# Portfolio optimization under the rough Bergomi model

by

Giorgia Bifronte

to obtain the degree of Master of Science

at the Delft University of Technology.

| | | |
|---|---|---|
| Student number: | 5850169 | |
| Thesis committee: | Dr. F. Yu, | TU Delft, supervisor |
| | Prof. A. Papapantoleon, | TU Delft |
| | Dr. Ir. E. Lorist, | TU Delft |
| Academic Year: | 2023-2024 | |

TU Delft

# Abstract

This thesis addresses the portfolio allocation problem within a financial market featuring one riskless asset and a risky asset exhibiting rough Bergomi volatility. The objective is to maximize the expected utility of terminal wealth with respect to power utility. The volatility process in the model is driven by fractional Brownian motion and does not fit within the Markovian or semimartingale frameworks. To address this issue, we explore Markovian approximations for fractional processes and apply them to the rough Bergomi model, resulting in a multi-factor stochastic volatility model. This approach facilitates the development of a practical simulation scheme employing Gaussian quadrature and Cholesky decomposition, and allows us to address the portfolio optimization problem within a Markovian context. We solve the optimization problem using the Hamilton-Jacobi-Bellman equation, deriving an implicit solution for the case where volatility and stock return are driven by correlated Brownian motions, and providing an explicit solution for the case where they are uncorrelated. The validity of these results is further confirmed through a numerical study.

# Acknowledgements

First, I would like to express my gratitude to Professor Fenghui Yu, my daily supervisor, for her availability and advice throughout the past academic year. I am deeply indebted to Professor Giorgia Callegaro for her encouragement and insightful feedback during our online meetings. My sincere thanks also go to Professor Kristin Kirchner and Professor Antonis Papapantoleon for their assistance with technical matters related to their areas of expertise. Each of these individuals has played a role in the development of this thesis, and their support and guidance are deeply appreciated. I would also like to thank the NAG library for generously providing me with a complimentary license, which enabled me to utilize specialized algorithms that greatly facilitated the implementation of thesis code. Lastly, I extend my deepest thanks to my mum for her unwavering encouragement, and to my dear friends from the department of mathematics, whose companionship and support have made the process of writing this thesis both enjoyable and rewarding.

# Contents

# Introduction

A fundamental problem in mathematical finance is determining an optimal trading strategy that maximizes an agent's expected utility from terminal wealth. Merton's pioneering work [49] addressed this problem within a continuous-time framework, considering a financial market with one riskless asset and one risky asset following geometric Brownian motion. Using stochastic control methods, Merton derived closed-form solutions for both the value function and the optimal portfolio strategy, assuming a utility function of constant relative risk aversion (CRRA) type. His methodology involved transforming the original stochastic optimal control problem into solving a nonlinear deterministic partial differential equation (PDE), namely the Hamilton-Jacobi-Bellman (HJB) equation.

Following Merton's seminal paper, the portfolio optimization problem has been extensively studied in financial markets subject to imperfections. For example, Karatzas et al. [39] introduced the possibility of bankruptcy; Fleming and Zariphopoulou [23] examined the impact of transaction costs; and Grossman and Zhou [28] explored portfolio constraints. In these studies, stock prices follow a geometric Brownian motion, and the authors characterize the value function as the unique smooth solution to the associated nonlinear HJB equation.

A key aspect of the portfolio optmization problem is modeling the behavior of the underlying risky asset. Empirical studies have highlighted the stochastic nature of volatility, leading to extensive investigation into portfolio optimization under stochastic volatility and semimartingale models. To accommodate more general price processes and relax the Markov assumption required by the Bellman approach, *martingale methods* were introduced by Karatzas et al. [40]. However, characterizing the optimal portfolio within this framework requires a representation theorem for martingales, with explicit solutions typically available only in specific cases. Moreover, in the context of incomplete markets, one typically needs to solve a dual optimization problem, which, within the Markovian framework, also yields a nonlinear partial differential equation.

Although Markovian models of volatility have traditionally been the focus ([23], [28], [39], [40] to name just a few), recent evidence suggests that non-Markovian models offer a more accurate description of market data. Gatheral et al. [25] demonstrated that realized volatility fluctuates more rapidly than Brownian motion across various time scales and markets, with log-volatility dynamics resembling those of a fractional Brownian motion with a small Hurst parameter $H$ of order 0.1. Additionally, the implied volatility skew for short maturities is notably steeper than that predicted by classical diffusion models. As maturity decreases, the slope of the at-the-money implied volatility skew follows a power law that diverges at zero. This behavior is accurately reproduced by volatility models involving the fractional kernel $K_H(t-s) = (t-s)^{H-1/2}$ with Hurst index $H < 1/2$, commonly referred to as rough stochastic volatility models [8], [20], [24].

These empirical findings are further substantiated by micro-structural analyses, which show that rough volatility models naturally arise as scaling limits of micro-structural pricing models incorporating self-exciting features driven by Hawkes processes [21]. This understanding has spurred the exploration and development of various rough volatility models, including the rough fractional stochastic volatility model [25], the rough Bergomi model [8], and the fractional and rough Heston models [18], [21].

A pertinent question is how the roughness of market volatility influences investment demands. Models that incorporate rough volatility, due to the presence of the fractional kernel, are generally neither Markovian nor semimartingales, which complicates both their theoretical analysis and practical tasks such as simulations and derivative pricing. Consequently, the literature on portfolio optimization under rough volatility models is still in its early stages. Recent research has started to address this gap, focusing on the rough Heston volatility model [5], [31]. These studies examine a financial market consisting of a bond $S^0$ and one risky asset $S$ respectively evolving according to

$$\begin{aligned} dS_t^0 &= r_t S_t^0, \quad S_0^0 > 0, \\ dS_t &= S_t(r_t + \lambda V_t)dt + S_t\sqrt{V_t}dB_t^S, \quad S_0 > 0, \end{aligned} \tag{1}$$

where $r_t > 0$ is the deterministic bounded risk-free rate, the parameter $\lambda \neq 0$, and $B^S = (B_t^S)_{t\geq 0}$ is standard Brownian motion. The variance process $V = (V_t)_{t\geq 0}$ is a continuous, adapted $\mathbb{R}_+$-valued process. It is generally non-Markovian and not a semimartingale, with its specific form dependent on the particular version of the rough Heston model being considered.

In [31], Han and Wong consider a Volterra Heston model for the variance process

$$V_t = V_0 + k\int_0^t K(t-s)(\phi - V_s)ds + \int_0^t K(t-s)\sigma\sqrt{V_s}dB_s^V, \tag{2}$$

where $K$ denotes the convolution kernel, $B^V = (B_t^V)_{t\geq 0}$ is a standard Brownian motion correlated to $B^S$, and the parameters $V_0, k, \phi, \sigma$ are positive constants. The rough Heston model presented in [21] is a special case of (2) with $K(\tau) \coloneqq \tau^{H-1/2}/\Gamma(H+1/2)$, $H \in (0, 1/2)$ and $\Gamma$ being the Gamma function. To address the challenges posed by the non-Markovian and non-semimartingale nature of the Volterra Heston model, the authors utilize the martingale optimality principle [40] and devise a trial solution, known as *Ansatz*, drawing on both the martingale distortion transformation [63] and exponential-affine representations [3]. They present explicit solutions for the case of correlated Brownian motions between stock and volatility.

The importance of this approach lies in its applicability beyond Markovian processes, effectively addressing the challenges posed by the non-Markovian nature of the Volterra Heston model, which limits the use of the HJB equation. However, increasing model complexity with this method proves challenging, as it heavily depends on specific parameter definitions and assumptions during solution derivation, ultimately limiting model flexibility. Bäuerle and Desmettre [5] propose a more adaptable approach to accommodate additional complexities, such as drift uncertainty and multi-factor models.

In their work, Bäuerle and Desmettre [5] consider a variance process of the form $V = a(\nu)$ where $a : \mathbb{R} \to \mathbb{R}_+$ is a sufficiently smooth function and the process $\nu = (\nu_t)_{t\geq 0}$ is based on the Marchaud fractional derivative:

$$\nu_t = \nu_0 + Z_t \frac{t^{-2H}}{\Gamma(1-2H)} + \frac{2H}{\Gamma(1-2H)} \int_0^t \frac{Z_t - Z_s}{(t-s)^{2H+1}} ds, \tag{3}$$

with

$$dZ = k(\theta - Z_t)dt + \sigma\sqrt{Z_t} dB_t^Z, \quad Z_0 \geq 0.$$

Here, $B^Z = (B_t^Z)_{t\geq 0}$ is again a standard Brownian motion correlated to $B^S$ and the parameters $\nu_0, k, \theta, \sigma$ are positive constants. Since the Marchaud fractional derivative is defined for Hölder $\alpha$-continuous functions with $2H < \alpha \leq 1$, and the paths of $Z$ are almost surely $\alpha$-Hölder continuous with $\alpha < 1/2$ [5, Lemma 7.1], the Hurst index $H$ is restricted to the interval $(0, 1/4)$. This interval still covers the most relevant case, $H \approx 0.1$, as found in empirical studies [8], [25].

To solve the portfolio optimization problem using the HJB equation, Bäuerle and Desmettre consider a Markovian representation of the fractional part of the process $\nu$ (3). Inspired by theoretical works on the approximation of fractional processes [15], [16], [33], they approximate this Markovian representation effectively. The solution to the original optimization problem is then obtained as the limit of the approximated problem.

Fractional processes, including fractional Brownian motion and Riemann-Liouville processes, can be represented as linear functionals of infinite-dimensional Markov processes. Carmona and Coutin introduced a practical and versatile method for approximating these processes within a Markovian framework [15]. Their approach is based on the observation that the fractional kernel $K_H$ associated with these processes is completely monotone on $(0, \infty)$, i.e. $K_H$ is infinitely differentiable on $(0, \infty)$ and satisfies $(-1)^n K_H^{(n)} \geq 0$ for all $n \geq 0$.

According to the Bernstein-Widder theorem (Theorem A.2), a completely monotone kernel $\tilde{K}$ can be represented through its Laplace transform. This implies the existence of a measure $\mu$ on the positive half-line such that:

$$\tilde{K}(\tau) = \int_0^\infty e^{-x\tau} \mu(dx), \quad \tau > 0. \tag{4}$$

This representation is particularly relevant for the fractional kernel $K$ characterizing rough volatility models, where $K(\tau) = \tau^{H-1/2}/\Gamma(H+1/2)$ with $H \in (0, 1/2)$. Let us now consider a Riemann-Liouville process $B^H = (B_t^H)_{t\geq 0}$. Substituting the Laplace transform of the fractional kernel and applying the Stochastic Fubini theorem (Theorem A.1) yields:

$$\begin{aligned} B_t^H = \int_0^t K(t-s) dB_s &= \int_0^t \int_0^\infty e^{-x(t-s)} \mu(dx) dB_s \\ &= \int_0^\infty \int_0^t e^{-x(t-s)} dB_s \mu(dx) = \int_0^\infty Y_t^x \mu(dx), \end{aligned} \tag{5}$$

where $Y^x$ is an Ornstein-Uhlenbeck process. Therefore, (5) expresses $B^H$ as a mixture of Ornstein-Uhlenbeck processes, each with speed of mean reversion $x \in (0, \infty)$ and dynamics:

$$dY_t = -xY_t dt + dB_t, \quad Y_0 = 0. \tag{6}$$

The representation of $B^H$ given in (5) is a Markovian representation, meaning that $B^H$ is Markov in the infinite dimensional process $Y_t = (Y_t^x)_{t\geq 0, x>0}$. This representation lends itself to a finite-dimensional approximation. The approach involves approximating the nonnegative measure $\mu$, which has infinite support, by a finite weighted sum of Dirac measures: $\hat{\mu} = \sum_{i=1}^N w_i \delta_{x_i}$ where $(w_i)_{i=1}^N, (x_i)_{i=1}^N$ are positive weights and nodes, respectively. As a result, we can approximate the integral representation of completely monotone kernels given in (4) with:

$$\hat{K}(\tau) = \sum_{i=1}^N w_i e^{-x_i \tau}. \tag{7}$$

Consequently, the Riemann-Liouville process $B^H$ in (5) can be approximated by:

$$\begin{aligned} \hat{B}_t^H = \int_0^t \hat{K}(t-s) dB_s &= \int_0^t \sum_{i=1}^N w_i e^{-x_i(t-s)} dB_s \\ &= \sum_{i=1}^N w_i \int_0^t e^{-x_i(t-s)} dB_s = \sum_{i=1}^N w_i Y_t^{x_i}. \end{aligned}$$

In this case, with $\mu$ finitely supported by $N$ points $(x_1, \ldots, x_N)$, the process $\hat{B}^H = (\hat{B}_t^H)_{t\geq 0}$ is Markov in $N$-dimensions $(Y^1, \ldots, Y^N)$. Note also that the factors $(Y^i)_{i=1}^N$ share the same dynamics (6) but mean-revert at different speeds $(x_i)_{i=1}^N$.

The convergence of $\hat{B}^H$ to $B^H$, as the number of factors $N$ increases, relies on how well the approximated kernel $\hat{K}$ converges to the true kernel $K$. This, in turn, is influenced by the choice of nodes $(x_i)_{i=1}^N$ and weights $(w_i)_{i=1}^N$. One effective method for constructing such point sets and corresponding weights was proposed by Carmona and Coutin [16] and has been adopted in subsequent works [2], [4], [5]. They approach the problem by truncating the positive half-line, where the measure $\mu$ is defined, to a finite interval $[\xi_0, \xi_N]$. Within this interval, they generate the auxiliary node set $(\xi_i)_{i=0}^N$ by further dividing $[\xi_0, \xi_N]$ into geometrically spaced subintervals. The weights and nodes are then determined as follows:

$$w_i = \int_{\xi_{i-1}}^{\xi_i} \mu(dx), \quad x_i = \frac{1}{w_i} \int_{\xi_{i-1}}^{\xi_i} x\mu(dx), \quad \text{for } i \in \{1, \ldots, N\}.$$

Alternatively, Harms [32] employs a Gaussian quadrature rule of level $m$ for spatial approximation. By considering a geometric sequence $(\xi_i)_{i=0}^N$, Harms uses Gaussian quadrature to determine the $m$ nodes $(x_j)_{j=1}^m$ and the corresponding weights $(w_j)_{j=1}^m$ such that

$$\hat{K}(\tau) = \sum_{j=1}^m w_j e^{-x_j \tau} \quad \approx \quad K(\tau) = \int_{\xi_i}^{\xi_{i+1}} e^{-x\tau} \mu(dx), \tag{8}$$

where $i = 0, \ldots, n-1$. This method achieves strong convergence rates of arbitrarily high polynomial order, with the discretization error being at most of order $n^{-2Hm/3}$, where $n = N/m$ denotes the number of spatial quadrature intervals.

However, as noted by Bayer and Breneis [6], although Harms achieves convergence of arbitrary order, selecting a sufficiently large value for $m$ is crucial to attain a significant order when $H$ is small, such as $H \approx 0.1$. This requirement may introduce additional challenges, as the constants in the error bound are likely to increase with $m$. Furthermore, Harms does not specify which value of $m$ should be chosen based on a given $N$, and his results are demonstrated specifically for the Riemann-Liouville process $B^H$ in (5) rather than more general fractional processes.

Motivated by these limitations, Bayer and Breneis [6] propose a refined version of the point set used in [32] to approximate $K$ with $\hat{K}$ as in (8). They achieve strong convergence results with an error bound of the form $\exp(-c\sqrt{N})$, where $c > 0$, and $N$ is the total number of nodes. Their method extends to a broader class of processes, specifically solutions of stochastic Volterra equations of the form:

$$V_t = V_0 + \int_0^t \tilde{K}(t-s)b(X_s)ds + \int_0^t \tilde{K}(t-s)\sigma(X_s)dB_s, \tag{9}$$

where $V_0 \in \mathbb{R}^d$, the coefficients $b : \mathbb{R}^d \to \mathbb{R}^d$ and $\sigma : \mathbb{R}^d \to \mathbb{R}^{d\times d}$ are Lipschitz continuous, $\tilde{K}$ is a completely monotone convolution kernel and $B$ is a $d$-dimensional standard Brownian motion. They obtain strong convergence results for approximating $\tilde{K}$ with $\hat{K}$, and thus for approximating $V$ in (9) with $\hat{V}$ as established in [4, Prop. 3.2]. They consider both the case of the fractional kernel $K(\tau) = \tau^{H-1/2}/\Gamma(H+1/2)$ [6, Theorem 2.1] and more general completely monotone kernels [6, Theorem 2.11].

Stochastic Volterra equations of the form (9) include the specific case of the Riemann-Liouville process (5). Furthermore, strong uniqueness results for the equation (9) are guaranteed [2, Lemma 5.29], as are convergence results for its approximation, such as those presented in [6, Theorem 2.1]. These results are particularly pertinent for the rough Bergomi model, where the volatility depends on a Riemann-Liouville process. In this model, the stock price evolves according to the dynamics specified in (1), and the volatility process is given by:

$$V_t = \xi(t)\exp\left(\eta\sqrt{2H}\int_0^t (t-s)^{H-1/2}dB_s^V - \frac{\eta^2}{2}t^{2H}\right),$$

where $\xi(t)$ denotes the forward variance curve, $\eta > 0$, $H \in (0, 1/2)$, and $B^V$ is a standard Brownian motion correlated with $B^S$, the Brownian motion driving the stock price.

In the case of the rough Heston volatility (2), the presence of the square root term precludes the process from satisfying Lipschitz conditions. So far, only weak solutions for the Volterra square root process have been established [2, Theorem 6.11]. Nonetheless, the rough Heston model is classified as an affine process. Specifically, as demonstrated in [21], the characteristic function of $\log(S_t/S_0)$ can be expressed in terms of the solution to a fractional Riccati equation:

$$\mathbb{E}\exp(z\log(S_t/S_0)) = \exp\left(\int_0^t F(z, \psi(t-s, z))g(s)ds\right),$$

where

$$g(t) = V_0 + \int_0^t K(t-s)\theta(s)ds,$$

and $\psi(\cdot, z)$ is the unique continuous solution to the fractional Riccati equation

$$\psi(t,z) = \int_0^t K(t-s)F(z,\psi(s,z))ds, \quad t \in [0,T], \tag{10}$$

with

$$F(z,x) \coloneqq \frac{1}{2}(z^2 - z) + (\rho\nu z - \lambda)x + \frac{\nu^2}{2}x.$$

Given that equation (10) needs to be solved numerically, only a semi-explicit formula for the characteristic function of $\log(S_t/S_0)$ is available. By approximating the fractional kernel $K$ with a kernel $\hat{K}$ of the form (7), Jaber [1] developed a multi-factor Markovian approximation of the rough Heston model, termed the *lifted Heston model*. For comprehensive details on existence, uniqueness, and convergence results of this model, we refer the reader to [2, Chapter 7].

Both the rough Bergomi and rough Heston models can be analyzed within the general framework of stochastic Volterra equations. By employing the Markovian approximation previously discussed, simplified multi-factor models can be constructed for both, which facilitate tasks such as simulation, option pricing, and solving stochastic control problems.

The objective of this thesis is to address the portfolio allocation problem within a financial market featuring a bond $S^0$ with a deterministic, bounded risk-free rate $r_t$, and a risky asset $S$, evolving according to

$$\begin{aligned} dS_t^0 &= r_t S_t^0, \quad S_0^0 > 0, \\ dS_t &= S_t\mu_t dt + S_t\sqrt{V_t}dB_t^S, \quad S_0 > 0, \end{aligned}$$

where $\mu_t$ is a predictable bounded process and $B^S$ is standard Brownian Motion. The stock price volatility $V$ is a rough Bergomi volatility of the form

$$V_t = V_0 \exp\left(\eta\sqrt{2H}\int_0^t (t-s)^{H-1/2}dB_s^V - \frac{\eta^2}{2}t^{2H}\right),$$

where the parameters $V_0, \eta > 0$, $H \in (0, 1/2)$, and $B^V$ is Brownian motion correlated to $B^S$. We apply the Markovian approximation to the process $V$, developing a multi-factor model, which allows us to use the HJB equation in the resulting finite-dimensional Markovian framework, as outlined in [5].

We address the portfolio problem for power utility functions, by leveraging the martingale distortion representation, detailed in [63], to express the value function in terms of the solution to the a linear parabolic equation. When the stock price and rough volatility are driven by correlated Brownian motion, we derive an implicit solution to the approximated optimization problem, which requires to be solved numerically. Conversely, in the uncorrelated case, we derive an explicit solution (Theorems 4.1 and 4.2). The solution to the original problem, featuring rough Bergomi volatility, is then obtained as the limit of the approximated problem (Theorem 4.3).

Regarding the numerical experiments, we adopt an $m$-point Gaussian quadrature to determine the positive weights and nodes for the fractional kernel approximation, following [6, Theorem 2.1]. The system of equations, derived from the Markovian approximation, is discretized in time and then simulated using Cholesky decomposition. We conduct simulations of the rough Bergomi model under various parameter settings and visualize the optimal terminal wealth in the uncorrelated scenario, illustrating the effect of rough volatility on the optimal terminal wealth process.

This thesis makes two primary contributions. First, it addresses the portfolio optimization problem under the rough Bergomi model, solving it through the HJB equation. The rough Bergomi model is chosen because existing literature on portfolio optimization has solely focused on the rough Heston model. The optimization problem is approached using a Markovian approximation of the rough model for its flexibility and tractability. This approximation is widely applicable to stochastic Volterra equations and, for rough volatility models, enables the development of manageable multi-factor models that are well-suited for simulations, pricing, hedging, and stochastic control.

Second, in both our theoretical results and numerical experiments related to the control problem, we use a Gaussian quadrature rule to approximate the fractional kernel. Gaussian quadrature provides fast convergence rates and is especially effective for approximating complete monotone kernels.

The thesis is structured as follows. Chapter 1 lays the groundwork for portfolio optimization within a Markovian framework. It starts with a review of stochastic differential equations (SDEs) and the Feynman-Kac formula, tailored for financial applications. The chapter then delves into controlled diffusion processes and employs partial differential equation techniques, particularly the HJB equation, to tackle stochastic control problems. It transitions to portfolio optimization, presenting it as an application of optimal stochastic control theory in finance. Additionally, the chapter outlines the market model considered in the thesis and concludes by presenting the martingale distortion transformation, which simplifies the second-order nonlinearity of the HJB equation for certain utility functions.

Chapter 2 delves into the Markovian representation and approximation of fractional Brownian motion (fBm), which is essential for developing rough volatility models. The chapter introduces fBm, covering its properties and integral representations, with a focus on the Mandelbrot-van Ness representation and Riemann-Liouville processes. It further explores the infinite-dimensional Ornstein-Uhlenbeck processes required to exploit the Markovian representation of fBm. Lastly, the chapter addresses the approximation of stochastic Volterra equations with completely monotone kernels, including a detailed error analysis and the application of Gaussian quadrature to specific fractional kernels.

Chapter 3 explores rough volatility models, emphasizing their effectiveness in capturing the fluctuations of realized volatility and the stylized features of the implied volatility surface seen in financial markets. It begins with an overview of these models and then examines the rough Bergomi model, addressing the challenges posed by its non-semimartingale and non-Markovian characteristics. We derive a Markovian representation

and approximation for the rough Bergomi model and present a comprehensive simulation scheme. This scheme employs Gaussian quadrature to approximate the fractional kernel and uses Cholesky decomposition for simulation. We validate its accuracy through comparisons with established literature. The chapter concludes with visualizations of the simulations across different parameter settings.

In Chapter 4, we address the portfolio optimization problem within a financial market that includes a bond and one risky asset modelled with the rough Bergomi model. We solve this problem by considering the multi-factor rough Bergomi model, derived from the Markovian approximation, which allows us to apply partial differential equation techniques effectively. The chapter presents the solution to the finite-dimensional optimization problem using the HJB equation, presenting both an implicit solution for the case where the stock and volatility Brownian motions are correlated and an explicit solution for the uncorrelated scenario. We then derive the solution to the original rough volatility problem as the limit of the finite-dimensional approximation. The chapter concludes with simulations of the optimal terminal wealth.

We conclude by highlighting the contributions and implications of this research and suggesting avenues for future exploration. Future studies might investigate a comparative analysis of the rough Heston and rough Bergomi models within the framework of portfolio optimization, utilizing Gaussian quadrature and PDE techniques. Empirical validation with real market data could refine the methodologies introduced in this thesis. Additionally, enhancing Gaussian quadrature convergence rates and exploring alternative numerical techniques could be valuable areas for further investigation.

# 1

# Portfolio Optimization

The objective of this chapter is to provide the essential background knowledge and conceptual framework to address the portfolio optimization problem within a Markovian setting. We will apply these tools to a financial market comprising a risk-free asset and a stock price driven by rough volatility, after developing a Markovian approximation for the volatility process, which is neither Markovian nor a semimartingale. The chapter is organized as follows: the first section addresses stochastic optimization problems in continuous time, while the subsequent section delves into the specifics of the portfolio optimization problem.

In Section 1.1, we revisit the fundamental concepts of stochastic differential equations (SDEs) and introduce a generalized formulation of the Feynman-Kac formula tailored for financial applications, as developed in [34]. We then explore controlled diffusion processes and define the corresponding control problem. To tackle this problem, we utilize partial differential equation (PDE) techniques, focusing specifically on the Hamilton-Jacobi-Bellman (HJB) equation. We derive the HJB equation using the Dynamic Programming Principle (DPP) and present the verification theorem, which, under suitable conditions, establishes the equivalence between the solution to the HJB equation and the solution to the control problem. This theorem provides a systematic method for solving the control problem in a Markovian setting through the HJB equation.

In Section 1.2, we introduce the portfolio optimization problem as a specific financial application of optimal stochastic control theory. We clarify the rationale behind our choice of the financial market model and formulate the control problem to be addressed using the HJB equation. Finally, we present the martingale distrortion transformation from the seminal work of Zariphopoulou [63] that allows us to eliminate the second-order nonlinearity of the HJB equation for specific types of utility functions.

# 1.1. Stochastic Optimal Control

## 1.1.1. Introduction

We now outline the fundamental structure of a stochastic optimization problem in continuous time, with detailed solutions to be provided later. Typically, a stochastic optimization problem is defined by the following features:

- State of the system: we consider a dynamic system characterized by its state at any time and evolving in a probability space $(\Omega, \mathcal{F}, \mathbb{P})$. We denote the state of the system at time $t$ in a world scenario $\omega \in \Omega$ as $X_t(\omega)$. The dynamics of the state system, represented by the mapping $t \to X_t(\omega)$ for all $\omega$, are described through a stochastic differential equation (SDE).
- Control: the evolution of the system's dynamics $t \to X_t$ is affected by a control process $\alpha = (\alpha_t)_{t \geq 0}$, with its value determined at each time $t$ in function of the available information. The control $\alpha$ must adhere to certain constraints and is called an *admissible control*. We denote by $\mathcal{A}$ the set of admissible controls.
- Performance criterion: the objective is to maximize (or minimize) over all admissible controls a functional $J(X, \alpha)$ defined as

$$J(X, \alpha) = \mathbb{E}\left[\int_0^T f(X_t, \alpha_t)dt + g(X_T)\right] \quad \text{on a finite time horizon } T < \infty,$$

and as

$$J(X, \alpha) = \mathbb{E}\left[\int_0^\infty e^{-\beta t} f(X_t, \alpha_t)dt\right] \quad \text{on an infinite time horizon.}$$

The function $f$ represents the running profit, $g$ is a terminal reward function, and $\beta > 0$ is a discounting factor. The maximum value, referred to as the value function, is given by:

$$v = \sup_{\alpha \in \mathcal{A}} J(X, \alpha).$$

The main goal in a stochastic optimization problem is to find the maximizing control process attaining the value function to be determined.

To further elucidate the conditions under which the state process $X$ exists and to provide the framework for solving stochastic control problems, we review some basic results of SDE theory.

## 1.1.2. Strong Solution of SDEs

Let $(\Omega, \mathcal{F}, \mathbb{F} = (\mathcal{F}_t)_{t \in \mathbb{T}}, \mathbb{P})$ be a filtered probability space, where $\mathbb{T} = [0, T]$ and $\mathbb{F}$ satisfies the *usual conditions*. A filtration $\mathbb{F} = (\mathcal{F}_t)_{t \in \mathbb{T}}$ meets the usual conditions if it is *right continuous*, i.e. $\mathcal{F}_{t^+} := \cap_{s \geq t} \mathcal{F}_s = \mathcal{F}_t$, $\forall t \in \mathbb{T}$, and if it is *complete*, i.e. $\mathcal{F}_0$ contains the negligible sets of $\mathcal{F}_T$. The right-continuity of $\mathcal{F}_s$ intuitively means that by observing all

the available information up to time $t$ inclusive; no further information is gained from an infinitesimal observation just beyond $t$. The completion of the filtration means that if an event is impossible, this impossibility is recognized from time 0.

Let $B = (B_t)_{t\in\mathbb{T}}$ be a $d$-dimensional Brownian motion, $B = (B^1, \dots, B^d)$, adapted with respect to $\mathbb{F}$, let also $b : \mathbb{T} \times \mathbb{R}^n \times \Omega \to \mathbb{R}^n$ and $\sigma : \mathbb{T} \times \mathbb{R}^n \times \Omega \to \mathbb{R}^{n\times d}$ be measurable functions. As usual, we suppress the dependence on $\omega$ in the coefficients processes $b$ and $\sigma$ for clarity in notation. Consider then the SDE valued in $\mathbb{R}^n$ of the form

$$dX_t = b(t, X_t)dt + \sigma(t, X_t)dB_t, \quad X_0 = \xi, \tag{1.1}$$

which must be interpreted as the stochastic integral equation (SIE)

$$X_t = \xi + \int_0^t b(s, X_s)ds + \int_0^t \sigma(s, X_s)dB_s, \quad \forall t \in \mathbb{T}. \tag{1.2}$$

We now explain what it means for a stochastic process $X = (X_t)_{t\in\mathbb{T}}$ to be a solution of the SIE in (1.2). For an $n \times 1$ column vector $v$ in $\mathbb{R}^n$, let $|v|$ denote the Euclidean norm of $v$, and for a matrix $n \times d$, let $\|\sigma\|^2$ denote the Hilbert-Schmidt norm, given by $\|\sigma\|^2 = \sum_{i=1}^n \sum_{j=1}^d \sigma_{i,j}^2$.

**Definition 1.1.** *[42, Definition 10.3.1] A jointly measurable stochastic process $X = (X_t)_{t\in\mathbb{T}}$ is called a solution of the SIE* (1.2) *if it satisfies the following conditions:*

*(i) The stochastic process $\sigma(t, X_t)$ belongs to $\mathcal{L}_{ad}(\Omega, L^2([0, T]))$ (see Notation 1.1), so that $\int_0^t \sigma(s, X_s)dB_s$ is an Itô integral for each $t \in \mathbb{T}$.*

*(ii) Almost all sample paths of the stochastic process $b(t, X_t)$ belong to $L^1([0, T])$.*

*(iii) For each $t \in [0, T]$, Equation* (1.2) *holds almost surely.*

**Notation 1.1.** $\mathcal{L}_{ad}(\Omega, L^2([a, b]))$ denotes the space of stochastic processes $f(t, \omega)$ satisfying: $f$ is adapted to the filtration $\mathbb{F}$ and $\int_a^b |f(s)|^2 ds < \infty$ almost surely.

We remark that there exists another concept of solution to the SDE (1.1), called weak, where the filtered probability space and the Brownian motion are part of the unknown of the SDE, in addition to $X$. Thus, a weak solution is not necessarily adapted to the Brownian filtration.

Concerning the coefficients $b$ and $\sigma$, we primarily consider two types of situations:

- $b$ and $\sigma$ are deterministic Borelian functions $b(t, x), \sigma(t, x)$ of $t$ and $x$, and we refer to SDE (1.1) as a diffusion process. If $b$ and $\sigma$ do not depend on time, we refer to SDE (1.1) as a (time-homogeneous) diffusion.
- The random coefficients are of the form $\tilde{b}(x, \alpha_t(\omega)), \tilde{\sigma}(x, \alpha_t(\omega))$, where $\tilde{b}, \tilde{\sigma}$ are deterministic Borelian functions on $\mathbb{R}^n \times A$. Here, $A$ is a subset of $\mathbb{R}^k$, and $\alpha = (\alpha_t)_{t\in\mathbb{T}}$ is a progressively measurable process valued in $A$. In this case, we say that the SDE (1.1) is a controlled diffusion by $\alpha$.

Now, we imppose conditions on the functions $b(t,x)$, $\sigma(t,x)$ in order to ensure the existence of a unique non-explosive solution of the SIE (1.2). The following theorem provides sufficient conditions on $b$ and $\sigma$.

**Theorem 1.1.** *[42, Theorem 10.4.1] Let $b(t,x)$ and $\sigma(t,x)$ be measurable functions. Assume that there exist constants $K, C > 0$ such that for all $t \in \mathbb{T}$ and $x, y \in \mathbb{R}^n$ the Lipschitz and linear growth conditions in $x$ are satisfied, i.e.*

$$|b(t,x) - b(t,y)| + \|\sigma(t,x) - \sigma(t,y)\| \leq K|x-y|, \tag{1.3}$$

$$|b(t,x)| + \|\sigma(t,x)\| \leq C(1+|x|). \tag{1.4}$$

*Let $\xi$ be a random vector which is independent of the $\sigma$-algebra generated by $B$ and such that $\mathbb{E}(|\xi|^2) < \infty$. Then the stochastic integral equation (1.2) has a unique continuous solution $X$.*

**Remark 1.1.** If $b(x)$ and $\sigma(x)$ are time-homogeneous, the linear growth condition (1.4) automatically follows from the Lipschitz condition (1.3). To see this, consider the Lipschitz condition for $b(x)$: $|b(x) - b(y)| \leq K|x-y|$. It implies that

$$|b(x)| \leq |b(x) - b(0)| + |b(0)| \leq K|x| + |b(0)| \leq C(1+|x|),$$

where $C = \max\{K, |b(0)|\}$.

In Theorem 1.1, the linear growth condition (1.4) can be relaxed further. It suffices to require the existence of a real-valued process $\kappa$ such that for all $t \in \mathbb{T}$ and $x \in \mathbb{R}^n$

$$|b(t,x)| + \|\sigma(t,x)\| \leq \kappa_t + K|x|,$$

with

$$\mathbb{E}\left[\int_0^t |\kappa_s|^2 ds\right] < \infty, \quad \forall t \in \mathbb{T}. \tag{1.5}$$

Under the Lipschitz conditions (1.3), a natural choice for $\kappa$ is $\kappa_t = |b(t,0)| + \|\sigma(t,0)\|$ once it satisfies the condition (1.5) since

$$|b(t,x)| + \|\sigma(t,x)\| \leq \kappa_t + |b(t,x) - b(t,0)| + \|\sigma(t,x) - \sigma(t,0)\| \leq \kappa_t + K|x|.$$

Let $x \in \mathbb{R}^n$ and $s \in [0,T]$, we denote by $X^{s,x} = \{X_t^{s,x}, s \leq t \in \mathbb{T}\}$ the strong solution to the SDE (1.1) starting from $x$ at time $s$. When $s = 0$, we simply write $X^x = X^{0,x}$.

We recall that any strong solution to the SDE (1.1) satisfies the Markov property: the future behaviour of the process given what has happened up to time $t$ is the same as the behaviour obtained when starting the process at $X_t$. Before stating the related theorems, we introduce some standard notation used in the context of Markov processes.

Let $\mathbb{Q}^x$ denote the probability law of a given (time-homogeneous) Itô diffusion $(X_t)_{t\in\mathbb{T}}$ when its initial value is $X_0 = x \in \mathbb{R}^n$. The expectation with respect to $\mathbb{Q}^x$ is denoted by $\mathbb{E}^x$. Hence, we have

$$\mathbb{E}^x[f_1(X_{t_1}) \cdots f_k(X_{t_k})] = \mathbb{E}[f_1(X_{t_1}^x) \cdots f_k(X_{t_k}^x)]$$

for all bounded Borel functions $f_1, \ldots, f_k$ and all times $t_1, \ldots, t_k \geq 0$ with $k = 1, 2, \ldots$, where $\mathbb{E}$ denotes the expectation with respect to the probability law $\mathbb{P} = \mathbb{P}^0$ for $(B_t)_{t\in\mathbb{T}}$ when $B_0 = 0$. We let $\mathbb{F} = (\mathcal{F}_t^B)_{t\in\mathbb{T}}$ be the filtration generated by $B$, namely, the $\sigma$-algebra $\mathcal{F}_t^B$ is defined by $\mathcal{F}_t^B = \sigma\{B(s); s \leq t\}$.

**Theorem 1.2.** *[52, Theorem 7.1.2] For any Borelian bounded function $f : \mathbb{R}^n \to \mathbb{R}$ and for $t, h \in \mathbb{T}$, the time-homogeneous Itô diffusion $X$ satisfies the Markov property*

$$\mathbb{E}^x[f(X_{t+h})|\mathcal{F}_t^B](\omega) = \mathbb{E}^x[f(X_{t+h})|X_t](\omega),$$

*and is also stationary Markov*

$$\mathbb{E}^x[f(X_{t+h})|X_t](\omega) = \mathbb{E}^{X_t(\omega)}[f(X_h)].$$

*Here, the r.h.s. means the function $\varphi(y) = \mathbb{E}^y[f(X_h)]$ evaluated at $y = X_t(\omega)$.*

**Remark 1.2.** In Theorem 1.2, $X$ is assumed to be a time-homogeneous Itô diffusions, the general case of time dependent coefficients can be reduced to this situation [52, Chapter 10].

It can be proved that the Itô diffusion $X$ satisfies a stronger version of the Markov property. Loosely speaking, the *strong* Markov property states that Theorem 1.2 continues to hold if the time $t$ is replaced by a finite stopping time $\tau(\omega)$ with respect to the filtration generated by the Brownian motion $B$ [42, Definition 5.4.1]. Let $\mathcal{F}_\tau$ be the $\sigma$-field generated by $\{B_{s\wedge\tau}; s \geq 0\}$.

**Theorem 1.3.** *[52, Theorem 7.2.4] Let $f$ be a bounded Borel function on $\mathbb{R}^n$, $\tau$ a stopping time w.r.t. $\mathcal{F}_t^B = \sigma\{B(s); s \leq t\}$ such that $\tau < \infty$ a.s. Then*

$$\mathbb{E}^x[f(X_{\tau+h})|\mathcal{F}_\tau](\omega) = \mathbb{E}^{X_\tau}[f(X_h)], \quad \textit{for all } h \geq 0.$$

### 1.1.3. Feynman-Kac formula

Consider the SDE (1.1) with deterministic coefficients $b(t, x)$ and $\sigma(t, x)$. For all $t \in \mathbb{T}$, we introduce the second order differential operator $\mathcal{L}_t$, called the *infinitesimal generator of the diffusion process* (1.1):

$$(\mathcal{L}_t\varphi)(x) = b(t, x)D_x\varphi + \frac{1}{2}\mathrm{tr}\Big(\sigma(t, x)\sigma^\mathsf{T}(t, x)D_x^2\varphi\Big), \quad \varphi \in C^2(\mathbb{R}^n),$$

where $D_x$ and $D_x^2$ are the gradient and Hessian operators and $C^k$ denotes the class of functions $f$ whose first $k$ derivatives $f'(x), f''(x), \ldots, f^{(k)}(x)$ all exist and are continuous. The component-wise expression of $\mathcal{L}_t$ is given by:

$$(\mathcal{L}_t\varphi)(x) = \sum_{i=1}^n b_i(t, x)\frac{\partial\varphi}{\partial x_i}(t, x) + \frac{1}{2}\sum_{i=1}^n\sum_{j=1}^m \sigma_{i,j}(t, x)\frac{\partial^2\varphi}{\partial x_i\partial x_j}(t, x), \quad \varphi \in C^2(\mathbb{R}^n). \tag{1.6}$$

Let now $C^{1,2}([0, T) \times \mathcal{O})$ denote the space of real-valued functions $f$ on $[0, T) \times \mathcal{O}$, where $\mathcal{O}$ is an open set of $\mathbb{R}^n$, and whose partial derivatives up to second order exist and are

continuous on $[0,T)$. If these partial derivatives of $f \in C^{1,2}([0,T)\times\mathcal{O})$ can be extended by continuity on $[0,T]\times\mathcal{O}$, we write $f \in C^{1,2}([0,T]\times\mathcal{O})$.

On a finite horizon interval $\mathbb{T}=[0,T]$, we aim to find a (real-valued) function $v(t,x)$ of class $C^{1,2}$ on $\mathbb{T}\times\mathbb{R}^n$ that satisfies the Cauchy problem:

$$\begin{cases} -\dfrac{\partial v}{\partial t} - \mathcal{L}_t v + cv = f, & \text{on } [0,T)\times\mathbb{R}^n, \\ v(T,\cdot) = g, & \text{on } \mathbb{R}^n, \end{cases} \tag{1.7}$$

where $f$ (resp. $g$) is a continuous function from $[0,T]\times\mathbb{R}^n$ (resp. $\mathbb{R}^n$) into $\mathbb{R}$ and $c$ is a continuous non-negative function. The Feynman-Kac formula, exploiting the intimate connection between SDEs and parabolic PDEs, provides the solution to (1.7) in terms of the solution to an SDE naturally associated with the Cauchy problem.

We assume that there actually exists a solution $v$ to (1.7) and fix a point in time $t$ and a point in space $x$. We apply the Itô formula to the process $\tilde{M}_t \coloneqq e^{-c(t,X_t)}v(t,X_t)$:

$$\begin{aligned} \tilde{M}_t = \tilde{M}_0 &+ \int_0^t e^{-\int_0^s c(u,X_u)du}\Big(\frac{\partial v}{\partial t} + \mathcal{L}_s v - cv\Big)(s,X_s)ds \\ &+ \int_0^t e^{-\int_0^s c(u,X_u)du} D_x v^{\intercal}(s,X_s)\sigma(s,X_s)dB_s, \end{aligned}$$

and note that the process

$$\begin{aligned} M_t &\coloneqq \tilde{M}_s - \int_0^t e^{-\int_0^s c(u,X_u)du}\Big(\frac{\partial v}{\partial t} + \mathcal{L}_s v - cv\Big)(s,X_s)ds \\ &= v(0,X_0) + \int_0^t e^{-\int_0^s c(u,X_u)du} D_x v^{\intercal}(s,X_s)\sigma(s,X_s)dB_s \end{aligned} \tag{1.8}$$

is a continuous local martingale. If the integrand of the stochastic integral in (1.8) is sufficiently integrable and we take expected values on both sides of the equation, the stochastic integral will vanish. The initial value $X_s = x$ and the boundary condition $v(T,x) = g(x)$ will eventually leave us with the Feynman-Kac representation formula. The following theorem presents a simplified version of it.

**Theorem 1.4.** *[53, Theorem 1.3.17] Let $v$ be a function $C^{1,2}([0,T)\times\mathbb{R}^d)\cap C^0([0,T]\times\mathbb{R}^d)$ with bounded derivative in $x$ and solution to the Cauchy problem* (1.7)*. Then $v$ admits the representation*

$$v(t,x) = \mathbb{E}\left[g(X_T^{t,x})e^{-\int_t^T c(u,X_u^{t,x})du} + \int_t^T e^{-\int_t^s c(u,X_u^{t,x})du} f(s,X_s^{t,x})ds\right] \tag{1.9}$$

*for all $(t,x)\in[0,T]\times\mathbb{R}^d$.*

The proof of the theorem relies on the observation that when $v$ has a bounded derivative in $x$, the integrand of the stochastic integral (1.8) lies in $\mathcal{L}_{ad}(\Omega, L^2[0,T])$ (see Notation 1.1). Hence, $M$ is a (square integrable) martingale and the representation (1.9) is simply derived by writing that $\mathbb{E}[M_T] = \mathbb{E}[M_t]$. We may also obtain this Feynman-Kac representation under other conditions on $v$, for example with $v$ satisfying a quadratic

growth condition. We will show such a result in the more general case of controlled diffusion, see Remark 1.4.

The application of Theorem 1.4 requires the existence of a smooth solution $v$ to the Cauchy problem (1.7). This type of result is typically obtained under an assumption of *uniform ellipticity* on the operator $\mathcal{L}_t$, boundedness conditions on $b, \sigma$, and polynomial growth condition on $f$ and $g$. However, many financial applications do not satisfy these restrictive assumptions imposed by standard results. For instance, $b$ and $\sigma$ may be unbounded or grow faster than linearly.

In this context, we present a broader-reaching result by Heath and Schweizer [34] which provides sufficient conditions on $X$, its domain, and coefficients, as well as on the functions $g, c$ and $f$, to guarantee that the function $v$ satisfies (1.7) and admits representation (1.9).

The ensuing theorem combines analytic and probabilistic assumptions that still allow us to derive (1.7) while being general enough to be applicable in a number of applications.

**Theorem 1.5.** *[34, Theorem 1] Let $\mathbb{T} = [0, T]$ be a fixed time horizon and $D$ a domain in $\mathbb{R}^n$, i.e., an open connected subset of $\mathbb{R}^n$. Suppose that the following conditions hold:*

($A1$) *The coefficients $b$ and $\sigma$ of the process $X$ with dynamics 1.1 are on $[0, T] \times D$ locally Lipschitz-continuous in $x$, uniformly in $t$, i.e., for each compact subset $F$ of $D$, there is a constant $K_F < \infty$ such that*

$$|b(t, x) - b(t, y)| \leq K_F|x - y|, \quad \|\sigma(t, x) - \sigma(t, y)\| \leq K_F|x - y|$$

*for all $t \in [0, T]$ and $x, y \in F$.*

($A2$) *For all $(t, x) \in [0, T) \times D$, the solution $X$ of (1.1) neither explodes nor leaves $D$ before $T$, i.e., $\mathbb{P}[\sup_{0<s<T} |X_s| < \infty] = 1 \ \wedge \ \mathbb{P}[X_s \in D, \forall s \in [0, T]] = 1$.*

($A3$) *There exists a sequence $(D_n)_{n\in\mathbb{N}}$ of bounded domains contained in $D$ such that $\cup_{n=1}^{\infty} D_n = D$ and such that for each $n$, the PDE*

$$-\frac{\partial G}{\partial t} - \mathcal{L}_t G + cG = f, \quad \text{on } (0, T) \times D_n,$$

*with boundary condition $G(t, x) = v(t, x)$ on $(0, T) \times \partial D_n \cup T \times \overline{D}_n$ has a classical solution $G_n(t, x)$.*

*Then $v$ satisfies the Cauchy problem (1.7). In particular $v \in C^{1,2}$ and there exists a unique classical solution to (1.7).*

To make ($A3$) more platable one can verify a set of conditions ($A3'$). Before introducing them, we provide the definition of Hölder continuity that will be employed.

**Definition 1.2.** *A function $f(x)$ defined on a bounded closed set $S$ of $\mathbb{R}^n$ is said to be Hölder continuous of exponent $\alpha$ $(0 < \alpha < 1)$ if there exists a constant $K$ such that*

$$|f(x) - f(y)| \leq K|x - y|^{\alpha} \tag{1.10}$$

*for all $x, y \in S$. The smallest $K$ for which (1.10) holds is called the Hölder coefficient.*

*If the function $f$ depends on a parameter $t$, i.e., $f = f(t, x)$, and the Hölder coefficient is independent of $t$, then we say that $f(t, x)$ is Hölder continuous in $x$, uniformly with respect to $t$.*

($A3'$) There exists a sequence $(D_n)_{n\in\mathbb{N}}$ of bounded domains with $\overline{D}_n \subseteq D$ such that $\cup_{n=1}^{\infty} D_n = D$, each $D_n$ has a $C^2$-boundary and for each $n$,

- ($A3'a$) $b$ and $\sigma$ are on $[0, T] \times \overline{D}_n$ Lipschitz-continuous in $x$, uniformly in $t$,
- ($A3'b$) $a(t, x) = \sigma(t, x)\sigma^{\mathsf{T}}(t, x)$ is uniformly elliptic on $\mathbb{R}^n$ for $(t, x) \in [0, T) \times D_n$, i.e., $\exists\, \varepsilon_n > 0$ such that $y^{\mathsf{T}} a(t, x) y \geq \varepsilon_n |y|^2$ for all $y \in \mathbb{R}^n$,
- ($A3'c$) $c$ is on $[0, T] \times \overline{D}_n$ Hölder-continuous in $x$, uniformly in $t$,
- ($A3'd$) $f$ is on $[0, T] \times \overline{D}_n$ Hölder-continuous in $x$, uniformly in $t$,
- ($A3'e$) $v$ is finite and continuous on $[0, T] \times \partial D_n \cup \{T\} \times \overline{D}_n$.

**Remark 1.3.** In cases where a smooth solution to the Cauchy problem (1.7) does not exist, the PDE can be formulated using the concept of a weak solution, known as a *viscosity solution*. The theory of viscosity solutions is beyond the scope of this thesis, and we refer the reader to [53, Chapter 4] for further reading.

### 1.1.4. Controlled Diffusion Processes

Consider a control model where the state of the system is governed by a stochastic differential equation (SDE) valued in $\mathbb{R}^n$:

$$dX_t = b(X_t, \alpha_t)dt + \sigma(X_t, \alpha_t)dB_t, \tag{1.11}$$

where $b : \mathbb{R}^n \times A \to \mathbb{R}^n, \sigma : \mathbb{R}^n \times A \to \mathbb{R}^{n\times d}$ and $B$ is a $d$-dimensional Brownian motion on the filtered probability space $(\Omega, \mathcal{F}, \mathbb{F} = (\mathcal{F}_t)_{t\in\mathbb{T}}, \mathbb{P})$, $\mathbb{T} = [0, T]$. Here, the control $\alpha$, which is valued in $A \subset \mathbb{R}^k$, is a parameter whose value we can choose in the given Borel set $A$ at any instant $t$ in order to control the process $X$. We denote by $a$ the value of the control at a certain time, and with $\alpha$ the mapping $\alpha_t = \alpha(t, X_t)$.

Our first modeling problem concerns the class of admissible control processes. Since our decision at time $t$ must be based upon what has happened up to time $t$, the function $\omega \to \alpha(t, \omega)$ must (at least) be measurable with respect to $\mathbb{F}$, i.e., the process $\alpha$ must be adapted. We will require a strictly stronger property that is $\alpha$ *progressively measurable*.

**Definition 1.3.** *[53, Definition 1.1.3] A process $(X_t)_{t\in\mathbb{T}}$ is progressively measurable (with respect to $\mathbb{F}$) if for any $t \in \mathbb{T}$, the mapping $(s, \omega) \to X_s(\omega)$ is measurable on $[0, t] \times \Omega$ equipped with the product $\sigma$-field $\mathcal{B}([0, t]) \otimes \mathcal{F}_t$.*

One natural way to obtain an adapted control process is by choosing a measurable function $g : \mathbb{T} \times \mathbb{R}^n \to A$ and then defining the control process $\alpha$ by $\alpha_t = g(t, X_t^{s,x})$. For mnemo-technical purposes it is common to denote control laws by $\alpha(t, x)$, rather than $g(t, x)$, and write $\alpha_t = \alpha(t, X_t)$. In this case we say that $\alpha$ is a Markovian control because

with such $\alpha$ the corresponding process $X$ (1.11) becomes an Itô diffusion, in particular, a Markov process [52]. Indeed, $\alpha$ does not depend on the starting point $y = (t, x)$; the value we choose at time $t$ only depends on the state of the system at that time.

We can now define the class of admissible control laws.

**Definition 1.4.** *$\mathcal{A}$ is a class of admissible control laws if*

*(i) $\alpha$ is a progressively measurable (with respect to $\mathbb{F}$) process, valued in $A \subset \mathbb{R}^k$.*

*(ii) For any given initial point $(t, x)$ the SDE (1.11) has a unique solution.*

For the remainder of this section, we require the coefficients $b$ and $\sigma$ to satisfy a uniform Lipschitz condition in $A$: $\exists\, K > 0$ such that for all $x, y \in \mathbb{R}^n$ and $a \in A$,

$$|b(x, a) - b(y, a)| + \|\sigma(x, a) - \sigma(y, a)\| < K|x - y|. \tag{1.12}$$

Moreover, we denote by $\mathcal{A}$ the set of control processes $\alpha$ such that

$$\mathbb{E}\left[\int_0^T |b(0, \alpha_t)|^2 + \|\sigma(0, \alpha_t)\|^2 \, ds\right] < \infty. \tag{1.13}$$

Conditions (1.12) and (1.13) are the counterpart of conditions (1.3) and (1.5), thus, they ensure for all $\alpha \in A$ and for any initial condition $(t, x) \in \mathbb{T} \times \mathbb{R}^n$, the existence and uniqueness of a strong solution to the SDE (with random coefficients) (1.11) starting from $x$ at $s = t$. Consequently, $\mathcal{A}$ is a class of admissible control laws.

We then denote by $X_s^{t,x}, t \le s \le T$ the solution with a.s. continuous paths starting from $x$ at $s = t$ and recall that under the specified conditions on $b, \sigma$ and $\alpha$, we have

$$\mathbb{E}\left[\sup_{t \le s \le T} |X_s^{t,x}|^2\right] < \infty. \tag{1.14}$$

We now proceed to define the objective function of the control problem. Consider as given a pair of measurable functions $f : [0, T] \times \mathbb{R}^n \times A \to \mathbb{R}$ and $g : \mathbb{R}^n \to \mathbb{R}$. We suppose that:

(**Hg**) $(i)$ $g$ is lower-bounded
or $(ii)$ $g$ satisfies a quadratic growth condition: $|g(x)| \le C(1 + |x|^2)$, $\forall x \in \mathbb{R}^n$, for some constant $C$ independent of $x$.

For $(t, x) \in [0, T] \times \mathbb{R}^n$, we denote by $\mathcal{A}(t, x)$ the subset of controls $\alpha$ in $\mathcal{A}$ such that

$$\mathbb{E}\left[\int_t^T \left|f(s, X_s^{t,x}, \alpha_s)\right| ds\right] < \infty,$$

and we assume that $\mathcal{A}(t, x)$ is not empty for all $(t, x) \in [0, T] \times \mathbb{R}^n$. Under (**Hg**) the performance function is defined by:

$$J(t, x, \alpha) = \mathbb{E}\left[\int_t^T f(s, X_s^{t,x}, \alpha_s) ds + g(X_T^{t,x})\right],$$

for all $(t,x) \in [0,T] \times \mathbb{R}^n$ and $\alpha \in \mathcal{A}(t,x)$. The objective is to maximize over admissible control processes the performance function $J$. The associated value function is:

$$v(t,x) = \sup_{\alpha \in \mathcal{A}(t,x)} J(t,X,\alpha).$$

The performance function $J(t,x,\alpha)$ represents the expected utility of employing the control law $\alpha$ over the time interval $[t,T]$, starting from state $x$ at time $t$. The optimal value function provides the optimal expected utility over $[t,T]$ under the same initial conditions. Equivalently, the value function represents the maximum value of the objective functional, varying with the initial states.

For an initial condition $(t,x) \in [0,T) \times \mathbb{R}^n$, we designate $\alpha^* \in \mathcal{A}(t,x)$ as the optimal control if

$$v(t,x) = J(t,x,\alpha^*).$$

We also note that, as for any optimization problem, the optimal law may not exist; see [52, Example 11.2.6].

### 1.1.5. The Hamilton-Jacobi-Bellman Equation

Given an optimal control problem, two fundamental questions emerge: does an optimal control law exist? Secondly, assuming its existence, how can it be determined? We will primarily focus on the second question, utilizing dynamic programming, which is a fundamental tool in the context of controlled diffusion processes and, more generally, for controlled Markov processes. For a fixed pair $(t,x)$ we now define the following control problem.

**Definition 1.5.** *The control problem $\mathcal{P}(t,x)$ is defined as the problem to maximize*

$$J(t,x,\alpha) = \mathbb{E}\left[\int_t^T f(s, X_s^{t,x}, \alpha_s)ds + g(X_T^{t,x})\right] \tag{1.15}$$

*over $\alpha$, given the dynamics of $X^{t,x}$ and the constraints $\alpha \in \mathcal{A}(t,x)$, $\forall (s,y) \in [t,T] \times \mathbb{R}^n$.*

Observe that we use the notation $s$ and $y$ above because the letters $t$ and $x$ are already used to denote the fixed chosen point.

To solve the control problem $\mathcal{P}(t,x)$ using the dynamic programming method, we first define the value function associated with it. Next, we derive an analytic characterization of the value function in terms of a partial differential equation known as the Hamilton-Jacobi-Bellman (HJB) equation. The control problem is then equivalent to finding a solution to the HJB equation [11, Section 19.3].

We define the value function associated to (1.15) by

$$v(t,x) = \sup_{\alpha \in \mathcal{A}(t,x)} \mathbb{E}\left[\int_t^T f(s, X_s^{t,x}, \alpha_s)ds + g(X_T^{t,x})\right], \quad (t,x) \in [0,T] \times \mathbb{R}^n. \tag{1.16}$$

The derivation of the PDE satisfied by the value function (1.16) relies on the dynamic programming principle (DPP), specifically its infinitesimal version. The DPP essentially

states that: if the state process is optimally controlled from any time $\tau$ ($t \leq \tau \leq T$) until $T$, then optimizing the control from $t$ until $\tau$, will ensure an optimal decision over the entire interval $[t, T]$ [53, Section 3.3]. Notably, the DPP holds for any stopping time $\tau$ valued in $[t, T]$.

**Assumption 1.1.** We assume the following.

1. There exists an optimal control law $\alpha^* \in \mathcal{A}(t, x)$.
2. The optimal value function $v$ is regular; $v \in C^{1,2}([0, T) \times \mathbb{R}^d) \cap C^0([0, T] \times \mathbb{R}^d)$.

We fix $(t, x) \in (0, T) \times \mathbb{R}^n$ and select a real number $h$ such that $t + h < T$. Then, we choose a fixed but arbitrary control law $\alpha \in \mathcal{A}(t, x)$, and define the control law $\bar{\alpha}$ by

$$\bar{\alpha}(s, y) = \begin{cases} \alpha(s, y), & (s, y) \in [t, t + h] \times \mathbb{R}^n, \\ \alpha^*(s, y), & (s, y) \in (t + h, T] \times \mathbb{R}^n. \end{cases}$$

In other words, if we use $\bar{\alpha}$ then we use the arbitrary control $\alpha$ during the time interval $[t, t + h]$, and then we switch to the optimal control law during the rest of the time period.

The infinitesimal version of the DPP boils down to the following procedure.

- First, given the point $(t, x)$ as above, we consider the following two strategies over the time interval $[t, T]$: Strategy I implements the optimal law $\alpha^*$, while Strategy II utilizes the control law $\bar{\alpha}$ defined above.
- We then compute the expected utilities obtained by the respective strategies.
- Finally, using the fact that Strategy I by definition has to be at least as good as Strategy II, and letting $h$ tend to zero, we obtain our fundamental PDE.

We now carry out this program.

The expected utility for Strategy I is straightforward to determine. By definition, the utility is the optimal one, given by $J(t, x, \alpha^*) = v(t, x)$. On the other hand, the expected utility for Strategy II requires a different approach. We divide the time interval $[t, T]$ into two parts, the intervals $[t, t + h]$ and $(t + h, T]$ respectively.

- The expected utility, using Strategy II, for the interval $[t, t + h)$ is given by

$$\mathbb{E}\left[\int_t^{t+h} f(s, X_s^{t,x}, \alpha_s) ds\right].$$

- In the interval $[t + h, T]$ we observe that at time $t + h$ we will be in the (stochastic) state $X_{t+h}$. Since, by definition, we will use the optimal strategy during the entire interval $[t + h, T]$ we see that the remaining expected utility at time $t + h$ is given by $v(t + h, X_{t+h})$. Thus the expected utility over the interval $[t + h, T]$, conditional on the fact that at time $y$ we are in state $x$, is given by

$$\mathbb{E}\left[v(t + h, X_{t+h}^{t,x})\right].$$

Thus the total expected utility for Strategy II is

$$\mathbb{E}\left[\int_t^{t+h} f(s, X_s^{t,x}, \alpha_s)ds + v(t+h, X_{t+h}^{t,x})\right].$$

We now go on to compare the two strategies, and since by definition Strategy I is the optimal one, we must have the inequality

$$v(t,x) \geq \mathbb{E}\left[\int_t^{t+h} f(s, X_s^{t,x}, \alpha_s)ds + v(t+h, X_{t+h}^{t,x})\right]. \tag{1.17}$$

We also note that the inequality sign is due to the fact that the arbitrarily chosen control law $\alpha$ which we use on the interval $[t, t+h]$ need not be the optimal one. We have equality in (1.17) if and only if the control law $\alpha$ is an optimal law $\alpha^*$.

Since, by assumption, $v \in C^{1,2}$ we now use the Itô formula to obtain

$$\begin{aligned} v(t+h, X_{t+h}^{t,x}) = v(t,x) &+ \int_t^{t+h} \Big(\frac{\partial v}{\partial t} + \mathcal{L}^a v\Big)(s, X_s^{t,x})ds \\ &+ \int_t^{t+h} D_x v^{\mathsf{T}}(s, X_s^{t,x})\sigma(s, X_s^{t,x})dB_s, \end{aligned} \tag{1.18}$$

where $\mathcal{L}^a$ is the second-order differential operator associated to the diffusion (1.11) for the control $\alpha$

$$(\mathcal{L}^a v)(x) = b(x,a)D_x v + \frac{1}{2}\mathrm{tr}\Big(\sigma(x,a)\sigma^{\mathsf{T}}(x,a)D_x^2 v\Big). \tag{1.19}$$

The assumptions $\alpha \in \mathcal{A}(t,x) \subset \mathcal{A}$ and $v \in C^{1,2}([0,T) \times \mathbb{R}^d)$ ensure that the stochastic integral in (1.18) is a martingale, because the integrability conditions on its integrand are satisfied.

We then insert (1.18) into the inequality (1.17) to obtain

$$0 \geq \mathbb{E}\left[\int_t^{t+h} \Big(\frac{\partial v}{\partial t} + \mathcal{L}^a v\Big)(s, X_s^{t,x}) + f(t, X_s^{t,x}, a)ds\right].$$

We divide it by $h$ and let $h$ tend to 0. Assuming enough regularity to allow us to take the limit within the expectation, using the fundamental theorem of integral calculus, and recalling that $X_t = x$, we get by the mean-value theorem

$$0 \geq \frac{\partial v}{\partial t}(t,x) + \mathcal{L}^a v(t,x) + f(t,x,a),$$

where $a$ denotes the value of the law $\alpha$ evaluated at $(t,x)$, i.e. $a = \alpha(t,x)$.

Since the control law $\alpha$ was arbitrary, this inequality will hold for all choices of $a \in A$, and we will have equality if and only if $a = \alpha^*(t,x)$. We thus have the following equation

$$-\frac{\partial v}{\partial t}(t,x) - \sup_{a \in A}[\mathcal{L}^a v(t,x) + f(t,x,a)] = 0. \tag{1.20}$$

During the discussion the point $(t,x)$ was fixed but chosen arbitrarily; therefore, equation (1.20) holds for all $(t,x) \in [0,T) \times \mathbb{R}^n$. We have a (nonstandard type of) PDE, and we obviously need some boundary conditions. One such condition is easily obtained, recalling that $v(T,x) = g(x)$ for all $x \in \mathbb{R}^n$.

**Theorem 1.6.** *[11, Theorem 19.5] Under Assumption 1.1, the following hold:*

*(i) $v$ satisfies the HJB equation.*

$$\begin{cases} -\dfrac{\partial v}{\partial t}(t,x) - \sup\limits_{a\in A}[\mathcal{L}^a v(t,x) + f(t,x,a)] = 0, \quad \forall (t,x) \in [0,T) \times \mathbb{R}^n, \\ \qquad\qquad\qquad\qquad\qquad\qquad v(T,x) = g(x), \quad x \in \mathbb{R}^n. \end{cases}$$

*(ii) For each $(t,x) \in [0,T] \times \mathbb{R}^n$ the supremum in the HJB equation above is attained by $a = \alpha^*(t,x)$.*

It is important to note that this theorem has the form of a necessary condition. It asserts that if $v$ is the optimal value function and $\alpha^*$ is the optimal control, then $v$ satisfies the HJB equation, and $\alpha^*$ realizes the supremum in the equation.

Equally significant is determining whether this theorem provides also a sufficient condition: if at each point $(t,x)$ we find $a = \alpha(t,x)$ such that $\mathcal{L}^a v(t,x) + f(t,x,a)$ is maximal and this maximum is 0, will $\alpha(t,x)$ be an optimal control? The next result confirms that this is indeed the case.

### 1.1.6. Verification Theorem

The crucial step in the classical approach to dynamic programming involves proving that, given a sufficiently regular solution to the HJB equation, this candidate solution coincides with the value function. This result, known as the verification theorem, also allows us to identify an optimal Markovian control as a byproduct.

The specific assertions regarding the sufficient conditions may vary slightly from one problem to another and should be adapted to the context of the particular problem under consideration. We present a general version of the verification theorem.

**Theorem 1.7.** *[53, Theorem 3.5.2] Let $G$ be a function $C^{1,2}([0,T)\times\mathbb{R}^n) \cap C^0([0,T]\times\mathbb{R}^n)$, and satisfying a quadratic growth condition, i.e. there exists a constant $K$ such that*

$$|G(t,x)| \leq K(1+|x|^2), \quad \forall (t,x) \in [0,T] \times \mathbb{R}^n.$$

*(i) Suppose that $G$ satisfies*

$$\begin{cases} -\dfrac{\partial G}{\partial t}(t,x) - \sup\limits_{a\in A}[\mathcal{L}^a G(t,x) + f(t,x,a)] \geq 0, \quad \forall (t,x) \in [0,T) \times \mathbb{R}^n, \\ \qquad\qquad\qquad\qquad\qquad\qquad G(T,x) \geq g(x), \quad x \in \mathbb{R}^n. \end{cases} \tag{1.21}$$

*Then $G \geq v$ on $[0,T] \times \mathbb{R}^n$.*

*(ii) Suppose further that $G(T,\cdot) = g$ and there exists a measurable function $\alpha^*(t,x)$, $(t,x) \in [0,T)\times\mathbb{R}^n$, valued in $A$. For each fixed $(t,x)$ the supremum in the expression*

$$-\frac{\partial G}{\partial t}(t,x) - \sup_{a\in A}[\mathcal{L}^a G(t,x) + f(t,x,a)]$$

*is attained by a, the value of $\alpha^*(t,x)$ at time t, and the SDE*

$$dX_s = b(X_s, \alpha^*(s, X_s))ds + \sigma(X_s, \alpha^*(s, X_s))dB_s,$$

*admits a unique solution, denoted by $\hat{X}^{t,x}$ given an initial state $(t,x)$. Then, the optimal value function v to the control problem is given by*

$$v(t,x) = G(t,x) \quad on\ [0,T] \times \mathbb{R}^n,$$

*and $\alpha^*$ is an optimal Markovian control.*

*Proof.*

(*i*) Since $w \in C^{1,2}([0,T) \times \mathbb{R}^n)$, we have for all $(t,x) \in [0,T) \times \mathbb{R}^n, \alpha \in \mathcal{A}(t,x)$, $s \in [t,T)$, and any stopping time $\tau$ valued in $[t,\infty)$, by Itô's formula

$$\begin{aligned} G(s \wedge \tau, X^{t,x}_{s\wedge\tau}) = G(t,x) &+ \int_t^{s\wedge\tau} \Big(\frac{\partial G}{\partial t} + \mathcal{L}^{a_u} G\Big)(u, X^{t,x}_u)du \\ &+ \int_t^{s\wedge\tau} D_x G^{\mathsf{T}}(u, X^{t,x}_u)\sigma(X^{t,x}_u, \alpha_u)dB_u. \end{aligned}$$

We choose $\tau = \tau_n = \inf\{s \geq t : \int_t^s |D_x G^{\mathsf{T}}(u, X^{t,x}_u)\sigma(X^{t,x}_u, \alpha_u)|^2 du \geq n\}$, and we notice that $\tau_n \nearrow \infty$ when $n$ goes to infinity. The stopped process

$$\left\{\int_t^{s\wedge\tau} D_x G^{\mathsf{T}}(u, X^{t,x}_u)\sigma(X^{t,x}_u, \alpha_u)dB_u, t \leq s \leq T\right\}$$

is then a martingale, and by taking the expectation, we get

$$\mathbb{E}[G(s \wedge \tau_n, X^{t,x}_{s\wedge\tau})] = G(t,x) + \mathbb{E}\left[\int_t^{s\wedge\tau_n} \Big(\frac{\partial G}{\partial t} + \mathcal{L}^{a_u} G\Big)(u, X^{t,x}_u)du\right].$$

Since $G$ satisfies (1.21), we have

$$\frac{\partial G}{\partial t}(t, X^{t,x}_u) + \mathcal{L}^a G(t, X^{t,x}_u) + f(X^{t,x}_u, \alpha_u) \leq 0, \quad \forall \alpha \in \mathcal{A}(t,x),$$

and so

$$\mathbb{E}[G(s \wedge \tau_n, X^{t,x}_{s\wedge\tau})] \leq G(t,x) - \mathbb{E}\left[\int_t^{s\wedge\tau_n} f(X^{t,x}_u, \alpha_u)du\right], \quad \forall \alpha \in \mathcal{A}(t,x). \tag{1.22}$$

We have

$$\left|\int_t^{s\wedge\tau_n} f(X^{t,x}_u, \alpha_u)du\right| \leq \int_t^T |f(X^{t,x}_u, \alpha_u)|du,$$

and the right-hand-side term is integrable by the integrability condition on $\mathcal{A}(t,x)$. Since $G$ satisfies a quadratic growth condition, we have

$$|G(s \wedge \tau, X^{t,x}_{s\wedge\tau})| \leq C(1 + \sup_{s\in[t,T]} |X^{t,x}_s|^2),$$

and the right-hand-side term is integrable from (1.14). We can then apply the dominated convergence theorem, and send $n$ to infinity into (1.22):

$$\mathbb{E}[G(s, X_s^{t,x})] \leq G(t,x) - \mathbb{E}\left[\int_t^s f(X_u^{t,x}, \alpha_u)du\right], \quad \forall \alpha \in \mathcal{A}(t,x).$$

Since $G$ is continuous on $[0,T] \times \mathbb{R}^n$, by sending $s$ to $T$, we obtain by the dominated convergence theorem and by (1.21)

$$\mathbb{E}[g(X_t^{t,x})] \leq G(t,x) - \mathbb{E}\left[\int_t^T f(X_u^{t,x}, \alpha_u)du\right], \quad \forall \alpha \in \mathcal{A}(t,x).$$

From the arbitrariness of $\alpha \in \mathcal{A}(t,x)$, we deduce that $G(t,x) \leq v(t,x)$, for all $(t,x) \in [0,T] \times \mathbb{R}^n$.

(*ii*) We apply Itô's formula to $G(u, \hat{X}_u^{t,x})$ between $t \in [0,T)$ and $s \in [t,T)$ (after an eventual localization for removing the stochastic integral term in the expectation):

$$\mathbb{E}[G(s, \hat{X}_s^{t,x})] = G(t,x) + \mathbb{E}\left[\int_t^s \big(\frac{\partial G}{\partial t} + \mathcal{L}^{\alpha^*(t,x)} G\big)(u, \hat{X}_u^{t,x})du\right].$$

Now, by definition of $\alpha(t,x)$, we have

$$\frac{\partial G}{\partial t} + \mathcal{L}^{\alpha^*(t,x)} G(t,x) + f(t,x,\alpha^*(t,x)) \leq 0, \quad \forall \alpha \in \mathcal{A}(t,x),$$

and so

$$\mathbb{E}[G(s, \hat{X}_s^{t,x})] = G(t,x) - \mathbb{E}\left[\int_t^s f(\hat{X}_u^{t,x}, \alpha^*(u, \hat{X}_u^{t,x}))du\right].$$

By sending $s$ to $T$, we then obtain

$$G(t,x) = \mathbb{E}\left[\int_t^s f(\hat{X}_u^{t,x}, \alpha^*(u, \hat{X}_u^{t,x}))du + g(\hat{X}_T^{t,x})\right] = J(t,x,\alpha^*).$$

This shows that $G(t,x) = J(t,x,\alpha^*) \leq v(t,x)$, and finally that $G = v$ with $\alpha^*$ as an optimal Markovian control.

□

**Remark 1.4.** In the particular case where the control space A is reduced to a singleton $\{a_0\}$, this verification theorem is a version of the Feynman-Kac formula: it states that if $G$ is a function in $C^{1,2}([0,T) \times \mathbb{R}^d) \cap C^0([0,T] \times \mathbb{R}^d)$ with a quadratic growth condition and is the solution to the Cauchy problem (1.7), then $G$ admits the representation

$$G(t,x) = \mathbb{E}\left[\int_t^T f(X_s^{t,x}, \alpha_0)ds + g(X_T^{t,x})\right].$$

Theorem 1.7 suggests a method for solving the optimization problem 1.5. Laid out explicitly, the approach can be delineated as follows.

1. Consider the HJB equation as a PDE for an unknown function $G$:

$$\begin{cases} -\dfrac{\partial G}{\partial t}(t,x) - \sup\limits_{a\in A}[\mathcal{L}^a G(t,x) + f(t,x,a)] = 0, \quad \forall (t,x) \in [0,T) \times \mathbb{R}^n, \\ G(T,x) = g(x), \quad x \in \mathbb{R}^n. \end{cases} \tag{1.23}$$

2. Fix an arbitrary point $(t,x) \in [0,T] \times \mathbb{R}^n$ and solve, for this fixed choice of $(t,x)$, the static optimization problem

$$\max_{a\in A}[\mathcal{L}^a G(t,x) + f(t,x,a)].$$

   Note that in this problem $a$ is the only variable, whereas $t$ and $x$ are considered to be fixed parameters. The functions $f, \mu, \sigma$ and $G$ are considered as given.

3. The optimal choice of $a$, denoted by $a^*$, will depend on our choice of $t$ and $x$, but it will also depend on the function $G$ and its partial derivatives. To highlight these dependencies we write $a^*$ as $\alpha^* = \alpha^*(t,x;G)$.

4. The function $\alpha^*(t,x;G)$ is our candidate for the optimal control law, but since we do not know $G$ this description is incomplete. Therefore, we substitute the expression for $\alpha^*$ into the HJB equation (1.23).

5. We solve the resulting PDE (see Remark 1.5) and substitute the solution $G$ into expression $\alpha^*(t,x;G)$. Using the verification Theorem 1.7 we can now identify $G$ as the optimal value function $v$, and $\alpha^*$ as the optimal control law.

**Remark 1.5.** The essence of dynamic programming involves addressing the highly nonlinear partial differential equation outlined in step 5 above. Due to the absence of general analytic methodologies, instances of optimal control problems with an analytically derived solution are exceedingly rare. In practice, one typically endeavors to propose a solution by formulating trial answer, commonly referred to as *Ansatz*, for $G$, parameterized by a finite set of parameters. Subsequently, the partial differential equation is employed to determine these parameters. The formulation of an Ansatz, is often facilitated by the intuitive observation that if an analytical solution exists, $G$ likely exhibits certain structural attributes inherited from both the boundary function $g$ and the instantaneous utility function $f$ [11, Remark 19.4.1].

## 1.2. The Portfolio Optimization Problem

### 1.2.1. Introduction

Consider a financial market comprising a risk-less asset with price process $S^0$, representing a savings account, and $m$ risky assets with price process $S$, representing the stocks. An agent can invest in this market at any time $t$, holding a number of shares $\alpha_t$ in the $m$ risky assets. The agent's decisions does not affect market prices, thus categorizing him as a "small investor". By denoting the agent's wealth at time $t$ as $W_t$, the number of

shares invested in the savings account at time $t$ can be calculated as $(W_t - \alpha_t S_t)/S_t^0$. The *self-financed* wealth process evolves according to:

$$dW_t = (W_t - \alpha_t S_t)\frac{dS_t^0}{S_t^0} + \alpha_t dS_t, \quad W_0 > 0. \tag{1.24}$$

Here, the control process, denoted by $\alpha$, is valued in $A$, a subset of $\mathbb{R}^m$. The portfolio allocation problem involves choosing the optimal investment in the financial market.

Classical models for describing the behavior and preferences of agents and investors include the mean-variance and expected utility criteria. The mean-variance criterion, introduced by Markowitz [47], assumes that an agent's preferences depend solely on the expectation and variance of his random incomes. The criterion's advantage lies in its simplicity and minimal requirement of probability knowledge, specifically expectations and covariances of random variables. However, it is criticized for measuring risk solely based on the variance of portfolio returns. The symmetric nature of variance limits both potential losses and gains. To address this issue, it is more appropriate to consider asymmetric preferences, which are typically represented using utility functions according to the von Neumann-Morgenstern theory [51].

In the expected utility criterion, the agent evaluates random incomes with known probability distributions. The agent's preferences are expressed through the expectation of a utility function, denoted by $U$. A random income $W$ is preferred to a random income $\overline{W}$ if

$$\mathbb{E}[U(W)] \geq \mathbb{E}[U(\overline{W})].$$

The utility function $U$ is non-decreasing, reflecting a preference for greater wealth, and concave, indicating the agent's risk aversion. For a risk-averse agent, the certainty of receiving $\mathbb{E}[W]$, the expected value of a random return $W$, is preferred over the uncertain return itself. This preference is formalized by Jensen's inequality, $U(\mathbb{E}[W]) \geq \mathbb{E}[U(W)]$, which is valid exclusively for concave functions.

For a risk-averse agent with concave utility function $U$, the *risk premium* associated with a random portfolio return $W$ is defined as the positive amount the agent is willing to pay to secure a certain gain. This premium satisfies: $U(\mathbb{E}[W] - \pi) = \mathbb{E}[U(W)]$. Let $\mu = \mathbb{E}[W]$ denote the expected return. If the portfolio return $W$ has low risk, the risk premium $\pi$ can be approximated by:

$$\pi \approx -\frac{1}{2}\frac{U''(\mu)}{U'(\mu)}\mathrm{Var}(W) = \frac{1}{2}\eta(\mu)\mathrm{Var}(W),$$

where $\eta(x) = -U''(x)/U'(x)$ is the local *absolute risk aversion* at the return level $x$ [54]. This approximation indicates that the variance of a portfolio return is a key indicator of its risk, with $\eta(\mu)$ representing the factor by which an economic agent with utility function $U$ weights the risk.

If we express the random return $W$ as $W = \mu(1 + \varepsilon)$, where $\varepsilon$ represents the relative deviation from the expected return, the *relative risk premium* $\rho$ is defined such that

$U(\mu(1-\rho)) = \mathbb{E}[U(W)] = \mathbb{E}[U(\mu(1+\varepsilon))]$. The relative risk premium $\rho$ indicates the proportion of the return that the investor is willing to forgo to obtain a certain gain. An approximation for $\rho$ is given by:

$$\rho \approx \frac{1}{2}\gamma(\mu)\mathrm{Var}(\varepsilon),$$

where $\gamma(x) = -xU''(x)/U'(x)$, is the *relative risk aversion* at level $x$ [54].

Two commonly studied classes of utility functions that illustrate these concepts are those with Constant Absolute Risk Aversion (CARA) and Constant Relative Risk Aversion (CRRA). In the CARA framework, the absolute risk aversion $\eta(x)$ is constant, $\eta > 0$, and the utility function can be represented as $U(x) = 1 - e^{-\eta x}$. Under CRRA, the relative risk aversion $\gamma(x)$ is constant, where $\gamma \in [0, 1)$ and utility function is given by:

$$U(x) = \begin{cases} x^\gamma/\gamma, & 0 < \gamma < 1, \\ \ln x, & \gamma = 0. \end{cases} \tag{1.25}$$

These utility functions adhere to the Inada conditions, which stipulate that the utility function $U : (0, \infty) \to \mathbb{R}$ is continuously differentiable, strictly increasing and strictly concave on $(0, \infty)$, and satisfies:

$$U'(0) \coloneqq \lim_{x \downarrow 0} U'(x) = \infty, \quad U'(\infty) \coloneqq \lim_{x \to \infty} U'(x) = 0.$$

Note that the decreasing marginal utility $U'(x)$ (due to the strict concavity of $U$) indicates that the agent is risk-averse, while the assumption $U'(0^+) = \infty$ guarantees that the constraint $W_T \geq 0$ on terminal wealth will never be active.

The use of stochastic control and dynamic programming methods for continuous-time portfolio optimization was pioneered by Merton in his seminal work [49]. He applied techniques and results from stochastic control theory to a financial setting, using the HJB equation to derive explicit solutions for specific cases.

In Merton's model, the bond price $S^0$ grows at the constant interest rate $r > 0$, and the stock price $S$ evolves according to the Black-Scholes model:

$$\begin{aligned} dS_t^0 &= rS_t^0 dt, \quad S_0^0 > 0, \\ dS_t &= \mu S_t dt + \sigma S_t dB_t, \quad S_0 > 0, \end{aligned}$$

where $\mu, \sigma > 0$ are constants, and $B$ is a Brownian motion on a filtered probability space. The non-negative wealth process $W$ is controlled by the proportion $\alpha$, invested in the risky asset and valued in $\mathbb{R}$, and is governed by the diffusion dynamics (1.24). The investor's objective is to maximize over portfolio strategy $\alpha$ his expected utility of terminal wealth at a finite horizon $T$. The value function for this control problem is defined by

$$v(t, w) = \sup_{\alpha \in \mathcal{A}} \mathbb{E}[U(W_T)],$$

where $U$ is an increasing and concave function on $\mathbb{R}_+ \coloneqq [0, \infty)$ and $w = W_t$.

Merton considered both CRRA and CARA utility functions. However, as he noted in his work, CARA utility functions are "behaviorially less plausible than constant relative risk aversion" [49]. Unlike the CRRA case where the proportion of wealth allocated to the risky asset remains constant, the optimal portfolio for a CARA function is inversely proportional to wealth. Thus, in the limit, as $W \to \infty$, one invests all wealth in the risk-free asset.

### 1.2.2. Motivations for Model Choice

We consider a financial market consisting of one risk-free asset and one risky asset. This model can be extended to a market with $m$ risky assets. The coefficients of the stock price process depend on $N$ stochastic processes, denoted in vector form as $\mathbf{Y} = (Y^1, \dots, Y^N)$, which we will refer to as the stochastic factor. This factor is driven by a single Brownian motion that is correlated with the Brownian motion driving the stock price. Each $Y^i$ with $i = 1, \dots, N$ is a diffusion process.

As the stochastic coefficients in the stock process are influenced by the latent factor $\boldsymbol{Y}$, the considered financial market is incomplete. Incompleteness arises when the number of stocks is strictly less than the number of Brownian motions ($m < d$); in our case $m = 1$ and $d = 2$. Under this condition, it is typically impossible to construct a portfolio consisting of the bond and the $m$ available stocks that can fully hedge the risk associated with these coefficient processes. In our setting, the stochastic factor is a non-tradable instrument that prevents the investor from effectively hedging risk.

A pivotal consideration revolves around the information available at decision time points. It seems natural to examine a market where agents can solely observe the stock price process $S$ only. The stock price process contains enough information to filter the evolution of the Brownian motion driving the stock price and $\boldsymbol{Y}$ from it. This stems from the fact that the *quadratic variation* of the stock price process (1.28)

$$\langle S, S \rangle_t = \int_0^t S_u^2 \sigma^2(u, \boldsymbol{Y}_u) du \tag{1.26}$$

can be observed and thus, if $\sigma$ is bijective, $\boldsymbol{Y}$ can also be inferred (for more details, see [55, Lemma 3.1]). Consequently, we assume that the agent knows the evolution of the Brownian motion driving the stock price and $\boldsymbol{Y}$ up to time $t$. Or, what is equivalent, at time $t$ he knows his current wealth and $\boldsymbol{Y}$.

Individual preferences are modeled using a power utility function $U(x) = x^\gamma/\gamma$, with trading occurring over a finite horizon. The power utility function is chosen because, in the limit case when the risk-aversion parameter $\gamma \to 0$, it converges to the log utility function $U(x) = \log(x)$, see (1.25). As already mentioned, the power utility is a CRRA function and, as such, is considered more plausible than CARA functions. Moreover, It allows us to express the value function as a specific power of the solution to a linear parabolic equation. As we will discuss, this exponent, known as the distortion power, depends solely on the risk aversion coefficient and the correlation between the Brownian motions driving the stock price and the stochastic factors.

### 1.2.3. The Financial Market

Consider a financial market under $(\Omega, \mathcal{F}, \mathbb{F} = (\mathcal{F}_t)_{0\leq t\leq T}, \mathbb{P})$, with $T < \infty$, consisting of one risk-less asset $S^0$ and one risky asset $S$ respectively evolving according to

$$dS_t^0 = r_t S_t^0 dt, \quad S_0^0 > 0, \tag{1.27}$$
$$dS_t = \mu(t, \boldsymbol{Y}_t)S_t dt + \sigma(t, \boldsymbol{Y}_t)S_t dB_t^S, \quad S_0 > 0, \tag{1.28}$$

where $r_t > 0$ is a deterministic bounded risk-free rate and $\boldsymbol{Y} = (Y^1, \dots, Y^N)$ is a stochastic factor assumed to satisfy

$$d\boldsymbol{Y}_t = c(t, \boldsymbol{Y}_t)dt + d(t, \boldsymbol{Y}_t)dB_t^Y, \quad \boldsymbol{Y}_0 \in \mathbb{R}^N. \tag{1.29}$$

The processes $B^S$ and $B^Y : [0, T] \to \mathbb{R}$ are correlated $\mathbb{F}$-Brownian motions with correlation coefficient $-1 < \rho < 1$. The initial prices $S_0^0, S_0$ and the initial values $\boldsymbol{Y}_0$ are deterministic constants. The coefficients $\mu, \sigma, c, d$ are functions of time and the factor $\boldsymbol{Y}$, and they are assumed to satisfy all the required regularity assumptions to guarantee the existence of a unique solution to (1.28) and (1.29).

It is reasonable to assume that at time $t$ the agent has knowledge of the stock price and, and, consequently, its quadratic variation (1.26). Thus, at time $t$, the agent knows the evolution of $B^S$ and $\boldsymbol{Y}$ up to that point. Therefore, we assume that the filtration $\mathbb{F}$ is generated by $B^S$ and $B^Y$, i.e. $\mathcal{F}_t = \sigma(B_s^S, B_s^Y, s \leq t)$.

The investor rebalances his portfolio dynamically by choosing at any time $t$, for $t \in [0, T]$, a proportion $\alpha_t$ of his wealth to be invested in the stock account $S$. $1 - \alpha_t$ is then the fraction of wealth invested in the bond $S^0$. Note that $\alpha_t$ is not restricted to the interval $[0, 1]$: $\alpha_t < 0$ means that the stock is sold short and $\alpha_t > 1$ means that money is borrowed from the bank at the interest rate $r$. The investor faces the portfolio constraint that at any time $t$, $\alpha_t$ is valued in $A$, a closed convex subset of $\mathbb{R}$. We denote by $\mathcal{A}$ the set of $\mathbb{F}$-progressively measurable processes $\alpha$ valued in $A$, and such that

$$\int_0^T \alpha_t^2 \sigma^2(t, \boldsymbol{Y}_t)dt < \infty \quad a.s.$$

This integrability condition ensures the existence and uniqueness of a strong solution to the SDE governing the wealth process controlled by $\alpha \in \mathcal{A}$ (cf. Definition 1.4).

Given a portfolio strategy $\alpha \in \mathcal{A}$, we denote by $W^{t,w}$ the corresponding wealth process starting from an initial capital $W_t = w > 0$ at time $t$, evolving according to:

$$\begin{aligned} dW_s &= W_s \alpha_s \frac{dS_s}{S_s} + W_s(1 - \alpha_s)\frac{dS_s^0}{S_s^0} \\ &= W_s[r_s + \alpha_s(\mu(s, \boldsymbol{Y}_s) - r_s)]ds + W_s \alpha_s \sigma(s, \boldsymbol{Y}_s)dB_s^S \end{aligned} \tag{1.30}$$

for $t \leq s \leq T$. This linear stochastic differential equation can be solved explicitly and the solution for $t \in [0, T]$ is given by

$$W^{t,w} = w \exp\left(\int_t^T \left[r_s + \alpha_s(\mu(s, \boldsymbol{Y}_s) - r_s) - \frac{1}{2}\sigma^2(s, \boldsymbol{Y}_s)\alpha_s^2\right]ds + \int_t^T \sigma(s, \boldsymbol{Y}_s)\alpha_s dB_s^S\right).$$

The investor's objective is to maximize his expected utility of terminal wealth:

$$J(t, w, \boldsymbol{y}) = \mathbb{E}_{t,w,\boldsymbol{y}}[U(W_T, \boldsymbol{Y}_T)],$$

where by $\mathbb{E}_{t,w,\boldsymbol{y}}$ we denote the conditional expectation, emphasizing that at time $t$, we are given $W_t = w > 0$ and $\boldsymbol{Y}_t = \boldsymbol{y} \in \mathbb{R}^N$. The value function of the investor for $t \in [0, T]$ is then defined by

$$v(t, w, \boldsymbol{y}) = \sup_{\alpha \in \mathcal{A}} \mathbb{E}_{t,w,\boldsymbol{y}}[U(W_T, \boldsymbol{Y}_T)].$$

The HJB equation together with the terminal condition of the stochastic control problem for a sufficiently regular function $G \in C^{1,2}([0, T) \times \mathbb{R}_+ \times \mathbb{R}^N) \cap C^0([0, T] \times \mathbb{R}_+ \times \mathbb{R}^N))$ are:

$$\begin{cases} -\dfrac{\partial G}{\partial t}(t, w, \boldsymbol{y}) - \sup\limits_{a \in A}[\mathcal{L}^a G(t, w, \boldsymbol{y})] = 0 \quad \forall (t, w, \boldsymbol{y}) \in [0, T) \times \mathbb{R}_+ \times \mathbb{R}^N, \\ G(T, w, \boldsymbol{y}) = U(w, \boldsymbol{y}), \quad w \in \mathbb{R}_+, \boldsymbol{y} \in \mathbb{R}^N. \end{cases} \tag{1.31}$$

Here, $\mathcal{L}^a$ is the operator associated with the state process $X = (W, \boldsymbol{Y})$ for the control $\alpha$ (cf. (1.19)). The state process $X$ has as its first component the controlled diffusion $W$, while the remaining $N$ components are the diffusion processes $Y^i$ for $i = 1, \ldots, N$. For clarity we rewrite the dynamics of $X$ in matrix form

$$\begin{aligned} d\begin{bmatrix} W_s \\ \boldsymbol{Y}_s \end{bmatrix} = &\begin{bmatrix} W_s[r_s + \alpha_s(\mu(s, \boldsymbol{Y}_s) - r_s] \\ c(s, \boldsymbol{Y}_s) \end{bmatrix} ds \\ &+ \begin{bmatrix} W_s \alpha_s \sigma(s, \boldsymbol{Y}_s) & 0 \\ d(s, \boldsymbol{Y}_s)\rho & d(s, \boldsymbol{Y}_s)\sqrt{1-\rho^2} \end{bmatrix} \begin{bmatrix} dB^1 \\ dB^2 \end{bmatrix}, \end{aligned}$$

where we identify the Brownian motion driving the wealth process $W$ with $B^S \equiv B^1$ and the Brownian motion driving the stochastic factor $\boldsymbol{Y}$ with $B^Y \equiv \rho B^1 + \sqrt{1-\rho^2} B^2$, where $B^1, B^2$ are independent.

We omit the verification that $B^1$ and $B^2$ are Brownian motions and instead demonstrate that they are independent. We begin by noting that the differential forms of $B^S$ and $B^V$ are

$$dB^S = dB^1, \quad dB^V = \rho dB^1 + \sqrt{1-\rho^2} dB^2.$$

Therefore,

$$dB^S dB^V = \rho dt + \sqrt{1-\rho^2} dB^1 dB^2.$$

Since $B^S$ and $B^V$ are correlated with coefficient $\rho$, i.e. $dB^S dB^V = \rho dt$, the term $\sqrt{1-\rho^2} dB^1 dB^2$ must vanish. Given that $\rho \in (-1, 0) \cup (0, 1)$, it follows that $dB^1 dB^2 = 0$. Thus, $B^1$ and $B^2$ are uncorrelated, and since uncorrelated Gaussian random variables are independent, we conclude that $B^1$ and $B^2$ are independent.

The HJB equation reads in this case

$$\begin{aligned} 0 = G_t + \sup_{a \in A} \Big\{ &w[r + a(\mu(t, \boldsymbol{y}) - r)] G_w + \sum_{i=1}^N c(t, y_i) G_{y_i} \\ &+ \frac{1}{2} w^2 a^2 \sigma^2(t, \boldsymbol{y}) G_{ww} + \sum_{i=1}^N w a \sigma(t, \boldsymbol{y}) d(t, y_i) \rho G_{y_i w} + \frac{1}{2} \sum_{i=1}^N \sum_{j=1}^N d^2(t, y_i) G_{y_i y_j} \Big\}. \end{aligned} \tag{1.32}$$

The goal herein is to find a smooth solution $G$ to (1.31), verify that it coincides with the value function $v$, and determine the optimal investment strategies $\alpha^*$ when the utility function is of power utility type

$$U(x) = \frac{x^\gamma}{\gamma}, \quad x \geq 0, \gamma \in (0,1).$$

The specific form of these utilities, combined with the linearity of the wealth dynamics with respect to the state and control processes (see (1.30)), allows us to represent the value function in a *separable* form. Specifically, the value function can be written as $v(t,w,\boldsymbol{y}) = w^\gamma/\gamma\ \Phi(t,\boldsymbol{y})$, where $\Phi$ is generally unknown, as it solves a nonlinear equation for which no closed-form solutions are available. However, with a simple transformation, it is possible to eliminate certain non-linearities that arise due to the stochastic factor, as the following result will demonstrate.

### 1.2.4. The Distortion Power

When considering exponential, power, or logarithmic utility functions, the investor's utility takes the form $U(W,\boldsymbol{Y}) = U(W)h(\boldsymbol{Y})$ here $W$ is the investor's wealth and $\boldsymbol{Y}$ is a random factor, not perfectly correlated with the market [62]. The value function becomes separable as well: $v(t,w,\boldsymbol{y}) = U(w)\Phi(t,\boldsymbol{y})$.

Remarkably, it is found that $\Phi$ can be represented as a power $\delta$ of the solution to a linear parabolic equation. The power $\delta$, referred to as the distortion power, depends solely on the risk aversion parameter $\gamma$ and the degree of correlation between the stock price and the stochastic factor. The following proposition by Zariphopoulou [63] addresses this result, specifically for the case where the utility function is $U(w,y) = w^\gamma/\gamma\ h(y)$ and the stochastic factor $Y$ is one-dimensional.

**Proposition 1.** *[63, Proposition 2.1]*

*(i) The value function $v$ is given by*

$$v(t,w,y) = \frac{x^\gamma}{\gamma}\varphi(t,y)^\delta$$

*where $\delta = \frac{1-\gamma}{1-\gamma+\rho^2\gamma}$ and $\varphi : [0,T]\times\mathbb{R} \to \mathbb{R}_+$ solves the linear parabolic equation*

$$\begin{cases} \varphi_t + \frac{1}{2}d^2(s,y)\varphi_{yy} + \Big[c(t,y) + \rho\frac{\gamma(\mu(t,y)-r)d(t,y)}{(1-\gamma)\sigma(t,y)}\Big]\varphi_y \\ \qquad\qquad + \frac{\gamma(1-\gamma+\rho^2\gamma)}{1-\gamma}\Big[r + \frac{(\mu(t,y)-r)^2}{2\sigma^2(t,y)(1-\gamma)}\Big]\varphi = 0 \end{cases}$$

*with terminal condition $\varphi(T,y) = h(y)$.*

*(ii) The optimal policy $\alpha^*$ is given in Markovian form $\alpha^* = \alpha^*(s,\hat{W}_s,Y_s)$, $t \leq s \leq T$, where the function $\alpha^* : [0,T]\times\mathbb{R}_+\times\mathbb{R} \to \mathbb{R}$ is defined by*

$$\alpha^*(t,w,y) = \left[\frac{\rho}{(1-\gamma)+\rho^2\gamma}\frac{d(t,y)\varphi_y(t,y)}{\sigma(t,y)\varphi(t,y)} + \frac{1}{1-\gamma}\frac{\mu(t,y)-r}{\sigma^2(t,y)}\right]w$$

*with $\hat{W}$ being the state wealth process given by* (1.30) *when the policy $\alpha^*$ is being used.*

The above proposition is rigorously proved in [63] using tools from the theories of stochastic control and viscosity solutions of the HJB equation. Here, we outline a procedure on how to prove result $(i)$ when $h(y) \equiv 1$ while considering an $N$-dimensional stochastic factor. This procedure will be adopted in Chapter 4.

1. Consider a candidate solution for (1.32) of the form $G(t, w, \boldsymbol{y}) = \frac{w^\gamma}{\gamma}\Phi(t, \boldsymbol{y})$ satisfying the terminal condition $G(T, \boldsymbol{y}) = 1$, and substitute $G$ directly into Equation (1.32).
2. Formally apply the first-order conditions to the obtained equation to determine $\alpha^*$.
3. Insert $\alpha^*$ and the transformation $\Phi(t, \boldsymbol{y}) = \varphi(t, \boldsymbol{y})^\delta$, where $\varphi$ satisfies $\varphi(T, \boldsymbol{y}) = 1$ and the parameter $\delta$ is not yet determined, into the previously obtained equation.
4. Determine $\delta$ so that the the resulting equation becomes linear parabolic.

# 2

# Markovian Approximation of Fractional Processes

Rough volatility models have garnered significant attention due to their efficacy in replicating both realized volatility and the stylized features of the implied volatility surface observed in financial markets. A stochastic process is termed *rough* if its sample paths are less Hölder continuous than those of Brownian motion. Unlike traditional stochastic volatility models driven by standard Brownian motion, rough volatility models employ fBm (fBm) with a small Hurst index $H$. For $H < 1/2$, fBm is a continuous yet less regular process than standard Brownian motion. Its non-Markovian and non-semimartingale nature further complicates both theoretical analysis and practical applications.

This chapter delves into the Markovian representation (Section 2.1) and approximation (Section 2.2) of fBm, both of which are essential for developing stochastic volatility models that approximate rough volatility models while preserving a Markovian structure. This approximation not only facilitates simulation but also enables the use of PDE techniques to address portfolio optimization problems, as will be further discussed in this thesis.

In Section 2.1, we introduce fBm, detailing its properties and integral representations. We focus particularly on the Mandelbrot-van Ness representation and Riemann-Liouville processes, which are commonly encountered in rough volatility models. We then present infinite-dimensional Ornstein-Uhlenbeck (OU) processes and explore the Markovian representation of Riemann-Liouville processes, which turn out to be a mixture of these infinite-dimensional OU processes, as thoroughly discussed in [33]. Lastly, we provide an overview of the more general case of stochastic Volterra equations, which encompass Riemann-Liouville processes. Under appropriate conditions on their convolution kernel, any solution to the stochastic Volterra equation can be represented as an infinite-dimensional system of processes that share the same Brownian motion and exhibit mean-reverting behavior at different rates [2].

In Section 2.2, we address the approximation of stochastic Volterra equations with completely monotone kernels and Lipschitz coefficients. We start by presenting theoretical results on the error between the stochastic Volterra equation and any multifactor Stochastic Differential Equation (SDE) approximation, drawing from [4]. Following this, we examine the *strong error analysis*, i.e. $L^2$-error, between fractional kernels of the form $K(\tau) = \tau^{H-1/2}/\Gamma(H+1/2)$ with $H \in (0, 1/2)$ and $\Gamma$ being the Gamma function, and their approximation using Gaussian quadrature rules, as detailed in [6].

## 2.1. Markovian Representation of Fractional Processes

### 2.1.1. Introduction

Fractional Brownian motion and the associated Riemann-Liouville processes are long-memory Gaussian processes distinguished by dependent increments and rough sample paths. The non-Markovian and non-semimartingale nature of these processes poses theoretical and practical challenges. To mitigate these issues, it is advantageous to represent them as linear functionals of an infinite-dimensional Markov process (cf. Remark 2.8). This Markovian representation facilitates the development of efficient algorithms for their approximation.

The key idea, originating from Carmona and Coutin [15], is to represent the fractional integral associated with these processes using a Laplace transform. For each $H < 1/2$, by the stochastic Fubini theorem A.1,

$$\int_0^t (t-s)^{H-1/2} dB_s \propto \int_0^t \int_0^\infty e^{-x(t-s)} \frac{dx}{x^{H+1/2}} dB_s = \int_0^\infty \int_0^t e^{-x(t-s)} dB_s \frac{dx}{x^{H+1/2}},$$

where, for each $x$, $Y_t^x = \int_0^t e^{-x(t-s)} dB_s$ is an OU process.

In this section, we will delve into the above representation in detail. We begin by reviewing the key properties that make fBm particularly appealing for financial applications.

### 2.1.2. Fractional Brownian Motion

Fractional Brownian motion (fBm), introduced by Kolmogorov in 1940, is a Gaussian process that generalizes standard Brownian motion. The term "fractional Brownian motion" was coined by Mandelbrot and Van Ness in 1968, when they provided a stochastic integral representation of the process using a two-sided standard Brownian motion [45]. Due to its self-similarity and long-memory properties, fBm is well-suited as input noise in various models where standard Brownian motion falls short.

**Definition 2.1.** *[10, Definition 1.1.1] Let $H$ be a constant belonging to $(0, 1)$. A fractional Brownian motion $(B_t^H)_{t\geq 0}$ of Hurst index $H$ is a continuous and centered Gaussian process with covariance function*

$$\mathbb{E}[B_t^H B_s^H] = \frac{1}{2}\Big(|s|^{2H} + |t|^{2H} - |t-s|^{2H}\Big), \quad s, t \geq 0. \tag{2.1}$$

To specify the distribution of a Gaussian process, it is sufficient to specify its mean and covariance function [19, Chapter 12]. Therefore, for each fixed value of the Hurst parameter $H$, the distribution of $B^H$ is uniquely determined by the above definition. The existence of the fBm follows from the general existence theorem of centered Gaussian processes with given covariance functions [19, Theorem 12.1.3].

Observe that for $H = 1/2$, the covariance function is $\mathbb{E}[B_t^{1/2} B_s^{1/2}] = t \wedge s$, meaning $B^{1/2}$ is a standard Brownian motion. This justifies the name "fractional Brownian motion": $B^H$ is a generalization of Brownian motion obtained by allowing the Hurst parameter to differ from $1/2$.

By Definition (2.1) we obtain that a standard fBm $B^H$ starts from zero, $B_0^H = 0$ almost surely, and enjoys the following properties:

(*i*) Stationary increments. Consider the process $(Y_t)_{t\geq 0} = (B_{t+s}^H - B_s^H)_{t\geq 0}$. Since

$$\mathbb{E}[(B_t^H - B_s^H)^2] = |t-s|^{2H}, \quad s,t \geq 0,$$

it follows that the covariance function of $Y$ is the same as that of $B^H$. As both processes are centered Gaussian, the equality of covariance functions implies that $(Y_t)_{t\geq 0} \overset{d}{=} (B_t^H)_{t\geq 0}$. Thus, the probability distributions of its increments remain invariant under a shift in time, and for this reason, $B^H$ is said to have stationary increments.

(*ii*) Self-similarity. Now consider, for a fixed $a > 0$, the process $(Z_t)_{t\geq 0} = (B_{at}^H)_{t\geq 0}$. From (2.1), it is apparent that $Z$ has the same covariance, and hence the same distribution, as $a^H B^H$. This property is called $H$-self-similarity, indicating the scale-invariance of the process: the probability distribution remains unchanged within each time interval under appropriate spatial scaling.

(*iii*) Dependence of increments. For $H = 1/2$, $B^H$ is a standard Brownian motion; hence, in this case the increments of the process are independent. On the contrary, for $H \neq 1/2$ the increments are not independent. More precisely, by Definition (2.1) we know that the covariance $\mathbb{E}[(B_{t+h}^H - B_t^H)(B_{s+h}^H - B_s^H)]$ with $s + h \leq t$ and $t - s = nh$ is

$$\begin{aligned}\rho_H(n) &= \frac{1}{2}h^{2H}\Big[(n+1)^{2H} + (n-1)^{2H} - 2n^{2H}\Big], \\ &= \frac{1}{2}h^{2H}\Big[\Big((n+1)^{2H} - n^{2H}\Big) - \Big(n^{2H} - (n-1)^{2H}\Big)\Big].\end{aligned}$$

Specifically, noting that the function $g(x) = x^{2H}$ is concave for $H \in (1/2, 1)$ and convex for $H \in (0, 1/2)$, we observe that two increments on disjoint time intervals are positively correlated when $H > 1/2$ and negatively correlated when $H < 1/2$.

The dependence of increments of fBm naturally leads to its long-memory property: the covariance between far apart increments decrease to zero as a power law.

**Definition 2.2.** *[10, Definition 1.4.1] A stationary sequence $(X_n)_{n\in N}$ exhibits long-range dependence if the autocovariance functions $\rho(n) \coloneqq Cov(X_{k+n}, X_k)$ satisfy*

$$\lim_{n\to\infty} \frac{\rho(n)}{cn^{-\alpha}} = 1$$

*for some constant $c$ and $\alpha \in (0,1)$. In this case, the dependence between $X_k$ and $X_{k+n}$ decays slowly as $n$ tends to infinity and $\sum_{n=1}^{\infty} \rho(n) = \infty$.*

We immediately observe that the increments $X_{k+n} \coloneqq B^H_{k+n} - B^H_{k+n-1}$ and $X_k \coloneqq B^H_k - B^H_{k-1}$ of $B^H$ have the long-range dependence property for $H > 1/2$. Applying Taylor's theorem to the second order with a small $x = 1/n$, we obtain the following as $n \to \infty$

$$\rho_H(n) = \frac{1}{2}h^{2H}\Big[(n+1)^{2H} + (n-1)^{2H} - 2n^{2H}\Big] \sim H(2H-1)n^{2H-2}.$$

In particular,

$$\lim_{n\to\infty} \frac{\rho(n)}{H(2H-1)n^{2H-2}} = 1.$$

Summarizing, we have

- $\sum_{n=1}^{\infty} \rho_H(n) = \infty, \quad \text{for } H > 1/2,$
- $\sum_{n=1}^{\infty} |\rho_H(n)| < \infty, \quad \text{for } H < 1/2.$

Thus far, we have observed that the Hurst index governs the dependence structure and long-memory properties of fBm. It also serves as a practical measure of the roughness of fBm paths.

According to the Kolmogorov criterion (for processes on $\mathbb{R}$) [52, Theorem 2.2.3], a process $X = (X_t)_{t\in\mathbb{R}}$ admits a continuous modification if there exist positive constants $\alpha, \beta$, and $K > 0$ such that

$$\mathbb{E}\Big[|X_t - X_s|^{\alpha}\Big] \le K|t-s|^{1+\beta}, \quad \forall s,t \in \mathbb{R}.$$

We recall that a stochastic process $\tilde{X}$ is a modification (or a version) of $X$ if, for each $t$, $\mathbb{P}(\tilde{X}_t = X_t) = 1$ [42, Definition 3.3.5].

**Theorem 2.1.** *[10, Theorem 1.6.1.] Let $H \in (0,1)$, the fBm $B^H$ admits a version whose sample paths are almost surely Hölder continuous of order strictly less than $H$.*

*Proof.* A function $f : \mathbb{R} \to \mathbb{R}$ is Hölder continuous of order $0 < \gamma \le 1$, and write $f \in C^{\gamma}(\mathbb{R})$, if there exists $M > 0$ such that

$$|f(t) - f(s)| \le M|t-s|^{\gamma},$$

for every $s, t \in \mathbb{R}$. Since $B^H_t - B^H_s$ is centered Gaussian with variance $|t-s|^H$ , we have for any $\alpha > 0$

$$\mathbb{E}\Big[|B^H_t - B^H_s|^{\alpha}\Big] = K_{\alpha}|t-s|^{\alpha H}.$$

Therefore, taking any $\alpha > 1/H$, we get the existence of continuous modification. We also get the Hölder continuity of the modification with exponent $\gamma \in (0, H - 1/\alpha)$. Choosing $\alpha$ sufficiently large, we arrive at the desired statement. □

Finally, two well-known features of fBm are that it is neither a Markov process nor a semimartingale. The first feature arises from fBm's long memory property, which inherently precludes it from being a Markov process, as demonstrated in [36]. The second one holds significant implications: it indicates that defining a stochastic integral of a random process with respect to a fBm is not feasible. Therefore, standard Itô calculus cannot be directly applied to fBm. The proof that fBm is not a semimartingale for $H \neq 1/2$ can be found in [58].

We further explore integral representations of fBm and assume the reader is familiar with Wiener processes and Itô stochastic calculus. It is often useful to represent $B^H$ for $0 < H < 1$ as a linear functional of Brownian motion [56]. Specifically, we seek a kernel $K_H(t,s)$ such that the Wiener integral

$$B_t^H = \int K_H(t,s) dB_s, \tag{2.2}$$

is a fBm. Due to the properties of the Wiener integral, the process (2.2) is a centered Gaussian process. Therefore, to verify that this representation defines a fBm, it is sufficient to show that it has the same covariance function.

With this in mind, we introduce and briefly compare the Mandelbrot-Van Ness representation of fBm and Riemann-Liouville processes, as they are frequently encountered in the financial literature.

**Definition 2.3.** *[45, Definition 2.1] Let $H \in (0,1)$, and let $b_0$ be an arbitrary real number. For $t \geq 0$, the Mandelbrot-Van Ness representation of fBm is defined by*

$$B_t^H = b_0 + \frac{1}{\Gamma(H+1/2)} \left\{ \int_{-\infty}^{0} \left[ (t-s)^{H-1/2} - (-s)^{H-1/2} \right] dB'_s + \int_0^t (t-s)^{H-1/2} dB'_s \right\},$$

*where $B' = (B'_t)_{t\in\mathbb{R}}$ is two-sided Brownian motion.*

**Remark 2.1.** We recall that the Gamma function, $\Gamma(\alpha) : (0,\infty) \to (0,\infty)$, is defined by

$$\Gamma(\alpha) = \int_0^\infty e^{-y} y^{\alpha-1} dy. \tag{2.3}$$

and that a two-sided Brownian motion $B' = \{B'(t) : t \in \mathbb{R}\}$ is a continuous random process such that $B'(0) = 0$ almost surely, and $B' = \{B(t) : t \geq 0\}$ and $B' = \{B(-t) : t \geq 0\}$ are independent standard Brownian motions on $[0,\infty)$ [61].

**Remark 2.2.** The Mandelbrot-Van Ness representation of fBm in Definition 2.3 can also be expressed in the form (2.2). To achieve this, we use the fractional kernel

$$K_H(t,s) \coloneqq \frac{1}{C'} \left( (t-s)_+^{H-1/2} - (-s)_+^{H-1/2} \right), \quad s,t \in \mathbb{R}, \quad \text{with}$$
$$C' \coloneqq \left( \int_0^\infty \left( (1+s)^{H-1/2} - s^{H-1/2} \right)^2 ds + \frac{1}{2H} \right)^{1/2},$$

and a two-sided standard Brownian motion as integrator. Here, $x_+^a = x^a \cdot 1_{(0,\infty)}(x)$, where $x, a \in \mathbb{R}$ [38].

**Definition 2.4.** *Let $t \geq 0$ and $H > 0$, the Riemann-Liouville process is defined as*

$$B_t^H = \frac{1}{\Gamma(H+1/2)} \int_0^t (t-s)^{H-1/2} dB_s, \tag{2.4}$$

*where $B = (B_t)_{t\geq 0}$ is standard Brownian motion.*

**Remark 2.3.** The constant $(\Gamma(H-1/2))^{-1}$ in front of the stochastic integral (2.4) is not unique: changing this constant does not alter the properties of the process, as these are determined by the fractional kernel $K(\tau) = \tau^{H-1/2}$.

**Remark 2.4.** In the literature, the Riemann-Liouville process is known by several names:

- Lévy's definition of fBm, named after Lévy who introduced it in [43],
- Volterra Brownian motion, as it belongs to the family of Volterra Gaussian processes,
- *Type 2* of fBm, due to its similarity with *type 1* of fBm defined earlier in 2.1. This convention and further details on the comparison between *type 1* and *2* of fBm are discussed in [46].

By computing the covariance function of the Mandelbrot-Van Ness representation of fBm (Definition 2.3), we can show that it coincides with the covariance function in Definition 2.1 [38, Theorem 3.4]. This confirms that the Mandelbrot-Van Ness representation is indeed fBm, inheriting all its associated properties: stationary increments, self-similarity, and long memory.

In the Mandelbrot-Van Ness representation of fBm, the filtrations generated by $B^H$ and $B'$ do not coincide, classifying it as a *non-canonical* representation. To address this issue, Molchan and Golosov [50] constructed fBm as a Wiener integral process with respect to a one-sided standard Brownian motion. This formulation established a *canonical* representation of fBm and laid the foundation for the family of Volterra Gaussian processes, which includes Riemann-Liouville processes (Definition 2.4). More precisely, the Volterra Gaussian process is a generalization of the Molchan and Golosov representation by allowing arbitrary integrand kernel functions.

**Definition 2.5.** *A kernel $z_X \in L^2_{loc}([0,\infty)^2)$, i.e. $\forall T, S > 0$, $\int_0^T \int_0^S \Big(z_X(t,s)\Big)^2 dsdt < \infty$, is called a Volterra kernel if it satisfies*

$$z_X(t,s) = 0, \quad 0 < t \leq s < \infty.$$

**Definition 2.6.** *[38, Definition 6.1] A Gaussian process $(X_t)_{t\geq 0}$ is called Volterra if there exists a standard Brownian motion $(B_t)_{t\geq 0}$ and a Volterra kernel $z_X \in L^2_{loc}([0,\infty)^2)$, such that for $t \geq 0$*

$$X_t = \int_0^t z_X(t,s) dB_s, \quad a.s.$$

Among the Volterra Gaussian processes, we focus on the specific instance of the Riemann-Liouville process, given by:

$$B_t^H = \sqrt{2H} \int_0^t (t-s)^{H-1/2} dB_s. \tag{2.5}$$

With the choice of the constant $\sqrt{2H}$ in front of the stochastic integral (cf. Remark 2.3), the variance of this process coincides with that of *type 1 fBm* (cf. Remark 2.4):

$$\mathbb{E}\left[\left(\sqrt{2H}\int_0^t (t-s)^{H-1/2}dB_s\right)^2\right] = \mathbb{E}\left[2H\int_0^t (t-s)^{2H-1}ds\right] = t^{2H}.$$

Despite this, the covariance function of the process (2.5) does not align with Definition 2.1, and its increments are non stationary [46, Section 3], indicating that it does not qualify as fBm.

There are several reasons why the Riemann-Liouville process remains of interest. As a Volterra gaussian process, it exhibits self-similarity [38, Section 6.1], and the filtrations generated by $B^H$ and $B$ coincide. It effectively models the long-memory property and the roughness of paths: although it is defined for $H > 0$, for a fixed $H \in (0, 1)$, the Riemann-Liouville process exhibits the same Hölder continuity of *type 1 fBm*, and its paths are sufficiently similar to those of fBm [56, Theorem 17].

### 2.1.3. Infinite-Dimentional OU Processes

The one-dimensional Gaussian Ornstein-Uhlenbeck (OU) process $\tilde{Y} = (\tilde{Y}_t)_{t\geq 0}$ can be defined as the solution to the stochastic differential equation

$$d\tilde{Y}_t = x(m - \tilde{Y}_t)dt + \sigma dB_t, \quad \tilde{Y}_0 > 0, \tag{2.6}$$

where $x, m$ and $\sigma \geq 0$ are real constants, $B$ is a standard Brownian Motion, and $\tilde{Y}_0$ is a given random variable, taken to be independent of $B$ [44]. The parameter $m$ can be formally eliminated from (2.6) by considering $Y := \tilde{Y} - m$ rather than $\tilde{Y}$. In our discussion, we will disregard $m$ and set $\sigma = 1$, focusing on the SDE:

$$dY_t = -xY_t dt + dB_t, \quad Y_0 \geq 0. \tag{2.7}$$

Alternatively, we could define $Y$ in terms of a stochastic integral:

$$Y_t = Y_0 e^{-xt} + \int_0^t e^{-x(t-s)}dB_s, \quad t \geq 0. \tag{2.8}$$

It is easily verified that $Y$ as defined by (2.8) satisfies (2.7) for any $x$ and choice of $Y_0$, and that $Y$ is the unique strong Markov solution to (2.7) (cf. Theorems 1.1 and 1.2).

Noting that the stochastic integral in (2.8) is a well-defined Wiener integral, basic properties of $Y$ are derived. In particular, and assuming that $Y_0$ has finite variance, $Y$ is Gaussian with expectation function and covariance functions given by

$$\mathbb{E}[Y_t] = e^{-xt}\mathbb{E}[Y_0], \quad t \geq 0,$$

and

$$\mathrm{Cov}(Y_u, Y_t) = \frac{1}{2x}(e^{xt} - e^{-xt}) + e^{-x(u+t)}\mathrm{Var}(Y_0), \quad u \geq t \geq 0.$$

We now introduce infinite-dimensional OU processes and present key results in Proposition 2 and Theorem 2.2 concerning their Markovian structure and integrability, which are essential for the Markovian representation of the fractional processes that we will discuss next.

**Definition 2.7.** *[33, Definition 2.1] Given a collection of $\mathcal{F}_0$-measurable $\mathbb{R}$-valued random variables $Y_0^x$ indexed by $x \in (0,\infty)$, let for each $t \geq 0$*

$$Y_t^x = Y_0^x e^{-xt} + \int_0^t e^{-x(t-s)} dB_s, \tag{2.9}$$

*and let $(Y_t^x)_{x>0,t\geq 0}$ denote the collection of OU processes indexed by the speed of mean reversion $x$.*

**Remark 2.5.** Observe that the random field $Y(t,x,\omega) : [0,\infty) \times (0,\infty) \times \Omega \to \mathbb{R}$ is a measurable map

$$Y : \Omega \to C([0,T], C^\infty((0,\infty),\mathbb{R})),$$

and for each $x \in (0,\infty)$, $(Y_t^x)_{t\geq 0}$ is a one-dimensional OU process that solves (2.7).

**Proposition 2.** *[15, Proposition 1] The collection of OU processes $(Y_t^x)_{x>0,t\geq 0}$ is an infinite dimensional Markov process.*

*Proof.* Let $\tau$ be an almost surely finite stopping time with respect to $\mathcal{F}_t = \sigma\{B(s); s \leq t\}$. Then, $\tilde{B} = (\tilde{B}_t = B_{t+\tau} - B_\tau)_{t\geq 0}$ is a standard Brownian motion independent of $(B_u, u \leq \tau)$ and, for each $x > 0$, $(Y_{t+\tau}^x)_{t\geq 0}$ is the solution of the SDE

$$dY_{t+\tau}^x = -xY_{t+\tau}^x dt + d\tilde{B}_t, \quad Y_\tau \geq 0. \tag{2.10}$$

Therefore,

$$Y_{t+\tau}^x = e^{-xt} Y_{t+\tau}^x + \int_0^t e^{-x(t-s)} d\tilde{B}_t.$$

Since for each $x > 0$, the process $(Y_{t+\tau}^x)_{t\geq 0}$ is the solution of the time-homogeneous diffusion (2.10), it satisfies the strong Markov property (cf. Theorem 1.3). Thus, for any bounded measurable function $f$,

$$\mathbb{E}[f((Y_{t+\tau})_{t\geq 0})|\mathcal{F}_\tau] = \mathbb{E}^{Y_\tau}[f(Y_t)_{t\geq 0})]. \tag{2.11}$$

Here, $(Y_t)_{t\geq 0}$ refers to the collection of OU processes $(Y_t^x)_{x>0,t\geq 0}$, i.e. $x$ is not fixed. The right hand side of (2.11) means the function $\mathbb{E}^y[f(Y_t)_{t\geq 0}]$ evaluated at $y = Y_\tau(\omega)$ where

$$\mathbb{E}^y[f(Y_t)_{t\geq 0}] = \mathbb{E}[f((Y_t^{x,y(x)})_{x>0,t\geq 0})],$$

and for fixed $x$, $(Y_t^{x,y})_{t\geq 0}$ is the OU process driven by $B$, of parameter $x$, starting from $y$. □

We now move on to the integrability conditions of the infinite-dimentional OU process.

**Theorem 2.2.** *[33, Theorem 2.4] Let $\mu$ be a $\sigma$-finite measure on $(0,\infty)$ such that for each $t > 0$*

$$\int_0^\infty (1 \wedge x^{-1/2}) \mu(ds) < \infty. \tag{2.12}$$

*and let $Y_0 \in L^1((0,\infty),\mu)$, i.e. $\int_0^\infty Y_0 \mu(dx) < \infty$. Then the process $(Y_t)_{t\geq 0}$ has a predictable $L^1((0,\infty),\mu)$-valued version and is Gaussian.*

### 2.1.4. Markovian Representations

The goal is to obtain a Markovian representation of fBm in terms of a infinite-dimentional OU process (Definition 2.7) for $H \in (0, 1/2)$. We omit from our discussion the case $H \in (1/2, 1)$, as in the next chapters we will consider fractional processes whose paths are "rougher" than standard Brownian motion. We will first focus on the Mandelbrot-Van Ness representation (Definition 2.3) and then on Riemann-Liouville processes (Definition 2.4).

**Assumption 2.1.** Let $(Y_t^x)_{x>0,t\geq 0}$ denote the collection of OU processes, as given in Definition 2.7, with initial value

$$Y_0^x = \int_{-\infty}^0 e^{sx} dB_s,$$

and let $\mu$ be the $\sigma$-finite measure on $(0, \infty)$ for $H < 1/2$ defined as follows:

$$\mu(dx) \coloneqq c_H x^{-H-1/2} dx, \quad c_H \coloneqq \frac{1}{\Gamma(H + 1/2)\Gamma(1/2 - H)}. \tag{2.13}$$

**Remark 2.6.** The constant $c_H$ in the definition of $\mu$ (2.13) is not unique: if $H < 1/2$, then $\mu$ may be multiplied by any positive constant without affecting the validity of the statements below [33, Remark 3.3].

**Theorem 2.3.** *[33, Theorem 3.5] Under Assumption 2.1, the Mandelbrot-Van Ness representation of fBm (Definition 2.3) admits the Markovian representation*

$$B_t^H = b_0 + \int_0^\infty (Y_t^x - Y_0^x)\mu(dx), \quad \textit{for } H < 1/2, \tag{2.14}$$

*where $(Y_t^x - Y_0^x)$ is a continuous process in $L^1((0, \infty), \mu)$.*

*Proof.* The function $\tau \to \tau^{H-\frac{1}{2}}/\Gamma(H + 1/2)$ appearing in the definition of $B^H$ is the Laplace transform (cf. Theorem A.2) of the measure $\mu$, i.e. for each $\tau > 0$,

$$\mathcal{L}(\mu)(\tau) = \int_0^\infty e^{-\tau x}\mu(dx) = \frac{\tau^{H-1/2}}{\Gamma(H + 1/2)}.$$

Equivalently, we can use the definition of gamma function (2.3), and by setting $y = \tau x$, $\alpha = \frac{1}{2} - H$, we obtain

$$\Gamma\Big(\frac{1}{2} - H\Big) = \int_0^\infty (\tau x)^{-H-1/2} e^{-\tau x} \tau \, dx,$$

which implies

$$\frac{\tau^{H-1/2}}{\Gamma(H + 1/2)} = \int_0^\infty e^{-\tau x} x^{-H-1/2} \frac{1}{\Gamma(H + 1/2)\Gamma(1/2 - H)} dx = \int_0^\infty e^{-\tau x}\mu(dx).$$

We recall that the Mandelbrot-Van Ness representation of fBm is given by

$$B_t^H = b_0 + \frac{1}{\Gamma(H + 1/2)} \left\{ \int_{-\infty}^0 \Big[(t - s)^{H-1/2} - (-s)^{H-1/2}\Big] dB_s + \int_0^t (t - s)^{H-1/2} dB_s \right\}.$$

Therefore,

$$B_t^H = b_0 + \int_{-\infty}^0 \int_0^\infty (e^{-x(t-s)} - e^{-x(-s)})\mu(dx)dB_s + \int_0^t \int_0^\infty e^{-x(t-s)}\mu(dx)dB_s.$$

By the Stochastic Fubini Theorem A.1,

$$B_t^H = b_0 + \int_0^\infty \int_{-\infty}^0 (e^{-x(t-s)} - e^{-x(-s)})dB_s\mu(dx) + \int_0^\infty \int_0^t e^{-x(t-s)}dB_s\mu(dx). \quad (2.15)$$

Condition (A.2) of Fubini's theorem guarantees that for almost all $\omega \in \Omega$ and for all $t \in [0,T]$

$$\int_{-\infty}^0 (e^{-x(t-s)} - e^{-x(-s)})dB_s \text{ and } \int_0^t e^{-x(t-s)}dB_s \quad \in L^1((0,\infty),\mu)$$

and that we can interchange the order of integration in the double integral.

In our case, condition (A.2) is satisfied because

$$\int_0^\infty \sqrt{\int_{-\infty}^0 (e^{-x(t-s)} - e^{-x(-s)})^2 ds}\mu(dx) = \int_0^\infty \frac{1-e^{-tx}}{\sqrt{2}x}\mu(dx) \le \int_0^\infty \sqrt{\frac{1-e^{-2xt}}{x}}\mu(dx)$$

$$\int_0^\infty \sqrt{\int_0^t e^{-2x(t-s)}ds}\mu(dx) = \int_0^\infty \sqrt{\frac{1-e^{-2xt}}{2x}}\mu(dx) \le \int_0^\infty \sqrt{\frac{1-e^{-2xt}}{x}}\mu(dx)$$

where we used $1 - e^{-tx} \le \sqrt{1-e^{-tx}}$. Moreover,

$$\int_0^\infty \sqrt{\frac{1-e^{-2xt}}{x}}\mu(dx) \le \left(1 \vee (2t)^{\frac{1}{2}}\right) \int_0^\infty \left(1 \wedge x^{-\frac{1}{2}}\right)\mu(dx) < \infty,$$

where we used the inequality

$$k_1(x,\tau) := \frac{1-e^{-\tau x}}{x} \le (1 \vee \tau)(1 \wedge x^{-1}). \quad (2.16)$$

The function $k_1(\cdot,\tau)$ is decreasing in $x$ for all $\tau > 0$, as its derivative w.r.t. $x$ is strictly negative for each $x \in (0,\infty)$: $\partial_x k_1(x,\tau) = (e^{-\tau x}x\tau - 1 + e^{-\tau x})/x^2 < 0$. The inequality (2.16) follows from

$$\lim_{x\to\infty} k_1(x,\tau) = 0, \quad \lim_{x\to 0^+} k_1(x,\tau) = \tau.$$

Inserting initial value $Y_0^x$ and the expression for $Y_t^x$ given in (2.9) into (2.15) yields the Markovian representation (2.14):

$$\begin{aligned} B_t^H &= b_0 + \int_0^\infty (e^{-xt} - 1)Y_0^x\mu(dx) + \int_0^\infty \int_0^t e^{-x(t-s)}dB_s\mu(dx) \\ &= b_0 + \int_0^\infty (Y_t^x - Y_0^x)\mu(dx). \end{aligned}$$

Lastly, the expressions

$$(e^{-xt} - 1)Y_0^x \text{ and } \int_0^t e^{-x(t-s)}dB_s$$

define continuous $L^1((0,\infty),\mu)$-valued processes: the first expression has majorant $(1 \vee t)(1 \wedge x)Y_0^x$ in $L^1((0,\infty),\mu)$, which allows one to apply the dominated convergence theorem, and the second expression is treated in [33, Theorem 2.11]. □

We now turn to the Riemann-Liouville process, as defined in Definition 2.4.

**Theorem 2.4.** *Let $Y = (Y_t^x)_{x>0,t\geq 0}$ denote the collection of OU processes, as given in Definition 2.7, with a constant initial value $Y_0^x = 0$, and $\mu$ as given in* (2.13)*. The Riemann-Liouville process (Definition 2.4) admits the Markovian representation*

$$B_t^H = \int_0^\infty Y_t^x \mu(dx), \quad \text{for } H < 1/2. \tag{2.17}$$

*Proof.* In analogy to the previous theorem, the function $\tau \to \tau^{H-\frac{1}{2}}/\Gamma(H+1/2)$ appearing in the definition of $B^H$ is the Laplace transform of the measure $\mu$. Therefore,

$$\begin{aligned} B_t^H &= \frac{1}{\Gamma(H+1/2)} \int_0^t (t-s)^{H-1/2} dB_s \\ &= \int_0^t \int_0^\infty e^{-x(t-s)} \mu(dx) dB_s = \int_0^\infty \int_0^t e^{-x(t-s)} dB_s \mu(dx), \end{aligned}$$

where in the last step we used the Stochastic Fubini Theorem A.1; for justification, refer to the preceding theorem. Lastly, from Equation (2.9), where $Y_t^x = \int_0^t e^{-x(t-s)} dB_s$ with $Y_0^x = 0$, we obtain the Markovian representation (2.17). □

The following lemma guarantees that there exists a version $Y$ which is continuous in the temporal variable and smooth in the spatial variable.

**Lemma 1.** *[32, Lemma 1] Let $Y = (Y_t^x)_{x>0,t\geq 0}$ denote the collection of OU processes, as given in Definition 2.7, with a constant initial value $Y_0^x = 0$, and $\mu$ as given in* (2.13)*. Then, $Y$ is a measurable mapping*

$$Y : \Omega \to C([0,T], C^\infty((0,\infty), \mathbb{R}) \cap L^1((0,\infty), \mu)).$$

**Remark 2.7.** Lemma 1 is a slightly variation of [32, Lemma 1]: we are considering $Y$ as given in Definition 2.7, with a constant initial value $Y_0^x = 0$, and $\mu$ as given in (2.13) instead of

$$Y_t^x = \frac{1}{\Gamma(1/2-H)} \int_0^t e^{-x(t-s)} dB_s \quad \forall t \in [0,\infty), x \in (0,\infty)$$

and $\mu(x) = x^{-H-1/2}dx$. In light of Remarks 2.3 and 2.6, altering the constants in front of $Y$ and $\mu$ do not affect the stated result.

**Remark 2.8.** When we say that the Riemann-Liouville process $B^H$ is a linear functional of $Y$, we mean that $\forall t \in [0,T]$,

$$B^H = \int_0^t (t-s)^{H-1/2} dB_s = \int_0^\infty Y_t^x \mu(dx), \quad \text{a.s.}$$

with $\mu$ as given in (2.13).

### 2.1.5. Markovian Representation of SVE

We now aim to elucidate the relationship between the Markovian representation of fBm and stochastic Volterra equations (SVE), with a focus on the class of affine Volterra processes.

Consider a one-dimensional stochastic Volterra equation of convolution type

$$V_t = \xi_0(t) + \int_0^t K(t-s)b(V_s)ds + \int_0^t K(t-s)\sigma(V_s)dB_s, \tag{2.18}$$

where $B$ is a $d$-dimensional Brownian motion, and the convolution kernel $K$, the function $\xi_0$ and coefficients $b$ and $\sigma$ satisfy suitable regularity conditions.

Setting $K \equiv 1$ and $\xi_0 \equiv V_0$ for some constant initial condition $V_0$, one recovers the stochastic integral equation (1.2) for $n = 1$ with time-independent coefficients. Consequently, stochastic Volterra equations extend standard stochastic differential equations allowing for more flexibility in modeling. However, they do not fall in general in the semimartingale and Markovian frameworks. We can easily see this by considering the Riemann-Liouville process (Definition 2.4) $V_t = \int_0^t K_H(t-s)dB_s$ where the kernel $K_H$ is defined by $K_H : \tau \to \tau^{H-1/2},\ H \in (0, 1/2)$.

The existence of *strong* and *weak* solutions of the form (2.18) can be established analogously to stochastic differential equations. By an abuse of terminology, we say that the stochastic Volterra equation (2.18) admits a weak solution if there exists a stochastic basis $(\Omega, \mathcal{F}, \mathbb{F} = (\mathcal{F}_t)_{t\in\mathbb{T}}, \mathbb{P})$ that satisfies the usual conditions, supports a Brownian motion $B$, and includes an adapted continuous process $V$ such that (2.18) holds. We call $V$ a strong solution if it is also adapted to the filtration generated by $B$.

Remarkably, existence results have been proved under mild assumptions on the kernel and coefficients in [2]. We place particular emphasis on affine Volterra processes, as they effectively characterize the affine and Markovian structures of rough volatility models, such as the rough Heston model, which will be discussed in the next chapter.

$V$ is an affine Volterra process when $b(x)$ and $\sigma^2(x)$ are affine in $x$. This definition naturally generalizes to higher dimension: if $b(x)$ is a vector and $\sigma(x)$ a matrix, and one then requires $b(x)$ and $\sigma(x)\sigma^\mathsf{T}(x) > 0$ to be affine in $x$. In the one-dimensional affine case:

$$b(x) = \beta - \lambda x \quad \text{and} \quad \sigma^2(x) = \alpha + ax \tag{2.19}$$

for some real parameters $\beta, \lambda, \alpha$ and $a$ such that $\alpha + aV_t \geq 0$ for all $t \geq 0$.

**Theorem 2.5.** *Consider the equation* (2.18) *with coefficients* $b(x)$ *and* $\sigma(x)$ *as in* (2.19) *and kernel* $K \in L^2_{loc}(\mathbb{R}_+)$*, i.e.* $\exists M > 0$ *such that* $\int_0^M (K(\tau))^2 d\tau < \infty$.

*(i) Assume that* $\alpha \geq 0$ *and* $a = 0$*. Then there exists a pathwise unique strong solution* $V$ *for any initial condition* $\xi_0 \in \mathbb{R}$*; the Volterra OU process.*

*(ii) Assume that* $\beta \geq 0, \alpha \geq 0, a > 0$*, and that* $K$ *is strictly positive and completely monotone (Definition A.1). Assume also that there is* $\gamma \in (0, 2]$ *such*

*that* $\int_0^h (K(\tau))^2\,d\tau = O(h^\gamma)$ *and* $\int_0^T (K(\tau+h)-K(\tau))^2\,dt = O(h^\gamma)$, *for every* $T<\infty$. *Then, there exists a unique in law* $\mathbb{R}_+$*-valued weak solution* $V$ *for any initial condition* $\xi_0 \in \mathbb{R}_+$*; the Volterra square-root process.*

*In either case, the trajectories of* $V$ *are Hölder continuous of any order less than* $\gamma/2$.

**Remark 2.9.** In light of Theorem 2.5, the Rienman-Liouville process (Definition 2.4) is then a *strong* solution of (2.18) with $b=0$ and $\sigma^2=\alpha$, where $\alpha$ is a constant.

To circumvent the fact that the process $V$ in (2.18) is generally neither a Markov process nor a semimartingale, one seeks a Markovian representation for $V$. The perspective we have adopted for fBm, viewing it as a mixture of (possibly infinite) mean-reverting processes, applies also in this more general context.

Assume that the kernel $K$ is the Laplace transform of some measure $\mu$:

$$K(\tau) = \int_0^\infty e^{-x\tau}\mu(dx), \quad \tau \geq 0. \tag{2.20}$$

If $\mu$ is a positive measure, then $K$ is completely monotone on $(0,\infty)$. Conversely, if $K$ admits the representation (2.20), it is completely monotone as a result of the Bernstein-Widder theorem (Theorem A.2). To understand how (2.20) can lead to (possibly infinite) mixture of mean-reverting processes, we assume $\xi_0 = 0$ and substitue (2.20) into (2.18) with $\xi_0 = 0$. A formal interchange of the time integrals and $\mu$ integrals yields the representation

$$\begin{aligned} V_t &= \int_0^t K(t-s)\Big(b(V_s)\,ds + \sigma(V_s)\,dB_s\Big) \\ &= \int_0^\infty \int_0^t e^{-x(t-s)}\Big(b(V_s)\,ds + \sigma(V_s)\,dB_s\Big)\,\mu(dx) \\ &= \int_0^\infty U_t(x)\,\mu(dx) \end{aligned} \tag{2.21}$$

where we have defined, for all $t \geq 0$,

$$U_t(x) = \int_0^t e^{-x(t-s)}b(V_s)ds + \int_0^t e^{-x(t-s)}\sigma(V_s)dB_s.$$

Crucially, each process $(U_t(x))_{t\geq 0}$ is a semimartingale, even if $V$ is not. To determine its dynamics, we move $e^{-xt}$ outside the integrals and apply the Itô formula to obtain

$$dU_t(x) = (-xU_t(x) + b(V_t))dt + \sigma(V_t)dB_t, \quad U_0(x) = 0.$$

Plugging (2.19) and (2.21) into this expression gives

$$\begin{aligned} dU_t(x) &= \Big(-xU_t(x) + \beta - \lambda V_t\Big)dt + \sqrt{\alpha + aV_t}dB_t \\ &= \Big(-xU_t(x) + \beta - \lambda\int_0^\infty U_t(x)\,\mu(dx)\Big)dt + \sqrt{\alpha + a\int_0^\infty U_t(x)\,\mu(dx)}dB_t. \end{aligned} \tag{2.22}$$

As $x$ ranges through the support of $\mu$, (2.22) defines a (possibly infinite) coupled system of mean-reverting processes, and (2.21) expresses $V$ as a mixture of these processes. Note

that for $a = 0$, $U_t(x)$ is Gaussian. If additionally $b = 0$, i.e. $\beta = \lambda = 0$, and $\sigma = \alpha = 1$, we retrieve the infinite-dimensional OU process (Definition 2.7); $U_t(x) \equiv Y_t^x$.

We emphasize that the representation (2.21) provides a natural Markovian approximation of $V$, which is suitable for numerical purposes. When the measure $\mu$ is replaced by an approximation $\hat{\mu}$ that is supported by finitely many points $x_1, \dots, x_N$, the dynamic (2.22) becomes a system of $N$ SDEs, corresponding to the dynamics of the $N$-dimensional Markov process $(U_t(x_1), \dots, U_t(x_N))_{t\geq 0}$. This system can then be used to approximate the affine Volterra process $V$ [41].

In the next section, we will establish stability results for the Markovian approximation of $d$-dimensional stochastic Volterra equations. Our analysis extends beyond the Gaussian and OU cases, as considering the more general scenario does not significantly complicate the discussion.

## 2.2. Approximation of Completely Monotone Kernels

### 2.2.1. Introduction

Consider a $d$-dimensional stochastic Volterra equation of the form

$$V_t = v_0 + \int_0^t K(t-s)b(V_s)ds + \int_0^t K(t-s)\sigma(V_s)dB_s, \tag{2.23}$$

where $v_0 \in \mathbb{R}^d$, and $b : \mathbb{R}^d \to \mathbb{R}^d, \sigma : \mathbb{R}^d \to \mathbb{R}^{d\times d}$ are Lipschitz continuous, i.e. $\exists L > 0$ s.t.

$$|b(x) - b(y)| + \|\sigma(x) - \sigma(y)\| \leq L|x - y|, \quad \forall x, y \in \mathbb{R}^d.$$

We recall that $|\cdot|$ denotes the Euclidean norm of $\mathbb{R}^d$, while $\|\cdot\|$ denotes the Hilbert–Schmidt on $\mathbb{R}^{d\times d}$ : $\|\sigma\|^2 = \sum_{i=1}^d \sum_{j=1}^d \sigma_{i,j}^2$. Moreover, $B$ is a $d$-dimensional Brownian motion, and the kernel $K : \mathbb{R}_+ \to \mathbb{R}^{d\times d}$ satisfies the conditions:

$$\begin{aligned} &K \in L^2_{\mathrm{loc}}(\mathbb{R}_+) \text{ and there is } \gamma \in (0, 2] \text{ such that } \textstyle\int_0^h (K(\tau))^2\, dt = O(h^\gamma) \text{ and} \\ &\textstyle\int_0^T (K(\tau + h) - K(\tau))^2\, d\tau = O(h^\gamma), \text{ for every } T < \infty. \end{aligned} \tag{2.24}$$

Then, we can apply the following theorem, which guarantees the existence and uniqueness of a strong solution to (2.23).

**Theorem 2.6.** *[2, Theorem 6.10] Under the kernel condition* (2.24)*, if $b$ and $\sigma$ are Lipschitz continuous,* (2.23) *admits a unique continuous strong solution $V$.*

We are interested in approximating the solution of (2.23) when there exists a measure $\mu$ on $(0, \infty)$ satisfying

$$K(\tau) = \int_0^\infty e^{-x\tau}\mu(dx) < \infty, \quad \forall \tau > 0. \tag{2.25}$$

In the one-dimensional case, such a kernel $K$ is completely monotone by the Bernstein-Widder Theorem A.2.

The principle of the approximation involves replacing the measure $\mu$ with a finite discrete measure $\hat{\mu} \coloneqq \sum_{i=1}^{N} w_i \delta_{x_i}$ where $\delta_x$ is the Dirac measure on $x$, and $(w_i)_{i=1}^N, (x_i)_{i=1}^N$ are positive weights and nodes, respectively. Consequently, the kernel in (2.25) is approximated by the finite sum

$$\hat{K}(\tau) \coloneqq \sum_{i=1}^{N} w_i e^{-x_i \tau}, \tag{2.26}$$

which leads to the approximation

$$\hat{V}_t = v_0 + \int_0^t \hat{K}(t-s) b(\hat{V}_s) ds + \int_0^t \hat{K}(t-s) \sigma(\hat{V}_s) dB_s, \tag{2.27}$$

of the stochastic Volterra equation in (2.23).

The following proposition shows that the approximation $\hat{V}$ in (2.27) reduces to an $N$-dimensional ordinary stochastic differential equation.

**Proposition 3.** *[4, Proposition 2.1] Assume that $\hat{\mu} = \sum_{i=1}^{N} w_i \delta_{x_i}$ with $w_i \geq 0$ and $x_i \geq 0$ where $x_1 < \ldots < x_N$. Let $v_0^1, \ldots, v_0^N \in \mathbb{R}^d$ be such that $\sum_{i=1}^{N} w_i v_0^i = v_0$. Then the solution to (2.27) is given by $\sum_{i=1}^{N} w_i V_t^i$, where $(V_t^1, \ldots, V_t^N)$ is the solution to the $(N \times d)$-dimensional SDE defined by*

$$V_t^i = v_0^i - \int_0^t x_i (V_t^i - v_0^i) ds + \int_0^t b\Big(\sum_{j=1}^{N} w_j V_s^j\Big) ds + \int_0^t \sigma\Big(\sum_{j=1}^{N} w_j V_s^j\Big) dB_s. \tag{2.28}$$

*Proof.* By assumption, the SDE (2.28) has Lipschitz coefficients and therefore has a unique strong solution (cf. Theorem 1.1 and Remark 1.1). Since

$$d\Big(e^{x_i t}(V_t^i - v_0^i)\Big) = e^{x_i t} b\Big(\sum_{j=1}^{N} w_j V_s^j\Big) dt + e^{x_i t} \sigma\Big(\sum_{j=1}^{N} w_j V_s^j\Big) dB_t,$$

we get

$$V_t^i = v_0^i + \int_0^t e^{x_i (t-s)} b\Big(\sum_{j=1}^{N} w_j V_s^j\Big) ds + \int_0^t e^{x_i (t-s)} \sigma\Big(\sum_{j=1}^{N} w_j V_s^j\Big) dB_s.$$

We left multiply this equation by $w_i$ and then sum over $i$ to obtain that $\sum_i^N w_i V_t^i$ solves (2.27). □

**Remark 2.10.** Proposition 3 presents a slight modification of the statement in [4]. In the original paper, the theorem allows the kernel $K$ to admit a signed measure. For instance, $K$ could take the form $K(\tau) = e^{-\tau} - 2e^{-3\tau}$, though not being completely monotone enters in this framework. To establish this broader result, it is necessary to assume the existence of a bounded measurable function $M : \mathbb{R}_+ \to \mathbb{R}^{d \times d}$ such that

$$K(\tau) = \int_0^\infty e^{-x\tau} M(x) \mu(dx), \quad \text{for } \tau \in [0, \infty).$$

### 2.2.2. Strong Error Analysis

To analyse the error between the stochastic Volterra equation (SVE) (2.23) and its approximation (2.27), we proceed in two steps. First, we analyze the truncation error incurred when replacing the kernel (2.25) with the truncated kernel $K^{tr}$, which is obtained by truncating the measure $\mu$ from $(0,\infty)$ to $[0,M]$, with $0 < M < \infty$. Second, we evaluate the discretization error between the SVE with the truncated kernel $K^{tr}$ and the approximating kernel $\hat{K}$ in (2.26), where $\mu$ has been discretized.

For any $M > 0$, the truncated kernels $K^{tr} : \mathbb{R}_+ \to \mathbb{R}^{d\times d}$ are defined as follows:

$$K^{tr}(\tau) = \int_0^M e^{-x\tau}\mu(dx), \quad \text{for all } \tau \geq 0.$$

Thus, the kernel $K^{tr}$ approximates the kernel $K$ defined by (2.25) as $M \to \infty$. We assume $K^{tr}$ satisfies conditions (2.24), ensuring the validity of Thorem 2.6 for the solution $(V_t^{tr})_{t\in[0,T]}$ to the stochastic convolution equation associated given by:

$$V_t^{tr} = v_0 + \int_0^t K^{tr}(t-s)b(V_s^{tr})ds + \int_0^t K^{tr}(t-s)\sigma(V_s^{tr})dB_s. \tag{2.29}$$

**Proposition 4.** *[4, Proposition 3.1] Let $\mu$ be a positive measure such that*

$$r(M) \coloneqq \int_M^\infty \int_M^\infty \frac{1}{x_1+x_2}\mu(dx_1)\mu(dx_2) < \infty, \quad \forall M > 0. \tag{2.30}$$

*Then, for any $T > 0$, there exists a constant $C > 0$ that depends on $T, \mu, L, |b(0)|$ and $\|\sigma(0)\|$ such that*

$$\mathbb{E}\Big[|V_t - V_t^{tr}|^2\Big] \leq Cr(M), \quad \forall t \in [0,T].$$

We now turn to the approximation of the truncated kernel $K^{tr}$ by $\hat{K}$. The following result demonstrates that the $L^2$-error between $V^{tr}$ and the Markovian approximation $\hat{V}$ on a finite interval $[0,T]$ is controlled by the $L^2$-error between $K^{tr}$ and $\hat{K}$.

**Proposition 5.** *[4, Proposition 3.2] Let $T > 0$. Suppose that for any $M > 0$, there is a kernel $K^{tr} : [0,T] \to \mathbb{R}^{d\times d}$ satisfying (2.24). Then, there exists a constant $C > 0$, depending on $T, L, |v_0|, |b(0)|, \|\sigma(0)\|$, such that*

$$\mathbb{E}\Big[|\hat{V}_t - V_t^{tr}|^2\Big] \leq C\int_0^t \Big\|\hat{K}(s) - K^{tr}(s)\Big\|^2 ds, \quad \forall t \in [0,T].$$

Combining Propositions 4 and 5, we obtain the following comprehensive result.

**Theorem 2.7.** *[4, Theorem 3.2] Let us assume that $\mu$ satisfies (2.30) and that $K^{tr} : [0,T] \to \mathbb{R}^{d\times d}$ satisfies (2.24). Then, there exists a constant $C > 0$ such that*

$$\mathbb{E}\Big[|\hat{V}_t - V_t|^2\Big] \leq C\Big(r(M) + \int_0^t \Big\|\hat{K}(s) - K^{tr}(s)\Big\|^2 ds\Big), \quad \forall t \in [0,T].$$

The term $r(M)$ and the integral in the right hand side correspond respectively to the truncation and discretization error.

The choice of positive weights $(w_i)_{i=1}^N$ and nodes $(x_i)_{i=1}^N$ in (2.26) is essential for ensuring the accuracy of the approximation. Specifically, they must be chosen such that

$$\left\|\hat{K} - K\right\|_{2,T} \to 0, \quad \text{as } N \to \infty,$$

where $\|\cdot\|_{2,T}$ is the $L^2([0,T])$ norm. As suggested by [32] and [6], we adopt an $m$-point Gaussian quadrature rule to determine weights and nodes for approximating $K$ in (2.25) with $\hat{K}$ in (2.26).

### 2.2.3. Strong Error Analysis under Gaussian Quadrature

We now determine the weights $(w_i)_{i=1}^N$ and nodes $(x_i)_{i=1}^N$ necessary for the approximating kernel $\hat{K}$ in (2.26) by means of Gaussian quadrature. For the delicate task of selecting optimal point sets and weights, we focus on the specific class of fractional kernels

$$K(\tau) \coloneqq \frac{\tau^{H-1/2}}{\Gamma(H+1/2)}, \tag{2.31}$$

with parameter $H \in (0, 1/2)$ and satisfying

$$K(\tau) = \int_0^\infty e^{-x\tau}\mu(dx) = c_H \int_0^\infty e^{-x\tau} x^{-H-1/2} dx. \tag{2.32}$$

We have set $\mu(dx) \coloneqq \nu(x)dx$, where for $x > 0$ the weight function $\nu$ is defined as

$$\nu(x) \coloneqq c_H x^{-H-1/2}, \quad c_H \coloneqq \frac{1}{\Gamma(H+1/2)\Gamma(1/2-H)}. \tag{2.33}$$

**Remark 2.11.** The fractional kernel $K$ (2.31) satisfies conditions (2.24) with $\gamma = 2H$. Indeed, it is locally square integrable, and we have $\int_0^h (K(\tau))^2 d\tau = h^{2H}/(2H)$ as well as

$$\int_0^T (K(\tau+h) - K(\tau))^2 dt \le h^{2H} \int_0^\infty \left((\tau+1)^{H-1/2} - (\tau)^{H-1/2}\right)^2 d\tau$$

where the constant on the right-hand side is bounded by $\frac{1}{2H} + \frac{1}{2-2H}$ [2, Example 5.3].

**Remark 2.12.** It should be noted that analogous results, regarding the selection of optimal point sets and weights by means of Gaussian quadrature, can be achieved for completely monotone kernels $K$, not necessarily of the form (2.31) (cf. [6, Section 2.2]).

The measure $\mu$ appearing in the integral representation of $K$ (2.32) is approximated by a weighted sum of Dirac measures ($\hat{\mu} \coloneqq \sum_{i=1}^N w_i \delta_{x_i}$). Specifically, for each $n \in \mathbb{N}$, we truncate the positive half-line on which $\mu$ is defined to a finite interval $[\xi_0, \xi_n]$. This interval is then divided into subintervals by a sequence of auxiliary nodes $(\xi_i)_{i=0}^n$. On each subinterval $[\xi_i, \xi_{i+1}]$ for $i = 0, \dots, n-1$, the measure $\mu$ is approximated by an $m$-point Gaussian quadrature rule. Thus, $N = nm$ is the total number of nodes.

Let $\mathcal{Z}^n \coloneqq \{0 < \xi_0 < \ldots < \xi_n < \infty\}$ denote the grid of auxiliary nodes, which we assume satisfies the following conditions

$$\begin{aligned} (\mathbf{Hz}) \quad & (i) \quad \xi_0 \to 0 \text{ and } \xi_n \to \infty \text{ for } N \to \infty, \\ & (ii) \quad \Delta \coloneqq \max_{i=0,\ldots,n-1} |\xi_{i+1} - \xi_i| \to 0 \text{ for } n \to \infty, \\ & (iii) \quad \mathcal{Z}^n \subset \mathcal{Z}^{n+1}. \end{aligned}$$

We adopt a *geometric grid* as suggested in [16], [32] and [6]. This choice satisfies (**Hz**) and simplifies the error analysis compared to more complex subdivisions that aim to distribute the error more uniformly. However, a better rate of convergence might be achievable with a more strategic selection of intervals.

By *geometric grid*, we mean that the sequence $(\xi_i)_{i=0}^n$ follows a geometric progression. Its definition is provided below.

**Definition 2.8** (Geometric grid)**.** *An interval of length $L$ exhibits a geometric partition if is divided into $n$ subintervals $I_k$ such that the ratio $I_{k+1}/I_k = q$ for $k = 0, \ldots, n-1$, where $q$ is known as the common ration. The length of each subinterval can be expressed as $cq^k$ for $k = 0, \ldots, n-1$, where $c > 0$ is a constant, and the total length of the interval is given by:*

$$L = \sum_{k=0}^{n-1} cq^k = \begin{cases} cn, & \text{for } q = 1, \\ c\dfrac{1-q^n}{1-q}, & \text{otherwise.} \end{cases}$$

**Notation 2.1.** For the remainder of this section, we will denote $a \approx b$ for $a \in \mathbb{N}$ and $b \in \mathbb{R}$ if $|a - b| < 1$, meaning that $a$ can be obtained from $b$ by rounding $b$ to the nearest integer.

We now describe the nodes and weights used in our approximation of the kernel $K$ in (2.32), which is the approximation adopted in [6, Theorem 2.1]. Let $N \in \mathbb{N}$ be the total number of nodes and $\alpha, \beta, a, b \in (0, \infty)$ be parameters of the scheme. Let also $A = A_H \coloneqq (\frac{1}{H} + \frac{1}{3/2-H})^{1/2}$, where $0 < H < 1/2$ denotes the Hurst index. With these parameters, we define

$$m \approx \frac{\beta}{A}\sqrt{N}, \quad n \approx \frac{N}{m},$$

and the sequence of auxiliary nodes

$$\begin{cases} \xi_0 \coloneqq a \exp\Big(-\dfrac{\alpha}{(3/2-H)A}\sqrt{N}\Big), \quad \xi_n \coloneqq b \exp\Big(\dfrac{\alpha}{HA}\sqrt{N}\Big), \\ \xi_i \coloneqq \xi_0 \left(\dfrac{\xi_n}{\xi_0}\right)^{i/n}, \quad i = 0, \ldots, n. \end{cases} \tag{2.34}$$

Subsequently, we establish the Gaussian rule of type $(H, N, \alpha, \beta, a, b)$ as the set of nodes $(x_i)_{i=1}^{nm}$ with corresponding weights $(w_i)_{i=1}^{nm}$ derived from the Gaussian quadrature rule of level $m$. This rule is applied to the intervals $[\xi_i, \xi_{i+1}]$ for $i = 0, \ldots, n-1$ with respect to the weight function $\nu$ as defined in (2.33). In addition, we set $x_0 = 0$ and

$$w_0 \coloneqq c_H \int_0^{\xi_0} x^{-H-1/2} dx = \frac{c_H}{1/2-H} \xi_0^{1/2-H}. \tag{2.35}$$

The approximation of $K$ in (2.32) is then given by

$$\hat{K} \coloneqq \sum_{i=0}^{nm} w_i e^{-x_i t}. \tag{2.36}$$

**Remark 2.13.** The positivity of the weights $w_i$ is ensured by the non-negativity property inherent to the weights obtained by Gaussian quadrature. A reminder of Gaussian quadrature is provided in Appendix B.

**Remark 2.14.** In geometric Gaussian rules, we have four parameters that can be freely chosen: $\alpha, \beta, a, b$. These parameters are interpreted as follows:

- The parameter $\alpha$ determines the cutoff point $\xi_n$ of the integral in (2.32), meaning we approximate the integral only over the interval $[0, \xi_n]$.
- The parameter $\beta$ sets the level of the Gaussian quadrature rule.
- The parameter $a$ specifies the size of the initial interval $[0, \xi_1]$, which is crucial due to the singularity in the weight function $\nu$.
- The parameter $b$ allows for fine-tuning the results.

In particular, parameter $\alpha$ primarily controls the approximation error on the interval $[\xi_n, \infty)$. Parameter $\beta$ is mainly responsible for managing the error on $[\xi_1, \xi_n]$ and parameter $a$ governs the error on $[0, \xi_1]$.

**Remark 2.15.** The partition defined in (2.34) is a geometric grid: we can identify the constants $c$ and $q$ in Definition 2.8 with

$$I_{k+1} - I_k = \xi_0 \left[ \left( \frac{\xi_n}{\xi_0} \right)^{\frac{1}{n}} - 1 \right] \left( \frac{\xi_n}{\xi_0} \right)^{\frac{k}{n}} = cq^k, \quad k \in \{0, \dots, n-1\}.$$

The following theorem demonstrates that the $L^2$-error between $V^{tr}$ and its approximation $\hat{V}$ (2.27) has an error bound of the form $\exp(-\alpha\sqrt{N}/A)$, when $\hat{K}$ in (2.36) is adopted.

**Theorem 2.8.** *[6, Theorem 2.1] Let $v_0 \in \mathbb{R}^d$, and let $b : \mathbb{R}^d \to \mathbb{R}^d$ and $\sigma : \mathbb{R}^{d\times d} \to \mathbb{R}^{d\times d}$ be globally Lipschitz continuous functions. Let $V^{tr}$ be the solution of* (2.29)*, and let $\hat{V}$ be the solution of* (2.27)*, where we use the Gaussian rule of type $(H, N, \alpha, \beta, 1, 1)$ with $\alpha \coloneqq 1.06418$ and $\beta \coloneqq 0.4275$. Then,*

$$\mathbb{E}\big[|\hat{V}_T - V_T^{tr}|^2\big] \le C c_H^2 \bigg( \frac{T^3}{(3/2-H)^2} + \frac{3}{2H^2} \tag{2.37}$$

$$+ \frac{5\pi^3}{48} (e^{\alpha\beta} - 1)^2 \frac{A^{2-2H} T^{2H}}{\beta^{2-2H} H} N^{1-H} \bigg) \exp\Big(-\frac{2\alpha}{A}\sqrt{N}\Big), \tag{2.38}$$

*where $C$ is the constant from Proposition 5.*

**Remark 2.16.** [6, Remark 2.2] Theorem 2.8 is not only valid for the specific values of $\alpha$ and $\beta$ given above, but for an entire set $\Gamma \subset \mathbb{R}^2_+$ of pairs $(\alpha, \beta)$. Specifically, the theorem asserts that $(1.06418, 0.4275) \in \Gamma$. While these values represent optimal choices

for the theorem, practical applications may achieve even better results with alternative selections of $(\alpha, \beta)$.

**Remark 2.17.** [6, Remark 2.5] The bound in Theorem 2.8 is completely non-asymptotic. In the proof, $m$ and $n$ are assumed to be real-valued, rather than integer-valued. One can still take the suggested point set of the theorem by simply rounding $m$ and $n$. As the theorem appears to overestimate the quadrature error, resulting in artificially small intervals $[\xi_0, \xi_n]$, it is convenient to round $m$ up to the next integer.

**Remark 2.18.** The positivity of the weights $(w_i)_{i=1}^N$ is ensured by the non-negativity property inherent to the weights obtained by Gaussian quadrature. A reminder of Gaussian quadrature is provided in Appendix B.

In [6, Remark 2.7], it is suggested to use a weight $w_0$ different from that is (2.35), which was employed in Theorem 2.8, to simplify the proof. The following proposition provides the exact $L^2$ error representation for $\hat{K}$ and $K^{tr}$, necessary to analyze the optimal $w_0$.

**Proposition 6.** *[6, Proposition 2.12.] Consider* $\hat{K}(t) = \sum_{i=0}^N w_i e^{-x_i t}$, *where* $x_0 = 0$ *and* $x_i > 0$ *for* $i = 1, \ldots, N$. *Then,*

$$\begin{aligned}\int_0^T |\hat{K}(t) - K^{tr}(t)|^2 dt = &\frac{T^{2H}}{2H\Gamma(H+1/2)^2} + w_0^2 T + 2w_0 \sum_{i=1}^N \frac{w_i}{x_i}(1 - e^{-x_i T}) \\ &+ \sum_{i=1}^N \sum_{j=1}^N \frac{w_i w_j}{x_i + x_j}(1 - e^{-(x_i+x_j)T}) - \frac{2w_0 T^{H+1/2}}{\Gamma(H+3/2)} \\ &- \frac{2}{\Gamma(H+1/2)} \sum_{i=1}^N \frac{w_i}{x_i^{H+1/2}} \int_0^{x_i T} t^{H-1/2} e^{-t} dt.\end{aligned} \tag{2.39}$$

**Remark 2.19.** [6, Remark 2.13] The expression (2.39) is almost explicit, except for the last integral term on the right-hand side. However, this term is a (lower) incomplete gamma function, which can be computed efficiently. The sum of the lower and upper incomplete gamma functions, referred to as "incomplete" because they integrate over only part of the region defined by the Gamma function, yields the Gamma function.

**Remark 2.20.** [6, Remark 2.14] The right hand side of (2.39) is a quadratic polynomial in $w_0$. The optimal $w_0$ is the one that minimizes (2.39). This choice of $w_0$ will be used in the numerical application instead of the $w_0$ specified in Theorem 2.8.

# 3

# Markovian Approximation of the Rough Bergomi Model

Stochastic volatility models, characterized by continuous yet less regular trajectories than Brownian motion and known as rough volatility models, have garnered significant attention in mathematical finance. These models, featuring fractional kernels, are well-suited to capture the stylized features of realized volatility time series and the implied volatility surface.

Recent statistical studies indicate that realized volatility oscillates more rapidly than Brownian motion across multiple time scales and markets [25]. The dynamics of the log-volatility are close to that of a fractional Brownian motion with a small Hurst parameter $H$ of order 0.1. Additionally, the implied volatility skew for short maturities is notably steeper than that predicted by classical diffusion models. As maturity decreases, the slope of the at-the-money (ATM) implied volatility skew follows a power law that diverges at zero. This behavior is accurately reproduced by rough volatility models [8], [20], [24]. Empirical findings are further supported by micro-structural considerations as rough volatility models naturally emerge as scaling limits of micro-structural pricing models with self-exciting features driven by Hawkes processes [21].

Such insights have driven the exploration of various rough volatility models in the literature, including the rough fractional stochastic volatility model [25], the rough Bergomi model [8], and the fractional and rough Heston models [18], [21]. The absence of semimartingale and Markov properties in these models complicates fundamental tasks such as simulation, pricing and control problems. To address these challenges, and drawing from the Markovian representation of fractional processes discussed in Chapter 2, we can develop tractable multi-factor stochastic volatility models that approximate rough volatility models while maintaining a Markovian structure.

In this chapter, we introduce rough volatility models (Section 3.1), with a particular focus on the rough Bergomi model, which serves as the central model of this work. We derive its Markovian representation and approximation (Section 3.2) and present the simulation scheme for the model based on the established Markovian approximation (Section 3.3).

In Section 3.1, we present rough volatility models, discussing the motivations behind their development and their suitability for modeling historical volatility dynamics as well as replicating the behavior of the implied volatility surface. We specifically focus on the rough Heston model, a specific instance of affine Volterra processes [3]. Thus, as outlined in the previous chapter, we can establish a Markovian approximation for this model, which leads to a multi-factor model known as the *lifted Heston model* [1].

In Section 3.2, we examine the rough Bergomi model and derive its Markovian approximation. Initially, we represent the fractional component of the rough Bergomi volatility as a linear functional of an infinite-dimensional Ornstein-Uhlenbeck process to obtain its Markovian representation. Next, we discretize this representation to derive the associated approximation, and establish an *ad hoc* convergence result in Theorem 3.1.

In Section 3.3, we outline the simulation scheme for the rough Bergomi model. This involves calculating the positive weights and nodes using Gaussian quadrature to approximate the fractional kernel in the model. We then use these weights and nodes to simulate the Markovian approximation of the rough Bergomi model. The accuracy of the implementation is validated by computing the $L^2$ error for the Riemann-Liouville process approximation and comparing it with existing literature results. Finally, simulations of the rough Bergomi model are performed across various parameter settings.

## 3.1. Rough Volatility Models

### 3.1.1. Introduction

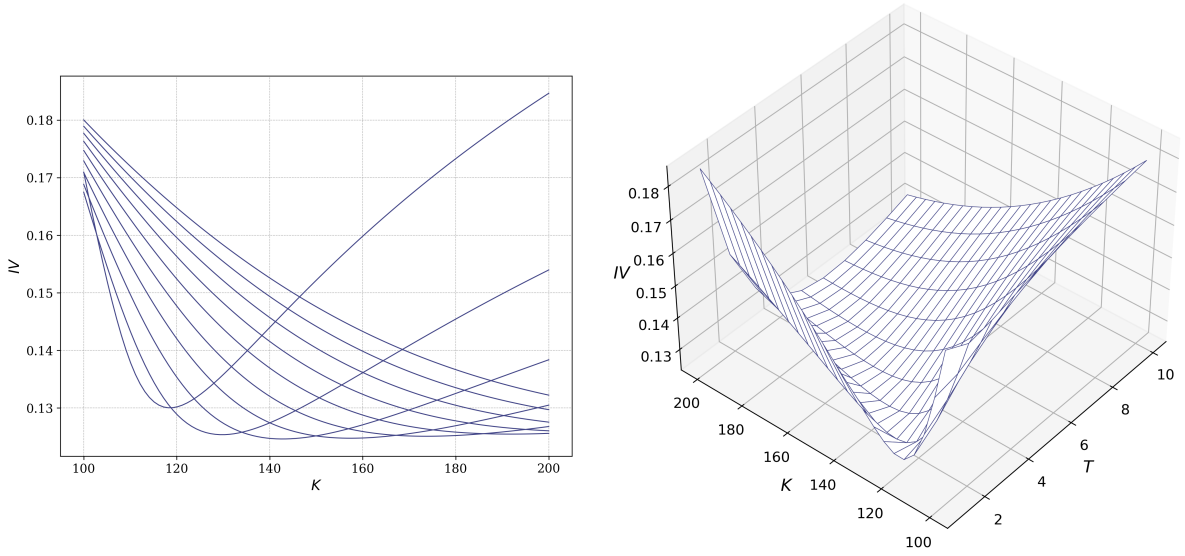
Financial derivatives are products that are based on the performance of some underlying assets such as stocks, interest rates, or commodity prices. The fundamental pricing partial differential equation for valuing European options on stocks, a type of derivative, is the renowned Black-Scholes equation, introduced by Black and Scholes in 1973 [12]. In their model, Black and Scholes consider a stock price $S$ that follows an Itô diffusion, described by:

$$dS_t = bS_t dt + \sigma S_t dB_t^S, \quad S_0 > 0,$$

where $b$ is the drift parameter, $\sigma$ denotes the volatility parameter and $B^S$ represents a standard Brownian motion. While this model facilitates efficient pricing and hedging of European options, it does not accurately reflect real-world market conditions. One significant limitation of the Black-Scholes model is the assumption of constant volatility, which fails to capture the *implied volatility smile* observed in many financial markets.

European options are often quoted in terms of their implied volatilities. The implied volatility, or Black-Scholes implied volatility denoted as $\sigma_{BS}$, is the unique volatility value that, when inserted as a parameter in the Black-Scholes option pricing model, matches the market price of the option. Numerical inversion of the Black-Scholes equation, based on market option prices for different strikes and a fixed maturity time, exhibits the so-called implied volatility skew or smile.

Figure 3.1a presents typical implied volatility shapes found in financial market data quotes, where the implied volatility is plotted against the strike price, $K$. Figure 3.1b depicts a typical implied volatility surface across both strike price (or moneyness) and time dimensions, demonstrating that the surface also varies over time. This variation is referred to as the *term-structure* of the implied volatility surface.




**(a)** Implied volatility shapes. **(b)** Implied volatility surface.

**Figure 3.1: (a)** Typical implied volatility shapes: smile and skew for fixed values of $T$. **(b)** Implied volatility surface exhibiting a pronounced smile for short maturities and a pronounced skew for longer maturities $T$.

To address the limitations of the Black-Scholes model, a new class of models known as stochastic volatility models has emerged. In these models, the volatility $V$ follows its own stochastic differential equation:

$$dV_t = a(t, V_t)dt + c(t, V_t)dB_t^V, \quad V_0 > 0,$$

where the coefficients $a$ and $c$ are some functions of $V$, and $B^V$ is a Brownian motion correlated with $B^S$. Stochastic volatility models, such as the Bergomi model or the Heston model, accurately reproduce the term structure of at-the-money skew (i.e., skew at the stike price $K = 0$) for long maturities. However, they fall short in explaining the observed behavior for very short maturities.

Specifically, let $\sigma_{BS}(k, \tau)$ represent the (Black-Scholes) implied volatility of an option with log-moneyness $k = \log K/S_t$ and time to maturity $\tau = T - t$. The term structure of at-the-money skew is defined as:

$$\psi(\tau) \coloneqq \left| \frac{\partial \sigma_{BS}(k, \tau)}{\partial k} \right|_{k=0}.$$

In conventional stochastic volatility models, the ATM volatility skew $\psi(\tau)$ is constant for short dates and inversely proportional to $\tau$ for long dates [8]. However, empirical studies

indicate a power-law decay of the ATM skew of option prices, with an explosion when $\tau$ approaches zero.

In [24], it is demonstrated that the term structure of a stochastic volatility model, where the volatility is driven by a fractional Brownian motion with Hurst parameter $H$, is given by $\psi(\tau) \sim \tau^{H-1/2}$, when $\tau \downarrow 0$. As shown in Figure 3.2, we observe that $\psi(\tau)$ is proportional to $\tau^{-\alpha}$ for some $0 < \alpha < 1/2$ over a wide range of expirations.

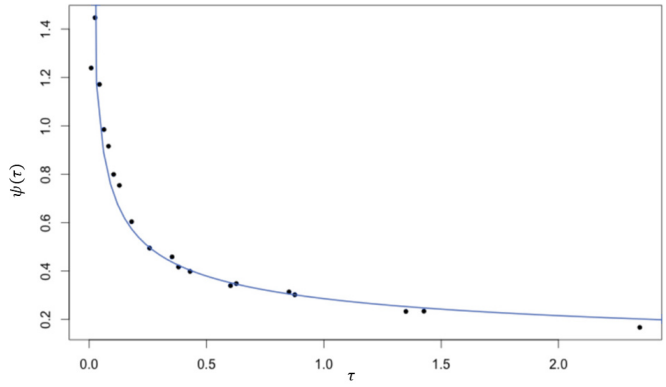



**Figure 3.2:** The black dots are non-parametric estimates of the S&P index at-the-money volatility skews as of August 14, 2013; the curve is the power-law fit $\psi(\tau) = A\tau^{-0.407}$ with $\tau$ measured in years [8, Figure 1.2].

The use of fractional Brownian motion was initiated by Comte and Renault [18], who proposed a log-volatility model based on the Mandelbrot-van Ness representation (Definition 2.3) of fractional Brownian motion:

$$B_t^H = b_0 + \frac{1}{\Gamma(H+1/2)} \left\{ \int_{-\infty}^{0} \left[ (t-s)^{H-1/2} - (-s)^{H-1/2} \right] dB_s + \int_0^t (t-s)^{H-1/2} dB_s \right\},$$

where $H \in (0,1)$ is the Hurst parameter governing the regularity of trajectories and the dependence structure of the increments of $B^H$, and $B$ is a two-sided Brownian motion. The sample paths of $B^H$ are Hölder continuous of any order strictly less than $H$ (Theorem 2.1) and for $H > 1/2$ the increments of $B^H$ have the long-range dependence property (Section 2.1). Comte and Renault chose the Hurst parameter $H > 1/2$ to ensure long memory. However, Gatheral et al. demonstrate in [25] that the log-volatility behaves essentially as a fractional Brownian motion with Hurst parameter $H$ around 0.1.

This observation has sparked growing interest in rough volatility models, as they can replicate the power-law decay of the ATM skew [8], [20], [24]. The term *rough volatility models* was coined to designate stochastic volatility models whose trajectories are continuous but exhibit rougher paths than those of a Brownian motion in terms of their Hölder regularity. Specifically, these trajectories are considered *rough* when the Hölder regularity is less than $1/2$.

Moreover, rough volatility models appear more consistent with financial time series data. Statistical evidence across multiple time scales supports the observation that realized

volatility time series oscillate more rapidly than those modeled by Brownian motion [25]. These insights are further supported by microstructural considerations, as rough volatility models emerge as scaling limits of microstructural pricing models driven by self-exciting dynamics driven by Hawkes processes [21], [37].

So far, the literature has examined various rough volatility models, including the rough fractional stochastic volatility (RFSV) model [25], the rough Bergomi model [8], and the rough Heston model [21]. A detailed exposition of the rough Bergomi model, which is the primary focus of this thesis, will be presented in the next section. Here, we provide a brief overview of the rough Heston model.

### 3.1.2. The Rough Heston Model

The rough Heston model, introduced by El Euch and Rosenbaum in [21], extends the Heston model by incorporating a fractional kernel in the variance dynamics, thereby serving as its rough analogue. In this model, the asset price $S$ evolves according to:

$$dS_t = S_t\sqrt{V_t}dB_t^S, \quad S_0 > 0, \tag{3.1}$$

while the instantaneous variance process $V$ is given by:

$$V_t = V_0 + \int_0^t \frac{(t-s)^{H-1/2}}{\Gamma(H+1/2)}\Big(\lambda(\theta - V_s)ds + \nu\sqrt{V_s}\Big)dB_s^V, \tag{3.2}$$

where $H \in (0, 1/2)$, the parameters $V_0, \lambda, \theta$ and $\nu$ are positive constants and $B^S, B^V$ are correlated Brownian motions. When $H = 1/2$, the model reverts to the standard Heston model.

**Remark 3.1.** Several variations of the rough Heston model have been developed since its introduction. Guennoun et al. [29] proposed an alternative formulation using fractional derivatives. Meanwhile, Jaber et al. [3] explored affine Volterra processes, leading to the development of the Volterra Heston model, of which the rough Heston model is a specific instance.

Similar to the Heston model, which is known for its closed-form expression of the characteristic function of the log-price, the rough Heston model benefits from a semi-closed formula for the same function. This facilitates fast pricing and calibration by Fourier inversion techniques. As demonstrated in [21], the characteristic function of $\log(S_t/S_0)$ can be expressed in terms of the solution to a fractional Riccati equation:

$$\mathbb{E}\exp(z\log(S_t/S_0)) = \exp\left(\int_0^t F(z, \psi(t-s, z))g(s)ds\right),$$

where

$$g(t) = V_0 + \int_0^t K(t-s)\theta(s)ds,$$

and $\psi(\cdot, z)$ is the unique continuous solution to the fractional Riccati equation

$$\psi(t, z) = \int_0^t K(t-s)F(z, \psi(s, z))ds, \quad t \in [0, T], \tag{3.3}$$

with

$$F(z,x) \coloneqq \frac{1}{2}(z^2 - z) + (\rho\nu z - \lambda)x + \frac{\nu^2}{2}x.$$

Given that equation (3.3) needs to be solved numerically, only a semi-explicit formula is available.

The same principles discussed in Section 2.2 apply here. We can approximate the fractional kernel $K$ underlying the processes (3.1) and (3.2) with a kernel $\hat{K}$ of the form $\hat{K}(t) = \sum_{i=1}^{N} w_i e^{-x_i t}$. Consequently, a multi-factor approximation $(\hat{S}, \hat{V})$ of the rough Heston model is obtained, the *lifted Heston model*, along with a semi-explicit formula for the characteristic function of $\log(\hat{S}/S_0)$ [2, Section 7.4].

### 3.1.3. The Lifted Heston Model

The Volterra Heston model[1] is an affine Volterra process $X = (\log S, V)$ with state space $\mathbb{R} \times \mathbb{R}_+$, where the price process $S$ and its variance process $V$ are described by:

$$\begin{cases} dS_t = S_t\sqrt{V_t}dB_t^S, \quad S_0 > 0, \\ V_t = g_0(t) + \displaystyle\int_0^t K(t-s)\Big(-\lambda V_s ds + \nu\sqrt{V_s}\Big)dB_s^V. \end{cases} \tag{3.4}$$

Here, $K \in L^2_{\text{loc}}(\mathbb{R}_+)$, $B^S$ and $B^V$ are two correlated Brownian motions, the parameters $\lambda, \nu \in \mathbb{R}_+$ and $g_0 : \mathbb{R}_+ \to \mathbb{R}$ is given by:

$$g_0(t) = V_0 + \int_0^t K(s)\lambda\theta ds, \quad t \geq 0, \quad \text{for some } V_0, \theta \geq 0.$$

Under conditions on the kernel and on the set of admissible input curves $g_0$, the existence of a weak unique solution $(S, V)$ valued in $\mathbb{R}_+^2$ of (3.4) is provided in [2, Theorem 6.1]. Given that $S$ is determined by $V$, this result primarily hinges on the weak existence of $V$.

The volatility process $V$ is itself a Volterra square root process. In the case of the rough Heston volatility (3.2), where $K(\tau) = \tau^{H-1/2}/\Gamma(H + 1/2)$, Theorem 2.5 ensures the weak existence of $V$. As outlined in Remark 2.11, the fractional kernel $K$ satisfies the assumptions of the mentioned theorem.

The fractional kernel $K$ is completely monotone (Definition A.1), thus, according to the Bernstein theorem (Theorem A.2), $K$ can be represented as the Laplace transform of a positive measure $\mu$

$$K(\tau) = \int_0^\infty e^{-x\tau}\mu(dx), \quad \mu(dx) = \frac{x^{-H-1/2}}{\Gamma(H+1/2)\Gamma(1/2-H)}.$$

Approximating the measure $\mu$ with a weighted sum of Dirac measures $\hat{\mu} \coloneqq \sum_{i=1}^{N} w_i \delta_{x_i}$, where $(w_i)_{i=1}^N$ are positive weights and $(x_i)_{i=1}^N$ are nodes, yields the following approximation for the kernel:

$$\hat{K}(\tau) = \sum_{i=1}^{N} w_i e^{-x_i\tau},$$

[1] More precisely, the extended Volterra Heston model, which allows arbitrary curves $g_0$ as input.

which, in turn, leads to a possible approximation of the rough Heston model

$$\begin{cases} d\hat{S}_t = \hat{S}_t\sqrt{\hat{V}_t}dB_t^S, \quad S_0 > 0, \\ \hat{V}_t = \hat{g}(t) + \sum_{i=1}^{N} w_i U_t^i, \end{cases} \tag{3.5}$$

with

$$\begin{aligned} dU_t^i &= (-x_i U_t^i - \lambda\hat{V}_t)dt + \nu\sqrt{\hat{V}_t}dB_t^V, \quad U_0^i = 0, \\ \hat{g}(t) &= V_0 + \int_0^t \hat{K}(t-s)\theta(s)ds. \end{aligned}$$

This formulation has been coined as the *lifted Heston model* in [1].

According to [2, Theorem 7.3], $(\hat{S}, \hat{V})$ is the unique $\mathbb{R}_+^2$-valued strong solution of (3.5). This result is analogous to Proposition 3, which establishes a strong uniqueness result for the approximation $\hat{V}$ when its coefficients are Lipschitz. However, this is not the case for $\sigma(x) = \nu\sqrt{x}$ in (3.2), thus, [2, Theorem 7.3] requires a weaker regularity condition (in terms of Hölder continuity) on $\sigma$.

The model (3.5) is Markovian and consists of $N+1$ state variables, including the spot price $\hat{S}$ and the variance factors $U = (U^i)_{i=1}^N$. As the factors $U$ are Markovian semimartigales, $\hat{V}$ represents the Markovian-semimartingale approximation of the volatility process $V$ (3.2) for a suitable choice of $\hat{g}$. This approximation is further supported by convergence results, which ensure that $(\hat{S}, \hat{V})$ converges in law to $(S, V)$ [2, Theorem 7.7].

The convergence of $(\hat{S}, \hat{V})$ is contingent upon the convergence of the approximated kernel $\hat{K}$ to $K$. The choice of positive weights $(w_i)_{i=1}^N$ and nodes $(x_i)_{i=1}^N$, as well as the set of auxiliary nodes $(\xi_i)_{i=1}^N$, is critical in this context. When the weights and nodes are defined as:

$$w_i = \int_{\xi_{i-1}}^{\xi_i} \mu(dx), \quad x_i = \frac{1}{w_i}\int_{\xi_{i-1}}^{\xi_i} x\mu(dx), \quad i \in \{1, \dots, N\}, \tag{3.6}$$

[2, Proposition 7.5] provides the conditions for the convergence of $\hat{K}$ to $K$ in $L^2([0,T])$, with a strong rate of convergence of order $N^{-4H/5}$. When the weights and nodes are derived using Gaussian quadrature, strong convergence is also assured, with an error bound of the form $\exp(-a\sqrt{N})$, as detailed in [6, Corollary 3.7].

## 3.2. Markovian Approximation of the Rough Bergomi Model

### 3.2.1. Introduction

The rough Bergomi model, introduced by Bayer et al. [8], has gained acceptance for stochastic volatility modeling due to its alignment with the stylized features of realized volatility time series and the implied volatility surface. However, the lack of semimartingale and Markov properties in this model complicates even fundamental tasks such as pricing European options. As discussed before, to overcome these difficulties, we

approximate the rough Bergomi model with a simpler, more practical version suitable for pricing, simulation, and stochastic control. Our primary focus is on the latter two applications.

Building on the Markovian representation from the previous chapter and inspired by prior works [2], [33], we introduce the rough Bergomi model and develop a finite Markovian approximation in two steps. First, we establish a suitable Markovian representation for its fractional component by expressing it as superposition of infinitely many Ornstein-Uhlenbeck processes, resulting in the Markovian representation $(S, V)$ of the model (Definition 3.2). Next, we discretize this representation to derive the corresponding Markovian approximation $(\hat{S}, \hat{V})$ (Definition 3.3).

As $S$ is determined by $V$, or more precisely by $(\int_0^t V_s ds, \int_0^t \sqrt{V}_s dB_s^S)$, see (3.16), $\hat{S}$ is likewise determined by $\hat{V}$. Therefore, when establishing convergence results our primary focus is on the approximated volatility $\hat{V}$. Concerning the portfolio optimization problem, we establish the requisite convergence result in Theorem 3.1. This theorem demonstrates the almost sure monotone convergence of $\hat{\mathcal{V}}$ to $\mathcal{V}$, where these processes share the same distributions as $\hat{V}$ and $V$, respectively. The process $\hat{V}$ is constructed using the fractional kernel $\hat{K}$, which employs Gaussian weights and nodes.

### 3.2.2. The Rough Bergomi Model

Building on the foundational work on rough volatility [25], Bayer et al., in their paper 'Pricing Under Rough Volatility' [8], illustrate how the rough fractional stochastic volatility (RFSV) model provides a natural framework for option pricing. We now outline the main steps leading to the development of the rough Bergomi model.

Consider the simple RFSV model initially proposed in [25], which assumes the following dynamics for the increments of log-volatility

$$\log \sigma_{t+\Delta} - \log \sigma_t = \nu(B^H_{t+\Delta} - B^H_t), \tag{3.7}$$

for $H < 1/2$ and $\nu > 0$. The adopted integral representation for $B_t^H$ is a slight modification to the Mandelbrot-Van Ness representation given in Definition 2.3. By setting the initial condition $b_0 = 0$, it is given by:

$$B_t^H = C_H \left\{ \int_{-\infty}^t (t-s)^{H-1/2} dB_s - \int_{-\infty}^0 (-s)^{H-1/2} dB_s \right\}, \tag{3.8}$$

where

$$C_H := \left( \frac{2H\Gamma(3/2 - H)}{\Gamma(H + 1/2)\Gamma(2 - 2H)} \right)^{1/2}. \tag{3.9}$$

The constant $C_H$ is chosen to ensure that the representation (3.8) defines a fractional Brownian motion, a mean-zero Gaussian process with the covariance function given in Definition 2.1.

Letting $V_t = \sigma_t^2$ and substituting equation (3.8) into (3.7), yields, for $t \geq u$,

$$\begin{aligned}\log V_t - \log V_u &= 2\nu C_H \left\{ \int_{-\infty}^{t} (t-s)^{H-1/2} dB_s - \int_{-\infty}^{u} (u-s)^{H-1/2} dB_s \right\} \\ &= 2\nu C_H \left\{ \int_{u}^{t} (t-s)^{H-1/2} dB_s + \int_{-\infty}^{u} \left( (t-s)^{H-1/2} - (u-s)^{H-1/2} \right) dB_s \right\} \\ &= 2\nu C_H \left\{ M_u(t) + Z_u(t) \right\},\end{aligned}$$

where the notation $M_u(t)$ indicates that $M = (M_t)_{t \geq u}$. We observe that the process $Z$ is $\mathcal{F}_u$-measurable, where $\mathcal{F}_u$ denotes the filtration supporting the two-sided Brownian motion $B$, while the process $M$ is independent of $\mathcal{F}_u$.

Introducing the Riemann-Liouville process (cf. Definition 2.4)

$$\tilde{B}_u^H(t) \coloneqq \sqrt{2H} \int_u^t (t-s)^{H-1/2} dB_s,$$

we define $\eta \coloneqq 2\nu C_H / \sqrt{2H}$, leading to

$$\log V_t - \log V_u = \eta \tilde{B}_u^H(t) + \eta \sqrt{2H} Z_u(t),$$

and

$$V_t = V_u \exp\left( \eta \tilde{B}_u^H(t) + \eta \sqrt{2H} Z_u(t) \right).$$

Thus, since $V$ and $Z$ are $\mathcal{F}_u$-measurable, we obtain

$$\mathbb{E}[V_t \mid \mathcal{F}_u] = V_u \exp\left( \eta \sqrt{2H} Z_u(t) \right) \mathbb{E}\left[ \exp\left( \eta \tilde{B}_u^H(t) \right) \right]. \tag{3.10}$$

As a consequence,

$$\begin{aligned}V_t &= V_u \exp\left( \eta \tilde{B}_u^H(t) + \eta \sqrt{2H} Z_u(t) \right) \\ &= \mathbb{E}[V_t \mid \mathcal{F}_u] \exp\left( \eta \tilde{B}_u^H(t) - \frac{1}{2} \eta^2 \mathbb{E}\left[ |\tilde{B}_u^H(t)|^2 \right] \right).\end{aligned}$$

Summarizing, Bayer et al. [8] proposed the following model, for $t \geq u$:

$$\begin{cases} dS_t = S_t(\mu_t dt + \sqrt{V_t} dB_t^S), \quad S_u > 0, \\ V_t = \mathbb{E}[V_t \mid \mathcal{F}_u] \exp\left( \eta \tilde{B}_u^H(t) - \frac{1}{2} \eta^2 \mathbb{E}\left[ |\tilde{B}_u^H(t)|^2 \right] \right), \end{cases} \tag{3.11}$$

where $B^S$ is another Brownian motion, which is correlated to the Brownian motion $B$ driving $Z$, and where $\mu$ is an appropriate drift term.

In the original paper, the model (3.11) is not employed for pricing purposes as it is framed under the real probability measure, $\mathbb{P}$. Therefore, it is reformulated with respect to the risk-neutral measure: $\mathbb{Q} \sim \mathbb{P}$ on $[u, T]$. Considering the change of measure

$dB_s = dB_s^{\mathbb{Q}} + \lambda(s)ds$[2], where $\lambda(s)$ is assumed to be a deterministic function of $s$, the rough Bergomi model is then expressed as follows, for $u \leq t \leq T$:

$$\begin{cases} dS_t = S_t\sqrt{V_t}dB_t^{S,\mathbb{Q}}, \quad S_u > 0, \\ V_t = \xi_u(t)\exp\left(\eta\tilde{B}_u^{\mathbb{Q}}(t) - \frac{1}{2}\eta^2\mathbb{E}\left[|\tilde{B}_u^{\mathbb{Q}}(t)|^2\right]\right). \end{cases} \tag{3.12}$$

Here, $\xi_u(t)$, termed forward variance, is defined as the expectation under the pricing measure of the instantaneous variance, that is $\xi_u(t) = \mathbb{E}^{\mathbb{Q}}[V_t|\mathcal{F}_u]$, and it holds that $\xi_t(t) = V_t$. Specifically, the forward variance has the form

$$\xi_u(t) = \mathbb{E}^{\mathbb{Q}}[V_t|\mathcal{F}_u] = \mathbb{E}[V_t|\mathcal{F}_u]\exp\left(\eta\sqrt{2H}\int_u^t (t-s)^{H-1/2}\lambda(s)ds\right).$$

The term $\mathbb{E}[V_t|\mathcal{F}_u]$ as given in (3.10) depends on the history of the driving Brownian motion $B$, with this dependency encapsulated in the reliance of the process $Z$ on the entire history of $B$. The second term involving $\lambda(s)$, alter the marginal distribution of $V$. Although the conditional distribution of $V$ under $\mathbb{P}$ is lognormal, it is not generally lognormal under $\mathbb{Q}$.

Since our focus is not on pricing, the rough Bergomi model we will use is derived from (3.11). Prior to defining it, we reformulate the rough Bergomi volatility to align its expression with existing literature that employs the Markovian approximation of the rough Bergomi model [4], [6], [64]. We set the simulation's origin at $u = 0$, assume $\mathbb{E}[V_t|\mathcal{F}_0] \equiv V_0 > 0$ for $t \geq 0$, and use standard notation for stochastic processes: $\tilde{B}^H = (\tilde{B}_t^H)_{t\geq 0}$ and $B^V = (B_t^V)_{t\geq 0}$. The volatility process in (3.11) then becomes:

$$V_t = V_0\exp\left(\eta\sqrt{2H}\int_0^t (t-s)^{H-1/2}dB_s^V - \frac{\eta^2}{2}t^{2H}\right), \quad V_0 > 0. \tag{3.13}$$

where we have used the Itô isometry to compute

$$\mathbb{E}\left[(\tilde{B}_t^H)^2\right] = \int_0^t 2H(t-s)^{2H-1}ds = t^{2H}. \tag{3.14}$$

Let $(\Omega, \mathcal{F}, \mathbb{F} = (\mathcal{F}_t)_{t\in[0,T]}, \mathbb{P})$ be a stochastic basis, and let the Hurst index $H \in (0, 1/2)$. We define the rough Bergomi model as follows.

**Definition 3.1.** *The rough Bergomi stochastic volatility model takes the form*

$$\begin{cases} dS_t = S_t\mu_t dt + S_t\sqrt{V_t}dB_t^S, \quad S_0 > 0, \\ V_t = V_0\exp\left(\eta\sqrt{2H}\int_0^t (t-s)^{H-1/2}dB_s^V - \frac{\eta^2}{2}t^{2H}\right), \quad V_0 > 0, \end{cases} \tag{3.15}$$

*where the parameter $\eta = 2\nu C_H/\sqrt{2H}$ with $C_H$ defined as in* (3.9).

[2]In our notation, expectations and stochastic processes without a superscript refer to the probability measure $\mathbb{P}$. Superscripts are used exclusively for the martingale measure $\mathbb{Q}$.

Here, $B^S, B^V : [0,T] \to \mathbb{R}$ are two correlated standard Brownian motions supported by the filtration $\mathbb{F}$ with correlation coefficient $\rho \in (-1,1)$, typically negative. The process $\mu : [0,T] \times \Omega \to \mathbb{R}$ is a predictable bounded process. The Hurst parameter $H$ governs the decay of the term structure of volatility skew for very short expirations and is usually around 0.1, more information on how to choose these parameters can be found in [8].

**Remark 3.2.** Note that the volatility $V$ in Definition 3.1 is given explicitly, ensuring its existence. Furthermore, its uniqueness follows from the fact that it is an injective function of a Riemann-Liouville process (Definition 2.4). The Riemann-Liouville process, being a Gaussian process with a specified covariance and mean, is unique [26]. Although the dynamics of the stock price in the definition do not satisfy the linear or quadratic growth condition due to the term $\sqrt{V}$, the existence of the solution $S$ is assured from its explicit expression (3.16). For further details on existence and pathwise uniqueness of the rough Bergomi model, we refer the reader to [35].

**Remark 3.3.** The solution to the stock price's SDE in Definition 3.1 can be obtained through the standard procedure consisting of applying the Itô formula to $\log S$, and it is given by:

$$S_t = S_0 \exp\left(\int_0^t \left(\mu_s - \frac{1}{2}V_s\right)ds + \int_0^t \sqrt{V_s}dB_s^S\right), \quad S_0 > 0. \tag{3.16}$$

Now, we aim to find the Markovian representation and approximation for the fractional component appearing in the rough Bergomi volatility (3.13). Since the fractional part is a Rienmann-Liouville process (Definition 2.4), its Markovian representation is a linear functional of infinitely many Ornstein-Uhlenbeck processes $Y_t^x$ with varying speed of mean reversion $x$ (Theorem 2.4). The Markovian approximation will follow from the discretization of this representation.

### 3.2.3. Markovian Representation and Approximation

From Definition 3.1 and Theorem 2.17, we formulate the Markovian representation of the rough Bergomi model. We replace the Rienmann-Liouville process appearing in the rough Bergomi volatility (3.13) with its Markovian representation:

$$\begin{aligned}\int_0^t (t-s)^{H-1/2}dB_s &= \Gamma(H+1/2)\int_0^\infty Y_t^x \mu(dx) \\ &= \Gamma(H+1/2)\int_0^\infty \int_0^t e^{-x(t-s)}dB_s\mu(dx),\end{aligned} \tag{3.17}$$

recalling that $\mu(dx) = \nu(x)dx$, where for $x > 0$

$$\nu(x) = c_H x^{-H-1/2}, \quad c_H = \frac{1}{\Gamma(H+1/2)\Gamma(1/2-H)}.$$

Motivated by (3.14), we also rewrite the term $t^{2H}$ appearing in (3.13) as follows:

$$\begin{aligned} t^{2H} &= 2H\int_0^t (t-s)^{2H-1}ds = 2H\mathbb{E}\left[\left(\int_0^t (t-s)^{H-1/2}dB_s\right)^2\right] \\ &= 2H\Gamma(H+1/2)^2\mathbb{E}\left[\left(\int_0^\infty Y_t^x\mu(dx)\right)^2\right] \\ &= 2H\Gamma(H+1/2)^2\mathbb{E}\left[\left(\int_0^\infty\int_0^t e^{-x(t-s)}dB_s\mu(dx)\right)^2\right], \end{aligned}$$

where in the second last step we have used (3.17). Applying the stochastic Fubini Theorem A.1 and the Itô isometry yields

$$\begin{aligned} t^{2H} &= 2H\Gamma(H+1/2)^2\mathbb{E}\left[\left(\int_0^t\int_0^\infty e^{-x(t-s)}\mu(dx)dB_s\right)^2\right] \\ &= 2H\Gamma(H+1/2)^2\int_0^t\left(\int_0^\infty e^{-x(t-s)}\mu(dx)\right)^2 ds. \end{aligned}$$

**Definition 3.2.** *The Markovian representation of the rough Bergomi model $(S,V)$ is given by*

$$\begin{cases} dS_t = S_t\mu_t dt + S_t\sqrt{V_t}dB_t^S, \quad S_0>0, \\ V_t = V_0\exp\left(C_1\int_0^\infty Y_t^x\mu(dx) - C_2\int_0^t\left(\int_0^\infty e^{-x(t-s)}\mu(dx)\right)^2 ds\right), \quad V_0>0, \end{cases}$$

*where $C_1 \coloneqq \eta\sqrt{2H}\Gamma(H+1/2)$, $C_2 \coloneqq \eta^2 H\Gamma(H+1/2)^2$, the parameter $\eta = 2\nu C_H/\sqrt{2H}$ with $C_H$ defined as in* (3.9).

Given the representation in Definition 3.2, we now proceed with its discretization. The measure $\mu$ appearing in Definition 3.2 is approximated by a weighted sum of Dirac measures $\hat{\mu} \coloneqq \sum_{i=1}^N w_i\delta_{x_i}$, where $(w_i)_{i=1}^N$ are positive weights and $(x_i)_{i=1}^N$ are the corresponding nodes.

**Definition 3.3.** *The Markovian approximation of the rough Bergomi model $(\hat{S},\hat{V})$ is given by*

$$\begin{cases} d\hat{S}_t = \hat{S}_t\mu_t dt + \hat{S}_t\sqrt{\hat{V}_t}dB_t^S, \quad S_0>0, \\ \hat{V}_t = V_0\exp\left(C_1\sum_{i=1}^N w_i\int_0^t e^{-x_i(t-s)}dB_s^V - C_2\int_0^t\left(\sum_{i=1}^N w_i e^{-x_i(t-s)}\right)^2 ds\right), \quad V_0>0, \end{cases}$$

*where $C_1 \coloneqq \eta\sqrt{2H}\Gamma(H+1/2)$, $C_2 \coloneqq \eta^2 H\Gamma(H+1/2)^2$, the parameter $\eta = 2\nu C_H/\sqrt{2H}$ with $C_H$ defined as in* (3.9).

**Remark 3.4.** It is important to note that neither $S$ nor $\hat{S}$ can be simulated exactly; both require time discretization of (3.16) that define $S$ and $\hat{S}$, respectively. The same applies to $V$ and $\hat{V}$. So far, a comprehensive weak error analysis of the Euler-Maruyama scheme for $S$ is still lacking [6, Remark 3.4].

We are now left to justify the convergence of $(\hat{S}, \hat{V})$ to the Markovian representation $(S, V)$ and, thus, to the original rough model. Since $\hat{S}$ is determined by $\hat{V}$,

$$\hat{S}_t = S_0 \exp\left(\int_0^t \left(\mu_s - \frac{1}{2}\hat{V}_s\right)ds + \int_0^t \sqrt{\hat{V}_s} dB_s^S\right), \quad S_0 > 0,$$

our focus lies on establishing convergence results for the approximated volatility $\hat{V}$.

The following theorem ensures the monotone almost sure convergence of $\hat{\mathcal{V}}$ to $\mathcal{V}$. The processes $\hat{\mathcal{V}}$ and $\mathcal{V}$ share the same distribution of $\hat{V}$ and $V$, more precisely they are the representations of $\hat{V}$ and $V$ we can construct using the Skorokhod representation theorem (Theorem A.3). Here, $\hat{V}$ is the approximated volatility process when its kernel $\hat{K}$ is constructed using Gaussian weights and nodes as described in Section 2.2.

**Remark 3.5.** In our thesis, we will focus on the convergence of $\hat{V}$ to $V$ only within the conditional expectation, cf. Theorem 4.3. Therefore, establishing the convergence of $\hat{\mathcal{V}}$ to $\mathcal{V}$ suffices for our purposes.

**Theorem 3.1.** *Let $V$ be the rough Bergomi volatility in* (3.13)*, and $\hat{V}$ the approximated volatility given in Definition 3.3 and obtained using Gaussian quadrature. On some probability space, there exist processes $\hat{\mathcal{V}}$ and $\mathcal{V}$ distributed as $V$ and $\hat{V}$, such that $\hat{\mathcal{V}}$ converges almost surely to $\mathcal{V}$ and this convergence is monotone: for $0 \leq t \leq T$,*

$$\hat{\mathcal{V}}_t \uparrow \mathcal{V}_t \ \ a.s., \quad as\ N \to \infty. \tag{3.18}$$

*Proof.* We start by proving that, as $N \to \infty$,

$$\hat{K}(\tau) = \sum_{i=1}^{N} w_i e^{-x_i \tau} \quad \uparrow \quad K(\tau) = \int_0^\infty e^{-x(\tau)} \mu(dx), \tag{3.19}$$

for $\tau \in [0, T]$. Both the monotonicity and the pointwise convergence of $\hat{K}$ to $K$ follow from the properties of Gaussian quadrature, as presented in Appendix B.

According to Section 2.2, we approximate $K$ using an $m$-point Gaussian quadrature rule:

$$\hat{K}(\tau) = \sum_{i=1}^{nm} w_i e^{-x_i \tau},$$

where $m \approx C\sqrt{N}$ for $C > 0$ and $n \approx N/m$. We round $m$ to the next integer and select $n$ such that $nm$ is as close as possible to $N$. Specifically, we truncate the domain $(0, \infty)$ of the measure $\mu$ to $[\xi_0, \xi_n]$. For each subinterval $[\xi_i, \xi_{i+1}]$ where $i = 0, \ldots, n-1$, we apply the Gaussian quadrature rule to approximate

$$K(\tau) = \int_{\xi_i}^{\xi_{i+1}} e^{-x\tau} \mu(dx) \quad \text{with} \quad \hat{K}(\tau) = \sum_{j=1}^{m} w_j e^{-x_j \tau},$$

for $\tau \in [0, T]$, with $\mu(dx) = \nu(x)dx = C_H x^{-H-1/2} dx$ and $C_H > 0$, as given in (3.9).

Theorem B.2 ensures monotonicity. To apply this theorem, we need to verify the assumption on the function $f$, specifically that the derivative of order $2m$, $f^{(2m)}$, is

continuous and non-negative, where $m$ denotes the level of the quadrature rule. In our case, this condition is met because in the expression for $\hat{K}$ we have $f(x) = e^{-x\tau}$, which is of class $C^\infty$ w.r.t $x$.

The pointwise convergence of $\hat{K}$ to $K$ is an immediate consequence of the definition of the integral as a limit of Riemann-Stieltjes sum (Definitions B.1 and B.2). Theorem B.3 allows us to demonstrate that the Gaussian quadrature rule is a Riemann-Stieltjes sum. To apply the theorem, we need to verify the strict positivity of the weight function $\nu(x)$ for $x \in [a, b]$. This assumption is met because, for $x \in [\xi_0, \xi_n]$ we have $\nu(x) = C_H x^{-H-1/2}$ with $C_H > 0$ as given in (3.9).

Consequently, on each subinterval $[\xi_i, \xi_i + 1]$ for $i = 0, \ldots, n-1$, there exist numbers $y_j$ for $j = 0, \ldots, m$ where

$$\xi_i = y_0 \leq x_1 \leq y_1 \leq x_2 \leq \ldots \leq x_m \leq y_m = \xi_{i+1},$$

such that

$$w_j = \int_{y_{j-1}}^{y_j} \nu(u)du \quad \text{for } j = 1, \ldots, m.$$

Thus, on each subinterval $[\xi_i, \xi_i + 1]$, $\hat{K}$ is a Riemann-Stieltjes sum:

$$\hat{K}(\tau) = \sum_{j=1}^{m} e^{-x_j\tau} \int_{y_{j-1}}^{y_j} \nu(u)du = \sum_{j=1}^{m} e^{-x_j\tau}[F(y_j) - F(y_{j-1})],$$

where $F$ denotes the primitive of $\nu$. Since $\nu(x)$ for $x \geq 0$ is monotonic and $f(x) = e^{-x\tau}$ is continuous w.r.t. $x$, $f$ is Riemann integrable with respect to $F$ on $[\xi_i, \xi_i + 1]$ [60, Theorem 6.9]. Therefore, for $i = 0, \ldots, n-1$, $\hat{K}$ converges to $K$ on each subinterval

$$\lim_{m\to\infty} \sum_{j=1}^{m} e^{-x_j\tau} \int_{y_{j-1}}^{y_j} \nu(u)du = \int_{\xi_i}^{\xi_{i+1}} e^{-x\tau}\nu(u)du$$

and (3.19) follows.

Recalling that for $t \in [0, T]$, $V_0 > 0$

$$\hat{V}_t = V_0 \exp\left(C_1 \int_0^t \sum_{i=1}^{N} w_i e^{-x_i(t-s)} dB_s^V - C_2 \int_0^t \Big(\sum_{i=1}^{N} w_i e^{-x_i(t-s)}\Big)^2 ds\right),$$
$$V_t = V_0 \exp\left(C_1 \int_0^t \int_0^\infty e^{-x_i(t-s)}\mu(dx)dB_s^V - C_2 \int_0^t \Big(\int_0^\infty e^{-x(t-s)}\mu(dx)\Big)^2 ds\right),$$

we begin by considering the first term in the exponent of $\hat{V}_t$ and $V_t$. We write

$$\int_0^t \sum_{i=1}^{N} w_i e^{-x_i(t-s)} dB_s^V = \int_0^t \hat{K}(t-s)dB_s^V,$$
$$\int_0^t \int_0^\infty e^{-x(t-s)}\mu(dx)dB_s^V = \int_0^t K(t-s)dB_s^V,$$

and note that,

$$\mathbb{E}\left[\Big(\int_0^t \hat{K}(t-s)dB_s^V - \int_0^t K(t-s)dB_s^V\Big)^2\right] = \mathbb{E}\left[\int_0^t \Big(\hat{K}(t-s) - K(t-s)\Big)^2 ds\right] \to 0,$$

due to the Itô isometry and the monotone convergence of $\hat{K}$ to $K$ on the product space $\Omega \times [0,T]$. Therefore,

$$\int_0^t \hat{K}(t-s)dB_s^V \xrightarrow{L^2} \int_0^t K(t-s)dB_s^V.$$

Since $L^2$ convergence implies convergence in probability, which in turn implies convergence in distribution [30, Theorem 3.1], the Skorokhod representation Theorem A.3 guarantees that, on a suitable probability space we can find two random variables $\hat{Y}$ and $Y$ distributed as $\int_0^t \hat{K}(t-s)dB_s^V$ and $\int_0^t K(t-s)dB_s^V$, such that

$$\lim_{N\to\infty} \hat{Y} = Y \quad a.s.$$

Regarding the second term in the exponent of $\hat{V}$ we have the pointwise limit

$$\lim_{N\to\infty} \int_0^t \Big(\hat{K}(t-s)\Big)^2 ds = \int_0^t \Big(K(t-s)\Big)^2 ds,$$

justified by monotone convergence (3.19). We can combine the above results to obtain:

$$\begin{aligned}\lim_{N\to\infty} \hat{\mathcal{V}}_t &= \lim_{N\to\infty} V_0 \exp\left(C_1\hat{Y} - C_2 \int_0^t \Big(\hat{K}(t-s)\Big)^2 ds\right) \\ &= V_0 \exp\left(C_1 Y - C_2 \int_0^t \Big(K(t-s)\Big)^2 ds\right) = \mathcal{V}_t.\end{aligned}$$

Here, for $0 \le t \le T$, $\hat{\mathcal{V}}$ and $\mathcal{V}$ denotes the two processes that share the same distribution as $\hat{V}$ and $V$ on some probability space.

□

**Remark 3.6.** If the approximation of $\hat{K}$ is not done using Gaussian quadrature, but instead with the choice of weights and nodes ([2], [4], [5]) given by

$$w_i = \int_{\xi_{i-1}}^{\xi_i} \mu(dx), \quad x_i = \frac{1}{w_i}\int_{\xi_{i-1}}^{\xi_i} x\mu(dx) \quad \text{for } i \in \{1,\dots,N\}$$

for an appropriate choice of the auxiliary nodes $(\xi_i)_{i=0}^N$ satisfying (**Hz**), the convergence of $\hat{K}$ to $K$ is still monotone and pointwise [5, Lemma 3.1].

## 3.3. The Simulation Scheme

### 3.3.1. Introduction

In the context of rough volatility, simulation becomes particularly challenging due to the model's non-Markovian nature. There is no consensus on a standardized method for approximation and simulation [48]. For the rough Bergomi model, one strategy employs the hybrid scheme developed by [9], which uses the power-law component of the

kernel without error for a certain number of $\kappa$ steps near the origin, approximating it as piecewise constant otherwise. Alternatively, as explored in the previous chapter, the fractional kernel can be approximated as a sum of exponentials, a method implemented by [32]. Hybrid methods that combine the strengths of both approaches, such as the hybrid multifactor scheme introduced by [59], are also available.

A comprehensive comparison of these methods is currently lacking in the literature. However, [32] demonstrates that its scheme is less effective in terms of complexity compared to the hybrid scheme of [9], which is further improved by the multifactor scheme of [59]. In this work, we build on [6] by adopting the approach of approximating the fractional kernel as a sum of exponentials. This method, inspired by [32], improves upon it by enhancing the rate of convergence. However, the scheme in [6] has not yet been compared with the hybrid scheme of [9] or the hybrid multifactor scheme of [59]. Notably, further developments to the approach in [6] are presented in [7].

The simulation scheme consists of two main parts: first, we compute the positive $(w_i)_{i=1}^N$ weights and nodes $(x_i)_{i=1}^N$ using Gaussian quadrature, second, we simulate the Markovian approximation of the rough Begormi model given in Definition 3.3 with the computed Gaussian weights and nodes.

Given a finite interval $[a, b]$, a continuous weight function $\nu : [a, b] \to [0, \infty)$, and a function $f : [a, b] \to \mathbb{R}$ that we wish to integrate, the $m$-point Gaussian rule quadrature establishes the weights and nodes such that:

$$\int_a^b f(x)\nu(x)dx \;\approx\; \sum_{i=1}^{m} w_i f(x_i). \tag{3.20}$$

More precisely, the Gaussian quadrature rule of level $m$ is the unique choice of nodes and weights that integrates all polynomials of degree at most $2m-1$ exactly. This means that for all polynomials of degree at most $2m-1$, we have equality in (3.20). A summary of Gaussian quadrature can be found in Appendix B.

Specifically, we are interested in the approximation:

$$K(\tau) = \int_a^b e^{-x\tau}\nu(x)dx \;\approx\; \hat{K}(\tau) = \sum_{i=1}^{m} w_i e^{-x_i\tau}. \tag{3.21}$$

Since, in practice, we must truncate the domain of the measure $\mu(x) = \nu(x)dx$ from $(0, \infty)$ to a finite interval $[\xi_0, \xi_n]$, in this section we refer to the volatility process

$$V_t = V_0 \exp\left(C_1 \int_{\xi_0}^{\xi_n} Y_t^x \mu(dx) - C_2 \int_0^t \left(\int_{\xi_0}^{\xi_n} e^{-x(t-s)}\mu(dx)\right)^2 ds\right),$$

with $V_0 > 0$ and $Y_t^x = \int_0^t e^{-x(t-s)}dB_s^V$. We also drop the superscript denoting the truncated non-approximated processes introduced in Section 2.2. We refer to $K$ instead of $K^{tr}$, $B^H$ instead of $B^{H,tr}$ and $V$ instead of $V^{tr}$ for less cumbersome notation.

After calculating the Gaussian weights and nodes via the Golub-Welsh algorithm, we proceed to simulate the Markovian approximation of the rough Bergomi model using

Cholesky decomposition. The time-homogeneous grid allows us to simulate the model through a single Cholesky decomposition of the covariance matrix of the associated Gaussian vector. Subsequently, we conduct multiple simulations of the rough Bergomi model (3.3) across various parameter settings.

We validate the correctness of our implementation by computing the $L^2$ error for the approximation of the Riemann-Liouville process, as given in Definition 2.4. This error is compared with the results obtained in [6], varying the total number of nodes. A Python implementation is made available at: `https://github.com/gbifronte/thesis_simulation_scheme`.

### 3.3.2. The Gaussian Quadrature

We implement the Gaussian rule of type $(H, N, \alpha, \beta, 1, 1)$ detailed in Theorem 2.8. We are given a finite interval $[\xi_0, \xi_n]$ on the positive half-line, which is divided into subintervals by a sequence of auxiliary nodes $(\xi_i)_{i=0}^n$. This interval represents the truncated domain of the measure $\mu(dx) = \nu(x)dx$, where for $x > 0$, the weight function $\nu$ is given by

$$\nu(x) = c_H x^{-H-1/2}, \quad c_H = \frac{1}{\Gamma(H+1/2)\Gamma(1/2-H)}. \tag{3.22}$$

On each subinterval $[\xi_i, \xi_{i+1}]$ for $i = 0, \dots, n-1$, the measure $\mu$ is approximated by the discrete measure $\hat{\mu} = \sum_{i=0}^m w_i \delta_{x_i}$, obtained from an $m$-point Gaussian quadrature rule. Across the entire interval $[\xi_0, \xi_n]$, this discretisation results in a total of $N = nm$ nodes with corresponding positive weights.

Given the total number of nodes $N$ and the Hurst index $H \in (0, 1/2)$, we define the parameters $A, \alpha, \beta, a$ and $b$ in accordance to the setting of Theorem 2.8, along with the level $m$ of the quadrature and the number $n$ of intervals into which $[\xi_0, \xi_n]$ is divided. It is important to note that $m$ is rounded up to the next integer (see Remark 2.17), while $n$ is chosen such that $nm$ is as close as possible to $N$. After initializing the geometric grid $(\xi_i)_{i=0}^n$, we proceed with the implementation of the Gaussian quadrature.

To compute Gaussian nodes and weights for the setting described in (3.20), one begins by finding orthogonal polynomials $p_n$ of degree $n$, where the orthogonality is defined with respect to the scalar product:

$$\langle f, g \rangle = \int_a^b f(x)g(x)\nu(x)dx.$$

These orthogonal polynomials can be efficiently obtained using the three-term recurrence relation:

$$p_{n+1}(x) = \left( x - \frac{\langle x p_n, p_n \rangle}{\langle p_n, p_n \rangle} \right) p_n(x) - \frac{\langle p_n, p_n \rangle}{\langle p_{n-1}, p_{n-1} \rangle} p_{n-1}(x), \tag{3.23}$$

for $n = 0, 1, \dots$, where $p_{-1}(x) \equiv 0$ and $p_0(x) \equiv 1$. Subsequently, the nodes for the $m$-point quadrature rule are determined by the $m$ roots of $p_m$. Once the nodes are found, the weight $w_i$, corresponding to the node $x_i$, can be computed [57].

Numerical challenges may arise when computing nodes and coefficients for general weight functions $\nu$, as is the case in our study. An efficient algorithm for computing these nodes and weights, accommodating general weight functions, is the Golub-Welsch algorithm, developed in [27]. Rather than implementing the Golub-Welsch algorithm ourselves, we define a function `gaussian_QR1` that utilizes two functions available in the NAG library: `dim1_gauss_recm` and `dim1_gauss_wrec`.

Given the $2m+1$ moments of the weight function $\nu$ in (3.22), `dim1_gauss_recm` generates the recursion coefficients of an underlying three-term recurrence formula (3.23). These coefficients are then used by `dim1_gauss_wrec` to compute the weights $(w_i)_{i=1}^m$ and nodes $(x_i)_{i=1}^m$ of the Gaussian quadrature rule using the Golub-Welsch method.

We then add the node $x_0 = 0$ and the weight $w_0$ that minimizes the quadratic polynomial from Proposition 6, instead of the $w_0$ provided in the setting of Theorem 2.8, as recommended in [6]. Following this, we proceed with the simulation of the rough Bergomi model (3.3).

### 3.3.3. The Simulation Scheme

Given the positive weights $(w_i)_{i=0}^N$ and nodes $(x_i)_{i=0}^N$, we aim to simulate the system:

$$\begin{cases} Y_t^i = \displaystyle\int_0^t e^{-x_i(t-s)} dB_s, \quad Y_0^i = 0, \\ \hat{V}_t = V_0 \exp\left( C_1 \displaystyle\sum_{i=0}^N w_i Y_t^i - C_2 \int_0^t \Big( \sum_{i=0}^N w_i e^{-x_i(t-s)} \Big)^2 ds \right), \quad V_0 > 0, \\ \hat{S}_t = S_0 \exp\left( \displaystyle\int_0^t \Big( \mu_s - \frac{1}{2}\hat{V}_s \Big) ds + \int_0^t \sqrt{\hat{V}_s} dB_s^S \right), \quad S_0 > 0, \end{cases} \tag{3.24}$$

where each $Y^i$ for $i \in \{0, \dots, N\}$ is an Ornstein-Uhlenbeck process (Definition 2.8) with speed of mean reversion $x_i$, $C_1 \coloneqq \eta\sqrt{2H}\Gamma(H+1/2)$, $C_2 \coloneqq \eta^2 H \Gamma(H+1/2)^2$, and the parameter $\eta$ is equal to $\eta = 2\nu C_H/\sqrt{2H}$ with $C_H$ defined as in (3.9).

Let $0 = t_0 < \cdots < t_k = T$ be a uniform partition with step size $\Delta t = T/k$. We discretize the system (3.24) in time and initialize the $N+3$ components $(\overline{Y}^i)_{i=0}^N, \overline{V}, \overline{S}$ by:

$$\overline{Y}_t^i = 0, \quad \overline{V}_t = V_0, \quad \overline{S}_t = S_0.$$

We then simulate for time $t_j$, where $j = 0, \dots, k-1$, the Gaussian vector underlying system (3.24):

$$\left( \int_{t_j}^{t_{j+1}} e^{-x_0(t_{j+1}-s)} dB_s, \dots, \int_{t_j}^{t_{j+1}} e^{-x_N(t_{j+1}-s)} dB_s, B_{t_{j+1}} - B_{t_j}, \right) \tag{3.25}$$

by computing a Cholesky decomposition of its covariance matrix $\Sigma$. Since the grid is chosen uniformly in time, the distribution of this vector is independent of $j$, requiring us to compute the Cholesky decomposition only once.

**Remark 3.7.** The algorithm employed by Bayer et al. [6] to simulate the rough Bergomi model (3.12), consists in discretizing the time interval $[0, T]$ into $k$ time steps and simulating the Gaussian vector

$$\left( \int_{t_j}^{t_{j+1}} (t_{j+1} - s)^{H-1/2} dB_s, B_{t_{j+1}} - B_{t_j} \right).$$

for $j = 0, \ldots, k-1$, by computing a Cholesky decomposition of its covariance matrix. Note that the rough kernel is therein not approximated.

We now provide a detailed explanation of how the simulation process works.

Let $Z$ be a vector of $N+2$ uncorrelated standard normal random variables, meaning $Z \sim N(0, \mathbf{I}_{N+2})$, where $\mathbf{I}_n$ is the identity matrix of order $n$. By applying an affine transformation $X = A + BZ$, the resulting vector $X$ has distribution $X \sim N(A, BB^\mathsf{T})$, since

$$\mathbb{E}[A + BZ] = A + B\mathbb{E}[Z] = A,$$

and

$$\mathbb{E}[(X - A)(X - A)^\mathsf{T}] = \mathbb{E}[BZZ^\mathsf{T}B^\mathsf{T}] = B\mathbb{E}[ZZ^\mathsf{T}]B^\mathsf{T} = B\mathbf{I}_{N+2}B^\mathsf{T} = BB^\mathsf{T}.$$

In our case, we identify $X$ with the Gaussian vector (3.25), resulting in $A = 0$ and $BB^\mathsf{T} = \Sigma$. Applying the Cholesky decomposition to $\Sigma$ provides a suitable matrix $B$, allowing us to simulate $X \sim N(0, \Sigma)$.

It is useful to note that the covariances within (3.25) are analytical and have to be computed with Itô's isometry and numerical integration. The covariance matrix $\Sigma$, of order $N+2$, is given by

$$\Sigma = \begin{bmatrix} \Sigma_{0,0} & \Sigma_{0,1} & \cdots & \Sigma_{0,N+1} \\ \Sigma_{1,0} & \Sigma_{1,1} & \cdots & \Sigma_{1,N+1} \\ \vdots & \vdots & \ddots & \vdots \\ \Sigma_{N+1,0} & \Sigma_{N+1,1} & \cdots & \Sigma_{N+1,N+1} \end{bmatrix},$$

where $\Sigma_{i,j} = \Sigma_{i,j} = \mathrm{Cov}(X_i, X_j) = \mathbb{E}[X_i X_j]$ for $i, j = 0, \ldots, N+1$, with $X_i$ denoting the $i$-th component of $X$ in (3.25). We compute this covariance matrix once for the time interval $[t_0, t_1]$, addressing three different cases.

(*i*) For $i = 0, \ldots, N$ and $j = 0, \ldots, N$,

$$\begin{aligned} \Sigma_{i,j} = \mathrm{Cov}(X_i, X_j) &= \mathbb{E}[\int_{t_0}^{t_1} e^{-x_i(t_1-s)} dB_s \int_{t_0}^{t_1} e^{-x_j(t_1-s)} dB_s] \\ &= \int_{t_0}^{t_1} e^{-x_i(t_1-s)} e^{-x_j(t_1-s)} ds = \frac{1}{x_i + x_j} \Big(1 - e^{-(x_i+x_j)(t_1-t_0)}\Big). \end{aligned}$$

(*ii*) For $i = 0, \ldots, N$ and $j = N+1$,

$$\begin{aligned} \Sigma_{i,N+1} = \mathrm{Cov}(X_i, X_{N+1}) &= \mathbb{E}[\int_{t_0}^{t_1} e^{-x_i(t_1-s)} dB_s \int_{t_0}^{t_1} dB_s] \\ &= \int_{t_0}^{t_1} e^{-x_i(t_1-s)} ds = \frac{1}{x_i} \Big(1 - e^{-x_i(t_1-t_0)}\Big). \end{aligned}$$

*(iii)* If $i = N+1$ and $j = N+1$,

$$\Sigma_{N+1,N+1} = \mathrm{Cov}(X_{N+1}, X_{N+1}) = \mathbb{E}[\int_{t_0}^{t_1} dB_s \int_{t_0}^{t_1} dB_s] = t_1 - t_0.$$

The Cholesky decomposition of a real positive-definite matrix $\Sigma$ is a decomposition of the form $\Sigma = BB^\mathsf{T}$, where $B$ is a real lower triangular matrix with positive diagonal entries. However, our $\Sigma$ is semi-definite rather than positive-definite, rendering the Cholesky decomposition non-unique.

To address this, we implement the function `chol_mod`, which applies the modified Cholesky algorithm of Cheng and Higham [17] to perturb the input covariance matrix slightly, rendering it positive-definite. Instead of implementing the algorithm from scratch, we utilize two functions from the NAG library: `real_modified_cholesky` and `real_mod_chol_perturbed_a`.

Given the symmetric semi-definite matrix $\Sigma$, the function `real_modified_cholesky` finds the Cheng-Higham modified Cholesky factorization

$$P^\mathsf{T}(\Sigma + E)P = LDL^\mathsf{T}.$$

Here, $L$ is a unit lower triangular matrix, specifically the lower triangular part of $\Sigma$, $P$ is a permutation matrix, $D$ is a symmetric block diagonal matrix with minimum eigenvalue $\delta$, and $E$ is a perturbation matrix of small norm chosen so that such a factorization can be found. The output is then passed to `real_mod_chol_perturbed_a` to compute the symmetric positive definite matrix $\Sigma + E$. We then apply the Cholesky decomposition to the perturbed matrix and simulate the Gaussian vector (3.25) as described above.

Then, we set for $j = 0, \ldots, k-1$,

$$\begin{cases} \overline{Y}^i_{t_{j+1}} = e^{-x_i \Delta t} \overline{Y}^i_{t_j} + \int_{t_j}^{t_{j+1}} e^{-x_i(t_{j+1}-s)} dB_s, \quad Y^i_0 = 0, \\ \overline{V}_{t_{j+1}} = V_0 \exp\left( C_1 \sum_{i=0}^{N} w_i \overline{Y}^i_{t_{j+1}} - C_2 \int_0^t \Big( \sum_{i=0}^{N} w_i e^{-x_i(t-s)} \Big)^2 ds \right), \\ \overline{S}_{t_{j+1}} = \overline{S}_{t_j} \exp\left( \Big(\mu_{t_j} - \frac{1}{2}\overline{V}_{t_j}\Big)\Delta t + \sqrt{\overline{V}_j}\Big(\rho(\tilde{B}_{t_{j+1}} - \tilde{B}_{t_j}) + \sqrt{1-\rho^2}(B_{t_{j+1}} - B_{t_j})\Big) \right), \end{cases}$$

with $V_0 > 0$ and $S_0 > 0$. We note that the deterministic integral in the exponent of $\overline{V}$ can be computed in closed form and that $B$ and $\tilde{B}$ are independent Brownian motions.

### 3.3.4. Validation and Visualization of the Simulation Scheme

To validate the correctness of our implementation, we compute the $L^2$ error of the Riemann-Liouville process $B^H$, as given in Definition (2.4), and its approximation $\hat{B}^H$:

$$B^H_t = \frac{1}{\Gamma(H+1/2)} \int_0^t (t-s)^{H-1/2} dB_s, \tag{3.26}$$
$$\hat{B}^H = \sum_{i=1}^{N} w_i \int_0^t e^{-x_i(t-s)} dB_s.$$

We compare our results with those presented in [6, Table 1].

We identify $B^H$ as the solution of a one-dimensional stochastic Volterra equation (2.29) with $v_0 = 0, b = 0, \sigma = 1$, and fractional kernel $K(\tau) = \tau^{H-1/2}/\Gamma(H+1/2)$. As previously mentioned, the domain of the measure $\mu$ is truncated from $(0, \infty)$ to $[\xi_0, \xi_n]$. Hence, the Riemann-Liouville process referred to as $B^H$ in (3.26) is the truncated process $B^{H,tr}$, consistent with the statements of Theorem 3.25 and Proposition 6.

We set $T = 1$, and choose $w_0$ to minimize the quadratic polynomial from Proposition 6. We vary $N$ within the set $\{1, 2, 4, 8, 16, 32, 64, 128\}$. The parameters $H, n, m$, the values of $\xi_0$ and $\xi_n$ are determined based on the following corollary:

**Corollary 1.** *[6, Corollary 3.1] For $H = 0.1$, consider the approximation with $N + 1$ nodes, including $x_0 = 0$, as suggested in the setting of Theorem 2.8. Let $N = nm$, where $m$ is the level of the Gaussian quadrature rule and $n$ is the number of intervals. Choose*

$$m \approx 0.1306\sqrt{N}, \quad n = N/m,$$

*and*

$$\xi_0 = 4.3679N^{0.1135}e^{-0.2322\sqrt{N}}, \quad \xi_n = 0.1421N^{-1.5889}e^{3.2511\sqrt{N}}.$$

*Then, we have*

$$\mathbb{E}\Big[|B_T^{H,tr} - \hat{B}_T^H|^2\Big] \leq 33.6483N^{0.3178}e^{-0.6502\sqrt{N}}.$$

**Remark 3.8.** Corollary 1 directly follows from Proposition 5, where the constant $C$ is set equal to 1, and [6, Theorem 2.15] with this specific choice of $C$. Theorem 2.15 in [6] is a variation of Theorem 2.8, which requires a Gaussian rule of type $(H, N, \alpha, \beta, a, b)$, with $a = b = 1$. In Theorem 2.15, the values of $a$ and $b$ are chosen as the solution an optimization problem. The proof of this result is technical and can be found in [6, Appendix C].

From Proposition 5 we recall that the $L^2$ error between $B^{H,tr}$ and $\hat{B}^H$ is controlled by the $L^2$ error of their respective kernels. We compute the $L^2$ error among $K^{tr}$ and $\hat{K}$ using Proposition 6 and compare our values with those in [6, Table 1]. Note that in practice the level $m$ of the Gaussian quadrature is rounded up to the next integer, and $n$ is chosen such that $nm$ is as close as possible to $N$. The comparison is detailed in Table (3.1), confirming the accuracy of our implementation.

Next, in Table 3.2, we compare the $L^2$ error among $K^{tr}$ and $\hat{K}$ (associated to the Riemann-Liouville process $B^{H,tr}$ and its approximation $\hat{B}^H$) under the settings of Theorem 2.8 and Corollary 1; see Remark 3.8. This comparison is of interest since we adopt the setting of Theorem 2.8 for the simulations of the model 3.24. For $N = 52$, we achieve an error tolerance of 0.1; thus, we will use this number of nodes for the simulations displayed in Figure 3.3.

| | | | Bayer and Breneis | | This work |
|---|---|---|---|---|---|
| $N$ | $m$ | $n$ | Error | Bound | Error |
| 1 | 1 | 1 | 0.996490 | 4.190757 | 0.99649093 |
| 2 | 1 | 2 | 0.975001 | 4.089250 | 0.97500068 |
| 4 | 1 | 4 | 0.899764 | 3.773728 | 0.89935636 |
| 8 | 1 | 8 | 0.757286 | 3.218395 | 0.75729121 |
| 16 | 1 | 16 | 0.571030 | 2.455046 | 0.57102870 |
| 32 | 1 | 32 | 0.372303 | 1.599424 | 0.37230131 |
| 64 | 2 | 32 | 0.195570 | 0.833625 | 0.19557919 |
| 128 | 2 | 64 | 0.075222 | 0.316916 | 0.07521947 |

**Table 3.1:** $L^2$-approximation errors for different values of $N, m$ and $n$ of the Riemann-Liouville process given in (3.26). We set $T = 1$ and $H = 0.1$, with $\xi_0$ and $\xi_1$ as suggested in Corollary 1. The "Error" columns show the actual errors for these values, computed using Proposition 6. The "Bound" column presents the error bounds from Corollary 1.

| $N$ | $m$ | $n$ | Error T2.8 | Error C1 |
|---|---|---|---|---|
| 1 | 1 | 1 | 0.77070377 | 0.99649093 |
| 2 | 1 | 2 | 0.68196321 | 0.97500068 |
| 4 | 1 | 4 | 0.56639290 | 0.89935636 |
| 8 | 1 | 8 | 0.43226404 | 0.75729121 |
| 16 | 1 | 16 | 0.29506523 | 0.57102870 |
| 32 | 1 | 32 | 0.17211350 | 0.37230131 |
| 64 | 2 | 32 | 0.08015850 | 0.19557919 |
| 128 | 2 | 64 | 0.02729483 | 0.07521947 |

**Table 3.2:** $L^2$-approximation errors for different values of $N, m$ and $n$ of the Riemann-Liouville process given in (3.26). We set $T = 1$ and $H = 0.1$. The "Error T2.8" and "Error C.1" columns show the actual errors for these values, computed under the settings of Theorem 2.8 and Corollary 1.

We illustrate the rough behavior of the model 3.24 in Figure 3.3. This figure displays sample paths of the rough Bergomi volatility process along with the corresponding stock price process for $H \in \{0.10, 0.25, 0.45\}$ and $N = 52$. The remaining model parameters are set as follows:

$$T = 1,\ \Delta t = 0.001,\ \eta = 1.9,\ \mu = 0,\ \rho = -0.7,\ S_0 = 1,\ V_0 = 0.235^2.$$

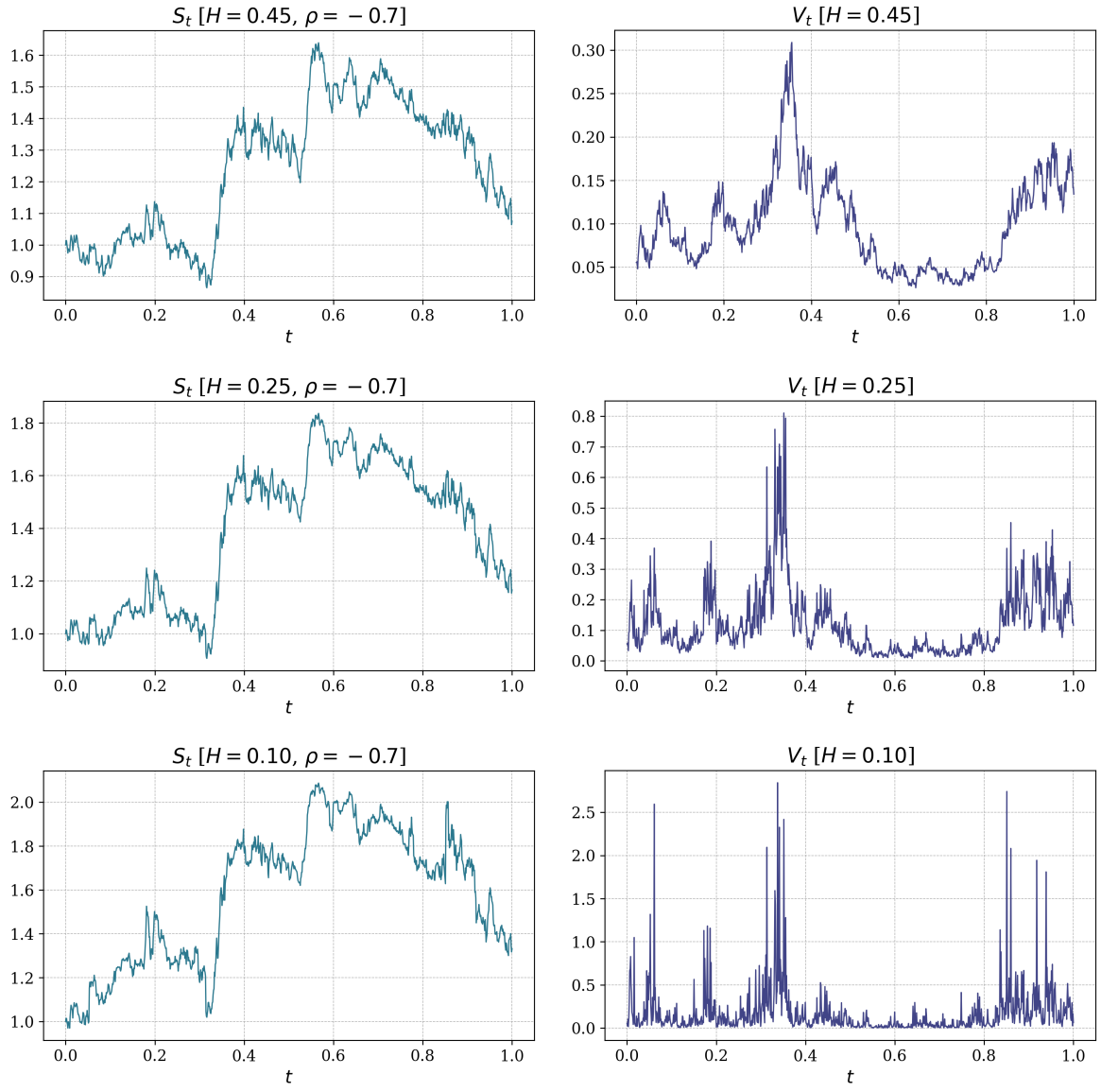



**Figure 3.3:** Sample paths of the rough Bergomi stock price and volatility for $H \in \{0.10, 0.25, 0.45\}$. The roughness of the paths increases as $H$ approaches its minimal possible value of 0.

# 4

# Solution to the Portfolio Optimization Problem

In this final chapter, we solve the portfolio optimization problem in a financial market comprising a bond and a risky asset featuring rough Bergomi volatility, as given in Definition 3.1. Specifically, we find the solution to the optimization problem as the limit of the approximated problem where the multi-factor rough Bergomi model given in 3.3 is considered. This approximation of the rough Bergomi model restores the Markovian property that the original model lacks, enabling us to use PDE techniques from classic dynamic programming, i.e. the Hamilton-Jacobi-Bellman equation, to solve the associated portfolio problem.

In Section 4.1 we briefly revisit our model choices, which are discussed in detail in Section 1.2, and outline the financial market setting. The market comprises one risk-free asset and one risky asset, whose coefficients are influenced by a stochastic factor $\mathbf{Y} = (Y^1, \dots, Y^N)$ driven by a Brownian motion that is correlated with the one driving the stock price.

In Section 4.2, we solve the approximated optimization problem, aiming to maximize the expected utility of terminal wealth with respect to a power utility function. We derive the HJB equation to address the optimization problem. When the Brownian motion driving the stock price and the Brownian motion driving the stochastic factor are correlated ($\rho \neq 0$) we find an implicit solution, which we further simplify using the martingale distortion transformation [63]. For $\rho = 0$, an explicit solution is derived (Theorems 4.1 and 4.2). The solution to the original rough path problem is then established in Theorem 4.3.

In Section 4.3, we visualize the the optimal terminal wealth process for different choices of the Hurst index. The process is simulated according to the approximation scheme described in Section 3.3. Gaussian quadrature is used to determine the nodes and weights for the fractional kernel approximation, and the simulations are performed using a Cholesky decomposition.

# 4.1. The Financial Market

## 4.1.1. Introduction

For a detailed exposition of our model choices and assumptions, we refer the reader to Section 1.2. Here, we provide a brief overview.

The market comprises one risk-free asset and one risky asset whose coefficients are influenced by a stochastic factor $\mathbf{Y} = (Y^1, \dots, Y^N)$. The stochastic factor $\mathbf{Y}$ is driven by a Brownian motion $B$, which is correlated with the Brownian motion $B^S$ driving the stock price. Each $Y^i$ is a diffusion process, specifically an Ornstein-Uhlenbeck process, which mean reverts at speed $x_i$. Due to the influence of the latent factor $\mathbf{Y}$ on the stock process's coefficients, the market is incomplete, i.e., it is impossible to fully hedge risk with just the bond and the available stock.

Regarding the information available, the agent observes only the stock price process. However, given that the stock price process contains sufficient information to infer the evolution of $B^S$ and $\mathbf{Y}$ through its quadratic variation, see Remark 4.1, it is reasonable to assume that the agent knows his wealth (driven by $B^S$) and the latent factor $\mathbf{Y}$ at any time $t$ [5, Remark 2.2].

Lastly, preferences are modeled using a power utility function $U(x) = x^\gamma/\gamma$ with $\gamma \in (0, 1)$, enabling the value function to be expressed as a power of the solution to a linear parabolic equation, as seen in Proposition 1. The power exponent, or distortion power, depends on the risk aversion coefficient and the correlation between the Brownian motions driving the stock price and the stochastic factor.

## 4.1.2. The Financial Market

Consider a financial market under $(\Omega, \mathcal{F}, \mathbb{F} = (\mathcal{F}_t)_{0\le t\le T}, \mathbb{P})$, with $T < \infty$, consisting of one risk-less asset $S^0$ and one risky asset $\hat{S}$ respectively evolving according to

$$\begin{aligned} dS_t^0 &= r_t S_t^0 dt, \quad S_0^0 > 0, \\ d\hat{S}_t &= \mu_t \hat{S}_t dt + \sqrt{\hat{V}_t}\hat{S}_t dB_t^S, \quad S_0 > 0, \end{aligned} \tag{4.1}$$

where $r_t \ge 0$ is a deterministic bounded risk-free rate, $B^S : [0, T] \to \mathbb{R}$ is an $\mathbb{F}$-Brownian motion, $\mu : [0, T] \times \mathbb{R}_+ \times \Omega \to \mathbb{R}$ is a predictable bounded process and is assumed to be a continuously differentiable function of $t$ and $\hat{V}$, we write $\mu_t = \mu(t, \hat{V})$. The approximated rough Bergomi volatility $\hat{V}$ is as given in Definition 3.3:

$$\hat{V}_t = V_0 \exp\left(C_1 \sum_{i=1}^N w_i \int_0^t e^{-x_i(t-s)} dB_s^V - C_2 t^{2H}\right), \quad V_0 > 0.$$

Here, $B^V : [0, T] \to \mathbb{R}$ is an $\mathbb{F}$-Brownian motion correlated to $B^S$ with correlation coefficient $\rho \in (-1, 1)$, $H$ is the Hurst index belonging to $(0, 1/2)$, $(w_i)_{i=1}^N$ are positive weights and $(x_i)_{i=1}^N$ are the corresponding nodes. The constants $C_1$ and $C_2$ are given

by $C_1 = \eta\sqrt{2H}\Gamma(H+1/2)$ and $C_2 = \eta^2 H\Gamma(H+1/2)^2$, where the parameter $\eta$ reads $\eta = 2\nu C_H/\sqrt{2H}$ with $C_H$ defined as in (3.9).

To underscore the dependence of $\hat{V}$ on the stochastic factor $\boldsymbol{Y} = (Y^1, \dots, Y^N)$, i.e. $\hat{V} = \hat{V}(t, \boldsymbol{Y})$, we align with the notation used in Section 1.2 and we rewrite $\hat{V}$ as follows:

$$\hat{V}_t = V_0 \exp\left(C_1 \sum_{i=1}^{N} w_i Y^i - C_2 t^{2H}\right), \quad V_0 > 0. \tag{4.2}$$

Each component $Y^i$ of the stochastic factor $\boldsymbol{Y}$ is an Ornstein-Uhlenbeck process (cf. Equation (2.8)) of the form

$$Y_t^i = \int_0^t e^{-x_i(t-s)} dB_s^V, \quad \text{with} \quad Y_0^i = 0, \tag{4.3}$$

and is the strong unique solution to the SDE

$$dY_t^i = -x_i Y_t^i dt + dB_t^V, \quad Y_0^i = 0. \tag{4.4}$$

**Remark 4.1.** It is reasonable to assume that at time $t$ the agent has knowledge of the stock price and, consequently, of its quadratic variation

$$\langle \hat{S}, \hat{S} \rangle_t = \int_0^t \hat{S}_u^2 \hat{V}_u du.$$

Thus, at time $t$, the agent is aware of the evolution of $B^S$ and $\boldsymbol{Y}$ up to that point [5, Remark 2.2]. We assume that the filtration $\mathbb{F}$ is generated by $B^S$ and $B^V$, i.e. $\mathcal{F}_t = \sigma(B_s^S, B_s^V, s \le t)$.

The investor rebalances his portfolio dynamically by choosing at any time $t$, for $t \in [0, T]$, a proportion $\alpha_t$ of his wealth to invest in the stock account $\hat{S}$. The remaining proportion $1 - \alpha_t$ is invested in the bond $S^0$. Note that $\alpha_t$ is not restricted to the interval $[0, 1]$. If $\alpha_t < 0$, the stock is sold short, and if $\alpha_t > 1$, money is borrowed from the bank at the interest rate $r_t$.

The investor faces the portfolio constraint that at any time $t$, $\alpha_t$ is valued in $A$, a closed convex subset of $\mathbb{R}$. We denote by $a$ the value of the control at a certain time, and with $\alpha$ the mapping $\alpha_t = \alpha(t, X_t)$. Let $\mathcal{A}$ be the set of $\mathbb{F}$-progressively measurable processes $\alpha$ valued in $A$ and satisfying the condition

$$\int_0^T \alpha_u^2 \hat{V}_u du < \infty \quad a.s., \tag{4.5}$$

Defined as such, $\mathcal{A}$ represents a class of admissible control laws. See Remark 4.2 for details.

Given an admissible portfolio strategy $\alpha \in \mathcal{A}$, we denote by $\hat{W} = (\hat{W}_t)_{t\in[0,T]}$ the corresponding wealth process starting from an initial capital $\hat{W}_t = w > 0$ at time $t$. For $t \le s \le T$ it evolves according to:

$$\begin{aligned} d\hat{W}_s &= \hat{W}_s \alpha_s \frac{d\hat{S}_s}{\hat{S}_s} + \hat{W}_s(1-\alpha_s)\frac{dS_s^0}{S_s^0} \\ &= \hat{W}_s[r_s + \alpha_s(\mu_s - r_s)]ds + \hat{W}_s \alpha_s \sqrt{\hat{V}_s} dB_s^S. \end{aligned} \tag{4.6}$$

The integrability condition (4.5) combined with the existence and uniqueness of a strong solution to the SDE (4.1) ensures the existence and uniqueness of the controlled wealth process. The stochastic differential equation (4.6) can be then solved explicitly and the solution for $t \in [0, T]$ is given by

$$\hat{W}_t = w \exp\left( \int_0^t \left[ r_s + \alpha_s(\mu_s - r_s) - \frac{1}{2}\hat{V}_s \alpha_s^2 \right] ds + \int_0^t \sqrt{\hat{V}_s}\alpha_s dB_s^S \right), \quad w > 0. \tag{4.7}$$

**Remark 4.2.** According to Definition 1.4, $\mathcal{A}$ is a class of admissible controls if: $(i)$ each element $\alpha \in \mathcal{A}$ is a progressively measurable process, valued in $A \subset \mathbb{R}$, and $(ii)$ for any given initial point the SDE of the state process has a unique solution. While the first condition is straightforward, the second warrants further explanation. The state process $X = (\hat{W}, \boldsymbol{Y})$ consists of the controlled diffusion $\hat{W}$ with dynamics (4.6), for which we have just verified existence and uniqueness. The remaining $N$ components, $Y^i$ for $i = 1, \ldots, N$, are Ornstein-Uhlenbeck processes, each of which is the strong unique solution of SDE (4.4). Consequently, given an initial point $(w, \boldsymbol{y})$, where $w \in \mathbb{R}_+$ and $\boldsymbol{y} \in \mathbb{R}^N$, the state process $X$ admits a unique strong solution, as each of its components is well-defined.

The investor's objective is to maximize his expected utility of terminal wealth:

$$J(t, w, \boldsymbol{y}) = \mathbb{E}_{t,w,\boldsymbol{y}}[U(\hat{W}_T, \boldsymbol{Y}_T)],$$

where by $\mathbb{E}_{t,w,\boldsymbol{y}}$ we denote the conditional expectation, emphasizing that at initial time $t$, we are given $\hat{W}_t = w > 0$ and $\boldsymbol{Y}_t = \boldsymbol{y} \in \mathbb{R}^N$. The utility function $U$ we consider is a power utility of the form

$$U(x) = \frac{x^\gamma}{\gamma}, \quad x \geq 0, \quad 0 < \gamma < 1.$$

The value function of the investor for $t \in [0, T]$ is then given by

$$v(t, w, \boldsymbol{y}) = \sup_{\alpha \in \mathcal{A}} \mathbb{E}_{t,w,\boldsymbol{y}}[U(\hat{W}_T, \boldsymbol{Y}_T)]. \tag{4.8}$$

## 4.2. Solution to the Portfolio Optimisation Problem

### 4.2.1. Introduction

Solving the portfolio optimization problem (4.8) involves finding, for each $(t, w, \boldsymbol{y})$ in $[0, T] \times \mathbb{R}_+ \times \mathbb{R}^N$, the value $v(t, w, \boldsymbol{y})$ and the admissible control $\alpha^* \in \mathcal{A}$ such that

$$v(t, w, \boldsymbol{y}) = \sup_{\alpha \in \mathcal{A}} \mathbb{E}_{t,w,\boldsymbol{y}}[U(\hat{W}_T, \boldsymbol{Y}_T)] = \mathbb{E}_{t,w,\boldsymbol{y}}[U(\hat{W}_T^*, \boldsymbol{Y}_T)],$$

where $\hat{W}^*$ denotes the wealth process (4.7) when the optimal control $\alpha^*$ is employed. As discussed in Section 1.1, finding a solution to the Hamilton-Jacobi-Bellmann (HJB) equation is equivalent to solving the optimization problem.

To solve the finite-dimensional optimization problem, we derive and solve the associated HJB equation. The second-order nonlinearities of the PDE are then eliminated using the distortion power $\delta$, according to Proposition 1. An implicit solution is found when the Brownian motion $B^S$ driving the wealth process and the Brownian motion $B^V$ driving the stochastic factor are correlated ($\rho \neq 0$). When $\rho = 0$, an explicit solution is derived.

We establish in Theorem 4.1 that the explicit solution is sufficiently smooth using the main result from [34], presented in Theorem 1.5. As a byproduct, we deduce that this solution can be represented as the Laplace transform of the integrated volatility process. Subsequently, in Theorem 4.2, we verify that it indeed corresponds to the value function.

By taking the limit of the finite-dimensional problem (4.8), we obtain the solution to the original optimization problem featuring rough Bergomi volatility:

$$v(t, w, \boldsymbol{y}) = \sup_{\alpha \in \mathcal{A}} \mathbb{E}_{t,w,\boldsymbol{y}}[U(W_T, \boldsymbol{Y}_T)], \tag{4.9}$$

where $W$ denotes the wealth process when considering the rBergomi volatility $V$ as given in Definition 3.1 instead of its approximation $\hat{V}$ in (4.2). This result is the main finding presented in Theorem 4.3.

### 4.2.2. Solution to the Finite Dimensional Optimisation Problem

We now solve the optimisation problem (4.8) using the HJB equation. The state process $X = (\hat{W}, \boldsymbol{Y})$ has as its first component the controlled diffusion $\hat{W}$ with dynamics (4.6), while the remaining $N$ components, $Y^i$ for $i = 1, \ldots, N$, are the diffusion processes evolving according to (4.4). Each $Y^i$ is a Ornstein-Uhlenbeck process that mean-reverts with speed $x_i$. For clarity, we rewrite the dynamics of $X$ in matrix form

$$d\begin{bmatrix} \hat{W}_s \\ \boldsymbol{Y}_s \end{bmatrix} = \begin{bmatrix} \hat{W}_s[r_s + \alpha_s(\mu_s - r_s)] \\ -\boldsymbol{x}\boldsymbol{Y}_s \end{bmatrix} ds + \begin{bmatrix} \hat{W}_s \alpha_s \sqrt{\hat{V}_s} & 0 \\ \mathbf{1}\rho & \mathbf{1}\sqrt{1-\rho^2} \end{bmatrix} \begin{bmatrix} dB_s^1 \\ dB_s^2 \end{bmatrix}, \tag{4.10}$$

where with $\boldsymbol{x}$ we denote the vector of nodes $\boldsymbol{x} = (x_1, \ldots, x_N)$ and with $\mathbf{1}$ the vector consisting of $N$ ones. We identify the Brownian motion driving the wealth process $\hat{W}$ with $B^S \equiv B^1$ and the Brownian motion driving the stochastic factor $\boldsymbol{Y}$ with $B^V \equiv \rho B^1 + \sqrt{1-\rho^2} B^2$; where $B^1, B^2$ are independent (cf. Section 1.2).

The HJB equation together with the terminal condition of the stochastic control problem for a sufficiently regular function $G \in C^{1,2}([0,T) \times \mathbb{R}_+ \times \mathbb{R}^N) \cap C^0([0,T] \times \mathbb{R}_+ \times \mathbb{R}^N))$ are given by:

$$\begin{cases} -\dfrac{\partial G}{\partial t}(t, w, \boldsymbol{y}) - \sup\limits_{a \in A}[\mathcal{L}_t^a G(t, w, \boldsymbol{y})] = 0 \quad \forall (t, w, \boldsymbol{y}) \in [0,T) \times \mathbb{R}_+ \times \mathbb{R}^N, \\ G(T, w, \boldsymbol{y}) = \dfrac{1}{\gamma} w^\gamma, \quad w \in \mathbb{R}_+. \end{cases} \tag{4.11}$$

Here, $\mathcal{L}_t^a$ is the second-order differential operator associated with the state process $X = (\hat{W}, \boldsymbol{Y})$ for the control $\alpha$:

$$(\mathcal{L}_t^a G)(x) = b(t, x, a) D_x G + \frac{1}{2} \mathrm{tr}\Big(\sigma(t, x, a) \sigma^\intercal(t, x, a) D_x^2 G\Big),$$

where $x = (w, \boldsymbol{y})$, $b$ and $\sigma$ are the coefficients of the state process $X$ in (4.10); $b$ is a vector with $N+1$ components and $\sigma$ is a $(N+1) \times 2$ matrix.

To solve the finite-dimensional optimization problem we carry out the following procedure.

1. We fix an arbitrary point $(t, w, \boldsymbol{y}) \in [0,T] \times \mathbb{R}_+ \times \mathbb{R}^N$ and solve, for this fixed choice of $(t, w, \boldsymbol{y})$, the static optimization problem $\sup_{a \in A}[\mathcal{L}^a G(t, w, \boldsymbol{y})] = 0$, where $a$ is the only variable.
2. The optimal choice of $a$, denoted by $a^*$, will depend on our choice of $t, w$ and $\boldsymbol{y}$, but it will also depend on the function $G$ and its partial derivatives. To highlight these dependencies we write $a^*$ as $\alpha^* = \alpha^*(t, w, \boldsymbol{y}; G)$.
3. We insert the separable Ansatz for $G$, i.e. $G(t, w, \boldsymbol{y}) = \frac{1}{\gamma} w^\gamma \Phi(t, \boldsymbol{y})$, in the HJB equation and in $\alpha^*(t, w, \boldsymbol{y}; G)$.
4. We simplify the non-linearities of the HJB equation utilizing the distortion power.
5. In the verification step (Theorem 4.2), we identify $G$ as the optimal value function $v$ (4.8), and $\alpha^*$ as the optimal control law.

Fixed an arbitrary point $(t, w, \boldsymbol{y}) \in [0,T] \times \mathbb{R}_+ \times \mathbb{R}^N$, the static optimization problem reads

$$\begin{aligned} 0 = \sup_{a \in A} \Big\{ & w\Big[(r_t + a(\mu_t - r_t)\Big] G_w + \sum_{i=1}^N (-x_i) y_i G_{y_i} \\ & + \frac{1}{2} w^2 a^2 \hat{V}_t G_{ww} + \sum_{i=1}^N w \rho a \sqrt{\hat{V}_t} G_{y_i w} + \frac{1}{2} \sum_{i=1}^N \sum_{j=1}^N G_{y_i y_j} \Big\}. \end{aligned} \tag{4.12}$$

Maximizing this expression in $a$ leads to our candidate for the optimal control law $\alpha^* = \alpha^*(t, w, \boldsymbol{y}; G)$:

$$\alpha^*(t, w, \boldsymbol{y}; G) = -\frac{1}{w} \frac{G_w}{G_{ww}} \frac{\mu_t - r_t}{\hat{V}_t} - \frac{\rho}{\sqrt{\hat{V}_t}} \sum_{i=1}^N \frac{1}{w} \frac{G_{y_i w}}{G_{ww}}. \tag{4.13}$$

Now, we substitute the separation Ansatz $G(t, w, \boldsymbol{y}) = \frac{1}{\gamma} w^\gamma \Phi(t, \boldsymbol{y})$ with $\Phi(T, \boldsymbol{y}) = 1$ in the HJB equation and divide by $\frac{1}{\gamma} w^\gamma$,

$$\begin{aligned} 0 = \Phi_t + \sup_{a \in A} \Big\{ & \Big[r_t + a(\mu_t - r_t)\Big] \gamma \Phi + \sum_{i=1}^N (-x_i) y_i \Phi_{y_i} \\ & + \frac{1}{2} a^2 \hat{V}_t (\gamma - 1) \gamma \Phi + \sum_{i=1}^N \rho a \sqrt{\hat{V}_t} \gamma \Phi_{y_i} + \frac{1}{2} \sum_{i=1}^N \sum_{j=1}^N \Phi_{y_i y_j} \Big\}. \end{aligned} \tag{4.14}$$

Substituting the Ansatz also in $\alpha^*$ (4.13) results in the following optimal trading strategy

$$\alpha_t^* = \frac{\mu_t - r_t}{1 - \gamma} \frac{1}{\hat{V}_t} + \frac{\rho}{(1-\gamma)} \frac{1}{\sqrt{\hat{V}_t}} \frac{\sum_{i=1}^N \Phi_{y_i}}{\Phi}. \tag{4.15}$$

Inserting the maximum point (4.15) into (4.14) yields

$$\begin{aligned}0 = \Phi_t &+ \Big(r_t + \Big[\frac{\mu_t - r_t}{1-\gamma}\frac{1}{\hat{V}_t} + \frac{\rho}{(1-\gamma)}\frac{1}{\sqrt{\hat{V}_t}}\frac{\sum_{i=1}^N \Phi_{y_i}}{\Phi}\Big](\mu_t - r_t)\Big)\gamma\Phi \\ &+ \sum_{i=1}^N (-x_i) y_i \Phi_{y_i} + \frac{1}{2}\sum_{i=1}^N\sum_{j=1}^N \Phi_{y_i y_j} \\ &- \frac{1}{2}\Big[\frac{\mu_t - r_t}{1-\gamma}\frac{1}{\hat{V}_t} + \frac{\rho}{(1-\gamma)}\frac{1}{\sqrt{\hat{V}_t}}\frac{\sum_{i=1}^N \Phi_{y_i}}{\Phi}\Big]^2 \hat{V}_t(1-\gamma)\gamma\Phi \\ &+ \sum_{i=1}^N \rho\Big[\frac{\mu_t - r_t}{1-\gamma}\frac{1}{\hat{V}_t} + \frac{\rho}{(1-\gamma)}\frac{1}{\sqrt{\hat{V}_t}}\frac{\sum_{i=1}^N \Phi_{y_i}}{\Phi}\Big]\sqrt{\hat{V}_t}\gamma\Phi_{y_i}.\end{aligned} \tag{4.16}$$

We expand the products and reorder the terms of the PDE (4.16)

$$\begin{aligned}0 = \Phi_t &+ \sum_{i=1}^N \left[(-x_i)y_i + (\rho-1)\frac{\gamma}{1-\gamma}\frac{\mu_t - r_t}{\sqrt{\hat{V}_t}}\right]\Phi_{y_i} \\ &+ \frac{1}{2}\sum_{i=1}^N\sum_{j=1}^N \Phi_{y_i y_j} + \left(\frac{r_t\gamma}{\delta} + \frac{1}{2}\frac{\gamma}{\delta(1-\gamma)}\frac{(\mu_t - r_t)^2}{\hat{V}_t}\right)\Phi \\ &+ \frac{1}{2}\rho^2\frac{\gamma}{1-\gamma}\frac{1}{\Phi}\Big(\sum_{i=1}^N \Phi_{y_i}\Big)^2.\end{aligned} \tag{4.17}$$

Unfortunately, PDE (4.17) is rather involved and has to be solved numerically. However, we can further reduce its complexity by eliminating the second-order terms through the application of the distortion power introduced in [63], see Proposition 1. We define a function $\varphi$ such that

$$\varphi(t,\boldsymbol{y})^\delta = \Phi(t,\boldsymbol{y}), \quad \text{with } \delta = \frac{1-\gamma}{1-\gamma+\rho^2\gamma},$$

and satisfying the terminal condition $\varphi(T,\boldsymbol{y}) = 1$. We substitute $\varphi^\delta$ and the derivatives

$$\begin{aligned}&\Phi_t = \delta\varphi^{\delta-1}\varphi_t, \quad \sum_{i=1}^N \Phi_{y_i} = \delta\varphi^{\delta-1}\sum_{i=1}^N \varphi_{y_i}, \\ &\sum_{i=1}^N\sum_{j=1}^N \Phi_{y_i,y_j} = \delta\varphi^{\delta-1}\sum_{i=1}^N\sum_{j=1}^N \varphi_{y_i y_j} + \delta(\delta-1)\varphi^{\delta-2}\Big(\sum_{i=1}^N \varphi_{y_i}\Big)^2,\end{aligned}$$

in (4.17). Dividing by $\delta\varphi^{\delta-1}$, we obtain the linear parabolic equation

$$\begin{aligned}0 = \varphi_t &+ \sum_{i=1}^N \left[(-x_i)y_i + (\rho-1)\frac{\gamma}{1-\gamma}\frac{\mu_t - r_t}{\sqrt{\hat{V}_t}}\right]\varphi_{y_i} \\ &+ \frac{1}{2}\sum_{i=1}^N\sum_{j=1}^N \varphi_{y_i y_j} + \left(\frac{r_t\gamma}{\delta} + \frac{1}{2}\frac{\gamma}{\delta(1-\gamma)}\frac{(\mu_t - r_t)^2}{\hat{V}_t}\right)\varphi.\end{aligned} \tag{4.18}$$

We remark again that the terms $\left(\sum_{i=1}^{N}\varphi_{y_i}\right)^2$ vanishes because the power distortion $\delta$ is chosen to nullify the coefficient of this term

$$\frac{1}{2}\varphi^{-1}\Big[(\delta-1)+\rho^2\frac{\gamma}{1-\gamma}\delta\Big]\Big(\sum_{i=1}^{N}\varphi_{y_i}\Big)^2.$$

The following theorem establishes that the PDE (4.18), when $\rho=0 \implies \delta=1$, admits a unique solution $\varphi$, which can be expressed as a Laplace transform of an integrated volatility via the Feynman-Kac theorem. In the proof, we use the main result of [34], presented in Theorem 1.5, which gives sufficient conditions to ensure that the function $\varphi \in C^{1,2}$ is the unique solution to the Cauchy problem

$$\begin{cases} -\dfrac{\partial\varphi}{\partial t}-\boldsymbol{\mathcal{L}}_t\varphi+c\varphi=0, \quad \text{on } [0,T)\times\mathbb{R}^N, \\ \qquad\qquad \varphi(T,\boldsymbol{y})=1 \quad \text{on } \mathbb{R}^N, \end{cases} \tag{4.19}$$

where the PDE is exactly (4.18). The Feynman-Kac theorem, Theorem 1.4, then provides the representation formula

$$\varphi(t,\boldsymbol{y})=\mathbb{E}\left[e^{-\int_t^T c(u,X_u^{t,x})du}\right].$$

**Theorem 4.1.** *A solution of the HJB equation* (4.17) *exists on a certain time interval* $[0,T]$ *and, for* $t\in[0,T], w\in\mathbb{R}_+, \boldsymbol{y}\in\mathbb{R}^N$*, it is given by*

$$G(t,w,\boldsymbol{y})=\frac{1}{\gamma}w^\gamma\varphi(t,\boldsymbol{y})^\delta, \quad \textit{with } \delta=\frac{1-\gamma}{1-\gamma+\gamma\rho^2}.$$

*Specifically,* $\varphi\in C^{1,2}([0,T)\times\mathbb{R}^N)\cap C^0([0,T]\times\mathbb{R}^N)$ *satisfies the Cauchy problem* (4.19) *and it can be written as*

$$\varphi(t,\boldsymbol{y})=\mathbb{E}_{t,w,\boldsymbol{y}}\Big[\exp\Big\{\int_t^T\Big(\frac{r_t\gamma}{\delta}+\frac{1}{2}\frac{\gamma}{\delta(1-\gamma)}\frac{(\mu_s-r_s)^2}{\tilde{V}_s}\Big)ds\Big\}\Big], \tag{4.20}$$

*where* $\tilde{V}_t=V_0\exp\left(C_1\sum_{i=1}^{N}w_i\tilde{Y}^i-C_2t^{2H}\right)$, $V_0>0$. *In the uncorrelated case* $\rho=0$*, we have* $\hat{V}=\tilde{V}$.

*Proof.* The existence of a classical solution for PDE (4.18) and its representation (4.20) follow from a direct application of Theorem 1.5. We let $[0,T]$ be the time horizon and $D=\mathbb{R}^N$ the domain. The diffusion process we consider is $X=\tilde{\boldsymbol{Y}}$, where each component $\tilde{Y}^i$ for $i=1,\dots,N$ is an Ornstein-Uhlenbeck process of the form (4.3) and evolving according to (4.4). Each $\tilde{Y}^i$ is driven by the same Brownian motion $B$, but with different speed of mean reversion $x_i$:

$$d\tilde{\boldsymbol{Y}}_s=-\boldsymbol{x}\tilde{\boldsymbol{Y}}_s ds+\mathbf{1}dB_s.$$

We identify the coefficient function $c:[0,T]\times\mathbb{R}^N\to\mathbb{R}^N$, in front of $\varphi$ in (4.18), as

$$c(t,\tilde{\boldsymbol{y}})=-\Big(\frac{r_t\gamma}{\delta}+\frac{1}{2}\frac{\gamma}{\delta(1-\gamma)}\frac{(\mu_t-r_t)^2}{\tilde{V}_t}\Big).$$

We now show that conditions $(A1), (A2)$ and $(A3)$ of Theorem 1.5 are satisfied.

In our case, $b : \mathbb{R}^N \to \mathbb{R}^N$ is given by $b(\tilde{\boldsymbol{Y}}) = -\boldsymbol{x}\tilde{\boldsymbol{Y}}$, while $\sigma$ is the $N$-dimensional vector of ones. Therefore, $b$ does not depend on $t$ and is $C^1$ in $\tilde{\boldsymbol{y}}$ on $\mathbb{R}^N$, and $\sigma$ does not depend on $t$ nor $\tilde{\boldsymbol{y}}$. Since continuously differentiable implies locally Lipschitz-continuous, assumption $(A1)$ is satisfied.

Each components of $X$ is the unique strong solution to SDE (4.4). Therefore, $X$ never explodes before $T$, ensuring that $(A2)$ is also satisfied.

Lastly, in order to verify $(A3)$ via $(A3')$, we take as domains $D_n = (-n, n)^N$ with smoothed corners so that they satisfy $(A3')$. Because $b$ and $\sigma$ are uniformly Lipschitz, $(A3a')$ is obvious, so is $(A3b')$: since $a = \sigma\sigma^\mathsf{T} = N$, we can choose $\epsilon_n = N$ to verify that $a$ is uniformly elliptic on $\mathbb{R}^N$.

We note that the approximated volatility

$$\tilde{V}(t, \tilde{\boldsymbol{y}}) = V_0 \exp\Big(C_1 \sum_{i=0}^{N} w_i \tilde{y}_i - C_2 t^{2H}\Big),\ V_0 > 0,$$

is $C^1$ in $\tilde{\boldsymbol{y}}$, $r_t$ does not depend on $\tilde{\boldsymbol{y}}$, the process $\mu_t = \mu(t, \tilde{V})$ is assumed to be a continuously differentiable function of $\tilde{V}$, thus,

$$\frac{\partial \mu(t, \tilde{V})}{\partial \tilde{\boldsymbol{y}}} = \frac{\partial \mu(t, \tilde{V})}{\partial \tilde{V}} \frac{\partial \tilde{V}(t, \tilde{\boldsymbol{y}})}{\partial \tilde{\boldsymbol{y}}}$$

and $\mu_t$ is $C^1$ in $\tilde{\boldsymbol{y}}$. Moreover, the derivative of $c$ with respect to $\tilde{\boldsymbol{y}}$ is

$$\frac{\partial c}{\partial \tilde{\boldsymbol{y}}} = \frac{1}{2} \frac{\gamma}{\delta(1-\gamma)} \frac{1}{\tilde{V}^2(t, \tilde{\boldsymbol{y}})} \left[ 2(\mu(t, \tilde{V}) - r) \frac{\partial \mu(t, \tilde{V})}{\partial \tilde{\boldsymbol{y}}} \tilde{V}(t, \tilde{\boldsymbol{y}}) - (\mu(t, \tilde{V}) - r)^2 \frac{\partial \tilde{V}(t, \tilde{\boldsymbol{y}})}{\partial \tilde{\boldsymbol{y}}} \right].$$

Consequently, the function $c$ has continuous derivatives on the compact $\overline{D}_n$, thus, is Lipschitz-continuous in $\tilde{\boldsymbol{y}}$ (i.e. Hölder-continuous with $\alpha = 1$) with respect to $t$, ensuring that condition $(A3c')$ holds.

Lastly, $(A3d')$ is apparent since $f \equiv 0$, and the finiteness condition $(A3e')$ is guaranteed because $\varphi$ does not have singularities. □

**Remark 4.3.** In Theorem 4.1 we denoted with $\tilde{V}$ the process whose $N$ factors $\tilde{Y}^i$ are driven by a standard Brownian motion $B$. This process differs from the correlated Brownian motion $B^V = \rho B^1 + \sqrt{1-\rho^2} B^2$ that drives the stochastic factor $\boldsymbol{Y}$ of the approximated volatility $\hat{V}$ as given in (4.2). Only when $\rho = 0$, we obtain $\hat{V} = \tilde{V}$.

**Remark 4.4.** Since $\varphi \in C^{1,2}([0,T) \times \mathbb{R}^N) \cap C^0([0,T] \times \mathbb{R}^N)$ and $G(t, w, \boldsymbol{y}) = \frac{1}{\gamma} w^\gamma \varphi(t, \boldsymbol{y})^\delta$ with $G(T, w, \boldsymbol{y}) = \frac{1}{\gamma} w^\gamma$, Theorem 4.1 guarantees that $G \in C^{1,2}([0,T) \times \mathbb{R}_+ \times \mathbb{R}^N) \cap C^0([0,T] \times \mathbb{R}_+ \times \mathbb{R}^N)$.

Now we need to verify that a smooth solution to the HJB equation of the form $G(t, w, \boldsymbol{y}) = \frac{1}{\gamma} w^\gamma \varphi(t, \boldsymbol{y})^\delta$ indeed yields the value function.

**Theorem 4.2** (Verification)**.** *Suppose that* $G(t,w,\boldsymbol{y}) = \frac{1}{\gamma} w^\gamma \varphi(t,\boldsymbol{y})^\delta$ *with* $\varphi$ *as in* (4.20) *and satisfying the boundary condition* $\varphi(T,\boldsymbol{y}) = 1$*. For* $t \in [0,T]$*,* $\rho = 0$*, the optimal investment strategy* $\alpha^*$ *is given by*

$$\alpha_t^* = \frac{\mu_t - r_t}{1-\gamma} \frac{1}{\hat{V}_t},$$

*and the value function can be written as*

$$v(t,w,\boldsymbol{y}) = \frac{1}{\gamma} w^\gamma \varphi(t,\boldsymbol{y}) = \frac{1}{\gamma} w^\gamma \mathbb{E}_{t,w,\boldsymbol{y}}\Big[\exp\Big\{\int_t^T \Big(r_s\gamma + \frac{1}{2}\frac{\gamma}{1-\gamma}\frac{(\mu_s - r_s)^2}{\hat{V}_s}\Big)ds\Big\}\Big]. \quad (4.21)$$

*Proof.* To obtain $v = G$, we show that for an arbitrary admissible investment strategy $\alpha$ we have that

$$\mathbb{E}_{t,w,\boldsymbol{y}}\Big[\frac{(\hat{W}_T)^\gamma}{\gamma}\Big] \le G(t,w,\boldsymbol{y}), \quad (4.22)$$

and that for the optimal policy $\alpha^*$ the equality in (4.22) is reached. Since Theorem (4.1) guarantees that $G \in C^{1,2}$, see Remark 4.4, we can apply Itô's formula. For an admissible investment strategy $\alpha$, we obtain

$$\begin{aligned} G\Big(T,\hat{W}_T,\boldsymbol{Y}_T\Big) &= G(t,w,\boldsymbol{y}) + \int_t^T G_w \hat{W}_s \alpha_s \sqrt{\hat{V}_s} dB_s^1 + \int_t^T \sum_{i=1}^N G_{y_i} dB_s^2 \\ &\quad + \int_t^T \Big\{ G_t + G_w \hat{W}_s \big[r_s + \alpha_s(\mu_s - r_s)\big] + \sum_{i=1}^N G_{y_i}(-x_i) Y_s^i \\ &\quad + \frac{1}{2} G_{ww} (\hat{W}_s)^2 \alpha_s^2 \hat{V}_s + \sum_{i=1}^N w\rho a \sqrt{\hat{V}_t} G_{y_i w} + \frac{1}{2}\sum_{i=1}^N \sum_{j=1}^N G_{y_i y_j} \Big\} ds \\ &\le G(t,w,\boldsymbol{y}) + \int_t^T G_w \hat{W}_s \alpha_s \sqrt{\hat{V}_s} dB_s^1 + \int_t^T \sum_{i=1}^N G_{y_i} dB_s^2, \end{aligned}$$

where we used (4.12) in the last line. We also remark that $B^1$ and $B^2$ are independent Brownian motions, and that the stochastic integrals are well-defined since $\alpha$ is as admissible portofolio and $G \in C^{1,2}([0,T) \times \mathbb{R}_+ \times \mathbb{R}^N)$. The right-hand side is a local martingale in $T$

$$L_T = G(t,w,\boldsymbol{y}) + \int_t^T G_w \hat{W}_s \alpha_s \sqrt{\hat{V}_s} dB_s^1 + \int_t^T \sum_{i=1}^N G_{y_i} dB_s^2.$$

For $\gamma \in (0,1)$ we have $G > 0$ and $L_T$ is in particular a supermartingale. Indeed, by taking the conditional expectation on both sides of the above inequality and using the boundary condition $\varphi(T,\boldsymbol{y}) = 1$, which yields $G(T,\hat{W}_T,\boldsymbol{Y}_T) = \hat{W}_T^\gamma/\gamma$, we obtain

$$\mathbb{E}_{t,w,\boldsymbol{y}}\Big[G\Big(T,\hat{W}_T,\boldsymbol{Y}_T\Big)\Big] = \mathbb{E}_{t,w,\boldsymbol{y}}\Big[\frac{(\hat{W}_T)^\gamma}{\gamma}\Big] \le G(t,w,\boldsymbol{y}),$$

since

$$\mathbb{E}_{t,w,\boldsymbol{y}}\Big[\int_t^T G_w \hat{W}_s \alpha_s \sqrt{\hat{V}_s} dB_s^1\Big] = \mathbb{E}_{t,w,\boldsymbol{y}}\Big[\int_t^T \sum_{i=1}^N G_{y_i} dB_s^2\Big] = 0.$$

Hence, relation (4.22) holds. When we consider the optimal strategy $\alpha^*$, we get

$$G\Big(T, \hat{W}_T^*, \boldsymbol{Y}_T\Big) = G(t, w, \boldsymbol{y}) + \int_t^T G_w \hat{W}_s^* \alpha_s^* \sqrt{\hat{V}_s} dB_s^1 + \int_t^T \sum_{i=1}^N G_{y_i} dB_s^2,$$

where we denoted with $\hat{W}^*$ the wealth process when the optimal policy is used. Taking the conditional expectation on both sides and using the assumption yields

$$\mathbb{E}_{t,w,\boldsymbol{y}}\Big[\frac{(\hat{W}_T^*)^\gamma}{\gamma}\Big] = G(t, w, \boldsymbol{y}).$$

□

Now that we have solved the finite-dimensional optimization problem, we proceed by taking its limit to find the solution of the original problem under the rough Bergomi volatility $V$ as given in Definition 3.1.

### 4.2.3. Solution to the Original Optimisation Problem

**Theorem 4.3.** *Suppose $\rho = 0$, and a classical solution $\Phi$ of the partial differential equation (4.17) with boundary condition $\Phi(T, \boldsymbol{y}) = 1$ exists for $t \in [0, T]$ and $\boldsymbol{y} \in \mathbb{R}^N$. The portfolio optimisation problem (4.9) has an optimal investment strategy given by*

$$\alpha_t^* = \frac{\mu_t - r_t}{1-\gamma} \frac{1}{V_t}, \tag{4.23}$$

*and the value function can be written as*

$$v(0, w, \boldsymbol{y}) = \frac{w^\gamma}{\gamma} \mathbb{E}_{0,\boldsymbol{y}}\Big[\exp\Big\{\int_0^T \Big(r_s\gamma + \frac{1}{2}\frac{\gamma}{1-\gamma}\frac{(\mu_s - r_s)^2}{V_s}\Big)ds\Big\}\Big]. \tag{4.24}$$

*Proof.* The first result (4.23) directly follows from the fact that $\alpha$ is an admissible strategy (see Definition 1.4), thus, is an $\mathbb{F}$-adapted process and does not depend on $N$.

Recall that the approximated final wealth process starting at $W_0 = w > 0$ is explicitly given by,

$$\hat{W}_T = w \exp\left(\int_0^T \Big[r_s + \alpha_s(\mu_s - r_s) - \frac{1}{2}\hat{V}_s\alpha_s^2\Big]ds + \int_0^T \sqrt{\hat{V}_s}\alpha_s dB_s^1\right). \tag{4.25}$$

Since $\alpha^*$ is optimal, from Theorem 4.2 we obtain

$$\mathbb{E}_{0,w,\boldsymbol{y}}\Big[\frac{(\hat{W}_T)^\gamma}{\gamma}\Big] \leq \mathbb{E}_{0,w,\boldsymbol{y}}\Big[\frac{(\hat{W}_T^*)^\gamma}{\gamma}\Big], \tag{4.26}$$

where we denoted with $\hat{W}^*$ the wealth process when the optimal policy is used. Using the Feynman-Kac result in Theorem 4.1 yields for the right-hand side of (4.26)

$$\mathbb{E}_{0,w,\boldsymbol{y}}\Big[\frac{(\hat{W}_T^*)^\gamma}{\gamma}\Big] = \frac{w^\gamma}{\gamma} \mathbb{E}_{0,\boldsymbol{y}}\Big[\exp\Big\{\int_0^T \Big(r_s\gamma + \frac{1}{2}\frac{\gamma}{1-\gamma}\frac{(\mu_s - r_s)^2}{\hat{V}_s}\Big)ds\Big\}\Big]. \tag{4.27}$$

By monotone convergence of $\hat{\mathcal{V}}$ to $\mathcal{V}$ from Theorem 3.1, and recalling that the processes $\hat{\mathcal{V}}$ and $\mathcal{V}$ are distributed as $\hat{V}$ and $V$, respectively, we obtain

$$\begin{aligned}
&\lim_{N\to\infty} \mathbb{E}_{0,\boldsymbol{y}}\Big[\exp\Big\{\int_0^T \Big(r_s\gamma + \frac{1}{2}\frac{\gamma}{1-\gamma}\frac{(\mu_s - r_s)^2}{\hat{V}_s}\Big)ds\Big\}\Big] \\
&\quad = \lim_{N\to\infty} \mathbb{E}_{0,\boldsymbol{y}}\Big[\exp\Big\{\int_0^T \Big(r_s\gamma + \frac{1}{2}\frac{\gamma}{1-\gamma}\frac{(\mu_s - r_s)^2}{\hat{\mathcal{V}}_s}\Big)ds\Big\}\Big] \\
&\quad = \mathbb{E}_{0,\boldsymbol{y}}\Big[\exp\Big\{\int_0^T r_s\gamma + \frac{1}{2}\frac{\gamma}{1-\gamma}\frac{(\mu_s - r_s)^2}{\mathcal{V}_s}\Big)ds\Big\}\Big] \\
&\quad = \mathbb{E}_{0,\boldsymbol{y}}\Big[\exp\Big\{\int_0^T r_s\gamma + \frac{1}{2}\frac{\gamma}{1-\gamma}\frac{(\mu_s - r_s)^2}{V_s}\Big)ds\Big\}\Big].
\end{aligned}$$

Let us now consider the left hand side of (4.26). Since we are taking the conditional expectation of $\hat{W}_T$, we use $\hat{\mathcal{V}}$ and $\mathcal{V}$ instead of $\hat{V}$ and $V$ to establish the limit of $\hat{W}_T$ as $N \to \infty$. Since by monotone convergence

$$\lim_{N\to\infty} \mathbb{E}\left[\int_0^T (\alpha_s)^2\Big(\sqrt{\hat{\mathcal{V}}_s} - \sqrt{\mathcal{V}_s}\Big)^2 ds\right] = 0,$$

applying the Itô isometry yields

$$\int_0^T \alpha_s\sqrt{\hat{\mathcal{V}}_s}dB_s^1 \xrightarrow{L^2} \int_0^T \alpha_s\sqrt{\mathcal{V}_s}dB_s^1,$$

which, in turn, implies $L^1$ convergence. Thus, the exponent of $\hat{W}_T$ in (4.25) becomes

$$\begin{aligned}
&\lim_{N\to\infty}\left(\int_0^T \big[r_s + \alpha_s(\mu_s - r_s) - \frac{1}{2}\hat{\mathcal{V}}_s\alpha_s^2\big]ds + \int_0^T \alpha_s\sqrt{\hat{\mathcal{V}}_s}dB_s^1\right) \\
&= \int_0^T \big[r_s + \alpha_s(\mu_s - r_s) - \frac{1}{2}\mathcal{V}_s\alpha_s^2\big]ds + \int_0^T \alpha_s\sqrt{\mathcal{V}_s}dB_s^1,
\end{aligned}$$

where in the first term the exchange between limit and integral is again justified by monotone convergence. Since $L^1$ convergence implies convergence in probability, which in turn implies convergence in distribution [30, Theorem 3.1], the Skorokhod representation theorem (Theorem A.3) and the continuity of the exponential function guarantee that, on a suitable probability space

$$\begin{aligned}
\lim_{N\to\infty} \hat{\mathcal{W}}_t &= \lim_{N\to\infty} w\exp\left(\int_0^T \big[r_s + \alpha_s(\mu_s - r_s) - \frac{1}{2}\hat{\mathcal{V}}_s\alpha_s^2\big]ds + \alpha_s\int_0^T \sqrt{\hat{\mathcal{V}}_s}dB_s^1\right) \\
&= w\exp\left(\int_0^T \big[r_s + \alpha_s(\mu_s - r_s) - \frac{1}{2}\mathcal{V}_s\alpha_s^2\big]ds + \alpha_s\int_0^T \sqrt{\mathcal{V}_s}dB_s^1\right) = \mathcal{W}_t \quad a.s.,
\end{aligned}$$

where the processes $\hat{\mathcal{W}}$ and $\mathcal{W}$ share the same distributions as $\hat{W}$ and $W$, with the latter being the wealth process while considering the rough Bergomi volatility $V$.

Noting that $U(\hat{W}_T) = (\hat{W}_T)^\gamma/\gamma$ is a sequence of measurable non-negative functions, from Fatou's Lemma, we obtain for the left hand side of (4.26) that

$$\begin{aligned}
\liminf_{N\to\infty} \mathbb{E}_{0,w,\boldsymbol{y}}\Big[\frac{(\hat{W}_T)^\gamma}{\gamma}\Big] &\geq \mathbb{E}_{0,w,\boldsymbol{y}}\Big[\liminf_{N\to\infty}\frac{(\hat{W}_T)^\gamma}{\gamma}\Big] = \mathbb{E}_{0,w,\boldsymbol{y}}\Big[\liminf_{N\to\infty}\frac{(\hat{\mathcal{W}}_T)^\gamma}{\gamma}\Big] \\
&= \mathbb{E}_{0,w,\boldsymbol{y}}\Big[\frac{(\mathcal{W}_T)^\gamma}{\gamma}\Big] = \mathbb{E}_{0,w,\boldsymbol{y}}\Big[\frac{(W_T)^\gamma}{\gamma}\Big],
\end{aligned}$$

where we have used that, since the limit of $\hat{\mathcal{W}}$ exists almost surely, $\liminf \hat{\mathcal{W}}_t = \limsup \hat{\mathcal{W}}_t = \mathcal{W}_t$ as $N \to \infty$. Finally, for all admissible strategies $\alpha$

$$\mathbb{E}_{0,w,\boldsymbol{y}}\Big[\frac{(W_T)^\gamma}{\gamma}\Big] \leq \frac{w^\gamma}{\gamma}\mathbb{E}_{0,\boldsymbol{y}}\Big[\exp\Big\{\int_0^T \Big(r_s\gamma + \frac{1}{2}\frac{\gamma}{1-\gamma}\frac{(\mu_s - r_s)^2}{V_s}\Big)ds\Big\}\Big],$$

where the equality is met if the optimal strategy $\alpha^*$ is adopted. □

The optimal terminal wealth $W^*$ on $[0,T]$ is then derived by substituting the optimal portfolio strategy $\alpha^*$ (4.23) into

$$W_T = w\exp\left(\int_0^T \Big[r_s + \alpha_s(\mu_s - r_s) - \frac{1}{2}V_s\alpha_s^2\Big]ds + \int_0^T \sqrt{V_s}\alpha_s dB_s^1\right).$$

This substitution yields

$$W_T^* = w\exp\left(\int_0^T \Big[r_t + (\mu_t - r_t)^2\frac{1-2\gamma}{2(1-\gamma)^2}\frac{1}{V_t}\Big]dt + \int_0^T \frac{\mu_t - r_t}{1-\gamma}\frac{1}{\sqrt{V}_t}dB_t^1\right), \tag{4.28}$$

where we recall that we are considering the case $\rho = 0$, meaning the Brownian motion $B^1 \equiv B^S$ driving the wealth process and the Brownian motion $B^V$ driving the volatility process are independent.

## 4.3. Simulation Results

In the case $\rho = 0$, we simulate the optimal terminal wealth $W^*$ (4.28) using the parameters:

$$w = 1000,\ r = 0.02,\ \mu = r + \lambda V_t,\ \lambda = 0.5,\ \gamma = -2,$$

and resulting in the following expression:

$$W_T^* = w\exp\left(\int_0^T \Big[r + \frac{\lambda^2(1-2\gamma)}{2(1-\gamma)^2}V_t\Big]dt + \int_0^T \frac{\lambda}{1-\gamma}\sqrt{V}_t dB_t^1\right).$$

**Remark 4.5.** The choices $r = 0.02$ and $\mu = r + \lambda V_t$ are admissible: $r$ is a constant, and $\mu$ is a continuously differentiable function of $t$ and $V$. Therefore, the assumptions we have made regarding these parameters in the financial market are satisfied.

The process is simulated based on the approximation scheme outlined in Section 3.3. Gaussian quadrature is employed to determine the nodes and weights for the fractional kernel approximation, and the simulations are conducted using the Cholesky decomposition of the Gaussian vector underlying the approximated system.

We compare the optimal terminal wealth $W^*$ with the wealth obtained from investing solely in the safe asset $S^0$, denoted as $\overline{W}$, which is given $\overline{W}_T = w\exp(rT)$ with $w > 0$.

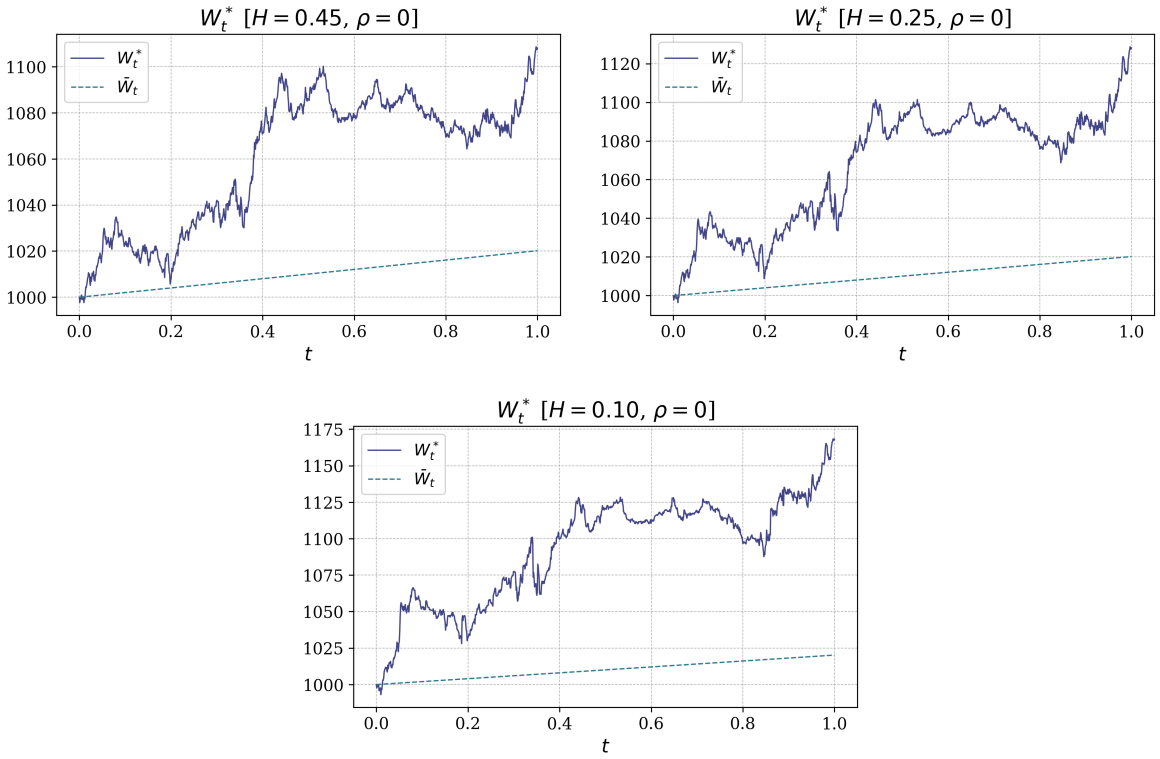



**Figure 4.1:** Rough optimal terminal wealth $W^*$ for $H \in \{0.10, 0.25, 0.45\}$ compared the wealth $\overline{W}$ obtained from investing solely in the safe asset.

# Conclusions

This thesis addresses the challenge of solving the portfolio optimization problem within a financial market framework comprising one risk-free asset and one risky asset featuring rough Bergomi volatility. According to the definition of rough volatility processes, the rough Bergomi volatility is governed by the fractional kernel $K(\tau) = \tau^{H-1/2}$ with Hurst index $H \in (0, 1/2)$. This specification forces the rough Bergomi model to leave both semimartingale and Markovian frameworks, thereby complicating its theoretical analysis and practical applications.

By leveraging Markovian approximations for fractional processes, as discussed in [16], [33], and [32], we approximate the fractional component of the rough volatility process and solve the associated optimization problem using PDE techniques. This method ultimately enables us to find the solution to the original optimization problem involving rough Bergomi volatility as the limit of the approximated problem, consistent with the methodology outlined in [5]. The findings are further validated through a numerical study utilizing Gaussian quadrature and Cholesky decomposition, based on the work [6].

The approximation involves two key steps: first, we represent the fractional component of the rough Bergomi volatility as a linear functional of an infinite-dimensional OU process to obtain its Markovian representation. Second, we discretize this representation, leading to an approximation dependent on a finite-dimensional Markovian stochastic factor. In the resulting approximated rough Bergomi model, the coefficients of the risky asset are influenced by a stochastic factor $\mathbf{Y} = (Y^1, \ldots, Y^N)$, driven by a Brownian motion correlated with the one driving the stock price. Each component of this factor shares the same dynamics but mean reverts at different rates. Using the HJB equation, we derive an implicit solution when the stochastic factor and the stock price are driven by correlated Brownian motions, and provide an explicit solution for the uncorrelated case.

The numerical study focuses on simulating the rough Bergomi model. The simulation scheme begins by approximating the fractional kernel of the rough volatility using Gaussian quadrature to compute the positive weights and nodes. These weights and nodes are then employed to simulate the Markovian approximation of the rough Bergomi model through Cholesky decomposition of the underlying Gaussian vector. The accuracy of this approach is validated by comparing the approximation error with results from existing literature and by visualizing the rough Bergomi model across various parameter settings.

This thesis contributes to the field of portfolio optimization by addressing the rough Bergomi model, thus filling a gap in the literature where analyses have been limited to the rough Heston model. The Markovian approximation employed herein enables the application of the HJB equation in a conventional manner, with its effectiveness further supported by the convergence of the approximated model to the original rough Bergomi model.

The approach of employing a Markovian approximation extends beyond the rough Bergomi model and is not limited to portfolio optimization. Its versatility and practicality effectively address the challenges posed by the non-Markovian and non-semimartingale limitations of processes featuring a fractional kernel. This approximation's effectiveness has been validated within the broader framework of stochastic Volterra equations [2] and has been applied to simulations, option pricing, and hedging problems, as discussed in [8] and [1].

The Markovian approximation of fractional processes is fundamentally connected to approximating their fractional kernels. In this thesis, we use Gaussian quadrature to approximate the fractional kernel associated with the rough volatility process, as detailed in [6]. The computational efficiency of this method has been validated in the same work, where the authors compare the Gaussian quadrature rule with other choices of nodes and weights for approximating the fractional kernel, including [32] and [4], which groups with [16], [2], [5].

Future research could explore several promising avenues. One potential direction is a comparative analysis of the rough Heston and rough Bergomi models in portfolio optimization. This could involve employing a consistent kernel approximation method, such as Gaussian quadrature, and applying PDE techniques to solve the portfolio optimization problem. Further empirical validation with real market data could offer deeper insights and refine the methodologies presented in this thesis. Additionally, future work might focus on improving the convergence rates of Gaussian quadrature approximations, as detailed in [7], or investigating alternative numerical techniques [48].

In terms of numerical implementation, particular attention could be given to the hybrid multifactor scheme proposed by Rømer [59]. This scheme integrates the Hybrid method developed by Bennedsen et al. [9], which advances the simulation of (truncated) Brownian semistationary processes, with the multifactor approximation for completely monotone kernels discussed in this thesis. By combining these methodologies, the hybrid scheme aims to leverage the strengths of both approaches, thereby improving the overall effectiveness and accuracy of numerical implementations.

# References


[1] E. Abi Jaber, "Lifting the Heston model," *Quantitative Finance*, vol. 19, no. 12, pp. 1995–2013, 2019.

[2] E. Abi Jaber, "Stochastic invariance and stochastic volterra equations," Ph.D. dissertation, Université Paris sciences et lettres, 2018.

[3] E. Abi Jaber, M. Larsson, and S. Pulido, "Affine Volterra processes," *The Annals of Applied Probability*, vol. 29, no. 5, pp. 3155–3200, 2019.

[4] A. Alfonsi and A. Kebaier, "Approximation of stochastic Volterra equations with kernels of completely monotone type," *Mathematics of Computation*, vol. 93, no. 346, pp. 643–677, 2023.

[5] N. Bäuerle and S. Desmettre, "Portfolio Optimization in Fractional and Rough Heston Models," *SIAM Journal on Financial Mathematics*, vol. 11, no. 1, pp. 240–273, 2020.

[6] C. Bayer and S. Breneis, "Markovian approximations of stochastic Volterra equations with the fractional kernel," *Quantitative Finance*, vol. 23, no. 1, pp. 53–70, 2023.

[7] C. Bayer and S. Breneis, "Weak Markovian approximations of rough Heston," *arXiv preprint arXiv:2309.07023*, 2023.

[8] C. Bayer, P. Friz, and J. Gatheral, "Pricing under rough volatility," *Quantitative Finance*, vol. 16, no. 6, pp. 887–904, 2016.

[9] M. Bennedsen, A. Lunde, and M. Pakkanen, "Hybrid scheme for Brownian semistationary processes," *Finance and Stochastics*, vol. 21, no. 4, pp. 931–965, 2017.

[10] F. Biagini, Y. Hu, B. Øksendal, and T. Zhang, *Stochastic Calculus for Fractional Brownian Motion and Applications*. London: Springer, 2008.

[11] T. Bjork, *Arbitrage Theory in Continuous Time*, 3rd ed. United States: Oxford University Press, 2009.

[12] F. Black and M. Scholes, "The Pricing of Options and Corporate Liabilities," *Journal of Political Economy*, vol. 81, no. 3, pp. 637–654, 1973.

[13] H. Brass, "Monotonie bei den quadraturverfahren von Gauss und Newton-Cotes," *Numerische Mathematik*, vol. 30, no. 3, pp. 349–354, 1978.

[14] H. Brass and K. Petras, *Quadrature theory: The theory of numerical integration on a compact interval*, 1st ed. United States: American Mathematical Society, 2012, vol. 178.

[15] P. Carmona and L. Coutin, "Fractional Brownian motion and the Markov property," *Electronic Communications in Probability*, vol. 3, pp. 95–107, 1998.

[16] P. Carmona, L. Coutin, and G. Montseny, "Approximation of some Gaussian processes," *Statistical Inference for Stochastic Processes*, vol. 3, no. 1, pp. 161–171, 2000.

[17] S. Cheng and N. Higham, "A Modified Cholesky Algorithm Based on a Symmetric Indefinite Factorization," *SIAM Journal on Matrix Analysis and Applications*, vol. 19, no. 4, pp. 1097–1110, 1998.

[18] F. Comte and E. Renault, "Long memory in continuous-time stochastic volatility models," *Mathematical Finance*, vol. 8, no. 4, pp. 291–323, 1998.

[19] R. Dudley, *Real Analysis and Probability*, 2nd ed. Cambridge: Cambridge University Press, 2002.

[20] O. El Euch, M. Fukasawa, J. Gatheral, and M. Rosenbaum, "Short-Term At-the-Money Asymptotics under Stochastic Volatility Models," *SIAM Journal on Financial Mathematics*, vol. 10, no. 2, pp. 491–511, 2019.

[21] O. El Euch and M. Rosenbaum, "The characteristic function of rough Heston models," *Mathematical Finance*, vol. 29, no. 1, pp. 3–38, 2019.

[22] G. Fasshauer, *Meshfree Approximation Methods with Matlab*. WORLD SCIENTIFIC, 2007.

[23] W. H. Fleming and T. Zariphopoulou, "An optimal investment/consumption model with borrowing," *Mathematics of Operations Research*, vol. 16, no. 4, pp. 802–822, 1991.

[24] M. Fukasawa, "Asymptotic analysis for stochastic volatility: Martingale expansion," *Finance and Stochastics*, vol. 15, pp. 635–654, 2011.

[25] J. Gatheral, T. Jaisson, and M. Rosenbaum, "Volatility is rough," *Quantitative Finance*, vol. 18, no. 6, pp. 933–949, 2018.

[26] S. Ghosal and A. van der Vaart, "Gaussian Process Priors," in *Fundamentals of Nonparametric Bayesian Inference* (Cambridge Series in Statistical and Probabilistic Mathematics), Cambridge Series in Statistical and Probabilistic Mathematics. Cambridge University Press, 2017, pp. 310–360.

[27] G. Golub and J. Welsch, "Calculation of Gauss Quadrature Rules," *Mathematics of Computation*, vol. 23, no. 106, pp. 221–230, 1969.

[28] S. Grossman and Z. Zhou, "Optimal investment strategies for controlling drawdowns," *Mathematical Finance*, vol. 3, no. 3, pp. 241–276, 1993.

[29] H. Guennoun, A. Jacquier, P. R. Patrick, and F. Shi, "Asymptotic Behavior of the Fractional Heston Model," *SIAM Journal on Financial Mathematics*, vol. 9, no. 3, pp. 1017–1045, 2018.

[30] A. Gut, *Probability: A Graduate Course* (Springer Texts in Statistics), 2nd ed. Springer New York, 2013.

[31] B. Han and H. Wong, "Merton's portfolio problem under Volterra Heston model," *Finance Research Letters*, vol. 39, pp. 101–115,

[32] P. Harms, "Strong convergence rates for Markovian representations of fractional processes," *Discrete and Continuous Dynamical Systems - B*, vol. 26, no. 10, pp. 5567–5579, 2021.

[33] P. Harms and D. Stefanovits, "Affine representations of fractional processes with applications in mathematical finance," *Stochastic Processes and their Applications*, vol. 129, no. 4, pp. 1185–1228, 2019.

[34] D. Heath and M. Schweizer, "Martingales versus PDEs in finance: An equivalence result with examples," *Journal of Applied Probability*, vol. 37, no. 4, pp. 947–957, 2000.

[35] B. Horvath, A. Jacquier, and A. Muguruza, *Functional central limit theorems for rough volatility*, 2017. arXiv: 1711.03078.

[36] D. Huy, "A remark on non-markov property of a fractional Brownian motion," *Vietnam Journal of Mathematics*, vol. 31, 2003.

[37] T. Jaisson and M. Rosenbaum, "Rough fractional diffusions as scaling limits of nearly unstable heavy tailed Hawkes processes," *The Annals of Applied Probability*, vol. 26, no. 5, pp. 2860–2882, 2016.

[38] C. Jost, "Integral transformations of Volterra Gaussian processes," Ph.D. dissertation, University of Helsinki, 2007.

[39] I. Karatzas, J. P. Lehoczky, S. P. Sethi, and S. E. Shreve, "Explicit solution of a general consumption/investment problem," *Mathematics of Operations Research*, vol. 11, no. 2, pp. 261–294, 1986.

[40] I. Karatzas, J. P. Lehoczky, S. E. Shreve, and G.-L. Xu, "Martingale and Duality Methods for Utility Maximization in an Incomplete Market," *SIAM Journal on Control and Optimization*, vol. 29, no. 3, pp. 702–730, 1991.

[41] M. Keller-Ressel, M. Larsson, and S. Pulido, "Rough affine models," in *Rough Volatility*, 2023, ch. 5, pp. 103–113.

[42] H. Kuo, *Introduction to stochastic integration*, 1st ed. New York: Springer, 2006.

[43] P. Lévy, *Random Functions: General Theory with Special Reference to Laplacian Random Functions*. Berkeley: University of California Press, 1953.

[44] R. A. Maller, G. Müller, and A. Szimayer, "Ornstein-Uhlenbeck Processes and Extensions," in *Handbook of Financial Time Series*. Springer Berlin Heidelberg, 2009, pp. 421–437.

[45] B. Mandelbrot and J. V. Ness, "Fractional Brownian motions, fractional noises and applications," *SIAM Review*, vol. 10, no. 4, pp. 422–437, 1968.

[46] D. Marinucci and P. Robinson, "Alternative forms of fractional Brownian motion," *Journal of Statistical Planning and Inference*, vol. 80, no. 1, pp. 111–122, 1999.

[47] H. Markowitz, "Portfolio selection," *The Journal of Finance*, vol. 7, no. 1, pp. 77–91, 1952.

[48] J. Matas and J. Pospíšil, *On simulation of rough Volterra stochastic volatility models*, 2022. arXiv: 2108.01999 [q-fin.CP].

[49] R. Merton, "Lifetime portfolio selection under uncertainty: The continuous-time case," *The Review of Economics and Statistics*, vol. 51, no. 3, pp. 247–257, 1969.

[50] G. Molchan and Y. Golosov, "Gaussian stationary processes with asymptotic power spectrum," *Soviet Mathematics Doklady*, vol. 10, no. 1, pp. 134–137, 1969.

[51] J. V. Neumann and O. Morgenstern, *Theory of Games and Economic Behavior*, 1st, Ed. Princeton, NJ, USA: Princeton University Press, 1944.

[52] B. Øksendal, *Stochastic Differential Equations: An Introduction with Applications*, 6th ed. Berlin, Heidelberg: Springer, 2014.

[53] H. Pham, *Continuous-time Stochastic Control and Optimization with Financial Applications*, 1st ed. Berlin, Heidelberg: Springer, 2009.

[54] H. Pham, *Optimization Methods in Portfolio Management and Option Hedging*, Lecture notes, 3rd cycle, Hanoi (Vietnam), pp. 27. ⟨cel-00392212⟩, 2007.

[55] H. Pham and M. Quenez, "Optimal portfolio in partially observed stochastic volatility models," *The Annals of Applied Probability*, vol. 11, no. 1, pp. 210–238, 2001.

[56] J. Picard, "Representation formulae for the fractional Brownian motion," in *Séminaire de Probabilités XLIII*. Berlin, Heidelberg: Springer Berlin Heidelberg, 2011, pp. 3–70.

[57] W. Press, S. Teukolsky, W. Vetterling, and B. Flannery, *Numerical Recipes: The Art of Scientific Computing*, 3rd. Cambridge: Cambridge University Press, 2007, ch. 4, pp. 147–161.

[58] C. Rogers, "Arbitrage with fractional Brownian motion," *Mathematical Finance*, vol. 7, no. 1, pp. 95–105, 1997.

[59] S. E. Rømer, *Hybrid Multifactor Scheme for Stochastic Volterra Equations with Completely Monotone Kernels*, Available at SSRN: `https://ssrn.com/abstract=3706253`, 2022.

[60] W. Rudin, "Riemann-stieltjes integral," in *Principles of Mathematical Analysis*, International Edition: McGraw-Hill, 1976, ch. 6.

[61] T. Seppäläinen and E. Sorensen, "Busemann process and semi-infinite geodesics in Brownian last-passage percolation," *Annales de l'Institut Henri Poincaré, Probabilités et Statistiques*, vol. 59, no. 1, pp. 117–165, 2023.

[62] M. Tehranchi, "Explicit solutions of some utility maximization problems in incomplete markets," *Stochastic Processes and their Applications*, vol. 114, no. 1, pp. 109–125, 2004.

[63] T. Zariphopoulou, "A solution approach to valuation with unhedgeable risks," *Finance and Stochastics*, vol. 5, no. 1, pp. 61–82, 2001.

[64] Q. Zhu, G. Loeper, W. Chen, and N. Langrené, "Markovian approximation of the rough Bergomi model for Monte Carlo option pricing," *Mathematics*, vol. 9, no. 5, 2021.

# A

# Auxiliary Results

## The Stochastic Fubini Theorem

Let $\mu$ be a $\sigma$-finite measure on $(0,\infty)$, i.e. $(0,\infty)$ is a countable union of measurable sets each with finite measure. Fix $T \geq 0$ and denote by $\mathcal{P}_r$ the $\sigma$-algebra on $[0,T]\times\Omega$ generated by all progressively measurable processes (Definition 1.3).

**Theorem A.1.** *[33, Theorem 6.1] Let $G : (0,\infty)\times[0,T]\times\Omega \to \mathbb{R}$ be measurable with respect to the product $\sigma$-algebra $\mathcal{B}(0,\infty)\otimes\mathcal{P}_r$. Define the processes $\zeta_{1,2} : (0,\infty)\times[0,T]\times\Omega\to\mathbb{R}$ and $\eta : [0,T]\times\Omega\to\mathbb{R}$ by*

$$\zeta_1(x,t,\omega) = \int_0^t G(x,s,\omega)ds, \quad \zeta_2(x,t,\omega) = \left(\int_0^t G(x,s,\cdot)dB_s\right)(\omega),$$
$$\eta(t,\omega) = \int_0^\infty G(x,t,\omega)\mu(dx).$$

*(i) Assume $G$ satisfies for almost all $\omega\in\Omega$*

$$\int_0^\infty \int_0^T |G(x,s,\omega)|ds\mu(dx) < \infty. \tag{A.1}$$

*Then, for almost all $\omega\in\Omega$ and for all $t\in[0,T]$ we have $\zeta_1(\cdot,t,\omega)\in L^1(\mu)$ and*

$$\int_0^\infty \zeta_1(x,t,\omega)\mu(dx) = \int_0^t \eta(s,\omega)ds.$$

*(ii) Assume $G$ satisfies for almost all $\omega\in\Omega$*

$$\int_0^\infty \sqrt{\int_0^T |G(x,s,\omega)|^2 ds}\mu(dx) < \infty. \tag{A.2}$$

*Then, for almost all $\omega\in\Omega$ and for all $t\in[0,T]$ we have $\zeta_2(\cdot,t,\omega)\in L^1(\mu)$ and*

$$\int_0^\infty \zeta_2(x,t,\omega)\mu(dx) = \left(\int_0^t \eta(s,\cdot)dB_s\right)(\omega).$$

## The Bernstein-Widder Theorem

In real analysis, Bernstein's theorem asserts that any real-valued function on the half-line $[0, \infty)$ that is completely monotone can be expressed as a mixture of exponential functions. More precisely, the Bernstein-Widder theorem, or Hausdorff-Bernstein-Widder theorem, gives an integral characterization of completely monotone functions.

**Definition A.1.** *[22, Definition 2.5.1] A function $\Psi : [0, \infty) \to \mathbb{R}$ which is in $C[0, \infty) \cap C^{\infty}(0, \infty)$ and which satisfies*

$$(-1)^n \Psi^{(n)}(t) \geq 0, \quad t > 0, \ n = 0, 1, 2, \ldots,$$

*is called completely monotone on $[0, \infty)$.*

**Theorem A.2.** *[22, Theorem 2.5.2] A function $\Psi : [0, \infty) \to \mathbb{R}$ is completely monotone on $[0, \infty)$ if and only if it is the Laplace transform of a finite non-negative Borel measure $\mu$ on $[0, \infty)$, i.e., $\Psi$ is of the form*

$$\Psi(t) = \mathcal{L}\mu(t) = \int_0^{\infty} e^{-tx} d\mu(x).$$

## The Skorokhod Representation Theorem

The Skorokhod's representation theorem is a result that shows that weakly convergent random variables can be represented as the distribution/law of a pointwise convergent sequence of random variables defined on a common probability space. The theorem only requires the support of the law of the limiting random variable to be separable.

Since we are interested in real-valued random variables and $\mathbb{R}$ is a separable metric space, we consider the version of the theorem that applies to $\mathbb{R}$. Before presenting the theorem, we first review the definitions related to the convergence of random variables involved.

**Definition A.2.** *[30, Definition 1.1] $X_n$ converges almost surely (a.s.) to the random variable $X$ as $n \to \infty$ if and only if*

$$P\left(\{\omega : X_n(\omega) \to X(\omega) \ \text{as} \ n \to \infty\}\right) = 1.$$

*Notation: $X_n \xrightarrow{a.s.} X$ as $n \to \infty$.*

**Definition A.3.** *[30, Definition 1.4] Let $C(F_X) = \{x : F_X(x) \text{ is continuous at } x\}$ denote the continuity set of $F_X$, where $F_X$ is the distribution function of $X$. Then, $X_n$ converges in distribution to the random variable $X$ as $n \to \infty$ if and only if*

$$F_{X_n}(x) \to F_X(x) \quad \text{as } n \to \infty, \quad \text{for all } x \in C(F_X).$$

*Notation: $X_n \xrightarrow{d} X$ as $n \to \infty$.*

**Theorem A.3.** *[30, Theorem 13.1] Let* $\{X_n, n \geq 1\}$ *be random variables such that*

$$X_n \xrightarrow{d} X \text{ as } n \to \infty.$$

*Then there exist random variables* $\tilde{X}$ *and* $\{\tilde{X}_n, n \geq 1\}$ *defined on a Lebesgue probability space, such that*

$$\tilde{X}_n \stackrel{d}{=} X_n \text{ for } n \geq 1, \quad \tilde{X} \stackrel{d}{=} X, \text{ and } \quad \tilde{X}_n \xrightarrow{a.s.} \tilde{X} \text{ as } n \to \infty.$$

**Remark A.1.** In probabilistic language, the Lebesgue probability space corresponds to a $U(0,1)$-distributed random variable.

# B

# Gaussian Quadrature

Let $[a, b]$ be a finite interval with $0 \geq a \geq b$, let $\nu : [a, b] \to [0, \infty)$ be a continuous weight function, and let $f : [a, b] \to \mathbb{R}$ be a function that we wish to integrate. The aim of quadrature theory is to compute an approximate value for the definite integral

$$I_\nu[f] = \int_a^b f(x)\nu(x)dx,$$

by means of a quadrature rule, that is a map $Q$, determined by $(x_i)_{i=1}^m$ (the nodes) that lie in $[a, b]$, and numbers $(w_i)_{i=1}^m$ (the weights) which assigns to a function $f$ the value

$$Q[f] = \sum_{i=1}^{m} w_i f(x_i). \tag{B.1}$$

If the $x_i$ and $w_i$ are chosen suitably, then $Q[f]$ is an approximation of $I[f]$; this approximation is mathematically meaningful only if one has a good understanding of the error

$$R[f] \coloneqq I[f] - Q[f].$$

We have a total of $2m$ degrees of freedom when selecting our nodes and weights. The Gaussian quadrature rule of level $m$, $Q_m^G$, is the unique choice of nodes and weights that integrates all polynomials of degree at most $2m - 1$ exactly, i.e.

$$R_m^G\Big[P_{2m-1}\Big] = \{0\}.$$

We now outline how the Gaussian nodes and weights can be computed, drawing on the exposition presented in [57]. Firstly, one needs to find orthogonal polynomials $p_n$ of degree $n$. Orthogonality is defined with respect to the scalar product:

$$\langle f, g\rangle = \int_a^b f(x)g(x)\nu(x)dx.$$

These orthogonal polynomials can be efficiently obtained using the three-term recurrence relation:

$$\begin{aligned} p_{-1}(x) &\equiv 0, \\ p_0(x) &\equiv 1, \\ p_{n+1}(x) &= (x - a_n)\, p_n(x) - b_n p_{n-1}(x), \quad n = 0, 1, 2 \ldots \end{aligned} \tag{B.2}$$

where

$$a_n = \frac{\langle xp_n, p_n \rangle}{\langle p_n, p_n \rangle}, \quad n = 0, 1, \dots$$
$$b_n = \frac{\langle p_n, p_n \rangle}{\langle p_{n-1}, p_{n-1} \rangle}, \quad n = 1, 2, \dots$$

The polynomials defined by (B.2) are monic, i.e., the coefficient of their leading term $x^n$ for $p_n(x)$ is unity. Moreover, the polynomial $p_n(x)$ can be shown to have exactly $n$ distinct real roots in the interval $(a, b)$.

Finding all the roots of an orthogonal polynomial $p_n(x)$ is essential because the abscissas of the $m$-point Gaussian quadrature formula (B.1) with the weighting function $\nu(x)$ over the interval $(a, b)$ correspond exactly to the roots of the orthogonal polynomial $p_n(x)$ for the same interval and weighting function [14, Theorem 6.1.2].

Once the abscissas $x_1, \dots, x_m$ , are determined, the next step is to compute the weights $w_i$ for $i = 1, \dots, m$. One method to evaluate the weights is given by the formula:

$$w_i = \frac{\langle p_{m-1}, p_{m-1} \rangle}{p_{m-1}(x_i) p'_m(x_i)},$$

where $p'_m(x_i)$ denotes the derivative of the orthogonal polynomial at the root $x_i$.

The computation of Gaussian quadrature rules thus involves two distinct phases:

(*i*) the generation of the orthogonal polynomials $p_0, \dots, p_m$ , i.e., the computation of the coefficients $a_n, b_n$ in (B.2),

(*ii*) the determination of the zeros of $p_m(x)$, and the computation of the associated weights.

For the case of the "classical" orthogonal polynomials (e.g. Legendre, Chebyshev, Hermite), the recurrence coefficients $a_n$ and $b_n$ are explicitly known, allowing us to omit phase (*i*). Instead, when dealing with a "non classical" weight function $\nu(x)$, the coefficients $a_n$ and $b_n$ are unknown and constructing the corresponding set of orthogonal polynomials becomes more complex. An efficient algorithm for computing Gaussian nodes and weights, applicable even for non-standard weight functions, is the one developed by Golub and Welsch in [27].

Moving forward, we highlight two fundamental properties of the Gaussian quadrature rule: the strict positivity of the weights and the monotonicity property.

**Theorem B.1.** *The weights $(w_i)_{i=1}^m$ of a Gaussian quadrature rule of level $m$, $Q_m^G$, are strictly positive.*

*Proof.* Consider the following polynomial of degree $2m - 2$

$$p_{2m-2}(x) = \prod_{\substack{j=1 \\ j \neq i}}^{m} \frac{(x - x_j)^2}{(x_i - x_j)^2},$$

where $(x_j)_{j=1}^m$ are the roots of the polynomial $p_m(x)$. Note that $p_{2m-2}(x) = (l_i(x))^2$, with $l_i(x)$ being the $i$th elementary Lagrange polynomial. Clearly, $p_{2m-2}(x_j) = \delta_{i,j}$. Since the degree of $p_{2m-2}$ is less than $2m-1$ and $Q_m^G$ is the unique choice of nodes and weights that integrates all polynomials of degree at most $2m-1$ exactly, we obtain

$$I_\nu[p_{2m-2}] = Q_m^G[p_{2m-2}] = \sum_{j=1}^m w_j^G p_{2m-2}(x_j^G).$$

Moreover, as the nodes of $Q_m^G$ coincide with the roots of $p_{2m-2}$, we have:

$$\sum_{j=1}^m w_j^G p_{2m-2}(x_j^G) = \sum_{j=1}^m w_j^G \delta_{i,j} = w_i^G > 0,$$

where the strict positivity of $w_i^G$ comes from the fact that $I_\nu[p_{2m-2}] = \int_a^b p_{2m-2}(x)\nu(x)dx$ is strictly positive as its integrand is non-zero almost everywhere on $[a,b]$. □

**Theorem B.2.** *[13] Let $f^{(2m)}$ denote the derivative of order $2m$ of $f$. If $f^{(2m)}$ is continuous and non-negative, then*

$$Q_m^G[f] \leq Q_n^G[f], \quad \textit{for } m < n.$$

*Proof.* Consider a polynomial of order $2m-1$, such that

$$\begin{aligned} p_{2m-1}(x_i^G) &= f(x_i^G), \quad i = 1, \ldots, m, \\ p'_{2m-1}(x_i^G) &= f'(x_i^G), \quad i = 1, \ldots, m. \end{aligned}$$

Since $Q_m^G$ is exact for polynomials of order $2m-1$, i.e. $R_m^G[P_{2m-1}] = \{0\}$,

$$I[f] \approx Q_m^G[f] \quad \leftrightarrow \quad I[f - p_{2m-1}] \approx Q_m^G[f - p_{2m-1}].$$

Consequently,

$$\begin{aligned} Q_n^G[f] - Q_m^G[f] &= Q_n^G[f - p_{2m-1}] - Q_m^G[f - p_{2m-1}] \\ &= \sum_{i=1}^n w_i^G \left[f(x_i^G) - p_{2m-1}(x_i^G)\right] - \sum_{i=1}^m w_i^G \left[f(x_i^G) - p_{2m-1}(x_i^G)\right] \\ &= \sum_{i=1}^n w_i^G \left[f(x_i^G) - p_{2m-1}(x_i^G)\right] \geq 0, \end{aligned}$$

where the latter inequality follows from the strict positivity of the Gaussian weights, $w_i^G > 0$ as seen in Theorem B.1, and the non-negativity of the residual representation

$$f(x) - p_{2m-1}(x) = \frac{f^{(2m)}(c(x))}{(2m)!} \prod_{j=1}^m (x - x_j)^2,$$

with $c(x)$ being a point in $(a,b)$. Here, $p_{2m-1}(x)$ is represented as an Hermite polynomial and $f(x) - p_{2m-1}(x)$ is the Hermite interpolation error. □

**Remark B.1.** Theorem B.2 can be directly applied to the case of generalized Gaussian quadrature with an arbitrary weight function $\nu$ [13].

We now present an additional result concerning the representation of Gaussian quadrature as a Riemann-Stieltjes sum, stated without proof. This follows the definitions of the Riemann-Stieltjes integral; for further details, we refer to [60].

**Definition B.1** (Riemann-Stieltjes sum)**.** *Let $[a, b]$ be a given interval. Define a partition $P$ of $[a, b]$ to be a set of points $x_0, x_1, \ldots, x_n$ where*

$$a = x_0 \leq x_1 \leq \ldots \leq x_n = b$$

*and let $\|P\| = \max_{1\leq i\leq n} |x_i - x_{i-1}|$ be the mesh. For each $1 \leq i \leq n$, choose a point to represent the subinterval $[x_{i-1}, x_i]$. A choice for the partition $P$ is a finite set $Y = \{y_0, \ldots, y_n\}$ of points with*

$$x_{i-1} \leq y_i \leq x_i \quad \text{for each} \quad 1 \leq i \leq n.$$

*The Riemann–Stieltjes sum with partition $P$ and choice $Y$ is*

$$S(P, Y, f, \alpha) = \sum_{i=1}^{n} f(t_i) \left[\alpha(x_i) - \alpha(x_{i-1})\right].$$

**Definition B.2** (Riemann-Stieltjes integral)**.** *A function $f : [a, b] \to \mathbb{R}$ is said to be Riemann integrable with respect to $\alpha$ on $[a, b]$, denoted $f \in \mathcal{R}(\alpha)$ on $[a, b]$, if there exists an $I \in \mathbb{R}$ such that*

$$\forall\, \varepsilon > 0, \exists\, a \text{ partition } P_\varepsilon \text{ of } [a, b] \text{ s.t.} \qquad \begin{array}{l} |S(P, Y, f, \alpha) - I| < \varepsilon \\ \text{for all partitions } P \text{ of } [a, b] \text{ finer than } P_\varepsilon, \\ \text{i.e. } P \subset P_\varepsilon, \text{ and all choices } Y \text{ for } P. \end{array}$$

*If so, $I$ is denoted by $\int_a^b f(x)d\alpha(x)$ or $\int_a^b f d\alpha$ and we write*

$$\lim_P S(P, Y, f, \alpha) = \int_a^b f(x)d\alpha(x).$$

**Theorem B.3.** *[14, Theorem 6.4.2] Let $Q_m^G$ be a Gaussian quadrature rule with nodes $x_i^G$ and weights $w_i^G$, let $\nu : [a, b] \to (0, \infty)$ be a continuous weight function, and let $f : [a, b] \to \mathbb{R}$ be a function that we wish to integrate. Then there exist numbers $y_i$ with*

$$a = y_0 \leq x_1 \leq y_1 \leq x_2 \leq \ldots \leq x_n \leq y_n = b,$$

*and*

$$w_i^G = \int_{y_{i-1}}^{y_i} \nu(u)du.$$

*This means than $Q_m^G$ is a Riemann-Stieltjes sum, i.e.*

$$Q_m^G[f] = \sum_{i=1}^{m} f(x_i) \int_{y_{i-1}}^{y_i} \nu(u)du.$$