The School of Mathematics

THE UNIVERSITY
of EDINBURGH

Numerical Analysis of the Cox-Ingersoll-Ross Model

by

Tim Howes

Dissertation Presented for the Degree of
MSc in Computational Mathematical Finance

August 2021

Industry Partner – Moody's Analytics

Industry Supervisor – Dr. Greig Smith

Academic Supervisor – Dr. David Šiška

Abstract

In mathematical finance, the Cox-Ingersoll-Ross (CIR) process is often used in financial modelling due to its several appealing features. A unique strong solution to this SDE is known to exist, however, an analytical form for this solution is unknown. As a result, simulating the CIR process on a discrete-time grid can prove to be a difficult task. In this paper, we study the CIR process and some of the available numerical approximation schemes that can be used for its simulation. We discuss the fundamental properties of the CIR process and we explore a range of available approximation schemes as found in mathematical literature. We study in detail the theoretical strong convergence in $\mathcal{L}^p$ of the truncated Milstein scheme for the full parameter range, as shown by Hefter and Herzwurm [12]. Strong and weak convergence rates of approximation schemes are analysed and compared through empirical analysis. We analyse strong convergence rates in $\mathcal{L}^2$ and weak convergence rates are analysed through both zero-coupon bond pricing under the CIR short-rate model and European call option pricing under the Heston model.

Acknowledgements

I would like to sincerely thank my supervisors, Dr. Greig Smith and Dr. David Šiška, for the continuous guidance and support that they gave to me throughout the course of this dissertation.

Own Work Declaration

I declare that this work is my own except where otherwise noted.

Contents

1 Introduction **1**

2 The CIR Process **3**
2.1 Strong Existence and Uniqueness 3
2.2 Probability Distribution 6
2.3 The Feller Condition 9

3 Numerical Schemes **12**
3.1 Available Approximation Schemes 13
3.1.1 Explicit Euler Schemes 13
3.1.2 Implicit Euler Schemes 15
3.1.3 Milstein Schemes 16
3.2 Strong Convergence of the Truncated Milstein Scheme 16

4 Empirical Analysis **25**
4.1 Strong Error Analysis 25
4.2 Weak Error Analysis 29
4.2.1 Zero-Coupon Bond Pricing 30
4.2.2 Option Pricing Under the Heston Model 32

5 Conclusions **35**

Appendices **38**

A More Properties of the CIR Process **38**

B Important Results **38**
B.1 Theorems and Lemmas 38
B.2 Other Results 39

C Additional Proofs **40**
C.1 Proofs from Section 2.3 40
C.2 Proofs from Section 3.2 41

D Code for Empirical Analysis **49**
D.1 Code for Section 4.1 53
D.2 Code for Section 4.2.1 58
D.3 Code for Section 4.2.2 60

List of Tables

1 Observed orders of convergence in $\mathcal{L}^2$ corresponding with Figure 3. 27

List of Figures

1 Plots showing the behaviour of the truncation schemes (driven by the same Wiener process) when their solutions are near 0 and when σ is relatively large ($\delta_0 = 0.2$). . . . 14
2 Plots showing the behaviour of the reflection schemes (driven by the same Wiener process) when their solutions are near 0 and when σ is relatively large ($\delta_0 = 0.2$). . . . 15
3 Log-log plots of RMSE vs. run-time of approximation schemes for $\delta_0 = 0.25, 0.45, 0.75, 1.15$. 26
4 Plot of observed order of convergence of the truncated Milstein scheme and the (truncated) modified Milstein scheme vs. Feller ratio for some $\delta_0 \leq 0.5$. 28
5 Plots of numerical probability of schemes attaining 0 (at time $T = 1$) vs. log run-time for $\delta_0 = 0.25, 0.45, 0.75, 1.15$. 29
6 Log-log plots of bond price error vs. run-time of approximation schemes for $\delta_0 = 0.25, 0.45, 0.75, 1.15$ under the CIR short-rate model. 31
7 Log-log plots of European call option price error vs. run-time of approximation schemes for $\delta_0 = 0.25, 0.45, 0.75, 1.15$ under the Heston model. 34

1 Introduction

In mathematical finance, SDEs are often used to model the dynamics of certain financial quantities (e.g. interest rates, volatility, price of an asset, etc.). Arguably one of the most well-known examples of this is in the Black-Scholes model, where the price of a stock at any given time is assumed to follow a geometric Brownian motion. This assumption is further used in this model to derive fair prices of options on the underlying stock. In many cases, modelling such quantities with SDEs allows one to deduce certain characteristics of the underlying quantity (e.g. its probability distribution) and it also provides a framework in which one can apply Monte Carlo methods to simulate possible paths of the underlying process. In cases where an analytical form of the solution to the SDE is not known, numerical schemes constructed on a discrete-time grid are often used in an attempt to closely approximate the true path of the process. One particular example of such an SDE is the Cox-Ingersoll-Ross (CIR) process, which we study in this paper.

Cox et. al. [7] introduced the CIR model in 1985, when they proposed that the instantaneous short-rate X_t^x at time $t \geq 0$ is the solution to the following SDE:

$$\mathrm{d}X_t^x = (a - kX_t^x)\mathrm{d}t + \sigma\sqrt{X_t^x}\mathrm{d}W_t,\ X_0^x = x \geq 0,\ \ t \geq 0, \tag{1.1}$$

where $a \geq 0$, $k \in \mathbb{R}$, $\sigma > 0$ and $W = (W_t)_{t\geq 0}$ is a Wiener process defined on the probability space $(\Omega, \mathcal{F}, \mathbb{P})$[1]. The CIR process has certain properties which can make it an attractive model. When $k > 0$, it is mean-reverting to mean $\frac{a}{k}$, and k corresponds to the speed of mean-reversion. The solution is also nonnegative, as a result of the parameter $a \geq 0$ preventing the process from going negative if it attains 0. When the Feller condition holds ($\sigma^2 \leq 2a$), the process stays strictly positive and when it is violated, 0 is attainable with a positive probability. More importantly, a unique strong solution to (1.1) is known to exist. Although an analytical form for this solution is not known, its probability distribution is known which allows for the CIR process to be simulated exactly. In particular, exact simulation for values of X_t^x requires sampling from a non-central χ^2-distribution.

Not only is the CIR process used to model the short-rate, but it was also used by Heston [14] to model the volatility of a stock's price. The nonnegativity of the CIR process is a useful property in the Heston model, as it ensures that the stock price's volatility does not take a negative value. Another key property of this model is that the Wiener processes used to drive the stock price and volatility processes are correlated. This model is just one example of how the CIR process is used to model just part of another diffusion process. Under such a setting, it can be challenging to perform exact simulation when the value of the CIR process is conditional on the whole diffusion. In some cases, the whole diffusion may need to be approximated anyway so exact simulation may even be unneccessary. In general, sampling from a non-central χ^2-distribution can often bring a computational burden with it anyway, especially when it is necessary to sample at many different time-points (e.g. when calculating the price of a path dependent option). In the case of the CIR process, it is therefore often preferred to implement numerical approximations that converge to the true solution in either a strong or weak sense, depending on the context of application.

In the context of numerical schemes, we assume the following setting. For a fixed $T > 0$ and $N \in \mathbb{N}$, we have $\Delta t = \frac{T}{N}$ with a uniform time grid

$$T = N\Delta t,\ \ t_n = n\Delta t,\ \ \forall n \in \{0, 1, ..., N\}.$$

We denote $\hat{X}_{t_i}^{x,N}$ for any approximation scheme $\hat{X}^{x,N} = (\hat{X}_t^{x,N})_{0\leq t\leq T}$ that uses constant interpolation between two grid points, i.e. $\hat{X}_t^{x,N} = \hat{X}_{t_i}^{x,N}$ for $t \in [t_i, t_{i+1})$. We also denote $\Delta W_i = W_{t_{i+1}} - W_{t_i}$ for a one dimensional $\mathcal{F}_t$-adapted Wiener process $W = (W_t)_{0\leq t\leq T}$. Under this setting, the Euler-Maruyama

[1] We denote by $(\mathcal{F}_t)_{t\geq 0}$ the Wiener filtration.

scheme for the CIR process should be given by

$$\hat{X}_{t_0}^{x,N} = x, \;\; \hat{X}_{t_{i+1}}^{x,N} = \hat{X}_{t_i}^{x,N} + (a - k\hat{X}_{t_i}^{x,N})\Delta t + \sigma\sqrt{\hat{X}_{t_i}^{x,N}}\Delta W_i,$$

for $i = 0, ..., N-1$. However, this scheme is not well-defined since the Wiener increments can cause the scheme to go negative with some positive probability, leaving the diffusion coefficient undefined. Although this is the case, Gyöngy et al. [11] have shown that Euler approximations to the CIR process do in fact converge strongly at a rate of $1/\log(N)$. Nevertheless, implementing the standard Euler scheme in practice remains a problem, which is exacerbated when the Feller condition does not hold. Similarly, the standard Milstein scheme given by

$$\hat{X}_{t_0}^{x,N} = x, \;\; \hat{X}_{t_{i+1}}^{x,N} = \left(\sqrt{\hat{X}_{t_i}^{x,N}} + \frac{\sigma}{2}\Delta W_i\right)^2 + (a - \frac{\sigma^2}{4} - k\hat{X}_{t_i}^{x,N})\Delta t,$$

for $i = 0, ..., N-1$, is only well-defined for $k \leq 0$ and $\sigma^2 \leq 4a$ because otherwise, we will have $a - \frac{\sigma^2}{4} - k\hat{X}_{t_i}^{x,N} < 0$. This can then bring the scheme into the negative values with some positive probability, leaving it undefined. As a result, several modifications to these schemes have been proposed. Standard fixes for the explicit Euler-Maruyama scheme include taking the absolute value or taking the positive part within the square root term when the scheme takes a negative value, as proposed by Higham et al. [15] and Deelstra et al. [9] respectively. Similar fixes involving the 'reflection' and 'truncation at 0' of the Euler-Maruyama scheme have been proposed by Berkaoui et al. [5] and Lord et al. [18] respectively. Brigo et al. [6] and Alfonsi [4] have proposed implicit Euler schemes (one of which is further studied by Dereich et al. [10], Alfonsi [2] and Neuenkirch [21]), while Alfonsi [4] has also proposed a modification to the Milstein scheme. These schemes have been shown to converge strongly under different parameter settings, since violation of the Feller ratio generally downgrades the performance of such numerical schemes. More recently, Hefter and Herzwurm [12] have proposed the truncated Milstein scheme, which has been proven to converge strongly for the full parameter range. However, the rate at which this scheme converges also downgrades as the Feller condition becomes more violated. Nevertheless, studying the relative performance of these schemes under different parameter settings could prove useful and might indicate the level of suitability of a given scheme depending on the context in which it is applied.

This paper is structured as follows. In Section 2, the main properties of the CIR process are studied. In Section 3, a selection of available numerical approximation schemes for the CIR process are discussed and the theoretical strong convergence of the truncated Milstein scheme is studied in detail. In Section 4, empirical analysis results on the convergence properties of some approximation schemes are discussed, whereas Section 5 concludes.

2 The CIR Process

As mentioned in Section 1, when $k > 0$ the CIR process is mean reverting to mean $\frac{a}{k}$ at a mean-reversion speed of k. Under this setting, it is common to rewrite (1.1) as

$$\mathrm{d}X_t^x = k(\theta - X_t^x)\mathrm{d}t + \sigma\sqrt{X_t^x}\mathrm{d}W_t,\ X_0^x = x \geq 0,\ \ t \geq 0,$$

where $\theta = \frac{a}{k}$. In practice, this more intuitive form of the SDE is often preferred. However, in this paper we will consider (1.1) since we do not make the assumption that $k > 0$. Firstly, we see that the CIR process differs from the Ornstein-Uhlenbeck process since a square root term in the diffusion coefficient is included. Unlike the Ornstein-Uhlenbeck process, the CIR process cannot take a negative value. In addition, the variability of the CIR process at any given time is proportional to the value of the process at that time, whereas the variability of the Ornstein-Uhlenbeck process does not have this dependence. In particular, the variability of the CIR process increases as it takes larger values and the opposite occurs as it takes smaller values. There is a nice connection between the CIR process and the Ornstein-Uhlenbeck process, in that the CIR process can be written as the sum of $d \in \mathbb{N}$ squared Ornstein-Uhlenbeck processes whenever we choose a and σ such that $\frac{4a}{\sigma^2} \in \mathbb{N}$ (as shown in [3]). To see this, we define the Ornstein-Uhlenbeck processes

$$\mathrm{d}Y_t^i = -\frac{k}{2}Y_t^i\mathrm{d}t + \frac{\sigma}{2}\mathrm{d}W_t^i,\ \ t \geq 0,\ \ Y_0^i = \sqrt{\frac{x}{d}},\ \ \text{for}\ \ 1 \leq i \leq d,$$

and set $X_t = \sum_{i=1}^d \left(Y_t^i\right)^2$ where $W_1, ..., W_d$ are independent Wiener processes. Applying Itô's formula to the process $Y = (Y_t)_{t\geq 0}$ and the function $x \mapsto x^2$ gives

$$\mathrm{d}(Y_t^i)^2 = \Big(- k(Y_t^i)^2 + \frac{\sigma^2}{4}\Big)\mathrm{d}t + \sigma Y_t^i\mathrm{d}W_t^i,\ \ t \geq 0,\ \ (Y_0^i)^2 = \frac{x}{d},\ \ \text{for}\ \ 1 \leq i \leq d.$$

Thus, we have

$$\mathrm{d}X_t = \sum_{i=1}^d \mathrm{d}(Y_t^i)^2 = \Big(d\frac{\sigma^2}{4} - kX_t\Big)\mathrm{d}t + \sigma\sum_{i=1}^d Y_t^i\mathrm{d}W_t^i,\ \ t \geq 0,\ \ X_0 = x.$$

Since $\langle\sum_{i=1}^d Y_t^i\mathrm{d}W_t^i\rangle = X_t\mathrm{d}t = \left(\sqrt{X_t}\mathrm{d}W_t\right)^2$ for some Wiener process $(W_t)_{t\geq 0}$, we can write $\sum_{i=1}^d Y_t^i\mathrm{d}W_t^i = \sqrt{X_t}\mathrm{d}W_t$. Hence, X_t is a CIR process starting from x with parameters $a = d\frac{\sigma^2}{4}$, k and σ. We can deduce from the fact that the probability distribution for the Ornstein-Uhlenbeck process is Gaussian that X_t is the sum of d squared Gaussian random variables that have mean $e^{-kt/2}\sqrt{\frac{x}{d}}$ and variance $\frac{\sigma^2}{4}\Big(\frac{1}{k}\big(1 - e^{-kt}\big)\cdot \mathbb{1}_{\{k\neq 0\}} + t\cdot\mathbb{1}_{\{k=0\}}\Big)$. We now look at the SDE (1.1) in more detail. Namely, we look at its strong existence and uniqueness.

2.1 Strong Existence and Uniqueness

Standard results for strong uniqueness of SDEs often require that the drift and diffusion coefficents are locally Lipschitz (e.g. [17], Section 5, Theorem 2.5). Similarly, results for the strong existence of SDEs usually require that the drift and diffusion coefficents are globally Lipschitz (e.g. [17], Section 5, Theorem 2.9). For the CIR process, we cannot apply either result since the square root that appears in the diffusion coefficient is not locally Lipschitz in the neighbourhood of 0 [2]. Even though this is the case, it has been shown that there still exists a unique strong solution to (1.1). Proving this result consists of two main steps. The first step involves showing that there exists a unique strong solution to the following SDE:

$$\mathrm{d}X_t^x = (a - kX_t^x)\mathrm{d}t + \sigma\sqrt{|X_t^x|}\mathrm{d}W_t,\ X_0^x = x \geq 0, t \geq 0. \tag{2.1}$$

[2] It is easy to see that the derivative of $x \mapsto \sqrt{x}$ blows up as x approaches 0. Indeed, we have $\frac{\mathrm{d}}{\mathrm{d}x}\sqrt{x} = \frac{1}{2\sqrt{x}} \to \infty$ as $x \to 0+$.

The second step involves showing that the solution of (2.1) remains nonnegative when x is nonnegative, which will allow us to remove the absolute value in the square root term. We give the theorem for uniqueness and existence of strong solution along with its proof (following the approach in [3]) below.

Theorem 2.1. *There exists a unique strong solution to (1.1).*

Proof. Consider the SDE (2.1) and suppose that $\tilde{X}_t^x = x + \int_0^t (a - k\tilde{X}_s^x)\mathrm{d}s + \int_0^t \sigma\sqrt{|\tilde{X}_s^x|}\mathrm{d}W_s,\ t \geq 0$ is another solution to this SDE. We wish to show that $\mathbb{P}(X_t^x = \tilde{X}_t^x, \forall t \geq 0) = 1$, i.e. the solution to (2.1) is pathwise unique. We first want to find the dynamics of $\left|\tilde{X}_t^x - X_t^x\right|$. Since the function $x \mapsto |x|$ is not $\mathcal{C}^2$, we cannot apply Itô's formula directly. Therefore, we construct smooth approximations of the absolute value function to which we can apply Itô's formula. We set $b_n = \exp(\frac{-n(n+1)}{2})$ and see that $e^{-n}b_{n-1} = \exp(-n - \frac{n(n-1)}{2}) = \exp(\frac{-n(n+1)}{2}) = b_n$. Letting $n \geq 1$, we see that $\int_{b_n}^{b_{n-1}} \frac{1}{nx}\mathrm{d}x = \frac{1}{2n}(-n(n-1) + n(n+1)) = 1$ which means we can find a continuous function $f_n : \mathbb{R} \to \mathbb{R}$ with compact support on (b_n, b_{n-1}) such that

$$\forall x \in \mathbb{R}, 0 \leq f_n(x) \leq \mathbb{1}_{\{x\in(b_n,b_{n-1})\}}\frac{2}{nx}, \text{ and } \int_{\mathbb{R}} f_n(x)\mathrm{d}x = 1.$$

Now, define $F_n(x) = \int_0^x f_n(y)dy$ and notice that $\mathbb{1}_{\{x\in(b_{n-1},\infty)\}} \leq \mathbb{1}_{\{x\in(b_n,\infty)\}}$. Hence, we have

$$\forall x \in \mathbb{R}, \mathbb{1}_{\{x\in(b_{n-1},\infty)\}} \leq F_n(x) \leq \mathbb{1}_{\{x\in(b_n,\infty)\}}.$$

Finally, we define

$$\psi_n(x) = \int_0^{|x|} F_n(y)dy. \tag{2.2}$$

Clearly, we have $|\psi_n(x)| \leq \int_0^{|x|} 1\mathrm{d}y = |x|$ and $|\psi_n'(x)| = |F_n(|x|)\cdot\frac{x}{|x|}| \leq |F_n(|x|)| \leq 1$. We also know that ψ_n is $\mathcal{C}^2$ since $\psi_n''(x) = (\frac{x}{|x|})^2 \cdot f_n(|x|) = f_n(|x|)$. Lastly, we can apply Lebesgue's dominated convergence theorem since $F_n(x) \xrightarrow[n\to\infty]{} \mathbb{1}_{\{x\in(0,\infty)\}}$, giving us $\lim_{n\to\infty}|\psi_n(x)| = \lim_{n\to\infty}\psi_n(x) = |x| - 0 = |x|$. By applying Itô's formula to the process $(\tilde{X}_t^x - X_t^x)_{t\geq 0}$ and the function ψ_n, we get

$$\begin{aligned}\mathrm{d}\psi_n(\tilde{X}_t^x - X_t^x) = \Big(&- k(\tilde{X}_t^x - X_t^x)\psi_n'(\tilde{X}_t^x - X_t^x) + \frac{\sigma^2}{2}\big(\sqrt{\tilde{X}_t^x} - \sqrt{X_t^x}\big)^2 f_n\big(|\tilde{X}_t^x - X_t^x|\big)\Big)\mathrm{d}t \\ &+ \sigma\big(\sqrt{\tilde{X}_t^x} - \sqrt{X_t^x}\big)\psi_n'(\tilde{X}_t^x - X_t^x)\mathrm{d}W_t.\end{aligned}$$

Integrating this gives us

$$\begin{aligned}\psi_n(\tilde{X}_t^x - X_t^x) = -k\int_0^t &\Big((\tilde{X}_s^x - X_s^x)\psi_n'(\tilde{X}_s^x - X_s^x) + \frac{\sigma^2}{2}\big(\sqrt{\tilde{X}_s^x} - \sqrt{X_s^x}\big)^2 f_n\big(|\tilde{X}_s^x - X_s^x|\big)\Big)\mathrm{d}s \\ &+ \sigma\int_0^t \big(\sqrt{\tilde{X}_s^x} - \sqrt{X_s^x}\big)\psi_n'(\tilde{X}_s^x - X_s^x)\mathrm{d}W_s.\end{aligned} \tag{2.3}$$

Furthermore, X and $\tilde{X}$ have uniformly bounded moments ([20], Chapter 2, Theorem 4.4), which means that $\mathbb{E}\Big[\sigma\int_0^t(\sqrt{\tilde{X}_s^x} - \sqrt{X_s^x})\psi_n'(\tilde{X}_s^x - X_s^x)\mathrm{d}W_s\Big] = 0$. Hence, taking expectation in (2.3) gives us

$$\begin{aligned}\mathbb{E}\Big[\psi_n(\tilde{X}_t^x - X_t^x)\Big] &= \mathbb{E}\Big[-k\int_0^t \Big((\tilde{X}_s^x - X_s^x)\psi_n'(\tilde{X}_s^x - X_s^x) + \frac{\sigma^2}{2}\big(\sqrt{\tilde{X}_s^x} - \sqrt{X_s^x}\big)^2 f_n\big(|\tilde{X}_s^x - X_s^x|\big)\Big)\mathrm{d}s\Big] \\ &\leq |k|\mathbb{E}\Big[\int_0^t \Big((\tilde{X}_s^x - X_s^x) + \frac{\sigma^2}{2}|\tilde{X}_s^x - X_s^x| f_n\big(|\tilde{X}_s^x - X_s^x|\big)\Big)\mathrm{d}s\Big] \\ &\leq |k|\mathbb{E}\Big[\int_0^t \Big((\tilde{X}_s^x - X_s^x) + \frac{\sigma^2}{2}\cdot\frac{2}{n}\Big)\mathrm{d}s\Big] \\ &\leq |k|\int_0^t \mathbb{E}\Big[|\tilde{X}_s^x - X_s^x|\Big]\mathrm{d}s + \frac{\sigma^2 t}{n},\end{aligned}$$

where we used the fact that $\big|\sqrt{|x_1|} - \sqrt{|x_2|}\big| \leq \sqrt{|x_1 - x_2|}$, $|\psi_n'(x_1)| \leq 1$ and $|x_1 - x_2| f_n\big(|x_1 - x_2|\big) \leq \frac{2}{n}$ for all $x_1, x_2 \in \mathbb{R}$. Applying Lebesgue's dominated convergence theorem gives

$$\lim_{n\to\infty} \mathbb{E}\Big[\psi_n(\tilde{X}_t^x - X_t^x)\Big] = \mathbb{E}\Big[|\tilde{X}_t^x - X_t^x|\Big] \leq |k|\int_0^t \mathbb{E}\Big[|\tilde{X}_s^x - X_s^x|\Big]\mathrm{d}s,$$

and Gronwall's inequality implies $\mathbb{E}\Big[|\tilde{X}_t^x - X_t^x|\Big] \leq 0$, which then implies $\mathbb{E}\Big[|\tilde{X}_t^x - X_t^x|\Big] = 0$. Hence, $\mathbb{P}(\tilde{X}_t^x = X_t^x) = 1$ for all $t \geq 0$ and by the continuity of both X and $\tilde{X}$, we have $\mathbb{P}(\tilde{X}_t^x = X_t^x, \forall t \geq 0) = 1$, i.e. the solution to (2.1) is pathwise unique.

Since $x \mapsto a - kx$ and $x \mapsto \sigma\sqrt{|x|}$ are continuous functions and satisfy the linear growth condition, we know that there exists a weak solution to (2.1) ([16], Theorem 0.1). Then by the pathwise uniqueness of (2.1) and the result obtained by a corollary to the Yamada-Watanabe theorem ([17], Chapter 5, Corollary 3.23), there exists a strong solution to (2.1).

We now complete the proof by showing that the solution X^x to (2.1) remains nonnegative when $x \geq 0$. Using the function ψ_n given by (2.2), we can apply Itô's formula to this function and $X^x = (X_t^x)_{t\geq 0}$ to get

$$\mathrm{d}\psi_n(X_t^x) = \Big((a - kX_t^x)\psi_n'(X_t^x) + \frac{\sigma^2}{2}|X_t^x| f_n\big(|X_t^x|\big)\Big)\mathrm{d}t + \sigma\sqrt{|X_t^x|}\psi_n'(X_t^x)\mathrm{d}W_t.$$

Integrating this gives us

$$\psi_n(X_t^x) = \psi_n(x) + \int_0^t \Big((a - kX_s^x)\psi_n'(X_s^x) + \frac{\sigma^2}{2}|X_s^x| f_n\big(|X_s^x|\big)\Big)\mathrm{d}s + \sigma\int_0^t \sqrt{|X_s^x|}\psi_n'(X_s^x)\mathrm{d}W_s.$$

Using again that X^x has uniformly bounded moments, $a\psi_n'(x) \leq a$ and $x f_n(x) \leq \frac{2}{n}$ for any $x \in \mathbb{R}$, taking expectation then gives us

$$\begin{aligned}\mathbb{E}\big[\psi_n(X_t^x)\big] &= \psi_n(x) + \mathbb{E}\Big[\int_0^t \Big((a - kX_s^x)\psi_n'(X_s^x) + \frac{\sigma^2}{2}|X_s^x| f_n\big(|X_s^x|\big)\Big)\mathrm{d}s\Big] \\ &\leq \psi_n(x) + \mathbb{E}\Big[\int_0^t \Big(a - k\psi_n'(X_s^x)X_s^x + \frac{\sigma^2}{2}\cdot\frac{2}{n}\Big)\mathrm{d}s\Big] \\ &= \psi_n(x) + \int_0^t \Big(a - k\mathbb{E}\big[\psi_n'(X_s^x)X_s^x\big]\Big)\mathrm{d}s + \frac{\sigma^2 t}{n}.\end{aligned}$$

Notice that $x\psi_n'(x) = x\cdot\frac{x}{|x|}\cdot F_n(|x|) = |x|F_n(x) \xrightarrow[n\to\infty]{} |x|$ and $x\psi_n'(x) \leq |x|$ for any $x \in \mathbb{R}$. Therefore, we can apply Lebesgue's dominated convergence theorem to get

$$\lim_{n\to\infty} \mathbb{E}\big[\psi_n(X_t^x)\big] = \mathbb{E}\big[|X_t^x|\big] \leq x + \int_0^t \Big(a - k\mathbb{E}\big[|X_s^x|\big]\Big)\mathrm{d}s, \tag{2.4}$$

and taking the expectation of (2.1) we get

$$\mathbb{E}\big[X_t^x\big] = x + \int_0^t \Big(a - k\mathbb{E}\big[X_s^x\big]\Big)\mathrm{d}s. \tag{2.5}$$

Now using the fact that $\mathbb{E}\big[|X_t^x|\big] - \mathbb{E}\big[X_t^x\big] \geq 0$, subtracting (2.5) from (2.4) then gives

$$\mathbb{E}\big[|X_t^x|\big] - \mathbb{E}\big[X_t^x\big] \leq -k\int_0^t \Big(\mathbb{E}\big[|X_s^x|\big] - \mathbb{E}\big[X_s^x\big]\Big)\mathrm{d}s \leq |k|\int_0^t \Big(\mathbb{E}\big[|X_s^x|\big] - \mathbb{E}\big[X_s^x\big]\Big)\mathrm{d}s.$$

Hence, Gronwall's inequality implies that $\mathbb{E}\big[|X_t^x|\big] = \mathbb{E}\big[X_t^x\big]$, which means that $\mathbb{P}\big(X_t^x = |X_t^x|\big) = 1$, i.e. $X_t^x \geq 0,\ \forall t \geq 0$ almost surely. This concludes the proof. □

An intuitive but non-rigorous argument as to why the CIR process remains nonnegative for all $t \geq 0$ when $x \geq 0$ is that in order for it to take negative values, it must cross zero because of its continuity. When the process is zero, the diffusion coefficient also becomes zero and the drift term just becomes $a \geq 0$. Therefore, the drift term prevents the process from going negative. When $a = 0$ and the process is zero, the process will remain at zero forever. On the other hand, when $a > 0$ the process will be instantly brought back into the positive values.

Remark 2.1. *In addition to the uniqueness and extistence of a strong solution for the CIR process, it also has uniformly bounded moments, i.e., for any $T > 0$ and for all $1 \leq p < \infty$, there exists a constant $C > 0$ depending only on T and p such that*

$$\left(\mathbb{E}\Big[\sup_{0\leq t\leq T}|X_t^x|^p\Big]\right)^{1/p} \leq C(1+x). \tag{2.6}$$

This follows from a general property of SDEs ([20]*, Chapter 2, Theorem 4.4) which states that (2.6) holds if for $2 \leq p < \infty$, $\hat{b}(x) = a - kx$, $\hat{\sigma}(x) = \sigma\sqrt{x}$, $t \in [0,T]$, and $x \geq 0$, there exists a constant $K > 0$ such that*

$$\max\Big(|\hat{b}(x)|^2, |\hat{\sigma}(x)|^2\Big) \leq K(1+x^2). \tag{2.7}$$

Indeed, we have

$$\begin{aligned}\max\Big(|\hat{b}(x)|^2, |\hat{\sigma}(x)|^2\Big) &\leq |\hat{b}(x)|^2 + |\hat{\sigma}(x)|^2 \\ &\leq 2a^2 + 2k^2x^2 + \sigma^2 x \\ &\leq 2a^2 + 2k^2x^2 + \sigma^2(1+x^2) \\ &\leq K(1+x^2)\end{aligned}$$

Hence, the linear growth condition (2.7) holds for $2 \leq p < \infty$ and by Hölder's inequality it holds for $1 \leq p < \infty$, i.e. we have

$$\mathbb{E}\Big[\sup_{0\leq t\leq T}|X_t^x|\Big] \leq \left(\mathbb{E}\Big[\sup_{0\leq t\leq T}|X_t^x|^p\Big]\right)^{1/p} \leq K(1+x^2).$$

2.2 Probability Distribution

As mentioned previously, there is a nice connection between the Ornstein-Uhlenbeck process and the CIR process under some rather strict parameter conditions. In particular, the value that the CIR process takes at time $t \geq 0$ is a random variable equal in distribution to the sum of $d = \frac{4a}{\sigma^2}$ squared normally

distributed random variables with mean $e^{-kt/2}\sqrt{\frac{x}{d}}$ and variance $\frac{\sigma^2}{4}\Big(\frac{1}{k}\big(1-e^{-kt}\big)\cdot \mathbb{1}_{\{k\neq 0\}} + t\cdot \mathbb{1}_{\{k=0\}}\Big)$ whenever $d\in\mathbb{N}$. It still remains to determine the probability distribution of the CIR process for the full parameter range. As is well known, the characteristic function allows us to identify the probability distribution of a random variable and we will see that the characteristic function for the CIR process is that of a non-central χ^2-distribution, as shown in [3].

Proposition 1. *Let X^x denote the solution to (1.1) and set*

$$\theta_k(t) = \begin{cases} t \text{ if } k=0, \\ \frac{1}{k}(1-e^{-kt}) \text{ otherwise.} \end{cases} \tag{2.8}$$

The characteristic function of X_t^x is well-defined on

$$\{i\in\mathbb{C}, \mathbb{E}\big[|e^{iX_t^x}|\big]<\infty\} = \{i\in\mathbb{C}, \mathrm{Re}(i) < \frac{2}{\sigma^2\theta_k(t)}\} \tag{2.9}$$

and is given by

$$\mathbb{E}\big[e^{iX_t^x}\big] = \big(1-\frac{\sigma^2}{2}i\theta_k(t)\big)^{-\frac{2a}{\sigma^2}}\exp\Big(\frac{ixe^{-kt}}{1-\frac{\sigma^2}{2}i\theta_k(t)}\Big) \tag{2.10}$$

Proof. (Following the approach in [3]) Firstly, we assume i is a nonpositive real number and as a basic property of affine diffusion processes (see e.g. [3], Proposition 1.3.1), we can write $\mathbb{E}\big[e^{iX_t^x}\big] = \exp(\phi_i(t)+x\psi_i(t))$ for some smooth functions ϕ_i and ψ_i, with $\phi_i(0)=0$ and $\psi_i(0)=i$. Since we know X_t^x is Markovian, we have for $0\le t\le T$

$$\mathbb{E}\big[e^{iX_T^x}|\mathcal{F}_t\big] = \mathbb{E}\big[e^{iX_T^{X_t^x}}\big] = \exp\big(\phi_u(T-t)+\psi_u(T-t)X_t^x\big).$$

Setting $M_t = \mathbb{E}\big[e^{iX_T^x}|\mathcal{F}_t\big]$, we see that it is an $\mathcal{F}_t$ martingale because it is clearly $\mathcal{F}_t$ measurable with

$$\mathbb{E}\Big[\Big|\mathbb{E}\big[e^{iX_T^x}|\mathcal{F}_t\big]\Big|\Big] = \mathbb{E}\big[e^{iX_T^x}\big] \le e^0 < \infty,$$

since we know $X_T^x\ge 0$ and $i\le 0$. For $0\le s\le t$, we have

$$\mathbb{E}\big[M_t|\mathcal{F}_s\big] = \mathbb{E}\Big[\mathbb{E}\big[e^{iX_T^x}|\mathcal{F}_t\big]|\mathcal{F}_s\Big] = \mathbb{E}\big[e^{iX_T^x}|\mathcal{F}_s\big] = M_s,$$

by the Tower property. Itô's formula also gives

$$\begin{aligned} dM_t = M_t\Big(&\Big(-\phi_i'(T-t) - \psi_i'(T-t)X_t^x + (a-kX_t^x)\psi_i(T-t) + \frac{\sigma^2}{2}\psi_i(T-t)^2X_t^x\Big)\mathrm{d}t \\ &+ \psi_i(T-t)\sigma\sqrt{X_t^x}\mathrm{d}W_t\Big). \end{aligned} \tag{2.11}$$

Since M_t is a martingale, we have

$$\begin{aligned} &-\phi_i'(T-t) - \psi_i'(T-t)X_t^x + (a-kX_t^x)\psi_i(T-t) + \frac{\sigma^2}{2}\psi_i(T-t)^2X_t^x = 0 \\ \iff\ &-\phi_i'(T-t) + a\psi_i(T-t) + X_t^x\Big(-\psi_i'(T-t) - k\psi_i(T-t) + \frac{\sigma^2}{2}\psi_i(T-t)^2\Big) = 0. \end{aligned}$$

Separating the terms with and without X_t^x leads to the following system of ODEs:

$$\begin{cases} -\phi_i'(t) + a\psi_i(t) = 0 \\ -\psi_i'(t) - k\psi_i(t) + \frac{\sigma^2}{2}\psi_i(t)^2 = 0 \end{cases},\ t\ge 0. \tag{2.12}$$

The second ODE is a Riccati differential equation and it can be solved explicitly since $y = 1/\psi_i$ solves

$y' - ky + \frac{\sigma^2}{2} = 0$. This gives $y(t) = e^{kt}(y(0) - \frac{\sigma^2}{2}\theta_k(t)) = e^{kt}(\frac{1}{i} - \frac{\sigma^2}{2}\theta_k(t))$. Solving for $\psi_i(t) = 1/y(t)$, we get

$$\psi_i(t) = \frac{ie^{-kt}}{1 - \frac{\sigma^2}{2}i\theta_k(t)}.$$

Now solving for ϕ_i in (2.12), we get

$$\begin{aligned}
\phi_i(t) &= a\int_0^t \psi_i(s)\mathrm{d}_s \\
&= a\int_0^t \frac{ie^{-kt}}{1 - \frac{\sigma^2}{2}i\theta_k(t)}\mathrm{d}s \\
&= -\frac{2a}{\sigma^2}\int_{b(t)}^{c(t)} \frac{1}{u}du \quad \text{where } u = 1 - \frac{\sigma^2}{2}i\theta_k(s), \\
&= -\frac{2a}{\sigma^2}\Big[\log\big(1 - \frac{\sigma^2}{2}i\theta_k(s)\big)\Big]_{s=0}^{s=t} \\
&= -\frac{2a}{\sigma^2}\log\big(1 - \frac{\sigma^2}{2}i\theta_k(t)\big).
\end{aligned}$$

This gives us the right hand side of (2.10). However, we have not shown the equality in (2.10) as we have only guessed a special form for $\mathbb{E}\big[e^{iX_t^x}\big]$ and derived the neccessary conditions for it. Nevertheless, the equality in (2.10) holds as follows. Integrating (2.11) from 0 to T with the drift term equal to 0 gives

$$e^{iX_T^x} = \exp\big(\phi_i(T) + \psi_i(T)x\big) + \int_0^T \psi_i(T-t)M_t\sigma\sqrt{X_t^x}\mathrm{d}W_t. \tag{2.13}$$

By noticing that $\phi_i(t) \le 0$ and $\psi_i(t) \le 0$ for all $t \ge 0$ (since $i \le 0$ and $\theta_k(t) \ge 0$ for all $t \ge 0$), we have that $0 \le M_t \le 1$. This means $\mathbb{E}\Big[\int_0^T \psi_i(T-t)M_t\sigma\sqrt{X_t^x}\mathrm{d}W_t\Big] = 0$ and taking expectation in (2.13) gives us (2.10).

It remains to find the set on which the characteristic function is well-defined. We do not prove this but we remark that it is easy to see that $\{i \in \mathbb{R} : \mathbb{E}\big[e^{iX_t^t}\big] < \infty\}$ is an interval that contains the negative real numbers (since X_t^x is nonnegative), and from (2.10) we can see that $\mathbb{E}\big[e^{iX_t^x}\big] \to \infty$ as $i \to \frac{2}{\sigma^2\theta_k(t)}$. $\square$

The characteristic function given in (2.10) is known to be that of a non-central χ^2-distribution (the characteristic and density functions of this distribution can be found in Appendix B). In particular, the CIR process defined by (1.1) has the following probability density function when $a > 0$:

$$f(t,x,X) = \sum_{n=0}^{\infty} \frac{e^{-\frac{d_t x}{2}}(\frac{d_t x}{2})^n}{n!}\frac{c_t/2}{\Gamma(n + \frac{2a}{\sigma^2})}\Big(\frac{c_t X}{2}\Big)^{n-1+\frac{2a}{\sigma^2}}e^{-c_t X/2}, \quad X \ge 0, \tag{2.14}$$

where $c_t = \frac{4}{\sigma^2\theta_k(t)}$, $d_t = c_t e^{-kt}$ and $\Gamma(\alpha) = \int_0^\infty X^{\alpha-1}e^{-X}\mathrm{d}X$, $\alpha > 0$. Hence, $c_t X_t^x$ has a non-central χ^2-distribution with $\frac{4a}{\sigma^2}$ degrees of freedom and non-centrality parameter xd_t. When $a = 0$, it is distributed according to the probability measure

$$e^{-d_t x/2}\delta_0(\mathrm{d}X) + \sum_{n=1}^{\infty} \frac{e^{-\frac{d_t x}{2}}(\frac{d_t x}{2})^n}{n!}\frac{c_t/2}{\Gamma(n)}\Big(\frac{c_t X}{2}\Big)^{n-1}e^{-c_t X/2}\mathrm{d}X, \quad X \ge 0, \tag{2.15}$$

where $\delta_0(\mathrm{d}X)$ is the Dirac mass at 0 and $\mathrm{d}X$ is the Lebesgue measure. To see this, we follow the approach in [3] and we first consider the case for $a > 0$. By letting $i \le 0$, we have

$$\mathbb{E}\big[e^{iX_t^x}\big] = \int_0^\infty e^{iX} f(t,x,X)\mathrm{d}X = \int_0^\infty \sum_{n=0}^\infty e^{iX} \frac{e^{-\frac{d_t x}{2}}(\frac{d_t x}{2})^n}{n!} \frac{c_t/2}{\Gamma(n+\frac{2a}{\sigma^2})} \Big(\frac{c_t X}{2}\Big)^{i-1+\frac{2a}{\sigma^2}} e^{-c_t X/2}\mathrm{d}X$$
$$= \sum_{n=0}^\infty \int_0^\infty e^{iX} \frac{e^{-\frac{d_t x}{2}}(\frac{d_t x}{2})^n}{n!} \frac{c_t/2}{\Gamma(n+\frac{2a}{\sigma^2})} \Big(\frac{c_t X}{2}\Big)^{n-1+\frac{2a}{\sigma^2}} e^{-c_t X/2}\mathrm{d}X$$
$$= \sum_{n=0}^\infty \frac{e^{-\frac{d_t x}{2}}(\frac{d_t x}{2})^n}{n!} \frac{c_t/2}{\Gamma(n+\frac{2a}{\sigma^2})} \Big(\frac{c_t}{2}\Big)^{n-1+\frac{2a}{\sigma^2}} \int_0^\infty X^{n-1+\frac{2a}{\sigma^2}} e^{-(c_t/2-i)X}\mathrm{d}X.$$

Now by using the definition of Γ, we get

$$\int_0^\infty X^{n-1+\frac{2a}{\sigma^2}} e^{-(c_t/2-i)y}\mathrm{d}X = \frac{1}{(c_t/2-i)^{n+\frac{2a}{\sigma^2}}} \int_0^\infty X^{n+\frac{2a}{\sigma^2}-1} e^{-X}\mathrm{d}X$$
$$= \frac{1}{(c_t/2-i)^{n+\frac{2a}{\sigma^2}}} \Gamma\Big(n+\frac{2a}{\sigma^2}\Big).$$

Hence, we have

$$\mathbb{E}[e^{iX_t^x}] = \sum_{n=0}^\infty \frac{e^{-\frac{d_t x}{2}}(\frac{d_t x}{2})^n}{n!} \Big(\frac{c_t}{c_t-2i}\Big)^{i+\frac{2a}{\sigma^2}}$$
$$= \Big(\frac{c_t}{c_t-2i}\Big)^{\frac{2a}{\sigma^2}} \exp\Big(\frac{d_t}{2} x \Big(\frac{c_t}{c_t-2i}-1\Big)\Big)$$
$$= \Big(\frac{c_t}{c_t-2i}\Big)^{\frac{2a}{\sigma^2}} \exp\Big(ixe^{-kt}\Big(\frac{c_t}{c_t-2i}\Big)\Big)$$
$$= \Big(1-\frac{\sigma^2}{2} i\theta_k(t)\Big)^{-\frac{2a}{\sigma^2}} \exp\Big(\frac{ixe^{-kt}}{1-\frac{\sigma^2}{2}i\theta_k(t)}\Big),$$

which is the characteristic function in (2.10). For $a = 0$, the calculation is the same once we notice that $\int_0^\infty e^{iX} e^{-d_t x/2}\delta_0(\mathrm{d}X) = e^0 e^{-d_t x/2} = e^{-d_t x/2}$. Thus, we know the probability distribution for the CIR process and that a direct approach to exactly simulating values of X_t^x would require sampling from a non-central χ^2-distribution. In Section 3, we will discuss the potential drawbacks that this can bring to exact simulation methods for the CIR process.

2.3 The Feller Condition

We know from Section 2.1 that the CIR process is non-negative. In fact, the CIR process is strictly positive when $\sigma^2 \leq 2a$, which can be quite a useful property in certain contexts, e.g. when modelling volatility in the Heston model. This particular parameter setting is known as the Feller condition. On the other hand, there is a positive probability that the CIR process attains the value 0 when the Feller condition does not hold. We will see in Sections 3 and 4 that the performance of numerical approximation schemes to the CIR process generally perform worse as the Feller ratio decreases. Following the approach in [3], we give the formal result and proof here.

Proposition 2. *Let $x > 0$ and $\tau_0 = \inf\{t \geq 0, X_t^x = 0\}$. Then $\tau_0 = \inf\emptyset = \infty$ a.s. if and only if*

$$\sigma^2 \leq 2a.$$

When $\sigma^2 > 2a$, we have $\tau_0 < \infty$ a.s. if and only if $k \geq 0$.

Before giving the proof for Proposition 2, we require some useful definitions and one particular lemma. Firstly, we define the scaling function for the CIR process for $x' > 0$ as

$$s(x) = \int_1^x e^{\frac{2ky}{\sigma^2}} y^{-\frac{2a}{\sigma^2}} \mathrm{d}y. \tag{2.16}$$

Then we clearly have

$$\lim_{x' \to 0+} s(x') = -\infty \iff \sigma^2 \leq 2a, \tag{2.17}$$

$$\lim_{x' \to +\infty} s(x') = \infty \iff k \geq 0 \text{ and } \sigma^2 \geq 2a. \tag{2.18}$$

Lastly, define $\tau_m = \inf\{t \geq 0, X_t^x = m\}$, $\tau_{m,m'} = \min\{\tau_m, \tau_{m'}\}$ for $m, m' \geq 0$ and we state the following result.

Lemma 2.1. *Let X^x denote the CIR process defined by (1.1) and consider $\underline{m}, \overline{m}$ such that*

$$0 < \underline{m} < x < \overline{m} < \infty.$$

Then

$$\mathbb{P}(\tau_{\underline{m}} < \tau_{\overline{m}}) = \frac{s(\overline{m}) - s(x)}{s(\overline{m}) - s(\underline{m})}. \tag{2.19}$$

The proof for Lemma 2.1 is postponed until Appendix C. We are now in a position to prove Proposition 2.

Proof of Proposition 2. We consider an increasing sequence $(\overline{m}_n)_{n\in\mathbb{N}} \in (x, \infty)$ and a decreasing sequence $(\underline{m}_n)_{n\in\mathbb{N}} \in (0, x)$ such that $\overline{m}_n \xrightarrow[n\to\infty]{} \infty$ and $\underline{m}_n \xrightarrow[n\to\infty]{} 0$. Clearly, $\tau_{\overline{m}_n} \leq \tau_{\overline{m}_{n+1}}$ and $\tau_{\underline{m}_n} \leq \tau_{\underline{m}_{n+1}}$ for any $n \in \mathbb{N}$, so $\tau_{\overline{m}_n}$ and $\tau_{\underline{m}_n}$ are increasing sequences of stopping times. We also have for any $T > 0$ that

$$\mathbb{P}(\tau_{\overline{m}_n} \leq T) = \mathbb{P}\Big(\sup_{0\leq t\leq T} X_t^x \geq \overline{m}_n \Big) \xrightarrow[n\to\infty]{} 0,$$

and so $\lim_{n\to\infty} \tau_{\overline{m}_n} = \infty$ almost surely. Similarly, we have $\lim_{n\to\infty} \tau_{\underline{m}_n} = \tau_0$ almost surely.

Case 1: $\sigma^2 \leq 2a$.
Since $0 < \underline{m} < x$, we clearly have $\mathbb{P}(\tau_0 < \tau_{\overline{m}}) \leq \mathbb{P}(\tau_{\underline{m}} < \tau_{\overline{m}})$. By letting $\underline{m} \to 0+$, we get from (2.19) that

$$\mathbb{P}(\tau_0 < \tau_{\overline{m}}) = \frac{s(\overline{m}) - s(x)}{s(\overline{m}) - \lim_{x'\to 0+} s(x')} = \frac{s(\overline{m}) - s(x)}{s(\overline{m}) + \infty} = 0. \tag{2.20}$$

This then gives us

$$\begin{aligned}
\mathbb{P}(\tau_0 < \infty) &= \mathbb{P}(\tau_0 < \lim_{n\to\infty} \tau_{\overline{m}_n}) \\
&= \mathbb{E}\Big[\mathbb{1}_{\{\tau_0 < \lim_{n\to\infty} \tau_{\overline{m}_n}\}} \Big] \\
&= \lim_{n\to\infty} \mathbb{E}\Big[\mathbb{1}_{\{\tau_0 < \tau_{\overline{m}_n}\}} \Big] \\
&= \lim_{n\to\infty} \mathbb{P}(\tau_0 < \tau_{\overline{m}_n}) \\
&= 0,
\end{aligned}$$

where we applied Lebesgue's dominated convergence theorem to $f_n = \mathbb{1}_{\{\tau_0 < \tau_{\overline{m}_n}\}}$.

Case 2: $\sigma^2 > 2a$.
We note that the process $s\big(X^x_{\tau_{\underline{m}_n,\overline{m}}}\big)$ is an almost surely bounded discrete-time martingale with respect to the filtration $\mathcal{F}_{\tau_{\underline{m}_n,\overline{m}}}$ (see the proof of Lemma 2.1). Hence, it converges almost surely ([17], Chapter 1, Theorem 3.15) and we have $\mathbb{E}\Big[\int_0^{\tau_{\underline{m}_n,\overline{m}}} \sigma^2 X_u^x s'(X_u^x)^2 \mathrm{d}u \Big] < \infty$ for all $n \in \mathbb{N}$. Hence,

$$\lim_{n\to\infty}\mathbb{E}\Big[\int_0^{\tau_{\underline{m}_n,\overline{m}}}\sigma^2X_u^xs'(X_u^x)^2\mathrm{d}u\Big]\geq\sigma^2\lim_{n\to\infty}\min_{x\in[\underline{m}_n,\overline{m}]}xs'(x)^2\mathbb{E}\Big[\int_0^{\tau_{\underline{m}_n,\overline{m}}}\mathrm{d}u\Big]$$
$$=\sigma^2\min_{x\in[0,\overline{m}]}xs'(x)^2\lim_{n\to\infty}\mathbb{E}\big[\tau_{\underline{m}_n,\overline{m}}\big]$$
$$=\sigma^2\min_{x\in[0,\overline{m}]}xs'(x)^2\mathbb{E}\big[\tau_{0,\overline{m}}\big],\tag{2.21}$$

where on the last line we apply Lebesgue's dominated convergence theorem to $\tau_{\underline{m}_n,\overline{m}}$. Therefore we have $\sigma^2\min_{x\in[0,\overline{m}]}xs'(x)^2\mathbb{E}\big[\tau_{0,\overline{m}}\big]<\infty$. Notice that for $\sigma^2\in(2a,4a]$, we have $\min_{x\in[0,\overline{m}]}xs'(x)^2=e^{\frac{4kx}{\sigma^2}}x^{1-\frac{4a}{\sigma^2}}>0$ and since (2.21) is finite we get $\mathbb{P}(\tau_{0,\overline{m}}<\infty)=1$. We now define

$$\hat{X}_t^{x/2}=\frac{x}{2}+\int_0^t\Big(\frac{a}{2}-k\hat{X}_s^{x/2}\Big)\mathrm{d}s+\int_0^t\sigma\sqrt{\hat{X}_s^{x/2}}\mathrm{d}W_s,$$
$$\hat{\tau}_m=\inf\{t\geq0,\hat{X}_t^{x/2}=m\},$$
$$\hat{\tau}_{m,m'}=\min\{\hat{\tau}_m,\hat{\tau}_{m'}\}\ \text{ for } m,m'\geq0,$$

and we have the following result (see Exercise 1.2.13, [3]),

$$\mathbb{P}(\tau_{0,\overline{m}}<\infty)=1\implies\mathbb{P}(\hat{\tau}_{0,\overline{m}/2}<\infty)=1.\tag{2.22}$$

Notice that for any $t\geq0$, we have

$$\hat{X}_t^{x/2}=\overline{m}/2\implies2\hat{X}_t^{x/2}=\overline{m},$$

and setting $\tilde{X}_t^x:=2\hat{X}_t^{x/2}$ gives us

$$\tilde{X}_t^x=x+\int_0^t(a-k\tilde{X}_s^x)\mathrm{d}s+\int_0^t\frac{2\sigma}{\sqrt{2}}\sqrt{\tilde{X}_s^x}\mathrm{d}W_s,$$

and we see that $(\frac{2\sigma}{\sqrt{2}})^2=2\sigma^2$. Thus, $\tilde{X}_t^x$ is the solution to a CIR process with parameters a, k and $\sqrt{2}\sigma$. Hence, from (2.22) we get that $\mathbb{P}(\tau_{0,\overline{m}}<\infty)=1$ for $\sigma^2\in(4a,8a]$ and induction through this argument gives $\mathbb{P}(\tau_{0,\overline{m}}<\infty)=1$ for $\sigma^2\in(2^ka,2^{k+1}a],\ k\in\mathbb{N}$. Thus $\mathbb{P}(\tau_{0,\overline{m}}<\infty)=1$ holds for any $\sigma^2>2a$. Therefore, we have

$$s(x)=\lim_{n\to\infty}\mathbb{E}\Big[s(X^x_{\tau_{\underline{m}_n,\overline{m}}})\Big]=\mathbb{E}\Big[\lim_{n\to\infty}s(X^x_{\tau_{\underline{m}_n,\overline{m}}})\Big]=\mathbb{E}\Big[s(X^x_{\tau_{0,\overline{m}}})\Big],$$

by the martingale property and then by Lebesgue's dominated convergence theorem. Finally, this gives

$$\mathbb{P}(\tau_0<\tau_{\overline{m}})=\frac{s(\overline{m})-s(x)}{s(\overline{m})-\lim\limits_{x'\to0+}s(x')}.$$

Using (2.18), we get for $\sigma^2>2a$:

$$\mathbb{P}(\tau_0<\infty)=\lim_{n\to\infty}\mathbb{P}(\tau_0<\tau_{\overline{m}_n})$$
$$=\begin{cases}1,\ \text{if } k\geq0,\\ \frac{\lim\limits_{x'\to+\infty}s(x')-s(x)}{\lim\limits_{x'\to+\infty}s(x')-\lim\limits_{x'\to0+}s(x')}\in(0,1),\ \text{otherwise.}\end{cases}$$

□

3 Numerical Schemes

Although there exists a unique strong solution to (1.1), an analytical form of the solution is not known. Since the distribution of the CIR process is known, exact simulation methods are available and they involve sampling from a non-central χ^2-distribution. As outlined by Shao [22], directly sampling from a non-central χ^2-distribution on a discrete-time grid when $4a/\sigma^2 < 1$ can be computationally expensive, since it requires sampling a Poisson random variable at each time-point. In practice, σ is often calibrated to be such that this condition does not always hold, which means applying the traditional exact simulation method is not always the best practical option. Shao [22] proposed an exact and relatively fast method for sampling from a non-central χ^2-distribution, which involves sampling a central χ^2 random variable, a uniformly distributed random variable on $[0, 1]$ and 2 standard normal random variables, all of which are independent. The algorithm avoids having to simulate a Poisson random variable and significantly improves computational efficiency. Such an exact simulation method can be useful when the CIR process is not being used as part of another more complex diffusion process, or if sampling on a discrete-time grid with many time points isn't necessary. One particular example of where the CIR process is used to model part of another diffusion process is in the Heston model, where the CIR process is used to model the volatility of a stock's price, which itself follows a geometric Brownian motion. In this model, the Wiener processes used for the volatility process and asset price are correlated according to some correlation coefficient $\rho \in [-1, 1]$. If one were to use exact simulation methods under this model, sampling exact values for the CIR process would require sampling values conditional on the correlated Wiener process, which is not an easy task. Although an algorithm for the exact simulation under the Heston model was recently proposed by Malham et. al. [19], there are still many cases where CIR process is used to model parts of higher dimensional and more complex diffusion processes. Exact simulation still remains a challenge in such cases due to the conditional sampling required. Instead, numerical approximations for the CIR process are often preferred since they are generally faster, easier to implement and can provide a sufficient level of accuracy. Therefore, it is good to have such approximation schemes at hand and it would be useful to know which schemes perform better than others under different parameter settings. For the remainder of this paper, we focus on some of the available numerical approximation schemes for simulating the CIR process.

As mentioned in Section 2.1, the diffusion coefficient in (1.1) fails to be Lipschitz in the neighbourhood of 0, which means that standard results on the rate of strong convergence of the corresponding Euler-Maruyama scheme do not apply. In fact, the standard Euler scheme is not well-defined due to the positive probability of the scheme taking a negative value and becoming undefined in the square root. However, Gyöngy et al. [11] have proven convergence rates for Euler approximations to SDEs with $(1/2 + \alpha)$-Hölder continuous diffusion coefficients in general, for $\alpha \geq 0$. The CIR process corresponds to the case where $\alpha = 0$ and the convergence rate proven for this particular case is $1/\log(N)$, which is considered to be quite a slow rate of convergence. In a similar way, the corresponding standard Milstein scheme is only well-defined for $k \leq 0$ and $\sigma^2 \leq 4a$ and standard results on the rate of strong convergence for this scheme do not apply. The schemes explored in this section can be seen as modifications or corrections to the Euler-Maruyama and Milstein schemes. For context, we give the definition of strong and weak convergence below (sourced from [3]).

Definition 3.1. *An approximation scheme $\hat{X}^{x,N} = (\hat{X}^{x,N}_t)_{0\leq t\leq T}$ for the process $X^x = (X^x_t)_{0\leq t\leq T}$ is said to converge strongly in $\mathcal{L}^p$ with order $\lambda > 0$ if there exists a constant $C > 0$, such that for all $N \in \mathbb{N}$,*

$$\sup_{0\leq t\leq T} \left(\mathbb{E}\Big[\big|\hat{X}^{x,N}_t - X^x_t\big|^p\Big]\right)^{1/p} \leq C \cdot N^{-\lambda}.$$

On the other hand, $\hat{X}^{x,N}$ is said to converge weakly with order $\lambda > 0$ if for any $\mathcal{C}^\infty$ function $f : \mathbb{R} \to \mathbb{R}$ with compact support and for all $N \in \mathbb{N}$, there exists a constant $C > 0$ such that

$$\big|\mathbb{E}\big[f(\hat{X}^{x,N}_T)\big] - \mathbb{E}\big[f(X^x_T)\big]\big| \leq C \cdot N^{-\lambda}.$$

Remark 3.1. *The weak order of convergence is equal to or higher than the strong order of convergence.*

This follows from the fact that f is Lipschitz, i.e. from Jensen's inequality and Hölder's inequality we have

$$\left|\mathbb{E}\big[f(\hat{X}_T^{x,N})\big] - \mathbb{E}\big[f(X_T^x)\big]\right| \leq \mathbb{E}\Big[\left|f(\hat{X}_T^{x,N}) - f(X_T^x)\right|\Big] \leq C\mathbb{E}\Big[\left|\hat{X}_T^{x,N} - X_T^x\right|\Big] \leq C\Big(\mathbb{E}\Big[\left|\hat{X}_T^{x,N} - X_T^x\right|^p\Big]\Big)^{1/p},$$

for Lipschitz constant $C > 0$ and $1 \leq p < \infty$.

Finally for notation purposes, we set

$$\delta_0 = \frac{2a}{\sigma^2} \tag{3.1}$$

$$\text{and} \quad \delta_1 = 2\delta_0, \tag{3.2}$$

and we refer to δ_0 as the Feller ratio.

3.1 Available Approximation Schemes

The approximation schemes mentioned here is a collection of some of the available schemes for simulating the CIR process, which we were particularly interested in studying.

3.1.1 Explicit Euler Schemes

The partial truncation scheme was proposed by Deelstra and Delbaen [9]. It is given by

$$\hat{X}_{t_0}^{x,N} = x, \;\; \hat{X}_{t_{i+1}}^{x,N} = \hat{X}_{t_i}^{x,N} + (a - k\hat{X}_{t_i}^{x,N})\Delta t + \sigma\sqrt{(\hat{X}_{t_i}^{x,N})^+}\Delta W_i,$$

for $i = 0, ..., N-1$. This scheme can take negative values but it is well-defined since the function $x \mapsto (x)^+$ is non-negative for all $x \in \mathbb{R}$. More importantly, it has been shown by Deelstra and Delbaen [9] that this scheme converges strongly with order 1/2 in $\mathcal{L}^1$ under the assumption that $k \geq 0$. If $k < 0$ and the scheme goes negative, it may stay negative. On the other hand, if $k \geq 0$ and if the scheme goes negative, it will be brought back towards the nonnegative values. This behaviour (when $k \geq 0$) is shown by Figure 1. The assumption that $k \geq 0$ is not an overly restrictive condition, since in practice it would generally hold more often than not.

A similar scheme, the full truncation scheme, was proposed by Lord et al. [18]. It was formulated whereby they utilise the following scheme:

$$\tilde{X}_{t_0}^{x,N} = x, \;\; \tilde{X}_{t_{i+1}}^{x,N} = \tilde{X}_{t_i}^{x,N} + (a - k(\tilde{X}_{t_i}^{x,N})^+)\Delta t + \sigma\sqrt{(\tilde{X}_{t_i}^{x,N})^+}\Delta W_i, \tag{3.3}$$

for $i = 0, ..., N-1$, to achieve the full truncation scheme given by $\hat{X}_t^{x,N} = (\tilde{X}_{t_i}^{x,N})^+$ whenever $t \in [t_i, t_{i+1})$ for $i = 0, ..., N-1$. It has been shown by Lord et al. [18] that this scheme converges strongly with order 1/2 in $\mathcal{L}^1$ and with order 1/4 in $\mathcal{L}^2$, while Cozma et al. [8] have shown that it converges in $\mathcal{L}^p$ with order 1/2 when $\delta_0 > 3$. The assumption that the Feller ratio is strictly greater than 3 is considered to be quite a restrictive condition on the parameters. Unlike the partial truncation scheme, the full truncation scheme preserves the non-negativity property of the CIR process. However, if the full truncation scheme solution is equal to 0, it may stay zero for quite some time. This is the case when the scheme given by (3.3) stays negative. This kind of behaviour does not mimic the true behaviour of the CIR process when it is near 0, since the true solution would be brought back into the positive values instantaneously whenever 0 is attained. The behaviour of the full truncation scheme near 0 is shown by Figure 1.

Higham and Mao [15] introduced a similar correction as that used in the partial truncation scheme, where instead they take the absolute value within the square root term in order to keep it defined.

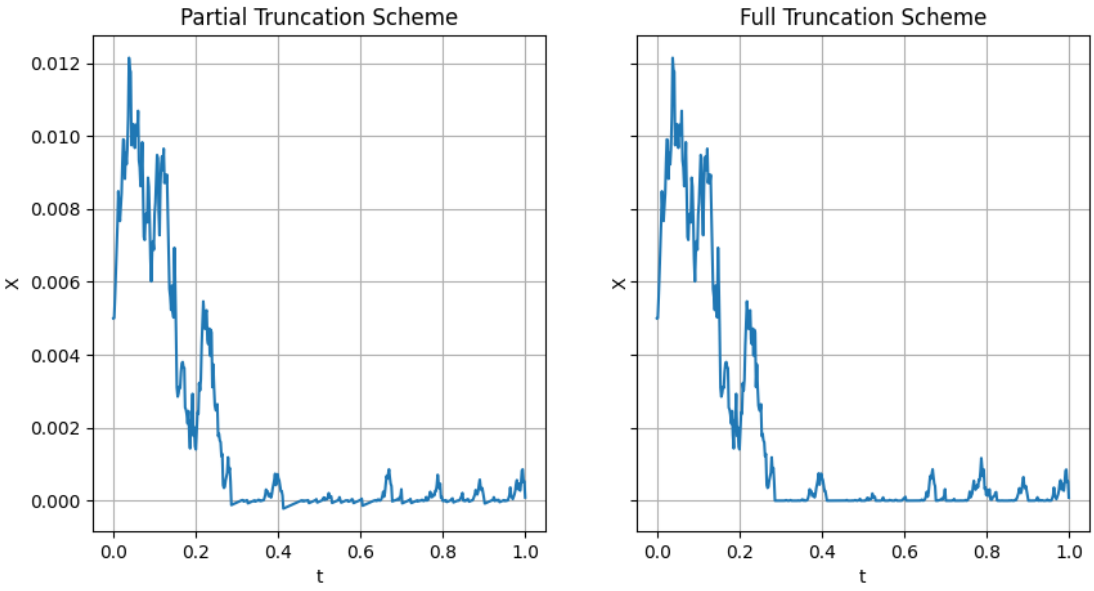

Figure 1: Plots showing the behaviour of the truncation schemes (driven by the same Wiener process) when their solutions are near 0 and when σ is relatively large ($\delta_0 = 0.2$).

They proposed the partial reflection scheme and it is given by

$$\hat{X}_{t_0}^{x,N} = x, \quad \hat{X}_{t_{i+1}}^{x,N} = \hat{X}_{t_i}^{x,N} + (a - k\hat{X}_{t_i}^{x,N})\Delta t + \sigma\sqrt{|\hat{X}_{t_i}^{x,N}|}\Delta W_i,$$

for $i = 0, ..., N-1$. This scheme appears to be similar to the partial truncation scheme as it can also take negative values and still be well-defined. However, when the two schemes are negative, the partial reflection scheme produces a noise whereas the partial truncation scheme does not. As a result, the partial reflection scheme is not necessarily brought back towards the non-negative values straight away and can remain negative for some time, as shown by Figure 2. This is not a desirable property of numerical schemes for the CIR process. This scheme has been shown by Higham and Mao [15] to converge strongly in $\mathcal{L}^1$ and $\mathcal{L}^2$ with order 1/2 .

Berkaoui et al. [5] proposed a scheme that employs a similar correction as that used in the partial reflection scheme, whereby the approximated solution is 'reflected' off the axis when it is about to be overshot into the negative values. They proposed the full reflection (or just reflection) scheme and it is given by

$$\hat{X}_{t_0}^{x,N} = x, \quad \hat{X}_{t_{i+1}}^{x,N} = \left|\hat{X}_{t_i}^{x,N} + (a - k\hat{X}_{t_i}^{x,N})\Delta t + \sigma\sqrt{\hat{X}_{t_i}^{x,N}}\Delta W_i\right|,$$

for $i = 0, ..., N-1$. Berkaoui et al. [5] have shown that this scheme converges strongly in $\mathcal{L}^p$ with order $1/2$ under some restrictive assumptions on the parameters. The behaviour of this scheme near 0 is shown by Figure 2. Clearly, this scheme preserves the non-negativity property of the CIR process. In fact, it is extremely likely that this scheme will remain strictly positive.

Remark 3.2. *For the explicit schemes mentioned thus far in this section, the result obtained by Gyöngy et al.* [11] *for the strong convergence rate of* $1/\log(N)$ *in* $\mathcal{L}^p$ *can be applied. This is due to the fact that the solution* X^x *to (1.1) is non-negative for all* $t \geq 0$*, as is shown in Section 2.1. Therefore, replacing* X^x *with* $(X^x)^+$ *or* $|X^x|$ *in (1.1) still yields the same solution.*

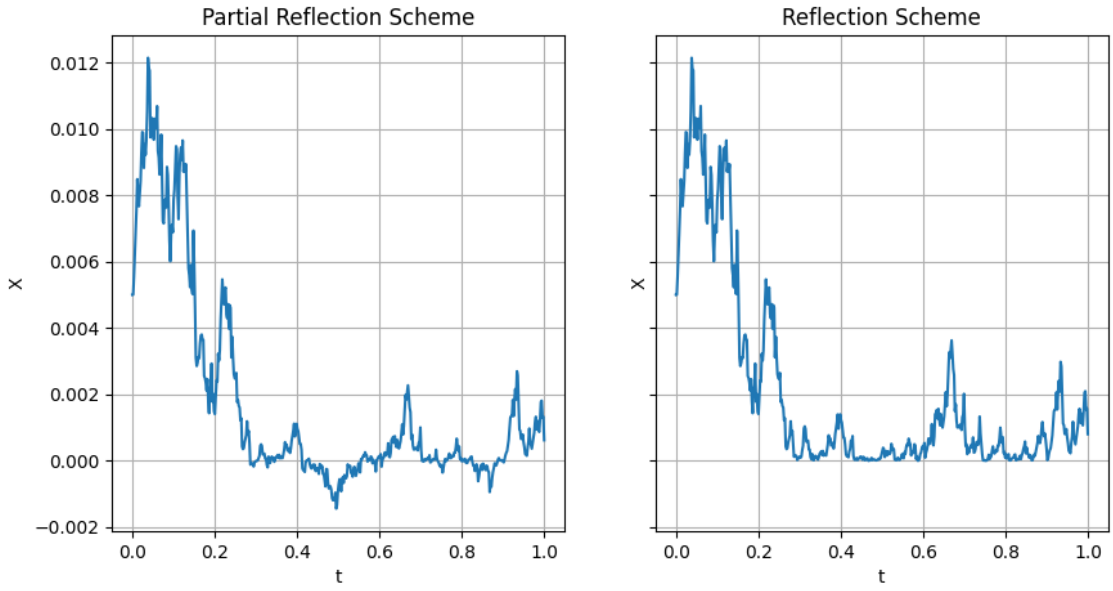

Figure 2: Plots showing the behaviour of the reflection schemes (driven by the same Wiener process) when their solutions are near 0 and when σ is relatively large ($\delta_0 = 0.2$).

3.1.2 Implicit Euler Schemes

The implicit Euler scheme is often used to approximate ordinary differential equations. It is also known as the backward Euler method. When applying the implicit Euler scheme to SDEs, it is necessary to include a quadratic variation term to compensate for the stochastic integral. Using that $\langle \mathrm{d}\sqrt{X_t^x}, \mathrm{d}W_t \rangle = \frac{\sigma}{2}\Delta t$, Brigo and Alfonsi [6] proposed the following implicit scheme:

$$\hat{X}_{t_0}^{x,N} = x, \quad \hat{X}_{t_{i+1}}^{x,N} = \hat{X}_{t_i}^{x,N} + (a - k\hat{X}_{t_{i+1}}^{x,N} - \frac{\sigma^2}{2})\Delta t + \sigma\sqrt{\hat{X}_{t_{i+1}}^{x,N}}\Delta W_i,$$

for $i = 0, ..., N-1$. It is clear that $\sqrt{\hat{X}_{t_{i+1}}^{x,N}}$ is the root of the second degree polynomial $\hat{X}_{t_{i+1}}^{x,N}(1+k\Delta t) - \sqrt{\hat{X}_{t_{i+1}}^{x,N}}(\sigma\Delta W_i) - \left(\hat{X}_{t_i}^{x,N} + (a - \sigma^2/2)\right) = 0$, and by applying Descartes' rule of signs, we see that there is only one positive root when $\hat{X}_{t_i}^{x,N} \leq 0$, $1 + k\Delta t > 0$ and $\delta_0 > 1$. This then gives

$$\hat{X}_{t_{i+1}}^{x,N} = \left(\frac{\sigma\Delta W_i + \sqrt{(\sigma\Delta W_i)^2 + 4(\hat{X}_{t_i}^{x,N} + (a - \sigma^2/2)\Delta t)(1 + k\Delta t)}}{2(1 + k\Delta t)} \right)^2. \tag{3.4}$$

This scheme has been shown by Alfonsi [4] to converge strongly in $\mathcal{L}^1$ at a rate of $1/\log(N)$.

By applying Itô's formula to the function $x \mapsto \sqrt{x}$ and the CIR process, we get

$$\mathrm{d}\sqrt{X_t^x} = \left(\frac{a - \sigma^2/4}{2\sqrt{X_t^x}} - \frac{k}{2}\sqrt{X_t^x}\right)\mathrm{d}t + \frac{\sigma}{2}\mathrm{d}W_t.$$

Using this as a tool for inspiration, Alfonsi [4] proposed another implicit scheme given by

$$\sqrt{\hat{X}_{t_0}^{x,N}} = x, \quad \sqrt{\hat{X}_{t_{i+1}}^{x,N}} = \sqrt{\hat{X}_{t_i}^{x,N}} + \left(\frac{a - \sigma^2/4}{2\sqrt{\hat{X}_{t_{i+1}}^{x,N}}} - \frac{k}{2}\sqrt{\hat{X}_{t_{i+1}}^{x,N}}\right)\Delta t + \frac{\sigma}{2}\Delta W_i,$$

for $i = 0, ..., N-1$. By multiplying this equation by $\sqrt{\hat{X}_{t_{i+1}}^{x,N}}$, we get another second degree polynomial

that has only one positive root if $\hat{X}_{t_i} \geq 0$, $\delta_1 > 1$ and $1 + \frac{k}{2}\Delta t > 0$. Similar to the previous scheme, we get

$$\hat{X}^{x,N}_{t_{i+1}} = \left(\frac{\frac{\sigma}{2}\Delta W_i + \sqrt{\hat{X}^{x,N}_{t_i}} + \sqrt{(\frac{\sigma}{2}\Delta W_i + \sqrt{\hat{X}^{x,N}_{t_i}})^2 + 4(\frac{a-\sigma^2/4}{2})(1+\frac{k}{2}\Delta t)\Delta t}}{2(1+\frac{k}{2}\Delta t)} \right)^2. \tag{3.5}$$

When $\delta_0 > 1$, Dereich et al. [10] have shown it converges strongly in $\mathcal{L}^p$ with order 1/2 up to a log term for $1 \leq p < \delta_0$. When $\delta_0 > 2$, Alfonsi [2] has shown that it converges with order 1 in $\mathcal{L}^p$ for $1 \leq p < \frac{\delta_1}{3}$, while Neuenkirch and Szpruch [21] have shown strong convergence of order 1 in $\mathcal{L}^p$ for all $p \geq 1$ when $\delta_0 > 3$. Again, these assumptions on δ_0 are considered to be quite restrictive.

3.1.3 Milstein Schemes

A modification to the standard Milstein scheme was proposed by Alfonsi [4]. It is an explicit scheme and its formulation is motivated by the implicit scheme (3.5). It is given by

$$\hat{X}^{x,N}_{t_0} = x, \;\; \hat{X}^{x,N}_{t_{i+1}} = \left((1 - \frac{k}{2}\Delta t)\sqrt{\hat{X}^{x,N}_{t_i}} + \frac{\sigma \Delta W_i}{2(1-\frac{k}{2}\Delta t)}\right)^2 + (a - \frac{\sigma^2}{4})\Delta t, \tag{3.6}$$

for $i = 0, ..., N-1$. This scheme is well defined when $\delta_1 \geq 1$ and $\Delta t < \frac{2}{k}$. Under this setting, it has been shown by Alfonsi [4] to converge strongly in $\mathcal{L}^1$ at a rate of $1/\log(N)$, although faster rates of convergence are empirically observed in [4]. Notice that when $k = 0$, this scheme is exactly the same as the standard Milstein scheme. When $\delta_1 < 1$, one can correct the scheme by applying a similar truncation technique to that used with the full truncation scheme to keep it defined and preserve its non-negativity, i.e. when $\delta_1 < 1$, the scheme becomes

$$\hat{X}^{x,N}_{t_0} = x, \;\; \hat{X}^{x,N}_{t_{i+1}} = \left(\left((1 - \frac{k}{2}\Delta t)\sqrt{(\hat{X}^{x,N}_{t_i})^+} + \frac{\sigma \Delta W_i}{2(1-\frac{k}{2}\Delta t)}\right)^2 + (a - \frac{\sigma^2}{4})\Delta t\right)^+. \tag{3.7}$$

Although there is no theoretical result (to the best of our knowledge) on the strong convergence of scheme (3.7), we still look at its potential convergence properties in Section 4.

More recently, Hefter and Herzwurm [12] have proposed another Milstein-type scheme, which they refer to as the truncated Milstein scheme. The scheme is given by

$$\begin{aligned} &\hat{X}^{x,N}_{t_0} = x, \\ &\hat{X}^{x,N}_{t_{i+1}} = \left(\left(\max\left(\sqrt{\frac{\sigma^2}{4}\cdot\Delta t}, \sqrt{\max(\frac{\sigma^2}{4}\cdot\Delta t, \hat{X}^{x,N}_{t_i})} + \frac{\sigma}{2}\cdot\Delta W_i\right)\right)^2 + (a - \frac{\sigma^2}{4} - k\hat{X}^{x,N}_{t_i})\cdot\Delta t\right)^+, \end{aligned} \tag{3.8}$$

for $i = 0, 1, ..., N-1$. This scheme clearly preserves the non-negativity property of the CIR process and it is well-defined for the full parameter range thanks to the truncation that occurs when the solution is close to 0. When the solution is far from 0 and when Δt is small enough, we see that it coincides with the standard Milstein scheme. In their work, Hefter and Herzwurm [12] show that this scheme converges strongly for the full parameter range at a rate that downgrades as the Feller ratio decreases from 0.5. This result is rather useful since it gives a convergence rate potentially better than the rate of $1/\log(N)$ for all Feller ratios (see Remark 3.3). However, the convergence rate does becomes arbitrarily slow as the Feller ratio approaches 0 since it is monotonically decreasing in $\frac{1}{\delta_1}$.

3.2 Strong Convergence of the Truncated Milstein Scheme

In this section, we study the theoretical strong convergence of the truncated Milstein scheme (3.8) in detail. The following result for the strong convergence in $\mathcal{L}^p$ was obtained by Hefter and Herzwurm [12] for the full parameter range.

Theorem 3.1. *Let $\delta_1 > 0$ be defined according to (3.2) and let $\hat{X}^{x,N}$ be the approximation scheme defined by (3.8). For every $1 \leq p < \infty$ and every $\epsilon > 0$ there exists a constant $C > 0$ such that*

$$\sup_{0\leq t\leq T}\left(\mathbb{E}\left[\left|X_t^x - \hat{X}_t^{x,N}\right|^p\right]\right)^{1/p} \leq C(1+x)\cdot\frac{1}{N^{\min(1,\delta_1)/(2p)-\epsilon}} \tag{3.9}$$

for all $N \in \mathbb{N}$ and for all $x \geq 0$.

Remark 3.3. *We can see that the convergence rate given in Theorem 3.1 is faster than $1/\log(N)$ under the assumption that $\min(1,\delta_1)/(2p) - \epsilon > 0$. indeed, we first notice that the functions $x \mapsto \log(x)$ and $x \mapsto x^\alpha$ are monotonically increasing for all $x \geq 1$ and some $\alpha > 0$. Then by L'Hôpital's rule, we have $\lim_{x\to+\infty} \frac{\log(x)}{x^\alpha} = \lim_{x\to+\infty} \frac{1}{\alpha}x^{-\alpha} = 0$. Hence, there exists an M such that $\log(m) \leq m^\alpha$ for all $m \geq M$. Since $N \geq 1$, we can set $C > 0$ such that $\log(N) \leq CN^\alpha$ and crucially, C only depends on α. Hence, we can see the result by setting $\alpha = \min(1,\delta_1)/(2p) - \epsilon$.*

We now follow the approach in [12] and give the proof of Theorem 3.1. We begin by introducing the setting in which we will work. Firstly, we focus only the case where $\sigma = 2$ and $T = 1$. This gives no loss of generality since firstly, the SDE given by

$$\mathrm{d}X_t^{x'} = (\delta_1 - kX_t^{x'})\mathrm{d}t + 2\sqrt{X_t^{x'}}\mathrm{d}W_t, \quad X_0^{x'} = x', t \geq 0,$$

with $x' = 4x/\sigma^2$, multiplied by $\sigma^2/4$ brings us back to (1.1). Secondly, we can always retain (1.1) through time scaling which allows us to restrict the analysis to $T = 1$. For example, consider the SDE

$$\mathrm{d}\tilde{X}_t^x = (\tilde{a} - \tilde{k}\tilde{X}_t^x)\mathrm{d}t + \tilde{\sigma}\sqrt{\tilde{X}_t^x}\mathrm{d}\tilde{W}_t, \quad \tilde{X}_0^x = x, \;\; t \geq 0,$$

with $\tilde{a} = aT$, $\tilde{k} = kT$, $\tilde{\sigma} = \sigma\sqrt{T}$ and a Wiener process $(\tilde{W}_t)_{t\geq 0}$ given by $\tilde{W}_t = 1/\sqrt{T}\cdot W_{tT}$. Then by the uniqueness of the CIR process, we have that $\tilde{X}_{t/T}^x$ solves (1.1). Therefore, for the rest of this section we only consider the SDE

$$\mathrm{d}X_t^x = (\delta_1 - kX_t^x)\mathrm{d}t + 2\sqrt{X_t^x}\mathrm{d}W_t, \quad X_0^x = x, \;\; 0 \leq t \leq 1. \tag{3.10}$$

We use the notion of a one-step approximation scheme for (3.10), which is defined to be a sequence of approximating processes $\hat{X}^{x,N,\Theta} = (\hat{X}_t^{x,N,\Theta})_{t\geq 0}$ that is itself defined by a Borel-measurable function $\Theta : \mathbb{R}_0^+ \times (0,1] \times \mathbb{R} \to \mathbb{R}_0^+$ such that

$$\hat{X}_0^{x,N,\Theta} = x, \quad \hat{X}_{(i+1)/N}^{x,N,\Theta} = \Theta(\hat{X}_{i/N}^{x,N,\Theta}, 1/N, \Delta W_i^N), \tag{3.11}$$

where $\Delta W_i^N = W_{(i+1)/N} - W_{i/N}$ for $i = 0, 1, ..., N-1$. Constant interpolation is used between grid points to achieve the continuous time solution $\hat{X}_t^{x,N,\Theta}$ for $t \in (0,1]$, i.e. $\hat{X}_t^{x,N,\Theta} = \hat{X}_{i/N}^{x,N,\Theta}$ for $t \in [i/N, (i+1)/N)$.

Remark 3.4. *In the case of the truncated Milstein scheme given by (3.8), we have*

$$\Theta_{MIL}(x,h,w) = \left(\left(\max\left(\sqrt{\frac{\sigma^2}{4}\cdot h}, \sqrt{\max(\frac{\sigma^2}{4}\cdot h, x)} + \frac{\sigma}{2}\cdot w\right)\right)^2 + (a - \frac{\sigma^2}{4} - kx)\cdot h\right)^+. \tag{3.12}$$

We use $\Delta_{local} : \mathbb{R}_0^+ \times (0,1] \to \mathbb{R}_0^+$ defined by

$$\Delta_{local}(x,h) = \begin{cases} h & \text{if } \; x \leq h, \\ x^{-1/2}h^{3/2} & \text{if } \; h \leq x \leq 1, \\ xh^{3/2} & \text{if } \; x \geq 1, \end{cases} \tag{3.13}$$

and we consider the following conditions for $p, q \in [1, \infty)$.

Condition 3.1. *There exists a constant $K > 0$ such that*

$$\left(\mathbb{E}\left[\left|\Theta(x_1, h, W_h) - \Theta(x_2, h, W_h)\right|^p\right]\right)^{1/p} \leq (1 + Kh)|x_1 - x_2|,$$

for all $x_1, x_2 \geq 0$ and $h \in (0, 1]$.

Condition 3.2. *There exists a constant $C > 0$ such that*

$$\left(\mathbb{E}\left[\left|\Theta(x, h, W_h) - X_h^x\right|^p\right]\right)^{1/p} \leq C\Delta_{local}(x, h),$$

for all $x \geq 0$ and $h \in (0, 1]$.

Condition 3.3. *There exists a constant $C > 0$ such that*

$$\sup_{0\leq t\leq 1}\left(\mathbb{E}\left[\left|\hat{X}_t^{x,N,\Theta}\right|^q\right]\right)^{1/q} \leq C(1 + x),$$

for all $x \geq 0$ and $N \in \mathbb{N}$.

Remark 3.5. *Condition 3.1 ensures that the one step scheme is $\mathcal{L}^p$-Lipschitz continuous, Condition 3.2 ensures that a sufficiently small local-discretisation error is obtained for a single step of the one-step scheme and Condition 3.3 ensures that the one-step scheme has uniformly bounded absolute moments.*

The first step in the proof of Theorem 3.1 involves showing that any one-step scheme that satisfies Conditions 3.1, 3.2 and 3.3 converges strongly at a rate in $\mathcal{L}^p$. It is first shown for the case $p = 1$ and this is then used to prove the result in $\mathcal{L}^p$. The final step of the proof then involves showing that the one-step scheme given by (3.12) is a suitable one-step scheme that satisfies all three conditions. To begin, we have the following theorem for the $\mathcal{L}_1$-convergence of one step schemes satisfying Conditions 3.1 and 3.2 for $p = 1$.

Theorem 3.2. *Let $\hat{X}^{x,N,\Theta}$ be a one-step scheme given by (3.11) and assume that Condition 3.1 and 3.2 hold for $p = 1$. Then there exists a constant $C > 0$ such that*

$$\sup_{0\leq t\leq 1}\mathbb{E}\left[\left|X_t^x - \hat{X}_t^{x,N,\Theta}\right|\right] \leq C(1 + x)\frac{1 + \mathbb{1}_{\{\delta_1=1\}} \cdot \log(N)}{N^{\min(\delta_1,1)/2}}$$

for all $N \in \mathbb{N}$ and $x \geq 0$.

For the proof of Theorem 3.2, we require the following result for the expected local error. We also remark that for the proofs in this section, the constant $C > 0$ is chosen suitably from line to line and it will be clear from the underlying proof to see what C is allowed to depend on.

Proposition 3. *There exists a constant $C > 0$ such that*

$$\mathbb{E}\left[\Delta_{local}(X_t^x, h)\right] \leq Ch(1 + x) \cdot \begin{cases} 1, & \text{if } \ t \leq h, \\ (h/t)^{\min(\delta_1,1)/2}(1 + \mathbb{1}_{\{\delta_1=1\}} \cdot \log(t/h)), & \text{if } \ h \leq t, \end{cases}$$

for all $t \in (0, 1], s \in (0, 1]$ and $x \geq 0$.

The proof of Proposition 3 is postponed until Appendix C.

Proof of Theorem 3.2. We have

$$\begin{aligned}
\mathbb{E}\Big[\big|X^x_{(i+1)/N} - \hat{X}^{x,N,\Theta}_{(i+1)/N}\big|\Big] &= \mathbb{E}\Big[\big|X^x_{(i+1)/N} - \Theta(X^x_{(i+1)/N}, 1/N, \Delta W^N_{(i+1)/N}) + \Theta(X^x_{(i+1)/N}, 1/N, \Delta W^N_{i+1}) \\
&\qquad - \hat{X}^{x,N,\Theta}_{(i+1)/N}\big|\Big] \\
&\leq \mathbb{E}\Big[\big|X^x_{(i+1)/N} - \Theta(X^x_{(i+1)/N}, 1/N, \Delta W^N_{(i+1)/N})\big|\Big] \\
&\quad + \mathbb{E}\Big[\big|\Theta(X^x_{(i+1)/N}, 1/N, \Delta W^N_{i+1}) - \hat{X}^{x,N,\Theta}_{(i+1)/N}\big|\Big] \quad \text{from Minkowski's inequality,} \\
&= \mathbb{E}\Bigg[\mathbb{E}\Big[\big|X^{x'}_{1/N} - \Theta(X^{x'}_{1/N}, 1/N, W_{1/N})\big|\Big]\Bigg| x' = X^x_{i/N}\Bigg] \\
&\quad + \mathbb{E}\Bigg[\mathbb{E}\Big[\big|\Theta(x', 1/N, W_{1/N}) - \Theta(\tilde{x}, 1/N, W_{1/N})\big|\Big]\Bigg| x' = X^x_{i/N}, \tilde{x} = \hat{X}^{x,N,\Theta}_{i/N}\Bigg] \\
&\leq C\mathbb{E}\Big[\Delta_{local}(X^x_{i/N}, 1/N)\Big] + (1 + K/N)\mathbb{E}\Big[\big|X^x_{i/N} - \hat{X}^{x,N,\Theta}_{i/N}\big|\Big] \\
&\qquad \text{using Conditions 3.1 and 3.2,}
\end{aligned}$$

for $i = 0, 1, ..., N-1$. Applying Proposition 3 gives

$$\begin{aligned}
\mathbb{E}\Big[\big|X^x_{(i+1)/N} - \hat{X}^{x,N,\Theta}_{(i+1)/N}\big|\Big] \leq &(1 + K/N)\mathbb{E}\Big[\big|X^x_{i/N} - \hat{X}^{x,N,\Theta}_{i/N}\big|\Big] \\
&+ \frac{C}{N}(1 + x) \cdot \begin{cases} 1, & \text{if } i = 0, \\ (1/i)^{\min(\delta_1,1)/2}(1 + \mathbb{1}_{\{\delta_1=1\}} \cdot \log(i)), & \text{if } i \geq 1, \end{cases}
\end{aligned}$$

and using that $2(\frac{1}{i+1})^y \geq (\frac{1}{i})^y$ for all $0 < y \leq 1/2$ and $i \geq \frac{1}{2^{1/y}-1} > 0$, we get

$$\mathbb{E}\Big[\big|X^x_{(i+1)/N} - \hat{X}^{x,N,\Theta}_{(i+1)/N}\big|\Big] \leq (1 + K/N)\mathbb{E}\Big[\big|X^x_{i/N} - \hat{X}^{x,N,\Theta}_{i/N}\big|\Big] + \frac{C}{N}(1 + x) \cdot \frac{1 + \log(N) \cdot \mathbb{1}_{\{\delta_1=1\}}}{(i+1)^{\min(\delta_1,1)/2}}.$$

Using this recursive relationship, we get the general expression

$$\mathbb{E}\Big[\big|X^x_{i/N} - \hat{X}^{x,N,\Theta}_{i/N}\big|\Big] \leq \frac{C}{N}(1 + x)(1 + \log(N) \cdot \mathbb{1}_{\{\delta_1=1\}}) \sum_{j=1}^{i} \frac{(1 + K/N)^{i-j}}{j^{\min(\delta_1,1)/2}}.$$

Now, we notice that $(1 + K/N)^{N-j} \leq (1 + K/N)^N \leq e^K$ for $j, K, N \geq 0$. This gives

$$\begin{aligned}
\mathbb{E}\Big[\big|X^x_{i/N} - \hat{X}^{x,N,\Theta}_{i/N}\big|\Big] &\leq C(1 + x)\frac{1 + \log(N) \cdot \mathbb{1}_{\{\delta_1=1\}}}{N} \sum_{j=1}^{N} \frac{1}{j^{\min(\delta_1,1)/2}} \\
&\leq C(1 + x)\frac{1 + \log(N) \cdot \mathbb{1}_{\{\delta_1=1\}}}{N} \int_0^N \frac{1}{j^{\min(\delta_1,1)/2}} \mathrm{d}j,
\end{aligned}$$

since for any decreasing function f, we have $\sum_{n=a}^{b} f(n) \leq \int_{a-1}^{b} f(n)\mathrm{d}n$.

Hence,

$$\begin{aligned}
\mathbb{E}\Big[\big|X^x_{i/N} - \hat{X}^{x,N,\Theta}_{i/N}\big|\Big] &\leq C(1 + x)\frac{1 + \log(N) \cdot \mathbb{1}_{\{\delta_1=1\}}}{N} \frac{N^{1-\min(\delta_1,1)/2}}{1 - \min(\delta_1, 1)/2} \\
&\leq C(1 + x)\frac{1 + \log(N) \cdot \mathbb{1}_{\{\delta_1=1\}}}{N^{\min(\delta_1,1)}},
\end{aligned}$$

since $\frac{1}{1-\min(\delta_1,1)/2} \leq 2$. This implies

$$\max_{i=0,1,\dots,N} \mathbb{E}\Big[\big|X^x_{i/N} - \hat{X}^{x,N,\Theta}_{i/N}\big|\Big] \leq C(1+x)\frac{1+\log(N)\cdot \mathbb{1}_{\{\delta_1=1\}}}{N^{\min(\delta_1,1)}},$$

for all $N \in \mathbb{N}$ and $x \geq 0$. Since the coefficients of (3.10) satisfy the linear growth condition (see Remark 2.1), there exists some constant $C > 0$ such that

$$\mathbb{E}\Big[\big|X^x_t - X^x_s\big|\Big] \leq C(1+x)\sqrt{|t-s|},$$

for all $x \geq 0$ and $t, s \in [0,1]$ (see [20], Chapter 2, Theorem 4.3). Using this result and constant interpolation gives

$$\begin{aligned}
\sup_{0\leq t\leq 1} \mathbb{E}\Big[\big|X^x_t - \hat{X}^{x,N,\Theta}_t\big|\Big] &= \sup_{0\leq t\leq 1} \mathbb{E}\Big[\big|X^x_t - \hat{X}^{x,N,\Theta}_{i_t/N}\big|\Big] \\
&= \sup_{0\leq t\leq 1} \mathbb{E}\Big[\big|X^x_t - X^x_{i_t/N} + X^x_{i_t/N} - \hat{X}^{x,N,\Theta}_{i_t/N}\big|\Big] \\
&\leq \sup_{0\leq t\leq 1} \mathbb{E}\Big[\big|X^x_t - X^x_{i_t/N}\big|\Big] + \max_{i=0,1,\dots,N} \mathbb{E}\Big[\big|X^x_{i/N} - \hat{X}^{x,N,\Theta}_{i/N}\big|\Big] \\
&\leq C(1+x)\frac{1+\log(N)\cdot \mathbb{1}_{\{\delta_1=1\}}}{N^{\min(\delta_1,1)}},
\end{aligned}$$

where i_t is such that $t \in [i_t/N, (i_t+1)/N)$. This completes the proof. □

Using Theorem 3.2 under the additional assumption that Condition 3.3 holds, the strong convergence of any one-step scheme is achieved for arbitrary $p \geq 1$ in the Corollary below.

Corollary 1. *Consider the setting of Theorem 3.2 and also assume that Condition 3.3 is satisfied. Let $1 \leq p \leq \infty$ and $\epsilon > 0$. Then there exists a constant $C > 0$ such that*

$$\sup_{0\leq t\leq 1} \left(\mathbb{E}\Big[\big|X^x_t - \hat{X}^{x,N,\Theta}_t\big|^p\Big]\right)^{1/p} \leq C(1+x)\frac{1}{N^{\min(\delta_1,1)/(2p)-\epsilon}},$$

for all $N \in \mathbb{N}$ and $x \geq 0$.

Proof. Let $\epsilon < 1/p$ and set $q = 1 + \frac{1-1/p}{\epsilon}$. Then $p \leq q < \infty$ and by choosing $\hat{p} = \frac{1}{\epsilon p}$ and $\hat{q} = \frac{1}{p(1/p-\epsilon)}$, Hölder's inequality yields

$$\begin{aligned}
\left(\mathbb{E}\Big[\big|X^x_t - \hat{X}^{x,N,\Theta}_t\big|^p\Big]\right)^{1/p} &= \left(\mathbb{E}\Big[\big|X^x_t - \hat{X}^{x,N,\Theta}_t\big|^{q(\epsilon p)} \cdot \big|X^x_t - \hat{X}^{x,N,\Theta}_t\big|^{p(1/p-\epsilon)}\Big]\right)^{1/p} \\
&\leq \left(\left(\mathbb{E}\Big[\big|X^x_t - \hat{X}^{x,N,\Theta}_t\big|^{qp\epsilon\cdot\hat{p}}\Big]\right)^{1/\hat{p}} \left(\mathbb{E}\Big[\big|X^x_t - \hat{X}^{x,N,\Theta}_t\big|^{p(1/p-\epsilon)\cdot\hat{q}}\Big]\right)^{1/\hat{q}}\right)^{1/p} \\
&= \left(\mathbb{E}\Big[\big|X^x_t - \hat{X}^{x,N,\Theta}_t\big|^q\Big]\right)^{\epsilon} \left(\mathbb{E}\big|X^x_t - \hat{X}^{x,N,\Theta}_t\big|\right)^{1/p-\epsilon},
\end{aligned}$$

and from Minkowski's inequality we get

$$\left(\mathbb{E}\Big[\big|X^x_t - \hat{X}^{x,N,\Theta}_t\big|^p\Big]\right)^{1/p} \leq \left(\mathbb{E}\Big[\big|X^x_t - \hat{X}^{x,N,\Theta}_t\big|\Big]\right)^{1/p-\epsilon} \left(\left(\mathbb{E}\Big[\big|X^x_t\big|^q\Big]\right)^{1/q} + \left(\mathbb{E}\Big[\big|\hat{X}^{x,N,\Theta}_t\big|^q\Big]\right)^{1/q}\right)^{q\epsilon}.$$

Hence,

$$\sup_{0\le t\le 1}\left(\mathbb{E}\Big[\big|X_t^x-\hat{X}_t^{x,N,\Theta}\big|^p\Big]\right)^{1/p}\le \sup_{0\le t\le 1}\left(\mathbb{E}\Big[\big|X_t^x-\hat{X}_t^{x,N,\Theta}\big|\Big]\right)^{1/p-\epsilon}\Bigg(\sup_{0\le t\le 1}\left(\mathbb{E}\Big[\big|X_t^x|^q\Big]\right)^{1/q} + \sup_{0\le t\le 1}\left(\mathbb{E}\Big[\big|\hat{X}_t^{x,N,\Theta}\big|^q\Big]\right)^{1/q}\Bigg)^{q\epsilon},$$

and from Theorem 3.2, (2.6) and Condition 3.3, we get

$$\begin{aligned}\sup_{0\le t\le 1}\left(\mathbb{E}\Big[\big|X_t^x-\hat{X}_t^{x,N,\Theta}\big|^p\Big]\right)^{1/p} &\le C\Big((1+x)\frac{1+\log(N)\cdot\mathbb{1}_{\{\delta_1=1\}}}{N^{\min(\delta_1,1)/2}}\Big)^{1/p-\epsilon}(1+x)^{q\epsilon}\\ &\le C(1+x)(1+\log(N)\cdot\mathbb{1}_{\{\delta_1=1\}})^{1/p-\epsilon}\cdot\frac{1}{N^{\min(\delta_1,1)/(2p)-\epsilon/2}},\end{aligned}$$

where we use that $1/p-\epsilon+q\epsilon=1$. For $\delta_1\neq 0$, the claim is trivial. For $\delta_1=1$, we recall from Remark 3.3 that for any $\alpha>0$, there exists a constant $C>0$ depending only on α such that $\log(N)\le CN^\alpha$. Since $N\ge 1$, we then have $1+\log(N)\le (C+1)N^\alpha$ and setting $\alpha=\frac{\epsilon}{2(1/p-\epsilon)}>0$ completes the proof. □

We are now in a position to show that the truncated Milstein scheme given by (3.8) satisfies Conditions 3.1, 3.2 and 3.3. First, we consider the function $\phi_{MIL}(x,h,w):\mathbb{R}_0^+\times(0,1]\times\mathbb{R}\to\mathbb{R}$ given by

$$\phi_{MIL}(x,h,w)=\nu(x,h,w)+(\delta_1-1-kx)\cdot h,$$

where $\nu(x,h,w)=(\sqrt{x}+w)^2$ and we have $\Theta_{MIL}(x,h,w):\mathbb{R}_0^+\times(0,1]\times\mathbb{R}\to\mathbb{R}_0^+$ defined by (3.12) which gives

$$\Theta_{MIL}(x,h,w)=\Big(\hat{\nu}(x,h,w)+(\delta_1-1-kx)\cdot h\Big)^+,$$

where $\hat{\nu}(x,h,w)=\Big(\max(\sqrt{h},\sqrt{\max(h,x)}+w)\Big)^2$ (using the parameters in (3.10)). Positivity is clearly preserved for ϕ_{MIL} only if $k\le 0$ and $\delta_1\ge 1$, so it is not a valid one-step scheme in general. Nevertheless, we have that ϕ_{MIL} satisfies Condition 3.2. In addition, we have that $\hat{\nu}$ is almost surely $\mathcal{L}^1$-Lipschitz continuous with Lipschitz constant 1 and that $\hat{\nu}$ approximates ν well enough. These results are given formally below.

Proposition 4. *ϕ_{MIL} satisfies Condition 3.2 for every $1\le p<\infty$.*

Lemma 3.1. *We have*

$$\mathbb{E}\Big[\big|\hat{\nu}(x_1,h,W_h)-\hat{\nu}(x_2,h,W_h)\big|\Big]\le|x_1-x_2|,$$

for all $x_1,x_2\ge 0$ and $h\in(0,1]$. In addition, for every $1\le p<\infty$ there exists a constant $C>0$ such that

$$\left(\mathbb{E}\Big[|\nu(x,h,W_h)-\hat{\nu}(x,h,W_h)|^p\Big]\right)^{1/p}\le C\Delta_{local}(x,h),$$

for all $x\ge 0$ and $h\in(0,1]$.

The proofs of Proposition 4 and Lemma 3.1 are postponed until Appendix C. We can now give the main result for $\mathcal{L}^1$-convergence of the truncated Milstein scheme by showing that the one-step scheme given by (3.12) satisfies Conditions 3.1 and 3.2.

Theorem 3.3. *Θ_{MIL} satisfies Condition 3.1 for $p=1$ and Condition 3.2 for every $1\le p<\infty$.*

Proof. We know that $|y^+ - z^+| \le |y - z|$ for every $y, z \in \mathbb{R}$. Using this, we get

$$\begin{aligned}
\mathbb{E}\Big[\big|\Theta_{MIL}(x_1,h,W_h) - \Theta_{MIL}(x_2,h,W_h)\big|\Big] &= \mathbb{E}\Big[\Big|\big(\hat{\nu}(x_1,h,w) + (\delta_1 - 1 - kx_1)\cdot h\big)^+ \\
&\qquad - \big(\hat{\nu}(x_2,h,w) + (\delta_1 - 1 - kx_2)\cdot h\big)^+\Big|\Big] \\
&\le \mathbb{E}\Big[\big|\hat{\nu}(x_1,h,w) + (\delta_1 - 1 - bx_1)\cdot h - \hat{\nu}(x_2,h,w) \\
&\qquad - (\delta_1 - 1 - kx_2)\cdot h\big|\Big] \\
&\le \mathbb{E}\Big[\big|\hat{\nu}(x_1,h,w) - \hat{\nu}(x_2,h,w)\big|\Big] \\
&\quad + \mathbb{E}\Big[\big|(\delta_1 - 1 - kx_1)\cdot h - (\delta_1 - 1 - kx_2)\cdot h\big|\Big] \\
&\le |x_1 - x_2| + kh|x_2 - x_1| \quad \text{from Lemma 3.1,} \\
&= (1 + kh)|x_1 - x_2|,
\end{aligned}$$

and Condition 3.1 is satisfied. Now using $|y^+ - z| \le |y - z|$ for $y \in \mathbb{R}$ and $z \ge 0$, we get

$$\begin{aligned}
\Big(\mathbb{E}\Big[\big|\Theta_{MIL}(x,h,W_h) - X_h^x\big|^p\Big]\Big)^{1/p} &\le \Big(\mathbb{E}\Big[\big|\hat{\nu}(x,h,W_h) + (\delta_1 - 1 - kx)\cdot h - X_h^x\big|^p\Big]\Big)^{1/p} \\
&= \Big(\mathbb{E}\Big[\big|\hat{\nu}(x,h,W_h) + (\delta_1 - 1 - kx)\cdot h \\
&\qquad + \nu(x,h,W_h) - \nu(x,h,W_h) - X_h^x\big|^p\Big]\Big)^{1/p} \\
&= \Big(\mathbb{E}\Big[\big|\hat{\nu}(x,h,W_h) + \phi_{MIL}(x,h,W_h) - \nu(x,h,W_h) - X_h^x\big|^p\Big]\Big)^{1/p} \\
&\le \Big(\mathbb{E}\Big[\big|\phi_{MIL}(x,h,W_h) - X_h^x\big|^p\Big]\Big)^{1/p} \\
&\quad + \Big(\mathbb{E}\Big[\big|\hat{\nu}(x,h,W_h) - \nu(x,h,W_h)\big|^p\Big]\Big)^{1/p} \text{ from Minkowski's inequality,} \\
&\le C\Delta_{local}(x,h),
\end{aligned}$$

from Proposition 4 and Lemma 3.1, which completes the proof. □

Finally, we show that Θ_{MIL} satisfies Condition 3.3 which shows the $\mathcal{L}^p$-convergence of the truncated Milstein scheme.

Proposition 5. *$\hat{X}_t^{x,N,\Theta_{MIL}}$ satisfies Condition 3.3 for every $1 \le q < \infty$.*

Proof. Let $x \ge 0, h \in (0,1]$ and $w \in \mathbb{R}$. First, we see that

$$\begin{aligned}
\hat{\nu}(x,h,w) &= \Big(\max(\sqrt{h}, \sqrt{\max(h,x)} + w)\Big)^2 \\
&\le \Big(\max(\sqrt{h}, \sqrt{h} + w)\Big)^2 + \Big(\max(\sqrt{h}, \sqrt{x} + w)\Big)^2 \\
&\le (\sqrt{h} + |w|)^2 + h + (\sqrt{x} + w)^2 \\
&\le 2h + 2w^2 + h + (\sqrt{x} + w)^2 \\
&= 3h + 2w^2 + (\sqrt{x} + w)^2,
\end{aligned}$$

and $(\delta_1 - 1 - kx)\cdot h \le (\delta_1 + |k|x)\cdot h$. Now define the function $g : \mathbb{R}_0^+ \times (0,1] \times \mathbb{R} \to \mathbb{R}_0^+$ by

$$\begin{aligned}
g(x,h,w) &= 3h + 2w^2 + (\sqrt{x} + w)^2 + (\delta_1 + |k|x)\cdot h \\
&= x + (\delta_1 + |k|x + 3)\cdot h + 3w^2 + 2w\sqrt{x}.
\end{aligned}$$

Then g is a one-step scheme and we have for $0 \le x_1 \le x_2$,

$$\begin{aligned}
\Theta_{MIL}(x_1,h,w) &= \Big(\hat{\nu}(x_1,h,w) + (\delta_1 - 1 - kx_1)\cdot h\Big)^+ \\
&\le \Big(\hat{\nu}(x_2,h,w) + (\delta_1 - 1 - kx_1)\cdot h\Big)^+ \\
&\le \Big(\hat{\nu}(x_2,h,w) + (\delta_1 + |k|x)\cdot h\Big)^+ \\
&= \hat{\nu}(x_2,h,w) + (\delta_1 + |k|x)\cdot h \\
&\le g(x_2,h,w).
\end{aligned}$$

Then we have

$$0 \le \hat{X}_t^{x,N,\Theta_{MIL}} \le \hat{X}_t^{x,N,g}, \tag{3.14}$$

for all $x, t \ge 0$ and $N \in \mathbb{N}$. For $t = 0, 1/N, 2/N, ...$, we have

$$\begin{aligned}
\hat{X}_t^{x,N,g} &= x + \frac{1}{N}\sum_{i=0}^{tN}\Big(\delta_1 + |k|\hat{X}_{i/N}^{x,N,g} + 3\Big) + 3\sum_{i=1}^{tN}\Big(W_{i/N} - W_{(i-1)/N}\Big)^2 \\
&\quad + 2\sum_{i=1}^{tN}\sqrt{\hat{X}_{i/N}^{x,N,g}}\Big(W_{i/N} - W_{(i-1)/N}\Big) \\
&= x + \int_0^t \Big(\delta_1 + |k|\hat{X}_s^{x,N,g} + 3\Big)\mathrm{d}s + 3\sum_{i=1}^{tN}\Big(W_{i/N}^2 - W_{(i-1)/N}^2 - 2\int_{(i-1)/N}^{i/N} W_{(i-1)/N}\mathrm{d}W_s\Big) \\
&\quad + 2\sum_{i=1}^{tN}\sqrt{\hat{X}_{i/N}^{x,N,g}}\int_{(i-1)/N}^{i/N}\mathrm{d}W_s \\
&= x + \int_0^t \Big(\delta_1 + |k|\hat{X}_s^{x,N,g} + 3\Big)\mathrm{d}s + 3\sum_{i=1}^{tN}\Big(2\int_{(i-1)/N}^{i/N} W_s\mathrm{d}W_s + \big(\frac{i}{N} - \frac{i-1}{N}\big)\Big) \\
&\quad - 6\sum_{i=1}^{tN}\int_{(i-1)/N}^{i/N} W_{(i-1)/N}\mathrm{d}W_s + 2\int_0^t\sqrt{\hat{X}_s^{x,N,g}}\mathrm{d}W_s \\
&= x + \int_0^t \Big(\delta_1 + |k|\hat{X}_s^{x,N,g} + 3\Big)\mathrm{d}s + 6\int_0^t W_s\mathrm{d}W_s + 6t - 6\int_0^t W_{\frac{\lfloor sN\rfloor}{N}}\mathrm{d}W_s + 2\int_0^t\sqrt{\hat{X}_s^{x,N,g}}\mathrm{d}W_s \\
&= x + \int_0^t \Big(\delta_1 + |k|\hat{X}_s^{x,N,g} + 6\Big)\mathrm{d}s + \int_0^t\Big(2\sqrt{\hat{X}_s^{x,N,g}} + 6\big(W_s - W_{\frac{\lfloor sN\rfloor}{N}}\big)\Big)\mathrm{d}W_s.
\end{aligned}$$

For $s = 0, 1/N, 2/N, ..., 1$ and $2 \le q < \infty$, we then have

$$\begin{aligned}
\big|\hat{X}_s^{x,N,g}\big|^q &= \Big|x + \int_0^t \Big(\delta_1 + |k|\hat{X}_s^{x,N,g} + 6\Big)\mathrm{d}s + \int_0^t\Big(2\sqrt{\hat{X}_s^{x,N,g}} + 6\big(W_s - W_{\frac{\lfloor sN\rfloor}{N}}\big)\Big)\mathrm{d}W_s\Big|^q \\
&\le \Big|x + \delta_1 + 6 + \int_0^t |k|\hat{X}_s^{x,N,g}\mathrm{d}s + \int_0^t\Big(2\sqrt{\hat{X}_s^{x,N,g}} + 6\big(W_s - W_{\frac{\lfloor sN\rfloor}{N}}\big)\Big)\mathrm{d}W_s\Big|^q \\
&\le x^q + \big(\delta_1 + 6\big)^q + \Big|\int_0^t |k|\hat{X}_s^{x,N,g}\mathrm{d}s\Big|^q \\
&\quad + \Big|\int_0^t\Big(2\sqrt{\hat{X}_s^{x,N,g}} + 6\big(W_s - W_{\frac{\lfloor sN\rfloor}{N}}\big)\Big)\mathrm{d}W_s\Big|^q \quad \text{from Minkowski's inequality,} \\
&\le x^q + \big(\delta_1 + 6\big)^q + |k|^q\int_0^t\Big|\hat{X}_s^{x,N,g}\Big|^q\mathrm{d}s \\
&\quad + \Big|\int_0^t\Big(2\sqrt{\hat{X}_s^{x,N,g}} + 6\big(W_s - W_{\frac{\lfloor sN\rfloor}{N}}\big)\Big)\mathrm{d}W_s\Big|^q \quad \text{from Hölder's inequality,}
\end{aligned}$$

and

$$\sup_{0\le s\le t}|\hat{X}_s^{x,N,g}|^q \le C\bigg(1+x^q+\int_0^t\big|\hat{X}_u^{x,N,g}\big|^q\mathrm{d}u+\sup_{0\le s\le t}\Big|\int_0^s\Big(2\sqrt{\hat{X}_u^{x,N,g}}+6\big(W_u-W_{\frac{\lfloor uN\rfloor}{N}}\big)\Big)\mathrm{d}W_u\Big|^q\bigg)$$

$$\begin{aligned}\implies \mathbb{E}\Big[\sup_{0\le s\le t}|\hat{X}_s^{x,N,g}|^q\Big] &= C\bigg(1+x^q+\int_0^t\mathbb{E}\Big[\big|\hat{X}_u^{x,N,g}\big|^q\Big]\mathrm{d}u\\ &\qquad+\mathbb{E}\Big[\sup_{0\le s\le t}\Big|\int_0^s\Big(2\sqrt{\hat{X}_u^{x,N,g}}+6\big(W_u-W_{\frac{\lfloor uN\rfloor}{N}}\big)\Big)\mathrm{d}W_u\Big|^q\Big]\bigg)\\ &\le C\bigg(1+x^q+\int_0^t\mathbb{E}\Big[\sup_{0\le s\le u}|\hat{X}_s^{x,N,g}|^q\Big]\mathrm{d}u\\ &\qquad+\mathbb{E}\Big[\sup_{0\le s\le t}\Big|\int_0^s\Big(2\sqrt{\hat{X}_u^{x,N,g}}+6(W_u-W_{\frac{\lfloor uN\rfloor}{N}})\Big)\mathrm{d}W_u\Big|^q\Big]\bigg)\\ &\le C\bigg(1+x^q+\int_0^t\mathbb{E}\Big[\sup_{0\le s\le u}\big|\hat{X}_s^{x,N,g}\big|^q\Big]\mathrm{d}u+\int_0^t\mathbb{E}\Big[\Big|2\sqrt{\hat{X}_u^{x,N,g}}+6\big(W_u-W_{\frac{\lfloor uN\rfloor}{N}}\big)\Big|^q\Big]\mathrm{d}u\bigg)\\ &\le C\bigg(1+x^q+\int_0^t\mathbb{E}\Big[\sup_{0\le s\le u}\big|\hat{X}_s^{x,N,g}\big|^q\Big]\mathrm{d}u\\ &\qquad+2^{q-1}\int_0^t\mathbb{E}\Big[\Big(2\sqrt{\hat{X}_u^{x,N,g}}\Big)^q+\big|6\big(W_u-W_{\frac{\lfloor uN\rfloor}{N}}\big)\big|^q\Big]\mathrm{d}u\bigg).\end{aligned}$$

Using Young's inequality with $\left(2\sqrt{\hat{X}_u^{x,N,g}}\right)^q$ and the fact that the increments of a Wiener process are normally distributed with mean 0 gives us

$$\begin{aligned}\mathbb{E}\Big[\sup_{0\le s\le t}\big|\hat{X}_s^{x,N,g}\big|^q\Big] &\le C\Big(1+x^q+\int_0^t\mathbb{E}\Big[\sup_{0\le s\le u}|\hat{X}_s^{x,N,g}|^q\Big]\mathrm{d}u\\ &\qquad+\int_0^t\Big(\mathbb{E}\Big[\big|\hat{X}_u^{x,N,g}\big|^q\Big]+\mathbb{E}\Big[\big|Z\big|^q\Big]\Big)\mathrm{d}u\Big),\quad \text{where } Z\sim N(0,1).\end{aligned}$$

Using that the absolute moments of a standard normal random variable are bounded, we get

$$\mathbb{E}\Big[\sup_{0\le s\le t}|\hat{X}_s^{x,N,g}|^q\Big]\le C\Big(1+x^q+\int_0^t\mathbb{E}\Big[\sup_{0\le s\le u}|\hat{X}_s^{x,N,g}|^q\Big]\mathrm{d}u\Big).$$

Gronwall's inequality then gives

$$\begin{aligned}\mathbb{E}\Big[\sup_{0\le s\le t}|\hat{X}_s^{x,N,g}|^q\Big] &\le C(1+x^q)+C(1+x^q)\cdot\int_0^t e^{C(t-s)}\mathrm{d}s\\ &= C(1+x^q)-(1+x^q)e^{Ct}(1-e^{-Ct})\\ &\le C(1+x^q),\end{aligned}$$

which completes the proof. □

Remark 3.6. *The proof for Theorem 3.1 shows that this convergence rate holds for a whole class of approximation schemes. In particular, any suitable one-step scheme that satisfies Conditions 3.1, 3.2 and 3.3 fall into this class.*

4 Empirical Analysis

In this section, we compare the performance of the approximation schemes mentioned in Section 3 by analysing their convergence properties under different Feller ratios. We analyse strong convergence by estimating the $\mathcal{L}^2$ error for different values of N and we analyse weak convergence by comparing estimates for zero-coupon bond prices under the CIR short-rate model and European call option prices under the Heston model with their respective fair prices. Overall, we expect performance to be worse for lower values of δ_0 and better for higher values. We recall that schemes (3.4) and (3.5) are only well-defined for $\delta_0 > 1$ and $\delta_1 > 1$ respectively, so they are omitted from the analysis when these conditions are not met. Similarly, the modified Milstein scheme given by (3.6) is only well-defined for $\delta_1 > 1$. However, it was still of interest to look at this scheme suitably truncated at 0 (scheme (3.7)) when this condition is violated.

In the following analysis, we refer to schemes (3.4), (3.5) and (3.6) as 'Implicit 1', 'Implicit 2' and the 'modified Milstein' respectively. In the figures in this section, scheme (3.7) is referred to as the '(truncated) modified Milstein' scheme but we will mostly refer to it as scheme (3.7) elsewhere. The antithetic variate technique was also employed as a tool for variance reduction in order to keep the Monte Carlo error sufficiently low. Further details on antithetic variates can be found in Appendix B.

4.1 Strong Error Analysis

For some fixed $T > 0$ and any approximation scheme $\hat{X}^{x,N} = (\hat{X}_t^{x,N})_{0 \le t \le T}$, the root mean square error (RMSE) at time $T > 0$ is given by

$$\left(\mathbb{E}\left[\left|\hat{X}_T^{x,N} - X_T^x\right|^2\right]\right)^{1/2}.$$

To estimate this numerically, the true solution X^x needs to be approximated as there is no known analytical form for the solution to (1.1). This is done by choosing a scheme that is known to converge strongly in $\mathcal{L}^2$ (under the given parameter setting) and simulating it for a very small time-step, δt . For this analysis, the truncated Milstein scheme was chosen for this approximation. A large number of runs, $N_{MC} \in \mathbb{N}$, of the approximated true solution are generated, and the Wiener processes used for each run are then used to generate the corresponding paths for a given approximation scheme $\hat{X}^{x,N}$. The RMSE can be then estimated as

$$\sqrt{\frac{1}{N_{MC}} \sum_{i=1}^{N_{MC}} \left|\left(\hat{X}_T^{x,N}\right)_i - \left(X_T^x\right)_i\right|^2}, \quad \text{for } i = 1, ..., N_{MC}.$$

When a larger time-step (compared to that used for approximating X^x), Δt, is used for the approximation scheme, the original Wiener increments are summed along each path in a way that produces more discrete Wiener increments, suitable to that of the approximation scheme. For example, if a time-step of 2^{-14} is used to approximate one path of the true solution and we wish to generate the corresponding path for the approximation scheme using time-step of 2^{-5}, then the Wiener increments used to generate the approximated true solution are summed every 2^9 time-points to produce new Wiener increments for a path with step-size 2^{-5}. Using this technique, we can estimate the rate of strong convergence for each scheme as $\Delta t \to \delta t$ and by plotting a log-log plot of RMSE vs. runtime, we can compare scheme convergence. Better performing schemes will have a steep descending line of best fit that is closer to the bottom left corner of the graph. We chose $N_{MC} = 2 \times 10^4$, $\delta t = 2^{-14}$ and the Monte Carlo error in this case is calculated as

$$\text{Error}_{MC} = \sqrt{\frac{\text{Var}\left(\hat{X}_T^{x,N}\right)}{N_{MC}}}.$$

Each scheme will have its own Monte Carlo error, but it was found that they were all significantly close to each other. Therefore, the average of all Monte Carlo errors was taken as a reliable measurement of error. We fix $x = 0.04$, $a = 0.02$, $k = 0.4$ and $T = 1$, while σ is adjusted according to the relationship $\sigma = \sqrt{\frac{2a}{\delta_0}}$ for Feller ratios $\delta_0 = 0.25, 0.45, 0.75, 1.15$.

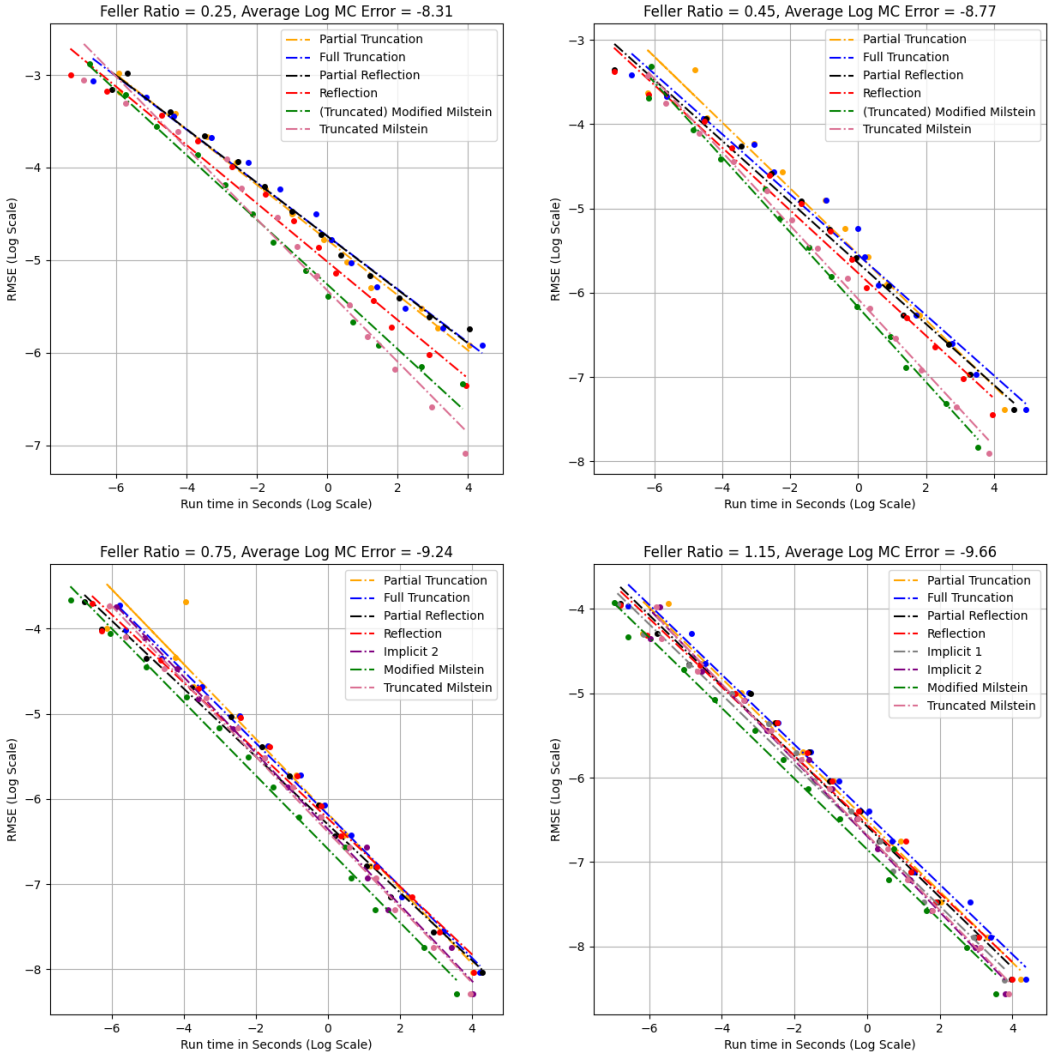

Figure 3: Log-log plots of RMSE vs. run-time of approximation schemes for $\delta_0 = 0.25, 0.45, 0.75, 1.15$.

Figure 3 illustrates the estimated rate of strong convergence for each approximation scheme. The corresponding values can be seen in Table 1. We can sum up these results as follows:

- For $\delta_0 = 0.25$, we observe that the rate of convergence of the truncated Milstein scheme is slightly better than that of scheme (3.7). Relatively poor convergence is observed for the remaining schemes. The full truncation scheme and the partial reflection scheme are shown to have almost identical rates of convergence here.

- For $\delta = 0.45$, we observe that the rate of scheme (3.7) is now slightly better than that of the truncated Milstein scheme, suggesting that one scheme may start to outperform the other for

some Feller ratio less than 0.5. Figure 4 illustrates the observed crossing point of convergence rates when the Feller ratio is less than or equal to 0.5. Again, the remaining schemes perform relatively poor here.

- For $\delta_0 = 0.75$, the modified Milstein scheme is shown to perform relatively well here as it has a convergence rate similar to the other schemes and it is observed to have faster computation times for practical values of N. The truncated Milstein scheme and the Implicit 2 scheme are observed to have higher convergence rates but not by a significant amount. The explicit Euler schemes are unsurprisingly observed to have relatively slower convergence rates here.
- For $\delta_0 = 1.15$, the truncated Milstein scheme and the Implicit 1 scheme are observed to have the highest rates of convergence (as shown in Table 1). The modified Milstein scheme still outperforms the other schemes in terms of computation time for practical values of N but it is observed to have a slightly lower convergence rate. Nevertheless, the gap in convergence rate is not significantly large considering the error associated with the line of best fit. On the other hand, we can see the relatively poor performance of the explicit Euler schemes once again.

Observed $\mathcal{L}^2$ Orders of Convergence				
	$\delta_0 = 0.25$	$\delta_0 = 0.45$	$\delta_0 = 0.75$	$\delta_0 = 1.15$
Partial Truncation	0.298	0.391	0.439	0.421
Full Truncation	0.287	0.358	0.42	0.415
Partial Reflection	0.289	0.362	0.399	0.418
Reflection	0.315	0.372	0.399	0.408
Implicit 1	N/A	N/A	N/A	0.418
Implicit 2	N/A	N/A	0.45	0.449
Modified Milstein (suitably truncated at 0)	0.349	0.445	0.431	0.419
Truncated Milstein	0.384	0.436	0.442	0.447

Table 1: Observed orders of convergence in $\mathcal{L}^2$ corresponding with Figure 3.

Table 1 contains the observed orders of convergence corresponding with Figure 3. We note here that the observed order of convergence of a given scheme is calculated as the negative of the slope of the line of best fit for that scheme. Firstly, we see that the observed order of convergence for each scheme (when well-defined) generally downgrades as the Feller ratio decreases below 0.5. Indeed, we observe all of the well-defined schemes to downgrade in terms of convergence rate when δ_0 decreases from 0.45 to 0.25. This behaviour is shown to be worse for the explicit Euler schemes than it is for the Milstein schemes. This and the comparatively poor performance observed in Figure 3 suggest that the explicit Euler schemes should not be preferred. We also note that the Implicit 2 scheme is observed to have a relatively high convergence rate whenever it is well-defined.

Figure 5 shows the estimated probabilities of the full truncation scheme, modified Milstein scheme (suitably truncated at 0) and the truncated Milstein scheme attaining 0 at a given time $T = 1$. We know from Section 2 that as soon as the CIR process attains 0, it is instantly brought away from 0 back to the positive values. Intuitively, this makes it highly unlikely that the CIR process will be equal to 0 at a given time. In Figure 5, we observe that this probability is non-zero for practical values of N and for $\delta_0 = 0.25, 0.45$ when implementing the full truncation scheme and scheme (3.7). In fact, the full truncation scheme is observed to have significantly high probabilities of attaining 0 for the low Feller ratios. On the other hand, the truncated Milstein scheme (more or less) retains a 0 probability throughout. The poor results here for scheme (3.7) are likely due to the truncation correction applied when $\delta_0 < 0.5$. This suggests that crude flooring at 0 is not a particularly good correction technique for the modified Milstein scheme when $\delta_1 < 1$ and maybe some other type of correction would yield better results. The significantly high probability observed for scheme (3.7) when $\delta_0 = 0.25$ may be an influencing factor as to why we see the truncated Milstein scheme outperform scheme (3.7) (in the strong sense) for the more extreme end of low Feller ratios.

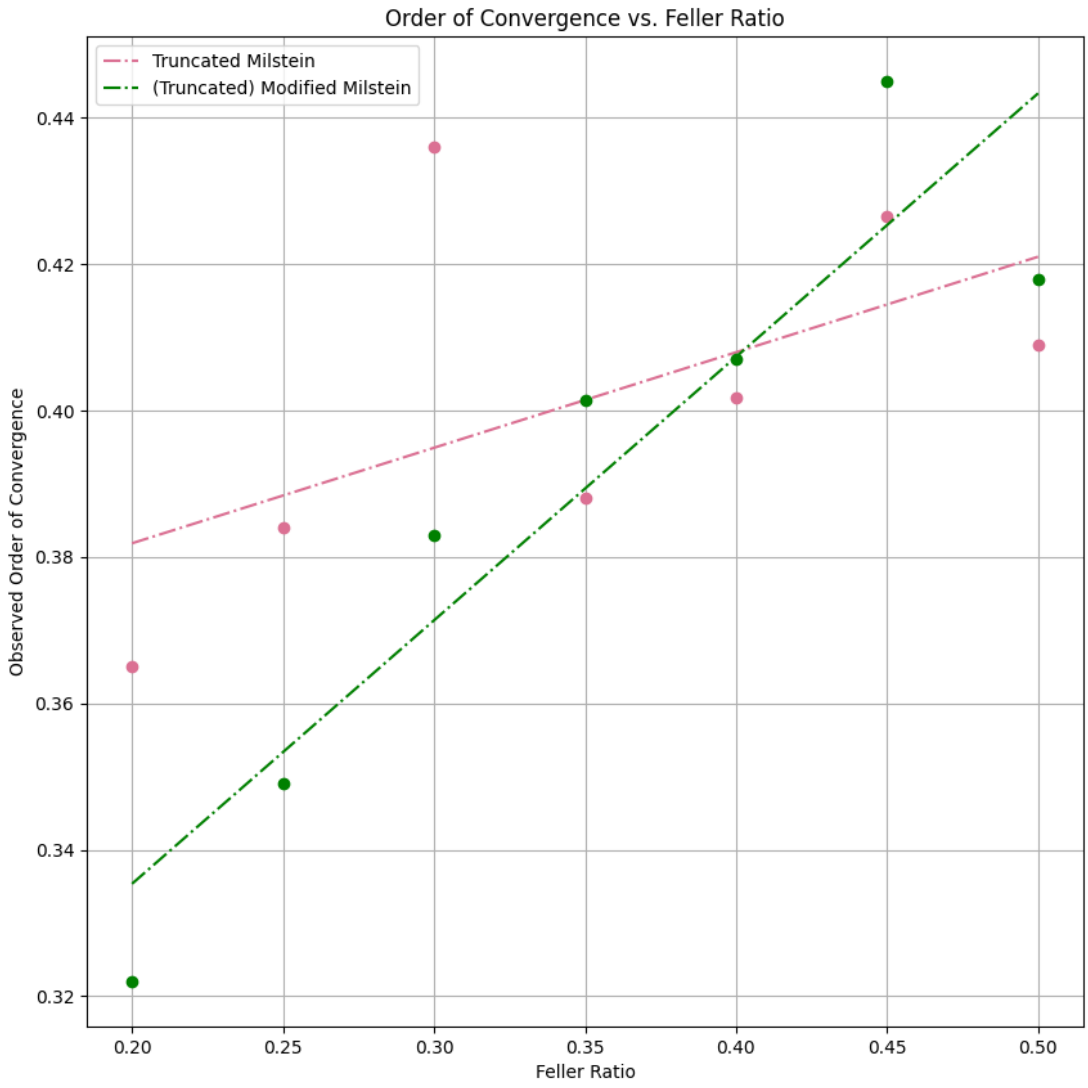

Figure 4: Plot of observed order of convergence of the truncated Milstein scheme and the (truncated) modified Milstein scheme vs. Feller ratio for some $\delta_0 \leq 0.5$.

Based on this analysis of the strong error, the truncated Milstein scheme and the modified Milstein scheme (even when suitably truncated at 0) are both good choices of approximation schemes for the full parameter range. The truncated Milstein scheme should be preferred for very low Feller ratios as indicated by the observed relationship in Figure 4, while the modified Milstein scheme should be preferred otherwise, with the truncation technique applied at 0 when $\delta_0 < 0.5$. Good convergence rates are also observed for the Implicit 2 scheme when it is well-defined, which indicates that it is also a good choice of scheme for relatively high Feller ratios. As indicated by the actual data points (the dots in Figure 3), the modified Milstein scheme and scheme (3.7) are observed to perform best in terms of computational time for the values of N that were tested, which is generally of more interest in a practical setting.

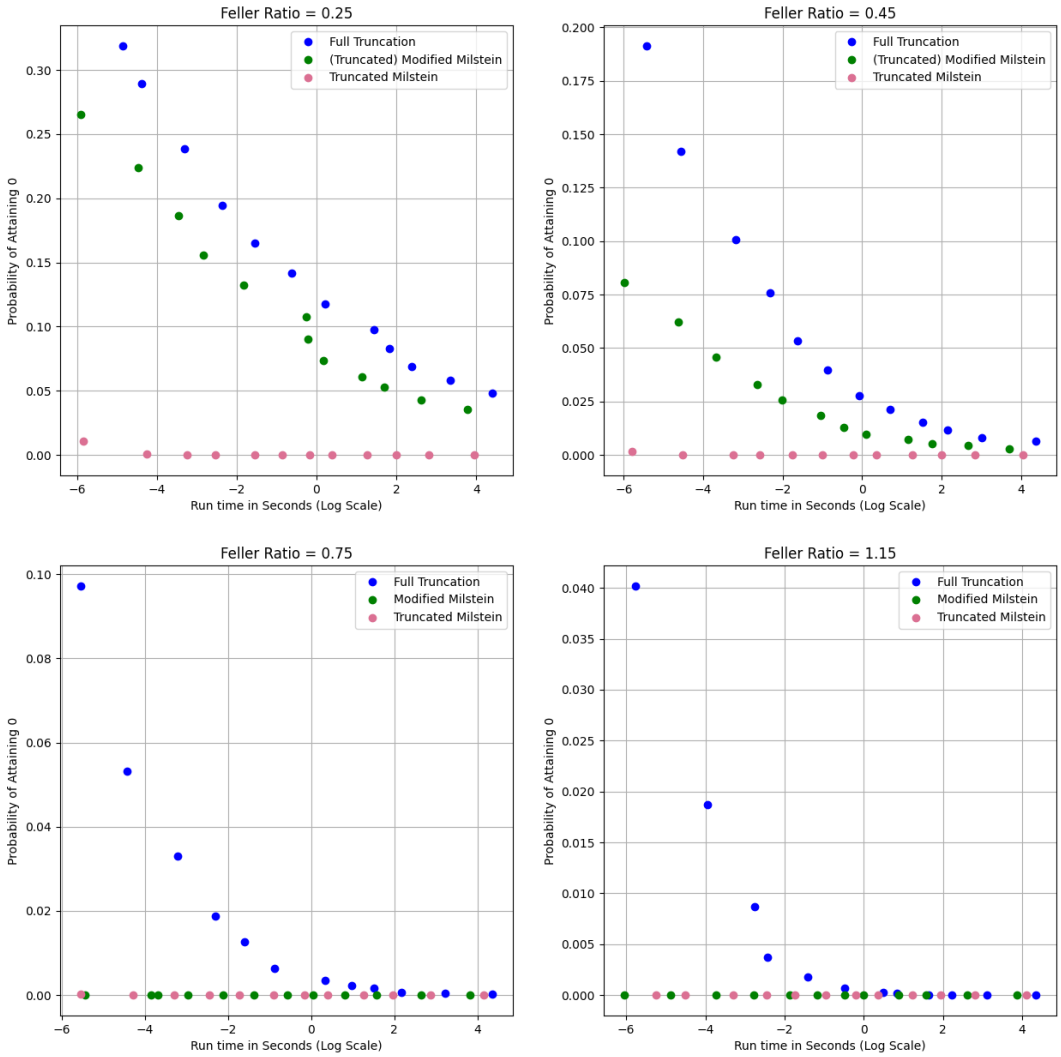

Figure 5: Plots of numerical probability of schemes attaining 0 (at time $T = 1$) vs. log run-time for $\delta_0 = 0.25, 0.45, 0.75, 1.15$.

4.2 Weak Error Analysis

Since there is no possible way to analyse weak convergence of approximation schemes for every $\mathcal{C}^\infty$ function $f : \mathbb{R} \to \mathbb{R}$ with compact support, we have decided out of interest to analyse weak convergence by looking at two widely used applications of the CIR process in financial mathematics. The first application is the short-rate model under which we analyse weak convergence through calculating zero-coupon bond prices. The second application is the Heston model under which we analyse weak-convergence through pricing European call options.

4.2.1 Zero-Coupon Bond Pricing

The original motivation for the CIR model was for modelling short-rate dynamics. Under this setting, one can determine the price a zero-coupon bond at time $t \in [0, T]$ maturing at time $T > 0$,

$$P_{ZCB}(t,T) = \mathbb{E}\Big[e^{-\int_t^T X_s^x \mathrm{d}s} \big| X_t^x\Big]. \tag{4.1}$$

The parameter k is also assumed to be positive under this setting. Thankfully, there is a known analytical form for (4.1) as shown in [7]. It is given by

$$P_{ZCB}(t,T) = A(t,T)e^{-B(t,T)X_t^x},$$

where

$$A(t,T) = \left(\frac{\sqrt{k^2+2\sigma^2}\exp\big(\frac{(k+\sqrt{k^2+2\sigma^2})(T-t)}{2}\big)}{2\sqrt{k^2+2\sigma^2}+(k+\sqrt{k^2+2\sigma^2})\big(\exp\big((T-t)\sqrt{k^2+2\sigma^2}\big)-1\big)}\right)^{2a/\sigma^2},$$

$$B(t,T) = \frac{2\big(\exp\big((T-t)\sqrt{k^2+2\sigma^2}\big)-1\big)}{2\sqrt{k^2+2\sigma^2}+(k+\sqrt{k^2+2\sigma^2})\big(\exp\big((T-t)\sqrt{k^2+2\sigma^2}\big)-1\big)}.$$

In order to numerically estimate (4.1) with an approximation scheme $\hat{X}^{x,N}$, we also need to approximate the integral within the expectation. Unfortunately, this requirement introduces an extra discretisation error in addition to the discretisation error already associated with the approximation schemes. The technique chosen to approximate this integral is the trapezoidal rule, which should be sufficiently accurate here since it is at least as accurate as any of the schemes we are analysing. Thus, the Monte Carlo estimate for (4.1) at time point t_n is given by

$$\hat{P}_{ZCB}^N(t_n,T) = \frac{1}{N_{MC}}\sum_{i=1}^{N_{MC}} \exp\Big(-\frac{T}{2N}\sum_{j=n+1}^{N}\big(\hat{X}_{t_j}^{x,N} - \hat{X}_{t_{j-1}}^{x,N}\big)\Big).$$

For the purposes of this study, we have only considered the case where $t = t_n$. Otherwise, interpolation between grid points contributes to arbitrage opportunities already introduced by the error associated with the approximation schemes. We are interested in the difference

$$\big|P_{ZCB}(t,T) - \hat{P}_{ZCB}^N(t,T)\big|,$$

for different N. In this analysis, we assume $t = 0$ and $T = 1$. W also fix $N_{MC} = 5 \times 10^5$, $a = 0.02$, $x = 0.03$ and $k = 0.4$. For Feller ratios $\delta_0 = 0.25, 0.45, 0.75, 1.15$, we adust σ accordingly. The Monte Carlo error is calculated by

$$\text{Error}_{MC} = \sqrt{\frac{\text{Var}\big(\hat{P}_{ZCB}^N(t,T)\big)}{N_{MC}}}.$$

Figure 6 illustrates the comparative scheme performance under the short-rate model. The results from this analysis can be summed up as follows:

- For $\delta_0 = 0.25$, scheme (3.7) is observed to have the best convergence by a significant margin. Indeed, the truncated Milstein scheme is shown to be outperformed here by scheme (3.7), which does not correspond to the results of strong convergence in Section 4.1. The reflection scheme also appears to produce a very slow rate of convergence in comparison to the other schemes.
- For $\delta_0 = 0.45$, the reflection scheme is shown to produce the best rate of convergence, which is surprising considering the results for $\delta_0 = 0.25$. On the other hand, scheme (3.7) shows a slower convergence rate than the truncated Milstein scheme here, which does not correspond with the results from Section 4.1.
- For $\delta_0 = 0.75$, the modified Milstein scheme is observed to dominate the other schemes in every sense. The full truncation scheme is observed to have the worst performance for this Feller

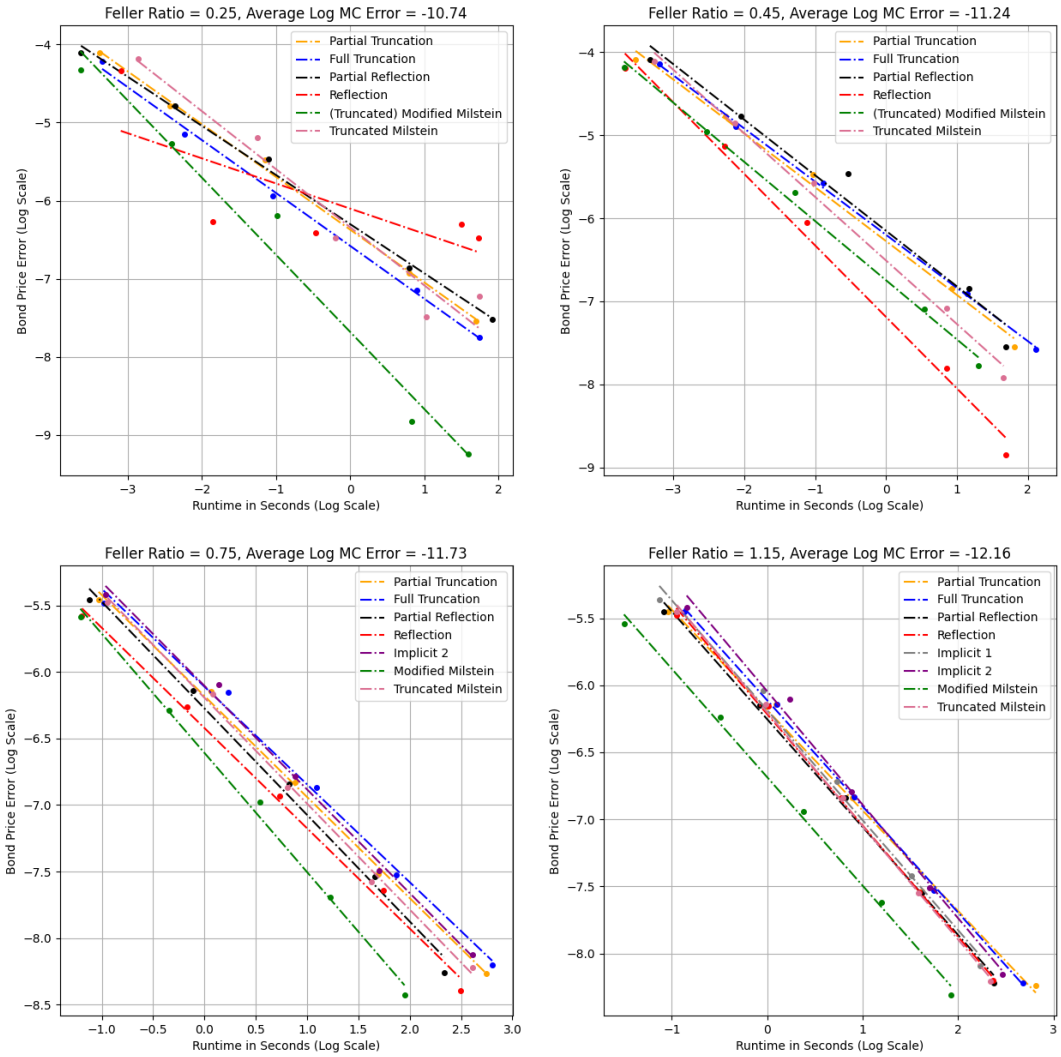

Figure 6: Log-log plots of bond price error vs. run-time of approximation schemes for $\delta_0 = 0.25, 0.45, 0.75, 1.15$ under the CIR short-rate model.

ratio. A similar result is observed for $\delta_0 = 1.15$, with a higher convergence rate observed for the Implicit 2 scheme.

From this analysis, we can see that the modified Milstein scheme (and scheme (3.7) where suitable), should be the preferred scheme(s) under this model with maybe the exception of Feller ratios of around 0.45. Under this setting, it is not too clear from this analysis which scheme should be preferred. One could argue that the reflection scheme should be preferred for $\delta_0 = 0.45$, however, one could also argue that the result observed for $\delta_0 = 0.25$ highlights the potential inconsistency associated with this scheme. Likewise, the relatively high rate of convergence observed for the truncated Milstein scheme for $\delta_0 = 0.45$ could be an argument for why it should be preferred under this setting.

4.2.2 Option Pricing Under the Heston Model

The Heston Model is an example of how the CIR process is used to model the stochastic volatility, $X^x = (X_t^x)_{t\geq 0}$, of a stocks price. The price of the stock itself, $S = (S_t)_{t\geq 0}$, follows a geometric Brownian motion driven by a source of risk that is correlated with that of its volatility and this correlation is measured by a constant $\rho \in [-1, 1]$. The Heston model dynamics are given by

$$\begin{aligned} \mathrm{d}S_t &= \mu S_t \mathrm{d}t + \sqrt{X_t^x} S_t \mathrm{d}W_1, \;\; t \geq 0, \;\; S_0 = s, \\ \mathrm{d}X_t^x &= (a - kX_t^x)\mathrm{d}t + \sigma\sqrt{X_t^x}\big(\rho \mathrm{d}W_1 + \sqrt{1-\rho^2}\mathrm{d}W_2\big), \;\; t \geq 0, \;\; X_0^x = x, \end{aligned}$$

where W_1 and W_2 are two independent $\mathcal{F}_t$-adapted Wiener processes, $\mu, k \in \mathbb{R}$, $s, x, \sigma > 0$ and $a \geq 0$. In order to analyse the weak convergence under this model, we look at pricing vanilla European call options through simulating S and X^x using the different approximation schemes and comparing the estimated prices to the respective fair price. For a fixed strike price K, time to maturity $T > 0$ and risk-free interest rate $r \in \mathbb{R}$, the arbitrage-free price of a European call option at time $t \in [0, T]$ under this model is

$$C(t, T) = e^{-r(T-t)}\mathbb{E}^{\mathbb{Q}}\Big[\big(S_T - K\big)^+ \big| \mathcal{F}_t\Big], \tag{4.2}$$

where $\mathbb{Q}$ is the risk-neutral probability measure. $\mathbb{Q}$ is not unique under this model, as the dynamics under $\mathbb{Q}$ are given by

$$\begin{aligned} \mathrm{d}S_t &= rS_t \mathrm{d}t + \sqrt{X_t^x} S_t \mathrm{d}W_1^{\mathbb{Q}}, \;\; S_0 = s, \\ \mathrm{d}X_t^x &= (a - (k+c)X_t^x)\mathrm{d}t + \sigma\sqrt{X_t^x}\big(\rho \mathrm{d}W_1^{\mathbb{Q}} + \sqrt{1-\rho^2}\mathrm{d}W_2^{\mathbb{Q}}\big), \;\; X_0^x = x, \end{aligned}$$

where c is an arbitrary constant and $W_1^{\mathbb{Q}}$ and $W_2^{\mathbb{Q}}$ are $\mathbb{Q}$-Wiener martingales. In practice, the free parameter c relates to the market price of risk and it is calibrated to market conditions. For the purpose of this analysis, we can assume $c = 0$ since market calibration is not required. To estimate (4.2) numerically, we therefore need to approximate X_t^x using the chosen approximation scheme $X_{x,N}$ and we use the log-Euler scheme to approximate S_t, i.e.

$$\hat{S}_0^N = s, \;\; \hat{S}_{t_{i+1}}^N = \hat{S}_{t_i}^N e^{(r - \hat{X}_{t_i}^{x,N}/2)\Delta t + \sqrt{\hat{X}_{t_i}^{x,N}}\Delta W_i},$$

for $i = 0, ..., N-1$. Thus, the Monte Carlo estimate for (4.2) is

$$\hat{C}^N(t, T) = \frac{e^{-r(T-t)}}{N_{MC}} \sum_{i=1}^{N_{MC}} \Big(\big(\hat{S}_{t_N}^N\big)_i - K\Big)^+. \tag{4.3}$$

The Black-Scholes formula cannot be used here to calculate the fair price since one of the key assumptions for that model is that the volatility of the underlying stock price is constant for all $t \geq 0$. As outlined in [1], we can consider two formulae for the fair price of a European call option under the Heston model. In this analysis, we apply the following choice of formula for the reasons outlined in [1].

$$C(t, T) = sP_1(t, \log(s), x) - Ke^{-r(T-t)}P_2(t, \log(s), x),$$

where for $j = 1, 2$,

$$
\begin{aligned}
P_j(t, L, x) &= \frac{1}{2} + \frac{1}{\pi}\int_0^\infty \mathrm{Re}\left(\frac{e^{-i\phi\log(K)}\psi_j(t, L, x, \phi)}{i\phi}\right)\mathrm{d}\phi, \qquad (4.4)\\
\psi_j(t, L, x, \phi) &= \exp\Big(C_j(T-t, \phi) + D_j(T-t, \phi)x + i\phi L\Big),\\
C_j(T-t, \phi) &= r\phi i(T-t) + \frac{a}{\sigma^2}\left(\big(b_j - \rho\sigma\phi i - d_j(\phi)\big)(T-t) - 2\log\left(\frac{1 - g_j(\phi)e^{-(T-t)d_j(\phi)}}{1 - g_j(\phi)}\right)\right),\\
D_j(T-t, \phi) &= \left(\frac{b_j - \rho\sigma\phi i - d_j(\phi)}{\sigma^2}\right)\left(\frac{1 - e^{-(T-t)d_j(\phi)}}{1 - g_j(\phi)e^{-(T-t)d_j(\phi)}}\right),\\
g_j(\phi) &= \frac{b_j - \rho\sigma\phi i - d_j(\phi)}{b_j - \rho\sigma\phi i + d_j(\phi)},\\
d_j(\phi) &= \sqrt{(\rho\sigma\phi i - b_j)^2 - \sigma^2(2u_j\phi i - \phi^2)},
\end{aligned}
$$

with $b_1 = k - \rho\sigma$, $b_2 = k$, $u_1 = 1/2$, $u_2 = -1/2$. We are interested in the difference

$$\big|C(t, T) - \hat{C}^N(t, T)\big|,$$

for different N. Numerical integration can be used to calculate (4.4) since the integrand decays quickly. In our analysis, the trapezoidal rule was used with step-size of approximately 0.02 to integrate from $\phi = 1 \times 10^{-7}$ to 175, providing a sufficient level of accuracy for the same reason outlined in Section 4.2.1. We also fixed $t = 0$, $T = 1$, $s = 1$, $x = 0.17$, $r = 0$, $\rho = -0.9$, $K = 1.1$, $a = 0.08$ and $k = 0.4$. We chose $N_{MC} = 4.5 \times 10^5$ and we ran the analysis for Feller ratios $\delta_0 = 0.25, 0.45, 0.75, 1.1$. The Monte Carlo error in this case is calculated by

$$\mathrm{Error}_{MC} = \sqrt{\frac{\mathrm{Var}\big(\hat{C}^N(t, T)\big)}{N_{MC}}}.$$

We also excluded any scheme that could take negative values from this analysis due to the square root term in the SDE for S_t.

Figure 6 shows the comparative scheme performance under the Heston option pricing model. For $\delta_0 = 0.25$, a rate of convergence was only observed for $N > 4$, hence why this part of the analysis was run for larger N. The results from this analysis can be summed up as follows:

- For $\delta_0 = 0.25$, it appears that the full truncation scheme (3.7) performs best here by a significant margin, followed by scheme (3.7). The observed difference between run-times for the two schemes at a given error level is significant, although the actual data points for scheme (3.7) show a clearer convergence pattern here. In contrast to the strong error results in Section 4.1, the truncated Milstein scheme is observed to have relatively poor convergence properties here, while the reflection scheme unsurprisingly performs the worst here.

- For $\delta_0 = 0.45$, the best convergence is seen through the reflection scheme, corresponding to the result obtained under the short-rate model in Section 4.2.1. The full truncation scheme appears to have relatively poor convergence here in contrast to the result obtained for $\delta_0 = 0.25$. At the same time, relatively poor performance is seen through the two Milstein schemes, although scheme (3.7) does appear to perform second best out of all the schemes here.

- For $\delta_0 = 0.75$ and $\delta_0 = 1.15$, we see similar convergence rates amongst the schemes. If anything, the Implicit 2 scheme is observed to have a faster rate of convergence for $\delta_0 = 1.15$. However, the modified Milstein scheme is observed to be computationally faster (for practical values of N) for both Feller ratios and it appears to have at least as good a rate of convergence as the other schemes for $\delta_0 = 0.75$. In both plots, we see that the rate of convergence of the reflection scheme is slightly worse than the other schemes, in contrast to the result obtained for $\delta_0 = 0.45$.

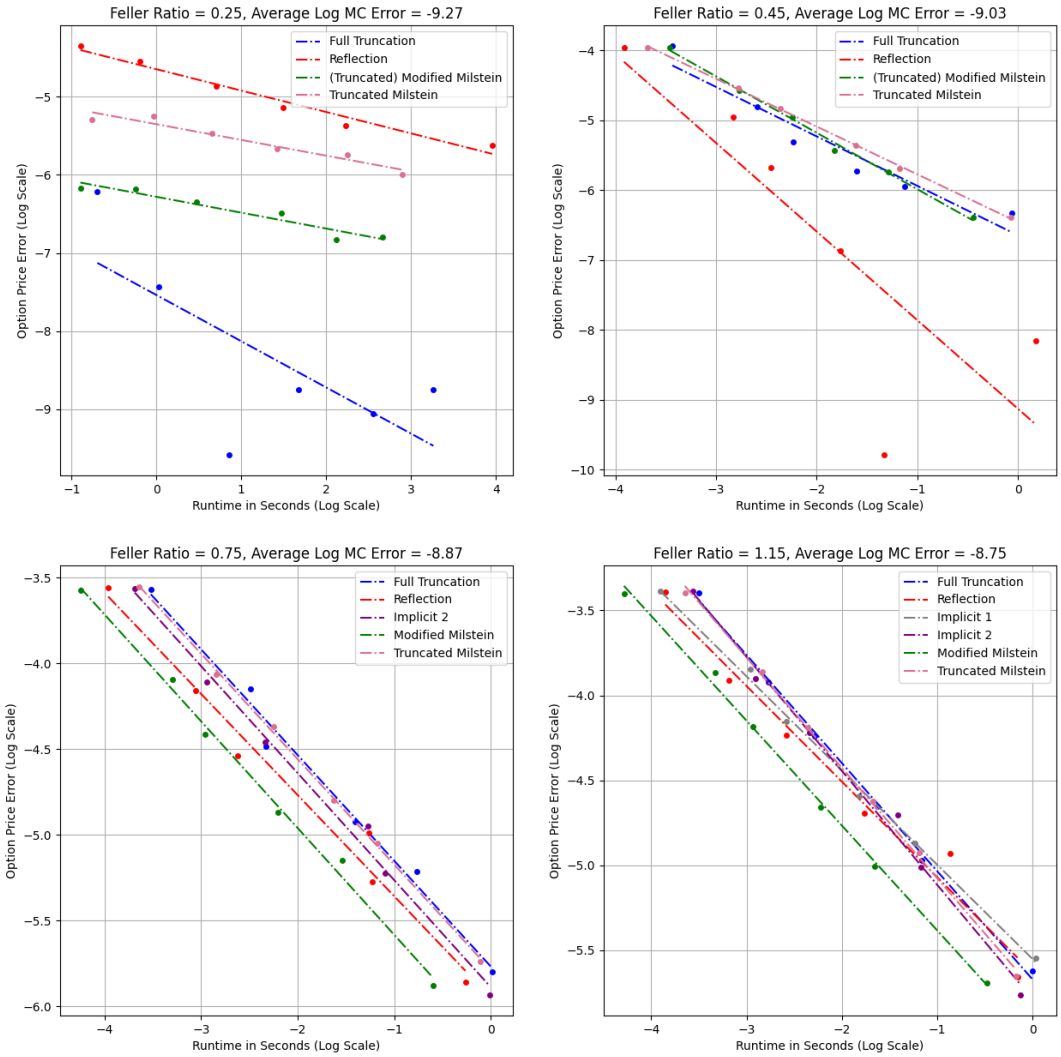

Figure 7: Log-log plots of European call option price error vs. run-time of approximation schemes for $\delta_0 = 0.25, 0.45, 0.75, 1.15$ under the Heston model.

Overall, similar results are obtained here compared to those obtained in Section 4.2.1. Based on the results obtained here, the modified Milstein scheme (and scheme (3.7) where suitable) should be the preferred scheme(s) under this model. We see the inconsistent behaviour of the reflection scheme here again for very low Feller ratios and we also see it with the full truncation scheme too. The good result observed for the full truncation scheme may be explained by the large amount of time its solution will be equal to 0, as indicated by Figure 1 and Figure 5. Perhaps the Heston option pricing model favours schemes that have solutions which spend lots of time around or equal to 0 when the Feller ratio is very low, as we also see scheme (3.7) perform well here too. This kind of inconsistent behaviour associated with the full truncation and reflection schemes would lead us to avoid picking either of this schemes under this model. On the other hand, the more consistent convergence of scheme (3.7) would lead us to prefer it under the lower Feller ratios. We also observe that the truncated Milstein scheme is one of the worst performing schemes under this model and therefore should not be the preferred scheme of choice here. The relatively poor performance of this scheme observed for the overall weak error analysis is surprising considering the good convergence observed for this scheme in Section 4.1.

5 Conclusions

We have studied the CIR process along with some of the available methods for its simulation, which we have seen can be a difficult task. In particular, we have discussed how exact simulation methods can be computationally challenging and often unnecessary, making numerical approximation schemes a preferable choice in many contexts. Although the standard Euler-Maruyama scheme is known to converge strongly, this scheme cannot generally be applied in practice since it is not well-defined. Similarly, the standard Milstein scheme is not well-defined for the full parameter range. As a result, we have studied some of the other available numerical approximation schemes which allow us to reasonably approximate the CIR process on a discrete-time grid. In particular, we have looked at four modified explicit Euler schemes, two implicit Euler schemes and two Milstein schemes. The performance of such schemes can be heavily influenced by the Feller ratio, $\delta_0 = \frac{2a}{\sigma^2}$, as low Feller ratios generally cause them to spend more time near 0, which can be troublesome when approximating the solution to (1.1) on a discrete-time grid. With this in mind, we have attempted to identify which approximation schemes perform best in terms of strong and weak convergence under different parameter settings.

In Section 2, we studied some of the fundamental properties of the CIR process. In particular, we have seen how a unique strong solution to (1.1) exists and through doing so, we saw how this solution is non-negative. Such a feature makes the CIR process attractive for financial modelling, along with the fact that the process is mean-reverting when $k > 0$. We have also seen how the solution can only attain value 0 if the Feller condition ($\delta_0 \geq 1$) is violated, which can often be the case in practice. In addition, we have studied the probability distribution of the CIR process, further seeing that exact simulation of the CIR process requires sampling from a non-central χ^2-distribution. As a result, exact simulation methods can often bring a computational burden with them.

In Section 3, we discussed some of the practical reasons for numerically approximating the CIR process instead of exact simulation. A selection of approximation schemes that are present in mathematical literature were then explored. At first, we discussed some of the modified explicit Euler schemes, which aim to correct the problem associated with the standard Euler scheme by ensuring that the square root term in the diffusion coefficient of (1.1) stays defined. These are rather simple corrections, but these schemes are known to converge strongly thanks to the result obtained by Gyöngy et al. [11]. At the same time, the slow convergence rate of $1/\log(N)$ for these schemes is unsurprising. Furthermore, we discussed two implicit Euler schemes, one of which is formulated from the dynamics of the square root of the CIR process. Then, two Milstein schemes were discussed, one of which was proposed by Alfonsi [4] and the other, proposed by Hefter and Herzwurm [12] that has been shown to converge strongly at a rate potentially better than $1/\log(N)$ for the full parameter range. This rate, however, is monotonically decreasing in $\frac{1}{2\delta_0}$ and we find that an arbitrarily slow convergence rate is achieved as the Feller ratio approaches 0. Nevertheless, we gave the proof for this result by following the approach in [12], which involves showing that this convergence rate holds for a whole class of approximation schemes (see Remark 3.6).

In Section 4, we compared the convergence properties of the aforementioned approximation schemes through empirical analysis. In Section 4.1, we analysed the strong error for each approximation scheme (see Figure 3 and Table 1) and found that the modified Milstein scheme proposed by Alfonsi [4] was observed to perform very well (even when applying a truncation correction for $\delta_0 < 0.5$) with the exception of a very low Feller ratio ($\delta_0 = 0.25$). Under this particular Feller ratio, the truncated Milstein scheme proposed by Hefter and Herzwurm [12] performed the best with a higher observed convergence rate. We suggested the possible existence of a linear relationship between the Feller ratio and strong convergence rate for the modified Milstein scheme suitably truncated at 0 (given by (3.7)), which might identify an approximate Feller ratio at which the truncated Milstein scheme starts to outperform scheme (3.7) (see Figure 4). To the best of our knowledge, it remains to see whether or not scheme (3.7) falls into the same class of approximation schemes used to prove the convergence rate for the truncated Milstein scheme. In addition, we also suggested that the truncation correction

applied to the modified Milstein scheme was a potential reason for the relatively poor performance of this scheme for low Feller ratios and perhaps a better correction technique is required (see Figure 5). In Section 4.2, we carried out weak error analysis by comparing estimated zero-coupon bond prices under the short-rate model and vanilla European call option prices under the Heston model to their respective fair prices. The results obtained for the modified Milstein scheme (suitably truncated at 0) were once again impressive here, while the results obtained for the truncated Milstein scheme were disappointing. Overall, we were impressed by the performance of the modified Milstein scheme, even when a truncation correction was applied. However, a general recommendation for which scheme to prefer would not be wise here since the context of application is important. In the context of strong approximation, we recommend that the truncated Milstein scheme should be preferred when the Feller ratio is very low (lower than approximately 0.4 based on Figure 4) and the modified Milstein scheme should be preferred elsewhere, with a truncation correction applied when the Feller ratio is less than 0.5. We also mention that the implicit scheme given by (3.5) is also a good scheme of choice for Feller ratios above 0.5 as it was observed to have relatively high rates of strong convergence. For practical reasons, however, we believe the modified Milstein scheme would be a better choice due to its computational efficiency. We also recommend that the modified Milstein scheme should be preferred for most (if not all) Feller ratios under the short-rate model and Heston model, with a truncation correction applied when the Feller ratio is less than 0.5. Under these models, we would avoid picking the reflection scheme or the full truncation scheme due to the inconsistency of their results for very low Feller ratios. All of these recommendations, of course, are based solely from the empirical analysis results in Section 4.

Although a strongly convergent approximation scheme is available for the full parameter range, simulation of the CIR process still remains a challenge when the Feller ratio becomes very small. The downgrading of convergence for an approximation scheme when the Feller ratio decreases remains a key limitation of such schemes. In this paper, we have studied some of the available schemes with the aim of identifying those that perform best for different Feller ratios. Of course, an approximation scheme that has good convergence properties invariant to a decreasing Feller ratio is desirable and could motivate further work in this area of research. In fact, recent work by Hefter et. al. [13] has shown that any numerical approximation scheme for the CIR process that uses equally spaced Wiener increments can at best achieve a strong convergence rate of δ_0 when $\delta_0 < 1$. As a result, this essentially puts a bound on how good schemes of this form can be. Moreover, convergence rates for these schemes become arbitrarily slow as the Feller ratio approaches 0. Furthermore, the convergence rate proven by Hefter and Herzwurm [12] for the truncated Milstein scheme is close to optimal and perhaps this is why we saw its relatively good performance for very low Feller ratios in Section 4.1. Following on from this work, it remains an open question as to how best to simulate the CIR process under settings with a small Feller ratio.

References

[1] Hansjörg Albrecher, Philipp Arnold Mayer, Wim Schoutens, and Jurgen Tistaert, *The little Heston trap*, Wilmott **1** (2007).

[2] Aurélien Alfonsi, *Strong order one convergence of a drift implicit Euler scheme: Application to the CIR process*, Statistics and Probability Letters **83** (2013), no. 2.

[3] Aurélien Alfonsi et al., *Affine diffusions and related processes: simulation, theory and applications*, Vol. 6, Springer, 2015.

[4] Aurélien Alfonsi, *On the discretization schemes for the CIR (and Bessel squared) processes*, Monte Carlo Methods Appl (2005).

[5] Abdel Berkaoui, Mireille Bossy, and Awa Diop, *Euler scheme for SDEs with non-Lipschitz diffusion coefficient: strong convergence*, ESAIM: Probability and Statistics **12** (2008).

[6] Damiano Brigo and Aurélien Alfonsi, *Credit default swap calibration and derivatives pricing with the SSRD stochastic intensity model*, Finance and Stochastics **9** (2005), no. 1.

[7] J. Cox, J. Ingersoll, and S. Ross, *A theory of the term structure of interest rates, Econometrica 53, 385-407* (1985).

[8] Andrei Cozma and Christoph Reisinger, *Strong order 1/2 convergence of full truncation Euler approximations to the Cox–Ingersoll–Ross process*, IMA Journal of Numerical Analysis **40** (2018), no. 1.

[9] G. Deelstra and F. Delbaen, *Convergence of discretized stochastic (interest rate) processes with stochastic drift term*, Applied Stochastic Models and Data Analysis **14** (1998).

[10] Steffen Dereich, Andreas Neuenkirch, and Lukasz Szpruch, *An Euler-type method for the strong approximation of the Cox-Ingersoll-Ross process*, Proceedings: Mathematical, Physical and Engineering Sciences **468** (2012), no. 2140.

[11] István Gyöngy and Miklós Rásonyi, *A note on Euler approximations for SDEs with Hölder continuous diffusion coefficients*, Stochastic Processes and their Applications **121** (2011), no. 10.

[12] Mario Hefter and André Herzwurm, *Strong convergence rates for Cox-Ingersoll-Ross processes — Full parameter range*, Journal of Mathematical Analysis and Applications **2** (2018).

[13] Mario Hefter and Arnulf Jentzen, *On arbitrarily slow convergence rates for strong numerical approximations of Cox-Ingersoll-Ross processes and squared Bessel processes*, Finance and Stochastics **23** (2019), no. 1, 139–172.

[14] S. Heston, *A Closed-Form Solution for Options with Stochastic Volatility with Applications to Bond and Currency Options*, Review of Financial Studies **6** (1993), 327–343.

[15] D. Higham and X. Mao, *Convergence of Monte Carlo Simulations involving the Mean-Reverting Square Root Process*, Journal of Computational Finance **8** (2005).

[16] Martina Hofmanová and Jan Seidler, *On Weak Solutions of Stochastic Differential Equations*, Stochastic Analysis and Applications **30** (2012), no. 1, 100–121.

[17] Ioannis Karatzas and Steven Shreve, *Brownian Motion and Stochastic Calculus*, Graduate Texts in Mathematics, vol. 113, Springer New York, 2004 (eng).

[18] R. Lord, R. Koekkoek, and D. V. Dijk, *A comparison of biased simulation schemes for stochastic volatility models*, Quantitative Finance **10** (2010), 177 –194.

[19] Simon John A. Malham, Jiaqi Shen, and Anke Wiese, *Series Expansions and Direct Inversion for the Heston Model*, SIAM Journal on Financial Mathematics **12** (2021), no. 1, 487–549.

[20] Xuerong Mao, *Stochastic differential equations and applications*, 2nd ed., 2007.

[21] Andreas Neuenkirch and Lukasz Szpruch, *First order strong approximations of scalar SDEs defined in a domain*, Numerische Mathematik **128** (2014), no. 1.

[22] Anqi Shao, *A fast and exact simulation for CIR process By Anqi Shao, A dissertation presented to the Graduate School of the University of Florida in partial fulfillment of the requirements for the degree of Doctor of Philosophy University of Florida*, 2012.

[23] Toshio Yamada and Shinzo Watanabe, *On the uniqueness of solutions of stochastic differential equations*, Journal of Mathematics of Kyoto University **11** (1971), no. 1, 155 –167.

Appendices

A More Properties of the CIR Process

Some of the following results are sourced from [12].

- We recall that a χ^2-distributed random variable Y with δ_1 degrees of freedom has the probability density function

$$\begin{aligned} f_Y(y;\delta_1) &= \frac{y^{\delta_1/2-1}e^{-y/2}}{2^{\delta_1}/2\Gamma(\delta_1/2)} \cdot \mathbb{1}_{\{y\geq 0\}} \\ &\propto y^{\delta_1/2-1}e^{-y/2} \cdot \mathbb{1}_{\{y\geq 0\}}, \end{aligned} \tag{A.1}$$

  and we notice from (2.14) and (2.15) that

$$f_{X_t^0}(X;\delta_1) \propto \left(\frac{X}{\theta_k(t)}\right)^{\delta_1/2-1} e^{-\frac{1}{2}(\frac{X}{\theta_k(t)})},$$

  with θ_k defined according to (2.8). Hence, we have that

$$X_t^0 \overset{d}{=} \theta_k(t) \cdot Y. \tag{A.2}$$

- For all $0 \leq x \leq y$ and $t \geq 0$, we have

$$X_t^x \leq X_t^y. \tag{A.3}$$

  This follows directly from the comparison theorem for SDEs.

- From Section 2.2 we have

$$\mathbb{E}\Big[\frac{4}{\sigma^2\theta_k(t)}X_t^x\Big] = \frac{4}{\sigma^2\theta_k(t)}\mathbb{E}\Big[X_t^x\Big] = \frac{4a}{\sigma^2} + x\Big(\frac{4}{\sigma^2\theta_k(t)}e^{-kt}\Big),$$

  and thus

$$\mathbb{E}\Big[X_t^x\Big] = xe^{-kt} + a\theta_k(t). \tag{A.4}$$

B Important Results

Some of the following results are sourced from [20] and [12].

B.1 Theorems and Lemmas

Theorem B.1 (Lebesgue's Dominated Convergence Theorem)**.** *Let X_n be a sequence of random variables converging to some random variable X. Assume that there exists a (non-negative) random variable Y such that $\mathbb{E}\big[Y\big]$ such that $\mathbb{E}\big[Y\big] < \infty$ and $\big|X_n\big| \leq Y$ for all n. Then $\mathbb{E}\big[X\big] < \infty$ and $\mathbb{E}\big[X_n\big] \to \mathbb{E}\big[X\big]$ as $n \to \infty$.*

Theorem B.2 (Hölder's Inequality)**.** *For any random variables X and Y,*

$$\mathbb{E}\Big[\big|XY\big|\Big] \leq \Big(\mathbb{E}\Big[\big|X\big|^p\Big]\Big)^{1/p} \cdot \Big(\mathbb{E}\Big[\big|Y\big|^p\Big]\Big)^{1/q},$$

for all $1 \leq p < \infty$ with $q = \frac{p}{p-1}$.

Theorem B.3 (Minkowski's Inequality). *For any random variables X and Y,*

$$\Big(\mathbb{E}\Big[\big|X+Y\big|^p\Big]\Big)^{1/p} \le \Big(\mathbb{E}\Big[\big|X\big|^p\Big]\Big)^{1/p} + \Big(\mathbb{E}\Big[\big|Y\big|^p\Big]\Big)^{1/p},$$

for all $1 \le p < \infty$.

Theorem B.4 (Jensen's Inequality). *If f is a convex function on $\mathbb{R}$ and X is a random variable such that $\mathbb{E}\big|X\big| < \infty$, then*

$$f\Big(\mathbb{E}\big[X\big]\Big) \le \mathbb{E}\Big[f(X)\Big].$$

Theorem B.5 (Yamada-Watanabe Theorem). *Pathwise uniqueness implies uniqueness in the sense of probability law.*

Theorem B.6 (Burkholder-Davis Type Inequality). *Let $p \ge 2$ and $X = (X_t)_{t\ge 0}$ be an $\mathcal{F}_t$-adapted process such that*

$$\mathbb{E}\bigg[\int_0^T |X_t|^p dt\bigg] < \infty,$$

for every $T > 0$. Then there exists a constant $C > 0$ depending only on p and T such that

$$\mathbb{E}\bigg[\sup_{0\le t\le T}\Big|\int_0^t X_s \mathrm{d}W_s\Big|^p\bigg] \le C\mathbb{E}\bigg[\int_0^T |X_s|^p \mathrm{d}s\bigg].$$

Theorem B.7 (Markov's Inequality). *If X is a non-negative random variable and $c > 0$ is constant, then for all $1 \le p < \infty$,*

$$\mathbb{P}\Big(\big|X\big| \ge c\Big) \le \frac{\mathbb{E}\Big[\big|X\big|^p\Big]}{c^p}.$$

Lemma B.1. *Let Z be a standard normal random variable. Then for all $x_1, x_2 \ge 1$, we have*

$$\mathbb{E}\bigg[\Big|\Big(\max(1,\sqrt{x_1}+Z)\Big)^2 - \Big(\max(1,\sqrt{x_2}+Z)\Big)^2\Big|\bigg] \le |x_1 - x_2|.$$

Furthermore, for every $1 \le p < \infty$ and for all $x \ge 1$, there exists a constant $C > 0$ such that

$$\bigg(\mathbb{E}\bigg[\Big|\Big(\max(1,\sqrt{x}+Z)\Big)^2 - \Big(\sqrt{x}+Z\Big)^2\Big|^p\bigg]\bigg)^{1/p} \le C\frac{1}{\sqrt{x}}.$$

B.2 Other Results

Some of the following results are sourced from [3] and [17].

Definition B.1 (One-dimensional Strong Solution). *Let $(\Omega, \mathcal{F}, \mathbb{P})$ be a probability space and consider the SDE*

$$\mathrm{d}X_t = b_t(X_t)\mathrm{d}t + \sigma_t(X_t)\mathrm{d}W_t, \quad t \in [0,T], \quad X_0 = x, \tag{B.1}$$

where $b : \Omega \times [0,T] \times \mathbb{R}$, $\sigma : \Omega \times [0,T] \times \mathbb{R}$ and $W = (W_t)_{t\in[0,T]}$ is a Wiener process. Then we say X is a strong solution to (B.1) if:

1. *The process X is continuous on $[0,T]$ and adapted to the natural filtration $(\mathcal{F}_t)_{t\in[0,T]}$.*
2. *We have $\mathbb{P}\Big(\int_0^T \big|b_t(X_t)\big|\mathrm{d}t < \infty\Big) = 1$ and $\mathbb{P}\Big(\int_0^T \big|\sigma_t(X_t)\big|^2\mathrm{d}t < \infty\Big) = 1$.*
3. *The process X satisfies (B.1) almost surely for all $t \in [0,T]$.*

Definition B.2 (One-dimensional Weak Solution). *A weak solution to the SDE (B.1) is a triple (X, W), $(\Omega, \mathcal{F}, \mathbb{P})$, $(\mathcal{F}_t)_{t\in[0,T]}$ where:*

1. *$(\Omega, \mathcal{F}, \mathbb{P})$ and $(\mathcal{F}_t)_{t\in[0,T]}$ is a filtration of sub σ-algebras of $\mathcal{F}$.*

2. *$X = (X_t)_{t\in[0,T]}$ is a continuous, $\mathcal{F}_t$-adapted, $\mathbb{R}$-valued process, $W = (W_t)_{t\in[0,T]}$ is a Wiener process and 2, 3 from Definition B.1 are satisfied.*

χ^2-Distribution
A non-central χ^2-distributed random variable X with $d > 0$ degrees of freedom and non-centrality $\lambda \geq 0$ has a probability density given by

$$f_X(x) = \sum_{n=0}^{\infty} \frac{e^{-\lambda/2}(\lambda/2)^n}{n!} \frac{1/2}{\Gamma(n+d)} \big(\frac{x}{2}\big)^{n+d-1} e^{-x/2}, \;\; x > 0,$$

where $\Gamma(\alpha) = \int_0^\infty x^{\alpha-1}e^{-x}\mathrm{d}x, \alpha > 0$. It also has the characteristic function given by

$$\psi(z) = \mathbb{E}\Big[e^{izX}\Big] = (1-2iz)^{-d/2} \exp\Big(\frac{\lambda iz}{1-2iz}\Big).$$

Antithetic Variates
The method of antithetic variates consists of considering N_{MC} pairs $(Y_1, \bar{Y}_1), ..., (Y_{N_{MC}}, \bar{Y}_{N_{MC}})$ that are i.i.d. but for each $1 \leq i \leq N_{MC}$, Y_i and $\bar{Y}_i$ are identically distributed but not independent. Then the antithetic variate estimator for $\mathbb{E}\big[X\big]$ is given by

$$\frac{1}{N_{MC}} \sum_{i=1}^{N_{MC}} \Big(\frac{Y_i + \bar{Y}_i}{2}\Big).$$

The variance of this estimator is

$$\frac{\mathrm{Var}(Y) + \mathrm{Var}(\bar{Y}) + 2\mathrm{Cov}(Y, \bar{Y})}{4} = \frac{\mathrm{Var}(Y) + \mathrm{Cov}(Y, \bar{Y})}{2},$$

so the method will decrease variance if $\mathrm{Cov}(Y, \bar{Y}) < 0$. In the case of getting estimates for the CIR process, generating N_{MC} Wiener paths is required. The antithetic variate approach used here is then multiplying each Wiener path by -1 to get a negatively correlated corresponding Wiener path, giving us the desired i.i.d pairs of negatively correlated random variables.

C Additional Proofs

C.1 Proofs from Section 2.3

The following proof follows the approach in [3].

Proof of Lemma 2.1. Since X^x is continuous, we know by the intermediate value theorem that $X_t^x \in [\underline{m}, \overline{m}]$ for $t \leq \tau_{\underline{m},\overline{m}}$. By Itô's formula, we get

$$\begin{aligned} \mathrm{d}s(X_t^x) &= \sigma\sqrt{X_t^x}\, s'(X_t^x)\mathrm{d}W_t \\ \implies s(X^x_{t\wedge\tau_{\underline{m},\overline{m}}}) &= s(x) + \int_0^{t\wedge\tau_{\underline{m},\overline{m}}} \sigma\sqrt{X_u^x}\, s'(X_u^x)\mathrm{d}W_u. \end{aligned} \tag{C.1}$$

Since $X_t^x \in [\underline{m}, \overline{m}]$ for $t \leq \tau_{\underline{m},\overline{m}}$, we have $\mathbb{E}\Big[\int_0^{t\wedge\tau_{\underline{m},\overline{m}}} \sigma^2 X_u^x s'(X_u^x)^2 \mathrm{d}u\Big] < \infty$. Hence, $s(X^x_{t\wedge\tau_{\underline{m},\overline{m}}})$ is an almost surely bounded martingale and it converges almost surely as $t \to \infty$ (see [17], Chapter 1, Theorem 3.15). Thus, we have

$$\lim_{t\to\infty}\mathbb{E}\Big[\int_0^{t\wedge\tau_{\underline{m},\overline{m}}}\sigma^2 X_u^x s'(X_u^x)^2\mathrm{d}u\Big] = \mathbb{E}\Big[\int_0^{\tau_{\underline{m},\overline{m}}}\sigma^2 X_u^x s'(X_u^x)^2\mathrm{d}u\Big]$$
$$\geq \sigma^2 \min_{x\in[\underline{m},\overline{m}]} xs'(x)^2\mathbb{E}\Big[\int_0^{\tau_{\underline{m},\overline{m}}}\mathrm{d}u\Big]$$
$$= \sigma^2 \min_{x\in[\underline{m},\overline{m}]} xs'(x)^2\mathbb{E}\big[\tau_{\underline{m},\overline{m}}\big],$$

where we used Lebesgue's dominated convergence theorem. We know $\min_{x\in[\underline{m},\overline{m}]} xs'(x)^2 > 0$, so we have $\mathbb{E}\big[\tau_{\underline{m},\overline{m}}\big] < \infty$. Hence, $\mathbb{P}(\tau_{\underline{m},\overline{m}} < \infty) = 1$ and $t\wedge\tau_{\underline{m},\overline{m}} \xrightarrow[t\to\infty]{} \tau_{\underline{m},\overline{m}}$. From (C.1), we get

$$s(X^x_{\tau_{\underline{m},\overline{m}}}) = s(x) + \int_0^{\tau_{\underline{m},\overline{m}}}\sigma\sqrt{X_u^x}s'(X_u^x)\mathrm{d}W_u.$$

Using again the fact that $X_t^x \in [\underline{m},\overline{m}]$for $t \leq \tau_{\underline{m},\overline{m}}$ and that $s(X^x_{\tau_{\underline{m},\overline{m}}})$ is a martingale, we can write

$$s(x) = \mathbb{E}\big[s(X^x_{\tau_{\underline{m},\overline{m}}})\big] = s(\underline{m})\mathbb{P}(\tau_{\underline{m}} < \tau_{\overline{m}}) + s(\overline{m})\mathbb{P}(\tau_{\underline{m}} > \tau_{\overline{m}})$$
$$\iff \mathbb{P}(\tau_{\underline{m}} < \tau_{\overline{m}}) = \frac{s(\overline{m}) - s(x)}{s(\overline{m}) - s(\underline{m})}.$$

□

C.2 Proofs from Section 3.2

The following proofs follow the approaches in [12].

Proof of Proposition 3. Let Y be a random variable distributed according to a χ^2 distribution with δ_1 degrees of freedom and let $\epsilon \in (0, c]$ for some $c > 0$. Then using (A.1), we have

$$\mathbb{P}(Y \leq \epsilon) \propto \int_0^\epsilon y^{\delta_1/2-1}e^{-y/2}\mathbb{1}_{\{y\geq 0\}}\mathrm{d}y$$
$$\leq \int_0^\epsilon y^{\delta_1/2-1}\mathrm{d}y \tag{C.2}$$
$$= \frac{2}{\delta_1}\epsilon^{\delta_1/2}.$$

If $\delta_1 > 1$, we have

$$\mathbb{E}\Big[Y^{-1/2}\mathbb{1}_{\{Y\geq\epsilon\}}\Big] \propto \int_\epsilon^\infty y^{(\delta_1-1)/2-1}e^{-y/2}\mathrm{d}y \leq C\cdot 1$$

using the kernel of the Gamma density function. On the other hand, if $\delta_1 \leq 1$, we have

$$\mathbb{E}\Big[Y^{-1/2}\mathbb{1}_{\{Y\geq\epsilon\}}\Big] \leq \int_\epsilon^\infty y^{(\delta_1-1)/2-1}\mathrm{d}y \leq C\epsilon^{(\delta_1-1)/2}.$$

Finally if $\delta_1 = 1$, we have

$$\begin{aligned}\mathbb{E}\Big[Y^{-1/2}\mathbb{1}_{\{Y\geq\epsilon\}}\Big] &= \int_{\epsilon}^{\infty} y^{-1}e^{-y/2}\mathrm{d}y \\ &= \int_{\epsilon/2}^{\infty} y^{-1}e^{-y}\mathrm{d}y \\ &\leq \int_{\epsilon/2}^{1} y^{-1}\mathrm{d}y + \int_{1}^{\infty} e^{-y}\mathrm{d}y \\ &= \log(2) - \log(\epsilon) + e^{-1} \\ &\leq C\log(c) - \log(\epsilon) + 1.\end{aligned}$$

Putting these together we get

$$\mathbb{E}\Big[Y^{-1/2}\mathbb{1}_{\{Y\geq\epsilon\}}\Big] \leq C \cdot \begin{cases} 1, & \text{if } \delta_1 > 1, \\ 1 + \log(\frac{c}{\epsilon}), & \text{if } \delta_1 = 1, \\ \epsilon^{(\delta_1-1)/2}, & \text{if } \delta_1 < 1. \end{cases} \tag{C.3}$$

Furthermore, we have

$$\frac{t}{\theta_k(t)} \leq \begin{cases} |k|t, & \text{if } k \neq 0, \\ 1, & \text{if } k = 0, \end{cases}$$

for $t \in [0,1]$. Thus, we have

$$\theta_k(t) \geq \begin{cases} t/|k|, & \text{if } k \neq 0, \\ t, & \text{if } k = 0, \end{cases} \tag{C.4}$$

for $t \in [0,1]$.

For $s \leq t$, the claim for $X_s^x \leq t$ is trivial. For $t \leq X_s^x \leq 1$, we have

$$\begin{aligned}\mathbb{E}\Big[\Delta_{local}(X_s^x, t)\Big] &= \mathbb{E}\Big[t^{3/2}\sqrt{X_s^x}\Big] \\ &\leq \mathbb{E}\Big[t(1 + X_s^x)\Big] \\ &\leq t + tC(1+x) \ \ (\text{by } (2.6))\end{aligned}$$

and the claim follows. For $X_s^x \geq 1$, we have

$$\begin{aligned}\mathbb{E}\Big[\Delta_{local}(X_s^x, t)\Big] &= \mathbb{E}\Big[t^{3/2}X_s^x] \\ &\leq tC(1+x) \quad \text{from } (2.6),\end{aligned}$$

which proves the claim for $s \leq t$.

Now consider $s \geq t$ and let $U_1, U_2 : R_0^+ \times [0,1] \to R_0^+$ be given by

$$U_1(x,t) = xt^{3/2}$$

and

$$U_2(x,t) = \begin{cases} t, & \text{if } x \leq t, \\ t^{3/2}x^{-1/2} & \text{if } t \leq x, \end{cases}$$

and we have that

$$\Delta_{local}(x,t) \le U_1(x,t) + U_2(x,t). \tag{C.5}$$

Using (2.6), we get

$$\mathbb{E}\Big[U_1(X_s^x,t)\Big] \le Ct^{3/2}(1+x), \tag{C.6}$$

for $t \in (0,1], s \in [0,1]$ and $x \ge 0$. Using (A.2), (C.2), (C.3) and (C.4) we get

$$\begin{aligned} t\mathbb{P}\big(X_s^0 \le t\big) &= t\mathbb{P}\big(\theta_k(s)Y \le t\big) \\ &= t\mathbb{P}\Big(Y \le \frac{t}{\theta_k(s)}\Big) \\ &\le Ct\Big(\frac{t}{\theta_k(s)}\Big)^{\delta_1/2} \\ &\le Ct\Big(\frac{t}{s}\Big)^{\delta_1/2} \end{aligned}$$

for $0 < t \le s \le 1$ and

$$\begin{aligned} \mathbb{E}\Big[t^{3/2}(X_s^0)^{-1/2}\mathbb{1}_{\{X_s^0 \ge t\}}\Big] &= t^{3/2}\theta_k(s)^{-1/2}\mathbb{E}\Big[Y^{-1/2}\mathbb{1}_{\{Y \ge \frac{t}{\theta_k(s)}\}}\Big] \\ &\le C\frac{t^{3/2}}{\sqrt{s}} \cdot \begin{cases} 1, & \text{if } \delta_1 > 1, \\ 1+\log(\frac{s}{t}), & \text{if } \delta_1 = 1, \\ (t/s)^{(\delta_1-1)/2}, & \text{if } \delta_1 < 1. \end{cases} \end{aligned}$$

Notice that for $0 \le x_1 \le x_2$ and $t \in [0,1]$ we have $U_2(x_1,t) \ge U_2(x_2,t)$. Then by (A.3), we get

$$\begin{aligned} \mathbb{E}\Big[U_2(X_s^x,t)\Big] &\le \mathbb{E}\Big[U_2(X_s^0,t)\Big] \\ &= t\mathbb{P}(X_s^0) + \mathbb{E}\Big[t^{3/2}(X_s^0)^{-1/2}\mathbb{1}_{\{X_s^0 \ge t\}}\Big] \\ &\le C\Bigg(t(t/s)^{\delta_1/2} + t \cdot \begin{cases} (t/s)^{1/2}, & \text{if } \delta_1 > 1, \\ (t/s)^{1/2}(1+\log(\frac{s}{t})), & \text{if } \delta_1 = 1, \\ (t/s)^{\delta_1/2}, & \text{if } \delta_1 < 1, \end{cases}\Bigg) \\ &= C\Big(t(t/s)^{\delta_1/2} + t \cdot \Big((t/s)^{\min(\delta_1,1)/2}(1+\log(s/t)\mathbb{1}_{\{\delta_1=1\}})\Big)\Big). \end{aligned} \tag{C.7}$$

Finally, from (C.5), (C.6) and (C.7) we obtain

$$\begin{aligned} \mathbb{E}\Big[\Delta_{local}(X_s^x,t)\Big] &\le C\Big(t^{3/2}(1+x) + t(t/s)^{\delta_1/2}(1+x) + t(1+x)\Big((t/s)^{\min(\delta_1,1)/2}(1+\log(s/t)\mathbb{1}_{\{\delta_1=1\}})\Big)\Big) \\ &\le Ct(1+x)\Big(1 + (t/s)^{\min(\delta_1,1)/2}(2+\log(s/t)\mathbb{1}_{\{\delta_1=1\}})\Big) \\ &\le Ct(1+x)\Big(1 + (t/s)^{\min(\delta_1,1)/2}(1+\log(s/t)\mathbb{1}_{\{\delta_1=1\}})\Big) \\ &\le C(1+x)\Big((t/s)^{\min(\delta_1,1)/2}(1+\log(s/t)\mathbb{1}_{\{\delta_1=1\}})\Big) \\ &\quad \text{where we use that } t \le 1. \end{aligned}$$

□

Proof of Proposition 4. From (A.4), we have

$$\begin{aligned}\mathbb{E}\Big[X_t^x\Big] &\le C\cdot\begin{cases} xe^{|k|}+at, & \text{if } k=0,\\ xe^{|k|}+\frac{a}{|k|}(e^{|k|t}-1) & \text{if } k\neq 0,\end{cases}\\ &\le C\cdot\begin{cases} xe^{|k|}+at, & \text{if } k=0,\\ xe^{|k|}+\frac{a}{|k|}(e^{|k|}-1)t & \text{if } k\neq 0 \quad \text{by convexity of } \frac{a}{|k|}(e^{|k|t}-1) \text{ on } [0,1],\end{cases}\\ &\le C(x+t).\end{aligned} \tag{C.8}$$

Now we have

$$\begin{aligned}\Big(\mathbb{E}\Big[\sup_{0\le s\le t}|X_s^x|^2\Big]\Big)^{1/2} &\le \Big(2x^2+4\mathbb{E}\Big[\sup_{0\le s\le t}\Big|\int_0^s(\delta_1-kX_u^x)\mathrm{d}u\Big|^2\Big]+4\mathbb{E}\Big[\sup_{0\le s\le t}\Big|\int_0^s 2\sqrt{X_u^x}\mathrm{d}W_u\Big|^2\Big]\Big)^{1/2}\\ &\le \sqrt{2}x+2\Big(\mathbb{E}\Big[\sup_{0\le s\le t}\Big|\int_0^s(\delta_1-kX_u^x)\mathrm{d}u\Big|^2\Big]\Big)^{1/2}+2\Big(\mathbb{E}\Big[\sup_{0\le s\le t}\Big|\int_0^s 2\sqrt{X_u^x}\mathrm{d}W_u\Big|^2\Big]\Big)^{1/2}\\ &\quad \text{from Minkowski's inequality,}\\ &\le \sqrt{2}x+2\Big(\mathbb{E}\Big[\sup_{0\le s\le t}\Big|\int_0^s\mathrm{d}u\int_0^s(\delta_1-kX_u^x)^2\mathrm{d}u\Big|\Big]\Big)^{1/2}+2\Big(4\mathbb{E}\int_0^t 4X_u^x\mathrm{d}u\Big)^{1/2}\\ &\quad \text{from Theorem (B.6) and Hölder's inequality,}\\ &\le \sqrt{2}x+2\sqrt{t}\Big(\int_0^t 2\delta_1^2\mathrm{d}u+2k^2\int_0^t\mathbb{E}\Big[(X_u^x)^2\Big]\mathrm{d}u\Big)^{1/2}+8\Big(\int_0^t\mathbb{E}\big[X_u^x\big]\mathrm{d}u\Big)^{1/2}\\ &\le \sqrt{2}x+2\sqrt{t\Big(2\delta_1^2+2k^2t(1+x)^2\Big)}+8\sqrt{t(x+t)} \quad \text{from (C.8) and (2.6),}\\ &\le \sqrt{2}x+2t\Big(\sqrt{2}\delta_1+\sqrt{2}k(1+x)\Big)+8\sqrt{xt}+8t\\ &\le x\Big(\sqrt{2}(1+2k)+8/\sqrt{2}\Big)+t\Big(2\sqrt{2}(\delta_1+k)+8(1+1/\sqrt{2})\Big)\\ &\quad \text{using } t\le 1 \text{ and Young's inequality,}\\ &\le C(x+t).\end{aligned}$$

This calculation can also be done for $p=4,8,16,...$ and then Jensen's inequality can be applied (since the function $x\mapsto x^{p/q}$ is concave for any $x\ge 0$ and for any p and q such that $1\le p<q<\infty$) to get

$$\Big(\mathbb{E}\Big[\sup_{0\le s\le t}|X_s^x|^p\Big]\Big)^{1/p}\le C(x+t) \tag{C.9}$$

for $1\le p<\infty$. Now assume $2\le p<\infty$. Then using the same steps as above we obtain

$$\Big(\mathbb{E}\Big[\sup_{0\le s\le t}|X_s^x-x|^p\Big]\Big)^{1/p}\le Ct(1+x)+\sqrt{xt}, \tag{C.10}$$

for $x\ge 0$ and $t\in[0,1]$. Similarly, we get

$$\begin{aligned}\Big(\mathbb{E}\Big[|\phi_{MIL}(x,t,\Delta W^h)-x|^p\Big]\Big)^{1/p} &= \Big(\mathbb{E}\Big[|2\Delta W^h\sqrt{x}+(\Delta W^h)^2+t(\delta_1-1-kx)|^p\Big]\Big)^{1/p}\\ &\le C\Big((1+x)h+\sqrt{xh}\Big)\end{aligned} \tag{C.11}$$

for $x \geq 0$ and $h \in (0,1]$. Thus, for $x \leq h$ we have

$$\begin{aligned}
\Big(\mathbb{E}\Big[|\phi_{MIL}(x,t,\Delta W^h) - X_h^x|^p\Big]\Big)^{1/p} &= \Big(\mathbb{E}\Big[|\phi_{MIL}(x,t,\Delta W^h) - x + x - X_h^x|^p\Big]\Big)^{1/p} \\
&\leq \Big(\mathbb{E}\Big[|\phi_{MIL}(x,t,\Delta W^h) - x|^p\Big]\Big)^{1/p} + \Big(\mathbb{E}\Big[|x - X_h^x|^p\Big]\Big)^{1/p} \\
&\leq C\Big((1+x)h + \sqrt{xh}\Big) \\
&\leq C(h + h^2 + h) \\
&\leq Ch \ \text{ since } \ h \leq 1,
\end{aligned}$$

which yields the claim for $x \leq h$. Furthermore, according to (C.10) we have

$$\begin{aligned}
\Big(\mathbb{E}\Big[\Big|\int_0^t (\delta_1 - kX_s^x)\mathrm{d}s - \int_0^t (\delta_1 - kx)\mathrm{d}s\Big|^p\Big]\Big)^{1/p} &= |k|\Big(\mathbb{E}\Big[\Big|\int_0^t (X_s^x - x)\mathrm{d}s\Big|^p\Big]\Big)^{1/p} \\
&\leq |k|\Big(\mathbb{E}\Big[\Big|\int_0^t (\sup_{0\leq s\leq t} X_s^x - x)\mathrm{d}u\Big|^p\Big]\Big)^{1/p} \\
&= |k|t\Big(\mathbb{E}\Big[\sup_{0\leq s\leq t} |X_s^x - x|^p\Big]\Big)^{1/p} \\
&\leq C\Big(t^2(1+x) + t^{3/2}\sqrt{x}\Big) \\
&\leq C\Big(t^{3/2}(t^{1/2} + x + \sqrt{x})\Big) \\
&\leq C\Delta_{local}(x,t)
\end{aligned} \tag{C.12}$$

for $x \geq 0$ and $t \in (0,1]$. Now define the stopping time

$$\tau^x = \inf\{s \geq 0 : |X_s^x - x| = x/2\}$$

for $x \geq 0$. Using Markov's inequality we have

$$\mathbb{P}(\tau^x \leq h) = \mathbb{P}\Big(\sup_{0\leq s\leq h} |X_s^h - x| \geq x/2\Big) \leq \frac{\mathbb{E}\Big[\sup_{0\leq s\leq h} |X_s^h - x|^p\Big]}{(x/2)^p}$$

for $h \geq 0$ and $x > 0$. Applying (C.10), we get for $x \geq h$ and $h \in (0,1]$,

$$\begin{aligned}
\mathbb{P}(\tau^x \leq h) &\leq C\frac{(h(1+x) + \sqrt{xh})^p}{x^p} \\
&\leq 2^{p-1}C\frac{h^p(1+x)^p + (xh)^{p/2}}{x^p} \\
&\leq 2^{p-1}C\frac{h^p(2^{p-1}(1+x^p) + (xh)^{p/2}}{x^p} \\
&\leq 2^{2(p-1)}C\frac{h^p + (xh)^p + (xh)^{p/2}}{x^p} \\
&\leq 2^{2(p-1)}C\Big(1 + \frac{(xh)^p + (xh)^{p/2}}{x^p}\Big) \\
&\leq C_p\Big(\frac{(xh)^p + (xh)^{p/2}}{x^p}\Big) \\
&= C(h^p + (h/x)^{p/2}) \\
&\leq C\Big(\frac{2h^{p/2}}{\min(1, x^{p/2})}\Big).
\end{aligned}$$

Since this holds for all $2 \le p < \infty$, we can quadruple p and we have

$$\mathbb{P}(\tau^x \le h) \le C\Big(\frac{h^{2p}}{\min(1, x^{2p})}\Big) \tag{C.13}$$

for $x \ge h$ and $h \in (0, 1]$. Now define the stopped process $\tilde{X}^x = (\tilde{X}^x_t)_{t\ge 0}$ by $\tilde{X}^x_t := X^x_{t\wedge\tau^x}$. Then $\tilde{X}^x_t \in [x/2, 3x/2]$ for $t \ge 0$ and Itô's lemma gives

$$\begin{aligned}
&\mathrm{d}\Big(\sqrt{\tilde{X}^x_t}\Big) = \Big(\frac{(\delta_1 - 1) - k\tilde{X}^x_t}{2\sqrt{\tilde{X}^x_t}}\Big)\mathrm{d}t + \mathrm{d}W_t \\
\implies &\sqrt{\tilde{X}^x_t} = \sqrt{x} + \int_0^{t\wedge\tau^x} \frac{(\delta_1 - 1) - k\tilde{X}^x_s}{2\sqrt{\tilde{X}^x_s}}\mathrm{d}s + W_{t\wedge\tau^x}.
\end{aligned}$$

Hence,

$$\begin{aligned}
\Big(\mathbb{E}\Big[\Big|\sqrt{\tilde{X}^x_t} - (\sqrt{x} + W_{t\wedge\tau^x})\Big|^p\Big]\Big)^{1/p} &= \Big(\mathbb{E}\Big[\Big|\int_0^{t\wedge\tau^x} \frac{(\delta_1 - 1) - k\tilde{X}^x_s}{2\sqrt{\tilde{X}^x_s}}\mathrm{d}s\Big|^p\Big]\Big)^{1/p} \\
&\le \Big(\mathbb{E}\Big[\Big|\int_0^{t\wedge\tau^x} \frac{(\delta_1 - 1) - 3kx/2}{2\sqrt{3x/2}}\mathrm{d}s\Big|^p\Big]\Big)^{1/p} \\
&= (t\wedge\tau^x)\Big(\frac{(\delta_1 - 1) - 3kx/2}{2\sqrt{3x/2}}\Big) \\
&\le t\Big(\frac{(\delta_1 - 1) + 3|k|x/2}{2\sqrt{3x/2}}\Big) \\
&\le C\frac{t(1+x)}{\sqrt{x}},
\end{aligned} \tag{C.14}$$

and

$$
\begin{aligned}
\Big(\mathbb{E}\Big[\Big|\sqrt{X_t^x}-(\sqrt{x}+W_h)\Big|^p\cdot\mathbb{1}_{\{\tau^x\le h\}}\Big]\Big)^{1/p} &\le \bigg(\Big(\mathbb{E}\Big[\Big|\sqrt{X_t^x}-(\sqrt{x}+W_h)\Big|^{2p}\Big]\Big)^{1/2}\Big(\mathbb{E}\Big[\mathbb{1}_{\{\tau^x\le h\}}\Big]\Big)^{1/2}\bigg)^{1/p}\\
&\quad \text{from Hölder's inequality,}\\
&= \Big(\mathbb{E}\Big[\Big|\sqrt{X_t^x}-(\sqrt{x}+W_h)\Big|^{2p}\Big]\Big)^{\frac{1}{2p}}\Big(\mathbb{P}(\tau^x\le h)\Big)^{\frac{1}{2p}}\\
&\le C\Big(\frac{h}{\min(1,x)}\Big)\bigg(\Big(\mathbb{E}\Big[|X_h^x|^p\Big]\Big)^{\frac{1}{p}\cdot\frac{1}{2}}+\Big(\mathbb{E}\Big[|\sqrt{x}+W_h|^{2p}\Big]\Big)^{\frac{1}{2p}}\bigg)\\
&\quad \text{from (C.13) and Minkowski's inequality,}\\
&\le C\Big(\frac{h}{\min(1,x)}\Big)\bigg(\sqrt{x+h}+\sqrt{x}+\Big(\mathbb{E}\Big[|W_h|^{2p}\Big]\Big)^{\frac{1}{2p}}\bigg)\\
&\quad \text{from Minkowski's inequality,}\\
&\le C\Big(\frac{h}{\min(1,x)}\Big)\bigg(\sqrt{x+h}+\sqrt{x}+\sqrt{h}\Big(\mathbb{E}\Big[|Z|^{2p}\Big]\Big)^{\frac{1}{2p}}\bigg),\\
&\quad \text{where } Z\sim N(0,1),\\
&= C\Big(\frac{h}{\min(1,x)}\Big)\bigg(\sqrt{x+h}+\sqrt{x}+\sqrt{h}\Big(2^{-p/2}\frac{p!}{(p/2)!}\Big)^{\frac{1}{2p}}\bigg)\\
&\quad \text{since 2p is even,}\\
&\le Ch\Big(\frac{\sqrt{x+h}+\sqrt{x}+\sqrt{h}}{\min(1,x)}\Big)\\
&\le Ch\Big(\frac{\sqrt{2x}+\sqrt{x}+\sqrt{x}}{\min(1,x)}\Big)\\
&\le Ch\frac{\sqrt{x}}{\min(1,x)}\\
&\le Ch\frac{1+x}{\sqrt{x}},
\end{aligned}
\tag{C.15}
$$

for $x\ge h$ and $h\in(0,1]$. Moreover, using Minkowski's inequality, (C.14) and (C.15) we have

$$
\begin{aligned}
\Big(\mathbb{E}\Big[\Big|\sqrt{X_t^x}-(\sqrt{x}+W_h)\Big|^p\Big]\Big)^{1/p} &\le \Big(\mathbb{E}\Big[\Big|\sqrt{X_t^x}-(\sqrt{x}+W_h)\Big|^p\cdot\mathbb{1}_{\{\tau^x\le h\}}\Big]\Big)^{1/p}\\
&\quad+\Big(\mathbb{E}\Big[\Big|\sqrt{X_t^x}-(\sqrt{x}+W_h)\Big|^p\cdot\mathbb{1}_{\{\tau^x> h\}}\Big]\Big)^{1/p}\\
&= \Big(\mathbb{E}\Big[\Big|\sqrt{X_t^x}-(\sqrt{x}+W_h)\Big|^p\cdot\mathbb{1}_{\{\tau^x\le h\}}\Big]\Big)^{1/p}\\
&\quad+\Big(\mathbb{E}\Big[\Big|\sqrt{\tilde{X}_t^x}-(\sqrt{x}+W_{t\wedge\tau^x})\Big|^p\Big]\Big)^{1/p}\\
&\le Ch\Big(\frac{1+x}{\sqrt{x}}\Big),
\end{aligned}
\tag{C.16}
$$

for $x \geq h$ and $h \in (0,1]$. Finally, we have

$$\begin{aligned}
\Big(\mathbb{E}\Big[|\phi(x,h,W_h) - X_h^x|^p\Big]\Big)^{1/p} &= \Big(\mathbb{E}\Big[|x + \int_0^h (\delta_1 - kX_s^x)\mathrm{d}s + \int_0^h 2\sqrt{X_s^x}\mathrm{d}W_s \\
&\qquad - (x + 2\sqrt{x}W_h + W_h^2 + h(\delta_1 - 1 - kx))|^p\Big]\Big)^{1/p} \\
&= \Big(\mathbb{E}\Big[\Big|\int_0^h (\delta_1 - kX_s^x)\mathrm{d}s - \int_0^h (\delta_1 - kx)\mathrm{d}s \\
&\qquad + \int_0^h \Big(2\sqrt{X_s^x} - 2(\sqrt{x} + W_s)\Big)\mathrm{d}W_s\Big|^p\Big]\Big)^{1/p} \\
&\leq \Big(\mathbb{E}\Big[\Big|\int_0^h (\delta_1 - kX_s^x)\mathrm{d}s - \int_0^h (\delta_1 - kx)\mathrm{d}s\Big|^p\Big]\Big)^{1/p} \\
&\quad + \Big(\mathbb{E}\Big[\Big|\int_0^h \Big(2\sqrt{X_s^x} - 2(\sqrt{x} + W_s)\Big)\mathrm{d}W_s\Big|^p\Big]\Big)^{1/p} \quad \text{from Minkowski's inequality,} \\
&\leq \Big(\mathbb{E}\Big[\Big|\int_0^h (\delta_1 - kX_s^x)\mathrm{d}s - \int_0^h (\delta_1 - kx)\mathrm{d}s\Big|^p\Big]\Big)^{1/p} \\
&\quad + \Big(\mathbb{E}\Big[\sup_{0\leq s\leq h}\Big|\int_0^s \Big(2\sqrt{X_u^x} - 2(\sqrt{x} + W_u)\Big)\mathrm{d}W_u\Big|^p\Big]\Big)^{1/p} \\
&\leq C\Delta_{local}(x,t) + \Big(\Big(\frac{p^3}{2(p-1)}\Big)^{p/2} h^{\frac{p-2}{2}} \int_0^h \mathbb{E}\Big|2\sqrt{X_u^x} - 2(\sqrt{x} + W_u)\Big|^p \mathrm{d}u\Big)^{1/p} \\
&\qquad \text{from Theorem (B.6) and (C.12),} \\
&\leq C\Big(\Delta_{local}(x,t) + h^{\frac{p-2}{2p}}\Big(h^{1/p} h(\frac{1+x}{\sqrt{x}})\Big)\Big) \quad \text{from (C.16),} \\
&= C\Big(\Delta_{local}(x,t) + h^{\frac{3}{2}}(\frac{1+x}{\sqrt{x}})\Big) \\
&\leq C\Delta_{local}(x,t),
\end{aligned}$$

which completes the proof. □

Proof of Lemma 3.1. Using Theorem B.1 and by writing

$$\nu(x,h,w) = h\Big(\sqrt{x/h} + w/\sqrt{h}\Big)^2$$

and

$$\hat{\nu}(x,h,w) = h\Big(\max\Big(1, \sqrt{\max(1, x/h)} + w/\sqrt{h}\Big)\Big)^2,$$

we get

$$\begin{aligned}
\mathbb{E}\Big|\hat{\nu}(x_1,h,W_h) - \hat{\nu}(x_2,h,W_h)\Big| &= h\mathbb{E}\Big|\Big(\max\Big(1, \sqrt{\max(1, x_1/h)} + W_h/\sqrt{h}\Big)\Big)^2 \\
&\qquad - \Big(\max\Big(1, \sqrt{\max(1, x_1/h)} + W_h/\sqrt{h}\Big)\Big)^2\Big| \\
&\leq h\Big|\max(1, x_1/h) - \max(1, x_2/h)\Big| \\
&\leq |x_1 - x_2|.
\end{aligned}$$

Again using Lemma B.1 we have

$$\begin{aligned}\left(\mathbb{E}\left[\left|\nu(x,h,W_h)-\hat{\nu}(x,h,W_h)\right|^p\right]\right)^{1/p} &= h\Bigg(\mathbb{E}\Bigg[\Bigg|\left(\sqrt{x/h}+W_h/\sqrt{h}\right)^2 \\ &\qquad -\left(\max\left(1,\sqrt{\max(1,x/h)}+W_h/\sqrt{h}\right)\right)^2\Bigg|^p\Bigg]\Bigg)^{1/p} \\ &\leq Ch\frac{1}{\sqrt{x/h}} \\ &= Ch^{3/2}\frac{1}{\sqrt{x}} \\ &\leq C\Delta_{local}(x,h),\end{aligned}$$

for $x \geq h$. Finally, by Minkowski's inequality we have

$$\begin{aligned}\left(\mathbb{E}\left[\left|\nu(x,h,W_h)-\hat{\nu}(x,h,W_h)\right|^p\right]\right)^{1/p} &\leq \left(\mathbb{E}\left[\left|\nu(x,h,W_h)\right|^p\right]\right)^{1/p}+\left(\mathbb{E}\left[\left|\hat{\nu}(x,h,W_h)\right|^p\right]\right)^{1/p} \\ &= h\left(\mathbb{E}\left[\left|\left(\sqrt{x/h}+W_h/\sqrt{h}\right)^2\right|^p\right]\right)^{1/p} \\ &\quad + h\left(\mathbb{E}\left[\left|\left(\max\left(1,\sqrt{\max(1,x/h)}+W_h/\sqrt{h}\right)\right)^2\right|^p\right]\right)^{1/p} \\ &\leq h\left(\mathbb{E}\left[\left|2+2W_h^2/h\right|^p\right]\right)^{1/p}+h\left(\mathbb{E}\left[\left|2+2W_h^2/h\right|^p\right]\right)^{1/p} \\ &= 2h\left(2+2\left(\mathbb{E}\left[|Z|^{2p}\right]\right)^{1/p}\right), \quad \text{where Z} \sim \text{N}(0,1), \\ &= 2h\left(2+2\left(2^{-p}\frac{(2p)!}{p!}\right)^{1/p}\right) \\ &\leq Ch = C\Delta_{local}(x,h),\end{aligned}$$

for $x \leq h$. □

D Code for Empirical Analysis

- Python was the chosen language for this study. The following code was run to install the necessary packages.

```
import time
import numpy as np
import matplotlib.pyplot as plt
```

- The following helper functions were defined.

```
def Wiener_inc(dt, N_MC=2000, T=1, ro=0):
    '''
    Generates Wiener increments into an array of size T/dt x N_MC.
    :param dt: (float) timestep size.
    :param N_MC: (int) number of Monte Carlo samples.
    :param T: (float) terminal time.
    :param ro: (float) correlation of Wiener processes to implement. Between -1
        and 1.
    :return: (ndarray) array of Wiener increments.
    '''

```

```
11          # simulate Wiener process
12          dW1 = np.zeros([N_MC, int(T / dt)])
13          dW1[:, 1:] = np.sqrt(dt) * np.random.standard_normal(size=(N_MC, int(T / dt)
                - 1))
14
15          if ro != 0:
16              # simulate correlated Wiener process
17              dW2 = np.zeros([N_MC, int(T / dt)])
18              dW2[:, 1:] = np.sqrt(dt) * np.random.standard_normal(size=(N_MC, int(T /
                    dt) - 1))
19              return ro * dW1 + np.sqrt(1 - ro**2) * dW2, dW1
20
21          else:
22              return dW1
23
24
25      def prob_of_nonpositive(X):
26          '''
27          Calculates probability of scheme attaining 0 or going negative at time T=1.
28          :param X: (ndarray) Simulated CIR processes.
29          :return:
30              (float) probability of attaining 0.
31              (float) probability of going negative.
32          '''
33
34          negative_prob = len([1 for i in range(X.shape[0]) if X[i, -1] < 0]) /
                X.shape[0]
35          zero_prob = len([1 for i in range(X.shape[0]) if X[i, -1] == 0]) / X.shape[0]
36
37          return zero_prob, negative_prob
38
39
40      def True_Bond_price(T=1, sigma=0.2, a=0.1, k=0.5, x=0.1):
41          '''
42          Calculates true ZCB price for under CIR short-rate model.
43          :param T: (float) Terminal time
44          :param sigma: (nonnegative float) parameter for CIR process
45          :param a: (nonnegative float) parameter for CIR process
46          :param k: (nonnegative float) parameter for CIR process
47          :param x: (nonnegative float) initial value for CIR process
48          :return: (float): true ZCB price
49          '''
50
51          # use analytical formula
52          h = np.sqrt(k ** 2 + 2 * (sigma ** 2))
53          A = ((2 * h * np.exp((k + h) * T / 2)) / (2 * h + (k + h)*(np.exp(h * T) -
                1))) ** (2 * a / (sigma ** 2))
54          B = (2 * (np.exp(h * T) - 1)) / (2*h + (k + h) * (np.exp(h * T) - 1))
55
56          return A * np.exp(-B * x)
57
58
59      def line_of_best_fit(X, Y, degree=1):
60          '''
61          Calculates line of best fit for polynomial of given degree.
62          :param X: (ndarray) input x values to use for regression.
63          :param Y: (ndarray) corresponding y values to use for regression.
64          :param degree: (int) degree of polynomial to fit.
65          :return:
66              (float) first coefficient of fitted line. For degree=1, this is the slope
                    of the line.
```

```
        (function): predictor function corresponding to line of best fit
    '''

    coeff = np.polyfit(X, Y, degree)
    return coeff[0], np.poly1d(coeff)


# Functions for Calculating true Heston call option price
def d(phi, ro, b, u, sigma):
    return np.sqrt((complex(0, ro * sigma * phi) - b) ** 2 - sigma**2 *
        (complex(0, 2 * u * phi) - phi**2))


def g(phi, ro, b, u, sigma):
    return (b - complex(0, ro * sigma * phi) - d(phi, ro, b, u, sigma)) / (b -
        complex(0, ro * sigma * phi) + d(phi, ro, b, u, sigma))


def C(phi, T, r, ro, a, sigma, b, u):
    return complex(0, r * phi * T) + a / (sigma**2) * (T * (b - complex(0, ro *
        sigma * phi) - d(phi, ro, b, u, sigma))
    - 2 * np.log((1 - g(phi, ro, b, u, sigma) * np.exp(-T * d(phi, ro, b, u,
        sigma))) / (1 - g(phi, ro, b, u, sigma))))


def D(phi, T, ro, sigma, b, u):
    return ((b - complex(0, ro * sigma * phi) - d(phi, ro, b, u, sigma)) /
        (sigma**2)) * ((1 - np.exp(-T * d(phi, ro, b, u, sigma))) / (1 - g(phi,
        ro, b, u, sigma) * np.exp(-T * d(phi, ro, b, u, sigma))))


def psi(phi, T, r, ro, a, sigma, b, u, S0, v0):
    return np.exp(C(phi, T, r, ro, a, sigma, b, u) + v0 * D(phi, T, ro, sigma, b,
        u) + complex(0, S0 * phi))


def Heston_call_option(ro, r, S0, K, v0, T, a, sigma, k):
    '''
    Calculates true price of call option under Heston model.
    :param ro: (float) correlation coefficient for Wiener processes.
    :param r: (float) risk-free interest rate.
    :param S0: (float) initial stock price.
    :param K: (float) strike price.
    :param v0: (float) initial volatility.
    :param T: (float) time to maturity.
    :param a: (float) parameter for CIR process.
    :param sigma: (float) parameter for CIR process.
    :param k: (float) parameter for CIR process.
    :return: (float) call option price.
    '''

    b = [k - ro * sigma, k]
    u = [1/2, -1/2]

    # define range to integrate over
    phi_values, step_size = np.linspace(0.0000001, 175, 8000, endpoint=True,
        retstep=True)

    P1 = 1/2
    P2 = 1/2

```

```
        # if P1 or P2 doesnt change by significant amount, break
        for i in range(1, len(phi_values)):
            if np.round(np.abs(step_size / 2 * ((np.exp(complex(0, -phi_values[i-1] *
                np.log(K))) * psi(phi=phi_values[i-1], T=T, r=r, ro=ro, a=a,
                sigma=sigma, b=b[0], u=u[0], S0=np.log(S0), v0=v0) / complex(0,
                phi_values[i - 1])).real
            + (np.exp(complex(0, -phi_values[i] * np.log(K))) * psi(phi=phi_values[i],
                T=T, r=r, ro=ro, a=a, sigma=sigma, b=b[0], u=u[0], S0=np.log(S0),
                v0=v0) / complex(0, phi_values[i])).real) / np.pi), 15) > 0:
                P1 += step_size / 2 * ((np.exp(complex(0, -phi_values[i-1] * np.log(K)))
                    * psi(phi=phi_values[i-1], T=T, r=r, ro=ro, a=a, sigma=sigma,
                    b=b[0], u=u[0], S0=np.log(S0), v0=v0) / complex(0, phi_values[i -
                    1])).real
                + (np.exp(complex(0, -phi_values[i] * np.log(K))) *
                    psi(phi=phi_values[i], T=T, r=r, ro=ro, a=a, sigma=sigma, b=b[0],
                    u=u[0], S0=np.log(S0), v0=v0) / complex(0, phi_values[i])).real) /
                    np.pi

                y1 = 0

            else:
                y1 = 1

            if np.round(np.abs(step_size / 2 * ((np.exp(complex(0, -phi_values[i - 1] *
                np.log(K))) * psi(phi=phi_values[i - 1], T=T, r=r, ro=ro, a=a,
                sigma=sigma, b=b[1], u=u[1], S0=np.log(S0), v0=v0) / complex(0,
                phi_values[i - 1])).real
            + (np.exp(complex(0, -phi_values[i] * np.log(K))) * psi(phi=phi_values[i],
                T=T, r=r, ro=ro, a=a, sigma=sigma, b=b[1], u=u[1], S0=np.log(S0),
                v0=v0) / complex(0, phi_values[i])).real) / np.pi), 15) > 0:
                P2 += step_size / 2 * ((np.exp(complex(0, -phi_values[i - 1] *
                    np.log(K))) * psi(phi=phi_values[i - 1], T=T, r=r, ro=ro, a=a,
                    sigma=sigma, b=b[1], u=u[1], S0=np.log(S0), v0=v0) / complex(0,
                    phi_values[i-1])).real
                + (np.exp(complex(0, -phi_values[i] * np.log(K))) *
                    psi(phi=phi_values[i], T=T, r=r, ro=ro, a=a, sigma=sigma, b=b[1],
                    u=u[1], S0=np.log(S0), v0=v0) / complex(0, phi_values[i])).real) /
                    np.pi

                y2 = 0

            else:
                y2 = 1

            if y1 and y2:
                break

        return S0 * P1 - K * np.exp(-r * T) * P2
```

- The following structure of code was used to implement the simulation of each numerical approximation scheme. The example used here is for the reflection scheme.

```
    def Reflection(dW, Antithetic=False, x=0.1, N_MC=2000, T=1, a=0.1, sigma=0.2,
        k=0.5):
        '''
        Implements explicit scheme for CIR process that reflects the process when it
            goes lower than 0.
        :param dW: (ndarray) Wiener increments to use for simulation
        :param Antithetic: (bool) whether to implement antithetic variate technique
            or not
```

```
        :param x: (nonnegative float) initial value for CIR process
        :param N_MC: (int) Number of Monte Carlo samples
        :param T: (float) Terminal time
        :param a: (float) parameter for CIR process
        :param sigma: (float) parameter for CIR process
        :param k: (float) parameter for CIR process
        :return: X (ndarray) values for simulated CIR processes
        '''

        N = dW.shape[1]

        if not Antithetic:
            # initialize process
            X = np.zeros([N_MC, N])
            X[:, 0] = x

            # simulate and time process
            t0 = time.time()
            for i in range(1, N):
                X[:, i] = np.abs(X[:, (i - 1)] + (a - k * X[:, (i - 1)]) * (T / N) +
                    sigma * np.sqrt(X[:, (i - 1)]) * dW[:, i])
            t1 = time.time()

            return X, t1 - t0

        else:
            # simulate correlated processes for antithetic variates
            X1 = np.zeros([N_MC, N])
            X2 = np.zeros([N_MC, N])
            X1[:, 0] = x
            X2[:, 0] = x
            dW2 = -dW
            # simulate and time process
            t0 = time.time()
            for i in range(1, N):
                X1[:, i] = np.abs(X1[:, (i - 1)] + (a - k * X1[:, (i - 1)]) * (T / N) +
                    sigma * np.sqrt(X1[:, (i - 1)]) * dW[:, i])
                X2[:, i] = np.abs(X2[:, (i - 1)] + (a - k * X2[:, (i - 1)]) * (T / N) +
                    sigma * np.sqrt(X2[:, (i - 1)]) * dW2[:, i])
            t1 = time.time()

            return X1, X2, t1 - t0
```

D.1 Code for Section 4.1

```
# function for calculating strong error for given scheme
def strong_error_analysis(dW, true_X1, true_X2, scheme, max_timestep=1/16,
    min_timestep_index=14, T=1, N_MC=2000, x=0.1, a=0.1, sigma=0.2, k=0.5):
    '''
    Calculates Strong error approximations for a given scheme. Automatically assumes use
        of antithetic variate technique. If not used, enter 0 for true_X2.
    :param scheme: (function.__name__) numerical scheme to implement.
    :param max_timestep: (float) biggest timestep to use
    :param min_timestep_index: (int) index which makes largest timestep 2 **
        (-min_timestep_index)
    :param T: (float) terminal time
    :param N_MC: (int) number of Monte Carlo samples
    :param x: (float) initial value for CIR process
    :param a: (float) parameter for CIR process
```

```
12        :param sigma: (float) parameter for CIR process
13        :param k: (float) parameter for CIR process
14        :return:
15           (ndarray) log of L^2 Error for each timestep
16           (list) runtimes for each simulation
17           (float) Monte Carlo error for final simulation
18        '''
19        # get list of timesteps to use for each simulation
20        dt = [2 ** (-i) for i in range(1, min_timestep_index + 1) if 2 ** (-i) <=
              max_timestep]
21        true_dt = T / dW.shape[1] # timestep for approximated true solution
22
23        strong_err = []
24        run_time = []
25
26        # join Wiener processes depending on timestep size
27        for i in range(len(dt)):
28           step = int(dt[i] / true_dt)
29           new_dW = np.transpose([np.sum(dW[:, (step*j):(step*(j + 1))], axis=1) for j in
                 range(int((true_dt**(-1)) / step))])
30
31        # simulate and time
32        X1, X2, t = scheme(new_dW, Antithetic=True, x=x, N_MC=N_MC, T=T, a=a, sigma=sigma,
              k=k)
33
34        run_time.append(t)
35
36        # calculate RMSE using antithetic variates
37        strong_err.append(np.sqrt(np.mean(((true_X1[:, -1] - X1[:, -1]) ** 2) / 2 +
              ((true_X2[:, -1] - X2[:, -1]) ** 2) / 2)))
38
39        # calculate Monte Carlo error
40        MC_err = np.log(np.sqrt(np.var((X1[:, -1] / 2 + X2[:, -1] / 2))/ N_MC))
41
42        return np.log(strong_err), run_time, MC_err
43
44
45        # plot results for Figure 3
46        # set parameters
47        x = 0.03
48        a = 0.02
49        FR = [0.25, 0.45, 0.75, 1.15] # Feller ratios
50        k = 0.4
51        T = 1
52        max_index = 14 # approximated true solution will have 2**(-14) timestep size
53        N_MC = 20000  # number of Monte Carlo simulations
54
55        # colors for plot
56        colors = {Partial_Truncation: 'orange', Full_Truncation: 'blue', Partial_Reflection:
              'black', Reflection: 'red', Implicit_1: 'grey', Implicit_2: 'purple',
57           Modified_Milstein: 'green', Adjusted_Modified_Milstein: 'green',
                 Truncated_Milstein: 'palevioletred'}
58
59        # labels for legend
60        label_list = {Partial_Truncation: 'Partial Truncation', Full_Truncation: 'Full
              Truncation', Partial_Reflection: 'Partial Reflection', Reflection: 'Reflection',
              Implicit_1: 'Implicit 1',
61           Implicit_2: 'Implicit 2', Modified_Milstein: 'Modified Milstein',
62           Adjusted_Modified_Milstein: '(Truncated) Modified Milstein', Truncated_Milstein:
                 'Truncated Milstein'}
63
```

```
    fig, ax = plt.subplots(nrows=2, ncols=2, figsize=(15, 15))
    axes = [ax[0, 0], ax[0, 1], ax[1, 0], ax[1, 1]]

    # simulate Wiener increments
    dW = Wiener_inc(2 ** (-max_index), N_MC=N_MC, T=T)

    # run simulations and plot results
    for i in range(len(FR)):
        sigma = np.sqrt(2 * a / FR[i]) # change sigma
        true_X1, true_X2, _ = Truncated_Milstein(dW, Antithetic=True, x=x, N_MC=N_MC, T=T,
            a=a, sigma=sigma, k=k) # get approximate true solution

        if FR[i] >= 1:
            scheme_list = [Partial_Truncation, Full_Truncation, Partial_Reflection,
                Reflection, Implicit_1, Implicit_2,
            Modified_Milstein, Truncated_Milstein]

        elif FR[i] >= 0.5:
            scheme_list = [Partial_Truncation, Full_Truncation, Partial_Reflection,
                Reflection, Implicit_2, Modified_Milstein,
            Truncated_Milstein]

        else:
            scheme_list = [Partial_Truncation, Full_Truncation, Partial_Reflection,
                Reflection, Adjusted_Modified_Milstein,
            Truncated_Milstein]

        MC_errs = []
        runtimes = []

        for scheme in scheme_list:
            err, runtime, MC_err = error_analysis(dW, true_X1, true_X2, scheme,
                max_timestep=1/2, min_timestep_index=max_index-1, T=T, N_MC=N_MC, x=x, a=a,
                sigma=sigma, k=k)
            MC_errs.append(MC_err)
            runtimes.append(runtime)
            slope, line = line_of_best_fit(np.log(runtime), err)
            print(str(FR[i]) + scheme.__name__ + str(-slope)) # print order of convergence
            axes[i].plot(np.log(runtime),
            line(np.log(runtime)), '-.', label=label_list[scheme], color=colors[scheme])
            axes[i].plot(np.log(runtime),
            err, 'o', ms=4, color=colors[scheme])


        axes[i].set_xlabel('Run time in Seconds (Log Scale)')
        axes[i].legend()
        axes[i].grid()
        axes[i].set_title(f'Feller Ratio = {FR[i]}, Average Log MC Error =
            {np.round(np.mean(MC_errs), 2)}')
        axes[i].set_ylabel('RMSE (Log Scale)')

    plt.savefig('Strong Convergence')


    # plot results for Figure 4
    # set parameters
    a = 0.02
    k = 0.4
    x = 0.03
    T = 1
    max_index = 14
```

```
N_MC = 20000
Feller_Ratios = [0.2, 0.3, 0.35, 0.4, 0.5]

schemes = [Truncated_Milstein, Adjusted_Modified_Milstein]
dW = Wiener_inc(2 ** (-max_index), N_MC=N_MC, T=T)

fig = plt.figure(figsize=(10, 10))
orders_1 = []
orders_2 = []

# same as before
for i in range(len(Feller_Ratios)):
   sigma = np.sqrt(2 * a / Feller_Ratios[i]) # change sigma
   true_X1, true_X2, _ = Truncated_Milstein(dW, Antithetic=True, x=x, N_MC=N_MC, T=T,
       a=a, sigma=sigma, k=k)
   err, runtime, _, _, _ = error_analysis(dW, true_X1, true_X2, schemes[0],
       max_timestep=1/2, min_timestep_index=max_index - 3, T=T, N_MC=N_MC, x=x, a=a,
   sigma=sigma, k=k)

   slope, _ = line_of_best_fit(np.log(runtime), err)
   orders_1.append(-slope)

   err, runtime, _, _, _ = error_analysis(dW, true_X1, true_X2, schemes[1],
       max_timestep=1 / 2,
   min_timestep_index=max_index - 2, T=T, N_MC=N_MC, x=x, a=a,
   sigma=sigma, k=k)

   slope, _ = line_of_best_fit(np.log(runtime), err)
   orders_2.append(-slope)
   if i == 0:
      orders_1.append(0.384)
      orders_2.append(0.349)

   elif i == 3:
      orders_1.append(0.436)
      orders_2.append(0.445)

Feller_Ratios = [0.2, 0.25, 0.3, 0.35, 0.4, 0.45, 0.5]
_, line = line_of_best_fit(Feller_Ratios, orders_1)
plt.plot(Feller_Ratios, line(Feller_Ratios), '-.', color='palevioletred',
    label='Truncated Milstein')
plt.plot(Feller_Ratios, orders_1, 'o', color='palevioletred')

_, line = line_of_best_fit(Feller_Ratios, orders_2)
plt.plot(Feller_Ratios, line(Feller_Ratios), '-.', color='green', label='(Truncated)
    Modified Milstein')
plt.plot(Feller_Ratios, orders_2, 'o', color='green')

plt.title('Truncated Milstein Scheme vs. Modified Milstein Scheme Truncated at 0')
plt.xlabel('Feller Ratio')
plt.ylabel('Observed Order of Convergence')
plt.legend()
plt.grid()

plt.savefig('Order of Convergence, FR < ' + str(0.5) + '.png')


# plot results for Figure 5
# set parameters
x = 0.03
a = 0.02
```

```
FR = [0.25, 0.45, 0.75, 1.15]
k = 0.4
T = 1
max_index = 13
N_MC = 20000

# colors for plot
colors = {Partial_Truncation: 'orange', Full_Truncation: 'blue', Partial_Reflection: 'black', Reflection: 'red', Implicit_1: 'grey', Implicit_2: 'purple', Modified_Milstein: 'green', Adjusted_Modified_Milstein: 'green', Truncated_Milstein: 'palevioletred'}

# labels for legend
label_list = {Partial_Truncation: 'Partial Truncation', Full_Truncation: 'Full Truncation', Partial_Reflection: 'Partial Reflection', Reflection: 'Reflection', Implicit_1: 'Implicit 1', Implicit_2: 'Implicit 2', Modified_Milstein: 'Modified Milstein', Adjusted_Modified_Milstein: '(Truncated) Modified Milstein', Truncated_Milstein: 'Truncated Milstein'}

fig, ax = plt.subplots(nrows=2, ncols=2, figsize=(15, 15))
axes = [ax[0, 0], ax[0, 1], ax[1, 0], ax[1, 1]]

for i in range(len(FR)):
    sigma = np.sqrt(2 * a / FR[i]) # change sigma
    if FR[i] < 0.5:
        scheme_list = [Full_Truncation, Adjusted_Modified_Milstein, Truncated_Milstein]

    else:
        scheme_list = [Full_Truncation, Modified_Milstein, Truncated_Milstein]

    dt = [2 ** (-i) for i in range(1, max_index + 1) if 2 ** (-i) <= 1 / 4]

    for scheme in scheme_list:
        zero_probs = []
        neg_probs = []
        runtimes = []
        for j in range(len(dt)):
            # get Wiener increments
            dW = Wiener_inc(dt[j], N_MC=N_MC, T=T)

            # simulate and time
            X1, X2, t = scheme(dW, Antithetic=True, x=x, N_MC=N_MC, T=T, a=a, sigma=sigma, k=k)

            # get probability of attaining 0
            prob_of_zero1, _ = prob_of_nonpositive(X1)
            prob_of_zero2, _ = prob_of_nonpositive(X2)

            zero_probs.append(prob_of_zero1 / 2 + prob_of_zero2 / 2)
            runtimes.append(t)

        axes[i].plot(np.log(runtimes), zero_probs, 'o', label=label_list[scheme], color=colors[scheme])

    axes[i].set_xlabel('Run time in Seconds (Log Scale)')
    axes[i].legend()
    axes[i].grid()
    axes[i].set_title(f'Feller Ratio = {FR[i]}')
    axes[i].set_ylabel('Probability of Attaining 0')

```

```
228     plt.savefig('Probability of Attaining 0')
```

D.2 Code for Section 4.2.1

```
1   # function for calculating bond price error for givens scheme
2   def Bond_error(true_price, scheme, dt, T=1, N_MC=2000, x=0.1, a=0.1, sigma=0.2, k=0.5):
3       '''
4       Calculates bond price error for a given scheme.
5       :param true_price: (float) true bond price to which that calculated through
            simulation is compared.
6       :param scheme: (function.__name__) numerical scheme to implement.
7       :param dt: (float) size of timestep.
8       :param T: (float) terminal time
9       :param N_MC: (int) number of Monte Carlo samples
10      :param x: (float) initial value for CIR process
11      :param a: (float) parameter for CIR process
12      :param sigma: (float) parameter for CIR process
13      :param k: (float) parameter for CIR process
14      :return:
15      (float) difference between true price and estmiated price
16      (float) runtime for given scheme
17      (float) Monte Carlo error
18      '''
19
20      bond_prices = []
21      err = []
22      run_time = []
23
24      for i in range(len(dt)):
25          dW = Wiener_inc(dt[i], N_MC=N_MC, T=T)
26
27          # simulate and time
28          X1, X2, t = scheme(dW, Antithetic=True, x=x, N_MC=N_MC, T=T, a=a, sigma=sigma, k=k)
29
30          # approximate integrated process
31          integral1 = np.zeros(N_MC)
32          integral2 = np.zeros(N_MC)
33
34          for n in range(1, X1.shape[1]):
35              # use trapazoidal rule
36              integral1 += (dt[i] / 2) * (X1[:, n] + X1[:, n - 1])
37              integral2 += (dt[i] / 2) * (X2[:, n] + X2[:, n - 1])
38
39          # estimate bond price with antithetic variates
40          bond_prices.append(np.mean(np.exp(-integral1) / 2 + np.exp(-integral2) / 2))
41          run_time.append(t)
42          err.append(np.abs(bond_prices[-1] - true_price))
43
44      # calculate Monte Carlo error
45      MC_error = np.sqrt(np.var(np.exp(-integral1) / 2 + np.exp(-integral2) / 2) / N_MC)
46      return err, run_time, MC_error
47
48
49  # plot results for Figure 6
50  # set parameters
51  a = 0.02
52  k = 0.4
53  x = 0.03
54  T = 1
```

```
55    N_MC = 500000
56    Feller_ratios = [0.25, 0.45, 0.75, 1.15]
57
58    my_list = [Partial_Truncation, Full_Truncation, Partial_Reflection, Reflection,
          Implicit_1, Implicit_2, Modified_Milstein, Truncated_Milstein]
59
60    # colors for plot
61    colors = {Partial_Truncation: 'orange', Full_Truncation: 'blue', Partial_Reflection:
          'black', Reflection: 'red', Implicit_1: 'grey', Implicit_2: 'purple',
          Modified_Milstein: 'green', Adjusted_Modified_Milstein: 'green', Truncated_Milstein:
          'palevioletred'}
62
63    # labels for legend
64    label_list = {Partial_Truncation: 'Partial Truncation', Full_Truncation: 'Full
          Truncation', Partial_Reflection: 'Partial Reflection', Reflection: 'Reflection',
          Implicit_1: 'Implicit 1',
65       Implicit_2: 'Implicit 2', Modified_Milstein: 'Modified Milstein',
66       Adjusted_Modified_Milstein: '(Truncated) Modified Milstein', Truncated_Milstein:
             'Truncated Milstein'}
67
68    fig, ax = plt.subplots(nrows=2, ncols=2, figsize=(15, 15))
69
70    for i in range(len(Feller_ratios)):
71       # timesteps which show more clear convergence
72       if i < 2:
73          dt = [1/2, 1/4, 1/8, 1/32, 1/64]
74
75       else:
76          dt = [1/8, 1/16, 1/32, 1/64, 1/128]
77
78       sigma = np.sqrt(2 * a / Feller_ratios[i]) # change sigma
79       true_price = True_Bond_price(T=T, sigma=sigma, a=a, k=k, x=x) # get true price
80
81       if Feller_ratios[i] <= 0.5:
82          scheme_list = [Partial_Truncation, Full_Truncation, Partial_Reflection, Reflection,
                Adjusted_Modified_Milstein,
83          Truncated_Milstein]
84
85       elif 0.5 < Feller_ratios[i] <= 1:
86          scheme_list = [Partial_Truncation, Full_Truncation, Partial_Reflection, Reflection,
                Implicit_2, Modified_Milstein,
87          Truncated_Milstein]
88
89       else:
90          scheme_list = my_list
91
92       MC_errors = []
93       for j in range(len(scheme_list)):
94          # call function for calculating bond price error
95          err, runtime, MC_err = Bond_error(true_price, scheme_list[j], dt, T=T, N_MC=N_MC,
                x=x, a=a, sigma=sigma, k=k)
96          _, line = line_of_best_fit(np.log(runtime), np.log(err), 1)
97
98          # plot results
99          if i <= 1:
100            ax[0, i].plot(np.log(runtime), line(np.log(runtime)), '-.',
                   color=colors[scheme_list[j]], label=label_list[scheme_list[j]])
101            ax[0, i].plot(np.log(runtime), np.log(err), 'o', color=colors[scheme_list[j]],
102            ms=4)
103         else:
```

```
            ax[1, i-2].plot(np.log(runtime), line(np.log(runtime)), '-.',
                color=colors[scheme_list[j]],
            label=label_list[scheme_list[j]])
            ax[1, i-2].plot(np.log(runtime), np.log(err), 'o', color=colors[scheme_list[j]],
            ms=4)

        MC_errors.append(MC_err)

    # get average of Monte Carlo error
    average_MC = np.log(np.mean(MC_errors))

    if i <= 1:
        ax[0, i].set_xlabel('Runtime in Seconds (Log Scale)')
        ax[0, i].set_ylabel('Bond Price Error (Log Scale)')
        ax[0, i].set_title(f'Feller Ratio = {Feller_ratios[i]}, Average Log MC Error =
            {np.round(average_MC, 2)}')
        ax[0, i].legend()
        ax[0, i].grid()
    else:
        ax[1, i-2].set_xlabel('Runtime in Seconds (Log Scale)')
        ax[1, i-2].set_ylabel('Bond Price Error (Log Scale)')
        ax[1, i-2].set_title(f'Feller Ratio = {Feller_ratios[i]}, Average Log MC Error =
            {np.round(average_MC, 2)}')
        ax[1, i-2].legend()
        ax[1, i-2].grid()

plt.savefig('Bond Pricing Comparison')
```

D.3 Code for Section 4.2.2

```
# plot results for Figure 7
# set parameters
a = 0.08
k = 0.4
x = 0.17
ro = -0.9
S0 = 1
r = 0
K = 1.1
T = 1
N_MC = 450000
Feller_ratios = [0.25, 0.45, 0.75, 1.15]

my_list = [Full_Truncation, Reflection, Implicit_1, Implicit_2, Modified_Milstein,
    Truncated_Milstein]

# colors for plot
colors = {Full_Truncation: 'blue', Reflection: 'red', Implicit_1: 'grey', Implicit_2:
    'purple', Modified_Milstein: 'green', Adjusted_Modified_Milstein: 'green',
    Truncated_Milstein: 'palevioletred'}

# labels for legend
label_list = {Full_Truncation: 'Full Truncation', Reflection: 'Reflection', Implicit_1:
    'Implicit 1', Implicit_2: 'Implicit 2', Modified_Milstein: 'Modified Milstein',
    Adjusted_Modified_Milstein: '(Truncated) Modified Milstein', Truncated_Milstein:
    'Truncated Milstein'}

fig, ax = plt.subplots(nrows=2, ncols=2, figsize=(15, 15))

```

```
    for i in range(len(Feller_ratios)):
        # timesteps which show clear convergence
        if i == 0:
            dt = [1/10, 1/16, 1/32, 1/64, 1/128, 1/200, 1/350]
        else:
            dt = [1/2, 1/3, 1/4, 1/6, 1/8, 1/16]

        sigma = np.sqrt(2 * a / Feller_ratios[i]) # change sigma
        true_price = Heston_call_option(ro=ro, r=r, S0=S0, K=K, v0=x, T=T, a=a, sigma=sigma,
            k=k) # get true price

        if Feller_ratios[i] < 0.5:
            scheme_list = [Full_Truncation, Reflection, Adjusted_Modified_Milstein,
                Truncated_Milstein]

        elif 0.5 <= Feller_ratios[i] <= 1:
            scheme_list = [Full_Truncation, Reflection, Implicit_2, Modified_Milstein,
            Truncated_Milstein]

        else:
            scheme_list = my_list

        MC_errors = []
        for j in range(len(scheme_list)):
            # call function for calculating option price error
            err, runtime, line, MC_err = Heston_analysis(true_price, scheme_list[j], dt, K=K,
                ro=ro, r=r, S0=S0, T=T, N_MC=N_MC, x=x, a=a, sigma=sigma, k=k)

            # plot results
            if i <= 1:
                ax[0, i].plot(np.log(runtime), line(np.log(runtime)), '-.',
                    color=colors[scheme_list[j]], label=label_list[scheme_list[j]])
                ax[0, i].plot(np.log(runtime), np.log(err), 'o', ms=4,
                    color=colors[scheme_list[j]])

            else:
                ax[1, i-2].plot(np.log(runtime), line(np.log(runtime)), '-.',
                    color=colors[scheme_list[j]],
                label=label_list[scheme_list[j]])
                ax[1, i-2].plot(np.log(runtime), np.log(err), 'o', ms=4,
                    color=colors[scheme_list[j]])

            MC_errors.append(MC_err)

        # get average of Monte Carlo errors
        average_MC = np.log(np.mean(MC_errors))

        if i <= 1:
            ax[0, i].set_xlabel('Runtime in Seconds (Log Scale)')
            ax[0, i].set_ylabel('Option Price Error (Log Scale)')
            ax[0, i].set_title(f'Feller Ratio = {Feller_ratios[i]}, Average Log MC Error =
                {np.round(average_MC, 2)}')
            ax[0, i].legend()
            ax[0, i].grid()

        else:
            ax[1, i - 2].set_xlabel('Runtime in Seconds (Log Scale)')
            ax[1, i - 2].set_ylabel('Option Price Error (Log Scale)')
            ax[1, i - 2].set_title(f'Feller Ratio = {Feller_ratios[i]}, Average Log MC Error =
                {np.round(average_MC, 2)}')
            ax[1, i - 2].legend()
```

```
76          ax[1, i - 2].grid()
77
78      plt.savefig('Heston Analysis')
```