Stochastic Invariance and Stochastic Volterra Equations

Eduardo Abi Jaber

THÈSE DE DOCTORAT

de l'Université de recherche Paris Sciences et Lettres
PSL Research University

Préparée à l'Université Paris-Dauphine

Stochastic Invariance and Stochastic Volterra Equations

École Doctorale de Dauphine – ED 543

Spécialité MATHÉMATIQUES APPLIQUÉES

COMPOSITION DU JURY :

Nizar TOUZI
Ecole Polytechnique, Président

Antoine JACQUIER
Imperial College, Rapporteur

Peter TANKOV
ENSAE, Rapporteur

Josef TEICHMANN
ETH Zürich, Rapporteur

Aurélien ALFONSI
Ecole des Ponts, Examinateur

Martin LARSSON
ETH Zürich, Examinateur

Sergio PULIDO
ENSIIE, Examinateur

Mathieu ROSENBAUM
Ecole Polytechnique, Examinateur

Soutenue par **Eduardo ABI JABER**
le 18 Octobre 2018

Dirigée par **Bruno BOUCHARD**
et par **Jean-David FERMANIAN**
Co-encadrée par **Camille ILLAND**

DAUPHINE UNIVERSITÉ PARIS | PSL

"L'univers est rempli de magie et il attend patiemment que notre intelligence s'affine."

Eden Phillpotts

Abstract

Stochastic Invariance and Stochastic Volterra Equations

The present thesis deals with the theory of finite dimensional stochastic equations.

In the first part, we derive necessary and sufficient geometric conditions on the coefficients of a stochastic differential equation for the existence of a constrained solution, under weak regularity on the coefficients.

In the second part, we tackle existence and uniqueness problems of stochastic Volterra equations of convolution type. These equations are in general non-Markovian. We establish their correspondence with infinite dimensional equations which allows us to approximate them by finite dimensional stochastic differential equations of Markovian type.

Finally, we illustrate our findings with an application to mathematical finance, namely rough volatility modeling. We design a stochastic volatility model with an appealing trade-off between flexibility and tractability.

Keywords: Stochastic invariance, stochastic convolution equations, affine processes, rough volatility.

Acknowledgements

Picture an adventure with an extremely rough path to start with. As time flows on, unexpected encounters combined with luck will tend to make the path as smooth as possible. In academic terms, such an adventure is called a PhD. I shall first acknowledge these key encounters.

First and foremost, I would like to express my deepest gratitude to everyone who supported me during my PhD and without whom this manuscript would not have seen the light of day. Indeed, the thesis at hand stems from multiple collaborations, several fruitful discussions around the globe, numerous advices at all levels, unconditional support from relatives, and, occasionally, some sleepless nights. I take this opportunity to thank everyone who made this journey a highly enjoyable life-changing experience for me.[1]

Above all, no words can describe my indebtedness to Bruno Bouchard who has been a tremendous advisor for me. Not only did he graciously offer me guidance to grow as a researcher, but also gave me precious advices for my career. He continuously encouraged me in my research and contributed immensely to it whenever I found myself desperately stuck. I admire his enthusiasm. Likewise, my profound gratitude goes to Jean-David Fermanian without whom this opportunity would not have been possible. He guided me in the choice of my career, and provided me with extensive help and valuable support during rough times. I would also like to express my special appreciation to Illand, my supervisor at AXA Investment Managers, for steering the project all the way from my internship to my PhD with patience and endurance. His invaluable creativity, intuition and expertise were key in finding new ways to explore when things fell apart. I could not be luckier to have them all as my mentors.

Secondly, I would like to thank the members of my defense committee. My deep appreciation goes first to the *rapporteurs* Antoine Jacquier, Peter Tankov, and Josef Teichmann for accepting to examine my manuscript, for their careful reading and for their interest in my work. All three of them occupy an influential place within my research. Indeed, my first contact with affine processes was through one of Antoine Jacquier's papers, which I had to present during my Master's degree at Université Paris Diderot. This is where, I first met Peter Tankov who immediately aroused my interest for research and encouraged me on embarking on such a journey, and so did I... It did not take me too long to realize that Josef Teichmann was a major key name in my line of research. I feel privileged to have all three of them act as *rapporteurs*. I am also very delighted that Aurélien Alfonsi is a member of the jury. Throughout my PhD I had the chance to have several illuminating discussions with him, at the Cermics seminar and the Summer School, regarding matrix-valued processes, affine processes and discretization schemes. I am very honored by the presence of Mathieu Rosenbaum whose recent findings stimulated and inspired the second part of this thesis. In this regard, I would like to thank him for supersiving my work with Omar El Euch and for his constant support and encouragement. I am also deeply touched that Nizar Touzi has agreed to be part of the committee. Last, but by no means least, I am very happy that Martin Larsson and Sergio Pulido are taking part in the committee. I was fortunate enough to have them as my first collaborators[2]. They taught me that math and humor can go along. I thank them for all the time we spent working together and for their precious advices.

[1] I apologize in advance to those I may have unintentionally omitted.

[2] besides my advisors.

I take this opportunity to extend my gratitude to my other collaborators Christa Cuhiero, Omar El Euch and Stefan Tappe. Working with them was a delightful and enriching experience, it helped me broaden and strengthen my knowledge. Needless to say that this would not have been possible without the solid background I acquired during my studies. In particular, I would like to acknowledge the excellent teaching staff of the Master Modélisation Aléatoire at Université Paris Diderot lead, at that time, by Laure Elie, Huyên Pham and Peter Tankov. I thank them for providing high quality courses and, most importantly, for conveying to me their passion for research.

Besides, I gratefully acknowledge the CIFRE funding received by AXA Investment Managers towards my PhD. I am very honored to have been part of their quantitative research team. My time at AXA-IM was made enjoyable in large part due to my remarkable colleagues Ichem Boudfar, Martin Deux, Maxence Jeunesse, Ethan Reiner and Jean-Benoit Woiselle. Working on such challenging subjects could not be more satisfying without them, they have been a great source of collaboration, making the environment a fun place to work in. I am also deeply thankful to Marie-Pierre Ravoteur for her unwavering support and trust. At last, I cannot forget the fun, yet competitive, Basketball matches with Chadi Naufal and Ramzi Djellas.

On another note, I am very proud of belonging to the mathematical finance community. I was extremely fortunate to meet genuinely passionate people with whom I shared joyful moments that will leave marks beyond this thesis. Complementing the names cited above, I would like to thank: Ahmed, Alexander, Alexandre, Alexandre, Antonis, Blanka, Benjamin, Carlo, Caroline, Charles-Albert, Claudio, Cyril, Daniel, Dirk, Dylan, Eyal, Fransesco, Hadrien, Heythem, Idris, Jean-François, Jean-Pierre, Jim, Johaness, Johaness, Marcel, Mark, Martin, Martino, Marvin, Matteo, Nicolas, Othmane, Pamela, Paolo, Paul, Philipp, Pierre, Roxana, Samuel, Sara, Stefano, Stéphane, Thibault, Thorsten, Todor, Wahid, Xialou, Zorana.

I would also like to profoundly thank Dave and Hassan for encouraging me to embark on such a stimulating adventure. I cannot forget to mention the administrative staff at Dauphine for their hard efforts and their reactivity. In particular, I am deeply thankful to César, Igor, Marie-Belle and Sylvie.

Finally, I could not have made it without the unconditional moral support and the love of my friends, my family and my parents. My final sweet thought goes to my lifetime companion Maya for always believing in me.

Paris, October 18, 2018
Eduardo Abi Jaber

Contents

Notations

$\mathbb{R}^d_+$	The non-negative orthant $[0, \infty[^d$
$\mathbb{R}^{n \times d}$	The space of $n \times d$ matrices
$\mathbb{S}^m$	The space of symmetric $d \times d$ matrices
$\mathbb{S}^d_+$	The cone of symmetric positive semi-definite matrices
I_d	The $d \times d$ identity matrix
$\operatorname{diag}[x]$	The diagonal matrix whose i-th diagonal component is x^i
$\operatorname{Tr}(A)$	The trace of matrix A
$A^\top$	The transpose of matrix A
A^+	The Moore-Penrose inverse of matrix A
$A^{1/2}$	The square-root of a symmetrix positive semi-definite matrix A
Df	The Jacobian of the function f
$D^2 f$	The Hessian of the function f
vec	The vectorization operator
$\otimes$	The Kronecker product
$\mathcal{D}$	A closed subset of $\mathbb{R}^d$
$\partial\mathcal{D}$	The boundary of the domain $\mathcal{D}$

Chapter 1

Introduction

The modeling puzzle: consistency and tractability

Models are everywhere. A model is an umbrella term that refers to a simplification of a complex reality embodying different concepts such as fashion models, scale models, scientific models... For scientists, a model is more precisely a mathematical representation of the world used to understand, describe and predict. Yet, it still has to verify certain properties to be recognized as a successful model before being adopted in practice. Two such key properties are consistency and tractability.

A model is said to be consistent if it captures, to some extent, the stylized facts of the observations. The more complex a model is, e.g. with a large number of parameters, the more detailed the description of the reality, the more simplistic the more it drifts away from real life. A model is said to be tractable if it produces its outputs in reasonable time[1]. Complex models will require more computational resources. It is worth noting that the concept of tractability evolves with time: what was considered un-tractable fifty years ago is now considered tractable due to technological progress and the increase in computational power. One can therefore distinguish between analytical tractability (closed-form solutions) and computational tractability (numerical approximations of solutions).

In a utopian world, ideal models would provide a perfect balance between consistency and tractability, that is, be able to replicate the stylized facts[2] of the empirical observations while predicting relevant outputs in a time-efficient fashion. Finding such models is without any doubt a myth. Yet, this has never discouraged scientists from going in search of models approximating this ideal concept:

How much consistency is one willing to sacrifice for tractability?

To answer this question, a picture is worth a thousand words.

[1] relatively to the problem at hand.

[2] at least some of them that are judged important for the problem considered.

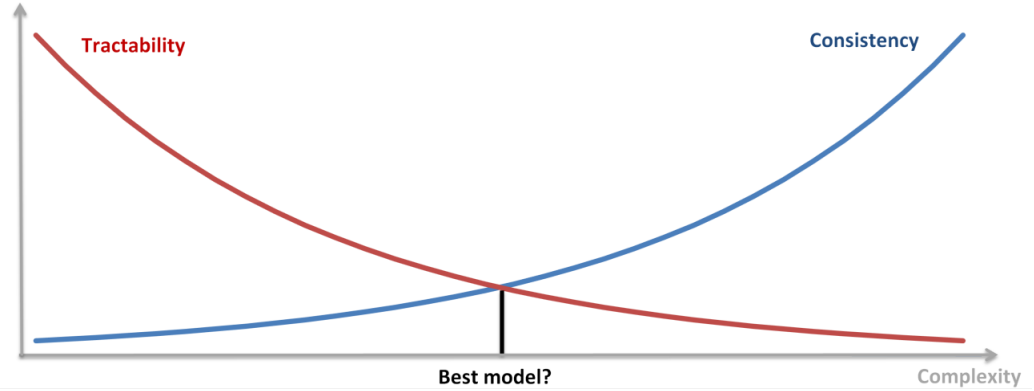

FIGURE 1.1: *"Everything should be made as simple as possible, but not simpler"* - Albert Einstein

The present thesis conceptually deals with consistency and tractability of probabilistic models, more precisely stochastic models in continuous time. These models are used to account for randomness phenomena with a wide range of applications in biology, economics, energy markets, finance, insurance, physics and much more. The motivation stems here mainly from mathematical finance. However, most of the chapters are concerned with the general theory of stochastic analysis and are not restricted to the exciting field of finance. The concepts and results developed in the sequel are in their most general form, perfectly suited for probabilistic modeling.

The first part of the thesis is devoted to one of the most basic qualitative concepts of consistency: picture a biologist wishing to model the evolution (in time) of the frequency of a specific gene in a certain population. The most minimal requirement that the biologist could ask for is a model with values in $[0, 1]$ in order to stay consistent with the physical constraint: a frequency is by definition a non-negative number less or equal than one. Similarly, before putting millions of euros at stake, a trader would, hopefully, require from his model to produce positive values for the asset price. In mathematical terms, starting from an initial set of physical constraints denoted by $\mathcal{D}$, we are asked to determine all stochastic models, more precisely all *stochastic differential equations*, that remain in $\mathcal{D}$ at all time. This is known as an *invariance* or *viability* problem and is the subject of Chapters 3 and 4. Once there, the biologist and the trader would often favor those consistent models having a high degree of tractability for practical use. Technically, this restriction is made by usually imposing an additional structure on the coefficients of the *stochastic differential equation*, for instance *affine* or *polynomial* structures. This often yields tractability by completely characterizing the underlying law of the stochastic process.

This leads to the second part of the thesis where the quantitative concept of tractability is tackled. We introduce a new class of stochastic models that are richer and more flexible than those studied in the first part, namely *stochastic Volterra equations*. The price to pay is that a priori these equations are highly un-tractable. Our main results show that looking at these equations through the *affine* lens allows one to recover tractability, making *stochastic Volterra equations* perfect candidates for successful probabilistic models with an appealing trade-off between consistency and tractability. This is highlighted in Chapters 5 and 6. Finally, we exemplify these results with an application to finance in Chapters 7 and 8.

From a mathematical viewpoint, the two parts of this thesis can be read independently.

1.1 Stochastic invariance

1.1.1 The invariance/viability paradigm

What is an invariance/viability problem? On an abstract level, we fix a set $\mathcal{D}$ of some topological space and we drop a particle inside $\mathcal{D}$. The set $\mathcal{D}$ is said to be invariant/viable if the particle stays in $\mathcal{D}$ at all time. The problem is then to characterize the invariance of $\mathcal{D}$ through conditions on the driving forces that govern the movement of the particle. Invariance and viability problems have been intensively studied in the literature, first in a deterministic setup and later in a random environment.

1.1.2 A physicist's insight in a deterministic setup

In order to gain insights, we start by making two restrictions on the set $\mathcal{D}$ and the movement of the particle. We consider a closed subset $\mathcal{D}$ of $\mathbb{R}^d$ equipped with its canonical inner product $\langle \cdot, \cdot \rangle$ and assume that the movement of the particle is governed by the following autonomous ordinary differential equation

$$dX_t = b(X_t)dt, \quad X_0 = x, \tag{1.1.1}$$

where $X_t \in \mathbb{R}^d$ denotes the position of the particle at time $t \geq 0$ and $b : \mathbb{R}^d \to \mathbb{R}^d$ is the driving force. Further, we assume that b is Lipschitz continuous so that (1.1.1) admits a unique solution for every initial point $x \in \mathbb{R}^d$. Starting the solution from $X_0 \in \mathcal{D}$, the invariance problem can now be formulated as follows.

On what conditions on b does the solution X remain in $\mathcal{D}$ at all time?

Assume that X stays in $\mathcal{D}$ up to the first time it hits the boundary at time $t \geq 0$. From a physicist's perspective, $b(X_t) = \frac{dX_t}{dt}$ denotes the instantaneous velocity of the particle at time t. In this context, it is useful to recall that the instantaneous velocity is always tangential to the path of the particle. Whence, intuitively, at a boundary point $X_t \in \mathcal{D}$ the trajectory will stay in the domain as long as

the velocity $b(X_t)$ is either tangential to the boundary or inward pointing,

or else the velocity would point outwards and force the particle outside the domain. When the particle is strictly inside the domain, the instantaneous velocity can point in any direction. This can be given a precise mathematical meaning by specifying a dual object to tangency: the first order normal cone defined as

$$\mathcal{N}^1_{\mathcal{D}}(x) = \Big\{ u \in \mathbb{R}^d : \langle u, y - x \rangle \leq o(\|y - x\|), \forall\, y \in \mathcal{D} \Big\} \tag{1.1.2}$$

consisting of all outward pointing normal vectors to the set $\mathcal{D}$ at a point x. The physicist's intuition translates into

$$\langle u, b(x) \rangle \leq 0, \quad x \in \mathcal{D},\ u \in \mathcal{N}^1_{\mathcal{D}}(x). \tag{1.1.3}$$

If the point x strictly lies in the interior of $\mathcal{D}$, $\mathcal{N}^1_{\mathcal{D}}(x) = \{0\}$ so that (1.1.3) is trivially satisfied for interior points independently of the behavior of b.

To confirm this assertion mathematically, assume that $\mathcal{D}$ is invariant under (1.1.1) started at $x \in \mathcal{D}$, i.e. $X_t \in \mathcal{D}$, for all $t \geq 0$. Let $\phi : \mathbb{R}^d \to \mathbb{R}$ be a smooth function such that

$\max_{\mathcal{D}} \phi = \phi(x)$. Since $\phi(X_t) \leq \phi(x)$, for all $t \geq 0$, the chain rule yields

$$\int_0^t D\phi(X_s) b(X_s) ds \leq 0, \quad t \geq 0,$$

where $D\phi(x)$ is the transpose of the gradient of ϕ at the point x. Dividing the above by t and sending $t \to 0$ yields by continuity

$$D\phi(x) b(x) \leq 0$$

which after a small additional effort can be shown to be equivalent to (1.1.3).[3] The condition is also sufficient for invariance. This result is known as Nagumo's tangency condition named after the Japanese mathematician Mitio Nagumo who was the first one to derive this characterization in a seminal paper in 1942 written in German [97]. The theorem was seemingly forgotten and rediscovered independently in different contexts during the next twenty years. For a detailed treatment of deterministic invariance and viability problems we refer to [9, 24].

1.1.3 A geometric intuition in a random environment

In a random environment, the autonomous ordinary differential equation (1.1.1) is extended to a time-homogeneous Markovian stochastic differential equation

$$dX_t = b(X_t)dt + \sigma(X_t) \underbrace{dW_t}_{\text{noise}}, \quad X_0 = x, \tag{1.1.4}$$

where $\sigma : \mathbb{R}^d \mapsto \mathbb{R}^{d \times d}$ is continuous with growth conditions and W is a d-dimensional Brownian motion W, i.e. the analogue of a random walk in continuous-time.

The rules of the game remain unchanged: find necessary and sufficient conditions on the drift b and the volatility matrix σ under which there exists a $\mathcal{D}$-valued weak solution of (1.1.4) given that $x \in \mathcal{D}$.

Extending the physicist's intuition, as long as the particle remains strictly inside the domain, it is clear that b and σ can point in any direction. The situation is now trickier at the boundary. Let us first consider the diffusive term $\sigma(X_t)dW_t$. Heuristically, $\frac{dW_t}{dt} \approx \frac{W_{t+h} - W_t}{h}$ is a Gaussian random variable with support equal to $\mathbb{R}^d$ meaning that it can take arbitrarily large positive and negative values with positive probability. For this reason, at the boundary, σ cannot have any transverse directions to the boundary, or else the particle can be pushed out quite far from the domain due to the unbounded fluctuations of the Gaussian random variable. Whence, the only possible directions for σ are those that are tangential to the boundary (highlighted in blue on Figure 1.2 below). At this stage, the geometry and the curvature of the domain enter into play for b as displayed on Figure 1.2:

- In the absence of curvature at point (ii), the volatility σ is keeping the particle on the boundary of the domain. Hence, the drift b can vanish here or point inward.
- For a locally convex boundary as in (i), although tangential, the volatility σ is pushing the particle outside the domain. Therefore, the drift b has to compensate this movement in order to bring back the particle inside the domain. Notice that in this case b cannot vanish.

[3] Think of a first order Taylor's expansion of ϕ around its maxima x with $u \equiv D\phi(x)^\top$.

- For a locally concave boundary as in (iii), the drift b can even be outward pointing since the volatility is pushing the particle inside the domain.

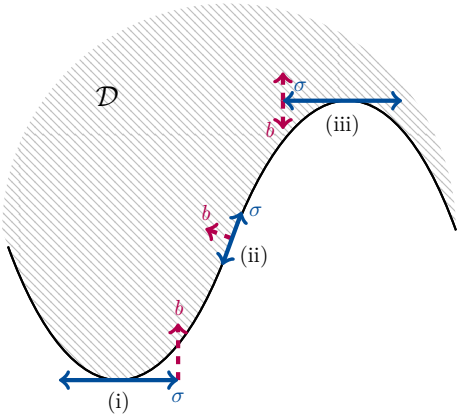

FIGURE 1.2: **Geometric intuitions:** Interplay between the geometry/curvature of $\mathcal{D}$ and the coefficients (b, σ).

All in all, the geometric intuitions can be formulated as follows.

An informal characterization

There exists a $\mathcal{D}$-valued solution to (1.1.4) started from any $X_0 \in \mathcal{D}$ if and only if, at all boundary points, the following geometric conditions hold

$$\textit{tangential volatility } \sigma \quad \text{and} \quad \textit{inward pointing compensated drift } (b - F(\sigma)), \tag{1.1.5}$$

where F quantifies the contribution due to σ in the tangent direction.

1.1.4 Existing results and sticking points

Second order characterization

We now turn to a more formal mathematical derivation mimicking that of Section 1.1.2. Assume that there exists a $\mathcal{D}$-valued solution X to (1.1.4) started from $x \in \mathcal{D}$. Let $\phi : \mathbb{R}^d \to \mathbb{R}$ be a smooth function such that $\max_{\mathcal{D}} \phi = \phi(x)$. Since $\phi(X_t) \leq \phi(x)$, by Itô's lemma

$$\int_0^t \mathcal{L}\phi(X_s)ds + \int_0^t (D\phi\sigma)(X_s)dW_s \leq 0, \quad t \geq 0, \tag{1.1.6}$$

where $\mathcal{L}\phi := D\phi b + \frac{1}{2}\operatorname{Tr}(D^2\phi\sigma\sigma^\top)$. Taking the expectation in the above, dividing by t and sending $t \to 0$ yield

$$\mathcal{L}\phi(x) \leq 0,$$

or equivalently, after a small effort[4],

$$\langle u, b(x)\rangle + \frac{1}{2}\operatorname{Tr}(v\sigma(x)\sigma(x)^\top) \leq 0 \quad x \in \mathcal{D},\ (u, v) \in \mathcal{N}^2_{\mathcal{D}}(x), \tag{1.1.7}$$

[4]Think of a second order Taylor's expansion of ϕ around its maxima x with $(u, v) \equiv (D\phi(x)^\top, D^2\phi(x))$.

where $\mathcal{N}^2_{\mathcal{D}}(x)$ is the second order normal cone to the set $\mathcal{D}$ at a point x defined as

$$\mathcal{N}^2_{\mathcal{D}}(x) = \{(u,v) \in \mathbb{R}^d \times \mathbb{S}^d, \langle u, y-x\rangle + \frac{1}{2}\langle y-x, v(y-x)\rangle \leq o(\|y-x\|^2), \forall\, y \in \mathcal{D}\}.$$

The condition (1.1.7) is also sufficient for invariance by virtue of the celebrated positive maximum principle of Ethier and Kurtz [54, Theorem 4.5.4].

This yields a characterization of the invariance in terms of the second order normal cone. Nevertheless, the formulation (1.1.7) has two main drawbacks. First, the geometric intuitions (1.1.5) cannot be read directly on (1.1.7). Second, computing the second order normal cone can be daunting in practice as opposed to the first order normal cone. Whence, it would be preferable to have a characterization using the first order normal cone as in the deterministic setting. Inspecting (1.1.7) and recalling that $(u,v) \equiv (D\phi(x), D^2\phi(x))$, one is tempted to perform an integration by parts on the term $\mathrm{Tr}(v\sigma(x)\sigma(x)^\top)$ to recover the gradient $u = D\phi(x)^\top$. In order to do so, additional regularity assumptions on σ such as differentiability are needed.

First order characterization

Looking back at (1.1.6) and assuming that σ is twice differentiable, one could reapply Itô's lemma on the term $(D\phi\sigma)(X)$ to get (recall that $X_0 = x$)

$$\begin{aligned} 0 \geq & \int_0^t \mathcal{L}\phi(X_s)ds + D\phi(x)\sigma(x)W_t \\ & + \int_0^t \left(\int_0^s D(D\phi\sigma)(X_u)\sigma(X_u)dW_u\right)^\top dW_s + \int_0^t\int_0^s \mathcal{L}(D\phi\sigma)(X_u)du\, dW_s. \end{aligned} \tag{1.1.8}$$

One could now study the pathwise small time behavior without taking the expectation. For illustration purposes we restrict to the one dimensional case $d = 1$ and set

$$\gamma : y \to D(D\phi\sigma)(y)\sigma(y) = D^2\phi(y)\sigma^2(y) + D\phi(y)D\sigma(y)\sigma(y).$$

The double stochastic integral term in (1.1.8) reads

$$\begin{aligned} \int_0^t\int_0^s \gamma(X_u)dW_u dW_s &= \gamma(x)\int_0^t\int_0^s dW_u dW_s + \int_0^t\int_0^s (\gamma(X_u)-\gamma(X_0))dW_u dW_s \\ &= \frac{\gamma(x)}{2}(W_t^2 - t) + O(t^{1+\epsilon}) \end{aligned}$$

where the estimate involving the random variable $\epsilon(\omega) > 0$ can be derived heuristically under suitable Hölder regularity on σ and the fact that W_t behaves very roughly like $\sqrt{t}$. With the same reasoning the last term in (1.1.8) can be shown to be dominated by $t^{3/2}$ as t goes to 0. Further, since

$$\liminf_{t\to 0} \frac{W_t^2}{t} = 0 \text{ and } \limsup_{t\to 0} \frac{W_t^2}{t} = +\infty,$$

dividing (1.1.8) by t and taking the lim sup for $t \to 0$ yields

$$0 \geq \mathcal{L}\phi(x) + \limsup_{t\to 0} D\phi(x)\sigma(x)\frac{W_t}{t} - \frac{\gamma(x)}{2} + \mathbb{1}_{\{\gamma(x)>0\}}\text{“}+\infty\text{”}.$$

Since $\frac{W_t}{t}$ is a centered Gaussian random variable with variance 1, the previous inequality is possible only if

(i) $D\phi(x)\sigma(x) = 0$,

(ii) $\gamma(x) \leq 0$,

(iii) $0 \geq \mathcal{L}\phi(x) - \frac{\gamma(x)}{2} = D\phi(x)\left(b(x) - \frac{1}{2}D\sigma(x)\sigma(x)\right)$.

The informal study of the small time path behavior of double stochastic integrals can be made rigorous by appealing to the law of the iterated logarithm for Brownian motion as shown in [30], see also [23, Lemma 2.1]. Dropping the restriction $d = 1$, (i) and (iii) lead to the following characterization first derived by Doss [47] and later on by Da Prato and Frankowska [38]. Under suitable regularity assumptions on (b, σ) there exists a $\mathcal{D}$-valued solution to (1.1.4) started from any $X_0 \in \mathcal{D}$ if and only if

$$\sigma(x)^\top u = 0 \quad \text{and} \quad \langle u, b(x) - \frac{1}{2}\sum_{j=1}^{d} D\sigma^j(x)\sigma^j(x)\rangle \leq 0, \quad x \in \mathcal{D},\ u \in \mathcal{N}_{\mathcal{D}}^1(x), \tag{1.1.9}$$

where $\sigma^j(x)$ denotes the j-th column of the matrix $\sigma(x)$ and $D\sigma^j(x)$ is the Jacobian of $\sigma^j(x)$ at x.

The conditions (1.1.9) reflect exactly the geometric intuitions (1.1.5) where the contribution $F(\sigma)$ due to the volatility σ in the tangent direction is quantified by means of the so-called Stratonovich correction term

$$F(\sigma) = \frac{1}{2}\sum_{j=1}^{d} D\sigma^j\sigma^j.$$

Historically, the correction term was named after the Russian mathematician Ruslan L. Stratonovich and allows one to recover the chain rule of ordinary calculus lost with Itô's theory. The Stratonovich correction term arises in a multitude of problems such as Wong and Zakaï approximation result [113], the support theorem of Stroock and Varadhan [107], the Milstein method for numerical simulations... and has numerous applications in Physics. Although disparate, these problems, along with invariance/viability problems, rely on the traditional chain rule, which explains the appearance of the Stratonovich correction term.

On the practical side, the formulation (1.1.9) requires strong regularity assumptions on the coefficients, namely on σ. The Stratonovich correction term makes sense only when σ is differentiable, which is too restrictive for applications. Indeed, set $\mathcal{D} = \mathbb{R}_+$ and consider the square-root process

$$dX_t = b(X_t)dt + \sqrt{X_t}dW_t.$$

Here, $\sigma : x \to \sqrt{x}$ is not differentiable at the boundary point $x = 0$. Although the Stratonovich term is not well defined in this case, one could still stubbornly consider a naive limit at the point 0 with $u = -1$ (since $\mathcal{N}_{\mathcal{D}}^1(0) = \mathbb{R}_-$) so that

$$0 \geq \lim_{x\to 0}\langle u, b(x) - \frac{1}{2}D\sigma(x)\sigma(x)\rangle = \lim_{x\to 0}(-1)\left(b(x) - \frac{1}{2}\frac{1}{2\sqrt{x}}\sqrt{x}\right),$$

yielding

$$b(0) \geq \frac{1}{4},$$

which is too strong since the well known invariance condition for the square-root process is $b(0) \geq 0$. Intuitively, when the process approaches zero the diffusion term $\sqrt{X_t}dW_t$ goes to 0 and X behaves as the deterministic equation (1.1.1) yielding the inward pointing drift

condition $b(0) \geq 0$. This example illustrates that on the one hand even simple diffusions as the square-root process do not satisfy the assumptions needed for (1.1.9). On the other hand approximating the Stratonovich term on the boundary of the domain by a sequence of the value it takes at interior points does not provide the correct invariance characterization.

Noting that $a := \sigma^2$ is differentiable for the square-root process, one is tempted to derive a first order characterization similar to (1.1.9) in terms of $a := \sigma\sigma^\top$ in place of σ by imposing the regularity assumptions on the level of a. This desire stems from the following two observations. As already seen on (1.1.7) the characterization depends on σ only through $a = \sigma\sigma^\top$ and an integration by parts would still be possible on the term $\mathrm{Tr}(v\sigma(x)\sigma(x)^\top)$ when only a is assumed to be differentiable. On a deeper probabilistic level, the problem of finding a $\mathcal{D}$-valued solution to (1.1.4) can be exclusively reformulated using the support of the law of the process X which is entirely determined by a and b. Second, imposing the regularity assumptions on a rather than on σ extends considerably the range of validity of (1.1.9) to a wider class of models useful in practice such as affine and polynomial diffusions (see [48] and [35]). For these processes the mapping $a = \sigma\sigma^\top$ is smooth (with affine or polynomial dependence in x) but σ may fail to be differentiable mainly on the boundary of the domain, as already highlighted by the example of the square-root process.

1.1.5 Our contributions

1.1.5.1 Guessing game: beyond the Stratonovich drift

At the present stage, the problem at hand is to reformulate the characterization (1.1.9) in terms of $a = \sigma\sigma^\top$. It is straightforward that the tangential volatility condition $\sigma(x)^\top u = 0$ is equivalent to the condition

$$a(x)u = 0. \tag{1.1.10}$$

Guessing the correction term when σ is not differentiable is a more challenging task. The example of the square-root process already revealed that taking a naive limit of the Stratonovich correction term does not yield the correct conditions. Therefore a new expression for quantifying the contribution of the diffusion term in (1.1.5) needs to be found. One could therefore think of taking more involved limits of some suitable Stratonovich correction terms. Let us make this more precise.

Assume that the mapping a is differentiable and denote by $(e_1, \ldots, e_d)$ the canonical basis of $\mathbb{R}^d$. Fix $\epsilon > 0$ and consider the following regularization preserving the tangential volatility conditions

$$\sigma_\epsilon : x \to a(x)(a(x) + \epsilon I_d)^{-\frac{1}{2}}.$$

It is clear that σ_ϵ is Fréchet differentiable with a derivative given by the product rule[5]

$$D\sigma_\epsilon(x)h = Da(x)h(a(x) + \epsilon I_d)^{-\frac{1}{2}} + a(x)D((a + \epsilon I_d)^{-\frac{1}{2}})(x)h, \quad h \in \mathbb{R}^d.$$

[5] we use the same notation for the Fréchet derivative and the Jacobian matrix where we identify $Da(x)e_j = Da^j(x)$.

In particular, evaluating at $h = e_j$ for $j = 1, \dots, d$ yields that the Stratonovich correction term $\langle u, \sum_{j=1}^d D\sigma_\epsilon^j(x)\sigma_\epsilon^j(x)\rangle$ is equal to

$$\langle u, \sum_{j=1}^d Da^j(x)(a(x)+\epsilon I_d)^{-\frac{1}{2}}\sigma_\epsilon^j(x)\rangle + \langle u, a(x)\sum_{j=1}^d D((a+\epsilon I_d)^{-\frac{1}{2}})(x)e_j\sigma_\epsilon^j(x)\rangle.$$

By virtue of the tangential condition (1.1.10) the second term vanishes so that

$$\langle u, \sum_{j=1}^d D\sigma_\epsilon^j(x)\sigma_\epsilon^j(x)\rangle = \langle u, \sum_{j=1}^d Da^j(x)(a(x)+\epsilon I_d)^{-\frac{1}{2}}a(x)(a(x)+\epsilon I_d)^{-\frac{1}{2}}e_j\rangle \tag{1.1.11}$$

Since $a(x) \in \mathbb{S}_+^d$, its spectral decomposition is given by

$$a(x) = Q(x)\mathrm{diag}(\lambda_1, \dots, \lambda_r, 0, \dots, 0)Q(x)^\top,$$

where $r \leq d$ denotes the rank of $a(x)$, $\lambda_1 \geq \cdots \geq \lambda_r > 0$ and $Q(x)$ is an orthogonal matrix. Therefore,

$$\begin{aligned}(a(x)+\epsilon I_d)^{-\frac{1}{2}}a(x)(a(x)+\epsilon I_d)^{-\frac{1}{2}} &= Q(x)\,\mathrm{diag}\left[\left(\frac{\lambda_1}{\lambda_1+\epsilon}, \dots, \frac{\lambda_r}{\lambda_r+\epsilon}, 0, \dots, 0\right)\right]Q(x)^\top \\ &\xrightarrow[\epsilon\to 0]{} Q(x)\,\mathrm{diag}\left[\underbrace{(1, \dots, 1}_{r \text{ times}}, 0, \dots, 0)\right]Q(x)^\top.\end{aligned}$$

The last expression is nothing else but the projection matrix on the span of $a(x)$ which can be re-expressed as $a(x)a(x)^+$ where $a(x)^+$ denotes the Moore-Penrose pseudo-inverse[6] of $a(x)$. Therefore, sending ϵ to 0 in (1.1.11) yields the correction term

$$\lim_{\epsilon\to 0}\langle u, \sum_{j=1}^d D\sigma_\epsilon^j(x)\sigma_\epsilon^j(x)\rangle = \langle u, \sum_{j=1}^d Da^j(x)(aa^+)^j(x)\rangle. \tag{1.1.12}$$

Heuristically, combining (1.1.10) and (1.1.12) we arrive at the following characterization.

Main result 1 - Invariance characterization

Under suitable assumptions such as differentiabilty of $a := \sigma\sigma^\top$. There exists a $\mathcal{D}$-valued solution to (1.1.4) started from $X_0 \in \mathcal{D}$ if and only if

$$a(x)u = 0 \quad \text{and} \quad \langle u, b(x) - \frac{1}{2}\sum_{j=1}^d Da^j(x)(aa^+)^j(x)\rangle \leq 0 \tag{1.1.13}$$

for all $x \in \mathcal{D}$ and $u \in \mathcal{N}_{\mathcal{D}}^1(x)$.

1.1.5.2 First order characterization under weak regularity assumptions

We prove in Chapter 3 that the characterization (1.1.13) holds. This constitutes the first main result of the present thesis. The successive applications of Itô's Lemma as done in

[6] The Moore-Penrose pseudoinverse of a $m \times n$ matrix A is the unique $n \times m$ matrix A^+ satisfying: $AA^+A = A$, $A^+AA^+ = A^+$, AA^+ and A^+A are Hermitian.

(1.1.8) cannot be simply replicated. The crux resides in the fact that $\sigma(X)$ fails to be a semimartingale in general. We therefore had to develop new ideas before revisiting the strategy.

To the best of our knowledge, as opposed to the Stratonovich term, the correction term entering the characterization (1.1.13) appears for the first time in the literature and quantifies the forcing action F of (1.1.5) as follows

$$F(a) = \frac{1}{2}\sum_{j=1}^{d} Da^j(x)(aa^+)^j(x).$$

The projection term aa^+ is needed to stay in line with the geometric intuitions: b only has to compensate the tangential directional derivatives of a to keep the particle inside the domain.

Further, if the volatility matrix σ is differentiable, it follows from (1.1.12) that

$$\langle u, \sum_{j=1}^{d} D\sigma^j(x)\sigma^j(x)\rangle = \lim_{\epsilon\to 0}\langle u, \sum_{j=1}^{d} D\sigma_\epsilon^j(x)\sigma_\epsilon^j(x)\rangle = \langle u, \sum_{j=1}^{d} Da^j(x)(aa^+)^j(x)\rangle,$$

proving that (1.1.13) reduces to the characterization (1.1.9). Therefore, our result generalizes the works [38, 47] under weaker regularity assumptions, mainly on the volatility matrix σ.

Finally, going back to the square-root example, we have $a(x) = x$ so that $aa^+(x) = \mathbb{1}_{\{x>0\}}$. Taking $u = -1$ the conditions (1.1.13) at the boundary point 0 read

$$a(0) = 0 \quad \text{and} \quad b(0) \geq 0,$$

yielding the correct characterization.

The following table summarizes our findings so far.

	Existing 2nd order	Existing 1st order	Our result
Conditions	(1.1.7)	(1.1.9)	(1.1.13)
Normal cone	Second order	First order	First order
Main assumption on σ	Continuity	Differentiability	Differentiability of σ^2
All affine/polynomial diffusions	✓	✗	✓
Geometric intuitions as in (1.1.5)	✗	✓	✓

TABLE 1.1: Summary of the characteristics of the different characterizations.

1.1.5.3 Extension to jumps

One can introduce an additional source of randomness to (1.1.4) to account for jumps in the movement of the particle as follows

$$dX_t = b(X_t)dt + \sigma(X_t)dW_t + \int_{\mathbb{R}^d} \rho(X_{t-}, z)\left(\mu(dt, dz) - F(dz)dt\right), \quad X_0 = x, \tag{1.1.14}$$

where μ is a Poisson random measure on $\mathbb{R}_+ \times \mathbb{R}^d$ with compensator $dt \otimes F(dz)$ governing the frequency of the jumps arriving at random times and ρ is a function determining the size of jumps with suitable continuity properties.

The sources of randomness can be shown to be independent, one can therefore focus on gaining intuitions on the discontinuous part for the particle to remain in $\mathcal{D}$. If the particle X is (strictly) inside the domain right before a jump at time t then one trivially sees that it is not allowed to jump outside the domain, that is

$$X_{t-} + (\textit{size of the jump}) \quad \textit{belongs to} \quad \mathcal{D}.$$

At the boundary, the situation is more involved. On a microscopic scale, the particle can have either big or small jumps. If one observes the movement of the particle on a macroscopic scale, small jumps fluctuating very rapidly become invisible to the naked eye and one only sees a continuous trajectory. One therefore expects these small jumps to have the same behavior as a continuous random process with infinite variation, say Brownian motion. Inspired by (1.1.5), the tangential volatility intuition carries over as follows:

rapidly fluctuating small jumps need to be tangential to the boundary.

Further, as in the continuous case,

b should compensate these tangential movements to keep the particle inside the domain.

Once again, the previous heuristics combined with (1.1.13) can be mathematically interpreted as follows.

Main result 2 - Extension to jumps

Under suitable assumptions such as differentiabilty of $a := \sigma\sigma^\top$. There exists a $\mathcal{D}$-valued solution to (1.1.14) started from $X_0 \in \mathcal{D}$ if and only if

$$\begin{aligned} &x + \rho(x,z) \in \mathcal{D}, \text{ for } F\text{-almost all } z, \\ &\int |\langle u, \rho(x,z)\rangle| F(dz) < \infty, \qquad a(x)u = 0, \\ &\langle u, b(x) - \int \rho(x,z) F(dz) - \frac{1}{2}\sum_{j=1}^{d} Da^j(x)(aa^+)^j(x)\rangle \leq 0, \end{aligned}$$

for all $x \in \mathcal{D}$ and $u \in \mathcal{N}_{\mathcal{D}}^1(x)$.

This leads to the second main result of the thesis proved in Chapter 4 where an equivalent characterization in the semimartingale framework in terms of the differential characteristics triplet is also provided.

1.2 Stochastic Volterra equations

The second part of the thesis is dedicated to the study of d-dimensional stochastic integral Volterra equations of convolution type

$$X_t = g_0(t) + \int_0^t K(t-s)b(X_s)ds + \int_0^t K(t-s)\sigma(X_s)dW_s, \tag{1.2.1}$$

where W is a multi-dimensional Brownian motion, and the convolution kernel K, the function g_0 and coefficients b and σ satisfy suitable regularity and integrability conditions.

Setting $K \equiv 1$ and $g_0 \equiv X_0$ for some constant initial condition X_0, one recovers (1.1.4) written in integral form. Consequently, stochastic Volterra equations extend standard stochastic differential equation allowing for more flexibility in modeling. However, they do not fall in general in the semimartingale and Markovian frameworks, as illustrated by the Riemann-Liouville fractional Brownian motion $X_t = \int_0^t K_H(t-s)dW_s$ where K_H is the fractional kernel defined by

$$K_H : t \to t^{H-1/2}, \quad H \in (0, 1/2). \tag{1.2.2}$$

Nevertheless, we develop in the sequel several techniques to treat existence and uniqueness for different state spaces and clarify the link with standard stochastic differential equations. Our arguments avoid stochastic integration with respect to non-semimartingales, relying instead on tools from the theory of finite-dimensional deterministic convolution equations.

Before moving to a more detailed exposition, let us first discuss the motivations.

1.2.1 Motivations

Our motivation for studying such convolution equations is twofold. Stochastic Volterra equations arise as scaling limits of branching processes in population genetics and self-exciting processes in mathematical finance.

A biologist's inspiration for a chemist's problem

A chemist is looking to model a chemical diffusion-reaction interaction between two substances: a *reactant* and a *catalyst*. The first substance, called the *reactant* diffuses in space with some random motion. The chemical reaction will only take place locally in the presence of a second *catalytic* substance spread out in space, proportionally to the concentration of the *catalyst* at the contact point.

In order to describe the microscopic picture, her[7] biologist colleague suggests the following simple model inspired from population genetics. Think of the *reactant* as a system of n particles in one dimension moving independently according to a standard Brownian motion. The *catalyst* region at a certain time t is defined as the support of some deterministic measure $\rho_t(dx)$. Whenever a particle enters in the *catalyst* region and after spending a random time in the vicinity of the *catalyst*, it will either die or split into two new particles, with equal probabilities. The measure $\rho_t(dx)$ determines the local branching rate in space and time depending on the location and the concentration of the *catalyst*. Two typical examples are $\rho_t(dx) \equiv \bar{\rho}$ where the branching occurs in the entire space with constant rate $\bar{\rho}$ and

[7] The sex of the chemist was determined by a coin flip ... with a biased coin.

$\rho_t(dx) = \delta_0(dx)$ for a branching occurring with infinite rate only when the particle hits a highly concentrated single point *catalyst* located at 0. In case of branching, the two offspring particles evolve independently with the same spatial movement and branching mechanism as their parent.

One can view the *reactant* as a rescaled measure-valued process $(\bar{Y}^n_t(dx))_{t\geq 0}$ defined by

$$\bar{Y}^n_t(B) = \frac{\text{number of particles in } B \text{ at time } t}{n}, \quad \text{for every Borel set } B.$$

Sending the number of particles to infinity, one can establish the convergence towards a measure-valued macroscopic *reactant* $\bar{Y}$, coined *catalytic super-Brownian motion*, which solves an infinite dimensional martingale problem. Once there, weak uniqueness follows from a duality argument on the Laplace functional which is shown to be exponentially affine

$$\mathbb{E}\left[\exp(\langle u, \bar{Y}_t\rangle)\right] = \exp(\langle \Psi_t, \bar{Y}_0\rangle), \quad u \leq 0, \tag{1.2.3}$$

where $\langle f, \nu\rangle = \int f(x)\nu(dx)$ and Ψ is the mild solution of the following Riccati partial differential equation

$$\frac{\partial \Psi_t(x)}{\partial t} = \frac{1}{2}\Delta\Psi_t(x) + \frac{1}{2}\Psi^2_t(x)\frac{\rho_t(dx)}{dx}, \quad \Psi_0(x) = u(x), \tag{1.2.4}$$

with $\Delta = \partial^2/\partial x^2$ the second order Laplace operator.

Moreover, in the presence of a suitable deterministic *catalyst* $\rho = (\rho_t(dx))_{t\geq 0}$ having no atoms, the measure-valued process $\bar{Y}$ admits a density $\bar{Y}_t(dx) = \int Y_t(x)dx$ solution to the following stochastic partial differential equation

$$\frac{\partial Y_t(x)}{\partial t} = \frac{1}{2}\Delta Y_t(x) + \sqrt{Y_t(x)}\dot{W}^\rho(t,x), \tag{1.2.5}$$

where $\dot{W}^\rho(t,x)$ is a space-time noise with covariance structure determined by ρ, refer to Zähle [116] for more details. Denoting by $p_t(x) = (2\pi t)^{-1/2}\exp(-(x-y)^2/(2t))$ the heat kernel, solutions to (1.2.5) started from an initial curve Y_0 are considered in mild form

$$Y_t(x) = \int_{\mathbb{R}} p_t(x-y)Y_0(y)dy + \int_0^t\int_{\mathbb{R}} p_{t-s}(x-y)\sqrt{Y_s(y)}W^\rho(ds,dy). \tag{1.2.6}$$

The previous equation is only valid if ρ has no atoms. One could still heuristically set $\rho_t(dx) = \delta_0(dx)$ in (1.2.6) for the extreme case of a single point *catalyst* at 0. Then, the space-time noise reduces to a standard Brownian motion W so that evaluation at $x = 0$ yields

$$Y_t(0) = g_0(t) + \int_0^t \frac{(t-s)^{-1/2}}{\sqrt{2\pi}}\sqrt{Y_s(0)}dW_s, \tag{1.2.7}$$

where $g_0(t) = \int_{\mathbb{R}} p_t(y)Y_0(y)dy$. Therefore, $(Y_t(0))_{t\geq 0}$ solves a stochastic Volterra equation of the form (1.2.1). Needless to say, one is not allowed to plug the Dirac measure in (1.2.6). Indeed, in the presence of a single point *catalyst*, the *catalytic super-Brownian motion* does not admit a density at the *catalyst* position as shown by Dawson and Fleischmann [41] and (1.2.5) breaks down. This can be seen directly on (1.2.7) since the kernel $K : t \to t^{-1/2}$ fails to be locally square-integrable and the stochastic integral is not well defined in the sense of Itô's L^2-theory. Nevertheless, square-integrability can be informally recovered by bumping

the power of the kernel with a small $H > 0$ leading to the fractional kernel K_H defined in (1.2.2). For the kernel K_H, (1.2.7) makes sense. At the level of the microscopic branching process, this translates into substituting the heat flow Δ by a suitable spatial motion Δ_H. These heuristics are made rigorous in Mytnik and Salisbury [96].

As a by-product, one could also deduce the Laplace transform of $(Y_t(0))_{t\geq 0}$ from (1.2.3). Indeed, because of the smoothing property of the semigroup generated by Δ_H, one could start the Riccati partial differential equation from measure-valued initial conditions. Since $\langle u, \bar{Y}_t\rangle = \langle u, Y_t\rangle$, setting $u(x) = u\delta_0(dx)$ for some non-positive constant u and $Y_0(x) \equiv Y_0 \geq 0$, straightforward manipulations of (1.2.3)-(1.2.4) with $\rho_t(dx) = \delta_0(dx)$ lead to

$$\mathbb{E}\left[\exp(uY_t(0))\right] = \exp(\chi(t)Y_0), \tag{1.2.8}$$

where $\chi(t) = u + \frac{1}{2}\int_0^t \Psi_s^2(0)ds$ and $(\Psi_t(0))_{t\geq 0}$ solves the following Riccati-Volterra equation

$$\psi(t) = uK_H(t) + \frac{1}{2}\int_0^t K_H(t-s)\psi^2(s)ds. \tag{1.2.9}$$

A trader's viewpoint: a tick-by-tick foundation to rough volatility

In the field of quantitative finance, due to the rapid development of algorithmic high frequency trading, the theory of market microstructure is flourishing. As its name suggests, this theory deals with issues such as price formation, transaction costs and liquidity by a refined study of the market on a microscopic scale.

Mimicking the biologist of the previous section, the trader can proceed in two steps. First, conceive a microscopic model encoding the main stylized facts of modern market microstructure and then look for the limiting macroscopic picture emerging from the microscopic model after suitable rescaling. Such approach has been adopted in Jaisson's PhD dissertation [78], relying on Hawkes processes for modeling the tick-by-tick price of an asset at the microscopic level. The thesis was then followed by a series of paper culminating in the elegant construction of the so-called *rough Heston model* in [51]. Hawkes processes are a generalization of Poisson processes where the intensity of the jumps depends on the past realizations of the process. This dependence structure is specified by some kernel. For suitable kernels, one can generate feedback effects allowing to jointly account for the following stylized facts of the market:

(i) clustering behavior of the order flow: one can observe periods with a high number of trades followed by periods of a low number of trades,

(ii) order splitting/metaorders: large orders are not executed at once, they are split in time by trading algorithms to optimize transaction costs,

(iii) high degree of endogeneity: most of the orders lack real economic motivation, they are only sent by algorithms in reaction to other orders.

After embedding these observations into the intensity of the Hawkes process, a suitable rescaling yields the convergence towards the following model for the stock price S and its

stochastic variance V

$$dS_t = S_t\sqrt{V_t}dB_t, \quad S_0 = 1, \tag{1.2.10}$$

$$V_t = V_0 + \int_0^t K_H(t-s)\eta\sqrt{V_s}dW_s, \tag{1.2.11}$$

where K_H is the fractional kernel defined in (1.2.2) with $H \in (0, 1/2)$, $(V_0, \eta) \in \mathbb{R}_+^2$ and $B = \rho W + \sqrt{1-\rho^2}W^\perp$ with $\rho \in [-1, 1]$ and $(W, W^\perp)$ a two dimensional Brownian motion. The two-dimensional process $X = (S, V)$ falls into the class of stochastic Volterra equations of the form (1.2.1) for the diagonal kernel $K = \text{diag}\,[(1, K_H)]$.

The macroscopic model (1.2.10)-(1.2.11) was coined *rough Heston model*. The appellation is justified as follows. On the one hand, sample paths of V are locally Hölder continuous of any order strictly less than H, thereby less regular than standard Brownian motion (which corresponds to the case $H = 1/2$). Consequently models involving the fractional kernel have been dubbed *rough volatility models*.[8] On the other hand, for $H = 1/2$, (1.2.10)-(1.2.11) reduces to the *standard Heston model* [72] owing its popularity to the closed-form expression of the characteristic function of the log-price allowing for fast pricing and calibration by Fourier inversion techniques. Remarkably, an analogous formula continue to hold for the *rough Heston model*. Indeed, taking limits of the Fourier transform of the microscopic Hawkes model, El Euch and Rosenbaum [50] derive the affine transform

$$\mathbb{E}\left[\exp(v \log S_t)\right] = \exp(\chi(t)V_0), \quad v \in i\mathbb{R}, \tag{1.2.12}$$

in terms of a Riccati-Volterra equation

$$\psi(t) = \int_0^t K_H(t-s)F(v, \psi(s))ds, \tag{1.2.13}$$

where

$$F(v, u) = \frac{1}{2}(v^2 - v) + \rho\nu u + \frac{u^2}{2} \quad \text{and} \quad \chi(t) = \int_0^t F(v, \psi(s))ds.$$

Clearly, equations (1.2.12)-(1.2.13) share some similarities with (1.2.8)-(1.2.9). Combining the two examples, one would expect to get an affine transform for the joint process $(\log S, V)$.

Our approach: the macroscopic picture

It is a common approach in probability theory to study various phenomena through the associated limiting macroscopic object in order to gain valuable insights on the behavior of the more realistic microscopic picture. We adopt this perspective in this thesis taking as starting point stochastic Volterra equations in $\mathbb{R}^d$ of the form (1.2.1). By doing so, we avoid the infinite-dimensional analysis used for super-processes. We also circumvent the need to study scaling limits of Hawkes processes allowing for a more generic treatment of the following issues:

- What about more general kernels than the fractional kernel K_H?
- Can one go into higher dimensions?
- What about arbitrary dynamics, not necessarily of square-root type?

[8] Empirical studies on a very wide range of assets volatility time-series in [66, 15] revealed that the dynamics of the log-volatility are close to that of a fractional Brownian motion with a small Hurst parameter H of order 0.1, which is inconsistent with standard semimartingale models.

- How can one clarify the underlying *affine structure* behind the square-root process and recover uniqueness as well as tractability?

This is mainly the object of Chapter 5 where general existence problems for (1.2.1) are tackled on different state spaces. Uniqueness is then established for a specific class, namely *affine Volterra processes*, by completely characterizing the Fourier–Laplace functional of the solution in terms of a Riccati-Volterra integral deterministic equation. Once there, we shed some light on the correspondence with stochastic partial differential equations through different lifting procedures of the stochastic Volterra equation to infinite dimension. This leads to the complete characterization of the Markovian structure of stochastic Volterra equations and consequently opens the door for numerical approximation schemes. These results are collected in Chapters 6, 7 and 8.

1.2.2 Existence

We provide new existence results for (1.2.1) on several state spaces under weak conditions on the kernel and coefficients.

Unconstrained solutions: from Itô to Skorokhod

The problem of existence of unconstrained solutions, i.e. on $\mathbb{R}^d$, is laid bare by revisiting the classical strategies for stochastic differential equations (1.1.4). Historically, the first existence and uniqueness result dates back to the pioneering work of Kiyosi Itô [75] under global Lipschitz coefficients b and σ. Later on, Anatoliy Volodymyrovych Skorokhod [104] was able to derive existence of solutions under continuity and linear growth conditions on b and σ.

It was discovered later that the two notions of solutions are not equivalent. Given a filtered probability space $(\Omega, \mathcal{F}, \mathbb{F} := (\mathcal{F}_t)_{t\geq 0}, \mathbb{P})$ supporting a Brownian motion as input, Itô constructed a *pathwise* solution X adapted to the filtration generated by the Brownian motion, which corresponds to the concept of *strong solution*. Whereas in Skorokhod's proof, one constructs a filtered probability space $(\Omega, \mathcal{F}, \mathbb{F} := (\mathcal{F}_t)_{t\geq 0}, \mathbb{P})$ supporting a Brownian motion W and a process X such that (1.1.4) holds almost surely, this is know as a *weak solution*. A *weak solution* X is not necessarily adapted to the Brownian filtration. Clearly, every *strong solution* is a *weak solution*, but the converse is not true as illustrated by Itô-Tanaka's example

$$dX_t = \operatorname{sign}(X_t)dW_t.$$

We prove that similar type of existence results continue to hold for the stochastic Volterra equation (1.2.1).

Main result 3 - Existence of unconstrained solutions

Under mild assumptions on g_0 and $K \in L^2_{\mathrm{loc}}(\mathbb{R}_+, \mathbb{R}^d)$:

(i) If b and σ are Lipschitz continuous, then (1.2.1) admits a unique continuous strong solution X.

(ii) If b and σ are continuous with growth conditions, then (1.2.1) admits a continuous weak solution X.

Constrained solutions: back to the invariance problem

Similar to stochastic differential equations, one could also ask for existence of constrained solutions. As one would expect, the invariance/viability problem turns out to be more involved in the non-Markovian framework. We provide sufficient conditions for existence of $\mathbb{R}^d_+$-valued solutions to stochastic Volterra equations (1.2.1).

Under continuity and growth conditions on (b, σ), the previous result guarantees the existence of an unconstrained continuous weak solution X to the following modified equation

$$X_t = g_0(t) + \int_0^t K(t-s)b(X_s^+)ds + \int_0^t K(t-s)\sigma(X_s^+)dW_s,$$

for a suitable input curve g_0 and kernel K, where $x^+ = \max(0, x)$. Clearly, one needs to impose additional assumptions on g_0 to ensure the nonnegativity of X and drop the positive part in the previous equation so that X solves (1.2.1). Indeed, since $X_0 = g_0(0)$, it is straightforward that $g_0(0)$ should be in $\mathbb{R}^d_+$ for instance. Whence, the problem can be now formulated as follows:

Can one find a set $\mathcal{G}_K$ of initial input curves g_0 such that X remains in $\mathbb{R}^d_+$?

In order to gain intuitions, one can go back to the standard one-dimensional square-root process

$$X_t = g_0(t) - \lambda \int_0^t X_s^+ ds + \int_0^t \sqrt{X_s^+} dW_s.$$

Putting on the physicist's glasses, the behavior of X right after a fixed time t can be approximated by

$$X_{t+h} \approx X_t + g_0(t+h) - g_0(t) - h\lambda X_t^+ + \sqrt{X_t^+}(W_{t+h} - W_t), \quad \text{for small } h > 0.$$

Since $g_0(0) \geq 0$, if X hits zero for the first time at $t \geq 0$, then

$$X_{t+h} \approx g_0(t+h) - g_0(t).$$

Whence, one sees that X_{t+h} would remain nonnegative for all $h > 0$ if the set of initial input curves is given by

$$\mathcal{G}_1 = \{g_0 \text{ non-decreasing such that } g_0(0) \geq 0\}.$$

For an arbitrary kernel K, one would expect that an analogous "non-decreasing" condition should take into account the variations of K. This is indeed the case, we provide a set of *admissible input curves*

$$\mathcal{G}_K = \{g_0 \text{ "non-decreasing w.r.t. } K\text{" such that } g_0(0) \geq 0\}. \tag{1.2.14}$$

such that (1.2.1) admits a $\mathbb{R}^d_+$-valued weak solution.[9]

[9] The explicit form of $\mathcal{G}_K$ is provided in Chapter 6.

Main result 4 - Existence of constrained solutions in $\mathbb{R}^d_+$

Under mild assumptions on $K \in L^2_{\text{loc}}(\mathbb{R}_+, \mathbb{R}^d)$ such as nonnegativity and non-increasing monotonicity, if the geometric conditions

$$x_i = 0 \text{ implies } b_i(x) \geq 0 \text{ and } \sigma_i(x) = 0, \quad i = 1, \dots, n$$

hold, then (1.2.1) admits an $\mathbb{R}^d_+$-valued continuous weak solution for any suitable initial curve $g_0 \in \mathcal{G}_K$.

1.2.3 Affine lens: uniqueness and tractability

Having established existence, one can turn to the uniqueness problem which is a crucial requirement before using the model in practice. Uniqueness is a rather challenging problem when the coefficients are not globally Lipschitz even for standard stochastic differential equations. Indeed, when $K \equiv \text{id}$ is constant and equal to the d-dimensional identity matrix, weak uniqueness can be obtained in the special case of affine coefficients $a := \sigma\sigma^\top$ and b:

$$\begin{aligned} a(x) &= A^0 + x_1 A^1 + \cdots + x_d A^d \\ b(x) &= b^0 + x_1 b^1 + \cdots + x_d b^d, \end{aligned} \tag{1.2.15}$$

for some d-dimensional symmetric matrices A^i and vectors b^i. In this case solutions of (1.1.4) are called *affine diffusions*, we refer to Duffie, Filipović and Schachermayer [48] for a systematic treatment. More precisely, let X be an *affine diffusion* of the form (1.1.4) on some state space $\mathcal{D}$. For a suitable d-dimensional row vector u, the conditional Fourier–Laplace transform of X displays an exponentially affine form

$$\mathbb{E}\left[\exp\left(uX_T\right) \;\middle|\; \mathcal{F}_t\right] = \exp\left(\phi(T-t) + \psi(T-t)X_t\right), \tag{1.2.16}$$

where the real-valued function ϕ and the row-vector-valued function ψ satisfy the following Riccati ordinary differential equations:

$$\begin{aligned} \phi(t) &= \int_0^t \left(\psi(s)b_0 + \frac{1}{2}\psi(s)A^0\psi(s)^\top\right) ds \\ \psi(t) &= u + \int_0^t \left(\psi(s)B + \frac{1}{2}A(\psi(s))\right) ds, \end{aligned}$$

with $A(u) = (uA^1u^\top, \dots, uA^du^\top)$ and $B = (b^1 \;\cdots\; b^d)$.

Weak uniqueness for (1.1.4) is then established as a consequence of (1.2.16) leading also to tractability. Indeed, *affine diffusions* arguably constitute the most popular framework for building tractable multi-factor models in finance. They have been used to model a vast range of risk factors such as credit and liquidity factors, inflation and other macro-economic factors, equity factors, and factors driving the evolution of interest rates. In particular, affine stochastic volatility models, such as the Heston model [72] already encountered in the motivation, are very popular.

Dropping the restriction $K \equiv \text{id}$ and recalling the two motivations, one would hope for a similar expression for the Fourier–Laplace transform to hold for stochastic Volterra equations

(1.2.1) with affine coefficients of the form (1.2.15). Nevertheless, (1.2.16) is clearly a Markovian expression and one cannot expect it to hold in a non-Markovian framework. Therefore, a natural first step would be to derive a good potential Ansatz for the Fourier–Laplace transform in a non-Markovian setting.

Guessing game: deriving a good Ansatz

In order to gain valuable insights, we revisit the one dimensional standard square-root process

$$dX_t = -\lambda X_t dt + \eta\sqrt{X_t} dW_t,$$

which is clearly an *affine diffusion* on $\mathbb{R}_+$ with coefficients $a(x) = \eta^2 x$ and $b(x) = -\lambda x$. It follows from (1.2.16) that

$$\mathbb{E}\left[\exp(uX_T)|\mathcal{F}_t\right] = \exp\left(\psi(T-t)X_t\right), \quad \Re(u) \leq 0,$$

where

$$\psi' = -\lambda\psi + \frac{\eta^2}{2}\psi^2, \quad \psi(0) = u.$$

The variation of constants formula on the level of X and ψ yields

$$\mathbb{E}[X_s|\mathcal{F}_t] = e^{-\lambda(t-s)}X_t \quad \text{and} \quad \psi(t) = ue^{-\lambda t} + \frac{\eta^2}{2}\int_0^t e^{-\lambda(t-s)}\psi^2(s)ds, \quad s \geq t,$$

so that

$$\begin{aligned}\psi(T-t)X_t &= ue^{-\lambda(T-t)}X_t + \frac{\eta^2}{2}\int_0^{T-t}\psi^2(s)e^{-\lambda(T-t-s)}X_t ds \\ &= u\mathbb{E}[X_T|\mathcal{F}_t] + \frac{\eta^2}{2}\int_t^T \psi^2(T-s)\mathbb{E}[X_s|\mathcal{F}_t]ds,\end{aligned}$$

leading to the alternative expression for the Fourier–Laplace transform

$$\mathbb{E}\left[\exp\left(uX_T\right) \;\middle|\; \mathcal{F}_t\right] = \exp\left(\mathbb{E}[uX_T \mid \mathcal{F}_t] + \frac{1}{2}\int_t^T \psi^2(T-s)a(\mathbb{E}[X_s \mid \mathcal{F}_t])ds\right).$$

Compared to (1.2.16), the previous expression has the advantage of completely hiding the Markovian property of the process. For this reason it constitutes a good potential Ansatz for the Volterra case.

Affine Volterra processes

Our main result shows that, remarkably, the same expression continue to hold for the class of *affine Volterra processes* where the function ψ now solves a deterministic integral equation of convolution type. An *affine Volterra process* with state space $\mathcal{D}$ and coefficients (1.2.15) is defined as a $\mathcal{D}$-valued weak solution to (1.2.1) with $a := \sigma\sigma^\top$ on $\mathcal{D}$.

Main result 5 - Affine transform

Under mild assumptions on $K \in L^2_{\mathrm{loc}}(\mathbb{R}_+, \mathbb{R}^d)$, assume that X is a $\mathcal{D}$-valued *affine Volterra process* with coefficients given by (1.2.15) and that the corresponding *Riccati-Volterra equation*

$$\psi(t) = uK(t) + \int_0^t \left(\psi(s)B + \frac{1}{2}A(\psi(s))\right) K(t-s)\, ds,$$

admits a solution, then under an integrability condition, the following affine transform holds

$$\mathbb{E}\left[\exp\left(uX_T\right) \;\middle|\; \mathcal{F}_t\right] = \exp\left(\mathbb{E}[uX_T \mid \mathcal{F}_t] + \frac{1}{2}\int_t^T \psi(T-s)a(\mathbb{E}[X_s \mid \mathcal{F}_t])\psi(T-s)^\top ds\right).$$

Again, weak uniqueness is established as a corollary. Furthermore, it is possible to express the Fourier-Laplace in a form that is exponential-affine in the past trajectory $\{X_s,\, s \le t\}$. In particular, when $g_0 \equiv X_0$ for some constant $X_0 \in \mathbb{R}^d$, setting $t = 0$ yields the unconditional Fourier–Laplace

$$\mathbb{E}[\exp(uX_T)] = \exp\left(\phi(T) + \chi(T)X_0\right),$$

where

$$\phi = \int_0^{\cdot} (\psi b^0 + \frac{1}{2}\psi A^0 \psi^\top)(s)ds \quad \text{and} \quad \chi = u + \int_0^{\cdot} (\psi B + \frac{1}{2}A(\psi))(s)ds.$$

These last expressions agree with the formulations (1.2.8) and (1.2.12) of the motivations. Finally, the characterization of the Fourier-Laplace is a major strength of *affine Volterra processes*, making these equations enjoy an appealing trade-off between consistency and tractability.

1.2.4 Recovering Markovianity: correspondence with SPDEs

The motivation arising from population genetics highlights the correspondence between stochastic Volterra equations and degenerate stochastic partial differential equations. More precisely, solving the existence and uniqueness problem for one of these classes leads to solving the problem for the other class and vice-versa. Starting from certain stochastic partial differential equations, one can recover a stochastic Volterra equation by a projection procedure. Conversely, any stochastic Volterra equation can be lifted to an infinite dimension equation. Having already tackled the existence and uniqueness problem for stochastic Volterra equations, it seems natural to adopt the latter approach. As a by-product, this leads to the complete characterization of the Markovian structure of stochastic Volterra equations. We provide two such representations in terms of infinite-dimensional objects and specify their state space.

To keep the exposition simple, we restrict to the one dimensional case of stochastic Volterra equations with state space $\mathbb{R}_+$. Recalling our Main result 4, we assume that b and σ satisfy the boundary conditions

$$b(0) \ge 0 \quad \text{and} \quad \sigma(0) = 0,$$

and that g_0 belongs to the set of *admissible input curves* $\mathcal{G}_K$ given by (1.2.14) so that (1.2.1) admits a nonnegative weak solution X.

First representation: forward process

One starts with the following observation: conditional on $\mathcal{F}_t$, the shifted process $X^t := (X_{t+x})_{x\geq 0}$ solves the same stochastic Volterra equation (1.2.1) provided that g_0 is replaced by the following adjusted forward process

$$g_t(x) = \mathbb{E}\left[X_{t+x} - \int_0^x K(x-s)b(X_{t+s})ds \;\middle|\; \mathcal{F}_t\right], \quad x \geq 0,$$

that is, the process X^t solves the equation

$$X_x^t = g_t(x) + \int_0^x K(x-s)b(X_s^t)ds + \int_0^x K(x-s)\sigma(X_s^t)dW_s^t, \quad x \geq 0,$$

with $W^t := W_{t+\cdot} - W_t$. This suggests that, on the one hand, X is Markovian in the infinite dimensional curve $(g_t)_{t\geq 0}$. On the other hand, because the process X^t is nonnegative, one is tempted to claim that g_t is again an *admissible input curve* belonging to $\mathcal{G}_K$ as defined in (1.2.14). Our main result shows that this is indeed the case as it states that $\mathcal{G}_K$ is stochastically invariant with respect to the family $(g_t)_{t\geq 0}$. In other words, if we start from an initial *admissible input curve* $g_0 \in \mathcal{G}_K$, then g_t belongs to $\mathcal{G}_K$, for all $t \geq 0$. This in turn enables us to characterize the Markovian structure of X in terms of the the adjusted forward process $(g_t)_{t\geq 0}$. Furthermore, $(g_t)_{t\geq 0}$ can be realized as a $\mathcal{G}_K$-valued mild solution of the following stochastic partial differential equation of Heath–Jarrow–Morton-type

$$dg_t(x) = \left(\frac{d}{dx}g_t(x) + K(x)b(g_t(0))\right)dt + K(x)\sigma(g_t(0))dW_t, \quad g_0 \in \mathcal{G}_K,$$

and displays an affine characteristic functional if in addition b and σ^2 are affine.

Second representation: completely monotone case

More can be said when K is completely monotone on $(0,\infty)$, that is K is infinitely differentiable on $(0,\infty)$ such that $(-1)^n K^{(n)} \geq 0$ for all $n \geq 0$. By Bernstein's theorem, this is equivalent to stating that K is the Laplace transform of a nonnegative measure μ

$$K(t) = \int_0^\infty e^{-xt}\mu(dx), \quad t > 0.$$

For instance, the fractional kernel K_H given in (1.2.2) is completely monotone and its associated measure reads $\mu_H(dx) = \frac{x^{-1/2-H}}{\Gamma(1/2-H)}dx$. Exploiting the complete monotonicity property, starting from a solution to (1.2.1) with $g_0 \equiv 0$, a formal interchange of the integration order leads to

$$\begin{aligned} X_t &= \int_0^t K(t-s)\left(b(X_s)ds + \sigma(X_s)dW_s\right) \\ &= \int_0^\infty \int_0^t e^{-x(t-s)}\left(b(X_s)ds + \sigma(X_s)dW_s\right)\mu(dx) \\ &= \int_0^\infty U_t(x)\mu(dx) \end{aligned} \tag{1.2.17}$$

where $U_t(x) := \int_0^t e^{-x(t-s)}\left(b(X_s)ds + \sigma(X_s)dW_s\right)$. In particular, one recognizes the mild formulation of the following equation

$$dU_t(x) = \left(-xU_t(x) + b\left(\int_0^\infty U_t(y)\mu(dy)\right)\right)dt + \sigma\left(\int_0^\infty U_t(y)\mu(dy)\right)dW_t, \quad U_0(x) = 0,$$

for all $x \in \text{supp}(\mu)$, where $\text{supp}(\mu)$ denotes the support of the measure μ. The processes $(U(x))_{x\in(\text{supp}\,\mu)}$ share the same dynamics except that they mean revert at different speeds. In case μ is finitely supported by n points $(x_1, \dots, x_n)$ then the equation reduces to a n-dimensional system of standard differential equation of the form (1.1.4). Moreover, it follows from the representation (1.2.17) that X is Markovian in n-dimension $(U(x_1), \dots, U(x_n))$.

The two representations are linked by the following formula

$$g_t(x) = \int_0^\infty e^{-yx} U_t(y)\mu(dy), \quad t, x \geq 0.$$

Further, in the case of an *affine Volterra process*, i.e. $b(x) = -\lambda x$ and $\sigma(x) = \eta\sqrt{x}$, the Laplace transform of X can be shown to be an exponential affine functional of U_t

$$\mathbb{E}\left[\exp(uX_T) \,\middle|\, \mathcal{F}_t\right] = \exp\left(\int_0^\infty \chi(T-t, x)U_t(x)\mu(dx)\right)$$

where χ solves the following Riccati partial differential equation

$$\partial_t\chi(t,x) = -x\chi(t,x) + F\left(\int_0^\infty \chi(t,y)\mu(dy)\right), \quad \chi(0,x) = u, \quad x \in \text{supp}(\mu),$$

with $F(u) = -\lambda u + \frac{\eta^2}{2}u^2$. As previously, if the support of μ is finite then the equation reduces to a finite system of Riccati ordinary differential equations.

1.2.5 Approximation procedure: only SDEs are needed

At this stage, one natural idea would be to approximate any nonnegative measure μ with infinite support by a finite weighted sum of Dirac measures and expect to get the convergence of the associated equations. More precisely, one starts with a completely monotone kernel K such that the associated measure has infinite support, for instance the fractional kernel. Approximating the kernel K by a sequence of smooth kernels $(K^n)_{n\geq 1}$, one would expect the convergence of the following corresponding sequence of stochastic Volterra equations

$$X_t^n = g_0^n(t) + \int_0^t K^n(t-s)b(X_s^n)ds + \int_0^t K^n(t-s)\sigma(X_s^n)dW_s, \quad n \geq 1, \tag{1.2.18}$$

towards the initial equation (1.2.1) with kernel K for a suitable choice of g_0^n.

This is made precise in the following abstract stability result for stochastic Volterra equations.

Main result 6 - Stability of stochastic Volterra equations

Fix $T > 0$, under mild integrability assumptions, assume that $\|K^n - K\|_{L^2(0,T)} \to 0$ and $g_0^n \to g_0$ pointwise on $[0, T]$, then the sequence $(X^n)_{n\geq 1}$ of solutions to (1.2.18) is tight for the uniform topology and any point limit is a solution of the stochastic Volterra equation (1.2.1).

As an application to mathematical finance, we show that the previous Markovian representations combined with the stability result are of particular importance in practice as they lead to new numerical approximation schemes mainly for rough volatility modeling. Indeed, we design tractable multi-factor stochastic volatility models approximating rough volatility

models while still enjoying a Markovian structure. Further, we apply our procedure to the specific case of the rough Heston model. This in turn enables us to derive a numerical method for solving the corresponding Riccati-Volterra equation in this setting.

1.2.6 Lifting the Heston model

In the final chapter, we turn the previous point of view around by taking the finite dimensional Markovian model as starting point. We introduce a lifted version of the Heston model with n multifactors sharing the same Brownian motion but mean reverting at different speeds. The model nests as extreme cases the classical Heston model (when $n = 1$) and the rough Heston model (when n goes to infinity). We show that the lifted model enjoys the best of both worlds: Markovianity and satisfactory fitting of implied volatility smiles for short maturities. Further, our approach speeds up the calibration time and opens the door to time-efficient simulation schemes.

1.2.7 Structure

The second part of this thesis is structured as follows.

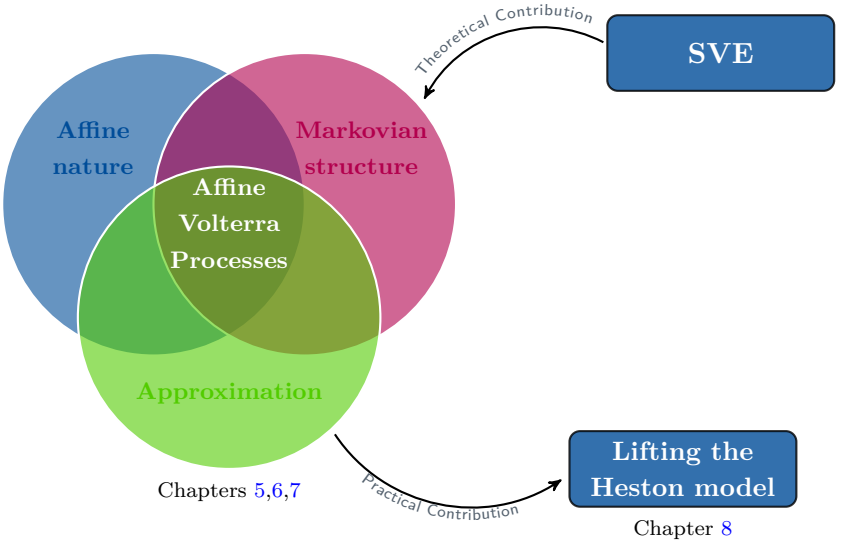

FIGURE 1.3: **Part II diagram**

Chapter 2

Introduction (Version française)

Résumé

La présente thèse traite de la théorie des équations stochastiques en dimension finie.

Dans la première partie, nous dérivons des conditions géométriques nécessaires et suffisantes sur les coefficients d'une équation différentielle stochastique pour l'existence d'une solution contrainte à rester dans un domaine fermé, sous de faibles conditions de régularité sur les coefficients.

Dans la seconde partie, nous abordons des problèmes d'existence et d'unicité d'équations de Volterra stochastiques de type convolutif. Ces équations sont en général non-Markoviennes. Nous établissons leur correspondance avec des équations en dimension infinie ce qui nous permet de les approximer par des équations différentielles stochastiques Markoviennes en dimension finie.

Enfin, nous illustrons nos résultats par une application en finance mathématique, à savoir la modélisation de la volatilité rugueuse. En particulier, nous proposons un modèle à volatilité stochastique assurant un bon compromis entre flexibilité et tractabilité.

Mots Clés: Invariance stochastique, équations de convolutions stochastiques, processus affines, volatilité rugueuse.

2.1 Invariance Stochastique

2.1.1 Le problème d'invariance ou de viabilité

Qu'est-ce qu'un problème d'invariance ou de viabilité? D'un point de vue abstrait, nous fixons un ensemble $\mathcal{D}$ d'un espace topologique et nous lâchons une particule à l'intérieur de $\mathcal{D}$. Nous dirons que l'ensemble $\mathcal{D}$ est invariant ou viable si la particule reste dans $\mathcal{D}$ à tout instant. Nous cherchons à caractériser l'invariance de $\mathcal{D}$ via des conditions sur les forces régissant le mouvement de la particule. Les problèmes d'invariance et de viabilité ont été intensivement étudiés dans la littérature, tout d'abord dans un cadre déterministe et plus tard dans un environnement aléatoire.

2.1.2 Le cadre déterminisite à la physicienne

Afin de mieux comprendre le problème, nous commençons par faire deux restrictions sur l'ensemble $\mathcal{D}$ et le mouvement de la particule. Nous considérons un sous-ensemble fermé $\mathcal{D}$ de $\mathbb{R}^d$ muni de son produit scalaire canonique $\langle \cdot, \cdot \rangle$ et nous supposons que le mouvement de la particule satisfait l'équation différentielle ordinaire autonome suivante

$$dX_t = b(X_t)dt, \quad X_0 = x, \tag{2.1.1}$$

où $X_t \in \mathbb{R}^d$ indique la position de la particule à l'instant $t \geq 0$ et $b : \mathbb{R}^d \to \mathbb{R}^d$ est la force motrice. De plus, nous supposons que b est Lipschitz, de sorte à ce que (2.1.1) admette une unique solution pour tout point initial $x \in \mathbb{R}^d$. En démarrant la solution d'un point $X_0 \in \mathcal{D}$, le problème d'invariance peut désormais être formulé comme suit:

A quelles conditions sur b, la solution X reste-t-elle, à tout instant, dans $\mathcal{D}$?

Supposons que X reste dans $\mathcal{D}$ jusqu'à l'intant t ou la particule touche le bord. Pour le physicien, $b(X_t) = \frac{dX_t}{dt}$ est le vecteur de vitesse instantanée de la particule à l'instant t, toujours tangent à la trajectoire de la particule. Par conséquent, à un point du bord $X_t \in \mathcal{D}$, la trajectoire reste dans le domaine tant que

le vecteur vitesse $b(X_t)$ est soit tangent au bord, soit pointe vers l'intérieur du domaine.

En effet, tout vecteur vitesse pointant vers l'extérieur pousse la particule à l'extérieur du domaine. Lorsque la particule est strictement à l'intérieur du domaine, la vitesse instantanée peut pointer dans n'importe quelle direction. Nous pouvons formuler rigouresement cette intuition en introduisant le concept de tangence via le cône normal du premier ordre défini par

$$\mathcal{N}^1_{\mathcal{D}}(x) = \left\{ u \in \mathbb{R}^d : \langle u, y - x \rangle \leq o(\|y - x\|), \forall\, y \in \mathcal{D} \right\}. \tag{2.1.2}$$

Le cône normal contient tous les vecteurs normaux qui pointent vers l'extérieur du domaine $\mathcal{D}$ en un point x. L'intuition du physicien se traduit donc par

$$\langle u, b(x) \rangle \leq 0, \quad x \in \mathcal{D},\ u \in \mathcal{N}^1_{\mathcal{D}}(x). \tag{2.1.3}$$

Si le point x se situe strictement à l'intérieur de $\mathcal{D}$, $\mathcal{N}^1_{\mathcal{D}}(x) = \{0\}$ et (2.1.3) est trivialement satisfaite pour les points situés strictement à l'intérieur du domaine et ceci indépendamment du comportement de b.

Afin de confirmer mathématiquement cette intuition, supposons que $\mathcal{D}$ est invariant sous (2.1.1) démarrée au point $x \in \mathcal{D}$, i.e. $X_t \in \mathcal{D}$, pour tout $t \geq 0$. Soit $\phi : \mathbb{R}^d \to \mathbb{R}$ une fonction régulière telle que $\max_{\mathcal{D}} \phi = \phi(x)$. Puisque $\phi(X_t) \leq \phi(x)$, pour tout $t \geq 0$, la règle de dérivation en chaîne entraîne

$$\int_0^t D\phi(X_s) b(X_s) ds \leq 0, \quad t \geq 0,$$

où $D\phi(x)$ est le vecteur transposé du gradient de ϕ évalué au point x. En divisant ce qui précède par t et en envoyant $t \to 0$, nous obtenons par continuité

$$D\phi(x) b(x) \leq 0$$

qui, après un petit effort supplémentaire, implique (2.1.3).[1] La condition est également suffisante pour l'invariance. Ce résultat est connu sous le nom de condition de tangence de Nagumo, nommée après le mathématicien japonais Mitio Nagumo, qui fut le premier à dériver cette caractérisation dans un article fondateur publié en 1942 en allemand [97]. Le théorème a été apparemment oublié et redécouvert dans différents contextes au cours des vingt prochaines années. Pour un traitement détaillé des problèmes d'invariance et de viabilité dans un cadre déterministe, nous renvoyons à [9, 24].

2.1.3 Une intuition géometrique dans un environnement aléatoire

Dans un environnement aléatoire, l'équation différentielle ordinaire autonome (2.1.1) est remplacée par une équation différentielle stochastique Markovienne homogène en temps

$$dX_t = b(X_t)dt + \sigma(X_t)\underbrace{dW_t}_{\text{bruit}}, \quad X_0 = x, \tag{2.1.4}$$

où $\sigma : \mathbb{R}^d \mapsto \mathbb{R}^{d\times d}$ est continu satisfaisant des conditions de croissance et W est un mouvement Brownien d-dimensionnel, i.e. l'analogue d'une marche aléatoire en temps continu.

Les règles du jeu restent inchangées: trouver des conditions nécessaires et suffisantes sur le vecteur de drift b et la matrice de volatilité σ sous lesquelles il existe une solution faible de (2.1.4) à valeur dans $\mathcal{D}$, pour tout point de départ $x \in \mathcal{D}$.

En reprenant l'intuition du physicien, tant que la particule reste strictement à l'intérieur du domaine, il est clair que b et σ peuvent pointer dans n'importe quelle direction. La situation est désormais plus délicate au bord du domaine. Considérons en premier le terme de diffusion $\sigma(X_t)dW_t$. Heuristiquement, $\frac{dW_t}{dt} \approx \frac{W_{t+h}-W_t}{h}$ est une variable aléatoire gaussienne dont le support est égal à $\mathbb{R}^d$, ce qui signifie qu'elle peut prendre des valeurs positives ou négatives arbitrairement grandes avec probabilité positive. Pour cette raison, au bord du domaine, σ ne peut pas avoir de direction transversale au bord, sinon la particule peut être projeter très loin du domaine en raison des fluctuations non bornées de la variable aléatoire gaussienne. Ainsi, les seules directions possibles pour σ sont celles qui sont tangentes au bord (illustrées en bleu sur la figure 2.1 ci-dessous). A ce stade, la géométrie et la courbure du domaine entrent en jeu pour déterminer les directions possibles pour le vecteur b, comme indiqué sur la figure 2.1:

- En l'absence de courbure au point (ii), la volatilité σ maintient la particule au bord du domaine. Par conséquent, le vecteur de drift b peut s'annuler en ce point ou alors pointer vers l'intérieur du domaine.
- Pour un bord localement convexe comme dans (i), bien que tangentielle, la volatilité σ pousse la particule hors du domaine. Par conséquent, le vecteur b doit compenser ce mouvement pour ramener la particule à l'intérieur du domaine. Dans ce cas, b ne peut pas s'annuler.
- Pour un bord localement concave comme dans (iii), le vecteur b peut même pointer vers l'extérieur vu que la volatilité pousse la particule à l'intérieur du domaine.

[1] Via la formule de Taylor au premier ordre pour ϕ autour de son maxima x avec $u \equiv D\phi(x)^\top$.

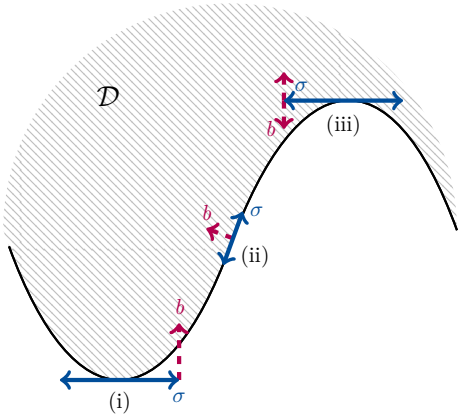

FIGURE 2.1: **Intuitions géometriques:** Géometrie/courbure de $\mathcal{D}$ et directions des coefficients (b, σ).

Dans l'ensemble, les intuitions géométriques peuvent être formulées comme suit.

Une caractérisation informelle

Il existe une solution à (2.1.4) à valeurs dans $\mathcal{D}$ pour n'importe quel point de départ $X_0 \in \mathcal{D}$ si et seulement si, en tout point du bord, les conditions géométriques suivantes sont satisfaites:

$$\textit{volatilité tangentielle } \sigma; \ \textit{drift compensé } (b - F(\sigma)) \ \textit{pointant vers l'intérieur}, \quad (2.1.5)$$

où F quantifie la contribution dûe à σ dans la direction tangentielle.

2.1.4 Résultats existant et points bloquants

Une caractérisation du second ordre

Nous passons désormais à une dérivation mathématique plus formelle imitant celle de la section 2.1.2. Supposons qu'il existe une solution X de (2.1.4) à valeurs dans $\mathcal{D}$ démarrée d'un point $x \in \mathcal{D}$. Soit $\phi : \mathbb{R}^d \to \mathbb{R}$ une fonction régulière telle que $\max_{\mathcal{D}} \phi = \phi(x)$. Comme $\phi(X_t) \leq \phi(x)$, le lemme d'Itô entraîne

$$\int_0^t \mathcal{L}\phi(X_s)ds + \int_0^t (D\phi\sigma)(X_s)dW_s \leq 0, \quad t \geq 0, \tag{2.1.6}$$

où $\mathcal{L}\phi := D\phi b + \frac{1}{2}\operatorname{Tr}(D^2\phi\sigma\sigma^\top)$. En prenant l'espérance de l'expression précèdente avant de diviser par t et d'envoyer $t \to 0$, nous obtenons

$$\mathcal{L}\phi(x) \leq 0,$$

ou, de manière équivalente, après un petit effort[2],

$$\langle u, b(x)\rangle + \frac{1}{2}\operatorname{Tr}(v\sigma(x)\sigma(x)^\top) \leq 0 \quad x \in \mathcal{D}, \ (u, v) \in \mathcal{N}^2_{\mathcal{D}}(x), \tag{2.1.7}$$

[2]via la formule de Taylor au second ordre de ϕ autour de son maxima x avec $(u, v) \equiv (D\phi(x)^\top, D^2\phi(x))$.

où $\mathcal{N}^2_{\mathcal{D}}(x)$ est le cône normal de second ordre de l'ensemble $\mathcal{D}$ en un point x défini par

$$\mathcal{N}^2_{\mathcal{D}}(x) = \{(u,v) \in \mathbb{R}^d \times \mathbb{S}^d, \langle u, y-x\rangle + \frac{1}{2}\langle y-x, v(y-x)\rangle \leq o(\|y-x\|^2), \forall\, y \in \mathcal{D}\}.$$

Par ailleurs, la condition (2.1.7) est suffisante pour l'invariance en vertu du principe du maximum d'Ethier et de Kurtz [54, Théorème 4.5.4].

Cela donne une caractérisation de l'invariance en termes du cône normal du second ordre. Néanmoins, la formulation (2.1.7) présente deux inconvénients majeurs. Premièrement, nous ne pouvons pas directement lire sur (2.1.7) les intuitions géométriques (2.1.5). Deuxièmement, le calcul du cône normal du second ordre peut être lourd en pratique, par opposition à celui du cône normal du premier ordre. Il serait donc préférable d'avoir une caractérisation exclusivement en terme du cône normal du premier ordre, comme dans le cadre déterministe. En inspectant (2.1.7) et en se rappelant que $(u,v) \equiv (D\phi(x), D^2\phi(x))$, nous sommes tentés d'effectuer une intégration par parties sur le terme $\mathrm{Tr}(v\sigma(x)\sigma(x)^\top)$ afin de récupérer le gradient $u = D\phi(x)^\top$. Pour ce faire, des hypothèses de régularité supplémentaires sur σ, telle que sa dérivabilité, sont nécessaires.

Une caractérisation du premier ordre

En observant (2.1.6) et en supposant que σ est deux fois dérivable, nous pouvons réappliquer le lemme d'Itô au terme $(D\phi\sigma)(X)$ pour obtenir (rappelons que $X_0 = x$)

$$\begin{aligned} 0 \geq & \int_0^t \mathcal{L}\phi(X_s)ds + D\phi(x)\sigma(x)W_t \\ & + \int_0^t \left(\int_0^s D(D\phi\sigma)(X_u)\sigma(X_u)dW_u\right)^\top dW_s + \int_0^t\int_0^s \mathcal{L}(D\phi\sigma)(X_u)du\, dW_s. \end{aligned} \tag{2.1.8}$$

L'idée est désormains d'étudier le comportement en temps court de l'expression précédente sans prendre l'espérance. Afin de simplifier les notations, nous nous limitons au cas unidimensionnel $d = 1$ et nous définissons

$$\gamma : y \to D(D\phi\sigma)(y)\sigma(y) = D^2\phi(y)\sigma^2(y) + D\phi(y)D\sigma(y)\sigma(y).$$

Le terme de double intégrales stochastiques dans (2.1.8) se décompose de la façon suivante

$$\begin{aligned} \int_0^t\int_0^s \gamma(X_u)dW_u dW_s &= \gamma(x)\int_0^t\int_0^s dW_u dW_s + \int_0^t\int_0^s (\gamma(X_u)-\gamma(X_0))dW_u dW_s \\ &= \frac{\gamma(x)}{2}(W_t^2 - t) + O(t^{1+\epsilon}) \end{aligned}$$

où l'estimation dépendant de la variable aléatoire $\epsilon(\omega) > 0$ peut être obtenue de manière heuristique sous des hypothèses de régularité Höldérienne convenables sur σ et le fait que W_t se comporte, très grossièrement, comme $\sqrt{t}$. Par le même raisonnement, nous pouvons montrer que le dernier terme de (2.1.8) est dominé par $t^{3/2}$ lorsque t tend vers 0. Par ailleurs, comme

$$\liminf_{t\to 0} \frac{W_t^2}{t} = 0 \quad \text{et} \quad \limsup_{t\to 0} \frac{W_t^2}{t} = +\infty,$$

en divisant (2.1.8) par t et en prenant la lim sup pour $t \to 0$, nous obtenons

$$0 \geq \mathcal{L}\phi(x) + \limsup_{t\to 0} D\phi(x)\sigma(x)\frac{W_t}{t} - \frac{\gamma(x)}{2} + \mathbb{1}_{\{\gamma(x)>0\}}\text{“}+\infty\text{”}.$$

Comme $\frac{W_t}{t}$ est une variable aléatoire gaussienne centrée avec variance 1, l'inégalité précédente n'est possible que si

(i) $D\phi(x)\sigma(x) = 0$,

(ii) $\gamma(x) \leq 0$,

(iii) $0 \geq \mathcal{L}\phi(x) - \frac{\gamma(x)}{2} = D\phi(x)\left(b(x) - \frac{1}{2}D\sigma(x)\sigma(x)\right)$.

L'étude informelle précédente du comportement en temps court des intégrales stochastiques doubles peut être rendue rigoureuse en faisant appel à la loi du logarithme itéré du mouvement Brownien, comme dans [30], voir aussi [23, Lemme 2.1]. En relâchant la restriction $d = 1$, (i) et (iii) conduisent à la caractérisation suivante, dérivée d'abord par Doss [47] et plus tard par Da Prato et Frankowska [38]. Sous des hypothèses de régularité appropriées sur (b, σ), il existe une solution à (2.1.4) à valeurs dans $\mathcal{D}$ pour tout point de départ $X_0 \in \mathcal{D}$ si et seulement si

$$\sigma(x)^\top u = 0 \quad \text{et} \quad \langle u, b(x) - \frac{1}{2}\sum_{j=1}^{d} D\sigma^j(x)\sigma^j(x)\rangle \leq 0, \quad x \in \mathcal{D},\ u \in \mathcal{N}_{\mathcal{D}}^1(x), \tag{2.1.9}$$

où $\sigma^j(x)$ désigne la j-ème colonne de la matrice $\sigma(x)$ et $D\sigma^j(x)$ est la matrice Jacobienne de $\sigma^j(x)$ évaluée au point x.

Les conditions (2.1.9) reflètent exactement les intuitions géométriques (2.1.5) où la contribution $F(\sigma)$ due à la volatilité σ dans la direction tangentielle est quantifiée via le terme correctif de Stratonovich

$$F(\sigma) = \frac{1}{2}\sum_{j=1}^{d} D\sigma^j\sigma^j.$$

Historiquement, le terme correctif a été nommé après le mathématicien russe Ruslan L. Stratonovich et permet de retrouver la règle de dérivation en chaîne perdue avec la théorie d'Itô. Le terme correctif de Stratonovich apparaît dans une multitude de problèmes tels que le résultat d'approximation de Wong et Zakaï [113], le théorème de support de Stroock et Varadhan [107], la méthode de Milstein pour les simulations numériques ... avec de nombreuses applications en physique. Bien que différents, les problèmes énumérées précédemment, tout comme les problèmes d'invariance et de viabilité, reposent sur la règle de dérivation en chaîne traditionnelle, ce qui explique l'apparition du terme correctif Stratonovich.

Sur le plan pratique, la formulation (2.1.9) nécessite de fortes hypothèses de régularité sur les coefficients, à savoir sur σ. Le terme correctif de Stratonovich n'a de sens que lorsque σ est dérivable, ce qui est très restrictif pour les applications. En effet, nous fixons $\mathcal{D} = \mathbb{R}_+$ et nous considérons le processus racine carrée suivant

$$dX_t = b(X_t)dt + \sqrt{X_t}dW_t.$$

Ici, $\sigma : x \to \sqrt{x}$ n'est pas dérivable au point du bord $x = 0$. Bien que le terme Stratonovich ne soit pas bien défini dans ce cas, nous pouvons toujours prendre une limite naïve au point 0 avec $u = -1$ (vu que $\mathcal{N}_{\mathcal{D}}^1(0) = \mathbb{R}_-$), ce qui donne

$$0 \geq \lim_{x\to 0}\langle u, b(x) - \frac{1}{2}D\sigma(x)\sigma(x)\rangle = \lim_{x\to 0}(-1)\left(b(x) - \frac{1}{2}\frac{1}{2\sqrt{x}}\sqrt{x}\right),$$

entraînant

$$b(0) \geq \frac{1}{4},$$

qui est une condition trop forte; puisque la bonne condition d'invariance pour le processus racine carrée est $b(0) \geq 0$. Intuitivement, lorsque le processus approche zéro, le terme de diffusion $\sqrt{X_t}dW_t$ tend vers 0 et X se comporte comme dans l'équation déterministe (2.1.1) donnant la condition de drift pointant vers l'intérieur $b(0) \geq 0$. Cet exemple illustre que d'une part, même pour des diffusions simples, en l'occurrence le processus racine carrée, les hypothèses requises pour (2.1.9) ne sont pas remplies. D'autre part, l'approximation naïve du terme correctif de Stratonovich sur le bord du domaine par les valeurs qu'il prend à l'intérieur du domaine ne fournit pas la bonne caractérisation de l'invariance.

En observant que $a := \sigma^2$ est dérivable pour le processus racine carrée, nous sommes tentés de dériver une caractérisation du premier ordre similaire à (2.1.9) en termes de la matrice $a := \sigma\sigma^\top$ à la place de σ, en imposant les hypothèses de régularité sur a plutôt que sur σ. Ceci peut être heuristiquement justifié par les deux observations suivantes. Comme nous l'avons déjà constaté sur (2.1.7), la caractérisation dépend de σ uniquement via $a = \sigma\sigma^\top$ et une intégration par parties serait toujours possible sur le terme $\mathrm{Tr}(v\sigma(x)\sigma(x)^\top)$ si seulement a est supposé dérivable. En langage probabiliste, le problème de trouver une solution à (2.1.4) à valeurs dans $\mathcal{D}$ peut être exclusivement reformulé via la loi du processus X, qui est entièrement déterminée par a et b. Deuxièmement, imposer les hypothèses de régularité sur a plutôt que sur σ étend considérablement la plage de validité de (2.1.9) à une classe plus large de modèles utiles en pratique, comme les diffusions affines et polynomiales (voir [48] et [35]). Pour ces processus, la fonction $a = \sigma\sigma^\top$ est régulière (avec une dépendance affine ou polynomiale en x) mais σ peut ne pas être dérivable, notamment sur le du bord domaine, comme l'exemple du processus racine carré l'a déjà illustré.

2.1.5 Nos contributions

2.1.5.1 Au delà du drift de Stratonovich

Au stade actuel, le problème consiste à reformuler la caractérisation (2.1.9) en termes de $a = \sigma\sigma^\top$. Il est immédiat de constater que la condition de volatilité tangentielle $\sigma(x)^\top u = 0$ est équivalente à la condition

$$a(x)u = 0. \tag{2.1.10}$$

Deviner le terme correctif lorsque σ n'est pas dérivable s'avère plus délicat. L'exemple du processus racine carrée a déjà révélé qu'une limite naïve du terme correctif de Stratonovich n'aboutit pas aux bonnes conditions. Par conséquent, une nouvelle expression quantifiant la contribution du terme de diffusion dans (2.1.5) doit être trouvée. L'idée naturelle serait de prendre des approximations plus appropriés de certains termes correctif de Stratonovich. Nous précisons ceci dans ce qui suit.

Supposons que la fonction a est dérivable et fixons $(e_1, \ldots, e_d)$ la base canonique de $\mathbb{R}^d$. Soit $\epsilon > 0$ et considérons la régularisation suivante qui préserve les conditions de volatilité tangentielle

$$\sigma_\epsilon : x \to a(x)(a(x) + \epsilon I_d)^{-\frac{1}{2}}.$$

Il est clair que σ_ϵ est dérivable au sens de Fréchet avec une dérivée donnée par la règle du produit[3]

$$D\sigma_\epsilon(x)h = Da(x)h(a(x)+\epsilon I_d)^{-\frac{1}{2}} + a(x)D((a+\epsilon I_d)^{-\frac{1}{2}})(x)h, \quad h \in \mathbb{R}^d.$$

En particulier, en évaluant au point $h = e_j$ pour $j = 1, \dots, d$, le terme de correction de Stratonovich $\langle u, \sum_{j=1}^d D\sigma_\epsilon^j(x)\sigma_\epsilon^j(x)\rangle$ est égal à

$$\langle u, \sum_{j=1}^d Da^j(x)(a(x)+\epsilon I_d)^{-\frac{1}{2}}\sigma_\epsilon^j(x)\rangle + \langle u, a(x)\sum_{j=1}^d D((a+\epsilon I_d)^{-\frac{1}{2}})(x)e_j\sigma_\epsilon^j(x)\rangle.$$

En vertu de la condition tangentielle (2.1.10), le deuxième terme disparaît de telle sorte à ce que

$$\langle u, \sum_{j=1}^d D\sigma_\epsilon^j(x)\sigma_\epsilon^j(x)\rangle = \langle u, \sum_{j=1}^d Da^j(x)(a(x)+\epsilon I_d)^{-\frac{1}{2}}a(x)(a(x)+\epsilon I_d)^{-\frac{1}{2}}e_j\rangle \tag{2.1.11}$$

Puisque $a(x) \in \mathbb{S}_+^d$, sa décomposition spectrale est donnée par

$$a(x) = Q(x)\mathrm{diag}(\lambda_1, \dots, \lambda_r, 0, \dots, 0)Q(x)^\top,$$

où $r \leq d$ dénote le rang de $a(x)$, $\lambda_1 \geq \cdots \geq \lambda_r > 0$ et $Q(x)$ est une matrice orthogonale. Il découle que

$$\begin{aligned}(a(x)+\epsilon I_d)^{-\frac{1}{2}}a(x)(a(x)+\epsilon I_d)^{-\frac{1}{2}} &= Q(x)\,\mathrm{diag}\left[\left(\frac{\lambda_1}{\lambda_1+\epsilon}, \dots, \frac{\lambda_r}{\lambda_r+\epsilon}, 0, \dots, 0\right)\right]Q(x)^\top \\ &\xrightarrow[\epsilon\to 0]{} Q(x)\,\mathrm{diag}\left[(\underbrace{1, \dots, 1}_{r \text{ times}}, 0, \dots, 0)\right]Q(x)^\top.\end{aligned}$$

La dernière expression n'est rien d'autre que la matrice de projection sur l'image de $a(x)$ qui peut être ré-exprimée sous la forme $a(x)a(x)^+$ où $a(x)^+$ dénote la matrice pseudo-inverse de Moore-Penrose[4] de $a(x)$. Par conséquent, en faisant tendre ϵ vers 0 dans (2.1.11), le terme correctif devient

$$\lim_{\epsilon\to 0}\langle u, \sum_{j=1}^d D\sigma_\epsilon^j(x)\sigma_\epsilon^j(x)\rangle = \langle u, \sum_{j=1}^d Da^j(x)(aa^+)^j(x)\rangle. \tag{2.1.12}$$

Heuristiquement, en combinant (2.1.10) et (2.1.12) nous arrivons à la caractérisation suivante.

[3] nous utilisons la même notation pour la dérivée de Fréchet et la matrice jacobienne en identifiant $Da(x)e_j = Da^j(x)$.

[4] La matrice pseudo-inverse de Moore-Penrose d'une $m \times n$-matrice A est l'unique $n \times m$-matrice A^+ satisfaisant: $AA^+A = A$, $A^+AA^+ = A^+$, tels que AA^+ et A^+A sont hermitiennes.

Résultat principal 1 - Caractérisation de l'invariance

Sous des hypothèses appropriées telles que la dérivabilité de $a := \sigma\sigma^\top$. Il existe une solution à (2.1.4) à valeurs dans $\mathcal{D}$ pour tout point de départ $X_0 \in \mathcal{D}$ si et seulement si

$$a(x)u = 0 \quad \text{et} \quad \langle u, b(x) - \frac{1}{2}\sum_{j=1}^{d} Da^j(x)(aa^+)^j(x)\rangle \leq 0 \tag{2.1.13}$$

pour tout $x \in \mathcal{D}$ et $u \in \mathcal{N}_{\mathcal{D}}^1(x)$.

2.1.5.2 Caractérisation du premier order sous de faibles hypothèses de régularité

Nous prouvons dans le Chapitre 3 que la caractérisation (2.1.13) est valide. Ceci constitue le premier résultat principal de la thèse. Les applications successives du Lemme d'Itô comme dans (2.1.8) ne peuvent pas être simplement répliquées. Le problème réside dans le fait que $\sigma(X)$ n'est pas une semimartingale en général. Nous avons donc été amenés à développer de nouvelles idées avant de revisiter la stratégie précédente.

À notre connaissance, contrairement au terme de Stratonovich, le terme correctif entrant dans la caractérisation (2.1.13) apparaît pour la première fois dans la littérature et quantifie la force tagentielle F de (2.1.5) comme suit

$$F(a) = \frac{1}{2}\sum_{j=1}^{d} Da^j(x)(aa^+)^j(x).$$

Le terme de projection aa^+ est nécessaire pour être en accord avec les intuitions géométriques: b ne doit compenser que les dérivées directionnelles tangentielles de a afin de garder la particule à l'intérieur du domaine.

De plus, si la matrice de volatilité σ est dérivable, il découle de (2.1.12) que

$$\langle u, \sum_{j=1}^{d} D\sigma^j(x)\sigma^j(x)\rangle = \lim_{\epsilon\to 0}\langle u, \sum_{j=1}^{d} D\sigma_\epsilon^j(x)\sigma_\epsilon^j(x)\rangle = \langle u, \sum_{j=1}^{d} Da^j(x)(aa^+)^j(x)\rangle,$$

prouvant que (2.1.13) se réduit à la caractérisation (2.1.9). Par conséquent, notre résultat généralise les travaux [38, 47] sous des hypothèses de régularité plus faibles, principalement sur la matrice de volatilité σ.

Enfin, pour revenir à l'exemple du processus racine carrée, $a(x) = x$ de sorte que $aa^+(x) = \mathbb{1}_{\{x>0\}}$. Pour $u = -1$ les conditions (2.1.13) au bord en 0 sont données par

$$a(0) = 0 \quad \text{et} \quad b(0) \geq 0,$$

ce qui aboutit à la bonne caractérisation.

Le tableau suivant résume nos résultats jusqu'à présent.

	Existant 2nd ordre	Existant 1er ordre	Notre résultat
Conditions	(2.1.7)	(2.1.9)	(2.1.13)
Cône normal	Second ordre	Premier ordre	Premier ordre
Hypothèses principales sur σ	Continuité	Dérivabilité	Dérivabilité de σ^2
Diffusions affines/polynomiales	✓	✗	✓
Intuitions géométriques selon (2.1.5)	✗	✓	✓

TABLE 2.1: Résumé des différentes caractérisations.

2.1.5.3 Extension aux sauts

Nous introduisons une source supplémentaire d'aléas dans (2.1.4) afin d'inclure des sauts dans le mouvement de la particule comme suit

$$dX_t = b(X_t)dt + \sigma(X_t)dW_t + \int_{\mathbb{R}^d} \rho(X_{t-}, z)\,(\mu(dt, dz) - F(dz)dt)\,, \quad X_0 = x, \tag{2.1.14}$$

où μ est une mesure aléatoire de Poisson sur $\mathbb{R}_+ \times \mathbb{R}^d$ avec comme compensateur $dt \otimes F(dz)$ contrôlant la fréquence d'arrivées des sauts, et ρ est une fonction qui détermine la taille des sauts et qui satisfait des conditions de continuité appropriées.

Nous pouvons montrer que les sources de bruits aléatoires sont indépendantes. De ce fait, nous nous concentrons uniquement sur les intuitions pour la partie discontinue afin que la particule reste dans $\mathcal{D}$. Si la particule X est (strictement) à l'intérieur du domaine juste avant un saut à l'instant t, alors il est évident de voir que la particule ne doit pas sauter à l'extérieur du domaine, c'est-à-dire

$$X_{t-} + (\textit{taille du saut}) \;\; \textit{appartient à } \mathcal{D}.$$

La situation s'avère plus délicate au bord du domaine. À l'échelle microscopique, la particule peut faire de grands ou petits sauts. Si l'on observe le mouvement de la particule à une échelle macroscopique, les petits sauts fluctuant très rapidement deviennent invisibles à l'œil nu et on ne voit plus qu'une trajectoire continue. Nous nous attendons donc à ce que ces petits sauts aient le même comportement qu'un processus aléatoire continu à variation infinie, tout comme le mouvement Brownien. En nous inspirant de (2.1.5), l'intuition de la volatilité tangentielle se traduit pour les sauts comme suit:

les petits sauts fluctuant rapidement doivent être tangents au bord.

De plus, comme dans le cas continu,

b devrait compenser ces mouvements tangentiels pour que la particule reste dans le domaine.

Encore une fois, les intuitions précédentes combinées avec (2.1.13) peuvent être interprétées mathématiquement comme suit.

Résultat principal 2 - Extension aux sauts

Sous des hypothèses appropriées telles que la dérivabilité de $a := \sigma\sigma^\top$. Il existe une solution à (2.1.14) à valeurs dans $\mathcal{D}$ pour tout point de départ $X_0 \in \mathcal{D}$ si et seulement si

$$x + \rho(x,z) \in \mathcal{D}, \text{ pour } F\text{-presque tout } z,$$
$$\int |\langle u, \rho(x,z)\rangle| F(dz) < \infty, \qquad a(x)u = 0,$$
$$\langle u, b(x) - \int \rho(x,z)F(dz) - \frac{1}{2}\sum_{j=1}^{d} Da^j(x)(aa^+)^j(x)\rangle \leq 0,$$

pour tout $x \in \mathcal{D}$ et $u \in \mathcal{N}^1_{\mathcal{D}}(x)$.

Cela conduit au second résultat principal de la thèse que nous prouvons dans le Chapitre 4. Nous dérivons également une caractérisation équivalente pour le cadre semimartingale en termes du triplet caractéristique.

2.2 Equations de Volterra stochastiques

La seconde partie de la thèse est consacrée à l'étude des équations intégrales de Volterra stochastiques d-dimensionnelles de la forme

$$X_t = g_0(t) + \int_0^t K(t-s)b(X_s)ds + \int_0^t K(t-s)\sigma(X_s)dW_s, \tag{2.2.1}$$

où W est un mouvement Brownien multidimensionnel, et le noyau de convolution K, la fonction g_0 et les coefficients b et σ satisfont des conditions de régularité et d'intégrabilité appropriées.

En fixant $K \equiv 1$ et $g_0 \equiv X_0$ pour une condition initiale constante X_0, nous récupèrons (2.1.4) sous forme intégrale. Par conséquent, les équations de Volterra stochastiques étendent les équations différentielles stochastiques traditionnelles et permettent une plus grande flexibilité pour la modélisation. Cependant, elles ne rentrent pas en général dans les cadres semimartingale et Markovien, comme l'illustre l'exemple du mouvement Brownien fractionnaire de Riemann-Liouville $X_t = \int_0^t K_H(t-s)$, où K_H est le noyau fractionnaire défini par

$$K_H : t \to t^{H-1/2}, \quad H \in (0, 1/2). \tag{2.2.2}$$

Néanmoins, nous développons dans la suite plusieurs techniques pour traiter l'existence et l'unicité pour différents espaces d'état et pour clarifier le lien avec les équations différentielles stochastiques standards. Nos arguments évitent l'intégration stochastique par rapport à des processus non-semimartingales, s'appuyant plutôt sur des outils issus de la théorie des équations de convolution déterministes en dimension finie.

Avant de passer à un exposé plus détaillé, nous motivons d'abord l'étude des équations de Volterra stochastiques.

2.2.1 Motivations

Notre motivation pour étudier ces équations de convolution est double. Les équations de Volterra stochastiques apparaissent comme limites de processus de branchement, convenablement normalisés, dans le domaine de la génétique des populations et des processus auto-excitants en mathématiques financières.

Interaction entre biologie et chimie

Un chimiste cherche à modéliser une interaction de type réaction-diffusion entre deux substances: un *réactif* et un *catalyseur*. La première substance, appelée le *réactif* diffuse dans l'espace suivant un mouvement aléatoire. La réaction chimique n'aura lieu localement qu'en présence d'une seconde substance, le *catalyseur*, répandu dans l'espace, proportionnellement à la concentration du *catalyseur* au point de contact.

Afin de décrire l'image microscopique, son collègue biologiste suggère le modèle simple suivant, inspiré de la génétique des populations. Le *reactif* peut être considéré comme un système de n particules unidimensionionelles se déplaçant dans l'espace suivant un mouvement Brownien standard, de façon indépendante. La région occupée par le *catalyseur* à un instant donnée t est définie par le support d'une mesure déterministe $\rho_t(dx)$. Chaque fois qu'une particule entre dans une région *catalyste*, et après avoir passé un temps aléatoire au voisinage du *catalyseur*, elle meurt ou se divise en deux nouvelles particules, avec probabilités égales. La mesure $\rho_t(dx)$ détermine le taux local de branchement en temps et en espace, en fonction de l'emplacement et de la concentration du *catalyseur*. Deux exemples typiques sont $\rho_t(dx) \equiv \bar{\rho}$, auquel cas le branchement se produit dans tout l'espace avec un taux constant $\bar{\rho}$, et $\rho_t(dx) = \delta_0(dx)$ correspondant à des branchements se produisant à un taux infini uniquement lorsque la particule atteint un unique point extrêmement concentré en *catalyseur* situé en 0. En cas de branchement, les deux particules filles évoluent de façon indépendante avec le même mouvement spatial et le même mécanisme de branchement que la particule mère.

Le *reactant* peut être vu comme un processus à valeurs dans l'espace des mesures $(\bar{Y}_t^n(dx))_{t\geq 0}$ défini par

$$\bar{Y}_t^n(B) = \frac{\text{nombre de particles dans } B \text{ à l'instant } t}{n}, \quad \text{pour tout Borélien } B.$$

En envoyant le nombre de particules vers l'infini, nous pouvons établir la convergence de la suite $(\bar{Y}^n)_{n\geq 1}$ vers un processus *réactif* limite $\bar{Y}$ à valeurs dans l'espace des mesures: le *super-mouvement Brownien avec catalyse*, solution d'un problème martingale en dimension infinie. Par ailleurs, l'unicité faible découle d'un argument de dualité sur la fonctionnelle de Laplace qui se révèle être exponentiellement affine

$$\mathbb{E}\left[\exp(\langle u, \bar{Y}_t\rangle)\right] = \exp(\langle \Psi_t, \bar{Y}_0\rangle), \quad u \leq 0, \tag{2.2.3}$$

où $\langle f, \nu\rangle = \int f(x)\nu(dx)$ et Ψ est la solution mild de l'équation aux dérivées partielles de type Riccati suivante

$$\frac{\partial \Psi_t(x)}{\partial t} = \frac{1}{2}\Delta\Psi_t(x) + \frac{1}{2}\Psi_t^2(x)\frac{\rho_t(dx)}{dx}, \quad \Psi_0(x) = u(x), \tag{2.2.4}$$

avec $\Delta = \partial^2/\partial x^2$ le Laplacien du second ordre.

De plus, en présence d'un *catalyseur* $\rho = (\rho_t(dx))_{t\geq 0}$ n'ayant aucun atome, le processus $\bar{Y}$ admet une densité $\bar{Y}_t(dx) = \int Y_t(x)dx$ solution de l'équation aux dérivées partielles stochastique suivante

$$\frac{\partial Y_t(x)}{\partial t} = \frac{1}{2}\Delta Y_t(x) + \sqrt{Y_t(x)}\dot{W}^{\rho}(t,x), \tag{2.2.5}$$

où $\dot{W}^{\rho}(t,x)$ est un bruit spatio-temporel dont la structure de covariance est déterminée par ρ; nous renvoyons vers Zähle [116] pour plus de détails. En notant

$$p_t(x) = (2\pi t)^{-1/2}\exp(-(xy)^2/(2t))$$

le noyau de la chaleur, les solutions à (2.2.5) démarrant d'une courbe initiale Y_0 sont considérées au sens mild, c'est à dire

$$Y_t(x) = \int_{\mathbb{R}} p_t(x-y)Y_0(y)dy + \int_0^t\int_{\mathbb{R}} p_{t-s}(x-y)\sqrt{Y_s(y)}W^{\rho}(ds,dy). \tag{2.2.6}$$

L'équation précédente n'est valide que si ρ n'a pas d'atomes. Nous pouvons toujours remplacer de manière heuristique $\rho_t(dx) = \delta_0(dx)$ dans (2.2.6) pour le cas extrême d'un seul point *catalyseur* en 0. Dans ce cas, le bruit spatio-temporel se réduit à un mouvement Brownien standard W, de telle sorte à ce que l'évaluation au point $x = 0$ donne

$$Y_t(0) = g_0(t) + \int_0^t \frac{(t-s)^{-1/2}}{\sqrt{2\pi}}\sqrt{Y_s(0)}dW_s, \tag{2.2.7}$$

où $g_0(t) = \int_{\mathbb{R}} p_t(y)Y_0(y)dy$. Par conséquent, $(Y_t(0))_{t\geq 0}$ résout une équation de Volterra stochastique de la forme (2.2.1). Inutile de dire que l'on n'est pas autorisé à introduire la mesure de Dirac dans (2.2.6). En effet, en présence d'un seul point *catalyseur*, le *super-mouvement Brownien avec catalyse* n'admet pas de densité au point *catalyseur*, comme le montrent Dawson et Fleischmann [41], et (2.2.5) n'est plus valide. Cela peut être vu directement sur (2.2.7) puisque le noyau $K : t \to t^{-1/2}$ n'est pas localement de carré intégrable et l'intégrale stochastique n'est pas bien définie au sens de la théorie L^2 d'Itô. Néanmoins, l'intégrabilité L^2 peut être récupérée de manière informelle en perturbant la puissance du noyau avec un petit $H > 0$ conduisant au noyau fractionnaire K_H défini dans (2.2.2). Pour le noyau K_H, (2.2.7) a désormais du sens. Au niveau microspcopique du processus de branchement, cela se traduit par la substitution du flux de la chaleur Δ par un mouvement spatial approprié Δ_H. Ces heuristiques sont rendues rigoureuses par Mytnik et Salisbury [96].

Par ailleurs, la transformée de Laplace de $(Y_t(0))_{t\geq 0}$ se déduit de (2.2.3). En effet, en raison de la propriété de régularisation du semi-groupe généré par Δ_H, l'équation aux dérivées partielles de Riccati peut être démarrée d'une condition initiale dans l'espace des mesures. Comme $\langle u, \bar{Y}_t\rangle = \langle u, Y_t\rangle$, et en définissant $u(x) = u\delta_0(dx)$ pour certaines constantes non positives u et $Y_0(x) \equiv Y_0 \geq 0$, des manipulations directes de (2.2.3)-(2.2.4) avec $\rho_t(dx) = \delta_0(dx)$ mènent à

$$\mathbb{E}\left[\exp(uY_t(0))\right] = \exp(\chi(t)Y_0), \tag{2.2.8}$$

où $\chi(t) = u + \frac{1}{2}\int_0^t \Psi_s^2(0)ds$ et $(\Psi_t(0))_{t\geq 0}$ résout l'équation Riccati-Volterra suivante

$$\psi(t) = uK_H(t) + \frac{1}{2}\int_0^t K_H(t-s)\psi^2(s)ds. \tag{2.2.9}$$

Le point de vue d'un trader: du tick-par-tick vers la volatilité rugueuse

Dans le domaine de la finance quantitative, en raison du développement rapide du trading algorithmique haute fréquence, la théorie de la microstructure des marchés est en plein essor. Comme son nom l'indique, cette théorie traite de questions telles que la formation des prix, les coûts de transaction et la liquidité, par une étude approfondie du marché à une échelle microscopique.

Imitant le biologiste de la section précédente, le trader peut procéder en deux étapes. Tout d'abord, concevoir un modèle microscopique codant les principaux faits stylisés de la microstructure moderne du marché, avant de regarder la limite macroscopique émanant du modèle microscopique, suite à une normalisation appropriée. Une telle approche a été adoptée dans la thèse de doctorat de Jaisson [78], en s'appuyant sur les processus de Hawkes pour modéliser le prix d'un actif au niveau microscopique. La thèse a ensuite été suivie par une série de travaux qui ont abouti à la construction élégante du *modèle d'Heston rugueux* dans [51]. Les processus de Hawkes sont une généralisation des processus de Poisson où l'intensité des sauts dépend des réalisations passées du processus. Cette structure de dépendance est spécifiée par un noyau. Pour des noyaux appropriés, des effets auto-excitants peuvent être générés, permettant une modélisation jointe des faits stylisés suivants:

(i) clustering du flux d'ordres: nous observons des périodes avec un nombre élevé de transactions suivies de périodes plus creuses,

(ii) découpage des ordres: les ordres importants ne sont pas exécutés en une fois, ils sont découpés dans le temps en plusieurs ordres de petite taille par des algorithmes de trading afin d'optimiser les coûts de transaction,

(iii) haut degré d'endogénéité: la plupart des ordres manquent de réelle motivation économique, ils ne sont envoyés que par des algorithmes en réaction à d'autres ordres.

Après avoir incorporé ces observations dans l'intensité du processus de Hawkes, une normalisation appropriée entraîne la convergence vers le modèle suivant pour le cours de l'action S et sa variance stochastique V

$$dS_t = S_t\sqrt{V_t}dB_t, \quad S_0 = 1, \tag{2.2.10}$$

$$V_t = V_0 + \int_0^t K_H(t-s)\eta\sqrt{V_s}dW_s, \tag{2.2.11}$$

où K_H est le noyau fractionnaire défini dans (2.2.2) avec $H \in (0, 1/2)$, $(V_0, \eta) \in \mathbb{R}_+^2$ et $B = \rho W + \sqrt{1-\rho^2}W^\perp$ avec $\rho \in [-1, 1]$ et $(W, W^\perp)$ un mouvement Brownien bidimensionnel. Le processus bidimensionnel $X = (S, V)$ appartient à la classe des équations de Volterra stochastiques de la forme (2.2.1) avec le noyau diagonal $K = \text{diag}\,[(1, K_H)]$.

Le modèle macroscopique (2.2.10) - (2.2.11) a été baptisé *modèle d'Heston rugueux*. L'appellation est justifiée comme suit. D'une part, les trajectoirs de V sont localement Höldériens pour tout ordre strictement inférieur à H, et donc moins régulier que le mouvement Brownien standard (qui correspond au cas $H = 1/2$). Par conséquent, les modèles impliquant le noyau fractionnaire ont été surnommé *modèles à volatilité rugueuse* (*rough volatility* en anglais).[5]

[5] Des études empiriques sur différentes séries temporelles de la volatilité réalisée des actifs dans [66, 15] ont révélé que la dynamique de la volatilité suit celui d'un mouvement brownien fractionnaire avec un paramètre de Hurst H de l'ordre de 0.1, ce qui est incompatible avec les modèles semimartingales standard.

D'autre part, pour $H = 1/2$, (2.2.10)-(2.2.11) se réduit au *modèle d'Heston standard* [72], qui est l'un des modèles les plus populaires grâce à sa formule fermée pour la fonction caractéristique du log-prix, permettant un pricing et une calibration rapides par des techniques d'inversion de Fourier. De façon assez remarquable, une formule analogue est disponible pour le *modèle d'Heston rugueux*. En effet, en prenant la limite de la suite des transformées de Fourier du modèle microscopique de Hawkes, El Euch et Rosenbaum [50] dérivent la transformation affine

$$\mathbb{E}\left[\exp(v \log S_t)\right] = \exp(\chi(t)V_0), \quad v \in i\mathbb{R}, \tag{2.2.12}$$

en fonction de l'équation de Riccati-Volterra

$$\psi(t) = \int_0^t K_H(t-s)F(v,\psi(s))ds, \tag{2.2.13}$$

où

$$F(v,u) = \frac{1}{2}(v^2 - v) + \rho\nu u + \frac{u2}{2} \quad \text{et} \quad \chi(t) = \int_0^t F(v,\Psi(s))ds.$$

Clairement, les équations (2.2.12)-(2.2.13) partagent certaines similitudes avec (2.2.8) - (2.2.9). En combinant les deux exemples, on s'attendrait à obtenir une expression similaire pour le processus joint $(\log S, V)$.

Notre approche: le point de vue macroscopique

Il est courant en théorie des probabilités d'étudier divers phénomènes à travers l'objet macroscopique limite associé afin d'obtenir de précieuses informations sur le comportement du modèle microscopique, qui lui est plus réaliste. Nous adoptons cette perspective dans cette thèse en prenant comme point de départ les équations de Volterra stochastiques dans $\mathbb{R}^d$ de la forme (2.2.1). Ce faisant, nous évitons l'analyse en dimensions infinie, comme celle utilisée pour étudier les super-processus. Nous évitons également les limites normalisés de de processus de Hawkes, ce qui permet un traitement plus générique des problèmes suivants:

- Qu'en est-il des noyaux plus généraux que le noyau fractionnaire K_H?
- Peut-on augmenter la dimension?
- Qu'en est-il des dynamiques arbitraires, pas nécessairement du type racine carrée?
- Comment peut-on clarifier la structure *affine* sous-jacente?

Ces questions font principalement l'objet du Chapitre 5 où les problèmes d'existence générale de (2.2.1) sont abordés pour différents espaces d'état. L'unicité est ensuite établie pour une classe spécifique, à savoir les *processus affines de Volterra*, en caractérisant complètement la transformé de Fourier-Laplace de la solution en termes d'une équation déterministe intégrale de type Riccati-Volterra. Par la suite, nous établissons la correspondance avec des équations aux dérivées partielles stochastiques. Cela conduit à la caractérisation complète de la structure markovienne des équations stochastiques de Volterra et ouvre la porte à des schémas d'approximations numériques. Ces résultats sont rassemblés dans les Chapitres 6, 7 et 8.

2.2.2 Existence

Nous fournissons de nouveaux résultats d'existence pour (2.2.1) pour plusieurs espaces d'état sous des conditions faibles sur le noyau et les coefficients.

Solutions non-contraintes: d'Itô à Skorokhod

Le problème d'existence de solutions non-contraintes, c'est-à-dire à valeurs dans $\mathbb{R}^d$, est traité en revisitant les stratégies classiques pour les équations différentielles stochastiques (2.1.4). Historiquement, le premier résultat d'existence et d'unicité remonte à Kiyosi Itô [75] pour des coefficients b et σ Lipschitz. Plus tard, Anatoliy Volodymyrovych Skorokhod [104] a pu établir l'existence de solutions sous des conditions de croissance linéaire et de continuité pour b et σ.

On a découvert plus tard que les deux notions de solutions n'étaient pas équivalentes. Étant donné un espace de probabilité filtré $(\Omega, \mathcal{F}, \mathbb{F} := (\mathcal{F}_t)_{t\geq 0}, \mathbb{P})$ avec un mouvement Brownien en entré, Itô construit une solution *trajectorielle* X adaptée à la filtration générée par le mouvement Brownien, ce qui correspond au concept de *solution forte*. Alors que dans la preuve de Skorokhod, on construit un espace de probabilité filtré $(\Omega, \mathcal{F}, \mathbb{F} := (\mathcal{F}_t)_{t\geq 0}, \mathbb{P})$ avec un mouvement Brownien W et un processus X tel que (2.1.4) est vérifiée presque sûrement, ce qui correspond au concept de *solution faible*. Nous remarquons qu'une *solution faible* X n'est pas nécessairement adaptée à la filtration Brownienne. Il est clair que toute *solution forte* est une *solution faible*, mais l'inverse n'est pas vrai, comme l'illustre le célèbre exemple d'Itô-Tanaka

$$dX_t = \text{sign}(X_t)dW_t.$$

Nous prouvons des résultats d'existence similaires pour l'équation de Volterra stochastique (2.2.1).

Résultat principal 3 - Existence de solutions non-contraintes

Sous des hypothèses faibles sur g_0 et $K \in L^2_{\text{loc}}(\mathbb{R}_+, \mathbb{R}^d)$:

(i) Si b et σ sont Lipschitz, alors (2.2.1) admet une unique solution forte continue X.

(ii) Si b et σ sont continus avec des conditions de croissance, alors (2.2.1) admet une solution faible continue x.

Solutions contraintes: retour au problème d'invariance

Comme pour les équations différentielles stochastiques, la question d'existence de solutions contraintes à rester dans un domaine se pose. Cependant, le problème d'invariance ou de viabilité se révèle plus délicat dans le cadre non-markovien. Nous fournissons des conditions suffisantes pour l'existence de solutions aux équations de Volterra stochastiques (2.2.1) à valeurs dans $\mathbb{R}^d_+$.

Sous des conditions de continuité et de croissance pour (b, σ), le résultat précédent garantit l'existence d'une solution faible non-contrainte X à l'équation modifiée suivante

$$X_t = g_0(t) + \int_0^t K(t-s)b(X_s^+)ds + \int_0^t K(t-s)\sigma(X_s^+)dW_s,$$

pour une courbe d'entrée g_0 et un noyau K appropriés, où $x^+ = \max(0, x)$. Clairement, il faudrait imposer des hypothèses supplémentaires sur g_0 pour assurer la positivité de X et pouvoir supprimer la partie positive dans l'équation précédente, afin d'obtenir que X est solution de (2.2.1). En effet, puisque $X_0 = g_0(0)$, il est clair que $g_0(0)$ devrait être dans $\mathbb{R}^d_+$.

Le problème peut désormais être formulé comme suit:

Peut-on trouver un ensemble $\mathcal{G}_K$ de courbes d'entrée initiales g_0 telles que X reste dans $\mathbb{R}^d_+$?

Afin de mieux comprendre le problème, revenons au processus racine carré standard en dimension un

$$X_t = g_0(t) - \lambda \int_0^t X_s^+ ds + \int_0^t \sqrt{X_s^+} dW_s.$$

En raisonnant à la physicienne, le comportement de X juste après un instant fixé t peut être approximé par

$$X_{t+h} \approx X_t + g_0(t+h) - g_0(t) - h\lambda X_t^+ + \sqrt{X_t^+}(W_{t+h} - W_t), \quad \text{pour } h > 0 \text{ petit.}$$

Comme $g_0(0) \geq 0$, si X touche zéro pour la première fois à l'instant $t \geq 0$, alors

$$X_{t+h} \approx g_0(t+h) - g_0(t).$$

D'où, X_{t+h} reste positif pour tout $h > 0$ si l'ensemble des courbes d'entrée initiales est donné par

$$\mathcal{G}_1 = \{g_0 \text{ croissante telle que } g_0(0) \geq 0\}.$$

Pour un noyau K arbitraire, on pourrait s'attendre à une condition de "croissance" analogue qui prenne en compte les variations de K. C'est en effet le cas, nous fournissons un ensemble de *courbes d'entrée admissibles*

$$\mathcal{G}_K = \{g_0 \text{ "croissante par rapport à } K\text{" telle que } g_0(0) \geq 0\}. \tag{2.2.14}$$

tel que (2.2.1) admette une solution faible à valeurs dans $\mathbb{R}^d_+$.[6]

Résultat principal 4 - Existence de solutions constraintes dans $\mathbb{R}^d_+$

Sous de faibles hypothèses sur $K \in L^2_{\text{loc}}(\mathbb{R}_+, \mathbb{R}^d)$ telles que la positivité et la décroissance, si les conditions géométriques

$$x_i = 0 \text{ implique } b_i(x) \geq 0 \text{ et } \sigma_i(x) = 0, \quad i = 1, \dots, n$$

sont satisfaites, alors (2.2.1) admet une solution faible continue à valeurs dans $\mathbb{R}^d_+$, pour toute courbe initiale admissible $g_0 \in \mathcal{G}_K$.

2.2.3 Cadre affine: unicité et tractabilité

Ayant établi l'existence, nous nous tournons vers le problème d'unicité qui est une étape cruciale pour l'utilisation du modèle dans la pratique. L'unicité est un problème assez difficile pour des coefficients non-Lipschitz, même pour les équations différentielles stochastiques standards. En effet, lorsque $K \equiv \text{id}$ est constant et égal à la matrice d'identité d-dimensionnelle, l'unicité faible peut être obtenue dans le cas particulier où les coefficients $a := \sigma\sigma^\top$ et b sont affines:

$$\begin{aligned} a(x) &= A^0 + x_1 A^1 + \cdots + x_d A^d \\ b(x) &= b^0 + x_1 b^1 + \cdots + x_d b^d, \end{aligned} \tag{2.2.15}$$

[6] La formulation explicite de $\mathcal{G}_K$ est donnée au Chapitre 6.

pour certaines matrices symétriques d-dimensionnelles A^i et vecteurs b^i. Dans ce cas, les solutions de (2.1.4) sont appelées *diffusions affines*, nous référons à Duffie, Filipović et Schachermayer [48] pour un traitement détaillé. Plus précisément, supposons que X soit une *diffusion affine* de la forme (2.1.4) à valeurs dans un fermé $\mathcal{D} \subset \mathbb{R}^d$. Pour un vecteur de ligne d-dimensionnel approprié u, la transformée de Fourier-Laplace conditionnelle de X est exponentiellement affine

$$\mathbb{E}\left[\exp\left(uX_T\right) \;\middle|\; \mathcal{F}_t\right] = \exp\left(\phi(T-t) + \psi(T-t)X_t\right), \tag{2.2.16}$$

où les fonctions ϕ et Ψ satisfont aux équations différentielles ordinaires de Riccati suivantes:

$$\begin{aligned}\phi(t) &= \int_0^t \left(\psi(s)b_0 + \frac{1}{2}\psi(s)A^0\psi(s)^\top\right) ds \\ \psi(t) &= u + \int_0^t \left(\psi(s)B + \frac{1}{2}A(\psi(s))\right) ds,\end{aligned}$$

avec $A(u) = (uA^1u^\top, \ldots, uA^du^\top)$ et $B = (b^1 \;\cdots\; b^d)$. L'unicité faible de (2.1.4) découle alors de (2.2.16), ce qui conduit également à la tractabilité. En effet, les *diffusions affines* constituent sans doute le cadre le plus populaire pour la construction de modèles multi-facteurs faciles à gérer en finance. En particulier, les modèles de volatilité stochastique affines, tels que le modèle de Heston [72] déjà rencontré dans la motivation, sont très populaires.

En supprimant la restriction $K \equiv \mathrm{id}$ et en renvoyant aux deux motivations, on pourrait espérer une expression similaire pour la transformée de Fourier-Laplace pour les équations de Volterra stochastiques (2.2.1) avec des coefficients affines de la forme (2.2.15). Néanmoins, (2.2.16) est clairement une expression Markovienne et ne peut pas rester valide dans un cadre non-Markovien. Par conséquent, une première étape naturelle serait de trouver un Ansatz potentiel pour la transformation de Fourier-Laplace dans un environnement non-Markovien.

Dérivation d'un bon Ansatz

Nous revisitons le processus standard racine carrée

$$dX_t = -\lambda X_t dt + \eta\sqrt{X_t}dW_t,$$

qui est clairement une *diffusion affine* à valeurs dans $\mathbb{R}_+$ avec comme coefficients $a(x) = \eta^2 x$ et $b(x) = -\lambda x$. Il découle de (2.2.16) que

$$\mathbb{E}\left[\exp(uX_T)|\mathcal{F}_t\right] = \exp\left(\psi(T-t)X_t\right), \quad \Re(u) \le 0,$$

où

$$\psi' = -\lambda\psi + \frac{\eta^2}{2}\psi^2, \quad \psi(0) = u.$$

La formule de la variation des constantes au niveau du processus X et de la fonction Ψ donne

$$\mathbb{E}[X_s|\mathcal{F}_t] = e^{-\lambda(t-s)}X_t \quad \text{et} \quad \psi(t) = ue^{-\lambda t} + \frac{\eta^2}{2}\int_0^t e^{-\lambda(t-s)}\psi^2(s)ds, \quad s \ge t,$$

ce qui entraîne

$$\begin{aligned}\psi(T-t)X_t &= ue^{-\lambda(T-t)}X_t + \frac{\eta^2}{2}\int_0^{T-t}\psi^2(s)e^{-\lambda(T-t-s)}X_t ds \\ &= u\mathbb{E}[X_T|\mathcal{F}_t] + \frac{\eta^2}{2}\int_t^T \psi^2(T-s)\mathbb{E}[X_s|\mathcal{F}_t]ds,\end{aligned}$$

conduisant à l'expression alternative pour la transformée de Fourier-Laplace

$$\mathbb{E}\left[\exp\left(uX_T\right) \;\middle|\; \mathcal{F}_t\right] = \exp\left(\mathbb{E}[uX_T \mid \mathcal{F}_t] + \frac{1}{2}\int_t^T \psi^2(T-s)a(\mathbb{E}[X_s \mid \mathcal{F}_t])ds\right).$$

Par rapport à (2.2.16), l'expression précédente à l'avantage de masquer complètement la propriété Markovienne du processus. Pour cette raison, elle constitue un bon potentiel Ansatz pour le cadre Volterra.

Processus affines de Volterra

Notre résultat principal montre que, remarquablement, cette même expression est toujours valide pour la classe des *processus affines de Volterra* où la fonction Ψ résout désormais une équation intégrale déterministe de type convolution. Un *processus affine de Volterra* avec un espace d'état $\mathcal{D}$ et des coefficients (2.2.15) est défini comme une solution faible évaluée de (2.2.1) à valeurs dans $\mathcal{D}$, avec $a := \sigma\sigma^\top$ sur $\mathcal{D}$.

Résultat principal 5 - Transformation affine

Sous des hypothèses appropriées sur $K \in L^2_{\text{loc}}(\mathbb{R}_+, \mathbb{R}^d)$, fixons X un *processus affine de Volterra* à valeurs dans $\mathcal{D}$ avec des coefficients donnés par (2.2.15) et supposons que l'équation de *Riccati-Volterra correspondant*

$$\psi(t) = uK(t) + \int_0^t \left(\psi(s)B + \frac{1}{2}A(\psi(s))\right)K(t-s)\,ds,$$

admette une solution. Alors, sous une condition supplémentaire d'intégrabilité, nous avons

$$\mathbb{E}\left[\exp\left(uX_T\right) \;\middle|\; \mathcal{F}_t\right] = \exp\left(\mathbb{E}[uX_T \mid \mathcal{F}_t] + \frac{1}{2}\int_t^T \psi(T-s)a(\mathbb{E}[X_s \mid \mathcal{F}_t])\psi(T-s)^\top ds\right).$$

Encore une fois, l'unicité faible en découle. De plus, il est possible d'exprimer la transformée de Fourier-Laplace sous une forme exponentiellement-affine en la trajectoire passée $\{X_s,\, s \leq t\}$. En particulier, lorsque $g_0 \equiv X_0$ pour une constante $X_0 \in \mathbb{R}^d$, et pour $t = 0$, la transformée de Fourier-Laplace non-conditionnelle est donnée par

$$\mathbb{E}[\exp(uX_T)] = \exp\left(\phi(T) + \chi(T)X_0\right),$$

où

$$\phi = \int_0^\cdot (\psi b^0 + \frac{1}{2}\psi A^0 \psi^\top)(s)ds \quad \text{and} \quad \chi = u + \int_0^\cdot (\psi B + \frac{1}{2}A(\psi))(s)ds.$$

Ces dernières expressions concordent avec les formulations (2.2.8) et (2.2.12) des motivations. Enfin, la caractérisation de Fourier-Laplace est une propriété clé des *processus affines de Volterra*, rendant ces équations intéressantes en pratique avec un compromis attrayant entre flexibilité et tractabilité.

2.2.4 Markovianité: correspondance avec les EDP stochastiques

La motivation provenant de la génétique des populations met en évidence la correspondance entre les équations stochastiques de Volterra et les équations aux dérivées partielles stochastiques dégénérées. Plus précisément, résoudre le problème d'existence et d'unicité pour l'une de ces classes conduit à résoudre le problème pour l'autre classe et vice-versa. En partant de certaines équations aux dérivées partielles stochastiques, on peut récupérer une équation de Volterra stochastique par une procédure de projection. Inversement, toute équation stochastique de Volterra peut être élevée vers une équation en dimension infinie. Ayant déjà abordé le problème d'existence et d'unicité des équations de Volterra stochastiques, il semble naturel d'adopter les équations de Volterra comme point de départ. Cela conduit à une caractérisation de la structure Markovienne des équations stochastiques de Volterra. Nous fournissons deux représentations Markoviennes en termes d'objets en dimension infinie et spécifions leur espace d'états.

Par soucis de clareté, nous nous limitons au cas des équations de Volterra stochastiques unidimensionnelles à valeurs dans $\mathbb{R}_+$. Nous supposons que b et σ satisfont les conditions au bord

$$b(0) \geq 0 \quad \text{et} \quad \sigma(0) = 0,$$

et que g_0 appartient à l'ensemble de *courbes d'entrée admissibles* $\mathcal{G}_K$, donné par (2.2.14) de sorte à ce que notre résultat principal 4 assure l'existence d'une solution faible non négative X à (2.2.1).

Première représentation: processus forward

Nous commençons par l'observation suivante: conditionnellement à $\mathcal{F}_t$, le processus shifté $X^t := (X_{t+x})_{x\geq 0}$ résout la même équation de Volterra stochastique (2.2.1) à condition que g_0 soit remplacée par le processus forward ajusté suivant

$$g_t(x) = \mathbb{E}\left[X_{t+x} - \int_0^x K(x-s)b(X_{t+s})ds \;\middle|\; \mathcal{F}_t\right], \quad x \geq 0.$$

Plus précisement, le processus X^t résout l'équation

$$X_x^t = g_t(x) + \int_0^x K(x-s)b(X_s^t)ds + \int_0^x K(x-s)\sigma(X_s^t)dW_s^t, \quad x \geq 0,$$

avec $W^t := W_{t+\cdot} - W_t$. Ceci suggère que, d'une part, X est Markovien en la courbe de dimension infinie $(g_t)_{t\geq 0}$. D'autre part, comme le processus X^t est positif, nous sommes tentés d'affirmer que g_t constitue de nouveau une *courbe d'entrée admissible* appartenant à $\mathcal{G}_K$ telle que définie dans (2.2.14). Notre résultat principal confirme cette intuition en montrant que $\mathcal{G}_K$ est invariant pour la famille $(g_t)_{t\geq 0}$. En d'autres termes, si on démarre d'une *courbe d'entrée admissible* $g_0 \in \mathcal{G}_K$, alors g_t appartient à $\mathcal{G}_K$, pour tout $t \geq 0$. Cela nous permet de caractériser la structure Markovienne de x en fonction du processus forward ajusté $(g_t)_{t\geq 0}$. De plus, $(g_t)_{t\geq 0}$ peut être vu comme la solution mild à valeurs dans $\mathcal{G}_K$ de l'équation aux dérivées partielles stochastique de type Heath-Jarrow-Morton

$$dg_t(x) = \left(\frac{d}{dx}g_t(x) + K(x)b(g_t(0))\right)dt + K(x)\sigma(g_t(0))dW_t, \quad g_0 \in \mathcal{G}_K,$$

et possède une fonctionnelle de Fourier exponentiellement affine lorsque b et σ^2 sont affines.

Deuxième représentation: cas complètement monotone

Nous pouvons en dire plus lorsque K est complètement monotone sur $(0,\infty)$, i.e. K est infiniment dérivable sur $(0,\infty)$ tel que $(-1)^n K^{(n)} \geq 0$ pour tout $n \geq 0$. D'après le théorème de Bernstein, cela revient à dire que K est la transformée de Laplace d'une mesure non négative μ

$$K(t) = \int_0^\infty e^{-xt}\mu(dx), \quad t > 0.$$

Par exemple, le noyau fractionnaire K_H donné dans (2.2.2) est complètement monotone et sa mesure associée est donnée par $\mu_H(dx) = \frac{x^{-1/2-H}}{\Gamma(1/2-H)}dx$. En exploitant cette propriété et en partant d'une solution à (2.2.1) avec $g_0 \equiv 0$, une interversion formelle de l'ordre d'intégration conduit à

$$\begin{aligned} X_t &= \int_0^t K(t-s)\left(b(X_s)ds + \sigma(X_s)dW_s\right) \\ &= \int_0^\infty \int_0^t e^{-x(t-s)}\left(b(X_s)ds + \sigma(X_s)dW_s\right)\mu(dx) \\ &= \int_0^\infty U_t(x)\mu(dx) \end{aligned} \tag{2.2.17}$$

où $U_t(x) := \int_0^t e^{-x(t-s)}\left(b(X_s)ds + \sigma(X_s)dW_s\right)$. En particulier, on reconnaît la formulation mild de l'équation suivante

$$dU_t(x) = \left(-xU_t(x) + b\left(\int_0^\infty U_t(y)\mu(dy)\right)\right)dt + \sigma\left(\int_0^\infty U_t(y)\mu(dy)\right)dW_t, \quad U_0(x) = 0,$$

pour tout $x \in \mathrm{supp}(\mu)$, où $\mathrm{supp}(\mu)$ indique le support de la mesure μ. Les processus $(U(x))_{x\in(\mathrm{supp}\,\mu)}$ partagent la même dynamique, sauf qu'ils retournent à leur moyenne à des vitesses différentes. Dans le cas où le support de μ est fini, constitué de n points $(x_1,\dots,x_n)$, l'équation se réduit à un système d'équation différentielle standard n-dimensionnel de la forme (2.1.4). De plus, il découle de la représentation (2.2.17) que X est Markovien en dimension n en $(U(x_1),\dots,U(x_n))$.

Les deux représentations sont liées par la formule suivante:

$$g_t(x) = \int_0^\infty e^{-yx}U_t(y)\mu(dy), \quad t,x \geq 0.$$

Par ailleurs, dans le cas d'un *processus affine de Volterra*, à savoir $b(x) = -\lambda x$ et $\sigma(x) = \eta\sqrt{x}$, nous montrons que la transformée de Laplace de X est une fonctionnelle exponentiellement affine en U_t

$$\mathbb{E}\left[\exp(uX_T)\;\middle|\;\mathcal{F}_t\right] = \exp\left(\int_0^\infty \chi(T-t,x)U_t(x)\mu(dx)\right)$$

où χ résout l'équation aux dérivées partielles de Riccati suivante

$$\partial_t\chi(t,x) = -x\chi(t,x) + F\left(\int_0^\infty \chi(t,y)\mu(dy)\right), \quad \chi(0,x) = u, \quad x \in \mathrm{supp}(\mu),$$

avec $F(u) = -\lambda u + \frac{\eta^2}{2}u^2$. Comme précédemment, lorsque le support de μ est fini, l'équation se réduit à un système fini d'équations différentielles ordinaires de Riccati.

2.2.5 Procédure d'approximation

À ce stade, une idée naturelle serait d'approcher toute mesure positive μ à support infini par une somme pondérée finie de mesures de Dirac, tout en espérerant obtenir la convergence des équations associées. Plus précisément, on fixe un noyau complètement monotone K de telle sorte à ce que la mesure associée est à support infinie, par exemple le noyau fractionnaire. En approchant le noyau K par une suite de noyaux réguliers $(K^n)_{n\geq 1}$, on pourrait s'attendre à obtenir la convergence de la suite des équations de Volterra stochastiques correspondante

$$X_t^n = g_0^n(t) + \int_0^t K^n(t-s)b(X_s^n)ds + \int_0^t K^n(t-s)\sigma(X_s^n)dW_s, \quad n \geq 1, \tag{2.2.18}$$

vers l'équation initiale (2.2.1) avec le noyau K, pour un choix approprié de g_0^n.

Ceci est rendu rigoureux par le résultat de stabilité pour les équations de Volterra stochastiques suivant.

Résultat principal 6 - Stabilité des équations de Volterra stochastique

Fixons $T > 0$, sous de faibles hypothèses d'intégrabilité, nous supposons que $\|K^n - K\|_{L^2(0,T)} \to 0$ et $g_0^n \to g_0$ point par point sur $[0,T]$, alors la suite $(X^n)_{n\geq 1}$ de solutions à (2.2.18) est tendue pour la topologie uniforme, et tout point limite est solution de l'équation stochastique de Volterra (2.2.1).

En tant qu'application aux mathématiques financières, nous montrons que les représentations Markoviennes précédentes, combinées avec le résultat de stabilité, ont une importance cruciale dans la pratique car elles conduisent à de nouveaux schémas d'approximation numériques principalement pour la modélisation de la volatilité rugueuse. En effet, nous concevons des modèles de volatilité stochastique multi-facteurs tractables approchant les modèles de volatilité rugueuse tout en bénéficiant d'une structure Markovienne. De plus, nous appliquons notre procédure au cas spécifique du modèle d'Heston rugueux. Cela nous permet de dériver une méthode numérique pour résoudre l'équation de Riccati-Volterra correspondante dans ce contexte.

2.2.6 Lifting du modèle d'Heston

Dans le dernier chapitre, nous inversons le point de vue précédent en prenant le modèle Markovien en dimension finie comme point de départ. Nous introduisons une version *liftée* du modèle d'Heston à n-facteurs partageant le même mouvement Brownien mais avec des vitesses de retour à la moyenne différents. Le modèle inclus comme cas extrêmes le modèle classique d'Heston (quand $n = 1$) ainsi que le modèle d'Heston rugueux (quand n tend vers l'infini). Nous montrons que le modèle lifté bénéficie du meilleur des deux mondes: la markovianité et une interpolation satisfaisante du smile de volatilité implicite pour les maturités courtes. De plus, notre approche accélère le temps de calibration et ouvre la porte à des schémas de simulation rapides.

Structure

La deuxième partie de cette thèse est structurée comme suit.

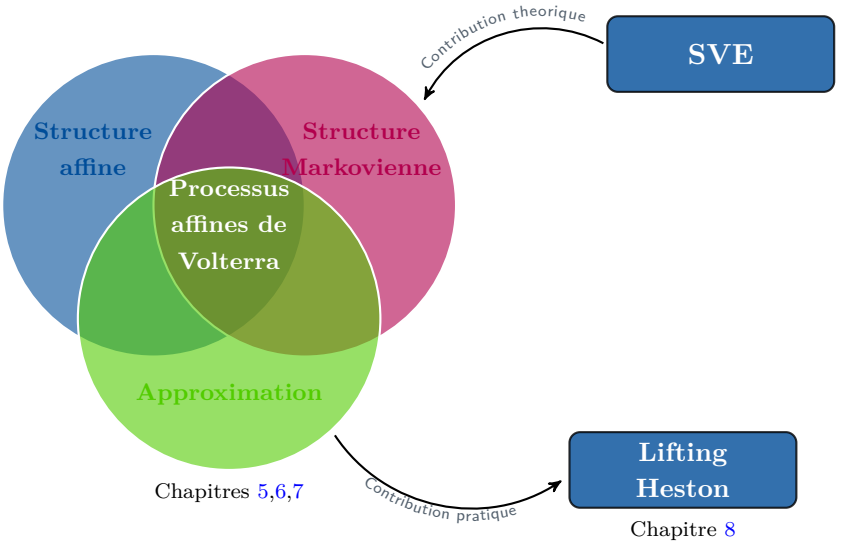

FIGURE 2.2: **Diagramme de la Partie II**

Part I

Stochastic invariance

Chapter 3

Stochastic invariance with with non-Lipschitz coefficients

Summary

This chapter provides a new characterization of the stochastic invariance of a closed subset of $\mathbb{R}^d$ with respect to a diffusion. We extend the well-known inward pointing Stratonovich drift condition to the case where the volatility matrix can fail to be differentiable: we only assume differentiability of the covariance matrix. In particular, our result can be applied to construct affine and polynomial diffusions on any arbitrary closed set.

Based on [3]: Abi Jaber, E., Bouchard, B., & Illand, C. (2016) Stochastic invariance of closed sets with non-Lipschitz coefficients. *Accepted for publication - Stochastic Processes and their Applications.*

3.1 Introduction

Let $b : \mathbb{R}^d \mapsto \mathbb{R}^d$ and $\sigma : \mathbb{R}^d \mapsto \mathbb{R}^{d\times d}$ be continuous functions, where $\mathbb{R}^{d\times d}$ denotes the space of $d \times d$ matrices. We assume that b and σ satisfy the following linear growth conditions: there exists $L > 0$ such that

$$\|b(x)\| + \|\sigma\sigma^\top(x)\|^{\frac{1}{2}} \leq L(1 + \|x\|), \quad \forall\, x \in \mathbb{R}^d, \tag{A_1}$$

and we consider a weak solution of the stochastic differential equation

$$dX_t = b(X_t)dt + \sigma(X_t)dW_t, \quad X_0 = x, \tag{3.1.1}$$

i.e. a d-dimensional Brownian motion W and an adapted process X such that the above equation holds.

The aim of this chapter is to provide a characterization of the *stochastic invariance* of a closed set $\mathcal{D} \subset \mathbb{R}^d$, i.e. find necessary and sufficient conditions on the instantaneous drift b

and the instantaneous covariance matrix $\sigma\sigma^\top$ under which there exists a weak solution of (3.1.1) that remains in $\mathcal{D}$ for all $t \geq 0$, almost surely, given that $x \in \mathcal{D}$. (See Definition 3.2 below for a precise formulation.)

The first stochastic invariance/viability results can be found in Stroock and Varadhan [107], Friedman [62] and Doss [47]. Since then, many extensions were considered in the literature. For an arbitrary closed set, the stochastic invariance was characterized through the second order normal cone in Bardi and Goatin [12] and Bardi and Jensen [13]. Aubin and Doss [10] used the notion of curvature, while Da Prato and Frankowska [38] provided a characterization in terms of the Stratonovich drift. For a closed convex set, the distance function was used in Da Prato and Frankowska [39], and the invariance was characterized for affine jump-diffusions in Tappe [109].

Although these approaches differ, they have at least one thing in common: the tradeoff one has to make between the assumptions on the topology/smoothness of the domain and the regularity of the coefficients b and σ. This makes all of these existing results difficult to apply in practice. Let us start by highlighting this difficulty through the two main contributions to the literature:

(i) In Bardi and Jensen [13], the stochastic invariance is characterized by using Nagumo-type geometric conditions on the second order normal cone. Their main result states that the closed set $\mathcal{D}$ is stochastically invariant if and only if

$$u^\top b(x) + \frac{1}{2}\operatorname{Tr}(va(x)) \leq 0, \quad x \in \mathcal{D},\ (u,v) \in \mathcal{N}_D^2(x),$$

in which $a := \sigma\sigma^\top$ on $\mathcal{D}$ and $\mathcal{N}_D^2(x)$ is the second order normal cone at the point x:

$$\mathcal{N}_{\mathcal{D}}^2(x) := \left\{(u,v) \in \mathbb{R}^d \times \mathbb{S}^d : \langle u, y-x\rangle + \frac{1}{2}\langle y-x, v(y-x)\rangle \leq o(\|y-x\|^2), \forall\, y \in \mathcal{D}\right\}. \tag{3.1.2}$$

Prior to deriving the conditions on b and σ, we have to determine the second order normal cone at all points of a given set. When the boundary is smooth, the computation of the second order normal cone is an easy task, see e.g. [13, Example 1]. However, it is much more challenging in general, by lack of efficient techniques. This renders the result of [13] difficult to use in practice. This also corresponds to the positive maximum principle of Ethier and Kurtz [54].

(ii) Building on Doss [47], Da Prato and Frankowska [38] give necessary and sufficient conditions for the stochastic invariance in terms of the Stratonovich drift and the first order normal cone:

$$\sigma(x)^\top u = 0 \text{ and } \langle u, b(x) - \frac{1}{2}\sum_{j=1}^{d} D\sigma^j(x)\sigma^j(x)\rangle \leq 0, \quad x \in \mathcal{D},\ u \in \mathcal{N}_{\mathcal{D}}^1(x), \tag{3.1.3}$$

where $\sigma^j(x)$ denotes the j-th column of the matrix $\sigma(x)$, $D\sigma^j$ is the Jacobian of σ^j, and the first order normal cone $\mathcal{N}_{\mathcal{D}}^1(x)$ at x (sometimes simply called *normal cone*) is defined as

$$\mathcal{N}_{\mathcal{D}}^1(x) := \left\{u \in \mathbb{R}^d : \langle u, y-x\rangle \leq o(\|y-x\|), \forall\, y \in \mathcal{D}\right\}. \tag{3.1.4}$$

In practice, the first order normal cone is much simpler to compute than the second order cone used in [13], see [11] and [102]. However, the price to pay is to impose a strong regularity condition on the diffusion matrix σ, which is assumed to be bounded

and differentiable on $\mathbb{R}^d$, with a bounded Lipschitz derivative. Again, this constitutes a sticking point for applications, it cannot be applied to simple models (think about square-root processes for instance, see below).

The aim of the present chapter is to extend the characterization (3.1.3), given in terms of the *easy-to-compute* first order normal cone, under weaker regularity conditions on the diffusion matrix σ. We make the following seemingly trivial observation: $a := \sigma\sigma^\top$ might be differentiable at a point x while σ is not. It is the case for the square-root process mentioned above, at the boundary point $x = 0$. Moreover, the terms $D\sigma^j(x)\sigma^j(x)$ can be rewritten in terms of the Jacobian of a whenever both quantities are well defined, see Proposition 3.4 for a precise formulation. This suggests to reformulate (3.1.3) with the Jacobian matrices of the columns of a instead of σ.

We prove that this is actually possible. Our main result, Theorem 3.3 below, states that the stochastic invariance is equivalent to the following conditions:

$$a(x)u = 0 \text{ and } \langle u, b(x) - \frac{1}{2}\sum_{j=1}^{d} Da^j(x)(aa^+)^j(x)\rangle \leq 0, \quad x \in \mathcal{D},\ u \in \mathcal{N}^1_{\mathcal{D}}(x). \tag{3.1.5}$$

Here, $(aa^+)^j(x)$ is the j-th column of $(aa^+)(x)$ with $a(x)^+$ defined as the Moore-Penrose pseudoinverse of $a(x)$, see Definition A.1 in the Appendix. We only assume that

$$a \text{ can be extended to a } \mathcal{C}^{1,1}_{loc}(\mathbb{R}^d, \mathbb{S}^d) \text{ function that coincides with } \sigma\sigma^\top \text{ on } \mathcal{D}, \tag{A_2}$$

in which $\mathcal{C}^{1,1}_{loc}$ means $\mathcal{C}^1$ with a locally Lipschitz derivative. Note that we do not impose the extension of a to be positive semi-definite outside $\mathcal{D}$, so that σ might only match with its square-root on $\mathcal{D}$. Also, it should be clear that the extension needs only to be local around $\mathcal{D}$.

The term aa^+ in (3.1.5) plays the role of the projection on the image of a, see Proposition A.3 in the Appendix and the discussion in Remark 3.5 below. This projection term cannot be removed. To see this, let us consider the square-root process with $a(x) = x$ and $\mathcal{D} = \mathbb{R}_+$, so that $\mathcal{N}^1_{\mathcal{D}}(0) = \mathbb{R}_-$. Then,

$$a(0)(-1) = 0 \quad \text{and} \quad \langle -1, b(0) - \frac{1}{2}Da(0)\rangle \leq 0$$

leads to $b(0) \geq 1/2$ while the correct condition for invariance is $b(0) \geq 0$, which is recovered from (3.1.5) by using the fact that $(aa^+)(0) = 0$.

This extension of the characterization of Da Prato and Frankowska [38] provides for the first time a unified criteria for the case where the volatility matrix may not be $\mathcal{C}^1$ on the whole domain, which is of importance in practical situations. In fact, many models used in practice, in mathematical finance for instance, do not have $\mathcal{C}^1$ volatility maps but satisfy our conditions. This is in particular the case of affine diffusions (see [48, 58]), or of polynomial diffusions that are characterized by a quadratic covariance matrix (see [35, 57]), etc. When applied to such processes, stochastic invariance results have been so far tweaked in order to fit in the previous set up, or have been proved under limiting conditions, on a case by case basis. For instance, in their construction of affine processes on the cone of symmetric semi-definite matrices, Cuchiero *et al.* [34] start by regularizing the martingale problem before applying the stochastic invariance characterization of [38] and then pass to the limit. In Spreij and Veerman [106], some stochastic invariance results are also derived for affine diffusions but

only on convex sets with smooth boundary. More recently, the mathematical foundation for polynomial diffusions was given in Filipović and Larsson [57]. Necessary conditions for the stochastic invariance are derived for basic closed semialgebraic sets. All the above cases can now be treated by using our characterization. See Section 3.5 for a generic example.

Our proof of the necessary condition is in the spirit of Buckdahn *et al.* [23]. They use a second order stochastic Taylor expansion together with small time behavior results for double stochastic integrals. However, in our case, the stochastic Taylor expansion cannot be applied directly since σ is not differentiable and $\sigma(X)$ fails to be a semi-martingale whenever an eigenvalue vanishes (see [93, Example 1.2]). We therefore need to develop new ideas. We first observe that, if σ is diagonal, then vanishing eigenvalues can be eliminated by taking the conditional expectation with respect to the path of the Brownian motion acting on the non-vanishing ones. This corresponds to the projection term aa^+ in (3.1.5). If σ is not diagonal, we can essentially reduce to the former case by considering its spectral decomposition and a suitable change of Brownian motion (based on the corresponding basis change), see Lemma 3.7 below. However, it requires a smooth spectral decomposition which is not guaranteed when repeated eigenvalues are present. To avoid this, we need an additional transformation of the state space, see Proposition 3.10.

Conversely, we show that the infinitesimal generator of our diffusion satisfies the *positive maximum principle* whenever (3.1.5) holds, see Section 3.4 below. Applying [54, Theorem 4.5.4] shows that this condition is indeed sufficient. (Note that the approach based on the comparison principle for viscosity solutions used in [13, 23] cannot be applied to our case since σ is not Lipschitz.)

The rest of the chapter is organized as follows. Our main result is stated in Section 3.2. The proofs are collected in Sections 3.3 and 3.4. In Section 3.5, we exemplify our characterization by deriving explicit stochastic invariance conditions for various typical examples of applications. Finally, Section 3.6 provides a complementary tractable sufficient condition ensuring the stochastic invariance of the interior of a domain. For the convenience of the reader, we collect some standard results of matrix calculus and differentiation in Appendix A.

From now on, all identities involving random variables have to be considered in the a.s. sense, the probability space and the probability measure being given by the context. Elements of $\mathbb{R}^d$ are viewed as column vectors, in particular the i-th element of the canonical basis of $\mathbb{R}^d$ is denoted by e_i.

3.2 Main result

In this section, we state our main result, Theorem 3.3, that extends Theorem 4.1 in Da Prato and Frankowska [38] to weaker regularity assumptions.

Since we are dealing with general coefficients b and σ, i.e. not necessarily Lipschitz coefficients, solutions to the stochastic differential equation (3.1.1) should be considered in the weak sense rather than in the strong sense. Existence is guaranteed by our condition (A_1), together with our standing assumption of continuity of b and σ: there exist a filtered probability space $(\Omega, \mathcal{F}, \mathbb{F} = (\mathcal{F}_t)_{t\geq 0}, \mathbb{P})$ satisfying the usual conditions, a d-dimensional $\mathbb{F}$-Brownian motion W and a $\mathbb{F}$-adapted process X with continuous sample paths such that (3.1.1) holds $\mathbb{P}-$a.s. See e.g. [74, Theorems IV.2.3 and IV.2.4].

For later use, note that Assumption (A_1) implies that, for any positive integer p, there exists $K_{p,x} > 0$ such that

$$\mathbb{E}\left[\|X_t - X_s\|^p\right] \leq K_{p,x}|t-s|^{\frac{p}{2}} \tag{3.2.1}$$

for all $0 \leq s,t \leq 1$. Hence, Kolmogorov's continuity criterion ensures that the sample paths of X are (locally) η-Hölder continuous for any $\eta \in (0, \frac{1}{2})$ (up to considering a suitable modification).

Remark 3.1. The collection $\mathcal{Q}$ of possible distributions of X is entirely determined by the infinitesimal generator $\mathcal{L}$ defined on the space of smooth functions ϕ by $\mathcal{L}\phi := D\phi\, b + \frac{1}{2}\mathrm{Tr}[\sigma\sigma^\top D^2\phi]$. Therefore, $\mathcal{Q}$ is the same if σ is replaced by $\tilde{\sigma}$ such that $\tilde{\sigma}\tilde{\sigma}^\top = \sigma\sigma^\top$, see e.g. [108, Remark 5.1.7]. Hence, we can reduce to the case where σ is the symmetric square-root of a on $\mathcal{D}$, which we will assume from now on.

Before stating our main result, let us make precise the definition of stochastic invariance.[1]

Definition 3.2 (*Stochastic invariance*)**.** A closed subset $\mathcal{D} \subset \mathbb{R}^d$ is said to be *stochastically invariant* with respect to the diffusion (3.1.1) if, for all $x \in \mathcal{D}$, there exists a weak solution (X, W) to (3.1.1) starting at $X_0 = x$ such that $X_t \in \mathcal{D}$ for all $t \geq 0$, almost surely.

Our characterization of stochastic invariance reads as follows (see Propositions 3.10 and 3.12 below for the proof). From now on we use the same notation a for a defined as $\sigma\sigma^\top$ on $\mathcal{D}$ and for its extension defined in Assumption (A_2).

Theorem 3.3 (Invariance characterization)**.** *Let $\mathcal{D}$ be closed. Assume that b, σ and a are continuous and satisfy assumptions (A_1)-(A_2). Then, the set $\mathcal{D}$ is stochastically invariant with respect to the diffusion (3.1.1) if and only if*

$$\begin{cases} a(x)u = 0 & \text{(3.2.2a)} \\ \langle u, b(x) - \frac{1}{2}\sum_{j=1}^{d} Da^j(x)(aa^+)^j(x)\rangle \leq 0 & \text{(3.2.2b)} \end{cases}$$

for every $x \in \mathcal{D}$ and for all $u \in \mathcal{N}^1_{\mathcal{D}}(x)$.

Clearly, the regularity conditions of Theorem 3.3 are much weaker than those of [38, Theorem 4.1]. Let us immediately exemplify this by considering the case of the square-root process already mentioned in the introduction. Let $\mathcal{D} = \mathbb{R}_+$, $a(x) = \eta^2 x$ with $\eta > 0$, and consider the diffusion $dX_t = b(X_t)dt + \eta\sqrt{X_t}dW_t$. Since $a(x)a(x)^+ = \mathbb{1}_{\{x>0\}}$ and $\mathcal{N}^1_{\mathbb{R}_+}(x) = \mathbb{1}_{\{x=0\}}\mathbb{R}_-$, Theorem 3.3 implies that $\mathbb{R}_+$ is stochastically invariant if and only if $b(0) \geq 0$, while $\sigma : x \in \mathbb{R}_+ \mapsto \eta\sqrt{x}$ is not differentiable at 0.

On the other hand, one can easily recover [38, Theorem 4.1] under their smoothness assumptions. This is the object of the next proposition (recall that, by Remark 3.1, the study can be reduced to the case $a = \sigma^2$ on $\mathcal{D}$).

[1] This concept is also known as viability. More precisely a set $\mathcal{D}$ is said to be viable if there exists a $\mathcal{D}$-valued solution started from $X_0 \in \mathcal{D}$. The set is said to be invariant if all solutions started from $X_0 \in \mathcal{D}$ remain in $\mathcal{D}$. In case of uniqueness of solutions, the two notions coincide. In this thesis, we will use the term invariance to stay coherent with the literature on affine and polynomial processes, where usually uniqueness of solutions hold, although our results go beyond these processes.

Proposition 3.4. *Fix $\sigma \in \mathcal{C}_b^{1,1}(\mathbb{R}^d, \mathbb{S}^d)$ (i.e. σ is differentiable with a bounded and a globally Lipschitz derivative). Then $a := \sigma^2 \in \mathcal{C}_{loc}^{1,1}(\mathbb{R}^d, \mathbb{S}_+^d)$ and*

$$\langle u, \sum_{j=1}^d D\sigma^j(x)\sigma^j(x)\rangle = \langle u, \sum_{j=1}^d Da^j(x)(aa^+)^j(x)\rangle, \quad x \in \mathcal{D},\ u \in \operatorname{Ker}\sigma(x).$$

Proof. Fix $x \in \mathcal{D}$ and $u \in \operatorname{Ker}\sigma(x)$. By using Definition A.7 and Proposition A.8 in the Appendix, we first compute that

$$Da(x) = D(\sigma(x)^2) = (\sigma(x) \otimes I_d)D\sigma(x) + (I_d \otimes \sigma(x))D\sigma(x),$$

which clearly shows that a is $\mathcal{C}_{loc}^{1,1}$. It then follows from Proposition A.5 and the fact that $u \in \operatorname{Ker}\sigma(x)$ that

$$(I_d \otimes u^\top)Da(x)a(x)a(x)^+ = (\sigma(x) \otimes u^\top)D\sigma(x)a(x)a(x)^+.$$

Observe now that $a(x)a(x)^+\sigma(x) = \sigma(x)$ since $a(x) = \sigma(x)^2$ (use the spectral decomposition of σ as in Proposition A.2). Using Proposition A.5 again, the above implies that

$$\begin{aligned}\operatorname{Tr}\left[(I_d \otimes u^\top)Da(x)a(x)a(x)^+\right] &= \operatorname{Tr}\left[\sigma(x)(I_d \otimes u^\top)D\sigma(x)a(x)a(x)^+\right] \\ &= \operatorname{Tr}\left[(I_d \otimes u^\top)D\sigma(x)\sigma(x)\right].\end{aligned}$$

Then, by Proposition A.5 and A.8,

$$\begin{aligned}\langle u, \sum_{j=1}^d D\sigma^j(x)\sigma^j(x)\rangle &= \sum_{j=1}^d u^\top D(\sigma(x)e_j)\sigma(x)e_j \\ &= \sum_{j=1}^d u^\top(e_j^\top \otimes I_d)D\sigma(x)\sigma(x)e_j \\ &= \sum_{j=1}^d e_j^\top(I_d \otimes u^\top)D\sigma(x)\sigma(x)e_j \\ &= \operatorname{Tr}\left[(I_d \otimes u^\top)D\sigma(x)\sigma(x)\right] \\ &= \operatorname{Tr}\left[(I_d \otimes u^\top)Da(x)a(x)a(x)^+\right] \\ &= \langle u, \sum_{j=1}^d Da^j(x)(aa^+)^j(x)\rangle, \end{aligned} \tag{3.2.3}$$

in which the last identity follows by reproducing exactly the same computations in the reverse order with a in place of σ. □

The following provides another formulation of (3.2.2b) that highlights the notion of projection on the image of a.

Remark 3.5 (Interpretation of the projection formulation). Fix $x \in \partial\mathcal{D}$ and assume that the spectral decomposition of a takes the form $a(x) = Q(x)\operatorname{diag}[\lambda_1(x), \dots, \lambda_r(x), 0, \dots, 0]\,Q(x)^\top$, where $Q(x)Q(x)^\top = I_d$ and $\lambda_j(x) > 0$ for all $1 \le j \le r$. Hence, the r-first columns of $Q(x)$, denoted by $(q_1, \dots, q_r) = (q_1(x), \dots, q_r(x))$, span the image of $a(x)$ and the projection matrix on the image of $a(x)$ is given by $a(x)a(x)^+ = \sum_{j=1}^r q_j q_j^\top$, see Propositions A.3 and A.2 in the

Appendix and recall that q_j is a column vector. Thus, by (3.2.3) in the proof of Proposition 3.4 and Proposition A.5 in the Appendix,

$$\begin{aligned}\langle u, \sum_{j=1}^{d} Da^j(x)(aa^+)^j(x)\rangle &= \mathrm{Tr}\left[(I_d \otimes u^\top)Da(x)a(x)a(x)^+\right] \\ &= \sum_{j=1}^{r} \mathrm{Tr}\left[(I_d \otimes u^\top)Da(x)q_j q_j^\top\right] \\ &= \sum_{j=1}^{r} \mathrm{Tr}\left[q_j^\top (I_d \otimes u^\top)Da(x)q_j\right] \\ &= \sum_{j=1}^{r} u^\top (q_j^\top \otimes I_d)Da(x)q_j\end{aligned}$$

so that, by Proposition A.8,

$$\langle u, \sum_{j=1}^{d} Da^j(x)(aa^+)^j(x)\rangle = \langle u, \sum_{j=1}^{r} D(aq_j)(x)q_j\rangle = \langle u, \sum_{j=1}^{r} D_{q_j}(aq_j)(x)\rangle$$

in which D_{q_j} is the directional derivative with respect to q_j:

$$D_{q_j}(aq_j)(x) := \lim_{t\to 0} \frac{a(x+tq_j)q_j - a(x)q_j}{t}.$$

Therefore (3.2.2b) reads $\langle u, b(x) - \frac{1}{2}\sum_{j=1}^{r} D_{q_j}(aq_j)(x)\rangle \le 0$. Otherwise stated, a is first projected onto the basis of the image of $a(x)$ before being derived only in the directions of $(q_1, ..., q_r)$. This is clearly consistent with (3.2.2a) that states that there cannot be any transverse diffusion of $a(x)$ to the boundary. Therefore, the drift $b(x)$ should only compensate the tangential diffusion given by the projection onto the image of $a(x)$ in order to keep the diffusion in the domain.

3.3 Necessary conditions

In this section, we prove that the conditions of Theorem 3.3 are necessary for $\mathcal{D}$ to be invariant.

Our general strategy is similar to [23]. We fix $x \in \mathcal{D}$ and we consider a smooth function $\phi : \mathbb{R}^d \mapsto \mathbb{R}$ such that $\max_{\mathcal{D}} \phi = \phi(x)$. Since $\mathcal{D}$ is stochastically invariant, let X be a $\mathcal{D}$-valued solution starting from $X_0 = x$. In particular, $\phi(X_t) \le \phi(x)$, for all $t \ge 0$. Then, if σ is sufficiently smooth, by applying Itô's Lemma twice, we obtain

$$\int_0^t \mathcal{L}\phi(X_s)ds + \int_0^t \left(D\phi\sigma(x) + \int_0^s \mathcal{L}(D\phi\sigma)(X_r)dr + \int_0^s D(D\phi\sigma)\sigma(X_r)dW_r\right)^\top dW_s \le 0.$$

Recall Remark 3.1 for the definition of the infinitesimal generator $\mathcal{L}$. Given (now standard) estimates on the small time behavior of single and double stochastic integrals, see e.g. [23, 30], this readily implies

$$D\phi(x)\sigma(x) = 0 \quad \text{and} \quad \langle D\phi(x), b(x) - \frac{1}{2}\sum_{j=1}^{d} D\sigma^j(x)\sigma^j(x)\rangle \le 0,$$

under appropriate regularity conditions. It remains to choose a suitable test function ϕ, i.e. such that $D\phi(x) = u^\top$, to deduce that (3.2.2a)-(3.2.2b) must hold when σ is differentiable, recall Proposition 3.4.

In our setting, one can however not differentiate σ^j in general. To surround this problem the above can be rewritten in term of the covariance matrix a. The projection term in (3.2.2a)-(3.2.2b) will appear through a conditioning argument.

In order to separate the difficulties, we shall first consider the case where a admits a locally smooth spectral decomposition. The general case will be handled in Section 3.3.2 below.

3.3.1 The case of distinct eigenvalues

As mentioned above, we shall first make profit of having distinct eigenvalues before considering the general case. The main idea consists in using the spectral decomposition of a in the form $Q\Lambda Q^\top$ in which Q is an orthogonal matrix and Λ is diagonal positive semi-definite. Then, the dynamics of X can be written as

$$dX_t = b(X_t)dt + Q(X_t)\Lambda(X_t)^{\frac{1}{2}}dB_t$$

in which $B = \int_0^\cdot Q(X_s)^\top dW_s$ is a Brownian motion. If Q and Λ are smooth enough, then we can apply the same ideas as the one exposed at the beginning of this section. An additional localization and conditioning argument will allow us to reduce to the case where Λ has only (strictly) positive entries.

Note that eigenvalues and the eigenvectors can always be chosen measurable. However, multiple eigenvalues and their corresponding eigenvectors can fail to have the same regularity as a. To ensure a sufficient regularity, we therefore assume in the following Lemma that non-zero eigenvalues are distinct. The general case will be treated later, thanks to a change of variable argument, see Section 3.3.2 below.

Lemma 3.6. *Assume that $a \in \mathcal{C}^{1,1}_{loc}(\mathbb{R}^d, \mathbb{S}^d)$. Let $x \in \mathcal{D}$ be such that the spectral decomposition of $a(x)$ is given by*

$$a(x) = Q(x)\mathrm{diag}\left[\lambda_1(x), \dots, \lambda_r(x), 0, \dots, 0\right] Q(x)^\top \tag{3.3.1}$$

with $\lambda_1(x) > \lambda_2(x) > \cdots > \lambda_r(x) > 0$ and $Q(x)Q(x)^\top = I_d$, $r \le d$. Then there exist an open (bounded) neighborhood $N(x)$ of x and two measurable $\mathbb{R}^{d\times d}$-valued functions on $\mathbb{R}^d$, $y \mapsto Q(y) := [q_1(y) \cdots q_d(y)]$ and $y \mapsto \Lambda(y) := \mathrm{diag}\left[\lambda_1(y), \dots, \lambda_d(y)\right]$ such that

(i) *$a(y) = Q(y)\Lambda(y)Q(y)^\top$ and $Q(y)Q(y)^\top = I_d$, for all $y \in \mathbb{R}^d$,*

(ii) *$\lambda_1(y) > \lambda_2(y) > \dots > \lambda_r(y) > \max\{\lambda_i(y), r+1 \le i \le d\} \vee 0$, for all $y \in N(x)$,*

(iii) *$\bar\sigma : y \mapsto \bar Q(y)\bar\Lambda(y)^{\frac{1}{2}}$ is $C^{1,1}(N(x), \mathbb{R}^{d\times d})$, in which $\bar Q := [q_1 \cdots q_r \ 0 \cdots 0]$ and $\bar\Lambda = \mathrm{diag}[\lambda_1, \dots, \lambda_r, 0, \dots, 0]$.*

Moreover, we have:

$$\langle u, \sum_{j=1}^d D\bar\sigma^j(x)\bar\sigma^j(x)\rangle = \langle u, \sum_{j=1}^d Da^j(x)(aa^+)^j(x)\rangle, \quad u \in \mathrm{Ker}(a(x)). \tag{3.3.2}$$

Proof. Note that the fact that $(q_i)_{i\leq d}$ can be chosen measurable is guaranteed when (C,Λ) is measurable by the fact that each eigenvector solves a quadratic minimization problem, see e.g. [19, Proposition 7.33(p.153)]. Moreover, the continuity of the eigenvalues follows from Weyl's perturbation theorem, [20, Corollary III.2.6], and the smoothness of $(\bar{\Lambda},\bar{Q})$ is a consequence of [89, Theorem 1] since all the positive eigenvalues are simple and a is $\mathcal{C}^{1,1}_{loc}(\mathbb{R}^d,\mathbb{S}^d)$. Let us now observe that any $u\in \mathrm{Ker}(a(x))$ satisfies

$$u^\top \bar{Q}(x) = u^\top \bar{\sigma}(x) = 0.$$

Since $\bar{a} := \bar{\sigma}\bar{\sigma}^\top$ is differentiable at x, the product rule of Proposition A.8 combined with Proposition A.5 yields

$$\begin{aligned}(I_d\otimes u^\top)D\bar{a}(x) &= (I_d\otimes u^\top)\left[(\bar{\sigma}(x)\otimes I_d)D\bar{\sigma}(x) + (I_d\otimes\bar{\sigma}(x))D\bar{\sigma}(x)^\top\right]\\ &= (\bar{\sigma}(x)\otimes u^\top)D\bar{\sigma}(x)\\ &= \bar{\sigma}(x)(I_d\otimes u^\top)D\bar{\sigma}(x).\end{aligned}$$

Observing that $\bar{a} = \bar{\sigma}\bar{\sigma}^\top = a\bar{Q}\bar{Q}^\top$ and that $\bar{Q}(x)\bar{Q}(x)^\top = a(x)a(x)^+$, we get by similar computations:

$$\begin{aligned}(I_d\otimes u^\top)D\bar{a}(x) &= (I_d\otimes u^\top)\left[(a(x)a(x)^+\otimes I_d)Da(x) + (I_d\otimes a(x))D\left(\bar{Q}\bar{Q}^\top\right)(x)\right]\\ &= a(x)a(x)^+(I_d\otimes u^\top)Da(x).\end{aligned}$$

Combining the above leads to

$$\mathrm{Tr}\left[(I_d\otimes u^\top)D\bar{\sigma}(x)\bar{\sigma}(x)\right] = \mathrm{Tr}\left[(I_d\otimes u^\top)Da(x)a(x)a(x)^+\right],$$

which proves (3.3.2) by similar computations as in the proof of (3.2.3). □

We can now adapt the arguments of [23]. In the following we use the notion of proximal normals. A vector $u\in\mathbb{R}^d$ is said to be a proximal normal to $\mathcal{D}$ at a point x if $\|u\| = d_{\mathcal{D}}(x+u)$, where $d_{\mathcal{D}}$ is the distance function to $\mathcal{D}$. We denote by $\mathcal{N}^{1,prox}_{\mathcal{D}}(x)$ the cone spanned by all proximal normals. Note however that (3.2.2a)-(3.2.2b) holds at x for all proximal normals $u\in\mathcal{N}^{1,prox}_{\mathcal{D}}(x)$ if and only if it holds for all $u\in\mathcal{N}^1_{\mathcal{D}}(x)$. Indeed,

$$\mathcal{N}^{1,prox}_{\mathcal{D}}(x)\subset\mathcal{N}^1_{\mathcal{D}}(x)\subset\bar{\mathrm{co}}\left(\limsup_{\mathcal{D}\ni y\to x}\mathcal{N}^{1,prox}_{\mathcal{D}}(y)\right), \tag{3.3.3}$$

where lim sup stands for the Painlevé-Kuratowski upper limit (see e.g. [11, 38]) and $\bar{\mathrm{co}}$ is the closed convex hull (see also [38, Remark 4.2 (a)]).

Lemma 3.7. *Assume that $\mathcal{D}$ is stochastically invariant with respect to the diffusion (3.1.1). Let $x\in\mathcal{D}$ and a be as in Lemma 3.6. Then, (3.2.2a) and (3.2.2b) hold at x for all $u\in\mathcal{N}^1_{\mathcal{D}}(x)$.*

Proof. It follows from the discussion before our lemma that it suffices to prove our claim for $u\in\mathcal{N}^{1,prox}_{\mathcal{D}}(x)$. Let (X,W) denote a weak solution starting at $X_0=x$ such that $X_t\in\mathcal{D}$ for all $t\geq 0$. If $x\notin\partial\mathcal{D}$, then $\mathcal{N}^{1,prox}_{\mathcal{D}}(x)=\{0\}$ and there is nothing to prove. We therefore assume from now on that $x\in\partial\mathcal{D}$. We fix $u\in\mathcal{N}^{1,prox}_{\mathcal{D}}(x)$.
Step 1. We first claim that there exists a function $\phi\in\mathcal{C}^\infty_b(\mathbb{R}^d,\mathbb{R})$ with compact support in $N(x)$ such that $\max_{\mathcal{D}}\phi = \phi(x) = 0$ and $D\phi(x) = u^\top$. Indeed, it follows from [102, Chapter

6.E] that one can find $\kappa > 0$ such that $\langle u, y - x\rangle \leq \frac{\kappa}{2}\|y - x\|^2$ for all $y \in \mathcal{D}$. Then, one can set $\psi := \langle u, \cdot - x\rangle - \frac{\kappa}{2}\|\cdot - x\|^2$ and define $\phi := \psi\rho$ in which ρ is a $\mathcal{C}_b^\infty$ function with values in $[0, 1]$, compact support included in $N(x)$, and satisfying $\rho = 1$ in a neighborhood of x.
Step 2. Since $\mathcal{D}$ is invariant under the diffusion X, $\phi(X_t) \leq \phi(x)$, for all $t \geq 0$. From now on, we use the notations of Lemma 3.6. By the above and Itô's lemma:

$$0 \geq \int_0^t \mathcal{L}\phi(X_s)ds + \int_0^t D\phi(X_s)\sigma(X_s)dW_s = \int_0^t \mathcal{L}\phi(X_s)ds + \int_0^t (D\phi Q\Lambda^{\frac{1}{2}}Q^\top)(X_s)dW_s$$

in which $\mathcal{L}$ is the infinitesimal generator of X. Let us define the Brownian motion $B = \int_0^\cdot Q(X_s)^\top dW_s$, recall that Q is orthogonal, together with

$$\bar{B} = \Lambda(x)\Lambda(x)^+B = (B^1, .., B^r, 0, ..., 0) \text{ and } \bar{B}^\perp = (I_d - \Lambda(x)\Lambda(x)^+)B = (0, ..., 0, B^{r+1}, ..., B^d),$$

recall Proposition A.2. Since $Q\bar{\Lambda}^{\frac{1}{2}} = \bar{Q}\bar{\Lambda}^{\frac{1}{2}}$, the above inequality can be written in the form

$$0 \geq \int_0^t \mathcal{L}\phi(X_s)ds + \int_0^t D\phi(X_s)\bar{\sigma}(X_s)d\bar{B}_s + \int_0^t (D\phi Q\Lambda^{\frac{1}{2}})(X_s)d\bar{B}_s^\perp.$$

Let $(\mathcal{F}_s^{\bar{B}})_{s\geq 0}$ be the completed filtration generated by $\bar{B}$. By [88, Corollaries 2 and 3 of Theorem 5.13], [82, Lemma 14.2], and the fact that the martingale $\bar{B}^\perp$ is independent of $\bar{B}$, we obtain

$$\begin{aligned} 0 &\geq \int_0^t \mathbb{E}_{\mathcal{F}_s^{\bar{B}}}[\mathcal{L}\phi(X_s)]ds + \int_0^t \mathbb{E}_{\mathcal{F}_s^{\bar{B}}}[D\phi(X_s)\bar{\sigma}(X_s)]d\bar{B}_s \\ &= \int_0^t \mathbb{E}_{\mathcal{F}_s^{\bar{B}}}[\mathcal{L}\phi(X_s)]ds + \int_0^t \mathbb{E}_{\mathcal{F}_s^{\bar{B}}}[D\phi(X_s)\bar{\sigma}(X_s)]dB_s, \end{aligned}$$

where the last equality holds because the $(d - r)$ columns of $\bar{\sigma}$ are 0. We now apply Lemma 3.8 below to $(D\phi\bar{\sigma})(X)$ and use [88, Corollaries 2 and 3 of Theorem 5.13] and [82, Lemma 14.2] again to find a bounded adapted process η such that

$$0 \geq \int_0^t \theta_s ds + \int_0^t \left(\alpha + \int_0^s \beta_r dr + \int_0^s \gamma_r dB_r\right)^\top dB_s \tag{3.3.4}$$

where

$$\begin{aligned} \theta := \mathbb{E}_{\mathcal{F}_\cdot^{\bar{B}}}[\mathcal{L}\phi(X_\cdot)] \quad &, \quad \alpha^\top := (D\phi\bar{\sigma})(x) = u^\top Q(x)\Lambda(x)^{\frac{1}{2}} \\ \beta := \mathbb{E}_{\mathcal{F}_\cdot^{\bar{B}}}[D(D\phi\bar{\sigma})(X_\cdot)b(X_\cdot) + \eta_\cdot] \quad &, \quad \gamma := \mathbb{E}_{\mathcal{F}_\cdot^{\bar{B}}}[D(D\phi\bar{\sigma})\bar{\sigma}(X_\cdot)], \end{aligned}$$

recall from Step 1 that $D\phi(x) = u^\top$.
Step 3. We now check that we can apply Lemma 3.9 below. First note that all the above processes are bounded. This follows from Lemma 3.6, Assumption (A_1) and the fact that ϕ has compact support. In addition, given $T > 0$, the independence of the increments of $\bar{B}$ implies that $\theta_s = \mathbb{E}_{\mathcal{F}_T^{\bar{B}}}[\mathcal{L}\phi(X_s)]$ for all $s \leq T$. It follows that θ is a.s. continuous at 0. Moreover, since $D\phi\bar{\sigma}$ is $\mathcal{C}^{1,1}$, $D(D\phi\bar{\sigma})\bar{\sigma}$ is Lipschitz which, combined with (3.2.1), implies condition (3.3.6) below.
Step 4. In view of Step 3, we can apply Lemma 3.9 to (3.3.4) to deduce that $\alpha = 0$ and $\theta_0 - \frac{1}{2}\mathrm{Tr}(\gamma_0) \leq 0$. Multiplying the first equation by $\Lambda(x)^{\frac{1}{2}}Q^\top(x)$ implies that $0 = \alpha^\top\Lambda(x)^{\frac{1}{2}}Q^\top(x) = u^\top Q(x)\Lambda(x)^{\frac{1}{2}}\Lambda(x)^{\frac{1}{2}}Q^\top(x) = u^\top a(x)$, or equivalently $a(x)u = 0$ since $a(x)$ is symmetric. The second identity combined with $D\phi(x) = u^\top$ and Proposition A.8 shows

that

$$0 \geq \mathcal{L}\phi(x) - \frac{1}{2}\operatorname{Tr}\left[\bar{\sigma}^\top D^2\phi\bar{\sigma} + (I_d \otimes u^\top)D\bar{\sigma}\bar{\sigma}\right](x) = u^\top b(x) - \frac{1}{2}\operatorname{Tr}\left[(I_d \otimes u^\top)D\bar{\sigma}\bar{\sigma}\right](x),$$

which is equivalent to (3.2.2b) by (3.3.2) and similar computations as in the proof of (3.2.3). □

The rest of this section is dedicated to the proof of the two technical lemmas that were used above. Our first result is a slight extension of Itô's lemma to only $\mathcal{C}^{1,1}$ function. It is based on a simple application of Komlós lemma (note that the assumption that f has a compact support in the following is just for convenience, it can obviously be removed by a localization argument, in which case the process η is only locally bounded).

Lemma 3.8. *Assume that b and σ are continuous and that there exists a solution (X, W) to (3.1.1). Let $f \in \mathcal{C}^{1,1}(\mathbb{R}^d, \mathbb{R})$ have compact support. Then, there exists an adapted bounded process η such that*

$$f(X_t) = f(x) + \int_0^t \left(Df(X_s)b(X_s) + \eta_s\right) ds + \int_0^t Df(X_s)\sigma(X_s)dW_s$$

for all $t \geq 0$.

Proof. Since $f \in \mathcal{C}^{1,1}$ has a compact support, we can find a sequence $(f_n)_n$ in $\mathcal{C}^\infty$ with compact support (uniformly) and a constant $K > 0$ such that

(i) $\|D^2 f_n\| \leq K$,

(ii) $\|f_n - f\| + \|Df_n - Df\| \leq \frac{K}{n}$,

for all $n \geq 1$. This is obtained by considering a simple mollification of f. By applying Itô's Lemma to $f_n(X)$, we get

$$f_n(X_t) = f_n(x) + \int_0^t Df_n(X_s)b(X_s)ds + \int_0^t \eta_s^n ds + \int_0^t Df_n(X_s)\sigma(X_s)dW_s$$

in which $\eta^n := \frac{1}{2}\mathrm{Tr}[D^2 f_n \sigma\sigma^\top](X)$. Since $\sigma\sigma^\top$ is continuous, (i) above implies that $(\eta^n)_n$ is uniformly bounded in $L^\infty(dt \times d\mathbb{P})$. By [43, Theorem 1.3], there exists $(\widetilde{\eta}^n) \in \mathrm{Conv}(\eta^k, k \geq n)$ such that $\widetilde{\eta}^n \to \eta$ $dt \otimes d\mathbb{P}$ almost surely. Let $N_n \geq 0$ and $(\lambda_k^n)_{n \leq k \leq N_n} \subset [0,1]$ be such that $\widetilde{\eta}^n = \sum_{k=n}^{N_n} \lambda_k^n \eta^k$ and $\sum_{k=n}^{N_n} \lambda_k^n = 1$. Set $\widetilde{f}_n := \sum_{k=n}^{N_n} \lambda_k^n f_k$. Then,

$$\widetilde{f}_n(X_t) = \widetilde{f}_n(x) + \int_0^t D\widetilde{f}_n(X_s)b(X_s)ds + \int_0^t \widetilde{\eta}_s^n ds + \int_0^t D\widetilde{f}_n(X_s)\sigma(X_s)dW_s. \tag{3.3.5}$$

By dominated convergence, $\int_0^t \widetilde{\eta}_s^n ds$ converges a.s. to $\int_0^t \eta_s ds$. Moreover, (ii) implies that

$$\|\widetilde{f}_n(X_t) - f(X_t)\| \leq \sum_{k=n}^{N_n} \lambda_k^n \|\widetilde{f}_k(X_t) - f(X_t)\| \leq \sum_{k=n}^{N_n} \lambda_k^n \frac{K}{k} \leq \frac{K}{n},$$

so that $\widetilde{f}_n(X_t)$ converges a.s. to $f(X_t)$. Similarly,

$$\int_0^t D\widetilde{f}_n(X_s)\sigma(X_s)dW_s \to \int_0^t Df(X_s)\sigma(X_s)dW_s, \int_0^t D\widetilde{f}_n(X_s)b(X_s)ds \to \int_0^t Df(X_s)b(X_s)ds$$

in $L^2(\Omega, \mathcal{F}, \mathbb{P})$ as $n \to \infty$, and therefore a.s. after possibly considering a subsequence. It thus remains to send $n \to \infty$ in (3.3.5) to obtain the required result. □

The following adapts [23, Lemma 2.1] to our setting, see also [22, 30].

Lemma 3.9. *Let $(W_t)_{t\geq 0}$ denote a standard d-dimensional Brownian motion on a filtered probability space $(\Omega, \mathcal{F}, (\mathcal{F}_t)_{t\geq 0}, \mathbb{P})$. Let $\alpha \in \mathbb{R}^d$ and $(\beta_t)_{t\geq 0}$, $(\gamma_t)_{t\geq 0}$ and $(\theta_t)_{t\geq 0}$ be predictable processes taking values respectively in $\mathbb{R}^d$, $\mathbb{R}^{d\times d}$ and $\mathbb{R}$ and satisfying*

(1) *β is bounded,*

(2) *$\int_0^t \|\gamma_s\|^2 ds < \infty$, for all $t \geq 0$,*

(3) *there exists $\eta > 0$ such that*

$$\int_0^t \int_0^s \mathbb{E}\left[\|\gamma_r - \gamma_0\|^2\right] dr ds = O(t^{2+\eta}), \tag{3.3.6}$$

(4) *θ is a.s. continuous at 0.*

Suppose that for all $t \geq 0$

$$\int_0^t \theta_s ds + \int_0^t \left(\alpha + \int_0^s \beta_r dr + \int_0^s \gamma_r dW_r\right)^\top dW_s \leq 0. \tag{3.3.7}$$

Then,

(a) *$\alpha = 0$,*

(b) *$-\gamma_0 \in \mathbb{S}_+^d$,*

(c) *$\theta_0 - \frac{1}{2}\operatorname{Tr}(\gamma_0) \leq 0$.*

Proof. Since $(W_t^i)^2 = 2\int_0^t W_s^i dW_s^i + t$, (3.3.7) reduces to

$$(\theta_0 - \frac{1}{2}\operatorname{Tr}(\gamma_0))t + \sum_{i=1}^d \alpha^i W_t^i + \sum_{i=1}^d \frac{\gamma_0^{ii}}{2}(W_t^i)^2 + \sum_{1\leq i\neq j\leq d} \gamma_0^{ij} \int_0^t W_s^i dW_s^j + R_t \leq 0,$$

where

$$\begin{aligned} R_t &= \int_0^t (\theta_s - \theta_0) ds + \int_0^t \left(\int_0^s \beta_r dr\right)^\top dW_s + \int_0^t \left(\int_0^s (\gamma_r - \gamma_0) dW_r\right)^\top dW_s \\ &=: R_t^1 + R_t^2 + R_t^3. \end{aligned}$$

In view of [23, Lemma 2.1], it suffices to show that $R_t/t \to 0$ in probability. To see this, first note that $R_t^1 = o(t)$ a.s. since θ is continuous at 0. Moreover, [30, Proposition 3.9] implies that $R_t^2 = o(t)$ a.s., as β is bounded. Finally, it follows from (3.3.6) that $\frac{R_t^3}{t} \to 0$ in L^2, and hence in probability. We conclude by applying [23, Lemma 2.1]. □

3.3.2 The general case

We can now turn to the general case.

Proposition 3.10 (Necessary conditions of Theorem 3.3)**.** *Let the conditions of Theorem 3.3 hold and assume that $\mathcal{D}$ is stochastically invariant with respect to the diffusion (3.1.1). Then conditions (3.2.2a) and (3.2.2b) hold for all $x \in \mathcal{D}$ and $u \in \mathcal{N}^1_{\mathcal{D}}(x)$.*

Proof. If x lies in the interior of $\mathcal{D}$, then $\mathcal{N}^1_{\mathcal{D}}(x) = \{0\}$ and there is nothing to prove. We therefore assume from now on that $x \in \partial\mathcal{D}$. Let Λ and Q be defined through the spectral decomposition of a, as in (3.3.1) but with only $\lambda_1(x) \geq \cdots \geq \lambda_d(x)$. We shall perform a change of variable to reduce to the conditions of Lemma 3.7. To do this, we fix $0 < \epsilon < 1$ and define

$$A^\epsilon = Q(x)\mathrm{diag}\left[\sqrt{(1-\epsilon)}, \sqrt{(1-\epsilon)^2}, \ldots, \sqrt{(1-\epsilon)^d}\right] Q(x)^\top.$$

Since $\mathcal{D}$ is invariant with respect to the diffusion X, $\mathcal{D}^\epsilon := A^\epsilon\mathcal{D}$ is invariant with respect to the diffusion $X^\epsilon := A^\epsilon X$. Note that

$$dX^\epsilon = b_\epsilon(X^\epsilon)dt + a_\varepsilon(X^\epsilon)^{\frac{1}{2}}dW$$

in which

$$b_\epsilon := A^\epsilon b((A^\epsilon)^{-1}\cdot) \quad \text{and} \quad a_\epsilon := A^\epsilon a((A^\epsilon)^{-1}\cdot)(A^\epsilon)^\top$$

have the same regularity and growth as b and a. Moreover, the positive eigenvalues of a_ϵ are all distinct at $x^\epsilon := A^\epsilon x$, as $a_\epsilon(x^\varepsilon) = Q(x)\mathrm{diag}\left[(1-\epsilon)\lambda_1(x), \ldots, (1-\epsilon)^d\lambda_d(x)\right] Q(x)^\top$. We can therefore apply Lemma 3.7 to $(X^\epsilon, \mathcal{D}^\epsilon)$:

$$\begin{cases} a_\epsilon(x^\epsilon)u_\epsilon = 0 & \text{(3.3.8a)} \\ \langle u_\epsilon, b_\epsilon(x^\epsilon) - \dfrac{1}{2}\displaystyle\sum_{j=1}^d Da^j_\epsilon(x^\epsilon)(a_\epsilon a_\epsilon^+)^j(x^\epsilon)\rangle \leq 0 & \text{(3.3.8b)} \end{cases}$$

for all $u_\epsilon \in \mathcal{N}^1_{A^\epsilon\mathcal{D}}(x^\epsilon)$. We now easily verify that $\mathcal{N}^1_{A^\epsilon\mathcal{D}}(x^\epsilon) = (A^\epsilon)^{-1}\mathcal{N}^1_{\mathcal{D}}(x)$, recall the definition in (3.1.4). Finally, by sending $\epsilon \to 0$ in (3.3.8a) and (3.3.8b), we get by continuity:

$$\begin{cases} a(x)u = 0 \\ \langle u, b(x) - \dfrac{1}{2}\displaystyle\sum_{j=1}^d Da^j(x)(aa^+)^j(x)\rangle \leq 0, \end{cases}$$

for all $u \in \mathcal{N}^1_{\mathcal{D}}(x)$, which ends the proof. □

3.4 Sufficient conditions

In this section, we prove that the necessary conditions of Proposition 3.10 are also sufficient. We start by showing in Proposition 3.11 that (3.2.2a) and (3.2.2b) imply that the generator $\mathcal{L}$ of X satisfies the *positive maximum principle*: $\mathcal{L}\phi(x) \leq 0$ for any $x \in \mathcal{D}$ and any function $\phi \in \mathcal{C}^2(\mathbb{R}^d, \mathbb{R})$ such that $\max_{\mathcal{D}} \phi = \phi(x) \geq 0$, see e.g. [54, p165]. Then, classical arguments,

mainly [54, Theorem 4.5.4], yield the existence of a solution to the corresponding martingale problem that stays in $\mathcal{D}$, see Proposition 3.12 below.

The following proposition is inspired by [38, Remark 5.6].

Proposition 3.11. *Under the assumptions of Theorem 3.3, assume that (3.2.2a)-(3.2.2b) hold for all $x \in \mathcal{D}$ and $u \in \mathcal{N}^1_{\mathcal{D}}(x)$. Then, the generator $\mathcal{L}$ satisfies the positive maximum principle.*

Proof. We fix $x \in \mathcal{D}$. For $1 \leq j \leq d$, let us consider the following deterministic control system:

$$\begin{cases} y'(t) = a(y(t))\sigma(x)^+ e_j \\ y(0) = x, \end{cases} \tag{3.4.1}$$

where $\sigma(x)^+$ is the pseudoinverse of $\sigma(x)$. Since a is locally Lipschitz and verifies condition (3.2.2a), [38, Proposition 2.5] combined with (3.3.3) implies that $\mathcal{D}$ is invariant with respect to the deterministic control system (3.4.1). Then, by definition of the second order normal cone in (3.1.2),

$$\langle u, y(\sqrt{h}) - x\rangle + \frac{1}{2}\langle v(y(\sqrt{h}) - x), y(\sqrt{h}) - x\rangle \leq o(||y(\sqrt{h}) - x||^2)$$

for any $(u, v) \in \mathcal{N}^2_{\mathcal{D}}(x)$. On the other hand, since a is $\mathcal{C}^{1,1}_{loc}$, a Taylor expansion around 0 yields

$$y(\sqrt{h}) = x + \sqrt{h}a(x)\sigma(x)^+ e_j + \frac{h}{2}(e_j^\top \sigma(x)^+ \otimes I_d)Da(x)a(x)\sigma(x)^+ e_j + o(h),$$

recall Proposition A.8 and note that $(\sigma^+)^\top = \sigma^+$ since σ is symmetric. Now observe that $u \in \mathcal{N}^1_{\mathcal{D}}(x)$ whenever $(u, v) \in \mathcal{N}^2_{\mathcal{D}}(x)$. In particular, $u^\top a(x) = 0$ under (3.2.2a). Combining the above, and recalling Proposition A.5 then leads to

$$\frac{h}{2}e_j^\top(\sigma(x)^+ \otimes u^\top)Da(x)a(x)\sigma(x)^+ e_j + \frac{h}{2}e_j^\top \sigma(x)^+ a(x)va(x)\sigma(x)^+ e_j \leq o(h).$$

Note that $\sigma^+\sigma^+ = a^+$ and that $a\sigma^+\sigma^+ a = aa^+a = a$, see e.g. Definition A.1 and Proposition A.2, and recall that $(\sigma(x)^+ \otimes u^\top) = \sigma(x)^+(I_d \otimes u^\top)$ by Proposition A.5. Then, dividing the above by $h/2$ and sending $h \to 0$ before summing over $1 \leq j \leq d$ yields

$$\mathrm{Tr}\left((I_d \otimes u^\top)Da(x)a(x)a(x)^+\right) + \mathrm{Tr}\left(va(x)\right) \leq 0.$$

In view of (3.2.2b) and (3.2.3), this shows that

$$\langle b(x), u\rangle + \frac{1}{2}\mathrm{Tr}(va(x)) \leq \langle u, b(x) - \frac{1}{2}\sum_{j=1}^d Da^j(x)(aa^+)^j(x)\rangle \leq 0$$

for all $(u, v) \in \mathcal{N}^2_{\mathcal{D}}(x)$. To conclude, it remains to observe that $(D\phi(x), D^2\phi(x)) \in \mathcal{N}^2_{\mathcal{D}}(x)$ whenever $\phi \in \mathcal{C}^2(\mathbb{R}^d, \mathbb{R})$ is such that $\max_{\mathcal{D}} \phi = \phi(x) \geq 0$. Hence, $\mathcal{L}\phi(x) \leq 0$. □

Proposition 3.12 (Sufficient conditions of Theorem 3.3)**.** *Under the assumptions of Theorem 3.3, assume that conditions (3.2.2a) and (3.2.2b) hold for all $x \in \mathcal{D}$ and $u \in \mathcal{N}^1_{\mathcal{D}}(x)$. Then, $\mathcal{D}$ is stochastically invariant with respect to the diffusion (3.1.1).*

Proof. We already know from Proposition 3.11 that $\mathcal{L}$ satisfies the positive maximum principle. Then, [54, Theorem 4.5.4] yields the existence of a solution to the martingale problem associated to $\mathcal{L}$ with sample paths in the space of càdlàg functions with values in $\mathcal{D}^\Delta := \mathcal{D} \cup \{\Delta\}$, the one-point compactification of $\mathcal{D}$. The discussion preceding [29, Proposition 3.2] and [54, Proposition 5.3.5], recall our linear growth conditions (A_1), then shows that the solution has a modification with continuous sample paths in $\mathcal{D}$. Finally, [54, Theorem 5.3.3] implies the existence of a weak solution (X, W) such that $X_t \in \mathcal{D}$ for all $t \geq 0$ almost surely. □

3.5 A generic application

We show in this section how Theorem 3.3 can be applied in various examples of application. We restrict to a two-dimensional setting for ease of computations and notations.

We first provide a generic tractable characterization for the stochastic invariance of all state spaces $\mathcal{D} \subset \mathbb{R}^2$ of the following form:

$$\mathcal{D} = \{(\bar{x}, \widetilde{x}) \in \mathbb{R}^2, \bar{x} \in \mathcal{D}_1 \text{ and } \phi(\bar{x}, \widetilde{x}) \in \mathcal{D}_2\}, \tag{3.5.1}$$

where $\mathcal{D}_1 \subset \mathbb{R}$ and $\mathcal{D}_2 \subset \mathbb{R}$ are closed subsets and ϕ is a continuously differentiable function.

Then, $\mathcal{D}$ can be characterized through $\Phi : (\bar{x}, \widetilde{x}) \mapsto (\bar{x}, \phi(\bar{x}, \widetilde{x}))$ by

$$\mathcal{D} = \Phi^{-1}(\mathcal{D}_1 \times \mathcal{D}_2),$$

and [102, Exercise 6.7 and Proposition 6.41] provides the following description of the normal cone whenever

$$\Phi \text{ is differentiable at } x \text{ and its Jacobian } D\Phi(x) \text{ has full rank} \tag{H_x}$$

holds at any point $x \in \mathcal{D}$.

Proposition 3.13. *Fix $x = (\bar{x}, \widetilde{x}) \in \mathcal{D}$ such that (H_x) holds. Then,*

$$\mathcal{N}^1_{\mathcal{D}}(x) = \left\{ \begin{pmatrix} \bar{u} + \partial_1\phi(x)\widetilde{u} \\ \partial_2\phi(x)\widetilde{u} \end{pmatrix}, \bar{u} \in \mathcal{N}^1_{\mathcal{D}_1}(\bar{x}) \text{ and } \widetilde{u} \in \mathcal{N}^1_{\mathcal{D}_2}(\phi(\bar{x}, \widetilde{x})) \right\},$$

in which $\partial_i\phi$ is the derivative with respect to the i-th component.

When x lies in the interior of $\mathcal{D}$, $\mathcal{N}^1_{\mathcal{D}}(x) = \{0\}$ and (3.2.2a)-(3.2.2b) are trivially verified. Hence, it suffices to control b and a on the boundary of the domain in order to ensure the stochastic invariance of $\mathcal{D}$ as stated by the following proposition, in which we use the notations

$$b = (\bar{b}, \widetilde{b})^\top, \ a = (a_{ij})_{ij} \ \text{ and } \ \partial_u = u_2\partial_1 - u_1\partial_2. \tag{3.5.2}$$

Proposition 3.14. *Let $\mathcal{D}$ be as in (3.5.1) and $x = (\bar{x}, \widetilde{x}) \in \partial\mathcal{D}$ be such that (H_x) holds. Fix $u = (u_1, u_2)^\top \in \mathcal{N}^1_{\mathcal{D}}(x)$ as in Proposition 3.13. Under the assumptions of Theorem 3.3, (3.2.2a)-(3.2.2b) are equivalent to the following:*

(a) Either $\widetilde{u} \neq 0$ *and*

$$\begin{cases} a(x) = a_{11}(x) \begin{pmatrix} 1 & -\frac{u_1}{u_2} \\ -\frac{u_1}{u_2} & \frac{u_1^2}{u_2^2} \end{pmatrix}, & \text{(3.5.3a)} \\ \langle u, b(x) \rangle - \frac{\mathbb{1}_{\{a_{11}(x) \neq 0\}}}{2(u_1^2 + u_2^2)} \left[u_1 u_2 \partial_u (a_{11} - a_{22})(x) + (u_2^2 - u_1^2) \partial_u a_{12}(x) \right] \leq 0 & \text{(3.5.3b)} \end{cases}$$

(b) Or, $\widetilde{u} = 0$, $u_1 = \bar{u}$ *and*

$$\begin{cases} a(x) \mathbb{1}_{\{\bar{u} \neq 0\}} = a_{22}(x) \begin{pmatrix} 0 & 0 \\ 0 & 1 \end{pmatrix} \mathbb{1}_{\{\bar{u} \neq 0\}}, & \text{(3.5.4a)} \\ \bar{u} \left(\bar{b}(x) - \frac{\mathbb{1}_{\{a_{22}(x) \neq 0\}}}{2} \partial_2 a_{12}(x) \right) \leq 0. & \text{(3.5.4b)} \end{cases}$$

Proof. Case (a), $\widetilde{u} \neq 0$: Since $D\Phi(x)$ has full rank, $\partial_2 \phi(x) \neq 0$ and therefore $u_2 \neq 0$. Since $a(x) \in \mathbb{S}^2$, (3.2.2a) is clearly equivalent to (3.5.3a).
If $a_{11}(x) \neq 0$, (3.5.3a) implies that $u = (\bar{u} + \partial_1 \phi(x) \widetilde{u}, \partial_2 \phi(x) \widetilde{u})^\top$ spans the kernel of $a(x)$. Therefore, by Proposition A.3,

$$a(x) a(x)^+ = I_2 - \frac{1}{\|u\|^2} u u^\top = \frac{1}{u_1^2 + u_2^2} \begin{pmatrix} u_2^2 & -u_1 u_2 \\ -u_1 u_2 & u_1^2 \end{pmatrix}.$$

Straightforward computations yield

$$\langle u, \sum_{j=1}^{2} Da^j(x) (aa^+)^j(x) \rangle = \frac{1}{u_1^2 + u_2^2} \left[u_1 u_2 \partial_u (a_{11} - a_{22})(x) + (u_2^2 - u_1^2) \partial_u a_{12}(x) \right],$$

recall the notations introduced in (3.5.2). This shows the equivalence between (3.2.2b) and (3.5.3b) when $a_{11}(x) \neq 0$.
If $a_{11}(x) = 0$, then (3.5.3a) implies that $a(x) a(x)^+ = 0$ and (3.2.2b) reads $\langle u, b(x) \rangle \leq 0$.
Case (b), $\widetilde{u} = 0$: If $\bar{u} = 0$, then $u = 0$ and there is nothing to prove. Otherwise, $u_1 = \bar{u} \neq 0$. Since $a(x) \in \mathbb{S}^2$, (3.2.2a) is clearly equivalent to $a_{11}(x) = 0$ and $a_{21}(x) = a_{12}(x) = 0$, that is (3.5.4a). If $a_{22}(x) \neq 0$, then (3.5.4a) provides

$$a(x) a(x)^+ = \begin{pmatrix} 0 & 0 \\ 0 & 1 \end{pmatrix},$$

and straightforward computations yield

$$\langle u, \sum_{j=1}^{2} Da^j(x) (aa^+)^j(x) \rangle = \bar{u} \partial_2 a_{12}(x),$$

which shows the equivalence between (3.2.2b) and (3.5.4b) when $a_{22}(x) \neq 0$. If $a_{22}(x) = 0$, then $a(x) a(x)^+ = 0$ and (3.2.2b) reads $\bar{u} \bar{b}(x) \leq 0$. □

Note that $\bar{u} = 0$ when $\mathcal{D}_1 = \mathbb{R}$, which will be the case from now on. In the sequel, we impose more structure on the coefficients, as it is usually done in the construction of invariant diffusions. This permits to deduce an explicit form of (a, b) on the whole domain from the boundary conditions (3.5.3a)-(3.5.3b). As already stated, Theorem 3.3 can be directly applied

to a large class of diffusions, e.g. affine diffusions [48, 58, 106] and polynomial diffusions [57, 85], not only for closed subsets of $\mathbb{R}^d$, but even when $\mathcal{D} \subset \mathbb{S}^d$ (as in [34]) since $\mathbb{S}^d$ can be identified with $\mathbb{R}^{\frac{d(d+1)}{2}}$ by using the half-vectorization operator. We start by defining these two main structures.

Definition 3.15 (Affine and polynomial diffusions)**.** X is a polynomial diffusion on $\mathcal{D}$ if:

(i) There exist $\bar{b}^i, \widetilde{b}^i \in \mathbb{R}$, $0 \leq i \leq 2$, and $A^i \in \mathbb{S}^2$, $1 \leq i \leq 5$, such that $b : x \mapsto b(x) := (\bar{b}(x), \widetilde{b}(x)) \in \mathbb{R}^2$ and $a : x \mapsto a(x) \in \mathbb{S}^2$ have the following form:

$$\left\{ \begin{array}{ll} \bar{b}(x) & = \bar{b}^0 + \bar{b}^1 \bar{x} + \bar{b}^2 \widetilde{x}, \\ \widetilde{b}(x) & = \widetilde{b}^0 + \widetilde{b}^1 \bar{x} + \widetilde{b}^2 \widetilde{x}, \\ a(x) & = A^0 + A^1 \bar{x} + A^2 \widetilde{x} + A^3 \bar{x}^2 + A^4 \bar{x}\widetilde{x} + A^5 \widetilde{x}^2, \end{array} \right. \tag{3.5.5}$$

for all $x = (\bar{x}, \widetilde{x}) \in \mathcal{D}$.

(ii) $a(x) \in \mathbb{S}^d_+$, for all $x \in \mathcal{D}$.

When $A^i = 0$ for all $3 \leq i \leq 5$, we say that X is an affine diffusion.

Then, it is clear that b and a are $\mathcal{C}^\infty$ and satisfy the linear growth conditions (A_1).

In what follows, we highlight the interplay between the geometry/curvature of the boundary and the coefficients b and a. The three explicit examples below characterize the invariance for flat, convex and concave boundaries.

Example 3.16 (Canonical state space)**.** *Fix $\mathcal{D}_1 = \mathbb{R}$, $\mathcal{D}_2 = \mathbb{R}_+$ and $\phi(\bar{x}, \widetilde{x}) = \widetilde{x}$. Then $\mathcal{D} = \mathbb{R} \times \mathbb{R}_+$ and $\mathcal{N}^1_{\mathcal{D}}(x) = \{0\} \times \mathbb{1}_{\{\widetilde{x}=0\}}\mathbb{R}_-$. Hence, (3.5.3a)-(3.5.3b) are equivalent to*

$$a(\bar{x}, 0) = a_{11}(\bar{x}, 0) \begin{pmatrix} 1 & 0 \\ 0 & 0 \end{pmatrix} \quad \text{and} \quad \widetilde{b}(\bar{x}, 0) - \frac{\mathbb{1}_{\{a_{11}(\bar{x},0)\neq 0\}}}{2} \partial_1 a_{12}(\bar{x}, 0) \geq 0, \quad \text{for all } \bar{x} \in \mathbb{R}.$$

If we now impose the structural condition (3.5.5), then straightforward computations lead to the characterization in [48] for affine diffusions. The case of polynomial diffusions can be treated similarly.

Example 3.17 (Parabolic convex state space)**.** *Let us consider the following parabolic state space:*

$$\mathcal{D} = \{(\bar{x}, \widetilde{x}) \in \mathbb{R}^2, \widetilde{x} \geq \bar{x}^2\}.$$

Then, with the previous notations, $\mathcal{D}_1 = \mathbb{R}$, $\mathcal{D}_2 = \mathbb{R}_+$ and $\phi(\bar{x}, \widetilde{x}) = \widetilde{x} - \bar{x}^2$. Therefore, the first order normal cone given by Proposition 3.13 reads

$$\mathcal{N}^1_{\mathcal{D}}(x) = \begin{pmatrix} -2\bar{x} \\ 1 \end{pmatrix} \mathbb{R}_-, \quad \text{for all } x = (\bar{x}, \bar{x}^2) \in \partial\mathcal{D}.$$

Conditions (3.5.3a)-(3.5.3b) are therefore equivalent to

$$\left\{ \begin{array}{ll} a(x) = a_{11}(x) \begin{pmatrix} 1 & 2\bar{x} \\ 2\bar{x} & 4\bar{x}^2 \end{pmatrix}, & \text{(3.5.6a)} \\ \langle u, b(x) \rangle - \dfrac{\mathbb{1}_{\{a_{11}(x)\neq 0\}}}{2(1+4\bar{x}^2)} \left[-2\bar{x}\partial_u(a_{11} - a_{22})(x) + (1 - 4\bar{x}^2)\partial_u a_{12}(x) \right] \geq 0, & \text{(3.5.6b)} \end{array} \right.$$

for all $\bar{x} \in \mathbb{R}$, $x = (\bar{x}, \bar{x}^2)$ *and* $u = (-2\bar{x}, 1)^\top$.
If we now impose an additional affine structure on the diffusion $X = (\bar{X}, \widetilde{X})$, *as in Duffie et al. [48, Section 12.2], we recover the characterization given in Gourieroux and Sufana [68, Proposition 2]. Indeed, Proposition 3.14 says that* $\mathcal{D}$ *is invariant if and only if there exists* $\alpha \geq 0$ *such that*

(a)

$$a(x) = \alpha \begin{pmatrix} 1 & 2\bar{x} \\ 2\bar{x} & 4\widetilde{x} \end{pmatrix}, \text{ for all } x = (\bar{x}, \widetilde{x}) \in \mathcal{D}, \tag{3.5.7}$$

(b) $\bar{b}^2 = 0$ *and*

$$\begin{cases} \widetilde{b}^2 > 2\bar{b}^1 \quad \text{and} \quad (\widetilde{b}^1 - 2\bar{b}^0)^2 \leq 4(\widetilde{b}^2 - 2\bar{b}^1)(\widetilde{b}^0 - \alpha) \\ \text{or} \\ \widetilde{b}^2 = 2\bar{b}^1, \quad \widetilde{b}^1 = 2\bar{b}^0 \quad \text{and} \quad \widetilde{b}^0 \geq \alpha. \end{cases} \tag{3.5.8}$$

Let us detail the computations: (a) The covariance matrix $a(x) \in \mathbb{S}_+^2$ *is of the form (3.5.6a) on the boundary. Since* a *is affine in* $(\bar{x}, \bar{x}^2)$, *then necessarily* $a_{11}(x)$ *is constant (or else* $a_{22}(x)$ *would have at least a polynomial dependence of order 3 in* $\bar{x}$*). Therefore, there exists* α *such that* $a(x)$ *has the form (3.5.7) at* $x = (\bar{x}, \bar{x}^2)$, *in which* $\alpha \geq 0$ *to ensure that* $a(0) \in \mathbb{S}_+^2$. *Finally,* a *needs to have the same form (3.5.7) on the whole state space* $\mathcal{D}$, *since it is affine. (b) We now derive the form of the drift vector* $b(x) = (\bar{b}(x), \widetilde{b}(x)) \in \mathbb{R}^2$ *by using (3.5.6b). From (3.5.7), elementary computations show that condition (3.5.6b) is equivalent to*

$$-2\bar{b}^2\bar{x}^3 + (\widetilde{b}^2 - 2\bar{b}^1)\bar{x}^2 + (\widetilde{b}^1 - 2\bar{b}^0)\bar{x} + \widetilde{b}^0 - \alpha \geq 0, \quad \text{for all } \bar{x} \in \mathbb{R},$$

which is equivalent to (3.5.8), when $\alpha > 0$. *If* $\alpha = 0$, *the same conclusion holds.*
Conversely, (3.5.7) clearly implies (3.2.2a) and (ii) of Definition 3.15 since $\det(a(x)) = 4\alpha(\widetilde{x} - \bar{x}^2) \geq 0$ *and* $\widetilde{x} \geq 0$ *for all* $(\bar{x}, \widetilde{x}) \in \mathcal{D}$. *Moreover, (3.5.8) leads to (3.5.3b) by the same computations as above.*

Example 3.18 (Parabolic concave state space)**.** *We now consider the epigraph of the concave function* $\bar{x} \mapsto -\bar{x}^2$,

$$\mathcal{D} = \{(\bar{x}, \widetilde{x}) \in \mathbb{R}^2, \widetilde{x} \geq -\bar{x}^2\}.$$

It follows that $\mathcal{D}_1 = \mathbb{R}$, $\mathcal{D}_2 = \mathbb{R}_+$, $\phi(\bar{x}, \widetilde{x}) = \widetilde{x} + \bar{x}^2$ *and*

$$\mathcal{N}_{\mathcal{D}}^1(x) = \begin{pmatrix} 2\bar{x} \\ 1 \end{pmatrix} \mathbb{R}_-, \quad \text{for all } x = (\bar{x}, -\bar{x}^2) \in \partial\mathcal{D},$$

from Proposition 3.13. Hence, conditions (3.5.3a)-(3.5.3b) are now equivalent to

$$\begin{cases} a(x) = a_{11}(x) \begin{pmatrix} 1 & -2\bar{x} \\ -2\bar{x} & 4\bar{x}^2 \end{pmatrix}, & \text{(3.5.9a)} \\ \langle u, b(x) \rangle - \dfrac{\mathbb{1}_{\{a_{11}(x) \neq 0\}}}{2(4\bar{x}^2 + 1)} \left[2\bar{x}\partial_u(a_{11} - a_{22})(x) + (1 - 4\bar{x}^2)\partial_u a_{12}(x) \right] \geq 0, & \text{(3.5.9b)} \end{cases}$$

for all $\bar{x} \in \mathbb{R}$, $x = (\bar{x}, -\bar{x}^2)$ *and* $u = (2\bar{x}, 1)^\top \in -\mathcal{N}_{\mathcal{D}}^1(x)$.
Let us first note that the above shows that we can not construct an affine diffusion living in

$\mathcal{D}$, that is not degenerate, unless it lives on the boundary only. Indeed, if a is affine then $a_{11} =: \alpha$ has to be constant, because of (3.5.9a), and a is of the form (3.5.9a) with $(-\widetilde{x})$ in place of $\bar{x}^2$. Since $a(x) \in \mathbb{S}_+^2$, we must have $\alpha \geq 0$ and $\det a(x) = -4\alpha^2(\widetilde{x} + \bar{x}^2) \geq 0$. Thus, $\alpha = 0$ unless we restrict to points $(\bar{x}, \widetilde{x})$ on the boundary. If we do so, it is not difficult to derive a necessary and sufficient condition on the coefficients from the identity $\widetilde{X} = -\bar{X}^2$. We now impose a polynomial structure on the diffusion $X = (\bar{X}, \widetilde{X})$, such that $\bar{X}$ is affine on its own, i.e. $\bar{b}$ and a_{11} are of affine form and only depend on $\bar{x}$. This extends [85, Example 5.2] and entirely characterizes the stochastic invariance of $\mathcal{D}$ with respect to this structure of diffusion. By Proposition 3.13, $\mathcal{D}$ is invariant if and only if there exist $\alpha, \beta \geq 0$, such that

(a)

$$a(x) = \begin{pmatrix} \alpha & -2\alpha\bar{x} \\ -2\alpha\bar{x} & (4\alpha + \beta)\bar{x}^2 + \beta\widetilde{x} \end{pmatrix}, \text{ for all } x = (\bar{x}, \widetilde{x}) \in \mathcal{D}, \tag{3.5.10}$$

(b) $\bar{b}^2 = 0$ *and*

$$\begin{cases} \widetilde{b}^2 < 2\bar{b}^1 \quad \text{and} \quad (\widetilde{b}^1 + 2\bar{b}^0)^2 \leq 4(-\widetilde{b}^2 + 2\bar{b}^1)(\widetilde{b}^0 + \alpha) \\ \text{or} \\ \widetilde{b}^2 = 2\bar{b}^1, \quad \widetilde{b}^1 = -2\bar{b}^0 \quad \text{and} \quad \widetilde{b}^0 \geq -\alpha \end{cases}. \tag{3.5.11}$$

Let us do the computations explicitly: (a) The covariance matrix $a(x) \in \mathbb{S}_+^2$ is of the form (3.5.9a) on the boundary. Therefore, $a_{11}(x) \geq 0$, for all $x \in \partial\mathcal{D}$. Since a_{11} is affine and only depends on $\bar{x} \in \mathbb{R}$, then necessarily a_{11} is a non-negative constant on the whole space $\mathcal{D}$. Therefore, there exists $\alpha \geq 0$ such that $a_{11}(.) = \alpha$ on $\mathcal{D}$. Moreover, (3.5.5) reads on the boundary

$$a(x) = A^0 + A^1\bar{x} + (A^3 - A^2)\bar{x}^2 - A^4\bar{x}^3 + A^5\bar{x}^4, \quad \text{for all } \bar{x} \in \mathbb{R}.$$

Therefore, comparing with (3.5.9a) leads to $A^4 = A^5 = 0$ and the existence of β, β' such that a is of the form

$$\begin{pmatrix} \alpha & -2\alpha\bar{x} \\ -2\alpha\bar{x} & 4\alpha\bar{x}^2 \end{pmatrix} + \begin{pmatrix} 0 & \beta' \\ \beta' & \beta \end{pmatrix} (\widetilde{x} + \bar{x}^2)$$

on the whole space $\mathcal{D}$. We now use the fact that $a(\mathcal{D}) \subset \mathbb{S}_+^2$. In particular, taking $\bar{x} = 0$ shows that we must have $\alpha\beta\widetilde{x} - (\beta')^2\widetilde{x}^2 \geq 0$ for all $\widetilde{x} \geq 0$, so that $\beta' = 0$. Similarly, $4\alpha\bar{x}^2 + \beta(\widetilde{x} + \bar{x}^2) \geq 0$ must hold for all $x \in \mathcal{D}$, which is equivalent to $\beta \geq 0$.
(b) We now derive the form of the drift vector $b(x) = (\bar{b}(x), \widetilde{b}(x)) \in \mathbb{R}^2$ by using (3.5.9b). Since X is affine on its own, $\bar{b}^2 = 0$. From (3.5.10), elementary computations show that condition (3.5.9b) is equivalent to

$$(-\widetilde{b}^2 + 2\bar{b}^1)\bar{x}^2 + (\widetilde{b}^1 + 2\bar{b}^0)\bar{x} + \widetilde{b}^0 + \alpha \geq 0, \quad \text{for all } \bar{x} \in \mathbb{R},$$

which is equivalent to (3.5.11), when $\alpha > 0$. If $\alpha = 0$, the same conclusion holds.
Conversely, (3.5.10)-(3.5.11) show that X is a polynomial diffusion such that $\bar{X}$ is affine on its own since $\det(a(x)) = \alpha\beta(\widetilde{x} + \bar{x}^2) \geq 0$ and $4\alpha\bar{x}^2 + \beta(\widetilde{x} + \bar{x}^2) \geq 4\alpha\bar{x}^2 \geq 0$ for all $(\bar{x}, \widetilde{x}) \in \mathcal{D}$. (3.5.10) clearly implies (3.2.2a). Moreover, (3.5.11) leads to (3.5.3b) by the same computations as above.

We conclude with a final remark on the interplay between the local geometry of the boundary, the coefficients a and b and the structure of the diffusion.

Remark 3.19. (i) Curvaceous boundary and covariance matrix: the curvature of the boundary plays a crucial role in determining the covariance structure. In Example 3.16, the canonical state space, which shows no curvature, imposes strict constraints on the covariance matrix. Whereas, for curved domains, as in Examples 3.17-3.18, the first order normal cone is a more complicated object and induces a richer covariance structure on the boundary.

(ii) Convexity and drift direction: Figure 3.1 visualizes the direction of the drift $b(0) = (\bar{b}^0, \widetilde{b}^0)$ with respect to the convexity of the domain. When the domain is convex, as in Example 3.17, the drift is necessarily inward pointing since $\widetilde{b}^0 \geq \alpha$, with $\alpha \geq 0$ from (3.5.8). However, when the domain is concave, as in Example 3.18, the drift could even be outward pointing. This follows from the fact that $\widetilde{b}^0 \geq -\alpha$, with $\alpha \geq 0$ in (3.5.11).

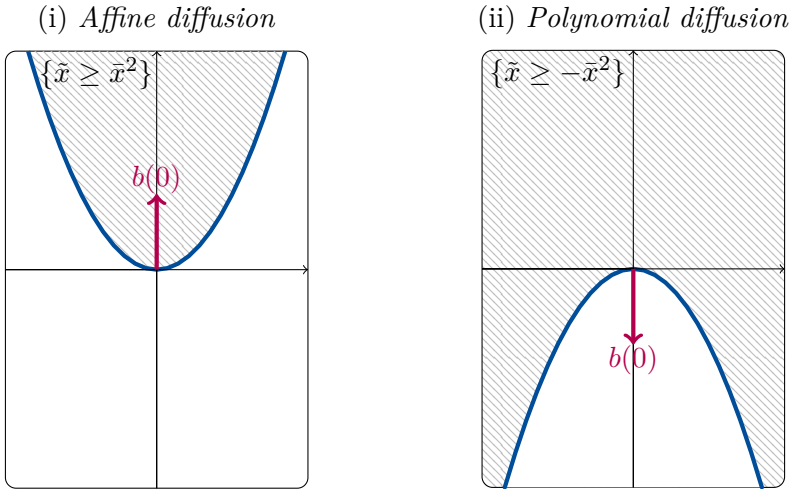

FIGURE 3.1: **Interplay between the convexity of the domain and the direction of the drift:** (i) Inward pointing drift for convex domains (Example 3.17). (ii) Possible outward pointing drift for concave domains (Example 3.18).

3.6 Additional remark on the boundary non-attainment

In this last section, we provide a sufficient condition for the stochastic invariance of the interior of $\mathcal{D}$, when $\mathcal{D}$ has a smooth boundary. The result is a direct implication of [106, Proposition 3.5] derived with the help of *McKean's argument* (see [92, Section 4]). Moreover, we extend the tractable conditions of [106, Proposition 3.7] given for affine diffusions. Our result could be easily used in the context of polynomial diffusions for instance.

Proposition 3.20. *Let $\mathcal{D} \subset \mathbb{R}^d$ be closed with a non-empty interior $\mathring{\mathcal{D}}$ that is a maximal connected subset of $\{x, \Phi(x) < 0\}$ where $\Phi \in \mathcal{C}^2(\mathbb{R}^d, \mathbb{R})$ such that $\partial\mathcal{D} = \Phi^{-1}(0)$. Assume that b and a are continuous and satisfy assumptions (A_1)-(A_2). Moreover, assume that $a \in \mathcal{C}^1(\mathbb{R}^d, \mathbb{S}^d_+)$. Then $\mathring{\mathcal{D}}$ is stochastically invariant if there exists $v \in \mathbb{R}^d$ such that*

$$\begin{cases} D\Phi(x)a(x) = \Phi(x)v^\top & \text{(3.6.1a)} \\ \langle D\Phi(x), b(x) - \frac{1}{2}\sum_{j=1}^{d} Da^j(x)e_j \rangle \leq 0 & \text{(3.6.1b)} \end{cases}$$

for all $x \in \mathcal{D}$.

Proof. Fix $x \in \mathring{\mathcal{D}}$. By differentiating (3.6.1a) with the help of Propositions A.8 and A.5, we obtain

$$vD\Phi(x) = (a(x) \otimes I_1)D^2\Phi(x) + (I_d \otimes D\Phi(x))Da(x) = a(x)D^2\Phi(x) + (I_d \otimes D\Phi(x))Da(x),$$

which, combined with (3.6.1b), leads to

$$\begin{aligned} \langle D\Phi(x), b(x)\rangle &\leq \frac{1}{2}\operatorname{Tr}\left[(I_d \otimes D\Phi(x)^\top)Da(x)\right] \\ &= -\frac{1}{2}\operatorname{Tr}\left[a(x)D^2\Phi(x)\right] + \frac{1}{2}D\Phi(x)v \\ &= -\frac{1}{2}\operatorname{Tr}\left[a(x)D^2\Phi(x)\right] + \frac{1}{2}\Phi(x)^{-1}D\Phi(x)a(x)D\Phi(x)^\top. \end{aligned}$$

We conclude by using [106, Proposition 3.5] (after a change of the sign, since $\mathcal{D}$ is assumed to be a connected subset of $\{x, \Phi(x) > 0\}$ in [106, Proposition 3.5]). □

Example 3.21. (i) Square root process: *Let us consider again the process defined by* $dX_t = b(X_t)dt + \eta\sqrt{X_t}dW_t$, *for some* $\eta > 0$, *on* $\mathcal{D} = \mathbb{R}_+$. *Then,* $\Phi : x \mapsto -x$ *and* (3.6.1a)-(3.6.1b) *are equivalent to* $v = \eta^2$ *and* $b(0) \geq \frac{\eta^2}{2}$. *These are the well known conditions for the boundary non-attainment of the square-root process.*

(ii) Affine diffusions: *More generally, let* $\mathcal{D} \subset \mathbb{R}^d$ *satisfy the assumptions of Proposition 3.20 and take* $a(x) = A^0 + \sum_{j=1}^d A^j x^j$ *for some* $A^j \in \mathbb{S}^d$, $1 \leq j \leq d$. *Then differentiating* a *shows that condition* (3.6.1b) *is equivalent to* $\langle D\Phi(x), b(x) - \frac{1}{2}\sum_{j=1}^d (A^j)^j\rangle \leq 0$ *yielding [106, Proposition 3.7].*

(iii) Jacobi diffusion: *Set* $\mathcal{D} = (0, 1]$ *and consider a polynomial diffusion* X *on* $\mathcal{D}$, *i.e.* b *is affine and* a *is a polynomial of degree two. Theorem 3.3 applied on* $[0, 1]$ *immediately yields that de dynamics of* X *must be of the form* $dX_t = \kappa(\theta - X_t)dt + \eta\sqrt{X_t(1 - X_t)}dW_t$ *where* $\kappa, \eta \geq 0$ *and* $0 \leq \theta \leq 1$. *Now a localized version of Proposition 3.20 shows that* $\mathcal{D} = (0, 1]$ *is stochastically invariant under the additional condition that* $\kappa\theta \geq \frac{\eta^2}{2}$.

Proposition 3.20 is important in practice since it gives, in many cases, the existence and the uniqueness of a global *strong solution* to (3.1.1) as discussed in the following remark.

Remark 3.22. Let $\mathcal{D}$ be as in Proposition 3.20. Assume that $a \in \mathcal{C}^2(\mathring{\mathcal{D}}, \mathbb{S}^d_+)$ and that b is locally Lipschitz (which is clearly the case for affine and polynomial diffusions). By [62, Remark 1 page 131], $\sigma = a^{\frac{1}{2}}$ is locally Lipschitz on $\mathring{\mathcal{D}}$. Therefore, when the boundary is never attained, (3.1.1) starting from any element $x \in \mathring{\mathcal{D}}$ admits a global *strong solution* and pathwise-uniqueness holds.

Chapter 4

Stochastic invariance with jumps

Summary

We extend the characterization of the previous chapter to account for jumps. We provide necessary and sufficient first order geometric conditions for the stochastic invariance of a closed subset of $\mathbb{R}^d$ with respect to a jump-diffusion under weak regularity assumptions on the coefficients. We also derive an equivalent formulation in the semimartingale framework.

Based on [1]: Abi Jaber, E. (2017) Stochastic invariance of closed sets for jump-diffusions with non-Lipschitz coefficients. *Electronic Communications in Probability, 22, paper no. 53, 15 pp.*

4.1 Introduction

We consider a weak solution to the following stochastic differential equation with jumps

$$dX_t = b(X_t)dt + \sigma(X_t)dW_t + \int \rho(X_{t-}, z)\left(\mu(dt, dz) - F(dz)dt\right), \quad X_0 = x, \tag{4.1.1}$$

that is: a filtered probability space $(\Omega, \mathcal{F}, \mathbb{F} = (\mathcal{F})_{t\geq 0}, \mathbb{P})$ satisfying the usual conditions and supporting a d-dimensional Brownian motion W, a Poisson random measure μ on $\mathbb{R}_+ \times \mathbb{R}^d$ with compensator $dt \otimes F(dz)$, and a $\mathbb{F}$-adapted process X with càdlàg sample paths such that (4.1.1) holds $\mathbb{P}$-almost surely.

Throughout this chapter, we assume that $b : \mathbb{R}^d \mapsto \mathbb{R}^d$, $\sigma : \mathbb{R}^d \mapsto \mathbb{R}^{d\times d}$ and $\rho : \mathbb{R}^d \times \mathbb{R}^d \mapsto \mathbb{R}^d$ are measurable. In addition, we assume that

$$b, \sigma \text{ and } \int \rho(., z)^\top H(\rho(., z))\rho(., z)F(dz) \text{ are continuous for any } H \in \mathcal{C}_b(\mathbb{R}^d, \mathbb{R}^{d\times d}), \quad (H_{\mathcal{C}})$$

where $\mathcal{C}_b(\mathbb{R}^d, \mathbb{R}^{d\times d})$ denotes the space of $\mathbb{R}^{d\times d}$-valued continuous bounded functions on $\mathbb{R}^d$. We also assume that there exist $q, L > 0$ such that, for all $x \in \mathbb{R}^d$,

$$\int_{\{\|\rho(x,z)\|>1\}} \|\rho(x,z)\|^q \ln \|\rho(x,z)\| F(dz) \leq L(1 + \|x\|^q), \tag{H_0}$$

$$\|b(x)\|^2 + \|\sigma\sigma^\top(x)\| + \int \|\rho(x,z)\|^2 F(dz) \leq L(1 + \|x\|^2). \tag{H_1}$$

Let $\mathcal{D}$ denote a closed subset of $\mathbb{R}^d$. Our aim is to characterize the stochastic invariance (a.k.a viability) of $\mathcal{D}$ under weak regularity assumptions, i.e. find necessary and sufficient conditions on the coefficients such that, for all $x \in \mathcal{D}$, there exists a $\mathcal{D}$-valued weak solution to (4.1.1) starting at x.

In the presence of jumps, invariance and viability problems have been studied in [103, 110, 59]. Note that a first order characterization for a smooth volatility matrix σ is given in [59], where the Stratonovich drift appears (see [38] for the diffusion case). For a second order characterization, we refer to [110, Propositions 2.13 and 2.15].

Combining the techniques used in Chapter 3 with the ones of [110], we derive for the first time in Theorem 4.2 below, a first order geometric characterization of the stochastic invariance with respect to (4.1.1) when the volatility matrix σ can fail to be differentiable. We also provide an equivalent formulation of the stochastic invariance with respect to semimartingales in Theorem 4.5. This extends Theorem 3.3 to the jump-diffusion case. From a practical perspective, this is the first known first order characterization that could be directly applied to construct affine [48, 80] and polynomial processes [35] on any arbitrary closed sets, since for these processes the volatility matrix can fail to be differentiable (on the boundary of the domain).

As in the previous chapter we only make the following assumption on the covariance matrix

$$a := \sigma\sigma^\top \text{ on } \mathcal{D} \text{ can be extended to a } C^{1,1}_{loc}(\mathbb{R}^d, \mathbb{S}^d) \text{ function}, \tag{H_2}$$

in which $\mathcal{C}^{1,1}_{loc}$ means $\mathcal{C}^1$ with a locally Lipschitz derivative and $\mathbb{S}^d$ denotes the set of $d \times d$ symmetric matrices. Note that we do not impose the extension of a to be positive semi-definite outside $\mathcal{D}$, so that σ might only match with its square-root on $\mathcal{D}$. Also, it should be clear that the extension needs only to be local around $\mathcal{D}$. We use the same notation a for $\sigma\sigma^\top$ on $\mathcal{D}$ and for its extension defined in Assumption (H_2).

The rest of the chapter is organized as follows. Our main results are stated and proved in Sections 4.2-4.3. In Section 4.4, we adapt some technical results from Chapter 3 to the jump-framework.

4.2 Stochastic invariance for stochastic differential equations with jumps

In order to ease the comparison with Chapter 3, we first provide in Theorem 4.2 below a characterization of the invariance for stochastic differential equations with jumps. An equivalent formulation in terms of semimartingales is also provided in the next section (see Theorem 4.5 below). We insist on the fact that the two formulations are equivalent by the representation theorem of semimartingales with characteristics as in (4.3.1) below in terms of a Brownian motion and a Poisson random measure (see [76, Theorem 2.1.2]).

We start by making precise the definition of stochastic invariance[1] for stochastic differential equations with jumps.

Definition 4.1 (*Stochastic invariance*)**.** A closed subset $\mathcal{D} \subset \mathbb{R}^d$ is said to be *stochastically invariant* with respect to the jump-diffusion (4.1.1) if, for all $x \in \mathcal{D}$, there exists a weak solution X to (4.1.1) starting at $X_0 = x$ such that $X_t \in \mathcal{D}$ for all $t \geq 0$, almost surely.

The following theorem provides a first order geometric characterization of the stochastic invariance in terms of the first order normal cone $\mathcal{N}^1_{\mathcal{D}}(x)$ as defined in (3.1.4). Recall that $\mathcal{N}^1_{\mathcal{D}}(x)$ consists of all outward pointing normal vectors to $\mathcal{D}$ at a point x.

Theorem 4.2. *Let $\mathcal{D} \subset \mathbb{R}^d$ be closed. Under the continuity assumptions ($H_{\mathcal{C}}$) and (H_0)-(H_2), the set $\mathcal{D}$ is stochastically invariant with respect to the jump-diffusion (4.1.1) if and only if*

$$\begin{cases} x + \rho(x,z) \in \mathcal{D}, \text{ for } F\text{-almost all } z, & \text{(4.2.1a)} \\ \int |\langle u, \rho(x,z)\rangle| F(dz) < \infty, & \text{(4.2.1b)} \\ a(x)u = 0, & \text{(4.2.1c)} \\ \langle u, b(x) - \int \rho(x,z)F(dz) - \frac{1}{2}\sum_{j=1}^{d} Da^j(x)(aa^+)^j(x)\rangle \leq 0, & \text{(4.2.1d)} \end{cases}$$

for all $x \in \mathcal{D}$ and $u \in \mathcal{N}^1_{\mathcal{D}}(x)$, in which $Da^j(x)$ denotes the Jacobian of the j-th column of $a(x)$ and $(aa^+)^j(x)$ is the j-th column of $(aa^+)(x)$ with $a(x)^+$ defined as the Moore-Penrose pseudoinverse of $a(x)$.

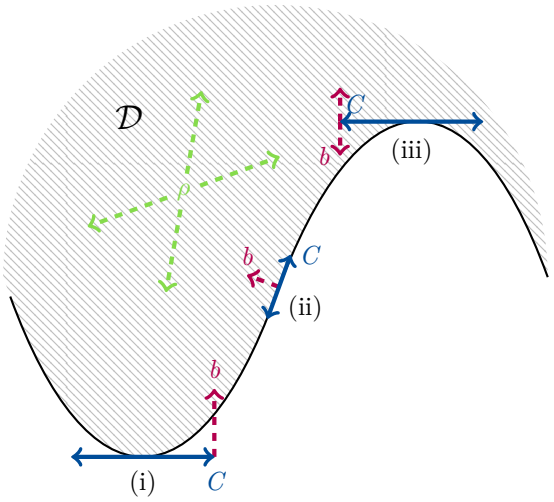

FIGURE 4.1: Interplay between the geometry/curvature of $\mathcal{D}$ and the coefficients (b, a, ρ).

Before moving to the proof, we start by giving the geometric interpretation of conditions (4.2.1a)-(4.2.1d), also shown in Figure 4.1. Condition (4.2.1c) states that at the boundary of the domain, the column of the covariance matrix should be tangential to the boundary,

[1] The concept is also often known as viability. We use the term invariance here in order to stay coherent with the affine/polynpmial literature.

while (4.2.1a) requires from $\mathcal{D}$ to capture all the jumps of the process. Moreover, at the boundary, the jumps can have infinite variation only if they are tangent to the boundary, by (4.2.1b). Finally, it follows from (4.2.1d) that the compensated drift should be inward pointing. We recall that the compensated drift extends the Stratonovich drift (see [38, 59]) when the volatility matrix can fail to be differentiable. In fact, if the volatility matrix is smooth, Proposition 3.4 yields

$$\langle u, \sum_{j=1}^{d} D\sigma^j(x)\sigma^j(x)\rangle = \langle u, \sum_{j=1}^{d} Da^j(x)(aa^+)^j(x)\rangle, \quad x \in \mathcal{D} \text{ and } u \in \operatorname{Ker}\sigma(x)^\top.$$

Conversely, the example of the square root process $a(x) = x$ and $\sigma(x) = \sqrt{x}$ on $\mathcal{D} := \mathbb{R}_+$ shows that σ may fail to be differentiable at 0 while a satisfies (H_2).

The proof of Theorem 4.2 adapts the argument of Chapter 3 combined with techniques taken from [110] to handle the jump component. For the necessity, we use the same conditioning/projection argument together with the small time behavior of double stochastic integrals as in the previous chapter. For the sufficiency, we show that conditions (4.2.1a)-(4.2.1d) imply the positive maximum principle for the infinitesimal generator and we conclude by applying [54, Theorem 4.5.4], which is possible by Lemma 4.3 below. The latter lemma highlights the role of the growth condition (H_0). In fact, (H_1) would only yield that $\mathcal{L}\phi$ is bounded. This is not enough to apply [54, Theorem 4.5.4].

We will need the continuity of the infinitesimal generator of (4.1.1) acting on smooth functions ϕ

$$\mathcal{L}\phi := D\phi b + \frac{1}{2}\operatorname{Tr}(D^2\phi\sigma\sigma^\top) + \int (\phi(. + \rho(., z)) - \phi - D\phi\rho(., z))\, F(dz), \tag{4.2.2}$$

where $D\phi^\top$ (resp. $D^2\phi$) is the gradient (resp. Hessian) of ϕ. In the sequel, we denote by $\mathcal{C}(\mathcal{D})$ the space of continuous functions on $\mathcal{D}$. We add the superscript p on $\mathcal{C}$ to denote functions with p-continuous derivatives for all $p \leq \infty$, and the subscript c (resp. 0) stands for functions with compact support (resp. vanishing at infinity). This is the object of the following lemma (a similar formulation in the semimartingale set-up can be found in [105, Lemma A.1]).

Lemma 4.3. *Under* ($H_{\mathcal{C}}$) *and* (H_0), $\mathcal{L}(\mathcal{C}_c^2(\mathcal{D})) \subset \mathcal{C}_0(\mathcal{D})$.

Proof. Let $\phi \in \mathcal{C}_c^2(\mathcal{D})$. We extend it to $\mathcal{C}_c^2(\mathbb{R}^d)$. Let $M > 0$ be such that $\phi(x) = 0$ if $\|x\| > M$ and fix $\|x\| > M + 1$. Then

$$\mathcal{L}\phi(x) = \int \phi(x + \rho(x, z))F(dz) = \int_{\{\|x+\rho(x,z)\|\leq M\}} \phi(x + \rho(x, z))F(dz).$$

On $\{\|x + \rho(x, z)\| \leq M\}$, $1 + M < \|x\| \leq M + \|\rho(x, z)\|$. Hence, ($H_0$) yields

$$\begin{aligned}|\mathcal{L}\phi(x)| &\leq \|\phi\|_\infty \int_{\{\|x+\rho(x,z)\|\leq M\}} \frac{\|\rho(x, z)\|^q \ln \|\rho(x, z)\|}{(\|x\| - M)^q \ln(\|x\| - M)} F(dz) \\ &\leq \|\phi\|_\infty L \frac{(1 + \|x\|^q)}{(\|x\| - M)^q} \frac{1}{\ln(\|x\| - M)},\end{aligned}$$

where $\|.\|_\infty$ is the uniform norm, which shows that $\mathcal{L}\phi(x) \to 0$ when $\|x\| \to \infty$. Moreover, denoting by $\Phi := \int(\phi(. + \rho(., z)) - \phi - D\phi\rho(., z))F(dz)$, we have for all $x, y \in \mathcal{D}$

$$\begin{aligned}\Phi(y) &= \int\int_0^1\int_0^t \rho(y,z)^\top D^2\phi(y + s\rho(y,z))\rho(y,z)dsdtF(dz)\\ &= \int\int_0^1\int_0^t \rho(y,z)^\top D^2\phi(x + s\rho(y,z))\rho(y,z)dsdtF(dz)\\ &\quad + \int\int_0^1\int_0^t \rho(y,z)^\top \left(D^2\phi(y + s\rho(y,z)) - D^2\phi(x + s\rho(y,z))\right)\rho(y,z)dsdtF(dz)\\ &=: I_1(x,y) + I_2(x,y).\end{aligned}$$

Observe that $I_2(x, y) \to 0$ when $y \to x$, since $D^2\phi$ is uniformly continuous (recall that ϕ has compact support). In addition, it follows from ($H_\mathcal{C}$) that $I_1(x, y) \to \Phi(x)$ when $y \to x$, which ends the proof. □

We can now move to the proof of Theorem 4.2.

Proof of Theorem (4.2). **Part a.** We first prove that our conditions are necessary. Let X denote a weak solution starting at $X_0 = x$ such that $X_t \in \mathcal{D}$ for all $t \geq 0$. If $x \notin \partial\mathcal{D}$, then $\mathcal{N}^1_\mathcal{D}(x) = \{0\}$ and there is nothing to prove. We therefore assume from now on that $x \in \partial\mathcal{D}$. Let $0 < \eta < 1$. Throughout the proof, we fix ψ_η a bounded continuous function on $\mathbb{R}^d$ such that $\psi_\eta = 0$ on $B_\eta(x)$ and $\psi_\eta \to \mathbb{1}_{\{\mathbb{R}^d\setminus\{0\}\}}$ for $\eta \downarrow 0$, where $B_\eta(x)$ is the open ball with center x and radius η.
Step 1. We start by proving (4.2.1a). Let $\epsilon > 0$ and $\phi_\epsilon : \mathbb{R}^d \mapsto [0, 1]$ be $\mathcal{C}^2$ such that $\phi_\epsilon = 0$ on $\mathcal{D} \cup B_\epsilon(x)$ and $\phi_\epsilon = 1$ on $(\mathcal{D} \cup B_{2\epsilon}(x))^c$. $\mathcal{D}$ is stochastically invariant, hence $\phi_\epsilon(X_t) = 0$, for all $t \geq 0$. Since ϕ_ϵ is twice differentiable and bounded, Itô's formula [77, Theorem I.4.57] yields

$$\int_0^t \mathcal{L}\phi_\epsilon(X_s) + \int_0^t D\phi_\epsilon(X_s)\sigma(X_s)dW_s + (\phi_\epsilon(X_{s-} + \rho(X_{s-}, .)) - \phi_\epsilon(X_{s-})) * (\mu - \nu) = 0,$$

where $*$ denotes the standard notation for stochastic integration with respect to a random measure (see [77]) and $\nu(dt, dz) := dtF(dz)$. By continuity of $\mathcal{L}\phi$ (see Lemma 4.3), taking the expectation, dividing by t and letting $t \to 0$ yield

$$\mathcal{L}\phi_\epsilon(x) = 0. \tag{4.2.3}$$

A change of probability measure with respect to the Doléans-Dade exponential $Z := \mathcal{E}(\psi_\eta * (\mu - \nu))$, which is uniformly integrable (see [86, Theorem IV.3] and the proof of [110, Proposition 2.13]), yields

$$\int_0^t \widetilde{\mathcal{L}}\phi_\epsilon(X_s)ds + \int_0^t D\phi_\epsilon(X_s)\sigma(X_s)dW_s + (\phi_\epsilon(X_{s-} + \rho(X_{s-}, .)) - \phi_\epsilon(X_{s-})) * (\mu - \widetilde{\nu}) = 0, \tag{4.2.4}$$

where

$$\widetilde{b} := b + \int \psi_\eta(z)\rho(., z)F(dz), \quad \widetilde{\nu}(dt, dz) := dt\widetilde{F}(dz), \quad \widetilde{F}(dz) := (1 + \psi_\eta(z))F(dz),$$
$$\widetilde{\mathcal{L}}\phi := D\phi\widetilde{b} + \frac{1}{2}\operatorname{Tr}(D^2\phi a) + \int (\phi(. + \rho(., z)) - \phi - D\phi\rho(., z))\,\widetilde{F}(dz).$$

By combining the above with (4.2.2), taking the expectation in (4.2.4), dividing by t and sending $t \to 0$, and invoking once again Lemma 4.3, we get

$$\mathcal{L}\phi_\epsilon(x) + \int \phi_\epsilon(x + \rho(x,z))\psi_\eta(z)F(dz) = 0.$$

It then follows from (4.2.3) that $\int \phi_\epsilon(x + \rho(x,z))\psi_\eta(z)F(dz) = 0$ for all $\eta \in (0,1)$. Sending $\eta \downarrow 0$ leads to $\int \phi_\epsilon(x + \rho(x,z))F(dz) = 0$, by monotone convergence (recall that $\phi_\epsilon \geq 0$). Hence

$$\int \mathbb{1}_{\{x+\rho(x+z)\in(\mathcal{D}\cup B_{2\epsilon}(x))^c\}}F(dz) = 0.$$

For $\epsilon \downarrow 0$, (4.2.1a) follows from monotone convergence again.

Step 2. By the proof of Proposition 3.10, it suffices to consider the case where the positive eigenvalues of the covariance matrix a at the fixed point $x \in \mathcal{D}$ are all distinct as in Lemma 3.6. We can also restrict the study to $\sigma = a^{\frac{1}{2}}$ (see Remark 3.1). We therefore use the notations of Lemma 3.6. We proceed as in Step 2 of the proof of Lemma 3.7 for the continuous part combined with the proof of [110, Proposition 2.13] for the jump part. Fix $u \in \mathcal{N}^1_{\mathcal{D}}(x)$ and let ϕ be a smooth function (with compact support in $N(x)$) such that $\max_{\mathcal{D}} \phi = \phi(x)$ and $D\phi(x) = u^\top$.[2] Since $\mathcal{D}$ is stochastically invariant, $\phi(X_t) \leq \phi(x)$, for all $t \geq 0$. Let $w_\eta := (\eta - 1)\psi_\eta$. By reapplying Step 1, with the test function ϕ (resp. w_η) instead of ϕ_ϵ (resp. ψ_η), we obtain

$$\begin{aligned} 0 &\geq \int_0^t \widetilde{\mathcal{L}}\phi(X_s)ds + \int_0^t D\phi(X_s)\sigma(X_s)dW_s + \widetilde{N}_t \\ &= \int_0^t \widetilde{\mathcal{L}}\phi(X_s)ds + \int_0^t (D\phi Q\Lambda^{\frac{1}{2}}Q^\top)(X_s)dW_s + \widetilde{N}_t, \end{aligned}$$

where $\widetilde{N}_s := (\phi(X_{s-} + \rho(X_{s-}, .)) - \phi(X_{s-})) * (\mu - \widetilde{\nu})$ is the pure-jump true martingale part under the new measure (since ϕ is Lipschitz and (H_1) holds). Let us define the Brownian motion $B = \int_0^\cdot Q(X_s)^\top dW_s$, recall that Q is orthogonal, together with $\bar{B} = (B^1, .., B^r, 0, ..., 0)^\top$ and $\bar{B}^\perp = (0, ..., 0, B^{r+1}, ..., B^d)$. Since $Q\bar{\Lambda}^{\frac{1}{2}} = \bar{Q}\bar{\Lambda}^{\frac{1}{2}}$, the above inequality can be written in the form

$$0 \geq \int_0^t \widetilde{\mathcal{L}}\phi(X_s)ds + \int_0^t D\phi(X_s)\bar{\sigma}(X_s)d\bar{B}_s + \int_0^t (D\phi Q\Lambda^{\frac{1}{2}})(X_s)d\bar{B}^\perp_s + \widetilde{N}_t.$$

Let $(\mathcal{F}^{\bar{B}}_s)_{s\geq 0}$ be the completed filtration generated by $\bar{B}$. Since $\bar{B}, \bar{B}^\perp$ are independent and $\bar{B}$ has independent increments, conditioning by $\mathcal{F}^{\bar{B}}_t$ yields, by Lemma 4.8 in the appendix,

$$0 \geq \int_0^t \mathbb{E}_{\mathcal{F}^{\bar{B}}_s}[\widetilde{\mathcal{L}}\phi(X_s)]ds + \int_0^t \mathbb{E}_{\mathcal{F}^{\bar{B}}_s}[D\phi(X_s)\bar{\sigma}(X_s)]d\bar{B}_s.$$

We now apply Lemma 4.7 of the Appendix to $(D\phi\bar{\sigma})(X)$ and reapply the same conditioning argument to find a bounded adapted process $\widetilde{\eta}$ such that

$$0 \geq \int_0^t \theta_s ds + \int_0^t \left(\alpha + \int_0^s \beta_r dr + \int_0^s \gamma_r dB_r\right)^\top dB_s, \tag{4.2.5}$$

[2] Such a function always exists (up to considering an element of the proximal normal cone), recall the discussion preceding Lemma 3.7 and Step 1 of the proof of the same Lemma.

where

$$\theta := \mathbb{E}_{\mathcal{F}_\cdot^{\bar{B}}}\left[\widetilde{\mathcal{L}}\phi(X_\cdot)\right] \quad , \quad \alpha^\top := (D\phi\bar{\sigma})(x) = D\phi(x)Q(x)\Lambda(x)^{\frac{1}{2}}$$
$$\beta := \mathbb{E}_{\mathcal{F}_\cdot^{\bar{B}}}\left[\widetilde{\eta}_\cdot\right] \quad , \quad \gamma := \mathbb{E}_{\mathcal{F}_\cdot^{\bar{B}}}\left[D(D\phi\bar{\sigma})\bar{\sigma}(X_\cdot)\right].$$

Step 3. We now check that we can apply Lemma 3.9 below. First note that all the above processes are bounded. This follows from Lemmas 3.6 and 4.3, (H_1) and the fact that ϕ has compact support. In addition, given $T > 0$, the independence of the increments of $\bar{B}$ implies that $\theta_s = \mathbb{E}_{\mathcal{F}_T^{\bar{B}}}\left[\widetilde{\mathcal{L}}\phi(X_s)\right]$ for all $s \leq T$. From Lemma 4.3 and since X has almost surely no jumps at 0, it follows that θ is a.s. continuous at 0. Moreover, since $D\phi\bar{\sigma}$ is $\mathcal{C}^{1,1}$, $D(D\phi\bar{\sigma})\bar{\sigma}$ is Lipschitz which, combined with (4.4.3), implies (3.3.6).
Step 4. In view of Step 3, we can apply Lemma 3.9 to (4.2.5) to deduce that $\alpha = 0$ and $\theta_0 - \frac{1}{2}\mathrm{Tr}(\gamma_0) \leq 0$. The first equation implies that $\alpha^\top\Lambda(x)^{\frac{1}{2}}Q^\top(x) = u^\top a(x) = 0$, or equivalently (4.2.1c) since $a(x)$ is symmetric. The second identity combined with $D\phi(x) = u^\top$ shows that

$$\begin{aligned}0 &\geq \widetilde{\mathcal{L}}\phi(x) - \frac{1}{2}\mathrm{Tr}\left[\bar{\sigma}^\top D^2\phi\bar{\sigma} + (I_d \otimes u^\top)D\bar{\sigma}\bar{\sigma}\right](x)\\ &= \mathcal{L}\phi(x) - \frac{1}{2}\mathrm{Tr}\left[\bar{\sigma}^\top D^2\phi\bar{\sigma} + (I_d \otimes u^\top)D\bar{\sigma}\bar{\sigma}\right](x) + (\eta - 1)\int \left(\phi(x + \rho(x,z)) - \phi(x)\right)\psi_\eta(z)F(dz),\end{aligned}$$

in which $\otimes$ stands for the Kronecker product (recall Definition A.4 and Proposition A.5 in the Appendix) and $D\bar{\sigma}$ is the Jacobian matrix of $\bar{\sigma}$ (see Definition A.7). Sending $\eta \downarrow 0$, by monotone convergence, we get

$$0 \geq \mathcal{L}\phi(x) - \frac{1}{2}\mathrm{Tr}\left[\bar{\sigma}^\top D^2\phi\bar{\sigma} + (I_d \otimes u^\top)D\bar{\sigma}\bar{\sigma}\right](x) + \int \left(\phi(x) - \phi(x + \rho(x + z))\right)F(dz). \tag{4.2.6}$$

In particular, since $\phi(x) = \max_{\mathcal{D}}\phi$, (4.2.1a) implies that $\int |\phi(x + \rho(x + z)) - \phi(x)|F(dz) = \int(\phi(x) - \phi(x + \rho(x + z)))F(dz) < \infty$. Moreover, the right hand side is equal to

$$-\int D\phi(x)\rho(x,z)F(dz) - \int\int_0^1\int_0^t \rho(x,z)^\top D^2\phi(x + s\rho(x,z))\rho(x,z)dsdtF(dz),$$

yielding (4.2.1b) (recall (H_1) and that ϕ has compact support). Combining (4.2.6), (3.3.2)-(4.2.2) and

$$\mathrm{Tr}\left[(I_d \otimes u^\top)D\bar{\sigma}\bar{\sigma}\right](x) = \langle u, \sum_{j=1}^d D\bar{\sigma}^j(x)\bar{\sigma}^j(x)\rangle,$$

we finally obtain (4.2.1d).
Part b. We now prove that our conditions are sufficient. It follows from (4.2.1c) and the proof of Proposition 3.11 that

$$\mathrm{Tr}(D^2\phi(x)a(x)) \leq -\langle D\phi(x)^\top, \sum_{j=1}^d Da^j(x)(aa^+)^j(x)\rangle,$$

for any smooth function ϕ such that $\max_{\mathcal{D}}\phi = \phi(x) \geq 0$. Moreover, after noticing that $D\phi(x)^\top \in \mathcal{N}_{\mathcal{D}}^1(x)$ (this is immediate from the Taylor expansion of ϕ around x), (4.2.1b)

yields

$$\int (\phi(x+\rho(x,z)) - \phi(x) + D\phi(x)\rho(x,z))\, F(dz) = \int (\phi(x+\rho(x,z)) - \phi(x))\, F(dz) + \int D\phi(x)\rho(x,z)F(dz).$$

In addition, it follows from (4.2.1a) that $\phi(x+\rho(x,z)) \leq \phi(x)$ for F-almost all z. Combining all the above with (4.2.1d) we finally get

$$\mathcal{L}\phi(x) \leq \langle D\phi(x)^\top, b(x) - \int \rho(x,z)F(dz) - \frac{1}{2}\sum_{j=1}^{d} Da^j(x)(aa^+)^j(x)\rangle \leq 0.$$

Therefore, $\mathcal{L}$ satisfies the positive maximum principle. In addition, since $\mathcal{L} : \mathcal{C}_c^\infty(\mathcal{D}) \mapsto \mathcal{C}_0(\mathcal{D})$ (see Lemma 4.3) and $\mathcal{C}_c^\infty(\mathcal{D})$ is dense in $\mathcal{C}_0(\mathcal{D})$, by [54, Theorem 4.5.4], there exists a càdlàg $(\mathcal{D}\cup\Delta)$-valued solution to the martingale problem for $\mathcal{L}$, where Δ denotes the one point compactification of $\mathcal{D}$. Δ is attained either by jump (killed by a potential) or by explosion. By the discussion preceding [29, Proposition 3.2], the process cannot jump to Δ. Moreover, the growth conditions (H_1) ensure that no explosion happens in finite time (see (4.4.3)). Hence Δ is never attained. We conclude by using the equivalence between martingale problems and stochastic differential equations, e.g. [81, Theorem 2.3]. □

4.3 Equivalent fomulation in the semimartingale framework

In this section, we provide an equivalent formulation of Theorem 4.2 in the semimartingale set-up which is more adapted to the construction of affine and polynomial jump-diffusions (see Remark 4.6 below). We stress once more that, by [53, 28], (4.1.1) is a very general formulation, equivalent to the semimartingale formulation (4.3.2) below (see also [76, Theorem 2.1.2]).

Let X denote a homogeneous diffusion with jumps in the sense of [77, Definition III.2.18] on a filtered probability space $(\widetilde{\Omega}, \widetilde{\mathcal{F}}, \widetilde{\mathbb{F}}, \widetilde{\mathbb{P}})$, i.e. its semimartingale characteristics $(\widetilde{B}, \widetilde{C}, \nu)$ are of the form

$$\widetilde{B}_t = \int_0^t \widetilde{b}(X_s)ds, \quad \widetilde{C}_t = \int_0^t \widetilde{a}(X_s)ds, \quad \nu(dt,dz) = dtK(X_t,dz), \tag{4.3.1}$$

with respect to a continuous truncation function h, i.e. h is bounded and equal to the identity on a neighborhood of 0. Here, $\widetilde{b} : \mathbb{R}^d \mapsto \mathbb{R}^d$, $\widetilde{a} : \mathbb{R}^d \mapsto \mathbb{S}_+^d$, K is a transition kernel from $\mathbb{R}^d$ into $\mathbb{R}^d \setminus \{0\}$ and

$$\widetilde{b}, \widetilde{a} \text{ and } \int f(z)\|z\|^2 K(.,dz) \text{ are continuous for any bounded continuous function } f. \quad (\widetilde{H}_\mathcal{C})$$

The triplet $(\widetilde{b}, \widetilde{a}, K)$ is called the differential characteristics of X. In addition we assume that there exist $\widetilde{q}, \widetilde{L} > 0$ such that

$$\int_{\{\|z\|>1\}} \|z\|^{\widetilde{q}} \ln\|z\| K(x,dz) \leq \widetilde{L}(1+\|x\|^{\widetilde{q}}), \tag{$\widetilde{H}_0$}$$

$$\|\widetilde{b}(x)\|^2 + \|\widetilde{a}(x)\| + \int \|z\|^2 K(x,dz) \leq \widetilde{L}(1+\|x\|^2), \tag{$\widetilde{H}_1$}$$

for all $x \in \mathbb{R}^d$. It follows that X is a locally square-integrable semimartingale (see [77, Definition II.2.27 and Proposition II.2.29]) and in particular X is a special semimartingale. Recall that ν is the compensated measure of the random jump measure μ of X. By [77, Theorem II.2.38], X admits the following canonical decomposition

$$X = X_0 + B + X^c + z * (\mu - \nu), \tag{4.3.2}$$

where X^c is a continuous local martingale with quadratic variation $\langle X^c \rangle_\cdot = \int_0^\cdot \widetilde{a}(X_s)ds$ and $B := \int_0^\cdot b(X_s)ds$, where $b := \widetilde{b} + \int (z - h(z))K(., dz)$. Finally, we assume that

$$\text{the restriction of } \widetilde{a} \text{ to } \mathcal{D} \text{ can be extended to a } C^{1,1}_{loc}(\mathbb{R}^d, \mathbb{S}^d) \text{ function,} \tag{$\widetilde{H}_2$}$$

and we denote by a this extended function.

We are now ready to state an equivalent formulation of Theorem 4.2 adapted to (4.3.2). We start by defining naturally the notion of stochastic invariance with respect to a semimartingale.

Definition 4.4 (*Stochastic invariance*)**.** A closed subset $\mathcal{D} \subset \mathbb{R}^d$ is said to be *stochastically invariant* with respect to the semimartingale (4.3.1) if, for all $x \in \mathcal{D}$, there exists a filtered probability space $(\Omega, \mathcal{F}, \mathbb{F} := (\mathcal{F}_t)_{t\geq 0}, \mathbb{P})$ supporting a semimartingale X with characteristics (4.3.1) starting at $X_0 = x$ and such that $X_t \in \mathcal{D}$ for all $t \geq 0$, $\mathbb{P}$-almost surely.

Theorem 4.5. *Let $\mathcal{D} \subset \mathbb{R}^d$ be closed. Under the continuity assumptions ($\widetilde{H}_\mathcal{C}$) and ($\widetilde{H}_0$)-($\widetilde{H}_2$), the set $\mathcal{D}$ is stochastically invariant with respect to the semimartingale (4.3.1) if and only if*

$$\begin{cases} \operatorname{supp} K(x, dz) \subset \mathcal{D} - x, & \text{(4.3.3a)} \\ \displaystyle\int |\langle u, z\rangle| K(x, dz) < \infty, & \text{(4.3.3b)} \\ a(x)u = 0, & \text{(4.3.3c)} \\ \displaystyle\langle u, b(x) - \int zK(x, dz) - \frac{1}{2}\sum_{j=1}^{d} Da^j(x)(aa^+)^j(x)\rangle \leq 0, & \text{(4.3.3d)} \end{cases}$$

for all $x \in \mathcal{D}$ and $u \in \mathcal{N}^1_\mathcal{D}(x)$.

Proof. Our proof is based on a (standard) representation of (4.3.2) in terms of (4.1.1). In this proof, we show the correspondence between the characteristics of (4.3.1) and the coefficients of (4.1.1), and between the assumptions and invariance conditions of the two settings. Then, Theorem 4.5 is deduced from a direct application of Theorem 4.2.

Part a. More precisely, let us fix F a σ-finite and infinite measure with no atom. By [28, Lemma 3.4] and the discussion preceding [28, Theorem 3.13], there exists a measurable function $\rho : \mathbb{R}^d \times \mathbb{R}^d \to \mathbb{R}^d \backslash \{0\}$ such that[3]

$$K(x, B) = \int \mathbb{1}_B(\rho(x, z))F(dz), \qquad \text{for all Borel sets } B. \tag{4.3.4}$$

Let us fix ρ for the rest of the proof, and recall that $b = \widetilde{b} + \int (z - h(z))K(., dz)$ and set $\sigma := \widetilde{a}^{\frac{1}{2}}$. We claim that the assumptions ($\widetilde{H}_\mathcal{C}$) and ($\widetilde{H}_0$)-($\widetilde{H}_2$) imply the assumptions ($H_\mathcal{C}$) and (H_0)-(H_2), and that the conditions (4.2.1a)-(4.2.1d) are equivalent to the conditions (4.3.3a)-(4.3.3d). To see this, recall that h is bounded and equal to the identity on a neighborhood of 0.

[3] There is a lot of freedom for ρ, see [28, Section 4] and [53, Theorem 6 and Corollary 7].

Hence $z \to (z-h(z))/(\|z\|^2 \mathbb{1}_{\{\|z\|\neq 0\}})$ is bounded and continuous (because it is equal to 0 on a neighborhood of 0). It follows, from $(\widetilde{H}_{\mathcal{C}})$ that b and σ are continuous. Therefore, combining the above with (4.3.4) and $(\widetilde{H}_0)$-$(\widetilde{H}_2)$ yields $(H_{\mathcal{C}})$ and (H_0)-(H_2). Finally, one easily deduce from (4.3.4) that (4.2.1a)-(4.2.1d) are equivalent to (4.3.3a)-(4.3.3d) since $\int g(z)K(x,dz) = \int g(\rho(x,z))F(dz)$, for any measurable function g.

Part b. To see that the conditions (4.3.3a)-(4.3.3d) of Theorem 4.5 are sufficient, it suffices to apply Theorem 4.2, whose assumptions and conditions are satisfied by Part a. above. Namely, under the conditions (4.3.3a)-(4.3.3d), (4.1.1) admits a $\mathcal{D}$-valued weak solution, which is also a semimartingale with characteristics (4.3.1).

Part c. We now prove that the conditions (4.3.3a)-(4.3.3d) are necessary. Assume that $\mathcal{D}$ is stochastically invariant with respect to (4.3.1). Fix $(\Omega, \mathcal{F}, \mathbb{F} := (\mathcal{F}_t)_{t\geq 0}, \mathbb{P})$ supporting a semimartingale X with characteristics (4.3.1) starting at $X_0 = x \in \mathcal{D}$ such that $\mathbb{P}(\{X_t \in \mathcal{D}, \forall t \geq 0\}) = 1$. By [76, Theorem 2.1.2], there exists a filtered extension $(\widetilde{\Omega}, \widetilde{\mathcal{F}}, \widetilde{\mathbb{F}} := (\widetilde{\mathcal{F}}_t)_{t\geq 0}, \widetilde{\mathbb{P}})$ supporting a d-dimensional Brownian motion W and a Poisson random measure μ with compensator $dt \otimes F(dz)$ such that X solves

$$X_t = x + \int_0^t \widetilde{b}(X_s)ds + \int_0^t \widetilde{\sigma}_s dW_s + (\delta \mathbb{1}_{\{\|\delta\|\leq 1\}}) * (\mu - dtF(dz)) + (\delta \mathbb{1}_{\{\|\delta\|>1\}}) * \mu, \quad (4.3.5)$$

where $(\widetilde{\sigma}, \delta)$ are such that $\widetilde{\sigma}_t(\widetilde{\omega})\widetilde{\sigma}_t(\widetilde{\omega})^\top = \widetilde{a}(X_t(\widetilde{\omega}))$ and $K(X_t(\widetilde{\omega}), B) = \int \mathbb{1}_B(\delta(\widetilde{\omega}, t, z))F(dz)$, for all Borel sets B, for all $t \geq 0$, for $\widetilde{\mathbb{P}}$-almost all $\widetilde{\omega}$. In view of (4.3.4),

$$\int \mathbb{1}_B(\rho(X.\widetilde{\omega}, z))F(dz) = K(X.(\widetilde{\omega}), B) = \int \mathbb{1}_B(\delta(\widetilde{\omega}, \cdot, z))F(dz),$$

for all Borel sets B, for $\widetilde{\mathbb{P}}$-almost all $\widetilde{\omega}$. Hence, $\delta = \rho(X., \cdot)$ $F \otimes \mathbb{P}$ almost everywhere. Similarly, $\widetilde{\sigma}$ can be taken to be equal to the square root of $\widetilde{a}$ (see Remark 3.1). Thus, (4.3.5) can be written in the form (4.1.1) with $(b, \sigma := \widetilde{a}^{\frac{1}{2}}, \rho)$. Moreover, $\widetilde{\mathbb{P}}(\{X_t \in \mathcal{D}, \forall t \geq 0\} = \mathbb{P}(\{X_t \in \mathcal{D}, \forall t \geq 0\}) = 1$, by the discussion following [76, Equation (2.1.26)]. In view of Part a., Theorem 4.2 implies that (4.2.1a)-(4.2.1d) should hold, so that (4.3.3a)-(4.3.3d) must be satisfied. □

Remark 4.6. As already mentioned above, in the presence of jumps, the semimartingale formulation given in Theorem 4.5 is more adapted to affine and polynomial processes than Theorem 4.2. In fact, affine (resp. polynomial) jump-diffusions are characterized by an affine (resp. polynomial) dependence of their triplet $(\widetilde{b}, \widetilde{a}, K)$ (e.g. [110, Definition 4.2]). Inspecting the identity in (4.3.4), it is not clear how this property translates to ρ.

4.4 Technical lemmas

For completeness, we provide in the sequel some technical lemmas with their proofs. They are either standard or minor modifications of already known results.

The generalized Itô's lemma derived in Lemma 3.8 can easily be extended to account for jumps in the following way.

Lemma 4.7. *Assume that σ is continuous and that there exists a solution X to (4.1.1). Let $f \in \mathcal{C}_c^{1,1}(\mathbb{R}^d, \mathbb{R})$. Then, there exists an adapted bounded process η such that*

$$f(X_t) = f(x) + \int_0^t \widetilde{\eta}_s ds + \int_0^t (Df\sigma)(X_s)dW_s + (f(X_{s-} + \rho(X_{s-}, z)) - f(X_{s-})) * (\mu - dtF(dz)),$$

for all $t \geq 0$, *with* $\widetilde{\eta}_s = (Dfb)(X_s) + \eta_s + \int \left(f(X_s + \rho(X_s, z)) - f(X_s) - Df(X_s)\rho(X_s, z)\right) F(dz)$.

Proof. Since $f \in \mathcal{C}^{1,1}$ has a compact support, we can find a sequence $(f_n)_n$ in $\mathcal{C}^\infty$ with compact support (uniformly) and a constant $K > 0$ such that

(i) $\|D^2 f_n\| \leq K$,

(ii) $\|f_n - f\| + \|Df_n - Df\| \leq \frac{K}{n}$,

for all $n \geq 1$. This is obtained by considering a simple mollification of f. Set $\widetilde{\mu} := \mu - dtF(dz)$. Since f_n is twice differentiable and bounded, Itô's formula [77, Theorem I.4.57] yields

$$f_n(X_t) = f_n(x) + \int_0^t Df_n(X_s)\sigma(X_s)dW_s + (f_n(X_{s-} + \rho(X_{s-}, z)) - f_n(X_{s-})) * \widetilde{\mu} + \int_0^t \left((Df_n b)(X_s) + \eta_s^n + \int \left(f_n(X_s + \rho(X_s, z)) - f_n(X_s) - Df_n(X_s)\rho(X_s, z)\right) F(dz)\right) ds$$

in which $\eta^n := \frac{1}{2}\mathrm{Tr}[D^2 f_n \sigma\sigma^\top](X)$. Since $\sigma\sigma^\top$ is continuous, (i) above implies that $(\eta^n)_n$ is uniformly bounded in $L^\infty(dt \times d\mathbb{P})$. By [43, Theorem 1.3], there exists $(\hat{\eta}^n) \in \mathrm{Conv}(\eta^k, k \geq n)$ such that $\hat{\eta}^n \to \eta$ $dt \otimes d\mathbb{P}$ almost surely. Let $N_n \geq 0$ and $(\lambda_k^n)_{n \leq k \leq N_n} \subset [0, 1]$ be such that $\hat{\eta}^n = \sum_{k=n}^{N_n} \lambda_k^n \eta^k$ and $\sum_{k=n}^{N_n} \lambda_k^n = 1$. Set $\hat{f}_n := \sum_{k=n}^{N_n} \lambda_k^n f_k$. Then,

$$\hat{f}_n(X_t) = \hat{f}_n(x) + \int_0^t \widetilde{\eta}^n ds + \int_0^t D\hat{f}_n(X_s)\sigma(X_s)dW_s + \left(\hat{f}_n(X_{s-} + \rho(X_{s-}, z)) - \hat{f}_n(X_{s-})\right) * \widetilde{\mu}, \tag{4.4.1}$$

in which $\widetilde{\eta}^n := (D\hat{f}_n b)(X_s) + \hat{\eta}_s^n + \int \left(\hat{f}_n(X_s + \rho(X_s, z)) - \hat{f}_n(X_s) - D\hat{f}_n(X_s)\rho(X_s, z)\right) F(dz)$. By dominated convergence, $\int_0^t \hat{\eta}_s^n ds$ converges a.s. to $\int_0^t \eta_s ds$. Moreover, (ii) implies that

$$\|\hat{f}_n(X_t) - f(X_t)\| \leq \sum_{k=n}^{N_n} \lambda_k^n \|\hat{f}_k(X_t) - f(X_t)\| \leq \sum_{k=n}^{N_n} \lambda_k^n \frac{K}{k} \leq \frac{K}{n},$$

so that $\hat{f}_n(X_t)$ converges a.s. to $f(X_t)$. Similarly,

$$\int_0^t \widetilde{\eta}_s^n ds \to \int_0^t \widetilde{\eta}_s ds, \quad \int_0^t D\hat{f}_n(X_s)\sigma(X_s)dW_s \to \int_0^t Df(X_s)\sigma(X_s)dW_s,$$
$$\left(\hat{f}_n(X_{s-} + \rho(X_{s-}, z)) - \hat{f}_n(X_{s-})\right) * \widetilde{\mu} \to (f(X_{s-} + \rho(X_{s-}, z)) - f(X_{s-})) * \widetilde{\mu},$$

in $L^2(\Omega, \mathcal{F}, \mathbb{P})$ as $n \to \infty$, and therefore a.s. after possibly considering a subsequence. It thus remains to send $n \to \infty$ in (4.4.1) to obtain the required result. □

We also used the following elementary lemma which extends [114, Lemma 5.4] to account for jumps (see also [88, Corollaries 2 and 3 of Theorem 5.13]).

Lemma 4.8. *Let* $B, B^\perp$ *denote two independent Brownian motions and* μ *a Poisson random measure on* $\mathbb{R}_+ \times \mathbb{R}^d$ *with compensator* $dt \otimes F$ *on a filtered probability space* $(\Omega, \mathcal{F}, (\mathcal{F}_t)_{t \geq 0}, \mathbb{P})$. *Let* $(\gamma_s)_{s \geq 0}$ *be an adapted square integrable process and* $\xi : \mathbb{R}_+ \times \mathbb{R}^d \mapsto \mathbb{R}^d$ *be a predictable process such that* $\mathbb{E}\left[\int_0^t \int \|\xi(s, z)\|^2 F(dz)ds\right] < \infty$, *for all* $t \geq 0$. *Define the sub-filtration*

$\mathcal{F}_t^B = \sigma\{B_s, s \le t\} \subset \mathcal{F}_t$ *and denote by* $\widetilde{\mu} = \mu - dtF(dz)$. *Then* $\mathbb{P}-a.s.$, *for all* $t \ge 0$,

$$\mathbb{E}_{\mathcal{F}_t^B}\left[\int_0^t \gamma_s dB_s\right] = \int_0^t \mathbb{E}_{\mathcal{F}_s^B}\left[\gamma_s\right] dB_s, \quad \mathbb{E}_{\mathcal{F}_t^B}\left[\int_0^t \gamma_s dB_s^{\perp}\right] = \mathbb{E}_{\mathcal{F}_t^B}\left[\xi * \widetilde{\mu}\right] = 0.$$

Moreover, it holds similarly for any integrable adapted process θ *that*

$$\mathbb{E}_{\mathcal{F}_t^B}\left[\int_0^t \theta_s ds\right] = \int_0^t \mathbb{E}_{\mathcal{F}_s^B}\left[\theta_s\right] ds.$$

Proof. We sketch the proof for the jump integral. See [114, Lemma 5.4] for the other identities. Let ξ be simple and predicable, *i.e.* $\xi(s,z) = \sum_{i=1}^n \xi_i \mathbb{1}_{(t_i,t_{i+1}]}(s)\mathbb{1}_{A_i}(z)$, in which ξ_i is bounded and $\mathcal{F}_{t_i}$-measurable, $(t_i)_{1\le i\le n}$ a subdivision of $[0,t]$ and $A_i \subset \mathbb{R}^d$ Borel sets such that $F(A_i) < \infty$. We can write $\mathcal{F}_t^B = \mathcal{F}_{t_i}^B \vee \mathcal{F}_{t_i,t}^B$ where $\mathcal{F}_{t_i,t}^B := \sigma(B_s - B_{t_i}, t_i \le s \le t)$. It follows from [74, Theorem II.6.3] that μ and B are independent and

$$\begin{aligned}
\mathbb{E}_{\mathcal{F}_t^B}\left[\xi * \widetilde{\mu}\right] &= \sum_{i=1}^n \mathbb{E}_{\mathcal{F}_{t_i}^B \vee \mathcal{F}_{t_i,t}^B}\left[\xi_i \widetilde{\mu}\left((t_i,t_{i+1}]\times A_i\right)\right] \\
&= \sum_{i=1}^n \mathbb{E}\left[\mathbb{E}\left[\xi_i \widetilde{\mu}\left((t_i,t_{i+1}]\times A_i\right) \mid \mathcal{F}_{t_i} \vee \mathcal{F}_{t_i,t}^B\right] \mid \mathcal{F}_{t_i}^B \vee \mathcal{F}_{t_i,t}^B\right] \\
&= \sum_{i=1}^n \mathbb{E}\left[\xi_i \mathbb{E}\left[\widetilde{\mu}\left((t_i,t_{i+1}]\times A_i\right) \mid \mathcal{F}_{t_i} \vee \mathcal{F}_{t_i,t}^B\right] \mid \mathcal{F}_{t_i}^B \vee \mathcal{F}_{t_i,t}^B\right] \\
&= 0.
\end{aligned}$$

For general ξ, the result follows from Itô's isometry and the fact that simple processes are dense in $L^2(dt \otimes F)$ (see [8, Lemma 4.1.4]). □

Proposition 4.9. *Let* X *denote a weak solution of* (4.1.1) *starting at* x. *Under the growth conditions* (H_1), *there exists* $M_{x,L}^1 > 0$ *such that the following moment estimates hold:*

$$\mathbb{E}\left[\sup_{s\le t}\|X_s\|^2\right] \le 4\left(\|x\|^2 + Lt(t+8)\right)e^{4Lt(t+8)}, \qquad t \ge 0, \tag{4.4.2}$$

$$\mathbb{E}\left[\|X_t - X_s\|^2\right] \le M_{x,L}^1 |t-s|, \qquad s,t \le 1. \tag{4.4.3}$$

Proof. Let $\tau_n = \inf\{t \ge 0 : \|X_t\| \ge n \text{ or } \|X_{t-}\| \ge n\} \wedge t$. Set $g_t^n := \mathbb{E}\left[\sup_{s\le t}\|X_s\|^2 \mathbb{1}_{\{s<\tau_n\}}\right]$. By convexity of $y \mapsto y^2$, $(a+b+c+d)^2 = 16(\frac{a+b+c+d}{4})^2 \le 4(a^2+b^2+c^2+d^2)$. Combined with Cauchy–Schwarz and Burkholder-Davis-Gundy inequalities, we get for all $u \le t$

$$\begin{aligned}
g_u^n \le 4\|x\|^2 &+ 4t\int_0^u \mathbb{E}\left[\|b(X_s\mathbb{1}_{\{s<\tau_n\}})\|^2\right] ds \\
&+ 16\int_0^u \mathbb{E}\left[\|C(X_s\mathbb{1}_{\{s<\tau_n\}})\|\right] ds + 16\int_{[0,u]\times\mathbb{R}^d} \mathbb{E}\left[\|\rho(X_{s-}\mathbb{1}_{\{s<\tau_n\}},z)\|^2\right] F(dz)ds.
\end{aligned}$$

The growth conditions (H_1) now yield

$$g_u^n \le 4\left(\|x\|^2 + Lt(t+8) + L(t+8)\int_0^u g_s^n ds\right), \quad u \le t.$$

Finally, (4.4.2) follows from Grönwall's Lemma and monotone convergence after sending $n \to \infty$. Moreover, for all $s,t \le 1$, by Cauchy-Schwarz inequality, Itô's isometry and (H_1)

we have

$$\begin{aligned}\mathbb{E}\left[\|X_t - X_s\|^2\right] &\leq 3\left(|t-s|\int_s^t \mathbb{E}\|b(X_r)\|^2 dr + \int_s^t \mathbb{E}[\|C(X_r)\| + \int_{\mathbb{R}^d} \|\rho(X_{r-}, z)\|^2 F(dz)]dr\right)\\ &\leq 3\left(L|t-s|^2(1+g_1) + L|t-s|(1+g_1)\right)\\ &\leq 6L(1+g_1)|t-s|.\end{aligned}$$

Hence, (4.4.3) follows from (4.4.2). □

Part II

Stochastic Volterra equations

Chapter 5

Affine Volterra processes

Summary

We introduce affine Volterra processes, defined as solutions of certain stochastic convolution equations with affine coefficients. Classical affine diffusions constitute a special case, but affine Volterra processes are neither semimartingales, nor Markov processes in general. We provide explicit exponential-affine representations of the Fourier–Laplace functional in terms of the solution of an associated system of deterministic integral equations, extending well-known formulas for classical affine diffusions. For specific state spaces, we prove existence, uniqueness, and invariance properties of solutions of the corresponding stochastic convolution equations. Our arguments avoid infinite-dimensional stochastic analysis as well as stochastic integration with respect to non-semimartingales, relying instead on tools from the theory of finite-dimensional deterministic convolution equations. Our findings generalize and clarify recent results in the literature on rough volatility models in finance.

Based on [6]: Abi Jaber, E., Larsson, M., & Pulido, S. (2017) Affine Volterra processes. *In revision - Annals of Applied Probability.*

5.1 Introduction

We study a class of d-dimensional stochastic convolution equations of the form

$$X_t = X_0 + \int_0^t K(t-s)b(X_s)ds + \int_0^t K(t-s)\sigma(X_s)dW_s, \tag{5.1.1}$$

where W is a multi-dimensional Brownian motion, and the convolution kernel K and coefficients b and σ satisfy regularity and integrability conditions that are discussed in detail after this introduction. We refer to equations of the form (5.1.1) as *stochastic Volterra equations* (of convolution type), and their solutions are always understood to be adapted processes defined on some stochastic basis $(\Omega, \mathcal{F}, (\mathcal{F}_t)_{t\geq 0}, \mathbb{P})$ satisfying the usual conditions. Stochastic Volterra equations have been studied by numerous authors; see e.g. [16, 17, 100, 99, 117, 96]

among many others. In Theorem 5.10 and Theorem 5.12 we provide new existence results for (5.1.1) under weak conditions on the kernel and coefficients.

We are chiefly interested in the situation where $a(x) = \sigma(x)\sigma(x)^\top$ and $b(x)$ are affine of the form

$$\begin{aligned} a(x) &= A^0 + x_1 A^1 + \cdots + x_d A^d \\ b(x) &= b^0 + x_1 b^1 + \cdots + x_d b^d, \end{aligned} \tag{5.1.2}$$

for some d-dimensional symmetric matrices A^i and vectors b^i. In this case we refer to solutions of (5.1.1) as *affine Volterra processes*. *Affine diffusions*, as studied in [48], are particular examples of affine Volterra processes of the form (5.1.1) where the convolution kernel $K \equiv \mathrm{id}$ is constant and equal to the d-dimensional identity matrix.

Stochastic models using classical affine diffusions are tractable because their Fourier–Laplace transform has a simple form. It can be written as an exponential-affine function of the initial state, in terms of the solution of a system of ordinary differential equations, known as the Riccati equations, determined by the affine maps (5.1.2). More precisely, let X be an affine diffusion of the form (5.1.1) with $K \equiv \mathrm{id}$. Then, given a d-dimensional row vector u and under suitable integrability conditions, we have

$$\mathbb{E}\left[\exp\left(uX_T\right) \;\middle|\; \mathcal{F}_t\right] = \exp\left(\phi(T-t) + \psi(T-t)X_t\right), \tag{5.1.3}$$

where the real-valued function ϕ and row-vector-valued function ψ satisfy the Riccati equations

$$\begin{aligned} \phi(t) &= \int_0^t \left(\psi(s)b_0 + \frac{1}{2}\psi(s)A_0\psi(s)^\top\right) ds \\ \psi(t) &= u + \int_0^t \left(\psi(s)B + \frac{1}{2}A(\psi(s))\right) ds, \end{aligned}$$

with $A(u) = (uA^1u^\top, \ldots, uA^du^\top)$ and $B = (b^1 \;\cdots\; b^d)$. Alternatively, using the variation of constants formula on X and ψ, one can write the Fourier–Laplace transform as

$$\mathbb{E}\left[\exp\left(uX_T\right) \;\middle|\; \mathcal{F}_t\right] = \exp\left(\mathbb{E}[uX_T \mid \mathcal{F}_t] + \frac{1}{2}\int_t^T \psi(T-s)a(\mathbb{E}[X_s \mid \mathcal{F}_t])\psi(T-s)^\top ds\right). \tag{5.1.4}$$

For more general kernels K, affine Volterra processes are typically neither semimartingales, nor Markov processes. Therefore one cannot expect a formula like (5.1.3) to hold in general. However, we show in Theorem 5.16 below that, remarkably, (5.1.4) does continue to hold, where now the function ψ solves the *Riccati–Volterra equation*

$$\psi(t) = uK(t) + \int_0^t \left(\psi(s)B + \frac{1}{2}A(\psi(s))\right) K(t-s)\, ds. \tag{5.1.5}$$

Furthermore, it is possible to express (5.1.4) in a form that is exponential-affine in the past trajectory $\{X_s,\, s \le t\}$. This is done in Theorem 5.18.

For the state spaces $\mathbb{R}^d$, $\mathbb{R}^d_+$, and $\mathbb{R} \times \mathbb{R}_+$, corresponding to the *Volterra Ornstein–Uhlenbeck*, *Volterra square-root*, and *Volterra Heston* models, we establish existence and uniqueness of global solutions of both the stochastic equation (5.1.1) and the associated Riccati–Volterra equation (5.1.5), under general parameter restrictions. For the state spaces $\mathbb{R}^d_+$ and $\mathbb{R} \times \mathbb{R}_+$, which are treated in Theorem 5.21 and Theorem 5.24, this involves rather delicate invariance

properties for these equations. While standard martingale and stochastic calculus arguments play an important role in several places, the key tools that allow us to handle the lack of Markov and semimartingale structure are the *resolvents of first and second kind* associated with the convolution kernel K. Let us emphasize in particular that no stochastic integration with respect to non-semimartingales is needed. Furthermore, by performing the analysis on the level of finite-dimensional integral equations, we avoid the infinite-dimensional analysis used, for instance, in [96]. We also circumvent the need to study scaling limits of Hawkes processes as in [51, 49, 52].

Our motivation for considering affine Volterra processes comes from applications in financial modeling. Classical affine processes arguably constitute the most popular framework for building tractable multi-factor models in finance. They have been used to model a vast range of risk factors such as credit and liquidity factors, inflation and other macro-economic factors, equity factors, and factors driving the evolution of interest rates. In particular, affine stochastic volatility models, such as the Heston model [72], are very popular.

However, a growing body of empirical research indicates that volatility fluctuates more rapidly than Brownian motion, which is inconsistent with standard semimartingale affine models. Fractional volatility models such as those in [70, 66, 14, 51, 15] have emerged as compelling alternatives, although tractability can be a challenge for these non-Markovian, non-semimartingales models. Nonetheless, it is shown in [51, 52] that there exist fractional adaptations of the Heston model where the Fourier–Laplace transform can be found explicitly, modulo the solution of a specific *fractional Riccati equation*. These models are of the affine Volterra type (5.1.1) involving singular kernels proportional to $t^{\alpha-1}$. Our framework subsumes and extends these examples.

The chapter is structured as follows. Section 5.2 covers preliminaries on convolutions and their resolvents, and in particular develops the necessary stochastic calculus. Section 5.3 gives existence theorems for stochastic Volterra equations on $\mathbb{R}^d$ and $\mathbb{R}^d_+$. Section 5.4 introduces affine Volterra processes on general state spaces and develops the exponential-affine transform formula. Section 5.5 contains detailed discussions for the state spaces $\mathbb{R}^d$, $\mathbb{R}^d_+$, and $\mathbb{R} \times \mathbb{R}_+$, which correspond to the Volterra Ornstein–Uhlenbeck, Volterra square-root, and Volterra Heston models, respectively. Additional proofs and supporting results are presented in Section 5.6. Our basic reference for the deterministic theory of Volterra equations is the excellent book [69].

Notation

Throughout the chapter we view elements of $\mathbb{R}^m$ and $\mathbb{C}^m = \mathbb{R}^m + \mathrm{i}\mathbb{R}^m$ as column vectors, while elements of the dual spaces $(\mathbb{R}^m)^*$ and $(\mathbb{C}^m)^*$ are viewed as row vectors. For any matrix A with complex entries, $A^\top$ denotes the (ordinary, not conjugate) transpose of A. The identity matrix is written id. The symbol $|\cdot|$ is used to denote the Euclidean norm on $\mathbb{C}^m$ and $(\mathbb{C}^m)^*$, as well as the operator norm on $\mathbb{R}^{m\times n}$. The shift operator Δ_h with $h \geq 0$, maps any function f on $\mathbb{R}_+$ to the function $\Delta_h f$ given by

$$\Delta_h f(t) = f(t+h).$$

If the function f on $\mathbb{R}_+$ is right-continuous and of locally bounded variation, the measure induced by its distribution derivative is denoted df, so that $f(t) = f(0) + \int_{[0,t]} df(s)$ for all $t \geq 0$. By convention, df does not charge $\{0\}$.

5.2 Stochastic calculus of convolutions and resolvents

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

$$(K * L)(t) = \int_{[0,t]} K(t-s)L(ds), \qquad (L * K)(t) = \int_{[0,t]} L(ds)K(t-s) \tag{5.2.1}$$

for $t > 0$ whenever these expressions are well-defined, and extended to $t = 0$ by right-continuity when possible. We allow K and L to be matrix-valued, in which case $K * L$ and $L * K$ may not both be defined (e.g. due to incompatible matrix dimensions), or differ from each other even if they are defined (e.g. if K and L take values among non-commuting square matrices). If F is a function on $\mathbb{R}_+$, we write $K * F = K * (F dt)$, that is,

$$(K * F)(t) = \int_0^t K(t-s)F(s)ds. \tag{5.2.2}$$

Further details can be found in [69], see in particular Definitions 2.2.1 and 3.2.1, as well as Theorems 2.2.2 and 3.6.1 for a number of properties of convolutions. In particular, if $K \in L^1_{\mathrm{loc}}(\mathbb{R}_+)$ and F is continuous, then $K * F$ is again continuous.

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

$$(K * dM)_t = \int_0^t K(t-s)dM_s \tag{5.2.3}$$

is well-defined as an Itô integral for every $t \geq 0$ such that $\int_0^t |K(t-s)|^2 d\,\mathrm{Tr}\langle M\rangle_s < \infty$. In particular, if $K \in L^2_{\mathrm{loc}}(\mathbb{R}_+)$ and $\langle M\rangle_s = \int_0^s a_u du$ for some locally bounded process a, then (5.2.3) is well-defined for every $t \geq 0$. We always choose a version that is jointly measurable in (t,ω). Just like (5.2.1)–(5.2.2), the convolution (5.2.3) is associative, as the following result shows.

Lemma 5.1. *Let $K \in L^2_{\mathrm{loc}}(\mathbb{R}_+, \mathbb{R}^{m\times d})$ and let L be an $\mathbb{R}^{n\times m}$-valued measure on $\mathbb{R}_+$ of locally bounded variation. Let M be a d-dimensional continuous local martingale with $\langle M\rangle_t = \int_0^t a_s ds$, $t \geq 0$, for some locally bounded process a. Then*

$$(L * (K * dM))_t = ((L * K) * dM)_t \tag{5.2.4}$$

for every $t \geq 0$. In particular, taking $F \in L^1_{\mathrm{loc}}(\mathbb{R}_+)$ we may apply (5.2.4) with $L(dt) = F dt$ to obtain $(F * (K * dM))_t = ((F * K) * dM)_t$.

Proof. By linearity it suffices to take $d = m = n = 1$ and L a locally finite positive measure. In this case,

$$(L * (K * dM))_t = \int_0^t \left(\int_0^t \mathbb{1}_{\{u<t-s\}} K(t-s-u)dM_u \right) L(ds).$$

Since

$$\int_0^t \left(\int_0^t \mathbb{1}_{\{u<t-s\}} K(t-s-u)^2 d\langle M\rangle_u \right)^{1/2} L(ds) \leq \max_{0\leq s\leq t} |a_s|^{1/2} \|K\|_{L^2(0,t)} L([0,t]),$$

which is finite almost surely, the stochastic Fubini theorem, see [111, Theorem 2.2], yields

$$(L*(K*dM))_t = \int_0^t \left(\int_0^t \mathbb{1}_{\{u<t-s\}} K(t-s-u)L(ds) \right) dM_u = ((L*K)*dM)_t,$$

as required. □

Under additional assumptions on the kernel K one can find a version of the convolution (5.2.3) that is continuous in t. We will use the following condition:

$$\begin{gathered} K \in L^2_{\mathrm{loc}}(\mathbb{R}_+, \mathbb{R}) \text{ and there is } \gamma \in (0,2] \text{ such that } \int_0^h K(t)^2 dt = O(h^\gamma) \\ \text{and } \int_0^T (K(t+h)-K(t))^2 dt = O(h^\gamma) \text{ for every } T < \infty. \end{gathered} \tag{5.2.5}$$

Remark 5.2. Other conditions than (5.2.5) have appeared in the literature. In [42], $dM = \sigma dW$ is defined on the Wiener space with coordinate process W under the requirement that $F \mapsto K * F$ is be continuous from certain L^p spaces to appropriate Besov spaces. In [95], K is assumed to be a function of smooth variation and M a semimartingale. See also [112, Theorem 1.3].

Example 5.3. *Let us list some examples of kernels that satisfy (5.2.5):*

1. *Locally Lipschitz kernels K clearly satisfy (5.2.5) with $\gamma = 1$.*
2. *The fractional kernel $K(t) = t^{\alpha-1}$ with $\alpha \in (\frac{1}{2}, 1)$ satisfies (5.2.5) with $\gamma = 2\alpha - 1$. Indeed, it is locally square integrable, and we have $\int_0^h K(t)dt = h^{2\alpha-1}/(2\alpha-1)$ as well as*
$$\int_0^T (K(t+h)-K(t))^2 dt \le h^{2\alpha-1} \int_0^\infty \left((t+1)^{\alpha-1} - t^{\alpha-1} \right)^2 dt,$$
where the constant on the right-hand side is bounded by $\frac{1}{2\alpha-1} + \frac{1}{3-2\alpha}$. Note that the case $\alpha \ge 1$ falls in the locally Lipschitz category mentioned previously.
3. *If K_1 and K_2 satisfy (5.2.5), then so does $K_1 + K_2$.*
4. *If K_1 satisfies (5.2.5) and K_2 is locally Lipschitz, then $K = K_1 K_2$ satisfies (5.2.5) with the same γ. Indeed, letting $\|K_2^2\|_{\infty,T}$ denote the maximum of K_2^2 over $[0,T]$ and $\mathrm{Lip}_T(K_2)$ the best Lipschitz constant on $[0,T]$, we have*
$$\int_0^h K(t)^2 dt \le \|K_2^2\|_{\infty,h} \int_0^h K_1^2(t) dt = O(h^\gamma)$$
and
$$\begin{aligned} \int_0^T (K(t+h)-K(t))^2 dt \le{}& 2\|K_2^2\|_{\infty,T+h} \int_0^T (K_1(t+h)-K_1(t))^2 dt \\ &+ 2\|K_1\|^2_{L^2(0,T)} \mathrm{Lip}_{T+h}(K_2)^2 h^2 \\ ={}& O(h^\gamma). \end{aligned}$$
5. *If K satisfies (5.2.5) and $f \in L^2_{\mathrm{loc}}(\mathbb{R}_+)$, then $f * K$ satisfies (5.2.5) with the same γ. Indeed, Young's inequality gives $\int_0^h (f*K)(t)^2 dt \le \|f\|^2_{L^1(0,h)} \|K\|^2_{L^2(0,h)} = O(h^\gamma)$ and,*

using also the Cauchy–Schwarz inequality,

$$\begin{aligned}\int_0^T ((f*K)(t+h)-(f*K)(t))^2 dt &\le 2T\|f\|^2_{L^2(0,T+h)}\|K\|^2_{L^2(0,h)}\\ &\quad + 2\|f\|^2_{L^1(0,h)}\|\Delta_h K - K\|^2_{L^2(0,h)}\\ &= O(h^\gamma).\end{aligned}$$

6. *If K satisfies (5.2.5) and is locally bounded on $(0,\infty)$, then $\Delta_\eta K$ satisfies (5.2.5) for any $\eta > 0$. Indeed, local boundedness gives $\|\Delta_\eta K\|^2_{L^2(0,h)} = O(h)$ and it is immediate that $\int_0^T (\Delta_\eta K(t+h) - \Delta_\eta K(t))^2 dt \le \int_0^{T+\eta} (K(t+h)-K(t))^2 dt = O(h^\gamma)$.*

7. *By combining the above examples we find that, for instance, exponentially damped and possibly singular kernels like the Gamma kernel $K(t) = t^{\alpha-1}e^{-\beta t}$ for $\alpha > \frac{1}{2}$ and $\beta \ge 0$ satisfy (5.2.5).*

Lemma 5.4. *Assume K satisfies (5.2.5) and consider a process $X = K*(bdt+dM)$, where b is a predictable process and M is a continuous local martingale with $\langle M\rangle_t = \int_0^t a_s ds$ for some predictable process a. Let $T \ge 0$ and $p > 2/\gamma$ be such that $\sup_{t\le T}\mathbb{E}[|a_t|^{p/2} + |b_t|^p]$ is finite. Then X admits a version which is Hölder continuous on $[0,T]$ of any order $\alpha < \gamma/2 - 1/p$. Denoting this version again by X, one has*

$$\mathbb{E}\left[\left(\sup_{0\le s<t\le T}\frac{|X_t - X_s|}{|t-s|^\alpha}\right)^p\right] \le c \sup_{t\le T}\mathbb{E}[|a_t|^{p/2} + |b_t|^p] \tag{5.2.6}$$

for all $\alpha \in [0, \gamma/2 - 1/p)$, where c is a constant that only depends on p, K, and T. As a consequence, if a and b are locally bounded, then X admits a version which is Hölder continuous of any order $\alpha < \gamma/2$.

Proof. For any $p \ge 2$ and any $s < t \le T < \infty$ we have

$$\begin{aligned}|X_t - X_s|^p &\le 4^{p-1}\left|\int_s^t K(t-u)b_u du\right|^p + 4^{p-1}\left|\int_0^s (K(t-u)-K(s-u))\, b_u du\right|^p\\ &\quad + 4^{p-1}\left|\int_s^t K(t-u) dM_u\right|^p + 4^{p-1}\left|\int_0^s (K(t-u)-K(s-u))\, dM_u\right|^p\\ &= 4^{p-1}(\mathbf{I}+\mathbf{II}+\mathbf{III}+\mathbf{IV}).\end{aligned}$$

Jensen's inequality applied twice yields $\mathbf{I} \le (t-s)^{p/2}\,\|K\|^{p-2}_{L^2(s,t)}\int_s^t |b_u|^p\, K(t-u)^2 du$. Taking expectations and changing variables we obtain

$$\mathbb{E}[\mathbf{I}] \le (t-s)^{p/2}\left(\int_0^{t-s} K(u)^2 du\right)^{p/2}\sup_{u\le T}\mathbb{E}[|b_u|^p]. \tag{5.2.7}$$

In a similar manner,

$$\mathbb{E}[\mathbf{II}] \le T^{p/2}\left(\int_0^s (K(u+t-s)-K(u))^2 du\right)^{p/2}\sup_{u\le T}\mathbb{E}[|b_u|^p]. \tag{5.2.8}$$

Analogous calculations relying also on the BDG inequalities applied to the continuous local martingale $\{\int_0^r K(t-u)dM_u \colon r \in [0,t]\}$ yield

$$\begin{aligned}\mathbb{E}\left[\mathbf{III}\right] &\leq C_p\,\mathbb{E}\left[\left(\int_s^t K(t-u)^2\, a_u\, du\right)^{p/2}\right] \\ &\leq C_p\left(\int_0^{t-s} K(u)^2 du\right)^{p/2} \sup_{u\leq T}\mathbb{E}[|a_u|^{p/2}]\end{aligned} \tag{5.2.9}$$

and

$$\mathbb{E}\left[\mathbf{IV}\right] \leq C_p\left(\int_0^s (K(u+t-s)-K(u))^2 du\right)^{p/2} \sup_{u\leq T}\mathbb{E}[|a_u|^{p/2}]. \tag{5.2.10}$$

Combining (5.2.7)–(5.2.10) with (5.2.5) leads to

$$\mathbb{E}\left[|X_t - X_s|^p\right] \leq c' \sup_{u\leq T}\mathbb{E}[|a_u|^{p/2} + |b_u|^p]\,(t-s)^{\gamma p/2},$$

where c' is a constant that only depends on p, K, and T, but not on s or t. Existence of a continuous version as well as the bound (5.2.6) now follow from the Kolmogorov continuity theorem; see [101, Theorem I.2.1].

Finally, if a and b are locally bounded, consider stopping times $\tau_n \to \infty$ such that a and b are bounded on $[\![0,\tau_n]\!]$. The process $X^n = K * (b\mathbb{1}_{[\![0,\tau_n]\!]}dt + a\mathbb{1}_{[\![0,\tau_n]\!]}dW)$ then has a Hölder continuous version of any order $\alpha < \gamma/2$ by the first part of the lemma, and one has $X_t = X_t^n$ almost surely on $\{t \leq \tau_n\}$ for each t. □

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

$$K * R = R * K = K - R.^{1} \tag{5.2.11}$$

The resolvent always exists and is unique, and a number of properties such as (local) square integrability and continuity of the original kernel K are inherited by its resolvent; see [69, Theorems 2.3.1 and 2.3.5]. Using the resolvent R one can derive a variation of constants formula as shown in the following lemma.

Lemma 5.5. *Let X be a continuous process, $F\colon \mathbb{R}_+ \to \mathbb{R}^m$ a continuous function, $B \in \mathbb{R}^{d\times d}$ and $Z = \int b\,dt + \int \sigma\,dW$ a continuous semimartingale with b, σ, and $K * dZ$ continuous. Then*

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

where R_B is the resolvent of $-KB$ and $E_B = K - R_B * K$.

Proof. Assume that $X = F + (KB) * X + K * dZ$. Convolving this with R_B and using Lemma 5.1 yields

$$X - R_B * X = (F - R_B * F) + (KB - R_B * (KB)) * X + E_B * dZ.$$

[1] Rather than (5.2.11), it is common to require $K * R = R * K = R - K$ in the definition of resolvent. We use (5.2.11) to remain consistent with [69].

The resolvent equation (5.2.11) states that $KB - R_B * (KB) = -R_B$, so that

$$X = F - R_B * F + E_B * dZ. \tag{5.2.12}$$

Conversely, assume that (5.2.12) holds. It follows from the resolvent equation (5.2.11) that $KB - (KB) * R_B = -R_B$ and

$$(KB) * E_B = (KB) * (K - R_B * K) = -R_B * K.$$

Hence, convolving both sides of (5.2.12) with KB and using Lemma 5.1 yields

$$\begin{aligned} X - (KB) * X &= F + (-R_B - KB + (KB) * R_B) * F + (E_B - (KB) * E_B) * dZ \\ &= F + (E_B + R_B * K) * dZ \\ &= F + K * dZ, \end{aligned}$$

which proves that $X = F + (KB) * X + K * dZ$. □

Another object related to K is its *resolvent of the first kind*, which is an $\mathbb{R}^{d\times d}$-valued measure L on $\mathbb{R}_+$ of locally bounded variation such that

$$K * L = L * K \equiv \mathrm{id}, \tag{5.2.13}$$

see [69, Definition 5.5.1]. Some examples of resolvents of the first and second kind are presented in Table 5.1. A resolvent of the first kind does not always exist. When it does, it has the following properties, which play a key role in several of our arguments.

Lemma 5.6. *Let X be a continuous process and $Z = \int b\,dt + \int \sigma\,dW$ a continuous semimartingale with b, σ, and $K * dZ$ continuous. Assume that K admits a resolvent of the first kind L. Then*

$$X - X_0 = K * dZ \qquad \Longleftrightarrow \qquad L * (X - X_0) = Z. \tag{5.2.14}$$

In this case, for any $F \in L^2_{\mathrm{loc}}(\mathbb{R}_+, \mathbb{C}^{m\times d})$ such that $F * L$ is right-continuous and of locally bounded variation, one has

$$F * dZ = (F * L)(0)X - (F * L)X_0 + d(F * L) * X \quad dt \otimes \mathbb{P}\text{-a.e.} \tag{5.2.15}$$

If $F * dZ$ has a right-continuous version, then with this version (5.2.15) holds up to indistinguishability.

Proof. Assume $X - X_0 = K * dZ$. Apply L to both sides to get

$$L * (X - X_0) = L * (K * dZ) = (L * K) * dZ = \mathrm{id} * dZ = Z,$$

where the second equality follows from Lemma 5.1. This proves the forward implication in (5.2.14). Conversely, assume $L * (X - X_0) = Z$. Then,

$$\begin{aligned} \mathrm{id} * (X - X_0) &= (K * L) * (X - X_0) \\ &= K * (L * (X - X_0)) \\ &= K * Z \\ &= K * (\mathrm{id} * dZ) \\ &= \mathrm{id} * (K * dZ), \end{aligned}$$

using [69, Theorem 3.6.1(ix)] for the second equality and Lemma 5.1 for the last equality. Since both $X - X_0$ and $K * dZ$ are continuous, they must be equal.

To prove (5.2.15), observe that the assumption of right-continuity and locally bounded variation entails that

$$F * L = (F * L)(0) + d(F * L) * \mathrm{id}.$$

Convolving this with K, using associativity of the convolution and (5.2.13), and inspecting the densities of the resulting absolutely continuous functions, we get

$$F = (F * L)(0)K + d(F * L) * K \quad \text{a.e.}$$

Using (5.2.4) and the fact that $K * dZ = X - X_0$ by assumption, it follows that

$$\begin{aligned} F * dZ &= (F * L)(0)K * dZ + d(F * L) * (K * dZ) \\ &= (F * L)(0)X - (F * L)X_0 + d(F * L) * X \end{aligned}$$

holds $dt \otimes \mathbb{P}$-a.e., as claimed. The final statement is clear from right-continuity of $F * L$ and $d(F * L) * X$. □

Remark 5.7. The previous lemma will be used with $F = \Delta_h K$, for fixed $h > 0$. If K is continuous on $(0, \infty)$, then $\Delta_h K * L$ is right-continuous. Moreover, if K is non-negative and L non-increasing in the sense that $s \to L([s, s+t])$ is non-increasing for all $t \geq 0$, then $\Delta_h K * L$ is non-decreasing since

$$(\Delta_h K * L)(t) = 1 - \int_{[0,h)} K(h-s)L(t+ds), \quad t > 0.$$

In particular, $\Delta_h K * L$ is of locally bounded variation.

	$K(t)$	$R(t)$	$L(dt)$
Constant	c	ce^{-ct}	$c^{-1}\delta_0(dt)$
Fractional	$c\,\frac{t^{\alpha-1}}{\Gamma(\alpha)}$	$ct^{\alpha-1}E_{\alpha,\alpha}(-ct^\alpha)$	$c^{-1}\,\frac{t^{-\alpha}}{\Gamma(1-\alpha)}dt$
Exponential	$ce^{-\lambda t}$	$ce^{-\lambda t}e^{-ct}$	$c^{-1}(\delta_0(dt) + \lambda\,dt)$
Gamma	$ce^{-\lambda t}\frac{t^{\alpha-1}}{\Gamma(\alpha)}$	$ce^{-\lambda t}t^{\alpha-1}E_{\alpha,\alpha}(-ct^\alpha)$	$c^{-1}\,\frac{1}{\Gamma(1-\alpha)}e^{-\lambda t}\frac{d}{dt}(t^{-\alpha} * e^{\lambda t})(t)dt$

TABLE 5.1: Some kernels K and their resolvents R and L of the second and first kind. Here $E_{\alpha,\beta}(z) = \sum_{n=0}^{\infty} \frac{z^n}{\Gamma(\alpha n+\beta)}$ denotes the Mittag–Leffler function, and the constant c may be an invertible matrix.

5.3 Stochastic Volterra equations

Fix $d \in \mathbb{N}$ and consider the stochastic Volterra equation (5.1.1) for a given kernel $K \in L^2_{\mathrm{loc}}(\mathbb{R}_+, \mathbb{R}^{d\times d})$, initial condition $X_0 \in \mathbb{R}^d$, and coefficients $b\colon \mathbb{R}^d \to \mathbb{R}^d$ and $\sigma\colon \mathbb{R}^d \to \mathbb{R}^{d\times m}$,

where W is m-dimensional Brownian motion. The equation (5.1.1) can be written more compactly as

$$X = X_0 + K * (b(X)dt + \sigma(X)dW).$$

We will always require the coefficients b and σ as well as solutions of (5.1.1) to be continuous in order to avoid problems with the meaning of the stochastic integral term. As for stochastic (ordinary) differential equations, we call X a strong solution if it is adapted to the filtration generated by W, and a weak solution otherwise.

The following moment bound holds for any solution of (5.1.1) under linear growth conditions on the coefficients.

Lemma 5.8. *Assume b and σ are continuous and satisfy the linear growth condition*

$$|b(x)| \vee |\sigma(x)| \leq c_{\mathrm{LG}}(1 + |x|), \qquad x \in \mathbb{R}^d, \tag{5.3.1}$$

for some constant c_{LG}. Let X be a continuous solution of (5.1.1) with initial condition $X_0 \in \mathbb{R}^d$. Then for any $p \geq 2$ and $T < \infty$ one has

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

for some constant c that only depends on $|X_0|$, $K|_{[0,T]}$, c_{LG}, p and T.

Proof. Let $\tau_n = \inf\{t \geq 0 \colon |X_t| \geq n\} \wedge T$, and observe that

$$|X_t|^p \mathbb{1}_{\{t < \tau_n\}} \leq \left| X_0 + \int_0^t K(t-s)\Big(b(X_s \mathbb{1}_{\{s<\tau_n\}})ds + \sigma(X_s \mathbb{1}_{\{s<\tau_n\}})dW_s\Big)\right|^p.$$

Routine application of the Jensen and BDG inequalities as well as the linear growth condition (5.3.1) show that the expectations $\mathbb{E}[|X_t|^p \mathbb{1}_{\{t<\tau_n\}}]$ satisfy the inequality

$$f_n \leq c' + c'|K|^2 * f_n$$

on $[0,T]$ for some constant c' that only depends on $|X_0|$, $\|K\|_{L^2(0,T)}$, c_{LG}, p and T. Consider now the scalar non-convolution kernel $K'(t,s) = c'|K(t-s)|^2 \mathbb{1}_{s \leq t}$. This is a Volterra kernel in the sense of [69, Definition 9.2.1], and for any interval $[u,v] \subset \mathbb{R}_+$, Young's inequality implies that

$$|||K'|||_{L^1(u,v)} \leq c' \|K\|_{L^2(0,v-u)}, \tag{5.3.2}$$

where $|||\cdot|||_{L^1(u,v)}$ is defined in [69, Definition 9.2.2]. Thus $-K'$ is of type L^1 on $(0,T)$. Next, we show that $-K'$ admits a resolvent of type L^1 on $(0,T)$ in the sense of [69, Definition 9.3.1]. For $v-u$ sufficiently small, the right-hand side in (5.3.2) is smaller than 1, whence $|||K'|||_{L^1(u,v)} < 1$. We now apply [69, Corollary 9.3.14] to obtain a resolvent of type L^1 on $(0,T)$ of $-K'$, which we denote by R'. Since $-c'|K|^2$ is nonpositive, it follows from [69, Proposition 9.8.1] that R' is also nonpositive. The Gronwall type inequality in [69, Lemma 9.8.2] then yields $f_n(t) \leq c'(1-(R'*1)(t)) \leq c'(1-(R'*1)(T))$ for $t \in [0,T]$. Sending n to infinity and using Fatou's lemma completes the proof. □

Remark 5.9. It is clear from the proof that the conclusion of Lemma 5.8 holds also for state and time-dependent predictable coefficients $b(x,t,\omega)$ and $\sigma(\omega,t,x)$, provided they satisfy a

linear growth condition uniformly in (t, ω), that is,

$$|b(x,t,\omega)| \vee |\sigma(x,t,\omega)| \leq c_{\mathrm{LG}}(1+|x|), \qquad x \in \mathbb{R}^d,\ t \in \mathbb{R}_+,\ \omega \in \Omega,$$

for some constant c_{LG}.

The following existence result can be proved using techniques based on classical methods for stochastic differential equations; the proof is given in Section 5.6.1.

Theorem 5.10. *Assume that K admits a resolvent of the first kind, that the components of K satisfy (5.2.5), and that b and σ are continuous and satisfy the linear growth condition (5.3.1). Then (5.1.1) admits a continuous weak solution for any initial condition $X_0 \in \mathbb{R}^d$.*

Remark 5.11. At the cost of increasing the dimension, (5.1.1) also covers the superficially different equation $X = X_0 + K_1 * (b(X)dt) + K_2 * (\sigma(X)dW)$ where the drift and diffusion terms are convolved with different kernels K_1 and K_2. Indeed, if one defines

$$\widetilde{K} = \begin{pmatrix} K_1 & K_2 \\ 0 & K_2 \end{pmatrix}, \qquad \widetilde{b}(x,y) = \begin{pmatrix} b(x) \\ 0 \end{pmatrix}, \qquad \widetilde{\sigma}(x,y) = \begin{pmatrix} 0 & \sigma(x) \\ 0 & 0 \end{pmatrix},$$

and obtains a solution $Z = (X, Y)$ of the equation $Z = Z_0 + \widetilde{K} * (\widetilde{b}(Z)dt + \widetilde{\sigma}(Z)d\widetilde{W})$ in $\mathbb{R}^{2d}$, where $Z_0 = (X_0, 0)$ and $\widetilde{W} = (W', W)$ is a $2d$-dimensional Brownian motion, then X is a solution of the original equation of interest. If K_1 and K_2 admit resolvents of the first kind L_1 and L_2, then

$$\widetilde{L} = \begin{pmatrix} L_1 & -L_1 \\ 0 & L_2 \end{pmatrix}$$

is a resolvent of the first kind of $\widetilde{K}$, and Theorem 5.10 is applicable.

Our next existence result is more delicate, as it involves an assertion about stochastic invariance of the nonnegative orthant $\mathbb{R}^d_+$. This forces us to impose stronger conditions on the kernel K along with suitable boundary conditions on the coefficients b and σ. We note that any nonnegative and non-increasing kernel that is not identically zero admits a resolvent of the first kind; see [69, Theorem 5.5.5].

Theorem 5.12. *Assume that K is diagonal with scalar kernels K_i on the diagonal that satisfy (5.2.5) as well as*

$$\begin{gathered} \text{K_i is nonnegative, not identically zero, non-increasing and continuous on} \\ \text{$(0,\infty)$, and its resolvent of the first kind L_i is nonnegative and} \\ \text{non-increasing in that $s \mapsto L_i([s, s+t])$ is non-increasing for all $t \geq 0$.} \end{gathered} \tag{5.3.3}$$

Assume also that b and σ are continuous and satisfy the linear growth condition (5.3.1) along with the boundary conditions

$$x_i = 0 \text{ implies } b_i(x) \geq 0 \text{ and } \sigma_i(x) = 0,$$

where $\sigma_i(x)$ is the ith row of $\sigma(x)$. Then (5.1.1) admits an $\mathbb{R}^d_+$-valued continuous weak solution for any initial condition $X_0 \in \mathbb{R}^d_+$.

Example 5.13. *If K_i is completely monotone on $(0,\infty)$ and not identically zero, then (5.3.3) holds due to [69, Theorem 5.5.4]. Recall that a function f is called completely monotone on $(0,\infty)$ if it is infinitely differentiable there with $(-1)^k f^{(k)}(t) \geq 0$ for all $t > 0$ and*

$k = 0, 1, \ldots$. *This covers, for instance, any constant positive kernel, the fractional kernel* $t^{\alpha-1}$ *with* $\alpha \in (\frac{1}{2}, 1)$*, and the exponentially decaying kernel* $e^{-\beta t}$ *with* $\beta > 0$*. Moreover, sums and products of completely monotone functions are completely monotone.*

Proof of Theorem 5.12. Define coefficients b^n and σ^n by

$$b^n(x) = b\left((x - n^{-1})^+\right), \qquad \sigma^n(x) = \sigma\left((x - n^{-1})^+\right),$$

and let X^n be the solution of (5.1.1) given by Theorem 5.10, with b and σ replaced by b^n and σ^n. Note that b^n and σ^n are continuous, satisfy (5.3.1) with a common constant, and converge to $b(x^+)$ and $\sigma(x^+)$ locally uniformly. Lemmas 5.30 and 5.31 therefore imply that, along a subsequence, X^n converges weakly to a solution X of the stochastic Volterra equation

$$X_t = X_0 + \int_0^t K(t-s)b(X_s^+)ds + \int_0^t K(t-s)\sigma(X_s^+)dW_s.$$

It remains to prove that X is $\mathbb{R}^d_+$-valued and hence a solution of (5.1.1). For this it suffices to prove that each X^n is $\mathbb{R}^d_+$-valued.

Dropping the superscript n, we are thus left with the task of proving the theorem under the stronger condition that, for some fixed $n \in \mathbb{N}$,

$$x_i \leq n^{-1} \text{ implies } b_i(x) \geq 0 \text{ and } \sigma_i(x) = 0. \tag{5.3.4}$$

Define $Z = \int b(X)dt + \int \sigma(X)dW$ and $\tau = \inf\{t \geq 0 \colon X_t \notin \mathbb{R}^d_+\}$. On $\{\tau < \infty\}$ we have

$$X_{\tau+h} = X_0 + (K * dZ)_{\tau+h} = X_0 + (\Delta_h K * dZ)_\tau + \int_0^h K(h-s)dZ_{\tau+s} \tag{5.3.5}$$

for all $h \geq 0$. We claim that

$$(\Delta_h K_i * L_i)(t) \text{ is nondecreasing in } t \text{ for any } h \geq 0. \tag{5.3.6}$$

Indeed, using that $K_i * L_i \equiv 1$ we have

$$\begin{aligned}(\Delta_h K_i * L_i)(t) &= \int_{[0,t]} K_i(t+h-u)L_i(du) \\ &= 1 - \int_{(t,t+h]} K_i(t+h-u)L_i(du) \\ &= 1 - \int_{(0,h]} K_i(h-u)L_i(t+du),\end{aligned}$$

and therefore, for any $s \leq t$,

$$(\Delta_h K_i * L_i)(t) - (\Delta_h K_i * L_i)(s) = \int_{(0,h]} K_i(h-u)\left(L_i(s+du) - L_i(t+du)\right).$$

This is nonnegative since K_i is nonnegative and L_i non-increasing, proving (5.3.6). Furthermore, since K_i is non-increasing and L_i nonnegative we obtain

$$0 \leq (\Delta_h K_i * L_i)(t) \leq (K_i * L_i)(t) = 1. \tag{5.3.7}$$

Since $\Delta_h K_i$ is continuous and of locally bounded variation on $\mathbb{R}_+$ for $h > 0$, it follows that $\Delta_h K_i * L_i$ is right-continuous and of locally bounded variation. Moreover, $\Delta_h K_i$ satisfies

(5.2.5) due to Example 5.36, so that $\Delta_h K_i * dZ$ has a continuous version by Lemma 5.4. Thus (5.2.15) in Lemma 5.6, along with (5.3.6)–(5.3.7) and the fact that X_t is $\mathbb{R}^d_+$-valued for $t \leq \tau$, yield

$$\begin{aligned} X_{i,0} + (\Delta_h K_i * dZ_i)_\tau &= (1 - (\Delta_h K_i * L_i)(\tau))\, X_{i,0} \\ &\quad + (\Delta_h K_i * L_i)(0) X_{i,\tau} \\ &\quad + (d(\Delta_h K_i * L_i) * X_i)_\tau \\ &\geq 0. \end{aligned}$$

In view of (5.3.5) it follows that

$$X_{i,\tau+h} \geq \int_0^h K_i(h-s)\,(b_i(X_{\tau+s})ds + \sigma_i(X_{\tau+s})dW_{\tau+s}) \tag{5.3.8}$$

on $\{\tau < \infty\}$ for all i and all $h \geq 0$.

Now, on $\{\tau < \infty\}$ there is an index i (depending on ω) such that $X_{i,\tau} = 0$ and $X_{i,\tau+h} < 0$ for arbitrarily small but positive values of h. On the other hand, by continuity there is some $\varepsilon > 0$ (again depending on ω) such that $X_{i,\tau+h} \leq n^{-1}$ for all $h \in [0, \varepsilon)$. Thus (5.3.4) and (5.3.8) yield $X_{i,\tau+h} \geq 0$ for all $h \in [0, \varepsilon)$. This contradiction shows that $\tau = \infty$, as desired. □

5.4 Affine Volterra processes

Fix a dimension $d \in \mathbb{N}$ and a kernel $K \in L^2_{\mathrm{loc}}(\mathbb{R}_+, \mathbb{R}^{d\times d})$. Let $a\colon \mathbb{R}^d \to \mathbb{S}^d$ and $b\colon \mathbb{R}^d \to \mathbb{R}^d$ be affine maps given by

$$\begin{aligned} a(x) &= A^0 + x_1 A^1 + \cdots + x_d A^d \\ b(x) &= b^0 + x_1 b^1 + \cdots + x_d b^d \end{aligned} \tag{5.4.1}$$

for some $A^i \in \mathbb{S}^d$ and $b^i \in \mathbb{R}^d$, $i = 0, \ldots, d$. To simplify notation we introduce the $d \times d$ matrix

$$B = \begin{pmatrix} b^1 & \cdots & b^d \end{pmatrix},$$

and for any row vector $u \in (\mathbb{C}^d)^*$ we define the row vector

$$A(u) = (uA^1u^\top, \ldots, uA^du^\top).$$

Let $\mathcal{D}$ be a subset of $\mathbb{R}^d$, which will play the role of state space for the process defined below, and assume that $a(x)$ is positive semidefinite for every $x \in \mathcal{D}$. Let $\sigma\colon \mathbb{R}^d \to \mathbb{R}^{d\times d}$ be continuous and satisfy $\sigma(x)\sigma(x)^\top = a(x)$ for every $x \in \mathcal{D}$. For instance, one can take $\sigma(x) = \sqrt{\pi(a(x))}$, where π denotes the orthogonal projection onto the positive semidefinite cone, and the positive semidefinite square root is understood.

Definition 5.14. An *affine Volterra process (with state space $\mathcal{D}$)* is a continuous $\mathcal{D}$-valued solution X of (5.1.1) with $a = \sigma\sigma^\top$ and b as in (5.4.1). In this chapter we always take X_0 deterministic.

Setting $K \equiv \mathrm{id}$ we recover the usual notion of an affine diffusion with state space $\mathcal{D}$; see e.g. [56]. Even in this case, existence and uniqueness is often approached by first fixing a state space $\mathcal{D}$ of interest, and then studying conditions on (a, b) under which existence

and uniqueness can be proved; see e.g. [48, 34, 106, 84]. A key goal is then to obtain explicit parameterizations that can be used in applications. In later sections we carry out this analysis for affine Volterra processes with state space $\mathbb{R}^d$, $\mathbb{R}^d_+$, and $\mathbb{R} \times \mathbb{R}_+$. In the standard affine case more general results are available. In [105], existence and uniqueness of affine jump-diffusions are characterized on closed convex state spaces, while Chapter 3 of this thesis provides necessary and sufficient first order geometric conditions for existence of affine diffusions on general closed state spaces. We do not pursue such generality here for affine Volterra processes.

Assuming that an affine Volterra process is given, one can however make statements about its law. In the present section we develop general results in this direction. We start with a formula for the conditional mean. This is an immediate consequence of the variation of constants formula derived in Lemma 5.5.

Lemma 5.15. *Let X be an affine Volterra process. Then for all $t \leq T$,*

$$\mathbb{E}[X_T \mid \mathcal{F}_t] = \left(\mathrm{id} - \int_0^T R_B(s)ds\right) X_0 + \left(\int_0^T E_B(s)ds\right) b^0 + \int_0^t E_B(T-s)\sigma(X_s)dW_s, \tag{5.4.2}$$

where R_B is the resolvent of $-KB$ and $E_B = K - R_B * K$. In particular,

$$\mathbb{E}[X_T] = \left(\mathrm{id} - \int_0^T R_B(s)ds\right) X_0 + \left(\int_0^T E_B(s)ds\right) b^0.$$

Proof. Since $X = X_0 + (KB) * X + K * (b^0 dt + \sigma(X)dW)$, Lemma 5.5 yields

$$X = (\mathrm{id} - R_B * 1)\, X_0 + E_B * (b^0 dt + \sigma(X)dW).$$

Consider the local martingale $M_t = \int_0^t E_B(T-s)\sigma(X_s)dW_s$, $t \in [0,T]$. Its quadratic variation satisfies

$$\mathbb{E}[|\langle M\rangle_T|] \leq \int_0^T |E_B(T-s)|^2\, \mathbb{E}[|\sigma(X_s)|^2]ds \leq \|E_B\|_{L^2(0,T)} \max_{s\leq T} \mathbb{E}[|\sigma(X_s)|^2],$$

which is finite by Lemma 5.8. Thus M is a martingale, so taking $\mathcal{F}_t$-conditional expectations completes the proof. □

The first main result of this section is the following theorem, which expresses the conditional Fourier–Laplace functional of an affine Volterra process in terms of the conditional mean in Lemma 5.15 and the solution of a quadratic Volterra integral equation, which we call a *Riccati–Volterra equation*.

Theorem 5.16. *Let X be an affine Volterra process and fix some $T < \infty$, $u \in (\mathbb{C}^d)^*$, and $f \in L^1([0,T], (\mathbb{C}^d)^*)$. Assume $\psi \in L^2([0,T], (\mathbb{C}^d)^*)$ solves the Riccati–Volterra equation*

$$\psi = uK + \left(f + \psi B + \frac{1}{2}A(\psi)\right) * K. \tag{5.4.3}$$

Then the process $\{Y_t,\, 0 \le t \le T\}$ *defined by*

$$Y_t = Y_0 + \int_0^t \psi(T-s)\sigma(X_s)dW_s - \frac{1}{2}\int_0^t \psi(T-s)a(X_s)\psi(T-s)^\top ds, \tag{5.4.4}$$

$$Y_0 = uX_0 + \int_0^T \left(f(s)X_0 + \psi(s)b(X_0) + \frac{1}{2}\psi(s)a(X_0)\psi(s)^\top \right) ds \tag{5.4.5}$$

satisfies

$$Y_t = \mathbb{E}[uX_T + (f * X)_T \mid \mathcal{F}_t] + \frac{1}{2}\int_t^T \psi(T-s)a(\mathbb{E}[X_s \mid \mathcal{F}_t])\psi(T-s)^\top ds \tag{5.4.6}$$

for all $0 \le t \le T$. *The process* $\{\exp(Y_t),\, 0 \le t \le T\}$ *is a local martingale and, if it is a true martingale, one has the exponential-affine transform formula*

$$\mathbb{E}\left[\exp\left(uX_T + (f * X)_T\right) \;\middle|\; \mathcal{F}_t\right] = \exp(Y_t), \quad t \le T. \tag{5.4.7}$$

Referring to (6.2.13) as an exponential-affine transform formula is motivated by the fact that Y_t depends affinely on the conditional expectations $\mathbb{E}[X_s \mid \mathcal{F}_t]$. We show in Theorem 5.18 below that under mild additional assumptions on K, Y_t is actually an affine function of the past trajectory $\{X_s,\, s \le t\}$. Before proving Theorem 5.16 we give the following lemma.

Lemma 5.17. *The Riccati–Volterra equation* (5.4.3) *is equivalent to*

$$\psi = uE_B + \left(f + \frac{1}{2}A(\psi)\right) * E_B, \tag{5.4.8}$$

where $E_B = K - R_B * K$ *and* R_B *is the resolvent of* $-KB$.

Proof. Assume (5.4.8) holds. Using the identity $E_B * (BK) = -R_B * K$ we get

$$\psi - \psi * (BK) = u(E_B + R_B * K) + \left(f + \frac{1}{2}A(\psi)\right) * (E_B + R_B * K),$$

which is (5.4.3). Conversely, assume (5.4.3) holds. With $\widetilde{R}_B$ being the resolvent of $-BK$, we obtain

$$\psi - \psi * \widetilde{R}_B = u(K - K * \widetilde{R}_B) + \left(f + \frac{1}{2}A(\psi)\right) * (K - K * \widetilde{R}_B) - \psi * \widetilde{R}_B.$$

To deduce (5.4.8) it suffices to prove $K * \widetilde{R}_B = R_B * K$. Equivalently, we show that for each $T < \infty$, there is some $\sigma > 0$ such that

$$(e^{-\sigma t}K) * (e^{-\sigma t}\widetilde{R}_B) = (e^{-\sigma t}R_B) * (e^{-\sigma t}K) \text{ on } [0,T], \tag{5.4.9}$$

where $e^{-\sigma t}$ is shorthand for the function $t \mapsto e^{-\sigma t}$. It follows from the definitions that $e^{-\sigma t}R_B$ is the resolvent of $-e^{-\sigma t}KB$, and that $e^{-\sigma t}\widetilde{R}_B$ is the resolvent of $-e^{-\sigma t}BK$; see [69, Lemma 2.3.3]. Choosing σ large enough that $\|e^{-\sigma t}KB\|_{L^1(0,T)} < 1$ we get, as in the proof of [69, Theorem 2.3.1],

$$e^{-\sigma t}R_B = -\sum_{k\ge 1}(e^{-\sigma t}KB)^{*k} \quad \text{and} \quad e^{-\sigma t}\widetilde{R}_B = -\sum_{k\ge 1}(e^{-\sigma t}BK)^{*k}$$

on $[0,T]$. This readily implies (5.4.9), as required. □

Proof of Theorem 5.16. Let $\widetilde{Y}_t$ be defined by the right-hand side of (6.2.14) for $0 \leq t \leq T$. We first prove that $\widetilde{Y}_0 = Y_0$. A calculation using the identity $va(x)v^\top = vA^0v^\top + A(v)x$ yields

$$\begin{aligned}\widetilde{Y}_0 - Y_0 = u\,\mathbb{E}[X_T - X_0] + (f * \mathbb{E}[X - X_0])(T) \\ + \left(\frac{1}{2}A(\psi) * \mathbb{E}[X - X_0]\right)(T) - \left(\psi * (b^0 + BX_0)\right)(T),\end{aligned} \tag{5.4.10}$$

where $\mathbb{E}[X - X_0]$ denotes the function $t \mapsto \mathbb{E}[X_t - X_0]$. This function satisfies

$$\mathbb{E}[X - X_0] = K * \left(b^0 + B\,\mathbb{E}[X]\right),$$

as can be seen by taking expectations in (5.1.1) and using Lemma 5.8. Consequently,

$$\begin{aligned}\frac{1}{2}A(\psi) * \mathbb{E}[X - X_0] &= \frac{1}{2}A(\psi) * K * \left(b^0 + B\,\mathbb{E}[X]\right) \\ &= (\psi - uK - (f + \psi B) * K) * \left(b^0 + B\,\mathbb{E}[X]\right) \\ &= \psi * \left(b^0 + B\,\mathbb{E}[X]\right) - u\,\mathbb{E}[X - X_0] - (f + \psi B) * \mathbb{E}[X - X_0].\end{aligned}$$

Substituting this into (5.4.10) yields $\widetilde{Y}_0 - Y_0 = 0$, as required.

We now prove that $\widetilde{Y} = Y$. In the following calculations we let C denote an $\mathcal{F}_0$-measurable quantity that does not depend on t, and may change from line to line. Using again the identity $va(x)v^\top = vA^0v^\top + A(v)x$ we get

$$\begin{aligned}\widetilde{Y}_t = C + u\,\mathbb{E}[X_T \mid \mathcal{F}_t] + \int_0^T \left(f + \frac{1}{2}A(\psi)\right)(T - s)\,\mathbb{E}[X_s \mid \mathcal{F}_t]\,ds \\ - \frac{1}{2}\int_0^t \psi(T - s)a(X_s)\psi(T - s)^\top ds.\end{aligned}$$

Lemma 5.15, the stochastic Fubini theorem, see [111, Theorem 2.2], and a change of variables yield

$$\begin{aligned}&\int_0^T \left(f + \frac{1}{2}A(\psi)\right)(T - s)\,\mathbb{E}[X_s \mid \mathcal{F}_t]\,ds \\ &= C + \int_0^T \left(f + \frac{1}{2}A(\psi)\right)(T - s) \int_0^t \mathbb{1}_{\{r<s\}}E_B(s - r)\sigma(X_r)dW_r\,ds \\ &= C + \int_0^t \left(\int_r^T \left(f + \frac{1}{2}A(\psi)\right)(T - s)E_B(s - r)ds\right)\sigma(X_r)dW_r \\ &= C + \int_0^t \left(\left(f + \frac{1}{2}A(\psi)\right) * E_B\right)(T - r)\sigma(X_r)dW_r,\end{aligned}$$

where the application of the stochastic Fubini theorem in the second equality is justified by the fact that

$$\begin{aligned}&\int_0^T \left(\int_0^t \left|\left(f + \frac{1}{2}A(\psi)\right)(T - s)\,\mathbb{1}_{\{r<s\}}E_B(s - r)\sigma(X_r)\right|^2 dr\right)^{1/2} ds \\ &\leq \max_{0\leq s\leq T}|\sigma(X_s)|\,\|E_B\|_{L^2(0,T)}\|f + \frac{1}{2}A(\psi)\|_{L^1(0,T)} < \infty.\end{aligned}$$

Since $\mathbb{E}[X_T \mid \mathcal{F}_t] = C + \int_0^t E_B(T-r)\sigma(X_r)dW_r$ by Lemma 5.15, we arrive at

$$\widetilde{Y}_t = \widetilde{Y}_0 + \int_0^t \left(uE_B + \left(f + \frac{1}{2}A(\psi)\right) * E_B\right)(T-r)\sigma(X_r)dW_r - \frac{1}{2}\int_0^t \psi(T-s)a(X_s)\psi(T-s)^\top ds.$$

Due to Lemma 5.17 and (6.2.18) we then obtain (6.2.17).

The final statements are now straightforward. Indeed, (6.2.17) shows that $Y + \frac{1}{2}\langle Y\rangle$ is a local martingale, so that $\exp(Y)$ is a local martingale by Itô's formula. In the true martingale situation, the exponential-affine transform formula then follows upon observing that $Y_T = uX_T + (f * X)_T$ by (6.2.14). □

In the particular case $f \equiv 0$ and $t = 0$, Theorem 5.16 yields two different expressions for the Fourier–Laplace transform of X,

$$\mathbb{E}[e^{uX_T}] = \exp\left(\mathbb{E}[uX_T] + \frac{1}{2}\int_0^T \psi(T-t)a(\mathbb{E}[X_t])\psi(T-t)^\top dt\right) \tag{5.4.11}$$

$$= \exp\left(\phi(T) + \chi(T)X_0\right), \tag{5.4.12}$$

where ϕ and χ are defined by

$$\phi'(t) = \psi(t)b^0 + \frac{1}{2}\psi(t)A^0\psi(t)^\top, \qquad \phi(0) = 0, \tag{5.4.13}$$

$$\chi'(t) = \psi(t)B + \frac{1}{2}A(\psi(t)), \qquad \chi(0) = u. \tag{5.4.14}$$

If K admits a resolvent of the first kind L, one sees upon convolving (5.4.3) by L and using (5.2.13) that $\chi = \psi * L$; see also Example 5.20 below. Note that (5.4.13)–(6.2.11) reduce to the classical Riccati equations when $K \equiv \mathrm{id}$, since in this case $L = \delta_0\mathrm{id}$ and hence $\psi = \chi$. While the first expression (5.4.11) does exist in the literature on affine diffusions in the classical case $K \equiv \mathrm{id}$, see [105, Proposition 4.2], the second expression (5.4.12) is much more common.

In the classical case one has a conditional version of (5.4.12), namely

$$\mathbb{E}[e^{uX_T} \mid \mathcal{F}_t] = \exp\left(\phi(T-t) + \psi(T-t)X_t\right).$$

This formulation has the advantage of showing clearly that the right-hand side depends on X_t in an exponential-affine manner. In the general Volterra case the lack of Markovianity precludes such a simple form, but using the resolvent of the first kind it is still possible to obtain an explicit expression that is exponential-affine in the past trajectory $\{X_s,\, s \le t\}$. Note that this property is not at all obvious either from (6.2.13) or from the expression

$$\mathbb{E}[e^{uX_T} \mid \mathcal{F}_t] = \mathcal{E}\Big(Y_0 + \int \psi(T-s)\sigma(X_s)dW_s\Big)_t,$$

which follows directly from (6.2.17)–(6.2.18) and where $\mathcal{E}$ denotes stochastic exponential. The second main result of this section directly leads to such an exponential-affine representation under mild additional assumptions on K.

Theorem 5.18. *Assume K is continuous on $(0,\infty)$, admits a resolvent of the first kind L, and that one has the total variation bound*

$$\sup_{h\leq T} \|\Delta_h K * L\|_{\mathrm{TV}(0,T)} < \infty \tag{5.4.15}$$

for all $T \geq 0$. Then the following statements hold:

(i) With the notation and assumptions of Lemma 5.15, the matrix function

$$\Pi_h = \Delta_h E_B * L - \Delta_h (E_B * L)$$

is right-continuous and of locally bounded variation on $[0,\infty)$ for every $h \geq 0$, and the conditional expectation (5.4.2) is given by

$$\mathbb{E}[X_T \mid \mathcal{F}_t] = (\mathrm{id} * E_B)(h) b^0 + (\Delta_h E_B * L)(0) X_t - \Pi_h(t) X_0 + (d\Pi_h * X)_t \tag{5.4.16}$$

with $h = T - t$.

(ii) With the notation and assumptions of Theorem 5.16, the scalar function

$$\pi_h = \Delta_h \psi * L - \Delta_h (\psi * L)$$

is right-continuous and of bounded variation on $[0, T-h]$ for every $h \leq T - t$, and the process Y in (6.2.14) is given by

$$Y_t = \phi(h) + (\Delta_h f * X)_t + (\Delta_h \psi * L)(0) X_t - \pi_h(t) X_0 + (d\pi_h * X)_t \tag{5.4.17}$$

with $h = T - t$ and

$$\phi(h) = \int_0^h \left(\psi(s) b_0 + \frac{1}{2} \psi(s) A_0 \psi(s)^\top \right) ds.$$

Proof. (i): We wish to apply Lemma 5.6 with $F = \Delta_h E_B$ for any fixed $h \geq 0$, so we first verify its hypotheses. Manipulations using Fubini's theorem give

$$(\Delta_h E_B * L)(t) = (\Delta_h K * L)(t) - \int_h^{t+h} R_B(s) ds - \int_0^h R_B(h-s)(\Delta_s K * L)(t) ds.$$

Owing to (5.4.15) we get the bound

$$\sup_{h\leq T} \|\Delta_h E_B * L\|_{\mathrm{TV}(0,T)} < \infty, \tag{5.4.18}$$

and using continuity of K on $(0,\infty)$ we also get that $\Delta_h E_B * L$ is right-continuous on $\mathbb{R}_+$. In particular, in view of the identity

$$E_B * L = \mathrm{id} - R_B * \mathrm{id}, \tag{5.4.19}$$

we deduce that Π_h is right-continuous and of locally bounded variation as stated. Now, observe that $E_B = K - R_B * K$ is continuous on $(0,\infty)$, since this holds for K and since R_B and K are both in L^2_{loc}. Moreover, Example 5.33 and 5 imply that the components of E_B satisfy (5.2.5). As a result, Example 5.36 shows that the components of $\Delta_h E_B$ satisfy

(5.2.5) for any $h \geq 0$. Fix $h = T - t$ and define

$$Z = \int b(X)dt + \int \sigma(X)dW.$$

It follows from Lemma 5.4 that $\Delta_h E_B * dZ$ has a continuous version. Lemma 5.6 with $F = \Delta_h E_B$ yields

$$\Delta_h E_B * dZ = (\Delta_h E_B * L)(0)X - (\Delta_h E_B * L)X_0 + d(\Delta_h E_B * L) * X.$$

Moreover, rearranging (5.4.2) and using (5.4.19) gives

$$\mathbb{E}[X_T \mid \mathcal{F}_t] = (E_B * L)(T)X_0 + (E_B * \mathrm{id})(h)b^0 + (\Delta_h E_B * (dZ - BXdt))_t.$$

Combining the previous two equalities and using the definition of Π_h yields

$$\begin{aligned}\mathbb{E}[X_T \mid \mathcal{F}_t] = (E_B * \mathrm{id})(h)b^0 + (\Delta_h E_B * L)(0)X_t - \Pi_h(t)X_0 \\ + ((d(\Delta_h E_B * L) - \Delta_h E_B B dt) * X)_t.\end{aligned}$$

The definition of E_B and the resolvent equation (5.2.11) show that $E_B B = -R_B$, which in combination with (5.4.19) gives $E_B B dt = d(E_B * L)$. Thus (5.4.16) holds as claimed. This completes the proof of i.

(ii): Recall that Lemma 5.17 gives $\psi = uE_B + G(\psi) * E_B$ where

$$G(\psi) = f + \frac{1}{2}A(\psi).$$

Manipulating this equation gives

$$\Delta_h \psi(t) = u\Delta_h E_B(t) + (G(\psi) * \Delta_t E_B)(h) + (G(\Delta_h \psi) * E_B)(t).$$

Convolving with L and using Fubini yields

$$\begin{aligned}(\Delta_h \psi * L)(t) = u(\Delta_h E_B * L)(t) + (G(\psi) * (\Delta_\bullet E_B * L)(t))(h) \\ + (G(\Delta_h \psi) * E_B * L)(t),\end{aligned} \tag{5.4.20}$$

where $(\Delta_\bullet E_B * L)(t)$ denotes the function $s \mapsto (\Delta_s E_B * L)(t)$. Similarly,

$$\begin{aligned}\Delta_h(\psi * L)(t) = u\Delta_h(E_B * L)(t) + (G(\psi) * \Delta_\bullet(E_B * L)(t))(h) \\ + (G(\Delta_h \psi) * E_B * L)(t).\end{aligned}$$

Computing the difference between the previous two expressions gives

$$\pi_h(t) = u\Pi_h(t) + (G(\psi) * \Pi_\bullet(t))(h). \tag{5.4.21}$$

In combination with (5.4.18) and (5.4.19), as well as the properties of Π_h that we have already proved, it follows that π_h is right-continuous and of bounded variation as stated. Now, using Fubini we get

$$\mathbb{E}[(f * X)_T \mid \mathcal{F}_t] = (\Delta_{T-t} f * X)_t + \int_0^{T-t} f(s)\mathbb{E}[X_{T-s} \mid \mathcal{F}_t]\, ds. \tag{5.4.22}$$

Combining (6.2.14), (5.4.16), and (5.4.22), we obtain after some computations

$$\begin{aligned} Y_t &= (\Delta_{T-t} f * X)_t + \frac{1}{2}\int_0^{T-t} \psi(s) A^0 \psi(s)^\top \, ds \\ &\quad + \left(u(\mathrm{id} * E_B)(T-t) + \int_0^{T-t} G(\psi(s))(\mathrm{id} * E_B)(T-t-s)\, ds \right) b^0 \\ &\quad + \left(u(\Delta_{T-t} E_B * L)(0) + \int_0^{T-t} G(\psi(s))(\Delta_{T-t-s} E_B * L)(0)\, ds \right) X_t \\ &\quad - \left(u\Pi_{T-t}(t) + \int_0^{T-t} G(\psi(s))\Pi_{T-t-s}(t)\, ds \right) X_0 \\ &\quad + u(d\Pi_{T-t} * X)_t + \int_0^{T-t} G(\psi(s))(d\Pi_{T-t-s} * X)_t \, ds \\ &= \mathbf{I} + \mathbf{II} + \mathbf{III} + \mathbf{IV} + \mathbf{V}. \end{aligned} \tag{5.4.23}$$

Here

$$\begin{aligned} \mathbf{I} + \mathbf{II} &= (\Delta_{T-t} f * X)_t + \frac{1}{2}\int_0^{T-t} \psi(s) A^0 \psi(s)^\top \, ds \\ &\quad + \left((uE_B + G(\psi) * E_B)\, b^0 * 1 \right)(T-t) \\ &= (\Delta_{T-t} f * X)_t + \phi(T-t). \end{aligned} \tag{5.4.24}$$

As a result of (5.4.19), $E_B * L$ is continuous on $\mathbb{R}_+$, whence $(G(\Delta_h \psi) * E_B * L)(0) = 0$. Evaluating (5.4.20) at $t = 0$ thus gives

$$\mathbf{III} = (\Delta_{T-t}\psi * L)(0) X_t. \tag{5.4.25}$$

As a consequence of (5.4.21),

$$\mathbf{IV} = -\pi_{T-t}(t) X_0. \tag{5.4.26}$$

Finally, it follows from (5.4.21) that $d\pi_h = u d\Pi_h + \mu_h$, where $\mu_h(dt) = (G(\psi) * d\Pi_\bullet(dt))(h)$. Since for any bounded function g on $[0,t]$ we have

$$\int_{[0,t]} g(r)\mu_h(dr) = \int_0^h G(\psi(s)) \left(\int_0^t g(r) d\Pi_{h-s}(dr) \right) ds,$$

we obtain

$$\mathbf{V} = (d\pi_{T-t} * X)_t. \tag{5.4.27}$$

Combining (5.4.23)–(5.4.27) yields (5.4.17) and completes the proof. □

Remark 5.19. Consider the classical case $K \equiv \mathrm{id}$. Then $L(dt) = \mathrm{id}\, \delta_0(dt)$, $R_B(t) = -Be^{Bt}$, and $E_B(t) = e^{Bt}$. Thus $(\Delta_h E_B * L)(t) = e^{B(t+h)} = \Delta_h (E_B * L)(t)$, so that (5.4.16) reduces to the well known expression $\mathbb{E}[X_T \mid \mathcal{F}_t] = e^{B(T-t)} X_t + \int_0^{T-t} e^{Bs} b^0 ds$. In addition, in (5.4.17) the correction π_h vanishes so that, if $f \equiv 0$, the expression for Y_t reduces to the classical form $\phi(T-t) + \psi(T-t) X_t$.

Example 5.20 (Fractional affine processes)**.** *Let* $K = \mathrm{diag}\,[(K_1, \dots, K_d)]$*, where*

$$K_i(t) = \frac{t^{\alpha_i - 1}}{\Gamma(\alpha_i)}$$

for some $\alpha_i \in (\frac{1}{2}, 1]$. *Then* $L = \operatorname{diag}[(L_1, \dots, L_d)]$ *with* $L_i(dt) = \frac{t^{-\alpha_i}}{\Gamma(1-\alpha_i)}dt$ *if* $\alpha_i < 1$, *and* $L_i(dt) = \delta_0(dt)$ *if* $\alpha_i = 1$. *It follows that* $\chi_i = \psi_i * L_i = I^{1-\alpha_i}\psi_i$, *where* $I^{1-\alpha_i}$ *denotes the Riemann-Liouville fractional integral operator. Hence,* (5.4.3) *and* (5.4.13) *reduce to the following system of fractional Riccati equations,*

$$\begin{aligned} \phi' &= \psi b^0 + \frac{1}{2}\psi A^0 \psi^\top, & \phi(0) &= 0, \\ D^{\alpha_i}\psi_i &= f_i + \psi b^i + \frac{1}{2}\psi A^i \psi^\top, \quad i = 1, \dots, d, & I^{1-\alpha}\psi(0) &= u, \end{aligned}$$

where $D^{\alpha_i} = \frac{d}{dt} I^{1-\alpha_i}$ *is the Riemann-Liouville fractional derivative. Moreover, for* $t = 0$, (6.2.13) *reads*

$$\mathbb{E}\left[e^{uX_T + (f*X)_T}\right] = \exp\left(\phi(T) + I^{1-\alpha}\psi(T)X_0\right)$$

where we write $I^{1-\alpha}\psi = (I^{1-\alpha_1}\psi_1, \dots, I^{1-\alpha_d}\psi_d)$. *This generalizes the expressions in [51, 52]. Notice that the identity* $L_{\alpha_i} * K_{\alpha_i} \equiv 1$ *is equivalent to the identity* $D^{\alpha_i}(I^{\alpha_i} f) = f$.

5.5 Examples

5.5.1 The Volterra Ornstein–Uhlenbeck process

The particular specification of (5.4.1) where $A^1 = \cdots = A^d = 0$, so that $a \equiv A^0$ is a constant symmetric positive semidefinite matrix, yields an affine Volterra process with state space $E = \mathbb{R}^d$ that we call the *Volterra Ornstein–Uhlenbeck process*. It is the solution of the equation

$$X_t = X_0 + \int_0^t K(t-s)(b^0 + BX_s)ds + \int_0^t K(t-s)\sigma dW_s,$$

where $\sigma \in \mathbb{R}^{d\times d}$ is a constant matrix with $\sigma\sigma^\top = A^0$. Here existence and uniqueness is no issue. Indeed, Lemma 5.6 with $T = t$ yields the explicit formula

$$X_t = \left(\mathrm{id} - \int_0^t R_B(s)ds\right) X_0 + \left(\int_0^t E_B(s)ds\right) b^0 + \int_0^t E_B(t-s)\sigma dW_s,$$

where R_B is the resolvent of $-KB$ and $E_B = K - R_B * K$. In particular X_t is Gaussian. Furthermore, the solution of the Riccati–Volterra equation (5.4.3) is obtained explicitly via Lemma 5.17 as

$$\psi = uE_B + f * E_B.$$

The quadratic variation of the process Y in (6.2.17) is given by

$$\langle Y \rangle_t = \int_0^t \psi(T-s)\sigma\sigma^\top \psi(T-s)^\top ds,$$

and is in particular deterministic. The martingale condition in Theorem 5.16 is thus clearly satisfied, and the exponential-affine transform formula (6.2.13) holds for any $T < \infty$, $u \in (\mathbb{C}^d)^*$, and $f \in L^1([0,T], (\mathbb{C}^d)^*)$.

5.5.2 The Volterra square-root process

We now consider affine Volterra processes whose state space is the nonnegative orthant $E = \mathbb{R}^d_+$. We let K be diagonal with scalar kernels $K_i \in L^2_{\mathrm{loc}}(\mathbb{R}_+, \mathbb{R})$ on the diagonal. The coefficients a and b in (5.4.1) are chosen so that $A^0 = 0$, A^i is zero except for the (i,i) element which is equal to σ_i^2 for some $\sigma_i > 0$, and

$$b^0 \in \mathbb{R}^d_+ \text{ and } B_{ij} \geq 0 \text{ for } i \neq j. \tag{5.5.1}$$

The conditions on a and b are the same as in the classical situation $K \equiv \mathrm{id}$, in which case they are necessary and sufficient for (5.1.1) to admit an $\mathbb{R}^d_+$-valued solution for every initial condition $X_0 \in \mathbb{R}^d_+$. With this setup, we obtain an affine Volterra process that we call the *Volterra square-root process*. It is the solution of the equation

$$X_{i,t} = X_{i,0} + \int_0^t K_i(t-s) b_i(X_s) ds + \int_0^t K_i(t-s)\sigma_i \sqrt{X_{i,s}} dW_{i,s}, \quad i = 1, \dots, d. \tag{5.5.2}$$

The Riccati–Volterra equation (5.4.3) becomes

$$\psi_i(t) = u_i K_i(t) + \int_0^t K_i(t-s) \left(f_i(s) + \psi(s) b^i + \frac{\sigma_i^2}{2} \psi_i(s)^2 \right) ds, \quad i = 1, \dots, d. \tag{5.5.3}$$

The following theorem is our main result on Volterra square-root processes.

Theorem 5.21. *Assume each K_i satisfies (5.2.5) and the shifted kernels $\Delta_h K_i$ satisfy (5.3.3) for all $h \in [0,1]$. Assume also that (5.5.1) holds.*

(i) The stochastic Volterra equation (5.5.2) has a unique in law $\mathbb{R}^d_+$-valued continuous weak solution X for any initial condition $X_0 \in \mathbb{R}^d_+$. For each i, the paths of X_i are Hölder continuous of any order less than $\gamma_i/2$, where γ_i is the constant associated with K_i in (5.2.5).

(ii) For any $u \in (\mathbb{C}^d)^*$ and $f \in L^1_{\mathrm{loc}}(\mathbb{R}_+, (\mathbb{C}^d)^*))$ such that

$$\operatorname{Re} u_i \leq 0 \text{ and } \operatorname{Re} f_i \leq 0 \text{ for all } i = 1, \dots, d,$$

the Riccati–Volterra equation (5.5.3) has a unique global solution $\psi \in L^2_{\mathrm{loc}}(\mathbb{R}_+, (\mathbb{C}^d)^*)$, which satisfies $\operatorname{Re}\psi_i \leq 0$, $i = 1, \dots, d$. Moreover, the exponential-affine transform formula (6.2.13) holds with Y given by (6.2.17)–(6.2.14).

Example 5.22. *A sufficient condition for K_i to satisfy the assumptions of Theorem 5.21 is that it satisfies (5.2.5) and is completely monotone and not identically zero; see Example 5.13. This covers in particular the gamma kernel $t^{\alpha-1}e^{-\beta t}$ with $\alpha \in (\frac{1}{2}, 1]$ and $\beta \geq 0$.*

Proof. Thanks to (5.5.1) and the form of $\sigma(x)$, Theorem 5.12 yields an $\mathbb{R}^d_+$-valued continuous weak solution X of (5.5.2) for any initial condition $X_0 \in \mathbb{R}^d_+$. The stated path regularity then follows from the last statement of Lemma 5.4.

Next, the existence, uniqueness, and non-positivity statement for the Riccati–Volterra equation (5.5.3) is proved in Lemma 5.23 below. Thus in order to apply Theorem 5.16 to obtain the exponential-affine transform formula, it suffices to argue that $\operatorname{Re} Y_t$ is bounded above

on $[0,T]$, since $\exp(Y)$ is then bounded and hence a martingale. This is done using Theorem 5.18, and we start by observing that

$$\pi^{\mathbf{r}}_{h,i}(t) = -\int_0^h \psi^{\mathbf{r}}_i(h-s)L_i(t+ds), \quad t \geq 0,$$

where $\pi_h = \Delta_h\psi * L - \Delta_h(\psi * L)$ and we write $\pi^{\mathbf{r}}_h = \operatorname{Re}\pi_h$ and $\psi^{\mathbf{r}} = \operatorname{Re}\psi$. Due to the assumption (5.3.3) on L_i and since $-\psi^{\mathbf{r}}_i \geq 0$, it follows that $\pi^{\mathbf{r}}_{h,i}$ is nonnegative and nonincreasing.

As in the proof of Theorem 5.12, each K_i satisfies (5.3.6) and (5.3.7). This implies that the total variation bound (5.4.15) holds, so that Theorem 5.18ii yields

$$\operatorname{Re} Y_t = \operatorname{Re}\phi(h) + (\operatorname{Re}\Delta_h f * X)_t + (\Delta_h\psi^{\mathbf{r}} * L)(0)X_t - \pi^{\mathbf{r}}_h(t)X_0 + (d\pi^{\mathbf{r}}_h * X)_t$$

where $h = T - t$ and, since $A^0 = 0$,

$$\phi(h) = \int_0^h \psi(s)b^0 ds.$$

Observe that $\psi^{\mathbf{r}}$, $(\Delta_h\psi^{\mathbf{r}} * L)(0)$, $\operatorname{Re}\Delta_h f$, $-\pi^{\mathbf{r}}_h$, and $d\pi^{\mathbf{r}}_h$ all have nonpositive components. Since b^0 and X take values in $\mathbb{R}^d_+$ we thus get

$$\operatorname{Re} Y_t \leq 0.$$

Thus $\exp(Y)$ is bounded, whence Theorem 5.16 is applicable and the exponential-affine transform formula holds.

It remains to prove uniqueness in law for X. This follows since the law of X is determined by the Laplace functionals $\mathbb{E}[\exp((f*X)_T)]$ as f ranges through, say, all $(\mathbb{R}^d)^*$-valued continuous functions f with nonpositive components, and T ranges through $\mathbb{R}_+$. □

Lemma 5.23. *Assume K is as in Theorem 5.21. Let $u \in (\mathbb{C}^d)^*$ and $f \in L^1_{\mathrm{loc}}(\mathbb{R}_+, (\mathbb{C}^d)^*)$ satisfy*

$$\operatorname{Re} u_i \leq 0 \text{ and } \operatorname{Re} f_i \leq 0 \text{ for all } i = 1, \dots, d.$$

Then the Riccati–Volterra equation (5.5.3) has a unique global solution $\psi \in L^2_{\mathrm{loc}}(\mathbb{R}_+, (\mathbb{C}^d)^*)$, and this solution satisfies $\operatorname{Re}\psi_i \leq 0$, $i = 1, \dots, d$.

Proof. By Theorem 5.32 there exists a unique non-continuable solution $(\psi, T_{\max})$ of (5.5.3). Let $\psi^{\mathbf{r}}$ and $\psi^{\mathbf{i}}$ denote the real and imaginary parts of ψ. They satisfy the equations

$$\psi^{\mathbf{r}}_i = (\operatorname{Re} u_i)K_i + K_i * \left(\operatorname{Re} f_i + \psi^{\mathbf{r}} b^i + \frac{\sigma_i^2}{2}\left((\psi^{\mathbf{r}}_i)^2 - (\psi^{\mathbf{i}}_i)^2\right)\right)$$
$$\psi^{\mathbf{i}}_i = (\operatorname{Im} u_i)K_i + K_i * \left(\operatorname{Im} f_i + \psi^{\mathbf{i}} b^i + \sigma_i^2\psi^{\mathbf{r}}_i\psi^{\mathbf{i}}_i\right)$$

on $[0, T_{\max})$. Moreover, on this interval, $-\psi^{\mathbf{r}}_i$ satisfies the linear equation

$$\chi_i = -(\operatorname{Re} u_i)K_i + K_i * \left(-\operatorname{Re} f_i + \chi b^i + \frac{\sigma_i^2}{2}\left((\psi^{\mathbf{i}}_i)^2 + \chi_i\psi^{\mathbf{r}}_i\right)\right).$$

Due to (5.5.1) and since $\operatorname{Re} u$ and $\operatorname{Re} f$ both have nonpositive components, Theorem 5.36 yields $\psi^{\mathbf{r}}_i \leq 0$, $i = 1, \dots, d$. Next, let $g \in L^2_{\mathrm{loc}}([0, T_{\max}), (\mathbb{R}^d)^*)$ and $h, \ell \in L^2_{\mathrm{loc}}(\mathbb{R}_+, (\mathbb{R}^d)^*)$ be

the unique solutions of the linear equations

$$\begin{aligned} g_i &= |\operatorname{Im} u_i| K_i + K_i * \left(|\operatorname{Im} f_i| + g b^i + \sigma_i^2 \psi_i^{\mathbf{r}} g_i \right) \\ h_i &= |\operatorname{Im} u_i| K_i + K_i * \left(|\operatorname{Im} f_i| + h b^i \right) \\ \ell_i &= (\operatorname{Re} u_i) K_i + K_i * \left(\operatorname{Re} f_i + \ell b^i - \frac{\sigma_i^2}{2} h_i^2 \right). \end{aligned}$$

These solutions exist on $[0, T_{max})$ thanks to Corollary 5.34. We now perform multiple applications of Theorem 5.36. The functions $g \pm \psi^{\mathbf{i}}$ satisfy the equations

$$\chi_i = 2(\operatorname{Im} u_i)^{\pm} K_i + K_i * \left(2(\operatorname{Im} f_i)^{\pm} + \chi b^i + \sigma_i^2 \psi_i^{\mathbf{r}} \chi_i \right)$$

on $[0, T_{\max})$, so $|\psi_i^{\mathbf{i}}| \le g_i$ on $[0, T_{\max})$ for all i. Similarly, $h - g$ satisfies the equation

$$\chi_i = K_i * \left(\chi b^i - \sigma_i^2 \psi_i^{\mathbf{r}} g_i \right)$$

on $[0, T_{\max})$, so $g_i \le h_i$ on $[0, T_{\max})$. Finally, $\psi^{\mathbf{r}} - \ell$ satisfies the equation

$$\chi_i = K_i * \left(\chi b^i + \frac{\sigma_i^2}{2} \left((\psi_i^{\mathbf{r}})^2 + h_i^2 - (\psi_i^{\mathbf{i}})^2 \right) \right),$$

on $[0, T_{\max})$, so $\ell_i \le \psi_i^{\mathbf{r}}$ on $[0, T_{\max})$. In summary, we have shown that

$$\ell_i \le \psi_i^{\mathbf{r}} \le 0 \text{ and } |\psi_i^{\mathbf{i}}| \le h_i \text{ on } [0, T_{\max}) \text{ for } i = 1, \dots, d.$$

Since ℓ and h are global solutions and thus have finite norm on any bounded interval, this implies that $T_{\max} = \infty$ and completes the proof of the lemma. □

5.5.3 The Volterra Heston model

We now consider an affine Volterra process with state space $\mathbb{R} \times \mathbb{R}_+$, which can be viewed as a generalization of the classical Heston stochastic volatility model in finance, and which we refer to as the *Volterra Heston model*. We thus take $d = 2$ and consider the process $X = (\log S, V)$, where the price process S and its variance process V are given by

$$\frac{dS_t}{S_t} = \sqrt{V_t} \, (\sqrt{1 - \rho^2} \, dW_{1,s} + \rho \, dW_{2,s}), \qquad S_0 \in (0, \infty), \tag{5.5.4}$$

and

$$V_t = V_0 + \int_0^t K(t-s) \left(\kappa(\theta - V_s) ds + \sigma \sqrt{V_s} \, dW_{2,s} \right), \tag{5.5.5}$$

with kernel $K \in L^2_{\mathrm{loc}}(\mathbb{R}_+, \mathbb{R})$, a standard Brownian motion $W = (W_1, W_2)$, and parameters $V_0, \kappa, \theta, \sigma \in \mathbb{R}_+$ and $\rho \in [-1, 1]$. Here the notation has been adapted to comply with established conventions in finance. Weak existence and uniqueness of V follows from Theorem 5.21 under suitable conditions on K. This in turn determines S. Moreover, observe that the log-price satisfies

$$\log S_t = \log S_0 - \int_0^t \frac{V_s}{2} \, ds + \int_0^t \sqrt{V_s} \, (\sqrt{1 - \rho^2} \, dW_{1,s} + \rho \, dW_{2,s}).$$

Therefore the process $X = (\log S, V)$ is indeed an affine Volterra process with diagonal kernel $\operatorname{diag}(1, K)$ and coefficients a and b in (5.4.1) given by

$$A^0 = A^1 = 0, \quad A^2 = \begin{pmatrix} 1 & \rho\sigma \\ \rho\sigma & \sigma^2 \end{pmatrix},$$
$$b^0 = \begin{pmatrix} 0 \\ \kappa\theta \end{pmatrix}, \quad B = \begin{pmatrix} 0 & -\frac{1}{2} \\ 0 & -\kappa \end{pmatrix}.$$

The Riccati–Volterra equation (5.4.3) takes the form

$$\psi_1 = u_1 + 1 * f_1, \tag{5.5.6}$$
$$\psi_2 = u_2 K + K * \left(f_2 + \frac{1}{2}\left(\psi_1^2 - \psi_1\right) - \kappa\psi_2 + \frac{1}{2}\left(\sigma^2\psi_2^2 + 2\rho\sigma\psi_1\psi_2\right)\right). \tag{5.5.7}$$

Theorem 5.24. *Assume K satisfies (5.2.5) and the shifted kernels $\Delta_h K$ satisfy (5.3.3) for all $h \in [0, 1]$.*

(i) The stochastic Volterra equation (5.5.4)-(5.5.5) has a unique in law $\mathbb{R} \times \mathbb{R}_+$-valued continuous weak solution $(\log S, V)$ for any initial condition $(\log S_0, V_0) \in \mathbb{R} \times \mathbb{R}_+$. The paths of V are Hölder continuous of any order less than $\gamma/2$, where γ is the constant associated with K in (5.2.5).

(ii) Let $u \in (\mathbb{C}^2)^*$ and $f \in L^1_{\mathrm{loc}}(\mathbb{R}_+, (\mathbb{C}^2)^*)$ be such that

$$\operatorname{Re}\psi_1 \in [0, 1], \ \operatorname{Re} u_2 \leq 0 \text{ and } \operatorname{Re} f_2 \leq 0.$$

where ψ_1 is given by (6.2.9). Then the Riccati–Volterra equation (6.2.10) has a unique global solution $\psi_2 \in L^2_{\mathrm{loc}}(\mathbb{R}_+, \mathbb{C}^*)$, which satisfies $\operatorname{Re}\psi_2 \leq 0$. Moreover, the exponential-affine transform formula (6.2.13) holds with Y given by (6.2.17)–(6.2.14).

(iii) The process S is a martingale.

Proof. As already mentioned above, part (i) follows directly from Theorem 5.21 along with the fact that S is determined by V. Part (iii) is proved in Lemma 5.26 below. The existence, uniqueness, and non-positivity statement for the Riccati–Volterra equation (6.2.10) is proved in Lemma 5.27 below. Thus in order to apply Theorem 5.16 to obtain the exponential-affine transform formula, it suffices to argue that $\exp(Y)$ is a martingale. This is done using Theorem 5.18 and part (iii). As the argument closely parallels that of the proof of Theorem 5.21, we only provide an outline. We use the notation of Theorem 5.18 and Theorem 5.21, in particular π_h and $\pi_h^{\mathbf{r}} = \operatorname{Re}\pi_h$, and let L be the resolvent of the first kind of K. Theorem 5.18 is applicable and gives

$$\begin{aligned}\operatorname{Re} Y_t = {} & \psi_1^{\mathbf{r}}(h)\log S_t + (\operatorname{Re}\Delta_h f_1 * \log S)_t + \operatorname{Re}\phi(h) + (\Delta_h\psi_2^{\mathbf{r}} * L)(0)V_t \\ & + (\operatorname{Re}\Delta_h f_2 * V)_t - \pi_{h,2}^{\mathbf{r}}(t)V_0 + (d\pi_{h,2}^{\mathbf{r}} * V)_t\end{aligned} \tag{5.5.8}$$

where $h = T - t$ and

$$\phi(h) = \kappa\theta \int_0^h \psi_2(s)\, ds.$$

Since $\psi_1^{\mathbf{r}} \in [0,1]$, integration by parts yields

$$\psi_1^{\mathbf{r}}(h)\log S_t + (\operatorname{Re}\Delta_h f_1 * \log S)_t = \psi_1^{\mathbf{r}}(T)\log S_0 + \int_0^t \psi_1^{\mathbf{r}}(T-s)\,d\log S_s$$
$$\leq \psi_1^{\mathbf{r}}(T)\log S_0 + U_t - \frac{1}{2}\langle U\rangle_t,$$

where

$$U_t = \int_0^t \psi_1^{\mathbf{r}}(T-s)\sqrt{V_s}\,(\sqrt{1-\rho^2}\,dW_{1,s} + \rho\,dW_{2,s}).$$

This observation and inspection of signs and monotonicity properties applied to (6.2.19) show that

$$|\exp(Y_t)| = \exp(\operatorname{Re} Y_t) \leq S_0^{\psi_1^{\mathbf{r}}(T)}\exp(U_t - \frac{1}{2}\langle U\rangle_t),$$

where the right-hand side is a true martingale by Lemma 5.26. Thus $\exp(Y)$ is a true martingale, Theorem 5.16 is applicable, and the exponential-affine transform formula holds. □

Example 5.25 (Rough Heston model)**.** *In the fractional case $K(t) = \frac{t^{1-\alpha}}{\Gamma(\alpha)}$ with $\alpha \in (\frac{1}{2}, 1)$ we recover the rough Heston model introduced and studied by [51, 52]. Theorem 5.24 generalizes some of their main results. For instance, with the notation of Example 5.20 and using that $L(dt) = \frac{t^{-\alpha}}{\Gamma(1-\alpha)}dt$, we have*

$$\chi = (\psi_1, I^{1-\alpha}\psi_2),$$

which yields the full Fourier–Laplace functional with integrated log-price and variance,

$$\mathbb{E}\left[e^{u_1\log S_T + u_2 V_T + (f_1 * \log S)_T + (f_2 * V)_T}\right] = \exp\left(\phi(T) + \psi_1(T)\log S_0 + I^{1-\alpha}\psi_2(T)V_0\right),$$

where ψ_1 is given by (6.2.9), and ϕ and ψ_2 solve the fractional Riccati equations

$$\phi' = \kappa\theta\psi_2, \qquad \phi(0) = 0,$$
$$D^\alpha\psi_2 = f_2 + \frac{1}{2}\left(\psi_1^2 - \psi_1\right) + (\rho\sigma\psi_1 - \kappa)\psi_2 + \frac{\sigma^2}{2}\psi_2^2, \qquad I^{1-\alpha}\psi_2(0) = u_2.$$

This extends some of the main results of [51, 52].

We now proceed with the lemmas used in the proof of Theorem 5.24.

Lemma 5.26. *Let $g \in L^\infty(\mathbb{R}_+, \mathbb{R})$ and define $U_t = \int_0^t g(s)\sqrt{V_s}(\sqrt{1-\rho^2}dW_{1,s} + \rho dW_{2,s}))$. Then the stochastic exponential $\exp(U_t - \frac{1}{2}\langle U\rangle_t)$ is a martingale. In particular, S is a martingale.*

Proof. Define $M_t = \exp(U_t - \frac{1}{2}\langle U\rangle_t)$. Since M is a nonnegative local martingale, it is a supermartingale by Fatou's lemma, and it suffices to show that $\mathbb{E}[M_T] \geq 1$ for any $T \in \mathbb{R}_+$. To this end, define stopping times $\tau_n = \inf\{t \geq 0 \colon V_t > n\} \wedge T$. Then M^{τ_n} is a uniformly integrable martingale for each n by Novikov's condition, and we may define probability measures $\mathbb{Q}^n$ by

$$\frac{d\mathbb{Q}^n}{d\mathbb{P}} = M_{\tau_n}.$$

By Girsanov's theorem, the process $dW_t^n = dW_{2,t} - \mathbb{1}_{\{t\leq\tau_n\}}\rho\, g(t)\sqrt{V_t}dt$ is Brownian motion under $\mathbb{Q}^n$, and we have

$$V = V_0 + K * ((\kappa\theta - (\kappa - \rho\sigma g\mathbb{1}_{[\![0,\tau_n]\!]})V)dt + \sigma\sqrt{V}dW^n).$$

Let γ be the constant appearing in (5.2.5) and choose p sufficiently large that $\gamma/2 - 1/p > 0$. Observe that the expression $\kappa\theta - (\kappa - \rho\sigma g(t)\mathbb{1}_{\{t\leq\tau_n(\omega)\}})v$ satisfies a linear growth condition in v, uniformly in (t,ω). Therefore, due to Lemma 5.8 and Remark 5.9, we have the moment bound

$$\sup_{t\leq T}\mathbb{E}_{\mathbb{Q}^n}[|V_t|^p] \leq c$$

for some constant c that does not depend on n. For any real-valued function f, write

$$|f|_{C^{0,\alpha}(0,T)} = \sup_{0\leq s<t\leq T}\frac{|f(t)-f(s)|}{|t-s|^\alpha}$$

for its α-Hölder seminorm. We then get

$$\begin{aligned}\mathbb{Q}^n(\tau_n < T) &\leq \mathbb{Q}^n\Big(\sup_{t\leq T} V_t > n\Big)\\ &\leq \mathbb{Q}^n\Big(V_0 + |V|_{C^{0,0}(0,T)} > n\Big)\\ &\leq \left(\frac{1}{n-V_0}\right)^p \mathbb{E}_{\mathbb{Q}^n}\left[|V|^p_{C^{0,0}(0,T)}\right]\\ &\leq \left(\frac{1}{n-V_0}\right)^p c'\end{aligned}$$

for a constant c' that does not depend on n, using Lemma 5.4 with $\alpha = 0$ for the last inequality. We deduce that

$$\mathbb{E}_{\mathbb{P}}[M_T] \geq \mathbb{E}_{\mathbb{P}}\left[M_T\mathbb{1}_{\{\tau_n=T\}}\right] = \mathbb{Q}^n(\tau_n = T) \geq 1 - \left(\frac{1}{n-V_0}\right)^p c',$$

and sending n to infinity yields $\mathbb{E}_{\mathbb{P}}[M_T] \geq 1$. This completes the proof. □

Lemma 5.27. *Assume K is as in Theorem 5.24. Let $u \in (\mathbb{C}^2)^*$ and $f \in L^1_{\mathrm{loc}}(\mathbb{R}_+,(\mathbb{C}^2)^*)$ be such that*

$$\operatorname{Re}\psi_1 \in [0,1],\ \operatorname{Re}u_2 \leq 0 \text{ and } \operatorname{Re}f_2 \leq 0,$$

with ψ_1 given by (6.2.9). Then the Riccati–Volterra equation (6.2.10) has a unique global solution $\psi_2 \in L^2_{\mathrm{loc}}(\mathbb{R}_+,\mathbb{C}^*)$, which satisfies $\operatorname{Re}\psi_2 \leq 0$.

Proof. The proof parallels that of Lemma 5.23. For any complex number z, we denote by $z^{\mathbf{r}}$ and $z^{\mathbf{i}}$ the real and imaginary parts of z. We rewrite equation (6.2.10) for ψ_2 as

$$\psi_2 = u_2K + K * \left(f_2 + \frac{1}{2}(\psi_1^2 - \psi_1) + (\rho\sigma\psi_1 - \kappa)\psi_2 + \frac{\sigma^2}{2}\psi_2^2\right). \tag{5.5.9}$$

By Theorem 5.32 there exists a unique non-continuable solution $(\psi_2, T_{\max})$ of (5.5.9). The functions $\psi_2^{\mathbf{r}}$ and $\psi_2^{\mathbf{i}}$ satisfy the equations

$$\begin{aligned}\psi_2^{\mathbf{r}} = u_2^{\mathbf{r}}K + K * \Big(&f_2^{\mathbf{r}} + \frac{1}{2}((\psi_1^{\mathbf{r}})^2 - \psi_1^{\mathbf{r}} - (\psi_1^{\mathbf{i}})^2) - \rho\sigma\psi_1^{\mathbf{i}}\psi_2^{\mathbf{i}}\\ &- \frac{\sigma^2}{2}(\psi_2^{\mathbf{i}})^2 + (\rho\sigma\psi_1^{\mathbf{r}} - \kappa)\psi_2^{\mathbf{r}} + \frac{\sigma^2}{2}(\psi_2^{\mathbf{r}})^2\Big)\\ \psi_2^{\mathbf{i}} = u_2^{\mathbf{i}}K + K * \Big(&f_2^{\mathbf{i}} + \frac{1}{2}\left(2\psi_1^{\mathbf{r}}\psi_1^{\mathbf{i}} - \psi_1^{\mathbf{i}}\right) + \rho\sigma\psi_1^{\mathbf{i}}\psi_2^{\mathbf{r}} + (\rho\sigma\psi_1^{\mathbf{r}} - \kappa + \sigma^2\psi_2^{\mathbf{r}})\psi_2^{\mathbf{i}}\Big)\end{aligned}$$

on $[0, T_{\max})$. After some rewriting, we find that on $[0, T_{\max})$, $-\psi_2^{\mathbf{r}}$ satisfies the linear equation

$$\chi = -u_2^{\mathbf{r}} K + K * \Big(- f_2^{\mathbf{r}} + \frac{1}{2}(\psi_1^{\mathbf{r}} - (\psi_1^{\mathbf{r}})^2 + (1-\rho^2)(\psi_1^{\mathbf{i}})^2) + \frac{(\sigma\psi_2^{\mathbf{i}} + \rho\psi_1^{\mathbf{i}})^2}{2} - \left(\rho\sigma\psi_1^{\mathbf{r}} - \kappa + \frac{\sigma^2}{2}\psi_2^{\mathbf{r}}\right)\chi\Big).$$

Due to (5.5.1) and since $\psi_1^{\mathbf{r}}, |\rho| \in [0,1]$, and $f_2^{\mathbf{r}}$ and $u_2^{\mathbf{r}}$ are nonpositive, Theorem 5.36 yields $\psi_2^{\mathbf{r}} \leq 0$ on $[0, T_{\max})$.

Now, if $\sigma = 0$, then (5.5.9) is a linear Volterra equation and thus admits a unique global solution $\psi_2 \in L^2_{\mathrm{loc}}(\mathbb{R}_+, \mathbb{C}^*)$ by Corollary 5.34. Therefore it suffices to consider the case $\sigma > 0$.

Following the proof of Lemma 5.23, we let $g \in L^2_{\mathrm{loc}}([0, T_{\max}), (\mathbb{R})^*)$ and $h, \ell \in L^2_{\mathrm{loc}}(\mathbb{R}_+, (\mathbb{R})^*)$ be the unique solutions of the linear equations

$$g = |u_2^{\mathbf{i}}|K + \left|\rho\sigma^{-1}u_1^{\mathbf{i}}\right| + K * \Big(\left| \rho\sigma^{-1}(L * f_1^{\mathbf{i}}) + f_2^{\mathbf{i}} + \frac{\psi_1^{\mathbf{i}}}{2}\left(2(1-\rho^2)\psi_1^{\mathbf{r}} - 1 + \frac{2\kappa\rho}{\sigma}\right)\right| + (\rho\sigma\psi_1^{\mathbf{r}} - \kappa + \sigma^2\psi_2^{\mathbf{r}})g\Big)$$

$$h = |u_2^{\mathbf{i}}|K + \left|\rho\sigma^{-1}u_1^{\mathbf{i}}\right| + K * \Big(\left| \rho\sigma^{-1}(L * f_1^{\mathbf{i}}) + f_2^{\mathbf{i}} + \frac{\psi_1^{\mathbf{i}}}{2}\left(2(1-\rho^2)\psi_1^{\mathbf{r}} - 1 + \frac{2\kappa\rho}{\sigma}\right)\right| + (\rho\sigma\psi_1^{\mathbf{r}} - \kappa)h\Big)$$

$$\ell = u_2^{\mathbf{r}} K + K * \Big(f_2^{\mathbf{r}} + \frac{1}{2}((\psi_1^{\mathbf{r}})^2 - \psi_1^{\mathbf{r}} - (\psi_1^{\mathbf{i}})^2) - |\rho\sigma\psi_1^{\mathbf{i}}|\left(h + \left|\rho\psi_1^{\mathbf{i}}\sigma^{-1}\right|\right) - \frac{\sigma^2}{2}\left(h + \left|\rho\psi_1^{\mathbf{i}}\sigma^{-1}\right|\right)^2 + (\rho\sigma\psi_1^{\mathbf{r}} - \kappa)\ell\Big).$$

These solutions exist on $[0, T_{\max})$ thanks to Corollary 5.34. We now perform multiple applications of Theorem 5.36. The functions $g \pm (\psi_2^{\mathbf{i}} + (\rho\psi_1^{\mathbf{i}}\sigma^{-1}))$ satisfy the equations

$$\chi = 2(u_2^{\mathbf{i}})^{\pm} K + 2\left(\rho\sigma^{-1}u_1^{\mathbf{i}}\right)^{\pm} + K * \Big(2\left(\rho\sigma^{-1}(L * f_1^{\mathbf{i}}) + f_2^{\mathbf{i}} + \frac{\psi_1^{\mathbf{i}}}{2}\left(2(1-\rho^2)\psi_1^{\mathbf{r}} - 1 + \frac{2\kappa\rho}{\sigma}\right)\right)^{\pm} + (\rho\sigma\psi_1^{\mathbf{r}} - \kappa + \sigma^2\psi_2^{\mathbf{r}})\chi\Big)$$

on $[0, T_{max})$, so that $0 \leq |\psi_2^{\mathbf{i}} + \left(\rho\psi_1^{\mathbf{i}}\sigma^{-1}\right)| \leq g$ on $[0, T_{max})$. Similarly, the function $h - g$ satisfies the equation

$$\chi = K * \left(-\sigma^2\psi_2^{\mathbf{r}} g + (\rho\sigma\psi_1^{\mathbf{r}} - \kappa)\chi\right)$$

on $[0, T_{\max})$, so that $g \leq h$ on $[0, T_{\max})$. This yields $|\psi_2^{\mathbf{i}}| \leq h + \left|\rho\psi_1^{\mathbf{i}}\sigma^{-1}\right|$ on $[0, T_{max})$. Finally, the function $\psi_2^{\mathbf{r}} - \ell$ satisfies the linear equation

$$\chi = K * \Big(|\rho\sigma\psi_1^{\mathbf{i}}|\left(h + \left|\rho\psi_1^{\mathbf{i}}\sigma^{-1}\right|\right) - \rho\sigma\psi_1^{\mathbf{i}}\psi_2^{\mathbf{i}} + \frac{\sigma^2}{2}\left(\left(h + \left|\rho\psi_1^{\mathbf{i}}\sigma^{-1}\right|\right)^2 - \left(\psi_2^{\mathbf{i}}\right)^2 + (\psi_2^{\mathbf{r}})^2\right) + (\rho\sigma\psi_1^{\mathbf{r}} - \kappa)\chi\Big)$$

on $[0, T_{\max})$, so that $\ell \leq \psi_2^{\mathbf{r}} \leq 0$ on $[0, T_{\max})$. Since h and ℓ are global solutions and thus have finite norm on any bounded interval, this implies that $T_{\max} = \infty$ and completes the proof of the lemma. □

We conclude this section with a remark on an alternative variant of the Volterra Heston model in the spirit of [70].

Example 5.28. *Let $\widetilde{K}$ denotes a scalar locally square integrable non-negative kernel. Consider the following variant of the Volterra Heston model*

$$\begin{aligned} dS_t &= S_t\sqrt{\widetilde{V}_t}dB_t, & S_0 &\in (0,\infty),\\ dV_t &= \kappa(\theta - V_t)dt + \sigma\sqrt{V_t}dB_t^{\perp}, & V_0 &\geq 0,\\ \widetilde{V}_t &= \widetilde{V}_0 + (\widetilde{K} * V)_t, \end{aligned}$$

where B and $B^{\perp}$ are independent Brownian motions. Since $\widetilde{K}$ is nonnegative, one readily sees that there exists a unique strong solution taking values in $\mathbb{R}\times\mathbb{R}^2_+$. The 3-dimensional process $X = (\log S, V, \widetilde{V})$ is an affine Volterra process with

$$K = \operatorname{diag}\left[(1,1,\widetilde{K})\right], \quad b^0 = \begin{pmatrix} 0\\ \kappa\theta\\ 0\end{pmatrix}, \quad B = \begin{pmatrix} 0 & 0 & -\frac{1}{2}\\ 0 & -\kappa & 0\\ 0 & 1 & 0\end{pmatrix},$$

$$A^0 = 0, \quad A^1 = 0, \quad A^2 = \operatorname{diag}\left[(0,\sigma^2,0)\right], \quad A^3 = \operatorname{diag}\left[(1,0,0)\right].$$

The Riccati–Volterra equation (5.4.3) reads

$$\begin{aligned} \psi_1' &= f_1, & \psi_1(0) &= u_1,\\ \psi_2' &= f_2 + \psi_3 - \kappa\psi_2 + \frac{\sigma^2}{2}\psi_2^2, & \psi_2(0) &= u_2,\\ \psi_3 &= u_3\widetilde{K} + \widetilde{K} * \left(f_3 + \frac{1}{2}\psi_1(\psi_1 - 1)\right). \end{aligned}$$

Under suitable conditions the solution exists and is unique, and the process e^Y with Y given by (6.2.17)–(6.2.14) is a true martingale. Hence by Theorem 5.16 the exponential-affine transform formula (6.2.13) holds. We omit the details. In particular, for $f \equiv 0$ we get, using Example 5.20,

$$\chi(t) = (\psi * L)(t) = \left(u_1, \psi_2(t), u_3 + \frac{(u_1^2 - u_1)t}{2}\right)$$

and

$$\mathbb{E}\left[e^{u_1\log S_T + u_2 V_T + u_3\widetilde{V}_T}\right] = \exp\left(\phi(T) + u_1\log S_0 + \psi_2(T)V_0 + \left(u_3 + \frac{(u_1^2-u_1)T}{2}\right)\widetilde{V}_0\right),$$

where ϕ and ψ_2 solve

$$\begin{aligned} \phi' &= \kappa\theta\psi_2, & \phi(0) &= 0,\\ \psi_2' &= u_3\widetilde{K} + \widetilde{K} * \frac{(u_1^2 - u_1)}{2} - \kappa\psi_2 + \frac{\sigma^2}{2}\psi_2^2, & \psi_2(0) &= u_2. \end{aligned}$$

Setting $\widetilde{K} = \frac{t^{\alpha-1}}{\Gamma(\alpha)}$ and $u_2 = 0$, this formula agrees with [70, Theorem 2.1]. If B and $B^{\perp}$ are correlated one loses the affine property, as highlighted in [70, Remark 2.2].

5.6 Supporting results

5.6.1 Weak existence for stochastic Volterra equations

Lemma 5.29. *Assume b and σ are Lipschitz continuous and the components of K satisfy (5.2.5). Then (5.1.1) admits a unique continuous strong solution X for any initial condition $X_0 \in \mathbb{R}^d$.*

Proof. The proof parallels that of [96, Proposition 2.1], using a Picard iteration scheme. We define $X^0 \equiv 0$ and for each $n \in \mathbb{N}$,

$$X^n = X_0 + K * (b(X^{n-1})dt + \sigma(X^{n-1})dW).$$

For any $p \geq 2$ and $T \geq 0$ we may combine the Lipschitz property of b and σ with the Jensen and BDG inequalities to obtain

$$\mathbb{E}[|X_t^n - X_t^{n-1}|^p] \leq c \int_0^t |K(t-s)|^2 \, \mathbb{E}[|X_s^{n-1} - X_s^{n-2}|^p] ds, \qquad t \leq T, \ n \geq 2,$$

where one can take $c = 2^{p-1}(T^{p/2} + C_p)c_{\mathrm{LIP}} \|K\|_{L^2(0,T)}^{p-2}$ with C_p the constant from the BDG inequality and c_{LIP} a common Lipschitz constant for b and σ. The extended Gronwall's lemma given in [40, Lemma 15] now yields that the series

$$\sum_{n \geq 2} \mathbb{E}[|X_t^n - X_t^{n-1}|^p]$$

converges, uniformly in $t \in [0,T]$. Consequently, for each $t \in [0,T]$ there exists a random variable X_t such that $X_t^n \to X_t$ in L^p, and one even has $\sup_{t \in [0,T]} \mathbb{E}[|X_t^n - X_t|^p] \to 0$. Passing to the limit in the identity

$$X_t^n - X_0 - \int_0^t K(t-s)(b(X_s^n)ds + \sigma(X_s^n)dW_s) = X_t^n - X_t^{n+1}$$

then shows that the random variables $\{X_t \colon t \in [0,T]\}$ satisfy (5.1.1) for each $t \in [0,T]$. Furthermore, it follows that $\sup_{t \leq T} \mathbb{E}[|X_t|^p]$ is finite, so that X has a continuous version by Lemma 5.4. This version is the desired strong solution.

To prove uniqueness, let X and X' be two solutions and define $f(t) = \mathbb{E}[|X_t - X_t'|^2]$ for $t \in [0,T]$, which is finite by Lemma 5.8. Relying on the Lipschitz continuity of b and σ, one find that f satisfies the inequality

$$f \leq c'|K|^2 * f \quad \text{on } [0,T]$$

for some constant c'. Arguing as in the proof of Lemma 5.8 we deduce that $f = 0$. This proves uniqueness. □

Lemma 5.30. *Fix an initial condition $X_0 \in \mathbb{R}^d$ and a constant c_{LG}. Let $\mathcal{X}$ denote the set of all continuous processes X that solve (5.1.1) for some continuous coefficients b and σ satisfying the linear growth bound (5.3.1) with the given constant c_{LG}. Then $\mathcal{X}$ is tight, meaning that the family $\{\text{law of } X \colon X \in \mathcal{X}\}$ of laws on $C(\mathbb{R}_+; \mathbb{R}^d)$ is tight.*

Proof. Let $X \in \mathcal{X}$ be any solution of (5.1.1) for some continuous b and σ satisfying the linear growth bound (5.3.1). Lemma 5.8 implies that $\sup_{u\leq T} \mathbb{E}[|b(X_u)|^p]$ and $\sup_{u\leq T} \mathbb{E}[|\sigma(X_u)|^p]$ are bounded above by a constant that only depends on $|X_0|$, $\|K\|_{L^2(0,T)}$, c_{LG}, p, and T. Therefore, since the components of K satisfy (5.2.5), we may apply Lemma 5.4 to obtain

$$\mathbb{E}\left[\left(\sup_{0\leq s<t\leq T} \frac{|X_t - X_s|}{|t-s|^\alpha}\right)^p\right] \leq c$$

for all $\alpha \in [0, \bar{\gamma}/2 - 1/p)$, where $\bar{\gamma}$ is the smallest of the constants γ appearing in (5.2.5) for the components of K, and where c is a constant that only depends on $|X_0|$, $\|K\|_{L^2(0,T)}$, c_{LG}, p, and T, but not on s or t, nor on the specific choice of $X \in \mathcal{X}$. Choosing p so that $\bar{\gamma}p/2 > 1$, and using that closed Hölder balls are compact in $C(\mathbb{R}_+;\mathbb{R}^d)$, it follows that $\mathcal{X}$ is tight. □

Lemma 5.31. *Assume that K admits a resolvent of the first kind L. For each $n \in \mathbb{N}$, let X^n be a weak solution of (5.1.1) with b and σ replaced by some continuous coefficients b^n and σ^n that satisfy (5.3.1) with a common constant c_{LG}. Assume that $b^n \to b$ and $\sigma^n \to \sigma$ locally uniformly for some coefficients b and σ, and that $X^n \Rightarrow X$ for some continuous process X. Then X is a weak solution of (5.1.1).*

Proof. Lemma 5.6 yields the identity

$$L * (X^n - X_0) = \int b^n(X^n)dt + \int \sigma^n(X^n)dW.$$

Moreover, [69, Theorem 3.6.1(ii) and Corollary 3.6.2(iii)] imply that the map

$$F \mapsto L * (F - F(0))$$

is continuous from $C(\mathbb{R}_+;\mathbb{R}^d)$ to itself. Using also the locally uniform convergence of b^n and σ^n, the continuous mapping theorem shows that the martingales

$$M^n = \int \sigma^n(X^n)dW = L * (X^n - X_0) - \int b^n(X^n)dt$$

converge weakly to some limit M, that the quadratic variations $\langle M^n\rangle = \int \sigma^n\sigma^{n\top}(X^n)dt$ converge weakly to $\int \sigma\sigma^\top(X)dt$, and that $\int b^n(X^n)dt$ converge weakly to $\int b(X)dt$.

Consider any $s < t$, $m \in \mathbb{N}$, any bounded continuous function $f\colon \mathbb{R}^m \to \mathbb{R}$, and any $0 \leq t_1 \leq \cdots \leq t_m \leq s$. Observe that the moment bound in Lemma 5.8 is uniform in n since the X^n satisfy the linear growth condition (5.3.1) with a common constant. Using [21, Theorem 3.5], one then readily shows that

$$\mathbb{E}[f(X_{t_1},\dots,X_{t_m})(M_t - M_s)] = \lim_{n\to\infty} \mathbb{E}[f(X^n_{t_1},\dots,X^n_{t_m})(M^n_t - M^n_s)] = 0,$$

and similarly for the increments of $M^n_i M^n_j - \langle M^n_i, M^n_j\rangle$. It follows that M is a martingale with respect to the filtration generated by X with quadratic variation $\langle M\rangle = \int \sigma\sigma^\top(X)dt$. This carries over to the usual augmentation. Enlarging the probability space if necessary, we may now construct a d-dimensional Brownian motion $\overline{W}$ such that $M = \int \sigma(X)d\overline{W}$.

The above shows that $L * (X - X_0) = \int b(X)dt + \int \sigma(X)d\overline{W}$. The converse direction of Lemma 5.6 then yields $X = X_0 + K * (b(X)dt + \sigma(X)d\overline{W})$, that is, X solves (5.1.1) with the Brownian motion $\overline{W}$. □

Proof of Theorem 5.10. Using [73, Proposition 1.1] we choose Lipschitz coefficients b^n and σ^n that satisfy the linear growth bound (5.3.1) with c_{LG} replaced by $2c_{\mathrm{LG}}$, and converge locally uniformly to b and σ as $n \to \infty$. Let X^n be the unique continuous strong solution of (5.1.1) with b and σ replaced by b^n and σ^n; see Lemma 5.29. Due to Lemma 5.30 the sequence $\{X^n\}$ is tight, so after passing to a subsequence we have $X^n \Rightarrow X$ for some continuous process X. The result now follows from Lemma 5.31. □

5.6.2 Local solutions of Volterra integral equations

Fix a kernel $K \in L^2_{\mathrm{loc}}(\mathbb{R}_+, \mathbb{R}^{d\times d})$ along with functions $g\colon \mathbb{R}_+ \to \mathbb{C}^d$ and $p\colon \mathbb{R}_+ \times \mathbb{C}^d \to \mathbb{C}^d$, and consider the Volterra integral equation

$$\psi = g + K * p(\,\cdot\,, \psi). \tag{5.6.1}$$

A *non-continuable solution* of (5.6.1) is a pair $(\psi, T_{\max})$ with $T_{\max} \in (0, \infty]$ and $\psi \in L^2_{\mathrm{loc}}([0, T_{\max}), \mathbb{C}^d)$, such that ψ satisfies (5.6.1) on $[0, T_{\max})$ and $\|\psi\|_{L^2(0,T_{\max})} = \infty$ if $T_{\max} < \infty$. If $T_{\max} = \infty$ we call ψ a *global solution* of (5.6.1). With some abuse of terminology we call a non-continuable solution $(\psi, T_{\max})$ *unique* if for any $T \in \mathbb{R}_+$ and $\widetilde{\psi} \in L^2([0,T], \mathbb{C}^d)$ satisfying (5.6.1) on $[0,T]$, we have $T < T_{\max}$ and $\widetilde{\psi} = \psi$ on $[0,T]$.

Theorem 5.32. *Assume that $g \in L^2_{\mathrm{loc}}(\mathbb{R}_+, \mathbb{C}^d)$, $p(\,\cdot\,, 0) \in L^1_{\mathrm{loc}}(\mathbb{R}_+, \mathbb{C}^d)$, and that for all $T \in \mathbb{R}_+$ there exist a positive constant Θ_T and a function $\Pi_T \in L^2([0,T], \mathbb{R}_+)$ such that*

$$|p(t,x) - p(t,y)| \le \Pi_T(t)|x-y| + \Theta_T|x-y|(|x| + |y|), \qquad x, y \in \mathbb{C}^d,\ t \le T. \tag{5.6.2}$$

The Volterra integral equation (5.6.1) has a unique non-continuable solution $(\psi, T_{\max})$. If g and p are real-valued, then so is ψ.

Remark 5.33. If $K \in L^{2+\varepsilon}_{\mathrm{loc}}$ for some $\varepsilon > 0$, then it is possible to apply [69, Theorem 12.4.4] with $p = 2 + \varepsilon$ to get existence.

Proof. We focus on the complex-valued case; for the real-valued case, simply replace $\mathbb{C}^d$ by $\mathbb{R}^d$ below. We first prove that a solution exists for small times. Let $\rho \in (0,1]$ and $\varepsilon > 0$ be constants to be specified later, and define

$$B_{\rho,\varepsilon} = \{\psi \in L^2([0,\rho], \mathbb{C}^d)\colon \|\psi\|_{L^2(0,\rho)} \le \varepsilon\}.$$

Consider the map F acting on elements $\psi \in B_{\rho,\varepsilon}$ by

$$F(\psi) = g + K * p(\,\cdot\,, \psi).$$

We write $\|\cdot\|_q = \|\cdot\|_{L^q(0,\rho)}$ for brevity in the following computations. The growth condition (5.6.2) along with the Young, Cauchy–Schwarz, and triangle inequalities yield for $\psi, \widetilde{\psi} \in B_{\rho,\varepsilon}$

$$\begin{aligned} \|F(\psi)\|_2 &\le \|g\|_2 + \|K\|_2 \|p(\,\cdot\,, \psi)\|_1 \\ &\le \|g\|_2 + \|K\|_2 \left(\|p(\,\cdot\,, 0)\|_1 + \|\Pi_1\|_2 \|\psi\|_2 + \Theta_1 \|\psi\|_2^2 \right) \\ &\le \|g\|_2 + \|K\|_2 \left(\|p(\,\cdot\,, 0)\|_1 + \|\Pi_1\|_{L^2(0,1)} \varepsilon + \Theta_1 \varepsilon^2 \right) \end{aligned} \tag{5.6.3}$$

and

$$\begin{aligned}\|F(\psi) - F(\widetilde{\psi})\|_2 &\leq \|K\|_2 \left(\|\Pi_1\|_2 + \Theta_1 \left(\|\psi\|_2 + \|\widetilde{\psi}\|_2\right)\right) \|\psi - \widetilde{\psi}\|_2 \\ &\leq \|K\|_2 \left(\|\Pi_1\|_{L^2(0,1)} + 2\Theta_1\varepsilon\right) \|\psi - \widetilde{\psi}\|_2.\end{aligned}$$

Choose $\varepsilon > 0$ so that $1 + \frac{\varepsilon}{2} + \|\Pi_1\|_{L^2(0,1)}\varepsilon + \Theta_1\varepsilon^2 < 2$ and $\varepsilon\left(\|\Pi_1\|_{L^2(0,1)} + 2\Theta_1\varepsilon\right) < 2$. Then choose $\rho > 0$ so that $\|g\|_2 \vee \|K\|_2 \vee \|p(\,\cdot\,, 0)\|_1 \leq \varepsilon/2$. This yields

$$\|F(\psi)\|_2 \leq \frac{\varepsilon}{2}\left(1 + \frac{\varepsilon}{2} + \|\Pi_1\|_{L^2(0,1)}\varepsilon + \Theta_1\varepsilon^2\right) \leq \varepsilon$$

and

$$\|F(\psi) - F(\widetilde{\psi})\|_2 \leq \kappa\|\psi - \widetilde{\psi}\|_2, \qquad \kappa = \frac{\varepsilon}{2}\left(\|\Pi_1\|_{L^2(0,1)} + 2\Theta_1\varepsilon\right) < 1.$$

Thus F maps $B_{\rho,\varepsilon}$ to itself and is a contraction there, so Banach's fixed point theorem implies that F has a unique fixed point $\psi \in B_{\rho,\varepsilon}$, which is a solution of (5.6.1).

We now extend this to a unique non-continuable solution of (5.6.1). Define the set

$$J = \{T \in \mathbb{R}_+ \colon \text{(5.6.1) has a solution } \psi \in L^2([0,T], \mathbb{C}^d) \text{ on } [0,T]\}.$$

Then $0 \in J$, and if $T \in J$ and $0 \leq S \leq T$, then $S \in J$. Thus J is a nonempty interval. Moreover, J is open in $\mathbb{R}_+$. Indeed, pick $T \in J$, let ψ be a solution on $[0,T]$, and set

$$h(t) = g(T+t) + \int_0^T K(T+t-s)p(s,\psi(s))ds, \quad t \geq 0,$$

which lies in $L^2_{\mathrm{loc}}(\mathbb{R}_+, \mathbb{C}^d)$ by a calculation similar to (5.6.3). By what we already proved, the equation

$$\chi = h + K * p(\,\cdot\, + T, \chi)$$

admits a solution $\chi \in L^2([0,\rho], \mathbb{C}^d)$ on $[0,\rho]$ for some $\rho > 0$. Defining $\psi(t) = \chi(t-T)$ for $t \in (T, T+\rho]$, one verifies that ψ solves (5.6.1) on $[0, T+\rho]$. Thus $T + \rho \in J$, so J is open in $\mathbb{R}_+$ and hence of the form $J = [0, T_{\max})$ for some $0 < T_{\max} \leq \infty$ with $T_{\max} \notin J$. This yields a non-continuable solution $(\psi, T_{\max})$.

It remains to argue uniqueness. Pick $T \in \mathbb{R}_+$ and $\widetilde{\psi} \in L^2([0,T], \mathbb{C}^d)$ satisfying (5.6.1) on $[0,T]$. Then $T \in J$, so $T < T_{\max}$. Let S be the supremum of all $S' \leq T$ such that $\widetilde{\psi} = \psi$ on $[0,S']$. Then $\widetilde{\psi} = \psi$ on $[0,S]$ (almost everywhere, as elements of L^2). If $S < T$, then for $\rho > 0$ sufficiently small we have $0 < \|\psi - \widetilde{\psi}\|_{L^2(0,S+\rho)} \leq \frac{1}{2}\|\psi - \widetilde{\psi}\|_{L^2(0,S+\rho)}$, a contradiction. Thus $S = T$, and uniqueness is proved. □

Corollary 5.34. *Let $K \in L^2_{\mathrm{loc}}(\mathbb{R}_+, \mathbb{C}^{d\times d})$, $F \in L^2_{\mathrm{loc}}(\mathbb{R}_+, \mathbb{C}^d)$ and $G \in L^2_{\mathrm{loc}}(\mathbb{R}_+, \mathbb{C}^{d\times d})$. Suppose that $p\colon \mathbb{R}_+ \times \mathbb{C}^d \to \mathbb{C}^d$ is a Lipschitz continuous function in the second argument such that $p(\,\cdot\,, 0) \in L^2_{\mathrm{loc}}(\mathbb{R}_+, \mathbb{C}^d)$. Then the equation*

$$\chi = F + K * (Gp(\,\cdot\,, \chi))$$

has a unique global solution $\chi \in L^2_{\mathrm{loc}}(\mathbb{R}_+, \mathbb{C}^d)$. Moreover, if K and F are continuous on $[0,\infty)$ then χ is also continuous on $[0,\infty)$ and $\chi(0) = F(0)$.

Proof. Theorem 5.32 implies the existence and uniqueness of a non-continuable solution $(\chi, T_{\max})$. If K and F are continuous on $[0,\infty)$, then this solution is continuous on $[0, T_{\max})$

with $\chi(0) = F(0)$. To prove that $T_{\max} = \infty$, observe that

$$|\chi| \leq |F| + |K| * (|G|(|p(\cdot\,, 0)| + \Theta|\chi|)) \tag{5.6.4}$$

for some positive constant Θ. Define the scalar non-convolution Volterra kernel $K'(t,s) = \Theta|K(t-s)||G(s)|\mathbb{1}_{s\leq t}$. This is a Volterra kernel in the sense of [69, Definition 9.2.1] and

$$\int_0^T \int_0^T \mathbb{1}_{s\leq t}|K(t-s)|^2|G(s)|^2 ds\, dt \leq \| K\|^2_{L^2(0,T)} \| G\|^2_{L^2(0,T)} \tag{5.6.5}$$

for all $T > 0$, by Young's inequality. Thus by [69, Proposition 9.2.7(iii)], K' is of type L^2_{loc}, see [69, Definition 9.2.2]. In addition, it follows from [69, Corollary 9.3.16] that $-K'$ admits a resolvent of type L^2_{loc} in the sense of [69, Definition 9.3.1], which we denote by R'. Since $-K'$ is nonpositive, it follows from [69, Proposition 9.8.1] that R' is also nonpositive. The Gronwall type inequality in [69, Lemma 9.8.2] and (5.6.4) then yield

$$|\chi(t)| \leq f'(t) - \int_0^t R'(t,s) f'(s)\, ds \tag{5.6.6}$$

for $t \in [0, T_{\max}]$, where

$$f'(t) = |F(t)| + \int_0^t |K(t-s)|\, |G(s)|\, |p(s,0)| ds.$$

Since the function on the right-hand side of (5.6.6) is in $L^2_{\mathrm{loc}}(\mathbb{R}_+, \mathbb{R})$ due to [69, Theorem 9.3.6], we conclude that $T_{\max} = \infty$. □

5.6.3 Invariance results for Volterra integral equations

Lemma 5.35. *Fix $T < \infty$. Let $u \in \mathbb{C}^d$, $G \in L^2([0,T], \mathbb{C}^{d\times d})$, as well as $F^n \in L^2([0,T], \mathbb{C}^d)$ and $K^n \in L^2([0,T], \mathbb{C}^{d\times d})$ for $n = 0, 1, 2, \ldots$. For each n, there exists a unique element $\chi^n \in L^2([0,T], \mathbb{C}^{d\times d})$ such that*

$$\chi^n = F^n + K^n * (G\chi^n).$$

Moreover, if $F^n \to F^0$ and $K^n \to K^0$ in $L^2(0,T)$, then $\chi^n \to \chi^0$ in $L^2(0,T)$.

Proof. For any $K \in L^2([0,T], \mathbb{C}^{d\times d})$, define $K'(t,s) = K(t-s)G(s)\mathbb{1}_{s\leq t}$. Arguing as in the proof of Corollary 5.34, K' is a Volterra kernel of type L^2 on $(0,T)$ since (5.6.5) still holds by Young's inequality. In particular,

$$|||K'|||_{L^2(0,T)} \leq \| K\|_{L^2(0,T)} \| G\|_{L^2(0,T)}, \tag{5.6.7}$$

where $|||\cdot|||_{L^2(0,T)}$ is defined in [69, Definition 9.2.2]. Invoking once again [69, Corollary 9.3.16], $-K'$ admits a resolvent R' of type L^2 on $(0,T)$. Due to [69, Theorem 9.3.6], the unique solution in $L^2(0,T)$ of the equation

$$\chi(t) = F(t) + \int_0^t K'(t,s)\chi(s) ds, \quad t \in [0,T],$$

for a given $F \in L^2([0,T],\mathbb{C}^d)$, is

$$\chi(t) = F(t) - \int_0^t R'(t,s)F(s)ds, \quad t \in [0,T].$$

This proves the existence and uniqueness statement for the χ^n. Next, assume $F^n \to F^0$ and $K^n \to K^0$ in $L^2(0,T)$. Applying (5.6.7) with $K = K^n - K^0$ shows that $(K')^n \to (K')^0$ with respect to the norm $|||\cdot|||_{L^2(0,T)}$. An application of [69, Corollary 9.3.12] now shows that $\chi^n \to \chi^0$ in $L^2(0,T)$ as claimed. □

Theorem 5.36. *Assume $K \in L^2_{\mathrm{loc}}(\mathbb{R}_+,\mathbb{R}^{d\times d})$ is diagonal with scalar kernels K_i on the diagonal. Assume each K_i satisfies (5.2.5) and the shifted kernels $\Delta_h K_i$ satisfy (5.3.3) for all $h \in [0,1]$. Let $u, v \in \mathbb{R}^d$, $F \in L^1_{\mathrm{loc}}(\mathbb{R}_+,\mathbb{R}^d)$ and $G \in L^2_{\mathrm{loc}}(\mathbb{R}_+,\mathbb{R}^{d\times d})$ be such that $u_i, v_i \geq 0$, $F_i \geq 0$, and $G_{ij} \geq 0$ for all $i,j = 1,\dots,d$ and $i \neq j$. Then the linear Volterra equation*

$$\chi = Ku + v + K * (F + G\chi) \tag{5.6.8}$$

has a unique solution $\chi \in L^2_{\mathrm{loc}}(\mathbb{R}_+,\mathbb{R}^d)$ with $\chi_i \geq 0$ for $i = 1,\dots,d$.

Proof. Define kernels $K^n = K(\cdot + n^{-1})$ for $n \in \mathbb{N}$, which are diagonal with scalar kernels on the diagonal that satisfy (5.3.3). Example 5.36 shows that the scalar kernels on the diagonal of K^n also satisfy (5.2.5). Lemma 5.35 shows that (5.6.8) (respectively (5.6.8) with K replaced by K^n) has a unique solution χ (respectively χ^n), and that $\chi^n \to \chi$ in $L^2(\mathbb{R}_+,\mathbb{R}^d)$. Therefore, we can suppose without loss of generality that K is continuous on $[0,\infty)$ with $K_i(0) \geq 0$. To shows that χ takes values in $\mathbb{R}^d_+$, it is therefore enough to consider the case where K is continuous on $[0,\infty)$ with $K_i(0) \geq 0$ for all i.

For $x \in \mathbb{R}^d$ define $b(x) = F + Gx$. For all positive n, Corollary 5.34 implies that there exists a unique solution $\chi^n \in L^2_{\mathrm{loc}}(\mathbb{R}_+,\mathbb{R}^d)$ of the equation

$$\chi^n = Ku + v + K * b((\chi^n - n^{-1})^+),$$

and that χ^n is continuous on $[0,\infty)$ with $\chi^n_i(0) = K_i(0)u_i + v_i \geq 0$ for $i = 1,\dots,d$. We claim that χ^n is $\mathbb{R}^d_+$ valued for all n. Indeed, arguing as in the proof of Theorem 5.12, we can show that if L_i denotes the resolvent of the first kind of K_i, then $(\Delta_h K_i * L_i)(t)$ is right-continuous, nonnegative, bounded by 1, and nondecreasing in t for any $h \geq 0$. Fix n and define $Z = \int b((\chi^n - n^{-1})^+)\,dt$. The argument of Lemma 5.6 shows that for all $h \geq 0$ and $i = 1,\dots,d$,

$$\begin{aligned}\Delta_h K_i * dZ_i &= (\Delta_h K_i * L_i)(0)K_i * dZ_i + d(\Delta_h K_i * L_i) * K_i * dZ_i \\ &= (\Delta_h K_i * L_i)(0)\chi^n_i + d(\Delta_h K_i * L_i) * \chi^n_i \\ &\quad - u_i\left((\Delta_h K_i * L_i)(0)K_i + d(\Delta_h K_i * L_i) * K_i\right) - v_i \Delta_h K_i * L_i.\end{aligned} \tag{5.6.9}$$

Convolving the quantity $d(\Delta_h K_i * L_i) * K_i$ first by L_i, then by K_i, and comparing densities of the resulting absolutely continuous functions, we deduce that

$$d(\Delta_h K_i * L_i) * K_i = \Delta_h K_i - (\Delta_h K_i * L_i)(0)K_i \quad \text{a.e.}$$

Plugging this identity into (5.6.9) yields

$$\Delta_h K_i * dZ_i = (\Delta_h K_i * L_i)(0)\chi^n_i + d(\Delta_h K_i * L_i) * \chi^n_i - u_i \Delta_h K_i - v_i \Delta_h K_i * L_i. \tag{5.6.10}$$

Define $\tau = \inf\{t \geq 0 \colon \chi^n_t \notin \mathbb{R}^d_+\}$ and assume for contradiction that $\tau < \infty$. Then

$$\chi^n(\tau+h) = \Delta_h K(\tau)u + v + (K * dZ)_{\tau+h} = \Delta_h K(\tau)u + v + (\Delta_h K * dZ)_\tau + \int_0^h K(h-s)dZ_{\tau+s} \tag{5.6.11}$$

for any $h \geq 0$. By definition of τ, the identities (5.6.10) and (5.6.11) imply

$$\chi^n_i(\tau+h) \geq \int_0^h K_i(h-s)b_i((\chi^n(\tau+s) - n^{-1})^+)\,ds, \quad i = 1, \dots, d.$$

As in the proof of Theorem 5.12, these inequalities lead to a contradiction. Hence $\tau = \infty$ and χ^n is $\mathbb{R}^d_+$-valued for all n.

To conclude that χ is $\mathbb{R}^d_+$-valued it suffices to prove that χ^n converges to χ in $L^2([0,T], \mathbb{R}^d)$ for all $T \in \mathbb{R}_+$. To this end we write

$$\chi - \chi^n = K * \Big(G(\chi^n - (\chi^n - n^{-1})^+) + G(\chi - \chi^n)\Big),$$

from which we infer

$$|\chi - \chi^n| \leq \frac{\sqrt{d}}{n}|K| * |G| + |K| * (|G||\chi - \chi^n|).$$

The same argument as in the proof of Corollary 5.34 shows that

$$|\chi - \chi^n| \leq \frac{\sqrt{d}}{n}\left(F' - \int_0^\cdot R'(\,\cdot\,, s)F'(s)\,ds\right), \tag{5.6.12}$$

where R' is the nonpositive resolvent of type L^2_{loc} of $K'(t,s) = |K(t-s)||G(s)|\mathbb{1}_{s\leq t}$, and $F' = |K|*|G|$. Since the right-hand side of (5.6.12) is in $L^2_{\mathrm{loc}}(\mathbb{R}_+, \mathbb{R})$ in view of [69, Theorem 9.3.6], we conclude that χ^n converges to χ in $L^2([0,T], \mathbb{R}^d)$ for all $T \in \mathbb{R}_+$. □

Chapter 6

Markovian structure

Summary

We characterize the Markovian and affine structure of the Volterra Heston model in terms of an infinite-dimensional adjusted forward process and specify its state space. More precisely, we show that it satisfies a stochastic partial differential equation and displays an exponentially-affine characteristic functional. As an application, we deduce an existence and uniqueness result for a Banach-space valued square-root process and provide its state space. This leads to another representation of the Volterra Heston model together with its Fourier-Laplace transform in terms of this possibly infinite system of affine diffusions.

Based on [4]: Abi Jaber, E., & El Euch, O. (2018) Markovian structure of the Volterra Heston model. *In revision - Statistics and Probability Letters.*

In this chapter, we restrict ourselves to the one-dimensional Volterra square-root process in order to highlight the key ideas without obscuring them with cumbersome notations. The correspondence between stochastic Volterra equations and stochastic partial differential equations established here remain valid for any b and σ. That between Riccati-Volterra equations and partial differential equations holds more generally for any affine Volterra process as studied in the previous chapter. Once one treats the one-dimensional case, the extension to higher dimension is straightforward.

6.1 Introduction

We recall the dynamics of the Volterra Heston model introduced in the previous chapter:

$$dS_t = S_t\sqrt{V_t}dB_t, \quad S_0 > 0, \tag{6.1.1}$$

$$V_t = g_0(t) + \int_0^t K(t-s)\left(-\lambda V_s ds + \nu\sqrt{V_s}dW_s\right), \tag{6.1.2}$$

with $K \in L^2_{\text{loc}}(\mathbb{R}_+, \mathbb{R})$, $g_0 : \mathbb{R}_+ \to \mathbb{R}$, $\lambda, \nu \in \mathbb{R}_+$ and $B = \rho W + \sqrt{1-\rho^2} W^\perp$ such that $(W, W^\perp)$ is a two-dimensional Brownian motion and $\rho \in [-1, 1]$. S_t typically represents a stock price at time t with instantaneous stochastic variance V_t.

This model nests as special cases the Heston model for $K \equiv 1$, and the rough Heston model of [51], obtained by setting $K(t) = \frac{t^{\alpha-1}}{\Gamma(\alpha)}$ for $\alpha \in (\frac{1}{2}, 1)$ and

$$g_0(t) = V_0 + \int_0^t K(s)\lambda\theta ds, \quad t \geq 0, \quad \text{for some } V_0, \theta \geq 0, \tag{6.1.3}$$

so that the only model parameters are $V_0, \theta, \lambda, \rho, \nu, \alpha$. Recall that the rough Heston model does not only fit remarkably well historical and implied volatilities of the market, but also enjoys a semi-closed formula for the characteristic function of the log-price in terms of a solution of a deterministic Riccati-Volterra integral equation.

In [52], the authors highlight the crucial role of (6.1.3) in the design of hedging strategies for the rough Heston model. Here we consider more general input curves g_0. Our motivation is twofold. In practice, the function g_0 is intimately linked to the forward variance curve $(\mathbb{E}[V_t])_{t\geq 0}$. More precisely, taking the expectation in (6.1.2) leads to the following relation

$$\mathbb{E}[V_t] + \lambda \int_0^t K(t-s)\mathbb{E}[V_s]ds = g_0(t), \quad t \geq 0.$$

Thus, allowing for more general input curves g_0 leads to more consistency with the market forward variance curve. From a mathematical perspective, this enables us to understand the general picture behind the Markovian and affine nature of the Volterra Heston model (6.1.1)-(6.1.2) and more generally, of stochastic Volterra equations as studied in Chapter 5.

More precisely, adapting the methods of the previous Chapter, we provide a set of admissible input curves $\mathcal{G}_K$ defined in (6.2.5) such that (6.1.1)-(6.1.2) admits a unique $\mathbb{R}^2_+$-valued weak solution for any $g_0 \in \mathcal{G}_K$. In particular, we show that the Fourier-Laplace transform of $(\log S, V)$ is exponentially affine in $(\log S_0, g_0)$. Then we prove that, conditional on $\mathcal{F}_t$, the shifted Volterra Heston model $(S_{t+\cdot}, V_{t+\cdot})$ still has the same dynamics as in (6.1.1)-(6.1.2) provided that g_0 is replaced by the following adjusted forward process

$$g_t(x) = \mathbb{E}\left[V_{t+x} + \lambda \int_0^x K(x-s)V_{t+s}ds \;\middle|\; \mathcal{F}_t\right], \quad x \geq 0. \tag{6.1.4}$$

This leads to our main result which states that $\mathcal{G}_K$ is stochastically invariant with respect to the family $(g_t)_{t\geq 0}$. In other words, if we start from an initial admissible input curve $g_0 \in \mathcal{G}_K$, then g_t belongs to $\mathcal{G}_K$, for all $t \geq 0$, see Theorem 6.4. This in turn enables us to characterize the Markovian structure of (S, V) in terms of the stock price and the adjusted forward process $(g_t)_{t\geq 0}$. Furthermore, $(g_t)_{t\geq 0}$ can be realized as the unique $\mathcal{G}_K$-valued mild solution of the following stochastic partial differential equation of Heath–Jarrow–Morton-type

$$dg_t(x) = \left(\frac{d}{dx}g_t(x) - \lambda K(x)g_t(0)\right)dt + K(x)\nu\sqrt{g_t(0)}dW_t, \quad g_0 \in \mathcal{G}_K,$$

and displays an affine characteristic functional.

As an application, we establish the existence and uniqueness of a Banach-space valued square-root process and provide its state space. This leads to another representation of $(V_t, g_t)_{t\geq 0}$ when the kernel K is completely monotone. Moreover, the Fourier-Laplace transform of

$(\log S, V)$ is shown to be an exponential affine functional of this process. These results are in the spirit of the Markovian representation of fractional Brownian motion, see [25, 71].

The chapter is organized as follows. In Section 6.2, we prove weak existence and uniqueness for the Volterra Heston model and provide its Fourier-Laplace transform. Section 6.3 characterizes the Markovian structure in terms of the adjusted forward variance process. Section 6.4 establishes the existence and uniqueness of a Banach-space valued square-root process and provides the link with the Volterra framework. In Section 6.5, we extend the existence results for stochastic Volterra equations of the previous Chapter.

Notations : As in Chapter 5, elements of $\mathbb{C}^m$ are viewed as column vectors, while elements of the dual space $(\mathbb{C}^m)^*$ are viewed as row vectors. For $h \geq 0$, Δ_h denotes the shift operator, i.e. $\Delta_h f(t) = f(t+h)$. If the function f on $\mathbb{R}_+$ is right-continuous and of locally bounded variation, the measure induced by its distribution derivative is denoted df, so that $f(t) = f(0) + \int_{[0,t]} df(s)$ for all $t \geq 0$. Finally, we use the notation $*$ for the convolution operation.

6.2 The extended Volterra Heston model

We study in this section the extended Volterra Heston model given by (6.1.1)-(6.1.2) allowing for arbitrary curves g_0 as input. Section 6.2.1 treats the existence part, while Section 6.2.2 tackles the uniqueness part and provides the exponential-affine transform.

6.2.1 Weak existence of the Volterra Heston model

When g_0 is given by (6.1.3), Theorem 5.24 provides the existence of a $\mathbb{R}^2_+$-valued weak solution to (6.1.1)-(6.1.2) under mild assumptions on K. We show in Theorem 6.1 below that weak existence in $\mathbb{R}^2_+$ continue to hold for (6.1.1)-(6.1.2) for a wider class of admissible input curves g_0 under the same assumptions on K. Since S is determined by V, it suffices to study the Volterra square-root equation (6.1.2). In the sequel, we assume that K satisfies (5.2.5) and (5.3.3). Theorem 6.10(ii) below guarantees the existence of an unconstrained continuous weak solution V to the following modified equation

$$V_t = g_0(t) + \int_0^t K(t-s)\left(-\lambda V_s ds + \nu\sqrt{V_s^+}dW_s\right), \tag{6.2.1}$$

for any locally Hölder continuous function g_0, where $x^+ = \max(0, x)$. Clearly, one needs to impose additional assumptions on g_0 to ensure the nonnegativity of V and drop the positive part in (6.2.1) so that V solves (6.1.2). Hence, weak existence of a nonnegative solution to (6.1.2) boils down to finding a set $\mathcal{G}_K$ of admissible input curves g_0 such that any solution V to (6.2.1) is nonnegative.

To get a taste of the admissible set $\mathcal{G}_K$, we start by assuming that g_0 and K are continuously differentiable on $[0, \infty)$. In that case, V is a semimartingale such that

$$dV_t = \left(g_0'(t) + (K' * dZ)_t - K(0)\lambda V_t\right)dt + K(0)\nu\sqrt{V_t^+}dW_t, \tag{6.2.2}$$

where $Z = \int_0^{\cdot}(-\lambda V_s ds + \nu\sqrt{V_s^+}dW_s)$. Relying on Lemma 5.6, we have[1]

$$K' = (K' * L)(0)K + d(K' * L) * K,$$

so that $K' * dZ$ can be expressed as a functional of (V, g_0) as follows

$$K' * dZ = (K' * L)(0)(V - g_0) + d(K' * L) * (V - g_0). \tag{6.2.3}$$

Since $V_0 = g_0(0)$, it is straightforward that $g_0(0)$ should be nonnegative. Now, assume that V hits zero for the first time at $\tau \geq 0$. After plugging (6.2.3) in the drift of (6.2.2), a first-order Euler scheme leads to the formal approximation

$$V_{\tau+h} \approx \left(g_0'(\tau) - (K' * L)(0)g_0(\tau) - (d(K' * L) * g_0)(\tau) + (d(K' * L) * V)_\tau\right) h,$$

for small $h \geq 0$. Since $K' * L$ is non-decreasing and $V \geq 0$ on $[0, \tau]$, it follows that $(d(K' * L) * V)_\tau \geq 0$ yielding the nonnegativity of $V_{\tau+h}$ if we impose the following additional condition

$$g_0' - (K' * L)(0)g_0 - d(K' * L) * g_0 \geq 0.$$

In the general case, V is not necessarily a semimartingale, and a delicate analysis should be carried on the integral equation (6.2.1) instead of the infinitesimal version (6.2.2). This suggests that the infinitesimal derivative operator should be replaced by the semigroup operator of right shifts leading to the following condition on g_0

$$\Delta_h g_0 - (\Delta_h K * L)(0)g_0 - d(\Delta_h K * L) * g_0 \geq 0, \quad h \geq 0,^{2} \tag{6.2.4}$$

and to the following definition of the set $\mathcal{G}_K$ of admissible input curves

$$\mathcal{G}_K = \left\{ g_0 \in \mathcal{H}^{\gamma/2} \text{ satisfying (8.6.4) and } g_0(0) \geq 0 \right\}, \tag{6.2.5}$$

where $\mathcal{H}^\alpha = \{g_0 : \mathbb{R}_+ \to \mathbb{R}, \text{ locally Hölder continuous of any order strictly smaller than } \alpha\}$. Recall that γ is the exponent associated with K in (5.2.5).

The following theorem establishes the existence of a $\mathbb{R}_+^2$-valued weak continuous solution to (6.1.1)-(6.1.2) on some filtered probability space $(\Omega, \mathcal{F}, \mathbb{F} = (\mathcal{F}_t)_{t\geq 0}, \mathbb{P})$ for any admissible input curve $g_0 \in \mathcal{G}_K$. Since S is determined by V, the proof follows directly from Theorems 6.10-6.11.

Theorem 6.1. *Assume that K satisfies (5.2.5) and (5.3.3). Then, the stochastic Volterra equation (6.1.1)-(6.1.2) has a $\mathbb{R}_+^2$-valued continuous weak solution (S, V) for any positive initial condition S_0 and any admissible input curve $g_0 \in \mathcal{G}_K$. Furthermore, the paths of V are locally Hölder continuous of any order strictly smaller than $\gamma/2$ and*

$$\sup_{t\leq T} \mathbb{E}[|V_t|^p] < \infty, \quad p > 0, \quad T > 0. \tag{6.2.6}$$

Example 6.2. *The following classes of functions belong to $\mathcal{G}_K$.*

(i) $g \in \mathcal{H}^{\gamma/2}$ non-decreasing such that $g(0) \geq 0$. Since K is non-increasing and L is nonnegative, we have $0 \leq \Delta_h K * L \leq 1$ for all $h \geq 0$ (see the proof of Theorem 5.12)

[1] Under (5.3.3) one can show that $K' * L$ is right-continuous, non-decreasing and of locally bounded variation (as in Remark 5.7), thus the associated measure $d(K' * L)$ is well defined.

[2] Recall that under assumption (5.3.3), one can show that $\Delta_h K * L$ is right-continuous and of locally bounded variation (see Remark 5.7), thus the associated measure $d(\Delta_h K * L)$ is well defined.

yielding, for all $t, h \geq 0$, *that* $\Delta_h g(t) - (\Delta_h K * L)(0)g(t) - (d(\Delta_h K * L) * g)(t)$ *is equal to*

$$\int_0^t (g(t) - g(t-s))(\Delta_h K * L)(ds) + g(t+h) - g(t) + g(t)(1 - (\Delta_h K * L)(t)) \geq 0$$

(ii) $g = V_0 + K * \theta$, *with* $V_0 \geq 0$ *and* $\theta \in L^2_{loc}(\mathbb{R}_+, \mathbb{R})$ *such that* $\theta(s)ds + V_0 L(ds)$ *is a nonnegative measure. First,* $g \in \mathcal{H}^{\gamma/2}$ *due to* (5.2.5) *and the Cauchy-Schwarz inequality*

$$(g(t+h) - g(t))^2 \leq 2\left(\int_0^t (K(s+h) - K(s))^2 ds + \int_0^h K(s)^2 ds\right) \int_0^{t+h} \theta(s)^2 ds.$$

Moreover, $g(0) = V_0 \geq 0$ *and*

$$\Delta_h g - (\Delta_h K * L)(0)g - d(\Delta_h K * L) * g \tag{6.2.7}$$

is equal to

$$V_0(1 - \Delta_h K * L) + \Delta_h(K * \theta) - (\Delta_h K * L)(0) K * \theta - d(\Delta_h K * L) * K * \theta.$$

(8.6.4) *now follows from Lemma 5.6 with* $F = \Delta_h K$, *after noticing that* (6.2.7) *becomes*

$$\Delta_h(K * (V_0 L + \theta)) - \Delta_h K * (V_0 L + \theta) = \int_\cdot^{\cdot + h} K(\cdot + h - s)(V_0 L(ds) + \theta(s)ds) \geq 0.$$

6.2.2 The Fourier-Laplace transform

We now tackle the weak uniqueness of (6.1.1)-(6.1.2) by entirely characterizing the Fourier-Laplace transform of the process $X = (\log S, V)$. Indeed, when g_0 is of the form (6.1.3), X is a two-dimensional affine Volterra process in the sense of Definition 5.14. For this particular g_0, Theorem 5.24(ii) provides the exponential-affine transform formula

$$\mathbb{E}[\exp(uX_T + (f * X)_T)] = \exp\left(\psi_1(T)\log S_0 + u_2 g_0(T) + \int_0^T F(\psi_1, \psi_2)(s) g_0(T-s)ds\right) \tag{6.2.8}$$

for suitable $u \in (\mathbb{C}^2)^*$ and $f \in L^1([0,T], (\mathbb{C}^*)^2)$ with $T > 0$, where $\psi = (\psi_1, \psi_2)$ solves the following system of Riccati-Volterra equations

$$\psi_1 = u_1 + 1 * f_1, \tag{6.2.9}$$
$$\psi_2 = u_2 K + K * F(\psi_1, \psi_2), \tag{6.2.10}$$

with

$$F(\psi_1, \psi_2) = f_2 + \frac{1}{2}\left(\psi_1^2 - \psi_1\right) + (\rho\nu\psi_1 - \lambda)\psi_2 + \frac{\nu^2}{2}\psi_2^2. \tag{6.2.11}$$

We prove in the following theorem that the affine transform (6.2.8) carries over for any admissible input curve $g_0 \in \mathcal{G}_K$ with the same Riccati equations (6.2.9)-(6.2.10).

Theorem 6.3. *Assume that* K *satisfies* (5.2.5) *and that the shifted kernels* $\Delta_h K$ *satisfy* (5.3.3) *for all* $h \in [0,1]$. *Fix* $g_0 \in \mathcal{G}_K, S_0 > 0$ *and denote by* (S, V) *a* $\mathbb{R}^2_+$*-valued continuous*

weak solution to (6.1.1)-(6.1.2). *For any* $u \in (\mathbb{C}^2)^*$ *and* $f \in L^1_{\mathrm{loc}}(\mathbb{R}_+, (\mathbb{C}^2)^*)$ *such that*

$$\operatorname{Re}\psi_1 \in [0,1],\ \operatorname{Re} u_2 \leq 0 \text{ and } \operatorname{Re} f_2 \leq 0, \tag{6.2.12}$$

with ψ_1 *given by* (6.2.9), *the Riccati–Volterra equation* (6.2.10) *admits a unique global solution* $\psi_2 \in L^2_{\mathrm{loc}}(\mathbb{R}_+, \mathbb{C}^*)$. *Moreover, the exponential-affine transform* (6.2.8) *is satisfied. In particular, weak uniqueness holds for* (6.1.1)-(6.1.2).

Proof. Uniqueness in law follows from part (6.2.8) since the law of V is determined by the Laplace functionals $\mathbb{E}[\exp((f_2 * V)_T)]$ as f_2 ranges through all nonpositive real valued continuous functions, and T ranges through $\mathbb{R}_+$. The existence, uniqueness, and non-positivity statement for the Riccati–Volterra equation is proved in Lemma 5.27. In order to obtain (6.2.8), we adapt the arguments of Theorems 5.16 and 5.24 to prove that the conditional Fourier-Laplace transform exhibits an exponentially affine form

$$\mathbb{E}\left[\exp\left(uX_T + (f * X)_T\right) \mid \mathcal{F}_t\right] = \exp(Y_t), \quad t \leq T, \tag{6.2.13}$$

and Y is defined as follows

$$Y_t = \mathbb{E}[uX_T + (f * X)_T \mid \mathcal{F}_t] + \frac{1}{2}\int_t^T \psi(T-s)a(\mathbb{E}[X_s \mid \mathcal{F}_t])\psi(T-s)^\top ds, \tag{6.2.14}$$

where a corresponds to the diffusion matrix

$$a(x) = \begin{pmatrix} 1 & \rho\nu \\ \rho\nu & \nu^2 \end{pmatrix} x_2, \quad x = (x_1, x_2) \in \mathbb{R} \times \mathbb{R}_+. \tag{6.2.15}$$

Step 1. Inspecting the proof of Lemmas (5.5) and 5.15, the conditional mean of V can be expressed as

$$\mathbb{E}[V_s \mid \mathcal{F}_t] = \left(g_0 - R_\lambda * g\right)(s) + \int_0^t E_\lambda(s-r)\nu\sqrt{V_r}dW_r, \quad s \leq t, \tag{6.2.16}$$

where R_λ is the resolvent of the second kind of $-\lambda K$, i.e. the unique solution of $R = -\lambda K + \lambda K * R = -\lambda K + R * \lambda K$ and $E_\lambda = K - R_\lambda * K$.
Step 2. Plugging back the expression (6.2.16) in (6.2.14) together with similar computations to those in the proof of Theorem 5.16 we get the following dynamics for Y:

$$Y_t = Y_0 + \int_0^t \psi(T-s)\sigma_s dW'_s - \frac{1}{2}\int_0^t \psi(T-s)a(X_s)\psi(T-s)^\top ds, \tag{6.2.17}$$

$$Y_0 = (u_1 + (1 * f_1)(T))\log S_0 + (u_2 g_0 + F(\psi_1, \psi_2) * g_0)(T), \tag{6.2.18}$$

where

$$\sigma_s = \begin{pmatrix} \sqrt{1-\rho^2} & \rho \\ 0 & 1 \end{pmatrix} \nu\sqrt{V_s},$$

$B = \sqrt{1-\rho^2}dW^\perp + \rho dW$, $W' = (W^\perp, W)$ is a standard two-dimensional Brownian motion and a, F are respectively given by (6.2.15) and (6.2.11). It follows that $Y + \frac{1}{2}\langle Y\rangle$ is a local martingale, so that $\exp(Y)$ is a local martingale by Itô's formula. Provided that $\exp(Y)$ is a true martingale, the exponential-affine transform formula (6.2.13) follows upon observing that $Y_T = uX_T + (f * X)_T$ by (6.2.14). Therefore, it suffices to argue that $\exp(Y)$ is indeed a true martingale.

Step 3. We start by observing that the scalar function

$$\pi_h = \Delta_h\psi * \bar{L} - \Delta_h(\psi * \bar{L})$$

is of locally bounded variation for every $h \geq 0$ (see Theorem 5.18(ii)), where $\bar{L} = \operatorname{diag}[\delta_0, L]$ and L is the resolvent of the first kind of K. Moreover, denoting by $(\cdot)^{\mathbf{r}} = \operatorname{Re}(\cdot)$, we observe that

$$\pi^{\mathbf{r}}_{h,2}(t) = -\int_0^h \psi^{\mathbf{r}}_2(h-s)L(t+ds), \quad t \geq 0,$$

Due to the assumption (5.3.3) on L and since $-\psi^{\mathbf{r}}_2 \geq 0$, it follows that $\pi^{\mathbf{r}}_{h,2}$ is nonnegative and non-increasing. In addition, a straightforward extension of Theorem 5.18(ii) shows that the process Y in (6.2.14) can be rewritten as

$$Y_t = (\Delta_h f * X)_t + (\Delta_h\psi * \bar{L})(0)(X_t - \bar{g}(t)) + (d\pi_h * (X - \bar{g}))_t$$

with $h = T - t$ and $\bar{g} = (\log S_0, g_0)^\top$. In particular,

$$\begin{aligned}\operatorname{Re} Y_t = {} & \psi^{\mathbf{r}}_1(h)\log S_t + (\operatorname{Re}\Delta_h f_1 * \log S)_t + (\Delta_h\psi^{\mathbf{r}}_2 * L)(0)V_t \\ & + (\operatorname{Re}\Delta_h f_2 * V)_t + (d\pi^{\mathbf{r}}_{h,2} * V)_t - (\Delta_h\psi^{\mathbf{r}}_2 * L)(0)g(t) - (d\pi^{\mathbf{r}}_{h,2} * g_0)(t). \end{aligned} \tag{6.2.19}$$

Since $\psi^{\mathbf{r}}_1 \in [0,1]$, integration by parts yields

$$\begin{aligned}\psi^{\mathbf{r}}_1(h)\log S_t + (\operatorname{Re}\Delta_h f_1 * \log S)_t &= \psi^{\mathbf{r}}_1(T)\log S_0 + \int_0^t \psi^{\mathbf{r}}_1(T-s)\, d(\log S)_s \\ &\leq \psi^{\mathbf{r}}_1(T)\log S_0 + U_t - \frac{1}{2}\langle U\rangle_t,\end{aligned}$$

where $U_t = \int_0^t \psi^{\mathbf{r}}_1(T-s)\sqrt{V_s}dB_s$. This observation together with the fact that $\psi^{\mathbf{r}}_2$, $(\Delta_h\psi^{\mathbf{r}}_2 * L)(0)$, $\operatorname{Re}\Delta_h f_2$, $-\pi^{\mathbf{r}}_{2,h}$, and $d\pi^{\mathbf{r}}_{2,h}$ all have nonpositive components in (6.2.19) and that g_0 is bounded on $[0,T]$ show that

$$|\exp(Y_t)| = \exp(\operatorname{Re} Y_t) \leq C_T S_0^{\psi^{\mathbf{r}}_1(T)} \exp(U_t - \frac{1}{2}\langle U\rangle_t),$$

for some constant $C_T \geq 0$, where the right-hand side is a true martingale by an obvious extension of Lemma 5.26. Thus $\exp(Y)$ is a true martingale and the exponential-affine transform (6.2.13) holds. Evaluating (6.2.13) at $t = 0$ leads to (6.2.8) thanks to (6.2.18). □

6.3 Markovian structure

Using the same methodology as in [52], we characterize the Markovian structure of the Volterra Heston model (6.1.1)-(6.1.2) in terms of the $\mathbb{F}$-adapted infinite-dimensional adjusted forward curve $(g_t)_{t\geq 0}$ given by (6.1.4) which is well defined thanks to (6.2.6). Furthermore, we prove that the set $\mathcal{G}_K$ is stochastically invariant with respect to $(g_t)_{t\geq 0}$.

Theorem 6.4. *Under the assumptions of Theorem 6.1, fix $g_0 \in \mathcal{G}_K$. Denote by (S,V) the unique solution to (6.1.1)-(6.1.2) and by $(g_t)_{t\geq 0}$ the process defined by (6.1.4). Then,*

(S^{t_0}, V^{t_0}) *satisfies*

$$dS_t^{t_0} = S_t^{t_0}\sqrt{V_t^{t_0}}dB_t^{t_0}, \quad S_0^{t_0} = S_{t_0},$$
$$V_t^{t_0} = g_{t_0}(t) + \int_0^t K(t-s)\left(-\lambda V_s^{t_0}ds + \nu\sqrt{V_s^{t_0}}dW_s^{t_0}\right),$$

where $(B^{t_0}, W^{t_0}) = (B_{t_0+\cdot} - B_{t_0}, W_{t_0+\cdot} - W_{t_0})$ *are two Brownian motions independent of* $\mathcal{F}_{t_0}$ *such that* $d\langle B^{t_0}, W^{t_0}\rangle_t = \rho dt$*. Moreover,* $\mathcal{G}_K$ *is stochastically invariant with respect to* $(g_t)_{t\geq 0}$*, that is*

$$g_t \in \mathcal{G}_K, \quad t \geq 0.$$

Proof. The part for V^{t_0} is immediate after observing that

$$g_{t_0}(t) = g_0(t_0+t) - \int_0^{t_0} K(t+t_0-s)\lambda V_s ds + \int_0^{t_0} K(t+t_0-s)\nu\sqrt{V_s}dW_s, \tag{6.3.1}$$

for all $t_0, t, h \geq 0$. The part for S^{t_0} is straightforward. We move to proving the claimed invariance. Fix $t_0, t, h \geq 0$ and define $Z = \int_0^{\cdot}(-\lambda V_s ds + \nu\sqrt{V_s}dW_s)$. By Lemma 5.6 and Remark 5.7,

$$\Delta_h K = (\Delta_h K * L)(0)K + d(\Delta_h K * L) * K, \tag{6.3.2}$$

so that

$$(\Delta_h K * dZ) = (\Delta_h K * L)(0)(V - g_0) + d(\Delta_h K * L) * (V - g_0).$$

Hence,

$$\begin{aligned}
V_{t+h}^{t_0} &= g_0(t_0+t+h) + (\Delta_h K * dZ)_{t_0+t} + \int_0^h K(h-s)dZ_{t_0+t+s} \\
&= g_0(t_0+t+h) + (\Delta_h K * L)(0)(V_t^{t_0} - g_0(t_0+t)) \\
&\quad + (d(\Delta_h K * L) * (V - g_0))_{t_0+t} + \int_0^h K(h-s)dZ_{t_0+t+s} \\
&= g_0(t_0+t+h) - (\Delta_h K * L)(0)g_0(t_0+t) - (d(\Delta_h K * L) * g_0)(t_0+t) \\
&\quad + (\Delta_h K * L)(0)V_t^{t_0} + (d(\Delta_h K * L) * V)_{t_0+t} + \int_0^h K(h-s)dZ_{t_0+t+s} \\
&\geq (\Delta_h K * L)(0)V_t^{t_0} + (d(\Delta_h K * L) * V)_{t_0+t} - \int_0^h K(h-s)\lambda V_{t+s}^{t_0}ds \\
&\quad + \int_0^h K(h-s)\nu\sqrt{V_{t+s}^{t_0}}dW_{t+s}^{t_0},
\end{aligned}$$

since $g_0 \in \mathcal{G}_K$. We now prove (8.6.4). Set $G_h^{t_0} = \Delta_h g_{t_0} - (\Delta_h K * L)(0) g_{t_0} - d(\Delta_h K * L) * g_{t_0}$. The previous inequality combined with (6.1.4) yields

$$\begin{aligned} G_h^{t_0}(t) &= \mathbb{E}\left[V_{t+h}^{t_0} + (\lambda K * V^{t_0})_{t+h} - (\Delta_h K * L)(0)(V_t^{t_0} + (\lambda K * V^{t_0})_t) \;\middle|\; \mathcal{F}_{t_0}\right] \\ &\quad - \mathbb{E}\left[\left(d(\Delta_h K * L) * (V^{t_0} + \lambda K * V^{t_0})\right)_t \;\middle|\; \mathcal{F}_{t_0}\right] \\ &\geq \mathbb{E}\left[(d(\Delta_h K * L) * V)_{t_0+t} - \left(d(\Delta_h K * L) * V^{t_0}\right)_t - \int_0^h K(h-s)\lambda V_{t+s}^{t_0} ds \;\middle|\; \mathcal{F}_{t_0}\right] \\ &\quad + \mathbb{E}\left[(\lambda K * V^{t_0})_{t+h} - \left(((\Delta_h K * L)(0)K + d(\Delta_h K * L) * K) * \lambda V^{t_0}\right)_t \;\middle|\; \mathcal{F}_{t_0}\right]. \end{aligned}$$

Relying on (6.3.2), we deduce

$$\begin{aligned} G_h^{t_0}(t) &\geq \mathbb{E}\left[\int_t^{t_0+t} (d(\Delta_h K * L))(ds) V_{t_0+t-s} - \int_0^h K(h-s)\lambda V_{t+s}^{t_0} ds \;\middle|\; \mathcal{F}_{t_0}\right] \\ &\quad + \mathbb{E}\left[\int_t^{t+h} K(t+h-s)\lambda V_s^{t_0} ds \;\middle|\; \mathcal{F}_{t_0}\right] \\ &= \mathbb{E}\left[\int_t^{t_0+t} (d(\Delta_h K * L))(ds) V_{t_0+t-s} \;\middle|\; \mathcal{F}_{t_0}\right]. \end{aligned}$$

Hence (8.6.4) holds for g_{t_0}, since $V \geq 0$ and $d(\Delta_h K * L)$ is a nonnegative measure, see Remark 5.7. Finally, by adapting the proof of Lemma 5.4, we can show that for any $p > 1, \epsilon > 0$ and $T > 0$, there exists a positive constant C_1 such that

$$\mathbb{E}\left[|V_{t+h} - V_t|^p\right] \leq C_1 h^{p(\gamma/2-\epsilon)}, \quad t, h \geq 0,\ t + h \leq T + t_0,$$

Relying on (5.2.5), (6.1.4) and Jensen inequality, there exists a positive constant C_2 such that

$$\mathbb{E}\left[|g_{t_0}(t+h) - g_{t_0}(t)|^p\right] \leq C_2 h^{p(\gamma/2-\epsilon)}, \quad t, h \geq 0,\ t + h \leq T,$$

By Kolmogorov continuity criterion, $g_{t_0} \in \mathcal{H}^{\gamma/2}$ so that $g_{t_0} \in \mathcal{G}_K$ since $g_{t_0}(0) = V_{t_0} \geq 0$. □

Theorem 6.4 highlights that V is Markovian in the state variable $(g_t)_{t\geq 0}$. Indeed, conditional on $\mathcal{F}_t$ for some $t \geq 0$, the shifted Volterra Heston model (S^t, V^t) can be started afresh from (S_t, g_t) with the same dynamics as in (6.1.1)-(6.1.2). Notice that g_t is again an admissible input curve belonging to $\mathcal{G}_K$. Therefore, applying Theorems 6.1 and 6.3 with (S^t, V^t, g_t) yields that the conditional Fourier-Laplace transform of $X = (\log S, V)$ is exponentially affine in $(\log S_t, g_t)$:

$$\begin{aligned} \mathbb{E}\left[\exp(uX_T + (f * X)_T) \;\middle|\; \mathcal{F}_t\right] &= \exp\left((\Delta_{T-t} f * X)_t\right) \\ &\quad \times \exp\left(\psi_1(T-t)\log S_t + (u_2 g_t + F(\psi_1, \psi_2) * g_t)(T-t)\right), \end{aligned} \tag{6.3.3}$$

for all $t \leq T$, where F is given by (6.2.11), under the standing assumptions of Theorem 6.3.

Moreover, it follows from (6.3.1) and the fact that $g_\cdot(0) = V$ that the process $(g_t)_{t\geq 0}$ solves

$$g_t(x) = \Delta_t g_0(x) + \int_0^t \Delta_{t-s}\left(-\lambda K g_s(0)\right)(x) ds + \int_0^t \Delta_{t-s}\left(K\nu\sqrt{g_s(0)}\right)(x) dW_s. \tag{6.3.4}$$

Recalling that $(\Delta_t)_{t\geq 0}$ is the semigroup of right shifts, (6.3.4) can be seen as a $\mathcal{G}_K$-valued mild solution of the following Heath–Jarrow–Morton-type *stochastic partial differential equation*

$$dg_t(x) = \left(\frac{d}{dx}g_t(x) - \lambda K(x) g_t(0)\right) dt + K(x)\nu\sqrt{g_t(0)}dW_t, \quad g_0 \in \mathcal{G}_K. \tag{6.3.5}$$

The following proposition provides the characteristic functional of $(g_t)_{t\geq 0}$ leading to the strong Markov property of $(g_t)_{t\geq 0}$. Define $\langle g, h\rangle = \int_{\mathbb{R}_+} g(x)h(x)dx$, for suitable functions f and g.

Theorem 6.5. *Under the assumptions of Theorem 6.3. Let $h \in \mathcal{C}_c^\infty(\mathbb{R}_+)$ and $g_0 \in \mathcal{G}_K$. Then,*

$$\mathbb{E}\left[\exp\left(\mathrm{i}\langle g_t, h\rangle\right)\right] = \exp\left(\langle H_t, g_0\rangle\right), \quad t \geq 0, \tag{6.3.6}$$

where H solves

$$H_t(x) = \mathrm{i}h(x-t)\mathbb{1}_{\{x>t\}} + \mathbb{1}_{\{x\leq t\}}\left(-\lambda\langle H_{t-x}, K\rangle + \frac{\nu^2}{2}\langle H_{t-x}, K\rangle^2\right), \quad t, x \geq 0. \tag{6.3.7}$$

In particular, weak uniqueness holds for (6.3.4) and $(g_t)_{t\geq 0}$ is a strong Markov process on $\mathcal{G}_K$.

Proof. Consider $\widetilde{S}_t = 1 + \int_0^t \widetilde{S}_u\sqrt{V_u}dW_u$, for all $t \geq 0$. Then, $(\widetilde{S}, V)$ is a Volterra Heston model of the form (6.1.1)-(6.1.2) with $\rho = 1$ and $\widetilde{S}_0 = 1$. Fix $t \geq 0$, $\langle g_t, h\rangle$ is well defined since $x \to g_t(x)$ is continuous. It follows from (6.3.1) together with stochastic Fubini theorem, see [111, Theorem 2.2], which is justified by (6.2.6), that

$$\begin{aligned}
\langle g_t, h\rangle &= \langle g_0(t+\cdot), h\rangle + \left(\frac{\nu}{2} - \lambda\right)\int_0^t \langle K(t-s+\cdot), h\rangle V_s ds + \nu\int_0^t \langle K(t-s+\cdot), h\rangle d(\log\widetilde{S})_s \\
&= \langle g_0, h(-t+\cdot)\rangle + \left(\frac{\nu}{2} - \lambda\right)\int_0^t \langle K, h(s-t+\cdot)\rangle V_s ds \\
&\quad + \nu\langle K, h\rangle \log\widetilde{S}_t - \nu\int_0^t \langle K, h'(s-t+\cdot)\rangle \log\widetilde{S}_s ds,
\end{aligned}$$

where the last identity follows from an integration by parts. Hence, setting

$$\begin{aligned}
u_2 = 0, \quad u_1 = \mathrm{i}\nu\langle K, h\rangle, \quad f_1(t) &= -\mathrm{i}\nu\langle K, h'(-t+\cdot)\rangle, \\
\psi_1(t) = u_1 + (1 * f_1)(t) &= \mathrm{i}\nu\langle K(t+\cdot), h\rangle, \\
f_2(t) = \mathrm{i}(\frac{\nu}{2} - \lambda)\langle K(t+\cdot), h\rangle, \quad \psi_2 &= K * F(\psi_1, \psi_2),
\end{aligned}$$

with F as in (6.2.11), the characteristic functional follows from Theorem 6.3

$$\mathbb{E}\left[\exp\left(\mathrm{i}\langle g_t, h\rangle\right)\right] = e^{\mathrm{i}\langle h(-t+\cdot), g_0\rangle}\mathbb{E}\left[\exp\left(u_1\log\widetilde{S}_t + (f_1 * \log\widetilde{S})_t + (f_2 * V)_t\right)\right] = \exp\left(\langle H_t, g_0\rangle\right)$$

where

$$H_t(x) = h(x-t)\mathbb{1}_{\{x>t\}} + \mathbb{1}_{\{0\leq x\leq t\}}F(\psi_1, \psi_2)(t-x), \quad x \geq 0,$$

and (6.2.11) reads

$$F(\psi_1,\psi_2)(t) = -\lambda\langle K(t+\cdot),h\rangle + \frac{\nu^2}{2}\langle K(t+\cdot),h\rangle^2 + (\nu^2\langle K(t+\cdot),h\rangle - \lambda)\psi_2(t) + \frac{\nu^2}{2}\psi_2(t)^2. \tag{6.3.8}$$

Now observe that

$$\langle H_t,K\rangle = \langle h(-t+\cdot),K\rangle + \int_0^t F(\psi_1,\psi_2)(t-x)K(x)dx = \langle h,K(t+\cdot)\rangle + \psi_2(t).$$

Hence, after plugging $\psi_2(t) = \langle H_t,K\rangle - \langle h,K(t+\cdot)\rangle$ back in (6.3.8) we get that

$$F(\psi_1,\psi_2)(t) = -\lambda\langle H_t,K\rangle + \frac{\nu^2}{2}\langle H_t,K\rangle^2,$$

yielding (6.3.7). Weak uniqueness now follows by standard arguments. In fact, thanks to (6.2.6) and stochastic Fubini theorem, $(g_t)_{t\geq 0}$ solves (6.3.5) in the weak sense, that is

$$\langle g_t,h\rangle = \langle g_0,h\rangle + \int_0^t \left(\langle g_s,-h'\rangle - \lambda\langle K,h\rangle g_s(0)\right)ds + \int_0^t \nu\langle K,h\rangle\sqrt{g_s(0)}dW_s, \quad h\in\mathcal{C}_c^\infty(\mathbb{R}).$$

Therefore, combined with Theorem 6.4, $(g_t)_{t\geq 0}$ solves a martingale problem on $\mathcal{G}_K$. In addition, (6.3.6) yields uniqueness of the one-dimensional distributions which is enough to get weak uniqueness for (6.3.4) and the strong Markov property by [54, Theorem 4.4.2]. □

We notice that (6.3.6)-(6.3.7) agree with [67, Proposition 4.5] when $\lambda = 0$. Moreover, one can lift (6.3.7) to a non-linear partial differential equation in duality with (6.3.5). Indeed, define the measure-valued function $\bar{H} : t \to \bar{H}_t(dx) = H_t(x)\mathbb{1}_{\{x\geq 0\}}dx$. Then, it follows from (6.3.7) that

$$\bar{H}_t(dx) = \mathrm{i}h(x-t)\mathbb{1}_{\{x>t\}}dx + \int_0^t \delta_0(dx-(t-s))(-\lambda\langle\bar{H}_s,K\rangle + \frac{\nu^2}{2}\langle\bar{H}_s,K\rangle^2)ds$$

which can be seen as the mild formulation of the following partial differential equation

$$d\bar{H}_t(dx) = (-\frac{d}{dx}\bar{H}_t(dx) + \delta_0(dx)(-\lambda\langle\bar{H}_t,K\rangle + \frac{\nu^2}{2}\langle\bar{H}_t,K\rangle^2))dt, \quad \bar{H}_0(dx) = \mathrm{i}h(x)\mathbb{1}_{\{x>t\}}dx. \tag{6.3.9}$$

We refer to [36] for similar results in the discontinuous setting where stochastic partial differential equations are taken as the starting point. The previous results highlight not only the correspondence between stochastic Volterra equations of the form (6.1.2) and stochastic partial differential equations (6.3.5) but also between their dual objects, that is the Riccati-Volterra equation (6.2.10) and the non-linear partial differential equation (6.3.9). As already illustrated in the introduction of this thesis, one can establish a correspondence between (6.1.2) and other related stochastic partial differential equations which, unlike $(g_t)_{t\geq 0}$, do not necessarily have a financial interpretation but for which the semigroup operator has the smoothing property, see in particular [96].

6.4 Application: square-root process in Banach space

As an application of Theorems 6.1, 6.3, 6.4, we obtain conditions for weak existence and uniqueness of the following (possibly) infinite-dimensional system of stochastic differential equations

$$dU_t(x) = \left(-xU_t(x) - \lambda \int_0^\infty U_t(z)\mu(dz)\right) dt + \nu\sqrt{\int_0^\infty U_t(z)\mu(dz)}dW_t, \quad x \in \operatorname{supp}(\mu), \tag{6.4.1}$$

for a fixed positive measure of locally bounded variation μ[3]. This is achieved by linking (6.4.1) to a stochastic Volterra equation of the form (6.1.2) with the following kernel

$$K(t) = \int_0^\infty e^{-xt}\mu(dx), \quad t > 0. \tag{6.4.2}$$

We will assume that μ is a positive measure of locally bounded variation such that

$$\int_0^\infty (1 \wedge (xh)^{-1/2})\mu(dx) \le Ch^{(\gamma-1)/2}, \quad \int_0^\infty x^{-1/2}(1 \wedge (xh))\mu(dx) \le Ch^{\gamma/2}; \quad h > 0, \tag{6.4.3}$$

for some $\gamma \in (0, 2]$ and positive constant C. The reader may check that in that case K satisfies (5.2.5). Furthermore, [69, Theorem 5.5.4] guarantees the existence of the resolvent of the first kind L of K and that (5.3.3) is satisfied for the shifted kernels $\Delta_h K$ for any $h \in [0, 1]$. Hence, K satisfies assumptions of Theorems 6.1 and 6.3.

	$K(t)$	Parameter restrictions	$\mu(d\gamma)$
Fractional	$c\frac{t^{\alpha-1}}{\Gamma(\alpha)}$	$\alpha \in (1/2, 1)$	$c\frac{x^{-\alpha}}{\Gamma(\alpha)\Gamma(1-\alpha)}dx$
Gamma	$ce^{-\lambda t}\frac{t^{\alpha-1}}{\Gamma(\alpha)}$	$\lambda \ge 0, \alpha \in (1/2, 1)$	$c\frac{(x-\lambda)^{-\alpha}\mathbb{1}_{(\lambda,\infty)}(x)}{\Gamma(\alpha)\Gamma(1-\alpha)}dx$
Exponential sum	$\sum_{i=1}^n c_i e^{-\gamma_i t}$	$c_i, \gamma_i \ge 0$	$\sum_{i=1}^n c_i\delta_{\gamma_i}(dx)$

TABLE 6.1: Some measures μ satisfying (6.4.3) with their associated kernels K. Here $c \ge 0$.

By a solution U to (6.4.1) we mean a family of continuous processes $(U(x))_{x\in\operatorname{supp}(\mu)}$ such that $x \to U_t(x) \in L^1(\mu)$ for any $t \ge 0$, $(\int_0^\infty U_t(x)\mu(dx))_{t\ge0}$ is a continuous process and (6.4.1) holds a.s. on some filtered probability space. If such solution exists, we set $V = \int_0^\infty U_\cdot(x)\mu(dx)$ and $g_0 = \int_0^\infty U_0(x)e^{-x(\cdot)}dx$. Thanks to (6.4.3), the stochastic Fubini theorem yields for each $t \ge 0$

$$V_t = g_0(t) + \int_0^t K(t-s)(-\lambda V_s ds + \nu\sqrt{V_s}dW_s). \tag{6.4.4}$$

The processes above being continuous, the equality holds in terms of processes. Thus, provided that g_0 belongs to $\mathcal{G}_K$, Theorem 6.3 leads to the weak uniqueness of (6.4.1) because

[3] We use the notation $\operatorname{supp}(\mu)$ to denote the support of a measure μ, that is the set of all points for which every open neighborhood has a positive measure. Here we assume that the support is in $\mathbb{R}_+$.

for each $x \in \mathrm{supp}(\mu)$,

$$U_t(x) = e^{-xt}U_0(x) + \int_0^t e^{-x(t-s)}(-\lambda V_s ds + \nu\sqrt{V_s}dW_s), \quad t \geq 0. \tag{6.4.5}$$

On the other hand, if we assume that $g_0 = \int_0^\infty U_0(x)e^{-x(\cdot)}\mu(dx) \in \mathcal{G}_K$ for some initial family of points $(U_0(x))_{x\in\mathrm{supp}(\mu)} \in L^1(\mu)$, there exists a continuous solution V for (6.4.4) by Theorem 6.1. In that case, we define for each $x \in \mathrm{supp}(\mu)$, the continuous process $U(x)$ as in (6.4.5). Thanks to (6.4.3) and (6.2.6), another application of the stochastic Fubini theorem combined with the fact that V satisfies (6.4.4) yields that, for each $t \geq 0$, $(U_t(x))_{x\in\mathrm{supp}(\mu)} \in L^1(\mu)$ and

$$V_t = \int_0^\infty U_t(x)\mu(dx). \tag{6.4.6}$$

Moreover, by an integration by parts, we get for each $x \in \mathrm{supp}(\mu)$,

$$U_t(x) = e^{-xt}U_0(x) + Z_t e^{-xt} + \int_0^t xe^{-x(t-s)}(Z_s - Z_t)ds,$$

with $Z = \int_0^\cdot(-\lambda V_s ds + \nu\sqrt{V_s}dB_s)$. We know that for fixed $T > 0$, $\eta \in (0, 1/2)$ and for almost any $\omega \in \Omega$ there exists a positive constant $C_T(\omega)$ such that $|Z_s - Z_t| \leq C_T(\omega)|t - s|^\eta$ for all $t, s \in [0, T]$. Hence for any $t \in [0, T]$ and $x \in \mathrm{supp}(\mu)$

$$|U_t(x)| \leq |U_0(x)| + C_T(\omega)e^{-xt}t^\eta + C_T(\omega)x\int_0^t e^{-xs}s^\eta ds = |U_0(x)| + C_T(\omega)\eta\int_0^t e^{-xs}s^{\eta-1}ds.$$

Then,

$$\sup_{t\in[0,T]}|U_t(x)| \leq |U_0(x)| + C_T(\omega)\eta\int_0^T e^{-xs}s^\eta ds \in L^1(\mu).$$

Therefore by dominated convergence theorem, the process $(\int_0^\infty U_t(x)\mu(dx))_{t\geq 0}$ is continuous. In particular, (6.4.6) holds in terms of processes and it follows from (6.4.5) that U is a solution of (6.4.1).

This leads to the weak existence and uniqueness of (6.4.1) if the initial family of points $(U_0(x))_{x\in\mathrm{supp}(\mu)}$ belongs to the following space $\mathcal{D}_\mu$ defined by

$$\mathcal{D}_\mu = \{(u_x)_{x\in\mathrm{supp}(\mu)} \in L^1(\mathrm{supp}(\mu)); \quad \int_0^\infty u_x e^{-x(\cdot)}\mu(dx) \in \mathcal{G}_K\}, \tag{6.4.7}$$

with K given by (6.4.2). Notice that for fixed $t_0 \geq 0$ and for any $t \geq 0$ and $x \in \mathrm{supp}(\mu)$,

$$U_{t+t_0}(x) = U_{t_0}(x)e^{-xt} + \int_0^t e^{-x(t-s)}\left(-\lambda\int_0^\infty U_{s+t_0}(z)\mu(dz) + \nu\sqrt{\int_0^\infty U_{s+t_0}(z)\mu(dz)}dW_{s+t_0}\right)$$

and then by stochastic Fubini theorem $\int_0^\infty U_{t+t_0}(y)\mu(dy)$ is equal to

$$g_{t_0}(t) + \int_0^t K(t-s)\left(-\lambda\int_0^\infty U_{s+t_0}(z)\mu(dz) + \nu\sqrt{\int_0^\infty U_{s+t_0}(z)\mu(dz)}dW_{s+t_0}\right),$$

with $g_{t_0}(t) = \int_0^\infty U_{t_0}(y)e^{-yt}\mu(dy)$. Thanks to Theorem 6.4, we deduce that $g_{t_0} \in \mathcal{G}_K$ and therefore $(U_{t_0}(x))_{x\in\mathrm{supp}(\mu)}$ belongs to $\mathcal{D}_\mu$. As a conclusion, the space $\mathcal{D}_\mu$ is stochastically invariant with respect to the family of processes $(U(x))_{x\in\mathrm{supp}(\mu)}$.

Theorem 6.6. *Fix μ a positive measure of locally bounded variation satisfying (6.4.3). There exists a unique weak solution U of (6.4.1) for each initial family of points $(U_0(x))_{x\in\mathrm{supp}(\mu)} \in \mathcal{D}_\mu$. Furthermore for any $t \geq 0$, $(U_t(x))_{x\in\mathrm{supp}(\mu)} \in \mathcal{D}_\mu$.*

Remark 6.7 (Representation of V in terms of U). In a similar fashion one can establish the existence and uniqueness of the following time-inhomogeneous version of (6.4.1)

$$dU_t(x) = \left(-xU_t(x) - \lambda\left(g_0(t) + \langle 1, U_t\rangle_\mu\right)\right)dt + \nu\sqrt{g_0(t) + \langle 1, U_t\rangle_\mu}dW_t, \quad x \in \mathrm{supp}(\mu), \tag{6.4.8}$$

whenever

$$g_0 = \widetilde{g_0} + \int_0^\infty e^{-x(\cdot)}U_0(x)\mu(dx) \in \mathcal{G}_K,$$

with $\widetilde{g_0} : \mathbb{R}_+ \to \mathbb{R}$. In this case,

$$\widetilde{g_0}(t+\cdot) + \int_0^\infty e^{-x(\cdot)}U_t(x)\mu(dx) \in \mathcal{G}_K, \quad t \geq 0.$$

In particular, for $U_0 \equiv 0$, $g_0 = \widetilde{g_0} \in \mathcal{G}_K$ and K as in (6.4.2), the solution V to the stochastic Volterra equation (6.1.2) and the forward process $(g_t)_{t\geq 0}$ admit the following representations

$$V_t = g_0(t) + \langle 1, U_t\rangle_\mu, \quad g_t(x) = g_0(t+x) + \langle e^{-x(\cdot)}, U_t\rangle_\mu, \quad t, x \geq 0, \tag{6.4.9}$$

where we used the notation $\langle f, g\rangle_\mu = \int_0^t f(x)g(x)\mu(dx)$. These results are in the spirit of [25, 71].

When μ has finite support, (6.4.8) is a finite dimensional diffusion with an affine structure in the sense of [48]. This underlying structure carries over to the case of infinite support and is the reason behind the tractability of the Volterra Heston model.

Remark 6.8 (Affine structure of $(\log S, V)$ in terms of U). Let the notations and assumptions of Remark 6.7 be in force. Relying on the existence and uniqueness of the Riccati-Volterra equation (6.2.10) one can establish the existence and uniqueness of a differentiable (in time) solution χ_2 to the following (possibly) infinite-dimensional system of Riccati ordinary differential equations

$$\partial_t\chi_2(t,x) = -x\chi_2(t,x) + F\left(\psi_1(t), \langle\chi_2(t,\cdot), 1\rangle_\mu\right), \quad \chi_2(0,x) = u_2, \quad x \in \mathrm{supp}\,\mu,\ t \geq 0, \tag{6.4.10}$$

such that $\chi_2(t,\cdot) \in L^1(\mu)$, for all $t \geq 0$ and $t \to \langle\chi_2(t,\cdot), 1\rangle_\mu \in L^2_{\mathrm{loc}}(\mathbb{R}_+)$ with ψ_1 given by (6.2.9) and F by (6.2.11). Moreover, the unique global solution $\psi_2 \in L^2_{\mathrm{loc}}(\mathbb{R}_+, \mathbb{C}^*)$ to the Riccati–Volterra equation (6.2.10) admits the following representation

$$\psi_2 = \int_0^\infty \chi_2(\cdot, x)\mu(dx),$$

where χ_2 is the unique solution to (6.4.10). In particular, combining the equality above with (6.3.3) and the representation of $(g_t)_{t\geq 0}$ in (6.4.9) leads to the exponentially-affine functional

$$\mathbb{E}\left[\exp\left(uX_T + (f * X)_T\right) \,\middle|\, \mathcal{F}_t\right] = \exp\left(\phi(t,T) + \psi_1(T-t)\log S_t + \langle\chi_2(T-t,\cdot), U_t\rangle_\mu\right) \tag{6.4.11}$$

for all $t \leq T$ where $\phi(t,T) = (u_2\Delta_t g_0 + F(\psi_1,\psi_2) * \Delta_t g_0)(T-t) + (\Delta_{T-t}f * X)_t$, (u, f) as in (6.2.12) and U solves (6.4.8).

The previous expression generalizes the well-known affine transform of the standard [72] case.

Remark 6.9. Set $K \equiv 1$ and $g_0 = V_0 + \int_0^{\cdot} \lambda\theta(s)ds$ for a locally square-integrable non-negative function θ. Then, $\mu = \delta_0$, $V^0 = V - g_0$ and $u_2 + \int_0^t F(\psi_1, \psi_2)(s)ds = \psi_2$. Define $\phi = \int_0^{\cdot} \lambda\theta(s)\psi_2(s)ds$. Direct computations show that we recover the standard affine transform

$$\mathbb{E}\left[\exp\left(u_1 \log S_T + u_2 V_T\right) \,\middle|\, \mathcal{F}_t\right] = \exp\left(\phi(T-t) + u_1 \log S_t + \psi_2(T-t)V_t\right).$$

The representations of this section lead to a generic approximation of the Volterra Heston model by finite-dimensional affine diffusions, see Chapter 7 for the rigorous treatment of these approximations.

6.5 Existence results for stochastic Volterra equations

In this section, we adapt the existence results of Chapter 5 to allow for arbitrary initial input curves in the equation

$$X_t = g(t) + \int_0^t K(t-s)b(X_s)ds + \int_0^t K(t-s)\sigma(X_s)dW_s, \tag{6.5.1}$$

where $K \in L^2_{\mathrm{loc}}(\mathbb{R}, \mathbb{R}^{d\times d})$, W is a m-dimensional Brownian motion, $g : \mathbb{R}^d \to \mathbb{R}^d$, $b : \mathbb{R}^d \to \mathbb{R}^d$, $\sigma : \mathbb{R}^d \to \mathbb{R}^{d\times m}$ are continuous with linear growth.

A straightforward adaptions of the proofs Lemma 5.29 and Theorem 5.10 lead to the following existence results of unconstrained solutions.

Theorem 6.10. *Under* (5.2.5), *assume that* $g \in \mathcal{H}^{\gamma/2}$.

(i) If b *and* σ *are Lipschitz continuous,* (6.5.1) *admits a unique continuous strong solution* X.

(ii) If b *and* σ *are continuous with linear growth and* K *admits a resolvent of the first kind* L, *then* (6.5.1) *admits a continuous weak solution* X.

In both cases, X *is locally Hölder continuous of any order strictly smaller than* $\gamma/2$ *and*

$$\sup_{t\le T} \mathbb{E}[|X_t|^p] < \infty, \quad p > 0, \quad T > 0.$$

The following theorem extends Theorem 5.12 for the existence of constrained solutions. Notice how the domain $\mathcal{G}_K$ defined in (6.2.5) enters in the construction of these solutions.

Theorem 6.11. *Assume that* $d = m = 1$ *and that the scalar kernel* K *satisfies* (5.2.5)-(5.3.3). *Assume also that* b *and* σ *are continuous with linear growth such that*

$$b(0) \ge 0 \quad \text{and } \sigma(0) = 0.$$

Then (6.5.1) *admits a nonnegative continuous weak solution for any* $g \in \mathcal{G}_K$.

Proof. Theorem 6.10(ii) yields the existence of an unsconstrained continuous weak solution X to the following modified equation

$$X_t = g(t) + \int_0^t K(t-s)b(X_s^+)ds + \int_0^t K(t-s)\sigma(X_s^+)dW_s.$$

As in the proof of of Theorem 5.12, it suffices to prove the nonnegativity of X under the stronger condition, that, for some fixed $n \in \mathbb{N}$,

$$x \le n^{-1} \text{ implies } b(x) \ge 0 \text{ and } \sigma(x) = 0. \tag{6.5.2}$$

Set $Z = \int(b(X)dt + \sigma(X)dW)$ and $\tau = \inf\{t \ge 0 \colon X_t < 0\}$. Since $g(0) \ge 0$, $\tau \ge 0$. On $\{\tau < \infty\}$,

$$X_{\tau+h} = g(\tau+h)+(K*dZ)_{\tau+h} = g(\tau+h)+(\Delta_h K*dZ)_\tau+\int_0^h K(h-s)dZ_{\tau+s}, \quad h \ge 0. \tag{6.5.3}$$

Using Lemma 5.6 and Remark 5.7, together with the fact that $X \ge 0$ on $[0,\tau]$,

$$\begin{aligned} g(\tau+h) + (\Delta_h K * dZ)_\tau &= g(\tau+h) + (\Delta_h K * L)(0)(X-g)(\tau) \\ &\quad + (d(\Delta_h K * L) * X)_\tau - (d(\Delta_h K * L) * g)(\tau) \\ &\ge g(\tau+h) - (d(\Delta_h K * L) * g)(\tau) - (\Delta_h K * L)(0)g(\tau), \end{aligned}$$

which is nonnegative. In view of (6.5.3) it follows that

$$X_{\tau+h} \ge \int_0^h K(h-s)\left(b(X_{\tau+s})ds + \sigma(X_{\tau+s})dW_{\tau+s}\right) \tag{6.5.4}$$

on $\{\tau < \infty\}$ for all $h \ge 0$. Now, on $\{\tau < \infty\}$, $X_\tau = 0$ and $X_{\tau+h} < 0$ for arbitrarily small h. On the other hand, by continuity there is some $\varepsilon > 0$ such that $X_{\tau+h} \le n^{-1}$ for all $h \in [0,\varepsilon)$. Thus (6.5.2) and (6.5.4) yield $X_{\tau+h} \ge 0$ for all $h \in [0,\varepsilon)$. This shows that $\tau = \infty$, ending the proof. □

Chapter 7

Approximation

Summary

Rough volatility models are very appealing because of their remarkable fit of both historical and implied volatilities. However, due to the non-Markovian and non-semimartingale nature of the volatility process, there is no simple way to simulate efficiently such models, which makes risk management of derivatives an intricate task. In this chapter, we design tractable multi-factor stochastic volatility models approximating rough volatility models and enjoying a Markovian structure. Furthermore, we apply our procedure to the specific case of the rough Heston model. This in turn enables us to derive a numerical method for solving fractional Riccati equations appearing in the characteristic function of the log-price in this setting.

Based on [5]: Abi Jaber, E., & El Euch, O. (2018) Multi-factor approximation of rough volatility models. *In revision - SIAM Journal on Financial Mathematics.* (BFS Junior Scholar Award for most outstanding paper.)

7.1 Introduction

Empirical studies of a very wide range of assets volatility time-series in [66] have shown that the dynamics of the log-volatility are close to that of a fractional Brownian motion W^H with a small Hurst parameter H of order 0.1. Recall that a fractional Brownian motion W^H can be built from a two-sided Brownian motion thanks to the Mandelbrot-van Ness representation

$$W_t^H = \frac{1}{\Gamma(H+1/2)} \int_0^t (t-s)^{H-\frac{1}{2}} dW_s + \frac{1}{\Gamma(H+1/2)} \int_{-\infty}^0 ((t-s)^{H-\frac{1}{2}} - (-s)^{H-\frac{1}{2}}) dW_s.$$

The fractional kernel $(t-s)^{H-\frac{1}{2}}$ is behind the $H-\varepsilon$ Hölder regularity of the volatility for any $\varepsilon > 0$. For small values of the Hurst parameter H, as observed empirically, stochastic volatility models involving the fractional kernel are called rough volatility models.

Aside from modeling historical volatility dynamics, rough volatility models reproduce accurately with very few parameters the behavior of the implied volatility surface, see [14, 50],

especially the at-the-money skew, see [63]. Moreover, microstructural foundations of rough volatility are studied in [49, 79].

In this paper, we are interested in a class of rough volatility models where the dynamics of the asset price S and its stochastic variance V are given by

$$dS_t = S_t\sqrt{V_t}dB_t, \quad S_0 > 0, \tag{7.1.1}$$

$$V_t = V_0 + \frac{1}{\Gamma(H+\frac{1}{2})}\int_0^t (t-u)^{H-\frac{1}{2}}(\theta(u)-\lambda V_u)du + \frac{1}{\Gamma(H+\frac{1}{2})}\int_0^t (t-u)^{H-\frac{1}{2}}\sigma(V_u)dW_u, \tag{7.1.2}$$

for all $t \in [0,T]$, on some filtered probability space $(\Omega, \mathcal{F}, \mathbb{F}, \mathbb{P})$. Here T is a positive time horizon, the parameters λ and V_0 are nonnegative, $H \in (0,1/2)$ is the Hurst parameter, σ is a continuous function and $B = \rho W + \sqrt{1-\rho^2}W^\perp$ with $(W, W^\perp)$ a two-dimensional $\mathbb{F}$-Brownian motion and $\rho \in [-1,1]$. Moreover, θ is a deterministic mean reversion level allowed to be time-dependent to fit the market forward variance curve $(\mathbb{E}[V_t])_{t\leq T}$ as explained in Section 7.2 and in [52]. Under some general assumptions, we establish in Section 7.2 the existence of a weak nonnegative solution to the fractional stochastic integral equation in (7.1.2) exhibiting $H-\varepsilon$ Hölder regularity for any $\varepsilon > 0$. Hence, this class of models is a natural rough extension of classical stochastic volatility models where the fractional kernel is introduced in the drift and stochastic part of the variance process V. Indeed, when $H = 1/2$, we recover classical stochastic volatility models where the variance process is a standard diffusion.

Despite the fit to the historical and implied volatility, some difficulties are encountered in practice for the simulation of rough volatility models and for pricing and hedging derivatives with them. In fact, due to the introduction of the fractional kernel, we lose the Markovian and semimartingale structure. In order to overcome theses difficulties, we approximate these models by simpler ones that we can use in practice.

In [51, 49, 52], the rough Heston model (which corresponds to the case of $\sigma(x) = \nu\sqrt{x}$) is built as a limit of microscopic Hawkes-based price models. This allowed the understanding of the microstructural foundations of rough volatility and also led to the formula of the characteristic function of the log-price. Hence, the Hawkes approximation enabled us to solve the pricing and hedging under the rough Heston model. However, this approach is specific to the rough Heston case and can not be extended to an arbitrary rough volatility model of the form (7.1.1)-(7.1.2).

Based on the representations of the previous Chapter and inspired by the works of [25, 26, 71, 94], we provide a natural Markovian approximation for the class of rough volatility models (7.1.1)-(7.1.2). The main idea is to write the fractional kernel $K(t) = \frac{t^{H-\frac{1}{2}}}{\Gamma(H+1/2)}$ as a Laplace transform of a positive measure μ

$$K(t) = \int_0^\infty e^{-xt}\mu(dx); \quad \mu(dx) = \frac{x^{-H-\frac{1}{2}}}{\Gamma(H+1/2)\Gamma(1/2-H)}dx. \tag{7.1.3}$$

We then approximate μ by a finite sum of Dirac measures $\mu^n = \sum_{i=1}^n c_i^n \delta_{x_i^n}$ with positive weights $(c_i^n)_{1\leq i\leq n}$ and mean reversions $(x_i^n)_{1\leq i\leq n}$, for $n \geq 1$. This in turn yields an approximation of the fractional kernel by a sequence of smoothed kernels $(K^n)_{n\geq 1}$ given by

$$K^n(t) = \sum_{i=1}^n c_i^n e^{-x_i^n t}, \quad n \geq 1.$$

This leads to a multi-factor stochastic volatility model $(S^n, V^n) = (S_t^n, V_t^n)_{t\leq T}$, which is Markovian with respect to the spot price and n variance factors $(U^{n,i})_{1\leq i\leq n}$ and is defined as follows

$$dS_t^n = S_t^n \sqrt{V_t^n} dB_t, \quad V_t^n = g^n(t) + \sum_{i=1}^n c_i^n U_t^{n,i}, \tag{7.1.4}$$

where

$$dU_t^{n,i} = (-x_i^n U_t^{n,i} - \lambda V_t^n)dt + \sigma(V_t^n)dW_t,$$

and $g^n(t) = V_0 + \int_0^t K^n(t-s)\theta(s)ds$ with the initial conditions $S_0^n = S_0$ and $U_0^{n,i} = 0$. Note that the factors $(U^{n,i})_{1\leq i\leq n}$ share the same dynamics except that they mean revert at different speeds $(x_i^n)_{1\leq i\leq n}$. Relying on our existence results of stochastic Volterra equations in the previous Chapter, we provide in Theorem 7.3 the strong existence and uniqueness of the model (S^n, V^n), under some general conditions. Thus the approximation (7.1.4) is uniquely well-defined. We can therefore deal with simulation, pricing and hedging problems under these multi-factor models by using standard methods developed for stochastic volatility models.

Theorem 7.7, which is the main result of this paper, establishes the convergence of the multi-factor approximation sequence $(S^n, V^n)_{n\geq 1}$ to the rough volatility model (S, V) in (7.1.1)-(7.1.2) when the number of factors n goes to infinity, under a suitable choice of the weights and mean reversions $(c_i^n, x_i^n)_{1\leq i\leq n}$. This convergence is obtained from a general result about stability of stochastic Volterra equations derived in Section 7.3.4.

As shown in Chapters 5 and 6, the characteristic function of the log-price for the specific case of the rough Heston model is obtained in terms of a solution of a fractional Volterra-Riccati equation. We highlight in Section 7.4.1 that the corresponding multi-factor approximation (7.1.4) inherits a similar affine structure as in the rough Heston model. More precisely, (S, V^n) is again a Volterra Heston model. Consequently, it displays the same characteristic function formula for which the Riccati-Volterra equation decouples in n-dimensional classical Riccati ordinary differential equations. This suggests solving numerically the fractional Riccati equation by approximating it through a n-dimensional classical Riccati equation with large n, see Theorem 7.9. In Section 7.4.2, we discuss the accuracy and complexity of this numerical method and compare it to the Adams scheme, see [44, 45, 46, 51].

The chapter is organized as follows. In Section 7.2, we define the class of rough volatility models (7.1.1)-(7.1.2) and discuss the existence of such models. Then, in Section 7.3, we build a sequence of multi-factor stochastic volatility models of the form of (7.1.4) and show its convergence to a rough volatility model. By applying this approximation to the specific case of the rough Heston model, we obtain a numerical method for computing solutions of fractional Riccati equations that is discussed in Section 7.4. Finally, some proofs are relegated to Section 7.5 and some useful technical results are given in Sections 7.6-7.8.

7.2 A definition of rough volatility models

We provide in this section the precise definition of rough volatility models given by (7.1.1)-(7.1.2). We discuss the existence of such models and more precisely of a nonnegative solution of the fractional stochastic integral equation (7.1.2). The existence of an unconstrained weak solution $V = (V_t)_{t\leq T}$ is guaranteed by Corollary 7.21 below when σ is a continuous function

with linear growth and θ satisfies the condition

$$\forall \varepsilon > 0, \quad \exists C_\varepsilon > 0; \quad \forall u \in (0,T] \quad |\theta(u)| \leq C_\varepsilon u^{-\frac{1}{2}-\varepsilon}. \tag{7.2.1}$$

Furthermore, the paths of V are Hölder continuous of any order strictly less than H and

$$\sup_{t\in[0,T]} \mathbb{E}[|V_t|^p] < \infty, \quad p > 0. \tag{7.2.2}$$

Moreover using Theorem 6.11 together with Remarks 7.23 and 7.24 below[1], the existence of a nonnegative continuous process V satisfying (7.1.2) is obtained under the additional conditions of non-negativity of V_0 and θ and $\sigma(0) = 0$. We can therefore introduce the following class of rough volatility models.

Definition 7.1. (Rough volatility models) We define a rough volatility model by any $\mathbb{R}_+^2$-valued continuous process $(S,V) = (S_t, V_t)_{t\leq T}$ satisfying

$$dS_t = S_t\sqrt{V_t}dB_t,$$

$$V_t = V_0 + \frac{1}{\Gamma(H+1/2)}\int_0^t (t-u)^{H-\frac{1}{2}}(\theta(u) - \lambda V_u)du + \frac{1}{\Gamma(H+1/2)}\int_0^t (t-u)^{H-\frac{1}{2}}\sigma(V_u)dW_u,$$

on a filtred probability space $(\Omega, \mathcal{F}, \mathbb{F}, \mathbb{P})$ with nonnegative initial conditions (S_0, V_0). Here T is a positive time horizon, the parameter λ is nonnegative, $H \in (0, 1/2)$ is the Hurst parameter and $B = \rho W + \sqrt{1-\rho^2}W^\perp$ with $(W, W^\perp)$ a two-dimensional $\mathbb{F}$-Brownian motion and $\rho \in [-1, 1]$. Moreover, to guarantee the existence of such model, $\sigma : \mathbb{R} \mapsto \mathbb{R}$ is assumed continuous with linear growth such that $\sigma(0) = 0$ and $\theta : [0,T] \mapsto \mathbb{R}$ is a deterministic nonnegative function satisfying (7.2.1).

As done in [52], we allow the mean reversion level θ to be time dependent in order to be consistent with the market forward variance curve. More precisely, the following result shows that the mean reversion level θ can be written as a functional of the forward variance curve $(\mathbb{E}[V_t])_{t\leq T}$.

Proposition 7.2. *Let (S,V) be a rough volatility model given by Definition 7.1. Then, $(\mathbb{E}[V_t])_{t\leq T}$ is linked to θ by the following formula*

$$\mathbb{E}[V_t] = V_0 + \int_0^t (t-s)^{\alpha-1}E_\alpha(-\lambda(t-s)^\alpha)\theta(s)ds, \quad t \in [0,T], \tag{7.2.3}$$

where $\alpha = H + 1/2$ and $E_\alpha(x) = \sum_{k\geq 0}\frac{x^k}{\Gamma(\alpha(k+1))}$ is the Mittag-Leffler function. Moreover, $(\mathbb{E}[V_t])_{t\leq T}$ admits a fractional derivative[2] of order α at each time $t \in (0,T]$ and

$$\theta(t) = D^\alpha(\mathbb{E}[V_\cdot] - V_0)_t + \lambda\mathbb{E}[V_t], \quad t \in (0,T]. \tag{7.2.4}$$

Proof. Thanks to (7.2.2) together with Fubini theorem, $t \mapsto \mathbb{E}[V_t]$ solves the following fractional linear integral equation

$$\mathbb{E}[V_t] = V_0 + \frac{1}{\Gamma(H+1/2)}\int_0^t (t-s)^{H-\frac{1}{2}}(\theta(s) - \lambda\mathbb{E}[V_s])ds, \quad t \in [0,T], \tag{7.2.5}$$

[1] Theorem 6.11 is used here with the fractional kernel $K(t) = \frac{t^{H-\frac{1}{2}}}{\Gamma(H+1/2)}$ together with $b(x) = -\lambda x$ and $g(t) = V_0 + \int_0^t K(t-u)\theta(u)du$.

[2] Recall that the fractional derivative of order $\alpha \in (0,1)$ of a function f is given by $\frac{d}{dt}\int_0^t \frac{(t-s)^{-\alpha}}{\Gamma(1-\alpha)}f(s)ds$ whenever this expression is well defined.

yielding (7.2.3) by Theorem 7.18 and Remark 7.20 below. Finally, (7.2.4) is obviously obtained from (7.2.5). □

Finally, note that uniqueness of the fractional stochastic integral equation (7.1.2) is a difficult problem. Adapting the proof in [96], we can prove pathwise uniqueness when σ is η-Hölder continuous with $\eta \in (1/(1+2H), 1]$. This result does not cover the square-root case, i.e. $\sigma(x) = \nu\sqrt{x}$, for which weak uniqueness has been established in Chapters 5 and 6, see also [96].

7.3 Multi-factor approximation of rough volatility models

Thanks to the small Hölder regularity of the variance process, models of Definition 7.1 are able to reproduce the rough behavior of the volatility observed in a wide range of assets. However, the fractional kernel forces the variance process to leave both the semimartingale and Markovian worlds, which makes numerical approximation procedures a difficult and challenging task in practice. The aim of this section is to construct a tractable and satisfactory Markovian approximation of any rough volatility model (S, V) of Definition 7.1. Because S is entirely determined by $(\int_0^\cdot V_s ds, \int_0^\cdot \sqrt{V_s} dB_s)$, it suffices to construct a suitable approximation of the variance process V. This is done by smoothing the fractional kernel.

More precisely, denoting by $K(t) = \frac{t^{H-\frac{1}{2}}}{\Gamma(H+1/2)}$, the fractional stochastic integral equation (7.1.2) reads

$$V_t = V_0 + \int_0^t K(t-s)\left((\theta(s) - \lambda V_s)ds + \sigma(V_s)dW_s\right),$$

which is a stochastic Volterra equation. Approximating the fractional kernel K by a sequence of smooth kernels $(K^n)_{n\geq 1}$, one would expect the convergence of the following corresponding sequence of stochastic Volterra equations

$$V_t^n = V_0 + \int_0^t K^n(t-s)\left((\theta(s) - \lambda V_s^n)ds + \sigma(V_s^n)dW_s\right), \quad n \geq 1,$$

to the fractional one.

The argument of this section runs as follows. First, exploiting the identity (7.1.3), we construct a family of potential candidates for $(K^n, V^n)_{n\geq 1}$ in Section 7.3.1 such that V^n enjoys a Markovian structure. Second, we provide convergence conditions of $(K^n)_{n\geq 1}$ to K in $L^2([0,T],\mathbb{R})$ in Section 7.3.2. Finally, the approximation result for the rough volatility model (S, V) is established in Section 7.3.3 relying on an abstract stability result of stochastic Volterra equations postponed to Section 7.3.4 for sake of exposition.

7.3.1 Construction of the approximation

In [25, 71, 94], a Markovian representation of the fractional Brownian motion of Riemann-Liouville type is provided by writing the fractional kernel $K(t) = \frac{t^{H-\frac{1}{2}}}{\Gamma(H+1/2)}$ as a Laplace transform of a nonnegative measure μ as in (7.1.3). This representation was extended in Chapter 6 for the Volterra square-root process. Adopting the same approach, we establish a similar representation for any solution of the fractional stochastic integral equation (7.1.2)

in terms of an infinite dimensional system of processes sharing the same Brownian motion and mean reverting at different speeds. Indeed by using the linear growth of σ together with the stochastic Fubini theorem, see [111], we obtain that

$$V_t = g(t) + \int_0^\infty U_t(x)\mu(dx), \quad t \in [0,T],$$

with

$$dU_t(x) = (-xU_t(x) - \lambda V_t)dt + \sigma(V_t)dW_t, \quad U_0(x) = 0, \quad x \geq 0,$$

and

$$g(t) = V_0 + \int_0^t K(t-s)\theta(s)ds. \tag{7.3.1}$$

Inspired by [25, 26], we approximate the measure μ by a weighted sum of Dirac measures

$$\mu^n = \sum_{i=1}^n c_i^n \delta_{x_i^n}, \quad n \geq 1,$$

leading to the following approximation $V^n = (V_t^n)_{t \leq T}$ of the variance process V

$$V_t^n = g^n(t) + \sum_{i=1}^n c_i^n U_t^{n,i}, \quad t \in [0,T], \tag{7.3.2}$$

$$dU_t^{n,i} = (-x_i^n U_t^{n,i} - \lambda V_t^n)dt + \sigma(V_t^n)dW_t, \quad U_0^{n,i} = 0,$$

where

$$g^n(t) = V_0 + \int_0^t K^n(t-u)\theta(u)du, \tag{7.3.3}$$

and

$$K^n(t) = \sum_{i=1}^n c_i^n e^{-x_i^n t}. \tag{7.3.4}$$

The choice of the positive weights $(c_i^n)_{1 \leq i \leq n}$ and mean reversions $(x_i^n)_{1 \leq i \leq n}$, which is crucial for the accuracy of the approximation, is studied in Section 7.3.2 below. Before proving the convergence of $(V^n)_{n \geq 1}$, we shall first discuss the existence and uniqueness of such processes. This is done by rewriting the stochastic equation (7.3.2) as a stochastic Volterra equation of the form

$$V_t^n = g^n(t) + \int_0^t K^n(t-s)\left(-\lambda V_s^n ds + \sigma(V_s^n)dW_s\right), \quad t \in [0,T]. \tag{7.3.5}$$

The existence of a continuous nonnegative weak solution V^n is ensured by Theorem 6.11 together with Remarks 7.23 and 7.24 below[3], because θ and V_0 are nonnegative and $\sigma(0) = 0$. Moreover, pathwise uniqueness of solutions to (7.3.5) follows by adapting the standard arugments of [115], provided a suitable Hölder continuity of σ, see Proposition 7.22 below. Note that this extension is made possible due to the smoothness of the kernel K^n. For instance, this approach fails for the fractional kernel because of the singularity at zero. This leads us to the following result which establishes the strong existence and uniqueness of a nonnegative solution of (7.3.5) and equivalently of (7.3.2).

Theorem 7.3. *Assume that $\theta : [0,T] \mapsto \mathbb{R}$ is a deterministic nonnegative function satisfying (7.2.1) and that $\sigma : \mathbb{R} \mapsto \mathbb{R}$ is η-Hölder continuous with $\sigma(0) = 0$ and $\eta \in [1/2, 1]$. Then,*

[3]Theorem 6.11 is used here with the smoothed kernel K^n given by (7.3.4) together with $b(x) = -\lambda x$ and g defined as in (7.3.1)

there exists a unique strong nonnegative solution $V^n = (V_t^n)_{t\leq T}$ *to the stochastic Volterra equation* (7.3.5) *for each* $n \geq 1$.

Due to the uniqueness of (7.3.2), we obtain that V^n is a Markovian process according to n state variables $(U^{n,i})_{1\leq i\leq n}$ that we call the factors of V^n. Moreover, V^n being nonnegative, it can be used to model a variance process. This leads to the following definition of multi-factor stochastic volatility models.

Definition 7.4. (Multi-factor stochastic volatility models). We define the following sequence of multi-factor stochastic volatility models $(S^n, V^n) = (S_t^n, V_t^n)_{t\leq T}$ as the unique $\mathbb{R}_+^2$-valued strong solution of

$$dS_t^n = S_t^n \sqrt{V_t^n} dB_t, \quad V_t^n = g^n(t) + \sum_{i=1}^n c_i^n U_t^{n,i},$$

with

$$dU_t^{n,i} = (-x_i^n U_t^{n,i} - \lambda V_t^n)dt + \sigma(V_t^n)dW_t, \quad U_0^{n,i} = 0, \quad S_0^n = S_0 > 0,$$

on a filtered probability space $(\Omega, \mathcal{F}, \mathbb{P}, \mathbb{F})$, where $\mathbb{F}$ is the canonical filtration a two-dimensional Brownian motion $(W, W^\perp)$ and $B = \rho W + \sqrt{1-\rho^2}W^\perp$ with $\rho \in [-1, 1]$. Here, the weights $(c_i^n)_{1\leq i\leq n}$ and mean reversions $(x_i^n)_{1\leq i\leq n}$ are positive, $\sigma : \mathbb{R} \mapsto \mathbb{R}$ is η-Hölder continuous such that $\sigma(0) = 0$, $\eta \in [1/2, 1]$ and g^n is given by (7.3.3), that is

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

with a nonnegative initial variance V_0, a kernel K^n defined as in (7.3.4) and a nonnegative deterministic function $\theta : [0, T] \mapsto \mathbb{R}$ satisfying (7.2.1).

Note that the strong existence and uniqueness of (S^n, V^n) follows from Theorem 7.3. This model is Markovian with $n+1$ state variables which are the spot price S^n and the factors of the variance process $U^{n,i}$ for $i \in \{1, \dots, n\}$.

7.3.2 An approximation of the fractional kernel

Relying on (7.3.5), we can see the process V^n as an approximation of V, solution of (7.1.2), obtained by smoothing the fractional kernel $K(t) = \frac{t^{H-\frac{1}{2}}}{\Gamma(H+1/2)}$ into $K^n(t) = \sum_{i=1}^n c_i^n e^{-x_i^n t}$. Intuitively, we need to choose K^n close to K when n goes to infinity, so that $(V^n)_{n\geq 1}$ converges to V. Inspired by [26], we give in this section a condition on the weights $(c_i^n)_{1\leq i\leq n}$ and mean reversion terms $0 < x_1^n < ... < x_n^n$ so that the following convergence

$$\|K^n - K\|_{2,T} \to 0,$$

holds as n goes to infinity, where $\|\cdot\|_{2,T}$ is the usual $L^2([0, T], \mathbb{R})$ norm. Let $(\eta_i^n)_{0\leq i\leq n}$ be auxiliary mean reversion terms such that $\eta_0^n = 0$ and $\eta_{i-1}^n \leq x_i^n \leq \eta_i^n$ for $i \in \{1, \dots, n\}$. Writing K as the Laplace transform of μ as in (7.1.3), we obtain that

$$\|K^n - K\|_{2,T} \leq \int_{\eta_n^n}^\infty \|e^{-x(\cdot)}\|_{2,T}\mu(dx) + \sum_{i=1}^n J_i^n,$$

with $J_i^n = \|c_i^n e^{-x_i^n(\cdot)} - \int_{\eta_{i-1}^n}^{\eta_i^n} e^{-x(\cdot)}\mu(dx)\|_{2,T}$. We start by dealing with the first term,

$$\int_{\eta_n^n}^{\infty} \|e^{-x(\cdot)}\|_{2,T}\mu(dx) = \int_{\eta_n^n}^{\infty} \sqrt{\frac{1-e^{-2xT}}{2x}}\mu(dx) \leq \frac{1}{H\Gamma(H+1/2)\Gamma(1/2-H)\sqrt{2}}(\eta_n^n)^{-H}.$$

Moreover by choosing

$$c_i^n = \int_{\eta_{i-1}^n}^{\eta_i^n} \mu(dx), \quad x_i^n = \frac{1}{c_i^n}\int_{\eta_{i-1}^n}^{\eta_i^n} x\mu(dx), \quad i \in \{1,\dots,n\}, \tag{7.3.6}$$

and using the Taylor-Lagrange inequality up to the second order, we obtain

$$\left| c_i^n e^{-x_i^n t} - \int_{\eta_{i-1}^n}^{\eta_i^n} e^{-xt}\mu(dx) \right| \leq \frac{t^2}{2}\int_{\eta_{i-1}^n}^{\eta_i^n} (x-x_i^n)^2\mu(dx), \quad t \in [0,T]. \tag{7.3.7}$$

Therefore,

$$\sum_{i=1}^n J_i^n \leq \frac{T^{5/2}}{2\sqrt{5}}\sum_{i=1}^n \int_{\eta_{i-1}^n}^{\eta_i^n} (x_i^n - x)^2\mu(dx).$$

This leads to the following inequality

$$\|K^n - K\|_{2,T} \leq f_n^{(2)}((\eta_i)_{0\leq i\leq n}),$$

where $f_n^{(2)}$ is a function of the auxiliary mean reversions defined by

$$f_n^{(2)}((\eta_i^n)_{1\leq i\leq n}) = \frac{T^{\frac{5}{2}}}{2\sqrt{5}}\sum_{i=1}^n \int_{\eta_{i-1}^n}^{\eta_i^n} (x-x_i^n)^2\mu(dx) + \frac{1}{H\Gamma(H+1/2)\Gamma(1/2-H)\sqrt{2}}(\eta_n^n)^{-H}. \tag{7.3.8}$$

Hence, we obtain the convergence of K^n to the fractional kernel under the following choice of weights and mean reversions.

Assumption 7.3.1. We assume that the weights and mean reversions are given by (7.3.6) such that $\eta_0^n = 0 < \eta_1^n < \ldots < \eta_n^n$ and

$$\eta_n^n \to \infty, \quad \sum_{i=1}^n \int_{\eta_{i-1}^n}^{\eta_i^n} (x_i^n - x)^2\mu(dx) \to 0, \tag{7.3.9}$$

as n goes to infinity.

Proposition 7.5. *Fix $(c_i^n)_{1\leq i\leq n}$ and $(x_i^n)_{1\leq i\leq n}$ as in Assumption 7.3.1 and K^n given by (7.3.4), for all $n \geq 1$. Then, $(K^n)_{n\geq 1}$ converges in $L^2[0,T]$ to the fractional kernel $K(t) = \frac{t^{H-1/2}}{\Gamma(H+\frac{1}{2})}$ as n goes to infinity.*

There exists several choices of auxiliary factors such that condition (7.3.9) is met. For instance, assume that $\eta_i^n = i\pi_n$ for each $i \in \{0,\dots,n\}$ such that $\pi_n > 0$. It follows from

$$\sum_{i=1}^n \int_{\eta_{i-1}^n}^{\eta_i^n} (x-x_i)^2\mu(dx) \leq \pi_n^2 \int_0^{\eta_n^n} \mu(dx) = \frac{1}{(1/2-H)\Gamma(H+1/2)\Gamma(1/2-H)}\pi_n^{\frac{5}{2}-H} n^{\frac{1}{2}-H},$$

that (7.3.9) is satisfied for

$$\eta_n^n = n\pi_n \to \infty, \quad \pi_n^{\frac{5}{2}-H} n^{\frac{1}{2}-H} \to 0,$$

as n tends to infinity. In this case,

$$\|K^n - K\|_{2,T} \leq \frac{1}{H\Gamma(H+1/2)\Gamma(1/2-H)\sqrt{2}} \left((\eta_n^n)^{-H} + \frac{HT^{\frac{5}{2}}}{\sqrt{10}(1/2-H)} \pi_n^2 (\eta_n^n)^{\frac{1}{2}-H} \right).$$

This upper bound is minimal for

$$\pi_n = \frac{n^{-\frac{1}{5}}}{T} \left(\frac{\sqrt{10}(1-2H)}{5-2H} \right)^{\frac{2}{5}}, \tag{7.3.10}$$

and

$$\|K^n - K\|_{2,T} \leq C_H n^{-\frac{4H}{5}},$$

where C_H is a positive constant that can be computed explicitly and that depends only on the Hurst parameter $H \in (0, 1/2)$.

Remark 7.6. Note that the kernel approximation in Proposition 7.5 can be easily extended to any kernel of the form

$$K(t) = \int_0^\infty e^{-xt} \mu(dx),$$

where μ is a nonnegative measure such that

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

7.3.3 Convergence result

We assume now that the weights and mean reversions of the multi-factor stochastic volatility model (S^n, V^n) satisfy Assumption 7.3.1. Thanks to Proposition 7.5, the smoothed kernel K^n is close to the fractional one for large n. Because V^n satisfies the stochastic Volterra equation (7.3.5), V^n has to be close to V and thus by passing to the limit, $(S^n, V^n)_{n\geq 1}$ should converge to the rough volatility model (S, V) of Definition 7.1 as n goes large. This is the object of the next theorem, which is the main result of this paper.

Theorem 7.7. *Let $(S^n, V^n)_{n\geq 1}$ be a sequence of multi-factor stochastic volatility models given by Definition 7.4. Then, under Assumption 7.3.1, the family $(S^n, V^n)_{n\geq 1}$ is tight for the uniform topology and any point limit (S, V) is a rough volatility model given by Definition 7.1.*

Theorem 7.7 states the convergence in law of $(S^n, V^n)_{n\geq 1}$ whenever the fractional stochastic integral equation (7.1.2) admits a unique weak solution. In order to prove Theorem 7.7, whose proof is in Section 7.5.2 below, a more general stability result for d-dimensional stochastic Volterra equations is established in the next subsection.

7.3.4 Stability of stochastic Volterra equations

As mentioned above, Theorem 7.7 relies on the study of the stability of more general d-dimensional stochastic Volterra equations of the form

$$X_t = g(t) + \int_0^t K(t-s)b(X_s)ds + \int_0^t K(t-s)\sigma(X_s)dW_s, \quad t \in [0,T], \tag{7.3.11}$$

where $b : \mathbb{R}^d \to \mathbb{R}^d$, $\sigma : \mathbb{R}^d \to \mathbb{R}^{d\times m}$ are continuous and satisfy the linear growth condition, $K \in L^2([0,T], \mathbb{R}^{d\times d})$ admits a resolvent of the first kind L, and W is a m-dimensional Brownian motion on some filtered probability space $(\Omega, \mathcal{F}, \mathbb{F}, \mathbb{P})$. From Section 7.7 below, $g : [0,T] \mapsto \mathbb{R}^d$ and $K \in L^2([0,T], \mathbb{R}^{d\times d})$ should satisfy Assumption 7.7.1, that is

$$|g(t+h) - g(t)|^2 + \int_0^h |K(s)|^2 ds + \int_0^{T-h} |K(h+s) - K(s)|^2 ds \leq Ch^{2\gamma}, \tag{7.3.12}$$

for any $t, h \geq 0$ with $t + h \leq T$ and for some positive constants C and γ, to guarantee the weak existence of a continuous solution X of (7.3.11).

More precisely, we consider a sequence $X^n = (X_t^n)_{t\leq T}$ of continuous weak solutions to the stochastic Volterra equation (7.3.11) with a kernel $K^n \in L^2([0,T], \mathbb{R}^{d\times d})$ admitting a resolvent of the first kind, on some filtered probability space $(\Omega^n, \mathcal{F}^n, \mathbb{F}^n, \mathbb{P}^n)$,

$$X_t^n = g^n(t) + \int_0^t K^n(t-s)b(X_s^n)ds + \int_0^t K^n(t-s)\sigma(X_s^n)dW_s^n, \quad t \in [0,T],$$

with $g^n : [0,T] \mapsto \mathbb{R}^d$ and K^n satisfying (7.3.12) for every $n \geq 1$. The stability of (7.3.11) means the convergence in law of the family of solutions $(X^n)_{n\geq 1}$ to a limiting process X which is a solution to (7.3.11), when (K^n, g^n) is close to (K, g) as n goes large.

This convergence is established by verifying first the Kolmogorov tightness criterion for the sequence $(X^n)_{n\geq 1}$. It is obtained when g^n and K^n satisfy (7.3.12) uniformly in n in the following sense.

Assumption 7.3.2. There exists positive constants γ and C such that

$$\sup_{n\geq 1}\left(|g^n(t+h) - g^n(t)|^2 + \int_0^h |K^n(s)|^2 ds + \int_0^{T-h} |K^n(h+s) - K^n(s)|^2 ds\right) \leq Ch^{2\gamma},$$

for any $t, h \geq 0$ with $t + h \leq T$,

The following result, whose proof is postponed to Section 7.5.1 below, states the convergence of $(X^n)_{n\geq 1}$ to a solution of (7.3.11).

Theorem 7.8. *Assume that*

$$\int_0^T |K(s) - K^n(s)|^2 ds \longrightarrow 0, \quad g_n(t) \longrightarrow g(t),$$

for any $t \in [0,T]$ as n goes to infinity. Then, under Assumption 7.3.2, the sequence $(X^n)_{n\geq 1}$ is tight for the uniform topology and any point limit X is a solution of the stochastic Volterra equation (7.3.11).

7.4 The particular case of the rough Heston model

The rough Heston model introduced in [51, 52] is a particular case of the class of rough volatility models of Definition 7.1, with $\sigma(x) = \nu\sqrt{x}$ for some positive parameter ν, that is

$$dS_t = S_t\sqrt{V_t}dB_t, \quad S_0 > 0,$$
$$V_t = g(t) + \int_0^t K(t-s)\left(-\lambda V_s ds + \nu\sqrt{V_s}dW_s\right),$$

where $K(t) = \frac{t^{H-\frac{1}{2}}}{\Gamma(H+1/2)}$ denotes the fractional kernel and g is given by (7.3.1). Aside from reproducing accurately the historical and implied volatility, the rough Heston model displays a closed formula for the characteristic function of the log-price in terms of a solution to a fractional Riccati equation allowing to fast pricing and calibration, see [50]. More precisely, the model belongs to the class of Volterra Heston models studied in Chapter 6 such that

$$L(t,z) = \mathbb{E}\left[\exp\left(z\log(S_t/S_0)\right)\right]$$

is given by

$$\exp\left(\int_0^t F(z,\psi(t-s,z))g(s)ds\right), \tag{7.4.1}$$

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

$$\psi(t,z) = \int_0^t K(t-s)F(z,\psi(s,z))ds, \quad t\in[0,T], \tag{7.4.2}$$

with $F(z,x) = \frac{1}{2}(z^2-z) + (\rho\nu z - \lambda)x + \frac{\nu^2}{2}x^2$ and $z\in\mathbb{C}$ such that $\Re(z)\in[0,1]$. Unlike the classical case $H = 1/2$, (7.4.2) does not exhibit an explicit solution. However, it can be solved numerically through the Adam scheme developed in [44, 45, 46, 51] for instance. In this section, we show that the multi-factor approximation applied to the rough Heston model gives rise to another natural numerical scheme for solving the fractional Riccati equation. Furthermore, we will establish the convergence of this scheme with explicit errors.

7.4.1 Multi-factor scheme for the fractional Riccati equation

We consider the multi-factor approximation (S^n, V^n) of Definition 7.4 with $\sigma(x) = \nu\sqrt{x}$, where the number of factors n is large, that is

$$dS_t^n = S_t^n\sqrt{V_t^n}dB_t, \quad V_t^n = g^n(t) + \sum_{i=1}^n c_i^n U_t^{n,i},$$

with

$$dU_t^{n,i} = (-x_i^n U_t^{n,i} - \lambda V_t^n)dt + \nu\sqrt{V_t^n}dW_t, \quad U_0^{n,i} = 0, \quad S_0^n = S_0.$$

Recall that g^n is given by (7.3.3) and it converges pointwise to g as n goes large, see Lemma 7.12.

We write the dynamics of (S^n, V^n) in terms of a Volterra Heston model with the smoothed kernel K^n given by (7.3.4) as follows

$$dS_t^n = S_t^n\sqrt{V_t^n}dB_t,$$

$$V_t^n = g^n(t) - \int_0^t K^n(t-s)\lambda V_s^n ds + \int_0^t K^n(t-s)\nu\sqrt{V_s^n}dW_s.$$

In particular, relying again on Chapter 6,

$$L^n(t,z) = \mathbb{E}\big[\exp\left(z\log(S_t^n/S_0)\right)\big]$$

is given by

$$\exp\left(\int_0^t F(z,\psi^n(t-s,z))g^n(s)ds\right), \tag{7.4.3}$$

where $\psi^n(\cdot,z)$ is the unique continuous solution of the Riccati-Volterra equation

$$\psi^n(t,z) = \int_0^t K^n(t-s)F(z,\psi^n(s,z))ds, \quad t\in[0,T], \tag{7.4.4}$$

for each $z\in\mathbb{C}$ with $\Re(z)\in[0,1]$.

Thanks to the weak uniqueness of the rough Heston model and to Theorem 7.7, $(S^n,V^n)_{n\geq 1}$ converges in law for the uniform topology to (S,V) when n tends to infinity. In particular, $L^n(t,z)$ converges pointwise to $L(t,z)$. Therefore, we expect $\psi^n(\cdot,z)$ to be close to the solution of the fractional Riccati equation (7.4.2). This is the object of the next theorem, whose proof is reported to Section 7.5.3 below.

Theorem 7.9. *There exists a positive constant C such that, for any $a\in[0,1]$, $b\in\mathbb{R}$ and $n\geq 1$,*

$$\sup_{t\in[0,T]}|\psi^n(t,a+ib)-\psi(t,a+ib)| \leq C(1+b^4)\int_0^T |K^n(s)-K(s)|ds,$$

where $\psi(\cdot,a+ib)$ (resp. $\psi^n(\cdot,a+ib)$) denotes the unique continuous solution of the Riccati Volterra equation (7.4.2) (resp. (7.4.4)).

Relying on the L^1-convergence of $(K^n)_{n\geq 1}$ to K under Assumption 7.3.1, see Proposition 7.5, we have the uniform convergence of $(\psi^n(\cdot,z))_{n\geq 1}$ to $\psi(\cdot,z)$ on $[0,T]$. Hence, Theorem 7.9 suggests a new numerical method for the computation of the fractional Riccati solution (7.4.2) where an explicit error is given. Indeed, set

$$\psi^{n,i}(t,z) = \int_0^t e^{-x_i^n(t-s)}F(z,\psi^n(s,z))ds, \quad i\in\{1,\dots,n\}.$$

Then,

$$\psi^n(t,z) = \sum_{i=1}^n c_i^n\psi^{n,i}(t,z),$$

and $(\psi^{n,i}(\cdot,z))_{1\leq i\leq n}$ solves the following n-dimensional system of ordinary Riccati equations

$$\partial_t\psi^{n,i}(t,z) = -x_i^n\psi^{n,i}(t,z) + F(z,\psi^n(t,z)), \quad \psi^{n,i}(0,z)=0, \quad i\in\{1,\dots,n\}. \tag{7.4.5}$$

Hence, (7.4.5) can be solved numerically by usual finite difference methods leading to $\psi^n(\cdot,z)$ as an approximation of the fractional Riccati solution.

7.4.2 Numerical illustrations

In this section, we consider a rough Heston model with the following parameters

$$\lambda = 0.3, \quad \rho = -0.7, \quad \nu = 0.3, \quad H = 0.1, \quad V_0 = 0.02, \quad \theta \equiv 0.02.$$

We discuss the accuracy of the multi-factor approximation sequence $(S^n, V^n)_{n\geq 1}$ as well as the corresponding Riccati Volterra solution $(\psi^n(\cdot, z))_{n\geq 1}$, for different choices of the weights $(c_i^n)_{1\leq i\leq n}$ and mean reversions $(x_i^n)_{1\leq i\leq n}$. This is achieved by first computing, for different number of factors n, the implied volatility $\sigma^n(k, T)$ of maturity T and log-moneyness k by a Fourier inversion of the characteristic function formula (7.4.3), see [27, 87] for instance. In a second step, we compare $\sigma^n(k, T)$ to the implied volatility $\sigma(k, T)$ of the rough Heston model. We also compare the Riccati Volterra solution $\psi^n(T, z)$ to the fractional one $\psi(T, z)$.

Note that the Riccati Volterra solution $\psi^n(\cdot, z)$ is computed by solving numerically the n-dimensional Riccati equation (7.4.5) with a classical finite difference scheme. The complexity of such scheme is $O(n \times n_{\Delta t})$, where $n_{\Delta t}$ is the number of time steps applied for the scheme, while the complexity of the Adam scheme used for the computation of $\psi(\cdot, z)$ is $O(n_{\Delta t}^2)$. In the following numerical illustrations, we fix $n_{\Delta t} = 200$.

In order to guarantee the convergence, the weights and mean reversions have to satisfy Assumption 7.3.1 and in particular they should be of the form (7.3.6) in terms of auxiliary mean reversions $(\eta_i^n)_{0\leq i\leq n}$ satisfying (7.3.9). For instance, one can fix

$$\eta_i^n = i\pi_n, \quad i \in \{0, \dots, n\}, \tag{7.4.6}$$

where π_n is defined by (7.3.10), as previously done in Section 7.3.2. For this particular choice, Figure 7.1 shows a decrease of the relative error $\left|\frac{\psi^n(T,ib)-\psi(T,ib)}{\psi(T,ib)}\right|$ towards zero for different values of b.

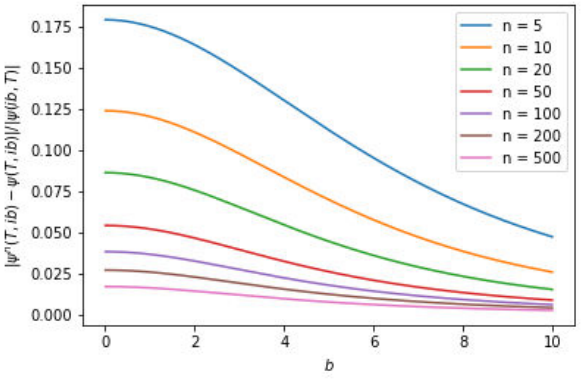

FIGURE 7.1: The relative error $\left|\frac{\psi^n(T,ib)-\psi(T,ib)}{\psi(T,ib)}\right|$ as a function of b under (7.4.6) and for different numbers of factors n with $T = 1$.

We also observe in the Figure 7.2 below that the implied volatility $\sigma^n(k, T)$ of the multi-factor approximation is close to $\sigma(k, T)$ for a number of factors $n \geq 20$. Notice that the approximation is more accurate around the money.

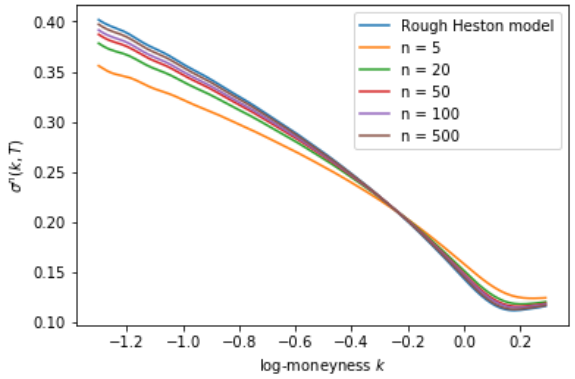

FIGURE 7.2: Implied volatility $\sigma^n(k,T)$ as a function of the log-moneyness k under (7.4.6) and for different numbers of factors n with $T = 1$.

In order to obtain a more accurate convergence, we can minimize the upper bound $f_n^{(2)}((\eta_i^n)_{0\leq i\leq n})$ of $\|K^n - K\|_{2,T}$ defined in (7.3.8). Hence, we choose $(\eta_i^n)_{0\leq i\leq n}$ to be a solution of the constrained minimization problem

$$\inf_{(\eta_i^n)_i\in\mathcal{E}_n} f_n^{(2)}((\eta_i^n)_{0\leq i\leq n}), \tag{7.4.7}$$

where $\mathcal{E}_n = \{(\eta_i^n)_{0\leq i\leq n};\quad 0 = \eta_0^n < \eta_1^n < ... < \eta_n^n\}$.

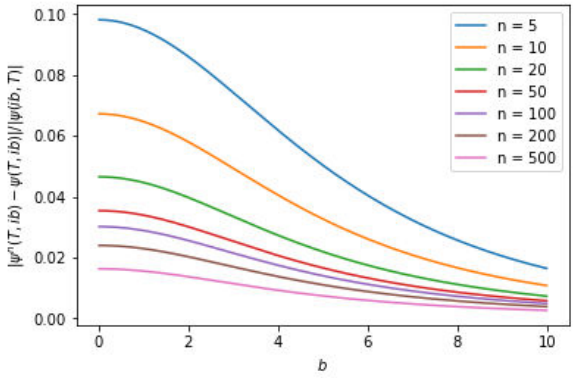

FIGURE 7.3: The relative error $\left|\frac{\psi^n(T,ib)-\psi(T,ib)}{\psi(T,ib)}\right|$ as a function of b under (7.4.7) and for different numbers of factors n with $T = 1$.

We notice from Figure 7.3, that the relative error $|\frac{\psi^n(T,ib)-\psi(T,ib)}{\psi(T,ib)}|$ is smaller under the choice of factors (7.4.7). Indeed the Volterra approximation $\psi^n(T,ib)$ is now closer to the fractional Riccati solution $\psi(T,ib)$ especially for small number of factors. However, when n is large, the accuracy of the approximation seems to be close to the one under (7.4.6). For instance when $n = 500$, the relative error is around 1% under both (7.4.6) and (7.4.7).

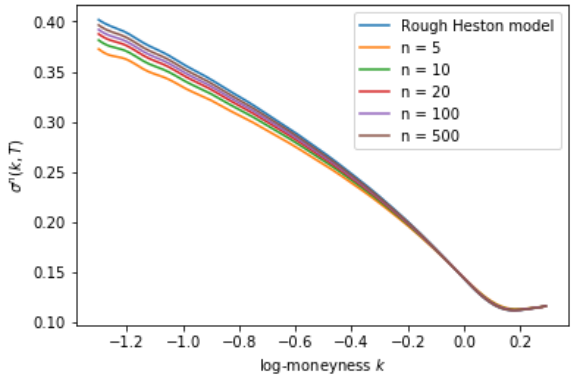

FIGURE 7.4: Implied volatility $\sigma^n(k,T)$ as a function of the log-moneyness k under (7.4.7) and for different numbers of factors n with $T = 1$.

In the same way, we observe in Figure 7.4 that the accuracy of the implied volatility approximation $\sigma^n(k,T)$ is more satisfactory under (7.4.7) especially for a small number of factors.

Theorem 7.9 states that the convergence of $\psi^n(\cdot,z)$ depends actually on the $L^1([0,T],\mathbb{R})$-error between K^n and K. Similarly to the computations of Section 7.3.2, we may show that,

$$\int_0^T |K^n(s) - K(s)|ds \leq f_n^{(1)}((\eta_i^n)_{0\leq i\leq n}),$$

where

$$f_n^{(1)}((\eta_i^n)_{0\leq i\leq n}) = \frac{T^3}{6}\sum_{i=1}^n \int_{\eta_{i-1}^n}^{\eta_i^n} (x - x_i^n)^2 \mu(dx) + \frac{1}{\Gamma(H+3/2)\Gamma(1/2-H)}(\eta_n^n)^{-H-\frac{1}{2}}.$$

This leads to choosing $(\eta_i^n)_{0\leq i\leq n}$ as a solution of the constrained minimization problem

$$\inf_{(\eta_i^n)_i \in \mathcal{E}_n} f_n^{(1)}((\eta_i^n)_{0\leq i\leq n}). \tag{7.4.8}$$

It is easy to show that such auxiliary mean-reversions $(\eta_i^n)_{0\leq i\leq n}$ satisfy (7.3.9) and thus Assumption 7.3.1 is met.

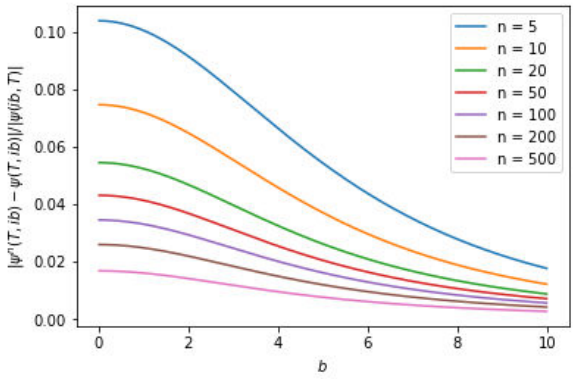

FIGURE 7.5: The relative error $\left|\frac{\psi^n(T,ib)-\psi(T,ib)}{\psi(T,ib)}\right|$ as a function of b under (7.4.8) and for different numbers of factors n with $T = 1$.

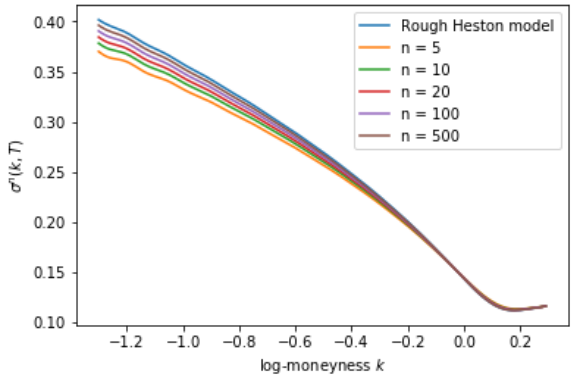

FIGURE 7.6: Implied volatility $\sigma^n(k,T)$ as a function of the log-moneyness k under (7.4.8) and for different numbers of factors n with $T = 1$.

Figures 7.5 and 7.6 exhibit similar results as the ones in Figures 7.3 and 7.4 corresponding to the choice of factors (7.4.7). In fact, we notice in practice that the solution of the minimization problem (7.4.7) is close to the one in (7.4.8).

7.4.3 Upper bound for call prices error

Using a Fourier transform method, we can also provide an error between the price of the call $C^n(k,T) = \mathbb{E}[(S_T^n - S_0 e^k)_+]$ in the multi-factor model and the price of the same call $C(k,T) = \mathbb{E}[(S_T - S_0 e^k)_+]$ in the rough Heston model. However, for technical reasons, this bound is obtained for a modification of the multi-factor approximation $(S^n, V^n)_{n\geq 1}$ of Definition 7.4 where the function g^n given initially by (7.3.3) is updated into

$$g^n(t) = \int_0^t K^n(t-s)\big(V_0 \frac{s^{-H-\frac{1}{2}}}{\Gamma(1/2-H)} + \theta(s)\big)ds, \tag{7.4.9}$$

where K^n is the smoothed approximation (7.3.4) of the fractional kernel. Note that the strong existence and uniqueness of V^n is still directly obtained from Proposition 7.22 and its non-negativity from Theorem 6.11 together with Remarks 7.23 and 7.24 below[4]. Although for g^n satisfying (7.4.9), $(V^n)_{n\geq 1}$ can not be tight[5], the corresponding spot price $(S^n)_{n\geq 1}$ converges as shown in the following proposition.

Proposition 7.10. *Let $(S^n, V^n)_{n\geq 1}$ be a sequence of multi-factor Heston models as in Definition 7.4 with $\sigma(x) = \nu\sqrt{x}$ and g^n given by (7.4.9). Then, under Assumption 7.3.1, $(S^n, \int_0^\cdot V_s^n ds)_{n\geq 1}$ converges in law for the uniform topology to $(S, \int_0^\cdot V_s ds)$, where (S,V) is a rough Heston model as in Definition 7.1 with $\sigma(x) = \nu\sqrt{x}$.*

Note that the characteristic function (7.4.3) still holds. Using Theorem 7.9 together with a Fourier transform method, we obtain an explicit error for the call prices. We refer to Section 7.5.5 below for the proof.

Proposition 7.11. *Let $C(k,T)$ be the price of the call in the rough Heston model with maturity $T > 0$ and log-moneyness $k \in \mathbb{R}$. We denote by $C^n(k,T)$ the price of the call in the*

[4]Note that Theorem 6.11 is used here for the smoothed kernel K^n, $b(x) = -\lambda x$ and g^n defined by (7.4.9).
[5]In fact, $V_0^n = 0$ while V_0 may be positive.

multi-factor Heston model of Definition 7.4 such that g^n is given by (7.4.9). If $|\rho| < 1$, then there exists a positive constant $c > 0$ such that

$$|C(k,T) - C^n(k,T)| \leq c \int_0^T |K(s) - K^n(s)| ds, \quad n \geq 1.$$

7.5 Proofs

In this section, we use the convolution notations together with the resolvent definitions of Section 7.6 below. We denote by c any positive constant independent of the variables t, h and n and that may vary from line to line. For any $h \in \mathbb{R}$, we will use the notation Δ_h to denote the semigroup operator of right shifts defined by $\Delta_h f : t \mapsto f(h+t)$ for any function f.

We first prove Theorem 7.8, which is the building block of Theorem 7.7. Then, we turn to the proofs of the results contained in Section 7.4, which concern the particular case of the rough Heston model.

7.5.1 Proof of Theorem 7.8

Tightness of $(X^n)_{n\geq 1}$: We first show that, for any $p \geq 2$,

$$\sup_{n\geq 1} \sup_{t\leq T} \mathbb{E}[|X_t^n|^p] < \infty. \tag{7.5.1}$$

Thanks to 7.7.2, we already have

$$\sup_{t\leq T} \mathbb{E}[|X_t^n|^p] < \infty. \tag{7.5.2}$$

Using the linear growth of (b, σ) and (7.5.2) together with Jensen and BDG inequalities, we get

$$\mathbb{E}[|X_t^n|^p] \leq c \left(\sup_{t\leq T} |g^n(t)|^p + \left(\int_0^T |K^n(s)|^2 ds \right)^{\frac{p}{2}-1} \int_0^t |K^n(t-s)|^2 (1 + \mathbb{E}[|X_s^n|^p]) ds) \right).$$

Relying on Assumption 7.3.2 and the convergence of $(g^n(0), \int_0^T |K^n(s)|^2 ds)_{n\geq 1}$, $\sup_{t\leq T} |g^n(t)|^p$ and $\int_0^T |K^n(s)|^2 ds$ are uniformly bounded in n. This leads to

$$\mathbb{E}[|X_t^n|^p] \leq c \left(1 + \int_0^t |K^n(t-s)|^2 \mathbb{E}[|X_s^n|^p] ds) \right).$$

By the Grönwall type inequality in Lemma 7.19 below, we deduce that

$$\mathbb{E}[|X_t^n|^p] \leq c \left(1 + \int_0^t E_c^n(s) ds) \right) \leq c \left(1 + \int_0^T E_c^n(s) ds) \right),$$

where $E_c^n \in L^1([0,T], \mathbb{R})$ is the canonical resolvent of $|K^n|^2$ with parameter c, as defined in Section 7.6 below, and the last inequality follows from the fact that $\int_0^{\cdot} E_c^n(s) ds$ is non-decreasing by Corollary 7.26. The convergence of $|K^n|^2$ to $|K|^2$ in $L^1([0,T], \mathbb{R})$ implies the

convergence of E_c^n to the canonical resolvent of $|K|^2$ with parameter c in $L^1([0,T],\mathbb{R})$, see [69, Theorem 2.3.1]. Thus, $\int_0^T E_c^n(s)ds$ is uniformly bounded in n, yielding (7.5.1).

We now show that $(X^n)_{n\geq 1}$ exhibits the Kolmogorov tightness criterion. In fact, using again the linear growth of (b,σ) and (7.5.1) together with Jensen and BDG inequalities, we obtain, for any $p\geq 2$ and $t,h\geq 0$ such that $t+h\leq T$,

$$\mathbb{E}[|X^n_{t+h}-X^n_t|^p]\leq c\Big(|g^n(t+h)-g^n(t)|^p+\big(\int_0^{T-h}|K^n(h+s)-K^n(s)|^2ds\big)^{p/2}+\big(\int_0^h|K^n(s)|^2ds\big)^{p/2}\Big).$$

Hence, Assumption 7.3.2 leads to

$$\mathbb{E}[|X^n_{t+h}-X^n_t|^p]\leq ch^{px},$$

and therefore to the tightness of $(X^n)_{n\geq 1}$ for the uniform topology.

Convergence of $(X^n)_{n\geq 1}$: Let $M^n_t=\int_0^t\sigma(X^n_s)dB^n_s$. As $\langle M^n\rangle_t=\int_0^t\sigma\sigma^*(X^n_s)ds$, $(\langle M^n\rangle)_{n\geq 1}$ is tight and consequently we get the tightness of $(M^n)_{n\geq 1}$ from [77, Theorem VI-4.13]. Let $(X,M)=(X_t,M_t)_{t\leq T}$ be a possible limit point of $(X^n,M^n)_{n\geq 1}$. Thanks to [77, Theorem VI-6.26], M is a local martingale and necessarily

$$\langle M\rangle_t=\int_0^t\sigma\sigma^*(X_s)ds,\quad t\in[0,T].$$

Moreover, setting $Y^n_t=\int_0^t b(X^n_s)ds+M^n_t$, the associativity property (5.2.4) yields

$$(L*X^n)_t=(L*g^n)(t)+\big(L*((K^n-K)*dY^n)\big)_t+Y^n_t, \tag{7.5.3}$$

where L is the resolvent of the first kind of K. By the Skorokhod representation theorem, we construct a probability space supporting a sequence of copies of $(X^n,M^n)_{n\geq 1}$ that converges uniformly on $[0,T]$, along a subsequence, to a copy of (X,M) almost surely, as n goes to infinity. We maintain the same notations for these copies. Hence, we have

$$\sup_{t\in[0,T]}|X^n_t-X_t|\to 0,\qquad \sup_{t\in[0,T]}|M^n_t-M_t|\to 0,$$

almost surely, as n goes to infinity. Relying on the continuity and linear growth of b together with the dominated convergence theorem, it is easy to obtain for any $t\in[0,T]$

$$(L*X^n)_t\to(L*X)_t,\qquad \int_0^t b(X^n_s)ds\to\int_0^t b(X_s)ds,$$

almost surely as n goes to infinity. Moreover for each $t\in[0,T]$

$$(L*g^n)(t)\to(L*g)(t),$$

by the uniform boundedness of g^n in n and t and the dominated convergence theorem. Finally thanks to the Jensen inequality,

$$\mathbb{E}[|\,(L*((K^n-K)*dY^n))_t\,|^2]\leq c\sup_{t\leq T}\mathbb{E}[|\,((K^n-K)*dY^n)_t\,|^2].$$

From (7.5.1) and the linear growth of (b,σ), we deduce

$$\sup_{t\leq T}\mathbb{E}[|\,((K^n-K)*dY^n)_t\,|^2]\leq c\int_0^T|K^n(s)-K(s)|^2ds,$$

which goes to zero when n is large. Consequently, we send n to infinity in (7.5.3) and obtain the following almost surely equality, for each $t \in [0,T]$,

$$(L * X)_t = (L * g)(t) + \int_0^t b(X_s)ds + M_t. \tag{7.5.4}$$

Recall also that $\langle M \rangle = \int_0^{\cdot} \sigma\sigma^*(X_s)ds$. Hence, by [101, Theorem V-3.9], there exists a m-dimensional Brownian motion W such that

$$M_t = \int_0^t \sigma(X_s)dB_s, \quad t \in [0,T].$$

The processes in (7.5.4) being continuous, we deduce that, almost surely,

$$(L * X)_t = (L * g)(t) + \int_0^t b(X_s)ds + \int_0^t \sigma(X_s)dB_s, \quad t \in [0,T].$$

We convolve by K and use the associativity property (5.2.4) to get that, almost surely,

$$\int_0^t X_s ds = \int_0^t g(s)ds + \int_0^t \left(\int_0^s K(s-u)(b(X_u)du + \sigma(X_u)dB_u) \right) ds, \quad t \in [0,T].$$

Finally it is easy to see that the processes above are differentiable and we conclude that X is solution of the stochastic Volterra equation (7.3.11) by continuity after taking the derivative.

7.5.2 Proof of Theorem 7.7

Theorem 7.7 is easily obtained once we prove the tightness of $(V^n)_{n\geq 1}$ for the uniform topology and that any limit point V is solution of the fractional stochastic integral equation (7.1.2). This is a direct consequence of Theorem 7.8, by setting $d = m = 1$, g and g^n respectively as in (7.3.1) and (7.3.3), $b(x) = -\lambda x$, K being the fractional kernel and $K^n(t) = \sum_{i=1}^n c_i^n e^{-x_i^n t}$ its smoothed approximation. Under Assumption 7.3.1, $(K^n)_{n\geq 1}$ converges in $L^2([0,T],\mathbb{R})$ to the fractional kernel, see Proposition 7.5. Hence, it is left to show the pointwise convergence of $(g^n)_{n\geq 1}$ to g on $[0,T]$ and that $(K^n, g^n)_{n\geq 1}$ satisfies Assumption 7.3.2.

Lemma 7.12 (Convergence of g^n)**.** *Define $g^n : [0,T] \mapsto \mathbb{R}$ and $g : [0,T] \mapsto \mathbb{R}$ respectively by (7.3.1) and (7.3.3) such that $\theta : [0,T] \mapsto \mathbb{R}$ satisfies (7.2.1). Under assumption (7.3.1), we have for any $t \in [0,T]$*

$$g^n(t) \to g(t),$$

as n tends to infinity.

Proof. Because θ satisfies (7.2.1), it is enough to show that for each $t \in [0,T]$

$$\int_0^t (t-s)^{-\frac{1}{2}-\varepsilon} |K^n(s) - K(s)| ds \tag{7.5.5}$$

converges to zero as n goes large, for some $\varepsilon > 0$ and K^n given by (7.3.4). Using the representation of K as the Laplace transform of μ as in (7.1.3), we obtain that (7.5.5) is bounded by

$$\int_0^t (t-s)^{-\frac{1}{2}-\varepsilon} \int_{\eta_n^n}^{\infty} e^{-xs}\mu(dx)ds + \sum_{i=1}^n \int_0^t (t-s)^{-\frac{1}{2}-\varepsilon} |c_i^n e^{-x_i^n s} - \int_{\eta_{i-1}^n}^{\eta_i^n} e^{-xs}\mu(dx)|ds. \tag{7.5.6}$$

The first term in (7.5.6) converges to zero for large n by the dominated convergence theorem because η_n^n tends to infinity, see Assumption 7.3.1. Using the Taylor-Lagrange inequality (7.3.7), the second term in (7.5.6) is dominated by

$$\frac{1}{2}\int_0^t (t-s)^{-\frac{1}{2}-\varepsilon} s^2 ds \sum_{i=1}^n \int_{\eta_{i-1}^n}^{\eta_i^n} (x - x_i^n)^2 \mu(dx),$$

which goes to zero thanks to Assumption 7.3.1. □

Lemma 7.13 (K^n satisfying Assumption 7.3.2)**.** *Under Assumption 7.3.1, there exists $C > 0$ such that, for any $t, h \geq 0$ with $t + h \leq T$,*

$$\sup_{n\geq 1} \left(\int_0^{T-h} |K^n(h+s) - K^n(s)|^2 ds + \int_0^h |K^n(s)|^2 ds \right) \leq C h^{2H},$$

where K^n is defined by (7.3.4).

Proof. We start by proving that for any $t, h \geq 0$ with $t + h \leq T$

$$\int_0^h |K^n(s)|^2 ds \leq c h^{2H}. \tag{7.5.7}$$

In fact we know that this inequality is satisfied for $K(t) = \frac{t^{H-\frac{1}{2}}}{\Gamma(H+1/2)}$. Thus it is enough to prove

$$\|K^n - K\|_{2,h} \leq c h^H,$$

where $\|\cdot\|_{2,h}$ stands for the usual $L^2([0,h],\mathbb{R})$ norm. Relying on the Laplace transform representation of K given by (7.1.3), we obtain

$$\|K^n - K\|_{2,h} \leq \int_{\eta_n^n}^{\infty} \|e^{-x(\cdot)}\|_{2,h} \mu(dx) + \sum_{i=1}^n J_{i,h}^n,$$

where $J_{i,h}^n = \|c_i^n e^{-x_i^n(\cdot)} - \int_{\eta_{i-1}^n}^{\eta_i^n} e^{-x(\cdot)} \mu(dx)\|_{2,h}$. We start by bounding the first term,

$$\begin{aligned}\int_{\eta_n^n}^{\infty} \|e^{-x(\cdot)}\|_{2,h} \mu(dx) &\leq \int_0^{\infty} \sqrt{\frac{1 - e^{-2xh}}{2x}} \mu(dx) \\ &= \frac{h^H}{\Gamma(H+1/2)\Gamma(1/2-H)\sqrt{2}} \int_0^{\infty} \sqrt{\frac{1-e^{-2x}}{x}} x^{-H-\frac{1}{2}} dx.\end{aligned}$$

As in Section 7.3.2, we use the Taylor-Lagrange inequality (7.3.7) to get

$$\sum_{i=1}^n J_{i,h}^n \leq \frac{1}{2\sqrt{5}} h^{\frac{5}{2}} \sum_{i=1}^n \int_{\eta_{i-1}^n}^{\eta_i^n} (x - x_i^n)^2 \mu(dx).$$

Using the boundedness of $\left(\sum_{i=1}^n \int_{\eta_{i-1}^n}^{\eta_i^n} (x - x_i^n)^2 \mu(dx)\right)_{n\geq 1}$ from Assumption 7.3.1, we deduce (7.5.7). We now prove

$$\int_0^{T-h} |K^n(h+s) - K^n(s)|^2 ds \leq c h^{2H}. \tag{7.5.8}$$

In the same way, it is enough to show

$$\|(\Delta_h K^n - \Delta_h K) - (K^n - K)\|_{2,T-h} \le c h^H,$$

Similarly to the previous computations, we get

$$\|(\Delta_h K^n - \Delta_h K) - (K^n - K)\|_{2,T-h} \le \int_{\eta_n^n}^{\infty} \|e^{-x(\cdot)} - e^{-x(h+\cdot)}\|_{2,T-h}\mu(dx) + \sum_{i=1}^{n} \widetilde{J}_{i,h}^n,$$

with $\widetilde{J}_{i,h}^n = \|c_i^n(e^{-x_i^n(\cdot)} - e^{-x_i^n(h+\cdot)}) - \int_{\eta_{i-1}^n}^{\eta_i^n}(e^{-x(\cdot)} - e^{-x(h+\cdot)})\mu(dx)\|_{2,T-h}$. Notice that

$$\begin{aligned}\int_{\eta_n^n}^{\infty} \|e^{-x(\cdot)} - e^{-x(h+\cdot)}\|_{2,T-h}\mu(dx) &= \int_{\eta_n^n}^{\infty}(1-e^{-xh})\sqrt{\frac{1-e^{-2x(T-h)}}{2x}}\mu(dx)\\ &\le c\int_0^{\infty}(1-e^{-xh})x^{-H-1}dx \le ch^H.\end{aligned}$$

Moreover, fix $h, t > 0$ and set $\chi(x) = e^{-xt} - e^{-x(t+h)}$. The second derivative reads

$$\chi''(x) = h(t^2 x e^{-xt}\frac{1-e^{-xh}}{xh} - he^{-x(t+h)} - 2te^{-x(t+h)}), \quad x > 0. \tag{7.5.9}$$

Because $x \mapsto xe^{-x}$ and $x \mapsto \frac{1-e^{-x}}{x}$ are bounded functions on $(0,\infty)$, there exists $C > 0$ independent of $t, h \in [0,T]$ such that

$$|\chi''(x)| \le Ch, \quad x > 0.$$

The Taylor-Lagrange formula, up to the second order, leads to

$$|c_i^n(e^{-x_i^n t} - e^{-x_i^n(t+h)}) - \int_{\eta_{i-1}^n}^{\eta_i^n}(e^{-xt} - e^{-x(t+h)})\mu(dx)| \le \frac{C}{2}h\int_{\eta_{i-1}^n}^{\eta_i^n}(x - x_i^n)^2\mu(dx).$$

Thus

$$\sum_{i=1}^{n}\widetilde{J}_{i,h}^n \le \frac{C}{2}h\sum_{i=1}^{n}\int_{\eta_{i-1}^n}^{\eta_i^n}(x-x_i^n)^2\mu(dx).$$

Finally, (7.5.8) follows from the boundedness of $\left(\sum_{i=1}^n \int_{\eta_{i-1}^n}^{\eta_i^n}(x-x_i^n)^2\mu(dx)\right)_{n\ge1}$ due to Assumption 7.3.1. □

Lemma 7.14 (g^n satisfying Assumption 7.3.2)**.** *Define $g^n : [0,T] \mapsto \mathbb{R}$ by (7.3.3) such that $\theta : [0,T] \mapsto \mathbb{R}$ satisfies (7.2.1). Under Assumption 7.3.1, for each $\varepsilon > 0$, there exists $C_\varepsilon > 0$ such that for any $t, h \ge 0$ with $t + h \le T$*

$$\sup_{n\ge1}|g^n(t) - g^n(t+h)| \le C_\varepsilon h^{H-\varepsilon}.$$

Proof. Because θ satisfies (7.2.1), it is enough to prove that, for each fixed $\varepsilon > 0$, there exists $C > 0$ such that

$$\sup_{n\ge1}\int_0^h (h-s)^{-\frac{1}{2}-\varepsilon}|K^n(s)|ds \le Ch^{H-\varepsilon}, \tag{7.5.10}$$

and

$$\sup_{n\ge1}\int_0^t (t-s)^{-\frac{1}{2}-\varepsilon}|K^n(s) - K^n(h+s)|ds \le Ch^{H-\varepsilon}, \tag{7.5.11}$$

for any $t, h \geq 0$ with $t+h \leq T$. (7.5.10) being satisfied for the fractional kernel, it is enough to establish

$$\int_0^h (h-s)^{-\frac{1}{2}-\varepsilon} |K^n(s) - K(s)| ds \leq c h^{H-\varepsilon}.$$

In the proof of Lemma 7.12, it is shown that

$$\int_0^h (h-s)^{-\frac{1}{2}-\varepsilon} |K^n(s) - K(s)| ds$$

is bounded by (7.5.6), that is

$$\int_0^h (h-s)^{-\frac{1}{2}-\varepsilon} \int_{\eta_n^n}^\infty e^{-xs} \mu(dx) ds + \sum_{i=1}^n \int_0^h (h-s)^{-\frac{1}{2}-\varepsilon} |c_i^n e^{-x_i^n s} - \int_{\eta_{i-1}^n}^{\eta_i^n} e^{-xs} \mu(dx)| ds.$$

The first term is dominated by

$$\int_0^h (h-s)^{-\frac{1}{2}-\varepsilon} \int_0^\infty e^{-xs} \mu(dx) ds = h^{H-\varepsilon} \frac{B(1/2-\varepsilon, H+1/2)}{B(1/2-H, H+1/2)},$$

where B is the usual Beta function. Moreover thanks to (7.3.7) and Assumption 7.3.1, we get

$$\sum_{i=1}^n \int_0^h (h-s)^{-\frac{1}{2}-\varepsilon} |c_i^n e^{-x_i^n s} - \int_{\eta_{i-1}^n}^{\eta_i^n} e^{-xs} \mu(dx)| ds \leq c h^{\frac{5}{2}-\varepsilon},$$

yielding (7.5.10). Similarly, we obtain (7.5.11) by showing that

$$\int_0^t (t-s)^{-\frac{1}{2}-\varepsilon} \left| (K^n(s) - \Delta_h K^n(s)) - (K(s) - \Delta_h K(s)) \right| ds \leq c h^{H-\varepsilon}.$$

By similar computations as previously and using (7.5.9), we get that

$$\int_0^t (t-s)^{-\frac{1}{2}-\varepsilon} \left| (K^n(s) - \Delta_h K^n(s)) - (K(s) - \Delta_h K(s)) \right| ds$$

is dominated by

$$c \left(\int_0^t (t-s)^{\frac{1}{2}-\varepsilon} \int_{\eta_n^n}^\infty (1-e^{-xh}) e^{-xs} \mu(dx) ds + h \sum_{i=1}^n \int_{\eta_{i-1}^n}^{\eta_i^n} (x - x_i^n)^2 \mu(dx) \right).$$

The first term being bounded by

$$\int_0^t (t-s)^{\frac{1}{2}-\varepsilon} \int_0^\infty (1-e^{-xh}) e^{-xs} \mu(dx) ds = \int_0^t (t-s)^{\frac{1}{2}-\varepsilon} (K(s) - K(h+s)) ds \leq c h^{H-\varepsilon},$$

Assumption 7.3.1 leads to (7.5.11). □

7.5.3 Proof of Theorem 7.9

Uniform boundedness : We start by showing the uniform boundedness of the unique continuous solutions $(\psi^n(\cdot, a+ib))_{n\geq 1}$ of (7.4.4).

Proposition 7.15. *For a fixed $T > 0$, there exists $C > 0$ such that*

$$\sup_{n\geq 1}\sup_{t\in[0,T]} |\psi^n(t, a + ib)| \leq C\left(1 + b^2\right),$$

for any $a \in [0, 1]$ and $b \in \mathbb{R}$.

Proof. Let $z = a + ib$ and start by noticing that $\Re(\psi^n(\cdot, z))$ is non-positive because it solves the following linear Volterra equation with continuous coefficients

$$\chi = K^n * \left(f + \left(\rho\nu\Re(z) - \lambda + \frac{\nu^2}{2}\Re(\psi^n(\cdot, z))\right)\chi\right),$$

where

$$f = \frac{1}{2}\left(a^2 - a - (1 - \rho^2)b^2\right) - \frac{1}{2}(\rho b + \nu\psi^n(\cdot, z))^2$$

is continuous non-positive, see Theorem 7.25. In the same way $\Re(\psi(\cdot, z))$ is also non-positive. Moreover, observe that $\psi^n(\cdot, z)$ solves the following linear Volterra equation with continuous coefficients

$$\chi = K^n * \left(\frac{1}{2}(z^2 - z) + (\rho\nu z - \lambda + \frac{\nu^2}{2}\psi^n(\cdot, z))\chi\right),$$

and

$$\Re\left(\rho\nu z - \lambda + \frac{\nu^2}{2}\psi^n(\cdot, z)\right) \leq \nu - \lambda.$$

Therefore, Corollary 7.28 leads to

$$\sup_{t\in[0,T]} |\psi^n(t, z)| \leq \frac{1}{2}|z^2 - z| \int_0^T E^n_{\nu-\lambda}(s)ds,$$

where $E^n_{\nu-\lambda}$ denotes the canonical resolvent of K^n with parameter $\nu - \lambda$, see Section 7.6 below. This resolvent converges in $L^1([0, T], \mathbb{R})$ because K^n converges in $L^1([0, T], \mathbb{R})$ to K, see [69, Theorem 2.3.1]. Hence, $(\int_0^T E^n_{\nu-\lambda}(s)ds)_{n\geq 1}$ is bounded, which ends the proof. □

End of the proof of Theorem 7.9 : Set $z = a + ib$ and recall that

$$\psi^n(\cdot, z) = K^n * F(z, \psi^n(\cdot, z)); \quad \psi(\cdot, z) = K * F(z, \psi(\cdot, z)).$$

with $F(z, x) = \frac{1}{2}(z^2 - z) + (\rho\nu z - \lambda)x + \frac{\nu^2}{2}x^2$. Hence, for $t \in [0, T]$,

$$\psi(t, z) - \psi^n(t, z) = h^n(t, z) + K * (F(z, \psi(\cdot, z)) - F(z, \psi^n(\cdot, z)))(t),$$

with $h^n(\cdot, z) = (K^n - K) * F(z, \psi^n(\cdot, z))$. Thanks to Proposition 7.15, we get the existence of a positive constant C such that

$$\sup_{n\geq 1}\sup_{t\in[0,T]} |h^n(t, a + ib)| \leq C(1 + b^4)\int_0^T |K^n(s) - K(s)|ds, \tag{7.5.12}$$

for any $b \in \mathbb{R}$ and $a \in [0, 1]$. Moreover notice that $(\psi - \psi^n - h^n)(\cdot, z)$ is solution of the following linear Volterra equation with continuous coefficients

$$\chi = K * \left((\rho\nu z - \lambda + \frac{\nu^2}{2}(\psi + \psi_n)(\cdot, z))(\chi + h^n(\cdot, z))\right),$$

and remark that the real part of $\rho\nu z - \lambda + \frac{\nu^2}{2}(\psi + \psi_n)(\cdot, z)$ is dominated by $\nu - \lambda$ because $\Re(\psi(\cdot, z))$ and $\Re(\psi^n(\cdot, z))$ are non-positive. An application of Corollary 7.28 together with (7.5.12) ends the proof.

7.5.4 Proof of Proposition 7.10

We consider for each $n \geq 1$, (S^n, V^n) defined by the multi-factor Heston model in Definition 7.4 with $\sigma(x) = \nu\sqrt{x}$.

Tightness of $(\int_0^\cdot V_s^n ds, \int_0^\cdot \sqrt{V_s^n} dB_s, \int_0^\cdot \sqrt{V_s^n} dW_s)_{n\geq 1}$ **:** Because the process $\int_0^\cdot V_s^n ds$ is non-decreasing, it is enough to show that

$$\sup_{n\geq 1} \mathbb{E}[\int_0^T V_t^n dt] < \infty, \tag{7.5.13}$$

to obtain its tightness for the uniform topology. Recalling that $\sup_{t\in[0,T]} \mathbb{E}[V_t^n] < \infty$ from (7.7.2) below, we get

$$\mathbb{E}\left[\int_0^t \sqrt{V_s^n} dW_s\right] = 0,$$

and then by Fubini theorem

$$E[V_t^n] = g^n(t) + \sum_{i=1}^n c_i^n \mathbb{E}[U_t^{n,i}],$$

with

$$\mathbb{E}[U_t^{n,i}] = \int_0^t (-x_i^n \mathbb{E}[U_s^{n,i}] - \lambda \mathbb{E}[V_s^n]) ds.$$

Thus $t \mapsto \mathbb{E}[V_t^n]$ solves the following linear Volterra equation

$$\chi(t) = \int_0^t K^n(t-s) \left(-\lambda\chi(s) + \theta(s) + V_0 \frac{s^{-H-\frac{1}{2}}}{\Gamma(1/2 - H)}\right) ds,$$

with K^n given by (7.3.4). Theorem 7.18 below leads to

$$\mathbb{E}[V_t^n] = \int_0^t E_\lambda^n(t-s) \left(\theta(s) + V_0 \frac{s^{-H-\frac{1}{2}}}{\Gamma(\frac{1}{2} - H)}\right) ds,$$

and then by Fubini theorem again

$$\int_0^t \mathbb{E}[V_s^n] ds = \int_0^t \left(\int_0^{t-s} E_\lambda^n(u) du\right) \left(\theta(s) + V_0 \frac{s^{-H-\frac{1}{2}}}{\Gamma(\frac{1}{2} - H)}\right) ds,$$

where E_λ^n is the canonical resolvent of K^n with parameter λ, defined in Section 7.6. Because $(K^n)_{n\geq 1}$ converges to the fractional kernel K in $L^1([0,T], \mathbb{R})$, we obtain the convergence of E_λ^n in $L^1([0,T], \mathbb{R})$ to the canonical resolvent of K with parameter λ, see [69, Theorem 2.3.1]. In particular thanks to Corollary 7.26 below, $\int_0^t E_\lambda^n(s) ds$ is uniformly bounded in $t \in [0,T]$ and $n \geq 1$. This leads to (7.5.13) and then to the tightness of $(\int_0^\cdot V_s^n ds, \int_0^\cdot \sqrt{V_s^n} dB_s, \int_0^\cdot \sqrt{V_s^n} dW_s)_{n\geq 1}$ by [77, Theorem VI-4.13].

Convergence of $(S^n, \int_0^\cdot V_s^n ds)_{n\geq 1}$ **:** We set $M_t^{n,1} = \int_0^t \sqrt{V_s^n} dB_s$ and $M_t^{n,2} = \int_0^t \sqrt{V_s^n} dW_s$. Denote by (X, M^1, M^2) a limit in law for the uniform topology of a subsequence of the tight

family $(\int_0^\cdot V_s^n ds, M^{n,1}, M^{n,2})_{n\geq 1}$. An application of stochastic Fubini theorem, see [111], yields

$$\int_0^t V_s^n ds = \int_0^t \int_0^{t-s} (K^n(u) - K(u)) du dY_s^n + \int_0^t K(t-s) Y_s^n ds, \quad t \in [0,T], \tag{7.5.14}$$

where $Y_t^n = \int_0^t (s^{-H-\frac{1}{2}} \frac{V_0}{\Gamma(1/2-H)} + \theta(s) - \lambda V_s^n) ds + \nu M_t^{n,2}$. Because $(Y^n)_{n\geq 1}$ converges in law for the uniform topology to $Y = (Y_t)_{t\leq T}$ given by $Y_t = \int_0^t (s^{-H-\frac{1}{2}} \frac{V_0}{\Gamma(\frac{1}{2}-H)} + \theta(s)) ds - \lambda X_t + \nu M_t^2$, we also get the convergence of $(\int_0^\cdot K(\cdot - s) Y_s^n ds)_{n\geq 1}$ to $\int_0^\cdot K(\cdot - s) Y_s ds$. Moreover, for any $t \in [0,T]$,

$$\left| \int_0^t \int_0^{t-s} (K^n(u) - K(u)) du \left(s^{-H-\frac{1}{2}} \frac{V_0}{\Gamma(\frac{1}{2}-H)} + \theta(s) - \lambda V_s^n \right) ds \right|$$

is bounded by

$$\int_0^t |K^n(s) - K(s)| ds \left(\int_0^t (s^{-H-\frac{1}{2}} \frac{V_0}{\Gamma(\frac{1}{2}-H)} + \theta(s)) ds + \lambda \int_0^t V_s^n ds \right),$$

which converges in law for the uniform topology to zero thanks to the convergence of $(\int_0^\cdot V_s^n ds)_{n\geq 1}$ together with Proposition 7.5. Finally,

$$\mathbb{E}\left[\left| \int_0^t \int_0^{t-s} (K^n(u) - K(u)) du dM_s^{n,2} \right|^2 \right] \leq c \int_0^T (K^n(s) - K(s))^2 ds \mathbb{E}\left[\int_0^t V_s^n ds \right],$$

which goes to zero thanks to (7.5.13) and Proposition 7.5. Hence, by passing to the limit in (7.5.14), we obtain

$$X_t = \int_0^t K(t-s) Y_s ds,$$

for any $t \in [0,T]$, almost surely. The processes being continuous, the equality holds on $[0,T]$. Then, by the stochastic Fubini theorem, we deduce that $X = \int_0^\cdot V_s ds$, where V is a continuous process defined by

$$V_t = \int_0^t K(t-s) dY_s = V_0 + \int_0^t K(t-s)(\theta(s) - \lambda V_s) ds + \nu \int_0^t K(t-s) dM_s^2.$$

Furthermore because $(M^{n,1}, M^{n,2})$ is a martingale with bracket

$$\int_0^\cdot V_s^n ds \begin{pmatrix} 1 & \rho \\ \rho & 1 \end{pmatrix},$$

[77, Theorem VI-6.26] implies that (M^1, M^2) is a local martingale with the following bracket

$$\int_0^\cdot V_s ds \begin{pmatrix} 1 & \rho \\ \rho & 1 \end{pmatrix}.$$

By [101, Theorem V-3.9], there exists a two-dimensional Brownian motion $(\widetilde{W}, \widetilde{B})$ with $d\langle \widetilde{W}, \widetilde{B} \rangle_t = \rho dt$ such that

$$M_t^1 = \int_0^t \sqrt{V_s} d\widetilde{W}_s, \quad M_t^2 = \int_0^t \sqrt{V_s} d\widetilde{B}_s, \quad t \in [0,T].$$

In particular V is solution of the fractional stochastic integral equation in Definition 7.1 with $\sigma(x) = \nu\sqrt{x}$. Because $S^n = \exp(M^{n,1} - \frac{1}{2}\int_0^{\cdot} V_s^n ds)$, we deduce the convergence of $(S^n, \int_0^{\cdot} V_s^n ds)_{n\geq 1}$ to the limit point $(S, \int_0^{\cdot} V_s ds)$ that displays the rough-Heston dynamics of Definition 7.1. The uniqueness of such dynamics, recall Chapters 5 and 6, enables us to conclude that $(S^n, V^n)_{n\geq 1}$ admits a unique limit point and hence converges to the rough Heston dynamics.

7.5.5 Proof of Proposition 7.11

We use the Lewis Fourier inversion method, see [87], to write

$$C^n(k,T) - C(k,T) = S_0 \frac{e^{\frac{k}{2}}}{2\pi} \int_{b\in\mathbb{R}} \frac{e^{-ibk}}{b^2+\frac{1}{4}} \left(L(T, \frac{1}{2}+ib) - L^n(T, \frac{1}{2}+ib) \right) dW.$$

Hence,

$$|C^n(k,T) - C(k,T)| \leq S_0 \frac{e^{\frac{k}{2}}}{2\pi} \int_{b\in\mathbb{R}} \frac{1}{b^2+\frac{1}{4}} \left| L(T, \frac{1}{2}+ib) - L^n(T, \frac{1}{2}+ib) \right| dW. \tag{7.5.15}$$

Because $L(T,z)$ and $L^n(T,z)$ satisfy respectively the formulas (7.4.1) and (7.4.3) with g and g^n given by

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

and $\psi(\cdot,z)$ and $\psi^n(\cdot,z)$ solve respectively (7.4.2) and (7.4.4), we use the Fubini theorem to deduce that

$$L(T,z) = \exp\left(\int_0^T \psi(T-s,z)(V_0 \frac{s^{-H-\frac{1}{2}}}{\Gamma(1/2-H)} + \theta(s))ds \right), \tag{7.5.16}$$

and

$$L^n(T,z) = \exp\left(\int_0^T \psi^n(T-s,z)(V_0 \frac{s^{-H-\frac{1}{2}}}{\Gamma(1/2-H)} + \theta(s))ds \right), \tag{7.5.17}$$

with $z = 1/2 + ib$. Therefore, relying on the local Lipschitz property of the exponential function, it suffices to find an upper bound for $\Re(\psi^n(\cdot,z))$ in order to get an error for the price of the call from (7.5.15). This is the object of the next proposition.

Upper bound of $\Re(\psi^n(\cdot,z))$: We denote by $\phi_\eta^n(\cdot,b)$ the unique continuous function satisfying the following Riccati Volterra equation

$$\phi_\eta^n(\cdot,b) = K^n * \left(-b + \eta\phi_\eta^n(\cdot,b) + \frac{\nu^2}{2}\phi_\eta^n(\cdot,b)^2 \right),$$

with $b \geq 0$ and $\eta, \nu \in \mathbb{R}$.

Proposition 7.16. *Fix $b_0, t_0 \geq 0$ and $\eta \in \mathbb{R}$. The functions $b \mapsto \phi_\eta^n(t_0,b)$ and $t \mapsto \phi_\eta^n(t,b_0)$ are non-increasing on $\mathbb{R}_+$. Furthermore*

$$\phi_\eta^n(t,b) \leq \frac{1 - \sqrt{1 + 2b\nu^2(\int_0^t E_\eta^n(s)ds)^2}}{\nu^2 \int_0^t E_\eta^n(s)ds}, \quad t > 0,$$

where E_η^n is the canonical resolvent of K^n with parameter η defined in Section 7.6 below.

Proof. The claimed monotonicity of $b \mapsto \phi_\eta^n(t_0, b)$ is directly obtained from Theorem 7.25. Consider now $h, b_0 > 0$. It is easy to see that $\Delta_h\phi_\eta^n(\cdot, b_0)$ solves the following Volterra equation

$$\Delta_h\phi_\eta^n(b_0, t) = \left(\Delta_t K^n * F(\phi_\eta^n(\cdot, b_0))\right)(h) + \left(K^n * F(\Delta_h\phi_\eta^n(\cdot, b_0))\right)(t)$$

with $F(b, x) = -b + \eta x + \frac{\nu^2}{2}x^2$. Notice that $t \to -\left(\Delta_t K^n * F(\phi_\eta^n(\cdot, b_0))\right)(h) \in \mathcal{G}_K$, as defined in (6.2.5), thanks to Theorem 7.25 below. $\phi_\eta^n(\cdot, b) - \Delta_h\phi_\eta^n(\cdot, b)$ being solution of the following linear Volterra integral equation with continuous coefficients,

$$x(t) = -\left(\Delta_t K^n * F(b, \phi_\eta^n(\cdot, b_0))\right)(h) + \left(K^n * \left(\left(\eta + \frac{\nu^2}{2}(\phi_\eta^n(\cdot, b) + \Delta_h\phi_\eta^n(\cdot, b))\right)x\right)\right)(t),$$

we deduce its non-negativity using again Theorem 7.25. Thus, $t \in \mathbb{R}_+ \to \phi_\eta^n(t, b_0)$ is non-increasing and consequently $\sup_{s\in[0,t]} |\phi_\eta(s, b)| = |\phi_\eta^n(t, b_0)|$ as $\phi_\eta^n(0, b) = 0$. Hence, Theorem 7.18 below leads to

$$\phi_\eta^n(t, b) = \int_0^t E_\eta^n(t-s)(-b + \frac{\nu^2}{2}\phi_\eta^n(s, b)^2) \le \int_0^t E_\eta^n(s)ds\left(-b + \frac{\nu^2}{2}\phi_\eta^n(t, b)^2\right).$$

We end the proof by solving this inequality of second order in $\phi_\eta^n(t, b)$ and using that ϕ_η^n is non-positive. Notice that $\int_0^t E_\eta^n(s)ds > 0$ for each $t > 0$, see Corollary 7.26 below. □

Corollary 7.17. *Fix $a \in [0, 1]$. We have, for any $t \in (0, T]$ and $b \in \mathbb{R}$,*

$$\sup_{n\ge1} \Re(\psi^n(t, a + ib)) \le \frac{1 - \sqrt{1 + (a - a^2 + (1-\rho^2)b^2)\nu^2 m(t)^2}}{\nu^2 m(t)}$$

where $m(t) = \inf_{n\ge1} \int_0^t E_{\rho\nu a-\lambda}^n(s)ds > 0$ for all $t \in (0, T]$ and E_η^n is the canonical resolvent of K^n with parameter η defined in Section 7.6.

Proof. Let $r = a - a^2 + (1-\rho^2)b^2$ and $\eta = \rho\nu a - \lambda$. $\phi_\eta^n(\cdot, r) - \Re(\psi^n(\cdot, a + ib))$ being solution of the following linear Volterra equation with continuous coefficients

$$\chi = K * \left(\frac{1}{2}\left(\rho b + \nu\Im(\psi^n(\cdot, a + ib))\right)^2 + \left(\rho\nu a - \lambda + \frac{\nu^2}{2}\left(\Re(\psi^n(\cdot, a + ib)) + \phi_\eta(\cdot, r)\right)\right)\chi\right),$$

we use Theorem 7.25 together with Proposition 7.16 to get, for all $t \in [0, T]$ and $b \in \mathbb{R}$,

$$\Re(\psi^n(t, a + ib)) \le \frac{1 - \sqrt{1 + 2r\nu^2(\int_0^t E_\eta^n(s)ds)^2}}{\nu^2 \int_0^t E_\eta^n(s)ds}. \tag{7.5.18}$$

Moreover for any $t \in [0, T]$, $\int_0^t E_\eta^n(s)ds$ converges as n goes to infinity to $\int_0^t E_\eta(s)ds$ because K^n converges to K in $L^1([0, T], \mathbb{R})$, see [69, Theorem 2.3.1], where E_η denotes the canonical resolvent of K with parameter η. Therefore, $m(t) = \inf_{n\ge1} \int_0^t E_\eta^n(s)ds > 0$, for all $t \in (0, T]$, because $\int_0^t E_\eta(s)ds > 0$ and $\int_0^t E_\eta^n(s)ds > 0$ for all $n \ge 1$, see Corollary 7.26. Finally we end the proof by using (7.5.18) together with the fact that $x \mapsto \frac{1-\sqrt{1+2r\nu^2x^2}}{\nu^2 x}$ is non-increasing on $(0, \infty)$.

End of the proof of Proposition 7.11 : Assume that $|\rho| < 1$ and fix $a = 1/2$. By dominated convergence theorem,

$$\int_0^T \frac{1-\sqrt{1+(a-a^2+(1-\rho^2)b^2)\nu^2 m(T-s)^2}}{\nu^2 m(T-s)}(\theta(s)+V_0\frac{s^{-H-\frac{1}{2}}}{\Gamma(\frac{1}{2}-H)})ds$$

is equivalent to

$$-|b|\frac{\sqrt{1-\rho^2}}{\nu}\int_0^T(\theta(s)+V_0\frac{s^{-H-\frac{1}{2}}}{\Gamma(\frac{1}{2}-H)})ds,$$

as b tends to infinity. Hence, thanks to Corollary 7.17, there exists $C>0$ such that for any $b\in\mathbb{R}$

$$\sup_{n\geq 1}\Re(\psi^n(t,a+ib))\leq C(1-|b|). \tag{7.5.19}$$

Recalling that

$$\forall z_1,z_2\in\mathbb{C} \text{ such that } \Re(z_1),\Re(z_2)\leq c,\quad |e^{z_1}-e^{z_2}|\leq e^c|z_1-z_2|,$$

we obtain

$$|L^n(a+ib,T)-L(a+ib,T)|\leq e^{C(1-|b|)}\sup_{t\in[0,T]}|\psi^n(t,a+ib)-\psi(t,a+ib)|\int_0^T(\theta(s)+V_0\frac{s^{-H-\frac{1}{2}}}{\Gamma(\frac{1}{2}-H)})ds,$$

from (7.5.16), (7.5.17) and (7.5.19). We deduce Proposition 7.11 thanks to (7.5.15) and Theorem 7.9 together with the fact that $\int_{b\in\mathbb{R}}\frac{b^4+1}{b^2+\frac{1}{4}}e^{C(1-|b|)}dW<\infty$. □

7.6 Resolvent of the second kind

We recall some properties of the resolvent of the second kind. We consider a kernel $K\in L^1_{\text{loc}}(\mathbb{R}_+,\mathbb{R})$ and define *the resolvent of the second kind of* K as the unique function $R_K\in L^1_{\text{loc}}(\mathbb{R}_+,\mathbb{R})$ such that

$$K-R_K=K*R_K.$$

For $\lambda\in\mathbb{R}$, we define *the canonical resolvent of* K *with parameter* λ as the unique solution $E_\lambda\in L^1_{\text{loc}}(\mathbb{R}_+,\mathbb{R})$ of

$$E_\lambda-K=\lambda K*E_\lambda.$$

This means that $E_\lambda=-R_{-\lambda K}/\lambda$, when $\lambda\neq 0$ and $E_0=K$. The existence and uniqueness of R_K and E_λ is ensured by [69, Theorem 2.3.1] together with the continuity of $K\to E_\lambda(K)$ in the topology of $L^1_{\text{loc}}(\mathbb{R}_+,\mathbb{R})$. Moreover, if $K\in L^2_{\text{loc}}(\mathbb{R}_+,\mathbb{R})$ so does E_λ due to [69, Theorem 2.3.5].

We recall [69, Theorem 2.3.5] regarding the existence and uniqueness of a solution of linear Volterra integral equations in $L^1_{\text{loc}}(\mathbb{R}_+,\mathbb{R})$.

Theorem 7.18. *Let* $f\in L^1_{\text{loc}}(\mathbb{R}_+,\mathbb{R})$. *The integral equation*

$$x=f+\lambda K*x$$

admits a unique solution $x\in L^1_{\text{loc}}(\mathbb{R}_+,\mathbb{R})$ *given by*

$$x=f+\lambda E_\lambda*f.$$

When K and λ are positive, E_λ is also positive, see [69, Proposition 9.8.1]. In that case, we have a Grönwall type inequality given by [69, Lemma 9.8.2].

Lemma 7.19. *Let $x, f \in L^1_{\mathrm{loc}}(\mathbb{R}_+, \mathbb{R})$ such that*

$$x(t) \leq (\lambda K * x)(t) + f(t), \quad t \geq 0, \ a.e.$$

Then,

$$x(t) \leq f(t) + (\lambda E_\lambda * f)(t), \quad t \geq 0, \ a.e.$$

Note that the definition of the resolvent of the second kind and canonical resolvent can be extended for matrix-valued kernels. In that case, Theorem 7.18 still holds.

Remark 7.20. The canonical resolvent of the fractional kernel $K(t) = \frac{t^{H-\frac{1}{2}}}{\Gamma(H+1/2)}$ with parameter λ is given by

$$t^{\alpha-1} E_\alpha(-\lambda t^\alpha),$$

where $E_\alpha(x) = \sum_{k\geq 0} \frac{x^k}{\Gamma(\alpha(k+1))}$ is the Mittag-Leffler function and $\alpha = H + 1/2$ for $H \in (0, 1/2)$.

7.7 Additional existence results for stochastic Volterra equations

We derive additional existence results for stochastic Volterra equations extending those of the previous chapters. We fix $T > 0$ and consider the d-dimensional stochastic Volterra equation

$$X_t = g(t) + \int_0^t K(t-s) b(X_s) ds + \int_0^t K(t-s) \sigma(X_s) dW_s, \quad t \in [0, T], \tag{7.7.1}$$

where $b : \mathbb{R}^d \mapsto \mathbb{R}^d$, $\sigma : \mathbb{R}^d \mapsto \mathbb{R}^{d\times m}$ are continuous functions with linear growth, $K \in L^2([0, T], \mathbb{R}^{d\times d})$ is a kernel admitting a resolvent of the first kind L, $g : [0, T] \mapsto \mathbb{R}^d$ is a continuous function and W is a m-dimensional Brownian motion on a filtered probability space $(\Omega, \mathcal{F}, \mathbb{F}, \mathbb{P})$. Recall that under the following regularity assumption:

Assumption 7.7.1. There exists $\gamma > 0$ and $C > 0$ such that for any $t, h \geq 0$ with $t + h \leq T$,

$$|g(t+h) - g(t)|^2 + \int_0^h |K(s)|^2 ds + \int_0^{T-h} |K(h+s) - K(s)|^2 ds \leq C h^{2\gamma}.$$

(7.7.1) admits a continuous weak solution $X = (X_t)_{t\leq T}$ thanks to Theorem 6.10(ii). Moreover X satisfies

$$\sup_{t\in[0,T]} \mathbb{E}[|X_t|^p] < \infty, \quad p > 0, \tag{7.7.2}$$

and admits Hölder continuous paths on $[0, T]$ of any order strictly less than γ.

In particular, for the fractional kernel, this yields the following result.

Corollary 7.21. *Fix $H \in (0, 1/2)$ and $\theta : [0, T] \mapsto \mathbb{R}$ satisfying*

$$\forall \varepsilon > 0, \quad \exists C_\varepsilon > 0; \quad \forall u \in (0, T] \quad |\theta(u)| \leq C_\varepsilon u^{-\frac{1}{2}-\varepsilon}.$$

The fractional stochastic integral equation

$$X_t = X_0 + \frac{1}{\Gamma(H+1/2)}\int_0^t (t-u)^{H-\frac{1}{2}}(\theta(u)+b(X_u))du + \frac{1}{\Gamma(H+1/2)}\int_0^t (t-u)^{H-\frac{1}{2}}\sigma(X_u)dW_u,$$

admits a weak continuous solution $X = (X_t)_{t\leq T}$ for any $X_0 \in \mathbb{R}$. Moreover X satisfies (7.7.2) and admits Hölder continuous paths on $[0,T]$ of any order strictly less than H.

Proof. It is enough to notice that the fractional stochastic integral equation is a particular case of (7.7.1) with $d = m = 1$, $K(t) = \frac{t^{H-\frac{1}{2}}}{\Gamma(H+1/2)}$ the fractional kernel, which admits a resolvent of the first kind, recall Table 5.1, and

$$g(t) = X_0 + \frac{1}{\Gamma(1/2+H)}\int_0^t (t-u)^{H-1/2}\theta(u)du.$$

As $t \mapsto t^{1/2+\varepsilon}\theta(t)$ is bounded on $[0,T]$, we may show that g is $H-\varepsilon$ Hölder continuous for any $\varepsilon > 0$. Hence, Assumption 7.7.1 is satisfied yielding the claimed result. □

We now establish the strong existence and uniqueness of (7.7.1) in the particular case of smooth kernels. This is done by extending the Yamada-Watanabe pathwise uniqueness proof in [115].

Proposition 7.22. *Fix $m = d = 1$ and assume that g is Hölder continuous, $K \in C^1([0,T],\mathbb{R})$ admitting a resolvent of the first kind and that there exists $C > 0$ and $\eta \in [1/2, 1]$ such that for any $x, y \in \mathbb{R}$,*

$$|b(x) - b(y)| \leq C|x-y|, \quad |\sigma(x) - \sigma(y)| \leq C|x-y|^\eta.$$

Then, the stochastic Volterra equation (7.7.1) admits a unique strong continuous solution.

Proof. We start by noticing that, K being smooth, it satisfies Assumption 7.7.1 yielding the existence of a weak continuous solution to (7.7.1). It is therefore enough to prove pathwise uniqueness. We proceed similarly to [115] by considering $a_0 = 1$, $a_{k-1} > a_k$ for $k \geq 1$ with $\int_{a_k}^{a_{k-1}} x^{-2\eta}dx = k$ and $\varphi_k \in C^2(\mathbb{R},\mathbb{R})$ such that $\varphi_k(x) = \varphi_k(-x)$, $\varphi_k(0) = 0$ and for $x > 0$

- $\varphi_k'(x) = 0$ for $x \leq a_k$, $\varphi_k'(x) = 1$ for $x \geq a_{k-1}$ and $\varphi_k'(x) \in [0,1]$ for $a_k < x < a_{k-1}$,
- $\varphi_k''(x) \in [0, \frac{2}{k}x^{-2\eta}]$ for $a_k < x < a_{k-1}$.

Let X^1 and X^2 be two solutions of (7.7.1) driven by the same Brownian motion W. Notice that, thanks to the smoothness of K, $X^i - g$ are semimartingales and for $i = 1, 2$

$$d(X_t^i - g(t)) = K(0)dY_t^i + (K' * dY^i)_t\, dt,$$

with $Y_t^i = \int_0^t b(X_s^i)ds + \int_0^t \sigma(X_s^i)dW_s$. Using Itô's formula, we write

$$\varphi_k(X_t^2 - X_t^1) = I_t^1 + I_t^2 + I_t^3,$$

where

$$I_t^1 = K(0)\int_0^t \varphi_k'(X_s^2 - X_s^1)d(Y_s^1 - Y_s^2),$$

$$I_t^2 = \int_0^t \varphi_k'(X_s^2 - X_s^1)(K' * d(Y^1 - Y^2))_s ds,$$

$$I_t^3 = \frac{K(0)^2}{2} \int_0^t \varphi_k''(X_s^2 - X_s^1)(\sigma(X_s^2) - \sigma(X_s^1))^2 ds.$$

Recalling that $\sup_{t\leq T} \mathbb{E}[(X_t^i)^2] < \infty$ for $i = 1, 2$ from (7.7.2), we obtain that

$$\mathbb{E}[I_t^1] \leq \mathbb{E}[K(0) \int_0^t |b(X_s^2) - b(X_s^1)|ds] \leq c \int_0^t \mathbb{E}[|X_s^2 - X_s^1|]ds,$$

and

$$\mathbb{E}[I_t^2] \leq c \int_0^t \mathbb{E}[(|K'| * |b(X^2) - b(X^1)|)_s]ds \leq c \int_0^t \mathbb{E}[|X_s^2 - X_s^1|]ds,$$

because b is Lipschitz continuous and K' is bounded on $[0, T]$. Finally by definition of φ_k and the η-Hölder continuity of σ, we have

$$\mathbb{E}[I_t^3] \leq \frac{c}{k},$$

which goes to zero when k is large. Moreover $\mathbb{E}[\varphi_k(X_t^2 - X_t^1)]$ converges to $\mathbb{E}[|X_t^2 - X_t^1|]$ when k tends to infinity, thanks to the monotone convergence theorem. Thus, we pass to the limit and obtain

$$\mathbb{E}[|X_t^2 - X_t^1|] \leq c \int_0^t \mathbb{E}[|X_s^2 - X_s^1|]ds.$$

Grönwall's lemma leads to $\mathbb{E}[|X_t^2 - X_t^1|] = 0$ yielding the claimed pathwise uniqueness. □

Under additional conditions on g and K, one can obtain the existence of nonnegative solutions to (7.7.1) in the case of $d = m = 1$. As in Theorem 6.11, the following assumption is needed.

Assumption 7.7.2. We assume that $K \in L^2([0, T], \mathbb{R})$ is nonnegative, non-increasing and continuous on $(0, T]$. We also assume that its resolvent of the first kind L is nonnegative and non-increasing in the sense that $0 \leq L([s, s + t]) \leq L([0, t])$ for all $s, t \geq 0$ with $s + t \leq T$.

Then, for a wide class of admissible input curves g belonging to the set $\mathcal{G}_K$ as defined in (6.2.5), Theorem 6.11 yields the existence of a nonnegative solution X.

Remark 7.23. Note that any locally square-integrable completely monotone kernel[6] that is not identically zero satisfies Assumption 7.7.2, see Example 5.13. In particular, this is the case for

- the fractional kernel $K(t) = \frac{t^{H-1/2}}{\Gamma(H+1/2)}$, with $H \in (0, 1/2)$.
- any weighted sum of exponentials $K(t) = \sum_{i=1}^n c_i e^{-x_i t}$ such that $c_i, x_i \geq 0$ for all $i \in \{1, \ldots, n\}$ and $c_i > 0$ for some i.

Remark 7.24. Theorem 6.11 will be used with functions g of the following form

$$g(t) = c + \int_0^t K(t - s)\xi(ds),$$

where ξ is a nonnegative measure of locally bounded variation and c is a nonnegative constant. In that case, we may show that g belongs to $\mathcal{G}_K$, under Assumption 7.7.2.

[6] Recall that a kernel $K \in L^2_{\text{loc}}(\mathbb{R}_+, \mathbb{R})$ is said to be completely monotone, if it is infinitely differentiable on $(0, \infty)$ such that $(-1)^j K^{(j)}(t) \geq 0$ for any $t > 0$ and $j \geq 0$.

7.8 Linear Volterra equation with continuous coefficients

In this section, we consider $K \in L^2_{\mathrm{loc}}(\mathbb{R}_+, \mathbb{R})$ satisfying Assumption 7.7.2 with $T = \infty$ and recall the definition of $\mathcal{G}_K$, that is

$$\mathcal{G}_K = \{g : \mathbb{R}_+ \mapsto \mathbb{R} \text{ continuous satisfying (8.6.4) and } g(0) \geq 0\}.$$

We denote by $\|.\|_{\infty,T}$ the usual uniform norm on $[0,T]$, for each $T > 0$.

Theorem 7.25. *Let $K \in L^2_{\mathrm{loc}}(\mathbb{R}_+, \mathbb{R})$ satisfying Assumption 7.7.2 and $g, z, w : \mathbb{R}_+ \mapsto \mathbb{R}$ be continuous functions. The linear Volterra equation*

$$\chi = g + K * (z\chi + w) \tag{7.8.1}$$

admits a unique continuous solution χ. Furthermore if $g \in \mathcal{G}_K$ and w is nonnegative, then χ is nonnegative and

$$\Delta_{t_0}\chi = g_{t_0} + K * (\Delta_{t_0} z \Delta_{t_0}\chi + \Delta_{t_0} w)$$

with $g_{t_0}(t) = \Delta_{t_0} g(t) + (\Delta_t K * (z\chi + w))(t_0) \in \mathcal{G}_K$, for all for $t_0, t \geq 0$.

Proof. The existence and uniqueness of such solution in $\chi \in L^1_{\mathrm{loc}}(\mathbb{R}_+, \mathbb{R})$ is obtained from Lemma 5.35. Because χ is solution of (7.8.1), it is enough to show the local boundedness of χ to get its continuity. This follows from Grönwall's Lemma 7.19 applied on the following inequality

$$|\chi(t)| \leq \|g\|_{\infty,T} + \left(K * \left(\|z\|_{\infty,T}|\chi|(\cdot) + \|w\|_{\infty,T}\right)\right)(t),$$

for any $t \in [0,T]$ and for a fixed $T > 0$.

We assume now that $g \in \mathcal{G}_K$ and w is nonnegative. The fact that $g_{t_0} \in \mathcal{G}_K$, for $t_0 \geq 0$, is proved by adapting the computations of the proof of Theorem 6.4 with $\nu = 0$ provided that χ is nonnegative. In order to establish the non-negativity of χ, we introduce, for each $\varepsilon > 0$, χ_ε as the unique continuous solution of

$$\chi_\varepsilon = g + K * (z\chi_\varepsilon + w + \varepsilon). \tag{7.8.2}$$

It is enough to prove that χ_ε is nonnegative, for every $\varepsilon > 0$, and that $(\chi_\varepsilon)_{\varepsilon>0}$ converges uniformly on every compact to χ as ε goes to zero.

Positivity of χ_ε : It is easy to see that χ_ε is nonnegative on a neighborhood of zero because, for small t,

$$\chi_\varepsilon(t) = g(t) + (z(0)g(0) + w(0) + \varepsilon)\int_0^t K(s)ds + o\left(\int_0^t K(s)ds\right),$$

as χ, z and w are continuous functions. Hence, $t_0 = \inf\{t > 0; \quad \chi_\varepsilon(t) < 0\}$ is positive. If we assume that $t_0 < \infty$, we get $\chi_\varepsilon(t_0) = 0$ by continuity of χ_ε. χ_ε being the solution of (7.8.2), we have

$$\Delta_{t_0}\chi_\varepsilon = g_{t_0,\varepsilon} + K * (\Delta_{t_0} z \Delta_{t_0}\chi_\varepsilon + \Delta_{t_0} w + \varepsilon),$$

with $g_{t_0,\varepsilon}(t) = \Delta_{t_0} g(t) + (\Delta_t K * (z\chi_\varepsilon + w + \varepsilon))(t_0)$. Then, by using Lemma 5.6 with $F = \Delta_t K$, we obtain

$$\begin{aligned} g_{t_0,\varepsilon}(t) = {} & \Delta_{t_0} g(t) - (d(\Delta_t K * L) * g)(t_0) - (\Delta_t K * L)(0)g(t_0) \\ & + (d(\Delta_t K * L) * \chi_\varepsilon)(t_0) + (\Delta_t K * L)(0)\chi_\varepsilon(t_0), \end{aligned}$$

which is continuous and nonnegative, because $g \in \mathcal{G}_K$ and $\Delta_t K * L$ is non-decreasing for any $t \geq 0$, see Remark 5.7. Hence, in the same way, $\Delta_{t_0}\chi_\varepsilon$ is nonnegative on a neighborhood of zero. Thus $t_0 = \infty$, which means that χ_ε is nonnegative.

Uniform convergence of χ_ε : We use the following inequality

$$|\chi - \chi_\varepsilon|(t) \leq \left(K * \left(\|z\|_{\infty,T}|\chi - \chi_\varepsilon| + \varepsilon\right)\right)(t), \quad t \in [0,T],$$

together with the Gronwall Lemma 7.19 to show the uniform convergence on $[0,T]$ of χ_ε to χ as ε goes to zero. In particular, χ is also nonnegative. □

Corollary 7.26. *Let $K \in L^2_{\mathrm{loc}}(\mathbb{R}_+,\mathbb{R})$ satisfying Assumption 7.7.2 and define E_λ as the canonical resolvent of K with parameter $\lambda \in \mathbb{R} - \{0\}$. Then, $t \mapsto \int_0^t E_\lambda(s)ds$ is nonnegative and non-decreasing on $\mathbb{R}_+$. Furthermore $\int_0^t E_\lambda(s)ds$ is positive, if K does not vanish on $[0,t]$*

Proof. The non-negativity of $\chi = \int_0^\cdot E_\lambda(s)ds$ is obtained from Theorem 7.25 and from the fact that χ is solution of the following linear Volterra equation

$$\chi = K * (\lambda\chi + 1),$$

by Theorem 7.18. For fixed $t_0 > 0$, $\Delta_{t_0}\chi$ satisfies

$$\Delta_{t_0}\chi = g_{t_0} + K * (\lambda\Delta_{t_0}\chi + 1),$$

with $g_{t_0}(t) = \left(\Delta_t K * (\lambda\Delta_{t_0}\chi + 1)\right)(t_0) \in \mathcal{G}_K$, see Theorem 7.25. It follows that $\Delta_{t_0}\chi - \chi$ solves

$$x = g_{t_0} + K * (\lambda x).$$

Hence, another application of Theorem 7.25 yields that $\chi \leq \Delta_{t_0}\chi$, proving that $t \to \int_0^t E_\lambda(s)ds$ is non-decreasing. □

We now provide a version of Theorem 7.25 for complex valued solutions.

Theorem 7.27. *Let $z, w : \mathbb{R}_+ \mapsto \mathbb{C}$ be continuous functions and $h_0 \in \mathbb{C}$. The following linear Volterra equation*

$$h = h_0 + K * (zh + w)$$

admits unique continuous solution $h : \mathbb{R}_+ \mapsto \mathbb{C}$ such that

$$|h(t)| \leq \psi(t), \quad t \geq 0,$$

where $\psi : \mathbb{R}_+ \mapsto \mathbb{R}$ is the unique continuous solution of

$$\psi = |h_0| + K * (\Re(z)\psi + |w|).$$

Proof. The existence and uniqueness of a continuous solution is obtained in the same way as in the proof of Theorem 7.25. Consider now, for each $\varepsilon > 0$, ψ_ε the unique continuous solution of

$$\psi_\varepsilon = |h_0| + K * (\Re(z)\psi + |w| + \varepsilon).$$

As done in the proof of Theorem 7.25, ψ_ε converges uniformly on every compact to ψ as ε goes to zero. Thus, it is enough to show that, for every $\varepsilon > 0$ and $t \geq 0$,

$$|h(t)| \leq \psi_\varepsilon(t).$$

We start by showing the inequality in a neighborhood of zero. Because z, h, w and ψ_ε are continuous, we get, taking $h_0 = 0$,

$$|h(t)| = |w(0)| \int_0^t K(s)ds + o(\int_0^t K(s)ds), \quad \psi_\varepsilon(t) = (|w(0)| + \varepsilon) \int_0^t K(s)ds + o(\int_0^t K(s)ds),$$

for small t. Hence, $|h| \leq \psi_\varepsilon$ on a neighborhood of zero. This result still holds when h_0 is not zero. Indeed in that case, it is easy to show that for t going to zero,

$$|h(t)|^2 = |h_0|^2 + 2\Re(\overline{h_0}(z(0)h_0 + w(0))) \int_0^t K(s)ds + o(\int_0^t K(s)ds),$$

and

$$|\psi_\varepsilon(t)|^2 = |h_0|^2 + 2(\Re(z(0))|h_0|^2 + |w(0)||h_0| + \varepsilon|h_0|)) \int_0^t K(s)ds + o(\int_0^t K(s)ds).$$

As $|h_0|$ is now positive, we conclude that $|h| \leq \psi_\varepsilon$ on a neighborhood of zero by the Cauchy-Schwarz inequality.

Hence, $t_0 = \inf\{t > 0; \quad \psi_\varepsilon(t) < |h(t)|\}$ is positive. If we assume that $t_0 < \infty$, we would get that $|h(t_0)| = \psi_\varepsilon(t_0)$ by continuity of h and ψ_ε. Moreover,

$$\Delta_{t_0} h = \phi_h + K * (\Delta_{t_0} z \Delta_{t_0} h + \Delta_{t_0} w),$$

and

$$\Delta_{t_0} \psi_\varepsilon = \phi_{\psi_\varepsilon} + K * (\Delta_{t_0} \Re(z) \Delta_{t_0} w + \Delta_{t_0} |w| + \varepsilon).$$

An application of Lemma 5.6 with $F = \Delta_t K$ for $t > 0$, yields

$$\phi_h(t) = h_0(1 - (\Delta_t K * L)(t_0)) + (d(\Delta_t K * L) * h)(t_0) + (\Delta_t K * L)(0)h(t_0),$$

and

$$\phi_{\psi_\varepsilon}(t) = |h_0|(1 - (\Delta_t K * L)(t_0)) + (d(\Delta_t K * L) * \psi_\varepsilon)(t_0) + (\Delta_t K * L)(0)|h(t_0)|.$$

Relying on the fact that $d(\Delta_t K * L)$ is a nonnegative measure and $\Delta_t K * L \leq 1$, by Remark 5.7, together with the fact that $|h(s)| \leq \psi_\varepsilon(s)$ for $s \leq t_0$, we get that $|\phi_h(t)| \leq \phi_{\psi_\varepsilon}(t)$. We now notice that in the case $h(t_0) = 0$, we have

$$\Delta_{t_0} h(t) = \phi_h(t) + w(t_0) \int_0^t K(s)ds + o(\int_0^t K(s)ds),$$

and

$$\Delta_{t_0} \psi_\varepsilon(t) = \phi_{\psi_\varepsilon}(t) + (|w(t_0)| + \varepsilon) \int_0^t K(s)ds + o(\int_0^t K(s)ds),$$

and in the case $|h(t_0)| > 0$, we have

$$\begin{aligned}|\Delta_{t_0} h(t)|^2 &= 2\big(\Re(z(t_0))|h(t_0)|^2 + \Re(w(t_0))\Re(h(t_0)) + \Im(w(t_0))\Im(h(t_0))\big)\int_0^t K(s)ds \\ &\quad + |\phi_h(t)|^2 + o(\int_0^t K(s)ds), \\ \Delta_{t_0}\psi_\varepsilon(t)^2 &= 2\big(\Re(z(t_0))|h(t_0)|^2 + |w(t_0)||h(t_0)| + \varepsilon|h(t_0)|\big)\big)\int_0^t K(s)ds \\ &\quad + \phi_{\psi_\varepsilon}(t)^2 + o(\int_0^t K(s)ds),\end{aligned}$$

for small t, thanks to the continuity of $z, w, h, \phi_h, \phi_{\psi_\varepsilon}$ and ψ_ε. In both cases, we obtain that $|h| \le \psi_\varepsilon$ on a neighborhood of t_0. Therefore $t_0 = \infty$ and for any $t \ge 0$

$$|h(t)| \le \psi_\varepsilon(t).$$

□

The following result is a direct consequence of Theorems 7.25 and 7.27.

Corollary 7.28. *Let $h_0 \in \mathbb{C}$ and $z, w : \mathbb{R}_+ \to \mathbb{C}$ be continuous functions such that $\Re(z) \le \lambda$ for some $\lambda \in \mathbb{R}$. We define $h : \mathbb{R}_+ \to \mathbb{C}$ as the unique continuous solution of*

$$h = h_0 + K * (zh + w).$$

Then, for any $t \in [0, T]$,

$$|h(t)| \le |h_0| + (\|w\|_{\infty,T} + \lambda|h_0|)\int_0^T E_\lambda(s)ds,$$

where E_λ is the canonical resolvent of K with parameter λ.

Proof. From Theorem 7.27, we obtain that $|h| \le \psi_1$, where ψ_1 is the unique continuous solution of

$$\psi_1 = |h_0| + K * (\Re(z)\psi_1 + |w|).$$

Moreover define ψ_2 as the unique continuous solution of

$$\psi_2 = |h_0| + K * (\lambda\psi_2 + \|w\|_{\infty,T}).$$

Then, $\psi_2 - \psi_1$ solves

$$\chi = K * (\lambda\chi + f),$$

with $f = (\lambda - \Re(z))\psi_1 + \|w\|_{\infty,T} - w$, which is a nonnegative function on $[0, T]$. Theorem 7.25 now yields

$$|h| \le \psi_1 \le \psi_2.$$

Finally, the claimed bound follows by noticing that, for $t \in [0, T]$,

$$\psi_2(t) = |h_0| + (\|w\|_{\infty,T} + \lambda|h_0|)\int_0^t E_\lambda(s)ds,$$

by Theorem 7.18 and that $\int_0^{\cdot} E_\lambda(s)ds$ is non-decreasing by Corollary 7.26. □

Chapter 8

Lifting the Heston model

Summary

How to reconcile the classical Heston model with its rough counterpart? We introduce a lifted version of the Heston model with n multi-factors, sharing the same Brownian motion but mean reverting at different speeds. Our model nests as extreme cases the classical Heston model (when $n = 1$), and the rough Heston model (when n goes to infinity). We show that the lifted model enjoys the best of both worlds: Markovianity and satisfactory fits of implied volatility smiles for short maturities with very few parameters. Further, our approach speeds up the calibration time and opens the door to time-efficient simulation schemes.

Based on [2]: Abi Jaber, E. (2018) Lifting the Heston model. *In revision - Quantitative Finance.*

8.1 Introduction

Conventional one-dimensional continuous stochastic volatility models, including the renowned Heston model [72]:

$$dS_t = S_t\sqrt{V_t}dB_t, \quad S_0 > 0, \tag{8.1.1}$$

$$dV_t = \lambda(\theta - V_t)dt + \nu\sqrt{V_t}dW_t, \quad V_0 \geq 0, \tag{8.1.2}$$

have struggled in capturing the risk of large price movements on a short timescale. In the pricing world, this translates into failure to reproduce the at-the-money skew observed in the market as illustrated on the following figure.

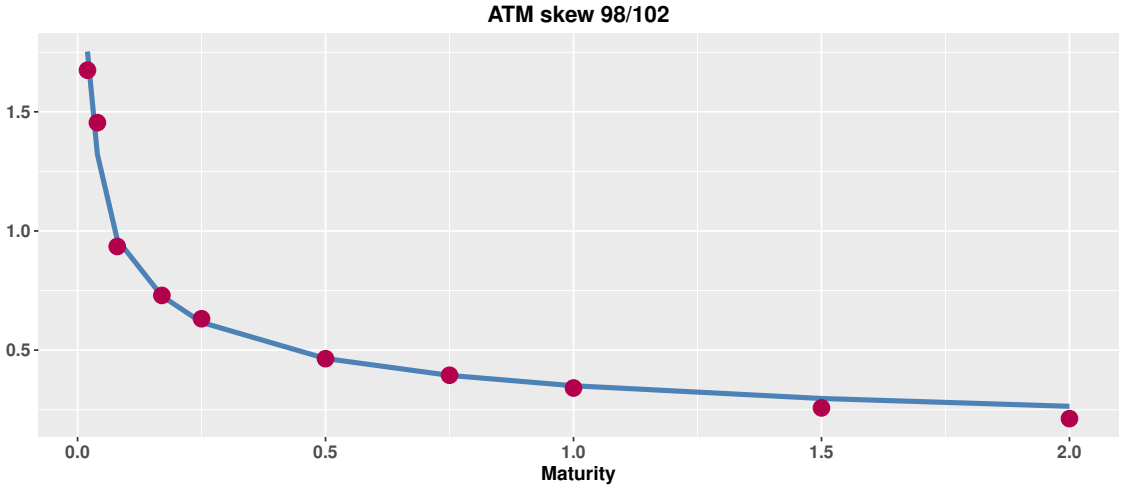

FIGURE 8.1: Term structure of the at-the-money skew for the S&P index on June 20, 2018 (red dots) and a power-law fit $t \to 0.35 \times t^{-0.41}$.

In view of improving the overall fit, several directions have been considered over the past decades. Two of the most common extensions are adding jumps [32, 65] and stacking additional random factors [18, 61], in order to jointly account for short and long timescales. While the two approaches have structural differences, they both suffer from the curse of dimensionality, as more parameters are introduced, slowing down the calibration process. Recently, rough volatility models have been introduced as a fresh substitute with remarkable fits of the implied volatility surface, see [66, 14, 50]. The rough variance process involves a one-dimensional Brownian motion, keeps the number of parameters small and enjoys continuous paths. However, the price to pay is that rough volatility models leave the realm of semimartingale and Markovian models, which makes pricing and hedging a challenging task, while degrading the calibration time. Here, the curse of dimensionality hits us straight in the face in the non-Markovianity of the process. Indeed, the rough model can be seen as an infinite dimensional Markovian model, as shown in Chapter 6.

Going back to the standard Heston model (8.1.1)-(8.1.2), despite its lack of fit for short maturities, it remains increasingly popular among practitioners. This is due to its high tractability, by virtue of the closed form solution of the characteristic function, allowing for fast pricing and calibration by Fourier inversion techniques [27, 55]. Recently, El Euch and Rosenbaum [51] combined the tractability of the Heston model with the flexibility of rough volatility models, to elegantly concoct a rough counterpart of (8.1.1)-(8.1.2), dubbed the rough Heston model. More precisely, the rough model is constructed by replacing the variance process (8.1.2) by a fractional square-root process as follows

$$dS_t = S_t\sqrt{V_t}dB_t, \quad S_0 > 0, \tag{8.1.3}$$

$$V_t = V_0 + \frac{1}{\Gamma(H+1/2)}\int_0^t (t-s)^{H-1/2}\left(\lambda(\theta - V_s)ds + \nu\sqrt{V_s}dW_s\right), \tag{8.1.4}$$

where $H \in (0, 1/2]$ has a physical interpretation, as it measures the regularity of the sample paths of V, see [66, 15], the case $H = 1/2$ corresponding to the standard Heston model. More precisely, the sample paths of V are locally Hölder continuous of any order strictly less than H. As for the standard Heston model, the characteristic function of the log-price is known, but only up to the solution of a certain fractional Riccati Volterra equation. Indeed, both

models belong to the tractable and unifying class of affine Volterra processes introduced in Chapter 5. The following table summarizes the characteristics of the two models.

Characteristics	Heston	Rough Heston
Markovian	✓	✗
Semimartingale	✓	✗
Simulation	Fast	Slow
Affine Volterra process	✓	✓
Characteristic function	Closed	Fractional Riccati
Calibration	Fast	Slower
Fit short maturities	✗	✓
Regularity of sample paths	$H = 0.5$	$0 < H \leq 0.5$

TABLE 8.1: Summary of the characteristics of the models.

In the present paper, we introduce a conventional multi-factor continuous stochastic volatility model: *the lifted Heston model*. The variance process is constructed as a weighted sum of n factors, driven by the same one-dimensional Brownian motion, but mean reverting at different speeds, in order to accommodate a full spectrum of timescales. At first glance, the model seems over-parametrized, with already $2n$ parameters for the mean reversions and the weights. Inspired by the approximation results of Chapter 7, we provide a good parametrization of these $2n$ parameters in terms of one single parameter H, which is nothing else but the Hurst index of a limiting rough Heston model (8.1.3)-(8.1.4), obtained after sending the numbers of factors to infinity.

The lifted model not only nests as extreme cases the classical Heston model (when $n = 1$) and the rough Heston model (when n goes to infinity), but also enjoys the best of both worlds: the flexibility of rough volatility models, and the Markovianity of their conventional counterparts. Further, the model remains tractable, as it also belongs to the class of affine Volterra processes. Here, the characteristic function of the log-price is known up to a solution of a finite system of Riccati ordinary differential equations. From a practical viewpoint, we demonstrate that the *lifted Heston model*:

- reproduces the same volatility surface as the rough Heston model for maturities ranging from one week to two years,
- mimics the explosion of the at-the-money skew for short maturities,
- calibrates twenty times faster than its rough counterpart,
- is easier to simulate than the rough model.

All in all, the *lifted Heston model* can be more easily implemented than its rough counterpart, while still retaining the precision of implied volatility fits of the rough Heston model. Further, the *lifted Heston model* is able to generate a volatility surface, which cannot be generated by the classical Heston model, with only one additional parameter. Finally, the stock price and

the variance process enjoy continuous paths and only depend on a two-dimensional Brownian motion, leading to simple and feasible hedging strategies.

The paper is outlined as follows. In Section 8.2 we introduce our *lifted Heston model* and provide its existence, uniqueness and its affine Fourier-Laplace transform. Exploiting the limiting rough model, we proceed in Section 8.3 to a reduction of the number of parameters to calibrate. Numerical experiments for the model, with $n = 20$ factors, are illustrated in Section 8.4, both for calibration and simulation. Finally, some technical material is postponed to Section 8.6.

8.2 The lifted Heston model

We fix $n \in \mathbb{N}$ and we define the *lifted Heston model* as a conventional stochastic volatility model, with n factors for the variance process:

$$dS_t^n = S_t^n \sqrt{V_t^n} dB_t, \quad S_0^n > 0, \tag{8.2.1}$$

$$V_t^n = g_0^n(t) + \sum_{i=1}^n c_i^n U_t^{n,i}, \tag{8.2.2}$$

$$dU_t^{n,i} = \left(-x_i^n U_t^{n,i} - \lambda V_t^n\right) dt + \nu \sqrt{V_t^n} dW_t, \quad U_0^{n,i} = 0, \quad i = 1, \dots, n, \tag{8.2.3}$$

with parameters the function g_0^n, $\lambda, \nu \in \mathbb{R}_+$, $c_i^n, x_i^n \geq 0$, for $i = 1, \dots, n$, and $B = \rho W + \sqrt{1-\rho^2} W^\perp$, with $(W, W^\perp)$ a two dimensional Brownian motion on a fixed filtered probability space $(\Omega, \mathcal{F}, \mathbb{F} := (\mathcal{F}_t)_{t \geq 0}, \mathbb{Q})$, with $\rho \in [-1, 1]$.

We stress that all the factors $(U^{n,i})_{1 \leq i \leq n}$ start from zero[1] and share the same dynamics, with the same one-dimensional Brownian motion W, except that they mean revert at different speeds $(x_i^n)_{1 \leq i \leq n}$. Further, the deterministic input curve g_0^n allows one to plug-in initial term-structure curves. More precisely, taking the expectation in (8.2.2) leads to the following relation

$$\mathbb{E}[V_t^n] + \lambda \sum_{i=1}^n c_i^n \int_0^t e^{-x_i^n (t-s)} \mathbb{E}[V_s^n] ds = g_0^n(t), \quad t \geq 0.$$

In practice, the forward variance curve, up to a horizon $T > 0$, can be extracted from variance swaps observed in the market and then plugged-in in place of $(\mathbb{E}[V_t^n])_{t \leq T}$ in the previous expression. For a suitable choice of continuous curves g_0^n, for instance if

$$g_0^n \text{ is non-decreasing such that } g_0^n(0) \geq 0, \tag{8.2.4}$$

or

$$g_0^n : t \to V_0 + \sum_{i=1}^n c_i^n \int_0^t e^{-x_i^n (t-s)} \theta(s) ds, \text{ with } V_0, \theta \geq 0, \tag{8.2.5}$$

there exists a unique continuous $\mathbb{F}$-adapted strong solution $(S^n, V^n, (U^{n,i})_{1 \leq i \leq n})$ to (8.2.1)-(8.2.3), such that $V_t^n \geq 0$, for all $t \geq 0$, and S^n is a $\mathbb{F}$-martingale. We refer to Section 8.6.1 below for more details and the exact definition of the set of admissible input curves g_0^n.

[1] Notice that the initial value of the variance process V^n is $g_0^n(0)$.

Since our main objective is to compare the lifted model to other existent models, we will restrict to the case of input curves of the form

$$g_0^n : t \to V_0 + \theta \sum_{i=1}^n c_i^n \int_0^t e^{-x_i^n (t-s)} ds, \quad \text{with } V_0, \theta \geq 0. \tag{8.2.6}$$

Setting $n = 1$, $c_1^1 = 1$ and $x_1^1 = 0$, the *lifted Heston model* degenerates into the standard Heston model (8.1.1)-(8.1.2). So far, the multi-factor extensions of the standard Heston model have been considered by stacking additional square-root processes as in the double Heston model[2] of [31] and the multi-scale Heston model of [60], or by considering a Wishart matrix-valued process as in [37]. In both cases, the dimension of the driving Brownian motion for the variance process, along with the number of parameters, grows with the number of factors. Clearly, the *lifted Heston model* differs from these extensions, one can compare (8.2.1)-(8.2.3) for $n = 2$ with (8.2.7)-(8.2.8).

Just like the classical Heston model, the *lifted Heston model* remains tractable. Specifically, fix $u \in \mathbb{C}$ such that $\mathrm{Re}(u) \in [0, 1]$. By virtue of Section 8.6.2 below, the Fourier-Laplace transform of the log-price is exponentially affine with respect to the factors $(U^{n,i})_{1 \leq i \leq n}$:

$$\mathbb{E}\left[\exp\left(u \log S_t^n\right) \,\middle|\, \mathcal{F}_t\right] = \exp\left(\phi^n(t,T) + u \log S_t^n + \sum_{i=1}^n c_i^n \psi^{n,i}(T-t) U_t^{n,i}\right), \tag{8.2.9}$$

for all $t \leq T$, where $(\psi^{n,i})_{1 \leq i \leq n}$ solves the following n-dimensional system of Riccati ordinary differential equations

$$(\psi^{n,i})' = -x_i^n \psi^{n,i} + F\left(u, \sum_{j=1}^n c_j^n \psi^{n,j}\right), \quad \psi^{n,i}(0) = 0, \quad i = 1, \ldots, n, \tag{8.2.10}$$

with

$$F(u,v) = \frac{1}{2}(u^2 - u) + (\rho\nu u - \lambda)v + \frac{\nu^2}{2} v^2,$$

and

$$\phi^n(t,T) = \int_0^{T-t} F\left(u, \sum_{i=1}^n c_i^n \psi^{n,i}(s)\right) g_0^n(T-s) ds, \quad t \leq T.$$

In particular, for $t = 0$, since $U_0^{n,i} = 0$ for $i = 1, \ldots, n$, the unconditional Fourier-Laplace transform reads

$$\mathbb{E}\left[\exp\left(u \log S_t^n\right)\right] = \exp\left(u \log S_0^n + \int_0^T F\left(u, \sum_{i=1}^n c_i^n \psi^{n,i}(s)\right) g_0^n(T-s) ds\right). \tag{8.2.11}$$

[2]The double Heston model is defined in [31] as follows

$$dS_t^n = S_t^n \left(\sqrt{U_t^1} dB_t^1 + \sqrt{U_t^2} dB_t^2\right), \tag{8.2.7}$$

$$dU_t^i = \lambda_i(\theta_i - U_t^i)dt + \nu_i \sqrt{U_t^i} dW_t^i, \quad U_0^i \geq 0, \quad i \in \{1, 2\}, \tag{8.2.8}$$

where $B^i = \rho_i W^i + \sqrt{1 - \rho_i^2} W^{i,\perp}$ with $\rho_i \in [-1, 1]$ and $(W^1, W^2, W^{1,\perp}, W^{2,\perp})$ a four-dimensional Brownian motion.

A similar formula holds for the Fourier-Laplace transform of the joint process $(\log S^n, V^n)$ with integrated log-price and variance, we refer to the Section 8.6.2 below for the precise expression.

Consequently, the Fourier-Laplace transform of the *lifted Heston model* is known in closed-form, up to the solution of a deterministic n-dimensional system of ordinary differential equations (8.2.10), which can be solved numerically. Once there, standard Fourier inversion techniques can be applied on (8.2.11) to deduce option prices. This is illustrated in the following sections.

8.3 Parameter reduction and the choice of the number of factors

In this section, we proceed to a reduction of the number of parameters to calibrate. Our inspiration stems from rough volatility. In a first step, for every n, we provide a parametrization of the weights and the mean reversions $(c_i^n, x_i^n)_{1\leq i\leq n}$ in terms of the Hurst index H of a limiting rough volatility model and one additional parameter r_n. Then, we specify the number of factors n and the value of the additional parameter r_n so that the lifted model reproduces the same volatility surface as the rough Heston model for maturities ranging from one week up to two years, while calibrating twenty times faster than its rough counterpart. Benchmarking against rough volatility models is justified by the fact that one of the main strengths of these models is their ability to achieve better fits of the implied volatility surface than conventional stochastic volatility models. This has been illustrated on real market data in [14, 50]. Finally, for the sake of completeness, we provide a comparison with the standard Heston model.

8.3.1 Parametrization in terms of the Hurst index

For an initial input curve of the form (8.2.6), the *lifted Heston model* (8.2.1)-(8.2.3) has the same five parameters $(V_0, \theta, \lambda, \nu, \rho)$ of the Heston model, plus $2n$ additional parameters for the weights and the mean reversions $(c_i^n, x_i^n)_{1\leq i\leq n}$.[3] At first sight, the model seems to suffer from the curse of dimensionality, as it requires the calibration of $(2n+5)$ parameters. This is where the exciting theory of rough volatility finally comes into play. Inspired by the approximation result Theorem 7.7, we suggest to use a parametrization of $(c_i^n, x_i^n)_{1\leq i\leq n}$ in terms of two well-chosen parameter. By doing so, we reduce the $2n$ additional parameters to calibrate to only two effective parameters.

Qualitatively, we choose the weights and mean reversions $(c_i^n, x_i^n)_{1\leq i\leq n}$ in such a way that sending the number of factors $n \to \infty$ would yield the convergence of the *lifted Heston model* towards a rough Heston model (8.1.3)-(8.1.4), with parameters $(V_0, \theta, \lambda, \nu, \rho, H)$. The additional parameter $H \in (0, 1/2)$ is the so-called Hurst index of the limiting fractional variance process (8.1.4), and it measures the regularity of its sample paths.

[3] If one chooses g_0^n to match the forward variance curve, then, the parameters (V_0, θ) can be eliminated from both models.

More precisely, for a fixed even number of factors n, we fix $r_n > 1$ and we consider the following parametrization for the weights and the mean reversions

$$c_i^n = \frac{(r_n^{1-\alpha} - 1) r_n^{(\alpha-1)(1+n/2)}}{\Gamma(\alpha)\Gamma(2-\alpha)} r_n^{(1-\alpha)i}, \quad x_i^n = \frac{1-\alpha}{2-\alpha} \frac{r_n^{2-\alpha} - 1}{r_n^{1-\alpha} - 1} r_n^{i-1-n/2}, \quad i = 1, \dots, n, \tag{8.3.1}$$

where $\alpha := H + 1/2$ for some $H \in (0, 1/2)$.[4]

If in addition, the sequence $(r_n)_{n\geq 1}$ satisfies

$$r_n \downarrow 1 \quad \text{and} \quad n \ln r_n \to \infty, \quad \text{as } n \to \infty, \tag{8.3.2}$$

then, Theorem 8.4 below ensures the convergence of the lifted model towards the rough Heston model, as n goes to infinity. We refer to Section 8.6.1 below for more details.

In order to visualize this convergence, we first define the following sequence

$$r_n = 1 + 10\, n^{-0.9}, \quad n \geq 1, \tag{8.3.3}$$

which clearly satisfies (8.3.2). Then, we generate our benchmark implied volatility surface,

$$\text{for 9 maturities } T \in \{1\text{w}, 1\text{m}, 2\text{m}, 3\text{m}, 6\text{m}, 9\text{m}, 1\text{y}, 1.5\text{y}, 2\text{y}\}, \tag{8.3.4}$$

$$\text{with up to 80 strikes } K \text{ per maturity}, \tag{8.3.5}$$

with a rough Heston model[5] with parameters $\Theta_0 := (V_0, \theta, \lambda, \nu, \rho, H)$ given by

$$V_0 = 0.02, \quad \theta = 0.02, \quad \lambda = 0.3, \quad \nu = 0.3, \quad \rho = -0.7 \quad \text{and} \quad H = 0.1. \tag{8.3.6}$$

The generated implied volatility is kept fixed and is denoted by $\sigma_\infty(K, T; \Theta_0)$, for every pair (K, T) in (8.3.4)-(8.3.5).

Then, for each $n \in \{10, 20, 50, 100, 500\}$, we generate the implied volatility surface of the *lifted Heston model*[6] with n-factors, with the same set of parameters Θ_0 as in (8.3.6), and (8.3.3) plugged in (8.3.1). For each n, the generated surface is denoted by $\sigma_n(K, T; r_n, \Theta_0)$, for every pair (K, T) in (8.3.4)-(8.3.5).

Because the sequence $(r_n)_{n\geq 1}$ defined in (8.3.3) satisfies condition (8.3.2), as n grows,

$$\sigma_n(K, T; r_n, \Theta_0) \to \sigma_\infty(K, T; \Theta_0),$$

[4] This corresponds to equation (7.3.6) of Chapter 7 with the geometric partition $\eta_i^n = r_n^{i-n/2}$ for $i = 0, \dots, n$, which is in the spirit of [26] for the approximation of the factional Brownian motion.

[5] The implied volatility surface is generated by first solving numerically the corresponding fractional Riccati equations with the Adams Predictor-Corrector scheme [44] with 200 time steps, see [50, Appendix A] for more details. Then, call prices are computed via the cosine method [55] for the inversion of the characteristic function. We note that other Fourier inversion techniques can be used for the second step, for instance, the Carr-Madan method [27], as done in [50]. As illustrated in [55], for the same level of accuracy, the cosine method is approximately 20 times faster than the Carr-Madan method, and needs drastically less evaluation points of the characteristic function $(E[\exp(u_i \log S_t^n)])_{i\in\mathcal{I}}$ ($|\mathcal{I}| = 160$ for the cosine methods and $|\mathcal{I}| = 4096$ for the Carr-Madan method). This latter point is crucial in our case since, for every $i \in \mathcal{I}$, evaluation of $E[\exp(u_i \log S_t^n)]$ requires a numerical discretization of the corresponding Riccati equation.

[6] The implied volatility surface is generated by first solving numerically the n-dimensional Riccati equations with the explicit-implicit scheme (8.6.11) detailed in the Appendix with a number of time steps $N = 300$. As before, the call prices are then computed via the cosine method [55] for the inversion of the characteristic function.

by virtue of Theorem 8.4 below. This convergence phenomenon is illustrated in Figure 8.2 below for two maturity slices, one week and one year.

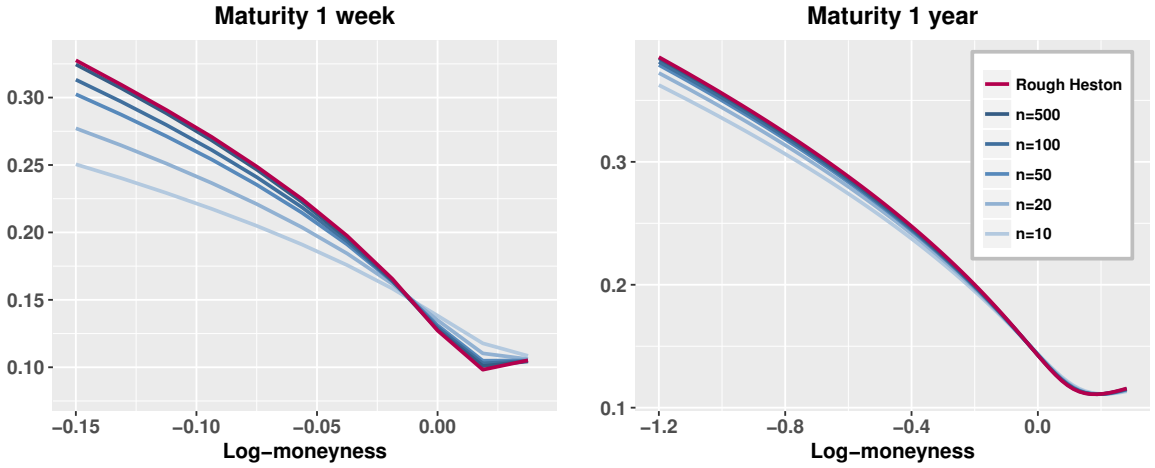

FIGURE 8.2: Convergence of the implied volatility surface of the lifted model $\sigma_n(k,T;r_n,\Theta_0)$, with $r_n = 1 + 10\,n^{-0.9}$, towards its rough counterpart $\sigma_\infty(k,T;\Theta_0)$, illustrated on two maturities slices $T \in \{1 \text{ week}, 1 \text{ year}\}$. Here $k := \ln(K/S_0)$ stands for the log-moneyness.

In view of assessing the proximity between the implied volatility surface $\sigma_n(K,T;r_n,\Theta_0)$ of the *lifted Heston model* and that of the rough Heston model $\sigma_\infty(K,T;\Theta_0)$, we compute the mean squared error (MSE) between the two volatility surfaces defined as follows

$$\frac{1}{\sum_{(K',T')} w(K',T')} \sum_{(K,T)} w(K,T)(\sigma_n(K,T;r_n,\Theta_0) - \sigma_\infty(K,T;\Theta_0))^2,$$

where we sum over all pairs (K,T) as in (8.3.4)-(8.3.5). Here, w stands for a matrix of weights, where we put more weight on options near the money and with short time to maturity (one could also set $w(K,T) = 1$ for all (K,T)).

The corresponding mean squared errors of Figure 8.2 are reported in Table 8.2 below, along with the computational time[7] for generating the whole volatility surface, for all pairs (K,T) as in (8.3.4)-(8.3.5), that is, for 9 maturities slices with up to 80 strikes per maturity.[8]

[7] All cpu times are computed on a laptop with Intel core i7 processor at 2.2GHz and 16GB of memory. The code, written in R, is far from being optimized.

[8] One cannot draw definite quantitative conclusions regarding the comparison between the computational times of the lifted surface and the one of the rough surface. Indeed, one needs a more careful study of the discretization errors of the corresponding Riccati equations before comparing the computational times needed to reach the same level of accuracy. We omit to do so here. However, even if one reduces the number of time steps from 200 to 150 in the Adams scheme, it still takes 67.2 seconds to compute the rough surface. Recall that we used $N = 300$ time steps for the n-dimensional Riccati equation of the lifted model. In any case, it should be clear that solving the 20-dimensional Riccati equations is considerably faster then solving the fractional Riccati equation.

	n	$r_n = 1 + 10\,n^{-0.9}$	Time (seconds)	MSE
Lifted Heston	10	2.26	3.9	1.20e-03
	20	1.67	4.4	1.85e-04
	50	1.3	5.2	6.81e-05
	100	1.16	6.6	2.54e-05
	500	1.04	17.4	3.66e-06
Rough Heston	$n \to \infty$	$r_n \downarrow 1$	106.8	

TABLE 8.2: Convergence of the lifted model towards its rough counterpart for $r_n = 1 + 10\,n^{-0.9}$, with the corresponding computational time in seconds for generating the implied volatility surface (8.3.4)-(8.3.5).

All in all, we notice that the number of effective parameters remains constant and does not depend on the number of factors n. This has to be contrasted with the usual multi-factor extensions: the double Heston model (8.2.7)-(8.2.8) already has 10 parameters $(U_0^i, \theta_i, \lambda_i, \nu_i, \rho_i)_{i\in\{1,2\}}$, the multi-scale model of [60] also suffers from over-parametrization.

In the subsequent subsection, we will explain how to fix n and r_n, so that the parameters to calibrate are reduced to only six effective parameters $(V_0, \theta, \lambda, \nu, \rho, H)$, one additional parameter than the standard Heston model!

8.3.2 Practical choice of n and r_n

We suggest to fix the following values

$$n = 20 \quad \text{and} \quad r_{20} = 2.5. \tag{8.3.7}$$

Our choice will be based on the numerical comparison with the rough Heston model of the previous section.

We start by explaining our choice for the number of factors n in (8.3.7). Based on Table 8.2, we choose n with a good trade-off between time-efficiency and proximity to the rough volatility surface. Fixing $n = 20$ seems to be a good choice. Visually, as already shown on Figure 8.2, the two implied volatility slices have almost identical shapes. Whence, one would expect that by letting the parameters r_{20} free, one could achieve a perfect fit of the rough surface with only $n = 20$ factors. This can be formulated as follows: keeping the six parameters of the lifted model fixed as in (8.3.6), can one find $r_{20}^*(\Theta_0) > 1$ such that

$$\sigma_{20}(K, T; r_{20}^*(\Theta_0), \Theta_0) \approx \sigma_\infty(K, T; \Theta_0), \quad \text{for all } K, T?$$

The next subsection provides a positive answer.

Mimicking roughness by increasing r_{20}

First, one needs to understand the influence of the parameter r_n on the *lifted Heston model*. Increasing r_n has the effect of boosting the parameters $(c_i^n, x_i^n)_{1\le i\le n}$ in (8.3.1), leading to an increase of the vol-of-vol parameter of the lifted model given by $\nu \sum_{i=1}^n c_i^n$, together with faster mean-reversions $(x_i^n)_{1\le i\le n}$ for the factors. In analogy with conventional stochastic

volatility models, such as the standard Heston model (8.1.1)-(8.1.2), increasing the vol-of-vol parameter together with the speed of mean reversion yields a steeper skew at the short-maturity end of the volatility surface. Consequently, increasing the parameter r_n in the lifted model should steepen the implied volatility slice for short-maturities. Figure 8.3 below confirms that this is indeed the case when one increases the value of r_{20} from 1.67 to 2.8, for the 20-dimensional lifted model, as the two slices now almost perfectly match:

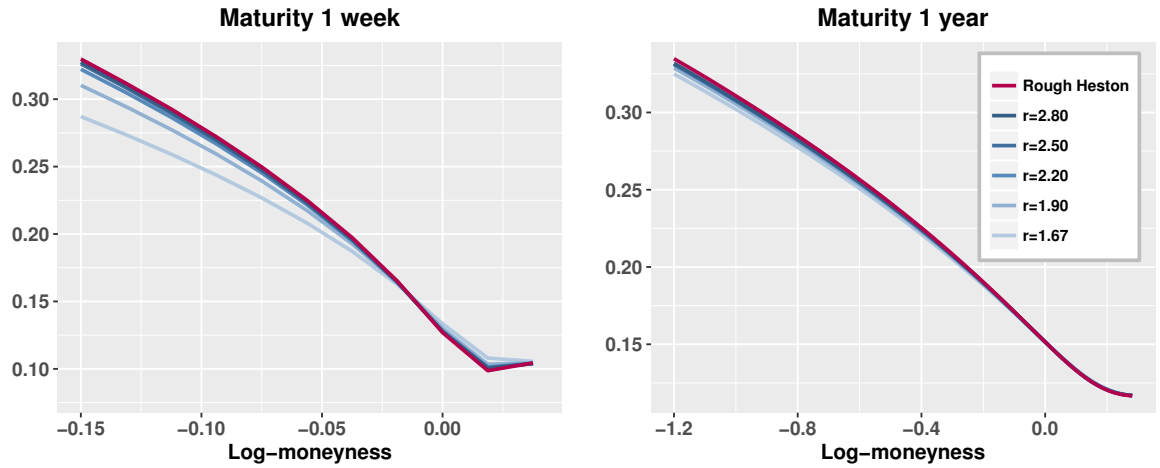

FIGURE 8.3: Implied volatility of the 20-dimensional lifted model $\sigma_{20}(k, T; r_{20}, \Theta_0)$, for different values of r_{20} ranging from 1.67 to 2.8, and the rough surface $\sigma_\infty(k, T; \Theta_0)$, for two maturities slices $T \in \{1 \text{ week}, 1 \text{ year}\}$.

The corresponding mean squared errors of Figure 8.3 are collected in Table 8.3 below.

Lifted Heston	$(n = 20)$
r_{20}	MSE
1.67	1.85e-04
1.90	4.16e-05
2.20	8.72e-06
2.50	3.64e-06
2.80	2.81e-06

TABLE 8.3: Mean squared errors between the 20-dimensional lifted model $\sigma_{20}(k, T; r_{20}, \Theta_0)$ and the rough model $\sigma_\infty(k, T; \Theta_0)$, for different values of r_{20}.

Because r_n has to converge to 1, when n goes to infinity, recall (8.3.2), we seek to keep r_n as small as possible. For $n = 20$, fixing $r_{20}^*(\Theta_0) = 2.5$ yields already satisfactory results, improving the mean squared error of 1.85e-04 in Table 8.2 to 3.64e-06. Further, this choice yields the same order of precision as with $n = 500$ factors given in Table 8.2.

Before moving to a physical justification of the choice of r_{20}, we proceed to the full calibration of the *lifted Heston model* with $n = 20$ and $r_{20} = 2.5$ to the rough volatility surface $\sigma_\infty(K, T; \Theta_0)$. That is, we let the six effective parameters $(V_0, \theta, \lambda, \nu, \rho, H)$ of the lifted model free. The calibrated values $\hat{\Theta}_0 := (\hat{V}_0, \hat{\theta}, \hat{\lambda}, \hat{\nu}, \hat{\rho}, \hat{H})$, provided in Table 8.4, agree with (8.3.6).

At the visual level, as shown on Figure 8.10 below, the calibrated lifted surface is indistinguishable from the rough surface $\sigma_\infty(K,T;\Theta_0)$ for all maturities ranging from one week to two years, with a mean squared error of order 4.01e-07.

Parameters	Calibrated values
$\hat{V}_0$	0.02012504
$\hat{\theta}$	0.02007956
$\hat{\lambda}$	0.29300681
$\hat{\nu}$	0.30527694
$\hat{\rho}$	-0.70241116
$\hat{H}$	0.09973346

TABLE 8.4: Calibrated *lifted Heston model* parameters.

We now provide another physical justification for the choice of r_{20} based on an infinite-dimensional Markovian representation of the limiting rough variance process (8.1.4) due to Chapter 6, which we recall in the following remark.

Remark 8.1 (Representation of the limiting rough process). The fractional kernel appearing in the limiting rough process (8.1.4) admits the following Laplace representation

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

so that the stochastic Fubini theorem, after setting $V_0 \equiv 0$ in (8.1.4), leads to

$$V_t = \int_0^\infty U_t(x)\mu(dx), \quad x > 0,$$

where, for all $x > 0$,

$$U_t(x) := \int_0^t e^{-x(t-s)}\left(\lambda(\theta - V_s)ds + \nu\sqrt{V_s}dW_s\right).$$

This can be seen as the mild formulation of the following stochastic partial differential equation

$$dU_t(x) = \left(-xU_t(x) + \lambda\left(\theta - \int_0^\infty U_t(y)\mu(dy)\right)\right)dt + \nu\sqrt{\int_0^\infty U_t(y)\mu(dy)}dW_t, \tag{8.3.8}$$

$$U_0(x) = 0, \quad x > 0. \tag{8.3.9}$$

Whence, the rough process can be reinterpreted as a superposition of infinitely many factors $(U_\cdot(x))_{x>0}$ sharing the same dynamics but mean reverting at different speeds $x \in (0,\infty)$. We refer to Chapter 6 for the rigorous treatment of this representation. One makes the following observations:

- multiple timescales are naturally encoded in rough volatility models, which can be a plausible explanation for their ability to achieve better fits than conventional one-dimensional models,
- the largest mean reversions going to infinity characterize the factors responsible of the roughness of the process.

We notice that for the lifted model, the mean reversions in (8.3.1) satisfy

$$x_i^n \geq r_n^{i-1-n/2}, \quad i = 1, \dots, n.$$

Therefore, based on Remark 8.1, for $n = 20$, one would like to force x_{20}^{20} to be large enough in order to mimic roughness and account for very short timescales, while having x_1^{20} small enough to accommodate a whole palette of timescales. Setting

$$r_{20} \approx 2.5,$$

would cover mean reversions between 10^{-4} and 10^4.

The previous justification suggests that once $n = 20$ is fixed, one can choose r_{20} independently of the parameters Θ. The next experiment shows that this is indeed the case.

Robustness of r_{20}: a numerical test

Throughout this section, we fix the three parameters $V_0, \theta = 0.02$ and $\lambda = 0$. In order to verify experimentally the robustness of $r_{20} = 2.5$, we proceed as follows.

1. Simulate $M = 500$ set of parameters $(\Theta_k := (0.02, 0.02, 0, \nu_k, \rho_k, H_k))_{k=1,\dots,M}$ uniformly distributed with the following bounds

$$0.05 \leq \nu \leq 0.5, \quad -0.9 \leq \rho \leq -0.5, \quad 0.05 \leq H \leq 0.2.$$

2. For each $k = 1, \dots, M$:
   (a) Generate the rough volatility surface $\sigma_\infty(K, T; \Theta_k)$, for all pairs (T, K) in (8.3.4)-(8.3.5),
   (b) Generate the lifted volatility surface $\sigma_{20}(K, T; r_{20} = 2.5, \Theta_k)$, for all pairs (T, K) in (8.3.4)-(8.3.5),
   (c) Compute the mean squared error between the two volatility surfaces:

$$\mathrm{MSE}_k := \frac{1}{\sum_{(K',T')} w(K', T')} \sum_{(K,T)} w(K, T)(\sigma_{20}(K, T; r_{20} = 2.5, \Theta_k) - \sigma_\infty(K, T; \Theta_k))^2.$$

The scatter plot and the empirical distribution of the mean squared error $(\mathrm{MSE}_k)_{k=1,\dots,M}$ are illustrated in Figure 8.4 below.

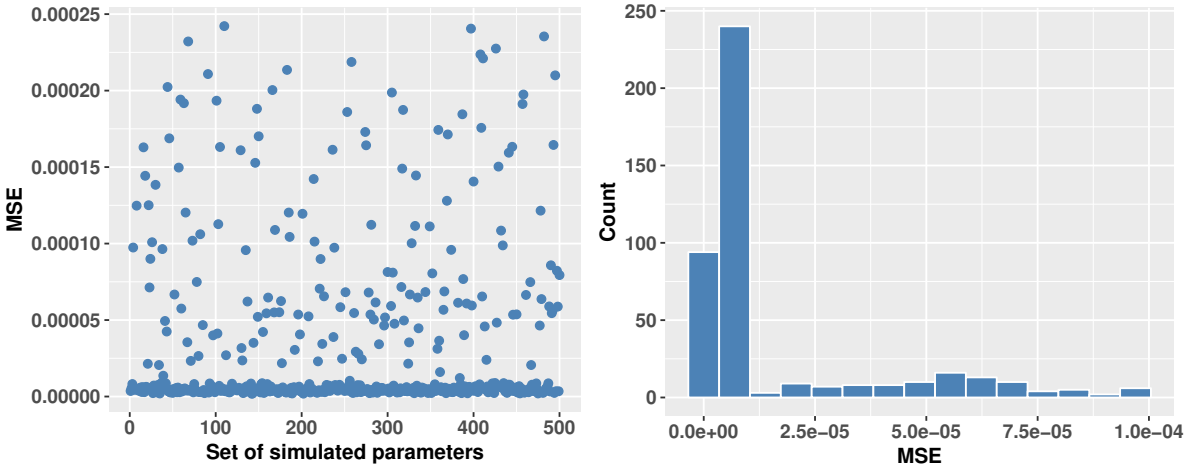

FIGURE 8.4: Scatter plot (left) and empirical distribution (right) of the mean squared error $(\text{MSE}_k)_{k=1,\dots,M}$ of the $M = 500$ simulated set of parameters $(\Theta_k)_{k=1,\dots,M}$.

The first twenty values of the simulated set of parameters with the corresponding mean squared error are provided in Table 8.8 below. We observe that the lifted surfaces are quite close to the rough surface, for any value of the simulated parameters. This is confirmed by Table 8.5 below, where we collect the descriptive statistics of the computed mean squared errors $(\text{MSE}_k)_{k=1,\dots,M}$.

	MSE
Minimum	1.81e-06
1st Quantile	3.83e-06
Median	5.48e-06
3rd Quantile	4.91e-05
Maximum	2.42e-04

TABLE 8.5: Descriptive statistics of the mean squared error $(\text{MSE}_k)_{k=1,\dots,M}$ of the $M = 500$ simulated set of parameters $(\Theta_k)_{k=1,\dots,M}$.

We now show that the mean squared errors can be improved by letting the three parameters (ν, ρ, H) of the lifted model free. Specifically, consider the worst mean squared error of Table 8.5

$$\max_{\Theta_k} \text{MSE}_k = 2.42\text{e-}04, \tag{8.3.10}$$

which is attained for the set of parameters Θ_{101} with

$$\nu_{101} = 0.1537099, \quad \rho_{101} = -0.8112745 \text{ and } H_{101} = 0.1892725.$$

Keeping the first three parameters fixed $V_0, \theta = 0.02$ and $\lambda = 0$, we proceed to the calibration of the lifted model to the rough surface $\sigma_\infty(K, T; \Theta_{101})$. The calibration yields

$$\hat{\nu} = 0.1647801, \quad \hat{\rho} = -0.7961080 \text{ and } \hat{H} = 0.1957235,$$

improving the previous mean squared error (8.3.10) to 1.62e-06. This shows that, by fine tuning the parameters of the lifted model, for any rough volatility surface $\sigma_\infty(K,T;\Theta)$ with a realistic set of parameters Θ, one can find a set of parameters $\hat{\Theta}$, not too far from Θ, such that

$$\sigma_{20}(K,T;r_{20}=2.5,\hat{\Theta}) \approx \sigma_\infty(K,T;\Theta), \quad \text{for any pair } (K,T) \text{ in (8.3.4)-(8.3.5)}.$$

To sum up, we showed so far that the *lifted Heston model*, with $n = 20$ and $r_{20} = 2.5$, is able to produce the same volatility surfaces of the rough Heston model, for any realistic set of parameters, for maturities ranging between one week and two years. Consequently, it can be used directly to fit real market data instead of the rough Heston model.

Why is it more convenient to use the lifted Heston model rather than its rough counterpart?

On the one hand, it speeds-up calibration time. Indeed, solving numerically the 20-dimensional system of Riccati ordinary differential equations (8.2.10) is up to twenty times faster than the Adams scheme for the fractional Riccati equation. On the other hand, the lifted model remains Markovian and semimartingale, which opens the door to time-efficient recursive simulation schemes for pricing and hedging more complex exotic options. Before testing the lifted model in practice, we compare it to the standard Heston model.

8.3.3 Comparison with the standard Heston model

For the sake of comparison, we calibrate a standard Heston model (8.1.1)-(8.1.2) to the full rough volatility surface $\sigma_\infty(K,T;\Theta_0)$, with Θ_0 as in (8.3.6). Recall that the standard Heston model corresponds to the case $n = 1$, $x_1^1 = 0$ and $c_1^1 = 1$. The calibrated parameters of the standard Heston are provided in Table 8.6 below. We observe that the calibrated values of $(\hat{V}_0, \hat{\theta}, \hat{\rho})$ have the same magnitude as the ones of (8.3.6). This is not surprising since these parameters have the same interpretation in the two models: the first two parameters $(\hat{V}_0, \hat{\theta})$ govern the level of the term structure of forward variance at time 0 while ρ dictates the leverage effect between the stock price and its variance.

Parameters	Calibrated values
$\hat{V}_0$	0.019841
$\hat{\theta}$	0.032471
$\hat{\lambda}$	3.480784
$\hat{\nu}$	0.908037
$\hat{\rho}$	-0.710067

TABLE 8.6: Calibrated Heston model parameters.

Despite the extreme values of the calibrated mean reversion and vol-of-vol parameters $(\hat{\lambda}, \hat{\nu})$, the Heston model is not able to reproduce the steepness of the skew for short maturities as shown on Figure 8.11 below, with a mean squared error of order 2.06e-03. For long maturities, the fit is fairly good.

In order to compare our findings with the observed stylized fact of Figure 8.1, we plot on Figure 8.5 below the term structure of the at-the-money skew of the three models: the rough

Heston with parameters as in (8.3.6), the calibrated *lifted Heston* model of Table 8.4 and the calibrated Heston model of Table 8.6. The Heston model fails in reproducing the explosive behavior of the term structure of the at-the-money skew observed in the market. On the contrary, this feature is captured by the lifted and rough counterparts. For long maturities, all three model have the same behavior.

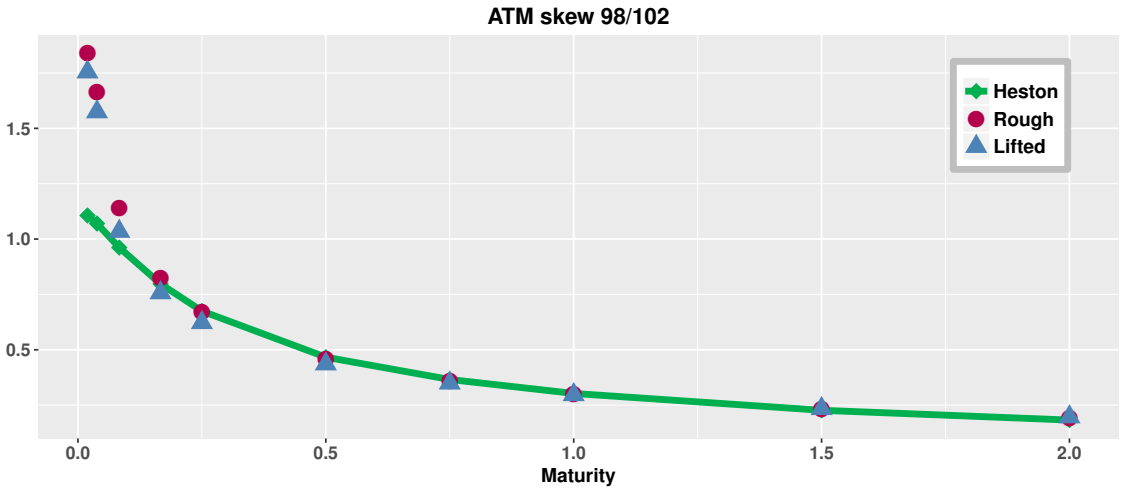

FIGURE 8.5: Term structure of the at-the-money skew of the rough Heston model $\sigma_\infty(K,T;\Theta_0)$ of (8.3.6) (red circles), the calibrated *lifted Heston model* $\sigma_{20}(K,T;r_{20} = 2.5,\hat{\Theta}_0)$ of Table 8.4 (blue triangles) and the calibrated Heston model of Table 8.6 (green line).

In the sequel, we will show that, for $n = 20$ factors, the *lifted Heston model* provides an appealing trade-off between consistency with market data and tractability. We stress that $r_{20} = 2.5$ is kept fixed in the lifted model, which now has only six effective parameters to calibrate $(V_0, \theta, \lambda, \nu, \rho, H)$. Again, in practice, V_0 and Θ_0 can be eliminated by specifying the initial forward variance curve as input and λ can be set to 0, as mean reversions at different speeds are naturally encoded in the lifted model through the family $(x_i^n)_{1\le i\le n}$. By doing so, one reduces the effective number of parameters to only three (ν, ρ, H), as already done in [50] for the rough Heston model.

8.4 Calibration on market data and simulation

In this section, we fix the number of factors to $n = 20$ and set $r_{20} = 2.5$ in (8.3.1). We demonstrate that the *lifted Heston model*:

- captures the explosion of the at-the-money skew observed in the market,
- is easier to simulate than the rough model,
- tricks the human eye as well as the statistical estimator of the Hurst index.

8.4.1 Calibration to the at-the-money skew

Going back to real market data, we calibrate the lifted model to the at-the-money skew of Figure 8.1. Keeping the parameters $V_0 = 0.02$, $\theta = 0.02$ and $\lambda = 0$ fixed, the calibrated parameters are given by

$$\hat{\nu} = 0.3161844, \quad \hat{\rho} = -0.6852625 \quad \text{and} \quad \hat{H} = 0.1104290. \tag{8.4.1}$$

The fit is illustrated on Figure 8.6 below.

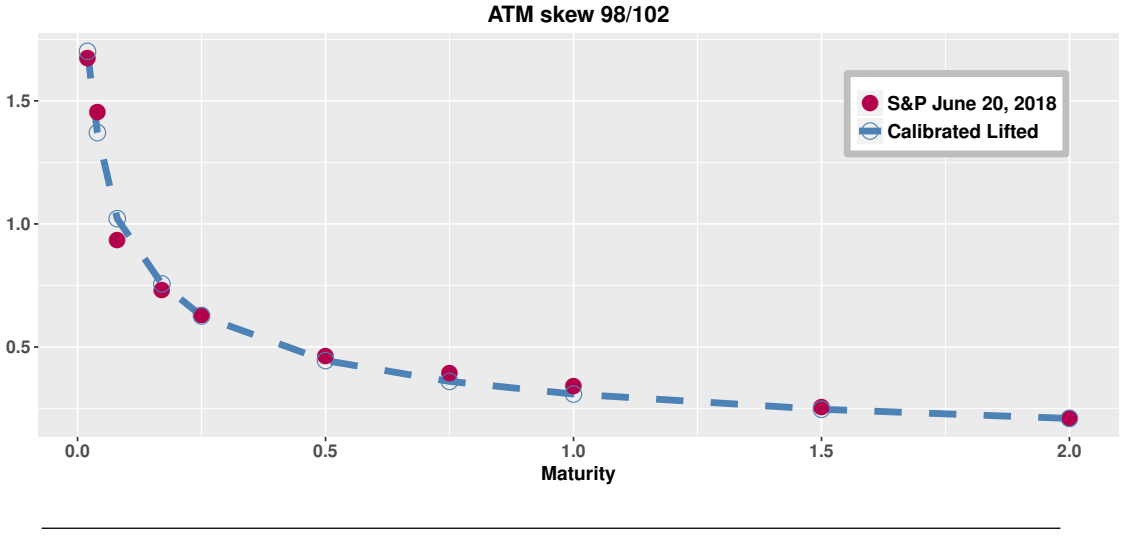

FIGURE 8.6: Term structure of the at-the-money skew for the S&P index on June 20, 2018 (red dots) and for the lifted model with calibrated parameters (8.4.1) (blue circles with dashed line).

We notice the calibrated value $\hat{H}$ in (8.4.1) is coherent with the value $(0.5 - 0.41) = 0.09$, which can be read off the power-law fit of Figure 8.1. Consequently, in the pricing world, the parameter H quantifies the explosion of the at-the-money skew through a power-law $t \to Ct^{0.5-H}$, see also [64].

We discuss briefly the simulation procedure of our lifted model in the next subsection.

8.4.2 Simulation and estimated roughness

Until now, there is no existing scheme to simulate the variance process (8.1.4) of the rough Heston model, the crux resides in the non-Markovianity of the variance process, the singularity of the kernel and the square-root dynamics. In contrast, numerous approximation schemes have been developed for the simulation of the standard square-root process (8.1.2), see [7, Chapters 3 and 4] and the references therein. Because the *lifted Heston model* (8.2.1)-(8.2.3) is a Markovian and semimartingale model, one can adapt standard recursive Euler-Maruyama schemes to simulate the variance process V^n first, and then the stock price S^n. For $T > 0$, we consider the modified explicit-implicit scheme (8.6.12)-(8.6.13) detailed in the Appendix for the variance process V^n.

We observe on Figure 8.7 below that the factors $(U^{20,i})_{1 \le i \le 20}$ are highly correlated. We can distinguish between the short-term factors with fast mean reversions, responsible of the

'roughness', and the long-term factors, with slower mean reversions, determining the level of the variance process. The variance process is then obtained by aggregating these factors with respect to (8.2.2). We also notice that some of the factors $(U^{n,i})_{1\leq i\leq n}$ become negative, but that the aggregated process V^n remains nonnegative at all time.

Remark 8.2 (Nonnegativity of the variance process). Looking at the stochastic differential equation (8.2.2)-(8.2.3), it is not straightforward at all why V^n should stay nonnegative at all time, even for the zero initial curve $g_0 \equiv 0$. Indeed, some of the factors $(U^{n,i})_{1\leq i\leq n}$ may become negative, but surprisingly enough, their aggregated sum V^n remains nonnegative, at all time. This is due to a very special underlying structure: equations (8.2.2)-(8.2.3) can be recast as a stochastic Volterra equation of convolution type for a suitable kernel, we refer to Section 8.6.1 below for more details.

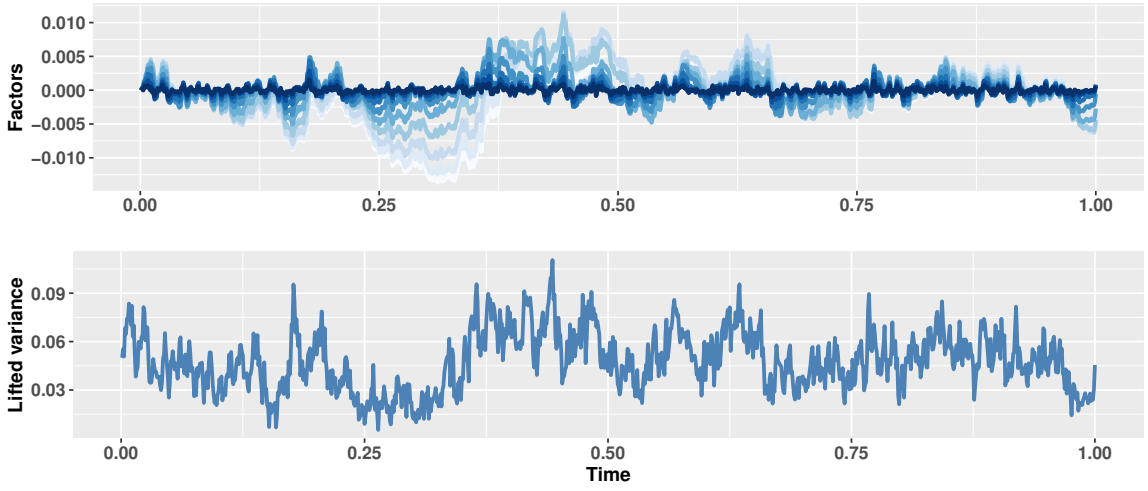

FIGURE 8.7: One sample path of the simulated factors $(U^{20,i})_{1\leq i\leq 20}$ with blue intensity proportional to the speed of mean reversions $(x_i^n)_{1\leq i\leq 20}$ (upper) and the corresponding aggregated variance process V^n (lower) with parameters $V_0 = 0.05$, $\theta = 0.05$, $\lambda = 0.3$, $\nu = 0.1$ and $H = 0.1$ for a time step of 0.001 and $T = 1$.

Visually, the sample path of the variance process seems rougher than the one of a standard Brownian motion. As shown on Figure 8.8 below, at the daily timescale, the simulated volatility process of the *lifted Heston model* not only tricks the human eye, but also misleads the statistical estimator of the Hurst index constructed in [15]. Specifically, the estimator recognizes a semimartingale model for the simulated volatility of the Heston model, with an estimated $\hat{H}$ close to 0.5. However, it fails to do so for the lifted model, the estimator displays $\hat{H} = 0.18$. The lifted model is therefore capable of mimicking, up to some extent, the 'roughness' of the volatility observed on the market, at least at the daily timescale. This should be paralleled with the explosive-like behavior of the at-the-money skew encountered earlier on Figures 8.5-8.6. Stated otherwise, if one is only provided the lower graph of Figure 8.8, one cannot say if the path has been generated by a rough volatility model with Hurst index $H = 0.18$ or by our lifted model with $H = 0.1$. As the step size of the discretization scheme goes to 0, the estimated H of the lifted model has to converge to 0.5, since V^n is a semimartingale, and therefore has the same regularity as a standard Brownian motion. The convergence is illustrated on Figure 8.12 below.

On another note, the upper graph of Figure 8.8 highlights the physical interpretation of the parameter H as it measures the roughness of the empirical realized volatility. Indeed,

empirical studies on a very wide range of assets volatility time-series in [66, 15] revealed that the dynamics of the log-volatility are close to that of a fractional Brownian motion with a 'universal' Hurst parameter H of order 0.1, from intra-day up to daily timescales.

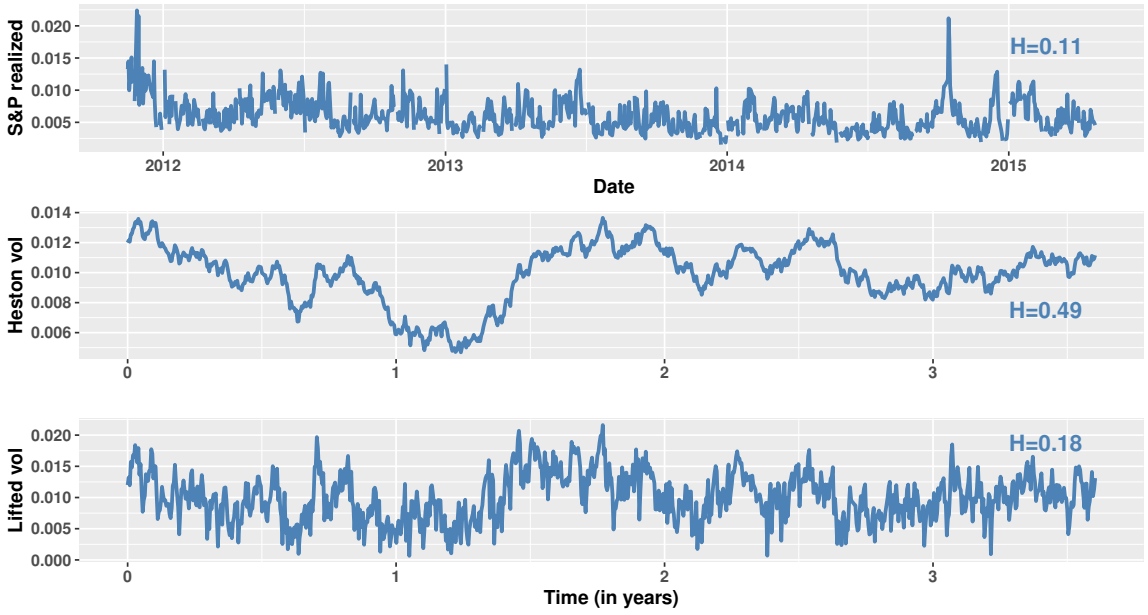

FIGURE 8.8: Estimated Hurst index of: the realized volatility of the S&P[(a)] (upper), a sample path of the volatility process in the Heston model (middle), and a sample path of the volatility process in the the lifted model with $H = 0.1$ (lower). The simulation is run with $N = 250$ time steps for each year.

[(a)]The realized volatility data series can be downloaded from https://realized.oxford-man.ox.ac.uk/.

8.5 Conclusion

We introduced *the lifted Heston model*, a conventional multi-factor stochastic volatility model, where the factors share the same one-dimensional Brownian motion but mean revert at different speeds corresponding to different timescales. The model nests as extreme cases the standard Heston model (for $n = 1$ factor), and the rough Heston model (when n goes to infinity). Inspired by rough volatility models, we provided a good parametrization of the model reducing the number of parameters to calibrate: the model has only one additional effective parameter than the standard Heston model, independently of the number of factors. The first five parameters have the same interpretation as in the standard Heston model, whereas the additional one has a physical interpretation as it is linked to the regularity of the sample paths and the explosion of the at-the-money skew.

This sheds some new light on the reason behind the remarkable fits of rough volatility models. Indeed, a rough variance process can be seen as a superposition of infinitely many factors sharing the same one-dimensional Brownian motion but mean reverting at different speeds

ranging from 0^+ to ∞. Each factor corresponds to a certain timescale. Therefore, time multi-scaling is naturally encoded in rough volatility models, which explains why these models are able to jointly handle different maturities in a satisfactory fashion.[9]

Finally, Table 8.7 below compares the characteristics of the three different models. As it can be seen, the *lifted Heston model* possesses an appealing trade-off between flexibility and tractability!

	Stochastic volatility models		
Characteristics	Heston	Rough Heston	Lifted Heston
Markovian	✓	✗	✓
Semimartingale	✓	✗	✓
Simulation	Fast	Slow	Fast
Affine Volterra process	✓	✓	✓
Characteristic function	Closed	Fractional Riccati	n-Riccati
Calibration	Fast	Slower	20x rough[(a)]
Fit short maturities	✗	✓	✓
Estimated daily regularity	$H \approx 0.5$	$H \approx 0.1$	$H \approx 0.2$

TABLE 8.7: Summary of the characteristics of the different models. [(a)]for $n = 20$.

8.6 Appendices

8.6.1 Existence and uniqueness

In the sequel, the symbol $*$ stands for the convolution operation, that is $(f * \mu)(t) = \int_0^t f(t-s)\mu(ds)$ for any suitable function f and measure μ. For a right-continuous function f of locally bounded variation, we denote by df the measure induced by its distributional derivative, that is $f(t) = f(0) + \int_{(0,t]} df(s)$.

We provide in this appendix the strong existence and uniqueness of (8.2.1)-(8.2.3), for a fixed $n \in \mathbb{N}$. We start by noticing that (8.2.1) is equivalent to

$$S_t^n = \mathcal{E}\left(\int_0^t V_s^n dB_s\right), \quad t \geq 0,$$

where $\mathcal{E}$ is the Doléans-Dade exponential. Therefore, it suffices to prove the existence and uniqueness of (8.2.2)-(8.2.3). Formally, starting from a solution to (8.2.2)-(8.2.3), the variation of constants formula on (8.2.3) yields

$$U_t^{n,i} = \int_0^t e^{-x_i^n(t-s)}\left(-\lambda V_s^n ds + \nu\sqrt{V_s^n}dW_s\right), \quad i = 1, \dots, n, \tag{8.6.1}$$

[9]Multiple timescales in the volatility process have been identified in the literature, see for instance [61, Section 3.4].

so that (8.2.2) reads

$$V_t^n = g_0^n(t) + \int_0^t K^n(t-s)\left(-\lambda V_s^n ds + \nu\sqrt{V_s^n}dW_s\right), \tag{8.6.2}$$

where K^n is the following completely monotone[10] kernel

$$K^n(t) = \sum_{i=1}^n c_i^n e^{-x_i^n t}, \quad t \geq 0. \tag{8.6.3}$$

Whence, if one proves the uniqueness of (8.6.2), then, uniqueness of (8.2.3) follows by virtue of (8.6.1). Conversely, if one proves the existence of a nonnegative solution V^n to (8.6.2), then, one can define $(U^{n,i})_{1\leq i\leq n}$ as in (8.6.1), showing that $(V^n, (U^{n,i})_{1\leq i\leq n})$ is a solution to (8.2.2)-(8.2.3). Therefore, the problem is reduced to proving the existence and uniqueness for the stochastic Volterra equation (8.6.2).

In Chapter 6, the existence of a nonnegative solution to (8.6.2) is proved, provided the initial input curve g_0^n satisfies a certain 'monotonicity' condition. This condition is related to the resolvent of the first kind L^n of the kernel (8.6.3), which is defined as the unique measure satisfying

$$\int_0^t K^n(t-s)L^n(ds) = 1, \quad t \geq 0.^{11}$$

More precisely, denoting by Δ_h the semigroup of right shifts acting on continuous functions, i.e. $\Delta_h f = f(h+\cdot)$ for $h \geq 0$, g_0^n should satisfy

$$\Delta_h g_0^n - (\Delta_h K^n * L^n)(0) g_0^n - d(\Delta_h K^n * L^n) * g_0^n \geq 0, \quad h \geq 0,^{12} \tag{8.6.4}$$

leading to the following definition of the set $\mathcal{G}_{K^n}$ of admissible input curves:

$$\mathcal{G}_{K^n} = \{g_0^n \text{ Hölder continuous of any order less than } 1/2, \text{ satisfying (8.6.4) and } g_0^n(0) \geq 0\}.$$

It is shown in Example 6.2 that the two specifications of input curves (8.2.4)-(8.2.5) provided earlier satisfy (8.6.4).

We now provide the rigorous existence and uniqueness result.

Theorem 8.3 (Existence and uniqueness)**.** *Fix $n \in \mathbb{N}$, $S_0^n > 0$ and assume that $g_0^n \in \mathcal{G}_{K^n}$. Then, the stochastic differential equation (8.2.1)-(8.2.3) has a unique continuous strong solution $(S^n, V^n, (U^{n,i})_{1\leq i\leq n})$ such that $V_t^n \geq 0$, for all $t \geq 0$, almost surely. Further, the process S^n is a martingale.*

Proof. By virtue of the variation of constants formula on the factors, the lifted Heston model is equivalent to a Volterra Heston model in the sense of Chapter 6 of the form

$$dS_t^n = S_t^n\sqrt{V_t^n}dB_t, \quad S_0^n > 0, \tag{8.6.5}$$

$$V_t^n = g_0^n(t) + \int_0^t K^n(t-s)\left(-\lambda V_s^n ds + \nu\sqrt{V_s^n}dW_s\right), \tag{8.6.6}$$

[10] A function f is said to be completely monotone, if it is infinitely differentiable on $(0,\infty)$ such that $(-1)^p f^{(p)} \geq 0$, for all $p \in \mathbb{N}$.

[11] The existence of L^n is ensured by the complete monoticity of K^n, see [69, Theorem 5.5.4].

[12] One can show that $\Delta_h K^n * L^n$ is right-continuous and of locally bounded variation, thus the associated measure $d(\Delta_h K^n * L^n)$ is well defined.

with K^n given by (8.6.3). Since K^n is locally Lipschitz and completely monotone, the assumptions of Theorem 6.1 are met. Consequently, the stochastic Volterra equation (8.6.5)-(8.6.6) has a unique $\mathbb{R}^2_+$-valued weak continuous solution (S^n, V^n) on some filtered probability space $(\Omega^n, \mathcal{F}^n, (\mathcal{F}^n_t)_{t\geq 0}, \mathbb{Q}^n)$ for any initial condition $S^n_0 > 0$ and admissible input curve $g^n_0 \in \mathcal{G}_{K^n}$. Moreover, since K^n is differentiable, strong uniqueness is ensured by Proposition 7.22. The claimed existence and uniqueness statement now follows from (8.6.1). Finally, the martingality of S^n follows along the lines of Theorem 5.24(iii). □

We now discuss the convergence of the *lifted Heston model* towards the rough Heston model (8.1.3)-(8.1.4), as the number of factors goes to infinity, we refer to Chapter 7 for more details. We fix $H \in (0, 1/2)$ and we denote by $K_H : t \to t^{H-\frac{1}{2}}/\Gamma(H+1/2)$ the fractional kernel of the rough Heston model appearing in (8.1.4). The kernel K_H can be re-expressed as a Laplace function

$$K_H(t) = \int_0^\infty e^{-xt}\mu(dx), \quad t \geq 0,$$

with $\mu(dx) = \frac{x^{-\alpha}}{\Gamma(\alpha)\Gamma(1-\alpha)}$ and $\alpha = H+1/2$. On the one hand, for a fixed n, the parametrization (8.3.1) is linked to μ as follows:

$$c^n_i = \int_{\eta^n_{i-1}}^{\eta^n_i} \mu(dx), \quad x^n_i = \frac{1}{c^n_i}\int_{\eta^n_{i-1}}^{\eta^n_i} \mu(dx), \quad i = 1, \dots, n, \tag{8.6.7}$$

where $\eta^n_i = r_n^{i-n/2}$, for $i = 0, \dots, n$. We will show that, under (8.3.2),

$$K^n \to K_H, \quad \text{as } n \text{ goes to infinity, in the } L^2 \text{ sense.} \tag{8.6.8}$$

On the other hand, for each $n \in \mathbb{N}$, we have proved the existence of a solution to (8.6.2). One would therefore expect from (8.6.8) the convergence of the sequence of solutions of (8.6.6) towards the solution of (8.1.4). This is indeed the case, as illustrated by the following theorem, which adapts Theorem 7.7 to the geometric partition.

Theorem 8.4 (Convergence towards the rough Heston model)**.** *Consider a sequence $(r_n)_{n\geq 1}$ satisfying (8.3.2), and set g^n_0 as in (8.2.6) and $(c^n_i, x^n_i)_{1\leq i\leq n}$ as in (8.3.1), for every even $n = 2p$, with $p \geq 1$. Assume $S^n_0 = S_0$, for all n, then, the sequence of solutions $(S^n, V^n)_{n=2p, p\geq 1}$ to (8.2.1)-(8.2.2) converges weakly, on the space of continuous functions on $[0, T]$ endowed with the uniform topology, towards the rough Heston model (8.1.3)-(8.1.4), for any $T > 0$.*

We will only sketch the proof for the L^2 convergence of the kernels (8.6.8), in order to highlight the small adjustments that one needs to make to the proof of Theorem 7.7. Indeed, since $\eta^n_0 \neq 0$ in our case, Theorem 7.7 cannot be directly applied, compare with Assumption 7.3.1 where the left-end point of the partition is zero. The following lemma adapts Proposition 7.5 to the geometric partition. The rest of the proof of Theorem 8.4 follows along the lines of Theorem 7.7 by making the same small adjustments highlighted below, mainly to treat the integral chunk between $[0, \eta^n_0]$.

Lemma 8.5 (Convergence of K^n towards K_H)**.** *Let $(r_n)_{n\geq 1}$ as in (8.3.2), and $(c^n_i, x^n_i)_{1\leq i\leq n}$ given by (8.3.1). Define K^n by (8.6.3), then,*

$$\|K^n - K_H\|_{L^2(0,T)} \to 0, \quad \text{as } n \to \infty, \tag{8.6.9}$$

for all $T > 0$.

Proof. Set $\eta_i^n = r_n^{i-n/2}$, for $i = 0, \dots, n$. Using (8.6.7), we start by decomposing $(K_H - K^n)$ as follows

$$\begin{aligned} K_H - K^n &= \int_0^\infty e^{-x(\cdot)}\mu(dx) - \sum_{i=1}^n c_i^n e^{-x_i^n(\cdot)} \\ &= \int_0^{\eta_0^n} e^{-x(\cdot)}\mu(dx) + \left(\sum_{i=1}^n \int_{\eta_{i-1}^n}^{\eta_i^n} \left(e^{-x(\cdot)} - e^{-x_i^n(\cdot)} \right) \mu(dx) \right) + \int_{\eta_n^n}^\infty e^{-x(\cdot)}\mu(dx) \\ &:= J_1^n + J_2^n + J_3^n, \end{aligned}$$

so that

$$\|K_H - K^n\|_{L^2(0,T)} \le I_1^n + I_2^n + I_3^n,$$

with $I_k^n = \|J_k^n\|_{L^2(0,T)}$, for $k = 1, 2, 3$. We now prove that each $I_k^n \to 0$, as n tends to ∞. Relying on a second order Taylor expansion, along the lines of the proof of [33, Proposition 7.1], we get the following bound

$$\left| \int_{\eta_{i-1}^n}^{\eta_i^n} \left(e^{-xt} - e^{-x_i^n t} \right) \mu(dx) \right| \le C\, t^2\, r_n^{1/2}\, (r_n - 1)^2 \int_{\eta_{i-1}^n}^{\eta_i^n} (1 \wedge x^{-1/2}) \mu(dx), \quad t \le T,$$

for all $i = 1, \dots, n$, where C is a constant independent of n, i and t. Summation over $i = 1, \dots, n$ leads to

$$I_2^n \le C\, \frac{T^{5/2}}{\sqrt{5}}\, r_n^{1/2}\, (r_n - 1)^2 \int_0^\infty (1 \wedge x^{-1/2}) \mu(dx),$$

so that $I_2^n \to 0$, as $n \to \infty$, by virtue of the first condition in (8.3.2). On another note,

$$I_1^n \le \int_0^{\eta_0^n} \mu(dx) = \frac{(\eta_0^n)^{1-\alpha}}{\Gamma(\alpha)\Gamma(2-\alpha)} = \frac{r_n^{-(1-\alpha)n/2}}{\Gamma(\alpha)\Gamma(2-\alpha)} \to 0, \quad \text{when } n \to \infty,$$

thanks to the second condition in (8.3.2). Similarly,

$$I_3^n \le \int_{\eta_n^n}^\infty \sqrt{\frac{1 - e^{-2xT}}{2x}} \mu(dx) \le \frac{r_n^{(1/2-\alpha)n/2}}{\Gamma(\alpha)\Gamma(1-\alpha)(1/2-\alpha)} \to 0, \quad \text{when } n \to \infty.$$

Combining the above leads to (8.6.9). □

8.6.2 The full Fourier-Laplace transform

We provide the full Fourier-Laplace transform for the joint process $X^n := (\log S^n, V^n)$ extending (8.2.9). The formula can be used to price path-dependent options on the stock price S^n and the variance process V^n.

Once again, this is a particular case of Section 6.4, by observing that K^n defined in (8.6.3) is the Laplace transform of the following nonnegative measure

$$\mu^n(dx) = \sum_{i=1}^n c_i^n \delta_{x_i^n}(dx).$$

Fix row vectors $u = (u_1, u_2) \in \mathbb{C}^2$ and $f \in L^1_{\mathrm{loc}}(\mathbb{R}_+, (\mathbb{C}^2))$ such that

$$\mathrm{Re}\,(u_1 + 1 * f_1) \in [0,1], \quad \mathrm{Re}\,u_2 \leq 0 \quad \text{and} \quad \mathrm{Re}\,f_2 \leq 0,$$

then, it follows from Remark 6.8 with $\mu = \sum_{i=1}^n c_i^n \delta_{x_i^n}$ that the Fourier-Laplace transform of $X^n = (\log S^n, V^n)$ is exponentially affine with respect to the family $(U^{n,i})_{1\leq i\leq n}$,

$$\mathbb{E}\left[\exp\left(uX_T^n + (f * X^n)_T\right) \,\middle|\, \mathcal{F}_t\right] = \exp\left(\phi^n(t,T) + \psi_1(T-t)\log S_t^n + \sum_{i=1}^n c_i^n \psi_2^{n,i}(T-t)U_t^{n,i}\right),$$

for all $t \leq T$, where $(\psi_1, (\psi_2^{n,i})_{1\leq i\leq n})$ are the unique solutions of the following system of Riccati ordinary differential equations

$$\begin{aligned}
\psi_1 &= u_1 + 1 * f_1, \\
(\psi_2^{n,i})' &= -x_i^n \psi_2^{n,i} + F\left(\psi_1, \sum_{j=1}^n c_j^n \psi_2^{n,j}\right), \quad \psi_2^{n,i}(0) = u_2, \quad i = 1, \ldots, n,
\end{aligned}$$

with

$$F(\psi_1, \psi_2) = f_2 + \frac{1}{2}\left(\psi_1^2 - \psi_1\right) + (\rho\nu\psi_1 - \lambda)\psi_2 + \frac{\nu^2}{2}\psi_2^2$$

and

$$\phi^n(t,T) = u_2 g_0^n(T) + \int_0^{T-t} F\left(\psi_1, \sum_{i=1}^n c_i^n \psi_2^{n,i}(s)\right) g_0^n(T-s)ds + \int_0^t f(T-s)X_s ds, \quad t \leq T.$$

8.6.3 Discretization schemes

Riccati equations

The aim of this section is to design an approximation scheme of the n-dimensional Riccati system of equations (8.2.10). In order to gain some insights, consider first the case where $F \equiv 0$ so that (8.2.10) reduces to

$$(\psi^{n,i})' = -x_i^n \psi^{n,i}, \quad i = 1, \ldots, n, \tag{8.6.10}$$

and the solution is given by

$$\psi^{n,i}(t) = \psi^{n,i}(0)e^{-x_i^n t}, \quad i = 1, \ldots, n.$$

One could start with an explicit Euler scheme for (8.6.10), that is

$$\hat{\psi}_{t_{k+1}}^{n,i} = \hat{\psi}_{t_k}^{n,i} - x_i^n \Delta t \hat{\psi}_{t_k}^{n,i} = (1 - x_i^n \Delta t)\,\hat{\psi}_{t_k}^{n,i}, \quad i = 1, \ldots, n,$$

for a regular time grid $t_k = (kT)/N$ for all $k = 1, \ldots, N$, where T is the terminal time, N the number of time steps and $\Delta t = T/N$. A sufficient condition for the stability of the scheme reads

$$\Delta t \leq \min_{1\leq i\leq n} \frac{1}{x_i^n}.$$

Recall from (8.3.1) that x_n^n grows very large as n increases. For instance, for $n = 20$, $r_{20} = 2.5$ and $H = 0.1$, $x_n^n = 6417.74$. Consequently, if one needs to ensure the stability of the explicit

scheme, one needs a very large number of time steps N. In contrast, the implicit Euler scheme

$$\hat{\psi}^{n,i}_{t_{k+1}} = \hat{\psi}^{n,i}_{t_k} - x^n_i \Delta t\, \hat{\psi}^{n,i}_{t_{k+1}}, \quad i = 1, \dots, n,$$

is stable for any number of time steps N and reads

$$\hat{\psi}^{n,i}_{t_{k+1}} = \frac{1}{1 + x^n_i \Delta t} \hat{\psi}^{n,i}_{t_k}, \quad i = 1, \dots, n.$$

For this reason, we consider the following explicit-implicit discretization scheme of the n-dimensional Riccati system of equations (8.2.10)

$$\hat{\psi}^{n,i}_0 = 0, \quad \hat{\psi}^{n,i}_{t_{k+1}} = \frac{1}{1 + x^n_i \Delta t} \left(\hat{\psi}^{n,i}_{t_k} + \Delta t\, F \left(u, \sum_{j=1}^{n} c^n_j \hat{\psi}^{n,j}_{t_k} \right) \right), \quad i = 1, \dots, n, \tag{8.6.11}$$

for a regular time grid $t_k = k\Delta t$ for all $k = 1, \dots, N$, with time step size $\Delta t = T/N$, terminal time T and number of time steps N. Alternatively, one could also consider the exponential scheme for the Riccati equations by replacing the term $1/(1 + x^n_i \Delta t)$ with $e^{-x^n_i \Delta t}$. One can also combine more involved discretization schemes for the explicit part involving the quadratic function F, for instance higher order Runge-Kutta methods can be used, see [83].

Stochastic process

Similarly, we suggest to consider the following modified explicit-implicit scheme for the variance process V^n:

$$\hat{V}^n_{t_k} = g^n_0(t_k) + \sum_{i=1}^{n} c^n_i \hat{U}^{n,i}_{t_k}, \quad \hat{U}^{n,i}_0 = 0, \tag{8.6.12}$$

$$\hat{U}^{n,i}_{t_{k+1}} = \frac{1}{1 + x^n_i \Delta t} \left(\hat{U}^{n,i}_{t_k} - \lambda \hat{V}^n_{t_k} \Delta t + \nu \sqrt{\left(\hat{V}^n_{t_k}\right)^+} \left(W_{t_{k+1}} - W_{t_k} \right) \right), \quad i = 1, \dots, n, \tag{8.6.13}$$

for a regular time grid $t_k = k\Delta t$, $k = 1 \dots N$, $\Delta t = T/N$ and $(W_{t_{k+1}} - W_{t_k}) \sim \mathcal{N}(0, \Delta t)$. Notice that we take the positive part $(\cdot)^+$ since the simulated process can become negative. Once there, simulating the spot-price process S^n is straightforward. We leave the theoretical study of convergence and stability for future work. Numerically, the scheme seems stable. Alternatively, one could also consider the exponential scheme for the stochastic process by replacing the term $1/(1 + x^n_i \Delta t)$ with $e^{-x^n_i \Delta t}$. As a final remark, one notices that (8.6.12)-(8.6.13) corresponds to the space-time discretization of the integro-differential stochastic partial differential equation (8.3.8)-(8.3.9). This is illustrated on Figure 8.9 below.

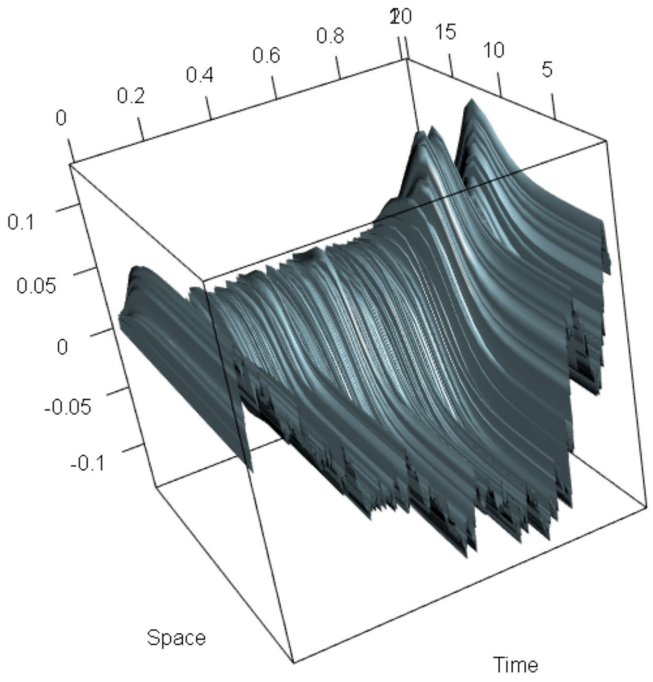

FIGURE 8.9: Simulated path of the stochastic partial differential equation (8.3.8)-(8.3.9) by using the scheme (8.6.12)-(8.6.13).

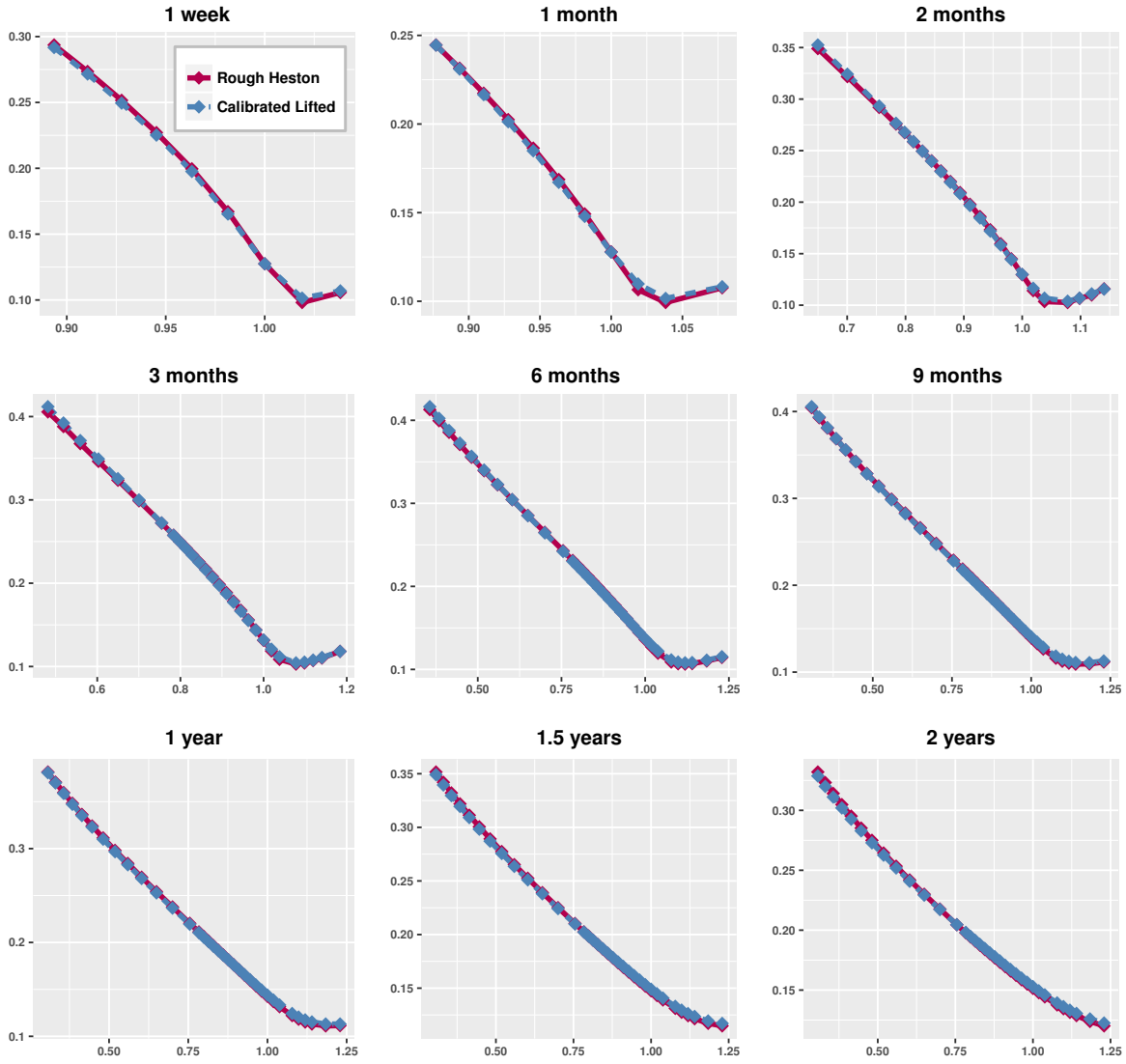

FIGURE 8.10: Implied volatility surface of the rough Heston model $\sigma_\infty(K,T;\Theta_0)$ of (8.3.6) (red) and the calibrated lifted Heston model $\sigma_{20}(K,T;r_{20}=2.5,\hat{\Theta}_0)$ of Table 8.4 (blue) for maturities ranging from 1 week to 2 years (MSE = 4.01e-07).

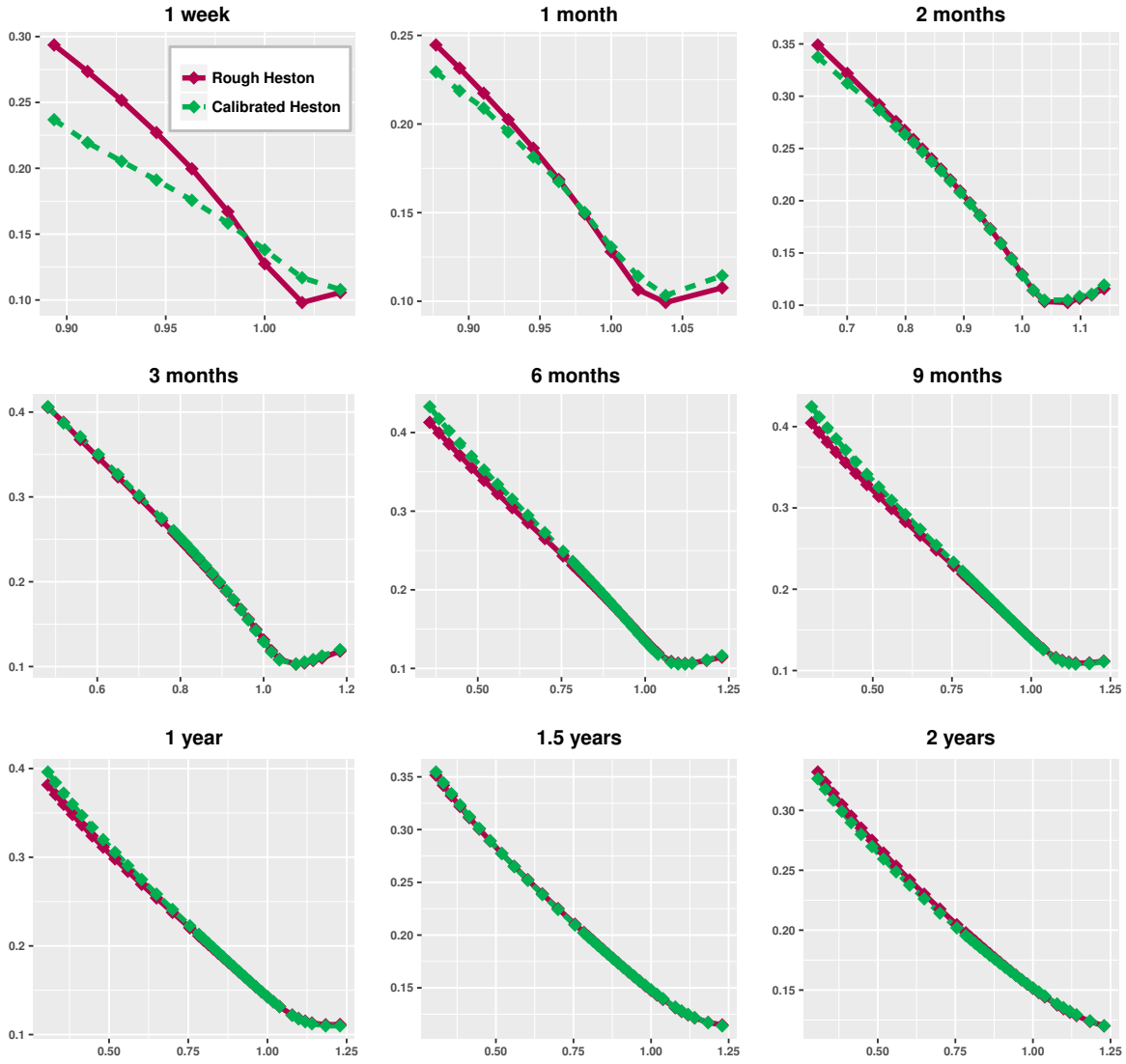

FIGURE 8.11: Implied volatility surface of the rough Heston model $\sigma_\infty(K,T;\Theta_0)$ (red) and the calibrated Heston model of Table 8.6 (green) for maturities ranging from 1 week to 2 years (MSE = 2.06e-03).

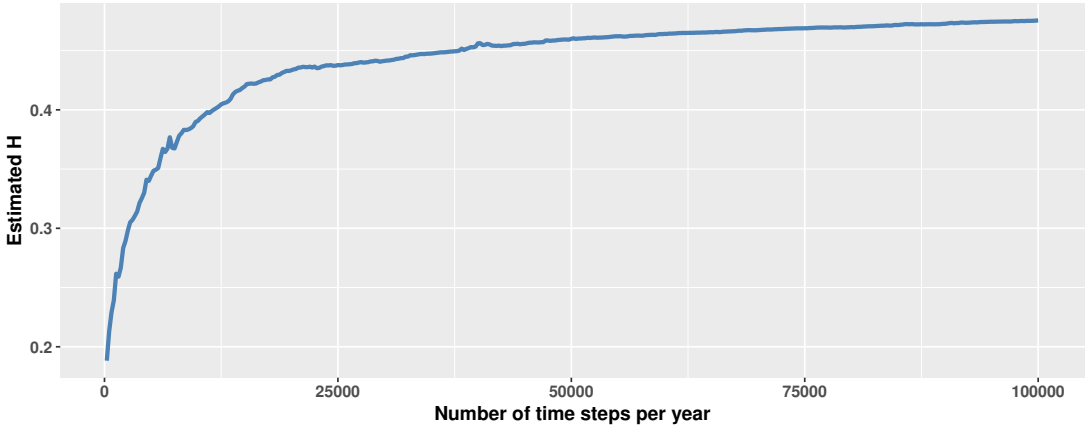

FIGURE 8.12: Estimated Hurst index of the simulated sample path of the volatility process in the lifted model with $H = 0.1$ as a function of the number of time steps per year.

ν	ρ	H	MSE
0.22	-0.67	0.09	3.63e-06
0.14	-0.54	0.19	5.34e-06
0.35	-0.65	0.19	8.17e-06
0.14	-0.83	0.06	9.74e-05
0.22	-0.59	0.15	4.60e-06
0.37	-0.50	0.12	4.55e-06
0.40	-0.53	0.11	4.56e-06
0.34	-0.85	0.08	3.45e-04
0.22	-0.89	0.09	1.25e-04
0.44	-0.76	0.11	2.79e-04
0.32	-0.70	0.12	4.56e-06
0.42	-0.63	0.08	5.22e-06
0.10	-0.61	0.17	3.69e-06
0.42	-0.64	0.11	4.81e-06
0.30	-0.69	0.17	5.96e-06
0.06	-0.71	0.17	2.98e-06
0.36	-0.71	0.16	6.14e-06
0.25	-0.80	0.18	1.63e-04
0.09	-0.77	0.06	2.87e-06
0.35	-0.74	0.13	1.44e-04

TABLE 8.8: **Robustness of $r_{20} = 2.5$:** First 20 values of the simulated parameters and the corresponding mean squared error between the implied volatility surface of the lifted model $\sigma_{20}(K, T; 2.5, \Theta_k)$ and the rough model $\sigma_\infty(K, T; \Theta_k)$, for $k = 1, \ldots, 20$.

Appendix A

Matrix tools

We collect in this Appendix some definitions and properties of matrix tools intensively used in the proofs of the first part of this thesis. For a complete review and proofs we refer to [90, 91, 98].

We start by recalling the definition of the Moore-Penrose pseudoinverse which generalizes the concept of invertibility of square matrices, to non-singular and non-square matrices. In the following, we denote by $\mathbb{R}^{m\times n}$ the collection of $m \times n$ matrices.

Definition A.1 (Moore-Penrose pseudoinverse)**.** Fix $A \in \mathbb{R}^{m\times n}$. The Moore-Penrose pseudoinverse of A is the unique $n \times m$ matrix A^+ satisfying: $AA^+A = A$, $A^+AA^+ = A^+$, AA^+ and A^+A are Hermitian.

Proposition A.2. *If $A \in \mathbb{R}^{d\times d}$ has the spectral decomposition $Q\Lambda Q^\top$ for some orthogonal matrix $Q \in \mathbb{R}^{d\times d}$ and a diagonal matrix $\Lambda = \operatorname{diag}[(\lambda_i)_{i\le d}] \in \mathbb{R}^{d\times d}$. Then, $A^+ = Q\Lambda^+Q^\top$ in which $\Lambda^+ = \operatorname{diag}\left[(\lambda_i^{-1}\mathbb{1}_{\{\lambda_i\neq 0\}})_{i\le d}\right]$, and $AA^+ = Q\operatorname{diag}\left[(\mathbb{1}_{\{\lambda_i\neq 0\}})_{i\le d}\right]Q^\top$. If moreover A is positive semi-definite and $B = A^{\frac{1}{2}}$, then $B^+ = Q(\Lambda^+)^{\frac{1}{2}}Q^\top$.*

Proposition A.3. *If $A \in \mathbb{R}^{m\times n}$, then AA^+ is the orthogonal projection on the image of A.*

We now collect some useful identities on the Kronecker product.

Definition A.4 (Kronecker product)**.** Let $A = (a_{ij})_{i\le m_1, j\le n_1} \in \mathbb{R}^{m_1\times n_1}$ and $B \in \mathbb{R}^{m_2\times n_2}$. The Kronecker product $(A \otimes B)$ is defined as the $m_1m_2 \times n_1n_2$ matrix

$$A \otimes B = \begin{pmatrix} a_{11}B & \cdots & a_{1n_1}B \\ \vdots & & \vdots \\ a_{m_11}B & \cdots & a_{m_1n_1}B \end{pmatrix}.$$

Proposition A.5. *Let A and B be as in Definition A.4, $C \in \mathbb{R}^{n_1\times n_3}$ and $D \in \mathbb{R}^{n_2\times n_4}$. Then,*

$$\begin{aligned} &(A \otimes B)(C \otimes D) = (AC \otimes BD), \\ &A \otimes B = A(I_{n_1} \otimes B) \text{ if } m_2 = 1, \\ &A \otimes B = B(A \otimes I_{n_2}) \text{ if } m_1 = 1. \end{aligned}$$

The following definitions extend the concept of Jacobian matrix and show how to nicely stack the partial derivatives of a matrix-valued function $F : \mathbb{R}^{n\times q} \mapsto \mathbb{R}^{m\times p}$ by using the vectorization operator (see [91, Chapter 9]).

Definition A.6 (Vectorization operator)**.** Let $A \in \mathbb{R}^{m\times n}$. The vectorization operator vec transforms the matrix into a vector in $\mathbb{R}^{mn}$ by stacking all the columns of the matrix A one underneath the other.

Definition A.7 (Jacobian matrix)**.** Let F be a differentiable map from $\mathbb{R}^{n\times q}$ to $\mathbb{R}^{m\times p}$. The Jacobian matrix $DF(X)$ of F at X is defined as the following $mp \times nq$ matrix:

$$DF(X) = \frac{\partial \operatorname{vec}(F(X))}{\partial \operatorname{vec}(X)^\top}.$$

Proposition A.8 (Product rule)**.** *Let G be a differentiable map from $\mathbb{R}^{n\times q}$ to $\mathbb{R}^{m\times p}$ and H be a differentiable map from $\mathbb{R}^{n\times q}$ to $\mathbb{R}^{p\times l}$. Then, $D(GH) = (H^\top \otimes I_m)DG + (I_l \otimes G)DH$.*

Bibliography

[1] Abi Jaber, E. Stochastic invariance of closed sets for jump-diffusions with non-lipschitz coefficients. *Electronic Communications in Probability 22* (2017).

[2] Abi Jaber, E. Lifting the Heston model. *arXiv preprint arXiv:1810.04868* (2018).

[3] Abi Jaber, E., Bouchard, B., and Illand, C. Stochastic invariance of closed sets with non-Lipschitz coefficients. *arXiv preprint arXiv:1607.08717. Accepted for publication in Stochastic Processes and their Applications* (2016).

[4] Abi Jaber, E., and El Euch, O. Markovian structure of the Volterra Heston model. *arXiv preprint arXiv:1803.00477* (2018).

[5] Abi Jaber, E., and El Euch, O. Multi-factor approximation of rough volatility models. *arXiv preprint arXiv:1801.10359* (2018).

[6] Abi Jaber, E., Larsson, M., and Pulido, S. Affine Volterra processes. *arXiv preprint arXiv:1708.08796* (2017).

[7] Alfonsi, A. *Affine diffusions and related processes: simulation, theory and applications*, vol. 6 of *Bocconi & Springer Series*. Springer, Cham; Bocconi University Press, Milan, 2015.

[8] Applebaum, D. *Lévy processes and stochastic calculus*, second ed., vol. 116 of *Cambridge Studies in Advanced Mathematics*. Cambridge University Press, Cambridge, 2009.

[9] Aubin, J.-P. *Viability theory*. Systems & Control: Foundations & Applications. Birkhäuser Boston, Inc., Boston, MA, 1991.

[10] Aubin, J.-P., and Doss, H. Characterization of stochastic viability of any nonsmooth set involving its generalized contingent curvature.

[11] Aubin, J.-P., and Frankowska, H. Set-valued analysis. systems and control: Foundations and applications, vol. 2. *Birkhuser, Boston* (1990).

[12] Bardi, M., and Goatin, P. Invariant sets for controlled degenerate diffusions: a viscosity solutions approach. In *Stochastic analysis, control, optimization and applications*. Springer, 1999, pp. 191–208.

[13] Bardi, M., and Jensen, R. A geometric characterization of viable sets for controlled degenerate diffusions. *Set-Valued Analysis 10*, 2-3 (2002), 129–141.

[14] Bayer, C., Friz, P., and Gatheral, J. Pricing under rough volatility. *Quantitative Finance 16*, 6 (2016), 887–904.

[15] Bennedsen, M., Lunde, A., and Pakkanen, M. S. Decoupling the short- and long-term behavior of stochastic volatility. *arXiv preprint arXiv:1610.00332* (2016).

[16] Berger, M. A., and Mizel, V. J. Volterra equations with Itô integrals. I. *J. Integral Equations 2*, 3 (1980), 187–245.

[17] Berger, M. A., and Mizel, V. J. Volterra equations with Itô integrals. II. *J. Integral Equations 2*, 4 (1980), 319–337.

[18] Bergomi, L. Smile dynamics II. *Risk 18* (2005), 67–73.

[19] Bertsekas, D. P., and Shreve, S. *Stochastic optimal control: the discrete-time case*. 2004.

[20] Bhatia, R. *Matrix analysis*, vol. 169. Springer Science & Business Media, 2013.

[21] Billingsley, P. *Convergence of probability measures*, second ed. Wiley Series in Probability and Statistics: Probability and Statistics. John Wiley & Sons, Inc., New York, 1999. A Wiley-Interscience Publication.

[22] Bruder, B. Super-replication of european options with a derivative asset under constrained finite variation strategies. *Preprint HAL http://hal. archives-ouvertes. fr/hal-00012183/fr* (2005), 14.

[23] Buckdahn, R., Quincampoix, M., Rainer, C., and Teichmann, J. Another proof for the equivalence between invariance of closed sets with respect to stochastic and deterministic systems. *Bull. Sci. Math. 134*, 2 (2010), 207–214.

[24] Carja, O., Necula, M., and Vrabie, I. I. *Viability, invariance and applications*, vol. 207. Elsevier, 2007.

[25] Carmona, P., and Coutin, L. Fractional Brownian motion and the Markov property. *Electron. Comm. Probab. 3* (1998), 95–107.

[26] Carmona, P., Coutin, L., and Montseny, G. Approximation of some Gaussian processes. *Stat. Inference Stoch. Process. 3*, 1-2 (2000), 161–171. 19th "Rencontres Franco-Belges de Statisticiens" (Marseille, 1998).

[27] Carr, P., and Madan, D. Option valuation using the fast Fourier transform. *Journal of Computational Finance 2*, 4 (1999), 61–73.

[28] Çinlar, E., and Jacod, J. Representation of semimartingale Markov processes in terms of Wiener processes and Poisson random measures. In *Seminar on Stochastic Processes, 1981 (Evanston, Ill., 1981)*, vol. 1 of *Progr. Prob. Statist.* Birkhäuser, Boston, Mass., 1981, pp. 159–242.

[29] Cheridito, P., Filipović, D., and Yor, M. Equivalent and absolutely continuous measure changes for jump-diffusion processes. *Ann. Appl. Probab. 15*, 3 (2005), 1713–1732.

[30] Cheridito, P., Soner, H. M., and Touzi, N. Small time path behavior of double stochastic integrals and applications to stochastic control. *Ann. Appl. Probab. 15*, 4 (2005), 2472–2495.

[31] Christoffersen, P., Heston, S., and Jacobs, K. The shape and term structure of the index option smirk: Why multifactor stochastic volatility models work so well. *Management Science 55*, 12 (2009), 1914–1932.

[32] Cont, R., and Tankov, P. *Financial modelling with jump processes*, vol. 2. CRC press, 2003.

[33] Coutin, L., and Pontier, M. Approximation of the fractional brownian sheet via ornstein-uhlenbeck sheet. *ESAIM: Probability and Statistics 11* (2007), 115–146.

[34] Cuchiero, C., Filipović, D., Mayerhofer, E., and Teichmann, J. Affine processes on positive semidefinite matrices. *Ann. Appl. Probab. 21*, 2 (2011), 397–463.

[35] Cuchiero, C., Keller-Ressel, M., and Teichmann, J. Polynomial processes and their applications to mathematical finance. *Finance Stoch. 16*, 4 (2012), 711–740.

[36] Cuchiero, C., and Teichmann, J. Generalized Feller processes and Markovian lifts of stochastic Volterra processes: the affine case. *arXiv preprint arXiv:1804.10450* (2018).

[37] Da Fonseca, J., Grasselli, M., and Tebaldi, C. A multifactor volatility heston model. *Quantitative Finance 8*, 6 (2008), 591–604.

[38] Da Prato, G., and Frankowska, H. Invariance of stochastic control systems with deterministic arguments. *J. Differential Equations 200*, 1 (2004), 18–52.

[39] Da Prato, G., and Frankowska, H. Stochastic viability of convex sets. *Journal of mathematical analysis and applications 333*, 1 (2007), 151–163.

[40] Dalang, R. Extending the martingale measure stochastic integral with applications to spatially homogeneous spde's. *Electronic Journal of Probability 4* (1999).

[41] Dawson, D. A., and Fleischmann, K. A super-brownian motion with a single point catalyst. *Stochastic Processes and their Applications 49*, 1 (1994), 3–40.

[42] Decreusefond, L. Regularity properties of some stochastic Volterra integrals with singular kernel. *Potential Analysis 16*, 2 (2002), 139–149.

[43] Delbaen, F., and Schachermayer, W. A compactness principle for bounded sequences of martingales with applications. In *Seminar on Stochastic Analysis, Random Fields and Applications (Ascona, 1996)*, vol. 45 of *Progr. Probab.* Birkhäuser, Basel, 1999, pp. 137–173.

[44] Diethelm, K., Ford, N. J., and Freed, A. D. A predictor-corrector approach for the numerical solution of fractional differential equations. *Nonlinear Dynamics 29*, 1-4 (2002), 3–22.

[45] Diethelm, K., Ford, N. J., and Freed, A. D. Detailed error analysis for a fractional Adams method. *Numerical algorithms 36*, 1 (2004), 31–52.

[46] Diethelm, K., and Freed, A. D. The fracpece subroutine for the numerical solution of differential equations of fractional order. In *Forschung und Wissenschaftliches Rechnen 1998*. Gesellschaft für Wisseschaftliche Datenverarbeitung Gottingen, Germany, 1999, pp. 57–71.

[47] Doss, H. Liens entre équations différentielles stochastiques et ordinaires. *Ann. Inst. H. Poincaré Sect. B (NS) 13*, 2 (1977), 99–125.

[48] Duffie, D., Filipović, D., and Schachermayer, W. Affine processes and applications in finance. *Ann. Appl. Probab. 13*, 3 (2003), 984–1053.

[49] El Euch, O., Fukasawa, M., and Rosenbaum, M. The microstructural foundations of leverage effect and rough volatility. *Finance and Stochastics 22*, 2 (2018), 241–280.

[50] El Euch, O., Gatheral, J., and Rosenbaum, M. Roughening Heston.

[51] El Euch, O., and Rosenbaum, M. The characteristic function of rough Heston models. *arXiv preprint arXiv:1609.02108* (2016).

[52] El Euch, O., and Rosenbaum, M. Perfect hedging in rough Heston models. *arXiv arXiv:1703.05049* (2017).

[53] El Karoui, N., and Lepeltier, J.-P. Représentation des processus ponctuels multivariés à l'aide d'un processus de Poisson. *Z. Wahrscheinlichkeitstheorie und Verw. Gebiete 39*, 2 (1977), 111–133.

[54] Ethier, S., and Kurtz, T. *Markov processes*. Wiley Series in Probability and Mathematical Statistics: Probability and Mathematical Statistics. John Wiley & Sons, Inc., New York, 1986. Characterization and convergence.

[55] Fang, F., and Oosterlee, C. W. A novel pricing method for european options based on fourier-cosine series expansions. *SIAM Journal on Scientific Computing 31*, 2 (2008), 826–848.

[56] Filipović, D. *Term-structure models*. Springer Finance. Springer-Verlag, Berlin, 2009. A graduate course.

[57] Filipović, D., and Larsson, M. Polynomial diffusions and applications in finance. *Finance and Stochastics 20*, 4 (2016), 931–972.

[58] Filipovic, D., and Mayerhofer, E. Affine diffusion processes: Theory and applications. *Advanced Financial Modelling 8* (2009), 1–40.

[59] Filipović, D., Tappe, S., and Teichmann, J. Invariant manifolds with boundary for jump-diffusions. *Electron. J. Probab. 19* (2014), no. 51, 28.

[60] Fouque, J.-P., and Lorig, M. J. A fast mean-reverting correction to heston's stochastic volatility model. *SIAM Journal on Financial Mathematics 2*, 1 (2011), 221–254.

[61] Fouque, J.-P., Papanicolaou, G., Sircar, R., and Sølna, K. *Multiscale stochastic volatility for equity, interest rate, and credit derivatives*. Cambridge University Press, 2011.

[62] Friedman, A. *Stochastic differential equations and applications*. Courier Corporation, 2012.

[63] Fukasawa, M. Asymptotic analysis for stochastic volatility: Martingale expansion. *Finance and Stochastics 15*, 4 (2011), 635–654.

[64] Fukasawa, M. Short-time at-the-money skew and rough fractional volatility. *Quantitative Finance 17*, 2 (2017), 189–198.

[65] Gatheral, J. *The volatility surface: a practitioner's guide*, vol. 357. John Wiley & Sons, 2011.

[66] Gatheral, J., Jaisson, T., and Rosenbaum, M. Volatility is rough. *Quantitative Finance* (2018), 1–17.

[67] Gatheral, J., and Keller-Ressel, M. Affine forward variance models. *arXiv preprint arXiv:1801.06416* (2018).

[68] Gourieroux, C., and Sufana, R. A classification of two-factor affine diffusion term structure models. *Journal of Financial Econometrics 4*, 1 (2005), 31–52.

[69] Gripenberg, G., Londen, S.-O., and Staffans, O. *Volterra integral and functional equations*, vol. 34 of *Encyclopedia of Mathematics and its Applications*. Cambridge University Press, Cambridge, 1990.

[70] Guennoun, H., Jacquier, A., and Roome, P. Asymptotic behaviour of the fractional Heston model. *arXiv preprint arXiv:1411.7653* (2017).

[71] Harms, P., and Stefanovits, D. Affine representations of fractional processes with applications in mathematical finance. *arXiv preprint arXiv:1510.04061* (2015).

[72] Heston, S. L. A closed-form solution for options with stochastic volatility with applications to bond and currency options. *Review of Financial Studies 6*, 2 (1993), 327–343.

[73] Hofmanová, M., and Seidler, J. On weak solutions of stochastic differential equations. *Stoch. Anal. Appl. 30*, 1 (2012), 100–121.

[74] Ikeda, N., and Watanabe, S. *Stochastic differential equations and diffusion processes*, vol. 24 of *North-Holland Mathematical Library*. North-Holland Publishing Co., Amsterdam-New York; Kodansha, Ltd., Tokyo, 1981.

[75] Itô, K. *On stochastic differential equations*, vol. 4. American Mathematical Soc., 1951.

[76] Jacod, J., and Protter, P. *Discretization of processes*, vol. 67 of *Stochastic Modelling and Applied Probability*. Springer, Heidelberg, 2012.

[77] Jacod, J., and Shiryaev, A. *Limit theorems for stochastic processes*, second ed., vol. 288 of *Grundlehren der Mathematischen Wissenschaften [Fundamental Principles of Mathematical Sciences]*. Springer-Verlag, Berlin, 2003.

[78] Jaisson, T. *Market activity and price impact throughout time scales*. PhD thesis, Ecole Polytechnique, 2015.

[79] Jaisson, T., and Rosenbaum, M. Rough fractional diffusions as scaling limits of nearly unstable heavy tailed hawkes processes. *The Annals of Applied Probability 26*, 5 (2016), 2860–2882.

[80] Kallsen, J. A didactic note on affine stochastic volatility models. In *From stochastic calculus to mathematical finance*. Springer, Berlin, 2006, pp. 343–368.

[81] Kurtz, T. Equivalence of stochastic equations and martingale problems. In *Stochastic analysis 2010*. Springer, Heidelberg, 2011, pp. 113–130.

[82] Kurtz, T. G. Lectures on stochastic analysis. *Department of Mathematics and Statistics, University of Wisconsin, Madison, WI* (2001), 53706–1388.

[83] Lambert, J. D. *Numerical methods for ordinary differential systems: the initial value problem*. John Wiley & Sons, Inc., 1991.

[84] Larsson, M., and Krühner, P. Affine processes with compact state space. *arXiv preprint arXiv:1706.07579* (2017).

[85] Larsson, M., and Pulido, S. Polynomial diffusions on compact quadric sets. *Stochastic Processes and their Applications 127*, 3 (2017), 901–926.

[86] Lépingle, D., and Mémin, J. Sur l'intégrabilité uniforme des martingales exponentielles. *Z. Wahrsch. Verw. Gebiete 42*, 3 (1978), 175–203.

[87] Lewis, A. L. A simple option formula for general jump-diffusion and other exponential lévy processes. *Available at SSRN 282110* (2001).

[88] Liptser, R., and Shiryayev, A. *Statistics of random processes. I.* Springer-Verlag, New York-Heidelberg, 1977. General theory, Translated by A. B. Aries, Applications of Mathematics, Vol. 5.

[89] Magnus, J. R. On differentiating eigenvalues and eigenvectors. *Econometric Theory 1*, 2 (1985), 179–191.

[90] Magnus, J. R., and Neudecker, H. The elimination matrix: some lemmas and applications. *SIAM Journal on Algebraic Discrete Methods 1*, 4 (1980), 422–449.

[91] Magnus, J. R., and Neudecker, H. Matrix differential calculus with applications in statistics and econometrics. *Wiley series in probability and mathematical statistics* (1988).

[92] Mayerhofer, E., Pfaffel, O., and Stelzer, R. On strong solutions for positive definite jump diffusions. *Stochastic processes and their applications 121*, 9 (2011), 2072–2086.

[93] Mijatović, A., and Urusov, M. On the loss of the semimartingale property at the hitting time of a level. *Journal of Theoretical Probability 28*, 3 (2015), 892–922.

[94] Muravlëv, A. A. Representation of fractal Brownian motion in terms of an infinite-dimensional Ornstein-Uhlenbeck process. *Uspekhi Mat. Nauk 66*, 2(398) (2011), 235–236.

[95] Mytnik, L., and Neuman, E. Sample path properties of Volterra processes. *arXiv preprint arXiv:1101.4969* (2011).

[96] Mytnik, L., and Salisbury, T. S. Uniqueness for Volterra-type stochastic integral equations. *arXiv preprint arXiv:1502.05513* (2015).

[97] Nagumo, M. Über die lage der integralkurven gewöhnlicher differentialgleichungen. *Proceedings of the Physico-Mathematical Society of Japan. 3rd Series 24* (1942), 551–559.

[98] Neudecker, H. Symmetry, 0-1 matrices and jacobians. *Econometric Theory 2* (1986), 157–190.

[99] Pardoux, E., and Protter, P. Stochastic Volterra equations with anticipating coefficients. *Ann. Probab. 18*, 4 (1990), 1635–1655.

[100] Protter, P. Volterra equations driven by semimartingales. *Ann. Probab. 13*, 2 (1985), 519–530.

[101] Revuz, D., and Yor, M. *Continuous martingales and Brownian motion*, third ed., vol. 293 of *Grundlehren der Mathematischen Wissenschaften [Fundamental Principles of Mathematical Sciences]*. Springer-Verlag, Berlin, 1999.

[102] Rockafellar, R. T., and Wets, R. J.-B. *Variational analysis*, vol. 317. Springer Science & Business Media, 2009.

[103] Simon, T. Support theorem for jump processes. *Stochastic Process. Appl. 89*, 1 (2000), 1–30.

[104] Skorokhod, A. V. On the existence and the uniqueness of solutions of stochastic diffusion equations. 1961.

[105] Spreij, P., and Veerman, E. The affine transform formula for affine jump-diffusions with a general closed convex state space. *arXiv preprint arXiv:1005.1099* (2010).

[106] Spreij, P., and Veerman, E. Affine diffusions with non-canonical state space. *Stochastic Analysis and Applications 30*, 4 (2012), 605–641.

[107] Stroock, D. W., and Varadhan, S. R. On the support of diffusion processes with applications to the strong maximum principle. In *Proceedings of the Sixth Berkeley Symposium on Mathematical Statistics and Probability (Univ. California, Berkeley, Calif., 1970/1971)* (1972), vol. 3, pp. 333–359.

[108] Stroock, D. W., and Varadhan, S. S. *Multidimensional diffusion processes*. Springer, 2007.

[109] Tappe, S. Stochastic invariance of closed, convex sets with respect to jumpdiffusions. *Preprint* (2009).

[110] Veerman, E. *Affine Markov processes on a general state space*. PhD thesis, University of Amsterdam, 2011.

[111] Veraar, M. The stochastic fubini theorem revisited. *Stochastics An International Journal of Probability and Stochastic Processes 84*, 4 (2012), 543–551.

[112] Wang, Z. Existence and uniqueness of solutions to stochastic Volterra equations with singular kernels and non-Lipschitz coefficients. *Statistics & Probability Letters 78*, 9 (2008), 1062–1071.

[113] Wong, E., and Zakai, M. On the convergence of ordinary integrals to stochastic integrals. *The Annals of Mathematical Statistics 36*, 5 (1965), 1560–1564.

[114] Xiong, J. *An introduction to stochastic filtering theory*, vol. 18 of *Oxford Graduate Texts in Mathematics*. Oxford University Press, Oxford, 2008.

[115] Yamada, T., and Watanabe, S. On the uniqueness of solutions of stochastic differential equations. *J. Math. Kyoto Univ. 11* (1971), 155–167.

[116] Zähle, H. Space-time regularity of catalytic super-brownian motion. *Mathematische Nachrichten 278*, 7-8 (2005), 942–970.

[117] Zhang, X. Stochastic Volterra equations in Banach spaces and stochastic partial differential equation. *J. Funct. Anal. 258*, 4 (2010), 1361–1425.

Résumé

La présente thèse traite de la théorie des équations stochastiques en dimension finie.
Dans la première partie, nous dérivons des conditions géométriques nécessaires et suffisantes sur les coefficients d'une équation différentielle stochastique pour l'existence d'une solution contrainte à rester dans un domaine fermé, sous de faibles conditions de régularité sur les coefficients.
Dans la seconde partie, nous abordons des problèmes d'existence et d'unicité d'équations de Volterra stochastiques de type convolutif. Ces équations sont en général non-Markoviennes. Nous établissons leur correspondance avec des équations en dimension infinie ce qui nous permet de les approximer par des équations différentielles stochastiques Markoviennes en dimension finie.
Enfin, nous illustrons nos résultats par une application en finance mathématique, à savoir la modélisation de la volatilité rugueuse. En particulier, nous proposons un modèle à volatilité stochastique assurant un bon compromis entre flexibilité et tractabilité.

Abstract

The present thesis deals with the theory of finite dimensional stochastic equations.
In the first part, we derive necessary and sufficient geometric conditions on the coefficients of a stochastic differential equation for the existence of a constrained solution, under weak regularity on the coefficients.
In the second part, we tackle existence and uniqueness problems of stochastic Volterra equations of convolution type. These equations are in general non-Markovian. We establish their correspondence with infinite dimensional equations which allows us to approximate them by finite dimensional stochastic differential equations of Markovian type.
Finally, we illustrate our findings with an application to mathematical finance, namely rough volatility modeling. We design a stochastic volatility model with an appealing trade-off between flexibility and tractability.

Mots Clés

Invariance stochastique, équations de convolutions stochastiques, processus affines, volatilité rugueuse.

Keywords

Stochastic invariance, stochastic convolution equations, affine processes, rough volatility.