**Technische Universiteit Delft**
**Faculteit Elektrotechniek, Wiskunde en Informatica**
**Delft Institute of Applied Mathematics**

# A numerical Fourier cosine method for forward backward stochastic differential equations with jumps.

Verslag ten behoeve van het
Delft Institute of Applied Mathematics
als onderdeel ter verkrijging

van de graad van

**BACHELOR OF SCIENCE**
**in**
**TECHNISCHE WISKUNDE**

**door**

**Sander Antonius Wilhelmus Blok**

**Delft, Nederland**
**Juli 2020**

**BSc verslag TECHNISCHE WISKUNDE**

**“A numerical Fourier cosine method for forward backward stochastic differential equations with jumps.”**


Sander Antonius Wilhelmus Blok

**Technische Universiteit Delft**


**Begeleider**

Prof.dr.ir C.W. Oosterlee

**Overige commissieleden**

Dr. N.V. Budko

Dr. B. van den Dries

Juli, 2020

Delft

# Contents

# Abstract

In financial and egineering problems, we are often faced with solving Partial-Integro Differential Equations (PIDEs). Rarely we can find an analytic solution in a closed form expression for these PIDEs, hence we turn to numerical schemes to accurately approximate the solution instead. Classically these methods are based on finite difference methods, however, we can turn certain kinds of PIDEs into a probabilistic representation, called Forward Backward Stochastic Differential Equations with Jumps (FBSDEJs). Solving the PIDE can now be done alternatively by solving a FBSDEJ.

In this thesis we will first investigate the stochastic framework behind FBSDEJs and we will look into the uniqueness and existence of their solutions. Furthermore we propose a new numerical method which can efficiently solve FBSDEJs. The semi-discretisation is based on the classical Backward Differentiation Formula (BDF) methods, for the computation of the conditional expectations we use the COS method which makes use of Fourier cosine expansions, exploiting the knowledge we have about characteristic functions. Finally we implement the new method and we investigate it extensively both numerically and theoretically.

We show that the BDFn schemes are highly stable and efficient for computing FBSDEJs, the initial steps still have to be investigated in greater detail so that we can make use of the high-order BDFn schemes.

# CHAPTER 1
# Introduction and Outline of this Thesis

Over the past decades, mathematical finance has become a new scientific discipline on its own. The theory of finance attempts to describe the mechanisms behind the financial markets, to make them more efficient, but also to regulate them. It tries to explain and enhance the important role the financial markets play in efficient capital allocation and risk reduction to facilitate economic activity. While still retaining its application to practical problems in finance, mathematical finance has grown out to be quite mathematically sophisticated, driven by the urge to model the financial markets more and more realistically.

One of the important topics in mathematical finance, is the valuation of so-called options. They are financial instruments which are based on the value of an underlying asset, such as a stock. An option contract offers the holder the opportunity to buy or sell, depending on the type of contract, the underlying asset. It should be stressed that the holder is under no obligation to exercise the contract, while the writer or seller of the contract has to oblige with the choice of the holder. The most common example of an option is the European call option. Given a maturity time $T$ and a strike price $K$, the holder of the option has the option to buy the underlying asset at time $T$ for the price $K$. Since the holder always has the most favourable outcome in terms of payoff, the writer asks a certain price for buying the option. The question which has been studied extensively, is how to compute a fair price for these kinds of options as the value of the underlying asset is of a stochastic nature and can sometimes vary wildly in between the purchase moment of the option and the maturity time. Nowadays there exists a big variety of options which have become increasingly difficult to price, necessitating a detailed study.

Traditionally, options have been priced under the Black-Scholes model, which won the 1997 Nobel Prize of Economics. Under this model, options like European call options can be computed analytically, and the model gives a very satisfying mathematical theory. However, the assumptions have been widely criticised as they tend to heavily underestimate the probability of extreme events on the market and insufficiently account for correlations on past events. The Black-Scholes model assumes among other things that the asset price moves continuously through time. One of the solutions which have been proposed, is adding jumps to the asset price dynamics. These kinds of models are mostly split into two classes, the jump-diffusion models and the infinite activity models. The jump-diffusion models assume that normally the price moves relatively tamely through time, but at rare occasions extreme events can happen in the form of jumps. The infinite activity models are different in that both the normal events as the extreme events are modelled with jumps. Consequently, infinite activity models pose additional mathematical challenges. In this thesis we will only develop numerical algorithms for jump-diffusion processes, infinite activity models still pose difficulties which are yet somewhat unresolved for the problems we will research.

### Connection PIDEs and FBSDEJs

Commonly in engineering problems, but also economic and financial problems, we are faced with solving a so-called Partial-Integro Differential Equation, PIDE in short, of the following form

$$\begin{aligned} \mathcal{L}v(t,x) &= -f\left(t,x,v,\sigma\frac{\partial v}{\partial x},\mathcal{M}u\right), && \forall (t,x)\in[0,T)\times\mathbb{R},\\ v(T,x) &= g(x) && \forall x\in\mathbb{R}. \end{aligned}$$

Here $\mathcal{L}$ is the second-order partial-integro differential operator defined as

$$\begin{aligned} \mathcal{L}v(t,x) = {}& \frac{\partial v}{\partial t}(t,x) + \mu(t,x)\frac{\partial v}{\partial x}(t,x) + \frac{1}{2}\sigma^2(t,x)\frac{\partial^2 v}{\partial x^2}(t,x)\\ &+ \int_{\mathbb{R}_0}\left(v(t,x+\beta(t,x,J)) - v(t,x) - \frac{\partial v}{\partial x}(t,x)\beta(t,x,J)\right)d\nu(J), \end{aligned}$$

and $\mathcal{M}$ is the integral operator defined as

$$\mathcal{M}v(t,x) = \int_{\mathbb{R}_0}(v(t,x+\beta(t,x,J)) - v(t,x))\eta(J)\,d\nu(J).$$

The functions $\mu, \sigma, \beta, \rho$ and $f$ and the $\sigma$-finite measure $\nu$ are all assumed to be known. The task is to solve the equation for $v$. Rarely we can find a solution analytically in a closed form expression. Hence we have to rely on numerical algorithms to approximate the solution. Even then, such a task is non-trivial.

Under certain regularity conditions on the known coefficients, as we will discuss later in the thesis, a probabilistic representation exists of this PIDE. The following set of equations is called a Forward Backward Stochastic Differential Equation with Jumps, FBSDEJ, in short. Formally it is given by

$$\begin{cases} X_t = x + \int_0^t \mu_s(X_s)\,ds + \int_0^t \sigma_s(X_s)\,dW_s + \int_0^t \int_{\mathbb{R}_0} \beta_s(X_{s-}, J)\,\widetilde{N}(ds, dJ), & \text{(FSDEJ)} \\ Y_t = g(X_T) + \int_t^T f_s(X_s, Y_s, Z_s, \Gamma_s)\,ds - \int_t^T Z_s\,dW_s - \int_t^T \int_{\mathbb{R}_0} U_s(J)\,\widetilde{N}(ds, dJ), & \text{(BSDEJ)} \end{cases}$$

which has the solution $(X, Y, Z, \Gamma)$ and $\Gamma_t = \int_{\mathbb{R}_0} U_t(J)\eta(J)\,d\nu(J)$. We furthermore assume that the system is Markovian, that is to say that all the randomness in $(Y_t, Z_t, \Gamma_t)$ is due to $X_t$. The relation with the PIDE is that $v(t, x) = Y_t(x)$ for every $x \in \mathbb{R}$, where we write the dependency of $Y_t$ on $x$ as the effect of changing $x$ in the FSDEJ. Furthermore we have $\sigma \frac{\partial v}{\partial x}(t, x) = Z_t(x)$ and $\mathcal{M}v(t, x) = \Gamma_t(x)$. The exact meaning of the integrals will have to wait for later chapters in the thesis, we can, however, give a heuristic explanation in terms of option valuation.

The FSDEJ is also sometimes called the state process, as it determines the entire randomness of the system due to the Markovian structure. For option pricing we can see $X$ as the log-asset price, where we call $\mu$ the drift term, $\sigma$ the diffusion term and $\beta$ the jump term. The drift term models the price changes, the diffusion term models the volatility of the asset price, how wildly the price changes, and finally the jump term models the size of the jump. There is a fourth parameter hidden inside the third integral which is, what we call, the Lévy measure $\nu$. The total mass $\lambda = \nu(\mathbb{R}_0)$ models the intensity of the jumps and $\lambda^{-1}\nu$ models the probability distribution of the jumps sizes. The BSDEJ is a bit more difficult to understand, the $Y$-process is the option price and the $Z$-process is equal to the Delta multiplied with the volatility and the underlying asset price. The Delta is the derivative of the option price with respect to the underlying asset price. Now the $U$-process is equal to the jumps in the option price, while the $\Gamma$-process is the expectation over all the possible jumps in the option price multiplied with $\eta$ and multiplied with the intensity $\lambda$. Alternatively we could see the $Z$ and $U$ processes as control processes. They represent the randomness which needs to be built into the $Y$-process such that it is steered towards the terminal condition $g(X_T)$ ensuring that we only use current knowledge of the underlying asset. The driver $f$ on itself is insufficient for these purposes, unless the diffusion term and the jump term are zero in the FSDEJ.

A major difficulty in solving Forward Backward Stochastic Differential Equations (FBSDEs) is that the process moves backwards in time, while our knowledge of the value of the state process grows incrementally forward in time. In classical PDE theory, we could do a change of variables in the time variable, to 'reverse' time. This is however not possible in this probabilistic framework, as we have a restriction on the available information of the underlying state process, which we call adaptedness to the filtration. Therefore analysing and solving these kinds of equations is a theory of its own, and turns out to be rather difficult.

## Outline of the thesis

The thesis is divided into two parts. Part I deals with the tools in stochastic analysis which we will need to properly define the FBSDEJs and prove their existence and uniqueness, so that in Part II we can give a detailed treatment on the numerical aspect of solving FBSDEJs.

We begin the thesis with the basic theory of stochastic processes which will be the backbone of the thesis in Chapter 2. A very special kind of stochastic processes, will be the class of martingales, which roughly means that they stay constant on average through time. They will be a suitable class of integrators on which we can define stochastic integrals. The important properties of martingales which we will use in this thesis are discussed in Chapter 3.

In Chapter 4 we will go a long way in setting up the stochastic calculus we will need to define and analyse our FBSDEJs. Integration with respect to continuous semimartingales has become standard in the field of stochastic calculus, and so is the theory of càdlàg semimartingales. We will, however, not need the full generality càdlàg semimartingales provides, as is treated in for example Protter [33], but we will need something more than continuous semimartingales as we want to introduce jump processes. The theory of what is called Lévy-Itô processes in Tankov [42] is not always documented in great detail. So we will fill in some gaps and combine the existing literature to get a satisfactory theory for our needs.

Finally in Chapter 5 we will define the stochastic differential equations. The structure of this chapter will be simple, we will analyse the existence and uniqueness of the solutions of FSDEJs and BSDEJs seperately.

In the second part we start off with defining and analysing the COS method as defined in Fang [16]. We will prove some new connections with Lévy-Itô processes and we will try to give a bit more detailed reasoning behind the numerical behaviour of the method. In Ruijter [35] the COS formulas necessary for the numerical computation of FBSDEs (so without jumps) were introduced, we will extend them to the case of jump processes.

In Chapter 7 we propose a new numerical method for computing FBSDEJs, the semi-discretisation is based on the Backward Differentiation Formula (BDF) methods for computing ODEs, while the full-discretisation is similar to the BCOS method as defined in Ruijter [35]. In the second part of the chapter we will discuss some convergence results for the new method and difficulties we still face in the BSDE literature to prove optimal convergence rates for the general case of FBSDEJs.

In the last chapter we thoroughly test our method and compare it to the scheme proposed in Ruijter [36]. The method as given by Ruijter only covered the FBSDEs, and only one example of an FBSDEJ. In Chapter 7 we will discuss how to extend this method to general FBSDEJs in an efficient and accurate manner.

Finally we present conclusions of the conducted research on this new numerical method and provide some topics for future research.

# Part I

# Stochastic Analysis

# CHAPTER 2
# Stochastic Processes

Stochastic processes can be thought of as an evolution of stochastic random variables. As an example assume that a stock price on the financial market follows a stochastic model, then we can view it also as a stochastic process, evolving through time. In the last decades, stochastic processes on the financial market have been studied extensively and the required stochastic calculus we will introduce later in this thesis, is fundamental for the modern language of finance. Everything we will discuss in this thesis we do by keeping the application to finance in mind, but we will treat the theory in generality as its applicability reaches further than just the world of finance.

We can either model stochastic processes as discrete-time processes, which can only change in value at a countable set of times, or we can model stochastic processes as continuous-time processes where the stochastic processes can now change in value at an uncountable set of times. For a thorough treatment of discrete-time stochastic processes we refer to Shreve [39], which is accessible for anyone having followed only some basic courses in mathematics. In this thesis we will however focus entirely on continuous-time processes as they form a more realistic framework for modelling in finance.

Contrary to discrete-time processes, continuous-time processes require a solid mathematical background to work with. An introduction is given in Shreve [38]. In this thesis we will however need a more elaborate theory of stochastic calculus, and for this we also need a precise theory of stochastic processes. This chapter is roughly based on the works Billingsley [4], Cohn [10], Le Gall [26] and Shreve [38].

## 2.1 Probability Spaces and Random Variables

A probability space is denoted as the triple $(\Omega, \mathcal{F}, \mathbb{P})$ consisting of the *sample space* $\Omega$ which contains all possible outcomes of a random event, we typically denote an element of $\Omega$ as $\omega$. Furthermore, $\mathcal{F}$ is called the *event space* containing all random events which can occur on this probability space, and finally $\mathbb{P}$ is called the *probability measure* which assigns a probability to each event in the event space.

To get a convenient structure on the probability space, the event space $\mathcal{F}$ and probability measure $\mathbb{P}$ have to satisfy a set of properties. We require the event space $\mathcal{F}$ to be a *$\sigma$-algebra* on $\Omega$, which consists of subsets of $\Omega$ and satisfies the following three properties

- The empty set $\varnothing$ is an element of $\mathcal{F}$;
- For any $A \in \mathcal{F}$, its complement $A^c := \Omega \setminus A$ is also in $\mathcal{F}$;
- For any countable sequence $(A_n)_{n\in\mathbb{N}} \subset \mathcal{F}$, its union $\bigcup_{n=1}^{\infty} A_n \in \mathcal{F}$.

For the probability space $\mathbb{P}$ we have the requirement that it has to be a measure on $\mathcal{F}$ and that $\mathbb{P}(\Omega) = 1$. For $\mathbb{P} : \mathcal{F} \to [0, 1]$ to be a measure, we need that

- $\mathbb{P}(\varnothing) = 0$;
- If a countable sequence $(A_n)_{n\in\mathbb{N}} \subset \mathcal{F}$ is disjoint, then

$$\mathbb{P}\left(\bigcup_{n=1}^{\infty} A_n\right) = \sum_{n=1}^{\infty} \mathbb{P}(A_n).$$

We call an event with probability zero, a *null event*, and in this framework it corresponds to the following. Suppose $A \in \mathcal{F}$ is such that $\mathbb{P}(A) = 0$, then $A$ is a null set. Consequently $\mathbb{P}(A^c) = 1$ and we say that $A^c \in \mathcal{F}$ occurs $\mathbb{P}$-*almost surely* or abbreviated as $\mathbb{P}$-a.s. When $\mathbb{P}$ is clear from the context we just write a.s. or almost surely.

Given a probability space $(\Omega, \mathcal{F}, \mathbb{P})$, write

$$\mathcal{N}(\mathcal{F}, \mathbb{P}) := \{A \subset \Omega : \text{ there exists } A \subset \widetilde{A} \in \mathcal{F} \text{ such that } \mathbb{P}(\widetilde{A}) = 0\}.$$

We say that $(\Omega, \mathcal{F}, \mathbb{P})$ is a *complete* probability space if $\mathcal{N}(\mathcal{F}, \mathbb{P}) \subset \mathcal{F}$. We can complete any probability space by enlarging the event space slightly with respect to the probability measure

$$\mathcal{F}^{\mathbb{P}} := \mathcal{F} \vee \mathcal{N}(\mathcal{F}, \mathbb{P}) := \sigma\left(\mathcal{F} \cup \mathcal{N}(\mathcal{F}, \mathbb{P})\right).$$

Here we denote $\sigma(\mathcal{G})$ as the smallest $\sigma$-algebra containing the family $\mathcal{G}$, or equivalently as the intersection of all the $\sigma$-algebras containing $\mathcal{G}$, as an arbitrary intersection of a family of $\sigma$-algebras is again a $\sigma$-algebra. Unfortunately a union of two $\sigma$-algebras does not necessarily have to be again a $\sigma$-algebra, hence we define $\mathcal{F} \vee \mathcal{G} = \sigma(\mathcal{F} \cup \mathcal{G})$.

The strength of this framework is that we can now use Lebesgue measure theory to do our analysis. Given a probability space $(\Omega, \mathcal{F}, \mathbb{P})$, note that now $(\Omega, \mathcal{F})$ is a measurable space. We call an $\mathcal{F}$-measurable mapping $X : \Omega \to \mathbb{R}$ a *random variable*, in this case we always equip $\mathbb{R}$ with the Euclidean topology and its corresponding Borel $\sigma$-algebra $\mathcal{B}(\mathbb{R})$. We write $L^0(\mathcal{F})$ for the space of all random variables, in the notation we emphasize the dependency on $\mathcal{F}$ for a function $X : \Omega \to \mathbb{R}$ to be a random variable.

Suppose we have a probability space $(\Omega, \mathcal{F}, \mathbb{P})$, given a random variable $X \in L^0(\mathcal{F})$, then we define $\sigma(X) := \{X^{-1}(A) : A \in \mathcal{B}(\mathbb{R})\}$, to be the $\sigma$-algebra generated by $X$. It is the smallest $\sigma$-algebra, which makes $X$ into a random variable on the probability space $(\Omega, \sigma(X), \mathbb{P})$ where $\mathbb{P}$ is now restricted to $\sigma(X)$. Again given a random variable $X \in L^0(\mathcal{F})$ we define the pushforward measure

$$\mu_X(A) := (\mathbb{P} \circ X^{-1})(A) = \mathbb{P}(X \in A), \ A \in \mathcal{B}(\mathbb{R}),$$

which makes $(\mathbb{R}, \mathcal{B}(\mathbb{R}), \mu_X)$ a probability space. We call $\mu_X$ the *distribution* of $X$. When $\int_{\mathbb{R}} |x| \, d\mu(x) < \infty$, we say that $X$ is integrable and we define its *expectation* by

$$\mathbb{E}(X) := \int_{\mathbb{R}} x \, d\mu_X(x) = \int_{\Omega} X(\omega) \, d\mathbb{P}(\omega).$$

Now we can use what we know about Lebesgue integrals to define for every $p \geq 1$ the Banach spaces $(L^p(\mathcal{F}, \mathbb{P}), \|\cdot\|_p)$, where $L^p(\mathcal{F}, \mathbb{P}) \subseteq L^0(\mathcal{F})$ consists of all random variables such that $\|\cdot\|_p := \mathbb{E}(|X|^p)^{1/p} < \infty$. Furthermore we define $L^\infty(\mathcal{F}, \mathbb{P}) \subseteq L^0(\mathcal{F})$ to be the space of all bounded random variables $X$, meaning that there exists an $M > 0$ satisfying $\mathbb{P}(|X| > M) = 0$.

For random variables $X, Y \in L^2(\mathcal{F}, \mathbb{P})$ we also define the variance and covariance by

$$\begin{aligned}
\mathrm{Var}(X) &:= \mathbb{E}\left((X - \mathbb{E}(X))^2\right) = \mathbb{E}(X^2) - \mathbb{E}(X)^2, \\
\mathrm{Cov}(X, Y) &:= \mathbb{E}((X - \mathbb{E}(X))(Y - \mathbb{E}(Y))) = \mathbb{E}(XY) - \mathbb{E}(X)\mathbb{E}(Y).
\end{aligned}$$

We will look at the relation between random variables all the time, so let $d \in \mathbb{N}$, then a *d-dimensional random vector* is a mapping $X = (X_1, \ldots, X_d) : \Omega \to \mathbb{R}^d$ such that $X_i \in L^0(\mathcal{F})$ for all $i = 1, \ldots, d$. We define $L^0(\mathcal{F}; \mathbb{R}^d)$ as the vector space of d-dimensional random vectors. Instead of random vectors we will always call the elements of $L^0(\mathcal{F}, \mathbb{R}^d)$ random variables. We define the *cumulative distribution function (cdf)* $F_X : \mathbb{R}^d \to [0, 1]$ as $F_X(x) := \mathbb{P}(X \leq x)$, where we will always consider the partial ordering $\leq$ on $\mathbb{R}^d$ as the standard component-wise extension from $\mathbb{R}$.

Suppose we have random variables $X_1, \ldots, X_n$, each of arbitrary dimension, then we say they are *independent* if

$$F_{X_1, \ldots, X_n}(x_1, \ldots, x_n) = F_{X_1}(x_1) \cdots F_{X_n}(x_n).$$

Just as with general measure theory, we define the product probability space as follows. Let $(\Omega_i, \mathcal{F}_i, \mathbb{P}_i)$ for $i = 1, \ldots, n$ be probability spaces, then $(\Omega, \mathcal{F}, \mathbb{P})$ is the product probability space where

- $\Omega = \Omega_1 \times \cdots \times \Omega_n := \{(\omega_1, \ldots, \omega_n) : \omega_i \in \Omega_i\}$;
- $\mathcal{F} = \mathcal{F}_1 \times \cdots \times \mathcal{F}_n := \omega\left(\{A_1 \times \cdots \times A_n : A_i \in \mathcal{F}_i\}\right)$;
- $\mathbb{P} = \mathbb{P}_1 \times \cdots \times \mathbb{P}_n$ is the unique probability measure such that for $A_i \in \mathcal{F}_i$ we have

$$\mathbb{P}(A_1 \times \cdots \times A_n) = \prod_{i=1}^{n} \mathbb{P}_i(A_i).$$

We can generalize independence to $\sigma$-algebras. Suppose $(\Omega, \mathcal{F}, \mathbb{P})$ is a probability space, and let $\mathcal{G}_1, \mathcal{G}_2$ be two sub-$\sigma$-algebras of $\mathcal{F}$, then $\mathcal{G}_1$ and $\mathcal{G}_2$ are called independent when for all $A \in \mathcal{G}_1, B \in \mathcal{G}_2$ we have $\mathbb{P}(A \cap B) = \mathbb{P}(A)\mathbb{P}(B)$. Now let $X \in L^0(\mathcal{F})$ and $\mathcal{G}$ a sub-$\sigma$-algebra of $\mathcal{F}$, then $X$ and $\mathcal{G}$ are called independent if $\sigma(X)$ and $\mathcal{G}$ are.

## 2.2 Conditional Expectation

We will need a way to model information, that is, at every time $t$ we want some mathematical structure to tell us whether we can already know if the true $\omega \in \Omega$ is contained in a certain event or not. In the analogue of option valuation, we want to create a coupling between the option price and the price of the underlying asset, based on the current knowledge of the asset price. For this purpose we can use a filtration.

**Definition 2.2.1.** *Let $\mathcal{F}$ be a $\sigma$-algebra and $(\mathcal{T}, \leq)$ a totally ordered index set. Suppose $\mathbb{F} = (\mathcal{F}_t)_{t\in\mathcal{T}}$ is a family of sub-$\sigma$-algebras of $\mathcal{F}$ such that for every $s, t \in \mathcal{T}$ with $s \leq t$ we have $\mathcal{F}_i \subseteq \mathcal{F}_j$. Then we call $\mathbb{F}$ a filtration for $\mathcal{F}$.*

Mostly we will work with only three index sets, being $\mathbb{N}, \mathbb{R}^+$ and $[0, T]$ for some terminal time $T > 0$. We call the quadruple $(\Omega, \mathcal{F}, \mathbb{F}, \mathbb{P})$ a filtered probability space. Denote

$$\mathcal{F}_{t+} := \bigcap_{s>t, s\in\mathcal{T}} \mathcal{F}_s, \ \ \mathbb{F}^+ := (\mathcal{F}_{t+})_{t\in\mathcal{T}}.$$

Then we say that $\mathbb{F}$ is *right-continuous* if $\mathbb{F}^+ = \mathbb{F}$ and we say given a probability measure $\mathbb{P}$ on $\mathcal{F}$ that $\mathbb{F}$ is complete if every $\mathcal{F}_t$ for $t \in \mathcal{T}$ is complete. If $\mathcal{T}$ is bounded from below (which is usually the case for our purposes) by for example $0 \in \mathcal{T}$, then if $\mathcal{F}_0$ is complete, the filtration $\mathbb{F}$ is complete by induction.

For the rest of the thesis we will assume that every filtration is both right-continuous and complete, in the literature these two assumptions are called the *usual hypothesis* on the filtration. In the context of this thesis, the usual hypothesis is only a small assumption. The use of the usual hypothesis will however still be mentioned, wherever it is used.

Suppose we already have some information about the state of a random variable, then we would like to have an expectation given our currently attained information. This can be done with what we call conditional expectations, and they will be the key link between filtrations and stochastic processes.

**Definition 2.2.2.** *Let $(\Omega, \mathcal{F}, \mathbb{P})$ be a probability space, let $\mathcal{G}$ be a sub-$\sigma$-algebra of $\mathcal{F}$ and let $X \in L^1(\mathcal{F}, \mathbb{P})$ or let $X$ be nonnegative $\mathbb{P}$-almost surely. The conditional expectation of $X$ given $\mathcal{G}$, denoted $\mathbb{E}(X|\mathcal{G})$, is a random variable that satisfies*

1. *(Measurability) The random variable $\mathbb{E}(X|\mathcal{G})$ is $\mathcal{G}$-measurable;*
2. *(Partial averaging) For all $A \in \mathcal{G}$ we have*

$$\int_A \mathbb{E}(X|\mathcal{G})(\omega)\, d\mathbb{P}(\omega) = \int_A X(\omega)\, d\mathbb{P}(\omega).$$

*If $\mathcal{G}$ is the $\sigma$-algebra generated by some other random variable $Y$, we generally write $E(X|Y)$ instead of $\mathbb{E}(X|\sigma(Y))$.*

Before we prove some important properties of the conditional expectation, we first have to show that it actually exists and is unique almost surely. Let $(\Omega, \mathcal{F})$ be a measurable space, and let $\mu$ and $\nu$ be positive measures (for example probability measures) on $(\Omega, \mathcal{F})$. Then $\nu$ is *absolutely continuous with respect to* $\mu$ if for each set $A \in \mathcal{F}$ with $\mu(A) = 0$ we also have $\nu(A) = 0$. We write $\nu \ll \mu$ when $\nu$ is aboslutely continuous with respect to $\mu$. Furthermore we call a measure $\mu$ on $(\Omega, \mathcal{F})$ a finite measure if $\mu(\Omega) < \infty$ and $\sigma$-finite if $\Omega$ is the union of a sequence $(A_n)_{n\in\mathbb{N}} \subseteq \mathcal{F}$ satisfying $\mu(A_n) < \infty$ for each $n \in \mathbb{N}$. We call the measure space $(\Omega, \mathcal{F}, \mu)$ finite or $\sigma$-finite when $\mu$ is finite or $\sigma$-finite respectively. An important theorem connecting absolutely continuous measures is the Radon-Nikodym theorem. The proof is given in Cohn [10, Theorem 4.2.2].

**Theorem 2.2.3** (Radon-Nikodym Theorem)**.** *Let $(\Omega, \mathcal{F})$ be a measurable space and let $\mu$ and $\nu$ be $\sigma$-finite positive measures on $(\Omega, \mathcal{F})$. If $\nu$ is absolutely continuous with respect to $\mu$, then there exists an $\mathcal{F}$-measurable function $g : \Omega \to [0, \infty)$ such that for each $A \in \mathcal{F}$ we have*

$$\nu(A) = \int_A g\, d\mu.$$

*The function $g$ is unique up to $\mu$-almost everywhere equality.*

Now that we have stated the Radon-Nikodym theorem, we can prove the existence and uniqueness of the conditional expectation, the proof follows Shreve [38, Theorem B.1].

**Theorem 2.2.4.** *Let $(\Omega, \mathcal{F}, \mathbb{P})$ be a probability space, let $\mathcal{G}$ be a sub-$\sigma$-algebra of $\mathcal{F}$, and let $X \in L^1(\mathcal{F}, \mathbb{P})$ or let $X$ be nonnegative $\mathbb{P}$-almost surely. Then there exists a unique $\mathcal{G}$-measurable random variable $\mathbb{E}(X|\mathcal{G})$ such that for every $A \in \mathcal{G}$ we have the equality*

$$\int_A \mathbb{E}(X|\mathcal{G})\, d\mathbb{P} = \int_A X\, d\mathbb{P}.$$

*Proof.*
We will first assume that $X$ is a nonnegative integrable random variable. In the case that $X \in L^1(\mathcal{F}, \mathbb{P})$ we only have to note that $X = X^+ - X^-$ where both $X^+, X^-$ are nonnegatie integrable random variables. In the case that $X$ is nonnegative $\mathbb{P}$-almost surely, we can take a sequence of nonnegative integrable random variables increasing to $X$, so that the theorem follows by the monotone convergence theorem (we will state the monotone convergence theorem later on).

First we will define the probability measure $\widetilde{\mathbb{P}}$ on $\mathcal{F}$ by

$$\widetilde{\mathbb{P}}(A) = \int_A \frac{X+1}{\mathbb{E}(X+1)}\, d\mathbb{P}.$$

Now note that the integrand $\frac{X+1}{\mathbb{E}(X+1)}$ is strictly positive, so if $\widetilde{\mathbb{P}}(A) = 0$, then $\mathbb{P}(A) = 0$ (and vice versa). Hence $\widetilde{\mathbb{P}} \ll \mathbb{P}$. Now we define two probability measures $\mathbb{Q}$ and $\widetilde{\mathbb{Q}}$ on $\mathcal{G}$ by restricting $\mathbb{P}$ to $\mathcal{G}$ and $\widetilde{\mathbb{P}}$ to $\mathcal{G}$ respectively. Hence we have the probability spaces $(\Omega, \mathcal{G}, \mathbb{Q})$ and $(\Omega, \mathcal{G}, \widetilde{\mathbb{Q}})$. Note that we still have $\widetilde{\mathbb{Q}} \ll \mathbb{Q}$. Theorem 2.2.3 guarantees we have a unique $Z \in L^0(\mathcal{G})$ up to $\mathbb{P}$-almost sure equality, such that for all $A \in \mathcal{G}$,

$$\widetilde{\mathbb{P}}(A) = \int_A Z\, d\mathbb{P}.$$

Rewriting this expression gives us for all $A \in \mathcal{G}$

$$\int_A \frac{X+1}{\mathbb{E}(X+1)}\, d\mathbb{P} = \int_A Z\, d\mathbb{P}$$
$$\int_A X\, d\mathbb{P} = \int_A (\mathbb{E}(X+1)Z - 1)\, d\mathbb{P}.$$

Finally let $\mathbb{E}(X|\mathcal{G}) = (\mathbb{E}(X+1)Z - 1)$ which is $\mathcal{G}$-measurable due to $Z$ and satisfies the partial averaging property. The uniqueness follows by the uniqueness of $Z$. $\square$

The conditional expectation satisfies a number of useful properties, we will sum a couple of them in the following proposition, later on we will prove more properties.

**Proposition 2.2.5.** *Let $(\Omega, \mathcal{F}, \mathbb{P})$ be a probability space and let $\mathcal{G}$ be a sub-$\sigma$-algebra of $\mathcal{F}$.*

*(i) (Linearity of conditional expectations) If $X, Y$ are integrable random variables and $\alpha, \beta \in \mathbb{R}$, then*

$$\mathbb{E}(\alpha X + \beta Y|\mathcal{G}) = \alpha\mathbb{E}(X|\mathcal{G}) + \beta\mathbb{E}(Y|\mathcal{G}).$$

*(ii) (Taking out what is known) If $X, Y$ and $XY$ are integrable random variables and $X$ is $\mathcal{G}$-measurable, then*

$$\mathbb{E}(XY|\mathcal{G}) = X\mathbb{E}(Y|\mathcal{G}).$$

*(iii) (Iterated conditioning) If $\mathcal{H}$ is a sub-$\sigma$-algebra of $\mathcal{G}$ and $X$ is an integrable random variable, then*

$$\mathbb{E}(\mathbb{E}(X|\mathcal{G})|\mathcal{H}) = \mathbb{E}(X|\mathcal{H}).$$

*(iv) (Independence) If $X$ is integrable and independent of $\mathcal{G}$, then*

$$\mathbb{E}(X|\mathcal{G}) = \mathbb{E}(X).$$

*(v) (Conditional Jensen's inequality) If $\varphi : \mathbb{R}^d \to \mathbb{R}$ is a convex function and $\varphi(X) \in L^1(\mathcal{F}, \mathbb{P}; \mathbb{R}^d)$, then*

$$\mathbb{E}(\varphi(X)|\mathcal{G}) \geq \varphi(\mathbb{E}(X|\mathcal{G})).$$

*Proof.*
**(i)** Linearity follows immediately by the partial averaging property. Note that $\mathbb{E}(\alpha X + \beta Y|\mathcal{G}), \mathbb{E}(X|\mathcal{G})$ and $\mathbb{E}(Y|\mathcal{G})$ all satisfy this property, then for all $A \in \mathcal{G}$

$$\begin{aligned}\int_A \alpha\mathbb{E}(X|\mathcal{G}) + \beta\mathbb{E}(Y|\mathcal{G})\, d\mathbb{P} &= \alpha \int_A \mathbb{E}(X|\mathcal{G})\, d\mathbb{P} + \beta \int_A \mathbb{E}(Y|\mathcal{G})\, d\mathbb{P} \\ &= \alpha \int_A X\, d\mathbb{P} + \beta \int_A Y\, d\mathbb{P} \\ &= \int_A \alpha X + \beta Y\, d\mathbb{P}.\end{aligned}$$

So indeed $\mathbb{E}(\alpha X + \beta Y|\mathcal{G}) = \alpha\mathbb{E}(X|\mathcal{G}) + \beta\mathbb{E}(Y|\mathcal{G})$.
**(ii)** It suffices to prove this for $X$ being a $\mathcal{G}$-measurable indicator random variable, as the general case follows by using the standard machinery (building simple functions and then taking the limit to general integrable random variables). Let $B \in \mathcal{G}$ and let $X = \mathbf{1}_B$, now note that $\mathbb{E}(Y|\mathcal{G})$ satisfies the partial averaging property itself, so for every $A \in \mathcal{G}$

$$\int_A X\mathbb{E}(Y|\mathcal{G})\, d\mathbb{P} = \int_{A\cap B} \mathbb{E}(Y|\mathcal{G})\, d\mathbb{P} = \int_{A\cap B} Y\, d\mathbb{P} = \int_A XY\, d\mathbb{P}$$

**(iii)** By definition $\mathbb{E}(X|\mathcal{G})$ is $\mathcal{F}$-measurable. Then for all $A \in \mathcal{H}$

$$\int_A \mathbb{E}(\mathbb{E}(X|\mathcal{G})|\mathcal{H})\, d\mathbb{P} = \int_A \mathbb{E}(X|\mathcal{G})\, d\mathbb{P}.$$

Now since $\mathcal{H} \subseteq \mathcal{G}$ we have for all $A \in \mathcal{H}$

$$\int_A \mathbb{E}(X|\mathcal{H})\, d\mathbb{P} = \int_A X\, d\mathbb{P} = \int_A \mathbb{E}(X|\mathcal{G})\, d\mathbb{P} = \int_A \mathbb{E}(\mathbb{E}(X|\mathcal{G})|\mathcal{H})\, d\mathbb{P}.$$

**(iv)** Again we only consider the case where $X$ is an indicator random variable $\mathbf{1}_B$ where $B$ is independent of $\mathcal{G}$, the general case follows again by standard machinery. Indeed, for all $A \in \mathcal{G}$,

$$\int_A X\, d\mathbb{P} = \mathbb{P}(A \cap B) = \mathbb{P}(A)\mathbb{P}(B) = \mathbb{P}(A)\mathbb{E}(X) = \int_A \mathbb{E}(X)\, d\mathbb{P}.$$

**(v)** This last property is less straightforward. We will first prove that $\varphi$ is the maximum of all linear functions that lie below it. So define $\mathcal{L} = \max\{\ell : \mathbb{R}^d \to \mathbb{R} : \ell \text{ is linear and } \ell \leq \varphi\}$, then for all $x \in \mathbb{R}^d$, we have $\varphi(x) \geq \max_{\ell\in\mathcal{L}} \ell(x)$. Conversely, because $\varphi$ is convex, there is always a linear function $\ell$ that lies below $\varphi$ and such that $\varphi(x) = \ell(x)$, hence $\varphi(x) = \max_{\ell\in\mathcal{L}} \ell(x)$. Then by property $(i)$ we have

$$\mathbb{E}(\varphi(X)|\mathcal{G}) \geq \max_{\ell\in\mathcal{L}} \mathbb{E}(\ell(X)|\mathcal{G}) = \max_{\ell\in\mathcal{L}} \ell\left(\mathbb{E}(X|\mathcal{G})\right) = \varphi(\mathbb{E}(X|\mathcal{G})).$$

□

## 2.3 Stochastic Processes

Having discussed the basic probabilistic framework, we can define stochastic processes.

**Definition 2.3.1.** *Let $(\Omega, \mathcal{F}, \mathbb{P})$ be a probability space, and let $(\mathcal{T}, \leq)$ be a totally ordered index set. A stochastic process is a collection of random variables $X = (X_t)_{t\in\mathcal{T}}$ where $X_t \in L^0(\mathcal{F})$ for each $t \in \mathcal{T}$.*

This is just one way to look at a stochastic process. We can also see $X$ as a function $\mathcal{T} \times \Omega \to \mathbb{R} : (t, \omega) \mapsto X_t(\omega)$, or we can see $X$ as a family of *paths* $\{t \mapsto X_t(\omega) : \omega \in \Omega)\}$. We will use these definitions interchangeably.

We will combine our model of information and stochastic processes as follows. We define an *adapted process* to be a stochastic process $X$ on a filtered probability space $(\Omega, \mathcal{F}, \mathbb{F}, \mathbb{P})$ such that for every $t \in \mathcal{T}$, the random variable $X_t$ is $\mathcal{F}_t$-measurable. Furthermore we say $X$ is $\mathbb{F}$-*progressively measurable*, if for every $t \in \mathcal{T}$, the map $\{s \in \mathcal{T} : s \leq t\} \times \Omega \to \mathbb{R}$ defined by $(s, \omega) \mapsto X_s(\omega)$ is $\mathcal{B}(\{s \in \mathcal{T} : s \leq t\}) \otimes \mathcal{F}_t$-measurable. We always assume that $\mathcal{T} \subseteq \mathbb{R}$, so that we can equip $\{s \in \mathcal{T} : s \leq t\}$ with the Euclidean topology.

Often we will construct stochastic processes with a certain joint distribution. However, since the index set $\mathcal{T}$ may be uncountable, we can run into measure-theoretic problems, it will often suffice to look at the joint distribution of all finite combinations of $X_t$. Let $n \in \mathbb{N}$ and let $t_1 \leq t_2 \leq \cdots \leq t_n$ be any partition of $\mathcal{T}$, let $\mu^X_{t_1,\ldots,t_n}$ be the joint distribution of $(X_{t_1}, \ldots, X_{t_n})$. We call the family of all these distributions $\mu^X_{t_1,\ldots,t_n}$ the *finite distribution* of $X$. We say that $X$ and $Y$ have the same distribution if they have the same finite distribution.

We can observe that the finite distribution of $X$ satisfies two important properties. Again take $n \in \mathbb{N}$, and let $t_1 \leq \cdots \leq t_n$ be a partition of $\mathcal{T}$, pick any $j \leq n$ and take $A_k \in \mathcal{B}(\mathbb{R})$ for $k = 1, \ldots, n$, furthermore let $\sigma$ be a permutation on $\{1, \ldots, n\}$, then

$$\begin{aligned} &\mu^X_{t_1,\ldots,t_n}(A_1 \times \cdots \times A_{j-1} \times \mathbb{R} \times A_{j+1} \times \cdots \times A_n) \\ &\qquad = \mu^X_{t_1,\ldots,t_{j-1},t_{j+1},\ldots,t_n}(A_1 \times \cdots \times A_{j-1} \times A_{j+1} \times \cdots \times A_n). \end{aligned} \tag{K1}$$

$$\mu^X_{t_{\sigma(1)},\ldots,t_{\sigma(n)}}(A_{\sigma(1)} \times \cdots \times A_{\sigma(n)}) = \mu^X_{t_1,\ldots,t_n}(A_1 \times \cdots \times A_k) \tag{K2}$$

We call these properties the Kolmogorov consistency criteria. It turns out that given a finite distribution adhering to the Kolmogorov consistency criteria, we can construct a stochastic process with the same finite distribution. We omit the proof of the following theorem, a proof can be found in for example Billingsley [4, Theorem 36.2].

**Theorem 2.3.2** (Kolmogorov's extension theorem). *For all $t_1 \leq \cdots \leq t_n \in \mathcal{T}, n \in \mathbb{N}$ let $\mu_{t_1,\ldots,t_n}$ be probability measures on $\mathbb{R}^n$ such that both* (K1) *and* (K2) *hold.*

*Then there exists a probability space $(\Omega, \mathcal{F}, \mathbb{P})$ and a stochastic process $X$ on this probability space such that $\mu^X_{t_1,\ldots,t_n} = \mu_{t_1,\ldots,t_n}$ for all $t_1 \leq \cdots \leq t_n \in \mathcal{T}, n \in \mathbb{N}$.*

We do lose some information about a stochastic process if we only look at its finite distribution, but many properties still carry over. Let $X$ and $Y$ be two stochastic processes with the same index set and on the same probability space $(\Omega, \mathcal{F}, \mathbb{P})$, then we call $X$ a *modification* of $Y$ if for all $t \in \mathcal{T}$ we have $\mathbb{P}(X_t = Y_t) = 1$. Furthermore we call $X$ and $Y$ *indistinguishable* if $\mathbb{P}(X_t = Y_t, \forall t \in \mathcal{T}) = 1$. There is a subtle measure theoretic difference between modifications and indistinguishability as $\mathcal{T}$ might be uncountable, which we will discuss later on.

Note that when $X$ is a modification of $Y$, they have the same finite distribution. When we construct a stochastic process through its finite distribution, we generally want to take a modification such that it has nice properties. We call a stochastic process $X$ *continuous* if for almost surely $\omega \in \Omega$, the paths $t \mapsto X_t(\omega)$ are continuous, we call a stochastic process $X$ *càdlàg* if for almost surely $\omega \in \Omega$, the paths $t \mapsto X_t(\omega)$ are càdlàg (right-continuous, limits exist from the left). Another theorem due to Kolmogorov gives a sufficient condition for a stochastic process to have a continuous modification. We will again omit a proof and refer to Le Gall [26, Theorem 2.9].

**Theorem 2.3.3** (Kolmogorov's continuity theorem). *Let $X = (X_t)_{t \in \mathcal{T}}$ be a stochastic process on a probability space $(\Omega, \mathcal{F}, \mathbb{P})$. Suppose that for every $T > 0$ there exists $\alpha, \beta, C > 0$ such that for all $s, t \in \mathcal{T} \cap [-T, T]$ we have*

$$\mathbb{E}(|X_s - X_t|^\alpha) \leq C|s - t|^{1+\beta}.$$

*Then there exists a continuous modification of $X$.*

In the following example we will construct a very important stochastic process by using the previous two theorems.

*Example* 2.3.4. Let $0 = t_0 < t_1 < t_2 < \cdots < t_n$ and set $x_0 = 0$. Then define the joint distribution

$$\mu_{t_1,\ldots,t_n}(A_1 \times \cdots \times A_n) = \int_{A_1} \cdots \int_{A_n} \prod_{i=1}^n \frac{1}{\sqrt{2\pi(t_i - t_{i-1})}} \exp\left(-\frac{(x_i - x_{i-1})^2}{2(t_i - t_{i-1})}\right) dx_1 \cdots dx_n.$$

Now K1 follows immediately as the joint distribution is the product of the marginal distributions and K2 follows by using Fubini-Tonelli to interchange the integrals, and changing the order of the product. Hence we can define a stochastic process $(\widetilde{W}_t)_{t \geq 0}$ on a probability space $(\Omega, \mathcal{F}, \mathbb{P})$ with the following properties.

- $\widetilde{W}_0 = 0$ almost surely;
- $\widetilde{W}_t - \widetilde{W}_s$ is independent of $\widetilde{W}_u$ for every $u \leq s$ and $s < t$;
- $\widetilde{W}_t - \widetilde{W}_s \sim \mathcal{N}(0, t-s)$ for $s < t$.

Such a process is called a *pre-Brownian motion*. Now let $s < t$, then $\widetilde{W}_t - \widetilde{W}_s$ has the same distribution as $\sqrt{t-s}\, Z$ where $Z \sim \mathcal{N}(0,1)$. It follows that

$$\mathbb{E}(|\widetilde{W}_t - \widetilde{W}_s|^4) = \mathbb{E}(|Z|^4)|t-s|^2 < \infty.$$

So then by applying Theorem 2.3.3 we can construct a continuous stochastic process $W$ such that $W$ is a modification of $\widetilde{W}$, hence has the same finite distribution. Such a process $W$ is called a *Brownian motion*, and is fundamental to many stochastic models. Remark that every Brownian motion is also a pre-Brownian motion.

Only in the case that $\mathcal{T} = \mathbb{R}^+$ or $\mathcal{T} = [0,T]$ for some $T > 0$, will we concern ourselves with the subtle differences of adaptedness, progressive-measurability, modifications and indistinguishability.

In the case that $\mathcal{T} = \mathbb{R}^+$ (the case $\mathcal{T} = [0,T]$ follows analogously) define $\mathscr{P}$ as the collection of subsets $A$ of $\Omega \times \mathbb{R}^+$ such that for every $t \geq 0$, $A \cap (\Omega \times [0,t])$ belongs to $\mathcal{F}_t \otimes \mathcal{B}([0,t])$. The family $\mathscr{P}$ is a $\sigma$-algebra, and we call it the *progressive-$\sigma$-algebra* of $(\Omega, \mathcal{F}, \mathbb{F}, \mathbb{P})$. It is not hard to see that a stochastic process $X$ is $\mathbb{F}$-progressive measurable if and only if $X$ is $\mathscr{P}$-measurable. In similar fashion as with random variables we define $L^0(\mathscr{P})$ as the space of $\mathbb{F}$-progressively measurable stochastic processes.

The following proposition relates adaptedness and progressive-measurability and it relates modifications and indistinguishability.

**Proposition 2.3.5.** *Let $(\Omega, \mathcal{F}, \mathbb{F}, \mathbb{P})$ be a probability space and let $X$ and $Y$ be càdlàg processes on this filtered probability space.*

1. *If $X$ is a modification of $Y$, then $X$ and $Y$ are indistinguishable.*
2. *If the filtration $\mathbb{F}$ is complete with index set $\mathcal{T} = \mathbb{R}^+$, then $X$ is $\mathbb{F}$-adapted if and only if $X \in L^0(\mathscr{P})$.*

*Proof.*
**(i)** Since $X$ is a modification of $Y$ we have that $\mathbb{P}(X_t \neq Y_t) = 0$ for all $t \in \mathcal{T}$. Now define the sets

$$A' = \{\omega : X_t(\omega) = Y_t(\omega), \forall t \in \mathcal{T} \cap \mathbb{Q}\} = \left( \bigcup_{t \in \mathcal{T} \cap \mathbb{Q}} \{\omega : X_t(\omega) \neq Y_t(\omega)\} \right)^c,$$

and

$$B = \{\omega : t \mapsto X_t(\omega) \text{ is càdlàg}\}, \qquad C = \{\omega : t \mapsto Y_t(\omega) \text{ is càdlàg}\}.$$

By assumption we have $\mathbb{P}(A') = \mathbb{P}(B) = \mathbb{P}(C) = 1$, so then define $A := A' \cap B \cap C$, which satisfies $\mathbb{P}(A) = 1$.

The set $\mathcal{T} \cap \mathbb{Q}$ is dense in $\mathcal{T}$, so for every $t \in \mathcal{T}$, there exists a sequence of rationals $(q_n)_{n \in \mathbb{N}} \subseteq \mathbb{Q}$ such that $q_n \downarrow t$. Then by the construction of the set $A$ we have for all $\omega \in A$

$$\lim_{n \to \infty} X_{q_n}(\omega) = X_t(\omega) \text{ and } \lim_{n \to \infty} Y_{q_n}(\omega) = Y_t(\omega).$$

Since $X_{q_n}(\omega) = Y_{q_n}(\omega)$, we must have $X_t(\omega) = Y_t(\omega)$. Hence $X$ and $Y$ are indistinguishable.
**(ii)** Let $\mathcal{N}$ be the null set where the paths of $X$ are not càdlàg, then set $\overline{X}_t(\omega) = 0$ for all $\omega \in \mathcal{N}$ and $t \geq 0$. Then $\overline{X}$ is still $\mathbb{F}$-adapted since $\mathcal{N} \in \mathcal{F}_0$, by the completeness of the filtration. So without loss of generality we can assume that the paths of $X$ are càdlàg on all of $\Omega$.

Fix $T > 0$, then define for every $n \in \mathbb{N}$ the stochastic processes

$$X_t^n(\omega) = \sum_{k=0}^{\lfloor nT \rfloor - 1} X_{\frac{k+1}{n}}(\omega) \mathbf{1}_{\left[\frac{k}{n}, \frac{k+1}{n}\right)}(t) + X_T(\omega) \mathbf{1}_{[T,\infty)}(t).$$

The right-continuity of the paths of $X$ guarantee that for every $t \leq T$ and all $\omega \in \Omega$ we have $X_t(\omega) = \lim_{n\to\infty} X_t^n(\omega)$. On the other hand, let $A \in \mathcal{B}(\mathbb{R})$, then

$$\{(\omega, t) \in \Omega \times [0,T] : X_t^n(\omega) \in A\} = (\{X_T \in A\} \times \{T\}) \cup \left( \bigcup_{k=0}^{[nT]-1} \left( \{X_{\frac{k+1}{n}} \in A\} \times \left[\frac{k}{n}, \frac{k+1}{n}\right) \right) \right),$$

which is an element of $\mathcal{F}_T \otimes \mathcal{B}([0,T])$. Hence for every $n \in \mathbb{N}$, the mapping $(\omega, t) \mapsto X_t^n(\omega)$ defined on $\Omega \times [0,t]$, is $\mathcal{F}_t \otimes \mathcal{B}([0,T])$-measurable. Since a pointwise limit of measurable functions is measurable, the same measurability holds for $(\omega, t) \mapsto X_t(\omega)$ defined on $\Omega \times [0,T]$. Then $X$ is $\mathbb{F}$-progressively measurable. The converse is obvious. □

## 2.4 Modes of Convergence

Often in probability theory when we have a sequence $X_1, X_2, \dots$ of random variables, we ask ourselves what happens to their sum if the number of summands increases, or what happens to the maximum of those random variables when $n \to \infty$. Can we interchange limits and integrals? What can we say about the limit of a sequence of random variables? All of those questions have to do with certain notions of convergence.

In this thesis we will often define constructions on a certain class of functions and then we generalise the them to a bigger class of functions through limits. Therefore it is important to know when these limits are well-defined and what they imply.

In the next definition we will define a couple of different notions, later on we will see more modes of convergence.

**Definition 2.4.1.** *Let $X_1, X_2, \dots$ be a sequence of random variables on the same probability space $(\Omega, \mathcal{F}, \mathbb{P})$.*

*(i) The sequence $X_n$ converges almost surely to $X$ if*

$$\mathbb{P}\left(\left\{\omega : X_n(\omega) \underset{n\to\infty}{\longrightarrow} X(\omega)\right\}\right) = 1.$$

*We will write $X_n \xrightarrow{a.s.} X$ as $n \to \infty$.*

*(ii) The sequence $X_n$ converges in probability to $X$ if for every $\varepsilon > 0$*

$$\lim_{n\to\infty} \mathbb{P}(\{\omega : |X_n(\omega) - X(\omega)| > \varepsilon\}) = 0.$$

*We will write $X_n \xrightarrow{\mathbb{P}} X$ as $n \to \infty$.*

*(iii) Let $p \in [1, \infty)$. The sequence $X_n$ converges in $L^p$ to $X$ if*

$$\lim_{n\to\infty} \mathbb{E}(|X_n - X|^p) = 0.$$

*We will write $X_n \xrightarrow{L^p} X$ as $n \to \infty$.*

*(iv) The sequence $X_n$ converges to $X$ in distribution if $F_{X_n}(x) \to F_X(x)$ for all $x$ such that $F_X$ is continuous. We will write $X_n \xrightarrow{d} X$ as $n \to \infty$. (Remark that the $X_n$ do not have to be defined on the same probability space, as we are only concerned about the distribution)*

*Remark* 2.4.2. Convergence in distribution is often called weak convergence in the literature as it corresponds to convergence of the distributions in the $\sigma(\mathcal{M}(\mathbb{R}), C_b(\mathbb{R}))$-weak topology, where $\mathcal{M}(\mathbb{R})$ is the space of finite signed measures on $\mathbb{R}$ equipped with the total variation norm $\|\mu\|_{TV} = |\mu|(\mathbb{R})$ and $C_b(\mathbb{R})$ the space of bounded continuous functions $\mathbb{R} \to \mathbb{R}$ equipped with the supremum norm. These details are unimportant for our purposes.

We will see that almost sure convergence implies convergence in probability which again implies convergence in distribution. There are also other connections between convergence we will use. Before we will discuss their implications, we will first state the following very important integral inequalities, which we will use often.

**Theorem 2.4.3.** *Let $(\Omega, \mathcal{F}, \mathbb{P})$ be a probability space, and $X$ and $Y$ random variables on this space.*

*(i) (Markov's inequality) Suppose $\varphi$ is non-negative, non-decreasing, such that $\mathbb{E}(\varphi(X)) < \infty, x > 0$, then*

$$\mathbb{P}(|X| > x) \leq \frac{\mathbb{E}(\varphi(|X|))}{\varphi(x)}.$$

*(ii) (Hölder's inequality) Let $p, q \in [1,\infty)$ such that $\frac{1}{p}+\frac{1}{q}=1$, and suppose $X \in L^p(\mathcal{F},\mathbb{P})$ and $Y \in L^q(\mathcal{F},\mathbb{P})$. Then*

$$\mathbb{E}(|XY|) \leq \|X\|_p \|Y\|_q .$$

*(iii) (Jensen's inequality) Let $\varphi : \mathbb{R}^d \to \mathbb{R}$ be a convex function, such that $X$ and $\varphi(X)$ are integrable, then*

$$\varphi(\mathbb{E}(X)) \leq \mathbb{E}(\varphi(X)).$$

*Proof.*
**(ii)** is a standard result in analysis, a proof can be found in Cohn [10, Proposition 3.3.2]. Also **(iii)** follows from Proposition 2.2.5 by taking the trivial $\sigma$-algebra $\mathcal{G}=\{\varnothing,\Omega\}$, in that case $\mathbb{E}(X|\mathcal{G})=\mathbb{E}(X)$. Now to prove **(i)**. Let $\varphi$ be a non-decreasing and non-negative function, then

$$\begin{aligned}\mathbb{E}(\varphi(|X|)) &= \mathbb{E}(\varphi(|X|)\mathbf{1}_{|X|>x}) + \mathbb{E}(\varphi(|X|)\mathbf{1}_{|X|\leq x})\\ &\geq \mathbb{E}(\varphi(|X|)\mathbf{1}_{|X|>x}) \geq \mathbb{E}(\varphi(x)\mathbf{1}_{|X|>x}) = \varphi(x)\mathbb{P}(|X|>x).\end{aligned}$$

□

When we talk about limits, we often need to have a notion of limits of sets. We define

$$\liminf_{n\to\infty} A_n = \bigcup_{n=1}^{\infty}\bigcap_{m=n}^{\infty} A_m, \qquad \limsup_{n\to\infty} A_n = \bigcap_{n=1}^{\infty}\bigcup_{m=n}^{\infty} A_m.$$

If the lim sup and lim inf agree we define $\lim_{n\to\infty} A_n = \liminf_{n\to\infty} A_n = \limsup_{n\to\infty} A_n$. A powerful tool in determining the probability of limiting events are the Borel-Cantelli lemmas.

**Theorem 2.4.4** (Borel-Cantelli lemmas)**.** *Let $A_1, A_2, \ldots \in \mathcal{F}$ for some $\sigma$-algebra $\mathcal{F}$.*

1. *If $\sum_{n=1}^{\infty}\mathbb{P}(A_n) < \infty$ then $\mathbb{P}(\limsup_{n\to\infty} A_n) = 0$*
2. *If $\sum_{n=1}^{\infty}\mathbb{P}(A_n) = \infty$ and $(A_n)_{n\in\mathbb{N}}$ is a sequence of independent events, then $\mathbb{P}(\limsup_{n\to\infty} A_n) = 1$.*

*Proof.*
For **(i)** note that by assumption $\sum_{n=1}^{\infty}\mathbb{P}(A_n) < \infty$, hence $\lim_{m\to\infty}\sum_{n=m}^{\infty}\mathbb{P}(A_n) = 0$. Let $m \in \mathbb{N}$, then by $\sigma$-subadditivity of the probability measure

$$\mathbb{P}(\limsup_{n\to\infty} A_n) \leq \mathbb{P}\left(\bigcup_{n=m}^{\infty} A_n\right) \leq \sum_{n=m}^{\infty}\mathbb{P}(A_n) \underset{m\to\infty}{\longrightarrow} 0.$$

To prove **(ii)** first write $(\limsup_{n\to\infty} A_n)^c = \bigcup_{n=1}^{\infty}\bigcap_{n=m}^{\infty} A_m^c$. It suffices to prove that for each $m$, we have $\mathbb{P}(\bigcap_{n=m}^{\infty} A_n^c) = 0$. Now since $1-x \leq e^{-x}$ for all $x \in \mathbb{R}$, we have

$$\begin{aligned}\mathbb{P}\left(\bigcap_{n=m}^{\infty} A_n^c\right) &= \lim_{k\to\infty}\mathbb{P}\left(\bigcap_{n=m}^{k} A_n^c\right) = \lim_{k\to\infty}\prod_{n=m}^{k}(1-\mathbb{P}(A_n))\\ &\leq \lim_{k\to\infty}\prod_{n=m}^{k} e^{-\mathbb{P}(A_n)} = \lim_{k\to\infty} e^{-\sum_{n=m}^{k}\mathbb{P}(A_n)} = 0\end{aligned}$$

□

The following theorem connects the four different modes of convergence we have defined for now.

**Theorem 2.4.5.** *Let $(\Omega,\mathcal{F},\mathbb{P})$ be a probability space and $(X_n)_{n\in\mathbb{N}}$ a sequence of random variables. We have the following implications.*

*(i) If $X_n \xrightarrow{a.s.} X$ as $n\to\infty$, then $X_n \xrightarrow{\mathbb{P}} X$.*

*(ii) If $X_n \xrightarrow{L_p} X$ as $n\to\infty$, then $X_n \xrightarrow{\mathbb{P}} X$.*

*(iii) If $X_n \xrightarrow{\mathbb{P}} X$ as $n\to\infty$, then $X_n \xrightarrow{d} X$.*

*(iv) If $X_n \xrightarrow{\mathbb{P}} X$ as $n\to\infty$, then there exists a subsequence $X_{n_k}$ such that $X_{n_k} \xrightarrow{a.s.} X$ as $k\to\infty$.*

*Proof.*
**(i)** Let $\varepsilon > 0$ and note that

$$\{\omega : X_n(\omega) \underset{n\to\infty}{\longrightarrow} X(\omega)\} = \bigcap_{k=1}^{\infty} \bigcup_{m=1}^{\infty} \bigcap_{i=m}^{\infty} \left\{\omega : |X_i(\omega) - X(\omega)| \le \frac{1}{k}\right\}.$$

So for almost sure convergence we need for every $k \in \mathbb{N}$ that

$$0 = \mathbb{P}\left(\left(\bigcup_{m=1}^{\infty} \bigcap_{i=m}^{\infty} \left\{\omega : |X_i(\omega) - X(\omega)| \le \frac{1}{k}\right\}\right)^c\right) = \mathbb{P}\left(\bigcap_{m=1}^{\infty} \bigcup_{i=m}^{\infty} \left\{\omega : |X_i(\omega) - X(\omega)| > \frac{1}{k}\right\}\right)$$
$$= \lim_{m\to\infty} \mathbb{P}\left(\bigcup_{i=m}^{\infty} \left\{\omega : |X_i(\omega) - X(\omega)| > \frac{1}{k}\right\}\right).$$

But then choose $k$ such that $\frac{1}{k} \le \varepsilon$, we have

$$\lim_{m\to\infty} \mathbb{P}(|X_m - X| > \varepsilon) \le \lim_{m\to\infty} \mathbb{P}\left(\bigcup_{i=m}^{\infty} \left\{|X_i - X| > \frac{1}{k}\right\}\right) = 0.$$

**(ii)** This is an immediate consequence of Markov's inequality, let $\varepsilon > 0$, then

$$\mathbb{P}(|X_n - X| > \varepsilon) \le \frac{\mathbb{E}(|X_n - X|^p)}{\varepsilon^p} \underset{n\to\infty}{\longrightarrow} 0.$$

**(iii)** Let $\varepsilon > 0$, then

$$\begin{aligned} F_{X_n}(x) = \mathbb{P}(X_n \le x) &= \mathbb{P}(\{|X_n - X| \le \varepsilon\} \cap \{X_n \le x\}) + \mathbb{P}(\{|X_n - X| > \varepsilon\} \cap \{X_n \le x\}) \\ &\le \mathbb{P}(\{|X_n - X| \le \varepsilon\} \cap \{X \le x + \varepsilon\}) + \mathbb{P}(|X_n - X| > \varepsilon) \\ &\le \mathbb{P}(X \le x + \varepsilon) + \mathbb{P}(|X_n - X| > \varepsilon) \end{aligned}$$

By using the convergence in probability we get

$$\limsup_{n\to\infty} F_{X_n}(x) \le F_X(x + \varepsilon).$$

Switch $X_n$ with $X$, switch $x$ with $x - \varepsilon$ on the left hand side and $X$ with $X_n$ and $x + \varepsilon$ with $x$ on the right hand side, it follows analogously that

$$\liminf_{n\to\infty} F_{X_n}(x) \ge F_X(x - \varepsilon).$$

Now we have

$$F_X(x - \varepsilon) \le \liminf_{n\to\infty} F_{X_n}(x) \le \limsup_{n\to\infty} F_{X_n}(x) \le F_X(x + \varepsilon).$$

If $F_X$ is continuous at $x$, the claim follows if we let $\varepsilon \downarrow 0$.
**(iv)** Choose a subsequence $X_{n_k}$ such that for every $k \in \mathbb{N}$ we have $\mathbb{P}(|X_{n_k} - X| > 2^{-k}) \le 2^{-k}$. Then define the sets $A_k = \{|X_{n_k} - X| > 2^{-k}\}$, we have

$$\sum_{k=1}^{\infty} \mathbb{P}(A_k) \le \sum_{k=1}^{\infty} 2^{-k} = 1 < \infty.$$

The Borel-Cantelli lemmas imply that

$$\mathbb{P}\left(\limsup_{k\to\infty} \{|X_{n_k} - X| > 2^{-k}\}\right) = \mathbb{P}\left(\bigcap_{k=1}^{\infty} \bigcup_{m=k}^{\infty} \{|X_{n_m} - X| > 2^{-m}\}\right) = 0.$$

Hence we get almost sure convergence by the same reasoning as in (i). □

Often we will work with expressions containing Lebesgue integrals, the following theorems are quite useful, the proofs can be found in any analysis book covering Lebesgue integration, for example Cohn [10].

**Theorem 2.4.6.** *Let $(\Omega, \mathcal{F}, \mathbb{P})$ be a probability space and let $X_1, X_2, \ldots$ be random variables.*

*(i) (Monotone convergence theorem) Suppose that $\mathbb{E}(X_1) > -\infty$ and $X_n \uparrow X$ almost surely. Then $\lim_{n\to\infty} \mathbb{E}(X_n) = \mathbb{E}(X)$.*

*(ii) (Fatou's lemma) Suppose $(X_n)_{n\in\mathbb{N}}$ are non-negative, then*

$$\mathbb{E}(\liminf_{n\to\infty} X_n) \leq \liminf_{n\to\infty} \mathbb{E}(X_n).$$

*(iii) (Dominated convergence theorem) Suppose that $|X_n| \leq Y$, for all $n \in \mathbb{N}$ where $\mathbb{E}(|Y|) < \infty$ and $X_n \to X$ almost surely as $n \to \infty$. Then $\mathbb{E}(|X_n - X|) \to 0$ as $n \to \infty$.*

We will see that $L^p$-convergence implies convergence in probability. However, the converse is in general not true, but if we impose a bit of extra structure on the sequence of random variables, we can have $L^p$-convergence. This extra structure is called *uniform integrability* and will also play a role besides this equivalence. A family $\mathcal{X}$ of random variables is called *uniformly integrable* if

$$\lim_{M\to\infty} \sup_{X\in\mathcal{X}} \mathbb{E}(|X|\mathbf{1}_{|X|\geq M}) = 0.$$

**Proposition 2.4.7.** *Let $(\Omega, \mathcal{F}, \mathbb{P})$ be a probability space. Given a sequence of random variables $(X_n)_{n\in\mathbb{N}} \subseteq L^p(\mathcal{F}, \mathbb{P})$ and $X \in L^p(\mathcal{F}, \mathbb{P})$. The following is equivalent:*

*(i) $X_n$ converges in probability to $X$ and $(X_n^p)_{n\in\mathbb{N}}$ is uniformly integrable*

*(ii) $X_n$ converges in $L^p(\mathcal{F}, \mathbb{P})$ to $X$.*

# CHAPTER 3

# Martingales

A very important class of stochastic processes is the class of martingales, which, as we will see, induces a lot of structure on the stochastic processes. It turns out that martingales give rise to a wide class of integrators for stochastic integrals, those stochastic integrals will be the final piece we will need, to understand FBSDEJs. Most of the theorems we will discuss in this chapter will be proven using the following structure. First we prove the theorem for discrete-time martingales, and then we will lift the result using some continuity properties of the stochastic process through density arguments.

We will start with a definition of martingales.

**Definition 3.0.1.** *Let $(\Omega, \mathcal{F}, \mathbb{F}, \mathbb{P})$ be a filtered probability space and let $X = (X_t)_{t\in\mathcal{T}}$ be an adapted stochastic process such that $X_t \in L^1(\mathcal{F}, \mathbb{P})$ for every $t \in \mathcal{T}$. Then*

- *$X$ is called a martingale if $\mathbb{E}(X_t|\mathcal{F}_s) = X_s$ for every $s < t$;*
- *$X$ is called a supermartingale if $\mathbb{E}(X_t|\mathcal{F}_s) \leq X_s$ for every $s < t$;*
- *$X$ is called a submartingale if $\mathbb{E}(X_t|\mathcal{F}_s) \geq X_s$ for every $s < t$.*

*Example* 3.0.2. An important example of a martingale is Brownian motion as introduced in Example 2.3.4. Indeed let $W = (W_t)_{t\geq 0}$ be Brownian motion. Then for $0 \leq s < t$ we have by the properties of conditional expectations, that

$$\mathbb{E}(W_t|\mathcal{F}_s) = \mathbb{E}(W_t - W_s|\mathcal{F}_s) + \mathbb{E}(W_s|\mathcal{F}_s) = \mathbb{E}(W_t - W_s) + W_s = W_s.$$

Given a martingale, we can construct submartingales by convex transformations, due to Jensen's inequality.

**Proposition 3.0.3.** *Let $X = (X_t)_{t\in\mathcal{T}}$ be an adapted process and let $\varphi : \mathbb{R}^d \to \mathbb{R}$ be a convex function such that $\varphi(X_t) \in L^1(\mathcal{F}, \mathbb{P})$ for all $t \in \mathcal{T}$.*

*(i) If $X$ is a martingale, then $\varphi(X) = (\varphi(X_t))_{t\in\mathcal{T}}$ is a submartingale.*

*(ii) If $X$ is a submartingale, and if in addition $\varphi$ is nondecreasing, then $\varphi(X)$ is a submartingale.*

*Proof.*
By Jensen's inequality for conditional expectations, we have for $s < t$ that

$$\mathbb{E}(\varphi(X_t)|\mathcal{F}_s) \geq \varphi(\mathbb{E}(X_t|\mathcal{F}_s)) \geq \varphi(X_s).$$

For the last (in)equality, we need the fact that $\varphi$ is nondecreasing, when $X$ is a submartingale. □

## 3.1 Stopping times

Often we will have to deal with stochastic processes which are not necessarily so nicely behaving. To amend these problems, we will often use a localisation argument. We want to 'stop' the stochastic process before it is going to misbehave, and then we hope to transfer the proof on the stopped process to the general process. To stop processes, we first have to define what it means to be a stopping time.

**Definition 3.1.1.** *Let $(\Omega, \mathcal{F}, \mathbb{F}, \mathbb{P})$ be a filtered probability space, and $\tau$ a random variable with values in $\mathcal{T} \cup \{\infty\}$ (where in the case of index sets bounded from above we mean by $\infty$ the upper bound). Then $\tau$ is a $\mathbb{F}$-stopping time, if for every $t \in \mathcal{T}$, we have $\{\tau \leq t\} \in \mathcal{F}_t$.*

The last measurability condition ensures that stopping times are not allowed to 'look into the future', they can only make decisions based on current knowledge. Given a filtered probability space $(\Omega, \mathcal{F}, \mathbb{F}, \mathbb{P})$, we define

$$\mathcal{F}_\infty := \bigvee_{t \in \mathcal{T}} \mathcal{F}_t,$$

to be the smallest $\sigma$-algebra, containing the union of the entire filtration. Furthermore, we define the $\sigma$-algebra with respect to an $\mathbb{F}$-stopping time $\tau$ as $\mathcal{F}_\tau = \{A \in \mathcal{F}_\infty : \forall t \in \mathcal{T}, A \cap \{\tau \leq t\} \in \mathcal{F}_t\}$. It follows that $\tau$ is $\mathcal{F}_\tau$-measurable. The following proposition contains some basic properties of stopping times.

**Proposition 3.1.2.** *Let $(\Omega, \mathcal{F}, \mathbb{F}, \mathbb{P})$ be a filtered probability space.*

*(i) A random variable $\tau$ is an $\mathbb{F}^+$-stopping time if and only if $\{\tau < t\} \in \mathcal{F}_t$ for every $t \in \mathcal{T}$. Equivalently, $\tau \wedge t$ is $\mathcal{F}_t$-measurable for every $t \in \mathcal{T}$.*

*(ii) Let $\tau$ be an $\mathbb{F}^+$-stopping time. Then*

$$\mathcal{F}_\tau = \{A \in \mathcal{F}_\infty : \forall t \in \mathcal{T}, A \cap \{\tau < t\} \in \mathcal{F}_t\}.$$

*We will write $\mathcal{F}_{\tau+}$ for the above.*

*(iii) Let $\tau, \rho$ be two $\mathbb{F}$-stopping times. The random variables $\tau \vee \rho$ and $\tau \wedge \rho$ are $\mathbb{F}$-stopping times.*

*(iv) If $(\tau_n)_{n \in \mathbb{N}}$ is a monotone increasing sequence of $\mathbb{F}$-stopping times, then $\tau = \lim_{n \to \infty} \tau_n$ is also an $\mathbb{F}$-stopping time.*

*(v) If $(\tau_n)_{n \in \mathbb{N}}$ is a monotone decreasing sequence of $\mathbb{F}$-stopping times, then $\tau = \lim_{n \to \infty} \tau_n$ is an $\mathbb{F}^+$-stopping time.*

*Proof.*
**(i)** Suppose that $\tau$ is an $\mathbb{F}^+$-stopping time. Then, for every $t \in \mathcal{T}$, we have

$$\{\tau < t\} = \bigcup_{q \in \mathbb{Q}^+, q < t} \{\tau \leq q\} \in \mathcal{F}_t.$$

Conversely, assume that $\{\tau < t\} \in \mathcal{F}_t$ for every $t \in \mathcal{T}$. Then for every $t \in \mathcal{T}$ and $s > t$, we have that

$$\{\tau \leq t\} = \bigcap_{q \in \mathbb{Q}^+, t < q < s} \{\tau < q\} \in \mathcal{F}_s.$$

Therefore by definition we have $\{\tau \leq t\} \in \mathcal{F}_{t+}$, so $\tau$ is an $\mathbb{F}^+$-stopping time.

If $\tau \wedge t$ is $\mathcal{F}_t$-measurable for every $t \in \mathcal{T}$, then equivalently for every $t \in \mathcal{T}$ and $s < t$, we have $\{\tau \leq s\} \in \mathcal{F}_t$. Now taking a sequence in $\mathcal{T}$ increasing to $t$, gives that $\{\tau < t\} \in \mathcal{F}_t$, and so $\tau$ is an $\mathbb{F}^+$-stopping time. Conversely, we have $\{\tau \leq s\} \in \mathcal{F}_{t+} \subseteq \mathcal{F}_t$ whenever $s < t$, and so $\tau \wedge t$ is $\mathcal{F}_t$-measurable.
**(ii)** If $A \in \mathcal{F}_\tau$, then for every $t \in \mathcal{T}$ we have $A \cap \{\tau \leq t\} \in \mathcal{F}_{t+}$. Hence, for $t \in \mathcal{T}$,

$$A \cap \{\tau < t\} = \bigcup_{q \in \mathbb{Q}^+, q < t} (A \cap \{\tau \leq q\}) \in \mathcal{F}_t,$$

Conversely, assume that $A \cap \{\tau < t\} \in \mathcal{F}_t$ for every $t \in \mathcal{T}$. Then, for every $t \in \mathcal{T}$ and $s > t$,

$$A \cap \{\tau \leq t\} = \bigcap_{q \in \mathbb{Q}^+, t < q < s} (A \cap \{\tau < q\}) \in \mathcal{F}_s.$$

So then $A \cap \{\tau \leq t\} \in \mathcal{F}_{t+}$ and so $A \in \mathcal{F}_{\tau+}$.
**(iii)** We have for all $t \in \mathcal{T}$, that

$$\{\tau \wedge \rho \leq t\} = \{\tau \leq t\} \cup \{\rho \leq t\} \in \mathcal{F}_t,$$
$$\{\tau \vee \rho \leq t\} = \{\tau \leq t\} \cap \{\rho \leq t\} \in \mathcal{F}_t,$$

hence $\tau \wedge \rho$ and $\tau \vee \rho$ are $\mathbb{F}$-stopping times.
**(iv)** For every $t \in \mathcal{T}$ we have

$$\{\tau \leq t\} = \bigcap_{n=1}^{\infty} \{\tau_n \leq t\} \in \mathcal{F}_t.$$

**(v)** Similarly,

$$\{\tau < t\} = \bigcup_{n=1}^{\infty} \{\tau_n < t\} \in \mathcal{F}_t,$$

and now we use (i) to conclude. □

We can already see some subtleties pop up when the filtration is not right-continuous, this will however not bother us as we wil assume every filtration to be right-continuous.

Often we will use stopping times in the following forms, which we call *hitting times*. The following proposition is much more general than we actually need, as we defined our stochastic processes to take values in $\mathbb{R}$, this can however be generalized to more general spaces.

**Proposition 3.1.3.** *Let $(\Omega, \mathcal{F}, \mathbb{F}, \mathbb{P})$ be a filtered probability space, and let $X$ be an adapted process taking values in a metric space $(S, d)$.*

*(i) Assume that $X$ is càdlàg, and let $G$ be an open subset of $S$. Then*

$$\tau = \inf\{t \geq 0 : X_t \in G\},$$

*is an $\mathbb{F}^+$-stopping time.*

*(ii) Assume that $X$ is continuous, and let $F$ be a closed subset of $S$. Then*

$$\tau = \inf\{t \geq 0 : X_t \in F\},$$

*is an $\mathbb{F}$-stopping time.*

*Proof.*
**(i)** For every $t \in \mathcal{T}$, we have

$$\{\tau < t\} = \bigcup_{s \in \mathcal{T} \cap \mathbb{Q}, s < t} \{X_s \in G\} \in \mathcal{F}_t$$

Then the result follows by Proposition 3.1.2(i).
**(ii)** For every $t \in \mathcal{T}$, by the continuity of $X$ and the metric $d$ we have through density

$$\{\tau \leq t\} = \left\{ \inf_{s \in \mathcal{T}, s \leq t} d(X_s, F) = 0 \right\} = \bigcap_{s \in \mathcal{T} \cap \mathbb{Q}, s \leq t} \{d(X_s, F) = 0\} \in \mathcal{F}_t.$$

□

## 3.2 Martingale theorems

A very important inequality for martingales, is the following theorem due to Doob. As mentioned in the introduction of this chapter, the idea is to prove a similar inequality for discrete-time martingales, then we require that the stochastic process is càdlàg, such that we can lift the result to continuous-time martingales.

**Theorem 3.2.1** (Doob's martingale inequality)**.** *Suppose $X = (X_t)_{t \in \mathcal{T}}$ is a càdlàg martingale. Then for every $T \in \mathcal{T}$ and $\lambda > 0$ we have that*

$$\mathbb{P}\left( \sup_{t \leq T} |X_t| > \lambda \right) \leq \frac{\mathbb{E}(|X_T|)}{\lambda}.$$

*If also for $p > 1$ we have $\mathbb{E}(|X_T|^p) < \infty$, then*

$$\mathbb{E}\left( \sup_{t \leq T} |X_t|^p \right) \leq \left( \frac{p}{p-1} \right)^p \mathbb{E}(|X_T|^p).$$

*Proof.*
**Step 1:** Let $\mathcal{T} = \mathbb{N}_0$ and $X = (X_n)_{n \in \mathbb{N}_0}$ be a discrete-time martingale, then for all $\lambda > 0$ we have for each $n \in \mathbb{N}$ that,

$$\mathbb{P}\left( \max_{0 \leq k \leq n} |X_k| > \lambda \right) \leq \frac{\mathbb{E}(|X_n|)}{\lambda}.$$

Let $\tau$ be a stopping time such that for some $N \in \mathbb{N}$ we have $\tau \leq N$. Note that $\{|X_\tau| > \lambda\} \cap \{\tau = n\} \in \mathcal{F}_n$, and that $|X|$ is a submartingale, so by the submartingale property of $|X|$ we have,

$$\begin{aligned}\mathbb{E}(|X_N|\mathbf{1}_{\{|X_\tau|>\lambda\}}) &= \sum_{n=0}^{N} \mathbb{E}(|X_N|\mathbf{1}_{\{|X_\tau|>\lambda\}\cap\{\tau=n\}}) \geq \sum_{n=0}^{N} \mathbb{E}(|X_n|\mathbf{1}_{\{|X_\tau|>\lambda\}\cap\{\tau=n\}}) \\ &= \sum_{n=0}^{N} \mathbb{E}(|X_\tau|\mathbf{1}_{\{|X_\tau|>\lambda\}\cap\{\tau=n\}}) = \mathbb{E}(|X_\tau|\mathbf{1}_{\{|X_\tau|>\lambda\}})\end{aligned}$$

Now define $\tau = \min\{k \geq 0 : |X_k| > \lambda\} \wedge n$ for some fixed $n \in \mathbb{N}$. Then $\tau$ is a stopping time and also

$$\left\{\max_{0\leq k\leq n} |X_k| > \lambda\right\} = \{|X_\tau| > \lambda\}.$$

Therefore we get the result

$$\begin{aligned}\mathbb{P}\left(\max_{0\leq k\leq n} |X_k| > \lambda\right) &= \mathbb{P}(|X_\tau| > \lambda) = \int_{\{|X_\tau|>\lambda\}} d\mathbb{P} \leq \frac{1}{\lambda}\int_{\{|X_\tau|>\lambda\}} |X_\tau|\, d\mathbb{P} \\ &= \frac{1}{\lambda}\int_\Omega |X_\tau|\mathbf{1}_{\{|X_\tau|>\lambda\}}\, d\mathbb{P} = \frac{\mathbb{E}(|X_\tau|\mathbf{1}_{\{|X_\tau|>\lambda\}})}{\lambda} \leq \frac{\mathbb{E}(|X_n|\mathbf{1}_{\{|X_\tau|>\lambda\}})}{\lambda} \leq \frac{\mathbb{E}(|X_n|)}{\lambda}.\end{aligned}$$

**Step 2:** Suppose $X = (X_t)_{t\in\mathcal{T}}$ is a càdlàg martingale. Then for every $T \in \mathcal{T}$ and $\lambda > 0$ we have that

$$\mathbb{P}\left(\sup_{t\leq T} |X_t| > \lambda\right) \leq \frac{\mathbb{E}(|X_T|)}{\lambda}.$$

Consider any sequence of partitions $\mathcal{P}_n = \{0 = t_0^n < t_1^n < \cdots < t_{p_n}^n = T\}$ with mesh tending to zero ($\max_{0\leq j\leq p_n-1} |t_{j+1}^n - t_j^n| \underset{n\to\infty}{\longrightarrow} 0$) and such that the partitions are nested, meaning that for $n_1 \leq n_2$ all the points in $\mathcal{P}_{n_1}$ are also contained in $\mathcal{P}_{n_2}$. Now for every $t \in \mathcal{T}$ define $X_t^n = X_{t_j^n}$ where $j = \max\{i : t_i < t\}$. Now as continuous-time process, $(X_t^n)_{t\in\mathcal{T}\cap[0,T]}$ is a martingale as for all $s < t$ we have

$$\mathbb{E}(X_t^n|\mathcal{F}_s) = \mathbb{E}(X_{t_j^n}|\mathcal{F}_s) = X_s.$$

But also for all $n \in \mathbb{N}$ as discrete-time process, $(X_{t_j^n})_{0\leq j\leq p_n}$ is again a martingale as

$$\mathbb{E}(X_{t_{j+1}^n}|\mathcal{F}_{t_j^n}) = X_{t_j^n}.$$

Fix $\varepsilon > 0$ arbitrary, then find $t_0 = t_0(\omega)$ such that $|X_{t_0}| > \sup_{t\leq T} |X_t| - \varepsilon/2$, which exists by the definition of the supremum. For every $n \in \mathbb{N}$ we can find $j = j(n)$ such that $t_{j(n)-1}^n \leq t_0 \leq t_{j(n)}^n$. Then $t_{j(n)} \to t_0$. As the mesh of the sequence of partitions goes to zero, we have $X_{t_{j(n)}} \xrightarrow{a.s.} X_{t_0}$, since $X$ is càdlàg. So for sufficiently large $n$, we have

$$\sup_{t\leq T} |X_t^n| > |X_{t_0}| - \varepsilon/2 > \sup_{t\leq T} |X_t| - \varepsilon.$$

Since $\varepsilon > 0$ was arbitrary, define $A_n = \left\{\sup_{t\leq T} |X_t^n| > \lambda\right\}$, then

$$\bigcup_{n=1}^{\infty} A_n = \left\{\sup_{t\leq T} |X_t| > \lambda\right\}.$$

Hence we have by using Step 1,

$$\mathbb{P}\left(\sup_{t\leq T} |X_t| > \lambda\right) = \lim_{n\to\infty} \mathbb{P}\left(\sup_{t\leq T} |X_t^n| > \lambda\right) \leq \lim_{n\to\infty} \frac{\mathbb{E}(|X_{p_n}^n|)}{\lambda} = \frac{\mathbb{E}(|X_T|)}{\lambda}.$$

**Step 3:** Suppose $X = (X_t)_{t\in\mathcal{T}}$ is a càdlàg martingale. Then for every $T \in \mathcal{T}$ and $\lambda > 0$ such that $X_T \in L^p(\mathcal{F}, \mathbb{P})$, we have that

$$\mathbb{E}\left(\sup_{t\leq T} |X_t|^p\right) \leq \left(\frac{p}{p-1}\right)^p \mathbb{E}(|X_T|^p).$$

This follows by a standard rewriting of the $L^p$-norm in terms of level sets. We know that

$$\begin{aligned}\mathbb{E}\left(\sup_{t\leq T}|X_t|^p\right) &= \int_0^\infty p\lambda^{p-1}\mathbb{P}\left(\sup_{t\leq T}|X_t|>\lambda\right)d\lambda\\ &= \int_0^\infty p\lambda^{p-1}\mathbb{P}\left(\sup_{t\leq T}\left(|X_t|\mathbf{1}_{\{\sup_{t\leq T}|X_t|>\lambda\}}\right)>\lambda\right)d\lambda\\ &\leq \int_0^\infty p\lambda^{p-2}\mathbb{E}\left(|X_T|\mathbf{1}_{\{\sup_{t\leq T}|X_t|>\lambda\}}\right)d\lambda\\ &= \int_0^\infty\int_{\{\sup_{t\leq T}|X_t|>\lambda\}} p\lambda^{p-2}|X_T|\,d\mathbb{P}\,d\lambda\\ &= \int_\Omega\int_0^{\sup_{t\leq T}|X_t|} p\lambda^{p-2}|X_T|\,d\mathbb{P}\,d\lambda\\ &= \frac{p}{p-1}\int_\Omega |X_T|\sup_{t\leq T}|X_t|^{p-1}\,d\mathbb{P}\end{aligned}$$

By Hölder's inequality we have that

$$\mathbb{E}\left(\sup_{t\leq T}|X_t|^p\right)\leq\frac{p}{p-1}\mathbb{E}(|X_T|^p)^{\frac{1}{p}}\mathbb{E}\left(\sup_{t\leq T}|X_t|^p\right)^{\frac{p-1}{p}},$$

from which the result follows. □

When we have an unbounded index set, say $\mathcal{T}=[0,\infty)$, we sometimes want to know what happens to a martingale when $t\to\infty$. If we impose uniform integrability on the collection $(X_t)_{t\geq 0}$, then we get the following results

**Theorem 3.2.2** (Doob's convergence theorem)**.** *Let $\mathcal{T}=[0,\infty)$ and $(\Omega,\mathcal{F},\mathbb{F},\mathbb{P})$ be a filtered probability space and $X$ a càdlàg martingale. The following properties are equivalent, if any holds then there exists a random variable $X_\infty\in L^1(\mathcal{F},\mathbb{P})$ such that*

*(i) For every $t\geq 0$ we have $X_t=\mathbb{E}(X_\infty|\mathcal{F}_t)$;*

*(ii) The collection $(X_t)_{t\geq 0}$ is uniformly integrable;*

*(iii) We have $X_t\xrightarrow{a.s}X_\infty$ and $X_t\xrightarrow{L^1}X_\infty$ as $t\to\infty$.*

*Proof.*
(i) $\Longrightarrow$ (ii): We have to show that the family $(\mathbb{E}(X_\infty|\mathcal{F}_t))_{t\geq 0}$ is uniformly integrable. Since $X_\infty\in L^1(\mathcal{F},\mathbb{P})$ we can just as in Proposition 2.4.7 choose a $\delta>0$ given $\varepsilon>0$ such that for all $\mathbb{P}(A)<\delta$, we have $\mathbb{E}(|X_\infty|\mathbf{1}_A)<\varepsilon$. Choose $M>0$ large enough such that by Markov's inequality $\mathbb{E}(|X_\infty|)/M<\delta$. Pick any $t\geq 0$, we have

$$\mathbb{P}(|E(X_\infty|\mathcal{F}_t)|\geq M)\leq\mathbb{E}(|\mathbb{E}(X_\infty|\mathcal{F}_T)|)/M\leq\mathbb{E}(\mathbb{E}(|X_\infty||\mathcal{F}_t))/M=\mathbb{E}(|X_\infty|)/M<\delta.$$

The set $\{|\mathbb{E}(X_\infty|\mathcal{F}_t)|\geq M\}$ is contained in $\mathcal{F}_t$, so it follows that

$$\mathbb{E}(|\mathbb{E}(X_\infty|\mathcal{F}_t)|\mathbf{1}_{\{|\mathbb{E}(X_\infty|\mathcal{F}_t)|\geq M\}})\leq\mathbb{E}(\mathbb{E}(|X_\infty||\mathcal{F}_t)\mathbf{1}_{\{|\mathbb{E}(X_\infty|\mathcal{F}_t)|\geq M\}})=\mathbb{E}(|X_\infty|\mathbf{1}_{\{|\mathbb{E}(X_\infty|\mathcal{F}_t)|\geq M\}})<\varepsilon.$$

The bound was independent of $t$ and of $\varepsilon$, so the uniform integrability follows.
(ii) $\Longrightarrow$ (iii): This proof requires a bit more work, see Le Gall [26, Theorem 3.21].
(iii) $\Longrightarrow$ (i): By dominated convergence, we can pass to the limit $s\to\infty$ for every $t\geq 0$ to get

$$X_t=\lim_{s\to\infty}\mathbb{E}(X_s|\mathcal{F}_t)=\mathbb{E}(X_\infty|\mathcal{F}_t).$$

□

When we stop a martingale with a stopping time, we would hope that process stays a martingale, this turns out to be the case. For the general case we need to impose uniform integrability on the entire stochastic process. We will not prove the theorem right now, instead see Le Gall [26, Theorem 3.22].

**Theorem 3.2.3** (Optional stopping theorem). *Let $(\Omega, \mathcal{F}, \mathbb{F}, \mathbb{P})$ be a filtered probability space and $X$ be a uniformly integrable càdlàg martingale. Let $\rho \leq \tau$ be two $\mathbb{F}$-stopping times, then $X_\rho$ and $X_\tau$ are in $L^1(\mathcal{F}, \mathbb{P})$ and*

$$X_\rho = \mathbb{E}(X_\tau | \mathcal{F}_\rho).$$

Often we will use this theorem in more particular forms. If we impose boundedness on the stopping times we do not have to require uniform integrability.

**Corollary 3.2.4.** *Let $(\Omega, \mathcal{F}, \mathbb{F}, \mathbb{P})$ be a filtered probability space and $X$ a càdlàg martingale. Let $\rho \leq \tau$ be two bounded $\mathbb{F}$-stopping times. Then $X_\rho$ and $X_\tau$ are in $L^1(\mathcal{F}, \mathbb{P})$ and*

$$X_\rho = \mathbb{E}(X_\tau | \mathcal{F}_\rho).$$

*Proof.*
Let $M > 0$ such that $\rho \leq \tau \leq M$. Now we apply Theorem 3.2.3 to the martingale $(X_{t\wedge M})_{t\geq 0}$, which is uniformly integrable because for $s < t$ we have $X_{s\wedge M} = \mathbb{E}(X_{t\wedge M}|\mathcal{F}_s)$, when we consider the cases $t \leq M$ and $t > M$ seperately. Now the uniform integrability follows by Theorem 3.2.2. □

The following corollary is what we use most of the times. This indeed tells us that a stopped martingale still is a martingale.

**Corollary 3.2.5.** *Let $(\Omega, \mathcal{F}, \mathbb{F}, \mathbb{P})$ be a filtered probability space and $X$ a càdlàg martingale. Let $\tau$ be an $\mathbb{F}$-stopping time, then*

*(i) The process $(X_{t\wedge\tau})_{t\geq 0}$ is still a martingale;*

*(ii) Suppose in addition that the martingale $(X_t)_{t\geq 0}$ is uniformly integrable. Then the process $(X_{t\wedge\tau})_{t\geq 0}$ is also a uniformly integrable martingale.*

*Proof.*
First we prove (ii). We know that $t\wedge\tau$ is a stopping time. So then $X_{t\wedge\tau}$ is in $L^1(\mathcal{F}, \mathbb{P})$ by Theorem 3.2.3. Since $\mathcal{F}_{t\wedge\tau} \subseteq \mathcal{F}_t$ we know that $X_{t\wedge\tau}$ is $\mathcal{F}_t$-measurable. If we show it is uniformly integrable, we are done.

By the partial averaging property of conditional expectations, we only have to show that for every $A \in \mathcal{F}_t$

$$\mathbb{E}(\mathbf{1}_A X_\tau) = \mathbb{E}(\mathbf{1}_A X_{t\wedge\tau}).$$

We trivially have $\mathbb{E}(\mathbf{1}_{A\cap\{\tau\leq t\}} X_\tau) = \mathbb{E}(\mathbf{1}_{A\cap\{\tau\leq t\}} X_{t\wedge\tau})$. Furthermore we know that we have $X_{t\wedge\tau} = \mathbb{E}(X_\tau|\mathcal{F}_{t\wedge\tau})$ by Theorem 3.2.3. Additionally both $A \cap \{\tau > t\} \in \mathcal{F}_t$ and $A \cap \{\tau > t\} \in \mathcal{F}_\tau$, so then $A \cap \{\tau > t\} \in \mathcal{F}_{t\wedge\tau}$. Then we have

$$\mathbb{E}(X\mathbf{1}_{A\cap\{\tau>t\}} X_\tau) = \mathbb{E}(\mathbf{1}_{A\cap\{\tau>t\}} X_{t\wedge\tau}).$$

Adding up the last two displays we indeed get

$$X_{t\wedge\tau} = \mathbb{E}(X_\tau|\mathcal{F}_t),$$

proving uniform integrability necessary for Theorem 3.2.3.

To prove (i) we apply (ii) to the uniformly integrable martingale $(X_{t\wedge M})_{t\geq 0}$ for every $M > 0$. □

## 3.3 Local martingales

Martingales have a lot of structure, and behave rather nicely. So much that, for stochastic integration, we can even weaken the structure a bit to still get a satisfactory integration theory, we only have to require a process to be a martingale 'locally'. Given a stochastic process $X = (X_t)_{t\geq 0}$ and a stopping time $\tau$, we define a stopped process $X^\tau = X_{t\wedge\tau}$. In the case where $X$ is a càdlàg martingale, we know that $X^\tau$ behaves well by Corollary 3.2.5, which motivates the following definition of local martingales.

**Definition 3.3.1.** *Let $(\Omega, \mathcal{F}, \mathbb{F}, \mathbb{P})$ be a filtered probability space and $M$ an adapted process such that $M_0 = 0$ almost surely. $M$ is called a local martingale if there exists a nondecreasing sequence $(\tau_n)_{n\in\mathbb{N}}$ of $\mathbb{F}$-stopping times such that $\tau_n \uparrow \infty$ and, for every $n$, the stopped process $M^{\tau_n}$ is a uniformly integrable martingale.*

*More generally, when we do not assume that $M_0 = 0$ almost surely, we say that $M$ is a local martingale if the process $N_t = M_t - M_0$ is a continuous local martingale.*

*In all cases, we say that the sequence of stopping times $(\tau_n)_{n\in\mathbb{N}}$ reduces $M$ if $\tau_n \uparrow \infty$ and, for every $n$, the stopped process $M^{\tau_n}$ is a uniformly integrable martingale.*

Remark that every càdlàg martingale $X$ is a càdlàg local martingale, as the sequence of stopping times $\tau_n = n$ makes the stopped process $X^{\tau_n}$ a uniformly integrable martingale. Indeed, we have $X^{\tau_n}_\infty = X_n$, so then by Theorem 3.2.2 we have that $X^{\tau_n}$ is a uniformly integrable martingale. The converse is in general false.

To test whether a càdlàg local martingale is in fact a càdlàg martingale, we can use the following proposition. It also gives a way to define reducing sequences for càdlàg local martingales, which will be very useful in abstract definitions of càdlàg local martingales.

**Proposition 3.3.2.** *Let $(\Omega, \mathcal{F}, \mathbb{F}, \mathbb{P})$ be a filtered probability space.*

*(i) A càdlàg local martingale $M$ is such that there exists a random variable $Z \in L^1(\mathcal{F}, \mathbb{P})$, satisfying $|M_t| \leq Z$ for every $t \geq 0$, which is a uniformly integrable martingale.*

*(ii) If $M$ is a càdlàg local martingale, $N_t = M_t - M_0$ and $M_0 \in L^1(\mathcal{F}, \mathbb{P})$, the sequence of stopping times*

$$\tau_n = \inf\{t \geq 0 : |N_t| > n\},$$

*reduces $M$, if the filtration is right-continuous.*

*Proof.*
**(i)** Write $M_t = M_0 + N$. By definition there exists a reducing sequence $(\tau_n)_{n\in\mathbb{N}}$ for $N$. By Corollary 3.2.5 we have that $N^{\tau_n}$ is a uniformly integrable martingale, so if $s \leq t$, we have for every $n$,

$$N_{s\wedge\tau_n} = \mathbb{E}(N_{t\wedge\tau_n}|\mathcal{F}_s).$$

Now since by definition $M_0 \in L^1(\mathcal{F}, \mathbb{P})$, we can add it to both sides to get

$$M_{s\wedge\tau_n} = \mathbb{E}(M_{t\wedge\tau_n}|\mathcal{F}_s).$$

Furthermore since $Z$ dominates $(M_{t\wedge\tau_n})_{n\in\mathbb{N}}$ for every $t \geq 0$, we can use dominated convergence to obtain the convergence of $M_{t\wedge\tau_n}$ to $M_t$ in $L^1(\mathcal{F}, \mathbb{P})$. We can therefore pass to the limit $n \to \infty$, to obtain the martingale property

$$M_s = \mathbb{E}(M_t|\mathcal{F}_s).$$

**(ii)** By Proposition 3.1.3 and the right-continuity of $\mathbb{F}$, we have that $(\tau_n)_{n\in\mathbb{N}}$ is a sequence of $\mathbb{F}$-stopping times. Now $|N^{\tau_n}| \leq n$, and $N^{\tau_n}$ is a càdlàg local martingale, so by (i) we have that $N^{\tau_n}$ is a uniformly integrable martingale, completing the proof. □

CHAPTER 4

# Stochastic Calculus

The theory of stochastic integration with respect to continuous semimartingales has become classical by now in the literature. To avoid making this chapter unnecessarily long, we omit most of the proofs in the first two sections of this chapter, in favor of a more detailed approach to modelling the jump part in our stochastic differential equations. A thorough treatment of continuous semimartingales can be found in for example Le Gall [26].

## 4.1 Finite variation processes

To start with stochastic calculus, we will first define integration with 'finite variation processes' as integrators. A suitable integral will be the pathwise Lebesgue-Stieltjes integral, that is to say we integrate pathwise with the Lebesgue-Stieltjes integral. We will not give a complete treatment of Lebesgue-Stieltjes integration here. For a more thorough treatment of finite variation we refer to [10] and [26].

Lebesgue-Stieltjes integration is in fact just Lebesgue integration as we will see, therefore theorems like the dominated convergence theorem carry over immediately. Let $(\Omega, \mathcal{F})$ be a measurable space, and let $\mu : \mathcal{F} \to [-\infty, \infty]$ be a set function such that

1. $\mu(\varnothing) = 0$;
2. $\mu$ attains at most one of the values $\infty$ or $-\infty$;
3. If $(A_n)_{n\in\mathbb{N}} \subseteq \mathcal{F}$ is pairwise disjoint, then

$$\mu\left(\bigcup_{n=1}^{\infty} A_n\right) = \sum_{n=1}^{\infty} \mu(A_n),$$

   where the series converges absolutely when the left hand side is finite.

If $\mu$ satisfies these properties, then we call $\mu$ a *signed measure*, furthermore we call it *finite* when $\mu(\mathcal{F}) \subset \mathbb{R}$. Suppose $\mu$ is a signed measure, then there exists a unique decomposition $\mu = \mu^+ - \mu^-$ where $\mu^+, \mu^- : \mathcal{F} \to [0, \infty]$ are positive measures, where atleast one of them is finite. This decomposition is called the *Jordan decomposition.* We call the positive measure $|\mu| = \mu^+ + \mu^-$ the *total variation measure.* Now finite variation functions are defined as follows.

**Definition 4.1.1.** *Let $T \geq 0$. A continuous function $a : [0, T] \to \mathbb{R}$ such that $a(0) = 0$ is said to have finite variation if there exists a finite signed Borel measure $\mu : \mathcal{B}([0, T]) \to \mathbb{R}$ such that $a(t) = \mu([0, t])$ for every $t \in [0, T]$. Furthermore a function $a : \mathbb{R}^+ \to \mathbb{R}$ is a finite variation function on $\mathbb{R}^+$, if the restriction of $a$ to $[0, T]$ has finite variation on $[0, T]$ for every $T \geq 0$.*

It turns out that this signed measure is unique and using the Jordan decomposition we can define the Lebesgue-Stieltjes integrals through Lebesgue integrals as follows. Let $a$ be a finite variation function with corresponding signed Borel measure $\mu$. Let $f : [0, T] \to \mathbb{R}$ be a measurable function such that $\int_{[0,T]} |f(s)|\, d|\mu|(s) < \infty$. Then we define

$$\int_0^T f(s)\, da(s) := \int_0^T f(s)\, d\mu(s) = \int_0^T f(s)\, d\mu^+(s) - \int_0^T f(s)\, d\mu^-(s),$$
$$\int_0^T f(s)\, |da(s)| = \int_{[0,T]} f(s)\, d|\mu|(s).$$

We have two important approximation results for Lebesgue-Stieltjes integrals

**Proposition 4.1.2.** *Let $a : [0,T] \to \mathbb{R}$ be a function having finite variation. For every $t \in (0,T]$ and every nested sequence of partitions $\mathcal{P}_n = \{0 = t_0^n < t_1^n < \cdots < t_{p_n}^n = t\}$ with mesh tending to zero, we have*

$$\int_0^t |da(s)| = \lim_{n\to\infty} \sum_{i=0}^{p_n-1} |a(t_{i+1}^n) - a(t_i^n)|.$$

*Furthermore if $f : [0,T] \to \mathbb{R}$ is a continuous function and we again let $\mathcal{P}_n$ be a sequence of partitions with mesh tending to zero (they do no longer have to be nested), then*

$$\int_0^T f(s)\, da(s) = \lim_{n\to\infty} \sum_{i=0}^{p_n-1} f(t_i^n)(a(t_{i+1}^n) - a(t_i^n)).$$

An important property of Lesbesgue-Stieltjes which all other integrals we will define in this chapter will also satisfy, is the 'associativity' property.

**Proposition 4.1.3.** *Let $a$ be of finite variation and $f$ an $a$-integrable function. Then define*

$$g(t) = \int_0^t f(s)\, da(s).$$

*We know that $g$ is of finite variation and that for any $g$-integrable function $h$ we have for all $t \geq 0$,*

$$\int_0^t h(s)\, dg(s) = \int_0^t h(s) f(s)\, da(s).$$

Going back to our stochastic setting, we can define finite variation processes to be as follows.

**Definition 4.1.4.** *An adapted continuous process $X$ is called a finite variation process if $X_0 = 0$ and the paths are of finite variation almost surely.*

By the following proposition we can define pathwise Lebesgue-Stieltjes integration, as is proven in Le Gall [26, Proposition 4.5].

**Proposition 4.1.5.** *Let $(\Omega, \mathcal{F}, \mathbb{F}, \mathbb{P})$ be a complete filtered probability space and let $A$ be a finite variation process. Let $X$ be a progressive process and $\mathcal{N}$ a null set such that*

$$\omega \in \mathcal{N}^c, \forall t \geq 0, \int_0^t |X_s(\omega)||dA_s(\omega)| < \infty.$$

*Then the process $X \cdot A = ((X \cdot A)_t)_{t\geq 0}$ defined by*

$$(X \cdot A)_t(\omega) = \begin{cases} \int_0^t X_s(\omega)\, dA_s(\omega), & \omega \in \bigcap_{n=1}^{\infty} \left\{ \int_0^n |X_s(\omega)|\, |dA_s(\omega)| < \infty \right\}, \\ 0, & \textit{otherwise}. \end{cases}$$

*is also a finite variation process.*

*Remark* 4.1.6. We have assumed the filtered probability space to be complete, we could have circumvented this by allowing the integrals to be infinite. So we would define

$$(X \cdot A)_t(\omega) = \int_0^t X_s\, dA_s(\omega),$$

for all $\omega \in \Omega$. However, we can not guarantee anymore that $(X \cdot A)$ is still a finite variation process, see Protter [33] for a counterexample.

To prove some identities with finite variation processes, it is often useful to have a way to approximate the finite variation integrals. We can do this for càdlàg progressive processes as follows.

**Proposition 4.1.7.** *Let $X$ be a càdlàg progressive process and $A$ a finite variation process, such that $X \cdot A$ is defined almost surely. Let $t \geq 0$ and $\mathcal{P}_n = \{0 = t_0^n < t_1^n < \cdots < t_{p_n}^n = t\}$ with mesh tending to zero. Then*

$$\sum_{i=0}^{p_n-1} X_{t_i^n}(A_{t_{i+1}^n} - A_{t_i^n}) \xrightarrow{\mathbb{P}} \int_0^t X_{s-}\, dA_s.$$

*Proof.*
Fix $t \geq 0$. Define the progressive process

$$X_s^n = \sum_{i=0}^{p_n-1} X_{t_i^n} \mathbf{1}_{(t_i^n, t_{i+1}^n]}(s),$$

then

$$\int_0^t X_s^n \, dA_s = \sum_{i=0}^{p_n-1} X_{t_i^n} (A_{t_{i+1}^n} - A_{t_i^n}).$$

So if we can prove that

$$\int_0^t (X_s^n - X_{s-}) \, dA_s \xrightarrow{\mathbb{P}} 0,$$

we are done. By the càdlàg property of $X$, we know that for every $t \geq 0$, we have $X_t^n \to X_{t-}$ as $n \to \infty$ almost surely. Now define $K_t = \sup_{s \leq t} |X_s|$, then $|X_t^n - X_{t-}| \leq 2K_t$ for all $t \geq 0$ and $n \in \mathbb{N}$.

Define for every $m \in \mathbb{N}$ the stopping time

$$\tau_m = \inf \left\{ r \in [0,t] : \int_0^r |K_s| \, |dA_s| \geq m \right\} \wedge t,$$

we know that $\tau_m \to t$ as $m \to \infty$ almost surely. Then for every $m \in \mathbb{N}$ we have that $\mathbb{E}(\int_0^{\tau_m} |K_s| \, |dA_s|) \leq m$, so then by the Dominated Convergence Theorem

$$\mathbb{E}\left( \int_0^{\tau_m} |X_s^n - X_{s-}| \, |dA_s| \right) \to 0,$$

as $n \to \infty$.

Given $\delta > 0$ and $\varepsilon > 0$, fix $m \in \mathbb{N}$ such that $\mathbb{P}(\tau_m < t) < \varepsilon/2$. Now pick $n$ large enough such that

$$\mathbb{E}\left( \int_0^{\tau_m} |X_s^n - X_{s-}| \, |dA_s| \right) < \frac{\delta\varepsilon}{4}.$$

Then we have that

$$\begin{aligned}
&\mathbb{P}\left( \left| \int_0^t (X_s^n - X_{s-}) \, dA_s \right| > \delta \right) \\
&\qquad \leq \mathbb{P}\left( \int_0^{\tau_m} |X_s^n - X_{s-}| \, |dA_s| > \frac{\delta}{2} \right) + \mathbb{P}\left( \int_{\tau_m}^t |X_s^n - X_{s-}| \, |dA_s| \mathbf{1}_{\tau_m < t} > \frac{\delta}{2} \right) \\
&\qquad \leq \frac{2}{\delta} \mathbb{E}\left( \int_0^{\tau_m} |X_s^n - X_{s-}| \, |dA_s| \right) + \mathbb{P}(\tau_m < t) < \varepsilon.
\end{aligned}$$

Hence the result follows. □

Using finite variation processes as integrators for our stochastic integrals is a good start, but we also want to be able to integrate with respect to martingales. It turns out that the concept of pathwise Lebesgue-Stieltjes integration is insufficient for this purpose and we have to consider a different type of stochastic integrals.

**Theorem 4.1.8.** *A continuous local martingale $M$ with $M_0 = 0$ almost surely is a process of finite variation if and only if $M$ is indistuinguishable from zero.*

## 4.2 Quadratic variation and covariation

In this section we will solely focus on continuous processes and we will introduce the concepts of quadratic variation and covariation which are important for stochastic integration with respect to continuous martingales. Later on we want to integrate with respect to a class of càdlàg martingales, however we will divide the integration in a continuous part and a jump part, instead of a fully general càdlàg semimartingale integration as done in Protter [33].

First we will look at a relation between a continuous local martingale and its squared process. A more general statement is known as the Doob-Meyer decomposition.

**Theorem 4.2.1.** *Let $(\Omega, \mathcal{F}, \mathbb{F}, \mathbb{P})$ be a complete filtered probability space. Let $M$ be a continuous local martingale. There exists a unique (up to indistinguishability) finite variation process $(\langle M\rangle_t)_{t\geq 0}$, which we call the quadratic variation of $M$, such that $(M_t^2 - \langle M\rangle_t)_{t\geq 0}$ is a continuous local martingale.*

*Furthermore, for any $t > 0$ and any sequence of partitions $\mathcal{P}_n = \{0 = t_0^n < t_1^n < \cdots < t_{p_n}^n = t\}$ with mesh tending to zero, we have*

$$\sum_{j=0}^{m-1} (M_{t_{j+1}^n} - M_{t_j^n})^2 \xrightarrow{\mathbb{P}} \langle M\rangle_t.$$

*Example* 4.2.2. Let $W$ be a Brownian motion, then $\langle W\rangle_t = t$. Indeed we know that $W_{t_{j+1}^n} - W_{t_j^n} \sim \mathcal{N}(0, t_{j+1}^n - t_j^n)$. So then $\mathbb{E}\left(\sum_{j=0}^{p_n-1}(W_{t_{j+1}^n} - W_{t_j^n})^2\right) = t$. Now the increments are independent and recall that for $X \sim \mathcal{N}(0,\sigma^2)$ we have $\mathrm{Var}(X^2) = \mathbb{E}(X^4) = 2\sigma^4$, so

$$\begin{aligned}
\mathbb{E}\left(\left(\sum_{j=0}^{p_n-1}(W_{t_{j+1}^n} - W_{t_j^n})^2 - t\right)^2\right) &= \mathrm{Var}\left(\sum_{j=0}^{p_n-1}(W_{t_{j+1}^n} - W_{t_j^n})^2\right)\\
&= \sum_{j=0}^{p_n-1} \mathrm{Var}((W_{t_{j+1}^n} - W_{t_j^n})^2)\\
&= 2\sum_{j=0}^{p_n-1}(t_{j+1}^n - t_j^n)^2 \leq 2 \max_{0\leq j\leq p_{n-1}} |t_{j+1}^n - t_j^n| t \to 0.
\end{aligned}$$

Hence we have convergence in $L^2$, in particular we have convergence in probability and we indeed have $\langle W\rangle_t = t$.

We will extend the definition of quadratic variation to general stochastic processes through the approximation property.

**Definition 4.2.3.** *Let $X$ be a stochastic process. The quadratic variation of $X$ up to time $t$ is defined as the stochastic process $\langle X\rangle$ such that*

$$\sum_{j=0}^{p_n-1} (X_{t_{j+1}} - X_{t_j})^2 \xrightarrow{\mathbb{P}} \langle X\rangle_t,$$

*when the limit exists, where $\mathcal{P}_n = \{0 = t_0^n < t_1^n < \cdots < t_{p_n}^n = t\}$ is a sequence of partitions with mesh tending to zero.*

We could see quadratic variation as a variant of variance for stochastic processes, similarly we can define covariation as a variant of covariance.

**Definition 4.2.4.** *If $M$ and $N$ are two continuous local martingales, the covariation $\langle M, N\rangle$ is the finite variation process defined by setting for every $t \geq 0$,*

$$\langle M, N\rangle_t = \frac{1}{2}\left(\langle M+N, M+N\rangle_t - \langle M, M\rangle_t - \langle N, N\rangle_t\right).$$

Again we can get similar results as in Theorem 4.2.1 through this polarization identity.

**Proposition 4.2.5.** *(i) $\langle M, N\rangle$ is the unique (up to indistinguishability) finite variation process such that $M_tN_t - \langle M, N\rangle_t$ is a continuous local martingale.*

*(ii) The mapping $(M, N) \mapsto \langle M, N\rangle$ is bilinear and symmetric.*

*(iii) If $\mathcal{P}_n = \{0 = t_0^n < t_1^n < \cdots < t_{p_n}^n = t\}$ is a sequence of partitions, we have*

$$\lim_{n\to\infty} \sum_{i=0}^{p_n-1} (M_{t_{i+1}^n} - M_{t_i^n})(N_{t_{i+1}^n} - N_{t_i^n}) \xrightarrow{\mathbb{P}} \langle M, N\rangle_t.$$

*(iv) For every stopping time $\tau$, we have $\langle M^\tau, N^\tau\rangle_t = \langle M^\tau, N\rangle_t = \langle M, N\rangle_{t\wedge\tau}$.*

(v) *If $M$ and $N$ are two continuous local martingales bounded in $L^2$, $M_tN_t - \langle M,N\rangle_t$ is a uniformly integrable martingale. Consequently, $\langle M,N\rangle_\infty$ is well defined as the almost sure limit of $\langle M,N\rangle_t$ as $t\to\infty$, is integrable, and satisfies*

$$\mathbb{E}(M_\infty N_\infty) = \mathbb{E}(M_0N_0) + \mathbb{E}(\langle M,N\rangle_\infty).$$

There exists a generalisation of the Cauchy-Schwarz inequality, but then for stochastic processes, which is called in the context of stochastic processes, the Kunita-Watanabe inequality. The proof is in essence the same as for the generalized Cauchy-Schwarz inequality.

**Proposition 4.2.6** (Kunita-Watanabe Inequality)**.** *Let $M$ and $N$ be two continuous local martingales and let $X$ and $Y$ be two progressive processes. Then, almost surely*

$$\int_0^\infty |X_s||Y_s||d\langle M,N\rangle_s| \leq \left(\int_0^\infty X_s^2\, d\langle M\rangle_s\right)^{\frac{1}{2}} \left(\int_0^\infty Y_s^2\, d\langle N\rangle_s\right)^{\frac{1}{2}}$$

The following class of stochastic processes will be nearly the biggest class of integrators which can be used for stochastic integration, in fact the biggest class of integrators, is the class of càdlàg semimartingales, Protter [33].

**Definition 4.2.7.** *A process $V=(V_t)_{t\geq 0}$ is a continuous semimartingale if it can be written in the form*

$$V_t = M_t + A_t,$$

*where $M$ is a continuous local martingale and $A$ a finite variation process.*

The decomposition $V = M + A$ is unique up to indistinguishability by Proposition 4.1.8, we say that this is the canonical decomposition of $V$. To rewrite stochastic integrals with semimartingales as integrators, we will often use the following proposition.

**Proposition 4.2.8.** *Let $V$ and $W$ be two continuous semimartingales and let $\mathcal{P}_n = \{0 = t_0^n < t_1^n < \ldots < t_{p_n}^n = t\}$ be an increasing sequence of partitions whose mesh tends to $0$. Then,*

$$\sum_{i=0}^{p_n-1}(V_{t_{i+1}^n} - V_{t_i^n})(W_{t_{i+1}^n} - W_{t_i^n}) \xrightarrow{\mathbb{P}} \langle V,W\rangle_t.$$

*Proof.*
Let $V = M + A$ and $W = M' + A'$ be the canonical decompositions. Then

$$\begin{aligned}
&\sum_{i=0}^{p_n-1}(V_{t_{i+1}^n} - V_{t_i^n})(W_{t_{i+1}^n} - W_{t_i^n}) \\
&\quad= \sum_{i=0}^{p_n-1}((M_{t_{i+1}^n} - M_{t_i^n}) + (A_{t_{i+1}^n} - A_{t_i^n}))((M'_{t_{i+1}^n} - M'_{t_i^n}) + (A'_{t_{i+1}^n} - A'_{t_i^n})) \\
&\quad= \sum_{i=0}^{p_n-1}(M_{t_{i+1}^n} - M_{t_i^n})(M'_{t_{i+1}^n} - M'_{t_i^n}) + \sum_{i=0}^{p_n-1}(A_{t_{i+1}^n} - A_{t_i^n})(A'_{t_{i+1}^n} - A'_{t_i^n}) \\
&\qquad+ \sum_{i=0}^{p_n-1}(M_{t_{i+1}^n} - M_{t_i^n})(A'_{t_{i+1}^n} - A'_{t_i^n}) + \sum_{i=0}^{p_n-1}(A_{t_{i+1}^n} - A_{t_i^n})(M'_{t_{i+1}^n} - M'_{t_i^n})
\end{aligned}$$

We already know that

$$\sum_{i=0}^{p_n-1}(M_{t_{i+1}^n} - M_{t_i^n})(M'_{t_{i+1}^n} - M'_{t_i^n}) \xrightarrow{\mathbb{P}} \langle X,Y\rangle_t,$$

so it remains to show all the other terms vanish. Indeed,

$$\begin{aligned}
\left|\sum_{i=0}^{p_n-1}(M_{t_{i+1}^n} - M_{t_i^n})(A'_{t_{i+1}^n} - A'_{t_i^n})\right| &\leq \sup_{0\leq i\leq p_n-1}|M_{t_{i+1}^n} - M_{t_i^n}| \sum_{i=0}^{p_n-1}|A'_{t_{i+1}^n} - A'_{t_i^n}| \\
&\leq \left(\int_0^t |dA'_s|\right) \sup_{0\leq i\leq p_n-1}|M_{t_{i+1}^n} - M_{t_i^n}| \xrightarrow{a.s.} 0,
\end{aligned}$$

by the continuity of $M$ on the compact interval $[0,t]$. The other two terms go in a similar manner and the result follows. □

## 4.3 Stochastic integration

Having discussed quadratic variation and covariation, we are now ready to define the stochastic integrals which we will use in the rest of the thesis extensively. First we will look at stochastic integrals with respect to continuous integrators, it turns out that the sum of a continuous local martingale and a finite variation process gives a satisfactory theory of integration.

Then we will define integrals for Poisson measures, these integrals will model the discontinuous part of our processes. Together we have a very general theory of integration which will be sufficient for our purposes.

### 4.3.1 Continuous semimartingale stochastic integrals

We write $\mathbb{H}^2$ as the space of all continuous martingales which are bounded in $L^2$ and such that $M_0 = 0$ almost surely, with the convention that two indistinguishable processes are identified. Proposition 4.2.5 shows that for $M, N \in \mathbb{H}^2$, the random variable $\langle M, N\rangle_\infty$ is well defined, and we have $\mathbb{E}(|\langle M, N\rangle_\infty|) < \infty$. So we can define the symmetric bilinear form on $\mathbb{H}^2$ by the formula

$$(M, N)_{\mathbb{H}^2} = \mathbb{E}(\langle M, N\rangle_\infty) = \mathbb{E}(M_\infty N_\infty).$$

It follows that $(M, N)_{\mathbb{H}^2}$ is an inner product on $\mathbb{H}^2$. Clearly $(M, M)_{\mathbb{H}^2} = 0$ if and only if $M = 0$, the other properties follow from Proposition 4.2.5. The inner product induces the norm on $\mathbb{H}^2$ given by

$$\|M\|_{\mathbb{H}^2} = (M, M)_{\mathbb{H}^2}^{1/2} = \mathbb{E}(\langle M, M\rangle_\infty)^{1/2} = \mathbb{E}((M_\infty)^2)^{1/2}.$$

**Proposition 4.3.1.** *The space $\mathbb{H}^2$ equipped with the inner product $(M, N)_{\mathbb{H}^2}$ is a Hilbert space.*

We define the space $L^2_M$ for each $M \in \mathbb{H}^2$ by the set of all progressive proccesses $X$ such that

$$\mathbb{E}\left(\int_0^T X_s^2 \, d\langle M\rangle_s\right) < \infty,$$

with the convention that two progressive processes $X$ and $Y$ satisfying this integrability condition are identified if $X = Y$, $\langle M\rangle_s$-almost everywhere, almost surely.

We can view $L^2_M$ as an ordinary $L^2$ space, namely the space

$$L^2_M = L^2\left(\Omega \times [0, T], \mathscr{P}, \mu_M\rangle\right),$$

where $\mu_M$ is the measure (called the Doléans measure) defined for all $A \in \mathscr{P}$ by $\mu_M(A) = \mathbb{E}\left(\int_A d\langle M\rangle_s\right)$. Therefore it inherits its structure, and we know that $L^2_M$ is a Hilbert space with the inner product

$$(X, Y)_{L^2_M} = \mathbb{E}\left(\int_0^T X_s Y_s \, d\langle M\rangle_s\right),$$

and the induced norm

$$\|X\|_{L^2_M} = \left(\mathbb{E}\left(\int_0^T X_s^2 \, d\langle M\rangle_s\right)\right)^{1/2}.$$

We will first define the stochastic integral with respect to simple processes which space we denote by $\mathscr{S}$, then we show it is dense in $L^2_M$ as vector subspace so that we can extend the integral through an isometry.

**Definition 4.3.2.** *A simple process is a progressive process of the form*

$$X_s(\omega) = \sum_{i=0}^{n-1} X_{(i)} \mathbf{1}_{(t_i, t_{i+1}]}(s),$$

*where $0 = t_0 < t_1 < t_2 < \cdots < t_p$ and for every $i = 0, 1, \ldots, n-1$, $X_{(i)}$ is a bounded $\mathcal{F}_{t_i}$-measurable random variable. Let $\mathscr{S}$ be the set of simple processes, it forms a linear subspace of $L^2_M$ under the identification of processes inherited from $L^2_M$ for every $M \in \mathbb{H}^2$.*

As we already forshadowed, the space $\mathscr{S}$ is dense in $L^2_M$.

**Proposition 4.3.3.** *For every $M \in \mathbb{H}^2$, the linear subspace $\mathscr{S}$ is dense in $L^2_M$.*

Given a simple process, we define the Itô integral with respect to a continuous martingale as follows.

**Definition 4.3.4.** *Let $M \in \mathbb{H}^2$, and $X \in \mathscr{S}$ be a simple process. Then define the Itô integral of $X$ with respect to $M$ for any $t \geq 0$ by*

$$X \cdot M = \int_0^t X_s \, dM_s = \sum_{i=0}^{n-1} X_{(i)}(M_{t_{i+1}\wedge t} - M_{t_i \wedge t}).$$

*We use both notations interchangeably.*

*Remark* 4.3.5. The fact that we use a similar notation $X \cdot A$ and $X \cdot M$ for the integrals with respect to a finite variation process $A$ and with respect to a continuous martingale $M$ creates no ambiguity, as by Proposition 4.1.8 only the zero process is both a finite variation process and a continuous martingale. From the context it should be clear which of the two stochastic integrals we use. We will also use $X \cdot M$ for stochastic integration where $M$ is a martingale-valued measure, but again from the context it should be clear what class $M$ lies in and which stochastic integral we use.

The following theorem proves the extension of the stochastic integral to general stochastic processes in $L^2_M$ and its characterizing properties.

**Theorem 4.3.6.** *Let $M \in \mathbb{H}^2$ and $X \in \mathscr{S}$ a simple process. Then the mapping $X \mapsto X \cdot M$ extends to a linear isometry from $L^2_M$ to $\mathbb{H}^2$. Furthermore, $X \cdot M$ is the unique martingale of $\mathbb{H}^2$ that satisfies the property*

$$\langle X \cdot M, N \rangle = X \cdot \langle M, N \rangle, \ \ \forall N \in \mathbb{H}^2.$$

*If $\tau$ is a stopping time, we have*

$$(\mathbf{1}_{[0,\tau]} X) \cdot M = (X \cdot M)^\tau = X \cdot M^\tau.$$

A useful consequence of the characterizing property of the Itô integral with a continuous martingale as integrator is the Itô isometry.

**Proposition 4.3.7** (Itô Isometry)**.** *Let $M, N \in \mathbb{H}^2, X \in L^2_M$ and $Y \in L^2_N$. Since $X \cdot M$ and $Y \cdot N$ are martingales in $\mathbb{H}^2$, we have*

$$\mathbb{E}\left(\left(\int_0^T X_s \, dM_s\right)\left(\int_0^T Y_s \, dN_s\right)\right) = \mathbb{E}\left(\int_0^T X_s Y_s \, d\langle M, N\rangle_s\right).$$

*In particular, we have that*

$$\mathbb{E}\left(\left(\int_0^T X_s \, dM_s\right)^2\right) = \mathbb{E}\left(\int_0^T X_s^2 \, d\langle M\rangle_s\right).$$

Similarly as with the stochastic integral for finite variation processes and the Lebesgue-Stieltjes integral, we again have an associativity property for the Itô integral.

**Proposition 4.3.8.** *Let $X \in L^2_M$. If $Y$ is a progressive process, we have $XY \in L^2_M$ if and only if $Y \in L^2_{X\cdot M}$. If the latter properties hold,*

$$(XY) \cdot M = Y \cdot (X \cdot M)$$

We can extend the definition of $X \cdot M$ to an arbitrary continuous local martingale. If $M$ is a continuous local martingale, we write $L^2_{M,loc}$ and $L^2_M$ for the set of progressive processes $X$ such that respectively

$$\mathbb{P}\left(\int_0^T X_s^2 \, d\langle M\rangle_s < \infty\right) = 1 \qquad \text{and} \qquad \mathbb{E}\left(\int_0^\infty X_s^2 \, d\langle M\rangle_s\right) < \infty.$$

Now $L^2_M$ can be seen as a Hilbert space as before.The following theorem, shows that we can indeed extend the definition of the Itô integral to continuous local martingales.

**Theorem 4.3.9.** *Let $M$ be a continuous local martingale. For every $X \in L^2_{M,loc}$ there exists a unique continuous local martingale with initial value 0, which is denoted by $X \cdot M$, such that, for every continuous local martingale $N$,*

$$\langle X \cdot M, N \rangle = X \cdot \langle M, N \rangle.$$

*If $\tau$ is a stopping time, we have*

$$(\mathbf{1}_{[0,\tau]} X) \cdot M = (X \cdot M)^\tau = X \cdot M^\tau.$$

*If $X \in L^2_{M,loc}$ and $Y$ is a progressive process, we have $Y \in L^2_{X \cdot M,loc}$ if and only if $XY \in L^2_{M,loc}$, and then*

$$X \cdot (Y \cdot M) = XY \cdot M.$$

*Finally, if $M \in \mathbb{H}^2$, and $X \in L^2_M$, the definition of $X \cdot M$ is consistent with our earlier definition of stochastic integration with respect to continuous martingales.*

Having defined a stochastic integral for finite variation processes and for continuous local martingales, we can now look at stochastic integrals for continuous semimartingales. First we will look at the suitable class of integrands. They turn out to be the class of locally bounded progressive processes.

**Definition 4.3.10.** *Let $X$ be a progressive process, we called it locally bounded if for all $t \geq 0$, we have almost surely*

$$\sup_{s \leq t} |X_s| < \infty.$$

In particular, any adapted process with continuous paths is a locally bounded progressive process. If $X$ is progressive and locally bounded, then for every finite variation process $A$, we have for all $t \geq 0$,

$$\int_0^t |X_s| \, |dA_s| < \infty, a.s.$$

and similarly $X \in L^2_{M,loc}$ for every continuous local martingale $M$. Now the stochastic integral of a locally bounded progressive process with respect to a continuous semimartingale is as follows.

**Definition 4.3.11.** *Let $V$ be a continuous semimartingale and let $V = M + A$ be its canonical decomposition. If $X$ is a locally bounded progressive process, the stochastic integral $X \cdot V$ is the continuous semimartingale with canonical decomposition*

$$X \cdot V = X \cdot M + X \cdot A,$$

*and we write*

$$(X \cdot V)_t = \int_0^t X_s \, dV_s.$$

The following proposition follows immediately from considering the canonical decomposition of the stochastic integral, and summarises most of the important properties of Itô integrals with respect to continuous local martingales

**Proposition 4.3.12.** *Let $V$ be a continuous local martingale and $X$ a locally bounded progressive process, then $X \cdot V$ satisfies the following properties*

*(i) The mapping $(X, V) \mapsto X \cdot V$ is bilinear.*

*(ii) $X \cdot (Y \cdot V) = (XY) \cdot V$, if $X$ and $Y$ are progressive and locally bounded.*

*(iii) For every stopping time $\tau$, $(X \cdot V)^\tau = X \mathbf{1}_{[0,\tau]} \cdot V = X \cdot V^\tau$.*

*(iv) If $V$ is a continuous local martingale, respectively if $V$ is a finite variation process, then the same holds for $X \cdot V$*

A very strong property from Lebesgue integrals is the dominated convergence theorem. It turns out that for stochastic integrals with respect to continuous local semimartingales we have a similar version called the stochastic dominated convergence theorem.

**Theorem 4.3.13** (Stochastic Dominated Convergence Theorem)**.** *Let $V = M + A$ be the canonical decomposition of a continuous semimartingale $V$, and let $t > 0$. Let $(X^n)_{n\in\mathbb{N}}$ and $X$ be locally bounded progressive processes, and let $Y$ be a nonnegative progressive process. Assume that the following properties hold almost surely:*

1. *$X^n_s \to X_s$ as $n \to \infty$, for every $s \in [0,t]$*
2. *$|X^n_s| \leq Y$, for every $n \in \mathbb{N}$ and $s \in [0,t]$*
3. *$\int_0^t Y_s^2\, d\langle M\rangle_s < \infty$ and $\int_0^t Y_s |dA_s| < \infty$*

*Then,*

$$\int_0^t X^n_s\, dV_s \xrightarrow{\mathbb{P}} \int_0^t X_s\, dV_s.$$

Using the stochastic dominated convergence theorem, we can prove some approximation results for the Itô integral in preparation for the most important theorem in this chapter, Itô's formula.

**Proposition 4.3.14.** *Let $V, W$ be continuous semimartingales and $X$ an adapted càdlàg process. Let $t \geq 0$ and assume that $\mathcal{P}_n = \{0 = t^n_0 < \cdots < t^n_{p_n} = t\}$ is a sequence of partitions of $[0,t]$ with mesh tending to zero. Then*

$$\sum_{i=0}^{p_n-1} X_{t_i^n}(V_{t^n_{i+1}} - V_{t^n_i}) \xrightarrow{\mathbb{P}} \int_0^t X_{s-}\, dV_s,$$

*and also*

$$\sum_{i=0}^{p_n-1} X_{t_i^n}(V_{t^n_{i+1}} - V_{t^n_i})(W_{t^n_{i+1}} - W_{t^n_i}) \xrightarrow{\mathbb{P}} \int_0^t X_{s-}\, d\langle V, W\rangle_s.$$

*Proof.*
Define the progressive processes

$$X^n_s = \sum_{i=0}^{p_n-1} X_{t^n_i} \mathbf{1}_{(t^n_i, t^n_{i+1}]}(s).$$

We will first show that

$$\int_0^t X^n_s\, dV_s = \sum_{i=0}^{p_n-1} X_{t^n_i}(V_{t^n_{i+1}} - V_{t^n_i}). \tag{4.1}$$

Let $V = M + A$ be the canonical decomposition. In the case $M = 0$, we have already shown this, so we can assume that $V = M$ and that $M_0 = 0$, by stopping $M$ we can also assume that $M \in \mathbb{H}^2$. Now fix $n \in \mathbb{N}$ and define the stopping times $\tau_m = \inf\{t^n_i : |X_{t^n_i}| \geq m\}$ and we have $\tau_m \uparrow \infty$ as $m \to \infty$. For every $m$, we have that

$$\sum_{i=0}^{p_n-1} X_{t^n_i} \mathbf{1}_{\{t^n_i \leq \tau_m\}} \mathbf{1}_{(t^n_i, t^n_{i+1}]},$$

is a simple process, then (4.1) follows by letting $m \to \infty$ as by the definition of the stochastic integral with respect to a martingale of $\mathbb{H}^2$, we have

$$(X^n \cdot M)_{t\wedge\tau_m} = (X^n \mathbf{1}_{[0,\tau_m]} \cdot M)_t = \sum_{i=0}^{p_n-1} X_{t^n_i} \mathbf{1}_{\{t^n_i \leq \tau_m\}} (M_{t^n_{i+1}\wedge t} - M_{t^n_i \wedge t}).$$

For the first assertion we have to show that $X^n \cdot M \xrightarrow{\mathbb{P}} X_{\cdot-} \cdot M$. By the càdlàg property of $X$, we know that almost surely $X^n_t \to X_{t-}$ as $n \to \infty$ for all $t \geq 0$. Now define $Y_t = \sup_{s\leq t} |X_s|$, then $|X^n_t - X_{t-}| \leq 2Y_t$ for all $t \geq 0$ and $n \in \mathbb{N}$. Since $Y$ is again a locally bounded process, the first assertion follows by Theorem 4.3.13.

Let $W = M' + A'$ be the canonical decomposition, then by similar arguments as in Theorem 4.2.8 we only have to show that

$$\sum_{i=0}^{p_n-1} X_{t^n_i}(M_{t^n_{i+1}} - M_{t^n_i})(M'_{t^n_{i+1}} - M'_{t^n_i}) \xrightarrow{\mathbb{P}} \int_0^t X_{s-}\, d\langle V, W\rangle_s.$$

Note that by Proposition 4.2.5 and the previous assertion, we have for every $t \geq 0$ and every sequence of partition of $[0,t]$, $0 = t_0^n < t_1^n < \cdots < t_{p_n}^n = t$ with mesh tending to zero, that in probability

$$\begin{aligned}\langle V, W\rangle_t &= \lim_{n\to\infty} \sum_{i=0}^{p_n-1} (M_{t_{i+1}^n} - M_{t_i^n})(M'_{t_{i+1}^n} - M'_{t_i^n}) \\ &= \lim_{n\to\infty} \sum_{i=0}^{p_n-1} (M_{t_{i+1}^n} M'_{t_{i+1}^n} - M_{t_i^n} M'_{t_i^n}) - \sum_{i=0}^{p_n-1} M_{t_i^n}(M'_{t_{i+1}^n} - M'_{t_i^n}) - \sum_{i=0}^{p_n-1} M_{t_i^n}(M'_{t_{i+1}^n} - M'_{t_i^N}) \\ &= M_t M'_t - \int_0^t M_s \, dM'_s - \int_0^t M'_s \, dM_s\end{aligned}$$

Then by Proposition 4.1.7 we have that

$$\sum_{i=0}^{p_n-1} X_{t_i^n} \left( M_{t_{i+1}^n} M'_{t_{i+1}^n} - M_{t_i^n} M'_{t_i^n} - \int_{t_i^n}^{t_{i+1}^n} M_s \, dM'_s - \int_{t_i^n}^{t_{i+1}^n} M'_s \, dM_s \right) \xrightarrow{\mathbb{P}} \int_0^t X_{s-} \, d\langle V, W\rangle_s.$$

Now in probability

$$\begin{aligned}&\lim_{n\to\infty} \sum_{i=0}^{p_n-1} X_{t_i^n}(M_{t_{i+1}^n} - M_{t_i^n})(M'_{t_{i+1}^n} - M'_{t_i^n}) \\ &\qquad = \lim_{n\to\infty} \sum_{i=0}^{p_n-1} X_{t_i^n}(M_{t_{i+1}^N} M'_{t_{i+1}^n} - M_{t_i^n} M'_{t_i^n}) - \sum_{i=0}^{p_n-1} X_{t_i^n} M_{t_i^n}(M'_{t_{i+1}^n} - M'_{t_i^n}) \\ &\qquad\quad - \sum_{i=0}^{p_n-1} X_{t_i^n} M'_{t_i^n}(M_{t_{i+1}^n} - M_{t_i^n})\end{aligned}$$

If we can show that

$$\sum_{i=0}^{p_n-1} X_{t_i^n} M_{t_i^n}(M'_{t_{i+1}^n} - M'_{t_i^n}) \xrightarrow{\mathbb{P}} \sum_{i=0}^{p_n-1} X_{t_i^n} \left( \int_{t_i^n}^{t_{i+1}^n} M_s \, dM'_s \right),$$

and

$$\sum_{i=0}^{p_n-1} X_{t_i^n} M'_{t_i^n}(M_{t_{i+1}^n} - M_{t_i^n}) \xrightarrow{\mathbb{P}} \sum_{i=0}^{p_n-1} X_{t_i^n} \left( \int_{t_i^n}^{t_{i+1}^n} M'_s \, dM_s \right),$$

we are done, by symmetry it suffices to prove only the first identity. This follows by the first assertion of the theorem and Proposition 4.3.9 In probability we have

$$\begin{aligned}\lim_{n\to\infty} \sum_{i=0}^{p_n-1} X_{t_i^n} M_{t_i^n}(M'_{t_{i+1}^n} - M'_{t_i^n}) &= \int_0^t X_s M_s \, dM'_s = \int_0^t X_s \, d(M \cdot M')_s \\ &= \lim_{n\to\infty} \sum_{i=0}^{p_n-1} X_{t_i^n} \left( \int_{t_i^n}^{t_{i+1}^n} M'_s \, dM_s \right).\end{aligned}$$

Therefore the second assertion of the theorem follows as well. □

### 4.3.2 Poisson stochastic integrals

To define stochastic integrals with integrators which are not necessarily continuous, we will use the method of integration against random measures. This theory has become standard for the jump parts of Lévy processes.

Less standard is the treatment of stochastic integrals against Brownian motion through random measures as done in Applebaum [1], which is a mere consequence of the theory in this section. Extending this theory to general continuous semimartingales is however non-trivial, hence why the continuous stochastic integration has been done seperately, through convential theory.

For this setting, progressive measurability will fail as we will have some subtle measurability problems. Instead we have to look at a more restricted type of measurability, called predictability. A stochastic process is predictable, when it is measurable to the predictable $\sigma$-algebra $\Sigma_p$ where $\Sigma_p =$

$\sigma\left(\{\mathcal{T}\times\Omega\to\mathbb{R}:\text{ left-continuous and adapted}\}\right)$. We will not prove here that $\Sigma_p\subset\mathscr{P}$, instead we refer to Chung and Williams [9].

When $\mathcal{T}=\mathbb{N}$, this means that a process $(X_n)_{n\in\mathbb{N}}$ is predictable if and only if $X_n$ is $\mathcal{F}_{n-1}$-measurable. Intuitively this means that the value of $X_n$ is only based on the information at the earlier time step $n-1$, hence the name predictability.

Let $(S,\mathcal{A})$ be a measurable space and $(\Omega,\mathcal{F},\mathbb{P})$ be a probability space. A *random measure* $M$ on $(S,\mathcal{A})$ is a collection of random variables $(M(B),B\in\mathcal{A})$ such that

(i) $M(\varnothing)=0$.

(ii) Given any sequence $(A_n)_{n\in\mathbb{N}}$ of mutually disjoint sets in $A$,

$$M\left(\bigcup_{n=1}^{\infty}A_n\right)=\sum_{n=1}^{\infty}M(A_n).$$

(iii) (independently scattered) For each disjoint family $(B_1,\ldots,B_n)$ in $\mathcal{A}$ the random variables $M(B_1),\ldots,M(B_n)$ are independent.

A special class of random measures is the class of martingae-valued measures.

**Definition 4.3.15.** *Let $E$ be a topological space equipped with its Borel $\sigma$-algebra $\mathcal{B}(E)$. Let $M$ be a random measure on $\mathbb{R}^+\times E$. For each $A\in\mathcal{B}(E)$, define a process $M^A=(M_t^A)_{t\geq 0}$ by $M_t^A=M([0,t],A)$. We say that $M$ is a martingale-valued measure if there exists a set $V\in\mathcal{B}(E)$ such that for all $A\in\mathcal{B}(E)$ with $\overline{A}\cap V=\varnothing$ we have that $M^A$ is a martingale. We call the set $V$ the associated forbidden set.*

We will only work in this thesis with sets $E\subset\mathbb{R}$ equipped with the standard Euclidean topology. Given a filtered probability space $(\Omega,\mathcal{F},\mathbb{F},\mathbb{P})$, we further specify martingale-valued measure to have $(2,\rho)$-*type* if the following three conditions hold

**(M1):** For all $A\in\mathcal{B}(E)$ we have $M(\{0\},A)=0$ almost surely;

**(M2):** for all $0\leq s<t<\infty$ and $A\in\mathcal{B}(E)$ we have that $M((s,t],A)$ is independent of $\mathcal{F}_s$;

**(M3):** There exists a $\sigma$-finite measure $\rho$ on $\mathcal{B}(\mathbb{R}^+)\otimes\mathcal{B}(E)$ for which

$$\mathbb{E}(M(t,A)^2)=\rho(t,A):=\rho((0,t],A).$$

Let $M$ be a $(2,\rho)$-type martingale-valued measure and fix $T>0$, then we define $L^2_M$ to be the vector space of all equivalence classes of predictable mappings $U:[0,T]\times E\times\Omega\to\mathbb{R}$ which coincide $\rho$-almost everywhere, almost surely and satisfy

$$\mathbb{E}\left(\int_0^T\int_E|U_t(x)|^2\,\rho(dt,dx)\right)<\infty.$$

We define the inner product $(U,V)_{L^2_M}$ on $L^2_M$ by

$$(U,V)_{L^2_M}=\mathbb{E}\left(\int_0^T\int_E U_t(x)V_t(x)\,\rho(dt,dx)\right),$$

which induces a norm $\|\cdot\|_{L^2_M}$.

**Lemma 4.3.16.** *$L^2_M$ equipped with the inner product $(\cdot,\cdot)_{L^2_M}$ is a Hilbert space.*

*Proof.*
Clearly $L^2_M$ is a subspace of $L^2([0,T]\times E\times\Omega,\Sigma_p\otimes\mathcal{B}(E),\rho\otimes\mathbb{P})$. We only need to show it is closed. Let $(U^n)_{n\in\mathbb{N}}$ be a sequence in $L^2_M$ converging to $U\in L^2$. By Markov's inequality we have $U_n\xrightarrow{\mathbb{P}}U$, so there exists a subsequence converging to $U$ almost surely. Since the subsequence consists of predictable mappings, the almost sure limit is predictable as well. Hence $U\in L^2_M$. □

We will only work with $(2,\rho)$-type martingale-valued measures such that the corresponding $\rho$ is of the form $\rho(t,A) = t\nu(A)$ where $\nu$ is a $\sigma$-finite measure on $\mathcal{B}(E)$, so from now on we will assume this to be the case.

Let $\mathscr{S}(E)$ be the space of simple processes $U : [0,T] \times E \times \Omega \to \mathbb{R}$ of the following form. For some $m,n \in \mathbb{N}$ there exists $0 \le t_1 \le t_2 \le \cdots \le t_{m+1} = T$ and disjoint Borel subsets $A_1, \dots, A_n$ of $E$ with $\mu(A_i) < \infty$, such that

$$U = \sum_{j,k=1}^{m,n} U_{t_j}^k \mathbf{1}_{(t_j,t_{j+1}]} \mathbf{1}_{A_k},$$

where for $j = 1,\dots,m$ and $k = 1,\dots,n$ we have $U_{t_j}^k = c_k U_{t_j}$ with $c_k \in \mathbb{R}$ and $U_{t_j}$ a bounded $\mathcal{F}_{t_j}$-measurable random variable. Note that $U$ is left-continuous and $\mathcal{B}(E)\otimes\mathcal{F}_t$-measurable, hence predictable. We will not prove the following lemma, as the proof digresses too far from the topic of this thesis, instead we refer to Applebaum [1, Lemma 4.1.4].

**Lemma 4.3.17.** *The space $\mathscr{S}(E)$ endowed with the $L^2_M$-norm is dense in $L^2_M$.*

We define the stochastic integral

$$U \cdot M = \sum_{j,k=1}^{m,n} U_{t_j}^k M((t_j,t_{j+1}], A_k).$$

It is not hard to see that this definition is linear in $U$, and we furthermore have an isometry on $\mathscr{S}(E)$.

**Lemma 4.3.18.** *Let $U \in \mathscr{S}(E)$. Then*

$$\mathbb{E}(U\cdot M) = 0, \qquad \mathbb{E}((U\cdot M)^2) = \|U\|_{L^2_M}.$$

*Proof.*
By the martingale property of $M$ for each $j = 1,\dots,m$ and $k = 1,\dots,n$ we have $\mathbb{E}(M((t_j,t_{j+1}],A_k)) = 0$. Hence by linearity and the fact that $M((t_j,t_{j+1}],A_k)$ is independent of $\mathcal{F}_{t_j}$ we have that

$$\mathbb{E}(U\cdot M) = \sum_{j,k=1}^{m,n} \mathbb{E}(U_{t_j}^k)\mathbb{E}(M((t_j,t_{j+1}],A_k)) = 0.$$

For the second moment, note that by linearity we again have

$$\mathbb{E}((U\cdot M)^2) = \sum_{j,k=1}^{m,n}\sum_{l,p=1}^{m,n} \mathbb{E}(U_{t_j}^k M((t_j,t_{j+1}]A_k) U_{t_l}^p M((t_l,t_{l+1}],A_p)$$

Now by the martingale property of $M$ and iterated conditioning, most terms vanish and we get

$$\begin{aligned}
&= \sum_{j,k=1}^{m,n}\sum_{p=1}^{n} \mathbb{E}(U_{t_j}^k U_{t_j}^p M((t_j,t_{j+1}],A_k) M((t_j,t_{j+1}],A_p) \\
&= \sum_{j,k=1}^{m,n}\sum_{p=1}^{n} \mathbb{E}(U_{t_j}^k U_{t_j}^p)\mathbb{E}(M((t_j,t_{j+1}],A_k) M((t_j,t_{j+1}],A_p)
\end{aligned}$$

Since the Borel sets $A_1,\dots,A_n$ are disjoint, we get by the scattering property of random measures and the martingale property of $M$ that

$$\begin{aligned}
&= \sum_{j,k=1}^{m,n} \mathbb{E}((U_{t_j}^k)^2)\mathbb{E}(M((t_j,t_{j+1}],A_k)^2) \\
&= \sum_{j,k=1}^{m,n} \mathbb{E}((U_{t_j}^k)^2)\mathbb{E}(M((0,t_{j+1}],A_k)^2 + M((0,t_j],A_k)^2 - 2M((0,t_j],A_k)M((0,t_{j+1}],A_k))
\end{aligned}$$

Again by the martingale property of $M$ and the fact that $M((0,t_{j+1}],A_k)$ is independent of $\mathcal{F}_{t_j}$.

$$
\begin{aligned}
&= \sum_{j,k=1}^{m,n} \mathbb{E}((U_{t_j}^k)^2)\mathbb{E}(M((0,t_{j+1}],A_k)^2 + M((0,t_j],A_k)^2) \\
&= \sum_{j,k=1}^{m,n} \mathbb{E}((U_{t_j}^k)^2)\rho((t_j,t_{j+1}],A_k) \\
&= \mathbb{E}\left(\int_0^T \int_E |U_t(x)|^2 \, \rho(dt,dx)\right).
\end{aligned}
$$

□

We can conclude that this stochastic integral is a linear isometry from $\mathscr{S}(E)$ into $L^2(\Omega,\mathcal{F},\mathbb{P})$ and by the density of $\mathscr{S}(E)$ in $L^2_M$, it extends to a linear isometry from $L^2_M$ into $L^2(\Omega,\mathcal{F},\mathbb{P})$. We will write for $U \in L^2_M$, the stochastic integral

$$U \cdot M = \int_0^T \int_E U_t(x)\, M(dt,dx),$$

defined through the extension. Just as with continuous integrators, it turns out these stochastic integrals are again square-integrable martingales.

We will need a local version of this type of stochastic integration. Define $L^2_{M,loc}$ as the set of all equivalence classes of indistinguishable predictable mappings $U : [0,T] \times E \times \Omega \to \mathbb{R}$ such that

$$\mathbb{P}\left(\int_0^T \int_E |U_t(x)|^2 \, \rho(dt,dx) < \infty\right) = 1.$$

Now we consider the topology on $L^2_{M,loc}$ generated by the sets

$$\left\{U \in L^2_{M,loc} : \mathbb{P}\left(\int_0^T \int_E |U_t(x) - V_t(x)|^2 \, \rho(dt,dx) < a\right) = 1\right\},$$

for $a > 0$ and $V \in L^2_{M,loc}$. It turns out this topology yields the convergence for sequences in $L^2_{M,loc}$ given by

$$\mathbb{P}\left(\lim_{n\to\infty} \int_0^T \int_E |U^n_t(x) - U_t(x)|^2 \, \rho(dt,dx) = 0\right) = 1.$$

Again $\mathscr{S}(E)$ is dense in $L^2_{M,loc}$ endowed with this topology on both spaces and we have the following estimate

**Lemma 4.3.19.** *If $U \in \mathscr{S}(E)$, then for all $C, K \geq 0$*

$$\mathbb{P}\left(\left|\int_0^T \int_E U_t(x)\, M(dt,dx)\right| > C\right) \leq \frac{K}{C^2} + \mathbb{P}\left(\int_0^T \int_E |U_t(x)|^2 \, \rho(dt,dx) > K\right).$$

*Proof.*
Fix $K > 0$ and define $\widetilde{U}^{(K)}$ by

$$\widetilde{U}^{p,(K)}_{t_j} = \begin{cases} U^p_{t_j} & \text{if } \sum_{i,l=1}^{j,p} (U^l_{t_i})^2 \rho((t_i,t_{i+1}],A_l) \leq K, \\ 0 & \text{otherwise.} \end{cases}$$

Then again $\widetilde{U}^{(K)} \in \mathscr{S}(E)$ and

$$\int_0^T \int_E |\widetilde{U}^{(K)}_t(x)|^2 \, \rho(dt,dx) = \sum_{i=1}^{m_K} \sum_{l=1}^{n_K} (U^l_{t_i})^2 \rho((t_i,t_{i+1}],A_l),$$

where $m_K$ and $n_K$ are the largest indices for which

$$\sum_{i=1}^{m_K}\sum_{l=1}^{n_K}(U^l_{t_i})^2\rho((t_i,t_{i+1}],A_l)\leq K.$$

By construction, we have that $U=\widetilde{U}^{(K)}$ if and only if

$$\int_0^T\int_E |U_t(x)|^2\,\rho(dt,dx)\leq K;$$

then by Markov's inequality, we have

$$\begin{aligned}\mathbb{P}\left(\left|\int_0^T\int_E U_t(x)\,M(dt,dx)\right|>C\right)&=\mathbb{P}\left(\left|\int_0^T\int_E \widetilde{U}^{(K)}_t(x)\,M(dt,dx)\right|>C\right)+\mathbb{P}(U\neq\widetilde{U}^{(K)})\\&\leq\frac{\mathbb{E}((F\cdot M)^2)}{C^2}+\mathbb{P}\left(\int_0^T\int_E|U_t(x)|^2\,\rho(dt,dx)>K\right)\\&\leq\frac{K}{C^2}+\mathbb{P}\left(\int_0^T\int_E|U_t(x)|^2\,\rho(dt,dx)>K\right),\end{aligned}$$

as required. □

Now let $U\in L^2_{M,loc}$, then we can find a sequence $U^n\in\mathscr{S}(E)$ such that almost surely we have

$$\lim_{n\to\infty}\int_0^T\int_E|U^n_t(x)-U_t(x)|^2\,\rho(dt,dx)=0.$$

Hence we also have this convergence in probability, and therefore it is Cauchy in probability. By Lemma 4.3.19 for any $m,n\in\mathbb{N}$ and $M,\beta>0$ we have,

$$\mathbb{P}\left(\left|\int_0^T\int_E(U^n_t(x)-U^m_t(x))\,M(dt,dx)\right|>\beta\right)\leq\frac{M}{\beta^2}+\mathbb{P}\left(\int_0^T\int_E|U^n_t(x)-U^m_t(x)|^2\,\rho(dt,dx)>M\right).$$

Given $\varepsilon>0$, we can find for any $\gamma>0$ an $N\in\mathbb{N}$ such that for $n,m\geq N$ we have

$$\mathbb{P}\left(\int_0^T\int_E|U^n_t(x)-U^m_t(x)|^2\,\rho(dt,dx)>\gamma\beta^2\right)<\varepsilon.$$

Now pick $M=\gamma\beta^2$, and we can see that $(U^n\cdot M)_{n\in\mathbb{N}}$ is Cauchy in probability and thus has a unique limit in probability almost surely. This limit will be the stochastic integral and we define

$$U\cdot M=\int_0^T\int_E U_t(x)\,M(dt,dx),$$

such that

$$\int_0^T\int_E U^n_t(x)\,M(dt,dx)\xrightarrow{\mathbb{P}}\int_0^T\int_E U_t(x)\,M(dt,dx).$$

**Theorem 4.3.20.** *Let $U\in L^2_{M,loc}$ and $M$ a $(2,\rho)$-type martingale-valued measure, then*

1. *$(U\cdot M)_{t\geq0}$ is a local martingale.*
2. *If the filtration satisfies the usual hypothesis, $(U\cdot M)_{t\geq0}$ has a càdlàg modification.*

*Proof.*
Define a sequence of stopping times $(\tau_n)_{n\in\mathbb{N}}$ by

$$\tau_n=\inf\left\{t\geq0:\int_0^t\int_E|U_t(x)|^2\,\rho(dt,dx)>n\right\}.$$

Then $\tau_n \xrightarrow{a.s.} \infty$. Now for all $x \in E, t \geq 0, n \in \mathbb{N}$, we have

$$\int_0^t \int_E |U_t^{\tau_n}(x)|^2 \, \rho(dt, dx) \leq n,$$

hence $U^n \in L^2_M$, so we know that $U^n \cdot M$ is a $L^2$-martingale, but also $(U^n \cdot M)_t = (U \cdot M)_{t \wedge \tau_n}$.

Now by the Itô isometry we have that $U^n \cdot M$ is a $L^2$-martingale, but $(U^n \cdot M)_t = (U \cdot M)_{t \wedge \tau_n}$, such that $U \cdot M$ is a local martingale.

For the càdlàg modification, note that every $(U \cdot M)^{\tau_n}$ is a martingale. It is a well-known fact that, for filtrations satisfying the usual hypothesis, every supermartingale $X$, such that $t \mapsto \mathbb{E}(X_t)$ is right-continuous, has a càdlàg modification, Dellacherie and Meyer [13, page 69]. For martingales this is trivially true as the mapping $t \mapsto \mathbb{E}(X_t)$ is constant. So for every $n \in \mathbb{N}$ we have that $(U \cdot M)^{\tau_n}$ has a càdlàg modification. Now since $(\tau_n)_{n \in \mathbb{N}}$ is increasing, for each $\omega \in \Omega$, and $t_0 \geq 0$, we can find $n_0(\omega) \in \mathbb{N}$ such that $\tau_{n_0}(\omega) > t_0$, so then $(U \cdot M)^{\tau_n}_{t_0} = (U \cdot M)_{t_0}$. Then it follows that the càdlàg property carries over. □

The most important example of a martingale-valued measure in this thesis is the compensated Poisson random measure. Let $(S, \mathcal{A})$ be some measurable space with a $\sigma$-finite measure $\nu$. We define a Poisson random measure with *intensity measure* $\nu$ to be a random measure such that for all $A \in \mathcal{A}$ with $\overline{A}$ lying outside the forbidden set, we have that $(N(t, A))_{t \geq 0}$ is a Lévy process (see Definition 4.4.6 in the next section) which is Poisson$(t\nu(A))$ distributed for each $t \geq 0$. The existence of a Poisson random measure is shown in Sato [37, Proposition 19.4]. Now let $(S, \mathcal{A}) = (\mathbb{R}_0, \mathcal{B}(\mathbb{R}_0))$, with forbidden set $\{0\}$, then we define the compensated Poisson random measure to be $\widetilde{N}(t, A) = N(t, A) - t\nu(A)$ for each $t \geq 0$ and $A \in \mathcal{B}(\mathbb{R}_0)$, with $0 \notin \overline{A}$. Since $(\widetilde{N}(t, A))_{t \geq 0}$ is a martingale, $\widetilde{N}$ extends to a martingale-valued measure with forbidden set $\{0\}$. Furthermore $\widetilde{N}$ is of $(2, \rho)$-type with $\rho(t, A) = t\nu(A)$. We call a set $A \in \mathcal{B}(\mathbb{R}_0)$ such that $0 \notin \overline{A}$ to be *bounded from below.*

We have already done most of the hard work in defininig a stochastic integral with respect to martingale-valued measures. An important approximation theorem which is called the 'interlacing construction' in Applebaum [1, Theorem 4.3.4], is the following.

**Theorem 4.3.21.** *Let $\widetilde{N}$ be a compensated Poisson random measure, with intensity measure $\nu$.*

*(i) Let $U \in L^2_{\widetilde{N},loc}$. For every sequence $(A_n)_{n \in \mathbb{N}}$ in $\mathcal{B}(E)$ with $A_n \uparrow E$, we have*

$$\int_0^T \int_{A_n} U_t(x) \, \widetilde{N}(dt, dx) \xrightarrow{\mathbb{P}} \int_0^T \int_E U_t(x) \, \widetilde{N}(dt, dx).$$

*(ii) Let $U \in L^2_{\widetilde{N}}$, then there exists a sequence $(A_n)_{n \in \mathbb{N}}$ in $\mathcal{B}(E)$ with each $\nu(A_n) < \infty$ and $A_n \uparrow E$ as $n \to \infty$ for which*

$$\int_0^T \int_{A_n} U_t(x) \, \widetilde{N}(dt, dx) \to \int_0^T \int_E U_t(x) \, \widetilde{N}(dt, dx),$$

*with uniform convergence almost surely.*

*Proof.*
It is possible to extend Lemma 4.3.19 to the entire $L^2_{M,loc}$ by density of $\mathscr{S}(E)$. Then we have for any $\delta, \varepsilon > 0, n \in \mathbb{N}$, that

$$\begin{aligned}\mathbb{P}\left(\left|\int_0^T \int_E U_t(x) \, \widetilde{N}(dt, dx) - \int_0^T \int_{A_n} U_t(x) \, \widetilde{N}(dt, dx)\right| > \varepsilon\right) \\ \leq \frac{\delta}{\varepsilon^2} + \mathbb{P}\left(\int_0^T \int_{E \setminus A_n} |U_t(x)|^2 \, d\nu(x) \, dt > \delta\right),\end{aligned}$$

so that (1) follows.

Now for (2) define a sequence $(\varepsilon_n)_{n \in \mathbb{N}}$ that decreases monotonically to zero, with $\varepsilon_1 = 1$ and, for $n \geq 2$,

$$\varepsilon_n = \sup\left\{y \geq 0 : \mathbb{E}\left(\int_0^T \int_{0<|x|<y} |U_t(x)|^2 \, d\nu(x) \, dt\right) \leq 8^{-n}\right\}.$$

Now define $A_n = \{x \in E : |x| > \varepsilon_n\}$ for each $n \in \mathbb{N}$. By Doob's martingale inequality, for each $n \in \mathbb{N}$, we have

$$\begin{aligned}
\mathbb{E}&\left(\sup_{s\le t}\left|\int_0^s\int_{A_{n+1}} U_u(x)\widetilde{N}(du,dx) - \int_0^s\int_{A_n} U_u(x)\widetilde{N}(du,dx)\right|^2\right)\\
&\le \mathbb{E}\left(\left|\int_0^t\int_{A_{n+1}\setminus A_n} U_u(x)\widetilde{N}(du,dx)\right|^2\right)\\
&= \int_0^t\int_{A_{n+1}\setminus A_n} \mathbb{E}(|U_u(x)|^2)\,d\nu(x)\,du \le 8^{-n}
\end{aligned}$$

Now by Markov's inequality

$$\mathbb{P}\left(\sup_{t\le T}\left|\int_0^s\int_{A_{n+1}} U_u(x)\widetilde{N}(du,dx) - \int_0^s\int_{A_n} U_u(x)\widetilde{N}(du,dx)\right| \ge \frac{1}{2^n}\right) \le \frac{1}{2^n}.$$

Hence by the Borel-Cantelli lemmas we have that

$$\mathbb{P}\left(\limsup_{n\to\infty}\sup_{t\le T}\left|\int_0^s\int_{A_{n+1}} U_u(x)\widetilde{N}(du,dx) - \int_0^s\int_{A_n} U_u(x)\widetilde{N}(du,dx)\right| < \frac{1}{2^n}\right) = 1.$$

Therefore given $\delta > 0$, there exists $N \in \mathbb{N}$, such that for $m, n > N$ we have

$$\begin{aligned}
&\sup_{t\le T}\left|\int_0^s\int_{A_n} U_u(x)\widetilde{N}(du,dx) - \int_0^s\int_{A_m} U_u(x)\widetilde{N}(du,dx)\right|\\
&\quad\le \sum_{k=m}^{n-1}\sup_{t\le T}\left|\int_0^s\int_{A_{k+1}} U_u(x)\widetilde{N}(du,dx) - \int_0^s\int_{A_k} U_u(x)\widetilde{N}(du,dx)\right|\\
&\quad< \sum_{k=m}^{n-1}\frac{1}{2^k} < \delta,
\end{aligned}$$

almost surely. Hence the convergence is almost surely uniformly Cauchy on compact intervals and therefore almost surely uniformly convergent on compact intervals. □

To conclude this section we give an example why predictability is necessary instead of just progressive measurability.

*Example* 4.3.22. Let $N$ be a Poisson process with intensity 1 and let $\widetilde{N}_t = N_t - t$. Then we know that $\widetilde{N}$ is a martingale. Assuming that the filtration is complete, we know that $N$ is progressive since it is adapted by Proposition 2.3.5. But it is not predictable, this is however rather difficult to prove, see for example Davis [12, A3.7]. The process $(N_{t-})_{t\ge0}$ is left-continuous and adapted and therefore by definition predictable.

Define $\Delta N_t = N_t - N_{t-}$. It turns out that we have

$$\int_0^t\int_{\mathbb{R}_0} \Delta N_s\,\widetilde{N}(ds,dJ) = \sum_{0\le s\le t}\Delta N_s^2 = N_t,$$

see Applebaum [1, page 207], hence we have

$$\begin{aligned}
\int_0^t\int_{\mathbb{R}_0} N_s\,\widetilde{N}(ds,dJ) &= \int_0^t\int_{\mathbb{R}_0} N_{s-}\,\widetilde{N}(ds,dJ) + \int_0^t\int_{\mathbb{R}_0} \Delta N_s\,\widetilde{N}(ds,dJ)\\
&= \int_0^t\int_{\mathbb{R}_0} N_{s-}\,\widetilde{N}(ds,dJ) + N_t.
\end{aligned}$$

But $N$ is not even a local martingale, so the stochastic integral $N \cdot \widetilde{N}$ is not a local martingale either. Therefore we see that predictability is necessary to make sure the stochastic integral is still a local martingale.

## 4.4 Itô's formula

Having built a theory of integration with respect to continuous semimartinagles and compensated Poisson measures, we lack only one very important ingredient of stochastic calculus. We have defined a notion of integration for stochastic processes, however we also want to look at the evolution or flow of a stochastic process through time to build differential equations.

For deterministic processes, we have a very satisfying theory of differentiation, it would be convenient if we could extend this notion to stochastic processes. However if we look at the paths of for example Brownian motion, then we know they are almost surely continuous, but wildly irregular, in fact they are nowhere differentiable almost surely, Billingsley [4, page 504-505]. This means that there is no hope of extending the classical theory of differentiation. Despite this setback, we can define a kind of stochastic differentiation through the stochastic integrals we have just defined. A stochastic analogue of the chain rule is known as Itô's formula, and is one of the most important theorems in stochastic calculus, giving a very satisfying theory of 'differentiation'.

Before we go into the details of the Itô formula, we introduce the class of Lévy processes, which will reappear briefly later in this thesis. The classes of Lévy processes and more generally of additive processes have been extensively studied in for example Sato [37], and Applebaum [1], and they satisfy a couple of nice properties. They have also been heavily researched in finance for modelling market incompleteness and heavy-tailed distributions, in for example Cont and Tankov [11].

Let $\mathcal{M}^+(\mathbb{R})$ denote the set of all finite positive Borel measures on $\mathbb{R}$. We define the *convolution* of two measures in $\mu_1, \mu_2 \in \mathcal{M}^+(\mathbb{R})$ as follows

$$(\mu_1 * \mu_2)(A) = \int_{\mathbb{R}} \mu_1(A - x)\, d\mu(x), \qquad A \in \mathcal{B}(\mathbb{R}).$$

It follows immediately that $(\mu_1 * \mu_2)(\mathbb{R}) \leq \mu_1(\mathbb{R})\mu_2(\mathbb{R}) < \infty$, so the convolution operation defines a binary operation on $\mathcal{M}^+(\mathbb{R})$. Furthermore we have the following properties which follow almost immediately from the definition.

**Proposition 4.4.1.** *If $f \in BM(\mathbb{R})$ (the space of bounded $(\mathcal{B}(\mathbb{R}), \mathcal{B}(\mathbb{R}))$-measurable functions), then for all $\mu_i \in \mathcal{M}^+(\mathbb{R}), i = 1, 2, 3$,*

*(i)*
$$\int_{\mathbb{R}} f(y)\, d(\mu_1 * \mu_2)(y) = \int_{\mathbb{R}} \int_{\mathbb{R}} f(x + y)\, d\mu_1(y)\, d\mu_2(x)$$

*(ii)*
$$\mu_1 * \mu_2 = \mu_2 * \mu_1$$

*(iii)*
$$(\mu_1 * \mu_2) * \mu_3 = \mu_1 * (\mu_2 * \mu_3)$$

The importance of convolutions in probability theory is in the following corollary, where we characterise the sum of two random variables through the convolution of their distributions.

**Corollary 4.4.2.** *For each $f \in BM(\mathbb{R})$, we have*

$$\mathbb{E}(f(X_1 + X_2)) = \int_{\mathbb{R}} f(z)\, d(\mu_{X_1} * \mu_{X_2})(z).$$

Let $X$ be a random variable taking values in $\mathbb{R}$ with law $\mu_X$. We say that $X$ is *infinitely divisible* if, for all $n \in \mathbb{N}$, there exists independent and identically distributed random variables $Y_1^{(n)}, \ldots, Y_n^{(n)}$ such that

$$X \overset{d}{=} Y_1^{(n)} + \ldots + Y_n^{(n)}.$$

Furthermore, let $\phi_\mu(s)$ denote the characteristic function of the probability measure $\mu$, where $u \in \mathbb{R}$. Then we define

$$\phi_\mu(\xi) = \int_{\mathbb{R}} e^{i\xi x}\, d\mu(x).$$

We define $\phi_X(\xi) = \phi_{\mu_X}(\xi)$, where $\mu_X$ is the distribution of $X$. If $X$ is infinitely divisible, then for each $n \in \mathbb{N}$, we have by independence

$$\mathbb{E}(\exp(i\xi X)) = \mathbb{E}\left(\prod_{i=1}^{n} \exp(i\xi Y_i^{(n)})\right) = \prod_{i=1}^{n} \mathbb{E}\left(\exp(i\xi Y_i^{(n)}\right).$$

Since the $Y_i^{(n)}$ are identically distributed, we can define the characteristic function of the $n$-th root of $X$ as $\phi_X^{1/n}(\xi) = \mathbb{E}(\exp(i\xi Y_1^{(n)}))$ such that $\left(\phi_X^{1/n}(\xi)\right)^n = \phi_X(\xi)$. Furthermore let $\mu_X^{1/n}$ be the distribution of $Y_1^{(n)}$, and define for a measure $\nu \in \mathcal{M}^+(\mathbb{R})$ the $n$-th convolution with itself $\nu^n = \nu * \nu * \cdots * \nu$, then also $\left(\mu_X^{1/n}\right)^n = \mu_X$.

*Example* 4.4.3 (Poisson random variables)*.* Consider a random variable $X \sim \text{Poisson}(\lambda)$. Then

$$\mathbb{P}(X = n) = \frac{\lambda^n}{n!} e^{-\lambda}.$$

It is not hard to show that

$$\phi_X(s) = \exp\left(\lambda(e^{is} - 1)\right),$$

from which we can deduce that $X$ is infinitely divisible with each $Y_j^{(n)} \sim \text{Poisson}\left(\frac{\lambda}{n}\right)$.

*Example* 4.4.4*.* Suppose that $(Z_n)_{n\in\mathbb{N}}$ is a sequence of independent and identically distributed random variables taking values in $\mathbb{R}$ with common law $\mu_Z$, further let $N \sim \text{Poisson}(\lambda)$ be a Poisson random variable. The *compound Poisson* random variable $X$ is defined as

$$X = \sum_{n=1}^{N} Z_n.$$

We say that $X \sim \text{CPoisson}(\lambda, \mu_Z)$, with intensity $\lambda$ and child distribution $\mu_Z$. Now let $\xi \in \mathbb{R}$, we have

$$\begin{aligned}\phi_X(\xi) &= \sum_{n=0}^{\infty} \mathbb{E}\left(\exp\left(i\xi \sum_{i=1}^{n} Z_i\right) \middle| N = n\right) \mathbb{P}(N = n) \\ &= e^{-\lambda} \sum_{n=0}^{\infty} \frac{(\lambda\phi_Z(\xi))^n}{n!} = \exp\left(\lambda \int_{\mathbb{R}} \left(e^{i\xi z} - 1\right) d\mu_Z(z)\right).\end{aligned}$$

It follows again that $X$ is infinitely divisible with each $Y_j^{(n)} \sim \text{CPoisson}\left(\frac{\lambda}{n}, \mu_Z\right)$.

The class of compound Poisson distributions constitute almost the entire class of infinitely divisible distributions.

**Theorem 4.4.5.** *The set of all infinitely divisible probability measures on $\mathbb{R}$ coincides with the weak closure of the set of all compound Poisson distributions on $\mathbb{R}$ (closure under convergence in distribution).*

*Proof.*
Let $\phi$ be the characteristic function of an arbitrary infinitely divisible probability measure $\mu$, so that $\phi^{1/n}$ is the characteristic function of $\mu^{1/n}$; then for each $n \in \mathbb{N}$, $\xi \in \mathbb{R}$, we may define

$$\phi_n(\xi) = \exp\left(n\left(\phi^{1/n}(\xi) - 1\right)\right) = \exp\left(n \int_{\mathbb{R}} (e^{i\xi x} - 1)\, d\mu^{1/n}(x)\right),$$

so that $\phi_n$ is the characteristic function of a compound Poisson distribution. We then have

$$\begin{aligned}\phi_n(\xi) &= \exp\left(n(e^{(1/n)\text{Log}(\phi(\xi))} - 1)\right) \\ &= \exp\left(\text{Log}(\phi(\xi)) + \mathcal{O}\left(\frac{1}{n}\right)\right) \overset{n\to\infty}{\longrightarrow} \phi(\xi),\end{aligned}$$

with Log the principal value of the logarithm. It is well known that the pointwise convergence of the characteristic functions is equivalent to convergence in distribution. Since all compound Poisson distributed random variables are infinitely divisible, the result follows. □

**Definition 4.4.6.** *Let $X$ be an adapted stochastic process defined on a probability space $(\Omega, \mathcal{F}, , \mathbb{F}, \mathbb{P})$. We say that $X$ is a Lévy process if*

**(L1):** $X_0 = 0$ *almost surely.*

**(L2):** *$X$ has independent increments, for every $n \in \mathbb{N}$ and $0 \le t_0 < t_1 < \cdots < t_n$, the random variables $X_{t_0}, X_{t_1} - X_{t_0}, X_{t_2} - X_{t_1}, \cdots, X_{t_n} - X_{t_{n-1}}$ are independent.*

**(L3):** *$X$ has stationary incremements, for every $s,t \geq 0$ we have that $X_{s+t} - X_s \stackrel{d}{=} X_t$.*

**(L4):** *$X$ is stochastically continuous, for every $\varepsilon$ we have for all $t \geq 0$, that $\lim_{s\downarrow t} \mathbb{P}(|X_t - X_s| > \varepsilon) = 0$.*

When $X$ only satisfies (L1), (L2) and (L4), then we call $X$ an *additive process.* It turns out that properties (L2) and (L3) imply that every Lévy process has a càdlàg modification, Protter [33, page 21]. Therefore we will always assume that Lévy processes are càdlàg from now on.

The following proposition is also true for additive processes, the proof is however less straightforward, Sato [37, page 47].

**Proposition 4.4.7.** *If $X$ is a Lévy process, then $X_t$ is infinitely divisible for each $t \geq 0$.*

*Proof.*
Let $n \in \mathbb{N}$ and $t \geq 0$ and define the random variables

$$Y_{k,t}^{(n)} = X_{\frac{kt}{n}} - X_{\frac{(k-1)t}{n}}.$$

Then we have that $X_t = Y_{1,t}^{(n)} + \ldots + Y_{n,t}^{(n)}$ where the $Y_{k,t}^{(n)}$ are independent and identically distributed. □

The properties of Lévy processes and also additive processes allow us to say a great deal about their characteristic functions, these families can actually be parameterised entirely by only three parameters, the characteristic triplet $(\mu, \sigma, \nu)$ for Lévy processes and the characteristic triplet $(\mu_t, \sigma_t, \nu_t)$ for additive processes.

**Lemma 4.4.8.** *If $X = (X_t)_{t\geq 0}$ is a stochastically continuous process, then the map $t \mapsto \phi_{X_t}(x)$ is continuous for each $x \in \mathbb{R}$.*

*Proof.*
Fix $\xi \in \mathbb{R}$. Since the map $x \mapsto \exp(i\xi x)$ is continuous at the origin, given any $\varepsilon > 0$ we can find $\delta_1 > 0$ such that

$$\sup_{0\leq |x| < \delta_1} |\exp(i\xi x) - 1| < \frac{\varepsilon}{2},$$

and by stochastic continuity, we can find $\delta_2 > 0$ such that whenever $0 < |t-s| < \delta_2$, we have $\mathbb{P}(|X_s - X_t| > \delta_1) < \varepsilon/4$. Hence for all $0 < |t-s| < \delta_2$ we have

$$\begin{aligned}
|\phi_{X_t}(\xi) - \phi_{X_s}(\xi)| &= \left| \int_\Omega e^{i\xi X_s(\omega)} \left( e^{i\xi(X_s(\omega) - X_t(\omega))} - 1 \right) d\mathbb{P}(\omega) \right| \\
&\leq \int_\mathbb{R} |e^{i\xi x} - 1| \, d\mu_{X_s - X_t}(x) \\
&= \int_{B_{\delta_1}(0)} |e^{i\xi x} - 1| \, d\mu_{X_s - X_t}(x) + \int_{B_{\delta_1}(0)^c} |e^{i\xi x} - 1| \, d\mu_{X_s - X_t}(x) \\
&\leq \sup_{0 \leq |x| < \delta_1} |e^{i\xi x} - 1| + 2\mathbb{P}(|X_s - X_t| > \delta_1) \\
&< \varepsilon
\end{aligned}$$

and so the result follows. □

**Theorem 4.4.9.** *If $X$ is a Lévy process, then*

$$\phi_{X_t}(x) = e^{t\eta(x)}, \tag{4.2}$$

*for each $x \in \mathbb{R}, t \geq 0$, for some function $\eta : \mathbb{R} \to \mathbb{C}$.*

*Proof.*
Suppose that $X$ is a Lévy process, now define for each $x \in \mathbb{R}$ and $t \geq 0$ the function $\phi_x(t) = \phi_{X_t}(x)$. Then by the definition of Lévy processes we have for all $s \geq 0$

$$\begin{aligned}
\phi_x(t+s) &= \mathbb{E}\left(e^{i(x, X_{t+s})}\right) = \mathbb{E}\left(e^{i(x, X_{t+s} - X_s)} e^{i(x, X_s)}\right) \\
&= \mathbb{E}\left(e^{i(x, X_{t+s} - X_s)}\right) \mathbb{E}\left(e^{i(x, X_s)}\right) = \phi_x(t)\phi_x(s).
\end{aligned}$$

Furthermore, we know that $\phi_x(0) = 1$, to prove the theorem we have to solve this functional equation.

Firt suppose that $\alpha \geq 0$ is such that $\phi_x(\alpha) = 0$, then for all $t \geq 0$ we have that $\phi_x(t) = \phi_x(\alpha+(t-\alpha)) = \phi_x(\alpha)\phi_x(t-\alpha) = 0$, but then $\phi_x \equiv 0$, which contradicts the result $\phi_x(0) = 1$. Hence, we know that $\phi_x(t) \neq 0$ for all $t \geq 0$.

Now let $n \in \mathbb{N}$, we know that $\phi_x(n) = \phi_x\left(\sum_{i=1}^n 1\right) = \phi_x(1)^n$, next take $p, q \in \mathbb{N}$, then we have that

$$\phi_x(1)^p = \phi_x(p) = \phi_x\left(\sum_{i=1}^q \frac{p}{q}\right) = \phi_x\left(\frac{p}{q}\right)^q.$$

So then for all $r \in \mathbb{Q}^+$ we have that $\phi_x(r) = \phi_x(1)^r$. By Lemma 4.4.8 we know that $\phi_x(t) = \phi_x(1)^t$ for all $t \geq 0$, since $\mathbb{Q}^+$ is dense $\mathbb{R}^+$.

Let Log be the principal branch of the logarithm and define $\eta(x) = \mathrm{Log}(\phi_x(1))$. Now for all $t \geq 0$ we get the result

$$\phi_x(t) = \phi_x(1)^t = \left(e^{\eta(x)}\right)^t = e^{t\eta(x)}.$$

□

The function $\eta$ as defined in (4.2) is called the Lévy symbol of $X$. It turns out we can actually say a great deal about $\eta$.

Let $\nu$ be a Borel measure defined on $\mathbb{R}_0$. We say that it is a *Lévy measure* if

$$\int_{\mathbb{R}_0} (|y|^2 \wedge 1)\, d\nu(y) < \infty.$$

Intuitively the Lévy measure measures the amount Equivalently we can define $\nu$ on $\mathbb{R}$ with $\nu(\{0\}) = 0$ as is sometimes done in the literature. Now let $\widetilde{N}$ be a compensated Poisson random measure with intensity measure $\nu$, being a Lévy measure. Then

$$\mathbb{E}\left(\left(\int_0^t \int_{|x|\leq 1} x\, \widetilde{N}(dt, dx)\right)^2\right) = \mathbb{E}\left(t \int_{|x|\leq 1} x^2\, d\nu(x)\right) < \infty.$$

**Theorem 4.4.10** (Lévy-Khintchine representation)**.** *Let $X$ be a Lévy process, then there exists $\mu \in \mathbb{R}, \sigma > 0$ and a Lévy measure $\nu$ such that*

$$\phi_{X_t}(\xi) = \exp\left(\xi\mu t - \frac{1}{2}\xi^2\sigma^2 t + t\int_{\mathbb{R}_0} \left(e^{i\xi x} - 1 - i\xi x \mathbf{1}_{|x|\leq 1}\right) d\nu(x)\right).$$

*We call the triplet $(\mu, \sigma, \nu)$ the characterising triplet of the Lévy process. When $X$ is an additive process, the same holds, but the parameters $\mu, \sigma$ and $\nu$ may still depend on time under some restrictions and we say that $(\mu_t, \sigma_t, \nu_t)$ is the characterising triplet of the additive process.*

*Remark* 4.4.11. It follows from the Lévy-Khintchine representation that every Lévy process is the sum of four independent stochastic processes, a deterministic linear drift term, a Brownian motion term, a compound Poisson process term with intensity $\nu((-1,1)^c)$ and child distribution $\nu((-1,1)^c)^{-1}\nu|_{(-1,1)^c}$ modelling the big jumps larger than 1 and a $L^2$-martingale which models the jumps smaller than 1.

For 'stochastic differentiation' we will look at a class of stochastic processes not unlike additive processes, which we call *Lévy-Itô processes* as in Tankov [42, page 20]. Let $\mu$ be locally bounded, $\sigma \in L^2_{W,loc}$ and $\gamma \in L^2_{\widetilde{N},loc}$ such that

$$X_t = X_0 + \int_0^t \mu_s\, ds + \int_0^t \sigma_s\, dW_s + \int_0^t \int_{\mathbb{R}_0} \gamma_s(J)\, \widetilde{N}(ds, dJ). \tag{4.3}$$

Note that every Lévy-Itô process has a càdlàg modification by Theorem 4.3.20, so we will always assume it to be càdlàg. The proof of Itô's formula is roughly based on Le Gall [26, Theorem 5.10], Applebaum [1, Lemma 4.4.6] and Cont and Tankov [11, page 263].

**Theorem 4.4.12** (Itô-formula)**.** *Let $X$ be a Lévy-Itô processes and let $f \in C^{1,2}$. Then for every $t_0 \le t \le T$, we have*

$$f(t, X_t) = f(t_0, X_{t_0}) + \int_{t_0}^{t} \mathcal{L} f(s, X_{s-})\, ds + \int_{t_0}^{t} \frac{\partial f}{\partial x}(s, X_{s-}) \sigma_s \, dW_s + \int_{t_0}^{t} \int_{\mathbb{R}_0} \left(f(s, X_{s-} + \gamma_s(J)) - f(s, X_{s-})\right) \widetilde{N}(ds, dJ),$$

*where $\mathcal{L}$ is the second-order partial-integro differential operator*

$$\mathcal{L} f(s,x) = \frac{\partial f}{\partial t}(s,x) + \frac{\partial f}{\partial x}(s,x)\mu_s + \frac{1}{2}\frac{\partial^2 f}{\partial x^2}(s,x)\sigma_s^2 + \int_{\mathbb{R}_0} \left( f(s, x + \gamma_s(J)) - f(s,x) - \frac{\partial f}{\partial x}(s,x)\gamma_s(J) \right) d\nu(J).$$

*Furthermore $(f(t, X_t))_{t \ge 0}$ is again a Lévy-Itô process.*

*Proof.*
First let $X$ and $\gamma$ be bounded for each $0 \le s \le t$ and $J \in \mathbb{R}_0$, then $f'$ and $f''$ are bounded on the range of $X_{s-} + \gamma_s(J)$ and $X_{s-}$ by constants $K_1$ and $K_2$ respectively. Since $x \mapsto f(t,x)$ is in $C^2$, we have by Taylor's theorem that

$$|f(s, x + \gamma_s(J)) - f(s,x) - \frac{\partial f}{\partial x}(s,x)\gamma_s(J)| = \frac{1}{2}|f''(s, \xi(s))||\gamma_s(J)|^2 \le \frac{K_2}{2}|\gamma_s(J)|^2,$$

where $\xi(s)$ is between $x$ and $x + \gamma_s(J)$. Furthermore by the mean-value theorem we have

$$|f(s, X_{s-} + \gamma_s(J)) - f(s, X_{s-}| = |\gamma_s(J)| \left| \frac{\partial f}{\partial x}(s, \xi(s)) \right| \le K_1^2 |\gamma_s(J)|,$$

where $\xi(s)$ is between $X_{s-}$ and $X_{s-} + \gamma_s(J)$. Hence all the integrals are well-defined. Now define $A_n = \{x \in \mathbb{R}_0 : |x| > \varepsilon_n\}$ where

$$\varepsilon_n = \sup \left\{ y \ge 0 : \mathbb{E}\left( \int_0^t \int_{0<|x|<y} |\gamma_s(x)|^2 \, d\nu(x)\, ds \right) \le 8^{-n} \right\},$$

and define the processes

$$X_t^n = \int_0^t \mu_s \, ds + \int_0^t \sigma_s \, dW_s + \int_0^t \int_{A_n} \gamma_s(x) \, \widetilde{N}(ds, dx).$$

Each $A_n$ is bounded from below then fix $n \in \mathbb{N}$ and write for convenience $A = A_n, X = X^n$ and let $\tau_m^A$ be the arrival times of the jumps. Futher drop for now the dependence of $f$ on $t$, we have

$$\begin{aligned} f(X_t) - f(X_0) &= \sum_{m=0}^{\infty} \left( f(X_{t \wedge \tau_{m+1}^A}) - f(X_{t \wedge \tau_m^A}) \right) \\ &= \sum_{m=0}^{\infty} \left( f(X_{t \wedge \tau_{m+1}^A -}) - f(X_{t \wedge \tau_m^A}) \right) + \sum_{m=0}^{\infty} \left( f(X_{t \wedge \tau_{m+1}^A}) - f(X_{t \wedge \tau_{m+1}^A -}) \right). \end{aligned}$$

For the second sum we have that

$$\begin{aligned} \sum_{m=0}^{\infty} \left( f(X_{t \wedge \tau_{m+1}^A}) - f(X_{t \wedge \tau_{m+1}^A -}) \right) &= \sum_{m=0}^{\infty} \left( f(X_{t \wedge \tau_{m+1}^A -} + \gamma_{t \wedge \tau_{m+1}^A}(\Delta X_{t \wedge \tau_{m+1}^A})) - f(X_{t \wedge \tau_{m+1}^A -}) \right) \\ &= \int_0^t \int_A \left(f(s, X_{s-} + \gamma_s(J)) - f(s, X_{s-})\right) N(ds, dJ) \\ &= \int_0^t \int_A \left(f(s, X_{s-} + \gamma_s(J)) - f(s, X_{s-})\right) \widetilde{N}(ds, dJ) \\ &\quad + \int_0^t \int_A \left(f(s, X_{s-} + \gamma_s(J)) - f(s, X_{s-})\right) d\nu(J)\, ds \end{aligned}$$

So we only have to show that

$$\sum_{m=0}^{\infty}\Big(f(t\wedge\tau^A_{m+1},X_{t\wedge\tau^A_{m+1}-})-f(t\wedge\tau^A_m,X_{t\wedge\tau^A_m})\Big)$$
$$=\int_0^t\frac{\partial f}{\partial t}(s,X_{s-})\,ds+\int_0^t\frac{\partial f}{\partial x}(s,X_{s-})\,dX^c_s+\frac{1}{2}\int_0^t\frac{\partial^2 f}{\partial x^2}(s,X_{s-})\,d\langle X^c\rangle_s$$
$$-\int_0^t\int_A\gamma_s(x)\frac{\partial f}{\partial x}(s,X_{s-})\,d\nu(J)\,ds$$

Since $A$ was bounded from below, the sum actually contains only finitely many elements, and hence by the linearity of the integrals, we can treat each element of the sum seperately. Fix $m\in\mathbb{N}$, we know that $X_{s-}$ is continuous for $s\in(\tau^A_m,\tau^A_{m+1}]$.

Let $0<\delta<\tau^A_{m+1}-\tau^A_m$ arbitrary, consider a sequence of partitions $\tau^A_m+\delta=t^n_0<\dots<t^n_{p_n}=\tau^A_{m+1}$ and assume for convenience that $\tau^A_{m+1}\le t$, then we have

$$f(\tau^A_{m+1},X_{\tau^A_{m+1}-})-f(\tau^A_m+\delta,X_{\tau^A_m+\delta})=\sum_{k=0}^{p_n-1}(f(t^n_{k+1},X_{t^n_{k+1}-})-f(t^n_k,X_{t^n_k-}))$$

Now observe that $(t)_{t\ge0}$ is a finite variation process, so write $Y=(X^1,X^2)=(t,X)$. For every $k=0,\dots,p_n-1$, apply Taylor's theorem to the function

$$[0,1]\ni\theta\mapsto f(Y_{t^n_k}+\theta(Y_{t^n_{k+1}}-Y_{t^n_k})).$$

Then there exists some $\xi\in[0,1]$ such that for

$$\mathcal{R}^{i,j}_{n,k}=\frac{\partial^2 f}{\partial x^i\partial x^j}(Y_{t^n_k}+\xi(Y_{t^n_{k+1}}-Y_{t^n_k}),$$

where $i,j\in\{1,2\}$, we have that

$$f(Y_{t^n_{k+1}})-f(Y_{t^n_k})=\frac{\partial f}{\partial t}(t^n_k,X_{t^n_k})(t^n_{k+1}-t^n_k)+\frac{\partial f}{\partial x}(t^n_k,X_{t^n_k})(X_{t^n_{k+1}-}-X_{t^n_k-})$$
$$+\frac{1}{2}\sum_{i,j=1}^{2}\mathcal{R}^{i,j}_{n,k}(X^i_{t^n_{k+1}}-X^i_{t^n_k})(X^j_{t^n_{k+1}}-X^j_{t^n_k}).$$

Note that $X_{t^n_{k+1}-}-X_{t^n_k-}=X^c_{t^n_{k+1}}-X^c_{t^n_k}$. Now by Proposition 4.3.14 we have that

$$\frac{\partial f}{\partial t}(t^n_k,X_{t^n_k})(t^n_{k+1}-t^n_k)+\frac{\partial f}{\partial x}(t^n_k,X_{t^n_k})(X_{t^n_{k+1}-}-X_{t^n_k-})\xrightarrow{\mathbb{P}}\int_{\tau^A_m+\delta}^{\tau^A_{m+1}}\frac{\partial f}{\partial t}(X_{s-})\,ds+\int_{\tau^A_m+\delta}^{\tau^A_{m+1}}\frac{\partial f}{\partial x}(X_{s-})\,dX^c_s.$$

It remains to check that for every $i,j=1,\dots d$ we have that

$$\mathcal{R}^{i,j}_{n,k}(X^i_{t^n_{k+1}}-X^i_{t^n_k})(X^j_{t^n_{k+1}}-X^j_{t^n_k})\xrightarrow{\mathbb{P}}\int_{\tau^A_m+\delta}^{\tau^A_{m+1}}\frac{\partial^2 f}{\partial x^i\partial x^j}(X_s)\,d\langle X^i,X^j\rangle_s.$$

First of all we have that

$$\sup_{0\le k\le p_n-1}\left|\mathcal{R}^{i,j}_{n,k}-\frac{\partial^2 f}{\partial x^i\partial x_j}(X_{t^n_k})\right|\le\sup_{0\le k\le p_n-1}\sup_{x\in[X_{t^n_k}\wedge X_{t^n_{k+1}},X_{t^n_k}\vee X_{t^n_{k+1}}]}\left|\frac{\partial^2 f}{\partial x^i\partial x^j}(x)-\frac{\partial^2 f}{\partial x^i\partial x_j}(X_{t^n_k})\right|,$$

where the later converges almost surely to zero as $n\to\infty$. This is due to the uniform continuity of $\frac{\partial^2 f}{\partial x^i\partial x^j}$ and the continuity of the paths of $X$ over the compact interval $[\tau^A_m+\delta,\tau^A_{m+1}]$. We also know that

$$\sum_{k=0}^{p_n-1}(X^i_{t^n_{k+1}}-X^i_{t^n_k})(X^j_{t^n_{k+1}}-X^j_{t^n_k})\xrightarrow{\mathbb{P}}\langle X^i,X^j\rangle_{\tau^A_{m+1}}-\langle X^i,X^j\rangle_{\tau^A_m+\delta},$$

by Proposition 4.2.8. Therefore we have that

$$\sum_{k=0}^{p_n-1}\left(\mathcal{R}^{i,j}_{n,k}-\frac{\partial^2 f}{\partial x^i\partial x^j}(X_{t^n_k})\right)(X^i_{t^n_{k+1}}-X^i_{t^n_k})(X^j_{t^n_{k+1}}-X^j_{t^n_k})\xrightarrow{\mathbb{P}}0.$$

It follows by Propostion 4.3.14 that,

$$\sum_{k=0}^{p_n-1} \frac{\partial^2 f}{\partial x^i \partial x^j}(X_{t_k^n})(X^i_{t^n_{k+1}} - X^i_{t^n_k})(X^j_{t^n_{k+1}} - X^j_{t^n_k}) \xrightarrow{\mathbb{P}} \int_{\tau^A_m+\delta}^{\tau^A_{m+1}} \frac{\partial^2 f}{\partial x^i \partial x^j}(X_s)\, d\langle X^i, X^j\rangle_s.$$

Since $X^1$ is a finite variation process, note that $\langle X^1, X^2\rangle = \langle X^2, X^1\rangle = \langle X^1, X^1\rangle = 0$. Furthermore the integrals are càdlàg, so then by letting $\delta \downarrow 0$ we have

$$\begin{aligned} & f(t, X^n_{\tau^A_{m+1}-}) - f(\tau^A_m, X_{\tau^A_m}) \\ &\quad = \int_{\tau^A_m}^{\tau^A_{m+1}} \frac{\partial f}{\partial t}(s, X_{s-})\, ds + \int_{\tau^A_m}^{\tau^A_{m+1}} \frac{\partial f}{\partial x}(s, X_{s-})\, dX^c_s + \frac{1}{2}\int_{\tau^A_m}^{\tau^A_{m+1}} \frac{\partial^2 f}{\partial x^2}(s, X_{s-})\, d\langle X^c\rangle_s \\ &\qquad - \int_{\tau^A_m}^{\tau^A_{m+1}} \int_A \gamma_s(x) \frac{\partial f}{\partial x}(s, X_{s-})\, d\nu(J)\, ds. \end{aligned}$$

Now adding all up gives for each $n \in \mathbb{N}$

$$\begin{aligned} f(X^n_t) - f(X^n_0) &= \int_0^t \frac{\partial f}{\partial t}(s, X^n_{s-}) + \frac{\partial f}{\partial x}(s, X^n_{s-})\mu_s + \frac{1}{2}\frac{\partial^2 f}{\partial x^2}(s, X^n_{s-})\sigma_s^2\, ds + \int_0^t \frac{\partial f}{\partial x}(s, X^n_{s-})\sigma_s\, dW_s \\ &\quad + \int_0^t \int_{A_n} \left( f(s, X^n_{s-} + \gamma_s(J)) - f(s, X^n_{s-}) - \gamma_s(x)\frac{\partial f}{\partial x}(s, X^n_{s-}) \right) d\nu(J)\, ds \\ &\quad + \int_0^t \int_{A_n} \big( f(s, X^n_{s-} + \gamma_s(J)) - f(s, X^n_{s-}) \big)\, \widetilde{N}(ds, dJ). \end{aligned}$$

The final result follows by passing to the limit $n \to \infty$ due to Lemma 4.3.21, the stochastic dominated convergence theorem and regular dominated convergence on the respective integrals. For extending to arbitrary $X$ and $\gamma$ which are not neccessarily bounded, we can use a localisation argument. □

# CHAPTER 5
# Stochastic Differential Equations

Before Itô's theory of stochastic integration for Brownian motion and later more general semimartingales, diffusions were studied through semigroup theory. Let $C_0(\mathbb{R})$ denote the space of all real-valued continuous functions on $\mathbb{R}$ vanishing at infinity. Equipping $C_0(\mathbb{R})$ with the supremum norm gives a Banach space. A transition semigroup or also known as a Feller semigroup, is a collection $(T_t)_{t\geq 0}$ of positive operators on $C_0(X)$ such that $\|T_t f\|_\infty \leq \|f\|_\infty$ for all $t \geq 0$ and $f \in C_0(X)$, moreover for all $s, t \geq 0$, the semigroup property is satisfied $T_{t+s} = T_t \circ T_s$ and finally for every $f \in C_0(X)$, we have $\lim_{t\to 0} \|T_t f - f\|_\infty = 0$.

Now let $X$ be a *Markov process*, that is for all $f \in BM(\mathbb{R})$ and $s, t \in \mathcal{T}$ with $s < t$ we have

$$\mathbb{E}(f(X_t)|\mathcal{F}_s) = \mathbb{E}(f(X_t)|X_s).$$

Then we can define the transition probabilities $p_{s,t}(x, A) = \mathbb{P}(X_t \in A|X_s = x)$, so that each $p_{s,t}(x, \cdot)$ is a probability measure on $\mathcal{B}(\mathbb{R})$. Furthermore we can define the operators

$$(T_{s,t}f)(x) = \int_{\mathbb{R}} f(y)p_{s,t}(x, dy),$$

for each $f \in BM(\mathbb{R})$ and $x \in \mathbb{R}$. If the Markov process $X$ is homogeneous, that is to say that for all $s, t \in \mathcal{T}$ such that $s < t$ we have $p_{s,t} = p_{0,t-s}$, then we have $T_{s,t} = T_{0,t-s}$ and we can write $T_{0,t}$ as $T_t$. It turns out that $(T_t)_{t\geq 0}$ satisfies the semigroup property when $T_t(BM(\mathbb{R})) \subseteq BM(\mathbb{R})$ for all $t \geq 0$. If it also satisfies the other two properties for being a Feller semigroup, then we call $X$ a Feller process. It turns out that all solutions to forward stochastic differential equations with Lipschitz continuous coefficients can be seen as Feller processes, Applebaum [1, Theorem 6.7.4].

We will not make use of any explicit semigroup theory, but in the numerical treatment of the stochastic differential equations we will make extensive use of the underlying semigroup structure. Furthermore a Feller process can be defined through its *infinitesimal generator* which is

$$\mathcal{A}f = \lim_{t\downarrow 0} \frac{T_t f - f}{t},$$

whenever the limit is well-defined, then $\mathcal{A}$ is a partial-integro differential operator in the case that $X$ is a solution of a stochastic differential equation. So investigating diffusions could be done with investigating partial-integro differential equations, however such an approach did not allow for an analysis for the paths of a diffusion process and its properties. The theory of Itô does, however, allow for the analysis of the paths of solutions, and this theory is what we will use in the rest of the thesis.

The following two sections will be structured rather simple, we will introduce two different types of stochastic differential equations, Forward Stochastic Differential Equations with Jumps (FSDEJs) and Backward Stochastic Differential Equations with Jumps (BSDEJs), and we will prove under certain regularity conditions the existence and uniqueness of their solutions. Contrary to deterministc differential equations, BSDEJs are a theory of its own, as we can no longer reverse time through a change of variables, due to the restriction of adaptedness to a filtration.

## 5.1 Forward Stochastic Differential Equations

This section will be devoted to the uniqueness and existence of solutions of FSDEJs. Just as with ordinary differential equations, the proof rests on the following lemma known as Grönwall's lemma.

**Lemma 5.1.1** (Grönwall's lemma)**.** *Let $T > 0$ and let $f$ be a nonnegative bounded measurable function on $[0, T]$. Assume that there exist two constants $\alpha \geq 0$ and $\beta \geq 0$ such that, for every $t \in [0, T]$ we have that*

$$f(t) \leq \alpha + \beta \int_0^t f(s)\, ds.$$

*Then, have for every $t \in [0,T]$ that*

$$f(t) \leq \alpha \exp(\beta t).$$

*Proof.*
Iterating the condition on $f$, we get,

$$f(t) \leq \alpha + \alpha(\beta t) + \beta^2 \int_0^t \int_0^s f(r)\,dr\,ds.$$

So by induction, for every $n \geq 1$ we have,

$$f(t) \leq \sum_{k=0}^{n} \frac{\alpha(\beta t)^k}{k!} + \beta^{n+1} \int_0^t \int_0^{s_1} \cdots \int_0^{s_n} f(s_{n+1})\,ds_{n+1}.$$

We have that $f$ is bounded, so let $M > 0$ such that $|f| \leq M$, then

$$\beta^{n+1} \int_0^t \int_0^{s_1} \cdots \int_0^{s_n} f(s_{n+1})\,ds_{n+1} \leq M\beta^{n+1} \int_0^t \int_0^{s_1} \cdots \int_0^{s_n} ds_{n+1}.$$

Now,

$$\int_0^t \int_0^{s_1} \cdots \int_0^{s_n} ds_{n+1} = \int_0^t \int_0^{s_1} \cdots \int_0^{s_{n-1}} s_n\,ds_n = \cdots = \frac{t^{n+1}}{(n+1)!}$$

Then the result follows,

$$f(t) \leq \lim_{n\to\infty} \sum_{k=0}^{n} \frac{\alpha(\beta t)^k}{k!} + \frac{(\beta t)^{n+1}}{(n+1)!} = \alpha \sum_{i=0}^{\infty} \frac{(\beta t)^k}{k!} = \alpha \exp(\beta t).$$

□

Let $(\Omega, \mathcal{F}, \mathbb{F}, \mathbb{P})$ be a filtered probability space which satisfies the usual hypothesis. Now let $W$ be a Brownian motion and $N$ a Poisson random measure on $\mathbb{R}^+ \times \mathbb{R}_0$ with compensator $\rho(dt, dJ) = d\nu(J)\,dt$ such that $\nu$ is a Lévy measure. Define the mappings $b : [0,T] \times \mathbb{R} \to \mathbb{R}$, $\sigma : [0,T] \times \mathbb{R} \to \mathbb{R}$ and $\gamma : [0,T] \times \mathbb{R} \times \mathbb{R} \to \mathbb{R}$ such that there exists a $K > 0$ which satisfies for every $t > 0$ the following Lipschitz and growth conditions

$$|b_t(x) - b_t(x')|^2 + |\sigma_t(x) - \sigma_t(x')|^2 + \int_{\mathbb{R}_0} |\gamma_t(x,J) - \gamma_t(x',J)|^2\,d\nu(J) \leq K|x - x'|^2,$$

$$|b_t(x)|^2 + |\sigma_t(x)|^2 + \int_{\mathbb{R}_0} |\gamma_t(x,J)|^2\,d\nu(J) \leq K(1 + |x|^2).$$

Given $X_0 \in L^2$, an adapted càdlàg stochastic process $X$ solves the (FSDEJ) if

$$X_t = X_0 + \int_0^t b_s(X_{s-})\,ds + \int_0^t \sigma_s(X_{s-})\,dW_s + \int_0^t \int_{\mathbb{R}_0} \gamma_s(X_{s-}, J)\,\widetilde{N}(ds, dJ).$$

The following proof is based on Le Gall [26, Theorem 8.3] and Applebaum [1, Theorem 6.2.3].

**Theorem 5.1.2.** *There exists a unique adapted and càdlàg solution $X$ which solves the FSDEJ.*

*Proof.*
We will start with uniqueness. We consider two solutions $X$ and $X'$ such that $X_0 = X'_0$ almost surely. Fix $M > 0$ and set

$$\tau = \inf\{t \geq 0 : |X_t| > M \text{ or } |X'_t| > M\}.$$

Then, for every $t \leq T$,

$$X_{t\wedge\tau} = X_0 + \int_0^{t\wedge\tau} b_s(X_{s-})\,ds + \int_0^{t\wedge\tau} \sigma_s(X_{s-})\,dW_s + \int_0^{t\wedge\tau} \int_E \gamma_s(X_{s-}, J)\,\widetilde{N}(ds, dJ).$$

$$\|X_{t\wedge\tau} - X'_{t\wedge\tau}\|_{L^2} \leq 3\mathbb{E}\left(\left|\int_0^{t\wedge\tau} (b_s(X_{s-}) - b_s(X'_{s-}))\,ds\right|^2\right) + 3\mathbb{E}\left(\left|\int_0^{t\wedge\tau} (\sigma_s(X_{s-}) - \sigma_s(X'_{s-}))\,dW_s\right|^2\right)$$

$$+ 3\mathbb{E}\left(\left|\int_0^{t\wedge\tau}\int_E (\gamma_s(X_{s-},J) - \gamma_s(X'_{s-},J))\,\widetilde{N}(ds,dJ)\right|^2\right)$$

Now use Jensen's inequality on the first integral and the Itô isometry on the second and third integral

$$\begin{aligned}
&\leq 3T\mathbb{E}\left(\int_0^{t\wedge\tau} |b_s(X_{s-}) - b_s(X'_{s-})|^2\,ds\right) + 3\mathbb{E}\left(\int_0^{t\wedge\tau} |\sigma_s(X_{s-}) - \sigma_s(X'_{s-})|^2\,ds\right)\\
&\quad + 3\mathbb{E}\left(\int_0^{t\wedge\tau}\int_E |\gamma_s(X_{s-},J) - \gamma_s(X'_{s-},J)|^2\,d\nu(J)\,ds\right)\\
&\leq K(3T+6)\mathbb{E}\left(\int_0^{t\wedge\tau} |X_{s-} - X'_{s-}|^2\,ds\right)\\
&= K(3T+6)\int_0^t \|X_{s\wedge\tau} - X'_{s\wedge\tau}\|_{L^2}\,ds
\end{aligned}$$

Since we have that $\|X_{s\wedge\tau} - X'_{t\wedge\tau}\|_{L^2} \leq 2\|X_{s\wedge\tau}\|_{L^2} + 2\|X'_{s\wedge\tau}\|_{L^2} \leq 4M^2$ for all $s \leq T$, we have by Lemma 5.1.1

$$\|X_{t\wedge\tau} - X'_{t\wedge\tau}\|_{L^2} \leq 0\exp(K(2T+4)t) = 0.$$

Now for $t \leq T$ we have $X_{t\wedge\tau} = X'_{t\wedge\tau}$ almost surely, letting $M \to \infty$ we get $X_t = X'_t$ almost surely for all $t \leq T$.

Since both $X$ and $X'$ are càdlàg, we have by Proposition 2.3.5 that $X$ and $X'$ are indistinguishable.

To show the existence of a solution, we use Picard's approximation method. Define by induction the scheme

$$\begin{cases}
X_t^0 &= X_0,\\
X_t^1 &= X_0 + \int_0^t b_s(X_{s-}^0)\,ds + \int_0^t \sigma_s(X_{s-}^0)\,dW_s + \int_0^t\int_E \gamma_s(X_{s-}^0,J)\,\widetilde{N}(ds,dJ),\\
X_t^n &= X_0 + \int_0^t b_s(X_{s-}^{n-1})\,ds + \int_0^t \sigma_s(X_{s-}^{n-1})\,dW_s + \int_0^t\int_E \gamma_s(X_{s-}^{n-1},J)\,\widetilde{N}(ds,dJ).
\end{cases}$$

By induction we have that the processes $X^n$ are adapted and càdlàg. For every $n \in \mathbb{N}$ and every $t \leq T$, define

$$g_n(t) = \mathbb{E}\left(\sup_{s\leq t} |X_s^n - X_s^{n-1}|^2\right).$$

For $g_1$ we have, using the growth condition,

$$\begin{aligned}
g_1(t) &= \mathbb{E}\left(\sup_{s\leq t}\left|\int_0^s b_r(X_{s-}^0)\,dr + \int_0^s \sigma_r(X_{s-}^0)\,dW_r + \int_0^s \gamma_s(X_{s-}^0,J)\,\widetilde{N}(dr,dJ)\right|^2\right)\\
&\leq 3t\mathbb{E}\left(\int_0^t |b_s(X_{s-}^0)|^2\,ds\right) + 3\mathbb{E}\left(\int_0^t |\sigma_s(X_{s-}^0)|^2\,ds\right) + 3\mathbb{E}\left(\int_0^t\int_E |\gamma_s(X_{s-}^0,J)|^2\,d\nu(J)\,ds\right)\\
&\leq K(3T+6)(1+\mathbb{E}(|X_0|^2))t.
\end{aligned}$$

Further we have that

$$\begin{aligned}
g_{n+1}(t) &\leq 3t\mathbb{E}\left(\int_0^t |b_s(X_{s-}^n) - b_s(X_{s-}^{n-1})|^2\,ds\right) + 3\mathbb{E}\left(\int_0^t |\sigma_s(X_{s-}^n) - \sigma_s(X_{s-}^{n-1})|^2\,ds\right)\\
&\quad + 3\mathbb{E}\left(\int_0^t\int_E |\gamma_s(X_{s-}^n,J) - \gamma_s(X_{s-}^{n-1},J)|^2\,d\nu(J)\,ds\right)\\
&\leq K(3t+6)\mathbb{E}\left(\int_0^t |X_{s-}^n - X_{s-}^{n-1}|^2\,ds\right)\\
&\leq K(3T+6)\int_0^t g_n(s)\,ds\\
&\leq K^n(3T+6)^n\int_0^t\int_0^{s_1}\cdots\int_0^{s_{n-1}} g_1(s_n)\,ds_n\\
&\leq K^{n+1}(3T+6)^{n+1}(1+\mathbb{E}(|X_0|^2))\int_0^t\int_0^{s_1}\cdots\int_0^{s_{n-1}} s_n\,ds_n
\end{aligned}$$

$$= \frac{K^n(3T+6)^{n+1}t^{n+1}(1+\mathbb{E}(|X_0|^2))}{(n+1)!}.$$

Hence we have

$$\mathbb{E}\left(\left(\sum_{n=0}^{\infty}\sup_{t\le T}|X_t^{n+1}-X_t^n|\right)^2\right)^{1/2} \le \sum_{n=0}^{\infty}\mathbb{E}\left(\sup_{t\le T}|X_t^{n+1}-X_t^n|^2\right)^{1/2} = \sum_{n=1}^{\infty} g_n(T)^{1/2}$$
$$\le \sum_{n=1}^{\infty}\frac{K^{n/2}(3T+6)^{n/2}T^{n/2}\sqrt{1+\mathbb{E}(|X_0|^2)}}{\sqrt{n!}} < \infty.$$

Therefore,

$$\sum_{n=0}^{\infty}\sup_{t\le T}|X_t^{n+1}-X_t^n| < \infty,$$

almost surely. Hence the sequence of processes $(X^n)_{n\in\mathbb{N}}$ converges uniformly on $[0,T]$ almost surely, to a limiting process $X$, which is càdlàg. By induction, $X^n$ is adapted and so the same holds for $X$. It remains to check that $X$ indeed solves the FSDEJ, if we define

$$\widetilde{X}_t = X_0 + \int_0^t b_s(X_{s-})\,ds + \int_0^t \sigma_s(X_{s-})\,dW_s + \int_0^t\int_{\mathbb{R}_0}\gamma_s(X_{s-},J)\,\widetilde{N}(ds,dJ),$$

then similarly since $X_s^n$ converges to $X_s$ in $L^2$ for every $s\ge 0$,

$$\left\|\widetilde{X}_t - X_t^n\right\|_{L^2} \le K(3T+6)\mathbb{E}\left(\int_0^t |X_t - X_t^n|^2\,ds\right)$$
$$\le K(3T+6)t\sup_{s\le t}\|X_s - X_s^n\|_{L^2}^2$$
$$\le K(3T+6)t\sup_{s\le t}\sum_{k=n+1}^{\infty}\left\|X_s^k - X_s^{k-1}\right\|_{L^2}^2$$
$$\le K(3T+6)t\sum_{k=n+1}^{\infty} g_n(t)^{1/2} \underset{n\to\infty}{\longrightarrow} 0,$$

So by the uniqueness of the $L^2$-limit, $X=\widetilde{X}$ solves the FSDEJ. □

## 5.2 Backward Stochastic Differential Equations

The theory of Backward Stochastic Differential Equations has its roots in optimal stochastic control. The linear BSDE was already proposed in Bismut [5] but the well-posedness of nonlinear BSDEs was estabilished only fairly recently in the pivotal paper of Peng and Pardoux [30]. Later Peng and Pardoux developed the theory and applications of BSDEs in a series of papers under global Lipschitz conditions. Tang and Li [40] extended the idea of BSDEs to existence of adapted solutions to BSDEs with Poisson jumps with a fixed terminal time, as we will call BSDEJs in this thesis. A thorough survey on the recent developments and applications has been given in Peng [31] and Zhang [45].

Besides optimal stochastic control, BSDEs have also been linked to various problems in mathematical finance, a detailed exposition of the basic theory and applications of BSDEs in mathematical finance has been given in El Karoui, Peng and Quenez [14]. Nowadays the theory BSDE has grown a lot since the intial paper of Peng and Pardoux, it has been extended to infinite horizon problems by Peng and Shi [32], random terminal times by Jeanblanc, Mastrolia, Possamai et al. [23] and infinite dimensional BSDEJs by Hassani and Ouknine [20] among other extensions. However, we will focus solely on decoupled FBSDEJs in this thesis, which is a combination of a FSDEJ and a BSDEJ, which have been studied thoroughly by Barsela, Buckdahn and Pardoux [2].

In addition to the stochastic interpretation of FBSDEJs, we will also see a deterministic interpretation later in this thesis. It turns out that FBSDEJs correspond to viscosity solutions of partial-integro differential equations. Hence studying partial-integro differential equations can be done with FBSDEJs and vice versa.

To prove the existence of solutions to BSDEs, one often uses the following Martingale Representation Theorem, which states that if the filtration is generated by a Brownian motion and a Poisson random measure, then we can represent every càdlàg martingale in terms of stochastic integrals with respect to Brownian motion and stochastic integrals with respect to compensated Poisson random measures. A proof can be found in Kunita [25].

**Theorem 5.2.1** (Martingale Representation Theorem)**.** *Let $(\Omega, \mathcal{F}, \mathbb{F}, \mathbb{P})$ be a filtered probability space where the filtration is generated by a Brownian motion $W$ and a Poisson random measure $N$, let $M$ be a càdlàg $L^2$-martingale, then there exists $Z \in L^2_W$ and $U \in L^2_{\widetilde{N}}$ such that*

$$M_t = M_0 + \int_0^t Z_s \, dW_s + \int_0^t \int_{\mathbb{R}_0} U_s(x) \, \widetilde{N}(ds, dx).$$

*Where $Z$ is $\mathbb{P} \otimes \lambda$-a.s. unique and $U$ is $\mathbb{P} \otimes \widetilde{N}$-a.s. unique.*

Let $(\Omega, \mathcal{F}, \mathbb{F}, \mathbb{P})$ be a filtered probability space which satisfies the usual hypothesis. Now let $W$ be a Brownian motion and $N$ a Poisson random measure on $\mathbb{R}^+ \times \mathbb{R}_0$ with compensator $\rho(dt, dJ) = d\nu(J)\, dt$ such that $\nu$ is a Lévy measure and assume the filtration is generated by $W$ and $N$. Consider the Backward Stochastic Differential Equation (BSDEJ)

$$Y_t = \xi + \int_t^T f(s, Y_s, Z_s, U_s) \, ds - \int_t^T Z_s dW_s - \int_t^T \int_{\mathbb{R}_0} U_s(x) \, \widetilde{N}(ds, dx), \tag{5.1}$$

where the terminal value is an $\mathcal{F}_T$-measurable random variable, $\xi : \Omega \to \mathbb{R}$, and the driver $f$ is a mapping $\Omega \times [0, T] \times \mathbb{R} \times \mathbb{R} \times L^2(\nu) \to \mathbb{R}$ which is progressively measurable.

A solution is a pair $(Y, Z, U) \in \mathcal{S}^2 \times L^2_W \times L^2_{\widetilde{N}}$ which satisfies (5.1), where $\mathcal{S}^2$ denotes the space of all progressively measurable and càdlàg processes $X : \Omega \times [0, T] \to \mathbb{R}$ such that

$$\|X\|^2_{\mathcal{S}^2} := \mathbb{E}\left(\sup_{t \le T} |X_t|^2\right) < \infty.$$

Suppose $\xi \in L^2$, $f(\cdot, 0, 0, 0) \in L^2_W$, and $f$ is Lipschitz; i.e., there exists $L_f > 0$ such that $\mathbb{P} \otimes \lambda$-a.s.

$$|f(\omega, t, y, z, u) - f(\omega, t, y', z', u')| \le L_f(|y - y'| + |z - z'| + \|u - u'\|_{L^2(\nu)}), \qquad \forall (y, z, u), \forall (y', z', u'),$$

Then $(f, \xi)$ are said to be standard parameters for the BSDEJ.

For $\beta > 0$ we define the normed space $(L^2_W, \|\cdot\|_\beta)$ where

$$\|X\|^2_\beta = \mathbb{E}\left(\int_0^T e^{\beta t} |X_t|^2 \, dt\right),$$

this is well-defined since $\|\cdot\|_\beta$ and $\|\cdot\|_{L^2_W}$ are equivalent, therefore $L^2_W$ is also a Banach space endowed with the norm $\|\cdot\|_\beta$. Similarly define the normed space $(L^2_{\widetilde{N}}, \|\cdot\|_\beta)$ where

$$\|X\|^2_\beta = \mathbb{E}\left(\int_0^T e^{\beta t} \, \|U_t\|^2_{L^2(\nu)} \; dt\right).$$

Since the driver $f$ depends on the solution $(Y, Z, U)$ we can not immediately apply the Martingale Representation Theorem, instead we will use a fixed point iteration proof, using the Banach fixed point theorem. The result is well-known so we omit a proof.

**Theorem 5.2.2** (Banach Fixed Point Theorem)**.** *Let $(X, d)$ be a non-empty complete metric space with a mapping (contraction) $T : X \to X$ and $\lambda \in [0, 1)$ such that*

$$d(T(x), T(y)) \le \lambda d(x, y),$$

*for all $x, y \in X$. Then $T$ admits a unique fixed point $x^*$ in $X$.*

Finally we also need the famous Burkholder-Davis-Gundy inequality, in the case where the martingale is not necessarily continuous, the proof is highly non-trivial, we refer to Dellacherie and Meyer [13, page 287].

**Theorem 5.2.3** (Burkholder-Davis-Gundy inequality)**.** *Let $X$ be a local martingale with $X_0 = 0$ and $p \in [1, \infty)$, then for every stopping time $\tau$ we have the inequality*

$$\mathbb{E}\left(\sup_{t\geq t} |M_{t\wedge\tau}|^p\right) \leq C_p \mathbb{E}\left(\langle M\rangle_{t\wedge\tau}^{p/2}\right),$$

*where*

$$\langle X\rangle_t = \langle X^c\rangle_t + \sum_{0\leq s\leq t} (\Delta X_s)^2,$$

*and $X^c$ is the continuous part of $X$.*

The rough idea of the proof is due to Wu [43] and is essentially an adaptation of the same proof for continuous BSDEs by El Karoui, Peng and Quenez [14].

**Theorem 5.2.4.** *Given standard parameters $(f, \xi)$, there exists a unique pair $(Y, Z, U) \in \mathcal{S}^2 \times L^2_W \times L^2_{\widetilde{N}}$ which solves the BSDEJ.*

*Proof.*
For the rest of the proof we will fix the standard parameters $(f, \xi)$. The entire proof rests on the application of the Banach fixed point theorem for the mapping

$$\begin{aligned} \Phi : (L^2_W, \|\cdot\|_\beta) \times (L^2_W, \|\cdot\|_\beta) \times (L^2_{\widetilde{N}}, \|\cdot\|_\beta) &\to ((L^2_W, \|\cdot\|_\beta) \times (L^2_W, \|\cdot\|_\beta)) \times (L^2_{\widetilde{N}}, \|\cdot\|_\beta), \\ (y, z, u) &\mapsto (Y, Z, U), \end{aligned}$$

where $(Y, Z, U)$ solves the BSDEJ with driver $f_t(y_t, z_t, u_t)$ and for some specific $\beta > 0$. First we will have to show that this mapping is in fact well defined, and then we will show that it is a contraction, concluding the proof.

The assumption that $(f, \xi)$ are standard parameters implies that $f(\cdot, y, z, u) \in L^2_W(\mathbb{R})$, indeed

$$\begin{aligned} \mathbb{E}\left(\int_0^T |f_s(y_s, z_s, u_s)|^2 \, ds\right) &\leq 2\mathbb{E}\left(\int_0^T |f_s(y_s, z_s, u_s) - f_s(0,0,0)|^2 \, ds\right) + 2\mathbb{E}\left(\int_0^T |f_s(0,0,0)|^2 \, ds\right) \\ &\leq 4L_f^2 \mathbb{E}\left(\int_0^T |y_s|^2 + |z_s|^2 + \|U_s\|^2_{L^2(\nu)} \, ds\right) + 2\mathbb{E}\left(\int_0^T |f_s(0,0,0)|^2 \, ds\right) < \infty. \end{aligned}$$

Furthermore for all $t \leq T$ we have by using Jensen's Inequality,

$$\begin{aligned} \mathbb{E}\left(\left|\int_t^T f_s(y_s, z_s, u_s) \, ds\right|^2\right) &\leq (T - t)\mathbb{E}\left(\int_t^T |f_s(y_s, z_s, u_s)|^2 \, ds\right) \\ &\leq T\mathbb{E}\left(\int_0^T |f_s(y_s, z_s, u_s)|^2 \, ds\right) \\ &< \infty \end{aligned}$$

Hence for all $t \leq T$ we have that $\int_t^T f_s(y_s, z_s, u_s) \, ds \in L^2$. Therefore

$$M_t = \mathbb{E}\left(\int_0^T f_s(y_s, z_s, u_s) \, ds + \xi \,\middle|\, \mathcal{F}_t\right),$$

is a continuous martingale bounded in $L^2$. By Theorem 5.2.1 we know that there exists a unique representation (up to almost sure equivalence)

$$M_t = M_0 + \int_0^t Z_s \, dW_s + \int_0^t \int_{\mathbb{R}_0} U_s(x) \, \widetilde{N}(ds, dx),$$

with unique stochastic processes $Z \in L^2_W$ and $U \in L^2_{\widetilde{N}}$. Define the adapted process $Y$ by

$$Y_t = M_t - \int_0^t f_s(y_s, z_s, u_s) \, ds.$$

We will show that $(Y, Z, U)$ is the solution we are searching for. First we check that $Y \in L^2_W$, indeed,

$$\begin{aligned}
\mathbb{E}\left(\int_0^T |Y_t|^2\,dt\right) &= \mathbb{E}\left(\int_0^T \left|\mathbb{E}\left(\int_0^T f_s(y_s,z_s,u_s)+\xi\,\middle|\,\mathcal{F}_t\right) - \int_0^t f_s(y_s,z_s,u_s)\,ds\right|^2 dt\right)\\
&\le \int_0^T \mathbb{E}\left(\mathbb{E}\left(\left|\int_t^T f_s(y_s,z_s,u_s)+\xi\right|^2 \,\middle|\, \mathcal{F}_t\right)\right) dt\\
&= \int_0^T \mathbb{E}\left(\left|\int_t^T f_s(y_s,z_s,u_s)+\xi\right|^2\right) dt\\
&\le 2\int_0^T \mathbb{E}\left(\left|\int_t^T f_s(y_s,z_s,u_s)\,ds\right|^2 + |\xi|^2\right) dt\\
&\le 2T^2\mathbb{E}\left(\int_0^T |f_s(y_s,z_s,u_s)|^2\,ds\right) + 2T\mathbb{E}(|\xi|^2)\\
&< \infty
\end{aligned}$$

Now we only have to show that $(Y, Z, U)$ actually solves the BSDE, the uniqueness is due to the martingale representation. We know

$$\begin{aligned}
\xi = Y_T = M_T - \int_0^T f_s(y_s,z_s,u_s)\,ds = M_0 &+ \int_0^t Z_s\,dW_s + \int_0^t\int_{\mathbb{R}_0} U_s(x)\,\widetilde{N}(ds,dx)\\
&- \int_0^T f_s(y_s,z_s,u_s)\,ds,
\end{aligned}$$

So then

$$\begin{aligned}
Y_t &= M_t - \int_0^t f_s(y_s,z_s,u_s)\,ds\\
&= M_0 + \int_0^t Z_s\,dW_s + \int_0^t\int_{\mathbb{R}_0} U_s(x)\,\widetilde{N}(ds,dx) - \int_0^t f_s(y_s,z_s,u_s)\,ds\\
&= \left(M_0 + \int_0^T Z_s\,dW_s + \int_0^T\int_{\mathbb{R}_0} U_s(x)\,\widetilde{N}(ds,dx) - \int_0^T f_s(y_s,z_s,u_s)\right)\\
&\qquad - \int_t^T Z_s\,dW_s - \int_t^T\int_{\mathbb{R}_0} U_s(x)\,\widetilde{N}(ds,dx) + \int_t^T f_s(y_s,z_s,u_s)\,ds\\
&= \xi + \int_t^T f_s(y_s,z_s,u_s)\,ds - \int_t^T Z_s\,dW_s - \int_t^T\int_{\mathbb{R}_0} U_s(x)\,\widetilde{N}(ds,dx)
\end{aligned}$$

and hence we get that $\Phi$ is well defined. Furthermore we have by Theorem 3.2.1 that

$$\begin{aligned}
&\mathbb{E}\left(\sup_{t\le T}\left|\int_0^t Z_s\,dW_s\right|^2 + \left|\int_0^t\int_{\mathbb{R}_0} U_s(x)\,\widetilde{N}(ds,dx)\right|^2\right)\\
&\qquad \le 4\mathbb{E}\left(\left|\int_0^T Z_s\,dW_s\right|^2 + \left|\int_0^T\int_{\mathbb{R}_0} U_s(x)\,\widetilde{N}(ds,dx)\right|^2\right) \le 4\mathbb{E}\left(\int_0^T |Z_s|^2 + \|U_s\|^2_{L^2(\nu)}\,ds\right).
\end{aligned}$$

Hence it follows after some calculations that $Y \in \mathcal{S}^2$.

Let $(y^1, z^1, u^1)$ and $(y^2, z^2, u^2)$ be two elements of $(L^2_W, \|\cdot\|_\beta) \times (L^2_W, \|\cdot\|_\beta) \times (L^2_{\widetilde{N}}, \|\cdot\|_\beta)$ and consider their images $\Phi(y^1, z^1, u^1) = (Y^1, Z^1, U^1)$ and $\Phi(y^2, z^2, u^2) = (Y^2, Z^2, U^2)$, respectively. Define $\delta X_t = X^1_t - X^2_t$ to be the difference of two processes. Now note that $\delta Y_T = \xi - \xi = 0$. We can use Theorem 4.4.12

on $f(t,x) = e^{\beta t}x^2$ since $\delta Y_t$ is a Lévy-Itô process. Then,

$$
\begin{aligned}
0 = e^{\beta T}|\delta Y_T|^2 = e^{\beta t}|\delta Y_t|^2 + \beta \int_t^T e^{\beta s}|\delta Y_s|^2\, ds - 2\int_t^T e^{\beta s}\delta Y_{s-}\left(f_s(y_s^1, z_s^1, u_s^1) - f_s(y_s^2, z_s^2, u_s^2)\right)\, ds \\
+ 2\int_t^T e^{\beta s}\delta Y_{s-}\delta Z_s\, dW_s - 2\int_t^T\int_{\mathbb{R}_0} \delta U_s(x)\delta Y_{s-}\, d\nu(x)\, ds + \int_t^T e^{\beta s}|\delta Z_s|^2\, ds \\
+ \int_t^T\int_{\mathbb{R}_0} e^{\beta s}\left((\delta Y_{s-} + \delta U_s(x))^2 - \delta Y_{s-}^2\right) N(ds, dx)
\end{aligned}
$$

Now $(\delta Y_{s-} + \delta U_s(x))^2 - \delta Y_{s-}^2 = \delta U_s^2(x) + 2\delta U_s(x)\delta Y_{s-}$. So then

$$
\begin{aligned}
&\int_t^T\int_{\mathbb{R}_0} e^{\beta s}\left((\delta Y_{s-} + \delta U_s(x))^2 - \delta Y_{s-}^2\right) N(ds, dx) \\
&\qquad = \int_t^T\int_{\mathbb{R}_0} e^{\beta s}\left((\delta Y_{s-} + \delta U_s(x))^2 - \delta Y_{s-}^2\right) \widetilde{N}(ds, dx) \\
&\qquad\quad \int_t^T\int_{\mathbb{R}_0} e^{\beta s}\left(\delta U_s^2(x) + 2\delta U_s(x)\delta Y_{s-}\right) d\nu(x)\, ds \\
&\qquad = \int_t^T\int_{\mathbb{R}_0} e^{\beta s}\left((\delta Y_{s-} + \delta U_s(x))^2 - \delta Y_{s-}^2\right) \widetilde{N}(ds, dx) \\
&\qquad\quad \int_t^T e^{\beta s}\left\|U_s\right\|_{L^2(\nu)}^2\, ds + 2\int_t^T\int_{\mathbb{R}_0} e^{\beta s}\delta U_s(x)\delta Y_{s-}\, d\nu(x)\, ds
\end{aligned}
$$

Hence we get

$$
\begin{aligned}
e^{\beta T}|\delta Y_T|^2 = e^{\beta t}|\delta Y_t|^2 + \int_t^T e^{\beta s}\left(\beta|\delta Y_s|^2 + |Z_s|^2 + \left\|U_s\right\|_{L^2(\nu)}^2\right) ds \\
- 2\int_t^T e^{\beta s}\delta Y_{s-}\left(f_s(y_s^1, z_s^1, u_s^1) - f_s(y_s^2, z_s^2, u_s^2)\right) ds + 2\int_t^T e^{\beta s}\delta Y_{s-}\delta Z_s\, dW_s \\
+ \int_t^T\int_{\mathbb{R}_0} e^{\beta s}\left((\delta Y_{s-} + \delta U_s(x))^2 - \delta Y_{s-}^2\right) \widetilde{N}(ds, dx)
\end{aligned}
$$

Now the last two integrals are both martingales. Indeed we have by Theorem 5.2.3 and the Cauchy-Schwarz inequality that

$$
\begin{aligned}
\mathbb{E}\left(\sup_{t\leq T}\left|\int_0^t e^{\beta s}\delta Y_{s-}\delta Z_s\, dW_s\right|\right) &\leq C\mathbb{E}\left(\left(\int_0^T |\delta Y_{s-}|^2|\delta Z_s|^2\, ds\right)^{1/2}\right) \\
&\leq C\mathbb{E}\left(\sup_{t\leq T}|\delta Y_t|\left(\int_0^T |\delta Z_s|^2\, ds\right)^{1/2}\right) \\
&\leq C\left\|\delta Y\right\|_{\mathcal{S}^2}\left\|\delta Z\right\|_{L_W^2} < \infty.
\end{aligned}
$$

Then by Proposition 3.3.2 we have that $\left(\int_\cdot^T e^{\beta s}\delta Y_{s-}\delta Z_s\, dW_s\right)_{t\leq T}$ is a uniformly integrable martingale with expectation zero and similarly note that by the mean value theorem $\left|(\delta Y_{s-} + \delta U_s(x))^2 - \delta Y_{s-}^2\right| \leq$

$2|\delta U_s(x)| \sup_{r\le T} |\delta Y_r|$. Hence we have by Theorem 5.2.3

$$\begin{aligned}
&\mathbb{E}\left(\sup_{t\le T}\left|\int_0^t \int_{\mathbb{R}_0} e^{\beta s}\left((\delta Y_{s-}+\delta U_s(x))^2-\delta Y_{s-}^2\right) \widetilde{N}(ds,dx)\right|\right) \\
&\qquad \le C\mathbb{E}\left(\left(\int_0^T \int_{\mathbb{R}_0} |(\delta Y_{s-}+\delta U_s(x))^2-\delta Y_{s-}^2|^2 \, N(ds,dx)\right)^{1/2}\right) \\
&\qquad \le C\mathbb{E}\left(2\sup_{t\le T}|\delta Y_t| \left(\int_0^T \int_{\mathbb{R}_0} |\delta U_s(x)|^2 \, N(ds,dx)\right)^{1/2}\right) \\
&\qquad \le 2C \left\|\delta Y\right\|_{\mathcal{S}^2} \mathbb{E}\left(\int_0^T \int_{\mathbb{R}_0} |\delta U_s(x)|^2 \, N(ds,dx)\right) < \infty.
\end{aligned}$$

Hence also $\left(\int_\cdot^T \int_{\mathbb{R}_0} e^{\beta s}\left((\delta Y_{s-}+\delta U_s(x))^2-\delta Y_{s-}^2\right) \widetilde{N}(ds,dx)\right)_{t\le T}$ is a uniformly integrable martingale with zero expectation. Therefore if we take expectations on both sides, rewrite a bit and substitute $t=0$ we get

$$\begin{aligned}
&\mathbb{E}\left(|\delta Y_0|^2\right)+\mathbb{E}\left(\int_0^T e^{\beta s}|\beta||\delta Y_s|^2+|\delta Z_s|^2+\left\|U_s\right\|^2 \, ds\right) \\
&\qquad = 2\mathbb{E}\left(\int_0^T e^{\beta s}\delta Y_{s-}\left(f_s(y_s^1,z_s^1,u_s^1)-f_s(y_s^2,z_s^2,u_s^2)\right) ds\right)
\end{aligned}$$

Now note that for any $\gamma>0$ and $x,y\in\mathbb{R}$ we have $\left(\sqrt{\gamma}x-\frac{1}{\sqrt{\gamma}y}\right)^2\ge 0$, so then $2xy\le \gamma x^2+\frac{1}{\gamma}y^2$, furthermore $(|x|+|y|+|z|)^2\le 3(|x|^2+|y|^2+|z|^2)$ hence we have

$$\begin{aligned}
2\delta Y_{s-}(f_s(y_s^1,z_s^1,u_s^1)-f_s(y_s^2,z_s^2,u_s^2)) &\le 2L_f|\delta Y_{s-}|(|\delta y_s|+|\delta z_s|+\left\|\delta u_s\right\|) \\
&\le \frac{L_f^2|\delta Y_{s-}|^2}{\gamma}+3\gamma\left(|\delta y_s|^2+|\delta z_s|^2+\left\|\delta u_s\right\|^2\right)
\end{aligned}$$

Now choose $\gamma=1/6$ and set $\beta=6L_f^2+1$, then

$$\begin{aligned}
&\mathbb{E}\left(\int_0^T e^{\beta s}|\beta||\delta Y_s|^2+|\delta Z_s|^2+\left\|U_s\right\|^2 \, ds\right) \\
&\qquad \le 2\mathbb{E}\left(\int_0^T e^{\beta s}\delta Y_{s-}\left(f_s(y_s^1,z_s^1,u_s^1)-f_s(y_s^2,z_s^2,u_s^2)\right) ds\right) \\
&\qquad \le (\beta-1)\mathbb{E}\left(\int_0^T e^{\beta s}|\delta Y_{s-}|^2 \, ds\right)+\frac{1}{2}\mathbb{E}\left(\int_0^T e^{\beta s}\left(|\delta y_s|^2+|\delta z_s|+\left\|\delta z_s\right\|^2\right) ds\right)
\end{aligned}$$

So

$$\left\|\delta Y\right\|_\beta^2+\left\|\delta Z\right\|_\beta^2+\left\|\delta U\right\|_\beta^2\le\frac{1}{2}\left(\left\|\delta y\right\|_\beta^2+\left\|\delta z\right\|_\beta^2+\left\|\delta u\right\|_\beta^2\right)$$

So for this choice of $\beta$ we have a contraction, hence $(Y,Z,U)$ is the unique solution of the BSDEJ. Since we had shown that $\Phi$ maps into the smaller space $(\mathcal{S}^2,\left\|\cdot\right\|_{\mathcal{S}^2})\times(L_W^2,\left\|\cdot\right\|_\beta)\times(L_{\widetilde{N}}^2,\left\|\cdot\right\|_\beta)$, the unique fixed point $(Y,Z,U)$ must also lie in this space. □

# Part II

# Numerical Analysis

CHAPTER 6

# COS Method

One of the main difficulties of solving FBSDEJs is that it is near inevitable we have to solve a series of conditional expectations, where the conditional densities are often unknown or numerically inefficient to use. Furthermore, they tend to have unbounded support.

The COS method is a highly efficient algorithm for solving conditional expectations where we have prior knowledge of the characteristic function of the conditional density, as it is based on a Fourier cosine expansion of the conditional density.

## 6.1 Smoothness transitional density

The conditional densities we will see in this thesis are closely related to the density of the FSDEJ. For the COS method to work effectively we need our density to be as smooth as possible, since the smoother the density, the faster the Fourier cosine series will converge. Hence the first problem we will investigate is to find out given the components of a FSDEJ, how smooth the density is. Since we have a lot of knowledge about the characteristic function of a FSDEJ, it is natural to investigate the smoothness through the characteristic function.

The characteristic function of a probability measure $\mu$ on $\mathbb{R}$ is in essence, just the inverse Fourier transform, as

$$\phi_\mu(\xi) = \int_{\mathbb{R}} e^{ix\xi}\, d\mu(x).$$

Therefore the theory of Fourier analysis can immediately be applied to characteristic functions. It turns out that there is a connection between the integrability of the distribution and the smoothness of the characteristic function, but also the integrability of the characteristic function and the existence of a continuous density. We will not prove the following proposition, for a proof, see for example [34, Proposition 11.0.1]. We will denote $\mathcal{P}(\mathbb{R})$ as the space of all Borel probability measures on $\mathbb{R}$, which should not be confused with the power set of $\mathbb{R}$.

**Proposition 6.1.1.** *Let $\mu \in \mathcal{P}(\mathbb{R})$.*

*(i) Let $n \in \mathbb{N}_0$, if we have*

$$\int_{\mathbb{R}} |x|^n\, d\mu(x) < \infty,$$

*then $\phi_\mu \in C^n$ and, for any nonnegative integer $k \leq n$ we have*

$$\int_{\mathbb{R}} x^k\, d\mu(x) = (-i)^k \frac{\partial^k \phi_\mu}{\partial \xi^k}(0).$$

*(ii) If $\int_{\mathbb{R}} |\phi_\mu(\xi)|\, d\xi < \infty$, then $\mu$ is absolutely continuous with respect to the Lebesgue measure, has a density $f \in C_b(\mathbb{R})$ and*

$$f(x) = \frac{1}{2\pi} \int_{\mathbb{R}} e^{-ix\xi} \phi_\mu(\xi)\, d\xi.$$

A Borel measure $\mu$ is said to be *discrete* if there exists a countable set $N$ such that $\mu(N^c) = 0$. The measure $\mu$ is said to be *continuous* if $\mu(\{x\}) = 0$ for every $x \in \mathbb{R}$, furthermore we say that $\mu$ is said to be *singular* if there is a set $A \in \mathcal{B}(\mathbb{R})$ such that $\mu(A^c) = 0$ and $|A| = 0$ (set of Lebesgue measure zero). Finally, we say that $\mu$ is *absolutely continuous* if $\mu(A) = 0$ for every $A \in \mathcal{B}(\mathbb{R})$ satisfying $|A| = 0$. We have the celebrated Lebesgue decomposition (see [6, Theorem 3.2.3] for a proof) of every finite measure $\mu$ giving $\mu = \mu_{as} + \mu_d + \mu_{cs}$ where $\mu_{as}$ is absolutely continuous, $\mu_d$ is discrete and $\mu_{cs}$ is continuous singular, we call $\mu_{as} + \mu_{cs}$ the continuous part.

Since FSDEJs consist of multiple components, it is useful to look at how the convolution operator preserves the discreteness, continuity and absolute continuity proprety of the measures.

**Lemma 6.1.2.** *Let $\mu_1, \mu_2 \in \mathcal{M}^+(\mathbb{R})$ be non-zero finite measures on $\mathbb{R}$. Let $\mu = \mu_1 * \mu_2$, then*

*(i) $\mu$ is continuous if and only if $\mu_1$ or $\mu_2$ is continuous*

*(ii) $\mu$ is discrete if and only if $\mu_1$ and $\mu_2$ are discrete*

*(iii) $\mu$ is absolutely continuous if $\mu_1$ or $\mu_2$ is absolutely continuous*

*(iv) $\mu_1$ or $\mu_2$ is continuous singular if $\mu$ is continuous singular*

*Proof.*
**(i)**, **(ii)**. If $\mu_1$ is continuous, then $\mu$ is continuous, because

$$\mu(\{x\}) = \int_{\mathbb{R}} \mu_1(\{x-y\})\, d\mu_2(y) = 0.$$

If $\mu_1$ and $\mu_2$ are discrete, then $\mu$ is discrete, since $\mu_1(N_1^c) = \mu_2(N_2^c) = 0$ with some countable sets $N_1$ and $N_2$ and for $N = N_1 + N_2$ we have $\mu(N^c) = 0$. The 'only if' part for (i) follows from the 'if' part of (ii), if both $\mu_1$ and $\mu_2$ are not continuous, then they have discrete parts, but then $\mu$ would have a discrete part as well, hence $\mu$ is not continuous either. Similarly if $\mu_1$ and $\mu_2$ are not discrete, then they must be continuous, hence $\mu$ is continuous and not discrete either.

**(iii)** Suppose that $\mu_1$ is absolutely continuous. If $A \in \mathcal{B}(\mathbb{R})$ satisfies $|A| = 0$, then $|A - y| = 0$ for every $y \in \mathbb{R}$ and $\mu(A) = \int_{\mathbb{R}} \mu_1(A-y)\, d\mu_2(y) = 0$.

**(iv)** Suppose that neither $\mu_1$ nor $\mu_2$ is continuous singular. Then $(\mu_1)_d + (\mu_1)_{ac} \neq 0$ and $(\mu_2)_d + (\mu_2)_{ac} \neq 0$. It follows from (ii) and (iii) that $((\mu_1)_d + (\mu_1)_{ac}) * ((\mu_2)_d + (\mu_2)_{ac})$ has a discrete or absolutely continuous part, hence $\mu$ is not continuous singular. □

Before we look at the smoothness of the density, it is good to note which distributions do not even admit a density. A necessary condition for the existence of a density is the continuity of the distribution. It should not be surprising that pure jump processes with finite Lévy measure are not continuous, however the continuity of jump processes with non-finite Lévy measure is perhaps less intuitive. The following proof is taken from [37, Theorem 27.4].

**Theorem 6.1.3.** *For a Lévy process $X$ on $\mathbb{R}$ with characteristic triplet $(\mu, \sigma^2, \nu)$, the following three statements are equivalent.*

*(i) $\mu_{X_t}$ is continuous for every $t > 0$.*

*(ii) $\mu_{X_t}$ is continuous for some $t > 0$.*

*(iii) Either $\sigma \neq 0$ or $\nu(\mathbb{R}) = \infty$ or both*

*Proof.*
**(ii) $\Longrightarrow$ (iii)** Suppose that $\mu = 0$ (not to be confused with the measure $\mu_{X_t}$) and $\nu(\mathbb{R}) < \infty$, then $X_t - \mu t$ is a compound Poisson process, and $\mathbb{P}(X_t - \mu t = 0) > 0$, hence the distribution has a discrete part and is therefore not continuous.

**(iii) $\Longrightarrow$ (i)** If $\sigma \neq 0$, then $X_t$ consists of a non-trivial Gaussian distribution for every $t > 0$, which has a continuous distribution, hence by Lemma 6.1.2 we have that $\mu_{X_t}$ is continuous. Now suppose that $\nu(\mathbb{R}) = \infty$. We will prove this statement, in the case where $\nu$ is discrete, the case where $\nu$ is continuous and the case where $\nu$ is neither discrete nor continuous.

First assume that $\nu$ is discrete, then let $x_1, x_2, \ldots$ be the points with non-zero $\nu$-measure and define $p_k = \nu(\{x_k\})$ and $p'_k = p_k \wedge 1$. We know that $\sum_{k=1}^{\infty} p_k = \infty$ and hence $\sum_{k=1}^{\infty} p'_k = \infty$. Now let $Y_t^n$ be the compound Poisson process with Lévy measure $\nu_n = \sum_{j=1}^{n} p'_k \delta_{x_k}$. Define in general for a probability measure $\mu \in \mathcal{P}(\mathbb{R})$,

$$\mathcal{S}(\mu) = \sup_{x \in \mathbb{R}} \mu(\{x\}).$$

When $\mu$ is the distribution of a random variable $X$, we write $\mathcal{S}(X) = \mathcal{S}(\mu)$. If $\mu = \mu_1 * \mu_2$, then $\mathcal{S}(\mu) \leq \mathcal{S}(\mu_1)$, since

$$\mu(\{x\}) = \int_{\mathbb{R}} \mu_1(\{x-y\})\, d\mu_2(y) \leq \mathcal{S}(\mu_1).$$

So it follows that $\mathcal{S}(X_t) \leq \mathcal{S}(Y_t^n)$. Let $c_n = \nu_n(\mathbb{R})$ and $\sigma_n = c_n^{-1}\nu_n$, then the $\sigma_n$ are probability measures. Recall the notation $\mu^k = \mu * \cdots * \mu$, $k$-times, then we have $\mathcal{S}(\sigma_n^k) \leq \mathcal{S}(\sigma_n) \leq c_n^{-1}$, then note that

$$\mathbb{P}(Y_t^n = x) = e^{-tc_n} \sum_{k=0}^{\infty} \frac{t^k}{k!} c_n^k \sigma_n^k(\{x\}) \leq e^{-tc_n} + c_n^{-1}.$$

Hence $\mathcal{S}(Y_t^n) \leq e^{-tc_n} + c_n^{-1}$. As we let $n \to \infty$ we have $c_n \to \infty$ and hence, we get $\mathcal{S}(X_t) = 0$, therefore $X_t$ has a continuous distribution.

Assume that $\nu$ is continuous, then let $Y^n$ be the compound Poisson process with Lévy measure $\nu_n = \nu|_{\{|x|>1/n\}}$, then we again have $\mathcal{S}(X_t) \leq \mathcal{S}(Y_t^n)$. For any $k \in \mathbb{N}$ we have that $\nu_n^k$ is continuous by Lemma 6.1.2. Hence, we have that $Y_t^n$ can only have a non-zero probability at 0, indeed taking again $c_n = \nu_n(\mathbb{R})$ and $\sigma_n = c_n^{-1}\nu_n$ we get

$$\mathbb{P}(Y_t^n = x) = e^{-tc_n} \sum_{k=0}^{\infty} \frac{t^k}{k!} c_n^k \sigma_n^k(\{x\}) = e^{-tc_n}\delta_0(\{x\})$$

It follows that $\mathcal{S}(Y_t^n) = e^{-tc_n}$ and since $c_n \to \infty$, we have $\mathcal{S}(X_t) = 0$.

In the last case where $\nu$ is neither discrete nor continuous, let $\nu_d$ and $\nu_c$ be the discrete and continuous parts of $\nu$, respectively. Then either $\nu_d$ or $\nu_c$ has infinite measure. If $\nu_d(\mathbb{R}) = \infty$, then letting $Y$ be the Lévy process with characteristic triplet $(0, 0, \nu_d)$, we see that $Y_t$ has a continuous distribution for any $t > 0$ by our above arguments. Then by Lemma 6.1.2 it follows that $X_t$ has a continuous distribution as well. If $\nu_c(\mathbb{R}) = \infty$, then let $Y$ be the Lévy process with characteristic triplet $(0, 0, \nu_c)$, then again $Y_t$ has a continuous distribution and so does $X_t$. □

We still have two big classes of Lévy processes left which might be able to admit a density. We have the class of *jump-diffusion processes*, which are Lévy processes which have a non-trivial Brownian motion part and finite Lévy measure, and the class of *infinite activity processes*, which are processes with a non-finite Lévy measure.

The following theorem is classical in Fourier analysis for limiting arguments.

**Theorem 6.1.4** (Riemann-Lebesgue Lemma)**.** *Let* $f \in L^1(\mathbb{R})$. *Then*

$$\lim_{|\xi|\to\infty} \int_{\mathbb{R}} f(x)e^{i\xi x}\, dx = 0.$$

*Proof.*
First assume $f(x) = \mathbf{1}_{(a,b)}(x)$ for some $a < b$. Then

$$\int_{\mathbb{R}} f(x)e^{i\xi x}\, dx = \int_a^b e^{i\xi x}\, dx = \frac{e^{i\xi b} - e^{i\xi a}}{i\xi} \underset{|\xi|\to\infty}{\longrightarrow} 0.$$

The result follows for simple functions by linearity of the integral, so pick $f \in L^1(\mathbb{R})$ arbitrary, since the space of simple functions is dense in $L^1(\mathbb{R})$, the result follows. □

We can formulate the following criterion for smoothness in terms of integrability of the characteristic function.

**Proposition 6.1.5.** *If* $\mu \in \mathcal{P}(\mathbb{R})$ *satisfies*

$$\int_{\mathbb{R}} |\phi_\mu(\xi)||\xi|^n\, d\xi < \infty,$$

*for some* $n \in \mathbb{N}_0$, *then* $\mu$ *has a density* $f$ *of class* $C^n$ *and the derivates vanish at infinity*

*Proof.*
By Proposition 6.1.1 we have that

$$f(x) = \frac{1}{2\pi} \int_{\mathbb{R}} e^{-ix\xi} \phi_\mu(\xi)\, d\xi,$$

is the density of $\mu$. Now the right-hand side is $n$-times differentiable by the Leibniz integral rule by our integrability assumption on $\phi_\mu$. Then $f \in C^n$ and for $k \leq n$, we have

$$f^{(k)}(x) = \frac{1}{2\pi} \int_{\mathbb{R}} (-i\xi)^k e^{-ix\xi} \phi_\mu(\xi)\, d\xi.$$

Then for all $k \le n$ we have that $|(i\xi)^k \phi_\mu(\xi)| = |\xi|^k |\phi_\mu(\xi)| \in L^1(\mathbb{R})$, so by Theorem 6.1.4, we have that $f^{(k)}(x) \underset{|x|\to\infty}{\longrightarrow} 0$. □

There also exist conditions for absolute continuity of the distribution and conditions for the smoothness of the density in Sato [37, pages 177-193] and Orey [29], but they are generally quite difficult to prove in practice and we therefore omit them in this thesis.

Before we state the main theorem of this section, we look at two important examples of Lévy processes, first we have the Merton jump-diffusion model, which is a jump-diffusion process.

*Example* 6.1.6 (Merton jump-diffusion). The Merton jump-diffusion model is a Lévy process

$$X_t = \int_0^t \mu \, dt + \int_0^t \sigma \, dW_t + \int_0^t \int_{\mathbb{R}_0} x \, \widetilde{N}(dt, dx).$$

with $\mu, \overline{\mu} \in \mathbb{R}, \sigma, \overline{\sigma}, \lambda > 0$ and

$$\nu(A) = \int_A \frac{\lambda}{\sqrt{2\pi\delta}} \exp\left(-\frac{(x-\overline{\mu})^2}{2\overline{\sigma}^2}\right) dx.$$

Here $\lambda$ is the jump intensity, $\overline{\mu}$ is the mean jump size and $\overline{\sigma}$ is the corresponding jump volatility. So we assume that jump size is normally distributed with intensity $\lambda$. The characteristic function is given as follows

$$\begin{aligned} \phi_{X_t}(\xi) &= \exp\left(i\mu t\xi - \frac{1}{2}\sigma^2 t\xi^2 + t\int_{\mathbb{R}} (e^{ix\xi} - 1)\, d\nu(x)\right) \\ &= \exp\left(i\mu t\xi - \frac{1}{2}\sigma^2 t\xi^2\right) \exp\left(\lambda t \left(\exp\left(i\overline{\mu}\xi - \frac{1}{2}\overline{\sigma}^2\xi^2\right) - 1\right)\right) \end{aligned}$$

Hence we have

$$|\phi_{X_t}(\xi)| = \exp\left(-\frac{1}{2}\sigma^2 t\xi^2\right) \exp\left(\lambda t \left(\cos(\overline{\mu}\xi) \exp\left(-\frac{1}{2}\overline{\sigma}^2\xi^2\right) - 1\right)\right) \le \exp\left(-\frac{1}{2}\sigma^2 t\xi^2\right).$$

It follows by Proposition 6.1.5 that the Merton jump-diffusion model has a density of class $C^\infty$ when $\sigma > 0$.

Next to the Merton jump-diffusion, there is a famous example of an infinite activity process, the variance gamma process.

*Example* 6.1.7 (Variance Gamma). A gamma process is a random process with independent gamma distributed increments, written as $\Gamma(t; \gamma, \lambda)$, which is a pure-jump increasing Lévy process with Lévy measure $\nu(x) = \gamma x^{-1} \exp(-\lambda x) \mathbf{1}_{(0,\infty)}(x)$. Now the Variance Gamma (VG) process is a sum of a gamma process and a time-changed Brownian motion,

$$X_t = \theta\gamma_t + \sigma W_{\gamma_t},$$

where $\gamma_t \sim \Gamma(t; 1, \kappa)$ and $\sigma, \kappa > 0, \theta \in \mathbb{R}$. It turns out that $X$ is again a Lévy process with characteristic function Fang [16, Table 1.1]

$$\phi_{X_t}(\xi) = \exp(i\mu t\xi) \left(1 - i\theta\kappa\xi + \frac{1}{2}\sigma^2\kappa\xi^2\right)^{-\frac{t}{\kappa}}.$$

Then we have

$$|\phi_{X_t}(\xi)| = \left(\left(1 + \frac{1}{2}\sigma^2\kappa\xi^2\right)^2 + \theta^2\kappa^2\xi^2\right)^{-\frac{t}{2\kappa}} \sim \mathcal{O}\left(|\xi|^{-\frac{2t}{\kappa}}\right),$$

as $|\xi| \to \infty$. It follows that the Variance Gamma model has a $C^n$ density for $n < \frac{2t}{\kappa} - 1$. So for big $\kappa$, we get a smoother density when $t$ is taken to be big as well. A sufficient condition for having a continuous density which tends to zero in the tails, is $t > \frac{\kappa}{2}$. Which means that for any $\kappa$, the density of $X_t$ for very small $t$ might be a serious problem.

The following theorem solves the problem of smoothness for a very large class of Lévy-Itô processes and also gives a generalized version of the Lévy-Khintchine representation (Theorem 4.4.10) as byproduct.

**Theorem 6.1.8.** *Let $X$ be a Lévy-Itô process where the coefficients $\mu, \sigma$ and $\gamma$ are deterministic and bounded. Suppose furthermore that $\nu(\mathbb{R}_0) < \infty$ and $\sigma_t \neq 0$ for all $t \geq 0$, then $X_t$ has a $C^\infty$ density for all $t \geq 0$ with all derivatives vanishing at infinity.*

*Proof.*
We will apply Itô's formula on $f(X_t) = e^{i\xi X_t}$ for each $\xi \in \mathbb{R}$, we have only proven Itô's formula for real-valued functions, however, we can just as well do the real part and imaginary part separately and add them up later to get the same formula. It follows that

$$\begin{aligned} e^{i\xi X_t} &= e^{i\xi X_0} + \int_0^t e^{i\xi X_s}\left(i\xi\mu_s - \frac{1}{2}\sigma_s^2\xi^2 + \int_{\mathbb{R}_0} e^{i\xi\gamma_s(J)} - 1 - i\xi\gamma_s(J)\, d\nu(J)\right) ds \\ &\quad + \int_0^t i\xi e^{i\xi X_s}\sigma_s\, dW_s + \int_0^t\int_{\mathbb{R}_0} e^{i\xi X_s}(e^{i\xi\gamma_s(J)} - 1)\, \widetilde{N}(ds, dJ). \end{aligned}$$

Let $M_1, M_2 > 0$ be such that $\sigma_t \leq M_1$ and $\gamma_t(J) \leq M_2$ for all $t \geq 0$ and $J \in \mathbb{R}_0$. By Theorem 5.2.3 we have

$$\mathbb{E}\left(\sup_{t\leq T}\left|\int_0^t i\xi e^{i\xi X_s}\sigma_s\, dW_s\right|\right) \leq C\mathbb{E}\left(\left(\int_0^T \xi^2\sigma_s^2\, ds\right)^{1/2}\right) \leq CTM_1|\xi| < \infty.$$

Furthermore by the Mean-Value Theorem we have

$$\left|e^{i\xi X_s}(e^{i\xi\gamma_s(J)} - 1)\right| \leq \sup_{r\geq 0, J\in\mathbb{R}_0}\left|e^{i\xi X_s} i\xi\gamma_r(J) e^{i\xi\gamma_r(J)}\gamma_s(J)\right| \leq M_2|\xi||\gamma_s(J)|.$$

Hence also

$$\mathbb{E}\left(\sup_{t\leq T}\left|\int_0^t\int_{\mathbb{R}_0} e^{i\xi X_s}(e^{i\xi\gamma_s(J)} - 1)\, \widetilde{N}(ds, dJ)\right|\right) \leq CM_2|\xi|\mathbb{E}\left(\left(\int_0^T N(ds, dJ)\right)^{1/2}\right) < \infty.$$

Therefore taking expectations on both sides and noting that the coefficients of the Lévy-Itô process are deterministic, it follows that

$$\phi_{X_t}(\xi) = \phi_{X_0}(\xi) + \int_0^t \phi_{X_s}(\xi)\left(i\xi\mu_s - \frac{1}{2}\sigma_s^2\xi^2 + \int_{\mathbb{R}_0} e^{i\xi\gamma_s(J)} - 1 - i\xi\gamma_s(J)\, d\nu(J)\right) ds.$$

By the Lebesgue Differentiation Theorem we have for almost everywhere $t \geq 0$,

$$\frac{d}{dt}\phi_{X_t}(\xi) = \phi_{X_t}(\xi)\left(i\xi\mu_t - \frac{1}{2}\sigma_t^2\xi^2 + \int_{\mathbb{R}_0} e^{i\xi\gamma_t(J)} - 1 - i\xi\gamma_t(J)\, d\nu(J)\right).$$

Therefore we get the generalized version of the Lévy-Khintchine formula

$$\phi_{X_t}(\xi) = \exp\left(i\xi\mu_t t - \frac{1}{2}\sigma_t^2\xi^2 t + t\int_{\mathbb{R}_0} e^{i\xi\gamma_t(J)} - 1 - i\xi\gamma_t(J)\, d\nu(J)\right). \tag{6.1}$$

First of all we have

$$\left|\int_{\mathbb{R}_0} e^{i\xi\gamma_t(J)} - 1\, d\nu(J)\right| \leq 2\nu(\mathbb{R}_0) < \infty.$$

Hence we get the estimate

$$|\phi_{X_t}(\xi)| \leq \exp\left(-\frac{1}{2}\sigma_t^2\xi^2 t + 2\nu(\mathbb{R}_0)t\right).$$

Since we have for every $n \in \mathbb{N}_0$ that

$$\int_{\mathbb{R}} |\xi|^n \exp\left(-\frac{1}{2}\sigma_t^2\xi^2 t + 2\nu(\mathbb{R}_0)t\right) d\xi < \infty,$$

the result follows by Proposition 6.1.5. □

## 6.2 Fourier cosine series

Another important part of the COS method is the theory of Fourier series representations, which has become quite classical theory by now. The exponential Fourier series representation of a continuous function $f$ on a finite interval $[a, b]$, defines a periodic extension of that function on all of $\mathbb{R}$, since the complex exponentials are periodic. However, if the function is non-periodic, the periodic extension will not be continuous, giving rise to numerical convergence problems known as the Gibb's phenomenom. This will severely slow down the convergence of the series, not only on the endpoints, but also further away from the endpoints.

One solution to this problem is to extend the interval $[a, b]$ to $[2a - b, b]$ and extend the function evenly. We will see that the sine part will vanish, giving a Fourier cosine series, which will be continuous at the end points due to its inherent periodicity.

Following the work of Oosterlee and Grzelak [28], extend $f \in L^2([0, \pi])$ to the even function $\widetilde{f} \in L^2([-\pi, \pi])$ where $\widetilde{f}(x) = f(|x|)$. Then we have the complex Fourier series of $\widetilde{f}$

$$\widetilde{f}(x) = \sum_{n \in \mathbb{Z}} \widehat{\widetilde{f}}(n) \exp(inx),$$

with coefficients

$$\widehat{\widetilde{f}}(n) = \frac{1}{2\pi} \int_{-\pi}^{\pi} \widetilde{f}(y) \exp(-iny)\, dy = \frac{1}{2\pi} \int_{-\pi}^{\pi} \widetilde{f}(y) \cos(ny)\, dy = \frac{1}{\pi} \int_{0}^{\pi} f(y) \cos(ny)\, dy,$$

since $\widetilde{f}$ is an even function.

Let $f \in L^2([a, b])$ and $\widetilde{f} \in L^2([2a - b, b])$ its even extension. Then we can do the change of variables

$$\theta = \frac{y - a}{b - a}\pi, \qquad y = \frac{b - a}{\pi}\theta + a,$$

to get

$$\widetilde{f}(y) = \sum_{n \in \mathbb{Z}} \widehat{\widetilde{f}}(n) \exp\left(in\pi \frac{y - a}{b - a}\right),$$

with coefficients

$$\widehat{\widetilde{f}}(n) = \frac{1}{2(b - a)} \int_{2a-b}^{b} \widetilde{f}(y) \exp\left(-in\pi \frac{y - a}{b - a}\right) dy = \frac{1}{b - a} \int_{a}^{b} f(y) \cos\left(n\pi \frac{y - a}{b - a}\right) dy.$$

By the even symmetry of the cos function, $\widehat{\widetilde{f}}(-n) = \widehat{\widetilde{f}}(n)$ and hence we have

$$\begin{aligned}
\widetilde{f}(y) &= \sum_{n \in \mathbb{Z}} \widehat{\widetilde{f}}(n) \exp\left(in\pi \frac{y - a}{b - a}\right) \\
&= \widehat{\widetilde{f}}(0) + \sum_{n=1}^{\infty} \widehat{\widetilde{f}}(n) \exp\left(in\pi \frac{y - a}{b - a}\right) + \sum_{n=1}^{\infty} \widehat{\widetilde{f}}(-n) \exp\left(-in\pi \frac{y - a}{b - a}\right) \\
&= \widehat{\widetilde{f}}(0) + \sum_{n=1}^{\infty} \widehat{\widetilde{f}}(n) \left(\exp\left(in\pi \frac{y - a}{b - a}\right) + \exp\left(-in\pi \frac{y - a}{b - a}\right)\right) \\
&= \widehat{\widetilde{f}}(0) + \sum_{n=1}^{\infty} 2\widehat{\widetilde{f}}(n) \cos\left(n\pi \frac{y - a}{b - a}\right)
\end{aligned}$$

Therefore we have for $f$,

$$f(y) = \widetilde{f}(y) = \widehat{\widetilde{f}}(0) + \sum_{n=1}^{\infty} 2\widehat{\widetilde{f}}(n) \cos\left(n\pi \frac{y - a}{b - a}\right).$$

Define for $n \in \mathbb{N}_0$ the coefficients

$$\mathcal{F}_n = \frac{2}{b - a} \int_{a}^{b} f(y) \cos\left(n\pi \frac{y - a}{b - a}\right) dy,$$

such that

$$f(y) = {\sum_{n=0}^{\infty}}' \mathcal{F}_n \cos\left(n\pi \frac{y-a}{b-a}\right).$$

where the prime in $\sum'$ indicates that the first term of the summation is halved. Let $f, g \in L^2([a,b])$, then by Parseval's relation (see Grafakos [19, Proposition 3.2.7]), we have that

$$\int_a^b f(x)g(x)\,dx = \frac{b-a}{2} {\sum_{n=0}^{\infty}}' \mathcal{F}_n \mathcal{G}_n, \tag{6.2}$$

where

$$\mathcal{F}_n = \frac{2}{b-a}\int_a^b f(y)\cos\left(in\pi\frac{y-a}{b-a}\right)\,dy,$$
$$\mathcal{G}_n = \frac{2}{b-a}\int_a^b g(y)\cos\left(in\pi\frac{y-a}{b-a}\right)\,dy.$$

The summation on the right-hand side of (6.2) has rarely an elementary analytic closed form, hence we truncate the series for numerical methods. One can wonder how many terms we need to get an accurate approximation and how fast the partial sums converges. For the latter question we have a very satisfying relation between the convergence speed and the smoothness of the function.

**Theorem 6.2.1.** *Let $k \in \mathbb{N}$ and suppose that $f \in C^k([a,b])$. Then for the Fourier cosine coefficients we have the asymptotic relation $\mathcal{F}_n = \mathcal{O}(n^{-k})$ as $k \to \infty$.*

*Proof.*
Recall that the Fourier cosine expansion of $f$ on $[a,b]$ is equal to the exponential Fourier expansion of the even extension $\widetilde{f}$ to $[2a-b, b]$. Hence we will first show this result for $\widetilde{f}$. Write for convenience $f = \widetilde{f}$ and let $[a,b] = [0,\pi]$, using integration by parts and noting that $f(-\pi) = f(\pi)$, due to the even extension, we have for $n \in \mathbb{Z}\setminus\{0\}$,

$$\begin{aligned}\widehat{f}(n) &= \frac{1}{2\pi}\int_{-\pi}^{\pi} f(x)e^{-inx}\,dx = \frac{1}{2\pi}\left[\frac{1}{-in}f(x)e^{-inx}\right]_{x=-\pi}^{\pi} + \frac{1}{in}\frac{1}{2\pi}\int_{-\pi}^{\pi} f'(x)e^{-inx}\,dx \\ &= \frac{1}{in}\frac{1}{2\pi}\int_{-\pi}^{\pi} f'(x)e^{-inx}\,dx = \frac{1}{in}\widehat{f'}(n).\end{aligned}$$

Since $f$ is $2\pi$-periodic, every derivative of $f$ is $2\pi$-periodic as well. Indeed,

$$f'(x+2\pi) = \lim_{h\to 0}\frac{f(x+2\pi+h) - f(x+2\pi)}{h} = \lim_{h\to 0}\frac{f(x+h)-f(x)}{h} = f'(x).$$

Then by induction the result follows for $\widetilde{f}$,

$$\widehat{f}(n) = \frac{1}{(in)^k}\widehat{f^{(k)}}(n), \qquad |\widehat{f}(n)| \le \frac{1}{|n|^k}\frac{1}{2\pi}\left\|f^{(k)}\right\|_{L^1}.$$

Therefore we also have the result for the Fourier cosine coefficients,

$$|\mathcal{F}_n| = |2\widehat{\widetilde{f}}(n)| \le \frac{1}{|n^k|}\frac{2}{\pi}\left\|f^{(k)}\right\|_{L^1}.$$

□

## 6.3 COS approximation formulae

Let $b, \sigma : [0,T]\times\mathbb{R}\to\mathbb{R}$ and $\beta : [0,T]\times\mathbb{R}\times\mathbb{R}\to\mathbb{R}$ be deterministic and bounded functions and let $\nu$ be a finite Lévy measure. Furthermore assume $\sigma_t \neq 0$ for all $t \geq 0$. We will apply the COS method to conditional expectations with respect to stochastic processes of the form

$$\begin{aligned}X_t = x &+ \mu_s(x)(t-s) + \sigma_s(x)(W_t - W_s) \\ &+ \sum_{\tau=N_s+1}^{N_t}\beta_s(x, J_\tau) - (t-s)\int_{\mathbb{R}_0}\beta_s(x,J)\,d\nu(J),\end{aligned} \tag{6.3}$$

for $t \geq s$ and some deterministic $x \in \mathbb{R}$. Define the shorthand notations

$$\begin{aligned} \Delta t_{m,k} &= t_{m+k} - t_m, & \Delta t_{r,s} &= t_s - t_r, \\ \Delta W_{m,k} &= W_{t_{m+k}} - W_{t_m}, & \Delta W_{r,s} &= W_s - W_r, \\ \Delta \widetilde{N}^*_{m,k} &= \int_{t_m}^{t_{m+k}} \int_{\mathbb{R}_0} \eta(J)\, \widetilde{N}(dt, dJ), & \Delta \widetilde{N}^*_{r,s} &= \int_r^s \int_{\mathbb{R}_0} \eta(J)\, \widetilde{N}(dt, dJ), \end{aligned} \tag{6.4}$$

where $\eta$ is some deterministic and bounded function. Then we can rewrite (6.3) into

$$\begin{aligned} X_t = x &+ \mu_s(x)\Delta t_{s,t} + \sigma_s(x)\Delta W_{s,t} \\ &+ \sum_{\tau=N_s+1}^{N_t} \beta_s(x, J_\tau) - \Delta t_{s,t} \int_{\mathbb{R}_0} \beta_s(x, J)\, d\nu(J) \end{aligned} \tag{6.5}$$

For the procedure of numerically evaluating conditional expectations of the form $\mathbb{E}(v(t, X_t)|X_s = x)$ we will follow Ruijter and Oosterlee [36]. Write $\mathbb{E}^x_s(v(t, X_t)) := \mathbb{E}(v(t, X_t)|X_s = x)$, and define $p_t(y|x)$ as the conditional distribution of $X_t|X_s = x$. We know that $p_t(y|x) \in C^\infty$ for all $t \geq s$ by Theorem 6.1.8, then

$$I := \mathbb{E}^x_s(v(t, X_t)) = \int_{\mathbb{R}} v(t, y) p_t(y|x)\, dy,$$

is well-defined for $v$ deterministic and not too irregular. So assume that $v(t, y)p_t(y|x)$ decays to zero rapidly as $|y| \to \infty$ for all $x$. Now fix $x$, then let $\varepsilon > 0$ and define $a < b$ such that

$$\left| \int_{\mathbb{R}\setminus[a,b]} v(t, y) p_t(y|x)\, dy \right| < \varepsilon.$$

We will look later into choosing a suitable $a, b$ such that this holds. Since $v(t, y)p_t(y|x)$ vanishes in infinity rapidly, we can truncate to a not too big finite interval $[a, b]$ without losing a significant amount of accuracy.

Let $I_i$ be a further approximation of $I$, to keep track of all the numerical errors we will make in each step. So we get

$$I_1 = \int_a^b v(t, y) p_t(y|x)\, dy.$$

Then given $x$ and $t \geq s$, we could see $y \mapsto p_t(y|x)$ as a function on the finite interval $[a, b]$.

Going back to our integral $I_1$, if we assume that $y \mapsto v(t, y)$ and $y \mapsto p_t(y|x)$ are both in $L^2([a, b])$, then from the discussion in the previous section we have that

$$I_1 = \frac{b-a}{2} {\sum_{k=0}^{\infty}}' \mathcal{V}_k(t) \mathcal{P}_k(x),$$

where

$$\begin{aligned} \mathcal{V}_k(t) &= \frac{2}{b-a} \int_a^b v(t, y) \cos\left(k\pi \frac{y-a}{b-a}\right) dy, \\ \mathcal{P}_k(x) &= \frac{2}{b-a} \int_a^b p_t(y|x) \cos\left(k\pi \frac{y-a}{b-a}\right) dy. \end{aligned}$$

Since the coefficients $\mathcal{P}_k(x)$ vanish at a faster rate than polynomially due to Theorem 6.2.1 we can truncate the series, to get the approximation

$$I_2 = \frac{b-a}{2} {\sum_{k=0}^{\mathcal{N}-1}}' \mathcal{V}_k(t_{m+1}) \mathcal{P}_k(x).$$

The reason why we rewrite the conditional expectation in a Fourier cosine expansion, is that we can now approximate $\mathcal{P}_k(x)$ with the characteristic function, which we will assume to be given. Indeed,

$$\begin{aligned}\mathcal{P}_k(x) &\approx \frac{2}{b-a}\int_{\mathbb{R}} p_t(y|x)\cos\left(k\pi\frac{y-a}{b-a}\right)dy \\ &= \frac{2}{b-a}\mathrm{Re}\left(\int_{\mathbb{R}}\exp\left(ik\pi\frac{y-a}{b-a}\right)p_t(y|x)\,dy\right) \\ &= \frac{2}{b-a}\mathrm{Re}\left(\exp\left(ik\pi\frac{-a}{b-a}\right)\int_{\mathbb{R}}\exp\left(i\frac{k\pi}{b-a}y\right)p_t(y|x)\,dy\right) \\ &= \frac{2}{b-a}\mathrm{Re}\left(\exp\left(ik\pi\frac{-a}{b-a}\right)\varphi\left(\frac{k\pi}{b-a}\middle|\,x\right)\right),\end{aligned}$$

where $\varphi(\cdot|x)$ is the conditional characteristic function of $X_t$, given $X_s = x$. By the generalized Lévy-Khintchine formula in the proof of Theorem 6.1.8 we have

$$\varphi(\xi|x) = \exp\left(i\xi x + \Delta t_{s,t}\left(i\xi\mu_s(x) - \frac{1}{2}\sigma_s^2(x)\xi^2 + \int_{\mathbb{R}_0} e^{i\xi\beta_s(x,J)} - 1 - i\xi\beta_s(x,J)\,d\nu(J)\right)\right)$$

Define $\phi(\xi) = \varphi(\xi|0)$, then $\varphi(\xi|x) = \phi(\xi)e^{i\xi x}$ and hence we get

$$\mathcal{P}_k(x) \approx \frac{2}{b-a}\mathrm{Re}\left(\phi\left(\frac{k\pi}{b-a}\right)e^{ik\pi\frac{x-a}{b-a}}\right),$$

and so we get the following approximation

$$I_3 = {\sum_{k=0}^{\mathcal{N}-1}}{}'\,\mathcal{V}_k(t)\mathrm{Re}\left(\phi\left(\frac{k\pi}{b-a}\right)e^{ik\pi\frac{x-a}{b-a}}\right) =: {\sum_{k=0}^{\mathcal{N}-1}}{}'\,\mathcal{V}_k(t)\Phi_k^Y(x). \tag{6.6}$$

Besides $\mathbb{E}_s^x(v(t,X_t))$, we also want to approximate the conditional expectations $\mathbb{E}_s^x(v(t,X_t)\Delta W_{s,t})$ and $\mathbb{E}_s^x(v(t,X_t)\Delta\widetilde{N}^*_{s,t})$. For $\mathbb{E}_s^x(v(t,X_t)\Delta W_{s,t})$ we have the following.

$$\mathbb{E}_m^x\left(v(t,X_t)\Delta W_{s,t}\right) \approx {\sum_{k=0}^{\mathcal{N}-1}}{}'\,\mathcal{V}_k(t)\mathbb{E}_m^x\left(\cos\left(k\pi\frac{X_t-a}{b-a}\right)\Delta W_{s,t}\right)$$

Now for $\mathbb{E}_m^x\left(\cos\left(k\pi\frac{X_t-a}{b-a}\right)\Delta W_{s,t}\right)$, we have by integration by parts

$$\begin{aligned}&\mathbb{E}_m^x\left(\cos(\xi(X_t-a))\Delta W_{s,t}\right) \\ &\quad= \mathrm{Re}\left(\mathbb{E}_m^x\left(\exp\left(i\xi\left((x-a)+\mu_s(x)\Delta t_{s,t}+\sigma_s(x)\Delta W_{s,t}+\int_s^t\beta_s(x,J)\,\widetilde{N}(ds,dJ)\right)\right)\right)\right) \\ &\quad= \mathrm{Re}\left(\frac{1}{\sqrt{2\pi\Delta t_{s,t}}}\int_{\mathbb{R}}\exp\left(i\xi\left((x-a)+\mu_s(x)\Delta t_{s,t}+\sigma_s(x)\zeta+\int_s^t\beta_s(x,J)\,\widetilde{N}(ds,dJ)\right)\right)\right. \\ &\qquad\qquad \left.\zeta e^{-\frac{\zeta^2}{2\Delta t_{s,t}}}\,d\zeta\right) \\ &\quad= \mathrm{Re}\left(\frac{i\xi\sigma_s(x)\Delta t_{s,t}}{\sqrt{2\pi\Delta t_{s,t}}}\int_{\mathbb{R}}\exp\left(i\xi\left((x-a)+\mu_s(x)\Delta t_{s,t}+\sigma_s(x)\zeta+\int_s^t\beta_s(x,J)\,\widetilde{N}(ds,dJ)\right)\right)\right. \\ &\qquad\qquad \left.e^{-\frac{\zeta^2}{2\Delta t_{s,t}}}\,d\zeta\right) \\ &\quad= \sigma_s(x)\Delta t_{s,t}\mathrm{Re}\left(i\xi\phi(\xi)e^{i\xi(x-a)}\right).\end{aligned}$$

Then we get the COS approximation formula

$$\begin{aligned}&\mathbb{E}_m^x(v(t,X_t)\Delta W_{s,t}) \\ &\quad\approx {\sum_{k=0}^{\mathcal{N}-1}}{}'\,\mathcal{V}_k(t)\sigma_s(x)\Delta t_{s,t}\mathrm{Re}\left(i\frac{k\pi}{b-a}\phi\left(\frac{k\pi}{b-a}\right)e^{ik\pi\frac{x-a}{b-a}}\right) =: {\sum_{k=0}^{\mathcal{N}-1}}{}'\,\mathcal{V}_k(t)\Phi_k^Z(x).\end{aligned} \tag{6.7}$$

Finally, we also want to calculate $\mathbb{E}_s^x(v(t,X_t)\Delta\widetilde{N}^*_{s,t})$, for which we again have to calculate $\mathbb{E}_m^x\left(\cos\left(k\pi\frac{X_t-a}{b-a}\right)\Delta N^*_{s,t}\right)$, where

$$\Delta N^*_{s,t} = \int_s^t\int_{\mathbb{R}_0}\eta(J)N(dr,dJ).$$

Then we have

$$\begin{aligned}
&\mathrm{Re}\left(\mathbb{E}_m^x\left(\Delta N^*_{s,t}\exp\left(i\xi(X_t-a)\right)\right)\right)\\
&\quad= \mathrm{Re}\Bigg(\mathbb{E}_m^x\left(\exp\left(i\xi\left((x-a)+\mu_s(x)\Delta t_{s,t}+\sigma_s(x)\Delta W_{s,t}-\Delta t_{s,t}\int_{\mathbb{R}_0}\beta_s(x,J)\,d\nu(J)\right)\right)\right)\\
&\qquad\quad \mathbb{E}_m^x\left(\Delta N^*_{s,t}\exp\left(i\xi\sum_{\tau=N_s+1}^{N_t}\beta_s(x,J_\tau)\right)\right)\Bigg)\\
&\quad= \mathrm{Re}\Bigg(\exp\left(i\xi\left((x-a)+\mu_s(x)\Delta t_{s,t}-\Delta t_{s,t}\int_{\mathbb{R}_0}\beta_s(x,J)\,d\nu(J)\right)-\frac{1}{2}\xi^2\sigma_s(x)^2\Delta t_{s,t}\right)\\
&\qquad\quad \mathbb{E}_m^x\left(\sum_{\tau=N_s+1}^{N_t}\eta(J_\tau)\exp\left(i\xi\sum_{\tau=N_s+1}^{N_t}\beta_s(x,J_\tau)\right)\right)\Bigg).
\end{aligned}$$

Now the final conditional expectation can be solved by iterated conditioning on the Poisson process, and we will assume $\varphi_J(\xi)=\mathbb{E}(\exp(i\xi\beta_s(x,J)))$ and $\lambda=\nu(\mathbb{R}_0)$ are known a priori in this thesis, then

$$\begin{aligned}
&\mathbb{E}_m^x\left(\sum_{\tau=N_s+1}^{N_t}\eta(J_\tau)\exp\left(i\xi\sum_{\tau=N_s+1}^{N_t}\beta_s(x,J_\tau)\right)\right)\\
&\quad=\sum_{q=0}^{\infty}e^{-\lambda\Delta t_{s,t}}\frac{(\lambda\Delta t_{s,t})^q}{q!}\mathbb{E}\left(\sum_{\tau=1}^{q}\eta(J_\tau)\exp\left(i\xi\sum_{\tau=1}^{q}\beta_s(x,J_\tau)\right)\right)\\
&\quad=\sum_{q=0}^{\infty}e^{-\lambda\Delta t_{s,t}}\frac{(\lambda\Delta t_{s,t})^q}{q!}\sum_{\tau=1}^{q}\mathbb{E}\left(\eta(J_\tau)\exp\left(i\xi\beta_s(x,J_\tau)\right)\right)\mathbb{E}\left(\exp\left(i\xi\sum_{\rho=1,\rho\neq\tau}^{q}\beta_s(x,J_\rho)\right)\right)\\
&\quad=\sum_{q=0}^{\infty}e^{-\lambda\Delta t_{s,t}}\frac{(\lambda\Delta t_{s,t})^q}{q!}q\mathbb{E}\left(\eta(J)\exp\left(i\xi\beta_s(x,J)\right)\right)\varphi_J(\xi)^{q-1}\\
&\quad=e^{-\lambda\Delta t_{s,t}(\varphi_J(\xi)-1)}\lambda\Delta t_{s,t}\mathbb{E}(\eta(J)\exp(i\xi\beta_s(x,J))).
\end{aligned}$$

Then we have

$$\mathbb{E}_m^x\left(\Delta\widetilde{N}^*_{s,t}\exp\left(ik\pi\frac{X_t-a}{b-a}\right)\right)=\mathrm{Re}\left(\phi\left(\frac{k\pi}{b-a}\right)e^{-ik\pi\frac{x-a}{b-a}}\lambda\Delta t_{s,t}\mathbb{E}\left(\eta(J)\exp\left(i\frac{k\pi}{b-a}\beta_s(x,J)\right)\right)\right).$$

Therefore we have the last COS approximation formula

$$\begin{aligned}
&\mathbb{E}_m^x(v(t,X_t)\Delta\widetilde{N}^*_{s,t})\\
&\quad\approx\sum_{k=0}^{\mathcal{N}-1}{}'\mathcal{V}_k(t)\mathrm{Re}\left(\phi\left(\frac{k\pi}{b-a}\right)e^{-ik\pi\frac{x-a}{b-a}}\lambda\Delta t_{s,t}\mathbb{E}\left(\eta(J)\left(\exp\left(i\frac{k\pi}{b-a}\beta_s(x,J)\right)-1\right)\right)\right) \qquad (6.8)\\
&\quad=:\sum_{k=0}^{\mathcal{N}-1}{}'\mathcal{V}_k(t)\Phi_k^{\Gamma}(x).
\end{aligned}$$

The expectation $\mathbb{E}\left(\eta(J)\left(\exp\left(i\frac{k\pi}{b-a}\beta_s(x,J)\right)-1\right)\right)$ still has to be approximated for general $\eta$ and $\beta$. This can again be done by the COS method, where instead of the conditional density, we now have the Lévy density. Since this expectation will be used extensively, it should be computed with high accuracy, such that the numerical error in the scheme does not amplify later on. Again recall that

$\varphi_J(\xi) = \mathbb{E}(\exp(i\xi\beta_s(x, J)))$,

$$\begin{aligned}
&\mathbb{E}\left(\eta(J)\left(\exp\left(i\frac{k\pi}{b-a}\beta_s(x,J)\right)-1\right)\right) \\
&\approx \sum_{k=0}^{\mathcal{N}-1}{}' \left(\int_a^b \eta(y)\left(\exp\left(i\frac{k\pi}{b-a}\beta_s(x,y)\right)-1\right)\cos\left(k\pi\frac{y-a}{b-a}\right)dy\right) \\
&\qquad\qquad \left(\frac{2}{b-a}\int_a^b \cos\left(k\pi\frac{y-a}{b-a}\right)\frac{d\nu(y)}{\lambda}\right) \\
&\approx \sum_{k=0}^{\mathcal{N}-1}{}' \mathcal{N}_k \mathcal{J}_k,
\end{aligned} \tag{6.9}$$

where

$$\begin{aligned}
\mathcal{N}_k &= \int_a^b \eta(y)\left(\exp\left(i\frac{k\pi}{b-a}\beta_s(x,y)\right)-1\right)\cos\left(k\pi\frac{y-a}{b-a}\right)dy, \\
\mathcal{J}_k &= \frac{2}{b-a}\mathrm{Re}\left(\varphi_J\left(\frac{k\pi}{b-a}\right)e^{ik\pi\frac{-a}{b-a}}\right).
\end{aligned}$$

The $\mathcal{N}_k$ can be efficiently computed with a Discrete Cosine Transform (DCT). We will come back to this later. Note that the $a$ and $b$ should not necessarily be chosen to be the same as with the other COS approximation formulae, as we have a different density here, so we also have a different mass distribution. Furthermore the convergence speed relies on the smoothness of the Lévy density $\nu$.

CHAPTER 7

# BCOS Method

## 7.1 Numerical discretisation FBSDEJs

In the filtered probability space $(\Omega, \mathcal{F}, \mathbb{F}, \mathbb{P})$ with index set $\mathcal{T} = [0, T]$ for some $T > 0$ we will look at the following forward backward stochastic differential equations with jumps (FBSDEJs) which are decoupled

$$\begin{cases} X_t = X_0 + \int_0^t \mu_s(X_s)\, ds + \int_0^t \sigma_s(X_s)\, dW_s + \int_0^t \int_{\mathbb{R}_0} \beta_s(X_{s-}, J)\, \widetilde{N}(ds, dJ), & \text{(FSDEJ)} \\ Y_t = \xi + \int_t^T f_s(\Theta_s)\, ds - \int_t^T Z_s\, dW_s - \int_t^T \int_{\mathbb{R}_0} U_s(J)\, \widetilde{N}(ds, dJ), & \text{(BSDEJ)} \end{cases}$$

where $\Theta_t := (X_t, Y_t, Z_t, \Gamma_t)$ such that for a given bounded function $\eta : \mathbb{R}_0 \to \mathbb{R}$ we have

$$\Gamma_s = \int_{\mathbb{R}_0} U_s(J)\eta(J)\, d\nu(J).$$

Under global Lipschitz and growth conditions, we have already shown the FBSDEJ has a unique solution.

Recall from Theorem 4.4.12 that for a given $g(t, x) \in C^{1,2}$ we have that

$$\begin{aligned} g(t, X_t) = g(t_0, X_{t_0}) &+ \int_{t_0}^t \mathcal{L}g(s, X_s)\, ds + \int_{t_0}^t \frac{\partial g}{\partial x}\sigma_s(X_s)\, dW_s \\ &+ \int_{t_0}^t \int_{\mathbb{R}_0} \left(g(s, X_{s-} + \beta_s(X_{s-}, J)) - g(s, X_{s-})\right) \widetilde{N}(ds, dJ), \end{aligned}$$

where $\mathcal{L}$ is the second-order partial-integro differential operator

$$\begin{aligned} \mathcal{L}g(s, x) = &\frac{\partial g}{\partial t}(s, x) + \frac{\partial g}{\partial x}(s, x)\mu_s(x) + \frac{1}{2}\frac{\partial^2 g}{\partial x^2}(s, x)\sigma_s^2(x) \\ &+ \int_{\mathbb{R}_0} \left( g(s, x + \beta_s(x, J)) - g(s, x) - \frac{\partial g}{\partial x}(s, x)\beta_s(x, J) \right) d\nu(J) \end{aligned}$$

Now consider the following partial-integro differential equation (PIDE):

$$\begin{cases} \mathcal{L}u + f_t(x, u, \sigma\nabla u, \mathcal{M}u) = 0, & (t, x) \in [0, T] \times \mathbb{R}, \\ u(T, x) = g(x), & x \in \mathbb{R}, \end{cases} \tag{7.1}$$

where $\mathcal{M}$ is the integral operator

$$\mathcal{M}u(t, x) = \int_{\mathbb{R}_0} (u(t, x + \beta_t(x, J)) - u(t, x))\eta(J)\, d\nu(J),$$

and $\nabla$ is the partial derivative with respect to $x$. We define a viscosity solution of (7.1) to be the following

**Definition 7.1.1.** *We say that $u \in C([0, T] \times \mathbb{R}; \mathbb{R})$ is*

*(i) a viscosity subsolution of* (7.1) *if*

$$u(T, x) \leq g(x), \qquad x \in \mathbb{R},$$

*and if, for any $\varphi \in C^2([0, T] \times \mathbb{R}; \mathbb{R})$, whenever $(t, x) \in [0, T] \times \mathbb{R}$ is a global maxima of $u - \varphi$,*

$$\mathcal{L}\varphi(t, x) + f_t(x, u(t, x), \sigma\nabla\varphi(t, x), \mathcal{M}\varphi(t, x)) \leq 0.$$

*(ii) a viscosity supersolution of* (7.1) *if*

$$u(T,x) \geq g(x), \qquad x \in \mathbb{R},$$

*and if, for any* $\varphi \in C^2([0,T]\times\mathbb{R};\mathbb{R})$*, whenever* $(t,x) \in [0,T]\times\mathbb{R}$ *is a global minima of* $u-\varphi$*,*

$$\mathcal{L}\varphi(t,x) + f_t(x, u(t,x), \sigma\nabla\varphi(t,x), \mathcal{M}\varphi(t,x)) \geq 0.$$

*(iii) a viscosity solution of* (7.1) *if it is both a viscosity subsolution and a viscosity supersolution of* (7.1)*.*

If $u \in C^{1,2}$, is a classical solution of (7.1) we can immediately see that, $u(t,X_t) = Y_t$ solves the FBSDEJ, by the Itô formula, where furthermore we have

$$\begin{cases} Z_t = \sigma_t(X_t)\nabla u(t,X_t), \\ U_t = u(t, X_{t-} + \beta_t(X_{t-}, J)) - u(t, X_{t-}), \\ \Gamma_t = \mathcal{M}u(t,X_t), \end{cases}$$

and where the quadruple $(X_t, Y_t, Z_t, \Gamma_t)$ is the solution of the FBSDEJs with terminal condition $\xi = g(X_T)$. However, given a solution $Y$, we in general do not have that $u$ is in $C^{1,2}$, but under some weak regularity conditions we do know that $u \in C([0,T]\times\mathbb{R};\mathbb{R})$ as shown in Barles, Buckdahn and Pardoux [2, Proposition 2.5]. Hence it is more natural to look at a viscosity solution of 7.1 instead of a classical solution. To prove that $u$ is indeed a viscosity solution, we need a comparison result for the FBSDEJs.

**Proposition 7.1.2.** *Suppose* $f$ *and* $\gamma$ *satisfy the regularity conditions in [2, Propositon 2.6] and suppose that* $\xi, \xi' \in L^2$ *are two terminal conditions. Then denote* $(Y,Z,U) \in \mathcal{S}^2 \times L^2_W \times L^2_{\mathcal{N}}$ *and* $(Y',Z',U') \in \mathcal{S}^2 \times L^2_W \times L^2_{\tilde{N}}$ *as the unique solutions of the FBSDEJ with terminal conditions* $\xi$ *and* $\xi'$ *respectively. If* $\xi \geq \xi'$*, it follows that*

$$Y_t \geq Y'_t, \qquad 0 \leq t \leq T.$$

Then under these regularity conditions we can prove that indeed $u$ is the unique viscosity solution of (7.1) and hence we have a connection between the FBSDEJs and the PIDEs. This connection is often known in the literature as the *Feynman-Kac representation* and is not limited to just FBSDEJs, but exists for nearly any type of stochastic differential equations.

*Remark* 7.1.3. We made the seemingly odd assumption that $f$ can only depend on $U$ through $\Gamma$, however we have a simple counterexample for which the comparison result in Proposition 7.1.2 fails to hold when this is not the case. Let $\nu = \delta_1$ and $f_t(\omega, x, y, z, u) = -2u(1)$. Then

$$N_t = \int_0^t \int_{\mathbb{R}_0} N(ds, dJ),$$

is a standard Poisson process and if we choose $\xi = N_T$ and $\xi' = 0$, then we have the solutions

$$\begin{aligned} (Y_t, Z_t, U_t) &= (N_t - (T-t), 0, \mathbf{1}_{J=1}), \\ (Y'_t, Z'_t, U'_t) &= (0,0,0). \end{aligned}$$

Clearly we have $\xi \geq \xi'$, but $\mathbb{P}(Y_t < Y'_t) = \mathbb{P}(N_t < T-t) \geq \mathbb{P}(N_t = 0) = e^{-1} > 0$ for all $0 \leq t < T$. Hence the assumption on $f$ is necessary.

### 7.1.1 Semi-discretisation

It turns out that the numerical analysis of FSDEJs is highly complex, due to its complex solution structure. One of the perhaps more surprising finds is that we can discretise the FSDEJ with an implicit Euler scheme without theoretically losing any accuracy over using the exact solution. We will explain and prove this through an application of semigroup theory.

**Definition 7.1.4.** *Let* $X$ *be a Lévy-Itô process with finite Lévy measure* $\nu$*, for a given measurable function* $g : [0,T]\times\mathbb{R} \to \mathbb{R}$*, the generator* $A^x_t$ *of* $X$ *on* $g$ *is defined by*

$$A^x_t g(\cdot, X_\cdot) = \lim_{s \downarrow t} \frac{\mathbb{E}^x_t(g(s,X_s)) - g(t,x)}{s-t}.$$

Given $g(t,x) \in C^{1,2}$ and $\sigma \in L^2_W, \gamma \in L^2_{\widetilde{N}}$, we have by the Itô formula that

$$\mathbb{E}^x_t(g(s,X_s)) = g(t,x) + \int_t^s \mathbb{E}^x_t\left(\mathcal{L}g(r,X_r)\right) dr.$$

Hence we obtain

$$A^x_t g(\cdot, X_\cdot) = \mathcal{L}g(t,x).$$

We could say that $A^x_t$ is a local operator, as its value only depends on the value of $\mathcal{L}g$ in $(t,x)$. Motivated by this locality, we can search for simpler to solve Lévy-Itô processes $\widetilde{X}$ such that $A^x_t g(\cdot,\widetilde{X}_\cdot) = A^x_t g(\cdot, X_\cdot)$. This is the case when we have $\widetilde{\mu}_t(x) = \mu_t(x)$, $\widetilde{\sigma}_t(x) = \sigma_t(x)$ and $\widetilde{\beta}_t(x,J) = \beta_t(x,J)$ for all $J \in \mathbb{R}_0$.

Consider the two processes

$$X_s = x + \int_t^s \mu_r(X_r)\,dr + \int_t^s \sigma_r(X_r)\,dW_r + \int_t^s \int_{\mathbb{R}_0} \beta_r(X_{r-},J)\,\widetilde{N}(dr,dJ),$$
$$\widetilde{X}_s = x + \int_t^s \widetilde{\mu}_r(\widetilde{X}_r)\,dr + \int_t^s \widetilde{\sigma}_r(\widetilde{X}_r)\,dW_r + \int_t^s \int_{\mathbb{R}_0} \widetilde{\beta}_r(\widetilde{X}_{r-},J)\,\widetilde{N}(dr,dJ).$$

We could for example choose an Euler implicit discretisation $\widetilde{\mu}_r(\widetilde{X}_r) = \mu_t(x), \widetilde{\sigma}_r(\widetilde{X}_r) = \sigma_t(x)$ and $\widetilde{\beta}_r(\widetilde{X}_r,J) = \beta_t(x,J)$ for all $r \in [t,s]$. Then the FSDEJ for $\widetilde{X}_s$ becomes

$$\widetilde{X}_s = x + \mu_t(x)(s-t) + \sigma_t(x)(W_s - W_t) + \sum_{\tau=N_t+1}^{N_s} \beta_t(x,J_\tau) - \int_{\mathbb{R}_0} \beta_t(x,J)\,d\nu(J)(s-t).$$

It follows that indeed $A^x_t g(\cdot,\widetilde{X}_\cdot) = A^x_t g(\cdot,X_\cdot)$, we will see in a moment why this is important.

Let $0 = t_0 < t_1 < t_2 < \cdots < t_{\mathcal{M}} = T$, be a partition. Then recalling the shorthand notations in (6.4). Given $m$, we have for $t \geq t_m$ that

$$\mathbb{E}^x_m(Y_m) = \mathbb{E}^x_m(Y_t) + \int_{t_m}^t \mathbb{E}^x_m(f_s(\Theta_s))\,ds, \tag{7.2}$$

then we have what we call a *reference ODE* by the Lebesgue differentiation theorem,

$$\frac{d}{dt}\mathbb{E}^x_m(Y_t)\bigg|_{t=t_m} = -f_{t_m}(x, y(t_m,x), z(t_m,x), \gamma(t_m,x)).$$

Similarly we want to have reference ODEs for the $Z$-process and the $\Gamma$-process. To free up $Z$ and $\Gamma$ from their stochastic integrals, we will multiply with $\Delta W_{t_m,t}$ and $\widetilde{N}^*_{t_m,t}$ respectively. Then it follows by the Itô isometry that

$$0 = \mathbb{E}^x_m(Y_m \Delta W_{t_m,t}) = \mathbb{E}^x_m(Y_t \Delta W_{t_m,t}) + \int_{t_m}^t \mathbb{E}^x_m(f_s(\Theta_s)\Delta W_{t_m,s})\,ds - \int_{t_m}^t \mathbb{E}^x_m(Z_s)\,ds, \tag{7.3}$$

such that we have the reference ODE

$$z(t_m,x) = \frac{d}{dt}\mathbb{E}^x_m(y(t,X_t)\Delta W_{t_m,t})\bigg|_{t=t_m}.$$

And also

$$0 = \mathbb{E}^x_m(Y_m \Delta\widetilde{N}^*_{t_m,t}) = \mathbb{E}^x_m(Y_t \Delta\widetilde{N}^*_{t_m,t}) + \int_{t_m}^t \mathbb{E}^x_m(f_s(\Theta_s)\Delta\widetilde{N}^*_{t_m,t})\,ds - \int_{t_m}^t \mathbb{E}^x_m(\Gamma_s)\,ds, \tag{7.4}$$

so that we have the final reference ODE

$$\gamma(t_m,x) = \frac{d}{dt}\mathbb{E}^x_m(y(t,X_t)\Delta\widetilde{N}^*_{t_m,t})\bigg|_{t=t_m}.$$

Motivated by the arguments with the generator $A^x_t g(\cdot,X_\cdot)$, define the Lévy-Itô process $\widetilde{X}^m$ on $[t_m, t_{m+1})$ by

$$\widetilde{X}^m_t = x + \mu_t(x)\Delta t_{t_m,t} + \sigma_t(x)\Delta W_{t_m,t} + \sum_{\tau=N_{t_m}+1}^{N_t} \beta_t(x,J_\tau) - \int_{\mathbb{R}_0} \beta_t(x,J)\,d\nu(J)\Delta t_{t_m,t}.$$

Assuming sufficient smoothness of the function $y(t,x)$, we know that

$$\frac{d}{dt}\mathbb{E}_m^x(y(t,X_t^m))\bigg|_{t=t_m} = A_{t_m}^x y(\cdot,X.) = A_{t_m}^x y(\cdot,\widetilde{X}.) = \frac{d}{dt}\mathbb{E}_m^x(y(t,\widetilde{X}_t^m))\bigg|_{t=t_m}.$$

Through similar arguments, we can argue that the same holds for the other two reference ODEs, hence we get

$$\begin{aligned}\frac{d}{dt}\mathbb{E}_m^x(y(t,\widetilde{X}_t^m))\bigg|_{t=t_m} &= -f_{t_m}(x,y(t_m,x),z(t_m,x),\gamma(t_m,x)),\\ z(t_m,x) &= \frac{d}{dt}\mathbb{E}_m^x(y(t,\widetilde{X}_t^m)\Delta W_{t_m,t})\bigg|_{t=t_m},\\ \gamma(t_m,x) &= \frac{d}{dt}\mathbb{E}_m^x(y(t,\widetilde{X}_t^m)\Delta\widetilde{N}_{t_m,t}^*)\bigg|_{t=t_m}.\end{aligned}$$

Therefore the solution of the FBSDEJ rests on the solution of the above system of first order ODEs. Ruijter and Oosterlee [36] proposed a theta discretisation scheme to solve the integral equations (7.2), (7.3) and (7.4). It is well known that theta discretisation schemes only have a first order convergence when applied to ODEs, with the exception of the Crank-Nicolson scheme which corresponds to $\theta = 1/2$ and has second order convergence. Let $(Y_m^\Delta, Z_m^\Delta, \Gamma_m^\Delta)$ be the numerical approximation of $(Y,Z,\Gamma)$ at time-step $t_m$. Then we get for the theta-discretisation the following scheme.

**Scheme 1** Semi-discrete theta scheme

**for** $m = \mathscr{M}-1,\dots,0$ **do**
$Z_m^\Delta = -\theta_2^{-1}(1-\theta_2)\mathbb{E}_m(Z_{m+1}^\Delta) + (\Delta t_{m,1}\theta_2)^{-1}\mathbb{E}_m(Y_{m+1}^\Delta\Delta W_{m,1}) + \theta_2^{-1}(1-\theta_2)\mathbb{E}_m(f_{t_{m+1}}(\Theta_{m+1}^\Delta)\Delta W_{m,1})$
$\Gamma_m^\Delta = -\theta_3^{-1}(1-\theta_3)\mathbb{E}_m(\Gamma_{m+1}^\Delta) + (\Delta t_{m,1}\theta_3)^{-1}\mathbb{E}_m(Y_{m+1}^\Delta\Delta\widetilde{N}_{m,1}^*) + \theta_3^{-1}(1-\theta_3)\mathbb{E}_m(f_{t_{m+1}}(\Theta_{m+1}^\Delta)\Delta\widetilde{N}_{m,1}^*)$
$Y_m^\Delta = \mathbb{E}_m(Y_{m+1}^\Delta) + \Delta t_{m,1}\theta_1 f_{t_m}(\Theta_m^\Delta) + \Delta t_{m,1}(1-\theta_1)\mathbb{E}_m(f_{t_{m+1}}(\Theta_{m+1}^\Delta))$
**end for**

To test the numerical properties of schemes for ODEs, we often look at the test equation $y' = \lambda y$ subject to the initial condition $y(0)=1$ for $\lambda\in\mathbb{C}$. Then for $\mathrm{Re}(\lambda)<0$ we have that $y(t)\to 0$ as $t\to\infty$. A numerical method is called *A-stable* if it has the same behaviour, when the step size is fixed. Applying the theta scheme on this test equation with discretisation points $t_0, t_1, \dots$ with equal distance $\Delta t$, yields $y_n = y_{n-1} + \Delta t\lambda(\theta y_{n-1} + (1-\theta)y_n)$, rewriting gives us the numerical solution

$$y_n = \frac{1+\Delta t\lambda\theta}{1-\Delta t\lambda(1-\theta)}y_{n-1} = \left(\frac{1+\Delta t\lambda\theta}{1-\Delta t\lambda(1-\theta)}\right)^n.$$

Hence for A-stability we need whenever $\mathrm{Re}(\lambda)<0$, that

$$\left|\frac{1+\Delta t\lambda\theta}{1-\Delta t\lambda(1-\theta)}\right| < 1.$$

It is not hard to show that this is satisfied when $\theta\le 1/2$. Another property which is called *L-stability* is especially important for solving stiff equations, it ascertains that the numerical solution of the test equation should go to zero in one step as the step size goes to infinity. This means that the amplification factor $R(\Delta t\lambda) = y_n/y_{n-1}$ should satisfy $\lim_{z\to\infty}|R(z)| = 0$. For the theta schemes this is only true when $\theta = 0$, and this is likely the reason why we will see instabilities in the solution of the FBSDEJ when using the Crank-Nicolson scheme. We could instead choose $\theta = 0$, the backward Euler method, but then we lose an order of convergence. Alternatively we can look at linear multistep methods which generalise the backward Euler method to higher order methods while still maintaining high stability. Specifically we will look at the Backward Differentiation Formula methods of order $n$ (BDFn methods).

The BDFn methods can be found by approximating the derivative with the derivative of the n-th order Lagrange interpolation polynomial through the points $(t_{m-n},y_{m-n}),(t_{m-n+1},y_{m-n+1}),\dots,(t_m,y_m)$. We get the n-th order Lagrange interpolation polynomial

$$L(t) = \sum_{i=0}^n y_{m-i}\ell_{m-i}(t) = \sum_{i=0}^n y_{m-i}\prod_{\substack{0\le k\le n\\ k\ne m-i}}\frac{t-t_{m-n+k}}{t_{m-n+i}-t_{m-n+k}},$$

which has the derivative in time-step $t_m$,

$$L'(t_m) = \sum_{j=0}^{n} \alpha_{n,j}^m y_{m-j},$$

for some real numbers $\alpha_{n,j}^m$. Suppose we have the ODE $y' = f(t, y(t))$, then we have at time step $t_m$ the approximation

$$y_m = -\frac{1}{\alpha_{n,0}^m} \sum_{j=1}^{n} \alpha_{n,j}^m y_{m-j} + \frac{1}{\alpha_{n,0}^m} f(t_m, y_m).$$

Since the method is implicit, we need to use another method to solve the nonlinear equation in each step, for that we will use Picard iteration. When the distances between the nodes is equal, we write $\alpha_{n,i}$, since they are independent of $j$. The coefficients $\alpha_{n,i}$ are given in Table 7.1.

| $n$ | $\alpha_{0,n}\Delta t$ | $\alpha_{1,n}\Delta t$ | $\alpha_{2,n}\Delta t$ | $\alpha_{3,n}\Delta t$ | $\alpha_{4,n}\Delta t$ | $\alpha_{5,n}$ | $\alpha_{6,n}\Delta t$ |
|---|---|---|---|---|---|---|---|
| 1 | -1 | 1 | | | | | |
| 2 | $-3/2$ | 2 | -1/2 | | | | |
| 3 | -11/6 | 3 | -3/2 | 1/3 | | | |
| 4 | -25/12 | 4 | -3 | 4/3 | -1/4 | | |
| 5 | -137/60 | 5 | -5 | 10/3 | -5/4 | 1/5 | |
| 6 | -49/20 | 6 | -15/2 | 20/3 | -15/4 | 6/5 | -1/6 |

Table 7.1: Coefficients Backward Differentiation Formula of order $n$

We have not included the coefficients for $n > 6$ as those methods fail to be convergent. The difficulty with the BDFn methods, is that we already need to have an approximation of the first $n-1$ steps before we can even start using it. Two advantages however, are that the BDFn methods are of $n$-th order convergence and are very stable. Yet, the higher $n$, the less stable the method. For a more detailed treatment of linear multistep methods and difference equations, see Butcher [7] and Henrici [21].

To apply the BDFn methods to our FBSDEJ, note that we are now going backwards in time, so we also have to 'reverse time' in the method itself. This gives the following scheme.

**Scheme 2** Semi-discrete BDFn scheme

**for** $m = \mathcal{M} - n, \ldots, 0$ **do**

$Z_m^\Delta = \sum_{j=1}^n \alpha_{n,j}^m \mathbb{E}_m(Y_{m+j}^\Delta \Delta W_{m,j}),$

$\Gamma_m^\Delta = \sum_{j=1}^n \alpha_{n,j}^m \mathbb{E}_m(Y_{m+j}^\Delta \Delta \widetilde{N}_{m,j}^*)$

$Y_m^\Delta = -\frac{1}{\alpha_{n,0}^m} \sum_{j=1}^n \alpha_{n,j}^m \mathbb{E}_m(Y_{m+j}^\Delta) - \frac{1}{\alpha_{n,0}^m} f_{t_m}(\Theta_m^\Delta)$

**end for**

### 7.1.2 Full-discretisation

We saw in Scheme 1 amd Scheme 2 that we still have to approximate conditional expectations in each time step. For these conditional expectations we will use the COS approximation formulae we have found in the previous chapter. Just as before define

$$\Phi_{k,j}^Y(t_m, x) = \text{Re}\left(\phi_{\widetilde{X}_{m+j}^m|\widetilde{X}_m^m=x}\left(\frac{k\pi}{b-a}\right) e^{ik\pi\frac{x-a}{b-a}}\right),$$

$$\Phi_{k,j}^Z(t_m, x) = \text{Re}\left(\phi_{\widetilde{X}_{m+j}^m|\widetilde{X}_m^m=x}\left(\frac{k\pi}{b-a}\right) e^{ik\pi\frac{x-a}{b-a}} i\frac{k\pi}{b-a}\sigma\Delta t_{m,j}\right),$$

$$\Phi_{k,j}^\Gamma(t_m, x) = \text{Re}\left(\phi_{\widetilde{X}_{m+j}^m|\widetilde{X}_m^m=x}\left(\frac{k\pi}{b-a}\right) e^{-ik\pi\frac{x-a}{b-a}} \lambda\Delta t_{m,j}\mathbb{E}\left(\eta(J)\left(\exp\left(i\frac{k\pi}{b-a}\beta_{t_m}(x,J)\right)-1\right)\right)\right),$$

where $\phi_{\widetilde{X}^m_{m+j}|\widetilde{X}^m_m=x}$ is the conditional characteristic function of $\widetilde{X}^m_{m+j}$, given $\widetilde{X}^m_m = x$. Then we have for some sufficiently smooth function $v(t,x)$ that

$$\begin{aligned}
\mathbb{E}^x_m(v(t_{m+j}, \widetilde{X}^m_{m+j})) &\approx \sum_{k=0}^{\mathcal{N}-1}{}' \mathcal{V}_k(t_{m+j})\Phi^Y_{k,j}(t_m, x),\\
\mathbb{E}^x_m(v(t_{m+j}, \widetilde{X}^m_{m+j})\Delta W_{m,j}) &\approx \sum_{k=0}^{\mathcal{N}-1}{}' \mathcal{V}_k(t_{m+j})\Phi^Z_{k,j}(t_m, x),\\
\mathbb{E}^x_m(v(t_{m+j}, \widetilde{X}^m_{m+j})\Delta \widetilde{N}^*_{m,j}) &\approx \sum_{k=0}^{\mathcal{N}-1}{}' \mathcal{V}_k(t_{m+j})\Phi^\Gamma_{k,j}(t_m, x)
\end{aligned}$$

where

$$\mathcal{V}_k(t_{m+j}) = \frac{2}{b-a}\int_a^b v(t_{m+j}, y)\cos\left(k\pi\frac{y-a}{b-a}\right)\,dy.$$

Since we know that $Y, Z$ and $\Gamma$ are deterministic functions of the state process $X$ and time $t$, we write $y(t, X_t) = Y_t, z(t, X_t) = Z_t, \gamma(t, X_t) = \Gamma_t$ and $\theta(t, X_t) = \Theta_t = (Y_t, Z_t, \Gamma_t)$ and write $\widehat{y}, \widehat{z}, \widehat{\gamma}, \widehat{\theta}$ for their numerical approximations. To fully calculate the conditional expectations, we still need a method to approximate the Fourier coefficients with respect to $v$.

Write $\mathcal{Y}, \mathcal{Z}, \mathcal{G}$ and $\mathcal{F}$ for the Fourier cosine coefficients of $y, z, \gamma$ and $f$ respectively. To approximate these Fourier coefficients accurately, we need to be able to evaluate the integrand at a large set of points. Let $v(t,x)$ again be a sufficiently smooth function, we will discretise $[a,b]$ as follows. Let $\Delta x = \frac{b-a}{\mathcal{N}}$, then we define the grid points $x^n = a + (n+\frac{1}{2})\Delta x$ such that we can use a composite midpoint rule on the integrals,

$$\begin{aligned}
\mathcal{V}_k(t_{m+j}) &= \frac{2}{b-a}\int_a^b v(t_{m+j}, y)\cos\left(k\pi\frac{y-a}{b-a}\right)\,dy\\
&= \frac{2}{b-a}\sum_{n=0}^{\mathcal{N}-1} v(t_{m+j}, x^n)\cos\left(k\pi\frac{x^n-a}{b-a}\right)\Delta x\\
&= \frac{2}{\mathcal{N}}\sum_{n=0}^{\mathcal{N}-1} v(t_{m+j}, x^n)\cos\left(k\pi\frac{2n+1}{\mathcal{N}}\right).
\end{aligned}$$

Now the final sum corresponds to a DCT-II of the data $v(t_{m+j}, x^n)$, and can hence be computed efficiently through Fast Cosine Transform (FCT) algorithms with $\mathcal{O}(\mathcal{N}\log(\mathcal{N}))$ time complexity instead of the normal $\mathcal{O}(\mathcal{N}^2)$ time complexity.

Finally in the case where $\theta_1 < 1, \theta_2 < 1$ or $\theta_3 < 1$ in Algorithm 3 we have for the theta scheme an implicit part in the $Y, Z$ and $\Gamma$ processes, we solve these through a Picard iteration method. Hence for the theta scheme we have the complete algorithm, where we start from $m = M-1$. So we assume that we can solve the first step. We will come back on how to do this later on in this thesis.

**Algorithm 3** Theta algorithm

Given $\mathcal{Y}_k(t_{\mathcal{M}}), \mathcal{Z}_k(t_{\mathcal{M}}), \mathcal{G}_k(t_{\mathcal{M}})$ and $\mathcal{F}_k(t_{\mathcal{M}})$.
Compute $\Phi^Y_{k,1}(t_{\mathcal{M}-1},x), \Phi^Z_{k,1}(t_{\mathcal{M}-1},x), \Phi^\Gamma_{k,1}(t_{\mathcal{M}-1},x)$.
**for** $m = \mathcal{M}-1,\dots,0$ **do**

$$\widehat{z}(t_m,x) = -\theta_2^{-1}(1-\theta_2)\sum_{k=0}^{\mathcal{N}-1}{}' \mathcal{Z}_k(t_{m+1})\Phi^Y_{k,1}(t_m,x) + (\Delta t_{m,1}\theta_2)^{-1}\sum_{k=0}^{\mathcal{N}-1}{}' \mathcal{Y}_k(t_{m+1})\Phi^Z_{k,1}(t_m,x) + \theta_2^{-1}(1-\theta_2)\sum_{k=0}^{\mathcal{N}-1}{}' \mathcal{F}_k(t_{m+1})\Phi^Z_{k,1}(t_m,x)$$

$$\widehat{\gamma}(t_m,x) = -\theta_3^{-1}(1-\theta_3)\sum_{k=0}^{\mathcal{N}-1}{}' \mathcal{G}_k(t_{m+1})\Phi^Y_{k,1}(t_m,x) + (\Delta t_{m,1}\theta_3)^{-1}\sum_{k=0}^{\mathcal{N}-1}{}' \mathcal{Y}_k(t_{m+1})\Phi^\Gamma_{k,1}(t_m,x) + \theta_3^{-1}(1-\theta_3)\sum_{k=0}^{\mathcal{N}-1}{}' \mathcal{F}_k(t_{m+1})\Phi^\Gamma_{k,1}(t_m,x)$$

$$\widehat{h}(t_m,x) = \sum_{k=0}^{\mathcal{N}-1}{}' \mathcal{Y}_k(t_{m+1})\Phi^Y_{k,1}(t_m,x) + \Delta t_{m,1}(1-\theta_1)\sum_{k=0}^{\mathcal{N}-1}{}' \mathcal{F}_k(t_{m+1})\Phi^Y_{k,1}$$

$$\widehat{y}^0(t_m,x) = \sum_{k=0}^{\mathcal{N}-1}{}' \mathcal{Y}_k(t_{m+1})\Phi^Y_{k,1}(t_m,x)$$

**for** $q = 1,\dots,P$ **do**

$$\widehat{y}^q(t_m,x) = \Delta t\theta_1 f_{t_m}(x,\widehat{y}^{q-1}(t_m,x),\widehat{z}(t_m,x),\widehat{\gamma}(t_m,x)) + \widehat{h}(t_m,x)$$

**end for**
Compute $\mathcal{Y}_k(t_m), \mathcal{Z}_k(t_m), \mathcal{G}_k(t_m)$ and $\mathcal{F}_k(t_m)$.
Compute $\Phi^Y_{k,1}(t_{m-1},x), \Phi^Z_{k,1}(t_{m-1},x), \Phi^\Gamma_{k,1}(t_{m-1},x)$.
**end for**

Similarly as with the theta-discretisation scheme, we will assume for the BDFn discretisation scheme that we can solve the first $n$ steps. Again we will come back later on how to compute these first steps.

**Algorithm 4** BDFn algorithm

Given $\mathcal{Y}_k(t_{\mathcal{M}-j})$, for $j = 0,\dots,n-1$.
Compute $\Phi^Y_{k,j}(t_{\mathcal{M}-n},x), \Phi^Z_{k,j}(t_{\mathcal{M}-n},x), \Phi^\Gamma_{k,j}(t_{\mathcal{M}-n},x)$ for $j = 1,\dots,n$.
**for** $m = \mathcal{M}-j,\dots,0$ **do**

$$\widehat{z}(t_m,x) = \sum_{j=1}^{n}\sum_{k=0}^{\mathcal{N}-1}{}' \alpha^m_{n,j}\mathcal{Y}_k(t_{m+j})\Phi^Z_{k,j}(t_m,x)$$

$$\widehat{\gamma}(t_m,x) = \sum_{j=1}^{n}\sum_{k=0}^{\mathcal{N}-1}{}' \alpha^m_{n,j}\mathcal{Y}_k(t_{m+j})\Phi^\Gamma_{k,j}(t_m,x)$$

$$\widehat{h}(t_m,x) = -\frac{1}{\alpha^m_{n,0}}\sum_{j=1}^{n}\sum_{k=0}^{\mathcal{N}-1}{}' \alpha^m_{n,j}\mathcal{Y}_k(t_{m+j})\Phi^Y_{k,j}(t_m,x)$$

$$\widehat{y}^0(t_m,x) = \sum_{k=0}^{\mathcal{N}-1}{}' \mathcal{Y}_k(t_{m+1})\Phi^Y_{k,1}(t_m,x)$$

**for** $q = 1,\dots,P$ **do**

$$\widehat{y}^q(t_m,x) = \Delta t\theta_1 f_{t_m}(x,\widehat{y}^{q-1}(t_m,x),\widehat{z}(t_m,x),\widehat{\gamma}(t_m,x)) + \widehat{h}(t_m,x)$$

**end for**
Compute $\mathcal{Y}_k(t_{m+j})$ for $j = 1,\dots,n$.
Compute $\Phi^Y_{k,j}(t_{m-1},x), \Phi^Z_{k,j}(t_{m-1},x), \Phi^\Gamma_{k,j}(t_{m-1},x)$ for $j = 1,\dots,n$.
**end for**

## 7.2 Convergence rate

Before we will look at some numerical examples, we will focus in this section on some error analysis of the schemes we presented. We will split the analysis into two parts, first we look at the semi-discretisation scheme for the BDFn methods as in Scheme 2. Afterwards we will give an error analysis of the COS method.

The theta-discretisation scheme for continuous BSDEs has been studied in Zhao, Wang and Peng [46] where the driver is assumed to be independent of $Z$. Second-order convergence rate has been proven for both the $Y$-process as the $Z$-process in the case where $\theta = 1/2$, for $\theta \neq 1/2$, the scheme has only first-order convergence. For general $f$ and BSDEJs, proving optimal convergence rates is still an open problem.

### 7.2.1 Semi-discretisation

Proving convergence rates of BDFn discretisation methods for general FSDEJs is highly non-trivial and is still an open problem. There are three main difficulties for solving this problem. The first problem is that the driver $f$ depends on $Z$ and $\Gamma$, while the solutions of $Z$ and $\Gamma$ again depend on $Y$. Since the driver $f$ is in general non-linear, the equations are coupled in a non-linear manner. The second problem is that the processes have to be computed using the solution of multiple previous steps at the same time, hence the error bound depends in a difficult manner on the error bounds of the previous steps. Finally the coefficients $\alpha_{n,j}$ generally lie outside the unit circle, which makes iteration procedures difficult, as $(\alpha_{n,j})^k$ is unbounded in $k$. In the case where the coefficients $\alpha_j$ for the linear multistep method are positive, we know that they all lie in the unit circle, and using this the problem has been solved by Chassagneux [8]. Therefore the following theorems are more a justification rather than a formal convergence proof as we will make a strong assumption on the driver $f$.

We will only look at a simpler FBSDEJ, given by

$$\begin{cases} X_t = X_0 + \int_0^t \mu_s(X_s)\,ds + \int_0^t \sigma_s(X_s)\,dW_s + \int_0^t \int_{\mathbb{R}_0} \beta_s(X_{s-}, J)\,\widetilde{N}(ds, dJ), & \text{(FSDEJ)} \\ Y_t = \xi + \int_t^T f_s(X_s, Y_s)\,ds - \int_t^T Z_s\,dW_s - \int_t^T \int_{\mathbb{R}_0} U_s(J)\,\widetilde{N}(ds, dJ), & \text{(BSDEJ)} \end{cases}$$

Here the driver $f$ does not depend on $Z$ and $\Gamma$ anymore, such that the numerical scheme for $Y$ is independent of $Z$ and $\Gamma$.

We write $(Y_m^\Delta, Z_m^\Delta, \Gamma_m^\Delta)$ for the numerical solution of Scheme 2 at time $t = t_m$ and let $(X_{t_m}, Y_{t_m}, Z_{t_m}, \Gamma_{t_m})$ be the exact solution at time $t = t_m$. We will furthermore assume that we can solve all the conditional expectations exact. Then we define the global errors

$$\begin{aligned} \varepsilon_y^m &= Y_{t_m} - Y_m^\Delta, & \varepsilon_z^m &= Z_{t_m} - Z_m^\Delta, \\ \varepsilon_\gamma^m &= \Gamma_{t_m} - \Gamma_m^\Delta, & \varepsilon_f^m &= f_{t_m}(X_{t_m}, \Theta_m^\Delta) - f_{t_m}(X_{t_m}, \Theta^n). \end{aligned}$$

Furthermore define the truncation errors

$$\begin{aligned} \mathcal{R}_{y,m}^n &= \left.\frac{d\mathbb{E}_m^x(Y_t)}{dt}\right|_{t=t_m} - \sum_{j=0}^n \alpha_{n,j}\mathbb{E}_m^x(Y_{t_{m+j}}), \\ \mathcal{R}_{z,m}^n &= \left.\frac{d\mathbb{E}_m^x(Y_t \Delta W_{t_m,t})}{dt}\right|_{t=t_m} - \sum_{j=0}^n \alpha_{n,j}\mathbb{E}_m^x(Y_{t_{m+j}}\Delta W_{t_m,t}), \\ \mathcal{R}_{\gamma,m}^n &= \left.\frac{d\mathbb{E}_m^x(Y_t \Delta \widetilde{N}_{t_m,t}^*)}{dt}\right|_{t=t_m} - \sum_{j=0}^n \alpha_{n,j}\mathbb{E}_m^x(Y_{t_{m+j}}\Delta \widetilde{N}_{t_m,t}^*). \end{aligned}$$

For the truncation errors we have the following convergence rate, the proofs are based on the proofs of Yang and Zhao [44] in the case of continuous BSDEs.

**Lemma 7.2.1.** *Suppose that the given data $\mu, \sigma, \gamma, f$ and $\xi$ are sufficiently smooth, such that $\mathcal{L}^{n+1}u(t,x), \mathcal{L}^n\sigma\nabla u(t,x)$ and $\mathcal{L}^n\mathcal{M}u(t,x)$ exist and are bounded for all $(t,x) \in [0,T] \times \mathbb{R}$. Then we have the local estimates*

$$\mathcal{R}_{y,m}^n = \mathcal{O}(\Delta t^n), \qquad \mathcal{R}_{z,m}^n = \mathcal{O}(\Delta t^n), \qquad \mathcal{R}_{\gamma,m}^n = \mathcal{O}(\Delta t^n)$$

*Proof.*
Suppose $0 \le t_0 < t_1 < \cdots < t_n \le T$, by Lagrange interpolation theory, we know that for $f \in C_b^{n+1}$ and every $i = 0, 1, \ldots, n$, that

$$|f'(t_i) - L'(t_i)| \le \sup_{t \le T} \prod_{\substack{j=0 \\ j \ne i}}^n |t_i - t_j| \frac{|f^{(n+1)}(t)|}{(n+1)!} = \mathcal{O}\left(\prod_{\substack{j=0 \\ j \ne i}}^n |t_i - t_j|\right).$$

Now it remains to prove that the conditional expectations are in $C^{n+1}$. We will write the derivatives with respect to the conditional expectations as right-derivatives, due to technicalities with the filtration. However, we will see that the right-derivatives are continuous by the smoothness conditions, and hence equal to the derivatives. It is well-known that $X$ is a Markov process see for example Applebaum [1, Theorem 6.4.6]. Hence define $\mathcal{F}_s^{t,x} = \sigma(X_r, t \le r \le s, X_t = x)$, and let $t_0 < t$ be a fixed time, and $x_0 \in \mathbb{R}$

a fixed point in space. If $g \in C^{1,2}$ it follows by the Leibniz integral rule that

$$\begin{aligned}
\frac{d}{dt}\mathbb{E}^{x_0}_{t_0}(g(t,X_t)) &= \lim_{s\downarrow t}\mathbb{E}^{x_0}_{t_0}\left(\frac{g(s,X_s)-g(t,X_t)}{s-t}\right)\\
&= \lim_{s\downarrow t}\mathbb{E}\left(\frac{g(s,X_s)-g(t,X_t)}{s-t}\middle|\,\mathcal{F}^{t_0,x_0}_s\right)\\
&= \lim_{s\downarrow t}\mathbb{E}\left(\mathbb{E}^{X_t}_t\left(\frac{g(s,X_s)-g(t,X_t)}{s-t}\right)\middle|\,\mathcal{F}^{t_0,x_0}_s\right)\\
&= \lim_{s\downarrow t}\mathbb{E}\left(\frac{\mathbb{E}^{X_t}_t(g(s,X_s))-g(t,X_t)}{s-t}\middle|\,\mathcal{F}^{t_0,x_0}_s\right)\\
&= \mathbb{E}\left(A^{X_t}_t g(\cdot,X_\cdot)\middle|\,\mathcal{F}^{t_0,x_0}_s\right)\\
&= \mathbb{E}\left(\mathcal{L}g(t,X_t)\middle|\,\mathcal{F}^{t_0,x_0}_s\right) = \mathbb{E}^{x_0}_{t_0}\left(\mathcal{L}g(t,X_t)\right).
\end{aligned}$$

Now $s \mapsto \mathbb{E}^{x_0}_{t_0}(\mathcal{L}g(s,X_s))$ is continuous at $t$, hence by our discussion above, we have by induction

$$\frac{d^{k+1}}{dt^{k+1}}\mathbb{E}^{x_0}_{t_0}\left(u(t,X_t)\right) = \mathbb{E}^{x_0}_{t_0}\left(\mathcal{L}^{k+1}u(t,X_t)\right),$$

where we now use the regular derivative, by our assumptions the derivative exists and is bounded. With similar arguments we have

$$\begin{aligned}
\frac{d}{dt}\mathbb{E}^{x_0}_{t_0}(g(t,X_t)) &= \mathbb{E}^{x_0}_{t_0}\left(\sigma_t(X_t)\nabla g(t,X_t)\right),\\
\frac{d}{dt}\mathbb{E}^{x_0}_{t_0}(g(t,X_t)) &= \mathbb{E}^{x_0}_{t_0}\left(\mathcal{M}g(t,X_t)\right),
\end{aligned}$$

hence the result follows. □

Fix $n$ and define $\alpha_j = \Delta t\alpha_{n,j}$. Then define the characteristic polynomial $\rho(\zeta) = \sum_{j=0}^n \alpha_j\zeta^{n-j}$. To prove actual convergence of the scheme, we also need to have so-called *zero-stability*, which means that the difference equation $\sum_{j=0}^n \alpha_j y_{m-j} = 0$ must have bounded solutions as $m \to \infty$. It turns out that the solutions are all polynomials and that the following root condition is necessary and sufficient for zero-stability. If all the roots of $\rho(\zeta)$ satisfy $|\zeta| \leq 1$ and any root with $|\zeta| = 1$ has multiplicity one, then the method satisfies the root condition, Henrici [21, Theorem 5.5]. For $1 \leq n \leq 6$ this is satisfied in the case of the BDFn methods, for $n > 6$ this fails. Numerical examples confirm that the method fails to converge, and FBSDEJs are no exception, as can be seen in Fu, Zhao and Zhou [18]. Therefore we will only consider $1 \leq n \leq 6$.

Assuming the root condition holds, it turns out that there exists a sequence $(\delta_\ell)_{\ell=0}^\infty$ such that

$$\frac{1}{\alpha_0 + \ldots + \alpha_{n-1}\zeta^{n-1} + \alpha_n\zeta^n} = \delta_0 + \delta_1\zeta + \delta_2\zeta^2 + \ldots, \tag{7.5}$$

and such that $\kappa := \sup_{\ell\geq 0}|\delta_\ell| < \infty$, Henrici [21, Lemma 5.5].

Furthermore, for the $Z$-process and the $\Gamma$-process we have the following non-optimal bound for the global error, as we lose a half-order of the convergence rate. In Yang and Zhao [44] an optimal convergence rate has been proven for the $Z$-process using variational and Malliavin derivative arguments. A similar proof should also be possible for the $\Gamma$-process, however, the necessary theory for either proof is outside the scope of this thesis. Furthermore we will see in the next chapter that we can empirically confirm the $n$-th order convergence which we should expect.

**Lemma 7.2.2.** *There exists a constant $C > 0$, such that for every $0 \leq m \leq \mathcal{N} - n$ we have the estimate*

$$\|\varepsilon^m_z\|_\infty + \|\varepsilon^m_\gamma\|_\infty \leq \frac{C}{\sqrt{\Delta t}}\sum_{j=1}^n \|\varepsilon^{m+j}_y\|_\infty + 2(\|\mathcal{R}^n_{z,m}\|_\infty + \|\mathcal{R}^n_{\gamma,m}\|_\infty).$$

*Proof.*
Note that

$$\varepsilon^m_z = \sum_{j=1}^k \alpha_{n,j}\mathbb{E}_{t_m}\left(\varepsilon^{m+j}_y\Delta W_{m,j}\right) + \mathcal{R}^n_{z,m}.$$

Squaring both sides and applying Hölder's inequality yields

$$\begin{aligned}|\varepsilon_z^m|^2 &= \left(\sum_{j=1}^{n} \alpha_{n,j}\mathbb{E}_{t_m}\left(\varepsilon_y^{m+j}\Delta W_{m,j}\right) + \mathcal{R}_{z,m}^n\right)^2 \\ &\leq 2\left(n\sum_{j=1}^{n}\alpha_{n,j}^2\left(\mathbb{E}_{t_m}\left(|\varepsilon_y^{m+j}\Delta W_{m,j}|\right)\right)^2 + |\mathcal{R}_{z,m}^n|^2\right) \\ &\leq 2\left(n\sum_{j=1}^{n}\alpha_{n,j}^2\Delta t_{m,j}\mathbb{E}_{t_m}(|\varepsilon_y^{m+j}|^2) + |\mathcal{R}_{z,m}^n|^2\right).\end{aligned}$$

Define $\alpha_{m,j}\Delta t_{m,j} = \alpha_m$, then $\sup_{0\leq m\leq n}|\alpha_m| \leq C$ for some $C > 0$. Hence there exists some $C > 0$ such that

$$|\varepsilon_z^m|^2\Delta t \leq C\sum_{j=1}^{n}\mathbb{E}_{t_m}(|\varepsilon_y^{m+j}|^2) + 2\Delta t\left\|\mathcal{R}_{z,m}^n\right\|_\infty^2.$$

Hence dividing by $\Delta t_{m,j}$, taking the square root and taking the $L^\infty$-norm gives for some other $C > 0$, the estimate

$$\|\varepsilon_z^m\|_\infty \leq \frac{C}{\sqrt{\Delta t}}\sum_{j=1}^{n}\left\|\varepsilon_y^{m+j}\right\|_\infty + 2\left\|\mathcal{R}_{z,m}^n\right\|_\infty.$$

We have a similar estimate for $\varepsilon_\gamma^m$.

$$\begin{aligned}|\varepsilon_\gamma^m|^2 &= \left(\sum_{j=1}^{n}\alpha_{n,j}\mathbb{E}_{t_m}\left(\varepsilon_y^{m+j}\Delta\widetilde{N}_{m,j}^*\right) + \mathcal{R}_{\gamma,m}^n\right)^2 \\ &\leq 2\left(n\sum_{j=1}^{n}\alpha_{n,j}^2\left(\mathbb{E}_{t_m}\left(|\varepsilon_y^{m+j}\Delta\widetilde{N}_{m,j}^*|\right)\right)^2 + |\mathcal{R}_{\gamma,m}^n|^2\right) \\ &\leq 2\left(n\sum_{j=1}^{n}\alpha_{n,j}^2\mathbb{E}_{t_m}(|\varepsilon_y^{m+j}|^2)\mathbb{E}_{t_m}(|\Delta\widetilde{N}_{m,j}^*|^2) + |\mathcal{R}_{\gamma,m}^n|^2\right).\end{aligned}$$

To compute $\mathbb{E}(|\Delta\widetilde{N}_{m,j}^*|^2)$ we can use the Itô isommetry to get

$$\mathbb{E}_{t_m}(|\Delta\widetilde{N}_{m,j}^*|^2) = \mathbb{E}_{t_m}\left(\int_{t_m}^{t_{m+j}}\eta^2(J)d\nu(J)\,dt\right) \leq \Delta t_{m,j}\|\eta\|_\infty^2\,\nu(\mathbb{R}_0).$$

Then it indeed follows that for some constant $C > 0$ we have

$$|\varepsilon_\gamma^m|^2\Delta t \leq C\sum_{j=1}^{n}\mathbb{E}_{t_m}(|\varepsilon_y^{m+j}|^2) + 2\Delta t|\mathcal{R}_{\gamma,m}^n|^2,$$

and the result follows. □

For the $Y$-process we can now prove the following bound.

**Theorem 7.2.3.** *Suppose $f(t,x,y)$ is Lipschitz continuous with respect to $y$ with Lipschitz constant $L_f$. Then for $0 < \Delta t \leq \frac{|\alpha_{n,0}|\Delta t}{2L_f}$, there exists constants $C_1, C_2 > 0$ depending on $T, L_f$ and $n$ such that*

$$\max_{0\leq m\leq N-n}\left\|\varepsilon_y^m\right\|_\infty \leq C_1\sum_{i=N-n+1}^{N}\left\|\varepsilon_y^i\right\|_\infty + C_2\Delta t\sum_{i=0}^{N-n}\left\|\mathcal{R}_{y,i}^n\right\|_\infty.$$

*Proof.*
First note that we can write $\varepsilon_y^m$ as

$$\alpha_{n,0}\varepsilon_y^m = -\sum_{j=1}^{n}\alpha_{n,j}\mathbb{E}_{t_m}(\varepsilon_y^{m+j}) - \varepsilon_f^m + \mathcal{R}_{y,m}^n.$$

Define $\beta_m = -\varepsilon_f^m/\varepsilon_y^m$, then by the Lipschitz continuity of $f$ we know $|\beta_m| \leq L_f$, furthermore define $\widetilde{\alpha}_j = \alpha_{n,j}\Delta t$, then note that $\sup_{0\leq j\leq n}|\widetilde{\alpha}_j| \leq C$ for some constant $C$ depending on $n$. Taking the conditional expectation $\mathbb{E}_{t_m}(\cdot)$ on both sides gives

$$\sum_{j=0}^{n} \widetilde{\alpha}_j \mathbb{E}_{t_m}(\varepsilon_y^{m+j}) = \beta_m \varepsilon_y^m \Delta t + \mathbb{E}_{t_m}(\mathcal{R}_{y,m}^n)\Delta t. \tag{7.6}$$

Because the $\widetilde{\alpha}_j$ are not bounded by 1, we cannot keep expanding $\mathbb{E}_{t_m}(\varepsilon_y^{m+j})$ until time $N$, as $\widetilde{\alpha}_j^k \to \infty$ as $k \to \infty$. Instead multiply both sides with $n := n + \ell$ by $\delta_\ell$ for $\ell = 0, 1, \ldots, N-m-n$ and sum over $\ell$ (as defined in (7.5)), which yields the following expression for $S_l$, the left-hand side,

$$\begin{aligned} S_l = &\sum_{\ell=0}^{N-m-n} \delta_\ell \sum_{j=0}^{n} \widetilde{\alpha}_j \mathbb{E}_{t_m}(\varepsilon_y^{m+j+\ell}) \\ &= \widetilde{\alpha}_0\delta_0\mathbb{E}_{t_m}(\varepsilon_y^m) + (\widetilde{\alpha}_0\delta_1 + \widetilde{\alpha}_1\delta_0)\mathbb{E}_{t_m}(\varepsilon_y^{m+1}) + \ldots \\ &\quad + (\widetilde{\alpha}_0\delta_{N-m-n} + \widetilde{\alpha}_1\delta_{N-m-n-1} + \ldots + \widetilde{\alpha}_n\delta_{N-m-2n+1})\mathbb{E}_{t_m}(\varepsilon_y^{N-n}) \\ &\quad + (\widetilde{\alpha}_1\delta_{N-m-n} + \widetilde{\alpha}_2\delta_{N-m-n-1} + \ldots + \widetilde{\alpha}_n\delta_{N-m-2n+1})\mathbb{E}_{t_m}(\varepsilon_y^{N-n+1}) \\ &\quad + \ldots + (\widetilde{\alpha}_{n-1}\delta_{N-m-n} + \widetilde{\alpha}_n\delta_{N-m-n-1})\mathbb{E}_{t_m}(\varepsilon_y^{N-1}) + \widetilde{\alpha}_n\delta_{N-m-n}\mathbb{E}_{t_m}(\varepsilon_y^N). \end{aligned}$$

For the right-hand side $S_r$ we have the expression

$$S_r = \sum_{\ell=0}^{N-n-m} \delta_\ell \left(\mathbb{E}_{t_m}(\beta_m\varepsilon_y^{m+\ell}) - \mathbb{E}_{t_m}(\mathcal{R}_{y,m+\ell}^n)\right)\Delta t.$$

Note that we have by definition for all $\lambda$ that

$$(\widetilde{\alpha}_0 + \ldots + \widetilde{\alpha}_{k-1}\lambda^{k-1} + \widetilde{\alpha}_k\lambda^k)(\delta_0 + \delta_1\lambda + \delta_2\lambda^2 + \ldots) = 1.$$

Now let $\delta_\ell = 0$ for $\ell < 0$ and $\widetilde{\alpha}_i = 0$ for $i > k$, then we have the Cauchy product

$$\sum_{\ell=0}^{\infty}\sum_{i=0}^{\ell} \widetilde{\alpha}_i\delta_{\ell-i}\lambda^i = 1.$$

Matching the coefficients of the polynomials on the left-hand side and on the right-hand side, we find the relation

$$\sum_{i=0}^{\ell} \widetilde{\alpha}_i\delta_{\ell-i} = \mathbf{1}_{\ell=0},$$

hence the expression $S_l$ simplifies significantly to

$$S_l = \varepsilon_y^m + \sum_{i=1}^{n}\left(\sum_{j=1}^{n+1-i} \widetilde{\alpha}_j\delta_{N-n-m+i-j}\right)\mathbb{E}_{t_m}(\varepsilon_y^{N-n+i}) := \varepsilon_y^m + A.$$

From the equality $S_l = S_r$ we derive

$$(1 - \delta_0\Delta t\beta_m)\varepsilon_y^m = -A + B_m + \sum_{i=m+1}^{N-n} \delta_{i-m}\mathbb{E}_{t_m}(\beta_i\varepsilon_y^i)\Delta t,$$

where

$$B_m = \sum_{i=m}^{N-n} \delta_{i-m}\mathbb{E}_{t_m}(\mathcal{R}_{y,i}^n)\Delta t.$$

Recall that $|\beta_m| \leq L_f$, so we have

$$(1 - \Delta t L_f|\delta_0|)|\varepsilon_y^m| \leq |A| + |B_m| + \sum_{i=m+1}^{N-n} |\delta_{i-m}|\Delta t L_f \mathbb{E}_{t_m}(|\varepsilon_y^i|).$$

Note that $\kappa := \sup_{\ell\geq 0}|\gamma_\ell| < \infty$ and $\frac{1}{1-\Delta t L_f|\delta_0|} \leq 2$, so then

$$|\varepsilon_y^m| \leq 2(|A| + |B_m|) + 2\kappa\Delta t L_f \sum_{i=m+1}^{N-n} \mathbb{E}_{t_m}(|\varepsilon_y^i|).$$

Taking the $L^\infty$-norm on both sides yields

$$\left\|\varepsilon_y^m\right\|_\infty \leq 2(\|A\|_\infty + \|B_m\|_\infty) + 2\kappa\Delta t L_f \sum_{i=m+1}^{N-n} \left\|\varepsilon_y^i\right\|_\infty .$$

Estimating $\|A\|_\infty$ and $\|B_m\|_\infty$ gives the inequalities

$$\|A\|_\infty \leq C\kappa \sum_{i=N-n+1}^{N} \left\|\varepsilon_y^i\right\|_\infty ,$$
$$\|B_m\|_\infty \leq \kappa \sum_{i=m}^{N-n} \left\|\mathcal{R}_{y,i}^n\right\|_\infty \Delta t.$$

Define $\alpha = 2(\|A\|_\infty + \|B_m\|_\infty)$ and $\beta = 2\kappa L_f$, furthermore define $\eta_m = \left\|\varepsilon_y^m\right\|_\infty$, then we have a discrete version of the Grönwall inequality,

$$\begin{aligned}
\eta_m &\leq \alpha + \beta\Delta t \sum_{i=m+1}^{N} \eta_i \\
&= \alpha + \beta\Delta t \sum_{i=m+2}^{N} \eta_i + \beta\Delta t \eta_{n+1} \\
&\leq (1+\beta\Delta t)\left(\alpha + \beta\Delta t \sum_{i=m+2}^{N} \eta_i\right) \\
&\leq (1+\beta\Delta t)^{N-m-n+1}\left(\alpha + \beta\Delta t \sum_{i=N-n+1}^{N} \eta_i\right) \\
&\leq e^{\beta N\Delta t}\left(\alpha + \beta\Delta t \sum_{i=N-n+1}^{N} \eta_i\right).
\end{aligned}$$

Hence the result follows

$$\begin{aligned}
\left\|\varepsilon_y^m\right\|_\infty &\leq e^{\beta T}\left(2C\kappa \sum_{i=N-n+1}^{N} \left\|\varepsilon_y^i\right\|_\infty + 2\kappa \sum_{i=m}^{N-n} \left\|\mathcal{R}_{y,i}^n\right\|_\infty \Delta t + 2\kappa L_f \Delta t \sum_{i=N-n+1}^{N} \left\|\varepsilon_y^i\right\|_\infty\right) \\
&= C_1 \sum_{i=N-n+1}^{N} \left\|\varepsilon_y^i\right\|_\infty + C_2\Delta t \sum_{i=m}^{N-n} \left\|\mathcal{R}_{y,i}^n\right\|_\infty .
\end{aligned}$$

□

We can summarise the previous results in the following corollary.

**Corollary 7.2.4.** *If the conditions in Lemma 7.2.1 and Theorem 7.2.3 hold, and we have that*

$$\max_{N-n<m\leq N} \left\|\varepsilon_y^m\right\|_\infty = \mathcal{O}(\Delta t^n),$$

*then for* $0 < \Delta t \leq \frac{|\alpha_{n,0}|\Delta t}{2L_f}$ *there exists a constant* $C > 0$ *such that*

$$\max_{0\leq m\leq N-n} \left\|\varepsilon_y^m\right\|_\infty \leq C\Delta t^n, \qquad \max_{0\leq m\leq N-n} \left\|\varepsilon_z^m\right\|_\infty \leq C\Delta t^{n-\frac{1}{2}}, \qquad \max_{0\leq m\leq N-n} \left\|\varepsilon_\gamma^m\right\|_\infty \leq C\Delta t^{n-\frac{1}{2}}$$

### 7.2.2 COS method

Most of the hard work has already been done in Chapter 6, however some problems we have not touched upon yet. Recall that we have made a couple of errors in the COS method for solving conditional expectations of the form

$$\int_{\mathbb{R}} v(t,y)p_t(y|x)\,dy.$$

(i) We truncated the integration range to the finite interval $[a,b]$:

$$\varepsilon_1 := \int_{\mathbb{R}\setminus[a,b]} v(t,y)p_t(y|x)\,dy.$$

(ii) Additionally we truncated the Fourier cosine series on $[a,b]$:

$$\varepsilon_2 := \frac{b-a}{2}\sum_{k=\mathcal{N}}^{\infty}{}' \mathcal{V}_k(t)\mathcal{P}_k(x).$$

(iii) Finally we approximated the Fourier cosine coefficients $\mathcal{P}_k(x)$:

$$\varepsilon_3 := \frac{b-a}{2}\sum_{k=0}^{\mathcal{N}-1}{}' \mathcal{V}_k(t)\int_{\mathbb{R}\setminus[a,b]} p_t(y|x)\cos\left(k\pi\frac{y-a}{b-a}\right)dy$$

Since we in general do not know much about the function $v$, we prefer to bound these errors in terms of the density. For $\varepsilon_2$ we can reason as follows. We know that

$$|\mathcal{V}_k(t)| \le \frac{2}{b-a}\int_a^b |v(t,y)|\,dy \le \frac{2C}{b-a}$$

Hence we have

$$|\varepsilon_2| \le C\sum_{k=\mathcal{N}}^{\infty}|\mathcal{P}_k(x)|.$$

Therefore the error $\varepsilon_2$, depends on the convergence speed of the above partial sums. Following Theorem 6.1.8 it is reasonable to assume that $p_t(\cdot|x)\in C^\infty$ for all $t\le T$ and $x\in\mathbb{R}$. Now we know that $\mathcal{P}_k(x)=\mathcal{O}(k^{-m})$ for every $m\in\mathbb{N}$, hence

$$\sum_{n=\mathcal{N}}^{\infty}\frac{1}{n^m} \le \sum_{n=\mathcal{N}}^{\infty}\int_{n-1}^{n}\frac{1}{x^m}\,dx = \int_{\mathcal{N}-1}^{\infty}\frac{1}{x^m}\,dx = \frac{1}{(m-1)(\mathcal{N}-1)^{m-1}},$$

So $\varepsilon_2=\mathcal{O}(\mathcal{N}^{1-m})$ for each $m\in\mathbb{N}$. Hence $\varepsilon$ decreases faster than polynomially to zero. For $\varepsilon_3$ we have

$$|\varepsilon_3| \le C\mathcal{N}\int_{\mathbb{R}\setminus[a,b]} p_t(y|x)\,dy.$$

The last bound can be improved when $y\mapsto v(t,y)$ is in $C^\alpha$ for some $\alpha\in\mathbb{N}$, then

$$\begin{aligned}
\varepsilon_3 &= \int_{\mathbb{R}\setminus[a,b]}\left(\frac{b-a}{2}\sum_{k=0}^{\mathcal{N}-1}{}'\mathcal{V}_k(t)\cos\left(k\pi\frac{y-a}{b-a}\right)\right)p_t(y|x)\,dy\\
&= \int_{\mathbb{R}\setminus[a,b]}\left(v(t,y)-\frac{b-a}{2}\sum_{k=\mathcal{N}}^{\infty}{}'\mathcal{V}_k(t)\cos\left(k\pi\frac{y-a}{b-a}\right)\right)p_t(y|x)\,dy\\
&= \varepsilon_1 - \frac{b-a}{2}\int_{\mathbb{R}\setminus[a,b]}\left(\sum_{k=\mathcal{N}}^{\infty}{}'\mathcal{V}_k(t)\cos\left(k\pi\frac{y-a}{b-a}\right)\right)p_t(y|x)\,dy
\end{aligned}$$

By Theorem 6.2.1 we have that

$$\left|\sum_{k=\mathcal{N}}^{\infty}{}'\mathcal{V}_k(t)\cos\left(k\pi\frac{y-a}{b-a}\right)\right| \le \sum_{k=\mathcal{N}}^{\infty}{}'|\mathcal{V}_k(t)| = \mathcal{O}(\mathcal{N}^{1-\alpha}).$$

Hence,

$$|\varepsilon_3| \leq |\varepsilon_1| + \mathcal{O}(\mathcal{N}^{1-\alpha}) \int_{\mathbb{R}\setminus[a,b]} p_t(y|x)\, dy.$$

If we choose $\mathcal{N}$ large enough, we only have to worry about choosing a suitable truncation range $[a, b]$.

Given a random variable $X$, the cumulants of $X$ are defined by

$$\kappa_n(X) = \frac{1}{i^n} \frac{\partial^n(\log(\phi_X(\xi)))}{\partial \xi^n}\bigg|_{\xi=0},$$

whenever the derivative exists. Fang and Oosterlee [17] and Ruijter and Oosterlee [36] proposed for some stochastic process $X_t$ and some $L > 0$ the following truncation range. Take the cumulants $\kappa_n$ from $X_T - X_0$, then

$$[a, b] = \left[X_0 + \kappa_1 - L\sqrt{\kappa_2 + \sqrt{\kappa_4}}, X_0 + \kappa_1 + L\sqrt{\kappa_2 + \sqrt{\kappa_4}}\right].$$

If we choose $\mathcal{N}$ sufficiently large, $L$ can be determined such that the error nears machine precision for a sufficiently wide class of distributions. $L = 10$ is sufficiently big for must standard models and we will use this for our numerical examples. Furthermore we take $\mathcal{N} = 2^{10}$, such that the COS errors are approximately of order $10^{-12}$. The values for $\mathcal{N}$ and $L$ can be fine-tuned beforehand by investigating the error in the truncated cosine expansion of the density. For some stochastic processes like geometric Brownian motion, $L = 6$ is enough, which allows us to take a lower $\mathcal{N}$ for the same accuracy, and thus increase the speed of the algorithm

# CHAPTER 8

# Numerical Examples

We will do five numerical experiments using MATLAB 9.8.0 with an Intel(R) Core(TM) i5-6600K CPU @4.20 GHz and 15.8 GB RAM. As mentoined in the previous chapter, we will take $\mathscr{N} = 2^{10}$ with 5 Picard iterations. The general setup is to gradually increase the complexity of the examples and pinpoint the difficulties we face with numerically computing FBSDEJs.

First we will look at continuous BSDEs with constant drift and diffusion terms, so that we can see what errors and instabilities are due to the extra jump terms.

## Example 1

The first example is taken from Ruijter [35], with the underlying process a standard Brownian motion. The BSDE is given by

$$\begin{cases} X_t = W_t, \\ Y_t = \sin(X_T + T) + \int_t^T \left( Y_s Z_s - Z_s + \frac{5}{2} Y_s - \sin(s + X_s)\cos(s + X_s) - 2\sin(s + X_s) \right) ds \\ \qquad\qquad - \int_t^T Z_s \, dW_s. \end{cases}$$

The exact solution is given by

$$(Y_t, Z_t) = (\sin(X_t + t), \cos(X_t + t)).$$

We will take $T = 1$, and note that $(Y_0, Z_0) = (0, 1)$. Note that the driver could have been simplified to $f_t(X_t, Y_t, Z_t) = 0.5Y_t - Z_t$, however with this driver we can test how well the method behaves when it also explicitly depends on $t$ and $X_t$.

Before we will study complete methods, we first want to look at the convergence rate of the BDF methods, given that the first initial steps are taken to be exact. Furthermore we take 10 Picard iterations instead of 5, so that the convergence rate of BDF6 will not be hampered by the errors due to the Picard iterations.

In Figure 8.1 the results are visualised, we can note that we indeed have a $n$-th order convergence rate as we expected. The oscillatory behaviour at errors of around $10^{-12}$ are due to the errors we made in the COS approximation formulae. This 'error plateau' can be lowered if we take $\mathscr{N}$ to be higher, alternatively we can also lower $L$ to 6 without increasing the computation time. As we mentioned before, $L = 6$ is sufficiently large for Brownian motion.

We have not discussed how to compute the initial steps necessary for the BDF methods yet. In this thesis we will only consider simple solutions to this problem. Given what we have already discussed, it is natural to use a combination of Theta schemes and BDF schemes to compute the initial values. For example, if we want to use BDF3, we can use a first step of BDF1, then a step of BDF2, so that we can finally use BDF3 for the rest of the steps. Alternatively, if we know the derivative of the terminal condition, we can also do a Crank-Nicolson step first and then progress in the same manner with BDF2, BDF3, and so on.

We will abbreviate the Crank-Nicolson to CN and we define the '+' to be a concatenation of one method for the first step on the left-hand side and a method for the rest on the right-hand side. In Figure 8.2 we can see first of all that starting with a Crank-Nicolson step is more accurate than starting with a BDF1 step, which is to be expected as Crank-Nicolson is second order while BDF1 is only first order. Furthermore BDF1 is $\mathcal{O}(\Delta t)$ while the CN, 'BDF1 + BDF2', 'CN + BDF2' and 'CN + BDF2 + BDF3' schemes are all of $\mathcal{O}(\Delta t^2)$. Which is more or less expected. In Corollary 7.2.4 we have only shown the $n$-th order convergence rate in the case the initial steps are also computed with another $n$-th order method. In the cases of 'BDF1 + BDF2', 'CN + BDF2 + BDF3' and 'CN + BDF2 + BDF3 + BDF4', this is not satisfied. It turns out that for this example we only need $n-1$-th order accuracy in the initial steps for the the rest of the scheme to convergence with $n$-th order. This will fail, however, if

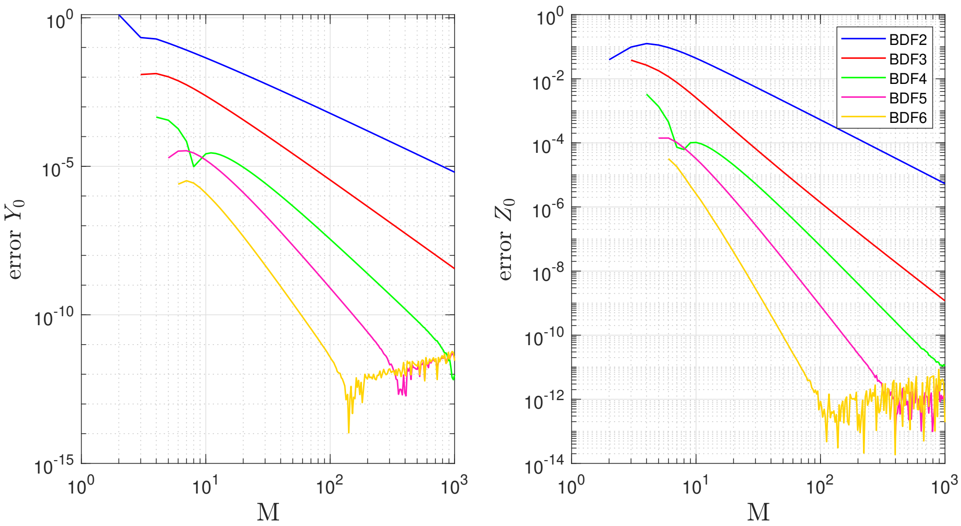


Figure 8.1: Results BDFn methods example 1, with exact initial steps for $\mathcal{N} = 2^{10}$, left: error $\widehat{y}(t_0, x_0)$, right: error $\widehat{z}(t_0, x_0)$.

the initial steps are only of $n-2$-th or lower order of accuracy. In those cases the method will only run at $\min(k+1, n)$-th order convergence where $k$ is the order of the lowest step.

For the BDFn schemes to work for $n \geq 4$, we need to have a method of at least order $n-1$ or preferably of order $n$. We will not go further into such methods, however, this should not be too hard to solve in future research. In Fu, Zhao and Zhou [18] and Tang, Zhao and Zhou [41], deferred correction methods have been used to guarantee $n$-th order convergence, this should also be the case for our schemes.

For the following examples, we will only look at the schemes 'CN', 'BDF1 + BDF2', 'CN + BDF2' and 'CN + BDF2 + BDF3'.

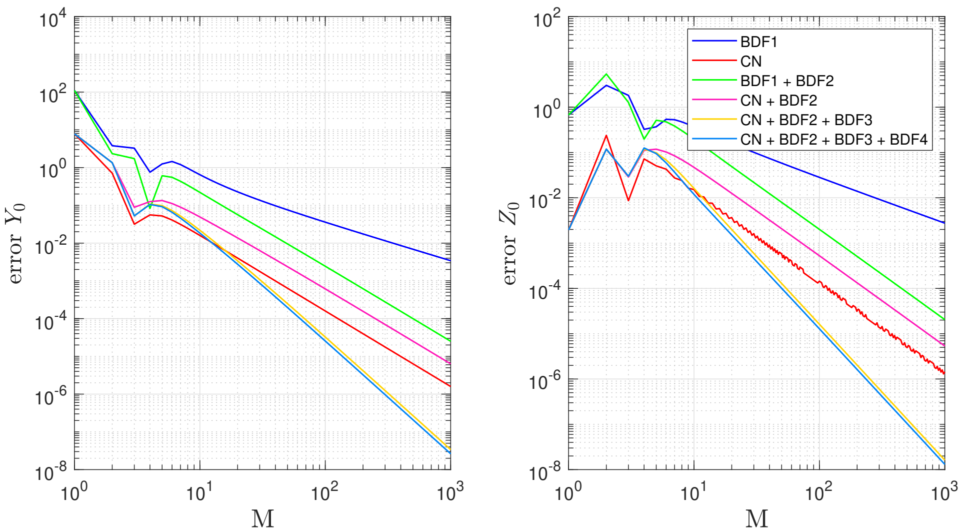


Figure 8.2: Results example 1 for $\mathcal{N} = 2^{10}$, left: error $\widehat{y}(t_0, x_0)$, right: error $\widehat{z}(t_0, x_0)$.

## Example 2: Black-Scholes call option

In this example we will price a European call option with terminal time $T$ and strike price $K$ under the so-called Black-Scholes model. The underlying asset is assumed to follow a geometric Brownian motion,

$$S_t = \int_0^t \mu S_s \, ds + \int_0^t \sigma S_s \, dW_s,$$

with payoff function given by $g(S_t) = (S_t - K)^+$. By the change of variables $X_t = \log(S_t)$, we have the following FBSDE for the option price $Y_t = v(t, \log(S_t))$.

$$\begin{cases} X_t = X_0 + \int_0^t \mu \, ds + \int_0^t \sigma \, dW_s \\ Y_t = (\exp(X_T) - K)^+ + \int_t^T \left( \frac{r-\mu}{\sigma} Z_t - rY_s \right) ds - \int_t^T Z_s \, dW_s, \end{cases}$$

The corresponding semilinear PDE is given by the famous Black-Scholes equation

$$\begin{cases} \frac{\partial v}{\partial t} + r\frac{\partial v}{\partial x} + \frac{1}{2}\sigma^2 \frac{\partial^2 v}{\partial x^2} - rv = 0, \\ v(T,x) = (\exp(x) - K)^+. \end{cases}$$

The exact solution of this PDE is given by the Black-Scholes formula

$$\begin{aligned} v(t, S_t) &= N(d_1)S_t - N(d_2)Ke^{-r(T-t)}, \\ N(x) &= \frac{1}{\sqrt{2\pi}} \int_{-\infty}^x e^{-z^2/2} \, dz \\ d_1 &= \frac{1}{\sigma\sqrt{T-t}} \left( \log\left(\frac{S_t}{K}\right) + \left(r + \frac{\sigma^2}{2}\right)(T-t) \right) \\ d_2 &= d_1 - \sigma\sqrt{T-t}. \end{aligned}$$

We will take the parameters

$$\mu = 0.2, \sigma = 0.25, K = 100, S_0 = 100, r = 0.1.$$

In Figure 8.3, the results can be found for the Black-Scholes option valuation of an European call. Again as in Figure 8.2 we can see mild oscillatory behaviour in the error of the $Z$-process for the Crank-Nicolson scheme. Furthermore, 'CN', 'BDF1 + BDF2' and 'CN + BDF2' again have second order convergence while 'CN + BDF2 + BDF3' has third order convergence.

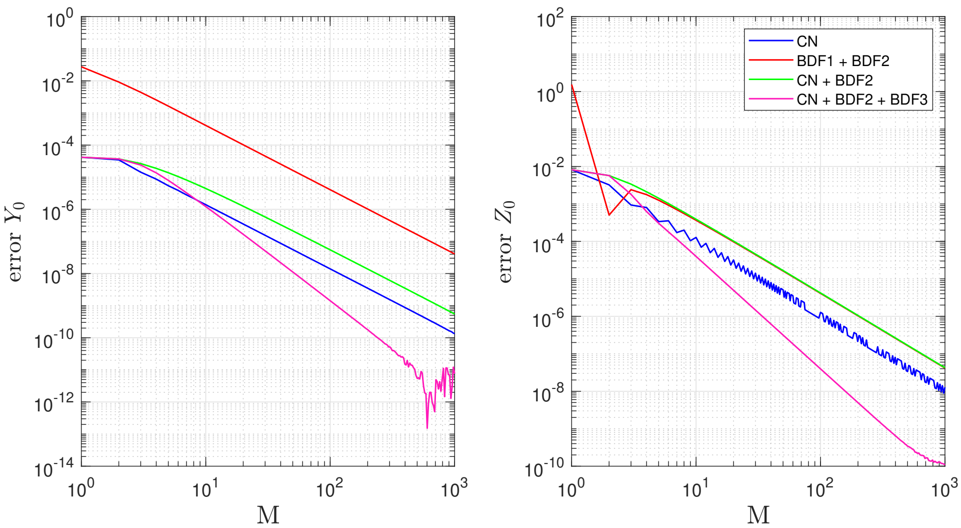


Figure 8.3: Results example 2 for $\mathcal{N} = 2^{10}$, left: error $\widehat{y}(t_0, x_0)$, right: error $\widehat{z}(t_0, x_0)$.

Table 8.1 shows the running time of the Crank-Nicolson scheme and the BDF2 scheme starting with one step of Crank-Nicolson, in seconds. Here, we observe that the Crank-Nicolson is twice as fast as the BDF2 scheme. This is due to the fact that the most time is spent calculating the characteristic function and the $\Phi_Y$ and $\Phi_Z$, because the BDF2 scheme has to do twice the amount of those calculations as the Crank-Nicolson scheme does. Furthermore we can see that the running time of both schemes increase linearly in $\mathcal{M}$ and is of order $\mathcal{O}(\mathcal{N} \log(\mathcal{N}))$ in $\mathcal{N}$, as expected.

| $\mathscr{M}\,(\mathscr{N} = 2^{10})$ | 8 | 16 | 32 | 64 | 128 | 256 | 512 |
|---|---|---|---|---|---|---|---|
| CN | 0.0584 | 0.0727 | 0.1077 | 0.1599 | 0.2633 | 0.4984 | 0.9548 |
| CN + BDF2 | 0.0861 | 0.1193 | 0.1700 | 0.2634 | 0.4884 | 0.8933 | 1.7147 |

| $\mathscr{N}(\mathscr{M} = 512)$ | $2^6$ | $2^7$ | $2^8$ | $2^9$ | $2^{10}$ |
|---|---|---|---|---|---|
| CN | 0.0400 | 0.0514 | 0.0639 | 0.1259 | 0.9548 |
| CN + BDF2 | 0.0657 | 0.0880 | 0.1176 | 0.4381 | 1.7147 |

Table 8.1: Running time example 4 in seconds. Upper table with $\mathscr{N} = 2^{10}$, lower table with $\mathscr{M} = 512$.

## Example 3: Black-Scholes bid-ask spread for interest rates

We now look at another option valuation problem introduced in Bender and Steiner [3]. Suppose that a trader can invest into a riskfree bond with rate $r \geq 0$ for investing and a rate $R \geq 0$ for borrowing from the bond. We want to value a European call spread option with payoff function $g(s) = (s-K_1)^+ - 2(s-K_2)^+$ for two strike prices $K_1, K_2 > 0$. We take the constants as in Lemor, Gobet and Warin [27]

$$x_0 = 100, \mu = 0.05, \sigma = 0.2, T = 0.25, r = 0.01, K_1 = 95, K_2 = 105.$$

Now we check for two borrowing rates $R = 0.06$ and $R = 3.01$ as in Bender and Steiner [3]. The borrowing rate $R = 3.01$ is not economically plausible, however the driver will have a much bigger Lipschitz constant, hence we can test the dependence of stability and convergence of the FBSDE on the Lipschitz constant of the driver. Under the Black-Scholes model, we have the following FBSDE for the option price $Y_t = v(t, \log(S_t))$,

$$\begin{cases} X_t = X_0 + \int_0^t \mu \, ds + \int_0^t \sigma \, dW_s \\ Y_t = (\exp(X_T) - K_1)^+ - 2(\exp(X_T) - K_2)^+ + \int_t^T \left( \frac{r-\mu}{\sigma} Z_t - rY_s + (R-r)(Z_t/\sigma - Y_t)^+ \right) ds \\ \qquad - \int_t^T Z_s \, dW_s, \end{cases}$$

Note that term $(R-r)(Z_t/\sigma - Y_t)^+$ will now make the driver non-differentiable in $Z_t = \sigma Y_t$. Or in terms of the option price, $\frac{\partial v}{\partial x} = v$. Furthermore there is no known closed form solution to this BSDE, hence we make use of numerically computed reference values. This we do by taking the 'CN + BDF2' scheme with $\mathscr{N} = 2^{12}$ and $M = 10^5$. For $R = 0.06$ we find $Z_0 = 0.553258567906919$ and $Y_0 = 2.958453653597437$, for $R = 3.01$ we find $Z_0 = -4.689739171048050$ and $Y_0 = 6.375127533471622$.

In Figure 8.4, the results for the case $R = 0.06$ are plotted. The oscillatory behaviour in the error of the $Z$-process for the Crank-Nicolson scheme is now much more pronounced. The error in the $Y$-process is of second-order for all the second-order schemes, as well for the 'CN + BDF2 + BDF3' scheme, despite the previous results. The error in the $Z$-process is however much more worrisome. The convergence rate is hard to read off for any of the schemes. Assuming monotonicity of the error for the 'CN + BDF2' scheme, our reference value is only reasonably accurate for the first 6 digits of the $Z_0$ value, after that it is possibly inaccurate. Hence, the schemes do not convergence well or much slower than expected. This is possibly due to discontinuities in the driver, as we did assume high smoothness conditions of $f$ in Lemma 7.2.1, which are violated for this problem.

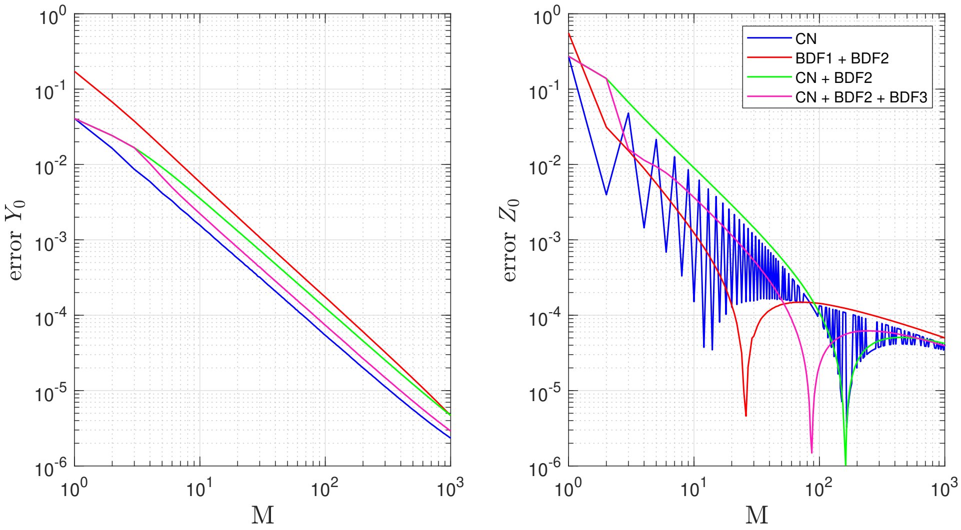


Figure 8.4: Results example 3 for $R = 0.06$ and $\mathcal{N} = 2^{10}$, left: error $\widehat{y}(t_0, x_0)$, right: error $\widehat{z}(t_0, x_0)$.

Figure 8.5 shows the results for the case $R = 3.01$. It is interesting to note that only the 'BDF1 + BDF2' scheme has second order convergence for the $Y$-process. It turns out that a first step of Crank-Nicolson gives bad results for the entire scheme even in the case 'CN + BDF2'. For the $Z$-process we again have odd behaviour, but it does seem to have convergent behaviour for $M = 10$ and higher. The order of convergence is hard to read off, and we again see some oscillatory behaviour for the Crank-Nicolson scheme.

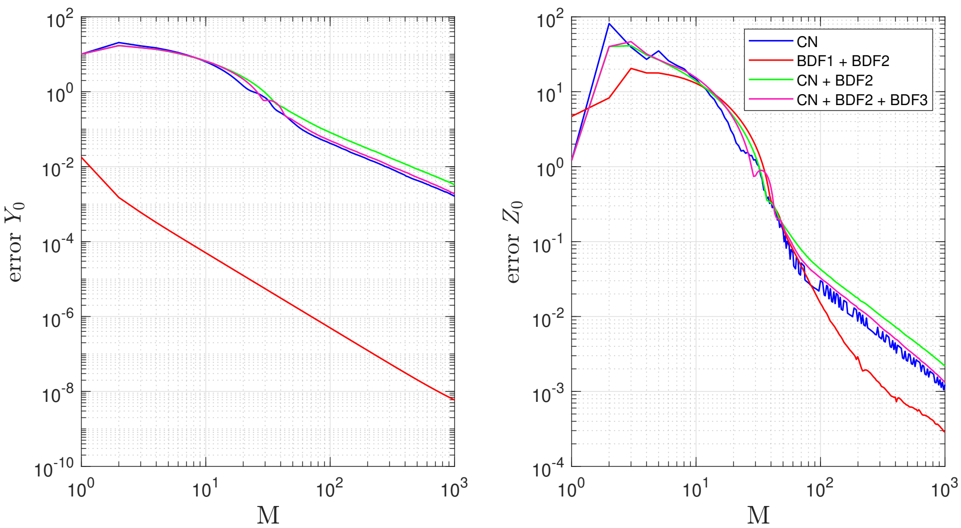


Figure 8.5: Results example 3 for $R = 3.01$ and $\mathcal{N} = 2^{10}$, left: error $\widehat{y}(t_0, x_0)$, right: error $\widehat{z}(t_0, x_0)$.

The following two examples are variants of Example 1, but then with an extra jump term, a drift term, and a more complicated driver.

## Example 4

The first example has a Merton jump-diffusion process as FSDEJ, Example 6.1.6. So we have $\mu, \overline{\mu} \in \mathbb{R}$ and $\sigma, \overline{\sigma}, \lambda > 0$, with Lévy measure

$$\nu(J) = \frac{\lambda}{\sqrt{2\pi\overline{\sigma}^2}} e^{-\frac{(J-\overline{\mu})^2}{2\overline{\sigma}^2}}.$$

Furthermore we let $\eta(J) = 1$ and we have the following FBSDEJ,

$$\begin{cases} X_t = X_0 + \int_0^t \mu\, ds + \int_0^t \sigma\, dW_s + \int_0^t \int_{\mathbb{R}_0} J\, \widetilde{N}(ds, dJ), \\ Y_t = \sin(X_T + T) + \int_t^T \left( \frac{1}{2}\sigma^2 Y_s - \frac{1 + \mu - \lambda\overline{\mu}}{\sigma} Z_s - \Gamma_s \right) ds - \int_t^T Z_s\, dW_s \\ \qquad - \int_t^T \int_{\mathbb{R}_0} U_s(J)\, \widetilde{N}(ds, dJ), \end{cases}$$

We will take the parameters

$$x_0 = \log(10), \mu = 0.2, \sigma = 0.25, \overline{\mu} = 0.5, \overline{\sigma} = 0.05, \lambda = 1.$$

The exact solution is given by

$$\begin{cases} Y_t = \sin(X_t + t), \\ Z_t = \sigma \cos(X_t + t), \\ \Gamma_t = \lambda \left( \sin(X_t + t + \overline{\mu}) \exp\left( -\frac{1}{2}\overline{\sigma}^2 \right) - \sin(X_t + t) \right). \end{cases}$$

In Figure 8.6 we have plotted the results. Again observe the oscillatory behaviour of the Crank-Nicolson scheme in both the $Z$-process and the $\Gamma$-process, furthermore we have stable second order convergence in all the three processes for the 'BDF1 + BDF2' and the 'CN + BDF2' schemes. We also have roughly third order convergence for the 'CN + BDF2 + BDF3' scheme. In Table 8.2 the running time of the schemes are shown. The running time of the Crank-Nicolson scheme is comparable with the running time for continuous BSDEs, the BDF scheme seems to be, however, a bit slower than for continuous BSDEs.

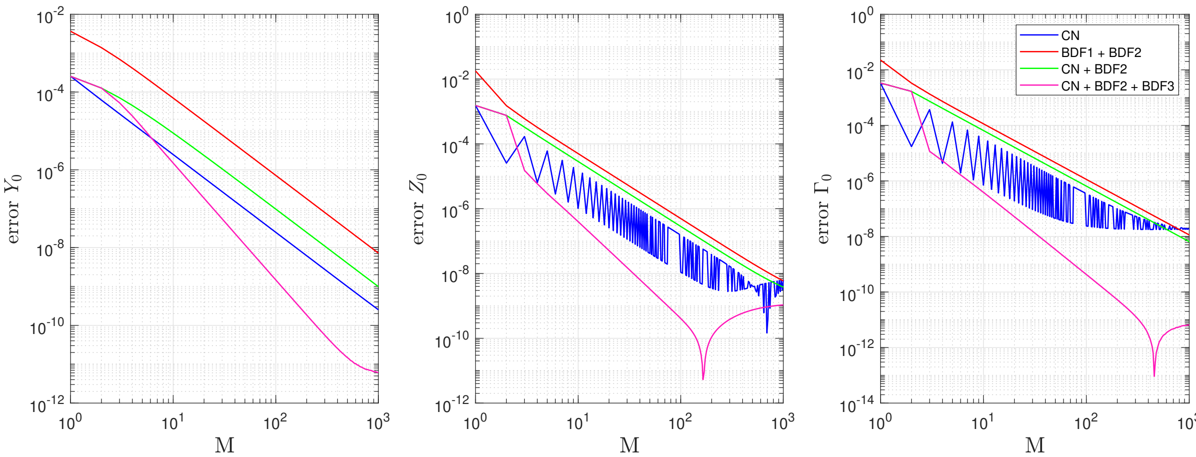


Figure 8.6: Results example 4 for $\mathcal{N} = 2^{10}$, from left to right: error $\widehat{y}(t_0, x_0)$, error $\widehat{z}(t_0, x_0)$ and error $\widehat{\gamma}(t_0, x_0)$.

| $\mathcal{M}(\mathcal{N} = 2^{10})$ | 8 | 16 | 32 | 64 | 128 | 256 | 512 |
|---|---|---|---|---|---|---|---|
| CN | 0.0528 | 0.0651 | 0.1009 | 0.1582 | 0.2690 | 0.4873 | 0.9485 |
| CN + BDF2 | 0.1111 | 0.1459 | 0.2053 | 0.3412 | 0.6297 | 1.1873 | 2.2206 |

| $\mathcal{N}(\mathcal{M} = 512)$ | $2^6$ | $2^7$ | $2^8$ | $2^9$ | $2^{10}$ |
|---|---|---|---|---|---|
| CN | 0.0534 | 0.0719 | 0.0941 | 0.2339 | 0.9548 |
| CN + BDF2 | 0.0588 | 0.0704 | 0.1344 | 0.6420 | 2.2206 |

Table 8.2: Running time example 5 in seconds. Upper table with $\mathcal{N} = 2^{10}$, lower table with $\mathcal{M} = 512$.

## Example 5

The final example is a generalized version of the Merton jump-diffusion model. Here we consider a drift term and diffusion term which are time-dependent. Furthermore the diffusion term is not always non-zero, so we could potentially run into problems, as at those time points we do not necessarily have a continuous density. Furthermore we have a driver which is nonlinear in both $Y$ and $Z$.

$$\begin{cases} X_t = X_0 + \int_0^t \sin(s)\,ds + \int_0^t \cos(s)\,dW_s + \int_0^t \int_{\mathbb{R}_0} J\,\widetilde{N}(ds, dJ), \\ Y_t = \sin(X_T + T) + \int_t^T \left( \frac{1}{2}\cos^2(t) Y_s + \frac{1}{1 + Y_s^2 + Z_s^2} - \frac{1}{1 + \sin^2(X_t + t) + \cos^2(t)\cos^2(X_t + t)} \right. \\ \qquad \left. - \cos(t + x)(1 + \sin(t) - \lambda\overline{\mu}) - \Gamma_s \right) ds - \int_t^T Z_s\,dW_s - \int_t^T \int_{\mathbb{R}_0} U_s(J)\,\widetilde{N}(ds, dJ). \end{cases}$$

The exact solution is given by

$$\begin{cases} Y_t = \sin(X_t + t), \\ Z_t = \cos(t)\cos(X_t + t), \\ \Gamma_t = \lambda\left(\sin(X_t + t + \overline{\mu})\exp\left(-\frac{1}{2}\overline{\sigma}^2\right) - \sin(X_t + t)\right). \end{cases}$$

In Figure 8.7 we can see that the Crank-Nicolson scheme has become a first order scheme for all three processes, while the BDF2 schemes are still of second order convergence. A possible reason for the Crank-Nicolson scheme to now be first order is that we use a forwawrd Euler discretisation in the $X$-process which is of only first order convergence. In the case of our BDF schemes we found that such a discretisation gives no extra error due to arguments with the generator. However, these arguments do not apply to the discretisation of the Theta method, as the discretisation was done with respect to the integral equations (7.2), (7.3) and (7.4) instead of the reference ODEs.

In contrary to the previous example, the 'CN + BDF2 + BDF3' is now also of second order convergence for all three processes. Finally in Table 8.3 we can observe that the running time is considerably higher than for our previous example, this is due to the fact that we now have to compute the characteristic functions and the $\Phi_Y, \Phi_Z, \Phi_\Gamma$ for every time-step.

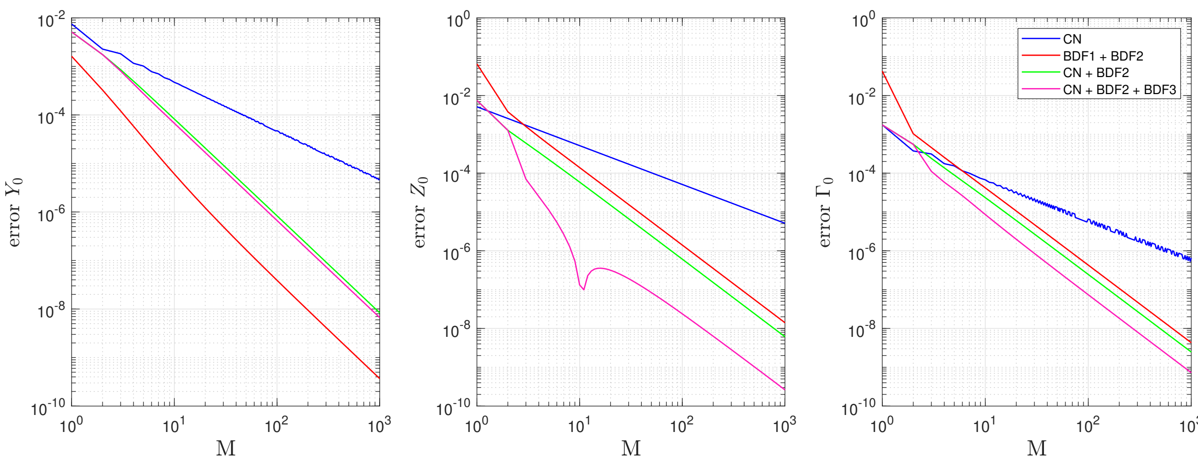


Figure 8.7: Results example 5 for $\mathcal{N} = 2^{10}$, from left to right: error $\widehat{y}(t_0, x_0)$, error $\widehat{z}(t_0, x_0)$ and error $\widehat{\gamma}(t_0, x_0)$.

| $\mathscr{M}(\mathscr{N} = 2^{10})$ | 8 | 16 | 32 | 64 | 128 | 256 | 512 |
|---|---|---|---|---|---|---|---|
| CN | 0.2197 | 0.3133 | 0.5933 | 1.1188 | 2.1864 | 4.3507 | 8.0924 |
| CN + BDF2 | 0.4542 | 0.6190 | 1.0871 | 1.9728 | 3.6929 | 7.6845 | 14.7482 |

| $\mathscr{N}(\mathscr{M} = 512)$ | $2^6$ | $2^7$ | $2^8$ | $2^9$ | $2^{10}$ |
|---|---|---|---|---|---|
| CN | 0.1197 | 0.2174 | 0.4465 | 2.2394 | 8.0924 |
| CN + BDF2 | 0.1669 | 0.3194 | 0.8073 | 4.5145 | 14.7482 |

Table 8.3: Running time example 2 in seconds. Upper table with $\mathscr{N} = 2^1 0$, lower table with $\mathscr{M} = 512$.

CHAPTER 9

# Conclusions and Further Research

## Conclusions

In this thesis we have presented new probabilistic numerical methods for solving Forward Backward Stochastic Differential Equations with Jumps (FBSDEJs). For the semi-discretisation we rewrote the FBSDEJ into a system of ODEs containing conditional expectations. For solving the ODEs, we used the Backward Differentiation Formula (BDF) methods. The computation of the conditional expectations was based on the COS method, which was developed in Fang and Oosterlee [17]. This method approximates conditional expectations based on Fourier cosine series expansions and the characteristic function of the underlying stochastic process.

For Lévy-Itô processes with bounded deterministic coefficients, non-zero diffusion term and finite Lévy measure, we found that the Fourier cosine series expansions converge faster than polynomially in the amount of cosine terms. From this we could conclude that under these conditions, the COS method is very efficient in the computation of the conditional expectations. Furthermore we have proven $n$-th order convergence of the BDFn methods under strict assumptions in the $Y$-process where the coefficients are sufficiently smooth. For the $Z$ and $\Gamma$-process we lose a half-order convergence rate, however these results are most likely non-optimal, as we also see $n$-th order convergence in the numerical results.

In the numerical examples we have seen the necessity of smoothness for the driver function, as otherwise $n$-th order convergence is not at all guaranteed. The BDF methods give very satisfying results, they are very stable and reliable. We pay for this robustness in terms of computational speed in a doubling of the running time for the BDF2 method with respect to a Crank-Nicolson scheme. When a suitable method has been found to approximate the initial steps sufficiently accurate for the BDFn methods, we have shown the promising high order convergence for large $n$. Since they scale much better with the time step size, they could solve the FBSDEJs much more efficiently.

To compare the BDFn methods with the Crank-Nicolson method, the scheme as presented in Ruijter and Oosterlee [36] had to be extended for the case of FBSDEJs rather than just FBSDEs. We found that doing a first step of Crank-Nicolson decreases the error in the first step significantly, instead of a fully implicit scheme like a BDFn method, at no extra cost in the running time. When creating a method to compute the initial steps of the BDFn methods, it might be worthwhile to propose an at least partially explicit scheme, so that we can make use of the terminal conditions of the $Y, Z$ and $\Gamma$-processes, due to the Feynman-Kac representation, rather than just the terminal condition on the $Y$-process.

## Further Research

There still remains a great deal to investigate for solving FBSDEJs in greater generality. Further, the method could be extended to different kinds of stochastic differential equations, which also have a backward component and an underlying forward stochastic process. We summarise the subjects for further research in the following:

- As we have discussed, the computation of the initial steps for the BDFn methods is unsatisfactory, we could only get BDF2 and sometimes BDF3 to work properly. A lot could be won, without having to do much more work, by choosing a more appropriate method to compute these intial values. In Fu, Zhao and Zhou [18] but also Tang, Zhao and Zhou [41], deferred correction methods have been proposed and shown to be working to compute these initial values. Alternatively we could see the problem as solving stiff differential equations. Most methods designed for stiff ODEs can be converted readily to our framework, without much work, as the necessary COS approximation formulae are already given in this thesis.

- For more general FBSDEJs where $\beta(J) \neq J$ and $\eta(J) \neq 1$, the expectations in (6.8) have to be approximated. We can do this again by a COS method, where we now work with a different density as before. The error propagation, suitable $\mathcal{N}$ and truncation range still have to be investigated, to ensure proper convergence of the method.

- The case of fully coupled FBSDEJs would be interesting to investigate. Here 'fully-coupled' means that the $\mu, \sigma$ and $\beta$ also depend on the solution $Y, Z$ and $\Gamma$. Since we built-in a forward Euler discretisation in the forward process, we face a nonlinear equation for solving the characteristic functions, due to the fact we compute the $Y, Z$ and $\Gamma$ process backwards in time. Alternatively we could look at a backward Euler discretisation of the forward process, so that we get an explicit method for solving the characteristic functions. Whether this affects the convergence rates for the $Y, Z$ and $\Gamma$ processes for the BDFn methods, should still be investigated. In Huijskens, Ruijter and Oosterlee [22] the fully-coupled FBSDEs have already been considered where the semi-discretisation is based on a theta-discretisation scheme and the conditional expectations are solved with the COS method.

- In this thesis we have only considered jump-diffusion models in the FSDEJ, the case of infinite activity models is still somewhat unresolved. The method as presented here, is unsuitable to extend directly to infinite activity models, as the jump term in the FSDEJ is no longer a compensated compound Poisson process. Hence the COS approximations are all invalid. We know from Theorem 4.3.21 that we can approximate the jump term by taking a smaller set $A \subset \mathbb{R}_0$ which is bounded below, meaning that we take a lower bound on the jump size. If we have a nonzero diffusion term, we can then approximate the infinite activity model with jump-diffusion models, for which we know how to solve the FBSDEJs. There still remain a lot of problems with this method. It is unclear how the new characteristic function and intensity measure relate to the infinite activity model and it is also unclear how this truncation affects the error in the $Y, Z$ and $\Gamma$-processes.

- There are still a lot of gaps in the numerical analysis of the presented methods. Especially in the convergence rate of the semi-discretisation. It turns out to be difficult to prove these for general drivers $f$, however most drivers are in practice of linear or quadratic nature, for these restrictions the results might be able to be proven in more detail and generality with respect to dependence on $Z$ and $\Gamma$. When we take a discretisation scheme different from forward Euler for the forward process (for example for solving fully-coupled FBSDEJs), the error in the FSDEJs also have to be taken into account.

- There exist many variations on the FBSDEJs which are similar in nature to what we have discussed in this thesis. The most straightforward would be to consider a multidimensional system. This has already been studied in the continuous case by using the BCOS method in Ruijter [35]. To name a few different variations. Reflected FBSDEJs (RFBSDEJs) have an additional restriction on $Y$, there is an additional obstacle process which restricts $Y$ from going below certain values. They correspond to obstacle problems for PIDEs, see [15]. 2BSDEJs are the probabilistic representation of certain fully nonlinear PIDEs, and are suprema of families of BSDEJs, they are discussed thoroughly in [24]. Finally we refer in the continuous case to the book of Zhang [45] which contains most of the recent developments of continuous BSDEs and its variants.

# Bibliography


[1] D Applebaum et al. *Lévy Processes and Stochastic Calculus*. Cambridge University Press, 2004. ISBN: 978-0-521-83263-2.

[2] Guy Barles, Rainer Buckdahn, and Etienne Pardoux. "Backward Stochastic Differential Equations and Integral-Partial Differential Equations". In: *Stochastics and Stochastics Reports* 60 (1996), pp. 57–83.

[3] Christian Bender and Jessica Steiner. "Least-Squares Monte Carlo for Backward SDEs". In: *Springer Proceedings in Mathematics* 12 (2012), pp. 257–289. ISSN: 21905614.

[4] Patrick Billingsley. *Probability and measure*. 3rd. New York: Wiley. ISBN: 0471007102.

[5] Jean Michel Bismut. "Conjugate convex functions in optimal stochastic control". In: *Journal of Mathematical Analysis and Applications* 44.2 (1973), pp. 384–404. ISSN: 10960813.

[6] V I Bogachev. *Measure Theory*. Springer, 2007, p. 1075. ISBN: 9783540345138.

[7] J C Butcher. *Numerical Methods for Ordinary Diffrencial Equations*. 2nd ed. Vol. 3. 2. Wiley, 2015, pp. 54–67. ISBN: 9780470723357.

[8] Jean François Chassagneux. "Linear multistep schemes for BSDEs". In: *SIAM Journal on Numerical Analysis* 52.6 (2014), pp. 2815–2836. ISSN: 00361429.

[9] Kai-lai Chung and Ruth J Williams. *Introduction to Stochastic Integration*. 2nd ed. 1990, p. 264.

[10] D L Cohn. *Measure Theory*. English. 2nd ed. Birkhäuser, 1980, p. 457. ISBN: 978-3-7643-3003-3.

[11] Rama Cont and Peter Tankov. *Financial Modelling with Jump Processes*. CRC Press, 2004, p. 552.

[12] M H A Davis. *Markov Models & Optimization*. Routledge, 2018, p. 308.

[13] Claude Dellacherie and Paul-André Meyer. *Martingales, Probabilités et Potential: Theorie des martingales*. Paris, 1987, p. 374.

[14] N El Karoui, Shige Peng, and M C Quenez. "Backward Stochastic Differential Equations in Finance". In: 7.1 (1997), pp. 1–71.

[15] N El Karoui et al. "Reflected solutions of backward SDE's, and related obstacle problems for PDE's". In: *Annals of Probability* 25.2 (1997), pp. 702–737. ISSN: 00911798.

[16] F Fang. "The COS Method: An Efficient Fourier Method for Pricing Financial Derivatives". PhD thesis. Delft University of Technology, 2010.

[17] F Fang and C W Oosterlee. "A novel pricing method for european options based on fourier-cosine series expansions". In: *SIAM Journal on Scientific Computing* 31.2 (2008), pp. 826–848. ISSN: 10648275.

[18] Yu Fu, Weidong Zhao, and Tao Zhou. "Multistep Schemes for Forward Backward Stochastic Differential Equations with Jumps". In: *Journal of Scientific Computing* 69.2 (2016), pp. 651–672. ISSN: 08857474.

[19] L Grafakos. *Classical Fourier analysis*. en. Third edit. New York: Springer, 2014, p. 638. ISBN: 978-1-4939-1193-6.

[20] Mohammed Hassani and Youssef Ouknine. "Infinite dimensional BSDE with jumps". In: *Stochastic Analysis and Applications* 20.3 (2002), pp. 519–565.

[21] Peter Henrici. *Discrete Variable Methods in Ordinary Differential Equations*. Wiley, 1962, p. 407.

[22] T P Huijskens, M J Ruijter, and C W Oosterlee. "Efficient numerical Fourier methods for coupled forward-backward SDEs". In: *Journal of Computational and Applied Mathematics* 296 (2016), pp. 593–612. ISSN: 03770427.

[23] Monique Jeanblanc et al. "Utility Maximization with Random Horizon: A BSDE Approach". In: *International Journal of Theoretical and Applied Finance* 18.7 (2015).

[24] Nabil Kazi-Tani, Dylan Possamaï, and Chao Zhou. "Second-order bsdes with jumps: Formulation and uniqueness". In: *Annals of Applied Probability* 25.5 (2015), pp. 2867–2908. ISSN: 10505164.

[25] Hiroshi Kunita. *Representation of Martingales with Jumps and Applications to Mathematical Finance.* Tech. rep. 2004, pp. 209–232.

[26] Jean-François Le Gall. *Brownian Motion, Martingales and Stochastic Calculus.* 1st ed. Vol. 274. Springer International Publishing, 2016, p. 273.

[27] Jean Philippe Lemor, Emmanuel Gobet, and Xavier Warin. "Rate of convergence of an empirical regression method for solving generalized backward stochastic differential equations". In: *Bernoulli* 12.5 (2006), pp. 889–916. ISSN: 13507265.

[28] Cornelis W Oosterlee and Lech A Grzelak. "Mathematical Modeling and Computation in Finance". In: *Mathematical Modeling and Computation in Finance* (2019).

[29] Steven Orey. *On Continuity Properties of Infinitely Divisible Distribution Functions.* Tech. rep. 3. 1968, pp. 936–937.

[30] Etienne Pardoux and Shige Peng. "Adapted solution of a backward stochastic differential equation". In: *Systems & Control Letters* 14 (1990), pp. 55–61.

[31] Shige Peng. "Backward stochastic differential equation, nonlinear expectation and their applications". In: *Proceedings of the International Congress of Mathematicians 2010, ICM 2010* 2007 (2010), pp. 393–432.

[32] Shige Peng and Yufeng Shi. "Infinite horizon forward-backward stochastic differential equations". In: *Stochastic Processes and their Applications* 85.1 (2000), pp. 75–92. ISSN: 03044149.

[33] P Protter. *Stochastic Modelling and Applied Probability.* Springer Berlin Heidelberg, 2013. ISBN: 9783642055607.

[34] J S Rosenthal. *A First Look at Rigorous Probability Theory.* World Scientific, 2000. ISBN: 978-981-02-4322-7.

[35] M J Ruijter. "Fourier Methods for Multidimensional Problems and Backward SDEs in Finance and Economics". PhD thesis. Delft University of Technology, 2015. ISBN: 9789462595262.

[36] M J Ruijter and C W Oosterlee. "A Fourier-cosine method for an efficient computation of solutions to BSDEs". en. In: (), p. 30.

[37] K Sato. *Lévy Processes and Infinitely Divisible Distributions.* Cambridge University Press. ISBN: 978-0-521-55302-5.

[38] S E Shreve. *Stochastic Calculus for Finance II: Continuous-Time Models.* Springer, 2004. ISBN: 978-0-387-40101-0.

[39] Steven E Shreve. *Stochastic Calculus for Finance I: The Binomial Asset Pricing Model.* 2003.

[40] Shanjian Tang and Xunjing Li. "Necessary Conditions for Optimal Control of Stochastic Systems with Random Jumps". In: *SIAM Journal on Control and Optimization* 32.5 (1994), pp. 1447–1475.

[41] Tao Tang, Weidong Zhao, and Tao Zhou. "Deferred Correction Methods for Forward Backward Stochastic Differential Equations". In: *Numerical Mathematics* 10.2 (2017), pp. 222–242. ISSN: 20797338.

[42] Peter Tankov. *Financial modeling with Lévy processes.*

[43] Zhen Wu. "Forward-Backward Stochastic Differential Equations with Brownian Motion and Poisson Process". In: *Acta Mathematicae Applicatae Sinica* 15.4 (1999).

[44] Jie Yang and Weidong Zhao. "Convergence of recent multistep schemes for a forward- backward stochastic differential equation". In: *East Asian Journal on Applied Mathematics* 5.4 (2015), pp. 387–404. ISSN: 20797370.

[45] Guannan Zhang et al. "Numerical methods for a class of nonlocal diffusion problems with the use of backward SDEs". In: *Computers and Mathematics with Applications* 71.11 (2016), pp. 2479–2496. ISSN: 08981221.

[46] Weidong Zhao, Jinlei Wang, and Shige Peng. "Error estimates of the $\theta$-scheme for backward stochastic differential equations". In: *Discrete and Continuous Dynamical Systems - Series B* 12.4 (2009), pp. 905–924. ISSN: 15313492.