$ we get that A is vanishing nowhere, and so we only need to check that it is point separating in $\mathcal{W}G\widehat{\Omega}^p_T(V)$. Theorem 4.7 yields the result. □

4.7.3 Second universal approximation theorem

In our case we will be interested in modelling a real one-dimensional stochastic process $Y = (Y_t)_{t \in [0,T]}$ like

$$Y_t = f((\mathbf{X}_s)_{s \in [0,t]}),$$

where $\mathbf{X} = (\mathbf{X}_t)_{t \in [0,T]}$ is a known stochastic process, with the trajectories being almost surely in $\mathcal{W}G\widehat{\Omega}^p_T(V)$, and where f is an unknown map. Notice that f is defined for weakly geometric p-rough paths with domains over different sub-intervals

[1] Compactness and continuity are defined with respect to the p-variation distance.

$[0,t] \subset [0,T]$. We therefore need to define such space and equip it with an appropriate topology.

Following [Bay+23] and [KLA20], we will use a distance motivated by Dupire's functional Itô calculus, see [Dup19]. In a nutshell, this means that when we compare a path x defined on $[0,s]$ and another path y defined on $[0,t]$, with $0 \le s \le t$, we will extend x to $[0,t]$ by defining $x_u := x_s$ for $s \le u \le t$. However, recall that in our case we are considering paths $u \mapsto (u, x_u)$, and extending the time component of such a path in a constant way does not make much sense. We will therefore apply the previous extension to the x-component, but use instead the linear extension for the time component.

We use a similar construction for time augmented weakly geometric p-rough paths. For any $\mathbf{X}|_{[0,s]} \in \mathcal{W}G\widehat{\Omega}^p_s(V)$ and $s \le t$, we will extend it to $[0,t]$ by considering the weakly geometric p-rough path

$$\mathbf{X}|_{[0,t]}(u) := \begin{cases} \mathbf{X}|_{[0,s]}(u) & \text{if } u \in [0,s] \\ \mathbf{X}|_{[0,s]}(s) & \text{if } u \in [s,t], \end{cases}$$

and applying the time augmentation given by Proposition 4.5. We will denote such extension by $\widetilde{\mathbf{X}}|_{[0,t]}$.

By construction, we have that $\widetilde{\mathbf{X}}|_{[0,t]}$ and $\mathbf{X}|_{[0,s]}$ coincide in $[0,s]$, which motivates the following definition.

Definition 4.19 (Space of stopped p-rough paths). *For $p \ge 1$, we set*

$$\Lambda^p_T(V) := \bigcup_{t\in[0,T]} \mathcal{W}G\widehat{\Omega}^p_t(V),$$

and call it the space of weakly geometric stopped p-rough paths. We equip it with the metric

$$d(\mathbf{X}|_{[0,t]}, \mathbf{Y}|_{[0,s]}) := d_{p\text{-var};[0,t]}(\mathbf{X}|_{[0,t]}, \widetilde{\mathbf{Y}}|_{[0,t]}) + |t-s|,$$

where we assume $s \le t$ and $\widetilde{\mathbf{Y}}|_{[0,t]}$ is the stopped rough path constructed as explained above.

We are now ready to state the second universal approximation theorem.

Theorem 4.9 (Second UAT). *Let $K \subset \mathcal{W}G\widehat{\Omega}^p_T(V)$ be a compact subset, and consider a continuous map $f : \Lambda^p_T(V) \to \mathbb{R}$. Then, for any $\epsilon > 0$ there exists some $\ell \in T(\mathbb{R} \oplus V)$ such that*

$$\sup_{\mathbf{X}\in K; t\in[0,T]} |f(\mathbf{X}|_{[0,t]}) - \langle \ell, S(\mathbf{X})_t\rangle| < \epsilon,$$

where $\mathbf{X}|_{[0,t]}$ denotes $\mathbf{X}$ restricted to $[0,t]$.

Proof. One can easily check that the map $\varphi : [0,T] \times \mathcal{W}G\widehat{\Omega}^p_T(V) \to \Lambda^p_T(V)$ given by $\varphi(t, \mathbf{X}) = \mathbf{X}|_{[0,t]}$ is continuous. In particular, this implies that the image of $[0,T] \times K \subset [0,T] \times \mathcal{W}G\widehat{\Omega}^p_T(V)$ is compact in $\Lambda^p_T(V)$.

The proof then follows by an application of the Stone-Weierstrass theorem, in a similar way as it was done in Theorem 4.8. $\square$

Chapter 5

A signature based model for volatility

In this last chapter we use the signature transform to model the volatility process of a market model, where the dynamics of the discounted price process of the underlying under the risk-neutral probability measure are given by

$$d\tilde{S}_t = \tilde{S}_t \sigma_t dB_t. \tag{5.1}$$

5.1 The model

Let $Y = (Y_t)_{t\in[0,T]}$ be a real stochastic process obtained by

$$Y_t = f((\mathbf{X}_s)_{s\in[0,t]}),$$

where $\mathbf{X} = (\mathbf{X}_t)_{t\in[0,T]}$ is a known stochastic process with the trajectories being almost surely in $\mathcal{W}G\widehat{\Omega}^p_T$, and f is an unknown map. Recall that f is defined over the space of weakly geometric stopped p-rough paths, which we denote by $\Lambda^p_T(V)$.

As we saw in Theorem 4.9, by equipping $\Lambda^p_T(V)$ with the appropriate topology, we have that linear functions composed with signature transform are able to approximate any such continuous map f over compact subsets.

Motivated by this result and following [Cuc+23], we propose to model the volatility $\sigma = (\sigma_t)_{t\in[0,T]}$ in (5.1) as a linear function of the signature of a primary underlying process $\mathbf{X}$. That is,

$$\begin{aligned} d\tilde{S}_t(\ell) &= \tilde{S}_t(\ell)\sigma_t(\ell)dB_t \\ \sigma_t(\ell) &:= \ell_\emptyset + \sum_{0<|I|\le N} \ell_I \langle e_I, S(\mathbf{X})_t\rangle, \end{aligned} \tag{5.2}$$

where

- $(S(\mathbf{X})_t)_{t\in[0,T]}$ is the signature of $\mathbf{X} = (\mathbf{X}_t)_{t\in[0,T]}$,
- $\ell := \{\ell_I \in \mathbb{R} \mid |I| \le N\}$ denotes the collection of parameters of the model.

However, we remark that, contrary to [Cuc+23], we don't impose $\mathbf{X}$ to be a semimartingale, which means that we include a broader class of models. For example, following rough volatility models, $\mathbf{X}$ can be the lift of a time-extended one dimensional fractional Brownian motion with Hurst parameter $H \in (0, 1/2)$.

5.2 Calibration of any model to market prices

Let us now explain how one would calibrate a model in practice, which will be useful in understanding the advantages of our approach in the next section.

In practice, we are given a set of market prices of call options with different strikes and maturities. We then choose a parametric family of market models, all under which the generated discounted price process $\tilde{S}$ is a local martingale. The goal is to find the parameters that make the generated prices match as well as possible the real ones.

More specifically, suppose that we have a parametric family of models for the volatility, which we denote by $\sigma^\theta := (\sigma_t^\theta)_{t\in[0,T]}$, with θ being the parameters, and let

$$d\tilde{S}_t^\theta = \tilde{S}_t^\theta \sigma_t^\theta dB_t \tag{5.3}$$

be the corresponding discounted price process under the risk neutral probability measure $\mathbb{P}^*$. Notice that we are making explicit the dependence of S on θ. Assume also that we are given prices of n call options

$$C^{\text{market}}(T_1, K_1), \dots, C^{\text{market}}(T_n, K_n),$$

with maturities T_i and strikes K_i. Then, following [CGS22], we translate the calibration task to finding the parameters θ such that

$$\theta \in \arg\min_\theta L(\theta),$$

with

$$L(\theta) = \sum_{i=1}^n \gamma_i \big(C^{\text{market}}(T_i, K_i) - C^{\text{model}}(T_i, K_i, \theta)\big)^2, \tag{5.4}$$

where γ_i are some pre-specified weights and $C^{\text{model}}(T_i, K_i, \theta)$ denotes the price of a call option with strike K_i and maturity time T_i under our model with parameters θ. Recall that this is computed as

$$C^{\text{model}}(T, K, \theta) = \mathbb{E}_{\mathbb{P}^*}(e^{-rT}(S_T^\theta - K)_+), \tag{5.5}$$

where $\mathbb{P}^*$ is the risk-neutral probability measure.

The loss function given by (5.4) is minimized using some iterative algorithm, like for example gradient descent. Therefore, it is very important to have a fast method to compute (or at least approximate) the expectation given by (5.5) for any set of parameters θ.

The most straightforward approach to approximate (5.5) would be to perform a Monte Carlo simulation. However, in most models this is highly inefficient, since one needs to use a numerical SDE solver to integrate the solution from $t = 0$ to $t = T$ every time one wants to compute (5.5) for different θ. This is the reason why for many models it is very convenient to have a closed (or semi-closed) formula, like in the Heston model.

Nevertheless, in the next section we will see that in our model we can take advantage of the linearity of the volatility process to express $\log S_T(\ell)$ as the sum of a quadratic function and a linear one. This way we are able to compute all the necessary samples in the Monte Carlo procedure in an offline manner (that is, before starting to optimize the parameters).

5.3 Calibration of our model to market prices

As explained in the last section, in order to calibrate the model to a given set of market prices of call options, we need to minimize

$$L(\ell) = \sum_{i=1}^{n} \gamma_i \big(C^{\text{market}}(T_i, K_i) - C^{\text{model}}(T_i, K_i, \ell)\big)^2, \tag{5.6}$$

where

$$C^{\text{model}}(T, K, \ell) = \mathbb{E}_{\mathbb{P}^*}(e^{-rT}(e^{rT}\tilde{S}_T(\ell) - K)_+), \tag{5.7}$$

with $\mathbb{P}^*$ being the risk-neutral probability measure.

In the next pages it will be useful to think of tensors in $T^N(V)$ like elements in $\mathbb{R}^{d_N}$, where $d_N := \sum_{k=0}^{N} d^k$. To that end, let us introduce some helpful notation.

Let $\mathcal{L} : \{I : |I| \leq N\} \to \{1, \ldots, d_N\}$ be a bijective labeling function. Let $\mathbf{vec}$: $T^N(V) \to \mathbb{R}^{d_N}$ be the map given by

$$\mathbf{vec}(\mathbf{a}) = (a_{\mathcal{L}^{-1}(1)}, \ldots, a_{\mathcal{L}^{-1}(d_N)}) \in \mathbb{R}^{d_N}$$

for any $\mathbf{a} = \sum_{|I|\leq N} a_I e_I$. We will also think of ℓ as a vector in $\mathbb{R}^{d_N}$, without further specification.

Proposition 5.1. *Let $\tilde{S}(\ell) := (\tilde{S}_t(\ell))_{t\in[0,T]}$ be the discounted price process given by (5.2). Then, we have that*

$$\tilde{S}_t(\ell) = S_0 \exp\Big(\ell' Q(t)\ell + \ell' \int_0^t \mathbf{vec}(S(\mathbf{X})_t^{\leq N}) dB_s\Big), \tag{5.8}$$

where ℓ' denotes the transpose of ℓ, and where the matrix $Q(t)$ is given by

$$Q_{\mathcal{L}(I)\mathcal{L}(J)}(t) := -\frac{1}{2}\langle (e_I \sqcup\!\sqcup e_J) \otimes e_0, S(\mathbf{X})_t^{\leq N}\rangle.$$

Proof. Recall that $\tilde{S}(\ell) := (\tilde{S}_t(\ell))_{t\in[0,T]}$ is given by

$$d\tilde{S}_t(\ell) = \tilde{S}_t(\ell)\sigma_t(\ell)dB_t.$$

By applying the Itô formula to the function $f(x) = \log x$, which is twice continuously differentiable in the domain of S_t given by $(0, +\infty)$, we have that

$$\begin{aligned} d\log\tilde{S}_t(\ell) &= -\frac{1}{2\tilde{S}_t(\ell)^2}(\tilde{S}_t(\ell)\sigma_t(\ell))^2 dt + \frac{1}{\tilde{S}_t(\ell)}\tilde{S}_t(\ell)\sigma_t(\ell)dB_t \\ &= -\frac{1}{2}\sigma_t(\ell)^2 dt + \sigma_t(\ell)dB_t. \end{aligned}$$

By integrating and applying the exponential function on both sides, we arrive at

$$\tilde{S}_t(\ell) = S_0 \exp\Big(-\frac{1}{2}\int_0^t \sigma_s(\ell)^2 ds + \int_0^t \sigma_s(\ell) dB_s\Big).$$

Recall that $\sigma_t(\ell) = \ell' \mathbf{vec}(S(\mathbf{X})_t^{\leq N})$. Therefore, we have that

$$\int_0^t \sigma_s(\ell) dB_s = \ell' \int_0^t \mathbf{vec}(S(\mathbf{X})_t^{\leq N}) dB_s.$$

As for the second term,

$$\sigma_t(\ell)^2 = (\ell' \mathbf{vec}(S(\mathbf{X})_t^{\leq N}))(\ell' \mathbf{vec}(S(\mathbf{X})_t^{\leq N}))' = \ell' \mathbf{vec}(S(\mathbf{X})_t^{\leq N}) \mathbf{vec}(S(\mathbf{X})_t^{\leq N})' \ell.$$

Notice that the elements of the matrix $\tilde{Q}(t) := \mathbf{vec}(S(\mathbf{X})_t^{\leq N}) \mathbf{vec}(S(\mathbf{X})_t^{\leq N})'$ are given by

$$\tilde{Q}_{\mathcal{L}(I)\mathcal{L}(J)}(t) = \langle e_I, S(\mathbf{X})_t \rangle \langle e_J, S(\mathbf{X})_t \rangle = \langle e_I \sqcup\!\sqcup\, e_J, S(\mathbf{X})_t \rangle,$$

where in the last equality we used the shuffle property of the signature of a weakly geometric p-rough path. By integrating against time and incorporating the $-\frac{1}{2}$, we arrive at

$$-\frac{1}{2}\int_0^t \tilde{Q}_{\mathcal{L}(I)\mathcal{L}(J)}(s) ds = -\frac{1}{2} \langle (e_I \sqcup\!\sqcup\, e_J) \otimes e_0, S(\mathbf{X})_t \rangle = Q_{\mathcal{L}(I)\mathcal{L}(J)}(t).$$

□

Remark 5.1. *Notice that the matrix* $\tilde{Q}(t) := \mathbf{vec}(S(\mathbf{X})_t^{\leq N}) \mathbf{vec}(S(\mathbf{X})_t^{\leq N})'$ *is positive semidefinite. By the monotonicity of the time integral, we also have that*

$$\ell' Q(t) \ell \geq 0$$

for any $\ell \in \mathbb{R}^{d_N}$. *Therefore,* $Q(t)$ *admits a Cholesky decomposition* $Q(t) = U(t)'U(t)$, *where* $U(t)$ *is an upper-triangular matrix with non-negative diagonal. Therefore, we can write*

$$\ell' Q(t) \ell = \ell' U(t)' U(t) \ell = (U(t)\ell)'(U(t)\ell) = \|U(t)\ell\|_2^2. \tag{5.9}$$

We will use this expression in practice, since computing the norm of a vector is cheaper than performing a vector-matrix product.

Let us now show how the previous proposition allows us to calibrate the model by Monte Carlo in a very efficient way. By substituting (5.9) at (5.8) and evaluating at $t = T$ we get that the marginal distribution of the discounted price process at maturity time T is

$$\tilde{S}_T(\ell) = S_0 \exp\Big(\|U(T)\ell\|_2^2 + \ell' \int_0^T \mathbf{vec}(S(\mathbf{X})_t^{\leq N}) dB_t \Big). \tag{5.10}$$

To estimate the price of a call option by a Monte Carlo simulation, we use the formula

$$C^{\text{model}}(T, K, \ell) = \mathbb{E}_{\mathbb{P}^*}(e^{-rT}(e^{rT}\tilde{S}_T(\ell) - K)_+) \approx \frac{1}{n_{MC}} \sum_{j=1}^{n_{MC}} (\tilde{S}_T(\ell)(\omega_j) - e^{-rT}K)_+,$$

where n_{MC} is the number of samples used in the Monte Carlo simulation and $\tilde{S}_T(\ell)(\omega_j)$ is the j-th sample of the marginal distribution $\tilde{S}_T(\ell)$.

Now, since we have that

$$\tilde{S}_T(\ell)(\omega) = S_0 \exp\Big(\|U(T)(\omega)\ell\|_2^2 + \ell' \int_0^T \mathbf{vec}(S(\mathbf{X})_t^{\leq N})(\omega) dB_t(\omega) \Big),$$

we have that we can draw n_{MC} samples of the distributions given by

$$U(T), \int_0^T \mathbf{vec}(S(\mathbf{X})_t^{\leq N}) dB_t$$

in an offline manner, and use them in every step of the calibration procedure. The following algorithm summarizes the process.

Algorithm 1 Calibration of $S(\ell)$ by a Monte Carlo simulation

for $j = 1, \ldots n_{MC}$ **do**
    Draw the j-th trajectory for B and X, which we denote by $B(\omega_j)$ and $X(\omega_j)$.
    Compute the Cholesky decomposition of

$$Q_{\mathcal{L}(I)\mathcal{L}(J)}(T)(\omega_j) = -\frac{1}{2}\langle (e_I \sqcup e_J) \otimes e_0, S(\mathbf{X})_T(\omega_j)\rangle,$$

which we denote by $U(T)(\omega_j)$. Also compute

$$\int_0^T \mathbf{vec}(S(\mathbf{X})_t^{\leq N})(\omega_j) dB_t(\omega_j).$$

end for
Initialize parameters of the model $\ell \in \mathbb{R}^{d_N}$.
while ℓ has not converged **do**
    Compute

$$\tilde{S}_T(\ell)(\omega_j) = S_0 \exp\Big(\|U(T)(\omega_j)\ell\|_2^2 + \ell' \int_0^T \mathbf{vec}(S(\mathbf{X})_t^{\leq N})(\omega_j) dB_t(\omega_j) \Big)$$

for $j = 1, \ldots, n_{MC}$, which simply accounts for two vector-matrix products and the computation of a norm.
    Compute

$$C^{\text{model}}(T_i, K_i, \ell) \approx \frac{1}{n_{MC}} \sum_{j=1}^{n_{MC}} (\tilde{S}_{T_i}(\ell)(\omega_j) - e^{-rT_i}K_i)_+.$$

    Compute the loss function

$$L(\ell) = \sum_{i=1}^{n} \gamma_i \big(C^{\text{market}}(T_i, K_i) - C^{\text{model}}(T_i, K_i, \ell)\big)^2. \tag{5.11}$$

    Update ℓ by using some numerical optimization method that minimizes $L(\ell)$.
end while

5.4 Empirical analysis

In the next pages we will calibrate our signature based model, which we recall is given by

$$\begin{aligned} dS_t(\ell) &= S_t(\ell)\sigma_t(\ell)dB_t \\ \sigma_t(\ell) &:= \ell_\emptyset + \sum_{0<|I|\leq N} \ell_I \langle e_I, S(\mathbf{X})_t \rangle, \end{aligned}$$

to three different options prices datasets, showing that it is able to provide highly accurate fits to the corresponding implied volatility surfaces[1].

Following [CGS22], we will set γ_i to be Vega weights, defined as the multiplicative inverse of

$$\mathcal{V}(T_i, K_i) := \frac{\partial C_{BS}(T_i, S_0, K_i, r, \sigma)}{\partial \sigma}.$$

In all cases we will consider call options of 9 different strikes, which vary from 80% to 120% of the initial spot price S_0, and 10 maturities, ranging from around 30 days to 1 year.

As a source noise $\mathbf{X}$ for the volatility process we will always use the lift to $\mathcal{W}G\widehat{\Omega}^p_T(\mathbb{R})$ given by Propositions 4.4 and 4.5 of a (time augmented) one-dimensional fractional Ornstein-Uhlenbeck process, which is given by the solution of the SDE

$$dX_t = \xi dW_t^H - \lambda(X_t - m)dt, \tag{5.12}$$

where W_t^H is a fractional Brownian motion in the Riemann-Liouville form, i.e.

$$W_t^H = \frac{1}{\Gamma(H + \frac{1}{2})} \int_0^t (t-s)^{H-\frac{1}{2}} dW_t,$$

with $dW_t = \rho dB_t + \sqrt{1-\rho^2} dZ_t$. One can easily check that there is an explicitly solution, which is given by

$$X_t = \xi \int_0^t e^{-\lambda(t-s)} dW_s^H + \theta,$$

where the integral is in the Young-Stieltjes sense. Following [GJR18], we simulate X by taking

$$X_{t_{i+1}} - X_{t_i} = \xi(W^H_{t_{i+1}} - W^H_{t_i}) + \lambda(t_{i+1} - t_i)(m - X_{t_i}).$$

[1] Recall that, as we mentioned in Section 3.3.2, implied volatilities are widely used as a way of quoting prices of calls more than the prices themselves, since it gives a better idea of their relative value.

5.4.1 Heston's model option prices

We start by calibrating to call options obtained from the classical Heston model, which we recall has dynamics given by

$$\begin{aligned} dS_t &= S_t(rdt + \sqrt{v_t}dB_t) \\ dv_t &= \kappa(\theta - v_t)dt + \nu\sqrt{v_t}dW_t \\ dW_t &= \rho dB_t + \sqrt{1-\rho^2}dZ_t. \end{aligned}$$

We select $r = 0$, $\kappa = 5$, $\nu = 0.6$, $\theta = 0.1$, $v_0 = 0.1$, $S_0 = 1$ and $\rho = -0.4$.

For the fractional Ornstein-Uhlenbeck process given by (5.12) we select $\xi = 0.6$, $\lambda = 12$, $m = 0.1$, $X_0 = 0.1$ and $\rho = -0.5$. We truncate the signature at depth $N = 3$, which means that the model has 15 learnable parameters. We also set $H = 0.5$, which means that the fBm in the Riemann-Liouville form is simply a Brownian motion. We set $n_{MC} = 5 \times 10^5$ for the Monte Carlo.

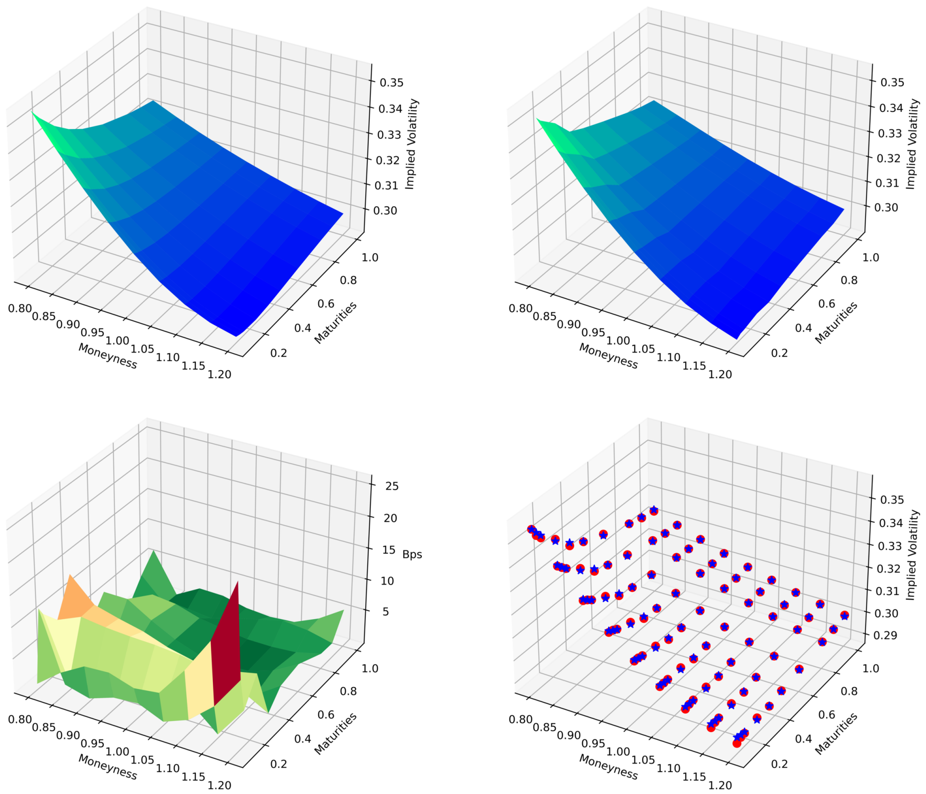

FIGURE 5.1: Top left: implied volatility surface generated from the Heston model. Top right: implied volatility surface generated by the calibrated $S(\ell)$. Bottom left: absolute errors between the two surfaces expressed in basis points (Bps)[2]. Bottom right: blue stars correspond to the implied volatilities of the Heston model, red dots denote the implied volatility surface generated by the calibrated $S(\ell)$.

[2]Basis points (Bps) are a unit of measure that is widely used in finance. One basis point is equivalent to 0.01%.

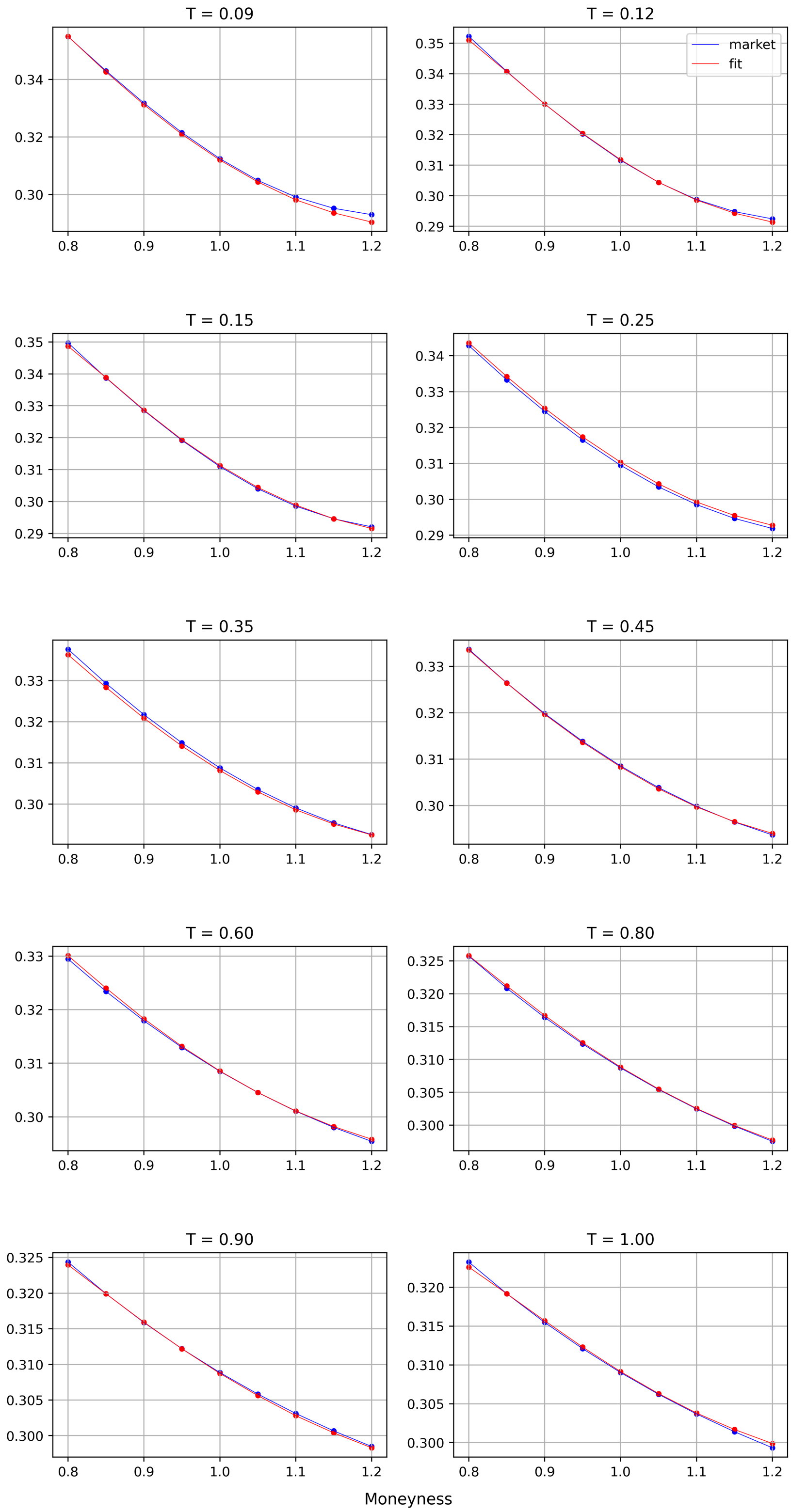

FIGURE 5.2: Volatility smiles for the different maturities, Heston generated data. Blue: market volatility smiles. Red: calibrated volatility smiles.

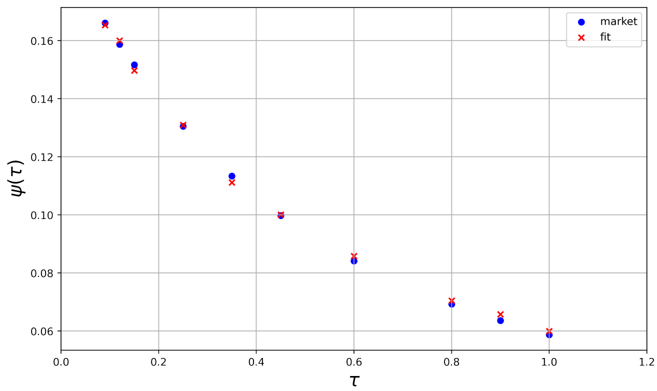

FIGURE 5.3: Comparison between market term structure of the at-the-money (blue circles) implied volatility skew and the calibrated one from $S(\ell)$ (red crosses), Heston generated data. Recall that, as we mentioned in Section 3.3.2, in the Heston model this converges to a constant as $\tau \to 0$.

5.4.2 Rough Heston's model option prices

Next we calibrate the model to call options obtained from the Rough Heston model[3], which we recall has dynamics given by

$$dS_t = S_t(rdt + \sqrt{v_t}dB_t)$$
$$v_t = v_0 + \frac{1}{\Gamma(\alpha)}\int_0^t (t-s)^{\alpha-1}\kappa(\theta - v_s)ds + \frac{1}{\Gamma(\alpha)}\int_0^t (t-s)^{\alpha-1}\kappa\nu\sqrt{v_s}dW_s$$
$$dW_t = \rho dB_t + \sqrt{1-\rho^2}dZ_t,$$

where $\alpha = H + \frac{1}{2}$. In our case, we select $r = 0$, $\alpha = 0.6$, $\kappa = 2$, $\nu = 0.4$, $\theta = 0.15^2$, $v_0 = 0.15^2$, $S_0 = 1$ and $\rho = -0.6$.

For the fractional Ornstein-Uhlenbeck process given by (5.12) we select $\xi = 0.6$, $\lambda = 10$, $m = 0.15$, $X_0 = 0.15$, $H = 0.1$ and $\rho = -0.7$. We truncate the signature at depth $N = 3$, which means that the model has 15 learnable parameters. We set $n_{MC} = 10^6$ for the Monte Carlo.

[3]This was done by using the Matlab code provided in https//:github.com/sigurdroemer/rough_heston.

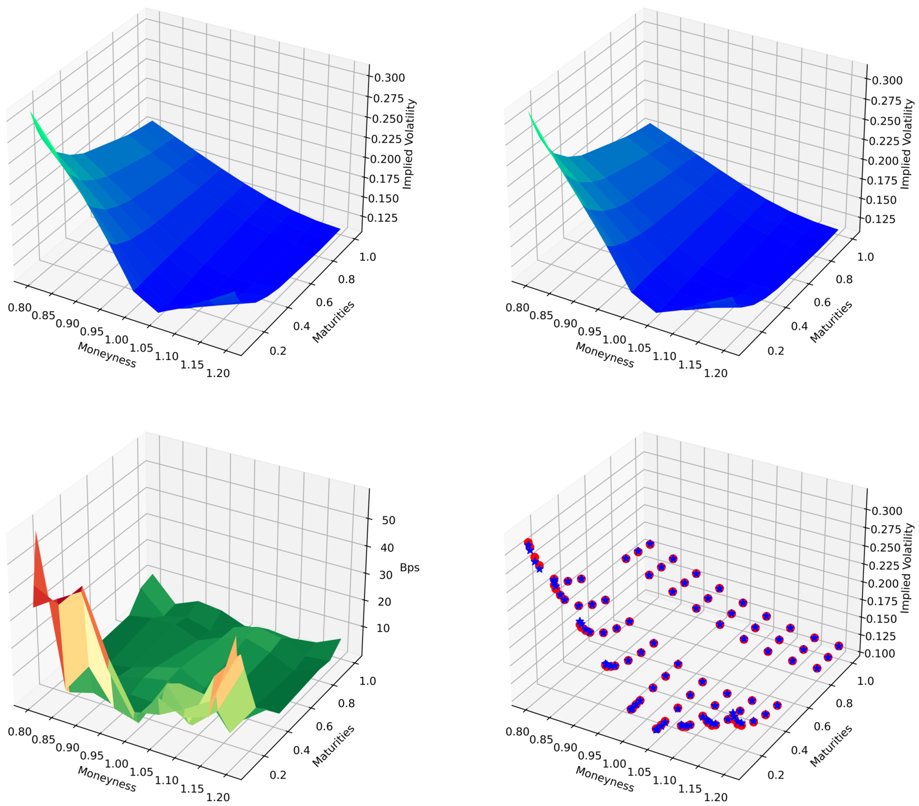

FIGURE 5.4: Top left: implied volatility surface generated from the Rough Heston model. Top right: implied volatility surface generated by the calibrated $S(\ell)$. Bottom left: absolute errors between the two surfaces expressed in basis points (Bps). Bottom right: blue stars correspond to the implied volatilities of the Rough Heston model, red dots denote the implied volatility surface generated by the calibrated $S(\ell)$.

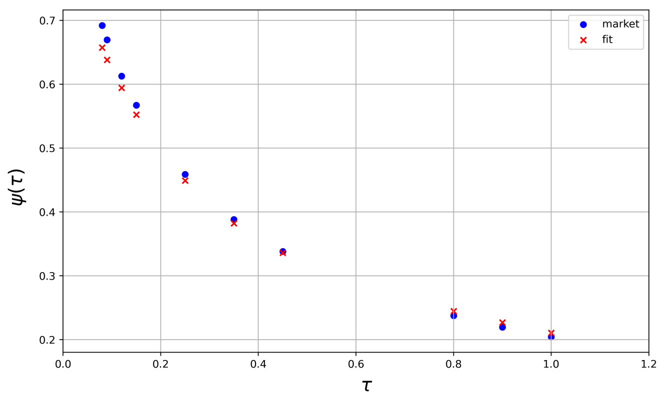

FIGURE 5.5: Comparison between market term structure of the at-the-money (blue circles) implied volatility skew and the calibrated one from $S(\ell)$ (red crosses), Rough Heston generated data.

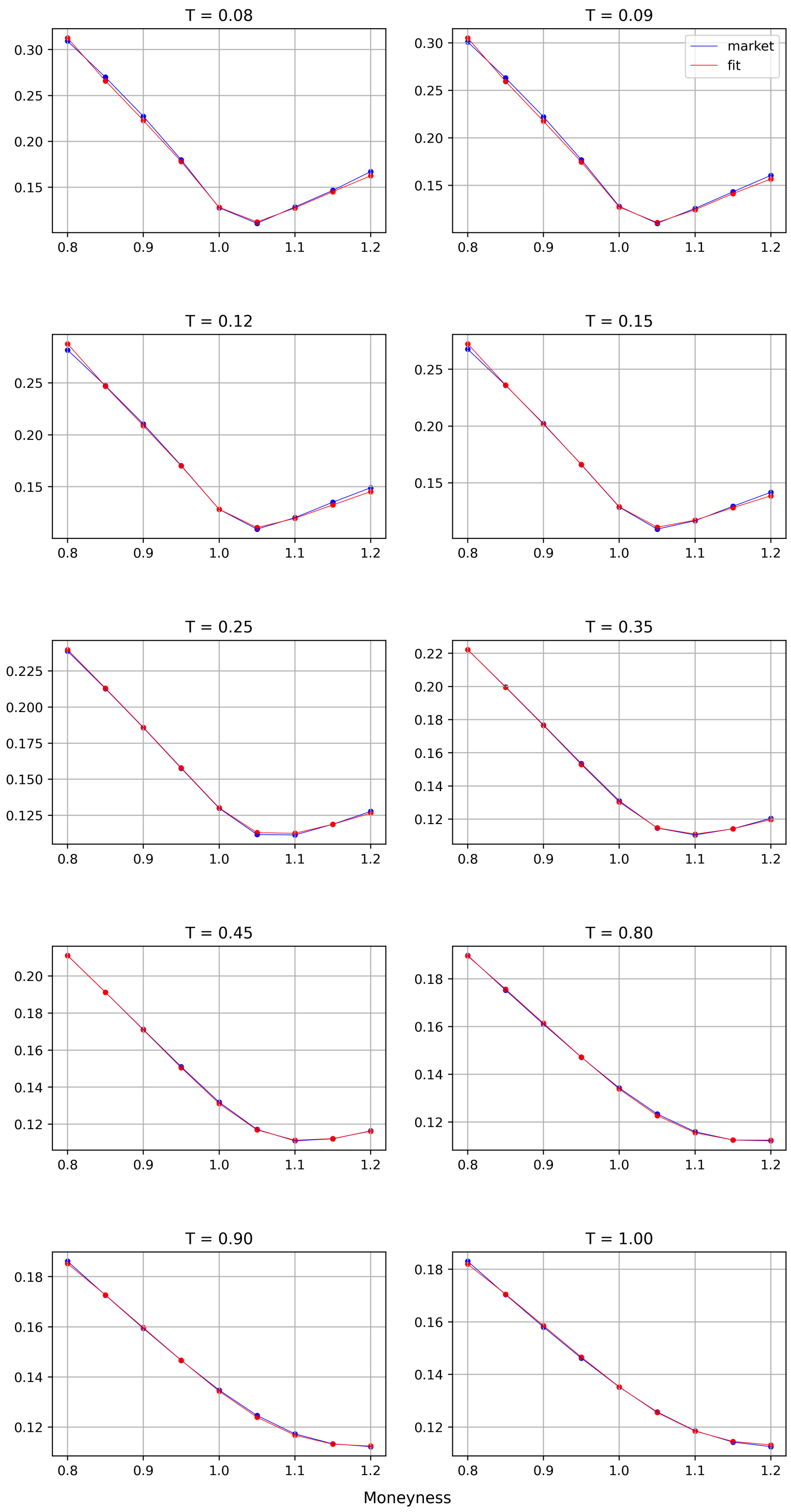

FIGURE 5.6: Volatility smiles for the different maturities, Rough Heston generated data. Blue: market volatility smiles. Red: calibrated volatility smiles.

5.4.3 Market data

Finally we calibrate the model to real market data. We consider the trading day 17/03/2021 for call options written on the S&P index. The data was obtained by CBOE DataShop, and consists in bid and ask prices of 9 strike prices, which vary from 80% to 120% of the initial spot price S_0, and 10 maturities, ranging from 30 days to 1 year.

For the fractional Ornstein-Uhlenbeck process given by (5.12) we select $\xi = 1$, $\lambda = 0.5$, $X_0 = 0$, $m = 0$, $H = 0.03$ and $\rho = -0.75$.

We truncate the signature at depth $N = 3$, which means that the model has 15 learnable parameters. We set $n_{MC} = 5 \times 10^5$ for the Monte Carlo. We take the interest rate r to be the 1 year *Daily Treasury Par Yield Curve Rate*[4], which in March 17, 2021 was $r = 0.07$.

We calibrate our model to match the midpoint prices (that is, the average between the ask and bid prices). However, we remark that any price between the ask and the bid can be considered correct.

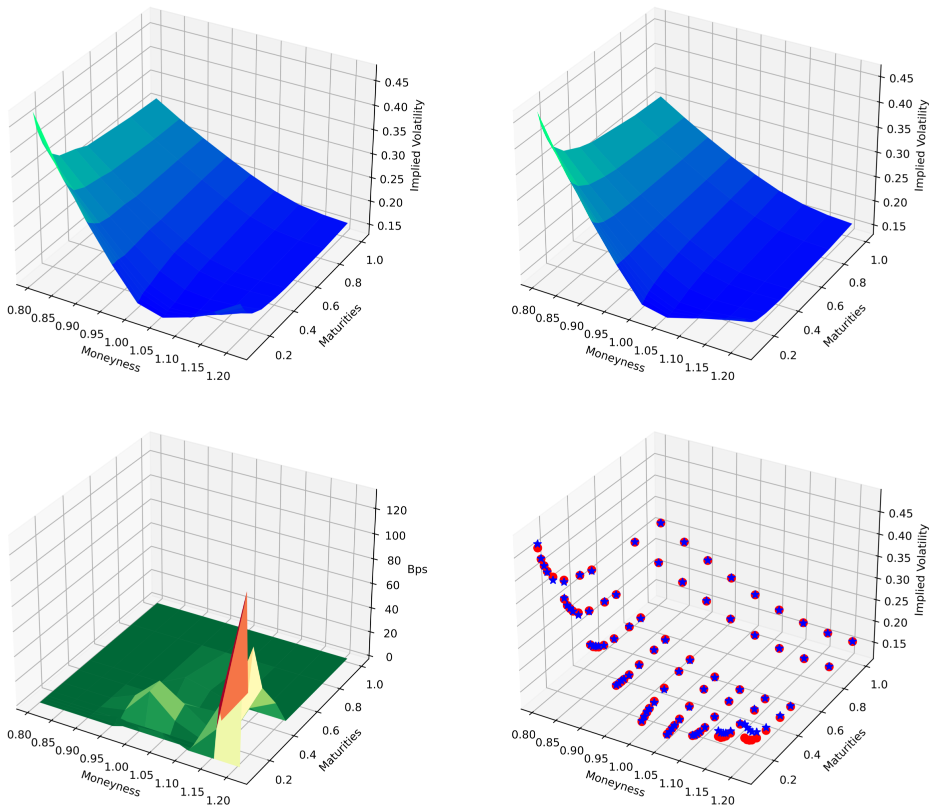

FIGURE 5.7: Top left: implied volatility surface of the S&P index as of 17/03/2021. Top right: implied volatility surface generated by the calibrated $S(\ell)$. Bottom left: absolute errors between the two surfaces expressed in basis points (Bps). Bottom right: blue stars correspond to the implied volatilities of the S&P index as of 17/03/2021, red dots denote the implied volatility surface generated by the calibrated $S(\ell)$.

[4]See https://home.treasury.gov/policy-issues/financing-the-government/interest-rate-statistics

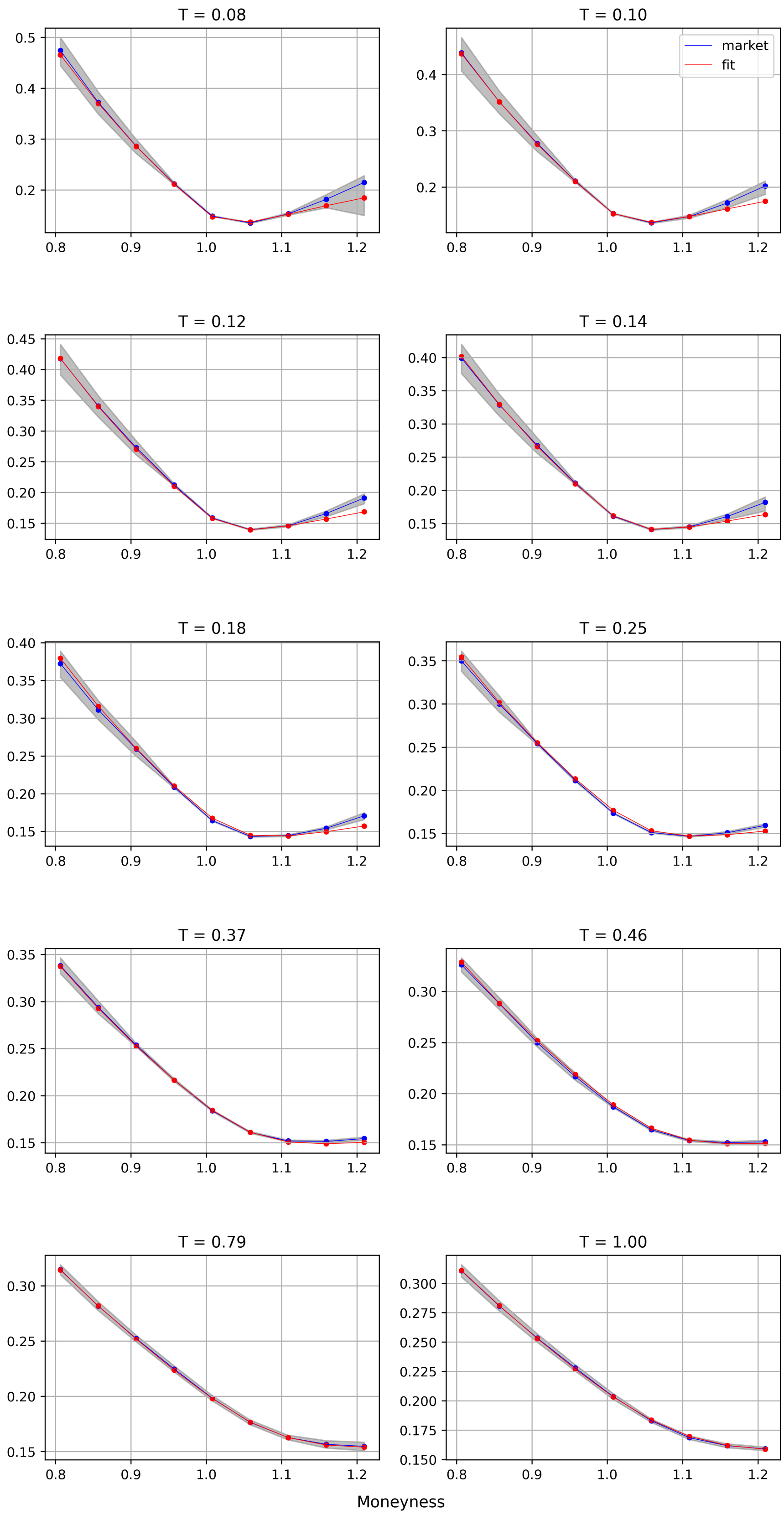

FIGURE 5.8: Volatility smiles for the different maturities, S&P index as of 17/03/2021. Blue: midpoint market volatility smiles. Red: calibrated volatility smiles. Shade of gray: bid-ask spread.

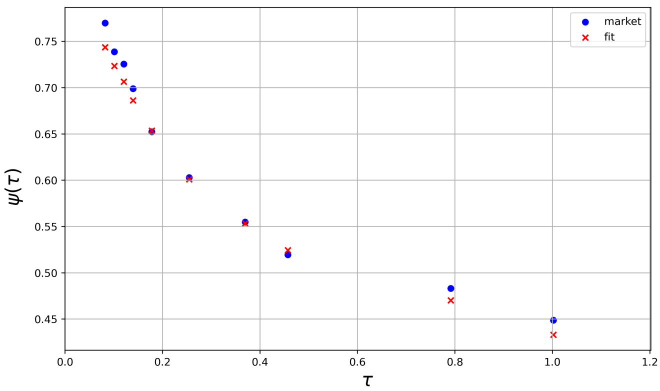

FIGURE 5.9: Comparison between the midpoint market term structure of the at-the-money (blue circles) implied volatility skew and the calibrated one from $S(\ell)$ (red crosses), S&P index as of 17/03/2021.

Chapter 6

Conclusions

After an introduction to stochastic processes, mathematical finance and the signature transform, we have studied a stochastic volatility model where the dynamics of the volatility are described by linear functions of the (time extended) signature of a primary underlying process X.

This was originally proposed in [Cuc+23]. However, they only considered the case where X was a continuous semimartingale. We have extended these models by allowing X to be any process with the trajectories being almost surely weakly geometric p-rough paths, in such a way that their signatures are well-defined and satisfy the shuffle product identity. In particular, X can be given by stochastic processes that are as rough as a fractional Brownian motion.

Then we showed how these models can be efficiently calibrated in practice by using a Monte Carlo simulation. Finally, we calibrated it to three datasets of call options: one with data generated from a Heston model, one from a rough Heston model, and one with real market data. In all three cases the calibrated prices adjusted well the implied volatility surface.

Appendix A

Introduction to stochastic analysis

By Proposition 2.11 and Proposition 4.1, we have that a Brownian motion $(B_t)_{t\in[0,T]}$ has sample paths that are almost surely not of finite p-variation for any $p \leq 2$. Therefore, one is not able to use the Young-Stieltjes integral to give a meaning to many simple integrals driven by Brownian motion, like for example

$$\int_0^t B_s dB_s.$$

In the 1950s, Kyoshi Itô introduced the Itô integral, which is an extension of the Young-Stieltjes integral for integrals driven by Brownian motion. The main goal of this section will be to define such integral and see some of its properties. We will not provide any proofs, and instead will reference them to other books and notes.

Through this chapter $(\Omega, \mathcal{F}, \mathbb{F}, \mathbb{P})$ will denote a filtered probability space. Unless specified, all stochastic processes will be assumed to be defined there, and moreover, we will always assume adaptability to the filtration $\mathbb{F} := (\mathcal{F}_t)_{t\geq 0}$.

A.1 The Itô integral

As we explained in the introduction of this chapter, our first goal will be to define integrals of the form

$$\int_0^T X_t dB_t$$

in the Itô sense, where $B = (B_t)_{t\in[0,T]}$ is a Brownian motion and $(X_t)_{t\in[0,T]}$ is a stochastic process. Besides giving a meaning to it, we will also see what conditions do we need X to satisfy in order for the integral to be well-defined.

The Itô integral for mean square integrable processes

We will first define the Itô integral over the space of mean square integrable processes.

Definition A.1 (Mean square integrable process)**.** *A process $X = (X_t)_{t\in[0,T]}$ is said to be mean square integrable if $\mathbb{E}[\int_0^T X_t^2 dt] < \infty$. We will denote this space as $\mathcal{S}_2$.*

It can be seen that $\mathcal{S}_2$ is a Hilbert space with norm given by

$$\|X\|_{\mathcal{S}_2} := \sqrt{\mathbb{E}[\int_0^T X_t^2 dt]}, \tag{A.1}$$

which is induced by the scalar product given by $\langle X, Y \rangle_{\mathcal{S}_2} := \mathbb{E}[\int_0^T X_t Y_t dt]$.

Similarly to the Lebesgue integral, we will first define the Itô integral over a class of very simple processes, and then extend it to any process in $\mathcal{S}_2$.

Definition A.2 (Elementary process)**.** *A process $X \in \mathcal{S}_2$ is an elementary process if there exists a partition $0 = t_0 < \cdots < t_n = T$ of $[0, T]$ for which $X_t = X_{t_i}$ for $t \in [t_i, t_{i+1})$. We denote such space as $\mathcal{E}$.*

Definition A.3 (Itô integral of an elementary process)**.** *We define the Itô integral of an elementary process X against a Brownian motion B as*

$$\int_0^T X_t dB_t := \sum_{i=0}^{n-1} X_{t_i}(B_{t_{i+1}} - B_{t_i}). \tag{A.2}$$

In order to extend this definition to any $X \in \mathcal{S}_2$, the standard approach is to prove that (A.2) gives a linear continuous functional between two metric spaces. By doing this, one is able to extend the definition to a more general class of processes given by the closure of $\mathcal{E}$ with respect to a suitable metric.

The linearity of (A.2) is trivial to show. In order to prove continuity, we have the following result.

Lemma A.1 (Isometry property for elementary processes)**.** *For any $X \in \mathcal{E}$,*

$$\mathbb{E}\Big[\Big(\int_0^T X_t dB_t\Big)^2\Big] = \mathbb{E}\Big[\int_0^T X_t^2 dt\Big].$$

Consequently, the mapping

$$\mathcal{E} \ni X \mapsto \int_0^T X_t dB_t$$

is continuous from $\mathcal{E}$ into $L^2(\Omega)$.

Proof. See [Bal17, Lemma 7.1]. □

Denote now by $\overline{\mathcal{E}}$ the closure of $\mathcal{E}$ in $\mathcal{S}_2$ with respect to the norm given by $\|\cdot\|_{\mathcal{S}_2}$. Since $\mathcal{S}_2$ is complete with respect to this norm (it is a Hilbert space), we have that $\overline{\mathcal{E}} \subset \mathcal{S}_2$. The converse inclusion is also true, as it is stated in the following proposition.

Proposition A.1. *The elementary processes $\mathcal{E}$ are dense in $\mathcal{S}_2$.*

Proof. See [Bal17, Lemma 7.2]. □

Using this approximation result, we can given the following definition.

Definition A.4. *The Itô integral of a process $X \in \mathcal{S}_2$ is*

$$\int_0^T X_t dB_t := L^2(\Omega) - \lim_{n\to\infty} \int_0^T X_t^n dB_t,$$

where $\{X^n\}_{n\geq 0} \subset \mathcal{E}$ is a sequence of elementary processes that converges to X in the norm given by $\|\cdot\|_{\mathcal{S}_2}$.

By using the Itô isometry for elementary processes, one easily checks that this definition does not depend on which sequence do we choose.

Proposition A.2. *The Itô integral in Definition A.4, defined as a map from $\mathcal{S}_2$ to $L^2(\Omega)$, satisfies the following properties:*

1. *It is an isometry.*
2. *It is a centered random variable.*
3. *It is a linear operator.*

Proof. See [Bal17, Theorem 7.1] for the first two properties. The third one is trivial to prove. □

The indefinite Itô integral as a stochastic process

Let us now consider the Itô integral as a stochastic process and see some of its most important properties. For $X \in \mathcal{S}_2$, this is defined as

$$\int_0^t X_s dB_s := \int_0^T X_s \mathbf{1}_{[0,t]}(s) dB_s, \quad t \in [0,T].$$

Theorem A.1. *If $X \in \mathcal{S}_2$, then the process given by $\int_0^t X_s dB_s$ is a square integrable martingale with respect to $\mathbb{F}$. Moreover, it has a continuous modification.*

Proof. See [Bal17, Theorem 7.3]. □

An extension of the Itô integral

Next we will extend the definition of the Itô integral to a larger class of integrands. In particular, we will define it for processes belonging to the bigger space given by all X such that

$$\mathbb{P}\Big(\int_0^T X_t^2 dt < \infty\Big) = 1,$$

which will be denoted by Λ_2. Note that clearly we have the inclusion $\mathcal{S}_2 \subset \Lambda_2$.

The main idea is to approximate $X \in \Lambda_2$ by processes in $\mathcal{S}_2$, by using a technique called *localization*.

Similarly to what we did in the definition of local martingales, we say that a sequence of stopping times $\{\tau_n\}_{n\geq 0}$ is *localizing* for X in $\mathcal{S}_2$ if:

1. The sequence $\{\tau_n\}_{n\geq 0}$ is increasing.
2. For all $n \geq 0$, the process given by $X_t^n := X_t \mathbf{1}_{\{t<\tau_n\}}$ belongs to $\mathcal{S}_2$.

3. $\mathbb{P}(\lim_{n\to\infty} \tau_n = T) = 1$.

For every $n > 0$, let us define the random variable

$$\tau_n := \inf\{t \in [0,T] \mid \int_0^t X_s^2 ds > n\},$$

with the convention that $\inf \emptyset = T$. One can easily check that these are indeed stopping times, and moreover, they are localizing in $\mathcal{S}_2$.

One is then able to define the Itô integral for any process $X \in \Lambda_2$ as the a.s. limit of the Itô integral of the sequence of processes given by $\{X_t^n\}_{n\geq 0}$, with $X_t^n := X_t \mathbf{1}_{\{t<\tau_n\}}$, i.e.

$$\int_0^T X_t dB_t := \text{a.s.} - \lim_{n\to\infty} \int_0^T X_t^n dB_t.$$

We refer the interested reader to [Bal17, Section 7.5] to show that this integral is well-defined.

Finally, let us mention that the indefinite Itô integral of a process in Λ_2 is not a martingale. However, it is a local martingale.

Proposition A.3. *If $X \in \Lambda_2$, then the process given by $\int_0^t X_s dB_s$ is a local martingale with respect to $\mathbb{F}$. Moreover, it has a continuous modification.*

Proof. See [Bal17, Propostion 7.7]. □

A.2 Fundamental results

In this section we will see some fundamental results that are necessary for the development of the theory in mathematical finance.

Itô processes

We start with the definition of a very important class of stochastic processes called Itô processes.

Definition A.5 (Itô process)**.** *An Itô process is a process that satisfies the equation*

$$X_t = X_0 + \int_0^t \alpha_s ds + \int_0^t \beta_s dB_s, \tag{A.3}$$

where X_0 is $\mathcal{F}_0$-measurable and $(\alpha_t)_{t\in[0,T]}$, $(\beta_t)_{t\in[0,T]}$ are stochastic processes with

$$\int_0^T (|\alpha_t| + |\beta_t|^2) dt < \infty \quad a.s.$$

Notice that an Itô process satisfies the following equation in differential notation:

$$dX_t = \alpha_t dt + \beta_t dB_t.$$

This allows us to extend the definition of the Itô integral in a natural way, allowing the integrator to be an Itô process, by defining

$$\int_0^t Y_s dX_s := Y_0 + \int_0^t Y_s \alpha_s ds + \int_0^t Y_s \beta_s dB_s.$$

Let us also mention that the more general class of (continuous) integrators such that one is able to define the Itô integral are continuous semi-martingales, which are defined as the sum of a process of bounded variation and a local martingale. Notice that an Itô process is also a continuous semimartingale.

Proposition A.4. *If $X = (X_t)_{t\in[0,T]}$ is a continuous martingale such that*

$$X_t = \int_0^t \alpha_s ds, \quad \text{with } \mathbb{P}\text{-a.s.} \quad \int_0^T |\alpha_t| dt < \infty,$$

then for all $t \in [0,T]$ we have that $X_t = 0$ $\mathbb{P}$-a.s.

In particular, this implies that

- *The decomposition of an Itô process is unique.*
- *If X is an Itô process and a local martingale with expression given by (A.3), then $\alpha_t = 0$ $\mathbb{P}$-a.s. As a consequence, if X is an Itô process and a local martingale, we have that*

  $$\int_0^t Y_s dX_s := \int_0^s Y_s \beta_s dB_s$$

  is also a local martingale, where we assume $\int_0^T |Y_t\beta_t|^2 ds < \infty$ $\mathbb{P}$-a.s.

Proof. See [Lam08, Proposition 3.4.9]. □

Itô's formula

Next we will introduce a version of the Itô formula, which is a fundamental tool in stochastic analysis.

Theorem A.2 (Itô's formula)**.** *Let $f \in \mathcal{C}^{1,2}([0,T] \times \mathbb{R}^d; \mathbb{R})$ and consider a d-dimensional Itô processes $X = (X^1, \dots, X^d)$, with $X^i = (X^i_t)_{t\in[0,T]}$ having decomposition*

$$dX^i_t = \alpha^i_t dt + \beta^i_t dB^i_t,$$

and where $B^1, \dots, B^d$ are d (possibly correlated) standard Brownian motions.

Then, the process given by $Y_t = f(t, X_t)$ is also an Itô process, and the following formula holds:

$$dY_t = \partial_t f(t, X_t) dt + \sum_{i=1}^d \partial_{x_i} f(t, X_t) dX^i_t + \frac{1}{2} \sum_{i,j=1}^d \partial_{x_i x_j} f(t, X_t) d\langle X^i, X^j \rangle_t \tag{A.4}$$

where $d\langle X^i, X^j\rangle_t$ denotes the differential of the covariance process, which for Itô processes is defined as

$$d\langle X^i, X^j \rangle_t := \beta^i_t \beta^j_t \rho_{ij} dt,$$

with ρ_{ij} denoting the correlation between B^i and B^j.

Proof. See [Bal17, Theorem 8.1]. $\square$

Notice that if not for the second summation term, the formula would be just like the classical chain rule. This is why sometimes the Itô formula is viewed as the particular version of the chain rule for the Itô integral.

As a consequence, we get the following integration by parts formula.

Corollary A.1 (Integration by parts)**.** *Let X^1 and X^2 be two Itô processes. Then we have that*

$$d(X_t^1 X_t^2) = dX_t^1 X_t^2 + X_t^1 dX_t^2 + d\langle X^1, X^2\rangle_t.$$

Stochastic differential equations

Stochastic differential equations are very important in mathematical finance, as they provide models for stochastic processes. We define them as follows.

Definition A.6 (Stochastic differential equation)**.** *We call a multidimensional stochastic differential equation (SDE) an equation of the type*

$$X_t = Z + \int_0^t b(s, X_s)ds + \int_0^t \sigma(s, X_s)dW_s, \tag{A.5}$$

where:

- $W = (W^1, \dots, W^p)$ *is a standard $\mathbb{R}^p$-valued Brownian motion.*
- $b : [0, T] \times \mathbb{R}^n \to \mathbb{R}^n$, $b(s, x) = (b_1(s, x), \dots, b_n(s, x))$.
- $\sigma : [0, T] \times \mathbb{R}^n \to \mathbb{R}^n \times \mathbb{R}^p$, $\sigma(s, x) = (\sigma_{i,j}(s, x))_{1\le i\le n, 1\le j\le p}$.
- $Z = (Z^1, \dots, Z^n)$ *is an $\mathcal{F}_0$-measurable random variable with values in $\mathbb{R}^n$.*

We say that a continuous process that is $\mathbb{F}$-adapted $X = (X_t)_{t\in[0,T]}$ is a solution if

$$\int_0^t |b(s, X_s)|ds < \infty, \quad \int_0^t |\sigma(s, X_s)|^2 ds < \infty \quad a.s.$$

where $|\cdot|$ denotes the Euclidean norm and for all $t \in [0, T]$ it satisfies (A.5) almost surely.

A sufficient condition for the existence and uniqueness of solution is the following.

Theorem A.3. *Assume that*

1. $|b(t, x) - b(t, y)| + |\sigma(t, x) - \sigma(t, y)| \le K|x - y|$.
2. $|b(t, x)| + |\sigma(t, x)| \le K(1 + |x|)$.
3. $\mathbb{E}[|Z|^2] < \infty$.

Then there exists a unique solution to (A.5).

Proof. See [Lam08, Theorem 3.5.5]. $\square$

A.3 The Stratonovich integral

We end this section with an elementary introduction to the Stratonovich integral, which provides an alternative to the Itô integral for stochastic processes. Although it is well-defined for continuous semimartingales, for simplicity we will only define it for Itô processes.

Definition A.7 (Stratonovich integral). *Let $X = (X_t)_{t\in[0,T]}$ and $Y = (Y_t)_{t\in[0,T]}$ be two Itô processes with decomposition*

$$dX_t = \alpha_t dt + \beta_t dB_t, \quad dY_t = \gamma_t dt + \eta_t dW_t,$$

where B and W are standard Brownian motions with correlation $\rho \in [-1, 1]$.

We then define the Stratonovich integral of Y against X as

$$\int_0^t Y_s \circ dX_s := \int_0^t Y_s dX_s + \frac{1}{2}\langle X, Y\rangle_t. \tag{A.6}$$

We remark that, contrary to the Itô integral and its Itô formula, the Stratonovich integral satisfies the usual chain rule. Indeed, let $f \in \mathcal{C}^{1,3}([0,T] \times \mathbb{R}^d; \mathbb{R})$ and consider a d-dimensional Itô processes $X = (X^1, \dots, X^d)$, with decomposition given like in Theorem A.2.

By writing $Y_t^i = \partial_{x_i} f(t, X_t)$ and applying the Itô formula to it, it can be readily verified that

$$d\langle Y^i, X^i\rangle_t = \sum_{j=1}^d \partial_{x_i x_j} f(t, X_t) d\langle X^i, X^j\rangle_t$$

In this case, Itô's formula for the process given by $f(t, X_t)$ becomes

$$\begin{aligned} df(t, X_t) &= \partial_t f(t, X_t) dt + \sum_{i=1}^d Y_t^i dX_t^i + \frac{1}{2}\sum_{i=1}^d d\langle Y^i, X^i\rangle_t \\ &= \partial_t f(t, X_t) dt + \sum_{i=1}^d \partial_{x_i} f(t, X_t) \circ dX_t^i, \end{aligned}$$

which is the usual chain rule.

Bibliography

[ALV07] Elisa Alòs, Jorge A. Leon, and Josep Vives. "On the short-time behavior of the implied volatility for jump-diffusion models with stochastic volatility". In: *Finance and stochastics* 11.4 (2007), pp. 571–589.

[AMN01] Elisa Alòs, Olivier Mazet, and David Nualart. "Stochastic Calculus with Respect to Gaussian Processes". In: *The Annals of Probability* 29.2 (2001), pp. 766–801.

[Bal17] P. Baldi. *Stochastic Calculus: An Introduction Through Theory and Exercises*. Universitext. Springer International Publishing, 2017.

[Bay+23] Christian Bayer et al. "Optimal stopping with signatures". In: *The Annals of Applied Probability* 33.1 (2023), pp. 238–273.

[BS73] Fischer Black and Myron Scholes. "The pricing of options and corporate liabilities". In: *Journal of political economy* 81.3 (1973), p. 637.

[CQ02] Laure Coutin and Zhongmin Qian. "Stochastic rough path Analysis and fractional Brownian motions". In: *Probability Theory and Related Fields* 122 (2002), pp. 108–140.

[CGS22] Christa Cuchiero, Guido Gazzani, and Sara Svaluto-Ferro. *Signature-based models: theory and calibration*. 2022. arXiv: 2207.13136.

[Cuc+23] Christa Cuchiero et al. *Joint calibration to SPX and VIX options with signature-based models*. 2023. arXiv: 2301.13235.

[DS94] Freddy Delbaen and Walter Schachermayer. "A General Version of the Fundamental Theorem of Asset Pricing". In: Springer Finance (1994), pp. 149–205.

[Dud02] R. M. Dudley. *Real Analysis and Probability*. 2nd ed. Cambridge Studies in Advanced Mathematics. Cambridge University Press, 2002.

[Dup94] Bruno Dupire. "Pricing with a Smile". In: 1994.

[Dup19] Bruno Dupire. "Functional Itô calculus". In: *Quantitative Finance* 19.5 (2019), pp. 721–729.

[ER19] Omar El Euch and Mathieu Rosenbaum. "The characteristic function of rough Heston models". In: *Mathematical finance* 29.1 (2019), pp. 3–38.

[FV06] Peter K. Friz and Nicolas B. Victoir. "A note on the notion of geometric rough paths". In: *Probability theory and related fields* 136.3 (2006), pp. 395–416.

[FV10] Peter K. Friz and Nicolas B. Victoir. *Multidimensional stochastic processes as rough paths : theory and applications*. Cambridge studies in advanced mathematics ; 120. Cambridge: Cambridge University Press, 2010.

[Fuk11] Masaaki Fukasawa. "Asymptotic analysis for stochastic volatility: martingale expansion". In: *Finance and stochastics* 15.4 (2011), pp. 635–654.

[GJR18] Jim Gatheral, Thibault Jaisson, and Mathieu Rosenbaum. "Volatility is rough". In: *Quantitative Finance* 18.6 (2018), pp. 933–949.

[GT06] Jim Gatheral and Nassim Nicholas Taleb. *The volatility surface: a practitioner's guide*. eng. Vol. 357. Wiley finance series. WILEY, 2006.

[Hes93] Steven L. Heston. "A Closed-Form Solution for Options with Stochastic Volatility with Applications to Bond and Currency Options". In: *The Review of financial studies* 6.2 (1993), pp. 327–343.

[Joh81] Richard. Johnsonbaugh. *Foundations of mathematical analysis / Richard Johnsonbaugh and W.E. Pfaffenberger*. Monographs and textbooks in pure and applied mathematics ; 62. Marcel Dekker, 1981.

[Kal97] O. Kallenberg. *Foundations of Modern Probability*. Probability and Its Applications. Springer New York, 1997.

[KLA20] Jasdeep Kalsi, Terry Lyons, and Imanol Perez Arribas. "Optimal Execution with Rough Path Signatures". In: *SIAM Journal on Financial Mathematics* 11.2 (2020), pp. 470–493.

[Lam08] Damien. Lamberton. *Introduction to stochastic calculus applied to finance / Damien Lamberton, Bernard Lapeyre*. eng. 2nd ed. Chapman & Hall/CRC financial mathematics series. Chapman & Hall/CRC, 2008.

[Leo14] Peter. Leoni. *The Greeks and Hedging Explained*. 1st ed. 2014. Financial Engineering Explained. Palgrave Macmillan UK, 2014.

[LCL04] Terry Lyons, Michael Caruana, and Thierry Lévy. *Differential equations driven by rough paths*. École d'Eté de Probabilités de Saint-Flour XXXIV, 2004.

[Lyo02] Terry. Lyons. *System control and rough paths*. Oxford mathematical monographs. New York: Oxford University Press, 2002.

[MN68] Benoit B. Mandelbrot and John W. Van Ness. "Fractional Brownian Motions, Fractional Noises and Applications". In: *Siam Review* 10 (1968), pp. 422–437.

[Mer73] Robert Merton. "Theory of rational option pricing". In: *The Bell Journal of Economics and Management Science* (1973), pp. 141–183.

[ML01] SV Muniandy and SC Lim. "Modeling of locally self-similar processes using multifractional Brownian motion of Riemann-Liouville type". In: *Physical review. E, Statistical, nonlinear, and soft matter physics* 63.4 (2001), p. 046104.

[MR05] Marek Musiela and Marek Rutkowski. *Martingale Methods in Financial Modelling*. 2. Aufl. Vol. 36. Stochastic Modelling and Applied Probability. Springer-Verlag, 2005, pp. xvi–xvi.

[Nou] Ivan Nourdin. *Selected Aspects of Fractional Brownian Motion*. Bocconi & Springer Series. Milano: Springer Milan.

[Ver12] Mark Veraar. "The stochastic Fubini theorem revisited". In: *Stochastics* 84.4 (2012), pp. 543–551.

[You36] L. C. Young. "An inequality of the Hölder type, connected with Stieltjes integration". In: *Acta Mathematica* 67.none (1936), pp. 251–282.