On Markovian Approximation Schemes of Jump Processes

A thesis presented to Imperial College London

for the degree of

Doctor of Philosophy in Mathematics

and the

Diploma of Imperial College

by

Francesco Mina

Department of Mathematics

Imperial College

180 Queen's Gate, London SW7 2BZ

MARCH 23, 2015

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

Signed:

Copyright

To my family: Mamma, Papà, Nonna Pina e Adriana.

Abstract

The topic of this thesis is the study of approximation schemes of jump processes whose driving noise is a Lévy process.

In the first part of our work we study properties of the driving noise. We present a novel approximation method for the density of a Lévy process. The scheme makes use of a continuous time Markov chain defined through a careful analysis of the generator. We identify the rate of convergence and carry out a detailed analysis of the error. We also analyse the case of multidimensional Lévy processes in the form of subordinate Brownian motion. We provide a weak scheme to approximate the density that does not rely on discretising the Lévy measure and results in better convergence rates.

The second part of the thesis concerns the analysis of schemes for BSDEs driven by Brownian motion and a Poisson random measure. Such equations appear naturally in hedging problems, stochastic control and they provide a natural probabilistic approach to the solution of certain semi linear PIDEs. While the numerical approximation of the continuous case has been studied in the literature, there has been relatively little progress in the study of such equations with a discontinuous driver. We present a weak Monte Carlo scheme in this setting based on Picard iterations. We discuss its convergence and provide a numerical illustration.

Acknowledgements

This is the result of research carried out under the supervision of Martijn Pistorius to whom I am thankful for providing expert and valuable guidance. I am very grateful for the encouragement he provided during the course of my PhD: his advice has improved the quality of this work immeasurably. I would like to thank the friends from office 617, in particular: William Giles, Eamon McMurray, Lukas Schimmer, Xin Dong, Christopher Warner and Michela Ottobre. Sharing the office, which has been my second home for the past four years, has made the time spent at Imperial a lot more enjoyable. I would also like to thank the friends from the Mathematical Finance section, in particular, Johannes Stolte, Grzegorz Andruszkiewicz, Benoît Pham Dang and Pawel Zaczkowski, for the company at many events at Imperial and abroad. A special thanks to Matilda – for teaching me something learned people have not taught me. I owe a special thanks to all the people I encountered who were generous enough to share their understanding of mathematics with me; in particular I would like to thank Prof Jacobo Pejsachowicz and Prof Simon Salamon for their help and for encouraging me to pursue further studies. Funding from the Engineering and Physical Sciences Research Council (EPSRC) and a Rotary Ambassadorial scholarship has made this research possible. Infine, un ringraziamento speciale alla mia famiglia: mamma, papà, nonna Pina e Adriana. I miei sforzi, il mio lavoro, sono dedicati a voi.

Francesco Mina

Table of contents

Introduction

Axiomatically, mathematics is concerned solely with relations among undefined things. This aspect is well illustrated by the game of chess. It is impossible to "define" chess otherwise than by stating a set of rules. The conventional shape of the pieces may be described to some extent but it will not always be obvious which piece is intended for "king"... The essential thing is to know how the pieces move and act.

(W. Feller, An introduction to Probability Theory and its Applications, Vol 1.)

The topic of this thesis is the study of weak discretisation schemes for continuous time stochastic processes with discontinuous paths. Discretisation of stochastic processes is a recurring theme in probability theory and has proved its fruitfulness in a plethora of applications. Historically, it has been used to deduce results on continuous-time processes from similar results for discrete processes: for instance, Markov processes may be considered as a limit of Markov chains, which are easier to analyse; the law of a stochastic differential equation can be approximated by solving an associated discrete-time problem; even the construction of Brownian motion, arguably the most fundamental of all stochastic processes, can be carried out by taking the limit of random walks in what has become known as the "Donsker invariance principle" – a pinnacle of the modern theory of probability.

Aside from its theoretical interest, the procedure of discretising a process has become increasingly popular due to the rise in interest in Monte-Carlo simulation, which enables one to compute expectations of random variables, often given in the form of complicated functionals of a stochastic process. Set in this background, weak approximation schemes, traditionally defined to approximate the law of the process, as opposed to its sample paths, become particularly relevant and can be converted into easily implementable algorithms since the original process and the sequence defined by the scheme need not live on the same probability space.

Within the class of processes with discontinuous paths, we will focus our attention to the subclass of processes for which the driving noise is given by a Lévy process. Lévy processes arise naturally in the theory of stochastic processes as the continuous time analogue of random walks. They serve as a basic example of both a Markov process and a general semimartingale and constitute the simplest class of stochastic processes that have discontinuous sample paths. Despite their simplicity, they exhibit a rich structure and their distributions display a wide range of statistical properties. It is this flexibility that has determined their use in applications and made them play a central role in several fields of science such as physics (study of turbulence, quantum field theory), engineering (study of networks, queues and dams), economics (time series analysis) and mathematical finance. Some standard references on Lévy processes and infinitely divisible distributions are given by [5, 13, 22, 119] while an overview of application of Lévy processes in option pricing can be found in Cont & Tankov [33].

Mathematical finance has perhaps witnessed the biggest surge of activity due to the ever-growing need for modelling irregular behaviour, jumps and extreme events which are regularly found in financial data. The introduction of jumps in the modelling of asset prices began with Merton [92] and it has led to many tractable and attractive models since these generally give rise to a more realistic description of the distribution of returns and tail behaviour. It is in fact well-known that the classical, diffusion based, models fail to capture in a natural way some commonly observed features in financial data such as heavy tails or asymmetry. Furthermore, introducing jumps in the modelling of these phenomena provides us with a tool to adequately describe sudden, discontinuous price fluctuations, a feature which any diffusion-based stochastic volatility models lack.

Although this provides sufficient support for incorporating discontinuities, it also leads to several problems of a different nature. In particular, the introduction of jumps results in incomplete markets, i.e., markets in which it is not always possible to construct a portfolio which replicates a given contingent claim. A convenient framework for the study of valuation and hedging of contingent claims in an incomplete market setting is that of backward stochastic differential equations (BSDEs). While this concept was originally proposed in the theory of optimal control, BSDEs have proved to be useful in a variety of applications, such as stochastic target problems and problems of valuation and hedging of contingent claims under uncertain volatility, transaction cost, or limited liquidity, or by using recursive utilities or non-linear risk measures. Outside the sphere of applications, another interest of BSDEs springs from their intrinsic connection to certain non-linear partial integro-differential equations. Non-linear Feynman-Kac formulae provide stochastic representations of solutions to related non-linear integro-differential equations in terms of the solutions of such BSDEs. Since analytical solutions to many of the above mentioned problems are typically not available, the evaluation of the solution becomes a numerical task, leading to a need for efficient numerical schemes to solve such stochastic differential equations.

The purpose of this work is to study approximation methods and develop numerical schemes for cases in which the driving noise is given by a Lévy process.

In the first part of our work we study properties of the driving noise. In the second part of our work we focus on developing numerical schemes for FBSDEs driven by a Brownian motion and a Poisson random measure. In particular, we will focus our attention on the case of a forward-backward stochastic difference equation (FBSΔE) and discuss the issue of convergence to the continuous case.

In Chapter 1 we introduce a discretization scheme based on a continuous-time Markov chain for a class of Lévy processes, and we investigate the convergence rate of this scheme for one dimensional distributions. In particular, we will start by examining Lévy processes with Gaussian component and establish the rate of convergence under mild assumptions on the behaviour of the Lévy measure. Rather than defining the chain starting from the characteristic exponent, we choose to work at the level of the generators: in the first section we give a definition of the chain, based on an approximation of the generator of the continuous state space process, and we fix the coefficients of the generator of the chain by matching moments. We then derive integral formulae for the transition probabilities of the processes: to this end, we obtain a spectral decomposition of the generator of the chain, given in terms of the semi-discrete Fourier transform. An accurate comparison of these integral expressions for the transition probabilities allows us to analyse the rate of convergence of the error: we shall see how this rate depends on the behaviour of the Lévy measure in a neighbourhood of the origin, thus making lattice models for jump processes more subtle to implement than the analogous models for diffusions. The remaining part of the chapter is devoted to the analysis of pure jump Lévy processes with paths of both bounded and unbounded variation. We discuss how, in a pure jump setting, it is either not possible or not useful to match the second moment and investigate how the absence of Brownian motion intrinsically reduces the speed of the scheme.

In Chapter 2 we study the problem of approximating the density of a multi-dimensional Lévy process given in the form of subordinated Brownian motion. Subordination is a type of time change introduced by Bochner [17, 18] in the context of harmonic analysis and is particularly useful since it leads to some analytically tractable expression for many quantities of interest. This is especially relevant since many Lévy processes used in applications take this form. After a brief review of the Phillips theorem and some basic facts on subordination, we tackle the problem of the approximation of the density of the Lévy process. Instead of approximating the Lévy process directly, we replace the Brownian noise by a random walk and study the subordinated random walk. We believe an advantage of our method over a direct approximation of the jumps lies in the fact that we not not attempt to approximate the law of the subordinator; this results in faster convergence of the transition kernels and provides an easy way to simulate transition probabilities since the approximating process is a continuous time Markov chain and the computation reduces to the exponentiation of a matrix. The main result of the chapter is to derive the first term in the expansion for the error given by the difference between the transition probabilities and identify the exact constant.

In Chapter 3 we move to the study of backward stochastic differential equations. At the beginning of the chapter, we start by reviewing some classical results in a Brownian setting; in particular, the connection with the solution of a Cauchy problem is presented. We then examine the case of BSDEs where the filtration is generated by a jump process, namely a Wiener process and an independent Poisson random measure. The following part of the chapter is devoted to the study of a weak discretisation scheme: we replace the continuous driving noise by suitably chosen random walks and analyse the associated FBSΔE. This discrete problem differs from the continuous one for the presence of an extra (martingale) term, due to the loss of the predictable representation property in the discretisation procedure. A solution to the BSΔE is found and we briefly discuss the issue of the convergence to the solution of the continuous BSDE. We then propose a numerical scheme using the regression on function bases technique for the computation of the conditional expectation; extensions to the case of a filtration generated by a forward stochastic equation are given.

The recent development of comparison theorems for BSDEs driven by discontinuous processes (see [8]) has made it possible to extend some results on g-expectations (in particular the inverse theorem) to a Lévy setting [117]. Drawing inspiration from the latter, we close the chapter with a numerical illustration where we carry out the computation of some non-linear expectation in a jump setting.

Chapter 1

Lattice approximation of Lévy processes

1.1 Introduction

In this chapter, we look at the problem of approximating a Lévy process by a Markov chain defined on a lattice and to determine the rate of convergence of the error term.[1] More precisely, we determine the rate of convergence of the transition kernel of the chain to that of the Lévy process. Instead of starting from the characteristic exponent, thus exploiting the time and space homogeneity of the process, we choose to work at the level of the generator. This approach suggests that the results allow for generalisation to the more general setting of space and time inhomogeneous Markov processes.

Literature review

The identification of conditions for the weak convergence of processes have received a considerable amount of attention in the literature. Their importance is both theoretical and practical since functionals of stochastic processes often arise in many areas of applications including finance, filtering and statistics. The article [82] discusses some sufficient conditions for the weak convergence of a sequence of semimartingales to a process of diffusion-type. Kurtz & Protter [77] show that, given a suitable sequence $\{(X_n, Y_n)\}$

[1] We thank Aleksandar Mijatović for useful conversations.

of processes converging weakly to (X, Y) (in the Skorokhod topology), the sequence $\{\int X_n \, \mathrm{d}Y_n\}$ converges weakly to $\int X \, \mathrm{d}Y$. In the spirit of the work of Gikhman & Skorokhod [60], Markov chain approximation methods were put forward by Kushner [78], who shows how a certain family of Markov chains, satisfying the so-called "local consistency" conditions, converges in law to a diffusion. The proof of this result relies on showing that a certain sequence of probability measures is tight. This fact is further exploited in the context of stochastic control by Kushner & Dupuis [79], where numerical methods based on Markov chains are refined. Stroock & Varadhan [120] developed a theory of diffusion approximation which makes use of the connection between solution to SDEs and the martingale problem and show a link between rescaled Markov chains and certain diffusion processes. This constituted a change of paradigm since most methods deployed until then to tackle approximation problems were analytical in nature. Platen [112], also constructs a Markov chain to approximate a diffusion process.

Discretization schemes for stochastic processes lie at the heart of computational finance too and their importance has been highlighted in the seminal paper by Cox, Ross & Rubinstein [36], who introduce a no-arbitrage pricing method based on recombining binomial trees. The convergence in law of the discrete-time approximation to the continuous-time Black-Scholes model was proved in [63] and [3]. This led to many attractive pricing algorithms for contingent claims for which closed-form solutions of continuous-time models are not available. Because of the importance and usefulness of discrete-time models, a considerable effort has been made to extend and generalise them, examples of which include [89, 44, 66]. The key issue of the rate of convergence of the discrete option prices to their continuous counterpart was solved by Heston & Zhou [64]. The rate of convergence measures the speed at which the numerical method converges to its limit and it is particularly useful since convergence-enhancing techniques, like the Richardson extrapolation, are no longer useful if the rate of convergence is unknown. They also show that the convergence rate of discrete-time multinomial option prices depends crucially on the smoothness of the option payoff. The latter is slower than generally believed, since often payoff profiles present a discontinuity in the first derivative. However, they proceed with proving that the convergence rate of a binomial model to the price of a derivative whose payoff is twice continuously differentiable and defined on a bounded interval, converges at the rate of $\mathcal{O}(n^{-1})$, with n being the number of time steps in the tree. A detailed study of the convergence of the binomial tree was also carried out by Walsh [124], who obtains sharper results using probabilistic methods. He embeds the tree scheme in the diffusion by means of Skorokhod embedding, thus defining the approximating sequence and the diffusion on the same probability space. The strong rate of convergence and the exact constant are thus determined.

Contribution

A number of authors, [76, 122], have given sufficient conditions for the convergence of a sequence of operator semigroups in terms of their infinitesimal generators. They use operator-theoretic techniques in their proofs, which seem ill-suited for the analysis of the accuracy. Mijatović & Pistorius [93] tackle the problem of pricing barrier options in a one-dimensional Markov model defining a continuous-time Markov chain that closely mimics the dynamic of the underlying. The related problem of determining the rate of convergence of the probability kernel of the discretized model to the probability density function of linear Brownian motion and to the probability density of a Lévy process has been examined by Albanese and Mijatović [2] and Mijatović *et al.* [94] respectively. The discretization scheme proposed in [94] is based on a continuous-time Markov chain whose infinitesimal generator is defined in such a way as to approximate the generator of the Lévy process. Their approach, however, differs from ours in that they restrict the analysis to a subclass of Lévy processes whose marginal densities have derivatives of all orders, namely the processes that have Gaussian component or satisfy the "Orey condition" while we impose weaker conditions on the decay of the characteristic exponent. Furthermore, by matching the moments of the 'small jumps' we obtain faster convergence rates in all cases covered, as shown in Table 1.1, provided that the conditions for the existence of the generator are satisfied.

The scheme we use in the case of a Lévy process with Gaussian component is summarised by Table 1.2 whereas in the case of a pure jump process we shall utilise the scheme in Table 1.3 and in Table 1.4 depending on whether the paths of the process are of bounded variation or not respectively.

	Bounded variation	Unbounded variation
$\sigma > 0$	$\mathcal{O}(h^2)$	$\mathcal{O}(h)$
$\sigma = 0$	$\mathcal{O}(h)$	$\mathcal{O}(h)$

Table 1.1: The table above shows the speed of convergence of the transition kernel of a continuous time Markov chain to that of a Lévy process with characteristic triplet (μ, σ, ν). The headings "Bounded variation" and "Unbounded variation" refer to whether or not the condition $\int_{(1,1)} |x|\, \nu(\mathrm{d}x) < \infty$ is satisfied.The parameter h represents the mesh of the lattice.

$$\nabla_h(x, y) = \frac{1}{2h}\begin{cases} 1 & \text{if } y = x + h, \\ -1 & \text{if } y = x - h, \\ 0 & \text{otherwise.} \end{cases}$$

$$\Delta_h(x, y) = \frac{1}{h^2}\begin{cases} 1 & \text{if } y = x \pm h, \\ -2 & \text{if } y = x, \\ 0 & \text{otherwise.} \end{cases}$$

$$\Lambda_h(x, y) = f(y - x)h, \quad x \neq y.$$

$$\mathcal{L}_h = \tilde{\mu}_h \nabla_h + \frac{\tilde{\sigma}_h^2}{2}\Delta_h + \Lambda_h.$$

Table 1.2: The exhibit describes the generator $\mathcal{L}_h$ of the continuous time Markov chain used in the case of a Lévy process with Gaussian component. The operators above approximate the drift, Brownian and jump part of the generator of the Lévy process. The expression for the parameters $\tilde{\mu}_h, \tilde{\sigma}_h$ are given in Equation (1.2.4) and (1.2.5) respectively. The states x and y lie on the lattice $h\mathbb{Z}$, where h represents the spacing.

We remark that the techniques used in this chapter are well suited to extensions of our work to a more general class of Markov processes since they do not rely essentially on the stationarity and independence of increments that Lévy processes enjoy. This suggests that there is a wider range of validity and that such an approximation could be developed for Markov processes for which the generator is known in closed form, like Feller processes; this, however, is left for future research. In a generator based approach, in the general case, sufficient conditions to guarantee the existence of a process given a generator, such as the ones given in Kolokoltsov [74] for the class of Feller processes, would be needed. For an exposition of the theory of Markov processes with an emphasis on the interplay between stochastic and functional analytic methods, we refer to the recent book by Kolokoltsov [75]. In [75] general results show that convergence of the generators under suitable conditions implies the convergence of semigroups (see for example Theorem

$$\nabla_h^+(x,y) = \frac{1}{h}\begin{cases} 1 & \text{if } y = x+h, \\ -1 & \text{if } y = x, \\ 0 & \text{otherwise.} \end{cases}$$

$$\nabla_h^-(x,y) = \frac{1}{h}\begin{cases} 1 & \text{if } y = x-h, \\ -1 & \text{if } y = x, \\ 0 & \text{otherwise.} \end{cases}$$

$$\Lambda_h(x,y) = f(y-x)h, \quad x \neq y.$$

$$\mathcal{L}_h = \tilde{\mu}_h \nabla_h^{\pm} + \Lambda_h.$$

Table 1.3: The table describes the generator of the continuous time Markov chain used in the case of a pure jump Lévy process with bounded variation paths. The operator $\nabla_h^{\pm}$ approximates the drift part of the generator depending on the sign of μ while Λ_h mimics the jump part of the generator of the Lévy process. The parameter $\tilde{\mu}_h \nabla_h^{\pm}$ is obtained by matching the first moment and it is given in Equation (1.3.4). The states x and y lie on the lattice $h\mathbb{Z}$, where h represents the spacing.

8.1.1 in [75]).

The remainder of this chapter is structured as follows: in section 1.2 we study the case of a Lévy process with Gaussian component: in section 1.2.1 we define the generator of the Markov chain and determine the parameters in such a way that the first two moments are matched; in section 1.2.2 we obtain a spectral representation of the generator of the chain and identify the characteristic exponent of the approximating process; this will be used to obtain an integral expression for the transition kernels. Finally, in section 1.2.4 we prove the error estimates. In section 1.3 we treat the case of a pure jump Lévy process with bounded variation paths: building upon the work in the previous sections, in 1.3.1 we explain the challenges arising from dealing with a pure jump process; in section 1.3.2 we define the generator of the chain and prove convergence estimates. In section 1.4, we examine the case of a pure jump, unbounded variation Lévy process and we prove convergence estimate for the transition probabilities. The proof relies on introducing an auxiliary process with Gaussian component and we are thus able to reduce some of the proof to a case covered in section 1.2. The chapter closes with an appendix where we collect some useful results used throughout the chapter.

$$\nabla_h(x,y) = \frac{1}{2h}\begin{cases} 1 & \text{if } y = x+h, \\ -1 & \text{if } y = x-h, \\ 0 & \text{otherwise.} \end{cases}$$

$$\Delta_h(x,y) = \frac{1}{h^2}\begin{cases} 1 & \text{if } y = x\pm h, \\ -2 & \text{if } y = x, \\ 0 & \text{otherwise.} \end{cases}$$

$$\Lambda_h(x,y) = f^G(y-x)h, \quad x \neq y.$$

$$\mathcal{L}_h^G = \tilde{\mu}_h \nabla_h + \frac{\tilde{\sigma}_h^2}{2}\Delta_h + \Lambda_h$$

Table 1.4: The table describes the generator of the continuous time Markov chain used in the approximation of a pure jump Lévy process with unbounded variation paths. The operator ∇_h approximates the drift part of the generator while Δ_h mimics the Brownian part of the generator of a Lévy process with Gaussian component X^G approximating the original process X. The parameter $\tilde{\mu}_h$ is obtained by matching the first moment and it is given in Equation (1.4.7). The states x and y lie on the lattice $h\mathbb{Z}$, where h represents the spacing.

1.2 Lattice approximation of Lévy processes with Gaussian component

1.2.1 Definition of the generator and construction of the chain

Let $X = \{X_t;\, t \geq 0\}$ be a Lévy process with characteristic triplet (μ, σ, ν) and such that $\nu(\mathrm{d}x) = f(x)\,\mathrm{d}x$. Our (sequence of) processes will be a discrete version of the process X taking values in $h\mathbb{Z}$ for some positive h in a sense that will be made clear later. We proceed to define a continuous time Markov chain (which has stationary and independent increments, i.e., a Lévy process) by defining its generator, which in this simple setting, can be thought of as an (infinite) matrix.

Let $l^2(h\mathbb{Z})$ denote the Hilbert space of maps from $h\mathbb{Z} \to \mathbb{C}$ which are square summable and let us consider the set $\{\delta_y, y \in h\mathbb{Z}\} \subset l^2(h\mathbb{Z})$ where

$$\delta_y(x) = \begin{cases} 1 & \text{if } x = y, \\ 0 & \text{else,} \end{cases}$$

which is a Hilbert basis for $l^2(h\mathbb{Z})$. Define $\nabla_h : l^2(h\mathbb{Z}) \to l^2(h\mathbb{Z})$ and $\Delta_h : l^2(h\mathbb{Z}) \to$

$l^2(h\mathbb{Z})$ in the following way:

$$\begin{aligned}(\nabla_h \phi)(x) &= \frac{\phi(x+h)-\phi(x-h)}{2h},\\ (\Delta_h \phi)(x) &= \frac{\phi(x+h)+\phi(x-h)-2\phi(x)}{h^2},\end{aligned}$$

where $x \in h\mathbb{Z}$. Note how the two operators defined above are respectively the discretisation of the first and second term in the expression for the generator of a Lévy process (1.A.1). The operator Δ_h defines a chain with equal probabilities of jumping one state up or one state down from the current state. Its expression in coordinates with respect to the standard basis $\{\delta_y; y \in h\mathbb{Z}\}$ is the following:

$$\Delta_h(x,y) = \Delta_h(\delta_y)(x) = \frac{1}{h^2}\begin{cases} 1 & \text{if } y = x+h, x-y, \\ -2 & \text{if } y = x, \\ 0 & \text{otherwise.}\end{cases}$$

On the other hand the discrete gradient operator ∇_h as defined above, is not a Markov generator since it has negative elements off the diagonal. It can be expressed in coordinates as follows:

$$\nabla_h(x,y) = \nabla_h(\delta_y)(x) = \frac{1}{2h}\begin{cases} 1 & \text{if } y = x+h, \\ -1 & \text{if } y = x-h, \\ 0 & \text{otherwise.}\end{cases}$$

Define $A \subset l^2(h\mathbb{Z})$ to be the space of compactly supported sequences in $h\mathbb{Z}$ (i.e., sequences which are eventually zero). A is clearly dense in $l^2(h\mathbb{Z})$. We are going to define

$$\Lambda_h : l^2(h\mathbb{Z}) \supseteq A \to l^2(h\mathbb{Z}), \quad h > 0,$$

where A is the domain of Λ_h, in the following way: we start by defining Λ_h on the elements of the Hilbert basis $\{\delta_y, y \in h\mathbb{Z}\}$.

$$\Lambda_h(x,y) = \Lambda_h(\delta_y)(x) = f(y-x)h, \quad x \neq y,\ x \in h\mathbb{Z}, \tag{1.2.1}$$

and

$$\Lambda_h(y,y) = \Lambda_h(\delta_y)(y) = -\sum_{x\neq y, x\in h\mathbb{Z}} \Lambda_h(y,x).$$

Lemma 1.2.1. *Under the conditions of Theorem 1.2.13, $\Lambda_h : A \subset l^2(h\mathbb{Z}) \to l^2(h\mathbb{Z})$ is a well-defined operator.*

The proof follows from an application of Theorem 1.A.6.

Remark 1.2.2. The definition of the jump part of the generator as above shows that $\Lambda(x,y)$ only depends on the difference $x-y$ and this means that the resulting Markov chain will be spatially homogeneous. Furthermore, it will also be time-homogeneous since there is no time dependence. The resulting Markov chain will therefore be a Lévy process.

Remark 1.2.3. The meaning of Equation (1.2.1) becomes clear when we think about the expected number of jumps of our process. Let $X = \{X_t; t \geq 0\}$ be a Lévy process and let $N(\cdot,\omega)$ be the associated Poisson random measure. The latter encompasses all the information about the jumps of the process. The expected number of jumps in $[0,t]$ of size $y-x-h/2 \leq \xi \geq y-x+h/2$, with x,y in $h\mathbb{Z}$ is given by

$$\mathbb{E}\left[N\left([0,t]\times B_{\frac{h}{2}}(y-x)\right)\right] = t\int_{B_{\frac{h}{2}}(y-x)} \nu(\mathrm{d}u) \approx tf(y-x)h, \tag{1.2.2}$$

where $B_{\frac{h}{2}}(y-x)$ denotes a ball of centre $y-x$ and radius $h/2$. In other words, the expected number of jumps per unit time of the Lévy process within the range $[y-x-h/2, y-x+h/2)$ is approximatively equal to with the expected number of jumps of the chain of size $y-x$. Differentiating with respect to t we obtain the rate of going from x to y. One can also think of this procedure as constructing another Lévy process whose jump measure is a sum of delta masses centred on the points of the grid $h\mathbb{Z}$ and with coefficients given by the right-hand side of (1.2.2).

We now define the operator Λ_h on the whole space A. This is done by setting:

$$\Lambda_h(\xi)(x) = \sum_{y\in h\mathbb{Z}} \langle \xi, \delta_y\rangle \Lambda_h(x,y), \quad \xi \in l^2(h\mathbb{Z}),$$

where $\langle\cdot,\cdot\rangle$ is the l^2 inner product. With this definition, it is easy to see that the mapping

$$\mathcal{L}_h = \tilde{\mu}_h \nabla_h + \frac{\tilde{\sigma}_h^2}{2}\Delta_h + \Lambda_h,$$

with $\tilde{\mu}_h, \tilde{\sigma}_h$ real coefficients that depend on h, is a densely defined unbounded operator. It now remains to define the coefficients $\tilde{\mu}_h$ and $\tilde{\sigma}_h$ in order to ensure that $\mathcal{L}_h$ is a generator. By the Lévy -Khintchine formula, the characteristic exponent of the Lévy process X with characteristic triplet (μ, σ, ν) can be written as $\Psi = \Psi_1 + \Psi_2$ where

$$\Psi_1(p) = \mathtt{i}\mu p - \frac{1}{2}\sigma^2 p^2 + \int_{\{|x|\leq 1+\frac{h}{2}\}} (\mathrm{e}^{\mathtt{i}px} - 1 - \mathtt{i}ux)\,\nu(\mathrm{d}x), \tag{1.2.3}$$

and

$$\Psi_2(p) = \int_{\{|x|>1+\frac{h}{2}\}} (\mathrm{e}^{\mathtt{i}px} - 1)\,\nu(\mathrm{d}x).$$

The choice of $1 + \frac{h}{2}$ as a cut-off point to separate the small jumps from the large ones is arbitrary and could be replaced by any real x such that $x = \frac{h}{2}$ mod h. The reason for doing this lies in the fact that while the value of the Riemann integral, unlike its stochastic counterpart, is not affected by the choice of the sampling point for the integrand, its rate of convergence is and it is maximised by choosing the midpoint (see Appendix). Note that Ψ_2 is the characteristic exponent of a compound Poisson process. In the following we set I to be the interval $[-1-h/2, 1+h/2]$ and J to be the set $\{k : |kh| \leq 1, k \neq 0\}$. We shall show in the next section that, with the choice of parameters $\tilde{\mu}_h, \tilde{\sigma}_h$ that we are about to make and under the conditions that we are going to specify, the approximating Markov chain X^h is a compound Poisson process with characteristic exponent:

$$\Psi^h(p) = \tilde{\mu}_h \frac{\sin(hp)}{h} + \tilde{\sigma}_h^2 \frac{\cos(hp) - 1}{h^2} + \sum_{k\neq 0} (\mathrm{e}^{\mathtt{i}khp} - 1) f(kh) h.$$

We can then treat Ψ^h in a similar fashion, namely write it as $\Psi^h = \Psi_1^h + \Psi_2^h$, where

$$\Psi_1^h(p) = \tilde{\mu}_h \frac{\sin(hp)}{h} + \tilde{\sigma}_h^2 \frac{\cos(hp) - 1}{h^2} + \sum_J (\mathrm{e}^{\mathrm{i}khp} - 1) f(kh) h,$$
$$\Psi_2^h(p) = \sum_{h\mathbb{Z}\setminus J} (\mathrm{e}^{\mathrm{i}khp} - 1) f(kh) h.$$

X^h can then be thought of as being the sum of two independent Lévy processes X_1^h and X_2^h, corresponding respectively to Ψ_1^h and Ψ_2^h. If the condition (BV) holds:

$$\int_{[-1,1]} |x| \, \nu(\mathrm{d}x) < \infty,$$

then (1.2.3) can be rewritten as

$$\Psi_1(p) = \mathrm{i}\hat{\mu}p - \frac{1}{2}\hat{\sigma}^2 p^2 + \int_I \left(\mathrm{e}^{\mathrm{i}px} - 1 - \mathrm{i}px + \frac{p^2 x^2}{2} \right) \nu(\mathrm{d}x),$$

where

$$\hat{\mu} = \mu + \int_I x \, \nu(\mathrm{d}x), \qquad \hat{\sigma}^2 = \sigma^2 + \int_I x^2 \, \nu(\mathrm{d}x).$$

and the latter makes sense because ν is a bounded variation Lévy measure. If the condition (UV) holds: $\int_{[-1,1]} |x| \, \nu(\mathrm{d}x) = \infty$, (1.2.3) reads

$$\Psi_1(p) = \mathrm{i}\mu p - \frac{1}{2}\hat{\sigma}^2 p^2 + \int_I \left(\mathrm{e}^{\mathrm{i}px} - 1 - \mathrm{i}px + \frac{p^2 x^2}{2} \right) \nu(\mathrm{d}x),$$

where

$$\hat{\sigma}^2 = \sigma^2 + \int_I x^2 \, \nu(\mathrm{d}x).$$

The key idea is to match the first two moments of the "small jumps" processes; in other words we want, for all $t \in \mathbb{R}_+$

$$\mathbb{E}[X_1(t)] = \mathbb{E}[X_1^h(t)], \qquad \mathbb{E}[X_1(t)^2] = \mathbb{E}[X_1^h(t)^2],$$

or equivalently

$$\mathrm{i}^{-1}\Psi_1'(0+) = \mathrm{i}^{-1}(\Psi_1^h)'(0+), \quad -\Psi_1''(0+) = -(\Psi_1^h)''(0+).$$

It is a matter of simple calculus to verify that the coefficients $\tilde{\mu}_h$, $\tilde{\sigma}_h$ should be set equal to

$$\tilde{\mu}_h = \begin{cases} \mu + \int_I x\,\nu(\mathrm{d}x) - \sum_J khf(kh)h & \text{if (BV)}, \\ \mu - \sum_J khf(kh)h & \text{if (UV)}, \end{cases} \tag{1.2.4}$$

and

$$\tilde{\sigma}_h^2 = \sigma^2 + \int_I x^2\,\nu(\mathrm{d}x) - \sum_J (hk)^2 f(kh)h, \tag{1.2.5}$$

in order to satisfy the previous equations. Since ∇_h is not a Markov generator, we need to check that $\mathcal{L}_h(x, x \pm h) \geq 0$, or, more explicitly:

$$\begin{aligned} \mathcal{L}_h(x, x-h) &= -\frac{\tilde{\mu}_h}{2h} + \frac{\tilde{\sigma}_h^2}{2h^2} + f(-h)h \geq 0, \\ \mathcal{L}_h(x, x+h) &= \frac{\tilde{\mu}_h}{2h} + \frac{\tilde{\sigma}_h^2}{2h^2} + f(h)h \geq 0. \end{aligned} \tag{1.2.6}$$

If (BV) holds, an application of a result on convergence of Riemann integrals, Theorem 1.A.6, shows that it is enough to require

$$\begin{cases} \int_{[-1,1]\setminus[-h,h]} |f(x) + xf'(x)|\,\mathrm{d}x = o\left(1/h\right), \\ \int_{[-1,1]\setminus[-h,h]} |xf(x) + x^2 f'(x)|\,\mathrm{d}x = o\left(1/h\right), \end{cases} \tag{A}$$

in order for

$$\left| \int_I x\,\nu(\mathrm{d}x) - \sum_J khf(kh)h \right| \to 0, \qquad \left| \int_I x^2\,\nu(\mathrm{d}x) - \sum_J (hk)^2 f(kh)h \right| \to 0.$$

The condition

$$-\frac{\sigma^2}{h} \leq \mu \leq \frac{\sigma^2}{h}, \tag{1.2.7}$$

is sufficient for (1.2.6) to be satisfied. This condition, however, will be satisfied for an arbitrary μ and positive σ, provided the lattice space h is small enough. Since we are interested in the behaviour of the discrete model in the limit as h goes to zero, we can assume without loss of generality that (1.2.7) condition holds. This implies that the operator

$\mathcal{L}_h$ is a Markov generator. If (UV) holds, then assume that

$$\begin{cases} \mu - \frac{\sigma^2}{2h} - 2f(-h)h^2 \leq \sum_J khf(kh)h \leq \mu + \frac{\sigma^2}{2h} + 2f(h)h^2, \\ \int_{[-1,1]\setminus[-h,h]} |xf(x) + x^2 f'(x)| \, \mathrm{d}x = o\left(1/h\right). \end{cases} \tag{B}$$

Note that the first condition in (B) is just restating explicitly that $\mathcal{L}_h(x, x \pm h) \geq 0$ and is satisfied by all alpha-stable processes and by any process whose Lévy measure ν is symmetric. The second condition in (B) guarantees that $|\int_I x^2 \, \nu(\mathrm{d}x) - \sum_J (kh)^2 f(kh)h|$ goes to zero. In the remaining part of the chapter we shall assume that (A) or (B) hold, so that it is guaranteed for $\mathcal{L}_h$ to be Markov generator.

1.2.2 Spectral representation and transition probabilities

In this section we discuss an application of the Kolmogorov backward equation, that we briefly review, and find the transition probabilities of the stochastic process that we defined in terms of the generator $\mathcal{L}_h$. More precisely, we want to find $\mathbf{P}_t^h(x,y) = \mathbb{P}(X_t^h = y | X_0^h = x)$, where the process $\{X_t^h; t \geq 0\}$ is the continuous-time Markov-chain corresponding to $\mathcal{L}_h$. The main tool that we are going to use in order to find those is the Kolmogorov backward equation. In the case of a Markov chain with infinite state space, the backward equation is an infinite system of differential equations:

$$p'_{ij}(t) = \sum_{k \in S} \mathcal{L}_h(i,k) p_{kj}(t), \quad p_{ij}(0) = \delta_{ij}, \tag{1.2.8}$$

and the standard results on matrix exponentials no longer apply. In fact, in the finite state-space setting, a solution to (1.2.8) is just given by $(\mathrm{e}^{t\mathcal{L}_h}\delta_i)_j$ where the latter is the standard matrix exponential. A simple way to compute this expression is to diagonalise (when possible) the matrix $\mathcal{L}_h = S^{-1}\Lambda S$ and to observe that $\mathrm{e}^{t\mathcal{L}_h} = S^{-1}\mathrm{e}^{t\Lambda}S$. In the infinite state space case, the generator $\mathcal{L}_h$ is a linear operator (on some linear space) which can be regarded as an "infinite matrix". Then, in our case, the Kolmogorov backward equation reads

$$\begin{cases} \dot{u} = \mathcal{L}_h u, \\ u(0) = \delta_x, \end{cases}$$

for a function $u : [0, \infty) \to \mathcal{D}(\mathcal{L}_h)$, where $\mathcal{D}(\mathcal{L}_h)$ represents the domain of the operator $\mathcal{L}_h$. The latter is an operator-valued differential equation in that $u(t_0)$ for any t_0 in $[0, \infty)$ is an element of a linear space, possibly infinite dimensional [2]. It is possible to show that the solution is again given by "$u(t) = \mathrm{e}^{t\mathcal{L}_h}\delta_x$" provided we make sense of $\mathrm{e}^{t\mathcal{L}_h}$, especially in the case of $\mathcal{L}_h$ unbounded. However, by using functional calculus (see chapters VII and VIII in Reed & Simons[114]), this turns out to be possible and, much as before, we shall see that in order to get a workable expressions we will need to "diagonalise" the operator $\mathcal{L}_h$. The analysis carries over in a surprisingly similar way. Then

$$\mathbb{P}(Y_t^h = y | Y_0^h = \cdot) = \mathrm{e}^{t\mathcal{L}_h}\delta_y(\cdot),$$

where the latter is an equality between two square summable sequences and the right hand side is to be interpreted as the operator $\mathrm{e}^{t\mathcal{L}_h}$ acting on δ_y. We now set out to do this.

Consider the family of functions defined on the interval $[-\frac{\pi}{h}, \frac{\pi}{h}]$ in the following way:

$$g_n : p \mapsto \sqrt{\frac{h}{2\pi}} \exp\left(-\mathrm{i}nhp\right), \quad n \in \mathbb{Z}.$$

The functions $\{g_n; n \in \mathbb{Z}\}$ form a countable orthonormal basis of the Hilbert space $L^2([-\frac{\pi}{h}, \frac{\pi}{h}])$ with respect to the Lebesgue measure. We will now define the semi-discrete Fourier transform which is an isometry $\mathcal{F}_h : l^2(h\mathbb{Z}) \to L^2([-\frac{\pi}{h}, \frac{\pi}{h}])$. Let f be an element of $l^2(h\mathbb{Z})$ and let $f(hn)$ denote the n^{th} element of the sequence, where $n \in \mathbb{Z}$. Then[3]

$$\mathcal{F}_h(f)(p) := \sum_{n\in\mathbb{Z}} f(hn) g_n(p), \qquad p \in \left[-\frac{\pi}{h}, \frac{\pi}{h}\right].$$

The transformation $\mathcal{F}_h$ maps sequences into periodic functions on $\mathbb{R}$ with period $\frac{2\pi}{h}$. It follows from the definitions of the respective inner products on $l^2(h\mathbb{Z})$ and $L^2(\left[-\frac{\pi}{h}, \frac{\pi}{h}\right])$ that $\mathcal{F}_h$ is a unitary transformation. It is a well-known fact that any function ϕ in $L^2(\left[-\frac{\pi}{h}, \frac{\pi}{h}\right])$

[2] In this case, $u(t_0)$ will be a square summable sequence specifying the transition probabilities from time zero to t_0.

[3] The following series converges in L^2 by using, for instance, the classical result (Riesz-Fisher): Let $\{e_i\}$ be a countable orthonormal basis in a Hilbert space H. The series $\sum_1^n \alpha_i e_i$ converges in H if and only if the sequence $\{\alpha_i\}$ is in l^2 .

can be expanded uniquely by a Fourier series:

$$\phi = \sum_{n\in\mathbb{Z}} \langle \phi, g_n \rangle g_n,$$

where $\langle \cdot, \cdot \rangle$ is the inner product in L^2 and $\langle \phi, g_n \rangle$ in turn coincides with $\hat{\phi}(n)$, the Fourier transform of ϕ evaluated at n. The series converges to ϕ in the L^2-norm and not necessarily pointwise[4]. This allows us to conclude that the inverse transform $\mathcal{F}_h^{-1}$ can be expressed as

$$\mathcal{F}_h^{-1}(\phi)(hn) = \langle \phi, g_n \rangle,$$

for any ϕ in $L^2\left(\left[-\frac{\pi}{h}, \frac{\pi}{h}\right]\right)$ and all points hn in $h\mathbb{Z}$. The following theorems are important to what follows; their proofs are inspired from the continuous state space case.

Theorem 1.2.4. *Let X_t^h be a Lévy process with state space $h\mathbb{Z}$ and characteristic exponent η. Let $f \in l^2(h\mathbb{Z})$ and $x \in h\mathbb{Z}$. Then the associated semigroup is given by:*

$$\mathbb{E}[f(X_t^h + x)] = \sqrt{\frac{h}{2\pi}} \int_{-\frac{\pi}{h}}^{\frac{\pi}{h}} e^{\mathtt{i}xp}\, e^{t\eta(p)}\, (\mathcal{F}_h f)(p)\, \mathrm{d}p. \tag{1.2.9}$$

Proof. We use the invertibility of the semi-discrete Fourier transform to write any $f \in l^2(h\mathbb{Z})$ as $f(hn) = \mathcal{F}_h^{-1}(\mathcal{F}_h f)(hn)$ for $n \in \mathbb{Z}$. Using the explicit formula for $\mathcal{F}_h^{-1}$ within the expectation sign, we obtain, for all $t \geq 0$, , $x \in h\mathbb{Z}$,

$$\mathbb{E}[f(X_t^h + x)] = \mathbb{E}\left[\sqrt{\frac{h}{2\pi}} \int_{-\frac{\pi}{h}}^{\frac{\pi}{h}} \mathrm{e}^{\mathtt{i}(X_t^h + x)p}\, (\mathcal{F}_h f)(p)\, \mathrm{d}p\right].$$

Moreover:

$$\mathbb{E}\left[\sqrt{\frac{h}{2\pi}} \int_{-\frac{\pi}{h}}^{\frac{\pi}{h}} \left|\mathrm{e}^{\mathtt{i}(X_t^h + x)p}\, (\mathcal{F}_h f)(p)\right|\, \mathrm{d}p\right] \leq \int_{-\frac{\pi}{h}}^{\frac{\pi}{h}} |(\mathcal{F}_h f)(p)|\, \mathrm{d}p < \infty,$$

where the last inequality follows from the fact that $L^2([-\frac{\pi}{h}, \frac{\pi}{h}])$ and is contained in $L^1([-\frac{\pi}{h}, \frac{\pi}{h}])$, since the interval $[-\frac{\pi}{h}, \frac{\pi}{h}]$ has finite Lebesgue measure. This entitles us to apply Fubini's

[4] However, note that Carleson theorem ensures convergence almost everywhere.

theorem and obtain:

$$\begin{aligned}\mathbb{E}[f(X_t^h+x)] &= \sqrt{\frac{h}{2\pi}}\int_{-\frac{\pi}{h}}^{\frac{\pi}{h}} \mathrm{e}^{\mathrm{i}xp}\,\mathbb{E}[\mathrm{e}^{\mathrm{i}X_t^h p}]\,(\mathcal{F}_h f)(p)\,\mathrm{d}p\\ &= \sqrt{\frac{h}{2\pi}}\int_{-\frac{\pi}{h}}^{\frac{\pi}{h}} \mathrm{e}^{\mathrm{i}xp}\,\mathrm{e}^{t\eta(p)}\,(\mathcal{F}_h f)(p)\,\mathrm{d}p.\end{aligned}$$

□

Theorem 1.2.5. *Let X_t^h be a Lévy process with state space $h\mathbb{Z}$ and characteristic exponent η. Let A be its infinitesimal generator, $f\in\mathcal{D}(A)\subset l^2(h\mathbb{Z})$ and $x\in h\mathbb{Z}$. Then the following holds:*

$$(Af)(x) = \sqrt{\frac{h}{2\pi}}\int_{-\frac{\pi}{h}}^{\frac{\pi}{h}} e^{\mathrm{i}xp}\,\eta(p)\,(\mathcal{F}_h f)(p)\,\mathrm{d}p.$$

Proof. By result (1.2.9), we have

$$\begin{aligned}(Af)(x) &= \lim_{t\downarrow 0}\frac{\mathbb{E}[f(X_t^h+x)]-f(x)}{t}\\ &= \lim_{t\downarrow 0}\sqrt{\frac{h}{2\pi}}\int_{-\frac{\pi}{h}}^{\frac{\pi}{h}} \mathrm{e}^{\mathrm{i}xp}\,\frac{\mathrm{e}^{t\eta(p)}-1}{t}\,(\mathcal{F}_h f)(p)\,\mathrm{d}p.\end{aligned}$$

Using $\lim_{t\downarrow 0}\frac{\mathrm{e}^{at}-1}{t}=a$, we have that the integrand converges point wise to $\mathrm{e}^{\mathrm{i}xp}\,\eta(p)\,(\mathcal{F}_h f)(p)$. Furthermore, by the mean value theorem

$$\left|\mathrm{e}^{\mathrm{i}xp}\,\frac{\mathrm{e}^{t\eta(p)}-1}{t}\,(\mathcal{F}_h f)(p)\right| \le |\eta(p)(\mathcal{F}_h f)(p)|, \qquad \forall p\in\mathbb{R},$$

and since η is continuous at every p we have that the right hand side is then bounded on the compact interval $[-\frac{\pi}{h},\frac{\pi}{h}]$ by an element in L^1. A straightforward application of the dominated convergence theorem yields the result we set out to prove. □

The result of Theorem 1.2.5 can be written in convenient shorthand form

$$Af = \mathcal{F}_h^{-1}(\eta\,\mathcal{F}_h f), \qquad \mathcal{F}_h Af = \eta\,\mathcal{F}_h f.$$

Furthermore by taking $f\in\mathrm{Im}(\mathcal{F}_h^{-1})$ i.e. $f=\mathcal{F}_h^{-1}\phi$, for some $\phi\in l^2(h\mathbb{Z})$, (1.2.10) becomes the so-called "spectral representation" of the operator A:

$$\mathcal{F}_h A \mathcal{F}_h^{-1} \phi = \eta \phi.$$

Theorem 1.2.5 says that for any Lévy process defined on a lattice the operator $\mathcal{F}_h A \mathcal{F}_h^{-1}$ is a "multiplication operator", in the sense that it simply multiplies the element it acts upon by a function of real variable which is its characteristic exponent. Incidentally, note the similarity of the previous two results to the next well known theorem, which is of great importance in the analytic study of Lévy processes and their generalisations.

Theorem 1.2.6. *Let X be a Lévy process with Lévy triplet (γ, σ, ν) and characteristic exponent Ψ. Let (P_t) be the associated (Feller) semigroup and let A be its infinitesimal generator.*

1) For each $t \geq 0$, $f \in S(\mathbb{R}^d)$, $x \in \mathbb{R}^d$,

$$(P_t f)(x) = (2\pi)^{-d/2} \int_{\mathbb{R}^d} e^{\mathrm{i}\langle u,x\rangle}\, e^{t\Psi(u)} \hat{f}(u)\,\mathrm{d}u,$$

where $S(\mathbb{R}^d)$ is the Schwartz space of rapidly decreasing functions on $\mathbb{R}^d$ and $\hat{f}$ is the Fourier transform of the function f.

2) For each $f \in S(\mathbb{R}^d)$, $x \in \mathbb{R}^d$,

$$(Af)(x) = (2\pi)^{-d/2} \int_{\mathbb{R}^d} e^{\mathrm{i}\langle u,x\rangle}\, \Psi(u) \hat{f}(u)\,\mathrm{d}u.$$

For a proof of the latter, the reader is referred to [5]. The previous result can be rephrased in the language of pseudo-differential operators and says that P_t and A are pseudo-differential operators with symbol $\mathrm{e}^{t\Psi}$ and Ψ respectively. We can now proceed to find a spectral representation of the generator $\mathcal{L}_h$. This has a twofold purpose since, by virtue of the backward equation and Theorem 1.2.5, will allow us to find the transition probabilities of the process and to identify its characteristic function.

Let a be a sequence in $l^2(h\mathbb{Z})$. Then, exploiting the relation $g_{n+1}(p) = \mathrm{e}^{-\mathtt{i}hp} g_n(p)$,

$$\begin{aligned}
\mathcal{F}_h \triangle_h(a)(p) &= \sum_{n\in\mathbb{Z}} \frac{a(h(n+1)) + a(h(n-1) - 2a(hn))}{h^2} g_n(p) \\
&= \frac{1}{h^2}\left(\mathrm{e}^{-\mathtt{i}hp} + \mathrm{e}^{\mathtt{i}hp} - 2\right) \sum_{n\in\mathbb{Z}} a(hn) g_n(p) \\
&= \frac{2(\cos(hp) - 1)}{h^2} \mathcal{F}_h(a)(p).
\end{aligned}$$

So that if we take $f = \mathcal{F}_h^{-1}(\phi)$ for some $\phi \in L^2$, we get

$$\mathcal{F}_h \triangle_h \mathcal{F}_h^{-1}(\phi)(p) = \frac{2(\cos(hp) - 1)}{h^2}\phi(p).$$

A similar calculation shows that $\mathcal{F}_h \nabla_h(a)(p) = \mathtt{i}\frac{\sin(hp)}{h}\mathcal{F}_h(a)(p)$. Let us now focus on the jump part of the generator, Λ_h. Denoting by $y = hm$ an element of the grid $h\mathbb{Z}$ and by δ_y the sequence in $l^2(h\mathbb{Z})$ that has one in the m^{th} position and zero everywhere else, we obtain:

$$\begin{aligned}
\mathcal{F}_h \Lambda_h(\delta_y)(p) &= \sum_{n\in\mathbb{Z}} \Lambda_h(\delta_y)(hn) g_n(p) \\
&= \sum_{n\neq m} \Lambda_h(hn, hm) g_n(p) + \Lambda_h(hm, hm) g_m(p) \\
&= h\sum_{k=1}^{\infty} f(kh) g_{m-k} + h\sum_{k=-\infty}^{-1} f(kh) g_{m-k} - h\sum_{k\neq 0} f(kh) g_m \\
&= h\sum_{k\neq 0} f(kh)\mathrm{e}^{\mathtt{i}khp} g_m - h\sum_{k\neq 0} f(kh) g_m \\
&= h\sum_{k\neq 0} f(kh)(\mathrm{e}^{\mathtt{i}khp} - 1)\mathcal{F}_h(\delta_y)(p).
\end{aligned}$$

It is then possible to show that, with a similar procedure, for $a \in A$ we obtain:

$$\mathcal{F}_h \Lambda_h(a)(p) = h\sum_{k\neq 0}(\mathrm{e}^{\mathtt{i}khp} - 1) f(kh)\mathcal{F}_h(a)(p).$$

Take $\mathcal{F}_h$ to be the semi-discrete Fourier transform. We have the following "spectral decomposition":

$$\mathcal{F}_h \mathcal{L}_h \mathcal{F}_h^{-1} \Phi = \widetilde{\mathcal{L}_h} \Phi,$$

where $\widetilde{\mathcal{L}_h}$ is a "multiplication operator" and $\Phi \in \mathcal{F}_h(A)$. $\widetilde{\mathcal{L}_h}$ is a "multiplication operator" in the sense that it simply multiplies the element it acts upon by a function of a real variable; in other words:

$$\widetilde{\mathcal{L}_h} \Phi = \Psi^h \Phi,$$

where

$$\Psi^h(p) = \mathrm{i}\tilde{\mu}_h \frac{\sin(hp)}{h} + \tilde{\sigma}_h^2 \frac{\cos(hp) - 1}{h^2} + \sum_{k \neq 0} (\mathrm{e}^{\mathrm{i}khp} - 1) f(kh) h. \tag{1.2.10}$$

Theorem 1.2.5 guarantees that (1.2.10) is in fact the characteristic exponent of X^h. A closer look at this expression tells us that the last term in the right hand side of (1.2.10) is nothing but the discretization of $\int_{\mathbb{R}} (\mathrm{e}^{\mathrm{i}ux} - 1)\, \nu(\mathrm{d}x)$, the characteristic exponent of a pure jump, bounded variation Lévy process (in particular, a compound Poisson process). Its form comes as no surprise as X^h is forced to live on the lattice $h\mathbb{Z}$ and hence its Lévy measure must take the form of a Dirac comb, turning the integral part of the Lévy-Khintchine formula into an infinite sum.

Remark 1.2.7. There are criteria that allow one to establish when certain linear operators admit a spectral decomposition i.e. can be modelled by multiplication operators, which are as simple as one can hope to find. This is always true, for instance, for the class of bounded, self-adjoint operators. Spectral theorems for self-adjoint operators may take several equivalent forms [5]; however, we did not invoke such results but rather found an explicit decomposition for our case.

More importantly, this enables us to define objects of the form $\psi(\mathcal{L}_h)$, where $\psi(\cdot)$ is

[5] For the sake of completeness, we give one possible formulation: For each self-adjoint operator T acting in a Hilbert space H, there exists a unitary operator, making an isometrically isomorphic mapping of the Hilbert space H onto the space $L^2(M, \mu)$, where the operator T is represented as a multiplication operator.

an entire function[6]. For any elements $x, y \in h\mathbb{Z}$, define

$$\mathbf{P}_t^h(x,y) := \mathbb{P}(X_t^h = y | X_0^h = x) = \left(\mathrm{e}^{t\mathcal{L}_h}(\delta_y)\right)(x).$$

The transition probabilities can then be obtained in the following way:

$$\begin{aligned} \mathbf{P}_t^h(x,y) &= \mathcal{F}_h^{-1}\mathrm{e}^{t\tilde{\mathcal{L}}_h}\mathcal{F}_h(\delta_y)(x) = \textstyle\int_{-\frac{\pi}{h}}^{\frac{\pi}{h}} \mathrm{e}^{t\Psi^h(p)}\, g_n(p)\overline{g_m}(p)\,\mathrm{d}p \\ &= \tfrac{h}{2\pi}\textstyle\int_{-\frac{\pi}{h}}^{\frac{\pi}{h}} \mathrm{e}^{t\Psi^h(p)}\,\mathrm{e}^{\mathrm{i}p(x-y)}\,\mathrm{d}p. \end{aligned} \tag{1.2.11}$$

Formula (1.2.11) has the property that the time variable t and the space variable x, y feature in an separated way.

Remark 1.2.8. The transition probabilities could be determined without invoking spectral theory if we were to start by defining the characteristic function. It is a well known fact that the characteristic function uniquely determines the distribution. This distribution can be computed by the so-called "inversion formula":

Theorem 1.2.9. *Let μ be a probability measure on $\mathbb{R}$ and let $\phi(t) = \int_{\mathbb{R}} e^{\mathtt{i}tx}\mu(\mathrm{d}x)$ be its characteristic function. If $a < b$, then:*

$$\lim_{T\to\infty}\frac{1}{2\pi}\int_{-T}^{T}\frac{e^{-\mathtt{i}ta}-e^{-\mathtt{i}tb}}{\mathtt{i}t}\phi(t)\,\mathrm{d}t = \mu(a,b) + \frac{1}{2}\mu(\{a,b\}).$$

1.2.3 Transition density

In order to define the transition density of our process Y let us take a step back and look at the problem from the point of view of Markov processes. For a Markov process taking values in $\mathbb{R}^d$, we can define its transition semigroup $T_{s,t}$ on a suitable space of measurable functions from $\mathbb{R}^d$ to $\mathbb{R}$ by

$$T_{s,t}f(x) = \mathbb{E}[f(X_t)|X_s = x].$$

[6]A function is said to be entire when it is complex valued and holomorphic over the whole complex plane.

In the same spirit it is possible to define the transition probabilities, also known as transition kernel or transition function:

$$P_{s,t}(x, A) = \mathbb{P}(X_t \in A | X_s = x).$$

The Markov property yields the following relation between transition functions, also known as Chapman-Kolmogorov equation:

$$P_{s,u}(x, A) = \int_{\mathbb{R}^d} P_{s,t}(x, \mathrm{d}y) P_{t,u}(y, A).$$

By the properties of conditional probability, each $P_{s,t}(x, \cdot)$ is a probability measure on $\mathcal{B}(\mathbb{R}^d)$. A transition function is said to be spatially and time homogeneous if, respectively:

$$P_{s,t}(x, A) = P_{s,t}(0, A - x) \quad \text{and} \quad P_{s,t}(x, A) = P_{0,t-s}(x, A).$$

It is easily seen that independence of increments yields spatial homogeneity while stationarity gives time homogeneity. Putting these two together we conclude that Lévy processes satisfy:

$$P_{s,t}(x, A) = P_{0,t-s}(0, A - x)\,. \tag{1.2.12}$$

In fact, Lévy processes are completely characterised by this condition, being the only Markov processes which are homogeneous in space and time[7]. It is now possible to define the transition density:

Definition 1.2.10. A Markov process is said to have a transition density if for each $x \in \mathbb{R}^d$, $0 \leq s \leq t < \infty$, there exists a measurable function $y \to \rho_{s,t}(x, y)$ such that

$$P_{s,t}(x, A) = \int_A \rho_{s,t}(x, y)\, \mathrm{d}y.$$

An application of (1.2.12) shows that any Lévy process $\{X_t; t \geq 0\}$ has a transition

[7] [119] offers a proof of this result (Theorem 10.5)

density if and only if $\mathbb{P}^{X_t}$ has a density f_t for each $t \geq 0$ and

$$\rho_{s,t}(x, y) = f_{t-s}(y - x), \tag{1.2.13}$$

for each $0 \leq s \leq t < \infty$, $x, y \in \mathbb{R}^d$. However, X_t does not always have a density: for instance, if X is a compound Poisson process (with parameter λ), we have

$$\mathbb{P}(X_t = 0) \geq \mathbb{P}(N_t = 0) = \mathrm{e}^{-\lambda t} > 0,$$

where N is the associated Poisson process; it follows that $\mathbb{P}^{X_t}$ has an atom at zero for all t. In general, given a Lévy process (on $\mathbb{R}^d$) generated by (γ, Σ, ν), it is enough to require that $\Sigma \neq 0$ or $\nu(\mathbb{R}^d) = \infty$ to make $\mathbb{P}^{X_t}$ continuous[8]; however, in the case $\Sigma = 0$, conditions for absolute continuity are a hard problem. Sufficient conditions are given in Tucker [123] and Orey [102]. Here we present the so-called Orey condition:

Theorem 1.2.11. *Let X be a real-valued Lévy process with characteristic triplet (γ, σ, ν). If there exists $\beta \in (0, 2)$ satisfying*

$$\liminf_{r \to 0} r^{-\beta} \int_{-r}^{r} |x|^2 \, \nu(\mathrm{d}x) > 0,$$

then X_t has a $\mathcal{C}^\infty$ density $p_t(\cdot)$ for every t.

In other words, continuity properties of the cumulative distribution function depend on the behaviour of ν near the origin. For a proof of this we refer to Orey [102] or Sato [119]. This condition however is sufficient but not necessary. To see this, consider (in the one dimensional case) the Lévy process X with triplet (γ, σ, ν), with $\sigma \neq 0$. The characteristic function of X_t is given by

$$\phi_t(u) = \mathrm{e}^{t\Psi(u)},$$

[8] A measure ρ on $\mathbb{R}^d$ is said to be continuous if $\rho(\{x\}) = 0$; a classic example of a measure which is continuous but not absolutely continuous is given by the measure induced by the Cantor function, which is mutually singular with respect to the Lebesgue measure.

where

$$\Psi(u) = \mathtt{i}\mu u - \frac{1}{2}\sigma^2 u^2 + \int_{\mathbb{R}} (\mathrm{e}^{\mathtt{i}ux} - 1 - \mathtt{i}ux\mathbb{1}_{[-1,1]}(x))\,\nu(\mathrm{d}x).$$

It is easy to check the integrability of ϕ:

$$\int_{\mathbb{R}} |\phi(u)|\,\mathrm{d}u \leq \int_{\mathbb{R}} \mathrm{e}^{-\frac{1}{2}\sigma^2 u^2}\,\mathrm{d}u < \infty,$$

since $|\exp(\mathtt{i}\gamma u + \int_{\mathbb{R}} \mathrm{e}^{\mathtt{i}ux} - 1 - \mathtt{i}ux\mathbb{1}_{[-1,1]}(x))\,\nu(\mathrm{d}x))| < 1$ and hence X_t admits a density that is uniformly continuous [9] and given by

$$f_1(x) = (2\pi)^{-1} \int_{\mathbb{R}} \mathrm{e}^{-\mathtt{i}ux} \mathrm{e}^{t\Psi(u)}\,\mathrm{d}u.$$

Remark 1.2.12. The example just considered shows that the decay at the tails makes the characteristic function integrable and this enables us to obtain a density. A similar calculation explains why Lévy processes with a Gaussian component admit a density, no matter what the Lévy measure does.

By (1.2.13), X_t admits a transition density $\rho_t(x, y)$ which is given by $f_t(y-x)$, which is:

$$\rho_t(x, y) = \frac{1}{2\pi} \int_{\mathbb{R}} \mathrm{e}^{t\Psi(p)}\,\mathrm{e}^{\mathtt{i}p(x-y)}\,\mathrm{d}p. \tag{1.2.14}$$

1.2.4 The convergence

In this section we examine the convergence of the discrete probability kernel in $\mathbf{P}_t^h$ to the conditional probability density function $\rho_t(x, y)$ of the Lévy process. It is clear that when trying to estimate

$$D_{x,y}(h) = 2\pi \left| \rho_t(x, y) - \frac{1}{h}\mathbf{P}_t^h(x, y) \right|, \tag{1.2.15}$$

[9] Let Y be a random variable and let $\phi : \mathbb{R} \to \mathbb{C}$ be its characteristic function. If $\phi \in L^1(\mathbb{R})$ i.e. $\int_{\mathbb{R}} |\phi(\xi)|\,\mathrm{d}\xi < \infty$, then Y admits a density f which is uniformly continuous and given by:

$$f(x) = (2\pi)^{-1} \int_{\mathbb{R}} \mathrm{e}^{-\mathtt{i}x\xi} \phi(\xi)\,\mathrm{d}\xi.$$

The integrability of the characteristic function is a sufficient but not necessary condition for the existence of a density: the uniform distribution (on, say, $(-1, 1)$) admits a density but its characteristic function is not Lebesgue integrable.

where

$$\rho_t(x,y) = \frac{1}{2\pi}\int_{\mathbb{R}} \mathrm{e}^{t\Psi(p)}\,\mathrm{e}^{\mathtt{i}p(x-y)}\,\mathrm{d}p, \quad \mathbf{P}_t^h(x,y) = \frac{h}{2\pi}\int_{-\frac{\pi}{h}}^{\frac{\pi}{h}} \mathrm{e}^{t\Psi^h(p)}\,\mathrm{e}^{\mathtt{i}p(x-y)}\,\mathrm{d}p,$$

the fundamental quantity is the difference between the two characteristic exponents Ψ^h and Ψ. This can be seen by writing:

$$\left|\rho_t(x,y) - \frac{1}{h}\mathbf{P}_t^h(x,y)\right| \le \int_{[-\frac{\pi}{h},\frac{\pi}{h}]} |\mathrm{e}^{t\Psi(p)}|\,|\mathrm{e}^{t(\Psi^h(p)-\Psi(p))}-1|\,\mathrm{d}p + \int_{\mathbb{R}\setminus[-\frac{\pi}{h},\frac{\pi}{h}]} |\mathrm{e}^{t\Psi(p)}|\,\mathrm{d}p,$$

and we can take this a little further by writing $|\Psi^h - \Psi| = |\Psi_1^h - \Psi_1 + \Psi_2^h - \Psi_2|$. Spelling out the latter a bit more, we note that we can write:

$$\begin{aligned}
\Psi_1^h(p) &= \mathtt{i}\tilde{\mu}_h\frac{\sin(hp)}{h} + \tilde{\sigma}_h^2\frac{\cos(hp)-1}{h^2} + \sum_J(\mathrm{e}^{\mathtt{i}khp}-1)f(kh)h \\
\Psi_1(p) &= \mathtt{i}\left(\tilde{\mu}_h + \sum_J khf(kh)h\right)p - \frac{1}{2}\left(\tilde{\sigma}_h^2 + \sum_J (hk)^2 f(kh)h\right)p^2 \\
&\quad + \int_I \left(\mathrm{e}^{\mathtt{i}px} - 1 - \mathtt{i}px + \frac{p^2x^2}{2}\right)\nu(\mathrm{d}x),
\end{aligned}$$

which in turn can be rearranged in the form:

$$\begin{aligned}
|\Psi_1^h(p) - \Psi_1(p)| \;\le\;& |\tilde{\mu}_h|\left|\frac{\sin(hp)}{h} - p\right| + \tilde{\sigma}_h^2\left|\frac{\cos(hp)-1}{h^2} + \frac{p^2}{2}\right| \\
+\;& \left|\int_I \left(\mathrm{e}^{\mathtt{i}px} - 1 - \mathtt{i}px + \frac{p^2x^2}{2}\right)\nu(\mathrm{d}x)\right. \\
-\;& \left.\sum_J \left(\mathrm{e}^{\mathtt{i}khp} - 1 - \mathtt{i}pkh - \frac{p^2(kh)^2}{2}\right)f(kh)h\right|,
\end{aligned}$$

the inequality simply being an application of the triangle inequality. Note how the expression within the last modulus sign should be small since the sum is nothing but a discretization of the integral. The same can be said for:

$$|\Psi_2(p) - \Psi_2^h(p)| = \left|\int_{\mathbb{R}\setminus I}(\mathrm{e}^{\mathtt{i}px}-1)\,\nu(\mathrm{d}x) - \sum_{h\mathbb{Z}\setminus J}(\mathrm{e}^{\mathtt{i}pkh}-1)f(kh)h\right|.$$

Define the quantities:

$$\varphi_h(p) = \frac{\cos(hp) - 1}{h^2} + \frac{p^2}{2}, \qquad \rho_h(p) = \frac{\sin(hp)}{h} - p,$$

and

$$\alpha_h(p) = \sum_J \left(\mathrm{e}^{\mathrm{i}khp} - 1 - \mathrm{i}khp + \frac{(khp)^2}{2} \right) f(kh)h - \int_I \left(\mathrm{e}^{\mathrm{i}px} - 1 - \mathrm{i}px + \frac{(px)^2}{2} \right) f(x)\,\mathrm{d}x,$$
$$\beta_h(p) = \sum_{h\mathbb{Z}\setminus J} (\mathrm{e}^{\mathrm{i}khp} - 1) f(kh)h - \int_{\mathbb{R}\setminus I} (\mathrm{e}^{\mathrm{i}px} - 1) f(x)\,\mathrm{d}x.$$

The strategy to find an upper bound of (1.2.15) will be to choose a positive function $K(h)$ which goes to infinity but is bounded above by π/h and to express the aforementioned difference as:

$$\int_{-K(h)}^{K(h)} |\mathrm{e}^{t\Psi(p)}|\,|\mathrm{e}^{t(\Psi^h(p)-\Psi(p))} - 1|\,\mathrm{d}p$$
$$+ 2\int_{K(h)}^{\frac{\pi}{h}} |\mathrm{e}^{t\Psi^h(p)}|\,\mathrm{d}p + 2\int_{K(h)}^{\infty} |\mathrm{e}^{t\Psi(p)}|\,\mathrm{d}p. \quad (1.2.16)$$

The reason for doing this is that the functions φ_h and ρ_h for a fixed value of h, can take large values as p approaches π/h. This can be seen for the function φ_h by developing the cosine with its Taylor polynomial:

$$\frac{\cos(hp) - 1}{h^2} = -\frac{p^2}{2} + \cos(\xi)\frac{h^2p^4}{4!},$$

for some $0 < \xi < hp$. Hence

$$\left| \frac{\cos(hp) - 1}{h^2} + \frac{p^2}{2} \right| \leq \frac{h^2p^4}{4!},$$

valid for all p in $\mathbb{R}$. As $p \to \pi/h$, the right hand side converges to $\frac{\pi^4}{4!h^2}$ and this in turns goes to infinity as h goes to zero. From this, emerges the need to find a function $K(h)$ such that the quantities φ_h, ρ_h are uniformly arbitrarily small on the set $[-K(h), K(h)]$ i.e. they converge to zero uniformly on the interval above. This turns out to be possible as long as $K(h)$ is chosen such that $\lim_{h\to 0} K^4(h)h^2 = 0$ or, equivalently, any

$K(h) = o(1/\sqrt{h})$ as $h \to 0$ will satisfy this requirement. On the other hand we want $K(h)$ to go to infinity (albeit, as we said, slower than π/h) since this makes the last two summands in Eqn (1.2.23) integrals of two integrable functions on sets whose measure is going to zero, and hence going to zero themselves.

Before moving to the main result, observe that the limit as h goes to zero involves a passage from a discrete state space $h\mathbb{Z}$ to the continuum of $\mathbb{R}$. This limiting procedure should be interpreted in the following way: we need to fix a strictly decreasing sequence of positive real numbers $(h_n)_{n\in\mathbb{N}}$ such that: (i) $\lim_n h_n = 0$ and (ii) $\frac{h_i}{h_j}$ is an integer for every $j > i$. The second property ensures that this gives rise to a nested grid, i.e., the lattice $h_j\mathbb{Z}$ contains the lattice $h_i\mathbb{Z}$. This is important since the domain of $\mathbf{P}_t^{h_i}$ is $h_i\mathbb{Z} \times h_i\mathbb{Z}$ so we want to ensure that when fixing two points x, y on the lattice $h_i\mathbb{Z}$, they will belong to all subsequent $h_j\mathbb{Z}$ so that $\mathbf{P}_t^{h_j}(x, y)$ is well defined for all $j > i$ and can thus be regarded as a function of the index j. An example of such a sequence is given by $h_n = (\frac{1}{2})^n$. In the sequel, we will drop the subscript n when dealing with the lattice spacing h_n; being clear that $h_n \to 0$ as $n \to \infty$, we will just take h to be the parameter and write $h \to 0$. In particular, when dealing with limits where spacing goes to zero, we will be assuming that the parameter h visits all elements of the sequence $(h_n)_{n\in\mathbb{N}}$ for some index n onwards.

Theorem 1.2.13. *Let X be a Lévy process with characteristic triplet given by (μ, σ, ν), where $\sigma \neq 0$. Assume that the Lévy measure is absolutely continuous with respect to the Lebesgue measure with Lévy density denoted by $\nu(\mathrm{d}x) = f(x)\mathrm{d}x$ and that f is $\mathcal{C}^2$ on $\mathbb{R} \setminus \{0\}$. Let $\mathbf{P}_t^h$ be the transition probability of a stochastic process that takes values in $h\mathbb{Z}$, given by expression (1.2.11). For any two elements x and y in $h\mathbb{Z}$, let $\rho_t(x, y)$ denote the transition density of the process X as defined in (1.2.14). The following hold true:*

(BV) If $\int_{[-1,1]} |x|\, \nu(\mathrm{d}x) < \infty$ assume that (A) holds and

$$\int_{-1}^{1} |x^3 f''(x) + x^2 f'(x) + x f(x)|\, \mathrm{d}x < \infty \quad \textit{and} \quad \int_{\mathbb{R}\setminus[-1,1]} |f'(x) + x f''(x)|\, \mathrm{d}x < \infty. \tag{1.2.17}$$

Then the following holds for any x and y in $h\mathbb{Z}$ and for any positive t:

$$\sup_{x,y\in h\mathbb{Z}} D_{x,y}(h) :=: \sup_{x,y\in h\mathbb{Z}} \left|\rho_t(x,y) - \frac{1}{h}\mathbf{P}_t^h(x,y)\right| = \mathcal{O}(h^2), \qquad h \to 0.$$

(UV) If $\int_{[-1,1]} |x|\,\nu(\mathrm{d}x) = \infty$ assume that (B) holds and

$$\int_{-1}^{1} |x^4 f''(x) + x^3 f'(x) + x^2 f(x)|\,\mathrm{d}x < \infty \quad \textit{and} \quad \int_{\mathbb{R}\setminus[-1,1]} |f'(x)|\,\mathrm{d}x < \infty. \tag{1.2.18}$$

Then the following holds for any x and y in $h\mathbb{Z}$ and for any positive t:

$$\sup_{x,y\in h\mathbb{Z}} D_{x,y}(h) :=: \sup_{x,y\in h\mathbb{Z}} \left|\rho_t(x,y) - \frac{1}{h}\mathbf{P}_t^h(x,y)\right| = \mathcal{O}(h), \qquad h \to 0.$$

Remark 1.2.14. Observe that for h small it is intuitively clear that the two quantities of interest should be comparable as it is shown by the following

$$\mathbf{P}_t^h(x,y) \approx \int_{y-h/2}^{y+h/2} \rho_t(x,z)\,\mathrm{d}z \approx h\rho_t(x,y).$$

Remark 1.2.15. Infinite activity Lévy densities have a singularity at the origin; the integral in (1.2.17) is then an improper integral and it has to be interpreted as:

$$\lim_{\epsilon\to 0^+} \int_{\epsilon}^{1} |x^3 f''(x) + x^2 f'(x) + x f(x)|\,\mathrm{d}x + \lim_{\eta\to 0^-} \int_{-1}^{\eta} |x^3 f''(x) + x^2 f'(x) + x f(x)|\,\mathrm{d}x.$$

and analogously for (1.2.18)

Remark 1.2.16. Condition (1.2.17) is saying that the Lévy density and its derivatives cannot have too much mass around the origin. In particular, we require the Lévy process to have jump part with paths of bounded variation.

Remark 1.2.17. The argument used in Theorem 2.4.1 can be applied to show that the convergence rate obtained is optimal; more precisely, it can be shown that for any function f such that $\lim_{h\to 0} f(h) = 0$, the convergence at the rate $\mathcal{O}(h^2 f(h))$ and $\mathcal{O}(hf(h))$ cannot be attained.

Remark 1.2.18. The fact that formulae (1.2.11) and (1.2.14) have separated space and time dependency is sufficient to have a uniform (in space) estimate of the difference between the transition kernels, rather than just $\left|\rho_t(x,y) - \frac{1}{h}\mathbf{P}_t^h(x,y)\right|$.

We are now ready to give a proof of the theorem.

Proof. The proof will be divided in two parts: in the first, we will show convergence of the characteristic exponents, while in the second we will prove the convergence rate for the transition kernels. More specifically, the goal of part one is to show that we can find a function $K(\cdot)$ such that

$$\lim_{h\to 0} \sup_{p\in[-K(h),K(h)]} |\Psi^h(p) - \Psi(p)| = 0.$$

Using the notation previously introduced, the difference between the characteristic exponents can be rewritten as $\Psi^h(p) - \Psi(p) = \Psi_1^h(p) - \Psi_1(p) + \Psi_2^h(p) - \Psi_2(p)$, where

$$\Psi_1^h(p) - \Psi_1(p) = \tilde{\mu}_h\rho_h(p) + \tilde{\sigma}_h^2\varphi_h(p) + \alpha_h(p), \qquad \Psi_2^h(p) - \Psi_2(p) = \beta_h(p).$$

Note the following inequality, which holds for all p in $(-\frac{1}{h}, \frac{1}{h})$:

$$\frac{\cos hp - 1}{h^2} + \frac{p^2}{2} = \frac{h^2p^4}{4!} + \epsilon_1(p), \quad \text{where} \quad -\frac{(ph)^6}{6!} \leq \epsilon_1(p) \leq 0, \tag{1.2.19}$$

from which it follows that

$$0 \leq \frac{h^2p^4}{4!} - \frac{h^4p^6}{6!} \leq \frac{\cos hp - 1}{h^2} + \frac{p^2}{2} \leq \frac{h^2p^4}{4!}.$$

The condition $p \in (-\frac{1}{h}, \frac{1}{h})$ ensures that $\cos(hp)$ can be written as an alternating series with monotonically decreasing terms. Equation (1.2.19) follows from the fact that the error of a partial sum of an alternating series with decreasing terms is bounded above by the absolute value of the first element in the series, after the partial sum[10]. In a completely

[10] In mathematical terms, given a positive sequence (a_n) such that $a_n \downarrow 0$, we have

$$\left|\sum_0^\infty (-1)^k a_k - \sum_0^n (-1)^k a_k\right| \leq |a_{n+1}|.$$

analogous fashion, we have that

$$\frac{\sin(hp)}{h} = p + \epsilon_2(p), \quad \text{where} \quad \begin{cases} 0 \le \epsilon_2(p) \le -\frac{h^2p^3}{3!} & \text{if } p \in (-\frac{1}{h}, 0), \\ -\frac{h^2p^3}{3!} \le \epsilon_2(p) \le 0 & \text{if } p \in (0, \frac{1}{h}). \end{cases}$$

Putting the two together it follows that the absolute value of the errors are in fact bounded by

$$|\rho_h| \le \frac{h^2p^3}{3!}, \qquad \varphi_h \le \frac{h^2p^4}{4!} \qquad \forall p \in \left(-\frac{1}{h}, \frac{1}{h}\right).$$

Let us now consider α_h, where, using the notation previously introduced we set I to be the interval $[-1-h/2, 1+h/2]$ and J to be the set $\{k : |kh| \le 1, k \ne 0\}$:

$$\left| \sum_J \left(\mathrm{e}^{\mathrm{i}khp} - 1 - \mathrm{i}khp + \frac{(khp)^2}{2} \right) f(kh)h - \int_I \left(\mathrm{e}^{\mathrm{i}px} - 1 - \mathrm{i}px + \frac{(px)^2}{2} \right) f(x)\,\mathrm{d}x \right|.$$

Define

$$\zeta(x) = (\mathrm{e}^{\mathrm{i}px} - 1 - \mathrm{i}px + (px)^2/2) f(x),$$

and consider only the intersection of I and J with the positive real line, $I^+ = I \cap \mathbb{R}^+$, $J^+ = J \cap \mathbb{R}^+$, I^- and J^- being defined analogously. We can rewrite

$$\left| \int_{I^+} \zeta(x)\,\mathrm{d}x - \sum_{J^+} \zeta(kh)h \right| = \left| \int_0^{h/2} \zeta(x)\,\mathrm{d}x + \int_{h/2}^{1+h/2} \zeta(x)\,\mathrm{d}x - \sum_{J^+} \zeta(kh)h \right|.$$

Set $D_h^{[h/2,1+h/2]} = \{kh + h/2; k = 0, \dots, 1/h\}$, a regular partition of $[h/2, 1+h/2]$ with spacing h. Note that the sum over J^+ in the previous expression is exactly the Riemann sums $R(\cdot)$ of ζ on $[h/2, 1+h/2]$, sampling at the midpoint. In other words

$$\sum_{J^+} \zeta(kh)h = R(\zeta, D_h^{[h/2,1+h/2]}, J^+).$$

For any given h, an application of the triangular inequality and part (d) of Theorem 1.A.6

to the difference on the interval $[h/2, 1+h/2]$ yields:

$$\left|\int_0^{\frac{h}{2}} \zeta(x)\,\mathrm{d}x + \int_{h/2}^{1+h/2} \zeta(x)\,\mathrm{d}x - \sum_{J^+} \zeta(kh)h\right| \\ \leq \int_0^{\frac{h}{2}} |\zeta(x)|\,\mathrm{d}x + TV_{[\frac{h}{2},1+\frac{h}{2}]}(\zeta')h^2, \quad (1.2.20)$$

where TV represents the total variation of ζ' over $[\frac{h}{2}, 1+\frac{h}{2}]$ and the inequality has to be considered as one between two functions of h. In the left hand side of (1.2.20) two opposite forces are at work: as h goes to zero, the mesh of the partition becomes smaller but at the same time the interval under consideration becomes larger. The integral around the origin is estimated in the following way:

$$\int_{-\frac{h}{2}}^{\frac{h}{2}} \left|\mathrm{e}^{\mathtt{i}px} - 1 - \mathtt{i}px + \frac{(px)^2}{2}\right| f(x)\,\mathrm{d}x \leq \int_{-\frac{h}{2}}^{\frac{h}{2}} |x|^3 f(x)\,\mathrm{d}x \leq \begin{cases} h^2L(h) & \text{if (BV) holds,} \\ hL(h) & \text{if (UV) holds,} \end{cases}$$

where in the last inequality we used that for all $x \in \mathbb{R}$ and $n \in \mathbb{N}$

$$\mathrm{e}^{\mathtt{i}x} = \sum_{k=0}^{n-1} \frac{(\mathtt{i}x)^k}{k!} + \theta\frac{|x|^n}{n!}, \qquad \text{with some } \theta \in \mathbb{C} \text{ satisfying } |\theta| < 1, \quad (1.2.21)$$

and $L(h) = o(1)$ is the function which is given by $\int_{-\frac{h}{2}}^{\frac{h}{2}} |x|f(x)\,\mathrm{d}x$ in case (BV) and $\int_{-\frac{h}{2}}^{\frac{h}{2}} x^2 f(x)\,\mathrm{d}x$ in case (UV). Note that by the dominated convergence theorem both functions go to zero being integrable over a set whose measure goes to zero. Since we assumed that f is twice differentiable away from the origin, an application of the mean value theorem yields the following

$$h^2 TV_{[\frac{h}{2},1+\frac{h}{2}]}(\zeta') = \frac{h^2}{8}\int_{h/2}^{1+h/2} |\zeta''(x)|\,\mathrm{d}x \leq h^2 \int_{h/2}^{1+\epsilon} |\zeta''(x)|\,\mathrm{d}x,$$

for some $\epsilon > 0$. Observe that, for a general Lévy measure, the quantity $TV_{[\frac{h}{2},1+\frac{h}{2}]}(\zeta')$ diverges as h goes to zero. An application of (1.2.21) gives

$$h^2 \int_{h/2}^{1+\epsilon} |\zeta''(x)|\,\mathrm{d}x \leq h^2 \int_{h/2}^{1+\epsilon} |xf(x) + x^2f'(x) + x^3f''(x)|\,\mathrm{d}x.$$

If (BV) holds we have that the right hand side of (1.2.22) can be bounded by $h^2 \int_0^{1+\epsilon} |xf(x)+x^2 f'(x) + x^3 f''(x)| \,\mathrm{d}x$, where the integral is finite by hypothesis. Analogously,

$$h^2 \int_{h/2}^{1+\epsilon} |xf(x) + x^2 f'(x) + x^3 f''(x)| \,\mathrm{d}x \leq h \int_{h/2}^{1+\epsilon} |x^2 f(x) + x^3 f'(x) + x^4 f''(x)| \,\mathrm{d}x,$$

and the last integral is in turn less than or equal to $\int_0^{1+\epsilon} |x^2 f(x) + x^3 f'(x) + x^4 f''(x)| \,\mathrm{d}x$, which is finite if (UV) holds. Summarising, we have proved that for all $p \in \mathbb{R}$

$$|\alpha_h(p)| \leq \begin{cases} C_{BV} h^2 & \text{if (BV) holds,} \\ C_{UV} h & \text{if (UV) holds,} \end{cases}$$

where C_{BV} and C_{UV} are two constants which can be identified explicitly. Define $\eta(x) = (\mathrm{e}^{\mathrm{i}px} - 1) f(x)$ and once again consider the positive part of the sets $h\mathbb{Z} \setminus J$ and $\mathbb{R} \setminus I$ for typographical reasons. The last term to consider is β_h:

$$|\beta_h| = \left| \sum_{k=2}^{\infty} \eta(kh)h - \int_{1+h/2}^{\infty} \eta(x) \,\mathrm{d}x \right|.$$

By the definition of limit it is possible to find two functions, $G(h)$ and $H(h)$ such that

$$\left| \int_{1+h/2}^{G(h)} \eta(x) \,\mathrm{d}x - \int_{1+h/2}^{\infty} \eta(x) \,\mathrm{d}x \right| < h^2, \qquad \left| \sum_{k=2}^{H(h)} \eta(kh)h - \sum_{k=2}^{\infty} \eta(kh)h \right| < h^2.$$

Set $J(h) = \lceil G(h) \rceil \vee H(h)$, where we use the function $\lceil x \rceil$ denotes the smallest integer larger than x. Then

$$\left| \sum_{k=2}^{\infty} \eta(kh)h - \int_{1+h/2}^{\infty} \eta(x) \,\mathrm{d}x \right| \leq \left| \sum_{k=2}^{J(h)} \eta(kh)h - \int_{1+h/2}^{J(h)} \eta(x) \,\mathrm{d}x \right| + 2h^2.$$

Another application of Theorem 1.A.6, provides us with the estimate:

$$\left| \sum_{k=2}^{J(h)} \eta(kh)h - \int_{1+h/2}^{J(h)} \eta(x) \,\mathrm{d}x \right| \leq \begin{cases} \frac{h^2}{8} \int_{1+h/2}^{J(h)} |\eta''(x)| \,\mathrm{d}x & \text{if} \quad \text{(BV),} \\ \frac{h}{2} \int_{1+h/2}^{J(h)} |\eta'(x)| \,\mathrm{d}x & \text{if} \quad \text{(UV),} \end{cases}$$

where both integrals on the right hand side are finite if (BV) and (UV) hold respectively.

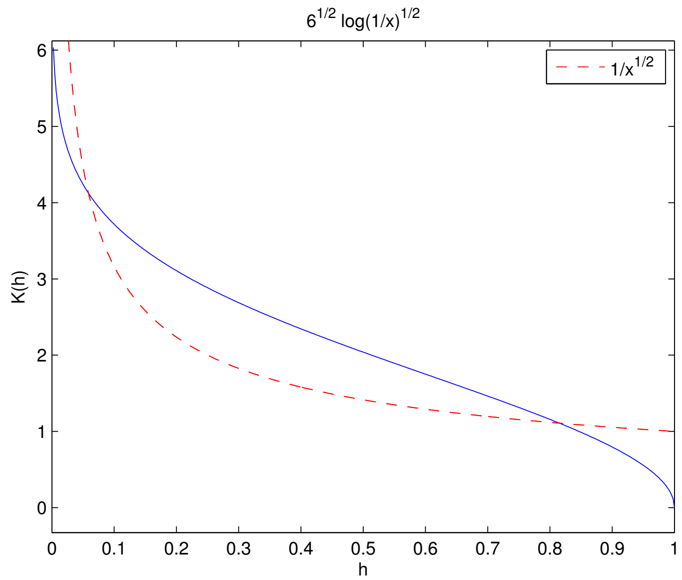

Figure 1.1: Graph of $K(h)$ for $\sigma = 1$ and $\frac{1}{\sqrt{x}}$. Observe the slower divergence of K at the origin.

Summarising, we have obtained that

$$|\Psi^h(p) - \Psi(p)| \leq \begin{cases} Mh^2\left(1 + \frac{p^3}{3!} + \frac{p^4}{4!}\right), \\ Mh\left(1 + \frac{p^3}{3!} + \frac{p^4}{4!}\right), \end{cases}$$

where M is a generic constant. Define the function $K : (0, \infty) \to \mathbb{R}$ to be:

$$K(h) = \sqrt{\frac{6}{\sigma^2} \log\left(\frac{1}{h}\right)}.$$

Then,

$$\sup_{p \in [-K(h), K(h)]} |\Psi^h(p) - \Psi(p)| \leq Mh^2\left(1 + \frac{K(h)^3}{3!} + \frac{K(h)^4}{4!}\right),$$

and the right hand side goes to zero since $K(h) = o(1/\sqrt{h})$. The second part of the proof consists of studying the convergence of

$$\left|\rho_t(x, y) - \frac{1}{h}\mathbf{P}_t^h(x, y)\right|. \tag{1.2.22}$$

For ease of exposition we will prove it when (BV) holds but exactly the same proof applies using the estimate of the characteristic exponents for case (UV), that we have provided. Since in part 1 we showed that $\sup_{p\in[-K(h),K(h)]} |\Psi^h(p) - \Psi(p)| \to 0$ as $h \searrow 0$, we split the estimation of (1.2.22) in three terms: the integral over $[-K(h), K(h)]$, which we can control since we have uniform convergence, and separate integrals over $(-\frac{\pi}{h}, -K(h)) \cup (K(h), \frac{\pi}{h})$ and $(-\infty, -K(h)) \cup (K(h), \infty)$, which will go to zero since the integrands are integrable. Note also that this is possible since $K(h)$ goes to infinity as h approaches zero but is bounded above by $\frac{\pi}{h}$. Following this programme, the application of the triangle inequality shows that (1.2.22) can be bounded from above by

$$\int_{-K(h)}^{K(h)} |\mathrm{e}^{t\Psi(p)}|\,|\mathrm{e}^{t(\Psi^h(p)-\Psi(p))} - 1|\,\mathrm{d}p + 2\int_{K(h)}^{\frac{\pi}{h}} |\mathrm{e}^{t\Psi^h(p)}|\,\mathrm{d}p + 2\int_{K(h)}^{\infty} |\mathrm{e}^{t\Psi(p)}|\,\mathrm{d}p,$$

where the factor 2 in front of the second and third integral follows from the fact that we are integrating over symmetric regions. Let us focus on the first integral. Observe the inequality

$$|\mathrm{e}^z - 1| \le \mathrm{e}^{|z|} - 1 \qquad \forall z \in \mathbb{C}. \tag{1.2.23}$$

This inequality follows from taking the absolute value of the Taylor series for the function $z \mapsto \mathrm{e}^z - 1$ and applying the triangle inequality to it: $|\mathrm{e}^z - 1| = |\sum_1^\infty \frac{z^n}{n!}| \le \sum_1^\infty \frac{|z|^n}{n!} = \mathrm{e}^{|z|} - 1$. Using (1.2.23), the inequality $e^x - 1 \le 2x$, which holds for x small and positive and the estimates obtained in part 1, we have:

$$\begin{aligned}
\int_{-K(h)}^{K(h)} &|\mathrm{e}^{t\Psi(p)}|\,|\mathrm{e}^{t(\Psi^h(p)-\Psi(p))} - 1|\,\mathrm{d}p \\
&\le \int_{-K(h)}^{K(h)} \mathrm{e}^{-t\frac{\sigma^2 p^2}{2}} \left(\mathrm{e}^{t|\Psi^h(p)-\Psi(p)|} - 1\right)\,\mathrm{d}p \\
&\le 2t\int_{-K(h)}^{K(h)} \mathrm{e}^{-t\frac{\sigma^2 p^2}{2}}\,|\Psi^h(p) - \Psi(p)|\,\mathrm{d}p \\
&\le 2h^2\,t\int_{-K(h)}^{K(h)} \mathrm{e}^{-t\frac{\sigma^2 p^2}{2}} \left(\tilde{\mu}_h\frac{p^4}{4!} + \tilde{\sigma}_h^2\frac{p^3}{3!} + C\right)\,\mathrm{d}p \\
&\le 2h^2\,t\int_{-\infty}^{\infty} \mathrm{e}^{-t\frac{\sigma^2 p^2}{2}} \left(\tilde{\mu}_h\frac{p^4}{4!} + \tilde{\sigma}_h^2\frac{p^3}{3!} + C\right)\,\mathrm{d}p.
\end{aligned} \tag{1.2.24}$$

Note that any function of the form $P(p)\mathrm{e}^{-\frac{1}{2}\sigma^2 p^2}$, where $P(\cdot)$ is a polynomial of any degree, is integrable over the real line, hence the integral (1.2.24) is finite. Now consider the second integral; since the Fourier transform of a probability measure is always less than or equal to one, the use of the elementary inequality $|\mathrm{e}^z| \leq \mathrm{e}^{|z|}$ yields

$$\int_{K(h)}^{\frac{\pi}{h}} |\mathrm{e}^{t\Psi^h(p)}|\,\mathrm{d}p \leq \int_{K(h)}^{\frac{\pi}{h}} \mathrm{e}^{t\sigma^2\frac{\cos(hp)-1}{h^2}}\mathrm{d}p.$$

Since the function $p \mapsto \frac{\cos(hp)-1}{h^2}$ is monotonically decreasing on the interval $[K(h), \frac{\pi}{h}]$ we can proceed with the estimate:

$$\int_{K(h)}^{\frac{\pi}{h}} \mathrm{e}^{\sigma^2 t\frac{\cos(hp)-1}{h^2}} \leq \frac{\pi}{h}\mathrm{e}^{\sigma^2 t\frac{\cos(hK(h))-1}{h^2}} \leq \frac{\pi}{h}\mathrm{e}^{\sigma^2 t(-\frac{K(h)^2}{2}+h^2K(h)^4)} \leq \frac{2\pi}{h}\mathrm{e}^{-\sigma^2 t\frac{K(h)^2}{2}},$$

where we used the estimate (1.2.19) in the third inequality and the fact that $h^2K(h)^4$ is less than $\frac{log(2)}{t\sigma^2}$ for small h in the last one (it goes to zero). Using the definition of $K(h)$ in the last expression then gives

$$\frac{2\pi}{h}\mathrm{e}^{-\frac{K(h)^2}{2}\sigma^2 t} \leq 2\pi h^2.$$

The third integral is handled in the following way:

$$\int_{K(h)}^{\infty} |\mathrm{e}^{t\Psi(p)}|\,\mathrm{d}p = \int_{K(h)}^{\infty} \mathrm{e}^{-t\frac{\sigma^2p^2}{2}}\,\mathrm{d}p \leq \sqrt{2\pi\sigma^2 t}\,\mathrm{e}^{-t\frac{K(h)^2}{2}\sigma^2} = \sqrt{2\pi\sigma^2 t}\,h^3,$$

where once again we used the fact that the modulus of the Fourier transform of a probability measure is less than or equal to one and the well-known estimate for the tail of a standard normal X that reads

$$\mathbb{P}(X > y) \leq \frac{1}{\sqrt{2\pi}}\mathrm{e}^{-y^2/2},$$

whenever y is greater than one. Putting the estimate of the three integrals all together and observing that the result does not depend on x and y proves the convergence estimate in the theorem. □

1.3 Lattice approximation of a pure jump, bounded variation Lévy process

1.3.1 Introduction

In the previous analysis, we examined the rate of convergence of the transition probabilities of the approximating process to a Lévy process which has Gaussian component. While the density of a Lévy process is a very useful quantity in a variety of problems, it can be especially hard to obtain in explicit form. This is even more true in the case of a pure jump Lévy process since we are not able to exploit the Gaussian decay of the characteristic function on the tails to ensure integrability away from the origin. We now turn to the study of convergence and its rate for a pure jump Lévy process i.e. a Lévy process whose Gaussian component is not present. Note that the paths of such a process will be almost surely discontinuous everywhere.

More specifically, let us consider a Lévy process $X = \{X_t; t \geq 0\}$ with characteristic triplet given by $(\mu, 0, \nu)$ and such that the jump process associated has paths of bounded variation. As we observed in the introduction this condition can be made explicit in terms of the Lévy measure ν which is required to integrate the function $|x|$ on the unit ball i.e.

$$\int_{(-1,1)} |x| \, \nu(\mathrm{d}x) < \infty.$$

This condition specifies how much mass the Lévy measure can have around the singularity at the origin. Observe that such a process can still have an infinite number of jumps on any compact subset of the positive real line and many processes of interest fall into this category. Building upon the work already done, we may wish to approximate the drift component of the generator of the Lévy process X by discretising using a central difference approximation for the first derivative as in equation (1.2.1). Although X has no Brownian component, taking inspiration from the work of Asmussen and Rosinski [6], we know that Brownian motion can well be used to approximate the small jumps of a Lévy process and we may want to include a discretization of the second derivative when defining the Markov chain as in (1.2.1). The interpretation of the Lévy measure $\nu(A)$ as the expected number of jumps *per unit time* with size in A supports the definition of Λ_h in equation (1.4.6), which corresponds to a simple discretization of of $\nu(A) = \int_A f(x)\,\mathrm{d}x$. With this setup, using the notation previously introduced, we can write the generator of our continuous time Markov chain as

$$\mathcal{L}_h = \tilde{\mu}_h \nabla_h + \frac{\tilde{\sigma}_h^2}{2}\Delta_h + \Lambda_h,$$

where $\tilde{\mu}_h$ and $\tilde{\sigma}_h$ are parameters, possibly dependent on h, which will be used to match the moments of X^h with those of X. Carrying out the moment matching procedure as in Section 1.2.1, we obtain the following expressions for the parameters $\tilde{\mu}_h$ and $\tilde{\sigma}_h$:

$$\begin{aligned} \tilde{\mu}_h &= \mu + \int_I |x|\,\nu(\mathrm{d}x) - \sum_J khf(kh)h, \\ \tilde{\sigma}_h^2 &= \int_I x^2\,\nu(\mathrm{d}x) - \sum_J (hk)^2 f(kh)h. \end{aligned}$$

Note, however, that there are several problems with this expression: even assuming that $\tilde{\sigma}_h^2 \to 0$ as $h \searrow 0$, its sign is not clear. This seemingly innocuous fact turns out to be problematic for the definition of the Markov chain. In fact, it is well known that the mapping $\mathcal{L}_h$ defines a generator (and hence a Markov chain) only if the off-diagonal elements of the matrix associated with $\mathcal{L}_h$ are non-negative. This enforces:

$$\begin{aligned} \mathcal{L}_h(x, x-h) &= -\frac{\tilde{\mu}_h}{2h} + \frac{\tilde{\sigma}_h^2}{2h^2} + f(-h)h \geq 0, \\ \mathcal{L}_h(x, x+h) &= \frac{\tilde{\mu}_h}{2h} + \frac{\tilde{\sigma}_h^2}{2h^2} + f(h)h \geq 0. \end{aligned} \tag{1.3.1}$$

Note also how the term $f(h)h$ is not sufficient to compensate the negativity of $-\frac{|\tilde{\mu}_h|}{2h}$, at least in the bounded variation case; an easy example is given by a process with alpha stable jumps, for which we have $f(h)h = 1/h^\alpha$, which does not go to infinity fast enough for $\alpha \in (0,1)$. This suggests to change the definition of ∇_h in the case when $f(\cdot)$ is of bounded variation since the source of the problem is the fact that ∇_h defined as above (i.e. using central difference approximation) is not a Markov generator, having negative elements off the diagonal. The solution that we will adopt will be to define ∇_h using one-sided derivatives: this has the advantage to make the discrete gradient operator a Markov generator but comes at the price of losing accuracy when approximating the first derivative. In fact, by using one-sided derivative approximation we make an error of order $\mathcal{O}(h)$ as it can be easily deduced by looking at the Taylor development

$$\frac{\phi(x+h)-\phi(x)}{h} = \phi'(x) + \mathcal{O}(h).$$

This means that the rate of convergence of the scheme cannot be faster than $\mathcal{O}(h)$ and therefore we will not deploy the discrete Laplacian operator Δ_h, since matching the second moments would not lead to any improvement.

1.3.2 Construction of the chain and main theorem

The discretization of the drift part of the generator of the Lévy process will now depend on the sign of the drift of the Lévy process. Namely, if $\mu > 0$ define

$$\nabla_h^+(\phi)(x) = \frac{\phi(x+h)-\phi(x)}{h},$$

whereas if $\mu < 0$

$$\nabla_h^-(\phi)(x) = \frac{\phi(x-h)-\phi(x)}{h}.$$

In coordinates (recall that $\nabla_h^{\pm}(x,y)$ is defined to be $\nabla_h^{\pm}(\delta_y)(x)$):

$$\nabla_h^+(x,y) = \frac{1}{h}\begin{cases} 1 & \text{if } y = x+h, \\ -1 & \text{if } y = x, \\ 0 & \text{otherwise.} \end{cases} \qquad \nabla_h^-(x,y) = \frac{1}{h}\begin{cases} 1 & \text{if } y = x-h, \\ -1 & \text{if } y = x, \\ 0 & \text{otherwise.} \end{cases}$$

The equations above implies that $\nabla_h^{\pm}$ is a Markov generator (on $l^2(h\mathbb{Z})$) and thus defines a Markov chain (with state space $l^2(h\mathbb{Z})$). The behaviour of this chain is very simple: if the drift of the Lévy process is positive(negative), the chain can only move upwards(downwards) at a rate of $1/h$. Summarising, the scheme we propose to approximate a Lévy process with bounded variation and characteristic triplet $(\mu, 0, \nu)$ is:

$$\mathcal{L}_h^{\pm} = \tilde{\mu}_h \nabla_h^{\pm} + \Lambda_h,$$

where Λ_h is defined as in (1.4.6). A calculation similar to that of Section 1.2.2 leads to

$$\mathcal{F}_h \nabla_h^+(a)(p) = \frac{\mathrm{e}^{\mathrm{i}hp} - 1}{h} \mathcal{F}_h(a)(p), \quad \mathcal{F}_h \nabla_h^-(a)(p) = \frac{\mathrm{e}^{-\mathrm{i}hp} - 1}{h} \mathcal{F}_h(a)(p), \tag{1.3.2}$$

for $a \in l^2(h\mathbb{Z})$. An application of Theorem 1.2.5 allows us to find the expression of the characteristic function of $\mathcal{L}_h^{\pm}$

$$\Psi_+^h(p) = \tilde{\mu}_h \frac{\mathrm{e}^{\mathrm{i}hp} - 1}{h} + \sum_{k \neq 0} (\mathrm{e}^{\mathrm{i}khp} - 1) f(kh) h,$$
$$\Psi_-^h(p) = \tilde{\mu}_h \frac{\mathrm{e}^{-\mathrm{i}hp} - 1}{h} + \sum_{k \neq 0} (\mathrm{e}^{\mathrm{i}khp} - 1) f(kh) h.$$

The characteristic exponent corresponding to the Lévy process with characteristic triplet $(\mu, 0, \nu)$ can be written as

$$\Psi(p) = \mathrm{i}\left(\mu + \int_I x f(x)\,\mathrm{d}x\right) p + \int_I (\mathrm{e}^{\mathrm{i}px} - 1 - \mathrm{i}px) f(x)\,\mathrm{d}x + \int_{\mathbb{R}\setminus I} (\mathrm{e}^{\mathrm{i}px} - 1) f(x)\,\mathrm{d}x. \tag{1.3.3}$$

Matching the first moments in order to determine the coefficient $\tilde{\mu}_h$ gives:

$$\tilde{\mu}_h = \begin{cases} \mu + \int_I x\,\nu(\mathrm{d}x) - \sum_J khf(kh)h & \text{if } \mu > 0, \\ -\mu - \int_I x\,\nu(\mathrm{d}x) + \sum_J khf(kh)h & \text{if } \mu < 0. \end{cases} \tag{1.3.4}$$

In the following, we shall assume that $\int_{[-1,1]\setminus[-h,h]} |f(x) + xf'(x)|\,\mathrm{d}x = o(1/h)$ as h goes to zero which ensures that $|\int_I x\,\nu(\mathrm{d}x) - \sum_J khf(kh)h| \to 0$ by Theorem 1.A.6. This implies that $\mathcal{L}_h^{\pm}$ with $\tilde{\mu}_h$ defined above satisfies the requirements for being a Markov generator, at least for h small enough, since $\lim_{h\to 0} \tilde{\mu}_h = |\mu|$ and hence the positivity of the off diagonal elements is preserved. Equation (1.3.3) rewritten using the "adjusted mean" $\tilde{\mu}_h$ reads:

$$\Psi(p) = \mathrm{i}\tilde{\mu}_h p + \int_I (\mathrm{e}^{\mathrm{i}px} - 1 - \mathrm{i}px) f(x)\,\mathrm{d}x + \int_{\mathbb{R}\setminus I} (\mathrm{e}^{\mathrm{i}px} - 1) f(x)\,\mathrm{d}x + \mathrm{i} \sum_J khf(kh)hp.$$

The main result of this section is the following:

Theorem 1.3.1. *Let X be a Lévy process with characteristic triplet given by $(\mu, 0, \nu)$ and let $t > 0$. Assume that the Lévy measure is absolutely continuous with respect to the Lebesgue measure with Lévy density denoted by $\nu(\mathrm{d}x) = f(x)\mathrm{d}x$ and that f is $\mathcal{C}^1$ on $\mathbb{R} \setminus \{0\}$. Let $\mathbf{P}_t^h$ be the transition probability of a stochastic process that takes values in $h\mathbb{Z}$, given by expression (1.2.11). For any two elements x and y in $h\mathbb{Z}$, let $\rho_t(x, y)$ denote the transition density of the process X as defined in (1.2.14). Assume also that the following conditions are satisfied:*

(a) $\int_{[-1,1]} |x^2 f'(x) + xf(x)|\,\mathrm{d}x < \infty$, *and* $\int_{[-1,1]\setminus[-h,h]} |f(x) + xf'(x)|\,\mathrm{d}x = o(\frac{1}{h})$,

(b) $\int_{\mathbb{R}\setminus[-1,1]} |f(x) + xf'(x)|\,\mathrm{d}x < \infty$,

(c) $e^{t\Psi(p)} = \mathcal{O}\left(\left(p^2(\log|p|)^{(1+\epsilon)}\right)^{-1}\right)$, *for some $\epsilon > 0$, as $p \to \infty$.*

Then

$$\sup_{x,y\in h\mathbb{Z}} D_{x,y}(h) := \sup_{x,y\in h\mathbb{Z}} \left| \rho_t(x, y) - \frac{1}{h}\mathbf{P}_t^h(x, y) \right| = \mathcal{O}(h), \qquad h \to 0.$$

Remark 1.3.2. The first two conditions are required to ensure that the characteristic exponent of the discretized process converges to the characteristic exponent of X with the prescribed rate. The third is simply a condition on the decay of the characteristic exponent of the Lévy process. Note that (c) is sufficient to guarantee the existence and differentiability of the density of X_t for all t. The latter follows since $\int |\mathrm{e}^{t\Psi(p)}| p\,\mathrm{d}p \leq \int (p(\log|p|)^{(1+\epsilon)})^{-1}\,\mathrm{d}p < \infty$ away from the origin and the same quantity around the origin is bounded since $\mathrm{e}^{t\Psi(p)}$ is uniformly continuous. We observe that condition (c) is indeed quite mild and it is satisfied, for instance, by all stable processes.

Proof. To fix the ideas, in this section we shall be working in the case of μ positive, the other being completely analogous. The idea of the proof is to split the quantity of interest in three parts and to control each one separately. The challenge here is to find a suitable function $K(\cdot)$ that grows slowly enough as to make $\sup_{p\in[-K(h),K(h)]} |\Psi(p) - \Psi^h(p)|$ small and yet fast enough that the quantity $\int_{K(h)}^{\infty} |\mathrm{e}^{t\Psi(p)}|\,\mathrm{d}p$ goes to zero with the required rate. This can be translated into a twofold requirement on the function K:

(1) for all $h \in (0, \infty)$ we have $\int_{K(h)}^{\infty} \frac{1}{p^2(\log p)^{1+\epsilon}}\,\mathrm{d}p = h$,

(2) the function $\Gamma(h) = \sup_{p\in[-K(h),K(h)]} |\Psi(p) - \Psi^h(p)|$ is such that $\lim_{h\to 0}\Gamma(h) = 0$.

The difference between the two characteristic exponents reads:

$$|\Psi_1^h(p) - \Psi_1(p)| = \Bigg|\tilde{\mu}_h\left(\frac{\mathrm{e}^{\mathrm{i}hp}-1}{h} - \mathrm{i}p\right) + \sum_J \left(\mathrm{e}^{\mathrm{i}khp} - 1 - \mathrm{i}pkh\right) f(kh)h$$
$$- \int_I \left(\mathrm{e}^{\mathrm{i}px} - 1 - \mathrm{i}px\right)\,\nu(\mathrm{d}x) + \sum_{h\mathbb{Z}\setminus J}\left(\mathrm{e}^{\mathrm{i}khp}-1\right) f(kh)h - \int_{\mathbb{R}\setminus I}\left(\mathrm{e}^{\mathrm{i}px}-1\right)\,\nu(\mathrm{d}x)\Bigg|. \tag{1.3.5}$$

We now show that, under conditions (a) and (b), there exist two constants M_1 and M_2 such that, for all $p\in\mathbb{R}$:

$$\left|\sum_J \left(\mathrm{e}^{\mathrm{i}khp} - 1 - \mathrm{i}khp\right) f(kh)h - \int_I \left(\mathrm{e}^{\mathrm{i}px} - 1 - \mathrm{i}px\right) f(x)\,\mathrm{d}x\right| \le M_1 h, \tag{1.3.6}$$

$$\left|\sum_{h\mathbb{Z}\setminus J} (\mathrm{e}^{\mathrm{i}khp}-1) f(kh)h - \int_{\mathbb{R}\setminus I} (\mathrm{e}^{\mathrm{i}px}-1) f(x)\,\mathrm{d}x\right| \le M_2 h. \tag{1.3.7}$$

We prove (1.3.6), (1.3.7) being similar in spirit. Set ζ to be the function $\zeta(x) = (\mathrm{e}^{\mathrm{i}px} - 1 - \mathrm{i}px)f(x)$; to avoid unnecessary complication in the notations, we consider only the intersection of I and J with the positive real line, $I^+ = I\cap\mathbb{R}^+$, $J^+ = J\cap\mathbb{R}^+$, the other case being similar. We can rewrite

$$\left|\int_{I^+}\zeta(x)\,\mathrm{d}x - \sum_{J^+}\zeta(kh)h\right| = \left|\int_0^{h/2}\zeta(x)\,\mathrm{d}x + \int_{h/2}^{1+h/2}\zeta(x)\,\mathrm{d}x - \sum_{J^+}\zeta(kh)h\right|.$$

Set $D_h^{[h/2,1+h/2]} = \{kh + h/2; k = 0,\dots,1/h\}$, a regular partition of $[h/2, 1+h/2]$ with spacing h. Since $\sum_{J^+}\zeta(kh)h$ is exactly the Riemann sums of the function ζ on the partition $D_h^{[h/2,1+h/2]}$ and sampled at the points in J^+, which we denote by $R(\zeta, D_h^{[h/2,1+h/2]}, J^+)$, we can then control the difference by applying Theorem 1.A.6(e)

$$\left|\int_{h/2}^{1+h/2}\zeta(x)\,\mathrm{d}x - R(\zeta, D_h^{[h/2,1+h/2]}, J^+)\right|$$
$$\le \frac{h}{2}\int_{h/2}^{1+h/2}|\zeta'(x)|\,\mathrm{d}x \le h\int_0^1 |\zeta'(x)|\,\mathrm{d}x, \tag{1.3.8}$$

where the last inequality holds for h sufficiently small. Note that the integral in the right hand side of the previous inequality is finite by hypothesis and does not depend on h. The integral around the origin is not compensated by an element in the sum. However, we can estimate this quantity from above to obtain the desired bound:

$$\int_{-\frac{h}{2}}^{\frac{h}{2}} \left|\mathrm{e}^{\mathrm{i}px} - 1 - \mathrm{i}px\right| f(x)\,\mathrm{d}x \leq \int_{-\frac{h}{2}}^{\frac{h}{2}} |x|^2 f(x)\,\mathrm{d}x \leq \frac{h}{2}\int_{-\frac{h}{2}}^{\frac{h}{2}} |x| f(x)\,\mathrm{d}x.$$

Observe that the integral on the right hand side of the previous display defines a function of h that goes to zero as h goes to zero by virtue of the integrability requirement in the hypothesis. By virtue of (1.3.6) and (1.3.7) we have that the second and third differences in the absolute value in (1.3.5) go to zero uniformly on the whole real line as h goes to zero. It then follows that

$$\begin{aligned}\sup_{p\in[-K(h),K(h)]} |\Psi(p) - \Psi^h(p)| \leq \sup_{p\in[-K(h),K(h)]} |\tilde{\mu}_h| \left|\frac{\mathrm{e}^{\mathrm{i}hp} - 1}{h} - \mathrm{i}p\right| + \epsilon(h) \\ \leq C\frac{1}{2}hK^2(h) + \epsilon(h),\end{aligned}$$

where the last inequality is a consequence of the elementary fact $\left|\frac{\mathrm{e}^{\mathrm{i}hp}-1}{h} - \mathrm{i}p\right| < \frac{1}{2}hp^2$, $\epsilon(h) = o(1), h \to 0$ and C is a generic constant. It therefore suffices to require for K to satisfy $\lim_{h\to 0}\sqrt{h}K(h) = 0$ in order to fulfil (2).

Let us now focus our attention on (1). Performing the change of variable $p = \mathrm{e}^u$ we obtain, for all $a > 0$, the identity

$$\int_a^\infty \frac{1}{p^2(\log p)^{1+\epsilon}}\,\mathrm{d}p = \int_{\log a}^\infty \mathrm{e}^{-u}u^{-\epsilon-1}\,\mathrm{d}u = \Gamma(-\epsilon, \log a).$$

The function Γ used in the previous equation is the upper incomplete gamma function. Its definition is given by

$$\Gamma(s,x) = \int_x^\infty \mathrm{e}^{-t}t^{s-1}\,\mathrm{d}t,$$

and we will write $\Gamma(x)$ to single out the dependency on the second variable. The above mentioned condition (1) then reads $\Gamma(\log K(h)) = h$, or equivalently:

$$K(h) = \mathrm{e}^{\Gamma^{-1}(h)}.$$

Furthermore, we need to check that the growth conditions on our function are satisfied. More specifically, we need to verify that $K(h) = o(h^{-1/2})$ as $h \to 0^+$. Observe that, by an application of L'Hospital's rule, $\Gamma(x) = o(\mathrm{e}^{-x})$ for $x \to \infty$. Since $\Gamma(x)$ goes to zero faster than the exponential function, we have that Γ^{-1} goes to ∞ slower than $-\log$; in other words, $\lim_{x\to 0^+} \Gamma^{-1}(x)/-\log(x) = 0$. In particular, for all $\eta > 0$ we have that $\Gamma^{-1}(x) = o(-\log(\sqrt{x}/\eta))$, as $x \to 0^+$ By the definition of Landau symbol little-o, it follows that, in a right neighbourhood of zero, we get $\Gamma^{-1}(x) \le -\log(\sqrt{x}/\eta)$ for all positive η. Thus, $\mathrm{e}^{\Gamma^{-1}(h)} < \eta/\sqrt{h}$ and yields the condition we set out to prove. Once the function $K(\cdot)$ has been determined, the estimation proceeds as in the previous case:

$$\begin{aligned} D_{x,y}(h) \le \int_{-K(h)}^{K(h)} |\mathrm{e}^{t\Psi(p)}|\, |\mathrm{e}^{t(\Psi^h(p)-\Psi(p))} - 1|\, \mathrm{d}p \\ + 2\int_{K(h)}^{\frac{\pi}{h}} |\mathrm{e}^{t\Psi^h(p)}|\, \mathrm{d}p + 2\int_{K(h)}^{\infty} |\mathrm{e}^{t\Psi(p)}|\, \mathrm{d}p. \end{aligned} \tag{1.3.9}$$

Let us now focus our attention on the first integral of the above. By use of the inequality $|\mathrm{e}^z - 1| \le \mathrm{e}^{|z|} - 1$, which holds for any complex z we obtain the upper bound:

$$\int_{-K(h)}^{K(h)} |\mathrm{e}^{t\Psi(p)}|\, |\mathrm{e}^{t(\Psi^h(p)-\Psi(p))} - 1|\, \mathrm{d}p \le \int_{-K(h)}^{K(h)} |\mathrm{e}^{t\Psi(p)}|\, (\mathrm{e}^{t|\Psi^h(p)-\Psi(p)|} - 1)\, \mathrm{d}p. \tag{1.3.10}$$

Since $\sup_{p\in[-K(h),K(h)]} |\Psi^h(p) - \Psi(p)| \to 0$ as h goes to zero, we can use an inequality of the exponential function $\mathrm{e}^x - 1 \le 2x$ for x small and positive to obtain an upper bound on the right hand side of (1.3.10):

$$\begin{aligned} 2\int_{-K(h)}^{K(h)} |\mathrm{e}^{t\Psi(p)}| t |\Psi^h(p) - \Psi(p)|\, \mathrm{d}p &\le 2th \int_{-K(h)}^{K(h)} |\mathrm{e}^{t\Psi(p)}| \left(\frac{p^2}{2} + M_1 + M_2\right) \mathrm{d}p \\ &\le 2th \int_{-\infty}^{\infty} |\mathrm{e}^{t\Psi(p)}| \left(\frac{p^2}{2} + M_1 + M_2\right) \mathrm{d}p, \end{aligned}$$

where in the second inequality we used $|\Psi^h(p) - \Psi(p)| \le h\,(p^2/2 + M_1 + M_2)$ for all p

in $[-K(h), K(h)]$. Note that the convergence of the integral on the right hand side of the inequality is guaranteed by condition (c). Let us now turn our attention to the estimation of the third integral in (1.3.9). Here, most of the work has already been done in our careful selection of the function $K(\cdot)$:

$$\int_{K(h)}^{\infty} |\mathrm{e}^{t\Psi(p)}|\,\mathrm{d}p \le \int_{K(h)}^{\infty} |\mathrm{e}^{t\Psi(p)}| \frac{p^2(\log p)^{1+\epsilon}}{p^2(\log p)^{1+\epsilon}}\mathrm{d}p \le M \int_{K(h)}^{\infty} \frac{1}{p^2(\log p)^{1+\epsilon}}\,\mathrm{d}p,$$

where in the latter, the constant M is given by the definition of $\mathcal{O}(\cdot)$ in condition (c) and the right hand side is equal to h by our careful choice of K. It now only remains to estimate the second integral in (1.3.9): $\int_{K(h)}^{\frac{\pi}{h}} |\mathrm{e}^{t\Psi^h(p)}|\,\mathrm{d}p$, where:

$$\Psi^h(p) = \tilde{\mu}_h \frac{\mathrm{e}^{\mathrm{i}hp}-1}{h} + \sum_{J} (\mathrm{e}^{\mathrm{i}khp} - 1 - \mathrm{i}khp) f(kh)h + \sum_{h\mathbb{Z}\setminus J} (\mathrm{e}^{\mathrm{i}khp} - 1) f(kh)h.$$

Observe that we can rewrite $\frac{\mathrm{e}^{\mathrm{i}hp}-1}{h} = \frac{1}{h}(\cos hp - \mathrm{i}\sin hp - 1)$, and the exponential becomes

$$\left|\exp\left(\tilde{\mu}_h \frac{\mathrm{e}^{\mathrm{i}hp}-1}{h}\right)\right| = \exp\left(\tilde{\mu}_h \frac{\cos hp - 1}{h}\right),$$

and the latter is less than or equal to one since $-\frac{2}{h} \le \frac{\cos hp-1}{h} \le 0$, and $\tilde{\mu}_h$ is positive for h small enough ($\tilde{\mu}_h \to |\mu|$). Using the elementary identity $|\mathrm{e}^{\rho+\mathrm{i}\theta}| = \mathrm{e}^{\rho}$, we obtain

$$\left|\exp\left(\sum_{J} (\mathrm{e}^{\mathrm{i}khp} - 1) f(kh)h\right)\right| = \exp\left(\sum_{J} (\cos khp - 1) f(kh)h\right).$$

From (1.3.6), the rather crude bound, $|z| \ge |\mathfrak{Re}(z)|$, valid for all complex z, yields

$$\left|\sum_{J} (\cos khp - 1) f(kh)h - \int_I (\cos px - 1) f(x)\,\mathrm{d}x\right| \le M_1 h,$$

and hence we obtain

$$\left|\exp\left(\sum_{J} (\mathrm{e}^{\mathrm{i}khp} - 1) f(kh)h\right)\right| \le \exp\left(\int_I (\cos px - 1) f(x)\,\mathrm{d}x + M_1 h\right). \quad (1.3.11)$$

In a similar fashion, condition (b) allows to get estimate (1.3.7), which in turn gives

$$\left|\exp\left(\sum_{h\mathbb{Z}\setminus J}(\mathrm{e}^{\mathtt{i}khp}-1)f(kh)h\right)\right| \leq \exp\left(\int_{\mathbb{R}\setminus I}(\cos px-1)f(x)\,\mathrm{d}x+M_2h\right). \quad (1.3.12)$$

Putting the inequalities (1.3.11) and (1.3.12) together:

$$\begin{aligned}\int_{K(h)}^{\frac{\pi}{h}}|\mathrm{e}^{t\Psi^h(p)}|\,\mathrm{d}p &\leq \int_{K(h)}^{\frac{\pi}{h}}\mathrm{e}^{\int_{\mathbb{R}}(\cos px-1)f(x)\,\mathrm{d}x+(M_1+M_2)h}\,\mathrm{d}p\\ &\leq C\int_{K(h)}^{\infty}\mathrm{e}^{\int_{\mathbb{R}}(\cos px-1)f(x)\,\mathrm{d}x}\,\mathrm{d}p = C\int_{K(h)}^{\infty}|\mathrm{e}^{\Psi(p)}|\,\mathrm{d}p,\end{aligned}$$

where C is a constant. By definition of $K(h)$ and assumption (c) the latter is less than or equal to a constant times h. $\square$

1.4 Lattice approximation of pure jump, unbounded variation Lévy process.

Introduction

In this section, we examine the problem of approximating a pure jump, unbounded variation Lévy process X with a continuous time Markov chain X^h with state space $h\mathbb{Z}$. We analyse this problem by introducing an auxiliary Lévy process with Gaussian component X^G. The estimate we derive makes use of the triangle inequality on the transition probabilities, informally written as

$$|X - X^h| \leq |X - X^G| + |X^G - X^h|,$$

so that the second term reduces to a case previously covered in section 1.2 with the only difference that additional caution is required since the Gaussian component of X^G goes to zero as h goes to zero.

Description of the method and proof of convergence

Let X be a Lévy process with characteristic triplet $(\mu, 0, \nu)$. Define a Lévy process with Gaussian component X^G, (μ^G, σ^G, ν^G). We choose the measure ν^G to be the truncated Lévy measure i.e. $\nu^G(\mathrm{d}x) = \nu(\mathrm{d}x)\mathbb{1}_{\mathbb{R}\setminus[-h,h]}$. Matching moments of X and X^G we obtain

$$\sigma^{G,2} = \int_{-1}^{1} x^2\, \nu(\mathrm{d}x) - \int_{-1}^{1} x^2\, \nu^G(\mathrm{d}x) = \int_{-h}^{h} x^2\, \nu(\mathrm{d}x) = L(h).$$

Summarizing, the characteristic triplet of X^G is given by $(\mu, L(h), \nu(\mathrm{d}x)\mathbb{1}_{\mathbb{R}\setminus[-h,h]})$.

Proposition 1.4.1. *Given a Lévy process X with characteristic triplet $(\mu, 0, \nu)$ and characteristic exponent $\Psi(\cdot)$, and $t > 0$, and assume that*

i) There exists a function $K(h) : (0, \infty) \to \mathbb{R}$ satisfying $K(h) = o(L(h)^{-\frac{1}{3}})$, as $h \to 0$,

ii) $\limsup_{h\to 0} \frac{1}{h} \int_{K(h)}^{\infty} |e^{t\Psi(p)}|\, \mathrm{d}p < \infty$, *iii)* $\int_{-\infty}^{\infty} |e^{t\Psi(p)}p^3|\,\mathrm{d}p < \infty$.

Then the Lévy process X^G with characteristic triplet $(\mu, L(h), \nu(\mathrm{d}x)\mathbb{1}_{\mathbb{R}\setminus[-h,h]})$ is such that

$$\sup_{x,y} |\rho_t(x,y) - \rho_t^G(x,y)| = \mathcal{O}(h), \quad h \to 0. \tag{1.4.1}$$

Before proving the proposition, we show the following Lemma:

Lemma 1.4.2. *Let Ψ^G be the characteristic exponent of X^G. If $ii)$ holds, then*

$$\limsup_{h\to 0} \frac{1}{h} \int_{K(h)}^{\infty} |e^{t\Psi^G(p)}|\mathrm{d}p < \infty. \tag{1.4.2}$$

Proof. The characteristic exponent Ψ^G can be written as

$$\begin{aligned}
\Psi^G(p) &= \Psi(p) - \int_{-h}^{h} (\mathrm{e}^{\mathtt{i}px} - 1 - \mathtt{i}px + \frac{1}{2}p^2x^2)\,\nu(\mathrm{d}x) \\
&= \Psi(p) - \int_{-h}^{h} \{(\cos px - 1 + \frac{1}{2}p^2x^2) + \mathtt{i}(\sin px - px)\}\,\nu(\mathrm{d}x).
\end{aligned}$$

An application of Fatou's lemma yields

$$\liminf_{p\to\infty} \int_{-h}^{h} (\cos px - 1 + \frac{1}{2}p^2x^2)\,\nu(\mathrm{d}x) \geq 0,$$

from which it follows, in particular, that $\int_{-h}^{h}(\cos px - 1 + \frac{1}{2}p^2x^2)\,\nu(\mathrm{d}x)$ is bounded from below as a function of p and hence

$$\limsup_{h\to 0} \frac{\int_{K(h)}^{\infty} |\mathrm{e}^{t\Psi^G(p)}|\,\mathrm{d}p}{h} = \limsup_{h\to 0} \frac{\int_{K(h)}^{\infty} |\mathrm{e}^{t(\Psi(p) - \int_{-h}^{h}(\cos px - 1 + \frac{1}{2}p^2x^2)\,\nu(\mathrm{d}x))}|\,\mathrm{d}p}{h} < \infty.$$

$\square$

Remark 1.4.3. Under the Orey condition, conditions i) and ii) are satisfied.

Proof of Proposition 1.4.1. The difference between characteristic exponents can be written as

$$\begin{aligned}\Psi(p)-\Psi^G(p) &= \int_{-h}^{h}(\mathrm{e}^{\mathtt{i}px}-1-\mathtt{i}px\mathbb{1}_{[-1,1]}(x))\,\nu(\mathrm{d}x)+\frac{1}{2}\int_{-h}^{h}x^2\,\nu(\mathrm{d}x)p^2\\ &= \int_{-h}^{h}(\mathrm{e}^{\mathtt{i}px}-1-\mathtt{i}px+\frac{1}{2}p^2x^2)\,\nu(\mathrm{d}x)\le M\int_{-h}^{h}|p^3x^3|\,\nu(\mathrm{d}x)\\ &\le Mh|p|^3L(h), \qquad (1.4.3)\end{aligned}$$

where M is a positive constant. By assumption $i)$ we have that $\sup_{p\in[-K(h),K(h)]}|\Psi(p)-\Psi^G(p)|\le hL(h)K^3(h)\to 0$, as h goes to zero. The difference between the transition probabilities is given by

$$\begin{aligned}|\rho_t(x,y)-\rho_t^G(x,y)| &= \frac{1}{2\pi}\left|\int_{\mathbb{R}}\mathrm{e}^{\mathtt{i}p(x-y)}\mathrm{e}^{t\Psi^G(p)}[\mathrm{e}^{t(\Psi(p)-\Psi^G(p))}-1]\,\mathrm{d}p\right|\\ &\le \int_{-K(h)}^{K(h)}|\mathrm{e}^{t\Psi^G(p)}||\mathrm{e}^{t(\Psi(p)-\Psi^G(p))}-1|\,\mathrm{d}p\\ &\quad+2\int_{K(h)}^{\infty}|\mathrm{e}^{t(\Psi(p)}|\,\mathrm{d}p+2\int_{K(h)}^{\infty}|\mathrm{e}^{t(\Psi^G(p)}|\,\mathrm{d}p. \qquad (1.4.4)\end{aligned}$$

Since we have uniform convergence of the characteristic exponents on $[-K(h),K(h)]$ we have that

$$|\mathrm{e}^{t(\Psi(p)-\Psi^G(p))}-1|\le 2t|\Psi(p)-\Psi^G(p)|\le 2MthL(h)|p|^3,$$

for all p in $[-K(h),K(h)]$. Furthermore, using the simple bound

$$|\mathrm{e}^{t\Psi^G(p)}|=|\mathrm{e}^{t(\Psi(p)-\int_{-h}^{h}(\mathrm{e}^{\mathtt{i}px}-1-\mathtt{i}px)\,\nu(\mathrm{d}x))}|\le C|\mathrm{e}^{t\Psi(p)}|, \qquad (1.4.5)$$

and hypothesis $iii)$, we have that the first integral is $\mathcal{O}(h)$ for some generic constant C. The second and third integral are $\mathcal{O}(h)$ by hypothesis $ii)$ and by the above Lemma respectively. □

Our strategy now consists of defining a continuous time Markov chain X^h that approximates the process X^G. The generator is defined to be

$$\mathcal{L}_h^G=\tilde{\mu}_h\nabla_h+\frac{\tilde{\sigma}_h^2}{2}\Delta_h+\Lambda_h,$$

where the operators expressed in coordinates are given by:

$$\Delta_h(x,y) = \frac{1}{h^2}\begin{cases} 1 & \text{if } y = x+h, x-y, \\ -2 & \text{if } y = x, \\ 0 & \text{otherwise.} \end{cases}$$

On the other hand the discrete gradient operator ∇_h as defined above, is not a Markov generator since it has negative elements off the diagonal. It can be expressed in coordinates as follows:

$$\nabla_h(x,y) = \nabla_h(\delta_y)(x) = \frac{1}{2h}\begin{cases} 1 & \text{if } y = x+h, \\ -1 & \text{if } y = x-h, \\ 0 & \text{otherwise.} \end{cases}$$

Define $A \subset l^2(h\mathbb{Z})$ to be the space of compactly supported sequences in $h\mathbb{Z}$ (i.e., sequences which are eventually zero). A is clearly dense in $l^2(h\mathbb{Z})$. We are going to define

$$\Lambda_h : l^2(h\mathbb{Z}) \supseteq A \to l^2(h\mathbb{Z}), \quad h > 0,$$

where A is the domain of Λ_h, in the following way: we start by defining Λ_h on the elements of the Hilbert basis $\{\delta_y, y \in h\mathbb{Z}\}$.

$$\Lambda_h(x,y) = \Lambda_h(\delta_y)(x) = f(y-x)h, \quad x \neq y,\, x \in h\mathbb{Z}, \tag{1.4.6}$$

We shall show that with the opportune choice of parameters $\tilde{\mu}_h$ and $\tilde{\sigma}^2$, X^h is a compound Poisson process with characteristic exponent given by

$$\Psi^h(p) = \mathtt{i}\tilde{\mu}_h \frac{\sin(hp)}{h} + \tilde{\sigma}_h^2 \frac{\cos(hp)-1}{h^2} + \sum_{k\neq 0} (\mathrm{e}^{\mathtt{i}khp} - 1) f^G(kh)h.$$

Furthermore, we can write $\Psi^h(p)$ as $\Psi^h(p) = \Psi_1^h(p) + \Psi_2^h(p)$, where

$$\Psi_1^h(p) = \mathtt{i}\tilde{\mu}_h \frac{\sin(hp)}{h} + \tilde{\sigma}_h^2 \frac{\cos(hp)-1}{h^2} + \sum_{J} (\mathrm{e}^{\mathtt{i}khp} - 1) f(kh)h.$$

In a similar fashion, write $\Psi^G(p) = \Psi_1^G(p) + \Psi_2^G(p)$, where

$$\Psi_1^G(p) = \mathtt{i}\mu p - \frac{1}{2}L(h)p^2 + \int_I (\mathrm{e}^{\mathtt{i}px} - 1 - \mathtt{i}ux)\,\nu^G(\mathrm{d}x).$$

Matching the first moments of X^G and X^h we obtain

$$\tilde{\mu}_h = \mu - \sum_J khf^G(kh)h,$$

while we do not match the second moment since we are aiming at a speed of convergence of $\mathcal{O}(h)$. In other words, we define

$$\tilde{\sigma}_h = \sigma^G.$$

In order to ensure that $\mathcal{L}_h^G$ is a generator we assume the following condition

a) $\mu - \frac{L(h)}{2h} - 2f(-h)h^2 \leq \sum_J khf^G(kh)h \leq \mu + \frac{L(h)}{2h} + 2f(h)h^2$

b) $\int_{[-1,1]\setminus[-h,h]} |xf^G(x) + x^2 f^{G\prime}(x)|\,\mathrm{d}x = o(1/h)$

In particular, $a)$ guarantees that $\mathcal{L}_h^G(x, x \pm h) \geq 0$ and $b)$ that

$$\left| \int_I x^2\,\nu^G(\mathrm{d}x) - \sum_J (kh)^2 f^G(kh)h \right|$$

goes to zero.

Remark 1.4.4. Note that the parameter $\tilde{\mu}_h$ can go to infinity but $\tilde{\mu}_h = \mathcal{O}(\frac{1}{h})$. Indeed,

$$\left| \sum_J khf^G(kh)h \right| \leq \frac{1}{h} \sum_J (kh)^2 f^G(kh)h$$

and the right-hand side is bounded for $b)$.

Remark 1.4.5. Observe that

- Condition $a)$ is satisfied whenever the Lévy measure ν is symmetric.
- the sum $\left|\sum_J khf^G(kh)h\right| \leq \frac{D}{h}$ for some constant D. In order to make condition $a)$ satisfied we can define the same approximation scheme on the lattice

$\{\ldots, -(c+1)h, -ch, ch, (c+1)h, \ldots\}$, for some $c \in \mathbb{N}$, instead of $h\mathbb{Z} \setminus \{0\}$. This will not change the rate of convergence of the scheme but will allow to have a smaller constant $D = D(c)$ and thus make condition $a)$ satisfied.

The difference between characteristic exponents can then be bounded by

$$|\Psi_1^h(p) - \Psi_1^G(p)| \leq |\tilde{\mu}_h| \left| \frac{\sin hp}{h} - p \right| + |\tilde{\sigma}_h^2| \left| \frac{\cos hp - 1}{h^2} + \frac{p^2}{2} \right| + \left| \int_I (\mathrm{e}^{\mathrm{i}px} - 1 - \mathrm{i}px)\nu^G(\mathrm{d}x) - \sum_J (\mathrm{e}^{\mathrm{i}khp} - 1 - \mathrm{i}khp) f^G(kh)h \right|.$$

We then bound each summand of the right-hand side like in previous cases: we have

$$\left| \frac{\sin hp}{h} - p \right| \leq \frac{h^2 p^3}{3!}, \quad \left| \frac{\cos hp - 1}{h^2} + \frac{p^2}{2} \right| \leq \frac{h^2 p^4}{4!}, \quad p \in \left(-\frac{1}{h}, \frac{1}{h} \right).$$

Furthermore, defining $\zeta(x) = (\mathrm{e}^{\mathrm{i}px} - 1 - \mathrm{i}px) f^G(x)$ and $I^\pm = I \cap \mathbb{R}^\pm$, $J^\pm = J \cap \mathbb{R}^\pm$, rewrite the last summand of the previous inequality as

$$\left| \int_I \zeta(x)\,\mathrm{d}x - \sum_J \zeta(kh)h \right| = \left| \int_{-\frac{h}{2}}^{\frac{h}{2}} \zeta(x)\,\mathrm{d}x + \int_{I \setminus [-\frac{h}{2}, \frac{h}{2}]} \zeta(x)\,\mathrm{d}x - \sum_J \zeta(kh)h \right|.$$

However, the first summand on the right-hand side is zero since f^G is zero on $[-h, h]$. Therefore, the previous expression can be bounded by

$$TV_{I \setminus [-\frac{h}{2}, \frac{h}{2}]}(\zeta') \frac{h^2}{8} \leq h^2 \int_{[-1-\epsilon, 1+\epsilon] \setminus [-\frac{h}{2}, \frac{h}{2}]} |\zeta''(x)|\,\mathrm{d}x,$$

for some $\epsilon > 0$, which in turn is less than

$$h \int_{[-1-\epsilon, 1+\epsilon] \setminus [-\frac{h}{2}, \frac{h}{2}]} |x^2 f^G(x) + x^3 f^{G\prime}(x) + x^4 f^{G\prime\prime}(x)|\,\mathrm{d}x.$$

Assume that

$$\int_{-1}^{1} |x^4 f''(x) + x^3 f'(x) + x^2 f(x)|\,\mathrm{d}x < \infty \quad \text{and} \quad \int_{\mathbb{R} \setminus [-1,1]} |f'(x)|\,\mathrm{d}x < \infty.$$

The difference $|\Psi_2^h(p) - \Psi_2^G(p)|$ is dealt with like in the previous sections and is $\mathcal{O}(h)$.

Summarizing, we have obtained that

$$
\begin{aligned}
|\Psi^h(p) - \Psi^G(p)| &\le |\tilde{\mu}_h| \frac{h^2 p^3}{3!} + |\tilde{\sigma}_h^2| \frac{h^2 p^4}{4!} + C_\alpha h + C_\beta h \\
&\le C \left(\frac{hp^3}{3!} + \frac{h^2 p^4}{4!} + h \right), \qquad (1.4.7)
\end{aligned}
$$

for some generic constant C. We are now ready to prove the following result

Theorem 1.4.6. *Let X be a Lévy process with characteristic triplet given by $(\mu, 0, \nu)$, where $\int_{[-1,1]} |x| \, \nu(\mathrm{d}x) = \infty$ and let $t > 0$. Assume that the Lévy measure is absolutely continuous with respect to the Lebesgue measure with Lévy density denoted by $\nu(\mathrm{d}x) = f(x)\mathrm{d}x$ and that f is $\mathcal{C}^2$ on $\mathbb{R} \setminus \{0\}$. Assume that conditions a) and b) for the well-definedness of the generator of the process X^h are satisfied. Let $\mathbf{P}_t^h$ be the transition probability of a stochastic process X^h that takes values in $h\mathbb{Z}$; for any two elements x and y in $h\mathbb{Z}$, let $\rho_t(x,y)$ denote the transition density of the process X. Assume that the following hold true:*

i) There exists a function $K(h) : (0,\infty) \to \mathbb{R}$ such that $K(h) = o(L(h)^{-\frac{1}{3}} \wedge h^{-\frac{1}{2}})$, as $h \to 0$,

ii) $\limsup_{h\to 0} \frac{1}{h} \int_{K(h)}^{\infty} |e^{t\Psi(p)}| \, \mathrm{d}p < \infty$,

iii) $\int_{-\infty}^{\infty} |e^{t\Psi(p)} p^4| \, \mathrm{d}p < \infty$,

iv) $\int_{-1}^{1} |x^4 f''(x) + x^3 f'(x) + x^2 f(x)| \, \mathrm{d}x < \infty$ and $\int_{\mathbb{R}\setminus[-1,1]} |f'(x)| \, \mathrm{d}x < \infty$.

Then the following holds for any x and y in $h\mathbb{Z}$:

$$
\sup_{x,y\in h\mathbb{Z}} \left| \rho_t(x,y) - \frac{1}{h} \mathbf{P}_t^h(x,y) \right| = \mathcal{O}(h), \qquad h \to 0.
$$

Proof. The difference between the transition probabilities can be estimated by

$$
\begin{aligned}
\left| \rho^G(x,y) - \frac{1}{h} \mathbf{P}^h(x,y) \right| \le & \int_{-K(h)}^{K(h)} |\mathrm{e}^{\Psi^G(p)}| |\mathrm{e}^{\Psi^h(p) - \Psi^G(p)} - 1| \, \mathrm{d}p \\
& + 2 \int_{K(h)}^{\frac{\pi}{h}} |\mathrm{e}^{\Psi^h(p)}| \, \mathrm{d}p + 2 \int_{K(h)}^{\infty} |\mathrm{e}^{\Psi^G(p)}| \, \mathrm{d}p,
\end{aligned}
$$

where $K(h)$ is the function in assumption i). Since we have uniform convergence of the characteristic exponents, namely $\sup_{p\in[-K(h),K(h)]}|\Psi^h(p)-\Psi^G(p)| = \mathcal{O}(h)$, the elementary bound $\mathrm{e}^x - 1 \leq 2x$, which holds for x small and positive, and assumption iii) yield

$$\begin{aligned}
\int_{-K(h)}^{K(h)} & |\mathrm{e}^{\Psi^G(p)}||\mathrm{e}^{\Psi^h(p)-\Psi^G(p)}-1|\,\mathrm{d}p \\
&\leq 2\int_{-K(h)}^{K(h)} |\mathrm{e}^{\Psi^G(p)}||\Psi^h(p)-\Psi^G(p)|\,\mathrm{d}p \\
&\leq 2h\int_{-K(h)}^{K(h)} |\mathrm{e}^{\Psi(p)}||p^4|\,\mathrm{d}p \\
&\leq 2h\int_{-\infty}^{\infty} |\mathrm{e}^{\Psi(p)}||p^4|\,\mathrm{d}p. \qquad (1.4.8)
\end{aligned}$$

The third summand $\int_{K(h)}^{\infty}|\mathrm{e}^{\Psi^G(p)}|\,\mathrm{d}p$ is $\mathcal{O}(h)$ by Lemma 1.4.2. We will make use of the following bound

$$|\mathrm{e}^{\Psi^h(p)}| \leq |\mathrm{e}^{\sum_{h\mathbb{Z}\setminus\{0\}}(\cos khp-1)f^G(kh)h}|.$$

Since

$$\left|\sum_{h\mathbb{Z}\setminus\{0\}}(\cos khp-1)f^G(kh)h - \int_{\mathbb{R}\setminus\{0\}}(\cos px-1)f^G(x)\,\mathrm{d}x\right| = \mathcal{O}(h), \qquad (1.4.9)$$

we then have

$$\int_{K(h)}^{\frac{\pi}{h}} |\mathrm{e}^{\Psi^h(p)}|\,\mathrm{d}p \leq \int_{K(h)}^{\frac{\pi}{h}} \mathrm{e}^{\int_{\mathbb{R}}(\cos px-1)f^G(x)\,\mathrm{d}x+Ch}\,\mathrm{d}p, \qquad (1.4.10)$$

where C is a generic constant. The right-hand side can be bounded from above by $C\int_{K(h)}^{\infty}\mathrm{e}^{\int_{\mathbb{R}}(\cos px-1)f(x)\,\mathrm{d}x}\,\mathrm{d}p$, where $f^G(x) = f(x)\mathbb{1}_{\mathbb{R}\setminus[-h,h]}(x)$. Observe now that the expression in (1.4.10) is less than or equal to $C\int_{K(h)}^{\infty}|\mathrm{e}^{\Psi(p)}|\,\mathrm{d}p$, which by hypothesis is $\mathcal{O}(h)$. $\square$

1.A Appendix

Preliminaries

A random variable taking values in $b + h\mathbb{Z}$ for some constants b and h is said to have a lattice distribution. If X is a lattice distribution and we can take $b = 0$ i.e. $\mathbb{P}(X \in h\mathbb{Z}) = 1$, X is said to be arithmetic. It is possible to show that the characteristic function ϕ of an arithmetic random variable is periodic of period $2\pi/h$, and

$$\mathbb{P}(X = x) = \frac{h}{2\pi} \int_{-\frac{\pi}{h}}^{\frac{\pi}{h}} \mathrm{e}^{-\mathrm{i}ux} \phi(u) \, \mathrm{d}u \quad \text{for } x \in h\mathbb{Z}.$$

For further information about arithmetic random variables and their properties check Durret [48] or Mukhin [99].

We shall now introduce the concept of infinitesimal generator and semigroup on an abstract Banach space. We will then restrict our attention to the important case of Lévy processes.

Definition 1.A.1. A family $(P_t)_{t\geq 0}$ of bounded linear operators on a Banach space $(B, \|\cdot\|)$ is called a strongly continuous semigroup if

$$P_t P_s = P_{t+s} \quad t, s \in [0, \infty)$$
$$P_0 = I$$
$$\lim_{t \downarrow 0} P_t f = f \quad f \in B,$$

where the limit in the last expression is to be considered with respect to the norm of the Banach space $\|\cdot\|$. Moreover, the semigroup is called a contraction if $\|P_t\| \leq 1$ for all t.

Definition 1.A.2. The infinitesimal generator L of a strongly continuous contraction semigroup (P_t) is defined by

$$Lf = \lim_{t \downarrow 0} \frac{P_t f - f}{t},$$

where the limit is once again with respect to the norm of B and f belongs to the subset of B such that the limit on the right-hand side of (1.A.1) exists. It is worth reminding that by virtue of the celebrated Hille-Yosida Theorem, the infinitiesmal generator determines the semigroup. That is, two strongly continuous contraction semigroups coincide if their infinitesimal generators coincide.

Remark 1.A.3. For any bounded linear operator L on B, the exponential e^L is defined to be a bounded linear operator such that

$$\mathrm{e}^L f = \sum_0^\infty \frac{L^k f}{k!}.$$

If a strongly continuous contraction semigroup (P_t) has a bounded operator L as its infinitesimal generator, then we can prove that $P_t = \mathrm{e}^{tL}$. Note that it is still possible to define functions of unbounded operators by finding a spectral representation of the operator.

The infinitesimal generator of a Lévy process is given by the following (Theorem 31.5 in [119]):

Theorem 1.A.4. *Let L be the infinitesimal generator of a Lévy process with characteristic triplet (γ, σ, ν). Then C_c^∞ is a core of L, $C_0^2 \subset \mathcal{D}(L)$, and, for $f \in C_0^2$*

$$\begin{aligned} Lf(x) = \gamma \frac{\mathrm{d}f}{\mathrm{d}x}(x) &+ \frac{1}{2}\sigma^2 \frac{\mathrm{d}^2 f}{\mathrm{d}x^2}(x) \\ &+ \int_{\mathbb{R}} (f(x+y) - f(x) - yf'(x)\mathbb{1}_{(-1,1)}(y))\, \nu(\mathrm{d}y). \end{aligned} \tag{1.A.1}$$

The first, second and third term are called respectively the drift, Brownian and jump part of the generator. Note how they are respectively given by a first and second order differential operator and by a superposition of difference operators. This enables to understand some of the path structure from the form of the generator and can be very useful for more general classes of processes where we no longer have a Lévy -Khinchine formula. From the general theory (see, for reference Ethier and Kurtz [53]), we know that convergence of generators (in some suitable sense) implies convergence of the semigroups. A more precise statement is given by the following theorem for the general case of Feller processes, of which Lévy processes are a subclass.

Theorem 1.A.5. *Let X_n be a sequence of Feller processes with semigroups $P_{n,t}$ and* *generators L_n, and let X be another Feller process with semigroup P_t and a generator L containing a core D. Then the following are equivalent:*

1. *If $f \in D$, there exists a sequence $f_n \in dom(L_n)$ such that $\|f_n - f\| \to 0$ and $\|L_n f_n - Lf\| \to 0$.*

2. *For every $t > 0$, $P_{n,t} f \to P_t f$ for every $f \in C_0$.*

3. *If $X_n(0) \to_d X(0)$ in distribution, then $X_n \to_d X$ in $\mathbf{D}$.*

Part 3 of the previous theorem also states that we have convergence of X^h to X as processes i.e., on the Skorokhod space $\mathbf{D}$; we will discuss this matter further in the sequel.

Approximation of indefinite Riemann integrals by Riemann sums

Let $[a, b]$ be a bounded closed interval. We take an partition Δ of $[a, b]$ defined by:

$$\Delta : a = s_0 \leq s_1 \cdots \leq s_{n-1} \leq s_n = b. \tag{1.A.2}$$

We denote by $D_n^{[a,b]}$ or simply D_n when it is clear from the context, the partition of $[a, b]$ defined by $s_i = a + i(b - a)/n$ and call it the regular partition. For a function f defined on $[a, b]$ and $s_{i-1} \leq \xi_i \leq s_i$, define the Riemann sum $R(f, \Delta, \xi_i)$ by

$$R(f, \Delta, \xi_i) = \sum_{i=1}^{n} (s_i - s_{i-1}) f(\xi_i). \tag{1.A.3}$$

The mesh $\|\Delta\|$ of the partition is defined as $\|\Delta\| = \max\{s_i - s_{i-1}; 1 \leq i \leq n\}$. The Riemann integral of f is defined as

$$\int_a^b f(x)\,\mathrm{d}x = \lim_{\|\Delta\| \to 0} R(f, \Delta, \xi_i), \tag{1.A.4}$$

and textbooks on calculus show that this limit exists for a continuous function f. It is of interest to consider limits of the expanded error term, such as

$$n \left| \int_a^b f(x)\,\mathrm{d}x - R(f, \Delta, \xi_i) \right|, \quad n^2 \left| \int_a^b f(x)\,\mathrm{d}x - R(f, \Delta, \xi_i) \right|, \tag{1.A.5}$$

as n goes to infinity. Recall that the total variation of a function f on the compact set $[a,b]$, denoted by $TV_{[a,b]}(f)$, is defined by $\lim_{\|\Delta\|\to 0}\sum_{\Delta}|f(t_{i+1})-f(t_i)|$. If $f\in\mathcal{C}^1[a,b]$, a simple application of Lagrange's theorem on the increments of the sum yields

$$TV_{[a,b]}(f)=\int_a^b|f'(x)|\,\mathrm{d}x. \tag{1.A.6}$$

Some well-known results are presented in the following theorem, where we denote the Riemann integral of f on $[a,b]$ by $\mathcal{I}$:

Theorem 1.A.6. *(a) If f is Riemann integrable on $[a,b]$, then*

$$R(f,\Delta,\xi_i)-\mathcal{I}=o(1),\quad n\to\infty.$$

(b) If f is a function of bounded variation on $[a,b]$, then

$$R(f,\Delta,\xi_i)-\mathcal{I}=\mathcal{O}(1/n),\quad n\to\infty.$$

(c) If f is absolutely continuous on $[a,b]$, then

$$R\left(f,D_n,\frac{s_{i-1}+s_i}{2}\right)-\mathcal{I}=o(1/n),\quad n\to\infty.$$

(d) If f is differentiable on $[a,b]$ and f' is of bounded variation on $[a,b]$, then

$$\left|R\left(f,D_n,\frac{s_{i-1}+s_i}{2}\right)-\mathcal{I}\right|\le TV_{[a,b]}(f')/8n^2\quad\textit{for all } n$$

where $TV_{[a,b]}(f')$ represents the total variation of f' on $[a,b]$.

(e) If f is a function of bounded variation on $[a,b]$, then

$$\left|R\left(f,D_n,\frac{s_{i-1}+s_i}{2}\right)-\mathcal{I}\right|\le TV_{[a,b]}(f)/2n\quad\textit{for all } n.$$

In light of (1.A.6), we can reformulate part (d) of Theorem 1.A.6 as follows: Assume

f is of class $\mathcal{C}^2[a,b]$ and $\int_a^b |f''(x)|\,\mathrm{d}x < \infty$, then

$$\left|R\left(f, D_n, \frac{s_{i-1}+s_i}{2}\right) - \mathcal{I}\right| \leq \int_a^b |f''(x)|\,\mathrm{d}x/8n^2 \quad \text{for all } n.$$

A proof of this theorem is given in [29], where it is also pointed out that the Riemann sums in points (c) and (d), $R\left(f, D_n, (s_{i-1}+s_i)/2\right)$, cannot be replaced by any of the $R(f, D_n, \xi_i)$. Moreover, [29] also shows that $\mathcal{O}(1/n^2)$ is the best possible estimate for $R\left(f, D_n, (s_{i-1}+s_i)/2\right) - I$ in (d), no matter how smooth the function is. Note that extending this to complex-valued integrands is straightforward since we simply apply the theory developed for real-valued integrands to real and imaginary parts, and no new ideas are involved.

Proof. Without loss of generality we prove the theorem on the interval $[0,1]$. Define the following function for all $n \in \mathbb{N}$:

$$s_n = \sum_{k=1}^{n} \mathbb{1}_{\left[\frac{k}{n}-\frac{1}{2n},1\right]}, \tag{1.A.7}$$

and consider the "saw tooth" function $v_n(t) = s_n(t) - nt$. For each n, we have $v_n(0) = v_n(1) = 0$, v_n jumps at each of the points $(k-1/2)n, k = 1, \ldots, n$ and it is linear in between. If f is Riemann integrable on $[0,1]$ then

$$R\left(f, D_n, \frac{s_{i-1}+s_i}{2}\right) - \mathcal{I} = \frac{1}{n}\int_0^1 f(t)\,\mathrm{d}v_n(t), \tag{1.A.8}$$

where the last integral is to be considered in the sense of Riemann-Stieltjes. Moreover, if f is of bounded variation on the unit interval, we are entitled to use the integration by parts formula which yields:

$$R\left(f, D_n, \frac{s_{i-1}+s_i}{2}\right) - \mathcal{I} = -\frac{1}{n}\int_0^1 v_n(t)\,\mathrm{d}f(t). \tag{1.A.9}$$

Since v_n lies in $[-1/2, 1/2]$ for all n, taking the modulus inside the integral gives

$$\left|R\left(f, D_n, \frac{s_{i-1}+s_i}{2}\right) - \mathcal{I}\right| \leq \frac{1}{2n}\int_0^1 |\mathrm{d}f(t)| = \frac{1}{2n} TV_{[0,1]}(f). \tag{1.A.10}$$

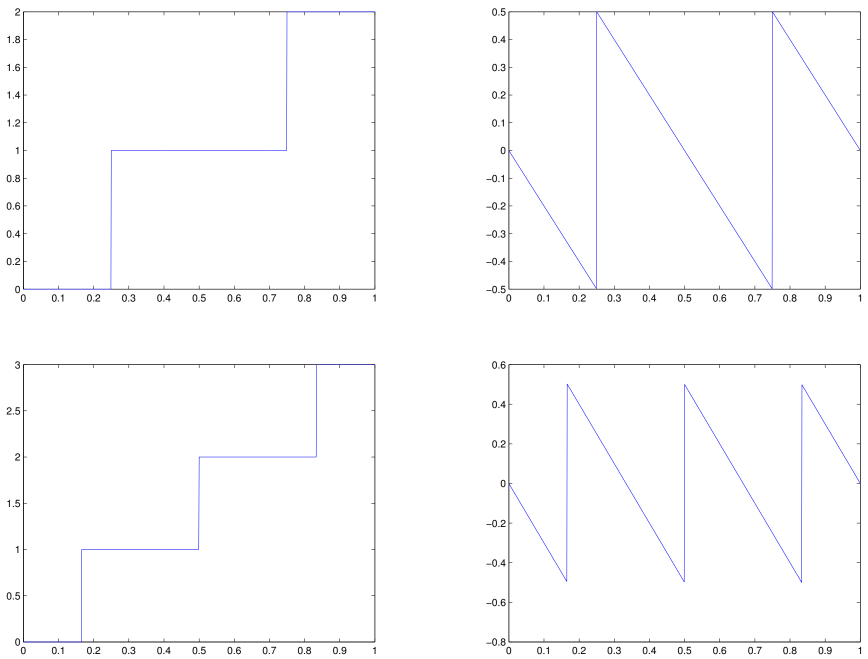

Figure 1.2: Graphs of s_2, s_3 and v_2, v_3.

If f is absolutely continuous on $[0,1]$, then f' is Lebesgue integrable there and

$$R_n(f) - \mathcal{I} = -\frac{1}{n}\int_0^1 f'(t)v_n(t)\,\mathrm{d}t, \tag{1.A.11}$$

so that by a proof similar to that of the Riemann-Lebesgue theorem, we see that $R_n(f) - I = o(1/n)$. Now, let f be differentiable with f' of bounded variation on $[0,1]$. Let $TV(f')$ denote the total variation of f' on $[0,1]$, and let

$$u_n(x) = \int_0^x v_n(t)\,\mathrm{d}t. \tag{1.A.12}$$

Then $u_n(k/n) = u_n(0)$ for $k = 1, \dots, n$, each u_n is periodic with period $1/n$ and

$$\max_{0 \le t \le 1} |u_n(t)| = \int_0^{\frac{1}{2n}} -v_n(t)\,\mathrm{d}t = \frac{1}{8n}. \tag{1.A.13}$$

But

$$R_n(f) - \mathcal{I} = -\frac{1}{n}\int_0^1 v_n(t)f'(t)\,\mathrm{d}t = \frac{1}{n}\int_0^1 u_n(t)\,\mathrm{d}f'(t). \tag{1.A.14}$$

Hence,

$$|R_n(f) - \mathcal{I}| \leq \frac{TV(f')}{8n^2}. \tag{1.A.15}$$

and this completes the proof of statement (d) of the Theorem. □

Convergence of processes

In the previous section we showed that we have convergence of the transition kernels of the approximating process to the transition kernel of the Lévy process. A natural question to ask is weather we have convergence in distribution of the marginals or convergence in distribution of the processes. Let us now recall what it means for a sequence of random elements of a general metric space to converge in distribution.

An $\mathcal{F}/\mathcal{E}$-measurable map from a probability space $(\Omega, \mathcal{F}, \mathbb{P})$ into a set E with σ-field $\mathcal{E}$ is called a random element of E. If E is a metric space the set of all bounded, continuous, $\mathcal{E}/\mathcal{B}(\mathbb{R})$-measurable, real valued (i.e. functionals) on E is denoted by $\mathcal{C}(E, \mathcal{E})$

Definition 1.A.7. A sequence $\{X^n\}$ of random elements on E converges in distribution to a random element X, denoted by $X^n \xrightarrow{\mathcal{L}} X$, if $\mathbb{E}[f(X^n)] \to \mathbb{E}[f(X)]$ for all f in $\mathcal{C}(E, \mathcal{E})$.

A sequence of probability measures $\mathbb{P}_n$ on $\mathcal{E}$ converges weakly to $\mathbb{P}$, denoted by $\mathbb{P}^n \xrightarrow{\mathcal{L}} \mathbb{P}$, if $\mathbb{P}_n(f) \to \mathbb{P}(f)$, for all f in $\mathcal{C}(E, \mathcal{E})$.

Note that this notion of convergence does not specify the limit random element uniquely, since, if X and Y have the same distribution, $X^n \xrightarrow{\mathcal{L}} X$ means the same as $X^n \xrightarrow{\mathcal{L}} Y$. The space $E = \mathcal{C}[0, 1]$ endowed with the uniform metric $\|\cdot\|_\infty$ is convenient when dealing with Brownian motion. In fact one of the main ingredients in the proof of the Donsker's invariance principle was Prohorov's Theorem which relates tightness of probability measures on $\mathcal{C}[0, 1]$ endowed with $\|\cdot\|_\infty$ to weak convergence. However, many important processes in practice are not continuous, due to discrete quantities involved. Let us then define the space $D[0, T]$ to be the set of càdlàg functions on $[0, T]$. In principle we could study this space using the uniform metric again, since all càdlàg functions on $[0, 1]$ are bounded and $\|\cdot\|_\infty$ is again a metric; let us consider two functions x, y in $D[0, T]$. We are given $a, a + \delta \in [0, T]$ and define $x(t) = \mathbb{1}\{t \geq a\}$ and $y(t) = \mathbb{1}\{t \geq a + \delta\}$.

Despite the fact that x and y coincide everywhere but on a small interval $[a, a+\delta)$, we have $\|x-y\|_\infty = 1$, for arbitrarily small δ. Since it makes sense to assume that x and y are 'close' to each other, we conclude that the uniform metric is inadequate to deal with discontinuous functions. The problem with the uniform metric here is that if we allowed for a deformation in time, the two functions would indeed be close (they would coincide in fact).This inspires the following definition:

Definition 1.A.8. Let Λ be the space of strictly increasing continuous functions $[0,T] \to [0,T]$. The Skorokhod metric σ on $D[0,T]$ is defined by

$$\sigma(x,y) = \inf_{\lambda \in \Lambda} (\|\lambda - I\|_\infty \vee \|x - y \circ \lambda\|_\infty)$$

for all x and y in $D[0,T]$ and where $I \in \Lambda$ is the identity function.

Thus, by definition, this distance is less than ϵ if there exists $\lambda \in \Lambda$ such that

$$\sup_{t \le T} \|\lambda(t) - t\| < \epsilon \quad \text{and} \quad \sup_{t \le T} \|x(t) - y(\lambda(t))\| < \epsilon.$$

The proof that σ is a metric contains no surprises[11]. The main difference between convergence in the uniform sense and convergence in the Skorokhod metric is appears most clearly at the discontinuity points of the limit function. If $\{x_n\}$ converges uniformly on compacta to x and if x has a jump at t_0, then each x_n, for n large, must have a jump of almost the same magnitude precisely at t_0. Skorokhod's metric still forces x_n to have a jump of almost the same magnitude, but not precisely at t_0. However, the Skorokhod metric and the uniform metric coincide on the subspace $\mathcal{C}[0,1]$. Necessary and sufficient conditions for convergence in distribution on $D[0,T]$ with the uniform and the Skorokhod metric, can be found in [113]. The following theorem says that for a Lévy process, convergence in distribution of the marginals ensures convergence in distribution of the process. For a proof, we refer to [69]:

Theorem 1.A.9. *Let $\{X^n\}$ and X be Lévy processes. If there is convergence in distribution of the marginals i.e.$X_1^n \xrightarrow{\mathcal{L}} X_1$ if and only if there is convergence of the processes*

[11] symmetry follows from the fact that λ has an inverse λ^{-1}: $\|x - y \circ \lambda\| = \|x \circ \lambda^{-1} - y\|$

$X^n \xrightarrow{\mathcal{L}} X$ in the Skorokhod topology.

Remark 1.A.10. Suppose we are given a Lévy process $X = \{X_t;\, t \geq 0\}$ and the continuous-time Markov chain $X^h = \{X^h_t;\, t \geq 0\}$, as described in the previous sections; the processes X and X^h are both processes with independent and stationary increments and can both be regarded as random elements taking values in $D[0, T]$. The proof of Theorem 1.2.13 hinges upon the fact that the characteristic exponent of the approximating process Ψ^h converges to the characteristic exponent of X as h goes to zero; an application of the Lévy continuity theorem guarantees that this is equivalent to having convergence of the one-dimensional distribution: $X^h_1 \xrightarrow{\mathcal{L}} X_1$. This yields the following simple result:

Corollary 1.A.11. *Let X be any Lévy process and consider a continuous-time Markov chain X^h with transition probability given by (1.2.11). Then $X^h \xrightarrow{\mathcal{L}} X$ in the Skorokhod topology.*

Chapter 2

Time change: Approximation of Subordinated Brownian Motion

2.1 Introduction

Literature review

The introduction of time change in probability theory can be traced back to the work of the French mathematician Doeblin who provided a way to transform a stochastic process into another by stretching and shrinking the time scale. This idea has proved very fruitful and a plethora of contributions have been obtained: Itô & McKean [67] carried out a profound study of the sample paths of diffusions and showed that time change allows to replace the study of diffusions by that of Brownian motion. Since then, time-changed Brownian motion emerged as a pivotal object in probability. We need to mention the theorems by Dubins & Schwarz [46] and Dambis [42] which show that every continuous martingale can be written as time-changed Brownian motion. Monroe [96] extended this important result and proved that every semimartingale can be written as a Brownian motion (possibly defined on some extended probability space) evaluated at a random time change.

Subordination is a type of time change introduced by Bochner [17, 18] in the context of harmonic analysis. In this case, the time change is given in terms of a non-decreasing Lévy process, independent of the driving process, called subordinator. Subordination transforms a temporally homogeneous Markov process into another temporally homogeneous Markov process, and hence a Lévy process into another Lévy process. We refer to Sato [119] and Bertoin [13, 14] for an exhaustive treatment of subordination of Lévy processes.

In the finance literature, time changed Brownian motion has first been used as a model for (logarithmic) asset prices by Clark [31]. He considered subordinated processes that evaluate geometric Brownian motion at a random time given by a process of independent and stationary increments which is itself independent of the Brownian motion. The economic interpretation of the subordinator is that it measures the level of the activity of the economy. German *et al.* [59] consider pure jump Lévy processes of finite variation as a model for the logarithm of asset prices. Such processes may be written as time-changed Brownian motions and the authors investigate the relationship between the time change and the price process. Note that many popular models in finance are based on time changed Brownian motion where the time change is chosen to be a subordinator: important examples are the Variance Gamma process [88], which can be represented as Brownian motion time changed by a gamma process and the Normal Tempered Stable process (including the Normal Inverse Gaussian process [9]), which can be written as Brownian motion time changed by a tempered stable subordinator. Simulation of stochastic processes has become increasingly important over the last two decades; in particular Monte Carlo methods have established themselves as a popular probabilistic method to solve PDEs (PIDEs) by virtue of the fact that the dimension of the problem has a very small effect on the complexity of the algorithm. However, simulation of Lévy processes can be challenging: we now list three of the main methods to simulate sample paths

i) The so-called *random walk approximation* relies on the fact that an infinitely divisible random variable can be written as a sum of i.i.d random variables and hence the law of a Lévy process at time t can be obtained as the convolution power of the law at time one. Simulation is then typically possible when the law of the increments lies in the same convolution closed class.

ii) *Series expansion* given in Rosiński [116] represent a Lévy process as an (infinite) sum of random variables and provide uniform along sample paths approximations;

while they are often easy to simulate, they have the drawback of working only in dimension one.

iii) The *small jump approximation* of a Lévy process consists in neglecting the jumps smaller than ϵ and in simulating the remaining compound Poisson process. When the intensity of small jumps is high (e.g. infinite variation paths) this method may lead to a significant error. It has been proposed in [6] to approximate such small jumps with a Brownian motion with matched variance; as it has been rigorously discussed, adding such a process improves the approximation in general.

There are several issues related to extending these simulation methods in higher dimension, as discussed in Cohen & Rosiński [32]; in fact contrary to the one-dimensional case, analytically tractable characterisations for simulation of increments of multidimensional Lévy processes are seldom available. The authors, extend the work of [6] to the multidimensional case and propose a method for simulating multivariate Lévy processes.

A general class of methods (sometimes denominated “lattice methods”), which found its origin in stochastic control problems with the work of Kushner [79], is to approximate a process with continuous state space with one with a discrete state space. Apart from being interesting in its own right as it concerns the interface between discrete and continuous processes, this approach has proved fruitful since complex functionals of the continuous state space process can often be approximated by a workable expression involving the discrete process. In the context of mathematical finance [93] exploit this technique to price barrier option in a very general framework. In his paper, Walsh [124] studies the detailed convergence of the binomial tree scheme; the analysis is carried out by embedding the tree scheme in the Black-Scholes diffusion using Skorokhod embedding. The work of [94] presents a continuous time Markov chain approximation of the density of a general Lévy process and provides rates of convergence. In this chapter we provide a novel Markov chain approximation for subordinated Brownian motion based on a comparison of the generator and that of the chain.

Contribution

In this chapter we study the problem of approximating the density of a multi dimensional Lévy process which has the form of a subordinated Brownian motion (with general covariance matrix Σ). Instead of approximating the jumps of the Lévy process we approximate the driving Brownian motion by a random walk in a classic fashion and then consider the subordinated random walk. We derive convergence of the transition probabilities of the subordinate random walk to that of the Lévy process, establish the optimal rate of convergence of the method and identify the exact constant. We believe an advantage of our method lies in the fact that we do not attempt to approximate the law of the subordinator but rather work solely at the level of the generators; this in turn provides an easy way to simulate transition probabilities since the approximating process is a continuous time Markov chain and the computation reduces to the exponentiation of a matrix. It is well known that the optimal rate of convergence of a Markov chain approximating linear Brownian motion is a quadratic function of the mesh of the lattice. It is a natural question to ask oneself whether such a rate can be extended to a more general class of Lévy processes: an answer in this direction has been given in [94] and in Chapter 2, where we prove that, in all cases apart from Lévy processes with Gaussian component and jumps of bounded variation, the rate worsens to linear or sub linear. In particular, for a pure jump Lévy process with Blumenthal-Getoor index $b > 1$, the rate becomes h^{2-b}, which is very slow for b close to 2, like in the case of a stable process with index $1 < b < 2$. In this chapter it is shown that for Lévy processes that take the form of subordinated Brownian motion, the quadratic rate can be maintained and substantial improvements on methods that discretise the Lévy measure can be achieved. An important difference of our work with the small jump approximation lies in the fact that our approach is based on Phillips theorem, thus avoiding a direct approximation of the small jumps.

This chapter is organised as follows: in section 2.2, we lay down some preliminaries, Phillips theorem and a characterisation of the processes arising as subordinated Brownian motion in terms of their characteristic triplet. In section 2.3, we explain the approximation procedure and derive the first term of the expansion of the density, which constitutes the main result of the chapter. At the end of the chapter we explain how to extend the validity of our result to the case of non rotationally invariant Lévy measure. The chapter closes with an appendix collecting some results on regularly varying functions.

2.2 Preliminaries

First of all, recall the following important Theorem (Sato [119], pp 196) is related in spirit to the seminal work of Phillips [111] and since it will be used in an important way in what follows.

Theorem 2.2.1 (Phillips)**.** *Let $\{Z_t; t \geq 0\}$ be a subordinator with Lévy measure ρ, drift β, and $\mathbb{P}_{Z_1} = \lambda$. Let $\{P_t; t \geq 0\}$ be a strongly continuous contraction semigroup of linear operator on a real Banach space $\mathbf{B}$ with infinitesimal generator L. Define*

$$Q_t f = \int_{[0,\infty)} P_s f \lambda^t(\mathrm{d}s), \qquad f \in \mathbf{B}.$$

Then $\{Q_t; t \geq 0\}$ is a strongly continuous contraction semigroup of linear operators on $\mathbf{B}$. Denote its infinitesimal generator by M. Then the domain of L, $\mathcal{D}(L)$, is a core of M and for all $f \in \mathcal{D}(L)$ we have:

$$Mf = \beta L f + \int_{[0,\infty)} (P_s f - f)\rho(\mathrm{d}s).$$

It is also possible to give an explicit expression for the characteristic function and characteristic triplet of a subordinated Lévy process.

Theorem 2.2.2. *Let $\{Z_t, t \geq 0\}$ be a subordinator (increasing Lévy process on $\mathbb{R}$) with Lévy measure ρ and drift β. Its Laplace transform can be written as*

$$\mathbb{E}[e^{-uZ_t}] = \mathrm{e}^{t\Phi(-u)}, \qquad u \geq 0,$$

where, for any complex number w with $\mathfrak{Re}(w) \leq 0$,

$$\Phi(w) = \beta w + \int_0^\infty (\mathrm{e}^{sw} - 1)\, \rho(\mathrm{d}s),$$

with

$$\beta \geq 0 \qquad \textit{and} \qquad \int_0^1 s\,\rho(\mathrm{d}s) < \infty.$$

Let X be a Lévy process (on $\mathbb{R}^d$) with generating triplet (μ, Σ, ν) and let its characteristic function be given by $\mathbb{E}[e^{\mathrm{i}uX_t}] = e^{t\Psi(u)}$. Suppose that X and Z are independent and define

$$Y_t = X_{Z_t}, \qquad t \geq 0.$$

Then Y is a Lévy process and its characteristic function is given by

$$\mathbb{E}[e^{\mathrm{i}\langle u, Y_t\rangle}] = e^{t\Phi(\Psi(u))}, \qquad u \in \mathbb{R}^d.$$

The generating triplet $(\mu^\#, \Sigma^\#, \nu^\#)$ of $\{Y_t\}$ is given by

$$\begin{aligned}
\mu^\# &= \int_0^\infty \rho(\mathrm{d}s) \int_{|x|\leq 1} x\,\mathbb{P}(X_s \in \mathrm{d}x),\\
\Sigma^\# &= \beta\,\Sigma,\\
\nu^\#(B) &= \mu\nu(B) + \int_0^\infty \mathbb{P}(X_s \in B)\,\rho(\mathrm{d}s) \quad B \in \mathcal{B}(\mathbb{R}^d).
\end{aligned}$$

Theorem 2.2.2 yields that the characteristic exponent of Y is given by the composition $\Psi_Y(u) = \Phi \circ \Psi_X(u)$. Note that this is expression is always well defined since

$$\mathfrak{Re}(\Psi_X(u)) = -\frac{\sigma^2 u^2}{2} + \int_{-\infty}^{\infty} (\cos(ux) - 1)\,\nu(\mathrm{d}x) \leq 0, \qquad \forall u \in \mathbb{R}.$$

Remark 2.2.3. Consider a Borel measure ν_α on $(0, \infty)$ with density

$$\nu_\alpha(\mathrm{d}x) = \frac{c}{x^{1+\alpha}} \mathbb{1}_{(0,\infty)}(x)\,\mathrm{d}x,$$

where $c > 0$ is a constant and $\mathrm{d}x$ denotes the Lebesgue measure. Then it is easy to show that ν_α is the Lévy measure of some Lévy process if and only if $\alpha \in (0, 2)$; the corresponding Lévy process is called stable process. Furthermore, ν_α is the Lévy measure of some subordinator if and only if $\alpha \in (0, 1)$. A stable subordinator with index $\alpha \in (0, 1)$ is a subordinator with zero drift and Lévy measure ν_α. It is easy to check that the Laplace exponent of a stable subordinator is proportional to

$$\Phi(-u) = -u^{\alpha} = \frac{\alpha}{\Gamma(1-\alpha)} \int_0^{\infty} (\mathrm{e}^{-ux} - 1) x^{-1-\alpha} \, \mathrm{d}x, \qquad \forall u \geq 0,$$

and where the constant of proportionality is given by some constant $c' > 0$. Without loss of generality we shall assume henceforth that Z is normalised so that $c' = 1$. Given Z and an independent standard Brownian motion W, an application of the strong Markov property shows that $Y_t = W_{Z_t}$ is a symmetric stable process of exponent 2α, where by symmetric we mean that Y_t has the same distribution as $-Y_t$. Further background on the rich and interesting class of stable processes can be found in Samorodnitsky & Taqqu [118].

Let W be Brownian motion and $Y_t = W_{Z_t}$ the subordinated process. Observe that

$$\left|\mathbb{E}[\mathrm{e}^{\mathrm{i}\langle u, Y_t\rangle}]\right| \leq \mathrm{e}^{\beta \Psi_W(u)},$$

which is integrable as a function of u on $\mathbb{R}^d$ and hence the probability measure $\mathbb{P}^{Y_t}$ is absolutely continuous with the Lebesgue measure (on $\mathbb{R}^d$) for all t. Sufficient conditions for the existence of a density in the case of a generic Lévy process are hard to find. For a partial solution we refer to Orey [102]. We now state the well-known conditions describing when a Lévy process is a subordinate Brownian motion (for a proof, we refer to Jacob [68], pp 190-192)

Proposition 2.2.4. *Let Y be a d-dimensional Lévy process with characteristic triplet (μ, Σ, ν). Then Y is a subordinate Brownian motion if and only if ν has a rotationally invariant density $x \mapsto u(|x|)$ such that $r \mapsto u(\sqrt{r})$ is a completely monotone[1] function on $(0, \infty)$, $\Sigma = cI_d$ for some $c \geq 0$ and $\mu = 0$.*

We will now define a continuous time Markov chain X^h with state space $h\mathbb{Z}$ that approximates W and we will do so by defining its generator $\frac{1}{2}\Delta_h$. The operator Δ_h defines

[1]A function f defined on $(0, \infty)$ is said to be completely monotone if it has derivatives of all orders and

$$(-1)^k f^{(k)}(t) > 0, \qquad t \in (0, \infty), k \in \mathbb{N}.$$

In particular, this implies that each completely monotone function on $(0, \infty)$ is positive, decreasing, convex and with concave first derivative. The interested reader is invited to consult [55].

a chain with equal probabilities of jumping one state up or one state down from the current state. Its expression in coordinates with respect to the standard basis $\{\delta_y; y \in h\mathbb{Z}\}$ is the following:

$$\Delta_h(x,y) = \Delta_h(\delta_y)(x) = \frac{1}{h^2}\begin{cases} 1 & \text{if } y = x+h, x-h, \\ -2 & \text{if } y = x, \\ 0 & \text{otherwise.} \end{cases} \tag{2.2.1}$$

The process X^h thus defined in terms of the generator Δ_h can be written as

$$X_t^h = \sum_{i=1}^{N(t)} \zeta_i,$$

where $\{N(t); t \geq 0\}$ is a Poisson process with rate $2/h^2$ and $\{\zeta_i\}$ is a sequence of i.i.d. Bernoulli random variables taking values $\pm h$ with probability $1/2$. For background material about Markov chains we refer to [100]. The process X^h is then a Lévy process, namely a compound Poisson process, and it is possible to show that its characteristic exponent is given by $\Psi_X(u) = \frac{2(\cos(hp)-1)}{h^2}$. Let us now define Υ^h to be the subordinated process of X^h, in other words: $\Upsilon_t^h = X_{Z_t}^h$. In the following we will also write X, Υ in place of X^h and Υ^h, the functional dependence on h being clear. Observe that, since we have convergence of the characteristic exponents, $\Phi(\Psi_{X^h}(p)) \to \Phi(\Psi_W(p))$ pointwise as h to zero, by virtue of Lévy continuity theorem, we obtain that $\Upsilon_t \xrightarrow{\mathcal{D}} Y_t$. Observe that, in the special case of a Lévy process, convergence in distribution of the marginals is equivalent to convergence in distribution of the processes on the Skorokhod space, see [113]. The rest of this chapter is devoted to studying the convergence of the transition kernels and deriving the exact rate and constants. Since Υ_t is a so-called arithmetic random variable, its transition probabilities are given by

$$\mathbf{P}_t^h(x,y) = \mathbb{P}(\Upsilon_t = y | \Upsilon_0 = x) = \frac{h}{2\pi}\int_{-\frac{\pi}{h}}^{\frac{\pi}{h}} \mathrm{e}^{\mathrm{i}p(x-y)}\mathbb{E}[\mathrm{e}^{\mathrm{i}p\Upsilon_t}]\,\mathrm{d}p, \quad x, y \in h\mathbb{Z}.$$

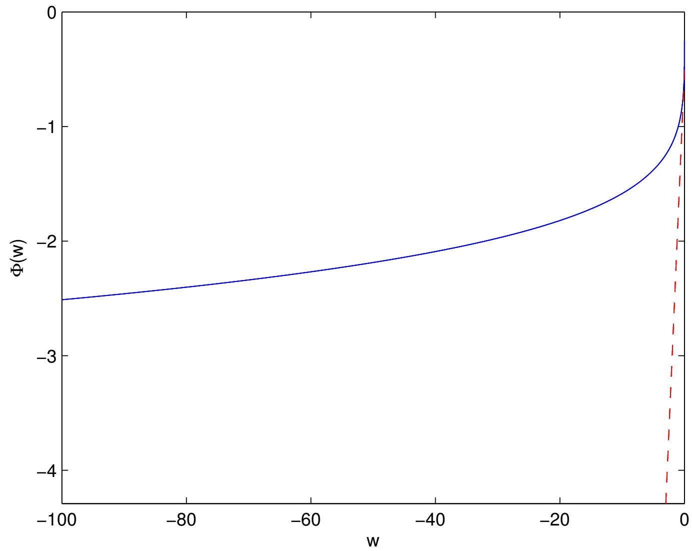

Figure 2.1: Example of the shape of the Laplace exponent $\Phi(w)$. The solid convex curve corresponds to the case of an alpha stable subordinator with index .2. The dashed line highlights the fact for a process with infinite mean we have $\Phi'(0-) = \infty$.

An application of the Fourier inversion formula yields the transition density of Y:

$$\rho_t(x, y) = \mathbb{P}(Y_t \in \mathrm{d}y | Y_0 = x) = \frac{1}{2\pi} \int_{-\infty}^{\infty} \mathrm{e}^{\mathrm{i}p(x-y)} \mathbb{E}[\mathrm{e}^{\mathrm{i}pY_t}] \, \mathrm{d}p.$$

2.3 Description of the method

In this section we tackle the extension of our previous result to the multidimensional case. This is particularly relevant since some of the main methods to simulate infinitely divisible distributions typically do not extend to higher dimensions. On this note, we refer to the literature concerning the so-called series expansion for the density of a Lévy process [116]. We first fix some notations: in the following $p = (p_1, \ldots, p_n)$ will denote the Euclidean vector in $\mathbb{R}^n$ and $\cdot$ the inner product on this space. Let W be Brownian motion in $\mathbb{R}^n$ i.e. an $\mathbb{R}^n$-valued random process with independent one dimensional Brownian motion as components. Let X be the $h\mathbb{Z}^n$-valued Markov chain consisting of n independent copies of the process defined by (2.2.1). Thus, it follows that the characteristic exponents are given by

$$\Psi_X(p) = \sum_1^n \frac{\cos(hp_i) - 1}{h^2}, \qquad \Psi_W(p) = -\frac{1}{2}\sum_1^n p_i^2, \qquad p \in \mathbb{R}^n.$$

Denote by $\circ : \mathbb{R}^n \times \mathbb{R}^n \to \mathbb{R}^n$ the Hadamard product, $p \circ q = (p_1 q_1, \ldots, p_n q_n)$. In the following, unless otherwise specified, $p^{\circ k} := p \circ \cdots \circ p$. Note the following inequality, which holds for all p in $(-\frac{1}{h}, \frac{1}{h})^n$:

$$\Psi_X(p) - \Psi_W(p) = \sum_{i=1}^n \frac{\cos hp_i - 1}{h^2} + \frac{p_i^2}{2} = \sum_{i=1}^n \frac{h^2 p_i^4}{4!} + \eta(p_i), \tag{2.3.1}$$

where $-\frac{(hp_i)^6}{6!} \leq \eta(p_i) \leq 0$, from which it follows that

$$0 \leq \sum_{i=1}^n \frac{h^2 p_i^4}{4!} - \frac{h^4 p_i^6}{6!} \leq \Psi_X(p) - \Psi_W(p) \leq \sum_{i=1}^n \frac{h^2 p_i^4}{4!}, \qquad \forall p \in \left(-\frac{1}{h}, \frac{1}{h}\right)^n.$$

The condition $p \in (-\frac{1}{h}, \frac{1}{h})$ ensures that $\cos(hp)$ can be written as an alternating series with monotonically decreasing terms. Equation (2.3.1) follows from the fact that the error of a partial sum of an alternating series with decreasing terms is bounded above by the absolute value of the first element in the series, after the partial sum.

Let $\overline{W}$ be Brownian motion on $\mathbb{R}^n$ with positive definite covariance matrix[2] $\Sigma = LL^*$, where $*$ denotes transposition and L is the unique (positive definite) lower triangular matrix given by the Cholesky decomposition. A simple computation yields $\overline{W} = LW$; furthermore, we set $\overline{X} = LX$. Consider the subordinated processes

$$\begin{aligned} \overline{Y}_t &= \overline{W}_{Z_t}, \quad Y_t = W_{Z_t}, \\ \overline{\Upsilon}_t &= \overline{X}_{Z_t}, \quad \Upsilon_t = X_{Z_t}, \end{aligned} \tag{2.3.2}$$

and their respective transition probabilities for x and y in $h\mathbb{Z}^n$:

$$\begin{aligned} \overline{\rho}_t(x,y) &= \mathbb{P}(\overline{Y}_t \in \mathrm{d}y | \overline{Y}_0 = x), \quad \rho_t(x,y) = \mathbb{P}(Y_t \in \mathrm{d}y | Y_0 = x), \\ \overline{\mathbf{P}}_t^h(x,y) &= \mathbb{P}(\overline{\Upsilon}_t = y | \overline{\Upsilon}_0 = x), \quad \mathbf{P}_t^h(x,y) = \mathbb{P}(\Upsilon_t = y | \Upsilon_0 = x). \end{aligned}$$

[2] A covariance matrix is always positive semi-definite and it is positive definite unless one variable is an exact linear combination of the others.

2.4 Main result

In this section we examine the convergence of the discrete probability kernel $\overline{\mathbf{P}}_t^h$ to the conditional probability density function $\overline{\rho}_t(x, y)$ of the Lévy process $\overline{Y}$. Observe that the limit as h goes to zero involves a passage from a discrete state space $h\mathbb{Z}^n$ to the continuum of $\mathbb{R}^n$. This limiting procedure has already been described in the previous chapter but we recall it here for the sake of completeness: consider a strictly decreasing sequence of positive real numbers $(h_n)_{n\in\mathbb{N}}$ such that: (i) $\lim_n h_n = 0$ and (ii) $\frac{h_i}{h_j}$ is an integer for every $j > i$. The second property ensures that this gives rise to a nested grid i.e.the lattice $h_j\mathbb{Z}^n$ contains the lattice $h_i\mathbb{Z}^n$. Since the domain of $\mathbf{P}_t^{h_i}$ is $h_i\mathbb{Z}^n \times h_i\mathbb{Z}^n$, we want to ensure that when fixing two points x, y on the lattice $h_i\mathbb{Z}^n$, they will belong to all subsequent $h_j\mathbb{Z}^n$ so that $\mathbf{P}_t^{h_j}(x, y)$ is well defined for all $j > i$ and can thus be regarded as a function of the index j. An example of such a sequence is given by $h_n = (\frac{1}{2})^n$. In the sequel, we will drop the subscript n when dealing with the lattice spacing h_n; being clear that $h_n \to 0$ as $n \to \infty$, we will just take h to be the parameter and write $h \to 0$. In particular, when dealing with limits where spacing goes to zero, we will be assuming that the parameter h visits all elements of the sequence $(h_n)_{n\in\mathbb{N}}$ for some index n onwards. We then have the following theorem

Theorem 2.4.1. *Let $\{Z_t; t \geq 0\}$ be a Lévy subordinator with Lévy measure ρ and drift β. Let the Laplace exponent of Z be given by the function Φ and assume that, in a neighbourhood of $+\infty$, the following growth condition is satisfied:*

$$\Phi(-u) \leq -(2(n+3)+1)\log(2u), \qquad u > 0. \tag{2.4.1}$$

Let $\Sigma = LL^*$ be a positive definite covariance matrix, $\overline{Y}, \overline{\Upsilon}$ be the processes defined as in (2.3.2) and denote their transition kernels by $\overline{\rho}_t, \overline{\mathbf{P}}_t^h$. Define the increasing function

$$U(x) = \int_0^x s\,\rho(\mathrm{d}s), \qquad x \geq 0,$$

and let us assume that U is regularly varying at infinity with index $1-\alpha$ ($\alpha \in (0, 1]$).

Then, for any x and y in the lattice $h\mathbb{Z}^n$, we have the following expansion

$$\overline{\rho}_t(x,y) = \frac{1}{h}\overline{\mathbf{P}}_t^h(x,y) + \epsilon(h),$$

where $\epsilon(h) = \epsilon(h,t,x,y)$ and, as $h \searrow 0$

$$\epsilon(h) = \frac{h^2 t}{(2\pi)^n}\int_{\mathbb{R}^n} \mathrm{e}^{\mathrm{i}q\cdot(x-y)}\mathrm{e}^{t\Phi(\Psi_W(L^*q))}\frac{(L^*q)^{\circ 4}\cdot\mathbf{1}}{4!}\left(\beta + \int_0^\infty \mathrm{e}^{s\Psi_W(L^*q)}\,\mathrm{d}U(s)\right)|\det L^*|\,\mathrm{d}q + \mathcal{O}(h^3).$$

Remark 2.4.2. The subordinators satisfying the assumptions in Theorem 2.4.1 include the subordinators Z with finite first moment and the stable and gamma subordinators, which are encountered in many applications. To see that the former is the case note that the first moment of Z_1 is finite if and only if

$$\int_1^\infty s\,\rho(\mathrm{d}s) < \infty,$$

which in particular means that the function U is slowly varying [3]. Note that requiring finite first moment is equivalent to $\Phi'(0-) < \infty$.

Remark 2.4.3. Explicit knowledge of the exact rate and constant allow us to use Romberg's extrapolation and thus obtain a method of order 3. Namely

$$\sup_{x,y\in 2h\mathbb{Z}^n}\left|\frac{4}{3}\frac{1}{h}\mathbf{P}_t^h(x,y) - \frac{1}{3}\frac{1}{2h}\mathbf{P}_t^{2h}(x,y) - \rho(x,y)\right| = \mathcal{O}(h^3).$$

We will now prove two preliminary lemmas. In order to ease the notations, we will denote the hypercube of side $2l$ by

$$H(l) = [-l,l]^n \quad \text{and by} \quad B(r) = \{x\in\mathbb{R}^n : |x| \le r\},$$

the ball centred at the origin with radius r.

Lemma 2.4.4. *Let us consider the following ODE:*

$$\frac{\mathrm{d}y}{\mathrm{d}x} = -x^{n+2}e^{-\Phi\left(-\frac{y^2}{2}\right)}. \tag{2.4.2}$$

[3] Any function with a finite limit at infinity is slowly varying at infinity.

If Φ satisfies (2.4.1), then there exists a solution $\bar{y}$ defined on a right neighbourhood of zero such that $\bar{y}(x) \to \infty$ and $\bar{y}(x) = o(\frac{1}{\sqrt[4]{x}})$, as $x \searrow 0$.

Proof. Let us first consider the autonomous ODE defined for $x > 0$

$$\frac{\mathrm{d}z}{\mathrm{d}x} = -\frac{1}{n+3}\mathrm{e}^{-\Phi\left(-\frac{z^2}{2}\right)}, \tag{2.4.3}$$

and let z be a solution to (2.4.3) defined on its maximal interval $I = (a, b)$, where $-\infty \leq a \leq b \leq \infty$. Observe that the ODE is autonomous and has no equilibrium solutions since $\mathrm{e}^{-\Phi(\Psi_W(u))} \neq 0$ for all u. It is then clear that $\lim_{x \to a^+} z(x) = \infty$ and $\lim_{x \to b^-} z(x) = -\infty$. By virtue of (2.4.1), $\int_{-\infty}^{\infty} \mathrm{e}^{\Phi\left(-\frac{z^2}{2}\right)} \mathrm{d}z < \infty$ we have that $b - a < \infty$ too and hence I is a finite interval. It then follows that $\bar{z}(x) = z(x - a)$ is a solution to (2.4.3) defined on $\bar{I} = (0, b - a)$ and such that $\lim_{x \to 0^+} \bar{z}(x) = \infty$. Moreover, the function $\bar{y}(x) = \bar{z}(x^{n+3})$ is a solution to (2.4.2) and once again we have $\lim_{x \to 0^+} \bar{y}(x) = \infty$. Let us take a couple of points $(x_0, \bar{y}_0)$ that belong to the graph of $\bar{y}$ i.e. $x_0 \in (0, \sqrt[n+3]{b - a})$ and $\bar{y}_0 = \bar{y}(x_0)$ and let us move backward in time. If the solution exists at time $x_1 < x_0$, then we have

$$\int_{\bar{y}_0}^{\bar{y}(x_1)} \exp\left(\Phi\left(-\frac{\bar{y}^2}{2}\right)\right) \mathrm{d}\bar{y} = \frac{1}{n+3}({x_0}^{n+3} - {x_1}^{n+3}).$$

For $x \to 0^+$ we have shown $\bar{y}(x) \to \infty$; by continuity and introducing an arbitrary $x > 0$ instead of x_0 we have

$$\int_{\bar{y}(x)}^{\infty} \exp\left(\Phi\left(-\frac{\bar{y}^2}{2}\right)\right) \mathrm{d}\bar{y} = \frac{x^{n+3}}{n+3}.$$

By (2.4.1), for sufficiently large $\bar{y}$ we have $\mathrm{e}^{\Phi\left(-\frac{\bar{y}^2}{2}\right)} \leq |\bar{y}|^{-4(n+3)-2}$, which in particular implies $\mathrm{e}^{\Phi\left(-\frac{\bar{y}^2}{2}\right)} = o\left(|\bar{y}|^{-4(n+3)-1}\right)$, as $y \to \infty$. Therefore, for x sufficiently close to zero and arbitrary positive ϵ, we have

$$\frac{x^{n+3}}{n+3} = \int_{\bar{y}(x)}^{\infty} \exp\left(\Phi\left(-\frac{\bar{y}^2}{2}\right)\right) \mathrm{d}\bar{y} \leq \epsilon \int_{\bar{y}(x)}^{\infty} \frac{1}{\bar{y}^{4(n+3)+1}} \mathrm{d}\bar{y} = \frac{\epsilon}{4(n+3)\bar{y}(x)^{4(n+3)}},$$

which in turn yields

$$\bar{y}^{4(n+3)}(x) \leq \frac{\epsilon}{4x^{n+3}},$$

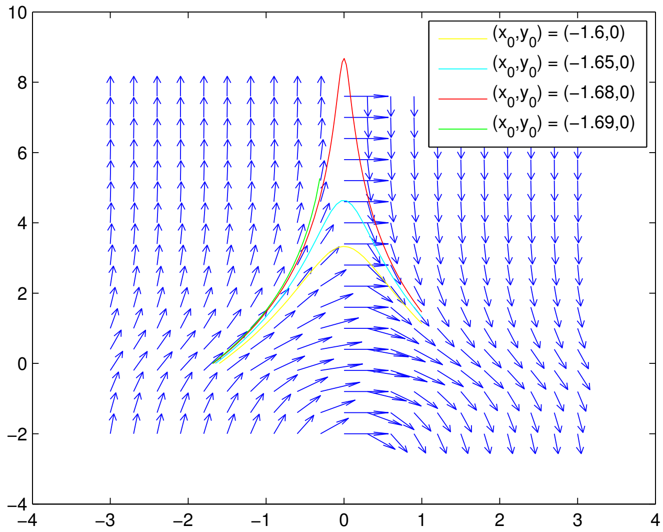

Figure 2.2: Some solutions to (2.4.2) for $n = 1$ and $\Phi(-u) = -u^\alpha$ $(u \geq 0)$ with $\alpha = .5$ and the corresponding direction field. Solutions corresponding to $(x_0, 0)$ with $|x_0| \leq \sqrt{2\sqrt{2}}$ (like the green one in the picture) lead to explosive behaviour.

which is what we set out to prove. $\square$

Lemma 2.4.5. *Let $\{Z_t; t \geq 0\}$ be a Lévy subordinator with Lévy measure ρ, drift β and Laplace exponent given by the function Φ. Define the increasing function*

$$U(x) = \int_0^x s\,\rho(\mathrm{d}s), \qquad x \geq 0,$$

and let us assume that U is regularly varying at infinity with index $1-\alpha$ ($\alpha \in (0,1]$) i.e.

$$U(x) \sim \frac{cx^{1-\alpha}l(x)}{\Gamma(2-\alpha)}, \quad x \to \infty,$$

where $c \geq 0$ and l varies slowly. Let $K(\cdot)$ be the solution to the ODE given by (2.4.2) and hence satisfying $K'(h) = -h^{n+2}e^{-\Phi\left(-K^2(h)/2\right)}$; then

$$\Phi(\Psi_X(p)) - \Phi(\Psi_W(p))\mathbb{1}_{H(K(h))} = h^2 \sum_{i=1}^n \left(\frac{p_i^4}{4!} + \bar{\eta}(p_i, h)\right) \times \times \left(\beta + \int_0^\infty \mathrm{e}^{s\Psi_W(p)}\,\mathrm{d}U(s)\right)(1 + \epsilon(p,h))\mathbb{1}_{H(K(h))}. \tag{2.4.4}$$

where $\epsilon(p,h)$ *and* $\bar{\eta}(p,h)$ *are* $o(1)$ *and* $|\bar{\eta}(p,h)| \leq \frac{h^2p^6}{6!}$. *Furthermore we have that* $\Phi(\Psi_X)$ *converges uniformly to* $\Phi(\Psi_W)$ *on the set* $H(K(h))$, *as* h *goes to zero:*

$$\sup_{p\in H(K(h))} \Phi(\Psi_X(p)) - \Phi(\Psi_W(p)) = o(h), \quad h \to 0. \tag{2.4.5}$$

Proof. By the definition of Φ, we can write

$$\Phi(\Psi_X(p)) - \Phi(\Psi_W(p)) = \beta(\Psi_X(p) - \Psi_W(p)) + \int_0^\infty (\mathrm{e}^{s\Psi_X(p)} - \mathrm{e}^{s\Psi_W(p)})\,\rho(\mathrm{d}s).$$

Using Fubini's theorem, the integral in the last expression can be written as

$$\int_0^\infty (\mathrm{e}^{s\Psi_X(p)} - \mathrm{e}^{s\Psi_W(p)})\,\rho(\mathrm{d}s) = \int_0^\infty s\int_{\Psi_W(p)}^{\Psi_X(p)} \mathrm{e}^{sy}\,\mathrm{d}y\,\rho(\mathrm{d}s) = \int_{\Psi_W(p)}^{\Psi_X(p)}\int_0^\infty s\mathrm{e}^{sy}\,\rho(\mathrm{d}s)\,\mathrm{d}y.$$

Since, for $p \in H(K(h))$, $0 \geq \Psi_X(p) \geq \Psi_W(p)$, we obtain the upper and lower bounds

$$\begin{aligned}(\Psi_X(p) - \Psi_W(p))\int_0^\infty s\mathrm{e}^{s\Psi_W(p)}\,\rho(\mathrm{d}s) &\leq \int_{\Psi_W(p)}^{\Psi_X(p)}\int_0^\infty s\mathrm{e}^{sy}\,\rho(\mathrm{d}s)\,\mathrm{d}y \\ &\leq (\Psi_X(p) - \Psi_W(p))\int_0^\infty s\mathrm{e}^{s\Psi_X(p)}\,\rho(\mathrm{d}s),\end{aligned}$$

and

$$\begin{aligned}(\Psi_X(p) - \Psi_W(p))\left(\beta + \int_0^\infty \mathrm{e}^{s\Psi_W(p)}\,\mathrm{d}U(s)\right) &\leq \Phi(\Psi_X(p)) - \Phi(\Psi_W(p)) \\ &\leq (\Psi_X(p) - \Psi_W(p))\left(\beta + \int_0^\infty \mathrm{e}^{s\Psi_X(p)}\,\mathrm{d}U(s)\right).\end{aligned}$$

Then, dividing by the left hand side,

$$1 \leq \frac{\Phi(\Psi_X(p)) - \Phi(\Psi_W(p))}{(\Psi_X(p) - \Psi_W(p))\left(\beta + \int_0^\infty \mathrm{e}^{s\Psi_W(p)}\,\mathrm{d}U(s)\right)} \leq \frac{\beta + \int_0^\infty \mathrm{e}^{s\Psi_X(p)}\,\mathrm{d}U(s)}{\beta + \int_0^\infty \mathrm{e}^{s\Psi_W(p)}\,\mathrm{d}U(s)}.$$

By a careful comparison of the characteristic functions we have that

$$\sup_{p\in H(K(h))} \frac{\beta + \int_0^\infty \mathrm{e}^{s\Psi_X(p)}\,\mathrm{d}U(s)}{\beta + \int_0^\infty \mathrm{e}^{s\Psi_W(p)}\,\mathrm{d}U(s)} = \frac{\beta + \int_0^\infty \mathrm{e}^{s\Psi_X(K(h)\cdot\mathbf{1})}\,\mathrm{d}U(s)}{\beta + \int_0^\infty \mathrm{e}^{s\Psi_W(K(h)\cdot\mathbf{1})}\,\mathrm{d}U(s)},$$

and the right hand side goes to one (for all $\beta \geq 0$) as h goes to zero by virtue of the dominated convergence theorem. This is equivalent to saying that, on $H(K(h))$

$$\Phi(\Psi_X(p)) - \Phi(\Psi_W(p)) = (\Psi_X(p) - \Psi_W(p)) \left(\beta + \int_0^\infty \mathrm{e}^{s\Psi_W(p)}\, \mathrm{d}U(s)\right)(1 + \epsilon(p,h)),$$

where $\lim_{h\to 0}\epsilon(p,h) = 0$ and $\epsilon(p,h) \le \bar\epsilon(h)$ with $\bar\epsilon$ a function that does not depend on p and $\bar\epsilon = o(1)$ as $h \searrow 0$. Observe that, although $\lim_{p\to 0}\int_0^\infty \mathrm{e}^{s\Psi_W(p)}\, \mathrm{d}U(s)$ can be infinite, the right hand side of the previous equation is bounded at zero since, by Karamata's tauberian theorem[4]

$$\int_0^\infty \mathrm{e}^{s\Psi_W(p)}\, \mathrm{d}U(s) \sim c(-\Psi_W(p))^{\alpha-1} l\left(-\frac{1}{\Psi_W(p)}\right), \qquad p \to 0,$$

and hence the integral cannot go to infinity faster than $(\sum_{i=1}^n p_i^2)^{-1} l(1/\sum_{i=1}^n p_i^2)$. Furthermore, we can write $\Psi_X(p) - \Psi_W(p) = \sum_{i=1}^n \frac{h^2 p_i^4}{4!} + h^2\bar\eta(p_i,h)$, with $\lim_{h\to 0}\bar\eta(p_i,h) = 0$[5], from which it follows that

$$\begin{aligned}\Phi(\Psi_X(p)) - \Phi(\Psi_W(p))\mathbb{1}_{H(K(h))} &= h^2 \sum_{i=1}^n \left(\frac{p_i^4}{4!} + \bar\eta(p_i,h)\right) \times \\ &\times \left(\beta + \int_0^\infty \mathrm{e}^{s\Psi_W(p)}\, \mathrm{d}U(s)\right)(1+\epsilon(p,h))\mathbb{1}_{H(K(h))}.\end{aligned}$$

Taking the supremum we obtain

$$\sup_{p\in H(K(h))} \Phi(\Psi_X(p)) - \Phi(\Psi_W(p)) \le h^2(1+\bar\epsilon(h)) \sup_{p\in H(K(h))} \frac{p^{\circ 4}\cdot \mathbf{1}}{4!}\left(\beta + \int_0^\infty \mathrm{e}^{s\Psi_W(p)}\, \mathrm{d}U(s)\right).$$

Since $\int_0^\infty \mathrm{e}^{s\Psi_W(p)}\, \mathrm{d}U(s) \to 0$ as $p \nearrow \infty$, there exists N (independent of h) such that for all p such that $|p| > N$, $\int_0^\infty \mathrm{e}^{s\Psi_W(p)}\, \mathrm{d}U(s) < 1$. Then, for h small enough

$$\begin{aligned}\sup_{p\in H(K(h))} \frac{p^4\cdot \mathbf{1}}{4!}\left(\beta + \int_0^\infty \mathrm{e}^{s\Psi_W(p)}\, \mathrm{d}U(s)\right) &\le \max_{p\in B(N)} \frac{p^{\circ 4}\cdot \mathbf{1}}{4!}\left(\beta + \int_0^\infty \mathrm{e}^{s\Psi_W(p)}\, \mathrm{d}U(s)\right) \\ &+ \max_{p\in H(K(h))\setminus B(N)} \frac{p^{\circ 4}\cdot \mathbf{1}}{4!}\left(\beta + \int_0^\infty \mathrm{e}^{s\Psi_W(p)}\, \mathrm{d}U(s)\right).\end{aligned}$$

[4] more specifically the abelian implication.

[5] and $\bar\eta(p,h) \le 0$ on $H(K(h))$.

Defining C to be the first term on the right hand side of the inequality[6], we obtain:

$$\sup_{p\in H(K(h))} \frac{p^{\circ 4}\cdot \mathbf{1}}{4!}\left(\beta+\int_0^\infty \mathrm{e}^{s\Psi_W(p)}\,\mathrm{d}U(s)\right) \leq \sup_{p\in H(K(h))} C+\frac{(\beta+1)}{4!}p^{\circ 4}\cdot\mathbf{1} = C+\frac{n(\beta+1)}{4!}K(h)^4.$$

Equation (2.4.5) then follows since $K(h)=o\left(\frac{1}{\sqrt[4]{h}}\right)$ and this completes the proof. □

Proof of Theorem 2.4.1. For typographical reasons we choose to prove the theorem for $t=1$ but exactly the same method of proof applies to any t. Let us begin by defining the function $K(\cdot)$ to be the solution to the ODE given by Lemma 2.4.2 and hence satisfying $K'(h)=-h^{n+2}\mathrm{e}^{-\Phi\left(-K(h)^2/2\right)}$. A simple application of Theorem 2.2.2 yields

$$(2\pi)^n\left(\rho_1(x,y)-\frac{1}{h}\mathbf{P}_1^h(x,y)\right) = \int_{H(\pi/h)} \mathrm{e}^{\mathrm{i}p\cdot(x-y)}\mathrm{e}^{\Phi(\Psi_X(p))}\,\mathrm{d}p-\int_{\mathbb{R}^n}\mathrm{e}^{\mathrm{i}p\cdot(x-y)}\mathrm{e}^{\Phi(\Psi_W(p))}\,\mathrm{d}p. \quad (2.4.6)$$

In Lemmas 2.4.4 and 2.4.5, we showed the existence of the function K which goes to infinity slowly enough so that we have uniform convergence i.e.

$$\sup_{p\in[-K(h),K(h)]^n} |\Phi(\Psi_X(p))-\Phi(\Psi_W(p))|\to 0, \qquad h\searrow 0.$$

This suggests to split our estimation of (2.4.6) in three parts: the integral over the hypercube in $\mathbb{R}^n$ with side $2K(h)$, $H(K(h))$ which we can control since we have uniform convergence of the characteristic exponents, and separate integrals over $A_1 := H(\frac{\pi}{h})\setminus H(K(h))$ and $A_2 := \mathbb{R}^n\setminus H(K(h))$, which will go to zero since integrals of integrable functions on sets whose measure goes to zero ($K(h)\to\infty$ and the integrands are integrable). In the following we will denote by $B(r)=\{x\in\mathbb{R}^n : |x|\leq r\}$ the ball centred at the origin with radius r and by S^{n-1} the area of the $n-1$ unit sphere (the sphere

[6] C is independent of h.

in $\mathbb{R}^n$ with radius one). The right hand side of (2.4.6) can be rewritten as

$$\int_{H(K(h))} \mathrm{e}^{\mathrm{i}p\cdot(x-y)} \left(\mathrm{e}^{\Phi(\Psi_X(p))} - \mathrm{e}^{\Phi(\Psi_W(p))}\right) \,\mathrm{d}p + \int_{A_1} \mathrm{e}^{\mathrm{i}p\cdot(x-y)}\mathrm{e}^{\Phi(\Psi_X(p))}\,\mathrm{d}p \\ - \int_{A_2} \mathrm{e}^{\mathrm{i}p\cdot(x-y)}\mathrm{e}^{\Phi(\Psi_W(p))}\,\mathrm{d}p. \quad (2.4.7)$$

The Fourier transform of even functions is a real even function. Consider the first integral in the above equation. An application of the elementary fact $\mathrm{e}^x - 1 = x(1 + q(x))$, where $q(x) = \mathcal{O}(x)$ yields:

$$\int_{H(K(h))} \mathrm{e}^{\mathrm{i}p\cdot(x-y)}\mathrm{e}^{\Phi(\Psi_W(p))} \left(\mathrm{e}^{\Phi(\Psi_X(p))-\Phi(\Psi_W(p))} - 1\right) \,\mathrm{d}p = \\ \int_{H(K(h))} \mathrm{e}^{\mathrm{i}p\cdot(x-y)}\mathrm{e}^{\Phi(\Psi_W(p))} \left(\Phi(\Psi_X(p)) - \Phi(\Psi_W(p)\right) \,\mathrm{d}p + \\ \int_{H(K(h))} \mathrm{e}^{\mathrm{i}p\cdot(x-y)}\mathrm{e}^{\Phi(\Psi_W(p))} \left(\Phi(\Psi_X(p)) - \Phi(\Psi_W(p)\right) q(\Phi(\Psi_X(p)) - \Phi(\Psi_W(p))\,\mathrm{d}p.$$

Observe that, since q is increasing for $x \geq 0$,

$$q(\Phi(\Psi_X(p)) - \Phi(\Psi_W(p)) \leq q(\sup_{H(K(h))} \Phi(\Psi_X(p)) - \Phi(\Psi_W(p)),$$

and hence $q(\Phi(\Psi_X(p)) - \Phi(\Psi_W(p)) = \mathcal{O}(h)$. Thus, for h small enough there exists a real constant M independent of h such that

$$\int_{H(K(h))} \mathrm{e}^{\mathrm{i}p\cdot(x-y)}\mathrm{e}^{\Phi(\Psi_W(p))} \left(\Phi(\Psi_X(p)) - \Phi(\Psi_W(p)\right) q(\Phi(\Psi_X(p)) - \Phi(\Psi_W(p))\,\mathrm{d}p \leq \\ Mh \int_{H(K(h))} \mathrm{e}^{\mathrm{i}p\cdot(x-y)}\mathrm{e}^{\Phi(\Psi_W(p))} \left(\Phi(\Psi_X(p)) - \Phi(\Psi_W(p)\right) \,\mathrm{d}p.$$

Using (2.4.4) and repeated applications of the dominated convergence theorem [7], we

[7] Using for instance the bound

$$\left|\mathbb{1}_{H(K(h))}(p)\mathrm{e}^{\Phi(\Psi_W(p))} \left(\beta + \int_0^\infty \mathrm{e}^{s\Psi_W(p)}\,\mathrm{d}U(s)\right) \epsilon(p,h)\right| \leq \mathrm{e}^{\Phi(\Psi_W(p))} \left(C + \frac{(\beta+1)}{4!}p^4\right) \bar{\epsilon}(h),$$

obtain

$$
\begin{aligned}
&\int_{H(K(h))} \mathrm{e}^{\mathrm{i}p\cdot(x-y)}\mathrm{e}^{\Phi(\Psi_W(p))}\left(\Phi(\Psi_X(p))-\Phi(\Psi_W(p))\right)\mathrm{d}p \\
&= h^2\int_{H(K(h)))} \mathrm{e}^{\mathrm{i}p\cdot(x-y)}\mathrm{e}^{\Phi(\Psi_W(p))}\left(\beta+\int_0^\infty \mathrm{e}^{s\Psi_W(p)}\,\mathrm{d}U(s)\right)\left(\frac{p^{\circ 4}\cdot\mathbf{1}}{4!}+\bar{\eta}(p,h)\right)(1+\epsilon(p,h))\,\mathrm{d}p \\
&\quad = h^2\int_{H(K(h))} \mathrm{e}^{\mathrm{i}p\cdot(x-y)}\mathrm{e}^{\Phi(\Psi_W(p))}\frac{p^{\circ 4}\cdot\mathbf{1}}{4!}\left(\beta+\int_0^\infty \mathrm{e}^{s\Psi_W(p)}\,\mathrm{d}U(s)\right)\mathrm{d}p+o(h^2) \\
&\qquad = h^2\int_{\mathbb{R}^n} \mathrm{e}^{\mathrm{i}p\cdot(x-y)}\mathrm{e}^{\Phi(\Psi_W(p))}\frac{p^{\circ 4}\cdot\mathbf{1}}{4!}\left(\beta+\int_0^\infty \mathrm{e}^{s\Psi_W(p)}\,\mathrm{d}U(s)\right)\mathrm{d}p+o(h^2),
\end{aligned}
$$

where the last equality follows since

$$
\begin{aligned}
&\left|\int_{\mathbb{R}^n\setminus H(K(h))} \mathrm{e}^{\mathrm{i}p\cdot(x-y)}\mathrm{e}^{\Phi(\Psi_W(p))}\frac{p^{\circ 4}\cdot\mathbf{1}}{4!}\left(\beta+\int_0^\infty \mathrm{e}^{s\Psi_W(p)}\,\mathrm{d}U(s)\right)\mathrm{d}p\right| \leq \\
&\left|\int_{\mathbb{R}^n\setminus B(K(h))} \mathrm{e}^{\mathrm{i}p\cdot(x-y)}\mathrm{e}^{\Phi(\Psi_W(p))}\frac{p^{\circ 4}\cdot\mathbf{1}}{4!}\left(\beta+\int_0^\infty \mathrm{e}^{s\Psi_W(p)}\,\mathrm{d}U(s)\right)\mathrm{d}p\right| \leq \\
&\qquad\qquad \frac{2(\beta+1)}{4!}\int_{\mathbb{R}^n\setminus B(K(h))} \mathrm{e}^{\Phi(\Psi_W(p))}p^{\circ 4}\cdot\mathbf{1}\,\mathrm{d}p,
\end{aligned}
$$

which is of order $o(h^2)$ as $h\to 0$. Indeed, note that by changing to polar coordinates and using the elementary estimate $|p_i|\leq r$

$$
\int_{\mathbb{R}^n\setminus B(K(h))} \mathrm{e}^{\Phi(\Psi_W(p))}p^{\circ 4}\cdot\mathbf{1}\,\mathrm{d}p \leq S^{n-1}\int_{r=K(h)}^\infty \mathrm{e}^{\Phi(-\frac{r^2}{2})}nr^4r^{n-1}\,\mathrm{d}r
$$

and, by l'Hôspital's rule and the definition of the function $K(h)$, we have

$$
\begin{aligned}
&\lim_{h\to 0}\frac{nS^{n-1}\int_{K(h)}^\infty \mathrm{e}^{\Phi(-\frac{r^2}{2})}r^{n+3}\,\mathrm{d}r}{h^2} = \\
&\qquad \lim_{h\to 0}\frac{-nS^{n-1}K'(h)\mathrm{e}^{\Phi(-\frac{K(h)^2}{2})}K^{n+3}(h)}{2h} = \lim_{h\to 0}\frac{h^{n+1}K^{n+3}(h)}{2} = 0.
\end{aligned}
$$

Next we show that the second integral in (2.4.7) is $o(h^2)$ as $h\to 0$. By an application of the triangle inequality we have

$$
\left|\int_{A_1} \mathrm{e}^{\mathrm{i}p\cdot(x-y)}\mathrm{e}^{\Phi(\Psi_X(p))}\,\mathrm{d}p\right| \leq \int_{A_1} \mathrm{e}^{\Phi(\Psi_X(p))}\,\mathrm{d}p.
$$

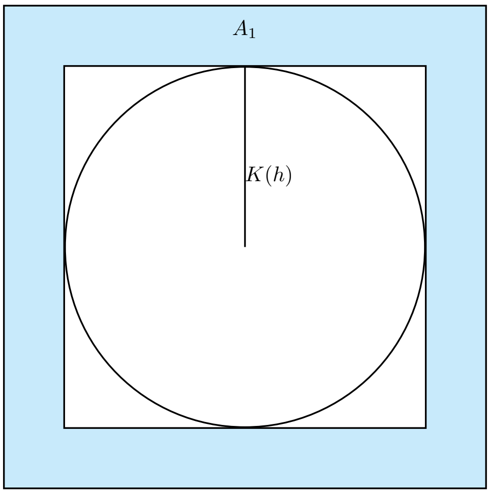

Figure 2.3: The regions $H(K(h))$ (in white) and A_1 (in cyan)

It is clear that $p \in H(\frac{\pi}{h}) \setminus H(K(h))$ if and only if $|p_k| \leq \frac{\pi}{h}$ for all k and there exists i such that $|p_i| > K(h)$. Since $p \mapsto \frac{\cos(hp)-1}{h^2}$ is monotonically decreasing on the interval $[0, \frac{\pi}{h}]$[8] we have that

$$\Psi_X(p) \leq \Psi_X(p^*), \qquad \forall p \in A_1,$$

where $p^* = (0, \ldots, 0, \pm K(h), 0, \ldots 0)$, the non-zero element being the ith entry and its sign being equal to the sign of p_i. Then, since Φ is monotonically increasing, we get

$$\int_{H(\frac{\pi}{h})\setminus H(K(h))} \mathrm{e}^{\Phi(\Psi_X(p))}\,\mathrm{d}p \leq \left(\frac{\pi}{h}\right)^n \mathrm{e}^{\Phi(\frac{\cos(hK(h))}{h^2})} \leq \left(\frac{\pi}{h}\right)^n \mathrm{e}^{\Phi\left(-\frac{K^2(h)}{2}+\frac{h^2K^4(h)}{4!}\right)}.$$

We now show that $\mathrm{e}^{\Phi\left(-\frac{K^2(h)}{2}+\frac{h^2K^4(h)}{4!}\right)} \leq C\mathrm{e}^{\Phi\left(-\frac{K^2(h)}{2}\right)}$ for some positive constant C in a neighbourhood of zero by proving that

$$\lim_{h\to 0} \mathrm{e}^{\Phi\left(-\frac{K^2(h)}{2}+\frac{h^2K^4(h)}{4!}\right)-\Phi\left(-\frac{K^2(h)}{2}\right)} = 1.$$

[8] and monotonically increasing on $[-\frac{\pi}{h}, 0]$.

By the mean value theorem we have that

$$\Phi\left(-\frac{K^2(h)}{2}+\frac{h^2K^4(h)}{4!}\right)-\Phi\left(-\frac{K^2(h)}{2}\right)=\Phi'(\xi)\frac{h^2K^4(h)}{4!},$$

where $-\frac{K^2(h)}{2}\le\xi\le-\frac{K^2(h)}{2}+\frac{h^2K^4(h)}{4!}$. Using the growth condition on $K(h)$ and the fact that $\Phi'(-\infty)=\beta$, we have that the above difference tends to zero as h goes to zero; the continuity of the exponential function proves our claim. By applying De l'Hôspital's rule to the following limit we obtain

$$\lim_{h\to0}\frac{\mathrm{e}^{\Phi\left(-\frac{K^2(h)}{2}\right)}}{h^{n+2}}=\lim_{h\to0}\frac{-\Phi'\left(-\frac{K^2(h)}{2}\right)K(h)K'(h)\mathrm{e}^{\Phi\left(-\frac{K^2(h)}{2}\right)}}{(n+2)h^{n+1}}$$
$$=\lim_{h\to0}\frac{\Phi'\left(-\frac{K^2(h)}{2}\right)K(h)h}{n+2}=0,$$

where in the previous equation we used the definition of the function K and the fact that $K=o(1/\sqrt[4]{h})$. In order to complete the proof, we are going to show that

$$\lim_{h\to0}\frac{\int_{A_2}\mathrm{e}^{\Phi(\Psi_W(p))}\,\mathrm{d}p}{h^2}=0,$$

which in particular imply that the numerators are bounded from above by h^2 on some interval around zero. De l'Hôspital's rule yields

$$\lim\frac{\int_{A_2}\mathrm{e}^{\Phi(\Psi_W(p))}\,\mathrm{d}p}{h^2}\le\lim\frac{\int_{\mathbb{R}^n\setminus B(K(h))}\mathrm{e}^{\Phi(\Psi_W(p))}\,\mathrm{d}p}{h^2}=$$
$$\lim\frac{S^{n-1}\int_{K(h)}^{\infty}\mathrm{e}^{\Phi(-\frac{r^2}{2})}r^{n-1}\,\mathrm{d}r}{h^2}=\lim\frac{-S^{n-1}K^{n-1}(h)K'(h)\mathrm{e}^{\Phi(-\frac{K(h)^2}{2})}}{2h}=0,$$

and the limit on the right hand side is equal to zero since $-K'(h)\mathrm{e}^{\Phi(\Psi_W(K(h)))}=h^{n+2}$.

To finish the proof, observe that

$$\overline{\rho}_t(x,y)=\rho_t(L^{-1}x,L^{-1}y),\qquad\overline{\mathbf{P}}_t^h(x,y)=\mathbf{P}_t^h(L^{-1}x,L^{-1}y),$$

where the matrix L is invertible since it is positive definite. By making the (linear) change of variable $q = (L^{-1})^* p$, we obtain the desired result. $\square$

Corollary 2.4.6. *Under the hypothesis of Theorem 2.4.1, the following uniform error estimate holds true*

$$\sup_{x,y\in h\mathbb{Z}} \left| \rho_t(x,y) - \frac{1}{h}\mathbf{P}_t^h(x,y) \right| = \mathcal{O}(h^2), \qquad h \searrow 0.$$

Furthermore, this convergence rate is exact in the sense that for any function $f = o(1)$, as $h \searrow 0$, the convergence rate $O(h^2 f(h))$ cannot be attained.

Proof. The claim that the error term is uniform with respect to the spacial variables x, y is easily obtained by using the triangle inequality and taking the absolute value inside the integral sign:

$$\begin{aligned}
&2\pi \left| \rho_1(x,y) - \frac{1}{h}\mathbf{P}_1^h(x,y) \right| \\
&= \left| \int_{-\frac{\pi}{h}}^{\frac{\pi}{h}} \mathrm{e}^{\mathrm{i}p(x-y)} \mathrm{e}^{\Phi(\Psi_X(p))}\, \mathrm{d}p - \int_{-\infty}^{\infty} \mathrm{e}^{\mathrm{i}p(x-y)} \mathrm{e}^{\Phi(\Psi_W(p))}\, \mathrm{d}p \right| \\
&\leq \int_{-K(h)}^{K(h)} \mathrm{e}^{\Phi(\Psi_W(p))} \left(\mathrm{e}^{\Phi(\Psi_X(p))-\Phi(\Psi_W(p))} - 1 \right) \mathrm{d}p \\
&+ 2 \int_{K(h)}^{\frac{\pi}{h}} \mathrm{e}^{\Phi(\Psi_X(p))}\, \mathrm{d}p + 2 \int_{K(h)}^{\infty} \mathrm{e}^{\Phi(\Psi_W(p))}\, \mathrm{d}p.
\end{aligned} \tag{2.4.8}$$

A calculation similar to that in the Proof of Theorem 2.4.1 then yields the claim. We shall now prove that the order of convergence h^2 is in fact the best attainable and we shall do so by contradiction. Take a function $f = o(1)$ and assume that the left hand side of (2.4.8) is $O(h^2 f(h))$. It suffices to show a contradiction for $n = 1$ and $x = y$ which we assume henceforth. Take the function K corresponding to the ODE $y' = -x^4 \exp(\Phi(-y^2/2))$ and such that the growth condition in (2.4.1) (with $n = 1$) is satisfied. Then we get, by a calculation similar to that in the proof of Theorem 2.4.1, that the second and third integral on the right hand side of (2.4.8) converge to zero like $O(h^3)$. Under our assumption this implies that $\int_{-K(h)}^{K(h)} \mathrm{e}^{\Phi(\Psi_W(p))} \left(\mathrm{e}^{\Phi(\Psi_X(p))-\Phi(\Psi_W(p))} - 1 \right) \mathrm{d}p$ is of order $O(h^2 g(h))$, where $g(h) = \max\{f(h), h\}$, since it can be bounded by a linear combination of the other terms. We shall obtain a contradiction by showing that

$$h^2 g(h) = o\left(\int_{-K(h)}^{K(h)} \mathrm{e}^{\Phi(\Psi_W(p))} \left(\mathrm{e}^{\Phi(\Psi_X(p))-\Phi(\Psi_W(p))} - 1\right) \, \mathrm{d}p \right), \tag{2.4.9}$$

which implies that $\int_{-K(h)}^{K(h)} \mathrm{e}^{\Phi(\Psi_W(p))} \left(\mathrm{e}^{\Phi(\Psi_X(p))-\Phi(\Psi_W(p))} - 1\right) \, \mathrm{d}p \le Ch^2 g(h)$, cannot hold (for all h) for positive C, no matter how big we choose it to be. Indeed, using equation (2.4.4) and the elementary inequality $\mathrm{e}^x - 1 \ge x$, valid for all positive x, we obtain

$$\begin{aligned}
&\int_{-K(h)}^{K(h)} \mathrm{e}^{\Phi(\Psi_W(p))} \left(\mathrm{e}^{\Phi(\Psi_X(p))-\Phi(\Psi_W(p))} - 1\right) \, \mathrm{d}p \\
&\qquad \ge \int_{-K(h)}^{K(h)} \mathrm{e}^{\Phi(\Psi_W(p))} \left(\Phi(\Psi_X(p)) - \Phi(\Psi_W(p))\right) \, \mathrm{d}p \\
&\qquad = h^2 \int_{-K(h)}^{K(h)} \mathrm{e}^{\Phi(\Psi_W(p))} \left(\beta + \int_0^\infty \mathrm{e}^{s\Psi_W(p)} \, \mathrm{d}U(s)\right) \left(\frac{p^4}{4!} + \bar{\eta}(p,h)\right) (1 + \epsilon(p,h)) \, \mathrm{d}p.
\end{aligned}$$

Therefore

$$\begin{aligned}
&\frac{1}{h^2 g(h)} \int_{-K(h)}^{K(h)} \mathrm{e}^{\Phi(\Psi_W(p))} \left(\mathrm{e}^{\Phi(\Psi_X(p))-\Phi(\Psi_W(p))} - 1\right) \, \mathrm{d}p \\
&\ge \frac{1}{g(h)} \int_{-K(h)}^{K(h)} \mathrm{e}^{\Phi(\Psi_W(p))} \left(\beta + \int_0^\infty \mathrm{e}^{s\Psi_W(p)} \, \mathrm{d}U(s)\right) \left(\frac{p^4}{4!} + \bar{\eta}(p,h)\right) (1 + \epsilon(p,h)) \, \mathrm{d}p,
\end{aligned}$$

and the limit of the right hand side goes to infinity since the integral converges to a finite positive value, proving (2.4.9). □

Remark 2.4.7. Observe that it is possible use the same technique to extend our result to the case of a subordinated linear Brownian motion $\mu t + \sigma W_t$, thus allowing the Lévy measure not to be rotationally invariant. In fact, it suffices to define the Markov chain X with generator given by

$$\mu \nabla_h + \frac{\sigma^2}{2} \Delta_h,$$

where

$$\Delta_h(x,y) = \frac{1}{h^2} \begin{cases} 1 & \text{if } y = x + h, x - y, \\ -2 & \text{if } y = x, \\ 0 & \text{otherwise.} \end{cases}$$

and

$$\nabla_h(x,y) = \frac{1}{2h}\begin{cases} 1 & \text{if } y = x+h, \\ -1 & \text{if } y = x-h, \\ 0 & \text{otherwise.} \end{cases}$$

Since ∇_h corresponds to a central difference approximation of the first derivative, the error term is of order $\mathcal{O}(h^2)$ and our result carries through. Here we chose not to include the drift term for clarity of exposition but we will be interested in getting the exact expression for the error term.

2.A Appendix

The basic theory of regular variation concerns the study of asymptotic relations of the type

$$\lim_{x\downarrow 0}\frac{f(\lambda x)}{f(x)} = g(\lambda), \qquad \forall\lambda\in(0,\infty),$$

for some function g, and their consequences. Aside from their intrinsic theoretical interest, regularly varying functions have proved to be useful in several branches of probability theory, for instance asymptotic first passage of Lévy processes.

Definition 2.A.1. A function $f:[0,\infty)\to(0,\infty)$ is said to be regularly varying at zero with index $\rho\in\mathbb{R}$ if, for all $\lambda>0$,

$$\lim_{x\downarrow 0}\frac{f(\lambda x)}{f(x)} = \lambda^\rho.$$

If the above limit holds as x tends to infinity then f is said to be regularly varying at infinity with index ρ. Moreover, if $\rho=0$, f is referred to as slowly varying.

Notice that any regularly varying function may always be written as

$$f(x) = x^\rho l(x),$$

where l is a slowly varying function. It is easy to check that any function that has strictly positive and finite limit at infinity is slowly varying and hence the set of regularly varying functions is clearly non-empty. Simple examples of slowly varying functions that tend to infinity are given by the logarithm and repeated composition of the logarithm with itself. We are now ready to state the following theorem that relates the behaviour of a function with that of its transform.

Theorem 2.A.2 (Karamata Tauberian Theorem)**.** *Let U be a non-decreasing right-continuous function on $\mathbb{R}$ with $U(x) = 0$ for all $x < 0$. If l varies slowly and $c \geq 0, \rho \geq 0$, the following are equivalent:*

$$U(x) \sim cx^{\rho} l(x)/\Gamma(1+\rho), \qquad x \to \infty, \tag{2.A.1}$$

$$\hat{U}(s) \sim cs^{-\rho} l(1/s), \qquad s \to 0^{+}, \tag{2.A.2}$$

where $\hat{U}(s) := \int_0^\infty e^{-sx}\,\mathrm{d}U(x)$ is the Laplace-Stieltjes transform of the measure U.

This theorem belongs to a class of classical results in which regular variation plays a role. The general idea behind it is that it is possible to put some condition on the measure at infinity to establish the behaviour of its transform near the origin and vice versa. [9] U is regularly varying at infinity with index $1-\alpha$ ($\alpha \in (0,1]$) i.e.

$$U(x) \sim \frac{cx^{1-\alpha} l(x)}{\Gamma(2-\alpha)}, \quad x \to \infty,$$

where $c \geq 0$ and l varies slowly. For a thorough discussion of the theory of regular variation the reader is referred to Bingham *et al.* [15]. An account of the theory and some of its applications in probability theory are given in Feller [56].

[9] More specifically, the passage from (2.A.1) to (2.A.2) is called Abelian and the converse Tauberian.

Chapter 3

Backward stochastic differential equations with jumps

Une intelligence qui, à un instant donné, connaîtrait toutes les forces dont la nature est animée et la situation respective des êtres qui la composent... rien ne serait incertain pour elle, et l'avenir, comme le passé, serait présent à ses yeux.

(P.S. Laplace, Essai philosophique sur les probabilités.)

3.1 Introduction

Literature review

Backward stochastic differential equations (BSDEs in the following) were pioneered in a paper by Bismut (1973) [16] who introduced the linear case in the stochastic version of the Pontryagin maximum principle. Other work in this area was done using linear BSDEs but their development took off only several years later with the seminal work of Pardoux & Peng (1990) [104], who introduced the general case and proved existence and uniqueness of solutions. Several people extended their results, in particular Antonelli [4], Ma, Protter & Yong [86], Pardoux & Peng [103, 105] and Peng [110, 108, 109, 106, 107]. The emergence of the theory of BSDEs has been accelerated by their use in applications: these equations have proved useful in control theory, stochastic finance theory, and for the probabilistic study of certain partial differential equations. Backward SDEs arise naturally in the problem of pricing and hedging contingent claims since the dynamics of wealth processes can be written naturally in such a form. These equations become then ideal tools to solve problems in a framework beyond the classical Black-Scholes-Merton paradigm. For instance, they have been used in the study of pricing contingent claims with trading constraints as in [41, 70] or in incomplete markets [52]. Work has also been done on the stochastic differential formulation of recursive utilities as in Duffie & Epstein [47] and in the context of utility maximisation, see Hu *et al.* [65], Klöppel & Schweizer [72], Mania & Schweizer [91] and Morlais [97]. The connection between the pricing of American option and reflected BSDEs was studied by El Karoui *et al.* [50]. An in-depth review article of applications of BSDEs in stochastic finance is given in [51]. Solutions of BSDEs provide a probabilistic representation to the solution of some nonlinear PDEs and their connection was studied by Pardoux & Peng [103, 105] and Peng [106, 107]. Crisan & Manolarakis provide an overview of probabilistic methods for semilinear PDEs in [38].

Numerical schemes to solve decoupled BSDEs can be divided in two categories, depending on whether they rely on a finite difference approximation of the associated PDE problem or whether they use a backward induction scheme typically involving the computation of conditional expectations. We do not discuss schemes that fall in the former category and refer the interested reader to the work of [86, 45, 95, 87]. The latter type of algorithm works backwards in time and tackles the problem by probabilistic means, i.e., avoiding the resolution of a PDE. Bally [7], Chevance [28] were the first to develop this kind of algorithm with a random time partition (given by the jump times of a Poisson process) and Zhang [125] and Bouchard & Touzi [21] independently proved convergence of this backward approach with deterministic partitions. A variety of algorithms based on this discretization has since then been developed and they differ mainly in the way the computation of the conditional expectation is carried out. These algorithms fall mainly in four categories, namely, the quantisation method, the Malliavin calculus regression method, the regression on function bases methods and the cubature method. The Malliavin calculus regression method was introduced by Bouchard & Touzi [21]; the algorithm makes use of the Malliavin integration by parts formula to obtain an expression for the conditional expectation which is amenable to Monte Carlo simulation. This approach builds upon the work done in [19] by Bouchard, Ekeland & Touzi. The regression on function bases method, (sometimes called regression based least squares Monte Carlo) was introduced by Gobet *et al.* [61] and computes the conditional expectation by projecting it on a subspace of function bases and then estimating the projection by Monte Carlo simulation. This method was first introduced by Longstaff & Schwartz [83] in the context of American option pricing. The cubature method was used by Crisan & Manolarakis [39, 40] and is based on the work of Lyons & Victoir [85]. This method relies on the fact that the solution of the BSDE can be expressed as an integral with respect to the law of the forward diffusion; the cubature of Lyons and Victoir provides an approximation of the Wiener measure and can therefore be used implicitly as an approximation of the law of the forward component of the BSDE. A different discretization scheme was developed by Bender & Denk in [11]: they propose a Picard type iteration method which has the advantage of avoiding nesting of conditional expectations through the time steps. They claim this to be a major advantage and makes it for a very implementable scheme. Higher order discretizations have also been studied by Chassagneaux & Crisan [26].

The development of discretization schemes for the fully coupled FBSDE case has been much slower. Most numerical algorithms for fully coupled FBSDEs exploit the relation to quasilinear parabolic PDEs via the so-called four step scheme by [86] and for further references see also [45, 95, 87]. The problem with this approach is that it becomes more and more difficult with increasing spacial dimension. To the best of our knowledge, the solution of this problem by probabilistic means has been tackled only by Delarue & Menozzi [43] and by Bender & Zhang [12].

While many methods have been developed for the classical case of decoupled FBSDE in a purely Brownian setting, much less work has been done in the presence of jumps. BSDEs driven not only by Brownian motion but also by a compensated Poisson random measure are considered by Barles, Buckdahn & Pardoux [8], Nualart and Schoutens [101], whereas existence and uniqueness of BSDEs driven by a general càdlàg martingale are addressed in El Karoui & Huang [49]. BSDEs with jumps play an important role in many optimal control problem, see for instance Tang & Li [121], Eyraud-Loisel [54] or Lim [81]. For the problem of utility maximisation in a setting with finite and infinite jump activity we refer to the article by Becherer [10] and Morlais [98]. Royer [117] studied BSDEs driven by Brownian motion and a Poisson random measure, and their application to g-expectations.

Contribution

We consider the problem of approximating the solution of a FBSDE driven by a Wiener process and an independent Poisson random measure. A common way to approximate FBSDEs is to discretise time and to replace the FBSDE by an appropriate backward stochastic difference equation (FBSΔE). Exact sampling methods from the distribution of the increments of the driving Poisson random measure are in general not available, which is an issue in the practical implementation of approximation schemes for BSDEs with jumps. Motivated by this observation, we present a weak approximation scheme for FBSDEs driven by a Wiener process and an independent Poisson random measure: we will consider the sequence of FBSΔEs driven by d_1-dimensional and d_2-dimensional random walks approximating a Brownian motion and a pure-jump Lévy martingale. We obtain a solution of the FBSΔE and give heuristic explanation of its convergence to the solution of the FBSDE with jumps. We also allow the driver to take a general functional form which is encountered in many applications. Convergence results for BSDEs driven by general random walks have been obtained in Briand *et al.* [24, 23], Madan *et al.* [90], and Cheridito & Stadje [27] typically using results on convergence of filtrations (see, e.g., Coquet *et al.* [35]).

Aazizi [1] recently has extended convergence results of Bouchard & Ellie [20] from a pure jump setting to an infinite jump activity process for forward backward SDEs with driver functions taking the form $f(t, y, \int_{\mathbb{R}^{d_2}\setminus\{0\}} \rho(x)\tilde{z}(x)x\nu(\mathrm{d}x))$, for a bounded functional ρ.

This chapter is structured in the following way: section 3.2 provides a review of classical results in the Brownian settings; section 3.3 presents a formulation and solution of the problem in discrete time; a heuristic motivation of the weak convergence is also given. in section 3.4 we put forward a numerical scheme for the solution of the FBSΔE using a regression on function bases for the computation of the conditional expectation. Extension to the case of a (decoupled) forward-backward SDEs are also presented. Finally, in section 3.5 we apply our method to compute numerically some non-linear expectations and check the behaviour of the solution qualitatively.

3.2 Backward stochastic differential equations

In this section we are going to present a brief overview to the theory of Backward SDEs and their extensions. Let us consider a probability space $(\Omega, \mathcal{F}, \mathbb{P})$ and an $\mathbb{R}^n$-valued Brownian motion W. First we fix some notations.

- $\mathbb{F} = \{\mathcal{F}_t; t \in [0,T]\}$, the filtration generated by Brownian motion and augmented and $\mathcal{P}$ the σ-field of predictable sets of $\Omega \times [0,T]$.

- L^2_T, the space of all $\mathcal{F}_T$-measurable random variables $X : \Omega \to \mathbb{R}^d$ satisfying

$$\|X\|^2_{L^2_T} := \mathbb{E}[|X|^2] < \infty.$$

- $\mathcal{H}^2(\mathbb{R}^d)$, the space of predictable processes $\phi : \Omega \times [0,T] \to \mathbb{R}^d$ such that

$$\|\phi\|^2_{\mathcal{H}^2} := \mathbb{E}\left[\int_0^T |\phi_t|^2 \,\mathrm{d}t\right] < \infty.$$

When no confusion arises we will denote $\mathcal{H}^2(\mathbb{R}^d)$ by $\mathcal{H}^2$.

A Backward SDE is written in differential form as

$$-\,\mathrm{d}Y_t = f(t, Y_t, Z_t)\,\mathrm{d}t - Z_t^* \mathrm{d}W_t, \qquad Y_T = \xi,$$

or equivalently

$$Y_t = \xi + \int_t^T f(s, Y_s, Z_s)\,\mathrm{d}s - \int_t^T Z_s^* \mathrm{d}W_s,$$

where the data is a pair (f, ξ) called the generator (sometimes driver) and terminal value such that $\xi : \Omega \to \mathbb{R}^d$ is $\mathcal{F}_T$-measurable and $f : \Omega \times \mathbb{R}^+ \times \mathbb{R}^d \times \mathbb{R}^{n\times d} \to \mathbb{R}^d$ and is $\mathcal{P} \otimes \mathcal{B}^d \otimes \mathcal{B}^{n\times d}$-measurable[1]. A solution is a pair (Y, Z) such that $\{Y_t; t \in [0,T]\}$ is a continuous $\mathbb{R}^d$-valued adapted process and $\{Z_t; t \in [0,T]\}$ is an $\mathbb{R}^{n\times d}$-valued predictable process satisfying $\int_0^T |Z_s|^2 \mathrm{d}s < \infty$, $\mathbb{P}$-a.s. Heuristically, one starts at the final time T with

[1] Note that the generator itself is possibly random.

$Y_T = \xi$ and then Y is constructed backwards in time for the whole of the interval $[0, T]$. At first glance, this problem seems ill-posed since finding a solution amounts to solving an equation in two unknowns. It is the constraint that the processes (Y, Z) are adapted to the filtration $(\mathcal{F})_{t\in[0,T]}$ that makes it possible to find a solution to the problem; for instance, in the case of $f = 0$, the pair $Y_t = \xi$ and $Z \equiv 0$ is not admissible. Let us consider the process defined by $Y_t = E[\xi|\mathcal{F}_t]$; this is an adapted process (martingale) satisfying the final condition. However, by virtue of the martingale representation theorem, Y_t satisfies

$$Y_t = Y_0 + \int_0^t Z_s \, \mathrm{d}W_s,$$

or, in differential notations: $dY_t = Z_t \, \mathrm{d}W_t$, $Y_T = \xi$. It is the freedom to choose the process Z that makes it possible to construct an adapted process Y and hence to find a solution to the problem.

Suppose that $\xi \in L^2_T$, $f(\cdot, 0, 0) \in \mathcal{H}^2$ and f is uniformly Lipschitz in its spacial variables, i.e., there exists a constant $C > 0$ such that $\mathbb{P} \otimes \mathrm{d}t$-a.s.

$$|f(\omega, t, y_1, z_1) - f(\omega, t, y_2, z_2)| \leq C(|y_1 - y_2| + |z_1 - z_2|) \qquad \forall (y_1, z_1), \forall (y_2, z_2).$$

Then (f, ξ) are said to be standard data, or parameters, for the BSDE.

The next theorem is a fundamental result, proving existence and uniqueness.

Theorem 3.2.1 (Pardoux-Peng,1990). *Given standard parameters* (f, ξ), *there exists a unique pair* $(Y, Z) \in \mathcal{H}^2(\mathbb{R}^d) \times \mathcal{H}^2(\mathbb{R}^{n\times d})$ *which solves (3.2.1).*

It is worth noting that it is possible to relax the Lipschitz condition on the generator f by simply requiring continuity in (y, z) and assuming some growth conditions on (y, z): we refer to [80, 73] for further details. The martingale representation property of W is key to proving that such equations have got solutions [2].

[2] If ξ is a scalar random variable in L^2_T, then $Y_t = E[\xi|\mathcal{F}_t^W]$ is a square integrable martingale that can be represented as a stochastic integral with respect to W of the unique adapted process $Z \in \mathcal{H}^2$:

$$\begin{aligned} \xi &= \mathbb{E}[\xi] + \int_0^T Z_s \, \mathrm{d}W_s, \\ Y_t &= \mathbb{E}[\xi] + \int_0^t Z_s \, \mathrm{d}W_s = \xi - \int_t^T Z_s \, \mathrm{d}W_s. \end{aligned}$$

We now state a very useful comparison principle for BSDEs.

Theorem 3.2.2 (Comparison Theorem). *Let (f^1, ξ^1) and (f^2, ξ^2) be two standard parameters of BSDEs, and let (Y^1, Z^1) and (Y^2, Z^2) be the associated square integrable solutions. Suppose that*

- $\xi^1 \geq \xi^2$ $\mathbb{P}$-*a.s.*,

- $\delta_2 f_t := f^1(t, Y^2_t, Z^2_t) - f^2(t, Y^2_t, Z^2_t) \geq 0, \quad \mathbb{P} \otimes \mathrm{d}t$-*a.s.*,

Then we have that almost surely for any time t, $Y^1_t \geq Y^2_t$.

It is now possible to deduce a sufficient condition for the non negativity of the BSDE solution.

Corollary 3.2.3. *If $\xi \geq 0$ a.s. and $f(t, 0, 0) \geq 0$ $\mathbb{P} \otimes \mathrm{d}t$-a.s., then $Y \geq 0$ $\mathbb{P} \otimes \mathrm{d}t$-a.s..*

This is an immediate consequence of the comparison theorem with $(f^2, \xi^2) = 0$, whose solution is obviously $(Y^2, Z^2) = 0$.

Connection with PDE and nonlinear Feynman-Kac formulae

Consider the linear parabolic PDE

$$\begin{aligned} &\partial_t u + \mathcal{L}u + f(t, x) = 0, \quad (t, x) \in [0, T] \times \mathbb{R}^n \\ &u(T, x) = g(x), \quad x \in \mathbb{R}^n. \end{aligned} \tag{3.2.1}$$

It is a well-known result that

$$v(t, x) = \mathbb{E}\left[g(X^{t,x}_T) + \int_t^T f(s, X^{t,x}_s)\,\mathrm{d}s \right],$$

where $\{X^{t,x}_s; s \in [t, T]\}$ is a solution to the SDE

$$\begin{aligned} &\mathrm{d}X_s = b(X_s)\mathrm{d}s + \sigma(X_s)\mathrm{d}W_s, \\ &X_t = x. \end{aligned}$$

is a solution to (3.2.1), where $\mathcal{L}$ is the infinitesimal generator of the process (3.2.2), which in this case is given by

$$\mathcal{L}v = \sum_i b_i(x)\partial_i v + \frac{1}{2}\sum_{i,j}(\sigma\sigma^*)_{i,j}(x)\partial^2_{i,j}v.$$

In this chapter, we seek to extend this probabilistic representation to solution of semi-linear PDEs and in order to do this we introduce the concept of Forward-Backward SDEs. To be more precise, we consider solutions of backward SDEs associated with some forward classical stochastic differential equations. Suppose that the randomness of the data (f, ξ) comes from the state of the forward equation. When the initial conditions of the forward equations (t, x) are taken into account, the backward solution (Y, Z) can be thought of as a parametrized BSDE, where the parameters are the initial conditions (t, x).

Suppose that f and Ψ are $\mathbb{R}^d$-valued Borel function defined on $[0, T] \times \mathbb{R}^p \times \mathbb{R}^d \times \mathbb{R}^{n\times d}$ and $\mathbb{R}^p$ respectively. Furthermore, suppose b and σ are $\mathbb{R}^p$-valued and $\mathbb{R}^{p\times n}$-valued functions defined on $[0, T] \times \mathbb{R}^p$. Note that now we assume the generator of the BSDE to be deterministic. For any $(t, x) \in [0, T] \times \mathbb{R}^p$ denote by $\{X^{t,x}_s; s \in [t, T]\}$ the solution to the (forward) SDE (3.2.2)

$$\begin{aligned} &\mathrm{d}X_s = b(s, X_s)\mathrm{d}s + \sigma(s, X_s)\mathrm{d}W_s, \\ &X_t = x. \end{aligned} \tag{3.2.2}$$

and consider the associated BSDE

$$\begin{aligned} &-\mathrm{d}Y_s = f(s, X^{t,x}_s, Y_s, Z_s)\mathrm{d}s - Z^*_s\mathrm{d}W_s, \\ &Y_T = \Psi(X^{t,x}_T). \end{aligned} \tag{3.2.3}$$

The solution to (3.2.3) is denoted by $\{(Y^{t,x}_s, Z^{t,x}_s); s \in [0, T]\}$ and the system (3.2.2) and (3.2.3) is said to be a (decoupled) Forward-Backward SDE (FBSDE), whose solution will be denoted by $\{(X^{t,x}_s, Y^{t,x}_s, Z^{t,x}_s); s \in [0, T]\}$. This equation is said to be "decoupled" since the coefficients b and σ of (3.2.2) do not depend on Y and Z. In some sense, the system consisting of (3.2.2) and (3.2.3) is made of two independent parts and can be solved in two stages: heuristically, this means that we can go forward in time to solve (3.2.2) first and then backward to find a solution to (3.2.3). The analysis of the "fully coupled" case from both an analytical and numerical point of view is much more complex due to the interdependency of backward and forward component. Lipschitz and growth conditions are required on the coefficient to ensure existence and uniqueness of solutions (both for the forward and backward part); namely there exists a constant $C > 0$ such that

$$|\sigma(t,x) - \sigma(t,y)| + |b(t,x) - b(t,y)| \leq C|x-y|, \tag{3.2.4}$$

$$|f(t,x,y_1,z_1) - f(t,x,y_2,z_2)| \leq C[|y_1-y_2| + |z_1-z_2|]. \tag{3.2.5}$$

Moreover we require that there exists a constant C such that for all (t,x,y,z):

$$|\sigma(t,x)| + |b(t,x)| \leq C(1+|x|), \tag{3.2.6}$$

$$|f(t,x,y,z)| + |\Psi(x)| \leq C(1+|x|). \tag{3.2.7}$$

Note that (3.2.4) and (3.2.6) guarantee that (3.2.2) admits a unique, adapted solution $\{X_t; 0 \leq t \leq T\}$ satisfying

$$\mathbb{E}\left[\sup_{0\leq t\leq T} |X_t|^2\right] < \infty,$$

see, e.g. [71]. Using the latter, (3.2.7) ensures that that $\xi \in L^2_T$, $f(\cdot,0,0) \in \mathcal{H}^2$ which, with the Lipschitz condition (3.2.5), guarantee existence and uniqueness of the solution of the BSDE. The following result investigates measurability properties of the solutions; for its proof, the interested reader is referred to [51], Proposition 4.2.

Theorem 3.2.4. *The solution $\{(Y_s^{t,x}, Z_s^{t,x}); s \in [t,T]\}$ of (3.2.3) is adapted to the future σ-algebra of W after t $\mathcal{F}_s^t$, where, for each $s \in [t,T]$, $\mathcal{F}_s^t = \sigma\{W_u - W_t; t \leq u \leq s\}$. In particular $Y_t^{t,x}$ is deterministic.*

An application of the last result allows us to conclude

$$\begin{aligned} Y_s^{t,x} &= \mathbb{E}[Y_s^{t,x}|\mathcal{F}_s^t] = \mathbb{E}\left[\Psi(X_T^{t,x}) + \int_s^T f(r, X_r^{t,x}, Y_r^{t,x}, Z_r^{t,x})\,\mathrm{d}r \,\Big|\mathcal{F}_s^t\right], \\ Y_t^{t,x} &= \mathbb{E}\left[\Psi(X_T^{t,x}) + \int_t^T f(r, X_r^{t,x}, Y_r^{t,x}, Z_r^{t,x})\,\mathrm{d}r\right]. \end{aligned} \tag{3.2.8}$$

Theorem 3.2.5. *There exist two measurable deterministic functions $u(t,x)$ and $v(t,x)$ such that the solution $(Y_s^{t,x}, Z_s^{t,x})$ of BSDE (3.2.3) is*

$$Y_s^{t,x} = u(s, X_s^{t,x}), \qquad Z_s^{t,x} = \sigma^*(s, X_s^{t,x})v(s, X_s^{t,x}), \qquad t \leq s \leq T \ \mathbb{P} \otimes \mathrm{d}s\text{-}a.s. \tag{3.2.9}$$

Furthermore, under additional smoothness assumptions on the data and coefficients of the FBSDE we have that

$$Z_s^{t,x} = \sigma^*(s, X_s^{t,x})\nabla u(s, X_s^{t,x}).$$

For additional details about the measurability of u and d and the smoothness requirements we refer to [51] Theorem 4.1 and Corollary 4.1. To get some intuition about the last result, consider the case where the generator f does not depend on y, z. Then the representation (3.2.9) follows from the Markov property of the forward diffusion $X^{t,x}$, namely:

$$Y_s^{t,x} = \mathbb{E}\left[\Psi(X_T^{t,x}) + \int_s^T f(r, X_r^{t,x})\,\mathrm{d}r \,\middle|\mathcal{F}_s^t\right] = \Phi(s, X_s^{t,x}),$$

where

$$\Phi(s,y) = \mathbb{E}\left[\Psi(X_T^{s,y}) + \int_s^T f(u, X_u^{t,y})\,\mathrm{d}u\right].$$

This result can be deduced from [30] but it should be intuitively clear. Observe that, by taking (3.2.8) and substituting the expressions for Y and Z given by Theorem 3.2.5, we obtain

$$Y_t^{t,x} = \mathbb{E}\left[\Psi(X_T^{t,x}) + \int_t^T f(r, X_r^{t,x}, u(r, X_r^{t,x}), \sigma(r, X_r^{t,x})^*\nabla u(r, X_r^{t,x}))\,\mathrm{d}r\right],$$

and thus we deduce the existence of a functional $\Lambda_t : \mathcal{C}([t,T];\mathbb{R}^d) \to \mathbb{R}$ such that

$$Y_t^{t,x} = \mathbb{E}[\Lambda_t(X_\cdot^{t,x})],$$

where $\mathcal{C}([t,T];\mathbb{R}^d)$ is the space of continuous functions from $[t,T]$ into $\mathbb{R}^d$ and $X_\cdot^{t,x}$ denotes a path-valued random variable. We thus obtain the rather surprising fact that the solution to a BSDE can be written entirely as an expectation of the future distribution of the forward component.

The next results were found by Pardoux and Peng [103] and provide a connection between BSDEs and nonlinear PDEs.

Theorem 3.2.6 (Generalization of the Feynman-Kac Formula)**.** *Let v be a function of class $\mathcal{C}^{1,2}$ or smooth enough to apply Itô's formula to $v(s, X_s^{t,x})$ and suppose that there exists a constant C such that, for each (s, x),*

$$|v(s,x)| + |\sigma^*(s,x)\nabla v(s,x)| \leq C(1 + |x|).$$

Suppose also that v is the solution to the following quasilinear parabolic partial differential equation:

$$\begin{aligned} &\partial_t v(t,x) + \mathcal{L}v(t,x) + f(t,x,v(t,x),\sigma^*(t,x)\nabla v(t,x)) = 0 \\ &v(T,x) = \Psi(x), \end{aligned} \tag{3.2.10}$$

where $\mathcal{L}$ denotes the second-order differential operator

$$\mathcal{L}v = \sum_i b_i(t,x)\partial_i v + \frac{1}{2}\sum_{i,j}(\sigma\sigma^*)_{i,j}(t,x)\partial^2_{i,j}v.$$

Then $(Y_s^{t,x}, Z_s^{t,x}) = (v(s, X_s^{t,x}), \sigma^*(s, X_s^{t,x})\nabla v(s, X_s^{t,x}))$, $t \leq s \leq T$, is the unique solution of the BSDE (3.2.3).

This result follows from a straightforward application of Itô's formula to the process $v(s, X_s^{t,x})$. Note that the last theorem allows us to construct a solution of a BSDE by knowing the solution to the semilinear PDE. While under the assumptions we made at the beginning of the section one cannot expect that equation (3.2.10) will have a solution in the classical sense, it can be shown that the converse of the last theorem also holds in the viscosity sense, as was found by [103].

Theorem 3.2.7. *We suppose that $d = 1$ and that f and Ψ are uniformly continuous with respect to x. Then the function u defined by $u(t,x) = Y_t^{t,x}$ is a viscosity solution of the PDE (3.2.10).*

For background material about viscosity solutions we refer to the book by Fleming and Soner [57] or to the original paper by Lions and Crandall [37] where it was applied to the study of the Hamilton-Jacobi equation.

3.3 Weak approximation of Forward-Backward SDEs with jumps

BSDEs with respect to Brownian motion and Lévy process

In the previous discussion, we dealt with (F)BSDEs where the filtration was generated by a Brownian motion. In this section, we generalise some of the previous results (in particular existence and uniqueness) to the case of BSDEs whose solution is measurable with respect to a filtration generated by a forward SDE driven by Brownian motion and a pure jump Lévy martingale. Let $(\Omega, \mathcal{F}, (\mathcal{F}_t), \mathbb{P})$ be a stochastic basis such that $\mathcal{F}_0$ contains all the $\mathbb{P}$-null elements of $\mathcal{F}$ and the filtration $(\mathcal{F}_t)$ is right-continuous and generated by the following two mutually independent processes:

- an d_1-dimensional standard Brownian motion W,
- a d_2 dimensional, zero-mean, square-integrable Lévy process without Gaussian component X. By the Lévy -Itô decomposition the following is given by

$$X_t = \int_{[0,t]\times\mathbb{R}^{d_2}\setminus\{0\}} x(N(\mathrm{d}s, \mathrm{d}x) - \nu(\mathrm{d}x)\mathrm{d}s) = \int_{[0,t]\times\mathbb{R}^{d_2}\setminus\{0\}} x\tilde{N}(\mathrm{d}s, \mathrm{d}x),$$

  where N is the Poisson random measure associated to the Poisson point process of jumps of X, $\tilde{N}(\mathrm{d}s, \mathrm{d}x)$ the corresponding compensated Poisson random measure and ν the Lévy measure of X.

Incidentally, we remind the reader that for any $A \in \mathbb{R}^{d_2} \setminus \{0\}$ with $\nu(A) < \infty$, the compensated poisson random measure $\tilde{N}([0,t] \times A) := (N - \nu \otimes \mathrm{d}s)([0,t] \times A)$ is a martingale.

Let us now introduce some notations.

- $\mathcal{S}^2$ be the set of $\mathcal{F}_t$-adapted, càdlàg processes $\{Y_t; 0 \leq t \leq T\}$ with values in $\mathbb{R}$

such that

$$\|Y\|_{\mathcal{S}^2}^2 = \mathbb{E}\left[\sup_{t\in[0,T]} |Y_t|^2\right] < \infty.$$

- $\mathcal{H}^2$, the space of predictable processes $\phi : \Omega \times [0,T] \to \mathbb{R}$ such that

$$\|\phi\|_{\mathcal{H}^2}^2 = \mathbb{E}\left[\int_0^T |\phi_t|^2 \,\mathrm{d}t\right] < \infty.$$

- $L^2(\mathbb{P} \otimes \nu \otimes \mathrm{d}s)$, the set of mappings $U : \Omega \times [0,T] \times (\mathbb{R}^{d_2} \setminus \{0\}) \to \mathbb{R}$ which are $\mathcal{P} \otimes \mathcal{B}(\mathbb{R}^{d_2} \setminus \{0\})$-measurable and such that

$$\|U\|_{L^2}^2 = \mathbb{E}\left[\int_0^T \int_{\mathbb{R}^{d_2}\setminus\{0\}} U(t,e)^2 \,\nu(\mathrm{d}e)\,\mathrm{d}t\right] < \infty.$$

A proof of the next result can be found in Tang and Li [121], Lemma 2.4.

Theorem 3.3.1. *Let $F \in L^2$ and $f : \Omega \times \mathbb{R} \times \mathbb{R}^{d_1} \times L^2(\nu) \to \mathbb{R}$ be $\mathcal{P} \otimes \mathcal{B}(\mathbb{R}) \otimes \mathcal{B}(\mathbb{R}^{d_1}) \otimes \mathcal{B}(L^2(\nu))$-measurable and $f(t,0,0,0) \in \mathcal{H}^2$. Furthermore, suppose the generator is uniformly Lipschitz in the state variables, i.e., there exists $K > 0$ such that, for all, $y, y' \in \mathbb{R}$, $z, z' \in \mathbb{R}^{d_1}$, $\tilde{z}, \tilde{z}' \in L^2(\nu)$*

$$|f(t,y,z,\tilde{z}) - f(t,y',z',\tilde{z}')| \le K\left(|y-y'| + |z-z'| + \sqrt{\int_{\mathbb{R}^{d_2}\setminus\{0\}} |\tilde{z}(x) - \tilde{z}'(x)|^2 \,\nu(\mathrm{d}x)}\right).$$

Then, there exists a unique triple $(Y, Z, \tilde{Z}) \in \mathcal{S}^2 \times \mathcal{H}^2 \times L^2(\mathbb{P} \otimes \nu \otimes \mathrm{d}s)$ which solves

$$Y_t = F + \int_t^T f(s, Y_s, Z_s, \tilde{Z}_s)\,\mathrm{d}s - \int_t^T Z_s \,\mathrm{d}W_s - \int_t^T \int_{\mathbb{R}^{d_2}\setminus\{0\}} \tilde{Z}_s(x)\,\tilde{N}(\mathrm{d}s, \mathrm{d}x), \quad 0 \le t \le T.$$

Such a triple is then called the solution of the BSDE with data (F, f).

Incidentally, we remark that the solution of an equation of such type provides a viscosity solution of a parabolic integral-partial differential equation. Further information can be found in [8]. Important work has been done by Royer [117] to introduce the notions of g-expectations and of non-linear expectations in the jump set-up.

The goal of this section is to study the forward-backward stochastic differential equation, for $0 \le t \le T$:

$$\begin{aligned} \xi_t &= x + \int_0^t \mu(\xi_{s-})\,\mathrm{d}s + \int_0^t \sigma(\xi_{s-})\,\mathrm{d}W_s + \int_0^t \int_{\mathbb{R}} \zeta(\xi_{s-}, e)\,\tilde{N}(\mathrm{d}s, \mathrm{d}e) \\ Y_t &= F + \int_t^T f(s, \xi_{s-}, Z_s, \tilde{Z}_s(\cdot))\,\mathrm{d}s - \int_t^T Z_s\,\mathrm{d}W_s - \int_t^T \int_{\mathbb{R}} \tilde{Z}_s(e)\,\tilde{N}(\mathrm{d}s, \mathrm{d}e), \end{aligned} \tag{3.3.1}$$

where $\mu : \mathbb{R} \to \mathbb{R}$, $\sigma : \mathbb{R}^{d_1} \to \mathbb{R}$, $\zeta : \mathbb{R} \times E \to \mathbb{R}$ and E denotes the the set $\mathbb{R}^{d_2} \setminus \{0\}$.

Before starting the discussion we briefly recall the condition that guarantee existence and uniqueness of the forward component. Assume that μ, σ are globally Lipschitz and let ζ be measurable and such that there exists a real constant C such that, for all $e \in E$,

$$\begin{aligned} &|\zeta(x, e)| \le C(1 \wedge |e|), \quad x \in \mathbb{R}, \\ &|\zeta(x, e) - \zeta(x', e')| \le C|x - x'|(1 \wedge |e|), \quad x, x' \in \mathbb{R}. \end{aligned}$$

Consider the following SDE:

$$\begin{aligned} &\mathrm{d}\xi_s = \mu(\xi_s)\mathrm{d}s + \sigma(\xi_s)\mathrm{d}W_s + \int_E \zeta(\xi_{s-}, e)\,\tilde{N}(\mathrm{d}s, \mathrm{d}e) \\ &X_t = x, \end{aligned} \tag{3.3.2}$$

or, equivalently, in integral form

$$\xi_t = x + \int_0^t \mu(\xi_s)\,\mathrm{d}s + \int_0^t \sigma(\xi_s)\,\mathrm{d}W_s + \int_0^t \int_E \zeta(\xi_{s-}, e)\,\tilde{N}(\mathrm{d}s, \mathrm{d}e).$$

We denote by $\{\xi_s^{t,x}; s \ge t\}$ the unique solution to (3.3.2). Some of its properties and existence and uniqueness are proved, for instance, by Fujiwara and Kunita [58].

The above SDE can model dynamics involving infinitely many small jumps per unit of time. The jump behaviour is controllable via the jump coefficient ζ that depends on the mark e. We remind that

$$\begin{aligned} \Delta I(t) &= \int_0^t \int_E \zeta(\xi_{s-}, e)\,N(\mathrm{d}s, \mathrm{d}e) - \int_0^{t-} \int_E \zeta(\xi_{s-}, e)\,N(\mathrm{d}s, \mathrm{d}e) \\ &= \int_{E \times \{t\}} \zeta(\xi_{t-}, e) N(\mathrm{d}s, \mathrm{d}e). \end{aligned}$$

This means that if at a jump time τ, the Poisson random measure generates an event with mark e, then the change in value of the corresponding integral will be $\beta(\tau-, e)$. We deal with time homogeneous dynamics but time dependence is easy to introduce.

BSΔEs, formulation and solution

As approximation to the FBSDE (3.3.1) we consider FBSDE in discrete time (also called FBSΔE) driven by processes with independent increments $(W^{(\pi)}, X^{(\pi)})$ that are constant outside uniform time grids $\pi = \{\pi_N\}$, where $\pi_N = \{t_0, t_1, \ldots, t_N\}$ where $t_i = iT/N$, $i = 0, \ldots, N$. Note that the approximating processes can be defined on filtrations that are different from the one associated to the BSDE. When no confusion arises we shall write $\pi = \pi_N$ and the denote the mesh of the grid by $\Delta = \Delta_N$ by T/N. We shall also identify the process $(W^{(\pi)}, X^{(\pi)})$ with the discrete time process $(W^{(\pi)}_{t_i}, X^{(\pi)}_{t_i})_{t_i \in \pi}$.

Assume that $W^{(\pi)}$ and $X^{(\pi)}$ are independent square-integrable martingales. In particular, let $W^{(\pi)}$ be a zero-mean random walk whose increments $\Delta W^{(\pi)}_{t_i} = W^{(\pi)}_{t_{i+1}} - W^{(\pi)}_{t_i}$ match the corresponding second moment of a Wiener process

$$\mathbb{E}_{t_i}\left[\left(\Delta W^{(\pi)}_{t_i}\right)^* \left(\Delta W^{(\pi)}_{t_i}\right)\right] = \Delta \cdot I_{d_1}, \quad i = 0, 1, \ldots, N-1, \tag{3.3.3}$$

where I_{d_1} denotes the $d_1 \times d_1$ identity matrix and $\mathbb{E}_{t_i}[\cdot]$ denotes the conditional expectation taken with respect to the standard filtration generated by $(W^{(\pi)}, X^{(\pi)})$, $\mathbf{F} := \{\mathcal{F}^{(\pi)}_{t_i}; t_i \in \pi\}$. As an example, $W^{(\pi)}$ could be chosen to be a random walk with suitably chosen Bernoulli, trinomial or Gaussian distributions, as in Briand *et al.* [24].

Assume that $X^{(\pi)}$ is a zero-mean random walk with distribution of the increments given by $G^{(\pi)}$, satisfying the moment conditions

$$\Delta^{-1/2}\mathbb{E}[|\Delta X^{(\pi)}_{t_i}|] \to 0, \qquad \Delta \to 0, \tag{3.3.4}$$

$$\mathbb{E}_{t_i}\left[\left(\Delta X^{(\pi)}_{t_i}\right)^* \left(\Delta X^{(\pi)}_{t_i}\right)\right] = \Delta \cdot (\nu_{k,l})_{k,l=1,\ldots,d_2}, \quad i = 0, \ldots, N-1, \tag{3.3.5}$$

where $*$ denotes the transpose of a matrix and

$$\nu_{k,l} = \int h_k(x) h_l(x)\, \nu(\mathrm{d}x), \quad h_k(x) = x_k, \quad k = 1, \ldots, d_2.$$

Furthermore, define measures

$$\nu^{(\pi)}(\mathrm{d}x) = \Delta^{-1} G^{(\pi)}(\mathrm{d}x),$$

and assume that

$$\nu^{(\pi)} \xrightarrow{v} \nu, \tag{3.3.6}$$

as the mesh goes to zero, where $\xrightarrow{v}$ denotes vague convergence of measures, namely $\nu^{(\pi)}(C) \to \nu(C)$ for all closed compact sets $C \subset \mathbb{R}^{d_2} \setminus \{0\}$. Observe that Equation (3.3.4) is satisfied when we take $\Delta X^{(\pi)}_{t_i}$ equal to the increment of X: since X is square integrable, we have that the first absolute moment satisfies $\mathbb{E}[|\Delta X_t|] = \mathcal{O}(t)$ for small t (see, e.g., Luschgy & Pagès [84]). Under the conditions in Equations (3.3.3), (3.3.5) and (3.3.6) there is functional weak convergence of $(W^{(\pi)}, X^{(\pi)})$ to the Lévy process (W, X) in the Skorokhod J_1- topology. This is a consequence of classical weak convergence theory (Jacod & Shiryaev [69] Theorem VII 3.7) given the conditions in Equations (3.3.3), (3.3.4) and (3.3.6).

As an approximation to the forward process X in Equation (3.3.1), we consider the stochastic difference equation given by $\xi^{(\pi)}_0 = x$ and, for $t_i \in \pi$:

$$\xi^{(\pi)}_{t_{i+1}} = \xi^{(\pi)}_{t_i} + \mu^{(\pi)}(\xi^{(\pi)}_{t_i})\Delta + \sigma^{(\pi)}(\xi^{(\pi)}_{t_i})\Delta W^{(\pi)}_{t_i} + \zeta^{(\pi)}(\xi^{(\pi)}_{t_i}, \Delta X^{(\pi)}_{t_i}) - \mathbb{E}_{t_i}[\zeta^{(\pi)}(\xi^{(\pi)}_{t_i}, \Delta X^{(\pi)}_{t_i})],$$

where $\mu^{(\pi)}, \sigma^{(\pi)}, \zeta^{(\pi)}$ are piecewise constant approximations of μ, σ, ζ respectively.

Observe that, in a discrete setting such as ours, the natural filtration generated by the process $\xi^{(\pi)}$, $\{\mathcal{F}^{\xi^{(\pi)}}_{t_i}\}$, and $\{\mathcal{F}^{(\pi)}_{t_i}\}$ do not coincide. This is intuitively clear in view of the fact that, in discrete time, it becomes more difficult to identify the contribution given by the jumps from the increments of $\xi^{(\pi)}$. We choose then to set the forward-backward difference equation on a probability space endowed with the filtration $\{\mathcal{F}^{(\pi)}_{t_i}\}$ rather than with the natural filtration generated by the forward process; conditional expectations will therefore be taken with respect to $\{\mathcal{F}^{(\pi)}_{t_i}\}$ in what follows. Observe that we are entitled to work on the latter since in the limit the two filtrations will coincide (under reasonable assumptions on the functions σ and ζ, e.g., invertibility).

As noted by [24], when switching from the Wiener process W to the process $W^{(\pi)}$ we lose the predictable representation property. Indeed, in the discrete case, a real martingale with independent increments has the predictable representation property [3] if an only if the increments are supported by two points. If follows that, in the formulation of the FBSDE driven by $W^{(\pi)}$ and $X^{(\pi)}$, we have to add an orthogonal martingale term $\{M^{(\pi)}\}$ so that the equation under consideration becomes:

$$
\begin{aligned}
\xi_{t_{i+1}}^{(\pi)} &= \xi_{t_i}^{(\pi)} + \mu^{(\pi)}(\xi_{t_i}^{(\pi)})\Delta + \sigma^{(\pi)}(\xi_{t_i}^{(\pi)})\Delta W_{t_i}^{(\pi)} + \zeta^{(\pi)}(\xi_{t_i}^{(\pi)}, \Delta X_{t_i}^{(\pi)}) - \mathbb{E}_{t_i}[\zeta^{(\pi)}(\xi_{t_i}^{(\pi)}, \Delta X_{t_i}^{(\pi)})],\\
Y_{t_i}^{(\pi)} &= F^{(\pi)} + \sum_{j=i}^{N-1} f^{(\pi)}(t_j, \xi_{t_i}^{(\pi)}, Y_{t_j}^{(\pi)}, Z_{t_j}^{(\pi)}, \tilde{Z}_{t_j}^{(\pi)})\Delta - \sum_{j=i}^{N-1} Z_{t_j}^{(\pi)}\Delta W_{t_j}^{(\pi)}\\
&\quad - \sum_{j=i}^{N-1}\left\{\tilde{Z}_{t_j}^{(\pi)}(\Delta X_{t_j}^{(\pi)})\mathbb{1}_{\{\Delta X_{t_j}^{(\pi)}\neq 0\}} - \mathbb{E}_{t_j}\left[\tilde{Z}_{t_j}^{(\pi)}(\Delta X_{t_j}^{(\pi)})\mathbb{1}_{\{\Delta X_{t_j}^{(\pi)}\neq 0\}}\right]\right\} + M_T^{(\pi)} - M_{t_i}^{(\pi)},
\end{aligned}
\tag{3.3.7}
$$

where $f^{(\pi)}$ is a $\nu^{(\pi)}$-regular driver that is piecewise constant, i.e., $f^{(\pi)}(s,\cdot) = f(t_i,\cdot)$ for $s \in [t_i, t_{i+1})$. The random variable $F^{(\pi)} \in L^2(\mathcal{F}_T)$ is the terminal condition and $Y_{t_i}^{(\pi)}, Z_{t_i}^{(\pi)} \in L^2(\mathcal{F}_{t_i}^{(\pi)})$, $\tilde{Z}_{t_i}^{(\pi)} \in L^2(G^{(\pi)}(\mathrm{d}x) \otimes \mathrm{d}\mathbb{P}, \mathcal{B}(\mathbb{R}^{d_2} \setminus \{0\}) \otimes \mathcal{F}_{t_i})$ and $M^{(\pi)} = \{M_{t_i}^{(\pi)}\}$ is a $\mathbf{F}$-martingale on π that is orthogonal to $\{W_{t_i}^{(\pi)}\}$ and to the martingales with increments

$$\tilde{U}_{t_i}^{(\pi)}(\Delta X_{t_i}^{(\pi)}) - \mathbb{E}_{t_i}[\tilde{U}_{t_i}^{(\pi)}(\Delta X_{t_i}^{(\pi)})],$$

for any function $\tilde{U}_{t_i}^{(\pi)}$ with $\tilde{U}_{t_i}^{(\pi)} \in L^\infty(\mathcal{F}_{t_i}^{(\pi)} \otimes \mathcal{B}(\mathbb{R}^{d_2} \setminus \{0\}))$.

Equation (3.3.7) can be written in differential notation as

$$
\begin{aligned}
\Delta\xi_{t_i}^{(\pi)} &= \mu^{(\pi)}(\xi_{t_i}^{(\pi)})\Delta + \sigma^{(\pi)}(\xi_{t_i}^{(\pi)})\Delta W_{t_i}^{(\pi)} + \zeta^{(\pi)}(\xi_{t_i}^{(\pi)}, \Delta X_{t_i}^{(\pi)}) - \mathbb{E}_{t_i}[\zeta^{(\pi)}(\xi_{t_i}^{(\pi)}, \Delta X_{t_i}^{(\pi)})],\\
\Delta Y_{t_i}^{(\pi)} &= -f^{(\pi)}(t_i, \xi_{t_i}^{(\pi)}, Y_{t_i}^{(\pi)}, Z_{t_i}^{(\pi)}, \tilde{Z}_{t_i}^{(\pi)})\Delta + Z_{t_i}^{(\pi)}\Delta W_{t_i}^{(\pi)}\\
&\qquad + \left\{\tilde{Z}_{t_i}^{(\pi)}(\Delta X_{t_i}^{(\pi)})\mathbb{1}_{\{\Delta X_{t_i}^{(\pi)}\neq 0\}} - \mathbb{E}_{t_i}\left[\tilde{Z}_{t_i}^{(\pi)}(\Delta X_{t_i}^{(\pi)})\mathbb{1}_{\{\Delta X_{t_i}^{(\pi)}\neq 0\}}\right]\right\} + \Delta M_{t_i}^{(\pi)},
\end{aligned}
\tag{3.3.8}
$$

$$Y_T^{(\pi)} = F^{(\pi)},$$

for $i = 0, \ldots, N-1$. Note that Equation (3.3.7) is equivalent to a backward recursive

[3] for background material on the predictable representation property see, for instance Revuz & Yor [115]

system beginning by $Y_T^{(\pi)} = F^{(\pi)}$ and followed by the expressions given in the following

Theorem 3.3.2. *Given a FBSΔE (3.3.7) with standard parameters $(f^{(\pi)}, F^{(\pi)})$ and Lipschitz constant of $f^{(\pi)}$ given by K. Then, for $\Delta < 1/K$ the BSΔE (3.3.7) has a unique solution $(Y^{(\pi)}, Z^{(\pi)}, \tilde{Z}^{(\pi)}, M^{(\pi)})$ that satisfies the relations:*

$$Y_{t_i}^{(\pi)} = \mathbb{E}_{t_i}\left[F^{(\pi)} + \sum_{j=i}^{N-1} f^{(\pi)}(t_j, \xi_{t_j}^{(\pi)}, Y_{t_j}^{(\pi)}, Z_{t_j}^{(\pi)}, \tilde{Z}_{t_j}^{(\pi)})\Delta\right] \tag{3.3.9}$$

$$= f^{(\pi)}(t_i, \xi_{t_i}^{(\pi)}, Y_{t_i}^{(\pi)}, Z_{t_i}^{(\pi)}, \tilde{Z}_{t_i}^{(\pi)})\Delta + \mathbb{E}_{t_i}[Y_{t_{i+1}}^{(\pi)}], \tag{3.3.10}$$

$$Z_{t_i}^{(\pi)} = \Delta^{-1}\,\mathbb{E}_{t_i}[Y_{t_{i+1}}^{(\pi)}\Delta W_{t_i}^{(\pi)}], \tag{3.3.11}$$

$$\tilde{Z}_{t_i}^{(\pi)}(x) = \mathbb{E}_{t_i}[Y_{t_{i+1}}^{(\pi)}|\Delta X_{t_i}^{(\pi)} = x] - \mathbb{E}_{t_i}[Y_{t_{i+1}}^{(\pi)}|\Delta X_{t_i}^{(\pi)} = 0], \quad x \in \mathbb{R}^{d_2}, \tag{3.3.12}$$

$$\begin{aligned}\Delta M_{t_i}^{(\pi)} &= Y_{t_{i+1}}^{(\pi)} - \mathbb{E}_{t_i}[Y_{t_{i+1}}^{(\pi)}] - Z_{t_i}^{(\pi)}\Delta W_{t_i}^{(\pi)} \\ &\quad - \{\tilde{Z}_{t_i}^{(\pi)}(\Delta X_{t_i}^{(\pi)})\mathbb{1}_{\{\Delta X_{t_i}^{(\pi)}\neq 0\}} - \mathbb{E}_{t_i}[\tilde{Z}_{t_i}^{(\pi)}(\Delta X_{t_i}^{(\pi)})\mathbb{1}_{\{\Delta X_{t_i}^{(\pi)}\neq 0\}}]\}.\end{aligned} \tag{3.3.13}$$

Proof. The proof is inspired by that of [90, Proposition 1]. First note that $Y_{t_i}^{(\pi)}$ is determined uniquely by the implicit equation (3.3.10): indeed, the map $\Psi : L^2(\mathcal{F}_{t_i}^{(\pi)}) \to L^2(\mathcal{F}_{t_i}^{(\pi)})$ given by $\Psi(Y) = f^{(\pi)}(t_i, Y, Z_{t_i}^{(\pi)}, \tilde{Z}_{t_i}^{(\pi)}) + \mathbb{E}_{t_i}[Y_{t_{i+1}}^{(\pi)}]$ is a contraction in the case $K\Delta < 1$, as a consequence of the Lipschitz continuity of $f^{(\pi)}$. Taking conditional expectations with respect to $\mathcal{F}_{t_i}^{(\pi)}$ on both sides of (3.3.8) yields (3.3.10). Moreover, multiplying both sides with $\Delta W_{t_i}^{(\pi)}$ and taking again conditional expectations with respect to $\mathcal{F}_{t_i}^{(\pi)}$ leads to (3.3.11). Observe that, given (3.3.10), (3.3.13) is equivalent to (3.3.8); in other words, if $\Delta M_{t_i}^{(\pi)}$ and $\tilde{Z}_{t_i}^{(\pi)}$, are uniquely determined, the expression in (3.3.10) yields (3.3.13).

As a consequence of the discrete martingale representation property (see Jacod & Shiryaev [69] Lemma III 4.24), given a martingale with increments given by $Y_{t_{i+1}}^{(\pi)} - \mathbb{E}_{t_i}[Y_{t_{i+1}}^{(\pi)}] - Z_{t_i}^{(\pi)}\Delta W_{t_i}^{(\pi)}$, there exist $\tilde{Z}_{t_i}^{(\pi)}(\cdot) \in L^\infty(\mathcal{F}_{t_i}^{(\pi)} \otimes \mathcal{B}(\mathbb{R}^{d_2} \setminus \{0\}))$ and $\Delta M_{t_i}^{(\pi)}$ such that

$$\begin{aligned}&Y_{t_{i+1}}^{(\pi)} - \mathbb{E}_{t_i}[Y_{t_{i+1}}^{(\pi)}] - Z_{t_i}^{(\pi)}\Delta W_{t_i}^{(\pi)} \\ &\qquad = \tilde{Z}_{t_i}^{(\pi)}(\Delta X_{t_i}^{(\pi)})\mathbb{1}_{\{\Delta X_{t_i}^{(\pi)}\neq 0\}} - \mathbb{E}_{t_i}[\tilde{Z}_{t_i}^{(\pi)}(\Delta X_{t_i}^{(\pi)})\mathbb{1}_{\{\Delta X_{t_i}^{(\pi)}\neq 0\}}] + \Delta M_{t_i}^{(\pi)}.\end{aligned}$$

The expression for $\tilde{Z}^{(\pi)}_{t_i}$ is then obtained by imposing the orthogonality requirement. To complete the proof, we then need to define $\tilde{Z}^{(\pi)}_{t_i}$ by the right hand side of (3.3.12) and with definition show the orthogonality of the right hand side of (3.3.13) to $W^{(\pi)}$ and the martingales induced by $X^{(\pi)}$ (see above). Once we have shown this fact it will follow that (a) $\tilde{Z}^{(\pi)}_{t_i}$ solves the BSΔE (which follows by combining (3.3.13) and (3.3.10)), and (b) that it is its unique solution.

Next we verify the orthogonality of the martingale $M^{(\pi)}$. We note, using (3.3.11) and the tower property of the conditional expectation

$$\mathbb{E}_{t_i}[\Delta M^{(\pi)}_{t_i} \Delta W^{(\pi)}_{t_i}] = \mathbb{E}_{t_i}[Y^{(\pi)}_{t_i} \Delta W^{(\pi)}_{t_i}] - \mathbb{E}_{t_i}[Z^{(\pi)}_{t_i} \Delta W^{(\pi)}_{t_i} \cdot \Delta W^{(\pi)}_{t_i}] = 0.$$

Furthermore, for any function $k \in L^\infty(\mathcal{F}^{(\pi)}_{t_i} \otimes \mathcal{B}(\mathbb{R}^{d_2} \setminus \{0\}))$ with $k(0) = 0$, it holds

$$\begin{aligned}
&\mathbb{E}_{t_i}[\Delta M^{(\pi)}_{t_i} \{k(\Delta X^{(\pi)}_{t_i}) - \mathbb{E}_{t_i}[k(\Delta X^{(\pi)}_{t_i})]\}] = \\
&\mathbb{E}_{t_i}[(Y^{(\pi)}_{t_{i+1}} - \mathbb{E}_{t_i}[Y^{(\pi)}_{t_{i+1}}])\{k(\Delta X^{(\pi)}_{t_i}) - \mathbb{E}_{t_i}[k(\Delta X^{(\pi)}_{t_i})]\}] - \\
&\mathbb{E}_{t_i}[(\tilde{Z}^{(\pi)}_{t_i}(\Delta X^{(\pi)}_{t_i})\mathbb{1}_{\{\Delta X^{(\pi)}_{t_i} \neq 0\}} - \mathbb{E}_{t_i}[\tilde{Z}^{(\pi)}_{t_i}(\Delta X^{(\pi)}_{t_i})\mathbb{1}_{\{\Delta X^{(\pi)}_{t_i} \neq 0\}}])\{k(\Delta X^{(\pi)}_{t_i}) - \mathbb{E}_{t_i}[k(\Delta X^{(\pi)}_{t_i})]\}].
\end{aligned} \tag{3.3.14}$$

Inserting the form of (3.3.12), using the tower property of conditional expectation and denoting by $\phi(x)$ the random variable $\mathbb{E}_{t_i}[Y^{(\pi)}_{t_{i+1}} | \Delta X^{(\pi)}_{t_i} = x]$, we obtain

$$\begin{aligned}
&\mathbb{E}_{t_i}[(\tilde{Z}^{(\pi)}_{t_i}(\Delta X^{(\pi)}_{t_i})\mathbb{1}_{\{\Delta X^{(\pi)}_{t_i} \neq 0\}} - \mathbb{E}_{t_i}[\tilde{Z}^{(\pi)}_{t_i}(\Delta X^{(\pi)}_{t_i})\mathbb{1}_{\{\Delta X^{(\pi)}_{t_i} \neq 0\}}])\{k(\Delta X^{(\pi)}_{t_i}) - \mathbb{E}_{t_i}[k(\Delta X^{(\pi)}_{t_i})]\}] \\
&\quad = \mathbb{E}_{t_i}[\tilde{Z}^{(\pi)}_{t_i}(\Delta X^{(\pi)}_{t_i})\mathbb{1}_{\{\Delta X^{(\pi)}_{t_i} \neq 0\}}\{k(\Delta X^{(\pi)}_{t_i}) - \mathbb{E}_{t_i}[k(\Delta X^{(\pi)}_{t_i})]\}] \\
&\quad = \mathbb{E}_{t_i}[\{\phi(\Delta X^{(\pi)}_{t_i}) - \phi(0)\}\{k(\Delta X^{(\pi)}_{t_i}) - \mathbb{E}_{t_i}[k(\Delta X^{(\pi)}_{t_i})]\}] \\
&\quad = \mathbb{E}_{t_i}[Y^{(\pi)}_{t_{i+1}}\{k(\Delta X^{(\pi)}_{t_i}) - \mathbb{E}_{t_i}[k(\Delta X^{(\pi)}_{t_i})]\}],
\end{aligned}$$

where in the third equality we used the fact that $\tilde{Z}(0) = 0$. It is a matter of collecting terms to verify that the first term on the left-hand side of (3.3.14) is equal to zero, so that $M^{(\pi)}$ is orthogonal to the martingale with increments $k(\Delta X^{(\pi)}_{t_i}) - \mathbb{E}_{t_i}[k(\Delta X^{(\pi)}_{t_i})]$. As k is taken arbitrary, this establishes the orthogonality and completes the proof. □

Donsker-type theorem for BSDEs with jumps

The solution to the FBSΔE given in Theorem 3.3.2 can be obtained by solving the backward recursive system. We choose to follow a "forward approach" like the one presented in Bender & Denk [11]. More precisely, the process $(Y^{(\pi)}, Z^{(\pi)}, \tilde{Z}^{(\pi)})$ satisfying the BSΔE can be obtained as the limit of a recursively defined Picard sequence $(Y^{(\pi,p)}, Z^{(\pi,p)}, \tilde{Z}^{(\pi,p)})$ which is initialised with $(Y^{(\pi,0)}, Z^{(\pi,0)}, \tilde{Z}^{(\pi,0)}) = (0,0,0)$ and is defined for $p \in \mathbb{N}$ and $t_i \in \pi$ by

$$Y_{t_i}^{(\pi,p)} = \mathbb{E}_{t_i}\left[F^{(\pi)} + \sum_{j=i}^{N-1} f^{(\pi)}(t_j, \xi_{t_j}^{(\pi)}, Y_{t_j}^{(\pi,p-1)}, Z_{t_j}^{(\pi,p-1)}, \tilde{Z}_{t_j}^{(\pi,p-1)}(\cdot))\Delta\right], \quad (3.3.15)$$

$$Z_{t_i}^{(\pi,p)} = \Delta^{-1}\mathbb{E}_{t_i}\left[\Delta W_{t_i}^{(\pi)}\left(F^{(\pi)} + \sum_{j=i+1}^{N-1} f^{(\pi)}(t_j, \xi_{t_j}^{(\pi)}, Y_{t_j}^{(\pi,p-1)}, Z_{t_j}^{(\pi,p-1)}, \tilde{Z}_{t_j}^{(\pi,p-1)}(\cdot))\Delta\right)\right],$$

$$\begin{aligned}\tilde{Z}_{t_i}^{(\pi,p)}(x) &= \mathbb{E}_{t_i}\left[F^{(\pi)} + \sum_{j=i+1}^{N-1} f^{(\pi)}(t_j, \xi_{t_j}^{(\pi)}, Y_{t_j}^{(\pi,p-1)}, Z_{t_j}^{(\pi,p-1)}, \tilde{Z}_{t_j}^{(\pi,p-1)}(\cdot))\Delta \,\middle|\, \Delta X_{t_i}^{(\pi)} = x\right] \\ &\quad - \mathbb{E}_{t_i}\left[F^{(\pi)} + \sum_{j=i+1}^{N-1} f^{(\pi)}(t_j, \xi_{t_j}^{(\pi)}, Y_{t_j}^{(\pi,p-1)}, Z_{t_j}^{(\pi,p-1)}, \tilde{Z}_{t_j}^{(\pi,p-1)}(\cdot))\Delta \,\middle|\, \Delta X_{t_i}^{(\pi)} = 0\right],\end{aligned}$$

for $x \in \mathbb{R}^{d_2}$. We may associate to the sequence $(Y^{(\pi,p)}, Z^{(\pi,p)}, \tilde{Z}^{(\pi,p)})_{p\in\mathbb{N}}$ a sequence of square-integrable orthogonal martingales $(M^{(\pi,p)})_{p\in\mathbb{N}}$ solving the corresponding BSΔE which is defined by $M^{(\pi,0)} = 0$ and for $p \in \mathbb{N}$ by

$$\begin{aligned}Y_{t_{i+1}}^{(\pi,p)} = \mathbb{E}_{t_i}[Y_{t_{i+1}}^{(\pi,p)}] + Z_{t_i}^{(\pi,p)}\Delta W_{t_i}^{(\pi)} + \{&\tilde{Z}_{t_i}^{(\pi,p)}(\Delta X_{t_i}^{(\pi)})\mathbb{1}_{\{\Delta X_{t_j}^{(\pi)}\neq 0\}} \\ &- \mathbb{E}_{t_i}[\tilde{Z}_{t_j}^{(\pi,p)}(\Delta X_{t_j}^{(\pi)})\mathbb{1}_{\{\Delta X_{t_j}^{(\pi)}\neq 0\}}]\} - \Delta M^{(\pi,p)}.\end{aligned}$$

A fixed point argument shows that the Picard sequence $(Y^{(\pi,p)}, Z^{(\pi,p)}, \tilde{Z}^{(\pi,p)}(\cdot), \Delta M^{(\pi,p)})$ converges to $(Y^{(\pi)}, Z^{(\pi)}, \tilde{Z}^{(\pi)}(\cdot), \Delta M^{(\pi)})$ as p tends to infinity. This is made more precise by the following

Corollary 3.3.3. *For some $C > 0$ and $n_0 \in \mathbb{N}$ it holds*

$$\sup_{\pi_N : N \geq n_0} \mathbb{E}\Bigg[\sup_{t_i \in \pi_N} |Y^{(\pi)}_{t_i} - Y^{(\pi,p)}_{t_i}|^2 + \sum_{t_i \in \pi_N} \{|Z^{(\pi)}_{t_i} - Z^{(\pi,p)}_{t_i}|^2 \Delta + (\Delta M^{(\pi)}_{t_i} - \Delta M^{(\pi,p)}_{t_i})^2\}$$
$$+ \sum_{t_i \in \pi_N} \{\tilde{Z}^{(\pi)}_{t_i}(\Delta X^{(\pi)}_{t_i}) \mathbb{1}_{\{\Delta X^{(\pi)}_{t_i} \neq 0\}} - \tilde{Z}^{(\pi,p)}_{t_i}(\Delta X^{(\pi)}_{t_i}) \mathbb{1}_{\{\Delta X^{(\pi)}_{t_i} \neq 0\}}$$
$$- \mathbb{E}_{t_i}[\tilde{Z}^{(\pi)}_{t_i}(\Delta X^{(\pi)}_{t_i}) \mathbb{1}_{\{\Delta X^{(\pi)}_{t_i} \neq 0\}} - \tilde{Z}^{(\pi,p)}_{t_i}(\Delta X^{(\pi)}_{t_i}) \mathbb{1}_{\{\Delta X^{(\pi)}_{t_i} \neq 0\}}]\}^2\Bigg] \leq C 4^{-p}.$$

The proof is omitted since it follows from a straightforward adaptation of Corollary 10 in Briand *et al.* [24].

With the results concerning the solution of the discrete time BSDE in hand, we now turn our attention to the question of weak convergence of BSΔEs to the limiting BSDE as the mesh size goes to zero.

Assume that the driver $f^{(\pi)}$ is a smooth functional with bounded derivatives (uniformly in t and π) and that F, $F^{(\pi)}$ are of the form

$$F = H(W_{s_1}, \ldots, W_{s_k}, X_{s_1}, \ldots, X_{s_k}),$$
$$F^{(\pi)} = H(W^{(\pi)}_{s_1}, \ldots, W^{(\pi)}_{s_k}, X^{(\pi)}_{s_1}, \ldots, X^{(\pi)}_{s_k}),$$

for some smooth function H. Let (π) be a sequence of partition whose mesh Δ tends to zero. If F converges to $F^{(\pi)}$ in L^2, then

$$Y^{(\pi)} \xrightarrow{\mathcal{L}} Y \text{ and in particular } Y^{(\pi)}_0 \to Y_0,$$

where $\xrightarrow{\mathcal{L}}$ denotes weak convergence in the Skorokhod J_1 topology.

We briefly sketch the main ideas of the proof of convergence and will present the full proof elsewhere.

Define the sequence of processes $(Y^{(\infty,p)}, Z^{(\infty,p)}, \tilde{Z}^{(\infty,p)}(\cdot))_{p \in \mathbb{N}}$ by setting

$$(Y^{(\infty,0)}, Z^{(\infty,0)}, \tilde{Z}^{(\infty,0)}(\cdot)) = (0, 0, 0),$$

and

$$Y_t^{(\infty,p+1)} = F + \int_t^T f(s, \xi_s, Y_s^{(\infty,p)}, Z_s^{(\infty,p)}, \tilde{Z}_s^{(\infty,p)})\,\mathrm{d}s \\ - \int_t^T Z_s^{(\infty,p+1)}\,\mathrm{d}W_s - \int_t^T \int_{\mathbb{R}^{d_2}\setminus\{0\}} \tilde{Z}_s^{(\infty,p+1)}(x)\,\tilde{N}(\mathrm{d}s, \mathrm{d}x).$$

The key point of the proof of convergence is to use the following decomposition

$$Y^{(\pi)} - Y = Y^{(\pi)} - Y^{(\pi,p)} + Y^{(\pi,p)} - Y^{(\infty,p)} + Y^{(\infty,p)} - Y. \tag{3.3.16}$$

Note that the first difference has been shown to go to zero in $\mathcal{S}^2$ norm in Corollary 3.3.3, while the third has been shown to go to zero in norm by Tang & Li [121]. To prove the convergence of $Y^{(\pi,p)}$ to $Y^{(\infty,p)}$ in the Skorokhod J_1-topology in probability for any p, the idea is to deploy results on the weak convergence of filtrations.

3.4 A numerical forward scheme with jumps

For numerical implementation of the iteration proposed in the previous section in Equation (3.3.15) one has to approximate conditional expectations. We adopt the point of view of conditional expectation as a projection operator in L^2. This is not the only possible choice but has the advantage to provide robust theoretical error estimates (see Appendix).

Orthogonal projection on subspaces

We will first replace the conditional expectation by orthogonal projections on (finite dimensional) subspaces of $L^2(\mathcal{F}_{t_i}^{(\pi)})$. In particular, choose d_1+1 subspaces $\Lambda_{d,i}$, $0 \le d \le d_1$ of $L^2(\mathcal{F}_{t_i}^{(\pi)})$ and d_2 subspaces $\Gamma_{d,i}$ of $L^2(\mathcal{F}_{t_i}^{(\pi)} \vee \sigma(\Delta X_{t_i}^{(\pi)}))$, $0 \le d < d_2$, where $\mathcal{F}_{t_i}^{(\pi)}$ is the filtration generated by $\{(W_{t_k}^{(\pi)}, X_{t_k}^{(\pi)})\}_{k=0,\dots,i}$ and $\Delta X_{t_i}^{(\pi)} = X_{t_{i+1}}^{(\pi)} - X_{t_i}^{(\pi)}$. Denote the projection operators by

$$P_{d,i} : L^2(\mathcal{F}_{t_i}) \to \Lambda_{d,i}, \qquad Q_{d,i} : L^2(\mathcal{F}_{t_i} \vee \sigma(\Delta X_{t_i})) \to \Gamma_{d,i}.$$

In the following, to avoid an overload in notation, we will focus on the case $d_1 = d_2 = 1$. Consider the approximations $(\mathcal{Y}_{t_i}^{(\pi,p)}, \mathcal{Z}_{t_i}^{(\pi,p)}, \tilde{\mathcal{Z}}_{t_i}^{(\pi,p)}(\cdot))_\pi$ obtained by replacing the conditional expectation in (3.3.15) by the projection operators $P_{0,i}, P_{1,i}, Q_{0,i}$.

Assume now that the projection spaces from the previous section are all finite dimensional and denote by

$$\{e_1^i, \dots, e_{K(i)}^i\}, \quad \{f_1^i, \dots, f_{H(i)}^i\}, \quad \{g_1^i, \dots, g_{J(i)}^i\}$$

a basis of $\Lambda_{0,i}, \Lambda_{1,i}, \Gamma_{0,i}$. Since g_k^i, $k = 0, \dots, J(i)$ is measurable with respect to the filtration $\mathcal{F}_i^{(\pi)} \vee \sigma(\Delta X_{t_i}^{(\pi)})$, we can write it as

$$g_k^i = \bar{g}_k^i(W_\cdot^{(\pi)}|_{\mathcal{F}_{t_i}^{(\pi)}}, X_\cdot^{(\pi)}|_{\mathcal{F}_{t_i}^{(\pi)}}, \Delta X_{t_i}^{(\pi)}),$$

where $\bar{g}_k^i$ is a scalar, deterministic function on $\mathbb{R}^i \times \mathbb{R}^i \times \mathbb{R}$ and $W_\cdot^{(\pi)}|_{\mathcal{F}_{t_i}^{(\pi)}}, X_\cdot^{(\pi)}|_{\mathcal{F}_{t_i}^{(\pi)}}$ are the discrete paths of the process $(W^{(\pi)}, X^{(\pi)})$ up to time t_i. Observe that when writing $\bar{g}_k^i(\cdot)$ we will only consider the dependency with respect to the third variable. Introduce the shorthand notation

$$\mathcal{L}^{(\pi)}(i, p-1) := F^{(\pi)} + \sum_{j=i}^{N-1} f^{(\pi)}(t_j, \xi_{t_j}^{(\pi)}, \mathcal{Y}_{t_j}^{(\pi,p-1)}, \mathcal{Z}_{t_j}^{(\pi,p-1)}, \tilde{\mathcal{Z}}_{t_j}^{(\pi,p-1)}(\cdot))\Delta.$$

Under these hypothesis the processes $\mathcal{Y}^{(\pi,p)}$, $\mathcal{Z}^{(\pi,p)}$ and $\tilde{\mathcal{Z}}^{(\pi,p)}(\cdot)$ may be rewritten as

$$\begin{aligned}
\mathcal{Y}_{t_i}^{(\pi,p)} &= \sum_{k=1}^{K(i)} \alpha_{i,k}^{(\pi,p)} e_k^i, \\
\mathcal{Z}_{t_i}^{(\pi,p)} &= \sum_{k=1}^{H(i)} \beta_{i,k}^{(\pi,p)} f_k^i, \\
\tilde{\mathcal{Z}}_{t_i}^{(\pi,p)}(x) &= \sum_{k=1}^{J(i)} \gamma_{i,k}^{(\pi,p)} (\bar{g}_k^i(x) - \bar{g}_k^i(0)),
\end{aligned}$$

where

$$(\alpha_{i,\cdot}^{(\pi,p)}) = \underset{\alpha\in\mathbb{R}^{K(i)}}{\operatorname{argmin}}\, \mathbb{E}\left[\left(\mathcal{L}^{(\pi)}(i,p-1) - \sum_{k=1}^{K(i)} \alpha_k e_k^i\right)^2\right],$$

$$(\beta_{i,\cdot}^{(\pi,p)}) = \underset{\beta\in\mathbb{R}^{H(i)}}{\operatorname{argmin}}\, \mathbb{E}\left[\left(\mathcal{L}^{(\pi)}(i+1,p-1)\frac{\Delta W_{t_i}^{(\pi)}}{\Delta} - \sum_{k=1}^{K(i)} \beta_k f_k^i\right)^2\right],$$

$$(\gamma_{i,\cdot}^{(\pi,p)}) = \underset{\gamma\in\mathbb{R}^{J(i)}}{\operatorname{argmin}}\, \mathbb{E}\left[\left(\mathcal{L}^{(\pi)}(i+1,p-1) - \sum_{k=1}^{K(i)} \gamma_k g_k^i\right)^2\right].$$

Define

$$_\lambda f^{(\pi)}(t_j) = f^{(\pi)}(t_j, {}_\lambda\xi_{t_j}^{(\pi)}, {}_\lambda\mathcal{Y}_{t_j}^{(\pi,p-1,M)}, {}_\lambda\mathcal{Z}_{t_j}^{(\pi,p-1,M)}, {}_\lambda\tilde{\mathcal{Z}}_{t_j}^{(\pi,p-1,M)}(\cdot)).$$

In practice, the expectations above will be replaced by their empirical estimators

$$(\alpha_{i,\cdot}^{(\pi,p,M)}) = \underset{\alpha\in\mathbb{R}^{K(i)}}{\operatorname{argmin}}\, \frac{1}{M}\sum_{\lambda=1}^{M}\left({}_\lambda F^{(\pi)} + \sum_{j=i}^{N-1} {}_\lambda f^{(\pi)}(t_j)\Delta - \sum_{k=1}^{K(i)} \alpha_k\, {}_\lambda e_k^i\right)^2,$$

$$(\beta_{i,\cdot}^{(\pi,p,M)}) = \underset{\beta\in\mathbb{R}^{H(i)}}{\operatorname{argmin}}\, \frac{1}{M}\sum_{\lambda=1}^{M}\left(\left({}_\lambda F^{(\pi)} + \sum_{j=i+1}^{N-1} {}_\lambda f^{(\pi)}(t_j)\Delta\right)\frac{\Delta_\lambda W_{t_i}^{(\pi)}}{\Delta} - \sum_{k=1}^{H(i)} \beta_k\, {}_\lambda f_k^i\right)^2,$$

$$(\gamma_{i,\cdot}^{(\pi,p,M)}) = \underset{\gamma\in\mathbb{R}^{J(i)}}{\operatorname{argmin}}\, \frac{1}{M}\sum_{\lambda=1}^{M}\left({}_\lambda F^{(\pi)} + \sum_{j=i+1}^{N-1} {}_\lambda f^{(\pi)}(t_j)\Delta - \sum_{k=1}^{K(i)} \gamma_k\, {}_\lambda g_k^i\right)^2.$$

where $(\Delta_\lambda W_i^{(\pi)}, {}_\lambda F^{(\pi)}, {}_\lambda e_k^i, {}_\lambda f_k^i, {}_\lambda g_k^i)$ are $M \geq \max_i\{K(i) \vee H(i) \vee J(i)\}$ independent copies of $(\Delta W_i^{(\pi)}, F^\pi, e_k^i, f_k^i, g_k^i)$ and $({}_\lambda\mathcal{Y}^{(\pi,p,M)}, {}_\lambda\mathcal{Z}^{(\pi,p,M)}, {}_\lambda\tilde{\mathcal{Z}}^{(\pi,p,M)}(\cdot))$ are M independent copies of the simulation based estimators

$$\mathcal{Y}_{t_i}^{(\pi,p,M)} := \sum_{k=1}^{K(i)} \alpha_{i,k}^{(\pi,p,M)} e_k^i,$$

$$\mathcal{Z}_{t_i}^{(\pi,p,M)} := \sum_{k=1}^{H(i)} \beta_{i,k}^{(\pi,p,M)} f_k^i,$$

$$\tilde{\mathcal{Z}}_{t_i}^{(\pi,p,M)}(x) := \sum_{k=1}^{J(i)} \gamma_{i,k}^{(\pi,p,M)} (\bar{g}_k^i(x) - \bar{g}_k^i(0)).$$

The column vectors (in $\mathbb{R}^M$) of these copies will be denoted by

$$(\Delta \mathbf{W_i}^{(\pi)}, \mathbf{F}^{(\pi)}, \mathbf{e}_k^i, \mathbf{f}_k^i, \bar{\mathbf{g}}_k^i(\cdot), \boldsymbol{\mathcal{Y}}^{(\pi,p,M)}, \boldsymbol{\mathcal{Z}}^{(\pi,p,M)}, \tilde{\boldsymbol{\mathcal{Z}}}^{(\pi,p,M)}(\cdot));$$

furthermore, we will define $\boldsymbol{f}^{(\pi)}(t_j) := ({}_\lambda f^{(\pi)}(t_j))$ The above can now be expressed concisely using vector notations: denote by

$$B_{e,i} = \mathbb{E}[e_k^i e_l^i]_{k,l=1,\dots,K(i)}, \qquad A_{e,i}^M = \frac{1}{\sqrt{M}}({}_\lambda e_k^i)_{\lambda=1,\dots,M,k=1,\dots K(i)},$$
$$B_{f,i} = \mathbb{E}[f_k^i f_l^i]_{k,l=1,\dots,H(i)}, \qquad A_{f,i}^M = \frac{1}{\sqrt{M}}({}_\lambda f_k^i)_{\lambda=1,\dots,M,k=1,\dots H(i)},$$
$$B_{g,i} = \mathbb{E}[g_k^i g_l^i]_{k,l=1,\dots,J(i)}, \qquad A_{g,i}^M = \frac{1}{\sqrt{M}}({}_\lambda g_k^i)_{\lambda=1,\dots,M,k=1,\dots J(i)},$$

the inner product matrices associated with the chosen basis. Then

$$\alpha_{i,\cdot}^{(\pi,p)} = B_{e,i}^{-1}\,\mathbb{E}\left[e_\cdot^i \cdot \mathcal{L}^{(\pi)}(i,p-1)\right],$$
$$\beta_{i,\cdot}^{(\pi,p)} = B_{f,i}^{-1}\,\mathbb{E}\left[f_\cdot^i \cdot \mathcal{L}^{(\pi)}(i+1,p-1)\frac{\Delta W_{t_i}^{(\pi)}}{\Delta}\right],$$
$$\gamma_{i,\cdot}^{(\pi,p)} = B_{g,i}^{-1}\,\mathbb{E}\left[g_\cdot^i \cdot \mathcal{L}^{(\pi)}(i+1,p-1)\right].$$

The simulation based analogue of $B_{\cdot,i}$ is given by

$$B_{\cdot,i}^M = (A_{\cdot,i}^M)^*(A_{\cdot,i}^M), \qquad \cdot = e, f, g.$$

Since the inverse of $B_{\cdot,i}^M$ may not exist, we shall make use of the pseudo-inverses $(\cdot)^+$. Note that

$$(A_{\cdot,i}^M)^+ = (B_{\cdot,i}^M)^{-1}(A_{\cdot,i}^M)^* \qquad \cdot = e, f, g.$$

If we denote by $\circ$ pointwise multiplication, the method can be expressed as

$$\begin{aligned}
\alpha_{i,\cdot}^{(\pi,0,M)} &= \beta_{i,\cdot}^{(\pi,0,M)} = \gamma_{i,\cdot}^{(\pi,0,M)} = 0,\\
\boldsymbol{\mathcal{Y}}_{t_i}^{(\pi,p,M)} &= ({}_{\lambda}\mathcal{Y}_{t_i}^{(\pi,p,M)})_{\lambda=1,\dots,M} = \sum_{k=1}^{K(i)} \alpha_{i,k}^{(\pi,p,M)} \mathbf{e}_k^i,\\
\boldsymbol{\mathcal{Z}}_{t_i}^{(\pi,p,M)} &= ({}_{\lambda}\mathcal{Z}_{t_i}^{(\pi,p,M)})_{\lambda=1,\dots,M} = \sum_{k=1}^{H(i)} \alpha_{i,k}^{(\pi,p,M)} \mathbf{f}_k^i,\\
\tilde{\boldsymbol{\mathcal{Z}}}_{t_i}^{(\pi,p,M)}(\cdot) &= ({}_{\lambda}\mathcal{Z}_{t_i}^{(\pi,p,M)}(\cdot)_{\lambda=1,\dots,M} = \sum_{k=1}^{J(i)} \alpha_{i,k}^{(\pi,p,M)} (\bar{\mathbf{g}}_k^i(\cdot) - \bar{\mathbf{g}}_k^i(0)),\\
\alpha_{i,\cdot}^{(\pi,p,M)} &= \frac{1}{\sqrt{M}} (A_{e,i}^M)^+ \left\{ \mathbf{F}^{(\pi)} + \sum_{j=i}^{N-1} \boldsymbol{f}^{(\pi)}(t_j)\Delta \right\},\\
\beta_{i,\cdot}^{(\pi,p,M)} &= \frac{1}{\sqrt{M}} (A_{f,i}^M)^+ \left\{ \left(\mathbf{F}^{(\pi)} + \sum_{j=i+1}^{N-1} \boldsymbol{f}^{(\pi)}(t_j)\Delta \right) \circ \frac{\Delta \mathbf{W}_i}{\Delta} \right\},\\
\gamma_{i,\cdot}^{(\pi,p,M)} &= \frac{1}{\sqrt{M}} (A_{g,i}^M)^+ \left\{ \mathbf{F}^{(\pi)} + \sum_{j=i+1}^{N-1} \boldsymbol{f}^{(\pi)}(t_j)\Delta \right\}.
\end{aligned} \tag{3.4.1}$$

The following Lemma is a direct consequence of the law of large numbers:

Lemma 3.4.1. *Under the Lipschitz condition outlined in the introduction we have that* $(\alpha_{i,\cdot}^{(\pi,p,M)}, \beta_{i,\cdot}^{(\pi,p,M)}, \gamma_{i,\cdot}^{(\pi,p,M)})$ *converges* $\mathbb{P}$*-a.s. to* $(\alpha_{i,\cdot}^{(\pi,p)}, \beta_{i,\cdot}^{(\pi,p)}, \gamma_{i,\cdot}^{(\pi,p)})$ *as* M *goes to infinity.*

The following Theorem is then an easy consequence

Theorem 3.4.2. *Under the Lipschitz condition outlined in the introduction we have that* $(\mathcal{Y}^{(\pi,p,M)}, \mathcal{Z}^{(\pi,p,M)}, \tilde{\mathcal{Z}}^{(\pi,p,M)})$ *converges* $\mathbb{P}$*-a.s. to* $(\mathcal{Y}^{(\pi,p)}, \mathcal{Z}^{(\pi,p)}, \tilde{\mathcal{Z}}^{(\pi,p)})$ *as* M *goes to infinity.*

Markovian setting

While the general setting concerns BSDEs with random terminal condition and generators which depend on a Brownian motion and a random measure in an arbitrary way, in this section we will assume that the randomness of the terminal condition and generator comes from the state of a Markovian forward SDE. These conditions are usually referred to as

"Markovian setting". Note that the solution of the backward component of the BSDE will now have to be measurable with respect to the filtration generated by the forward component. Intuitively, in the Markovian setting, the solution to the BSDE can be represented as a function of time and the (forward) state process. The results of the previous section can now be made more explicit. To avoid unnecessary overloads in notations once again we shall deal with the case $d_1 = d_2 = 1$.

Let us now consider the forward-backward stochastic differential equation, for $0 \leq t \leq T$:

$$\xi_t = x + \int_0^t \mu(\xi_{s-})\,\mathrm{d}s + \int_0^t \sigma(\xi_{s-})\,\mathrm{d}W_s + \int_0^t \int_{\mathbb{R}} \zeta(\xi_{s-}, e)\,\tilde{N}(\mathrm{d}s, \mathrm{d}e),$$
$$Y_t = \Psi(X(T)) + \int_t^T f(s, \xi_{s-}, Z_s, \tilde{Z}_s(\cdot))\,\mathrm{d}s - \int_t^T Z_s\,\mathrm{d}W_s - \int_t^T \int_{\mathbb{R}} \tilde{Z}_s(e)\,\tilde{N}(\mathrm{d}s, \mathrm{d}e).$$

From now on we shall restrict to a strong scheme (which is a particular case of a weak) and detail the implementation of the method in this setting.

We choose a partition $\pi = \{t_i; 0, \ldots, N\}$ of $[0, T]$ with mesh size $|\pi| = \max_i |t_{i+1} - t_i|$. In the following denote by $\Delta_i = t_{i+1} - t_i$. We discretize ξ by the Euler scheme

$$\begin{aligned} \xi_0^{(\pi)} &= x, \\ \xi_{t_{i+1}}^{(\pi)} &= \xi_{t_i}^{(\pi)} + \mu(\xi_{t_i}^{(\pi)})\Delta_i + \sigma(\xi_{t_i}^{(\pi)})\Delta W_i + \int_{\mathbb{R}} \zeta(\xi_{t_i}^{(\pi)}, e)\,\tilde{N}(t_i, t_{i+1}]\,\mathrm{d}e. \end{aligned} \tag{3.4.2}$$

Clearly, there exists a unique $\mathcal{F}^{X^{(\pi)}}$-adapted, square integrable solution $\xi^{(\pi)}$ to (3.4.2). We extend $\xi^{(\pi)}$ to a càdlàg process by piecewise constant interpolation. The terminal condition $F^{(\pi)}$ of the previous section will be of the form

$$F^{(\pi)} = \Psi^{(\pi)}(\Theta_{t_N}^{(\pi)}),$$

where $\{\Theta_{t_i}^{(\pi)}, \mathcal{F}_{t_i}^{(\pi)}\}$ is a Markov chain with $\xi_{t_i}^{(\pi)}$ as first components and can be extended to take into accounts functionals of the paths of the process. By the Markov property, the conditional expected values in (3.3.15) can be represented as functions of the state process $\xi^{(\pi)}$ only ($\xi^{(\pi)}$ and $\Delta\xi^{(\pi)}$ respectively); in other words, instead of conditioning on the whole filtration generated by the (discretised) forward process $\mathcal{F}_{t_i}^{(\pi)}$, it is enough to condition on the sigma algebra generated by the terminal value $\Theta_{t_i}^{(\pi)}$. Introduce the shorthand notation

$$L^{(\pi)}(i,p) := F^{(\pi)} + \sum_{j=i}^{N-1} f^{(\pi)}(t_j, \xi_{t_j}^{(\pi)}, Y_{t_j}^{(\pi,p)}, Z_{t_j}^{(\pi,p)}, \tilde{Z}_{t_j}^{(\pi,p)}(\cdot))\Delta.$$

Then, we have

$$Y_{t_i}^{(\pi,p)} = \mathbb{E}\left[L^{\pi}(i,p-1)\Big|\Theta_{t_i}^{(\pi)}\right], \quad Z_{t_i}^{(\pi,p)} = \mathbb{E}\left[\frac{\Delta W_{t_i}}{\Delta} L^{(\pi)}(i+1,p-1)\Big|\Theta_{t_i}^{(\pi)}\right],$$

and

$$\begin{aligned}\tilde{Z}_{t_i}^{(\pi,p)}(x) = \mathbb{E}&\left[L^{(\pi)}(i+1,p-1)\,\Big|\Theta_{t_i}^{(\pi)}, \int_{(t_i,t_{i+1}]\times\mathbb{R}} e\,\tilde{N}(\mathrm{d}s,\mathrm{d}e) = x\right]\\ &- \mathbb{E}\left[L^{(\pi)}(i+1,p-1)\,\Big|\Theta_{t_i}^{(\pi)}, \int_{(t_i,t_{i+1}]\times\mathbb{R}} e\,\tilde{N}(\mathrm{d}s,\mathrm{d}e) = 0\right].\end{aligned}$$

Choose two set of functions $\{\eta_1(x), \eta_2(x), \ldots, \eta_\kappa(x)\}$ and $\{\tilde{\eta}_1(x,y), \tilde{\eta}_2(x,y), \ldots, \tilde{\eta}_\iota(x,y)\}$ and define the basis

$$e_\cdot^i = f_\cdot^i = \eta_\cdot(\Theta_{t_i}^{\pi}), \qquad g_\cdot^i = \tilde{\eta}_\cdot(\Theta_{t_i}^{\pi}, \Delta X_{t_i})$$

In the multidimensional setting the choice of basis functions η and $\tilde{\eta}$ may depend on d_1 and d_2 but for simulation purposes it is more convenient to work with one set of functions only.

Y	$Y^{(\pi)}$	$Y^{(\pi,p)}$	$\mathcal{Y}^{(\pi,p)}$	$\mathcal{Y}^{(\pi,p,M)}$	$\boldsymbol{\mathcal{Y}}^{(\pi,p,M)}$
Z	$Z^{(\pi)}$	$Z^{(\pi,p)}$	$\mathcal{Z}^{(\pi,p)}$	$\mathcal{Z}^{(\pi,p,M)}$	$\boldsymbol{\mathcal{Z}}^{(\pi,p,M)}$
$\tilde{Z}(\cdot)$	$\tilde{Z}^{(\pi)}(\cdot)$	$\tilde{Z}^{(\pi,p)}(\cdot)$	$\tilde{\mathcal{Z}}^{(\pi,p)}(\cdot)$	$\tilde{\mathcal{Z}}^{(\pi,p,M)}(\cdot)$	$\tilde{\boldsymbol{\mathcal{Z}}}^{(\pi,p,M)}(\cdot)$

Table 3.1: This table summarises some of the notations used in this section. The columns from left to right denote respectively: the solution of the BSDE with jumps, the solution of the BSΔE, the solution of the forward scheme at the p^{th} iteration, the projection of the solution of the forward scheme on a finite dimensional subspace, the Monte Carlo estimator of the the projection of the solution of the forward scheme with M paths and the vector column of the latter.

3.5 Numerical Illustration

In this section we present some simulations for a g-expectation with respect to $(\mathcal{F}_t)_{0\leq t\leq T}$, the natural filtration generated by an exponential Lévy process S. Throughout the section, S is one dimensional representing a stock with ($\mathbb{Q}$ risk neutral) dynamics given by

$$S_t = S_0 \mathrm{e}^{\beta W_t + X_t - \kappa t},$$

where W is standard one-dimensional Brownian motion, X is a VG process independent of W and κ is the "convexity adjustment" given by

$$\kappa = \frac{\beta^2}{2} - \frac{1}{\nu} \log\left(1 - \theta\nu - \frac{1}{2}\sigma^2\nu\right).$$

The convexity adjustment is equal to the characteristic exponent of $\beta W + X$ evaluated at $-\mathtt{i}$ and makes S into a martingale. If we denote by $\gamma(t;\mu,\nu)$ a gamma process with mean μ and variance ν, and by B a Brownian motion independent of γ, and by

$$b(t;\theta,\sigma) = \theta t + \sigma B_t,$$

an arithmetic Brownian motion, then X is defined to be

$$X_t = b(\gamma(t;1,\nu);\theta,\sigma).$$

In other words, the VG process is obtained by evaluating arithmetic Brownian motion with drift θ and volatility σ at a random time given by a gamma process having a mean rate of 1 and a variance rate of ν. The VG process is a pure jump Lévy process of infinite activity and bounded variation. This means that it has an infinite number of jumps in any time interval but, unlike Brownian motion, the sum of absolute increments converges to a finite value. For the use of the VG process in option pricing we refer, for instance, to [88]. We discretise the process S by the log-Euler scheme on an equidistant partition of the interval $[0,T]$ consisting on $N+1$ points and denoted by π_N. In the following, we set the parameters

Table 3.2: Parameters

$S_0 = 100$	$\beta = 0.2$	$\theta = -0.33$	$\sigma = 0.12$	$\nu = 0.16$	$T = 1$

For any $\xi \in L^2(\mathcal{F}_t)$ denote the solution of

$$Y_t = \xi + \int_t^T f(s, Y_s, Z_s, \tilde{Z}_s)\,\mathrm{d}s - \int_t^T Z_s\,\mathrm{d}W_s - \int_t^T \int_{\mathbb{R}} \tilde{Z}_s(x)\,\tilde{N}(\mathrm{d}s, \mathrm{d}x),$$

by $(Y^\xi, Z^\xi, \tilde{Z}^\xi)$ and we define $\mathcal{E}_f(\xi) = Y_0^\xi$. It is possible to show that under suitable conditions on f, namely the ones that guarantee uniqueness and under which a comparison theorem holds, $\mathcal{E}_f$ is a non-linear expectation called f-expectation. A non-linear expectation is an operator preserving monotonicity and constants; it is very close to the classical concept of expectation in the sense that it satisfies almost all the classical properties except linearity. General non-linear expectations and the notion of g-expectation (in a Brownian setting) were introduced by Coquet *et al.* [34]; their results were subsequently extended to the setting of a filtration generated by Brownian motion and a compensated Poisson random measure by Royer [117]. Following [117], we denote by

$$\mathcal{E}^{\mu,C_1} = \mathcal{E}_{f_{\mu,C_1}} \qquad \text{and} \qquad \bar{\mathcal{E}}^{\mu,C_1} = \mathcal{E}_{\bar{f}_{\mu,C_1}}$$

two particular non-linear expectations, with f_{μ,C_1} and $\bar{f}_{\mu,C_1}$ defined by

$$\begin{aligned}
f_{\mu,C_1}(t,z,u) &= \mu\|z\| + |\mu| \int_{\mathbb{R}} (1 \wedge x) u^+(x)\,\lambda(\mathrm{d}x) - C_1 \int_{\mathbb{R}} (1 \wedge x) u^-(x)\,\lambda(\mathrm{d}x), \\
\bar{f}_{\mu,C_1}(t,z,u) &= -\mu\|z\| - |\mu| \int_{\mathbb{R}} (1 \wedge x) u^-(x)\,\lambda(\mathrm{d}x) + C_1 \int_{\mathbb{R}} (1 \wedge x) u^+(x)\,\lambda(\mathrm{d}x),
\end{aligned}$$

where λ is the Lévy measure associated to the compensated Poisson random measure. In our setting λ will therefore be the Lévy measure of the VG process which is given by the expression

$$\lambda(\mathrm{d}x) = \left(\frac{\mu_-^2}{\nu_-} \frac{\mathrm{e}^{-\frac{\mu_-}{\nu_-}|x|}}{|x|} \mathbb{1}_{(-\infty,0)}(x) + \frac{\mu_+^2}{\nu_+} \frac{\mathrm{e}^{-\frac{\mu_+}{\nu_+}x}}{x} \mathbb{1}_{(0,\infty)}(x) \right) \mathrm{d}x,$$

with $\mu_{\pm} = \frac{1}{2}\sqrt{\theta^2 + \frac{2\sigma^2}{\nu}} \pm \frac{\theta}{2}$ and $\nu_{\pm} = \mu_{\pm}^2 \nu$. Let us denote by $\mathcal{M}^2$ the set of square integrable martingales. By the martingale representation theorem we can define the mapping

$$\Phi : M \mapsto (\theta, \upsilon), \qquad M_t = \int_0^t \theta_s \, \mathrm{d}W_s + \int_0^t \int_{\mathbb{R}} \upsilon_s(x) \, \tilde{N}(\mathrm{d}s, \mathrm{d}x).$$

The following is Proposition 3.6 in [117], which we recall

Proposition 3.5.1. *Fix* $\mu > 0$ *and* $-1 < C_1 \leq 0$. *Consider the set of probabilities* $\tilde{\mathcal{Q}} = \{\mathrm{d}\tilde{\mathbb{Q}} := \mathcal{E}_T(\tilde{M})\mathrm{d}\mathbb{Q} \,|\, \widetilde{M} \in \widetilde{\mathcal{M}}\}$ *with*

$$\widetilde{\mathcal{M}} = \{\widetilde{M} \in \mathcal{M}^2 \,|\, \|\tilde{\theta}_s\| \leq \mu, \tilde{\upsilon}_s^+ \leq \mu(1 \wedge |x|), \tilde{\upsilon}_s^- \leq C_1(1 \wedge |x|) \text{ where } (\tilde{\theta}, \tilde{\upsilon}) = \Phi(\widetilde{M})\}.$$

Then

$$\bar{\mathcal{E}}^{\mu, C_1}(\eta|\mathcal{F}_t) = \inf_{\tilde{\mathbb{Q}} \in \tilde{\mathcal{Q}}} \mathbb{E}^{\tilde{\mathbb{Q}}}[\xi|\mathcal{F}_t], \qquad \mathcal{E}^{\mu, C_1}(\xi|\mathcal{F}_t) = \sup_{\tilde{\mathbb{Q}} \in \tilde{\mathcal{Q}}} \mathbb{E}^{\tilde{\mathbb{Q}}}[\xi|\mathcal{F}_t].$$

As illustration, we numerically evaluate the non-linear expectation $\mathcal{E}^{\mu, C_1}(\xi)$ with $\xi = |S_T - \mathcal{K}|$, $\mathcal{K} = 100$. Observe that by making the choice of $\mu = C_1 = 0$, we have that Y_0 is given by the expectation of the final payoff and therefore is the (risk neutral) price of an at the money straddle. By virtue of the previous proposition $\mathcal{E}^{\mu, C_1}(\xi)$ can then be interpreted as the supremum of linear expectations over a set of measures that "move away" from $\mathbb{Q}$ in a manner dictated by the parameters μ and C_1. Observe that, when rates are zero, by a no-arbitrage argument, the price of a straddle at time 0 is given by: $2C_T(\mathcal{K}) - S_0 + \mathcal{K}$, where $C_T(\mathcal{K})$ is the price at time zero of a call option maturing at T with strike $\mathcal{K}$. The price of the call option is known analytically using the inverse Fourier transform

$$C_T(\log(\mathcal{K})) = \frac{\mathrm{e}^{-\alpha \log(\mathcal{K})}}{2\pi} \int_{-\infty}^{\infty} \mathrm{e}^{-\mathtt{i}v \log(K)} \frac{\psi_T(v - (\alpha+1)\mathtt{i})}{\alpha^2 + \alpha - v^2 + \mathtt{i}(2\alpha+1)v} \, \mathrm{d}v,$$

where ψ_T is the characteristic exponent of $\log(S_T)$

$$\psi_T(u) = \mathrm{e}^{\mathtt{i}u(\log(S_0) - \kappa T) - \frac{1}{2}\beta^2 T u^2} \left(1 - \theta\nu u + \frac{1}{2}\sigma^2 \nu u^2\right)^{-T/\nu},$$

and α is a fixed positive value called dampening factor. Setting $\alpha = 1.5$ and performing the Fourier inversion numerically we obtain a value for the straddle of 20.93, which we shall use as a benchmark. For further information on option valuation using Fourier techniques we refer to Carr & Madan [25]. The bases function for the empirical regression on Y and Z are given by monomials and the straddle payoff; more precisely:

$$\eta_1(x) = |x - \mathcal{K}|, \qquad \eta_k = (x - S_0)^{k-2}, \quad 2 \leq k \leq K.$$

For the computation of $\tilde{Z}$ we used the bases $\tilde{\eta}_k(x, y), 1 \leq k \leq J$ that need to satisfy the condition

$$\tilde{\eta}_k(\Theta^{\pi}_{t_i}, 0) = 0,$$

which from a financial viewpoint corresponds to the fact that if there is no jump in the underlying, the jump-delta should be zero:

$$\begin{gathered}
\tilde{\eta}_1(x, y) = y, \\
\tilde{\eta}_2(x, y) = xy, \quad \tilde{\eta}_3(x, y) = y^2, \\
\tilde{\eta}_4(x, y) = x^2 y, \quad \tilde{\eta}_5(x, y) = xy^2, \quad \tilde{\eta}_6(x, y) = y^3. \\
\dots
\end{gathered} \tag{3.5.1}$$

We chose $K = J = 7$ (independently of i) and denote the total number of calculated iteration by n_{Picard}. Typically the convergence in the number of Picard iterations is quite fast and in the following simulations we set $n_{Picard} = 5$. Fig. 3.1 shows the empirical mean and empirical standard deviation of $\mathcal{E}^{\mu, C_1}(\xi)$ with $\mu = 0.5$ and $C_1 = 0$ from 30 launches of the algorithm as a function of the number of simulated paths per launch. For this, we chose $N = 10$. Fig 3.2 represents $\mathcal{E}^{\mu, C_1}(\xi)$ as a function of the number of partition points. For each run of the algorithm, we simulate 100000 paths of the forward component. As expected in view of the representation of $\mathcal{E}^{\mu, C_1}(\xi)$ given in Proposition 3.5.1 we notice that $\mathcal{E}^{\mu, C_1}(\xi)$ is indeed an increasing function of μ.

3.A Appendix

Recall that, by the L^2 theory of conditional expectation, if a scalar random variable $X \in L^2$, then its conditional expectation given another (possibly multidimensional) random

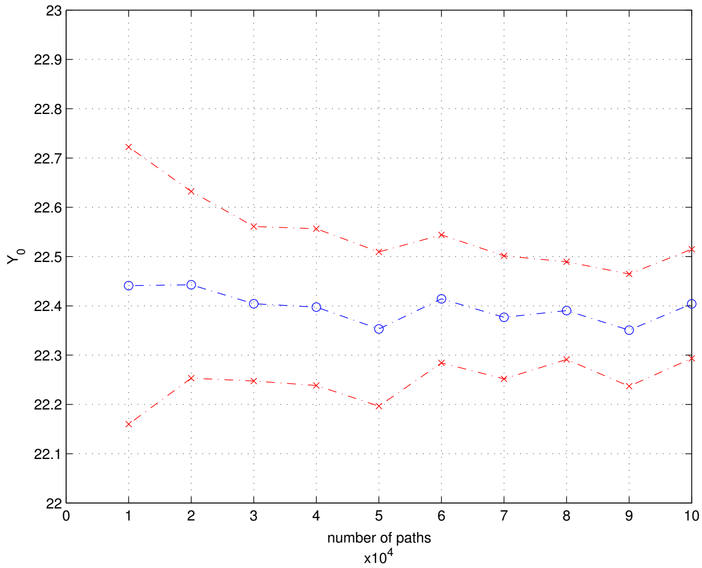

Figure 3.1: Empirical mean (in blue) and empirical standard deviation (in red) of 30 launches as a function of the number of paths in the case of $\mu = 0.5$ and $C_1 = 0$.

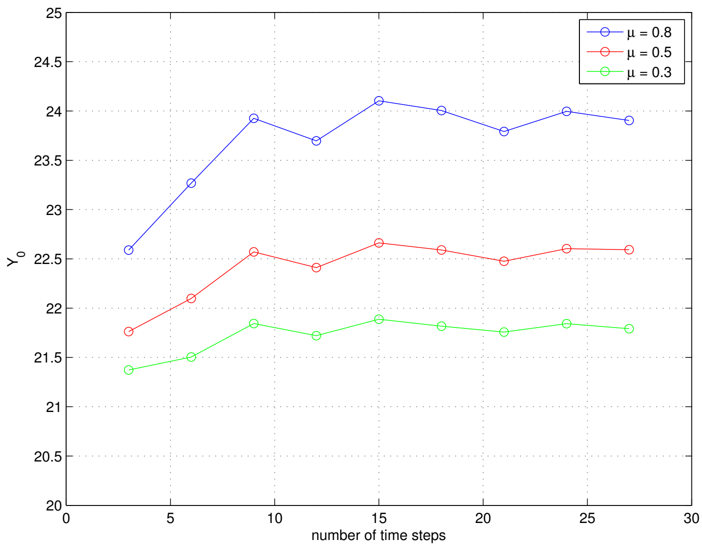

Figure 3.2: Y_0 computed with $M = 100000$ paths and $n_{Picard} = 7$ as a function of the number of partition points for different values of μ. We verify the theoretical result that the solution is an increasing function of μ.

variable Y is given by its projection on $L^2(\sigma(Y))$ i.e.

$$\mathbb{E}[X|Y] = \underset{\varphi(Y)}{\operatorname{argmin}}\, \mathbb{E}[|X - \varphi(Y)|^2]$$

where $\varphi(\cdot)$ is a measurable function such that $\mathbb{E}[\varphi(Y)^2] < \infty$. This is a least square problem in infinite dimension known in the literature as regression problem. Since L^2 is a Hilbert space, it has a countable orthonormal basis and the conditional expectation can be represented as a (limit) of linear combinations of elements of the basis. To approximate the regression function φ we project on a finite dimensional space by considering the first K elements of the basis $\{\eta_i\}_{i=1}^K$. We then proceed to compute the coefficients of the basis by solving a empirical least square problem

$$(\alpha_k^M) = \underset{\alpha \in \mathbb{R}^K}{\operatorname{argmin}} \frac{1}{M} \sum_{i=1}^M (X_i - \sum_{j=1}^K \alpha_j \eta_j(Y_i))^2,$$

where (X_i, Y_i), $i = 1, \ldots, M$ are independent simulations of the couple (X, Y). For the approximation of $\varphi(\cdot) = \mathbb{E}[X|Y = \cdot]$ we set

$$\varphi_M(\cdot) = \sum_{j=1}^K \alpha_j^M \eta_j(\cdot).$$

Without further assumptions it is possible to derive error estimates on the approximation. For further details, see [62].

Theorem 3.A.1. *Assume that* $\mathbb{E}[X|Y] = \varphi(Y)$ *and that* (X_i, Y_i), $i = 1, \ldots, M$ *are independent simulations of the couple* (X, Y). *Furthermore, let* $\sigma^2 = \sup_y Var(X|Y = y) < \infty$ *and let* V *be the linear space given by the span of the* K *basis elements* $\{\eta_i\}_{i=1}^K$. *Then, we have*

$$\mathbb{E}\left[\frac{1}{M}\sum_{i=1}^M (\varphi_M(Y_i) - \varphi(Y_i))^2\right] \le \sigma^2 \frac{K}{M} + \min_{\psi \in V} \|\psi - \varphi\|^2_{L^2(\mathbb{P}^Y)}.$$

The first term in the right hand side above is interpreted as the statistical error term, while the second summand is the error due to the projection onto a finite dimensional space. As K goes to infinity, the first term goes to infinity while the second one converges to zero.

He did not know that the new life would not be given him for nothing... But that is the beginning of a new story – the story of the gradual renewal of a man, the story of his gradual regeneration, of his passing from one world into another, of his initiation into a new unknown life. That might be the subject of a new story, but our present story is ended.

(F. Dostoevsky, Crime and Punishment, Epilogue.)

References

[1] S. AAZIZI, Discrete-time approximation of decoupled forward-backward stochastic differential equations driven by pure jump Lévy processes, Advances in Applied Probability, 45 (2013), pp. 791–821.

[2] C. ALBANESE AND A. MIJATOVIĆ, Convergence rates for diffusions on continuous-time lattices, Available at SSRN : http://ssrn.com/abstract=1018609 or http://dx.doi.org/10.2139/ssrn.1018609.

[3] K. AMIN AND A. KHANNA, Convergence of American option values from discrete- to continuous-time financial models, Mathematical Finance, 4 (1994), pp. 289–304.

[4] F. ANTONELLI, Backward-forward stochastic differential equations, The Annals of Applied Probability, 3 (1993), pp. 777–793.

[5] D. APPLEBAUM, Lévy Processes and Stochastic Calculus, Cambridge Studies in Advanced Mathematics, Cambridge University Press, 2009.

[6] S. ASMUSSEN AND J. ROSIŃSKI, Approximations of small jumps of Lévy processes with a view towards simulation, Journal of Applied Probability, 38 (2001), pp. 482–493.

[7] V. BALLY, Approximation scheme for solutions of BSDE, in Backward stochastic differential equations (Paris, 1995–1996), vol. 364 of Pitman Research Notes in Mathematics Series, Longman, Harlow, 1997, pp. 177–191.

[8] G. BARLES, R. BUCKDAHN, AND É. PARDOUX, Backward stochastic differential equations and integral-partial differential equations, Stochastics and Stochastics Reports, 60 (1997), pp. 57–83.

[9] O. E. BARNDORFF-NIELSEN, Processes of normal inverse Gaussian type, Finance and Stochastics, 2 (1997), pp. 41–68.

[10] D. BECHERER, Bounded solutions to backward SDE's with jumps for utility optimization and indifference hedging, The Annals of Applied Probability, 16 (2006), pp. 2027–2054.

[11] C. BENDER AND R. DENK, A forward scheme for backward SDEs, Stochastic Processes and their Applications, 117 (2007), pp. 1793–1812.

[12] C. BENDER AND J. ZHANG, Time discretization and Markovian iteration for coupled FBSDEs, The Annals of Applied Probability, 18 (2008), pp. 143–177.

[13] J. BERTOIN, Lévy Processes, vol. 121 of Cambridge Tracts in Mathematics, Cambridge University Press, 1996.

[14] J. BERTOIN, Subordinators: examples and applications, in Lectures on probability theory and statistics (Saint-Flour, 1997), vol. 1717 of Lecture Notes in Mathematics, Springer, Berlin, 1999.

[15] N. H. BINGHAM, C. M. GOLDIE, AND J. L. TEUGELS, Regular variation, vol. 27 of Encyclopedia of Mathematics and its Applications, Cambridge University Press, Cambridge, 1989.

[16] J. M. BISMUT, Conjugate convex functions in optimal stochastic control, Journal of Mathematical Analysis and Applications, 44 (1973), pp. 384–404.

[17] S. BOCHNER, Diffusion equation and stochastic processes, Proceedings of the National Academy of Sciences of the USA, 35 (1949), pp. 368–370.

[18] S. BOCHNER, Harmonic analysis and the theory of probability, University of California Press, Berkeley and Los Angeles, 1955.

[19] B. Bouchard, I. Ekeland, and N. Touzi, On the Malliavin approach to Monte Carlo approximation of conditional expectations, Finance and Stochastics, 8 (2004), pp. 45–71.

[20] B. Bouchard and R. Elie, Discrete-time approximation of decoupled forward–backward sde with jumps, Stochastic Processes and their Applications, 118 (2008), pp. 53–75.

[21] B. Bouchard and N. Touzi, Discrete-time approximation and Monte-Carlo simulation of backward stochastic differential equations, Stochastic Processes and their applications, 111 (2004), pp. 175–206.

[22] S. I. Boyarchenko and S. Z. Levendorskiĭ, Non-Gaussian Merton-Black-Scholes theory, vol. 9 of Advanced Series on Statistical Science & Applied Probability, World Scientific Publishing Co., 2002.

[23] P. Briand, B. Delyon, and J. Mémin, Donsker-type theorem for BSDEs, Electronic Communications in Probability, 6 (2001), pp. 1–14.

[24] P. Briand, B. Delyon, and J. Mémin, On the robustness of backward stochastic differential equations, Stochastic Processes and their Applications, 97 (2002), pp. 229–253.

[25] P. Carr and D. B. Madan, Option valuation using the fast fourier transform, Journal of Computational Finance, 2 (1999), pp. 61–73.

[26] J.-F. Chassagneux and D. Crisan, Runge-Kutta schemes for backward stochastic differential equations, The Annals of Applied Probability, 24 (2014), pp. 679–720.

[27] P. Cheridito and M. Stadje, BSΔEs and BSDEs with non-Lipschitz drivers: comparison, convergence and robustness, Bernoulli, 19 (2013), pp. 1047–1085.

[28] D. Chevance, Numerical methods for backward stochastic differential equations, in Numerical methods in finance, Publications of the Newton Institute, Cambridge University Press, Cambridge, 1997, pp. 232–244.

[29] C. CHUI, Concerning rates of convergence of Riemann sums, Journal of Approximation Theory, 4 (1971), pp. 279–287.

[30] E. ÇINLAR, J. JACOD, P. PROTTER, AND M. J. SHARPE, Semimartingales and Markov processes, Zeitschrift für Wahrscheinlichkeitstheorie und Verwandte Gebiete, 54 (1980), pp. 161–219.

[31] P. K. CLARK, A subordinated stochastic process model with finite variance for speculative prices, Econometrica, 41 (1973), pp. 135–156.

[32] S. COHEN AND J. ROSIŃSKI, Gaussian approximation of multivariate Lévy processes with applications to simulation of tempered stable processes, Bernoulli, 13 (2007), pp. 195–210.

[33] R. CONT AND P. TANKOV, Financial Modelling with Jump Processes, Financial Mathematics Series, Chapman & Hall/CRC, Boca Raton, FL, second ed., 2009.

[34] F. COQUET, Y. HU, J. MÉMIN, AND S. PENG, Filtration-consistent nonlinear expectations and related g-expectations, Probability Theory and Related Fields, 123 (2002), pp. 1–27.

[35] F. COQUET, J. MÉMIN, AND L. SŁOMINSKI, On weak convergence of filtrations, in Séminaire de probabilités XXXV, Springer, 2001, pp. 306–328.

[36] J. C. COX, S. A. ROSS, AND M. RUBINSTEIN, Option pricing: A simplified approach, Journal of Financial Economics, 7 (1979), pp. 229–263.

[37] M. G. CRANDALL AND P.-L. LIONS, Viscosity solutions of Hamilton-Jacobi equations, Transactions of the American Mathematical Society, 277 (1983), pp. 1–42.

[38] D. CRISAN AND K. MANOLARAKIS, Probabilistic methods for semilinear partial differential equations. Applications to finance, M2AN Mathematical Modelling and Numerical Analysis, 44 (2010), pp. 1107–1133.

[39] ——, Solving backward stochastic differential equations using the cubature method: application to nonlinear pricing, SIAM Journal on Financial Mathematics, 3 (2012), pp. 534–571.

[40] D. CRISAN AND K. MANOLARAKIS, Second order discretization of backward SDEs and simulation with the cubature method, The Annals of Applied Probability, 24 (2014), pp. 652–678.

[41] J. CVITANIĆ AND I. KARATZAS, Hedging contingent claims with constrained portfolios, The Annals of Applied Probability, 3 (1993), pp. 652–681.

[42] K. È. DAMBIS, On decomposition of continuous submartingales, Theory of Probability and its Applications, 10 (1965), pp. 401–410.

[43] F. DELARUE AND S. MENOZZI, A forward-backward stochastic algorithm for quasi-linear PDEs, The Annals of Applied Probability, 16 (2006), pp. 140–184.

[44] E. DERMAN, I. KANI, AND N. CHRISS, Implied trinomial trees of the volatility smile, The Journal of Derivatives, 3 (1996), pp. 7–22.

[45] J. DOUGLAS, J. MA, AND P. PROTTER, Numerical methods for forward-backward stochastic differential equations, The Annals of Applied Probability, 6 (1996), pp. 940–968.

[46] L. DUBINS AND G. SCHWARZ, On continuous martingales, Proceedings of the National Academy of Sciences USA, 53 (1965), pp. 913–916.

[47] D. DUFFIE AND L. G. EPSTEIN, Stochastic differential utility, Econometrica, 60 (1992), pp. 353–394.

[48] R. DURRETT, Probability: theory and examples, Cambridge Series in Statistical and Probabilistic Mathematics, Cambridge University Press, Cambridge, fourth ed., 2010.

[49] N. EL KAROUI AND S.-J. HUANG, A general result of existence and uniqueness of backward stochastic differential equations, in Backward stochastic differential

equations (Paris, 1995–1996), vol. 364 of Pitman Research Notes in Mathematics Series, Longman, Harlow, 1997, pp. 27–36.

[50] N. El Karoui, E. Pardoux, and M. C. Quenez, Reflected backward SDEs and American options, in Numerical methods in finance, vol. 13 of Publications of the Newton Institute, Cambridge University Press, Cambridge, 1997, pp. 215–231.

[51] N. El Karoui, S. Peng, and M. C. Quenez, Backward stochastic differential equations in finance, Mathematical Finance, 7 (1997), pp. 1–71.

[52] N. El Karoui and M.-C. Quenez, Dynamic programming and pricing of contingent claims in an incomplete market, SIAM Journal on Control and Optimization, 33 (1995), pp. 29–66.

[53] S. N. Ethier and T. G. Kurtz, Markov Processes: Characterization and Convergence, Wiley, 1986.

[54] A. Eyraud-Loisel, Backward stochastic differential equations with enlarged filtration: option hedging of an insider trader in a financial market with jumps, Stochastic Processes and their Applications, 115 (2005), pp. 1745–1763.

[55] W. Feller, Completely monotone functions and sequences, Duke Mathematical Journal, 5 (1939), pp. 661–674.

[56] W. Feller, An introduction to probability theory and its applications. Vol. II., John Wiley & Sons Inc., New York, second ed., 1971.

[57] W. H. Fleming and H. M. Soner, Controlled Markov processes and viscosity solutions, vol. 25 of Applications of Mathematics (New York), Springer-Verlag, New York, 1993.

[58] T. Fujiwara and H. Kunita, Stochastic differential equations of jump type and Lévy processes in diffeomorphisms group, Journal of Mathematics of Kyoto University, 25 (1985), pp. 71–106.

[59] H. Geman, D. B. Madan, and M. Yor, Time changes for Lévy processes, Mathematical Finance, 11 (2001), pp. 79–96.

[60] I. GIKHMAN AND A. SKOROKHOD, Introduction to the Theory of Random Processes, Saunders, Philadelphia, 1969.

[61] E. GOBET, J.-P. LEMOR, AND X. WARIN, A regression-based Monte Carlo method to solve backward stochastic differential equations, The Annals of Applied Probability, 15 (2005), pp. 2172–2202.

[62] L. GYÖRFI, M. KOHLER, A. KRZYŻAK, AND H. WALK, A distribution-free theory of nonparametric regression, Springer Series in Statistics, Springer-Verlag, New York, 2002.

[63] H. HE, Convergence from discrete to continuous-time contingent claims prices, Review of Financial Studies, 3 (1990), pp. 523–546.

[64] S. HESTON AND G. ZHOU, On the rate of convergence of discrete-time contingent claims, Mathematical Finance, 10 (2000), pp. 53–75.

[65] Y. HU, P. IMKELLER, AND M. MÜLLER, Utility maximization in incomplete markets, The Annals of Applied Probability, 15 (2005), pp. 1691–1712.

[66] J. HULL AND A. WHITE, The use of the control variate technique in option pricing, Journal of Financial and Quantitative Analysis, 23 (1988), pp. 237–251.

[67] K. ITÔ AND H. P. MCKEAN, Diffusion processes and their sample paths, Die Grundlehren der Mathematischen Wissenschaften, Band 125, Academic Press Inc., Publishers, New York, 1965.

[68] N. JACOB, Pseudo differential operators and Markov processes. Vol. I, Imperial College Press, London, 2001. Fourier analysis and semigroups.

[69] J. JACOD AND A. SHIRYAEV, Limit theorems for stochastic processes, vol. 288 of Grundlehren der Mathematischen Wissenschaften, Springer-Verlag, Berlin, second ed., 2003.

[70] E. JOUINI AND H. KALLAL, Arbitrage in securities markets with short-sales constraints, Mathematical Finance, 5 (1995), pp. 197–232.

[71] I. KARATZAS AND S. SHREVE, Brownian motion and stochastic calculus, Springer, 1991.

[72] S. KLÖPPEL AND M. SCHWEIZER, Dynamic indifference valuation via convex risk measures, Mathematical Finance, 17 (2007), pp. 599–627.

[73] M. KOBYLANSKI, Résultats d'existence et d'unicité pour des équations différentielles stochastiques rétrogrades avec des générateurs à croissance quadratique, Comptes Rendus de l'Académie des Sciences. Série I. Mathématique, 324 (1997), pp. 81–86.

[74] V. N. KOLOKOLTSOV, The Lévy-Khintchine type operators with variable Lipschitz continuous coefficients generate linear or nonlinear Markov processes and semigroups, Probability Theory and Related Fields, 151 (2011), pp. 95–123.

[75] ——, Markov processes, semigroups and generators, vol. 38 of de Gruyter Studies in Mathematics, Walter de Gruyter & Co., Berlin, 2011.

[76] T. G. KURTZ, Extensions of Trotter's operator semigroup approximation theorems, Journal of Functional Analysis, 3 (1969), pp. 354–375.

[77] T. G. KURTZ AND P. PROTTER, Weak limit theorems for stochastic integrals and stochastic differential equations, The Annals of Probability, 19 (1991), pp. 1035–1070.

[78] H. J. KUSHNER, On the weak convergence of interpolated Markov chains to a diffusion, The Annals of Probability, 2 (1974), pp. 40–50.

[79] H. J. KUSHNER AND P. DUPUIS, Numerical Methods for Stochastic Control Problems in Continuous Time, Applications of Mathematics, Springer, 2000.

[80] J. P. LEPELTIER AND J. SAN MARTIN, Backward stochastic differential equations with continuous coefficient, Statistics & Probability Letters, 32 (1997), pp. 425–430.

[81] A. E. B. LIM, Mean-variance hedging when there are jumps, SIAM Journal on Control and Optimization, 44 (2005), pp. 1893–1922.

[82] R. LIPTSER AND A. N. SHIRYAEV, On a problem of necessary and sufficient conditions in the functional central limit theorem for local martingales, Probability Theory and Related Fields, 59 (1982), pp. 311–318.

[83] F. A. LONGSTAFF AND E. S. SCHWARTZ, Valuing american options by simulation: A simple least-squares approach, Review of Financial studies, 14 (2001), pp. 113–147.

[84] H. LUSCHGY AND G. PAGÈS, Moment estimates for Lévy processes, Electronic Communications in Probability, 13 (2008), pp. 422–434.

[85] T. LYONS AND N. VICTOIR, Cubature on Wiener space, Proceedings of the Royal Society of London. Series A: Mathematical, Physical and Engineering Sciences, 460 (2004), pp. 169–198.

[86] J. MA, P. PROTTER, AND J. M. YONG, Solving forward-backward stochastic differential equations explicitly—a four step scheme, Probability Theory and Related Fields, 98 (1994), pp. 339–359.

[87] J. MA, J. SHEN, AND Y. ZHAO, On numerical approximations of forward-backward stochastic differential equations, SIAM Journal on Numerical Analysis, 46 (2008), pp. 2636–2661.

[88] D. MADAN, P. CARR, AND E. CHANG, The variance gamma process and option pricing., European Finance Review, 2 (1998), pp. 79–105.

[89] D. MADAN, F. MILNE, AND H. SHEFRIN, The multinomial option pricing model and its Brownian and Poisson limits, Review of Financial Studies, 2 (1989), pp. 251–265.

[90] D. MADAN, M. PISTORIUS, AND M. STADJE, Convergence of BSΔEs driven by random walks to BSDEs: the case of (in)finite activity jumps with general driver, arXiv preprint: arXiv:1406.7145, (2014).

[91] M. MANIA AND M. SCHWEIZER, Dynamic exponential utility indifference valuation, The Annals of Applied Probability, 15 (2005), pp. 2113–2143.

[92] R. C. MERTON, Option pricing when underlying stock returns are discontinuous, Journal of Financial Economics, 3 (1976), pp. 125–144.

[93] A. MIJATOVIĆ AND M. PISTORIUS, Continuously monitored barrier options under Markov processes, Mathematical Finance, 23 (2013), pp. 1–38.

[94] A. MIJATOVIĆ, M. VIDMAR, AND S. JACKA, Markov chain approximations for transition densities of Lévy processes, arXiv preprint arXiv:1211.0476, (2012).

[95] G. N. MILSTEIN AND M. V. TRETYAKOV, Numerical algorithms for forward-backward stochastic differential equations, SIAM Journal on Scientific Computing, 28 (2006), pp. 561–582.

[96] I. MONROE, Processes that can be embedded in Brownian motion, The Annals of Probability, 6 (1978), pp. 42–56.

[97] M.-A. MORLAIS, Quadratic BSDEs driven by a continuous martingale and applications to the utility maximization problem, Finance and Stochastics, 13 (2009), pp. 121–150.

[98] M.-A. MORLAIS, A new existence result for quadratic BSDEs with jumps with application to the utility maximization problem, Stochastic Processes and their Applications, 120 (2010), pp. 1966–1995.

[99] A. B. MUKHIN, Local limit theorems for lattice random variables, Theory of Probability & Its Applications, 36 (1991), pp. 698–713.

[100] J. R. NORRIS, Markov chains, vol. 2 of Cambridge Series in Statistical and Probabilistic Mathematics, Cambridge University Press, Cambridge, 1997.

[101] D. NUALART AND W. SCHOUTENS, Backward stochastic differential equations and Feynman-Kac formula for lévy processes, with applications in finance, Bernoulli, 7 (2001), pp. 761–776.

[102] S. OREY, On continuity properties of infinitely divisible distribution functions, Annals of Mathematical Statistics, 39 (1968), pp. 936–937.

[103] É. PARDOUX AND S. PENG, Backward stochastic differential equations and quasilinear parabolic partial differential equations, in Stochastic partial differential equations and their applications (Charlotte, NC, 1991), vol. 176 of Lecture Notes in Control and Information Science, Springer, Berlin, 1992, pp. 200–217.

[104] É. PARDOUX AND S. G. PENG, Adapted solution of a backward stochastic differential equation, Systems & Control Letters, 14 (1990), pp. 55–61.

[105] ——, Backward doubly stochastic differential equations and systems of quasilinear SPDEs, Probability Theory and Related Fields, 98 (1994), pp. 209–227.

[106] S. G. PENG, A nonlinear Feynman-Kac formula and applications, in Control theory, stochastic analysis and applications (Hangzhou, 1991), World Scientific Publishing, River Edge, NJ, 1991, pp. 173–184.

[107] ——, Probabilistic interpretation for systems of quasilinear parabolic partial differential equations, Stochastics and Stochastics Reports, 37 (1991), pp. 61–74.

[108] ——, A generalized dynamic programming principle and Hamilton-Jacobi-Bellman equation, Stochastics and Stochastics Reports, 38 (1992), pp. 119–134.

[109] ——, Stochastic Hamilton-Jacobi-Bellman equations, SIAM Journal on Control and Optimization, 30 (1992), pp. 284–304.

[110] ——, Backward stochastic differential equations and applications to optimal control, Applied Mathematics and Optimization, 27 (1993), pp. 125–144.

[111] R. S. PHILLIPS, On the generation of semigroups of linear operators., Pacific Journal of Mathematics, 2 (1952), pp. 343–369.

[112] E. PLATEN, Higher order weak approximation of Itô diffusion by Markov chains, Probability in Engineering and Informational Sciences, 6 (1992), pp. 391–408.

[113] D. POLLARD, Convergence of Stochastic Processes, Springer Series in Statistics, Springer-Verlag, 1984.

[114] M. C. Reed and B. Simon, Methods of modern mathematical physics I. Functional analysis, Academic Press Inc., New York, second ed., 1980.

[115] D. Revuz and M. Yor, Continuous martingales and Brownian motion, vol. 293 of Grundlehren der Mathematischen Wissenschaften [Fundamental Principles of Mathematical Sciences], Springer-Verlag, Berlin, third ed., 1999.

[116] J. Rosiński, Series representations of Lévy processes from the perspective of point processes, in Lévy processes, Birkhäuser Boston, Boston, MA, 2001, pp. 401–415.

[117] M. Royer, Backward stochastic differential equations with jumps and related non-linear expectations, Stochastic Processes and their Applications, 116 (2006), pp. 1358–1376.

[118] G. Samorodnitsky and M. S. Taqqu, Stable non-Gaussian random processes, Stochastic Modeling, Chapman & Hall, New York, 1994. Stochastic models with infinite variance.

[119] K. Sato, Lévy Processes and Infinitely Divisible Distributions, vol. 68 of Cambridge Studies in Advanced Mathematics, Cambridge University Press, 1999.

[120] D. W. Stroock and S. R. S. Varadhan, Multidimensional Diffusion Processes, vol. 233 of Grundlehren Der Mathematischen Wissenschaften, Springer-Verlag, New York, 1979.

[121] S. J. Tang and X. J. Li, Necessary conditions for optimal control of stochastic systems with random jumps, SIAM Journal on Control and Optimization, 32 (1994), pp. 1447–1475.

[122] H. F. Trotter, Approximation of semi-groups of operators, Pacific Journal of Mathematics, 8 (1958), pp. 887–919.

[123] H. G. Tucker, Absolute continuity of infinitely divisible distributions, Pacific Journal of Mathematics, 12 (1962), pp. 1125–1129.

[124] J. B. WALSH, The rate of convergence of the binomial tree scheme, Finance and Stochastics, 7 (2003), pp. 337–361.

[125] J. ZHANG, A numerical scheme for BSDEs, The Annals of Applied Probability, 14 (2004), pp. 459–488.