Technische Universität Berlin

Institut für Mathematik

Fakultät II
Straße des 17. Juni 136
10623 Berlin
https://www.tu.berlin/math

M.Sc. Thesis

Second-order Weak Schemes and Error Bounds for Rough Heston and Quadratic Rough Heston Models

Runzhe Yang

Matriculation Number: 0508536

March 24th

Supervised by

Dr. Christian Bayer &
Prof. Dr. Peter Friz

Declaration of authorship

I hereby declare that the thesis submitted is my own, unaided work, completed without any external help. Only the sources and resources listed were used. All passages taken from the sources and aids used, either unchanged or paraphrased, have been marked as such.

Where generative AI tools were used, I have indicated the product name, manufacturer, the software version used, as well as the respective purpose (e.g. checking and improving language in the texts, systematic research). I am fully responsible for the selection, adoption, and all results of the AI-generated output I use.

I have taken note of the Principles for Ensuring Good Research Practice at TU Berlin dated 15 February 2023. https://www.static.tu.berlin/fileadmin/www/10002457/K3-AMBl/Amtsblatt_2023/Amtliches_Mitteilungsblatt_Nr._16_vom_30.05.2023.pdf

I further declare that I have not submitted the thesis in the same or similar form to any other examination authority.

Berlin, March 24th

. . . .

Abstract

This thesis studies and solves two problems arising in rough volatility models, focusing on the rough Heston model and the quadratic rough Heston model, which extend the classical Heston framework in mathematical finance. We provide the necessary background, motivation, and fundamental properties of these models in Chapter 1.

The first problem is addressed in Chapter 2, where we prove that the weak scheme introduced in [BB23a] for the rough Heston model achieves a second-order weak convergence rate, thereby establishing a result not proven in the original work. The analysis is based on the weak Markovian approximation introduced in [BB23c] and relies on the Talay–Tubaro's analysis method. We also discuss potential applications of this second-order scheme.

The second problem is studied in Chapter 3, concerning the derivation of an error bound for the weak Markovian approximation of the quadratic rough Heston model. Due to the lack of explicit structural information on the characteristic function, standard techniques are not applicable, making the analysis significantly more challenging. To overcome this, we introduce polynomial Volterra processes to describe the volatility and extend the path-dependent PDE (PPDE) method in [BMM25] and [BJP25] to a broader class of stochastic Volterra equations by developing the corresponding Malliavin calculus framework. These results allow the establishment of a path-dependent Feynman–Kac formula and a functional Itô formula, which are essential ingredients for obtaining the error bound. We show that this method can be applied on the quadratic rough Heston model to yield an error bound of the order

$$\int_0^T \Big\{ \|K - K^N\|_{L^1([0,t])} + \|K^2 - (K^N)^2\|_{L^1([0,t])} \Big\} dt.$$

An efficient weak scheme for simulating the quadratic rough Heston model is also proposed in the last of Chapter 3.

Finally, in Chapter 4, numerical experiments are conducted to validate the theoretical results. In particular, we testify the second-order weak convergence of the proposed scheme and investigate the weak convergence behavior of numerical methods for the quadratic rough Heston model.

Keywords: Rough Volatility; Rough Heston Model; Quadratic Rough Heston Model; Error Bound; Malliavin Calculus; Functional Stochastic Calculus; Polynomial Volterra Processes; Weak Markovian Approximation

Zusammenfassung

Diese Arbeit untersucht und löst zwei Probleme, die in Rough-Volatility-Modellen auftreten, mit Schwerpunkt auf dem Rough-Heston-Modell und dem Quadratic Rough-Heston-Modell, die das klassische Heston-Modell im Bereich der Finanzmathematik erweitern. In Kapitel 1 werden die notwendige Hintergrundinformationen, die Motivation und die grundlegenden Eigenschaften dieser Modelle vorgestellt.

Das erste Problem wird in Kapitel 2 behandelt, in dem gezeigt wird, dass das in [BB23a] eingeführte schwache Schema für das Rough-Heston-Modell eine schwache Konvergenz zweiter Ordnung erreicht und damit ein Ergebnis liefert, das in der ursprünglichen Arbeit nicht bewiesen wurde. Die Analyse basiert auf der in [BB23c] vorgestellten schwachen Markov-Approximation und stützt sich auf die Methode von Talay–Tubaro. Außerdem werden mögliche Anwendungen dieses Schemas zweiter Ordnung diskutiert.

Das zweite Problem wird in Kapitel 3 untersucht und betrifft die Herleitung einer Fehlerabschätzung für die schwache Markov-Approximation des Quadratic Rough-Heston-Modells. Aufgrund fehlender expliziter Strukturdaten der Charakteristikfunktion sind die Standardtechniken nicht anwendbar, was die Analyse erheblich erschwert. Zur Lösung dieses Problems werden polynomiale Volterra-Prozesse zur Beschreibung der Volatilität eingeführt, und der pfadabhängige PDE-Ansatz (PPDE) aus [BMM25] und [BJP25] wird auf eine breitere Klasse stochastischer Volterra-Gleichungen erweitert, wobei ein entsprechendes Malliavin-Kalkül-Rahmenwerk entwickelt wird. Diese Ergebnisse ermöglichen die Formulierung der pfadabhängigen Feynman–Kac-Formel sowie der funktionalen Itô-Formel, die für die Herleitung der Fehlerabschätzung unerlässlich sind. Es wird gezeigt, dass diese Methode auf das Quadratic Rough-Heston-Modell angewendet werden kann, um eine Fehlerabschätzung der Größenordnung

$$\int_0^T \Big\{ \|K - K^N\|_{L^1([0,t])} + \|K^2 - (K^N)^2\|_{L^1([0,t])} \Big\} dt$$

zu erhalten. Am Ende von Kapitel 3 wird außerdem ein effizientes schwaches Schema zur Simulation des Quadratic Rough-Heston-Modells vorgestellt.

Abschließend werden in Kapitel 4 numerische Experimente durchgeführt, um die theoretischen Ergebnisse zu validieren. Insbesondere wird die schwache Konvergenz zweiter Ordnung des vorgeschlagenen Schemas überprüft, und das Verhalten der schwachen Konvergenz numerischer Methoden für das Quadratic Rough-Heston-Modell wird untersucht.

Schlüsselwörter: Rough Volatility; Rough-Heston-Modell; Quadratic Rough-Heston-Modell; Fehlerabschätzung; Malliavin-Kalkül; Funktionales stochastisches Kalkül; Polynomiale Volterra-Prozesse; Schwache Markov-Approximation

Acknowledgements

I would like to sincerely thank my master's thesis supervisor, Dr. Christian Bayer, for introducing me to this fascinating topics, which marked the beginning of my journey into mathematical finance and stochastic volatility models.

I would like to express my sincere gratitude to the Berlin Mathematical School (BMS) for generously providing me with a full scholarship and travel support, which made my studies in Berlin possible. During my time there, I had the opportunity to attend a wide range of advanced courses and gained valuable mathematical insights.

I am also grateful to my Phase I supervisor, Prof. Peter Friz at BMS, for his continuous support, which has provided me with an enriching and rewarding academic experience in Berlin.

I greatly appreciate the valuable weekly discussions with Sorelle and Dr. Bayer, which have been immensely helpful throughout my research.

Finally, I would like to thank my family and my friends, on whom I have always been able to rely for support and encouragement.

Contents

List of Figures

List of Tables

1. Introduction to the Heston Models

1.1. Classical Heston Model

The Heston model, first introduced by Steve Heston in 1993 in [Hes93], is a well-known option pricing model in financial mathematics. The dynamics are given by

$$d\widetilde{S}_t = \sqrt{\widetilde{V}_t}\widetilde{S}_t\left(\rho dW_t + \sqrt{1-\rho^2}dB_t\right), \qquad \widetilde{S}_0 = S_0, \tag{1.1}$$

$$d\widetilde{V}_t = (\theta - \lambda\widetilde{V}_t)dt + \nu\sqrt{\widetilde{V}_t}dW_t, \qquad \widetilde{V}_0 = V_0, \tag{1.2}$$

where $\theta > 0$ is long-term variance level , $\lambda \geq 0$ is mean reversion speed , $\nu > 0$ is volatility of volatility, $\rho \in [-1, 1]$ is the correlation between spot and volatility, and (B_t, W_t) is a two-dimensional standard Brownian motion.

By the Yamada–Watanabe theorem (see [KS12]), the model admits a unique strong solution. Moreover, the variance process can be written in the following integral form:

$$\widetilde{V}_u = \widetilde{V}_t e^{-\lambda(u-t)} + \theta\left(1 - e^{-\lambda(u-t)}\right) + \nu\int_t^u e^{-\lambda(u-s)}\sqrt{\widetilde{V}_s}\,dW_s, \quad \text{for any } t \leq u. \tag{1.3}$$

This model is widely adopted for several compelling reasons. From a statistical perspective, it is capable of capturing key stylized features of low-frequency asset price data, including the leverage effect, time-varying volatility, and heavy-tailed return distributions. From the perspective of option pricing, as a stochastic volatility framework, it provides a realistic and flexible description of the implied volatility surface.

Moreover, compared with other stochastic volatility models, one of its most attractive features is its affine structure: the joint process $(\widetilde{S}, \widetilde{V})$ is affine. As a consequence, the characteristic function of the log-price admits an explicit representation, as shown in [Hes93]. In particular, for any $u \in \mathbb{R}$,

$$\mathbb{E}\left[\left(\frac{\widetilde{S}_t}{\widetilde{S}_0}\right)^{iu}\right] = \exp\left(g(u,t) + \widetilde{V}_0 h(u,t)\right), \tag{1.4}$$

where h is solution of the Riccati equation

$$\partial_t h(u,t) = -\frac{u(u+i)}{2} + (iu\varrho\nu - \lambda)h(u,t) + \frac{\nu^2}{2}h^2(u,t), \tag{1.5}$$

with boundary condition $h(u,0)=0$ and

$$g(u,t)=\theta\lambda\int_0^t h(u,s)ds. \tag{1.6}$$

1.2. Rough Heston Model

One limitation of the classical Heston model is its inability to capture certain stylized facts observed in financial markets. For instance, it fails to account for the explosion of the implied volatility skew, see [ER19]. Moreover, empirical studies suggest that the paths of the volatility process typically exhibit rougher behavior, see [GJR18]. Consequently, attention has shifted to a rough version of the Heston model, which arises naturally as a suitable framework for price dynamics based on microstructural considerations and no-arbitrage arguments. This model, introduced in [ER19], takes the form

$$dS_t=\sqrt{V_t}S_t\left(\rho dW_t+\sqrt{1-\rho^2}dB_t\right),\qquad S_0=S_0, \tag{1.7}$$

$$V_t=V_0+\int_0^t K(t-s)(\theta-\lambda V_s)ds+\int_0^t K(t-s)\nu\sqrt{V_s}dW_s, \tag{1.8}$$

where K is the fractional kernel

$$K(t):=\frac{t^{H-1/2}}{\Gamma(H+1/2)}=\int_0^\infty e^{-xt}\mu(dx), \tag{1.9}$$

$$\mu(dx):=\frac{x^{-H-1/2}dx}{\Gamma(H+1/2)\Gamma(1/2-H)}, \tag{1.10}$$

for all $t>0$, with Hurst parameter $H\in(-1/2,1/2)$, and Γ is the Gamma function. Setting the Hurst parameter to $\frac{1}{2}$ recovers the classical Heston model. Rough volatility models, such as the rough Heston model (1.7)–(1.8), have become a standard framework for modeling equity markets, as they provide an accurate representation of the observed market dynamics; see, e.g., [BFF$^+$23]. Moreover, they yield remarkable fits to the implied volatility surface while remaining almost as tractable as the classical Heston model.

Similar to the classical Heston model, the rough Heston model constitutes a Volterra affine process, which allows for the quasi–closed-form computation of the characteristic function of the log-price. This is achieved through a connection between Hawkes processes and the rough Heston model, as established in [ER19]. We define the characteristic function of the log-price $\log(S_T)$ as

$$\varphi(z):=\mathbb{E}\left[\exp\left(z\log(S_T/S_0)\right)\right].$$

Then φ can be explicitly expressed in semi-closed form as

$$\varphi(z) = \exp\left(\int_0^T F(z, \psi(T-t, z))g(t)dt\right), \tag{1.11}$$

where g is defined by

$$g(t) = V_0 + \theta \int_0^t K(s)ds = V_0 + \theta \frac{t^{H+1/2}}{\Gamma(H+3/2)}, \tag{1.12}$$

and ψ is the solution to the fractional Riccati equation

$$\psi(t, z) = \int_0^t K(t-s)F(z, \psi(s, z))ds, \tag{1.13}$$

with

$$F(z, x) = \frac{1}{2}(z^2 - z) + (\rho\nu z - \lambda)x + \frac{\nu^2}{2}x^2. \tag{1.14}$$

1.3. Quadratic Rough Heston Model

Although the rough Heston model provides remarkable fits for most volatility features, Jean-Philippe Bouchaud raised the following question: Can a rough volatility model reproduce the so-called Zumbach effect, an empirical phenomenon first identified by Gilles Zumbach (see [Zum09]) that reflects the lack of time-reversal invariance in financial time series? To address this question, two formal definitions of the Zumbach effect have been introduced.

- The weak Zumbach effect (typically considered in the econophysics literature; see [Zum09]): Past squared returns forecast future integrated volatilities better than past integrated volatilities forecast future squared returns. This property is not satisfied in classical stochastic volatility models. However, rough stochastic volatility models are consistent with the weak Zumbach effect.

- The strong Zumbach effect: Conditional dynamics of volatility with respect to the past depend not only on the past volatility trajectory but also on the historical price path; specifically, price trends tend to increase volatility; see [Zum10]. Such feedback of the historical price path on volatility also occurs on implied volatility as illustrated in [Zum10]. Rough stochastic volatility models such as the rough Heston model are not consistent with the strong Zumbach effect, see [ER18]

The quest for a rough volatility model consistent with the strong Zumbach effect, together with the empirical success of quadratic Hawkes process–based models documented in [BDB17], led to the development of super-Heston rough volatility models in [DJR21]. These models extend the rough Heston framework and arise as scaling limits of quadratic Hawkes process–based microstructural models, in a manner analogous to the emergence of the rough Heston model as the continuous-time limit of a linear Hawkes process–based microstructural model.

The quadratic rough Heston model considered here is a special case of the super-Heston rough volatility models introduced in [DJR21]. The joint dynamics of the asset S and its spot variance V are given by

$$dS_t = S_t\sqrt{V_t}dW_t, \tag{1.15}$$

$$V_t = a(Z_t - b)^2 + c, \tag{1.16}$$

where W is a Brownian motion, a, b and c some positive constants and Z_t follows a SDE given by

$$Z_t = \int_0^t (t-s)^{\alpha-1}\frac{\lambda}{\Gamma(\alpha)}(\theta - Z_s)ds + \int_0^t (t-s)^{\alpha-1}\frac{\lambda}{\Gamma(\alpha)}\eta\sqrt{V_s}dW_s, \tag{1.17}$$

with $\alpha \in (\frac{1}{2}, 1)$, $\lambda > 0$, $\eta > 0$. In this special case of a quadratic rough Heston model, the asset S and its volatility depend on the history of only one Brownian motion W. The model therefore constitutes a pure feedback framework, in which volatility is entirely driven by price dynamics, without any additional exogenous source of randomness. In general, the volatility process need not depend solely on the Brownian motion driving the asset price S. For the sake of simplicity, however, we will refer to this pure feedback specification simply as the quadratic rough Heston model.

As in the general case of super-Heston rough volatility models, the quadratic rough Heston model also exhibits the strong Zumbach effect, since the effect of past returns on Z cannot be reduced to an influence of past volatility dynamics on Z, see [DJR21].

1.3.1. Parameter Interpretation

As discussed in [GJR20], the parameters a, b and c in the specification

$$V_t = a(Z_t - b)^2 + c$$

can be interpreted in the following way.

- c represents the minimal instantaneous variance. When calibrating the model, we use c to shift upward or downward the smiles of SPX options.
- $b > 0$ encodes the asymmetry of the feedback effect. Indeed for the same absolute value of Z, volatility is higher when Z is negative than when it is positive. Such asymmetry aims at reproducing the empirical behavior of the VIX. From a calibration viewpoint, the higher b the more SPX options smiles are shifted to the right.

- a is the sensitivity of the volatility to the feedback of price returns. The greater a, the greater the role of feedback in the model and the higher is volatility of volatility. Consistent with this SPX smiles become more extreme as a increases.

1.3.2. SPX and VIX

Although the rough Heston model provides an excellent fit to the SPX implied volatility surface, it fails to achieve satisfactory calibration when applied to VIX options. By definition, the VIX index is a derivative of the SPX index S, and can be represented as

$$\mathrm{VIX}_t = \sqrt{-2\mathbb{E}[\log(S_{t+\Delta}/S_t)|\mathcal{F}_t]} \times 100,$$

where $\Delta = 30$ days and $\mathbb{E}$ is the risk-neutral expectation. Consequently, VIX options are also derivatives of SPX. Finding a model which jointly calibrates the prices of SPX and VIX options is known to be very challenging, especially for short maturities. As indicated in [Guy20]:

> The very large negative skew of short-term SPX options, which in classical continuous SV models implies a very large volatility of volatility, seems inconsistent with the comparatively low levels of VIX implied volatilities.

As explained in [GJR20], combining rough volatility and the Zumbach effect can be key to solving the joint calibration problem with continuous-time models with continuous sample paths. Along with [RZ22], they showed that the quadratic rough Heston model enables us to obtain the following with only six parameters:

- VIX smiles in the bid-ask spread,
- a global shape of the implied volatility surface of the SPX that is very well reproduced, in particular the ATM skew, which can be exactly matched,
- model SPX skews typically of orders -1.5 (shortest maturities) to -1 (longer maturities) as for market data.

1.3.3. The Volatility Process

We observe that the volatility process V_t is a quadratic transformation of process Z_t which can be interpreted as a weighted moving average of past log-returns of the asset price. Indeed, from [ER18, Lemma A.1], we have

$$Z_t = \int_0^t \theta f^{\alpha,\lambda}(t-s)ds + \int_0^t f^{\alpha,\lambda}(t-s)\eta\sqrt{V_s}dW_s, \tag{1.18}$$

where $f^{\alpha,\lambda}(t)$ is the Mittag-Leffler density function defined for $t \geq 0$ as

$$f^{\alpha,\lambda}(t) = \lambda t^{\alpha-1}E_{\alpha,\alpha}(-\lambda t^{\alpha}), \tag{1.19}$$

with

$$E_{\alpha,\beta}(z) = \sum_{n\geq 0} \frac{z^n}{\Gamma(\alpha n + \beta)}. \tag{1.20}$$

is thus path-dependent, representing a weighted average of past returns of the type typically considered in path-dependent volatility models; see [HR98]. As explained in [Guy14], modeling with path-dependent variables provides a natural framework for capturing the empirical observation that volatility depends on recent price changes. Consequently, under the quadratic rough Heston dynamics, we obtain a model that exhibits

- rough volatility,
- essentially lognormal volatility,
- strong Zumbach effect.

However the kernels used to model this dependency are typically exponential, see for example [HR98]. Here a crucial idea, motivated by [DJR21], is to use a rough kernel, more precisely the Mittag-Leffler density function. Thanks to this kernel, the "memory" of Z decays as a power law and Z is highly sensitive to recent returns since

$$f^{\alpha,\lambda}(t) \underset{t\to+\infty}{\sim} \frac{\alpha}{\lambda\Gamma(1-\alpha)} t^{-\alpha-1} \quad \text{and} \quad f^{\alpha,\lambda}(t) \underset{t\to 0^+}{\sim} \frac{\lambda}{\Gamma(\alpha)} t^{\alpha-1}.$$

This essentially means that long periods of trends or sudden upwards or downwards moves of the price generate large values for $|Z|$ and so high volatility, in particular when Z is negative. Finally the choice of $f^{\alpha,\lambda}$ as kernel ensures that the volatility process is rough, with Hurst parameter equal to $H = \alpha - 1/2$. As shown in [GJR18], this enables us to reproduce the behavior of historical volatility time series provided H is taken of order 0.1.

As discussed above, an immediate implication of the feedback effect is that negative price trends lead to high volatility levels. Moreover, such trends also affect the instantaneous variance of volatility in our model. To see this, consider the classical case with $\alpha = 1$. In that case, an application of Itô's formula yields, up to a drift term,

$$dV_t = 2a(Z_t - b)\eta\sqrt{V_t}dW_t. \tag{1.21}$$

Hence, the coefficient governing the "variance of instantaneous variance" is proportional to $a(Z_t - b)^2$ which, up to c, is equal to the variance of $\log S$. Consequently, when volatility is high, the volatility of volatility is also high.

1.3.4. Infinite Dimensional Markovian Representation

Although the quadratic rough Heston model is not Markovian in the variables (S, V), it admits an infinite-dimensional Markovian representation. Following the computations in [ER18], we obtain that for any t, $t_0 > 0$,

$$Z_{t_0+t} = \int_0^t (t-s)^{\alpha-1} \frac{\lambda}{\Gamma(\alpha)} (\theta_{t_0}(s) - Z_{t_0+s}) ds + \int_0^t (t-s)^{\alpha-1} \frac{\lambda}{\Gamma(\alpha)} \eta \sqrt{V_{t_0+s}} dW_{t_0+s}, \quad (1.22)$$

with θ_{t_0} a $\mathcal{F}_{t_0}$-measurable function. More precisely, θ_{t_0} is given by

$$\theta_{t_0}(u) = \theta + \frac{\alpha}{\lambda\Gamma(1-\alpha)} \int_0^{t_0} (t_0 - v + u)^{-1-\alpha} (Z_v - Z_{t_0}) dv. \quad (1.23)$$

i.e., θ_{t_0} encodes all the information of Z_t over the time interval $[0, t_0]$. Mathematically, it means $\theta_{t_0} \in \sigma(Z_s, s \leq t_0)$.

The equation (1.23) implies that the law of $(S_t, V_t)_{t \geq t_0}$ only depends on S_{t_0} and $\theta_{t_0}(\cdot)$. In view of [AJLP19], it means that we can express the VIX at time t as a function of θ_t and S_t. So we can write the price of any European option with pay-off depending on SPX and VIX as a function of time, S and θ.

2. Rough Heston Models

In this chapter, we first introduce the weak Markovian approximation for the rough Heston model, as proposed in [BB23c], along with the corresponding weak simulation method developed in [BB23a]. We then prove that the weak simulation scheme presented in Section 2.2 achieves a strongly potential second-order weak convergence rate—a result not established in [BB23a]. After deriving this second-order weak scheme, we present several applications of the scheme.

2.1. Weak Markovian Approximation

Although the rough Heston model resolves several shortcomings of the classical Heston model, the singular Volterra-type dynamics governing the volatility process V in (1.8) pose substantial challenges for both theoretical analysis and practical tasks such as simulation and option pricing. In particular, the rough Heston model is neither a semimartingale nor a Markov process, and the sample paths of V are only $(H-\varepsilon)$-Hölder continuous, for all $\varepsilon > 0$, where $H \in (0, 1/2)$ is the Hurst parameter in (1.9). These features create significant difficulties in both theoretical analysis and the simulation of sample paths, as well as in option pricing more generally.

Nevertheless, the rough Heston model is an affine Volterra process, as introduced in [AJCP+24], and can be analyzed accordingly. This framework implies weak existence and uniqueness of solutions, as well as a semi-explicit formula for the characteristic function in terms of a fractional Riccati equation, see (1.11).

One approach to addressing these challenging properties of the model is to employ Markovian approximations of V. More precisely, we first observe that K satisfies the following properties.

Proposition 2.1.1. *The kernel K is completely monotone, and can thus be written as*

$$K(t) = \int_0^\infty e^{-xt}\mu(dx),\ \ \mu(dx) := c_H x^{-H-1/2}dx,\ \ c_H := \frac{1}{\Gamma(H+1/2)\Gamma(1/2-H)}, \quad (2.1)$$

for all $t > 0$, where Γ is the gamma function.

Remark 2.1.2. A notable example is the fractional Brownian motion W^H. By applying the stochastic Fubini theorem, it can be expressed as

$$\begin{aligned} W_t^H &= \int_0^t K(t-s)dW_s = \int_0^t \int_0^\infty e^{-x(t-s)}\mu(dx)dW_s \\ &= \int_0^\infty \int_0^t e^{-x(t-s)}dW_s\mu(dx) =: \int_0^\infty Y_t(x)\mu(dx), \end{aligned} \tag{2.2}$$

where W is a Brownian motion. Note that $Y_t(x)$ is an Ornstein-Uhlenbeck process driven by W with mean-reversion rate x. it was shown that $(Y_t(x))_{x>0}$ is an infinite-dimensional Markov process in [CC98].

The representation (2.2) suggests a natural approach for approximating fractional Brownian motion by a finite-dimensional Markov process. Specifically, we discretize the integral over $\mu(dx)$ in (2.2), yielding an approximation of W^H as a linear functional of a finite-dimensional Ornstein–Uhlenbeck (and thus Markov) process. Moreover, it is straightforward to see that discretizing the integral in (2.2) is equivalent to discretizing the integral in (2.1), that is, to approximating K by a discrete sum of exponential functions K^N given by

$$K^N(t) := \sum_{i=1}^N w_i e^{-x_i t}, \tag{2.3}$$

for some non-negative nodes $(x_i)_{i=1}^N$ and non-negative weights $(w_i)_{i=1}^N$.

Besides, using the approximation K^N of K, we can define the approximation (S^N, V^N) of (S, V) by

$$\begin{aligned} dS_t^N &= \sqrt{V_t^N}S_t^N\left(\rho dW_t + \sqrt{1-\rho^2}dB_t\right), \qquad S_0 = S_0, \\ V_t^N &= V_0 + \int_0^t K^N(t-s)(\theta - \lambda V_s^N)ds + \int_0^t K^N(t-s)\nu\sqrt{V_s^N}dW_s. \end{aligned} \tag{2.4}$$

We also remark that the solution (S^N, V^N) can also be represented in terms of the solution $(S^N, \boldsymbol{V}^N)$ to another $(N+1)$ - dimensional stochastic differential equation by applying [AJEE19a, Theorem 3.1], which is given by

$$dS_t^N = \sqrt{V_t^N}S_t^N\left(\rho dW_t + \sqrt{1-\rho^2}dB_t\right), \quad S_0 = S_0, \tag{2.5}$$

$$dV_t^{(i)} = -x_i(V_t^{(i)} - v_0^i)dt + (\theta - \lambda V_t^N)dt + \nu\sqrt{V_t^N}dW_t, \quad V_0^{(i)} = v_0^i, \tag{2.6}$$

for $i = 1, \ldots, N$, satisfying $V_t^N = \boldsymbol{w}\cdot\boldsymbol{V}_t^N$, $\boldsymbol{w}\cdot\boldsymbol{v}_0 = V_0$ and $\boldsymbol{V}^N$ is an N-dimensional diffusion. This representation will be very useful when deriving the weak simulation for (S^N, V^N) lately. And we can see that for dimension $N = 1$, it's just the classical Heston model.

Given the Markovian approximation (S^N, V^N) of (S, V), in the following two sections, we will give a error bound for this approximation and the attainable convergence rate for K^N of K by some quadrature rules.

2.1.1. Error Bound for Markovian Approximation

Assumption 2.1.3. Assume that K^N are chosen such that

$$e_N := \int_0^T \left|K(t) - K^N(t)\right| dt \to 0,$$

and that there exists a constant $C > 0$ such that for all $t \in [0,T]$ and $N \geq 1$, we have

$$K^N(t) \leq CK(t), \qquad \left|(K^N)'(t)\right| \leq C\left|K'(t)\right|. \tag{2.7}$$

In fact, the derivation of the error bound for the Markovian approximation relies heavily on the semi-closed form of the characteristic function of the volatility $\log S_t$ because $(\log S_t, V_t)$ is a affine Volterra process. Recall the characteristic function of $\log S_t$ can be expressed in

$$\varphi(z) = \exp\left(\int_0^T F(z, \psi(T-t,z))g(t)dt\right), \tag{2.8}$$

where g is defined by

$$g(t) = V_0 + \theta\int_0^t K(s)ds = V_0 + \theta\frac{t^{H+1/2}}{\Gamma(H+3/2)}, \tag{2.9}$$

and ψ is the solution to the fractional Riccati equation

$$\psi(t,z) = \int_0^t K(t-s)F(z,\psi(s,z))ds, \tag{2.10}$$

with

$$F(z,x) = \frac{1}{2}(z^2 - z) + (\rho\nu z - \lambda)x + \frac{\nu^2}{2}x^2. \tag{2.11}$$

By replacing the kernel K with K^N and defining g^N and ψ^N as in (2.9) and (2.10), respectively, we obtain a formula for the characteristic function φ^N of the log-stock price $\log(S_T^N)$ that is analogous to (1.11).

First, we present a local Lipschitz error bound for the characteristic function, $\left|\varphi(z) - \varphi^N(z)\right|$, in terms of e_N. The following theorem is taken from [BB23c, Theorem 2.2].

Theorem 2.1.4. *Under the Assumption 2.1.3, for all $T \geq 0$ there exists a constant C_0 such that for all $N \geq 0$ and $z = a + bi$ with $a \in [0,1]$ and $b \in \mathbb{R}$ satisfying*

$$C_0(1 + b^6)e_N \leq 1, \tag{2.12}$$

we have

$$\left|\varphi(z) - \varphi^N(z)\right| \leq 6C_0\varphi(a)(1+b^6)e_N. \tag{2.13}$$

Remark 2.1.5. We see that the Markovian approximations φ^N converge weakly at the same speed as e_N.

And this result can be extended to the weak error in $\log(S_T)$. To prove it, we apply the Fourier method for option pricing. We first introduce some notations and definitions related to the Fourier transform.

Definition 2.1.6. Given a function $f : \mathbb{R} \to \mathbb{R}$, we denote its generalized Fourier transform by

$$\widehat{f}(z) \coloneqq \int_{\mathbb{R}} e^{izx} f(x) dx$$

for all $z \in \mathbb{C}$ for which the above integral is well-defined.

Definition 2.1.7. For $R \in \mathbb{R}$, we define the damped function

$$f_R(x) \coloneqq e^{-Rx} f(x).$$

Define the sets

$$\mathcal{I} \coloneqq \left\{ R \in \mathbb{R} \colon f_R \in L^1_{\mathrm{bc}},\ \widehat{f_R} \in L^1 \right\}, \qquad \mathcal{J} \coloneqq \left\{ R \in \mathbb{R} \colon \varphi(R) < \infty \right\}, \tag{2.14}$$

where L^1_{bc} denotes the set of L^1-functions that are bounded and continuous.

Remark 2.1.8. Note that $[0, 1] \subseteq \mathcal{J}$. Besides, for each $R \in \mathcal{I} \cap \mathcal{J}$, we have the Fourier pricing formula

$$\mathbb{E} f(X_T) = \frac{1}{2\pi} \int_{\mathbb{R}} \varphi(R - iu) \widehat{f}(u + iR) du, \tag{2.15}$$

in [EGP10, Theorem 2.2].

Using formula (2.15), we can establish [BB23c, Theorem 2.7], which is stated as follows.

Theorem 2.1.9. *Let $f : \mathbb{R} \to \mathbb{R}$ be a payoff function of the log-stock price, and let $a \in \mathcal{I} \cap [0, 1] \neq \emptyset$. Assume that there are $C_1, \delta > 0$ such that*

$$\left| \widehat{f}(b + ia) \right| \le C_1 (1 + |b|)^{-(1+\delta)}. \tag{2.16}$$

Then there exist constants $\widetilde{C}, \varepsilon > 0$ independent of f, C_1, δ, such that if $e_N < \varepsilon$ then, for $\delta \neq 6$,

$$\left| \mathbb{E} f(X_T) - \mathbb{E} f(X_T^N) \right| \le \widetilde{C} \left(\frac{1}{|6 - \delta|} + \frac{1}{\delta} \right) C_1 e_N^{\frac{\delta}{6} \wedge 1}.$$

For $\delta = 6$, we get

$$\left| \mathbb{E} f(X_T) - \mathbb{E} f(X_T^N) \right| \le \widetilde{C} C_1 \log(e_N^{-1}) e_N.$$

Remark 2.1.10. The proof of Theorem 2.1.9 relies solely on Theorem 2.1.4 and does not require any additional properties of the rough Heston model.

We also present a special case of Theorem 2.1.9, which is straightforward to apply.

Corollary 2.1.11. *Let $h \in C_K^8(\mathbb{R}_+;\mathbb{R})$. Then there is $C > 0$ such that*

$$\left|\mathbb{E}h(S_T) - \mathbb{E}h(S_T^N)\right| \leq Ce_N. \tag{2.17}$$

Remark 2.1.12. L^1-error in K performs very welll, whereas the L^2-error converges very slowly for small Hurst parameters $1/2 \gg H > 0$, as discussed in [AK21]. However, the approach in [AK21] is not applicable in our model because $\sqrt{x}$ is not Lipschitz.

Since $K(t) \approx t^{H-1/2}$, the singularity at $t = 0$ of $K(t)^2 \approx t^{2H-1}$ is barely integrable, leading to slow convergence rates. If using Gaussian quadrature rules to get K^N, [BB23b] shows one can achieve a convergence rate

$$\left(\int_0^T \left|K(t) - K^N(t)\right|^2 dt\right)^{1/2} \leq C \exp\left(-1.064 \left(1 + \frac{H}{3/2 - H}\right)^{-1/2} \sqrt{HN}\right). \tag{2.18}$$

It is evident that this rate becomes very slow for small $H \approx 0$. And the method is not applicable in the hyper-rough case $-1/2 < H \leq 0$, see [JR20].

2.1.2. Convergence Rate for K^N

In [BB23c], the authors improve upon the geometric Gaussian quadrature method of [BB23b] by choosing non-geometrically spaced intervals $[\xi_i, \xi_{i+1}]$ to construct an approximate kernel K^N, such that

$$e_N := \int_0^T \left|K(t) - K^N(t)\right| dt$$

will converge rapidly for $N \to \infty$.

The definition of this quadrature rule in [BB23c] is given as follows.

Definition 2.1.13 (Non-geometric Gaussian rules)**.** Let $N \in \mathbb{N}$ be the total number of nodes and $\beta, a, c \in (0, \infty)$ be parameters of the scheme not depending on N, where $c > 1$. We define

$$m := \mathrm{rd}\left(\beta\sqrt{(1/2 + H)N}\right), \qquad n := \mathrm{rd}\left(\frac{1}{\beta}\sqrt{\frac{N}{1/2 + H}}\right),$$

$$\xi_0 := 0, \quad \xi_1 := a, \quad \xi_{i+1} = \left(\frac{c + \xi_i^{\frac{1/2+H}{2m}}}{c - \xi_i^{\frac{1/2+H}{2m}}}\right)^2 \xi_i, \qquad i = 1, \ldots, n-1, \tag{2.19}$$

where $\mathrm{rd} : \mathbb{R} \to \mathbb{N}_+$ is the function rounding to the nearest positive integer. Here, we assume that $\xi_i^{\frac{1/2+H}{2m}} < c$ for all $i = 1, \ldots, n-1$, which is true if we choose c large enough. Define $(x_i)_{i=1}^m$ to be the nodes and $(\widetilde{w}_i)_{i=1}^m$ to be the weights of Gaussian quadrature of level m on the interval $[0, \xi_1]$ with the weight function $w(x) = c_H x^{-H-1/2}$. Furthermore, let $(x_i)_{i=m+1}^{mn}$ be the nodes and $(w_i)_{i=m+1}^{mn}$ be the weights of Gaussian quadrature of level

m on to the intervals $[\xi_i, \xi_{i+1}]$ for $i = 1, \dots, n-1$ with the weight function $w(x) \equiv 1$. Then we define the non-geometric Gaussian rule of type (H, N, c, β, a) to be the set of nodes $(x_i)_{i=1}^{mn}$ with weights $(w_i)_{i=1}^{mn}$ defined by

$$w_i := \begin{cases} \widetilde{w}_i, & \text{if } i = 1, \dots, m, \\ c_H \widetilde{w}_i x_i^{-H-1/2}, & \text{else.} \end{cases}$$

Remark 2.1.14. (i) Non-geometric Gaussian rules have three free parameters: β, a, c which can be interpreted as in [BB23c].

- The parameter c determines the cutoff point ξ_n of the interval in (2.1), similar to the parameter α in geometric Gaussian rules. It is hence mainly used to control the error on $[\xi_n, \infty)$.
- The parameter β determines the degree of the Gaussian quadrature rule, similar to the parameter β in geometric Gaussian rules. It is hence mainly used to control the error on $[\xi_1, \xi_n]$.
- The parameter a determines the size of the interval $[0, a]$, which is again special due to the singularity in the weight function w. Similarly to the parameter a in geometric Gaussian rules, it is mainly used to control the error on $[0, \xi_1]$.

(ii) The reason for the specific definition of ξ_{i+1} in (2.19) can be seen in the proof of [BB23c, Lemma 3.13].

[BB23c, Theorem 3.16] provides the error bound for non-geometric Gaussian rules, given in the following.

Theorem 2.1.15. *Define the constants $\beta_0 = 0.92993273$, and $c_0 = 3.60585021$. Let K^N be a Gaussian approximation coming from a non-geometric Gaussian rule with parameters $c \geq c_0$, $\beta \geq \beta_0$, and $a > 0$, where $c > c_0$ or $\beta > \beta_0$. Then,*

$$\begin{aligned} \int_0^T \left| K(t) - K^N(t) \right| dt &\lesssim \frac{c_H}{H + 1/2} \xi_n^{-1/2-H} \\ &= \mathcal{O}(\exp\left(-\frac{\eta_1(c, \beta)}{\beta} \sqrt{(H + 1/2)N} \right)). \end{aligned}$$

In the first inequality, the higher order terms explode as $(c, \beta) \to (c_0, \beta_0)$, and in the second inequality, the leading constant explodes for the same limit. Furthermore, we have

$$\frac{\eta_1(c_0, \beta_0)}{\beta_0} = 2.3853845446404978.$$

Remark 2.1.16. Basically, throughout the proof of Theorem 2.1.15, the authors make heavy use of the representation of K in terms of its inverse Laplace transform, i.e.,

$$K(t) = \frac{t^{H-1/2}}{\Gamma(H + 1/2)} = c_H \int_0^\infty e^{-xt} x^{-H-1/2} dx, \quad c_H = \frac{1}{\Gamma(H + 1/2)\Gamma(1/2 - H)} \tag{2.20}$$

to obtain the estimation.

2.2. Weak Scheme for the rough Heston Model

In this section, we first introduce the weak scheme for simulating the volatility process $\boldsymbol{V}_t^N$, obtained in [BB23a]. And then we prove this scheme actually achieves (strongly potential) second-order weak convergence rate which was not established and but conjectured in [BB23a, Section 2.2].

2.2.1. Weak Scheme for the Volatility V

Note that the volatility process $\boldsymbol{V}_t^N$ of the Markovian approximation (2.5)-(2.6) does not depend on the stock price S^N. So we can first try to simulate the volatility process $\boldsymbol{V}_t^N$. Recall the SDE for weak Markovian approximation $\boldsymbol{V}_t^N$ of the volatility process is

$$dV_t^{(i)} = -x_i(V_t^{(i)} - v_0^i)dt + (\theta - \lambda V_t^N)dt + \nu\sqrt{V_t^N}dW_t,\ V_0^{(i)} = v_0^i,\ i = 1, \dots, N,$$

where $\boldsymbol{w} \coloneqq (w_i)_{i=1}^N$ is the vector of weights, $\boldsymbol{v}_0 = (\boldsymbol{v}_0)_{i=1}^N$ is any initial condition with $\boldsymbol{w} \cdot \boldsymbol{v}_0 = V_0$, $\boldsymbol{V}^N \coloneqq (V^{(i)})_{i=1}^N$, and $V_t^N = \boldsymbol{w} \cdot \boldsymbol{V}_t^N$.

Write the SDEs into vector form, we obtain

$$d\boldsymbol{V}_t^N = -\mathrm{diag}(\boldsymbol{x})\left(\boldsymbol{V}_t^N - \boldsymbol{v}_0\right)dt + \left(\theta - \lambda \boldsymbol{w}^\top \boldsymbol{V}_t^N\right)\mathbf{1}dt + \nu\sqrt{\boldsymbol{w}^\top \boldsymbol{V}_t^N}\mathbf{1}dW_t. \tag{2.21}$$

The authors used the idea in [LM20] to get a weak scheme for it, which studied weak approximations using discrete-valued random variables of the CIR process, and showed that their approximations converge weakly with second order. Because the dynamics of $\boldsymbol{V}^N$ are quite similar in nature to the dynamics of the CIR process and actually when $N = 1$, we obtain the CIR process.

We split the SDE (2.21) into two parts. We denote by $D(\boldsymbol{z}, h) \coloneqq \boldsymbol{Z}_h \coloneqq (Z_h^i)_{i=1}^N$, the solution at time h of the ODE

$$d\boldsymbol{Z}_t^N = -\mathrm{diag}(\boldsymbol{x})\left(\boldsymbol{Z}_t^N - \boldsymbol{v}_0\right)dt + \left(\theta - \lambda \boldsymbol{w}^\top \boldsymbol{Z}_t^N\right)\mathbf{1}dt,$$

and by $S(\boldsymbol{y}, h) \coloneqq \boldsymbol{Y}_h \coloneqq (Y_h^i)_{i=1}^N$ the solution at time h of the SDE

$$d\boldsymbol{Y}_t^N = \nu\sqrt{\boldsymbol{w}^\top \boldsymbol{Y}_t^N}\mathbf{1}dW_t. \tag{2.22}$$

We remark that the SDE is split into two parts mainly to overcome the effect of mean reversion in the deterministic component, which would otherwise lead to a stiffness issue when approximating the solution of the SDE (2.21).

Next, we give schemes $\widehat{D}$ and $\widehat{S}$ for approximating D and S, respectively. First, note that D satisfies a linear ODE and can hence be solved exactly. The solution is given by

$$D(\boldsymbol{z}, h) = \boldsymbol{Z}_h = e^{Ah}\boldsymbol{z} + A^{-1}(e^{Ah} - \mathrm{Id})b,$$

where

$$A := -\lambda \mathbf{1}\boldsymbol{w}^T - \mathrm{diag}(\boldsymbol{x}), \quad \text{and} \quad b := \theta \mathbf{1} + \mathrm{diag}(\boldsymbol{x})\boldsymbol{v}_0,$$

and we denote $\mathbf{1} := (1, \dots, 1)^T \in \mathbb{R}^N$, $\mathrm{Id} \in \mathbb{R}^{N \times N}$ the identity matrix, and $\mathrm{diag}(\boldsymbol{x}) \in \mathbb{R}^{N \times N}$ is the diagonal matrix with entries $\boldsymbol{x}$. Therefore, we can simply set $\widehat{D}(\boldsymbol{z}, h) := D(\boldsymbol{z}, h)$.

Next, we provide a scheme $\widehat{S}$ for approximating S. Define $\overline{w} := \mathbf{1}^T \boldsymbol{w}$. The strategy is to consider the inner product of the SDE (2.22) with $\boldsymbol{w}$ to get the SDE

$$\mathrm{d}Y_t = \nu \overline{w} \sqrt{Y_t}\, \mathrm{d}W_t, \qquad Y_0 = \boldsymbol{w} \cdot \boldsymbol{y}.$$

This is exactly the SDE studied in [LM20], with $x := \boldsymbol{w} \cdot \boldsymbol{y}$ and $z := \nu^2 \overline{w}^2 h$. Likewise, we perform moment matching for Y_h to obtain a three-valued approximation, which leads to the following second-order scheme. Define the quantities

$$\begin{aligned}
m_1 &= x, \qquad m_2 = x^2 + xz, \qquad m_3 = x^3 + 3x^2 z + \frac{3}{2} x z^2, \\
p_1 &= \frac{m_1 x_2 x_3 - m_2(x_2 + x_3) + m_3}{x_1(x_3 - x_1)(x_2 - x_1)}, \\
p_2 &= \frac{m_1 x_1 x_3 - m_2(x_1 + x_3) + m_3}{x_2(x_3 - x_2)(x_1 - x_2)}, \\
p_3 &= \frac{m_1 x_1 x_2 - m_2(x_1 + x_2) + m_3}{x_3(x_1 - x_3)(x_2 - x_3)}, \\
x_1 &= x + Az - \sqrt{(3x + A^2 z)\, z}, \\
x_2 &= x + \left(A - \frac{3}{4}\right) z, \\
x_3 &= x + Az + \sqrt{(3x + A^2 z)\, z}, \\
A &= \frac{6 + \sqrt{3}}{4}.
\end{aligned} \tag{2.23}$$

We define $\widehat{Y}_h$ to be the random variable which is x_i with probability p_i, $i = 1, 2, 3$. We remark that for $i = 1, 2, 3$, $m^i = \mathbb{E} Y_h^i$ and $x_i \geq 0$, and that the x_i and p_i are chosen such that $p_1 + p_2 + p_3 = 1$ and $\mathbb{E} \widehat{Y}_h^k = \mathbb{E} Y_h^k$ for $k = 1, \dots, 5$.

Now we can get the scheme for the entire process $\boldsymbol{Y}$. Fortunately, we can easily reconstruct an approximation $\widehat{\boldsymbol{Y}}$ from $\widehat{Y}$. Indeed, we note that each component of the right-hand side of SDE (2.22) is identical for all $i = 1, \dots, N$. Hence, the solution of (2.22) should be of the form

$$Y_h^i = y^i + Q, \qquad i = 1, \dots, N, \tag{2.24}$$

for some scalar random variable Q. Taking the inner product of (2.24) with $\boldsymbol{w}$, we get

$$Y_h = \boldsymbol{w} \cdot \boldsymbol{y} + \overline{w} Q, \quad \text{implying} \quad Q = \frac{Y_h - \boldsymbol{w} \cdot \boldsymbol{y}}{\overline{w}}.$$

Hence, we set

$$\widehat{S}(\boldsymbol{y}, h) \coloneqq \widehat{\boldsymbol{Y}}_h \coloneqq \boldsymbol{y} + \frac{\widehat{Y}_h - \boldsymbol{w} \cdot \boldsymbol{y}}{\overline{w}}. \tag{2.25}$$

Finally, we combine the schemes for drift and diffusion by Strang splitting, and get

$$\widehat{A}(\boldsymbol{v}, h) \coloneqq \widehat{D}\left(\widehat{S}\left(\widehat{D}\left(\boldsymbol{v}, \frac{h}{2}\right), h\right), \frac{h}{2}\right) \tag{2.26}$$

for approximating $\boldsymbol{V}_h$ given $\boldsymbol{v}$.

Therefore, we get a simulation algorithm

$$\boldsymbol{V}^{N,M}_{t_{j+1}} \coloneqq \widehat{A}(\boldsymbol{V}^{N,M}_{t_j}, t_{j+1} - t_j), \qquad j = 0, \dots, M-1,$$

where $0 = t_0 < t_1 < \cdots < t_M = T$. But we should also notice that the scheme may not be well-defined because the occurrence of square roots in (2.23), as [BB23a, Section 2.2] pointed out. To resolve this problem, the authors of [BB23a] replace V by $V^+ := \max(V, 0)$ or $|V|$. While recently, [AJBB24, Theorem 3.1] asserted that the scheme is truly well-defined and $\boldsymbol{V}^{N,M}$ will even stay in a cone domain of the half-space $\boldsymbol{w} \cdot \boldsymbol{V} \geq 0$ as long as $\boldsymbol{w} \cdot \boldsymbol{v}_0$ stay non-negative, or more precisely $\boldsymbol{w} \cdot \widehat{D}\left(\boldsymbol{v}_0, \frac{t_1 - t_0}{2}\right) \geq 0$.

2.2.2. Analysis of Weak Error

In this subsection, we will introduce the Talay and Tubaro's method on analyzing the weak error, basically following [Tal86].

We consider a d_W-dimensional standard Brownian motion $(W_t)_{t\geq 0}$ and denote by $(\mathcal{F}_t)_{t\geq 0}$ its augmented filtration satisfying the usual conditions. Let $d \in \mathbb{N}^+$, and $\mathbb{D} \subset \mathbb{R}^d$ be a domain, assumed for simplicity to be a product of d intervals.

Notations. For any multi-index $\alpha = (\alpha_1, \dots, \alpha_d) \in \mathbb{N}^d$, define $\partial^\alpha = \partial_1^{\alpha_1} \cdots \partial_d^{\alpha_d}$ and $|\alpha| = \sum_{l=1}^d \alpha_l$. And $M_{d\times d_W}(\mathbb{R})$ denotes the space of $d \times d_W$ matrices on $\mathbb{R}$. Next we introduce the following space

$$\begin{aligned} C^\infty_{\mathrm{pol}}(\mathbb{D}) = \Big\{ f \in C^\infty(\mathbb{D}, \mathbb{R}) \,\Big|\, & \forall \alpha \in \mathbb{N}^d,\ \exists C_\alpha > 0,\ e_\alpha \in \mathbb{N}^*, \\ & \forall x \in \mathbb{D},\ |\partial^\alpha f(x)| \leq C_\alpha (1 + \|x\|^{e_\alpha}) \Big\}, \end{aligned}$$

where $\|\cdot\|$ is a norm in $\mathbb{R}^d$. And we denote $C^\infty_K(\mathbb{D}, \mathbb{R})$ the set of the C^∞ real valued functions with a compact support in $\mathbb{D}$. We say that $(C_\alpha, e_\alpha)_{\alpha\in\mathbb{N}^d}$ is a *good sequence* for f if the above bound holds.

We will write $g(x, h) = O(h^n)$ if, for some $C > 0$, $k \in \mathbb{N}$, and $h_0 > 0$,

$$|g(x, h)| \leq C(1 + |x|^k) h^n,\ x \geq 0,\ 0 < h \leq h_0.$$

For $k \in \mathbb{N}$, we denote by $C^k(\mathbb{R}\times\mathbb{R}_+)$ the space of continuous functions $f : \mathbb{R}\times\mathbb{R}_+ \to \mathbb{R}$ such that the partial derivatives $\partial_x^\alpha \partial_y^\beta f(x,y)$ exist and are continuous with respect to (x,y) for all $(\alpha,\beta) \in \mathbb{N}^2$ such that $\alpha + 2\beta \le k$. We then define for $L \in \mathbb{N}$,

$$\mathbf{f}_L(x,y) = (1 + x^{2L} + y^{2L}),\ x \in \mathbb{R}, y \in \mathbb{R}_+, \tag{2.27}$$

and introduce

$$\begin{aligned}\mathcal{C}^{k,L}_{\text{pol}}(\mathbb{R}\times\mathbb{R}_+) = \{f \in C^k(\mathbb{R}\times\mathbb{R}_+) \mid \exists C > 0 \text{ such that } \forall(\alpha,\beta) \in \mathbb{N}^2, \alpha + 2\beta \le k,\\ |\partial_x^\alpha \partial_y^\beta f(x,y)| \le C\mathbf{f}_L(x,y)\},\end{aligned} \tag{2.28}$$

the space of continuously differentiable functions up to order k with derivatives with polynomial growth of order $2L$. Note that we assume twice less differentiability on the y component. Furthermore, we set

$$\mathcal{C}^{k}_{\text{pol}}(\mathbb{R}\times\mathbb{R}_+) = \cup_{L\in\mathbb{N}} \mathcal{C}^{k,L}_{\text{pol}}(\mathbb{R}\times\mathbb{R}_+) \text{ and } \mathcal{C}^{\infty}_{\text{pol}}(\mathbb{R}\times\mathbb{R}_+) = \cap_{k\in\mathbb{N}} \mathcal{C}^{k}_{\text{pol}}(\mathbb{R}\times\mathbb{R}_+).$$

Last, we endow $\mathcal{C}^{k,L}_{\text{pol}}(\mathbb{R}\times\mathbb{R}_+)$ with the following norm

$$\|f\|_{k,L} = \sum_{\alpha+2\beta\le k} \sup_{(x,y)\in\mathbb{R}\times\mathbb{R}_+} \frac{|\partial_x^\alpha \partial_y^\beta f(x,y)|}{\mathbf{f}_L(x,y)}. \tag{2.29}$$

Assumptions. Assume that $b : \mathbb{D} \to \mathbb{R}^d$ and $\sigma : \mathbb{D} \to M_{d\times d_W}(\mathbb{R})$ are such that, for $1 \le i,j \le d$, the functions $x \mapsto b_i(x)$ and $x \mapsto (\sigma\sigma^\top)_{i,j}(x)$ belong to $C^\infty_{\text{pol}}(\mathbb{D})$.

Then for $x \in \mathbb{D}$, we consider the SDE

$$X_t^x = x + \int_0^t b(X_s^x)\,ds + \int_0^t \sigma(X_s^x)\,dW_s, \quad t \ge 0. \tag{2.30}$$

We assume that this SDE admits a unique weak solution staying in $\mathbb{D}$, i.e. $\mathbb{P}(\forall t \ge 0, X_t^x \in \mathbb{D}) = 1$, so that it satisfies the strong Markov property. The infinitesimal generator associated with the SDE is

$$Lf(x) = \sum_{i=1}^d b_i(x)\partial_i f(x) + \frac{1}{2}\sum_{i,j=1}^d \sum_{k=1}^{d_W} \sigma_{i,k}(x)\sigma_{j,k}(x)\partial_{ij} f(x). \tag{2.31}$$

We remark here that we see if $f \in C^\infty_{\text{pol}}(\mathbb{D})$, then $L^k f \in C^\infty_{\text{pol}}(\mathbb{D})$ for any $k \in \mathbb{N}$. So L is a linear operator on space $C^\infty_{\text{pol}}(\mathbb{D})$.

Definition 2.2.1. We say that the operator L satisfies the *required assumptions* on $\mathbb{D}$ if it is given by (2.31) for some b and σ satisfying all assumptions above.

Now, let us turn to the definition of discretization schemes for the SDE (2.30). Let us fix a time horizon $T > 0$. We will consider in the whole subsection the time interval $[0,T]$ and the regular time discretization $t_i^n = iT/n$ for $i = 0,1,\dots,n$.

Definition 2.2.2. A family of transition probabilities $(\widehat{p}_x(t)(dz), t > 0, x \in \mathbb{D})$ on $\mathbb{D}$ is such that $\widehat{p}_x(t)$ is a probability law on $\mathbb{D}$ for $t > 0$ and $x \in \mathbb{D}$.

A discretization scheme with transition probabilities $(\widehat{p}_x(t)(dz), t > 0, x \in \mathbb{D})$ is a sequence $(\widehat{X}^n_{t^n_i}, 0 \leq i \leq n)$ of $\mathbb{D}$-valued random variables such that

- for $0 \leq i \leq n$, $\widehat{X}^n_{t^n_i}$ is a $\mathcal{F}_{t^n_i}$-measurable random variable on $\mathbb{D}$,
- the law of $\widehat{X}^n_{t^n_{i+1}}$ is given by $\mathbb{E}[f(\widehat{X}^n_{t^n_{i+1}}) \mid \mathcal{F}_{t^n_i}] = \int_{\mathbb{D}} f(z)\widehat{p}_{\widehat{X}^n_{t^n_i}}(T/n)(dz)$ and thus only depends on $\widehat{X}^n_{t^n_i}$ and T/n.

For convenience, we will denote, for $t > 0$ and $x \in \mathbb{D}$, $\widehat{X}^x_t$ a random variable distributed according to the probability law $\widehat{p}_x(t)(dz)$ since the law of a discretization scheme $(\widehat{X}^n_{t^n_i}, 0 \leq i \leq n)$ is entirely determined by its initial value and its transition probabilities by the definition. And note that the initial value is quite always taken equal to the initial value of the SDE, we will identify with a slight abuse of language the scheme $(\widehat{X}^n_{t^n_i}, 0 \leq i \leq n)$ with its transition probabilities ($\widehat{p}_x(t)(dz)$ or $\widehat{X}^x_t$).

Definition 2.2.3. Let $x \in \mathbb{D}$. A discretization scheme $(\widehat{X}^n_{t^n_i}, 0 \leq i \leq n)$ is a *weak ν-th-order scheme* for the SDE $(X^x_t, t \in [0,T])$ if

$$\forall f \in C^\infty_K(\mathbb{D}, \mathbb{R}),\ \exists K > 0,\ \left|\mathbb{E}(f(X^x_T)) - \mathbb{E}(f(\widehat{X}^n_{t^n_n}))\right| \leq K/n^\nu.$$

The quantity $\mathbb{E}(f(X^x_T)) - \mathbb{E}(f(\widehat{X}^n_{t^n_n}))$ is called the *weak error* associated to f.

Definition 2.2.4. A discretization scheme $(\widehat{X}^n_{t^n_i},\ 0 \leq i \leq n)$ has *uniformly bounded moments* if one has

$$\exists n_0 \in \mathbb{N}^+,\ \forall q \in \mathbb{N}^+, \qquad \sup_{n \geq n_0,\, 0 \leq i \leq n} \mathbb{E}\left[\|\widehat{X}^n_{t^n_i}\|^q\right] < \infty.$$

The proofs of most of the propositions or theorems in what follows are deferred to Appendix A.1.

Proposition 2.2.5. *Let us suppose that there is $\eta > 0$ such that for $t \in (0, \eta)$,*

$$\forall q \in \mathbb{N}^*,\ \exists C_q > 0,\ \forall x \in \mathbb{D},\ \mathbb{E}\left[\|\widehat{X}^x_t\|^q\right] \leq \|x\|^q(1 + C_q t) + C_q t. \tag{2.32}$$

Then, the discretization scheme has uniformly bounded moments.

Definition 2.2.6. Let us consider a mapping $f \in C^\infty_{\text{pol}}(\mathbb{D}) \mapsto Rf$ such that $Rf : \mathbb{R}_+ \times \mathbb{D} \to \mathbb{R}$. It is a *remainder of order $\nu \in \mathbb{N}$* if for any function $f \in C^\infty_{\text{pol}}(\mathbb{D})$,

$$\forall t \in (0, \eta),\ \forall x \in \mathbb{D},\ |Rf(t,x)| = O(t^\nu).$$

Definition 2.2.7. For any scheme $(\widehat{p}_x(t)(dz),\, t > 0,\, x \in \mathbb{D})$ we define

$$\forall f \in C^\infty, \quad R^{\widehat{p}(t)}_{\nu+1} f(x) = \mathbb{E}\big[f(\widehat{X}^x_t)\big] - \left[f(x) + \sum_{k=1}^{\nu} \frac{1}{k!} t^k L^k f(x)\right]$$

as soon as $\mathbb{E}\big[|f(\widehat{X}^x_t)|\big] < \infty$.

We will say that $\widehat{p}_x(t)(dz)$ is a *potential weak ν-th-order scheme* for the operator L if $R^{\widehat{p}(t)}_{\nu+1} f(x)$ is defined for $f \in C^\infty_{\text{pol}}(\mathbb{D})$ and $t > 0$, and is a remainder of order $\nu + 1$.

Remark 2.2.8. Iterating the Dynkin's formula $\mathbb{E}f(X^x_{t+h}) = f(x) + \int_0^h \mathbb{E}Lf(X^x_{t+s})\, ds$, we have

$$\begin{aligned}\mathbb{E}f(X^x_h) = f(x) &+ \sum_{k=1}^{\nu} \frac{L^k f(x)}{k!} h^k \\ &+ \int_0^h \int_0^{s_1} \cdots \int_0^{s_\nu} \mathbb{E}L^{\nu+1} f(X^x_{s_{\nu+1}})\, ds_{\nu+1} \cdots ds_2 ds_1,\end{aligned}$$

which motivates the Definition 2.2.7 above.

Proposition 2.2.9. *Let $b : \mathbb{D} \to \mathbb{R}^d$ and $\sigma : \mathbb{D} \to M_{d\times d_W}(\mathbb{R})$ such that $\|b(x)\| + \|\sigma(x)\| \le C(1 + \|x\|)$ for some $C > 0$, and assume that the associated operator L satisfies the required assumption on $\mathbb{D}$. Then, for any $\nu \in \mathbb{N}$, the exact scheme is a potential weak νth-order scheme for L.*

We introduce [Tal86, Theorem 1.9] as follows, which establishes the link between a *weak ν-th order scheme* and a *potential weak ν-th order scheme.*

Theorem 2.2.10 (Talay–Tubaro)**.** *Let us consider an operator L that satisfies the required assumptions on $\mathbb{D}$ and a discretization scheme $(\widehat{X}^n_{t_i},\, 0 \le i \le n)$ with transition probabilities $\widehat{p}_h(t)(dz)$ on $\mathbb{D}$ that starts from $\widehat{X}^n_{t_0} = x \in \mathbb{D}$. If*

(i) the scheme has uniformly bounded moments and is a potential weak νth-order discretization scheme *for the operator L.*

(ii) $f : \mathbb{D} \to \mathbb{R}$ is a function such that $u(t,x) = \mathbb{E}\left[f(X^{t,x}_t)\right]$ is defined on $[0,T] \times \mathbb{D}$, C^∞, solves $\forall t \in [0,T], \forall x \in \mathbb{D}, \partial_t u(t,x) = -Lu(t,x)$, and satisfies:

$$\forall l \in \mathbb{N}, \alpha \in \mathbb{N}^d, \exists C_{l,\alpha}, e_{l,\alpha} > 0, \forall x \in \mathbb{D}, t \in [0,T], \left|\partial^\alpha_x \partial^l_t u(t,x)\right| \le C_{l,\alpha}\left(1 + \|x\|^{e_{l,\alpha}}\right).$$

Then, there exist $K > 0$ and $n_0 \in \mathbb{N}$ such that $\left|\mathbb{E}[f(\widehat{X}^n_{t_n})] - \mathbb{E}[f(X^x_T)]\right| \le K/n^\nu$ for all $n \ge n_0$ and $f \in C^\infty_{pol}(\mathbb{D})$.

Remark 2.2.11. (i) When $\mathbb{D} = \mathbb{R}^d$, if the diffusion coefficients b and σ are both smooth with bound derivatives, then the condition (ii) will hold automatically, which is shown in [Tal86].

(ii) We say that a potential νth-order weak approximation is a strongly potential νth-order weak approximation if it has uniformly bounded moments, i.e., condition (i) holds. Typically, a strongly potential νth-order discretization is a νth-order weak approximation, at least, we do not know any counterexample.

Definition 2.2.12. Let us consider two transition probabilities $\widehat{p}_x^1(t)(dz)$ and $\widehat{p}_x^2(t)(dz)$ on $\mathbb{D}$. Then, we define the composition $\widehat{p}^2(t_2) \circ \widehat{p}_x^1(t_1)(dz)$ simply as

$$\widehat{p}^2(t_2) \circ \widehat{p}_x^1(t_1)(dz) = \int_{\mathbb{D}} \widehat{p}_y^1(t_2)(dz)\, \widehat{p}_x^1(t_1)(dy).$$

This amounts to first use the scheme 1 with a time step t_1 and then the scheme 2 with a time step t_2 with independent samples. We name $\widehat{X}_{t_2,t_1}^{2\circ 1,x} = \widehat{X}_{t_2}^{2,\widehat{X}_{t_1}^{1,x}}$ a random variable with the law $\widehat{p}^2(t_2) \circ \widehat{p}_x^1(t_1)(dz)$. More generally, if one has m transition probabilities $\widehat{p}_x^1, \dots, \widehat{p}_x^m$ on $\mathbb{D}$, we define $\widehat{p}^m(t_m) \circ \dots \circ \widehat{p}_x^1(t_1)(dz)$ as the composition of $\widehat{p}^{m-1}(t_{m-1}) \circ \dots \circ \widehat{p}_x^1(t_1)(dz)$ and then $\widehat{p}_x^m(t_m)$.

Remark 2.2.13. The criterion (2.32) that ensures the uniform boundedness of the moments is easy to use with the scheme composition. Indeed, let us fix $\lambda_1, \lambda_2 > 0$. One checks easily that if $\widehat{p}_x^1(t)(dz)$ and $\widehat{p}_x^2(t)(dz)$ satisfy (2.32), then $\widehat{p}^2(\lambda t) \circ \widehat{p}_x^1(\lambda t)(dz)$ satisfies also (2.32) and has thus uniformly bounded moments.

Proposition 2.2.14. *Let L_1 and L_2 be two operators that satisfy the required assumptions on $\mathbb{D}$, and assume that $\widehat{p}_x^1(t)(dz)$ and $\widehat{p}_x^2(t)(dz)$ are respectively potential weak ν-th-order discretization schemes on $\mathbb{D}$ for these operators. Then, for $\lambda_1, \lambda_2 > 0$, $\widehat{p}^2(\lambda_2 t) \circ \widehat{p}_x^1(\lambda_1 t)(dz)$ is such that for $f \in C_{pol}^\infty(\mathbb{D})$,*

$$\mathbb{E}\big[f(\widehat{X}_{\lambda_2 t,\lambda_1 t}^{2\circ 1,x})\big] = \sum_{l_1+l_2\le\nu} \frac{\lambda_1^{l_1}\lambda_2^{l_2}}{l_1!l_2!} t^{l_1+l_2} L_1^{l_1} L_2^{l_2} f(x) + R^{\widehat{p}^2(\lambda_2 t)\circ\widehat{p}^1(\lambda_1 t)} f(x)$$

where $R^{\widehat{p}^2(\lambda_2 t)\circ\widehat{p}^1(\lambda_1 t)} f(x)$ is a remainder of order $\nu+1$.

Corollary 2.2.15. *If $L_1L_2 = L_2L_1$, then $\widehat{p}^2(t) \circ \widehat{p}_x^1(t)$ is a potential weak ν-th order scheme for $L_1 + L_2$.*

The following theorem, taken from [Alf10b, Theorem 1.17], states that the composition of several potential second-order schemes remains potential second order.

Theorem 2.2.16. *Let $L_1, \dots, L_m$ be m operators that satisfy the required assumption on $\mathbb{D}$. Let us consider $\widehat{p}_x^1, \dots, \widehat{p}_x^m$ m potential second order discretization schemes on $\mathbb{D}$ for the operators $L_1, \dots, L_m$. Then, the transition probabilities*

$$\widehat{p}^m(t/2) \circ \dots \circ \widehat{p}^2(t/2) \circ \widehat{p}^1(t) \circ \widehat{p}^2(t/2) \circ \dots \circ \widehat{p}^m(t/2) \tag{2.33}$$

and

$$\frac{1}{2}\big(\widehat{p}^m(t) \circ \dots \circ \widehat{p}^2(t) \circ \widehat{p}_x^1(t) + \widehat{p}^1(t) \circ \widehat{p}^2(t) \circ \dots \circ \widehat{p}^m(t)\big) \tag{2.34}$$

are potential second order discretization schemes for the operator $\Sigma L = L_1 + L_2 + \dots + L_m$.

2.2.3. Proof of Second Order Weak Rate

Let $\mathbb{D} = \mathbb{R}^N$, $m = 2$. Since we apply the splitting method (2.33) in Theorem 2.2.16 to construct the scheme A, and $\widehat{D}$ is the exact scheme, which can attain arbitrary potential weak order by Proposition 2.2.9, it suffices to show that $\widehat{S}$ is indeed a potential second-order weak scheme and that the scheme A has uniformly bounded moments. We begin by deriving the generator of the SDE

$$d\boldsymbol{Y}_t^N = \nu\sqrt{\boldsymbol{w}^\top \mathbf{Y}_t^N}\;\mathbf{1}\;dW_t. \tag{2.35}$$

Apply Itô's formula on $f(\mathbf{Y}_t^N)$ for $f \in C^\infty(\mathbb{R}^N)$, we obtain

$$\begin{aligned}\mathbb{E}\big(f(\mathbf{Y}_t^N)\big) &= f(\boldsymbol{y}) + \mathbb{E}\left[\sum_{i=1}^N \int_0^t \partial_i f(\mathbf{Y}_s^N)\,dY_s^i\right] + \frac{1}{2}\mathbb{E}\left[\sum_{i,j=1}^N \int_0^t \partial_i\partial_j f(\mathbf{Y}_s^N)\,d\langle Y^i, Y^j\rangle_s\right]\\ &= f(\boldsymbol{y}) + \frac{1}{2}\mathbb{E}\left[\sum_{i,j=1}^N \int_0^t \nu^2\,\boldsymbol{w}^\top \mathbf{Y}_s^N\,\partial_i\partial_j f(\mathbf{Y}_s^N)\,ds\right].\end{aligned}$$

So we can define the generator L for $\boldsymbol{Y}^N$ by

$$Lf(\boldsymbol{x}) = \frac{\nu^2}{2}\sum_{i,j}(\boldsymbol{w}^\top\boldsymbol{x})\partial_i\partial_j f(\boldsymbol{x}).$$

Thus, we have

$$L^2 f(\boldsymbol{x}) = \frac{\nu^4}{4}\Big[\sum_{i,j,k,l}(\boldsymbol{w}^\top\boldsymbol{x})^2\partial_k\partial_l\partial_i\partial_j f(\boldsymbol{x}) + 2\sum_{i,j,l}\overline{w}(\boldsymbol{w}^\top\boldsymbol{x})\partial_l\partial_i\partial_j f(\boldsymbol{x})\Big].$$

Apply Taylor expansion formula on $f(\widehat{\mathbf{Y}}_t^N)$ for $f \in C^\infty(\mathbb{R}^N)$, recall that all $\widehat{Y}^i$ share the same stochastic part, i.e. $\widehat{Y}_t^i - y^i = \widehat{Q}$ where $\widehat{Q} = \frac{\widehat{Y}_t - \boldsymbol{w}\cdot\boldsymbol{y}}{\overline{w}}$, we have

$$\mathbb{E}\big(f(\widehat{\mathbf{Y}}_t^N)\big) = f(\boldsymbol{y}) + \sum_{|\alpha|\le 5}\frac{1}{|\alpha|}\mathbb{E}(\widehat{Q}^{|\alpha|})\partial^\alpha f(\boldsymbol{y}) + \frac{1}{5!}\sum_{|\alpha|=5}\sum_i \int_{y^i}^{Y_t^i}\partial_i\partial^\alpha f(y)\mathbb{E}(\widehat{Q}^5)dy, \tag{2.36}$$

$$\frac{1}{5!}\sum_{|\alpha|=5}\sum_i \int_{y^i}^{Y_t^i}\partial_i\partial^\alpha f(y)\mathbb{E}(\widehat{Q}^5)dy \le \frac{1}{6!}\sum_{|\alpha|=5}\sum_i \mathbb{E}\big(\sup_{y^i\le y\le Y_t^i}|\partial_i\partial^\alpha f(y)|\widehat{Q}^6\big).$$

Recall $\widehat{Q}$ also satisfies the equations below from [LM20]

$$
\begin{aligned}
\mathbb{E}(\widehat{Q}) &= O(t^3), \\
\mathbb{E}(\widehat{Q}^2) &= \frac{xz}{\overline{w}^2}, \\
\mathbb{E}(\widehat{Q}^3) &= \frac{3xz^2}{2\,\overline{w}^3} + O(t^3), \\
\mathbb{E}(\widehat{Q}^4) &= \frac{3x^2z^2}{\overline{w}^4} + O(t^3), \\
\mathbb{E}(\widehat{Q}^5) &= O(t^3), \\
\sum_{|\alpha|=5}\sum_i \mathbb{E}\Big[\sup_{y^i\le y\le Y_t^i}\big|\partial_i\partial^\alpha f(y)\big|\,\widehat{Q}^6\Big] & \\
&= O(t^3).
\end{aligned}
\tag{2.37}
$$

Next we check whether the first two order terms of the residual $R_3^{\widehat{S}}f(\boldsymbol{y})$ for $f\in C^\infty_{\text{pol}}$ can be erased so that it will have local third order, where $R_3^{\widehat{S}}f(\boldsymbol{y})$ is defined by

$$
R_3^{\widehat{S}}f(\boldsymbol{y}) := \mathbb{E}(f(\widehat{\mathbf{Y}}_t^N)) - [f(\boldsymbol{y}) + tLf(\boldsymbol{y}) + \frac{t^2}{2}L^2 f(\boldsymbol{y})].
$$

Insert the Taylor expansion (2.36) to the definition of $R_3^{\widehat{S}}f$, we have

$$
\begin{aligned}
R_3^{\widehat{S}}f(\boldsymbol{y}) = &\sum_{i,j}\partial_i\partial_j f(\boldsymbol{y})\left(\frac{1}{2!}\,\mathbb{E}(\widehat{Q}^2) - \frac{\nu^2}{2}\,xt\right) \\
&+ \sum_{i,j,l}\partial_i\partial_j\partial_l f(\boldsymbol{y})\left(\frac{1}{3!}\,\mathbb{E}(\widehat{Q}^3) - \frac{\nu^4}{4}\cdot 2\,\overline{w}x\cdot\frac{t^2}{2}\right) \\
&+ \sum_{i,j,k,l}\partial_i\partial_j\partial_k\partial_l f(\boldsymbol{y})\left(\frac{1}{4!}\,\mathbb{E}(\widehat{Q}^4) - \frac{\nu^4}{4}\cdot\frac{t^2}{2}\cdot x^2\right) \\
&+ O(t^3) = O(t^3),
\end{aligned}
$$

where the last equation follows from the equations (2.37) for moments of $\widehat{Q}$. So we truly have local potential third order weak rate.

We can further extend this result to a more general model, which is stated in the following lemma.

Lemma 2.2.17. *Consider the $\widehat{\boldsymbol{Y}}$ is obtained by the same way as in (2.25), which is a weak scheme for the solution of a more general SDE*

$$
d\mathbf{Y}_t^N = \eta\sqrt{\sigma(\boldsymbol{w}^\top\mathbf{Y}_t^N)}\;\mathbf{1}\;dW_t, \tag{2.38}
$$

where $\sigma:\mathbb{R}\to\mathbb{R}^{\ge 0}$ is a C^2 function. And we assume $\widehat{Y}$ in the equation (2.25) is a potential second order approximation for the projected SDE

$$
dY_t = \eta\sqrt{\sigma(Y_t)}\;\mathbf{1}\;dW_t, \tag{2.39}
$$

where $Y_t = \boldsymbol{w}^\top\cdot\mathbf{Y}_t^N$.

Then we have $\widehat{\boldsymbol{Y}}$ is a scheme of potential second order convergence rate for $\boldsymbol{Y}^N$.

Proof. Recall the scheme $\widehat{Y}$ is obtained by the following strategy: we consider the inner product of the SDE (2.38) with $\boldsymbol{w}$ to get (2.39). So the generator L for the SDE (2.39) is

$$Lf(x) = \frac{1}{2}(\bar{w}\eta)^2\sigma(x)f''(x)$$

with

$$Lf(x) = \frac{1}{4}(\bar{w}\eta)^4\sigma(x)^2 f^{(4)}(x) + \frac{1}{2}(\bar{w}\eta)^4\sigma(x)\sigma'(x)f'''(x) + \frac{1}{4}(\bar{w}\eta)^4\sigma(x)\sigma''(x)f''(x).$$

Since $\widehat{Y}$ is our potentially second order approximation starting at $y = \boldsymbol{w}\cdot\boldsymbol{y}$, it should satisfy: for $f \in C^\infty_{\mathrm{pol}}$,

$$R_3 f(y) = \mathbb{E}(f(\widehat{Y}_h)) - [f(y) + hLf(y) + \frac{h^2}{2}L^2 f(y)] = O(h^3).$$

By Taylor expansion of $f(\widehat{Y}_h)$ at y, we have

$$\begin{aligned} R_3 f(y) =& f'(y)\mathbb{E}(\widehat{Y}_h - y) + \frac{f''(y)}{2}[\mathbb{E}[(\widehat{Y}_h - y)^2] - (\bar{w}\eta)^2 h\sigma(y) - \frac{1}{4}(\bar{w}\eta)^4 h^2\sigma(y)\sigma''(y)] \\ &+ \frac{f'''(y)}{6}[\mathbb{E}[(\widehat{Y}_h - y)^3] - 3(\bar{w}\eta)^4 h^2\sigma'(y)\sigma(y)] \\ &+ \frac{f''''(y)}{4!}[\mathbb{E}[(\widehat{Y}_h - y)^4] - 6(\bar{w}\eta)^4 h^2\sigma^2(y)] \\ &+ \frac{f^{(5)}(y)}{5!}\mathbb{E}[(\widehat{Y}_h - y)^5] + r(y,h), \end{aligned} \tag{2.40}$$

where

$$|r(y,h)| = \frac{1}{5!}|\mathbb{E}[\int_y^{\widehat{Y}_h} f^{(6)}(s)(\widehat{Y}_h - s)^5 ds]| \le \frac{1}{6!}\mathbb{E}[\max_{0\le s\le \widehat{Y}_h}|f^{(6)}(s)|(\widehat{Y}_h - s)^6].$$

So $\widehat{Y}$ should satisfy

$$\begin{aligned} &\mathbb{E}(\widehat{Y}_h - y) = O(h^3), \\ &\mathbb{E}[(\widehat{Y}_h - y)^2] = (\bar{w}\eta)^2 h\sigma(y) + \frac{1}{4}(\bar{w}\eta)^4 h^2\sigma(y)\sigma''(y) + O(h^3), \\ &\mathbb{E}[(\widehat{Y}_h - y)^3] = 3(\bar{w}\eta)^4 h^2\sigma'(y)\sigma(y) + O(h^3), \\ &\mathbb{E}[(\widehat{Y}_h - y)^4] = 6(\bar{w}\eta)^4 h^2\sigma^2(y) + O(h^3), \\ &\mathbb{E}[(\widehat{Y}_h - y)^5] = O(h^3), \\ &\mathbb{E}[(\widehat{Y}_h - y)^6] = O(h^3). \end{aligned} \tag{2.41}$$

Recall the definition of $\widehat{\boldsymbol{Y}}$, which is given by

$$\widehat{\boldsymbol{Y}} := \boldsymbol{y} + \frac{\widehat{Y}_h - \boldsymbol{w}\cdot\boldsymbol{y}}{\overline{w}}.$$

Let $y = \boldsymbol{w} \cdot \boldsymbol{y}$, we go back to the SDE (2.38) which has generator L defined by

$$Lf(\boldsymbol{y}) = \frac{\eta^2}{2} \sum_{i,j} \sigma(y) \partial_i \partial_j f(\boldsymbol{y})$$

with

$$\begin{aligned} L^2 f(\boldsymbol{y}) =& \frac{\eta^4}{4} [\sum_{i,j,k,l} \sigma^2(y) \partial_i \partial_j \partial_k \partial_l f(\boldsymbol{y}) + 2 \sum_{i,j,k} \bar{w} \sigma(y) \sigma'(y) \partial_i \partial_j \partial_k f(\boldsymbol{y}) \\ &+ \sum_{i,j} \bar{w}^2 \sigma(y) \sigma''(y) \partial_i \partial_j f(\boldsymbol{y})]. \end{aligned} \tag{2.42}$$

Therefore, we only need to check $\widehat{\boldsymbol{Y}}$ will satisfy

$$R_3 f(\boldsymbol{y}) := \mathbb{E}(f(\widehat{\boldsymbol{Y}}_h)) - [f(\boldsymbol{y}) + hLf(\boldsymbol{y}) + \frac{h^2}{2} L^2 f(\boldsymbol{y})] = O(h^3).$$

By Taylor expansion, it's equivalent with

$$\begin{aligned} R_3 f(\boldsymbol{y}) = & \sum_{i,j} \partial_i \partial_j f(\boldsymbol{y}) \left(\frac{1}{2!} \mathbb{E}(\widehat{Q}^2) - \frac{\eta^2}{2} h \sigma(y) - \frac{\eta^4}{4} \bar{w}^2 h^2 \sigma(y) \sigma''(y) \right) \\ &+ \sum_{i,j,l} \partial_i \partial_j \partial_l f(\boldsymbol{y}) \left(\frac{1}{3!} \mathbb{E}(\widehat{Q}^3) - \frac{\eta^4}{2} \bar{w} h^2 \sigma(y) \sigma'(y) \right) \\ &+ \sum_{i,j,k,l} \partial_i \partial_j \partial_k \partial_l f(\boldsymbol{y}) \left(\frac{1}{4!} \mathbb{E}(\widehat{Q}^4) - \frac{\eta^4}{4} h^2 \sigma^2(y) \right) \\ &+ O(h^3) = O(h^3), \end{aligned} \tag{2.43}$$

where $\widehat{Q} := \frac{\widehat{Y}_h - \boldsymbol{w} \cdot \boldsymbol{y}}{\bar{w}}$ satisfying

$$\begin{aligned} &\mathbb{E}(\widehat{Q}) = O(h^3), \\ &\mathbb{E}[(\widehat{Q})^2] = \eta^2 h \sigma(y) + \frac{1}{4} \bar{w}^2 \eta^4 h^2 \sigma(y) \sigma''(y) + O(h^3), \\ &\mathbb{E}[(\widehat{Q})^3] = 3 \bar{w} \eta^4 h^2 \sigma'(y) \sigma(y) + O(h^3), \\ &\mathbb{E}[(\widehat{Q})^4] = 6 \eta^4 h^2 \sigma^2(y) + O(h^3), \\ &\mathbb{E}[(\widehat{Q})^5] = O(h^3), \\ &\mathbb{E}[(\widehat{Q})^6] = O(h^3). \end{aligned} \tag{2.44}$$

It's easy to see we get the conclusion after inserting (2.44) to the equation (2.43). $\square$

For the part of proving $\widehat{A}$ has uniform bounded moments: for any $p \geq 1$, we repeatedly use c_p-inequality and Minkovski's inequality

$$\begin{aligned}
\sup_{0\le h\le T}\mathbb{E}[\|\widehat{A}(\boldsymbol{v},h)\|^p] &= \sup_{0\le h\le T}\mathbb{E}[\|e^{\frac{Ah}{2}}\widehat{S}\left(\widehat{D}\left(\boldsymbol{v},\frac{h}{2}\right),h\right)+A^{-1}(e^{\frac{Ah}{2}}-\mathrm{Id})b\|^p] \\
&\le c_{p,T,w,x,N}\sup_{0\le h\le T}\mathbb{E}[\|\widehat{S}\left(\widehat{D}\left(\boldsymbol{v},\frac{h}{2}\right),h\right)\|^p]+c'_{p,T,w,x,N} \\
&\le c_{p,T,w,x,N}\sup_{0\le h\le T}\sum_{i=1}^{3}\|x_i^{\boldsymbol{w}\cdot\widehat{D}\left(\boldsymbol{v},\frac{h}{2}\right)}\|^p \\
&\quad + c''_{p,T,w,x,N}\sup_{0\le h\le T}\|\widehat{D}\left(\boldsymbol{v},\frac{h}{2}\right)\|^p+c'_{p,T,w,x,N} \\
&\le c_{p,T,w,x,N}\sup_{0\le h\le T}\|\boldsymbol{w}\cdot\widehat{D}\left(\boldsymbol{v},\frac{h}{2}\right)\|^p+c'_{p,T,w,x,N} \\
&\le c_{p,T,w,x,N}\|v\|^p+c'_{p,T,w,x,N}<\infty.
\end{aligned} \tag{2.45}$$

Here, $x_i^{\boldsymbol{w}\cdot\widehat{D}\left(\boldsymbol{v},\frac{h}{2}\right)}$ denotes that the choice of x_i in the moment-matching depends on the quantity $\boldsymbol{w}\cdot\widehat{D}\left(\boldsymbol{v},\frac{h}{2}\right)$. These inequalities hold because each $\widehat{Y}^i = y^i+\widehat{Q}$ shares the same random component

$$\widehat{Q}=\frac{\widehat{Y}_t-\boldsymbol{w}\cdot\boldsymbol{y}}{\overline{w}}$$

where $\widehat{Y}_t$ is a three-valued random variable taking values such that

$$|x_i|\le c_w(|x|+t),$$

for $i=1,2,3$, which follows from the AM–GM inequality.

Finally, we show that the scheme $\widehat{A}$ has uniformly bounded moments. Alternatively, note that $\widehat{D}$ satisfies the criterion in Proposition 2.2.5. By adapting the argument in the last paragraph of the proof of [LM20, Theorem 2], one can verify that $\widehat{S}$ also satisfies the criterion in Proposition 2.2.5. Therefore, by Remark 2.2.13, the composition scheme $\widehat{A}$ has uniformly bounded moments. Consequently, $\widehat{A}$ is a strongly potential second-order scheme.

However, to rigorously establish that the scheme $\widehat{A}$ achieves second-order weak convergence, it is still necessary to prove that the rough Heston model admits polynomial estimates for the corresponding Cauchy problem (see item (ii) of Theorem 2.2.10). This is a highly challenging task. Although such estimates can be obtained for the classical log-Heston model (see [BCT21]), the proof strategy involves first deriving a representation formula for the derivatives $\partial_x^\alpha\partial_t u$ and then estimating this formula to verify that the derivatives of u indeed satisfy polynomial growth conditions.

We remark that in Chapter 4, we will investigate whether the numerical scheme $\widehat{A}$ exhibits the expected second-order weak convergence in practice.

2.2.4. Weak Scheme for the Stock Price S

Having constructed the scheme $\widehat{A}$ for simulating the Markovian approximations $\boldsymbol{V}^N$, we now turn to deriving a scheme for simulating the stock price S. The main challenge is that the Brownian motion driving the stock price is correlated with the Brownian motion driving the volatility, whereas the scheme $\widehat{A}$ does not require simulating increments of W. To address this, we introduce the weak scheme proposed in [BB23a], which follows the strategy of [Alf10a] to construct an appropriate scheme for S.

Define the processes

$$Y_t^i = \int_0^t V_t^i dt, \qquad i = 1, \dots, N.$$

And note that the solution S to the SDE (2.5) is given by

$$S_t = S_0 \exp\left(\int_0^t \sqrt{V_s^N}\left(\rho dW_s + \sqrt{1-\rho^2} dB_s\right) - \frac{1}{2}\int_0^t V_s^N ds\right),$$

where the latter term in the exponent is precisely $-\boldsymbol{w}^\top \cdot \boldsymbol{Y}/2$. So we can surmise the weak scheme $\widehat{S}$ for the stock price S may depend on both $\widehat{V}^N$ and $\widehat{Y}$.

Now, the SDE $S^{\text{rHeston}}((s, \boldsymbol{v}, \boldsymbol{y}), h)$ is given by

$$\begin{aligned}
dS_t &= \sqrt{V_t^N} S_t \left(\rho dW_t + \sqrt{1-\rho^2} dB_t\right), \qquad S_0 = s,\\
dV_t^i &= -x_i(V_t^i - v_0^i)dt + (\theta - \lambda V_t^N)dt + \nu\sqrt{V_t^N} dW_t, \qquad V_t^i = v^i,\\
dY_t^i &= V_t^i dt, \qquad Y_0^i = y^i,
\end{aligned}$$

for $i = 1, \dots, N$, where $V_t^N = w^\top \cdot \boldsymbol{V}_t^N$. It can again be split into two parts, namely the SDE $S^W((s, \boldsymbol{v}, \boldsymbol{y}), h)$ given by

$$\begin{aligned}
dS_t &= \sqrt{V_t^N} S_t \rho dW_t, \qquad S_0 = s,\\
dV_t^i &= -x_i(V_t^i - v_0^i)dt + (\theta - \lambda V_t^N)dt + \nu\sqrt{V_t^N} dW_t, \qquad V_t^i = v^i,\\
dY_t^i &= V_t dt, \qquad Y_0^i = y^i,
\end{aligned}$$

and the SDE $S^B((s, \boldsymbol{v}, \boldsymbol{y}), h)$ given by

$$\begin{aligned}
dS_t &= \sqrt{V_t^N} S_t \sqrt{1-\rho^2} dB_t, \qquad S_0 = s,\\
dV_t^i &= 0, \qquad V_t^i = v^i,\\
dY_t^i &= 0, \qquad Y_0^i = y^i.
\end{aligned}$$

The SDE S^B can be solved explicitly. Indeed, since the volatility $\boldsymbol{V}^N$ is constant, this is essentially the Black-Scholes model. The solution is given by

$$\begin{aligned}
S_h &= s\exp\left(\sqrt{v}\sqrt{1-\rho^2} B_h - \frac{1}{2} v(1-\rho^2)h\right),\\
\boldsymbol{V}_h^{\boldsymbol{N}} &= \boldsymbol{v},\\
\boldsymbol{Y}_h &= \boldsymbol{y}.
\end{aligned}$$

For the SDE S^W, we first approximate $\widehat{\boldsymbol{V}}_h^N := \widehat{A}(\boldsymbol{v}, h)$. Then we approximate $\boldsymbol{Y}$ using the trapezoidal rule, i.e. $\widehat{\boldsymbol{Y}}_h := \boldsymbol{y} + (\boldsymbol{v} + \widehat{\boldsymbol{V}}_h)h/2$. We want to express S_h in terms of $\boldsymbol{V}_h^N$ and $\boldsymbol{Y}_h$. We make the Ansatz

$$S_t = s \exp\left(at + \sum_{i=1}^N b_i(Y_t^i - Y_0^i) + \sum_{i=1}^N c_i(V_t^i - V_0^i)\right). \tag{2.46}$$

Using the Itô formula, we get

$$\begin{aligned} dS_t &= S_t\left[adt + \sum_{i=1}^N b_i dY_t^i + \sum_{i=1}^N c_i dV_t^i + \frac{1}{2}\sum_{i,j=1}^N c_i c_j d[V^i, V^j]_t\right] \\ &= S_t\Bigg[\left(a + \sum_{i=1}^N c_i x_i v_0^i + \theta\sum_{i=1}^N c_i\right)dt \\ &\quad + \sum_{i=1}^N\left(b_i - c_i x_i - \lambda w_i \sum_{j=1}^N c_j + \frac{1}{2}\nu^2 w_i \sum_{j,k=1}^N c_j c_k\right)V_i dt + \nu\sum_{i=1}^N c_i\sqrt{V_t}dW_t\Bigg]. \end{aligned}$$

To retain the SDE S^W, we thus require that

$$\begin{aligned} a + \sum_{i=1}^N c_i x_i v_0^i + \theta\sum_{i=1}^N c_i &= 0, \\ b_i - c_i x_i - \lambda w_i\sum_{j=1}^N c_j + \frac{1}{2}\nu^2 w_i\sum_{j,k=1}^N c_j c_k &= 0, \qquad i = 1,\dots,N, \\ \nu\sum_{i=1}^N c_i &= \rho. \end{aligned}$$

This is achieved with

$$\begin{aligned} a &= -\sum_{i=1}^N c_i x_i v_0^i - \theta\sum_{i=1}^N c_i, \\ b_i &= c_i x_i + \lambda w_i\sum_{j=1}^N c_j - \frac{1}{2}\nu^2 w_i\sum_{j,k=1}^N c_j c_k, \qquad i = 1,\dots,N, \\ \sum_{i=1}^N c_i &= \frac{\rho}{\nu}. \end{aligned}$$

Note that the last equation has many solutions. We want to choose the (c_i) to be such that the representation (3.151) becomes as numerically stable as possible. In the paper [BB23a], the authors tested several different choices of (c_i), and got a relatively good choice for (c_i) and identified a relatively effective one – particularly for larger mean-reversions $\boldsymbol{x}$, which are given by

$$c_1 = \frac{\rho}{\nu}, \qquad c_i = 0, \qquad i = 2,\dots,N. \tag{2.47}$$

This further yields

$$a = -\left(x_1 v_0^1 + \theta\right) \frac{\rho}{\nu}, \qquad b_i = \frac{\rho}{\nu} x_1 \delta_{i,1} + \lambda w_i \frac{\rho}{\nu} - \frac{1}{2} w_i \rho^2.$$

Therefore, we get the solution

$$\begin{aligned} S_t = s \exp\Bigg(\frac{\rho}{\nu}\Bigg(&- \left(x_1 v_0^1 + \theta\right) t + x_1 (Y_t^1 - Y_0^1) \\ &+ \left(\lambda - \frac{1}{2}\rho\nu\right)(Y_t - Y_0) + (V_t^1 - V_0^1)\Bigg)\Bigg), \end{aligned}$$

where $Y_t := \boldsymbol{w} \cdot \boldsymbol{Y}$.

Hence, we may solve the SDE S^W approximately with the algorithm $\widehat{S}^W$ given by

$$\begin{aligned} \widehat{\boldsymbol{V}}_h^N &= \widehat{A}(\boldsymbol{v}, h), \\ \widehat{\boldsymbol{Y}}_h &= \boldsymbol{y} + (\boldsymbol{v} + \widehat{\boldsymbol{V}}_h)\frac{h}{2}, \\ \widehat{S}_h^W &= s \exp\Bigg(\frac{\rho}{\nu}\Bigg(- \left(x_1 v_0^1 + \theta\right) h + x_1 (\widehat{Y}_h^1 - Y_0^1) \\ &\qquad + \left(\lambda - \frac{1}{2}\rho\nu\right)(\widehat{Y}_h - Y_0) + (\widehat{V}_h^1 - V_0^1)\Bigg)\Bigg). \end{aligned}$$

So we solved the SDEs S^B and S^W separately, and now want to combine them again using some splitting scheme. While the algorithm for S^B is exact, the algorithm for $\widehat{S}^W$ is of second order, since $\widehat{A}$ is (strongly potential) second order scheme as proved in the Section 2.2.3, and $\widehat{\boldsymbol{Y}}$ is computed using the trapezoidal rule. Hence, we would like to use one splitting scheme that is also of second order, one can see the article [MQ02] which contains a lot of splitting schemes. While in [BB23a], the authors show that a splitting scheme called the randomized Leapfrog splitting performs relatively better. This splitting scheme is given by

$$\widehat{S}^{\text{rHeston}}((s, \boldsymbol{v}, \boldsymbol{y}), h) = \begin{cases} \widehat{S}^W(S^B((s, \boldsymbol{v}, \boldsymbol{y}), h), h), & \text{if } U \leq \frac{1}{2}, \\ S^B(\widehat{S}^W((s, \boldsymbol{v}, \boldsymbol{y}), h), h), & \text{else}, \end{cases}$$

where U is a uniform random variable on $[0, 1]$ independent of everything else.

Finally we have a (strongly potential) second order weak scheme for $(S^N, \boldsymbol{V}^N)$ which is $(\widehat{S}^N, \widehat{\boldsymbol{V}}^N) := (\widehat{S}^{\text{rHeston}}, \widehat{A})$.

2.2.5. Higher Order Approximation

In this subsection, we present a method to obtain a higher-order scheme for the Markovian approximations (2.5) and (2.6) of the rough Heston model, following the approach in [AL25], where the authors constructed a higher-order scheme by iterating on a second-order scheme for the log-Heston model. Recall that we have already obtained a (not rigorously) second-order weak scheme for (S, V) in Section 2.2.4.

In the following, we will use the same notations in Section 2.2.2. We consider a time horizon $T > 0$ and a time step $h = T/n$, with $n \in \mathbb{N}^{>0}$. We denote $(\widehat{S}_h^{s,v}, \widehat{V}_h^v)$ an approximate scheme for the solution (S, V) of some SDEs starting from (s, v) with time-step h, and

$$\widehat{P}_h f(s, v) := \mathbb{E}[f(\widehat{S}_h^{s,v}, \widehat{V}_h^v)]$$

the associated semigroup approximation. We also define the notation P_h in the same way for solution (S, V) by

$$P_h f(s, v) := \mathbb{E}[f(S_h^{s,v}, V_h^v)].$$

Next we consider the SDE on Markovian approximation for rough log-Heston model

$$dS_t^N = \sqrt{V_s^N}\left(\rho dW_s + \sqrt{1-\rho^2} dB_s\right) - \frac{1}{2} V_s^N ds,$$

$$d\boldsymbol{V}_t^N = -\mathrm{diag}(\boldsymbol{x})\left(\boldsymbol{V}_t^N - \boldsymbol{v}_0\right) dt + \left(\theta - \lambda V_t^N\right) \mathbf{1} dt + \nu \sqrt{V_t^N}\, \mathbf{1}\, dW_t,$$

where $\boldsymbol{w} \cdot \boldsymbol{v}_0 = V_0$, $\boldsymbol{V}^N := (V^{(i)})_{i=1}^N$ and $V^N = w^\top \cdot \boldsymbol{V}^N$.

For $l \in \mathbb{N}^{>0}$, let us define the time step $h_l = \frac{T}{n^l}$. We set $Q_l = \widehat{P}_{h_l}$ the operator obtained by using the approximation scheme with the time step h_l. The principle is to iterate the identity proposed by Talay and Tubaro in [TT90], consisting in writing

$$\widehat{P}_h^{[n]} - P_T = \widehat{P}_h^{[n]} - P_h^{[n]} = \sum_{i=0}^{n-1} \widehat{P}_h^{[n-(i+1)]}(\widehat{P}_h - P_h) P_h^{[i]} = \sum_{i=0}^{n-1} \widehat{P}_h^{[n-(i+1)]}(\widehat{P}_h - P_h) P_{ih}, \quad (2.48)$$

where $\widehat{P}_h^{[0]} = Id$ and $\widehat{P}_h^{[i]} = \widehat{P}_h^{[i-1]} \widehat{P}_h$ for $i \geq 1$, and $P_h^{[i]} = P_{ih}$ by the semigroup property. Using this idea, we establish the following proposition. The proof is given in Appendix A.2.

Proposition 2.2.18. *If the semigroups (P_t) for (S^N, V^N) and the family of operators (Q_l) for $(\widehat{S}^N, \widehat{V}^N)$ satisfy*

(i) $\forall l, k, L \in \mathbb{N}, \exists C > 0, \forall f \in \mathcal{C}_{pol}^{k+12,L}(\mathbb{R} \times \mathbb{R}_+)$,

$$\|(P_{h_l} - Q_l) f\|_{k,L+3} \leq C \|f\|_{k+12,L} h_l^3. \quad (2.49)$$

(ii) $\forall l, k, L \in \mathbb{N}, \exists C > 0, \forall f \in \mathcal{C}_{pol}^{k,L}(\mathbb{R} \times \mathbb{R}_+)$,

$$\max_{0 \leq j \leq n^l} \|Q_l^{[j]} f\|_{k,L} + \sup_{t \leq T} \|P_t f\|_{k,L} \leq C \|f\|_{k,L}. \quad (2.50)$$

Then we have

$$\widehat{\mathcal{P}}^{2,n} := \widehat{P}_{h_1}^{[n]} + \sum_{i=0}^{n-1} \widehat{P}_{h_1}^{[n-(i+1)]}(\widehat{P}_{h_2}^{[n]} - \widehat{P}_{h_1}) \widehat{P}_{h_1}^{[i]} \quad (2.51)$$

is an approximation of fourth order. Namely, we get

$$\forall f \in \mathcal{C}_{pol}^{24}(\mathbb{R} \times \mathbb{R}_+),\ \exists C > 0, L \in \mathbb{N},\ \|\widehat{\mathcal{P}}^{2,n} f - P_T f\|_{0,L+6} \leq C \|f\|_{24,L} \left(\frac{T}{n}\right)^4. \quad (2.52)$$

Remark 2.2.19. One should note that $\widehat{P}_{h_1}^{[n-(i+1)]}\widehat{P}_{h_2}^{[n]}\widehat{P}_{h_1}^{[i]}$ to applying the scheme on a time grid that is uniform except for a refinement of the $(i+1)$-th time step. This grid therefore contains $2n$ time steps. If all the terms in the sum of (2.51) were computed explicitly, the overall computational cost would be of order $O(n^2)$. Consequently, the method would not be more efficient than using the second-order scheme with n^2 time steps.

To overcome this drawback, we introduce random grids. More precisely, the authors in [AB21] consider a random variable κ uniformly distributed on $\{0,\dots,n-1\}$ and define the following randomized scheme

$$\widehat{\mathcal{P}}^{2,n} = \widehat{P}_{h_1}^{[n]} + n\mathbb{E}[\widehat{P}_{h_1}^{[n-(\kappa+1)]}(\widehat{P}_{h_2}^{[n]} - \widehat{P}_{h_1})\widehat{P}_{h_1}^{[\kappa]}]. \tag{2.53}$$

Thus, for the correcting term, we consider a random time grid that is the obtained from the uniform time grid with time step T/n by refining uniformly the $(\kappa+1)$-th time step with a time step $h_2 = T/n^2$.

We have presented how $\widehat{\mathcal{P}}^{2,n}$ improves the convergence of $\widehat{\mathcal{P}}^{1,n} = \widehat{P}_{h_1}^{[n]}$ in the Proposition 2.2.18. Furthermore, for $\nu \geq 2$, [AB21, Lemma 2.4] defines by induction approximations $\widehat{\mathcal{P}}^{\nu,n}$, such that

$$\forall f \in \mathcal{C}_{\text{pol}}^{12\nu}(\mathbb{R}\times\mathbb{R}_+),\ \exists C>0, L\in\mathbb{N},\ \|\widehat{\mathcal{P}}^{\nu,n}f - P_T f\|_{0,L+3\nu} \leq C\|f\|_{12\nu,L}\left(\frac{T}{n}\right)^{2\nu}. \tag{2.54}$$

Therefore, we see that $\widehat{\mathcal{P}}^{\nu,n}$ is a scheme with rate 2ν for every $\nu \geq 1$. While we remark here again, to rigorously obtain a true high order scheme, we should first prove that the SDEs on Markovian approximation for rough log-Heston model have polynomial estimates on the derivatives for the Cauchy problem.

2.2.6. Optimal Choice on N and M

Assume that the scheme $\widehat{A}$ is rigorously of second order convergence. Then, roughly we can see that

- To compute the weak Markovian approximation (S^N, V^N) for (S,V), its computation cost is $\mathcal{O}(N)$ when applying the non-geometric Gaussian quadrature for number of quadrature points N with the error $\mathcal{O}(e^{-c\sqrt{N}})$, where $c = \sqrt{H+1/2}$.
- To compute the weak scheme $(S^{N,M}, V^{N,M})$ for (S^N, V^N), where M is the number of time stepping, it costs $\mathcal{O}(N^2M)$ (which is discussed in [BB23c, Section 3.1.3]) with the error $\mathcal{O}(NM^{-2})$.

Then, if allowing the Markovian dimension N and the number of time stepping M can depend on each others, then it leads us to find a optimal pair (N,M) such that

$$\min\{N^2M | C_1e^{-c\sqrt{N}} + C_2NM^{-2} \leq \epsilon\}, \tag{2.55}$$

where the constants C_i, $i = 1, 2$ inside are both independent of N, M.

The Lagrangian of this optimization problem is given by

$$F(M, N, \lambda) := N^2 M + \lambda(C_1 e^{-c\sqrt{N}} + C_2 N M^{-2} - \epsilon) \tag{2.56}$$

From $\partial F/\partial M = 0$, we have

$$\lambda \approx N M^3.$$

And from $\partial F/\partial N = 0$, we have

$$NM + C_2 \lambda M^{-2} \approx \lambda e^{-c\sqrt{N}}/\sqrt{N} \approx N M^3 e^{-c\sqrt{N}}/\sqrt{N}.$$

Since $NM + C_2 \lambda M^{-2} \approx NM$, we get the optimal relation of N and M is

$$M^{-2} \approx e^{-\sqrt{(H+1/2)N}}/\sqrt{N},$$

where $H \in (-1/2, 1/2)$ is the Hurst parameter.

It seems workable and plausible to do it. However, one of the important point we ignore here is that the constant C_2 will definitely depend on the Markovian dimension N. Recall the SDE for the Markovian approximation is

$$\mathrm{d}\boldsymbol{V}_t^N = -\mathrm{diag}(\boldsymbol{x}) \left(\boldsymbol{V}_t^N - \boldsymbol{v}_0\right) \mathrm{d}t + \left(\theta - \lambda \boldsymbol{w}^\top \boldsymbol{V}_t^N\right) \mathbf{1}\, \mathrm{d}t + \nu \sqrt{\boldsymbol{w}^\top \boldsymbol{V}_t^N} \mathbf{1}\, \mathrm{d}W_t,$$

where the multiple constant coefficients are related to N (e.g. $\mathrm{diag}(\boldsymbol{x})$ is a $N \times N$ matrix). It can be readily verified from the proof of Theorem 2.2.10 (see [Alf10b, Appendix A]) that the parameter N in the constant $C_2(N)$ arises from the following sources.

- The constant of the residual $R_3^{\widehat{S}}$ when proving the scheme $\widehat{A}$ has potentially second order weak convergence rate.
- The constants of the estimates for uniform bounded moments for the scheme $\widehat{A}$.
- The constants of the polynomial estimates on the derivatives of the Cauchy problem for the model (2.21), i.e. the constant in Theorem 2.2.10 (ii).

Therefore, the constant $C_2(N)$ is essentially intractable. Even if it were computable, determining an optimal pair (N, M) would be highly complicated. Nevertheless, in Chapter 4, we present numerical results for the scheme under different values of N to examine the relationship between the maximal relative error and the computational time.

3. Quadratic Rough Heston Model

Recall the quadratic rough Heston model has the following dynamics for the stock price S and the volatility V

$$\begin{aligned} dS_t &= S_t\sqrt{V_t}\,dW_t, \\ V_t &= a(Z_t - b)^2 + c, \end{aligned} \tag{3.1}$$

where W is a Brownian motion, a, b and c some positive constants and Z_t follows a integral equation given by

$$Z_t = \int_0^t (t-s)^{\alpha-1}\frac{\lambda}{\Gamma(\alpha)}(\theta - Z_s)ds + \int_0^t (t-s)^{\alpha-1}\frac{\lambda}{\Gamma(\alpha)}\eta\sqrt{V_s}dW_s,$$

with $\alpha \in (\frac{1}{2}, 1)$, $\lambda > 0$, $\eta > 0$.

Denote the fractional kernel $(t-s)^{\alpha-1}\frac{\lambda}{\Gamma(\alpha)}$ by $K(t-s)$ and its log stock price $logS_t$ by X_t, then it reduces to

$$dX_t = -\frac{1}{2}p(Z_t)dt + \sqrt{p(Z_t)}dW_t, \tag{3.2}$$

$$Z_t = \int_0^t K(t-s)b(Z_s)ds + \int_0^t K(t-s)\eta\sqrt{p(Z_t)}\ dW_s, \tag{3.3}$$

where $b(x) = \theta - x$ and $p(x) = a(x-b)^2 + c$, $a, b, c > 0$. We remark here $\sigma(x) := \eta\sqrt{p(x)}$ is a smooth function having linear growth with bounded first order derivative (so $\sigma(x) \in \mathrm{Lip}$).

Theorem 3.0.1. *The Stochastic Volterra equation*

$$Z_t = \int_0^t K(t-s)b(Z_s)ds + \int_0^t K(t-s)\eta\sqrt{p(Z_t)}\ dW_s,$$

where $b(x) = \theta - x$ and $p(x) = a(x-b)^2 + c$, $a, b, c > 0$ admits a unique continuous strong solution Z.

Proof. Note that $b, \sigma := \eta\sqrt{p(x)}$ satisfy linear growth condition and Lipschitz condition. So we can apply [AJLP19, Theorem 3.3] since the fractional kernel K also satisfies the assumption of [AJLP19, Theorem 3.3]. $\square$

Plan for this Chapter

In this chapter, we first introduce a stochastic process called the polynomial Volterra process, mainly based on [AJCP+24]. The process Z in our quadratic rough Heston model belongs to this class of processes and therefore inherits many properties of polynomial Volterra processes. We then extend the results of [BJP25] and [BMM25] to models involving a more general volatility process. As an application, we obtain an estimate of the error bound for the expectation of payoff functions in the quadratic rough Heston model. Finally, we derive a weak simulation method for this model as well, which is theoretically of second order weak rate.

3.1. Polynomial Volterra Process

In this section, we introduce a class of stochastic processes called polynomial Volterra processes. Most of the following content is based on [AJCP+24].

3.1.1. Definition and Existence

We first fix a dimension $d \in \mathbb{N}$ and consider a filtered probability space $(\Omega, \mathcal{F}, (\mathcal{F}_t)_{t\geq 0}, \mathbb{P})$, where $(\mathcal{F}_t)_{t\geq 0}$ satisfies the usual conditions and $\mathcal{F}_0$ is the trivial σ-algebra on Ω. We denote $\mathrm{Pol}_k(\mathbb{R}^d)$ is set of all $(x_1, ..., x_d)$-polynomials on $\mathbb{R}$ with degree $\leq k$.

Definition 3.1.1. A continuous polynomial Volterra process of convolution type is a d-dimensional adapted process X with continuous trajectories solving a stochastic Volterra equation of the form

$$X_t = g_0(t) + \int_0^t K(t-s)b(X_s)ds + \int_0^t K(t-s)\sigma(X_s)dW_s, \quad t \geq 0, \tag{3.4}$$

where

- W is a d-dimensional Brownian motion,
- the *initial condition* $g_0 \colon \mathbb{R}_+ \to \mathbb{R}^d$ is in $C(\mathbb{R}_+, \mathbb{R}^d)$,
- the *convolution kernel* $K \colon \mathbb{R}_+ \to \mathbb{R}^{d\times d}$ is in $L^2_{\mathrm{loc}}(\mathbb{R}_+, \mathbb{R}^{d\times d})$,
- the map $b \colon \mathbb{R}^d \to \mathbb{R}^d$ has components in $\mathrm{Pol}_1(\mathbb{R}^d)$, and $\sigma \colon \mathbb{R}^d \to \mathbb{R}^{d\times d}$ is a continuous map such that $a(x) = \sigma(x)\sigma(x)^\top$ has entries in $\mathrm{Pol}_2(\mathbb{R}^d)$. More precisely,

$$b(x) = b_0 + \sum_{i=1}^d b_i x_i, \quad a(x) = A_0 + \sum_{i=1}^d A_i x_i + \sum_{i,j=1}^d A_{ij} x_i x_j \tag{3.5}$$

  for some $b_i \in \mathbb{R}^d$ and $A_i, A_{ij} \in \mathbb{R}^{d\times d}$.

Remark 3.1.2. (i) If $A_{ij} = 0$ for any i, j, then X is an affine Volterra process as in [AJLP19].

(ii) For $K = 1$, X is just a polynomial process.

(iii) If $A_{ij} = 0$ for any i, j and $K = 1$, then X is just an affine process.

Example 3.1.3. (i) The volatility of the Heston model is an affine process.

(ii) The volatility of the rough Heston model is an affine Voltera process.

(iii) The volatility of the quadratic rough Heston model is a polynomail Voltera process.

Remark 3.1.4. In contrast to the Heston model and the rough Heston model, whose main feature is the availability of an explicit or semi-explicit representation of their characteristic functions, the quadratic rough Heston model completely lacks such a representation. As a consequence, we lose access to the characteristic function in explicit form, which creates significant difficulties when deriving error bounds in the next subsection.

The next proposition is an extension for [AJCP+24, Proposition 2.1], which gives one sharper estimate for the moments of solution to (3.4).

Proposition 3.1.5 (Estimate for Moments)**.** *Let X be a continuous solution to (3.4), st., the drift function b and the volatility function σ are continuous and have linear growth. Then, for any $p \in \mathbb{N}$ and $T \geq 0$,*

$$\sup_{0\leq t\leq T} \mathbb{E}[|X_t|^p] \leq \mathbb{E}[\sup_{0\leq t\leq T} |X_t|^p] \leq c < \infty \tag{3.6}$$

for some constant c which depends only on $\sup_{0\leq t\leq T} |g_0(t)|$, p, $K|_{[0,T]}$, b, σ, and T.

Corollary 3.1.6. *Let X be a continuous solution to (3.4) and a polynomial Volterra process, then Inequality (3.6) holds for some constant c which depends only on $\sup_{0\leq t\leq T} |g_0(t)|$, p, $K|_{[0,T]}$, b_i, A_i, A_{ij}, and T.*

Remark 3.1.7. This extended proposition plays a crucial role in proving Lemma 3.2.53, which establishes that the control functions for the volatility model are polynomial.

Remark 3.1.8. A smart resolution to obtain the inequality

$$\sup_{0\leq t\leq T} \mathbb{E}[|X_t|^p] \leq c$$

in the Corollary 3.1.6 is to use the Theorem 3.1.20 in the following, which represents the moments of X by polynomial of its initial value X_0. The result follows from the fact that the coefficients $(c_\beta(t))_{|\beta|\leq p}$ are continuous so bounded on the compact interval.

Proof. WLOG, assume the initial curve $g_0 = 0$ since it is continuous and hence bounded on compact interval. Then the proof follows a similar argument as in the proof of [AJLP19, Lemma 3.1]. We can set $g_n(t) = \mathbb{E}[\sup_{0\le s\le t} |X_s|^p \mathbb{1}_{t<\tau_n}]$ and get a similar inequality

$$g_n(t) \le c' + c' \sup_{0\le s\le t} |K|^2(s-\cdot) * f_n(s),$$

where $f_n(s) = \mathbb{E}[|X_s|^p \mathbb{1}_{s<\tau_n}]$ defined as in the proof of [AJLP19, Lemma 3.1].

And we note that

$$\sup_{r\le t} |K|^2(r-\cdot) * f_n = \sup_{r\le t} \int_0^r K^2(r-s) f_n(s) ds \tag{3.7}$$

$$\le \sup_{r\le t} \int_0^r K^2(r-s) g_n(s) ds \tag{3.8}$$

$$\le \int_0^t K^2(t-s) g_n(s) ds \tag{3.9}$$

follows from

$$\int_0^t K^2(t-s) g_n(s) ds \ge \int_r^t K^2(t-s) g_n(s) ds \tag{3.10}$$

$$\ge \int_r^t K^2(t-s) g_n(s-r) ds = \int_0^{t-r} K^2(t-r-s) g_n(s) ds, \tag{3.11}$$

for any $r \ge 0$ since $g_n \ge 0$ is a non-decreasing function.

Finally, we can obtain the same inequality in [AJLP19, Lemma 3.1]

$$g_n \le c' + c'|K|^2 * g_n.$$

Then we can use the same argument at the last paragraph of proof for [AJLP19, Lemma 3.1] to obtain

$$g_n \le c < \infty,$$

let $n \to \infty$ by using the monotone convergence Theorem to obtain the result. □

And we have the following result for the weak existence of solutions to (3.4).

Assumption 3.1.9. There exists a constant $\gamma \in (0, 2]$ such that $\int_0^h |K(t)|^2 dt = O(h^\gamma)$ and $\int_0^T |K(t+h) - K(t)|^2 dt = O(h^\gamma)$ for every $T < \infty$.

Example 3.1.10. Assumption 3.1.9 is satisfied for the fractional kernel $K(t) = t^{H-1/2}$ with $H \in (0, 1)$. In particular, the corresponding polynomial Volterra process (3.4) fails to be a semimartingale whenever $H \ne 1/2$.

[AJCP+24, Theorem 2.3] establishes the weak existence of polynomial Volterra processes, which is stated below.

Theorem 3.1.11 (Existence of polynomial Volterra processes)**.** *Let Assumption 3.1.9 hold. Then, for any $0 < \alpha < \gamma/2$, (3.4) admits a weak solution X such that $X - g_0$ has α-Hölder continuous trajectories.*

3.1.2. Moment Formula

In the following, we derive moment formulas for a continuous Volterra process X solving (3.4). More precisely, we aim to obtain formulas for moments of the form

$$\mathbb{E}[X_t^\alpha] = \mathbb{E}[X_{1,t}^{\alpha_1} \cdots X_{d,t}^{\alpha_d}], \quad t \geq 0, \tag{3.12}$$

with $\alpha = (\alpha_i)_{i=1}^d \in \mathbb{N}_0^d$ a multi-index. One of the main difficulties in characterizing these moments, compared with the classical framework where the kernel K is equal to I_d, stems from the fact that X is not necessarily a Markovian semimartingale. To overcome this issue, and following the approach of [AJEE19a, JO19], we introduce for each $T \geq 0$ the following adjusted forward process associated with the polynomial Volterra process

$$g_t(T) = g_0(T) + \int_0^t K(T-s)b(X_s)ds + \int_0^t K(T-s)\sigma(X_s)dW_s, \quad t \leq T. \tag{3.13}$$

The following proposition is an extension of [AJCP+24, Lemma 3.1] which shows that we can control the moments of the adjusted forward process g uniformly.

Proposition 3.1.12. *Suppose that X is a continuous process solving (3.4) and define the processes g as in (3.13). Then, for any $p \in \mathbb{N}$ and $0 \leq s \leq T$,*

$$\sup_{0 \leq s \leq T} \sup_{0 \leq t \leq s} \mathbb{E}\left[|g_t(s)|^p\right] \leq c < \infty \tag{3.14}$$

for some constant c which depends only on $\sup_{0 \leq t \leq T} |g_0(t)|$, p, $K|_{[0,T]}$, b_i, A_i, A_{ij}, and T.

Remark 3.1.13. This extended proposition is very useful and also plays a significant role in proving Lemma 3.2.53.

Proof. It is sufficient to prove the inequality (3.14) for $p \geq 2$. Given $T \geq 0$, using the Burkholder-Davis-Gundy and Jensen's inequalities

$$\begin{aligned} \mathbb{E}\left[\left|\int_0^t K(s-r)\sigma(X_r)dW_r\right|^p\right] &\leq C\,\mathbb{E}\left[\left(\int_0^t |K(s-r)|^2 |a(X_r)| dr\right)^{\frac{p}{2}}\right] \\ &\leq C\left(\int_0^T |K(t)|^2 dt\right)^{\frac{p}{2}} \sup_{0 \leq t \leq T} \mathbb{E}[|a(X_t)|^{\frac{p}{2}}] \end{aligned} \tag{3.15}$$

for some constant $C > 0$. Similarly, multiple applications of Jensen's inequality yield

$$\mathbb{E}\left[\left(\left|\int_0^t K(s-r)b(X_r)dr\right|\right)^p\right] \leq t^{\frac{p}{2}} \left(\int_0^t |K(r)|^2 dr\right)^{\frac{p}{2}} \sup_{0 \leq r \leq t} \mathbb{E}[|b(X_r)|^p] \tag{3.16}$$

$$\leq T^{\frac{p}{2}} \left(\int_0^T |K(r)|^2 dr\right)^{\frac{p}{2}} \sup_{0 \leq r \leq T} \mathbb{E}[|b(X_r)|^p] \tag{3.17}$$

since the functions b and $a := \sigma^2$ have the form (3.5), (3.15) and (3.17) together with the Proposition 3.1.5 yield (3.14).

Therefore, note that the right hand sides of the two inequalities (3.15) and (3.17) are independent of t and s. Then we can take $\sup_{0\le s\le T}\sup_{0\le t\le s}$ on both hand sides to get the conclusion. $\square$

Notice that $g_T(T) = X_T$ and more generally

$$g_t(T) = \mathbb{E}_t\left[X_T - \int_t^T K(T-s)b(X_s)ds\right], \quad t \le T, \tag{3.18}$$

It follows from the inequality (3.15) – the process $M_t = \int_0^t K(T-s)\sigma(X_s)dW_s$, $t \le T$, is a martingale. Moreover, for each $T \ge 0$, $g_t(T)$ is a semimartingale with dynamics

$$dg_t(T) = K(T-t)dZ_t = K(T-t)b(X_t)dt + K(T-t)\sigma(X_t)dW_t, \quad t < T. \tag{3.19}$$

To study the moments of X in (3.12), we need to understand the behavior of more general moments of the processes g defined in (3.13), we will consider expressions of the form in the following

$$\mathfrak{m}^{(p)}(t, T_1, \dots, T_p; w) = \mathbb{E}\left[\prod_{n=1}^{p} g_{i_n,t}(T_n)\right], \tag{3.20}$$

where $p \in \mathbb{N}$, $0 \le t \le \min\{T_1, \dots, T_p\}$, $w = (i_n)_{n=1}^p \in \{1, \dots, d\}^p$, and $g_{i_n,t}(T_n)$ is the i_n-th coordinate of $g_t(T_n)$. If $d = 1$, we can omit the argument w and write $\mathfrak{m}^{(p)}(t, T_1, \dots, T_p)$ for $p \in \mathbb{N}$. We shall use the convention $\mathfrak{m}^{(0)} \equiv 1$.

Remark 3.1.14. Notice that

$$\mathfrak{m}^{(p)}(t, t, \dots, t; w) = \mathbb{E}[X_t^{\alpha(w)}], \tag{3.21}$$

where $\alpha(w)$ is the multi-index given by $\alpha_k(w) = \#\{n : i_n = k\}$, $k = 1, \dots, d$. In particular, $|\alpha(w)| = p$.

We also introduce some notations. Given $N \in \mathbb{N}$, we define the set

$$\mathcal{I}^{(N)} = \{(p, w) : p \in \{0, 1, \dots, N\} \text{ and } w \in \{1, \dots, d\}^p\}. \tag{3.22}$$

Let D_N be the cardinality of the set $\mathcal{I}^{(N)}$. Notice that $D_N = N + 1$ for $d = 1$ and $D_N = \sum_{p=0}^N d^p = (d^{N+1} - 1)/(d-1)$ for $d > 1$. For $T \ge 0$, let

$$\mathcal{D}_T^{(N)} = \{(t, T_1, \dots, T_N) \in [0, T]^{N+1} : t \le \min\{T_1, \dots, T_N\}\}, \tag{3.23}$$

and let π be an enumeration of $\mathcal{I}^{(N)}$.

Definition 3.1.15. Let $\mathcal{X}_T^{(N)}$ be the space of $\mathbb{R}^{D_N}$-valued bounded functions $\mathbf{f}$ on $\mathcal{D}_T^{(N)}$ such that the $\pi(p, w)$-th component $\mathbf{f}_{\pi(p,w)}$ only depends on the variables $(t, T_1, \dots, T_p)$ for any $(p, w) \in \mathcal{I}^{(N)}$, and $\mathbf{f}_{\pi(0,\emptyset)} \equiv 1$.

Remark 3.1.16. As we will see in the Theorem 3.1.18 below, the vector-valued function

$$(t, T_1, \dots, T_N) \mapsto \{\mathfrak{m}^{(p)}(t, T_1, \dots, T_p; w) : p \in \{0, \dots, N\} \text{ and } w \in \{1, \dots, d\}^p\} \quad (3.24)$$

belongs to the space $\mathcal{X}_T^{(N)}$.

Given $\mathbf{f} \in \mathcal{X}_T^{(N)}$, define the function $\mathcal{M}_T^{(N)}\mathbf{f}$ on $\mathcal{D}_T^{(N)}$ as follows: $(\mathcal{M}_T^{(N)}\mathbf{f})_{\pi(0,\emptyset)} \equiv 0$, and for $(p, w) \in \mathcal{I}^{(N)}$ with $1 \le p \le N$ and $w = (i_n)_{n=1}^p$,

$$\begin{aligned}
(\mathcal{M}_T^{(N)}\mathbf{f})_{\pi(p,w)}(t, T_1, \dots, T_N) &= \sum_{n=1}^p \int_0^t e_{i_n}^\top K(T_n - r) b_0 \mathbf{f}_{\pi(p-1, w_{-n})}(r, (T_m)_{m \neq n}) dr \\
&+ \sum_{n=1}^p \sum_{j=1}^d \int_0^t e_{i_n}^\top K(T_n - r) b_j \mathbf{f}_{\pi(p, w_{-n}^j)}(r, r, (T_m)_{m \neq n}) dr \\
&+ \sum_{1 \le n < m \le p} \int_0^t e_{i_n}^\top K(T_n - r) A_0 K(T_m - r)^\top e_{i_m} \mathbf{f}_{\pi(p-2, w_{-n,-m})}(r, (T_l)_{l \neq m,n}) dr \\
&+ \sum_{1 \le n < m \le p} \sum_{j=1}^d \int_0^t e_{i_n}^\top K(T_n - r) A_j K(T_m - r)^\top e_{i_m} \mathbf{f}_{\pi(p-1, w_{-n,-m}^j)}(r, r, (T_l)_{l \neq m,n}) dr \\
&+ \sum_{1 \le n < m \le p} \sum_{j,k=1}^d \int_0^t e_{i_n}^\top K(T_n - r) A_{jk} K(T_m - r)^\top e_{i_m} \mathbf{f}_{\pi(p, w_{-n,-m}^{j,k})}(r, r, r, (T_l)_{l \neq m,n}) dr,
\end{aligned} \quad (3.25)$$

where e_i is the i-th canonical vector in $\mathbb{R}^d$, w_{-n} is the vector obtained by erasing the n-th coordinate of w, $w_{-n,-m}$ is the vector obtained by erasing the n-th and m-th coordinates of w, w_{-n}^j is the vector whose first coordinate is j and the other coordinates are given by w_{-n}, and $w_{-n,-m}^{j,k}$ is the vector whose first two coordinates are j, k and the other coordinates are given by $w_{-n,-m}$.

Remark 3.1.17. The right hand side of (3.25) is well-defined for $\mathbf{f} \in \mathcal{X}_T^{(N)}$ since K is locally square-integrable. Moreover, $\mathcal{M}_T^{(N)}\mathbf{f} \in \mathcal{X}_T^{(N)}$.

The moment formula for (3.20) is given in [AJCP+24, Theorem 3.3] which is stated in the following.

Theorem 3.1.18 (Main moment formula). *Fix $N \in \mathbb{N}$ and $T \ge 0$. Define $\mathbf{m} : \mathcal{D}_T^{(N)} \to \mathbb{R}^{D_N}$ by $\mathbf{m}_{\pi(0,\emptyset)} = \mathfrak{m}^{(0)} \equiv 1$, and for $(p, w) \in \mathcal{I}^{(N)}$ with $1 \le p \le N$, $\mathbf{m}_{\pi(p,w)}(t, T_1, \dots, T_N) = \mathfrak{m}^{(p)}(t, T_1, \dots, T_p; w)$ as in (3.20). Then $\mathbf{m} \in \mathcal{X}_T^{(N)}$ and $\mathbf{m}$ solves the integral equation*

$$\mathbf{m}(t, T_1, \dots, T_N) = \mathbf{m}(0, T_1, \dots, T_N) + (\mathcal{M}_T^{(N)}\mathbf{m})(t, T_1, \dots, T_N), \quad (t, T_1, \dots, T_N) \in \mathcal{D}_T^{(N)}, \quad (3.26)$$

with $\mathcal{M}_T^{(N)}$ as in (3.25). Furthermore, $\mathbf{m}$ is the unique solution in $\mathcal{X}_T^{(N)}$ of (3.26) with initial condition $\mathbf{m}(0, T_1, \dots, T_N)$, and $\mathbf{m}$ is continuous on $\mathcal{D}_T^{(N)}$.

Remark 3.1.19. The well-posedness for Volterra-type integral equation in the Theorem 3.1.18 can be found in [AJCP+24, Appendix A].

3.1.3. Representation for Moments in Polynomials

Starting at an example, if $K \equiv \mathrm{I}_d$ and $g_0 \equiv X_0$ in our framework, i.e., when X degenerates as a polynomial process, [FL16, Theorem 3.1] gives the following explicit formula for the moments of X

$$\mathbb{E}[p(X_t)] = H(X_0)^T \mathrm{e}^{tG}\vec{p}, \quad p \in \mathrm{Pol}_n(\mathbb{R}^d),\ t \geq 0, \tag{3.27}$$

where $H(x)$ is a vector whose components are elements of a basis for $\mathrm{Pol}_n(\mathbb{R}^d)$, G is the matrix of the infinitesimal generator of X restricted to $\mathrm{Pol}_n(\mathbb{R}^d)$, and $\vec{p}$ are the coordinates of the polynomial p with respect to the basis in $H(x)$. So we know that the moments of X with initial value X_0 are again polynomials in the variable X_0, i.e., we have representation for moments in polynomials

$$\mathbb{E}[X_t^\alpha] = \sum_{\beta \in \mathbb{N}_0^d, |\beta| \leq |\alpha|} c_\beta(t) X_0^\beta, \tag{3.28}$$

for some deterministic and time-dependent family of coefficients $\{c_\beta : \beta \in \mathbb{N}_0^d, |\beta| \leq |\alpha|\}$, which can be computed explicitly from (3.27).

In the following theorem, it also asserts that this structural property still holds for the moments $\mathfrak{m}^{(p)}$ in (3.20) of a polynomial Volterra process X starting at X_0, i.e. when $g_0 \equiv X_0$ in (3.4), by applying Theorem 3.1.18.

Theorem 3.1.20 (Moments are polynomials)**.** *Suppose that X is a continuous solution to (3.4) with $g_0 \equiv X_0$ and fix $T \geq 0$. Then, for all $p \in \mathbb{N}_0$ and all $w \in \{1, \dots, d\}^p$, we can express the functions $\mathfrak{m}^{(p)}$ in (3.20) as*

$$\mathfrak{m}^{(p)}(t, T_1, \dots, T_p; w) = \sum_{\beta \in \mathbb{N}_0^d, |\beta| \leq p} C_\beta^{(p)}(t, T_1, \dots, T_p; w) X_0^\beta, \quad (t, T_1, \dots, T_p) \in \mathcal{D}_T^{(p)}, \tag{3.29}$$

where the functions $C_\beta^{(p)}(\cdot, \cdots, \cdot; w) \in C(\mathcal{D}_T^{(p)}, \mathbb{R})$ are independent of X_0.
In particular, for any $\alpha \in \mathbb{N}_0^d$, there exists a family $\{c_\beta\}_{\beta \in \mathbb{N}_0^d, |\beta| \leq |\alpha|}$ of real-valued continuous functions on $[0, T]$, independent of X_0, such that

$$\mathbb{E}[X_t^\alpha] = \sum_{\beta \in \mathbb{N}_0^d, |\beta| \leq |\alpha|} c_\beta(t) X_0^\beta, \quad t \geq 0. \tag{3.30}$$

Remark 3.1.21. The coefficients can be obtained by solving an integral equation similar to (3.26), see [AJCP+24, Theorem 3.8].

3.1.4. Weak Uniqueness Characterized by Moments

The ideas in Theorem 3.1.20 can be extended to characterize moments of the finite dimensional distributions of a weak solution to (3.4). With similar argument in the proof of [FL16, Lemma 4.1], we have the following result.

Theorem 3.1.22 (Uniqueness in law for polynomial Volterra processes)**.** *Suppose that for any weak solution X to (3.4) and for any $t \geq 0$, the law of X_t is determined by its moments. Then uniqueness in law (i.e., uniqueness of all finite-dimensional distributions) holds for (3.4).*

Remark 3.1.23. It claims uniqueness of moments can imply uniqueness in law, which doesn't hold for general random variable, e.g. for one class of random variables $X = e^Z$ where $Z \sim N(0,1)$.

Remark 3.1.24. For affine Volterra processes, [AJLP19] provides uniqueness in law via the Fourier-Laplace transform. Pathwise uniqueness for stochastic Volterra equations has been established for Lipschitz coefficients, e.g., in [BM80] and, for certain coefficients in the one-dimensional case, when the kernel K is regular [PS23] and when the kernel is singular in [Wan08].

Remark 3.1.25. As explained in the proof of [FL16, Lemma 4.1], the hypothesis of Theorem 3.1.22 holds for instance if for any weak solution X to (3.4) and $t \geq 0$, there is $\epsilon > 0$ such that $\mathbb{E}[\exp(\epsilon|X_t|)] < \infty$. For example, this is the case if any such solution X is bounded.

Example 3.1.26. Let $\alpha_1 \leq \alpha_2$, $b \in [\alpha_1, \alpha_2]$, $\lambda \geq 0$ and $c > 0$. Fix a scalar kernel $K : [0,T] \to \mathbb{R}$ that satisfies Assumptions 3.1.9 and K is nonnegative, not identically zero, non-increasing and continuous on $(0,\infty)$, and its resolvent of the first kind L is nonnegative and non-increasing in the sense that $s \mapsto L([s, s+t])$ is non-increasing for all $t \geq 0$.

Then, there exists a unique weak $[\alpha_1, \alpha_2]$-valued solution to the equation

$$\begin{aligned} &Y_t = Y_0 + \lambda \int_0^t K(t-s)(b - Y_s)ds + c\int_0^t K(t-s)\sqrt{(Y_s - \alpha_1)(\alpha_2 - Y_s)}\,dW_s, \\ &Y_0 \in [\alpha_1, \alpha_2]. \end{aligned} \tag{3.31}$$

which we call Jacobi Volterra process on $[\alpha_1, \alpha_2]$. It's easy to see the weak existence is from Theorem 3.1.11. And the weak uniqueness is from its boundedness and Theorem 3.1.22.

3.1.5. Moments Formula for the Quadratic Rough Heston Model

Return to our quadratic rough Heston model, since the volatility process Z_t is an one dimensional polynomial Volterra process. Let dimension $d = 1$, we can first derive the moment formula for one dimensional polynomial Volterra process.

Proposition 3.1.27. *For $g_s(t)$ an adjusted forward process for X_t which is an one dimensional polynomial Volterra process. Then the moment formula of $g_s(t)$ can be* *expressed in*

$$\begin{aligned}\mathbb{E}(g_s(t)^N) =& g_0(t)^N + N\int_0^s b_0K(s-r)\mathbb{E}[g_r^{N-1}(t)]dr + N\int_0^s b_1K(s-r)\mathbb{E}[X_rg_r^{N-1}(t)]dr \\ &+ \frac{N(N-1)}{2}\Big(\int_0^s A_0K^2(s-r)\mathbb{E}[g_r^{N-2}(t)]dr \\ &+ \int_0^s A_1K^2(s-r)\mathbb{E}[X_rg_r^{N-2}(t)]dr + \int_0^s A_{11}K^2(s-r)\mathbb{E}[X_r^2g_r^{N-2}(t)]dr\Big),\end{aligned} \tag{3.32}$$

for $N \geq 2$*, where* A_0, A_1, A_{11} *and* b_0, b_1 *are associated coefficients for* X_t *as in Definition 3.1.1.*

In particular, let $s = t$, we can now express the moments for X_t in

$$\begin{aligned}\mathbb{E}(X_t^N) =& g_0(t)^N + N\int_0^t b_0K(t-r)\mathbb{E}[g_r^{N-1}(t)]dr + N\int_0^s b_1K(t-r)\mathbb{E}[X_rg_t^{N-1}(t)]dr \\ &+ \frac{N(N-1)}{2}\Big(\int_0^t A_0K^2(t-r)\mathbb{E}[g_r^{N-2}(t)]dr \\ &+ \int_0^t A_1K^2(t-r)\mathbb{E}[X_rg_r^{N-2}(t)]dr + \int_0^t A_{11}K^2(t-r)\mathbb{E}[X_r^2g_r^{N-2}(t)]dr\Big).\end{aligned} \tag{3.33}$$

And when $N = 1$,

$$\mathbb{E}(X_t) = \int_0^t b_0K(t-r)dr + \int_0^t b_1K(t-r)\mathbb{E}(X_r)dr.$$

In fact, let B be a matrix with columns equal to $b_1, \dots, b_d$ and let R_B be the resolvent of $-KB$, i.e. the solution to the linear equation $KB * R_B = KB + R_B$. Due to [AJLP19, Lemma 2.5], (3.4) is equivalent to the integral equation

$$X_t = g_0(t) - \int_0^t R_B(t-s)g_0(s)ds + \left(\int_0^t E_B(s)ds\right)b_0 + \int_0^t E_B(t-s)\sigma(X_s)dW_s,$$

where $E_B = K - R_B * K$. We see it offers us another way to express expectation of X in

$$\mathbb{E}(X_t) = g_0(t) - \int_0^t R_B(t-s)g_0(s)ds + b_0\int_0^t E_B(s)ds.$$

Now, Let $\overline{Z}$ be an approximation of Z. The dynamics of $\overline{Z}$ is given by

$$\overline{Z}_t = \int_0^t \overline{K}(t-s)b(\overline{Z}_s)ds + \int_0^t \overline{K}(t-s)\eta\sqrt{p(\overline{Z}_t)}dW_s, \tag{3.34}$$

where $b(x) = \theta - x$, $p(x) = a(x-b)^2 + c$, $a, b, c > 0$ and $\overline{K}$ is some approximation of K. Since $\overline{Z}$ is also a polynomial Volterra process, so it also has the moment formulas. rom the structure of the moment formulas, we conjecture that for any $n \in \mathbb{Z}^+$,

$$\left|\mathbb{E}[(Z_T)^n - (\overline{Z}_T)^n]\right| \lesssim \|K - \overline{K}\|_{L^1[0,T]} + \|K^2 - \overline{K}^2\|_{L^1[0,T]} \tag{3.35}$$

Moreover, let $\overline{X}$ denote the process with the same dynamics as in (3.2), where Z is replaced by $\overline{Z}$. We further conjecture that

$$\left|\mathbb{E}[X_T^n - \overline{X}_T^n]\right| \lesssim \int_0^T \left(\|K - \overline{K}\|_{L^1[0,t]} + \|K^2 - \overline{K}^2\|_{L^1[0,t]} \right) dt \tag{3.36}$$

since the diffusion coefficients of X has only polynomial growth. In the next section, we show that this conjecture indeed holds and can even be extended to more general test functions, not only power functions.

3.2. Error Bound for the Quadratic Rough Heston Model

In this section, we derive an error bound for the quadratic rough Heston model whose form is in a way similar to Corollary 2.1.11 in Chapter 2. However, as discussed in the Remark 3.1.4 in the previous subsection, the quadratic rough Heston model does not admit an semi-explicit characterization of its characteristic function. As a consequence, we cannot follow the same strategy as in the rough Heston case, where we estimated the characteristic function and then applied Fourier methods to derive the error bound.

Instead, we adopt a path-dependent PDE framework inspired by [BJP25] and [BMM25], and extend it to a broader class of volatility processes characterized as solutions to stochastic Volterra equations. This extension relies on developing the corresponding Malliavin calculus framework for stochastic Volterra equations, together with precise estimates of the associated Malliavin derivatives, which are essential for obtaining sharper error bounds since we will make extensive use of tools from Malliavin calculus. The resulting improved path-dependent PDE approach accommodates a wider range of rough volatility models with deterministic drift, and in particular applies to the quadratic rough Heston model.

First let see an example where we can derive a simple error bound for X. Assume the approximation (X^N, Z^N) satisfies the equation

$$dX_t^N = -\frac{1}{2} p(Z_t^N) dt + \sqrt{p(Z_t^N)} dW_t, \tag{3.37}$$

$$Z_t^N = \int_0^t K^N(t-s) b(Z_s^N) ds + \int_0^t K^N(t-s) \eta \sqrt{p(Z_t^N)} dW_s, \tag{3.38}$$

where K^N is an approximation for K satisfying the conditions in [AK21, Theorem 3.1]. Then we have the following result:

Proposition 3.2.1. *Assume $K^N(t) \le CK(t)$ on $(0,T]$ for every N. Then there exists a constant $C \in \mathbb{R}_+$ depending on T, b, σ, $K|_{[0,T]}$ such that*

$$\mathbb{E}[|X_T^N - X_T|] \le C \int_0^T \|K^N(s) - K(s)\|_{L^2[0,t]} dt. \tag{3.39}$$

Proof. Note the coefficients b, σ of the SVEs (3.3) and (3.38) satisfy Lipschitz continuous condition of [AK21, Theorem 3.1], so we have

$$\mathbb{E}[|Z_t^N - Z_t|^2] \leq C \int_0^t |K^N(s) - K(s)|^2 ds$$

and

$$\mathbb{E}[|Z_t^N - Z_t|^2] \geq \mathbb{E}[|Z_t^N - Z_t|]^2.$$

So

$$\mathbb{E}[|Z_t^N - Z_t|] \leq C\|K^N - K\|_{L^2[0,t]}.$$

Eventually, with using the Lipschitz continuity of $\sqrt{p(x)}$ and applying BDG inequality and Hölder inequality, we have

$$\begin{aligned}
\mathbb{E}[|X_T^N - X_T|] &\leq \frac{1}{2}\mathbb{E}[\int_0^T |p(Z_t) - p(Z_t^N)|dt] + \mathbb{E}[\int_0^T |\sqrt{p(Z_t)} - \sqrt{p(Z_t^N)}|dt] \\
&\leq C\mathbb{E}[(\int_0^T |Z_t - Z_t^N|^2 dt)^{1/2}]\mathbb{E}[(\int_0^T |Z_t + Z_t^N|^2 dt)^{1/2}] \\
&\quad + C\int_0^T \mathbb{E}[|Z_t - Z_t^N|]dt \\
&\leq C_{K,T}\mathbb{E}[(\int_0^T |Z_t - Z_t^N|^2 dt)^{1/2}] + C\int_0^T \mathbb{E}[|Z_t - Z_t^N|]dt \\
&\leq C_{K,T}\int_0^T \mathbb{E}[|Z_t - Z_t^N|]dt \\
&\leq C_{K,T}\int_0^T \|K^N - K\|_{L^2[0,t]}dt,
\end{aligned} \tag{3.40}$$

where we use Proposition 3.1.5 and the assumption $K^N \leq K$. □

In the following, we extend the result to a more general setting and obtain an estimate given by

$$|\mathbb{E}[\phi(X_T)] - \mathbb{E}[\phi(X_T^N)]| \leq C\int_0^T \left\{\|K - K^N\|_{L^1[0,t]} + \|K^2 - (K^N)^2\|_{L^1[0,t]}\right\}dt.$$

3.2.1. Preliminary

Notations

If $(X, \mathcal{X})$ and $(Y, \mathcal{Y})$ are two topological spaces, we denote $C^0(X, Y)$ (resp. $D^0(X, Y)$) by the space of continuous functions (resp. càdlàg functions) from X to Y. If $(X, \mathcal{X})$ and $(Y, \mathcal{Y})$ are two normed vector spaces and X is also a measure space with measure μ_X, we denote by $L^p(X, Y)$ the space of p-integrable functions from (X, μ_X) to Y. For $Y = \mathbb{R}$, we simply write $L^p(X, Y) = L^p(X)$.

If x and y are two real number, we note $x \wedge y$ (resp. $x \vee y$) the minimum (resp. the maximum) between x and y.

We write $a \lesssim b$ if there exists a constant $C > 0$ such that $a \leq Cb$.

Let $\partial^j_{x^j} := d^j/dx^j$, $j-$th order partial differential operator with respect to the variable x. And we denote $\partial_{x_j} := \partial_j$ for convention. If $x \in \mathbb{R}$, then we denote the $j-th$ derivative of $f(x)$ by $f^{(j)}(x)$.

Let $\omega : [0,T] \to \mathbb{R}$ and $\theta : [0,T] \to \mathbb{R}$ be Borel-measurable functions. We define their concatenation path at time t by

$$\forall u \in [0,T], \qquad (\omega \oplus_t \theta)_u = \begin{cases} \omega_u, & \text{if } u \leq t, \\ \theta_u, & \text{if } u > t. \end{cases}$$

For a measurable function K on $\mathbb{R}_+$ and a measure L on $\mathbb{R}_+$ of locally bounded variation, the convolutions $K * L$ and $L * K$ are defined by

$$(K * L)(t) = \int_{[0,t]} K(t-s)L(ds), \quad (L * K)(t) = \int_{[0,t]} L(ds)K(t-s)$$

for $t \geq 0$. If F is a function on $\mathbb{R}_+$, we write $K * F = K * (Fdt)$, that is,

$$(K * F)(t) = \int_0^t K(t-s)F(s)ds.$$

Fix $d \in \mathbb{N}$ and let M be a d-dimensional continuous local martingale. If K is $\mathbb{R}^{m \times d}$-valued for some $m \in \mathbb{N}$, the convolution

$$(K * dM)_t = \int_0^t K(t-s)dM_s$$

is well-defined as an Itô integral for any $t \geq 0$ if

$$\int_0^t |K(t-s)|^2 d\ \mathrm{tr}[M]_s < \infty.$$

We note $\mathbf{D}$ the Malliavin derivative operator with respect to the 1-dim Brownian motion B and $\mathbb{D}^{1,2}(B)$ its domain of application in $L^2(\Omega)$.

We consider a function $K \in L^2([0,T])$, called kernel, and a function $\overline{K} \in L^2([0,T])$, called approximate kernel. We will extensively make use of the following notations in the proofs:

$$\begin{aligned} \Delta K &:= \overline{K} - K, \qquad & \Delta(K^2) &:= \overline{K}^2 - K^2, \\ \Sigma K &:= \overline{K} + K, \qquad & \Sigma(K^2) &:= \overline{K}^2 + K^2. \end{aligned} \tag{3.41}$$

Malliavin Calculus

The treatment of singular kernels in the next section requires the use of Malliavin calculus, we will introduce some basic results for Malliavin calculus in the following, including the basic properties of the Malliavin derivative and several useful theorems and propositions, following the lecture notes [Kun13] and [Hai21].

Definition 3.2.2 (Isonormal Gaussian process over H)**.** Let H be a real separable Hilbert space with inner product $\langle\cdot,\cdot\rangle$. An isonormal Gaussian process over H is a family of real-valued random variables $W = \{W(h) : h \in H\}$ defined on a probability space $(\Omega, \mathcal{F}, \mathbb{P})$ such that:

- for any $h_1, \ldots, h_n \in H$, the vector $(W(h_1), \ldots, W(h_n))$ is jointly Gaussian;
- for all $h, g \in H$,
$$\mathbb{E}[W(h)W(g)] = \langle h, g\rangle.$$

Remark 3.2.3. Gaussian white noise is the special case of H-isonormal Gaussian process, where we take $H = L^2(\mathbb{R}_+)$. So the Brownian motions belong to the class of H-isonormal Gaussian processes. More generally, fractional Brownian motions also belong to the class of H-isonormal Gaussian processes as long as we choose another inner product $\langle\cdot,\cdot\rangle_H$.

Fix a H-isonormal Gaussian process W, we denote the underlying probability space by $(\Omega, \Sigma, \mathbb{P})$, where $\Sigma := \sigma(W(h) : h)$, the $\sigma-$algebra generated by the H-isonormal Gaussian process W. By $\mathcal{S}$, the set of "smooth random variables" (resp. $\mathcal{S}_b$, resp. $\mathcal{S}_c$) we denote the class of random variables X such that there exists an $n \in \mathbb{N}$, vectors $h_1, \ldots, h_n$ and a function $f \in C^\infty_{\text{pol}}(\mathbb{R}^n)$ (resp. $f \in C^\infty_B(\mathbb{R}^n)$, resp. $f \in C^\infty_c(\mathbb{R}^n)$) such that
$$X = f(W(h_1), \ldots, W(h_n)).$$
And we remark here all the three spaces are dense in $L^2(\Omega)$.

Definition 3.2.4. Let $X = f(W(h_1), \ldots, W(h_n)) \in \mathcal{S}$. The Malliavin derivative of X is the H-valued random variable $\mathbf{D}X$, defined by
$$\mathbf{D}X := \sum_{j=1}^{n} \partial_j f(W(h_1), \ldots, W(h_n))h_j,$$
where $\partial_j f$ denotes the j-th partial derivative of f.

Note that for $h \in H$, we have
$$\langle \mathbf{D}X, h\rangle = \lim_{t\to 0} \frac{f(W(h_1) + t\langle h_1, h\rangle, \ldots, W(h_n) + t\langle h_n, h\rangle) - f(W(h_1), \ldots, W(h_n))}{t}.$$

Remark 3.2.5. The derivative is independent of choice of the representation for
$$X = f(W(h_1), \ldots, W(h_n)) \quad \text{or} \quad g(W(e_1), ..., W(e_m)).$$

Example 3.2.6. Consider the case of Brownian motion, i.e., take $H = L^2(\mathbb{R}_+)$, we have $B_t = W(\mathbf{1}_{[0,t]})$, so $\mathbf{D}_s B_t = \mathbf{1}_{[0,t]}(s)$. More generally, we often write $X = W(h) := \int_0^\infty h(t)dB_t$, the Ito integral, where B_t is a Brownian motion. Then $\mathbf{D}_s X = h(s)$.

Next, we introduce some basic results for Malliavin derivatives. Most of proofs can be found in Appendix A.3.

Lemma 3.2.7 (Integration by part)**.** *Let $X \in \mathcal{S}$ and $h \in H$. Then* $\mathbb{E}(\langle \mathbf{D}X, h\rangle) = \mathbb{E}(XW(h))$.

Corollary 3.2.8. *Let $X, Y \in \mathcal{S}$ and $h \in H$. Then*

$$\mathbb{E}(Y\langle \mathbf{D}X, h\rangle) = \mathbb{E}(XYW(h) - X\langle \mathbf{D}Y, h\rangle).$$

Proposition 3.2.9. *The operator $\mathbf{D}$ viewed as an operator from $L^p(\Omega)$ to $L^p(\Omega; H)$, defined on $\mathcal{S}$ is closable.*

Definition 3.2.10. we denote the closure of $\mathbf{D}$ in $L^p(\Omega)$ again by $\mathbf{D}$ and call it Malliavin differential operator. The domain of $\mathbf{D}$ in $L^p(\Omega)$ is denoted by $\mathbb{D}^{1,p}$. Thus $\mathbb{D}^{1,p}$ is the closure of $\mathcal{S}$ with respect to the norm

$$\|X\|_{1,p} := \left(\mathbb{E}|X|^p + \mathbb{E}\|\mathbf{D}X\|_H^p\right)^{\frac{1}{p}}.$$

Remark 3.2.11. For $p = 2$, the space $\mathbb{D}^{1,2}$ is a Hilbert space with respect to the inner product

$$\langle X, Y\rangle_{1,2} = \mathbb{E}(XY) + \mathbb{E}\langle \mathbf{D}X, \mathbf{D}Y\rangle_H.$$

Remark 3.2.12. For $p \in (1, \infty)$ the space $\mathbb{D}^{1,p}$ is reflexive since it's isometrically isomorphic to a closed subspace of the reflexive space $L^p(\Omega) \times L^p(\Omega; H)$.

Definition 3.2.13. The *divergence operator*, also called the *Skorohod integral*, is the adjoint δ of $\mathbf{D}$ on $L^2(\Omega)$ with the domain $D(\delta)$ consisting of those $u \in L^2(\Omega; H)$ such that there exists $X \in L^2(\Omega)$ with

$$\mathbb{E}\langle \mathbf{D}Y, u\rangle_H = \mathbb{E}(Y \cdot X)$$

for all $Y \in \mathbb{D}^{1,2}$. Since $\mathbb{D}^{1,2}$ is dense in $L^2(\Omega)$, there is at most one such element X. We write $\delta(u) := X$.

Remark 3.2.14. For $u \in L^2(\Omega; H)$, it belongs to the domain $D(\delta)$ if and only if there exists a constant $c \geq 0$ st.

$$\left|\mathbb{E}\langle \mathbf{D}Y, u\rangle\right| \leq c\left(\mathbb{E}|Y|^2\right)^{\frac{1}{2}}$$

for all $X \in \mathbb{D}^{1,2}$.

Proposition 3.2.15. *The divergence operator δ is local on $L^2(\Omega; H)$.*

Proposition 3.2.16 (Chain rule for Malliavin derivatives)**.** *Let $p \geq 1$, $X = (X_1, \dots, X_m)$ be a vector of random variables with $X_j \in \mathbb{D}^{1,p}$ for $j = 1, \dots, m$ and $\varphi : \mathbb{R}^m \to \mathbb{R}$ be continuously differentiable with bounded partial derivatives. Then $\varphi(X) \in \mathbb{D}^{1,p}$ and*

$$\mathbf{D}\varphi(X) = \sum_{j=1}^{m} \partial_j \varphi(X) \mathbf{D}X_j. \tag{3.42}$$

Lemma 3.2.17. *Let $p \in (1, \infty)$ and $X_n \in \mathbb{D}^{1,p}$ be such that $X_n \to X$ in $L^p(\Omega)$ and such that*

$$\sup_{n\in\mathbb{N}} \mathbb{E}\|\mathbf{D}X_n\|_H^p < \infty.$$

Then $X \in \mathbb{D}^{1,p}$ and $\mathbf{D}X_n$ converges weakly to $\mathbf{D}X$.

Next we study the special case where $H = L^2([0,T], \mathscr{B}_{[0,T]}, dt)$. Given a random variable $X \in \mathbb{D}^{1,2}$, the Malliavin derivative $\mathbf{D}X$ can be identified with the element in $L^2([0,T] \times \Omega)$. Thus the Malliavin derivative can be viewed as a stochastic process $\{\mathbf{D}_t X : t \in [0,T]\}$ where $\mathbf{D}_t X$ is defined almost everywhere with respect to the measure $dt \otimes \mathbb{P}$. Consider the case $X = f(W(h_1), \ldots, W(h_n)) \in \mathcal{S}$, we have

$$\mathbf{D}_t X = \sum_{i=1}^{n} (\partial_i f)(W(h_1), \ldots, W(h_n))h_i(t).$$

And similar arguments also apply to the divergence operator δ. If we view the Malliavin derivative as an operator from $L^2(\Omega)$ to $L^2([0,T] \times \Omega)$, then the adjoint δ is an operator on $L^2([0,T] \times \Omega)$ taking values in $L^2(\Omega)$. Thus the domain of δ consists of certain stochastic processes. In this case, we have the following conditional integration by part formula.

Lemma 3.2.18 (Conditional integration by part)**.** *The (conditional) Malliavin integration by parts formula is*

$$\mathbb{E}\left[F \int_t^T h_s \, dB_s \Big| \mathcal{F}_t\right] = \mathbb{E}\left[\int_t^T h_s \mathbf{D}_s F ds \Big| \mathcal{F}_t\right], \tag{3.43}$$

for any $F \in \mathbb{D}^{1,2}$ and $h \in L^2(\Omega \times [0,T])$.

Proposition 3.2.19. *For $A \in \mathscr{B}_{[0,T]}$, $\Sigma_A := \sigma(W(\mathbf{1}_B) : B \subset A, B \in \mathscr{B}_{[0,T]})$ and $X \in \mathbb{D}^{1,2}$. We have $\mathbb{E}(X|\Sigma_A) \in \mathbb{D}^{1,2}$ and*

$$\mathbf{D}_t \mathbb{E}(X|\Sigma_A) = \mathbb{E}(\mathbf{D}_t X|\Sigma_A)\mathbf{1}_A.$$

Corollary 3.2.20. *Let $A \in \mathscr{B}_{[0,T]}$ and assume that $X \in \mathbb{D}^{1,2}$ is Σ_A-measurable. Then*

$$\mathbf{D}_t X = 0 \quad a.s. \text{ on } \quad A^c \times \Omega.$$

Remark 3.2.21. It asserts that if X is $\mathscr{F}_t$-measurable, then the Malliavin derivative $\mathbf{D}_s X$ is only supported in the interval $[0,t]$, i.e. $\mathbf{D}_s X = 0$ for $s > t$.

We now denote the space $\mathbb{D}^{1,2}(L^2([0,T], \mathscr{B}, dt)) =: \mathbb{L}^{1,2}$, i.e., $\mathbb{L}^{1,2} = \{u \in L^2([0,T] \times \Omega) : u(t) \in \mathbb{D}^{1,2}, \mathbf{D}_s u(t) \in L^2([0,T]^2 \times \Omega)\}$, which is contained in $D(\delta)$. And $\mathbb{L}^{1,2}$ is a Hilbert space with norm

$$\|u\|_{\mathbb{L}^{1,2}}^2 := \|u\|_{L^2([0,T]\times\Omega)}^2 + \|\mathbf{D}u\|_{L^2([0,T]^2\times\Omega)}^2.$$

We set $B_t := W(\mathbf{1}_{(0,t]})$ and consider the natural filtration $\mathbb{F} = (\mathscr{F}_t)_{T \geq t \geq 0}$ where $\mathscr{F}_t = \Sigma_{(0,t]} = \sigma(B_s : s \leq t)$. Let $L^2_{\mathbb{F}}([0,T] \times \Omega)$ be the space of the adapted, square integrable processes.

Proposition 3.2.22. *We have $L^2_{\mathbb{F}}([0,T]\times\Omega) \subset D(\delta)$.*

Remark 3.2.23. For $u \in L^2_{\mathbb{F}}((0,T)\times\Omega)$ it is customary to write

$$\int_0^T u(t)dB_t$$

for $\delta(u)$ and to call this the Itô-integral of u. Actually, any Itô-integrable stochastic process is in the domain of the Skorohod integral and the Skorohod integral coincides with the Itô integral.

Proposition 3.2.24 (Malliavin derivative for Itô integral)**.** *Let $u \in L^2_{\mathbb{F}}((0,T)\times\Omega)$ and define $X := \int_0^T u(s)dB_s$. Then $X \in \mathbb{D}^{1,2}$ if and only if $u \in \mathbb{L}^{1,2}$. In that case, $t \mapsto \mathbf{D}_t u(s)$ belongs to $L^2_{\mathbb{F}}$ and for $t \in (0,T)$ we have*

$$\mathbf{D}_t X = u(t) + \int_t^T \mathbf{D}_t u(s) dB_s$$

almost surely.

Remark 3.2.25. It will be useful to be applied together with the Lemma 3.2.17.

Hadamard derivatives in $L^p(\Omega)$

Next, we introduce Hadamard derivatives, which provide the weakest notion of differentiability ensuring the validity of the chain rule in $L^p(\Omega)$. In this setting, the chain rule does not generally hold for Gâteaux derivatives, and the functions under consideration are not necessarily Fréchet differentiable. We recall the definition of Hadamard differentiability and some basic properties from [BMM25, Appendix A.1] and [Sha90].

Definition 3.2.26. Let (X, d_X) and (Y, d_Y) be two metric vector spaces and $f : X \to Y$. f is called Hadamard-differentiable at $x \in X$ in the direction $h \in X$ if, for any sequence $(h_n)_{n\in\mathbb{N}}$ of elements of X and $(t_n)_{n\in\mathbb{N}}$ of elements of $\mathbb{R}_+$ such that $h_n \to h$ and $t_n \to 0$, the limit of the sequence

$$\left(\frac{f(x + t_n h_n) - f(x)}{t_n}\right)_{n\in\mathbb{N}}$$

exists in Y. We denote this limit by $Df(x,h)$ and call it the Hadamard derivative of f at x in the direction h.

Remark 3.2.27. When f is Hadamard differentiable at x in the direction h, then f is also Gâteaux differentiable at x in the direction h, and the Gâteaux derivative coincides with the Hadamard derivative. Actually, we have the following relation for the three commonly used differentiabilities

Fréchet differentiable $\Rightarrow$ Hadamard differentiable $\Rightarrow$ Gâteaux differentiable.

Besides, for locally Lipschitz mappings in normed spaces, the Gâteaux derivative will coincide with the Hadamard derivative.

Proposition 3.2.28. *If f is Hadamard differentiable at x, then its Hadamard derivative $Df(x, \cdot)$ is is sequentially continuous.*

Remark 3.2.29. So if the domain space X is metrizable, then Hadamard directional differentiability implies continuity of the directional derivative Df_x.

We have the chain rule and product rule for the Hadamard derivatives in the following. Those results are from [BMM25, Appendix A.1].

Proposition 3.2.30 (The chain rule for Hadamard derivatives)**.** *Let X, Y, Z be metric spaces, $f : X \to Y$ and $g : Y \to Z$. If f is Gâteaux-differentiable at $x \in X$ in the direction $h \in X$ and g is Hadamard-differentiable at $f(x) \in Y$ in the direction $Df(x, h) \in Y$, then $g \circ f$ is Gâteaux-differentiable at x in the direction h and*

$$D(g \circ f)(x, h) = Dg(f(x), Df(x, h)).$$

Lemma 3.2.31 (The product rule for Hadamard derivatives)**.** *Let (X,d) be a metric space. If two functions $f, g : X \to L^2(\Omega)$ are Gâteaux-differentiable at $x \in X$ in the direction $h \in X$, then their product $fg : X \to L^1(\Omega)$ is Gâteaux-differentiable for the norm $\| \cdot \|_1$ at x in the direction h and the Gâteaux derivative is given by $D(fg)(x, h) = f(x)Dg(x, h) + g(x)Df(x, h)$.*

The following lemma will be vital in the proof of Proposition 3.54 in the next subsection, which provides a sufficient condition ensuring the Hadamard differentiability of the extension of a smooth function $f : \mathbb{R} \to \mathbb{R}$ to the appropriate probabilistic spaces in our specific context.

Lemma 3.2.32. *Let $f \in C^2(\mathbb{R}, \mathbb{R})$, and assume that, for $g \in \{f, f', f''\}$, there exists a real $p \geq 1$ such that $|g(x)| \lesssim 1 + |x|^p$ for any $x \in \mathbb{R}$. Fix $r \geq 1$. For $q \geq 2pr \vee 4r$, we consider the functions $F : L^q(\Omega) \to L^r(\Omega)$ defined by $F(X)(\omega) = f(X(\omega))$ for $X \in L^q(\Omega)$ and $\omega \in \Omega$. Then F is a well-defined Hadamard-differentiable function between the Banach spaces $(L^q(\Omega), \|\cdot\|_q)$ and $(L^r(\Omega), \|\cdot\|_r)$, and the Hadamard derivative of F at $X \in L^q(\Omega)$ in the direction $H \in L^q(\Omega)$ is given by $F'(X, H) = f'(X)H$.*

Resolvent

Note that stochastic Volterra equations can be viewed as stochastic differential equations convolved with a kernel K. It is therefore useful to introduce some results concerning the resolvent.

Definition 3.2.33. Consider a kernel $K \in L^1_{\text{loc}}(\mathbb{R}_+, \mathbb{R}^{d \times d})$. The *resolvent*, or *resolvent of the second kind*, corresponding to K is the kernel $R \in L^1_{\text{loc}}(\mathbb{R}_+; \mathbb{R}^{d \times d})$ such that

$$K * R = R * K = K - R. \tag{3.44}$$

Remark 3.2.34. The resolvent always exists and is unique, and several properties of the original kernel K, such as local integrability and continuity, are inherited by the resolvent; see [GLS90, Theorems 2.3.1 and 2.3.5].

Definition 3.2.35. The *resolvent of the first kind* associated with the kernel K is an $\mathbb{R}^{d\times d}$-valued measure L on $\mathbb{R}_+$ of locally bounded variation such that

$$K * L = L * K \equiv I,$$

where I stands for the identity matrix.

Remark 3.2.36. A resolvent of the first kind does not always exist.

Using the concept of resolvent of the second kind, one can derive a variation of constants formula as shown in the following lemma, which is a special case of [AJLP19, Lemma 2.5] with $Z = 0$.

Lemma 3.2.37 (Resolvent Representation)**.** *Let X be a continuous process, $F\colon \mathbb{R}_+ \to \mathbb{R}^m$ a continuous function, $B \in \mathbb{R}^{d\times d}$. Then*

$$X = F + (KB) * X \qquad \Longleftrightarrow \qquad X = F - R_B * F,$$

where R_B is the resolvent of the second kind for $-KB$.

As in the classical ODE setting, where Gronwall's inequality is frequently used to derive estimates, we will need an analogue in the present context. A suitable Volterra-type Gronwall inequality is given in [GLS90, Lemma 9.8.2].

Lemma 3.2.38 (Volterra-type Grownwall's inequality)**.** *Let $k \in L^p_{\mathrm{loc}}$ be a scalar Volterra kernel (i.e., $k(t,s) = 0$ for $s > t$ and $(k * x)(t) := \int_0^t k(t,s)x(s)ds$), for some p, $1 \le p \le \infty$, on an interval $J := [S,T) \subset \mathbb{R}$, with $-\infty < S < T \le \infty$, and assume that $-k$ has a nonpositive resolvent of the second kind $r \in L^p_{\mathrm{loc}}$. Let $x, f \in L^p_{\mathrm{loc}}(J)$, and suppose that*

$$x(t) \le (k * x)(t) + f(t)$$

for a.e. $t \in J$. Then $x(t) \le y(t)$ for a.e. $t \in J$, where y is the solution of the comparison equation

$$y(t) = (k * y)(t) + f(t).$$

Remark 3.2.39. By applying Lemma 3.2.37, we can express solution y of the comparison equation explicitly in terms of the resolvent of k.

Function Spaces and Derivatives

We fix a real $t \in [0,T]$, and define

$$\begin{aligned}
&\mathcal{X} \coloneqq C^0([0,T],\mathbb{R}), \qquad \mathcal{X}_t \coloneqq C^0([t,T],\mathbb{R}), \qquad \overline{\mathcal{X}} \coloneqq D^0([0,T],\mathbb{R});\\
&\Lambda \coloneqq [0,T]\times\mathbb{R}\times C^0([0,T],\mathbb{R});\\
&\bar{\Lambda} \coloneqq \Big\{(t,x,\omega) \in [0,T]\times\mathbb{R}\times D^0([0,T],\mathbb{R})\ ;\ \omega_{|_{[t,T]}} \in \mathcal{X}_t\Big\};\\
&\|\omega\|_{[0,T]} \coloneqq \sup_{t\in[0,T]} |\omega_t|, \qquad d_{\bar{\Lambda}}(\ (t,x,\omega),(t',x',\omega')\ ) \coloneqq |t-t'| + |x-x'| + \|\omega-\omega'\|_{[0,T]}.
\end{aligned}$$

Let $C^0(\bar{\Lambda}) \coloneqq C^0(\bar{\Lambda},\mathbb{R})$ be the space of functions $u : \bar{\Lambda} \to \mathbb{R}$ such that u is continuous with respect to the topology induced by the distance $d_{\bar{\Lambda}}$. Next, we define several derivatives for functions $u \in C^0(\bar{\Lambda})$ in the sense of Gâteaux. We define the time derivative of a function $u \in C^0(\bar{\Lambda})$ at $(t,x,\omega) \in \bar{\Lambda}$ as the following limit

$$\partial_t u(t,\omega) = \lim_{\delta\to 0} \frac{u(t+\delta,\omega) - u(t,\omega)}{\delta},$$

if it exists. Let $(t,x) \in [0,T]\times\mathbb{R}$ and $\eta \in \mathcal{X}_t$. We define the path derivative of u at $\omega \in \mathcal{X}$ in the direction η by

$$\langle \partial_\omega u(t,x,\omega), \eta\rangle = \lim_{\varepsilon\to 0} \frac{u(t,x,\omega+\varepsilon\eta\mathbb{1}_{[t,T]}) - u(t,x,\omega)}{\varepsilon},$$

if the limit exists. For any $r < t$ and $\eta \in \mathcal{X}_r$, we fix the convention that

$$\langle \partial_\omega u(t,x,\omega), \eta\rangle \coloneqq \langle \partial_\omega u(t,x,\omega), \eta\mathbb{1}_{[t,T]}\rangle.$$

Similarly, we define the second path derivative of u at point $(t,x,\omega) \in \bar{\Lambda}$, in the directions $\eta, \zeta \in \mathcal{X}_t$ by

$$\langle \partial^2_{\omega\omega} u(t,x,\omega), (\eta,\zeta)\rangle = \lim_{\varepsilon\to 0} \frac{\langle \partial_\omega u(t,x,\omega+\varepsilon\zeta\mathbb{1}_{[t,T]}),\eta\rangle - \langle \partial_\omega u(t,x,\omega),\eta\rangle}{\varepsilon},$$

if the limit exists.

Remark 3.2.40. More generally, if we consider $U(t,\boldsymbol{\omega}) : [0,T]\times V \mapsto \mathbb{R}$, V is some n.v.s., and we define its Gâteaux derivatives at variable $\boldsymbol{\omega}$ as above. Then we have the following results: if $\boldsymbol{\omega}$ is $\mathbb{R}^d$-valued, then for $\boldsymbol{\eta}$ is $\mathbb{R}^d$-valued, we have

$$\begin{aligned}
\langle \partial_{\boldsymbol{\omega}} U(t,\boldsymbol{\omega}), \boldsymbol{\eta}\rangle &= \sum_{i=1}^d \langle \partial_{\omega_i} U(t,\boldsymbol{\omega}), \eta_i\rangle,\\
\langle \partial^2_{\boldsymbol{\omega}\boldsymbol{\omega}} U(t,\boldsymbol{\omega}), (\boldsymbol{\eta},\boldsymbol{\eta})\rangle &= \sum_{i,j=1}^d \langle \partial^2_{\omega_i\omega_j} U(t,\boldsymbol{\omega}), (\eta_i,\eta_j)\rangle.
\end{aligned}$$

Furthermore, if $\boldsymbol{\eta} \in \mathbb{R}^{d\times m}$ is matrix valued, then

$$\langle \partial_{\boldsymbol{\omega}} u(t,\boldsymbol{\omega}), \boldsymbol{\eta}\rangle = \left(\sum_{i=1}^d \langle \partial_{\omega_i} u(t,\boldsymbol{\omega}), \eta_{i,j}\rangle\right)_{j=1}^m,$$

$$\langle \partial^2_{\boldsymbol{\omega}\boldsymbol{\omega}} u(t,\boldsymbol{\omega}), (\boldsymbol{\eta},\boldsymbol{\eta})\rangle = \sum_{k=1}^{m}\sum_{i,j=1}^{d} \langle \partial^2_{\omega_i\omega_j} u(t,\boldsymbol{\omega}), (\eta_{j,k},\eta_{i,k})\rangle.$$

Definition 3.2.41. We say that $u \in C^{0,2,2}(\bar{\Lambda}) \subset C^0(\bar{\Lambda})$ if the derivatives $\partial_x u$, $\partial^2_{xx} u$, ∂_ω, $\partial^2_{\omega\omega} u$ and $\partial_\omega(\partial_x u)$ exist and are continuous functions (at fixed direction) on $\bar{\Lambda}$ with respect to the distance $d_{\bar{\Lambda}}$, and if the maps $\eta \mapsto \langle \partial_\omega u(t,x,\omega), \eta\rangle$, $\zeta \mapsto \langle \partial^2_\omega u(t,x,\omega), (\eta,\zeta)\rangle$ and $\eta \mapsto \langle \partial_\omega(\partial_x u)(t,x,\omega), \eta\rangle$ are linear and continuous on $\mathcal{X}$.

And we say that $u \in C^{1,2,2}(\bar{\Lambda})$ if $u \in C^{0,2,2}(\bar{\Lambda})$ and, moreover, its time derivative $\partial_t u$ is continuous.

Remark 3.2.42. (i) Due to the definition of $d_{\bar{\Lambda}}$, it is enough to show that the derivatives of u are continuous respectively with respect to t, x and ω separately to ensure that they are continuous under $d_{\bar{\Lambda}}$ on $\bar{\Lambda}$.

(ii) Fréchet-differentiable condition of u, $\partial_\omega u$ and $\partial_x u$ can imply the linearity and the continuity of the path derivatives with respect to the direction. However, the conditions in Definition 3.2.41 are sufficient in our setting, so we choose to work with the Hadamard derivative in the proofs. This allows the chain rule to be applied while remaining well-suited to the framework of our study.

Definition 3.2.43. Given $\boldsymbol{X}_t$ satisfying some stochastic Volterra equation on $[0,T]$, we denote the space of control functions by

$$\mathcal{G} := \left\{ G : \mathbb{R}^d \to \mathbb{R} : \mathbb{E}\left[G(\|\boldsymbol{X}\|^p_{[0,T]})\right] < \infty \text{ for all } p > 1 \right\}.$$

Remark 3.2.44. We use the same definition as [BJP25, Definition 2.2] here. We will see that $f(t,x,\omega) = 1 + |x|^q + \|\omega\|^q_{[0,T]}$ belongs to the space of control functions from the following Lemma 3.2.53.

Definition 3.2.45. Let $u \in C^{0,2,2}(\bar{\Lambda})$.

(i) **Controlled growth for** u, $\partial_x u$, $\partial^2_{xx} u$. We say that u, $\partial_x u$ or $\partial^2_{xx} u$ has *controlled growth* if there exists a constant $q > 0$ such that

$$\left|\partial^i_{x^i} u(t,x,\omega)\right| \lesssim 1 + |x|^q + \|\omega\|^q_{[0,T]}, \qquad \forall (t,x,\omega) \in \bar{\Lambda},$$

for $i = 0$, 1 or 2 respectively.

(ii) **Controlled growth for** $\partial_\omega u$, $\partial_\omega(\partial_x u)$. We say that $\partial_\omega u$ or $\partial_\omega(\partial_x u)$ has *controlled growth* if there exists $q > 0$ such that

$$\left|\langle \partial_\omega \partial^j_{x^j} u(t,x,\omega),\, \eta\rangle\right| \lesssim \left(1 + |x|^q + \|\omega\|^q_{[0,T]}\right) \|\eta\, \mathbf{1}_{[t,T]}\|_{[0,T]},$$

for all $(t,x,\omega) \in \bar{\Lambda}$ and $\eta \in \mathcal{X}$, with $j = 0$ or 1.

(iii) **Controlled growth for** $\partial^2_{\omega\omega}u$. We say that $\partial^2_{\omega\omega}u$ has *controlled growth* if there exists $q > 0$ such that

$$\left|\langle \partial^2_{\omega\omega}u(t,x,\omega),(\eta,\zeta)\rangle\right| \lesssim \left(1+|x|^q+\|\omega\|^q_{[0,T]}\right)\|\eta\,\mathbf{1}_{[t,T]}\|_{[0,T]}\,\|\zeta\,\mathbf{1}_{[t,T]}\|_{[0,T]},$$

for all $(t,x,\omega)\in\bar\Lambda$ and $\eta,\zeta\in\mathcal{X}$.

We say $u\in C^{0,2,2}_{+}(\bar\Lambda)$ if u and all the derivatives above have controlled growth.

Remark 3.2.46. The conditions of the Definition 3.2.45 are stronger than [BJP25, Definition 2.4] since we restrict on a smaller family of functions now. We will use the same family of of control functions for Definition 3.2.47 in the following.

To ensure that the functional Itô formula and the path-dependent PDE hold in the Section 3.2.5, we impose additional regularity assumptions on u. The following conditions ensure that the path derivatives of u are negligible when evaluated in directions supported near the time singularity.

Definition 3.2.47. We say that $u\in C^{0,2,2}_{+,\alpha}(\Lambda)$, with $\alpha\in(0,1)$, if there exists a continuous extension of $u\in C^{0,2,2}_{+}(\overline{\Lambda})$, a constant $q>0$ and a bounded modulus of continuity function ϱ such that, for any $x\in\mathbb{R}$, for any $0\le t<T$, $0<\delta\le T-t$, and $\eta,\zeta\in\mathcal{X}_t$ with supports contained in $[t,t+\delta]$,

(i) for any $\omega\in\overline{\mathcal{X}}$ such that $\omega\mathbf{1}_{[t,T]}\in\mathcal{X}_t$,

$$\begin{aligned}|\langle\partial_\omega u(t,x,\omega),\eta\rangle| &\le (1+|x|^q+\|\omega\|^q_{[0,T]})\,\|\eta\|_{[0,T]}\,\delta^\alpha,\\ \left|\langle\partial^2_{\omega\omega}u(t,x,\omega),(\eta,\zeta)\rangle\right| &\le (1+|x|^q+\|\omega\|^q_{[0,T]})\,\|\eta\|_{[0,T]}\,\|\zeta\|_{[0,T]}\,\delta^{2\alpha}.\end{aligned}$$

(ii) for any other $\omega'\in\overline{\mathcal{X}}$ such that $\omega'\mathbf{1}_{[t,T]}\in\mathcal{X}_t$,

$$\begin{aligned}&|\langle\partial_\omega u(t,x,\omega)-\partial_\omega u(t,x,\omega'),\eta\rangle|\\ &\qquad\le (1+|x|^q+\|\omega\|^q_{[0,T]}+\|\omega'\|^q_{[0,T]})\,\|\eta\|_{[0,T]}\,\varrho(\|\omega-\omega'\|_{[0,T]})\delta^\alpha,\\ &\left|\langle\partial^2_{\omega\omega}u(t,x,\omega)-\partial^2_{\omega\omega}u(t,x,\omega'),(\eta,\zeta)\rangle\right|\\ &\qquad\le (1+|x|^q+\|\omega\|^q_{[0,T]}+\|\omega'\|^q_{[0,T]})\,\|\eta\|_{[0,T]}\,\|\zeta\|_{[0,T]}\,\varrho(\|\omega-\omega'\|_{[0,T]})\,\delta^{2\alpha}.\end{aligned}$$

(iii) For any $\omega\in\overline{\mathcal{X}}$, $t\mapsto\langle\partial_\omega u(t,x,\omega),\eta\rangle$ and $t\mapsto\langle\partial^2_{\omega\omega}u(t,x,\omega),(\eta,\zeta)\rangle$ are continuous.

Besides, we define $u\in C^{1,2,2}_{+,\alpha}(\Lambda)\subset C^1(\overline{\Lambda})$ if $u\in C^{0,2,2}_{+,\alpha}(\Lambda)$ and the time derivative $\partial_t u$ is continuous.

3.2.2. Main Result

Fix the filtrated probability space $(\Omega, \mathcal{F}, \mathbb{P}, \{\mathcal{F}_t\}_{t\in[0,T]})$ where $\{\mathcal{F}_t\}_{t\in[0,T]}$ is the filtration generated by one dimensional Brownian motion B.

In our context, we consider the following rough volatility model which is a extension for the model introduced in the [BJP25, Section 2.4] since it has more general structure for the volatility process that is no more a Volterra Gaussian, the dynamics is described by the stochastic Volterra equation (SVE) below

$$\begin{aligned} X_t &= x_0 + \int_0^t \psi(r, Z_r)\, dB_r + \zeta \int_0^t \psi(r, Z_r)^2\, dr, \\ Z_t &= \int_0^t K(t,r) \big(b(Z_r)\, dr + \sigma(Z_r)\, dB_r\big) \end{aligned} \tag{3.45}$$

with $x_0, \zeta \in \mathbb{R}$. The kernel $K : [0,T]^2 \to \mathbb{R}$ can be singular and Z is a Volterra process. We will focus on the Riemann-Liouville case $K(t,s) = \frac{(t-s)^{H-1}}{\Gamma(H)}$ for $H \in (\frac{1}{2}, 1)$ or at least the kernel K should have the identical property as this fractional kernel.

Also the SVE (3.45) of the two-dimensional $\boldsymbol{X} = (X, Z)^\top$ is a particular case of [BJP25, SVE (2.1)] with $d = m = 2$ and

$$b(t, r, x_1, x_2) = \begin{bmatrix} \zeta\psi(r, x_2)^2 \\ K(t,r)b(x_2) \end{bmatrix}, \qquad \sigma(t, r, x_1, x_2) = \begin{bmatrix} 0 & \psi(r, x_2) \\ 0 & K(t,r)\sigma(x_2) \end{bmatrix}. \tag{3.46}$$

In financial models, X is the log stock price, i.e., the log of an exponential martingale and thus $\zeta = -\frac{1}{2}$. Actually, our method also works for more general setting on dynamics for X with diffusion coefficients $\tilde{b}(r,x), \tilde{\sigma}(r,x)$ and there's no need to satisfy the relation $\tilde{b}(r,x) = \zeta\tilde{\sigma}(r,x)^2$, see [BMM25].

Definition 3.2.48. We say that $f : \mathbb{R} \to \mathbb{R}$ belongs to $\mathcal{C}^m_{\text{poly}}(\mathbb{R})$ with $m \in \mathbb{N}$ if there exists $k_f > 0$ such that $\max_{j\le m} \left|f^{(j)}(x)\right| \lesssim 1 + |x|^{k_f}$ for all $x \in \mathbb{R}$.

For $g : [0,T] \times \mathbb{R} \to \mathbb{R}$ belongs to $\mathcal{C}^{0,m}_{\text{poly}}([0,T] \times \mathbb{R})$ with $m \in \mathbb{N}$ if there exists $k_g > 0$ such that $\max_{j\le m} \left|\partial^j_{x^j} g(t,x)\right| \lesssim 1 + (|x| + t)^{k_g}$ for all $x \in \mathbb{R}$.

Assumption 3.2.49.

(i) $b, \sigma \in C^1$ having linear growth and bounded first order derivatives.

(ii) The payoff function ϕ belongs to $\mathcal{C}^3_{\text{poly}}(\mathbb{R})$.

(iii) The volatility function ψ belongs to $\mathcal{C}^{0,3}_{\text{poly}}([0,T] \times \mathbb{R})$.

(iv) $K(t,s) = K(t-s)$, $K : (0,T] \to \mathbb{R}^{>0}$. The function $K(t)$ is non-increasing, continuous on $(0,T]$ and differentiable a.e. on $(0,T]$ and satisfies: exist $H \in (\frac{1}{2}, 1)$,

$$|K(t,s)| \lesssim (t-s)^{H-1}, \quad |\partial_t K(t,s)| \lesssim (t-s)^{H-2},$$

for all $s < t$. Assume we also have another comparison kernel $\overline{K}(t,s) = \overline{K}(t-s)$, s.t., $\overline{K} \in C([0,T];\mathbb{R}^{>0}) \cap L^2([0,T])$, and there exists $c_{K,\overline{K}} > 0$, for any $t \in (0,T]$, we have

$$\overline{K}(t) \leq c_{K,\overline{K}} K(t).$$

Remark 3.2.50. (i) Assumption 3.2.49 (iv) can imply $K, \overline{K} \in L^{2+\epsilon}[0,T]$ for $\epsilon > 0$ small enough (depending on H).

(ii) In the case of quadratic rough Heston model, the kernel is given by $K(t,s) = \frac{(t-s)^{H-1}}{\Gamma(H)}$ and $\overline{K} = K^N$ denote an approximation obtained via some quadrature method. In this setting, Assumption 3.2.49 (iv) should hold with a universal constant c_K, uniformly in N. This ensures that the constant appearing on the right-hand side of the error estimate in the Theorem 3.2.51 below remains independent of N.

(iii) Assumption 3.2.49 (i) implies b, σ are Lipschitz. So we will have strong wellposedness for the model (3.45) by Theorem 3.0.1.

(iv) The assumption "$K(t,s) = K(t-s)$" is required to apply the Volterra type Grownwall's inequality in the following.

We consider the approximate processes $(\overline{X_t}, \overline{Z_t})$ for (X_t, Z_t), satisfying a SVE associated with an approximate kernel $\overline{K}$:

$$\begin{aligned}
\overline{X_t} &= x_0 + \int_0^t \psi(r, \overline{Z_r})\, dB_r + \int_0^t \zeta\psi(r, \overline{Z_r})^2\, dr, \\
\overline{Z_t} &= \int_0^t \overline{K}(t,r)\left(b(\overline{Z_r})\, dr + \sigma(\overline{Z_r})\, dB_r\right).
\end{aligned} \tag{3.47}$$

Theorem 3.2.51 (Error Bound)**.** *Let Assumption 3.2.49 (i)-(iv) hold. Then, there exists a constant $C \in \mathbb{R}_+$ (depending on T, H, ψ, ϕ, b, σ, $K|_{[0,T]}$, $\overline{K}|_{[0,T]}$), for the stochastic Volterra processes (3.45) and (3.47), (X,Z) and $(\overline{X},\overline{Z})$, we have*

$$|\mathbb{E}[\phi(X_T)] - \mathbb{E}[\phi(\overline{X}_T)]| \leq C \int_0^T \left\{ \|K - \overline{K}\|_{L^1([0,t])} + \|K^2 - \overline{K}^2\|_{L^1([0,t])} \right\} dt.$$

Remark 3.2.52. (i) In the case of Remark 3.2.50 (ii), $c_{\overline{K},K}$ in the Assumption 3.2.49 (iv) only depends on K when N is large enough, then the constant C in Theorem 3.2.51 won't depend on K^N. It means in this case it really gives us an estimation on error bound for some approximations K^N, as [BB23b, Corollary 2.12] did.

(ii) The estimation of error bound consists of two terms. The term $\int_0^T \|K - \overline{K}\|_{L^1([0,t])} dt$ usually behaves better than the latter $\int_0^T \|K^2 - \overline{K}^2\|_{L^1([0,t])} dt$, i.e., yields a sharper bound, especially when H is small, one can see [BMM25, Example (1.4)].

3.2.3. Representation for Conditional Expectation

For $(t,x,\omega) \in [0,T] \times \mathbb{R} \times C^0([0,T],\mathbb{R})$, we define the stochastic flow

$$\begin{aligned} X_T^{t,x,\omega} =& x + \int_t^T \zeta\psi^2(s, Z_s^{t,\omega})ds + \int_t^T \psi(s, Z_s^{t,\omega})dB_s, \\ Z_s^{t,\omega} =& \omega_s + \int_t^s K(s,r)\big(b(Z_r)\,dr + \sigma(Z_r)\,dB_r\big). \end{aligned} \tag{3.48}$$

We define the forward process

$$\Theta_s^t = \int_0^t K(s,r)\,\big(b(Z_r)\,dr + \sigma(Z_r)\,dB_r\big), \quad s \geq t. \tag{3.49}$$

We remark here the process Θ_s^t is a semimartingale for $s > t$. This process arises from the decomposition of Z as

$$Z_s = \Theta_s^t + \int_t^s K(s,r)\,\big(b(Z_r)\,dr + \sigma(Z_r)\,dB_r\big) =: \Theta_s^t + I_s^t, \quad \text{for all } 0 \leq t \leq s \leq T.$$

In particular, $\Theta_t^t = Z_t$. And similarly we also define $\overline{\Theta}_s^t$ and $\overline{I}_s^t$ for the approximate processes $(\overline{X}_t, \overline{Z}_t)$.

We are interested in representations of the value function $U : \overline{\Lambda} \to \mathbb{R}$ such that $U(t,\boldsymbol{\omega}) = \mathbb{E}[\Phi(\boldsymbol{X}^{t,\boldsymbol{\omega}})]$, which fits for our model. By the discussion of the [BJP25, Remark 2.15], the strong well-posedness of (X,Z) (from Theorem 3.0.1) can imply that the conditional expectation $\mathbb{E}[\phi(X_T)|\mathcal{F}_t] = U(t, \boldsymbol{X} \otimes_t \boldsymbol{\Theta}^t)$ where $\boldsymbol{X}$ is the solution for SVE with coefficients (3.46) and $\boldsymbol{\Theta}^t$ is defined in the same way as (3.49) but replace Z by $\boldsymbol{X}$. Also by [BJP25, Remark 2.17 (2),(3)], the semi-martingality of X and the state-dependent payoff function ϕ imply that the conditional expectation $\mathbb{E}[\phi(X_T)|\mathcal{F}_t]$ is only a function of (t, X_t, Θ^t), i.e., there exists

$$\begin{aligned} u : \Big\{(t,x,\omega) \in [0,T] \times \mathbb{R} \times \big(D^0([t,T]) \cap C^0(t,T]\big)\Big\} \to \mathbb{R}, \\ \text{s.t.} \quad u\big(t, \omega_t^{(1)}, \omega^{(2)}|_{[t,T]}\big) = U(t,\boldsymbol{\omega}). \end{aligned} \tag{3.50}$$

Therefore we have the following representation

$$u(t, X_t, \Theta_{[t,T]}^t) = \mathbb{E}[\phi(X_T)|\mathcal{F}_t],$$

which is a function defined on the space $\overline{\Lambda}$.

3.2.4. Regularity for $u(t,x,\omega)$

In this subsection, we will prove $u(t,x,\omega)$ has enough regularity in our context and then calculate its Hadamard derivatives.

First, we want to prove a lemma to obtain the moment control uniformly in time for the stochastic flow (3.48) started at fixed (x,ω) and the stochastic flow started at the approximate processes $(\overline{X}_t, \overline{\Theta}^t)$, which will allow us to prove the regularity of $(x, \omega) \mapsto \mathbb{E}[\phi(X_T^{t,x,\omega})]$ since it's helpful for us to use the Hölder inequality throughout our proofs.

Lemma 3.2.53. *Let Assumptions 3.2.49 (i)-(iv) hold. For any $p \geq 1$ and $(x, \omega) \in \mathbb{R} \times C^0([0,T], \mathbb{R})$. There exists positive constants $c_1, c_2, \bar{c}_1$ and $\bar{c}_2$ depending on $K|_{[0,T]}$, $\overline{K}|_{[0,T]}$, b, σ, ψ and T, s.t.*

(i) We have

$$\sup_{t\in[0,T]} \mathbb{E}[|X_T^{t,x,\omega}|^p] \leq c_1\big(1 + |x|^p + \|\omega\|_{[0,T]}^p\big),$$

$$\sup_{t\in[0,T]} \sup_{s\in[t,T]} \mathbb{E}[|Z_s^{t,\omega}|^p] \leq c_2\big(1 + \|\omega\|_{[0,T]}^p\big),$$

$$\sup_{t\in[0,T]} \mathbb{E}[|X_T^{t,\overline{X}_t,\overline{\Theta}^t}|^p] \leq \bar{c}_1\big(|x|^p + 1\big),$$

$$\sup_{t\in[0,T]} \sup_{s\in[t,T]} \mathbb{E}[|Z_s^{t,\overline{\Theta}^t}|^p] \leq \bar{c}_2.$$

(ii) We have

$$\mathbb{E}[|\sup_{0\leq t\leq T} Z_t|^p] \vee \mathbb{E}[|\sup_{0\leq t\leq T} X_t|^p] < \infty.$$

Proof. (i) First we define the forward process for Z, which is given by

$$g_t^Z(r) = \mathbb{E}_t\left[Z_r - \int_t^r K(r-s)b(Z_s)ds\right], \quad t \leq r, \tag{3.51}$$

where here $\mathbb{E}_t$ means the conditional expectation under the filtration $\mathcal{F}_t$. And we only prove the last two inequalities, i.e., the case that the stochastic flow starts at the approximate processes $(\overline{X}_t, \overline{\Theta}^t)$. The another two inequalities will be obtained in the same way.

First we see

$$\begin{aligned} \mathbb{E}[|Z_s^{t,\overline{\Theta}^t}|^p] &\leq c_p\left\{\mathbb{E}[|\overline{\Theta}_s^t|^p + \mathbb{E}[|Z_s - g_t^Z(s)|^p]\right\} \\ &\leq c_p\left\{\mathbb{E}[|\overline{\Theta}_s^t|^p + \mathbb{E}[|Z_s|^p] + \mathbb{E}[|g_t^Z(s)|^p]\right\} \end{aligned} \tag{3.52}$$

follows from c_p-inequality. Then we can apply Proposition 3.1.5 and Proposition 3.1.12 to obtain

$$\sup_{t\in[0,T]} \sup_{s\in[t,T]} \mathbb{E}[|Z_s^{t,\overline{\Theta}^t}|^p] \leq \bar{c}_2.$$

And similarly, using c_p-inequality, BDG inequality and Hölder inequality, we have

$$\begin{aligned}
\mathbb{E}[|X_T^{t,\overline{X}_t,\overline{\Theta}^t}|^p] &\le c_p\left\{\mathbb{E}[|\overline{X}_t|^p] + \int_t^T \mathbb{E}[|\psi(s, Z_s^{t,\overline{\Theta}^t})|^{2p}]ds + \int_t^T \mathbb{E}[|\psi(s, Z_s^{t,\overline{\Theta}^t})|^p]ds\right\} \\
&\le c_p\left\{1 + |x|^p + c_{k_\psi}(T^{k_\psi} + \sup_{t\in[0,T]}\sup_{s\in[t,T]}\mathbb{E}[|Z_s^{t,\overline{\Theta}^t}|^{2k_\psi p}])\right\} \\
&\le c_{p,T}(1 + |x|^p + \overline{c}_2),
\end{aligned} \tag{3.53}$$

where we also use the following estimate for $\mathbb{E}[|\overline{X}_t|^p]$:

$$\begin{aligned}
\mathbb{E}[|\overline{X}_t|^p] &\le c_p\mathbb{E}\left\{|X_0|^p + \left|\int_0^t \psi(s, \overline{Z}_s)\,ds\right|^p + \left|\int_0^t \psi^2(s, \overline{Z}_s)\,dB_s\right|^p\right\} \\
&\le c_p\left\{\mathbb{E}[|X_0|^p] + T^{p-1}\int_0^T \mathbb{E}\left[|\psi(s, \overline{Z}_s)|^p\right]ds + c_pT^{p/2-1}\int_0^T \mathbb{E}\left[|\psi^2(s, \overline{Z}_s)|^p\right]ds\right\} \\
&\le c_{p,T}(|x|^p + 1),
\end{aligned}$$

due to the identity $\overline{Z}_s = Z_s^{s,\overline{\Theta}^t}$, we can use the same argument as above.

Take $\sup_{t\le T}$ on the both hand sides to obtain

$$\sup_{t\in[0,T]}\mathbb{E}[|X_T^{t,\overline{X}_t,\overline{\Theta}^t}|^p] \le \overline{c}_1(|x|^p + 1).$$

(ii) It's trivial from Proposition 3.1.5 and similar estimations from above. $\square$

The following lemma gives the first and second Hadamard derivatives in the space $L^p(\Omega)$ for the flow process $X_T^{t,x,\omega}$.

Lemma 3.2.54. *Let Assumptions 3.2.49 (i)-(iv) hold. Fix $(t,x) \in [0,T] \times \mathcal{X}_t$ and $p \ge 1$.*

(i) The function $\omega \mapsto X_T^{t,x,\omega}$ is Hadamard-differentiable in $L^p(\Omega)$, and the Hadamard derivative at $\omega \in \mathcal{X}_t$ in the direction $\eta \in \mathcal{X}_t$ is given by

$$\langle\partial_\omega X_T^{t,x,\omega}, \eta\rangle = \zeta\int_t^T \partial_x(\psi^2)(s, Z_s^{t,\omega})\eta_s ds + \int_t^T \partial_x\psi(s, Z_s^{t,\omega})\eta_s dB_s.$$

(ii) The function $\omega \mapsto \langle\partial_\omega X_T^{t,x,\omega}, \eta\rangle$ is Hadamard-differentiable in $L^p(\Omega)$, and the Hadamard derivative at $\omega \in \mathcal{X}_t$ in the direction $\xi \in \mathcal{X}_t$ is given by

$$\langle\partial^2_{\omega\omega} X_T^{t,x,\omega}, (\eta,\xi)\rangle = \zeta\int_t^T \partial^2_{xx}(\psi^2)(s, Z_s^{t,\omega})\eta_s\xi_s ds + \int_t^T \partial^2_{xx}\psi(s, Z_s^{t,\omega})\eta_s\xi_s dB_s.$$

Proof. The proof is same as the proof for [BMM25, Lemma 4.4] since we assume the same growth conditions on ψ and ϕ and we also have Lemma 3.2.53 to apply. $\square$

Next we can check the regularity of function $u(t,x,\omega)=\mathbb{E}[\phi(X_T^{t,x,\omega})]$ and obtain the representation for the derivatives of u in the following proposition.

Proposition 3.2.55. *Let $t\in[0,T]$ and $\eta,\xi\in\mathcal{X}_t$, and Assumptions 3.2.49 (i)-(iii) hold. The function $u:\bar{\Lambda}\to\mathbb{R}$ defined by $u(t,x,\omega)=\mathbb{E}[\phi(X_T^{t,x,\omega})]$ belongs to the space $C_+^{0,2,2}(\bar{\Lambda})$. Moreover, the following representations hold*

$$\begin{aligned}
\partial_x u(t,x,\omega) =&\mathbb{E}[\phi'(X_T^{t,x,\omega})],\\
\partial^2_{xx} u(t,x,\omega) =&\mathbb{E}[\phi''(X_T^{t,x,\omega})],\\
\langle\partial_\omega u(t,x,\omega),\eta\rangle =&\mathbb{E}[\phi'(X_T^{t,x,\omega})\langle\partial_\omega X_T^{t,x,\omega},\eta\rangle],\\
\langle\partial_\omega\partial_x u(t,x,\omega),\eta\rangle =&\mathbb{E}[\phi''(X_T^{t,x,\omega})\langle\partial_\omega X_T^{t,x,\omega},\eta\rangle],\\
\langle\partial^2_{\omega\omega} u(t,x,\omega),(\eta,\xi)\rangle =&\mathbb{E}[\phi''(X_T^{t,x,\omega})\langle\partial_\omega X_T^{t,x,\omega},\eta\rangle\langle\partial_\omega X_T^{t,x,\omega},\xi\rangle]\\
&+\mathbb{E}[\phi'(X_T^{t,x,\omega})\langle\partial^2_{\omega\omega} X_T^{t,x,\omega},(\eta,\xi)\rangle].
\end{aligned}\tag{3.54}$$

Remark 3.2.56. In particular, the map $(\eta,\zeta)\mapsto\langle\partial^2_{\omega\omega}u(t,x,\omega),(\eta,\zeta)\rangle$ is a bilinear symmetric form on $\mathcal{X}_t\times\mathcal{X}_t$.

Proof. The proof is similar as the proof for [BMM25, Proposition 4.5] for the same reasons. The only part needed to be additionally checked falls on proving the time continuity of derivatives of u. It can be reduced to prove: for any $q\geq 2$,

$$\|Z_r^{u,\omega}-Z_r^{t,\omega}\|_{L^q(\Omega)}\to 0\quad(u\to t).$$

And it can be implied by the following inequalities

$$\begin{aligned}
\|Z_r^{u,\omega}-Z_r^{t,\omega}\|_{L^q(\Omega)}&=\|\int_u^t K(r-l)[b(Z_l)dl+\sigma(Z_l)dB_l]\|_{L^q(\Omega)}\\
&\leq\int_u^t K(r-l)\|b(Z_l)\|_{L^q(\Omega)}dl+(\int_u^t K^2(r-l)\|\sigma^2(Z_l)\|_{L^{q/2}(\Omega)}dl)^{1/2}\\
&\leq\sup_l\|b(Z_l)\|_{L^q(\Omega)}\int_u^t K(r-l)dl\\
&+\sup_l\|\sigma^2(Z_l)\|^{1/2}_{L^{q/2}(\Omega)}(\int_u^t K^2(r-l)dl)^{1/2}\\
&\to 0,
\end{aligned}\tag{3.55}$$

follows from BDG inequality, Minkovsky's inequality and Lemma 3.1.5, along with the fact that b and σ have only linear growth. ☐

Since now we're in a model where the volatility process Z satisfying a more complicated stochastic Volterra equation. Its Malliavin derivative $\mathbf{D}_sZ_t$ can't be calculated in an explicit form any more as in [BJP25] where $\mathbf{D}_sZ_t=K(t,s)\mathbb{1}_{\{s<t\}}$. So in the following lemma, we require to develop the Malliavin calculus theory for stochastic Volterra equation and get a sharp estimation for the Malliavin derivative $\mathbf{D}_sZ_t$, which is vital to obtain the second order path derivative at the singular direction K.

Lemma 3.2.57. *Consider the stochastic volterra equation*

$$Z_t = \int_0^t K(t,r)\left(b(Z_r)\,dr + \sigma(Z_r)\,dB_r\right). \tag{3.56}$$

Let the Assumptions 3.2.49 (i) and (iv) hold. Then the strong solution $Z_t \in \mathbb{D}^{1,2}$ for $0 \le t \le T$ and the Malliavin derivative $\mathbf{D}_r Z_t$ satisfies

$$\mathbf{D}_r Z_t = K(t,r)\sigma(Z_r) + \int_r^t K(t,s)\big(\partial_x b(Z_s)\mathbf{D}_r Z_s ds + \partial_x \sigma(Z_s)\mathbf{D}_r Z_s dB_s\big), \tag{3.57}$$

for $r < t$, while $\mathbf{D}_r Z_t = 0$ for $T \ge r > t$. Besides we also have the estimate

$$\sup_{0\le r\le T} \int_r^T \mathbb{E}(|\mathbf{D}_r Z_t|^2) dt < \infty. \tag{3.58}$$

Remark 3.2.58. It is worth noting that the following observation will be repeatedly used in what follows: if we consider the stochastic flow

$$Z_t^{t_0,w} = w + \int_{t_0}^t K(t,r)\left(b(Z_r)\,dr + \sigma(Z_r)\,dB_r\right), \tag{3.59}$$

for some path w on $[0,t_0]$. Then $Z_t^{t_0,w} \in \mathbb{D}^{1,2}$ and

$$\mathbf{D}_r Z_t^{t_0,w} = K(t,r)\sigma(Z_r) + \int_r^t K(t,s)\big(\partial_x b(Z_s)\mathbf{D}_r Z_s ds + \partial_x \sigma(Z_s)\mathbf{D}_r Z_s dB_s\big) \tag{3.60}$$

for any $t_0 < r \le t$ by chain rule. Besides, we still have the same estimate for $\mathbf{D}_r Z_t^{t_0,w}$ follows from

$$\begin{aligned}
\int_r^T \mathbb{E}(|\mathbf{D}_r Z_t^{t_0,w}|^2) dt &\lesssim \sup_r \int_r^T K^2(t,r)dt \cdot \sup_r \mathbb{E}[\sigma^2(Z_r)] \\
&\quad + C_T \int_r^T \int_r^t K^2(t,s)\mathbb{E}[|\mathbf{D}_r Z_s|^2] ds dt \\
&= C_K + C_T \int_r^T \int_s^T K^2(t,s)\mathbb{E}[|\mathbf{D}_r Z_s|^2] dt ds \\
&\lesssim C_K + C_T[\sup_{r\le s\le T} \int_s^T K^2(t,s)dt] \int_r^T \mathbb{E}[|\mathbf{D}_r Z_s|^2] ds < \infty,
\end{aligned} \tag{3.61}$$

where we apply Fubini-Tonelli Theorem to get rid of integral of K^2. In particular, it works for $w = \overline{\Theta}^{t_0}$.

Proof. We first consider the Picard iterates Z_n, defined inductively by $Z_0 = 0$ and

$$Z_{n+1}(t) = \mathcal{T}(Z_n)(t) := \int_0^t K(t,r)\left(b(Z_n(r))\,dr + \sigma(Z_n(r))\,dB_r\right).$$

Set

$$\|X\|_{p,T} := \sup_{t\le T} \mathbb{E}[|X_t|^p]^{1/p}$$

and

$$\|X\|_{p,T,\lambda} := \sup_{t\le T} \mathbb{E}[|e^{-\lambda t}X_t|^p]^{1/p}, \quad \lambda > 0.$$

We let $\mathcal{H}_{p,T}$ denote the space of all X, s.t., $\|X\|_{p,T} < \infty$, modulo the equivalence relation obtained by identifying processes that are versions of each other. Then $(\mathcal{H}_{p,T}, \|\cdot\|_{p,T})$ is obviously a Banach space.

It's easy to see $\|\cdot\|_{p,T,\lambda} \sim \|\cdot\|_{p,T}$. And by induction on n, $Z_0 \in (\mathcal{H}_{p,T}, \|\cdot\|_{p,T})$ is trivial, along with the estimate

$$\begin{aligned}\|Z_{n+1}\|_{p,T} &\le \sup_{t\le T}\{\int_0^t K(t,r)\mathbb{E}[|b(Z_n(r))|^p]^{\frac{1}{p}}dr + (\mathbb{E}[(\int_0^t K(t,r)^2|\sigma(Z_n(r)|^2 dr)^{\frac{p}{2}}]^{\frac{1}{p}}\}\\ &\le \sup_{t\le T}\{\int_0^t K(t,r)\mathbb{E}[|b(Z_n(r))|^p]^{\frac{1}{p}}dr + (\int_0^t K(t,r)^2\mathbb{E}[|\sigma(Z_n(r)|^p]^{\frac{2}{p}} dr)^{\frac{1}{2}}\}\\ &\le C(1+\|Z_n\|_{p,T})\sup_{t\le T}\{\int_0^t K(t,r)dr + (\int_0^t K(t,r)^2 dr)^{\frac{1}{2}}\} < \infty\end{aligned} \tag{3.62}$$

follows from Minkovsky's inequality, BDG inequality and Jensen's inequality, along with the fact that $K \in L^2$. So $(Z_n) \subset (\mathcal{H}_{p,T}, \|\cdot\|_{p,T})$.

Therefore, from [AJLP19, Appendix A], we see we can choose λ large enough to let $\mathcal{T}$ be a contraction. Then we can prove that $Z_n \to Z$ in $\|\cdot\|_{p,T}$, So it's also L^p-convergence and (Z_n) is L^p bounded, where Z is the unique strong solution of SVE (3.56).

Next we will prove $(Z(t))_{[0,T]} \subset \mathbb{D}^{1,2}$ and the Malliavin derivatives satisfy the equation (3.60). We first prove these results will hold for $(Z_n(t))_{[0,T]}$ by induction on n. For $n = 0$, it's trivial. We assume the case $k = n$ holds. For the case $n+1$: since $b, \sigma \in C^1$, apply Proposition 3.2.16, we have

$$\mathbf{D}_r[b(s, Z_n(s))] = \partial_x b(s, Z_n(s))\mathbf{D}_r Z_n(s), \qquad \mathbf{D}_r(\sigma(s, Z_n(s))) = \partial_x\sigma(s, Z_n(s))\mathbf{D}_r Z_n(s)$$

For a Itô integral,

$$\mathbf{D}_r\Big(\int_0^t u_s\, dB_s\Big) = u_r\mathbf{1}_{\{r<t\}} + \int_r^t \mathbf{D}_r u_s\, dB_s,$$

for $r < t$ and $\mathbf{D}_r\Big(\int_0^t u_s\, dW_s\Big) = 0$, for $r > t$. Applying $\mathbf{D}_r$ on $Z_{n+1}(t)$, we obtain

$$\begin{aligned}\mathbf{D}_r Z_{n+1}(t) =& K(t,r)\sigma(Z_n(r)) + \int_r^t K(t,s)\big(\partial_x b(Z_n(r))\mathbf{D}_r Z_n(s)ds\\ &+ \partial_x\sigma(Z_n(s))\mathbf{D}_r Z_n(s)dB_s\big),\end{aligned} \tag{3.63}$$

for $r < t$ and $\mathbf{D}_r Z_{n+1}(t) = 0$, for $r > t$. Therefore, $(Z_n(t))_{[0,T]} \subset \mathbb{D}^{1,2}$ and the their Malliavin derivatives satisfy the equation (3.60).

Denote $\phi_k(t) := \int_0^t \|\mathbf{D}_r Z_k(t)\|^2_{L^2(\Omega)}dr$. By induction again on k, we can prove that

$\phi_k(t) \in L^1$ and $\phi_k(t) < \infty$ for any $t \in [0,T]$, follow from the following estimates

$$\begin{aligned}\phi_{n+1}(t) \leq & (\sup_{0\leq t\leq T}\int_0^t |K(t-r)|^2 dr) \sup_n \sup_{t\leq T} \|\sigma(Z_n(t))\|^2_{L^2(\Omega)} \\ & + \int_0^t \mathbb{E}[(\int_r^t |K(t,s)\mathbf{D}_r Z_n(s)|ds)^2]dr + C\int_0^t\int_r^t K(t,s)^2\|\mathbf{D}_r Z_n(s)\|^2_{L^2(\Omega)} ds dr \\ & \leq C_1 + C_T \int_0^t\int_r^t K(t,s)^2 \|\mathbf{D}_r Z_n(s)\|^2_{L^2(\Omega)} ds dr \\ & = C_1 + C_T\int_0^t K(t-s)^2 \int_0^s \|\mathbf{D}_r Z_n(s)\|^2_{L^2(\Omega)} dr ds \\ & = C_1 + C_2 K^2 * \phi_n(t),\end{aligned} \tag{3.64}$$

where the first term of the first inequality is bounded from the fact that (Z_n) is L^p bounded and σ has only linear growth. We also use Jensen inequality, BDG inequality and Fubini Theorem here, along with the fact that $\partial_x b, \partial_x \sigma$ are bounded to get these estimations. Besides, C_1, C_2 only depend on $T, K|_{[0,T]}, b, \sigma$.

So $\phi_k(t)$'s are well-defined. Denote $\tilde{K} := K^2 \in L^1[0,T]$. Iteratively use the inequality (3.64), we have

$$\phi_{n+1}(t) \leq C_2 \sum_{i\geq 0}^{n} C_1 * \tilde{K}^{(*i)} \leq C_2 \sum_{i\geq 0}^{\infty} C_1 * \tilde{K}^{(*i)}, \tag{3.65}$$

since $\tilde{K}$ is nonnegative. And we also have $C_1 * \tilde{K}^{(*n)} = C_1 \int_0^t \tilde{K}^{(*n)}(s)ds$. Combined with the equation

$$\int_0^t \tilde{K}^{(*n)}(s)ds \leq \int_{0\leq \sum_1^n s_i \leq t} \prod_{i=1}^n \tilde{K}(s_i) ds_1 ds_2 \cdots ds_n \leq \frac{\|\tilde{K}\|^n_{L^1[0,T]}}{n!},$$

which follows from the symmetry of integration domain. Therefore,

$$\phi_n(t) \leq C_1 C_2 e^{\|\tilde{K}\|_{L^1[0,T]}} < \infty,$$

for any n. It implies

$$\sup_n \mathbb{E}[\|\mathbf{D}.Z_n(t)\|^2_{L^2([0,T])}] = \sup_n \int_0^T \|\mathbf{D}_r Z_n(t)\|^2_{L^2(\Omega)} dr = \sup_n \int_0^t \|\mathbf{D}_r Z_n(t)\|^2_{L^2(\Omega)} dr < \infty.$$

Combined with $Z_n(t) \to Z(t)$ in $L^2(\Omega)$. We have $(Z(t))_{[0,T]} \subset \mathbb{D}^{1,2}$ from Lemma 3.2.17 with $H = L^2([0,T])$ and get the equation (3.60) from Proposition 3.2.24.

As for the estimate, similarly, we can obtain

$$\psi_r(t) := \mathbb{E}(|\mathbf{D}_r Z(t)|^2) \leq CK(t,r)^2 + CK^2 * \psi_r(t).$$

where C only depends on K, T, b, σ. Apply the Lemma 3.2.38, we have

$$\psi_r(t) \leq y_r(t),$$

where $y_r(t)$ satisfies equation $y_r(t) = CK^2 * y_r(t) + CK(t,r)^2$. We can solve $y_r(t)$ by Lemma 3.2.37, which is given by

$$y_\cdot(t) = CK(t,\cdot)^2 - CR * K^2(t,\cdot),$$

where R is resolvent of the second kind of $-K^2$. So R is negative and belongs to L^1. Eventually we get

$$\sup_{0\leq r\leq T}\int_r^T \mathbb{E}(|\mathbf{D}_r Z_t|^2)dt \leq \sup_{0\leq r\leq T}\int_r^T y_r(t)dt \tag{3.66}$$

$$\leq C\sup_{0\leq r\leq T}\int_r^T K(t,r)^2 dt + C\sup_{0\leq r\leq T}\int_r^T |R * K^2(t,r)|dt \tag{3.67}$$

$$\leq C_{K,T} + C\|R\|_{L^1[0,T]}\sup_{0\leq r\leq T}\int_r^T K(t,r)^2 dt < \infty, \tag{3.68}$$

follows from Young's inequality. □

Remark 3.2.59. The Lemma 3.2.57 is crucial, as it allows us to apply Hölder's inequality to separate the term $|\mathbf{D}_r Z_t|^2$ and another L^p random variable inside the same expectation.

Corollary 3.2.60. *Under the conditions of Lemma 3.2.57. There exists a $\epsilon_H > 0$, st.*

$$\sup_{0\leq r\leq T}\int_r^T \mathbb{E}[|\mathbf{D}_r Z_t|^{(2+\epsilon_H)}]dt \leq C_{K,T,H,b,\sigma} < \infty.$$

Proof. Similarly, we can derive the inequality below

$$\psi_r(t) := \mathbb{E}[|\mathbf{D}_r Z(t)|^{(2+\epsilon_H)}] \leq CK(t,r)^{(2+\epsilon_H)} + CK^{(2+\epsilon_H)} * \psi_r(t),$$

where C only depends on $K|_{[0,T]}, H, T, b, \sigma$. Since K satisfies the Assumption 3.2.49 (iv), so there exists a $\epsilon_H > 0$, s.t., $\exists H' \in (1/2, 1)$,

$$|K^{1+\epsilon_H/2}(t,s)| \lesssim |t-s|^{H'-1}, \quad |\partial_t(K^{1+\epsilon_H/2})(t,s)| \lesssim |t-s|^{H'-2},$$

and $K^{1+\epsilon_H/2} \in L^2$. Therefore we can repeat the argument above with $\tilde{K} := K^{1+\epsilon_H/2}$. □

Remark 3.2.61. For $H > \frac{3}{4}$, we can obtain a stronger estimate by applying Hölder's inequality to bound the kernel K in the integral of the SVE (3.56). In this case, $K \in L^4$, which allows us to handle the term K^2 arising from Itô's isometry. Moreover, the assumption $K(t,s) = K(t-s)$ in Assumption 3.2.49 (iv) is no longer required.

Denote $\phi_n(t) := \mathbb{E}[\int_0^t |\mathbf{D}_r Z_n(t)|^4 dr]$, then apply the BDG inequality, Hölder inequality, c_p-inequality and Fubini Theorem, and use the same argument above, we can obtain

$$\phi_{n+1}(t) \le C_1 + C_2 \int_0^t \phi_n(s) ds, \tag{3.69}$$

where C_1, C_2 only depend on b, σ, K, T. Iteratively use the inequality (3.69), then we have

$$\phi_n(t) \le C_1 \sum_0^n C_2^j t^j / j! \le C_1 e^{C_2 T},$$

follows from the same argument in the proof above. Therefore, we can conclude $Z_t \in \mathbb{D}^{1,4}$.

Besides, note that we have estimate

$$\phi(t) := \mathbb{E}[\int_0^t |\mathbf{D}_r Z(t)|^4 dr] \le C_1 + C_2 \int_0^t \phi(s) ds$$

by using Grownwall's inequality. It can provide us a nice bound by the identical argument as above

$$\mathbb{E}[\int_0^t |\mathbf{D}_r Z(t)|^4 dr] \le C_{b,\sigma,K,T}.$$

Next, to derive formulas for the path derivative of u in singular directions, we approximate them using truncation functions. Let $K_{(1)}, K_{(2)} \in L^2([0,T], \mathbb{R}_+)$ such that $K_{(i)}$ is continuous on $(0,T]$ for $i = 1,2$. For $t \in [0,T]$ and $K \in \{K_{(1)}, K_{(2)}\}$, we set for all $u \in [t,T]$, $K^t(u) := K(u-t)$ so that $K^t \in L^2([t,T], \mathbb{R}_+)$ and we define the truncated directions following [BMM25].

$$K^{t,\delta}(s) = K^t(s \vee (t+\delta)) = \begin{cases} K(s-t) & \text{if } s \ge t+\delta, \\ K(\delta) & \text{if } s \in [0, t+\delta]. \end{cases}$$

We can now define the path derivatives of a function $u : \bar{\Lambda} \to \mathbb{R}$ by the following limits

$$\begin{aligned} \langle \partial_\omega u(t,x,\omega), K^t_{(i)} \rangle &= \lim_{\delta \to 0} \langle \partial_\omega u(t,x,\omega), K^{t,\delta}_{(i)} \rangle, \quad \text{for} \quad i = 1,2, \\ \langle \partial^2_{\omega\omega} u(t,\omega), (K^t_{(1)}, K^t_{(2)}) \rangle &= \lim_{\delta \to 0} \langle \partial^2_{\omega\omega} u(t,\omega), (K^{t,\delta}_{(1)}, K^{t,\delta}_{(2)}) \rangle, \end{aligned} \tag{3.70}$$

if they exist.

Lemma 3.2.62. *For $(K_{(1)}, K_{(2)})$ defined as above. Let Assumptions 3.2.49 (i)-(iv) hold and $K_{(i)}$ satisfies $\lim_{\delta\to 0} \delta K^2_{(i)}(\delta) = \lim_{\delta \to 0} \int_0^\delta K^2_{(i)}(r) dr = 0$ for $i = 1,2$. Then for any $t \in [0,T]$, we have the following representations for the path derivatives of $u(t,x,\omega) =$*

$\mathbb{E}[\phi(X_T^{t,x,\omega})] \in C_+^{0,2,2}(\bar{\Lambda})$ *in the singular direction* $(K_{(1)}^t, K_{(2)}^t) \in L^2([t,T], \mathbb{R}_+)^{\otimes 2}$*:*

$$\langle \partial_\omega u(t,x,\omega), K_{(i)}^t \rangle = \mathbb{E}[\phi'(X_T^{t,x,\omega}) \langle \partial_\omega X_T^{t,x,\omega}, K_{(i)}^t \rangle], \tag{3.71}$$

$$\langle \partial_\omega \partial_x u(t,x,\omega), K_{(i)}^t \rangle = \mathbb{E}[\phi''(X_T^{t,x,\omega}) \langle \partial_\omega X_T^{t,x,\omega}, K_{(i)}^t \rangle], \tag{3.72}$$

$$\begin{aligned}\langle \partial^2_{\omega\omega} u(t,x,\omega), (K_{(1)}^t, K_{(2)}^t) \rangle =& \mathbb{E}[\phi''(X_T^{t,x,\omega}) \langle \partial_\omega X_T^{t,x,\omega}, K_{(1)}^t \rangle \langle \partial_\omega X_T^{t,x,\omega}, K_{(2)}^t \rangle] \\ &+ \mathbb{E}\Big[\zeta \phi'(X_T^{t,x,\omega}) \int_t^T \partial^2_{xx}(\psi^2)(s, Z_s^{t,\omega})(K_{(1)}K_{(2)})(t-s)ds \\ &+ \phi''(X_T^{t,x,\omega}) \int_t^T \mathbf{D}_s X_T^{t,x,\omega} \, \partial^2_{xx}\psi(s, Z_s^{t,\omega})(K_{(1)}K_{(2)})(t-s)ds\Big],\end{aligned} \tag{3.73}$$

where

$$\langle \partial_\omega X_T^{t,x,\omega}, K_{(i)}^t \rangle = \zeta \int_t^T \partial_x(\psi^2)(s, Z_s^{t,\omega}) K_{(i)}^t(s) ds + \int_t^T \partial_x \psi(s, Z_s^{t,\omega}) K_{(i)}^t(s) dB_s$$

holds in the sense of $L^2(\Omega)$ *for* $i = 1, 2$.

Proof. To prove the equations (3.71) and (3.72), we prove that

$$\langle \partial_\omega X_T^{t,x,\omega}, K_{(i)}^{t,\delta} \rangle \to \langle \partial_\omega X_T^{t,x,\omega}, K_{(i)}^t \rangle \quad \text{in } L^2(\Omega) \quad \text{for} \quad \delta \to 0,$$

i.e.,

$$\begin{aligned}&\mathbb{E}[(\langle \partial_\omega X_T^{t,x,\omega}, K_{(i)}^{t,\delta} \rangle - \langle \partial_\omega X_T^{t,x,\omega}, K_{(i)}^t \rangle)^2] \\ &\lesssim \mathbb{E}[(\int_t^T \partial_x(\psi^2)(s, Z_s^{t,\omega})(K_{(i)}^{t,\delta} - K_{(i)}^t(s))ds)^2)] \\ &\quad + \mathbb{E}[(\int_t^T \partial_x \psi(s, Z_s^{t,\omega})(K_{(i)}^{t,\delta} - K_{(i)}^t(s))dB_s)^2] \\ &\le \int_t^T (K_{(i)}^{t,\delta} - K_{(i)}^t(s))^2 ds \int_t^T \mathbb{E}[(\partial_x(\psi^2)(s, Z_s^{t,\omega}))^2] ds \\ &\quad + \sup_{t \le s \le T} \mathbb{E}[(\partial_x \psi(s, Z_s^{t,\omega}))^2] \int_t^T (K_{(i)}^{t,\delta} - K_{(i)}^t(s))^2 ds \\ &\le C_T \int_t^T (K_{(i)}^{t,\delta} - K_{(i)}^t(s))^2 ds \to 0,\end{aligned} \tag{3.74}$$

follows from BDG inequality, Minkovsky inequality and Hölder inequality and Lemma 3.2.53.

In the following we denote $\mathbb{E}[F(X_s^{t,x,\omega}, Z_{s'}^{t,\omega})] := \mathbb{E}_{t,x,\omega}[F(X_s, Z_{s'})]$ for $s \ge t$ for convenience. To prove (3.73), most of the argument follows the proof of Proposition 2.23 in [BJP25, Section 7.3], since we assume similar growth conditions on ψ and ϕ, and we also have Lemma 3.2.53. The only part that needs to be handled differently in our proof is

$$\begin{aligned}
&\mathbb{E}_{t,x,\omega}\left[\phi'(X_T)\int_t^T \partial^2_{xx}\psi(s,Z_s)(K^\delta_{(1)}K^\delta_{(2)})(t-s)\,dB_s\right]\\
&=\mathbb{E}_{t,x,\omega}\left[\int_t^T \mathbf{D}_s\phi'(X_T)\partial^2_{xx}\psi(s,Z_s)(K^\delta_{(1)}K^\delta_{(2)})(t-s)ds\right]\\
&\to \mathbb{E}_{t,x,\omega}\left[\int_t^T \mathbf{D}_s\phi'(X_T)\partial^2_{xx}\psi(s,Z_s)(K_{(1)}K_{(2)})(t-s)ds\right](\delta\to 0),
\end{aligned}$$

here we use Malliavin integration by parts in Lemma 3.2.18. Because when taking $\delta\downarrow 0$, we would obtain the ill-defined integral $\int_t^T(K_{(1)}K_{(2)})(t-s)\,dB_s$ since $K_{(i)}$may $\notin L^4$ for $i=1,2$. Note that

$$\begin{aligned}
&|\mathbb{E}_{t,x,\omega}\left[\int_t^T \mathbf{D}_s\phi'(X_T)\partial^2_{xx}\psi(s,Z_s)[(K^\delta_{(1)}K^\delta_{(2)})(t-s)-(K_{(1)}K_{(2)})(t-s)]ds\right]|\\
&\le (\sup_{t\le s\le T}\mathbb{E}_{t,x,\omega}[|\mathbf{D}_s\phi'(X_T)\partial^2_{xx}\psi(s,Z_s)|])\int_t^T[(K^\delta_{(1)}K^\delta_{(2)})(t-s)-(K_{(1)}K_{(2)})(t-s)]ds
\end{aligned}\tag{3.75}$$

and

$$\begin{aligned}
&\int_t^T[(K^\delta_{(1)}K^\delta_{(2)})(t-s)-(K_{(1)}K_{(2)})(t-s)]ds\\
&\le \|K_{(1)}\|_{L^2[0,T]}\|K^\delta_{(2)}-K_{(2)}\|_{L^2[0,T]}+\|K_{(2)}\|_{L^2[0,T]}\|K^\delta_{(1)}-K_{(1)}\|_{L^2[0,T]}\to 0\quad(\delta\to 0).
\end{aligned}\tag{3.76}$$

So we only need to prove

$$\sup_{t\le s\le T}\mathbb{E}_{t,x,\omega}[|\mathbf{D}_s\phi'(X_T)\partial^2_{xx}\psi(s,Z_s)|]<\infty.$$

Next we simply denote the norm $\|\cdot\|_{L^p_{t,x,\omega}(\Omega)}$ of conditional expectation starting at (t,x,ω) by $\|\cdot\|_{L^p(\Omega)}$. Fix $p_1\in(1,2)$, first we see

$$\begin{aligned}
&\sup_{t\le s\le T}\mathbb{E}_{t,x,\omega}[|\mathbf{D}_s\phi'(X_T)\partial^2_{xx}\psi(s,Z_s)|]\\
&=\sup_{t\le s\le T}\mathbb{E}_{t,x,\omega}[|\mathbf{D}_sX_T||\partial^2_{xx}\phi(X_T)\partial^2_{xx}\psi(s,Z_s)|]\\
&\le(\sup_{t\le s\le T}\|\mathbf{D}_sX_T\|_{L^{p_1}(\Omega)}\|)(\sup_{t\le s\le T}\|\partial^2_{xx}\phi(X_s)\partial^2_{xx}\psi(s,Z_s)\|_{L^{q_1}(\Omega)}]\\
&\le C_{T,\psi,\phi}\sup_{t\le s\le T}\|\mathbf{D}_sX_T\|_{L^{p_1}(\Omega)},
\end{aligned}\tag{3.77}$$

by Proposition 3.2.16 and Hölder's inequality. Combined with

$$\begin{aligned}
\mathbf{D}_sX^{t,x,\omega}_T&=\psi(s,Z^{t,\omega}_s)+\int_s^T\mathbf{D}_s\psi(r,Z^{t,\omega}_r)dB_r+\zeta\int_t^T\mathbf{D}_s\psi^2(r,Z^{t,\omega}_r)dr\\
&=\psi(s,Z^{t,\omega}_s)+\int_s^T\partial_x\psi(r,Z^{t,\omega}_r)\mathbf{D}_sZ_rdB_r+\zeta\int_s^T\partial_x(\psi^2)(r,Z^{t,\omega}_r)\mathbf{D}_sZ^{t,\omega}_rdr,
\end{aligned}\tag{3.78}$$

for $s > t$, follows from Proposition 3.2.16 and 3.2.24. Then using BDG inequality, Minkovsky's inequality and Hölder's inequality and Lemma 3.2.53, we have

$$\begin{aligned}
&\sup_{t\le s\le T}\|\mathbf{D}_s X_T\|_{L^{p_1}(\Omega)}\\
&\lesssim \sup_{s\le T}\|\psi(s,Z_s)\|_{L^2(\Omega)}+\sup_{s\le T}\|\int_s^T\partial_x\psi(r,Z_r)\mathbf{D}_sZ_rdB_r\|_{L^2(\Omega)}\\
&\quad+\sup_{t\le s\le T}\|\int_s^T\partial_x(\psi^2)(r,Z_r)\mathbf{D}_sZ_rdr\|_{L^{p_1}(\Omega)}\\
&\le C_{T,\psi}(1+|x|+\|\omega\|_{[0,T]})^{k_\psi}+C_\psi\sup_{t\le s\le T}(\int_s^T\|\partial_x\psi(r,Z_r)\mathbf{D}_sZ_r\|^2_{L^2(\Omega)}dr)^{\frac12}\\
&\quad+\sup_{t\le s\le T}\int_s^T\|\partial_x(\psi^2)(r,Z_r)\mathbf{D}_sZ_r\|_{L^{p_1}(\Omega)}dr\\
&\le C_{T,\psi}(1+|x|+\|\omega\|_{[0,T]})^{k_\psi}\\
&\quad+C_\psi\sup_{t\le s\le r\le T}(\|\partial_x\psi(r,Z_r)\|_{L^{2(2+\epsilon_H)/\epsilon_H}}\sup_{t\le s\le T}(\int_s^T\|\mathbf{D}_sZ_r\|^{1/2}_{L^{2+\epsilon_H}(\Omega)}dr)^{\frac12}\\
&\quad+(\sup_{t\le s\le T}\int_s^T\|\mathbf{D}_sZ_r\|_{L^2(\Omega)}dr)(\sup_{t\le s\le r\le T}\|\partial_x(\psi^2)(r,Z_r)\|_{L^q(\Omega)})\\
&\le C_{H,T,x,\|\omega\|_{[0,T]},\psi}\{1+[\sup_{t\le s\le T}(\int_s^T\|\mathbf{D}_sZ_r\|^2_{L^2(\Omega)}dr)^{\frac12}+\sup_{t\le s\le T}(\int_s^T\|\mathbf{D}_sZ_r\|^{1/2}_{L^{2+\epsilon_H}(\Omega)}dr)^{\frac12}]\}\\
&<\infty,
\end{aligned}\tag{3.79}$$

where we apply the Hölder inequality with $\frac{1}{2}+\frac{1}{q}=\frac{1}{p_1}$ and $\epsilon_H/(2+\epsilon_H)+2/(2+\epsilon_H)=1$ for the third inequality, along with the results from Corollary 3.2.60 and Lemma 3.2.57

$$\sup_{t\le r\le T}\int_r^T\mathbb{E}[|\mathbf{D}_rZ_s|^{(2+\epsilon_H)}]ds<\infty$$

and

$$\sup_{t\le r\le T}\int_r^T\mathbb{E}[|\mathbf{D}_rZ_s|^2]ds<\infty.$$

We complete the proof. □

Corollary 3.2.63. *We thereby get an estimate on $\mathbf{D}_sX_T^{t,x,\omega}$: for any $p\in[1,2)$,*

$$\sup_{0\le t\le T}\sup_{t\le s\le T}\|\mathbf{D}_sX_T^{t,x,\omega}\|_{L^p(\Omega)}\le C_{H,p,T,\psi,\|\omega\|_{[0,T]}}<\infty.\tag{3.80}$$

Corollary 3.2.64. *For any $p\in[1,2)$,*

$$\sup_{0\le t\le T}\sup_{t\le s\le T}\|\mathbf{D}_sX_T^{t,\overline{X}_t,\overline{\Theta}^t}\|_{L^p(\Omega)}\le C_{p,T,\psi,K,\overline{K}}<\infty.\tag{3.81}$$

Proof. It follows from Lemma 3.2.53, Remark 3.2.58 and growth condition on ϕ, ψ. □

Corollary 3.2.65. *For any $p \in [1, 2)$,*

$$\sup_{0\le t\le T}\sup_{t\le s\le T} \|\mathbf{D}_s\phi'(X_T^{t,x,\omega})\|_{L^p(\Omega)} \le C_{H,p,T,\psi,\phi,\|\omega\|_{[0,T]}} < \infty, \tag{3.82}$$

$$\sup_{0\le t\le T}\sup_{t\le s\le T} \|\mathbf{D}_s\phi'(X_T^{t,\overline{X}_t,\overline{\Theta}^t})\|_{L^p(\Omega)} \le C_{p,T,\psi,K,\overline{K}} < \infty. \tag{3.83}$$

Proof. It follows from Proposition 3.2.16 and note that $\phi''(X_T^{t,x,\omega}) \in L^q$ for any $\infty > q \ge 1$ by the growth condition on derivative of ϕ. $\square$

Remark 3.2.66. These estimates will be vital to prove regularity properties for $u(t, x, \omega)$ in the following proposition.

With these tools at hand, we can prove that u possesses sufficient regularity to handle singular directions by using the representation of its derivatives.

Proposition 3.2.67. *Let Assumption 3.2.49 (i)-(iv) hold. Then $u \in C^{0,2,2}_{+,\frac{1}{2}}(\Lambda)$.*

Proof. We have $u \in C^{0,2,2}_{+}(\bar{\Lambda})$ by Proposition 3.2.55 and the existence of the derivatives for u in singular directions follows from Lemma 3.2.62. So it's sufficient to check that u will satisfy the conditions of Definition 3.2.47. Most parts of the proof can follow the proof for Proposition 2.24 in [BJP25, Section 7.4].

Part (i) of [BJP25, Section 7.4] is totally workable in our case. Since we have Corollary 3.2.65 and similar growth conditions on ϕ and ψ.

When it comes to the part (ii) of [BJP25, Section 7.4], The only point that needs to be proved slightly differently lies in proving the ω-continuity for the second order path derivative of u where Hölder's inequality must be applied with care since the estimate in Corollary 3.2.65 only holds for $1 \le p < 2$. Actually we only need to prove it for the last term of the representation (3.73) since the proof for other terms is identical as [BJP25, Section 7.4].

We denote the conditional $L^p(\Omega)-$norm starting at (t, x) by $\|\cdot\|_{L^p_{t,x}}$. Let $G \in \mathcal{G}$ denote some control function having polynomial growth which may be different line to line. For $\eta_s^{(i)} \in \mathcal{X}$ supported on $[t, T]$ for $i = 1, 2$ and (t, x, ω), $(t, x, \omega') \in \bar{\Lambda}$. Let $1 \le p < 2$

and apply Hölder inequality on $1/p + 1/q = 1$, we have

$$
\begin{aligned}
&\mathbb{E}_{t,x}\left[\int_t^T \mathbf{D}_s\phi'(X_T^{\omega})\partial_{xx}^2\psi(s,Z_s^{\omega})\eta_s^{(1)}\eta_s^{(2)}ds\right] - \mathbb{E}_{t,x}\left[\int_t^T \mathbf{D}_s\phi'(X_T^{\omega'})\partial_{xx}^2\psi(s,Z_s^{\omega'})\eta_s^{(1)}\eta_s^{(2)}ds\right]\\
&= \mathbb{E}_{t,x}\Bigg[\int_t^{t+\delta}\Bigg\{\Big(\mathbf{D}_s\phi'(X_T^{\omega}) - \mathbf{D}_s\phi'(X_T^{\omega'})\Big)\partial_{xx}^2\psi(s,Z_s^{\omega})\\
&\qquad + \mathbf{D}_s\phi'(X_T^{\omega'})\left(\partial_{xx}^2\psi(s,Z_s^{\omega}) - \partial_{xx}^2\psi(s,Z_s^{\omega'})\right)\Bigg\}\eta_s^{(1)}\eta_s^{(2)}ds\Bigg]\\
&\le \int_t^{t+\delta}\left\|\mathbf{D}_s\phi'(X_T^{\omega}) - \mathbf{D}_s\phi'(X_T^{\omega'})\right\|_{L^p_{t,x}}\left\|\partial_{xx}^2\psi(s,Z_s^{\omega})\right\|_{L^q_{t,x}}ds\left\|\eta^{(1)}\right\|_{[0,T]}\left\|\eta^{(2)}\right\|_{[0,T]}\\
&\quad + \int_t^{t+\delta}\left\|\mathbf{D}_s\phi'(X_T^{\omega}) - \mathbf{D}_s\phi'(X_T^{\omega'})\right\|_{L^p_{t,x}}\left\|\partial_{xx}^2\psi(s,Z_s^{\omega}) - \partial_{xx}^2\psi(s,Z_s^{\omega'})\right\|_{L^q_{t,x}}ds\\
&\qquad\qquad \times\left\|\eta^{(1)}\right\|_{[0,T]}\left\|\eta^{(2)}\right\|_{[0,T]}.
\end{aligned}
$$

Concerning the second term on the right-hand side, first note that $\mathbf{D}_s\phi'(X_T^{\omega}), \mathbf{D}_s\phi'(X_T^{\omega'}) \in L^p$ from Corollary 3.2.65. Besides note that $\partial_{xx}^2\psi$ is C^1, we have

$$
\left\|\partial_{xx}^2\psi(s,Z_s^{\omega}) - \partial_{xx}^2\psi(s,Z_s^{\omega'})\right\|_{L^q} \le \left\|\int_0^1 \partial_{xxx}^3\psi_s(\theta Z_s^{\omega} + (1-\theta)Z_s^{\omega'})d\theta\right\|_{L^q}\|\omega-\omega'\|_{[0,T]} \tag{3.84}
$$

$$
\lesssim \left(\|G(x,\omega)\|_{[0,T]} + \|G(x,\omega')\|_{[0,T]}\right)\|\omega-\omega'\|_{[0,T]} \tag{3.85}
$$

by using intermediate value Theorem and Lemma 3.2.53.

On the other hand, for the first term, apply Hölder inequality on $p < p' < 2$, $1/p' + 1/q' = 1/p$, we have

$$
\begin{aligned}
&\left\|\mathbf{D}_s\phi'(X_T^{\omega}) - \mathbf{D}_s\phi'(X_T^{\omega'})\right\|_{L^p_{t,x}} = \left\|\phi''(X_T^{\omega})\mathbf{D}_sX_T^{\omega} - \phi''(X_T^{\omega'})\mathbf{D}_sX_T^{\omega'}\right\|_{L^p_{t,x}}\\
&\le \left\|\big(\phi''(X_T^{\omega}) - \phi''(X_T^{\omega'})\big)\mathbf{D}_sX_T^{\omega}\right\|_{L^p_{t,x}} + \left\|\phi''(X_T^{\omega'})\big(\mathbf{D}_sX_T^{\omega} - \mathbf{D}_sX_T^{\omega'}\big)\right\|_{L^p_{t,x}} \qquad (3.86)\\
&\lesssim \left\|\phi''(X_T^{\omega}) - \phi''(X_T^{\omega'})\right\|_{L^{q'}_{t,x}}\left\|\mathbf{D}_sX_T^{\omega}\right\|_{L^{p'}_{t,x}} + \left\|\phi''(X_T^{\omega'})\right\|_{L^{q'}_{t,x}}\Bigg[\left\|\psi(s,Z_s^{\omega}) - \psi(s,Z_s^{\omega'})\right\|_{L^{p'}_{t,x}}\\
&+ \left\|\int_s^T \mathbf{D}_sZ_r\big(\partial_x\psi_r(Z_r^{\omega}) - \partial_x\psi_r(Z_r^{\omega'})\big)dB_r\right\|_{L^{p'}_{t,x}}\\
&+ \left\|\int_s^T \mathbf{D}_sZ_r\big(\partial_x(\psi_r^2)(Z_r^{\omega}) - \partial_x(\psi_r^2)(Z_r^{\omega'})\big)dr\right\|_{L^{p'}_{t,x}}\Bigg]. \qquad (3.87)
\end{aligned}
$$

Using monotonicity of $p' \mapsto \|\cdot\|_{L^{p'}_{t,x}}$ $(p' < 2)$ and applying BDG inequality, Hölder inequality on $2/(2+\epsilon_H) + \epsilon_H/(2+\epsilon_H) = 1$, $1/2 + 1/p'' = 1/p'$, Jensen inequality with intermediate value Theorem on the third term of (3.87), we have the following inequalities for the third term and the fourth term of (3.87)

$$\left\| \int_s^T \mathbf{D}_s Z_r \big(\partial_x \psi_r(Z_r^\omega) - \partial_x \psi_r(Z_r^{\omega'})\big) dB_r \right\|_{L^2_{t,x}} + \left\| \int_s^T \mathbf{D}_s Z_r \big(\partial_x(\psi_r^2)(Z_r^\omega) - \partial_x(\psi_r^2)(Z_r^{\omega'})\big) dr \right\|_{L^{p'}_{t,x}} \tag{3.88}$$

$$\leq \Big(\int_s^T \mathbb{E}_{t,x}[|\mathbf{D}_s Z_r|^2 \big(\partial_x \psi_r(Z_r^\omega) - \partial_x \psi_r(Z_r^{\omega'})\big)^2] dr \Big)^{1/2}$$
$$+ \int_s^T \|\mathbf{D}_s Z_r \big(\partial_x(\psi_r^2)(Z_r^\omega) - \partial_x(\psi_r^2)(Z_r^{\omega'})\big)\|_{L^{p'}_{t,x}} dr \tag{3.89}$$

$$\lesssim \Big(\int_s^T \mathbb{E}_{t,x}[|\mathbf{D}_s Z_r|^{2+\epsilon_H} dr \Big)^{1/(2+\epsilon_H)}$$
$$\cdot \Big(\int_0^1 \|\partial^2_{xx}(\psi_r)(\theta Z_r^{\omega'} + (1-\theta) Z_r^\omega)\|^2_{L^{2(2+\epsilon_H)/\epsilon_H}_{t,x}} d\theta \Big)^{1/2} \|\omega - \omega'\|_{[0,T]}$$
$$+ \sup_{t \leq r \leq T} \| \int_0^1 \|\partial^2_{xx}(\psi_r^2)(\theta Z_r^{\omega'} + (1-\theta) Z_r^\omega) d\theta\|_{L^{p''}_{t,x}} \|\omega - \omega'\|_{[0,T]} \int_s^T \|\mathbf{D}_s Z_r\|_{L^2_{t,x}} dr \tag{3.90}$$

$$\lesssim \Big[\sup_{t \leq r \leq T} \| \sup_{\theta \in [0,1]} \partial^2_{xx}(\psi_r)(\theta Z_r^{\omega'} + (1-\theta) Z_r^\omega)\|_{L^{2(2+\epsilon_H)/\epsilon_H}}$$
$$+ \sup_{t \leq r \leq T} \| \sup_{\theta \in [0,1]} \partial^2_{xx}(\psi_r^2)(\theta Z_r^{\omega'} + (1-\theta) Z_r^\omega)\|_{L^{p''}_{t,x}} \Big] \|\omega - \omega'\|_{[0,T]} \tag{3.91}$$

$$\lesssim (\|G(x,\omega)\|_{[0,T]} + \|G(x,\omega')\|_{[0,T]}) \|\omega - \omega'\|_{[0,T]} \tag{3.92}$$

follows from Corollary 3.2.60 and Lemma 3.60. Besides, the first two terms of (3.87) can be estimated similarly by the regularity of ψ and ϕ and Lemma 3.2.53. Therefore, we can conclude that

$$\mathbb{E}_{t,x}\left[\int_t^T \mathbf{D}_s \phi'(X_T^\omega) \partial^2_{xx} \psi(s, Z_s^\omega) \eta_s^{(1)} \eta_s^{(2)} ds \right] - \mathbb{E}_{t,x}\left[\int_t^T \mathbf{D}_s \phi'(X_T^{\omega'}) \partial^2_{xx} \psi(s, Z_s^{\omega'}) \eta_s^{(1)} \eta_s^{(2)} ds \right]$$
$$\lesssim (\|G(x,\omega)\|_{[0,T]} + \|G(x,\omega')\|_{[0,T]}) \|\omega - \omega'\|_{[0,T]} \left\|\eta^{(1)}\right\|_{[0,T]} \left\|\eta^{(2)}\right\|_{[0,T]} \delta. \tag{3.93}$$

Next we're going to prove the time-continuity for $\partial_\omega u$ and $\partial^2_{\omega\omega} u$ in Definition 3.2.47(iii), as was done in the part (iii) of [BJP25, Section 7.4]. We can see from the representation (3.71), the case for $\partial_\omega u$ is easy and can follow the case for $\partial^2_{\omega\omega} u$. So we only prove the time-continuity for $\partial^2_{\omega\omega} u$.

We fix $(t, x, \omega) \in \bar{\Lambda}$. Note that to prove the first term of (3.73) has time continuity, WLOG, let $\zeta = 0$ since it suffices to treat the trickiest term containing the two stochastic integrals. For $\delta > 0$,

$$
\begin{aligned}
&\mathbb{E}_{x,\omega}\left[\phi''(X_T^{t+\delta})\left(\int_{t+\delta}^T \partial_x\psi(s,Z_s^{t+\delta})\eta_s^{(1)}\,dB_s\right)\left(\int_{t+\delta}^T \partial_x\psi(s,Z_s^{t+\delta})\eta_s^{(2)}\,dB_s\right)\right]\\
&-\mathbb{E}_{x,\omega}\left[\phi''(X_T^{t})\left(\int_{t}^T \partial_x\psi(s,Z_s^{t})\eta_s^{(1)}\,dB_s\right)\left(\int_{t}^T \partial_x\psi(s,Z_s^{t})\eta_s^{(2)}\,dB_s\right)\right]\\
&=\mathbb{E}_{x,\omega}\left[\left(\phi''(X_T^{t+\delta})-\phi''(X_T^t)\right)\left(\int_{t+\delta}^T \partial_x\psi(s,Z_s^{t+\delta})\eta_s^{(1)}\,dB_s\right)\left(\int_{t+\delta}^T \partial_x\psi(s,Z_s^{t+\delta})\eta_s^{(2)}\,dB_s\right)\right] && (3.94)\\
&+\mathbb{E}_{x,\omega}\left[\phi''(X_T^{t})\left(\int_{t+\delta}^T \partial_x\psi(s,Z_s^{t+\delta})\eta_s^{(1)}\,dB_s-\int_{t}^T \partial_x\psi(s,Z_s^{t})\eta_s^{(1)}\,dB_s\right)\right.\\
&\qquad\qquad\left.\times\left(\int_{t+\delta}^T \partial_x\psi(s,Z_s^{t+\delta})\eta_s^{(2)}\,dB_s\right)\right] && (3.95)\\
&+\mathbb{E}_{x,\omega}\left[\phi''(X_T^{t})\left(\int_{t}^T \partial_x\psi(s,Z_s^{t})\eta_s^{(1)}\,dB_s\right)\right.\\
&\qquad\qquad\left.\times\left(\int_{t+\delta}^T \partial_x\psi(s,Z_s^{t+\delta})\eta_s^{(2)}\,dB_s-\int_{t}^T \partial_x\psi(s,Z_s^{t})\eta_s^{(2)}\,dB_s\right)\right]. && (3.96)
\end{aligned}
$$

For (3.94), we apply Hölder inequality on $1/2+1/4+1/4=1$ to focus on the difference

$$
\left\|\phi''(X_T^{t+\delta})-\phi''(X_T^t)\right\|_{L^2_{x,\omega}} \le \left\|X_T^{t+\delta}-X_T^t\right\|_{L^4_{x,\omega}}\left\|\int_0^1 \phi'''(\lambda X_T^{t+\delta}+(1-\lambda)X_T^t)\,d\lambda\right\|_{L^4_{x,\omega}}
$$

since the other two terms of (3.94) belong to $L^4_{x,\omega}$ thereby they can be bounded by Lemma 3.2.53. Furthermore, apply BDG and Jensen's inequalities, we have

$$
\begin{aligned}
&\left\|X_T^{t+\delta}-X_T^t\right\|^4_{L^4_{x,\omega}} = \mathbb{E}_{x,\omega}\left[\left(\int_{t+\delta}^T\left(\psi(s,Z_s^{t+\delta})-\psi(s,Z_s^t)\right)dB_s+\int_t^{t+\delta}\psi(s,Z_s^t)\,dB_s\right)^4\right]\\
&\le CT\int_t^T \mathbb{E}_{x,\omega}\left[\left|\psi(s,Z_s^{t+\delta})-\psi(s,Z_s^t)\right|^4\right]ds + C\delta\int_t^{t+\delta}\mathbb{E}_{x,\omega}\left[\psi(s,Z_s)^4\right]ds, && (3.97)
\end{aligned}
$$

where the second term of (3.97) clearly goes to zero as $\delta\to 0$. For the first term of (3.97), we write

$$
\begin{aligned}
&\mathbb{E}_{x,\omega}\left[\left|\psi(s,Z_s^{t+\delta})-\psi(s,Z_s^t)\right|^4\right]\\
&=\mathbb{E}_{x,\omega}\left[\left|\left(Z_s^{t+\delta}-Z_s^t\right)\int_0^1\partial_x\psi\left(s,Z_s^t+\lambda\left(Z_s^{t+\delta}-Z_s^t\right)\right)d\lambda\right|^4\right]\\
&\le \mathbb{E}_{x,\omega}\left[\left|Z_s^{t+\delta}-Z_s^t\right|^8\right]^{1/2}\mathbb{E}_{x,\omega}\left[\left|\int_0^1\partial_x\psi\left(s,Z_s^t+\lambda\left(Z_s^{t+\delta}-Z_s^t\right)\right)d\lambda\right|^8\right]^{1/2} && (3.98)\\
&\le C\mathbb{E}_{x,\omega}\left[\left|Z_s^{t+\delta}-Z_s^t\right|^8\right]^{1/2},
\end{aligned}
$$

where the integral can be bounded in $L^8_{x,\omega}$ from Lemma 3.2.53 and the growth condition on ψ. Besides apply c_p-inequality, BDG inequality, Hölder inequality and Jensen's inequality, we can obtain

$$\begin{aligned}
&\mathbb{E}_{x,\omega}\left[\left|Z_s^{t+\delta}-Z_s^t\right|^8\right]=\mathbb{E}_{x,\omega}\left[\left|\int_t^{t+\delta}K(s,r)[b(Z_r)dr+\sigma(Z_r)\,dB_r]\right|^8\right]\\
&\lesssim \mathbb{E}_{x,\omega}\left[\left|\int_t^{t+\delta}K(s,r)b(Z_r)dr\right|^8\right]+\mathbb{E}_{x,\omega}\left[\left|\int_t^{t+\delta}K(s,r)\sigma(Z_r)\,dB_r\right|^8\right]\\
&\lesssim (\int_t^{t+\delta}K^2(s,r)dr)^4\int_t^{t+\delta}\mathbb{E}_{x,\omega}\left[|b(Z_r)|^8\right]dr+\mathbb{E}_{x,\omega}\left[\left|\int_t^{t+\delta}K^2(s,r)\sigma^2(Z_r)dr\right|^4\right]\\
&\lesssim \delta+(\int_t^{t+\delta}K^{2+\epsilon_H}(s,r)dr)^{2(2+\epsilon_H)}(\int_t^{t+\delta}\mathbb{E}[\sigma^{2(2+\epsilon_H)/\epsilon_H}(Z_r)dr])^{4\epsilon_H/(2+\epsilon_H)}\\
&\to 0\quad(\delta\to 0)
\end{aligned}\tag{3.99}$$

since there actually exists a $\epsilon_H>0$ small enough, s.t., $(2+\epsilon_H)(H-1)>-1$ so that $K^{2+\epsilon_H}\in L^1$. Similarly we can also see

$$\left\|\int_0^1\phi'''(\lambda X_T^{t+\delta}+(1-\lambda)X_T^t)\,d\lambda\right\|_{L^4_{x,\omega}}<\infty$$

which yields that (3.94) goes to zero with δ.

And note that the cases for (3.95) and (3.96) are identical. Next we only deal with the second one. We decompose further

$$\begin{aligned}
&\int_{t+\delta}^T\partial_x\psi(s,Z_s^{t+\delta})\eta_s^{(1)}\,dB_s-\int_t^T\partial_x\psi(s,Z_s^t)\eta_s^{(1)}\,dB_s\\
&=\int_{t+\delta}^T\big(\partial_x\psi(s,Z_s^{t+\delta})-\partial_x\psi(s,Z_s^t)\big)\eta_s^{(1)}\,dB_s+\int_t^{t+\delta}\partial_x\psi(s,Z_s^t)\eta_s^{(1)}\,dB_s.
\end{aligned}$$

By the same argument as in the part (iii) of [BJP25, Section 7.4], the integral between t and $t+\delta$ was already showed that it tends to zero in $L^p_{x,\omega}$ norm as $\delta\downarrow 0$. Since $\partial_x\psi$ is also C^1, the same computations as above show that $\mathbb{E}_{x,\omega}[\left|\partial_x\psi(s,Z_s^{t+\delta})-\partial_x\psi(s,Z_s^t)\right|^4]$ goes to zero, which concludes the time continuity of the first term of the representation (3.73). The second term in the representation (3.73) can be handled easily, as it follows from the previous argument. Regarding the last term of representation (3.73),

$$\begin{aligned}
&\mathbb{E}_{x,\omega}\left[\int_{t+\delta}^{T}\mathbf{D}_s\phi'(X_T^{t+\delta})\partial_{xx}^2\psi(s,Z_s^{t+\delta})K(s,t+\delta)^2ds-\int_t^T\mathbf{D}_s\phi'(X_T^t)\partial_{xx}^2\psi(s,Z_s^t)K(s,t)^2ds\right]\\
&=\mathbb{E}_{x,\omega}\bigg[\int_{t+\delta}^{T}\Big\{\mathbf{D}_s\phi'(X_T^{t+\delta})\partial_{xx}^2\psi(s,Z_s^{t+\delta})K(s,t+\delta)^2-\mathbf{D}_s\phi'(X_T^t)\partial_{xx}^2\psi(s,Z_s^t)K(s,t)^2\Big\}ds\\
&\qquad+\int_t^{t+\delta}\mathbf{D}_s\phi'(X_T^t)\partial_{xx}^2\psi(s,Z_s^t)K(s,t)^2ds\bigg] \qquad (3.100)\\
&=\mathbb{E}_{x,\omega}\left[\int_{t+\delta}^{T}\Big\{\mathbf{D}_s\phi'(X_T^{t+\delta})-\mathbf{D}_s\phi'(X_T^t)\Big\}\partial_{xx}^2\psi(s,Z_s^{t+\delta})K(s,t+\delta)^2ds\right]\\
&\qquad+\mathbb{E}_{x,\omega}\left[\int_{t+\delta}^{T}\mathbf{D}_s\phi'(X_T^t)\Big\{\partial_{xx}^2\psi(s,Z_s^{t+\delta})-\partial_{xx}^2\psi(s,Z_s^t)\Big\}K(s,t+\delta)^2ds\right]\\
&\qquad+\mathbb{E}_{x,\omega}\left[\int_{t+\delta}^{T}\mathbf{D}_s\phi'(X_T^t)\partial_{xx}^2\psi(s,Z_s^t)\Big\{K(s,t+\delta)^2-K(s,t)^2\Big\}ds\right]\\
&\qquad+\mathbb{E}_{x,\omega}\left[\int_t^{t+\delta}\mathbf{D}_s\phi'(X_T^t)\partial_{xx}^2\psi(s,Z_s^t)K(s,t)^2ds\right]. \qquad (3.101)
\end{aligned}$$

For the first term of (3.101), apply Hölder inequality for $p\in(1,2)$,

$$\begin{aligned}
&\mathbb{E}_{x,\omega}\left[\int_{t+\delta}^{T}\Big\{\mathbf{D}_s\phi'(X_T^{t+\delta})-\mathbf{D}_s\phi'(X_T^t)\Big\}\partial_{xx}^2\psi(s,Z_s^{t+\delta})K(s,t+\delta)^2ds\right]\\
&\leq\int_{t+\delta}^{T}\|\mathbf{D}_s\phi'(X_T^{t+\delta})-\mathbf{D}_s\phi'(X_T^t)\|_{L^p_{x,\omega}}\sup_{s\in[t,T]}\|\partial_{xx}^2\psi(s,Z_s^{t+\delta})\|_{L^q_{x,\omega}}K(s,t+\delta)^2ds.
\end{aligned} \tag{3.102}$$

Similarly to the estimate in (3.87), we can show that

$$\sup_{s\in[t,T]}\|\mathbf{D}_s\phi'(X_T^{t+\delta})-\mathbf{D}_s\phi'(X_T^t)\|_{L^p_{x,\omega}}\to 0 \quad (\delta\to 0).$$

Indeed, the first two terms in (3.87) converge to zero by Corollary 3.2.65. Moreover, applying the intermediate value theorem to ϕ'', together with

$$\|X_T^t-X_T^{t+\delta}\|_{L^p_{x,\omega}}\to 0,$$

which follows from Lemma 3.2.53, yields the desired control. Finally, the remaining terms can be handled using the BDG inequality, Corollary 3.2.60, Lemma 3.60, Lemma 3.2.53, and the inequality (3.103) below.

The estimation for the second term of (3.101) follows from using Hölder inequality and

$$\sup_{s\in[t,T]}\|\mathbf{D}_s\phi'(X_T^t)\|_{L^{3/2}_{x,\omega}}<\infty,$$

along with

$$\begin{aligned}\sup_{T\geq s\geq t+\delta}\|\partial^2_{xx}\psi(s,Z_s^{t+\delta})-\partial^2_{xx}\psi(s,Z_s^t)\|_{L^4_{x,\omega}} &\lesssim \delta+\sup_{T\geq s\geq t+\delta}(\int_t^{t+\delta}K^{2+\epsilon_H}(s,r)dr)^{2(2+\epsilon_H)}\\ &\lesssim \delta+\sup_{T\geq s\geq t+\delta}(\int_t^{t+\delta}(s-r)^{2H'-2}dr)^{2(2+\epsilon_H)}\\ &\to 0 \quad (\delta\to 0)\end{aligned} \tag{3.103}$$

by the same estimate as in (3.98), applied with ψ replaced by $\partial^2_{xx}\psi$, where $H'\in(1/2,1)$ chosen in Corollary 3.2.60 for ϵ_H. For the third one and the last one, they both tend to zero since we have

$$\sup_{t\leq s\leq T}\mathbb{E}_{x,\omega}[\big|\mathbf{D}_s\phi'(X_T^t)\partial^2_{xx}\psi(s,Z_s^t)\big|]<\infty,$$

follows from the argument in the proof for Corollary 3.2.65, along with

$$\left|\int_{t+\delta}^T\Big(K(s,t+\delta)^2-K(s,t)^2\Big)ds\right|\lesssim\delta^{2H}$$

and

$$\int_t^{t+\delta}K(s,t)^2ds\lesssim\delta^{2H}$$

from the Assumption 3.2.49 (iv). This concludes Definition 3.2.47 (iii) and hence the proof of the whole proposition. □

3.2.5. Path-dependent Feynman-Kac Formula and Functional Itô Formula

Having established sufficient regularity for the function $u(t,w,\omega)$, we now establish two key tools: the path-dependent Feynman-Kac formula and the functional Itô formula. These results will be used to estimate the error bound between our model and the approximate model in the previous subsection.

Proposition 3.2.68 (Path dependent Feynman-Kac formula)**.** *Let Assumptions 3.2.49 (i)-(iv) hold. Then the time derivative of $u:(t,x,\omega)\to\mathbb{E}[\phi(X_T^{t,x,\omega})]$ exists and $u\in C^{1,2,2}_{+,\frac{1}{2}}(\Lambda)$. Besides, u is the unique classical solution of the path-dependent PDE:*

$$\begin{aligned}&\partial_t u(t,x,\omega)+\zeta\psi^2(t,\omega_t)\partial_x u(t,x,\omega)+b(\omega_t)\langle\partial_\omega u(t,x,\omega),K(\cdot-t)\rangle\\ &\qquad+\psi(t,\omega_t)\sigma(\omega_t)\langle\partial_\omega(\partial_x u)(t,x,\omega),K(\cdot-t)\rangle+\frac{1}{2}\psi^2(t,\omega_t)\partial^2_{xx}u(t,x,\omega)\\ &\qquad+\frac{1}{2}\sigma^2(\omega_t)\langle\partial^2_{\omega\omega}u(t,x,\omega),(K(\cdot-t),K(\cdot-t))\rangle=0,\end{aligned}$$

with terminal condition $u(T,x,\omega)=\phi(x)$, for any $(t,x,\omega)\in\bar{\Lambda}$.

Proof. Note $U(t,\boldsymbol{\omega}) := u(t,\omega_t^{(1)},\omega^{(2)}|_{[t,T]}) \in C^{0,2}_{+,\frac{1}{2}}$ in the sense that it satisfies [BJP25, Definition 2.7 (i), (ii)], which follows from Proposition 3.2.67. The remaining conditions of [BJP25, Proposition 2.13, 2.14] are satisfied by Lemma 3.2.53 and Assumption 3.2.49 (iv). Therefore, we can directly apply [BJP25, Proposition 2.13, 2.14] to our model (3.46). Therefore, $U(t,\boldsymbol{\omega})$ is the unique solution to the path-dependent PDE

$$\partial_t U + \frac{1}{2}\langle \partial^2_{\boldsymbol{\omega}\boldsymbol{\omega}} U, (\sigma^{t,\boldsymbol{\omega}},\sigma^{t,\boldsymbol{\omega}})\rangle + \langle \partial_{\boldsymbol{\omega}} U, b^{t,\boldsymbol{\omega}}\rangle = 0, \tag{3.104}$$

with terminal condition $U(T,\boldsymbol{\omega}) = \phi(\omega_T^{(1)})$, where the coefficients are given by

$$b(t,r,x_1,x_2) = \begin{bmatrix} \zeta\psi(r,x_2)^2 \\ K(t,r)b(x_2)\end{bmatrix}, \qquad \sigma(t,r,x_1,x_2) = \begin{bmatrix} 0 & \psi(r,x_2) \\ 0 & K(t,r)\sigma(x_2)\end{bmatrix}. \tag{3.105}$$

Hense, with the help of Remark 3.2.40, we have

$$\begin{aligned}\langle \partial_{\boldsymbol{\omega}} U, b^{t,\boldsymbol{\omega}}\rangle &= \langle \partial_{\omega^{(1)}} U, \zeta\psi(t,\omega_t^{(2)})^2\rangle + \langle \partial_{\omega^{(2)}} U, K(\cdot - t)b(\omega^{(2)})\rangle \\ &= \zeta\psi(t,\omega_t)^2\partial_x u + b(\omega_t)\langle \partial_\omega u(t,x,\omega), K(\cdot - t)\rangle\end{aligned}$$

and

$$\begin{aligned}\langle \partial^2_{\boldsymbol{\omega}\boldsymbol{\omega}} U, (\sigma^{t,\boldsymbol{\omega}},\sigma^{t,\boldsymbol{\omega}})\rangle &= \langle \partial^2_{\omega^{(1)}\omega^{(1)}} U, (\psi(t,\omega_t^{(2)}),\psi(t,\omega_t^{(2)}))\rangle \\ &\quad + 2\langle \partial^2_{\omega^{(1)}\omega^{(2)}} U, (\psi(t,\omega_t^{(2)}), K(\cdot - t)\sigma(\omega_t^{(2)}))\rangle \\ &\quad + \sigma(\omega_t^{(2)})^2\langle \partial^2_{\omega^{(2)}\omega^{(2)}} U, (K(\cdot - t), K(\cdot - t))\rangle \\ &= \psi(t,\omega_t)^2\partial^2_{xx}u + 2\psi(t,\omega_t)\sigma(\omega_t)\langle \partial_\omega(\partial_x u), K(\cdot - t)\rangle \\ &\quad + \sigma(\omega_t)^2\langle \partial^2_{\omega\omega}u, (K(\cdot - t), K(\cdot - t))\rangle.\end{aligned}$$

Substituting them into the equation (3.104) for U, we obtain the desired result. ☐

Proposition 3.2.69 (Functional Itô formula). *Let Assumptions 3.2.49 (i)-(iv) hold. The following functional Itô formula holds for the function $u : (t,x,\omega) \to \mathbb{E}[\phi(X_T^{t,x,\omega})]$ and the semimartingale $(\overline{X}_t, \overline{\Theta}^t)$:*

$$\begin{aligned}u(t,\overline{X}_t,\overline{\Theta}^t) = u(0,X_0,0) &+ \int_0^t \partial_t u(s,\overline{X}_s,\overline{\Theta}^s)ds + \int_0^t \zeta\psi^2(s,\overline{Z}_s)\partial_x u(s,\overline{X}_s,\overline{\Theta}^s)ds \\ &+ \int_0^t b(\overline{Z}_s)\langle \partial_\omega u(s,\overline{X}_s,\overline{\Theta}^s), \overline{K}(\cdot - s)\rangle ds \\ &+ \int_0^t \psi(s,\overline{Z}_s)\sigma(\overline{Z}_s)\langle \partial_\omega(\partial_x u)(s,\overline{X}_s,\overline{\Theta}^s), \overline{K}(\cdot - s)\rangle ds \\ &+ \frac{1}{2}\int_0^t \psi^2(s,\overline{Z}_s)\partial^2_{xx}u(s,\overline{X}_s,\overline{\Theta}^s)ds \\ &+ \frac{1}{2}\int_0^t \sigma^2(\overline{Z}_s)\langle \partial^2_{\omega\omega}u(s,\overline{X}_s,\overline{\Theta}^s), (\overline{K}(\cdot - s),\overline{K}(\cdot - s))\rangle ds \\ &+ \int_0^t \psi(s,\overline{Z}_s)\partial_x u(s,\overline{X}_s,\overline{\Theta}^s)dB_s + \int_0^t \sigma(\overline{Z}_s)\langle \partial_\omega u(s,\overline{X}_s,\overline{\Theta}^s), \overline{K}(\cdot - s)\rangle dB_s.\end{aligned}$$

Proof. Due to Proposition 3.2.68, $U(t,\boldsymbol{\omega}) \in C^{1,2}_{+,\frac{1}{2}}$ in the sense of [BJP25, Definition 2.7]. So we can now apply [BJP25, Theorem 2.10], i.e. the functional Itô formula to $U(t,\boldsymbol{\omega})$, which yields

$$dU_t = \left(\partial_t U_t + \frac{1}{2}\langle \partial^2_{\boldsymbol{\omega}\boldsymbol{\omega}} U_t, (\sigma^{t,\boldsymbol{\omega}}, \sigma^{t,\boldsymbol{\omega}})\rangle + \langle \partial_{\boldsymbol{\omega}} U_t, b^{t,\boldsymbol{\omega}}\rangle\right) dt + \langle \partial_{\boldsymbol{\omega}} U_t, \sigma^{t,\boldsymbol{\omega}}\rangle dB_t,$$

then we can replace $\langle \partial^2_{\boldsymbol{\omega}\boldsymbol{\omega}} U_t, (\sigma^{t,\boldsymbol{\omega}}, \sigma^{t,\boldsymbol{\omega}})\rangle$, $\langle \partial_{\boldsymbol{\omega}} U_t, b^{t,\boldsymbol{\omega}}\rangle$ and $\langle \partial_{\boldsymbol{\omega}} U_t, \sigma^{t,\boldsymbol{\omega}}\rangle$ by the calculations above to get the formula for u.

□

3.2.6. Estimation for Error Bound

Denote the weak error associated with the approximate model (3.47) by

$$\mathcal{E}_T := \mathbb{E}[\phi(\overline{X}_T)] - \mathbb{E}[\phi(X_T)].$$

We now estimate this error using the tools introduced in the previous sections. The main ideas follow [BMM25, Section 5].

First, note that

$$\mathbb{E}[\phi(X_T)] = \mathbb{E}[u(0,x,0)]$$

and

$$\mathbb{E}[\phi(\overline{X}_T)] = \mathbb{E}[u(T,\overline{X}_T,\overline{\Theta}^T)].$$

Applying the functional Itô formula in Proposition 3.2.69 on function $u \in C^{1,2,2}_{+,\frac{1}{2}}(\Lambda)$ and the semi-martingale $(\overline{X},\overline{\Theta})$ for $t = T$. Taking expectations on both sides, and noting that all the Itô integrals are martingales and thus have zero expectation by Assumption 3.2.49, we obtain

$$\begin{aligned}
\mathcal{E}_T = \mathbb{E}\bigg(&\int_0^T \partial_t u(s,\overline{X}_s,\overline{\Theta}^s)ds + \int_0^T \zeta\psi^2(s,\overline{Z}_s)\partial_x u(s,\overline{X}_s,\overline{\Theta}^s)ds \\
&+ \int_0^T b(\overline{Z}_s)\langle \partial_\omega u(s,\overline{X}_s,\overline{\Theta}^s), \overline{K}(\cdot - s)\rangle ds \\
&+ \int_0^T \psi(s,\overline{Z}_s)\sigma(\overline{Z}_s)\langle \partial_\omega(\partial_x u)(s,\overline{X}_s,\overline{\Theta}^s), \overline{K}(\cdot - s)\rangle ds \\
&+ \frac{1}{2}\int_0^T \psi^2(s,\overline{Z}_s)\partial^2_{xx} u(s,\overline{X}_s,\overline{\Theta}^s)ds \\
&+ \frac{1}{2}\int_0^T \sigma^2(\overline{Z}_s)\langle \partial^2_{\omega\omega} u(s,\overline{X}_s,\overline{\Theta}^s), (\overline{K}(\cdot - s),\overline{K}(\cdot - s))\rangle ds\bigg).
\end{aligned} \tag{3.106}$$

Using the path-dependent Feynman–Kac formula in Proposition 3.2.68, we can rewrite the first integral on the right-hand side of (3.106) in terms of the path derivatives of u.

All derivatives that are not taken in the direction of the kernel cancel with those in (3.106), and we are left with

$$\begin{aligned}\mathcal{E}_T =&\mathbb{E}\bigg(\int_0^T b(\overline{Z}_s)\langle \partial_\omega u(s,\overline{X}_s,\overline{\Theta}^s),\Delta K(\cdot - s)\rangle ds \\ &+ \int_0^T \psi(s,\overline{Z}_s)\sigma(\overline{Z}_s)\langle \partial_\omega(\partial_x u)(s,\overline{X}_s,\overline{\Theta}^s),\Delta K(\cdot - s)\rangle ds \\ &+ \frac{1}{2}\int_0^T \sigma^2(\overline{Z}_s)\langle \partial^2_{\omega\omega} u(s,\overline{X}_s,\overline{\Theta}^s),(\Delta K(\cdot - s),\Sigma K(\cdot - s))\rangle ds\bigg),\end{aligned}$$

where we used the bilinearity and the symmetry of the map $\partial^2_{\omega\omega}u(t,\overline{X}_t,\overline{\Theta}^t)$ to factorise the last term. We use now the representation formulas for the derivatives from Proposition 3.2.55 and Lemma 3.2.62 with $K_{(1)} = \Delta K$ and $K_{(2)} = \Sigma K$, both belonging to $L^2([0,T],\mathbb{R}_+)\cap C(0,T]$ and satisfying $\lim_{\delta\to 0}\delta K^2_{(i)}(\delta) = \lim_{\delta\to 0}\int_0^\delta K^2_{(i)}(r)dr = 0$ for $i=1,2$ by Assumption 3.2.49 (iv). From this, each of the above integrals can be decomposed into a sum of integrals, which we analyze separately. From the representation formulas of $\partial_\omega\partial_x u$ and $\partial_\omega u$, we get

$$\begin{aligned}&\mathbb{E}\left[\int_0^T b(\overline{Z}_t)\langle \partial_\omega u(t,\overline{X}_t,\overline{\Theta}^t),\Delta K(\cdot - t)\rangle dt\right] \\ &= \int_0^T\int_t^T \mathbb{E}\Big[b(\overline{Z}_t)\ \mathbb{E}\Big[\phi'(X_T^{t,\overline{X}_t,\overline{\Theta}^t})\ \partial_x(\psi^2)(s,Z_s^{t,\overline{\Theta}^t})\ \Big|\ \mathcal{F}_t\Big]\Big]\Delta K(s-t)dsdt \\ &\quad+ \int_0^T \mathbb{E}\Big[b(\overline{Z}_t)\mathbb{E}\Big[\phi'(X_T^{t,\overline{X}_t,\overline{\Theta}^t})\int_t^T \partial_x\psi(s,Z_s^{t,\overline{\Theta}^t})\Delta K(s-t)dB_s\ \Big|\ \mathcal{F}_t\Big]\Big]dt \\ &:= F_1 + F_2.\end{aligned}$$

$$\begin{aligned}&\mathbb{E}\left[\int_0^T \psi(t,\overline{Z}_t)\sigma(\overline{Z}_t)\langle \partial_\omega(\partial_x u(t,\overline{X}_t,\overline{\Theta}^t)),\Delta K(\cdot - t)\rangle dt\right] \\ &= \int_0^T\int_t^T \mathbb{E}\Big[\psi(t,\overline{Z}_t)\sigma(\overline{Z}_t)\ \mathbb{E}\Big[\phi''(X_T^{t,\overline{X}_t,\overline{\Theta}^t})\ \partial_x(\psi^2)(s,Z_s^{t,\overline{\Theta}^t})\ \Big|\ \mathcal{F}_t\Big]\Big]\Delta K(s-t)dsdt \\ &\quad+ \int_0^T \mathbb{E}\Big[\psi(t,\overline{Z}_t)\sigma(\overline{Z}_t)\mathbb{E}\Big[\phi''(X_T^{t,\overline{X}_t,\overline{\Theta}^t})\int_t^T \partial_x\psi(s,Z_s^{t,\overline{\Theta}^t})\Delta K(s-t)dB_s\ \Big|\ \mathcal{F}_t\Big]\Big]dt \\ &:= F'_1 + F'_2.\end{aligned}$$

From the representation formula of $\partial^2_{\omega\omega}u$, we get

$$
\begin{aligned}
&\mathbb{E}\left[\int_0^T \sigma^2(\overline{Z}_t)\langle \partial^2_{\omega\omega}u(t,\overline{X}_t,\overline{\Theta}^t),(\Delta K(\cdot-t),\Sigma K(\cdot-t))\rangle dt\right]\\
&=\int_0^T \mathbb{E}\Bigg[\sigma^2(\overline{Z}_t)\mathbb{E}\Bigg[\phi''(X_T^{t,\overline{X}_t,\overline{\Theta}^t})\times\\
&\Bigg\{\left(\int_t^T \partial_x(\psi^2)(s,Z_s^{t,\overline{\Theta}^t})\Delta K(s-t)ds\right)\left(\int_t^T \partial_x(\psi^2)(s,Z_s^{t,\overline{\Theta}^t})\Sigma K(s-t)ds\right)\\
&+\left(\int_t^T \partial_x(\psi^2)(s,Z_s^{t,\overline{\Theta}^t})\Delta K(s-t)ds\right)\left(\int_t^T \partial_x\psi(s,Z_s^{t,\overline{\Theta}^t})\Sigma K(s-t)dB_s\right)\\
&+\left(\int_t^T \partial_x(\psi^2)(s,Z_s^{t,\overline{\Theta}^t})\Sigma K(s-t)ds\right)\left(\int_t^T \partial_x\psi(s,Z_s^{t,\overline{\Theta}^t})\Delta K(s-t)dB_s\right)\\
&+\left(\int_t^T \partial_x\psi(s,Z_s^{t,\overline{\Theta}^t})\Sigma K(s-t)dB_s\right)\left(\int_t^T \partial_x\psi(s,Z_s^{t,\overline{\Theta}^t})\Delta K(s-t)dB_s\right)\Bigg\}\;\Big|\;\mathcal{F}_t\Bigg]\Bigg]dt\\
&+\int_0^T\int_t^T \mathbb{E}\Big[\sigma^2(\overline{Z}_t)\mathbb{E}\big[\phi'(X_T^{t,\overline{X}_t,\overline{\Theta}^t})\partial^2_{xx}(\psi^2)(s,Z_s^{t,\overline{\Theta}^t})\;\big|\;\mathcal{F}_t\big]\Big]\Delta(K^2)(s-t)dsdt\\
&+\int_0^T\int_t^T \mathbb{E}\Big[\sigma^2(\overline{Z}_t)\mathbb{E}\big[\phi''(X_T^{t,\overline{X}_t,\overline{\Theta}^t})\partial^2_{xx}\psi(s,Z_s^{t,\overline{\Theta}^t})\mathbf{D}_s X_T^{t,\overline{X}_t,\overline{\Theta}^t}\;\big|\;\mathcal{F}_t\big]\Big]\Delta(K^2)(s-t)dsdt\\
&:=F_3+F_4+F_5+F'+H_1+H_2.
\end{aligned}
$$

So, in total, we get a decomposition of $\mathcal{E}_T$

$$
\mathcal{E}_T = F_1+F_2+F_1'+F_2'+\frac{1}{2}\left(F_3+F_4+F_5\right)+\frac{1}{2}F'+\frac{1}{2}\left(H_1+H_2\right). \tag{3.107}
$$

In the following, we will analyze each terms separately since they have different structure. In fact, we will show that the terms F_1, F_2, F_3, F_4, and F_5 contribute to the error term of the type $\|K-\overline{K}\|_{L^1([0,T])}$, whereas the terms H_1 and H_2 contribute to the error term of the type $\|(K^2)-(\overline{K}^2)\|_{L^1([0,T])}$. The term F', slightly more technical to analyze, leads to both types of contributions.

Estimate of F_1'

The estimate for F_1' is straightforward. It suffices to bound the stochastic term, which contributes an error of order $\|K - \overline{K}\|_{L^1([0,T])}$.

Using Hölder's inequality directly to obtain

$$\begin{aligned}
|F_1'| &= \left| \int_0^T \int_t^T \mathbb{E}\Big[\psi(t, \overline{Z}_t)\sigma(\overline{Z}_t)\, \mathbb{E}\Big[\phi''(X_T^{t,\overline{X}_t,\overline{\Theta}^t})\ \partial_x(\psi^2)(s, Z_s^{t,\overline{\Theta}^t}) \ \Big|\ \mathcal{F}_t\Big]\Big] \Delta K(s-t) ds dt \right| \\
&\lesssim \int_0^T \int_t^T \|\psi(t, \overline{Z}_t)\sigma(\overline{Z}_t)\|_{L^2(\Omega)} \|\mathbb{E}\Big[\phi''(X_T^{t,\overline{X}_t,\overline{\Theta}^t})\ \partial_x(\psi^2)(s, Z_s^{t,\overline{\Theta}^t}) \ \Big|\ \mathcal{F}_t\Big]\|_{L^2(\Omega)} \\
&\quad \times |\Delta K(s-t)| ds dt \\
&\lesssim \sup_{t\in[0,T]} \|\psi(t, \overline{Z}_t)\sigma(\overline{Z}_t)\|_{L^2(\Omega)} \sup_{s\in[t,T]} \|\phi''(X_T^{t,\overline{X}_t,\overline{\Theta}^t})\ \partial_x(\psi^2)(s, Z_s^{t,\overline{\Theta}^t})\|_{L^2(\Omega)} \\
&\quad \times \int_0^T \int_t^T |\Delta K(s-t)| ds dt \\
&\lesssim \int_0^T \|K - \overline{K}\|_{L^1([0,t])} dt
\end{aligned} \tag{3.108}$$

follows from the relation $\overline{Z}_t = Z_t^{t,\overline{\Theta}^t}$, the growth conditions on ϕ and ψ, the contraction property of conditional expectation and Lemma 3.2.53.

Estimate of F_2'

Unlike F_1', the term F_2' contains a stochastic integral in which ΔK appears. Applying the Itô isometry directly would lead to a time integral of $(\Delta K)^2$, which is not desirable. Therefore, we introduce Malliavin calculus in the estimation, namely by applying the conditional Malliavin integration by parts from Proposition 3.2.18, which yields

$$\begin{aligned}
|F_2'| &= \left| \int_0^T \mathbb{E}\left[\psi(t, \overline{Z}_t)\sigma(\overline{Z}_t)\mathbb{E}\left[\phi''(X_T^{t,\overline{X}_t,\overline{\Theta}^t}) \int_t^T \partial_x \psi(s, Z_s^{t,\overline{\Theta}^t})\Delta K(s-t)dB_s \;\middle|\; \mathcal{F}_t \right]\right] dt \right| \\
&= \left| \int_0^T \mathbb{E}\left[\psi(t, \overline{Z}_t)\sigma(\overline{Z}_t) \int_t^T \mathbb{E}\left[\mathbf{D}_s \phi''(X_T^{t,\overline{X}_t,\overline{\Theta}^t})\partial_x \psi(s, Z_s^{t,\overline{\Theta}^t}) \;\middle|\; \mathcal{F}_t \right] \Delta K(s-t) ds \right] dt \right| \\
&= \left| \int_0^T \int_t^T \mathbb{E}\left[\psi(t, \overline{Z}_t)\sigma(\overline{Z}_t)\mathbb{E}\left[\mathbf{D}_s \phi''(X_T^{t,\overline{X}_t,\overline{\Theta}^t})\partial_x \psi(s, Z_s^{t,\overline{\Theta}^t}) \;\middle|\; \mathcal{F}_t \right]\right] \Delta K(s-t) ds dt \right| \\
&\leq \int_0^T \int_t^T \sup_{t\in[0,T]} \|\psi(t, \overline{Z}_t)\sigma(\overline{Z}_t)\|_{L^{q'}(\Omega)} \|\mathbf{D}_s \phi''(X_T^{t,\overline{X}_t,\overline{\Theta}^t})\partial_x \psi(s, Z_s^{t,\overline{\Theta}^t})\|_{L^{p'}(\Omega)} \\
&\quad \times |\Delta K(s-t)| ds dt \\
&\leq \int_0^T \int_t^T \sup_{t\in[0,T]} \|\psi(t, \overline{Z}_t)\sigma(\overline{Z}_t)\|_{L^{q'}(\Omega)} \sup_{t\in[0,T]} \sup_{s\in[t,T]} \|\mathbf{D}_s \phi''(X_T^{t,\overline{X}_t,\overline{\Theta}^t})\|_{L^{p}(\Omega)} \\
&\quad \times \sup_{t\in[0,T]} \sup_{s\in[t,T]} \|\partial_x \psi(s, Z_s^{t,\overline{\Theta}^t})\|_{L^{r}(\Omega)} |\Delta K(s-t)| ds dt \\
&\lesssim \int_0^T \|K - \overline{K}\|_{L^1([0,t])} dt
\end{aligned} \tag{3.109}$$

follows from Lemma 3.2.53 and Corollary 3.2.65, where we apply Hölder inequality on $1/p' + 1/q' = 1$, and $1/p + 1/r = 1/p'$ s.t. $p \in (1, 2)$ and notice that conditional expectation is a contraction on $L^p(\Omega)$.

Estimates of F_1 and F_2

We observe that F_1 and F_2 have the same structure as F_1' and F_2'. Their estimates can therefore be obtained in the same way using Lemma 3.2.53, yielding an error of order $\|K - \overline{K}\|_{L^1([0,T])}$.

Estimate of F_3

We can use a similar strategy as for F_1' to bound F_3. However, now F_3 involves the integral of a product of the time integral of ΔK and the time integral of ΣK. We first apply Fubini's theorem to interchange the expectation and the integral, leading to

$$
\begin{aligned}
F_3 = \int_0^T \mathbb{E}\Bigg[\sigma^2(\overline{Z}_t)\mathbb{E}\Bigg[\phi''(X_T^{t,\overline{X}_t,\overline{\Theta}^t}) \left(\int_t^T \partial_x(\psi^2)(s, Z_s^{t,\overline{\Theta}^t})\Delta K(s-t)ds\right) \\
\times \left(\int_t^T \partial_x(\psi^2)(s, Z_s^{t,\overline{\Theta}^t})\Sigma K(s-t)ds\right) \;\Bigg|\; \mathcal{F}_t\Bigg]\Bigg] dt \\
= \int_0^T \int_t^T \int_t^T \mathbb{E}\Bigg[\sigma^2(\overline{Z}_t)\mathbb{E}\Bigg[\phi''(X_T^{t,\overline{X}_t,\overline{\Theta}^t})\partial_x(\psi^2)(s_1, Z_{s_1}^{t,\overline{\Theta}^t})\partial_x(\psi^2)(s_2, Z_{s_2}^{t,\overline{\Theta}^t}) \;\Bigg|\; \mathcal{F}_t\Bigg]\Bigg] \\
\times \Delta K(s_1-t)\Sigma K(s_2-t)ds_1 ds_2 dt.
\end{aligned}
$$

Using the similar techniques as before, we assert

$$
\sup_{t\in[0,T]} \sup_{s_1\in[t,T]} \sup_{s_2\in[t,T]} \left|\mathbb{E}\Bigg[\sigma^2(\overline{Z}_t)\mathbb{E}\Bigg[\phi''(X_T^{t,\overline{X}_t,\overline{\Theta}^t})\partial_x(\psi^2)(s_1, Z_{s_1}^{t,\overline{\Theta}^t})\partial_x(\psi^2)(s_2, Z_{s_2}^{t,\overline{\Theta}^t}) \;\Bigg|\; \mathcal{F}_t\Bigg]\Bigg]\right| < \infty,
$$

follows from Hölder inequality, the growth conditions on ϕ, ψ and Lemma 3.2.53. Therefore, we have

$$
\begin{aligned}
|F_3| &\le C_{T,\phi,\psi} \int_t^T \int_t^T |\Delta K(s_1-t)\Sigma K(s_2-t)| ds_1 ds_2 dt \\
&\le C_{T,\phi,\psi} \int_0^T \|K-\overline{K}\|_{L^1([0,t])} dt \times \sup_{t\in[0,T]} \int_t^T |\Sigma K(s_2-t)| ds_2 \qquad (3.110) \\
&\le C_{T,K,\overline{K},\phi,\psi} \int_0^T \|K-\overline{K}\|_{L^1([0,t])} dt.
\end{aligned}
$$

Thanks to the Assumption 3.2.49 (iv), we can bound the time integral of ΣK in the last inequality.

Estimate of F_4

Different from F_3, the term F_4 also contains a stochastic integral. Fortunately, this stochastic integral involves ΣK rather than ΔK, which allows us to apply Itô's isometry directly to bound it.

First notice that the process $\left(\int_t^r \partial_x\psi(s,Z_s)\Sigma K(s-t)dB_s\right)_{r\in[t,T]}$ is a $(\mathcal{F}_r)_{r\in[t,T]}$-local martingale since $s\mapsto \partial_x\psi(s,Z_s)\Sigma K(s-t)$ is $(\mathcal{F}_s)_{s\in[t,T]}$-adapted and left-continuous on $[t,T]$, hence $(\mathcal{F}_s)_{s\in[t,T]}$-predictable. We can thereby apply Hölder inequality with $1/p+1/q=1$, $1/r'+1/p'+1/2=1/p$ along with Minkowsky inequality and BDG inequality to obtain

$$
\begin{aligned}
F_4 = &\int_0^T \mathbb{E}\Bigg[\sigma^2(\overline{Z}_t)\mathbb{E}\Bigg[\phi''(X_T^{t,\overline{X}_t,\overline{\Theta}^t})\left(\int_t^T \partial_x(\psi^2)(s,Z_s^{t,\overline{\Theta}^t})\Delta K(s-t)ds\right) \\
&\qquad\qquad \times\left(\int_t^T \partial_x\psi(s,Z_s^{t,\overline{\Theta}^t})\Sigma K(s-t)dB_s\right)\;\Big|\;\mathcal{F}_t\Bigg]\Bigg]dt \\
\lesssim &\int_0^T \|\sigma^2(\overline{Z}_t)\|_{L^p(\Omega)}\|\phi''(X_T^{t,\overline{X}_t,\overline{\Theta}^t})\left(\int_t^T \partial_x(\psi^2)(s,Z_s^{t,\overline{\Theta}^t})\Delta K(s-t)ds\right) \\
&\qquad\qquad \times\left(\int_t^T \partial_x\psi(s,Z_s^{t,\overline{\Theta}^t})\Sigma K(s-t)dB_s\right)\|_{L^q(\Omega)}dt \\
\le &\sup_{t\in[0,T]}\|\sigma^2(\overline{Z}_t)\|_{L^p(\Omega)}\int_0^T \|\phi''(X_T^{t,\overline{X}_t,\overline{\Theta}^t})\|_{L^{r'}(\Omega)} \\
&\qquad\qquad \times\left(\int_t^T \|\partial_x(\psi^2)(s,Z_s^{t,\overline{\Theta}^t})\|_{L^{p'}(\Omega)}\Delta K(s-t)ds\right) \\
&\qquad\qquad \times\left(\int_t^T \mathbb{E}[(\partial_x\psi)^2(s,Z_s^{t,\overline{\Theta}^t})](\Sigma K)^2(s-t)ds\right)^{1/2}dt \\
\lesssim &\sup_{t\in[0,T]}\|\phi''(X_T^{t,\overline{X}_t,\overline{\Theta}^t})\|_{L^4(\Omega)}\sup_{T\ge s\ge t\ge 0}\|\partial_x(\psi^2)(Z_s^{t,\overline{\Theta}^t})\|_{L^4(\Omega)}\sup_{T\ge s\ge t\ge 0}\|(\partial_x\psi)^2(Z_s^{t,\overline{\Theta}^t})\|_{L^1(\Omega)} \\
&\qquad\qquad \times\sup_{t\in[0,T]}\left(\int_t^T(\Sigma K)^2(s-t)ds\right)^{1/2}\int_0^T\int_t^T|\Delta K(s-t)|dsdt \\
\le &\, C_{T,K,\overline{K},\phi,\psi}\int_0^T \|K-\overline{K}\|_{L^1([0,t])}dt.
\end{aligned}
\tag{3.111}
$$

Estimate of F_5

We observe that ΔK appears inside the stochastic integral, which creates difficulties in the estimation. To avoid producing a term of the form $(\Delta K)^2$, we instead apply the conditional Malliavin integration by parts in Lemma 3.2.18 and Fubini's theorem, which yields

$$\begin{aligned}
F_5 = \int_0^T \mathbb{E}\Bigg[\sigma^2(\overline{Z}_t)\mathbb{E}\Bigg[\phi''(X_T^{t,\overline{X}_t,\overline{\Theta}^t})\left(\int_t^T \partial_x(\psi^2)(s, Z_s^{t,\overline{\Theta}^t})\Sigma K(s-t)ds\right) \\
\times \left(\int_t^T \partial_x\psi(s, Z_s^{t,\overline{\Theta}^t})\Delta K(s-t)dB_s\right) \Bigg| \mathcal{F}_t\Bigg]\Bigg]dt
\end{aligned} \tag{3.112}$$

$$\begin{aligned}
= \int_0^T \int_t^T \mathbb{E}\Bigg[\sigma^2(\overline{Z}_t)\mathbb{E}\Bigg[\mathbf{D}_s\Big\{\phi''(X_T^{t,\overline{X}_t,\overline{\Theta}^t})\left(\int_t^T \partial_x(\psi^2)(s, Z_s^{t,\overline{\Theta}^t})ds\right)\Big\} \\
\times \partial_x\psi(s, Z_s^{t,\overline{\Theta}^t}) \Bigg| \mathcal{F}_t\Bigg]\Bigg]\Delta K(s-t)dsdt.
\end{aligned} \tag{3.113}$$

By the product rule for the Malliavin derivative and by interchanging the Lebesgue integral with the Malliavin derivative, we have

$$\begin{aligned}
&\mathbf{D}_s\Big\{\phi''(X_T^{t,\overline{X}_t,\overline{\Theta}^t})\left(\int_t^T \partial_x(\psi^2)(s, Z_s^{t,\overline{\Theta}^t})ds\right)\Big\} \\
&= \mathbf{D}_s\phi''(X_T^{t,\overline{X}_t,\overline{\Theta}^t})\left(\int_t^T \partial_x(\psi^2)(r, Z_r^{t,\overline{\Theta}^t})\Sigma K(r-t)dr\right) \\
&\quad + \phi''(X_T^{t,\overline{X}_t,\overline{\Theta}^t})\left(\int_t^T \mathbf{D}_s\{\partial_x(\psi^2)(r, Z_r^{t,\overline{\Theta}^t})\}\Sigma K(r-t)dr\right).
\end{aligned} \tag{3.114}$$

Combined with the calculations

$$\mathbf{D}_s\phi''(X_T^{t,\overline{X}_t,\overline{\Theta}^t}) = \phi'''(X_T^{t,\overline{X}_t,\overline{\Theta}^t})\mathbf{D}_s X_T^{t,\overline{X}_t,\overline{\Theta}^t} \tag{3.115}$$

and

$$\mathbf{D}_s\{\partial_x(\psi^2)(r, Z_r^{t,\overline{\Theta}^t})\} = \partial_{xx}^2(\psi^2)(r, Z_r^{t,\overline{\Theta}^t})\mathbf{D}_s Z_r^{t,\overline{\Theta}^t}. \tag{3.116}$$

We can now rewrite F_5 into

$$F_5 = F_5^{(1)} + F_5^{(2)}, \tag{3.117}$$

where

$$\begin{aligned}
F_5^{(1)} = \int_0^T \int_t^T \mathbb{E}\Bigg[\sigma^2(\overline{Z}_t)\mathbb{E}\Bigg[\phi'''(X_T^{t,\overline{X}_t,\overline{\Theta}^t})\mathbf{D}_s X_T^{t,\overline{X}_t,\overline{\Theta}^t}\left(\int_t^T \partial_x(\psi^2)(r, Z_r^{t,\overline{\Theta}^t})\Sigma K(r-t)dr\right) \\
\times \partial_x\psi(s, Z_s^{t,\overline{\Theta}^t}) \Bigg| \mathcal{F}_t\Bigg]\Bigg]\Delta K(s-t)dsdt,
\end{aligned} \tag{3.118}$$

and

$$F_5^{(2)} = \int_0^T \int_t^T \mathbb{E}\Bigg[\sigma^2(\overline{Z}_t)\mathbb{E}\Big[\phi''(X_T^{t,\overline{X}_t,\overline{\Theta}^t}) \left(\int_t^T \partial^2_{xx}(\psi^2)(r, Z_r^{t,\overline{\Theta}^t})\mathbf{D}_s Z_r^{t,\overline{\Theta}^t} \Sigma K(r-t)dr\right) \\ \times \partial_x \psi(s, Z_s^{t,\overline{\Theta}^t}) \,\Big|\, \mathcal{F}_t\Big]\Bigg] \Delta K(s-t)dsdt. \tag{3.119}$$

As before, we use Corollary 3.2.64 and 3.2.65 and apply Hölder's inequality repeatedly to bound the expectations appearing in the two double integrals. Apply Hölder's inequality for $1/p + 1/q + 1/r = 1$ and $1/q' + 1/q'' + 1/q''' = 1/q$ where $q'' \in (1,2)$ and the contraction property of conditional expectation, we have

$$\begin{aligned}
|F_5^{(1)}| &\leq \int_0^T \int_t^T \|\phi'''(X_T^{t,\overline{X}_t,\overline{\Theta}^t})\mathbf{D}_s X_T^{t,\overline{X}_t,\overline{\Theta}^t} \left(\int_t^T \partial_x(\psi^2)(r, Z_r^{t,\overline{\Theta}^t})\Sigma K(r-t)dr\right) \|_{L^q(\Omega)} \\
&\quad \times \|\sigma^2(\overline{Z}_t)\|_{L^p(\Omega)} \|\partial_x \psi(s, Z_s^{t,\overline{\Theta}^t})\|_{L^r(\Omega)} \Delta K(s-t)dsdt \\
&\leq \sup_{t\in[0,T]} \|\sigma^2(\overline{Z}_t)\|_{L^p(\Omega)} \sup_{0\leq t\leq s\leq T} \|\partial_x \psi(s, Z_s^{t,\overline{\Theta}^t})\|_{L^r(\Omega)} \sup_{t\in[0,T]} \|\phi'''(X_T^{t,\overline{X}_t,\overline{\Theta}^t})\|_{L^{q'}(\Omega)} \\
&\quad \times \sup_{0\leq t\leq s\leq T} \|\mathbf{D}_s X_T^{t,\overline{X}_t,\overline{\Theta}^t}\|_{L^{q''}(\Omega)} \sup_{t\in[0,T]} \|\int_t^T \partial_x(\psi^2)(r, Z_r^{t,\overline{\Theta}^t})\Sigma K(r-t)dr\|_{L^{q'''}(\Omega)} \\
&\quad \times \int_0^T \int_t^T |\Delta K(s-t)|dsdt \\
&\leq C_{T,K,\overline{K},\phi,\psi} \int_0^T \|K - \overline{K}\|_{L^1([0,t])}dt,
\end{aligned} \tag{3.120}$$

follows from Corollary 3.2.64 and the estimate

$$\begin{aligned}
&\sup_{t\in[0,T]} \|\int_t^T \partial_x(\psi^2)(r, Z_r^{t,\overline{\Theta}^t})\Sigma K(r-t)dr\|_{L^{q'''}(\Omega)} \\
&\leq \sup_{t\in[0,T]} \int_t^T |\Sigma K(r-t)|dr \times \sup_{t\in[0,T]} \|\partial_x(\psi^2)(r, Z_r^{t,\overline{\Theta}^t})\|_{L^{q'''}(\Omega)} < \infty.
\end{aligned} \tag{3.121}$$

Again, apply Hölder's inequality for $1/p + 1/q + 1/r = 1$ and $1/q' + 1/q'' = 1/q$ where $q'' \in (1, 2)$ to obtain

$$\begin{aligned}
|F_5^{(2)}| \leq & \int_0^T \int_t^T \|\phi''(X_T^{t,\overline{X}_t,\overline{\Theta}^t}) \left(\int_t^T \partial^2_{xx}(\psi^2)(r, Z_r^{t,\overline{\Theta}^t}) \mathbf{D}_s Z_r^{t,\overline{\Theta}^t} \Sigma K(r-t) dr \right) \|_{L^q(\Omega)} \\
& \times \|\sigma^2(\overline{Z}_t)\|_{L^p(\Omega)} \|\partial_x \psi(s, Z_s^{t,\overline{\Theta}^t})\|_{L^r(\Omega)} \Delta K(s-t) ds dt \\
\leq & \sup_{t\in[0,T]} \|\sigma^2(\overline{Z}_t)\|_{L^p(\Omega)} \sup_{0\leq t\leq s\leq T} \|\partial_x \psi(s, Z_s^{t,\overline{\Theta}^t})\|_{L^r(\Omega)} \sup_{t\in[0,T]} \|\phi''(X_T^{t,\overline{X}_t,\overline{\Theta}^t})\|_{L^{q'}(\Omega)} \\
& \times \sup_{0\leq t\leq s\leq T} \| \int_t^T \partial^2_{xx}(\psi^2)(r, Z_r^{t,\overline{\Theta}^t}) \mathbf{D}_s Z_r^{t,\overline{\Theta}^t} \Sigma K(r-t) dr \|_{L^{q''}(\Omega)} \\
& \times \int_0^T \int_t^T |\Delta K(s-t)| ds dt \\
\leq & C_{T,\phi,\psi} \sup_{0\leq t\leq s\leq T} \| \int_t^T \partial^2_{xx}(\psi^2)(r, Z_r^{t,\overline{\Theta}^t}) \mathbf{D}_s Z_r^{t,\overline{\Theta}^t} \Sigma K(r-t) dr \|_{L^{q''}(\Omega)} \\
& \times \int_0^T \int_t^T |\Delta K(s-t)| ds dt.
\end{aligned} \tag{3.122}$$

Combined with the estimate

$$\begin{aligned}
& \sup_{0\leq t\leq s\leq T} \| \int_t^T \partial^2_{xx}(\psi^2)(r, Z_r^{t,\overline{\Theta}^t}) \mathbf{D}_s Z_r^{t,\overline{\Theta}^t} \Sigma K(r-t) dr \|_{L^{q''}(\Omega)} \\
\leq & \sup_{0\leq t\leq s\leq T} \int_t^T \|\partial^2_{xx}(\psi^2)(r, Z_r^{t,\overline{\Theta}^t}) \mathbf{D}_s Z_r^{t,\overline{\Theta}^t}\|_{L^{q''}(\Omega)} \Sigma K(r-t) dr \\
\leq & \sup_{0\leq t\leq r\leq T} \|\partial^2_{xx}(\psi^2)(r, Z_r^{t,\overline{\Theta}^t})\|_{L^{q'''}(\Omega)} \sup_{0\leq t\leq s\leq T} \int_t^T \|\mathbf{D}_s Z_r^{t,\overline{\Theta}^t}\|_{L^2(\Omega)} \Sigma K(r-t) dr \\
\lesssim & \sup_{0\leq t\leq s\leq T} (\int_t^T \|\mathbf{D}_s Z_r^{t,\overline{\Theta}^t}\|^2_{L^2(\Omega)} dr)^{1/2} \times \sup_{0\leq t\leq s\leq T} (\int_t^T |\Sigma K(r-t)|^2 dr)^{1/2} \\
= & \sup_{s\in[0,T]} (\int_s^T \|\mathbf{D}_s Z_r\|^2_{L^2(\Omega)} dr)^{1/2} \times \sup_{0\leq t\leq s\leq T} (\int_t^T |\Sigma K(r-t)|^2 dr)^{1/2} < \infty,
\end{aligned} \tag{3.123}$$

follows from Lemma 3.2.57 and Remark 3.2.58, where we apply Hölder's inequality on $1/q'' = 1/2 + 1/q'''$.

Therefore, collecting these estimates together, we conclude that

$$|F_5| \leq C_{T,K,\overline{K},\phi,\psi} \int_0^T \|K - \overline{K}\|_{L^1([0,t])} dt. \tag{3.124}$$

Estimate of F'

The estimation of F' is the most delicate part, since it involves the product of two stochastic integrals, which are multiplied by the stochastic process $\phi''(X_T^{t,\overline{X}_t,\overline{\Theta}^t})$, preventing the direct use of Itô's isometry. As in the estimation of F_5, we first apply the Malliavin integration by parts to remove one of the stochastic integrals. We remark here this will yield errors of order $\|K-\overline{K}\|_{L^1([0,T])}$ and $\|K^2-\overline{K}^2\|_{L^1([0,T])}$.

Using the Malliavin integration by part in Proposition 3.2.18 on F', we have

$$\begin{aligned} F' &= \int_0^T \mathbb{E}\bigg[\sigma^2(\overline{Z}_t)\mathbb{E}\bigg[\phi''(X_T^{t,\overline{X}_t,\overline{\Theta}^t})\bigg(\int_t^T \partial_x\psi(s,Z_s^{t,\overline{\Theta}^t})\Sigma K(s-t)dB_s\bigg) \\ &\qquad\qquad \times\bigg(\int_t^T \partial_x\psi(s,Z_s^{t,\overline{\Theta}^t})\Delta K(s-t)dB_s\bigg) \;\bigg|\; \mathcal{F}_t\bigg]\bigg]dt \\ &= \int_0^T\int_t^T \mathbb{E}\bigg[\sigma^2(\overline{Z}_t)\mathbb{E}\bigg[\mathbf{D}_s\Big\{\phi''(X_T^{t,\overline{X}_t,\overline{\Theta}^t})\int_t^T \partial_x\psi(r,Z_r^{t,\overline{\Theta}^t})\Sigma K(r-t)dB_r\Big\} \\ &\qquad\qquad \times\partial_x\psi(s,Z_s^{t,\overline{\Theta}^t}) \;\bigg|\; \mathcal{F}_t\bigg]\bigg]\Delta K(s-t)dsdt. \end{aligned} \tag{3.125}$$

By product rule of Malliavin calculus, we have

$$\begin{aligned} &\mathbf{D}_s\Big\{\phi''(X_T^{t,\overline{X}_t,\overline{\Theta}^t})\int_t^T \partial_x\psi(r,Z_r^{t,\overline{\Theta}^t})\Sigma K(r-t)dB_r\Big\} \\ &= \big(\phi'''(X_T^{t,\overline{X}_t,\overline{\Theta}^t})\mathbf{D}_sX_T^{t,\overline{X}_t,\overline{\Theta}^t}\big)\int_t^T \partial_x\psi(r,Z_r^{t,\overline{\Theta}^t})\Sigma K(r-t)dB_r \\ &\quad+\phi''(X_T^{t,\overline{X}_t,\overline{\Theta}^t})\bigg[\partial_x\psi(s,Z_s^{t,\overline{\Theta}^t})\Sigma K(s-t)+\int_s^T\partial^2_{xx}\psi(r,Z_r^{t,\overline{\Theta}^t})\mathbf{D}_sZ_r^{t,\overline{\Theta}^t}\Sigma K(r-t)dB_r\bigg]. \end{aligned} \tag{3.126}$$

Insert (3.126) into F', we have decomposition

$$\begin{aligned} F' &= \int_0^T\int_t^T \mathbb{E}\bigg[\sigma^2(\overline{Z}_t)\mathbb{E}\bigg[\phi'''(X_T^{t,\overline{X}_t,\overline{\Theta}^t})\mathbf{D}_sX_T^{t,\overline{X}_t,\overline{\Theta}^t}\partial_x\psi(s,Z_s^{t,\overline{\Theta}^t}) \\ &\qquad\qquad \times\int_t^T \partial_x\psi(r,Z_r^{t,\overline{\Theta}^t})\Sigma K(r-t)dB_r \;\bigg|\; \mathcal{F}_t\bigg]\bigg]\Delta K(s-t)dsdt \\ &\quad+\int_0^T\int_t^T \mathbb{E}\bigg[\sigma^2(\overline{Z}_t)\mathbb{E}\bigg[\phi''(X_T^{t,\overline{X}_t,\overline{\Theta}^t})(\partial_x\psi)^2(s,Z_s^{t,\overline{\Theta}^t}) \;\bigg|\; \mathcal{F}_t\bigg]\bigg]\Delta(K^2)(s-t)ds \\ &\quad+\int_0^T\int_t^T \mathbb{E}\bigg[\sigma^2(\overline{Z}_t)\mathbb{E}\bigg[\phi''(X_T^{t,\overline{X}_t,\overline{\Theta}^t})\partial_x\psi(s,Z_s^{t,\overline{\Theta}^t}) \\ &\qquad\qquad \times\int_s^T \partial^2_{xx}\psi(r,Z_r^{t,\overline{\Theta}^t})\mathbf{D}_sZ_r^{t,\overline{\Theta}^t}\Sigma K(r-t)dB_r \;\bigg|\; \mathcal{F}_t\bigg]\bigg]\Delta K(s-t)dsdt. \end{aligned} \tag{3.127}$$

For the first term of (3.127), to bound the expectation inside, we apply the Hölder's inequality with $1/p+1/q+1/r+1/u+1/r(H)=1$, where $r\in(1,2)$ and $r(H)=2+\epsilon_H$ is chosen as in Corollary 3.2.60, s.t. $K^{r(H)}\in L^1$, we have

$$
\begin{aligned}
&\sup_{T\geq s\geq t\geq 0}\mathbb{E}\Bigg[\sigma^2(\overline{Z}_t)\mathbb{E}\Bigg[\phi'''(X_T^{t,\overline{X}_t,\overline{\Theta}^t})\mathbf{D}_s X_T^{t,\overline{X}_t,\overline{\Theta}^t}\partial_x\psi(s,Z_s^{t,\overline{\Theta}^t}) \\
&\qquad\qquad\times\int_t^T\partial_x\psi(r,Z_r^{t,\overline{\Theta}^t})\Sigma K(r-t)dB_r\;\Big|\;\mathcal{F}_t\Bigg]\Bigg] \\
&\leq \sup_{t\in[0,T]}\|\sigma^2(\overline{Z}_t)\|_{L^p(\Omega)}\sup_{T\geq s\geq t\geq 0}\|\phi'''(X_T^{t,\overline{X}_t,\overline{\Theta}^t})\|_{L^q(\Omega)}\sup_{T\geq s\geq t\geq 0}\|\mathbf{D}_s X_T^{t,\overline{X}_t,\overline{\Theta}^t}\|_{L^r(\Omega)} \\
&\quad\times\sup_{T\geq s\geq t\geq 0}\|\partial_x\psi(s,Z_s^{t,\overline{\Theta}^t})\|_{L^u(\Omega)}\sup_{t\in[0,T]}\|\int_t^T\partial_x\psi(r,Z_r^{t,\overline{\Theta}^t})\Sigma K(r-t)dB_r\|_{L^{r(H)}(\Omega)} \\
&\leq C_{T,\psi,\phi}\sup_{t\in[0,T]}\mathbb{E}\big[\big(\int_t^T(\partial_x\psi)^2(r,Z_r^{t,\overline{\Theta}^t})(\Sigma K)^2(r-t)dr\big)^{r(H)/2}\big]^{1/r(H)} \\
&\leq C_{T,\psi,\phi}\sup_{T\geq r\geq t\geq 0}\|(\partial_x\psi)^2(r,Z_r^{t,\overline{\Theta}^t})\|^{1/2}_{L^{r(H)/2}(\Omega)}\sup_{t\in[0,T]}\Big(\int_t^T(\Sigma K)^{r(H)}(r-t)dr\Big)^{1/2} \\
&\leq C_{T,H,\phi,\psi,K,\overline{K}}<\infty,
\end{aligned}
\tag{3.128}
$$

where we also use the Minkovski's inequality on the third inequality. So it yields an error of order $\int_0^T\|K-\overline{K}\|_{L^1([0,t])}dt$.

And as before, the expectation inside the second term of (3.127) can be bounded using Lemma 3.2.53, which yields an error of order $\int_0^T\|K^2-\overline{K}^2\|_{L^1([0,t])}dt$.

For the third term of (3.127), we obtain a bound for the expectation inside the double integral. This is done by applying Hölder's inequality with $q\in(1,2)$, $1/p+1/q=1$ and $1/q=1/p'+1/q'+1/2$, and use Lemma 3.2.53 and using Lemma 3.2.53 together with the BDG inequality. Indeed,

$$
\begin{aligned}
&\mathbb{E}\Big[\sigma^2(\overline{Z}_t)\mathbb{E}\Big[\phi''(X_T^{t,\overline{X}_t,\overline{\Theta}^t})\partial_x\psi(s,Z_s^{t,\overline{\Theta}^t})\int_s^T \partial_{xx}^2\psi(r,Z_r^{t,\overline{\Theta}^t})\mathbf{D}_sZ_r^{t,\overline{\Theta}^t}\Sigma K(r-t)dB_r \;\Big|\; \mathcal{F}_t\Big]\Big]\\
&\leq \sup_{0\leq t\leq T}\|\sigma^2(\overline{Z}_t)\|_{L^p(\Omega)}\\
&\quad\times\|\phi''(X_T^{t,\overline{X}_t,\overline{\Theta}^t})\partial_x\psi(s,Z_s^{t,\overline{\Theta}^t})\int_s^T \partial_{xx}^2\psi(r,Z_r^{t,\overline{\Theta}^t})\mathbf{D}_sZ_r^{t,\overline{\Theta}^t}\Sigma K(r-t)dB_r\|_{L^q(\Omega)}\\
&\leq \sup_{0\leq t\leq T}\|\sigma^2(\overline{Z}_t)\|_{L^p(\Omega)}\sup_{0\leq t\leq T}\|\phi''(X_T^{t,\overline{X}_t,\overline{\Theta}^t})\|_{L^{p'}(\Omega)}\sup_{0\leq t\leq s\leq T}\|\partial_x\psi(s,Z_s^{t,\overline{\Theta}^t})\|_{L^{q'}(\Omega)}\\
&\quad\times\|\int_s^T \partial_{xx}^2\psi(r,Z_r^{t,\overline{\Theta}^t})\mathbf{D}_sZ_r^{t,\overline{\Theta}^t}\Sigma K(r-t)dB_r\|_{L^2(\Omega)}\\
&\leq C_{T,\phi,\psi}\Big(\int_s^T\mathbb{E}\Big[(\partial_{xx}^2\psi)^2(r,Z_r^{t,\overline{\Theta}^t})|\mathbf{D}_sZ_r^{t,\overline{\Theta}^t}|^2\Big](1+c_{K,\overline{K}}^2)K^2(r-t)dr\Big)^{1/2}\\
&\leq C_{T,H,\phi,\psi,K,\overline{K}}K(s-t)\Big(\int_s^T\mathbb{E}\Big[(\partial_{xx}^2\psi)^2(r,Z_r^{t,\overline{\Theta}^t})|\mathbf{D}_sZ_r^{t,\overline{\Theta}^t}|^2\Big]dr\Big)^{1/2}\\
&\leq C_{T,H,\phi,\psi,K,\overline{K}}\Sigma K(s-t),
\end{aligned}
\tag{3.129}
$$

where we use $0<\overline{K}(x)\leq c_{K,\overline{K}}K(x)$ on the third inequality and K is non-increasing from Assumption 3.2.49 (iv) on the forth inequality, and $C_{T,H,\phi,\psi,K,\overline{K}}$ doesn't depend on s and t because of the uniform bound

$$
\sup_{0\leq t\leq s\leq T}\int_s^T\mathbb{E}\Big[(\partial_{xx}^2\psi)^2(r,Z_r^{t,\overline{\Theta}^t})|\mathbf{D}_sZ_r^{t,\overline{\Theta}^t}|^2\Big]dr\leq C_{T,H,\psi}<\infty,
$$

as we proved before by using Corollary 3.2.60 and Lemma 3.2.53.

Therefore, in total, we have

$$
\begin{aligned}
|F'| &\lesssim \int_0^T\|K-\overline{K}\|_{L^1([0,t])}dt+\int_0^T\|K^2-\overline{K}^2\|_{L^1([0,t])}dt\\
&\quad+\int_0^T\int_t^T\Sigma K(s-t)\times\Delta K(s-t)dsdt \qquad (3.130)\\
&=\int_0^T\|K-\overline{K}\|_{L^1([0,t])}dt+2\int_0^T\|K^2-\overline{K}^2\|_{L^1([0,t])}dt. \qquad (3.131)
\end{aligned}
$$

Estimate of H_1 and H_2

Recall that

$$
H_1=\int_0^T\int_t^T\mathbb{E}\Big[\sigma^2(\overline{Z}_t)\mathbb{E}\big[\phi'(X_T^{t,\overline{X}_t,\overline{\Theta}^t})\partial_{xx}^2(\psi^2)(s,Z_s^{t,\overline{\Theta}^t})\;\big|\;\mathcal{F}_t\big]\Big]\Delta(K^2)(s-t)dsdt
$$

and

$$
H_2=\int_0^T\int_t^T\mathbb{E}\Big[\sigma^2(\overline{Z}_t)\mathbb{E}\big[\phi''(X_T^{t,\overline{X}_t,\overline{\Theta}^t})\partial_{xx}^2\psi(s,Z_s^{t,\overline{\Theta}^t})\mathbf{D}_sX_T^{t,\overline{X}_t,\overline{\Theta}^t}\;\big|\;\mathcal{F}_t\big]\Big]\Delta(K^2)(s-t)dsdt.
$$

Note that H_1 have same structure as F_1, and can therefore be bounded using the same strategy, which yields an error of order $\int_0^T \|K^2 - \overline{K}^2\|_{L^1([0,t])}dt$. As for H_2, it can be bounded in the same way as $F_5^{(1)}$ by using Hölder's inequality repeatedly, since $\sigma^2(\overline{Z}_t)$, $\phi''(X_T^{t,\overline{X}_t,\overline{\Theta}^t})$, $\partial^2_{xx}\psi(s, Z_s^{t,\overline{\Theta}^t})$ are uniformly bounded in $L^p(\Omega)$ for any $p > 1$ by Lemma 3.2.53, and $\mathbf{D}_s X_T^{t,\overline{X}_t,\overline{\Theta}^t}$ is uniformly bounded in $L^p(\Omega)$ for $p \in (1, 2)$ by Corollary 3.2.65. Eventually, it provides an error bound of order $\int_0^T \|K^2 - \overline{K}^2\|_{L^1([0,t])}dt$ as well.

Error Bound for $\mathcal{E}_T$

In summary, we have gathered the estimates for all the terms in the decomposition (3.107) of $\mathcal{E}_T$. We can now incorporate these estimates to conclude that

$$|\mathcal{E}_T| \leq C \int_0^T \left\{\|K - \overline{K}\|_{L^1([0,t])} + \|K^2 - \overline{K}^2\|_{L^1([0,t])}\right\}dt,$$

which ends the proof for the main Theorem 3.2.51.

3.2.7. Application on the Quadratic Rough Heston Model

Recall the dynamics for the log stock price X and the volatility process Z in the quadratic rough Heston model are given by

$$\begin{aligned} dX_t &= -\frac{1}{2}p(Z_t)\,dt + \sqrt{p(Z_t)}\,dW_t, \\ Z_t &= \int_0^t K(t-s)b(Z_s)\,ds + \int_0^t K(t-s)\eta\sqrt{p(Z_s)}\,dW_s, \end{aligned} \tag{3.132}$$

where $b(x) = \theta - x$ and $p(x) = a(x-b)^2 + c$, $a, b, c > 0$.

We now consider the case where $\zeta = -1/2$, $\psi(x) = \sqrt{p(x)}$, $b(x) = \theta - x$, $\sigma(x) = \eta\sqrt{p(x)}$, which clearly satisfy Assumptions 3.2.49(i) and (iii). The fractional kernel K also satisfies the Assumption 3.2.49 (iv). Let K^N be an approximation for K obtained from quadrature method, i.e.

$$K^N(t) := \sum_{i=1}^N w_i e^{-x_i t}, \tag{3.133}$$

as introduced in Chapter 2, and let X_N and Z_N be the approximations of X and Z with dynamics

$$dX_t^N = -\frac{1}{2}p(Z_t^N)dt + \sqrt{p(Z_t^N)}dW_t, \tag{3.134}$$

$$Z_t^N = \int_0^t K^N(t-s)b(Z_s^N)ds + \int_0^t K^N(t-s)\eta\sqrt{p(Z_t^N)}dW_s, \tag{3.135}$$

where $b(x) = \theta - x$ and $p(x) = a(x-b)^2 + c$, $a, b, c > 0$. As discussed in Remark 3.2.52 (ii), we have the following theorem.

Theorem 3.2.70 (Error bound for the quadratic rough Heston model)**.** *If $\exists C_K > 0$, s.t., $K^N(t) \leq C_K K(t)$ for any $t \in (0, T]$. Then there exists a constant $C = C_{T,H} > 0$, for payoff function $\phi \in \mathcal{C}^3_{\text{poly}}(\mathbb{R})$, we have*

$$|\mathbb{E}[\phi(X_T)] - \mathbb{E}[\phi(X_T^N)]| \leq C \int_0^T \left\{ \|K - K^N\|_{L^1([0,t])} + \|K^2 - (K^N)^2\|_{L^1([0,t])} \right\} dt.$$

3.2.8. Weak Simulation for the Quadratic Rough Heston Model

Since we have already shown that the Markovian approximation provides a good approximation with L^1-error of $K - K^N$ and $K^2 - (K^N)^2$, we next seek an efficient scheme to simulate the quadratic rough Heston model.

Note that there are many similarities between the rough Heston model and the quadratic rough Heston model. We can thereby use the same strategy to construct a weak scheme for simulating the quadratic rough Heston model, namely by applying the Strang splitting. While we remark here the moment-matching technique is not easy to apply, we will show it in the following.

First, we're going to simulate V_t as in the Chapter 2. Replacing the kernel K by its approximation $K^N = . \sum_{i=1}^N w_i e^{-x_i t}$ in equation (3.3) yields the SDE corresponding to the Markovian approximation

$$dZ_t^{(i)} = -x_i(Z_t^{(i)} - z_0^i)dt + (\theta - Z_t^N)dt + \eta\sqrt{a(Z_t^N - b)^2 + c}dW_t,\ Z_0^{(i)} = z_0^i,\ i = 1, \dots, N,$$

follows from [AJEE19a, Theorem 3.1], where $\boldsymbol{w} \coloneqq (w_i)_{i=1}^N$ is the vector of weights, $\boldsymbol{z}_0 = (\boldsymbol{z}_0^i)_{i=1}^N$ is any initial condition with $\boldsymbol{w} \cdot \boldsymbol{z}_0 = Z_0$, $\boldsymbol{Z}^N \coloneqq (Z^{(i)})_{i=1}^N$, and $Z_t^N = \boldsymbol{w} \cdot \boldsymbol{Z}_t^N$.

Rewrite it into the vector form

$$d\boldsymbol{Z}_t^N = -\text{diag}(\boldsymbol{x}) \left(\boldsymbol{Z}_t^N - \boldsymbol{z}_0\right) dt + \left(\theta - \boldsymbol{w}^\top \boldsymbol{Z}_t^N\right) \mathbf{1}\ dt + \left(\eta\sqrt{a(\boldsymbol{w}^\top \boldsymbol{Z}_t^N - b)^2 + c}\right) \mathbf{1}\ dW_t. \tag{3.136}$$

Then we still split SDE (3.136) into two parts to avoid the effect of mean-reversion from the deterministic part.

The deterministic part is identical to that in the rough Heston model, and is given by

$$d\boldsymbol{Z}_t^N = -\text{diag}(\boldsymbol{x}) \left(\boldsymbol{Z}_t^N - \boldsymbol{v}_0\right) dt + \left(\theta - \boldsymbol{w}^\top \boldsymbol{Z}_t^N\right) \mathbf{1} dt,$$

The exact solution at time h is

$$D(\boldsymbol{z}, h) = \boldsymbol{Z}_h = e^{Ah}\boldsymbol{z} + A^{-1}(e^{Ah} - \text{Id})b,$$

where

$$A \coloneqq -\mathbf{1}\boldsymbol{w}^T - \text{diag}(\boldsymbol{x}), \quad \text{and} \quad b \coloneqq \theta\mathbf{1} + \text{diag}(\boldsymbol{x})\boldsymbol{v}_0,$$

denote $\mathbf{1} := (1,\dots,1)^T \in \mathbb{R}^N$, $\mathrm{Id} \in \mathbb{R}^{N\times N}$ is the identity matrix, and $\mathrm{diag}(\boldsymbol{x}) \in \mathbb{R}^{N\times N}$ is the diagonal matrix with entries $\boldsymbol{x}$. Therefore, we can simply set the scheme $\widehat{D}(\boldsymbol{z},h) := D(\boldsymbol{z},h)$.

The stochastic part is given by

$$d\mathbf{Y}_t^N = \left(\eta\sqrt{a(\boldsymbol{w}^\top \mathbf{Y}_t^N - b)^2 + c}\right)\mathbf{1}\; dW_t. \tag{3.137}$$

We can still consider the the inner product of the SDE (3.137) with $\boldsymbol{w}$ to get the SDE

$$dY_t = \eta\overline{w}\sqrt{a(Y_t-b)^2+c}dW_t, \qquad Y_0 = \boldsymbol{w}\cdot\boldsymbol{z}. \tag{3.138}$$

Since it no longer coincides with the CIR equation in [LM20], unlike in the rough Heston model, we need to apply the moment-matching technique to obtain an approximation. This requires choosing suitable parameters again.

Next we try to construct a scheme $\widehat{S}$ for approximating Y, by following the moment-matching technique in [LM20]. Denote $\alpha := a\overline{w}^2\eta^2$ and $\beta := c\overline{w}^2\eta^2$, we first write down the diffusion generator for SDE (3.138)

$$Lf(x) = \frac{1}{2}\left(\alpha(x-b)^2+\beta\right)f''(s),$$

$$L^2 f(x) = \frac{1}{4}\left(\alpha(x-b)^2+\beta\right)\left[2\alpha f''(x) + 4\alpha(x-b)f'''(x) + \left(\alpha(x-b)^2+\beta\right)f''''(s)\right].$$

Assume $\widehat{Y}$ is our required second order approximation starting at $y = \boldsymbol{w}\cdot\boldsymbol{z}$, then it should satisfy: for $f \in C^\infty_{\mathrm{pol}}$

$$R_3 f(y) = \mathbb{E}(f(\widehat{Y}_h)) - [f(y) + hLf(y) + \frac{h^2}{2}L^2 f(y)] = O(h^3).$$

By Taylor expansion of $f(\widehat{Y}_h)$ at y, we have

$$\begin{aligned} R_3 f(y) =& f'(y)\mathbb{E}(\widehat{Y}_h - y) + \frac{f''(y)}{2}[\mathbb{E}[(\widehat{Y}_h-y)^2] - (h + \frac{\alpha h^2}{2})\left(\alpha(y-b)^2+\beta\right)] \\ &+ \frac{f'''(y)}{6}[\mathbb{E}[(\widehat{Y}_h-y)^3] - 3\alpha h^2(y-b)\left(\alpha(y-b)^2+\beta\right)] \\ &+ \frac{f''''(y)}{4!}[\mathbb{E}[(\widehat{Y}_h-y)^4] - 3h^2\left(\alpha(y-b)^2+\beta\right)^2] \\ &+ \frac{f^{(5)}(y)}{5!}\mathbb{E}[(\widehat{Y}_h-y)^5] + r(y,h), \end{aligned} \tag{3.139}$$

where $|r(y,h)| = \frac{1}{5!}|\mathbb{E}[\int_y^{\widehat{Y}_h} f^{(6)}(s)(\widehat{Y}_h - s)^5 ds]| \le \frac{1}{6!}\mathbb{E}[\max_{0\le s\le \widehat{Y}_h}|f^{(6)}(s)|(\widehat{Y}_h - s)^6]$.

So $\widehat{Y}$ should satisfy

$$\begin{aligned}
&\mathbb{E}(\widehat{Y}_h - y) = O(h^3),\\
&\mathbb{E}[(\widehat{Y}_h - y)^2] = (h + \frac{\alpha h^2}{2})\left(\alpha(y-b)^2 + \beta\right) + O(h^3),\\
&\mathbb{E}[(\widehat{Y}_h - y)^3] = 3\alpha h^2 (y-b)\left(\alpha(y-b)^2 + \beta\right) + O(h^3),\\
&\mathbb{E}[(\widehat{Y}_h - y)^4] = 3h^2\left(\alpha(y-b)^2 + \beta\right)^2 + O(h^3),\\
&\mathbb{E}[(\widehat{Y}_h - y)^5] = O(h^3),\\
&\mathbb{E}[\max_{0\le s\le \widehat{Y}_h} |f^{(6)}(s)|(\widehat{Y}_h - s)^6] = O(h^3).
\end{aligned} \tag{3.140}$$

We can equivalently replace the condition $\mathbb{E}[\max_{0\le s\le \widehat{Y}_h} |f^{(6)}(s)|(\widehat{Y}_h - s)^6] = O(h^3)$ with

$$\mathbb{E}[(\widehat{Y}_h - s)^6] = O(h^3).$$

Since if we choose a three- or four-valued random variable $\widehat{Y}$ to approximate Y, then $\widehat{Y}$ will be bounded.

Let $\gamma := \alpha(y-b)^2 + \beta$, we can then express all these conditions on non-central moments in terms of conditions on central moments.

$$\begin{aligned}
&\mathbb{E}[\widehat{Y}_h] = y + O(h^3),\\
&\mathbb{E}[\widehat{Y}_h^2] = y^2 + \Big(h + \tfrac{\alpha}{2}h^2\Big)\gamma + O(h^3),\\
&\mathbb{E}[\widehat{Y}_h^3] = y^3 + 3y\Big(h + \tfrac{\alpha}{2}h^2\Big)\gamma + 3\alpha h^2(y-b)\gamma + O(h^3),\\
&\mathbb{E}[\widehat{Y}_h^4] = y^4 + 6y^2\Big(h + \tfrac{\alpha}{2}h^2\Big)\gamma + 12\alpha h^2 y(y-b)\gamma + 3h^2\gamma^2 + O(h^3),\\
&\mathbb{E}[\widehat{Y}_h^5] = y^5 + 10y^3\Big(h + \tfrac{\alpha}{2}h^2\Big)\gamma + 30\alpha h^2 y^2(y-b)\gamma + 15h^2 y\gamma^2 + O(h^3),\\
&\mathbb{E}[\widehat{Y}_h^6] = y^6 + 15y^4\Big(h + \tfrac{\alpha}{2}h^2\Big)\gamma + 60\alpha h^2 y^3(y-b)\gamma + 45h^2 y^2\gamma^2 + O(h^3).
\end{aligned} \tag{3.141}$$

Next we denote

$$\begin{aligned}
&\widehat{m}_1 = y,\\
&\widehat{m}_2 = y^2 + \Big(h + \tfrac{\alpha}{2}h^2\Big)\gamma,\\
&\widehat{m}_3 = y^3 + 3y\Big(h + \tfrac{\alpha}{2}h^2\Big)\gamma + 3\alpha h^2(y-b)\gamma,\\
&\widehat{m}_4 = y^4 + 6y^2\Big(h + \tfrac{\alpha}{2}h^2\Big)\gamma + 12\alpha h^2 y(y-b)\gamma + 3h^2\gamma^2,\\
&\widehat{m}_5 = y^5 + 10y^3\Big(h + \tfrac{\alpha}{2}h^2\Big)\gamma + 30\alpha h^2 y^2(y-b)\gamma + 15h^2 y\gamma^2,\\
&\widehat{m}_6 = y^6 + 15y^4\Big(h + \tfrac{\alpha}{2}h^2\Big)\gamma + 60\alpha h^2 y^3(y-b)\gamma + 45h^2 y^2\gamma^2.
\end{aligned} \tag{3.142}$$

Therefore, it suffices to look for a three-valued random variable $\widehat{Y}_h$ taking three values x_1, x_2 and x_3 with probabilities p_1, p_2 and p_3 such that

$$\mathbb{E}[(\widehat{Y}_h)^i] = \widehat{m}_i + O(h^3), \quad i = 1, 2..., 6.$$

We remark that the values x_i need not be non-negative, since Y_t is allowed to take negative values. This is unlike the rough Heston model, where the volatility must remain non-negative, and therefore simplifies the implementation compared with the rough Heston model. Therefore, we're going to solve the system

$$x_1^i p_1 + x_2^i p_2 + x_3^i p_3 = \widehat{m}_i, \; i = 1, 2, 3,$$

with respect to unknowns x_1, x_2, x_3. In [LM20], solutions for p_i were obtained, expressed in terms of x_i

$$\begin{aligned}
p_1 &= \frac{\widehat{m}_1 x_2 x_3 - \widehat{m}_2(x_2 + x_3) + \widehat{m}_3}{x_1(x_3 - x_1)(x_2 - x_1)}, \\
p_2 &= \frac{\widehat{m}_2(x_1 + x_3) - \widehat{m}_1 x_1 x_3 - \widehat{m}_3}{x_2(x_2 - x_1)(x_3 - x_2)}, \\
p_3 &= \frac{\widehat{m}_1 x_1 x_2 - \widehat{m}_2(x_1 + x_2) + \widehat{m}_3}{x_3(x_3 - x_2)(x_3 - x_1)}.
\end{aligned}$$

In our context, we aim to express $x_i's$ in terms of $p_i's$, since x_i can take arbitrary values, whereas p_i must remain non-negative. This, however, is a very difficult task. Hence, we choose the values of x_i to ensure that p_i remain non-negative and satisfy

$$\mathbb{E}[(\widehat{Y}_h)^i] = \widehat{m}_i + O(h^3), \quad i = 4, ...6$$

as well. In [LM20], the values $x_i's$ are chosen in the form

$$x_{1,3} = x + \left(A + \frac{3}{4}\right) z \mp \sqrt{\left(3x + \left(A + \frac{3}{4}\right)^2\right) z}, \; x_2 = x + Az,$$

where $A \in [3/4, 3/2]$. We remark that the generic forms of the solutions x_i proposed by the authors in [LM20] are quite reasonable and are given by

$$x_{1,3} = x + A_1 z \mp \sqrt{(Bx + Cz)\, z}, \; x_2 = x + A_2 z. \tag{3.143}$$

Recall the moments $\widehat{m}_i$ in [LM20] are given by

$$\begin{aligned}
\widehat{m}_1 &= x, \\
\widehat{m}_2 &= x^2 + xz, \\
\widehat{m}_3 &= x^3 + 3x^2 z + \frac{3}{2} x z^2, \\
\widehat{m}_4 &= x^4 + 6x^3 z + 9x^2 z^2, \\
\widehat{m}_5 &= x^5 + 10x^4 z + 30x^3 z^2, \\
\widehat{m}_6 &= x^6 + 15x^5 z + 75x^4 z^2,
\end{aligned} \tag{3.144}$$

where we ignore the residual terms $O(h^3)$. And for each k, we observe $\widehat{m}_k$ is a homogeneous polynomial $p(x,z)$ of degree k. The solutions (3.143) can thus be interpreted as consisting of a first-order contribution, so that all the p_i contribute zero order. We thus denote

$$A^2 := (h + \frac{\alpha}{2}h^2)\gamma, \quad B^3 := 3\alpha h^2(y-b)\gamma,$$

then the equations (3.142) can be reduced to

$$\begin{aligned}
\widehat{m}_1 &= y, \\
\widehat{m}_2 &= y^2 + A^2, \\
\widehat{m}_3 &= y^3 + 3yA^2 + B^3 \\
\widehat{m}_4 &= y^4 + 6y^2A^2 + 4B^3 + 3h^2\gamma^2, \\
\widehat{m}_5 &= y^5 + 10y^3A^2 + 10B^3 + 15h^2y\gamma^2, \\
\widehat{m}_6 &= y^6 + 15y^4A^2 + 20B^3 + 45h^2y^2\gamma^2.
\end{aligned} \tag{3.145}$$

In our case, although we only need to guarantee the non-negativity of p_i, it remains a technically challenging and lengthy task. Even with the help of MAPLE, the introduction of three additional parameters a, b and c and the quadratic transformation on the volatility term implies that, for each k, $\widehat{m}_k$ is a homogeneous polynomial $p(A,B,y,\gamma,h)$ of degree k. Hence, the algebraic computations are computationally intensive, even though the procedure is theoretically feasible.

Similarly, observing that y and B appear only in the form of power terms in the equations, we can propose that the solutions x_i have forms in the following

$$x_{1,3} = y + k_1A + k_2B + k_3\gamma + k_4h \mp \sqrt{(k_5y + k_6A + k_7\gamma + k_8h)(y + k_9B)},$$

$$x_2 = y + k_1'A + k_2'B + k_3'\gamma + k_4'h$$

with undetermined k_i and k_i'. Although we could consider a four-valued random variable to gain more freedom, it remains technically challenging to solve, even with the help of MAPLE.

Since our goal is to derive a second-order approximation for Y, we can consider either a stochastic Runge–Kutta scheme or a stochastic weak Taylor scheme. We first perform moment matching on the increments of the Brownian motion ΔW. Define

$$\xi = \begin{cases} +\sqrt{3}, & p = 1/6, \\ 0, & p = 2/3, \\ -\sqrt{3}, & p = 1/6, \end{cases}$$

which is a three-valued random variable satisfying

$$\Delta W_h \approx \sqrt{h}\xi$$

in the sense of

$$\mathbb{E}[(\Delta W_h)^k] = \mathbb{E}[(\sqrt{h}\xi)^k] \qquad \text{for} \quad k = 1, 2, 3, 4.$$

Therefore, since they both have identical moments up to degree four, this results in a second-order weak scheme for Y when replacing ΔW is replaced in the second-order weak Taylor scheme for Y, see [KP92, Section 14.2], which provides a second order weak Taylor scheme for SDEs. In our setting, the corresponding scheme for the SDE (3.137) is given by

$$\begin{aligned} \widehat{Y}^i_{t_{n+1}} =& \widehat{Y}^i_{t_n} + \eta\sqrt{p(\widehat{\mathbf{Y}}^N_{t_n})}\ \Delta W + \frac{a\overline{w}\eta^2}{2}(w^\top \cdot \widehat{\mathbf{Y}}^N_{t_n} - b)[(\Delta W)^2 - \Delta t] \\ &+ \eta^3(\frac{a\overline{w}^2}{2}\sqrt{p(\widehat{\mathbf{Y}}^N_{t_n})} - \frac{(a\overline{w})^2}{2}\frac{(w^\top \cdot \widehat{\mathbf{Y}}^N_{t_n} - b)^2}{\sqrt{p(\widehat{\mathbf{Y}}^N_{t_n})}})[\Delta W \Delta t - \Delta Z], \end{aligned} \tag{3.146}$$

for $i = 1, 2, ..., N$, where $\Delta W = W_{t_{n+1}} - W_{t_n}$, $\Delta t = t_{n+1} - t_n$ and ΔZ is the double Ito integral

$$I_{1,0} = \int_{t_n}^{t_{n+1}} \int_{t_n}^{s} dW_r\, ds.$$

Following the argument in [KP92, Section 14.2] and applying moment-matching techniques, we replace ΔW with $\sqrt{\Delta t}\ \xi$ and ΔZ with $\frac{(\sqrt{\Delta t})^3}{2}\ \xi$. We then rewrite the scheme into

$$\begin{aligned} \widehat{Y}^i_{t_{n+1}} =& \widehat{Y}^i_{t_n} + \eta\sqrt{p(\widehat{\mathbf{Y}}^N_{t_n})}\ \sqrt{\Delta t}\ \xi + \frac{a\overline{w}\eta^2}{2}(w^\top \cdot \widehat{\mathbf{Y}}^N_{t_n} - b)[\Delta t\ \xi^2 - \Delta t] \\ &+ \eta^3(\frac{a\overline{w}^2}{4}\sqrt{p(\widehat{\mathbf{Y}}^N_{t_n})} - \frac{(a\overline{w})^2}{4}\frac{(w^\top \cdot \widehat{\mathbf{Y}}^N_{t_n} - b)^2}{\sqrt{p(\widehat{\mathbf{Y}}^N_{t_n})}})(\sqrt{\Delta t})^3\ \xi, \end{aligned} \tag{3.147}$$

for $i = 1, 2, ..., N$. Since the diffusion coefficient $\sqrt{p(\boldsymbol{x})}$ in our model is smooth and Lipschitz, [KP92, Theorem 14.2.4] guarantees that this scheme achieves second-order weak convergence. We denote this scheme by $\widehat{S}(\boldsymbol{y}, \Delta t)$, i.e.,

$$\widehat{\mathbf{Y}}^N_{t_{n+1}} = \widehat{S}(\widehat{\mathbf{Y}}^N_{t_n}, \Delta t),$$

for $i = 0, 1, ..., N-1$. Finally, combining the schemes for drift and diffusion via Strang splitting yields

$$\widehat{A}'(\boldsymbol{z}, h) := \widehat{D}\left(\widehat{S}\left(\widehat{D}\left(\boldsymbol{z}, \frac{h}{2}\right), h\right), \frac{h}{2}\right), \tag{3.148}$$

for approximating $\boldsymbol{Z}_h$ given $\boldsymbol{z}$.

We thus obtain a simulation algorithm for the quadratic rough Heston model, defined by

$$\boldsymbol{Z}^{N,M}_{t_{j+1}} := \widehat{A}'(\boldsymbol{Z}^{N,M}_{t_j}, t_{j+1} - t_j), \qquad j = 0, \ldots, M-1,$$

where $0 = t_0 < t_1 < \cdots < t_M = T$. Unlike in the rough Heston model, this scheme is always well-defined, as Z is not required to be non-negative in our setting.

Theorem 3.2.71. *The scheme* $\widehat{A}'(\boldsymbol{z},h) := \widehat{D}\left(\widehat{S}\left(\widehat{D}\left(\boldsymbol{z},\frac{h}{2}\right),h\right),\frac{h}{2}\right)$ *defined in* (3.148) *has second order weak rate.*

Proof. First, note that the scheme $\widehat{A}'$ achieves a potential second-order weak convergence rate by the argument above. The uniform boundedness follows from the fact that $\sqrt{p(\boldsymbol{x})}$ is a Lipschitz function with a positive lower bound. Hence, the scheme exhibits a strongly potential second-order weak rate.

Recall that the main difficulty in the rough Heston model lies in establishing polynomial estimates for the derivatives of the solution of the Cauchy problem, i.e., Theorem 2.2.10 (ii). This is difficult to prove for the rough Heston model; however, thanks to Remark 2.2.11 (i), the quadratic rough Heston model automatically satisfies this condition. Therefore, we can conclude that the scheme $\widehat{A}'$ indeed achieves second-order weak convergence. □

Having obtained the scheme $\widehat{A}'$ for simulating $\boldsymbol{Z}$, we now derive a scheme for simulating S^N. Note that the source of randomness for the stock price S^N comes solely from the Brownian motion W, which is the same as for Z. Hence, for simulating S^N, we only need to treat the same challenging part as in Section 2.2.4, which means that we can follow the same strategy as in Section 2.2.4.

We remark that one could also introduce another independent Brownian motion B and define the dynamics of the stock price S as

$$dS_t = S_t\sqrt{p(Z_t)}\left(\rho dW_s + \sqrt{1-\rho^2}dB_s\right),$$

as in the rough Heston model. The proof of Theorem 3.2.70 remains valid under this modification.

As in the Section 2.2.4, we first define the processes

$$Y_t^{(1,i)} = \int_0^t Z_t^{(i)}dt, \qquad i = 1,\ldots,N, \tag{3.149}$$

$$Y_t^{(2)} = \int_0^t (Z_t)^2 dt, \tag{3.150}$$

where $Z_t = \boldsymbol{w}^\top \cdot \boldsymbol{Z}_t$. And note that the solution S to

$$dS_t^N = S_t^N\sqrt{V_t^N}dW_t$$

is given by

$$S_t = S_0 \exp\left(\int_0^t \sqrt{V_s^N}dW_s - \frac{1}{2}\int_0^t V_s^N ds\right),$$

where $V_s = p(Z_s)$. And we notice that the latter term in the exponent is the linear combination of $\widehat{\boldsymbol{Y}}^{(1)}$ and $\widehat{Y}^{(2)}$. So the weak scheme $\widehat{S}$ for the stock price S may depend on $\widehat{\boldsymbol{Z}}$, $\widehat{\boldsymbol{Y}}^{(1)}$ and $\widehat{Y}^{(2)}$.

The SDE $S((s, \boldsymbol{z}, \boldsymbol{y}^{(1)}, y^{(2)}), h)$ is then given by

$$\begin{aligned} dS_t &= \sqrt{V_t} S_t dW_t, \qquad V_t = p(Z_t), \qquad S_0 = s, \\ dY_t^{(1,i)} &= Z_t^{(i)} dt, \qquad Y_0^{(1,i)} = y^{(1,i)}, \\ dY_t^{(2)} &= Z_t^2 dt, \qquad Y_0^{(2)} = y^{(2)}, \\ dZ_t^{(i)} &= -x_i(Z_t^{(i)} - z_0^{(i)})dt + (\theta - Z_t)dt + \eta\sqrt{p(Z_t)}dW_t, \qquad Z_t^{(i)} = z^{(i)}. \end{aligned}$$

For approximating S_t, we first approximate $\boldsymbol{Z}$ by $\widehat{\boldsymbol{Z}}_h := \widehat{A'}(\boldsymbol{z}, h)$. Then, we approximate $\boldsymbol{Y}^{(1)}$ and $Y^{(2)}$ by using the trapezoidal rule, i.e.

$$\widehat{\boldsymbol{Y}}_h^{(1)} := \boldsymbol{y}^{(1)} + h(\boldsymbol{z} + \widehat{\boldsymbol{Z}}_h)/2$$

and

$$\widehat{Y}_h^{(2)} := y^{(2)} + h(z^2 + (w \cdot \widehat{\boldsymbol{Z}}_h)^2)/2.$$

Next, we aim to express S_h in terms of $\boldsymbol{Z}_h$, $\mathbf{Y}_h^{(1)}$ and $Y_h^{(2)}$, in the form

$$S_t = s \exp\left(\alpha t + \sum_{i=1}^N b_i(Y_t^{(1,i)} - Y_0^{(1,i)}) - b'(Y_t^{(2)} - Y_0^{(2)}) + \sum_{i=1}^N c_i(Z_t^{(i)} - Z_0^{(i)})\right). \tag{3.151}$$

Using Itô's formula, we obtain

$$\begin{aligned} dS_t &= S_t\left[\alpha dt + \sum_{i=1}^N b_i dY_t^{(1,i)} - b' dY_t^{(2)} + \sum_{i=1}^N c_i dZ_t^{(i)} + \frac{1}{2}\sum_{i,j=1}^N c_i c_j d[Z^{(i)}, Z^{(j)}]_t\right] \\ &= S_t\Bigg[\left(\alpha + \sum_{i=1}^N c_i x_i z_0^{(i)} + \theta\sum_{i=1}^N c_i + \frac{\eta^2}{2}(ab^2 + c)\sum_{i,j}^N c_i c_j\right) dt \\ &\quad + \sum_{i=1}^N \left(b_i - c_i x_i - w_i \sum_{j=1}^N c_j - ab\eta^2 w_i \sum_{j,k=1}^N c_j c_k\right) Z_t^{(i)} dt \\ &\quad + \left(\frac{a\eta^2}{2}\sum_{j,k=1}^N c_j c_k - b'\right) Z_t^2 dt + \eta \sum_{i=1}^N c_i \sqrt{V_t} dW_t\Bigg]. \end{aligned}$$

To preserve the SDE S, we thus require that

$$\begin{aligned} \alpha + \sum_{i=1}^N c_i x_i z_0^{(i)} + \theta\sum_{i=1}^N c_i + \frac{\eta^2}{2}(ab^2 + c)\sum_{i,j}^N c_i c_j &= 0, \\ b_i - c_i x_i - w_i\sum_{j=1}^N c_j - ab\eta^2 w_i \sum_{j,k=1}^N c_j c_k &= 0, \qquad i = 1, \dots, N, \\ \frac{a\eta^2}{2}\sum_{j,k=1}^N c_j c_k - b' &= 0, \\ \eta\sum_{i=1}^N c_i &= 1. \end{aligned}$$

Since a, $\eta > 0$, we can reduce these equations to

$$\begin{aligned} \alpha + \sum_{i=1}^N c_i x_i z_0^{(i)} + \theta \sum_{i=1}^N c_i + \frac{b'}{a}(ab^2 + c) = 0,& \\ b_i - c_i x_i - w_i \sum_{j=1}^N c_j = 2bb' w_i,& \qquad i = 1, \dots, N, \\ \sum_{j}^N c_j = \sqrt{\frac{2b'}{a\eta^2}},& \\ \sum_{i=1}^N c_i = 1/\eta.& \end{aligned}$$

It's achieved with

$$\begin{aligned} \alpha &= -\sum_{i=1}^N c_i x_i z_0^{(i)} - \frac{\theta}{\eta} - \frac{1}{2}(ab^2 + c), \\ b_i &= c_i x_i + \frac{w_i}{\eta} + abw_i, \qquad i = 1, \dots, N, \\ b' &= a/2, \\ \sum_{i=1}^N c_i &= \frac{1}{\eta}, \end{aligned}$$

Observe that the last equation admits multiple solutions. We aim to choose (c_i) so that the representation (3.151) is numerically as stable as possible. Following the approach in [BB23a], to mitigate the mean-reversion effects from $\boldsymbol{x}$ (noting that the two models share the same deterministic component inducing mean-reversion), we select a reasonable and potentially effective choice, given by

$$c_1 = \frac{1}{\eta}, \qquad c_i = 0, \qquad i = 2, \dots, N. \tag{3.152}$$

This further yields

$$\begin{aligned} \alpha &= -\frac{\left(x_1 z_0^{(1)} + \theta\right)}{\eta} - \frac{1}{2}(ab^2 + c), \\ b_i &= \frac{x_1}{\eta}\delta_1 + \frac{w_i}{\eta} + abw_i, \qquad i = 1, \dots, N. \end{aligned}$$

Therefore, we get the solution

$$\begin{aligned} S_t = s \exp\Bigg((-\frac{\left(x_1 z_0^{(1)} + \theta\right)}{\eta} - \frac{1}{2}(ab^2 + c))t + (ab + \frac{1}{\eta})(Y_t^{(1)} - Y_0^{(1)}) \\ + \frac{x_1}{\eta}(Y_t^{(1,1)} - Y_0^{(1,1)}) - \frac{a}{2}(Y_t^{(2)} - Y_0^{(2)}) + \frac{Z_t^{(1)} - Z_0^{(1)}}{\eta} \Bigg), \end{aligned}$$

where $Y_t^{(1)} \coloneqq \boldsymbol{w} \cdot \boldsymbol{Y}^{(1)}$.

Hence, we approximate the solution of the SDE S using the algorithm $\widehat{S}$, given by

$$\begin{aligned}
\widehat{\boldsymbol{Z}}_h &= \widehat{A}'(\boldsymbol{z}, h),\\
\widehat{\boldsymbol{Y}}_h^{(1)} &= \boldsymbol{y}^{(1)} + (\boldsymbol{z} + \widehat{\boldsymbol{Z}}_h)\frac{h}{2},\\
\widehat{Y}_h^{(2)} &= y^{(2)} + (z^2 + (w^\top \cdot \widehat{\boldsymbol{Z}}_h)^2)\frac{h}{2},\\
\widehat{S}_h &= s\exp\Bigg((-\frac{\left(x_1 z_0^{(1)} + \theta\right)}{\eta} - \frac{1}{2}(ab^2 + c))h + (ab + \frac{1}{\eta})(\widehat{Y}_h^{(1)} - Y_0^{(1)})\\
&\quad + \frac{x_1}{\eta}(\widehat{Y}_h^{(1,1)} - Y_0^{(1,1)}) - \frac{a}{2}(\widehat{Y}_h^{(2)} - Y_0^{(2)}) + \frac{\widehat{Z}_h^{(1)} - Z_0^{(1)}}{\eta}\Bigg).
\end{aligned}$$

Since $\widehat{A}'$ is second order scheme, as proved in Theorem 3.2.71, and $\widehat{\boldsymbol{Y}}^{(1)}$, $\widehat{Y}^{(2)}$ are computed using the trapezoidal rule, the scheme $\widehat{S}$ might be also of second order.

And we remark here even if the dynamics of S are extended to

$$dS_t = S_t\sqrt{p(Z_t)}\left(\rho dW_s + \sqrt{1-\rho^2}dB_s\right),$$

the same splitting scheme (e.g., randomized Leapfrog splitting as in Section 2.2.4) can be applied to obtain a second-order weak scheme.

Consequently, we obtain a second-order weak scheme for (S, V), denoted by $(\widehat{S}, \widehat{V})$.

4. Numerics

In this chapter, we first verify numerically that the weak scheme $\widehat{A}$ introduced in Section 2.2.4 exhibits second-order weak convergence, and we provide numerical illustrations related to the discussion in Section 2.2.6. We then simulate sample paths of the quadratic rough Heston model using the Euler–Markovian scheme and the weak scheme derived in Section 3.2.8, and compare their convergence efficiency.

4.1. Weak Simulation for the Rough Heston Model

In the following, we will price various kinds of vanilla and exotic options under different N in the rough Heston model. In all our examples, we used the same parameters as in [Gat22, Section 4.2] and [BB23a, Section 3], namely

$$\lambda = 0.3,\ \nu = 0.3,\ \theta = 0.02,\ V_0 = 0.02,\ \rho = -0.7,\ S_0 = 1,$$

while the values for N, T and H will vary. And we also need to choose one kind of Markovian approximation K^N of K of the form (2.3) for our weak scheme. The paper [BB23c] indicates the algorithm "BL2" shows promising results and it's more robust. Hence, we will use the approximations K^N generated by this algorithm BL2. Throughout, we will explicitly write down the nodes $\boldsymbol{x}$ and weights $\boldsymbol{w}$ that were generated using BL2.

4.1.1. European Call Options

We consider European call options as our first example since it's the easiest kind of options. Note that the characteristic functions of the stock price in the rough Heston model and its Markovian approximation are known in semi-closed form, we can compute reference prices using Fourier inversion, as showed in [AJEE19b]. Our goal is then to numerically observe convergence rates for the weak scheme under $N = 2, 3, 4$ as the number of time steps $M \to \infty$.

In Figure 4.1, we display the implied volatility smiles of European call options generated by the weak scheme for $N = 2, 3, 4$ with a fixed number of time steps $M = 32$. We observe that the schemes with $N = 2$ and $N = 3$ are sufficiently accurate, so that their implied volatility smiles are practically indistinguishable from the reference curve

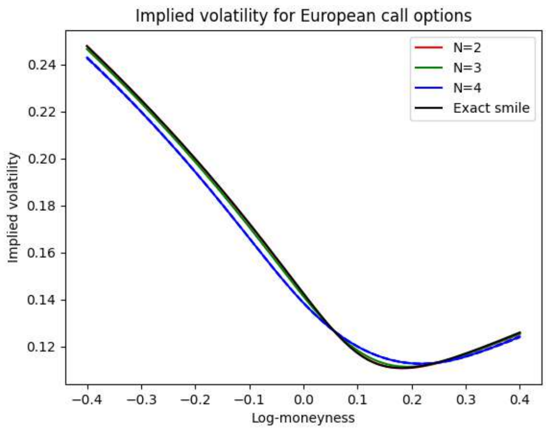

Figure 4.1.: Implied volatility smiles for European call option using $m = 2^{20}$ samples and $M = 32$ time steps, with $H = 0.1$, $T = 1$, and using the nodes and weights given in Table 4.1 for $N = 2, 3, 4$. The exact rough Heston smile (black) is obtained by Fourier inversion are not visible because it almost overlaps with the smiles from the weak scheme under $N = 2$ (red) and the weak scheme under $N = 3$ (green).

in the plot. By contrast, the scheme with $N = 4$ still exhibits a noticeable deviation for in-the-money strikes. This is mainly because the scheme with $N = 4$ employs a relatively large node $x_{\max} = 175.38958$ in Table 4.1, which induces strong mean reversion from $\boldsymbol{x}$.

Next, we investigate the numerical convergence rates of the implied volatility (IV) smiles under the weak scheme. One important issue to note is that the Monte Carlo (MC) error can be a major limiting factor. If the MC error is too large, it may dominate the total error and thus prevent us from accurately identifying the convergence rate.

In Figure 4.2, which shows the approximate IV smiles together with the reference smiles on the interval $[-1.5, 0.75]$, we observe that the MC error becomes significantly larger near the left and right boundaries. To mitigate this effect, we restrict the log-moneyness to the interval $[-0.05, 0.1]$ and consider 30 linearly spaced points within this range.

Furthermore, to reduce the sampling error, we employ randomized quasi-Monte Carlo (RQMC) methods introduced in [L'E09] instead of standard MC. The total number of samples is given by $m = m_1 \cdot m_2$, where m_2 denotes the number of Sobol points used, and m_1 is the number of random shifts applied to these Sobol points.

We consider two different Hurst parameters, namely $H = 0.1$ and $H = -0.2$. For each case, we use $N = 2$, $N = 3$, and $N = 4$ dimensions in the Markovian approximation of V. The corresponding nodes and weights are reported in Table 4.1.

The results for $H = 0.1$ are reported in Table 4.2 and illustrated in Figure 4.3. We observe second-order convergence in the middle region of the three curves. In the

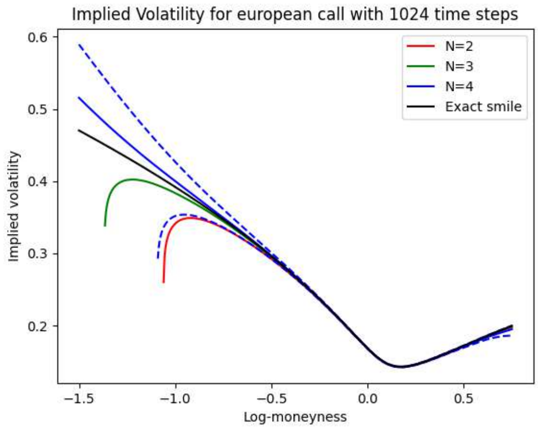

Figure 4.2.: Implied volatility smiles for European call option using $m = 2^{22}$ samples and $M = 1024$ time steps, with $H = 0.1$, $T = 1$, and using the nodes and weights given in Table 4.1 for $N = 2, 3, 4$ with MC method. And log-moneyness stays in $[-1.5, 0.75]$ with 450 linearly spaced values.

	$N=2$		$N=3$		$N=4$	
	Nodes	Weights	Nodes	Weights	Nodes	Weights
$H = 0.1$	0.05000	0.76733	0.033333	0.55543	0.02500	0.17062
	8.71709	3.22943	2.24161	1.11096	0.70001	0.99788
			46.83081	6.08578	9.59468	1.67259
					175.38958	10.29837
$H = -0.2$	1.66312	1.34831	1.37409	1.13509	1.19406	1.00577
	36.27832	9.38907	19.38019	4.89048	13.84629	3.38902
			299.63589	40.98515	125.09583	17.66860
					1911.61912	149.95173

Table 4.1.: Nodes and weights from the algorithm BL2 in [BB23c] for $T = 1$ and $H = 0.1$, $H = -0.2$, with $N = 2$, $N = 3$ and $N = 4$.

rightmost region (i.e., for sufficiently large M), however, the curves are strongly affected by the MC error, which lies in the range $(0.05, 0.1)$. This occurs because the Markovian approximation errors of the schemes with $N = 2, 3, 4$ are all significantly smaller than the MC error. Consequently, the total error eventually fluctuates around the MC error (and may even increase, e.g., for $N = 2$) or decreases only at a very slow rate.

Furthermore, by comparing the cases $N = 2$, $N = 3$, and $N = 4$, we observe that higher dimensions initially lead to slower convergence in the early stage, namely when $M \leq 64$. This effect is likely caused by the larger nodes (i.e., stronger mean reversion) appearing in higher-dimensional approximations, although the Markovian approximation error itself decreases as N increases.

Indeed, the weak scheme appears to require at least $M \geq 8$ (for $N = 2$), $M \geq 16$ (for $N = 3$), and $M \geq 64$ (for $N = 4$) time steps to approach its expected second-order convergence rate. This threshold increases with the size of the largest nodes, which are 8.71709, 46.83081, and 175.38958, respectively. Hence, when choosing the dimension N, there is a clear trade-off between achieving a more accurate Markovian approximation and maintaining an accurate time discretization.

To further isolate the effect of MC errors, we also investigate the discretization errors of the weak schemes for $N = 2, 3, 4$, reported in Table 4.3 and illustrated in Figure 4.4. The curves relatively clearly exhibit second-order convergence, although MC errors still affect the rightmost region, causing some instability.

Since it is difficult to determine the optimal pair (N, M) analytically (see Section 2.2.6), we present Figure 4.5 to illustrate the maximal relative error (over strikes) versus the computational time. We observe that, for a fixed error level, the computational time increases with N. Conversely, achieving a smaller error generally requires a larger value of N. It should also be noted that these results are still significantly influenced by MC errors.

The results for the Hurst parameter $H = -0.2$ are reported in Table 4.4 and illustrated in Figure 4.6, where the nodes and weights listed in Table 4.1 are used for $N = 2, 3, 4$. This corresponds to a much rougher regime, in which the rough Heston model can only be defined in a weak sense; see [AJDC23]. Nevertheless, the Markovian approximations remain standard SDEs.

In this regime, the Markovian approximation errors are larger and the corresponding nodes are also larger, leading to stronger mean reversion effects. Consequently, we restrict the log-moneyness to the interval $[0, 0.1]$ and consider 10 linearly spaced values. In this setting, the Markovian approximation error becomes the dominant component of the total error, making the convergence behavior of the weak schemes for $N = 2, 3, 4$ in Figure 4.6 easier to distinguish.

From the Figure 4.6, we observe that at least $M \geq 16$ (for $N = 2$), $M \geq 64$ (for $N = 3$), and $M \geq 512$ (for $N = 4$) time steps are required for the schemes to approach

	$N = 2$		$N = 3$		$N = 4$	
M	Error	Time	Error	Time	Error	Time
1	15.72	3.29	19.35	3.46	20.94	3.68
2	10.09	6.01	15.34	6.25	18.01	6.78
4	5.266	12.17	10.59	12.14	14.42	12.86
8	2.100	22.92	5.970	23.56	9.994	25.33
16	0.6498	45.24	2.714	47.07	5.800	50.45
32	0.1938	90.44	0.9563	94.00	2.840	100.96
64	0.0454	182.43	0.2634	189.06	1.138	202.00
128	0.0356	364.99	0.0717	381.55	0.3467	407.69
256	0.0403	721.15	0.0345	748.65	0.1592	799.72
512	0.0223	1493.84	0.0343	1553.11	0.0504	1648.48
1024	0.0372	3274.23	0.0326	3398.14	0.0507	3611.27

Table 4.2.: Maximal (over the strike) relative errors in % and computation time in s for the implied volatility smiles for European call option, where $H = 0.1$. The error of the Markovian approximation is 0.01232% for $N = 2$, 0.009917% for $N = 3$ and 0.001733% for $N = 4$. Most of the MC errors are under 0.05% and at most 0.07416%, where we used $m = 64 \cdot 2^{17}$ RQMC samples.

	$N = 2$	$N = 3$	$N = 4$
M	Error	Error	Error
1	15.71	19.34	20.94
2	10.08	15.34	18.01
4	5.254	10.58	14.42
8	2.088	5.964	9.994
16	0.6376	2.708	5.800
32	0.2022	0.9663	2.840
64	0.0534	0.2733	1.140
128	0.0440	0.0817	0.3460
256	0.0280	0.0284	0.1585
512	0.0347	0.0244	0.0498
1024	0.0288	0.0227	0.0500

Table 4.3.: Discretization relative errors in % for the implied volatility smiles for European call option, where $H = 0.1$. The error of the Markovian approximation is 0.01232% for $N = 2$, 0.009917% for $N = 3$ and 0.001733% for $N = 4$. Most of the MC errors are under 0.05% and at most 0.07416%, where we used $m = 64 \cdot 2^{17}$ RQMC samples.

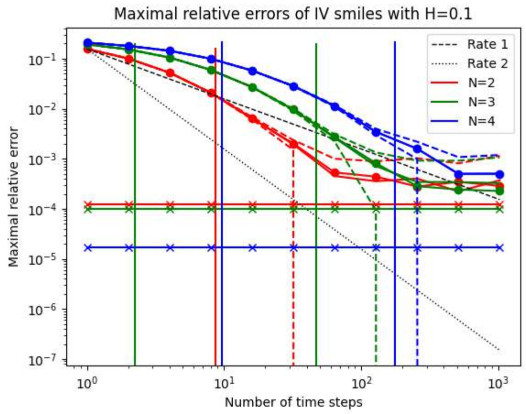

Figure 4.3.: Errors of implied volatility smiles of European call option with $H = 0.1$ under $N = 2, 3, 4$. The horizontal (blue/green/red) lines are the error of the Markovian approximation, while the vertical lines (blue/green/red) are the nodes x ($x > 1$) under $N = 2, 3, 4$. The solid (blue/green/red) lines represent the maximal errors between the true smile and the smile using simulation. The dashed (blue/green/red) lines indicate the 95% MC confidence intervals of these errors.

second-order convergence. In the final few steps for $N = 2$, however, the error remains almost constant (and may even fluctuate around the same level), which is caused by the MC error becoming dominant as the number of time steps increases.

We also report the maximal relative discretization errors of the weak schemes for $N = 2, 3, 4$ in Table 4.5 and Figure 4.7, and present Figure 4.8 to illustrate the maximal relative error (over strikes) versus the computational time. The middle region of the three curves in Figure 4.7 also exhibits second-order convergence. Moreover, the curves corresponding to $N = 3$ and $N = 4$ show a clear tendency toward second-order convergence as M increases.

From Figure 4.8, we observe that the weak scheme with $N = 2$ is the most efficient for achieving errors above 0.01%. However, for the schemes with $N = 2$ and $N = 3$, increasing the computational time beyond a certain point may lead to an increase in the maximal relative error. This behavior is mainly caused by the growing influence of MC errors, as the number of samples is also increased.

4.1.2. Implied Volatility Surfaces

Next, we consider the entire implied volatility surface of European call options. Under the standard set of parameters, we take the maturities and log-strikes as

$$T^{(i)} = \frac{i}{16}, \qquad i = 1, \ldots, 16,$$

	$N=2$		$N=3$		$N=4$	
M	Error	Time	Error	Time	Error	Time
1	25.89	2.04	28.52	1.80	29.34	1.80
2	24.26	3.45	29.34	2.95	30.49	2.97
4	18.59	6.10	27.82	5.21	30.21	5.41
8	11.33	11.20	24.49	9.83	29.19	10.15
16	5.001	21.88	19.14	18.98	27.16	19.74
32	1.647	43.16	12.65	36.58	23.65	38.82
64	1.868	82.69	6.659	74.93	18.34	78.02
128	2.003	166.62	2.562	153.44	12.09	161.16
256	1.992	324.82	0.9721	312.23	6.71	333.82
512	1.947	647.30	1.031	630.15	2.843	676.33
1024	1.984	1303.32	1.033	1291.18	0.8904	1453.01

Table 4.4.: Maximal (over the strike) relative errors in % and computation time in s for the implied volatility smiles of European call options with $H=-0.2$. The Markovian errors are 2.043% for $N=2$, 1.063% for $N=3$, and 0.603% for $N=4$. Most of the MC errors are below 0.13% and at most 0.1544%, where we used $m=25\cdot 2^{17}$ RQMC samples.

	$N=2$	$N=3$	$N=4$
M	Error	Error	Error
1	25.72	28.46	29.30
2	24.08	29.27	30.45
4	18.40	27.75	30.18
8	11.13	24.42	29.16
16	4.784	19.07	27.13
32	1.465	12.57	23.62
64	0.389	0.3084	18.30
128	0.0962	2.468	12.05
256	0.0664	0.6382	6.664
512	0.0944	0.153	2.794
1024	0.0573	0.0301	0.8414

Table 4.5.: Maximal relative discretization errors in % for the implied volatility smiles for European call options with $H=-0.2$. The Markovian error is 2.043% for $N=2$, 1.063% for $N=3$ and 0.603% for $N=4$. Most of the MC errors are below 0.13% and at most 0.1544%, and we used $m=25\cdot 2^{17}$ RQMC samples.

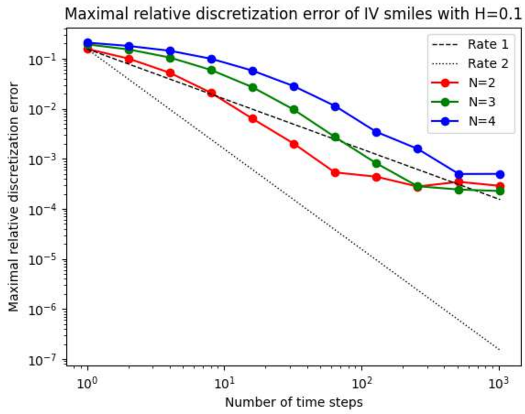

Figure 4.4.: Maximal relative discretization errors of implied volatility smiles for European call options with $H = 0.1$ under $N = 2, 3, 4$. The solid (blue/-green/red) lines represent the maximal relative discretization errors between the true smile and the smile using simulation

Nodes	0.20000	34.86835
Weights	1.33599	5.62277

Nodes	0.08399	5.64851	118.00625
Weights	0.80386	1.60786	8.80776

Table 4.6.: Nodes and weights from the algorithm BL2 in [BB23c] for $H = 0.1$, the vector of maturities $T = (1/16, 2/16, \dots, 1)$, and $N = 2, 3$.

and let $\boldsymbol{k}^{(i)}$ lie in the interval $[-0.05, 0.05]$, with 10 linearly spaced values for each $i = 1, \dots, 16$.

Furthermore, the corresponding nodes and weights obtained using the BL2 quadrature rule for the above maturity vector with $N = 2$, $N = 3$, and $N = 4$ are reported in Tables 4.6 and 4.7. Since we consider the entire vector of maturities, we adapt the BL2 algorithm as recommended in [BB23a, Section 3.3.2].

In Figure 4.9, we display the implied volatility surfaces produced by the weak scheme for $N = 2, 3, 4$ with a fixed number of time steps $M = 1024$. Since the figure is three-dimensional, the differences between the surfaces are not easily distinguishable. The quantitative results are therefore reported in Table 4.8 and illustrated in Figure 4.10.

Overall, the schemes with different values of N exhibit similar convergence behavior. In particular, the weak scheme approaches the expected second-order convergence rate

Nodes	0.05000	1.40001	19.18936	350.77917
Weights	0.22513	1.31672	2.20700	13.58878

Table 4.7.: Nodes and weights from the algorithm BL2 in [BB23c] for $H = 0.1$, the vector of maturities $T = (1/16, 2/16, \dots, 1)$, and $N = 4$.

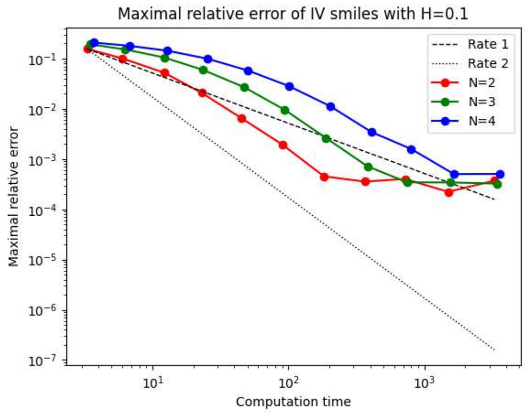

Figure 4.5.: Maximal relative errors of implied volatility smiles for European call options versus computation time, with $H = 0.1$ under $N = 2, 3, 4$. The solid (blue/green/red) lines represent the maximal relative errors between the true smile and the smile using simulation

when the number of time steps satisfies $M \geq 32$, $M \geq 64$, and $M \geq 128$ for $N = 2$, $N = 3$, and $N = 4$, respectively.

	$N = 2$		$N = 3$		$N = 4$	
M	Error	Time	Error	Time	Error	Time
16	9.899	12.22	14.87	13.09	19.28	12.70
32	4.663	20.88	8.423	21.12	13.15	20.82
64	2.194	36.82	3.691	37.59	7.144	37.36
128	1.581	68.08	1.221	74.41	3.076	72.22
256	1.717	135.84	0.4427	140.52	0.8642	144.35
512	1.619	268.20	0.2364	276.17	0.2472	285.91
1024	1.624	566.71	0.2204	577.21	0.2561	598.95

Table 4.8.: Maximal (over the strikes and maturities) relative errors in % for the implied volatility surfaces. The error of the Markovian approximation are 1.63% for $N = 2$, 0.1867% for $N = 3$, and 0.1841% for $N = 4$. Most of the MC errors are below 0.23%., where we used $m = 25 \cdot 2^{17}$ RQMC samples.

In Table 4.9 and Figure 4.11, we report the discretization errors for the weak schemes with $N = 2, 3, 4$, while Figure 4.12 illustrates the maximal relative errors (over both strikes and maturities) versus computational time. In the middle region of all three curves in Figure 4.11, second-order convergence can be roughly observed. However, as M increases further, the error decay slows down, indicating that the total error becomes increasingly dominated by the MC error component. Figure 4.12 exhibits a behavior similar to that observed for the implied volatility smiles with $H = 0.1$. But note that in this figure, only the range $M = 16$ to $M = 1024$ is displayed.

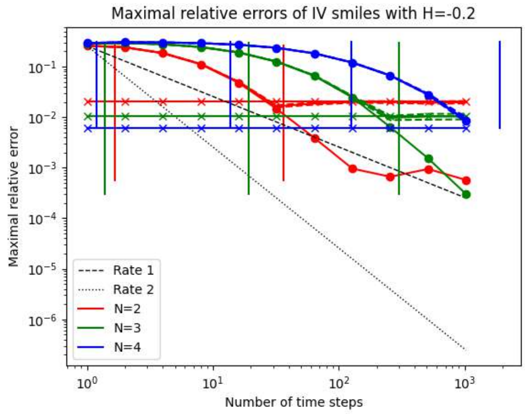

Figure 4.6.: Maximal relative error of implied volatility smiles of European call options with $H = -0.2$ under $N = 2, 3, 4$. The horizontal (blue/green/red) lines are the error of the Markovian approximation, while the vertical lines (blue/green/red) are the nodes x ($x > 1$) under $N = 2, 3, 4$. The solid (blue/green/red) lines represent the maximal errors between the true smile and the smile using simulation. The dashed (blue/green/red) lines indicate the 95% MC confidence intervals of these errors.

4.1.3. Geometric Asian Options

Here, we also apply the weak scheme to a path-dependent option, namely geometric Asian options. We remark that [BB23c] highlights an important advantage of geometric Asian options in the rough Heston model: the Fourier transform of the random variable $\int_0^T \log(S_t), \mathrm{d}t$ is available in semi-explicit form. This allows for an efficient computation of reference prices, which can then be compared with our simulation results. And given the simulated sample paths, the Asian option prices are computed using the trapezoidal rule in the following.

Under the same set of parameters, we consider a maturity $T = 1$ and log-strikes K in the interval $[-0.05, 0.05]$, with 20 linearly spaced values. We again report numerical results based on the Markovian approximations with dimensions $N = 2, 3, 4$, using the same nodes and weights as in Table 4.1 for the implied volatility smiles.

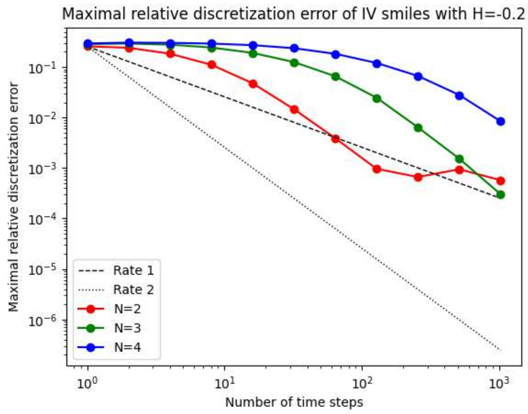

Figure 4.7.: Maximal relative discretization errors of implied volatility smiles for European call options with $H = -0.2$ under $N = 2, 3, 4$. The solid (blue/-green/red) lines represent the maximal relative discretization errors between the true smile and the smile using simulation.

	$N = 2$	$N = 3$	$N = 4$
M	Error	Error	Error
16	9.449	14.76	19.32
32	4.142	8.301	13.19
64	1.408	3.563	7.192
128	0.397	1.093	3.126
256	0.2436	0.3213	0.9274
512	0.1366	0.1794	0.236
1024	0.1186	0.1141	0.2282

Table 4.9.: Maximal relative discretization errors in % for the implied volatility surfaces for European call options with $H = 0.1$. The Markovian error is 1.63% for $N = 2$, 0.1867% for $N = 3$ and 0.1841% for $N = 4$. Most of the MC errors are below 0.23%, where we use $m = 25 \cdot 2^{17}$ RQMC samples.

In Figure 4.13, we display the implied volatility smiles generated by the weak scheme for $N = 2, 3, 4$ with a fixed number of time steps $M = 32$. As in the case of European call options, the schemes with $N = 2$ and $N = 3$ are more accurate than the scheme with $N = 4$. This is due to the stronger mean reversion induced by the larger maximal node $x_{\max} = 175.38958$ used when $N = 4$.

The numerical results are reported in Table 4.10 and illustrated in Figure 4.14. We observe that the weak schemes approach second-order convergence for $M \geq 8$, $M \geq 128$, and $M \geq 128$ time steps for $N = 2$, $N = 3$, and $N = 4$, respectively. However, all curves gradually exhibit a slowdown in convergence and eventually approach a first-order rate as M becomes large.

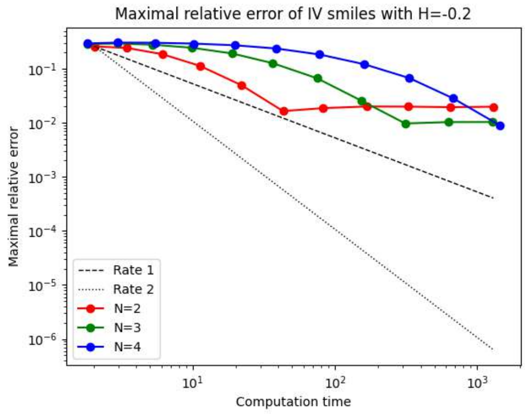

Figure 4.8.: Maximal relative errors of implied volatility smiles for European call options versus computation time, with $H = -0.2$ under $N = 2, 3, 4$. The solid (blue/green/red) lines represent the maximal errors between the true smile and the smile using simulation.

	$N=2$		$N=3$		$N=4$	
M	Error	Time	Error	Time	Error	Time
1	30.11	2.42	33.68	2.13	35.01	2.25
2	14.04	3.22	19.22	3.34	21.81	3.54
4	6.967	5.65	12.44	5.86	16.22	6.19
8	3.403	10.36	7.10	10.85	11.20	11.50
16	1.219	19.97	5.364	21.19	6.734	22.67
32	0.5047	40.00	2.567	43.64	5.234	46.50
64	0.4053	84.09	0.8136	88.77	2.943	95.80
128	0.3524	176.27	0.2734	183.82	0.928	197.66
256	0.6458	272.84	0.06947	295.24	0.2791	304.82
512	0.6135	808.74	0.05072	881.52	0.178	918.61
1024	0.6032	1223.68	0.02217	1493.25	0.1189	1558.14

Table 4.10.: Maximal (over the strikes) relative errors in % and computation time in s for the implied volatility smiles of geometric Asian call options with $H = 0.1$. The Markovian errors are 0.5291% for $N = 2$, 0.01256% for $N = 3$, and 0.003018% for $N = 4$. Most of the MC errors are below 0.28%, where we use $m = 25 \cdot 2^{17}$ RQMC samples.

In Table 4.11 and Figure 4.15, we report the maximal relative discretization errors for the weak schemes with $N = 2, 3, 4$, while Figure 4.16 illustrates the maximal relative errors (over strikes) versus computational time for $N = 2, 3, 4$. From Figure 4.15, we observe that all three curves initially exhibit first-order convergence, then approach second-order convergence over an intermediate range, and eventually experience a significant slowdown in the convergence rate. This behavior is mainly caused by the

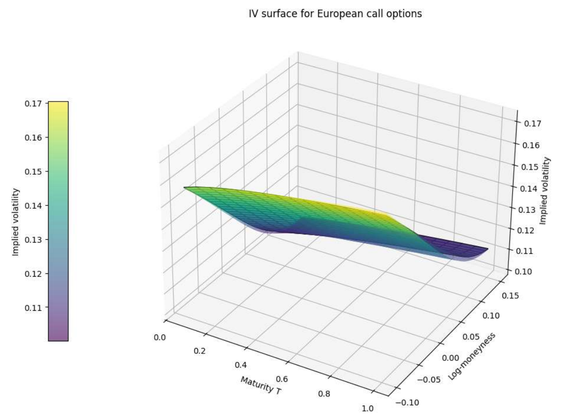

Figure 4.9.: Implied volatility surfaces for European call options using $m = 2^{17}$ samples and $M = 1024$ time steps, with $H = 0.1$, $T^{(i)} = \frac{i}{16}, \quad i = 1, \ldots, 16$, and using the nodes and weights given in Table 4.6 and Table 4.7 for $N = 2, 3, 4$. The black wireframe represents the true implied volatility surface, while the colored semi-transparent surfaces correspond to the approximate surfaces of the weak schemes under $N = 2, 3, 4$.

increasing influence of MC errors. Figure 4.16 shows a similar tendency to that observed for the implied volatility smiles of European call options with $H = 0.1$. Moreover, in the initial regime, we roughly observe a proportional relationship between the maximal relative error and the computational time, with an empirical rate of approximately 2 for $N = 2$ and 1 for $N = 3, 4$.

4.2. Weak Simulation for the Quadratic Rough Heston Model

In this section, we introduce two numerical schemes: the weak scheme presented in Section 3.2.8 and the Euler–Markovian scheme. We then simulate sample paths of (S, V) in the quadratic rough Heston model over the time interval $[0, 1]$ (i.e., $T = 1$) using these schemes. Finally, we assess their weak convergence rates via option pricing for European options.

We consider the following set of parameters:

$$H = 0.1,\ a = 0.5,\ b = 0.2,\ c = 0.3,\ \eta = 0.3,\ \theta = 0.02,\ Z_0 = 0.02,\ S_0 = 1.$$

We also use the approximations K^N generated by the BL2 algorithm, as given in Table 4.1.

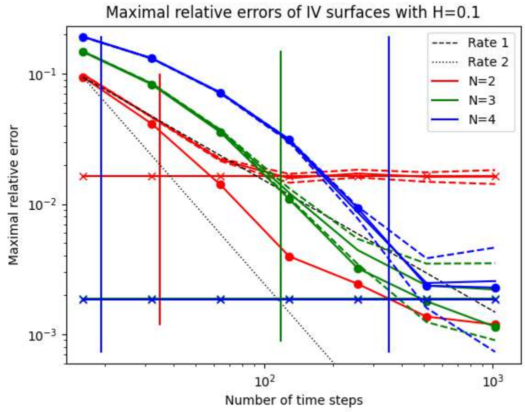

Figure 4.10.: Maximal relative errors of implied volatility surfaces of European call options with $H = 0.1$ under $N = 2, 3, 4$. The horizontal (blue/green/red) lines are the error of the Markovian approximation, while the vertical lines (blue/green/red) are the nodes x ($x > 1$) under $N = 2, 3, 4$. The solid (blue/green/red) lines represent the maximal errors between the true smile and the smile using simulation. The dashed (blue/green/red) lines indicate the 95% MC confidence intervals of these errors.

	$N = 2$	$N = 3$	$N = 4$
M	Error	Error	Error
1	29.81	33.69	35.01
2	13.90	19.23	21.81
4	6.859	12.45	16.22
8	3.953	7.108	11.20
16	1.757	5.351	6.734
32	0.5897	2.554	5.231
64	0.1556	0.801	2.94
128	0.1777	0.2608	0.925
256	0.1173	0.0569	0.2761
512	0.0848	0.0382	0.1807
1024	0.0745	0.0324	0.1219

Table 4.11.: Maximal relative dicretization errors in % for the implied volatility smiles for geometric Asian call options with $H = 0.1$. The Markovian error is 0.5291% for $N = 2$, 0.01256% for $N = 3$, and 0.003018% for $N = 4$. Most of the MC errors are below 0.28%, where we use $m = 25 \cdot 2^{17}$ RQMC samples.

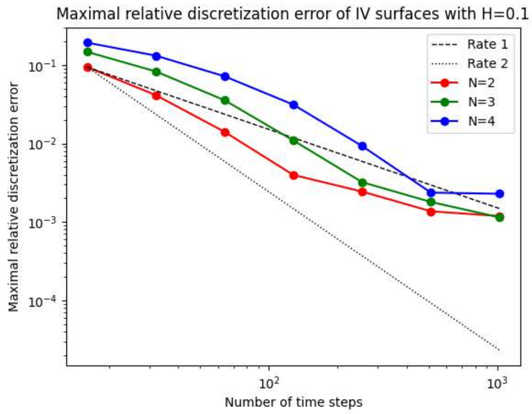

Figure 4.11.: Maximal relative discretization errors of implied volatility surfaces for European call options with $H = 0.1$ under $N = 2, 3, 4$. The solid (blue/-green/red) lines represent the maximal discretization errors between the true surface and the surface using simulation.

4.2.1. Algorithms

Next, for a given number of time steps M, we discretize the interval $[0, 1]$ into M subintervals of equal length $\Delta t \coloneqq 1/M$. This yields a partition $(t_m)_{m=0}^{M}$ of $[0, 1]$, where $t_m = m\Delta t$.

Euler-Markovian Scheme

The Euler-Markovian scheme is directly using the implicit Euler scheme for the weak Markovian approximation $(S^N, \boldsymbol{Z}^N)$ for (S, V). Recall the dynamics of them are given by

$$dS_t^N = S_t^N \sqrt{V_t^N} dW_t,\ V_t^N = a(Z_t^N - b)^2 + c,\ Z_t^N = w^\top \boldsymbol{Z}_t^N,$$

$$dZ_t^{(i)} = -x_i(Z_t^{(i)} - z_0^i)dt + (\theta - Z_t^N)dt + \eta\sqrt{a(Z_t^N - b)^2 + c}\, dW_t,\ Z_0^{(i)} = z_0^i,\ i = 1, \dots, N.$$

We remark here we apply the implicit Euler scheme rather than the explicit Euler scheme on the drift part of $\boldsymbol{Z}^N$ due to the stiffness of the SDE governing $\boldsymbol{Z}^N$. Note that the mean-reversion $\boldsymbol{x}$ will lead to unstable behavior if using the explicit scheme, whereas the implicit Euler scheme effectively mitigates large oscillations and yields improved stability.

On the time grid, we define the approximation $(S^{N,M}, \boldsymbol{Z}^{N,M})$ of $(S^N, \boldsymbol{Z}^N)$ using the drift-implicit Euler scheme given by

$$\begin{aligned}
S_{t_{m+1}}^{N,M} &= S_{t_m}^{N,M} \sqrt{V_{t_m}^{N,M}} \Delta W_m,\\
V_{t_{m+1}}^{N,M} &= a(Z_{t_{m+1}}^{N,M} - b)^2 + c,\\
Z_{t_{m+1}}^{(i),M} &= Z_{t_m}^{(i),M} - x_i(Z_{t_{m+1}}^{(i),M} - z_0^i)\Delta t + (\theta - Z_{t_{m+1}}^{N,M})\Delta t + \eta\sqrt{a(Z_{t_m}^{N,M} - b)^2 + c}\, \Delta W_m,
\end{aligned}$$

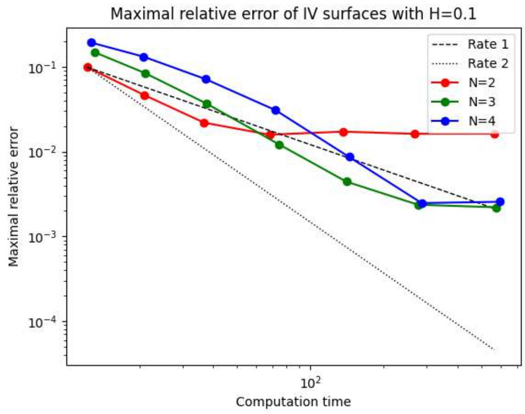

Figure 4.12.: Maximal relative errors of implied volatility surfaces for European call options versus computation time, with $H = 0.1$ under $N = 2, 3, 4$. The solid (blue/green/red) lines represent the maximal discretization errors between the true surface and the surface using simulation.

for $m = 0, \dots, M-1$, where $Z^{N,M}_{t_{m+1}} = w^\top \cdot \boldsymbol{Z}^{N,M}_{t_{m+1}}$, $S^{N,M}_0 = S_0$, $Z^{(i),M}_0 = z^i_0$, and $\boldsymbol{Z}^{N,M}_{t_m} := (Z^{(i),M}_{t_m})^N_{i=1}$. Also, we denote the Brownian increments by $\Delta W_m := W_{t_{m+1}} - W_{t_m}$. We notice that $Z^{(i),M}_{t_{m+1}}$ can be obtained by solving a linear system here. The sample paths of stock price S_t and volatility V_t that simulated by the Euler-Markovian scheme under $N = 3$ are illustrated in Figures 4.17 and 4.18, respectively.

Weak Approximate scheme

The second scheme is the algorithm introduced in Section 3.2.8, where we apply a Strang splitting scheme to simulate the volatility process V, together with a tailored scheme—following the same strategy as in the rough Heston case—to simulate the stock price S. On the time grid, the scheme can be written as follows.

$$
\begin{aligned}
\boldsymbol{Z}^{N,M}_{t_{m+1}} &= \widehat{A}'(\boldsymbol{Z}^{N,M}_{t_m}, t_{m+1} - t_m),\\
V^{N,M}_{t_{m+1}} &= a(Z^{N,M}_{t_{m+1}} - b)^2 + c,\\
\widehat{\boldsymbol{Y}}^{(1,N,M)}_{t_{m+1}} &= \boldsymbol{y}^{(1)} + (\boldsymbol{Z}^{N,M}_{t_m} + \boldsymbol{Z}^{N,M}_{t_{m+1}})\frac{\Delta t}{2},\\
\widehat{Y}^{(2,N,M)}_{t_{m+1}} &= y^{(2)} + ((w^\top \cdot \boldsymbol{Z}^{N,M}_{t_m})^2 + (w^\top \cdot \boldsymbol{Z}^{N,M}_{t_{m+1}})^2)\frac{\Delta t}{2},\\
\widehat{S}^{N,M}_{t_{m+1}} &= \widehat{S}^{N,M}_{t_m} \exp\Bigg((-\frac{\left(x_1 z^{(1)}_0 + \theta\right)}{\eta} - \frac{1}{2}(ab^2 + c))\Delta t + (ab + \frac{1}{\eta})(\widehat{Y}^{(1,N,M)}_{t_{m+1}} - \widehat{Y}^{(1,N,M)}_{t_m})\\
&\qquad + \frac{x_1}{\eta}(\widehat{Y}^{((1,1),N,M)}_{t_{m+1}} - \widehat{Y}^{((1,1),N,M)}_{t_m}) - \frac{a}{2}(\widehat{Y}^{(2,N,M)}_{t_{m+1}} - \widehat{Y}^{(2,N,M)}_{t_m}) + \frac{Z^{(1),M}_{t_{m+1}} - Z^{(1),M}_{t_m}}{\eta}\Bigg),
\end{aligned}
$$

for $m = 0, \dots, M-1$, where $\widehat{\boldsymbol{Y}}^{(1,N,M)}_{t_m} := (\widehat{Y}^{((1,i),N,M)}_{t_m})_{i=1,\dots,N}$. The sample paths of stock price S_t and volatility V_t are illustrated in Figures 4.19 and 4.20, respectively.

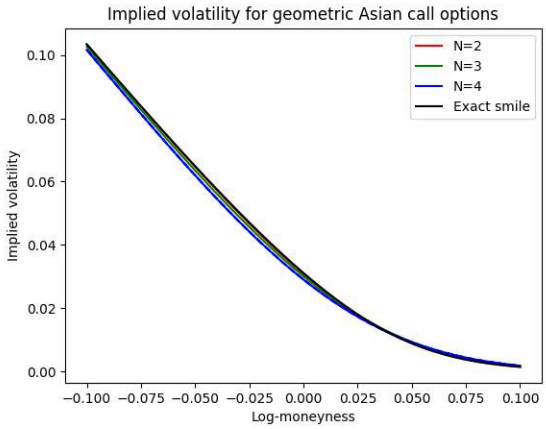

Figure 4.13.: Implied volatility smiles for geometric Asian call options using $m = 2^{17}$ samples and $M = 32$ time steps, with $H = 0.1$, $T = 1$, and using the nodes and weights given in Table 4.1 for $N = 2, 3, 4$ with QRMC method. And log-moneyness stays in $[-0.1, 0.1]$ with 40 linearly spaced values.

$N = 2$		$N = 3$	
Nodes	Weights	Nodes	Weights
0.00313	0.77264	0.00208	0.85368
1.32367	0.48397	1.29373	0.29388
		19.73078	0.43614

Table 4.12.: Nodes and weights from the algorithm BL2 in [BB23c] for $T = 1$ and $H = 0.4$ with $N = 2$ and $N = 3$.

Computational Cost of Simulation

We remark here the Euler-Markovian Scheme and the weak approximation scheme both cost $\mathcal{O}(N^2 M)$. But in practical, $N \in \{2, 3\}$ is enough. So they will both exhibit linear cost.

4.2.2. Benchmark

In this part, we use the following set of parameters:

$$T = 1,\ H = 0.4,\ a = 0.5,\ b = 0.2,\ c = 0.3,\ \eta = 0.3,\ \theta = 0.02,\ Z_0 = 0.02,\ S_0 = 1.$$

Here, we choose a relatively larger value $H = 0.4$ to avoid the extremely rough regime. We again use the approximations K^N generated by the BL2 algorithm, as given in Table 4.12.

We vary the dimension N to compute European call option prices for the quadratic rough Heston model by using the two schemes introduced above, with the aim of

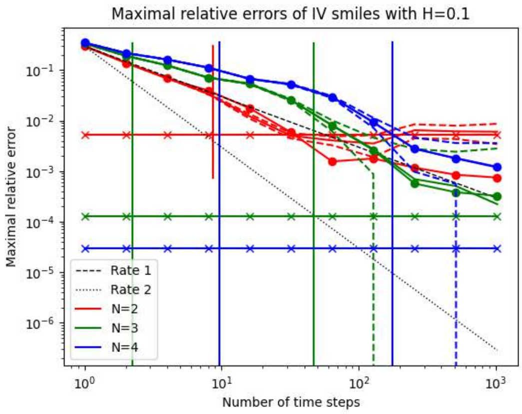

Figure 4.14.: Maximal relative errors of implied volatility smiles of geometric Asian call options with $H = 0.1$ under $N = 2, 3, 4$. The horizontal (blue/green/red) lines are the error of the Markovian approximation, while the vertical lines (blue/green/red) are the nodes x ($x > 1$) under $N = 2, 3, 4$. The solid (blue/green/red) lines represent the maximal errors between the true smile and the smile using simulation. The dashed (blue/green/red) lines indicate the 95% MC confidence intervals of these errors.

investigating their weak convergence rates. Since the characteristic function of the volatility process V in the quadratic rough Heston model is not available in semi-closed form, Fourier methods for pricing cannot be applied. In particular, unlike in the rough Heston model, we are unable to compute a reference option price via Fourier method.

Consequently, we assess the convergence rates of the two simulation schemes in Section 4.2 by comparing their numerical results with a more refined solution taken as a reference. More precisely, we use a Richardson-type extrapolation error to estimate the convergence rates of the schemes. The underlying principle is described as follows.

Let $u(M)$ denote a family of numerical approximations to the European call option price u, where M is the number of time steps. Since the exact value u is not available in closed form in our model, the global error

$$\varepsilon(M) := \|u(M) - u\|$$

cannot be computed directly. A standard remedy is to replace u with a finer numerical approximation and analyze the resulting *adjacent-difference* sequence instead.

We remark that the IV smile of the European call option is not used to estimate the weak convergence rate, as its computation requires additional calculations that can amplify other sources of error unrelated to the discretization, potentially distorting the convergence assessment.

Consider two consecutive grid levels $M_m := 2^m$ and $M_{m-1} = 2^{m-1}$. Define the *adjacent difference*

$$d_m := \big\|u(M_m) - u(M_{m-1})\big\|. \tag{4.1}$$

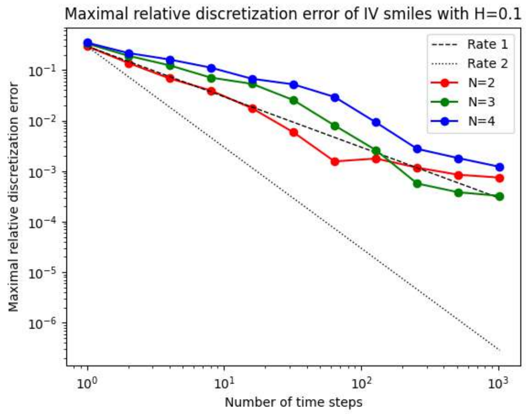

Figure 4.15.: Maximal relative dicretization errors of implied volatility silmes for geometric Asian call options with $H = 0.1$ under $N = 2, 3, 4$. The solid (blue/green/red) lines represent the maximal discretization errors between the true smile and the smile using simulation.

We see that the adjacent difference satisfies

$$d_m \;=\; \big\|u(M_m) - u(M_{m-1})\big\| \;\leq\; \varepsilon(M_m) + \varepsilon(M_{m-1}) \;\approx\; C\, M_m^{-\alpha}(1 + 2^{\alpha})\,, \tag{4.2}$$

if we assume the approximation scheme has weak rate α, where we use the triangle inequality here. In particular,

$$\frac{d_{m-1}}{d_m} \;\approx\; \frac{M_m^{\alpha}}{M_{m-1}^{\alpha}} \;=\; 2^{\alpha}, \tag{4.3}$$

so the ratio of adjacent differences converges to the same limit 2^{α} as the ratio of true errors. Taking logarithms of (4.3) yields

$$\widehat{\alpha}_m \;:=\; \frac{\log(d_{m-1}/d_m)}{\log(M_m/M_{m-1})} \;=\; \frac{\log d_{m-1} - \log d_m}{\log M_m - \log M_{m-1}} \;\approx\; \alpha. \tag{4.4}$$

This is the slope of the log-log plot of d_m against M_m, estimated from two consecutive points. As $M_m \to \infty$ (i.e. as m grows), the higher-order terms will become negligible and $\widehat{\alpha}_m \to \alpha$.

We remark that if the MC error at step M_m is of the same order as the adjacent difference d_m—or even larger—then the estimator (4.4) becomes unreliable. In particular,

$$\big\|u(M_m) - u(M_{m-1})\big\| \approx \epsilon_{discret} + \epsilon_{MC}(M) + \epsilon_{MC}(M-1),$$

where $\epsilon_{\text{discret}}$ denotes the discretization error at steps M_m and M_{m-1}, and $\epsilon_{\text{MC}}(M)$ denotes the MC error at step M.

To reduce the MC error, we restrict the log-strikes k in the interval $[0, 0.2]$ with 20 linearly spaced values. And we use the same random number sequence when simulating

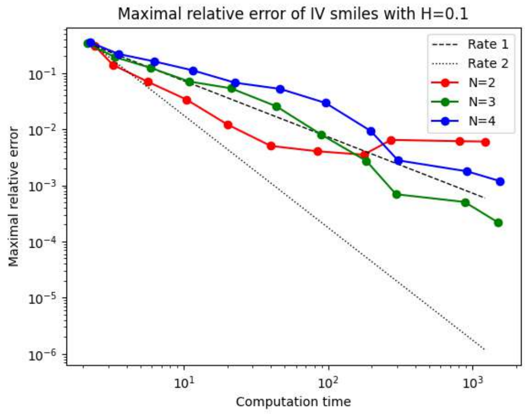

Figure 4.16.: Maximal relative errors of implied volatility smiles for geometric Asian call options with $H = 0.1$ under $N = 2, 3, 4$. The solid (blue/green/red) lines represent the maximal errors between the true smile and the smile using simulation.

Figure 4.17.: Sample paths of the stock price S_t in the quadratic rough Heston model with $H = 0.1$ and $N = 3$, generated by using the Euler–Markovian scheme. We display three representative sample paths, selected from 2^{15} simulated paths, using different colors. The number of time steps is chosen as $M = 256$.

$u(M_m)$ and $u(M_{m-1})$. That is, we generate M_m random numbers to simulate $u(M_m)$ and derive the M_{m-1} random numbers for the coarser grid by summing adjacent pairs (for the Euler-Markovian scheme) or taking the even-indexed elements (for the weak scheme) to compute $u(M_{m-1})$. We then compute the difference $u(M_m) - u(M_{m-1})$. This approach largely cancels out the MC error, leaving primarily the discretization error, which enables a reliable estimation of the weak convergence rate.

The implied volatility smiles for European call options simulated using the two schemes under $N = 2$ and $N = 3$ in Section 4.2 are shown in Figures 4.21 and 4.22, respectively. In our framework, each $u(M_m)$ is itself a quasi-Monte Carlo estimator. The figures also display the corresponding statistical confidence intervals.

We observe that the central portions of the two curves no matter under $N = 2$ or $N = 3$ both nearly coincide and are indistinguishable, whereas the left and right tails

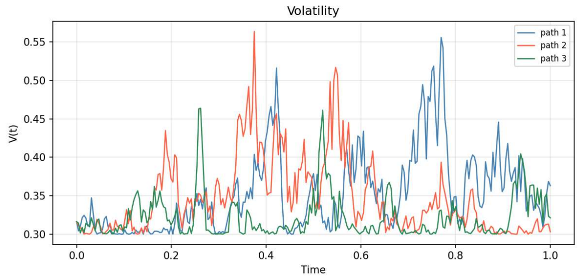

Figure 4.18.: Sample paths of volatility V_t in the quadratic rough Heston model with $H = 0.1$ and $N = 3$, generated by using the Euler–Markovian scheme. We display three representative sample paths, selected from 2^{15} simulated paths, using different colors. The number of time steps is chosen as $M = 256$.

Figure 4.19.: Sample paths of stock price S_t in the quadratic rough Heston model with $H = 0.1$ and $N = 3$, generated by using the weak scheme in Section 3.2.8. We display three representative sample paths, selected from 2^{15} simulated paths, using different colors. The number of time steps is chosen as $M = 256$.

exhibit slight differences, which are influenced by the massive MC error. Additionally, the curves generated by the weak scheme described in Section 3.2.8 display a "wavy" pattern. This phenomenon arises primarily because the simulations use three-point random variables as substitutes for Brownian motion.

We report the main numerical results for $N = 2$ and $N = 3$ in Tables 4.13 and 4.14, and in Figures 4.23 and 4.24, respectively. It should be emphasized that d_m is a quantity used to measure the convergence rate, rather than the maximal relative error as in Section 4.1. Therefore, the magnitude of d_m cannot be directly used to assess the overall performance of a given scheme. This explains why the values of d_m for the weak scheme are significantly larger than those for the Euler–Markovian scheme. We also note that Tables 4.13 and 4.14 start from $M = 2$ instead of $M = 1$, due to the computation of adjacent differences of $u(M_m)$.

From Figures 4.23 and 4.24, we observe that the Euler–Markovian scheme achieves weak convergence of order 1 for both $N = 2$ and $N = 3$, which is consistent with the

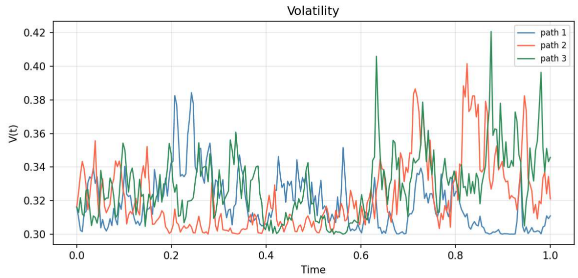

Figure 4.20.: Sample paths of volatility V_t in the quadratic rough Heston model with $H = 0.1$ and $N = 3$, generated by using the weak scheme in Section 3.2.8. We display three representative sample paths, selected from 2^{15} simulated paths, using different colors. The number of time steps is chosen as $M = 256$.

theoretical results. The weak scheme under both $N = 2$ and $N = 3$ exhibits similar behavior: it achieves a convergence rate slightly higher than 1, but does not reach order 2. Moreover, both the weak scheme and the Euler–Markovian scheme exhibit a slowdown in the tail region, due to the increasing MC error.

We also remark the presence of an abnormal point at the fourth adjacent difference d_4 for both $N = 2$ and $N = 3$, indicated by the *-entries in Tables 4.13 and 4.14. This is mainly due to pre-asymptotic effects occurring for certain log-strike values $k \in [0, 0.2]$. When the range of log-strikes is modified, the location of this abnormal point shifts accordingly. However, this phenomenon has a negligible impact on the evaluation of the convergence rate of the scheme, since it occurs only at a very early stage.

	Weak		Euler	
M_m	d_m	Time	d_m	Time
2	3.5079e-02	0.82	7.5872e-04	0.51
4	1.7984e-02	1.26	4.0381e-04	0.64
8	7.1432e-03	2.01	2.0077e-04	1.04
16	1.1614e-02*	3.24	9.7455e-05	1.69
32	6.0438e-03	5.93	4.7688e-05	3.23
64	3.0709e-03	11.68	2.7641e-05	5.67
128	1.3148e-03	23.49	1.4397e-05	11.13
256	6.3604e-04	48.98	7.2305e-06	23.84
512	3.4846e-04	104.04	4.3303e-06	46.73
1024	2.5214e-04	220.08	2.1906e-06	104.91

Table 4.13.: Maximal (over the strike) adjacent differences d_m in \$ and computation time in s for European call option prices with $H = 0.4$, under $N = 2$. Most of the MC errors are around 0.001%, where we use 2^{17} samples.

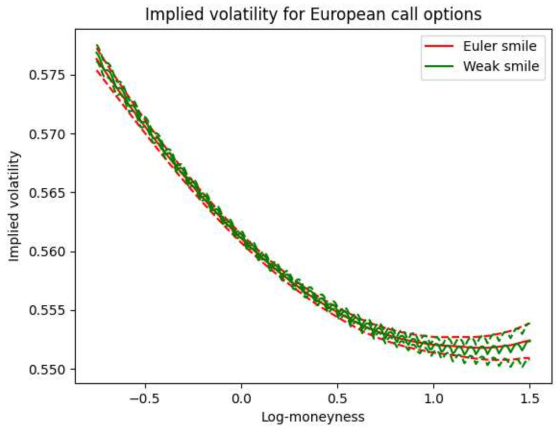

Figure 4.21.: Implied volatility smiles for European call option using $m = 2^{19}$ samples and $M = 256$ time steps, with $H = 0.4$, $T = 1$, and using the nodes and weights given in Table 4.12 for $N = 2$ with RQMC method. The log-moneyness stays in $[-0.75, 1.5]$ with 225 linearly spaced values. The solid red curve represents the approximate IV smile simulated by the Euler-Markovian scheme, while the solid green curve represents the approximate IV smile simulated by the weak scheme in Section 3.2.8. The dashed curves (green/red) indicate the 95% MC confidence intervals for the respective smiles.

	Weak		Euler	
M_m	d_m	Time	d_m	Time
2	3.0529e-02	0.96	7.6043e-04	0.70
4	1.6667e-02	1.33	3.9952e-04	0.81
8	7.9302e-03	2.00	1.9359e-04	1.19
16	1.2356e-02*	3.52	9.0288e-05	1.94
32	6.7654e-03	6.37	4.1782e-05	3.38
64	3.5114e-03	12.90	2.4350e-05	6.42
128	1.5452e-03	23.02	1.2348e-05	12.39
256	7.4460e-04	47.18	6.3129e-06	25.20
512	4.0655e-04	111.42	4.0026e-06	51.16
1024	2.8151e-04	249.37	2.0866e-06	122.80

Table 4.14.: Maximal (over the strike) adjacent differences d_m in \$ and computation time in s for European call option prices with $H = 0.4$, under $N = 3$. Most of the MC errors are around 0.001%, where we use 2^{17} samples.

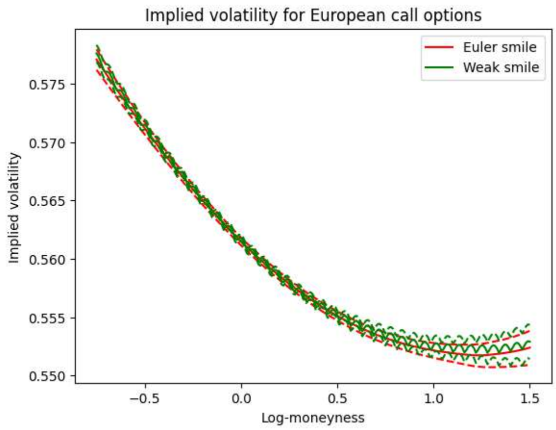

Figure 4.22.: Implied volatility smiles for European call option using $m = 2^{19}$ samples and $M = 256$ time steps, with $H = 0.4$, $T = 1$, and using the nodes and weights given in Table 4.12 for $N = 3$ with RQMC method. The log-moneyness stays in $[-0.75, 1.5]$ with 225 linearly spaced values. The solid red curve represents the approximate IV smile simulated by the Euler-Markovian scheme, while the solid green curve represents the approximate IV smile simulated by the weak scheme in Section 3.2.8. The dashed curves (green/red) indicate the 95% MC confidence intervals for the respective smiles.

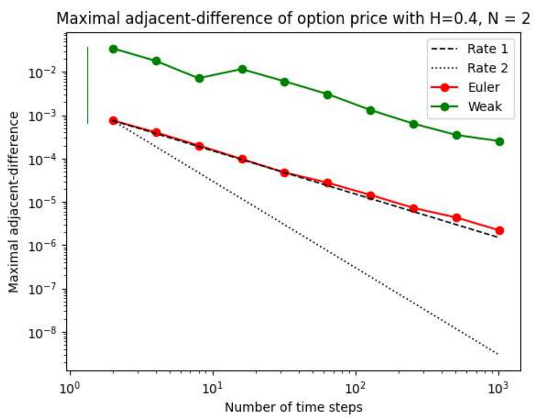

Figure 4.23.: Maximal adjacent differences of European call option prices with $H = 0.4$ under $N = 2$. The vertical green line indicates the node x $(x > 1)$ for $N = 2$. The solid green and red lines represent the maximal differences between adjacent option prices obtained from the simulations.

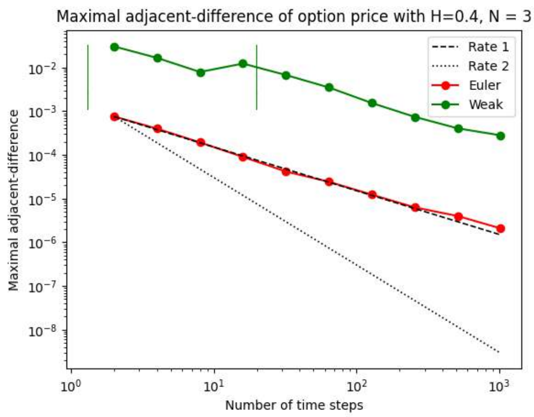

Figure 4.24.: Maximal adjacent differences of European call option prices with $H = 0.4$ under $N = 3$. The vertical green lines indicate the nodes x $(x > 1)$ for $N = 3$. The solid green and red lines represent the maximal differences between adjacent option prices obtained from the simulations.

A. Appendix

A.1. Analysis of Weak Error

Proof of Proposition 2.2.5

If $n > T/\eta$, we have clearly

$$\mathbb{E}\big[\|\widehat{X}^n_{t^n_{i+1}}\|^q\big] \leq (1 + C_q T/n)\mathbb{E}\big[\|\widehat{X}^n_{t^n_i}\|^q\big] + C_q T/n,$$

and thus $\mathbb{E}\big[\|\widehat{X}^n_{t^n_i}\|^q\big] \leq u_i$ where $u_0 = \|\widehat{X}^n_{t^n_0}\|^q$ and $u_{i+1} = (1 + C_q T/n)u_i + C_q T/n$. Therefore, we iteratively use them to derive

$$u_i = (1 + C_q T/n)^i u_0 - 1 \leq \|\widehat{X}^n_{t^n_0}\|^q e^{C_q T},$$

we get the desired result eventually.

Proof of Theorem 2.2.16

We have the following expansions due to the Proposition 2.2.14. The first scheme gives:

$$\begin{aligned}
&\left(I + \frac{t}{2}L_m + \frac{t^2}{8}L_m^2 + \text{rem}\right) \times \cdots \times \left(I + \frac{t}{2}L_2 + \frac{t^2}{8}L_2^2 + \text{rem}\right) \\
&\quad \times \left(I + tL_1 + \frac{t^2}{2}L_1^2 + \text{rem}\right)\left(I + \frac{t}{2}L_2 + \frac{t^2}{8}L_2^2 + \text{rem}\right) \times \cdots \times \left(I + \frac{t}{2}L_m + \frac{t^2}{8}L_m^2 + \text{rem}\right) \\
&= I + t\Sigma L + \frac{t^2}{2}(\Sigma L)^2 + \text{rem}
\end{aligned}$$

where rem denotes a remainder of order 3.

Similarly,

$$\begin{aligned}
&\left(I + tL_1 + \frac{t^2}{2}L_1^2 + \text{rem}\right)\cdots\left(I + tL_m + \frac{t^2}{2}L_m^2 + \text{rem}\right) \\
&\quad = I + t\sum_{j=1}^{m} L_j + \frac{t^2}{2}\left(\sum_{j=1}^{m} L_j^2 + 2\sum_{1\leq j<k\leq m} L_j L_k\right) + \text{rem} \\
&\quad = I + t\sum L + \frac{t^2}{2}\left(\sum L\right)^2 + \text{rem},
\end{aligned}$$

and therefore the second scheme is also a potential second order discretization scheme for ΣL.

A.2. Higher Order Approximation

Proof of Proposition 2.2.18

First note that from the equation (2.48), we have

$$\widehat{P}_{h_1}^{[i]} - P_{ih_1} = \sum_{i_1=0}^{i-1} \widehat{P}_{h_1}^{[i-(i_1+1)]}(\widehat{P}_{h_1} - P_{h_1})P_{h_1}^{[i_1]}$$

and

$$\widehat{P}_{h_2}^{[n]} - P_{h_1} = \widehat{P}_{h_2}^{[n]} - P_{h_2}^{[n]} = \sum_{j=0}^{n-1} \widehat{P}_{h_2}^{[n-(j+1)]}(\widehat{P}_{h_2} - P_{h_2})P_{h_2}^{[j]}.$$

Plugging these two identities successively in the equation (2.48), we obtain

$$\widehat{P}_{h_1}^{[n]} - P_T = \sum_{i=0}^{n-1} \widehat{P}_{h_1}^{[n-(i+1)]}(\widehat{P}_{h_1} - \widehat{P}_{h_2}^{[n]})\widehat{P}_{h_1}^{[i]} + R, \tag{A.1}$$

with the remainder given by

$$\begin{aligned} R = &\sum_{i=0}^{n-1} \widehat{P}_{h_1}^{[n-(i+1)]} \left(\sum_{j=0}^{n-1} \widehat{P}_{h_2}^{[n-(j+1)]}(\widehat{P}_{h_2} - P_{h_2})P_{h_2}^{[j]} \right) \widehat{P}_{h_1}^{[i]} \\ &- \sum_{i=0}^{n-1} \widehat{P}_{h_1}^{[n-(i+1)]}(\widehat{P}_{h_1} - P_{h_1}) \sum_{i_1=0}^{i-1} \widehat{P}_{h_1}^{[i-(i_1+1)]}(\widehat{P}_{h_1} - P_{h_1})P_{h_1}^{[i_1]}. \end{aligned}$$

Then, we can upper bound the remainder for $f \in \mathcal{C}_{\mathrm{pol}}^{k+24,L}(\mathbb{R} \times \mathbb{R}_+)$ by

$$\|Rf\|_{k,L+6} \le C^3 n^2 \|f\|_{k+12,L+3} h_2^3 + C^5 \frac{n(n-1)}{2} \|f\|_{k+24,L} (h_1^3)^2 \le \tilde{C} \|f\|_{k+24,L} \left(\frac{T}{n}\right)^4,$$

where we have used the equation (2.50) twice and the equation (2.49) once for the first sum, and the equation (2.50) three times and the equation (2.49) twice for the second one.

A.3. Malliavin Calculus

Proof of Lemma 3.2.7

WLOG, we may assume that $\|h\| = 1$ and assume that X is given as

$$X = f(W(e_1), \ldots, W(e_n))$$

where $f \in C_{\mathrm{pol}}^{\infty}(\mathbb{R}^n)$ and $e_1, \ldots, e_n$ is an o.n.b. with $e_1 = h$. Thus, the vector $W(e_1), \ldots, W(e_n)$ has standard normal distribution on $\mathbb{R}^n$. Note that in this case $\langle DX, h \rangle = \partial_1 f(W(e_1), \ldots, W(e_n))$.

We denote by ρ the density of the standard normal distribution, i.e.

$$\rho(x) = (2\pi)^{-\frac{n}{2}} \exp\left(-\frac{1}{2}\sum_{j=1}^{n} x_j^2\right).$$

Using integration by parts

$$\begin{aligned}\mathbb{E}(\langle \mathbf{D}X, h\rangle) &= \int_{\mathbb{R}^n} (\partial_1 f(x))\rho(x)dx \\ &= -\int_{\mathbb{R}^n} f(x)\rho(x)(-x_1)dx = \mathbb{E}(XW(h)).\end{aligned}$$

Proof of Proposition 3.2.9

Let X_n be a sequence in $\mathcal{S}$ which converges to 0 in $L^p(\Omega)$ and such that $\mathbf{D}X_n$ converges to ξ in $L^p(\Omega; H)$. We need to show that $\xi = 0$.

To that end, let $h \in H$ and $Y \in \mathcal{S}_b$ be given such that $YW(h)$ is bounded. By Corollary 3.2.8, we have

$$\begin{aligned}\mathbb{E}(Y\langle \xi, h\rangle) &= \lim \mathbb{E}(Y\langle \mathbf{D}X_n, h\rangle) \\ &= \lim \mathbb{E}\big(YX_nW(h) - X_n\langle \mathbf{D}Y, h\rangle\big) = 0\end{aligned}$$

since $X_n \to 0$ in $L^p(\Omega)$ and $YW(h)$ and $\langle \mathbf{D}Y, h\rangle$ are bounded. It follows that $\mathbb{E}\langle \xi, h\rangle = 0$. Since $h \in H$ was arbitrary, it follows that $\xi = 0$.

Proof of Proposition 3.2.16

If $X_j \in \mathcal{S}$ for $j = 1, \dots, m$ and $\varphi \in C^\infty(\mathbb{R}^m)$ with bounded partial derivatives, then equation (3.42) follows directly from the definition and the classical chain rule.

For the general case, pick sequences $X_j^{(n)}$ in $\mathcal{S}$ with $X_j^{(n)} \to X_j$ in $\mathbb{D}^{1,p}$ for $j = 1, \dots, m$ and a sequence $\varphi_n \in C^\infty(\mathbb{R})$ such that $\varphi_n \to \varphi$ uniformly on compact sets, $\nabla\varphi_n \to \nabla\varphi$ uniformly on compact sets and $|\varphi_n| \le C(1 + |x|)$ for a certain constant C.

It then follows from dominated convergence that $\varphi_n(X^{(n)}) \to \varphi(X)$ in $L^p(\Omega)$ and that $\mathbf{D}\varphi_n(X^{(n)}) = \sum_{j=1}^m \partial_j\varphi_n(X^{(n)})\mathbf{D}X_j^{(n)}$ converges to $\sum_{j=1}^m \partial_j\varphi(X)\mathbf{D}X_j$ in $L^p(\Omega; H)$. By closedness of the Malliavin derivative, it follows that $\varphi(X) \in \mathbb{D}^{1,p}$ and that (3.42) holds.

Proof of Lemma 3.2.17

The sequence X_n, being convergent in L^p is bounded in L^p. It thus follows that X_n is bounded in $\mathbb{D}^{1,p}$. Since bounded subsets of reflexive spaces are relatively weakly compact, we can extract a subsequence X_{n_k} which converges weakly in $\mathbb{D}^{1,p}$ to some $Y \in \mathbb{D}^{1,p}$. In particular, X_{n_k} converges weakly in L^p to Y, whence $Y = X$. It follows that $X \in \mathbb{D}^{1,p}$.

By subsequence convergence theorem, since we can apply the above to any subsequence to see that every subsequence of X_n has a subsequence converging to X weakly in $\mathbb{D}^{1,p}$. It then follows that X_n converges weakly to X in $\mathbb{D}^{1,p}$. In particular, $\mathbf{D}X_n \to \mathbf{D}X$ weakly in $L^p(\Omega; H)$.

Bibliography

[AB21] Aurélien Alfonsi and Vladislav Bally. A generic construction for high order approximation schemes of semigroups using random grids. *Numerische Mathematik*, 148(4):743–793, 2021.

[AJBB24] Eduardo Abi Jaber, Christian Bayer, and Simon Breneis. State spaces of multifactor approximations of nonnegative volterra processes. *SSRN Electronic Journal*, December 2024. Available at SSRN: https://ssrn.com/abstract=5068765.

[AJCP+24] Eduardo Abi Jaber, Christa Cuchiero, Luca Pelizzari, Sergio Pulido, and Sara Svaluto-Ferro. Polynomial volterra processes. *Electronic Journal of Probability*, 29:1–37, 2024.

[AJDC23] Eduardo Abi Jaber and Nathan De Carvalho. Reconciling rough volatility with jumps, 2023. Available at SSRN: 4387574.

[AJEE19a] E. Abi Jaber and O. El Euch. Markovian structure of the volterra heston model. *Statistics Probability Letters*, 149:63–72, 2019.

[AJEE19b] Eduardo Abi Jaber and Omar El Euch. Multifactor approximation of rough volatility models. *SIAM Journal on Financial Mathematics*, 10(2):309–349, 2019.

[AJLP19] Eduardo Abi Jaber, Martin Larsson, and Sergio Pulido. Affine volterra processes. *The Annals of Applied Probability*, 29(5):3155–3200, 2019.

[AK21] Aurélien Alfonsi and Ahmed Kebaier. Approximation of stochastic volterra equations with kernels of completely monotone type. *arXiv preprint arXiv:2102.13505*, 2021.

[AL25] Aurélien Alfonsi and Edoardo Lombardo. High order approximations and simulation schemes for the log-heston process. *SIAM Journal on Financial Mathematics*, 16(2):516–544, 2025.

[Alf10a] Aurélien Alfonsi. High order discretization schemes for the cir process: application to affine term structure and heston models. *Mathematics of Computation*, 79(269):209–237, 2010.

[Alf10b] Aurélien Alfonsi. High order discretization schemes for the cir process: Application to affine term structure and heston models. *Mathematics of Computation*, 79(269):209–237, 2010. Accessed October 9, 2025.

[BB23a] Christian Bayer and Simon Breneis. Efficient option pricing in the rough heston model using weak simulation schemes, 2023.

[BB23b] Christian Bayer and Simon Breneis. Markovian approximations of stochastic volterra equations with the fractional kernel. *Quantitative Finance*, 23(1):53–70, 2023.

[BB23c] Christian Bayer and Simon Breneis. Weak markovian approximations of rough heston, 2023.

[BCT21] Marco Briani, Lucia Caramellino, and Giorgio Terenzi. Convergence rate of markov chains and hybrid numerical schemes to jump-diffusion with application to the bates model. *SIAM Journal on Numerical Analysis*, 59(1):477–502, 2021.

[BDB17] Pierre Blanc, Jonathan Donier, and Jean-Philippe Bouchaud. Quadratic hawkes processes for financial prices. *Quantitative Finance*, 17(2):171–188, 2017.

[BFF+23] Christian Bayer, Peter K. Friz, Masaaki Fukasawa, Jim Gatheral, Antoine Jacquier, and Mathieu Rosenbaum. *Rough Volatility.* Society for Industrial and Applied Mathematics, 2023. ISBN: 978-1-61197-777-6.

[BJP25] Ofelia Bonesini, Antoine Jacquier, and Alexandre Pannier. Rough volatility, path-dependent pdes and weak rates of convergence, 2025.

[BM80] M. A. Berger and V. J. Mizel. Volterra equations with itô integrals—i. *The Journal of Integral Equations*, pages 187–245, 1980.

[BMM25] Mireille Bossy, Kerlyns Martinez, and Paul Maurer. Weak rough kernel comparison via ppdes for integrated volterra processes, 2025.

[CC98] Laure Coutin and Philippe Carmona. Fractional brownian motion and the markov property. *Electronic Communications in Probability*, 3:1–12, 1998.

[DJR21] Aditi Dandapani, Paul Jusselin, and Mathieu Rosenbaum. From quadratic hawkes processes to super-heston rough volatility models with zumbach effect. *Quantitative Finance*, 21(8):1235–1247, 2021.

[EGP10] Ernst Eberlein, Kathrin Glau, and Antonis Papapantoleon. Analysis of fourier transform valuation formulas and applications. *Applied Mathematical Finance*, 17(3):211–240, 2010.

[ER18] Omar El Euch and Mathieu Rosenbaum. Perfect hedging in rough heston models. *The Annals of Applied Probability*, 28(6):3813–3856, 2018.

[ER19] Omar El Euch and Mathieu Rosenbaum. The characteristic function of rough heston models. *Mathematical Finance*, 29(1):3–38, 2019.

[FL16] Damir Filipović and Martin Larsson. Polynomial diffusions and applications in finance. *Finance and Stochastics*, 20(4):931–972, 2016.

[Gat22] Jim Gatheral. Efficient simulation of affine forward variance models. *Risk*, February 2022.

[GJR18] Jim Gatheral, Thibault Jaisson, and Mathieu Rosenbaum. Volatility is rough. *Quantitative Finance*, 18(6):933–949, 2018.

[GJR20] Jim Gatheral, Paul Jusselin, and Mathieu Rosenbaum. The quadratic rough heston model and the joint s&p 500/vix smile calibration problem. *SSRN Electronic Journal*, January 2020. Available at SSRN: https://ssrn.com/abstract=3514894.

[GLS90] Gustaf Gripenberg, Stig-Olof Londen, and Olof Staffans. *Volterra Integral and Functional Equations*, volume 34 of *Encyclopedia of Mathematics and its Applications*. Cambridge University Press, Cambridge, 1990.

[Guy14] Jim Guyon. Path-dependent volatility. https://ieor.columbia.edu/files/seas/content/docs/pathdepvol_slides_columbia_ieor_17sep2018.pdf, 2014. Lecture slides, Columbia University.

[Guy20] Julien Guyon. The joint s&p 500/vix smile calibration puzzle solved. *Risk*, April 2020.

[Hai21] Martin Hairer. Introduction to malliavin calculus, 2021. Lecture notes.

[Hes93] Steven L. Heston. A closed-form solution for options with stochastic volatility with applications to bond and currency options. *The Review of Financial Studies*, 6(2):327–343, 1993.

[HR98] D. G. Hobson and L. C. G. Rogers. Complete models with stochastic volatility. *Mathematical Finance*, 8(1):27–48, 1998.

[JO19] Antoine Jacquier and Mohamed Oumgari. Deep ppdes for rough local stochastic volatility, 2019. Available at SSRN: 3400035.

[JR20] Paul Jusselin and Mathieu Rosenbaum. No-arbitrage implies power-law market impact and rough volatility. *Mathematical Finance*, 30(4):1309–1336, 2020.

[KP92] Peter E. Kloeden and Eckhard Platen. *Numerical Solution of Stochastic Differential Equations*, volume 23 of *Stochastic Modelling and Applied Probability*. Springer Berlin Heidelberg, Berlin, Heidelberg, 1992.

[KS12] Ioannis Karatzas and Steven E. Shreve. *Brownian Motion and Stochastic Calculus*, volume 113 of *Graduate Texts in Mathematics*. Springer-Verlag, New York, 2012.

[Kun13] Markus Kunze. An introduction to malliavin calculus. 2013.

[L'E09] Pierre L'Ecuyer. Quasi-monte carlo methods with applications in finance. *Finance and Stochastics*, 13:307–349, 2009.

[LM20] G. Lileika and V. Mackevičius. Weak approximation of ckls and cev processes by discrete random variables. *Lithuanian Mathematical Journal*, 60:208–224, 2020.

[MQ02] Robert I. McLachlan and G. Reinout W. Quispel. Splitting methods. *Acta Numerica*, 11:341–434, 2002.

[PS23] David J. Prömel and Daniel Scheffels. Stochastic volterra equations with hölder diffusion coefficients. *Stochastic Processes and their Applications*, 161:291–315, 2023.

[RZ22] Mathieu Rosenbaum and Jianfei Zhang. Deep calibration of the quadratic rough heston model. *Risk.net*, September 2022.

[Sha90] Alexander Shapiro. On concepts of directional differentiability. *Journal of Optimization Theory and Applications*, 66(3):477–487, September 1990.

[Tal86] Denis Talay. Discrétisation d'une équation différentielle stochastique et calcul approché d'espérances de fonctionnelles de la solution. *ESAIM: Mathematical Modelling and Numerical Analysis - Modélisation Mathématique et Analyse Numérique*, 20(1):141–179, 1986.

[TT90] Denis Talay and Lucio Tubaro. Expansion of the global error for numerical schemes solving stochastic differential equations. *Stochastic Analysis and Applications*, 8(4):483–509, 1990.

[Wan08] Zhenyu Wang. Existence and uniqueness of solutions to stochastic volterra equations with singular kernels and non-lipschitz coefficients. *Statistics Probability Letters*, 78(9):1062–1071, 2008.

[Zum09] Gilles Zumbach. Time reversal invariance in finance. *Quantitative Finance*, 9(5):505–515, 2009.

[Zum10] Gilles Zumbach. Volatility conditional on price trends. *Quantitative Finance*, 10(4):431–442, 2010.