Limit theorems for non-Markovian and fractional processes

a thesis presented for the degree of
Doctor of Philosophy of Imperial College London
and the Diploma of Imperial College
by

Alexandre Pannier

Department of Mathematics
Imperial College London

September 2021

Imperial College
London

Candidate: Alexandre Pannier

Title: *Limit theorems for non-Markovian and fractional processes*

Submitted for the degree of Ph.D.

Department of Mathematics, Imperial College London

Supervisor: Dr. Antoine Jacquier

Examiners: Prof. Johannes Muhle-Karbe and Prof. Josef Teichmann

Submission date: September 10, 2021

Doctoral studies funded by EPSRC CDT in Financial Computing and Analytics

Author email: aqp17@ic.ac.uk

This dissertation was typeset in LaTeX.

Statement of originality

I, Alexandre Pannier, certify that this thesis, and the research it contains, are the product of my own work, and that any ideas or quotations from the work of other people, published or otherwise, are fully acknowledged in accordance with the standard referencing practices of the discipline.

Signature ....

Limit theorems for non-Markovian and fractional processes

Abstract

This thesis examines various non-Markovian and fractional processes—rough volatility models, stochastic Volterra equations, Wiener chaos expansions—through the prism of asymptotic analysis.

Stochastic Volterra systems serve as a conducive framework encompassing most rough volatility models used in mathematical finance. In Chapter 2, we provide a unified treatment of pathwise large and moderate deviations principles for a general class of multidimensional stochastic Volterra equations with singular kernels, not necessarily of convolution form. Our methodology is based on the weak convergence approach by Budhiraja, Dupuis and Ellis [46, 88].

This powerful approach also enables us to investigate the pathwise large deviations of families of white noise functionals characterised by their Wiener chaos expansion as $X^\varepsilon = \sum_{n=0}^{\infty} \varepsilon^n I_n\big(f_n^\varepsilon\big)$. In Chapter 3, we provide sufficient conditions for the large deviations principle to hold in path space, thereby refreshing a problem left open in [213]. Hinging on analysis on Wiener space, the proof involves describing, controlling and identifying the limit of perturbed multiple stochastic integrals.

In Chapter 4, we come back to mathematical finance via the route of Malliavin calculus. We present explicit small-time formulae for the at-the-money implied volatility, skew and curvature in a large class of models, including rough volatility models and their multi-factor versions. Our general setup encompasses both European options on a stock and VIX options. In particular, we develop a detailed analysis of the two-factor rough Bergomi model.

Finally, in Chapter 5, we consider the large-time behaviour of affine stochastic Volterra equations, an under-developed area in the absence of Markovianity. We leverage on a measure-valued Markovian lift introduced by Cuchiero and Teichmann [73] and the associated notion of generalised Feller property. This setting allows us to prove the existence of an invariant measure for the lift and hence of a stationary distribution for the affine Volterra process, featuring in the rough Heston model.

This page was intentionally left blank.

Acknowledgements

One may think it is with great pride that I hand over this manuscript. The truth is I am more humbled than ever. I feel like this work emerges from the joint effort of a huge number of people, the unconditional kindness of a subset of them, and a significant amount of luck. I am tremendously thankful for everyone who supported me during this incredible experience, both mathematically and emotionally. And I apologise in advance to whomever I will unintentionally forget in the following lines.

The transition from the competitive mindset of academic studies to the understanding that mathematical research is a collaborative effort took place with its share of anxiety. Still, when it was tough, when I felt like an impostor, like I was failing and not up to the task, I reminded myself of the preciousness of this opportunity. The opportunity to have a shot at a fulfilling job, in an exciting city, surrounded by people I can invariably rely on.

I am infinitely grateful for the perfectly balanced supervision of Antoine (Jack) Jacquier. His passion inspired me to explore the subtle harmony of mathematical rigour and to read the books sitting on his shelves. He always adopted a calm and joyful attitude, even when reluctant problems were fighting back. But most importantly, he put his trust in me when I did not even believe in myself and, for the past three years, I knew I could always count on him. Needless to say I am impatient to continue learning at his side.

I would like to sincerely thank Johannes Muhle-Karbe and Josef Teichmann for their interest in my work, for taking the time to read this thesis and be the examiners of my viva.

I thank my collaborators, Aitor Muguruza and Kostas Spiliopoulos, from whom I have learned a lot over stimulating discussions, as well as my PhD companions Alessandro, Chloé, Claudio, Emilio, Henry, Marco, Žan, who were loyal lunch buddies and a priceless support during this journey. I repeat my appreciation for all the conference organisers who allowed me to present my research in front of experts in the field. It is a sensible place to thank the teachers who shared their passion for mathematics and inspired me to choose this path: Mme Postic, Wing-Sum Cheung, Siu Pang Yung, François Simenhaus, Jimmy Lamboley. I express my profound appreciation to Richard Craster and Richard Martin who introduced me to research and gave me the opportunity to come to Imperial College in the first place. I did not forget that Anderson Santos showed care and provided support during difficult times; I extend my gratitude to him.

Terence Tao's blog and Cédric Villani's books helped me understand with wit what the mind of a mathematician looks like. Mathoverflow and tex.stackexchange have always been there in times of struggle. I also acknowledge that I could not have finished this thesis without the assistance of Teams and Zoom, and a certain amount of tea.

On a more personal note, I would like to deeply thank Silvia and Luca for enduring me and whom I miss, as well as Guillaume, without whom London is not really the same.

For teaching me the most important lessons, there are no words to express the extent of my gratitude for Lucie.

Every day, I thank the family I chose, Antoine, Elliot, Kévin, Laurent, Océane, Robin, Théo B., Théo E., Valentine, for being part of my life, for making me laugh and making me love. I also thank Vincent, Nesrine; Baldé, Mickey and all the Phinedo Team; and Antoine R., Gabin, Mathieu, Thomas and the rest of the TDS for bearing with me through pain and laughter.

A special expression of my gratitude goes to the IC Buddhist Society and the BKUK Sangha, in particular Karuna, Thomas, Clément, Nitima, and Pratistha.

For your patience and self-compassion, I would like to thank the person writing this. We had a lot of fun.

To my family, and especially my parents, I am indebted for raising me. I am forever grateful that they always believed in me, supported me, fought for my happiness, and gracefully let me go.

Thank you, reader, for taking the time to read these words. I hope you enjoy reading this thesis as much as I enjoyed writing it.

We have to allow ourselves to realize that we are complete fools; otherwise, we have nowhere to begin.

Chögyam Trungpa

La logique mène à tout, à condition d'en sortir.

Alphonse Allais

Contents

List of Figures

Notations

$\mathbb{N}$	The integers $\{1, 2, \cdots\}$
$[\![N_1, N_2]\!]$	The integers $\{N_1, \cdots, N_2\}$, for $N_1 < N_2 \in \mathbb{N}$
$\mathbb{R}_+$	$[0, +\infty)$
$\overline{\mathbb{R}}_+$	$[0, +\infty]$
$\mathbb{T}$	A subset of $\mathbb{R}_+$
$\mathcal{C}(\boldsymbol{X}, \boldsymbol{Y})$	The space of $\boldsymbol{Y}$-valued continuous functions on $\boldsymbol{X}$, for two given spaces $\boldsymbol{X}$, $\boldsymbol{Y}$
$\mathcal{C}_b(\boldsymbol{X}, \boldsymbol{Y})$	The space of bounded functions of $\mathcal{C}(\boldsymbol{X}, \boldsymbol{Y})$
$\mathcal{C}(\boldsymbol{X})$	$\mathcal{C}(\boldsymbol{X}, \mathbb{R})$
$L^p(\boldsymbol{X}, \boldsymbol{Y})$	The space of $\boldsymbol{Y}$-valued p-integrable functions on $\boldsymbol{X}$, for $p \geq 1$
$L^p(\boldsymbol{X})$	$L^p(\boldsymbol{X}, \mathbb{R})$

The norms and filtered probability spaces will be defined in each chapter.

By convention, the infimum over an empty set is equal to $+\infty$.

1

Introduction

What better place than the first sentence to remind the reader that a piece of work is never finished? A thesis, once submitted, may seem permanent. A theorem, once proven, can appear as eternal. Yet, a creation of any sort always depends on the observer, the roots from which it stems, and the further sprouts it gives rise to. In such a dynamic and relational world, if it is set in stone then it is dead, claimed Pablo Picasso:

To finish a work? To finish a painting? What nonsense! To finish it means to be through with it, to kill it, to rid it of its soul.

That being said, let us go for a gentle walk across the landscape relevant to this thesis.

1.1 Laplace, Markov, and Mandelbrot walk into a pub

According to Pierre-Simon de Laplace, a computer vast enough to know all the information about the present state of the universe and to analyse it could deduce with absolute certainty everything there is to know about the past and the future [174].

This thought experiment served as an argument in favour of determinism and remained influential over the centuries. The caveat is that such a hypothetical algorithm—later dubbed “Laplace’s demon”—lies far beyond the limits of human abilities. The chaotic behaviour of dynamical systems uncovered by Poincaré bears witness to it [216, 217]. Nevertheless, scientists and mathematicians have not given up on their dream to describe reality; they just take a more pragmatic approach and consider ‘models’ which are skilful simplifications of Laplace’s demon.

Skilful here means finding the right balance between consistency—with the features of reality one wants to capture, and tractability—the capacity to analyse and get outputs. The demon is at the far end of the spectrum: perfect representation of reality but not a clue how it works. It should also be noted that the notion of right balance itself varies with the problem at hand and the tools one possesses to tackle it.

When observing the evolution of natural phenomena, a widely accepted simplification consists in accounting for the apparent randomness, and in incorporating this aspect in the model. Such a random dynamical model is called a *stochastic process* [85]. This is a slap in the face of Laplace’s determinism, but it allows to abstract away a huge chunk of the necessary data and it is also consistent with our point of view: we may have beliefs about future events but we cannot know for sure. A famous historical example is due to the botanist Robert Brown who described the fluctuations of pollen in water in a random fashion, giving rise to the Brownian motion (BM) [42]. It is today central in the theory of continuous stochastic processes [166]—where it is also known as the Wiener process, and in mathematical finance [226]—where Louis Bachelier introduced it to model equity prices [12].

A more subtle feature of Laplace’s demon is the supposed independence between the past and the future. In other words, the probability of a future event depends only on the present state; in probability theory such a model is called a *Markovian process*, after Andrey Markov. Because of their tractability and elegant properties, Markovian models have become omnipresent in virtually every field of applied mathematics, from thermodynamics to finance, passing through statistical mechanics, chemistry, biology, computer science, and queuing theory [209, 237]. However, unless one is a goldfish or has spent too much time at the pub, memory has a crucial influence on one’s present behaviour, and we observe in many systems that past causes and conditions participate in determining their future evolution. A further, less widely accepted, modification thus consists in incorporating memory in the dynamics of a model. Once again, it is a pragmatic choice for two reasons. As we just discussed, it coincides with our perception of reality. And even though it complicates the dynamics, it actually reduces the amount of information necessary at a given instant for the model to be consistent.

Inspired by the interdependence between distant samples and the fractal behaviour measured in natural phenomena, Mandelbrot and Van Ness introduced a generalisation of the Brownian motion which they named *fractional Brownian motion (fBM)* [185]. With this example, one shifts the balance towards consistency as the ever-growing width of applications testifies [34, Preface]. It was proved by Decreusefond and Üstünel [80] that, if $(W_t)_{t\geq 0}$ is a BM, there exists a deterministic kernel $K : \mathbb{R}_+^2 \to \mathbb{R}_+$ such that $(W^H)_{t\geq 0}$ is an fBM with Hurst exponent $H \in (0,1)$ where

$$W_t^H := \int_0^t K^H(t,s)\,\mathrm{d}W_s. \tag{1.1.1}$$

If $H = 1/2$ then $K^H \equiv 1$ and one recovers the standard BM. The realisations of the fBM have almost sure Hölder regularity of $H - \varepsilon$ for all $\varepsilon > 0$, hence the lower H is the more erratic the paths look like, as Figure 1.1 displays.

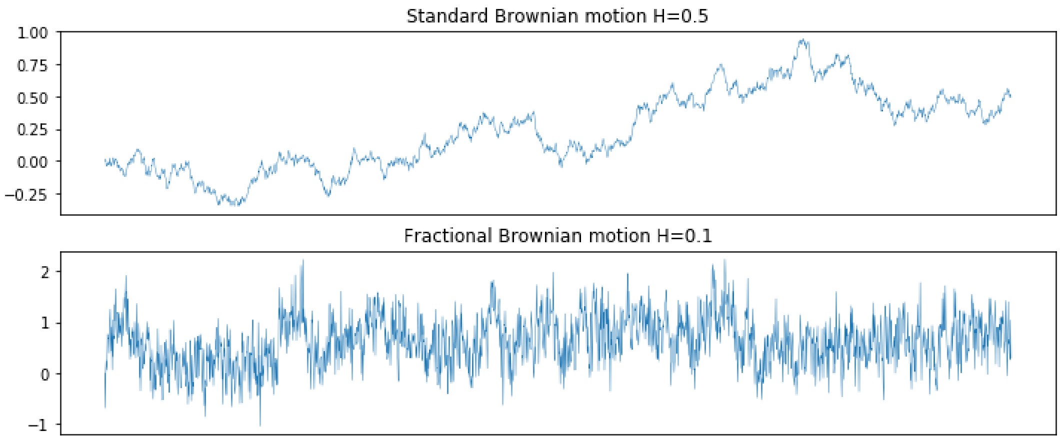

Figure 1.1: Simulated paths of standard and fractional Brownian motions

When $H \neq 1/2$, it is not difficult to prove that W^H is neither Markovian nor a semi-martingale [221]. The cost to pay in tractability for these more realistic dynamics is quite high since many tools were developed to suit the previous generation of models. The lack of Markovianity implies the loss of Feynman-Kac type formulas, which describe expectations of functionals of Markovian processes as the solution to a Partial Differential Equation (PDE). It also greatly increases the computational time of Euler schemes, and therefore restricts the efficiency of Monte-Carlo methods. In quantitative finance, these two approaches prevail when one is concerned with pricing options [238]. On the other hand, semimartingality is a requirement for Itô's calculus which itself lies at the core of classical stochastic analysis [222].

Therefore, challenging the Markovian paradigm allows for more accurate representations of the reality and is also an impetus for the development of new mathematical theories and techniques. In particular, asymptotic methods provide efficient ways to understand the behaviour of these processes. Explicit expressions often become available when passing to the limit, and they can be used to compute quick approximations of the quantities of interest, gain insights on the impact of the model parameters, improve numerical schemes, or compare the asymptotic behaviour with Markovian models. In option pricing, these can lead to a more efficient model choice and faster calibration to the data. Recently, limit theorems describing the short-time behaviour of option prices in equity markets were essential in the shift from traditional stochastic volatility models to non-Markovian fractional models, and in understanding how to fit the model parameters (especially the Hurst exponent H) with observed data [8, 112, 113].

The stochastic calculus of variations initiated by Paul Malliavin [181] played a fundamental role in the proofs of [8] to represent the processes of interest in a conducive form for studying their convergence. This infinite-dimensional calculus (called Malliavin calculus) is suited to the analysis of random variables defined on Gaussian probability spaces [201]; like many asymptotic methods, it does not weaken when Markovianity is absent. In this thesis, we will combine the powerful insights of asymptotic analysis with the universal representations of Malliavin calculus to study non-Markovian processes. The Wiener chaos expansion, an abstract representation of a large classes of processes, is the object of Chapter 3, while in Chapter 4 we extend the framework of [8] to volatility options.

1.2 Three V's for motivations

1.2.1 V for Volatility

The quantitative finance industry supplies interesting problems to academics and they in turn contribute with practical tools and insights. This back and forth relationship acts as a safeguard against dogmatic approaches. Mathematical modelling in particular is often justified by one of two concrete motivations: (i) predicting future trends, or (ii) understanding and managing risks.

The former application requires to study historical prices with statistical tools. Option pricing is concerned with the latter; in this situation one does not make assumptions about the "true probability" of the future dynamics of the price of an asset, but rather proposes a model that is consistent with the other traded assets, such that no opportunity of arbitrage exists—discounted prices of traded assets are martingales, and risk can be virtually eliminated—a hedging strategy is available. In this probability world, called "risk-neutral", Black, Scholes, and Merton [36, 190] introduced a model where the price of the underlying asset $(S_t)_{t\geq 0}$ satisfies the stochastic differential equation (SDE)

$$\mathrm{d}S_t/S_t = \sigma\,\mathrm{d}W_t, \quad S_0 > 0. \tag{1.2.1}$$

Without loss of generality, we consider the interest rates to be zero and $S_0 = 1$. This model, where the positive constant σ represents the volatility, remains the universal benchmark. Indeed, today's option prices are quoted in terms of the *implied volatility*, a quantity defined as the number σ one has to input in equation (1.2.1) such that the model and market prices coincide. For time to maturity T and log-strike price k, we denote the Black-Scholes Call option price by $\mathrm{BS}(T, k, \sigma)$ and consider a market with observed prices $\mathrm{C}(T, k)$. The implied volatility is therefore the unique non-negative solution $\sigma_{\mathrm{BS}}(T, k)$ to

$$\mathrm{BS}(T, k, \sigma_{\mathrm{BS}}(T, k)) = \mathrm{C}(T, k). \tag{1.2.2}$$

The solution map induced from quoted prices—the volatility surface—is clearly not constant across different values of T and k, hence it calls for more accurate models which capture the fluctuations of the implied volatility [121]. The local volatility model introduced by Dupire [87] replaces σ in (1.2.1) by a function $\sigma(t, S)$ of time and spot price. It is designed to calibrate perfectly the observed Call prices of all strikes and maturities at a fixed date, and therefore allows to price other derivatives consistently. The idea is reminiscent of Laplace's demon: an algorithm which takes as input all the available data and outputs a flawless representation of reality. In practice however, the necessary data is delicate to recover and the model needs re-calibration as time passes.

Paralleling the previous section, a natural evolution consists in integrating a new source of randomness and upgrading σ to a (Markovian) stochastic process $(\sigma_t)_{t\geq 0}$, often driven by a standard BM (although this progression fits nicely into our story, let us emphasise that it is not chronologically accurate). The aforementioned stochastic volatility models take different names for different dynamics of the volatility (SABR, Heston, Bergomi), but they all equally manage to fit the implied volatility smile at one given date—the map $k \mapsto \sigma_{\mathrm{BS}}(T, k)$ for a fixed $T > 0$. Hence, looking at a single smile does not give much information on the dynamics of the underlying. Instead, the term-structure of the at-the-money (ATM) implied volatility skew—the map $T \mapsto \left|\frac{\partial}{\partial k}\sigma_{\mathrm{BS}}(T, k)\right|$ evaluated at $k = 0$, is very sensitive to the dynamics. It obeys a power-law proportional to $T^{-\alpha}$ for some $\alpha > 0$, as Figure 1.2 illustrates, and is notoriously hard to reproduce under traditional stochastic volatility models, especially for small T [121, Chapter 3]. Additionally, many sources (e.g. [63, 66, 96]) argued that a consistent model should exhibit memory of the past. The time is ripe to let go of the Markovian assumption.

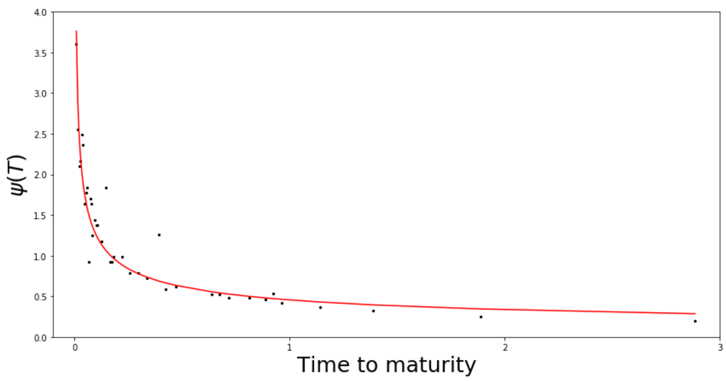

Figure 1.2: Term-structure of ATM implied volatility skew as of January 28, 2020 (black dots), plotted against the power-law fit $\psi(T) = 0.45\,T^{-0.44}$ (red curve).

Certainly inspired by the fBM's momentum in mathematical finance, the statistical analysis performed by [123] on high-frequency data can be considered as the last nail in the coffin of Markovian stochastic volatility models. With its provocative and self-explanatory title "Volatility is rough", the study shows that instantaneous log-volatility across a wide range of asset prices (complemented by [26]) exhibits a low Hölder regularity and is well modelled by an fBM with $H \approx 0.1$, as depicted in Figure 1.1. Other nails were hit by earlier analyses [63, 64, 65, 66], modelling the volatility as a function of an fBM with $H > 0.5$, a process with long memory. As alluded to in the previous section, the term-structure was investigated through asymptotic methods; the implied volatility is in general not available in closed form hence novel techniques were necessary. The short-time results of [8, 112, 113] demonstrated that fBM-driven models generated a term-structure proportional to $T^{H-\frac{1}{2}}$ as T goes to zero and were able to mimic the observed power-law with $H \approx 0.1$.

Before moving forward, we must give credit to Mandelbrot (and his co-authors) who was the first to predict the role the fBM could play in financial modelling [182, 183]. There is of course an emotional bias involved as the author of this thesis discovered the link between fractals and finance in one of his books [184].

In summary, this new class of *rough volatility* models improves the consistency with the data via (at least) three aspects:

1. The roughness of the time-series / historical prices / motivation (i) / Figure 1.1;
2. The short-time implied volatility skew / risk-neutral world / motivation (ii) / Figure 1.2;
3. The memory of the past.

Rough volatility models also have the potential to be consistent with other traded assets, thereby eliminating arbitrage opportunities. Traders appreciate volatility derivatives both for market insights—as a measure of market risk, and for hedging purposes—because of their negative correlation with equity prices. VIX derivatives in particular, which are designed to represent the volatility of the S&P index, became so liquid that they took a life of their own. The question of their joint modelling will be addressed in Chapter 4.

Arbitrage is however possible in a market with power-law volatility skew if the volatility model is *not* rough [114]. Building on the insights of [123], Bayer, Friz, and Gatheral [19] attempted to bridge the gap between historical and risk-neutral worlds by proposing a model consistent in both; they gave birth to the rough Bergomi model. At the core of [93, 165] lied the motivation of reconciling high- and low-frequency: El Euch, Fukasawa, Jaisson, and Rosenbaum built an asset price model based on market microstructure features and made the transition to longer term. A rough volatility model appeared in the long-time limit, which they dubbed the rough Heston model, providing a microstructural justification.

1.2.2 V for Volterra

The variance process $(\sigma_t^2)_{t\geq 0}$ in the rough Heston model satisfies an equation which can be seen as an extension of (1.1.1) with state-dependent drift and diffusion coefficients. We shall be chiefly interested in the class of (multidimensional) Stochastic Volterra Equations (SVEs) of the form

$$X_t = X_0 + \int_0^t K(t,s) b(X_s)\,\mathrm{d}s + \int_0^t K(t,s)\sigma(X_s)\,\mathrm{d}W_s, \quad t \geq 0, \tag{1.2.3}$$

which is a conducive framework to study rough volatility models. The key difference with Itô SDEs is the inclusion of a singular kernel K which blows up as s tends to t, just like the fBM's. Recall that, through the fBM case, we brought to light several aspects of non-Markovian and fractional processes—richer structure and loss of tractability, which extend beyond this particular example to SVEs, and serve as a motivation for this thesis. Figure 1.3 summarises the interplay between the different models we have presented so far. The processes in the second and third columns have strong rigorous mathematical links and sometimes overlap, while we insist that the first column only represents an informal analogy. The double arrows highlight a deep interdependence: new techniques allow for the emergence of new models, while the necessity for new models motivate the development of new techniques. It is often futile to try to determine which came first.

Interestingly enough, the original (deterministic) Volterra equations do not fit easily in this picture. The motivation came from population dynamics, where ordinary differential equations were used to model the evolution of the number of individuals alive. The observation that the latter depends on the state of the environment, which is itself a consequence of the past history of the population, led Vito Volterra to consider equations of the type (1.2.3), with $\sigma \equiv 0$ [180, Chapter 2]. While our story line suggests to include first randomness and then memory, he jumped directly to the latter.

The question of well-posedness of SVEs appeared rather early with [28, 29, 71, 210, 218] for Lipschitz coefficients, while pathwise uniqueness with non-Lipschitz coefficients turned out to be highly intricate without the support of Itô's calculus [195]. Abi Jaber, Larsson and Pulido [3] established powerful methods for SVEs with convolutional kernels, leveraging on deterministic resolvent analysis. The rough Heston model belongs to this class of affine Volterra processes, granting it weak existence and

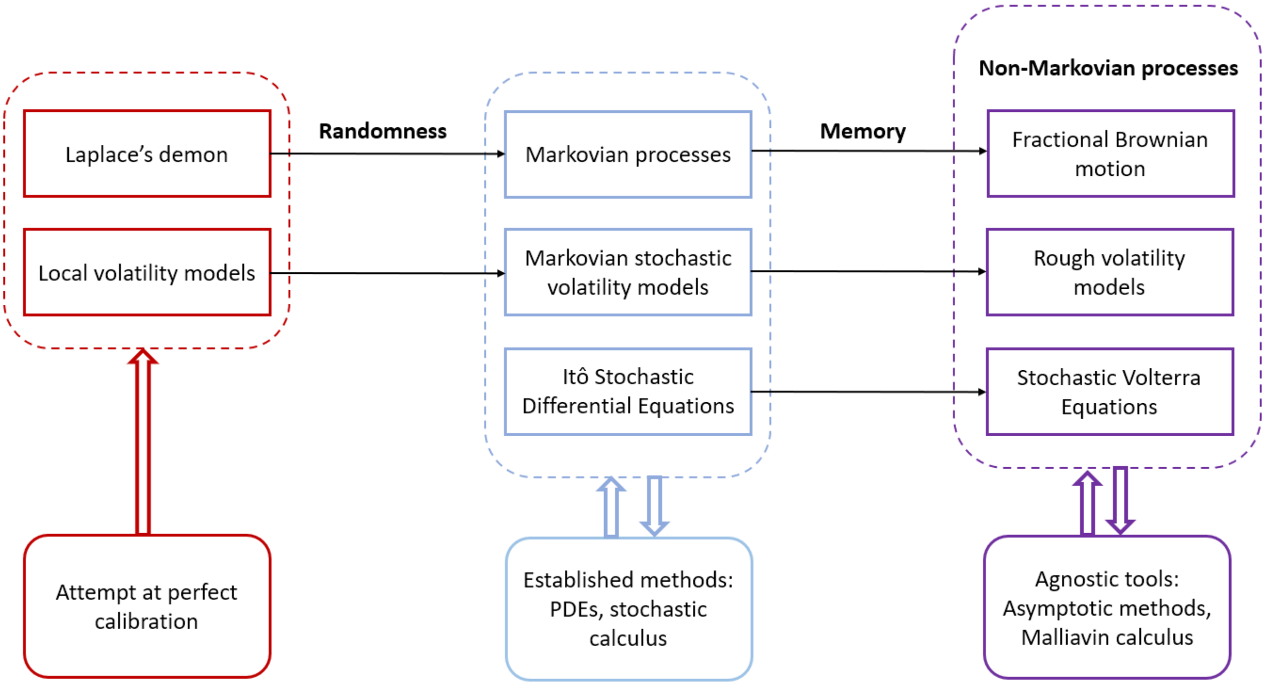

Figure 1.3: Informal summary of presented models

uniqueness, and a semi-closed form characteristic function. Like the classical Heston model in its time, the rough version proved to be a fertile soil for theoretical progresses [1, 2, 94, 95, 125, 131, 167].

An SVE-friendly fractional stochastic calculus has not emerged (yet?) to compensate the loss of semimartingality, but path-dependent PDEs now exist as an extension of the Feynman-Kac theory to SVEs [232]. We already mentioned that numerical methods suffer from the lack of Markovianity, hence a myriad of new techniques came to life, including but not limited to hybrid Monte-Carlo schemes [25, 188], Fourier methods [95], Donsker schemes [152], neural networks applications [23, 154, 160], multi-factor approximations [2, 59, 241], weak error rates [21], and quantization [37]. The rough volatility literature is expanding at a tremendous speed both to hasten their industrial use and to explore uncharted mathematical territories. The rough volatility network keeps an account of all the papers in the field [150].

As advertised in Figure 1.3, asymptotic analysis of rough volatility is particularly prolific and insightful. In addition to the aforementioned papers, this subfield includes vol-of-vol expansions [5], $H \downarrow 0$ convergences [22, 99, 100], small-time formulas based on Malliavin calculus methods [6], a rough SABR formula [115], Edgeworth expansions [92, 102], and in overwhelming majority large deviations results. Large Deviations Principles (LDPs) were first derived in finite-dimensional settings [18, 100, 103, 130, 136, 138, 139, 173], establishing the ground for higher-order approximations of the short-time implied volatility [20, 108, 109]. They provide insights on the role of the parameters and their accuracy competes with exact numerical methods, albeit with an almost instantaneous computation.

LDPs on path space are valuable when one is interested in the probability that the whole path of a stochastic process hits a given set. Such results grant a greater flexibility on the asymptotic regime, providing small-time, small-noise, or large-strike expressions, and yield the finite-dimensional version by a simple projection argument [56, 151, 161, 162]. They shall also lead to variance reduction methods for Monte-Carlo schemes, as in the Markovian case [90, 128, 133, 135, 194, 215, 220, 227]. The fractional processes we have been discussing are not remote from former Markovian models but generalise them, hence these limit theorems also allow for an efficient comparison between the two and for a measurement of the impact of the Hurst exponent ($H = 0.5$ versus $H < 0.5$). Let us now present a number of reasons for the popularity of large deviations techniques in mathematical finance.

1.2.3 V for Varadhan

The naked eye suffices to recognise that deviations from the mean are more abrupt in the second graph of Figure 1.1 than in the first. Furthermore, instantaneous volatility consistently features fatter tails than the Gaussian distribution, in other words the probability of an event with high impact is much higher than in Gaussian models. Classical limit theorems such as the Law of Large Numbers or the Central Limit Theorem, being focused on the most likely value or path, often miss these crucial aspects of modelling. Large deviations theory offers subtler insights by characterising the decay of the probability of rare events, hence assigning a quantitative interpretation to these unlikely scenarios.

Pinned by Varadhan in 1966 [229], the LDP is a universal property which describes the asymptotic behaviour of a family of random variables on an exponential scale through only two functions—the speed and the rate function. For the proper definition we refer to (2.1.2); in an informal way a family of $\mathcal{X}$-valued random variables $(X^\varepsilon)_{\varepsilon>0}$ satisfies an LDP with speed ε^{-1} and rate function $I : \mathcal{X} \to \overline{\mathbb{R}}_+$ if, for all Borel subsets $A \subset \mathcal{X}$,

$$\mathbb{P}(X^\varepsilon \in A) \overset{\varepsilon\downarrow 0}{\sim} \exp\left(-\inf_{x\in A} I(x)/\varepsilon\right). \tag{1.2.4}$$

To illustrate our earlier observation, let us notice that the family of standard and fractional Brownian motions $(W_\varepsilon)_{\varepsilon>0}$ and $(W^H_\varepsilon)_{\varepsilon>0}$ satisfy an LDP with speed ε^{-1} and ε^{-2H} respectively, meaning that in short-time the fBM with $H < 0.5$ allocates larger probabilities to large values.

The importance of large deviations is best explained by Varadhan himself [230]:

In computing expectations of random variables that can assume large values with small probabilities, contributions from such values cannot be ignored.

This resonates nicely with option pricing theory where one is concerned with computing expectations of the type $\mathbb{E}[(S_T - \mathrm{e}^k)^+]$, for time to maturity T, log-strike k, and $(x)^+ = \max(x, 0)$. Out-of-the-money options, where S_0 is much smaller than e^k, require an important divergence from the mean to be exercised, with the small probability $\mathbb{P}(S_T > \mathrm{e}^k)$. Nevertheless, the price of the option cannot be overlooked because multiplying a large value with a small probability can still yield a large expectation. For a concrete example, let S^M and S^R represent the spot prices in Markovian and rough volatility models, and satisfy a small-time LDP with rate functions I^M and I^R, respectively. Then one can extend the heuristic (1.2.4) and the remark underneath to option prices:

$$\begin{aligned}
&\mathbb{E}\left[(S^M_\varepsilon - \mathrm{e}^k)^+\right] \overset{\varepsilon\downarrow 0}{\sim} \exp\left(-I^M(k)/\varepsilon\right), &&\text{for all } k > \log(S^M_0);\\
&\mathbb{E}\left[(S^R_\varepsilon - \mathrm{e}^{k_\varepsilon})^+\right] \overset{\varepsilon\downarrow 0}{\sim} \exp\left(-I^R(k)/\varepsilon^{2H}\right), &&\text{for all } k > \log(S^R_0),\ k_\varepsilon := k\varepsilon^{1/2-H}.
\end{aligned}$$

We note the necessary rescaling of the log-strike in the rough case, but more on that later. We refer to [215] for an overview of large deviations theory in finance.

This thesis is mainly concerned with *pathwise* large deviations, where the state space $\mathcal{X}$ is a path space, and therefore deals with continuous stochastic processes. We owe the traditional theory to Friedlin and Wentzell [106] who decided to regard stochastic systems $(X^\varepsilon)_{\varepsilon>0}$ on $\mathcal{C}([0, T], \mathbb{R})$, for some fixed $T > 0$, as perturbations of deterministic ones $\overline{X}$, such as

$$X^\varepsilon_t = x_0 + \int_0^t b(X^\varepsilon_s)\,\mathrm{d}s + \varepsilon \int_0^t \sigma(X^\varepsilon_s)\,\mathrm{d}W_s, \qquad \overline{X}_t = x_0 + \int_0^t b(\overline{X}_s)\,\mathrm{d}s,$$

for all $\varepsilon > 0$, $x_0 \in \mathbb{R}$, $t \in [0,T]$, and with some regularity conditions on b and σ. From this perspective, the LDP is a way to measure the distance between the rescaled SDEs satisfied by $(X^\varepsilon)_{\varepsilon>0}$ and their deterministic limit $\overline{X}$. One may draw satisfaction from the universal form of the rate function in this pathwise setting. It always displays the infimum of an energy function over a range of functions which act as controls, or perturbations, of the limit equation. Typically, for all continuous paths ϕ, the rate function associated to $(X^\varepsilon)_{\varepsilon>0}$ reads

$$I^X(\phi) = \inf\left\{\frac{1}{2}\int_0^T v_t^2 \,\mathrm{d}t : v \in L^2([0,T],\mathbb{R}),\ \phi_t = x_0 + \int_0^t \big(b(\phi_s) + \sigma(\phi_s)v_s\big)\,\mathrm{d}s\right\}. \tag{1.2.5}$$

In the same paper, Varadhan [229] proved a central result in large deviations theory. If $(X^\varepsilon)_{\varepsilon>0}$ satisfies an LDP with rate function I^X then for all continuous bounded functions F the following limit holds

$$\lim_{\varepsilon\downarrow 0} -\varepsilon \log \mathbb{E}\left[\exp\left\{-\frac{F(X^\varepsilon)}{\varepsilon}\right\}\right] = \inf_{x\in\mathcal{X}} \left\{I^X(x) + F(x)\right\}. \tag{1.2.6}$$

We say in that case that $(X^\varepsilon)_{\varepsilon>0}$ satisfies the Laplace principle (yes, the same fellow who summoned a demon a few pages back). The reverse statement was later proved by Bryc [43]. Inspired by this equivalence, Dupuis and Ellis introduced a route to derive the Laplace principle directly [88]. As far as we are concerned, their fundamental argument is the variational representation proved by Boué and Dupuis [38]. Equipped with this tool, Budhiraja and Dupuis [45, 46] showed that the Laplace principle immediately follows from the convergence in law of a perturbed version of $(X^\varepsilon)_{\varepsilon>0}$, thereby trading the methods of Freidlin and Wentzell for the established theory of weak convergence [35].

Simply stated, they represented X^ε as the functional of the Brownian motion $\mathcal{G}^\varepsilon(W)$ and then studied the convergence of the perturbed processes $X^{\varepsilon,v^\varepsilon} := \mathcal{G}^\varepsilon\left(W + \varepsilon^{-1}\int_0^\cdot v_s^\varepsilon \,\mathrm{d}s\right)$, for some convergent sequence of processes $(v^\varepsilon)_{\varepsilon>0}$. The choice of scaling ε^{-1} is crucial and suggests the speed of the LDP. Heuristically, the form of the rate function can often be guessed early on, it resembles (1.2.5) where the limit equation corresponds to the original stochastic one with $\varepsilon \mathrm{d}W_t$ replaced by $v_t\,\mathrm{d}t$. The universality of this approach made it incredibly successful, as the monograph [46] and the references therein indicate. Chapters 2 and 3 present new applications of this method, to SVEs and Wiener chaos expansions respectively.

1.3 Outline of the contributions

This thesis lies in an interconnected web made of all the ideas presented above and of an enormous number of people. We think not only of the many mathematicians cited in this introduction, giants whose names are forever tied to their theorems, but also of the many who contributed and whose names are too often forgotten. This thesis could not have existed without them. Before we present our contributions in more details, the following picture (Figure 1.4) helps to put them into context. It highlights how the insights gathered in the first project are stretched into various directions, following curiosity and opportunities rather than linearity.

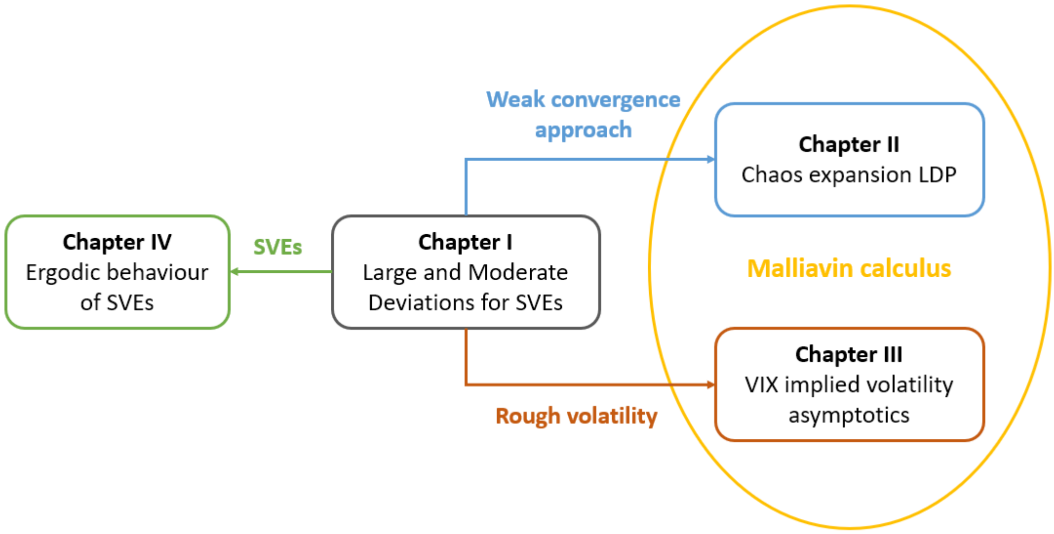

Figure 1.4: Relationships between the chapters

Each chapter corresponds to one project and one paper, and can therefore be read independently of the others. Respectively, Chapters 2, 3, 4, and 5 coincide with the papers [162, 206, 159, 163].

1.3.1 Large and moderate deviations for stochastic Volterra systems

This first chapter, based on the preprint [162], is the central piece of the thesis as it combines the three main aspects of rough volatility, SVEs, and large deviations. It aims at proving pathwise large and moderate deviations principles on the state space $\mathcal{C}([0,T],\mathbb{R}^d)$ for a broad class of rescaled SVEs that encompasses the most popular rough volatility models, where $T > 0$ is a fixed time horizon and $d \in \mathbb{N}$. We consider the family of processes $(X^\varepsilon)_{\varepsilon>0}$ defined as the solutions to the d-dimensional SVEs

$$X_t^\varepsilon = X_0^\varepsilon + \int_0^t K(t,s) b_\varepsilon(s, X_s^\varepsilon)\,\mathrm{d}s + \vartheta_\varepsilon \int_0^t K(t,s)\sigma_\varepsilon(s, X_s^\varepsilon)\,\mathrm{d}W_s, \quad t\in[0,T],$$

where $\vartheta_\varepsilon > 0$ tends to zero as ε goes to zero, and W is an m-dimensional Brownian motion. Moreover, K is a kernel that may be singular, not necessarily of convolution type, and the coefficients are such that a unique pathwise solution exists.

Moderate deviations are tightly connected to large deviations. If X^ε converges to $\overline{X}$, we say that $(X^\varepsilon)_{\varepsilon>0}$ satisfies a Moderate Deviations Principle (MDP) if $(\eta^\varepsilon)_{\varepsilon>0}$ satisfies an LDP with speed h_ε^2, where

$$\eta^\varepsilon := \frac{X^\varepsilon - \overline{X}}{\vartheta_\varepsilon h_\varepsilon}, \quad \text{for all } \varepsilon > 0,$$

with $\lim_{\varepsilon\downarrow 0} h_\varepsilon = +\infty$ and $\lim_{\varepsilon\downarrow 0}\vartheta_\varepsilon h_\varepsilon = 0$. By choosing h_ε appropriately, the MDP interpolates between the LDP regime (with $h_\varepsilon = \vartheta_\varepsilon$) and the central limit theorem regime (with $h_\varepsilon = 1$). Hence, moderate deviations are relevant to study less unlikely events than large deviations. Interestingly enough, the MDP rate function is often—if not always—much more tractable than the LDP one, paving the way for explicit asymptotic results and effective numerical methods.

The proof of the LDP, based on the weak convergence approach [46], consists in showing the convergence of the perturbed family $(X^{\varepsilon,v^\varepsilon})_{\varepsilon>0}$. Recall that if $X^\varepsilon = \mathcal{G}^\varepsilon(W)$ then $X^{\varepsilon,v^\varepsilon} := \mathcal{G}^\varepsilon\left(W + \vartheta_\varepsilon^{-1}\int_0^\cdot v_s^\varepsilon\,\mathrm{d}s\right)$ for all $\varepsilon > 0$, where $(v^\varepsilon)_{\varepsilon>0}$ is a convergent sequence of adapted processes taking values in $L^2([0,T],\mathbb{R}^m)$. They turn out to be the unique solutions to the following controlled SVEs

$$X_t^{\varepsilon,v^\varepsilon} = X_0^\varepsilon + \int_0^t K(t,s)\Big(b_\varepsilon(s, X_s^{\varepsilon,v^\varepsilon}) + \sigma_\varepsilon(s, X_s^{\varepsilon,v^\varepsilon})v_s^\varepsilon\Big)\,\mathrm{d}s + \vartheta_\varepsilon\int_0^t K(t,s)\sigma_\varepsilon(s, X_s^{\varepsilon,v^\varepsilon})\,\mathrm{d}W_s.$$

The tightness of this family in $\mathcal{C}[0,T],\mathbb{R}^d)$ is a consequence of Kolmogorov continuity theorem combined with Arzelá-Ascoli theorem. Keeping in mind our objective we slightly relax the linear growth assumption on the coefficients, traditionally used to apply Grönwall's lemma and obtain the moment bounds. This is necessary to include the rough Bergomi model which features exponential coefficients. Frameworks for deriving the Hölder continuity of SVEs can be found in [79] for non-convolution kernels and in [3] for convolution kernels. Leveraging on their insights, we obtain the tightness and then need to identify the limit in law.

As in the Freidlin-Wentzell theory (1.2.5), the latter turns out to be the solution to some deterministic equation. However, if uniqueness of this equation fails, then the proof of the LDP cannot go through. Since we do not impose Lipschitz continuity of the coefficients to accommodate for the rough Heston model, we end up facing this pitfall in the LDP regime. To circumvent it, we devise in Section 2.3.1 a modification of the original proof of [45, Theorem 4.4] which only requires the solutions of the limit equation to be asymptotically close to a class of equations with a unique solution—this is possible because the right-hand-side of the Laplace principle (1.2.6) is an infimum and thus allows for "an ε of a room". This relaxation applies in particular to equations with locally Lipschitz coefficients such as the square root function present in the rough Heston model. Passed these hurdles, we reach our first main result, combination of Theorems 2.3.11 and 2.3.29.

LDP for SVEs

Under mild assumptions on the coefficients, strong existence and uniqueness of (1.3.1), and an "almost-uniqueness" condition for the limit equation, the family $(X^\varepsilon)_{\varepsilon>0}$ satisfies an LDP with speed $\vartheta_\varepsilon^{-2}$ and rate function $I : \mathcal{C}([0,T],\mathbb{R}^d) \to \overline{\mathbb{R}}_+$ given by

$$I(\phi) = \inf\left\{\frac{1}{2}\int_0^T |v_s|^2 \,\mathrm{d}s : v \in L^2([0,T],\mathbb{R}^m),\right.$$
$$\left.\phi_t = x_0 + \int_0^t K(t,s)\Big(b(s,\phi_s) + \sigma(s,\phi_s)v_s\Big)\,\mathrm{d}s\right\}.$$

Let us recall that pathwise uniqueness with non-Lipschitz coefficients was achieved in [195], hence this result extends [240] by lifting two important assumptions: Lipschitz continuity and linear growth.

The proof of the MDP unfolds in an analogous way albeit with a different speed. Because of the rescaling of η^ε, the shifted BM now reads $W + h_\varepsilon \int_0^\cdot v_s \,\mathrm{d}s$. A more intricate SVE emerges which requires additional care since the rescaling may induce a blow-up as ε goes to zero, such as the following term which shall give ∇b in the limit

$$\frac{b_\varepsilon\big(s, \overline{X} + \vartheta_\varepsilon h_\varepsilon \eta_s^{\varepsilon,v^\varepsilon}\big) - b(s,\overline{X}_s)}{\vartheta_\varepsilon h_\varepsilon}.$$

This gives an insight on our second main result, which corresponds to Theorems 2.3.20 and 2.3.29.

MDP for SVEs

Under strong existence and uniqueness of (1.3.1) and slightly stronger assumptions on the coefficients, the family $(X^\varepsilon)_{\varepsilon>0}$ satisfies an MDP with speed h_ε^2 and rate function $\Lambda : \mathcal{C}([0,T],\mathbb{R}^d) \to \overline{\mathbb{R}}_+$ given by

$$\Lambda(\psi) = \inf\left\{\frac{1}{2}\int_0^T |v_s|^2\,\mathrm{d}s : v \in L^2([0,T],\mathbb{R}^m),\right.$$
$$\left.\psi_t = \int_0^t K(t,s)\Big(\nabla b(s,\overline{X}_s)\psi_s + \sigma(s,\overline{X}_s)v_s\Big)\,\mathrm{d}s\right\}.$$

Crucially, the limit equation is here linear and therefore uniqueness is granted. This simpler form also makes it more amenable for further computations, as advertised earlier.

Let us now consider a general rough volatility model of the form

$$\begin{cases} X_t &= -\dfrac{1}{2}\displaystyle\int_0^t \Sigma(Y_s)\,\mathrm{d}s + \int_0^t \sqrt{\Sigma(Y_s)}\mathrm{d}B_s, \\ Y_t &= y_0 + \displaystyle\int_0^t K(t-s)\mathfrak{b}(Y_s)\,\mathrm{d}s + \int_0^t (t-s)^{H-\frac{1}{2}}\zeta(Y_s)\mathrm{d}W_s, \end{cases}$$

where B and W are two Brownian motions with correlation $\rho \in [-1,1]$, the Hurst exponent $H \in (0,1)$ corresponds to the almost Hölder regularity of the variance Y, and the convolution kernel satisfies $K(\lambda t) = \lambda^{\varpi} t$ (we think of $\varpi = \frac{1}{2}$ to recover a smooth kernel or $\varpi = H$ to have a singular kernel in the drift as well). We need this property for the rescaled processes $X_t^\varepsilon = \varepsilon^{H-\frac{1}{2}}X_{\varepsilon t}$ and $Y_t^\varepsilon = Y_{\varepsilon t}$ to fit in the setting of (1.3.1) with $\vartheta_\varepsilon = \varepsilon^H$. The LDPs, gathered in Proposition 2.4.3, are then a corollary of our first main result.

LDPs for small-time rough volatility

Under mild assumptions on the coefficients, LDPs with speed ε^{-2H} hold for $(X^\varepsilon, Y^\varepsilon)$, X^ε, and Y^ε in a pathwise sense, and for $\varepsilon^{H-\frac{1}{2}}X_\varepsilon$ and Y_ε on the real line.

The rate function associated to $(X^\varepsilon, Y^\varepsilon)$ has an explicit integral form but the others involve infima over infinite-dimensional spaces and are thus not available in closed form, unlike in the MDP case, where $\eta^\varepsilon = \varepsilon^{\beta-H}\big(X^\varepsilon, Y^\varepsilon - y_0\big)$ for any $\beta \in (0,H)$. This is the content of Proposition 2.4.6.

MDPs for small-time rough volatility

Under slightly stronger assumptions on the coefficients, MDPs with speed $\varepsilon^{-2\beta}$ hold for $(X^\varepsilon, Y^\varepsilon)$, X^ε, and Y^ε in a pathwise sense, and for $\varepsilon^{H-\frac{1}{2}}X_\varepsilon$ and Y_ε on the real line.

All the rate functions are explicit, in particular for ϕ absolutely continuous and $x \in \mathbb{R}$,

$$\Lambda^X(\phi) = \frac{1}{2\Sigma(y_0)} \int_0^T \dot{\phi}_t^2 \,\mathrm{d}t \qquad \text{and} \qquad \Lambda_1^X(x) = \frac{x^2}{2\Sigma(y_0)},$$

associated to X^ε and $\varepsilon^{H-\frac{1}{2}}X_\varepsilon$ respectively, are independent of H and coincide with Markovian LDP rate functions. Nevertheless, the roughness (read H) still appears in the scaling of the process and the speed of the MDP.

These small-time asymptotic results apply in Section 2.3.2 to the rough Stein-Stein, (multi-factor) rough Bergomi, and rough Heston models (under the assumption that pathwise uniqueness holds for the latter).

We also propose to study tail asymptotics in Corollaries 2.4.15 and 2.4.17, which generally have the form $X^\varepsilon = \varepsilon X$, such that an LDP provides asymptotic estimates on $\mathbb{P}(X^\varepsilon \geq 1) = \mathbb{P}(X \geq \varepsilon^{-1})$.

Tail LDPs for rough volatility

In the rough Stein-Stein model, LDPs with speed ε^{-2} hold for the families $(\varepsilon^2 X, \varepsilon Y)$, $\varepsilon^2 X$ in a pathwise sense and $\varepsilon^2 X_t$ on the real line (for a fixed $t \in [0,T]$).

In the rough Heston model, LDPs with speed ε^{-2} hold for the families $(\varepsilon^2 X, \varepsilon^2 Y)$, $\varepsilon^2 X$ in a pathwise sense and $\varepsilon^2 X_t$ on the real line (for a fixed $t \in [0,T]$).

The small-time LDPs (respectively MDPs) for rough volatility models translate into implied volatility asymptotics for $\sigma_{\mathrm{BS}}(t, kt^{H-\frac{1}{2}})$ (respectively $\sigma_{\mathrm{BS}}(t, kt^{\beta-\frac{1}{2}})$) as t goes to zero, see Section 2.4.1.3. In the LDP case, the limit is cumbersome to compute while it is simply equal to $\Sigma(y_0)^2$ in the MDP case. Meanwhile, Corollary 2.4.18 shows the large-tail LDP yields a limit for $\sigma_{\mathrm{BS}}(t,k)^2 t/k$ as k goes to infinity.

We mentioned previously that one outlook of this project lies in variance reduction techniques for Monte-Carlo schemes. Importance sampling consists in simulating an unbiased estimator under a different probability measure and finding the optimal change of probability measure that minimises its variance, which is equivalent to the optimisation problem

$$\inf_{\mathbb{Q}\sim\mathbb{P}} \mathbb{E}^{\mathbb{P}}\left[G(X)^2\frac{d\mathbb{P}}{d\mathbb{Q}}\right]. \tag{1.3.1}$$

Simulating under this new measure can yield a significant gain when confronted to rare events with big impacts by increasing the probability of the paths which contribute the most to the expectation. However, the raw problem (1.3.1) cannot be solved explicitly in practice. Large and moderate deviations, by design, provide insight on events with low probability and were found useful to derive asymptotically optimal change of measure. Restricting the range of equivalent measures to the Cameron-Martin space, (1.3.1) can be approximated by the following proxy featured in the Laplace principle (1.2.6)

$$\mathbb{E}\left[\exp\left\{\frac{1}{\varepsilon}\Big(2\log(G(X^\varepsilon)) - \sqrt{\varepsilon}\int_0^T \dot{h}_t \mathrm{d}W_t + \frac{1}{2}\int_0^T |\dot{h}_t|^2\,\mathrm{d}t\Big)\right\}\right].$$

The minimisation problem then boils down to a simpler variational one, but unfortunately the gains of this approach are often mitigated by the cost of computing the intricate LDP rate function. This challenge encourages to use moderate deviations where the rate functions are explicit, as in [194].

1.3.2 Pathwise large deviations for white noise chaos expansions

The second chapter corresponds to the preprint [206] and takes a different angle to tackle non-Markovian systems. It exploits the same methods as the first chapter but applies them to a universal object that comes up in Malliavin calculus. Any Gaussian functional $F \in L^2(\Omega)$ can be represented via its Wiener chaos expansion

$$F = \mathbb{E}F + \sum_{n=1}^{\infty} I_n(f_n), \tag{1.3.2}$$

where the terms of this (orthogonal) decomposition are multiple Wiener–Itô integrals and the kernels $(f_n)_{n\geq 1}$ are deterministic and uniquely determined by F, see [201, Chapter 1]. A variety of limit theorems related to the Wiener chaos exist [202, 199, 204] but very few relate to large deviations. Pathwise large deviations were derived for the family $(\varepsilon^n I_n(f_n))_{\varepsilon>0}$ in [83, 175, 187]. However only one LDP deals with the whole expansion

$$X^{\varepsilon} = \sum_{n=0}^{\infty} \varepsilon^n I_n(f_n^{\varepsilon})$$

and it is limited both to the state space $\mathbb{R}$ and to the Brownian motion case [214]. We extend this result to path space and to a larger class of Gaussian processes by leveraging on the weak convergence approach to large deviations. Notice that chaos expansions are a universal representation, aloof from Markovianity considerations.

For tightness criteria to be available, we need to consider the space of continuous functions $\mathcal{C}(\mathrm{K}, \mathbb{R})$ where K is a compact subset of $\mathbb{R}^d$. As the name of this chapter suggests, we are interested in a white noise measure $\dot{W}$, an $L^2(\Omega)$-Gaussian measure on $(\mathbf{T}, \mathcal{T})$ based on the (atomless) product measure $\widehat{\mu} := \lambda \otimes \mu$. Here λ is the Lebesgue measure on $\mathbb{R}_+$, μ is a sigma-finite measure on a measurable space $(\boldsymbol{E}, \mathcal{E})$ and the parameter space is $\mathbf{T} := \mathbb{R}_+ \times \boldsymbol{E}$ with σ-algebra $\mathcal{T}$.

The white noise measure $\dot{W}$ has zero mean and takes independent values on any family of disjoint subsets of $\mathbf{T}$. It is more precisely characterised by its covariance structure. For all $A_1, A_2 \in \mathcal{T}$,

$$\mathbb{E}\Big[\dot{W}(A_1)\dot{W}(A_2)\Big] = \widehat{\mu}(A_1 \cap A_2).$$

Interestingly enough, one can recover well-known Gaussian processes as functionals of this white noise measure such as space-time white noise, isonormal processes, fractional Brownian motion, multidimensional and infinite-dimensional Brownian motions, and Q-Wiener processes.

Having defined our Gaussian measure, we can identify the multiple Itô integrals $I_n : L^2(\mathbf{T}^n) \to L^2(\Omega)$ as

$$I_n(f_n) = \int_{\mathbf{T}^n} f_n(\boldsymbol{t_n}, \boldsymbol{x_n})\, W(\mathrm{d}t_1\, \mathrm{d}x_1) \cdots W(\mathrm{d}t_n\, \mathrm{d}x_n).$$

Since the weak convergence approach requires to shift the Gaussian process we need to study such multiple integrals where $W(\mathrm{d}t\, \mathrm{d}x)$ is replaced by $W^{v/\varepsilon}(\mathrm{d}t\, \mathrm{d}x) := W(\mathrm{d}t\, \mathrm{d}x) + \frac{1}{\varepsilon}\int v(t,x)\widehat{\mu}(\mathrm{d}t\, \mathrm{d}x)$. The multiple integrals are not independent any longer and developing one in full generates no less than 2^n integrals. Section 3.4.2 provides a rigorous account; let us illustrate this informally here. First, notice that there is a combinatorial flavour in this development and it can be helpful to think of the binomial formula for $(x + y)^n$. Heuristically, if we suppose that all the integrals are well-defined, this translates as

$$I_n^{\varepsilon,v^\varepsilon}(f) := \int_{\mathbf{T}^n} f(\boldsymbol{t_n},\boldsymbol{x_n})\,\varepsilon W^{v^\varepsilon/\varepsilon}(\mathrm{d}t_1\,\mathrm{d}x_1)\cdots\varepsilon W^{v^\varepsilon/\varepsilon}(\mathrm{d}t_n\,\mathrm{d}x_n) = \sum_{k=0}^{n}\binom{n}{k} m_k^\varepsilon(f),$$

where m_k^ε is a multiple integral which integrates k times against $v^\varepsilon\,\mathrm{d}\widehat{\mu}$ and $n-k$ times against εW:

$$m_k^\varepsilon(f) = \int_{\mathbf{T}^n} f(\boldsymbol{t_n},\boldsymbol{x_n})\big(v^\varepsilon(t_1,x_1)\,\widehat{\mu}(\mathrm{d}t_1\,\mathrm{d}x_1)\big) \\ \cdots\big(v^\varepsilon(t_k,x_k)\,\widehat{\mu}(\mathrm{d}t_k\,\mathrm{d}x_k)\big)\,\varepsilon W(\mathrm{d}t_{k+1}\,\mathrm{d}x_{k+1})\cdots\varepsilon W(\mathrm{d}t_n\,\mathrm{d}x_n).$$

Suppose that v^ε converges weakly to v. One would then expect that, as $\varepsilon\downarrow 0$, if $k<n$ then m_k^ε goes to zero, and if $k=n$ then m_k^ε converges towards

$$J_n^v(f_n) := \int_{\mathbf{T}^n} f_n(\boldsymbol{t_n},\boldsymbol{x_n})\,\big(v(t_1,x_1)\,\widehat{\mu}(\mathrm{d}t_1\,\mathrm{d}x_1)\big)\cdots\big(v(t_n,x_n)\,\widehat{\mu}(\mathrm{d}t_n\,\mathrm{d}x_n)\big).$$

Hence, it boils down to one multiple integral for each $n\geq 0$.

We suppose that X^ε is a stochastic process with chaos expansion of the form

$$X^\varepsilon(z) = \mathcal{G}^\varepsilon(W)(z) = \sum_{n=0}^{\infty} I_n(f_n^{\varepsilon,z}),\qquad \text{for each } z\in\mathrm{K},$$

where $f_n^{\varepsilon,z}$ tends to f_n^z for each $n\geq 0$. Therefore, $\mathcal{G}^\varepsilon(W^{v^\varepsilon/\varepsilon})(z) = \sum_{n=0}^{\infty} I_n^{\varepsilon,v^\varepsilon}(f_n^{\varepsilon,z})$ shall naturally converge to $\sum_{n=0}^{\infty} J_n^v(f_n^z)$ as ε goes to zero. Although this series of arguments lack rigour, this is precisely what happens. It leads to our main result, Theorem 3.3.3.

LDP chaos expansion

Under deterministic assumptions on the coefficients $\{f_n^{\varepsilon,z} : n\geq 0, \varepsilon>0, z\in\mathrm{K}\}$, an LDP holds for $(X^\varepsilon)_{\varepsilon>0}$ with speed ε^{-2} and rate function $\Lambda : \mathcal{C}(\mathrm{K})\to\overline{\mathbb{R}}_+$ given by

$$\Lambda(\psi) := \inf\left\{\frac{1}{2}\,\|v\|^2_{L^2(\mathbf{T})} : v\in L^2(\mathbf{T}),\ \psi = \sum_{n=0}^{\infty} J_n^v(f_n)\right\}.$$

The rigorous proof unfolds in three steps, presented in Sections 3.4.2 and 3.4.3. First, one needs to give a proper meaning to the integrals m_k^ε which mix integration against white noise and Lebesgue measures; this is achieved from elementary functions and by passing to the limit. Then we derive proper moment bounds for these integrals, using that $\|v^\varepsilon\|_{L^2(\mathbf{T})} \leq M$ almost surely for some $M > 0$. The argument can be summarised as follows, for all $z \in \mathrm{K}$ and $p \geq 2$,

$$\begin{aligned}
\|m_k^\varepsilon(f_n^{\varepsilon,z})\|_{L^p(\Omega)} &\leq M^{k/2}\,\varepsilon^{n-k}\left\|I_{n-k}\Big(\|f_n^{\varepsilon,z}\|_{L^2(\mathbf{T}^k)}\Big)\right\|_{L^p(\Omega)} && \text{(Cauchy-Schwarz)}\\
&\leq M^{k/2}\,\varepsilon^{n-k}(p-1)^{\frac{n-k}{2}}\left\|I_{n-k}\Big(\|f_n^{\varepsilon,z}\|_{L^2(\mathbf{T}^k)}\Big)\right\|_{L^2(\Omega)} && \text{(hypercontractivity)}\\
&= M^{k/2}\,\varepsilon^{n-k}(p-1)^{\frac{n-k}{2}}\sqrt{(n-k)!}\,\|f_n^{\varepsilon,z}\|_{L^2(\mathbf{T}^n)} && \text{(isometry)}.
\end{aligned}$$

Our assumptions ensure that this estimate and its analogue for $\|m_k^\varepsilon(f_n^{\varepsilon,z} - f_n^{\varepsilon,y})\|_{L^p(\Omega)}$ grant tightness of $(X^{\varepsilon,v^\varepsilon})_{\varepsilon>0}$. Finally, we take the limit ε to zero and identify the limit properly.

To illustrate our main result and show how our abstract assumptions can be satisfied, we propose a variety of applications in Section 3.5. The first one deals with processes that are infinitely Malliavin differentiable and for which the kernels can be computed explicitly. Then we turn our attention to Skorohod integrals, and apply the LDP to multiple Skorohod integrals and linear Skorohod equations. We note that these processes are not necessarily adapted and that LDPs for anticipative processes are very scarce. As a typical example of non-Markovian processes we propose an application to stochastic integrals driven by a fractional Brownian motion in the Wick-Itô sense. Finally, we consider a class of processes inspired from and generalising closed martingales in $\mathbb{R}_+$.

1.3.3 Smile asymptotics in multi-factor rough volatility models

In Chapter 4, based on [159], we study VIX options under rough volatility models and derive small-time explicit formulas which can be helpful for model selection and calibration procedures. This contributes to exhibiting the consistency of this class of models across various traded assets.

In continuous volatility models, the VIX can be defined as the forward price of a log-contract on the SPX with Δ equal to one month

$$\mathrm{VIX}_T := \sqrt{\mathbb{E}\left[-\frac{2}{\Delta}\ln\left(\frac{S_{T+\Delta}}{S_T}\right)\,\Big|\,\mathcal{F}_T\right]},$$

which also turns out to be equal to the implied volatility of the same log-contract. In a stochastic volatility model $\mathrm{d}S_t/S_t = \sigma_t \, \mathrm{d}W_t$, this boils down to

$$\mathrm{VIX}_T = \sqrt{\frac{1}{\Delta}\mathbb{E}\left[\int_T^{T+\Delta} \sigma_t^2 \, \mathrm{d}t \,\middle|\, \mathcal{F}_T\right]},$$

hence it measures the 30-day forward looking volatility of the SPX. One thus needs to choose the variance process v appropriately to model both SPX and VIX derivatives. This joint calibration problem is famously challenging and has attracted a lot of attention [14, 68, 101, 104, 124, 132, 140, 153, 158]. Despite the large number of innovations, the spread of VIX options (options on VIX futures to be precise) remains pretty high, betraying the lack of a convincing answer. To capture the shape of the smile, we would like to focus on two important stylised facts depicted in Figure 1.5: its positive skew and negative curvature.

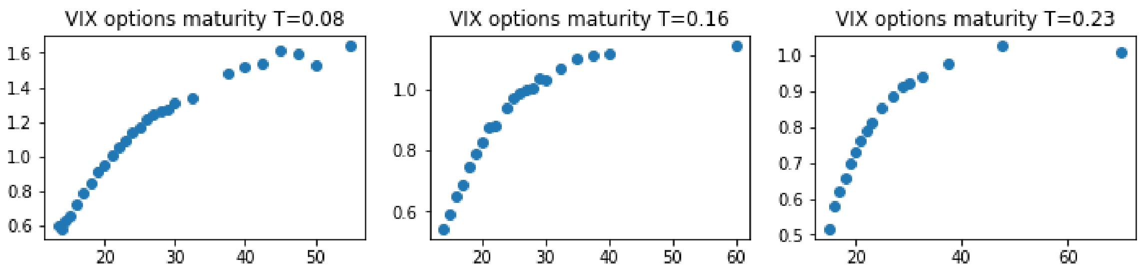

Figure 1.5: VIX implied volatilities with respect to strike as of March 23, 2016.

Combining methods from asymptotic analysis and Malliavin calculus originating from [8], Alós, García-Lorite, and Muguruza [6] showed that the mixed rough Bergomi model

$$\sigma_t^2 = \sigma_0^2 \left[\chi \mathcal{E}\left(\nu W_t^H\right) + (1-\chi)\mathcal{E}\left(\eta W_t^H\right)\right], \tag{1.3.3}$$

where $W_t^H := \int_0^t (t-s)^{H-\frac{1}{2}} \, \mathrm{d}W_s$ is a Riemann-Liouville fBM and $\mathcal{E}(X) := \exp\left(X - \frac{1}{2}\mathbb{E}[X^2]\right)$, produces a positive at-the-money (ATM) skew consistent with the observations. More generally, their framework allows to dismiss other models and parameter choices.

We integrate their insights with those of Bergomi [30, 31] and Gatheral [122] and decide to extend their framework to multi-factor models (driven by an N-dimensional BM) and to compute an asymptotic formula for both the skew and the curvature. Considering a generic $\mathcal{F}_T$-measurable random variable A_T, we apply the Clark-Ocone formula to the martingale $M_t := \mathbb{E}_t[A_T] := \mathbb{E}[A_T|\mathcal{F}_t]$, for $0 \le t \le T$.

$$M_t = \mathbb{E}(A_T) + \sum_{i=1}^{N} \int_0^t \phi_s^i M_s \,\mathrm{d}W_s^i, \qquad \text{where} \quad \phi_s^i = \frac{\mathbb{E}_s[\mathrm{D}_s^i A_T]}{M_s}.$$

This stochastic volatility representation allows to define the ATM skew $\mathcal{S}_T := \partial_k \sigma_{\mathrm{BS}}(T,k)\big|_{k=\ln(M_0)}$ and curvature $\mathcal{C}_T := \partial_k^2 \sigma_{\mathrm{BS}}(T,k)\big|_{k=\ln(M_0)}$ in this general setting. It also yields a price decomposition formula for the ATM option price $V_t := \mathbb{E}_t[(A_T - M_t)^+]$, which is instrumental in obtaining implied volatility representations and short-time asymptotics. The main results are gathered in Theorems 4.3.1, 4.3.2, and 4.3.3.

Main asymptotic results

Under technical assumptions, there exists $\lambda \in (-\frac{1}{2}, 0]$ and $\gamma \in (-1, 0]$ such that

$$\begin{cases} \lim\limits_{T\downarrow 0} \sigma_{\mathrm{BS}}(T,0) & = \lim\limits_{T\downarrow 0} \mathbb{E}[u_0]; \\ \lim\limits_{T\downarrow 0} \dfrac{\mathcal{S}_T}{T^\lambda} & = \lim\limits_{T\downarrow 0} \dfrac{1}{T^\lambda} \mathbb{E}\left[\dfrac{\int_0^T |\mathbf{\Theta}_s| \,\mathrm{d}s}{2u_0^3 T^2}\right]; \\ \lim\limits_{T\downarrow 0} \dfrac{\mathcal{C}_T}{T^\gamma} & = \lim\limits_{T\downarrow 0} \dfrac{1}{T^\gamma} \left\{ \mathcal{S}_T(0) - \dfrac{15}{2}\mathbb{E}\left[\dfrac{1}{u_0^7 T^4} \displaystyle\int_0^T |\mathbf{\Theta}_r| \left(\int_r^T |\mathbf{\Theta}_y| \,\mathrm{d}y \right) \mathrm{d}r \right] \right. \\ & \qquad \left. + \dfrac{3}{2}\mathbb{E}\left[\dfrac{1}{u_0^5 T^3} \displaystyle\int_0^T \sum_{j=1}^N \left\{ \phi_s^j \mathrm{D}_s^j \left(\int_s^T |\mathbf{\Theta}_y| \,\mathrm{d}y \right) \right\} \mathrm{d}s \right] \right\}, \end{cases}$$

where $u_t := \sqrt{\frac{1}{T-t}\int_t^T \|\phi_s\|^2 \,\mathrm{d}s}$ and $|\mathbf{\Theta}_s| := \sum_{i=1}^N \left(\int_s^T \mathrm{D}_s^i \|\phi_r\|^2 \,\mathrm{d}r \right) \phi_s^i$.

The assumptions involve multiple integrals of Malliavin derivatives of ϕ and are left as general as possible. We can then apply this to the VIX case with assumptions that are easier to verify. This is the content of Proposition 4.4.1. The regime $H \in (0, 1/6)$ is the most relevant to rough volatility modelling. Thanks to the nice rescaling $T^{\frac{1}{2}-3H}$ only the third term of the curvature formula contributes here.

Asymptotic results for the VIX

Let $A_T = \mathrm{VIX}_T$. Under four sufficient conditions, for $H \in (0, 1/6)$ we have

$$\begin{cases} \lim\limits_{T\downarrow 0} \sigma_{\mathrm{BS}}(T,0) &= \dfrac{\|\boldsymbol{J}\|}{2\Delta\mathrm{VIX}_0^2}, \\ \lim\limits_{T\downarrow 0} \mathcal{S}_T &= \dfrac{\sum_{i,j=1}^N J_i J_j \left(G_{ij} - \frac{J_i J_j}{\Delta\mathrm{VIX}_0^2}\right)}{2\|\boldsymbol{J}\|^3}, \\ \lim\limits_{T\downarrow 0} \dfrac{\mathcal{C}_T}{T^{3H-\frac{1}{2}}} &= \dfrac{2\Delta\mathrm{VIX}_0^2}{3\|\boldsymbol{J}\|^5} \displaystyle\sum_{i,j,k=1}^N J^i J^j J^k \lim_{T\downarrow 0} \dfrac{\int_T^{T+\Delta} \mathbb{E}_T\left[\mathrm{D}_t^k \mathrm{D}_s^j \mathrm{D}_y^i \sigma_r^2\right] \mathrm{d}r}{T^{3H-\frac{1}{2}}}, \end{cases}$$

where for all $i, j = 1, \cdots, N$, $\|\boldsymbol{J}\| := \left(\sum_{i=1}^N J_i^2\right)^{\frac{1}{2}}$ and

$$J_i := \int_0^\Delta \mathrm{D}_0^i \sigma_r^2 \, \mathrm{d}r, \quad G_{ij} := \int_0^\Delta \mathrm{D}_0^j \mathrm{D}_0^i \sigma_r^2 \, \mathrm{d}r.$$

Finally, we define the two-factor rough Bergomi model, extension of (1.3.3) with two independent fBMs $W^{1,H}$ and $W^{2,H}$

$$\sigma_t^2 = \sigma_0^2 \left[\chi \mathcal{E}\left(\nu W_t^{1,H}\right) + (1-\chi)\mathcal{E}\left(\eta\left(\rho W_t^{1,H} + \overline{\rho} W_t^{2,H}\right)\right)\right],$$

and show in Proposition 4.4.3 that the limits can be computed explicitly and only involve the parameters of the model.

Asymptotic results for the VIX in the two-factor rough Bergomi model

Let $\psi(\rho,\nu,\eta,\chi) := \sqrt{(\chi\nu + (1-\chi)\eta\rho)^2 + (1-\chi)^2\eta^2\overline{\rho}^2}$. For $H \in (0, 1/6)$, as T tends to zero, we have the following limits

$$\begin{cases} \lim\limits_{T\downarrow 0} \sigma_{\mathrm{BS}}(T,0) = \dfrac{\Delta^{H-\frac{1}{2}}}{2(H+\frac{1}{2})} \psi(\rho,\nu,\eta,\chi), \\ \lim\limits_{T\downarrow 0} \mathcal{S}_T = \dfrac{(H+\frac{1}{2})\Delta^{H-\frac{1}{2}}}{2\psi(\rho,\nu,\eta,\chi)^3} \Bigg\{ (1-\chi)^3 \eta^4 \overline{\rho}^4 \left(\dfrac{1}{2H} - \dfrac{1}{(H+\frac{1}{2})^2}\right) \\ \qquad + (\chi\nu + (1-\chi)\eta\rho)^2 \left[\frac{\chi\nu^2 + (1-\chi)\eta^2\rho^2}{2H} - \left(\frac{\chi\nu + (1-\chi)\eta\rho}{H+\frac{1}{2}}\right)^2\right] \\ \qquad + 2(\chi\nu + (1-\chi)\eta\rho)(1-\chi)^2\eta^2\overline{\rho}^2 \left[\dfrac{\eta\rho}{2H} - \dfrac{\nu+\eta\rho}{(H+\frac{1}{2})^2}\right] \Bigg\}, \\ \lim\limits_{T\downarrow 0} \dfrac{\mathcal{C}_T}{T^{3H-\frac{1}{2}}} = \dfrac{128\Delta^{-2H}(H+\frac{1}{2})^2}{3\psi(\rho,\nu,\eta,\chi)^5(1-6H)} \Big\{ (\chi\nu + (1-\chi)\eta\rho)^3 (\chi\nu^3 + (1-\chi)\eta^3\rho^3) \\ + 3(\chi\nu + (1-\chi)\eta\rho)^2 (1-\chi)^2 \eta^4 \overline{\rho}^2 \rho^2 + 3(\chi\nu + (1-\chi)\eta\rho)(1-\chi)^3 \eta^5 \overline{\rho}^4 \rho + (1-\chi)^4 \eta^6 \overline{\rho}^6 \Big\}. \end{cases}$$

These closed-form expressions, combined with the skew short-time limit of the spot implied volatility

$$\lim_{T\downarrow 0} T^{\frac{1}{2}-H}\mathcal{S}_T^{\text{spot}} = \frac{\rho_1\chi\nu + \eta(1-\chi)(\rho_1\rho+\rho_2\overline{\rho})}{(2H+1)(H+\frac{3}{2})},$$

bring new insights on the behaviour of these models and how they should be used for pricing options.

1.3.4 On the large-time behaviour of affine Volterra processes

In this last chapter, we present preliminary results on the large-time behaviour of a class of affine SVEs represented by

$$V_t = V_0 + \int_0^t K(t-s)\beta(\theta - V_s)\,\mathrm{d}s + \int_0^t K(t-s)\sigma\sqrt{V_s}\,\mathrm{d}W_s, \tag{1.3.4}$$

where the kernel is the Laplace transform of some signed measure on $\mathbb{R}_+$. If large-time asymptotic methods for non-Markovian processes are rare, in the case of SVEs this is completely uncharted territory. An overwhelming section of ergodic theory is concerned with Markovian processes, hence parting with this framework requires innovative techniques.

We rely upon a Markovian lift introduced by Cuchiero and Teichmann in [73] which takes the form of a Stochastic Partial Differential Equation (SPDE). More precisely, let $Y := \mathcal{C}_b(\mathbb{R}_+)$, such that its dual $Y^\star$ is the space of signed measures on $\overline{\mathbb{R}}_+$. We consider the (well-posed) Y^*-valued SPDE

$$\mathrm{d}\lambda_t(\mathrm{d}x) = -x\lambda_t(\mathrm{d}x)\,\mathrm{d}t + \nu(\mathrm{d}x)\mathrm{d}X_t, \tag{1.3.5}$$

where ν is a signed measure and X is an Itô semimartingale of the form

$$X_t = \int_0^t \beta(\theta - \langle 1, \lambda_s\rangle)\,\mathrm{d}s + \int_0^t \sigma\sqrt{\langle 1, \lambda_s\rangle}\mathrm{d}W_s,$$

where $\langle g, \lambda\rangle := \int_0^\infty g(x)\,\lambda(\mathrm{d}x)$ for all $g \in \mathcal{C}_b(\mathbb{R}_+)$. The total mass $\langle 1, \lambda\rangle$ thus solves the affine SVE (1.3.4), with $K(t) = \int_0^\infty \mathrm{e}^{-tx}\,\nu(\mathrm{d}x)$. We note that, passing to the multivariate case, the rough Heston model is also included in this framework.

The Feller property, a crucial component of traditional Markovian ergodic theory, is extended to a class of weighted spaces conducive for the study of the Markovian processes satisfying (1.3.5). This generalised Feller property, introduced in [86], plays a central role in our analysis. Leveraging on it, we first obtain a version of Krylov-Bogoliubov theorem on these weighted spaces. Combined with a tightness criterion on measure space, it entails that a generalised Feller process λ has an invariant measure provided the condition $\sup_{t\geq 0}\mathbb{E}[\|\lambda_t\|_{Y^*}] < \infty$ holds.

We prove that the solution to the SPDE (1.3.5) is indeed a generalised Feller process (only the case $\theta = 0$ was treated in [73]) and that it satisfies the condition above where the kernel is of Riemann-Liouville type $K(t) = t^{\alpha-1}/\Gamma(\alpha)$, with $\alpha \in (1/2, 3/2)$. This yields our main result, Theorem 5.1.4.

Existence of an invariant measure

Let $\lambda_0(\mathrm{d}x) = V_0\,\delta_0(\mathrm{d}x)$ for some $V_0 > 0$ and $\nu(\mathrm{d}x) = c\,x^{-\alpha}$, where $\alpha \in (1/2, 3/2)$ and $c > 0$. Then the solution of the SPDE (1.3.5) has an invariant measure. Furthermore, the solution of the SVE (1.3.4) has a stationary distribution, for any $V_0 > 0$, $K(t) = C\,t^{\alpha-1}$, with $\alpha \in (1/2, 3/2)$ and $C > 0$.

The natural continuation of this work consists in extending our framework to the multi-dimensional case, determining whether the invariant measure is unique and then identifying it. This would be the first result of this type for an SVE and a rough volatility model.

2

Large and moderate deviations for stochastic Volterra systems

This chapter, based on [162], is a joint work with Antoine Jacquier, submitted to *Stochastic Processes and their Applications*.

2.1 Introduction

This chapter sheds new light on the asymptotic behaviour of the class of stochastic Volterra equations (SVEs)

$$X_t = X_0 + \int_0^t K(t,s)b(s,X_s)\mathrm{d}s + \int_0^t K(t,s)\sigma(s,X_s)\mathrm{d}W_s, \quad t \in [0,T], \tag{2.1.1}$$

for some fixed time horizon $T > 0$, where $X_0 \in \mathbb{R}^d$, $d \in \mathbb{N}$, W is a multidimensional Brownian motion, K is a kernel that may be singular, and the coefficients are such that a unique pathwise solution exists. This class of models has been investigated in many fields, including nonlinear filtering [70] using fractional Brownian motion kernels, pharmacokinetic models [186] (Langevin equation driven by fractional Brownian motion), fluid turbulence [60], and turbulence modelling in atmospheric winds or energy prices [16, 69] using Brownian Semistationary processes.

As the literature review of Section 1.2 testifies, mathematical finance has however been the most dynamic area by far in terms of applications of SVEs, where they are a natural framework for rough volatility models. We already explained how the latter reconcile the stylised facts of the markets from both the statistical and the option pricing viewpoints and that, due to the loss of tractability, asymptotic methods have flourished, both to provide clearer understanding of models in extreme parameter configurations and to act as proxies to numerical schemes.

Large Deviations Principles (LDP), in particular, have been widely explored in mathematical finance, and we refer the interested reader to [110] for an overview. Let $(X^\varepsilon)_{\varepsilon>0}$ be a sequence of random variables in some Polish space $\mathcal{X}$, converging in probability to a deterministic limit $\overline{X}$ as ε goes to zero. This sequence is said to satisfy an LDP with speed ε^{-1} and rate function $I : \mathcal{X} \to \overline{\mathbb{R}}_+$ if for all Borel subsets $B \subset \mathcal{X}$, the inequalities

$$-\inf_{x\in B^\circ} I(x) \le \liminf_{\varepsilon\downarrow 0} \varepsilon \log \mathbb{P}\big(X^\varepsilon \in B\big) \le \limsup_{\varepsilon\downarrow 0} \varepsilon \log \mathbb{P}\big(X^\varepsilon \in B\big) \le -\inf_{x\in \overline{B}} I(x) \quad (2.1.2)$$

hold, and the level sets $\{x \in \mathcal{X} : I(x) \le M\}$ of I are compact for all $M > 0$. This rate function encompasses in a (relatively) concise formula first-order information about the asymptotic behaviour of complex dynamical systems. If X satisfies (2.1.1) and $\mathcal{X} = \mathbb{R}^d$, one can consider finite-dimensional LDP for $(X_t)_{t\ge 0}$ (also called small-time LDP if the limit takes place as t goes to zero), or pathwise LDP for some rescaling of X with $\mathcal{X} = \mathcal{C}([0,T],\mathbb{R}^d)$. The former is easily recovered from the latter by a projection argument. Moderate deviations on the other hand are concerned with deviations of a lower order than large deviations, and thus apply to 'less rare events'. We recall that $(X^\varepsilon)_{\varepsilon>0}$ satisfies a moderate deviations principle (MDP) if $(\eta^\varepsilon)_{\varepsilon>0}$ satisfies an LDP with speed h_ε^2, where

$$\eta^\varepsilon := \frac{X^\varepsilon - \overline{X}}{\sqrt{\varepsilon} h_\varepsilon}, \qquad \text{for all } \varepsilon > 0,$$

with $\lim_{\varepsilon\downarrow 0} h_\varepsilon = +\infty$ and $\lim_{\varepsilon\downarrow 0} \sqrt{\varepsilon} h_\varepsilon = 0$. Since the speed of convergence of h_ε is not fixed, an MDP essentially bridges the gap between the central limit regime where $h_\varepsilon = 1$ and the LDP regime where $h_\varepsilon = \varepsilon^{-1/2}$. An edifying example of the relevance of moderate deviations appears in [111] where, interested in option pricing asymptotics, the authors judiciously rescale the strikes with respect to time to expiry. Indeed, as time to expiry becomes smaller, the range of pertinent strikes naturally shrinks, and this 'moderately-out-of-the-money' regime becomes more realistic.

Large deviations for SVEs were originally studied in [203, 224] with regular kernels. In the context of rough volatility, Forde and Zhang [103] introduced the first finite-dimensional LDP where the log-volatility is modelled by a fractional Brownian motion, and refined versions followed in [18, 20, 109], while pathwise LDP for similar models were studied in [56, 137]. Departing from regular conditions on the behaviour of the coefficients led to specific requirements, and finite-dimensional large deviations for the fractional Heston model were carried out in [100, 136], while pathwise LDPs were derived for the rough Stein-Stein model with random starting point [151], for the rough Bergomi model [161], and small-time LDPs for the multi-factor rough Bergomi appeared in [173]. We emphasise at this point that no pathwise MDP was previously known in the context of rough volatility.

The Gärtner-Ellis theorem [82, Theorem 2.3.6] is the main ingredient of a finite-dimensional LDP and depends on explicit computations of certain limits of the Laplace transform. This is only available though, when the process is either Gaussian [103] or affine [100]. Pathwise LDP on the other hand, have mainly been derived using the Freidlin-Wentzell approach [106]: starting from known large deviations for the driving (Gaussian) process [83, Theorem 3.4.5], they follow from a combination of approximations and continuous mapping, keeping track of the rate function. While this methodology is clear, it requires a case-by-case tailored path for each model, and in general leads to a cumbersome rate function. Furthermore, pathwise moderate deviations are so far out of reach in this approach, partially explaining the small number of related results compared to LDP.

The weak convergence approach, presented in Section 1.2, was introduced by Dupuis and Ellis in the monograph [88] and developed further by Budhiraja and Dupuis [46]. We recall that this alternative consists in proving a Laplace principle (1.2.6) where the left-hand side pre-limit of can be represented as a variational principle for expectations of functionals of Brownian motion [38, Theorem 3.1].

Lemma 2.1.1 (Boué-Dupuis). *Let W be an $\mathbb{R}^m$-Brownian motion and F be a bounded*

Borel-measurable function mapping $\mathcal{C}([0,T],\mathbb{R}^m)$ into $\mathbb{R}$. Then

$$-\log\mathbb{E}\left[\mathrm{e}^{-F(W)}\right]=\inf_{v\in\mathcal{A}}\mathbb{E}\left[\frac{1}{2}\int_0^T|v_s|^2\,\mathrm{d}s+F\left(W+\int_0^{\cdot}v_s\,\mathrm{d}s\right)\right], \tag{2.1.3}$$

where

$$\mathcal{A}:=\left\{v:[0,T]\to\mathbb{R}^m\ \textit{progressively measurable},\ \mathbb{E}\left[\int_0^T|v_t|^2\,\mathrm{d}t\right]<+\infty\right\}. \tag{2.1.4}$$

The representation (2.1.3) contains in a single formula the usual tools used in the proof of an LDP. The first term on the right-hand side comes from the relative entropy between the Wiener measure and the measure shifted by $\int_0^{\cdot}v_s\,\mathrm{d}s$ via Girsanov's theorem, under which $W+\int_0^{\cdot}v_s\,\mathrm{d}s$ is a Brownian motion. It can be interpreted as the cost of deviating from the original path and clearly indicates where the form of the rate function comes from. In essence, this representation replaces the non-linear analysis of the Freidlin-Wentzell approach with the linear theory of weak convergence. Instead of exponential estimates, only qualitative properties of the shifted process need to be established, such as strong existence and uniqueness and tightness.

The extensive literature on the topic, summarised in [46] and the references therein, demonstrates the strength of this generic approach which can be applied to a variety of models without appealing to their particular features. It has been used to derive LDPs, in the continuous-time case, for diffusions [62], multiscale systems [48, 89, 227], SVEs with singular kernels and Lipschitz continuous coefficients [240], SDEs driven by infinite-dimensional Brownian motions [49], by Poisson random measures [44] or both [50], including stochastic PDEs. Contrary to the Freidlin-Wentzell approach, this method has also proved efficient to obtain MDPs. SVEs with Lipschitz continuous kernels [177], SDEs with jumps [47] and slow-fast systems [193] are a few relevant examples. The latter were then tailored to the setting of stochastic volatility models in [164], developing the first application of the weak convergence approach in mathematical finance, and extending the MDP results in [111] to a pathwise setting. A further appealing feature of moderate deviations is the simple form, often quadratic, of the rate function, as opposed to that provided by large deviations, thereby opening the gates to the use of importance sampling and variance reduction techniques [90, 194, 220].

Building on this powerful approach, we provide a unified treatment of (finite-dimensional and pathwise) large and moderate deviations in the general framework (2.1.1) by showing the weak convergence of a perturbed system. We relax the uniqueness requirement for the limiting Volterra equation, as in [67, 84] for the diffusion case, allowing us to consider coefficients that are not Lipschitz continuous and do not necessarily have sublinear growth.

The chapter is organised in the following way. Section 2.2 introduces the framework and useful definitions. In Section 2.3, we present abstract criteria for the validity of an LDP, extending the results by Budhiraja and Dupuis [45, Theorem 4.1]. Our main results, Theorem 2.3.11 for LDP and Theorem 2.3.20 for MDP, are then stated in the case of convolution kernels and extended to non-convolution kernels in Theorem 2.3.29. In Section 2.4, we show how these results apply to rough volatility models, and give precise formulae for the rough Stein-Stein, the (multi-factor) rough Bergomi and the rough Heston models. We finally gather technical proofs in the final section.

2.2 General framework

2.2.1 Notations

We consider a fixed time horizon $T > 0$, and denote $\mathbb{T} := [0, T]$, $\mathbb{R}_+ := [0, +\infty)$ and $\overline{\mathbb{R}}_+ := [0, +\infty]$. For $d_1 \geq 1$, $d_2 \geq 1$, $|\cdot|$ denotes the Euclidean norm in $\mathbb{R}^{d_1}$ and the Frobenius norm in $\mathbb{R}^{d_1 \times d_2}$. For $p \geq 1$, L^p stands short for $L^p(\mathbb{T})$, and $\|\cdot\|_2$ is the usual L^2 norm. Furthermore, for $d \geq 1$, $\mathcal{W}^d := \mathcal{C}(\mathbb{T} : \mathbb{R}^d)$ represents the space of continuous functions from $\mathbb{T}$ to $\mathbb{R}^d$, equipped with the supremum norm $\|\varphi\|_{\mathbb{T}} := \sup_{t\in\mathbb{T}} |\varphi_t|$ for any $\varphi \in \mathcal{W}^d$. Finally, for any $d_1 \geq 1$, $d_2 \geq 1$, $M > 0$, $f : \mathbb{R}^{d_1} \to \mathbb{R}^{d_2}$, we write $\|f\|_R := \sup\{|f(x)| : |x| \leq R\}$. Unless stated otherwise, constants will be denoted by C (with possible subscript) and may be different from one proof to another. Every statement involving ε stands for all $\varepsilon > 0$ small enough. A family of random variables will be called tight if the corresponding measures are tight [88, Appendix A]. We also use the classical convention that the infimum over an empty set is equal to infinity. Finally, we recall the following definitions for clarity and notations:

Definition 2.2.1. Let g be a function from $\mathbb{R}^d$ to $\mathbb{R}^n$.

- It has linear growth if there exists $C_L > 0$ such that $|g(x)| \leq C_L(1 + |x|)$, for all $x \in \mathbb{R}^d$;

- if it is uniformly continuous, it admits a continuous and increasing modulus of continuity $\rho_g : \mathbb{R}_+ \to \mathbb{R}_+$, with $\rho_g(0) = 0$ and $|g(x) - g(y)| \le \rho_g(|x-y|)$, for all $x, y \in \mathbb{R}^d$;
- it is locally δ-Hölder continuous with $\delta \in (0,1)$ if, for all $R > 0$, there exists $C_R > 0$ such that $|g(x) - g(y)| \le C_R |x-y|^\delta$, for all $|x| \vee |y| \le R$.

2.2.2 Framework

We consider small-noise convolution stochastic Volterra equations (SVE)

$$X_t^\varepsilon = X_0^\varepsilon + \int_0^t K(t-s) b_\varepsilon(s, X_s^\varepsilon) \mathrm{d}s + \vartheta_\varepsilon \int_0^t K(t-s) \sigma_\varepsilon(s, X_s^\varepsilon) \mathrm{d}W_s, \quad t \in \mathbb{T}, \quad (2.2.1)$$

taking values in $\mathbb{R}^d$ with $d \ge 1$, where $\varepsilon > 0$, and $\vartheta_\varepsilon > 0$ tends to zero as ε goes to zero. For each $\varepsilon > 0$, $X_0^\varepsilon \in \mathbb{R}^d$, $b_\varepsilon : \mathbb{T} \times \mathbb{R}^d \to \mathbb{R}^d$ and $\sigma_\varepsilon : \mathbb{T} \times \mathbb{R}^d \to \mathbb{R}^{d\times m}$ are Borel-measurable functions, and W is an m-dimensional Brownian motion on the filtered probability space $(\Omega, \mathcal{F}, \{\mathcal{F}_t\}_{t\in\mathbb{T}}, \mathbb{P})$ satisfying the usual conditions. The kernel function $K : \mathbb{T} \to \mathbb{R}^d \times \mathbb{R}^d$, of convolution type, is allowed to be singular, thus encompassing fractional processes, in particular the recent literature on rough volatility [8, 19, 93, 123]. Components of the system are in general correlated, the correlation matrix being implicitly encoded in the diffusion coefficient σ_ε. General existence and uniqueness results for such stochastic Volterra equations are so far out of reach, and our conditions below are sufficient and general enough for most applications. In order to state them precisely, we first introduce several definitions and concepts:

Definition 2.2.2. For any $\varepsilon > 0$, a solution to (2.2.1) is an $\mathbb{R}^d$-valued progressively measurable stochastic process X^ε satisfying (2.2.1) almost surely and such that

$$\mathbb{P}\left(\int_0^t \left\{ |K(t-s) b_\varepsilon(s, X_s^\varepsilon)| + |K(t-s)\sigma_\varepsilon(s, X_s^\varepsilon)|^2 \right\} \mathrm{d}s < \infty, \text{ for all } t \in \mathbb{T} \right) = 1. \quad (2.2.2)$$

We shall assume that the (singular) convolution kernel satisfies the following condition, which is essentially a multivariate version of the one given in [3, Condition (2.5)]:

Assumption 2.2.3. The kernel $K : \mathbb{T} \to \mathbb{R}^{d\times d}$ is an upper triangular matrix satisfying the following conditions: $K \in L^2(\mathbb{T} : \mathbb{R}^{d\times d})$ and there exists $\gamma \in (0, 2]$ such that, for h

small enough,

$$\int_0^h |K(t)|^2 \,\mathrm{d}t + \int_0^T |K(t+h) - K(t)|^2 \,\mathrm{d}t = \mathcal{O}(h^\gamma).$$

We refer to [3, Example 2.3] for a broad range of kernels that satisfy this assumption. Of particular interest in mathematical finance is the Riemann-Liouville kernel $K(t) = t^{H-\frac{1}{2}}$, for $H \in (0, \frac{1}{2})$ implying $\gamma = 2H$. Moreover if $\widetilde{K}$ is locally Lipschitz and K satisfies Assumption 2.2.3 then so does the product $K\widetilde{K}$; this includes the gamma and power-law kernels which are related to the class of Brownian semistationary processes [17].

Remark 2.2.4. This setup covers in particular the following two useful forms for the kernel:

- $K = \mathrm{Diag}(k_1, \cdots, k_d)$ is a diagonal matrix, where each $k_i : \mathbb{T} \to \mathbb{R}$ satisfies Assumption 2.2.3.
- The drift and diffusion coefficients of any sub-system of (2.2.1) can be convoluted with different kernels. As an example, the one-dimensional SVE

$$Y_t^\varepsilon = Y_0^\varepsilon + \int_0^t K_1(t-s) b_\varepsilon^Y(s, Y_s^\varepsilon) \,\mathrm{d}s + \int_0^t K_2(t-s) \sigma_\varepsilon^Y(s, Y_s^\varepsilon) \,\mathrm{d}W_s^{(1)}$$

is the first component of (2.2.1) with $d = 2$ and

$$K = \begin{pmatrix} K_1 & K_2 \\ 0 & 0 \end{pmatrix}, \quad b_\varepsilon = \begin{pmatrix} b_\varepsilon^Y \\ 0 \end{pmatrix}, \quad \sigma_\varepsilon = \begin{pmatrix} 0 & 0 \\ \sigma_\varepsilon^Y & 0 \end{pmatrix}.$$

One could in fact use general matrices to eliminate the auxiliary state, such as $K = (K_1\ K_2) : \mathbb{T} \to \mathbb{R}^{1,2}$ in the example above. We stick to square matrices for consistency.

Volterra systems appearing in the literature, and in particular in the mathematical finance one, have a specific structure in the sense that only one component satisfies an SVE with (singular) kernel, and can be dealt with independently of the other component. This particular structure allows us to relax some conditions on the coefficients, and we shall leverage on it whenever needed. We make this more specific through the following two definitions:

Definition 2.2.5. Let $\Upsilon \subset [\![1, d]\!]$ and $\Gamma : \mathbb{R}^{|\Upsilon|} \to [0, \infty)$. We define $\mathcal{S}_\Upsilon^\Gamma$ as the set of functions f for which there exists a strictly positive constant C_Υ such that, for all $x \in \mathbb{R}^d$,

$$|f(x)| \leq C_\Upsilon \Big(1 + |x|_{\Upsilon^c} + \Gamma\big(x^{(\Upsilon)}\big)\Big), \tag{2.2.3}$$

where $|x|_{\Upsilon^c} := \sum_{i \in \Upsilon^c} |x^{(i)}|$ and $x^{(\Upsilon)} := (x^{(i)})_{i \in \Upsilon}$.

Definition 2.2.6. The process X^ε admits an autonomous $\mathcal{S}_\Upsilon^\Gamma$-subsystem $\{X^{\varepsilon,(l)}\}_{l \in \Upsilon}$ if for all $1 \leq i, j \leq d$ such that $K_{ij} \neq 0$, $b_\varepsilon^{(j)}(t, \cdot)$ and all the components of the row $\sigma_\varepsilon^{(j)}(t, \cdot)$ satisfy the following for small enough ε and uniformly in $t \in \mathbb{T}$:

- if $i \in \Upsilon$, they have linear growth and do not depend on $X^{\varepsilon,(k)}$ for $k \in \Upsilon^c$;
- if $i \in \Upsilon^c$, they belong to $\mathcal{S}_\Upsilon^\Gamma$.

Example 2.2.7. *The motivation for Definition 2.2.6 is to be able to handle (rough) stochastic volatility models, ubiquitous in mathematical finance, where linear growth of all the coefficients may not hold. Consider for example the rough Bergomi model [19]*

$$\begin{cases} X_t^{(1)} & = -\dfrac{1}{2}\displaystyle\int_0^t \exp\left(X_s^{(2)}\right) \mathrm{d}s + \int_0^t \exp\left(\frac{1}{2}X_s^{(2)}\right) \mathrm{d}B_s, \\ X_t^{(2)} & = y_0 - at^{2H} + \displaystyle\int_0^t (t-s)^{H-\frac{1}{2}} \mathrm{d}W_s, \end{cases}$$

where B and W are Brownian motions with correlation $\rho \in (-1, 1)$. After dropping the dependence in ε, this fits into the setup of (2.2.1) with $d = 3$, $y_0 \in \mathbb{R}$, $a > 0$, $H \in (0, \frac{1}{2}), \overline{\rho} = \sqrt{1-\rho^2}$ and

$$K(t) = \begin{pmatrix} 1 & 0 & 0 \\ 0 & t^{2H-1} & t^{H-\frac{1}{2}} \\ 0 & 0 & 0 \end{pmatrix}, \quad b\big(t, (x_1, x_2)\big) = \begin{pmatrix} -\mathrm{e}^{x_2}/2 \\ -a/(2H) \\ 0 \end{pmatrix}, \quad \sigma\big(t, (x_1, x_2)\big) = \begin{pmatrix} \overline{\rho}\,\mathrm{e}^{x_2/2} & \rho\,\mathrm{e}^{x_2/2} & 0 \\ 0 & 0 & 0 \\ 1 & 0 & 0 \end{pmatrix},$$

where the third component is meaningless but allows us to handle the two different kernels. Here X admits $X^{(2)}$ as autonomous subsystem with $\Upsilon = \{2\}$ and $\Gamma(x_2) = 1 + \mathrm{e}^{x_2}$.

The following set of assumptions, inspired by [62], completes our framework:

H1. X_0^ε converges to $x_0 \in \mathbb{R}^d$ as ε tends to zero.

H2. For all $\varepsilon > 0$ small enough, the coefficients b_ε and σ_ε are measurable maps on $\mathbb{T} \times \mathbb{R}^d$ and converge pointwise to b and σ as ε goes to zero. Moreover, $b(t, \cdot)$ and $\sigma(t, \cdot)$ are continuous on $\mathbb{R}^d$, uniformly in $t \in \mathbb{T}$.

H3. Either **a)** or **b)** holds:

- **a)** For all $\varepsilon > 0$ small enough, b_ε and σ_ε have linear growth uniformly in ε and in $t \in \mathbb{T}$.
- **b)** The process X^ε admits an autonomous $\mathcal{S}_\Upsilon^\Gamma$-subsystem.

H4. The SVE (2.2.1) is pathwise unique for small enough $\varepsilon > 0$.

H2 ensures that, on compact subsets of $\mathbb{T} \times \mathbb{R}^d$, the convergence of b_ε and σ_ε is uniform and that b and σ are uniformly continuous. **H1, H2, H3a** are standard and easily verifiable. **H3b** is unusual but includes a large number of functions; Assumption 2.3.9 will complete it to indicate the role of Γ such as to include Example 2.2.7. Moreover, the growth conditions from **H3** are uniform in ε and therefore apply to the limits b and σ. The main restrictions arise from **H4**, although the latter is satisfied if, for instance, the coefficients b_ε and σ_ε are locally Lipschitz continuous for small enough $\varepsilon > 0$. This condition was relaxed in [195] to the one-dimensional case where $K(t) = t^{-\alpha}$, for $\alpha \in (0, \frac{1}{2})$ and $\sigma(x) = x^\gamma$, for $\gamma \in (\frac{1}{2(1-\alpha)}, 1]$, which is clearly not Lispchitz continuous. Furthermore, to the best of our knowledge there currently exists no pathwise LDP for stochastic equations where pathwise uniqueness fails.

One can compare our setup with the SVE considered by Zhang in [240], where a pathwise LDP was derived under the assumptions of Lipschitz continuity and linear growth of the coefficients. We relax both these assumptions. Indeed, our framework covers Hölder-continuous diffusion coefficients, as mentioned in the previous paragraph, and functions with non-linear growth through the concept of autonomous $\mathcal{S}_\Upsilon^\Gamma$-subsystem, as in Example 2.2.7.

2.3 Large and moderate deviations

As discussed in the introduction, our goal is to provide pathwise large and moderate deviations for the general convolution stochastic Volterra system (2.2.1), and then extend these to non-convolution kernels. The classical Freidlin-Wentzell approach, used in [106], has limitations regarding the behaviour of the coefficients, and the rate function is often rather cumbersome to write. We follow here instead the weak convergence approach developed by Dupuis and Ellis [88]. We first introduce the reader to their abstract setting, and refine the large deviations result by Budhiraja and Dupuis [45] to our general setup. We then show how this abstract framework applies to the small-noise stochastic Volterra system (2.2.1), first proving pathwise large deviations, and then the moderate deviations counterpart.

2.3.1 Weak convergence approach: the abstract setting

Given a family of Borel-measurable functions $\{\mathcal{G}^\varepsilon\}_{\varepsilon>0}$ from $\mathcal{W}^m$ to $\mathcal{W}^d$, we enquire about the large deviations behaviour of the family of random variables $\{\mathcal{G}^\varepsilon(W)\}_{\varepsilon>0}$ as ε tends to zero, where W is a standard Brownian motion on the filtered probability space above. For each $M>0$, the spaces of bounded deterministic and stochastic controls

$$\mathcal{S}_M := \left\{v \in L^2 : \int_0^T |v_s|^2\,\mathrm{d}s \le M\right\} \qquad \text{and} \qquad \mathcal{A}_M := \left\{v \in \mathcal{A} : v \in \mathcal{S}_M \text{ almost surely}\right\}, \tag{2.3.1}$$

with $\mathcal{A}$ introduced in (2.1.4), are equipped with the weak topology on $L^2(\mathbb{T}\times\Omega)$ such that they are closed and even compact (by Banach-Alaoglu-Bourbaki theorem). Budhiraja and Dupuis [45] assume, for any sequence $\{v^\varepsilon\}_{\varepsilon>0}$ in $\mathcal{A}_M$ converging weakly to $v \in \mathcal{A}_M$, the existence of a limit in distribution of $\mathcal{G}^\varepsilon\left(W + \frac{1}{\varepsilon}\int_0^\cdot v_s^\varepsilon\,\mathrm{d}s\right)$ which is uniquely characterised by v. However, such uniqueness may fail when the coefficients of the system (2.1.1) (in particular the diffusion coefficient σ) are not locally Lipschitz, as is the case for the Feller diffusion for example (in this case without singular kernel, a dedicated analysis was carried out in [67, 84] using the Freidlin-Wentzell approach). We relax here this uniqueness assumption by replacing the limiting trajectory by a perturbed version.

Definition 2.3.1. For all $M \in \mathbb{N}$ and $v \in \mathcal{A}_M$ we define

$$\begin{aligned}\mathcal{G}^0_{v,M} := \Big\{\phi : \Omega \to \mathcal{W}^d \;\Big|\; &\text{there exist } \{\varepsilon_n\}_{n\in\mathbb{N}} \subset \mathbb{R}_+ \text{ with } \lim_{n\uparrow\infty} \varepsilon_n = 0 \\ &\text{and a sequence } \{v^{\varepsilon_n}\}_{n\in\mathbb{N}} \subset \mathcal{A}_M \text{ with } \lim_{n\uparrow\infty} v^{\varepsilon_n} = v \text{ in distribution such that} \\ &\phi = \lim_{n\uparrow\infty} \mathcal{G}^{\varepsilon_n}\left(W + \varepsilon_n^{-1}\int_0^\cdot v_s^{\varepsilon_n}\,\mathrm{d}s\right) \text{ in distribution}\Big\},\end{aligned}$$

and $\mathcal{G}^0_{v,M}$ is empty if v is not in $\mathcal{A}_M$. For all $v \in \mathcal{A}$, we also denote $\mathcal{G}^0_v := \bigcup_{M \in \mathbb{N}} \mathcal{G}^0_{v,M}$.

Then we define the functional $I : \mathcal{W}^d \to \overline{\mathbb{R}}_+$ given by

$$I(\phi) := \inf\left\{ \frac{1}{2}\int_0^T |v_s|^2 \, \mathrm{d}s : v \in L^2 \text{ such that } \phi \in \mathcal{G}^0_v \right\}. \tag{2.3.2}$$

Definition 2.3.2. We say that $\phi \in \mathcal{W}^d$ is *uniquely characterised* if there exists a sequence $\{v^n\}_{n \in \mathbb{N}} \subset L^2$ such that

$$\mathcal{G}^0_{v^n} = \{\phi\} \qquad \text{and} \qquad \frac{1}{2}\int_0^T |v^n_s|^2 \, \mathrm{d}s \leq I(\phi) + \frac{1}{n}, \qquad \text{for all } n \in \mathbb{N}. \tag{2.3.3}$$

In particular, if there exists $\widetilde{v} \in L^2$ which attains the infimum in (2.3.2) and $\mathcal{G}^0_{\widetilde{v}} = \{\phi\}$ then ϕ is uniquely characterised, because one can choose $v^n = \widetilde{v}$ for all $n \in \mathbb{N}$.

In the display (2.3.2) and Definition 2.3.2, ϕ, v, v^n and $\widetilde{v}$ are all deterministic.

Assumption 2.3.3. For any $\delta > 0$ and any $\phi \in \mathcal{W}^d$ such that $I(\phi) < +\infty$, there exists ϕ^δ uniquely characterised such that $\left\|\phi - \phi^\delta\right\|_{\mathbb{T}} \leq \delta$ and $\left|I(\phi) - I(\phi^\delta)\right| \leq \delta$.

Remark 2.3.4. This assumption is reminiscent of [84, Proposition 3.3], where the authors resolve the non-uniqueness issue in the diffusion case. A similar problem is also at the core of [51, Lemma 5.1] in an infinite-dimensional setting.

Our abstract large deviations result is the following, extending [45, Theorem 4.4], at least when the underlying Hilbert space is $L^2(\mathbb{T} : \mathbb{R}^m)$, to the non-uniqueness case.

Theorem 2.3.5. *Assume that*

(i) For all $M > 0$, all $v \in \mathcal{A}_M$ and all families $\{v^\varepsilon\}_{\varepsilon>0}$ in $\mathcal{A}_M$ converging in distribution to v as ε tends to zero, $\left\{\mathcal{G}^\varepsilon\left(W + \varepsilon^{-1}\int_0^\cdot v^\varepsilon_s \, \mathrm{d}s\right)\right\}_{\varepsilon>0}$ is tight.

(ii) The functional I defined by (2.3.2) has compact level sets.

(iii) Assumption 2.3.3 holds.

Then the family $\{\mathcal{G}^\varepsilon(W)\}_{\varepsilon>0}$ satisfies the Laplace principle and, by equivalence, the Large Deviations Principle with rate function I and speed ε^{-2}.

Remark 2.3.6. Item (i) entails that, for all $M > 0$, all $v \in \mathcal{A}_M$ and all families $\{v^\varepsilon\}_{\varepsilon>0}$ in $\mathcal{A}_M$ converging in distribution to v, there exists a subsequence such that $\lim_n \mathcal{G}^{\varepsilon_n}\left(W + \varepsilon_n^{-1}\int_0^\cdot v^{\varepsilon_n}_s \, \mathrm{d}s\right)$ exists and by definition belongs to $\mathcal{G}^0_{v,M}$.

Remark 2.3.7. In the large deviations literature, a rate function is sometimes called 'good' if it has compact level sets. All the rate functions in the present paper satisfy this requirement (item (ii) above takes care of that) therefore we drop the adjective 'good'.

We defer the proof to Section 2.5.1; the lower bound can be tackled as in [45], and we therefore concentrate on the upper bound. The idea is that the Laplace principle (1.2.6) upper bound involves an infimum so deriving it only requires a δ-optimal path. Hence a perturbation will also do the trick, provided one knows how to handle the control associated to it. In [45, Theorem 4.4], unique characterisation of the limiting element in (i) is granted, and the set $\mathcal{G}_v^0$ is a singleton that takes the form $\mathcal{G}^0(\int_0^\cdot v_s\,\mathrm{d}s)$, where they view $\mathcal{G}^0$ as a map. In that case Assumption 2.3.3 is clearly satisfied since ϕ^δ can be taken as ϕ itself.

2.3.2 Application to stochastic Volterra systems

We now show how the abstract setting developed above in Section 2.3.1 applies to the small-noise stochastic Volterra system (2.2.1) and why pathwise uniqueness is so fundamental. If **H4** holds, define the functional $\mathcal{G}^\varepsilon$ as the Borel-measurable map associating the multidimensional Brownian motion W to the solution of the stochastic Volterra equation (2.2.1), that is: $\mathcal{G}^\varepsilon(W) = X^\varepsilon$. For any control $v \in \mathcal{A}_M$, $M > 0$ (introduced in (2.3.1)) and any $\varepsilon > 0$, the process $\widetilde{W} := W + \vartheta_\varepsilon^{-1}\int_0^\cdot v_s\,\mathrm{d}s$ is a $\widetilde{\mathbb{P}}$-Brownian motion by Girsanov's theorem, where

$$\frac{\mathrm{d}\widetilde{\mathbb{P}}}{\mathrm{d}\mathbb{P}} := \exp\left\{-\frac{1}{\vartheta_\varepsilon}\sum_{i=1}^m \int_0^T v_s^{(i)}\mathrm{d}W_s^{(i)} - \frac{1}{2\vartheta_\varepsilon^2}\int_0^T |v_s|^2\,\mathrm{d}s\right\}.$$

Hence the shifted version $X^{\varepsilon,v} := \mathcal{G}^\varepsilon(\widetilde{W})$ appearing in Theorem 2.3.5(i) is the strong unique solution of (2.2.1) under $\widetilde{\mathbb{P}}$, with X^ε and W replaced by $X^{\varepsilon,v}$ and $\widetilde{W}$. Because $\mathbb{P}$ and $\widetilde{\mathbb{P}}$ are equivalent, $X^{\varepsilon,v}$ is also the unique strong solution, under $\mathbb{P}$, of the controlled equation

$$X_t^{\varepsilon,v} = X_0^\varepsilon + \int_0^t K(t-s)\Big[b_\varepsilon(s, X_s^{\varepsilon,v}) + \sigma_\varepsilon(s, X_s^{\varepsilon,v})v_s\Big]\mathrm{d}s + \vartheta_\varepsilon\int_0^t K(t-s)\sigma_\varepsilon(s, X_s^{\varepsilon,v})\mathrm{d}W_s. \tag{2.3.4}$$

Under appropriate conditions, and using the notations set in **H1, H2**, we heuristically observe that taking ε to zero, the system (2.3.4) reduces to the deterministic Volterra equation

$$\phi_t = x_0 + \int_0^t K(t-s)\Big[b(s,\phi_s) + \sigma(s,\phi_s)v_s\Big]\mathrm{d}s. \tag{2.3.5}$$

We will show later that the set $\mathcal{G}_v^0$ corresponds to the set of solutions of (2.3.5).

Example 2.3.8. *To illustrate the need for a set $\mathcal{G}_v^0$ rather than a singleton, consider the Feller diffusion*

$$X_t = x_0 + \kappa\int_0^t (\theta - X_s)\,\mathrm{d}s + \int_0^t \sqrt{X_s}\mathrm{d}W_s,$$

for $t \in \mathbb{T}$, with $x_0, \kappa, \theta > 0$. Letting $t \mapsto \varepsilon t$ and denoting $X_t^\varepsilon := X_{\varepsilon t}$ yields, by scaling,

$$X_t^\varepsilon = x_0 + \kappa\varepsilon\int_0^t (\theta - X_s^\varepsilon)\ \mathrm{d}s + \sqrt{\varepsilon}\int_0^t \sqrt{X_s^\varepsilon}\mathrm{d}W_s,$$

which is exactly (2.2.1) with $d=1$, $K\equiv 1$, $\vartheta_\varepsilon = \sqrt{\varepsilon}$, $b_\varepsilon(x) = \kappa\varepsilon(\theta - x)$, $\sigma_\varepsilon(x) = \sqrt{x}$. For $v \in L^2$, taking limits as ε tends to zero in the corresponding controlled equation (2.3.4) yields (2.3.5), or

$$\phi_t = x_0 + \int_0^t \sqrt{\phi_s}v_s\,\mathrm{d}s, \quad t \in \mathbb{T}.$$

Uniqueness of this Volterra equation does not hold in general because of the non-Lipschitz coefficient, and thus $\mathcal{G}_v^0$ corresponds to the set of non-negative solutions. Consider for example $x_0 = 1$, $\mathbb{T} = [0,4]$ and the control

$$v_t := \begin{cases} -1, & \text{if } t \in [0,2) \\ 1, & \text{if } t \in [2,4]. \end{cases}$$

The function $\phi_t := \frac{(t-2)^2}{4}$ is clearly a solution, but so is φ equal to ϕ on $[0,2]$ and null on $[2,4]$. The square root function is indeed locally Lipschitz away from zero, and uniqueness can thus be guaranteed as long as the solution remains positive. The perturbation $\phi^\delta := \phi + \delta t$ is now the unique solution to

$$\phi_t^\delta = 1 + \int_0^t \sqrt{\phi_s^\delta}v_s^\delta\,\mathrm{d}s,$$

for all $t \in [0,4]$, where $v^\delta := \dot{\phi}^\delta/\sqrt{\phi^\delta}$. The infimum in (2.3.2) is attained by v^δ and $\mathcal{G}_{v^\delta}^0 = \{\phi^\delta\}$, thus ϕ^δ is uniquely characterised. Furthermore, [84, Proposition 3.3] shows that ϕ^δ satisfies Assumption 2.3.3.

In [33], the authors were also confronted to a limiting equation with multiple solutions. Instead of perturbing the path ϕ, they perturb the control in a way that the resulting equation has a unique solution which is precisely ϕ, i.e. $\mathcal{G}^0_{v^\delta} = \{\phi\}$. This approach may seem more natural; however, it is not always obvious how to perturb the control ensuring uniqueness of the ODE, while our formulation makes it more straightforward. Before stating the main large and moderate deviations results for small-noise stochastic Volterra equations, we introduce the following assumption, monitoring the moments of the controlled equation:

Assumption 2.3.9. Let $X^{\varepsilon,v}$ be the pathwise unique solution to (2.3.4). If **H3a** holds then the present assumption is satisfied. If instead **H3b** holds, then there exists $\varepsilon_0 > 0$ such that, for any $p \geq 1$ and $M > 0$,

$$\sup\left\{\mathbb{E}\left[\left|\Gamma\left((X^{\varepsilon,v}_t)^{(\Upsilon)}\right)\right|^p\right] : t \in \mathbb{T}, v \in \mathcal{A}_M, \varepsilon \in (0,\varepsilon_0)\right\} < \infty, \tag{2.3.6}$$

$$\sup\left\{\left|\Gamma\left((\phi_t)^{(\Upsilon)}\right)\right| : t \in \mathbb{T}, v \in \mathcal{S}_M, \phi \in \mathcal{G}^0_v\right\} < \infty. \tag{2.3.7}$$

Remark 2.3.10. In the following, **H3b** will always be complemented by Assumption 2.3.9.

2.3.3 Large Deviations

Armed with the abstract setting in Section 2.3.1, and its application to the stochastic Volterra system (2.2.1) in Section 2.3.2, we can at last show large deviations for the latter:

Theorem 2.3.11 (Large Deviations)**.** *Under* **H1 - H4***, Assumptions 2.2.3, 2.3.3 and 2.3.9, the family $\{X^\varepsilon\}_{\varepsilon>0}$, unique solution of* (2.2.1)*, satisfies a Large Deviations Principle with rate function* (2.3.2) *and speed $\vartheta_\varepsilon^{-2}$, where $\mathcal{G}^0_v$ is the set of solutions of the limiting equation* (2.3.5).

Remark 2.3.12. We recall that Assumption 2.3.3 is automatically satisfied if the limiting equation (2.3.5) has a unique solution. Also, it is only necessary to check Assumption 2.3.9 if **H3a** does not hold.

2.3.3.1 Technical preliminary results

The proof will rely on the following results: Lemma 2.3.13 (proved in Section 2.5.2) shows the moment bound of the controlled process defined by (2.3.4), Lemma 2.3.15

(proved in Section 2.5.3) demonstrates the tightness and Lemma 2.3.18 (proved in Section 2.5.4) deals with the compactness of the level sets of the rate function. They will then allow the use of Theorem 2.3.5.

Lemma 2.3.13 (LDP Moment bound)**.** *Under* **H1 - H4***, Assumptions 2.2.3 and 2.3.9, for all $p \geq 2$, $M > 0$, $v \in \mathcal{A}_M$ and $\varepsilon > 0$ small enough, there exists a constant $\bar{c} > 0$ independent of ε, v, t such that*

$$\sup_{t\in\mathbb{T}} \mathbb{E}\left[|X_t^{\varepsilon,v}|^p\right] \leq \bar{c}. \tag{2.3.8}$$

Remark 2.3.14. This bound also holds for any solution ϕ of (2.3.5) under the same assumptions, therefore there also exists $\bar{c}_0 > 0$ such that $\sup\{\|\phi\|_{\mathbb{T}} : v \in \mathcal{S}_M \text{ such that } \phi \in \mathcal{G}_v^0\} \leq \bar{c}_0$.

The following lemma deals with 2.3.11(i) by showing tightness of $X^{\varepsilon,v^\varepsilon} = \mathcal{G}^\varepsilon\left(W + \vartheta_\varepsilon^{-1}\int_0^\cdot v_s^\varepsilon \,\mathrm{d}s\right)$.

Lemma 2.3.15 (LDP Tightness)**.** *Consider* **H1 - H4***, Assumptions 2.2.3 and 2.3.9. If $p > 2\vee 2/\gamma$, $M > 0$ and $\{v^\varepsilon\}_{\varepsilon>0}$ is a family in $\mathcal{A}_M$, then $X^{\varepsilon,v^\varepsilon}$ admits a version which is Hölder continuous on $\mathbb{T}$ of any order $\alpha < \gamma/2 - 1/p$, uniformly for all $\varepsilon > 0$. Denoting again this version by $X^{\varepsilon,v^\varepsilon}$, one has for all $\varepsilon > 0$ small enough*

$$\mathbb{E}\left[\left(\sup_{0\leq s<t\leq T} \frac{\left|X_t^{\varepsilon,v^\varepsilon} - X_s^{\varepsilon,v^\varepsilon}\right|}{|t-s|^\alpha}\right)^p\right] \leq \bar{C}, \tag{2.3.9}$$

for all $\alpha \in [0, \gamma/2 - 1/p)$, where $\bar{C}$ is a constant independent of $\varepsilon, v^\varepsilon, t$. Moreover, the family of random variables $\{X^{\varepsilon,v^\varepsilon}\}_{\varepsilon>0}$ is tight in $\mathcal{W}^d$.

Remark 2.3.16. This lemma entails that for all $M > 0, v \in \mathcal{S}_M$, any solution to (2.3.5) also has Hölder continuous paths of the same order.

Lemma 2.3.17. *The set $\mathcal{G}_v^0$ from Definition 2.3.1 is characterised by*

$$\mathcal{G}_v^0 = \left\{\phi : \Omega \to \mathcal{W}^d \,\middle|\, \phi_t = x_0 + \int_0^t K(t-s)\Big[b(s,\phi_s) + \sigma(s,\phi_s)v_s\Big]\mathrm{d}s, \textit{ for all } t \in \mathbb{T}\right\}.$$

The following lemma proves Theorem 2.3.5(ii) and its proof can be found in Section 2.5.4.

Lemma 2.3.18 (LDP Compactness). *Under* **H2, H3**, *Assumptions 2.2.3 and 2.3.9, the functional I in* (2.3.2) *has compact level sets.*

Leveraging on the above lemmas, the Large Deviations Principle (Theorem 2.3.11) is a direct consequence of Theorem 2.3.5.

2.3.3.2 Proof of Lemma 2.3.17

For $M \in \mathbb{N}$ and $v \in \mathcal{A}_M$, we first need to identify $\mathcal{G}^0_{v,M}$, defined in (2.3.1). Consider a subsequence $\{\varepsilon_n\}_{n\in\mathbb{N}} \subset \mathbb{R}_+$ with $\lim_{n\uparrow\infty} \varepsilon_n = 0$ and a sequence $\{v^{\varepsilon_n}\}_{n\in\mathbb{N}} \subset \mathcal{A}_M$ such that $\lim_{n\uparrow\infty} v^{\varepsilon_n} = v$ in distribution, and assume that $X^n := X^{\varepsilon_n, v^{\varepsilon_n}}$ converges in distribution to some random variable ϕ with values in $\mathcal{W}^d$. We also denote $v^n, \varepsilon_n, X^n_0, b_n, \sigma_n$ along this subsequence.

By Skorohod representation theorem we can work with almost sure convergence for the purpose of identifying the limit. Hence $\{X^n, v^n\}_{n\geq 0}$ converges almost surely in the product topology on $\mathcal{W}^d \times \mathcal{S}_M$, and the limit is the $\mathcal{W}^d \times \mathcal{S}_M$-valued random variable (ϕ, v). The convergence of the couple also takes place in distribution, so that we can follow the technique in [62] to identify the limit. For $t \in \mathbb{T}$, define $\Phi_t : \mathcal{S}_M \times \mathcal{W}^d \to \mathbb{R}$ as

$$\Phi_t(f,\omega) := \left|\omega_t - x_0 - \int_0^t K(t-s)\Big[b(s,\omega_s) + \sigma(s,\omega_s)f_s\Big]\mathrm{d}s\right| \wedge 1.$$

Clearly, Φ_t is bounded and we show that it is also continuous. Indeed, let $\omega^n \to \omega$ in $\mathcal{W}^d$ and $f^n \to f$ in $\mathcal{S}_M$ with respect to the weak topology. **H2** implies the existence of continuous moduli of continuity ρ_b and ρ_σ for both coefficients on compact subsets (see Definition 2.2.1). Since the paths $\omega^n, n \geq 1$ and ω are continuous, they are also uniformly bounded and hence these moduli are available. Then, using Cauchy-Schwarz inequality and the fact that $|x \wedge 1 - y \wedge 1| \leq |x - y|$ for all $x, y > 0$,

$$\begin{aligned}
|\Phi_t(f,\omega) - \Phi_t(f^n,\omega^n)| &\leq |\omega_t - \omega^n_t| + \int_0^t |K(t-s)|\,|b(s,\omega_s) - b(s,\omega^n_s)|\,\mathrm{d}s \\
&\quad + \int_0^t |K(t-s)|\,\big|\big(\sigma(s,\omega_s) - \sigma(s,\omega^n_s)\big)f^n_s + \sigma(s,\omega_s)\big(f_s - f^n_s\big)\big|\,\mathrm{d}s \\
&\leq \|\omega - \omega^n\|_{\mathbb{T}} + \|\rho_b(|\omega-\omega^n|)\|_{\mathbb{T}}\,\|K\|_1 + \|\rho_\sigma(|\omega-\omega^n|)\|_{\mathbb{T}}\,\|K\|_2\,\|f^n\|_2 \\
&\quad + \|\sigma(\cdot,\omega)\|_{\mathbb{T}} \int_0^t |K(t-s)|\,|f_s - f^n_s|\,\mathrm{d}s.
\end{aligned}$$

Since $K(t-\cdot) \in L^2$ and f_n tends to f weakly in L^2 then the last integral converges to zero as n goes to infinity. Moreover $\lim_{n\uparrow\infty} \|\omega - \omega^n\|_{\mathbb{T}} = 0$, $\|f^n\|_2 \leq \sqrt{M}$ for all $n \geq 0$ and $\|K\|_2 + \|\sigma(\cdot,\omega)\|_{\mathbb{T}} < \infty$, which proves that Φ_t is continuous, and therefore

$$\lim_{n\uparrow\infty} \mathbb{E}\left[\Phi_t(v^n, X^n)\right] = \mathbb{E}\left[\Phi_t(v,\phi)\right].$$

We now prove that the left-hand side is actually equal to zero. We start with the observation that, using BDG inequality,

$$\begin{aligned}
\mathbb{E}\left[\Phi_t(v^n, X^n)\right] \leq |X_0^n - x_0| &+ \int_0^t |K(t-s)|\,\mathbb{E}\big[\,|b_n(s,X_s^n) - b(s,X_s^n)|\,\big]\,\mathrm{d}s \\
&+ \int_0^t |K(t-s)|\,\mathbb{E}\left[|\sigma_n(s,X_s^n) - \sigma(s,X_s^n)|\,|v_s^n|\right]\mathrm{d}s \\
&+ \vartheta_{\varepsilon_n}\mathbb{E}\left[\int_0^t |K(t-s)\sigma_n(s,X_s^n)|^2\,\mathrm{d}s\right]^{\frac{1}{2}}. \qquad (2.3.10)
\end{aligned}$$

The bounds (2.5.3) and (2.5.5) show how to control the last term under **H3a** and **H3b** respectively, hence there exists $C_1 > 0$ independent of t and n such that $\mathbb{E}\left[\int_0^t |K(t-s)\sigma_n(s,X_s^n)|^2\,\mathrm{d}s\right]^{\frac{1}{2}} \leq C_1$.

However the convergence of b_n, σ_n only occurs on compact subsets so we use a localisation argument. For all $n \geq 0$, $R > 0$ we introduce

$$A_n^R := \Big\{\omega \in \Omega : \|X^n(\omega)\|_{\mathbb{T}} > R\Big\}. \qquad (2.3.11)$$

The uniform (in $n \in \mathbb{N}$) Hölder regularity of X^n, encompassed by (2.3.9), entails the existence, for all $p > 2 \vee 2/\gamma$, of $C_2(p), C_3(p) > 0$ independent of n such that

$$\mathbb{E}\left[\sup_{t\in\mathbb{T}} |X_t^n|^p\right] \leq C_2(p)\big(\,|X_0^n|^p + T^{p\alpha}\big) \leq C_3(p),$$

for some $0 < \alpha < \gamma/2 - 1/p$ and where X_0^n is uniformly bounded by $2\,|x_0|$ for n large enough. Markov's inequality then implies that

$$\lim_{R\uparrow\infty} \sup_{n\in\mathbb{N}} \mathbb{P}\big(A_n^R\big) \leq \lim_{R\uparrow\infty} \sup_{n\in\mathbb{N}} \frac{C_3(p)}{R^p} = 0. \qquad (2.3.12)$$

Moreover, for all $n \in \mathbb{N}$, $\omega \in \Omega \setminus A_n^R$, and $t \in \mathbb{T}$, $|X_t^n(\omega)|$ is bounded by R, which means

$$|b_n(t, X_t^n(\omega)) - b(t, X_t^n(\omega))| \leq \|b_n(t,\cdot) - b(t,\cdot)\|_R\,,$$

which tends to zero uniformly on $\mathbb{T}$ as n goes to infinity (and likewise for σ_n) from **H2**.

Define now

$$I_n := \int_0^t |K(t-s)| \Big(|b_n(s, X_s^n) - b(s, X_s^n)| + |\sigma_n(s, X_s^n) - \sigma(s, X_s^n)| \, |v_s^n| \Big) \, \mathrm{d}s$$

and observe that, using Jensen and Cauchy-Schwarz inequalities, the growth condition on the coefficients from **H3** and the moment bounds on X^n from (2.3.8), there exists $C_4 > 0$ independent of n such that

$$\begin{aligned}
\mathbb{E}\big[\,|I_n|^2\,\big] &\leq 2t \int_0^t |K(t-s)|^2 \, \mathbb{E}\big[\,|b_n(s, X_s^n) - b(s, X_s^n)|^2\,\big] \, \mathrm{d}s \\
&\quad + 2M \int_0^t |K(t-s)|^2 \, \mathbb{E}\left[|\sigma_n(s, X_s^n) - \sigma(s, X_s^n)|^2\right] \, \mathrm{d}s \\
&\leq 2 \, \|K\|_2^2 \sup_{s \leq t} \left\{ t\mathbb{E}\big[\,|b_n(s, X_s^n) - b(s, X_s^n)|^2\,\big] + M\mathbb{E}\left[|\sigma_n(s, X_s^n) - \sigma(s, X_s^n)|^2\right] \right\} \\
&\leq C_4.
\end{aligned} \tag{2.3.13}$$

Let us fix $\epsilon > 0$ and choose $R_\epsilon > 0$ large enough such that $\sup_{n \in \mathbb{N}} \mathbb{P}\big(A_n^{R_\epsilon}\big) \leq \epsilon^2 / C_4$; this choice is possible because of (2.3.12). Therefore, using the bound (2.3.13) and Cauchy-Schwarz inequality to separate I_n and $\mathbb{1}_{A_n^{R_\varepsilon}}$, one obtains

$$\begin{aligned}
\limsup_{n \uparrow \infty} \mathbb{E}[I_n] &= \limsup_{n \uparrow \infty} \mathbb{E}\Big[I_n \big(\mathbb{1}_{A_n^{R_\varepsilon}} + \mathbb{1}_{\Omega \setminus A_n^{R_\varepsilon}} \big) \Big] \\
&\leq \limsup_{n \uparrow \infty} \left\{ \sqrt{C_4 \mathbb{P}\big(A_n^{R_\varepsilon}\big)} + \|K\|_1 \, \big\| \|b_n - b\|_{R_\varepsilon} \big\|_{\mathbb{T}} + \sqrt{M} \, \|K\|_2 \, \big\| \|\sigma_n - \sigma\|_{R_\varepsilon} \big\|_{\mathbb{T}} \right\} \leq \epsilon.
\end{aligned}$$

It follows from (2.3.10) that

$$\lim_{n \uparrow \infty} \mathbb{E}\big[\Phi_t(v^n, X^n)\big] \leq \lim_{n \uparrow \infty} \Big\{ |X_0^n - x_0| + \mathbb{E}[I_n] + \vartheta_{\varepsilon_n} C_1 \Big\} \leq \epsilon,$$

hence $\lim_{n \uparrow \infty} \mathbb{E}\left[\Phi_t(v^n, X^n)\right] = 0$ since $\epsilon > 0$ was chosen arbitrarily. The equality $\mathbb{E}\left[\Phi_t(v, \phi)\right] = 0$ implies that ϕ satisfies (2.3.5) almost surely, for all $t \in \mathbb{T}$. Since ϕ has continuous paths, it satisfies (2.3.5) for all $t \in \mathbb{T}$, almost surely, which means $\mathcal{G}_{v,M}^0$ consists of all the solutions of (2.3.5). Since this definition is independent of M, it extends to $\mathcal{G}_v^0$, which yields the claim. □

2.3.4 Moderate Deviations

Let h_ε tend to infinity such that $\vartheta_\varepsilon h_\varepsilon$ tends to zero as ε goes to zero and define $\overline{X}$ to be the limit in law of X^ε, which we identified in the previous subsection as a solution of the Volterra equation

$$\overline{X}_t = x_0 + \int_0^t K(t-s)b(s, \overline{X}_s)\mathrm{d}s. \tag{2.3.14}$$

Then the MDP for $\{X^\varepsilon\}_{\varepsilon>0}$ is equivalent to the LDP for the family $\{\eta^\varepsilon\}_{\varepsilon>0}$ defined as

$$\eta^\varepsilon := \frac{X^\varepsilon - \overline{X}}{\vartheta_\varepsilon h_\varepsilon} = \frac{\mathcal{G}^\varepsilon(W) - \overline{X}}{\vartheta_\varepsilon h_\varepsilon} =: \mathcal{T}^\varepsilon(W),$$

where $\mathcal{T}^\varepsilon : \mathcal{W}^m \to \mathcal{W}^d$ is a Borel-measurable map for each $\varepsilon > 0$. Therefore η^ε satisfies the following SVE for all $\varepsilon > 0$, and is its unique solution if **H4** holds.

$$\eta_t^\varepsilon = \frac{X_0^\varepsilon - x_0}{\vartheta_\varepsilon h_\varepsilon} + \int_0^t K(t-s)\frac{b_\varepsilon\big(s, \overline{X}_s + \vartheta_\varepsilon h_\varepsilon \eta_s^\varepsilon\big) - b(s, \overline{X}_s)}{\vartheta_\varepsilon h_\varepsilon}\mathrm{d}s + \int_0^t K(t-s)\frac{\sigma_\varepsilon\big(s, \overline{X}_s + \vartheta_\varepsilon h_\varepsilon \eta_s^\varepsilon\big)}{h_\varepsilon}\mathrm{d}W_s. \tag{2.3.15}$$

Similarly to the LDP case we are interested in a certain shift of the driving Brownian motion, controlled by $v \in \mathcal{A}$. For all $\varepsilon > 0$, let

$$\eta^{\varepsilon,v} := \mathcal{T}^\varepsilon\left(W + h_\varepsilon \int_0^\cdot v_s \,\mathrm{d}s\right) = \frac{\mathcal{G}^\varepsilon\big(W + h_\varepsilon \int_0^\cdot v_s \,\mathrm{d}s\big) - \overline{X}}{\vartheta_\varepsilon h_\varepsilon}. \tag{2.3.16}$$

For convenience we introduce the sequence $\{\Theta^{\varepsilon,v}\}_{\varepsilon>0}$ defined, for all $\varepsilon > 0$, $v \in \mathcal{A}$, $t \in \mathbb{T}$, by

$$\begin{aligned}
\Theta_t^{\varepsilon,v} :=& \mathcal{G}^\varepsilon\left(W + h_\varepsilon \int_0^\cdot v_s \,\mathrm{d}s\right)(t) \\
=& X_0^\varepsilon + \int_0^t K(t-s)\Big[b_\varepsilon(s, \Theta_s^{\varepsilon,v}) + \vartheta_\varepsilon h_\varepsilon \sigma_\varepsilon(s, \Theta_s^{\varepsilon,v})v_s\Big]\mathrm{d}s + \vartheta_\varepsilon \int_0^t K(t-s)\sigma_\varepsilon(s, \Theta_s^{\varepsilon,v})\mathrm{d}W_s.
\end{aligned}$$

This sequence satisfies the bound (2.3.8) and converges weakly towards $\overline{X}$ since $\vartheta_\varepsilon h_\varepsilon \sigma_\varepsilon$ tends to zero as ε tends to zero. Finally the process defined by (2.3.16) satisfies

$$\begin{aligned}
\eta_t^{\varepsilon,v} = & \frac{X_0^\varepsilon - x_0}{\vartheta_\varepsilon h_\varepsilon} + \int_0^t K(t-s)\frac{b_\varepsilon\big(s, \Theta_s^{\varepsilon,v}\big) - b(s, \overline{X}_s)}{\vartheta_\varepsilon h_\varepsilon}\mathrm{d}s \\
& + \int_0^t K(t-s)\sigma_\varepsilon\big(s, \Theta_s^{\varepsilon,v}\big)v_s\mathrm{d}s + \frac{1}{h_\varepsilon}\int_0^t K(t-s)\sigma_\varepsilon\big(s, \Theta_s^{\varepsilon,v}\big)\mathrm{d}W_s.
\end{aligned} \tag{2.3.17}$$

For all $v \in \mathcal{A}$, we define $\mathcal{T}_v^0$ to be the solution of the limiting equation

$$\psi_t = \int_0^t K(t-s)\Big[\nabla b(s, \overline{X}_s)\psi_s + \sigma(s, \overline{X}_s)v_s\Big]\,\mathrm{d}s. \tag{2.3.18}$$

The form of the limit equation is dramatically simpler than for the LDP and much easier to compute. Moreover $\mathcal{T}_v^0$ is well defined because the linearity of the equation and Assumption 2.2.3 grant uniqueness for free, provided ∇b exists. Hence we will need the following assumptions:

H5. For each $t \in \mathbb{T}$, the function $b(t, \cdot)$ is continuously differentiable and b is Lipschitz continuous.

H6. There exists $\delta > 0$ such that $\sigma(t, \cdot)$ is locally δ-Hölder continuous, uniformly for all $t \in \mathbb{T}$.

H7. $\lim_{\varepsilon \downarrow 0} \left(\vartheta_\varepsilon h_\varepsilon\right)^{-1} |X_0^\varepsilon - x_0| = 0$.

H8. There exist $\varepsilon_0 > 0$, a sequence $\{\nu_\varepsilon\}_{\varepsilon>0}$ with $\lim_{\varepsilon \downarrow 0} \nu_\varepsilon(\vartheta_\varepsilon h_\varepsilon)^{-1} = 0$ and a function $\Xi : \mathbb{R}^d \to \mathbb{R}$ such that $|b_\varepsilon(t, x) - b(t, x)| \leq \nu_\varepsilon \Xi(x)$ for all $t \in \mathbb{T}$, $\varepsilon \in (0, \varepsilon_0)$, where for all $p \geq 1$, $M > 0$,

$$\sup\left\{\mathbb{E}\Big[\left|\Xi(\Theta_t^{\varepsilon,v})\right|^p\Big], \varepsilon \in (0, \varepsilon_0), v \in \mathcal{A}_M, t \in \mathbb{T}\right\} < \infty. \tag{2.3.19}$$

Remark 2.3.19. **H5** entails that (2.3.14) has a unique solution and yields the bound $\left\|\nabla b(\cdot, \overline{X})\right\|_{\mathbb{T}} < \infty$ by continuity. **H7** implies **H1**. We have already proved in Lemma 2.3.8 that the moments of all orders of $\Theta^{\varepsilon,v}$ are bounded hence (2.3.19) is automatically satisfied if Ξ is of polynomial growth. This is however not sufficient for the applications we have in mind where Ξ is of exponential growth.

The main theorem of this section is the following.

Theorem 2.3.20 (Moderate Deviations)**.** *Under* **H2 - H8***, Assumptions 2.2.3 and 2.3.9, the family $\{\eta^\varepsilon\}_{\varepsilon>0}$ satisfies a Large Deviations Principle (equivalently $\{X^\varepsilon\}_{\varepsilon>0}$ satisfies a Moderate Deviations Principle) with speed h_ε^2 and rate function*

$$\Lambda(\psi) := \inf\left\{\frac{1}{2}\int_0^T |v_t|^2 \, \mathrm{d}t : v \in L^2, \psi = \mathcal{T}_v^0\right\}, \tag{2.3.20}$$

and $\Lambda(\psi) = +\infty$ if this set is empty.

The proof of the moderate deviations theorem follows a similar structure to that of Theorem 2.3.11, making use of Theorem 2.3.5. It will rely on moment bounds in Lemma 2.3.21 (proved in Section 2.6.1), tightness in Lemma 2.3.22 (proved in Section 2.6.2), weak convergence in Lemma 2.3.23 (proved in Section 2.6.3), and finally compactness of the level sets in Lemma 2.3.24.

Lemma 2.3.21 (MDP Moment bound). *Under* **H2 - H5, H7, H8**, *Assumptions 2.2.3 and 2.3.9, for all* $p \geq 2$, $M > 0$, $v \in \mathcal{A}_M$ *and* $\varepsilon > 0$ *small enough, there exists* $\widehat{c} > 0$ *independent of* ε, v, t *such that*

$$\sup_{t \in \mathbb{T}} \mathbb{E}\left[|\eta_t^{\varepsilon,v}|^p\right] \leq \widehat{c}. \tag{2.3.21}$$

Lemma 2.3.22 (MDP tightness). *Let* $p > 2 \vee 2/\gamma$, $M > 0$ *and a family* $\{v^\varepsilon\}_{\varepsilon>0}$ *in* $\mathcal{A}_M$*. Under* **H2 - H5, H7, H8**, *Assumptions 2.2.3 and 2.3.9,* $\eta^{\varepsilon,v^\varepsilon}$ *admits a version which is Hölder continuous on* $\mathbb{T}$ *of any order* $\alpha < \gamma/2 - 1/p$*, uniformly for all* $\varepsilon > 0$*. Denoting again this version by* $\eta^{\varepsilon,v^\varepsilon}$*, one has for all* $\varepsilon > 0$ *small enough,*

$$\mathbb{E}\left[\left(\sup_{0 \leq s < t \leq T} \frac{\left|\eta_t^{\varepsilon,v^\varepsilon} - \eta_s^{\varepsilon,v^\varepsilon}\right|}{|t-s|^\alpha}\right)^p\right] \leq \widehat{C}, \tag{2.3.22}$$

for all $\alpha \in [0, \gamma/2 - 1/p)$*, where* $\widehat{C}$ *is a constant independent of* $\varepsilon, v^\varepsilon, s, t$*. Moreover, the family* $\{\eta^{\varepsilon,v^\varepsilon}\}_{\varepsilon>0}$ *is tight in* $\mathcal{W}^d$*.*

We recall that $\eta^{\varepsilon,v^\varepsilon} = \mathcal{T}^\varepsilon\left(W + h_\varepsilon \int_0^\cdot v_s \, \mathrm{d}s\right)$ and $\psi = \mathcal{T}_v^0$, hence the lemma above deals with Theorem 2.3.5 *(i)*. The following one identifies the limit set as the unique solution to (2.3.18). It is thus more precise than in the LDP case, and justifies the form of the rate function (2.3.20).

Lemma 2.3.23 (MDP weak convergence). *Let* $M > 0$*, a family* $\{v^\varepsilon\}_{\varepsilon>0}$ *such that, for all* $\varepsilon > 0$, $v^\varepsilon \in \mathcal{A}_M$ *and* v^ε *converges in distribution to* $v \in \mathcal{A}_M$*, and* ψ *the unique solution of* (2.3.18)*. Under* **H2 - H8**, *Assumptions 2.2.3 and 2.3.9,* $\eta^{\varepsilon,v^\varepsilon}$ *converges in distribution to* ψ *as* ε *goes to zero.*

Item *(ii)* is dealt with in the following lemma.

Lemma 2.3.24 (MDP compactness). *Under* **H2, H3, H5**, *Assumptions 2.2.3 and 2.3.9, the functional* Λ *defined by* (2.3.20) *has compact level sets.*

Proof. Noticing that $\nabla b(t, \overline{X}_t)$ and $\sigma(t, \overline{X}_t)$ are uniformly bounded on $\mathbb{T}$ by continuity, this lemma boils down to a particular case of Lemma 2.3.18. □

Theorem 2.3.5(iii) is immediate by uniqueness of (2.3.18), therefore all the conditions are met and Theorem 2.3.20 follows as a direct application of Theorem 2.3.5.

2.3.5 Extension to non-convolution kernels

The analysis undertaken in this paper is based, both for notational convenience and with a view towards application, on convolution kernels. Different assumptions were studied in the literature, in particular Decreusefond [79] considered the properties of the map $f \mapsto \int_0^\cdot K(\cdot, s) f(s)\,\mathrm{d}s$ in order to include the fractional Brownian motion in his setting.

2.3.5.1 Setting

We call a kernel a map $K : \mathbb{T}^2 \to \mathbb{R}$ for which both $\int_0^t K(t,s)^2 \mathrm{d}s$ and $K(t,s)$ are finite for all $t \in \mathbb{T}$ and $s \neq t$. The associated space is defined as

$$\mathcal{K} := \left\{ u : \mathbb{T} \to \mathbb{R},\ \{\mathcal{F}_t\}\text{-progressively measurable, such that } \mathbb{E} \int_0^t \big[K(t,s)u(s)\big]^2 \mathrm{d}s < \infty,\ \text{for all } t \in \mathbb{T} \right\}.$$

Hence, for all $u \in \mathcal{K}$ the stochastic integral

$$\widetilde{M}_t^K(u) := \int_0^t K(t,s)u(s)\mathrm{d}W_s$$

is well defined for all $t \in \mathbb{T}$ in the Itô sense. For any $\alpha \in (0,1)$, we denote the Riemann-Liouville integral I^α and derivative D^α as

$$(\mathrm{I}^\alpha f)(t) := \frac{1}{\Gamma(\alpha)} \int_0^t (t-s)^{\alpha-1} f(s)\mathrm{d}s, \qquad (\mathrm{D}^\alpha f)(t) := \frac{\mathrm{d}}{\mathrm{d}t}(\mathrm{I}^{1-\alpha} f)(t), \quad \text{for } f \in L^1, t \in \mathbb{T}. \tag{2.3.23}$$

Define $\mathcal{I}_{\alpha,p} := \mathrm{I}^\alpha(L^p)$ equipped with the norm $\|f\|_{\mathcal{I}_{\alpha,p}} := \|D^\alpha f\|_{L^p}$. If $\alpha > \frac{1}{p}$, then $\mathcal{I}_{\alpha,p} \subset \mathcal{C}_0^{\alpha-\frac{1}{p}}$, the space of $(\alpha - \frac{1}{p})$-Hölder continuous functions null at time 0. Let K denote the linear map associated to $K(t,s)$ by

$$\mathrm{K}f(t) := \int_0^t K(t,s) f(s)\mathrm{d}s, \tag{2.3.24}$$

and introduce, for $x \in (0,\infty) \setminus \{2\}$,

$$\theta(x) := \frac{2x}{2-x}. \tag{2.3.25}$$

Given the space inclusions above, the following assumption implies precise Hölder regularity for the integral (2.3.24):

Assumption 2.3.25. There exist $\chi \in (1, 2)$ and $\gamma > 1/\theta(\chi)$ for which K is continuous from L^2 to $\mathcal{I}_{\gamma+\frac{1}{2},2}$ and from L^χ to $\mathcal{I}_{\gamma,\theta(\chi)}$.

Example 2.3.26. *The operators associated to the following kernels satisfy Assumption 2.3.25:*

- *The Riemann-Liouville kernel*

$$J_H(t, s) = \frac{(t - s)_+^{H-\frac{1}{2}}}{\Gamma(H + \frac{1}{2})}, \qquad \text{with } H \in (0, 1),$$

  satisfies this assumption with $\gamma = H$ and any $\chi < 2$ [79, Theorem 4.1].

- *The fractional Brownian motion kernel*

$$K_H(t, s) = \frac{(t - s)_+^{H-\frac{1}{2}}}{\Gamma(H + \frac{1}{2})} F\left(H - \frac{1}{2}, \frac{1}{2} - H, H + \frac{1}{2}, 1 - \frac{t}{s}\right),$$

  where F is the Gauss hypergeometric function, also satisfies this assumption with the same parameters as above [79, Theorem 4.2].

Decreusefond's main result yields the Hölder regularity of the stochastic Volterra integral [79, Theorem 3.1]:

Theorem 2.3.27. *Let Assumption 2.3.25 hold and $u \in \mathcal{K} \cap L^{\theta(\chi)}(\Omega \times \mathbb{T})$. Then $\widetilde{M}^K(u)$ has a measurable version $M^K(u)$ which is α-Hölder continuous for all $\alpha < \gamma - 1/\theta(\chi)$.*

From now on, we only consider the measurable version of the stochastic integral. Although this theorem was proved in a one-dimensional setting, it also covers multi-dimensional stochastic Volterra integrals by considering their components individually and summing them.

2.3.5.2 Large and moderate deviations

For each $\varepsilon > 0$ consider the stochastic Volterra equation

$$X_t^\varepsilon = X_0^\varepsilon + \int_0^t K(t, s) b_\varepsilon(s, X_s^\varepsilon)\, \mathrm{d}s + \vartheta_\varepsilon \int_0^t K(t, s) \sigma_\varepsilon(s, X_s^\varepsilon) \mathrm{d}W_s, \quad t \in \mathbb{T}, \qquad (2.3.26)$$

which was studied in [71] without the ε-dependence, and where the coefficients live in the same spaces as those from (2.2.1). To complete the non-convolution setup we also need the following condition.

Assumption 2.3.28. There exists $q > 2$ such that

$$\sup_{t\in\mathbb{T}} \left\{ \int_0^t |K(t,s)|^{-\theta(q)} \,\mathrm{d}s \right\} < \infty, \tag{2.3.27}$$

with $\theta(q)$ introduced in (2.3.25).

Let $p > q > 2$, then $2 < -\theta(p) < -\theta(q)$ and Hölder's and Jensen's inequalities yield

$$\begin{aligned}\left[\int_0^t |K(t,s)f(s)|^2 \,\mathrm{d}s\right]^{\frac{p}{2}} &\le \left[\int_0^t |K(t,s)|^{-\theta(p)} \,\mathrm{d}s\right]^{\frac{p-2}{2}} \int_0^t |f(s)|^p \,\mathrm{d}s \\ &\le t^{\frac{p-q}{q}} \left[\int_0^t |K(t,s)|^{-\theta(q)} \,\mathrm{d}s\right]^{\frac{p}{-\theta(q)}} \int_0^t |f(s)|^p \,\mathrm{d}s.\end{aligned}$$

This replaces the Gronwall-type inequality derived for convolution kernels in Lemma 2.5.1. Hence replacing Assumption 2.2.3 by the condition (2.3.27) one recovers the moments bounds of Lemmata 2.3.13 and 2.3.21 for the processes $\{X^{\varepsilon,v}\}_{\varepsilon>0}$ and $\{\eta^{\varepsilon,v}\}_{\varepsilon>0}$ and for any $p \ge 1$. Setting in particular $p = \theta(\chi)$ from Theorem 2.3.27, then for any $v \in \mathcal{A}_M$, $M > 0$ and $\xi \in \{b_\varepsilon, \sigma_\varepsilon\}$ we have $\xi(X^{\varepsilon,v}) \in \mathcal{K} \cap L^p(\Omega\times\mathbb{T})$ thanks to the growth conditions **H3**. Therefore Assumption 2.3.25, Theorem 2.3.27 and Assumption 2.3.28 yield the almost sure Hölder regularity of the following processes defined on $\mathbb{T}$:

$$\int_0^\cdot K(\cdot,s)\xi(X_s^{\varepsilon,v})\,\mathrm{d}s \qquad \text{and} \qquad \int_0^\cdot K(\cdot,s)\xi(X_s^{\varepsilon,v})\,\mathrm{d}W_s.$$

Hence we recover the tightness of Lemmata 2.3.15 and 2.3.22 under this new set of assumptions. Notice that we can consider kernels consisting of both convolution and non-convolution components. Finally we can extend the LDP and MDP results without further modifications:

Theorem 2.3.29. *The conclusions of Theorems 2.3.11 and 2.3.20 stand for $\{X^\varepsilon\}_{\varepsilon>0}$ defined by (2.3.26) when each component of the kernel K satisfies either Assumption 2.2.3 or Assumptions 2.3.25 and 2.3.28.*

2.4 Application to rough volatility

We now show how our results (Theorems 2.3.11, 2.3.20 and 2.3.29) apply to a large class of models recently developed in mathematical finance. Originally proposed by Comte and Renault [66] with financial econometrics applications in mind, rough volatility models were rediscovered later in the context of option pricing in [8, 19, 112, 123], developed and extended widely, and have now become the new standards of volatility modelling. They usually take the following form:

$$\begin{cases} X_t &= -\frac{1}{2}\int_0^t \Sigma(Y_s)\,\mathrm{d}s + \int_0^t \sqrt{\Sigma(Y_s)}\mathrm{d}B_s, \\ Y_t &= y_0 + \int_0^t K_1(t-s)\mathfrak{b}(Y_s)\,\mathrm{d}s + \int_0^t K_2(t-s)\zeta(Y_s)\mathrm{d}W_s, \end{cases} \tag{2.4.1}$$

where both X and Y are one-dimensional, $K_1, K_2 \in L^2(\mathbb{T} : \mathbb{R}_+)$ and B and W are two standard Brownian motions with $\mathrm{d}\langle B, W\rangle_t = \rho\,\mathrm{d}t$, for some correlation parameter $\rho \in (-1, 1)$. We further define $\overline{\rho} := \sqrt{1-\rho^2}$, and set $X_0 = 0$ without loss of generality. Here X denotes the logarithm of a stock price process, and $\sqrt{\Sigma(Y)}$ its instantaneous volatility. We adopt a slight abuse of notation, as X previously denoted the multi-dimensional system, but writing now X as the log-stock price is consistent with the mathematical finance literature and should not create any confusion. We summarise in Table 2.1 the most common rough volatility models used in mathematical finance, indicating where their asymptotic behaviours were covered, and where our framework not only encompasses those, but fills the gaps so far missing. As discussed below, our application to the rough Heston model is conditional on the latter to have a unique pathwise solution, a problem that remains open so far. The detailed analysis of these cases is then provided in Section 2.4.2 in the small-time case, and in Section 2.4.3 for their tail behaviours.

2.4.1 Small-time rescaling (general)

In the small-time case, we need to assume some scaling behaviour for the kernel functions. We say that a function $f : \mathbb{R} \to \mathbb{R}$ is homogeneous of degree $\alpha \in \mathbb{R}$ if $f(\lambda x) = \lambda^\alpha f(x)$ holds for all $x, \lambda \in \mathbb{R}$.

Assumption 2.4.1. K_1 and K_2 are homogeneous of degrees $\varpi \in (-\frac{1}{2}, \frac{1}{2}]$ and $H - \frac{1}{2} \in (-\frac{1}{2}, \frac{1}{2}]$.

Models	Rough Stein-Stein				Rough Bergomi		multi-factor rough Bergomi		Rough Heston		
	Small-time		Tail		Small-time		Small-time		Small-time [100]		Tail
	LDP [151]	MDP	LDP [151]	MDP	LDP [161]	MDP	LDP [173]	MDP	LDP	MDP	LDP
$(X^\varepsilon, Y^\varepsilon)$	IF	IF	IF	-	IF	IF	IF	IF	IF	IF	IF
X^ε	OP	IF	OP	-	OP	IF	OP	OP	OP	IF	OP
X_ε	OP	CF	OP	-	OP	CF	OP	OP	OP	CF	OP
$\widehat{\sigma}$	OP	CF	OP	-	OP	CF	OP	OP	OP	CF	OP
Y^ε	IF	IF	IF	IF	IF	IF	IF	IF	IF	IF	IF
Y_ε	CF	CF	IF	IF	CF	CF	CF	CF	IF	CF	IF

Table 2.1: Summary of rough volatility results and form of the rate functions (CF=closed-form; IF=integral form; OP=optimisation problem; shadowed cells are new contributions from this paper). $\widehat{\sigma}$ corresponds to the implied volatility, defined precisely in Section 2.4.1.3.

Since K_1 is homogeneous of degree ϖ, then

$$\begin{aligned} \int_0^h K_1(t)^2\,\mathrm{d}t &= \int_0^h K_1(1)^2 t^{2\varpi}\,\mathrm{d}t \leq \frac{K_1(1)^2}{1+2\varpi} h^{1+2\varpi}, && \text{for any } h>0,\\ \int_0^T \left(K_1(t+h)-K_1(t)\right)^2\mathrm{d}t &= K_1(1)^2 \int_0^T \left((t+h)^\varpi - t^\varpi\right)^2 \mathrm{d}t = \mathcal{O}\left(h^{1+2\varpi}\right), && \text{for } h \text{ small enough,} \end{aligned}$$

and so Assumption 2.2.3 is satisfied with $\gamma = 1+2\varpi \in (0,2]$, and likewise for K_2 with $\gamma = 2H$. Under this assumption, the rescalings $X_t^\varepsilon := \varepsilon^{H-\frac{1}{2}} X_{\varepsilon t}$ and $Y_t^\varepsilon := Y_{\varepsilon t}$ turn (2.4.1) into

$$\left\{ \begin{aligned} X_t^\varepsilon &= -\frac{\varepsilon^{H+\frac{1}{2}}}{2}\int_0^t \Sigma(Y_s^\varepsilon)\,\mathrm{d}s + \varepsilon^H \int_0^t \sqrt{\Sigma(Y_s^\varepsilon)}\mathrm{d}B_s,\\ Y_t^\varepsilon &= y_0 + \varepsilon^{1+\varpi}\int_0^t K_1(t-s)\mathfrak{b}(Y_s^\varepsilon)\,\mathrm{d}s + \varepsilon^H \int_0^t K_2(t-s)\zeta(Y_s^\varepsilon)\mathrm{d}W_s, \end{aligned}\right. \tag{2.4.2}$$

so that we are precisely in the framework of (2.2.1) with $d = 3$, $\vartheta_\varepsilon = \varepsilon^H$,

$$K(t) = \begin{pmatrix} 1 & 0 & 0 \\ 0 & K_1(t) & K_2(t) \\ 0 & 0 & 0 \end{pmatrix}, \quad b_\varepsilon(t,(x,y)) = \begin{pmatrix} -\frac{1}{2}\varepsilon^{H+\frac{1}{2}}\Sigma(y) \\ \varepsilon^{1+\varpi}\mathfrak{b}(y) \\ 0 \end{pmatrix}, \quad \sigma_\varepsilon(t,(x,y)) = \begin{pmatrix} \rho\sqrt{\Sigma(y)} & \overline{\rho}\sqrt{\Sigma(y)} & 0 \\ 0 & 0 & 0 \\ \zeta(y) & 0 & 0 \end{pmatrix},$$

where, similarly to Example 2.2.7, the additional dimension allows to handle the two different kernels. Note that σ_ε does not depend on ε but encodes the correlation. The controlled equation (2.3.4) for the second component reads

$$Y_t^{\varepsilon,v} = y_0 + \varepsilon^{1+\varpi}\int_0^t K_1(t-s)\mathfrak{b}(Y_s^{\varepsilon,v})\,\mathrm{d}s + \varepsilon^H\int_0^t K_2(t-s)\zeta(Y_s^{\varepsilon,v})\mathrm{d}W_s + \int_0^t K_2(t-s)\zeta(Y_s^{\varepsilon,v})v_s\,\mathrm{d}s,$$

for each $t \in \mathbb{T}, \varepsilon > 0$ and $v \in \mathcal{A}$. Note that the dynamics of X^ε do not feed back into Y^ε and that $\Sigma \in \mathcal{S}_{\{2\}}^{|\Sigma|}$ in the sense of Definition 2.2.5. The following assumption stands throughout this section:

Assumption 2.4.2 (Small-time assumptions)**.**

- $K_2(t) := t^{H-\frac{1}{2}}/\Gamma(H+\frac{1}{2})$;
- Σ, ζ and $\mathfrak{b}$ are continuous on $\mathbb{R}$;
- $\mathfrak{b}$ and ζ are of linear growth;
- Σ is either of linear growth or such that for all $p \geq 1$, $M > 0$ and $\varepsilon > 0$ small enough,
$$\sup_{t\in\mathbb{T}, v\in\mathcal{A}_M} \mathbb{E}\Big[\big|\Sigma\big(Y_t^{\varepsilon,v}\big)\big|^p\Big] < \infty; \tag{2.4.3}$$
- the equation for Y^ε in (2.4.2) is pathwise unique for small enough $\varepsilon > 0$.

The choice of kernel K_2 is a common setup in rough volatility models and allow for more explicit results. These conditions ensure that **H2** holds with limit coefficients $b \equiv (0,0,0)^\top$ and $\sigma = \sigma_\varepsilon$. Furthermore, Y^ε is an autonomous subsystem in the sense of Definition 2.2.6 and **H3** and the bound (2.3.6) hold. An pathwise unique solution of the system (2.4.1) exists since X^ε is explicit from Y^ε, and **H4** is satisfied.

2.4.1.1 Large deviations

For each control $v \in \mathcal{S}_M$ with $M > 0$, the limit equation (2.3.5) of the volatility in the large deviations regime reads

$$\varphi_t = y_0 + \int_0^t \frac{(t-s)^{H-\frac{1}{2}}}{\Gamma(H+\frac{1}{2})} \zeta(\varphi_s) v_s \, \mathrm{d}s, \qquad \text{for } t \in \mathbb{T}. \tag{2.4.4}$$

From the uniform bound on φ derived in Remark 2.3.14 and the continuity of Σ, we obtain that $|\Sigma(\varphi_t)|$ is uniformly bounded in $t \in \mathbb{T}$ and in $v \in \mathcal{S}_M$, hence (2.3.7) holds. Therefore Assumption 2.3.9 and **H1 - H4** follow from Assumption 2.4.2. Mimicking the fractional integral notation from Section 2.3.5, we introduce for convenience the notations

$$\mathrm{I}_x^{H+\frac{1}{2}}(f) := x + \int_0^t \frac{(t-s)^{H-\frac{1}{2}}}{\Gamma(H+\frac{1}{2})} f_s \, \mathrm{d}s \qquad \text{and} \qquad \mathrm{I}_x^{H+\frac{1}{2}}(L^1) := \left\{ \mathrm{I}_x^{H+\frac{1}{2}}(f), f \in L^1 \right\},$$

and the fractional derivative D is defined in (2.3.23). From now on, to simplify the statements, we write $Z^\varepsilon \sim \mathrm{LDP}(I, \varepsilon^{-1})$ to express that the family of random variables $\{Z^\varepsilon\}_{\varepsilon>0}$ satisfies an LDP with rate function I and speed ε^{-1}, as ε tends to zero.

Proposition 2.4.3 (Large deviations). *Under Assumptions 2.3.3, 2.4.1 and 2.4.2, the following hold:*

(L1) $(X^\varepsilon, Y^\varepsilon) \sim \mathrm{LDP}\left(I, \varepsilon^{-2H}\right)$, *where* $I : \mathcal{W}^2 \to \overline{\mathbb{R}}_+$ *is given by*

$$I(\phi, \varphi) = \inf \left\{ \frac{1}{2} \int_0^T \left(u_t^2 + v_t^2\right) \mathrm{d}t : u, v \in L^2, \phi_t = \int_0^t \sqrt{\Sigma(\varphi_s)} \left(\bar{\rho} u_s + \rho v_s\right) \mathrm{d}s, \; \varphi = \mathrm{I}_{y_0}^{H+\frac{1}{2}}(\zeta(\varphi) v) \right\};$$

(L2) $X^\varepsilon \sim \mathrm{LDP}\left(I^X, \varepsilon^{-2H}\right)$, *where*

$$I^X(\phi) = \inf \left\{ I(\phi, \varphi) : \varphi \in \mathrm{I}_{y_0}^{H+\frac{1}{2}}(L^1) \right\},$$

if $\phi \in \mathcal{AC}_0$ *and infinity otherwise;*

(L3) $\varepsilon^{H-\frac{1}{2}} X_\varepsilon \sim \mathrm{LDP}\left(I_1^X, \varepsilon^{-2H}\right)$ *where* $I_1^X(x) = \inf\left\{I^X(\phi) : \phi_1 = x\right\}$ *for all* $x \in \mathbb{R}$*;*

(L4) $Y^\varepsilon \sim \mathrm{LDP}\left(I^Y, \varepsilon^{-2H}\right)$, *where*

$$I^Y(\varphi) = \frac{1}{2} \int_0^T \left(\frac{\mathrm{D}^{H+\frac{1}{2}}(\varphi - y_0)(t)}{\zeta(\varphi_t)} \right)^2 \mathbb{1}_{\zeta(\varphi_t) \neq 0} \, \mathrm{d}t,$$

if $\varphi \in \mathrm{I}_{y_0}^{H+\frac{1}{2}}(L^1)$ *and infinity otherwise;*

(L5) $Y_\varepsilon \sim \mathrm{LDP}\left(I_1^Y, \varepsilon^{-2H}\right)$, *with* $I_1^Y(y) = \inf\left\{I^Y(\varphi) : \varphi_1 = y\right\}$ *for* $y \in \mathbb{R}$.

While **(L1)**, **(L2)** and **(L4)** deal with pathwise large deviations, **(L3)** and **(L5)** are one-dimensional large deviations statements, about the marginal distributions of X and Y. In this small-time behaviour case, we recover the same scaling as in [100, 103].

Proof.

(L1) As discussed above, the assumptions of Theorem 2.3.11 are satisfied, so that the three-dimensional process $(X^\varepsilon, Y^\varepsilon, Z^\varepsilon)$, where $Z^\varepsilon \equiv 0$ for all $\varepsilon > 0$, satisfies an LDP with rate function

$$J(\phi, \varphi, \psi) = \inf\left\{\frac{1}{2}\int_0^T \left(u_t^2 + v_t^2 + w_t^2\right) \mathrm{d}t : \quad u, v, w \in L^2, \right.$$
$$\left. \phi_t = \int_0^t \sqrt{\Sigma(\varphi_s)}\left(\bar{\rho} u_s + \rho v_s\right) \mathrm{d}s, \quad \varphi = \mathrm{I}_{y_0}^{H+\frac{1}{2}}\left(\zeta(\varphi) v\right), \quad \psi \equiv 0\right\}.$$

The map $(X^\varepsilon, Y^\varepsilon, Z^\varepsilon) \mapsto (X^\varepsilon, Y^\varepsilon)$ is continuous so that the contraction principle [82, Theorem 4.2.1] yields an LDP for $(X^\varepsilon, Y^\varepsilon)$ with rate function $\inf\left\{J(\phi, \varphi, \psi) : \psi \equiv 0\right\} = J(\phi, \varphi, 0)$, which corresponds to I.

(L2) Since the map $(X^\varepsilon, Y^\varepsilon) \mapsto X^\varepsilon$ is continuous, the claim follows from the contraction principle.

(L3) Projecting the pathwise large deviations **(L2)** onto the last coordinate point $t = 1$ is equivalent to applying the contraction principle, and the claim follows immediately.

(L4) A direct application of Theorem 2.3.11 yields an LDP with rate function

$$I^Y(\varphi) = \inf\left\{\frac{1}{2}\int_0^T v_t^2 \,\mathrm{d}t : v \in L^2, \varphi = \mathrm{I}_{y_0}^{H+\frac{1}{2}}(\zeta(\varphi) v)\right\};$$

Inverting it as above ends the proof and **(L5)** follows from the contraction principle. □

We observe that in some special cases one can reach a more explicit expression for I.

Corollary 2.4.4. *Under the same assumptions as Proposition 2.4.3, if $\rho = 0$ or $\zeta(\varphi_t) \neq 0$ almost everywhere, the rate function I can be written*

$$I(\phi, \varphi) = \frac{1}{2\overline{\rho}^2} \int_0^T \left(\frac{\dot{\phi}_t^2}{\Sigma(\varphi_t)} \mathbb{1}_{\Sigma(\varphi_t) \neq 0} - \frac{2\rho\dot{\phi}_t \mathrm{D}^{H+\frac{1}{2}}(\varphi - y_0)(t)}{\zeta(\varphi_t)\sqrt{\Sigma(\varphi_t)}} \mathbb{1}_{\zeta(\varphi_t)\Sigma(\varphi_t) \neq 0} + \left(\frac{\mathrm{D}^{H+\frac{1}{2}}(\varphi - y_0)(t)}{\zeta(\varphi_t)} \right)^2 \mathbb{1}_{\zeta(\varphi_t) \neq 0} \right) \mathrm{d}t,$$

if $\phi \in \mathcal{AC}_0$ and $\varphi \in \mathrm{I}_{y_0}^{H+\frac{1}{2}}(L^1)$, and infinity otherwise.

Proof. We start from the definition of I given in Proposition 2.4.3**(L1)**. For each $\phi \in \mathcal{AC}_0$ and $\varphi \in \mathrm{I}_{y_0}^{H+\frac{1}{2}}(L^1)$, reverting the integral which defines ϕ gives

$$u_t = \frac{1}{\overline{\rho}} \left(\frac{\dot{\phi}_t}{\sqrt{\Sigma(\varphi_t)}} - \rho v_t \right) \mathbb{1}_{\Sigma(\varphi_t) \neq 0}, \quad \text{for all } t \in \mathbb{T},$$

because whenever $\Sigma(\varphi_t) = 0$, although u is not uniquely determined by ϕ, the optimal choice of control (the one minimising the cost) is $u_t = 0$, see [67, Remark 2.3] for more details. In the uncorrelated case $\rho = 0$, the same reasoning for the equation that φ solves yields for all $t \in \mathbb{T}$

$$v_t = \frac{1}{\zeta(\varphi_t)} \mathrm{D}^{H+\frac{1}{2}}(\varphi - y_0)(t) \mathbb{1}_{\zeta(\varphi_t) \neq 0}.$$

Furthermore, in the special case where $\zeta(\varphi_t) \neq 0$ almost everywhere the equality above holds almost everywhere, which is sufficient for the optimisation problem (because of the correlation, v may not be equal to zero even if $\zeta(\varphi)$ is). Plugging these into the rate function yields the claim. The last condition stands because if $\phi \notin \mathcal{AC}_0$ or $\varphi \notin \mathrm{I}_{y_0}^{H+\frac{1}{2}}(L^1)$ then they cannot satisfy the equations and therefore the infimum takes place over an empty set. □

2.4.1.2 Moderate deviations

We now show how our moderate deviations results apply to the rough volatility model (2.4.1). Let $h_\varepsilon = \varepsilon^{-\beta}$ for any $\beta \in (0, H)$, and define the two-dimensional process

$$\eta^\varepsilon := \frac{1}{\vartheta_\varepsilon h_\varepsilon} (X^\varepsilon, Y^\varepsilon - y_0) = \frac{1}{\varepsilon^{H-\beta}} (X^\varepsilon, Y^\varepsilon - y_0).$$

The case $\beta = 0$ corresponds to the Central Limit Theorem, whereas $\beta = H$ is the LDP regime, so that MDP precisely corresponds to some interpolation between the two.

Regarding the assumptions note that $y_0^\varepsilon = y_0$ and b_ε clearly tends to zero as ε goes to zero, hence it is trivial that $b \equiv 0$ is continuously differentiable and Lipschitz continuous, and thus **H5** and **H7** hold.

Now let Assumption 2.4.2 hold. Denoting the i-th component of b with $b^{(i)}$, one notices that $\left|b_\varepsilon^{(1)}(t,(x,y)) - b^{(1)}(t,(x,y))\right| \le \varepsilon^{H+\frac{1}{2}} |\Sigma(y)|$ and $\left|b_\varepsilon^{(2)}(t,(x,y)) - b^{(2)}(t,(x,y))\right| \le \varepsilon^{1+\varpi} C_L(1+|y|)$ by linear growth of $\mathfrak{b}$. For **H8** to hold, one then requires that $\varepsilon^{1+\varpi-(H-\beta)}$ and $\varepsilon^{H+\frac{1}{2}-(H-\beta)}$ both tend to zero as ε goes to zero. Moreover the bound (2.4.3) implies (2.3.19).

Assumption 2.4.5 (Moderate deviations assumptions)**.**

- The parameters H, ϖ and β are such that $(1+\varpi) \wedge (H+\frac{1}{2}) > H - \beta$;
- There exists $\delta > 0$ such that Σ and ζ are locally δ-Hölder continuous.

Notice that the first inequality is always satisfied if $H \le \frac{1}{2}$. Therefore, Assumptions 2.4.1, 2.4.2, 2.4.5 imply **H1 - H8** and Assumptions 2.2.3 and 2.3.9. Similarly to the LDP case, and recalling the definition of MDP from the introduction, we write $Z^\varepsilon \sim \mathrm{MDP}(\Lambda, l_\varepsilon)$ if in fact $\varepsilon^{\beta-H}(Z^\varepsilon - \overline{Z}) \sim \mathrm{LDP}(\Lambda, l_\varepsilon)$, for any $l_\varepsilon > 0$ converging to zero as ε tends to zero, where $\overline{Z}$ is the limit in distribution of Z^ε. We also denote the subset of $\mathcal{W}^d$ of absolutely continuous functions by $\mathcal{AC}$, and $\mathcal{AC}_0 := \{\phi \in \mathcal{AC}, \phi_0 = 0\}$, and refer to (2.3.23) for the definition of the Riemann-Liouville fractional derivative.

Proposition 2.4.6. *Under Assumptions 2.4.1, 2.4.2, 2.4.5 and the condition $\zeta(y_0)\Sigma(y_0) \neq 0$, the following moderate deviations hold:*

(M1) $(X^\varepsilon, Y^\varepsilon) \sim \mathrm{MDP}\left(\Lambda, \varepsilon^{-2\beta}\right)$, *where $\Lambda : \mathcal{W}^2 \to \overline{\mathbb{R}}_+$ is given by*

$$\Lambda(\phi,\varphi) = \frac{1}{2\overline{\rho}^2} \int_0^T \left(\frac{\dot{\phi}_t^2}{\Sigma(y_0)} - \frac{2\rho \dot{\phi}_t \mathrm{D}^{H+\frac{1}{2}}(\varphi - y_0)(t)}{\zeta(y_0)\sqrt{\Sigma(y_0)}} + \left(\frac{\mathrm{D}^{H+\frac{1}{2}}(\varphi - y_0)(t)}{\zeta(y_0)} \right)^2 \right) \mathrm{d}t, \tag{2.4.5}$$

if $\phi \in \mathcal{AC}_0$ and $\varphi \in \mathrm{I}_{y_0}^{H+\frac{1}{2}}(L^1)$, and infinity otherwise;

(M2) $X^\varepsilon \sim \mathrm{MDP}\left(\Lambda^X, \varepsilon^{-2\beta}\right)$, *where $\Lambda^X(\phi) = \frac{1}{2\Sigma(y_0)} \int_0^T \dot{\phi}_t^2 \,\mathrm{d}t$, if $\phi \in \mathcal{AC}_0$ and infinity otherwise;*

(M3) $\varepsilon^{H-\frac{1}{2}} X_\varepsilon \sim \mathrm{MDP}\left(\Lambda_1^X, \varepsilon^{-2\beta}\right)$, *with $\Lambda_1^X(x) = \frac{x^2}{2\Sigma(y_0)}$, for $x \in \mathbb{R}$;*

(M4) $Y^\varepsilon \sim \mathrm{MDP}\left(\Lambda^Y, \varepsilon^{-2\beta}\right)$, *where* $\Lambda^Y : \mathcal{W} \to \overline{\mathbb{R}}_+$ *is given by*

$$\Lambda^Y(\varphi) = \frac{1}{2}\int_0^T \left(\frac{\mathrm{D}^{H+\frac{1}{2}}(\varphi - y_0)(t)}{\zeta(y_0)}\right)^2 \mathbb{1}_{\zeta(y_0)\neq 0}\,\mathrm{d}t, \qquad \text{if } \varphi \in \mathrm{I}_{y_0}^{H+\frac{1}{2}}(L^1) \text{ and infinity otherwise;}$$

(M5) $Y_\varepsilon \sim \mathrm{MDP}(\Lambda_1^Y, \varepsilon^{-2\beta})$, *where* $\Lambda_1^Y(y) = \frac{1}{2}y^2$ *for* $y \in \mathbb{R}$.

Proof.

(M1) As discussed above, the assumptions of Theorem 2.3.20 are satisfied, thus it yields an MDP with rate function

$$\Lambda(\phi, \varphi) = \left\{\frac{1}{2}\int_0^T \left(u_t^2 + v_t^2\right)\,\mathrm{d}t : u, v \in L^2, \phi_t = \int_0^t \sqrt{\Sigma(y_0)}\left(\overline{\rho}u_s + \rho v_s\right)\mathrm{d}s, \varphi = \mathrm{I}_{y_0}^{H+\frac{1}{2}}(\zeta(y_0)v)\right\},$$

and inverting it as in the LDP case gives the claim.

(M2) The contraction principle implies that an MDP for X^ε holds with rate function $\Lambda^X(\phi) = \inf\left\{\Lambda(\phi, \varphi) : \varphi \in \mathrm{I}_{y_0}^{H+\frac{1}{2}}(L^1)\right\}$. Let $\psi \in \mathcal{AC}_0$ such that $\dot{\psi} := \frac{\mathrm{D}^{H+\frac{1}{2}}(\varphi - y_0)}{\zeta(y_0)}$, then the rate function translates to

$$\Lambda^X(\phi) = \inf\left\{\frac{1}{2\overline{\rho}^2}\int_0^T \left(\frac{\dot{\phi}_t^2}{\Sigma(y_0)} - \frac{2\rho\dot{\phi}_t\dot{\psi}_t}{\sqrt{\Sigma(y_0)}} + \dot{\psi}_t^2\right)\mathrm{d}t, \quad \psi \in \mathcal{AC}\right\},$$

which can be solved as a variational problem as in [164, Corollary 2.4]. The corresponding Euler-Lagrange equation reads $\ddot{\psi} = \rho\ddot{\phi}/\sqrt{\Sigma(y_0)}$ hence $\dot{\psi} = \rho\dot{\phi}/\sqrt{\Sigma(y_0)}$ because $\dot{\psi}_0 = 0$ by definition. Plugging into the above equation finishes the proof.

(M3) The rate function is given by contraction principle as

$$\Lambda_1^X(x) = \inf\left\{\Lambda^X(\phi) : \phi \in \mathcal{AC}_0, \phi_1 = x\right\} = \inf\left\{\frac{1}{2\Sigma(y_0)}\int_0^T \dot{\phi}_t^2\,\mathrm{d}t : \phi \in \mathcal{AC}_0, \phi_1 = x\right\}.$$

Setting $T = 1$ the optimal path under the constraint $\phi_1 = x$ is $\phi_t = xt$ by the Euler-Lagrange equation. Again, plugging it into the rate function ends the proof.

(M4) Theorem 2.3.20 gives an MDP for Y^ε with rate function

$$\Lambda^Y(\varphi) = \inf\left\{\frac{1}{2}\int_0^T v_t^2\,\mathrm{d}t : v \in L^2, \varphi = \mathrm{I}_{y_0}^{H+\frac{1}{2}}(\zeta(y_0)v)\right\}.$$

Inverting it yields **(M4)**.

(M5) By contraction principle one obtains $\Lambda_1^Y(y) = \inf\left\{\Lambda^Y(\varphi), \varphi_1 = y\right\}$. Then setting ψ as in **(M2)** it boils down to the same optimisation problem as for **(M3)**. $\square$

As in the large deviations results, **(M1)**, **(M2)** and **(M4)** correspond to pathwise statements, whereas **(M3)** and **(M5)** are finite-dimensional results about the marginal distributions. For the log-stock price, **(M3)** corresponds precisely to the moderately-out-of-the-money regime presented and justified in [111] (for diffusion volatility models), based on the observation that the range of observable strikes grows with maturity. Furthermore, one can always apply Theorem 2.3.20 in the degenerate case $\Sigma(y_0)\zeta(y_0) = 0$ although the rate functions take slightly different forms.

2.4.1.3 Implied volatility asymptotics

We can easily deduce from the above results the asymptotic behaviour of the implied volatility, a standard norm for quoting option prices. For each maturity $t \geq 0$ and log-moneyness $k \in \mathbb{R}$, the implied volatility $\widehat{\sigma}(t,k)$ is the unique non-negative solution to $C_{\mathrm{BS}}\big(t,k,\widehat{\sigma}(t,k)\big) = C(t,k)$, where C_{BS} corresponds to the price of a European Call option under the Black-Scholes model, and C a given Call option price (for example in a rough volatility model). This notion is only well defined if the underlying stock price is a true martingale, which we have not assumed so far, and may require additional conditions on the coefficients. This will be the case though in all our examples below, but for now, with the current level of generality, we assume it:

Assumption 2.4.7. The process $\exp(X^\varepsilon)$ in (2.4.2) is a true martingale for all small enough $\varepsilon > 0$.

Small-time implied volatility asymptotics can be derived from Properties 2.4.3 and 2.4.6 in a similar fashion. The explicit form of the MDP rate function allows a closed form expression.

Corollary 2.4.8. *Let Assumption 2.4.7 hold.*

(LDP) *Under the same assumptions as Proposition 2.4.3,*

$$\lim_{t\downarrow 0} \widehat{\sigma}\big(t, kt^{\frac{1}{2}-H}\big)^2 = \begin{cases} \dfrac{k^2}{2\inf_{x\geq k} I_1^X(x)}, & \text{if } k > 0, \\ \dfrac{k^2}{2\inf_{x\leq k} I_1^X(x)}, & \text{if } k < 0. \end{cases} \tag{2.4.6}$$

(MDP) *Under the same assumptions as Proposition 2.4.6 and for any $\beta \in (0, H)$, $k \neq 0$,*

$$\lim_{t\downarrow 0} \widehat{\sigma}\big(t, kt^{\frac{1}{2}-\beta}\big)^2 = \Sigma(y_0). \tag{2.4.7}$$

Proof.

(LDP) Consider the case $k > 0$. Proposition 2.4.3**(L3)** translates into

$$\lim_{t\downarrow 0} t^{2H} \log \mathbb{P}(t^{H-\frac{1}{2}} X_t \geq k) = -\inf_{x\geq k} I_1^X(x).$$

Meanwhile in the Black-Scholes model with constant volatility $\sigma > 0$ the log-price process satisfies $X_t = -\frac{\sigma^2 t}{2} + \sigma B_t$ for all $t \in \mathbb{T}$, and simple Gaussian computations yield the large deviations behaviour

$$\lim_{t\downarrow 0} t^{2H} \log \mathbb{P}(X_t \geq kt^{\frac{1}{2}-H}) = -\frac{k^2}{2\sigma^2}.$$

The claim then follows directly from [117, Corollary 7.1], and by symmetry for the case $k < 0$.

(MDP) Following the same arguments as above, we obtain

$$\lim_{t\downarrow 0} \widehat{\sigma}\big(t, kt^{\frac{1}{2}-\beta}\big)^2 = \begin{cases} \dfrac{k^2}{2\inf_{x\geq k} \Lambda_1^X(x)}, & \text{if } k > 0, \\ \dfrac{k^2}{2\inf_{x\leq k} \Lambda_1^X(x)}, & \text{if } k < 0. \end{cases}$$

Plugging in the expression of Λ_1^X from **(M3)** finishes the proof. □

This concludes the presentation of the general results for rough volatility models. The next sections display the diversity of the models found in the literature and how large and moderate deviations principles apply to them.

2.4.2 Small-time rescaling (examples)

2.4.2.1 Rough Stein-Stein

The rough Stein-Stein, suggested in [151] is an extension of the classical Stein-Stein volatility model [228] to the fractional setting. It corresponds to (2.4.1) with $K_1 \equiv 1$ (hence $\varpi = 0$), $K_2(t) = t^{H-\frac{1}{2}}/\Gamma(H+\frac{1}{2})$, $H \in (0, \frac{1}{2})$, $y_0 > 0$, $\Sigma(y) = y^2$, $\mathfrak{b}(y) =$

$\kappa(\theta - y)$, $\kappa, \theta > 0$ and $\zeta(y) \equiv \xi > 0$. The coefficients are Lipschitz continuous and well-behaved, hence Assumptions 2.4.2 and 2.4.5 are easily checkable and the limit equation (2.4.4) has a unique solution, hence Propositions 2.4.3 and 2.4.6 apply. Note that because ζ is a positive constant, Corollary 2.4.4 gives the rate function I in integral form and one can solve **(L5)** in closed-form using the Euler-Lagrange equation in a similar way as in the proof of Proposition 2.4.6. Furthermore, since Y is Gaussian its exponential moments are finite and Novikov's condition [166, Section 3.5.D] ensures that Assumption 2.4.7 holds. Therefore, Corollary 2.4.8 yields the small-time behaviour of the implied volatility. Notice that the LDP and MDP for this model still hold when replacing the Riemann-Liouville kernel with the standard fractional Brownian motion by virtue of Theorem 2.3.29. The pathwise LDP for this model was first derived in [151] albeit with the different scaling $X_t^\varepsilon := \varepsilon^{H-\frac{1}{2}+2\beta} X_{\varepsilon t}$ and $Y_t^\varepsilon := \varepsilon^\beta Y_{\varepsilon t}$, for $\beta > 0$.

2.4.2.2 Rough Bergomi

The rough Bergomi model as presented in [19] reads

$$\begin{cases} X_t &= -\frac{1}{2}\int_0^t V_s \,\mathrm{d}s + \int_0^t \sqrt{V_s}\mathrm{d}B_s, \\ V_t &= V_0 \exp\left(\int_0^t \frac{(t-s)^{H-\frac{1}{2}}}{\Gamma(H+\frac{1}{2})}\mathrm{d}W_s - at^{2H}\right), \end{cases}$$

with $V_0 > 0$ and $a \in \mathbb{R}$. A pathwise LDP for this model first appeared in [161] using the Freidlin-Wentzell approach and a tailored proof. This case is quite intricate because the exponential does not satisfy the linear growth bound but we circumvented this issue by introducing the notion of autonomous system, illustrated in Example 2.2.7 and completed by Assumption 2.3.9 and **H3b**. Not only does this framework unifies the result of [161] with other rough volatility models, but it also leads to a pathwise MDP.

With $Y := \log(V)$, the system (X, Y) fits into (2.4.1) where $K_1(t) = t^{2H-1}$, $K_2(t) = t^{H-\frac{1}{2}}/\Gamma(H+\frac{1}{2})$, for $H \in (0, \frac{1}{2})$, $\Sigma(y) = \exp(y)$, $\mathfrak{b}(y) = -a/(2H)$, $a > 0$, $\zeta(y) \equiv 1$ and $y_0 = \log(V_0)$. The bound (2.4.3) is then satisfied since $\int_0^t (t-s)^{H-\frac{1}{2}}\mathrm{d}W_s$ is a Gaussian process with exponentional moments bounded uniformly in $t \in \mathbb{T}$, and for each $v \in \mathcal{A}_M$, $M > 0$:

$$\int_0^t (t-s)^{H-\frac{1}{2}} v_s \,\mathrm{d}s \le M\frac{t^{2H}}{2H},$$

almost surely, by Cauchy-Schwarz inequality. Therefore $\sup_{t\in\mathbb{T}} \mathbb{E}\big[\exp\big(Y_t^{\varepsilon,v}\big)\big]$ is finite, yielding the claim. Moreover, the volatility equation is explicit so we shall not be concerned with uniqueness and the rest of Assumptions 2.4.2 and 2.4.5 is straightforward to check. This implies that Propositions 2.4.3 and 2.4.6 apply, and so does Corollary 2.4.4. Again, Theorem 2.3.29 guarantees that the LDP and MDP still hold when K_2 is replaced with the non-convolution fractional Brownian motion kernel. Gassiat [120] showed that, if $\rho \leq 0$, then the stock price process is a true martingale, ensuring that Assumption 2.4.7 holds, and implied volatility asymptotics thus follow from Corollary 2.4.8.

2.4.2.3 Rough Heston

As introduced in [95] the rough Heston model fits into the framework of (2.4.1) with $K_1(t) = K_2(t) = t^{H-\frac{1}{2}}/\Gamma(H+\frac{1}{2})$, for $H \in (0, \frac{1}{2})$, $y_0 > 0$, $\Sigma(y) = y$, $\mathfrak{b}(y) = \kappa\big(\theta - y\big)$, $\kappa > 0$, $\theta \geq 0$ and $\zeta(y) = \xi\sqrt{y}$, $\xi \in \mathbb{R}^d$. Linear growth and local Hölder continuity of the coefficients clearly hold. The weak existence and uniqueness was proved in [3], however the square-root coefficient brings an issue for pathwise uniqueness of the SVE. We assume here that there exists a set $\mathcal{U}$ of coefficients $(H, \kappa, \theta, \xi, \rho, y_0)$ such that pathwise uniqueness indeed stands. The only known result so far is due to [239] in the smooth case $H = \frac{1}{2}$. We also recall that pathwise uniqueness was proved for $\zeta(y) = y^\gamma$ where $\gamma > \frac{1}{2H+1}$ in [195], but does not encompass the square root case. Therefore Assumption 2.4.2 holds in those two cases. On a heuristic note remark that, even if pathwise uniqueness fails, there is a unique candidate for $\mathcal{G}^\varepsilon$ since there exists a unique strong solution until the first hitting time of zero. The issue is it may not satisfy the SVE anymore after that time, but should be consistent for small-time LDP.

Moreover, uniqueness of the limit equation (2.4.4) only holds up to first hitting time of zero. Hence we will make use of the uniqueness relaxation presented in Section 2.3.1 and similar arguments as in Example 2.3.8 to prove that Assumption 2.3.3 holds. The suggested rate function (2.3.2) reads now

$$I(\phi,\varphi) = \inf\left\{\frac{1}{2}\int_0^t \left(u_t^2 + v_t^2\right)\,\mathrm{d}t : u, v \in L^2,\, \phi_t = \int_0^t \sqrt{\varphi_s}\left(\overline{\rho} u_s + \rho v_s\right)\mathrm{d}s,\, \varphi_t = \mathrm{I}_{y_0}^{H+\frac{1}{2}}(\xi\sqrt{\varphi}v)\right\}. \tag{2.4.8}$$

We emphasise that φ above solves the Volterra equation

$$\varphi_t = y_0 + \xi \int_0^t \frac{(t-s)^{H-\frac{1}{2}}}{\Gamma(H+\frac{1}{2})} \sqrt{\varphi_s} v_s \, \mathrm{d}s, \quad \text{for all } t \in \mathbb{T}. \tag{2.4.9}$$

Lemma 2.4.9. *Let $v \in L^2$ and φ satisfying the Volterra equation (2.4.9) with $H \in (0, \frac{1}{2})$. Then the set $\mathcal{D} := \{t \in [0,T] : \varphi_t > 0\}$ has Lebesgue measure T.*

Proof. We follow some arguments in the proof of [3, Theorem 3.6]. Let us drop the subscript in the kernel and write it K for clarity, and introduce its resolvent of the first kind $L(\mathrm{d}t) := \frac{t^{-H-\frac{1}{2}}}{\Gamma(\frac{1}{2}-H)} \mathrm{d}t$ [134, Definition 5.5.1]. Moreover, for $h > 0$, define $\Delta_h K(t) := K(t+h)$ and for every measurable function f on $\mathbb{R}_+$ and measure g on $\mathbb{R}_+$, $(f * g)(t) := \int_0^t f(t-s)\mathrm{d}g(s)$. It is proved, in [3, Equation (3.9)], that $\Delta_h K * L$ is non-decreasing but in fact in this special (rough) case, it is strictly increasing. Indeed, the authors show that in the general case, for all $0 \le s \le t \le T$,

$$(\Delta_h K * L)(t) - (\Delta_h K * L)(s) = \int_0^h K(h-u)\big(L(s+\mathrm{d}u) - L(t+\mathrm{d}u)\big),$$

which is positive because $K > 0$ and L is decreasing. Furthermore K is decreasing and $L > 0$ thus

$$0 < (\Delta_h K * L)(t) < (K * L)(t) = 1,$$

where the equality holds by definition. Let $\varphi_t = y_0 + \int_0^t K(t-s)\xi\sqrt{\varphi_s} v_s \, \mathrm{d}s =: y_0 + (K * z)(t)$, where z is trivially a semimartingale, hence from [3, Equation (2.15)]:

$$y_0 + (\Delta_h K * \mathrm{d}z)(t) = \Big(1 - (\Delta_h K * L)(t)\Big) y_0 + \Big(\Delta_h K * L\Big)(0)\varphi_t + \Big(\mathrm{d}(\Delta_h K * L) * \varphi\Big)(t),$$

which is strictly positive because because $y_0, \varphi \ge 0$ and the two lines before. Now let us suppose there exists an interval $[t, t+h] \subset \mathbb{T}$ on which $\varphi = 0$. Then

$$\varphi_{t+h} = y_0 + \int_0^t K(t+h-s)\sqrt{\varphi_s} v_s \, \mathrm{d}s = y_0 + (\Delta_h K * \mathrm{d}z)(t) > 0,$$

which is a contradiction. Hence no such interval exists and the claim follows. □

Remark 2.4.10. This argument works for any rough kernel but not for the diffusion case $H = \frac{1}{2}$. We refer to [84, Proposition 3.3] for the latter.

The previous lemma allows to invert the integrals as showed in the proof of Corollary 2.4.4 and yields a more explicit form for I:

$$I(\phi,\varphi) = \int_0^T \frac{\mathbb{1}_{\varphi_t>0}}{2\overline{\rho}^2\varphi_t}\left(\dot{\phi}_t^2 - 2\rho\dot{\phi}_t\frac{\mathrm{D}^{H+\frac{1}{2}}(\varphi-y_0)(t)}{\xi} + \left(\frac{\mathrm{D}^{H+\frac{1}{2}}(\varphi-y_0)(t)}{\xi}\right)^2\right)\mathrm{d}t, \tag{2.4.10}$$

if the integral is well defined, $\phi \in \mathcal{AC}_0$, $\varphi \in \mathrm{I}_{y_0}^{H+\frac{1}{2}}(L^1)$, and $I = +\infty$ otherwise. We can now prove the following:

Lemma 2.4.11. *Let $H \in (0, \frac{1}{2})$, then the functional I satisfies Assumption 2.3.3.*

Proof. Note that any solution φ of (2.4.9) is non-negative. Let (ϕ,φ) be such that $I(\phi,\varphi)$ is finite. Then, for each $\delta > 0$, define $\varphi_t^\delta := \varphi_t + \delta t^{H+\frac{1}{2}}$ such that φ^δ is strictly positive. Therefore from definition (2.3.23) we have:

$$\begin{aligned}\mathrm{D}^{H+\frac{1}{2}}(\varphi^\delta - y_0)(t) - \mathrm{D}^{H+\frac{1}{2}}(\varphi - y_0)(t) &= \frac{1}{\Gamma(\frac{1}{2}-H)}\frac{\mathrm{d}}{\mathrm{d}t}\int_0^t (t-s)^{-H-\frac{1}{2}}(\varphi_s^\delta - \varphi_s)\,\mathrm{d}s \\ &= \frac{\delta}{\Gamma(\frac{1}{2}-H)}\frac{\mathrm{d}}{\mathrm{d}t}\int_0^t (t-s)^{-H-\frac{1}{2}}s^{H+\frac{1}{2}}\,\mathrm{d}s = \delta\,\frac{\pi\left(H+\frac{1}{2}\right)}{\Gamma(\frac{1}{2}-H)\cos(\pi H)},\end{aligned}$$

which entails convergence as δ goes to zero, uniformly on $\mathbb{T}$. Now define the control

$$v_t := \frac{\mathrm{D}^{H+\frac{1}{2}}(\varphi - y_0)(t)}{\sqrt{\varphi_t}}\mathbb{1}_{\varphi_t>0},$$

and v belongs to L^2 since $I(\phi,\varphi)$ is finite. Then for each $\delta > 0$ the control v^δ defined as

$$v_t^\delta := \frac{\mathrm{D}^{H+\frac{1}{2}}(\varphi^\delta - y_0)(t)}{\sqrt{\varphi_t^\delta}} \le \frac{\mathrm{D}^{H+\frac{1}{2}}(\varphi^\delta - y_0)(t)}{\sqrt{\varphi_t}}$$

is also in L^2 because $\mathbb{1}_{\varphi_t>0} = 1$ almost everywhere. Furthermore, for all $t \in \mathcal{D}$, $\lim_{\delta\downarrow 0}\left(\varphi_t^\delta\right)^{-1} = (\varphi_t)^{-1}$ and therefore $\lim_{\delta\downarrow 0} v_t^\delta = v_t$. Let $P(\varphi)$ denote the term between brackets in (2.4.10) divided by $2\overline{\rho}^2$, which is non-negative by design since it corresponds to $\varphi(u^2+v^2)$. Therefore, by Lemma 2.4.9,

$$I(\phi,\varphi) - I(\phi,\varphi^\delta) = \int_{\mathcal{D}}\left(\frac{1}{\varphi_t} - \frac{1}{\varphi_t^\delta}\right)P(\varphi)_t\,\mathrm{d}t + \int_{\mathcal{D}}\frac{1}{\varphi_t^\delta}\Big(P(\varphi)_t - P(\varphi^\delta)_t\Big)\,\mathrm{d}t,$$

where the first integrand is smaller than $P(\varphi)_t/\varphi_t$ for all $t \in \mathcal{D}$ and this upper bound belongs to L^1 by assumptions. Hence the dominated convergence theorem implies that the first integral goes to zero. From the calculations above we deduce that $P(\varphi)_t - P(\varphi^\delta)_t$ tends to zero uniformly as δ goes to zero, hence the second integrand converges pointwise and, for δ small enough, is dominated by $P(\varphi)/\varphi$. A second application of DCT yields convergence of the integral, and the claim follows. $\square$

Therefore the large and moderate deviations from Propositions 2.4.3 and 2.4.6 apply if the coefficients belong to $\mathcal{U}$ and $H \in (0, \frac{1}{2})$. Observe that Proposition 2.4.6**(M3)** agrees with [100, Section 3.5], although the routes taken differ significantly. El Euch and Rosenbaum [94, Appendix B] showed that Assumption 2.4.7 is satisfied, and the implied volatility behaviour thus follows from Corollary 2.4.8.

2.4.2.4 Multi-factor rough Bergomi

Let W be an $\mathbb{R}^{m+1}$-Brownian motion, $\boldsymbol{Z} := \left(Z^{(1)}, \cdots, Z^{(m)}\right)^{\dagger}$ where

$$Z_t^{(j)} := \int_0^t K^{(j)}(t-s)\mathrm{d}W_s^{(j)},$$

and we allow $K^{(j)} \in L^2(\mathbb{T}, \mathbb{R}_+)$, $j \in [\![1, m]\!]$, to be homogeneous of different degrees $H_j - \frac{1}{2}$ with $H_j \in (0, 1]$. Therefore, the variance of $\boldsymbol{Z}$ is proportional to $\boldsymbol{A}_t := \left(t^{2H_1}/(2H_1), \dots, t^{2H_m}/(2H_m)\right)^{\dagger}$, for all $t \in \mathbb{T}$. Assume without loss of generality that the H_j are ordered by increasing values, then we will design the rescaling at the speed ε^{-2H_1}. Denote $m^\star := \max\{j \in [\![1, m]\!] : H_j = H_1\}$. Let $\mathcal{U}$ and $\mathcal{V}$ be m-dimensional square matrices, and $\boldsymbol{Y}$ an m-dimensional process defined for all $t \in \mathbb{T}$ by

$$\boldsymbol{Y}_t := y_0 + \mathcal{U}\boldsymbol{Z}_t - \frac{1}{2}\mathcal{V}\boldsymbol{A}_t, \quad y_0 \in \mathbb{R}^m.$$

The log-price reads

$$X_t = -\frac{1}{2m}\int_0^t \sum_{i=1}^m \exp\left(Y_s^{(i)}\right) \mathrm{d}s + \int_0^t \sqrt{\frac{1}{m}\sum_{i=1}^m \exp\left(Y_s^{(i)}\right)}\mathrm{d}B_s,$$

where $B = \overline{\rho}W^{(m+1)} + \sum_{j=1}^m \rho_j W^{(j)}$, $\overline{\rho}^2 + \sum_{i=1}^m \rho_i^2 = 1$. The rescaling $X_t^\varepsilon = \varepsilon^{H_1 - \frac{1}{2}} X_{\varepsilon t}, \boldsymbol{Y}_t^\varepsilon := \boldsymbol{Y}_{\varepsilon t}$ yields

$$\begin{cases} X_t^\varepsilon = -\dfrac{\varepsilon^{H_1+\frac{1}{2}}}{2m}\displaystyle\int_0^t \sum_{i=1}^m \exp\left(Y_s^{\varepsilon,(i)}\right) \mathrm{d}s + \varepsilon^{H_1}\int_0^t \sqrt{\frac{1}{m}\sum_{i=1}^m \exp\left(Y_s^{\varepsilon,(i)}\right)}\mathrm{d}B_s, \\ Y_t^{\varepsilon,(i)} = y_0^{(i)} + \displaystyle\sum_{j=1}^m \left(\varepsilon^{H_j}\mathcal{U}_{ij}Z_t^{(j)} - \mathcal{V}_{ij}\frac{(\varepsilon t)^{2H_j}}{2H_j}\right), \quad \text{for all } i \in [\![1, m]\!]. \end{cases}$$

As we will shift each BM by $\varepsilon^{-H_1}\int v^{(j)}$, we notice that $\varepsilon^{H_j-H_1}\mathcal{U}_{ij}$ goes to zero if $H_j > H_1$, i.e. if $j > m^\star$. It means that the roughest component(s) (the one(s) with H_1) will outweigh the others, and only the former will make a contribution to the rate function.

Although similar to its one-dimensional counterpart, this model does not fit into the framework of (2.4.1). Regarding the assumptions of Theorems 2.3.11, we only check **H3b** and Assumption 2.3.9 because the others are standard and similar to the one-dimensional case. Clearly $(Y^{\varepsilon,(1)},\cdots,Y^{\varepsilon,(m)})$ is an autonomous subsystem. As a Gaussian process, $\mathcal{U}\boldsymbol{Z}$ has exponential moments of all orders and for all $M>0$, $j\in[\![1,m]\!]$ and $v^{(j)}\in\mathcal{A}_M$:

$$\int_0^t K^{(j)}(t-s)v_s^{(j)}\,\mathrm{d}s \le \sqrt{M}\|K^{(j)}\|_2 \quad \text{almost surely,}$$

thus $\exp\left(Y^{\varepsilon,(i),v}\right)\in L^p(\Omega)$ for all $p\ge 1$. Therefore the bound (2.3.6) and Assumption 2.3.9 are satisfied. This estimate also checks that (2.3.19) and thus **H8** stand. Since $\vartheta_\varepsilon=\varepsilon^{H_1}$, we define for the moderate deviations regime $h_\varepsilon=\varepsilon^\beta, \beta\in(0,H_1)$.

Corollary 2.4.12. *The pathwise LDP and MDP hold.*

- $(X^\varepsilon,Y^\varepsilon)\sim \mathrm{LDP}\left(I,\varepsilon^{-2H_1}\right)$ *where for all* $\phi\in\mathcal{W}$ *and* $\varphi\in\mathcal{W}^m$*:*

$$I(\phi,\varphi)=\inf\Bigg\{\frac{1}{2}\int_0^T\left(u_t^2+|v_t|^2\right)\mathrm{d}t: \quad u\in L^2(\mathbb{T},\mathbb{R}), v\in L^2(\mathbb{T},\mathbb{R}^m),$$
$$\phi_t=\int_0^t\left(\frac{1}{m}\sum_{i=1}^m \mathrm{e}^{\varphi_s^{(i)}}\right)^{\frac{1}{2}}\left(\overline{\rho}u_s+\sum_{j=1}^m\rho_j v_s^{(j)}\right)\mathrm{d}s,\quad \varphi_t^{(i)}=y_0^{(i)}+\sum_{j=1}^{m^\star}\mathcal{U}_{ij}\int_0^t K^{(j)}(t-s)v_s^{(j)}\,\mathrm{d}s\Bigg\}.$$

- $(X^\varepsilon,Y^\varepsilon)\sim \mathrm{MDP}\left(\Lambda,\varepsilon^{-2\beta}\right)$ *where for all* $\phi\in\mathcal{W}$ *and* $\varphi\in\mathcal{W}^m$*:*

$$\Lambda(\phi,\varphi)=\inf\Bigg\{\frac{1}{2}\int_0^T\left(u_t^2+|v_t|^2\right)\mathrm{d}t: \quad u\in L^2(\mathbb{T},\mathbb{R}), v\in L^2(\mathbb{T},\mathbb{R}^m), \tag{2.4.11}$$
$$\phi_t=\int_0^t\left(\frac{1}{m}\sum_{i=1}^m \mathrm{e}^{y_0^{(i)}}\right)^{\frac{1}{2}}\left(\overline{\rho}u_s+\sum_{j=1}^m\rho_j v_s^{(j)}\right)\mathrm{d}s,\quad \varphi_t^{(i)}=\sum_{j=1}^{m^\star}\mathcal{U}_{ij}\int_0^t K^{(j)}(t-s)v_s^{(j)}\,\mathrm{d}s\Bigg\}.$$

Proof. The LDP is a direct application of Theorem 2.3.11 and the MDP of Theorem 2.3.20. □

One can also recover the LDP and MDP for $(X^\varepsilon)_{\varepsilon>0}$ as well as the small-time LDP and MDP by contraction principle, as in Propositions 2.4.3 and 2.4.6.

If $\mathcal{U}$ is lower triangular (i.e. $\mathcal{U}_{ij} = 0$ for all $i < j$), for instance if it arises from the Cholesky decomposition of a covariance matrix, then for all $\phi \in \mathcal{W}$, $\varphi \in \mathcal{W}^m$, one can derive the vector v recursively, followed by u. Note that if $m^\star < m$, φ may not be attainable by the restrained number of controls $\{v^{(j)}, j \in [\![1, m^\star]\!]\}$.

Example 2.4.13. *Consider the case $m = 2$. Let $K^{(1)}(t) = K^{(2)}(t) = t^{H-\frac{1}{2}}/\Gamma(H + \frac{1}{2})$ for $H \in (0, \frac{1}{2})$, hence $m^\star = 2$, and $\mathcal{U}$ be lower triangular (i.e. $\mathcal{U}_{12} = 0$). Then in the moderate deviations setting v and u are explicit from* (2.4.11)*:*

$$
\begin{aligned}
v^{(1)} &= \frac{1}{\mathcal{U}_{11}} \mathrm{D}^{H+\frac{1}{2}} \left(\varphi^{(1)}\right), \qquad v^{(2)} = \frac{1}{\mathcal{U}_{22}} \mathrm{D}^{H+\frac{1}{2}} \left(\varphi^{(2)}\right) - \frac{\mathcal{U}_{21}}{\mathcal{U}_{22}} v^{(1)}, \\
u &= \frac{1}{\bar{\rho}} \left[\sqrt{2}\, \dot{\phi} \left\{ \exp\left(y_0^{(1)}\right) + \exp\left(y_0^{(2)}\right) \right\}^{-\frac{1}{2}} - \rho_1 v^{(1)} - \rho_2 v^{(2)} \right].
\end{aligned}
$$

Remark 2.4.14. We can similarly consider multidimensional versions of the other models presented in this chapter and derive large and moderate deviation principles. We only work out the computations for the multi-factor rough Bergomi model because it is the most relevant in the literature.

2.4.3 Tail rescaling

We now investigate tail rescalings, which generally have the form $X^\varepsilon = \varepsilon X$, such that an LDP provides asymptotic estimates on $\mathbb{P}(X^\varepsilon \geq 1) = \mathbb{P}(X \geq \varepsilon^{-1})$. The MDP for the whole system is not available in this case because $\overline{Y} := \lim_{\varepsilon \downarrow 0} Y^\varepsilon \equiv 0$ hence the limit equation for X^ε, arising from (2.3.18), would be independent of the control. Note that the theory does not break down but the rate function is trivial (equals zero at zero and $+\infty$ everywhere else). Furthermore, the exponential function prevents the study of such a rescaling in the rough Bergomi model.

2.4.3.1 Rough Stein-Stein

This model was defined in Section 2.4.2.1, but with the rescaling $Y_t^\varepsilon := \varepsilon Y_t$ and $X_t^\varepsilon := \varepsilon^2 X_t$, the system becomes

$$
\begin{cases}
X_t^\varepsilon &= -\frac{1}{2} \int_0^t (Y_s^\varepsilon)^2 \, \mathrm{d}s + \varepsilon \int_0^t Y_s^\varepsilon \mathrm{d}B_s, \\
Y_t^\varepsilon &= \varepsilon y_0 + \int_0^t \kappa \left(\varepsilon\theta - Y_s^\varepsilon \right) \mathrm{d}s + \varepsilon \int_0^t \xi \frac{(t-s)^{H-\frac{1}{2}}}{\Gamma(H + \frac{1}{2})} \mathrm{d}W_s,
\end{cases}
\tag{2.4.12}
$$

where the coefficients are identical to the small-time case. Although the rescaling is different, Assumption 2.2.3 and **H1 - H4** are easily satisfied in a similar way, the limit equation (2.3.5) has a unique solution, and therefore Theorem 2.3.11 applies.

Corollary 2.4.15. *The following hold:*

(L1) $(X^\varepsilon, Y^\varepsilon) \sim \mathrm{LDP}\left(I, \varepsilon^{-2}\right)$ *with*

$$I(\phi,\varphi) = \frac{1}{2}\int_0^T \left(u_t^2 + v_t^2\right) \mathrm{d}t, \quad \textit{where} \quad \begin{cases} u &= \dfrac{1}{\overline{\rho}}\left(\dfrac{\dot{\phi}}{\varphi} + \dfrac{1}{2}\varphi - \rho v\right) \mathbb{1}_{\varphi\neq 0}, \\ v &= \dfrac{1}{\xi}\left(\mathrm{D}^{H+\frac{1}{2}}(\varphi) + \kappa \mathrm{I}^{\frac{1}{2}-H}(\varphi)\right), \end{cases}$$

if $\phi \in \mathcal{AC}_0$, $\varphi \in \mathrm{I}_0^{H+\frac{1}{2}}$ *and infinity otherwise.*

(L2) $X^\varepsilon \sim \mathrm{LDP}\left(I^X, \varepsilon^{-2}\right)$ *with* $I^X(\phi) = \inf\left\{I(\phi,\varphi) : \varphi \in \mathrm{I}_0^{H+\frac{1}{2}}\right\}$ *if* $\phi \in \mathcal{AC}_0$ *and infinity otherwise.*

(L3) *For each* $t \in \mathbb{T}$, $\varepsilon^2 X_t \sim \mathrm{LDP}\left(I_t^X, \varepsilon^{-2}\right)$*, where* $I_t^X(x) = \inf\left\{I^X(\phi) : \phi_t = x\right\}$.

Proof. For **(L1)**, Theorem 2.3.11 entails that the rate function is

$$I(\phi,\varphi) = \inf\left\{\frac{1}{2}\int_0^T \left(u_t^2 + v_t^2\right) \mathrm{d}t : \quad u, v \in L^2, \right.$$
$$\left. \phi_t = -\frac{1}{2}\int_0^t \varphi_s^2 \,\mathrm{d}s + \int_0^t \varphi_s\left(\overline{\rho} u_s + \rho v_s\right) \mathrm{d}s,\ \varphi_t = -\int_0^t \kappa\, \varphi_s \,\mathrm{d}s + \int_0^t \xi \frac{(t-s)^{H-\frac{1}{2}}}{\Gamma(H+\frac{1}{2})} v_s \,\mathrm{d}s \right\}.$$

Inverting the integrals as in Corollary 2.4.4 to obtain the unique controls and using $\mathrm{D}^{H+\frac{1}{2}}\mathrm{I}^1 = \mathrm{I}^{\frac{1}{2}-H}$ yields the claim. Similarly to the small-time case, **(L2)** follows from the contraction principle, and one only needs to fix $t \in \mathbb{T}$ to prove **(L3)**. □

One can prove an LDP for Y^ε in a similar way; a more interesting problem is the moderate deviations setting. Recall that an MDP for the couple $(X^\varepsilon, Y^\varepsilon)$ would have a trivial rate function because the limit equation of X^ε is independent of the control. However, since the diffusion coefficient of Y^ε is constant equal to ξ, one can obtain an MDP for Y^ε. More surprisingly, the limit equations in the large deviations (2.3.5) and moderate deviations (2.3.18) regimes coincide, which leads to identical rate functions. Notice that $\vartheta_\varepsilon = \varepsilon$ in this example, and therefore let $h_\varepsilon = \varepsilon^{-\beta}$ where $\beta \in (0,1)$.

Corollary 2.4.16. *$Y^\varepsilon \sim \mathrm{MDP}\left(\Lambda^Y, \varepsilon^{-2\beta}\right)$ where*

$$\Lambda^Y(\varphi) = \frac{1}{2\xi^2}\int_0^T \left(\mathrm{D}^{H+\frac{1}{2}}(\varphi)(t) + \kappa \mathrm{I}^{\frac{1}{2}-H}(\varphi)(t)\right)^2 \mathrm{d}t,$$

if $\varphi \in \mathrm{I}_0^{H+\frac{1}{2}}$ and infinity otherwise.

Proof. From (2.4.12), $b_\varepsilon(y) = \kappa(\varepsilon\theta - y)$ converges to $b(y) = -\kappa y$ and the diffusion coefficient is constant, hence **H2 - H6** are easily satisfied. Moreover, $b_\varepsilon - b \equiv \varepsilon\kappa\theta$ and $\varepsilon^{1-(H-\beta)}$ tends to zero therefore **H7** and **H8** also hold. Theorem 2.3.20 thus yields an MDP with rate function

$$\Lambda^Y(\varphi) = \inf\left\{\frac{1}{2}\int_0^T v_t^2\,\mathrm{d}t : v \in L^2,\ \varphi_t = -\int_0^t \kappa\varphi_s\,\mathrm{d}s + \xi\int_0^t \frac{(t-s)^{H-\frac{1}{2}}}{\Gamma(H+\frac{1}{2})} v_s\,\mathrm{d}s\right\}.$$

Inverting the integral yields the claim. □

2.4.3.2 Rough Heston

After the rescaling $Y_t^\varepsilon := \varepsilon^2 Y_t$ and $X_t^\varepsilon := \varepsilon^2 X_t$, this model introduced in Section 2.4.2.3 takes the form

$$\begin{cases} X_t^\varepsilon &= -\dfrac{1}{2}\displaystyle\int_0^t Y_s^\varepsilon\,\mathrm{d}s + \varepsilon\int_0^t \sqrt{Y_s^\varepsilon}\mathrm{d}B_s, \\ Y_t^\varepsilon &= \varepsilon^2 y_0 + \displaystyle\int_0^t \kappa\frac{(t-s)^{H-\frac{1}{2}}}{\Gamma(H+\frac{1}{2})}\left(\varepsilon^2\theta - Y_s^\varepsilon\right)\mathrm{d}s + \varepsilon\int_0^t \xi\frac{(t-s)^{H-\frac{1}{2}}}{\Gamma(H+\frac{1}{2})}\sqrt{Y_s^\varepsilon}\mathrm{d}W_s. \end{cases}$$

Clearly **H1 - H3** hold and we recall that $\mathcal{U}$ is the set of coefficients such that pathwise uniqueness, and hence **H4**, hold. We appeal to the uniqueness relaxation in the same way as the small-time case to prove the following result, which extends [67, Theorem 1.1] to the rough case.

Corollary 2.4.17. *If the rough Heston coefficients belong to $\mathcal{U}$ and $H \in (0, \frac{1}{2})$, then the following hold*

(L1) $(X^\varepsilon, Y^\varepsilon) \sim \mathrm{LDP}\left(I, \varepsilon^{-2}\right)$ *where*

$$I(\phi,\varphi) = \frac{1}{2}\int_0^T \left(u_t^2 + v_t^2\right)\mathrm{d}t, \quad \textit{where} \begin{cases} u &= \dfrac{1}{\rho}\left(\dfrac{\dot{\phi}}{\sqrt{\varphi}} + \dfrac{1}{2}\sqrt{\varphi} - \rho v\right)\mathbb{1}_{\varphi>0}, \\ v &= \dfrac{1}{\xi\sqrt{\varphi}}\left(\mathrm{D}^{H+\frac{1}{2}}(\varphi) + \kappa\varphi\right)\mathbb{1}_{\varphi>0}, \end{cases}$$

where $\phi \in \mathcal{AC}_0$ and $\varphi \in \mathrm{I}_0^{H+\frac{1}{2}}(L^1)$ and infinity otherwise.

(L2) *$X^\varepsilon \sim \mathrm{LDP}\left(I^X, \varepsilon^{-2}\right)$ with $I^X(\phi) = \inf\left\{I(\phi,\varphi) : \varphi \in \mathrm{I}_0^{H+\frac{1}{2}}\right\}$ if $\phi \in \mathcal{AC}_0$ and infinity otherwise.*

(L3) *For each $t \in \mathbb{T}$, $\varepsilon^2 X_t \sim \mathrm{LDP}\left(I_t^X, \varepsilon^{-2}\right)$, where $I_t^X(x) = \inf\left\{I^X(\phi) : \phi_t = x\right\}$.*

Proof. The proof is similar to the small-time case. The potential rate function for the couple is

$$I(\phi,\varphi) = \inf\left\{\frac{1}{2}\int_0^T \left(u_t^2 + v_t^2\right) \mathrm{d}t : \quad u, v \in L^2, \right.$$
$$\left. \phi_t = -\frac{1}{2}\int_0^t \varphi_s \,\mathrm{d}s + \int_0^t \sqrt{\varphi_s}\left(\overline{\rho} u_s + \rho v_s\right) \mathrm{d}s, \quad \varphi_t = \int_0^t \frac{(t-s)^{H-\frac{1}{2}}}{\Gamma(H+\frac{1}{2})}\Big(-\kappa\varphi_s + \xi\sqrt{\varphi_s} v_s\Big) \,\mathrm{d}s\right\}.$$

The same arguments that were used to prove Lemmata 2.4.9 and 2.4.11 in the small-time case can be applied again here. They entail that Assumption 2.3.3 holds and hence Theorem 2.3.11 applies, and the form of the rate function in **(L1)** follows by inverting the relationships between (u, v) and (ϕ, φ). **(L2)** and **(L3)** follow from the same steps as in Corollary 2.4.15. □

2.4.3.3 Implied volatility asymptotics

We can also obtain implied volatility asymptotics since, by the same arguments as before, $\exp(X^\varepsilon)$ is a martingale in both the rough Stein-Stein and rough Heston models.

Corollary 2.4.18. *In both the rough Stein-Stein and rough Heston models, for each $t \in \mathbb{T}$, the implied volatility $\widehat{\sigma}$ satisfies*

$$\lim_{k\uparrow\infty} \frac{\widehat{\sigma}(t,k)^2 t}{k} = \frac{1}{2}\left(\inf_{y\geq 1} I_t^X(y)\right)^{-1},$$

where I_t^X is the respective rate function, given in Corollaries 2.4.15(**L3**) *and 2.4.17*(**L3**).

Proof. Mapping ε^2 to $1/k$ we have from Corollaries 2.4.15 and 2.4.17 respectively that, for each $t \in \mathbb{T}$,

$$\lim_{k\uparrow\infty} \frac{1}{k} \log \mathbb{P}(X_t \geq k) = -\inf_{y\geq 1} I_t^X(y).$$

In the Black-Scholes model with constant volatility $\sigma > 0$, we can directly compute

$$\lim_{k\uparrow\infty} \frac{1}{k^2} \log \mathbb{P}(X_t \geq k) = -\frac{1}{2\sigma^2 t},$$

and, similarly to the small-time case, the proof follows from [117, Corollary 7.1]. $\square$

2.5 Technical large deviations proofs

2.5.1 Abstract relaxation: Proof of Theorem 2.3.5

The proof follows [45, Theorem 4.4]. The lower bound proof stands as it is until the last series of inequalities. For any sequence $\{v^\varepsilon\}_{\varepsilon>0}$ in $\mathcal{A}_M, M > 0$ converging weakly to v and any $F \in \mathcal{C}_b(\mathcal{W}^d : \mathbb{R})$, define the function $x : \mathbb{R}_+ \to \mathbb{R}$ by $x(\varepsilon) := \mathbb{E}\left[F \circ \mathcal{G}^\varepsilon\left(W + \varepsilon^{-1}\int_0^\cdot v_t^\varepsilon \,\mathrm{d}t\right)\right]$. Theorem 2.3.5(i) entails the existence of $\phi \in \mathcal{G}^0_{v,M}$ such that the subsequence $y : \mathbb{R}_+ \to \mathbb{R}$ defined as $y(\varepsilon) := \inf_{\alpha<\varepsilon} x(\alpha)$ has a subsequence converging to $\mathbb{E}\left[F(\phi)\right]$. By definition $\liminf_{\varepsilon\downarrow 0} x_\varepsilon = \lim_{\varepsilon\downarrow 0} y_\varepsilon$, which implies that y has a limit in $[-\infty, +\infty]$ and by uniqueness this limit must be $\mathbb{E}\left[F(\phi)\right]$. Therefore we deduce:

$$\begin{aligned}
&\liminf_{\varepsilon\downarrow 0} \mathbb{E}\left[\frac{1}{2}\int_0^T |v_t^\varepsilon|^2 \,\mathrm{d}t + F\circ\mathcal{G}^\varepsilon\left(W + \varepsilon^{-1}\int_0^\cdot v_t^\varepsilon \,\mathrm{d}t\right)\right] \\
&\qquad \geq \mathbb{E}\left[\frac{1}{2}\int_0^T |v_t|^2 \,\mathrm{d}t + F(\phi)\right] \geq \inf\left\{\frac{1}{2}\int_0^T |v_t|^2 \,\mathrm{d}t + F(\phi) : v \in L^2 \text{ such that } \phi \in \mathcal{G}_v^0 \cap \mathcal{W}^d\right\},
\end{aligned}$$

which suggests the potential rate function I defined in (2.3.2) and concludes the proof of the lower bound.

Then we prove the Laplace principle upper bound, for all $F \in \mathcal{C}_b(\mathcal{W}^d : \mathbb{R})$:

$$\limsup_{\varepsilon\downarrow 0} -\varepsilon^2 \log \mathbb{E}\left[\mathrm{e}^{-F\circ\mathcal{G}^\varepsilon(W)/\varepsilon^2}\right] \leq \inf_{\psi\in\mathcal{W}^d}\{I(\psi) + F(\psi)\}.$$

Assume that the right-hand side is finite otherwise there is nothing to prove. Fix $\epsilon > 0$ and let $\phi \in \mathcal{W}^d$ such that

$$I(\phi) + F(\phi) \leq \inf_{\psi\in\mathcal{W}^d}\{I(\psi) + F(\psi)\} + \epsilon.$$

Since F is continuous at ϕ, there exists $\delta \in (0, \epsilon)$ such that $|F(\phi) - F(\varphi)| \leq \epsilon$ for all $\varphi \in \mathcal{W}^d$ such that $\|\phi - \varphi\|_{\mathbb{T}} \leq \delta$. If ϕ is uniquely characterised then the proof is the same as in [45]. Otherwise, by Theorem 2.3.5(iii), we can choose ϕ^δ uniquely characterised such that $\left\|\phi-\phi^\delta\right\|_{\mathbb{T}} \leq \delta$ and $\left|I(\phi)-I(\phi^\delta)\right| \leq \delta$, which implies $\left|I(\phi)+F(\phi)-I(\phi^\delta)-F(\phi^\delta)\right| \leq 2\epsilon$. Hence, combining inequalities we obtain

$$I(\phi^\delta)+F(\phi^\delta) \leq \inf_{\psi\in\mathcal{W}^d}\{I(\psi)+F(\psi)\}+3\epsilon.$$

Moreover, there exist $\{v^n\}_{n\in\mathbb{N}}$ in L^2 such that (2.3.3) is satisfied with ϕ^δ and $m \geq 1/\epsilon$ such that

$$\mathcal{G}^0_{v^m}=\{\phi^\delta\} \quad \text{and} \quad \frac{1}{2}\int_0^T |v^m_t|^2\,\mathrm{d}t \leq I(\phi^\delta)+\frac{1}{m} \leq I(\phi^\delta)+\epsilon,$$

and therefore the remainder of the upper bound proof unfolds identically.

Along the subsequence $\{\varepsilon_n\}_{n\geq 0}$, $\mathcal{G}^{\varepsilon_n}(W+\varepsilon_n^{-1}\int_0^\cdot v^m_t\,\mathrm{d}t)$ converges in distribution to ϕ^δ by item (i). Using the variational representation formula (2.1.3) and the convergence we obtain

$$\begin{aligned}
\limsup_{n\uparrow\infty} -\varepsilon_n^2\log\mathbb{E}\left[\exp\left\{-\frac{F\circ\mathcal{G}^{\varepsilon_n}(W)}{\varepsilon_n^2}\right\}\right] &= \limsup_{n\uparrow\infty}\inf_{v\in\mathcal{A}}\mathbb{E}\left[\frac{1}{2}\int_0^T |v_t|^2\,\mathrm{d}t + F\circ\mathcal{G}^{\varepsilon_n}\left(W+\varepsilon_n^{-1}\int_0^\cdot v_t\,\mathrm{d}t\right)\right]\\
&\leq \limsup_{n\uparrow\infty}\mathbb{E}\left[\frac{1}{2}\int_0^T |v^m_t|^2\,\mathrm{d}t + F\circ\mathcal{G}^{\varepsilon_n}\left(W+\varepsilon_n^{-1}\int_0^\cdot v^m_t\,\mathrm{d}t\right)\right]\\
&= \frac{1}{2}\int_0^T |v^m_t|^2\,\mathrm{d}t + F(\phi^\delta)\\
&\leq I(\phi^\delta)+F(\phi^\delta)+\epsilon \leq \inf_{\psi\in\Omega}\{I(\psi)+F(\psi)\}+4\epsilon.
\end{aligned}$$

Since $\epsilon>0$ is arbitrary this concludes the proof. □

2.5.2 LDP moment bounds: Proof of Lemma 2.3.13

Let us fix $p\geq 2$, $M>0$, $v\in\mathcal{A}_M$, $\varepsilon>0$ and $t\in\mathbb{T}$. Let $\tau_n := \inf\{t\geq 0 : |X^{\varepsilon,v}_t|\geq n\}\wedge T$ for all $n\in\mathbb{N}$. For clarity we write $b^n_s := b_\varepsilon(s, X^{\varepsilon,v}_s \mathbb{1}_{s\leq\tau_n})$ and $\sigma^n_s := \sigma_\varepsilon(s, X^{\varepsilon,v}_s\mathbb{1}_{s\leq\tau_n})$. We start by assuming that all the coefficients satisfy the linear growth condition **H3a**. We fix $n\in\mathbb{N}$ and observe that, almost surely:

$$\begin{aligned}
|X^{\varepsilon,v}_t|^p\,\mathbb{1}_{t\leq\tau_n} \leq& 4^{p-1}\left[|X^\varepsilon_0|^p + \left|\int_0^t K(t-s)b^n_s\,\mathrm{d}s\right|^p + \left|\int_0^t K(t-s)\sigma^n_s v_s\,\mathrm{d}s\right|^p + \vartheta^p_\varepsilon\left|\int_0^t K(t-s)\sigma^n_s\mathrm{d}W_s\right|^p\right]\\
=:& 4^{p-1}\Big[|X^\varepsilon_0|^p + \mathrm{I}_n + \mathrm{II}_n + \mathrm{III}_n\Big],
\end{aligned} \tag{2.5.1}$$

because if $t > \tau_n$ then the left-hand side is zero while the right-hand side is non-negative, and if $t \leq \tau_n$ then $s \leq \tau_n$ for all $s \in [0, t]$ and the τ_n dependence vanishes on both sides of the inequality. For ε small enough we can bound $|X_0^\varepsilon|$ by $2|X_0|$ and ϑ_ε by 1 and we will do so repetitively in the sequel. Using Hölder's and Jensen's inequalities, we obtain the following estimates almost surely:

$$\mathrm{I}_n \leq \left[\int_0^t |K(t-s)|^{\frac{4}{p}} |b_s^n|^2 \, \mathrm{d}s\right]^{\frac{p}{2}} \left[\int_0^t |K(t-s)|^{2-\frac{4}{p}} \, \mathrm{d}s\right]^{\frac{p}{2}} \leq t^{\frac{p^2+4}{2(p-2)}} \|K\|_2^{p-2} \int_0^t |K(t-s)|^2 |b_s^n|^p \, \mathrm{d}s, \tag{2.5.2}$$

and

$$\mathrm{II}_n \leq M^{\frac{p}{2}} \left[\int_0^t |K(t-s)|^{2-4/p} |K(t-s)|^{\frac{4}{p}} |\sigma_s^n|^2 \, \mathrm{d}s\right]^{\frac{p}{2}} \leq M^{\frac{p}{2}} \|K\|_2^{p-2} \int_0^t |K(t-s)|^2 |\sigma_s^n|^p \, \mathrm{d}s, \tag{2.5.3}$$

where we also used that $\int_0^T |v_s|^2 \, \mathrm{d}s \leq M$ almost surely. Notice that for fixed $t \in \mathbb{T}$, $\{\int_0^u K(t-s)\sigma_s^n \mathrm{d}W_s : u \in [0,t]\}$ is a continuous local martingale and is bounded in $L^2(\Omega)$. Hence, using Burkholder-Davis-Gundy (BDG) inequality and similar calculations as (2.5.3) there exists $C_p > 0$ such that

$$\mathbb{E}[\mathrm{III}_n] \leq C_p \|K\|_2^{p-2} \int_0^t |K(t-s)|^2 \, \mathbb{E}|\sigma_s^n|^p \, \mathrm{d}s.$$

From the linear growth condition on b_ε and σ_ε (uniform in $\varepsilon > 0$) we deduce that there exists $C_1 > 0$ independent of ε, v, n, t such that, for all $n \in \mathbb{N}$, $f_t^n := \mathbb{E}\left[|X_t^{\varepsilon,v}|^p \, \mathbb{1}_{t \leq \tau_n}\right]$ satisfies the inequality

$$f_t^n \leq C_1 + C_1 \int_0^t |K(t-s)|^2 f_s^n \, \mathrm{d}s. \tag{2.5.4}$$

The following lemma (Lemma 2.5.1) yields a uniform bound in both $n \in \mathbb{N}$ and $t \in \mathbb{T}$ for f_t^n. Taking the limit as n goes to infinity and using Fatou's lemma concludes the first part of the proof.

Lemma 2.5.1. *Let $f : \mathbb{T} \to \mathbb{R}_+$ and K a kernel satisfying Assumption 2.2.3. If there exists $c_1, c_2 \geq 0$ such that*

$$f(t) \leq c_1 + c_2 \int_0^t |K(t-s)|^2 f(s) \, \mathrm{d}s, \qquad \text{for all } t \in \mathbb{T},$$

then f is uniformly bounded on $\mathbb{T}$ by a constant depending only on $c_1, c_2, \|K\|_2, T$. If $c_1 = 0$ then $f = 0$.

Proof of Lemma 2.5.1. By definition $|K(t-s)|^2 = \sum_{i,j=1}^d |K_{ij}(t-s)|^2$, and $\widetilde{K}(t,s) := c_2 |K(t-s)|^2 \mathbb{1}_{s\le t}$ is a Volterra kernel in the sense of [134, Definition 9.2.1]. Following similar arguments as in the proof of [3, Lemma 3.1], the generalised Gronwall lemma [134, Theorem 9.8.2] yields the bound

$$f(t) \le c_1 - c_1 \int_0^t \widetilde{R}(s)\,\mathrm{d}s \le c_1 - c_1 \int_0^T \widetilde{R}(s)\,\mathrm{d}s,$$

where $\widetilde{R}$ is the (non-positive) resolvent of second kind of $-\widetilde{K}$ [3, Equation (2.11)], proving the lemma. $\square$

If only **H3b** and Assumption 2.3.9 hold with an autonomous sub-system Υ (see Definition 2.2.6), then by the previous calculations for all $l \in \Upsilon$, the components $(X^{\varepsilon,v})^{(l)}$ satisfy the bound (2.3.8) because their coefficients have linear growth. Then we turn our attention to the components $(X^{\varepsilon,v})^{(i)}$, $i \notin \Upsilon$. Using (2.3.6) and Hölder's inequality as in (2.5.3), we obtain that for all $1 \le j \le d$ such that $K_{ij} \neq 0$ and for all $1 \le k \le m$:

$$\begin{aligned}
&\mathbb{E}\left[\left(\int_0^t \left|K_{ij}(t-s)\sigma_\varepsilon^{(jk)}\left(s, X_s^{\varepsilon,v}\mathbb{1}_{s\le\tau_n}\right)\right|^2 \mathrm{d}s\right)^{p/2}\right] \\
&\le C_\Upsilon^p \mathbb{E}\left[\left(\int_0^t |K_{ij}(t-s)|^2 \left(1 + \left|X_s^{\varepsilon,v}\mathbb{1}_{s\le\tau_n}\right|_{\Upsilon^c} + \left|\Gamma\left((X_s^{\varepsilon,v}\mathbb{1}_{s\le\tau_n})^{(\Upsilon)}\right)\right|\right)^2 \mathrm{d}s\right)^{p/2}\right] \\
&\le C_\Upsilon^p \mathbb{E}\left[\int_0^t |K_{ij}(t-s)|^2 \left(1 + |X_s^{\varepsilon,v}\mathbb{1}_{s\le\tau_n}|_{\Upsilon^c} + \left|\Gamma\left((X_s^{\varepsilon,v}\mathbb{1}_{s\le\tau_n})^{(\Upsilon)}\right)\right|\right)^p \mathrm{d}s\right] \left(\int_0^t |K_{ij}(t-s)|^2\,\mathrm{d}s\right)^{p/2-1} \\
&\le 3^{p-1} C_\Upsilon^p \|K\|_2^{p-2} \int_0^t |K_{ij}(t-s)|^2 \,\mathbb{E}\left[1 + \left|X_s^{\varepsilon,v}\mathbb{1}_{s\le\tau_n}\right|_{\Upsilon^c}^p + \left|\Gamma\left((X_s^{\varepsilon,v}\mathbb{1}_{s\le\tau_n})^{(\Upsilon)}\right)\right|^p\right] \mathrm{d}s \\
&\le C_2 + C_2 \int_0^t |K_{ij}(t-s)|^2 \,\mathbb{E}\left[|X_s^{\varepsilon,v}\mathbb{1}_{s\le\tau_n}|_{\Upsilon^c}^p\right]\,\mathrm{d}s,
\end{aligned} \tag{2.5.5}$$

for some $C_2 > 0$. Applying the same calculations to the other terms and summing all the coefficients we fall back on (2.5.4). Taking the limit and applying Fatou's lemma again conclude the proof. $\square$

2.5.3 LDP tightness: Proof of Lemma 2.3.15

Let us fix $p > 2 \vee 2/\gamma$, $M > 0$, a family $\{v^\varepsilon\}_{\varepsilon>0}$ in $\mathcal{A}_M$ and $\varepsilon > 0$. For clarity we will write $b_u := b_\varepsilon(u, X_u^{\varepsilon,v})$ and $\sigma_u := \sigma_\varepsilon(u, X_u^{\varepsilon,v})$ for all $u \in \mathbb{T}$. Then, for all $0 \le s < t \le T$, using Cauchy-Schwarz and BDG inequalities as in the previous proof we obtain:

$$\begin{aligned}
\mathbb{E}\left[\left|X_t^{\varepsilon,v^\varepsilon}-X_s^{\varepsilon,v^\varepsilon}\right|^p\right] \leq & \, 6^{p-1}\mathbb{E}\left[\left|\int_0^s \big(K(t-u)-K(s-u)\big)b_u\,\mathrm{d}u\right|^p\right] \\
& + 6^{p-1}\mathbb{E}\left[\left|\int_s^t K(t-u)b_u\,\mathrm{d}u\right|^p\right] \\
& + 6^{p-1}M^{p/2}\mathbb{E}\left[\left(\int_0^s \left|\big(K(t-u)-K(s-u)\big)\sigma_u\right|^2\,\mathrm{d}u\right)^{p/2}\right] \\
& + 6^{p-1}M^{p/2}\mathbb{E}\left[\left(\int_s^t \left|K(t-u)\sigma_u\right|^2\,\mathrm{d}u\right)^{p/2}\right] \\
& + 6^{p-1}\vartheta_\varepsilon^p C_p\mathbb{E}\left[\left(\int_0^s \left|\big(K(t-u)-K(s-u)\big)\sigma_u\right|^2\,\mathrm{d}u\right)^{p/2}\right] \\
& + 6^{p-1}\vartheta_\varepsilon^p C_p\mathbb{E}\left[\left(\int_s^t \left|K(t-u)\sigma_u\right|^2\,\mathrm{d}u\right)^{p/2}\right].
\end{aligned}$$

In a first step we assume that all the coefficients satisfy the linear growth condition **H3a**. Analogous calculations to the proof of Lemma 2.3.13, bounds on $\sup_{t\in\mathbb{T},\varepsilon>0}\mathbb{E}\left[\left|X_t^{\varepsilon,v^\varepsilon}\right|^p\right]$, linear growth from **H3a** and Assumption 2.2.3 lead to

$$\begin{aligned}
\mathbb{E}\left[\left|X_t^{\varepsilon,v^\varepsilon}-X_s^{\varepsilon,v^\varepsilon}\right|^p\right] & \leq C_1\left(\left|\int_0^s \big(K(t-u)-K(s-u)\big)^2\,\mathrm{d}u\right|^{p/2}+\left|\int_s^t K(t-u)^2\,\mathrm{d}u\right|^{p/2}\right) \\
& = C_1\left(\left|\int_0^s \big(K(u+t-s)-K(u)\big)^2\,\mathrm{d}u\right|^{p/2}+\left|\int_0^{t-s} K(u)^2\,\mathrm{d}u\right|^{p/2}\right) \\
& \leq C_2(t-s)^{\gamma p/2},
\end{aligned}$$

for some $C_1, C_2 > 0$ independent of $\varepsilon, v^\varepsilon, s, t$. Again, if there are components such that only **H3b** holds with Assumption 2.3.9 then following the example of (2.5.5) yields the same result. Then Kolmogorov continuity theorem asserts that $X^{\varepsilon,v^\varepsilon}$ admits a version which is Hölder continuous on $\mathbb{T}$ of any order $\alpha < \gamma/2 - 1/p$, uniformly in $\varepsilon > 0$ because C_2 does not depend on ε, and which satisfies (2.3.9). Furthermore, Aldous theorem [35, Theorem 16.10] states that the sequence $\{X^{\varepsilon,v^\varepsilon}\}_{\varepsilon>0}$ is tight. □

2.5.4 LDP compactness: Proof of Lemma 2.3.18

We prove that for all $M > 0$, the sublevel sets

$$L_M := \{\phi \in \mathcal{W}^d : I(\phi) \leq M\}$$

of the map $I : \mathcal{W}^d \to \mathbb{R}$ given by (2.3.2) or more precisely by

$$I(\phi) = \inf\left\{\frac{1}{2}\int_0^T |v_s|^2 \,\mathrm{d}s : v \in L^2, \phi_t = x_0 + \int_0^t K(t-s)\Big[b(s,\phi_s) + \sigma(s,\phi_s)v_s\Big]\mathrm{d}s\right\} \tag{2.5.6}$$

are compact. Fix $M > 0$ and consider an arbitrary sequence $\mathcal{J} := \{\phi^n\}_{n\in\mathbb{N}} \subset L_M$; we will show that there exists a converging subsequence the limit of which belongs to L_M. Interestingly enough, the proof parallels, in a deterministic context, the proofs of bound, Hölder continuity and convergence of $X^{\varepsilon,v}$.

Relative compactness. According to Arzelà-Ascoli's theorem, the family $\mathcal{J}$ is relatively compact in $\mathcal{W}^d$ if and only if $\{\phi^n_t\}$ is bounded uniformly in $n \in \mathbb{N}$ and in $t \in \mathbb{T}$ and $\mathcal{J}$ is equicontinuous. Moreover, for all $n \in \mathbb{N}$ and all $t \in \mathbb{T}$, there exists $v^n \in L^2$ such that $\frac{1}{2}\int_0^T |v^n_t|^2 \,\mathrm{d}t \le M$ and $\phi^n \in \mathcal{G}^0_{v^n}$, which means $v^n \in \mathcal{S}_{2M}$ and

$$\phi^n_t = x_0 + \int_0^t K(t-s)\Big[b(s,\phi^n_s) + \sigma(s,\phi^n_s)v^n_s\Big]\mathrm{d}s.$$

Hence Remarks 2.3.14 and 2.3.16 grant the uniform bound and equicontinuity respectively. Therefore $\mathcal{J}$ is relatively compact which entails that L_M is relatively compact for any $M > 0$.

Closure. Let $\{\phi^n\}_{n\in\mathbb{N}}$ be a converging sequence of L_M and denote its limit by $\phi \in \mathcal{W}^d$. The controls v^n associated to ϕ^n through (2.5.6) belong to $\mathcal{S}_{2M}$ which is a compact space with respect to the weak topology. Hence there exists a subsequence $\{n_k\}_{k\in\mathbb{N}}$ such that v^{n_k} converges weakly in L^2 to a limit $v \in \mathcal{S}_{2M}$ and $\lim_{k\uparrow\infty}\phi^{n_k} = \phi$. Now let us prove that $\phi \in L_M$. For clarity we replace n_k by n from now on. The convergence as n goes to $+\infty$ and the continuity of the paths entail

$$\sup_{n\in\mathbb{N}}\sup_{t\in\mathbb{T}}\big(|\phi^n_t| + |\phi_t|\big) < +\infty,$$

such that the paths lie in compact subsets of $\mathbb{R}^d$ and **H2** asserts that uniform continuity of the coefficients b and σ hold. Therefore they admit continuous moduli of continuity that we respectively name ρ_b and ρ_σ. Using Cauchy-Schwarz inequality and **H3** we get that for all $t \in \mathbb{T}$:

$$\left|\int_0^t K(t-s)b(s,\phi_s^n)\mathrm{d}s - \int_0^t K(t-s)b(s,\phi_s)\mathrm{d}s\right| \le \|K\|_{L^1} \left\|\rho_b\big(|\phi_\cdot^n - \phi_\cdot|\big)\right\|_{\mathbb{T}},$$

$$\begin{aligned}
&\left|\int_0^t K(t-s)\sigma(s,\phi_s^n)v_s^n\mathrm{d}s - \int_0^t K(t-s)\sigma(s,\phi_s)v_s\mathrm{d}s\right| \\
&\quad \le \int_0^t \left|K(t-s)\big(\sigma(s,\phi_s^n)-\sigma(s,\phi_s)\big)v_s^n\right|\mathrm{d}s + \int_0^t |K(t-s)\sigma(s,\phi_s)(v_s^n - v_s)|\,\mathrm{d}s \\
&\quad \le \|K\|_2\,\|v^n\|_2\,\|\rho_\sigma(|\phi_\cdot^n - \phi_\cdot|)\|_{\mathbb{T}} + \|\sigma(\phi)\|_{\mathbb{T}} \int_0^t |K(t-s)(v_s^n - v_s)|\,\mathrm{d}s,
\end{aligned}$$

and both estimates converge towards zero as n tends to infinity. Therefore, for all $t \in \mathbb{T}$

$$\begin{aligned}
\phi_t = \lim_{n\uparrow\infty}\phi_t^n &= \lim_{n\uparrow\infty}\left(x_0 + \int_0^t K(t-s)\Big[b(s,\phi_s^n) + \sigma(s,\phi_s^n)v_s^n\Big]\mathrm{d}s\right) \\
&= x_0 + \int_0^t K(t-s)\Big[b(s,\phi_s) + \sigma(s,\phi_s)v_s\Big]\mathrm{d}s,
\end{aligned}$$

so that $\phi \in L_M$ since $v \in \mathcal{S}_{2M}$, which concludes the proof of the closure and therefore of the compactness of L_M. $\square$

2.6 Technical moderate deviations proofs

2.6.1 MDP moment bounds: Proof of Lemma 2.3.21

Let $p \ge 2$, $M > 0$, $v \in \mathcal{A}_M$, $\varepsilon > 0$ and $t \in \mathbb{T}$. Starting from (2.3.17), we use Cauchy-Schwarz and BDG inequalities to obtain

$$\begin{aligned}
\mathbb{E}\left[|\eta_t^{\varepsilon,v}|^p\right] \le{}& 5^{p-1}\frac{|X_0^\varepsilon - x_0|^p}{(\vartheta_\varepsilon h_\varepsilon)^p} \\
&+ 5^{p-1}\mathbb{E}\left|\int_0^t K(t-s)\frac{b_\varepsilon\big(s,\Theta_s^{\varepsilon,v}\big) - b\big(s,\Theta_s^{\varepsilon,v}\big)}{\vartheta_\varepsilon h_\varepsilon}\,\mathrm{d}s\right|^p \\
&+ 5^{p-1}\mathbb{E}\left|\int_0^t K(t-s)\frac{b\big(s,\Theta_s^{\varepsilon,v}\big) - b(s,\overline{X}_s)}{\vartheta_\varepsilon h_\varepsilon}\mathrm{d}s\right|^p \\
&+ 5^{p-1}M^{p/2}\mathbb{E}\left[\left(\int_0^t \left|K(t-s)\sigma_\varepsilon\big(s,\Theta_s^{\varepsilon,v}\big)\right|^2\mathrm{d}s\right)^{p/2}\right] \\
&+ \frac{5^{p-1}C_p}{h_\varepsilon^p}\mathbb{E}\left[\left(\int_0^t \left|K(t-s)\sigma_\varepsilon\big(s,\Theta_s^{\varepsilon,v}\big)\right|^2\,\mathrm{d}s\right)^{p/2}\right].
\end{aligned} \tag{2.6.1}$$

The first term converges by **H7** and is thus bounded. Notice that **H8** entails

$$\mathbb{E}\left|\int_0^t K(t-s)\frac{b_\varepsilon\big(s,\Theta_s^{\varepsilon,v}\big)-b\big(s,\Theta_s^{\varepsilon,v}\big)}{\vartheta_\varepsilon h_\varepsilon}\,\mathrm{d}s\right|^p \le \left(\frac{\nu_\varepsilon}{\vartheta_\varepsilon h_\varepsilon}\right)^p \mathbb{E}\left|\int_0^t K(t-s)\Xi\big(\Theta_s^{\varepsilon,v}\big)\,\mathrm{d}s\right|^p,$$

which is bounded because $\nu_\varepsilon(\vartheta_\varepsilon h_\varepsilon)^{-1}$ tends to zero and, using Cauchy-Schwarz and Jensen's inequalities in the same way as (2.5.2) and the bound (2.3.19),

$$\mathbb{E}\left|\int_0^t K(t-s)\Xi\big(\Theta_s^{\varepsilon,v}\big)\,\mathrm{d}s\right|^p \le \|K\|_2^p\, t^{p/2-1}\sup_{s\le T}\mathbb{E}\Big[\big|\Xi\big(\Theta_s^{\varepsilon,v}\big)\big|^p\Big] \le C_1,$$

where C_1 is a positive constant that does not depend on ε. Since b is globally Lipschitz continuous, there exists $C_b > 0$ such that for all $s \in \mathbb{T}$:

$$\big|b\big(s,\overline{X}_s+\vartheta_\varepsilon h_\varepsilon \eta_s^{\varepsilon,v}\big)-b\big(s,\overline{X}_s\big)\big| \le C_b\,\vartheta_\varepsilon h_\varepsilon\,|\eta_s^{\varepsilon,v}|\,. \tag{2.6.2}$$

Therefore, using Cauchy-Schwarz and Jensen's inequalities again

$$\begin{aligned}\mathbb{E}\left|\int_0^t K(t-s)\frac{b\big(s,\Theta_s^{\varepsilon,v}\big)-b\big(s,\overline{X}_s\big)}{\vartheta_\varepsilon h_\varepsilon}\mathrm{d}s\right|^p &\le \mathbb{E}\left|\int_0^t K(t-s)C_b\;|\eta_s^{\varepsilon,v}|\;\mathrm{d}s\right|^p\\ &\le C_b^p t^{\frac{p^2+4}{2(p-2)}}\,\|K\|_2^{p-2}\int_0^t |K(t-s)|^2\,\mathbb{E}\big[\,|\eta_s^{\varepsilon,v}|^p\,\big]\,\mathrm{d}s.\end{aligned}$$

The last two terms of (2.6.1) are also uniformly bounded in n and ε because, similarly to (2.5.3),

$$\begin{aligned}&\mathbb{E}\left[\left(\int_0^t \big|K(t-s)\sigma_\varepsilon\big(s,\Theta_s^{\varepsilon,v}\big)\big|^2\,\mathrm{d}s\right)^{p/2}\right]\\ &\le \mathbb{E}\left[\left(\int_0^t |K(t-s)|^2\,\big|\sigma_\varepsilon\big(s,\Theta_s^{\varepsilon,v}\big)\big|^p\;\mathrm{d}s\right)\left(\int_0^t |K(t-s)|^2\;\mathrm{d}s\right)^{\frac{p-2}{2}}\right]\\ &\le \|K\|_2^p\, C_L\big(1+\sup_{t\in\mathbb{T}}\mathbb{E}\left[|\Theta_t^{\varepsilon,v}|^p\right]\big),\end{aligned} \tag{2.6.3}$$

by Hölder's inequality and the linear growth condition **H3a**. If the latter fails we rely on **H3b**, Assumption 2.3.9 and the same calculations as in (2.5.5) to obtain a similar bound.

Overall this results in the existence of a constant $C_2 > 0$ independent of ε, v, t such that

$$\mathbb{E}\left[|\eta_t^{\varepsilon,v}|^p\right] \le C_2 + C_2\int_0^t |K(t-s)|^2\,\mathbb{E}\left[|\eta_s^{\varepsilon,v}|^p\right]\mathrm{d}s, \qquad \text{for all } t\in\mathbb{T},$$

and Lemma 2.5.1 yields the bound uniform in ε and t. ☐

2.6.2 MDP tightness: Proof of Lemma 2.3.22

Let $p > 2 \vee 2/\gamma$, $M > 0$, a sequence $\{v^\varepsilon\}_{\varepsilon>0}$ in $\mathcal{A}_M$, $\varepsilon > 0$, and $0 \le s < t \le T$. We proceed as in Lemma 2.3.15; starting from (2.3.17), applying consecutively **H8**, then (2.6.2), Cauchy-Schwarz and BDG inequalities we obtain

$$
\begin{aligned}
&\mathbb{E}\left[\left|\eta_t^{\varepsilon,v^\varepsilon} - \eta_s^{\varepsilon,v^\varepsilon}\right|^p\right] \\
\le& 8^{p-1}\mathbb{E}\left|\frac{\nu_\varepsilon}{\vartheta_\varepsilon h_\varepsilon}\int_0^s (K(t-u)-K(s-u))\Xi\big(\Theta_u^{\varepsilon,v^\varepsilon}\big)\,\mathrm{d}u\right|^p \\
&+ 8^{p-1}\mathbb{E}\left|\frac{\nu_\varepsilon}{\vartheta_\varepsilon h_\varepsilon}\int_s^t K(t-u)\Xi\big(\Theta_u^{\varepsilon,v^\varepsilon}\big)\,\mathrm{d}u\right|^p \\
&+ 8^{p-1}\mathbb{E}\left|\int_0^s (K(t-u)-K(s-u))C_b\left|\eta_u^{\varepsilon,v^\varepsilon}\right|\mathrm{d}u\right|^p \\
&+ 8^{p-1}\mathbb{E}\left|\int_s^t K(t-u)C_b\left|\eta_u^{\varepsilon,v^\varepsilon}\right|\mathrm{d}u\right|^p \\
&+ 8^{p-1}M^{p/2}\mathbb{E}\left[\left(\int_0^s \left|(K(t-u)-K(s-u))\sigma_\varepsilon\big(u,\Theta_u^{\varepsilon,v^\varepsilon}\big)\right|^2\mathrm{d}u\right)^{p/2}\right] \\
&+ 8^{p-1}M^{p/2}\mathbb{E}\left[\left(\int_s^t \left|K(t-u)\sigma_\varepsilon\big(u,\Theta_u^{\varepsilon,v^\varepsilon}\big)\right|^2\mathrm{d}u\right)^{p/2}\right] \\
&+ \frac{8^{p-1}C_p}{h_\varepsilon^p}\mathbb{E}\left[\left(\int_0^s \left|(K(t-u)-K(s-u))\sigma_\varepsilon\big(u,\Theta_u^{\varepsilon,v^\varepsilon}\big)\right|^2\mathrm{d}u\right)^{p/2}\right] \\
&+ \frac{8^{p-1}C_p}{h_\varepsilon^p}\mathbb{E}\left[\left(\int_s^t \left|K(t-u)\sigma_\varepsilon\big(u,\Theta_u^{\varepsilon,v^\varepsilon}\big)\right|^2\mathrm{d}u\right)^{p/2}\right].
\end{aligned}
$$

In the first four terms, Cauchy-Schwarz inequality allows to separate the kernels from the random variables. For the last four terms, analogous calculations to (2.5.3) achieve a similar separation of kernels and random variables. Then linear growth or (2.3.9) and bounds on $\sup_{t\in\mathbb{T},\varepsilon>0}\mathbb{E}\left[\left|\Xi\big(\Theta_t^{\varepsilon,v^\varepsilon}\big)\right|^p\right]$ and $\sup_{t\in\mathbb{T},\varepsilon>0}\mathbb{E}\left[\left|\eta_t^{\varepsilon,v^\varepsilon}\right|^p\right]$ lead to the existence of $C_1 > 0$ independent of t and ε such that

$$
\begin{aligned}
\mathbb{E}\left[\left|\eta_t^{\varepsilon,v^\varepsilon} - \eta_s^{\varepsilon,v^\varepsilon}\right|^p\right] &\le C_1\left(\left|\int_0^s (K(t-u)-K(s-u))^2\,\mathrm{d}u\right|^{p/2} + \left|\int_s^t K(t-u)^2\,\mathrm{d}u\right|^{p/2}\right) \\
&= C_1\left(\left|\int_0^s (K(u+t-s)-K(u))^2\,\mathrm{d}u\right|^{p/2} + \left|\int_0^{t-s} K(u)^2\,\mathrm{d}u\right|^{p/2}\right).
\end{aligned}
$$

Hence Assumption 2.2.3 yields the existence of a constant $C_2 > 0$ such that

$$\mathbb{E}\left[\left|\eta_t^{\varepsilon,v^\varepsilon} - \eta_s^{\varepsilon,v^\varepsilon}\right|^p\right] \leq C_2(t-s)^{\gamma p/2}.$$

Then Kolmogorov continuity theorem asserts that $\eta^{\varepsilon,v^\varepsilon}$ admits a version which is Hölder continuous on $\mathbb{T}$ of any order $\alpha < \gamma/2 - 1/p$, uniformly in $\varepsilon > 0$ and which satisfies (2.3.22). Furthermore, Aldous theorem [35, Theorem 16.10] states that the sequence $\{\eta^{\varepsilon,v^\varepsilon}\}_{\varepsilon>0}$ is tight. □

2.6.3 MDP weak convergence: Proof of Lemma 2.3.23

We have shown in Lemma 2.3.22 that for any subsequence $\{\varepsilon_k\}_{k\in\mathbb{N}}$, $\{\eta^{\varepsilon_k,v^{\varepsilon_k}}\}_{k\in\mathbb{N}}$ and $\{v^{\varepsilon_k}\}_{k\in\mathbb{N}}$ are tight as families of random variables with values in $\mathcal{W}^d$ and $\mathcal{S}_M$ respectively. By Skorohod representation theorem we can work with almost sure convergence for the purpose of identifying the limit. Hence there exists a subsubsequence, denoted hereafter $\{\eta^k, v^k\}$, that converges almost surely in the product topology on $\mathcal{W}^d \times \mathcal{S}_M$ to some $\mathcal{W}^d \times \mathcal{S}_M$-valued limit (η^0, v) in a possibly different probability space $(\Omega^0, \mathcal{F}^0, \mathbb{P}^0)$ as n tends to $+\infty$. We also denote $\varepsilon_k, b_k, \sigma_k, X_0^k, \Theta^k$ along this subsequence.

The convergence of the couple also takes place in distribution, and we follow the same method as in the LDP case which comes from [62]. For all $t \in [0,T]$, let $\Psi_t : \mathcal{S}_M \times \mathcal{W}^d \to \mathbb{R}$ such that

$$\Psi_t(f,\omega) := \left|\omega_t - \int_0^t K(t-s)\left[\nabla b(s,\overline{X}_s)\omega_s + \sigma(s,\overline{X}_s)f_s\right]\mathrm{d}s\right| \wedge 1.$$

Clearly, Ψ_t is bounded and one can show its continuity along the same lines as in the LDP proof but in a simpler way because of the linearity. Therefore

$$\lim_{k\uparrow\infty} \mathbb{E}\left[\Psi_t(v^k,\eta^k)\right] = \mathbb{E}^0\left[\Psi_t(v,\eta^0)\right],$$

and we prove that the left-hand side is actually equal to zero. By **H5** and Taylor's formula there exists a sequence of $\mathbb{R}^d$-valued stochastic processes $\{R^\varepsilon\}_{\varepsilon>0}$ such that

$$\frac{b\big(s,\overline{X}_s + \vartheta_\varepsilon h_\varepsilon \eta_s^{\varepsilon,v}\big) - b\big(s,\overline{X}_s\big)}{\vartheta_\varepsilon h_\varepsilon} = \nabla b\big(s,\overline{X}_s\big)\eta_s^{\varepsilon,v} + R^\varepsilon(s), \tag{2.6.4}$$

and a constant $C_R > 0$ such that

$$|R^{\varepsilon}(s)| \leq C_R \vartheta_{\varepsilon} h_{\varepsilon} \, |\eta_s^{\varepsilon,v}|^2 \, .$$

We recall that $\left\|\nabla b(\cdot, \overline{X})\right\|_{\mathbb{T}}$ is finite by **H5** and observe that, by (2.3.21):

$$\mathbb{E}\big[\, |R^{\varepsilon}(u)|^p \,\big] \leq \big(C_R \vartheta_{\varepsilon} h_{\varepsilon}\big)^p \mathbb{E}\left[\left|\eta_s^{\varepsilon,v^{\varepsilon}}\right|^{2p}\right] < \infty. \tag{2.6.5}$$

Again starting from (2.3.17), we use **H8**, the Taylor estimate (2.6.4) and Itô isometry to get

$$\begin{aligned}
\mathbb{E}\left[\Psi_t(\eta^k, v^k)^2\right] \leq & \frac{5\left|X_0^k - x_0\right|^2}{\vartheta_{\varepsilon_k}^2 h_{\varepsilon_k}^2} \\
& + 5\left(\frac{\nu_{\varepsilon_k}}{\vartheta_{\varepsilon_k} h_{\varepsilon_k}}\right)^2 \mathbb{E}\left|\int_0^t K(t-s)\Xi\big(\Theta_s^k\big)\,\mathrm{d}s\right|^2 \\
& + 5\mathbb{E}\left|\int_0^t K(t-s)R^{\varepsilon}(s)\,\mathrm{d}s\right|^2 \\
& + 5\mathbb{E}\left|\int_0^t K(t-s)\Big[\sigma_k\big(s,\Theta_s^k\big) - \sigma\big(s,\overline{X}_s\big)\Big]v_s^k\,\mathrm{d}s\right|^2 \\
& + \frac{5}{h_{\varepsilon_k}^2}\mathbb{E}\left[\int_0^t \left|K(t-s)\sigma_k\big(s,\Theta_s^k\big)\right|^2 \mathrm{d}s\right] \\
=: & 5\big(\mathrm{I}_k + \mathrm{II}_k + \mathrm{III}_k + \mathrm{IV}_k + h_{\varepsilon_k}^{-2}\mathrm{V}_k\big).
\end{aligned}$$

H7 and **H8** tell us that $\mathrm{I}_k + \mathrm{II}_k$ tends to zero as ε goes to zero while an application of Cauchy-Schwarz inequality and the bound (2.6.5) yields the same conclusion for III_k.

To deal with IV_k, recall that Θ^k converges in distribution towards $\overline{X}$ as k tends to infinity. Since $\overline{X}$ is deterministic, the convergence actually takes place in probability, with respect to the topology of uniform convergence. Moreover Θ^k is uniformly bounded in $L^r(\Omega)$ for all $r > p$, thus the family $\{\left|\Theta^k\right|^p\}_{k\geq 0}$ is uniformly integrable. Therefore the convergence also occurs with respect to the $L^p(\Omega)$-norm.

The modulus of continuity of σ is only available on compact sets of $\mathbb{T}\times\mathbb{R}^d$ so we define a constant $N > \left\|\overline{X}\right\|_{\mathbb{T}}$ and introduce the following sets, for each $k \in \mathbb{N}$:

$$E_k := \left\{\omega \in \Omega : \left\|\Theta^k(\omega) - \overline{X}\right\|_{\mathbb{T}} \leq N\right\},$$

with the observation that $\lim_{k\uparrow\infty}\mathbb{P}(E_k) = 1$ thanks to the previous argument. Since $\overline{X}$ is uniformly bounded, $\left|\Theta_t^k(\omega)\right| \leq 2N$ for all $t \in \mathbb{T}, \omega \in E_k$, $k \geq 0$. Therefore using

Cauchy-Schwarz inequality,

$$\begin{aligned}\mathbb{E}\left[\mathrm{IV}_k\right] \leq 2N\mathbb{E}\left[\int_0^t |K(t-s)|^2 \left|\sigma_k\big(s,\Theta^k_s\big) - \sigma\big(s,\Theta^k_s\big)\right|^2 \mathrm{d}s(\mathbb{1}_{E_k} + \mathbb{1}_{E^c_k})\right] \\ + 2N\mathbb{E}\left[\int_0^t |K(t-s)|^2 \left|\sigma\big(s,\Theta^k_s\big) - \sigma(s,\overline{X}_s)\right|^2 \mathrm{d}s(\mathbb{1}_{E_k} + \mathbb{1}_{E^c_k})\right],\end{aligned}$$

where we will use the localisation to obtain convergence in the first term and Hölder continuity in the second. Let us assume for the moment that linear growth **H3a** holds. We use that Θ^k is uniformly bounded by $2N$ in E_k and linear growth for both σ_k and σ to obtain

$$\begin{aligned}&\sup_{s\in\mathbb{T}}\mathbb{E}\left[\left|\sigma_k\big(s,\Theta^k_s\big) - \sigma\big(s,\Theta^k_s\big)\right|^2\right] \\ &= \sup_{s\in\mathbb{T}}\mathbb{E}\left[\left|\sigma_k\big(s,\Theta^k_s\big) - \sigma\big(s,\Theta^k_s\big)\right|^2 \mathbb{1}_{E_k}\right] + \sup_{s\in\mathbb{T}}\mathbb{E}\left[\left|\sigma_k\big(s,\Theta^k_s\big) - \sigma\big(s,\Theta^k_s\big)\right|^2 \mathbb{1}_{E^c_k}\right] \\ &\leq \| \|\sigma_k - \sigma\|_{2N}\|^2_{\mathbb{T}} + \sup_{s\in\mathbb{T}}\mathbb{E}\left[\mathbb{1}_{E^c_k} C_L^2\big(2+2|\Theta^k_s|\big)^2\right],\end{aligned} \tag{2.6.6}$$

which tends to zero as k goes to infinity because of **H2** for the first term and because $\mathbb{P}(\Omega\setminus E_k)$ tends to zero for the second. Moreover, by **H6**, there exists $\delta>0$ such that σ is locally δ-Hölder continuous thus there exist $C_{2N}>0$ such that

$$\sup_{s\in\mathbb{T}}\mathbb{E}\left[\mathbb{1}_{E_k}\left|\sigma\big(s,\Theta^k_s\big) - \sigma(s,\overline{X}_s)\right|^2\right] \leq \sup_{s\in\mathbb{T}}\mathbb{E}\left[\mathbb{1}_{E_k} C_{2N}\left|\Theta^k_s - \overline{X}_s\right|^{2\delta}\right],$$

which tends to zero. Finally, linear growth leads to

$$\sup_{s\in\mathbb{T}}\mathbb{E}\left[\mathbb{1}_{E^c_k}\left|\sigma\big(s,\Theta^k_s\big) - \sigma(s,\overline{X}_s)\right|^2\right] \leq \mathbb{E}\left[\mathbb{1}_{E^c_k} C_L^2\big(2+|\Theta^k_s|+|\overline{X}_s|\big)^2\right], \tag{2.6.7}$$

which also tends to zero because $\mathbb{P}(\Omega\setminus E_k)$ tends to zero. If only **H3b** with Assumption 2.3.9 hold then a different bound depending on (2.2.3) would replace those in (2.6.6) and (2.6.7), by noticing that $\overline{X} = \mathcal{G}^0(0)$. In both cases the above estimates tend to zero as k tends to infinity, hence $\mathbb{E}\left[\mathrm{IV}_k\right]$ converges towards zero. Finally, $\{\mathrm{V}_k\}_{k\in\mathbb{N}}$ is uniformly bounded across $k\geq 0$ as (2.6.3) shows, thus $h^{-p}_{\varepsilon_k}\mathrm{V}_k$ tends to zero. We have proved that

$$\lim_{k\uparrow\infty}\mathbb{E}\left[\Psi_t(v^k,\eta^k)\right] = 0,$$

and this entails that the limit η^0 satisfies (2.3.18) $\mathbb{P}^0$-almost surely, for all $t\in\mathbb{T}$. Since η^0 has continuous paths, this holds for all $t\in\mathbb{T}$, $\mathbb{P}^0$-almost surely and the solution is unique therefore we conclude that $\eta^0=\psi$. Every subsequence has a subsequence for which this convergence holds therefore $\eta^{\varepsilon,v^\varepsilon}$ converges weakly towards ψ as ε goes to zero. □

3

Pathwise large deviations for Wiener chaos expansions

This chapter is based on the paper [206], accepted for publication in *Bernoulli*.

3.1 Introduction

This chapter contributes to the large body of work dedicated to limit theorems in relation with the Wiener chaos. In a probability space $(\Omega, \mathcal{F}, \mathbb{P})$, any Gaussian functional $F \in L^2(\Omega)$ can be represented via its Wiener chaos expansion

$$F = \mathbb{E}F + \sum_{n=1}^{\infty} I_n(f_n), \tag{3.1.1}$$

where, in the setup we are interested in, the terms of this (orthogonal) decomposition are multiple Wiener–Itô integrals and the kernels $(f_n)_{n\in\mathbb{N}}$ are deterministic and uniquely determined by F. We refer to the monograph [201] for more details. By exposing the inner structure of Gaussian functionals in a systematic way, this route generated a fruitful field of asymptotic methods. In particular, the Fourth Moment Theorem of Nourdin and Peccati [202], later combined with Stein's method [91, 199, 200], is an active field of research which studies the convergence of elements of a fixed Wiener chaos, i.e. multiple Wiener–Itô integrals. In the same line of research, the Breuer–Major theorem [41] was recently revisited by means of Malliavin calculus and Wiener chaos in infinite-dimensional spaces [54, 204]. These results and the related limit theorems from [11, 13, 198] fall, for the most part, into the scope of central and noncentral limit theorems on Wiener chaos, and were successfully applied to stochastic partial differential equations [81, 155, 204].

However, to the best of our knowledge, the literature on large deviations for such expansions has not evolved since the nineties. The LDP definition was given in (2.1.2). The LDP for families of a fixed Wiener chaos, even Banach space-valued, is well-known from [83, 175] and was derived again in [187] in the space of continuous functions. Around the same time, an LDP for a whole, real-valued, Wiener chaos expansion was obtained in [214] in the Brownian motion case. Our main contribution, Theorem 3.3.3, reconciles the best of both worlds: we establish an LDP in the space of continuous functions *and* for the whole Wiener chaos expansion. Moreover, we consider the case of a general white noise measure.

Among the many limit theorems on Wiener space cited above, very few are concerned with random variables on path space. The lift from the real-valued LDP of [214] to our sample-path version parallels the transition from the Gärtner-Ellis theorem for finite-dimensional LDPs to the theory of Friedlin and Wentzell [106]: both aim at studying the probability that the whole path of a stochastic process hits a given set. Furthermore, the pathwise LDP is stronger in the sense that an application of the contraction principle suffices to recover the finite-dimensional one.

The proof of [214] relied on a type of exponentially good approximation theorem, which preserves the LDP with a change of the rate function. Although these techniques are nowadays ubiquitous in sample-path large deviations for e.g. solutions of stochastic equations, applying this approximation to a Wiener chaos expansion in the space of continuous functions is challenging since it requires a good estimate of

$$\mathbb{P}\left(\sup_{t\in[0,1]} \left|I_n(f_n^t)\right| > \delta\right), \quad \text{for all } n \in \mathbb{N}.$$

This highlights the difficulty of deriving bounds for such objects, in contrast with stochastic equations where one can exploit the local martingale structure of the stochastic integrals. It also exposes the limits of the standard approach based on exponential approximations, especially in infinite-dimensional spaces.

Deviating from this approach, we will instead follow the methods presented in the monographs [46, 88] which rely upon the equivalence between the LDP (2.1.2) and the Laplace principle (1.2.6). As mentioned in the main Introduction, this route introduced by Dupuis and Ellis was dubbed the weak convergence approach because it essentially boils down to proving the convergence in distribution of a family of processes related to $(X^\varepsilon)_{\varepsilon>0}$. On top of the applications cited in the introduction of Chapter 2, it enabled to prove LDPs for infinite-dimensional Gaussian processes [32, 45, 49, 225]. Our choice of setup therefore lies at the intersection between the domains of application of the Wiener chaos expansion and the weak convergence approach. A conducive integrator is the white noise measure displayed in [201, Section 1.1.2], which, as we will see in the next section, covers a wide range of infinite-dimensional Gaussian processes. We note at this point that spatially homogeneous coloured noise and Poisson random measures also offer potential research directions.

Regarding the choice of state space, we are only limited by the tightness criteria available to us. We consider the space of continuous functions $\mathcal{C}(\mathrm{K},\mathbb{R})$, where K is a compact subset of $\mathbb{R}^d$. To the best of our knowledge, the latter condition is necessary to obtain a type of Kolmogorov continuity criterion; in this chapter we will rely upon the multidimensional version derived by Ibragimov in [156]. There is no apparent reason why the theory should not also apply to $\mathcal{C}(\mathrm{K},\mathbb{R}^N)$ for $N \geq 2$ but tracking the indices becomes cumbersome so we restrict to the one-dimensional case for clarity. Nevertheless, our state space covers infinite-dimensional processes living in $\mathcal{C}\big([0,1],\mathcal{C}([0,1],\mathbb{R})\big)$ for instance. Relatively to the classical discretisation and approximation argument, the weak convergence approach has proved efficient in deriving LDPs in such spaces.

In the framework of stochastic (partial differential) equations, pathwise uniqueness is a necessary and sometimes tricky condition, while in this chapter the difficulty lies elsewhere. Here, we must consider the convergence in distribution of a family of processes where the stochastic integration takes place with respect to a *shifted* white noise, such as $W^v(\mathrm{d}x) = W(\mathrm{d}x) + \int v(x)\,\mathrm{d}x$. Hence, after replacing $W(\mathrm{d}x)$ by $W^v(\mathrm{d}x)$ in the multiple stochastic integrals, the terms of the decomposition (3.1.1) are not independent any longer, and developing $W^v(\mathrm{d}x_1)\cdots W^v(\mathrm{d}x_n)$ generates 2^n multiple integrals. We then classify and analyse these new integrals in order to obtain the necessary estimates.

We stress that the assumptions laid down in this chapter only involve the kernels of the chaos expansion and are therefore purely deterministic. Moreover, in the context of pathwise large deviations, the chaos expansion is more agnostic than the stochastic equations paradigm, and hence this representation opens the LDP gates to uncommon continuous models. In particular, we provide applications to (multiple) Skorohod integrals and linear Skorohod equations driven by space-time white noise, and to Wick–Itô integrals driven by fractional Brownian motion (fBm). So far, LDPs for anticipating integrals have only been considered in the context of anticipating SDEs driven by a standard Brownian motion [189, 191, 192]. Related asymptotic results include central limit theorems for multiple [197] and fBm-driven [24, 149] Skorohod integrals. For completeness, we also mention the LDPs for stochastic integrals derived in [116, 118].

The outline of the chapter is the following. Section 3.2 introduces the white noise and the chaos expansion. In Section 3.3, we describe our model, assumptions, and main result. The proof is presented in Section 3.4: first we adapt sufficient conditions for the LDP of Gaussian functionals to our framework, then we lay down the foundations of the integrals under consideration, and finally we split the proof of the conditions in different lemmas. Section 3.5 gathers our applications and we make some concluding remarks in Section 3.6.

3.2 Definitions and notations

3.2.1 The white noise measure

Let us fix a probability space $(\Omega, \mathcal{F}, \mathbb{P})$, a convex set $\mathbb{T} \subseteq \mathbb{R}_+$ that includes 0, and a sigma-finite measure μ on a measurable space $(\boldsymbol{E}, \mathcal{E})$. Then we define the parameter space $\mathbf{T} := \mathbb{T} \times \boldsymbol{E}$ and its σ-algebra $\mathcal{T} := \mathcal{B}(\mathbb{T}) \otimes \mathcal{E}$. Following [201], we define a white noise measure $\dot{W}$ as an $L^2(\Omega)$-Gaussian measure on $(\mathbf{T}, \mathcal{T})$ based on the (atomless) product measure $\widehat{\mu} := \lambda \otimes \mu$, where λ is the Lebesgue measure on $\mathbb{T}$. This means $\dot{W}$ has zero mean and the following covariance structure for all $A_1, A_2 \in \mathcal{T}$,

$$\mathbb{E}\left[\dot{W}(A_1)\dot{W}(A_2)\right] = \widehat{\mu}(A_1 \cap A_2).$$

This exhibits the main feature of the white noise measure, namely that it takes independent values on any family of disjoint subsets of $\mathbf{T}$. The following examples demonstrate how one can recover well-known Gaussian processes from the white noise measure. For all $n \in \mathbb{N}$, we will henceforth write $L^2(\mathbf{T}^n)$ to denote the Hilbert space $L^2\big(\mathbf{T}^n, \widehat{\mu}^{\otimes n}\big)$ of real-valued functions.

Example 3.2.1 (Space-time white noise)**.** *In the case where $(\boldsymbol{E}, \mathcal{E}) = \big(\mathbb{R}^m, \mathcal{B}(\mathbb{R}^m)\big)$, $m \geq 1$, and μ is the Lebesgue measure on $\mathbb{R}^m$, $\dot{W}$ is called space-time white noise. When μ is not the Lebesgue measure one speaks of spatially inhomogeneous white noise [196]. A similar setup with $\boldsymbol{E} = \mathbb{T}^m$ also gives rise to the Brownian sheet [76, Example 3.10].*

Example 3.2.2 (Isonormal process)**.** *The theory of integration presented in [201] allows to define an isonormal process X over $L^2(\mathbf{T})$ as the stochastic integral*

$$X(h) := \int_{\mathbf{T}} h(t,x) W(\mathrm{d}t\,\mathrm{d}x), \quad h \in L^2(\mathbf{T}),$$

where, following standard notations, we drop the dot when integrating. This generates the covariance structure $\mathbb{E}\big[X(h)X(g)\big] = \langle h, g\rangle_{L^2(\mathbf{T})}$, for $h, g \in L^2(\mathbf{T})$.

We can recover the cylindrical Brownian motion B defined in [46, Definition 11.3] by setting $B_t(\phi) := X\big(\mathbb{1}_{[0,t]}\,\phi\big)$, for all $\phi \in L^2(\boldsymbol{E})$ and $t \geq 0$. Another variant consists in picking $\phi \in C_c^\infty(\mathbf{T})$.

Example 3.2.3 (Multidimensional Brownian motion)**.** *We reproduce Example 1.1.2 in [201]: let $\boldsymbol{E} = [\![1, N]\!]$ for some $N \in \mathbb{N}$ and μ be the uniform measure which gives mass one to each point $1, 2, \cdots, N$. Then $W^i(t) := \dot{W}([0,t] \times \{i\})$ is a standard Brownian motion for all $t \in \mathbb{T}$, $i \in \boldsymbol{E}$, with covariance $\mathbb{E}[W^i(t)W^j(s)] = (s \wedge t)\mathbb{1}_{i=j}$. For any $h \in L^2(\mathbf{T})$, one can also define its stochastic integral, which coincides with the Itô integral:*

$$W(h) = \sum_{i=1}^{N} \int_{\mathbb{T}} h^i(t)\,\mathrm{d}W^i(t).$$

Example 3.2.4 (Fractional Brownian motion)**.** *For this example, we refer to [201, Section 5.1] and set $\mathbf{T} = \mathbb{T}$. The process W defined as $W_t := \dot{W}([0,t])$ is a standard Brownian motion, and there exists a deterministic function $K_H : \mathbf{T}^2 \to \mathbf{T}$ such that $(W_t^H)_{t\in\mathbf{T}}$ defined as*

$$W_t^H := \int_0^t K_H(t,s)\,\mathrm{d}W_s \tag{3.2.1}$$

is a fractional Brownian motion with Hurst exponent $H \in (0,1)$ *and covariance function*

$$R_H(t,s) = \frac{1}{2}\big(t^{2H} + s^{2H} - |t-s|^{2H}\big).$$

3.2.2 The Wiener chaos

Let $\mathcal{F}^W := \sigma\big(\{\dot{W}(A) : A \in \mathcal{T}\}\big)$ denote the σ-field generated by $\dot{W}$—it is then clear that $\mathcal{F}^W$ coincides with the σ-fields generated by the examples presented above. From now on, we will write $L^2(\Omega)$ as short for $L^2\big(\Omega, \mathcal{F}^W, \mathbb{P}\big)$. It is well-known that any random variable $F \in L^2(\Omega)$ has the Wiener chaos expansion (3.1.1), where the convergence takes place in $L^2(\Omega)$, see for instance [201, Theorem 1.1.2]. For all $n \in \mathbb{N}$, and for the remainder of the chapter, $f_n \in L^2(\mathbf{T}^n)$ are symmetric functions and $I_n \colon L^2(\mathbf{T}^n) \to L^2(\Omega)$ is a continuous linear map that can be interpreted as a multiple integral:

$$I_n(f_n) := \int_{\mathbf{T}^n} f_n(\boldsymbol{t_n}, \boldsymbol{x_n})\, W(\mathrm{d}t_1\, \mathrm{d}x_1) \cdots W(\mathrm{d}t_n\, \mathrm{d}x_n), \tag{3.2.2}$$

with $\boldsymbol{t_n} = (t_1, \cdots, t_n)$ and $\boldsymbol{x_n} = (x_1, \cdots, x_n)$, where $(t_i, x_i) \in \mathbf{T}$ for all $i \in [\![1, n]\!]$. The nth Wiener chaos $\mathcal{H}_n$ is the subset of $L^2(\Omega)$ made of all the elements with the representation (3.2.2). The notion of integration is recalled in subsection 3.4.2, but we already highlight the important identity

$$\|I_n(f_n)\|_{L^2(\Omega)} = \sqrt{n!}\, \|f_n\|_{L^2(\mathbf{T}^n)}\,. \tag{3.2.3}$$

In the next section we specify our framework and the main large deviations result.

3.3 Statement of the main result

We aim at proving a large deviations principle for a family $(X^\varepsilon)_{\varepsilon>0}$ of processes taking values in $\mathcal{C}(\mathrm{K}, \mathbb{R})$, where $\mathrm{K} \subset \mathbb{R}^d$ is a compact subspace equipped with the Euclidean norm denoted $\|\cdot\|$, and $d \geq 1$. For all $\varepsilon > 0$ and $z \in \mathrm{K}$, we define these processes as

$$X^\varepsilon(z) := \sum_{n=0}^{\infty} \varepsilon^n I_n(f_n^{\varepsilon,z}). \tag{3.3.1}$$

For all $n \in \mathbb{N}$, $\varepsilon > 0$ and $z \in \mathrm{K}$, the kernels $f_n^{\varepsilon,z}$ are symmetric functions belonging to $L^2(\mathbf{T}^n)$, $f_0^{\varepsilon,z} = \mathbb{E}[X^\varepsilon(z)]$ and I_0 denotes the identity mapping on constants. Since the $(I_n)_{n\geq 0}$ are independent, it is clear that $X^\varepsilon(z) \in L^2(\Omega)$ for all $z \in \mathrm{K}$. We consider the following sufficient conditions, which in particular ensure the continuity of X^ε, as pointed out in Remark 3.4.12.

Assumption 3.3.1.

(A) There exist $(f_n^z)_{n\geq 0, z\in \mathrm{K}}$, where $f_n^z \in L^2(\mathbf{T}^n)$, such that $\lim_{\varepsilon\downarrow 0} f_0^{\varepsilon,z} = f_0^z$ for all $z \in \mathrm{K}$ and

$$\lim_{\varepsilon\downarrow 0} \sup_{z\in\mathrm{K}} \sup_{n\in\mathbb{N}} \|f_n^{\varepsilon,z} - f_n^z\|_{L^2(\mathbf{T}^n)} = 0.$$

(B) For all $\varepsilon > 0$ small enough, $z \in \mathrm{K}$ and $\kappa > 0$ the following holds

$$\sum_{n=0}^{\infty} \sqrt{n!}\,\kappa^n \Big(\|f_n^z\|_{L^2(\mathbf{T}^n)} + \|f_n^{\varepsilon,z}\|_{L^2(\mathbf{T}^n)} \Big) < \infty. \tag{3.3.2}$$

(C) There exists a concave modulus of continuity $\omega\colon \mathbb{R}_+ \to \mathbb{R}_+$ satisfying, for some $\alpha_0 > 0$,

$$\int_0^1 \frac{\omega(s)}{s^{1+\alpha_0}}\,\mathrm{d}s < \infty \qquad \text{and} \qquad \lim_{s\downarrow 0} \frac{\omega(s)}{s} = \infty,$$

such that, for all $\varepsilon > 0$ small enough, $y, z \in \mathrm{K}$ and $\kappa > 0$,

$$\sum_{n=0}^{\infty} \sqrt{n!}\,\kappa^n \Big(\|f_n^z - f_n^y\|_{L^2(\mathbf{T}^n)} + \|f_n^{\varepsilon,z} - f_n^{\varepsilon,y}\|_{L^2(\mathbf{T}^n)} \Big) \leq \omega(\|z-y\|). \tag{3.3.3}$$

Remark 3.3.2.

(A) The first assumption is satisfied in the case where $\mathbb{E}[X^\varepsilon(z)]$ converges as ε goes to zero and $(f_n^{\varepsilon,z})$ is independent of ε, which correspond to a purely small-noise problem. Example 3.3.6 clarifies the need (or not) for the kernels to be ε-dependent.

(B) The bound (3.3.2) holds in particular in the case where (f_n^z) and $(f_n^{\varepsilon,z})$ are of *exponential type*, as defined in [214, Equation (3.2)]: there exists $\Delta > 0$ such that for all $z \in \mathrm{K}$ and all $n \geq 0$ large enough,

$$\|f_n^z\|_{L^2(\mathbf{T}^n)} + \|f_n^{\varepsilon,z}\|_{L^2(\mathbf{T}^n)} \leq \frac{\Delta^n}{n!}. \tag{3.3.4}$$

(C) The condition (3.3.3) allows us to recover the equicontinuity necessary in Kolmogorov's continuity theorem and therefore the tightness of a family of processes related to (X^ε). The assumption can be satisfied with the modulus of continuity $\omega : s \to s^\gamma$, $\gamma \in (0,1)$, and is also related to the "slow-growth" assumption in [187], which yields the continuity of the map $z \mapsto I_n(f_n^z)$.

(D) Various examples of processes satisfying all three assumptions are given in Section 3.5.

For convenience, and despite a mild but common abuse of notations, we define the white noise process $(W_t)_{t\in\mathbb{T}}$ by $W_t(E) := \dot{W}([0,t] \times E)$ for all $E \in \mathcal{E}$, along with its filtration $\mathcal{F}_t := \sigma\big(\{W_s(E) : 0 \le s \le t, E \in \mathcal{E}\}\big)$. We think of W as a process when we need the concept of filtration but both points of view are equivalent. This allows us to introduce the set of stochastic controls

$$\mathcal{A} := \left\{ v \colon \Omega \times \mathbf{T} \to \mathbb{R}, \{\mathcal{F}_t\}\text{–progressively measurable}, \mathbb{E}\left[\int_{\mathbf{T}} v(t,x)^2\, \widehat{\mu}(\mathrm{d}t\,\mathrm{d}x)\right] < \infty \right\}, \tag{3.3.5}$$

and, for all $v \in \mathcal{A}$, $z \in \mathrm{K}$:

$$\overline{X}^v(z) := \sum_{n=0}^{\infty} J_n^v(f_n^z), \quad \text{where} \tag{3.3.6}$$

$$J_n^v(f_n^z) := \int_{\mathbf{T}^n} f_n^z(\boldsymbol{t_n}, \boldsymbol{x_n}) \big(v(t_1,x_1)\,\widehat{\mu}(\mathrm{d}t_1\,\mathrm{d}x_1)\big) \cdots \big(v(t_n,x_n)\,\widehat{\mu}(\mathrm{d}t_n\,\mathrm{d}x_n)\big). \tag{3.3.7}$$

We recall that infimum over empty spaces are equal to $+\infty$. The main result of the chapter is the following.

Theorem 3.3.3. *Under Assumption 3.3.1, the family $(X^\varepsilon)_{\varepsilon>0}$ defined in (3.3.1) satisfies an LDP in $\mathcal{C}(\mathrm{K})$ with rate function $\Lambda \colon \mathcal{C}(\mathrm{K}) \to \overline{\mathbb{R}}_+$ given by*

$$\Lambda(\psi) := \inf\left\{ \frac{1}{2}\|v\|^2_{L^2(\mathbf{T})} : v \in L^2(\mathbf{T}),\, \psi = \overline{X}^v \right\}. \tag{3.3.8}$$

Remark 3.3.4. The form of the rate function (3.3.8) is reminiscent of the rate function of the finite-dimensional LDP proved by Pérez–Abreu and Tudor in [214, Theorem 3.1].

In fact, for a fixed $z \in \mathrm{K}$, Assumption 3.3.1 (A) and (B) alone are sufficient to derive an LDP for $(X^\varepsilon(z))_{\varepsilon>0}$ on $\mathbb{R}$, with rate function $\Lambda_R \colon \mathbb{R} \to \overline{\mathbb{R}}_+$ given by

$$\Lambda_R(r) := \inf\left\{ \frac{1}{2}\|v\|^2_{L^2(\mathbf{T})} : v \in L^2(\mathbf{T}),\, r = \overline{X}^v(z) \right\}.$$

Remark 3.3.5. A variety of Gaussian processes can be expressed as $g(W)$ where g is a linear and measurable map. As examples, we already mentioned the fractional Brownian motion (Example 3.2.4) and will soon introduce a infinite collection of Brownian motions $\beta = (\beta_i)_{i\in\mathbb{N}}$ (subsection 3.4.1). Furthermore, β is itself connected through a measurable map to other infinite-dimensional Gaussian processes such as Q-Wiener processes and cylindrical Brownian motions [46, Section 11.1]. Hence the process $\mathcal{G}(\varepsilon B)$, where B is one of the aforementioned Gaussian processes, can be written as $(\mathcal{G}\circ g)(\varepsilon W)$ and be covered by our framework.

The following example demonstrates the interest of the ε-dependence in the kernels.

Example 3.3.6. *We take as an example the Brownian motion case with $\mathrm{K} = \mathbf{T} = [0,1]$. For any $\varepsilon > 0$, $x \in \mathbb{R}$ and $t \in \mathrm{K}$, we consider the following small-noise SDE*

$$X^\varepsilon(t) = x + \int_0^t b(X^\varepsilon(s))\,\mathrm{d}s + \varepsilon\int_0^t \sigma(X^\varepsilon(s))\,\mathrm{d}W_s, \tag{3.3.9}$$

where b and σ are real-valued functions such that (3.3.9) has a unique strong solution. Therefore, there exists a measurable map $\mathcal{G}^\varepsilon\colon \mathcal{C}[0,1]\to\mathcal{C}[0,1]$ such that $\mathcal{G}^\varepsilon(\varepsilon W) := X^\varepsilon$. This yields an expansion for X^ε in the form of (3.3.1), where ε^n arises from εW. We notice that, for $\varepsilon_1 \neq \varepsilon_2$, the distributions of X^{ε_1} and X^{ε_2} are mutually singular, which means the map $\mathcal{G}^\varepsilon$ can depend explicitly on ε, hence so do the kernels $(f_n^{\varepsilon,t})$.

3.4 Proof of the main result

The proof of Theorem 3.3.3 consists of three parts. In the first one, we adapt the sufficient conditions for an LDP of Gaussian functionals, originally derived in [45], to our setup. In the second, we give meaning to some multiple stochastic integrals arising from the first part and finally, we verify said conditions. The game is afoot.

3.4.1 The weak convergence approach

As mentioned in the introduction, we rely upon the weak convergence approach to large deviations which is conducive for proving LDPs of functionals of infinite-dimensional Brownian motions [45, 46]. In order to deduce abstract sufficient conditions for the LDP as in [46, Theorem 11.13], we need to cast our white noise process in this framework. This entails first expressing $W = (W_t(E))_{t\in\mathbb{T},E\in\mathcal{E}}$ as a functional of an infinite sequence of standard Brownian motions. Building on this connection, we then recover the variational representation formula for functionals of white noise, analogue of [46, Theorem 11.11] as cornerstone of the proof.

We start by observing that $\dot{W}$ (and W) is a random variable taking values in the space of signed measures on $\mathbf{T}$, which we denote by $\mathcal{M}(\mathbf{T})$. Now, consider the Hilbert space $L^2(\boldsymbol{E}) := \{f\colon \boldsymbol{E} \to \mathbb{R} : \int_{\boldsymbol{E}} f(x)^2\mu(\mathrm{d}x) < \infty\}$ equipped with the usual inner product, and let $(\phi_i)_{i\in\mathbb{N}}$ be a complete orthonormal sequence. Let $\beta := (\beta_i)_{i\in\mathbb{N}}$ defined for all $t \in \mathbb{T}$ by $\beta_i(t) := \int_{[0,t]\times\boldsymbol{E}} \phi_i(x)W(\mathrm{d}s\,\mathrm{d}x)$, and note that it is a sequence of independent standard Brownian motions. Furthermore, for all $E \in \mathcal{E}$, we have

$$W_t(E) = \sum_{i=1}^{\infty} \beta_i(t) \int_E \phi_i(x)\mu(\mathrm{d}x),$$

where the series converges in $L^2(\Omega)$. The sequence β can be regarded as a random variable with values in $\mathcal{C}([0,T],\mathbb{R}^\infty)$, where $\mathbb{R}^\infty$ denotes the product space of countably infinite copies of $\mathbb{R}$. Therefore the σ-fields generated by β and W (and $\dot{W}$) coincide and there exists a measurable function $g : \mathcal{C}(\mathbb{T},\mathbb{R}^\infty) \to \mathcal{M}(\mathbf{T})$ such that $\dot{W} = g(\beta)$ almost surely.

We introduce the following spaces of deterministic and stochastic controls, and recall that $\mathcal{A}$ was defined by (3.3.5)

$$\begin{aligned}
\mathcal{S}_M &:= \left\{v\colon \mathbf{T} \to \mathbb{R} : \int_{\mathbf{T}} v(t,x)^2\,\widehat{\mu}(\mathrm{d}t\,\mathrm{d}x) \le M\right\},\\
\mathcal{A}_M &:= \left\{v \in \mathcal{A} : v \in \mathcal{S}_M,\ \mathbb{P}\text{−almost surely}\right\}, \qquad \text{and} \qquad \mathcal{A}_b := \bigcup_{M\in\mathbb{N}} \mathcal{A}_M,
\end{aligned}$$

where $\mathcal{S}_M$ is endowed with the weak topology on $L^2(\mathbf{T})$, not to be confused with the weak convergence of measures. For all $v \in \mathcal{A}_b$, let us define $W^v = \left(W_t^v(E)\right)_{t\in\mathbb{T},E\in\mathcal{E}}$ by

$$W_t^v(E) := W_t(E) + \int_{[0,t]\times E} v(s,x)\,\widehat{\mu}(\mathrm{d}s\,\mathrm{d}x). \tag{3.4.1}$$

We can now state the corresponding variational representation formula.

Lemma 3.4.1. *For any bounded measurable map $G\colon \mathcal{M}(\mathbf{T}) \to \mathbb{R}$, the following holds:*

$$-\log\mathbb{E}\left[\mathrm{e}^{-G(W)}\right] = \inf_{v\in\mathcal{A}_b} \mathbb{E}\left[\frac{1}{2}\int_{\mathbf{T}} v(t,x)^2\,\widehat{\mu}(\mathrm{d}t\,\mathrm{d}x) + G(W^v)\right]. \tag{3.4.2}$$

Proof. The proof, based on the corresponding formula for β and its relationship with W, is identical to the proof of [46, Theorem 11.11] after replacing the Brownian sheet with the present white noise. $\square$

From this lemma arise abstract sufficient conditions yielding an LDP for functionals of white noise. The functionals take the form $\mathcal{G}^\varepsilon(\varepsilon W)$, where $\mathcal{G}^\varepsilon \colon \mathcal{M}(\mathbf{T}) \to \mathcal{X}$ is a measurable map and $\mathcal{X}$ is a Polish space ($\mathcal{X} = \mathcal{C}(\mathrm{K})$ in the case we are interested in). With the tools that we just described, one can translate [46, Condition 11.12 and Theorem 11.13] into the white noise framework, as follows.

Assumption 3.4.2. There exists a measurable map $\mathcal{G}^0 \colon L^2(\mathbf{T}) \to \mathcal{X}$ such that the following hold.

(a) Consider $M > 0$ and a family $(v^\varepsilon)_{\varepsilon>0} \subset \mathcal{A}_M$ such that v^ε converges in distribution to v as ε goes to zero. Then the following convergence also holds in distribution

$$\lim_{\varepsilon\downarrow 0} \mathcal{G}^\varepsilon\left(\varepsilon W^{v^\varepsilon/\varepsilon}\right) = \mathcal{G}^0(v).$$

(b) For every $M > 0$, the following sets are compact subsets of $\mathcal{X}$:

$$\Gamma_M := \left(\mathcal{G}^0(v) : v \in \mathcal{S}_M\right).$$

Remark 3.4.3. Since $\mathcal{S}_M$ is compact for every $M > 0$ in the weak topology, Assumption 3.4.2(b) is satisfied as soon as $\mathcal{G}^0$ is continuous.

Theorem 3.4.4. *Let Assumption 3.4.2 hold. Then* $\left(\mathcal{G}^\varepsilon\left(\varepsilon W\right)\right)_{\varepsilon>0}$ *satisfies an LDP with rate function* $\Lambda \colon \mathcal{X} \to \overline{\mathbb{R}}_+$ *given by*

$$\Lambda(\psi) = \inf\left\{\frac{1}{2}\|v\|^2_{L^2(\mathbf{T})} : v \in L^2(\mathbf{T}),\, \psi = \mathcal{G}^0(v)\right\}.$$

3.4.2 Perturbing the chaos expansion

We observe that, for each $\varepsilon > 0$, there exist a measurable map $\mathcal{G}^\varepsilon \colon \mathcal{M}(\mathbf{T}) \to \mathcal{C}(\mathrm{K})$ such that, for all $z \in \mathrm{K}$,

$$\mathcal{G}^\varepsilon(W)(z) = \sum_{n=0}^\infty I_n(f_n^{\varepsilon,z}),$$

where the kernels come from (3.3.1), and therefore we have $X^{\varepsilon} = \mathcal{G}^{\varepsilon}(\varepsilon W)$ by linearity of I_n. Our objective is to apply the abstract Theorem 3.4.4 to the map $\mathcal{G}^{\varepsilon}$ defined above, hence we must study the controlled processes $X^{\varepsilon,v} := \mathcal{G}^{\varepsilon}\left(\varepsilon W^{v/\varepsilon}\right)$, for $v \in \mathcal{A}_b$. Perturbing the chaos expansion leads us to revisit some notions of integration with respect to white noise.

Let $n \in \mathbb{N}$. We recall how the multiple integral operator I_n is defined, following [201, Section 1.1.2]. Let us call $f \in L^2(\mathbf{T}^n)$ an elementary function if it has the form

$$f(\boldsymbol{t_n}, \boldsymbol{x_n}) = \sum_{i_1,\cdots,i_n=1}^{N} a_{i_1\cdots i_n} \mathbb{1}_{A_{i_1}\times\cdots\times A_{i_n}}(\boldsymbol{t_n}, \boldsymbol{x_n}), \tag{3.4.3}$$

where $N \geq n$, $A_1, \cdots, A_N \in \mathcal{T}$ are pairwise disjoint sets with finite measure, and the real coefficients $a_{i_1,\cdots,i_n}$ are zero if and only if two of the indices $i_1, \cdots, i_n$ are equal. For such a function, the integral is defined as

$$I_n(f) = \sum_{i_1,\cdots,i_n=1}^{N} a_{i_1\cdots i_n} \dot{W}(A_{i_1}) \times \cdots \times \dot{W}(A_{i_n}).$$

Let $v \in \mathcal{A}_b$, and let the shifted integral read

$$I_n^v(f) := \sum_{i_1,\cdots,i_n=1}^{N} a_{i_1\cdots i_n} \dot{W}^v(A_{i_1}) \times \cdots \times \dot{W}^v(A_{i_n}), \tag{3.4.4}$$

where, similarly to (3.4.1), $\dot{W}^v(A) := \dot{W}(A) + \int_A v(t,x)\, \widehat{\mu}(\mathrm{d}t\, \mathrm{d}x)$ for all $A \in \mathcal{T}$, and such that $\dot{W}^v$ is also $\mathcal{M}(\mathbf{T})$-valued.

Let us define the random measure $\nu\colon A \to \int_A v(t,x)\, \widehat{\mu}(\mathrm{d}t\, \mathrm{d}x)$, for $A \in \mathcal{T}$. In order to study (3.4.4) and extend it to the whole space $L^2(\mathbf{T}^n)$, we develop $\dot{W}^v(A_{i_1}) \times \cdots \times \dot{W}^v(A_{i_n})$ in full, which is suggestive of the binomial formula for $(x+y)^n$. We obtain the sum of 2^n terms, each of them being the product of all the elements of an n-tuple. We can regard these n-tuples as words of length n over an alphabet made of two distinct elements: the measures ν and $\dot{W}$. Moreover the order of these letters, encoded by the sets $(A_{i_j})_{j=1}^n$, matters. Nevertheless, the 2^n n-tuples can be classified in the following way. For each $k \in [\![0, n]\!]$ there are $\binom{n}{k}$ terms made of exactly k times ν and $n-k$ times $\dot{W}$; therefore let $\Theta(k,n)$ be the set of such n-tuples. An element θ of $\Theta(k,n)$ is thus a sequence made of ν and $\dot{W}$ and we can write, for a fixed sequence $(i_1, \cdots, i_n)$:

$$\prod_{j=1}^{n} \dot{W}^v(A_{i_j}) = \sum_{k=0}^{n} \sum_{\theta \in \Theta(k,n)} \prod_{j=1}^{n} \theta_j(A_{i_j}).$$

For an elementary function f of the form (3.4.3) and a given n-tuple $\theta \in \Theta(k, n)$, we define

$$m_\theta(f) := \sum_{i_1, \cdots, i_n = 1}^{N} a_{i_1 \cdots i_n} \prod_{j=1}^{n} \theta_j(A_{i_j}).$$

It is well-known that I_n can be extended to a linear map from $L^2(\mathbf{T}^n)$ to $L^2(\Omega)$. The purpose of the following lemma is to extend the maps m_θ and I_n^v in a similar way and to identify this extension as a multiple integral.

Lemma 3.4.5. *For all $p \geq 2$, $M > 0$, $v \in \mathcal{A}_M$ and $n \geq 0$, the following hold.*

1. *For all $k \in [\![0, n]\!]$ and $\theta \in \Theta(k, n)$, the map m_θ can be extended to a linear and continuous map from $L^2(\mathbf{T}^n)$ to $L^p(\Omega)$ with*

$$\|m_\theta(f)\|_{L^p(\Omega)} \leq \sqrt{(n-k)!}\, M^{k/2} (p-1)^{\frac{n-k}{2}} \|f\|_{L^2(\mathbf{T}^n)} . \tag{3.4.5}$$

2. *The same extension holds for I_n^v with the bound*

$$\|I_n^v(f)\|_{L^p(\Omega)} \leq \sqrt{n!} \Big(4(M+1)(p-1)\Big)^{n/2} \|f\|_{L^2(\mathbf{T}^n)} . \tag{3.4.6}$$

3. *If $\theta \in \Theta(n, n)$ and $f \in L^2(\mathbf{T}^n)$, then $m_\theta(f)$ is also a multiple stochastic integral over $\mathbf{T}^n$ where integration against ν takes place in the Lebesgue sense, almost surely. Furthermore, we have $m_\theta(f) = J_n^v(f)$, almost surely, where J_n^v is defined in (3.3.7).*

Remark 3.4.6. The bound (3.4.5) does not depend on θ but only on k and n.

Remark 3.4.7. Like I_n in (3.2.2), we can interpret m_θ as an $L^2(\Omega)$-multiple integral in the following sense:

$$m_\theta(f) = \int_{\mathbf{T}^n} f(\boldsymbol{t_n}, \boldsymbol{x_n})\, \nu(\mathrm{d}t_1\, \mathrm{d}x_1) \cdots \nu(\mathrm{d}t_k\, \mathrm{d}x_k) W(\mathrm{d}t_{k+1}\, \mathrm{d}x_{k+1}) \cdots W(\mathrm{d}t_n\, \mathrm{d}x_n). \tag{3.4.7}$$

Likewise, we can interpret I_n^v as an $L^2(\Omega)$-multiple integral in the following sense:

$$I_n^v(f) = \int_{\mathbf{T}^n} f_n(\boldsymbol{t_n}, \boldsymbol{x_n})\, W^v(\mathrm{d}t_1\, \mathrm{d}x_1) \cdots W^v(\mathrm{d}t_n\, \mathrm{d}x_n).$$

Proof. Let us fix $p \geq 2$, $M > 0$, $v \in \mathcal{A}_M$ and $n \geq 0$.

1. Let $k \in [\![0, n]\!]$ and $\theta \in \Theta(k, n)$. The map I_n is linear for elementary functions and therefore so is m_θ since they are designed in an identical way.

   Notice that for all $A \in \mathcal{T}$ with finite measure, one has $\nu(A) = \int_{\mathbf{T}} \mathbb{1}_A \, \nu(\mathrm{d}t \, \mathrm{d}x) = \int_{\mathbf{T}} \mathbb{1}_A \, v(t,x) \, \widehat{\mu}(\mathrm{d}t \, \mathrm{d}x)$. Without loss of generality, and for more clarity, assume $\theta_j = \nu$ for all $l \in [\![1, k]\!]$ and $\theta_j = \dot{W}$ for all $l \in [\![k+1, n]\!]$. Otherwise, just split the i_j's in a different way in the computation below. Let f be an elementary function in $L^2(\mathbf{T}^n)$, then we have:

$$
\begin{aligned}
m_\theta(f) &= \sum_{i_{k+1},\cdots,i_n=1}^{N} \sum_{i_1,\cdots,i_k=1}^{N} a_{i_1\cdots i_n} \left(\prod_{j=1}^{k} \int_{\mathbf{T}} \mathbb{1}_{A_{i_j}}(t,x) \nu(\mathrm{d}t \, \mathrm{d}x) \right) \left(\prod_{j=k+1}^{n} \dot{W}(A_{i_j}) \right) \\
&= \sum_{i_{k+1},\cdots,i_n=1}^{N} \left(\int_{\mathbf{T}^k} \sum_{i_1,\cdots,i_k=1}^{N} a_{i_1\cdots i_n} \left(\prod_{j=1}^{k} \mathbb{1}_{A_{i_j}}(t_j,x_j) \right) \nu^{\otimes k}(\mathrm{d}\boldsymbol{t_k} \mathrm{d}\boldsymbol{x_k}) \right) \left(\prod_{j=k+1}^{n} \dot{W}(A_{i_j}) \right) \\
&\leq \sum_{i_{k+1},\cdots,i_n=1}^{N} M^{k/2} \left(\int_{\mathbf{T}^k} \left(\sum_{i_1,\cdots,i_k=1}^{N} a_{i_1\cdots i_n} \prod_{j=1}^{k} \mathbb{1}_{A_{i_j}}(t_j,x_j) \right)^2 \widehat{\mu}^{\otimes k}(\mathrm{d}\boldsymbol{t_k} \mathrm{d}\boldsymbol{x_k}) \right)^{\frac{1}{2}} \left(\prod_{j=k+1}^{n} \dot{W}(A_{i_j}) \right),
\end{aligned}
$$

   almost surely, where we used the linearity of the integral, Cauchy–Schwarz inequality k times and the fact that $v \in \mathcal{A}_M$, which yields an almost sure bound for its $L^2(\mathbf{T})$-norm. Now we observe that, since the A_{i_j}'s are disjoint, the cross terms vanish

$$
\begin{aligned}
\Psi &:= \int_{\mathbf{T}^k} \left(\sum_{i_1,\cdots,i_k=1}^{N} a_{i_1\cdots i_n} \prod_{j=1}^{k} \mathbb{1}_{A_{i_j}}(t_j,x_j) \right)^2 \widehat{\mu}^{\otimes k}(\mathrm{d}\boldsymbol{x_k} \mathrm{d}\boldsymbol{t_k}) \\
&= \int_{\mathbf{T}^k} \sum_{i_1,\cdots,i_k=1}^{N} (a_{i_1\cdots i_n})^2 \left(\prod_{j=1}^{k} \mathbb{1}_{A_{i_j}}(t_j,x_j) \right) \widehat{\mu}^{\otimes k}(\mathrm{d}\boldsymbol{x_k} \mathrm{d}\boldsymbol{t_k}) \\
&= \sum_{i_1,\cdots,i_k=1}^{N} (a_{i_1\cdots i_n})^2 \prod_{j=1}^{k} \widehat{\mu}(A_{i_j}).
\end{aligned}
$$

   We observe that, for any fixed $(i_1, \cdots, i_k)$,

$$
h := \sum_{i_{k+1},\cdots,i_n=1}^{N} \sqrt{\Psi} \prod_{j=k+1}^{n} \mathbb{1}_{A_{i_j}} \tag{3.4.8}
$$

   is an elementary function in $L^2(\mathbf{T}^{n-k})$, and therefore $m_\theta(f) \leq M^{k/2} I_{n-k}(h)$, almost surely. We then recover an element of the $(n-k)$th Wiener chaos and appeal to the hypercontractivity property [201, Theorem 1.4.1]. It implies that $L^p(\Omega)$ norms are equivalent in a given Wiener chaos for all $p \geq 2$ and, in particular, that for all $p \geq 2$

$$\|I_{n-k}(h)\|_{L^p(\Omega)} \leq (p-1)^{\frac{n-k}{2}} \|I_{n-k}(h)\|_{L^2(\Omega)} = (p-1)^{\frac{n-k}{2}} \sqrt{(n-k)!}\, \|h\|_{L^2(\mathbf{T}^{n-k})} ,$$

where we also used (3.2.3). Once more, the cross terms vanish to give

$$\|h\|^2_{L^2(\mathbf{T}^{n-k})} = \sum_{i_1,\cdots,i_n=1}^{N} (a_{i_1\cdots i_n})^2 \prod_{j=1}^{n} \widehat{\mu}(A_{i_j}) = \|f\|^2_{L^2(\mathbf{T}^n)} ,$$

which yields the desired bound and thus the continuity of m_θ for elementary functions. Moreover, the space of elementary functions is dense in $L^2(\mathbf{T}^n)$ and therefore m_θ and its bound extend as claimed [201, Section 1.1.2].

2. Let f be an elementary function. This is a corollary of the previous item which follows by noticing that

$$I_n^v(f) = \sum_{k=0}^{n} \sum_{\theta\in\Theta(k,n)} m_\theta(f).$$

Then, bounding k and $n-k$ by n and and using the identity $\sum_{k=0}^{n} \binom{n}{k} = 2^n$, this yields

$$\|I_n^v(f)\|_{L^p(\Omega)} \leq \sum_{k=0}^{n} \binom{n}{k} \sqrt{(n-k)!}\, M^{k/2} (p-1)^{\frac{n-k}{2}} \|f\|_{L^2(\mathbf{T}^n)} \leq \sqrt{n!} \Big(4(M+1)(p-1)\Big)^{n/2} \|f\|_{L^2(\mathbf{T}^n)} ,$$

as desired.

3. Let $f \in L^2(\mathbf{T}^n)$ and a sequence of elementary functions $(f^l)_{l\in\mathbb{N}}$ tending to f in $L^2(\mathbf{T}^n)$. Let $\theta \in \Theta(n,n)$, from (3.4.5), we know that $m_\theta(f-f^l)$ converges to zero in $L^2(\Omega)$. Moreover, for every subsequence of $(m_\theta(f-f^l))_{l\in\mathbb{N}}$, there exists a further subsubsequence that tends to zero almost surely. Therefore, the original sequence converges to zero almost surely, which implies that integration against ν takes place in the Lebesgue sense, almost surely.

Therefore, it is clear that, almost surely, we have

$$\begin{aligned} m_\theta(f) &= \int_{\mathbf{T}^n} f(\boldsymbol{t_n}, \boldsymbol{x_n})\, \nu(\mathrm{d}t_1\, \mathrm{d}x_1) \cdots \nu(\mathrm{d}t_n\, \mathrm{d}x_n) \\ &= \int_{\mathbf{T}^n} f(\boldsymbol{t_n}, \boldsymbol{x_n}) \left(v(t_1, x_1)\, \widehat{\mu}(\mathrm{d}t_1\, \mathrm{d}x_1)\right) \cdots \left(v(t_n, x_n)\, \widehat{\mu}(\mathrm{d}t_n\, \mathrm{d}x_n)\right) = J_n^v(f). \square \end{aligned}$$

Remark 3.4.8. We observe that h, defined in (3.4.8), is equal to $\|f(\boldsymbol{t_{n-k}}, \boldsymbol{x_{n-k}}, \bullet)\|_{L^2(\mathbf{T}^k)}$.

Remark 3.4.9. It is natural to wonder whether representing m_θ in the sense of Walsh [76, 233] would yield similar bounds. Because of the necessity of adaptedness, we would have to write (3.4.7) as

$$m_\theta(f_n) = n! \int_{\mathbf{T}} \int_{[0,t_n]\times \boldsymbol{E}} \cdots \int_{[0,t_2]\times \boldsymbol{E}} f(\boldsymbol{t_n}, \boldsymbol{x_n})\, \nu(\mathrm{d}t_1\, \mathrm{d}x_1) \cdots \nu(\mathrm{d}t_k\, \mathrm{d}x_k) W(\mathrm{d}t_{k+1}\, \mathrm{d}x_{k+1}) \cdots W(\mathrm{d}t_n\, \mathrm{d}x_n),$$

with $\theta \in \Theta(k, n)$ and $f \in L^2(\mathbf{T}^n)$. Applying Cauchy–Schwarz inequality k times yields

$$\begin{aligned} \mathbb{E}\left[m_\theta(f)^2\right] &\leq (n!)^2 M^k \mathbb{E}\left[\left(\int_{\mathbf{T}} \int_{[0,t_n]\times \boldsymbol{E}} \cdots \int_{[0,t_{k+1}]\times \boldsymbol{E}} \|f\|_{L^2(\mathbf{T}^{n-k})}\, W(\mathrm{d}t_{k+1}\, \mathrm{d}x_{k+1}) \cdots W(\mathrm{d}t_n\, \mathrm{d}x_n)\right)^2\right] \\ &= \left(\frac{n!}{(n-k)!}\right)^2 M^k \|f\|_{L^2(\mathbf{T}^n)}, \end{aligned}$$

which is not as sharp as (3.4.5).

3.4.3 Proof of convergence

Now that we have set the stage, the game consists in verifying that Assumption 3.4.2 holds and then applying Theorem 3.4.4. In this subsection we will then study, for all $z \in \mathrm{K}$,

$$X^{\varepsilon,v}(z) = \sum_{n=0}^{\infty} I_n^{\varepsilon,v}(f_n^{\varepsilon,z}),$$

where $I_0^{\varepsilon,v} = I_0$ and for all $n \in \mathbb{N}$, $f_n \in L^2(\mathbf{T}^n)$,

$$I_n^{\varepsilon,v}(f_n) = \int_{\mathbf{T}^n} f_n(\boldsymbol{t_n}, \boldsymbol{x_n})\, \varepsilon W^{v/\varepsilon}(\mathrm{d}t_1\, \mathrm{d}x_1) \cdots \varepsilon W^{v/\varepsilon}(\mathrm{d}t_n\, \mathrm{d}x_n), \tag{3.4.9}$$

which meaning flows from Lemma 3.4.5 and Remark 3.4.7. Note that in this subsection, the n-tuples θ are sequences made of the measures $\varepsilon\dot{W}$ (instead of $\dot{W}$) and ν, but we retain the same notations for simplicity. Hence, for $f \in L^2(\mathbf{T}^n)$, $\theta \in \Theta(k, n)$ and $v \in$

$\mathcal{A}_M$, the linearity of m_θ entails that the moment bound (3.4.5) becomes

$$\|m_\theta(f)\|_{L^p(\Omega)} \leq \varepsilon^{n-k}\sqrt{(n-k)!}\, M^{k/2}(p-1)^{\frac{n-k}{2}}\, \|f\|_{L^2(\mathbf{T}^n)}\,. \tag{3.4.10}$$

We start by deriving a moment bound for $X^{\varepsilon,v}(z)$ in Lemma 3.4.10, then the tightness of $(X^{\varepsilon,v^\varepsilon})_{\varepsilon>0}$ in $\mathcal{C}(\mathrm{K})$ in Lemma 3.4.11, which, combined with the convergence of the marginals proved in Lemma 3.4.13, yields the weak convergence in $\mathcal{C}(\mathrm{K})$. Finally we show the compactness of the sets Γ_M, for all $M>0$, in Lemma 3.4.14.

Lemma 3.4.10 (Moment bound)**.** *Let Assumption 3.3.1(B) hold. For all $p \geq 2$, $v \in \mathcal{A}_M$, $M>0$ and $\varepsilon>0$ small enough, we have $\sup_{z\in\mathrm{K}} \|X^{\varepsilon,v}(z)\|_{L^p(\Omega)} < \infty$.*

Proof. Let us fix $\varepsilon>0$, $v\in\mathcal{A}_M$, $M>0$ and $z\in\mathrm{K}$, and recall the sequence of kernels $(f_n^{\varepsilon,z})_{n\geq 0}$ from (3.3.1). As in the proof of Lemma 3.4.5, the representation

$$I_n^{\varepsilon,v}(f_n^{\varepsilon,z}) = \sum_{k=0}^{n} \sum_{\theta\in\Theta(k,n)} m_\theta(f_n^{\varepsilon,z}),$$

combined with (3.4.10), yields

$$\begin{aligned}
\|I_n^{\varepsilon,v}(f_n^{\varepsilon,z})\|_{L^p(\Omega)} &\leq \sum_{k=0}^{n} \sum_{\theta\in\Theta(k,n)} \|m_\theta(f_n^{\varepsilon,z})\|_{L^p(\Omega)} \\
&\leq \sum_{k=0}^{n} \binom{n}{k} \varepsilon^{n-k}\sqrt{(n-k)!}\, M^{k/2}(p-1)^{\frac{n-k}{2}}\, \|f_n^{\varepsilon,z}\|_{L^2(\mathbf{T}^n)}\,.
\end{aligned}$$

Hence, bounding k and $n-k$ by n, ε by one and using the identity $\sum_{k=0}^n \binom{n}{k} = 2^n$, we conclude that

$$\|X^{\varepsilon,v}(z)\|_{L^p(\Omega)} \leq \sum_{n=0}^{\infty} \|I_n^{\varepsilon,v}(f_n^{\varepsilon,z})\|_{L^p(\Omega)} \leq \sum_{n=0}^{\infty} \sqrt{n!}\Big(4(M+1)(p-1)\Big)^{n/2} \|f_n^{\varepsilon,z}\|_{L^2(\mathbf{T}^n)}\,,$$

which is finite by Assumption 3.3.1(B). □

Lemma 3.4.11 (Tightness)**.** *Assume that Assumption 3.3.1 (B) and (C) hold, let $M>0$ and a family $(v^\varepsilon)_{\varepsilon>0}\subset\mathcal{A}_M$, then the family $(X^{\varepsilon,v^\varepsilon})_{\varepsilon>0}$ is tight in $\mathcal{C}(\mathrm{K})$.*

Proof. Let $y,z\in\mathrm{K}$ and $\varepsilon>0$. Using the same computations as in the proof of

Lemma 3.4.10 and Assumption 3.3.1(C), one obtains that, for all $p > 2$,

$$\left\|X^{\varepsilon,v^\varepsilon}(z) - X^{\varepsilon,v^\varepsilon}(y)\right\|_{L^p(\Omega)} \leq \sum_{n=0}^{\infty} \sqrt{n!}\,\Big(4(M{+}1)(p{-}1)\Big)^{n/2} \|f_n^{\varepsilon,z} - f_n^{\varepsilon,y}\|_{L^2(\mathbf{T}^n)} \leq \omega\big(\,\|z-y\|\,\big).$$

We appeal to [156, Theorem 6], a multidimensional version of Kolmogorov's continuity theorem, which then states that for all $\alpha > 0$ and some constant $C > 0$,

$$\mathbb{E}\left[\sup_{\|z-y\|\leq 1} \frac{\left|X^{\varepsilon,v^\varepsilon}(z) - X^{\varepsilon,v^\varepsilon}(y)\right|}{\|z-y\|^\alpha}\right] \leq C \int_0^1 \frac{\omega(u)}{u^{1+\alpha+d/p}}\,\mathrm{d}u.$$

The latter is finite by choosing α and p such that $\alpha + d/p < \alpha_0$, where we recall that $\alpha_0 > 0$ comes from Assumption 3.3.1(C) and d is such that $\mathrm{K} \subset \mathbb{R}^d$. This is possible because we can choose $\alpha > 0$ as small and p as large as we want, for instance $\alpha < \alpha_0$ and $p = 1 + d/(\alpha_0 - \alpha)$. Combining this with Lemma 3.4.10, Aldous theorem [35, Theorem 7.3] yields the tightness of $(X^{\varepsilon,v^\varepsilon})_{\varepsilon>0}$ in $\mathcal{C}(\mathrm{K})$. $\square$

Remark 3.4.12. Suppose there exists a family of kernels $(f_n^z)_{n\geq 0, z\in \mathrm{K}}$ such that $f_n^z \in L^2(\mathbf{T}^n)$ for all $n \geq 0$, and which satisfy conditions (3.3.2) and (3.3.3). Define the real-valued process Z via the Wiener chaos expansion

$$Z(z) := \sum_{n=0}^{\infty} I_n(f_n^z), \quad z \in \mathrm{K}.$$

Then the version of the Kolmogorov's continuity theorem used above entails that Z admits a version which is Hölder continuous of any order $\alpha < \alpha_0$.

Lemma 3.4.13 (Convergence)**.** *Let Assumption 3.3.1 (A) and (B) hold. Consider $M > 0$ and a family $(v^\varepsilon)_{\varepsilon>0} \subset \mathcal{A}_M$ such that v^ε converges in distribution to v as ε goes to zero. Then, for all $z \in \mathrm{K}$, $X^{\varepsilon,v^\varepsilon}(z)$ converges in distribution to $\overline{X}^v(z)$, defined in (3.3.6), as ε goes to zero.*

Proof. We start by recalling, as Lemma 3.4.5(3) shows, that the multiple integral J_n^v defined in (3.3.7) arises from the unique n-tuple belonging to $\Theta(n,n)$, made exclusively of ν. On the other hand, all the n-tuples containing one or more $\varepsilon\dot{W}$ will converge to zero.

Let us fix $z \in \mathrm{K}$. We decompose the quantity of interest in the following way:

$$X^{\varepsilon,v^\varepsilon}(z) - \overline{X}^v(z) = \boldsymbol{S}_1^\varepsilon + \boldsymbol{S}_2^\varepsilon,$$

where

$$\boldsymbol{S}_1^\varepsilon := \sum_{n=0}^{\infty}\sum_{k=0}^{n-1}\sum_{\theta\in\Theta(k,n)} m_\theta(f_n^{\varepsilon,z}) \qquad \text{and} \qquad \boldsymbol{S}_2^\varepsilon := \sum_{n=0}^{\infty}\sum_{\theta\in\Theta(n,n)} m_\theta(f_n^{\varepsilon,z}) - \overline{X}^v(z).$$

Similarly to the proof of Lemma 3.4.10, the bound (3.4.10) and the identity $\sum_{k=0}^n \binom{n}{k} = 2^n$ yield

$$\begin{aligned}\|\boldsymbol{S}_1^\varepsilon\|_{L^2(\Omega)} \le \sum_{n=0}^{\infty}\sum_{k=0}^{n-1}\sum_{\theta\in\Theta(k,n)} \|m_\theta(f_n^{\varepsilon,z})\|_{L^2(\Omega)} &\le \sum_{n=0}^{\infty}\sum_{k=0}^{n-1}\binom{n}{k} M^{k/2}\varepsilon^{n-k}\sqrt{(n-k)!}\,\|f_n^{\varepsilon,z}\|_{L^2(\mathbf{T}^n)}\\ &\le \varepsilon\sum_{n=0}^{\infty}\sqrt{n!}\,\big(4(M+1)\big)^{n/2}\,\|f_n^{\varepsilon,z}\|_{L^2(\mathbf{T}^n)},\end{aligned}$$

for all $\varepsilon < 1$. Assumption 3.3.1(B) entails that the series is finite and therefore $\|\boldsymbol{S}_1^\varepsilon\|_{L^2(\Omega)}$ tends to zero as ε goes to zero.

For all $n \in \mathbb{N}$, let us define

$$\mathfrak{I}_n^{\varepsilon,z} := J_n^{v^\varepsilon}(f_n^{\varepsilon,z}) - J_n^v(f_n^z), \quad \text{such that} \quad \boldsymbol{S}_2^\varepsilon = f_0^{\varepsilon,z} - f_0^z + \sum_{n=1}^{\infty}\mathfrak{I}_n^{\varepsilon,z}.$$

According to the Skorohod representation theorem [88, Theorem A.3.9], there exists a probability space on which are defined $(\widetilde{v}^\varepsilon)_{\varepsilon>0}$ and $\widetilde{v}$ having the same law as $(v^\varepsilon)_{\varepsilon>0}$ and v on the original probability space, and such that $\widetilde{v}^\varepsilon$ converges towards $\widetilde{v}$ in an almost sure sense. This is a standard trick in the weak convergence literature; we retain the original labels and continue with almost sure convergence. However, we recall that this convergence takes place in the weak topology in $L^2(\mathbf{T})$, and therefore we need to work out the convergence of $\mathfrak{I}_n^{\varepsilon,z}$ by induction.

Let us fix $n \in \mathbb{N}$ for the moment and recall that J_n^v, defined in (3.3.7), maps $L^2(\mathbf{T}^n)$ to $\mathbb{R}$. We claim that for all $k \in [\![1, n-1]\!]$, the following function, mapping $\mathbf{T}$ to $\mathbb{R}$, tends to zero in $L^2(\mathbf{T})$, almost surely and for almost all $(\boldsymbol{t_{n-k-1}}, \boldsymbol{x_{n-k-1}}) \in \mathbf{T}^{n-k-1}$:

$$J_k^{v^\varepsilon}\big(f_n^{\varepsilon,z}(\boldsymbol{t_{n-k-1}}, \boldsymbol{x_{n-k-1}}, \bullet)\big) - J_k^v\big(f_n^z(\boldsymbol{t_{n-k-1}}, \boldsymbol{x_{n-k-1}}, \bullet)\big). \tag{3.4.11}$$

We name this claim $\mathfrak{C}(k)$.

With this in mind, we start by pointing out a uniform bound. Assumption (3.3.1)(A) entails that, for $\varepsilon > 0$ small enough, $\|f_n^{\varepsilon,z}\|_{L^2(\mathbf{T}^n)} \le 2\,\|f_n^z\|_{L^2(\mathbf{T}^n)}$ for all $n \ge 0$ and $z \in \mathrm{K}$.

Hence, by using Cauchy–Schwarz inequality k times, the following bound stands almost surely, for all $k \le n$ and all $(\boldsymbol{t_{n-k}}, \boldsymbol{x_{n-k}}) \in \mathbf{T}^{n-k}$:

$$\left|J_k^{v^\varepsilon}\left(f_n^{\varepsilon,z}(\boldsymbol{t_{n-k}}, \boldsymbol{x_{n-k}}, \bullet)\right) - J_k^v\left(f_n^z(\boldsymbol{t_{n-k}}, \boldsymbol{x_{n-k}}, \bullet)\right)\right| \le 3M^{k/2} \left\|f_n^z(\boldsymbol{t_{n-k}}, \boldsymbol{x_{n-k}}, \bullet)\right\|_{L^2(\mathbf{T}^k)} . \tag{3.4.12}$$

We recall that if (a_ε) and (b_ε) are sequences in a Hilbert space $(H, \langle\cdot,\cdot\rangle)$ such that $a_\varepsilon \to a$ (strongly) and $b_\varepsilon \rightharpoonup b$ (weakly) as ε goes to zero, then $\langle a_\varepsilon, b_\varepsilon\rangle \to \langle a, b\rangle$; we call this property $\mathfrak{P}(\rightharpoonup)$ for further reference.

Starting with $k = 1$, we know that $\lim_{\varepsilon\downarrow 0} \|f_n^{\varepsilon,z} - f_n^z\|_{L^2(\mathbf{T}^n)} = 0$, which implies that for almost every $(\boldsymbol{t_{n-1}}, \boldsymbol{x_{n-1}}) \in \mathbf{T}^{n-1}$,

$$\lim_{\varepsilon\downarrow 0} \|f_n^{\varepsilon,z}(\boldsymbol{t_{n-1}}, \boldsymbol{x_{n-1}}, \bullet) - f_n^z(\boldsymbol{t_{n-1}}, \boldsymbol{x_{n-1}}, \bullet)\|_{L^2(\mathbf{T})} = 0,$$

and we also know that $v^\varepsilon \rightharpoonup v$, almost surely. Property $\mathfrak{P}(\rightharpoonup)$ entails that, almost surely and for almost every $(\boldsymbol{t_{n-1}}, \boldsymbol{x_{n-1}}) \in \mathbf{T}^{n-1}$, the following goes to zero as ε goes to zero:

$$\begin{aligned} &J_1^{v^\varepsilon}\left(f_n^{\varepsilon,z}(\boldsymbol{t_{n-1}}, \boldsymbol{x_{n-1}}, \bullet)\right) - J_1^v\left(f_n^z(\boldsymbol{t_{n-1}}, \boldsymbol{x_{n-1}}, \bullet)\right) \\ &\qquad = \left\langle f_n^{\varepsilon,z}(\boldsymbol{t_{n-1}}, \boldsymbol{x_{n-1}}, \bullet), v^\varepsilon\right\rangle_{L^2(\mathbf{T})} - \left\langle f_n^z(\boldsymbol{t_{n-1}}, \boldsymbol{x_{n-1}}, \bullet), v\right\rangle_{L^2(\mathbf{T})}. \end{aligned} \tag{3.4.13}$$

Therefore, dominated convergence, induced by (3.4.12), yields that $\mathfrak{C}(1)$ holds.

Now suppose that $\mathfrak{C}(k-1)$ holds, for some $k \in [\![2, n-1]\!]$. We observe that, for any $f \in L^2(\mathbf{T}^k)$,

$$J_k^v(f) = \int_{\mathbf{T}} J_{k-1}^v\big(f(t_1, x_1, \bullet)\big)\, v(t_1, x_1)\, \widehat{\mu}(\mathrm{d}t_1\, \mathrm{d}x_1) = \langle J_{k-1}^v(f), v\rangle_{L^2(\mathbf{T})},$$

which helps us see the following equality:

$$\begin{aligned} &J_k^{v^\varepsilon}\left(f_n^{\varepsilon,z}(\boldsymbol{t_{n-k}}, \boldsymbol{x_{n-k}}, \bullet)\right) - J_k^v\left(f_n^z(\boldsymbol{t_{n-k}}, \boldsymbol{x_{n-k}}, \bullet)\right) \\ &\qquad = \left\langle J_{k-1}^{v^\varepsilon}\left(f_n^{\varepsilon,z}(\boldsymbol{t_{n-k}}, \boldsymbol{x_{n-k}}, \bullet)\right), v^\varepsilon\right\rangle_{L^2(\mathbf{T})} - \left\langle J_{k-1}^v\left(f_n^z(\boldsymbol{t_{n-k}}, \boldsymbol{x_{n-k}}, \bullet)\right), v\right\rangle_{L^2(\mathbf{T})}. \end{aligned}$$

From $\mathfrak{P}(\rightharpoonup)$ and the induction hypothesis $\mathfrak{C}(k-1)$, the quantity above tends to zero almost surely and for almost every $(\boldsymbol{t_{n-k-1}}, \boldsymbol{x_{n-k-1}}) \in \mathbf{T}^{n-k-1}$. By dominated convergence (from (3.4.12)), we deduce that $\mathfrak{C}(k)$ holds, and therefore the claim is proved for all $k \in [\![1, n-1]\!]$. Combining $\mathfrak{C}(n-1)$ with $\mathfrak{P}(\rightharpoonup)$ again yields that $\mathfrak{J}_n^{\varepsilon,z} = J_n^{v^\varepsilon}(f_n^{\varepsilon,z}) - J_n^v(f_n^z)$ tends to zero almost surely as ε goes to zero.

To conclude this proof we observe that, from (3.4.12), $\mathfrak{J}_n^{\varepsilon,z}$ is dominated in the following sense

$$|\mathfrak{J}_n^{\varepsilon,z}| \leq 3M^{n/2} \, \|f_n^z\|_{L^2(\mathbf{T}^n)} =: \mathfrak{J}_n^z,$$

and that $\sum_{n=1}^\infty |\mathfrak{J}_n^z| < \infty$, by Assumption 3.3.1(B). This entails, by dominated convergence, the almost sure convergence:

$$\lim_{\varepsilon\downarrow 0} \boldsymbol{S}_2^\varepsilon = \lim_{\varepsilon\downarrow 0}\left(f_0^{\varepsilon,z} - f_0^z + \sum_{n=1}^\infty \mathfrak{J}_n^{\varepsilon,z}\right) = \sum_{n=1}^\infty \lim_{\varepsilon\downarrow 0} \mathfrak{J}_n^{\varepsilon,z} = 0.$$

Coming back from Skorohod's world, we in fact proved that, as ε goes to zero, $\sum_{n=0}^\infty m_\theta(f_n^{\varepsilon,z})$ converges towards $\overline{X}^v(z)$ in distribution in the original probability space, where $\theta \in \Theta(n,n)$. We also showed that $\boldsymbol{S}_1^\varepsilon$ tends to zero in probability. Put together, they imply that $\boldsymbol{S}_1^\varepsilon + \sum_{n=0}^\infty m_\theta(f_n^{\varepsilon,z})$ converges towards $\overline{X}^v(z)$ in distribution, as claimed. □

Lemma 3.4.14 (Compactness). *Under Assumption 3.3.1, the following sets are compact subsets of $\mathcal{C}(\mathrm{K})$ for all $M > 0$:*

$$\Gamma_M = \left\{ \overline{X}^v : v \in \mathcal{S}_M \right\}.$$

Proof. The proofs of relative compactness and closure parallel the proofs of tightness and convergence of the marginals, only in a deterministic and therefore simpler setting. Let us fix $M > 0$ and consider a sequence $(\overline{X}^{v_n})_{n\in\mathbb{N}} \subset \Gamma_M$. Similar calculations as in the proof of Lemma 3.4.10 yield

$$\sup_{z\in\mathrm{K},\, n\in\mathbb{N}} \left|\overline{X}^{v_n}(z)\right| < \infty.$$

Furthermore, the equicontinuity follows as in the proof of Lemma 3.4.11:

$$\left|\overline{X}^{v_n}(z) - \overline{X}^{v_n}(y)\right| \leq \omega\big(\|z-y\| \big).$$

Therefore an application of Arzelá–Ascoli's theorem yields the relative compactness of Γ_M.

Consider a converging sequence $(\overline{X}^{v_n})_{n\in\mathbb{N}} \subset \Gamma_M$ and denote its limit by X^∞. Since $\mathcal{S}_M$ is compact with respect to the weak topology, there exists a converging subsequence $(v_{n_k})_{k\geq 1}$ of (v_n), with limit $v \in \mathcal{S}_M$. Then using the arguments of convergence of $\boldsymbol{S}_2^\varepsilon$ in the proof of Lemma 3.4.13, one can prove that for all $z \in \mathrm{K}$,

$$\lim_{k\uparrow\infty} \left|X^{v_{n_k}}(z) - \overline{X}^v(z)\right| = 0.$$

Hence, we deduce that $X^\infty = \overline{X}^v$, which implies $X^\infty \in \Gamma_M$. This entails that Γ_M is closed and thus compact. □

Proof of Theorem 3.3.3. Lemmas 3.4.11 and 3.4.13 respectively show the tightness of $(X^{\varepsilon,v^\varepsilon})_{\varepsilon>0}$ in $\mathcal{C}(\mathrm{K})$ and the convergence of its finite-dimensional distributions to those of $\overline{X}^v$. Together they yield the weak convergence condition of Assumption 3.4.2(a). Lemma 3.4.14 takes care of the other condition that needs to be satisfied for Theorem 3.4.4 to hold. Therefore a direct application of the latter yields the desired claim. □

3.5 Applications

The kernels of a random variable's chaos expansion are sometimes delicate to compute explicitly, making Assumption 3.3.1 difficult to verify, but there exist interesting cases meeting this requirement. In this section, we exhibit families of processes which are covered by our framework along with the sufficient conditions for the pathwise LDP to hold.

We start by displaying in subsection 3.5.1 a formula to compute the kernels of infinitely Malliavin differentiable processes. The following two subsections introduce a route to apply our main result (Theorem 3.3.3) to multiple Skorohod integrals and linear Skorohod equations driven by a white noise measure. Large deviations for anticipating equations were derived by exploiting their flow property in [189, 192] and [191], for Stratonovich and Skorohod integrals respectively. However we note that these results are limited to the case of Brownian motion and single integral.

The Wick–Itô integral is a conducive way of defining stochastic integrals driven by a fractional Brownian motion; we present a way of computing the kernels of such an integral in subsection 3.5.4. This also leads to sufficient conditions for an LDP to hold. Integrating against the fractional Brownian motion is a tricky business: large deviations were studied recently in the case $H > \frac{1}{2}$ [52], while our approach covers the whole range $H \in (0,1)$. Finally, subsection 3.5.5 introduces a class of processes inspired from closed martingales in $\mathbb{R}_+$ where the continuity criterion (Assumption 3.3.1(C)) is simplified.

For clarity of exposition, we will replace (t,x) by t, even where $\mathbf{T}$ represents time-space.

3.5.1 Malliavin derivatives

Adopting notations and definitions from [201, Section 1.2], we denote by D the Malliavin derivative operator and by $\mathbb{D}^{1,2}$ its domain of application in $L^2(\Omega)$. For all $n \in \mathbb{N}$, we also consider the iterated derivatives D^n and their domain of application $\mathbb{D}^{n,2}$, which are dense in $L^2(\Omega)$. If $F \in \mathbb{D}^{\infty,2} := \bigcap_{n\in\mathbb{N}} \mathbb{D}^{n,2}$, then the kernels of its chaos expansion are given explicitly by (see for instance [201, Exercise 1.2.6]):

$$f_n = \frac{1}{n!}\,\mathbb{E}[\mathrm{D}^n F]. \tag{3.5.1}$$

This yields a verifiable condition for Assumption 3.3.1 to hold, provided one can compute the $L^2(\mathbf{T}^n)$-norm of the kernels. We note that in the white noise case, the Malliavin derivative $\mathrm{D}F$ is a stochastic process denoted $\{\mathrm{D}_t F : t \in \mathbf{T}\}$, and similarly $\mathrm{D}^n F$ can be written $\{\mathrm{D}^n_{\boldsymbol{t_n}} F : \boldsymbol{t_n} \in \mathbf{T}^n\}$.

Example 3.5.1. *Let* K *and* $\mathbf{T}$ *be as general as possible and consider for all* $\varepsilon > 0$ *and* $z \in \mathrm{K}$ *the small-noise process*

$$X^\varepsilon(z) = \exp\left(\varepsilon \int_{\mathbf{T}} h^z(t)\, W(\mathrm{d}t) - \frac{1}{2}\varepsilon^2 \left\|h^z\right\|^2_{L^2(\mathbf{T})}\right),$$

where $h^z \in L^2(\mathbf{T})$ *for all* $z \in \mathrm{K}$ *and* $\sup_{z\in\mathrm{K}} \|h^z\|_{L^2(\mathbf{T})} < \infty$. *Notice that* $\mathbf{T}$ *is not necessarily equal to* K*, hence the question of adaptation to a filtration is here meaningless. Straightforward computations yield* $\mathbb{E}[X^\varepsilon(z)] = 1$ *and* $\mathrm{D}^n_{\boldsymbol{t_n}} X^\varepsilon(z) = \varepsilon^n X^\varepsilon(z) \prod_{i=1}^n h^z(t_i)$ *for any* $\boldsymbol{t_n} \in \mathbf{T}^n$. *These give us the form of the kernels of the chaos expansion of* X^ε *by* (3.5.1)*:*

$$f_n^{\varepsilon,z}(\boldsymbol{t_n}) = \frac{\varepsilon^n}{n!}\prod_{i=1}^n h^z(t_i), \quad \text{for all } n \in \mathbb{N},\ \varepsilon > 0,\ z \in \mathrm{K}, \boldsymbol{t_n} \in \mathbf{T}^n,$$

which shows that $f_0^{\varepsilon,z}$ *is independent of* ε *and hence Assumption 3.3.1(A) holds. So does*

Assumption 3.3.1(B) since we have

$$\|f_n^{\varepsilon,z}\|_{L^2(\mathbf{T}^n)} = \frac{\varepsilon^n}{n!}\,\|h^z\|^n_{L^2(\mathbf{T})}\,.$$

From the observation that $\prod_{i=1}^n a_i - \prod_{i=1}^n b_i = \sum_{i=1}^n \Big((a_i - b_i)\prod_{j\neq i} c_j\Big)$*, where* c_j *is either* a_j *or* b_j*, we can derive*

$$\|f_n^{\varepsilon,z} - f_n^{\varepsilon,y}\|_{L^2(\mathbf{T}^n)} \leq \frac{\varepsilon^n}{n!} n\,\|h^z - h^y\|_{L^2(\mathbf{T})} \max\left\{\, \|h^z\|_{L^2(\mathbf{T})}\,, \|h^y\|_{L^2(\mathbf{T})}\,\right\}^{n-1}.$$

This entails Assumption 3.3.1(C) holds as soon as $\|h^z - h^y\|_{L^2(\mathbf{T})} \leq \omega(\|z-y\|)$ *for an appropriate modulus of continuity* ω*.*

Example 3.5.2. *Let* $\mathrm{K} = \mathbf{T} = [0,1]$ *such that we recover the Freidlin-Wentzell setting of Example 3.3.6, and consider the small-noise SDE*

$$X^\varepsilon(t) = x + \varepsilon\int_0^t \sigma(X^\varepsilon(s))\,\mathrm{d}W_s, \qquad \textit{for all } t\in\mathbf{T},\ \varepsilon>0,$$

where $x\in\mathbb{R}$*,* $\sigma\in C^\infty(\mathbb{R})$ *has linear growth and* W *is a standard Brownian motion. We ignore the drift for simplicity.*

Let us consider the following induction hypothesis: for a given $n\geq 0$ *and for all* $p\geq 1$*, there exists* $C_p<\infty$ *such that* $\sup_{t\in\mathbf{T}}\mathbb{E}\left[\left|\mathrm{D}^k_{\boldsymbol{t_k}}X^\varepsilon(t)\right|^p\right] \leq C_p^k$ *for all* $k\leq n$*. This is clearly true for* $n=0$ *by a standard argument based on Grönwall and Burkholder-Davis-Gundy inequalities. Using the chain rule* $\mathrm{D}\sigma(X) = \mathrm{D}X\sigma'(X)$*, Hölder's inequality and* $\sigma\in C^\infty(\mathbb{R})$*, one can conclude that a similar uniform bound holds for* $\mathrm{D}^n_{\boldsymbol{t_n}}\sigma(X^\varepsilon(t))$*, with a possibly different constant than* C_p*. Then one can show that*

$$\mathrm{D}^{n+1}_{\boldsymbol{t_{n+1}}}X^\varepsilon(t) = \varepsilon\int_0^t \mathrm{D}^{n+1}_{\boldsymbol{t_{n+1}}}\sigma(X^\varepsilon(s))\,\mathrm{d}W_s + \varepsilon\sum_{i=1}^{n+1}\mathrm{D}^n_{\widehat{t_n}(i)}\sigma(X^\varepsilon(t_i))\mathbb{1}_{t_i\leq t} \tag{3.5.2}$$

where $\widehat{t_n}(i) := (t_1,\cdots,t_{i-1},t_{i+1},\cdots,t_{n+1})$ *for all* $n\geq 0$*. The terms in the sum are clearly bounded in* $L^p(\Omega)$ *and since* X^ε *is adapted they are non zero only if* $t_i\geq t_k$ *for all* $k\leq n+1$*. Then using the chain rule again, Hölder's inequality and* $\sigma\in C^\infty(\mathbb{R})$ *we deduce that there exists yet another constant* C_p *such that*

$$\mathbb{E}\left[\left|\int_0^t \mathrm{D}^{n+1}_{\boldsymbol{t_{n+1}}}\sigma(X^\varepsilon(s))\,\mathrm{d}W_s\right|^p\right] \leq C_p^n\left(1+\int_0^t \mathrm{D}^{n+1}_{\boldsymbol{t_{n+1}}}X^\varepsilon(s)\ \mathrm{d}s\right)$$

Grönwall's inequality thus entails that the induction hypothesis holds at $n+1$*, with* *a possibly different constant. Therefore the kernels* $(f_n^{\varepsilon,t})$ *are well-defined by* (3.5.1). *Clearly* $f_0^{\varepsilon,t} = x$*, and moreover, for* $0 \leq s \leq t \leq 1$*, some constant* C *and all* $n \in \mathbb{N}$*, we have by* (3.5.2)

$$\left\|\mathbb{E}\left[\mathrm{D}^n_{\boldsymbol{t_n}} X^\varepsilon(t)\right]\right\|^2_{L^2(\mathbf{T}^n)} \leq n \sum_{i=1}^n \int_{\mathbf{T}^n} \mathbb{E}\left[\left|\mathrm{D}^{n-1}_{\widehat{t_{n-1}(i)}} \sigma(X^\varepsilon(t_i))\right|^2\right] \mathrm{d}\boldsymbol{t_n} \leq C^n;$$

$$\left\|\mathbb{E}\left[\mathrm{D}^n_{\boldsymbol{t_n}} \left(X^\varepsilon(t) - X^\varepsilon(s)\right)\right]\right\|^2_{L^2(\mathbf{T}^n)} \leq n \sum_{i=1}^n \int_s^t \int_{\mathbf{T}^{n-1}} \mathbb{E}\left[\left|\mathrm{D}^{n-1}_{\widehat{t_{n-1}(i)}} \sigma(X^\varepsilon(t_i))\right|^2\right] \mathrm{d}\widehat{t_n}(i)\, \mathrm{d}t_i \leq C^n(t-s).$$

These estimate prove that Assumption 3.3.1 (B) and (C) are satisfied and thus the LDP from Theorem 3.3.3 holds in $\mathcal{C}(\mathrm{K})$*. Of course, this was already known since Freidlin and Wentzell [105] with the seemingly different rate function*

$$I(\psi) = \inf\left\{\frac{1}{2}\int_0^1 v_t^2\, \mathrm{d}t : v \in L^2([0,1]),\ \psi(t) = x + \int_0^t \sigma(\psi(s))v_s\, \mathrm{d}s\right\}.$$

Yet, Lemma 4.1.4 in [82] certifies that these two rate functions must be equal. While the representation above depends on an implicit ODE, (3.3.8) *enforces an explicit expansion for* ψ*. This provides an alternative viewpoint on families of processes which LDP can be derived from both approaches.*

3.5.2 Skorohod integrals

We introduce the divergence operator δ as the adjoint of the Malliavin derivative D, and refer to [201, Section 1.3] for the details. In the white noise case, the domain of δ is a subset of $L^2(\Omega, L^2(\mathbf{T}))$ and for $Y \in \mathrm{Dom}(\delta)$, $\delta(Y)$ is called the Skorohod integral. If, for all $t \in \mathbf{T}$, Y_t has the chaos expansion $Y_t = \sum_{n=0}^\infty I_n(f_n^t)$, then the following expansion holds [201, Proposition 1.3.7]

$$\delta(Y) = \sum_{n=1}^\infty I_n\big(\widetilde{f}_{n-1}\big),$$

where $\widetilde{f}_n$ is the symmetrisation of $f_n^\cdot$

$$\widetilde{f}_n(\boldsymbol{t_n}, t) := \frac{1}{n+1}\left(f_n^t(\boldsymbol{t_n}) + \sum_{i=1}^n f_n^{t_i}(t_1, \cdots, t_{i-1}, t, t_{i+1}, \cdots, t_n)\right).$$

The symmetrisation allows us to recover our initial setting from subsection 3.2.2.

Naturally, we can also consider multiple integrals of the form $\delta^k(Y)$, for some $k \in \mathbb{N}$

and $Y \in \mathrm{Dom}(\delta^k) \subset L^2(\Omega, L^2(\mathbf{T}^k))$, where we define by iteration $\delta^k(Y) := \delta(\delta^{k-1}(Y))$. We introduce $(X(z))_{z\in\mathrm{K}}$ as $X(z) := \delta^k(Y^z(W))$, for some process $Y^z_{\boldsymbol{t_k}}(W) := \sum_{n=0}^{\infty} I_n(f_n^{z,\boldsymbol{t_k}})$ such that $Y^z(W) \in \mathrm{Dom}(\delta^k)$, for all $z \in \mathrm{K}$.

For simplicity, we restrict this example to "small-noise" large deviations, that means processes that can be written as $\mathcal{G}(\varepsilon W)$ where $\mathcal{G}$ is a measurable map from $\mathcal{M}(\mathbf{T})$ to $\mathcal{C}(\mathrm{K})$, as the kernels of their expansion does not depend on ε and they easily meet Assumption 3.3.1(A). One could of course study processes of the type $\mathcal{G}^\varepsilon(\varepsilon W)$ with kernels depending on ε.

Let us define the measurable map $\mathcal{G}\colon \mathcal{M}(\mathbf{T}) \to \mathcal{C}(\mathrm{K})$ such that $X = \mathcal{G}(W)$, and the small-noise perturbation given by

$$X^\varepsilon(z) := \mathcal{G}(\varepsilon W)(z) = \varepsilon^k I_k\big(\widetilde{f}_0^{\varepsilon,z}\big) + \sum_{n=k+1}^{\infty} \varepsilon^n I_n\big(\widetilde{f}_{n-k}^z\big), \tag{3.5.3}$$

where $f_0^{\varepsilon,z}(\boldsymbol{t_k}) := \mathbb{E}\big[Y^z_{\boldsymbol{t_k}}(\varepsilon W)\big]$ and the kernels require k successive symmetrisations. The kernels of the white noise chaos expansion of X^ε are thus explicit functions of the kernels of Y. Therefore Theorem 3.3.3 ensures that if $\widetilde{f}_0^{\varepsilon,z}$ converges to some $\widetilde{f}_0^z$ and the family $(f_n^z)_{z\in\mathrm{K},n\in\mathbb{N}}$ satisfies Assumption 3.3.1(B,C) as functions of $L^2(\mathbf{T}^{n+k})$, then a pathwise LDP for $(X^\varepsilon)_{\varepsilon>0}$ holds with rate function

$$\begin{aligned}\Lambda(\psi) &:= \inf\left\{\frac{1}{2}\|v\|^2_{L^2(\mathbf{T})} : v \in L^2(\mathbf{T}),\, \psi(z) = \sum_{n=k}^{\infty} J_n^v(\widetilde{f}_{n-k}^z)\right\}\\ &= \inf\left\{\frac{1}{2}\|v\|^2_{L^2(\mathbf{T})} : v \in L^2(\mathbf{T}),\, \psi(z) = J_k^v\left(\sum_{n=0}^{\infty} J_n^v(\widetilde{f}_n^z)\right)\right\}.\end{aligned} \tag{3.5.4}$$

Let us consider the case where $\mathbf{T}^k = \mathrm{K}$, f_n^z are already symmetric functions of $L^2(\mathbf{T}^{n+k})$ and $f_n^{z,\boldsymbol{t_k}}$ satisfy Assumption 3.3.1(B,C) as functions of $L^2(\mathbf{T}^n)$. Then for each $z \in \mathrm{K}$, $(Y^z(\varepsilon W))_{\varepsilon>0}$ satisfies an LDP in $\mathcal{C}(\mathbf{T}^k)$ with rate function

$$\widehat{\Lambda}(\phi^z) := \inf\left\{\frac{1}{2}\|v\|^2_{L^2(\mathbf{T})} : v \in L^2(\mathbf{T}),\, \phi^z(\boldsymbol{t_k}) = \sum_{n=0}^{\infty} J_n^v(f_n^{z,\boldsymbol{t_k}})\right\}.$$

The display (3.5.4) leads to the new representation

$$\Lambda(\psi) = \inf\left\{\widehat{\Lambda}(\phi^z) : \phi^z \in \mathcal{C}(\mathrm{K}), \psi(z) = J_k^v(\phi^z)\right\},$$

which is reminiscent of the contraction principle and of [118, Theorem 2.1] where conditions on the integrand yield an LDP for the stochastic integral.

3.5.3 Linear Skorohod equations

We let $\mathrm{K} = \mathbf{T} = [0,1]^d$ and consider for all $\varepsilon > 0$ the small-noise equation

$$X^\varepsilon(t) = x_0^t + \varepsilon\delta(a^t X^\varepsilon), \tag{3.5.5}$$

where $x_0^t \in \mathbb{R}$ and $a^t : \mathbf{T} \to \mathbb{R}$ for all $t \in \mathbf{T}$. Plugging in the putative chaos expansion of X, the equation becomes

$$\sum_{n=0}^{\infty} \varepsilon^n I_n(f_n^t) = x_0^t + \sum_{n=1}^{\infty} \varepsilon^n I_n(g_{n-1}^t), \tag{3.5.6}$$

where $g_{n-1}^t(\boldsymbol{t_n}) := a^t(t_n) f_{n-1}^{t_n}(\boldsymbol{t_{n-1}})$ for all $n \in \mathbb{N}$. Ignoring the symmetrisation for now, we identify the Wiener chaos on both sides and deduce that $f_0^t = x_0^t$ and $f_n^t(\boldsymbol{t_n}) = g_{n-1}^t(\boldsymbol{t_n})$ for all $n \in \mathbb{N}$. Hence by induction we obtain

$$f_n^t(\boldsymbol{t_n}) = a^t(t_n) a^{t_n}(t_{n-1}) \cdots a^{t_2}(t_1) x_0^{t_1}, \tag{3.5.7}$$

which we can symmetrise in the following way:

$$\widetilde{f}_n^t(\boldsymbol{t_n}) = \frac{1}{n!} \sum_{\sigma} f_n^t(t_{\sigma(1)}, \cdots, t_{\sigma(n)}), \tag{3.5.8}$$

with σ running over all permutations of $\{1, \cdots, n\}$. This allows to recover the chaos expansion of our framework and give a proper meaning to the solution of (3.5.5). We note that the kernels are independent of ε so that we can define the map $\mathcal{G}$ as

$$X^\varepsilon := \mathcal{G}(\varepsilon W) = \mathbb{E}[\mathcal{G}(\varepsilon W)] + \sum_{n=1}^{\infty} \varepsilon^n I_n(\widetilde{f}_n^t),$$

Theorem 3.3.3 yields explicit criteria for the LDP of $(X^\varepsilon)_{\varepsilon>0}$ to hold with rate function (3.3.8).

Example 3.5.3. *To illustrate this, let us look at the special case* $\mathrm{K} = \mathbf{T} = [0,1]$ *and* $a^t(s) = a^t \mathbb{1}_{s \le t}$ *for some* $a^t \in \mathbb{R}$. *Then* (3.5.7) *yields* $f_n^t(\boldsymbol{t_n}) = \mathbb{1}_{t_1 \le t_2 \le \cdots \le t_n \le t}\, a^t a^{t_n} \cdots a^{t_2} x_0^{t_1}$. *We* *first notice that for almost all $\boldsymbol{t_n} \in \mathbf{T}^n$*

$$\Big|\sum_{\sigma} \mathbb{1}_{t_{\sigma(1)}\le t_{\sigma(2)}\le\cdots\le t_{\sigma(n)}\le t}\, a^t a^{t_{\sigma(n)}}\cdots a^{t_{\sigma(2)}} x_0^{t_{\sigma(1)}}\Big|^2 = \sum_{\sigma} \mathbb{1}_{t_{\sigma(1)}\le t_{\sigma(2)}\le\cdots\le t_{\sigma(n)}\le t} \left|a^t a^{t_{\sigma(n)}}\cdots a^{t_{\sigma(2)}} x_0^{t_{\sigma(1)}}\right|^2$$

because the cross terms vanish whenever $t_1, \cdots, t_n$ are distinct. Hence from the symmetrisation (3.5.8) *we can compute explicitly*

$$\begin{aligned}
\left\|\widetilde{f}_n^t\right\|^2_{L^2(\mathbf{T}^n)} &= \frac{1}{(n!)^2}\int_{\mathbf{T}^n}\Big|\sum_{\sigma} \mathbb{1}_{t_{\sigma(1)}\le t_{\sigma(2)}\le\cdots\le t_{\sigma(n)}\le t}\, a^t a^{t_{\sigma(n)}}\cdots a^{t_{\sigma(2)}} x_0^{t_{\sigma(1)}}\Big|^2 \mathrm{d}\boldsymbol{t_n} \\
&= \left(\frac{a^t}{n!}\right)^2 \sum_{\sigma}\int_0^t\int_0^{t_{\sigma(n)}}\cdots\int_0^{t_{\sigma(2)}} \left|a^{t_{\sigma(n)}}\cdots a^{t_{\sigma(2)}} x_0^{t_{\sigma(1)}}\right|^2 \mathrm{d}\boldsymbol{t_n} \\
&= \left(\frac{a^t}{n!}\right)^2 \sum_{\sigma}\frac{1}{n!}\int_{[0,t]^n} \left|a^{t_{\sigma(n)}}\cdots a^{t_{\sigma(2)}} x_0^{t_{\sigma(1)}}\right|^2 \mathrm{d}\boldsymbol{t_n} \\
&= \left(\frac{a^t}{n!}\right)^2 \|a\|^{2(n-1)}_{L^2[0,t]}\, \|x_0\|^2_{L^2[0,t]},
\end{aligned}$$

which yields Assumption 3.3.1(B) as the kernels are of exponential type (3.3.4). *Analogously we see that*

$$\left\|\widetilde{f}_n^t - \widetilde{f}_n^s\right\|_{L^2(\mathbf{T}^n)} \le \left|a^t - a^s\right| \frac{\|a\|^{n-1}_{L^2(\mathbf{T})}\,\|x_0\|_{L^2(\mathbf{T})}}{n!},$$

which shows that the continuity criterion depends explicitly on the regularity of the map $t \mapsto a^t$. For instance, if the latter is Hölder continuous then Assumption 3.3.1(C) is automatically satisfied.

One could generalise (3.5.5) by considering multiple Skorohod integrals or taking a^t to belong to a given Wiener chaos and applying [201, Proposition 1.1.3].

3.5.4 Fractional integrals in the Wick–Itô sense

Let $\mathrm{K} = \mathbf{T} = [0,1]$ and $\widehat{\mu}$ be the Lebesgue measure, such that W is a standard Brownian motion. We will use the Hida space $(\mathcal{S})$ of stochastic test functions and its dual $(\mathcal{S})^*$ the Hida space of stochastic distributions; we refer to [34, Definition A.1.4] for the definitions. Consider the Hermite functions $(\xi_k)_{k\ge1}$ [34, Equation (A.6)] and the operator M [34, Equation (4.2)] which, in particular, has the property $\int_{\mathbb{R}} \mathrm{M}\mathbb{1}_{[0,t]}(x)\mathrm{M}\mathbb{1}_{[0,s]}(x)\,\mathrm{d}x = R_H(t,s)$, the covariance function of the fractional Brownian motion. The latter was introduced in Example 3.2.4 and we denote it by W^H.

The fractional white noise $\widetilde{W}^H$, not to be confused with the white noise *measure*, is seen as the derivative of W^H in $(\mathcal{S})^*$ and is defined in [34, Equation (4.16)]:

$$\widetilde{W}^H(t) := \sum_{k=1}^{\infty} \mathrm{M}\xi_k(t) I_1(\xi_k).$$

Let $Y : \mathbf{T} \to L^2(\Omega) \subset (\mathcal{S})^*$ have the Wiener chaos expansion $Y_t = \sum_{n=0}^{\infty} I_n(f_n^t)$ with respect to the standard Brownian motion. Then the Wick–Itô integral with respect to the fractional Brownian motion W^H is defined as [34, Definition 4.2.2]

$$\int_{\mathbf{T}} Y_t \,\mathrm{d}W_t^H := \int_{\mathbf{T}} Y_t \diamond \widetilde{W}^H(t) \,\mathrm{d}t,$$

where $\diamond$ denotes the Wick product. We then notice by [34, Theorem A.3.7] and [201, Proposition 1.1.2] that for all $f \in L^2(\mathbf{T}^n)$ and $g \in L^2(\mathbf{T})$

$$I_n(f)\diamond I_1(g) = I_n(f)I_1(g) - \int_{\mathbf{T}} g(t)\mathrm{D}_t I_n(f)\,\mathrm{d}t = I_{n+1}(f\otimes g) + nI_{n-1}(f\otimes_1 g) - \int_{\mathbf{T}} g(t)\, nI_{n-1}\big(f(\cdot,t)\big)\,\mathrm{d}t,$$

where $(f \otimes_1 g)(\boldsymbol{t_{n-1}}) := \int_{\mathbf{T}} f(\boldsymbol{t_{n-1}}, t) g(t)\,\mathrm{d}t$. Therefore we obtain $I_n(f) \diamond I_1(g) = I_{n+1}(f \otimes g)$. By linearity this yields, if the series are convergent,

$$\int_{\mathbf{T}} Y_s \diamond \widetilde{W}^H(s)\,\mathrm{d}t = \int_{\mathbf{T}} \sum_{\substack{k=1\\n=0}}^{\infty} \Big(I_{n+1}(f_n^s \otimes \xi_k)\mathrm{M}\xi_k(s)\Big)\,\mathrm{d}s = \sum_{n=1}^{\infty} I_n(g_n), \tag{3.5.9}$$

where $g_n := \int_{\mathbf{T}} \sum_{k=1}^{\infty} (f_{n-1}^s \otimes \xi_k)\mathrm{M}\xi_k(s)\,\mathrm{d}s$. One can symmetrise the kernels as in (3.5.8) to recover the white noise chaos expansion.

For all $\varepsilon > 0$, $t \in \mathbf{T}$, let us introduce the family of processes $X^\varepsilon(t) = \varepsilon \int_{\mathbf{T}} Y_s^t(\varepsilon W)\,\mathrm{d}W_s^H$, where the integral is taken in the Wick–Itô sense and $Y_s^t(W) := \sum_{n=0}^{\infty} I_n(f_n^{s,t})$. There exists a measurable map $\mathcal{G} : \mathcal{M}(\mathbf{T}) \to \mathbb{R}$ such that $\mathcal{G}(\varepsilon W) := X^\varepsilon$ and similarly to (3.5.9) we find that

$$X^\varepsilon(t) = \varepsilon I_1\left(g_1^{\varepsilon,t}\right) + \sum_{n=2}^{\infty} \varepsilon^n I_n\left(g_n^t\right),$$

where $g_1^{\varepsilon,t} := \int_{\mathbf{T}} \sum_{k=1}^{\infty} \left(\mathbb{E}[Y_s^t(\varepsilon W)] \otimes \xi_k\right)\mathrm{M}\xi_k(s)\,\mathrm{d}s$, and $g_n^t := \int_{\mathbf{T}} \sum_{k=1}^{\infty} \left(f_{n-1}^{s,t} \otimes \xi_k\right)\mathrm{M}\xi_k(s)\,\mathrm{d}s$ for all $n \geq 2$. After symmetrisation of the kernels, sufficient conditions for the large deviations of $(X^\varepsilon)_{\varepsilon>0}$ follow from Theorem 3.3.3 and can be reduced to estimates on $\|f_{n-1}^{s,t} \otimes \xi_k\|_{L^2(\mathbf{T}^n)}$ and $\|(f_{n-1}^{s,t_1} - f_{n-1}^{s,t_2}) \otimes \xi_k\|_{L^2(\mathbf{T}^n)}$. These conditions imply in particular that the series are convergent and that X^ε has continuous paths.

Similarly to the previous section we can consider fractional equations of the type

$$X^{\varepsilon}(t) = x_0^t + \varepsilon \int_{\mathbf{T}} a^t(s) X^{\varepsilon}(s)\, \mathrm{d}W_s^H,$$

where a is deterministic, and identify the kernels of the chaos expansion as in (3.5.6). They satisfy $f_0^t = x_0^t$ and

$$f_n^t(\boldsymbol{t_n}) = \int_{\mathbf{T}} a^t(s) f_{n-1}^s(\boldsymbol{t_{n-1}}) \sum_{k=1}^{\infty} \xi_k(t_n) \mathrm{M}\xi_k(s)\, \mathrm{d}s, \quad \text{for all } n \in \mathbb{N}.$$

Even though the expression is more intricate than in the white noise case (3.5.7), after symmetrisation this also yields sufficient conditions for the pathwise LDP of the small-noise version of X.

3.5.5 Martingales

We come back to the general case for $\mathrm{K}, \mathbf{T}$ and $\widehat{\mu}$ and illustrate the regularity condition (3.3.3) through an example where the z-dependence of the kernels is more explicit, and show the connection with closed martingales. Consider the case where the kernels have the form $f_n^{\varepsilon,z} = \mathfrak{f}_n^{\varepsilon}\, \mathfrak{f}_n^z$, for some functions $\mathfrak{f}_n^{\varepsilon}, \mathfrak{f}_n^z \colon \mathbf{T}^n \to \mathbb{R}$ and for all $\varepsilon > 0$, $z \in \mathrm{K}$ and $n \geq 0$. Then, for all $q > 2$, we can write by Hölder's inequality

$$\|f_n^{\varepsilon,z} - f_n^{\varepsilon,y}\|_{L^2(\mathbf{T}^n)} \leq \|\mathfrak{f}_n^{\varepsilon}\|_{L^q(\mathbf{T}^n)} \|\mathfrak{f}_n^z - \mathfrak{f}_n^y\|_{L^{\frac{2q}{q-2}}(\mathbf{T}^n)},$$

which allows to split the responsibilities of Assumption 3.3.1 (B) and (C). Indeed, both are satisfied if there exist $q > 2$ and $\Delta > 0$ such that, for all n large enough, all ε small enough and all $y, z \in \mathrm{K}$,

$$\|\mathfrak{f}_n^{\varepsilon}\|_{L^q(\mathbf{T}^n)} \leq \frac{\Delta^n}{n!} \quad \text{and} \quad \|\mathfrak{f}_n^z - \mathfrak{f}_n^y\|_{L^{\frac{2q}{q-2}}(\mathbf{T}^n)} \leq \omega(\|z-y\|). \tag{3.5.10}$$

For all $A \in \mathcal{T}$, let $\mathcal{F}_A$ be the σ-field generated by $\{\dot{W}(B) : B \subset A,\, B \in \mathcal{T}\}$. Moreover, assume there exists an $L^2(\Omega)$-random variable $\widetilde{X}^{\varepsilon} = \sum_{n=0}^{\infty} \varepsilon^n I_n(\mathfrak{f}_n^{\varepsilon})$ such that, for all $z \in \mathrm{K}$, there exists $A_z \in \mathcal{T}$ such that $X^{\varepsilon}(z) = \mathbb{E}\big[\widetilde{X}^{\varepsilon} | \mathcal{F}_{A_z}\big]$. Then, Lemma 1.2.5 in [201] entails that

$$X^{\varepsilon}(z) = \sum_{n=0}^{\infty} \varepsilon^n I_n\big(\mathfrak{f}_n^{\varepsilon}\, \mathbb{1}_{A_z}^{\otimes n}\big).$$

An LDP for $(X^{\varepsilon})_{\varepsilon>0}$ thus holds if (3.5.10) is satisfied with $\mathfrak{f}_n^z = \mathbb{1}_{A_z}^{\otimes n}$.

Example 3.5.4. *Let* $\mathrm{K} = \mathbf{T} = [0,1]^d$ *and* $A_z = [0,z] := \{y \in \mathbf{T} : y(i) \leq z(i) \text{ for all } i\}$. *Then, for* $\|z-y\| \leq 1$, *we have*

$$\|\mathfrak{f}_n^z - \mathfrak{f}_n^y\|_{L^{\frac{2q}{q-2}}(\mathbf{T}^n)} = \left\|\mathbb{1}_{A_z}^{\otimes n} - \mathbb{1}_{A_y}^{\otimes n}\right\|_{L^{\frac{2q}{q-2}}(\mathbf{T}^n)} = \left(\prod_{i=1}^d |z(i)-y(i)|\right)^{n\frac{q-2}{2q}} \leq \|z-y\|^{n\frac{q-2}{2q}} \leq \|z-y\|^{\frac{q-2}{2q}}.$$

If $d=1$, X^ε *is a closed martingale and this type of kernel corresponds to the "adapted case" in [187].*

3.6 Conclusion

This chapter provides a pathwise large deviations principle for Wiener chaos expansions, and thus establishes an intermediate level of generality between the abstract machinery of Budhiraja and Dupuis [45] and specific models such as stochastic equations. We also show that this approach embraces and sheds new light on the asymptotic behaviour of Gaussian functionals that can be examined through the chaos expansion lens. Finally, we believe it opens up room for future research both horizontally—for different types of driving noise, and vertically—for exploring the scope of applications.

4

Smile asymptotics in multi-factor rough volatility models

The results presented in this chapter correspond to a working paper written jointly with Antoine Jacquier and Aitor Muguruza [159].

4.1 Introduction

Exposure to the uncertain dynamics of volatility is a desirable feature of most current trading strategies, and has naturally generated a growing interest in volatility derivatives. From a theoretical viewpoint, an adequate financial model should reproduce the volatility dynamics accurately and consistently with those of the asset price; any dicrepancy may otherwise lead to an arbitrage opportunity. Despite the extensive academic literature, options on VIX and S&P 500 index still display different volatility surfaces, betraying the lack of a proper modelling framework. This issue is well-known as a the SPX-VIX joint calibration problem and has motivated a number of creative modelling innovations in the past 15 years.

Reconciling both implied volatilities requires additional factors to enrich the variance curve dynamics, argued Bergomi in [30, 31] where he proposed what we now call the two-factor Bergomi model. In forward variance form and with W^1, W^2 two correlated Brownian motions, it reads

$$\frac{\mathrm{d}\xi_t(T)}{\xi_t(T)} = c_1 \mathrm{e}^{-\kappa_1(T-t)} \mathrm{d}W_t^1 + c_2 \mathrm{e}^{-\kappa_2(T-t)} \mathrm{d}W_t^2, \tag{4.1.1}$$

where $T \geq t$ and $c_1, c_2, \kappa_1, \kappa_2 > 0$. Gatheral [122], seduced by the additional factor that allows to disentangle different aspects of implied volatility and allows humps in the variance curve, introduced a mean-reverting volatility model, the double CEV model, where the mean of the variance follows a CEV model itself. Although promising, these attempts fall short of reproducing jointly the short-time behaviour of the SPX and VIX implied volatilities. A variety of new models were suggested to tackle this issue, both with continuous paths [15, 104, 132] and with jumps [14, 55, 68, 170, 205, 208], incorporating novel ideas and increased complexity such as regime switching volatility dynamics. We note that model-free bounds and requirements were also obtained in [78, 141, 140, 207], shedding light on the links between VIX and SPX and the difficulty of capturing them both simultaneously.

As we have seen in Section 1.2, getting rid of the restraining Markovian assumption that burdens classical stochastic volatility models has permitted the emergence of rough volatility models, which consistency with stylised facts under both the historical and pricing measures remains unmatched [8, 19, 26, 112, 123]. A large portion of the toolbox developed for Markovian diffusion models is not available any longer hence asymptotic methods play a prominent role in the literature [136, 151, 161, 162]. Since the fit of the spot implied volatility skew is extremely accurate under this class of models, it seems reasonable to expect good results when calibrating VIX options. Moreover, the newly established hedging formula of Viens and Zhang [232] shows that a rough volatility market is complete if it also contains a proxy of the volatility of the asset; this acts as an additional motivation for our work.

Still, [158] demonstrates that the rough Bergomi model is too close to lognormal to jointly calibrate both markets. Its younger sister, born in [153] with a stochastic vol-of-vol component, generates a smile sandwiched between the bid-ask prices when calibrating VIX, but the joint calibration is not provided. By incorporating a Zumbach effect, the quadratic rough Heston model of [124] achieves good results for the joint calibration at one given date. Further numerical methods were developed in [37, 39, 223]. However, the lack of analytical tractability of rough volatility models is holding back the progress of theoretical results on the VIX, with the notable exception of the LDPs from [101, 173] and the small-time asymptotics of [6].

In the latter, $\mathcal{F}_T$-measurable random variables (with volatility derivatives in mind) are written in the form of exponential martingales thanks to the Clark-Ocone formula, allowing the application of the established asymptotic methods from [8]. An expression for short-time limit at-the-money (ATM) implied volatility skew is derived, yielding an analytical criterion that a model should satisfy to reproduce the correct short-time behaviour. The proposed mixed rough Bergomi model does meet the requirement of positive skew of the VIX implied volatility, backing its implementation with theoretical evidence. And indeed, the fits are rather satisfying. This model is built by replacing the exponential kernels of Bergomi (4.1.1) ($t \mapsto \mathrm{e}^{-\kappa t}$) with fractional kernels of the type $t \mapsto t^{H-\frac{1}{2}}$ where $H \in (0, \frac{1}{2})$, but is limited to a single factor, i.e. $W^1 = W^2$. As a result, numerical computations under this model induce a linear smile, or equivalently a null curvature, which is inconsistent with market observations.

To remedy this, we remind ourselves of the early insights of Bergomi and Gatheral in favour of multi-factor models, and integrated by [77, 173] to rough volatility models. We therefore extend the results of [6] to the multi-factor case and also compute the short-time ATM curvature of the implied volatility, deriving a second criterion for a more accurate model choice. The main results of this chapter, the short-time limits of the implied volatility level, skew, and curvature are gathered in Section 4.3. Our framework covers a wide range of underlying assets; we apply these asymptotic results to VIX and stock options in Propositions 4.4.1 and 4.5.1 respectively. We further analyse in details the two-factor rough Bergomi model; under this model, the variance process reads

$$v_t = v_0 \left[\chi \mathcal{E}\left(\nu W_t^{1,H}\right) + (1-\chi)\mathcal{E}\left(\eta\left(\rho W_t^{1,H} + \sqrt{1-\rho^2} W_t^{2,H}\right)\right)\right], \quad t \geq 0,$$

where $\chi \in [0,1]$, $W_t^{i,H} := \int_0^t (t-s)^{H-\frac{1}{2}} \mathrm{d}W_s^i$, W^1, W^2 are independent Brownian motions, the Wick exponential is defined as $\mathcal{E}(X) := \exp\left(X - \frac{1}{2}\mathbb{E}[X^2]\right)$, and $\rho \in (-1,1)$ encodes the correlation. Unlike in the main introduction, the variance will be denoted by v in this chapter. Closed-form expressions that depend explicitly on the parameters of the model are provided in Proposition 4.4.3 for the VIX and Corollary 4.5.4 for the stock. They give insights on the interplay between the different parameters, and make the calibration task easier by allowing to fit some stylised facts prior to performing numerical computations. For instance, different combinations of parameters can yield positive or negative curvature.

The outline of the chapter is as follows. Section 4.2 presents our abstract framework and assumptions. The main results are displayed in Section 4.3; they are applied to the VIX case in Section 4.4 and the stock case in Section 4.5. All the proofs are gathered in the appendices, starting with useful lemmas and then following the order of the sections.

Notations We set an integer $N \in \mathbb{N}$. For a vector $\boldsymbol{x} \in \mathbb{R}^N$, we denote $|\boldsymbol{x}| := \sum_{i=1}^N x_i$ and $\|\boldsymbol{x}\|^2 := \sum_{i=1}^N x_i^2$. We fix a finite time horizon $T > 0$ and denote $\mathbb{T} := [0, T]$. For all $p \geq 1$, L^p stand for $L^p(\Omega)$.

4.2 Framework

We consider a square-integrable process $(A_t)_{t\in\mathbb{T}}$, adapted to the natural filtration $(\mathcal{F}_t)_{t\in\mathbb{T}}$ of an N-dimensional Brownian motion $\mathbf{W} = (W^1, ..., W^N)$ defined on a probability space $(\Omega, \mathcal{F}, \mathbb{P})$. We further introduce the true $(\mathcal{F}_t)_{t\in\mathbb{T}}$-martingale conditional expectation process

$$M_t := \mathbb{E}_t[A_T] := \mathbb{E}[A_T|\mathcal{F}_t], \qquad \text{for all } t \in \mathbb{T}.$$

The set $\mathbb{D}^{1,2}$ will denote the domain of the Malliavin derivative operator D with respect to the Brownian motion $\mathbf{W}$, while D^i indicates the Malliavin derivative operator with respect to W^i. It is well-known that $\mathbb{D}^{1,2}$ is a dense subset of $L^2(\Omega)$ and that D is a closed and unbounded operator from $L^2(\Omega)$ into $L^2(\mathbb{T}\times\Omega)$. We will also consider the iterated derivatives D^n, for $n \in \mathbb{N}$, whose domains will be denoted by $\mathbb{D}^{n,2}$. Analogously we define the sets of Malliavin differentiable processes $\mathbb{L}^{n,2} := L^2(\mathbb{T}; \mathbb{D}^{n,2})$. We refer to [201] for more details on Malliavin calculus. Assuming $A_T \in \mathbb{D}^{1,2}$, the Clark-Ocone formula [201, Theorem 1.3.14] reads, for each $t \in \mathbb{T}$,

$$M_t = \mathbb{E}[M_t] + (\mathbf{m} \bullet \mathbf{W})_t := \mathbb{E}[M_t] + \sum_{i=1}^N \int_0^t m_s^i \mathrm{d}W_s^i, \tag{4.2.1}$$

where each component of $\mathbf{m}$ is $m_s^i := \mathbb{E}\left[\mathrm{D}_s^i A_T | \mathcal{F}_s\right]$. Since M is a martingale, we may rewrite (4.2.1) as

$$M_t = M_0 + (M\boldsymbol{\phi} \bullet \mathbf{W})_t, \tag{4.2.2}$$

where $\boldsymbol{\phi}_s := \mathbf{m}_s / M_s$ is defined whenever $M_s \neq 0$ almost surely.

Furthermore, under the condition that $\boldsymbol{\phi} = (\phi^1, ..., \phi^N)$ belongs to $\mathbb{L}^{n,2}$, the following processes are well-defined for all $t < T$:

$$Y_t := \int_t^T \|\boldsymbol{\phi}_r\|^2 \,\mathrm{d}r, \qquad \mathfrak{u}_t := \sqrt{Y_t}, \qquad u_t := \frac{\mathfrak{u}_t}{\sqrt{T-t}}; \tag{4.2.3}$$

$$\Theta_t^i := \left(\int_t^T \mathrm{D}_t^i \|\boldsymbol{\phi}_r\|^2 \,\mathrm{d}r\right) \phi_t^i, \qquad \text{and} \qquad |\boldsymbol{\Theta}| := \sum_{i=1}^n \Theta^i. \tag{4.2.4}$$

Note that all the processes we introduced depend implicitly on T; this will be crucial when we study the short-time limit as T tends to zero.

4.2.1 Level, Skew and Curvature

Since M is a martingale process, we can use it as an underlying to introduce options. A standard practice is to work with its logarithm $\mathfrak{M} := \log(M)$, so that $\mathfrak{M}_T = \log \mathbb{E}_T[A_T] = \log(A_T)$ and $\mathfrak{M}_0 = \log \mathbb{E}[A_T]$. Under no-arbitrage arguments, the price V_t at time t of a European Call option with maturity T and log-strike $k \geq 0$ is equal to

$$V_t(k) := \mathbb{E}_t\left[\left(M_T - \mathrm{e}^k\right)^+\right] = \mathbb{E}_t\left[\left(A_T - \mathrm{e}^k\right)^+\right], \tag{4.2.5}$$

and the at-the-money value is denoted by $V_t := V_t(\mathfrak{M}_0) = \mathbb{E}_t[(A_T - M_t)^+]$. We adapt the usual definitions of at-the-money implied volatility Level, Skew and Curvature for the case where the underlying is a general process (later we will analyse VIX options and S&P options). Denote by $\mathrm{BS}(t,x,k,\sigma)$ the Black-Scholes price of a European Call option at time $t \in \mathbb{T}$, with maturity T, log-stock x, log-strike k and volatility σ. Its closed-form expression reads

$$\mathrm{BS}(t,x,k,\sigma) = \mathrm{e}^x \mathcal{N}(d_+(k,\sigma)) - \mathrm{e}^k \mathcal{N}(d_-(k,\sigma)), \quad \text{with} \quad d_\pm(k,\sigma) := \frac{x-k}{\sigma\sqrt{T-t}} \pm \frac{\sigma}{2}\sqrt{T-t},$$

where $\mathcal{N}$ denotes the cumulative distribution function of the standard Normal distribution.

Definition 4.2.1. For any $k \in \mathbb{R}$, the implied volatility $\sigma_{\mathrm{BS}}(T,k)$ is the unique non-negative solution to $V_0(k) = \mathrm{BS}\big(0, \mathfrak{M}_0, k, \sigma_{\mathrm{BS}}(T,k)\big)$, and we will omit the k-dependence when considering it at-the-money (i.e. $k = \mathfrak{M}_0$).

Remark 4.2.2. In the first two chapters, the first argument of BS was the time to maturity. In this chapter, it corresponds to the time at which the option is evaluated, with the maturity T being implicit.

Definition 4.2.3. The at-the-money implied skew $\mathcal{S}$ and curvature $\mathcal{C}$ at time zero are defined as

$$\mathcal{S}_T := \left|\partial_k \sigma_{\mathrm{BS}}(T,k)\right|_{k=\mathfrak{M}_0} \qquad \text{and} \qquad \mathcal{C}_T := \left|\partial_k^2 \sigma_{\mathrm{BS}}(T,k)\right|_{k=\mathfrak{M}_0},$$

where ∂_k and ∂_k^2 stand for $\partial/\partial k$ and $\partial^2/\partial k^2$ respectively.

4.2.2 Examples of underlyings covered by the framework

The general framework (4.2.2) encompasses a large class of models, in particular stochastic volatility models ubiquitous in quantitative finance. Consider the setup

$$\frac{dS_t}{S_t} = \sqrt{v_t}\,\mathrm{d}B_t = \sqrt{v_t}\sum_{i=1}^{N} \rho_i\,\mathrm{d}W_t^i,$$

for a stock price process $(S_t)_{t\in\mathbb{T}}$, where v is a stochastic process adapted to $(\mathcal{F}_t)_{t\in\mathbb{T}}$, $\boldsymbol{\rho} := (\rho_1, \cdots, \rho_N) \in [-1,1]^N$ and $\boldsymbol{\rho}^\top \boldsymbol{\rho} = 1$.

4.2.2.1 Asset price

For $N = 2$, the model (4.2.2) corresponds to a one-dimensional stochastic volatility model by identifying $A = M = S$, $\phi^1 = \rho_1\sqrt{v}$ and $\phi^2 = \rho_2\sqrt{v}$, and v is a stochastic process generated by W^1. Our analysis here generalises the approach in [8, Equation (2.1)] to the multi-factor case (in the continuous-path case). We refer to Section 4.5 for the details in the multi-factor setting and the analysis of the implied volatility.

4.2.2.2 VIX

The VIX process is defined as $\mathrm{VIX}_T = \sqrt{\frac{1}{\Delta}\int_T^{T+\Delta}\mathbb{E}_T[v_t]\,\mathrm{d}t}$. The representation (4.2.1) yields that the underlying is the VIX future

$$M_t^{\mathrm{VIX}} := \mathbb{E}_t[\mathrm{VIX}_T] = \mathbb{E}[\mathrm{VIX}_T] + (\mathbf{m} \bullet \mathbf{W})_t$$

where

$$m_s^i = \frac{1}{2\Delta}\mathbb{E}_s\left[\frac{1}{\mathrm{VIX}_T}\int_T^{T+\Delta} \mathrm{D}_s^i v_r \,\mathrm{d}r\right].$$

4.2.2.3 Average process

The average process is defined as $\mathrm{AV}_T = \frac{1}{T}\int_0^T S_t\,\mathrm{d}t$. Using (4.2.1) we find

$$M_t^{\mathrm{AV}} := \mathbb{E}_t[\mathrm{AV}_T] = \mathbb{E}[\mathrm{AV}_T] + (\mathbf{m} \bullet \mathbf{W})_T,$$

where $m_s^i = \frac{1}{T}\int_s^T \mathbb{E}_s[\mathrm{D}_s^i S_r]\,\mathrm{d}r$.

4.2.2.4 Multi-factor rough Bergomi model

Rough volatility models can be written as $v_t = f(\mathbf{W}_t^H)$, where $\mathbf{W}^H$ is a multidimensional fractional Brownian motion with correlated components and $f : \mathbb{R}^N \to \mathbb{R}$. For instance in the two-factor rough Bergomi model,

$$v_t = v_0\left(\chi \exp\left\{\nu W_t^{1,H} - \frac{\nu^2}{2}\mathbb{E}\left[\left(W_t^{1,H}\right)^2\right]\right\} + (1-\chi)\exp\left\{\eta W_t^{2,H} - \frac{\eta^2}{2}\mathbb{E}\left[\left(W_t^{2,H}\right)^2\right]\right\}\right),$$

with $\chi \in (0,1)$, $\nu, \eta, v_0 > 0$. In Example 4.2.2.2 we set $A = \mathrm{VIX}$ and hence $N = 2$, but in the asset price case we set $A = S$ and therefore $N = 3$ even though the variance only depends on two factors.

4.2.3 General assumptions

We introduce the following broad assumptions, key for the whole analysis, and provide in Section 4.4 sufficient conditions to simplify them in the VIX case.

$(\mathbf{H_1})$ $A \in \mathbb{L}^{4,p}$.

$(\mathbf{H_2})$ $\dfrac{1}{M_t} \in L^p$, for all $p > 1$, and all $t \in \mathbb{T}$.

$(\mathbf{H_3})$ The term $\mathbb{E}_t\left[\int_t^T \frac{|\boldsymbol{\Theta}_s|}{\mathfrak{u}_s^2}\,\mathrm{d}s\right]$ is well-defined for all $t \in \mathbb{T}$.

$(\mathbf{H_4})$ The term $\frac{1}{\sqrt{T}}\mathbb{E}\left[\int_0^T \frac{|\boldsymbol{\Theta}_s|}{\mathfrak{u}_s^2}\,\mathrm{d}s\right]$ tends to zero as T tends to zero.

$(\mathbf{H_5})$ There exists $p \geq 1$ such that $\sup_{T\in[0,1]} \mathfrak{u}_0^p < \infty$ almost surely and, for all random variables $Z \in L^p$ and all $i \in [\![1, N]\!]$, the terms

$$\int_0^T \mathbb{E}\left[Z\left(\mathbb{E}_s\left[\frac{1}{u_0}\int_0^T \mathrm{D}_s^i \left\|\boldsymbol{\phi}_r\right\|^2 \mathrm{d}r\right]\right)^2\right]\mathrm{d}s$$

are well defined and tend to zero as T tends to zero.

There exists $\lambda \in (-\frac{1}{2}, 0]$ such that:

$(\mathbf{H_6^\lambda})$ These expressions converge to zero as T tends to zero

$$\frac{1}{T^{\frac{1}{2}+\lambda}}\mathbb{E}\left[\int_0^T \frac{|\boldsymbol{\Theta}_s|\int_s^T|\boldsymbol{\Theta}_r|\,\mathrm{d}r}{\mathfrak{u}_s^6}\,\mathrm{d}s\right] \quad \text{and} \quad \frac{1}{T^{\frac{1}{2}+\lambda}}\mathbb{E}\left[\int_0^T \frac{1}{\mathfrak{u}_s^4}\sum_{k=1}^N\left\{\phi_s^k \mathrm{D}_s^k\left(\int_s^T |\boldsymbol{\Theta}_r|\,\mathrm{d}r\right)\right\}\mathrm{d}s\right].$$

$(\mathbf{H_7^\lambda})$ The process $\mathfrak{K}_T := \dfrac{\int_0^T|\boldsymbol{\Theta}_s|\,\mathrm{d}s}{T^{\frac{1}{2}+\lambda}\mathfrak{u}_0^3}$ is such that $\mathbb{E}[\mathfrak{u}_0^2\mathfrak{K}_T]$ tends to zero and $\mathbb{E}[\mathfrak{K}_T]$ has a finite limit, as T tends to zero.

There exists $\gamma \in (-1, 0]$ such that:

$(\mathbf{H_8^\gamma})$ The following expressions converge to zero as T tends to zero

$$\begin{aligned}
&\frac{1}{T^{\frac{1}{2}+\gamma}}\mathbb{E}\left[\int_0^T \mathfrak{u}_s^{-10}|\boldsymbol{\Theta}_s|\left(\int_s^T |\boldsymbol{\Theta}_r|\left(\int_r^T |\boldsymbol{\Theta}_y|\,\mathrm{d}y\right)\mathrm{d}r\right)\mathrm{d}s\right],\\
&\frac{1}{T^{\frac{1}{2}+\gamma}}\mathbb{E}\left[\int_0^T \mathfrak{u}_s^{-8}\sum_{j=1}^N\left(\phi_s^j\mathrm{D}_s^j\left(\int_s^T |\boldsymbol{\Theta}_r|\left(\int_r^T |\boldsymbol{\Theta}_y|\,\mathrm{d}y\right)\mathrm{d}r\right)\mathrm{d}s\right)\mathrm{d}s\right],\\
&\frac{1}{T^{\frac{1}{2}+\gamma}}\mathbb{E}\left[\int_0^T \mathfrak{u}_s^{-8}|\boldsymbol{\Theta}_s|\int_s^T\sum_{j=1}^N\left\{\phi_r^j\mathrm{D}_r^j\left(\int_r^T |\boldsymbol{\Theta}_y|\,\mathrm{d}y\right)\right\}\mathrm{d}r\,\mathrm{d}s\right],\\
&\frac{1}{T^{\frac{1}{2}+\gamma}}\mathbb{E}\left[\int_0^T \mathfrak{u}_s^{-6}\sum_{k=1}^N\left\{\phi_s^k\mathrm{D}_s^k\left(\int_s^T\sum_{j=1}^N\left\{\phi_r^j\,\mathrm{D}_r^j\left(\int_r^T |\boldsymbol{\Theta}_y|\,\mathrm{d}y\right)\right\}\mathrm{d}r\right)\right\}\mathrm{d}s\right].
\end{aligned}$$

$(\mathbf{H}_9^\gamma)$ The processes

$$\mathfrak{H}_T^1 := \frac{1}{T^{\frac{1}{2}+\gamma}\mathfrak{u}_0^7}\int_0^T |\mathbf{\Theta}_s| \left(\int_s^T |\mathbf{\Theta}_r|\,\mathrm{d}r\right)\mathrm{d}s \quad \text{and} \quad \mathfrak{H}_T^2 := \frac{1}{T^{\frac{1}{2}+\gamma}\mathfrak{u}_0^5}\int_0^T \sum_{j=1}^N \left\{\phi_s^j \mathrm{D}_s^j \left(\int_s^T |\mathbf{\Theta}_r|\,\mathrm{d}r\right)\right\}\mathrm{d}s,$$

are such that $\mathbb{E}\big[(\mathfrak{u}_0^6+\mathfrak{u}_0^4+\mathfrak{u}_0^2)\mathfrak{H}_T^1+(\mathfrak{u}_0^4+\mathfrak{u}_0^2)\mathfrak{H}_T^2\big]$ tend to zero and both $\mathbb{E}[\mathfrak{H}_T^1]$ and $\mathbb{E}[\mathfrak{H}_T^2]$ have a finite limit, as T tends to zero.

Remark 4.2.4. $(\mathbf{H_1})$ requires A to be four times Malliavin differentiable. This is necessary because we apply the Clark-Ocone formula (4.2.1) and three times the anticipative Itô's formula for the proof of the short-time curvature.

Remark 4.2.5. These assumptions can be replaced by direct conditions on ϕ when only considering the stock price as underlying martingale price process, as in [7, 8, 9].

4.3 Main results

We gather here our main asymptotic results for the general framework described above. The proofs are postponed to Appendix 4.6.2 in an attempt to ease the flow of the chapter. The first theorem states that the small-time limit of the implied volatility is equal to the limit of the forward volatility. This is a rather well-known result which was proved for Markovian stochastic volatility models in [9, 27] and in a one-factor setting in [6].

In order to streamline the call to the assumptions, we shall group them using mixed subscript notations, for example $(\mathbf{H_{123}})$ corresponds to $(\mathbf{H_1})$-$(\mathbf{H_2})$-$(\mathbf{H_3})$ and we further write $(\overline{\mathbf{H}}^\lambda)$ to mean $(\mathbf{H_{12345}})$-$(\mathbf{H}_{67}^\lambda)$ and $(\overline{\mathbf{H}}^{\lambda\gamma})$ as short for $(\mathbf{H_{12345}})$-$(\mathbf{H}_{67}^\lambda)$-$(\mathbf{H}_{89}^\gamma)$.

Theorem 4.3.1. *If* $(\mathbf{H_{12345}})$ *hold, then*

$$\lim_{T\downarrow 0}\Big(\sigma_{\mathrm{BS}}(T)-\mathbb{E}[u_0]\Big)=0.$$

Note that we did not assume the limit of $\mathbb{E}[u_0]$ to be finite. The proof, in Appendix 4.6.2.1, builds on arguments from [9, Proposition 3.1].

We then turn our attention to the ATM skew, defined in 4.2.3. This short-time asymptotic is reminiscent of [8, Proposition 6.2] and [6, Theorem 8].

Theorem 4.3.2. *If there exist $\lambda \in (-\frac{1}{2}, 0]$ such that $(\overline{\mathbf{H}}^{\lambda})$ are satisfied, then*

$$\lim_{T\downarrow 0} \frac{\mathcal{S}_T}{T^{\lambda}} = \frac{1}{2} \lim_{T\downarrow 0} \mathbb{E}\left[\frac{1}{T^{\frac{1}{2}+\lambda}} \frac{\int_0^T |\mathbf{\Theta}_s| \, \mathrm{d}s}{\mathfrak{u}_0^3} \right]. \tag{4.3.1}$$

Note that (4.3.1) still holds without $(\mathbf{H}_{\mathbf{7}}^{\lambda})$, but in that case both sides are infinite. In the stochastic volatility setting of Section 4.2.2.1 with $v_t = f(\mathbf{W}_t^H)$, λ corresponds to $H - 1/2$ such that (4.3.1) matches the slope of the observed ATM skew of SPX implied volatility, displayed in Figure 1.2. We prove this theorem in Appendix 4.6.2.2.

We also provide the short-term curvature, proved in Appendix 4.6.2.3.

Theorem 4.3.3. *If there exist $\lambda \in (-\frac{1}{2}, 0]$ and $\gamma \in (-1, \lambda]$ ensuring $(\overline{\mathbf{H}}^{\lambda\gamma})$, then*

$$\begin{aligned}\lim_{T\downarrow 0} \frac{\mathcal{C}_T}{T^{\gamma}} = \lim_{T\downarrow 0} \frac{1}{T^{\gamma}} \Bigg\{ \mathcal{S}_T - \frac{15}{2\sqrt{T}} \mathbb{E}\left[\frac{1}{\mathfrak{u}_0^7} \int_0^T |\mathbf{\Theta}_r| \left(\int_r^T |\mathbf{\Theta}_y| \, \mathrm{d}y \right) \mathrm{d}r \right] \\ + \frac{3}{2\sqrt{T}} \mathbb{E}\left[\frac{1}{\mathfrak{u}_0^5} \int_0^T \sum_{j=1}^N \left\{ \phi_s^j \mathrm{D}_s^j \left(\int_s^T |\mathbf{\Theta}_y| \, \mathrm{d}y \right) \right\} \mathrm{d}s \right] \Bigg\}.\end{aligned} \tag{4.3.2}$$

The limit still holds without $(\mathbf{H}_{\mathbf{9}}^{\gamma})$ but in that case the second and third term are infinite. Similarly, $(\mathbf{H}_{\mathbf{7}}^{\lambda})$ with $\lambda \geq \gamma$ guarantees that the first term $T^{-\gamma}\mathcal{S}_T$ converges. By Theorem 4.3.2,

$$\lim_{T\downarrow 0} T^{-\gamma} \mathcal{S}_T = \begin{cases} 0, & \text{if } \lambda > \gamma, \\ \dfrac{1}{2} \lim_{T\downarrow 0} \mathbb{E}\left[\dfrac{1}{T^{\frac{1}{2}+\lambda}} \dfrac{\int_0^T |\mathbf{\Theta}_s| \, \mathrm{d}s}{\mathfrak{u}_0^3} \right], & \text{if } \lambda = \gamma, \\ +\infty, & \text{if } \lambda < \gamma. \end{cases}$$

4.4 Asymptotic results in the VIX case

As advertised, our framework includes the case where

$$A_T = \mathrm{VIX}_T = \sqrt{\frac{1}{\Delta} \int_T^{T+\Delta} \mathbb{E}_T[v_r] \, \mathrm{d}r},$$

for $v_r \in \mathbb{D}^{3,2}$ for all $r \in [0, T+\Delta]$, and we provide simple sufficient conditions for $(\overline{\mathbf{H}}^{\lambda\gamma})$ to be satisfied in this case.

4.4.1 A generic volatility model

Let us consider the following four conditions which we will gather under the notation $(\overline{\mathbf{C}})$. There exists $X \in L^p$ for all $p > 1$ such that for any $T > 0$:

$(\mathbf{C_1})$ For all $t \geq 0$, $\frac{1}{M_t} \leq X$ almost surely;

$(\mathbf{C_2})$ For all $i, j, k \in [\![1, N]\!]$ and $t \leq s \leq y \leq T \leq r$, we have, almost surely

- $v_r \leq X$,
- $\mathrm{D}^i_y v_r \leq X(r-y)^{H-\frac{1}{2}}$,
- $\mathrm{D}^j_s \mathrm{D}^i_y v_r \leq X(r-s)^{H-\frac{1}{2}}(r-y)^{H-\frac{1}{2}}$,
- $\mathrm{D}^k_t \mathrm{D}^j_s \mathrm{D}^i_y v_r \leq X(r-t)^{H-\frac{1}{2}}(r-s)^{H-\frac{1}{2}}(r-y)^{H-\frac{1}{2}}$;

$(\mathbf{C_3})$ For all $p > 1$, $\mathbb{E}[u_s^{-p}]$ is uniformly bounded in s and T, with $s \leq T$.

$(\mathbf{C_4})$ For all $i, j, k \in [\![1, N]\!]$ and $r \geq 0$, the mappings $y \mapsto \mathrm{D}^i_y v_r$, $s \mapsto \mathrm{D}^j_s \mathrm{D}^i_y v_r$, and $t \mapsto \mathrm{D}^k_t \mathrm{D}^j_s \mathrm{D}^i_y v_r$ are almost surely continuous in a neighbourhood zero.

We compute level, skew, and curvature of the VIX implied volatility in a model which satisfies the sufficient conditions. Let us define the following limits

$$J_i := \int_0^\Delta \mathbb{E}[\mathrm{D}^i_0 v_r]\, \mathrm{d}r, \qquad G_{ij} := \int_0^\Delta \mathbb{E}\big[\mathrm{D}^j_0 \mathrm{D}^i_0 v_r\big]\, \mathrm{d}r, \quad \text{for all } i, j \in [\![1, N]\!]. \tag{4.4.1}$$

Proposition 4.4.1. *Let $(\overline{\mathbf{C}})$ hold. For $H \in (0, \frac{1}{6})$, the following limits hold*

$$\begin{aligned}
\lim_{T\downarrow 0} \sigma_{\mathrm{BS}}(T) &= \frac{\|\boldsymbol{J}\|}{2\Delta \mathrm{VIX}_0^2},\\
\lim_{T\downarrow 0} \mathcal{S}_T &= \frac{\sum_{i,j=1}^N J_i J_j \left(G_{ij} - \frac{J_i J_j}{\Delta \mathrm{VIX}_0^2}\right)}{2\,\|\boldsymbol{J}\|^3},\\
\lim_{T\downarrow 0} \frac{\mathcal{C}_T}{T^{3H-\frac{1}{2}}} &= \frac{2\Delta \mathrm{VIX}_0^2}{3\,\|\boldsymbol{J}\|^5} \sum_{i,j,k=1}^N J^i J^j J^k \lim_{T\downarrow 0} \frac{\int_T^{T+\Delta} \mathbb{E}\left[\mathrm{D}^k_0 \mathrm{D}^j_0 \mathrm{D}^i_0 v_r\right]\, \mathrm{d}r}{T^{3H-\frac{1}{2}}}.
\end{aligned} \tag{4.4.2}$$

The first two limits still hold for any $H \in (0, \frac{1}{2})$.

We split the proof in two steps, collected in Appendix 4.6.3. In the first one, we show that $(\mathbf{C_1})$, $(\mathbf{C_2})$, $(\mathbf{C_3})$ are sufficient conditions to apply our main theorems because they imply $(\overline{\mathbf{H}}^{\lambda\gamma})$ for any $\lambda \in (-1/2, 0]$ and $\gamma \in (-1, 3H - 1/2]$. Thanks to $(\mathbf{C_4})$ we can also compute the limits—after a rigorous statement of convergence results—starting with σ_{BS} and the skew with $\lambda = 0$. Restricting H to the range $(0, 1/6)$, which is the most relevant regime for rough volatility models, we can set $\gamma = 3H - 1/2 < \lambda$ and compute the short-time curvature, with only the second term in $(\mathbf{H_9^\gamma})$ contributing to the limit. A different choice of H would lead to another γ and therefore different quantities would appear in the curvature formula. The limit in (4.4.2) is finite by the last point of $(\mathbf{C_2})$.

4.4.2 The two-factor rough Bergomi model

We consider the following two-factor exponential model

$$v_t = v_0 \left[\chi \mathcal{E}\left(\nu W_t^{1,H}\right) + \overline{\chi}\mathcal{E}\left(\eta\left(\rho W_t^{1,H} + \overline{\rho} W_t^{2,H}\right)\right)\right] =: v_0(\chi\mathcal{E}_t^1 + \overline{\chi}\mathcal{E}_t^2), \tag{4.4.3}$$

where $H \in (0, \frac{1}{2}]$, $W_t^{i,H} = \int_0^t (t-s)^{H-\frac{1}{2}} \mathrm{d}W_s^i$, W^1, W^2 are independent Brownian motions, the Wick exponential is defined as $\mathcal{E}(X) := \exp\left(X - \frac{1}{2}\mathbb{E}[X^2]\right)$, for any random variable X, and $\chi \in [0,1]$, $\overline{\chi} = 1 - \chi$, $v_0, \nu, \eta > 0$, $\rho \in [-1,1]$, $\overline{\rho} = \sqrt{1-\rho^2}$. This model is an extension of the Bergomi model [31], where we replace the exponential kernel with a fractional one and an extension of the rough Bergomi model [19] to the two-factor case. It combines Bergomi's insights on the need for several factors with the benefits of rough volatility.

Proposition 4.4.2. *The model* (4.4.3) *satisfies conditions* $(\overline{\mathbf{C}})$.

Proof. See Appendix 4.6.4.1. □

We can therefore apply Proposition 4.4.1 to this model and obtain the following closed-form limits. The proof is postponed to Appendix 4.6.4.2.

Proposition 4.4.3. *Let* $\psi(\rho,\nu,\eta,\chi) := \sqrt{(\chi\nu + \overline{\chi}\eta\rho)^2 + \overline{\chi}^2\eta^2\overline{\rho}^2}$. *For* $H \in (0, 1/6)$, *we have the following limits:*

$$\lim_{T\downarrow 0} \sigma_{\mathrm{BS}}(T) = \frac{\Delta^{H-\frac{1}{2}}}{2H+1}\psi(\rho,\nu,\eta,\chi),$$

$$\begin{aligned}\lim_{T\downarrow 0} \mathcal{S}_T = \frac{(H+\frac{1}{2})\Delta^{H-\frac{1}{2}}}{2\psi(\rho,\nu,\eta,\chi)^3} \Bigg\{ &(\chi\nu + \overline{\chi}\eta\rho)^2 \left[\frac{\chi\nu^2 + \overline{\chi}\eta^2\rho^2}{2H} - \left(\frac{\chi\nu + \overline{\chi}\eta\rho}{H + \frac{1}{2}}\right)^2\right] \\ &+ 2(\chi\nu + \overline{\chi}\eta\rho)\overline{\chi}^2\eta^2\overline{\rho}^2\left[\frac{\eta\rho}{2H} - \frac{\nu+\eta\rho}{(H+\frac{1}{2})^2}\right] + \overline{\chi}^3\eta^4\overline{\rho}^4\left(\frac{1}{2H} - \frac{1}{(H+\frac{1}{2})^2}\right)\Bigg\},\end{aligned}$$

$$\lim_{T\downarrow 0}\frac{\mathcal{C}_T}{T^{3H-\frac{1}{2}}} = \frac{128\Delta^{-2H}(H+\frac{1}{2})^2}{3\psi(\rho,\nu,\eta,\chi)^5(1-6H)}\Big\{(\chi\nu+\overline{\chi}\eta\rho)^3(\chi\nu^3+\overline{\chi}\eta^3\rho^3) + 3(\chi\nu+\overline{\chi}\eta\rho)^2\overline{\chi}^2\eta^4\overline{\rho}^2\rho^2 + 3(\chi\nu+\overline{\chi}\eta\rho)\overline{\chi}^3\eta^5\overline{\rho}^4\rho + \overline{\chi}^4\eta^6\overline{\rho}^6\Big\}.$$

The limits depend explicitly on the parameters of the model (H,χ,ν,η,ρ) and can be used to gain insight on their impact over the quantities of interest.

Remark 4.4.4. In the case $\rho = 1$ (hence $\overline{\rho} = 0$) we have the following simplifications:

$$\lim_{T\downarrow 0}\sigma_{\mathrm{BS}}(T) = \frac{\Delta^{H-\frac{1}{2}}}{2H+1}\left|\chi\nu+\overline{\chi}\eta\right|, \quad \lim_{T\downarrow 0} T^{\frac{1}{2}-3H}\mathcal{C}_T = \frac{128\Delta^{-2H}(H+\frac{1}{2})^2}{3-24H}\frac{\chi\nu^3+\overline{\chi}\eta^3}{(\chi\nu+\overline{\chi}\eta)^2},$$
$$\lim_{T\downarrow 0}\mathcal{S}_T = \frac{1}{2}\frac{(H+\frac{1}{2})\Delta^{H-\frac{1}{2}}}{\chi\nu+\overline{\chi}\eta}\left[\frac{\chi\nu^2+\overline{\chi}\eta^2}{2H} - \left(\frac{\chi\nu+\overline{\chi}\eta}{H+\frac{1}{2}}\right)^2\right].$$

The case $\rho = -1$ follows by replacing η with $-\eta$. If we set $\overline{\rho} = 1$ (hence $\rho = 0$) we get

$$\lim_{T\downarrow 0}\sigma_{\mathrm{BS}}(T) = \frac{\Delta^{H-\frac{1}{2}}}{2H+1}\sqrt{\chi^2\nu^2+\overline{\chi}^2\eta^2}, \quad \lim_{T\downarrow 0} T^{\frac{1}{2}-3H}\mathcal{C}_T = \frac{128\Delta^{-2H}(H+\frac{1}{2})^2}{3-24H}\frac{\chi^4\nu^6+\overline{\chi}^4\eta^6}{(\chi^2\nu^2+\overline{\chi}^2\eta^2)^{5/2}}.$$

4.5 The stock smile under multi-factor models

We use the setting of Section 4.2.2.1 to apply our results to an asset price of the form

$$S_t = S_0 + \int_0^t S_r\sqrt{v_r}\,\mathrm{d}B_r,$$

where B is correlated with the other N Brownian motions as $B = \sum_{i=1}^N \rho_i W^i$ with $\sum_{i=1}^N \rho_i^2 = 1$, $\rho_i \in [-1,1]$ for all $i \in [\![1,N]\!]$. The volatility is a function of $(N-1)$ Brownian motions, such that the stock price features one additional and independent source of randomness.

To fit this model into the framework of (4.2.2) we set $A = S$ and identify ϕ^i with $\rho_i\sqrt{v}$. We modify slightly the notations to differentiate from the VIX framework; the implied volatility is denoted $\widehat{\sigma}_{\mathrm{BS}}(T)$ and the skew $\widehat{\mathcal{S}}_T$. We do not consider the curvature in this setting, by lack of an explicit formula. The proof of this Proposition and the following Corollary are postponed to Appendix 4.6.5.

Proposition 4.5.1. *Assume that there exists $H \in (0,\frac{1}{2})$ and $X \in L^p$ for all $p \geq 1$ such that, for all $0 \leq s \leq y$, $j \in [\![1,N]\!]$, and $p > 1$,* *(i)* $v_s \le X$ *;*

(ii) $\mathrm{D}_s^j v_y \le X(y-s)^{H-\frac{1}{2}}$*;*

(iii) $\sup_{s\le T}\mathbb{E}[u_s^{-p}] < \infty$*;*

(iv) $\limsup_{T\downarrow 0}\mathbb{E}\left[(\sqrt{v_T/v_0}-1)^2\right] = 0$.

Then the short-time limits of the implied volatility and skew are

$$\begin{cases} \lim\limits_{T\downarrow 0}\widehat{\sigma}_{\mathrm{BS}}(T) &= \sqrt{v_0}; \\ \lim\limits_{T\downarrow 0}\dfrac{\widehat{\mathcal{S}}_T}{T^{H-\frac{1}{2}}} &= \displaystyle\sum_{j=1}^{N}\frac{\rho_j}{2v_0}\lim_{T\downarrow 0}\frac{\int_0^T\int_s^T\mathbb{E}\left[\mathrm{D}_s^j v_y\right]\mathrm{d}y\,\mathrm{d}s}{T^{H+\frac{1}{2}}}. \end{cases}$$

Remark 4.5.2. The second limit is finite because of condition (ii).

Remark 4.5.3. The one-dimensional version ($N=2$) agrees with [8, Theorem 6.3] up to the sign because they derive with respect to the spot x and not the log-strike k.

In the two-factor rough Bergomi model (4.4.3) we can compute the short-time skew more explicitly. We recall from Example 4.2.2.4 that, for all $t\ge 0$, it entails setting $N=3$ and defining

$$\begin{cases} S_t &= S_0 + \displaystyle\int_0^t S_r\sqrt{v_r}\,\mathrm{d}B_r, \\ v_t &= v_0\left[\chi\mathcal{E}(\nu W_t^{1,H}) + \overline{\chi}\mathcal{E}\big(\eta(\rho W_t^{1,H}+\overline{\rho}W_t^{2,H})\big)\right], \end{cases}$$

where $W_t^{i,H} = \int_0^t (t-s)^{H-\frac{1}{2}}\,\mathrm{d}W_s^i$, for $i=1,2$ and $B=\sum_{i=1}^3\rho_i W^i$, with W^1,W^2,W^3 being independent Brownian motions. Hence W^3 only influences the asset price but not the variance.

Corollary 4.5.4. *In the two-factor rough Bergomi model we have the short-time skew limit*

$$\lim_{T\downarrow 0} T^{\frac{1}{2}-H}\widehat{\mathcal{S}}_T = \frac{\rho_1\chi\nu + \eta\overline{\chi}(\rho_1\rho+\rho_2\overline{\rho})}{(2H+1)(H+\frac{3}{2})}. \tag{4.5.1}$$

4.5.1 Tips for joint calibration in the two-factor rough Bergomi model

Assuming we can observe the short-time limits of the spot ATM implied volatility, it grants us v_0 for free, while the slope of its skew gives us H by (4.5.1).

Next, we simplify the expressions from Proposition 4.4.3 in the case $\chi = \frac{1}{2}$. Call I_0, $\mathcal{S}_0$ and $\mathcal{C}_0$ the three limits of Proposition 4.4.3 and denote $H_\pm := H \pm \frac{1}{2}$, $\alpha := \eta\rho$, $\beta := \eta\overline{\rho}$. Introduce further the normalised parameters

$$\widetilde{\alpha} := \frac{\alpha}{\nu}, \qquad \widetilde{\beta} := \frac{\beta}{\nu},$$

so that, denoting $\widetilde{\psi}(\widetilde{\alpha}, \widetilde{\beta}) := \sqrt{(1+\widetilde{\alpha})^2 + \widetilde{\beta}^2}$, we have

$$\begin{aligned}
I_0 &= \frac{\Delta^{H_-}}{4H_+}\sqrt{(\nu+\alpha)^2+\beta^2} = \frac{\nu\Delta^{H_-}}{4H_+}\sqrt{(1+\widetilde{\alpha})^2+\widetilde{\beta}^2} =: \nu C_I \widetilde{\psi}(\widetilde{\alpha}, \widetilde{\beta}),\\
\mathcal{S}_0 &= \frac{H_+\Delta^{H_-}}{2}\frac{(\nu+\alpha)^2\left[\frac{\nu^2+\alpha^2}{2H} - \left(\frac{\nu+\alpha}{H_+}\right)^2\right] + 2(\nu+\alpha)\beta^2\left[\frac{\alpha}{2H} - \frac{\nu+\alpha}{H_+}\right] + \alpha^4\left[\frac{1}{2H} - \frac{1}{H_+^2}\right]}{\left((\nu+\alpha)^2+\beta^2\right)^{3/2}}\\
&= \frac{\nu H_+\Delta^{H_-}}{2}\frac{(1+\widetilde{\alpha})^2\left[\frac{1+\widetilde{\alpha}^2}{2H} - \left(\frac{1+\widetilde{\alpha}}{H_+}\right)^2\right] + 2(1+\widetilde{\alpha})\widetilde{\beta}^2\left[\frac{\widetilde{\alpha}}{2H} - \frac{1+\widetilde{\alpha}}{H_+}\right] + \widetilde{\alpha}^4\left[\frac{1}{2H} - \frac{1}{H_+^2}\right]}{\left((1+\widetilde{\alpha})^2+\widetilde{\beta}^2\right)^{3/2}}\\
&=: \nu C_{\mathcal{S}}\frac{\Phi_{\mathcal{S}}(\widetilde{\alpha}, \widetilde{\beta})}{\widetilde{\psi}(\widetilde{\alpha}, \widetilde{\beta})^3},\\
\mathcal{C}_0 &= \frac{128\Delta^{-2H}H_+^2}{3\left((\nu+\alpha)^2+\beta^2\right)^{5/2}(1-6H)}\left\{(\nu+\alpha)^3(\nu^3+\alpha^3) + 3(\nu+\alpha)^2\alpha^2\beta^2 + 3(\nu+\alpha)\alpha\beta^4 + \beta^6\right\}\\
&= \frac{128\nu\Delta^{-2H}H_+^2}{3\left((1+\widetilde{\alpha})^2+\widetilde{\beta}^2\right)^{5/2}(1-6H)}\left\{(1+\widetilde{\alpha})^3(1+\widetilde{\alpha}^3) + 3(1+\widetilde{\alpha})^2\widetilde{\alpha}^2\widetilde{\beta}^2 + 3(1+\widetilde{\alpha})\widetilde{\alpha}\widetilde{\beta}^4 + \widetilde{\beta}^6\right\}\\
&=: \nu C_{\mathcal{C}}\frac{\Phi_{\mathcal{C}}(\widetilde{\alpha}, \widetilde{\beta})}{\widetilde{\psi}(\widetilde{\alpha}, \widetilde{\beta})^5},
\end{aligned}$$

where the constants $C_I, C_{\mathcal{S}}, C_{\mathcal{C}}$ only depend on Δ and H. Provided one can observe an approximation of these three limits, one can numerically solve for $\nu, \widetilde{\alpha}, \widetilde{\beta}$ in a system with three equations. Alternatively, since the three quantities have a factor ν, any quotient of two of them is a function of only $\widetilde{\alpha}, \widetilde{\beta}$, which we can plot and match to the observed number.

Both methods allow us to deduce $\nu, \widetilde{\alpha}, \widetilde{\beta}$ which also entail the values of η and ρ. Finally, we are left with ρ_1 and ρ_2 to play with such that the right-hand-side of (4.5.1) matches the market observations.

4.6 Proofs

4.6.1 Useful results

We start by adapting to the multivariate case a well-known decomposition formula and then prove a lemma which will be used extensively in the rest of the proofs. Both proofs build on the multidimensional anticipative Itô formula [201, Theorem 3.2.4].

Proposition 4.6.1 (Price decomposition). *Under* $(\mathbf{H_{123}})$*, the following decomposition formula holds, for all* $t \in \mathbb{T}$*, for the at-the-money price* (4.2.5)*, with* u_t *defined in* (4.2.3) *and* $G := (\partial_x^2 - \partial_x)\mathrm{BS}$*,*

$$V_t = \mathbb{E}_t\left[\mathrm{BS}(t, \mathfrak{M}_t, \mathfrak{M}_0, u_t)\right] + \frac{1}{2}\mathbb{E}_t\left[\int_t^T \partial_x G(s, \mathfrak{M}_s, \mathfrak{M}_0, u_s)|\boldsymbol{\Theta}_s|\,\mathrm{d}s\right].$$

Proof. Note that $V_T = \mathrm{BS}(T, \mathfrak{M}_T, \mathfrak{M}_0, u_T)$, hence $V_t = \mathbb{E}_t\left[\mathrm{BS}(T, \mathfrak{M}_T, \mathfrak{M}_0, u_T)\right]$ by no-arbitrage arguments. Define $\widehat{\mathrm{BS}}(t, x, k, \theta^2(T-t)) := \mathrm{BS}(t, x, k, \theta)$ and write for simplicity $\widehat{\mathrm{BS}}_t := \widehat{\mathrm{BS}}\left(t, \mathfrak{M}_t, \mathfrak{M}_0, Y_t\right) = \mathrm{BS}\left(t, \mathfrak{M}_t, \mathfrak{M}_0, u_t\right)$, where we recall that $Y_t = u_t^2(T-t)$. Thanks to $(\mathbf{H_1})$ and $(\mathbf{H_2})$, we then apply a multidimensional anticipative Itô's formula [201, Theorem 3.2.4] with respect to $(t, \mathfrak{M}, Y)$:

$$\begin{aligned}
\mathrm{BS}(T, \mathfrak{M}_T, \mathfrak{M}_0, u_T) = \widehat{\mathrm{BS}}_T = \widehat{\mathrm{BS}}_t &+ \int_t^T \partial_s\widehat{\mathrm{BS}}_s\,\mathrm{d}s + \int_t^T \partial_x\widehat{\mathrm{BS}}_s\left(\mathrm{d}(\boldsymbol{\phi}\bullet\mathbf{W})_s - \frac{1}{2}\left\|\boldsymbol{\phi}_s\right\|^2\mathrm{d}s\right) \\
&- \int_t^T \partial_y\widehat{\mathrm{BS}}_s\left\|\boldsymbol{\phi}_s\right\|^2\mathrm{d}s + \frac{1}{2}\int_t^T \partial_x^2\widehat{\mathrm{BS}}_s\left\|\boldsymbol{\phi}_s\right\|^2\mathrm{d}s + \int_t^T \partial_{xy}\widehat{\mathrm{BS}}_s|\boldsymbol{\Theta}_s|\,\mathrm{d}s.
\end{aligned}$$

with $\boldsymbol{\Theta}$ in (4.2.3). Derivatives of the Black-Scholes price (omitting the argument for simplicity) yield

$$\partial_s\widehat{\mathrm{BS}}_s = \partial_s\mathrm{BS}_s + \frac{u_s\partial_u\mathrm{BS}}{2(T-s)}, \qquad \partial_y\widehat{\mathrm{BS}}_s = \frac{\partial_u\mathrm{BS}}{2u_s(T-s)}, \qquad \text{and} \qquad G = \frac{\partial_u\mathrm{BS}}{u_s(T-s)}.$$

Putting everything together, using the Gamma-Vega-Delta relation

$$\frac{\partial_\sigma\mathrm{BS}(t,x,k,\sigma)}{\sigma(T-t)} = (\partial_x^2 - \partial_x)\mathrm{BS}(t,x,k,\sigma), \tag{4.6.1}$$

and applying the conditional expectation, one obtains

$$V_t = \mathbb{E}_t\left[\mathrm{BS}(t, \mathfrak{M}_t, \mathfrak{M}_0, u_t)\right] + \mathbb{E}_t\left[\int_t^T \mathcal{L}_{\mathrm{BS}}(s, u_s)\,\mathrm{d}s\right] + \mathbb{E}_t\left[\int_t^T \frac{\partial_{xu}\mathrm{BS}(s, \mathfrak{M}_s, \mathfrak{M}_0, u_s)}{2u_s(T-s)}|\boldsymbol{\Theta}_s|\,\mathrm{d}s\right], \tag{4.6.2}$$

where $\mathcal{L}_{\mathrm{BS}}(s,u_s) := \left[\frac{u_s^2}{2}\left(\partial_x^2-\partial_x\right)+\partial_s\right]\mathrm{BS}(s,\mathfrak{M}_s,\mathfrak{M}_0,u_s)$ is the Black-Scholes operator applied to the Black-Scholes function. Since $\mathcal{L}_{\mathrm{BS}}(s,u_s)=0$ by construction and

$$\partial_x G(s,x,k,\sigma) = \frac{\mathrm{e}^x \mathcal{N}'(d_+(k,\sigma))}{\sigma\sqrt{T-s}}\left(1-\frac{d_+(k,\sigma)}{\sigma\sqrt{T-s}}\right),$$

the last term in (4.6.2) is well defined by $(\mathbf{H_3})$, and the proposition follows. $\square$

Lemma 4.6.2. *For all $t\in\mathbb{T}$, let $J_t := \int_t^T a_s\,\mathrm{d}s$, for some adapted process $a\in\mathbb{L}^{1,2}$, and $\mathfrak{L} := \sum_{i=1}^n c_i\partial_x^i$ be a linear combination of partial derivatives, with weights $c_i\in\mathbb{R}$. Then, writing for clarity $\mathrm{BS}_t := \mathrm{BS}(t,\mathfrak{M}_t,\mathfrak{M}_0,u_t)$, we have*

$$\mathbb{E}\left[\int_0^T \mathfrak{L}\mathrm{BS}_s a_s\,\mathrm{d}s\right] = \mathbb{E}\left[\mathfrak{L}\mathrm{BS}_0 J_0 + \int_0^T\left(\partial_x^3-\partial_x^2\right)\mathfrak{L}\mathrm{BS}_s|\boldsymbol{\Theta}_s|J_s\,\mathrm{d}s + \int_0^T \partial_x\mathfrak{L}\mathrm{BS}_s\sum_{k=1}^N\left(\phi_s^k \mathrm{D}_s^k J_s\right)\,\mathrm{d}s\right]. \tag{4.6.3}$$

Remark 4.6.3. We will use this lemma freely in the proofs with the justification that the condition $a\in\mathbb{L}^{1,2}$ will always be satisfied thanks to $(\mathbf{H_1})$.

Proof. As in the proof of Proposition 4.6.1, we define $\widehat{\mathrm{BS}}(t,x,k,\theta^2(T-t)) := \mathrm{BS}(t,x,k,\theta)$ and write for simplicity $\widehat{\mathrm{BS}}_t := \widehat{\mathrm{BS}}\left(t,\mathfrak{M}_t,\mathfrak{M}_0,Y_t\right) = \mathrm{BS}\left(t,\mathfrak{M}_t,\mathfrak{M}_0,u_t\right)$.

Let us define $\widehat{\mathrm{P}}(t,x,k,y,j) := \mathfrak{L}\widehat{\mathrm{BS}}(t,x,k,u)j$, and denote $\widehat{\mathrm{P}}\left(t,\mathfrak{M}_s,\mathfrak{M}_0,u_s,J_s\right)$ by $\widehat{\mathrm{P}}_t$ for simplicity. We then apply the multidimensional anticipative Itô's formula [201, Theorem 3.2.4] with respect to $(t,\mathfrak{M},Y,J)$:

$$\begin{aligned}\widehat{\mathrm{P}}_T =& \widehat{\mathrm{P}}_0 + \int_0^T \partial_s\widehat{\mathrm{P}}_s\,\mathrm{d}s + \int_0^T \partial_x\widehat{\mathrm{P}}_s\left(\mathrm{d}(\boldsymbol{\phi}\bullet\mathbf{W})_s - \frac{1}{2}\|\boldsymbol{\phi}_s\|^2\,\mathrm{d}s\right) - \int_t^T \partial_y\widehat{\mathrm{P}}_s\|\boldsymbol{\phi}_s\|^2\,\mathrm{d}s\\ &+ \frac{1}{2}\int_t^T \partial_x^2\widehat{\mathrm{P}}_s\|\boldsymbol{\phi}_s\|^2\,\mathrm{d}s + \int_t^T \partial_{xy}\widehat{\mathrm{P}}_s|\boldsymbol{\Theta}_s|\,\mathrm{d}s + \int_0^T \partial_j\widehat{\mathrm{P}}_s\,\mathrm{d}J_s + \int_0^T \partial_{xj}\widehat{\mathrm{P}}_s\sum_{k=1}^N\left(\phi_s^k\mathrm{D}_s^kJ_s\right)\,\mathrm{d}s.\end{aligned}$$

One first notices that $\widehat{\mathrm{P}}_0 = \mathfrak{L}\widehat{\mathrm{BS}}_0 J_0$ and $\widehat{\mathrm{P}}_T = 0$. Moreover we observe that $\int_0^T \partial_j\widehat{\mathrm{P}}_s\,\mathrm{d}J_s = -\int_0^T \mathfrak{L}\widehat{\mathrm{BS}}_s a_s\,\mathrm{d}s$, which corresponds to the left-hand-side of (4.6.3), and

$$\int_0^T \partial_{xj}\widehat{\mathrm{P}}_s\sum_{k=1}^N\left(\phi_s^k\mathrm{D}_s^kJ_s\right)\,\mathrm{d}s = \int_0^T \partial_x\mathfrak{L}\widehat{\mathrm{BS}}_s\sum_{k=1}^N\left(\phi_s^k\,\mathrm{D}_s^kJ_s\right)\,\mathrm{d}s.$$

Since $\mathfrak{L}$ is a linear operator the partial derivatives in s, x and u cancel as in the proof of

Proposition 4.6.1. That means we are left with

$$\int_0^T \mathfrak{L}\widehat{\mathrm{BS}}_s a_s \,\mathrm{d}s = \mathfrak{L}\widehat{\mathrm{BS}}_0 J_0 + \int_0^T \left(\partial_x^3 - \partial_x^2\right) \mathfrak{L}\widehat{\mathrm{BS}}_s |\boldsymbol{\Theta}_s| J_s \,\mathrm{d}s + \int_0^T \partial_x \widehat{\mathrm{P}}_s \mathrm{d}(\boldsymbol{\phi} \bullet \mathbf{W})_s$$
$$+ \int_0^T \partial_x \mathfrak{L}\widehat{\mathrm{BS}}_s \sum_{k=1}^N \left(\phi_s^k \, \mathrm{D}_s^k J_s\right) \mathrm{d}s.$$

Since $\partial_x^n \mathrm{BS}(s,x,u) = \partial_x^n \widehat{\mathrm{BS}}(s,x,u^2(T-s))$ for any $n \in \mathbb{N}$, summing everything and taking expectations yields the claim. □

We also adapt and clarify [8, Lemma 4.1] to our purposes, yielding a convenient bound for the partial derivatives of G.

Proposition 4.6.4. *Let us fix $t \in \mathbb{T}$ (we will omit this argument for clarity), and write σ instead of $\sigma\sqrt{T-t}$. For any $n \in \mathbb{N}$ and $p \in \mathbb{R}$, there exists a strictly positive constant $C_{n,p}$ independent of x and σ such that, for all $\sigma > 0$ and $x \in \mathbb{R} \setminus \left\{0, \frac{\sigma^2}{2}\right\}$,*

$$\partial_x^n G(x,k,\sigma) \le \frac{C_{n,p}\,\mathrm{e}^k}{\sigma^{p+1}}. \tag{4.6.4}$$

If $x = 0$, then for any $n \in \mathbb{N}$ the bound (4.6.4) holds with $p = n$.

If $x = \frac{\sigma^2}{2}$, there exists a strictly positive constant C_n independent of σ such that

$$\partial_x^n G\left(\frac{\sigma^2}{2}, k, \sigma\right) = \begin{cases} \dfrac{C_n\,\mathrm{e}^k}{\sigma^{n+1}}, & \text{if } n \text{ is even,} \\ 0, & \text{if } n \text{ is odd.} \end{cases}$$

Furthermore, taking derivatives with respect to k instead of x does not alter the form of the right-hand-side bounds.

One could simplify and extend the statement as follows.

Corollary 4.6.5. *For any $n \in \mathbb{N}$, there exists a non-negative $C_{n,k}$ independent of x and σ such that, for all $\sigma > 0$ and $x \in \mathbb{R}$,*

$$|\partial_x^n G(x,k,\sigma)| \le \frac{C_{n,k}}{\sigma^{n+1}}; \qquad |\partial_k \partial_x^n G(x,k,\sigma)| \le \frac{C_{n,k}}{\sigma^{n+2}}; \qquad \left|\partial_k^2 \partial_x^n G(x,k,\sigma)\right| \le \frac{C_{n,k}}{\sigma^{n+3}}.$$

Proof of Proposition 4.6.4. We start by proving the proposition with $k = 0$ (omitting this argument in G). We start from the expression for the function G.

$$G(x,\sigma) := (\partial_{xx} - \partial_x)\mathrm{BS}(x,\sigma) = \frac{1}{\sigma\sqrt{2\pi}} \exp\left\{x - \frac{1}{2}d_+(x,\sigma)^2\right\},$$

where $d_+(x,\sigma) := \frac{x}{\sigma} + \frac{\sigma}{2}$. Direct computation (proof by recursion) yields, for any $n \in \mathbb{N}$,

$$\partial_x^n G(x,\sigma) = \exp\left\{-\frac{(\sigma^2 - 2x)^2}{8\sigma^2}\right\} \sum_{j=0}^{n} \alpha_j \frac{P_j(x)}{\sigma^{2j+1}}, \tag{4.6.5}$$

where, for each j, P_j is a polynomial of degree j independent of σ.

In the case $x = \sigma^2/2$, we see that $d_+(x,\sigma)|_{x=\frac{\sigma^2}{2}} = \partial_x d_+(x,\sigma)|_{x=\frac{\sigma^2}{2}} = 0$, and $\partial_x^2 d_+(x,\sigma)|_{x=\frac{\sigma^2}{2}} = -\sigma^{-2}$, hence the induction simplifies to

$$\partial_x^n G(x,\sigma)|_{x=\frac{\sigma^2}{2}} = \begin{cases} \dfrac{C_n}{\sigma^{n+1}}, & \text{if } n \text{ is even}, \\ 0, & \text{if } n \text{ is odd}, \end{cases}$$

for some constant $C_n > 0$ independent of σ, proving the third statement in the proposition.

Similarly, if $x = 0$, simplifications occur which yield for any $n \in \mathbb{N}$,

$$\partial_x^n G(x,\sigma)|_{x=0} = \exp\left\{-\frac{\sigma^2}{8}\right\} \sum_{j=0}^{n} \frac{\alpha_j}{\sigma^{j+1}} = \frac{1}{\sigma^{n+1}} \exp\left\{-\frac{\sigma^2}{8}\right\} \sum_{j=0}^{n} \alpha_j \sigma^{n-j},$$

and the second statement in the proposition follows.

Finally, in the general case $x \in \mathbb{R} \setminus \left\{0, \frac{\sigma^2}{2}\right\}$, we can rewrite (4.6.5) for any $p \in \mathbb{R}$ as

$$\begin{aligned} \partial_x^n G(x,\sigma) &= \frac{1}{\sigma^{p+1}} \exp\left\{-\frac{(\sigma^2 - 2x)^2}{8\sigma^2}\right\} \sum_{j=0}^{n} \alpha_k P_j(x) \sigma^{p-2j} \\ &=: \frac{1}{\sigma^{p+1}} \exp\left\{-\frac{(\sigma^2 - 2x)^2}{8\sigma^2}\right\} H_{n,p}(x,\sigma). \end{aligned}$$

For each $n \in \mathbb{N}, p \in \mathbb{N}$, $H_{n,p}$ is a two-dimensional function only consisting of powers of σ^2 and x^2/σ^2. Since the exponential factor contains these very same terms, there exists then a strictly positive constant $C_{n,p}$, independent of x and σ, such that

$$\exp\left\{-\frac{(\sigma^2 - 2x)^2}{8\sigma^2}\right\} H_{n,p}(x,\sigma) \le C_{n,p},$$

proving the proposition in the case $k=0$.

The case $k \in \mathbb{R}$ follows directly by noticing that $G(x,k,\sigma) = G(x-k,0,\sigma)\mathrm{e}^k$. Finally, since $\partial_k d_+ = -\partial_x d_+$ and $\partial_k^2 d_+ = -\partial_x^2 d_+$, the same simplifications occur when taking a partial derivative with respect to k instead of x. □

4.6.2 Proofs of the main results

4.6.2.1 Proof of Theorem 4.3.1: Level

To prove this result, we draw insights from the proofs of [6, Theorem 8] and [9, Proposition 3.1].

By definition
$$\sigma_{\mathrm{BS}}(T) = \mathrm{BS}^{\leftarrow}(0, \mathfrak{M}_0, \mathfrak{M}_0, V_0) =: \overleftarrow{\mathrm{BS}}(V_0),$$

and we write $\widetilde{\mathrm{BS}}(x) := \mathrm{BS}(0,x,x,u_0)$. Using Proposition 4.6.1 at time 0 we see that $V_0 = \Gamma_T$ where

$$\Gamma_t := \mathbb{E}\left[\widetilde{\mathrm{BS}}(\mathfrak{M}_0) + \frac{1}{2}\int_0^t \partial_x G(s, \mathfrak{M}_s, \mathfrak{M}_0, u_s)|\boldsymbol{\Theta}_s| \,\mathrm{d}s\right], \qquad \text{for } t \in \mathbb{T},$$

which is a deterministic path. The fundamental theorem of integration reads

$$\begin{aligned}\sigma_{\mathrm{BS}}(T) = \overleftarrow{\mathrm{BS}}(\Gamma_T) = \overleftarrow{\mathrm{BS}}(\Gamma_0) + \int_0^T \partial_t \overleftarrow{\mathrm{BS}}(\Gamma_t)\,\mathrm{d}t = \overleftarrow{\mathrm{BS}}(\Gamma_0) + \int_0^T \overleftarrow{\mathrm{BS}}'(\Gamma_t)\partial_t\Gamma_t\,\mathrm{d}t \\ = \overleftarrow{\mathrm{BS}}(\Gamma_0) + \frac{1}{2}\int_0^T \overleftarrow{\mathrm{BS}}'(\Gamma_t)\,\mathbb{E}\big[|\boldsymbol{\Theta}_t|\partial_x G_t\big]\,\mathrm{d}t,\end{aligned}$$

where we note $G_t := G(t, \mathfrak{M}_t, \mathfrak{M}_0, u_t)$ for clarity. We can deal with the integral by computing $\overleftarrow{\mathrm{BS}}'$ and $\partial_x G$ explicitly

$$\begin{aligned}\overleftarrow{\mathrm{BS}}'(\Gamma_t) &= \left(e^{\mathfrak{M}_0}\mathcal{N}'\left(d_+\left(\mathfrak{M}_0, \overleftarrow{\mathrm{BS}}(\Gamma_t)\right)\right)\sqrt{T}\right)^{-1}, \\ \partial_x G(s,x,k,\sigma) &= \frac{e^x \mathcal{N}'\big(d_+(k,\sigma)\big)}{\sigma\sqrt{T-s}}\left(1 - \frac{d_+(k,\sigma)}{\sigma\sqrt{T-s}}\right).\end{aligned}$$

Since $\Gamma : \mathbb{R}_+ \to \mathbb{R}$ and $\overleftarrow{\mathrm{BS}} : \mathbb{R} \to \mathbb{R}$ are continuous, the following is uniformly bounded for all $T \le 1$

$$\frac{\mathcal{N}'\big(d_+(\mathfrak{M}_0, u_0)\big)}{\mathcal{N}'\left(d_+\left(\mathfrak{M}_0, \overleftarrow{\mathrm{BS}}(\Gamma_s)\right)\right)} = \exp\left(\frac{1}{8}\left((T-s)\,\overleftarrow{\mathrm{BS}}(\Gamma_s)^2 - \mathrm{u}_0^2\right)\right).$$

Therefore, we obtain by $(\mathbf{H_4})$

$$\lim_{T\downarrow 0}\mathbb{E}\left[\int_0^T \overleftarrow{\mathrm{BS}}'(\Gamma_t)|\boldsymbol{\Theta}_t|\partial_x G_t\,\mathrm{d}t\right] = \lim_{T\downarrow 0}\mathbb{E}\left[\int_0^T \frac{\mathcal{N}'\left(d_+\left(,\mathfrak{M}_0,u_t\right)\right)}{\mathcal{N}'\left(d_+\left(,\mathfrak{M}_0,\overleftarrow{\mathrm{BS}}(\Gamma_t)\right)\right)}\frac{|\boldsymbol{\Theta}_t|}{2\mathfrak{u}_t^2\sqrt{T}}\,\mathrm{d}t\right] = 0.$$

Since $\Gamma_0 = \mathbb{E}\left[\widetilde{\mathrm{BS}}(\mathfrak{M}_0)\right]$ and $u_0 = \overleftarrow{\mathrm{BS}}\left(\widetilde{\mathrm{BS}}(\mathfrak{M}_0)\right)$, we have

$$\overleftarrow{\mathrm{BS}}(\Gamma_0) = \overleftarrow{\mathrm{BS}}\left(\mathbb{E}\left[\widetilde{\mathrm{BS}}(\mathfrak{M}_0)\right]\right) - \mathbb{E}[u_0 - u_0] = \mathbb{E}\left[\overleftarrow{\mathrm{BS}}\left(\mathbb{E}\left[\widetilde{\mathrm{BS}}(\mathfrak{M}_0)\right]\right) - \overleftarrow{\mathrm{BS}}\left(\widetilde{\mathrm{BS}}(\mathfrak{M}_0)\right)\right] + \mathbb{E}[u_0]. \tag{4.6.6}$$

By Clark-Ocone's formula we get $\widetilde{\mathrm{BS}}(\mathfrak{M}_0) = \mathbb{E}\left[\widetilde{\mathrm{BS}}(\mathfrak{M}_0)\right] + \sum_{i=1}^N \int_0^T \mathbb{E}_s\left[\mathrm{D}_s^i\widetilde{\mathrm{BS}}(\mathfrak{M}_0)\right]\mathrm{d}W_s^i$.
By the Gamma-Vega-Delta relation (4.6.1) we have

$$\partial_\sigma \mathrm{BS}(0,x,x,\sigma) = \exp\left(x - \frac{\sigma^2}{8}T\right)\sqrt{\frac{T}{2\pi}}, \tag{4.6.7}$$

which yields in turn

$$\begin{aligned} U_s^i := \mathbb{E}_s\left[\mathrm{D}_s^i\widetilde{\mathrm{BS}}(\mathfrak{M}_0)\right] &= \mathbb{E}_s\left[\frac{\partial_\sigma\widetilde{\mathrm{BS}}(\mathfrak{M}_0)}{2u_0\sqrt{T}}\int_0^T \mathrm{D}_s^i\left\|\boldsymbol{\phi}_r\right\|^2\,\mathrm{d}r\right] \\ &= \mathbb{E}_s\left[\frac{\mathrm{e}^{\mathfrak{M}_0 - \mathfrak{u}_0^2/8}}{2\sqrt{2\pi}u_0}\int_0^T \mathrm{D}_s^i\left\|\boldsymbol{\phi}_r\right\|^2\,\mathrm{d}r\right]. \end{aligned} \tag{4.6.8}$$

Define $\Lambda_r := \mathbb{E}_r\left[\widetilde{\mathrm{BS}}(\mathfrak{M}_0)\right]$ so that the difference we are interested in from (4.6.6) reads, after applying the standard Itô's formula,

$$\overleftarrow{\mathrm{BS}}(\Gamma_0) - \mathbb{E}[u_0] = \mathbb{E}\left[\overleftarrow{\mathrm{BS}}(\Lambda_0) - \overleftarrow{\mathrm{BS}}(\Lambda_T)\right] = \sum_{i=1}^N \mathbb{E}\left[-\int_0^T \overleftarrow{\mathrm{BS}}'(\Lambda_s)U_s^i\mathrm{d}W_s^i - \frac{1}{2}\int_0^T \overleftarrow{\mathrm{BS}}''(\Lambda_s)(U_s^i)^2\,\mathrm{d}s\right]. \tag{4.6.9}$$

The stochastic integral above has zero expectation by the same argument as [9, Proposition 3.1]. Moreover, $(\mathbf{H_5})$ states that $\mathfrak{u}_0$ is dominated almost surely by $Z \in L^p$, and therefore so are Λ and

$$\overleftarrow{\mathrm{BS}}''(\Lambda_s) = \frac{\overleftarrow{\mathrm{BS}}(\Lambda_s)}{4e^{2x}\mathcal{N}'\left(d_+\left(x,\overleftarrow{\mathrm{BS}}(\Lambda_s)\right)\right)^2},$$

by continuity. Plugging in the expression of U^i from (4.6.8), we apply $(\mathbf{H_5})$ to conclude that the second integral of (4.6.9) tends to zero. □

4.6.2.2 Proof of Theorem 4.3.2: Skew

This proof follows similar arguments as [8, Proposition 5.1]. We recall that $V_0(k) = \mathrm{BS}\big(0, \mathfrak{M}_0, k, \sigma_{\mathrm{BS}}(T,k)\big)$. On the one hand, by the chain rule we have

$$\partial_k V_0(k) = \partial_k \mathrm{BS}\big(0, \mathfrak{M}_0, k, \sigma_{\mathrm{BS}}(T,k)\big) + \partial_\sigma \mathrm{BS}\big(0, \mathfrak{M}_0, k, \sigma_{\mathrm{BS}}(T,k)\big)\partial_k \sigma_{\mathrm{BS}}(T,k). \tag{4.6.10}$$

On the other hand, the decomposition obtained in Proposition 4.6.1 yields

$$\partial_k V_0(k) = \mathbb{E}\big[\partial_k \mathrm{BS}(0, \mathfrak{M}_T, k, u_0)\big] + \mathbb{E}\left[\int_0^T \frac{1}{2}\partial_{xk} G(s, \mathfrak{M}_s, \mathfrak{M}_0, u_s)|\mathbf{\Theta}_s|\,\mathrm{d}s\right], \tag{4.6.11}$$

where ∂_{xk} stands for $\partial_x\partial_k$. Equating (4.6.10) and (4.6.11) gives

$$\begin{aligned}\partial_k \sigma_{\mathrm{BS}}(T,k) =& \frac{\mathbb{E}\left[\partial_k \mathrm{BS}(0, \mathfrak{M}_T, k, u_0)\right] - \partial_k \mathrm{BS}(0, \mathfrak{M}_0, k, \sigma_{\mathrm{BS}}(T,k))}{\partial_\sigma \mathrm{BS}(0, \mathfrak{M}_0, k, \sigma_{\mathrm{BS}}(T,k))} \\ &+ \frac{\mathbb{E}\left[\int_0^T \frac{1}{2}\partial_{xk} G(s, \mathfrak{M}_s, \mathfrak{M}_0, u_s)|\mathbf{\Theta}_s|\,\mathrm{d}s\right]}{\partial_\sigma \mathrm{BS}(0, \mathfrak{M}_0, k, \sigma_{\mathrm{BS}}(T,k))},\end{aligned} \tag{4.6.12}$$

which in particular also holds for $k = \mathfrak{M}_0$. Performing simple algebraic manipulations and using the derivatives of the Black-Scholes function ATM as in [8, Proposition 5.1] we obtain that the difference (remember we drop the k-dependence in $\sigma_{\mathrm{BS}}(T)$ when ATM)

$$\mathbb{E}\big[\partial_k \mathrm{BS}(0, \mathfrak{M}_0, \mathfrak{M}_0, u_0)\big] - \partial_k \mathrm{BS}(0, \mathfrak{M}_0, \mathfrak{M}_0, \sigma_{\mathrm{BS}}(T)) = \frac{1}{2}\mathbb{E}\left[\int_0^T \frac{1}{2}\partial_x G(s, \mathfrak{M}_s, \mathfrak{M}_0, u_s)|\mathbf{\Theta}_s|\,\mathrm{d}s\right],$$

which in turn using (4.6.12) yields

$$\partial_k \sigma_{\mathrm{BS}}(T) = \frac{\mathbb{E}\left[\int_0^T L(s, \mathfrak{M}_s, \mathfrak{M}_0, u_s)|\mathbf{\Theta}_s|\,\mathrm{d}s\right]}{\partial_\sigma \mathrm{BS}\Big(0, \mathfrak{M}_0, \mathfrak{M}_0, \sigma_{\mathrm{BS}}(T)\Big)}, \tag{4.6.13}$$

where $L := (\frac{1}{2} + \partial_k)\frac{1}{2}\partial_x G$. We denote $L_s = L(s, \mathfrak{M}_s, \mathfrak{M}_0, u_s)$ for simplicity and apply Lemma 4.6.2 to $L_s \int_s^T |\mathbf{\Theta}_r|\,\mathrm{d}r$, which yields

$$\begin{aligned}\mathbb{E}\left[\int_0^T L_s|\mathbf{\Theta}_s|\,\mathrm{d}s\right] &= \mathbb{E}\left[L_0\int_0^T |\mathbf{\Theta}_s|\,\mathrm{d}s\right] + \mathbb{E}\left[\int_0^T (\partial_x^3 - \partial_x^2)L_s|\mathbf{\Theta}_s|\left(\int_s^T |\mathbf{\Theta}_r|\,\mathrm{d}r\right)\mathrm{d}s\right] \\ &+ \mathbb{E}\left[\int_0^T \partial_x L_s \sum_{j=1}^N \left\{\phi_s^j \mathrm{D}_s^j\left(\int_s^T |\mathbf{\Theta}_r|\,\mathrm{d}r\right)\right\}\mathrm{d}s\right] =: R_1 + R_2 + R_3.\end{aligned}$$

We combine (4.6.7) with the bound $\partial_k \partial_x^n G(t, x, k, \sigma) \leq C\big(\sigma\sqrt{T-t}\big)^{-n-2}$ derived in Corollary 4.6.5 to obtain

$$\frac{R_2}{T^\lambda \partial_\sigma \mathrm{BS}(0, \mathfrak{M}_0, \mathfrak{M}_0, \sigma_{\mathrm{BS}}(T))} \leq \frac{C}{T^{\frac{1}{2}+\lambda}} \mathbb{E} \int_0^T \mathfrak{u}_s^{-6} |\mathbf{\Theta}_s| \left(\int_s^T |\mathbf{\Theta}_r| \, \mathrm{d}r \right) \mathrm{d}s,$$
$$\frac{R_3}{T^\lambda \partial_\sigma \mathrm{BS}(0, \mathfrak{M}_0, \mathfrak{M}_0, \sigma_{\mathrm{BS}}(T))} \leq \frac{C}{T^{\frac{1}{2}+\lambda}} \mathbb{E} \int_0^T \mathfrak{u}_s^{-4} \sum_{j=1}^N \left\{ \phi_s^j \mathrm{D}_s^j \left(\int_s^T |\mathbf{\Theta}_r| \, \mathrm{d}r \right) \right\} \mathrm{d}s,$$

and both converge to zero by $(\mathbf{H}_6^\lambda)$. We are left with R_1. From Appendix 4.6.6, we have

$$L(0, x, x, u) = \frac{\exp(x - \frac{u^2}{8} T)}{u\sqrt{2\pi T}} \left(\frac{1}{4} + \frac{1}{2u^2 T} \right),$$

and therefore by (4.6.7)

$$\frac{L_0}{T^\lambda \partial_\sigma \mathrm{BS}(0, \mathfrak{M}_0, \mathfrak{M}_0, \sigma_{\mathrm{BS}}(T))} = T^{-\lambda} \left(\frac{1}{4} + \frac{1}{2u_0^2 T} \right) \frac{1}{u_0 T} \exp\left\{ -\frac{T}{8}\Big(u_0^2 - \sigma_{\mathrm{BS}}(T)^2\Big) \right\}.$$

This yields

$$\frac{R_1}{T^\lambda \partial_\sigma \mathrm{BS}(0, \mathfrak{M}_0, \mathfrak{M}_0, \sigma_{\mathrm{BS}}(T))} = \mathbb{E}\left[\left(\frac{\mathfrak{u}_0^2}{2} + 1 \right) \exp\left\{ -\frac{T}{8}\Big(u_0^2 - \sigma_{\mathrm{BS}}(T)^2\Big) \right\} \mathfrak{K}_T \right],$$

where $\mathfrak{K}_T := \dfrac{\int_0^T |\mathbf{\Theta}_s| \, \mathrm{d}s}{2T^{\frac{1}{2}+\lambda} \mathfrak{u}_0^3}$. Furthermore,

$$\sup_{\omega \in \Omega} \left| \exp\left\{ -\frac{T}{8}\Big(u_0^2 - \sigma_{\mathrm{BS}}(T)^2\Big) \right\} - 1 \right| = \sup_{\omega \in \Omega} \left| \exp\left\{ \frac{1}{8}\left(T\sigma_{\mathrm{BS}}(T)^2 - \mathfrak{u}_0^2\right) \right\} - 1 \right|$$

not only is finite but converges to zero as T goes to zero. Hence,

$$\lim_{T \downarrow 0} \mathbb{E}\left[\left(\frac{\mathfrak{u}_0^2}{2} + 1 \right) \exp\left\{ -\frac{T}{8}\Big(u_0^2 - \sigma_{\mathrm{BS}}(T)^2\Big) \right\} \mathfrak{K}_T \right] = \lim_{T \downarrow 0} \mathbb{E}\left[\left(\frac{\mathfrak{u}_0^2}{2} + 1 \right) \mathfrak{K}_T \right].$$

We can finally conclude by $(\mathbf{H}_7^\lambda)$ that

$$\lim_{T \downarrow 0} \frac{R_1}{T^\lambda \partial_\sigma \mathrm{BS}(0, \mathfrak{M}_0, \mathfrak{M}_0, \sigma_{\mathrm{BS}}(T))} = \lim_{T \downarrow 0} \mathbb{E}[\mathfrak{K}_T],$$

which has a finite limit. □

4.6.2.3 Proof of Theorem 4.3.3: Curvature

Step 1. Let us start by simply taking a second derivative with respect to k. We write $\mathrm{BS}(\mathfrak{M}_0, \sigma_{\mathrm{BS}}(T,k))$ as short for $\mathrm{BS}(0, \mathfrak{M}_0, \mathfrak{M}_0, \sigma_{\mathrm{BS}}(T,k))$.

$$\begin{aligned}\partial_k\Big(\partial_\sigma \mathrm{BS}(\mathfrak{M}_0, \sigma_{\mathrm{BS}}(T,k))\,\partial_k\sigma_{\mathrm{BS}}(T,k)\Big) &= \partial_\sigma \mathrm{BS}\big(\mathfrak{M}_0, \sigma_{\mathrm{BS}}(T,k)\big)\partial_k^2\sigma_{\mathrm{BS}}(T,k)\\ &+\Big[\partial_{k\sigma}\mathrm{BS}\big(\mathfrak{M}_0, \sigma_{\mathrm{BS}}(T,k)\big) + \partial_\sigma^2\mathrm{BS}\big(\mathfrak{M}_0, \sigma_{\mathrm{BS}}(T,k)\big)\partial_k\sigma_{\mathrm{BS}}(T,k)\Big]\partial_k\sigma_{\mathrm{BS}}(T,k).\end{aligned}$$

Taking the derivative with respect to k in (4.6.13) and equating with the above formula yields

$$\begin{aligned}\partial_k^2\sigma_{\mathrm{BS}}(T,k)\partial_\sigma \mathrm{BS}(\mathfrak{M}_0, \sigma_{\mathrm{BS}}(T)) &= -\,\partial_\sigma^2\mathrm{BS}\big(\mathfrak{M}_0, \sigma_{\mathrm{BS}}(T,k)\big)\partial_k\sigma_{\mathrm{BS}}(T,k)^2\\ &\quad-\partial_{k\sigma}\mathrm{BS}\big(\mathfrak{M}_0, \sigma_{\mathrm{BS}}(T,k)\big)\partial_k\sigma_{\mathrm{BS}}(T,k)\\ &\quad+\mathbb{E}\left[\int_0^T \partial_k L(s, \mathfrak{M}_s, \mathfrak{M}_0, u_s)|\mathbf{\Theta}_s|\,\mathrm{d}s\right]\\ &=:\, T_1 + T_2 + T_3.\end{aligned}$$

A similar expression is presented in [7] and we notice that T_1 and T_2 in the expression above are, after multiplying them by $T^{-\lambda}$, identical to those from [7, Equation (25)] and can therefore be dealt with in the same way. Step 1 shows that $T^{-\lambda}T_1$ tends to zero as $T \downarrow 0$ and step 2 yields $T_2 = -\frac{1}{2}\partial_k\sigma_{\mathrm{BS}}(T,k)$.

Step 2. We recall that $L = \frac{1}{2}\left(\frac{1}{2} + \partial_k\right)\partial_x G$. We will need to use the anticipating Itô's formula (Lemma 4.6.2) twice on T_3. Indeed, even though the bound on $\partial_x^n G$ worsens as n increases, it is more than compensated by the additional integrations. The terms with the more integrals, i.e. the more regularity, will tend to zero as T geos to zero by $(\mathbf{H}_8^\gamma)$ and we will compute the others in closed-form.

For clarity we write $L_s = L(s, \mathfrak{M}_s, \mathfrak{M}_0, u_s)$ for all $s \geq 0$. By a first application of Lemma 4.6.2 on $\partial_k L_s \int_s^T |\mathbf{\Theta}_s|\,\mathrm{d}s$ we get

$$\begin{aligned}T_3 =& \,\mathbb{E}\left[\partial_k L_0 \int_0^T |\mathbf{\Theta}_s|\,\mathrm{d}s\right] + \mathbb{E}\left[\int_0^T (\partial_x^3 - \partial_x^2)\partial_k L_s |\mathbf{\Theta}_s| \left(\int_s^T |\mathbf{\Theta}_r|\,\mathrm{d}r\right)\mathrm{d}s\right]\\ &+\mathbb{E}\left[\int_0^T \partial_{xk}L_s \sum_{j=1}^N \left\{\phi_s^j\,\mathrm{D}_s^j\left(\int_s^T |\mathbf{\Theta}_r|\,\mathrm{d}r\right)\right\}\mathrm{d}s\right] =: S_1 + S_2 + S_3.\end{aligned}$$

To deal with S_2, we apply Lemma 4.6.2 again on

$$(\partial_x^3 - \partial_x^2)\partial_k L_s \int_s^T |\mathbf{\Theta}_r| \left(\int_r^T |\mathbf{\Theta}_y|\,\mathrm{d}y\right)\,\mathrm{d}r =: H_s Z_s,$$

and this yields

$$S_2 = \mathbb{E}\left[H_0 Z_0 + \int_0^T (\partial_x^3 - \partial_x^2) H_s |\mathbf{\Theta}_s| Z_s\,\mathrm{d}s + \int_0^T \partial_x H_s \sum_{j=1}^N \left(\phi_s^j \mathrm{D}_s^j Z_s\right)\,\mathrm{d}s\right] =: S_2^a + S_2^b + S_2^c.$$

We will deal with those three terms in the last step. Regarding S_3, similarly, we apply Lemma 4.6.2 once more to

$$\partial_{xk} L_s \int_s^T \sum_{j=1}^N \left\{\phi_r^j\,\mathrm{D}_r^j \left(\int_r^T |\mathbf{\Theta}_y|\,\mathrm{d}y\right)\right\}\,\mathrm{d}r =: \widetilde{H}_s \widetilde{Z}_s,$$

and similarly we obtain

$$S_3 = \mathbb{E}\left[\widetilde{H}_0 \widetilde{Z}_0 + \int_0^T (\partial_x^3 - \partial_x^2) \widetilde{H}_s |\mathbf{\Theta}_s| \widetilde{Z}_s\,\mathrm{d}s + \int_0^T \partial_x \widetilde{H}_s \sum_{j=1}^N \left(\phi_s^j \mathrm{D}_s^j \widetilde{Z}_s\right)\,\mathrm{d}s\right] =: S_3^a + S_3^b + S_3^c.$$

Step 3. We now evaluate the derivative at $k = \mathfrak{M}_0$ and therefore drop the k dependence. To summarise we have

$$\partial_k^2 \sigma_{\mathrm{BS}}(T) = \frac{T_1 + T_2 + S_1 + S_2^a + S_2^b + S_2^c + S_3^a + S_3^b + S_3^c}{\partial_\sigma \mathrm{BS}\big(0, \mathfrak{M}_0, \mathfrak{M}_0, \sigma_{\mathrm{BS}}(T)\big)},$$

where

$$S_1 = \mathbb{E}\left[\partial_k L_0 \int_0^T |\mathbf{\Theta}_s|\,\mathrm{d}s\right], \qquad S_2^a = \mathbb{E} H_0 \int_0^T |\mathbf{\Theta}_r| \left(\int_r^T |\mathbf{\Theta}_y|\,\mathrm{d}y\right)\,\mathrm{d}r$$

$$S_3^a = \mathbb{E}\widetilde{H}_0 \int_0^T \sum_{j=1}^N \left\{\phi_r^j\,\mathrm{D}_r^j \left(\int_r^T |\mathbf{\Theta}_y|\,\mathrm{d}y\right)\right\}\,\mathrm{d}r$$

$$S_2^b = \mathbb{E}\int_0^T (\partial_x^3 - \partial_x^2) H_s |\mathbf{\Theta}_s| \left(\int_s^T |\mathbf{\Theta}_r| \left(\int_r^T |\mathbf{\Theta}_y|\,\mathrm{d}y\right)\,\mathrm{d}r\right)\,\mathrm{d}s$$

$$S_2^c = \mathbb{E}\int_0^T \partial_x H_s \sum_{j=1}^N \left(\phi_s^j \mathrm{D}_s^j \left(\int_s^T |\mathbf{\Theta}_r| \left(\int_r^T |\mathbf{\Theta}_y|\,\mathrm{d}y\right)\,\mathrm{d}r\right)\,\mathrm{d}s\right)\,\mathrm{d}s$$

$$S_3^b = \mathbb{E}\int_0^T (\partial_x^3 - \partial_x^2)\widetilde{H}_s|\mathbf{\Theta}_s| \int_s^T \sum_{j=1}^N \left\{\phi_r^j \, \mathrm{D}_r^j \left(\int_r^T |\mathbf{\Theta}_y| \, \mathrm{d}y\right)\right\} \mathrm{d}r \, \mathrm{d}s$$

$$S_3^c = \mathbb{E}\int_0^T \partial_x \widetilde{H}_s \sum_{k=1}^N \left\{\phi_s^k \mathrm{D}_s^k \left(\int_s^T \sum_{j=1}^N \left\{\phi_r^j \, \mathrm{D}_r^j \left(\int_r^T |\mathbf{\Theta}_y| \, \mathrm{d}y\right)\right\} \mathrm{d}r\right)\right\} \mathrm{d}s.$$

We recall once again the bound $\partial_x^n G(t,x,k,\sigma) \le C\big(\sigma\sqrt{T-t}\big)^{-n-1}$ as T goes to zero. We observe that H and $\widetilde{H}$ consist of derivatives of G up to the 6th and 4th order respectively, therefore $S_2^b, S_2^c, S_3^b, S_3^c$ tend to zero by $(\mathbf{H_8^\gamma})$. In order to deal with S_1, S_2^a and S_3^a, we use the explicit partial derivatives from Appendix 4.6.6 and (4.6.7) and, as in the proof of Theorem 4.3.2, $(\mathbf{H_9^\gamma})$ entails that only the higher derivatives of $\mathfrak{u}_0$ remain in the limit.

$$\begin{aligned}
\lim_{T\downarrow 0} \frac{S_1}{T^\lambda \partial_\sigma \mathrm{BS}\big(0, \mathfrak{M}_0, \mathfrak{M}_0, \sigma_{\mathrm{BS}}(T)\big)} &= \lim_{T\downarrow 0} \frac{1}{T^\lambda} \frac{\mathbb{E}\left[\partial_k L_0 \int_0^T |\mathbf{\Theta}_s| \, \mathrm{d}s\right]}{\partial_\sigma \mathrm{BS}\big(0, \mathfrak{M}_0, \mathfrak{M}_0, \sigma_{\mathrm{BS}}(T)\big)} \\
&= \lim_{T\downarrow 0} \frac{1}{T^\lambda} \mathbb{E}\left[\frac{1}{u_0 T}\left(\frac{1}{8} + \frac{1}{2u_0^2 T}\right) \int_0^T |\mathbf{\Theta}_s| \, \mathrm{d}s\right] \\
&= \lim_{T\downarrow 0} \mathbb{E}\left[\frac{1}{2\mathfrak{u}_0^3 T^{\frac{1}{2}+\lambda}} \int_0^T |\mathbf{\Theta}_s| \, \mathrm{d}s\right] = \lim_{T\downarrow 0} \frac{\mathcal{S}_T}{T^\lambda}.
\end{aligned}$$

$$\begin{aligned}
\lim_{T\downarrow 0} \frac{S_2^a}{T^\lambda \partial_\sigma \mathrm{BS}\big((0, \mathfrak{M}_0, \mathfrak{M}_0, \sigma_{\mathrm{BS}}(T))\big)} &= \lim_{T\downarrow 0} \frac{1}{T^\lambda} \mathbb{E}\left[\frac{Z_0}{\mathfrak{u}_0\sqrt{T}}\left(-\frac{15}{2\mathfrak{u}_0^6} - \frac{3}{2\mathfrak{u}_0^4} - \frac{5}{32\mathfrak{u}_0^2} - \frac{1}{64}\right)\right] \\
&= \lim_{T\downarrow 0} \mathbb{E}\left[\frac{-15}{2\mathfrak{u}_0^7 T^{\frac{1}{2}+\lambda}} \int_0^T |\mathbf{\Theta}_r| \left(\int_r^T |\mathbf{\Theta}_y| \, \mathrm{d}y\right) \mathrm{d}r\right].
\end{aligned}$$

$$\begin{aligned}
\lim_{T\downarrow 0} \frac{S_3^a}{T^\lambda \partial_\sigma \mathrm{BS}\big(0, \mathfrak{M}_0, \mathfrak{M}_0, \sigma_{\mathrm{BS}}(T)\big)} &= \lim_{T\downarrow 0} \frac{1}{T^\lambda} \mathbb{E}\left[\frac{\widetilde{Z}_0}{\mathfrak{u}_0\sqrt{T}}\left(\frac{3}{2\mathfrak{u}_0^4} + \frac{3}{8\mathfrak{u}_0^2} + \frac{1}{16}\right)\right] \\
&= \lim_{T\downarrow 0} \frac{3}{2} \mathbb{E}\left[\frac{1}{\mathfrak{u}_0^5 T^{\frac{1}{2}+\lambda}} \int_0^T \sum_{j=1}^N \left\{\phi_r^j \, \mathrm{D}_r^j \left(\int_r^T |\mathbf{\Theta}_y| \, \mathrm{d}y\right)\right\} \mathrm{d}r\right].
\end{aligned}$$

Hence to conclude we have

$$\lim_{T\downarrow 0} \frac{\mathcal{C}_T}{T^\lambda} = \lim_{T\downarrow 0} \frac{T_2 + S_1 + S_2^a + S_3^a}{T^\lambda \partial_\sigma \mathrm{BS}(\mathfrak{M}_0, \sigma_{\mathrm{BS}}(T))},$$

and this yields the claim. □

4.6.3 Proof of Proposition 4.4.1: VIX asymptotics

In this section, we will repeatedly interchange Malliavin derivative and conditional expectation, for which a justification can be found in [201, Proposition 1.2.8].

Proposition 4.6.6. *In the case where* $A = \mathrm{VIX}$*, the conditions in* $(\overline{\mathbf{C}})$ *imply assumptions* $(\overline{\mathbf{H}}^{\lambda\gamma})$ *for any* $\lambda \in (-1/2, 0]$ *and* $\gamma \in (-1, 3H - 1/2]$.

Proof. We will use $a \lesssim b$ when there exists $X \in L^p$ such that $a \leq Xb$ almost surely, and $a \approx b$ if $a \lesssim b$ and $b \lesssim a$. $(\mathbf{H_1})$ is granted by the first point of $(\mathbf{C_2})$ and $(\mathbf{H_2})$ corresponds to $(\mathbf{C_1})$. Since $1/M$ is dominated, then so is $1/\mathrm{VIX}$. We then have, for $i = 1, 2$ and by Cauchy-Schwarz:

$$m^i_y = \mathbb{E}_y \left[\frac{\int_T^{T+\Delta} \mathrm{D}^i_y v_r \,\mathrm{d}r}{2\Delta \mathrm{VIX}_T} \right] \lesssim \int_T^{T+\Delta} (r-y)^{H-\frac{1}{2}} \,\mathrm{d}r = \frac{(T+\Delta-y)^{H+\frac{1}{2}} - (T-y)^{H+\frac{1}{2}}}{H+\frac{1}{2}}.$$

If $H < \frac{1}{2}$ then the incremental function $x \mapsto (x+\Delta)^{H+\frac{1}{2}} - x^{H+\frac{1}{2}}$ is decreasing by concavity. For $j = 1, 2$ and $t \leq s$, this implies by domination of $1/M$ that $\phi^i \approx m^i$ is also dominated and

$$\begin{aligned} \mathrm{D}^j_s \phi^i_y &= \frac{\mathrm{D}^j_s m^i_y}{M_y} - \frac{m^i_y \mathrm{D}^j_s M_y}{M_y^2} \qquad (4.6.14) \\ &\lesssim \int_T^{T+\Delta} (r-y)^{H-\frac{1}{2}} (r-s)^{H-\frac{1}{2}} \,\mathrm{d}r + \int_T^{T+\Delta} (r-y)^{H-\frac{1}{2}} \,\mathrm{d}r \int_T^{T+\Delta} (r-s)^{H-\frac{1}{2}} \,\mathrm{d}r \\ &\leq \frac{\Delta^{2H}}{2H} + \frac{\Delta^{H+1}}{(H+\frac{1}{2})^2}. \end{aligned}$$

Combining these two estimates we obtain

$$\Theta^j_s = 2\phi^j_s \int_s^T \left(\sum_{i=1}^N \phi^i_y \mathrm{D}^j_s \phi^i_y \right) \mathrm{d}y \lesssim T - s.$$

It is clear by now that indices and sums do not influence the estimates, hence we informally drop them for more clarity and continue with the higher derivatives:

$$\mathrm{D}_t \Theta_s = \mathrm{D}_t \phi_s \int_s^T \phi_r \mathrm{D}_s \phi_r \,\mathrm{d}r + \phi_s \int_s^T \mathrm{D}_t \phi_r \mathrm{D}_s \phi_r \,\mathrm{d}r + \phi_s \int_s^T \phi_r \mathrm{D}_t \mathrm{D}_s \phi_r \,\mathrm{d}r,$$

where the first and second terms behave like $(T-s)$. For $t \leq s \leq y \leq T$, we deduce from (4.6.14) that $\mathrm{D}_t \mathrm{D}_s \phi_y$ consists of the sum of five terms; four of them behave like $(T-s)$, and only one features three derivatives, it is

$$\mathrm{D}_t\mathrm{D}_s m_y \lesssim \int_T^{T+\Delta}(r-s)^{H-\frac{1}{2}}(r-t)^{H-\frac{1}{2}}(r-y)^{H-\frac{1}{2}}\,\mathrm{d}r \le \int_T^{T+\Delta}(r-y)^{3H-3/2}\,\mathrm{d}y$$
$$\approx (T+\Delta-y)^{3H-\frac{1}{2}}-(T-y)^{3H-\frac{1}{2}}.$$

If $H\ge 1/6$ we have by concavity $\mathrm{D}_t\Theta_s \lesssim (T-s)$. Otherwise, if $H\le 1/6$ the following holds

$$\mathrm{D}_t\Theta_s \lesssim (T-s)+\left[(T+\Delta-s)^{3H+\frac{1}{2}}-\Delta^{3H+\frac{1}{2}}\right]+(T-s)^{3H+\frac{1}{2}} \le (T-s)+2(T-s)^{3H+\frac{1}{2}}.$$

When looking at the second derivative of Θ, the first and second terms behave as $(T-s)$ and $\mathrm{D}_t\Theta_s \lesssim (T-s)+(T-s)^{(3H+\frac{1}{2})\wedge 1}$ respectively, hence we focus on the integral $\int_s^T \mathrm{D}_w\mathrm{D}_t\mathrm{D}_s\phi_y\,\mathrm{d}y$, where the new term is

$$\mathrm{D}_w\mathrm{D}_t\mathrm{D}_s m_y \approx \int_T^{T+\Delta}(r-w)^{H-\frac{1}{2}}(r-s)^{H-\frac{1}{2}}(r-t)^{H-\frac{1}{2}}(r-y)^{H-\frac{1}{2}}\,\mathrm{d}r \lesssim (T+\Delta-y)^{4H-1}-(T-y)^{4H-1}.$$

If $H\ge 1/4$ we have by concavity $\mathrm{D}_t\Theta_s \lesssim (T-s)$. Otherwise, if $H\le 1/4$ the following holds

$$\begin{aligned}\mathrm{D}_w\mathrm{D}_t\Theta_s &\lesssim (T-s)+(T-s)^{(3H+\frac{1}{2})\wedge 1}+\left[(T+\Delta-s)^{4H}-\Delta^{4H}\right]+(T-s)^{4H}\\ &\le (T-s)+(T-s)^{(3H+\frac{1}{2})\wedge 1}+2(T-s)^{4H},\end{aligned}$$

where the last inequality holds by yet again the same concavity argument if $H\le 1/4$.

This yields a rule for checking that the quantities in our assumptions indeed converge. We summarise the above estimates in the case $H\le \frac{1}{2}$: there exists $Z\in L^p$ such that for $s\le T$ and T small enough the following hold almost surely

$$\begin{aligned}\boldsymbol{\Theta}_s &\le Z(T-s)\\ \mathrm{D}\boldsymbol{\Theta}_s &\le Z(T-s)^{(3H+\frac{1}{2})\wedge 1}\\ \mathrm{D}\mathrm{D}\boldsymbol{\Theta}_s &\le Z(T-s)^{(4H)\wedge 1}.\end{aligned}$$

Thanks to Cauchy-Schwarz inequality we can disentangle the numerators (integrals and derivatives of $\boldsymbol{\Theta}$) and denominators (powers of u) of the assumptions, which are both uniformly bounded in L^p. We can easily deduce that $(\mathbf{H_3})$, $(\mathbf{H_4})$, $(\mathbf{H_5})$, $(\mathbf{H_6^\lambda})$, $(\mathbf{H_8^\gamma})$ are satisfied (convergence to zero). In $(\mathbf{H_7^\lambda})$, $\mathbb{E}[\mathfrak{K}_T]$ behaves as $T^{-\lambda}$, hence it converges for any $\lambda \in (-\frac{1}{2}, 0]$, and the uniform L^2 bound is satisfied thanks to $(\mathbf{C_3})$. Moreover, in the limit the first term in $(\mathbf{H_9^\gamma})$ behaves as $T^{-\gamma}$ and the second behaves as $T^{3H-\frac{1}{2}-\gamma}$, therefore both assumptions are satisfied for any $\lambda \in (-1/2, 0]$ and $\gamma \in (-1, 3H - 1/2]$. Similarly, $(\mathbf{C_3})$ ensures the uniform L^2 bounds. $\square$

4.6.3.1 Convergence Lemmas

We require some preliminaries before diving into the computations. We tailor three versions of integral convergence fitted for our purposes, and essential to compute the limits appearing in our main Theorems 4.3.1, 4.3.2 and 4.3.3. The conditions they require will hold thanks to the continuity of $(\mathbf{C_4})$.

Recall the local Taylor theorem: if a function $g(\cdot)$ is continuous on $[0, \delta]$ for some $\delta > 0$, then there exists a continuous function ε on $[0, \delta]$ satisfying $\lim_{x\downarrow 0} \varepsilon(x) = 0$ such that, for any $x \in [0, \delta]$,

$$g(x) = g(0) + \varepsilon(x).$$

Lemma 4.6.7. *Let* $f : \mathbb{R}_+^2 \to \mathbb{R}$ *such that* $f(T, \cdot)$ *is continuous on* $[0, \delta_0]$ *and* $\lim_{T\downarrow 0} f(T, 0) = f(0, 0)$. *Then we have*

$$\lim_{T\downarrow 0} \frac{1}{T} \int_0^T f(T, y)\, \mathrm{d}y = f(0, 0). \tag{4.6.15}$$

Proof. For $T < \delta_0$, we can write

$$\frac{1}{T} \int_0^T f(T, y)\,\mathrm{d}y = \frac{1}{T} \int_0^T [f(T, 0) + \varepsilon_0(y)]\,\mathrm{d}y = f(T, 0) + \frac{1}{T} \int_0^T \varepsilon_0(y)\,\mathrm{d}y,$$

where the function ε_0 is continuous on $[0, \delta_0]$ and converges to zero at the origin. Hence, for any $\eta_0 > 0$, there exists $\widetilde{\delta}_0 > 0$ such that, for any $y \le \widetilde{\delta}_0$, $|\varepsilon_0(y)| < \eta_0$. For all $T < \widetilde{\delta}_0 \wedge \delta_0$,

$$\left| \frac{1}{T} \int_0^T \varepsilon_0(y)\,\mathrm{d}y \right| \le \eta_0.$$

Since η_0 can be taken as small as desired, the fact that $\lim_{T\downarrow 0} f(T, 0) = f(0, 0)$ concludes the proof. $\square$

Lemma 4.6.8. *Let* $f : \mathbb{R}_+^3 \to \mathbb{R}$ *be such that, for each* $y \le T$, $f(T, y, \cdot)$ *is continuous on* $[0, \delta_0]$, $f(T, \cdot, 0)$ *is continuous on* $[0, \delta_1]$ *and* $\lim_{T\downarrow 0} f(T, 0, 0) = f(0, 0, 0)$. *Then the* *following limit holds*

$$\lim_{T\downarrow 0}\frac{1}{T^2}\int_0^T\int_0^y f(T,y,s)\,\mathrm{d}s\,\mathrm{d}y=\frac{f(0,0,0)}{2}. \tag{4.6.16}$$

Proof. For $T<\delta_0\wedge\delta_1$, we can write

$$\begin{aligned}
\frac{1}{T^2}\int_0^T\left\{\int_0^y f(T,y,s)\mathrm{d}s\right\}\mathrm{d}y &= \frac{1}{T^2}\int_0^T\left\{\int_0^y [f(T,y,0)+\varepsilon_0(s)]\,\mathrm{d}s\right\}\mathrm{d}y\\
&= \frac{1}{T^2}\int_0^T\left\{f(T,y,0)y+\int_0^y\varepsilon_0(s)\mathrm{d}s\right\}\mathrm{d}y\\
&= \frac{1}{T^2}\int_0^T\left\{\Big(f(T,0,0)+\varepsilon_1(y)\Big)y+\int_0^y\varepsilon_0(s)\mathrm{d}s\right\}\mathrm{d}y\\
&= \frac{f(T,0,0)}{2}+\frac{1}{T^2}\int_0^T\left\{\varepsilon_1(y)y+\int_0^y\varepsilon_0(s)\mathrm{d}s\right\}\mathrm{d}y,
\end{aligned}$$

where the function ε_1 is continuous on $[0,\delta_1]$ and the function ε_0 is continuous on $[0,\delta_0]$, both converging to zero at the origin.

Since $\varepsilon_1(\cdot)$ tends to zero at the origin, then for any $\eta_1>0$, there exists $\widetilde{\delta}_1>0$ such that, for any $y\leq 0,\widetilde{\delta}_1$, $|\varepsilon_1(y)|<\eta_1$. Therefore, for the first integral, we have, for $T<\widetilde{\delta}_1\wedge\delta_0\wedge\delta_1$,

$$\left|\frac{1}{T^2}\int_0^T\varepsilon_1(y)y\mathrm{d}y\right|\leq\frac{1}{T^2}\int_0^T|\varepsilon_1(y)|\,y\mathrm{d}y\leq\frac{1}{T^2}\int_0^T\eta_1 y\mathrm{d}y\leq\frac{\eta_1}{2}.$$

Likewise, since $\varepsilon_0(\cdot)$ tends to zero at the origin, then for any $\eta_0>0$, there exists $\widetilde{\delta}_0>0$ such that, for any $y\in[0,\widetilde{\delta}_0]$, $|\varepsilon_0(y)|<\eta_0$. Therefore, for the second integral, we have, for $T<\widetilde{\delta}_0\wedge\delta_0\wedge\delta_1$,

$$\left|\frac{1}{T^2}\int_0^T\int_0^y\varepsilon_0(s)\mathrm{d}s\mathrm{d}y\right|\leq\frac{1}{T^2}\int_0^T\int_0^y|\varepsilon_0(s)|\,\mathrm{d}s\mathrm{d}y\leq\frac{1}{T^2}\int_0^T\int_0^y\eta_0\mathrm{d}s\mathrm{d}y\leq\frac{\eta_0}{2}.$$

Since η_1 and η_0 can be taken as small as desired, taking the limit of $f(T,0,0)$ as T goes to zero concludes the proof. □

Lemma 4.6.9. *Let $f:\mathbb{R}_+^4\to\mathbb{R}$ be such that, for all $0\leq s\leq y\leq T$, $f(T,y,s,\cdot)$, $f(T,y,\cdot,0)$, and $f(T,\cdot,0,0)$ are continuous on $[0,\delta_0]$, $[0,\delta_1]$, and $[0,\delta_2]$ respectively, and $\lim_{T\downarrow 0}f(T,0,0,0)=f(0,0,0,0)$. Then the following limit holds*

$$\lim_{T\downarrow 0}\frac{1}{T^3}\int_0^T\int_0^y\int_0^s f(T,y,s,t)\,\mathrm{d}t\,\mathrm{d}s\,\mathrm{d}y=\frac{f(0,0,0,0)}{6}. \tag{4.6.17}$$

Proof. For $T < \delta_0 \wedge \delta_1 \wedge \delta_2$, we can write

$$
\begin{aligned}
&\frac{1}{T^3}\int_0^T \left\{\int_0^y \left(\int_0^s f(T,y,s,t)\,\mathrm{d}t\right)\,\mathrm{d}s\right\}\,\mathrm{d}y \\
&\quad= \frac{1}{T^3}\int_0^T \left\{\int_0^y \left(\int_0^s [f(T,y,s,0)+\varepsilon_0(t)]\,\mathrm{d}t\right)\,\mathrm{d}s\right\}\,\mathrm{d}y \\
&\quad= \frac{1}{T^3}\int_0^T \left\{\int_0^y \left(f(T,y,s,0)s+\int_0^s \varepsilon_0(t)\,\mathrm{d}t\right)\,\mathrm{d}s\right\}\,\mathrm{d}y \\
&\quad= \frac{1}{T^3}\int_0^T \left\{\int_0^y \left([f(T,y,0,0)+\varepsilon_1(s)]s+\int_0^s \varepsilon_0(t)\,\mathrm{d}t\right)\,\mathrm{d}s\right\}\,\mathrm{d}y \\
&\quad= \frac{1}{T^3}\int_0^T \left\{f(T,y,0,0)\frac{y^2}{2}+\int_0^y \left(\varepsilon_1(s)s+\int_0^s \varepsilon_0(t)\,\mathrm{d}t\right)\,\mathrm{d}s\right\}\,\mathrm{d}y \\
&\quad= \frac{1}{T^3}\int_0^T \left\{[f(T,0,0,0)+\varepsilon_2(y)]\frac{y^2}{2}+\int_0^y \left(\varepsilon_1(s)s+\int_0^s \varepsilon_0(t)\,\mathrm{d}t\right)\,\mathrm{d}s\right\}\,\mathrm{d}y \\
&\quad= \frac{f(T,0,0,0)}{6}+\frac{1}{T^3}\int_0^T \left\{\varepsilon_2(y)\frac{y^2}{2}+\int_0^y \left(\varepsilon_1(s)s+\int_0^s \varepsilon_0(t)\,\mathrm{d}t\right)\,\mathrm{d}s\right\}\,\mathrm{d}y,
\end{aligned}
$$

where the function ε_2 is continuous on $[0,\delta_2]$, the function ε_1 is continuous on $[0,\delta_1]$ and the function ε_0 is continuous on $[0,\delta_0]$, all converging to zero at the origin. By the same argument as in the previous proof, for any $\eta_0,\eta_1,\eta_2 > 0$, there exists $\widetilde{\delta} > 0$ such that for all $T \leq \widetilde{\delta}$, $|\varepsilon_0(T)| \leq \eta_0$, $|\varepsilon_1(T)| \leq \eta_1$, and $|\varepsilon_2(T)| \leq \eta_2$. This entails

$$
\left|\frac{1}{T^3}\int_0^T \left\{\varepsilon_2(y)\frac{y^2}{2}+\int_0^y \left(\varepsilon_1(s)s+\int_0^s \varepsilon_0(t)\,\mathrm{d}t\right)\,\mathrm{d}s\right\}\,\mathrm{d}y\right| \leq \frac{\eta_2+\eta_1+\eta_0}{6}.
$$

Since η_2, η_1 and η_0 can be taken as small as desired, taking the limit of $f(T,0,0,0)$ as T goes to zero concludes the proof. □

To apply these lemmas, we will use a modified version of the martingale convergence theorem. It holds in our setting thanks to domination provided by $\mathbf{(C_1)}$ and $\mathbf{(C_2)}$ and the continuity of $\mathbf{(C_4)}$.

Lemma 4.6.10. *Let $(X_t)_{t\geq 0}$ be almost surely continuous in a neighbourhood of zero, with $\sup_{t\leq 1}|X_t| \leq Z \in L^1$. Then the conditional expectation process $(\mathbb{E}_t[X_t])_{t\geq 0}$ is also almost surely continuous in a neighbourhood of zero. In particular,*

$$
\lim_{t\downarrow 0}\mathbb{E}_t[X_t] = \mathbb{E}[X_0].
$$

Remark 4.6.11. The process $(X_t)_{t\geq 0}$ is not necessarily adapted.

Proof. All the limits are taken in the almost sure sense. Let $\delta > 0$ be such that X is continuous on $[0, \delta]$, and fix $t < \delta$. We set a sequence $\{t_n\}_{n\in\mathbb{N}}$ on $[0, \delta]$ which converges to t as n goes infinity.

Assume in the first step that t_n is monotone. Since $\mathcal{F}_{t_n}$ tends monotonically to $\mathcal{F}_t$ and X is dominated, the classical martingale convergence theorem (MCT) asserts $\lim_{n\uparrow\infty} \mathbb{E}_{t_n}[X_t] = \mathbb{E}_t[X_t]$. Let us fix $n \in \mathbb{N}$, then for any $\mathfrak{q} \geq |t_n - t|$,

$$|X_{t_n} - X_t| \leq \sup_{|\mathfrak{p}-t|\leq\mathfrak{q}} |X_{\mathfrak{p}} - X_t| . \tag{4.6.18}$$

Let us fix $\varepsilon > 0$, by MCT, there exists $n_0 \in \mathbb{N}$ such that, if $n \geq n_0$ then

$$\left|\mathbb{E}_{t_n}\left[\sup_{|\mathfrak{p}-t|\leq\mathfrak{q}} |X_{\mathfrak{p}} - X_t|\right] - \mathbb{E}_t\left[\sup_{|\mathfrak{p}-t|\leq\mathfrak{q}} |X_{\mathfrak{p}} - X_t|\right]\right| < \varepsilon,$$

and by the dominated convergence theorem (DCT) there exists $\delta' > 0$ such that $\mathbb{E}_t\left[\sup_{|\mathfrak{p}-t|\leq\delta'} |X_{\mathfrak{p}} - X_t|\right] < \varepsilon$. There exists $n_1 \in \mathbb{N}$ such that for all $n \geq n_1$, $|t_n - t| \leq \delta'$, hence if $n \geq n_0 \vee n_1$, (4.6.18) yields $\mathbb{E}_{t_n}[|X_{t_n} - X_t|] < 2\varepsilon$. Therefore,

$$\lim_{n\uparrow\infty} \mathbb{E}_{t_n}[X_{t_n}] = \mathbb{E}_t[X_t]. \tag{4.6.19}$$

Now we consider the general case where $\{t_n\}_{n\in\mathbb{N}}$ is not monotone. From every subsequence of $\{t_n\}_{n\in\mathbb{N}}$, one can extract a further subsequence which is monotone. Let us call this subsubsequence $\{t_{n_k}\}_{k\in\mathbb{N}}$. Therefore, (4.6.19) holds with t_{n_k} instead of t_n. Since every subsequence of $(\mathbb{E}_{t_n}[X_{t_n}])_{n\in\mathbb{N}}$ has a further subsequence that converges to the same limit, the original sequence also converges to this limit. $\square$

For convenience, we will use the following definition.

Definition 4.6.12. Let $k, n \in \mathbb{N}$ with $k \leq n$. For a function $f : \mathbb{R}_+^n \to \mathbb{R}$, we denote

$$\lim_{x_1\leq x_2\leq\cdots\leq x_k\downarrow 0} f(x_1, \cdots, x_n) := \lim_{x_k\downarrow 0} \cdots \lim_{x_2\downarrow 0} \lim_{x_1\downarrow 0} f(x_1, \cdots, x_n).$$

Notice that the right-hand sides of (4.6.15), (4.6.16), and (4.6.17) correspond to $\lim_{y\leq T\downarrow 0} f(T, y)$, $\lim_{s\leq y\leq T\downarrow 0} f(T, y, s)/2$, and $\lim_{t\leq s\leq y\leq T\downarrow 0} f(T, y, s, t)/6$ respectively.

4.6.3.2 Proof of Proposition 4.4.1

Let us recall some important quantities:

$$M_y = \mathbb{E}_y\left[\mathrm{VIX}_T\right] = \mathbb{E}_y\left[\sqrt{\frac{1}{\Delta}\int_T^{T+\Delta}\mathbb{E}_T v_r\,\mathrm{d}r}\right],$$

$$m_y^i = \mathbb{E}_y[\mathrm{D}_y^i M_y] = \mathbb{E}_y\left[\frac{\int_T^{T+\Delta}\mathrm{D}_y^i\mathbb{E}_T v_r\,\mathrm{d}r}{2\Delta\mathrm{VIX}_T}\right] = \mathbb{E}_y\left[\frac{\int_T^{T+\Delta}\mathrm{D}_y^i v_r\,\mathrm{d}r}{2\Delta\mathrm{VIX}_T}\right],$$

$$\phi_y^i = \frac{m_y^i}{M_y} = \frac{\mathbb{E}_y\left[\left(\int_T^{T+\Delta}\mathrm{D}_y^i v_r\,\mathrm{d}r\right)/(2\Delta\mathrm{VIX}_T)\right]}{\mathbb{E}_y[\mathrm{VIX}_T]}. \tag{4.6.20}$$

We also recall that J_i and G_{ij}, $i,j\in[\![1,N]\!]$ were defined in (4.4.1). In this proof we will define $f(0):=\lim_{x\downarrow 0}f(x)$, for every $f:\mathbb{R}_+\to\mathbb{R}$, as soon as the limit exists and even if f is not actually continuous around zero. This way we make it continuous and it allows us to apply the convergence lemmas.

Level. By $(\mathbf{C_1})$ and the martingale convergence theorem, $\lim_{y\downarrow 0}\mathbb{E}_y[\mathrm{VIX}_T]=\mathbb{E}[\mathrm{VIX}_T]$ and $(M_y)_{y\geq 0}$ is continuous around zero, almost surely. By $(\mathbf{C_4})$ and the dominated convergence theorem (DCT), $\lim_{y\downarrow 0}\int_T^{T+\Delta}\mathrm{D}_y^i v_r\,\mathrm{d}r=\int_T^{T+\Delta}\mathrm{D}_0^i v_r\,\mathrm{d}r$ and $\left(\int_T^{T+\Delta}\mathrm{D}_y^i v_r\,\mathrm{d}r\right)_{y\geq 0}$ is continuous around zero, almost surely. Let $i\in[\![1,N]\!]$, from $(\mathbf{C_1})$ and $(\mathbf{C_2})$ we also obtain that almost surely

$$\frac{1}{\mathrm{VIX}_T}\int_T^{T+\Delta}\mathrm{D}_y^i v_r\,\mathrm{d}r \leq X^2\left[(T+\Delta-y)^{H+\frac{1}{2}}-(T-y)^{H+\frac{1}{2}}\right],$$

for some $X\in L^2$. Therefore it is dominated and by Lemma 4.6.10 , almost surely m_y^i is continuous around zero and $\lim_{y\downarrow 0}m_y^i=\mathbb{E}\left[\frac{\int_T^{T+\Delta}\mathrm{D}_0^i v_r\,\mathrm{d}r}{2\Delta\mathrm{VIX}_T}\right]$. Since $M_y>0$ for all $y\leq T$, ϕ^i is also continuous around zero and

$$\lim_{y\leq T\downarrow 0}\phi_y^i = \frac{J_i}{2\Delta\mathrm{VIX}_0^2}.$$

By virtue of Theorem 4.3.1 and Lemma 4.6.7, we obtain

$$\lim_{T\downarrow 0}\sigma_{\mathrm{BS}}(T) = \lim_{T\downarrow 0}\mathbb{E}[u_0] = \lim_{y\leq T\downarrow 0}\|\boldsymbol{\phi}_y\| = \frac{\|\boldsymbol{J}\|}{2\Delta\mathrm{VIX}_0^2}.$$

Skew. In order to obtain the skew limit we will need to compute a few Malliavin derivatives. For all $i, j \in [\![1, N]\!]$,

$$
\begin{aligned}
\mathrm{D}_s^j m_y^i &= \mathbb{E}_y \left[\frac{\int_T^{T+\Delta} \mathrm{D}_s^j \mathrm{D}_y^i v_r \,\mathrm{d}r \,\mathrm{VIX}_T - \int_T^{T+\Delta} \mathrm{D}_y^i v_r \,\mathrm{d}r \,\mathrm{D}_s^j \mathrm{VIX}_T}{2\Delta \mathrm{VIX}_T^2} \right] \\
&= \mathbb{E}_y \left[\frac{\int_T^{T+\Delta} \mathrm{D}_s^j \mathrm{D}_y^i v_r \,\mathrm{d}r}{2\Delta \mathrm{VIX}_T} - \frac{\int_T^{T+\Delta} \mathrm{D}_y^i v_r \,\mathrm{d}r \int_T^{T+\Delta} \mathrm{D}_s^j v_r \,\mathrm{d}r}{4\Delta^2 \mathrm{VIX}_T^3} \right],
\end{aligned}
$$

which yields

$$
\begin{aligned}
\mathrm{D}_s^j \phi_y^i &= \frac{\mathrm{D}_s^j m_y^i}{M_y} - \frac{m_y^i \mathrm{D}_s^j M_y}{M_y^2} \\
&= \mathbb{E}_y \left[\frac{\int_T^{T+\Delta} \mathrm{D}_s^j \mathrm{D}_y^i v_r \,\mathrm{d}r}{2\Delta \mathrm{VIX}_T M_y} - \frac{\int_T^{T+\Delta} \mathrm{D}_y^i v_r \,\mathrm{d}r \int_T^{T+\Delta} \mathrm{D}_s^j v_r \,\mathrm{d}r}{4\Delta^2 \mathrm{VIX}_T^3 M_y} - \frac{m_y^i \int_T^{T+\Delta} \mathrm{D}_s^j v_r \,\mathrm{d}r}{2\Delta \mathrm{VIX}_T M_y^2} \right] \\
&=: \mathbb{E}_y \left[A_T^{ij}(y,s) + B_T^{ij}(y,s) + C_T^{ij}(y,s) \right].
\end{aligned}
$$

Based on $(\mathbf{C_1})$, $(\mathbf{C_2})$ and $(\mathbf{C_4})$, for each $T \geq 0$, A_T^{ij}, B_T^{ij} and C_T^{ij} are dominated and almost surely continuous in both arguments. For each $s \geq 0$, Lemma 4.6.10 and DCT yield, almost surely, that $(\mathrm{D}_s^j \phi_y^i)_{y \geq 0}$ and $(\mathrm{D}_s^j \phi_0^i)_{s \geq 0}$ are continuous around zero. In particular,

$$
\begin{aligned}
\lim_{s \downarrow 0} \mathbb{E}_y \big[A_T^{ij}(y,s) + B_T^{ij}(y,s) + C_T^{ij}(y,s) \big] &= \mathbb{E} \big[A_T^{ij}(y,0) + B_T^{ij}(y,0) + C_T^{ij}(y,0) \big], \\
\lim_{y \downarrow 0} \mathbb{E}_y \big[A_T^{ij}(y,0) + B_T^{ij}(y,0) + C_T^{ij}(y,0) \big] &= \mathbb{E} \big[A_T^{ij}(0,0) + B_T^{ij}(0,0) + C_T^{ij}(0,0) \big].
\end{aligned}
$$

By DCT again this yields

$$
\lim_{T \downarrow 0} \mathbb{E}[A_T^{ij}(0,0)] = \frac{G_{ij}}{2\Delta \mathrm{VIX}_0^2}, \quad \text{and} \quad \lim_{T \downarrow 0} \mathbb{E}[B_T^{ij}(0,0)] = \lim_{T \downarrow 0} \mathbb{E}[C_T^{ij}(0,0)] = -\frac{J_i J_j}{4\Delta^2 \mathrm{VIX}_0^4}.
$$

Therefore, $\phi_s^j \mathrm{D}_s^j (\phi_y^i)^2$ satisfies the continuity requirements of $f(T, y, s)$ in Lemma 4.6.8. We combine this Lemma with the limits above to see that, almost surely,

$$
\begin{aligned}
\lim_{T \downarrow 0} \frac{1}{T^2} \int_0^T \phi_s^j \int_s^T \mathrm{D}_s^j (\phi_y^i)^2 \,\mathrm{d}y \,\mathrm{d}s &= \lim_{T \downarrow 0} \frac{1}{T^2} \int_0^T \int_0^y \phi_s^j \mathrm{D}_s^j (\phi_y^i)^2 \,\mathrm{d}s \,\mathrm{d}y \\
&= \frac{1}{2} \lim_{s \leq y \leq T \downarrow 0} \phi_s^j \mathrm{D}_s^j (\phi_y^i)^2 \\
&= \frac{J_j}{4\Delta \mathrm{VIX}_0^2} \left[\frac{J_i G_{ij}}{2\Delta^2 \mathrm{VIX}_0^4} - \frac{J_i^2 J_j}{2\Delta^3 \mathrm{VIX}_0^6} \right].
\end{aligned}
$$

We also recall that $\lim_{T \downarrow 0} u_0 = \frac{\|\boldsymbol{J}\|}{2\Delta \mathrm{VIX}_0^2}$ almost surely, hence thanks to $(\mathbf{C_2})$ and $(\mathbf{C_3})$, the dominated convergence theorem entails

$$\lim_{T\downarrow 0}\mathcal{S}_T = \sum_{i,j=1}^{N}\lim_{T\downarrow 0}\frac{1}{2}\mathbb{E}\left[\frac{\int_0^T \phi_s^j \int_s^T \mathrm{D}_s^j(\phi_y^i)^2\,\mathrm{d}y\,\mathrm{d}s}{u_0^3 T^2}\right] = \frac{\sum_{i,j=1}^{N} J_i J_j\left(G_{ij} - \frac{J_i J_j}{\Delta \mathrm{VIX}_0^2}\right)}{2\left\|\boldsymbol{J}\right\|^3}. \tag{4.6.21}$$

Curvature. We now turn our attention to the curvature. By the same arguments as above we have

$$\lim_{T\downarrow 0}\mathbb{E}\left[\frac{\left(\sum_{i,j=1}^{N}\int_0^T \phi_s^i \int_s^T \mathrm{D}_s^j(\phi_y^i)^2\,\mathrm{d}y\,\mathrm{d}s\right)^2}{u_0^7 T^4}\right] = 2\Delta\mathrm{VIX}_0^2\,\frac{\left(\sum_{i,j=1}^{N} J_i J_j\left(G_{ij} - \frac{J_i J_j}{\Delta\mathrm{VIX}_0^2}\right)\right)^2}{\left\|\boldsymbol{J}\right\|^7}.$$

For the last term of (4.3.2) we need to go one step further and compute more Malliavin derivatives since

$$\begin{aligned}
\mathrm{D}_t^k\Theta_s^j &= \sum_{i=1}^{N}\left(\mathrm{D}_t^k\phi_s^j\int_s^T \mathrm{D}_s^j(\phi_y^i)^2\,\mathrm{d}y + 2\phi_s^j\int_s^T \mathrm{D}_t^k\phi_y^i\mathrm{D}_s^j\phi_y^i\,\mathrm{d}y + 2\phi_s^j\int_s^T \phi_y^i\mathrm{D}_t^k\mathrm{D}_s^j\phi_y^i\,\mathrm{d}y\right)\\
&=: \sum_{i=1}^{N}\int_s^T \Upsilon^{ijk}(t,s,y,T)\,\mathrm{d}y.
\end{aligned}$$

Thus we zoom in on the last term of the display above,

$$\begin{aligned}
\mathrm{D}_t^k\mathrm{D}_s^j\phi_y^i &= \frac{\mathrm{D}_t^k\mathrm{D}_s^j m_y^i\,M_y - \mathrm{D}_s^j m_y^i\mathrm{D}_t^k M_y}{M_y^2} - \frac{\mathrm{D}_t^k m_y^i\mathrm{D}_s^j M_y + m_y^i\mathrm{D}_t^k\mathrm{D}_s^j M_y}{M_y^2} + \frac{2m_y^i\mathrm{D}_s^j M_y\mathrm{D}_t^k M_y}{M_y^3}\\
&=: \sum_{n=1}^{5} Q_n^{ijk}(t,s,y,T),
\end{aligned}$$

and zoom in again on $Q_1^{ijk}(t,s,y,T)$,

$$\begin{aligned}
\mathrm{D}_t^k\mathrm{D}_s^j m_y^i = \mathrm{D}_t^k\mathrm{D}_s^j\mathrm{D}_y^i M_y = \mathrm{D}_t^k\mathbb{E}_y&\left[\frac{\int_T^{T+\Delta}\mathrm{D}_s^j\mathrm{D}_y^i v_r\,\mathrm{d}r}{2\Delta\mathrm{VIX}_T} - \frac{\int_T^{T+\Delta}\mathrm{D}_y^i v_r\,\mathrm{d}r\int_T^{T+\Delta}\mathrm{D}_s^j v_r\,\mathrm{d}r}{4\Delta^2\mathrm{VIX}_T^3}\right]\\
&=: \mathbb{E}_y\left[\alpha_T^{ijk} + \beta_T^{ijk}\right].
\end{aligned}$$

Some tedious computations lead to

$$\begin{aligned}
\alpha_T^{ijk} &= \frac{\mathrm{VIX}_T\int_T^{T+\Delta}\mathrm{D}_t^k\mathrm{D}_s^j\mathrm{D}_y^i v_r\,\mathrm{d}r - \mathrm{D}_t^k\mathrm{VIX}_T\int_T^{T+\Delta}\mathrm{D}_s^j\mathrm{D}_y^i v_r\,\mathrm{d}r}{2\Delta\mathrm{VIX}_T^2}\\
&= \frac{\int_T^{T+\Delta}\mathrm{D}_t^k\mathrm{D}_s^j\mathrm{D}_y^i v_r\,\mathrm{d}r}{2\Delta\mathrm{VIX}_T} - \frac{\int_T^{T+\Delta}\mathrm{D}_t^k v_r\,\mathrm{d}r\int_T^{T+\Delta}\mathrm{D}_s^j\mathrm{D}_y^i v_r\,\mathrm{d}r}{4\Delta^2\mathrm{VIX}_T^3};
\end{aligned}$$

$$
\begin{aligned}
\beta_T^{ijk} &= -\frac{\int_T^{T+\Delta} \mathrm{D}_t^k \mathrm{D}_y^i v_r \,\mathrm{d}r \int_T^{T+\Delta} \mathrm{D}_s^j v_r \,\mathrm{d}r + \int_T^{T+\Delta} \mathrm{D}_y^i v_r \,\mathrm{d}r \int_T^{T+\Delta} \mathrm{D}_t^k \mathrm{D}_s^j v_r \,\mathrm{d}r}{4\Delta^2 \mathrm{VIX}_T^3} \\
&\quad + \frac{\mathrm{D}_t^k \mathrm{VIX}_T^3 \int_T^{T+\Delta} \mathrm{D}_y^i v_r \,\mathrm{d}r \int_T^{T+\Delta} \mathrm{D}_s^j v_r \,\mathrm{d}r}{4\Delta^2 \mathrm{VIX}_T^6} \\
&= -\frac{\int_T^{T+\Delta} \mathrm{D}_t^k \mathrm{D}_y^i v_r \,\mathrm{d}r \int_T^{T+\Delta} \mathrm{D}_s^j v_r \,\mathrm{d}r + \int_T^{T+\Delta} \mathrm{D}_y^i v_r \,\mathrm{d}r \int_T^{T+\Delta} \mathrm{D}_t^k \mathrm{D}_s^j v_r \,\mathrm{d}r}{4\Delta^2 \mathrm{VIX}_T^3} \\
&\quad + \frac{3\int_T^{T+\Delta} \mathrm{D}_t^k v_r \,\mathrm{d}r \int_T^{T+\Delta} \mathrm{D}_y^i v_r \,\mathrm{d}r \int_T^{T+\Delta} \mathrm{D}_s^j v_r \,\mathrm{d}r}{8\Delta^3 \mathrm{VIX}_T^5}.
\end{aligned}
$$

We notice, crucially, that we have already justified the continuity of ϕ and $\mathrm{D}\phi$ around zero in the proofs of level and skew respectively. Furthermore, by Lemma 4.6.10 the first two terms in Υ^{ijk} as well as Q_2, Q_3, Q_4, Q_5 all converge to some finite limit as $t \le s \le y \downarrow 0$ and are continuous around zero, almost surely. Similarly, β_T and the second term in α_T are almost surely continuous around zero, and their conditional expectation converges almost surely to some finite limit as $t \le s \le y \downarrow 0$ by DCT and Lemma 4.6.10. Taking the limit T to zero afterwards, all the aforementioned terms tend to a finite limit. On the other hand, by $(\mathbf{C_4})$, DCT, and Lemma 4.6.10 we know that the conditional expectation of the first term in α_T is almost surely continuous around zero, and its limit

$$
\lim_{t\le s\le y\downarrow 0} \mathbb{E}_y \left[\frac{\int_T^{T+\Delta} \mathrm{D}_t^k \mathrm{D}_s^j \mathrm{D}_y^i v_r \,\mathrm{d}r}{2\Delta \mathrm{VIX}_T} \right] = \mathbb{E}\left[\frac{\int_T^{T+\Delta} \mathrm{D}_0^k \mathrm{D}_0^j \mathrm{D}_0^i v_r \,\mathrm{d}r}{2\Delta \mathrm{VIX}_T} \right].
$$

Since $\gamma < 0$, only this term contributes in the limit

$$
\begin{aligned}
\lim_{t\le s\le y\le T\downarrow 0} \frac{\phi_t^k\, \Upsilon^{ijk}(t,s,y,T)}{T^\gamma} &= \lim_{t\le s\le y\le T\downarrow 0} 2\phi_t^k \phi_s^j \phi_y^i \mathbb{E}_y \left[\frac{\int_T^{T+\Delta} \mathrm{D}_t^k \mathrm{D}_s^j \mathrm{D}_y^i v_r \,\mathrm{d}r}{2T^\gamma \Delta \mathrm{VIX}_T M_y} \right] \\
&= \frac{J_i J_j J_k}{8\Delta^4 \mathrm{VIX}_0^8} \lim_{T\downarrow 0} \frac{\mathbb{E}\left[\int_T^{T+\Delta} \mathrm{D}_0^k \mathrm{D}_0^j \mathrm{D}_0^i v_r \,\mathrm{d}r \right]}{T^\gamma},
\end{aligned}
$$

where we applied DCT at the end. Moreover, we know by $(\mathbf{C_2})$ that this limit is finite for $\gamma = 3H - \frac{1}{2}$, hence the conditions of Lemma 4.6.9 are satisfied. We also recall that $\lim_{T\downarrow 0} u_0 = \frac{\|\boldsymbol{J}\|}{2\Delta \mathrm{VIX}_0^2}$ almost surely, hence Lemma 4.6.9 yields the almost sure limit

$$
\begin{aligned}
\lim_{T\downarrow 0} \frac{1}{u_0^5 T^{3+\gamma}} \int_0^T \sum_{k=1}^N \left\{ \phi_t^k \mathrm{D}_t^k \left(\int_t^T |\boldsymbol{\Theta}_s| \,\mathrm{d}s \right) \right\} \mathrm{d}t &= \sum_{i,j,k=1}^N \lim_{T\downarrow 0} \frac{1}{u_0^5 T^3} \int_0^T \int_0^y \int_0^s \frac{\phi_t^k\, \Upsilon^{ijk}(t,s,y,T)}{T^{3H-\frac{1}{2}}} \,\mathrm{d}y \,\mathrm{d}s \,\mathrm{d}t \\
&= \frac{2\Delta \mathrm{VIX}_0^2}{3\|\boldsymbol{J}\|^5} \sum_{i,j,k=1}^N J_i J_j J_k \lim_{T\downarrow 0} \frac{\mathbb{E}\left[\int_T^{T+\Delta} \mathrm{D}_0^k \mathrm{D}_0^j \mathrm{D}_0^i v_r \,\mathrm{d}r \right]}{T^{3H-\frac{1}{2}}}.
\end{aligned}
$$

The first two terms in (4.3.2) tend to zero too because $\gamma < 0$, hence Theorem 4.3.3 and DCT yield the final result

$$\lim_{T\downarrow 0} T^{\frac{1}{2}-3H}\mathcal{C}_T = \frac{2\Delta \mathrm{VIX}_0^2}{3\left\|\boldsymbol{J}\right\|^5}\sum_{i,j,k=1}^{N} J_i J_j J_k \lim_{T\downarrow 0}\frac{\int_T^{T+\Delta}\mathbb{E}\left[\mathrm{D}_0^k\mathrm{D}_0^j\mathrm{D}_0^i v_r\right]\,\mathrm{d}r}{T^{3H-\frac{1}{2}}}.$$

4.6.4 Proofs in the two-factor rough Bergomi model

4.6.4.1 Proof of Proposition 4.4.2

We start with a useful Lemma for Gaussian processes.

Lemma 4.6.13. *For a Gaussian process B on $[0,T]$ with $\|B\|_T := \sup_{t\le T}|B_t|$, $\mathbb{E}\left[\mathrm{e}^{p\|B\|_T}\right]$ is finite for any $p\in\mathbb{R}$.*

Proof. The Borell-TIS inequality asserts that $\mathbb{E}[\|B\|_T] < \infty$ and $\mathbb{P}(\|B\|_T - \mathbb{E}\|B\|_T > x) \le \exp\left(-\frac{x^2}{2\sigma_T^2}\right)$, where $\sigma_T^2 := \sup_{t\le T}\mathbb{E}[B_t^2]$, see [4, Theorem 2.1.1]. We then follow the proof of [4, Theorem 2.1.2].

$$\mathbb{E}\left[\mathrm{e}^{p\|B\|_T}\right] = \int_0^\infty \mathbb{P}\left(\mathrm{e}^{p\|B\|_T} > x\right)\mathrm{d}x \le \mathrm{e}^p + \mathbb{E}[\|B\|_T] + \int_{\mathrm{e}^p\vee\mathbb{E}[\|B\|_T]}^\infty \mathbb{P}\left(\|B\|_T > \frac{\log(x)}{p}\right)\mathrm{d}x.$$

The Borell-TIS inequality in particular reads

$$\mathbb{P}\left(\|B\|_T > \log(x^{1/p})\right) \le \exp\left(-\frac{\left(\log(x^{1/p}) - \mathbb{E}[\|B\|_T]\right)^2}{2\sigma_T^2}\right), \qquad \text{for all } u > \mathbb{E}[\|B\|_T].$$

After a change of variable this yields

$$\int_{\mathrm{e}^p\vee\mathbb{E}[\|B\|_T]}^\infty \mathbb{P}\left(\|B\|_T > \frac{\log(x)}{p}\right)\mathrm{d}x \le \int_{\frac{\log(\mathrm{e}^p\vee\mathbb{E}[\|B\|_T])}{p}}^\infty \exp\left(-\frac{\left(x-\mathbb{E}[\|B\|_T]\right)^2}{2\sigma_T^2}\right)p\mathrm{e}^{px}\,\mathrm{d}x < \infty,$$

as desired. □

By the above lemma $\|v\|_T \in L^p$. We can compute its Malliavin derivatives

$$\mathrm{D}_y^1 v_r = v_0(r-y)^{H-\frac{1}{2}}\left(\chi\nu\mathcal{E}_r^1 + \overline{\chi}\eta\rho\mathcal{E}_r^2\right), \qquad \mathrm{D}_y^2 v_r = v_0\overline{\chi}\eta\overline{\rho}(r-y)^{H-\frac{1}{2}}\mathcal{E}_r^2. \quad (4.6.22)$$

Without computing explicitly further derivatives, one notices that $(\mathbf{C_4})$ holds and that there exist a constant $C > 0$ and a random variable $X = C\left\|\mathcal{E}_r^1 + \mathcal{E}_r^2\right\|_T \in L^p$ for all $p > 1$ such that $\mathrm{D}^i_y v_r \leq X(r-y)^{H-\frac{1}{2}}$, $\mathrm{D}^j_s \mathrm{D}^i_y v_r \leq X(r-s)^{H-\frac{1}{2}}(r-y)^{H-\frac{1}{2}}$ and $\mathrm{D}^k_t \mathrm{D}^j_s \mathrm{D}^i_y v_r \leq X(r-t)^{H-\frac{1}{2}}(r-s)^{H-\frac{1}{2}}(r-y)^{H-\frac{1}{2}}$, implying $(\mathbf{C_2})$.

The following lemma grants $(\mathbf{C_1})$.

Lemma 4.6.14. *In the two factor rough Bergomi model* (4.4.3) *with* $0 \leq T_1 < T_2$*, we have*

$$\mathbb{E}\left[\sup_{y \leq T_1} \left(\mathbb{E}_y\left[\frac{1}{T_2 - T_1}\int_{T_1}^{T_2} v_r \, \mathrm{d}r\right]\right)^{-p}\right] < \infty,$$

for all $p > 1$*. In particular,* $1/M$ *is dominated in* L^p*.*

Proof. We first apply an $\exp - \log$ identity, and from the concave property of the logarithm function we may use Jensen's inequality, to get that

$$\begin{aligned}\left(\mathbb{E}_y\left[\frac{1}{T_2 - T_1}\int_{T_1}^{T_2} v_r \, \mathrm{d}r\right]\right)^{-p} &= \exp\left\{-p \log\left(\frac{1}{T_2 - T_1}\int_{T_1}^{T_2} v_r \, \mathrm{d}r\right)\right\} \\ &\leq \exp\left\{-\frac{p}{T_2 - T_1}\int_{T_1}^{T_2} \log(\mathbb{E}_y[v_r]) \, \mathrm{d}r\right\}.\end{aligned}$$

We further bound $\log(\mathbb{E}_y[v_r])$ by the concavity of the logarithm and (4.4.3)

$$-\log(\mathbb{E}_y[v_r]) \leq -\frac{1}{2}\left(\log\left(2\chi v_0 \mathbb{E}_y[\mathcal{E}^1_r]\right) + \log\left(2\overline{\chi} v_0 \mathbb{E}_y[\mathcal{E}^2_r]\right)\right), \tag{4.6.23}$$

which we now compute

$$\begin{aligned}\mathbb{E}_y[\mathcal{E}^1_r] &= \exp\left(-\frac{\nu^2 r^{2H}}{4H} + \nu\int_0^y (r-s)^{H-\frac{1}{2}} \mathrm{d}W^1_s\right)\mathbb{E}_y\left[\exp\left(\nu\int_y^r (r-s)^{H-\frac{1}{2}} \mathrm{d}W^1_s\right)\right] \\ &= \exp\left(\frac{\nu^2}{4H}\left[(r-u)^{2H} - r^{2H}\right] + \nu\int_0^u (r-s)^{H-\frac{1}{2}} \mathrm{d}W^1_s\right); \\ \mathbb{E}_y[\mathcal{E}^2_r] &= \exp\left(\frac{\eta^2}{4H}\left[(r-y)^{2H} - r^{2H}\right] + \eta\int_0^u (r-s)^{H-\frac{1}{2}} \mathrm{d}(\rho W^1_s + \overline{\rho} W^2_s)\right).\end{aligned} \tag{4.6.24}$$

Let us deal with the first term of (4.6.23), as the second one is analogous. We have

$$\int_{T_1}^{T_2}\left[(r-y)^{2H} - r^{2H}\right] \, \mathrm{d}r = \frac{(T_2-u)^{2H+1} - (T_1-y)^{2H+1} - T_2^{2H+1} + T_1^{2H+1}}{2H+1}$$

is clearly bounded below for all $0 \leq u \leq T_1$. Moreover by Fubini's theorem

$$\begin{aligned}\int_{T_1}^{T_2}\int_0^y (r-t)^{H-\frac{1}{2}} \mathrm{d}W^1_t \, \mathrm{d}r &= \int_0^y \int_{T_1}^{T_2} (r-t)^{H-\frac{1}{2}} \, \mathrm{d}r \mathrm{d}W^1_t \\ &= \int_0^y \frac{(T_2-t)^{H+\frac{1}{2}} - (T_1-t)^{H+\frac{1}{2}}}{H+\frac{1}{2}} \mathrm{d}W^1_t =: \overline{B}_t\end{aligned}$$

is a Gaussian process. Since the exponential function is increasing, $\sup_t \exp(\overline{B}_t) = \exp(\sup_t \overline{B}_t)$ holds, and thus Lemma 4.6.13 entails

$$\mathbb{E}\left[\sup_{y\leq T_1} \exp\left(-\frac{p}{T_2-T_1}\int_{T_1}^{T_2}\int_0^y (r-s)^{H-\frac{1}{2}}\mathrm{d}W_s^1\,\mathrm{d}r\right)\right] \leq \mathbb{E}\left[\exp\left(\frac{p}{T_2-T_1}\left\|\overline{B}\right\|_T\right)\right] < \infty,$$

which concludes the proof. □

Combining (4.6.22), (4.6.24) we obtain $\mathbb{E}_y[\mathrm{D}_y^i v_r]$, $i=1,2$. The following lemma proves that $(\mathbf{C_3})$ is satisfied.

Lemma 4.6.15. *For any $p>1$, $\mathbb{E}[u_s^{-p}]$ is uniformly bounded in s and T, with $s\leq T$.*

Proof. We start by observing that, since VIX and $1/\text{VIX}$ are dominated by some $X\in L^p$ for all $p>1$, then almost surely and independently of the sign of the numerator we obtain

$$m_y^i = \mathbb{E}_y\left[\frac{\int_T^{T+\Delta}\mathrm{D}_y^i v_r\,\mathrm{d}r}{2\Delta\text{VIX}_T}\right] \geq \mathbb{E}_y\left[\frac{\int_T^{T+\Delta}\mathrm{D}_y^i v_r\,\mathrm{d}r}{2\Delta X}\right],$$

and therefore, using that $1/M$ is dominated by X and Jensen's inequality we get

$$\begin{aligned}\frac{1}{u_s^2} &= \frac{T-s}{\int_s^T\sum_{i=1}^N(\phi_y^i)^2\,\mathrm{d}y} \leq \frac{X^2(T-s)}{\int_s^T\sum_{i=1}^N(m_y^i)^2\,\mathrm{d}y} \leq \frac{X^2N(T-s)}{\int_s^T\left(\sum_{i=1}^N m_y^i\right)^2} \leq X^2N\left(\frac{T-s}{\int_s^T\sum_{i=1}^N m_y^i\,\mathrm{d}y}\right)^2\\ &\leq 4X^2N\left(\frac{\int_s^T\int_T^{T+\Delta}\sum_{i=1}^N\mathbb{E}_y\left[\mathrm{D}_y^i v_r/X\right]\,\mathrm{d}r\,\mathrm{d}y}{\Delta(T-s)}\right)^{-2}. \qquad (4.6.25)\end{aligned}$$

Hence we turn our attention to

$$\begin{aligned}&\mathbb{E}\left[\left(\frac{1}{\Delta(T-s)}\int_s^T\int_T^{T+\Delta}\mathbb{E}_y\left[\frac{\mathrm{D}_y^1v_r+\mathrm{D}_y^2v_r}{X}\right]\,\mathrm{d}r\,\mathrm{d}y\right)^{-p}\right]\\ &=\mathbb{E}\left[\exp\left\{-p\log\left(\frac{1}{\Delta(T-s)}\int_s^T\int_T^{T+\Delta}\mathbb{E}_y\left[\frac{\mathrm{D}_y^1v_r+\mathrm{D}_y^2v_r}{X}\right]\,\mathrm{d}r\,\mathrm{d}y\right)\right\}\right]\\ &\leq\mathbb{E}\left[\exp\left\{-\frac{p}{\Delta(T-s)}\int_s^T\int_T^{T+\Delta}\mathbb{E}_y\left[\log\left(\mathrm{D}_y^1v_r+\mathrm{D}_y^2v_r\right)-\log(X)\right]\,\mathrm{d}r\,\mathrm{d}y\right\}\right]\\ &\leq\left(\mathbb{E}\left[\exp\left\{-\frac{2p}{\Delta(T-s)}\int_s^T\int_T^{T+\Delta}\mathbb{E}_y\left[\log\left(\mathrm{D}_y^1v_r+\mathrm{D}_y^2v_r\right)\right]\,\mathrm{d}r\,\mathrm{d}y\right\}\right]\right)^{\frac{1}{2}}\left(\mathbb{E}[X^{2p}]\right)^{\frac{1}{2}},\end{aligned}$$
(4.6.26)

where we applied Jensen's and Cauchy-Schwarz inequalities and $\exp\left(p\mathbb{E}_y\log(X)\right) \leq$

$\mathbb{E}_y[X^p]$. Then by convexity we get

$$
\begin{aligned}
-\log\left(\mathrm{D}_y^1 v_r + \mathrm{D}_y^2 v_r\right) &\leq -\frac{1}{2}\left(\log(2\mathrm{D}_y^1 v_r) + \log(2\mathrm{D}_y^2 v_r)\right) \\
&\leq -\frac{1}{4}\left\{\log\left(4v_0\chi\nu(r-y)^{H-\frac{1}{2}}\mathcal{E}_r^1\right) + \log\left(4v_0\overline{\chi}\eta\rho(r-y)^{H-\frac{1}{2}}\mathcal{E}_r^2\right)\right\} \\
&\quad -\frac{1}{2}\log\left(2v_0\overline{\chi}\eta\overline{\rho}(r-y)^{H-\frac{1}{2}}\mathcal{E}_r^2\right).
\end{aligned}
$$

We focus on the first term and the others can be treated identically. From (4.6.24) we get

$$
\mathbb{E}_y\left[\log\left(4v_0\chi\nu(r-y)^{H-\frac{1}{2}}\mathcal{E}_r^1\right)\right] = \log\left(4v_0\chi\nu(r-y)^{H-\frac{1}{2}}\right) - \frac{\nu^2 r^{2H}}{4H} + \nu\int_0^y (r-t)^{H-\frac{1}{2}}\mathrm{d}W_t^1. \tag{4.6.27}
$$

Let us start with

$$
\begin{aligned}
&\int_s^T \int_T^{T+\Delta} \log\left(4v_0\chi\nu(r-y)^{H-\frac{1}{2}}\right)\,\mathrm{d}r\,\mathrm{d}y \\
&= (T-s)\Delta 4v_0\chi\nu + \left(H-\frac{1}{2}\right)\int_s^T\int_T^{T+\Delta}\log(r-y)\,\mathrm{d}r\,\mathrm{d}y \\
&= (T-s)\Delta 4v_0\chi\nu + \left(H-\frac{1}{2}\right)\int_s^T \Big((T+\Delta-y)\log(T+\Delta-y) - (T+\Delta-y) \\
&\qquad\qquad - (T-y)\log(T-y) + (T-y)\Big)\,\mathrm{d}y \\
&= (T-s)\Delta 4v_0\chi\nu + \left(H-\frac{1}{2}\right)\left\{-\Delta(T-s) - \int_\Delta^{T+\Delta-s} x\log(x)\,\mathrm{d}x + \int_0^{T-s} x\log(x)\,\mathrm{d}x\right\} \\
&= (T-s)\Delta 4v_0\chi\nu + \left(H-\frac{1}{2}\right)\Bigg\{-\Delta(T-s) + \left(\frac{(T-s)^2\log(T-s)}{2} - \frac{(T-s)^2}{4}\right) \\
&\qquad - \left(\frac{(T+\Delta-s)^2\log(T+\Delta-s)}{2} - \frac{(T+\Delta-s)^2}{4}\frac{\Delta^2\log(\Delta)}{2} + \frac{\Delta^2}{4}\right)\Bigg\} \\
&= (T-s)\Delta 4v_0\chi\nu + (T-s)\left(H-\frac{1}{2}\right)\Bigg\{-\Delta + \left(\frac{(T-s)\log(T-s)}{2} - \frac{(T-s)}{4}\right) + \frac{2\Delta+(T-s)}{4} \\
&\qquad - \Delta\log(T+\Delta-s) - \frac{(T-s)\log(T+\Delta-s)}{2}\Bigg\} - \frac{\Delta^2}{2}\big(\log(T+\Delta-s) - \log(\Delta)\big).
\end{aligned}
$$

By Taylor's theorem $\log(T+\Delta-s) - \log(\Delta) = \frac{T-s}{\Delta} + \varepsilon(T-s)$, where $\varepsilon : \mathbb{R}_+ \to \mathbb{R}_+$ is such that $\varepsilon(x)/x$ tends to zero at the origin. We conclude that

$$
-\frac{p}{2\Delta(T-s)}\int_s^T\int_T^{T+\Delta}\log\left(4v_0\chi\nu(r-y)^{H-\frac{1}{2}}\right)\,\mathrm{d}r\,\mathrm{d}y
$$

is uniformly bounded. Now we study the second term of (4.6.27)

$$-\int_s^T \int_T^{T+\Delta} r^{2H}\,\mathrm{d}r\,\mathrm{d}y = (T-s)\frac{T^{2H+1}-(T+\Delta)^{2H+1}}{2H+1}.$$

Therefore the following is uniformly bounded

$$\frac{p}{2\Delta(T-s)}\int_s^T \int_T^{T+\Delta} \frac{\nu^2 r^{2H}}{4H}\,\mathrm{d}r\,\mathrm{d}y.$$

And for the last term we get by stochastic Fubini's theorem [219, Theorem 65]

$$\begin{aligned}\int_s^T \int_T^{T+\Delta} \int_0^y (r-t)^{H-\frac{1}{2}}\mathrm{d}W_t^1\,\mathrm{d}r\,\mathrm{d}y &= \int_s^T \int_0^y \int_T^{T+\Delta} (r-t)^{H-\frac{1}{2}}\,\mathrm{d}r\mathrm{d}W_t^1\,\mathrm{d}y \\ &= \int_0^T \int_{s\vee t}^T \frac{(T+\Delta-t)^{H+\frac{1}{2}}-(T-t)^{H+\frac{1}{2}}}{H+\frac{1}{2}}\,\mathrm{d}y\mathrm{d}W_t^1.\end{aligned}$$

Standard Gaussian computations then yield

$$\begin{aligned}&\mathbb{E}\left[\exp\left\{-\frac{p}{4\Delta(T-s)}\int_s^T \int_T^{T+\Delta} \nu \int_0^y (r-t)^{H-\frac{1}{2}}\mathrm{d}W_t^1\,\mathrm{d}r\,\mathrm{d}y\right\}\right] \\ &= \exp\left\{\frac{1}{2}\left(\frac{p\nu}{4\Delta(T-s)}\right)^2 \int_0^T \left(\int_{s\vee t}^T \frac{(T+\Delta-t)^{H+\frac{1}{2}}-(T-t)^{H+\frac{1}{2}}}{H+\frac{1}{2}}\,\mathrm{d}y\right)^2 \mathrm{d}t\right\}.\end{aligned} \tag{4.6.28}$$

The incremental function $x \mapsto (x+\Delta)^{H+\frac{1}{2}} - x^{H+\frac{1}{2}}$ is decreasing by concavity, hence $(T+\Delta-t)^{H+\frac{1}{2}} - (T-t)^{H+\frac{1}{2}} \leq \Delta^{H+\frac{1}{2}}$ and we obtain

$$\int_0^T (T-s\vee t)^2\,\mathrm{d}t = \int_0^s (T-s)^2\,\mathrm{d}t + \int_s^T (T-t)^2\,\mathrm{d}t = s(T-s)^2 + \frac{(T-s)^3}{3},$$

which entails that (4.6.28) is uniformly bounded. We thus showed that (4.6.26) is uniformly bounded in s, T.

Coming back to (4.6.25) we have by Cauchy-Schwarz inequality

$$\mathbb{E}[u_s^{-p}]^2 \leq 2^p \mathbb{E}[X^p]\,\mathbb{E}\left[\left(\frac{T-s}{\int_s^T (m_y^1+m_y^2)\,\mathrm{d}y}\right)^{2p}\right],$$

which is uniformly bounded for all $s \leq T$, and this concludes the proof. □

4.6.4.2 Proof of Proposition 4.4.3

Level. We start with the following derivatives:

$$\mathrm{D}_s^1 v_t = v_0\Big[\chi\nu(t-s)^{H-\frac{1}{2}}\mathcal{E}_t^1 + \overline{\chi}\eta\rho(t-s)^{H-\frac{1}{2}}\mathcal{E}_t^2\Big], \qquad \mathrm{D}_s^2 v_t = v_0\overline{\chi}\eta\overline{\rho}(t-s)^{H-\frac{1}{2}}\mathcal{E}_t^2.$$

and recall the definitions (4.4.1)

$$\begin{aligned}
J_1 &= \int_0^\Delta v_0\mathbb{E}\left[\chi\nu r^{H-\frac{1}{2}}\mathcal{E}_r^1 + \overline{\chi}\eta\rho r^{H-\frac{1}{2}}\mathcal{E}_r^2\right]\,\mathrm{d}r = v_0(\chi\nu+\overline{\chi}\eta\rho)\frac{\Delta^{H+\frac{1}{2}}}{H+\frac{1}{2}};\\
J_2 &= v_0\overline{\chi}\eta\overline{\rho}\frac{\Delta^{H+\frac{1}{2}}}{H+\frac{1}{2}}.
\end{aligned}$$

We also note that $\mathbb{E}[\mathcal{E}_t^i]=1$. This yields the norm

$$\|\boldsymbol{J}\| = \left[(J_1)^2+(J_2)^2\right]^{\frac{1}{2}} = \frac{v_0\Delta^{H+\frac{1}{2}}}{H+\frac{1}{2}}\sqrt{(\chi\nu+\overline{\chi}\eta\rho)^2+\overline{\chi}^2\eta^2\overline{\rho}^2} = \frac{v_0\Delta^{H+\frac{1}{2}}}{H+\frac{1}{2}}\psi(\rho,\nu,\eta,\chi),$$

with the function ψ defined in the proposition, which grants us the first limit by Proposition 4.4.1. To simplify the notations below, we introduce $\mathfrak{w} := \chi\nu+\overline{\chi}\eta\rho$.

Skew. We compute the further derivatives:

$$\begin{aligned}
\mathrm{D}_0^1\mathrm{D}_0^1 v_t &= v_0\left(\chi\nu^2 t^{2H-1}\mathcal{E}_t^1 + \overline{\chi}\eta^2\rho^2 t^{2H-1}\mathcal{E}_t^2\right);\\
\mathrm{D}_0^1\mathrm{D}_0^2 v_t &= v_0\overline{\chi}\eta^2\rho\overline{\rho}t^{2H-1}\mathcal{E}_t^2;\\
\mathrm{D}_0^2\mathrm{D}_0^2 v_t &= v_0\overline{\chi}\eta^2\overline{\rho}^2 t^{2H-1}\mathcal{E}_t^2;
\end{aligned}$$

Similarly to J, we recall that $G_{ij} = \int_0^\Delta \mathbb{E}\big[\mathrm{D}_0^j\mathrm{D}_0^i v_r\big]\,\mathrm{d}r$, such that

$$G_{11} = \frac{\Delta^{2H}}{2H}v_0(\chi\nu^2+\overline{\chi}\eta^2\rho^2), \qquad G_{12} = \frac{\Delta^{2H}}{2H}v_0\overline{\chi}\eta^2\rho\overline{\rho}, \qquad G_{22} = \frac{\Delta^{2H}}{2H}v_0\overline{\chi}\eta^2\overline{\rho}^2.$$

Notice that $\mathrm{VIX}_0^2 = v_0$, thus we have

$$\begin{aligned}
(J_1)^2\left(G_{11} - \frac{(J_1)^2}{\Delta\mathrm{VIX}_0^2}\right) &= \frac{v_0^3\Delta^{4H+1}}{2H(H+\frac{1}{2})^2}\mathfrak{w}^2(\chi\nu^2+\overline{\chi}\eta^2\rho^2) - \frac{v_0^3\Delta^{4H+1}}{(H+\frac{1}{2})^4}\mathfrak{w}^4\\
&= v_0^3\frac{\Delta^{4H+1}}{(H+\frac{1}{2})^2}\mathfrak{w}^2\left[\frac{\chi\nu^2+\overline{\chi}\eta^2\rho^2}{2H} - \left(\frac{\mathfrak{w}}{H+\frac{1}{2}}\right)^2\right];
\end{aligned}$$

$$
\begin{aligned}
J_1 J_2 \left(G_{12} - \frac{J_1 J_2}{\Delta \mathrm{VIX}_0^2} \right) &= \frac{v_0^3 \Delta^{4H+1}}{2H(H+\frac{1}{2})^2} \mathfrak{w} \overline{\chi}^2 \eta^3 \rho \overline{\rho}^2 - \frac{v_0^3 \Delta^{4H+1}}{(H+\frac{1}{2})^4} \mathfrak{w}^2 \overline{\chi}^2 \eta^2 \overline{\rho}^2 \\
&= v_0^3 \frac{\Delta^{4H+1}}{(H+\frac{1}{2})^2} \mathfrak{w} \overline{\chi}^2 \eta^2 \overline{\rho}^2 \left[\frac{\eta\rho}{2H} - \frac{\mathfrak{w}}{(H+\frac{1}{2})^2} \right]; \\
(J_2)^2 \left(G_{22} - \frac{(J_2)^2}{\Delta \mathrm{VIX}_0^2} \right) &= v_0^3 \frac{\Delta^{4H+1}}{2H(H+\frac{1}{2})^2} \overline{\chi}^3 \eta^4 \overline{\rho}^4 - v_0^3 \overline{\chi}^4 \frac{\Delta^{4H+1}}{(H+\frac{1}{2})^4} \eta^4 \overline{\rho}^4 \\
&= v_0^3 \frac{\Delta^{4H+1}}{(H+\frac{1}{2})^2} \overline{\chi}^3 \eta^4 \overline{\rho}^4 \left(\frac{1}{2H} - \frac{1-\chi}{(H+\frac{1}{2})^2} \right).
\end{aligned}
$$

Finally by Proposition 4.4.1 we obtain

$$
\begin{aligned}
\lim_{T \downarrow 0} \mathcal{S}_T = & \frac{(H+\frac{1}{2}) \Delta^{H-\frac{1}{2}}}{2\psi(\rho,\nu,\eta,\chi)^3} \left\{ \mathfrak{w}^2 \left[\frac{\chi \nu^2 + \overline{\chi} \eta^2 \rho^2}{2H} - \left(\frac{\mathfrak{w}}{H+\frac{1}{2}} \right)^2 \right] \right. \\
& \left. + 2\mathfrak{w} \overline{\chi}^2 \eta^2 \overline{\rho}^2 \left[\frac{\eta\rho}{2H} - \frac{\nu + \eta\rho}{(H+\frac{1}{2})^2} \right] + \overline{\chi}^3 \eta^4 \overline{\rho}^4 \left(\frac{1}{2H} - \frac{1}{(H+\frac{1}{2})^2} \right) \right\}.
\end{aligned}
$$

Curvature. For the last step we go one step further:

$$
\mathrm{D}_0^1 \mathrm{D}_0^1 \mathrm{D}_0^1 v_t = v_0 \left(\chi \nu^3 t^{3H-3/2} \mathcal{E}_t^1 + \overline{\chi} \eta^3 \rho^3 t^{3H-3/2} \mathcal{E}_t^2 \right); \qquad \mathrm{D}_0^2 \mathrm{D}_0^1 \mathrm{D}_0^1 v_t = v_0 \overline{\chi} \eta^3 \rho^2 \overline{\rho} t^{3H-3/2} \mathcal{E}_t^2;
$$

$$
\mathrm{D}_0^2 \mathrm{D}_0^2 \mathrm{D}_0^1 v_t = v_0 \overline{\chi} \eta^3 \rho \overline{\rho}^2 t^{3H-3/2} \mathcal{E}_t^2; \qquad \mathrm{D}_0^2 \mathrm{D}_0^2 \mathrm{D}_0^2 v_t = v_0 \overline{\chi} \eta^3 \overline{\rho}^3 t^{3H-3/2} \mathcal{E}_t^2.
$$

We notice that

$$
\lim_{T \downarrow 0} \frac{\int_T^{T+\Delta} r^{3H-3/2} \,\mathrm{d}r}{T^{3H-\frac{1}{2}}} = \lim_{T \downarrow 0} \frac{(T+\Delta)^{3H-\frac{1}{2}} - T^{3H-\frac{1}{2}}}{T^{3H-\frac{1}{2}}(3H-\frac{1}{2})} = \frac{2}{1-6H}.
$$

By (4.4.2) this entails the following limit

$$
\begin{aligned}
\lim_{T \downarrow 0} \frac{\mathcal{C}_T}{T^{3H-\frac{1}{2}}} \quad &= \frac{2\Delta v_0}{3 \left(\frac{v_0 \Delta^{H+\frac{1}{2}}}{2(H+\frac{1}{2})} \right)^5 \psi(\rho,\nu,\eta,\chi)^5} \left\{ v_0^3 \frac{\Delta^{3H+3/2}}{(H+\frac{1}{2})^3} \mathfrak{w}^3 v_0 \frac{\chi \nu^3 + \overline{\chi} \eta^3 \rho^3}{\frac{1}{2} - 3H} \right. \\
& \left. + 3v_0^3 \frac{\Delta^{3H+3/2}}{(H+\frac{1}{2})^3} \mathfrak{w}^2 \overline{\chi} \eta \overline{\rho} v_0 \frac{\overline{\chi} \eta^3 \rho^2 \overline{\rho}}{\frac{1}{2} - 3H} + 3v_0^3 \frac{\Delta^{3H+3/2}}{(H+\frac{1}{2})^3} \mathfrak{w} \overline{\chi}^2 \eta^2 \overline{\rho}^2 v_0 \frac{\overline{\chi} \eta^3 \rho \overline{\rho}^2}{\frac{1}{2} - 3H} + v_0 \frac{\Delta^{3H+3/2}}{(H+\frac{1}{2})^3} \overline{\chi}^3 \eta^3 \overline{\rho}^3 v_0 \frac{\overline{\chi} \eta^3 \overline{\rho}^3}{\frac{1}{2} - 3H} \right\} \\
&= \frac{128 \Delta^{-2H} (H+\frac{1}{2})^2}{3\psi(\rho,\nu,\eta,\chi)^5 (1-6H)} \left\{ \mathfrak{w}^3 (\chi \nu^3 + \overline{\chi} \eta^3 \rho^3) + 3\mathfrak{w}^2 \overline{\chi}^2 \eta^4 \overline{\rho}^2 \rho^2 + 3\mathfrak{w} \overline{\chi}^3 \eta^5 \overline{\rho}^4 \rho + \overline{\chi}^4 \eta^6 \overline{\rho}^6 \right\},
\end{aligned}
$$

which yields the claim. □

4.6.5 Proofs in the stock price case

4.6.5.1 Proof of Proposition 4.5.1

Since ϕ and u^{-p} are dominated by conditions (i) and (iii) respectively, with the same notations as in the proof of Proposition 4.6.6, we obtain by (ii), as T goes to zero,

$$\mathrm{D}\phi_s \lesssim (T-s)^{H-\frac{1}{2}}; \quad \boldsymbol{\Theta}_s \lesssim (T-s)^{H+\frac{1}{2}}; \quad \mathrm{D}\boldsymbol{\Theta}_s \lesssim (T-s)^{2H}; \quad \mathrm{DD}\boldsymbol{\Theta}_s \lesssim (T-s)^{3H-\frac{1}{2}}.$$

Under our three assumptions it is straightforward to see that $(\mathbf{H_{12345}})$ are satisfied. Moreover, the terms in $(\mathbf{H_6^\lambda})$ behave as $T^{2H-\lambda}$ and the one $(\mathbf{H_7^\lambda})$ behaves as $T^{H-\frac{1}{2}-\lambda}$, which means that by setting $\lambda = H-\frac{1}{2}$ the former vanishes and the second yields a non-trivial behaviour.

Let us have a look at the short-time implied volatility. By Lemma 4.6.7 and the continuity of v we have $\lim_{T\downarrow 0} u_0 = \sqrt{\sum_{i=1}^N v_0\rho_i^2} = \sqrt{v_0}$ almost surely, hence by Theorem 4.3.1 and dominated convergence

$$\lim_{T\downarrow 0}\widehat{\sigma}_{\mathrm{BS}}(T) = \lim_{T\downarrow 0}\mathbb{E}[u_0] = \sqrt{v_0}.$$

We then turn our attention to the short-time skew. Setting $\lambda = H-\frac{1}{2}$ we have by Theorem 4.3.2 and dominated convergence

$$\lim_{T\downarrow 0} T^{\frac{1}{2}-H}\mathcal{S}_T = \sum_{i,j=1}^N \lim_{T\downarrow 0}\frac{1}{2}\mathbb{E}\left[\frac{\int_0^T \phi_s^j \int_s^T \mathrm{D}_s^j(\phi_y^i)^2\,\mathrm{d}y\,\mathrm{d}s}{u_0^3 T^{\frac{3}{2}+H}}\right] = \sum_{j=1}^N \frac{\rho_j}{2v_0^{3/2}}\mathbb{E}\left[\lim_{T\downarrow 0}\frac{\int_0^T\int_s^T \sqrt{v_s}\,\mathrm{D}_s^j v_y\,\mathrm{d}y\,\mathrm{d}s}{T^{\frac{3}{2}+H}}\right],$$

where we used $\sum_{i=1}^N \rho_i^2 = 1$. For any $j \in [\![1,N]\!]$, Cauchy-Schwarz inequality yields

$$\mathbb{E}\left[\left(\sqrt{\frac{v_s}{v_0}}-1\right)\mathrm{D}_s^j v_y\right] \le \mathbb{E}\left[\left(\sqrt{\frac{v_s}{v_0}}-1\right)^2\right]^{\frac{1}{2}} \mathbb{E}\big[(\mathrm{D}_s^j v_y)^2\big]^{\frac{1}{2}},$$

where $\mathbb{E}\big[(\mathrm{D}_s^j v_y)^2\big]^{\frac{1}{2}} \le C(y-s)^{H-\frac{1}{2}}$ for some finite constant C by (ii). Therefore,

$$\lim_{T\downarrow 0}\left(\frac{\int_0^T\int_s^T \mathbb{E}[\sqrt{v_s}\,\mathrm{D}_s^j v_y]\,\mathrm{d}y\,\mathrm{d}s}{\sqrt{v_0}\,T^{\frac{3}{2}+H}} - \frac{\int_0^T\int_s^T \mathbb{E}[\mathrm{D}_s^j v_y]\,\mathrm{d}y\,\mathrm{d}s}{T^{\frac{3}{2}+H}}\right)$$
$$\le C\lim_{T\downarrow 0}\left(\sup_{t\le T}\mathbb{E}\left[\left(\sqrt{\frac{v_t}{v_0}}-1\right)^2\right]^{\frac{1}{2}}\frac{\int_0^T\int_s^T (y-s)^{H-\frac{1}{2}}\,\mathrm{d}y\,\mathrm{d}s}{T^{\frac{3}{2}+H}}\right).$$

Since the fraction is equal to $((H+\frac{1}{2})(H+\frac{3}{2}))^{-1}$ and $\limsup_{T\downarrow 0}\mathbb{E}\big[(\sqrt{v_t/v_0}-1)^2\big]$ is equal to zero by (iv), we obtain

$$\lim_{T\downarrow 0} T^{\frac{1}{2}-H}\widehat{\mathcal{S}}_T = \sum_{j=1}^N \frac{\rho_j}{2v_0}\mathbb{E}\left[\lim_{T\downarrow 0}\frac{\int_0^T\int_s^T \mathrm{D}_s^j v_y\,\mathrm{d}y\,\mathrm{d}s}{T^{\frac{3}{2}+H}}\right].$$

4.6.5.2 Proof of Corollary 4.5.4

Noting that $\mathbb{E}[u_s^{-p}] = \mathbb{E}\left[\left(\frac{1}{T-s}\int_s^T v_r\,\mathrm{d}r\right)^{-p/2}\right]$, Lemmas 4.6.13 and 4.6.14 show that assumptions (i)-(iii) of Proposition 4.5.1 are satisfied under this model. Moreover, v has almost sure continuous paths, hence $\sqrt{v_t/v_0}$ tends to one almost surely, and (iv) holds by reverse Fatou's lemma.

Let $0\le s\le y$. From (4.6.22) we know that

$$\mathbb{E}[\mathrm{D}_s^1 v_y] = v_0(y-s)^{H-\frac{1}{2}}\big(\chi\nu+\overline{\chi}\eta\rho\big),\quad \mathbb{E}[\mathrm{D}_s^2 v_y] = v_0(y-s)^{H-\frac{1}{2}}\overline{\chi}\eta\overline{\rho},$$

and clearly $\mathbb{E}[\mathrm{D}_s^3 v_y]=0$. Therefore, Proposition 4.5.1 entails

$$\lim_{T\downarrow 0} T^{\frac{1}{2}-H}\widehat{\mathcal{S}}_T = \frac{\rho_1}{2v_0}\frac{v_0(\chi\nu+\overline{\chi}\eta\rho)}{(H+\frac{1}{2})(H+\frac{3}{2})} + \frac{\rho_2}{2v_0}\frac{v_0\overline{\chi}\eta\overline{\rho}}{(H+\frac{1}{2})(H+\frac{3}{2})} = \frac{\rho_1\chi\nu+\eta\overline{\chi}(\rho_1\rho+\rho_2\overline{\rho})}{(2H+1)(H+\frac{3}{2})}.$$

4.6.6 Partial derivatives of the Black-Scholes function

We recall the Black-Scholes formula

$$\mathrm{BS}(t,x,k,\sigma) = \mathrm{e}^x\mathcal{N}(d_+) - \mathrm{e}^k\mathcal{N}(d_-),$$

where $\mathcal{N}'(x) = \mathrm{e}^{-x^2/2}/\sqrt{2\pi}$ and $d_\pm = \frac{x-k}{\sigma\sqrt{T-t}} \pm \frac{\sigma}{2}\sqrt{T-t}$. The first partial derivatives are

$$\partial_x\mathrm{BS} = \mathrm{e}^x\mathcal{N}(d_+) \qquad\text{and}\qquad \partial_x^2\mathrm{BS} = \mathrm{e}^x\left\{\mathcal{N}(d_+)+\frac{\mathcal{N}'(d_+)}{\sigma\sqrt{T-t}}\right\},$$

such that

$$G := (\partial_x^2-\partial_x)\mathrm{BS} = \frac{\mathrm{e}^{x-d_+^2/2}}{\sqrt{2\pi}\sigma\sqrt{T-t}} = \frac{\mathrm{e}^{k-d_-^2/2}}{\sqrt{2\pi}\sigma\sqrt{T-t}}.$$

Now let us define $\beta := (\sqrt{2\pi}\sigma\sqrt{T-t})^{-1}$ and

$$f(x,k) := x - \frac{d_+^2}{2} = k - \frac{d_-^2}{2} = \frac{x+k}{2} - \frac{(x-k)^2}{2\sigma^2(T-t)} - \frac{\sigma^2(T-t)}{8}.$$

We then have

$$\partial_x f = \frac{1}{2} - \frac{x-k}{\sigma^2(T-t)}; \qquad \partial_k f = \frac{1}{2} + \frac{x-k}{\sigma^2(T-t)}; \qquad \partial_x^2 f = \partial_k^2 f = -\partial_{xk}^2 f = -\frac{1}{\sigma^2(T-t)}.$$

For the partial derivatives, noting that $\partial_x G = \beta\,\partial_x f\,\mathrm{e}^f$ implies the ATM formula

$$\partial_x G(x,x) = \frac{1}{2}\beta\,\exp\left(x - \frac{\sigma^2}{8}(T-t)\right).$$

Furthermore, we have

$$\begin{aligned}
\partial_{xk} G &= \beta \mathrm{e}^f \Big(\partial_{xk} f + \partial_x f \partial_k f\Big),\\
\partial_{xk} G(x,x) &= \beta \mathrm{e}^f \Big(\frac{1}{\sigma^2(T-t)} + \frac{1}{4}\Big).
\end{aligned}$$

We further define the partial derivatives appearing in the proof of Theorem 4.3.2

$$\begin{aligned}
L &:= \left(\frac{1}{4}\partial_x + \frac{1}{2}\partial_{xk}\right) G = \beta \mathrm{e}^f \left(\frac{1}{4} + \frac{1}{4}\partial_k f - \frac{1}{2}(\partial_k f)^2 - \frac{1}{2}\partial_{kk} f\right),\\
L(x,x) &= \beta \mathrm{e}^f \left(\frac{1}{4} + \frac{1}{2\sigma^2(T-t)}\right).
\end{aligned}$$

Using $\partial_k f = 1 - \partial_x f$ and $\partial_{xk} f = -\partial_{xx} f = -\partial_{kk} f$ we compute

$$\begin{aligned}
\partial_k L &= \beta \mathrm{e}^f \left[\frac{3}{4}\partial_x f - \frac{5}{4}(\partial_x f)^2 + \frac{1}{2}(\partial_x f)^3 - \frac{5}{4}\partial_{xx} f + \frac{3}{2}\partial_x f \partial_{xx} f\right],\\
\partial_k L(x,x) &= \beta \mathrm{e}^f \left(\frac{1}{8} + \frac{1}{2\sigma^2(T-t)}\right).
\end{aligned}$$

Finally, we need the derivatives featuring in the proof of Theorem 4.3.3

$$\begin{aligned}
\widetilde{H} = \partial_{xk} L &= \beta \mathrm{e}^f \left[\frac{3}{4}(\partial_x f)^2 - \frac{5}{4}(\partial_x f)^3 + \frac{1}{2}(\partial_x f)^4 + \frac{3}{4}\partial_{xx} f - \frac{15}{4}\partial_x f \partial_{xx} f + 3(\partial_x f)^2 \partial_{xx} f + \frac{3}{2}(\partial_{xx} f)^2\right]\\
\partial_{xk} L(x,x) &= \beta \mathrm{e}^f \left(\frac{1}{16} + \frac{3}{8}\frac{1}{\sigma^2(T-t)} + \frac{3}{2}\frac{1}{\sigma^4(T-t)^2}\right)
\end{aligned}$$

We need to compute

$$\begin{aligned}
\partial_{xxk} L = \beta \mathrm{e}^f &\left[\frac{3}{4}(\partial_x f)^3 - \frac{5}{4}(\partial_x f)^4 + \frac{1}{2}(\partial_x f)^5 + \frac{9}{4}\partial_{xx} f \partial_x f - \frac{15}{2}\partial_{xx} f (\partial_x f)^2\right.\\
&\left. - \frac{15}{4}(\partial_{xx} f)^2 + 5\partial_{xx} f (\partial_x f)^3 + \frac{15}{2}(\partial_{xx} f)^2 \partial_x f\right],\\
\partial_{xxk} L(x,x) = \beta \mathrm{e}^f &\left(\frac{1}{32} + \frac{1}{8}\frac{1}{\sigma^2(T-t)}\right),
\end{aligned}$$

and one more

$$
\begin{aligned}
\partial_{xxxk} L = \beta \mathrm{e}^f \Bigg[& \frac{3}{4}(\partial_x f)^4 - \frac{5}{4}(\partial_x f)^5 + \frac{1}{2}(\partial_x f)^6 + \frac{9}{2}\partial_{xx} f(\partial_x f)^2 - \frac{25}{2}\partial_{xx} f(\partial_x f)^3 - \frac{75}{4}(\partial_{xx} f)^2 \partial_x f, \\
& + \frac{15}{2}\partial_{xx} f(\partial_x f)^4 + \frac{45}{2}(\partial_{xx} f)^2(\partial_x f)^2 + \frac{9}{4}(\partial_{xx} f)^2 + \frac{15}{2}(\partial_{xx} f)^3 \Bigg] \\
\partial_{xxxk} L(x,x) = \beta \mathrm{e}^f & \left(\frac{1}{64} - \frac{1}{32}\frac{1}{\sigma^2(T-t)} - \frac{3}{2}\frac{1}{\sigma^4(T-t)^2} - \frac{15}{2}\frac{1}{\sigma^6(T-t)^3} \right).
\end{aligned}
$$

We conclude with

$$
H(x,x) = (\partial_{xxxk} - \partial_{xxk})L(x,x) = \beta \mathrm{e}^f \left(-\frac{1}{64} - \frac{5}{32}\frac{1}{\sigma^2(T-t)} - \frac{3}{2}\frac{1}{\sigma^4(T-t)^2} - \frac{15}{2}\frac{1}{\sigma^6(T-t)^3} \right).
$$

5

On the large-time behaviour of affine Volterra processes

This chapter is based on a work in progress with Antoine Jacquier and Konstantinos Spiliopoulos [163].

5.1 Introduction

5.1.1 The problem

We are interested in the large-time behaviour of Stochastic Volterra Equations (SVEs) of affine type which take the form

$$V_t = V_0 + \int_0^t K(t-s)\beta(\theta - V_s)\,\mathrm{d}s + \int_0^t K(t-s)\sigma\sqrt{V_s}\,\mathrm{d}W_s, \quad t \geq 0, \tag{5.1.1}$$

where the kernel is the Laplace transform of some signed measure on $\mathbb{R}_+$, such as the Riemann-Liouville kernel $K(t) = t^{\alpha-1}$, $\alpha \in (1/2, 1)$. The main result of this chapter is the existence of an invariant measure for the Markovian lift of (5.1.1) introduced by Cuchiero and Teichmann [73]; it entails the existence of a stationary measure for V. The renewed interest in SVEs stems from the emergence of a new class of stochastic volatility models which surrender the comfort of Markovianity for the sake of consistency with the data. We refer to the main introduction for more information on these rough volatility models, and stress that (5.1.1) represents the variance of the rough Heston model [93]. We note here however that despite the large number of asymptotic results, very little is known about the ergodic behaviour of rough volatility models.

In fact, to the best of our knowledge only two results exist on their large-time behaviour. The first one is a large deviation principle for the rescaled log-price under the rough Heston model [100]; it is derived from computing the limit of its characteristic function, which is known in semi-closed form. Meanwhile in [129], the authors proved the existence of an invariant measure for the asset price under the assumption that the non-Markovian variance process has a stationary distribution.

5.1.2 Literature

When the process is Markovian, one gains access to transition semigroups which in turn allow to study the ergodic behaviour. A multitude of tools have been developed to prove existence and uniqueness of invariant measures, as well as asymptotic stability and rates of convergence. The most prominent in the literature are based either on showing that the transition semigroup has the (strong) Feller property or on dissipativity methods, which are related to Lyapunov function techniques. We refer to [74, 75] for an overview of these approaches.

A radically different point of view is given by the theory of Random Dynamical Systems where Markovianity is replaced by a cocycle property for the driving noise, e.g. (fractional) Brownian motion [10, 119]. Furthermore, in this context one analyses the asymptotic behaviour by taking the starting time to $-\infty$ and looking at the system at $t = 0$, instead of starting at $t = 0$ and looking at t goes to $+\infty$.

This inspired Hairer's theory of Stochastic Dynamical Systems, which aimed at studying the large-time behaviour of an SDE driven by additive fractional noise [142]. To reconcile the Markovian ergodic theory with non-Markovian processes, his main idea was to augment the state space with all the past of the fractional noise, and to define a Feller semigroup on this augmented space. This led to further advances with multiplicative noise in the case $H > 1/3$ [147, 148], and for the discretised version of the SDE [231].

The more recent literature on large-time behaviour of fractional processes has flourished under the umbrella of rough paths theory, in particular regarding multiscale systems [61, 126, 127, 176, 211, 212]. Their asymptotic properties were also investigated thanks to Malliavin calculus [40] and the stochastic sewing lemma [145]. As usual with fractional stochastic integrals however, these frameworks do not accomodate for the highly irregular paths found in rough volatility models.

Hairer's idea of augmenting the state space to recover Markovianity made its way to rough volatility modelling, albeit with a mild adaptation. Motivated by hedging applications in the rough Heston model, El Euch and Rosenbaum [94] showed that since the variance process is non-Markovian one needs to include the whole forward variance curve; this suggests to consider the system $(S_t, (\mathbb{E}[V_{s+t}|\mathcal{F}_t])_{s\geq 0})$. The forward variance even satisfies a Stochastic Partial Differential Equation (SPDE) [2], and this allowed to further characterise the Markovian structure of the model. A similar SPDE lift for another singular SVE was in fact derived in an earlier work [195]. The analysis of the rough Heston model is facilitated by its affine structure, unlike the rough Bergomi model, its main competitor. Yet, the forward variance curve of the latter also satisfies an SPDE as it belongs to the realm of polynomial processes [72]. The most conducive SPDE lift for our purposes, though, was introduced by Cuchiero and Teichmann in [73]. Unlike the previous approaches, they argue it is more sensible to start from the Markovian lift before solving the SVE. More precisely, they consider the (well-posed) measure-valued SPDE

$$\mathrm{d}\lambda_t(\mathrm{d}x) = -x\lambda_t(\mathrm{d}x)\,\mathrm{d}t + \nu(\mathrm{d}x)\mathrm{d}X_t, \tag{5.1.2}$$

where ν is a signed measure and X is an Itô semimartingale of the form

$$X_t = \int_0^t \beta(\theta - \langle 1, \lambda_s\rangle)\,\mathrm{d}s + \int_0^t \sigma\sqrt{\langle 1, \lambda_s\rangle}\mathrm{d}W_s,$$

where $\langle y, \lambda\rangle := \int_0^\infty y(x)\,\lambda(\mathrm{d}x)$ for all $y \in \mathcal{C}_b(\mathbb{R}_+)$. In fact, one should interpret the SPDE (5.1.2) in the mild sense such that

$$\langle y, \lambda_t\rangle = \int_0^\infty \mathrm{e}^{-tx} y(x)\lambda_0(\mathrm{d}x) + \int_0^\infty \int_0^t \mathrm{e}^{-(t-s)x} y(x)\,\mathrm{d}X_s\,\nu(\mathrm{d}x).$$

Applying the stochastic Fubini theorem, the total mass $\langle 1, \lambda \rangle$ thus solves the affine SVE (5.1.1), with $K(t) = \int_0^\infty \mathrm{e}^{-tx}\,\nu(\mathrm{d}x)$ and λ_0 being the Dirac mass at zero multiplied by V_0.

This setting is not limited to the univariate case and, recalling our initial motivation, we observe that the rough Heston model implies the following lift

$$\begin{cases} \mathrm{d}\lambda_t^1(\mathrm{d}x) & = -x\lambda_t^1(\mathrm{d}x)\,\mathrm{d}t + \delta_0(\mathrm{d}x)\Big(\frac{1}{2}\langle 1, \lambda_t^2\rangle\,\mathrm{d}t + \sqrt{\langle 1, \lambda_t^2\rangle}\mathrm{d}W_t^1\Big), \\ \mathrm{d}\lambda_t^2(\mathrm{d}x) & = -x\lambda_t^2(\mathrm{d}x)\,\mathrm{d}t + \nu(\mathrm{d}x)\Big(\beta(\theta - \langle 1, \lambda_t^2\rangle)\,\mathrm{d}t + \sigma\sqrt{\langle 1, \lambda_t^2\rangle}\mathrm{d}W_t^2\Big), \end{cases}$$

where W^1 and W^2 are correlated Brownian motions, the log asset price is $\langle 1, \lambda_t^1\rangle$ and the variance $\langle 1, \lambda_t^2\rangle$.

The added value of this approach lies in the generalised Feller property, a notion introduced in [86], which is satisfied by the solution to (5.1.2) (albeit with $\theta = 0$) and opens the gates to many parallels with Markovian ergodic theory. Indeed, it is an extension of the standard Feller property to spaces that are not locally compact. Interestingly enough, the authors also prove weak existence and uniqueness of the mild solution of (5.1.2) by computing its Laplace transform

$$\mathbb{E}[\exp(\langle y_0, \lambda_t\rangle)] = \exp(\langle y_t, \lambda_0\rangle),$$

where y is the unique solution of a non-linear PDE.

The ergodic theory of space-time SPDEs is classical by now, with the monographs [74, 75], and in particular the wide literature on the 2D stochastic Navier-Stokes equation which was the stage of several profound advances such as the asymptotic coupling technique and the asymptotic Feller property [98, 143, 144, 146, 236, 235]. More recent results include [53, 57, 169].

As one can expect, the literature on measure-valued SPDEs is much less developed. However, there exist connections with measure-valued branching processes (aka superprocesses) which are also Markovian processes characterised by their Laplace transform, only the PDE satisfied by y takes a different form [178]. We note that the ergodic behaviour of superprocesses has been studied extensively thanks to their Laplace transform [58, 97, 107, 157, 171], which is a hopeful message for us. Furthermore, the analogy with our SPDE lift does not stop there: the density field of some of these superprocesses also satisfies an SPDE with an affine structure [172, 179], and in the continuous-state case branching processes share properties with affine SDEs [168]. However, superprocesses take values in spaces of non-negative measures, which can be locally compact, in which case the standard Feller property makes sense.

5.1.3 Framework and main results

We recall some of the notations introduced in [73]. Let $\boldsymbol{X}$ be a completely regular Hausdorff topological space. We say that $\varrho\colon \boldsymbol{X} \to (0,\infty)$ is an admissible weight function if the sets $K_R := \{x \in \boldsymbol{X} : \varrho(x) \leq R\}$ are compact for all $R > 0$ [73, Definition 2.1]. The supremum norm on $\mathbb{R}$ is denoted $\|\cdot\|_\infty$. The vector space

$$\mathrm{B}^\varrho(\boldsymbol{X}) := \left\{ f\colon \boldsymbol{X} \to \mathbb{R} : \sup_{x\in\boldsymbol{X}} \varrho(x)^{-1} \|f(x)\|_\infty < \infty \right\},$$

equipped with the norm

$$\|f\|_\varrho := \sup_{x\in\boldsymbol{X}} \varrho(x)^{-1} \|f(x)\|_\infty, \tag{5.1.3}$$

is a Banach space, and $\mathcal{C}_b(\boldsymbol{X}) \subset \mathrm{B}^\varrho(\boldsymbol{X})$. Moreover, the space $\mathcal{B}^\varrho(\boldsymbol{X})$ is defined as the closure of $\mathcal{C}_b(\boldsymbol{X})$ in $\mathrm{B}^\varrho(\boldsymbol{X})$ [73, Definition 2.3], and is also a Banach space when equipped with the norm (5.1.3). Generalised Feller semigroups are the analogue of standard Feller semigroups on the space of continuous functions vanishing at infinity on locally compact spaces. They are bounded, positive, linear, strongly continuous operators.

Definition 5.1.1 (Definition 2.5 of [73]). A family of bounded linear operator P_t : $\mathcal{B}^\varrho(\boldsymbol{X}) \to \mathcal{B}^\varrho(\boldsymbol{X})$ for $t \geq 0$ is called generalised Feller semigroup if

(i) $P_0 = I$, the identity on $\mathcal{B}^\varrho(\boldsymbol{X})$,

(ii) $P_{t+s} = P_t P_s$, for all $t, s \geq 0$,

(iii) For all $f \in \mathcal{B}^\varrho(\boldsymbol{X})$ and $x \in \boldsymbol{X}$, $\lim_{t\downarrow 0} P_t f(x) = f(x)$,

(iv) There exist $C \in \mathbb{R}$ and $\varepsilon > 0$ such that for all $t \in [0,\varepsilon]$, $\|P_t\|_{L(\mathcal{B}^\varrho(\boldsymbol{X}))} \leq C$,

(v) P_t is positive for all $t \geq 0$, that is, for $f \in \mathcal{B}^\varrho(\boldsymbol{X})$, $f \geq 0$, we have $P_t f \geq 0$.

Let $\mathcal{P}(\boldsymbol{X})$ be the space of probability measures on $\boldsymbol{X}$.

Definition 5.1.2. For all $t \geq 0$, we define P_t^* as the adjoint of P_t, that is, for all $\varphi \in \mathcal{B}^{\varrho}(\boldsymbol{X})$ and $\mu \in \mathcal{P}(\boldsymbol{X})$, they satisfy

$$\int_X P_t\varphi(x)\mu(\mathrm{d}x) = \langle P_t\varphi, \mu\rangle = \langle \varphi, P_t^*\mu\rangle.$$

We will call $(\lambda_t)_{t\geq 0}$ the process associated to the semigroup $(P_t)_{t\geq 0}$ (and reciprocally) if, for all $t \geq 0$, $P_t^\star\gamma$ is the law of λ_t whenever γ is the law of λ_0. From the display above, one deduces that in that case $P_t\varphi(\lambda_0) = \mathbb{E}_\gamma[\varphi(\lambda_t)]$.

Definition 5.1.3. We say that μ is an invariant measure if $P_t^*\mu = \mu$ for all $t \geq 0$.

Cuchiero and Teichmann provide two types of lifts for the SVE (5.1.1) in [73, Section 5]. The forward curve lift (Section 5.2) lies in $\mathcal{C}(\mathbb{R}_+^2)$ where a tightness criterion depends on Arzelá-Ascoli theorem, as for the original SVE. However, the Kolmogorov criterion seems unsuitable for large time, see e.g. [3, Lemma 2.4] or [162, Lemma 3.12]. The second type of lift, introduced in Section 5.1 and displayed in (5.1.2), takes value in $Y^\star := \mathcal{M}(\overline{\mathbb{R}}_+)$, the space of signed measures on $\overline{\mathbb{R}}_+$, which is the dual space of $Y := \mathcal{C}_b(\overline{\mathbb{R}}_+)$, and for which there exists a nice tightness criterion. Note that in this setting $\boldsymbol{X}$ is a subspace of $Y^\star$.

This chapter aims at proving the following theorem.

Theorem 5.1.4. *Let $\lambda_0(\mathrm{d}x) = V_0\,\delta_0(\mathrm{d}x)$, for some $V_0 > 0$, and $\nu(\mathrm{d}x) = c\,x^{-\alpha}\,\mathrm{d}x$, where $\alpha \in (1/2, 3/2)$ and $c > 0$. Then, there is a unique analytically mild and probabilistically weak solution to (5.1.2), and it has an invariant measure.*

Remark 5.1.5. Existence and uniqueness of (5.1.2) was only established by [73] in the case $\theta = 0$.

Proof. The proof is a combination of Propositions 5.2.1, 5.3.1 and 5.4.1. More precisely, Proposition 5.2.1 states that, for a generalised Feller process $(\lambda_t)_{t\geq 0}$, the bound (5.2.1) is a sufficient condition for the existence of an invariant measure. Proposition 5.3.1 shows that there exists a unique solution $(\lambda_t)_{t\geq 0}$ to (5.1.2) and that it is indeed a generalised Feller process. Finally, Proposition 5.4.1 ensures the condition holds. □

Theorem 5.1.4 delivers a stationary distribution for the SVE (5.1.1) even though it is not Markovian. Note that this choice of ν with $c = \big(\Gamma(1-\alpha)\Gamma(\alpha)\big)^{-1}$ yields the

Riemann-Liouville kernel $K(t) = t^{\alpha-1}/\Gamma(\alpha)$, and that weak existence and uniqueness were derived in [3].

Corollary 5.1.6. *If K is the Riemann-Liouville kernel, then the unique weak solution of (5.1.1) has a stationary distribution.*

Proof. Let us call $\mu^\star$ the invariant measure of λ. If the law of λ_t is $\mu^\star$ for some $t \geq 0$, then $V_{t+s} = \langle 1, \lambda_{t+s}\rangle$ is equal in distribution to $\langle 1, \lambda_t\rangle = V_t$. ☐

The rest of the chapter consists of the proof of Theorem 5.1.4. Naturally, in Section 5.2 we show that Proposition 5.2.1 holds, in Section 5.3 we prove Proposition 5.3.1, and Section 5.4 deals with Proposition 5.4.1.

5.2 A condition for existence

Prior to examining the SVE itself, we devise a simple condition for the existence of an invariant measure for the lift. On the dual space $Y^\star$ we will use the strong norm $\|\lambda\|_{Y^\star} := \sup_{y\in Y,\, \|y\|\leq 1}\langle y, \lambda\rangle$.

Proposition 5.2.1. *If $(\lambda_t)_{t\geq 0}$ is a generalised Feller process taking values in $Y^\star$ and*

$$\sup_{t\geq 0} \mathbb{E}[\|\lambda_t\|_{Y^\star}] < \infty, \tag{5.2.1}$$

then it has an invariant measure.

Remark 5.2.2. Equation (5.2.1) corresponds to $\sup_{t\geq 0}\mathbb{E}[\varrho(\lambda_t)] < \infty$ for the choice of weight function $\varrho(\lambda) := 1 + \|\lambda\|_{Y^\star}$.

We start by stating and proving an extension of Krylov-Bogoliubov theorem (see e.g. [74, Theorem 11.7]) to the setting of generalised Feller processes on $\mathcal{B}^\varrho(\boldsymbol{X})$, for any completely regular Hausdorff topological space $\boldsymbol{X}$.

Lemma 5.2.3. *Let $(P_t)_{t\geq 0}$ be a generalised Feller semigroup. Suppose that there exists $\gamma \in \mathcal{P}(\boldsymbol{X})$ and a sequence T_n going to $+\infty$ as n goes to $+\infty$ such that*

(i) The sequence of measures $P^*_{T_n}\gamma$ converges weakly to some $\mu \in \mathcal{P}(\boldsymbol{X})$,

(ii) For any $\boldsymbol{X}$-valued random variable λ with distribution γ, $\sup_{t\geq 0} P_t\varrho(\lambda) < \infty$.

Then μ is an invariant measure for $(P_t)_{t\geq 0}$.

Remark 5.2.4. By Prokhorov's theorem [74, Theorem 2.3], condition (i) is equivalent to tightness of $(P_t^{\star}\gamma)_{t\geq 0}$. The original Krylov-Bogoliubov theorem for standard Feller semigroups does not require condition (ii).

Proof. Fix $t > 0$ and $\varphi \in \mathcal{B}^{\varrho}(\boldsymbol{X})$, then by definition $P_t\varphi \in \mathcal{B}^{\varrho}(\boldsymbol{X})$. Let $R > 0$, since $f \in \mathcal{B}^{\varrho}(\boldsymbol{X})$ if and only if $f|_{K_R} \in \mathcal{C}(K_R)$ (see [73, Equation (2.3)]), we need to work on K_R to apply the weak convergence of (i).

$$\langle \varphi, P_t^* \mu\rangle = \langle P_t\varphi, \mu\rangle = \int_{\boldsymbol{X}\setminus K_R} P_t\varphi \,\mathrm{d}\mu + \int_{K_R} P_t\varphi \,\mathrm{d}\mu = \int_{\boldsymbol{X}\setminus K_R} P_t\varphi \,\mathrm{d}\mu + \lim_{n\uparrow\infty}\int_{K_R} P_t\varphi \,\mathrm{d}P_{T_n}^*\gamma.$$

Then we go back to $\boldsymbol{X}$ to apply the adjoint property

$$\int_{K_R} P_t\varphi \,\mathrm{d}P_{T_n}^*\gamma = \langle P_t\varphi, P_{T_n}^*\gamma\rangle - \int_{\boldsymbol{X}\setminus K_R} P_t\varphi \,\mathrm{d}P_{T_n}^*\gamma = \langle \varphi, P_{T_n+t}^*\gamma\rangle - \int_{\boldsymbol{X}\setminus K_R} P_t\varphi \,\mathrm{d}P_{T_n}^*\gamma,$$

and back to K_R for the weak convergence

$$\lim_{n\uparrow\infty}\langle \varphi, P_{T_n+t}^*\gamma\rangle = \lim_{n\uparrow\infty}\left(\int_{K_R}\varphi \,\mathrm{d}P_{T_n+t}^*\gamma + \int_{\boldsymbol{X}\setminus K_R}\varphi \,\mathrm{d}P_{T_n+t}^*\gamma\right) = \int_{K_R}\varphi \,\mathrm{d}\mu + \lim_{n\uparrow\infty}\int_{\boldsymbol{X}\setminus K_R}\varphi \,\mathrm{d}P_{T_n+t}^*\gamma.$$

Overall, this yields our objective plus a remainder

$$\begin{aligned}\langle \varphi, P_t^* \mu\rangle &= \langle \varphi, \mu\rangle - \int_{\boldsymbol{X}\setminus K_R}\varphi \,\mathrm{d}\mu + \int_{\boldsymbol{X}\setminus K_R} P_t\varphi \,\mathrm{d}\mu + \lim_{n\uparrow+\infty}\left(\int_{\boldsymbol{X}\setminus K_R}\varphi \,\mathrm{d}P_{T_n+t}^*\gamma - \int_{\boldsymbol{X}\setminus K_R} P_t\varphi \,\mathrm{d}P_{T_n}^*\gamma\right)\\ &=: \langle \varphi, \mu\rangle + \varepsilon(R).\end{aligned}$$

To deal with $\varepsilon(R)$, we apply a Fatou-type lemma for measures [88, Theorem A.3.12]. For any $f \in \mathcal{B}^{\varrho}(\boldsymbol{X})$,

$$\begin{aligned}\left|\int_{\boldsymbol{X}\setminus K_R} f \,\mathrm{d}\mu\right| \leq \sup_{x\in\boldsymbol{X}\setminus K_R}\frac{|f(x)|}{\varrho(x)}\int_{\boldsymbol{X}}\varrho \,\mathrm{d}\mu &\leq \sup_{x\in\boldsymbol{X}\setminus K_R}\frac{|f(x)|}{\varrho(x)}\liminf_{n\uparrow+\infty}\int_{\boldsymbol{X}}\varrho \,\mathrm{d}P_{T_n}^*\gamma\\ &= \sup_{x\in\boldsymbol{X}\setminus K_R}\frac{|f(x)|}{\varrho(x)}\liminf_{n\uparrow+\infty} P_{T_n}\varrho(\lambda),\end{aligned}$$

where λ is a $\boldsymbol{X}$-valued random variable with distribution γ. Therefore,

$$\varepsilon(R) \leq 2\left(\sup_{x\in\boldsymbol{X}\setminus K_R}\frac{|\varphi(x)| + |P_t\varphi(x)|}{\varrho(x)}\right)\left(\sup_{t\geq 0} P_t\varrho(\lambda)\right).$$

By Equation (2.3) in [73] we have for any $f \in \mathcal{B}^{\varrho}(\boldsymbol{X})$,

$$\lim_{R\uparrow\infty} \sup_{\boldsymbol{X}\setminus K_R} \frac{|f(x)|}{\varrho(x)} = 0.$$

The uniform bound $\sup_{t\geq 0} P_t\varrho(\lambda) < \infty$ thus ensures that $\varepsilon(R)$ vanishes as R goes to $+\infty$. □

Proof of Proposition 5.2.1. Let $(P_t)_{t\geq 0}$ be the generalised Feller semigroup associated to $(\lambda_t)_{t\geq 0}$. Hence, for any $\boldsymbol{X}$-valued random variable λ_0, $P_t\varrho(\lambda_0) = \mathbb{E}[\varrho(\lambda_t)]$.

In [73], the analysis is performed with the weight function $\varrho(\lambda) = 1 + \|\lambda\|^2_{Y^\star}$. The only necessary estimate is $\mathbb{E}[\varrho(\lambda_t)] \leq C\varrho(\lambda_0)$, for some finite constant C, and we note that the inequality $\mathbb{E}[1 + \|\lambda_t\|^2_{Y^\star}] \leq C(1 + \|\lambda_0\|^2_{Y^\star})$ implies $\mathbb{E}[1 + \|\lambda_t\|_{Y^\star}] \leq \sqrt{2C}(1 + \|\lambda_0\|_{Y^\star})$. Therefore, the weight function $\varrho(\lambda) = 1 + \|\lambda\|_{Y^\star}$ is also admissible, and the assumption (5.2.1) implies

$$\sup_{t\geq 0} \mathbb{E}[\varrho(\lambda_t)] < \infty. \tag{5.2.2}$$

This yields condition (ii) of Lemma 5.2.3.

Furthermore, by [46, Lemma 2.9], $(\lambda_t)_{t\geq 0}$ is tight if and only if there exists a tightness function $G : Y^\star \to \mathbb{R}_+$ such that $\sup_{t\geq 0} \mathbb{E}[G(\lambda_t)] < \infty$. We observe that an admissible weight function is also a tightness function by design, hence proving tightness of a generalised Feller process $(\lambda_t)_{t\geq 0}$ boils down to showing (5.2.2). Therefore, both conditions of Lemma 5.2.3 are satisfied as soon as (5.2.1) holds. □

5.3 The generalised Feller condition holds

Let us recall the measure-valued SPDE introduced in [73, Section 5.1]

$$\mathrm{d}\lambda_t(\mathrm{d}x) = -x\lambda_t(\mathrm{d}x)\,\mathrm{d}t + \nu(\mathrm{d}x)\mathrm{d}X_t, \tag{5.3.1}$$

where, ignoring jumps and denoting $\overline{\lambda} = \langle 1, \lambda\rangle$, X is an Itô semimartingale of the form

$$X_t = \int_0^t \beta(\theta - \overline{\lambda}_s)\,\mathrm{d}s + \int_0^t \sigma\sqrt{\overline{\lambda}_s}\mathrm{d}B_s. \tag{5.3.2}$$

The initial condition λ_0 belongs to $\mathcal{E} \subset Y^\star$, a closed cone left invariant by the solution of (5.3.1) with $\theta = 0$, see [73, Definition 4.12], and ν is a signed measure on $\mathbb{R}_+$. Recall that this SPDE should be understood in the (analytically) mild sense

$$\langle y, \lambda_t \rangle = \langle \mathrm{e}^{-t\cdot} y, \lambda_0 \rangle + \int_0^t \langle \mathrm{e}^{-(t-s)\cdot} y, \nu \rangle \, \mathrm{d}X_s, \quad \text{for any } y \in Y.$$

The SPDE (5.3.1) with $\theta = 0$ has a unique probabilistically weak and analytically mild solution and is a generalised Feller process [73, Theorem 5.7]. However, it generates an SVE reverting to zero, so we would like to consider $\theta > 0$ by perturbing its generator, and hence showing that (5.3.1) is also a generalised Feller process.

5.3.1 On summing generators

We want to perturb the generator A of the SPDE (5.3.1) (with mean-reversion parameter $\theta = 0$) by applying one of the theorems in Section 3 of [73], which all unfold from their Theorem 2.10 for bounded perturbations. Theorem 3.1 of [73] requires a Cauchy-type convergence with respect to the norm of a dense subset D of $\mathcal{B}^{\varrho}$. In the paper, it is applied with $D = \mathcal{B}^{\sqrt{\varrho}}$ but does not work here because it leads to an estimate of the type

$$\|A^n f - A^m f\|_{\varrho} \leq a_{nm} \|f\|_{\sqrt{\varrho}} \frac{\overline{\lambda}}{\sqrt{\|\lambda\|_{\varrho}}},$$

and we picked $\varrho(x) = 1 + \|x\|$ instead of $1 + \|x\|^2$, hence this is unbounded. Theorem 3.3 of [73] is tailored for jumps perturbations. Theorem 3.2 of [73] is promising: one needs a subset D such that its *span* is dense in $\mathcal{B}^{\varrho}$. In the paper, the set D is defined as the Fourier basis elements of the form

$$f_y : \mathcal{E} \to [0,1];\ \lambda \mapsto \exp(\langle y, \lambda \rangle). \tag{5.3.3}$$

Moreover, $\mathrm{span}(D)$ is equipped with the uniform norm, such that heuristically at least we can expect

$$\|A^n f_y - A^m f_y\|_{\varrho} \leq a_{nm} \|f_y\|_{\infty} \frac{\overline{\lambda}}{\|\lambda\|_{\varrho}} \leq a_{nm} \|f_y\|_{\infty},$$

for some sequence a_{nm} going to zero as n, m go to $+\infty$. With some rigour, we attain our objective.

Proposition 5.3.1. *For every $\lambda_0 \in \mathcal{E}$, the SPDE (5.3.1) has a unique probabilistically weak and analytically mild solution in $\mathcal{E}$, and its semigroup satisfies the generalised Feller property.*

Remark 5.3.2. For clarity, we work with the lift of the SVE directly (the SPDE (5.3.1)), but notice that one could replace it by the more general SPDE displayed in (4.19) of [73]. The proof would still hold with $\langle y, \lambda\rangle$ and $\mathcal{S}_t^\star\lambda$ instead of $\overline{\lambda}$ and $\mathrm{e}^{-t\cdot}\lambda$, respectively.

Proof. For all $y \in Y$ and $x \in \mathbb{R}_+$, define $\mathcal{A}y(x) := -xy(x)$. Let us start from the generator of the $(\theta = 0)$-SPDE (5.3.1),

$$Af_y(\lambda) = f_y(\lambda)\left(\langle \mathcal{A}y, \lambda\rangle - \beta\langle y, \nu\rangle\overline{\lambda} + \frac{\sigma^2}{2}\langle y, \nu\rangle^2\overline{\lambda}\right),$$

where f_y is defined in (5.3.3) and $\lambda \in \mathcal{E}$. The corresponding generalised Feller semigroup is defined on $\mathcal{B}^\varrho(\mathcal{E})$. The generator we aim at (with $\theta > 0$) is

$$A^\theta f_y(\lambda) := f_y(\lambda)\left(\langle \mathcal{A}y, \lambda\rangle - \beta(\langle y, \nu\rangle - \theta)\overline{\lambda} + \frac{\sigma^2}{2}\langle y, \nu\rangle^2\overline{\lambda}\right) =: (A+B)f_y(\lambda).$$

Inspired by the proof of [73, Proposition 3.3], we define for all $n \in \mathbb{N}$ the approximate sequence of generators $B^n f_y(\lambda) := \beta\theta f_y(\lambda)\overline{\lambda}\frac{n}{\varrho(\lambda)\vee n}$, which is bounded by $n\,|\beta\theta|\,\|f_y\|_\infty$ because $\varrho(\lambda) = 1 + \sup_{y\in Y}\langle y, \lambda\rangle$. Hence, for all $n \in \mathbb{N}$, Theorem 2.10 of [73] declares that $A^n := A + B^n$ is the generator of a generalised Feller semigroup on $\mathcal{B}^\varrho(\mathcal{E})$. We call λ^n and P^n the associated process and semigroup, respectively.

This paragraph borrows elements from the proof of [73, Proposition 4.14]. Let us fix $n \in \mathbb{N}$, $\lambda \in \mathcal{E}$ and $\lambda_0^n = \lambda$. Define $\mathcal{E}_\star := \{y \in Y : \langle y, \lambda\rangle \leq 0 \text{ for all } \lambda \in \mathcal{E}\}$, the polar cone of $\mathcal{E}$, and set $y_0 \in \mathcal{E}_\star$. Then $\mathbb{E}[\exp(\langle y_0, \lambda_t^n\rangle)]$ solves uniquely the Kolmogorov PDE

$$\partial_t u(t,\lambda) = A^n u(t,\lambda), \quad u(0,\lambda) = \exp(\langle y_0, \lambda\rangle). \tag{5.3.4}$$

We introduce the function y^n, which is such that $y_t^n \in Y$ for all $t \geq 0$, $y_0^n = y_0$, and follows the dynamics

$$\partial_t y_t^n = R^n(y_t^n), \quad \text{where } R^n(y)(x) := \left(\mathcal{A}y(x) - \beta\langle y, \nu\rangle + \frac{\sigma^2}{2}\langle y, \nu\rangle^2\right)\overline{\lambda} + \beta\theta\overline{\lambda}\frac{n}{\varrho(\lambda)\vee n}.$$

Hence $u(t,\lambda) := f_{y_t^n}(\lambda) = \exp(\langle y_t^n, \lambda\rangle)$ solves (5.3.4) since $\partial_t u(t,\lambda) = \langle \partial_t y_t^n, \lambda\rangle u(t,\lambda) = A^n u(t,\lambda)$. We just showed that $\exp(\langle y_t^n, \lambda\rangle) = \mathbb{E}[\exp(\langle y_0, \lambda_t^n\rangle)] \leq 1$, hence $y_t^n \in \mathcal{E}_\star$ for all $t \geq 0$. Recall that D is the set of the Fourier elements (5.3.3) equipped with the supremum norm $\|\cdot\|_\infty$. Therefore P^n leaves D invariant, and $\mathrm{span}(D)$ is dense in $\bigcap_{n\in\mathbb{N}}\mathrm{dom}(A^n) = \mathcal{B}^\varrho(\mathcal{E})$. This grants condition (i) of [73, Theorem 3.2].

We conclude the proof with an argument from the proof of [73, Proposition 4.15]. For any $m \in \mathbb{N}$, $t \geq 0$, $y \in \mathcal{E}_\star$ and $\lambda \in \mathcal{E}$, we showed that

$$P_t^m f_y(\lambda) = \mathbb{E}[\exp(\langle y, \lambda_t^m \rangle)] = \exp(\langle y_t^m, \lambda \rangle) = f_{y_t^m}(\lambda),$$

with $y_0^m = y$ and $\lambda_0^m = \lambda$. Since $y_t^m \in \mathcal{E}_\star$ it is clear that $\left\| f_{y_t^m} \right\|_\infty \leq 1$, and therefore we can compute

$$\frac{|A^n P_t^m f_y(\lambda) - A^m P_t^m f_y(\lambda)|}{\varrho(\lambda)} \leq \beta\theta \left\| f_{y_t^m} \right\|_\infty \frac{|\overline{\lambda}|}{\varrho(\lambda)} \left| \frac{n}{\varrho(\lambda) \vee n} - \frac{m}{\varrho(\lambda) \vee m} \right| \leq a_{nm}^y \left\| f_y \right\|_\infty,$$

where a_{nm}^y tends to zero as $n, m \uparrow +\infty$. Hence, condition (ii) of Theorem 3.2 holds and we conclude that $A^\theta = \lim_{n \uparrow \infty} A^n$ is the generator of a generalised Feller semigroup on $\mathcal{B}^\varrho(\mathcal{E})$. $\square$

5.4 Verifying the condition for existence

We proved in the previous section that the solution of (5.3.1) is indeed a generalised Feller process, hence all that is left to apply Proposition 5.2.1 is to show that the bound

$$\sup_{t \geq 0} \mathbb{E}[\|\lambda_t\|_{Y_\star}] < \infty \tag{5.4.1}$$

holds. A simple application of Proposition 5.2.1 then yields Theorem 5.1.4. Recall that $V_t = \langle 1, \lambda_t \rangle$ for all $t \geq 0$. The main result of this section is the following.

Proposition 5.4.1. *Let $\lambda_0(\mathrm{d}x) = V_0\, \delta_0(\mathrm{d}x)$ and $\nu(\mathrm{d}x) = \left(x^\alpha \Gamma(1-\alpha)\Gamma(\alpha)\right)^{-1} \mathrm{d}x$ for some $V_0 > 0$, $\alpha \in (1/2, 3/2)$, so that $K(t) = t^{\alpha-1}/\Gamma(\alpha)$ is the Riemann-Liouville kernel. Then the solution $(\lambda_t)_{t\geq 0}$ to (5.3.1) satisfies (5.4.1).*

Remark 5.4.2. We make some comments on the possible choices of initial condition λ_0; recall that λ_0 is admissible if it belongs to $\mathcal{E}$. By Assumption 5.2 and Remark 5.3 in [73], $\mathrm{e}^{-r\cdot}\nu \in \mathcal{E}$ for all $r > 0$. By Remark 5.4 in [73] and Example 2.2 in [1], are allowed $\lambda_0 \in Y^\star$ such that $t \mapsto \int_0^\infty \mathrm{e}^{-tx} \lambda_0(\mathrm{d}x) \in \mathcal{G}_K$, where $\mathcal{G}_K$ includes in particular Hölder continuous, non-decreasing functions g with $g(0) \geq 0$, and functions of the form $g = V_0 + K * \theta$ such that $\theta(s)\, \mathrm{d}s + V_0 L(\mathrm{d}s)$ is a non-negative measure, where $V_0 > 0$ and L is the resolvent of first kind of K. This yields at least two options:

- $\lambda_0(\mathrm{d}x) = V_0\, \delta_0(\mathrm{d}x)$, with $V_0 > 0$;

- $\lambda_0(\mathrm{d}x) = V_0\,\delta_0(\mathrm{d}x) + x^{-\alpha-\mu}\,\mathrm{d}x$ where $0 < \mu < 1-\alpha$, such that

$$\int_0^\infty \mathrm{e}^{-xt}\lambda_0(\mathrm{d}x) = V_0+\Gamma(1-\alpha-\mu)t^{\alpha+\mu-1} = V_0+\frac{\Gamma(1-\alpha-\mu)\Gamma(\alpha+\mu)}{\Gamma(\alpha)\Gamma(\mu)}\int_0^t (t-s)^{\alpha-1}s^{\mu-1}\,\mathrm{d}s,$$

which is equal to $V_0 + (K*\theta)(t)$ with θ and L non-negative.

Proposition 5.4.1 unfolds directly from the two following lemmas.

Lemma 5.4.3. *Suppose $\lambda_0 \in \mathcal{E}$ and $\nu \in Y^\star$ are such that ν and λ_0 are non-negative and the function $t \mapsto \mathbb{E}[V_t] - \theta$ keeps the same sign for all $t \geq 0$. Then the solution $(\lambda_t)_{t\geq 0}$ to (5.3.1) satisfies (5.4.1).*

Lemma 5.4.4. *Let $\lambda_0(\mathrm{d}x) = V_0\,\delta_0(\mathrm{d}x)$ and $\nu(\mathrm{d}x) = \big(x^\alpha\Gamma(1-\alpha)\Gamma(\alpha)\big)^{-1}\mathrm{d}x$ for some $V_0 > 0$, $\alpha \in (1/2, 3/2)$, so that $K(t) = t^{\alpha-1}/\Gamma(\alpha)$ is the Riemann-Liouville kernel. The function $t \mapsto \mathbb{E}[V_t] - \theta$ keeps the same sign for all $t \geq 0$.*

Proof. Denoting $v(t) = \int_0^\infty \mathrm{e}^{-tx}\lambda_0(\mathrm{d}x)$ and following [3, Lemma 4.2], we have

$$\mathbb{E}[V_t] = v(t) - \int_0^t v(t-s)f^{\alpha,\beta}(s)\,\mathrm{d}s, \tag{5.4.2}$$

where $f^{\alpha,\beta}(s) = \beta s^{\alpha-1}\mathrm{E}_{\alpha,\alpha}(-\beta s^\alpha)$, and $\mathrm{E}_{\alpha,\zeta}(z) = \sum_{n=0}^\infty \frac{z^n}{\Gamma(\alpha n+\zeta)}$ is the Mittag-Leffler function. In our setting this yields

$$\mathbb{E}[V_t] = V_0 - (V_0-\theta)\int_0^t f^{\alpha,\beta}(s)\,\mathrm{d}s,$$

and

$$\begin{aligned}\int_0^t f^{\alpha,\beta}(s)\,\mathrm{d}s = \beta\sum_{n=0}^\infty \frac{(-\beta)^n}{\Gamma(\alpha(n+1))}\int_0^t s^{\alpha(n+1)-1}\,\mathrm{d}s &= \beta\sum_{n=0}^\infty \frac{(-\beta)^n\,t^{\alpha(n+1)}}{\Gamma(\alpha(n+1))\,\alpha(n+1)}\\ &= \beta t^\alpha \mathrm{E}_{\alpha,\alpha+1}(-\beta t^\alpha).\end{aligned}$$

We write $f(z) \sim g(z)$ if $f(z)/g(z)$ tends to one as z goes to some limit. From Lemma 1.1 in [234] we know that, as z goes to $-\infty$

$$\mathrm{E}_{\alpha,\beta}(z) \sim -\frac{z^{-1}}{\Gamma(\beta-\alpha)}.$$

Hence, as t goes to $+\infty$, this entails

$$\beta t^\alpha \mathrm{E}_{\alpha,\alpha+1}(-\beta t^\alpha) \sim \beta t^\alpha\left(-\frac{(-\beta t^\alpha)^{-1}}{\Gamma(1)}\right) = 1.$$

Therefore $\mathbb{E}[V_t]$ goes from V_0 to θ, and we wonder whether $\mathbb{E}[V_t]$ passes on the other side of θ. Notice that $\mathbb{E}[V_t]$ satisfies the Volterra equation

$$x_t = x_0 + \int_0^t K(t-s)\beta(\theta - x_s)\,\mathrm{d}s. \tag{5.4.3}$$

Let $x_0 < \theta$ and $\tau_1 := \inf\{t > 0 : x_t > \theta\}$ and suppose there exists a non-empty interval $(\tau_1, \tau_1 + h]$ where $x_t > \theta$. Then

$$x_{\tau_1+h} = \left[x_0 + \int_0^{\tau_1} K(\tau_1 + h - s)\beta(\theta - x_s)\,\mathrm{d}s\right] + \int_{\tau_1}^{\tau_1+h} K(\tau_1 + h - s)\beta(\theta - x_s)\,\mathrm{d}s, \tag{5.4.4}$$

where the first term is smaller than $x_{\tau_1} = \theta$ because K is decreasing and $\theta - x_s > 0$, and the second term is negative because $x_s > \theta$ by assumption. Hence $x_{\tau_1+h} < \theta$ which is a contradiction.

Now let $x_0 > \theta$, $\tau_2 := \inf\{t > 0 : x_t < \theta\}$ and suppose there exists a non-empty interval $(\tau_2, \tau_2 + h]$ where $x_t < \theta$. The signs are reversed compared to the previous situation. Consider (5.4.4) with τ_2 instead of τ_1: the first term is greater than $x_{\tau_2} = \theta$, and the second term is positive. Again, this yields a contradiction. Overall this entails that $\theta - x_s$ has a constant sign for all $s \geq 0$. $\square$

Proof of Lemma 5.4.3. Let $\mathfrak{B}(\overline{\mathbb{R}}_+)$ denote the Borel subsets of $\overline{\mathbb{R}}_+$. For a signed measure λ on $\overline{\mathbb{R}}_+$, define the upper variation $\mathfrak{U}$ (resp. lower variation $\mathfrak{L}$) as $\mathfrak{U}(\lambda) := \sup\{\lambda(A) : A \in \mathfrak{B}(\overline{\mathbb{R}}_+)\}$ (resp. inf), and such that the total variation corresponds to $\|\lambda\|_{Y^\star} = \mathfrak{U}(\lambda) - \mathfrak{L}(\lambda)$.

Let us fix $t > 0$, then there exist two increasing sequences of sets $(U_n)_{n\geq 1}$ and $(L_n)_{n\geq 1}$, both in $\mathfrak{B}(\overline{\mathbb{R}}_+)$, such that $\lambda_t(U_n)$ is non-negative for all $n \in \mathbb{N}$ and increases towards $\mathfrak{U}(\lambda_t)$ as n goes to $+\infty$, and analogously, $\lambda_t(L_n)$ is non-positive and decreases to $\mathfrak{L}(\lambda_t)$. We will use the representation

$$\|\lambda_t\|_{Y^\star} = \lim_{n\uparrow\infty} \big(\lambda_t(U_n) - \lambda_t(L_n)\big) = \lim_{n\uparrow\infty} \left(\int_{U_n} \lambda_t(\mathrm{d}x) - \int_{L_n} \lambda_t(\mathrm{d}x)\right). \tag{5.4.5}$$

Intuitively, one should think of the limit of U_n (resp. L_n) as the subset of $\overline{\mathbb{R}}_+$ on which λ_t is positive (resp. negative). This way, the function maximising $\sup_{\|y\|_\infty \leq 1}\langle y, \lambda_t\rangle$ (if it exists) corresponds to the limit of $\mathbb{1}_{U_n} - \mathbb{1}_{L_n}$.

We did not assume that $\|\lambda_t\|_{Y^\star}$ should be finite. Yet, we can apply the monotone convergence theorem on the representation (5.4.5)

$$\begin{aligned}\mathbb{E}[\|\lambda_t\|_{Y^\star}] &= \lim_{n\uparrow+\infty}\mathbb{E}\left[\int_{U_n}\lambda_t(\mathrm{d}x)-\int_{L_n}\lambda_t(\mathrm{d}x)\right]\\ &= \lim_{n\uparrow+\infty}\mathbb{E}\left[\int_0^\infty(\mathbb{1}_{U_n}-\mathbb{1}_{L_n})\left(\mathrm{e}^{-xt}\lambda_0(\mathrm{d}x)+\int_0^t \mathrm{e}^{-x(t-s)}\,\mathrm{d}X_s\,\nu(\mathrm{d}x)\right)\right].\end{aligned}$$

Let $Y_t := \int_0^t K(t-s)\sigma\sqrt{V_s}\,\mathrm{d}W_s$. Recalling the SVE (5.1.1), variation of constants on $V_t-\theta$ yields

$$V_t-\theta = (V_0-\theta+Y_t)-R_\beta*(V_0-\theta+Y_t) = \big(\mathbb{E}[V_t]-\theta\big)+\left(Y_t-\int_0^t R_\beta(t-s)Y_s\,\mathrm{d}s\right),$$

where R_β is the resolvent of the second kind of $-\beta K$. This implies

$$\begin{aligned}\mathbb{E}\big[(\mathbb{1}_{U_n}-\mathbb{1}_{L_n})(\theta-V_t)\big] &= \mathbb{E}[\mathbb{1}_{U_n}-\mathbb{1}_{L_n}](\theta-\mathbb{E}[V_t])-\mathbb{E}\left[\int_0^t(\mathbb{1}_{U_n}-\mathbb{1}_{L_n})K(t-s)\sigma\sqrt{V_s}\,\mathrm{d}W_s\right]\\ &\quad+\int_0^t R_\beta(t-s)\mathbb{E}\left[\int_0^s(\mathbb{1}_{U_n}-\mathbb{1}_{L_n})K(s-r)\sigma\sqrt{V_r}\,\mathrm{d}W_r\right]\mathrm{d}s, \qquad (5.4.6)\end{aligned}$$

where the second and third terms are equal to zero by the property of the local martingale. Similarly, we note that

$$\mathbb{E}\left[\int_0^t \mathrm{e}^{-x(t-s)}(\mathbb{1}_{U_n}-\mathbb{1}_{L_n})\sigma\sqrt{V_s}\,\mathrm{d}W_s\right]=0. \qquad (5.4.7)$$

By (5.4.6) and (5.4.7) we deduce

$$\begin{aligned}\mathbb{E}\left[\int_0^\infty(\mathbb{1}_{U_n}-\mathbb{1}_{L_n})\int_0^t \mathrm{e}^{-x(t-s)}\,\mathrm{d}X_s\,\nu(\mathrm{d}x)\right] &= \int_0^\infty\int_0^t \mathrm{e}^{-x(t-s)}\mathbb{E}\big[(\mathbb{1}_{U_n}-\mathbb{1}_{L_n})\beta(\theta-V_s)\big]\,\mathrm{d}s\\ &= \mathbb{E}\left[\left(\int_{U_n}-\int_{L_n}\right)\int_0^t \mathrm{e}^{-x(t-s)}(\theta-\mathbb{E}[V_s])\,\mathrm{d}s\right].\end{aligned}$$

We want to recover $\int_{U_n}+\int_{L_n}$ which tends to $\int_0^\infty$. In the case $\theta<V_0$ (thus $\theta<\mathbb{E}[V_s]$ for all $s\geq 0$) we have

$$\int_{U_n}\int_0^t \mathrm{e}^{-x(t-s)}\beta(\theta-\mathbb{E}[V_s])\,\mathrm{d}s\,\nu(\mathrm{d}x)\leq\int_{U_n}\int_0^t \mathrm{e}^{-x(t-s)}\beta(\mathbb{E}[V_s]-\theta)\,\mathrm{d}s\,\nu(\mathrm{d}x),$$

and since λ_0 is non-negative this yields

$$\left(\int_{U_n} - \int_{L_n}\right)\left(\mathrm{e}^{-xt}\lambda_0(\mathrm{d}x) + \int_0^t \mathrm{e}^{-x(t-s)}\beta(\theta - \mathbb{E}[V_s])\,\mathrm{d}s\,\nu(\mathrm{d}x)\right)$$
$$\leq \left(\int_{U_n} + \int_{L_n}\right)\left(\mathrm{e}^{-xt}\lambda_0(\mathrm{d}x) + \int_0^t \mathrm{e}^{-x(t-s)}\beta(\mathbb{E}[V_s] - \theta)\,\mathrm{d}s\,\nu(\mathrm{d}x)\right),$$

Recalling that $\mathbb{1}_{U_n} + \mathbb{1}_{L_n}$ converges increasingly to one, the monotone convergence theorem gives us

$$\lim_{n\uparrow\infty}\mathbb{E}\left[\left(\int_{U_n} + \int_{L_n}\right)\left(\mathrm{e}^{-xt}\lambda_0(\mathrm{d}x) + \int_0^t \mathrm{e}^{-x(t-s)}\beta(\mathbb{E}[V_s] - \theta)\,\mathrm{d}s\,\nu(\mathrm{d}x)\right)\right]$$
$$= \int_0^\infty \mathrm{e}^{-xt}\lambda_0(\mathrm{d}x) + \int_0^t K(t-s)\beta(\mathbb{E}[V_s] - \theta)\,\mathrm{d}s = 2\int_0^\infty \mathrm{e}^{-xt}\lambda_0(\mathrm{d}x) - \mathbb{E}[V_t].$$

The case $\theta \geq V_0$ (hence $\theta > \mathbb{E}[V_s]$ for all $s \geq 0$) is very similar. First notice that

$$-\int_{L_n}\int_0^t \mathrm{e}^{-x(t-s)}\beta(\theta - \mathbb{E}[V_s])\,\mathrm{d}s\,\nu(\mathrm{d}x) \leq \int_{L_n}\int_0^t \mathrm{e}^{-x(t-s)}\beta(\theta - \mathbb{E}[V_s])\,\mathrm{d}s\,\nu(\mathrm{d}x),$$

which yields

$$\lim_{n\uparrow+\infty}\mathbb{E}\left[\left(\int_{U_n} - \int_{L_n}\right)\left(\mathrm{e}^{-xt}\lambda_0(\mathrm{d}x) + \int_0^t \mathrm{e}^{-x(t-s)}\beta(\theta - \mathbb{E}[V_s])\,\mathrm{d}s\,\nu(\mathrm{d}x)\right)\right]$$
$$\leq \int_0^\infty \mathrm{e}^{-xt}\lambda_0(\mathrm{d}x) + \int_0^t K(t-s)\beta(\theta - \mathbb{E}[V_s])\,\mathrm{d}s = \mathbb{E}[V_t].$$

Since $\mathbb{E}[V_t]$ is uniformly bounded over all $t \geq 0$, this proves that condition (5.4.1) holds. $\square$

Remark 5.4.5. In the case where λ_0 is non-negative and $\theta \geq V_0$, the only possible source of negative variation comes from the local martingale part, which has zero expectation. This explains the rather counter-intuitive conclusion $\mathbb{E}[\|\lambda_t\|_{Y^\star}] \leq \mathbb{E}[\langle 1, \lambda_t\rangle]$.

5.5 Outlook

Extending our results to the rough Heston model will require to cast our analysis into a multidimensional framework with $Y = \mathcal{C}_b(\mathbb{R}_+^2)$.

Building on the generalised Feller property, we aim at showing the uniqueness of the invariant measure. The analogue of the strong Feller property is granted in our setting since $\mathcal{B}^\varrho(\boldsymbol{X})$ already includes the space of bounded functions; uniqueness shall follow if $(\lambda_t)_{t\geq 0}$ satisfies a sort of recurrence property. One can also hope to characterise the unique invariant measure using tools from Malliavin calculus or the form of the Laplace transform. Once ergodicity is established, it is interesting to identify the rate of convergence and how it varies with α, thereby linking roughness of the volatility and convergence to the stationary measure.

Bibliography

[1] ABI JABER, E. AND EL EUCH, O. Markovian structure of the Volterra Heston model. *Statistics and Probability Letters*, **149** (2018). `doi:10.1016/j.spl.2019.01.024`.

[2] ABI JABER, E. AND EL EUCH, O. Multi-factor approximation of rough volatility models. *SIAM Journal on Financial Mathematics*, **10** (2019). `doi:10.1137/18M1170236`.

[3] ABI JABER, E., LARSSON, M., AND PULIDO, S. Affine Volterra processes. *The Annals of Applied Probability*, **29** (2017). `doi:10.1214/19-AAP1477`.

[4] ADLER, R. J. AND TAYLOR, J. E. *Random Fields and Geometry*. Springer, New York (2007). `doi:10.1007/978-0-387-48116-6`.

[5] AKDOGAN, O. Vol-of-vol expansion for (rough) stochastic volatility models (2018). Preprint, `arXiv:1910.03245`.

[6] ALÒS, E., GARCÍA-LORITE, D., AND MUGURUZA, A. On smile properties of volatility derivatives and exotic products: understanding the VIX skew (2018). Preprint, `arXiv:1808.03610`.

[7] ALÒS, E. AND LEÓN, J. A. On the curvature of the smile in stochastic volatility models. *SIAM Journal on Financial Mathematics*, **8** (2017). `doi:10.1137/16M1086315`.

[8] ALÒS, E., LEÓN, J. A., AND VIVES, J. On the short-time behavior of the implied volatility for jump-diffusion models with stochastic volatility. *Finance and Stochastics*, **11** (2007). `doi:10.1007/s00780-007-0049-1`.

[9] ALÒS, E. AND SHIRAYA, K. Estimating the Hurst parameter from

short term volatility swaps. *Finance and Stochastics*, **23** (2019). doi:10.1007/s00780-019-00384-5.

[10] ARNOLD, L. *Random Dynamical Systems*. Springer-Verlag, Berlin (1998). doi:10.1007/978-3-662-12878-7.

[11] AZMOODEH, E. AND NOURDIN, I. Almost sure limit theorems on Wiener chaos: the non-central case. *Electronic Communications in Probability*, **24** (2019). doi:10.1214/19-ECP212.

[12] BACHELIER, L. Théorie de la spéculation. *Annales Scientifiques de l'École Normale Supérieure*, **17** (1900). numdam.org/item/10.24033/asens.476.pdf.

[13] BAI, S. AND TAQQU, M. S. Limit theorems for long-memory flows on Wiener chaos. *Bernoulli*, **26** (2020). doi:10.3150/19-BEJ1168.

[14] BALDEAUX, J. AND BADRAN, A. Consistent modelling of VIX and equity derivatives using a 3/2 plus jumps model. *Applied Mathematical Finance*, **21** (2014). doi:10.1080/1350486X.2013.868631.

[15] BARLETTA, A., NICOLATO, E., AND PAGLIARANI, S. The short-time behavior of VIX-implied volatilities in a multifactor stochastic volatility framework. *Mathematical Finance*, **29** (2018). doi:10.1111/mafi.12196.

[16] BARNDORFF-NIELSEN, O. E., PAKKANEN, M. S., AND SCHMIEGEL, J. Assessing relative volatility/intermittency/energy dissipation. *Electronic Journal of Statistics*, **8** (2014). doi:10.1214/14-EJS942.

[17] BARNDORFF-NIELSEN, O. E. AND SCHMIEGEL, J. Brownian semistationary processes and volatility/intermittency. *Radon Series on Computational and Applied Mathematics*, **8** (2009). doi:10.1515/9783110213140.1.

[18] BAYER, C., FRIZ, P. K., GASSIAT, P., MARTIN, J., AND STEMPER, B. A regularity structure for rough volatility. *Mathematical Finance*, **30** (2019). doi:10.1111/mafi.12233.

[19] BAYER, C., FRIZ, P. K., AND GATHERAL, J. Pricing under rough volatility. *Quantitative Finance*, **16** (2016). doi:10.1080/14697688.2015.1099717.

[20] BAYER, C., FRIZ, P. K., GULISASHVILI, A., HORVATH, B., AND STEMPER, B. Short-time near-the-money skew in rough fractional volatility models. *Quantitative Finance*, **19** (2019). doi:10.1080/14697688.2018.1529420.

[21] BAYER, C., HALL, E. J., AND TEMPONE, R. Weak error rates for option pricing under the rough Bergomi model (2020). Preprint, `arXiv:2009.01219`.

[22] BAYER, C., HARANG, F. A., AND PIGATO, P. Log-modulated rough stochastic volatility models (2020). Preprint, `arXiv:2008.03204`.

[23] BAYER, C., HORVATH, B., MUGURUZA, A., STEMPER, B., AND TOMAS, M. On deep calibration of (rough) stochastic volatility models (2019). Preprint, `arXiv:1908.08806`.

[24] BELL, D., BOLAÑOS, R., AND NUALART, D. Limit theorems for singular Skorohod integrals. *Theory of Probability and Mathematical Statistics*, **102** (2020), 21. `doi:10.1090/tpms/1126`.

[25] BENNEDSEN, M., LUNDE, A., AND PAKKANEN, M. Hybrid scheme for Brownian semistationary processes. *Finance and Stochastics*, **21** (2017). `doi:10.1007/s00780-017-0335-5`.

[26] BENNEDSEN, M., LUNDE, A., AND PAKKANEN, M. Decoupling the short- and long-term behavior of stochastic volatility. *Journal of Financial Econometrics*, (2021). `doi:10.1093/jjfinec/nbaa049`.

[27] BERESTYCKI, H., BUSCA, J., AND FLORENT, I. Computing the implied volatility in stochastic volatility models. *Communications in Pure and Applied Mathematics*, **57** (2004). `doi:10.1002/cpa.20039`.

[28] BERGER, M. AND MIZEL, V. Volterra equations with Itô integrals I. *Journal of Integral Equations*, **2** (1980).

[29] BERGER, M. AND MIZEL, V. Volterra equations with Itô integrals II. *Journal of Integral Equations*, **2** (1980).

[30] BERGOMI, L. Smile dynamics II. *Risk*, (2005). `doi:10.2139/ssrn.1493302`.

[31] BERGOMI, L. Smile dynamics III. *Risk*, (2008). `doi:10.2139/ssrn.1493308`.

[32] BESSAIH, H. AND MILLET, A. Large deviation principle and inviscid shell models. *Electronic Journal of Probability*, **14** (2009). `doi:10.1214/EJP.v14-719`.

[33] BHAMIDI, S., BUDHIRAJA, A., DUPUIS, P., AND WU, R. Rare event asymptotics for exploration processes for random graphs (2019). Preprint, `arXiv:1912.04714`. Forthcoming in *The Annals of Applied Probability*.

[34] BIAGINI, F., HU, Y., ØKSENDAL, B., AND ZHANG, T. *Stochastic Calculus for Fractional Brownian Motion and Applications*. Springer-Verlag London (2008). doi:10.1007\%2F978-1-84628-797-8.

[35] BILLINGSEY, P. *Convergence of probability measures*. John Wiley & Sons, Inc., New York (1999). doi:10.1002/9780470316962.

[36] BLACK, F. AND SCHOLES, M. The pricing of options and corporate liabilities. *Journal of Political Economy*, **81** (1973). doi:10.1086/260062.

[37] BONESINI, O., CALLEGARO, G., AND JACQUIER, A. Functional quantization of rough volatility and applications to the VIX (2021). Preprint, arXiv:2104.04233.

[38] BOUÉ, M. AND DUPUIS, P. A variational representation for certain functionals of Brownian motion. *The Annals of Probability*, **26** (1998). doi:10.1214/aop/1022855876.

[39] BOURGEY, F. AND DE MARCO, S. Multilevel Monte Carlo simulation for VIX options in the rough Bergomi model. (2021). Preprint, arXiv:2105.05356.

[40] BOURGUIN, S., GAILUS, S., AND SPILIOPOULOS, K. Typical dynamics and fluctuation analysis of slow-fast systems driven by fractional Brownian motion (2019). Preprint, arXiv:1906.02131.

[41] BREUER, P. AND MAJOR, P. Central limit theorems for non-linear functionals of Gaussian fields. *Journal of Multivariate Analysis*, **13** (1983). doi:10.1016/0047-259X(83)90019-2.

[42] BROWN, R. A brief account of microscopical observations made in the months of June, July and August, 1827, on the particles contained in the pollen of plants; and on the general existence of active molecules in organic and inorganic bodies. *Philosophical Magazine*, **4** (1828). doi:10.1080/14786442808674769.

[43] BRYC, W. Large deviations by the asymptotic value method. *Diffusion processes and related problems in analysis*, (1990). doi:10.1007/978-1-4684-0564-4_25.

[44] BUDHIRAJA, A., CHEN, J., AND DUPUIS, P. Large deviations for stochastic partial differential equations driven by a Poisson random measure. *Stochastic Processes and Applications*, **123** (2013). doi:10.1016/j.spa.2012.09.010.

[45] BUDHIRAJA, A. AND DUPUIS, P. A variational representation for positive functionals of infinite dimensional Brownian motion. *Probability and Mathematical Statistics*, **20** (2001).

[46] BUDHIRAJA, A. AND DUPUIS, P. *Analysis and Approximation of Rare Events*. Springer, New York (2019). doi:10.1007\%2F978-1-4939-9579-0.

[47] BUDHIRAJA, A., DUPUIS, P., AND GANGULY, A. Moderate deviations principles for stochastic differential equations with jumps. *The Annals of Probability*, **44** (2016). doi:10.1214/15-AOP1007.

[48] BUDHIRAJA, A., DUPUIS, P., AND GANGULY, A. Large deviations for small noise diffusions in a fast Markovian environment. *Electronic Journal of Probability*, **23** (2018). doi:10.1214/18-EJP228.

[49] BUDHIRAJA, A., DUPUIS, P., AND MAROULAS, V. Large deviations for infinite dimensional stochastic dynamical systems. *The Annals of Probability*, **36** (2008). doi:10.1214/07-AOP362.

[50] BUDHIRAJA, A., DUPUIS, P., AND MAROULAS, V. Variational representations for continuous time processes. *Annales de l'Institut Henri Poincaré - Probabilités et Statistiques.*, **47** (2011). doi:10.1214/10-AIHP382.

[51] BUDHIRAJA, A., FRIEDLANDER, E., AND WU, R. Many-server asymptotics for join-the-shortest-queue: Large deviations and rare events (2020). Preprint, arXiv:1904.04938. Forthcoming in *The Annals of Applied Probability*.

[52] BUDHIRAJA, A. AND SONG, X. Large deviation principles for stochastic dynamical systems with a fractional Brownian noise (2020). Preprint, arXiv:2006.07683.

[53] BUTKOVSKY, O., KULIK, A., AND SCHEUTZOW, M. Generalized couplings and ergodic rates for SPDEs and other Markov models. *The Annals of Applied Probability*, **30** (2020). doi:10.1214/19-AAP1485.

[54] CAMPESE, S., NOURDIN, I., AND NUALART, D. Continuous Breuer–Major theorem: Tightness and nonstationarity. *The Annals of Probability*, **48** (2020). doi:10.1214/19-AOP1357.

[55] CARR, P. AND MADAN, D. Joint modeling of VIX and SPX options at a single and common maturity with risk management applications. *IIE Transactions*, **46** (2014). doi:10.1080/0740817X.2013.857063.

[56] CELLUPICA, M. AND PACCHIAROTTI, B. Pathwise asymptotics for Volterra type rough volatility models. *Journal of Theoretical Probability*, (2020). doi:10.1007/s10959-020-00992-4.

[57] CHEN, L., KHOSHNEVISAN, D., NUALART, D., AND PU, F. Spatial ergodicity for SPDEs via Poincaré-type inequalities (2019). Preprint: arXiv:1907.11553.

[58] CHEN, Z.-Q., REN, Y.-X., AND YANG, T. Law of large numbers for branching symmetric Hunt processes with measure-valued branching rates. *Journal of Theoretical Probability*, **30** (2017). doi:10.1007/s10959-016-0671-y.

[59] CHEVALIER, E., PULIDO, S., AND ZUNIGA, E. American options in the Volterra Heston model (2021). Preprint, arXiv:2103.11734.

[60] CHEVILLARD, L. Regularized fractional Ornstein-Uhlenbeck processes and their relevance to the modeling of fluid turbulence. *Physical Review E*, **97** (2017). doi:10.1103/PhysRevE.96.033111.

[61] CHEVYREV, I., FRIZ, P. K., KOREPANOV, A., MELBOURNE, I., AND ZHANG, H. Multiscale systems, homogenization, and rough paths. In *Probability and Analysis in Interacting Physical Systems* (edited by P. K. Friz, W. König, C. Mukherjee, and S. Olla). Springer International Publishing, Cham (2019). ISBN 978-3-030-15338-0. doi:10.1007/978-3-030-15338-0_2.

[62] CHIARINI, F. AND FISCHER, M. On large deviations for small noise Itô processes. *Advances in Applied Probability*, **46** (2014). doi:10.1239/aap/1418396246.

[63] CHRONOPOULOU, A. AND VIENS, F. G. Estimation and pricing under long-memory stochastic volatility. *Annals of Finance*, **8** (2012). doi:10.1007/s10436-010-0156-4.

[64] CHRONOPOULOU, A. AND VIENS, F. G. Stochastic volatility and option pricing with long-memory in discrete and continuous time. *Quantitative Finance*, **12** (2012). doi:10.1080/14697688.2012.664939.

[65] COMTE, E., COUTIN, L., AND RENAULT, E. Affine fractional stochastic volatility models with application to option pricing. *Annals of Finance*, **8** (2012). doi:10.1007/s10436-010-0165-3.

[66] COMTE, E. AND RENAULT, E. Long memory in continuous-time stochastic volatility models. *Mathematical Finance*, **8** (1998). doi:10.1111/1467-9965.00057.

[67] CONFORTI, G., DE MARCO, S., AND DEUSCHEL, J.-D. On small-noise equations with degenerate limiting system arising from volatility models. *In: Friz P. K., Gatheral J., Gulisashvili A., Jacquier A., Teichmann J. (eds) Large Deviations and Asymptotic Methods in Finance. Springer Proceedings in Mathematics & Statistics*, (2015). doi:10.1007/978-3-319-11605-1_17.

[68] CONT, R. AND KOKHOLM, T. A consistent pricing model for index options and volatility derivatives. *Mathematical Finance*, **23** (2013). doi:10.1111/j.1467-9965.2011.00492.x.

[69] CORCUERA, J. M., HEDEVANG, E., PAKKANEN, M. S., AND PODOLSKIJ, M. Asymptotic theory for Brownian semi-stationary processes with application to turbulence. *Stochastic Processes and their Applications*, **123** (2017). doi:10.1016/j.spa.2013.03.011.

[70] COUTIN, L. AND DECREUSEFOND, L. Abstract nonlinear filtering theory in the presence of fractional Brownian motion. *The Annals of Applied Probability*, **9** (1999). doi:10.1214/aoap/1029962865.

[71] COUTIN, L. AND DECREUSEFOND, L. Volterra differential equations with singular kernels. *Proceedings of the Workshop on Mathematical Physics and Stochastic Analysis*, (2000).

[72] CUCHIERO, C. AND SVALUTO-FERRO, S. Infinite-dimensional polynomial processes. *Finance and Stochastics*, **25** (2021). doi:10.1007/s00780-021-00450-x.

[73] CUCHIERO, J. AND TEICHMANN, J. Generalized Feller processes and Markovian lifts of stochastic Volterra processes: the affine case. *Journal of Evolution Equations*, **20** (2020). doi:10.1007/s00028-020-00557-2.

[74] DA PRATO, G. AND ZABCZYK, J. *Stochastic Equations in Infinite Dimensions*. Cambridge University Press (1992). doi:10.1017/CBO9781107295513.

[75] DA PRATO, G. AND ZABCZYK, J. *Ergodicity for Infinite Dimensional Systems*. Cambridge University Press (1996). doi:10.1017/CBO9780511662829.

[76] DALANG, R., KHOSHNEVISAN, D., MUELLER, C., NUALART, D., AND XIAO, Y. *A Minicourse on Stochastic Partial Differential Equations*. Springer Berlin Heidelberg (2008). doi:10.1007/978-3-540-85994-9.

[77] De Marco, S. VIX derivatives in rough forward variance models. *Presentation, Bachelier Congress, Dublin*, (2018).

[78] De Marco, S. and Henry-Labordère, P. Linking vanillas and VIX options: a constrained martingale optimal transport problem. *SIAM Journal on Financial Mathematics*, **6** (2015). doi:10.1137/140960724.

[79] Decreusefond, L. Regularity properties of some stochastic Volterra integrals with singular kernels. *Potential Analysis*, **16** (2002). doi:10.1023/A:1012628013041.

[80] Decreusefond, L. and Üstünel, A. Stochastic analysis of the fractional Brownian motion. *Potential Analysis*, **10** (1999). doi:10.1023/A:1008634027843.

[81] Delgado-Vences, F., Nualart, D., and Zheng, G. A central limit theorem for the stochastic wave equation with fractional noise. *Annales de l'Institut Henri Poincaré - Probabilités et Statistiques*, **56** (2020). doi:10.1214/20-AIHP1069.

[82] Dembo, A. and Zeitouni, O. *Large Deviations Techniques and Applications*. Springer-Verlag Berlin Heidelberg (1998). doi:10.1007/978-3-642-03311-7.

[83] Deuschel, J.-D. and Stroock, D. W. *Large Deviations*. Academic Press Inc. (1989). doi:10.1090/chel/342.

[84] Donati-Martin, C., Rouault, A., Yor, M., and Zani, M. Large deviations for squares of Bessel and Ornstein-Uhlenbeck processes. *Probability Theory and Related Fields*, **129** (2004). doi:10.1007/s00440-004-0338-y.

[85] Doob, J. L. *Stochastic Processes*. New York: John Wiley and Sons (1953). doi:10.1126/science.118.3074.659.

[86] Dörsek, P. and Teichmann, J. A semigroup point of view on splitting schemes for stochastic (partial) differential equations (2010). Preprint, arXiv:1011.2651.

[87] Dupire, B. Pricing with a smile. *Risk Magazine*, (1994).

[88] Dupuis, P. and Ellis, R. S. *A Weak Convergence Approach to the Theory of Large Deviations*. Wiley (1997). doi:10.1002/9781118165904.

[89] DUPUIS, P. AND SPILIOPOULOS, K. Large deviations for multiscale diffusions via weak convergence methods. *Stochastic Processes and their Applications*, **122** (2012). doi:10.1016/j.spa.2011.12.006.

[90] DUPUIS, P., SPILIOPOULOS, K., AND WANG, H. Importance sampling for multiscale diffusions. *SIAM Journal on Multiscale Modeling and Simulation*, **10** (2012). doi:10.1137/110842545.

[91] EICHELSBACHER, P. AND THÄLE, C. Malliavin-Stein method for variance-gamma approximation on Wiener space. *Electronic Journal of Probability*, **20** (2015). doi:10.1214/EJP.v20-4136.

[92] EL EUCH, O., FUKASAWA, M., GATHERAL, J., AND ROSENBAUM, M. Short-term at-the-money asymptotics under stochastic volatility models. *SIAM Journal on Financial Mathematics*, (2019). doi:10.1137/18M1167565.

[93] EL EUCH, O., FUKASAWA, M., AND ROSENBAUM, M. The microstructural foundations of leverage effect and rough volatility. *Finance and Stochastics*, **22(2)** (2018). doi:10.1007/s00780-018-0360-z.

[94] EL EUCH, O. AND ROSENBAUM, M. Perfect hedging in rough Heston models. *The Annals of Applied Probability*, **28** (2018). doi:10.1214/18-AAP1408.

[95] EL EUCH, O. AND ROSENBAUM, M. The characteristic function of rough Heston models. *Mathematical Finance*, **29** (2019). doi:10.1111/mafi.12173.

[96] ENGLE, R. AND PATTON, A. What good is a volatility model? *Quantitative Finance*, **1** (2001). doi:10.1016/B978-075066942-9.50004-2.

[97] ETHERIDGE, A. M. Asymptotic behaviour of measure-valued critical branching processes. *Proceedings of the American mathematical society*, **118** (1993). doi:10.1090/S0002-9939-1993-1100650-X.

[98] FLANDOLI, F. AND MASLOWSKI, B. Ergodicity of the 2-D Navier-Stokes equation under random perturbations. *Communications in Mathematical Physics*, **171** (1995). doi:10.1007/BF02104513.

[99] FORDE, M., FUKASAWA, M., GERHOLD, S., AND SMITH, B. The Riemann-Liouville field and its GMC as $H \to 0$, and skew flattening for the rough Bergomi model (2020). Preprint, nms.kcl.ac.uk/martin.forde.

[100] FORDE, M., GERHOLD, S., AND SMITH, B. Small-time, large-time, and $H \to 0$ asymptotics for the rough Heston model. *Mathematical Finance*, **31** (2020). doi:10.1111/mafi.12290.

[101] FORDE, M., GERHOLD, S., AND SMITH, B. Rough Heston with jumps - small-time VIX smile and exact calibration to SPX/VIX level and skew (2021). Preprint, nms.kcl.ac.uk/martin.forde.

[102] FORDE, M., SMITH, B., AND VIITASAARI, L. Rough unbounded volatility with CGMY jumps with a finite history and the rough Heston model - small-time asymptotics in the $k\sqrt{t}$ regime. *Quantitative Finance*, **21** (2021). doi:10.1080/14697688.2020.1790634.

[103] FORDE, M. AND ZHANG, H. Asymptotics for rough stochastic volatility models. *SIAM Journal on Financial Mathematics*, **8** (2017). doi:10.1137/15M1009330.

[104] FOUQUE, J. P. AND SAPORITO, Y. F. Heston stochastic vol-of-vol model for joint calibration of VIX and S&P 500 options. *Quantitative Finance*, **18** (2018). doi:10.1080/14697688.2017.1412493.

[105] FREIDLIN, M. I. AND WENTZELL, A. D. On small random perturbations of dynamical systems. *Russian Mathematical Surveys*, **25** (1970). doi:10.1070/RM1970v025n01ABEH001254.

[106] FREIDLIN, M. I. AND WENTZELL, A. D. *Random Perturbations of Dynamical Systems*. Springer-Verlag New York (1984). doi:10.1007/978-3-642-25847-3.

[107] FRISEN, M. Long-time behavior for subcritical measure-valued branching processes with immigration (2019). Preprint, arXiv:1903.05546.

[108] FRIZ, P. K., GASSIAT, P., AND PIGATO, P. Short dated smile under rough volatility: asymptotics and numerics (2020). Preprint, arXiv:2009.08814.

[109] FRIZ, P. K., GASSIAT, P., AND PIGATO, P. Precise asymptotics: robust stochastic volatility models. *The Annals of Applied Probability*, **2** (2021). doi:10.1214/20-AAP1608.

[110] FRIZ, P. K., GATHERAL, J., GULISASHVILI, A., JACQUIER, A., AND TEICHMANN, J. *Large Deviations and Asymptotic Methods in Finance*. Springer Proceedings in Mathematics and Statistics (2015). doi:10.1007/978-3-319-11605-1.

[111] FRIZ, P. K., GERHOLD, S., AND PINTER, A. Option pricing in the moderate deviations regime. *Mathematical Finance*, **28** (2018). doi:10.1111/mafi.12156.

[112] FUKASAWA, M. Asymptotic analysis for stochastic volatility: martingale expansion. *Finance and Stochastics*, **15** (2011). doi:10.1007/s00780-010-0136-6.

[113] FUKASAWA, M. Short-time at-the-money skew and rough fractional volatility. *Quantitative Finance*, **17** (2017). doi:10.1080/14697688.2016.1197410.

[114] FUKASAWA, M. Volatility has to be rough (2020). Preprint, arXiv:2002.09215.

[115] FUKASAWA, M. AND GATHERAL, J. A rough SABR formula (2021). Preprint, arXiv:2105.05359.

[116] GANGULY, A. Large deviation principle for stochastic integrals and stochastic differential equations driven by infinite-dimensional semimartingales. *Stochastic Processes and their Applications*, **128** (2018). doi:10.1016/j.spa.2017.09.011.

[117] GAO, K. AND LEE, R. Asymptotics of implied volatility to arbitrary order. *Finance and Stochastics*, **18(2)** (2014). doi:10.1007/s00780-013-0223-6.

[118] GARCIA, J. A large deviation principle for stochastic integrals. *Journal of Theoretical Probability*, **21** (2008). doi:10.1007/s10959-007-0136-4.

[119] GARRIDO-ATIENZA, M. J., KLOEDEN, P. E., AND NEUENKIRCH, A. Discretization of stationary solutions of stochastic systems driven by fractional Brownian motion. *Applied Mathematics and Optimization*, **60** (2009). doi:10.1007/s00245-008-9062-9.

[120] GASSIAT, P. On the martingale property in the rough Bergomi model. *Electronic Communications in Probability*, **24** (2019). doi:10.1214/19-ECP239.

[121] GATHERAL, J. *The Volatility Surface: A Practitioner's Guide*. Wiley (2006).

[122] GATHERAL, J. Consistent modeling of SPX and VIX options. *Presentation, Bachelier Congress, London, July 18*, (2008).

[123] GATHERAL, J., JAISSON, T., AND ROSENBAUM, M. Volatility is rough. *Quantitative Finance*, **18** (2018). doi:10.1080/14697688.2017.1393551.

[124] GATHERAL, J., JAISSON, T., AND ROSENBAUM, M. The quadratic rough Heston model and the joint S&P 500/VIX smile calibration problem. *Risk*, (2020).

[125] GATHERAL, J. AND KELLER-RESSEL, M. Affine forward variance models. *Finance and Stochastics*, **23** (2019). doi:10.1007/s00780-019-00392-5.

[126] GEHRINGER, J. AND LI, X.-M. Functional limit theorems for the fractional Ornstein-Uhlenbeck process. *Journal of Theoretical Probability*, (2020). doi:10.1007/s10959-020-01044-7.

[127] GEHRINGER, J., LI, X.-M., AND SIEBER, J. Functional limit theorems for Volterra processes and applications to homogenization (2021). Preprint, arXiv:2104.06364.

[128] GENIN, A. AND TANKOV, P. Optimal importance sampling for Lévy processes. *Stochastic Processes and their Applications*, **130** (2020). doi:10.1016/j.spa.2018.12.019.

[129] GERENCSÉR, B. AND RÁSONYI, M. Invariant measures for fractional stochastic volatility models (2020). Preprint, arXiv:2002.04832.

[130] GERHOLD, S., GERSTENECKER, C., AND GULISASHVILI, A. Large deviations for fractional volatility models with non-Gaussian volatility driver (2020). Preprint, arXiv:2003.12825.

[131] GERHOLD, S., GERSTENECKER, C., AND PINTER, A. Moment explosions in the rough Heston model. *Decisions in Economics and Finance*, **42** (2019). doi:10.1007/s10203-019-00267-6.

[132] GOUTTE, S., ISMAIL, A., AND PHAM, G. Regime-switching stochastic volatility model: estimation and calibration to VIX options. *Applied Mathematical Finance*, **24** (2017). doi:10.1080/1350486X.2017.1333015.

[133] GRBAC, Z., KRIEF, D., AND TANKOV, P. Long-time trajectorial large deviations and importance sampling for affine stochastic volatility models. *Advances in Applied Probability*, **53** (2021). doi:10.1017/apr.2020.58.

[134] GRIPENBERG, G., LONDEN, S.-O., AND STAFFANS, O. *Volterra integral and functional equations*. Encyclopedia of Mathematics and its Applications. Cambridge University Press (1990). doi:10.1017/CBO9780511662805.

[135] GUASONI, P. AND ROBERTSON, S. Optimal importance sampling with explicit formulas in continuous time. *Finance and Stochastics*, **12** (2008). doi:10.1007/s00780-007-0053-5.

[136] Guennoun, H., Jacquier, A., Roome, P., and Shi, F. Asymptotic behavior of the fractional Heston model. *SIAM Journal on Financial Mathematics*, **9** (2018). doi:10.1137/17M1142892.

[137] Gulisashvili, A. Large deviation principle for Volterra type fractional stochastic volatility models. *SIAM Journal on Financial Mathematics*, **9** (2018). doi:10.1137/17M116344X.

[138] Gulisashvili, A., Viens, F., and Zhang, X. Small-time asymptotics for Gaussian self-similar stochastic volatility models. *Applied Mathematics and Optimization*, (2018). doi:10.1007/s00245-018-9497-6.

[139] Gulisashvili, A., Viens, F., and Zhang, X. Extreme-strike asymptotics for general Gaussian stochastic volatility models. *Annals of finance*, **15** (2019). doi:10.1007/s10436-018-0338-z.

[140] Guyon, J. The joint S&P 500/VIX smile calibration puzzle solved. *Risk*, (2020).

[141] Guyon, J., Menegaux, R., and Nutz, M. Bounds for VIX futures given S&P 500 smiles. *Finance and Stochastics*, **21** (2017). doi:10.1007/s00780-017-0334-6.

[142] Hairer, M. Exponential mixing properties of stochastic PDEs through asymptotic coupling. *Probability Theory and Related Fields*, **124** (2002). doi:10.1007/s004400200216.

[143] Hairer, M. Ergodicity of stochastic differential equations driven by fractional Brownian motion. *The Annals ofProbability*, **33** (2005). doi:10.1214/009117904000000892.

[144] Hairer, M. Ergodicity of stochastic differential equations driven by fractional Brownian motion. *The Annals of Probability*, **33** (2005). doi:10.1214/009117904000000892.

[145] Hairer, M. and Li, X.-M. Averaging dynamics driven by fractional Brownian motion. *The Annals of Probability*, **48** (2020). doi:10.1214/19-AOP1408.

[146] Hairer, M. and Mattingly, J.-C. Ergodicity of the 2D Navier-Stokes equations with degenerate stochastic forcing. *Annals of Mathematics*, **164** (2006). doi:10.4007/annals.2006.164.993.

[147] HAIRER, M. AND OHASHI, A. Ergodic theory for SDEs with extrinsic memory. *The Annals of Probability*, **35** (2007). doi:10.1214/009117906000001141.

[148] HAIRER, M. AND PILLAI, N. S. Ergodicity of hypoelliptic SDEs driven by fractional Brownian motion. *Annales de l'Institut Henri Poincaré - Probabilités et Statistiques*, **47** (2011). doi:10.1214/10-AIHP377.

[149] HARNETT, D. AND NUALART, D. Central limit theorem for an iterated integral with respect to fBm with $H > 1/2$. *Stochastics*, **86** (2014). doi:10.1080/17442508.2013.774403.

[150] HORVATH, B. AND JACQUIER, A. Rough volatility network – rough volatility literature. https://sites.google.com/site/roughvol/home/risks-1 (2021). Online; accessed 20 July 2021.

[151] HORVATH, B., JACQUIER, A., AND LACOMBE, C. Asymptotic behaviour of randomised fractional volatility models. *Journal of Applied Probability*, **56** (2019). doi:10.1017/jpr.2019.27.

[152] HORVATH, B., JACQUIER, A., AND MUGURUZA, A. Functional central limit theorems for rough volatility (2017). Preprint, arXiv:1711.03078.

[153] HORVATH, B., JACQUIER, A., AND TANKOV, P. Volatility options in rough volatility models. *SIAM Journal on Financial Mathematics*, **11** (2020). doi:10.1137/18M1169242.

[154] HORVATH, B., TEICHMANN, J., AND ZURIC, Z. Deep hedging under rough volatility (2021). Preprint, arXiv:2102.01962.

[155] HUANG, J., NUALART, D., AND VIITASAARI, L. A central limit theorem for the stochastic heat equation. *Stochastic Processes and their Applications*, **130** (2020). doi:10.1016/j.spa.2020.07.010.

[156] IBRAGIMOV, I. A. On smoothness conditions for trajectories of random functions. *Theory of Probability and its Applications*, **28** (1983). doi:10.1137/1128023.

[157] ISCOE, I. Ergodic theory and a local occupation time for measure-valued critical branching Brownian motion. *Stochastics: An International Journal of Probability and Stochastic Processes*, **18** (1986). doi:10.1080/17442508608833409.

[158] JACQUIER, A., MARTINI, C., AND MUGURUZA, A. On VIX futures in the rough Bergomi model. *Quantitative Finance*, **18** (2018). doi:10.1080/14697688.2017.1353127.

[159] JACQUIER, A., MUGURUZA, A., AND PANNIER, A. Smile asymptotics in multi-factor rough volatility models (2021). In preparation.

[160] JACQUIER, A. AND OUMGARI, M. Deep curve-dependent PDEs for affine rough volatility (2019). Available at: arXiv:1906.02551.

[161] JACQUIER, A., PAKKANEN, M., AND STONE, H. Pathwise large deviations for the rough Bergomi model. *Journal of Applied Probability*, **55** (2018). doi:10.1017/jpr.2018.72.

[162] JACQUIER, A. AND PANNIER, A. Large and moderate deviations for stochastic Volterra systems (2020). Preprint, arXiv: 2004.10571.

[163] JACQUIER, A., PANNIER, A., AND SPILIOPOULOS, K. On the large-time behaviour of affine Volterra processes (2021). In preparation.

[164] JACQUIER, A. AND SPILIOPOULOS, K. Pathwise moderate deviations for option pricing. *Mathematical Finance*, **30** (2020). doi:10.1111/mafi.12228.

[165] JAISSON, T. AND ROSENBAUM, M. Rough fractional diffusions as scaling limits of nearly unstable heavy tailed Hawkes processes. *The Annals of Applied Probability*, **26** (2016). doi:10.1214/15-AAP1164.

[166] KARATZAS, I. AND SHREVE, S. E. *Brownian motion and stochastic calculus*. Springer-Verlag New York (1998). doi:10.1007\%2F978-1-4612-0949-2.

[167] KELLER-RESSEL, M., LARSSON, M., AND PULIDO, S. Affine rough models (2019). Preprint, arXiv:1812.08486.

[168] KELLER-RESSEL, M. AND MIJATOVIC, A. On the limit distributions of continuous-state branching processes with immigration. *Stochastic Processes and their Applications*, **122** (2012). doi:10.1016/j.spa.2012.03.012.

[169] KHOSHNEVISAN, D., KIM, K., MUELLER, C., AND SHIU, S.-Y. Phase analysis for a family of stochastic reaction-diffusion equations (2020). Preprint, arXiv:2012.12512.

[170] KOKHOLM, T. AND STISEN, M. Joint pricing of VIX and SPX options with stochastic volatility and jump models. *Journal of Risk Finance*, **16** (2015). doi:10.1108/JRF-06-2014-0090.

[171] KOLKOVSKA, E. T. AND LÓPEZ-MIMBELA, J. A. Survival of some measure-valued Markov branching processes. *Stochastic Models*, **35** (2019). doi:10.1080/15326349.2019.1578238.

[172] KONNO, N. AND SHIGA, T. Stochastic partial differential equations for some measure-valued diffusions. *Probability Theory and Related Fields*, **79** (1988). doi:10.1007/BF00320919.

[173] LACOMBE, C., MUGURUZA, A., AND STONE, H. Asymptotics for volatility derivatives in multi-factor rough volatility models. *Mathematics and Financial Economics*, **15** (2021). doi:10.1007/s11579-020-00288-5.

[174] LAPLACE, P.-S. *Essai philosophiques sur les probabilités*. Courcier (1814).

[175] LEDOUX, M. A note on large deviations for Wiener chaos. *Séminaire de probabilités (Strasbourg)*, **24** (1990). doi:10.1007/BFb0083753.

[176] LI, X.-M. AND SIEBER, J. Slow-fast systems with fractional environment and dynamics (2020). Preprint, arXiv:2012.01910.

[177] LI, Y., WANG, R., YAO, N., AND ZHANG, S. A moderate deviation principle for stochastic Volterra equation. *Statistics and Probability Letters*, **122** (2017). doi:10.1016/j.spl.2016.10.033.

[178] LI, Z. *Measure-Valued Branching Markov Processes*. Springer, Probability and Its Applications (2011). doi:10.1007/978-3-642-15004-3.

[179] LI, Z., WANG, H., AND XIONG, J. A degenerate stochastic partial differential equation for superprocesses with singular interaction. *Probability Theory and Related Fields*, **130** (2004). doi:10.1007/s00440-003-0313-z.

[180] LINZ, P. *Analytical and Numerical Methods for Volterra Equations*. SIAM Studies in Applied and Numerical Mathematics (1985). doi:10.1137/1.9781611970852.

[181] MALLIAVIN, P. Stochastic calculus of variation and hypoelliptic operators. In *Proceedings of the International Symposium on Stochastic Differential Equations, Res. Inst. Math. Sci., Kyoto Univ., Kyoto, 1976*. Wiley, New York-Chichester-Brisbane (1978). doi:10.1007/978-1-4612-2054-1_2.

[182] MANDELBROT, B. B. The variation of some other speculative prices. *The Journal of Business*, **40** (1967).

[183] MANDELBROT, B. B. *Fractals and Scaling in Finance*. Springer New York (1997). ISBN 9781475727630. doi:10.1007/978-1-4757-2763-0.

[184] MANDELBROT, B. B. AND HUDSON, R. L. *The (mis) Behaviour of Markets: A Fractal View of Risk, Ruin and Reward*. Profile (2004).

[185] MANDELBROT, B. B. AND VAN NESS, J. W. Fractional Brownian motions, fractional noises and applications. *SIAM Review*, **10** (1968). doi:10.1137/1010093.

[186] MARIE, N. A pathwise fractional one compartment intra-veinous bolus model. *International Journal of Statistics and Probability*, **3** (2014). doi:ff10.5539/ijsp.v3n3p65.

[187] MAYER-WOLF, E., NUALART, D., AND PÉREZ-ABREU, V. Large deviations for multiple Wiener-Itô integral processes. *Séminaire de probabilités (Strasbourg)*, **26** (1992). doi:10.1007/BFb0084307.

[188] MCCRICKERD, R. AND PAKKANEN, M. Turbocharging Monte–Carlo pricing for the rough Bergomi model. *Quantitative Finance*, **18** (2018). doi:10.1080/14697688.2018.1459812.

[189] MELLOUK, M. AND MILLET, A. Large deviations for stochastic flows and anticipating SDEs in Besov-Orlicz spaces. *Stochastics*, **63** (1998). doi:10.1080/17442509808834151.

[190] MERTON, R. C. Theory of rational option pricing. *The Bell Journal of Economics and Management Science*, **4** (1973). doi:10.2307/3003143.

[191] MILLET, A., NUALART, D., AND SANZ, M. Composition of large deviation principles and applications. *Stochastic Analysis*, (1991). doi:10.1016/B978-0-12-481005-1.50026-0.

[192] MILLET, A., NUALART, D., AND SANZ, M. Large deviations for a class of anticipating stochastic differential equations. *The Annals of Probability*, **20** (1992). doi:10.1214/aop/1176989535.

[193] MORSE, M. R. AND SPILIOPOULOS, K. Moderate deviations principle for systems of slow-fast diffusions. *Asymptotic Analysis*, **105** (2017). doi:10.3233/ASY-171434.

[194] Morse, M. R. and Spiliopoulos, K. Importance sampling for slow-fast diffusions based on moderate deviations. *SIAM Journal on Multiscale Modeling and Simulation*, **18** (2018). doi:10.1137/18M1192962.

[195] Mytnik, L. and Salisbury, T. S. Uniqueness for Volterra-type stochastic integral equations (2015). Preprint, arXiv:1502.05513.

[196] Neuman, E. Pathwise uniqueness of the stochastic heat equations with spatially inhomogeneous white noise. *The Annals of Probability*, **46** (2018). doi:10.1214/17-AOP1239.

[197] Nourdin, I. and Nualart, D. Central limit theorems for multiple Skorohod integrals. *Journal of Theoretical Probability*, **23** (2010). doi:10.1007/s10959-009-0258-y.

[198] Nourdin, I., Nualart, D., and Poly, G. Absolute continuity and convergence of densities for random vectors on Wiener chaos. *Electronic Journal of Probability*, **18** (2013). doi:10.1214/EJP.v18-2181.

[199] Nourdin, I. and Peccati, G. Stein's method on Wiener chaos. *Probability Theory and Related Fields*, **145** (2009). doi:10.1007/s00440-008-0162-x.

[200] Nourdin, I. and Poly, G. Convergence in law in the second Wiener/Wigner chaos. *Electronic Communications in Probability*, **17** (2012). doi:10.1214/ECP.v17-2023.

[201] Nualart, D. *The Malliavin calculus and related topics*. Springer (2006). doi:10.1007/3-540-28329-3.

[202] Nualart, D. and Peccati, G. Central limit theorems for sequences of multiple stochastic integrals. *The Annals of Probability*, **33** (2005). doi:10.1214/009117904000000621.

[203] Nualart, D. and Rovira, C. Large deviations for stochastic Volterra equations. *Bernoulli*, **6** (2000). doi:10.2307/3318580.

[204] Nualart, D. and Zheng, G. Averaging Gaussian functionals. *Electronic Journal of Probability*, **25** (2020). doi:10.1214/20-EJP453.

[205] Pacati, C., Pompa, G., and Reno, R. Smiling twice: The Heston ++ model. *Journal of Banking and Finance*, **96** (2018). doi:j.jbankfin.2018.08.010.

[206] PANNIER, A. Pathwise large deviations for white noise chaos expansions (2021). Preprint, arXiv:2011.14851. Forthcoming in *Bernoulli*.

[207] PAPANICOLAOU, A. Extreme-strike comparisons and structural bounds for SPX and VIX options. *SIAM Journal on Financial Mathemtics*, **9** (2018). doi:10.1137/141001615.

[208] PAPANICOLAOU, A. AND SIRCAR, R. A regime-switching Heston model for VIX and S&P 500 implied volatilities. *Quantitative Finance*, **14** (2014). doi:10.1080/14697688.2013.814923.

[209] PARDOUX, E. *Markov Processes and Applications: Algorithms, Networks, Genome and Finance*. Wiley Series in Probability and Statistics (2008). doi:10.1002/9780470721872.

[210] PARDOUX, E. AND PROTTER, P. Stochastic Volterra equations with anticipating coefficients. *The Annals of Probability*, **18** (1990). doi:10.1214/aop/1176990638.

[211] PEI, B., INAHAMA, Y., AND XU, Y. Averaging principle for fast-slow system driven by mixed fractional Brownian rough path (2020). Preprint, arXiv:2010.06788.

[212] PEI, B., INAHAMA, Y., AND XU, Y. Averaging principles for mixed fast-slow systems driven by fractional Brownian motion (2020). Preprint, arXiv:2001.06945.

[213] PÉREZ-ABREU, V. Chaos expansions: A review. *Resehans, Journal of IME-USP*, **1** (1993).

[214] PÉREZ-ABREU, V. AND TUDOR, C. Large deviations for a class of chaos expansions. *Journal of Theoretical Probability*, **7** (1994). doi:10.1007/BF02214370.

[215] PHAM, H. Large deviations in finance. *Third SMAI European Summer School in Financial Mathematics*, (2010). Lecture notes, lpsm.paris/pageperso/pham/GD-finance.

[216] POINCARÉ, H. Sur le problème des trois corps et les équations de la dynamique. *Acta Mathematica*, **13** (1890).

[217] POINCARÉ, H. *Science et Méthode*. Flammarion (1908).

[218] PROTTER, P. Volterra equations driven by semimartingales. *The Annals of Probability*, **13** (1985). doi:10.1214/aop/1176993006.

[219] PROTTER, P. *Stochastic Integration and Differential Equations*. Springer-Verlag (2005). doi:10.1007/978-3-662-10061-5.

[220] ROBERTSON, S. Sample path large deviations and optimal importance sampling for stochastic volatility models. *Stochastic Processes and their Applications*, **120** (2010). doi:10.1016/j.spa.2009.10.010.

[221] ROGERS, L. Arbitrage with fractional Brownian motion. *Mathematical Finance*, **7(1)** (1997). doi:10.1111/1467-9965.00025.

[222] ROGERS, L. AND WILLIAMS, D. *Diffusions, Markov Processes and Martingales, Volume 2*. Cambridge University Press (2000).

[223] ROSENBAUM, M. AND ZHANG, J. Deep calibration of the quadratic rough Heston model (2021). Preprint, arXiv:2107.01611.

[224] ROVIRA, C. AND SANZ-SOLÉ, M. Large deviations for stochastic Volterra equations in the plane. *Potential Analysis*, **12** (2000). doi:10.1023/A:1008662409325.

[225] SALINS, M., BUDHIRAJA, A., AND DUPUIS, P. Uniform large deviation principles for Banach space valued stochastic evolution equations. *Transactions of the American Mathematical Society*, **372** (2019). doi:10.1090/tran/7872.

[226] SHREVE, S. E. *Stochastic Calculus for Finance II: Continuous-Time Models*. Springer Finance (2004).

[227] SPILIOPOULOS, K. Large deviations and importance sampling for systems of slow-fast motion. *Applied Mathematics and Optimization*, **67** (2013). doi:10.1007/s00245-012-9183-z.

[228] STEIN, E. AND STEIN, J. Stock price distributions with stochastic volatility - an analytic approach. *The Review of Financial studies*, **4** (1991). doi:10.1093/rfs/4.4.727.

[229] VARADHAN, S. R. S. Asymptotic probabilities and differential equations. *Communications on Pure and Applied Mathematics*, **19** (1966). doi:10.1002/cpa.3160190303.

[230] Varadhan, S. R. S. Large deviations. *The Annals of Probability*, **36** (2008). doi:10.1214/07-AOP348.

[231] Varvenne, M. *Ergodicity of fractional stochastic differential equations and related problems*. Ph.D. thesis, Université Paul Sabatier - Toulouse III (2019). tel.archives-ouvertes.fr/tel-02737000/.

[232] Viens, F. and Zhang, J. A martingale approach for fractional Brownian motions and related path dependent PDEs. *The Annals of Applied Probability*, **29** (2019). doi:10.1214/19-AAP1486.

[233] Walsh, J. An introduction to stochastic partial differential equations. *Éc. Été Probab. St.-Flour, XIV—1984, in: Lecture Notes in Math. vol. 1180, Springer, Berlin*, (1986). doi:10.1007/BFb0074920.

[234] Wang, J., Zhou, Y., and O'Regan, D. A note on asymptotic behaviour of Mittag–Leffler functions. *Integral Transforms and Special Functions*, **29** (2018). doi:10.1080/10652469.2017.1399373.

[235] Weinan, E. and Mattingly, J.-C. Ergodicity for the Navier-Stokes equation with degenerate random forcing: Finite-dimensional approximation. *Communications on Pure and Applied Mathematics*, **54** (2001). doi:10.1002/cpa.10007.

[236] Weinan, E., Mattingly, J.-C., and Sinai, Y. Gibbsian dynamics and ergodicity for the stochastically forced Navier-Stokes equation. *Communications in Mathematical Physics*, **224** (2001). doi:10.1007/s002201224083.

[237] Wikipedia contributors. Markov chain – Wikipedia, the free encyclopedia. https://en.wikipedia.org/wiki/Markov_chain (2021). Online; accessed 20 July 2021.

[238] Wilmott, P. *Wilmott introduces Quantitative Finance*. Wiley (2007). Second edition.

[239] Yamada, T. and Watanabe, S. On the uniqueness of solutions of stochastic differential equations. *Journal of Mathematics of Kyoto University*, **11** (1971).

[240] Zhang, X. Euler schemes and large deviations for stochastic Volterra equations with singular kernels. *Journal of Differential Equations*, **244** (2008). doi:10.1016/j.jde.2008.02.019.

[241] Zhu, Q., Loeper, G., Chen, W., and Langrené, N. Markovian approximation of the rough Bergomi model for Monte Carlo option pricing (2020). Preprint, arXiv:2007.02113.