Numerical Methods of Stochastic Differential Equations in Finance

S. Llamazares Elías

July 12, 2021

MÁSTER UNIVERSITARIO EN MODELIZACIÓN MATEMÁTICA
Facultad de Ciencias
Universidad de Salamanca
e-mail: mumoma@usal.es
https://www.usal.es/modelizacion-matematica

VISTO BUENO DEL TUTOR DEL TRABAJO FIN DE GRADO/TRABAJO FIN DE MÁSTER

TÍTULO DEL TFG/TFM

Numerical Methods of Stochastic Differential Equations in Finance

1.- Alumno:

D/Dª:	Samir Llamazares Elías

TITULACIÓN:

Master en Modelización Matemática

2.- Tutor / Profesor responsable del TFG/TFM:

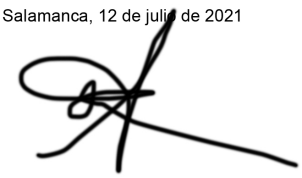
Nombre y Apellidos:	Ángel Tocino García
Departamento:	Matemáticas

Como Tutor/Profesor responsable considero que el TFG/TFM anteriormente mencionado

☒ **CUENTA con mi Visto Bueno** ☐ **NO CUENTA con mi Visto Bueno**

para ser presentado ante LA Comisión Evaluadora.

Salamanca, 12 de julio de 2021



Firma del tutor/profesor responsable

SR/A. PRESIDENTE DE LA COMISIÓN EVALUADORA DE LOS TRABAJOS FIN DE MÁSTER

Contents

1 Introduction

The study of stochastic differential equations (SDEs) originated with a groundbreaking paper published by Kiyosi Itô in the 1940's. In this paper, Itô defined a kind of stochastic integral, called the Itô integral, which allowed the integration of a certain class of functions with respect to a stochastic process [12]. The Itô integral was later used as the basis of Itô calculus, which allowed for the definition and study of SDEs [13]. Since then SDEs have been used to model a wide array of phenomena of stochastic nature in all kinds of scientific fields. In Physics, stochastic models are in many cases a more appropriate way to describe systems that cannot be isolated from their environment. SDEs allow the generalization of many deterministic models to their random counterpart such as noisy radioactive decay, the noisy pendulum [5] and the stochastic oscillator [23]. Models based on SDEs are also used in Biology for population growth where there are many factors operating (immigration, diseases, war...) and a deterministic model may therefore fail to provide accurate results [8]. The field in which SDEs have seen perhaps the biggest success is mathematical finance, where models based on SDEs have been used to give accurate pricing of a wide range of financial assets [2] and to model interest rates [6].

An SDE is a generalization of a deterministic differential equation. By incorporating a noise term and random coefficients, the differential equation becomes an SDE whose solution is no longer deterministic but rather a stochastic process. For example, let's consider the simplest interest rate model

$$\frac{dX(t)}{dt} = rX(t) \tag{1}$$

where $X(t)$ is the currency at time t and r is the interest. The model given by (1) has solution

$$X(t) = X(0)e^r t.$$

However, in practice, the interest rate$r$ might be slightly more complex, say a random variable of the form

$$r(t) = \mu(t) + \text{"noise"}$$

so that (1) turns into the SDE

$$\frac{dX(t)}{dt} = (\mu(t) + \text{"noise"})X(t)$$

of the sort we will analyze more in depth in the following sections. The solutions of these SDEs are no longer deterministic functions, but rather stochastic processes. For example, in Figure 1 we see ten trajectories of a geometric Brownian motion, i.e., a solution of the equation

$$dX_t = \mu X_t\, dt + \sigma X_t\, dW_t, \tag{2}$$

that in Finance is named as Black-Scholes equation. In this paper we will

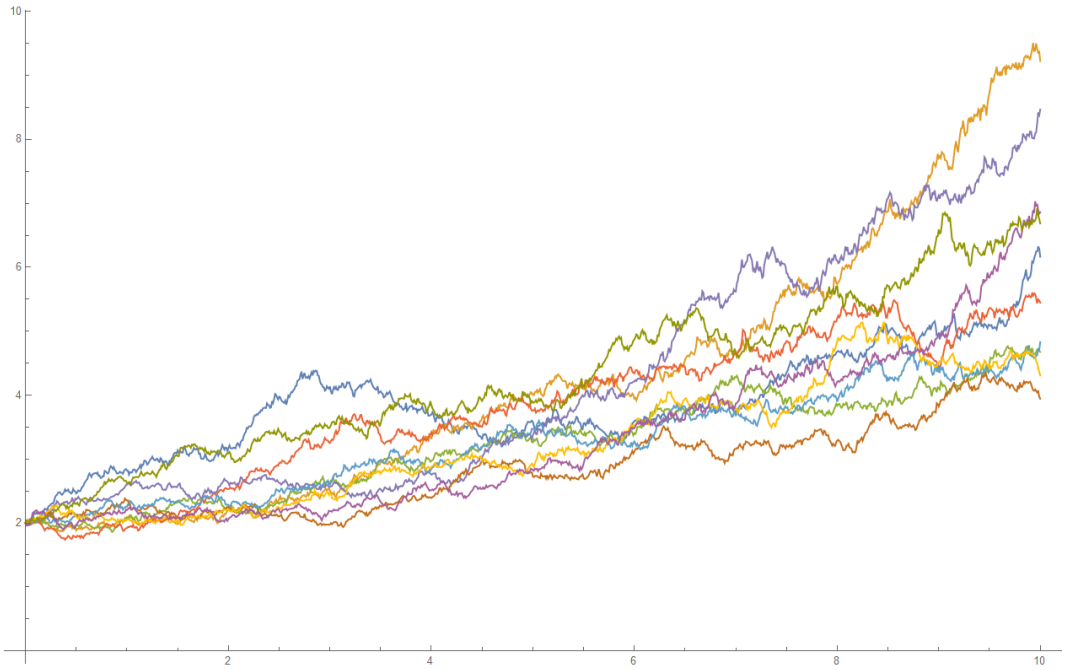


Figure 1: Ten trajectories of the solution to the Black-Scholes equation (2).

develop the fundamental theory of SDEs and use this theory to study the numerical solution of an important model in finance, the CIR model. In the first chapter we construct and study some of the basic properties of one of the most important stochastic processes, Brownian motion, which has been studied by some of the most influential scientific minds and has a fundamental role in Itô calculus. The second chapter is devoted to analyzing the central concept of Itô calculus, the Itô integral, and to prove its most important result, the Itô formula, which is the stochastic analog of the deterministic chain rule. In the third chapter we examine one of the only classes of SDEs with a closed form solution, linear SDEs, and study when and how a more general SDE can be reduced to a linear one. In the fifth chapter we go through several of the

most commonly used numerical methods used to approximate the solutions of SDEs. The sixth chapter is used to introduce some of the basic concepts found in mathematical finance as well as to survey some of the main models that have been developed in this field. Finally, in the seventh chapter we perform a comparison of some of the most used ways to simulate the CIR model as well as analyzing another way to approximate the CIR process.

2 Stochastic Processes and Brownian Motion

Throughout this paper we work fundamentally with stochastic processes, a mathematical object which is used to model systems which evolve in a random fashion.

Definition 2.1 (Stochastic Process)**.** *A stochastic process X on a probability space $(\Omega, \mathcal{F}, \mathcal{P})$ is a function*

$$X : T \times \Omega \to S$$

such that $X(t)$ is a measurable function for all $t \in T$. T is the index set, often thought of as time, in this paper it will be a real interval of the form $T = [t_0, t_f]$. S is called the state space, in this paper we will consider that $S = \mathbb{R}^n$.

We say that a stochastic process X is sample path continuous if the function

$$\begin{aligned} X(w) : T &\to S \\ t &\mapsto X(t, w) \end{aligned}$$

is continuous for all $w \in W$. We will also denote the function $X(w)$ as X_w and use the analogous notation for

$$\begin{aligned} X(t) : \Omega &\to S \\ w &\mapsto X(t, w). \end{aligned}$$

Definition 2.2 (Filtration)**.** *A filtration on a measurable space $(\Omega, \mathcal{F})$ is a sequence of sub-σ-algebras $\{\mathcal{M}_t\}_{t \in T}$ of $\mathcal{F}$ such that*

$$\mathcal{M}_{t_1} \subseteq \mathcal{M}_{t_2} \iff t_1 \leqslant t_2.$$

That is, a filtration on a measurable space is an non-decreasing sequence of sub-σ-algebras.

Definition 2.3 (Adapted Stochastic Process)**.** *A stochastic process* $\{X_t\}_{t\in T}$*, defined on a probability space* $(\Omega, \mathcal{F}, \mathcal{P})$*, is said to be adapted with respect to the filtration* $\{\mathcal{M}_t\}_{t\in T}$ *if* X_t *is* $\mathcal{M}_t$ *-measurable for every* $t \in T$*. This is commonly summarized by saying that* $\{X_t, \mathcal{M}_t\}_{t\in T}$ *is an adapted process.*

The intuition behind an adapted process is that X_t depends only upon the information contained in the sub-σ-algebra $\mathcal{M}_t$.

Definition 2.4 (Martingale)**.** *A stochastic process* $\{X_t\}_{t\in T}$*, defined on a probability space* $(\Omega, \mathcal{F}, \mathcal{P})$ *is a martingale with respect to a filtration if* $\{X_t, \mathcal{M}_t\}_{t\in T}$ *is an adapted process,*

$$\mathbb{E}[|X_t|] < \infty \quad \forall t \in T$$

and

$$\mathbb{E}\left(X_t \mid \mathcal{M}_s\right) = X_s \quad \forall s \leq t \in T.$$

This is commonly summarized by saying that $\{X_t, \mathcal{M}_t\}_{t\in T}$ *is a martingale.*

Intuitively, the fact that $\mathrm{E}\left(X_t \mid \mathcal{M}_s\right) = X_s$ for $s \leq t$ means that the best bet for the future is the present.

Definition 2.5 (Progressively Measurable)**.** *An adapted process* $\{X_t, \mathcal{M}_t\}_{t\in T}$ *defined on a probability space* $(\Omega, \mathcal{F}, \mathcal{P})$ *is said to be progressively measurable if* X *is jointly* $\mathcal{B}([t_0, t]) \times \mathcal{M}_t$ *measurable for every* $t \in T$ *where* $\mathcal{B}([t_0, t])$ *denotes the Borel* σ*-algebra on the interval* $[t_0, t]$*.*

2.1 Brownian Motion

In 1825 the botanist Robert Brown was the first to describe the chaotic displacement of pollen particles in water [4]. This movement is due to the collision of the pollen grains with water particles and was later studied by Albert Einstein [7].

From the mathematical perspective, Brownian motion is described by a stochastic Wiener process, which owes its name to Norbert Wiener [24].

The Wiener process $\{B_t\}_{t\geq 0}$ starting at x can be constructed as follows. We first look for a family of probability measures that agree with experimental results. This family is defined as follows:

$$\begin{aligned}&\nu_{t_1,\ldots,t_k}(F_1, \ldots, F_k)\\ &= \int_{F_1\times\ldots\times F_k} p\left(t_1, x, x_1\right) p\left(t_2 - t_1, x_1, x_2\right) \cdots p\left(t_k - t_{k-1}, x_{k-1}, x_k\right) dx_1 \cdots dx_k\end{aligned}$$

for all, $k \in \mathbb{N}$, $0 \leqslant t_1 \leqslant ... \leqslant t_k$, $F_1, ..., F_k \subseteq \mathbb{R}^n$ measurable sets,

$$p(t,x,y) = \begin{cases} (2\pi t)^{-\frac{n}{2}} e^{-\frac{|x-y|^2}{2t}} & \text{si } t > 0 \\ \delta_x(y) & \text{if } t = 0 \end{cases} .$$

where δ_x is the Dirac delta at x. Intuitively, this means that if a pollen particle starts at point x, then the probability that it reaches some point in F_1 after t_1 seconds, then reaches the set F_2 after another $t_2 - t_1$ seconds and so on is $\nu_{t_1,...,t_k}(F_1, ..., F_k)$.

Now, since the family of probability measures that we have just defined verify the two hypotheses of the Kolmogorov extension theorem (see Appendix A Lemma 10.1), there is a probability space $(\Omega, \mathcal{F}, \mathcal{P}^x)$ and a stochastic process $\{B_t\}_{t\geq 0}$ defined on this probability space such that

$$\mathcal{P}^x(B_{t_1} \in F_1, ..., B_{t_k} \in F_k) = \nu_{t_1,...,t_k}(F_1, ..., F_k) \tag{3}$$

for all $k \in \mathbb{N}$, $0 \leqslant t_1 \leqslant ... \leqslant t_k$ and $F_1, ..., F_k \subseteq \mathbb{R}^n$ measurable sets .

Theorem 2.6 (Properties of the Wiener Process)**.** *The Wiener process $\{B_t\}$ starting at x fulfills the following properties:*

1. $B_0 = x$, *i.e the process starts at point x.*
2. $\{B_t\}_{t\in\mathbb{R}^+}$ *is a Gaussian process.*
3. $\{B_t\}_{t\in\mathbb{R}^+}$ *has independent increments, or equivalently,*

   $$B_{t_1} - B_{t_0}, B_{t_2} - B_{t_1}, B..., Bt_k - B_{t_{k-1}}$$

   are independent for all $k > 1$ and $0 \leqslant t_0 < t_1 < ... < t_k$.
4. $B_t - B_s \sim \mathcal{N}((0, ..., 0), min(t-s)I_n)$.
5. *There is a stochastic process $\{X_t\}_{t\in\mathbb{R}^+}$ equivalent to the Wiener process $\{B_t\}_{t\in\mathbb{R}^+}$ and almost certainly continuous in time, that is, such that the function $t \mapsto X_t$ is continuous for almost every $w \in \Omega$.*

Proof. 1. This follows from the definition of the transition density.

$$\mathcal{P}^x(B_0 \in F) = \nu_0(F) = \int_F p(0,x,y)dy = \int_F \delta_x(y)dy = x$$

2. We wish to show that

$$B_{t_1,t_2,...,t_k} := (B_{t_1}, B_{t_2}, ..., B_{t_k})$$

is a Gaussian random vector for all k$\in \mathbb{R}^+$ and $0 \leqslant t_1 \leqslant ... \leqslant t_k$. Given that a random variable's distribution is uniquely determined by its characteristic function, it is enough to show that there exist $\mu \in \mathbb{R}^{nk}$ and $\Sigma \in \mathbb{R}^{nk\times nk}$ such that

$$\varphi_{B_{t_1,...,t_k}(A)} = E^x\left[\exp\left(i\sum_{j=1}^{\mathrm{n}k} A_j B_{t_j}\right)\right] = \exp\left(-\frac{1}{2}\sum_{j,m=1}^{nk} A_j\Sigma_{jm}A_m + i\sum_{j=1}^{nk} A_j\mu_j\right) \tag{4}$$

for all $A \in \mathbb{R}^{nk}$. Expanding the expected value using (3) it can be shown that (4) holds for

$$\mu = E^x[Z] = (x, x, \cdots, x) \in \mathbb{R}^{nk} \tag{5}$$

and

$$\Sigma = \begin{pmatrix} t_1I_n & t_1I_n & \cdots & t_1I_n \\ t_1I_n & t_2I_n & \cdots & t_2I_n \\ \vdots & \vdots & & \vdots \\ t_1I_n & t_2I_n & \cdots & t_kI_n \end{pmatrix}. \tag{6}$$

A representation of this covariance matrix for a one-dimensional Brownian motion can be seen in Figure 2.

3. As we have just seen, $B_{t_1,t_2,...,t_k}$ is a random Gaussian vector and hence so its linear combinations are. Consequently

$$(B_{t_1} - B_{t_0}, B_{t_2} - B_{t_1}, ..., B_{t_k} - B_{t_{k-1}})$$

is a Gaussian vector. Therefore, to see that these increments are independent, it suffices to prove that they are uncorrelated. We have that

$$\begin{aligned} &E^x\left[\left(B_{t_i} - B_{t_{i-1}}\right)\left(B_{t_j} - B_{t_{j-1}}\right)\right] \\ &= E^x\left[B_{t_i}B_{t_j} - B_{t_{i-1}}B_{t_j} - B_{t_i}B_{t_{j-1}} + B_{t_{i-1}}B_{t_{j-1}}\right] \\ &= n(t_i - t_{i-1} - t_i + t_{i-1}) = 0 \end{aligned}$$

where in the second equality we have used (6). From this we deduce that $B_s - B_t$ is independent of $\mathcal{F}_t$ for all $s > t$ where

$$\mathcal{F}_t := \sigma\{B_k; k < t\}$$

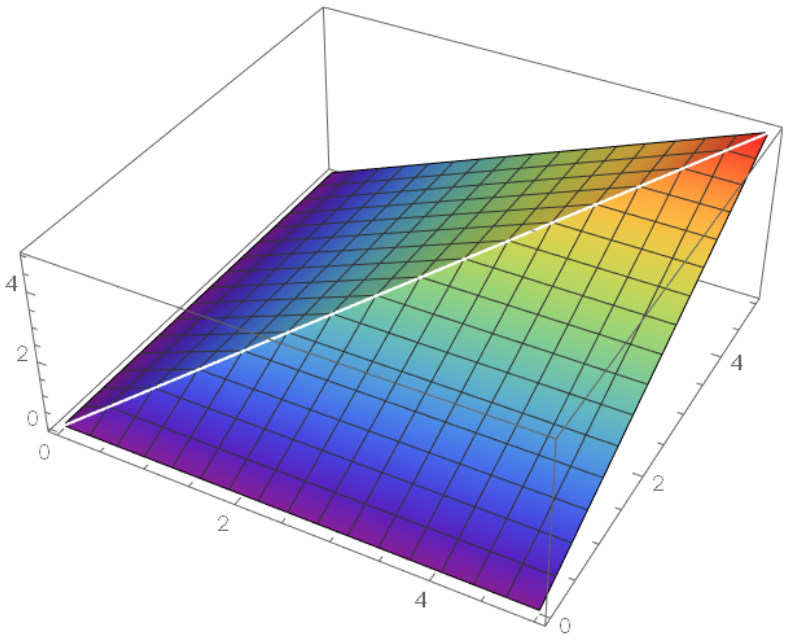


Figure 2: Covariance function of a one-dimensional Wiener process.

is the sigma-algebra generated by the Wiener process before t. Intuitively, this corresponds to the fact that $B_s - B_t$ is independent of what happened in moments before t.

4. We have already seen that $\{B_t\}_{t\in\mathcal{R}^+}$ has Gaussian increments. In addition by (3) these increments mean is

$$\mu = (0, 0, ..., 0)$$

and their covariance matrix is

$$\Sigma = \min(t - s)I_n.$$

5. We have shown that $B_t - B_s \sim \mathcal{N}((0, ..., 0), \min(t - s)I_n)$ therefore, its fourth order is

$$E^x[(B_t - B_s)^4] = n(n + 2)(t - s)^2.$$

Hence, by Kolmogorov's continuity theorem (see Appendix A, Lemma 10.2) applied to the complete metric space $L^1(\mathbb{R}^n)$ and the stochastic

process $B_t : [0, \infty] \times \mathbb{R}^n \to \mathbb{R}^n$ we conclude that the Wiener process has a continuous version.

□

In Figure 3 five trajectories of a three-dimensional Wiener process are shown.

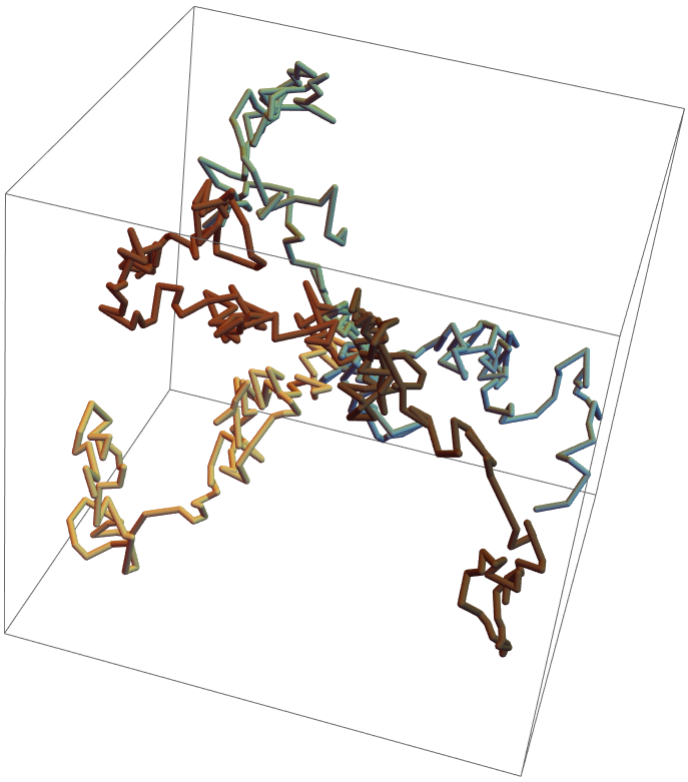


Figure 3: Five simulations of a 3D Brownian motion.

3 Stochastic Calculus

3.1 The Itô Integral

In the introduction we mentioned that SDEs were generalizations of deterministic differential equations, for example the stochastic interest rate growth model

$$\frac{dX(t)}{dt} = (\mu_t + \text{"noise"})X(t).$$

Now let's try to make sense of an equation of the form

$$\mathrm{d}X_t = \mu\left(t, X_t, \right)\mathrm{d}t + \sigma\left(t, X_t, \right)W_t$$

that we mentioned in the introduction. This equation is more commonly written as

$$\mathrm{d}X_t = \mu\left(t, X_t, \right)\mathrm{d}t + \sigma\left(t, X_t, \right)\mathrm{d}B_t \tag{7}$$

using the fact that the white noise process W_t is the distributional derivative of the Wiener process B_t. That is,

$$(W, f) = \int_{\mathbb{R}} W_t f_t dt = -\int_{\mathbb{R}} B_t f_t' = (dB, f)$$

for every test function $f \in \mathcal{D}(\mathbb{R}, \mathbb{R}^n)$. The question that arises naturally is how we should interpret (7)? If we choose to view (7) in a discrete sense, by partitioning our time interval then we would have

$$X_{t_k} = X_{t_0} + \sum_{j=0}^{k-1} \mu\left(t_j, X_{t_j}\right)\Delta t_j + \sum_{j=0}^{k-1} \sigma\left(t_j, X_{t_j}\right)\Delta B_{t_j}$$

where $\Delta B_{t_j} = B_{t_{j+1}} - B_{t_j}$. Now, if we could take the limit when $\Delta t_j \downarrow 0$ then intuitively we would arrive at an expression of the form

$$X_t = X_0 + \int_T \mu\left(s, X_s\right) ds + \int_T \sigma\left(s, X_s\right) dB_s; \tag{8}$$

which is why if $\{X_t\}_{t\in T}$ is a solution of the integral equation (8) it is said that $\{X_t\}_{t\in T}$ is a solution of the SDE (7) . However, while the first integral in (8) is a commonplace integral, the second one

$$\int_T \sigma\left(s, X_s\right) dB_s \tag{9}$$

is not. This second integral is a stochastic integral which we must now define and analyze in order to continue our study of SDEs. At first glance, one could believe that (9) is a Riemann-Stieljes integral. However, the fact that Brownian motion is not of bounded variation means that it cannot be. In the following pages we will go through the construction of the Itô integral, a particular type of stochastic integral.

First of all, to define the set of functions for which this integral makes sense, we give the following definitions.

Definition 3.1 (Square Integrable Function)**.** *Given a probability space* $(\Omega, \mathcal{F}, \mathcal{P})$*, a stochastic process* $f : T \times \Omega \to \mathbb{R}^n$ *is said to be square integrable if*

$$||f||_{2,T} := \sqrt{\int_T \mathbb{E}\left[|f_t|^2\right] dt} < \infty.$$

We denote this by $f \in \mathcal{L}^2(T \times \Omega; \mathbb{R}^n)$*. Likewise, we say that* f *is integrable, and denote this by* $f \in \mathcal{L}^1(T \times \Omega; \mathbb{R}^n)$*, if* f *verifies*

$$||f||_{1,T} := \int_T \mathbb{E}\left[|f_t|\right] dt < \infty.$$

These last two definitions are crucial for the Itô integral. This is because, as we shall see later on, for a function to be Itô integrable, it must be both square integrable and progressively measurable with respect to a certain filtration.

To start off, we will consider the one-dimensional case. Given a probability space $(\Omega, \mathcal{F}, \mathcal{P})$ and a one-dimensional Wiener process $\{B_t\}_{t\in T}$ we define the set $\mathcal{I}(T \times \Omega; \mathbb{R})$ of real valued integrable processes with respect to $\{B_t\}_{t\in T}$ as the set of functions

$$f : T \times \Omega \to \mathbb{R}$$

verifying that:

1. There exists a filtration $\{\mathcal{M}_t\}_{t\in T}$ such that $\{f_t, \mathcal{M}_t\}_{t\in T}$ is progressively measurable and $\{B_t, \mathcal{M}_t\}_{t\in T}$ is a martingale.
2. $f \in \mathcal{L}^2(T \times \Omega; \mathbb{R})$.

In the future, whenever we refer to an integrable process $\mathcal{M}_{tt\in T}$ will denote the filtration verifying the above conditions.

The Itô integral construction is similar in some sense to that of the Lebesgue integral. First of all, we define the stochastic integral for step functions. We say that a function $f \in \mathcal{I}(T \times \Omega; \mathbb{R})$ is a step function if it can be written as

$$f(t, w) = \sum_{j=0}^{n-1} \phi_j(w) \mathcal{X}_{[t_j, t_{j+1})}(t)$$

where $t_0 \leqslant t_1 \leqslant \cdots \leqslant t_n = t_f$ and $\phi_j \in L^\infty\left(\mathcal{M}_{t_j}\right)$ are bounded $\mathcal{M}_{t_j}$ measurable random variables. We denote the set of step functions as $\mathcal{S}_T$. If $f \in \mathcal{S}_T$ we define its Itô integral with respect to B_t as

$$\int_T f\left(t,w\right) dB_t(w) := \sum_{j=0}^{n-1} \phi_j(w)(B_{t_{j+1}} - B_{t_j})(w).$$

Lemma 3.2. *Itô Isometry for Step Functions. If $f \in \mathcal{S}_T$ then*

$$\mathbb{E}\left[\left(\int_T f(t,w)dB_t(w)\right)^2\right] = \mathbb{E}\left[\int_T f^2(t,w)dt\right]$$

Proof. Since $f \in \mathcal{S}_T$ then f has the form

$$f = \sum_{j=0}^{n-1} \phi_j(w)\mathcal{X}_{[t_j,t_{j+1})}(t).$$

Therefore we have that

$$\begin{aligned}\mathbb{E}\left[\left(\int_T fdB\right)^2\right] &= \sum_{j=0}^{n-1} \mathbb{E}[\phi_j^2(B_{t_{j+1}} - B_{t_j})^2] + 2\sum_{i=0}^{n-2}\sum_{j=i+1}^{n-1} \mathbb{E}\left[\phi_i\phi_j(B_{t_{i+1}} - B_{t_i})(B_{t_{j+1}} - B_{t_j})\right] \\ &= \sum_{j=0}^{n-1} \mathbb{E}[\phi_j^2(B_{t_{j+1}} - B_{t_j})^2] = \sum_{j=0}^{n-1} \mathbb{E}[\phi_j^2(t_{j+1} - t_j)] = \mathbb{E}\left[\int_T f^2 dt\right],\end{aligned}$$

where we have used that Brownian motion has independent increments of mean zero and variance the time elapsed. □

Our goal is to approximate any integrable function f by a sequence of step functions $\{f_n\}_{n\in\mathbb{N}}$ and then define the integral of f as the limit of the integrals of the step functions, which we have just defined. To do this we must prove that the set of step functions $\mathcal{S}_T$ is dense in the set of integrable functions $\mathcal{I}(T\times\Omega;\mathbb{R})$ with the norm $||\cdot||_{2,T}$, that such a limit exists and that it does not depend on the choice of sequence $\{f_n\}_{n\in\mathbb{N}}$. The proof of the density of the set of step functions in the set of integrable functions is divided into four parts. First we show that for any integrable function, there is a sequence of bounded integrable functions that converge to it.

Lemma 3.3. *Let $f \in \mathcal{I}(T\times\Omega;\mathbb{R})$, then there exists $\{f_n\}_{n\in\mathbb{N}} \subset \mathcal{S}_T$ such that* $\lim_{n\to\infty} ||f - f_n||_{2,T} = 0$.

Proof. We define f_n as follows:

$$f_n(t) := \min\{\max\{-n, f(t)\}, n\}$$

Clearly f_n is bounded and square integrable. It is also $\{\mathcal{M}_t\}_{t\in T}$ adapted because $f_n(t)$ is the minimum of two $\mathcal{M}_t$ measurable functions for each $t \in T$. Furthermore we have that

$$|f_n(t,\omega) - f(t,\omega)| \leqslant |f_n(t,\omega)| + |f(\omega)| \leqslant 2|f(t,\omega)|$$

hence we conclude that

$$\lim_{n\to\infty} ||f_n - f||_{2,T} = 0$$

by the dominated convergence theorem applied to the sequence

$$\left\{\mathbb{E}\left[|f_n(t) - f(t)|^2\right]\right\}_{n\in\mathbb{N}}$$

which is dominated by

$$4\mathbb{E}[|f(t)|^2].$$

□

The second step is to prove that given a bounded integrable function, there is a sequence of bounded and pathwise continuous functions that converge to said function.

Lemma 3.4. *Suppose that $f \in \mathcal{I}(T \times \Omega; \mathbb{R})$ is bounded. Then there exists a sequence of bounded, pathwise-continuous processes $\{f_n\}_{n\in\mathbb{N}} \subset \mathcal{I}(T \times \Omega; \mathbb{R})$ such that*

$$\lim_{n\to\infty} ||f - f_n||_{2,T} = 0$$

Proof. For each $n \in \mathbb{N}$, we define f_n as follows

$$f_n(t,\omega) := \int_T \varphi_n(s-t) f(s,\omega) ds.$$

where $\varphi_n : \mathbb{R} \to \mathbb{R}^+$ is a continuous function such that $\varphi_n(s) = 0$ if $s \in \left(-\frac{1}{n}, 0\right)^C$ and

$$\int_{-\infty}^{\infty} \varphi_n(s) ds = 1.$$

Then for every $\omega \in \Omega$ it can be shown that $f_n(t,\omega)$ is continuous in t, $|f_n(t,\omega)| \leqslant K$ and $||f_n||_{2,T} < \infty$. Additionally $\{f_n(t), \mathcal{M}_t\}_{t\in T}$ is an adapted process for every $n \in \mathbb{N}$ and

$$\lim_{n\to\infty} \int_T |f(t,\omega) - f_n(t,\omega)|^2 \, dt = 0$$

for all $\omega \in \Omega$. These results are technically complex and not very connected with SDEs in general, hence their proof will be omitted proof can be found in [15]). They allow us to use the dominated convergence theorem to show that

$$\lim_{n\to\infty} ||f - f_n||_{2,T} = 0.$$

□

Finally, given an integrable, bounded and pathwise continuous function, we will show that there is a sequence of step functions that converge to it.

Lemma 3.5. *Suppose that $f \in \mathcal{I}(T\times\Omega;\mathbb{R})$ is bounded and pathwise-continuous. Then there exists a sequence $\{f_n\}_{n\in\mathbb{N}} \in {\mathcal{S}_T}^{\mathbb{N}}$ such that*

$$\lim_{n\to\infty} ||f - f_n||_{2,T} = 0$$

.

Proof. We define f_n as

$$f_n(t,w) := \sum_{j=0}^{\infty} f\left(t_j, w\right) \mathcal{X}_{[t_j,t_{j+1})}(t)$$

where $t_j := j2^{-n}$. Then $\lim_{n\to\infty} f_n(t) - f(t) = 0$ for all $t \in\ T$ and since

$$|f_n(t,\omega) - f(\omega)| \leqslant |f_n(t,\omega)| + |f(t,\omega)| \leqslant 2\|f_t\|_\infty < \infty$$

we conclude that

$$\lim_{n\to\infty} ||f_n - f||_{2,T} = 0$$

by the dominated convergence theorem. □

Lemma 3.6. *The set of step functions $\mathcal{S}_T$ is dense in the metric space*

$$(\mathcal{I}(T \times \Omega;\mathbb{R}), ||\cdot||_{2,T}).$$

Proof. This is immediate by using the previous three lemmas. □

Lemma 3.7. *Let f be an integrable function and let $\{f_n\}_{n\in\mathbb{N}}$ be a sequence of step functions converging to f as in Lemma 3.4. Then*

$$\lim_{n\to\infty} \int_T f_n(t,w) dB_t(w)$$

exists and does not depend on the sequence of step functions converging to f.

Proof. By Lemma 3.2 we have that

$$\begin{aligned}
&\mathbb{E}\left[\left|\int_T f_n(t,w)dB_t(w) - \int_T f_m(t,w)dB_t(w)\right|^2\right] \\
&= \mathbb{E}\int_T |f_n(t,w) - f_m(t,w)|^2\, dt \\
&\leq 2\int_T \left(\mathbb{E}\left[|f_n(t,w) - f(t,w)|^2\right] + \mathbb{E}\left[|f_m(t,w) - f(t,w)|^2\right]\right) dt
\end{aligned}$$

where in the last inequality we have used that

$$(a+b)^2 = a^2 + b^2 + 2ab \leqslant a^2 + b^2 + a^2 + b^2 = 2(a^2+b^2).$$

This last expression converges to 0 when $n, m \to \infty$ so we have that

$$\left\{\int_T f_n(t,w)dB_t(w)\right\}_{n\in\mathbb{N}}$$

is a Cauchy sequence in the complete space $L^2(\Omega;\mathbb{R})$. Therefore $\lim_{n\to\infty}\int_T f_n(t,w)dB_t(w)$ exists.

Now suppose that $\{\bar{f}_n\}$ is another sequence of step functions converging to f as in Lemma 3.4. Then clearly the alternating sequence $\{g_n\}_{n\in\mathbb{N}}$, where $g_{2n-1} = f_n$ and $g_{2n} = \bar{f}_n$ also converges to f in this sense. Consequently the sequence

$$\left\{\int_T g_n(t,w)dB_t(w)\right\}_{n\in\mathbb{N}}$$

is also convergent, say to it's limit I, in $L^2(\Omega;\mathbb{R})$. Therefore it's subsequences with increasing sub-index, in particular $\{g_{2n-1}\}_{n\in\mathbb{N}} = \{f_n\}_{n\in\mathbb{N}}$ and $\{g_{2n}\}_{n\in\mathbb{N}} = \{\bar{f}_n\}_{n\in\mathbb{N}}$, converge to I which conclude the proof.

□

Now we have the tools necessary to define the Itô integral.

Definition 3.8 (Itô Integral)**.** *Given a function $f \in \mathcal{I}(T\times\Omega;\mathbb{R})$ we define its Itô integral as follows*

$$\int_T f(t,w)dB_t(w) := \lim_{n\to\infty}\int_T f_n(t,w)dB_t(w)$$

where $\{f_n\}_{n\in\mathbb{N}}$ is a sequence of step functions converging to f.

Now that we have defined the Itô integral, let's look at some of its properties.

Lemma 3.9 (Properties of the Itô Integral)**.** *For any $f \in \mathcal{I}(T\times\Omega;\mathbb{R})$, $a, b \in \mathbb{R}$ and $t_1 < t_2 < t_3 \in T$ the following statements hold:*

1. $\int_T f(t,w)dB_t(w)$ *is $\mathcal{M}_T$ measurable.*
2. $\int_T (af(t,w) + bg(t,w))dB_t(w) = a\int_T f(t,w)dB_t(w) + b\int_T g(t,w)dB_t(w)$ *almost surely.*
3. $\mathbb{E}\left[f(t,w)dB_t(w)\right] = 0$.
4. $\mathbb{E}\left[\left(\int_T f(t,w)dB_t(w)\right)^2\right] = \mathbb{E}\left[\int_T f^2(t)dt\right]$ *(Itô isometry).*
5. $\int_{t_1}^{t_3} f(t,w)dB_t(w) = \int_{t_1}^{t_2} f(t,w)dB_t(w) + \int_{t_2}^{t_3} f(t,w)dB_t(w)$ *for almost all $w \in \Omega$.*

Proof. It is easy to check that the properties hold for step functions. By Lemma 3.4 there exist sequences of step functions $\{f_n\}_{n\in\mathbb{N}}$ and $\{g_n\}_{n\in\mathbb{N}}$ that converge to f and g respectively; from here the properties hold for f and g. □

Another important property of the Itô integral is that it is a martingale. This is a crucial property that distinguishes the Itô integral from other stochastic integrals.

Theorem 3.10. *Suppose that $f \in \mathcal{I}(T \times \Omega;\mathbb{R})$ and $T = [t_0, t_f]$. Then, the stochastic process $I_f : T \times \Omega \to \mathbb{R}$ defined by*

$$I_f(t,w) := \int_{t_0}^{t} f(s,w)dB_s(w).$$

is a martingale.

Proof. By Lemma 3.9 I_f has expected value zero. It remains to see that

$$\mathbb{E}\left[I_f(t)|\mathcal{M}_s\right] = I_f(s)$$

for all $t_0 \leqslant s < t \leqslant t_f$. Using Lemma 3.9 we have that

$$\mathbb{E}\left[I_f(t)|\mathcal{M}_s\right] = \mathbb{E}\left[\int_{t_0}^{t} f(u)dB_u|\mathcal{M}_s\right] = \mathbb{E}\left[\int_{t_0}^{s} f(u)dB_u|\mathcal{M}_s\right] + \mathbb{E}\left[\int_{s}^{t} f(u)dB_u|\mathcal{M}_s\right].$$

Since, again by Lemma 3.9, $\int_{t_0}^{s} f(u)dB_u$ is $\mathcal{M}_s$ measurable with null expectation,

$$\mathbb{E}\left[I_f(t)|\mathcal{M}_s\right] = \int_{t_0}^{s} f(u)dB_u + \mathbb{E}\left[\int_s^t f(u)dB_u|\mathcal{M}_s\right] = \mathbb{E}\left[\int_s^t f(u)dB_u|\mathcal{M}_s\right].$$

If f were a step function, then by the definition of Itô integral for step functions and the property of independent increments of the Wiener process, we would have that $\int_s^t f(u)dB_u$ would be independent of the σ-algebra $\mathcal{M}_s$. Given any integrable function f, by Lemma 3.4 there is a sequence of step functions that converge to f; then the limit $\int_s^t f(u)dB_u$ is independent of $\mathcal{M}_s$. Therefore

$$\mathbb{E}\left[I_f(t)|\mathcal{M}_s\right] = \mathbb{E}\left[\int_s^t f(u)dB_u|\mathcal{M}_s\right] = \int_s^t f(u)dB_u = I_f(s)$$

just as we wanted. □

Finally we will see that there exists a continuous version of the Itô integral. For this reason whenever we consider an Itô integral we will suppose it to be continuous.

Theorem 3.11. *Let $f \in \mathcal{I}(T \times \Omega; \mathbb{R})$ an integrable function. Then there exists an almost surely pathwise-continuous version of the stochastic process*

$$I_f(t, w) = \int_{t_0}^{t} f(s, w)dB_s(w).$$

Proof. Our goal will be to construct a sequence of continuous functions that converges almost surely to the desired stochastic process. We construct this sequence in the most natural way possible. Let $\{\phi_n\}_{n\in\mathbb{N}}$ be a sequence of step functions that converge to f we define the sequence $\{I_n\}_{n\in\mathbb{N}}$ as

$$I_n(t, w) := \int_{t_0}^{t} \phi_n(s, w)dB_s(w).$$

As we have seen the Itô integral is a martingale with respect to the filtration generated by the Brownian motion. In addition, by the definition of the Itô integral for step functions $I_n(t, w)$ is almost surely continuous with respect to t for all $n \in \mathbb{N}$. It remains to see that this sequence converges uniformly to I. Since $I_n - I_m$ is an almost surely pathwise continuous martingale by Doob's

martingale inequality 10.4 we have that

$$P\left[\sup_{t\in T}|I_n(t,\omega)-I_m(t,\omega)|>\epsilon\right]\leqslant\frac{1}{\epsilon^2}\cdot\mathbb{E}\left[|I_n(t_f,\omega)-I_m(t_f,\omega)|^2\right]$$
$$=\frac{1}{\epsilon^2}\mathbb{E}\left[\int_T(\phi_n-\phi_m)^2\,ds\right]\to 0$$

as $n,m\to\infty$. Therefore we can choose a subsequence such that

$$P\left[\sup_{0\leqslant t\leqslant T}\left|I_{n_{k+1}}(t,\omega)-I_{n_k}(t,\omega)\right|>2^{-k}\right]<2^{-k}.$$

Applying the Borel-Cantelli lemma 10.5 to the sequence of events

$$\left\{\sup_{0\leqslant t\leqslant T}\left|I_{n_{k+1}}(t,\omega)-I_{n_k}(t,\omega)\right|>2^{-k}\right\}_{k\in\mathbb{N}}$$

it follows that only a finite number of these events can occur. Hence, for almost every $w\in\Omega$ there exists $k(w)\in\mathbb{N}$ such that

$$\sup_{0\leqslant t\leqslant T}\left|I_{n_{k+1}}(t,\omega)-I_{n_k}(t,\omega)\right|\leqslant 2^{-k};\quad k\geq k(w).$$

Consequently $I_n(t,w)$ converges to uniformly, to its continuous limit, for all $t\in T$ and for almost all $w\in\Omega$. Furthermore since

$$I_f(t)=\lim_{n\to\infty}I_n(t)$$

we conclude that $I_f(t)$ is almost surely continuous. □

In the future we will use that any Itô integral is an almost surely pathwise continuous martingale. In particular, it allows the use of Doob's martingale inequality (see Lemma 10.4 in Appendix A) as in the following proof:

Lemma 3.12 (Maximal Inequality)**.** *Let $(\Omega,\mathcal{F},\mathcal{P})$ be a probability space and $f\in\mathcal{I}(T\times\Omega;\mathbb{R})$ an integrable function. Then*

$$\mathbb{E}\left[\int_T|f(s)|^2ds\right]\leqslant\mathbb{E}\left[\sup_{t\in T}\left(\int_{t_0}^t f(s)dB_s\right)^2\right]\leqslant 4\mathbb{E}\left[\int_T|f(s)|^2ds\right].$$

Proof. From Theorem 3.10 and Theorem 3.11 the random variable

$$I_f(t,w)=\int_{t_0}^t f(s,w)dB_s(w)$$

is almost surely a continuous martingale. Therefore we can apply Doob's martingale inequality. Then

$$\mathbb{E}\left[\int_T |f(s)|^2 ds\right] = \mathbb{E}\left[\left(\int_T f(s)dB_s\right)^2\right] \leqslant \mathbb{E}\left[\sup_{t\in T}\left(\int_{t_0}^t f(s)dB_s\right)^2\right]$$
$$\leqslant 4\mathbb{E}\left[\left(\int_T f(s)dB_s\right)^2\right] = 4\mathbb{E}\left[\int_T \left|f(s)^2\right| ds\right],$$

where the Itô isometry has been used. □

The Itô integral has been defined for a certain class of $\mathbb{R}$-valued functions with respect to a one dimensional Wiener process. This definition can be generalized to the multidimensional case in the following way.

Definition 3.13. *Let $B = (B_1, ..., B_m)$ be an m-dimensional Wiener process and*

$$f = \begin{pmatrix} f_{11} & \cdots & f_{1m} \\ \vdots & & \vdots \\ f_{n1} & \cdots & f_{nm} \end{pmatrix} \in \mathcal{I}\left(T \times \Omega; \mathbb{R}^{nm}\right),$$

we define

$$\int_T f(t,w)dB(t,w) := \left(\sum_{j=1}^m \int_T f_{1j}(t,w)dB_j(t,w), ..., \sum_{j=1}^m \int_T f_{nj}(t,w)dB_j(t,w)\right)$$

It is easy to check that all the properties of Lemma 3.9 hold for the generalized definition. While, as we have seen, the Itô integral is a martingale, which is an important property for modelling processes whose expected values only depend on the current value of the process, the derivation rules resulting from this definition do not follow those of ordinary calculus. Transforming a process and then differentiating it leads to a surprising result if we consider the chain rule of ordinary calculus. This stochastic chain rule is given by the the Itô lemma, also called Itô formula, which we shall prove in the one-dimensional case.

3.2 The Itô Formula

Definition 3.14 (Itô Process)**.** *Given a probability space $(\Omega, \mathcal{F}, \mathcal{P})$, an $n-$dimensional stochastic process $\{X_t\}_{t\in T}$ is said to be a Itô process with respect to the filtration $\{\mathcal{M}_t\}_{t\in T}$ if there exists:*

1. *an $m-$dimensional Wiener process $\{B_t\}_{t\in T}$ which is a martingale with respect to $\{\mathcal{M}_t\}_{t\in T}$*

2. *An adapted process $\{\mu_t, \mathcal{M}_t\}_{t\in T}$ called drift such that $\mu \in \mathcal{L}^1(T\times\Omega;\mathbb{R}^n)$.*

3. *An adapted process $\{\sigma_t, \mathcal{M}_t\}_{t\in T}$ called diffusion such that $\sigma \in \mathcal{I}(T\times\Omega;\mathbb{R}^{nm})$*

verifying the relation

$$dX_t = \mu_t dt + \sigma_t dB_t. \tag{10}$$

In this case we will say that X is an n-dimensional Itô process driven by an m-dimensional Brownian motion.

Theorem 3.15 (One-dimensional Itô Formula)**.** *Let*

$$\begin{aligned} f :&\mathbb{R}^+\times\mathbb{R}\to\mathbb{R}\\ &(t,x)\to f(t,x)\end{aligned}$$

be continuously differentiable with respect to t and twice continuously differentiable with respect to x and suppose that $\{X_t\}$ is the one-dimensional Itô process driven by a one-dimensional Brownian motion given in (10)*. Then $f(t,X_t)$ is an Itô process verifying the identity*

$$\begin{aligned} df(t,X_t) = &\left(\frac{\partial f}{\partial t}(t,X_t) + \frac{\partial f}{\partial x}(t,X_t)\mu(t,X_t) + \frac{1}{2}\frac{\partial^2 f}{\partial x^2}(t,X_t)\sigma^2(t,X_t)\right)dt\\ &+ \frac{\partial f}{\partial x}(t,X_t)\sigma(t,X_t)dB_t.\end{aligned}$$

Proof. Part of the proof is quite technical and will therefore be abbreviated; more details can be found in [17]. The first step is to show that it suffices to prove the theorem for bounded stochastic processes by considering a sequence of bounded stochastic processes that approximate X. Secondly, it is possible to show that there exists a sequence of functions $\{f_n\}_{n\in\mathbb{N}}$ which are C^2 in both t and x, instead of just x, and whose partial derivatives f, $\frac{\partial f}{\partial t}$, $\frac{\partial f}{\partial x}$, $\frac{\partial f}{\partial x^2}$, converge to those of f uniformly on compact subsets of $T\times\mathbb{R}$. Next, as we have already done previously, we can approximate σ and μ by step functions. In summary, it suffices to prove the theorem for bounded, twice continuously differentiable functions and stochastic processes whose drift and diffusion are step functions.

After this simplification of the problem we consider the second order Taylor expansion of $f(t, X_t)$:

$$f(t, X_t) = f(0, X_0) + \sum_i \frac{\partial f}{\partial t}\Delta t_i + \sum_i \frac{\partial f}{\partial x}\Delta X_i + \frac{1}{2}\sum_i \frac{\partial^2 f}{\partial t^2}\Delta t_i^2 + \sum_i \frac{\partial^2 f}{\partial t \partial x}\Delta t_i \Delta X_i + \frac{1}{2}\sum_i \frac{\partial^2 f}{\partial x^2}\Delta X_i^2 + \sum_i R_i$$

where $\Delta t_i = t_{i+1} - t_i$, $\Delta X_i = X(t_{i+1}) - X(t_i)$, $R_i = o(\Delta t_i^2 + \Delta x_i^2)$ and the partial derivatives are evaluated at (t_i, X_{t_i}). Our goal now is to study each member of this expression in an attempt to simplify it. When the size of the partition converges to zero we have that

$$\sum_i \frac{\partial f}{\partial t}\Delta t_i = \sum_i \frac{\partial f}{\partial t}(t_i, X_i)\,\Delta t_i \to \int_{t_0}^{t} \frac{\partial f}{\partial t}(s, X_s)\,ds, \tag{11}$$

$$\sum_i \frac{\partial f}{\partial x}\Delta X_i = \sum_i \frac{\partial f}{\partial x}(t_i, X_i)\,\Delta X_i \to \int_{t_0}^{t} \frac{\partial f}{\partial x}(s, X_s)\,dX_s, \tag{12}$$

$$\sum_i R_i \to 0, \tag{13}$$

and

$$\frac{1}{2}\sum_i \frac{\partial^2 f}{\partial t^2}\Delta t_i^2 \to 0, \tag{14}$$

because $\frac{\partial^2 f}{\partial t^2}$ is bounded. Let us see what happens with the remaining two terms. On the one hand

$$\sum_i \frac{\partial^2 f}{\partial t \partial x}\Delta t_i \Delta X_i = \sum_i \frac{\partial^2 f}{\partial t \partial x}\mu_i (\Delta t_i)^2 + \sum_i \frac{\partial^2 f}{\partial t \partial x}\sigma_i \Delta t_i \Delta B_i$$

where $\mu_i := \mu(t_i, X(t_i))$ and $\sigma_i := \sigma(t_i, X(t_i))$. When the size of the partition converges to zero, the first addend does too; for the second one we have

$$\mathbb{E}\left[\left(\sum_i \frac{\partial^2 f}{\partial t \partial x}\sigma_i \Delta t_i \Delta B_i\right)^2\right] = \sum_i \mathbb{E}\left[\left(\frac{\partial^2 f}{\partial t \partial x}\sigma_i\right)^2\right](\Delta t_i)^3 \longrightarrow 0$$

due to the boundedness of $\frac{\partial^2 f}{\partial t \partial x}$; this means that

$$\sum_i \frac{\partial^2 f}{\partial t \partial x}\Delta t_i \Delta X_i \xrightarrow{\mathcal{L}_T^2} 0. \tag{15}$$

On the other hand

$$\frac{1}{2}\sum_i \frac{\partial^2 f}{\partial x^2}\Delta X_i^2 = \frac{1}{2}\sum_i \frac{\partial^2 f}{\partial x^2}\mu_i^2(\Delta t_i)^2 + \sum_i \frac{\partial^2 f}{\partial x^2}\mu_i\sigma_i\Delta t_i\Delta B_i + \frac{1}{2}\sum_i \frac{\partial^2 f}{\partial x^2}\sigma_i^2(\Delta B_i)^2.$$

The first two addends converge to zero because $\frac{\partial^2 f}{\partial x^2}$ is bounded and

$$\mathbb{E}\left[\left(\sum_i \frac{\partial^2 f}{\partial x^2}\mu_i\sigma_i\Delta t_i\Delta B_i\right)^2\right] = \sum_i \mathbb{E}\left[\left(\frac{\partial^2 f}{\partial x^2}\mu_i\sigma_i\right)^2\right](\Delta t_i)^3.$$

Since $\mathbb{E}\left[\Delta B_i^2\right] = \Delta t_i$ intuitively it would seem that the final term verifies

$$\frac{1}{2}\sum_i \frac{\partial^2 f}{\partial x^2}\sigma_i^2(\Delta B_i)^2 \to \frac{1}{2}\int_T \frac{\partial^2 f}{\partial x^2}\sigma^2 dt.$$

Let us see that this in fact true by subtracting $\frac{1}{2}\sum_i \frac{\partial^2 f}{\partial x^2}\sigma_i^2\Delta t_i$ from $\frac{1}{2}\sum_i \frac{\partial^2 f}{\partial x^2}\sigma_i^2(\Delta B_i)^2$ and showing that this subtraction converges to zero in $\mathcal{L}^2(T\times\Omega;\mathbb{R})$. To make the expression a bit simpler, we abbreviate $\frac{1}{2}\frac{\partial^2 f}{\partial x^2}\sigma_i^2$ by a_i. Proceeding as we explained we obtain

$$\begin{aligned}\mathbb{E}\left[\left(\sum_i a_i(\Delta B_i)^2 - \sum_i a_i\Delta t_i\right)^2\right] &= \mathbb{E}\left[\sum_i\sum_j a_i a_j\left((\Delta B_i)^2 - \Delta t_i\right)\left((\Delta B_j)^2 - \Delta t_j\right)\right] \\ &= \mathbb{E}\left[\sum_i a_i^2\left((\Delta B_i)^2 - \Delta t_i\right)^2\right]\end{aligned}$$

because when $i \neq j$ then $(\Delta B_i)^2 - \Delta t_i$ and $(\Delta B_j)^2 - \Delta t_j$ are independent and their expected values are 0. Now using the formula for the expected value of a normal distribution we have that

$$\begin{aligned}\mathbb{E}\left[\sum_i a_i^2((\Delta B_i)^2 - \Delta t_i)^2\right] &= \sum_i \mathbb{E}\left[a_i^2\right]\left(3\left(\Delta t_i\right)^2 - 2\left(\Delta t_i\right)^2 + \left(\Delta t_i\right)^2\right) \\ &= 2\sum_i \mathbb{E}\left[a_i^2\right](\Delta t_i)^2\end{aligned}$$

which converges to zero as we wished to show since $\frac{\partial^2 f}{\partial x^2}$ is bounded. This means that

$$\frac{1}{2}\sum_i \frac{\partial^2 f}{\partial x^2}\Delta X_i^2 \xrightarrow{\mathcal{L}_T^2} \frac{1}{2}\int_{t_0}^t \frac{\partial^2 f}{\partial x^2}\sigma^2 dt. \tag{16}$$

Using (11)-(16) we obtain that

$$\begin{aligned} f\left(t, X_t\right) &= f(0, X(0)) \\ &+ \int_{t_0}^{t} \left(\frac{\partial f}{\partial t}\left(s, X_s\right) + \frac{\partial f}{\partial x}\left(s, X_s\right) \mu\left(s, X_s\right) + \frac{\partial^2 f}{\partial x^2}(s, X_s)\sigma^2(s, X_s) \right) ds \\ &+ \int_{t_0}^{t} \frac{\partial f}{\partial x}(s, X_s)\sigma(s, X_s) dB_s(X_s). \end{aligned}$$

□

The above lemma can be generalized to higher dimensions.

Lemma 3.16 (Multidimensional Itô Lemma [12])**.** *Suppose that X is an n-dimensional Itô process with respect to the Brownian motion B and let $f = (f_1, ..., f_n) : T \times \mathbb{R}^n \to \mathbb{R}^p$ be a function whose components are C^1 with respect to t and C^2 with respect to x. Then $f(t, X_t)$ is an Itô process with*

$$df_k(t, X_t) = \frac{\partial f_k}{\partial t}(t, X_t)dt + \sum_{i=1}^{n} \frac{\partial f_k}{\partial x_i}(t, X_t)dX_i + \frac{1}{2} \sum_{1 \le i,j \le n} \frac{\partial^2 f_k}{\partial x_i \partial x_j}(t, X_t)dX_i dX_j \tag{17}$$

where $dX_i \cdot dX_j$ is computed using the rules

$$dtdt = dtdB_i = dB_i dt = 0; \quad dB_i dB_j = \delta_{ij} dt \tag{18}$$

for all $1 \le i, j \le m$.

This result can be expressed more compactly as

$$\begin{aligned} df(t, X_t) &= \left(\frac{\partial f}{\partial t}(t, X_t) + \nabla_x f \mu(t, X_t) + \frac{1}{2} Tr\left[\sigma^T(t, X_t) H_x f(t, X_t) \sigma(t, X_t)\right] \right) dt \\ &+ \nabla_x f(t, X_t)\sigma(t, X_t) dB_t \end{aligned}$$

where $\nabla_x f$ and $H_x f$ denote the gradient and the Hessian matrix of f respectively:

$$\nabla_x f = \left(\frac{\partial f}{\partial x_1}, ..., \frac{\partial f}{\partial x_n} \right) \quad H_x f = \begin{pmatrix} \frac{\partial f}{\partial x_1 x_1} & \cdots & \frac{\partial f}{\partial x_1 x_n} \\ \vdots & \ddots & \vdots \\ \frac{\partial f}{\partial x_n x_1} & \cdots & \frac{\partial f}{\partial x_n x_n} \end{pmatrix}.$$

4 Stochastic Differential Equations

In this section we define in a more rigorous manner what an SDE is and what its solutions are, concepts that were intuitively introduced above. Then we will study under what conditions the solutions of an SDE exist and are unique, as well as what properties they verify. In the field of SDEs, we also differentiate between weak and strong solutions. However, this notion is very different from the one found in ODEs and PDEs.

4.1 Definitions and Existence and Uniqueness Theorem

Let $(\Omega, \mathcal{F}, \mathcal{P})$ be a probability space, $\{B_t\}_{t\in T}$ an m-dimensional Wiener process with respect to to the filtration $\{\mathcal{M}_t\}_{t\in T}$.

Definition 4.1. *Given $\mu, \sigma_k : T \times \mathbb{R}^n \to \mathbb{R}^n$, $k = 1, \dots, m$, a stochastic differential equation (SDE) is an equation of the form*

$$dX_t = \mu(t, X_t)\, dt + \sum_{k=1}^{m} \sigma_k(t, X_t)\, dB_k(t); \quad t \in T, \tag{19}$$

also denoted in a compact form as

$$dX_t = \mu(t, X_t)\, dt + \sigma(t, X_t)\, dB_t, \tag{20}$$

where μ and σ are called the drift and diffusion coefficients respectively.

Notice that the SDE (19) is equivalent to the integral equation

$$X_t = X_{t_0} + \int_{t_0}^{t} \mu(s, X_s)\, ds + \sum_{k=1}^{m} \int_{t_0}^{t} \sigma_k(s, X_s)\, dB_k(s).$$

Definition 4.2. *Given a random variable $\xi : \Omega \to \mathbb{R}^n$, a stochastic process $Y : T \times \Omega \to \mathbb{R}^n$ is a strong solution of the SDE (20) with initial value $X_{t_0} = \xi$ if it verifies:*

1. *Y is progressively measurable with respect to $\{\mathcal{M}_t\}_{t\in T}$.*
2. *Y is sample path continuous.*
3. *$\mu(t, Y_t) \in \mathcal{L}^1(T \times \Omega; \mathbb{R}^n)$*
4. *$\sigma(t, Y_t) \in \mathcal{I}(T \times \Omega; \mathbb{R}^n)$*

5. $Y_t = \xi + \int_{t_0}^{t} \mu(s, Y_s)ds + \int_{t_0}^{t} \sigma(s, Y_s)dB_s$ *with probability* 1 *for all* $t \in T$.

Clearly if Y is the solution of an SDE then it is an Itô process. In the case of the strong solutions, we have fixed the probability space, the Wiener process, the filtration and the initial value. However, in the case of the weak solution we will let them vary.

Definition 4.3. *A weak solution of an SDE of the form* (20) *is a probability space* $(\Omega, \mathcal{F}, \mathcal{P})$*, a filtration* $\{\mathcal{M}_t\}_{t\in T}$*, a Wiener process* $\{B_t\}_{t\in T}$ *which is a martingale with respect to* $\{\mathcal{M}_t\}_{t\in T}$ *and a stochastic process* $Y : T \times \Omega \to \mathbb{R}^n$ *such that*

1. Y *is progressively measurable with respect to* $\{\mathcal{M}_t\}_{t\in T}$.
2. Y *is sample path continuous.*
3. $\mu(t, Y_t) \in \mathcal{L}^1(T \times \Omega; \mathbb{R}^n)$
4. $\sigma(t, Y_t) \in \mathcal{I}(T \times \Omega; \mathbb{R}^n)$
5. $Y_t = Y_{t_0} + \int_{t_0}^{t} \mu(s, Y_s)ds + \int_{t_0}^{t} \sigma(s, Y_s)dB_s$ *almost surely for all* $t \in T$.

Since the solution of an SDE is a stochastic process, we have to slightly alter the concept of uniqueness of solutions from ODEs. In this case we consider two strong solutions Y, Z of (20) to be equivalent if Z is a version of Y, that is, if

$$\mathcal{P}(Y_t = Z_t) = 1$$

for all $t \in T$. Therefore, a strong solution Y is unique if and only if all other solutions are modifications of Y. Unfortunately, we can't use the same notion of uniqueness for weak solutions because two weak solutions don't need to be defined on the same probability space. Instead, we say that two weak solutions X,Y are equivalent if they have the same finite-dimensional distribution. In symbols,

$$\mathcal{P}_X \{\omega \in \Omega_X \mid X_{t_i}(\omega) \in A_j \text{ for } 1 \leqslant i \leqslant k\} = \mathcal{P}_Y \{\omega \in \Omega_Y \mid Y_{t_i}(\omega) \in A_j \text{ for } 1 \leqslant i \leqslant k\}$$

for all $k \in \mathbb{N}$, $t_1, ..., t_k \in T$ and $A_1, ..., A_k \in \mathcal{B}(\mathbb{R}^n)$.

We shall now see under what conditions we can assure that an SDE has a solution that is unique.

Definition 4.4. *A function* $f : T \times \mathbb{R}^n \to \mathbb{R}^n$ *satisfies a*

1. *Lipschitz condition in x if there exists a constant $K \in \mathbb{R}$ such that*

$$|f(t,x) - f(t,y)| \leqslant K|x-y| \quad \textit{for all } t \in T \textit{ and for all } x, y \in \mathbb{R}^n \tag{21}$$

2. *Linear growth condition if there exists a constant $K \in \mathbb{R}$ such that*

$$|f(t,x) - f(t,y)| \leqslant K(1+|x|) \quad \textit{for all } t \in T \textit{ and for all } x, y \in \mathbb{R}^n \tag{22}$$

Notice that the Lipschitz condition implies uniform continuity in x. On the other hand, linear growth condition can be written as

$$|f(t,x) - f(t,y)| \leqslant K^2(1+|x|^2) \text{ for all } t \in T \text{ and for all } x \in \mathbb{R}^n.$$

Theorem 4.5 (Existence and Uniqueness)**.** *Consider the SDE* (19) *with coefficients μ, σ_k, $k = 1, \ldots, m$ satisfying the Lipschitz and linear growth conditions. If ξ is $\mathcal{F}_0$ measurable then there exists a unique strong solution X to* (19) *such that $X_{t_0} = \xi$. If, in addition, $\mathbb{E}\left[|\xi|^2\right] < \infty$ then $X \in \mathcal{I}(T \times \Omega; \mathbb{R}^n)$.*

Proof. Suppose that X_t and Y_t are solutions of (19).

$$\begin{aligned}
\mathbb{E}[|X_t - Y_t|^2] &= \mathbb{E}\left[\left|\int_{t_0}^t (\mu(s,X_s) - \mu(s,Y_s))ds + \int_{t_0}^t (\sigma(s,X_s) - \sigma(s,Y_s))dBs\right|^2\right] \\
&\overset{(*)}{\leqslant} 2\mathbb{E}\left[\left|\int_{t_0}^t (\mu(s,X_s) - \mu(s,Y_s))ds\right|^2\right] + 2\mathbb{E}\left[\left|\int_{t_0}^t (\sigma(s,X_s) - \sigma(s,Y_s))dBs\right|^2\right] \\
&\overset{(**)}{\leqslant} (t-t_0)\mathbb{E}\left[\int_{t_0}^t |\mu(s,X_s) - \mu(s,Y_s)|^2 ds\right] + \mathbb{E}\left[\int_{t_0}^t |\sigma(s,X_s) - \sigma(s,Y_s)|^2 ds\right] \\
&\overset{(***)}{\leqslant} K^2(t-t_0)\int_{t_0}^t \mathbb{E}\left[|X_s - Y_s|^2 ds\right] + K^2 \int_{t_0}^t \mathbb{E}\left[|X_s - Y_s|^2 ds\right] \\
&= K^2(t-t_0+1)\int_{t_0}^t \mathbb{E}\left[|X_s - Y_s|^2 ds\right],
\end{aligned}$$

where in (*) we have used that $(a+b)^2 \leqslant 2(a^2+b^2)$, in (**) we have used the Cauchy-Schwarz inequality and the Itô isometry while in (***) we have used the Lipschitz condition on μ and σ. Using Gronwall's integral inequality (see Lemma 10.3 in Appendix A) we conclude that

$$\mathbb{E}[|X_t - Y_t|^2] \leqslant 0.$$

Then X_t is almost surely equal to Y_t and the uniqueness is proved.

To prove that a solution Y_t exists, we shall use the Picard iteration method. We define

$$Y_t^{(k+1)} := \xi + \int_{t_0}^{t} \mu\left(s, Y_s^{(k)}\right) ds + \int_{t_0}^{t} \sigma\left(s, Y_s^{(k)}\right) dB_s.$$

Our goal is to show that $Y_t^{(k+1)}$ converges to a solution of (19). This proof is similar to the one used to show that there was a continuous version of the Itô integral. The method consists in bounding the probabilities

$$P\left(\max_{t\in T} \left|Y^{(k+1)}(t) - Y^{(k)}(t)\right| > 2^{-k}\right) \leqslant C_k$$

by constants C_k such that

$$\sum_{k\in\mathbb{N}} C_k < \infty$$

and applying Borel-Cantelli to conclude that the sequence is uniformly convergent. For this reason the first step of proving the existence will be to show that

$$\mathbb{E}\left[\left|Y^{(k+1)}(t) - Y^{(k)}(t)\right|^2\right] \leqslant C_{k,t} := \frac{(\alpha(t-t_0))^{k+1}}{(k+1)!}. \tag{23}$$

for some constant $\alpha \geq 2K^2(t_f - t_0 + 1) \in \mathbb{R}_+$, for all $t \in T$ and $k \in \mathbb{N}$. We shall do this by induction.

1. $k = 0$: We have that

$$\begin{aligned}
\mathbb{E}\left[\left|Y^{(1)}(t) - Y^{(0)}(t)\right|^2\right] &= \mathbb{E}\left[\left|\int_{t_0}^{t} \mu\left(s, Y^{(0)}(s)\right) ds + \int_{t_0}^{t} \sigma\left(s, Y^{(0)}(s)\right) dB_s\right|^2\right] \\
&\overset{(*)}{\leqslant} 2\mathbb{E}\left[\left|\int_{t_0}^{t} K(1 + |\xi(s)|) ds\right|^2\right] + 2\mathbb{E}\left[\int_{t_0}^{t} K^2\left(1 + |\xi(s)|^2\right) ds\right] \\
&\overset{(**)}{\leqslant} 2K^2(t - t_0 + 1)\mathbb{E}\left[\left|\int_{t_0}^{t} (1 + |\xi(s)|)^2 ds\right|\right] \\
&\leqslant \alpha
\end{aligned}$$

for an appropriately large α, which exists because $\mathbb{E}[|\xi|] < \infty$. In (*) we have used that $(a+b)^2 \leqslant 2(a^2+b^2)$, the Itô isometry and the bounded growth condition on μ and σ. For (**) we have used Cauchy-Schwarz.

2. Suppose now (23). Then abbreviating $\mu\left(s, Y^{(k)}(s)\right)$ by $\mu_k(s)$ and likewise for σ we have that

$$\begin{aligned}
&\mathbb{E}\left[\left|Y^{(k+2)}(t) - Y^{(k+1)}(t)\right|^2\right] \\
&= \mathbb{E}\left[\left|\int_{t_0}^{t} (\mu_{k+1}(s) - \mu_k(s))ds + \int_{t_0}^{t} (\sigma_{k+1}(s) - \sigma_k(s))dB_s\right|^2\right] \\
&\overset{(***)}{\leqslant} 2K^2(t_f - t_0)\mathbb{E}\left[\int_{t_0}^{t} |Y^{(k+1)}(s) - Y^{(k)}(s)|^2 ds\right] \\
&\quad + 2K^2\mathbb{E}\left[\int_{t_0}^{t} |Y^{(k+1)}(s) - Y^{(k)}(s)|^2 ds\right] \\
&\overset{(****)}{\leqslant} 2K^2(t_f - t_0 + 1)\int_{t_0}^{t} \frac{(\alpha(t - t_0))^{k+1}}{(k+1)!} ds \\
&\leqslant \frac{(\alpha(t - t_0))^{k+2}}{(k+2)!},
\end{aligned}$$

where in (***) we have used $(a+b)^2 \leqslant 2(a^2+b^2)$, the Itô isometry, the Lipschitz condition on μ and σ and Cauchy-Schwarz and in (****) we have used the hypothesis of induction.

Now by Doob's martingale inequality we have that

$$\begin{aligned}
P\left(\sup_{t\in T}\left|Y^{(k+1)}(t) - Y^{(k)}(t)\right|^2 > 2^{-k}\right) &\leqslant 2^{2k}\mathbb{E}\left[\left|Y^{(k+1)}(t_f) - Y^{(k)}(t_f)\right|^2\right] \\
&\leqslant 2^{2k}\frac{(\alpha t_f)^{k+1}}{(k+1)!};
\end{aligned}$$

so with the same procedure used when we showed that there exists a continuous modification of the Itô integral, by the Borel-Cantelli lemma we conclude that $\{Y^n\}_{n\in\mathbb{N}}$ converges uniformly for all $t \in T$ and almost all $w \in \Omega$ to a limit, denoted by Y. Passing to the limit it is clear that Y verifies (19). Clearly Y is progressively measurable with respect to $\{\mathcal{M}_t\}_{t\in T}$ and pathwise continuous as it is the uniform limit of the sequence $\{Y^{(n)}\}_{n\in\mathbb{N}}$ with these properties. Furthermore, by (23), $\{Y^{(n)}\}_{n\in\mathbb{N}}$ is a Cauchy sequence in $\mathcal{L}^2(T\times\Omega;\mathbb{R}^n)$. Hence its limit Y also belongs to $\mathcal{L}^2(T\times\Omega;\mathbb{R}^n)$ and we have that $Y \in \mathcal{I}(T\times\Omega;\mathbb{R}^n)$, which concludes the proof. □

4.2 Stability of Solutions

As is the case in ODE, when the conditions of the theorem of existence and uniqueness are verified, then the SDE has stable solutions. Simply put, this

means that a slight disturbance in the solution's initial conditions produces only a small change in the solution

Theorem 4.6. *Suppose that X_t and Y_t are solutions of the SDE* (19) *with initial conditions $X_{t_0} = \xi$ and $X_{t_0} = \phi$ respectively. Then if the coefficient functions μ and σ of* (19) *verify the Lipschitz condition* (21) *and ξ and ϕ are $\mathcal{F}_0$ measurable variables such that $\xi, \phi \in \mathcal{L}^2(\Omega; \mathbb{R}^n)$, then it holds that*

$$\mathbb{E}\left[\sup_{t\leqslant T}|X_t - Y_t|^2\right] \leqslant 3e^{3K^2(t_f-t_0+4)(t_f-t_0)}\mathbb{E}\left[|\xi-\eta|^2\right].$$

Proof.

$$\begin{aligned}
&\mathbb{E}\left[\sup_{t\in T}|X_t - Y_t|^2\right] \\
&= \mathbb{E}\left[\sup_{t\in T}\left|\xi - \phi + \int_{t_0}^{t}(\mu(s,X_s)-\mu(s,Y_s))ds| + \int_{t_0}^{t}(\sigma(s,X_s)-\sigma(s,Y_s))dBs\right|^2\right] \\
&\overset{(*)}{\leqslant} 3\mathbb{E}\left[|\xi-\phi|^2\right] + 3\mathbb{E}\left[\sup_{t\in T}\left|\int_{t_0}^{t}(\mu(s,X_s)-\mu(s,Y_s))ds\right|^2\right] \\
&\quad + 3\mathbb{E}\left[\sup_{t\in T}\left|\int_{t_0}^{t}(\sigma(s,X_s)-\sigma(s,Y_s))dBs\right|^2\right] \\
&\overset{(**)}{\leqslant} 3\mathbb{E}\left[|\xi-\phi|^2\right] + (t_f-t_0)3\mathbb{E}\left[\int_T|\mu(s,X_s)-\mu(s,Y_s)|^2ds\right] \\
&\quad + 12\mathbb{E}\left[\int_T|\sigma(s,X_s)-\sigma(s,Y_s)|^2ds\right] \\
&\overset{(***)}{\leqslant} 3\mathbb{E}\left[|\xi-\phi|^2\right] + 3K^2(t_f-t_0)\int_T\mathbb{E}\left[|X_s-Y_s|^2ds\right] + 12K^2\int_T\mathbb{E}\left[|X_s-Y_s|^2ds\right] \\
&= 3\mathbb{E}\left[|\xi-\phi|^2\right] + 3K^2(t_f-t_0+4)\int_T\mathbb{E}\left[|X_s-Y_s|^2ds\right] \\
&\leqslant 3\mathbb{E}\left[|\xi-\phi|^2\right] + 3K^2(t_f-t_0+4)\int_T\mathbb{E}\left[\sup_{s\in T}|X_s-Y_s|^2ds\right],
\end{aligned}$$

where in (*) we have used that $(a+b+c)^2 \leqslant 3a^2+3b^2+3c^2$, in (**) we have used the Cauchy-Schwarz inequality and Lemma 3.12 and in (***) we have used the Lipschitz condition on μ and σ. Now we conclude, by Gronwall

inequality, that

$$\mathbb{E}\left[\sup_{t\leqslant T}|X_t - Y_t|^2\right] \leqslant 3e^{3K^2(t_f-t_0+4)(t_f-t_0)}\mathbb{E}\left[|\xi-\phi|^2\right].$$

as we wished to see. □

5 Explicitly solvable SDEs

5.1 Linear SDEs

As is the case with ODEs the analytical solution of a linear SDE can be given in closed form, which we examine the procedure in this section.

Definition 5.1 (Linear SDE)**.** *An SDE is said to be linear if it can be written as*

$$dX_t = (F(t)X_t + f(t))dt + \sum_{k=1}^{m}\left(G_k(t)X_t + g_k(t)\right)dB_k(t); \quad t \in T \tag{24}$$

where $f, g\colon T \to \mathbb{R}^n$; $\quad F, G_k : T \to \mathbb{R}^{nn}$; $\quad 1 \leq k \leq m$. *Furthermore, a linear SDE is said to be homogeneous if it is of the form*

$$dX_t = F(t)X_t dt + \sum_{k=1}^{m} G_k(t)X_t dB_k(t); \quad t \in T. \tag{25}$$

In the remainder of this section we will suppose that the functions f, g, F_k, G_k verify the existence and uniqueness conditions of Theorem 4.5.

Definition 5.2. *The fundamental solution* Φ *of the homogeneous linear equation* (25) *is the stochastic process*

$$\Phi : T \times \Omega \to \mathbb{R}^{nn}$$

such that Φ_i *is the solution to said SDE with initial condition* $X(t_0) = (0, ..., 0, \overset{i}{1}, 0, ..., 0)$.

This solution receives the name of fundamental the solution of (25) with any other initial condition can be expressed in terms of Φ.

Lemma 5.3. *Suppose that* Y *is the solution to* (25) *with initial condition* $Y(t_0) = y_0$ *then*

$$Y = \Phi y_0$$

Proof. We define $Z := \Phi y_0$ and wish to see that $Z = Y$. By definition of Φ we have that $\Phi(t_0) = I_n$ hence

$$Z(t_0) = y_0 = Y(t_0).$$

Now by Theorem 4.5 it remains to see that Z is a solution to (25). This is also simple to see since

$$\begin{aligned} dZ(t) = d\Phi(t)y_0 &= F(t)\Phi(t)y_0 dt + \sum_{k=1}^{m} G_k(t)\Phi(t)y_0 dB_k(t) \\ &= F(t)Z(t)dt + \sum_{k=1}^{m} G_k(t)Z(t)dB_k(t). \end{aligned}$$

□

Using the fundamental solution, Φ of (25) we can obtain the solution of the general equation (24) using the following theorem.

Theorem 5.4. *The solution of* (24) *with initial condition* y_0 *is*

$$Y(t) = \Phi(t)\Psi(t)$$

where

$$\Psi(t) = y_0 + \int_{t_0}^{t} \Phi^{-1}(s)\left(f(s) - \sum_{k=1}^{m} G_k(s)g_k(s)\right)ds + \sum_{k=1}^{m}\int_{t_0}^{t} \Phi^{-1}(s)g_k(s)dB_k(s)$$

Proof. It is convenient to note that that Φ is invertible on T since using the Itô formula it is possible to show that $\mathrm{Det}(\Phi(t)) > 0$ almost surely for all $t \in T$. Therefore the function $\Psi(t)$ makes sense and

$$d\Psi(t) = \Phi^{-1}(t)\left(f(t) - \sum_{k=1}^{m} G_k(t)g_k(t)\right)dt + \sum_{k=1}^{m} \Phi^{-1}(t)g_k(t)dB_k(t)$$

Using the Itô formula to calculate $dY_t = d(h(\Phi(t), \Psi(t)))$ and the function with $h(x_1, x_2) = x_1 x_2$ we obtain

$$
\begin{aligned}
dY(t) &= d(\Phi(t)\Psi(t)) = d\Phi(t)\Psi(t) + \Phi(t)d\Psi(t) + d\Phi(t)d\Psi(t). \\
&= \left(F(t)\Phi(t)dt + \sum_{k=1}^{m} G_k(t)\Phi(t)dB_k(t) \right) \Psi(t) \\
&\quad + \Phi(t)\Phi^{-1}(t) \left(f(t)dt - \sum_{k=1}^{m} G_k(t)g_k(t)dt + \sum_{k=1}^{m} g_k(t)dB_k(t) \right) \\
&\quad + \left(F(t)\Phi(t)dt + \sum_{k=1}^{m} G_k(t)\Phi(t)dB_k(t) \right) \Phi^{-1}(t) \\
&\qquad\qquad \times \left(f(t)dt - \sum_{k=1}^{m} G_k(t)g_k(t)dt + \sum_{k=1}^{m} g_k(t)dB_k(t) \right) \\
&= (F(t)Y(t) + f(t))dt + \sum_{k=1}^{m} \left(G_k(t)Y(t) + g_k(t) \right) dB_k(t)
\end{aligned}
$$

where in the last equality we have used (18). Hence by uniqueness $Y(t) = \Phi(t)\Psi(t)$ is the solution to (24) with initial condition y_0. □

When the SDE (25) is one-dimensional then we can actually calculate $\Phi(t)$ as follows.

Theorem 5.5. *The fundamental solution $\Phi(t)$ of the homogeneous one-dimensional linear SDE* (25) *is*

$$
\Phi(t) = \exp\left(\int_{t_0}^{t} \left(F(s) - \frac{1}{2}\sum_{k=1}^{m} G_k^2(s) \right) ds + \sum_{k=1}^{m} \int_{t_0}^{t} G_k(s)dB_k(s) \right)
$$

Proof. In the scalar case the fundamental solution is the solution to the homogeneous equation with initial condition $y_0 = 1$. By Itô lemma applied to the stochastic process

$$
Y(t) := \int_{t_0}^{t} \left(F(s) - \frac{1}{2}\sum_{k=1}^{m} G_k^2(s) \right) ds + \sum_{k=1}^{m} \int_{t_0}^{t} G_k(s)dB_k(s)
$$

and the function $x \mapsto e^x$ we have that

$$\begin{aligned} d\Phi(t) &= Y(t)dY(t) + \frac{1}{2}Y(t)\sum_{k=1}^{m} G_k(t)^2 dt \\ &= Y(t)\left(\left(F(t) - \frac{1}{2}\sum_{k=1}^{m} G_k^2(t)\right)dt + \sum_{k=1}^{m} G_k(t)dB_k(t)\right) + \frac{1}{2}Y(t)\sum_{k=1}^{m} G_k^2(t)dt \\ &= Y(t)\left(F(t) + \sum_{k=1}^{m} G_k(t)dB_k(t)\right) \end{aligned}$$

as we wished to see. □

5.2 Reducible SDEs

Given an Itô process X_t as the solution of an SDE, Itô lemma gives the SDE fulfilled by the process $f(X_t)$; in some cases the transformed equation is a linear SDE.

Suppose that $f : T \times \mathbb{R} \to \mathbb{R}$ verifies the conditions of Lemma 3.16 and X_t is a solution of the one-dimensional SDE

$$dX_t = \mu(t, X_t)\, dt + \sigma(t, X_t)dB_t. \tag{26}$$

Then, by Itô lemma we have that $Y_t := f(t, X_t)$ is an Itô process with drift coefficient

$$\overline{\mu}(t, X_t) = \frac{\partial f}{\partial t}(t, X_t) + \frac{\partial f}{\partial x}(t, X_t)\mu(t, X_t) + \frac{1}{2}\frac{\partial^2 f}{\partial x^2}(t, X_t)\sigma^2(t, X_t)$$

and diffusion coefficient

$$\overline{\sigma}(t, X_t) = \frac{\partial f}{\partial x}(t, X_t)\sigma(t, X_t).$$

Our desire is to find the conditions for the existence of an invertible f such that the resulting stochastic differential equation

$$dY_t = \overline{\mu}(t, Y_t)\, dt + \overline{\sigma}(t, Y_t)dB_t \tag{27}$$

has a explicitly solvable solution. If this were the case, we could obtain the explicit solution of (26) by applying the inverse transform f^{-1} to the solution of (27). As we have just seen, we can calculate the solution of one-dimensional linear SDEs; then, the possibility of transforming our equation into a linear SDE is highly desirable. Next lemma shows a simple procedure to check if we can transform our SDE into an homogeneous linear SDE of the form

$$dY_t = \overline{\mu}(t)dt + \overline{\sigma}(t)dB_t. \tag{28}$$

Lemma 5.6. *The one dimensional SDE* (26) *can be reduced to a one-dimensional homogeneous equation of the form* (28) *if and only if*

$$\frac{\partial}{\partial x}\left(\frac{1}{\sigma(t,x)}\frac{\partial\sigma}{\partial t}(t,x)-\sigma(t,x)\frac{\partial}{\partial x}\left(\frac{\mu(t,x)}{\sigma(t,x)}-\frac{1}{2}\frac{\partial\sigma}{\partial x}(t,x)\right)\right)=0. \tag{29}$$

Proof. Suppose that (26) can be reduced to (28) by means of a function $f : T\times\mathbb{R}\to\mathbb{R}$ verifying the conditions of Lemma 3.16. Then, by Itô formula

$$\overline{\mu}(t)=\frac{\partial f}{\partial t}(t,X_t)+\frac{\partial f}{\partial x}(t,X_t)\mu(t,X_t)+\frac{1}{2}\frac{\partial^2 f}{\partial x^2}(t,X_t)\sigma^2(t,X_t) \tag{30}$$

$$\overline{\sigma}(t)=\frac{\partial f}{\partial x}(t,X_t)\sigma(t,X_t). \tag{31}$$

By (30) we have that

$$\frac{\partial\overline{\mu}}{\partial x}(t)=0=\frac{\partial^2 f}{\partial t\partial x}(t,x)+\frac{\partial}{\partial x}\left(\mu(t,x)\frac{\partial f}{\partial x}(t,x)+\frac{1}{2}\sigma^2(t,x)\frac{\partial^2 f}{\partial x^2}(t,x)\right) \tag{32}$$

and by (31) we know that

$$\begin{aligned}\frac{\partial f}{\partial x}(t,x)&=\frac{\overline{\sigma}(t)}{\sigma(t,x)}\\ \frac{\partial^2 f}{\partial x^2}(t,x)&=-\frac{\overline{\sigma}(t)\frac{\partial\sigma}{\partial x}(t,x)}{\sigma^2(t,x)}\\ \frac{\partial^2 f}{\partial t\partial x}(t,x)&=\frac{1}{\sigma^2(t,x)}\left(\frac{\partial\overline{\sigma}}{\partial t}(t)\sigma(t,x)-\overline{\sigma}(t)\frac{\partial\sigma}{\partial t}(t,x)\right);\end{aligned}$$

so substituting the values of these derivatives in (32) we obtain

$$\frac{\partial\overline{\sigma}(t)}{\partial t}\frac{1}{\overline{\sigma}(t)}=\frac{1}{\sigma(t,x)}\frac{\partial\sigma}{\partial t}(t,x)-\sigma(t,x)\frac{\partial}{\partial x}\left(\frac{\mu(t,x)}{\sigma(t,x)}-\frac{1}{2}\frac{\partial\sigma}{\partial x}(t,x)\right).$$

Finally differentiating with respect to x we conclude that a necessary condition for the existence of f is

$$0=\frac{\partial}{\partial x}\left(\frac{1}{\sigma(t,x)}\frac{\partial\sigma}{\partial t}(t,x)-\sigma(t,x)\frac{\partial}{\partial y}\left(\frac{\mu(t,x)}{\sigma(t,x)}-\frac{1}{2}\frac{\partial\sigma}{\partial y}(t,x)\right)\right).$$

Reciprocally if (29) holds then working backward we have that

$$\begin{aligned}&U(t,x)\\ &=C\exp\left(\int_0^t\left(\frac{1}{\sigma(s,x)}\frac{\partial\sigma}{\partial s}(s,x)-\sigma(s,x)\frac{\partial}{\partial x}\left(\frac{\mu(s,x)}{\sigma(s,x)}-\frac{1}{2}\frac{\partial\sigma}{\partial x}(s,x)\right)\right)ds\right)\int_0^x\frac{1}{\sigma(t,z)}dz\end{aligned} \tag{33}$$

reduces the SDE (26) to the desired SDE (28). □

6 Numerical Methods for Solving SDEs

Given an SDE

$$dX_t = \mu(t, X_t)dt + \sigma(t, X_t)dB_t \tag{34}$$

with solution X a numerical method is an algorithm that is used to find an approximation of the exact solution X. In this section we analyze some of the most commonly used numerical methods used for solving SDEs. But before doing so we will define the stochastic analogous of the properties we desire of our numerical method. Throughout this section we consider that the SDE whose solution we wish to approximate is of the form (19) and verifies the hypothesis of the existence and uniqueness theorem. its solution will be denoted X and we shall consider a uniform discretization

$$P^h = \{t_0, t_0 + h, ..., t_f - h, t_f\}$$

of the time interval $T = [t_0, t_f]$ and a numerical scheme

$$Y^h : P^h \times \Omega \to \mathbb{R}^n$$

whose goal is to approximate the values of X at each point of P. For convenience we will denote $Y^h(t_0 + nh)$ as $Y^h(t_n)$, Y_n^h or Y_n if h is clear from the context. In addition, we will denote the cardinal of P^h as $|P^h|$. The most simple example of a numerical method for solving an SDE is the Euler-Maruyama scheme. This numerical method, named after Leonhard Euler and Gisiro Maruyama, is a generalization of the Euler method for ODEs and can be written as

$$\begin{cases} Y_0 = \xi \\ Y_n = Y_{n-1} + h\mu(t_{n-1}, Y_{n-1}) + \sigma(t_{n-1}, Y_{n-1})\Delta B_n; \quad 1 \leqslant n \leqslant |P^h| \end{cases} \tag{35}$$

where B is a Wiener process and $\Delta B_n := B_{t_{n+1}} - B_{t_n}$. It is worth noting that the definitions we will now give are easily generalizable in the case that the discretization is not uniform by setting h to be the maximum step size.

6.1 Convergence of Numerical Methods

The convergence of a numerical scheme is a property that tells us how well this scheme approximates in some sense the exact one when the step size goes to 0. We can differentiate between strong convergence and weak convergence of numerical methods.

Definition 6.1 (Strong Convergence at t)**.** *We shall say that the numerical solution Y^h converges strongly to X at time $t \in P^h$ if*

$$\lim_{h \downarrow 0} \mathbb{E}\left[\left|X(t) - Y^h(t)\right|\right] = 0.$$

We say that this convergence is of order p if

$$\mathbb{E}\left[\left|X(t) - Y^h(t)\right|\right] = \mathcal{O}(h^p)$$

i.e. if there exists a constant $K \in \mathbb{R}$ and a step size h_0 such that

$$\mathbb{E}\left[\left|X(t) - Y^h(t)\right|\right] \leqslant Kh^p$$

for all $h \in [0, h_0]$.

Definition 6.2 (Weak Convergence at t)**.** *We shall say that Y^h converges weakly to X at time $t \in P^h$ with respect to a class of functions $\mathcal{D} \subset \mathbb{R}^{\mathbb{R}^n}$ if*

$$\lim_{h \downarrow 0} \left|\mathbb{E}\left[f(X(t))\right] - \mathbb{E}\left[f(Y^h(t))\right]\right| = 0$$

for all $f \in \mathcal{D}$. Analogously we say that the weak convergence is of order p

$$\left|\mathbb{E}\left[f\left(X(t)\right)\right] - \mathbb{E}\left[f(Y^h(t))\right]\right| = \mathcal{O}(h^p)$$

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

For these definitions we can also consider (instead of convergence at only one instant of time) convergence in the whole time interval T by requiring the respective bound to hold for each $t \in P^h$. For example, Y^h converges strongly to X if

$$\mathbb{E}\left[\left|X(t) - Y^h(t)\right|\right] = 0; \quad \forall t \in P^h$$

6.2 Convergence of the Euler-Maruyama Method

Theorem 6.3. *The Euler-Maruyama method* (35) *has strong order of convergence of $\frac{1}{2}$.*

Proof. For simplicity we shall conduct the proof in the scalar case. The method for proving this statement is to define a step function $\overline{Y^h}$ which is equal to the Euler-Maruyama approximation Y^h at each point in the discretization P^h and then applying Gronwall's inequality on

$$f(t) := \sup_{s \in [t_0, t]} \mathbb{E}(|\overline{Y^h}(s) - X(s)|^2).$$

Defining the step function $\overline{Y^h}$ facilitates the proof as it is defined on the real interval T, as opposed to Y^h which is only defined on P, thus allowing us to use Gronwall's inequality on f in order to reach the desired bound. The step function $\overline{Y^h}$ is defined as follows

$$\overline{Y^h}(t) := \sum_{k=0}^{|P^h|-1} Y_n^h \mathcal{X}_{[t_n,t_{n+1})}(t).$$

Let $t \in [t_n, t_{n+1})$ we have that

$$\begin{aligned}
\overline{Y}(t) = Y_n &= Y_0 + \sum_{k=0}^{n-1} \left(Y_{k+1} - Y_k\right) \\
&= Y_0 + \sum_{k=0}^{n-1} \left(h\mu\left(t_k, Y_k\right) + \sigma\left(t_k, Y_k\right) \Delta B_k\right) \\
&= Y_0 + \sum_{k=0}^{n-1} \int_{t_k}^{t_{k+1}} \mu\left(t_k, Y_k\right) ds + \sum_{k=0}^{n-1} \int_{t_k}^{t_{k+1}} \sigma\left(t_k, Y_k\right) dB_s \\
&= Y_0 + \int_{t_0}^{t_n} \mu(t_s, \overline{Y}(s)) ds + \int_{t_0}^{t_n} \mu(t_s, \overline{Y}(s)) dB_s
\end{aligned}$$

so abbreviating $\mu(t_s, \overline{Y^h}(t_s))$ by $\mu_s^{\overline{Y^h}}$ and likewise for σ we have that

$$
\begin{aligned}
&\mathbb{E}\left[\left|\overline{Y^h}(t) - X(t)\right|^2\right] \\
&= \mathbb{E}\left[\left|\int_{t_0}^{t_n}\left(\mu_s^{\overline{Y^h}} - \mu_s^X\right)ds + \int_{t_0}^{t_n}\left(\sigma_s^{\overline{Y^h}} - \sigma_s^X\right)dB_s - \int_{t_n}^{t}\mu_s^X ds - \int_{t_n}^{t}\sigma_s^X dB_s\right|^2\right] \\
&\overset{(*)}{\leqslant} 4(t_n - t_0)\mathbb{E}\left[\int_{t_0}^{t_n}\left|\mu_s^{\overline{Y^h}} - \mu_s^X\right|^2 ds\right] + 4\mathbb{E}\left[\int_{t_0}^{t_n}\left|\sigma_s^{\overline{Y^h}} - \sigma_s^X\right|^2 ds\right] \\
&\quad + 4(t - t_n)\mathbb{E}\left[\int_{t_n}^{t}\left|\mu_s^X\right|^2 ds\right] + 4\mathbb{E}\left[\int_{t_n}^{t}\left|\sigma_s^X\right|^2 dB_s\right] \\
&\overset{(**)}{\leqslant} 4(t_n - t_0 + 1)K^2\mathbb{E}\left[\int_{t_0}^{t_n}\left|\overline{Y}(s) - X(s)\right|^2 ds\right] \\
&\quad + 4K(t - t_n + 1)\mathbb{E}\left[\int_{t_n}^{t}(1 + |X_s|^2)ds\right] \\
&\overset{(***)}{\leqslant} 4(t_f - t_0 + 1)K^2\int_{t_0}^{t} f(s)ds + 4Kh(h+1)\sup_{s\in T}E[|X_s|^2] \\
&:= h(B_1 + hB_2) + C\int_{t_0}^{t} f(s)ds
\end{aligned}
$$

where in (*) we have used that $(a+b+c+d)^2 \leqslant 4(a^2+b^2+c^2+d^2)$, the Cauchy-Schwarz inequality and the Itô isometry, in (**) we employed the Lipschitz and linear boundedness condition of μ and σ and in (***) we used Theorem 4.5. Therefore we arrive at

$$f(t) \leqslant h(B_1 + hB_2) + C\int_{t_0}^{t} f(s)ds$$

which means by Gronwall's inequality that

$$f(t) \leqslant h(B_1 + hB_2)e^{C(t-t_0)} \leqslant h(B_1 + hB_2) + +e^{C(t_f - t_0)} := h(B_1 + hB_2)K^2.$$

Hence

$$
\begin{aligned}
\max_{t\in P^h}\mathbb{E}\left[\left|X(t) - Y^h(t)\right|\right] &\leqslant \sup_{t\in T}\mathbb{E}\left[\left|X(t) - \overline{Y^h}(t)\right|\right] \\
&\leqslant \sqrt{f(t)} \leqslant K\sqrt{h(B_1 + hB_2)} = \mathcal{O}(\sqrt{h})
\end{aligned}
$$

which concludes the theorem. □

There is also a well known result regarding the weak convergence of the Euler-Maruyama scheme. We will state it without the proof which can be found at [20].

Theorem 6.4. *The Euler-Maruyama method 35 applied to an SDE which verifies the conditions of the theorem of existence an uniqueness has a weak order of convergence of* 1.

6.3 The Milstein Method

The Milstein scheme is a higher order numerical method that owes its name to Grigori Milstein [19]. The Milstein approximation $Y_t := Y_t$ for a one dimensional SDE can be defined as follows

$$\begin{cases} Y_0 = \xi \\ Y_n = Y_{n-1} + h\mu(t_{n-1}, Y_{n-1}) + \sigma(t_{n-1}, Y_{n-1})\Delta B_n \\ \qquad\qquad + \frac{1}{2}\sigma(t_{n-1}, Y_{n-1})\frac{\partial \sigma}{\partial x}(t_{n-1}, Y_{n-1})\left(\Delta B_n^2 - \Delta t\right). \end{cases} \tag{36}$$

In the most general case, where we have an n-dimensional SDE driven by an m-dimensional Wiener process the i-th component of the Milstein method is

$$Y_i(t_{p+1}) = Y_i(t_p) + h\mu_i + \sum_{j=1}^{m} \sigma_{j,i}\Delta B_i + \sum_{j,k=1}^{m}\sum_{l=1}^{n} \sigma_{jl}\frac{\partial \sigma_{ki}}{\partial x_l}\int_{t_p}^{t_{k+1}}\int_{t_p}^{t} dB_j(s)dB_k(t) \quad 1 \leq i \leq n$$

where μ_i, σ_i and their derivatives are evaluated at $(t_p, Y(t_p))$. For the Milstein method we have the following theorem

Theorem 6.5. *The numerical scheme* (6.3) *has a strong and weak order of convergence of* 1.

This theorem tells us that the Milstein method improves on the Euler-Maruyama scheme in the sense that its strong order of convergence is higher. However, the order of weak convergence remains the same.

6.4 General Itô-Taylor approximations

In this section we will generalize how to obtain numerical methods of any strong order based on here Itô expansion of the SDE in question. We will present here a series of results since they are exceedingly useful. However their proof is outside of the scope of this work, a complete exposition can be found

at [16]. First of all we must start with a series of definitions. A multi-index α of length $k \in \mathbb{N}_0$ is a vector of length k whose components are natural numbers. We denote by v the multi index of length 0 and define $\mathcal{M}_m$ to be the set of all multi-indices whose indices are at most m, that is,

$$\mathcal{M}_m := \left(\bigcup_{n \in \mathbb{N}} \prod_{n} \{1, 2, ..., m\} \right) \bigcup v.$$

We define l and z as the functions defined on $\mathcal{M}_m$ that return the length and number of zeroes of multi-indices. Given a subset of multi-indices $\mathcal{A} \in \mathcal{M}_m$ we define its remainder set $\mathcal{B}(\mathcal{A})$ as

$$\mathcal{B}(\mathcal{A}) := \{\alpha = (\alpha_1, \alpha_2..., \alpha_n) \in \mathcal{M}_m \backslash \mathcal{A} \,|\quad (\alpha_2..., \alpha_n) \in \mathcal{A}\}\,.$$

We say that set of multi-indices $\mathcal{A}$ is hierarchical if

1. $\sup_{\alpha \in \mathcal{A}} l(\alpha) < \infty$.
2. $(\alpha_1, \alpha_2, ..., \alpha_n) \in \mathcal{A}$ and $l(\alpha) \geq 2 \implies (\alpha_2, ..., \alpha_n) \in \mathcal{A}$.

Given a multi-index $\alpha = (\alpha_1, \alpha_2, ..., \alpha_n)$, and a Brownian motion B we define the iterated integral $I_{\alpha,t_1,t_2}(f)$ of the stochastic process f between the times t_1 and t_2 as

$$I_{\alpha,t_1,t_2}(f) := \int_{t_1}^{t_2} \int_{t_1}^{s_2} ... \int_{t_1}^{s_n} f(s_1) dB^{\alpha_1}(s_1) dB^{\alpha_2}(s_2) ... dB^{\alpha_n}(s_n)$$

and

$$I_{\alpha,t_1,t_2} := I_{\alpha,t_1,t_2}(1)$$

where we have denoted $dB^0(s) = ds$ so as to not over complicate the notation. Additionally f must obviously be a stochastic process such that the successive Itô integrals are defined. With these definitions we can give a fairly compact construction for the Itô-Taylor expansion of a given Itô process.

Theorem 6.6. *If*

$$X_t = X_{t_0} + \int_{t_0}^{t} \mu\left(s, X_s\right) ds + \sum_{j=1}^{m} \int_{t_0}^{t} \sigma_j\left(s, X_s\right) dB_j(s); \quad t \in T$$

is an n-dimensional Itô process with respect to an m-dimensional Wiener process and $\mathcal{A} \in \mathcal{M}_m$ is a hierarchical set of multi-indices such that then

$$f(t, X_t) = f(t_0, X_{t_0}) + \sum_{\alpha \in \mathcal{A}} L^{\alpha}(f)(t_0, X_{t_0}) I_{\alpha,t_0,t} + \sum_{\alpha \in \mathcal{B}(\mathcal{A})} I_{\alpha,t_0,t}(L^{\alpha}(f)) \quad (37)$$

where $f : T \times \mathbb{R}^n \to \mathbb{R}$ *is a process such that* (37) *is well defined and*

$$L^0 := \frac{\partial}{\partial t} + \sum_{i=1}^{n} \mu_i \frac{\partial}{\partial x^i} + \frac{1}{2} \sum_{i,j=1}^{n} \sum_{k=1}^{m} \sigma_{k,i}\sigma_{k,j} \frac{\partial^2}{\partial x^i \partial x^j}$$

$$L^k := \sum_{i=1}^{n} \sigma_{k,i} \frac{\partial}{\partial x_i}; \quad k \in \mathbb{N}$$

$$L^\alpha := L^{\alpha_1} L^{\alpha_2} ... L^{\alpha_k}; \quad \alpha = (\alpha_1, \alpha_2 ..., \alpha_k) \in \mathcal{M}_m.$$

(37) *is called the Itô-Taylor expansion of* $f(t, X_t)$ *with respect to* $\mathcal{A}$.

For example if X_t is a two dimensional process with respect to a one-dimensional Brownian motion B and

$$\mathcal{A} = \{v, (0), (1)\}$$

with remainder set

$$\mathcal{B}(\mathcal{A}) = \{(0,0), (0,1), (1,0), (1,1)\}$$

then applying Theorem 6.6 to

$$dX_t = \mu(t, X_t)dt + \sigma(t, X_t)dB_t; \quad X_{t_0} = \xi,$$

we have that

$$\begin{aligned}
f(t, X_t) &= \sum_{\alpha \in \mathcal{A}} I_{\alpha, t_0, t}(L^\alpha(f)) + \sum_{\alpha \in \mathcal{B}(\mathcal{A})} I_{\alpha, t_0, t}(L^\alpha(f)) \\
&= f\left(t_0, X_{t_0}\right) + \int_{t_0}^{t} L^0 f\left(t_0, X_{t_0}\right) ds + \int_{t_0}^{t} L^1 f\left(t_0, X_{t_0}\right) dB_s \\
&\quad + \int_{t_0}^{t} \int_{t_0}^{s_2} L^0(L^0(f))(s_1, X_{s_1}) ds_1 dB_{s_2} + \int_{t_0}^{t} \int_{t_0}^{s_2} L^0(L^1(f))(s_1, X_{s_1}) ds_1 dB_{s_2} \\
&\quad + \int_{t_0}^{t} \int_{t_0}^{s_2} L^1(L^0(f))(s_1, X_{s_1}) dB_{s_1} ds_2 + \int_{t_0}^{t} \int_{t_0}^{s_2} L^1(L^1(f))(s_1, X_{s_1}) dB_{s_1} dB_{s_2} \\
&= f\left(t_0, X_{t_0}\right) + L^0(f)(t_0, X_{t_0})(t - t_0) + L^1(f)(t_0, X_{t_0})(B_t - B_{t_0}) + R
\end{aligned}$$

where R are the four remaining double integrands. Notice that if we remove R from this identity we obtain the Euler-Maruyama method. This provides motivation for the following result. For each $k \in \mathbb{N}$ we define the set

$$\mathcal{A}_k := \left\{ \alpha \in \mathcal{M}_m | \quad l(\alpha) + n(\alpha) \leq 2k \quad \text{or} \quad l(\alpha) = n(\alpha) = k + \frac{1}{2} \right\},$$

these are the sets we will use for our desired theorem.

Theorem 6.7. *If X_t is the solution of an SDE of the form* (19)*, Y_n^h is the numerical method given by*

$$\begin{cases} Y_0^h := X_{t_0} \\ Y_n^h := \sum_{\alpha \in \mathcal{A}_k} L^\alpha(X)(t_0, X_{t_0}) I_{\alpha, t_{n-1}, t_n} \end{cases} \tag{38}$$

and suppose that $L^\alpha(X)$ verifies the Lipschitz condition (21) *for all $\alpha \in \mathcal{A}_k$. Suppose additionally that $L^\alpha(X)$ verifies the linear growth condition* (22) *and the conditions of the Itô lemma for all $\alpha \in \mathcal{A}_k \cup \mathcal{B}(\mathcal{A}_k)$. Then there exists $K_1, K_2 \in \mathbb{R}$ independent of the step size h such that*

$$\mathbb{E}\left(\max_{t_n \in P^h} \left| X_{t_n} - Y_n^h \right|^2 \mid \mathcal{F}_0 \right) \leq K_1 \left(1 + |X_{t_0}|^2 \right) h^{2k} + K_2 \left| X_{t_0} - Y_0^h \right|^2$$

i.e Y_n has strong order $2k$.

The numerical method Y_n given in the previous theorem is called the strong Itô-Taylor approximation of order k of the Itô process $f(t, X_t)$.

In other words the strong Itô-Taylor approximation of order k converges strongly with order k to the solution. The analogous result for the weak convergence order uses the sets

$$\mathcal{A}'_k := \{ \alpha \in \mathcal{M}_m | \ l(\alpha) < k \}$$

and states the following

Theorem 6.8. *If X_t is the solution of an SDE of the form* (19)*, Y_n^h is the numerical method given by*

$$\begin{cases} Y_0^h := X_{t_0} \\ Y_n^h := \sum_{\alpha \in \mathcal{A}'_k} L^\alpha(X)(t_0, X_{t_0}) I_{\alpha, t_{n-1}, t_n} \end{cases} \tag{39}$$

and suppose that the drift and diffusion coefficients $\mu_i, \sigma_{j,i}$ verify the Lipschitz condition (21)*, the linear growth condition* (22) *and have derivatives of order up to $2(k+1)$ that are continuous and have polynomial growth. Then for each $g : \mathbb{R}^n \to \mathbb{R}$ whose derivatives of order up to $2(k+1)$ are continuous and have polynomial growth there exist $K \in \mathbb{R}^+$ and $r \in \mathbb{N}$ such that*

$$\left| E\left(g\left(X_{t_n} \right) - g\left(Y_n^h \right) \mid \mathcal{F}_0 \right) \right| \leq K \left(1 + |X_0|^{2r} \right) h^{k+1}$$

for all $t_n \in P^h$ i.e Y_n has weak order $k+1$.

6.5 Implicit Methods

Implicit algorithms differentiate themselves from explicit ones by needing to solve an equation in each time step. Despite requiring this extra calculation in some cases, specifically when we are dealing with a stiff equation, this can come accompanied by an improved numerical stability. A family of semi-implicit methods studied in [22] and [11] corresponding to the Euler-Maruyama scheme is

$$\begin{cases} Y(t_0 =)\xi \\ Y_k(t_{i+1}) = Y_k(k + \left[\alpha_k \mu_k \left(t_{i+1}, Y(t_{i+1})\right) + \left(1-\alpha_k\right)\mu_k\right] h \\ \qquad\qquad + \sum_{j=1}^{m} \sigma_{k,j} \Delta B_j(t_i); \quad 1 \le k \le n \end{cases} \tag{40}$$

where $\alpha \in [0,1]^n$. Notably when $\alpha = 0$, $\alpha = \frac{1}{2}$ and $\alpha = 1$ (40) becomes the explicit Euler method, the trapezoidal rule and the fully implicit or backward Euler method respectively. The one-dimensional implicit Milstein scheme is

$$\begin{cases} Y_0 = \xi \\ Y_{n+1} = Y_n + \left[\alpha\mu(t_n, Y_n) + (1-\alpha)\mu(t_{n+1}, Y_{n+1})\right] h \\ \qquad\qquad + \sigma(t_n, Y_n)\Delta B(t_n) + \frac{1}{2}\sigma(t_n, Y_n)\frac{\partial \sigma}{\partial x}(t_n, Y_n)\left(\Delta B_n^2 - \Delta t\right). \end{cases}$$

In the case that our SDE is n-dimensional with an m-dimensional Brownian motion the component number k of the implicit Milstein scheme is

$$\begin{cases} Y_0 = \xi \\ Y_k(t_{i+1}) = Y_k(t_i) + \left[\alpha_k \mu_k \left(t_{i+1}, Y(t_{i+1})\right) + \left(1-\alpha_k\right)\mu_k\right] h \\ \qquad\qquad + \sum_{j=1}^{m} \sigma_{j,k} \Delta B_j(t_i) + \frac{1}{2} \sum_{j_1, j_2 = 1}^{m} L^{j_1} \sigma_{j_2,k} I_{((j_1,j_2),t_i,t_{i+1})}; \quad 1 \le k \le n \end{cases}$$

where the drift and diffusion coefficients are evaluated at $(t_i, Y(t_i))$ unless otherwise specified. Let us take a look at a stiff equation in which the advantages of implicit stochastic methods are manifest. Suppose $\{X_t\}_{t \in T}$ is a one-dimensional process satisfying the equation

$$X_t = \mu X_t + dB_t; \quad t \in T$$

where $\mu \in \mathbb{R}^-$ and $\{B_t\}_{t \in T}$ is a one-dimensional Brownian motion. The numerical solution Y_n with initial condition ξ given by the Euler-Maruyama algorithm is

$$\begin{cases} Y_0 = \xi \\ Y_{n+1} = Y_n^{h,\xi}(1 + \mu h) + \Delta B_n. \end{cases}$$

Thus we have that

$$\left|Y_n^{h,\xi_1} - Y_n^{h,\xi_2}\right| \leq |1+\mu h|^n \, |\xi_1 - \xi_2|$$

For the error to not increase in every subsequent iteration of this algorithm it would then be necessary for

$$|1+\mu h| \leq 1 \iff h \in \left[0, -\frac{2}{\mu}\right].$$

If however we consider the implicit Euler method we have that

$$\begin{cases} Y_0 = \xi \\ Y_{n+1} = Y_{n+1}^{h,\xi}(1+\mu h) + \Delta B_n. \end{cases}$$

so

$$\left|Y_{n+1}^{h,\xi_1} - \bar{Y}_n^{h,\xi_2}\right| \leq |1-\mu h|^{-n} \, |\xi_1 - \xi_2|$$

which means that to preserve the stability of (6.5) we are not restricted on the step size h using the implicit Euler method.

6.6 Runge-Kutta methods

The generalization to SDEs of Runge-Kutta schemes provide us with a familiy of iterative single-step methods that do not rely on the derivatives of the drift and diffusion of the SDE. The drawback that they present is that it is necessary to evaluate the drift and diffusion several times for each time step. Here we present two schemes of orders 1 and 1.5 respectively as schemes of higher order become increasingly complex.

The order 1 scheme is not valid for the more general case in which the Brownian motion is m-dimensional, however we will showcase it as it is similar to a two-stage deterministic Runge-Kutta method. Using the usual notation suppose we have an n-dimensional SDE with respect to a 1-dimensional Brownian motion. Let us consider the approximation

$$\begin{cases} Y_0 = \xi \\ Y_n = \frac{1}{2}(K_1 + K_2) \end{cases} \tag{41}$$

where

$$K_1 := h\mu\left(t_n, Y_n\right) + \left(\Delta B_n - S_n\sqrt{h}\right)\sigma\left(t_n, Y_n\right)$$

$$K_2 := h\mu\left(t_{n+1}, Y_n + K_1\right) + \left(\Delta B_n + S_n\sqrt{h}\right)\sigma\left(t_{n+1}, Y_n + K_1\right)$$

and S_n is a Bernouilli process with $\mathcal{P}(S_n = 1) = \mathcal{P}(S_n = -1) = \frac{1}{2}$ for each n. In the deterministic case, $\sigma = 0$,it is easy to see that this scheme corresponds to Heun's method.

Theorem 6.9. *The numerical scheme* (41) *has strong order* 1.*[21]*

To obtain Runge-Kutta methods of higher order is substantially more difficult than to obtain Taylor approximations, since there is no known general procedure to do so. In [16] Runge-Kutta methods of strong order 1, 1.5 and 2 can be found For example for solving a scalar autonomous SDE of the form

$$dX_t = \mu(X_t)dt + \sigma(X_t)dB_t$$

consider the approximation Y_n

$$\begin{aligned} Y_{n+1} = Y_n &+ \sigma\Delta B_n + \frac{1}{2\sqrt{h}}\left(\mu\left(\bar{\Upsilon}_+\right) - \mu\left(\bar{\Upsilon}_-\right)\right) I_{(1,0),t_n,t_{n+1}} \\ &+ \frac{1}{4}\left(\mu\left(\bar{\Upsilon}_+\right) + 2\mu + \mu\left(\bar{\Upsilon}_-\right)\right) h \\ &+ \frac{1}{4\sqrt{h}}\left(\sigma\left(\bar{\Upsilon}_+\right) - \sigma\left(\bar{\Upsilon}_-\right)\right)\left((\Delta B_n)^2 - h\right) \\ &+ \frac{1}{2h}\left(\sigma\left(\bar{\Upsilon}_+\right) - 2\sigma + \sigma\left(\bar{\Upsilon}_-\right)\right)\left(\Delta B_n h - I_{(1,0),t_n,t_{n+1}}\right) \\ &+ \frac{1}{4h}\left(\sigma\left(\bar{\Phi}_+\right) - \sigma\left(\bar{\Phi}_-\right) - \sigma\left(\bar{\Upsilon}_+\right) + \sigma\left(\bar{\Upsilon}_-\right)\right)\left(\frac{1}{3}(\Delta B_n)^2 - h\right)\Delta B_n \end{aligned} \tag{42}$$

where

$$\bar{\Upsilon}_\pm = Y_n + \mu h \pm \sigma\sqrt{h}; \quad \bar{\Phi}_\pm = \bar{\Upsilon}_+ \pm \sigma\left(\bar{\Upsilon}_+\right)\sqrt{h}.$$

and μ, σ are evaluated on (t_n, Y_n) unless otherwise specified.

Theorem 6.10. *[16] The numerical scheme given by* (42) *has strong order* $\frac{3}{2}$ *for autonomous SDEs .*

6.7 Predictor-Corrector Methods

This class of methods aims to improve on single step methods by initially calculating a "predictor" $\overline{Y}_n$ which is then used to calculate the "corrector" Y_n. In general predictor-corrector numerical schemes have better numerical stability than the single step explicit methods without needing to solve an equation for each step of the algorithm. In the scalar case the family of predictor-corrector

schemes corresponding to the semi-implicit Euler-Maruyama schemes takes the form

$$\begin{cases} Y_0 = \xi; \\ Y_{n+1} = Y_n + \left\{\alpha\bar{\mu}_\beta\left(t_{n+1}, \bar{Y}_{n+1}\right) + (1-\alpha)\bar{\mu}_\beta\left(t_n, Y_n\right)\right\} h \\ \qquad\qquad + \left\{\beta\sigma\left(t_{n+1}, \bar{Y}_{n+1}\right) + (1-\beta)\sigma\left(t_n, Y_n\right)\right\} \Delta B_n \end{cases}$$

where

$$\bar{Y}_{n+1} = Y_n + \mu(t_n, Y_n)h + \sigma(t_n, Y_n)\Delta B_n; \quad \bar{\mu}_\beta = \mu - \beta\sigma\frac{\partial\sigma}{\partial x}.$$

and $\alpha, \beta \in [0,1]$. So essentially we take the Euler method as a predictor and then take a weighted sum of the diffusion and a slightly modified drift coefficient. In the most general case where the solution of the SDE is an n-dimensional Itô process driven by an $m-$dimensional Brownian motion this scheme can be written as

$$\begin{cases} Y_0 = \xi; \\ Y_i(t_{k+1}) = Y_i(t_k) + \left\{\alpha\bar{\mu}_{i,\beta}\left(t_{n+1}, \bar{Y}_{n+1}\right) + (1-\alpha)\bar{\mu}_{i,\beta}\left(t_n, Y_n\right)\right\} h \\ \qquad\qquad + \sum_{j=1}^{m}\left(\beta\sigma_{j,i}\left(t_{n+1}, \bar{Y}_{n+1}\right) + (1-\beta)\sigma_{j,i}\left(t_n, Y_n\right)\right)\Delta B_i(t_n) \end{cases}$$

where $\bar{Y}_{n+1}$ is given by the Euler method starting at $(t_n, Y(t_n))$,

$$\bar{\mu}_{i,\beta} = \mu_i - \beta\sum_{j=1}^{m}\sigma_{j,i}\frac{\partial\sigma_j}{\partial x_i}$$

and $\alpha, \beta \in [0,1]$. Additionally it can be shown that this family of methods has weak order 1 [16].

7 SDEs in Finance

A financial market is a marketplace in which the trade of financial assets takes place. These financial assets are liquid, in the sense that they can be converted quickly into cash, and derive their value from a contractual claim. There are many types of financial assets, e.g., bank deposits, stocks, bonds and derivatives. Due to the vast amount of capital that continuously changes hands in financial markets such as the New York and Shanghai Stock Exchanges and

the growing interest and participation of the population in financial markets, several fields have arisen with the goal of studying these markets. The field of mathematical finance aims to do so by attempting to model the evolution of a financial market in a mathematically rigorous way. In 1973 the economists Fischer-Black and Myron Scholes wrote an article in which, using stochastic calculus, they derived a partial differential equation to model the price of European options in a financial market[2]. This paper sparked the interest in mathematical finance and established a method for the fair pricing of options that is still used today.

One of the most common tools to model a market in mathematical finance is the Brownian model. In this model a market X on a probability space $(\Omega, \mathcal{F}, \mathcal{P})$ driven by an m-dimensional Brownian motion $\{B_t\}_{t\in T}$ is an $(n+1)$-dimensional Itô process $X = (X_0, X_1, ..., X_n)$ where

$$\begin{cases} X_0(t) = 1 \\ dX_0(t,\omega) = \mu_0(t,\omega)X_0(t,\omega)dt; \quad t \in T \end{cases}$$

and

$$\begin{cases} X_i(t) = x_i \\ dX_i(t,\omega) = \mu_i(t,\omega)dt + \sum_{j=1}^{m} \sigma_j(t,\omega)dB_j(t,\omega); \quad 1 \leq i \leq n \quad t \in T. \end{cases}$$

Each component of the market X is the price of a financial asset. The underlying asset of price X_0 is called a riskless asset with riskless rate μ_0 because its price is not driven by a Brownian motion. The rest of the assets are called risky assets. An important feature of this model is the assumption that financial assets have prices that change continuously and are driven by a Wiener process.

A portfolio or strategy $\{P_t\}_{t\in T}$ is an $(n+1)$-dimensional process adapted to the filtration generated by $\{B_t\}_{t\in T}$. The i-th component of the portfolio P represents the share of the asset i. Thus, its value $V^P(t)$ at time $t \in T$ is given by the formula

$$V^P(t) := \sum_{i=0}^{n} P_i(t)X_i(t).$$

The portfolio P is said to be self-financing if

$$dV^P(t) = \sum_{i=0}^{n} P_i(t) dX_i(t).$$

This condition makes the continuous model to be consistent with discrete ones, where a change in price $\Delta X_i(t_k)$ of the asset i would correspond to an increase of

$$P_i(t_k)\Delta X_i(t_k)$$

in the value of the portfolio. A portfolio $\{P_t\}_{t\in T}$ is said to be admissible if it is self-financing and there exists $K \in \mathbb{R}$ such that

$$\mathcal{P}(V^P(t) > K) = 1; \quad \text{for all } t \in T.$$

The condition of admissibility means that the value of the portfolio is almost surely lower bounded, which rules out strategies where one invests without limit.

An arbitrage is an admissible portfolio $\{P_t\}_{t\in T}$ such that

$$V^P(t_0) = 0; \quad V^P(t_f) \geq 0; \quad \mathcal{P}(V^P(t_f) > 0) = 0.$$

Subsequently an arbitrage is a strategy that guarantees that the investor will never lose money and will occasionally turn a profit.

7.1 The Black-Scholes Model

Let's take a look at the most well known contribution of mathematical finance. The Black-Scholes model consists of a market $X = (X_0, X_1)$ formed by a riskless asset of price X_0 and a risky asset whose price is X_1. These prices follow the equations

$$\begin{aligned} dX_0(t) &= rX_0 dt \\ dX_1(t) &= \mu dt + \sigma dB_t \end{aligned}$$

where r is the rate of the riskless asset and $\mu, \sigma \in \mathbb{R}$ correspond respectively to the expected rate of return and the volatility of $X_1(t)$. The price of the risky asset $X_1(t)$ is an Itô process called geometric Brownian motion. In Figure 1 10 trajectories of the risky asset are shown for $\mu = \sigma = 0.1$. The market X is supposed to present no arbitrage and have no cost associated to transactions.

The goal of the Black-Scholes model is to give an accurate price for the European call option associated with the asset of price X_1. This call option is a contract that gives the buyer the option of buying the asset of price X_1 from the seller at a strike price of K and at the time of maturity t_m. This option then pays the buyer

$$(X_1(t_m) - K)_+ := \max(X_1(t_m) - K, 0).$$

What Black and Scholes proved, using the Itô formula and the assumption that there can exist no arbitrages in the market, was that the price $C(t, X_1(t))$ of the option must verify the PDE

$$\left(\frac{\partial}{\partial t} + \frac{\sigma^2 x^2}{2}\frac{\partial^2}{\partial x^2} + rx\frac{\partial}{\partial x} - r\right) C(t, x) = 0.$$

Solving this equation one arrives at the Black-Scholes formula

$$C\left(t, X_1(t)\right) = X_1(t)N\left(d_1(t)\right) - Ke^{-r(t_m - t)}N\left(d_2(t)\right),$$

where N is the CDF of a Gaussian $\mathcal{N}(0, 1)$ distribution and

$$\begin{aligned} d_1(t) &:= \frac{\log\left(X_1(t)/K\right) + \left(r + \frac{1}{2}\sigma^2\right)(t_m - t)}{\sigma\sqrt{t_m - t}} \\ d_2(t)_: &= d_1 - \sigma\sqrt{t_m - t}. \end{aligned}$$

7.2 The CIR Model

One of the most important models in mathematical finance is the Cox-Ingersoll-Ross model (CIR)[6]. In the Black Scholes model the riskless asset has a constant rate of return r. However, in practice this interest rate may vary in time. The CIR model, which seeks to explain the evolution of this interest rate, is the SDE

$$dX_t = s(m - X_t)dt + v\sqrt{X_t}dB_t, \tag{43}$$

where

1. $\{X_t\}_{t\in[0,\infty)}$ is a one-dimensional stochastic process representing the interest rate.

2. $\{B_t\}_{t\in[0,\infty)}$ is a one-dimensional Brownian motion starting at zero.

3. $m, s, v \in [0, \infty)$ are respectively the mean, the speed of reversion to the mean and the financial volatility.

The first aspect we should analyze of the stochastic differential equation (43) is is the existence and uniqueness of solutions, due to it is not obvious: notice that in (43) the drift and diffusion coefficients are respectively

$$\mu(t,x) = \mu(x) = s(m-x)$$
$$\sigma(t,x) = \sigma(x) = v\sqrt{x};$$

then μ and σ verify the Lipschitz condition (21) which guarantees the uniqueness of solutions, but the linear growth condition (22) is not verified for σ and Theorem 4.5 can not be applied. Fortunately, in the one-dimensional case the theorem of existence and uniqueness has been refined to include a larger class of SDEs, as the following theorem, proved in [25], shows.

Theorem 7.1. *Given a one dimensional SDE*

$$\begin{cases} X_0 = \xi \\ dX_t = \mu\left(t, X_t\right) dt + \sigma\left(t, X_t\right) dB_t \end{cases} \tag{44}$$

if there exists $K \in \mathbb{R}^+$ such that μ and σ verify the inequalities

$$|\mu(t,x) - \mu(t,y)| \leq K|x-y| \quad \textit{for every } t \in \mathbb{R}^+ \textit{ and for all } x, y \in \mathbb{R}$$
$$|\sigma(t,x) - \sigma(t,y)| \leq h(|x-y|) \quad \textit{for every } t \in \mathbb{R}^+ \textit{ and for all } x, y \in \mathbb{R}$$

where and $h : \mathbb{R}^+ \to \mathbb{R}^+$ is a strictly increasing function such that $h(0) = 0$ and

$$\int_0^\epsilon \frac{1}{h^2(x)} dx = \infty; \quad \forall \epsilon > 0,$$

then the equation (44) has a unique strong solution.

Applying Theorem 7.1 to the SDE (43) with $h(x) := \sqrt{x}$ we obtain the existence of a unique strong equation to the CIR model. From now on we will refer to this solution as the CIR process and will denote it as Y_t^ξ or simply as Y_t. As a matter of fact, the cumulative distribution function (CDF) of the solution of (43) is known:

Proposition 7.2. *The cumulative distribution function of Y_t given Y_t' with $t' < t$ is given by*

$$\mathcal{P}(Y_t < x \mid Y_t') = F_{\chi^2}\left(f_1(t', t, x); d, f_2(t', t)\right), \tag{45}$$

where

$$f_1(t', t, x) := \frac{4sx}{v^2\left(1 - e^{-s(t-t')}\right)}; \quad d := \frac{4sm}{v^2}; \quad f_2(t', t) := Y_{t'} \frac{4se^{-s(t-t')}}{v^2\left(1 - e^{-s(t-t')}\right)},$$

and where $F_{\chi^2}(x; d, \lambda)$ is the CDF of the non-central chi-squared distribution with d degrees of freedom and non-centrality parameter λ.

Using this CDF it follows that

$$\mathbb{E}[Y_t \mid Y_{t'}] = m + (Y_{t'} - m)e^{-s(t-t')} \tag{46}$$

$$\mathrm{Var}[Y_t \mid Y_{t'}] = \frac{Y_{t'}v^2e^{-s(t-t')}}{s}\left(1 - e^{-s(t-t')}\right) + \frac{mv^2}{2s}\left(1 - e^{-s(t-t')}\right)^2. \tag{47}$$

The expression (46) gives an important property of the CIR process: its mean reversion. The expected value of Y_t converges to m as time goes to infinity and the velocity increases with the speed of reversion s. On the other hand, (47) tells us that that the variance of the CIR process increases with the volatility parameter and decreases with its speed of reversion to the mean. Another important property of the CIR model, derived from (45), is the so called *Feller condition*:

Proposition 7.3. *[Feller Condition][9] Suppose that $\xi > 0$ and $2sm \geq v^2$, then the CIR process Y_t^ξ with initial condition ξ verifies*

$$Y_t^\xi > 0 \quad \textit{for all } t \in \mathbb{R}^+. \tag{48}$$

So, essentially, if the coefficients of (43) verify a certain bound then the CIR process will always remain positive. However, in practice the parameters associated to a financial market seldom verify this inequality. This means that the boundary value zero is accessible in most cases.

7.3 Stochastic Volatility Models

Stochastic volatility models are financial models used for options pricing in which the underlying asset's volatility is an Itô process. In the Black-Scholes model the price of the asset follows the equation

$$dX_t = \mu_t dt + \sigma dB_t,$$

where the volatility σ was constant. The general stochastic volatility model is of the form

$$\begin{aligned} dX_t &= \mu_t X_t dt + \sigma(t, Y_t)X_t dB_1(t) \\ dY_t &= \overline{\mu}(t, Y_t)dt + \overline{\sigma}(t, Y_t)dB_2(t). \end{aligned}$$

where B_1 and B_2 are scalar Brownian motions with correlation $c \in (-1, 1)$.

The most well know stochastic volatility model is the Heston model, developed by Steven Heston [10]. The Heston model of parameter $c \in (-1, 1)$ is

$$\begin{aligned} dX_t &= \mu X_t dt + \sqrt{Y_t} X_t dB_1(t) \\ dY_t &= s(m - Y_t)dt + v\sqrt{X_t} dB_2(t) \end{aligned}$$

where $\mu \in \mathbb{R}$, $m, s, v \in [0, \infty)$ and B_1, B_2 are stndard scalar Brownian motions with correlation c. So in the Heston Model the volatility of the asset of price X_t is given by the square root of the CIR process. This model can also be written as

$$\begin{aligned} d\ln(X_t) &= (\mu - \frac{1}{2})Y_t dt + \sqrt{Y_t} dB_1(t) \\ dY_t &= s(m - Y_t)dt + v\sqrt{Y_t} dB_2(t) \end{aligned}$$

by applying Itô formula to $\ln(X_t)$. Using this form of the equation, Heston [10] calculated the characteristic function of $\ln(X_t)$. This characteristic function can be used to prove the following

Proposition 7.4. *[1] The call option with strike price K and time of maturity t_m has expected value given by the formula*

$$\begin{aligned} &\mathbb{E}\left((X(t_m) - K)^+\right) \\ &= X(t_0) - \frac{K}{2\pi} \int_{-\infty}^{\infty} \frac{\exp\left\{(1/2 - ix)\ln(X(t_0)/K) + h_1 - (x^2 + 1/4)\, h_2\right\} V(t_0)}{x^2 + 1/4} dx \end{aligned}$$

where $i = \sqrt{-1}$ and

$$h_1 = -\frac{sm}{v^2}\left(\partial_+ t_m + 2\ln\left(\frac{\partial_- + \partial_+ e^{-pt_m}}{2p}\right)\right), \quad h_2 = \frac{1 - e^{-pt_m}}{\partial_- + \partial_+ e^{-pt_m}}, \quad \hat{s} = s - cv/2$$
$$\partial_\pm = p \mp (ixcv + \hat{s}), \quad p = \sqrt{x^2 v^2 (1 - c^2) + 2ixvc\hat{s} + \hat{s}^2 + v^2/4}.$$

8 Simulation of the CIR Model

The CIR model is quite interesting numerically speaking. This is because if we apply a naive method to obtain a numerical approximation Y_n on a finite interval T, for example, the Euler-Maruyama method with initial condition ξ.

$$\begin{cases} Y_0 = \xi \\ Y_n = Y_{n-1} + s(m - Y_{n-1})h + v\sqrt{Y_{n-1}}\Delta B_n; \quad 1 \leqslant n \leqslant |P^h| \end{cases}$$

we may encounter significant problems, particularly when Feller inequality (see Proposition 7.3) is not verified and $m \sim 0$. These obstacles are encountered when the stochastic process approaches zero as small imprecisions can make the numerical approximation reach negative values and hence have no real square root. There are a few relatively simple ways of eliminating these problems, at a cost. The first modification consists of replacing $\sqrt{Y_{n-1}}$ by $\sqrt{|Y_{n-1}|}$; so our algorithm becomes

$$\begin{cases} Y_0 = \xi \\ Y_n = Y_{n-1} + s(m - Y_{n-1})h + v\sqrt{|Y_{n-1}|}\Delta B_n; \quad 1 \leqslant n \leqslant |P^h|. \end{cases} \tag{49}$$

The second modification is similar, replacing now $\sqrt{Y_{n-1}}$ by $\sqrt{[Y_{n-1}]_+}$, where $[Y_{n-1}]_+ = \max\{Y_{n-1}, 0\}$; the method is written

$$\begin{cases} Y_0 = \xi \\ Y_n = Y_{n-1} + s(m - Y_{n-1})h + v\sqrt{[Y_{n-1}]_+}\Delta B_n; \quad 1 \leqslant n \leqslant |P^h|. \end{cases} \tag{50}$$

Lastly, in this vein, instead of forcing the diffusion term to be zero, we can apply this same principle to avoid the approximation from ever being negative:

$$\begin{cases} Y_0 = \xi \\ Y_n = \left[Y_{n-1} + s(m - Y_{n-1})h + v\sqrt{Y_{n-1}}\Delta B_n\right]_+; \quad 1 \leqslant n \leqslant |P^h|. \end{cases} \tag{51}$$

We will refer to the numerical schemes (49)-(51) as the Euler absolute value method, the Euler square root truncation method and the Euller full truncation method respectively.

Another way of dealing with the problems posed by discretizing the time interval, proposed in [3], is to consider an exact simulation using the cumulative function distribution (45). The issue that this method has is that it can be inefficient. This is because if we wish to sample the exact solution knowing its CDF it is necessary to perform inverse transform sampling, a process which relies on using the inverse of the CDF given in (45), for which there is no closed form solution. In addition, often it will be desirable to simulate the CIR model along a large amount of times; so the step size h will be very small. But when h converges to zero the degrees of freedom of the chi-squared distribution that the CIR process's CDF follows diverge to infinity, as Proposition 7.2 establishes, which also results problematic.

In [14] the authors propose to use the scheme

$$\begin{cases} Y_0 = \xi \\ Y_n = \frac{Y_{n-1}+smh+v\sqrt{Y_{n-1}}h\Delta B_{n-1}+\frac{1}{4}v^2(\Delta B_{n-1}^2-h)}{1+sH} \end{cases}$$

with the advantage over the naive method that it cannot become negative when $4ms > v^2$. Unfortunately this bound, like the Feller one, is seldom verified in practical applications and thus is of limited use.

What we suggest is generalizing these methods of avoiding negative square roots to the Milstein scheme. The Milstein scheme is particularly suited to the task because its added term with respect to the Euler scheme is

$$\frac{1}{2}\sigma(t,x)\frac{\partial\sigma(t,x)}{\partial x}(\Delta B_n^2 - h)$$

which in the case of the CIR model (43) is equal to

$$\frac{v^2}{4}(\Delta B_n^2 - h),$$

This means that the Milstein scheme does not necessitate any simplification of its added term in order eliminate negative square roots unlike what would occur in most other numerical methods. The Milstein versions of methods (49)-(51) are

$$\begin{cases} Y_0 = \xi \\ Y_n = Y_{n-1} + s(m - Y_{n-1})h + v\sqrt{|Y_{n-1}|}\Delta B_n + \frac{v^2}{4}(\Delta B_n^2 - h); \quad 1 \leqslant n \leqslant |P^h|, \end{cases} \tag{52}$$

$$\begin{cases} Y_0 = \xi \\ Y_n = Y_{n-1} + s(m - Y_{n-1})h + v\sqrt{[Y_{n-1}]_+}\Delta B_n + \frac{v^2}{4}(\Delta B_n^2 - h); \quad 1 \leqslant n \leqslant |P^h| \end{cases} \tag{53}$$

and

$$\begin{cases} Y_0 = \xi \\ Y_n = \left[Y_{n-1} + s(m - Y_{n-1})h + v\sqrt{Y_{n-1}}\Delta B_n + \frac{v^2}{4}(\Delta B_n^2 - h)\right]_+; \quad 1 \leqslant n \leqslant |P^h|. \end{cases} \tag{54}$$

In the following we will compare the numerical schemes (49)-(54) to try to determine which one of them is the most suited to the CIR model. The simulation we will present were carried out in Wolfram Mathematica 12.1 for which the code can be found in the Appendix B. The parameters m, s, v were chosen to be in line with those calculated by maximum likelihood estimation in [18], although we will experiment with different ones as well to test the flexibility of our conclusions. For each set of parameters 10000 simulations were carried out for each numerical method. To start off our parameters are

$$s = 0.44; \quad m = 0.06; \quad v = 0.32,$$

the time interval is $T = [0, 5]$, the step size is $h = 2^{-5}$ and the starting point is $\xi = m$. First of all we will carry out a detailed analysis of the Euler schemes (49)-(51). In Figure 4 we represent the mean of each numerical method's 10000

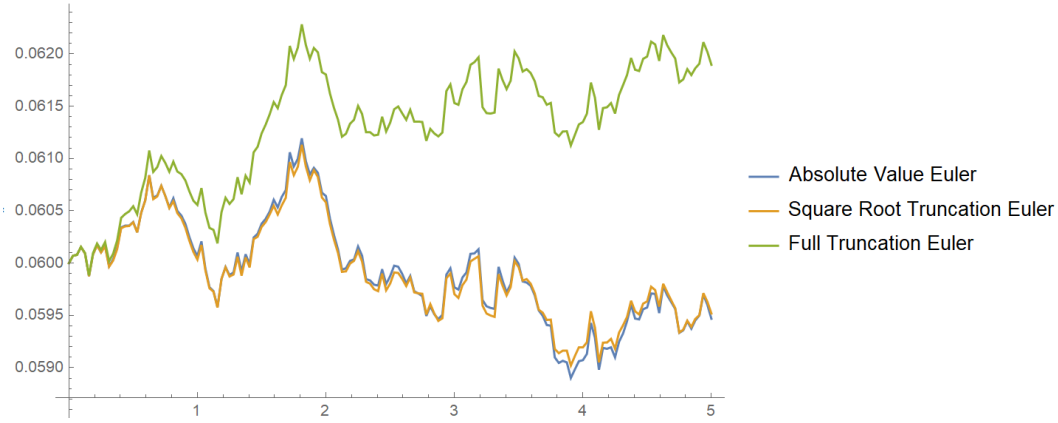


Figure 4: Comparison of the three modified Euler methods.

trajectories. As we can see the means of methods (49) and (50) are quite similar, however, method (51) is notably larger than the other two. Figure 5, which represents the tenth trajectory of the methods (49) and (51), serves to explain the reason of the differences shown in Figure 4. At the start both trajectories are positive, so they are indistinguishable. Later on, when the Euler method would take a negative value, the full truncation scheme substitutes it by zero whereas the absolute value scheme keeps it and modifies the algorithm by using the square root of its absolute value to calculate the next step. This leads to values in (49) lesser or equal to those in (51). The same is true when comparing (50) to (51).

Let's compare these means with the expected value of the true solution. Since the starting point is equal to m, the CIR process $\{Y_t\}_{t\in T}$ has constant

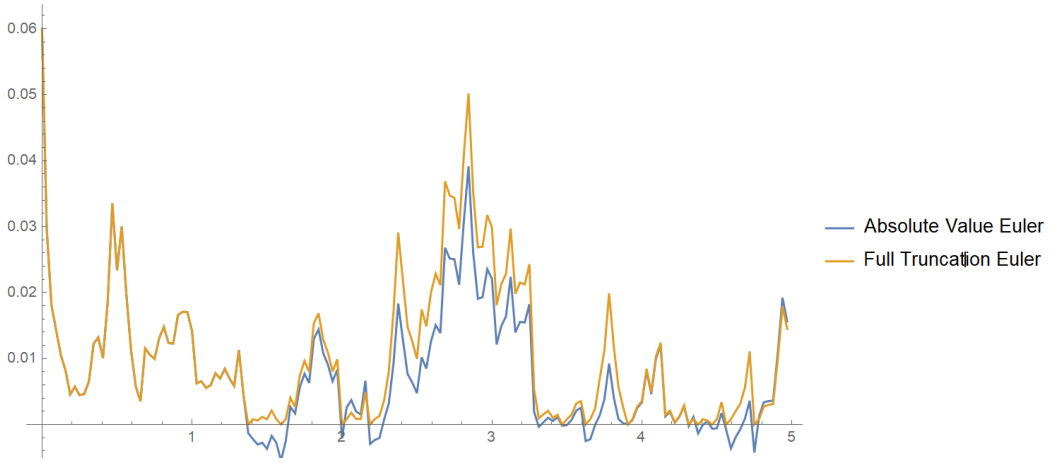


Figure 5: Comparison of a trajectory of the absolute value and full truncation methods.

expected value, see (46),

$$\mathbb{E}[Y_t] = m,$$

We will compare the numerical schemes calculating the mean square error between the numerical means and the expected value of the CIR process at each point of the time discretization. That is, if $\mathbb{E}\left[\bar{Y}_n\right]$ is the mean of the 10^4 simulations of numerical scheme $\bar{Y}_n$ and $\mathbb{E}\left[Y_n\right]$ is the mean of the CIR process then the mean square error between both is

$$\frac{1}{|P^h|}\sqrt{\sum_{k=1}^{|P^h|}(\bar{Y}_k - Y_k)^2}.$$

Computing these distances results in Figure (6) shows that methods (49)

0.0000708747	0.0000722276	0.000111958
Absolute Value Euler	Square Root Truncation Euler	Full Truncation Euler

Figure 6: Mean square error between the expected values of the numerical Euler methods and the CIR process

and (50) have means more similar to the mean of the exact solution than (51).

Another way of contrasting the numerical schemes is to compare their empirical distributions with the distribution of the CIR process, which was given in (45). By empirical distribution F of the data points $D = \{y_1, ..., y_n\} \subset \mathbb{R}$ we mean the step function

$$\begin{aligned} F :&\mathbb{R} \to [0, 1] \\ &x \mapsto F(x) = \frac{|\{y \in D : y \leq x\}|}{|D|}. \end{aligned}$$

In other words, the empirical distribution evaluated at x is the probability that a data point chosen at random (with equal likelihood) is less than or equal to x. So when we say empirical distribution of, for example, method (49) at time t we mean the empirical distribution of the values of its 10000 trajectories at time t. We graphically compare the empirical distributions of the numerical methods and the exact distribution at time $t = 5$ in the following image

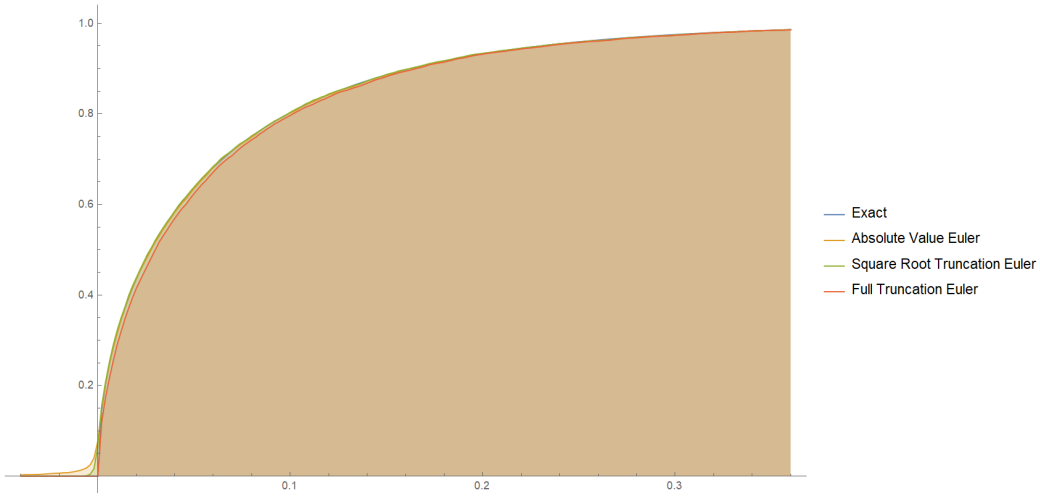


Figure 7: Comparison of the exact and empirical CDFs at time $t = 5$.

It is difficult to come to any conclusion regarding Figure 7 as all the distributions are very similar. We can affirm at first glance that the empirical CDF of method (51) is below the others. However this could be compensated by the fact that, in contrast to methods (49) and (50), method (51) can't take negative values, as is the case with the exact solution. To determine the methods with empirical distributions most similar to the exact distribution at the instant t we calculate the distance between them, or error, as follows. Suppose

that $\bar{F}_t$ is an empirical distribution of the type obtained from our simulations at time t and F_t is the exact distribution at time t. We define the distance between $\bar{F}_t$ and F_t as

$$d(F_t, \bar{F}_t) := ||F_t - \bar{F}_t||_2.$$

It is worth noting that $F_t - \bar{F}_t$ is in fact square integrable for all $t \geq 0$ because the data points empirical distribution is bounded and due to the convergence speed of $F_t - 1$ to 0 as $t \to \infty$. We will approximate this distance as follows

$$d(F_t, \bar{F}_t) = \sqrt{\int_{-\infty}^{\infty} (F_t(x) - \bar{F}_t(x))^2 dx} \sim \sqrt{\int_a^b (F_t(x) - \bar{F}_t(x))^2 dx}$$

where a and b are such that

$$F_t(a) = \bar{F}_t(a) = 0; \quad F_t(b) \geq 1 - 10^{-5}; \quad \bar{F}_t(b) = 1.$$

So essentially we choose the integration limits so that the data points are contained within the interval $(a, b]$ and the exact solution is almost certainly contained in the same interval. Finally we compute this second integral numerically using the trapezoid rule with uniform partition size of 0.02. Doing so results in the following Although all three errors are low it is worth noting

Absolute Value Euler	Square Root Truncation Euler	Full Truncation Euler
0.000196276	0.000140557	0.000246356

Figure 8: The distances between the empirical Euler and exact distribution at time $t = 5$.

method (50) outperforms the rest. To confirm this result we take the mean of the distances for times $t \in \{1, 2, 3, 4, 5\}$. This result is found in Figure 9 We

Absolute Value Euler	Square Root Truncation Euler	Full Truncation Euler
0.00018287	0.000134516	0.000189363

Figure 9: The average distance between the empirical Euler and exact distribution.

have that methods (49) and (51) perform roughly equivalently while (50) is about 30% smaller than the rest.

Now we shall analyze our proposed Milstein schemes (52), (53) and (54). We will proceed identically to in the case of the Euler schemes. In Figure 10 we graph the mean of 10000 trajectories of each of the modified Milstein schemes. It is notable that in this case, unlike the case of the the Euler

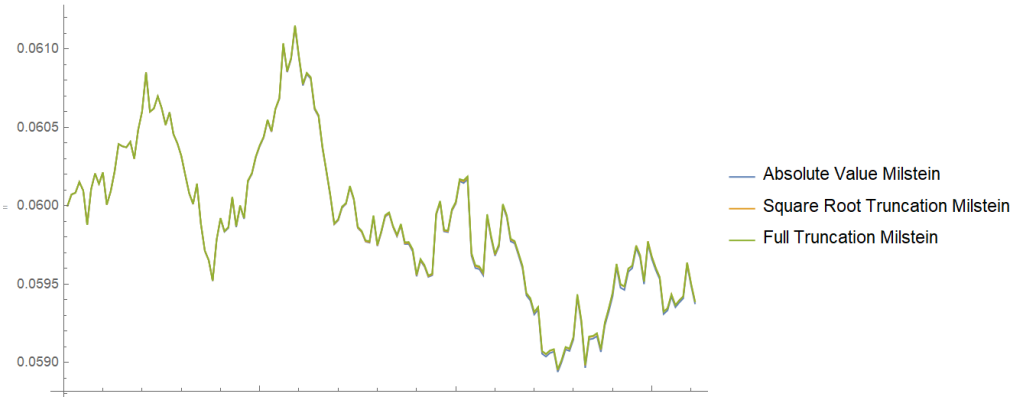


Figure 10: Comparison of the three modified Euler methods.

methods, all three modified Milstein methods have practically identical means. The mean square error between the means of all the numerical schemes and the expected value of the exact solution are shown in Figure 11. So the Milstein

0.0000708747	0.0000722276	0.000111958	0.0000620109	0.0000619906	0.0000618042
Absolute Value Euler	Square Root Truncation Euler	Full Truncation Euler	Absolute Value Milstein	Square Root Truncation Milstein	Full Truncation Milstein

Figure 11: Distance between the means of the numerical methods and the CIR process

methods we proposed have means more similar to that of the CIR process, although still quite similar to the means given by the Euler methods. Now let's compare their empirical distributions. We don't doing so graphically as they are indistinguishable, just like with the Euler schemes. Computing the distance of the empirical distributions to the distribution of the exact solution at time $t = 5$ we obtain In Figure 12 it is clear that our Milstein methods have done substantially better than the others. Let's take a look at the average between these distances at times $t \in \{1, 2, 3, 4, 5\}$. The difference still remains: with these parameters the square root truncation method is the best of the modified Euler methods, while all the Milstein methods perform comparably with an error almost 2 times smaller than the best Euler method. However,

0.000196276	0.000140557	0.000246356	0.0000752976	0.0000738652	0.0000737623
Absolute Value Euler	Square Root Truncation Euler	Full Truncation Euler	Absolute Value Milstein	Square Root Truncation Milstein	Full Truncation Milstein

Figure 12: The distances between the empirical and exact distribution at time $t = 5$.

0.00018287	0.000134516	0.000189363	0.0000768092	0.0000759761	0.00007537
Absolute Value Euler	Square Root Truncation Euler	Full Truncation Euler	Absolute Value Milstein	Square Root Truncation Milstein	Full Truncation Milstein

Figure 13: The average distance between the empirical distributions and the exact distribution.

since it would be unwise to draw any hasty conclusion, in the following we will vary the parameters s, m, v, h and ξ to see if these early results still hold.

In the next test we increase the mean and the speed of reversions to $m = 0.1$ and $s = 0.5$ but this time start slightly below the mean $\xi = 0.9$. The rest of the parameters are unchanged. We now proceed as we did previously. The means of all the numerical methods and the expected value of the exact solution are shown in Figure 14. Now the trajectories means are increasing as the interest rate is returning to the parameter m. The full truncation methods have means larger than the rest of the methods, which have quite similar means. It is obvious at first sight that the mean of method (51) is quite dissimilar to the mean to the CIR process whereas the other methods have quite similar means. This is corroborated by calculating the mean square error, as we did previously, and which is shown in Figure 15.

Now let's examine the distance between the empirical and exact distributions. In Figure 16 we appreciate a very similar pattern to the one we saw in the first test. The method (50) is still the best of the modified Euler, in this case even comparing favorably to the modified Milstein method (52). However, the other two Milstein scheme's empirical distributions are even closer to the exact distribution than (50).

In the next test we use as parameters $s = 0.44$, $m = 0.2$, $v = 0.5$, $\xi = 0.24$, $T = [0, 10]$, $h = 2^{-4}$ so we've essentially made the time interval larger, decreased the step size and increased by quite a bit the volatility and the mean. Once more, the mean of method (51) is clearly above the rest. However, this time the mean of (54) is more in line with the rest, which are, as we've seen previously, very similar. The mean square error between the numerical and the exact expected value are given in Figure 18.

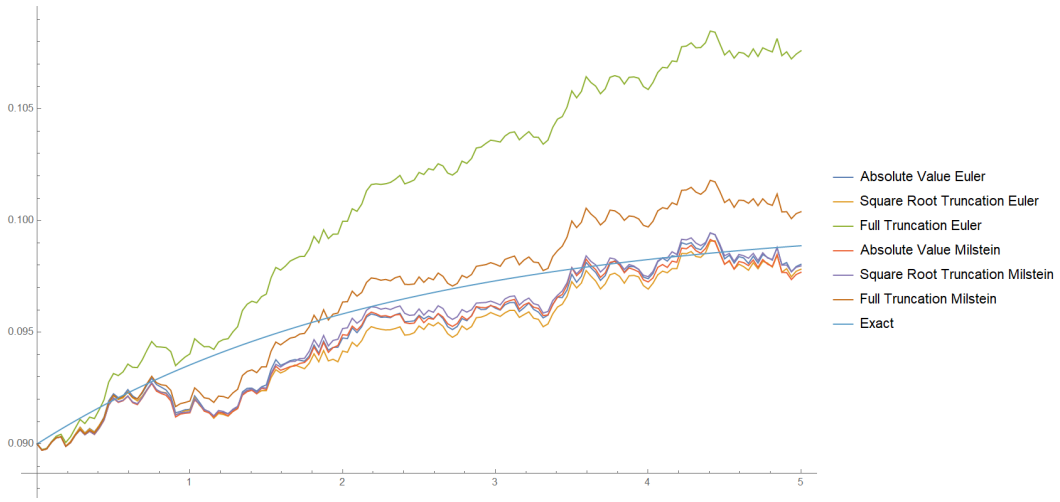


Figure 14: Comparison of the modified numerical methods.

0.000485653	0.000577609	0.00229441	0.000505172	0.000456778	0.000576651
Absolute Value Euler	Square Root Truncation Euler	Full Truncation Euler	Absolute Value Milstein	Square Root Truncation Milstein	Full Truncation Milstein

Figure 15: The mean square error between the expected values of the exact and numerical solutions

0.00536209	0.00302911	0.00559154	0.00322563	0.00262494	0.00230419
Absolute Value Euler	Square Root Truncation Euler	Full Truncation Euler	Absolute Value Milstein	Square Root Truncation Milstein	Full Truncation Milstein

Figure 16: The average distance between the empirical distributions and exact distribution.

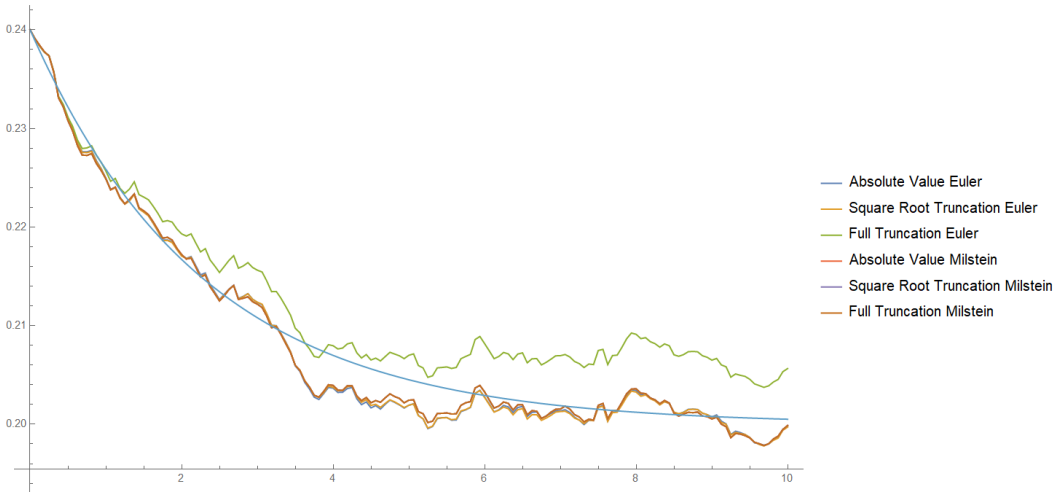


Figure 17: Comparison of the means of the modified numerical methods.

0.000766903	0.000733859	0.00157566	0.000692672	0.000706778	0.000701651
Absolute Value Euler	Square Root Truncation Euler	Full Truncation Euler	Absolute Value Milstein	Square Root Truncation Milstein	Full Truncation Milstein

Figure 18: The mean square error between the numerical and exact means.

The distance between the distributions is given in Figure 19 Figure 19

0.00271423	0.00217714	0.00314236	0.00169866	0.00169872	0.00169871
Absolute Value Euler	Square Root Truncation Euler	Full Truncation Euler	Absolute Value Milstein	Square Root Truncation Milstein	Full Truncation Milstein

Figure 19: The average distance between the empirical distributions and exact distribution.

shows how the empirical distributions of the Milstein methods are nearly twice as close to those of the Euler methods. In addition, once again (50) is the best performing Euler scheme.

There is also a different way to compare these methods: if the parameters m, s, v verify certain relation then we are able to find the Itô process that solves equation (43) and use it to contrast the numerical solutions. The idea is to determine when (43) can be reduced to a linear SDE as in section 5.2. Using (29) we find that the condition for (43) to be reducible is

$$v = 2\sqrt{ms}$$

and using (33) we have that the function $U(t, y)$ that transforms (43) into a

linear equation is

$$U : T \times \mathbb{R}^+ \to \mathbb{R}^+$$
$$(t, y) \mapsto \frac{\exp(\frac{ts}{2})\sqrt{y}}{\sqrt{ms}}. \quad t \in T.$$

Moreover U is continuously differentiable with respect to t and twice continuously differentiable with respect to y so we may apply Itô's lemma on $U(t, Y_t)$, where Y_t is the CIR process with initial condition ξ. Doing so, we obtain

$$dU(t, Y_t) = e^{\frac{ts}{2}} dB_t$$

with exact solution

$$U(t, Y_t) = U(0, \xi) + \int_0^t e^{\frac{ts}{2}} dB_t; \quad t \in T. \tag{55}$$

Applying the inverse transform of U

$$V : T \times \mathbb{R}^+ \to \mathbb{R}^+$$
$$(t, y) \mapsto e^{-ts} y^2 ms.$$

on (55) to recover the solution of (43) we have that

$$Y_t = V(t, U(t, Y_t)) = e^{-ts} ms \left(U(0, \xi) + \int_0^t e^{\frac{ts}{2}} dB_t \right)^2 = e^{-ts} ms \left(\frac{\xi}{\sqrt{ms}} + \int_0^t e^{\frac{ts}{2}} dB_t \right)^2 \tag{56}$$

is the CIR process with initial condition ξ. Approximating

$$Z_t := \int_0^t e^{\frac{ts}{2}} dB_t$$

with the Euler method

$$\begin{cases} Z_0 = 0 \\ Z_n = Z_{n-1} + e^{\frac{t_n s}{2}} \Delta B_n \end{cases}$$

we obtain the numerical solution

$$Y_n = e^{-t_n s} ms \left(\frac{\xi}{\sqrt{ms}} + Z_n \right)^2 \tag{57}$$

of (56). In (57) we don't have the issue we had in (43) where at some point we could have our numerical solution take complex values. Therefore it is logical

to choose the numerical solution (57) as a reference point to contrast the numerical schemes (49)-(54). The comparison we will conduct is the following. Suppose $\bar{Y}_n$ is a numerical approximation of (43), we estimate the absolute error between Y_n and $\bar{Y}_n$ at the end of the time interval t_f as

$$E(\bar{Y}) = \frac{1}{10^4} \sum_{k=1}^{10^4} |\bar{Y}_{t_f,k} - Y_{t_f,k}| \tag{58}$$

where $\bar{Y}_{t_f,k}$ is the simulation number k of $\bar{Y}$ at time t_f.

For our test we choose as parameters

$$s = 0.4363; \quad m = 0.06; \quad v = 2\sqrt{ms}; \quad \xi = m; \quad T = [0, 10]; \quad h = 2^{-5}.$$

In Figure 20 and Figure 21 respectively we juxtapose the means of the Euler and Milstein methods with the mean of (57) Once again the mean of (51)

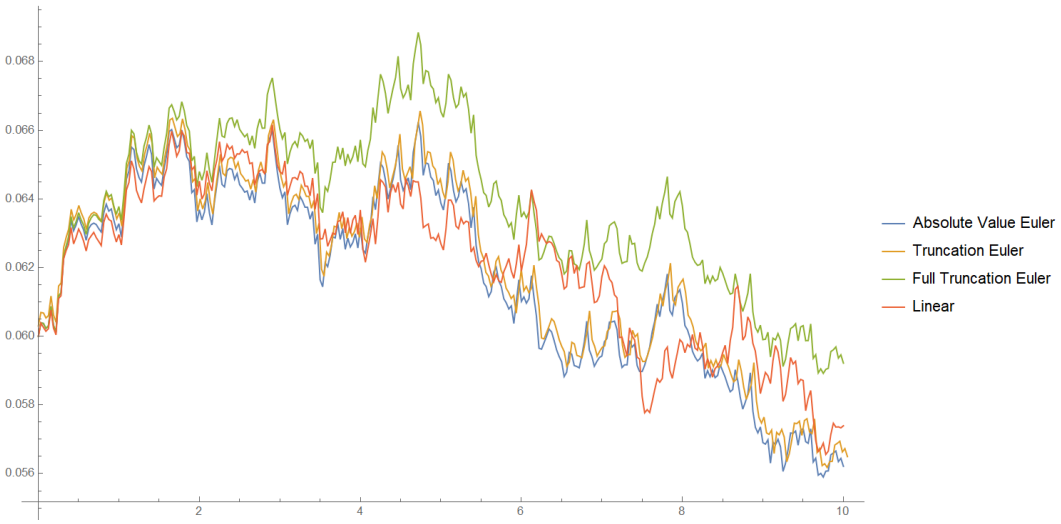


Figure 20: Means of the modified and exact Euler methods

is greater than the rest and the Milstein methods are almost identical. The computation of (58) leads to So, as in previous experiments, the proposed Milstein methods perform the best and the method (51) is the Euler method with the highest error.

As our final test we choose the parameters

$$s = 0.6663; \quad m = 0.02; \quad v = 2\sqrt{ms}; \quad \xi = 1.3m; \quad T = [0, 10]; \quad h = 2^{-5}.$$

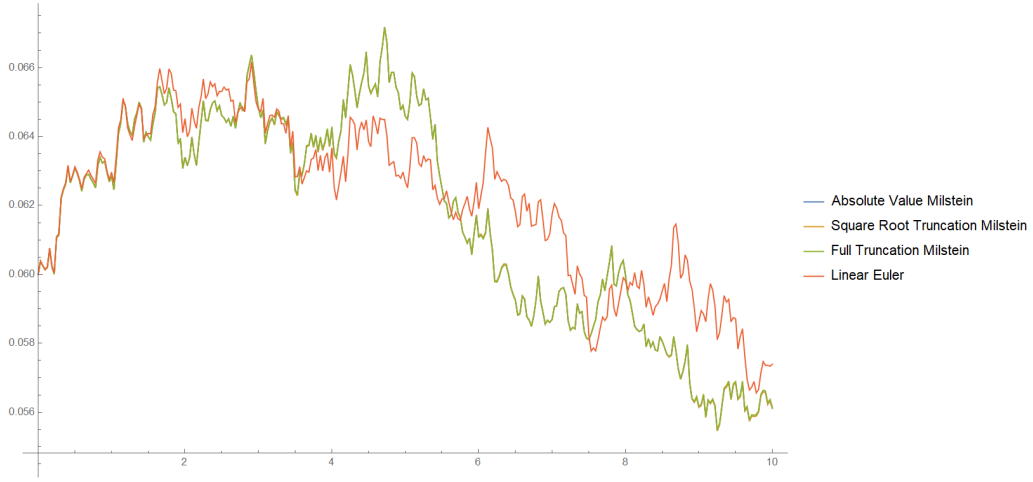


Figure 21: Means of the modified Milstein and exact Euler methods

0.0458548	0.0458548	0.0463213	0.0431984	0.0431984	0.0432173
Absolute Value Euler	Square Root Truncation Euler	Full Truncation Euler	Absolute Value Milstein	Square Root Truncation Milstein	Full Truncation Milstein

Figure 22: The absolute error of each numerical method

Which means we've increased the speed of reversion, chosen the lowest mean yet and started at a value greater than the mean. We proceed as in the previous test. The results are shown in Figures 23-25.

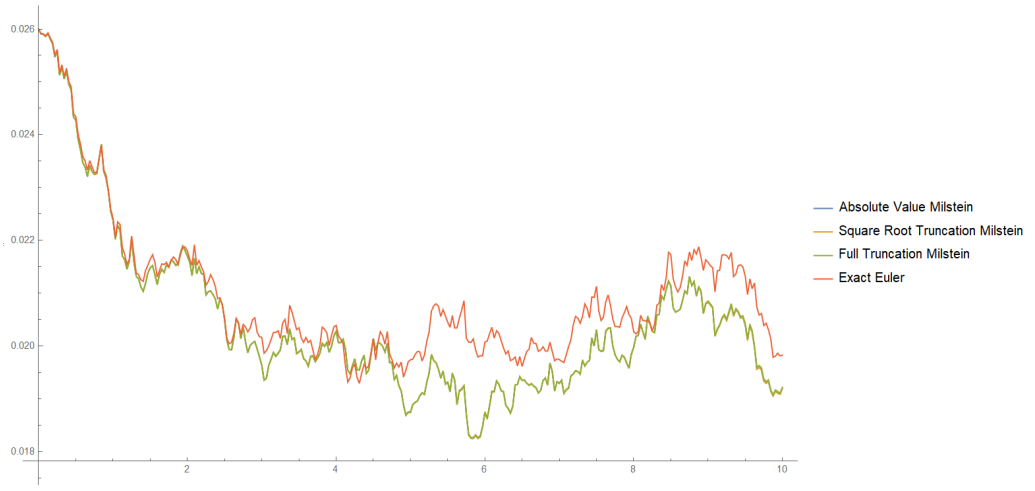


Figure 24: Means of the modified Milstein and exact Euler methods

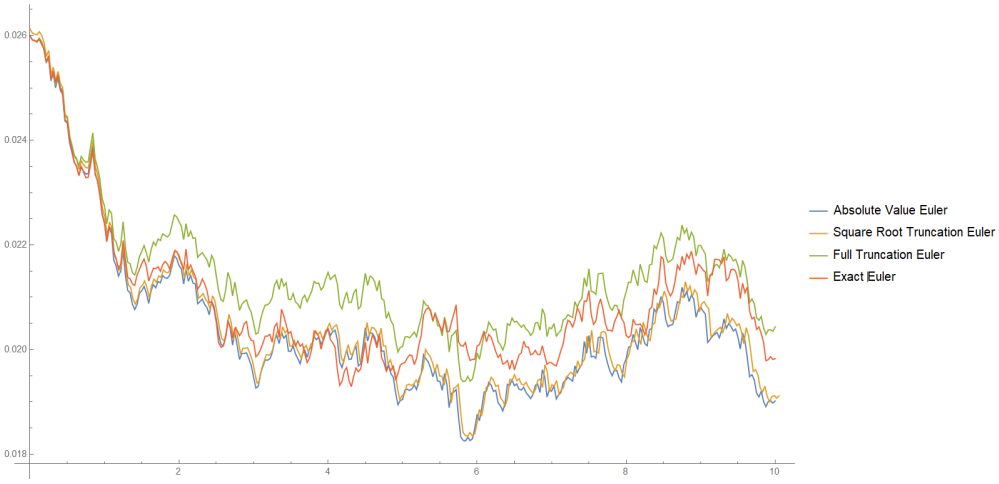


Figure 23: Means of the modified and exact Euler methods

0.0176892	0.0176892	0.0177751	0.0168088	0.0168088	0.0168284
Absolute Value Euler	Square Root Truncation Euler	Full Truncation Euler	Absolute Value Milstein	Square Root Truncation Milstein	Full Truncation Milstein

Figure 25: The absolute error of each numerical method

The results are the same as in the previous experiments: Milstein methods have the lowest error and the method (51) is the Euler method with the highest error.

9 Conclusions

In this paper we have carried out a comparison of three of the most used schemes to simulate the evolution of interest rates given by the CIR model. The tests we carried out showed that the absolute value and square root Euler methods are equally good in the sense that their Euclidean and absolute distance to the CIR process's expected value are extremely similar. The full truncation Euler method performed worse, in this metric, than the other two modified Euler schemes. In terms of the proximity to the CIR process's distribution, the square root truncation Euler method gives the best results out of the Euler methods, followed by the absolute value Euler method which performed better in turn than the full truncation Euler method.

We also proposed our own Milstein schemes, based on these three Euler methods and obtained very satisfactory results. In terms of the distance to the

expected value of the CIR process, the square root and truncation Milstein methods performed better than the Euler ones in all cases, whereas the full truncation Milstein method, while better for the most part than the Euler methods, in one case performed equally well. All three Milstein methods also are superior to their Euler counterparts when it comes to approximating the distribution of the CIR process, although in some cases the square root truncation method may perform equally well.

In conclusion, the square root truncation Euler method is the best Euler method when it comes to simulating the CIR process. In addition, the three proposed Milstein schemes are better choices to simulate the CIR process than the existing Euler methods, although in some cases the square root truncation Euler method may perform just as well.

10 Appendix A

It is not uncommon for the information we have about a stochastic process be limited to its finite dimensional distributions. Kolmogorov's extension theorem gives us conditions under which these finite dimensional distributions do in fact induce a stochastic process.

Lemma 10.1. *Kolmogorov's Extension Theorem for $\mathbb{R}^n$-valued stochastic processes Let $T \subset \mathbb{R}^n$ be a real interval, and let $n \in \mathbb{N}$. For each $k \in \mathbb{N}$ and each finite sequence of distinct times $t_1, \ldots, t_k \in T$ let $\nu_{t_1 \ldots t_k}$ be a probability measure on $(\mathbb{R}^n)^k$ such that the following two conditions are verified:*

1. $$\nu_{t_1 \ldots t_k}(F_1 \times \cdots \times F_k) = \nu_{t_1 \ldots t_k, t_{k+1}, \ldots, t_{k+m}}(F_1 \times \cdots \times F_k \times \underbrace{\mathbb{R}^n \times \cdots \times \mathbb{R}^n}_{m})$$

   for all $m \in \mathbb{N}$.

2. $$\nu_{t_{\sigma(1)} \ldots t_{\sigma(k)}}\left(F_{\sigma(1)} \times \cdots \times F_{\sigma(k)}\right) = \nu_{t_1 \ldots t_k}(F_1 \times \cdots \times F_k)$$

   for all permutations $\sigma \in S_k$.

Then there exists a probability space $(\Omega, \mathcal{F}, P)$ and a stochastic process

$$X : T \times \Omega \to \mathbb{R}^n$$

such that

$$\nu_{t_1 \ldots t_k}(F_1 \times \cdots \times F_k) = \mathbb{P}(X_{t_1} \in F_1, \ldots, X_{t_k} \in F_k)$$

for all $k \in \mathbb{N}$, $(t_1, ..., t_k) \in \mathbb{R}^{nk}$ and all $(F_1, ..., F_k) \in \mathcal{B}(\mathbb{R}^n)^k$

Kolmogorov also established how to prove the continuity of a stochastic process by bounding its moments.

Lemma 10.2. *Kolmogorov's Continuity Theorem Let (S, d) be a complete metric space , and let*
$X\colon [0, +\infty) \times \Omega \mapsto S$ be a stochastic process. If there exist α, β, K such that

$$\mathbb{E}[d(X_s, X_t)^\alpha] \leqslant K|t - s|^{1+\beta}$$

for all $T > 0$, $0 \leqslant s, t \leqslant T$ then there exists a continuous version of X.

Gronwall's integral inequality gives us conditions under which we can bound a function satisfying an integral inequality by the solution of an appropiate integral equation.

Lemma 10.3 (Gronwall's Integral Inequality)**.** *Let I denote an interval of the real line of the form $[a, \infty)$ or $[a, b]$ or $[a, b)$ with $a < b$. Assume $\phi : I \to \mathbb{R}$ is a bounded nonnegative measurable function, $C : I \to \mathbb{R}$ is a non negative integrable function and $B : I \to \mathbb{R}$ is a non-decreasing integrable function with the property that*

$$\phi(t) \leqslant B(t) + \int_{t_1}^{t} C(s)\phi(s)ds \quad \text{for all } t_1 \leqslant t \in I.$$

Then

$$\phi(t) \leqslant B(t) \exp\left(\int_{t_1}^{t} C(s)ds\right) \quad \text{for all } t_1 \leqslant t \in I.$$

Doob's martingale inequalities give us bounds on the likelihood that a martingale will exceed another certain bound and on the expected value of the martingales supremum.

Lemma 10.4 (Doob's Martingale Inequality)**.** *Suppose that $(\Omega, \mathcal{F}, \mathcal{P})$ is a probability space and $T = [t_0, t_f] \subset \mathbb{R}$. Let $X : T \times \Omega \to \mathbb{R}^n$ be a martingale which is almost surely pathwise continuous, then*

$$\mathcal{P}(\sup_{t\in T}|X_t| \geq \lambda) \leqslant \frac{1}{\lambda^p}\mathbb{E}[|X_{t_f}|^p]; \quad \mathbb{E}[\sup_{t\in T}|X_t|^p] \leqslant q^p \sup_{t\in T}\mathbb{E}[|X_t|^p] \leqslant q^p\mathbb{E}[|X_{t_f}|^p]$$

for all $p \geq 1$, $\lambda > 0$ and where $q = \frac{p}{p-1}$.

The Borel-Cantelli lemma is a measure theoretical result that tells us that if the sum of probabilities of a sequence of events is finite then only a finite number of them can occur at once.

Lemma 10.5 (Borel-Cantelli Lemma)**.** *Let $(\Omega, \mathcal{F}, \mathcal{P})$ be a probability space. If $\{A_n\}_{n\in\mathbb{N}} \subset \mathcal{F}^{\mathbb{N}}$ is a sequence of measurable sets such that*

$$\sum_{k\in\mathbb{N}} \mathcal{P}(A_k) < \infty$$

then

$$\mathcal{P}\left(\limsup_{n\to\infty} A_n\right) = \mathcal{P}\left(\bigcap_{n\in\mathbb{N}}\bigcup_{k=n}^{\infty} A_k\right) = 0. \tag{59}$$

11 Appendix B

```
a[t_, x_] := K1 (K2 - x);
b[t_, x_] := K3 Sqrt[x];
```

$$U[t\_, y\_] := \frac{e^{\frac{t K_1}{2}} \sqrt{y}}{\sqrt{K_1 K_2}};$$

```
V[t_, y_] := e^(-t K1) y^2 * K1 * K2;
```

```
V[t, U[t, y]] == y
FullSimplify[U[t, V[t, y]] == y, Assumptions → {K1 > 0, K2 > 0, c > 0, y > 0, t ∈ Reals}]
```

```
α[t_] := Simplify[D[U[t, y], t] + a[t, y] * D[U[t, y], y] +
    1/2 (b[t, y])^2 * D[U[t, y], {y, 2}] /. {K3 → 2 Sqrt[K1 K2]}]
β[t_] := b[t, y] * D[U[t, y], y] /. {K3 → 2 Sqrt[K1 K2]}
```

Parameters

```
tf = 10;
Δ = 1/2^5;
niter = tf / Δ;
tra = 10000;
dist := RandomVariate[NormalDistribution[0, Sqrt[Δ]]];
tabled = Table[dist, {i, 0, tra}, {j, 1, niter}];
```

```
K1 = 0.4363; K2 = 0.06;
K3 = 2 Sqrt[K1 K2];
x0 = K2;
```

Modified and Exact Euler

```
f[t_, x_] := Exp[-t * K1] (Sqrt[x0] + Sqrt[K1 K2] * x)^2;
```

```
In[•]:= trajectoriesE0 = {};
        trajectoriesE1 = {};
        trajectoriesE2 = {};
        trajectoriesE3 = {};
        For[j = 1, j ≤ tra, j++,
        valuesE0 = {{0, x0}};
         valuesE1 = {{0, x0}};
         valuesE2 = {{0, x0}};
         valuesE3 = {{0, x0}};
         xe = 0;
         x1 = x0;
         x2 = x0;
         x3 = x0;
        For[i = 1, i ≤ niter, i++,
          xe = (xe + Exp[Δ * i * K1/2] tabled[[j, i]]);
          x1 = x1 + K1 (K2 - x1) * Δ + K3 Sqrt[Abs[x1]] * tabled[[j, i]];
          x2 = x2 + K1 (K2 - x2) * Δ + K3 Sqrt[Max [Abs[x2], 0]] * tabled[[j, i]];
          x3 = Max[0, x3 + K1 (K2 - x3) * Δ + K3 Sqrt[x3] * tabled[[j, i]]];
          AppendTo[valuesE0, {i * Δ, f[i * Δ, xe]}];
          AppendTo[valuesE1, {i * Δ, x1}];
          AppendTo[valuesE2, {i * Δ, x2}];
          AppendTo[valuesE3, {i * Δ, x3}]
         ];
         AppendTo[trajectoriesE0, valuesE0];
         AppendTo[trajectoriesE1, valuesE1];
         AppendTo[trajectoriesE2, valuesE2];
         AppendTo[trajectoriesE3, valuesE3]
        ]

In[•]:= mean[t_] := K2 + (x0 - K2) Exp[-K1 * t]

In[•]:= ListLinePlot[{Mean[trajectoriesE1], Mean[trajectoriesE2], Mean[trajectoriesE3],
          Mean[trajectoriesE0], Table[{i * Δ, mean[i * Δ]}, {i, 0, niter}]},
         PlotLegends → {"Absolute Value Euler", "Square Root Truncation Euler",
           "Full Truncation Euler", "Exact Euler", "Exact"}]
```

Modified Milstein

```
trajectoriesM1 = {};
trajectoriesM2 = {};
trajectoriesM3 = {};
For[j = 1, j ≤ tra, j++,
 valuesM1 = {{0, x0}};
 valuesM2 = {{0, x0}};
 valuesM3 = {{0, x0}};
 x1 = x0;
 x2 = x0;
 x3 = x0;
 For[i = 1, i ≤ niter, i++,
  x1 = x1 + K1 (K2 - x1) * Δ +
    K3 Sqrt[Abs[x1]] * tabled[[j, i]] + (K3^2 / 4) * (tabled[[j, i]]^2 - Δ);
  x2 = x2 + K1 (K2 - x2) * Δ + K3 Sqrt[Max [Abs[x2], 0]] * tabled[[j, i]] +
    (K3^2 / 4) * (tabled[[j, i]]^2 - Δ);
  x3 = Max[0, x3 + K1 (K2 - x3) * Δ + K3 Sqrt[x3] * tabled[[j, i]] +
     (K3^2 / 4) * (tabled[[j, i]]^2 - Δ)];
  AppendTo[valuesM1, {i * Δ, x1}];
  AppendTo[valuesM2, {i * Δ, x2}];
  AppendTo[valuesM3, {i * Δ, x3}]
 ];
 AppendTo[trajectoriesM1, valuesM1];
 AppendTo[trajectoriesM2, valuesM2];
 AppendTo[trajectoriesM3, valuesM3]
]
```

```
ListLinePlot[{Mean[trajectoriesM1], Mean[trajectoriesM2], Mean[trajectoriesM3],
  Mean[trajectoriesE0], Table[{i * Δ, mean[i * Δ]}, {i, 0, niter}]},
 PlotLegends → {"Absolute Value Milstein", "Square Root Truncation Milstein",
   "Full Truncation Milstein", "Exact Euler", "Exact"}]
```

```
ListLinePlot[{Mean[trajectoriesE1], Mean[trajectoriesE2], Mean[trajectoriesE3],
  Mean[trajectoriesM1], Mean[trajectoriesM2], Mean[trajectoriesM3],
  Mean[trajectoriesE0], Table[{i * Δ, mean[i * Δ]}, {i, 0, niter}]}, PlotLegends →
  {"Absolute Value Euler", "Square Root Truncation Euler", "Full Truncation Euler",
   "Absolute Value Milstein", "Square Root Truncation Milstein",
   "Full Truncation Milstein", "Exact Euler", "Exact"}]
```

Mean square error

```
meansquareerror[x_] :=
 1/niter * EuclideanDistance[Table[mean[i * Δ], {i, 0, niter}], x[[All, 2]]]
```

```
means = {Mean[trajectoriesE1], Mean[trajectoriesE2], Mean[trajectoriesE3],
   Mean[trajectoriesM1], Mean[trajectoriesM2], Mean[trajectoriesM3]};
```

```
Grid[{Table[meansquareerror[means[[i]]], {i, 1, 6}], {"Absolute Value Euler",
   "Square Root Truncation Euler", "Full Truncation Euler", "Absolute Value Milstein",
   "Square Root Truncation Milstein", "Full Truncation Milstein"}}, Frame → All]
```

Absolute Error

```
In[•]:= Trajectories = {trajectoriesE1, trajectoriesE2, trajectoriesE3,
          trajectoriesM1, trajectoriesM2, trajectoriesM3, trajectoriesE0};

In[•]:= AbsoluteError[traj_] := Mean[Abs[traj[[All, -1, 2]] - trajectoriesE0[[All, -1, 2]]]]

In[•]:= Grid[{Table[AbsoluteError[Trajectories[[i]]], {i, 1, 6}], {"Absolute Value Euler",
          "Square Root Truncation Euler", "Full Truncation Euler", "Absolute Value Milstein",
          "Square Root Truncation Milstein", "Full Truncation Milstein"}}, Frame → All]
```

Distribution Error

The CDF of the CIR process.

```
In[•]:= cdf[t_, x1_, x2_] := CDF[NoncentralChiSquareDistribution[4 K1 * K2 / K3^2,
           x1 * 4 K1 Exp[-K1 * t] / (K3^2 * (1 - Exp[-K1 * t]))], 4 K1 * x2 / (K3^2 * (1 - Exp[-K1 * t]))];
```

The empirical distribution of methodnumber at time step t.

```
In[•]:= ED[methodnumber_, t_] :=
         EmpiricalDistribution[Trajectories[[methodnumber]][[All, t, 2]]]
```

The distribution error at time step t for method.

```
In[•]:= CDFError[t_, method_] := .002 Sqrt[
           Total[Table[(cdf[(5 * (t - 1))/(niter + 1), x0, x] - CDF[ED[method, t], x])^2, {x, -0.1, 0.6, .002}]]]


In[•]:= DiscretePlot[
         {cdf[10, x0, x], CDF[ED[1, niter], x], CDF[ED[2, niter], x], CDF[ED[3, niter], x],
          CDF[ED[4, niter], x], CDF[ED[5, niter], x], CDF[ED[6, niter], x]},
         {x, K2 - 0.1, K2 + 0.2, .002}, PlotLegends → {Exact, "Absolute Value Euler",
           "Square Root Truncation Euler", "Full Truncation Euler", "Absolute Value Milstein",
           "Square Root Truncation Milstein", "Full Truncation Milstein"}]
```

The distribution error at the end of the time interval.

```
In[•]:= Grid[{Table[CDFError[niter, j], {j, 1, 6}], {"Absolute Value Euler",
          "Square Root Truncation Euler", "Full Truncation Euler", "Absolute Value Milstein",
          "Square Root Truncation Milstein", "Full Truncation Milstein"}}, Frame → All]
```

The mean distribution error.

```
In[•]:= Grid[{Table[Mean[Table[CDFError[i, j], {i, niter/5, niter, niter/5}]], {j, 1, 6}],
           {"Absolute Value Euler", "Square Root Truncation Euler",
            "Full Truncation Euler", "Absolute Value Milstein",
            "Square Root Truncation Milstein", "Full Truncation Milstein"}}, Frame → All]
```

References

[1] Andersen, B.G., Efficient simulation of the Heston stochastic volatility model, *SSRN* (2007) 946405

[2] Black, F., Scholes, M., The pricing of options and corporate liabilities, *Journal of Political Economy* 81 (1973) 637-654.

[3] Broadie, M., Kaya, O., Exact simulation of stochastic volatility and other affine jump diffusion processes, *Operations Research* 54 (2006), 217-231.

[4] Brown, R., A brief account of microscopical observations made in the months of June, July and August 1827, on the particles contained in the pollen of plants; and on the general existence of active molecules in organic and inorganic bodies, *The Philosophical Magazine* 4 (1828), 161-173.

[5] Calin, O. An Informal Introduction to Stochastic Calculus with Applications *World Scientific*, (2015).

[6] Cox, J., Ingersoll, J., Ross, S. A theory of the term structure of interest rates, *World Scientific* (2005), 129-164.

[7] Einstein, A., Uber die von der molekularkinetischen theorie der warme geforderte bewegung von in ruhenden, *Annalen der Physik* 4 (1905), 549-560.

[8] Evans, S., Stochastic population growth in spatially heterogeneous environments, *Journal of Mathematical Biology* 66 (2013), 423-476.

[9] Feller, W., Two singular diffusion problems *Annals of Mathematics* 54 (1951), 173-182.

[10] Heston, S., A closed-form solution for options with stochastic volatility with applications to bond and currency options, *The Review of Financial Studies* 6 (1993), 327-343.

[11] Higham, D. Mean-square and asymptotic stability of the stochastic theta method, *SIAM Journal on Numerical Analysis* 38 (2000) 753-769.

[12] Itô, K., Stochastic integral, *Proceedings of the Imperial Academy* 20 (1944), 519-524.

[13] Itô, K., On stochastic differential Equations, *American Mathematical Soc.* (1951).

[14] Kahl, C. Jackel, P, Not-so-complex logarithms in the Heston model, *Wilmott Magazine* 19 (2005), 94-103.

[15] Karatzas, I., Shreve, S., *Brownian Motion and Stochastic Calculus*, Springer, 2014.

[16] Kloeden, P., Platen, E., *Stochastic Differential Equations*, Springer, 1992.

[17] Mao, X., *Stochastic Differential Equations and Applications*, Elsevier, 2007.

[18] Miao, Z. CIR Modeling of Interest Rates, Ph. D. thesis, Linnaeus University, 2018.

[19] Milstein, G.N., Approximate integration of stochastic differential equations, *Theory of Probability & Its Applications* 19 (1975) 557-562.

[20] Milstein, G.N., Tretyakov, M., Stochastic *Numerics for Mathematical Physics*, Springer, 2013.

[21] Roberts, A.J., Modify the improved Euler scheme to integrate stochastic differential equations, preprint arXiv:1210.0933 (2012).

[22] Saito, Y., Mitsui, T., Stability analysis of numerical schemes for stochastic differential equations, *SIAM Journal on Numerical Analysis*, 33 (1996) 2254-2267.

[23] Senosiaín, M.J., Tocino, A., A review on numerical schemes for solving a linear stochastic oscillator, *BIT Numerical Mathematics* 55 (2015) 515-529.

[24] Wiener, N., *Collected Works. Vol. I: Mathematical Philosophy and Foundations; Potential Theory; Brownian Movement, Wiener Integrals, Ergodic and Chaos Theories, Turbulence and Statistical Mechanics*, (P. Masani, ed.), MIT Press, 1976.

[25] Yamada, Y., Watanabe, S., On the uniqueness of solutions of stochastic differential equations, *Journal of Mathematics of Kyoto University* 11 (1971) 155-167.