Implementation of Numerical Fourier Method for Second order Taylor Schemes

Qaphela Mashalaba

A dissertation submitted to the Faculty of Commerce, University of Cape Town, in partial fulfilment of the requirements for the degree of Master of Philosophy.

August 1, 2019

MPhil in Mathematical Finance,
University of Cape Town.

Declaration

I declare that this dissertation is my own, unaided work. It is being submitted for the Degree of Master of Philosophy in the University of the Cape Town. It has not been submitted before for any degree or examination in any other University.

Signed by candidate

August 1, 2019

Abstract

The problem of pricing contingent claims in a complete market has received a significant amount of attention in literature since the seminal work of Black, Fischer and Scholes, Myron (1973). It was also in 1973 that the theory of backward stochastic differential equations (BSDEs) was developed by Bismut, Jean-Michel (1973), but it was much later in the literature that BSDEs developed links to contingent claim pricing.

This dissertation is a thorough exposition of the survey paper Ruijter, Marjon J and Oosterlee, Cornelis W (2016) in which a highly accurate and efficient Fourier pricing technique compatible with BSDEs is developed and implemented. We prove our understanding of this technique by reproducing some of the numerical experiments and results in Ruijter, Marjon J and Oosterlee, Cornelis W (2016), and outlining some key implementationl considerations.

Acknowledgements

I express deep gratitude to my mother, who with all her energies and love, reared my siblings and I through some of the most adverse set of circumstances. Amongst loving friends and family who have been a constant support throughout this journey, special thanks to Sinqobile Mashalaba for his timely encouragement as the time to submit drew near. To the Mandela Rhodes Foundation for their friendship, willingness to accommodate my unusual circumstances and their generous financial support, I offer my gratitude.

Last, but by no means the least, I offer the highest gratitude to Professor Thomas McWalter for his patience, willingness to assist me and for the provision of insights which invariably aided the completion of this work.

Contents

List of Figures

List of Tables

Chapter 1

Introduction

The theory of pricing contingent claims in a complete market made its first significant step with the seminal work of Black, Fischer and Scholes, Myron (1973) in which a formula was derived for pricing European claims. Alongside this was the publication of Merton's pioneering idea that one can synthetically create a European option by holding continuously rebalanced quantities of stocks and bonds. These early contributions in pricing claims in complete markets were followed by several works, among the most influential being Harrison, J Michael and Kreps, David M (1979) and Harrison, J Michael and Pliska, Stanley R (1981).

The development of the theory of contingent claim pricing in 1973 coincided with the development of the theory of backward stochastic differential equations (BSDEs) by Bismut, Jean-Michel (1973). A BSDE is a stochastic differential equation of the form

$$dY_s = -f(s, Y_s, Z_s)ds + Z_s dW_s,\ 0 \leq s \leq T,\ Y_T = \xi, \tag{1.1}$$

where a terminal condition corresponding to some time $T > 0$ is specified; in contrast, for forward stochastic differential equations (FSDEs), an initial condition is specified. Bismut, Jean-Michel (1973) introduced these equations in the case where f was linear, whereas Pardoux, Etienne and Peng, Shige (1990) extended this to allow for a general form of f, proving an important existence and uniqueness result in the process.

It was against this background that connections between finance and BSDEs emerged (see, for instance, El Karoui, Nicole and Peng, Shige and Quenez, Marie Claire (1997)). In particular, many European pricing problems can be cast in the BSDE framework so that $Y = \{Y_s\}_{s\in[0,T]}$ represents the value of the replicating portfolio and $Z = \{Z_s\}_{s\in[0,T]}$ relates to the hedging strategy. To see an instance of this, suppose that $\{a_s\}_{s\in[0,T]}$ represents the delta of some European claim that is replicated with simultaneous holdings in the underlying stock $\{S_s\}_{s\in[0,T]}$ and a bond

$B = \{B_s\}_{s\in[0,T]}$ returning a risk-free rate of $r > 0$. Then the replicating portfolio $Y = \{Y_s\}_{s\in[0,T]}$ has dynamics

$$dY_s = a_s\left(rS_s ds + \sigma S_s dW_s\right) + r(Y_s - a_s S_s)ds,\ 0 \le s \le T,\ Y_T = \xi, \tag{1.2}$$

where σ is the diffusion of the stock and ξ is the option payoff function. One then observes that $f(s, y, z) = ry$ and $Z_s = a_s \sigma S_s$.

Part of the aim of this dissertation is to price claims by solving the corresponding BSDE problem numerically. Attempts in the literature to solve BSDEs have relied on empirical regression and the use of Monte Carlo simulation (see, for instance, Gobet, Emmanuel and Lemor, Jean-Philippe and Warin, Xavier and others (2005), Lemor, Jean-Philippe and Gobet, Emmanuel and Warin, Xavier and others (2006) and Bender, Christian and Steiner, Jessica (2012)). A new, efficient method, the 'BCOS method' for numerically solving BSDEs was developed by Ruijter, Marjon J and Oosterlee, Cornelis W (2015); this was in fact an extension of the 'COS method' developed in Fang, Fang and Oosterlee, Cornelis W (2008), a European option pricing method that relies on Fourier-cosine expansions. The BCOS method, similar to the COS method, relied on the availability of the transition characteristic function of the FSDE process. This characteristic function is available for the family of Levy processes (a large class of processes) and can be computed using the Levy-Khintchine result. However, in instances where this characteristic function does not exist, this version of the BCOS method is inapplicable. To circumvent this restriction, Ruijter, Marjon J and Oosterlee, Cornelis W (2016) then extended the BCOS method by using the transition characteristic function of the *discrete* process, that they prove to exist whenever the FSDE is discretized using either one of the Euler, Milstein or second order weak simplified weak Taylor schemes.

Numerically solving BSDEs results in backward induction schemes. Since in most settings, the processes involved are adapted and cannot anticipate the future, conditional expectations arise naturally. The BCOS method approximates these conditional expectations by relying on Fourier-cosine expansions (as opposed to a regression and Monte Carlo based method).

The claim pricing procedure requires the discretization of the FSDE process using either one of the Euler, Milstein or second order weak simplified weak Taylor schemes (see Ruijter, Marjon J and Oosterlee, Cornelis W (2016)). The idea is to compute the pricing errors due to Euler and Milstein discretization schemes, and to benchmark these against second order weak simplified weak Taylor scheme pricing errors in the context where the pricing method is the extended version of the

BCOS method given in Ruijter, Marjon J and Oosterlee, Cornelis W (2016).

The remainder of the dissertation is organized as follows. Chapter 2 develops the mathematical foundation that forms the basis of the BCOS method. Chapter 3 provides a comprehensive development of the version of the BCOS method in Ruijter, Marjon J and Oosterlee, Cornelis W (2016). Chapter 4 summarizes the numerical results and is followed by the conclusion.

Chapter 2

Mathematical Preliminaries

The aim of this chapter is to provide the background theory required for the numerical experiments that follow later. This development can be found in the survey paper Ruijter, Marjon J and Oosterlee, Cornelis W (2016).

2.1 Backward Stochastic Differential Equation

Henceforth, let $(\Omega, \mathcal{F}, \mathbb{P})$ be a probability space that supports a one-dimensional standard Brownian motion $W = \{W_s\}_{s\in[0,T]}$ over a finite time horizon $T > 0$. Let $(\Omega, \mathcal{F}, \mathbb{P})$ be equipped with $\mathrm{F} = \{\mathrm{F}_s\}_{s\in[0,T]}$, the augmented natural filtration generated by W. The BSDEs we are interested in have the form

$$dY_s = -f(s, Y_s, Z_s)ds + Z_s dW_s,\ 0 \leq s \leq T,\ Y_T = \xi, \tag{2.1}$$

where $f : \Omega \times [0, T] \times \mathbb{R} \times \mathbb{R} \to \mathbb{R}$ is $\mathcal{P} \otimes \mathbb{B} \otimes \mathbb{B}$-measurable. As usual, $\mathbb{B}$ is the Borel σ-field on $\mathbb{R}$, while $\mathcal{P}$ is the σ-field generated by the set of progressively measurable scalar-valued processes on $\Omega \times [0, T]$. It may be worth noting that the definition of $\mathcal{P}$ as given in Ruijter, Marjon J and Oosterlee, Cornelis W (2016) is less restrictive than a similar definition, but one based on predictable (equivalently, left-continuous) processes instead, given in El Karoui, Nicole and Peng, Shige and Quenez, Marie Claire (1997). The function f is the *generator* or driver of the BSDE, and ξ, an F_T-measurable random variable, is the terminal condition. While the processes $Y = \{Y_s\}_{s\in[0,T]}$, $Z = \{Z_s\}_{s\in[0,T]}$ and W are one dimensional in this formulation, a corresponding multidimensional development can be found in El Karoui, Nicole and Peng, Shige and Quenez, Marie Claire (1997). We are now ready to render a formal definition of a solution to a BSDE whose form is the same as that of (2.1).

Definition 2.1.1. (Solution of a BSDE).
Let $(\Omega, \mathcal{F}, \mathbb{P})$ be a probability space that supports a one-dimensional standard Brownian motion W over a finite time horizon $T > 0$. Let $(\Omega, \mathcal{F}, \mathbb{P})$ be equipped with $\mathrm{F} = \{\mathrm{F}_s\}_{s\in[0,T]}$, the augmented natural filtration generated by W. Then a solution to (2.1) is a

pair $(Y, Z) = (Y_s, Z_s)_{s\in[0,T]}$ *such that*
i. Y *is a continuous and adapted real-valued process;*
ii. Z *is a predictable real-valued process satisfying* $\int_0^T |Z_s|^2 ds < \infty$, $\mathbb{P}$*-almost surely;*
iii. $\mathbb{P}$*-almost surely, we have that*

$$Y_t = \xi + \int_t^T f(s, Y_s, Z_s)ds - \int_t^T Z_s dW_s,\ 0 \le t \le T. \tag{2.2}$$

Let $\mathbb{L}_T^2(\mathbb{R})$ denote the set of all F_T-measurable square integrable random variables, and let $\mathbb{H}_T^2(\mathbb{R})$ denote the set of all predictable processes $\eta = \{\eta_s\}_{s\in[0,T]}$ such that $\mathbb{E}[\int_0^T |\eta_s|^2 ds] < \infty$. If $\xi \in \mathbb{L}_T^2(\mathbb{R})$, $f(\cdot, 0, 0) \in \mathbb{H}_T^2(\mathbb{R})$ and $f(t, y, z)$ satisfies a uniform Lipschitz condition in the y- and z-components, then the pair (f, ξ) is known as the *standard parameters* for (2.1). According to El Karoui, Nicole and Peng, Shige and Quenez, Marie Claire (1997) (see Theorem 2.1 and corresponding proof) and Pham, Huyên (2009), if (2.1) has standard parameters, then a unique (strong) solution such that $(Y, Z) \in \mathbb{H}_T^2(\mathbb{R}) \times \mathbb{H}_T^2(\mathbb{R})$ is guaranteed to exist. An alternative proof for the existence of a unique solution can also be found in Pardoux, Etienne and Peng, Shige (1990).

Furthermore, as mentioned in Ruijter, Marjon J and Oosterlee, Cornelis W (2016), the adaptedness of (Y, Z) implies that these processes cannot anticipate the future; although it cannot anticipate the future, the solution to (2.1) always has to terminate at ξ. Thus 'backward' refers only to the imposition of a terminal condition (as opposed to the usual initial condition) on the SDE, and not to a solution that is obtained by evolving the BSDE backwards in time. It is important to bear this distinction in mind.

2.2 Forward Backward Stochastic Differential Equation

The pricing technique that is implemented in the numerical experiments arrives at contingent claim prices by solving a decoupled system of an FSDE and a BSDE over $[0, T]$. The FSDE has the form

$$dX_s = \mu(X_s)ds + \sigma(X_s)dW_s,\ 0 \le s \le T,\ X_0 = x, \tag{2.3}$$

where $x \in \mathbb{R}$. The functions $\mu : \mathbb{R} \to \mathbb{R}$ and $\sigma : \mathbb{R} \to \mathbb{R}$ are assumed to be Lipschitz continuous and to satisfy a linear growth condition. The BSDE has the form

$$dY_s = -f(s, X_s, Y_s, Z_s)ds + Z_s dW_s,\ 0 \le s \le T,\ Y_T = \xi, \tag{2.4}$$

with the additional specification that $\xi = g(X_T)$, for some function $g : \mathbb{R} \to \mathbb{R}$. Notice that now $f : \Omega \times [0, T] \times \mathbb{R} \times \mathbb{R} \times \mathbb{R} \to \mathbb{R}$. In order to ensure the existence of a unique solution to (2.4), the function $f(\omega,t,x,y,z):\mathbb{R}\to\mathbb{R}$, when considered as a function of $x\in\mathbb{R}$, and $g(x)$, are assumed to be uniformly continuous in x. Additionally, assume that f satisfies a Lipschitz condition in (y,z) as before, and

$$|f(\omega,t,x,y,z)|+|g(x)|\leq C(1+|x|^p+|y|+|z|),\text{ with } p\geq\frac{1}{2}, \tag{2.5}$$

for $C\in\mathbb{R}$. By the linear growth condition on μ and σ, and by (2.5), it then follows that we have

$$\mathbb{E}[g^2(X_T)]\leq\mathbb{E}[C^2(1+X_T)^{2p}]<\infty$$

(that is, $\xi\in\mathbb{L}^2(\mathbb{R})$); finally, by the linear growth condition on μ and σ and (2.5),

$$\mathbb{E}[\int_0^T f^2(\omega,t,X_t,0,0)dt]\leq\mathbb{E}[\int_0^T C^2(1+2X_t)^{2p}dt]<\infty$$

(that is, $f(\omega,t,x,0,0)\in\mathbb{H}^2(\mathbb{R})$). Therefore, (2.4) has a pair of standard parameters (f,ξ) such that f is Lipschitz continuous in (y,z). By El Karoui, Nicole and Peng, Shige and Quenez, Marie Claire (1997) (Theorem 2.1), (2.4) has a unique solution.

2.3 A Link between FBSDEs and PDEs

This FSDE-BSDE (henceforth, FBSDE) couple is a probabilistic representation of the semi-linear parabolic PDE

$$\begin{aligned}&D_t(t,x)+\mu(x)D_xv(t,x)+\frac{1}{2}\sigma^2(x)D_{xx}v(t,x)+f(t,x,v(t,x),\sigma(x)D_xv(t,x))=0,\\&(t,x)\in[0,T)\times\mathbb{R},\\&v(T,x)=g(x),\ x\in\mathbb{R},\end{aligned} \tag{2.6}$$

where D_x and D_{xx} are first and second derivative operators with respect to x respectively, and D_t is the first derivative operator with respect to time. PDEs, such as the one in (2.6) for example, may or may not have a classical solution (that is, a function v together with its partial derivatives, which satisfy the corresponding PDE). This has led to numerous theoretical developments of weaker notions of solutions to PDEs. One such notion is the so-called *viscosity* solution to a PDE. Let $\mathcal{G}:C^{1,2}([0,T]\times\mathbb{R})\to\mathbb{R}$ be the operator

$$\mathcal{G}(\phi)(t,x)=\mu(x)D_x\phi(t,x)+\frac{1}{2}\sigma^2(x)D_{xx}\phi(t,x)+f(t,x,\phi(t,x),\sigma(x)D_x\phi(t,x)),$$

for $\phi\in C^{1,2}([0,T]\times\mathbb{R})$.

Definition 2.3.1. (Viscosity solution to a PDE) *Let* $v : [0,T] \times \mathbb{R} \to \mathbb{R} \in C([0,T] \times \mathbb{R})$. *Consider the PDE*

$$\begin{aligned} &D_t w(t,x) + \mathcal{G}(w)(t,x) = 0, \ (t,x) \in [0,T) \times \mathbb{R}, \\ &w(T,x) = g(x), \ x \in \mathbb{R}. \end{aligned} \tag{2.7}$$

We say that an uppersemicontinuous function (respectively, a lowersemicontinuous function) v is a viscosity sub-solution (respectively, viscosity super-solution) of (2.7) at $(t,x) \in [0,T) \times \mathbb{R}$ if:

i. $D_t v(t,x) + \mathcal{G}(v)(t,x) \geq 0(\text{respectively}, \leq 0), \ (t,x) \in [0,T) \times \mathbb{R};$
ii. $v(T,x) \leq g(x) \ (\text{respectively}, \geq g(x)), \ x \in \mathbb{R},$
whenever $v - w$ has a maximum (respectively, minimum) at (t,x), where
$w \in C^{1,2}([0,T] \times \mathbb{R})$, with $v(t,x) = w(t,x)$.

If v is both a sub-solution and a super-solution to (2.7) on $[0,T] \times \mathbb{R}$, then it is a viscosity solution.

Remark 2.3.1. One immediate consequence of the viscosity framework is that merely continuous functions are allowed to be solutions to (2.7). It must be noted that not all definitions of viscosity solutions insist on the continuity of the solution. For a comprehensive overview of viscosity solutions, we refer the interested reader to Crandall, Michael G and Ishii, Hitoshi and Lions, Pierre-Louis (1992), Fleming, Wendell H. and Soner, H. Mete (2006) and Pham, Huyên (2009).

The following result will be useful in the forthcoming chapter.

Theorem 2.3.1. *(see Pham, Huyên (2009), p. 145) Let $v \in C^{1,2}([0,T] \times \mathbb{R})$ be a classical solution to (2.6) and suppose there exists a constant $C > 0$ such that, for all $(t,x) \in [0,T] \times \mathbb{R}$,*

$$|v(t,x)| + |\sigma(t,x) D_x v(t,x)| \leq C(1 + |x|).$$

Then the pair (Y,Z), defined by

$$Y_s = v(s, X_s), \ Z_s = \sigma(s, X_s) D_x v(s, X_s), \ t \leq s \leq T, \tag{2.8}$$

is a solution to (2.4), where $X_t = x$.

Conversely, suppose (Y,Z) is the solution to (2.4), then the function defined by $v(t,x) = Y_t^{t,x}$ is a viscosity solution to (2.6).

The above result provides an important link between BSDEs of the form (2.1) and semi-linear PDEs of the form (2.6). Interestingly, the converse statement suggests a candidate solution, in the viscosity framework, to a PDE by means of solving a related BSDE problem. The complete proof is rendered below.

Proof. (see Pham, Huyên (2009)) The forward direction can be proven by a simple application of Itô's lemma on $v \in C^{1,2}([0,T] \times \mathbb{R})$:

$$\begin{aligned} dY_s &= D_t v(s,X_s)ds + D_x v(s,X_s)\Big[\mu(X_s)ds + \sigma(X_s)dW_s\Big] + \frac{D_{xx}v(s,X_s)\sigma^2(X_s)}{2}ds, \\ &= -f(s,X_s,v(s,X_s),\sigma(X_s)D_x v(s,X_s)) + D_x v(s,X_s)\sigma(X_s)dW_s, \text{ (using (2.6))}. \end{aligned} \tag{2.9}$$

For the converse direction, we prove that $v(t,x) = Y_t^{t,x}$ is continuous; doing so ensures the required uppersemicontinuity (respectively, lowersemicontinuity) for the sub-solution (respectively, super-solution). For $(t_1,x_1),(t_2,x_2) \in [0,T] \times \mathbb{R}$, define $X_s^i := X_s^{t_i,x_i}$, with $0 \le t_1 \le s \le t_2$ for $i = 1,2$. By extension, we have that $(Y_s^i, Z_s^i) := (Y_s^{t_i,x_i}, Z_s^{t_i,x_i})$ for $i = 1,2$. By applying Itô's formula to $|Y_s^1 - Y_s^2|^2$ for $s \in [t_1,T]$, we obtain:

$$\begin{aligned} |Y_t^1 - Y_t^2|^2 = |g(X_T^1) - g(X_T^2)|^2 &- \int_t^T |Z_s^1 - Z_s^2|^2 ds \\ &+ 2\int_t^T (Y_s^1 - Y_s^2)\cdot\big(f(s,X_s^1,Y_s^1,Z_s^1) - f(s,X_s^2,Y_s^2,Z_s^2)\big)ds \\ &- 2\int_t^T (Y_s^1 - Y_s^2)(Z_s^1 - Z_s^2)dWs. \end{aligned} \tag{2.10}$$

It can be shown that the last term on the right hand side of the above equation is in fact a local martingale (see Pham, Huyên (2009)), so that by taking expectations on both sides, and applying Fubini-Tonelli theorems to interchange the expectation with integration sign, we obtain:

$$\begin{aligned} \mathbb{E}\big[|Y_t^1 - Y_t^2|^2\big] = \mathbb{E}\big[|g(X_T^1) - g(X_T^2)|^2\big] &- \int_t^T \mathbb{E}\big[|Z_s^1 - Z_s^2|^2\big]ds \\ &+ 2\int_t^T \mathbb{E}\Big[(Y_s^1 - Y_s^2)\cdot(f(s,X_s^1,Y_s^1,Z_s^1) - f(s,X_s^2,Y_s^2,Z_s^2))\Big]ds, \end{aligned} \tag{2.11}$$

recalling that the expectation of the above local martingale is zero. Rearranging, we obtain:

$$
\begin{aligned}
&\mathbb{E}\big[|Y_t^1 - Y_t^2|^2\big] + \int_t^T \mathbb{E}\big[|Z_s^1 - Z_s^2|^2\big]ds \\
&= 2\int_t^T \mathbb{E}\Big[(Y_s^1 - Y_s^2)\cdot(f(s,X_s^1,Y_s^1,Z_s^1) - f(s,X_s^2,Y_s^2,Z_s^2))\Big]ds \\
&\quad +\mathbb{E}\big[|g(X_T^1) - g(X_T^2)|^2\big] \\
&\le 2\int_t^T \mathbb{E}\Big[|Y_s^1 - Y_s^2||f(s,X_s^1,Y_s^1,Z_s^1) - f(s,X_s^2,Y_s^2,Z_s^2)|\Big]ds.
\end{aligned}
\tag{2.12}
$$

With reference to the inequality

$$
(x+y)^2 \le 4x^2 + 4y^2,
$$

note that

$$
\begin{aligned}
&|Y_s^1 - Y_s^2||f(s,X_s^1,Y_s^1,Z_s^1) - f(s,X_s^2,Y_s^2,Z_s^2)| \\
&\le \frac{3(|Y_s^1 - Y_s^2|^2 + |f(s,X_s^1,Y_s^1,Z_s^1) - f(s,X_s^2,Y_s^2,Z_s^2)|^2)}{2}.
\end{aligned}
\tag{2.13}
$$

From this and (2.12), we have that

$$
\begin{aligned}
&\mathbb{E}\big[|Y_t^1 - Y_t^2|^2\big] \\
&\le 3\int_t^T \mathbb{E}\Bigg[(|Y_s^1 - Y_s^2|^2 + |f(s,X_s^1,Y_s^1,Z_s^1) - f(s,X_s^2,Y_s^2,Z_s^2)|^2)\Bigg]ds \\
&\quad + \mathbb{E}\big[|g(X_T^1) - g(X_T^2)|^2\big].
\end{aligned}
\tag{2.14}
$$

Put

$$
h(t) = 3\int_t^T \mathbb{E}\big[|f(s,X_s^1,Y_s^1,Z_s^1) - f(s,X_s^2,Y_s^2,Z_s^2)|^2\big]ds + \mathbb{E}\big[|g(X_T^1) - g(X_T^2)|^2\big].
$$

We now have

$$
\mathbb{E}\big[|Y_t^1 - Y_t^2|^2\big] \le h(t) + 3\int_t^T \mathbb{E}\big[|Y_s^1 - Y_s^2|^2\big]ds. \tag{2.15}
$$

By Gronwall's lemma (see Theorem B.0.2, with $\beta = 3$ and $\nu(t) = \mathbb{E}\big[|Y_t^1 - Y_t^2|^2\big]$, we obtain that

$$
\begin{aligned}
\mathbb{E}\big[|Y_t^1 - Y_t^2|^2\big] &\le h(t) + 3\int_t^T h(s)e^{3(T-t)}ds \\
&\le h(t) + h(t) * T * e^{3T}.
\end{aligned}
\tag{2.16}
$$

Observe that $h(t)$ is continuous in the variation of x since both f and g are continuous when considered functions of x. Therefore, by the continuity of f, g in x and the continuity of $X_t^{t,x}$ in (t,x) we have obtained the mean-square continuity of

$(t, x) \to v(t, x) = Y_t^{t,x}$ and hence the continuity of $v(t, x)$.

Moreover, $v(T, x) = Y_T = g(X_T) = g(x)$. Therefore, condition (ii) of the definition of both the sub-solution and the super-solution are satisfied. What remains to be proven is condition (i). Here, we provide the proof for the sub-solution property; the same idea can be used to prove the super-solution property.

For the sake of contradiction, assume that $\exists\ (t, x) \in [0, T] \times \mathbb{R}$ with the property that $v - w$ has a maximum at (t, x) such that $v(t, x) = w(t, x)$, but

$$D_t w(t, x) + \mathcal{G}(w)(t, x) < 0.$$

By the continuity properties of w, its derivatives, and f, the generator, $\exists\, \epsilon > 0, h > 0$ such that for (s, y) satisfying $t \le s \le t+h$, $|x-y| \le \epsilon$, (that is, if (s, y) is a sufficiently small deviation from (t, x)),

$$D_t w(s, y) + \mathcal{G}(w)(s, y) < 0, \text{ and } v(s, y) < w(s, y). \tag{2.17}$$

Define $\tau := \inf\{s \ge t : |X_s - x| \ge \epsilon\} \wedge (t + h)$. It is straightforward to confirm that this is a finite stopping time. Let $Y_s^1 := Y_{s\wedge\tau}$ and $Z_s^1 := \mathbb{I}_{[0,\tau]}(s) Z_s$ for $t \le s \le t + h$. By observation, we see that

$$\begin{aligned} dY_s^1 &= -\mathbb{I}_{[0,\tau]}(s)\left[f(s, X_s, v(s, X_s), Z_s)ds - Z_s dW_s\right], \\ &= -\mathbb{I}_{[0,\tau]}(s) f(s, X_s, v(s, X_s), Z_s^1)ds + Z_s^1 dW_s,\ t \le s \le t + h, \\ Y_{t+h}^1 &= v(\tau, X_\tau) \text{ (since } \tau \le t + h). \end{aligned} \tag{2.18}$$

Now define $Y_s^2 := w(s \wedge \tau, X_{s\wedge\tau})$ and $Z_s^2 := \mathbb{I}_{[0,\tau]}(s) D_x w(s, X_s)\sigma(X_s)$ for $t \le s \le t + h$. By applying Itô's formula to Y^2, we obtain the BSDE

$$\begin{aligned} dY_s^2 &= \mathbb{I}_{[0,\tau]}(s) D_t w(s, X_s)ds + D_x w(s, X_s)\Big(\mathbb{I}_{[0,\tau]}(s)[\mu(X_s)ds + \sigma(X_s)dW_s]\Big) + \\ &\quad \frac{1}{2} D_{xx} w(s, X_s)\mathbb{I}_{[0,\tau]}(s)\sigma^2(X_s)ds \\ &= \mathbb{I}_{[0,\tau]}(s)\Big(D_t w(s\, X_s) + \frac{1}{2} D_{xx} w(s, X_s)\sigma^2(X_s) + D_x w(s, X_s)\mu(X_s)\Big)ds \\ &\quad + Z_s^2 dW_s, \end{aligned} \tag{2.19}$$

for $t \le s \le t + h$, with $Y_{t+h}^2 = w(\tau, X_\tau)$.

Put $\xi^1 = Y_T^1$ and $\xi_2 = Y_T^2$. Observe that $v(\tau, X_\tau) < w(\tau, X_\tau)$; that is, $\xi_1 < \xi_2$ and so $\mathrm{P}(\xi_1 < \xi_2) > 0$.

Therefore, by the strict version of the comparison theorem (see theorem B.0.1), we have that $Y_t^1 < Y_t^2$; that is $v(t, x) < w(t, x)$, a contradiction. □

2.4 Discretization Schemes for the FSDE

The objective of this dissertation is to implement and benchmark the performance of the pricing technique as given in Ruijter, Marjon J and Oosterlee, Cornelis W (2016) when the FSDE process is discretized using a second order Taylor scheme. In order to benchmark this scheme, the same procedure is implemented under the Milstein or Euler discretization. The purpose of this section is to present the form of the discretization schemes.

2.4.1 Multiple Itô Integrals

For a given FSDE of the form (2.3), we can define the following diffusion operators $L^i : C^{1,2}([0,T] \times \mathbb{R}) \to \mathbb{R}$, for $i = \{0, 1\}$, by

$$L^0\psi(t,x) := D_t\psi(t,x)+\mu(x)D_x\psi(t,x)+\frac{1}{2}\sigma^2(x)D_{xx}\psi(t,x),\ \ L^1\psi(t,x) := \sigma(x)D_x\psi(t,x), \tag{2.20}$$

with $\psi\ inC^{1,2}([0,T] \times \mathbb{R})$. Also, we define the Itô integral operators:

$$I_{(0)}^{(\beta,\rho)} := \int_\beta^\rho ds,\ \ I_{(1)}^{(\beta,\rho)} := \int_\beta^\rho dW_s, \tag{2.21}$$

where $\beta \leq \rho \leq T$ are stopping times and $\{W_s\}_{\beta\leq s\leq\rho}$ is a standard Brownian motion. We can now define a multiple Itô integral operator for a function $\phi : [0,T] \times \mathbb{R} \to \mathbb{R}$:

$$I_\alpha^{(\beta,\rho)}[\phi(\cdot,X)] = \begin{cases} \phi(\rho, X_\rho), & \text{if } l(\alpha) = 0 \\ \int_\beta^\rho I_{\alpha-}^{(\beta,s)}[\phi(\cdot,X)]ds, & \text{if } l(\alpha) \geq 1 \text{ and } j_l = 0 \\ \int_\beta^\rho I_{\alpha-}^{(\beta,s)}[\phi(\cdot,X)]dW_s, & \text{if } l(\alpha) \geq 1 \text{ and } j_l = 1, \end{cases} \tag{2.22}$$

where $X = \{X_s\}_{s\in[0,T]}$, $\alpha = (j_1, ..., j_l)$ is a multi-index vector with $j_i \in \{0,1\}$ for $1 \leq i \leq l$. The vectors $-\alpha$ and $\alpha-$ are defined as $-\alpha := (j_2, ..., j_l)$ and $\alpha- := (j_1, ..., j_{l-1})$ respectively. The length of α is denoted by $l(\alpha)$; $l(\alpha)$ is also representative of the number of iterative integrals that are computed by the integral operator in (2.24).

Denote by $\mathcal{M}$ the set of all multi-indices α. A subset $\mathcal{A} \subset \mathcal{M}$ is called a *hierarchical* set if and only if it is non-empty, $\sup_{\alpha\in\mathcal{A}} l(\alpha) < \infty$ and $-\alpha \in \mathcal{A}$ for all $\alpha \in \mathcal{A}$, where $l(\alpha) \geq 1$. Finally, according to Theorem 5.5.1 in Kloeden, Peter E. and Platen, Eckhard (1992), the Itô expansion of a sufficiently differentiable function $f : [0,T] \times \mathbb{R} \to \mathbb{R}$ when computed from $t_0 \leq t$ is given by

$$f(t, X_t) = \sum_{\alpha\in\mathcal{A}} I_\alpha^{t_0,t}[f_\alpha(t_0, X_{t_0})] + \sum_{\alpha\in\mathcal{B}(\mathcal{A})} I_\alpha^{t_0,t}[f_\alpha(\cdot, X)], \tag{2.23}$$

where $\mathcal{B}(\mathcal{A}) := \{\alpha \in \mathcal{M}\backslash\mathcal{A} : -\alpha \in \mathcal{A}\}$, $\mathcal{A}$ is a hierarchical set and

$$f_\alpha = \begin{cases} f, & \text{if } l(\alpha) = 0 \\ f_{-\alpha}, & \text{if } l(\alpha) \geq 1. \end{cases} \tag{2.24}$$

For a more comprehensive overview of multiple Itô integrals, see Kloeden, Peter E. and Platen, Eckhard (1992).

2.4.2 Derivation of the Order 2.0 Simplified Weak Taylor Scheme

A description of a second order Taylor scheme is given in Ruijter, Marjon J and Oosterlee, Cornelis W (2016), termed 'Order 2.0 continuous simplified weak Taylor scheme', henceforth '2.0-continuous-weak-Taylor scheme'. It is a slight variant of the well-known order 2.0-weak-Taylor scheme. We summarize the theoretical development of the 2.0-continuous-weak-Taylor scheme scheme in what follows.

Let $t_0 < t_1 < \ldots < t_m < \ldots < t_M$ denote the discrete time grid. For convenience, denote $X_m = X_{t_m}$ and $W_m = W_{t_m}$. Also, let $\Delta t = t_{m+1} - t_m$ for $m \in \{0, 1, ..., M-1\}$ (since the time steps are assumed to be equidistant) and put $\Delta W_{m+1} = W_{m+1} - W_m$. The approximated process obtained by a numerical method is denoted as X_m^Δ at t_m, with $X_0^\Delta = X_0$.

The 2.0-weak-Taylor scheme, as a discretization technique of (2.3), has the form

$$\begin{aligned} X_{m+1}^\Delta &= X_m^\Delta + \mu(X_m^\Delta) I_{(0)}^{m,m+1} + \sigma(X_m^\Delta) I_{(1)}^{m,m+1} + L^1\sigma(X_m^\Delta) I_{(1,1)}^{m,m+1} \\ &\quad + L^1\mu(X_m^\Delta) I_{(1,0)}^{m,m+1} + L^0\sigma(X_m^\Delta) I_{(0,1)}^{m,m+1} + L^0\mu(X_m^\Delta) I_{(0,0)}^{m,m+1} \\ &= X_m^\Delta + \mu(X_m^\Delta)\Delta t + \sigma(X_m^\Delta)\Delta W_{m+1} + \frac{1}{2}\sigma(X_m^\Delta)\sigma_x(X_m^\Delta)[(\Delta W_{m+1})^2 - \Delta t] \\ &\quad + \mu_x(X_m^\Delta)\sigma(X_m^\Delta)\Delta\overline{W}_{m+1} + \frac{1}{2}(\Delta t)^2\left(\mu(X_m^\Delta)\mu_x(X_m^\Delta) + \frac{1}{2}\mu_{xx}(X_m^\Delta)\sigma^2(X_m^\Delta)\right) \\ &\quad + \left(\mu(X_m^\Delta)\sigma_x(X_m^\Delta) + \frac{1}{2}\sigma_{xx}(X_m^\Delta)\sigma^2(X_m^\Delta)\right)\left[\Delta W_{m+1}\Delta t - \Delta\overline{W}_{m+1}\right], \end{aligned} \tag{2.25}$$

where

$$\Delta\overline{W}_{m+1} := I_{(1,0),m+1} \sim N\left(0, \frac{(\Delta t)^3}{3}\right), \text{ (by (2.21) and (2.24))}, \tag{2.26}$$

and

$$\text{cov}(\Delta\overline{W}_{m+1}, \Delta W_{m+1}) = \int_{t_m}^{t_{m+1}} (s - t_m)ds = \frac{1}{2}(\Delta t)^2. \tag{2.27}$$

The 2.0-continuous-weak-Taylor scheme is obtained by replacing $\Delta\overline{W}_{m+1}$ in (2.25) by $\Delta\widetilde{W}_{m+1} := \frac{1}{2}\Delta W_{m+1}\Delta t$.

Remark 2.4.1. Using the analysis in Kloeden, Peter E. and Platen, Eckhard (1992), it can be shown that the 2.0-continuous-weak-Taylor scheme is weakly convergent to the FSDE it approximates with order $\delta = 2$ (see Appendix A.0.2 for the definition of weak convergence).

To summarize, we note that the 2.0-continuous-weak-Taylor method can be abbreviated by first writing

$$m(x) = \mu(x) - \frac{1}{2}\sigma(x)\sigma_x(x) + \frac{1}{2}\left(\mu(x)\mu_x(x) + \frac{1}{2}\mu_{xx}\sigma^2(x)\right)\Delta t, \quad \kappa(x) = \frac{1}{2}\sigma(x)\sigma_x(x),$$

and

$$s(x) = \sigma(x) + \frac{1}{2}\left(\mu_x(x)\sigma(x) + \mu(x)\sigma_x(x) + \frac{1}{2}\sigma_{xx}(x)\sigma^2(x)\right)\Delta t.$$

Then the 2.0-continuous-weak-Taylor scheme has the form

$$X_{m+1}^{\Delta,m,x} = x + m(x)\Delta t + s(x)\Delta W_{m+1} + \kappa(x)(\Delta W_{m+1})^2, \tag{2.28}$$

where the superscript in $X_{m+1}^{\Delta,m,x}$ denotes the progression of the state process conditional on $X_m = x$. On the other hand, the Milstein scheme is specified by setting

$$m(x) = \mu(x) - \frac{1}{2}\sigma(x)\sigma_x(x), \; \kappa(x) = \frac{1}{2}\sigma(x)\sigma_x(x) \text{ and } s(x) = \sigma(x),$$

and the Euler scheme is specified by

$$m(x) = \mu(x), \; \kappa(x) = 0 \text{ and } s(x) = \sigma(x).$$

In order to benchmark the 2.0-continuous-weak-Taylor scheme (against the Euler and Milstein schemes), we price instruments under each of the discretizations. We then compare the resulting errors in terms of relative absoluteness and weak convergence.

Chapter 3

BCOS Method

While the method for discretizing FSDEs was provided in Section 2.4, the aim of this section is to present the method given in Ruijter, Marjon J and Oosterlee, Cornelis W (2016) for numerically solving the BSDEs that we encounter later. One of the key features of the so-called 'BCOS' method is its reliance on Fourier techniques for solving BSDEs.

3.1 Fourier Theory and Characteristic Functions in Option Pricing

One of the early uses of characteristic functions for option pricing can be found in Heston, Steven L (1993) wherein he derived, in semi-closed form, a technique to price European call options in the case where the underlying asset is equipped with a stochastic volatility. In particular, the in-the-money probabilities similar to those that appear in a standard Black-Scholes-Merton model were computed using the characteristic function of $\log S$, where S is the corresponding underlying asset. A slight extension of this approach, using instead the characteristic function of $\log \frac{S}{K}$ in Bates, David S (1996), was implemented in the context of pricing American options in a stochastic volatility and jump-diffusion setting.

The above-mentioned approaches to option pricing, together with the works Chen, Ren-Raw and Scott, Louis (1992) and Scott, Louis O (1997) for example, critically relied on Fourier theory. However, as later mentioned in Carr, Peter and Madan, Dilip (1999), these applications of Fourier theory were relatively inefficient and could be significantly improved by making use of the Fast Fourier Transform (FFT) algorithm (along with the characteristic function of $\log S$). In doing so, the cost of computing the numerical integrals required for the option prices is reduced to $O(N \log_2 N)$ from $O(N^2)$, where N is the number of strikes being simultaneously considered. This computational cost saving is immensely beneficial, among other things, for efficient day-to-day parameter calibration.

Carr and Madan's method, based on the Fourier transform, was later followed by the influential work of Fang, Fang and Oosterlee, Cornelis W (2008) that was instead based on Fourier cosine expansions. In particular, the Fourier cosine expansions were for the density of $\log \frac{S}{K}$. This so-called 'COS method', an alternative to Carr and Madan's and other methods that relied on the FFT algorithm, was notably more efficient at pricing vanilla and some exotic options Fang, Fang and Oosterlee, Cornelis W (2008). Similar to the foregoing methods mentioned thus far, the COS method relies on the existence of the characteristic function of $\log \frac{S}{K}$ in an appropriate probability space. According to Fang, Fang and Oosterlee, Cornelis W (2008), when compared to other efficient techniques such as the Fast Gauss Transform or the Double Exponential Transformation, the COS method had the added advantage that it could accommodate more general dynamics of the underlying process and allow for efficient recovery of the density (of the log-stock process) from the characteristic function.

The common feature in all of the above-mentioned methods is the assumption of the existence of the characteristic function for either one of $\log S$ or $\log \frac{S}{K}$, thus placing some restriction on their respective applicability with respect to problem formulations in which the appropriate characteristic function does not exist.

3.2 BCOS Method: Version I

The 'BSDE-COS method' or BCOS method, which can be viewed as an extension of the COS method, was developed by Ruijter, Marjon J and Oosterlee, Cornelis W (2015) as an efficient technique for pricing options by numerically solving an FBSDE system; the FSDE would characterize the value of the underlying, and the BSDE pair would respectively characterize the value of the option and the hedging strategy. Recall that by the adaptedness of this pair, the future value one time step ahead in a time discretization cannot be anticipated, even though the numerical solution needs to be arrived at backwards in time (in order to ensure the satisfaction of the terminal condition). In this setting, conditional expectations arise naturally, and how these are treated is a point of divergence in the literature for BSDEs.

Zhao, Weidong and Chen, Lifeng and Peng, Shige (2006) proposed a method to approximate the conditional expectations on a set of grid points directly using Monte Carlo, and further augmented this by applying a local space interpolating technique for the non-grid points. This method, however, is not to be confused with the well-known least-squares Monte Carlo techniques, such as the Longstaff and Schwartz method, that have found some application to BSDEs through the work of Bender, Christian and Steiner, Jessica (2012), for example. For more examples of approaches to BSDEs, we refer the reader to Gobet, Emmanuel and Lemor, Jean-Philippe and Warin, Xavier and others (2005), Lemor, Jean-Philippe and Gobet, Emmanuel and Warin, Xavier and others (2006) and Hyndman, Cody B and Ngou, Polynice Oyono (2017).

The BCOS method of Ruijter, Marjon J and Oosterlee, Cornelis W (2015), or 'Version 1' as we shall henceforth refer to it, is at its core a Fourier technique for computing the conditional expectations that arise. For its application, it is necessary that the *characteristic function of the transition density function* of the solution to the FSDE exists and is known; this is not always the case in practice, as exemplified in the case where the FSDE process is a constant elasticity of variance (CEV) process with an elasticity parameter value of 3. In Ruijter, Marjon J and Oosterlee, Cornelis W (2016), the authors generalized 'Version 1'; in particular, they replaced the characteristic function of the transition density function with the characteristic function of the discrete approximation of the FSDE process. We shall refer to this extended method as 'Version 2'. In what follows, we provide an exposition of Version 2 as given in Ruijter, Marjon J and Oosterlee, Cornelis W (2016).

3.3 BCOS Method: Version 2

3.3.1 Discretization of the BSDE

The FSDE discretization procedure has been given in Section 2.4.2. In this section, the aim is to relay the discretization of the BSDE in (2.4) as is done in Ruijter, Marjon J and Oosterlee, Cornelis W (2015). For convenience, put $\Lambda_s = (X_s, Y_s, Z_s)$. We also reserve the time discretization given in Section 2.4.2, as well as the convenient abbreviation of the time indices.

The exact integral of the BSDE over $[t_m, t_{m+1}]$ is given by

$$Y_m = Y_{m+1} + \int_{t_m}^{t_{m+1}} f(s, \Lambda_s) ds - \int_{t_m}^{t_{m+1}} Z_s dW_s. \tag{3.1}$$

If we take expectations on both sides of (3.1) conditional on the information at t_m, apply the θ-method (see Iserles, Arieh (2009)) to approximate the integral, and apply Fubini's Theorem to interchange the expectation and the integral, then we obtain

$$
\begin{aligned}
Y_m &= \mathbb{E}_m[Y_{m+1}] + \int_{t_m}^{t_{m+1}} E_m[f(s,\Lambda_s)]ds \\
&\approx \mathbb{E}_m[Y_{m+1}] + \Delta t\theta f(t_m,\Lambda_m) + \Delta t(1-\theta)E_m[f(t_{m+1},\Lambda_{m+1})],
\end{aligned} \tag{3.2}
$$

where $\mathbb{E}_m[\cdot]$ is shorthand for $\mathbb{E}[\cdot|\mathcal{F}_{t_m}]$. Now multiply each side in (3.1) by ΔW_{m+1}, apply the operator $\mathbb{E}_m[\cdot]$ on both sides and apply the θ-method and Fubini's Theorem to obtain

$$
\begin{aligned}
0 &= \mathbb{E}_m[Y_{m+1}\Delta W_{m+1}] + \mathbb{E}_m\left[\int_{t_m}^{t_{m+1}} f(s,\Lambda_s)(W_{m+1}-W_s+W_s-W_m)ds\right] \\
&\quad - \mathbb{E}_m\left[\int_{t_m}^{t_{m+1}} Z_s dW_s \int_{t_m}^{t_{m+1}} dW_s\right] \\
&= \mathbb{E}_m[Y_{m+1}\Delta W_{m+1}] + \int_{t_m}^{t_{m+1}} \mathbb{E}_m[f(s,\Lambda_s)(W_{m+1}-W_s)]ds \\
&\quad + \int_{t_m}^{t_{m+1}} \mathbb{E}_m[f(s,\Lambda_s)(W_s-W_m)]ds - \operatorname{cov}\left(\int_{t_m}^{t_{m+1}} Z_s dW_s, \int_{t_m}^{t_{m+1}} dW_s\right) \\
&\approx \mathbb{E}_m[Y_{m+1}\Delta W_{m+1}] + \theta\Delta t f(t_m,\Lambda_m)\mathbb{E}_m[W_{m+1}-W_m] \\
&\quad + (1-\theta)\Delta t\mathbb{E}_m[f(t_{m+1},\Lambda_{m+1})(W_{m+1}-W_m)] - \int_{t_m}^{t_{m+1}} \mathbb{E}_m[Z_s]ds \\
&\approx \mathbb{E}_m[Y_{m+1}\Delta W_{m+1}] + \Delta t(1-\theta)\mathbb{E}_m[f(t_{m+1},\Lambda_{m+1})\Delta W_{m+1}] - \Delta t\theta Z_m \\
&\quad - \Delta t(1-\theta)\mathbb{E}_m[Z_{m+1}].
\end{aligned} \tag{3.3}
$$

We are now ready to provide the discretization for the BSDE. Theorem 2.3.1 suggests the following terminal conditions:

$$
Y_M^{\Delta} = g(X_M^{\Delta}), \quad Z_M^{\Delta} = \sigma(X_M^{\Delta})D_x g(X_M^{\Delta}). \tag{3.4}
$$

The equations (3.2) and (3.3) together suggest the following scheme for $m = M-1, M-2, ..., 0$, by isolating Z_m and Y_m respectively:

$$
\begin{aligned}
Z_m^{\Delta} &= \frac{1}{\Delta t}\theta^{-1}\mathbb{E}_m[Y_{m+1}^{\Delta}\Delta W_{m+1}] + \theta^{-1}(1-\theta)\mathbb{E}_m[f(t_{m+1},\Lambda_{m+1}^{\Delta})\Delta W_{m+1}] \\
&\quad - \theta^{-1}(1-\theta)\mathbb{E}_m[Z_{m+1}^{\Delta}], \\
Y_m^{\Delta} &= \mathbb{E}_m[Y_{m+1}^{\Delta}] + \Delta t\theta f(t_m,\Lambda_m^{\Delta}) + \Delta t(1-\theta)\mathbb{E}_m[f(t_{m+1},\Lambda_{m+1}^{\Delta})].
\end{aligned} \tag{3.5}
$$

Remark 3.3.1. For $\theta > 0$, the above scheme is implicit in Y_m^{Δ}. Picard iterations are used to compute Y_m^{Δ}, with an initial guess of $\mathbb{E}_m[Y_{m+1}^{\Delta}]$. In particular, the p^{th} Picard iteration is defined as

$$
Y_m^{\Delta,p} = \mathbb{E}_m[Y_{m+1}^{\Delta}] + \Delta t\theta f(t_m, X_m^{\Delta}, Y_m^{\Delta,p-1}, Z_m^{\Delta}) + \Delta t(1-\theta)\mathbb{E}_m[f(t_{m+1},\Lambda_{m+1}^{\Delta})],
$$

where $Y_m^{\Delta,p}$ is the value of Y_m^{Δ} after p iterations. The iterations are terminated once consecutive approximations are within 10^{-12} of each other (see Ruijter, Marjon J and Oosterlee, Cornelis W (2016)). In most cases, this takes between five to six iterations and is one of the more expensive jobs in implementing Version 2.

As mentioned in both Ruijter, Marjon J and Oosterlee, Cornelis W (2015) and Ruijter, Marjon J and Oosterlee, Cornelis W (2016), there exist deterministic functions $y_m(x)$ and $z_m(x)$ such that $Y_m^{\Delta} = y_m(X_m^{\Delta})$ and $Z_m^{\Delta} = z_m(X_m^{\Delta})$ for $m \in \{0, 1, ..., M\}$. This is trivially true for $m = M$ by (3.4). By induction, we can assume the existence of such functions for $m + 1 < M$, so that for $m < M$, (3.5) can be written as

$$
\begin{aligned}
z_m(X_m^{\Delta}) &= \frac{1}{\Delta t}\theta^{-1}\mathbb{E}_m^x[y_m(X_{m+1}^{\Delta})\Delta W_{m+1}] - \theta^{-1}(1-\theta)\mathbb{E}_m^x[z_{m+1}(X_{m+1}^{\Delta})] \\
&\quad + \theta^{-1}(1-\theta)\mathbb{E}_m^x[f(t_{m+1}, \Lambda_{m+1}^{\Delta}(X_{m+1}^{\Delta}))\Delta W_{m+1}], \\
y_m(X_m^{\Delta}) &= \mathbb{E}_m^x[y_{m+1}(X_{m+1}^{\Delta})] + \Delta t(1-\theta)\mathbb{E}_m^x[f(t_{m+1}, \Lambda_{m+1}^{\Delta}(X_{m+1}^{\Delta}))] \\
&\quad + \Delta t\theta f(t_m, \Lambda_m^{\Delta}),
\end{aligned}
\tag{3.6}
$$

where $z_m(X_m^{\Delta}) = z_m(x)$, $y_m(X_m^{\Delta}) = y_m(x)$, where the superscript in $\mathbb{E}_m^x[\cdot]$ expresses the fact that we condition on $X_m^{\Delta} = x$, and X_{m+1}^{Δ} is X_{m+1}^{Δ} conditional on $X_m^{\Delta} = x$. The notion of solving the BSDE backwards in time by computing, at each time point t_m, the functions $y_m(\cdot)$ and $z_m(\cdot)$ plays an important role in the numerical implementation. We will revisit this point in Section 3.3.4.

Now, it is obvious that the deterministic function for $m < M$ is given by the right hand side of (3.6): a function of $x = X_m^{\Delta}$. Furthermore, in order to determine fully $z_m(X_m^{\Delta})$ and $y_m(X_m^{\Delta})$, it is necessary to compute the conditional expectations in (3.6). Collectively, these have the form

$$
\mathbb{E}_m^x[h(t_{m+1}, X_{m+1}^{\Delta})] \text{ and } \mathbb{E}_m^x[h(t_{m+1}, X_{m+1}^{\Delta})\Delta W_{m+1}], \tag{3.7}
$$

where h is a general function. This is the subject of the next subsection.

3.3.2 A Characteristic Function Result

To compute either one of the conditional expectations in (3.7), Ruijter, Marjon J and Oosterlee, Cornelis W (2015) take an approach that relies on the availability of the characteristic function of the transition density of $X = \{X_t\}_{t \in [0,T]}$. In practice, at t_m, this means that they compute $\varphi_{X_{m+1}}(\cdot|F_m) = \varphi_{X_{m+1}}(\cdot|X_m = X_m^{\Delta})$, where $\varphi_{X_{m+1}}(\cdot|X_m)$ is the conditional characteristic function of X_{m+1} given the available information at time t_m. In Ruijter, Marjon J and Oosterlee, Cornelis W (2016), this approach is generalized by using the conditional characteristic function of the discrete form of the FSDE in the place of $\varphi_{X_{m+1}}(\cdot|X_m)$. That is, $\varphi_{X_{m+1}}(\cdot|X_m)$ is replaced by the characteristic function $\varphi^{\Delta}_{X_{m+1}}(\cdot|X^{\Delta}_{t_m})$ of $X^{\Delta}_{t_{m+1}}$ given the information $\mathbb{F}_{t_m}$ at t_m. The main reason is that while $\varphi_{X_{m+1}}(\cdot|X_m)$ may not always be available in closed form, $\varphi^{\Delta}_{X_{m+1}}(\cdot|X^{\Delta}_{t_m})$ can always be computed if either one of the Euler, Milstein or 2.0-continuous-weak-Taylor schemes are used to discretize the FSDE process.

We now state the characteristic function result as a theorem.

Theorem 3.3.1. *The characteristic function of X^{Δ}_{m+1} in (2.28), is given by*

$$\begin{aligned}\varphi_{X^{\Delta}_{m+1}}(u|X^{\Delta}_m = x) &= \mathbb{E}\left[e^{iuX^{\Delta}_{m+1}}\middle|X^{\Delta}_m = x\right]\\ &= \exp\left(iux + ium(x)\Delta t - \frac{\frac{1}{2}u^2 s^2(x)\Delta t}{1-2iu\kappa(x)\Delta t}\right)(1-2iu\kappa(x)\Delta t)^{-\frac{1}{2}}.\end{aligned} \tag{3.8}$$

If $\kappa(x) = 0$, then

$$\varphi_{X^{\Delta}_{m+1}}(u|X^{\Delta}_m = x) = \exp\left(iux + ium(x)\Delta t - \frac{1}{2}u^2 s^2(x)\Delta t\right). \tag{3.9}$$

Proof. If $\kappa(x) = 0$, then (2.28) simplifies so that $X^{\Delta}_{m+1} \sim N(x + m(x)\Delta t, s^2(x)\Delta t)$, in which case (3.9) follows easily.

In order to prove (3.8), first observe that an equivalent representation of (2.28) is

$$\begin{aligned}X^{\Delta}_{m+1} &= x + m(x)\Delta t + \kappa(x)\left(\Delta W_{m+1} + \frac{s(x)}{2\kappa(x)}\right)^2 - \frac{s^2(x)}{4\kappa^2(x)}\\ &= x + m(x)\Delta t - \frac{s^2(x)}{4\kappa^2(x)} + \kappa(x)\Delta t\left(\varepsilon_{m+1} + \sqrt{\lambda(x)}\right)^2,\end{aligned} \tag{3.10}$$

where $\varepsilon_{m+1} \sim N(0,1)$ and $\lambda(x) = \frac{s^2(x)}{4\Delta t\kappa^2(x)}$. The random variable $\left(\varepsilon_{m+1} + \sqrt{\lambda(x)}\right)^2$ has a non-central chi-squared distribution, with $\nu = 1$ degrees of freedom and a non-centrality parameter $\lambda(x)$. The corresponding characteristic function

$$\varphi^{\chi,\lambda}(u) = \exp\left(\frac{i\lambda u}{1-2iu}\right)(1-2iu)^{-\nu/2}$$

is well-known. Finally, it follows from (3.10) that

$$\begin{aligned}\varphi^{\Delta}_{X_{m+1}}(u|X^{\Delta}_m = x) &= \exp\left(iux + ium(x)\Delta t - \frac{s^2(x)}{4\kappa^2(x)}\right)\varphi^{\chi,\lambda(x)}(u\kappa(x)\Delta t)\\ &= \exp\left(iux + ium(x)\Delta t - \frac{\frac{1}{2}u^2 s^2(x)\Delta t}{1-2iu\kappa(x)\Delta t}\right)(1-2iu\kappa(x)\Delta t)^{-\frac{1}{2}}.\end{aligned} \tag{3.11}$$

□

The characteristic function $\varphi^{\Delta}_{X_{m+1}}(\cdot|X^{\Delta}_{m})$ replaces the canonical conditional characteristic function. This is beneficial in settings where the latter does not exist. For instance, if the FSDE is a CEV process with elasticity parameter $\gamma \notin \{0, 0.5, 1, 1.5\}$, there is no closed form solution for the conditional characteristic function (Ruijter, Marjon J and Oosterlee, Cornelis W (2016)). Practically, one would not be able to compute $\varphi_{X_{m+1}}(\cdot|X_m)$, which is needed for the actual computation of the conditional expectations.

Remark 3.3.2. At this point, it is essential to note that the discretization schemes are not in fact used to simulate sample paths as one would initially expect. Instead, the schemes are used only in conjunction with Theorem 3.3.1 on a static grid. The use of a static grid (for the stock prices) in the place of simulation is one of the main reasons for the high efficiency of Version 2.

3.3.3 Computing the Conditional Expectations

In this section, explicit formulae for the computation of the conditional expectations are given. By extension, with the conditional expectations at hand, it should not be difficult to solve (3.6).

For $\mathbb{E}^x_m[h(t_{m+1}, X^{\Delta}_{m+1})]$, the computation is based on the so-called 'COS formula' which was first suggested in Ruijter, Marjon J and Oosterlee, Cornelis W (2015). For details on the COS formula, see Appendix B. In particular, inspired by the COS formula,

$$\mathbb{E}^x_m[h(t_{m+1}, X^{\Delta}_{m+1})] = {\sum_{k=0}^{N-1}}' \mathcal{H}_k(t_{m+1}) \mathrm{Re}\left[\varphi_{X^{\Delta}_{t_{m+1}}}\left(\frac{k\pi}{b-a}\Big|X^{\Delta}_{t_m} = x\right) \exp\left(\frac{-ik\pi a}{b-a}\right)\right], \tag{3.12}$$

where $\{\mathcal{H}_k(t_{m+1})\}_{k=0}^{\infty}$ is the set of coefficients corresponding to the Fourier cosine expansion of $h(t_{m+1}, \cdot)$, where $\mathrm{Re}[\cdot]$ refers to the real part of the subject and where the stroke in the summation denotes the fact the first term in the summation is weighted by a half.

Remark 3.3.3. The conditional characteristic function used in (3.12) is not the same as the one used in its development; $\varphi_{X_{m+1}}(\cdot|X_m)$ was used in Ruijter, Marjon J and Oosterlee, Cornelis W (2015) to derive the COS formula, as already alluded to. This fact distinguishes Version 1 from Version 2 and generalizes the applicability of the former.

On the other hand, using integration by parts (see Ruijter, Marjon J and Oosterlee, Cornelis W (2015)), one can show that

$$\begin{aligned} &\mathbb{E}_m^x[h(t_{m+1}, X_{m+1}^\Delta)\Delta W_{m+1}] \\ &= \sum_{k=0}^{N-1}{}' \mathcal{H}_k(t_{m+1})\mathrm{Re}\left[\mathbb{E}_m^x\left[\exp\left(\frac{ik\pi X_{m+1}^\Delta}{b-a}\right)\Delta W_{m+1}\right]\exp\left(\frac{-ik\pi a}{b-a}\right)\right]. \end{aligned} \tag{3.13}$$

With a further application of integration by parts, one can also show that (see Ruijter, Marjon J and Oosterlee, Cornelis W (2016))

$$\begin{aligned} &\mathbb{E}_m^x\left[\exp\left(iuX_{m+1}^\Delta\right)\Delta W_{m+1}\right] = iu\Delta t s(x) + \mathbb{E}_m^x\left[\exp\left(iuX_{m+1}^\Delta\right)\right] \\ &+ 2(iu\Delta t)^2 s(x)\kappa(x)\mathbb{E}_m^x\left[\exp\left(iuX_{m+1}^\Delta\right)\right] \\ &+ 2(iu\Delta t)^3 s(x)\kappa(x)\mathbb{E}_m^x\left[\exp\left(iuX_{m+1}^\Delta\right)\right] \\ &+ 2(iu\Delta t)^4 s(x)\kappa(x)\mathbb{E}_m^x\left[\exp\left(iuX_{m+1}^\Delta\right)\right] + \dots \end{aligned} \tag{3.14}$$

Ruijter, Marjon J and Oosterlee, Cornelis W (2016) mention that for numerical experiments, the first two terms in (3.14) suffice due to the order in Δt. Now put $u = \frac{k\pi}{b-a}$ and notice that $\mathbb{E}_m^x\left[\exp\left(iuX_{m+1}^\Delta\right)\right] = \varphi_{X_{m+1}}^\Delta(u|X_m^\Delta)$. This leads to the following computation:

$$\begin{aligned} &\mathbb{E}_m^x[h(t_{m+1}, X_{m+1}^\Delta)\Delta W_{m+1}] \\ &= \sum_{k=0}^{N-1}{}' \mathcal{H}_k(t_{m+1})\mathrm{Re}\Bigg[\left(\frac{ik\pi}{b-a}s(x)\Delta t + 2\left(\frac{ik\pi}{b-a}\Delta t\right)^2 s(x)\kappa(x)\right) \\ &\varphi_{X_{m+1}}^\Delta\left(\frac{k\pi}{b-a}\Big|X_m^\Delta = x\right)\exp\left(\frac{-ik\pi a}{b-a}\right)\Bigg]. \end{aligned} \tag{3.15}$$

In both (3.12) and (3.15), what should be an infinite Fourier sum has been truncated since the terms in the summation converge quickly to zero. Together, (3.12) and (3.15) provide the formulae for computing the conditional expectations that arise when we numerically solve the BSDEs. What remains to be specified is the set of Fourier coefficients $\{\mathcal{H}_k(t_{m+1})\}_{k=0}^\infty$. In particular, we need the first N coefficients at every time point, where N is usually a power of 2. In our numerical experiments, $N = 2^9$.

3.3.4 Recovering the Fourier Coefficients

The definition of a Fourier coefficient involves integration over the entire real line. However, due to rapid decay to zero of the integrands in our consideration, it suffices to consider integration on a truncated range $[a, b]$. In particular, we take

$a = X_0 + \mu(X_0) - L\sqrt{\sigma^2(X_0)}$ and $b = X_0 + \mu(X_0) + L\sqrt{\sigma^2(X_0)}$ where $L = 10$.

$$\begin{aligned} \mathcal{H}_k(t_{m+1}) &:= \frac{2}{b-a}\int_a^b h(t_{m+1},x)\cos\left(k\pi\frac{x-a}{b-a}\right)dx \\ &\approx \frac{2}{b-a}\sum_{n=0}^{N-1} h(t_{m+1},x_n)\cos\left(k\pi\frac{x_n-a}{b-a}\right)\Delta x \\ &= \frac{2}{N}\sum_{n=0}^{N-1} h(t_{m+1},x_n)\cos\left(k\pi\frac{2n+1}{2N}\right), \end{aligned} \tag{3.16}$$

where the approximation follows from the mid-point integration rule, and the last equality is easily deduced from the fact that $\Delta x = \frac{b-a}{N}$, and $x_n = a + (n + \frac{1}{2})\Delta x$ for $n = 0, 1, ..., N-1$. The static grid, which is defined by $\{x_n : n = 0, ..., N\}$, used to solve the BSDE numerically and provides the integration range for computing the relevant Fourier coefficients. The last expression in (3.16) can be computed efficiently using MATLAB's `dct` function.

Recall that $h(t_m, x)$ is a generic function that may represent any one of $f(t_m, \Lambda_m)$, $y_m(x)$, or $z_m(x)$. This function is needed both for the numerical solution of BSDE (3.6) as well as the computation of the coefficients $\mathcal{F}_k^\Delta(t_{m+1})$, $\mathcal{Y}_k^\Delta(t_{m+1})$, and $\mathcal{Z}_k^\Delta(t_{m+1})$. In particular, computing functions y and z in (3.6) (as opposed to merely computing values) allows one to use the same functions in order to compute the coefficients according to (3.16) easily as needed.

3.3.5 A Summary of Version 2

In this section, we give a summary of the implementation of Version 2 for the solution of the FBSDE process as an option pricing technique.

Initial step: compute the terminal coefficients using (3.4) and (3.16). To do this, we need $y_M(x_n) = g(x_n)$, $z_M(x_n) = \sigma(x)D_x g(x_n)$ and $f(t_M, x_n, y_M(x_n), z_M(x_n))$ for $n = 0, 1, ..., N$.

For $m = M-1, M-2, ..., 1$, use the computed coefficients $\mathcal{F}_k^\Delta(t_{m+1})$, $\mathcal{Y}_k^\Delta(t_{m+1})$, and $\mathcal{Z}_k^\Delta(t_{m+1})$ to compute the functions $y_m(\cdot)$ and $z_m(\cdot)$. Then evaluate $z_m(x_n)$, $y_m(x_n)$ and $f(x_n, y_m(x_n), z_m(x_n))$ for $n = 0, 1, ..., N-1$. The equations in (3.6) as well as the observation made in Remark 3.3.1 are helpful for these computations. Now, with $y_m(x_n)$ and $z_m(x_n)$ at hand, compute the coefficients $\mathcal{F}_k^\Delta(t_m)$, $\mathcal{Y}_k^\Delta(t_m)$, and $\mathcal{Z}_k^\Delta(t_m)$ which will be used in the next iteration. To compute the coefficients, refer to (3.16).

Finally, for $m = 0$, compute the terminal value $y_0(X_0^\Delta)$.

Note that computing the BSDE as a pair of functions as opposed to merely as values enables efficient computation, especially for $y_m(\cdot)$. Moreover, the scheme for discretization affects every computation of $y_m(x)$ and $z_m(x)$.

3.3.6 Benchmarking

In order to benchmark the performance of a discretization scheme (be it Euler, Milstein or 2.0-continuous-weak-Taylor), we compute the expressions $|X_0 - X_0^\Delta|$ and $|v(t_0, X_0) - y_0(X_0^\Delta)|$. It is worth noting that these error measures are not an average of some statistic as the implementation is simulation independent.

An essential feature of the benchmarking process is analysing the error behaviour as we increase the number of time steps M. From Kloeden, Peter E. and Platen, Eckhard (1992), we know that the Euler and Milstein schemes converge weakly with order 1, while the 2.0-continuous-weak-Taylor and the 2.0-weak-Taylor scheme converge weakly with order 2 (except when we set $\theta = 1$, in which case the order of weak convergence is 1). These expectations are confirmed (but not necessarily proven) in the Chapter 4.

Finally, the definition of weak convergence given in Appendix A.0.2 is defined with respect to a specific time point T. In Chapter 4, we will display weak convergence on the numerical procedure with respect to $T = 0$. Also, as alluded to in the beginning of this section, we will not verify the weak convergence conditions for *every* smooth function g of the underlying process; in part, the definition of weak convergence requires consideration of every smooth function g. Since our purpose is not to prove weak convergence but rather to implement examples in which it occurs, the only error measures we assess are $|X_0 - X_0^\Delta|$ and $|v(t_0, X_0) - y_0(X_0^\Delta)|$.

Chapter 4

Numerical Experiments

In this chapter, we present a discussion of our results from the numerical experiments conducted. As mentioned before, the goal is to compute $|v(t_0, X_0) - y_0(X_0^\Delta)|$. For the ease of visualisation, we will instead plot $\log_{10}\left(|v(t_0, X_0) - y_0(X_0^\Delta)|\right)$. The computations are performed on MATLAB (2018a) on a machine with a Intel(R) Core(TM) i7-6700 CPU @ 3.40GHz processor and 15.7GB RAM.

4.1 Source of Errors

The BCOS methods (Version 1 and Version 2) have several errors built into it by its design. In this section, we briefly identify these errors. For a more detailed discussion, we refer the interested reader to Ruijter, Marjon J and Oosterlee, Cornelis W (2015) and Ruijter, Marjon J and Oosterlee, Cornelis W (2016).

The first source of error we encounter is due to the discretization of time into increments of size Δt. In the benchmarking process, we expect this error to decrease as we take smaller values of Δt (equivalently, larger values of M). The second source of error is in the approximation of the integrals in the BSDE (see (3.2) and (3.3)) using the θ-method. Thirdly, when approximating the Fourier coefficients, the infinite integrand is truncated to $[a, b]$ and the mid-point integration rule used to approximate the integral; it may well be that using Simpson's rule (for example) leads to a significant improvement in the overall error behaviour. The fourth source of error is the truncation of the summation for computing the conditional expectations in Section 3.3.3 to N terms. Finally, the last source of error is due to the Picard iterations used to solve implicit equations, though this error is adequately controlled for by setting a tolerance of 10^{-12}.

4.2 Experiment 1: CIR Bond Price

In this experiment, taken from Ruijter, Marjon J and Oosterlee, Cornelis W (2016), we consider the problem of pricing a bond where the interest rate process follows Cox-Ingersoll-Ross (CIR) dynamics. These have the form

$$dX_s = \alpha(\beta - X_s)ds + \eta\sqrt{X_s}dW_s,\ 0 \leq s \leq T, \tag{4.1}$$

where $X_0 = 0.04$, $\alpha = 0.2$, $\beta = 0.01$, $\eta = 0.1$ and $T = 0.25$. From the PDE of the bond price,

$$\begin{aligned}\frac{\partial v}{\partial t}(t,x) + \alpha(\beta - x)D_x v(t,x) + \frac{\eta^2 x}{2} D_{xx} v(t,x) - xv(t,x) = 0\ (t,x) \in [0,T) \times \mathbb{R}_+,\\ v(T,x) = 1,\ x \in \mathbb{R}_+,\end{aligned} \tag{4.2}$$

we can deduce the form of the driver function to be $f(t,x,y,z) = -xy$ by using (2.6). This deduction also gives us the BSDE

$$dY_s = X_s Y_s ds + Z_s dW_s,\ Y_T = 1. \tag{4.3}$$

Finally, the analytical price of this bond is given by $v(t,x) = A(t,T)\exp(-xB(t,T))$ where

$$\begin{aligned}A(t,T) &= \left(\frac{2he^{\frac{1}{2}(\alpha+h)(T-t)}}{2h + (\alpha+h)(e^{h(T-t)} - 1)}\right),\\ B(t,T) &= \frac{2(e^{(T-t)h} - 1)}{2h + (\alpha+h)(e^{h(T-t)} - 1)},\\ h &= \sqrt{\alpha^2 + 2\eta^2}.\end{aligned} \tag{4.4}$$

In order to prevent the interest rate from becoming negative, it suffices to set $a = 0$ (otherwise, $a = X_0 + \mu(X_0) - L\sqrt{\sigma^2(X_0)} < 0$) for the implementation. Negative interest rates may occur in this experiment since the Feller condition is not satisfied (Ruijter, Marjon J and Oosterlee, Cornelis W (2016)). While this may be an issue with Monte Carlo methods, the BCOS method can make accommodation for this by using a restricted x_n-grid. As expected, in Figure 4.1 we observe first order weak convergence for the Euler and Milstein scheme, as well as for the 2.0-continuous-weak-Taylor scheme (when $\theta = 1$); we also observe second order weak convergence for the 2.0-continuous-weak-Taylor scheme (when $\theta = 0.5$).

We summarize the run times in Table 4.1, which are in line with those observed in Ruijter, Marjon J and Oosterlee, Cornelis W (2016) and exemplify the efficiency of Version 2.

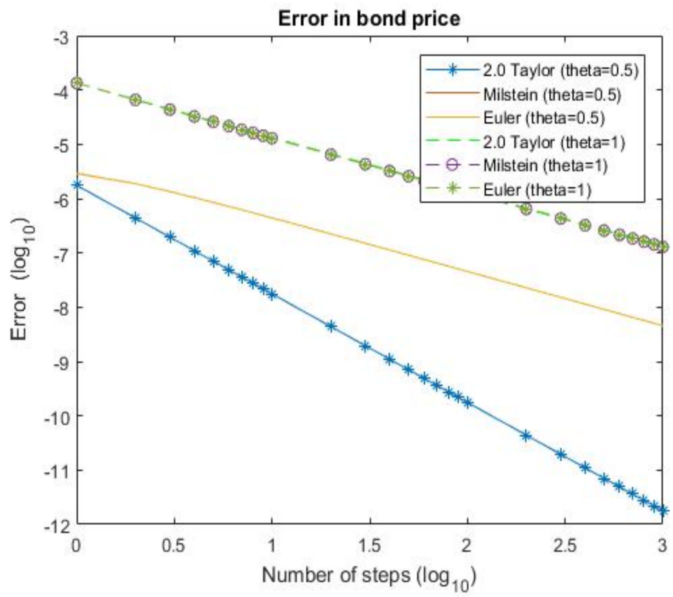

Fig. 4.1: $\log_{10}$ pricing errors from CIR bond.

Tab. 4.1: CPU times for Experiment 1 (seconds)

	$\theta = 0.5$			$\theta = 1$		
	Euler	**Milstein**	**Version 2**	**Euler**	**Milstein**	**Version 2**
$M = 5$	0.01201	0.11798	0.10894	0.12441	0.11839	0.11225
$M = 10$	0.12629	0.11342	0.11061	0.13118	0.11712	0.01132
$M = 40$	0.15376	0.14326	0.01360	0.01526	0.14488	0.13757
$M = 80$	0.18126	0.16992	0.17524	0.01872	0.18049	0.17529
$M = 100$	0.20390	0.19540	0.18740	0.21640	0.19639	0.19603
$M = 400$	0.43064	0.45875	0.44866	0.46440	0.45087	0.44984
$M = 800$	0.77827	0.74682	0.74084	0.80516	0.79886	0.77822
$M = 1000$	0.90591	0.92652	0.93871	0.95094	0.94795	0.94528

4.3 Experiment 2: Handling Complex Underlying

This example can be found in Ma, Jin and Shen, Jie and Zhao, Yanhong (2008) and Milstein, GN and Tretyakov, Michael V (2006). The drift and diffusion terms are

$$\begin{aligned} \mu(x) &= \frac{x(1+x^2)}{(2+x^2)^3}, \\ \sigma(x) &= \frac{1+x^2}{2+x^2}, \end{aligned} \tag{4.5}$$

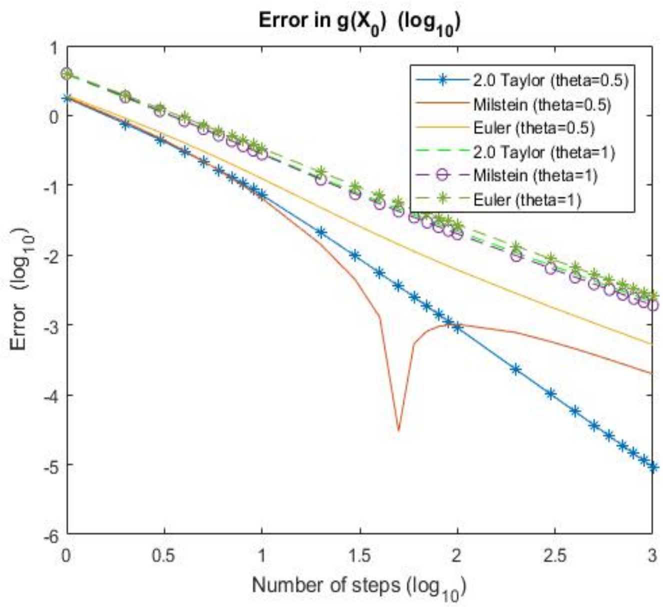

Fig. 4.2: $\log_{10}$ errors in analytical solution.

respectively. The driver function related to the BSDE is given by

$$f(t,x,y,z) = \frac{1}{t+1}\exp\left(-\frac{x^2}{t+1}\right)\left(\frac{4x^2(1+x^2)}{(2+x^2)^3} + \left(\frac{1+x^2}{2+x^2}\right)^2\left(1-\frac{2x^2}{t+1}\right) - \frac{x^2}{t+1}\right) + \frac{zx}{(2+x^2)^2}\sqrt{\frac{1+y^2+\exp\left(-\frac{2x^2}{t+1}\right)}{1+2y^2}}, \tag{4.6}$$

with the terminal condition $Y_T = g(X_T) = \exp\left(-\frac{x^2}{T+1}\right)$. The analytical solution is known to be $v(t,x) = \exp\left(-\frac{x^2}{t+1}\right)$. The results of the experiment, as seen in Figure 4.2, confirm the order 1 convergence under the Euler, Milstein and 2.0-continuous-weak-Taylor (when $\theta = 1$) discretization schemes; we also observe order 2 convergence behaviour for 2.0-continuous-weak-Taylor scheme (when $\theta = 0.5$), as outlined in Section 3.3.6. Moreover, the high level of accuracy achieved in this experiment testifies to the considerable ability of Version 2 to handle extremely complicated dynamics.

The run times for this experiment are summarized in Table 4.2.

Tab. 4.2: CPU times for Experiment 2 (seconds)

	$\theta = 0.5$			$\theta = 1$		
	Euler	**Milstein**	**Version 2**	**Euler**	**Milstein**	**Version 2**
$M = 5$	0.11718	0.10362	0.09688	0.11872	0.10663	0.10217
$M = 10$	0.11823	0.10405	0.10366	0.11597	0.11109	0.09986
$M = 40$	0.13483	0.11624	0.11822	0.13276	0.12638	0.11503
$M = 80$	0.15689	0.13649	0.15037	0.15146	0.15161	0.14106
$M = 100$	0.15967	0.15381	0.16190	0.16721	0.15322	0.15221
$M = 400$	0.30832	0.29063	0.30191	0.30644	0.29807	0.29423
$M = 800$	0.53878	0.54857	0.49927	0.53898	0.52617	0.47439
$M = 1000$	0.60873	0.59867	0.59543	0.63185	0.64820	0.57826

4.4 Experiment 3: Pricing a European Call Option under the CEV Model

Let $v(t, x)$ denote the option value at time t when $S_t = x$. The model dynamics for $S = \{S_t\}_{t\in[0,T_0]}$ are

$$dS_s = rX_s ds + \hat{\sigma} X_s^{\gamma} dW_s,$$

where $S_0 = 100$, $K = 100$, $r = 0.1$, $T = 0.1$ and $\hat{\sigma}$ is chosen so that $\sigma(S_0) = \hat{\sigma} S_0^{\gamma} = 25$. In the experiment, we consider the cases $\gamma = 0.2$ and $\gamma = 0.8$. For both these cases, as alluded to previously, Version 1 cannot be implemented since the applicable characteristic function does not exist. However, the characteristic function derived in Section 3.3.2 has been used to implement Version 2. In this we appreciate the generality of Version 2, not just for this example but also for a large variety of problems for which the underlying dynamics are not tractable.

We display the results in Figure 4.3. Again, we observe results that are mostly consistent with the foregoing theoretical development. In particular, we observe order 1 convergence in the same class of discretization schemes as before. However, under the 2.0-continuous-weak-Taylor scheme (when $\theta = 0.5$), we don't observe order 2 convergence as we increase the number of steps in the horizon T. This result differs from the findings of the Ruijter, Marjon J and Oosterlee, Cornelis W (2016), who by altering the implementation outlined in Section 3.3.5, were able to achieve order 2 convergence. Owing to the lack of clarity in their alterations, we were not able to reproduce the results in their reference paper.

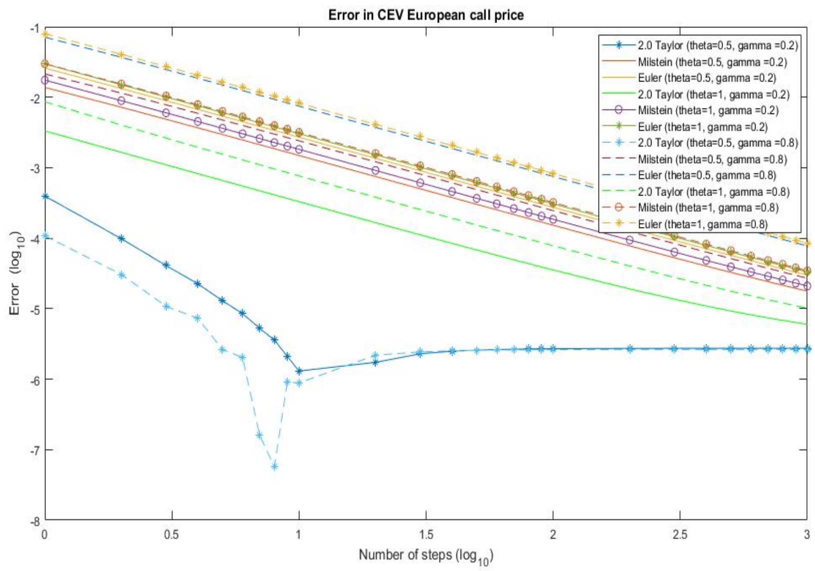

Fig. 4.3: $\log_{10}$ errors in analytical solution.

The run times for this experiment are summarized in Table 4.3, displaying the efficiency of Version 2.

Tab. 4.3: CPU times for Experiment 3 (seconds)

	$\theta = 0.5$			$\theta = 1$		
	Euler	**Milstein**	**Version 2**	**Euler**	**Milstein**	**Version 2**
$M = 5$	0.13374	0.11671	0.11003	0.13505	0.11793	0.11425
$M = 10$	0.12804	0.11921	0.10871	0.13999	0.11346	0.09986
$M = 40$	0.15016	0.13705	0.13794	0.14704	0.14941	0.14629
$M = 80$	0.18541	0.17526	0.20861	0.18905	0.17675	0.17385
$M = 100$	0.19527	0.19128	0.18630	0.21191	0.19443	0.19109
$M = 400$	0.41960	0.40207	0.40842	0.45915	0.43090	0.44178
$M = 800$	0.72152	0.70078	0.72182	0.75694	0.73435	0.72247
$M = 1000$	0.90036	0.87842	0.89276	0.93145	0.86078	0.90263

4.5 Concerns about Version 2 Applicability

The key ingredients to implementing Version 2 are the functions $\mu(x)$ and $\sigma(x)$ from the FSDE, the terminal value function $g(x,t)$ and its derivative $\frac{dg}{dx}$, as well as the driver function. By definition, we know that f maintains (2.6). This gives guidance in terms of constructing the function f, at least in the case where v is in fact known.

Though the above results display the high level of accuracy and speed of Version 2, the major stumbling block with regard to applying this method to a problem with new dynamics and (unknown) value function v is determining the form of the driver function f in (2.6), and to be able to determine f in the *absence* of the knowledge of the form of the value function v, since in any case the intention is to solve for $v(t_0, X_0)$ numerically.

Chapter 5

Conclusion

In this work, we surveyed Ruijter, Marjon J and Oosterlee, Cornelis W (2016) and in particular, provided a detailed exposition of the BCOS method for numerically solving backward stochastic differential equations in the process of valuing European claims. Where implicit equations arose, these were solved by using Picard iterations. Importantly, we highlighted how Ruijter, Marjon J and Oosterlee, Cornelis W (2016) builds from the foregoing work in Ruijter, Marjon J and Oosterlee, Cornelis W (2015) by distinguishing between Version 1 and Version 2. Finally, to a large extent we were able to reproduce the results in Ruijter, Marjon J and Oosterlee, Cornelis W (2016) and obtained order 1 and order 2 convergence that was in line with the theoretical results.

An obvious way in which the framework in Ruijter, Marjon J and Oosterlee, Cornelis W (2016) could be extended is by applying it to claims that depend on more than one stock price. Furthermore, as mentioned in Chapter 4, future research could pay attention to development of a framework to select a driver function that maintains (2.6) and leads to convergent results, given an FSDE and a contingent claim.

With this at hand, practitioners and academics will be able to apply Version 2 to problems for which the price of the claim is not known a-priori.

Bibliography

Bates, David S (1996). Jumps and Stochastic Volatility: Exchange Rate Processes Implicit in Deutsche Mark Options, *The Review of Financial Studies* **9**(1): 69–107.

Bender, Christian and Steiner, Jessica (2012). Least-Squares Monte Carlo for Backward SDEs, *Numerical Methods in Finance*, Springer, pp. 257–289.

Bismut, Jean-Michel (1973). Conjugate Convex Functions in Optimal Stochastic Control, *Journal of Mathematical Analysis and Applications* **44**(2): 384–404.

Black, Fischer and Scholes, Myron (1973). The Pricing of Options and Corporate Liabilities, *Journal of Political Economy* **81**(3): 637–654.

Carr, Peter and Madan, Dilip (1999). Option Valuation Using the Fast Fourier Transform, *Journal of Computational Finance* **2**(4): 61–73.

Chen, Ren-Raw and Scott, Louis (1992). Pricing Interest Rate Options in a Two-Factor Cox-Ingersoll-Ross Model of the Term Structure, *The Review of Financial Studies* **5**(4): 613–636.

Crandall, Michael G and Ishii, Hitoshi and Lions, Pierre-Louis (1992). User's Guide to Viscosity Solutions of Second Order Partial Differential Equations, *Bulletin of the American Mathematical Society* **27**(1): 1–67.

El Karoui, Nicole and Peng, Shige and Quenez, Marie Claire (1997). Backward Stochastic Differential Equations in Finance, *Mathematical Finance* **7**(1): 1–71.

Fang, Fang and Oosterlee, Cornelis W (2008). A Novel Pricing Method for European Options Based on Fourier-Cosine Series Expansions, *SIAM Journal on Scientific Computing* **31**(2): 826–848.

Fleming, Wendell H. and Soner, H. Mete (2006). Controlled Markov Processes and Viscosity Solutions, Vol. 25 of *Stochastic Modelling and Applied Probability*, Springer, New York.

Gobet, Emmanuel and Lemor, Jean-Philippe and Warin, Xavier and others (2005). A Regression-based Monte Carlo Method to Solve Backward Stochastic Differential Equations, *The Annals of Applied Probability* **15**(3): 2172–2202.

Harrison, J Michael and Kreps, David M (1979). Martingales and Arbitrage in Multiperiod Securities Markets, *Journal of Economic Theory* **20**(3): 381–408.

Harrison, J Michael and Pliska, Stanley R (1981). Martingales and Stochastic Integrals in the Theory of Continuous Trading, *Stochastic Processes and their Applications* **11**(3): 215–260.

Heston, Steven L (1993). A Closed-form Solution for Options with Stochastic Volatility with Applications to Bond and Currency Options, *The Review of Financial Studies* **6**(2): 327–343.

Hyndman, Cody B and Ngou, Polynice Oyono (2017). A Convolution Method for Numerical Solution of Backward Stochastic Differential Equations, *Methodology and Computing in Applied Probability* **19**(1): 1–29.

Iserles, Arieh (2009). A First Course in the Numerical Analysis of Differential Equations, number 44, Cambridge university press.

Kloeden, Peter E. and Platen, Eckhard (1992). Numerical Solution of Stochastic Differential Equations, Applications of Mathematics Series, Springer.

Lemor, Jean-Philippe and Gobet, Emmanuel and Warin, Xavier and others (2006). Rate of Convergence of an Empirical Regression Method for Solving Generalized Backward Stochastic Differential Equations, *Bernoulli* **12**(5): 889–916.

Ma, Jin and Shen, Jie and Zhao, Yanhong (2008). On Numerical Approximations of Forward-Backward Stochastic Differential Equations, *SIAM Journal on Numerical Analysis* **46**(5): 2636–2661.

Milstein, GN and Tretyakov, Michael V (2006). Numerical Algorithms for Forward-Backward Stochastic Differential Equations, *SIAM Journal on Scientific Computing* **28**(2): 561–582.

Pardoux, Etienne and Peng, Shige (1990). Adapted Solution of a Backward Stochastic Differential Equation, *Systems & Control Letters* **14**(1): 55–61.

Pham, Huyên (2009). Continuous-time Stochastic Control and Optimization with Financial Applications, Vol. 61 of *Stochastic Modelling and Applied Probability*, Springer-Verlag, Berlin.

Ruijter, Marjon J and Oosterlee, Cornelis W (2015). A Fourier Cosine Method for an Efficient Computation of Solutions to BSDEs, *SIAM Journal on Scientific Computing* **37**(2): A859–A889.

Ruijter, Marjon J and Oosterlee, Cornelis W (2016). Numerical Fourier Method and Second-order Taylor Scheme for Backward SDEs in Finance, *Applied Numerical Mathematics* **103**: 1–26.

Scott, Louis O (1997). Pricing Stock Options in a Jump-Diffusion Model with Stochastic Volatility and Interest Rates: Applications of Fourier Inversion Methods, *Mathematical Finance* **7**(4): 413–426.

Zhao, Weidong and Chen, Lifeng and Peng, Shige (2006). A New Kind of Accurate Numerical Method for Backward Stochastic Differential Equations, *SIAM Journal on Scientific Computing* **28**(4): 1563–1581.

Appendix A

Helpful definitions

Definition A.0.1. (Progressively measurable process).
Let $(\Omega, \mathbb{F}, \{\mathcal{F}_t\}_{t\in\mathbb{T}}, \mathbb{P})$ *be a filtered probability space,* $X = \{X_t\}_{t\in\mathbb{T}}$ *a stochastic process on this space, and* $\mathbb{T}$ *a continuous time index. The process* X *is called progressively measurable if* $\forall\ s \geq 0$*, the map* $(\omega, r) \to X(\omega, r) = X_r(\omega)$ *on* $\Omega \times [0, s]$ *into* $\mathbb{R}^n$ *is* $\mathcal{F}_s \otimes \mathbb{B}([0, s])$ *measurable.*

Definition A.0.2. (Weak convergence of a discrete approximation to an SDE, see Kloeden, Peter E. and Platen, Eckhard (1992), p. 327)
We shall say that a discrete time approximation Y^Δ *converges weakly with order* δ *to a continuous time stochastic process* X *at time* T *as* $\Delta t \downarrow 0$ *if for each* $g \in C_P^{2(\delta+1)}$ *(where the subscript* P *denotes the polynomial growth property of* g *and its first* $2(\delta + 1)$ *derivatives) there exists* $C > 0$*, independent of* Δt*, and a finite* $\delta_0 > 0$ *such that*

$$\left|\mathbb{E}[g(X_T)] - \mathbb{E}[g(Y_T^\Delta)\right| \leq C(\Delta t)^\delta$$

for $t \in (0, \delta_0)$.

Appendix B

Auxiliary results

Theorem B.0.1. *(Comparison theorem, see Pham, Huyên (2009) p. 142)*
Let $(\Omega, \mathbb{F}, \mathbb{P})$ be given and suppose that (ξ_1, f_1) and (ξ_2, f_2) are two pairs of terminal conditions and generators satisfying, for $i = \{1, 2\}$, that $\xi_i \in \mathbb{L}^2_T(\mathbb{R})$, $f_i : [0, T] \times \Omega \times \mathbb{R} \times \mathbb{R} \times \to \mathbb{R}$ such that $f_i(t, \cdot, x, y, z)$ is progressively measurable for all (y, z), $f_i(t, \cdot, 0, 0) \in \mathbb{H}^2_T(\mathbb{R})$ and f_i satisfies a uniform Lipschitz condition in (y, z); that is, $\exists\, C_f$ such that

$$|f(t, \omega, y_1, z_1) - f(t, \omega, y_2, z_2)| \leq C_f(|y_1 - y_2| + |z_1 - z_2|),$$

$\forall\, y_1, y_2, \forall\, z_1, z_2$, $dt \otimes dP$ almost surely. Furthermore, let (Y^1, Z^1), (Y^2, Z^2) be the solutions (2.1) which correspond to (ξ_1, f_1) and (ξ_2, f_2) respectively.

If $\xi_1 \leq \xi_2$ almost surely, $f_1(t, \omega, X^1, Y^1, Z^1) \leq f_2(t, \omega, X^2, Y^2, Z^2)$ $dt \otimes \mathbb{P}$ almost everywhere and $f_2(t, \cdot, X^1, Y^1_t, Z^1_t) \in H^2_T(\mathbb{R})$, then $Y^1_t \leq Y^2_t$ for all $0 \leq t \leq T$ almost surely.

Furthermore, if $Y^2_0 \leq Y^1_0$, then $Y^1_t = Y^2_t$, $0 \leq t \leq T$. In particular, if $P(\xi_1 < \xi_2) > 0$, or $f_1(t, \cdot, \cdot, \cdot) < f_2(t, \cdot, \cdot, \cdot)$ on a set of strictly positive measure $dt \otimes dP$, then $Y^1_0 < Y^2_0$.

Theorem B.0.2. *(Gronwall Theorem)*
Let ν be a continuous positive function on $\mathbb{R}_+$ such that

$$g(t) \leq h(t) + \beta \int_0^t \nu(s) ds,\ 0 \leq t \leq T,$$

where $\beta > 0$, and h is an integrable function on $[0, T]$, $T > 0$. Then

$$g(t) \leq h(t) + \beta \int_0^t h(s) e^{\beta(t-s)} ds,\ 0 \leq t \leq T.$$

COS formula

Let $X = \{X_s\}_{s \in [0,T]}$ be the solution of

$$dX_s = \mu(X_s)ds + \sigma(X_s)dW_s,\ 0 \leq s \leq T,\ X_s = x. \tag{B.1}$$

Let $\mathcal{M} = \{0, t_1, .., t_M\}$ be the discretized time, and let $X^\Delta = \{X^\Delta_s\}_{s \in \mathcal{M}}$ denote a discretization of X. Finally, let $\mathbb{E}^x_m[\cdot]$ denote the conditional expectation on condition that $X_{t_m} = x$. If $h : \mathbb{R} \times \mathbb{R} \to \mathbb{R}$ is a general function of time and of X respectively, then

$$\mathbb{E}_m^x \left[h(t_{m+1}, X_{t_{m+1}}^{\Delta}) \right] \approx \sum_{k=0}^{N-1}{}' \mathcal{H}_k(t_{m+1}) \mathrm{Re} \left[\varphi_{X_{t_{m+1}}} \left(\frac{k\pi}{b-a} \Big| X_{t_m} = x \right) \exp \left(\frac{-ik\pi a}{b-a} \right) \right], \tag{B.2}$$

where the stroke in the summation denotes that the first term in the sum is weighted by a half, where 'Re' denotes the real part of the input argument, and where $\varphi_{X_{t_{m+1}}}(\cdot|X_{t_m})$ is the characteristic function of $X_{t_{m+1}}$ given X_{t_m} with $t_m, t_{m+1} \in \mathcal{M}$. Furthermore, $\{\mathcal{H}_k(t_{m+1})\}_{k=0}^{\infty}$ is the set of Fourier coefficients for the Fourier cosine expansion of $h(t_{m+1}, \cdot)$.