Pricing Exotic Options using Improved Strong Convergence

Klaus E. Schmitz Abe
St Catherine's College
University of Oxford

A thesis submitted for the degree of
Doctor of Philosophy
Michaelmas term 2007

.

.

.

.

to dR kaOs
and his ambitious dreams.
and of course , my mama, Vatty and my grandmother

.

.

.

.

In memory of all my friends, brothers and cousins: "In so far as the universe and the chaos are one and the same thing, the nature of our friendship relies on such a process of causality. Hence, I am grateful to the infinite for having allowed these causal encounters"
In special, my best friends:
Katie, Lynne, Cesar, Agustin, Topo, Mario, Rodolfo, Emily, Laura, Cristina, Elisa, Lola, Armando, Marcus, Amarillo, Karol, Rafa and my English mam.

.

.

.

.

I am not afraid of tomorrow
for I have seen yesterday,
and I love today....

.

.

.

.

Acknowledgements

Many people have played a part in the production of this thesis and I am very grateful to all of them. I would like to thank my supervisors, Prof. Mike Giles and Prof. William Shaw, they have been unfailingly supportive and an inspiration in my work. Thanks also to Prof. Terry Lyons for several helpful meetings and support for chapter 4. I am thankful to Prof. Paul Malliavin for his helpful discussion about his magnificent idea [6] in the common room of my department. Their comments were key in obtaining θ scheme (orthogonal Milstein scheme). I would like to mention my colleagues from OCIAM, the mathematical department and friends who have made excellent and useful suggestions.

Last but not least, thanks are due to Katie Miller and Jason Lee for their help and time editing this thesis. It was a pleasure to work with both of you.

The research was funded by CONACYT, Mexico and Microsoft Corporation, I am grateful for the funding.

Abstract

Today, better numerical approximations are required for multi-dimensional SDEs to improve on the poor performance of the standard Monte Carlo integration. With this aim in mind, the material in the thesis is divided into two main categories, stochastic calculus and mathematical finance. In the former, we introduce a new scheme or discrete time approximation based on an idea of Paul Malliavin where, for some conditions, a better strong convergence order is obtained than the standard Milstein scheme without the expensive simulation of the Lévy Area. We demonstrate when the conditions of the 2−Dimensional problem permit this and give an exact solution for the orthogonal transformation (θ Scheme or Orthogonal Milstein Scheme). Our applications are focused on continuous time diffusion models for the volatility and variance with their discrete time approximations (ARV). Two theorems that measure with confidence the order of strong and weak convergence of schemes without an exact solution or expectation of the system are formally proved and tested with numerical examples. In addition, some methods for simulating the double integrals or Lévy Area in the Milstein approximation are introduced.

For mathematical finance, we review evidence of non-constant volatility and consider the implications for option pricing using stochastic volatility models. A general stochastic volatility model that represents most of the stochastic volatility models that are outlined in the literature is proposed. This was necessary in order to both study and understand the option price properties. The analytic closed-form solution for a European/Digital option for both the Square Root Model and the 3/2 Model are given. We present the Multilevel Monte Carlo path simulation method which is a powerful tool for pricing exotic options. An improved/updated version of the ML-MC algorithm using multi-schemes and a non-zero starting level is introduced. To link the contents of the thesis, we present a wide variety of pricing exotic option examples where considerable computational savings are demonstrated using the new θ Scheme and the improved Multischeme Multilevel Monte Carlo method (MSL-MC). The computational cost to achieve an accuracy of $O(\epsilon)$ is reduced from $O(\epsilon^{-3})$ to $O(\epsilon^{-2})$ for some applications.

Contents

List of Figures

Chapter 1

Introduction

The Black-Scholes exponential Brownian motion model provides an approximate description of the behaviour of asset prices and a benchmark against which other models can be compared. However, volatility does not behave in the way the Black-Scholes equation assumes; it is not constant, it is not predictable, it is not even directly observable. Plenty of evidence exists that returns on equities, currencies and commodities are not normally distributed, they have higher peaks and fatter tails. Volatility has a key role to play in the determination of risk and in the valuation of options and other derivative securities.

As observed in empirical studies, stochastic volatility aims to reflect the apparent randomness of the level of volatility. Stochastic volatility models (SVMs) change the skewness and kurtosis of the return distribution, and option prices depend largely on these effects. SVMs are useful because they explain in a self-consistent way, why it is that options with different strikes and expirations have different Black-Scholes implied volatilities (the volatility smile). More interestingly for us, the prices of exotic options given by models based on Black-Scholes assumptions can be wildly wrong.

At the beginning of the thesis, we review evidence of non-constant volatility and consider the implications for option pricing using stochastic volatility models. A general stochastic volatility model that represents most of the stochastic volatility models that are outlined in the literature is proposed. This was necessary in order to both study and understand the option price properties. The analytic closed-form solution for a European/Digital option for both the Square Root Model [14] and the 3/2 Model [25] are given.

Any financial instrument can be priced using the exact solution for its corresponding stochastic differential equations (SDEs) and the payoff of the option. Since a closed-form expression for the arbitrage price of a claim is not always available, an important issue is the study of numerical methods which give approximations of

arbitrage prices and hedging strategies. One method uses the corresponding partial differential equations (PDEs). This method is easy and efficient to implement when one works in one or two dimensions. Unfortunately for higher dimensions, the implementation becomes more difficult and computationally very expensive. The same problem arises if one uses multinomial lattices (trees) to approximate continuous-time models of security price. The most general and famous method in the literature for pricing exotic options is the Monte Carlo method together with a discrete time approximation of the SDE. It is easy to implement and can be applied for higher dimensions without any problem.

In finance, the convergence properties of discretizations of stochastic differential equations (SDEs) are very important for hedging and the valuation of exotic options. The Milstein scheme gives first order strong convergence for all $1-$dimensional systems (one Wiener process). However, for two or more Wiener processes, such as correlated portfolios and stochastic volatility models, there is no exact solution for the iterated integrals of second order (Lévy area) and the Milstein scheme neglecting the Lévy area usually gives the same order of convergence as the Euler-Maruyama scheme.

In the middle of the thesis, we introduce a new scheme or discrete time approximation based on an idea of Paul Malliavin where, for some conditions, a better strong convergence order is obtained than the standard Milstein scheme without the expensive simulation of the Lévy Area. We demonstrate when the conditions of the $2-$Dimensional problem permit this and give an exact solution for the orthogonal transformation (θ Scheme or Orthogonal Milstein Scheme). Our applications are focused on continuous time diffusion models for the volatility and variance with their discrete time approximations (ARV). Two theorems that measure with confidence the order of strong and weak convergence of schemes without an exact solution or expectation of the system are formally proved and tested with numerical examples. In addition, some methods for simulating the double integrals or Lévy Area in the Milstein approximation are introduced.

The convergence analysis in this thesis requires the SDE to satisfy global Lipschitz conditions in the drift and diffusion coefficients. This is a standard requirement for this type of analysis (e.g. as in Kloeden and Platen [22]). However, most of the SDE models that are mentioned in the thesis, and used in the computational experiments, do not satisfy such global Lipschitz conditions; problems arise at the origin and/or at infinity. Numerical results in Chapters 4 and 5 give numerical evidence that the conclusions regarding strong order remain true in circumstances where no theory

currently exists. Though there has been some work completed on convergence analysis under non-global Lipschitz conditions [40], this topic has not been covered in the thesis.

When analyzing the option pricing problem in depth, the accuracy or error ϵ between the price option and the estimated price depends mainly on the characteristics or importance of the problem. The stochastic volatility model (SVM) and its parameters depend on the stock market data for the asset S. However, the scheme, the number of time steps and how many Monte Carlo paths are used to estimate the option price depends only on the method or algorithm applied. On the other hand, the thesis proves (as is well known in practice) that a single optimal scheme does not exist for general purposes. The selection of the scheme and the number of time steps depends totally on both the required accuracy of the problem and the parameters of the SVM. Therefore, the construction of an intelligent algorithm that can use different time approximations for different inputs will be found to be helpful.

At the end of the thesis, we present the Multilevel Monte Carlo path simulation method [10] which is a powerful tool for pricing exotic options. An improved/updated version of the ML-MC algorithm using multi-schemes and a non-zero starting level is introduced. To link the contents of the thesis, we present a wide variety of exotic option examples where considerable computational savings are demonstrated using the new θ Scheme [30] and the improved Multischeme Multilevel Monte Carlo method (MSL-MC). The computational cost to achieve an accuracy of $O(\epsilon)$ is reduced from $O(\epsilon^{-3})$ to $O(\epsilon^{-2})$ for some applications.

At the beginning of the thesis, in Chapter 2, we introduce implied, local and stochastic volatility, to review evidence of non-constant volatility and to consider the implications for option pricing of alternative random or stochastic volatility models. In the middle of the chapter, the theoretical or analytic closed-form solution for a European/Digital option for both the Heston and 3/2 Models are given. This solution has inherent advantages for pricing exotic options. At the end of the chapter we propose a general stochastic volatility model that represents most of the stochastic volatility models that are outlined in the literature. We focus on continuous time diffusion models for the volatility and variance but also briefly discuss some classes of discrete time models, such as ARV or ARCH.

Chapter 3 demonstrates how one can obtain a discrete time approximation for a $2-$Dimensional SDE using strong Taylor approximations. Later on, definitions for

both Euler and Milstein schemes for a $N-$Dimensional SDE are presented. In addition to the subject, some methods for simulating the double integrals or Lévy Area in the Milstein approximation are demonstrated. The main purpose of this chapter is to show how to measure the strong and weak order of convergence, in cases where there may, or may not, be an exact solution or expectation of our system.

The purpose of Chapter 4 is to show that if certain conditions are satisfied, one can avoid the calculation of the Lévy area and obtain first convergence order by applying an orthogonal transformation. We demonstrate when the conditions of the $2-$Dimensional problem permit this and give an exact solution for the orthogonal transformation (θ Scheme or Orthogonal Milstein Scheme).

Chapter 5 demonstrates how the use of stochastic volatility models and the θ scheme can improve the convergence of the multi-level Monte Carlo method (ML-MC [10]), so that the computational cost to achieve an accuracy of $O(\epsilon)$ is reduced from $O(\epsilon^{-3})$ to $O(\epsilon^{-2})$ for a Lipschitz payoff. We present a modification to the ML-MC algorithm that can be used to achieve better savings in some cases. To illustrate these, various examples of pricing exotic options using a wide variety of payoffs and the new Multischeme Multilevel Monte Carlo method (MSL-MC) are given. For standard payoffs, both European and Digital options are presented. For complex payoffs, such as combinations of European options, examples are also given (Butterfly Spread, Strip and Strap options). Finally, for path dependent payoffs, both Asian and Swap options are demonstrated.

In Chapter 6 we present conclusions and observations of the complete thesis. In addition, recommendations and future research are indicated. The Appendix is divided into three parts. The first section outlines the fundamental financial and mathematical definitions required to understand the thesis. In the second section are the theorems and mathematical operations required to understand both the Milstein and θ schemes. At the end of the Appendix, plots and definitions required to explain the MSL-MC are presented.

Chapter 2

Implied, Local and Stochastic Volatility

Volatility has a key role to play in the determination of risk and in the valuation of options and other derivative securities. The widespread Black-Scholes model for asset prices assumes constant volatility. The phenomenon of the implied volatility smile shows that the Black-Scholes (1973) formula tends to systematically misprice out-of-the-money[1] and in-the-money[1] options if the volatility implied from the at-the-money[1] option is used. Stochastic volatility models are useful because they explain in a self-consistent way, why it is that options with different strikes and expirations have different Black-Scholes implied volatilities (the volatility smile). More interestingly for us, the prices of exotic options given by models based on Black-Scholes assumptions can be wildly wrong.

At the beginning of the chapter, we introduce implied, local and stochastic volatility, to review evidence of non-constant volatility and to consider the implications for option pricing of alternative random or stochastic volatility models. In the middle of the chapter, the theoretical or analytic closed-form solution for a European/Digital option for both the Heston and 3/2 Models are given. This solution has inherent advantages for pricing exotic options. At the end of the chapter we propose a general stochastic volatility model that represents most of the stochastic volatility models that are outlined in the literature. We focus on continuous time diffusion models for the volatility and variance but also briefly discuss some classes of discrete time models, such as ARV or ARCH.

[1] We can find a formal definition in the Appendix.

2.1 Black-Scholes World

This section reviews the Black and Scholes arbitrage argument from option valuation under constant volatility. This allows us to introduce some frequently used notation and provides a basis for the generalization to stochastic volatility.

We start by assuming that the stock price S satisfies the following stochastic differential equation (SDE):

$$dS = S(\mu - D)dt + S\sigma dW \ , \tag{2.1}$$

where μ is the deterministic instantaneous drift or return of the stock price which pays the owner a continuous dividend D, and σ is the volatility for the stock price S. The SDE (2.1) has a solution:

$$S_T = S_0 e^{\left(\mu - D - \frac{\sigma^2}{2}\right)T + \sigma W_T} \ .$$

Besides the stock, there are two assumptions:

- Assumption 1: There is a money market security (Bank) that pays at the continuously compounded annual rate r.
- Assumption 2: Security markets are perfect. This means that you can trade continuously with no transaction costs and there are no arbitrage[2] opportunities.

Let us construct a portfolio Π consisting of one European option V with arbitrary payoff $V(S,T) = \Psi(S)$ and a number $-\phi$ of an underlying asset. The value of the portfolio at time t is:

$$\Pi = V - \phi S \ ,$$

where ϕ is constant and makes Π instantaneously risk-free. Let us consider that the dividend yield is defined as the proportion of the asset price paid out per unit time, so then, at time dt, the underlying asset pays out a dividend $D * S * dt$. The jump in the value of this portfolio in one infinitesimal time step is:

$$d\Pi = dV - \phi dS - \phi DSdt \ .$$

Hence by the principle of no arbitrage, Π must instantaneously earn the risk-free bank rate r:

$$d\Pi = r\Pi dt \ .$$

[2]There are never any opportunities to make an instantaneous risk-free profit. "There's no such thing as free lunch" [38]. There is a formal definition in page 107.

The central idea of the Black-Scholes argument is to eliminate the stochastic component of risk dW by making the number of shares equal to:

$$\phi = \frac{\partial V}{\partial S} \ .$$

Applying Itô's lemma to $V(S,t)$ and with some substitutions, one gets:

$$\frac{dV}{\partial t} + \frac{1}{2}S^2\sigma^2\frac{\partial^2 V}{\partial S^2} + S(r-D)\frac{\partial V}{\partial S} = rV \ . \tag{2.2}$$

This is **the Black-Scholes equation** and is a linear parabolic partial differential equation. In fact, almost all partial differential equations in finance are of a similar form. One of the attractions of (2.2) is that the option price function is independent of the expected return of the stock μ (hard to estimate). The Black-Scholes equation was first written down in 1969 but a few years passed, with Fisher Black and Myron Scholes justifying the model, before it was published. The derivation of the equation was finally published in 1973, although the call and put formula had been published a year earlier.

The payoffs for European (vanilla) options with strike K are:

$$V(S,T) = \left\{ \begin{array}{l} \max(S-K,0) \rightarrow \text{For } Call \text{ options} \\ \max(K-S,0) \rightarrow \text{For } Put \text{ options} \end{array} \right\} \tag{2.3}$$

and they have an analytic or closed solution in the form:

$$\begin{array}{l} Call = Se^{-D(T-t)}N(d_1) - Ke^{-r(T-t)}N(d_2) \ , \\ Put = Ke^{-r(T-t)}N(-d_2) - Se^{-D(T-t)}N(-d_1) \ , \end{array} \tag{2.4}$$

where:

$$\begin{array}{l} d_1 = \dfrac{\log\left(\frac{S}{K}\right) + \left(r - D + \frac{1}{2}\sigma^2\right)(T-t)}{\sigma\sqrt{T-t}} \ , \\ d_2 = d_1 - \sigma\sqrt{T-t} \ , \end{array}$$

and $N(d)$ is the standard normal cumulative distribution function (cdf):

$$N(d) = \frac{1}{\sqrt{2\pi}}\int_{-\infty}^{d} e^{-x^2/2}dx \ .$$

On the other hand, if one knows the value of the options, one can calculate the volatility for these instruments using the last explicit solution and a numerical method that solves (2.4) to converge to **the unique implied volatility** for this option price (e.g. use the Newton-Raphson Method). If one computes the implied volatility for

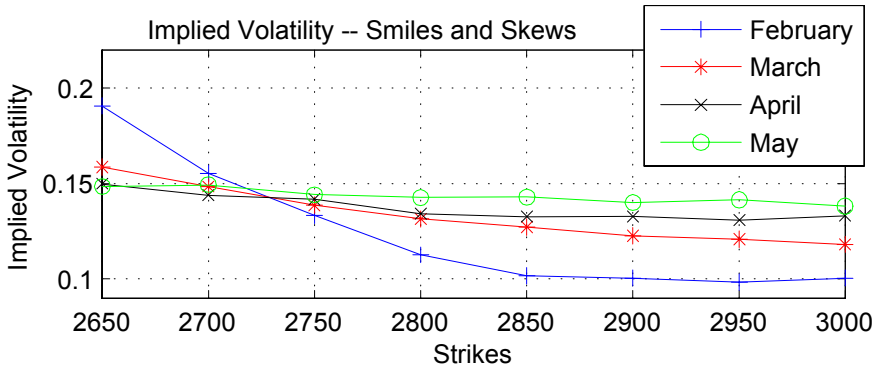

Figure 2.1: Implied volatility from the European call options (table 2.1).

market data using the option prices from Table 2.1, one would expect the same volatility for all strikes and maturities for options with the same underlying price. However, it is well known that this is not what is observed (Figure 2.1). In practice, either the term "**volatility smile**" or "**volatility skew**" may be used to refer to the general phenomena of volatilities varying by strike.

Strikes → **Maturity ↓**	**2850**	**2700**	**2750**	**2800**	**2850**	**2900**	**2950**	**3000**
Feb.	233	183	135	89	50	24	9	3
Mar.	243	197	153	113	79	51	31	17
Apr.	254	210	170	131	99	73	51	36
May	266	226	186	151	121	93	72	52
June	–	235	–	164	–	107	–	67
Dec.	–	–	–	235	–	187	–	130

Table 2.1.- Example of European call option prices obtained from the newspaper [38]. FTSE−100 index (*2872), contracts 31, 257 (Calls 21, 861, Puts 9, 396), Feb. 3, 1993.

Black and Scholes in [3] give one of the most important results of hedging, the replication argument. Using the three securities: the stock, the option and the money market security, any two of them could be used to exactly replicate the third by trading strategies. The replicating portfolio must be self-financing, which means you neither consume from it or nor add money to it beyond an initial deposit. In their original paper [3], they replicate the money market security by creating a portfolio consisting of the stock and the financial claims. Nowadays, traders replicate the option from the stock and money market account.

2.1.1 Local Volatility

Given the prices of call or put options across all strikes and maturities, one may deduce the volatility which produces those prices via the full Black-Scholes equation (Dupire, 1994 and Derman and Kani, 1994). This function has come to be known as local volatility. Unlike the naive volatility produced by applying the Black-Scholes formula to market prices, the local volatility is the volatility implied by the market prices and the one factor Black-Scholes.

In 1994, Dupire [7] showed that if the spot price follows a risk-neutral random walk of the form:

$$\frac{dS}{S} = (r - D)dt + \sigma(S, t)dW \ ,$$

and if no-arbitrage market prices for European vanilla options are available for all strikes K and expiries T, then $\sigma_L(K, T)$ can be extracted analytically from these option prices. If $C(S, t, K, T)$ denotes the price of a European call with strike K and expiry T, **Dupire's famous equation** is obtained:

$$\frac{\partial C}{\partial T} = \sigma_L^2(K, T)\frac{K^2}{2}\frac{\partial^2 C}{\partial K^2} - (r - D)K\frac{\partial C}{\partial K} - DC \ .$$

Rearranging this equation, the direct expression to calculate the local volatility (**Dupire formula**) is obtained:

$$\sigma_L(K, T) = \sqrt{\frac{\frac{\partial C}{\partial T} + (r - D)K\frac{\partial C}{\partial K} + DC}{\frac{K^2}{2}\frac{\partial^2 C}{\partial K^2}}} \ . \tag{2.5}$$

One potential problem of using the Dupire formula (2.5) is that, for some financial instruments, the option prices of different strikes and maturities are not available or are not enough to calculate the right local volatility. Another problem is for strikes far in- or out-the-money, the numerator and denominator of this equation may become very small, which could lead to numerical inaccuracies. In Figure 2.2 the Local volatility is plotted using (2.5) and Table 2.1.

"Implied volatility is the wrong number to put into the wrong formula to obtain the correct price. Local volatility on the other hand has the distinct advantage of being logically consistent. It is a volatility function which produces, via the Black-Scholes equation, prices which agree with those of the exchange traded options".

Rebonato 1999

Figure 2.2: Local volatility smile using European options (table 2.1).

2.2 Stochastic Volatility

If the Black-Scholes assumptions are correct, then the implied volatilities of options (those backed out of the Black-Scholes pricing formula given the other parameters) should fall on a horizontal line when plotted against strike prices of the options used. However, the conclusive patterns include smiles and skewed lines depending upon the underlying asset and the time period. Fifteen years ago, smiles were typical when you plotting the implied volatilities against strikes. Nowadays one is more likely to get skews or smirks.

What is happening may be viewed in some different and related ways. Options prices are determined by supply and demand, not by theoretical formula. The traders who are determining the option prices are implicitly modifying the Black-Scholes assumptions to account for volatility that changes both with time and with stock price level. This is contrary to the Black and Scholes (1973) assumptions of constant volatility irrespective of stock price or time to maturity. That is, traders assume $\sigma = \sigma\left(S(t), t\right)$, whereas Black-Scholes assume σ is just a constant.

If volatility is changing with both levels of the underlying and time to maturity, then the distribution of future stock price is no longer Lognormal. Black-Scholes option pricing takes discounted expectation payoffs relative to a Lognormal distribution. As volatility changes through time, you are likely to get periods of little activity and periods of intense activity. These periods produce peakedness and fat tails respectively (together called "leptokurtosis"), in stock return distributions. Fat tails are likely to lead to some sort of smile effect, because they increase the chance of payoffs away-from-the-money. The interaction of skewness and kurtosis of returns gives to many different possible smile effects (Hull [20]).

These irregularities have led to "stochastic volatility" models that account for volatility changing as a function of both time and stock price level. The effect of stochastic volatility on option values is similar to the effect of a jump component: both increase the probability that out-of-the-money options will finish in-the-money and vice versa (Wiggins [36]). Whether the smile is skewed left, skewed right, or symmetrical in a stochastic volatility model depends upon the sign of the correlation between changes in volatility and changes in stock price (Hull [20]).

In principle, if the continuous time model can be observed perfectly then it is possible to read off the instantaneous value of the volatility from the asset price. In practice however the volatility must be estimated from data. Suppose that the data consists of a series of daily observations of the price of an asset $(S_k)_{k<N}$. Our first estimate of the volatility, $\widehat{\sigma}$, is called the **historic volatility**. At time n, the historic volatility based on the last J days can be calculated by the maximum likelihood estimator obtained from (2.1) and the data, which is [18]:

$$\widehat{\sigma}_{n,J} = \sqrt{\frac{365}{J-1}\sum_{k=0}^{J-1}\left(\ln\left(\frac{S_{n-k}}{S_{n-k-1}}\right) - \frac{1}{J}\sum_{m=0}^{J-1}\ln\left(\frac{S_{n-m}}{S_{n-m-1}}\right)\right)^2}\,.$$

The factor of 365 converts daily volatility into an annualized term. Typically J is taken to be 90, 180 or 360 days. Figure 2.3 shows a plot of the Stock Exchange index of the stock price of the 100 leading UK companies (FTSE−100), Germany (Dax), Japan (Nikkei) and USA (Dow Jones) from December 1998 to November 2005. Figure 2.4 shows an estimate of the 360−day historic volatility based on the above data. This limited evidence supports that stock volatility is not constant at all and moreover that volatility shocks persist through time. This conclusion was reached by many authors in the literature. Stochastic volatility models are needed to describe and explain volatility patterns. Note that for the historic volatility of the Stock Exchange indices (Figure 2.4) exists a high positive correlation, one of the indices has to be the principal or dominant market that move the other ones. If we calculate the empirical correlation, we obtain:

$$corr\,[S_k] = \begin{bmatrix} 1 & 0.91 & 0.71 & 0.90 \\ 0.91 & 1 & 0.70 & 0.82 \\ 0.71 & 0.70 & 1 & 0.67 \\ 0.90 & 0.82 & 0.67 & 1 \end{bmatrix}, \quad corr\,[\widehat{\sigma}_k] = \begin{bmatrix} 1 & 0.95 & 0.96 & 0.62 \\ 0.95 & 1 & 0.87 & 0.59 \\ 0.96 & 0.87 & 1 & 0.64 \\ 0.62 & 0.59 & 0.64 & 1 \end{bmatrix},$$

where $k = 1, 2, 3, 4$ are the FTSE−100, Dax, Dow Jones and Nikkei Stock Exchange index respectively.

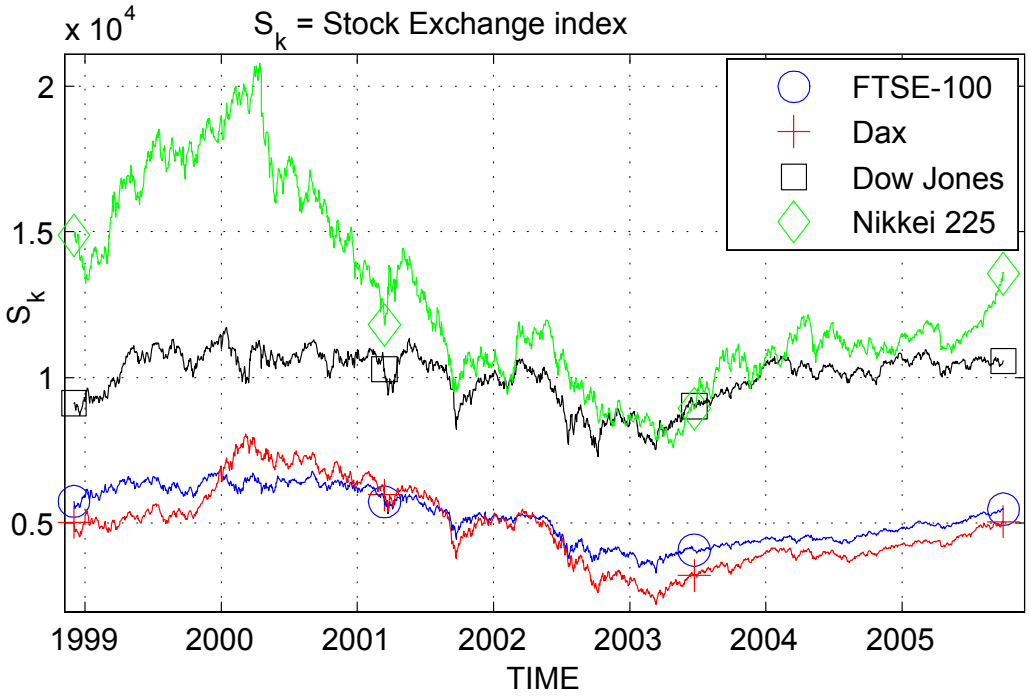

Figure 2.3: Stock Exchange index (www.londonstockexchange.com)

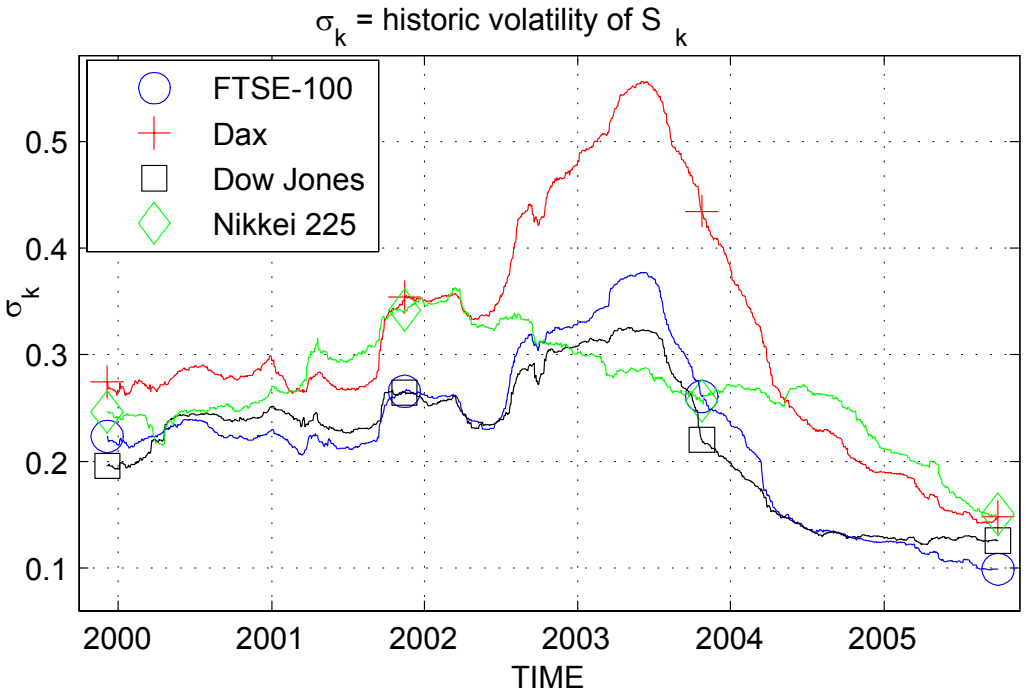

Figure 2.4: 360−day historic volatility of the Stock Exchange index

In summary, the aim with a stochastic volatility model is to incorporate the empirical observation that volatility appears not to be constant and indeed varies, at least in part, randomly. The idea is to make the volatility itself a stochastic process. The candidate models have generally been motivated by intuition, convenience and a desire for tractability. In particular the following stochastic models with their respective variance models (using Itô formula (A.1)) have all appeared in the literature:

$$\begin{aligned} dS &= S(\mu - D)dt + S\sqrt{\nu}dW_1 \ , \\ \nu &= \sigma^2 \ , \qquad E\left[dW_1 dW_2\right] = \rho\, dt \ . \end{aligned}$$

- Hull and White ([20], $\rho = 0$, 1987) and Wiggins ([36], $\rho \neq 0$, 1987):

$$d\sigma = \sigma\left[\overline{w}dt + \overline{\xi}dW_2\right] \quad \text{or}^3 \quad d\nu = \nu\left[wdt + \xi dW_2\right] \ . \tag{2.6}$$

- Scott ([31], $\rho \neq 0$, 1987) using $y = \ln\nu$:

$$dy = \left(\overline{\overline{w}} - \zeta y\right)dt + \xi dW_2 \ ,$$

$$d\sigma = \sigma\left(\overline{w} - \zeta \ln\sigma\right)dt + \overline{\xi}\sigma dW_2 \quad \text{or}^4 \quad d\nu = \nu(w - \zeta\ln\nu)dt + \xi\nu dW_2 \ . \tag{2.7}$$

- Stein and Stein ([35], $\rho = 0$, 1991):

$$d\sigma = \left(\overline{w} - \overline{\zeta}\sigma\right)dt + \overline{\xi}dW_2 \quad \text{or}^5 \quad d\nu = \left(\sqrt{\nu}w - \zeta\nu\right)dt + \xi\sqrt{\nu}dW_2 \ . \tag{2.8}$$

Variance models ($\nu = \sigma^2$) that have appeared the literature:

- The Square Root Model (Heston [14], $\rho \neq 0$, 1993):

$$d\nu = (w - \zeta\nu)dt + \xi\sqrt{\nu}dW_2 \quad \text{or}^6 \quad d\sigma = \left(\frac{\overline{w}}{\sigma} - \overline{\zeta}\sigma\right)dt + \overline{\xi}dW_2 \ . \tag{2.9}$$

- The GARCH Diffusion Model (Lewis [25], $\rho \neq 0$, 2000):

$$d\nu = \left(w - \zeta\nu\right)dt + \xi\nu dW_2 \quad \text{or}^7 \quad d\sigma = \left(\frac{\overline{w}}{\sigma} - \overline{\zeta}\sigma\right)dt + \overline{\xi}\sigma dW_2 \ . \tag{2.10}$$

[3] Using Itô formula and letting $\left(w = 2\overline{w} + \overline{\xi}^2;\ \xi = 2\overline{\xi}\right)$.

[4] Using Itô formula and letting $\left(\overline{w} = \frac{\overline{\overline{w}}}{2} + \frac{\xi^2}{8};\ \overline{\xi} = \frac{\xi}{2}\right)$ or $\left(w = \overline{\overline{w}} + \frac{\xi^2}{2}\right)$.

[5] Using Itô formula and letting $\left(w = 2\overline{w} + \frac{\overline{\xi}^2}{\sqrt{\nu}};\ \zeta = 2\overline{\zeta};\ \xi = 2\overline{\xi}\right)$.

[6] Using Itô formula and letting $\left(\overline{w} = \frac{1}{2}\left(w - \frac{\xi^2}{4}\right);\ \overline{\zeta} = \frac{\zeta}{2};\ \overline{\xi} = \frac{\xi}{2}\right)$.

[7] Using Itô formula and letting $\left(\overline{w} = \frac{w}{2};\ \overline{\zeta} = \frac{\zeta}{2} + \frac{\xi^2}{8};\ \overline{\xi} = \frac{\xi}{2}\right)$.

- The 3/2 Model (Heston [15], Lewis [25], $\rho \neq 0$, 1997):

$$d\nu = \nu(w - \zeta\nu)dt + \xi\nu^{3/2}dW_2 \quad \text{or}^{8} \quad d\sigma = \sigma^2\left[\left(\frac{\overline{w}}{\sigma} - \overline{\zeta}\sigma\right)dt + \overline{\xi}dW_2\right] \ . \tag{2.11}$$

The first model (2.6) was introduced by Hull and White (1987) who used the correlation ρ between the asset price and volatility equal to zero ($\rho = 0$). The second case was presented by Wiggins (1987) who considered the general case ($\rho \neq 0$). Here the volatility is an exponential Brownian motion, and it can grow indefinitely (or equivalently the logarithm of the volatility is a drifting Brownian motion). Scott (1987) considered the case (2.7) in which the logarithm of the volatility is an Ornstein Uhlenbeck (OU) process or a Gauss-Markov process[9]. The models (2.6) and (2.7) have the advantage that the volatility is strictly positive all the time. The third model (2.8) was proposed by Scott (1987) and further investigated by Stein and Stein (1991). These authors specialized in the case ($\rho = 0$). In this model, the volatility process itself is an OU process with a mean reversion level ω. However, the disadvantage of this model is that the volatility σ could easily become negative.

The next model (2.9) was proposed by Heston in 1993. The volatility is related to a square root process and can be interpreted as the radial distance from the origin of a multidimensional OU process [18]. For small dt, this model keeps the volatility positive and is the most popular among them because of its two main features: it has a semi-analytical pricing formula which is easy to implement and the solution is typical (it displays the same qualitative properties that one expects in general time homogenous cases). Furthermore, it can be used to understand how volatility models that do not have analytical solutions behave in many respects.

(2.10) is described as the diffusion limit of a GARCH-type process, a GARCH Diffusion, for short. From a practical point of view, the advantage of this model is that you can estimate its parameters using well known algorithms that are available as computer software, although no closed form solution is available for option pricing.

The 3/2 Model (2.11) is an important model in finance, not only because it has a closed form solution for option pricing as simple as the square root model (Heston Model), but it also displays a feature of many stochastic volatility models that you do not see in the square root model. That is, even after a change of measure to the

[8] Using Itô formula and letting $\left(\overline{w} = \frac{w}{2};\ \overline{\zeta} = \frac{\zeta}{2} + \frac{\xi^2}{8};\ \overline{\xi} = \frac{\xi}{2}\right)$.

[9] A stochastic process $\{X_T : t \succeq 0\}$ is an Ornstein-Uhlenbeck (OU) process or a Gauss-Markov process if it is stationary, Gaussian, Markovian, and continuous in probability. There is a formal definition in the Appendix of the thesis.

risk-adjusted process[10], option prices (relative to the bond price) under the 3/2 model are sometimes not martingales but merely local martingales [25]. When option prices are not martingales, this means that they are not given by the standard expected value formula, for example, $e^{-rT}E\left[\max\left(S_T - K\right)\right]$ for a call option. The failure of the usual martingale pricing relation can also occur in the GARCH Diffusion model[11]. So the 3/2 Model, with its closed form solution, is one of the simplest illustrations of this important phenomenon for financial theory. It was first used by Cox, Ingersoll, and Ross ([4], 1985) and further investigated by Heston ([15], 1997) and Lewis ([25], 2000).

Even though continuous time models provide the natural framework for an analysis of option pricing, discrete time models are ideal for the statistical and descriptive analysis of the patterns of daily price changes. There are two main classes of discrete time models for stock prices with volatility. The first class, the autoregressive random variance[12] (ARV) or stochastic volatility models, are a discrete time approximation of the continuous time diffusion models that are outlined in (2.6-2.11). The second class is the autoregressive conditional heteroskedastic (ARCH) models introduced by Engle (1982), and its descendents (GARCH, NARCH, etc.) can be defined in a variety of contents. Generally speaking, one can say that they attempt to model persistence in volatility shocks by assuming an autoregressive structure for the conditional variances (time series). A large number of parameters are often needed to approximate the behaviour of financial prices. Both the ARCH and ARV models give similar option prices (when the model parameters are appropriately matched). As shown in [1], these two models yield observational equivalents with respect to pricing options. They also notice that numerical procedures for computing option prices are faster for ARV, but estimation is simpler for ARCH. We concentrate our research on continuous time diffusion models (2.6-2.11) and in the discrete time approximation of them (ARV).

There is a simple economic argument in the literature which justifies the mean reversion of volatility. Consider the distribution of the volatility of IBM in 100 years time as an example. If the volatility was not mean-reverting (if the distribution of

[10] To value an option, you do not use (2.18), but a closely related process which is often call the risk-adjusted process $\widetilde{P}$ (2.19). For more information see Risk-Neutral Valuation for Stochastic volatility models in the Appendix of the thesis.

[11] The failure of the martingale formula for the stock price in the GARCH diffusion model was first shown by Sin [34] in 1998. These failures are specific examples of the notion that the absence of arbitrage implies that financial claim prices are, in general, only strictly local martingales, not martingales [25].

[12] An autoregressive model of a random variable is one where the random variable is assumed to exhibit a tendency to revert back to a long run mean value or distribution.

volatility was not stable), the probability of the volatility of IBM being between 1% and 100% would be rather low. Since we believe that this is overwhelmingly likely and that the volatility of IBM would, in fact, lie in that range, one can deduce that volatility must be mean-reverting.

2.2.1 Stochastic Volatility World

We begin by writing down the usual Geometric Brownian Motion SDE where the volatility σ is written as the square root of a variance ν:

$$dS = S\left(\mu - D\right)dt + S\sqrt{\nu}d\widehat{W}_1 \ , \tag{2.12}$$

where μ is the deterministic instantaneous drift or return of the stock price which pays the owner a continuous dividend D. The variance ν is constant in the original Black-Scholes model (1973) however now it is assumed to follow its own SDE in the form:

$$d\nu = f(\nu)dt + g(\nu)d\widehat{W}_2 \ , \tag{2.13}$$

where ρ is the correlation between $d\widehat{W}_1$ and $d\widehat{W}_2$. We cannot hold or "short" volatility as it is, but we can hold a position in a second option to do hedging. So let us consider the valuation of the volatility dependent instrument V (e.g. Volatility Swaps), assuming that one can take long or short positions in a second instrument U as well as in the underlying S. Now our candidate for an instantaneously risk-neutral portfolio Π is:

$$\Pi = V - \phi_1 S - \phi_2 U \ .$$

The jump in the value of this portfolio in one time step is:

$$d\Pi = dV - \phi_1 dS - \phi_2 dU - \phi_1 DSdt \ ,$$

where D is the dividend or yield on the asset S. As is by now standard, one applies Itô's Lemma to this portfolio to obtain:

$$d\Pi = adS + bd\nu + cdt \ , \tag{2.14}$$

where:

$$a = \frac{\partial V}{\partial S} - \phi_1 - \phi_2 \frac{\partial U}{\partial S} \ ,$$
$$b = \frac{\partial V}{\partial \nu} - \phi_2 \frac{\partial U}{\partial \nu} \ ,$$

$$c = \left(\frac{\partial V}{\partial t} + \frac{1}{2} S^2 \nu \frac{\partial^2 V}{\partial S^2} + \rho S \sqrt{\nu} g(\nu) \frac{\partial^2 V}{\partial S \partial \nu} + \frac{1}{2} g(\nu)^2 \frac{\partial^2 V}{\partial \nu^2} \right) - \phi_1 DS$$
$$- \phi_2 \left(\frac{\partial U}{\partial t} + \frac{1}{2} S^2 \nu \frac{\partial^2 U}{\partial S^2} + \rho S \sqrt{\nu} g(\nu) \frac{\partial^2 U}{\partial S \partial \nu} + \frac{1}{2} g(\nu)^2 \frac{\partial^2 U}{\partial \nu^2} \right) .$$

Clearly we wish to eliminate the stochastic component of risk by setting $a = b = 0$, so one can rearrange the hedge parameters in the form:

$$\phi_1 = \frac{\partial V}{\partial S} - \phi_2 \frac{\partial U}{\partial S} , \qquad \phi_2 = \left(\frac{\partial V}{\partial \nu} \right) / \left(\frac{\partial U}{\partial \nu} \right) ,$$

to eliminate the dS term and the $d\nu$ term in (2.14). The avoidance of the arbitrage, once these choices of ϕ_1, ϕ_2 are made, is the condition:

$$d\Pi = r\Pi dt ,$$

$$d\Pi = r \left(V - \phi_1 S - \phi_2 U \right) dt , \tag{2.15}$$

where we have used the fact that the return on a risk-free portfolio must be equal to the risk-free bank rate r which we will assume to be deterministic for our purposes. Combining equations (2.14) and (2.15), collecting all V terms on the left hand side and all U terms on the right hand side, one gets:

$$\left(\begin{array}{c} \frac{\partial V}{\partial t} + \frac{1}{2} S^2 \nu \frac{\partial^2 V}{\partial S^2} + \rho S \sqrt{\nu} g(\nu) \frac{\partial^2 V}{\partial S \partial \nu} \\ + \frac{1}{2} g(\nu)^2 \frac{\partial^2 V}{\partial \nu^2} + S(r - D) \frac{\partial V}{\partial S} - rV \end{array} \right) / \frac{\partial V}{\partial \nu} =$$

$$\left(\begin{array}{c} \frac{\partial U}{\partial t} + \frac{1}{2} S^2 \nu \frac{\partial^2 U}{\partial S^2} + \rho S \sqrt{\nu} g(\nu) \frac{\partial^2 U}{\partial S \partial \nu} \\ + \frac{1}{2} g(\nu)^2 \frac{\partial^2 U}{\partial \nu^2} + S(r - D) \frac{\partial U}{\partial S} - rU \end{array} \right) / \frac{\partial U}{\partial \nu} .$$

Now V, U are an arbitrary pair of derivative contracts. The only way that this can occur is when both sides of the equation are equal to some function depending only on S, ν, t. So, if one writes both sides as $F(s, \nu, t)$, in doing so, one arrives at the **General PDE for stochastic volatility**:

$$\frac{\partial V}{\partial t} + \frac{1}{2} S^2 \nu \frac{\partial^2 V}{\partial S^2} + \rho S \sqrt{\nu} g(\nu) \frac{\partial^2 V}{\partial S \partial \nu} + \frac{1}{2} g(\nu)^2 \frac{\partial^2 V}{\partial \nu^2} + S(r - D) \frac{\partial V}{\partial S} + F(\cdot) \frac{\partial V}{\partial \nu} = rV . \tag{2.16}$$

This allows us to consider how to solve (2.16) without reference to a particular volatility.

If $F(s, \nu, t)$ is written as:

$$F(s, \nu, t) = (\omega - \zeta \nu) - \overline{\Lambda} ,$$

then (2.13) becomes:

$$d\nu = \left((\omega - \zeta \nu) - \overline{\Lambda} \right) dt + g(\nu) d\widehat{W}_2 .$$

This representation models mean-reversion in the volatility σ or variance ν. Conventionally $(\omega - \zeta\nu)$ is called **the real world drift,** $\overline{\Lambda}(S, \nu, t)$ is **the market price of volatility risk**, and it tells us how much of the expected return of V is explained by the risk (standard deviation) of ν in the Capital Asset Pricing Model framework. Various economic arguments can be made (see reference [25],[32] for examples) that the market price of volatility risk $\overline{\Lambda}$ should be proportional to the variance ν. Then let $\overline{\Lambda} = \Lambda\nu$ for some constant or function Λ. Furthermore, if the real world drift is changed to:

$$k\nu^{\lambda_1}(\varpi - \nu) \quad \text{and} \quad g(\nu) = \xi\nu^{\lambda_2} \ ,$$

one gets:

$$d\nu = \nu^{\lambda_1}\left(\kappa(\varpi - \nu) - \Lambda\nu\right)dt + \xi\nu^{\lambda_2}d\widehat{W}_2 \ , \tag{2.17}$$

where κ is the mean-reverting speed, ϖ is the long-run mean, Λ is the market price of risk function, ξ is the volatility of volatility, and $d\widehat{W}_1$, $d\widehat{W}_2$ are two Wiener processes (Brownian motion) with correlation coefficient ρ. The seven parameters κ, ϖ, Λ, ξ, λ_1, λ_2, and ρ are assumed to be constant. This mean reverting variance model is a more general case compared to the traditional models because of the use of λ_2 and it can be interpreted as the radial distance from the origin of a multidimensional Ornstein Uhlenbeck process. For example, for $(\lambda_1 = 0;\ \lambda_2 = 0)$ one obtains the Stein and Stein model [35], if one uses $\lambda_2 = \frac{1}{2}$, the Heston model [14] is obtained, for $\lambda_2 = 1$, the GARCH Diffusion Model, for $(\lambda_1 = 1;\ \lambda_2 = 1.5)$ the 3/2 Model [25], and so on. For our purpose, we will use the stochastic model (2.17) as our main model in the rest of the thesis because it is a general representation for all stochastic models that are outlined above (2.6-2.11).

2.2.2 Analytic Solution for European and Digital Options

Consider the following volatility model or probability measure P under which dW_i are Brownian motions:

$$P : \left\{ \begin{array}{l} dS = S(\mu - D)dt + S\sqrt{\nu}d\widehat{W}_1 \\ d\nu = \nu^{\lambda_1}\left(\kappa(\varpi - \nu)\right)dt + \xi\nu^{\lambda_2}d\widehat{W}_2 \end{array} \right\} \ . \tag{2.18}$$

To value an option or financial security V, do not use (2.18), but a closely related process which is often call the risk-adjusted process $\widehat{P}$ (replace the expected return μ by the interest rate r, and use the risk-adjusted volatility drift φ). This procedure is carried out explicitly for a class of equilibrium models (see Risk-Neutral Valuation for Stochastic volatility models in the Appendix). The risk-adjusted process $\widehat{P}$ will

be in the risk-neutral world or equivalent martingale measure and will produce the theoretical fair price of the financial security V.

$$\widehat{P}: \left\{ \begin{array}{l} dS = S(r-D)dt + S\sqrt{\nu}d\widehat{W}_1 \\ d\nu = \nu^{\lambda_1}\left(\kappa(\varpi-\nu) - \Lambda\nu\right)dt + \xi\nu^{\lambda_2}d\widehat{W}_2 \end{array} \right\} , \tag{2.19}$$

where the parameters r, D, κ, ϖ, Λ, ξ, λ_1, λ_2, and ρ are assumed to be constant. The Heston Model $\left(\lambda_1 = 0,\ \lambda_2 = \frac{1}{2}\right)$ and the 3/2 Model $\left(\lambda_1 = 1,\ \lambda_2 = \frac{3}{2}\right)$ have a theoretical or analytic closed form solution for a European or Digital option in the form:

$$V(S,T) = \frac{1}{2\pi}e^{-r(T-t_0)}\int_{ic-\infty}^{ic+\infty}\widehat{U}(w,\nu,0)G(w,\nu,\tau)e^{-iwx}dw\ , \tag{2.20}$$

where for each model the fundamental transform G is equal to:

	$G(w,\nu,\tau)$
Heston Model	$e^{A+B\nu(t_0)}$
3/2 Model	$\frac{\Gamma_C(\beta-\alpha)}{\Gamma_C(\beta)}E^{\alpha}M\left(\alpha,\beta,-E\right)$

and $\widehat{U}$ is equal to:

Type of option	Payoff	$\widehat{U}(w,\nu,0)$	Conditions
European Call	$\max(S-K,0)$	$\frac{K^{(1+iw)}}{iw-w^2}$	$\mathrm{Im}(w) > 1$
European Put	$\max(K-S,0)$	$\frac{K^{(1+iw)}}{iw-w^2}$	$\mathrm{Im}(w) < 0$
Digital Call	$H(S-K,0)$	$\frac{-K^{iw}}{iw}$	$\mathrm{Im}(w) > 0$
Digital Put	$H(K-S,0)$	$\frac{K^{iw}}{iw}$	$\mathrm{Im}(w) < 0$

.

$H(x)$ is the Heaviside function ($H(x) = 1$ if $x > 0$, else $H(x) = 0$), $\Gamma_C(\cdot)$ is the Gamma function for complex numbers and $M(\cdot)$ is a confluent hypergeometric function[13]. If one wants to differentiate V with respect to S to obtain the option sensitivities or so-called Greeks, one merely multiplies the integral in (2.20) by:

$\Delta_S = \frac{\partial V}{\partial S}$	$\Gamma_S = \frac{\partial^2 V}{\partial S^2}$
$-\frac{iw}{S}$	$-\frac{w^2}{S^2}$

.

For the Heston Model, the constants A, B are:

$$\tau = T - t_0\ ,\quad x = \log(S) + (r-D)(T-t_0)\ ,$$

[13] The confluent hypergeometric function is a degenerate form of the hypergeometric function which arises as a solution the confluent hypergeometric differential equation. It is also known as Kummer's function of the first kind. The confluent hypergeometric function of the first kind is implemented in Mathematica as $Hypergeometric1F1[a,b,z]$ or in Matlab as $KummerComples(a,b,z)$.

$$A = \frac{k\varpi}{\xi^2}\left(\left(\kappa + \Lambda + iw\rho\xi + c_1\right)\tau - 2\log\left(\frac{1 - c_2 e^{c_1\tau}}{1 - c_2}\right)\right) \ ,$$

$$B = \frac{\left(\kappa + \Lambda + iw\rho\xi + c_1\right)}{\xi^2}\left(\frac{1 - e^{c_1\tau}}{1 - c_2 e^{c_1\tau}}\right) \ ,$$

$$c_1 = \sqrt{\left(w^2 - iw\right)\xi^2 + \left(\kappa + \Lambda + iw\rho\xi\right)^2} \ , \qquad c_2 = \frac{\kappa + \Lambda + iw\rho\xi + c_1}{\kappa + \Lambda + iw\rho\xi - c_1} \ .$$

Using the risk-aversion parameter γ, the market price of risk function Λ is equal to:

$$\Lambda\left(\gamma\right) = -\kappa + \left(1 - \gamma\right)\rho\xi + \sqrt{\kappa^2 - \gamma\left(1 - \gamma\right)\xi^2} \ ,$$

with the restriction on the parameters:

$$\gamma \leq 1 \text{ and } \kappa^2 \leq \gamma\left(1 - \gamma\right)\xi^2 \ .$$

For the 3/2 Model, the principal functions are equal to:

$$\tau = T - t_0 \ , \quad \alpha = d_2 - d_1 \ , \quad \beta = 1 + 2d_2 \ ,$$

$$E = \frac{1}{\nu\left(t_0\right)\xi^2}\frac{2w}{\left(e^{w\tau} - 1\right)} \ , \quad \widehat{\varpi} = \frac{2d_3}{\xi^2} + \frac{2\rho\left(1 - \gamma + iw\right)}{\xi} - 1 \ ,$$

$$d_1 = \frac{\left(1 + \widehat{\varpi}\right)}{2} \ , \quad d_2 = \sqrt{d_1^2 + \frac{w^2 - iw}{\xi^2}} \ , \quad d_3 = \sqrt{\left(\kappa + \frac{\xi^2}{2}\right)^2 - \gamma\left(1 - \gamma\right)\xi^2} \ .$$

Using the risk-aversion parameter γ, the market price of risk function Λ is equal to:

$$\Lambda\left(\gamma\right) = -\left(\kappa + \frac{\xi^2}{2}\right) + \left(1 - \gamma\right)\rho\xi + d_3 \ ,$$

with the restriction on the parameters:

$$\gamma \leq 1 \text{ and } \gamma\left(1 - \gamma\right)\xi^2 \leq \left(\kappa + \frac{\xi^2}{2}\right)^2 \ .$$

The Heston solution is the usual martingale-style or expected value formula. The solution of the 3/2 Model is more general and sometimes includes an additional term that relates to volatility explosions. When a European option price is not a martingale, the solution of the 3/2 Model yields the desired arbitrage-free fair value. Both solutions preserve useful properties of the Black-Scholes formula (2.4). In particular it predicts that increasing the current level of variance is equivalent to increasing the maturity of a European option. For further information or implementation, see [25], [32] or [33]. The derivation of the Heston solution is explained in more detail in the Appendix of the thesis.

2.2.3 Implementation

Mainly, the implementation of stochastic volatility models is to match the smile of the market and use this information to price exotic options. For example, one simple way that traders price their options is: using real Call and/or Put European options prices obtained from the market (newspapers, banks or Internet), then they estimate the parameters κ, ϖ, Λ, ξ, λ_1, λ_2, ρ of (2.19) using the analytical solution (2.20) and other tools. The selection on the non-trivial parameters can be time-consuming. The use of an analytic expansion in terms of the volatility of volatility is discussed by Lewis [25]. After they have estimated the parameters from (2.19) that match the real market options prices, then they price their exotic options. If there is not just a simple European or Digital payoff that does not have an analytic or closed form solution then Monte Carlo simulation is called for. For example, options with path dependent payoffs that can be found in the market do not have exact solutions, e.g. Barriers Options, Arithmetic Asian options, Variance Swap Options, etc. Nowadays, there exist hundreds of different financial derivatives in the US market.

Empirically, estimates for the long-run mean ϖ are quite variable against stock indices and can require data over a range from a few weeks to more than a year. In [25] it states that for the GARCH Diffusion Model and the US stock indices, the volatility of volatility ξ is typically in the range of 1.0 to 2.5 on an annualized basis, which represents volatility uncertainty of 100 to 250% over a year. The correlation ρ captures the association between security price and volatility changes. Typically, negative price shocks are associated with higher volatility. For the same indices, $\rho = -0.5$ to -0.8.

2.2.4 Steady-State Probability Distribution

Having so many stochastic volatility models outlined in the literature (2.6-2.11), it is important to make a comparison between all of them. One way to do it is to integrate the Fokker-Planck equation and obtain the steady-state distribution. The solution to the SDE:

$$dy = f(y,t)dt + g(y,t)dW_t \ , \qquad f(y_{t_0}) = y_0 \ ,$$

has a probability density function $p(y,t)$ which satisfies the Fokker-Planck equation (A.2) also known as the Kolmogorov forward equation:

$$\frac{\partial p}{\partial t} + \frac{\partial}{\partial y}\left(fp\right) - \frac{\partial^2}{\partial y^2}\left(\frac{1}{2}g^2 p\right) = 0 \ .$$

Under certain conditions, this evolves towards a steady-state distribution (A.3) in which $\partial p/\partial t = 0$ and hence:

$$\frac{d}{dy}(fp) - \frac{d^2}{dy^2}\left(\frac{1}{2}g^2 p\right) = 0 \ .$$

Integrating once, with the boundary condition that $p, \frac{dp}{dy} \to 0$ at infinity, gives:

$$fp - \frac{d}{dy}\left(\frac{1}{2}g^2 p\right) = 0 \quad \Longrightarrow \quad \frac{1}{g^2 p}\ \frac{d}{dy}\left(g^2 p\right) = \frac{2f}{g^2} \ .$$

Integrating this gives:

$$\log\left(g^2 p\right) = \int^y \frac{2f}{g^2} ds \ ,$$

and hence:

$$p(y) \propto \frac{1}{g^2(y)} \exp\left(\int^y \frac{2f(s)}{g^2(s)} ds\right) \ .$$

If $f\left(y_0\right) = 0$, $f'(y_0) < 0$ and $g(y_0)$ is very small, then the asymptotic approximation:

$$f(y) \approx \frac{\partial f(y_0)}{\partial y}\left(y - y_0\right), \quad g(y) \approx g(y_0) \ , \tag{2.21}$$

leads to:

$$p\left(y\right) \propto \exp\left(\frac{\left(y - y_0\right)^2}{g^2(y_0)} \frac{\partial f(y_0)}{\partial y}\right) \ . \tag{2.22}$$

The SVM (2.18) can be represented as:

$$\begin{aligned} \frac{dx}{x} &= \overline{\mu} dt + \overline{\sigma} d\widehat{W}_{1,t} \ , \\ dy &= k_j y^{\lambda_3}\left(\varpi_j^{\lambda_0} - y\right) dt + \beta_j y^{\lambda_2} d\widehat{W}_{2,t} \ , \\ \overline{\sigma} &= y^{\lambda_1} \ , \qquad\qquad j = case. \end{aligned} \tag{2.23}$$

Using Itô's lemma:

$$d\overline{\sigma} = f_j\left(\overline{\sigma}\right) dt + g_j\left(\overline{\sigma}\right) d\widehat{W}_{2,t} \ ,$$

where:

$$\begin{aligned} f_j\left(\overline{\sigma}\right) &= \lambda_1 k_j\left(\overline{\sigma}^{\frac{\lambda_1+\lambda_3-1}{\lambda_1}} \varpi_j^{\lambda_0} - \overline{\sigma}^{\frac{\lambda_1+\lambda_3}{\lambda_1}}\right) + \frac{\lambda_1\left(\lambda_1 - 1\right)\beta_j^2}{2} \overline{\sigma}^{\frac{\lambda_1+2\lambda_2-2}{\lambda_1}} \ , \\ g_j\left(\overline{\sigma}\right) &= \lambda_1 \beta_j \overline{\sigma}^{\frac{\lambda_1+\lambda_2-1}{\lambda_1}} \ . \end{aligned}$$

To make a comparison between the steady-state distribution for different cases, one can set the following equilibrium. For any choice of $\overline{\overline{\sigma}}$, using the asymptotic approximation (2.21) and taking the square root Model (Heston model, $j = H$) as the master model, we have:

- Same reversion value:

$$f_j\left(\overline{\overline{\sigma}}\right) = 0 \ . \tag{2.24}$$

- Same reversion rate:

$$\frac{\partial f_j\left(\overline{\overline{\sigma}}\right)}{\partial \overline{\sigma}} = \frac{\partial f_H\left(\overline{\overline{\sigma}}\right)}{\partial \overline{\sigma}} \ .$$

- Same volatility:

$$g_j\left(\overline{\overline{\sigma}}\right) = g_H\left(\overline{\overline{\sigma}}\right) \ .$$

Doing some operations (Appendix (A.19)), one gets:

$$\beta_j = \frac{\beta_H}{2\lambda_1}\left(\frac{k_H}{k_H\varpi_H^2 - \frac{1}{4}\beta_H^2}\right)^{\frac{\lambda_1+\lambda_2-1}{2\lambda_1}} \ , \tag{2.25}$$

$$k_j = \left(\frac{k_H}{k_H\varpi_H^2 - \frac{1}{4}\beta_H^2}\right)^{\frac{2\lambda_1+\lambda_3}{2\lambda_1}}\left(k_H\varpi_H^2 - \frac{C_{\lambda_j}\beta_H^2}{8}\right) \ , \tag{2.26}$$

$$\varpi_j = \left(\left(\varpi_H^2 - \frac{\beta_H^2}{4k_H}\right)^{\frac{1}{2\lambda_1}}\left(1 - \frac{(\lambda_1-1)\,\beta_H^2}{\lambda_1^2\left(8k_H\varpi_H^2 - C_{\lambda_j}\beta_H^2\right)}\right)\right)^{\frac{1}{\lambda_0}} \ , \tag{2.27}$$

$$C_{\lambda_j} = 2 + \frac{(\lambda_1-1)\,(\lambda_3 - 2\lambda_2 + 1)}{\lambda_1^2} \ .$$

We have taken the Heston model as the master model because it is the most well known in the literature. Setting this equilibrium allow us to introduce a mathematical relation between all stochastic volatility models SVMs outlined in (2.6-2.11). For example, in Figure 2.5 we have plotted the approximation of the steady-state distribution (2.22) using the following initial conditions from the Heston model:

$$k_H = 1, \ \varpi_H = 0.4, \ \beta_H = 0.35 \ . \tag{2.28}$$

For small vol-of-vol, $\beta_H = 0.04$, all cases match (Figure 2.9). To obtain the SVMs (2.6-2.11) one needs:

Case	λ_0	λ_1	λ_2	λ_3	k_j	ϖ_j	β_j
Hull and White, Wiggins (2.6)	1	1	1	1	2.58	0.3873	0.25
Stein and Stein (2.8)	1	1	0	0	1.00	0.3873	0.10
Heston Model (2.9)	2	0.5	1/2	0	1.00	0.4000	0.20
GARCH Diff. Model (2.10)	2	0.5	1	0	0.93	0.4009	0.51
3/2 Model (2.11)	2	1	3/2	1	6.22	0.4009	1.33

. (2.29)

The constants k_j, ϖ_j, β_j from other cases are calculated using (2.25-2.27). On the other hand, if one tries to match the expectation and variance for all cases using Newton-Raphson iteration:

$$E[\sigma] \ , \ E\left[\sigma^2\right] \ , \ g(E[\sigma]) \ , \tag{2.30}$$

you obtain Figure 2.7[14]. The distributions become more similar between them, nevertheless they are not equal. To prove that the approximation of the steady-state distribution (2.22) is good enough, we have simulated the discrete histogram in Figure 2.6. We have used the SVM (2.23) with a simple Euler scheme, Monte Carlo integration, (2.28-2.29) and $T = 10$; $y(0) = 0.3^2$. As one can see, Figure 2.6 converges to its continuous approximation (Figure 2.5).

Figures 2.5-2.6 show some important results. The equilibrium (2.24) gives as expected, the expectations and the distributions are different for all cases. All SVMs are important and have different properties. It is important the incorporation of a more general SVM that include all these features (2.18). The selection of the parameters λ_i will depend on the properties of the real data one wants to match or simulate.

2.3 Conclusions

The prices of exotic options given by models based on Black-Scholes assumptions can be wildly inaccurate because they are frequently even more sensitive to levels of volatility than standard European calls and puts. Therefore currently traders or dealers of these financial instruments are motivated to find models to price options which take the volatility smile and skew into account. To this extent, stochastic volatility models are partially successful because they can capture, and potentially explain, the smiles, skews and other structures which have been observed in market prices for options. Indeed, they are widely used in the financial community as a refinement of the Black-Scholes model.

A strong example of the existence of random correlated volatility is when the historic volatility of the Stock Exchange index is plotted (Figure 2.4). This evidence shows that stock volatility is not constant at all and moreover that volatility shocks persistently through time. This conclusion was reached by many authors in the

[14]Matlab code from M. Giles.

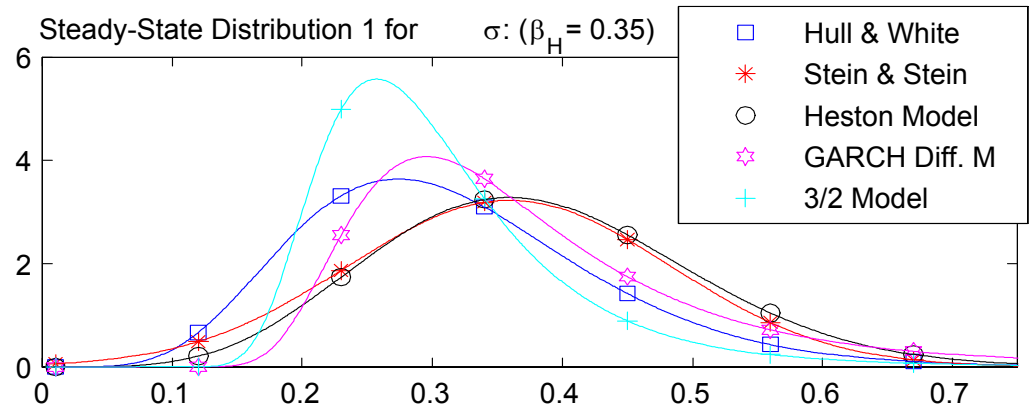

Figure 2.5: Steady-State Probability Distribution using (2.24) and $\beta_H = 0.35$.

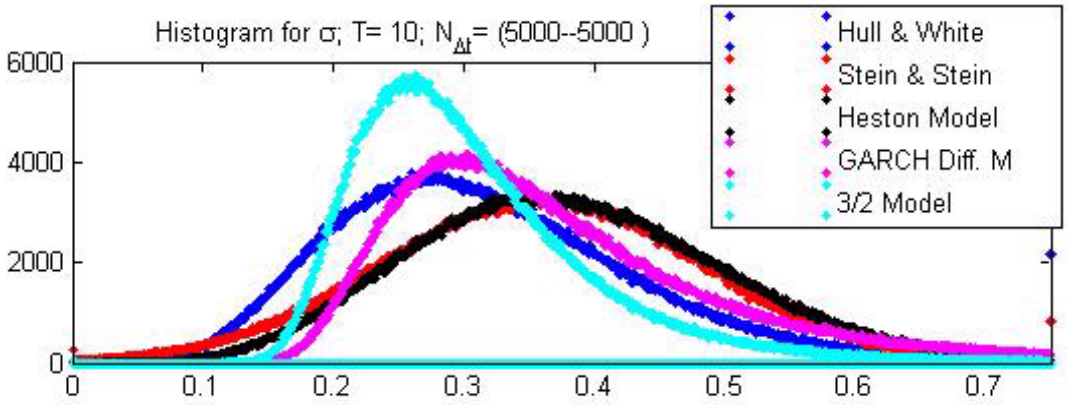

Figure 2.6: Histogram of sigma using Monte Carlo and (2.28-2.29).

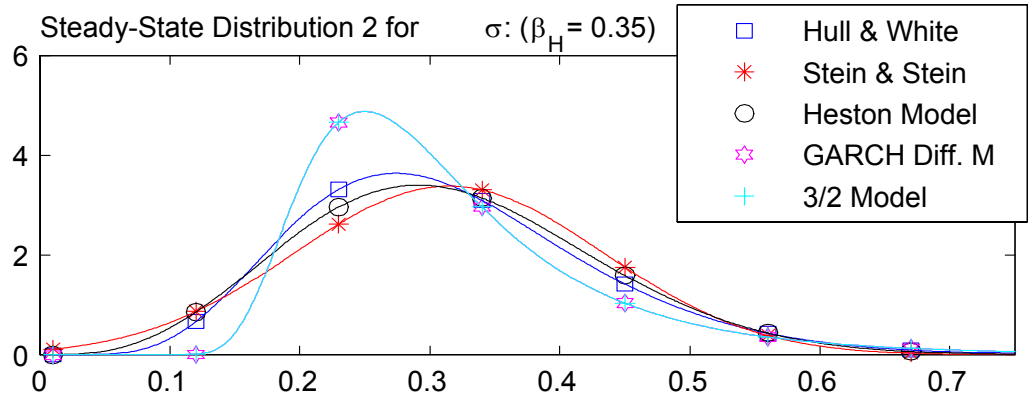

Figure 2.7: Steady-State Probability Distribution using (2.30) and $\beta_H = 0.35$.

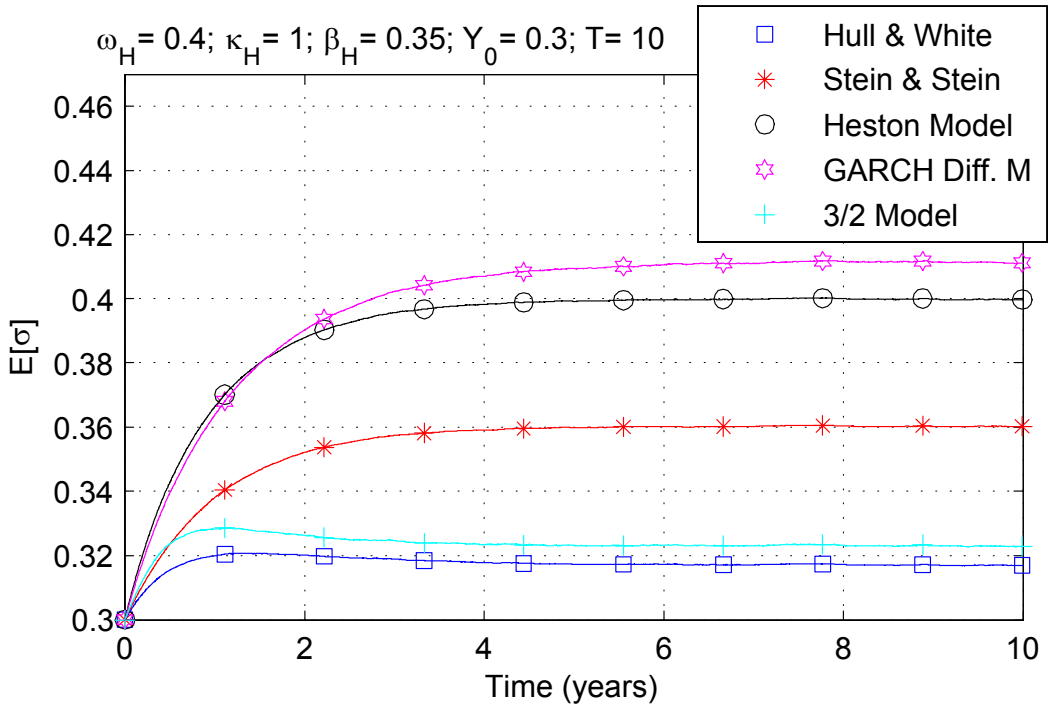

Figure 2.8: Expectation of sigma ($E[\sigma]$) using Monte Carlo and (2.28- 2.29).

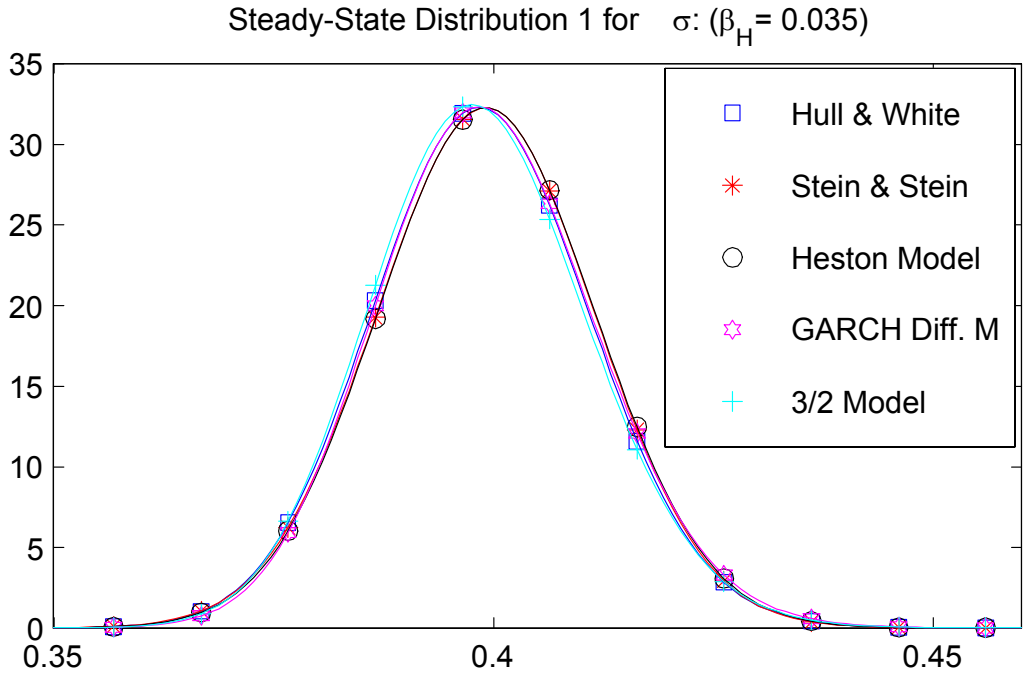

Figure 2.9: Steady-State Probability Distribution using (2.24) and $\beta_H = 0.035$.

literature; stochastic volatility models are needed to describe and explain volatility patterns.

When one analyses the steady-state probability distribution of the stochastic volatility models that are outlined in the literature, you can conclude that despite some similarities, all SVMs are important and have different properties. The definition of a more general stochastic volatility model (2.18) that represents all of them is necessary for the study and understanding of the option price properties. The selection of the parameters in (2.18) will depend on the properties of the real data one wants to match or simulate.

Chapter 3

Convergence of Discrete Time Approximations

Any financial instrument can be priced using the exact solution for its corresponding stochastic differential equations (SDEs) and the payoff of the option. Since a closed-form expression for the arbitrage price of a claim is not always available, an important issue is the study of numerical methods which give approximations of arbitrage prices and hedging strategies. One method uses the corresponding partial differential equations (PDEs). This method is easy and efficient to implement when one works in one or two dimensions. Unfortunately for higher dimensions, the implementation becomes more difficult and computationally very expensive. The same problem arises if one uses multinomial lattices (trees) to approximate continuous-time models of security price. The most general and famous method in the literature for pricing exotic options is the Monte Carlo method together with a discrete time approximation of the SDE. It is easy to implement and can be applied for higher dimensions without any problem.

At the beginning of the chapter we demonstrate how one can obtain a discrete time approximation for a $2-$Dimensional SDE using strong Taylor approximations. Later on, definitions for both Euler and Milstein schemes for a $N-$Dimensional SDE are presented. In addition to the subject, some methods for simulating the double integrals or Lévy Area in the Milstein approximation are demonstrated. The main purpose of this chapter is to show how to measure the strong and weak order of convergence, in cases where there may, or may not, be an exact solution or expectation of our system.

3.1 Introduction

Following Einstein's explanation of physically observed Brownian motion during the first decade of the $1900s$, the physicists Langevin, Smoluchowski, and others attempted to model the dynamics of such motion in terms of differential equations. Instead of a deterministic ordinary differential equation:

$$\frac{dX_t}{dt} = a(X_t, t) \ ,$$

they obtained a noisy differential equation of the form:

$$\frac{dX_t}{dt} = a(X_t, t) + b(X_t, t)Z_t \ , \tag{3.1}$$

with a deterministic or averaged drift term $a(X_t, t)$ perturbed by a noise intensity term $b(X_t, t)Z_t$, where $b(X_t, t)$ is the intensity factor and Z_t are independent normal distributed Gaussian random variables for each time t. The driving process Z_t, which is called Gaussian white noise, appears formally to be the path-wise derivative of a mathematical Brownian motion or Wiener process W_t. A Gaussian process with $W_{t_0} = 0$, continuous paths and $N(0, t)-$distributed for each W_t:

$$E\left[W_t\right] = 0 \ , \qquad E\left[W_t^2\right] = t \ ,$$

which has independent increments for any $t_1 < t_2 < t_3 < t_4 \in [t_0, T]$:

$$E\left[\left(W_{t_4} - W_{t_3}\right)\left(W_{t_2} - W_{t_1}\right)\right] = 0 \ .$$

However, the Gaussian white noise process is not a conventional process, having, for example, covariance equal to a constant multiple of the Dirac delta function. Moreover, it is now known that the sample paths of a Wiener process W_t are nowhere differentiable. This suggests that the stochastic differential equation (3.1), which might be written symbolically in terms of differentials as:

$$dX_t = a\left(X_t, t\right) dt + b\left(X_t, t\right) dW_t \ ,$$

should be interpreted in some sense as an integral equation:

$$X_t = X_{t_0} + \int_{t_0}^{t} a\left(X_s, s\right) ds + \int_{t_0}^{t} b\left(X_s, s\right) dW_s \ .$$

The first integral here is just path-wise an ordinary Riemann integral, while it might seem that the second integral could be a Riemann-Stieltjes integral for each sample path. This is not possible because the sample paths of a Wiener process are not

just not differentiable, but also are not even of bounded variation on any bounded time interval. In the $1940s$, the Japanese mathematician K. Itô proposed a means to overcome this difficulty with the definition of a new type of integral, a stochastic integral, which is now called an Itô stochastic integral. Later, in the $1960s$, the Russian physicist R. L. Stratonovich proposed another kind of stochastic integral, now called the Stratonovich stochastic integral, which is distinguished from the Itô integral by a "$\circ$" before the differential dW_t, e.g. written symbolically in the differential form:

$$dX_t = a(X_t, t)dt + b(X_t, t) \circ dW_t \ .$$

However it should be interpreted as an integral equation:

$$X_t = X_{t_0} + \int_{t_0}^{t} a\left(X_s, s\right) ds + \int_{t_0}^{t} b\left(X_s, s\right) \circ dW_s \ .$$

There are thus two types of stochastic calculus, the Itô stochastic calculus and the Stratonovich stochastic calculus, depending on the type of stochastic integral used. Both have their advantages as well as their disadvantages. Which one should be used is more a modelling than mathematical issue, but once one has been chosen, a corresponding equation of the other type with the same solutions can be determined. Therefore, it is possible to switch between the two stochastic calculi.

3.2 Stochastic Taylor Series

In this section we shall use stochastic Taylor expansions to derive discrete time approximations with respect to the strong convergence criterion[1], which we shall call strong Taylor approximations. These expansions are derived through an iterated application of the stochastic chain rule, known as the Itô formula. We shall see that the desired order of strong convergence determines the truncation to be applied. Consider the following $2-$dimensional stochastic differential equation:

$$\begin{aligned} dX_t &= A(X_t, Y_t, t)dt + B(X_t, Y_t, t)d\widehat{W}_{1,t} \ , \qquad X_{t_0} = X_0 \ , \\ dY_t &= C(X_t, Y_t, t)dt + D(X_t, Y_t, t)d\widehat{W}_{2,t} \ , \qquad Y_{t_0} = Y_0 \ , \end{aligned} \tag{3.2}$$

where $d\widehat{W}_{1,t}$ and $d\widehat{W}_{2,t}$ are two correlated Wiener processes. The definition of a correlation matrix for a $2D$ system is:

$$\Omega = \begin{bmatrix} 1 & \rho \\ \rho & 1 \end{bmatrix} \ . \tag{3.3}$$

[1]When we talk about strong convergence, we are referring to how fast our discrete time approximation converges to the exact solution as it is refined by Δt. We can find a formal definition in page 44.

Using a Cholesky factorization ($\Omega = LL^T$ with L lower triangular), dW can be defined in 2 ways. The standard and most common is:

$$\left[\begin{array}{c} d\widehat{W}_{1,t} \\ d\widehat{W}_{2,t} \end{array} \right]_{Std} = \left[\begin{array}{cc} 1 & 0 \\ \rho & \widehat{\rho} \end{array} \right] \left[\begin{array}{c} dW_{1,t} \\ dW_{2,t} \end{array} \right] , \qquad \hat{\rho} = \sqrt{1-\rho^2} . \tag{3.4}$$

Using the standard definition of correlation (3.4), (3.2) can be transformed into two equations with independent noise sources:

$$\begin{aligned} dX_t &= A(Z_t)dt + B(Z_t)dW_{1,t} , \\ dY_t &= C(Z_t)dt + \rho D(Z_t)dW_{1,t} + \widehat{\rho} D(Z_t)dW_{2,t} , \\ Z_t &= [X_t, Y_t, t] , \qquad \langle dW_{1,t}, dW_{2,t} \rangle = 0 , \end{aligned}$$

which are the short hand for the integral equations:

$$\begin{aligned} X_t &= X_{t_0} + \int_{t_0}^{t} A(Z_s)ds + \int_{t_0}^{t} B(Z_s)dW_{1,s} , \\ Y_t &= Y_{t_0} + \int_{t_0}^{t} C(Z_s)ds + \rho \int_{t_0}^{t} D(Z_s)dW_{1,s} + \hat{\rho} \int_{t_0}^{t} D(Z_s)dW_{2,s} . \end{aligned} \tag{3.5}$$

The first integrals are deterministic Riemann integrals and the rest are Itô stochastic integrals. More generally, if f is a differentiable function of Z, one obtains Itô's lemma:

$$\begin{aligned} df(Z_t) &= \left(\frac{\partial f(Z_t)}{\partial t} + A(Z_t) \frac{\partial f(Z_t)}{\partial X} + C(Z_t) \frac{\partial f(Z_t)}{\partial Y} \right) dt \\ &+ \left(\frac{B^2(Z_t)}{2} \frac{\partial f^2(Z_t)}{\partial X^2} + \rho B(Z_t) D(Z_t) \frac{\partial f^2(Z_t)}{\partial X \partial Y} + \frac{D^2(Z_t)}{2} \frac{\partial f^2(Z_t)}{\partial Y^2} \right) dt \\ &+ \left(B(Z_t) \frac{\partial f(Z_t)}{\partial X} + \rho D(Z_t) \frac{\partial f(Z_t)}{\partial Y} \right) dW_{1,t} + \hat{\rho} D(Z_t) \frac{\partial f(Z_t)}{\partial Y} dW_{2,t} . \end{aligned} \tag{3.6}$$

One can write Itô's lemma (3.6) in its integrated form:

$$\begin{aligned} f(Z_t) &= f(Z_{t_0}) + \int_{t_0}^{t} (L_{de} f)(Z_s) ds \\ &\quad + \int_{t_0}^{t} (L_{W_1} f)(Z_s) dW_{1,s} + \int_{t_0}^{t} (L_{W_2} f)(Z_s) dW_{2,s} , \end{aligned} \tag{3.7}$$

where L_{de}, L_{W_1}, L_{W_2} are deterministic operators. Applying the results (3.7) to $f(Z_t) = A(Z_t)$:

$$A(Z_t) = A(Z_{t_0}) + \int_{t_0}^{t} (L_{de} A)(Z_u) du + \sum_{i=1}^{2} \left(\int_{t_0}^{t} (L_{W_i} A)(Z_u) dW_{i,u} \right) .$$

Doing the same to $f(Z_t) = B(Z_t)$ and substituting into (3.5), X_t can be represented by:

$$\begin{aligned} X_t &= X_{t_0} + A(Z_{t_0}) \int_{t_0}^{t} ds + B(Z_{t_0}) \int_{t_0}^{t} dW_{1,s} \\ &\quad + \left[\widetilde{I}_{(W,W)} + \widetilde{I}_{(dt,W)} + \widetilde{I}_{(W,dt)} + \widetilde{I}_{(dt,dt)} \right]_{t_0}^{t} , \end{aligned} \tag{3.8}$$

where the Itô integrals are defined by:

$$\left[\widetilde{I}_{(W,W)} \right]_{t_0}^{t} = \int_{t_0}^{t} \sum_{i=1}^{2} \left(\int_{t_0}^{s} \left(L_{W_i} B \right) (Z_u) \, dW_{i,u} \right) dW_{1,s} \quad = O(\Delta t) , \tag{3.9}$$

$$\left[\widetilde{I}_{(dt,W)} \right]_{t_0}^{t} = \int_{t_0}^{t} \int_{t_0}^{s} \left(L_{de} B \right) (Z_u) \, du dW_{1,s} \quad = O\left(\Delta t^{\frac{3}{2}} \right) , \tag{3.10}$$

$$\left[\widetilde{I}_{(W,dt)} \right]_{t_0}^{t} = \int_{t_0}^{t} \sum_{i=1}^{2} \left(\int_{t_0}^{s} \left(L_{W_i} A \right) (Z_u) \, dW_{i,u} \right) ds \quad = O\left(\Delta t^{\frac{3}{2}} \right) , \tag{3.11}$$

$$\left[\widetilde{I}_{(dt,dt)} \right]_{t_0}^{t} = \int_{t_0}^{t} \int_{t_0}^{s} \left(L_{de} A \right) (Z_u) \, du ds \quad = O\left(\Delta t^{2} \right) . \tag{3.12}$$

The traditional Euler approximation is essentially equivalent to ignoring all the last double integrals. If $\Delta t = t - t_0$ is a fixed stepsize, (3.8) leads to:

$$\widehat{X}_{t+\Delta t} = \widehat{X}_t + A\left(\widehat{Z}_t \right) \Delta t + B\left(\widehat{Z}_t \right) \Delta W_{1,t} .$$

If a better approximation is required, one needs to approximate one or more of the double integrals (3.9,3.10,3.11 & 3.12). Because one is dealing with Brownian motion, the next approximation can be used:

$$\left[I_{(i,i)} \right]_{t}^{t+\Delta t} = \int_{t}^{t+\Delta t} \int_{t}^{s} dW_{i,u} dW_{i,s} \; = \int_{t}^{t+\Delta t} W_{i,s} dW_{i,s} - W_{i,t} \int_{t}^{t+\Delta t} dW_{i,s} ,$$

using Itô's lemma[2]:

$$\begin{aligned} \left[I_{(i,i)} \right]_{t}^{t+\Delta t} &= \frac{1}{2} \int_{t}^{t+\Delta t} \left(dW_{i,s}^2 - ds \right) - W_{i,t} \left(W_{i,t+\Delta t} - W_{i,t} \right) \\ &= \frac{1}{2} \left(W_{i,t+\Delta t}^2 - t - \Delta t \right) - \frac{1}{2} \left(W_{i,t}^2 - t \right) - W_{i,t+\Delta t} W_{i,t} + W_{i,t}^2 \\ &= \frac{1}{2} \left(\left(W_{i,t+\Delta t} - W_{i,t} \right)^2 - \Delta t \right) = \frac{1}{2} \left(\Delta W_{i,t}^2 - \Delta t \right) . \end{aligned} \tag{3.13}$$

Note that the essence of the method is to use the substitution repeatedly to obtain constant integrands in higher order terms. For example, if one repeats this argument

[2] If we apply Itô's lemma to $f = W^2$; we obtain $d(W)^2 = 2WdW + dt$, then $WdW = \frac{1}{2}\left(dW^2 - dt \right)$.

applying (3.7) to $f(Z_t) = (L_{W_1}B)(Z_t)$ and $f(Z_t) = (L_{W_2}B)(Z_t)$, and one uses (3.13), the double Itô integrals (3.9) can be approximated by:

$$\left[\widetilde{I}_{(W,W)}\right]_t^{t+\Delta t} = \frac{1}{2}(L_{W_1}B)(Z_t)\left(\Delta W_{1,t}^2 - \Delta t\right) + (L_{W_2}B)(Z_t)\left[I_{(2,1)}\right]_t^{t+\Delta t} + O\left(\Delta t^{\frac{3}{2}}\right) . \tag{3.14}$$

(3.14), usually called the Milstein correction, is a stochastic effect (a result of Itô's lemma[1] if you like). Unfortunately, there is not a solution or approximation for the other double integral:

$$\left[I_{(2,1)}\right]_t^{t+\Delta t} = \int_t^{t+\Delta t}\int_t^s dW_{2,u}dW_{1,s} . \tag{3.15}$$

By truncating X_t at $O(t)$, the 1.0 strong order of convergence, usually called the Milstein scheme is:

$$\begin{aligned}\widehat{X}_{t+\Delta t} &= \widehat{X}_t + A\left(\widehat{Z}_t\right)\Delta t + B\left(\widehat{Z}_t\right)\Delta W_{1,t} \qquad (3.16)\\ &+ \frac{1}{2}(L_{W_1}B)(Z_t)\left(\Delta W_{1,t}^2 - \Delta t\right) + (L_{W_2}B)(Z_t)\left[I_{(2,1)}\right]_t^{t+\Delta t} ,\end{aligned}$$

where:

$$\begin{aligned}(L_{W_1}\bullet)(Z_t) &= B\frac{\partial \bullet}{\partial X} + \rho D\frac{\partial \bullet}{\partial Y} , \qquad (3.17)\\ (L_{W_2}\bullet)(Z_t) &= \hat{\rho} D\frac{\partial \bullet}{\partial Y} .\end{aligned}$$

It is well known in the literature the concept of the Lévy Area [24] which is defined by:

$$\left[\overline{L}_{(1,2)}\right]_t^{t+\Delta t} := \int_t^{t+\Delta t}\int_t^s dW_{1,u}dW_{2,s} - \int_t^{t+\Delta t}\int_t^s dW_{2,u}dW_{1,s} . \tag{3.18}$$

In addition, if one applies Itô's lemma[3]:

$$\begin{aligned}\left[I_{(1,2)}\right]_t^{t+\Delta t} &= \int_t^{t+\Delta t}(W_{1,s} - W_{1,t})\,dW_{2,s} \qquad (3.19)\\ &= \int_t^{t+\Delta t} W_{1,s}dW_{2,s} - W_{1,t}W_{2,t+\Delta t} + W_{1,t}W_{2,t}\\ &= -\int_t^{t+\Delta t} W_{2,s}dW_{1,s} + W_{1,t+\Delta t}W_{2,t+\Delta t} - W_{1,t}W_{2,t+\Delta t} ,\end{aligned}$$

and for the other Itô integral:

$$\left[I_{(2,1)}\right]_t^{t+\Delta t} = \int_t^{t+\Delta t} W_{2,s}dW_{1,s} - W_{1,t+\Delta t}W_{2,t} + W_{1,t}W_{2,t} . \tag{3.20}$$

[3] If we apply Itô's lemma to $f = W_1W_2$; we obtain $d(W_1W_2) = W_2dW_1 + W_1dW_2$.

Adding both equations, (3.19) and (3.20), one obtains:

$$\left[I_{(1,2)}\right]_t^{t+\Delta t} + \left[I_{(2,1)}\right]_t^{t+\Delta t} = \left(W_{1,t+\Delta t} - W_{1,t}\right)\left(W_{2,t+\Delta t} - W_{2,t}\right) = \Delta W_{1,t}\Delta W_{2,t} \ . \quad (3.21)$$

Using (3.18) and (3.21), (3.15) can be expressed as:

$$\begin{aligned}\left[I_{(2,1)}\right]_t^{t+\Delta t} &= \frac{1}{2}\left[\left(\left[I_{(1,2)}\right] + \left[I_{(2,1)}\right]\right) - \left(\left[I_{(1,2)}\right] - \left[I_{(2,1)}\right]\right)\right]_t^{t+\Delta t} \\ &= \frac{1}{2}\left(\Delta W_{1,t}\Delta W_{2,t} - \left[\underline{\overline{L}}_{(1,2)}\right]_t^{t+\Delta t}\right) \ .\end{aligned}$$

So, using the concept of Lévy Area (3.18), the Milstein scheme (3.16) leads to:

$$\widehat{X}_{t+\Delta t} = \widehat{X}_t + A\left(\widehat{Z}_t\right)\Delta t + B\left(\widehat{Z}_t\right)\Delta W_{1,t} + \frac{1}{2}\left(L_{W_1}B\right)\left(\widehat{Z}_t\right)\left(\Delta W_{1,t}^2 - \Delta t\right)$$

$$+\frac{1}{2}\left(L_{W_2}B\right)\left(\widehat{Z}_t\right)\left(\Delta W_{1,t}\Delta W_{2,t} - \left[\underline{\overline{L}}_{(1,2)}\right]_t^{t+\Delta t}\right) \ .$$

Doing the same argument above for $Y(t)$, the discrete time strong approximation of order 1.0 (Milstein scheme) for $Y(t)$ is:

$$\widehat{Y}_{t+\Delta t} = \widehat{Y}_t + C\left(\widehat{Z}_t\right)\Delta t + D\left(\widehat{Z}_t\right)\Delta\widehat{W}_{2,t} + \left[\rho\widetilde{I}_{Y,1} + \hat{\rho}\ \widetilde{I}_{Y,2}\right]_t^{t+\Delta t} \ ,$$

where the Itô integrals are:

$$\begin{aligned}\left[\ \widetilde{I}_{Y,1}\right]_t^{t+\Delta t} &= \frac{1}{2}\left(L_{W_1}D\right)\left(\widehat{Z}_t\right)\left(\Delta W_{1,t}^2 - \Delta t\right) + \left(L_{W_2}D\right)\left(\widehat{Z}_t\right)\left[I_{(2,1)}\right]_t^{t+\Delta t} \ , \\ \left[\ \widetilde{I}_{Y,2}\right]_t^{t+\Delta t} &= \frac{1}{2}\left(L_{W_2}D\right)\left(\widehat{Z}_t\right)\left(\Delta W_{2,t}^2 - \Delta t\right) + \left(L_{W_1}D\right)\left(\widehat{Z}_t\right)\left[I_{(1,2)}\right]_t^{t+\Delta t} \ .\end{aligned}$$

If one applies back the Itô's operators (3.17), the 1.0 strong order scheme (Milstein scheme) for our original system (3.2) is:

$$\widehat{X}_{t+\Delta t} = \widehat{X}_t + A\Delta t + B\Delta W_{1,t} + \frac{1}{2}B\frac{\partial B}{\partial X}\left(\Delta W_{1,t}^2 - \Delta t\right) + \frac{1}{2}D\frac{\partial B}{\partial Y}\left[\Delta W_L^-\right]_t^{t+\Delta t} \ ,$$

$$\widehat{Y}_{t+\Delta t} = \widehat{Y}_t + C\Delta t + D\Delta\widehat{W}_{2,t} + \frac{1}{2}D\frac{\partial D}{\partial Y}\left(\Delta\widehat{W}_{2,t}^2 - \Delta t\right) + \frac{1}{2}B\frac{\partial D}{\partial X}\left[\Delta W_L^+\right]_t^{t+\Delta t} \ ,$$

where:

$$\left[\Delta W_L^{\pm}\right]_t^{t+\Delta t} = \rho\left(\Delta W_{1,t}^2 - \Delta t\right) + \hat{\rho}\left(\Delta W_{1,t}\Delta W_{2,t} \pm \left[\underline{\overline{L}}_{(1,2)}\right]_t^{t+\Delta t}\right) \ .$$

Milstein integration includes all the $O(\Delta t)$ terms neglecting $O\left(\Delta t^{\frac{3}{2}}\right)$ and higher terms. As noted above, the Milstein correction is peculiar to SDEs (as opposed to ODEs) and is a consequence of Itô's lemma and/or the definition of an Itô integral. It can be shown in a more rigorous proof (see Kloeden and Platen [22]) that the

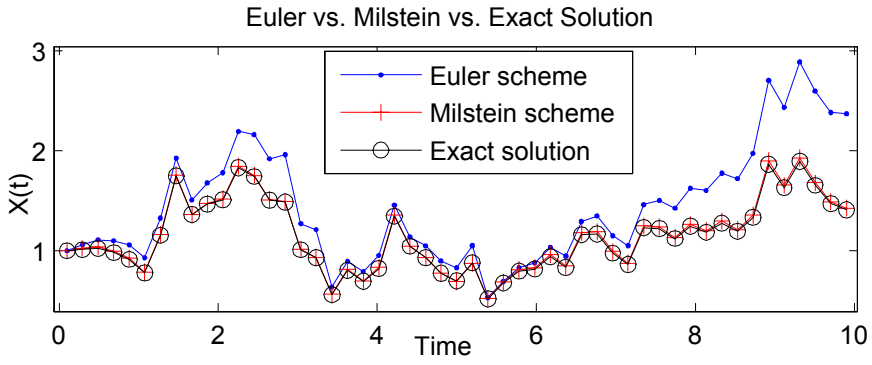

Figure 3.1: One random simulation for an Exponential Brownian Motion process (3.43) using Euler and Milstein scheme ($N_{\triangle t} = 50$).

Milstein scheme gives 1.0 strong order convergence, and in general, is a better scheme than Euler approximation for integrating a SDE (Figure 3.1). An integration scheme which is more accurate than the accuracy of the Milstein scheme requires the evaluation of the integrals (3.10,3.11 & 3.12). Unfortunately, these integrals and (3.15) from Milstein scheme cannot be accurately expressed in terms of the under laying random process ΔW_t alone. Rather, accurate evaluation of these integrals requires the generation of additional random numbers.

3.3 Discrete Time Approximations ($d-$Dimensional)

Appropriate stochastic Taylor expansions can give consistent numerical schemes of an arbitrarily desired higher order. These expansions are derived through an iterated application of the stochastic chain rule (Itô formula). In this section we shall define Euler and Milstein schemes for a general case: a $d-$Dimensional Itô stochastic differential equation with a $M-$Dimensional Wiener process. Both schemes can be defined in many different ways and all of them are necessary for different applications. The most important two representations in the literature are using "Itô operators" and the "Vector form". Both have their advantages as well as their disadvantages depending on the dimension and where they are used. Both representations give exactly the same solution.

3.3.1 Euler and Milstein Scheme (Itô Operators)

All models can be described through the following $d-$Dimensional Itô stochastic differential equation (SDE) with a $M-$Dimensional independent Wiener process:

$$dX_t = \mu\left(X_t, t\right) dt + \sigma\left(X_t, t\right) dW_t \ , \qquad X_{t_0} = X_0 \ , \tag{3.22}$$

where:

$$X_t \in \mathbb{R}^d, \ W_t \in \mathbb{R}^M, \ t \in [t_0, T] \ ,$$
$$E\left[dW_{j,t} dW_{k,t}\right] = 0 \ , \qquad for \ j \neq k \ ,$$

and μ, σ are sufficiently smooth functions of X and t, e.g. satisfy Lipschitz conditions (Theorem 1, page 46):

$$\mu\left(X_t, t\right) = \left[\begin{array}{cccc} a_1 & a_2 & ... & a_d \end{array} \right]^T \in \mathbb{R}^d \ ,$$

$$\sigma\left(X_t, t\right) = \left[\begin{array}{cccc} b_{1,1} & b_{1,2} & ... & b_{1,M} \\ b_{2,1} & b_{2,2} & ... & b_{2,M} \\ ... & ... & ... & ... \\ b_{d,1} & b_{d,2} & ... & b_{d,M} \end{array} \right] \in \mathbb{R}^{d \times M} \ .$$

The 0.5 strong order Euler scheme for (3.22) with time step Δt ([22], page 340) is:

$$\widehat{X}_{i,t+\Delta t} = \widehat{X}_{i,t} + a_i \Delta t + \sum_{j=1}^{M} b_{i,j} \Delta W_{j,t} \ .$$

The 1 strong order Milstein scheme for (3.22) with time step Δt using Itô operators ([22], page 345) is:

$$\widehat{X}_{i,t+\Delta t} = \widehat{X}_{i,t} + a_i \Delta t + \sum_{j=1}^{M} b_{i,j} \Delta W_{j,t} + R_M \ ,$$

where if one uses the double Itô integrals, R_M is equal to:

$$R_M = \sum_{j_1, j_2 = 1}^{M} \left(L_{j_1} b_{i,j_2}\right) \left[I_{(j_1,j_2)}\right]_t^{t+\Delta t} \ ,$$

$$\left[I_{(j_1,j_2)}\right]_t^{t+\Delta t} = \int_t^{t+\Delta t} \int_t^{U_1} dW_{j_1,U_2} dW_{j_2,U_1} \ ,$$

or using Lévy Areas, R_M is equal to:

$$\begin{aligned} R_M \ &= \ \frac{1}{2} \sum_{j_1, j_2 = 1}^{M} L_{j_1} b_{i,j_2} \left(\Delta W_{j_1,t} \Delta W_{j_2,t} - \widetilde{\delta}_{j_1,j_2} \Delta t\right) \\ &+ \frac{1}{2} \sum_{j_1=1}^{M} \sum_{j_2=j_1+1}^{M} \left(L_{j_1} b_{i,j_2} - L_{j_2} b_{i,j_1}\right) \left[\underline{I}_{(j_1,j_2)}\right]_t^{t+\Delta t} \ , \end{aligned}$$

$$\left[\underline{L}_{(j_1,j_2)}\right]_t^{t+\Delta t} = -\left[\underline{L}_{(j_2,j_1)}\right]_t^{t+\Delta t} = \left[I_{(j_1,j_2)}\right]_t^{t+\Delta t} - \left[I_{(j_2,j_1)}\right]_t^{t+\Delta t} \ .$$

The Itô operators are defined by:

$$L_j := \sum_{k=1}^{d} b_{k,j} \frac{\partial}{\partial X_k} \ ,$$

and $\widetilde{\delta}_{j_1,j_2}$ is the Kronecker symbol defined by $\widetilde{\delta}_{j_1,j_2} = 1$ if $j_1 = j_2$ and zero otherwise. Both expressions for R_M can be seen to be equivalent if one uses the following properties:

$$\begin{aligned} \left[I_{(j,j)}\right]_t^{t+\Delta t} &= \frac{1}{2}\left((\Delta W_{j,t})^2 - \Delta t\right) \ , \\ \left[I_{(j_1,j_2)}\right]_t^{t+\Delta t} &= \frac{1}{2}\left(\Delta W_{j_1,t}\Delta W_{j_2,t} + \left[\underline{L}_{(j_1,j_2)}\right]_t^{t+\Delta t}\right) \ , \\ \left[I_{(j_1,j_2)}\right]_t^{t+\Delta t} + \left[I_{(j_2,j_1)}\right]_t^{t+\Delta t} &= \Delta W_{j_1,t}\Delta W_{j_2,t} \ . \end{aligned} \tag{3.23}$$

3.3.2 Euler and Milstein Scheme (Vector Form)

The stochastic process (3.22) can be represented in vector form by:

$$\begin{aligned} dX_t = A_0\left(X_t,t\right)dt + \sum_{j=1}^{m} A_j\left(X_t,t\right)dW_{j,t} \ , \qquad X_{t_0} = X_0 \ , \\ A_0\left(X_t,t\right) = \mu\left(X_t,t\right) \ , \qquad A_j\left(X_t,t\right) = \sigma_{:,j}\left(X_t,t\right) \ . \end{aligned} \tag{3.24}$$

The 0.5 strong order Euler scheme for (3.24) with time step Δt using Vector Form [27] is:

$$\widehat{X}_{t+\Delta t} = \widehat{X}_t + A_0\Delta t + \sum_{j=1}^{m} A_j \Delta W_{j,t} \ .$$

The 1 strong order Milstein scheme for (3.24) with time step Δt using Vector Form [27] is:

$$\widehat{X}_{t+\Delta t} = \widehat{X}_t + A_0\Delta t + \sum_{j=1}^{m} A_j \Delta W_{j,t} + R_M \ ,$$

where if one uses the double Itô integrals, R_M is equal to:

$$R_M = \sum_{j_1,j_2=1}^{M} \left(\partial_{A_{j_2}} A_{j_1}\right)\left[I_{(j_1,j_2)}\right]_t^{t+\Delta t} \ ,$$

or using Lévy Areas, R_M is equal to:

$$\begin{aligned} R_M &= \frac{1}{2}\sum_{j_1,j_2=1}^{M}\left(\partial_{A_{j_2}} A_{j_1}\right)\left(\Delta W_{j_1,t}\Delta W_{j_2,t} - \widetilde{\delta}_{j_1,j_2}\Delta t\right) \\ &\quad + \frac{1}{2}\sum_{\substack{j_1,j_2=1 \\ j_1<j_2}}^{M}\left(\partial_{A_{j_2}} A_{j_1} - \partial_{A_{j_1}} A_{j_2}\right)\left[\underline{L}_{(j_1,j_2)}\right]_t^{t+\Delta t} \ . \end{aligned}$$

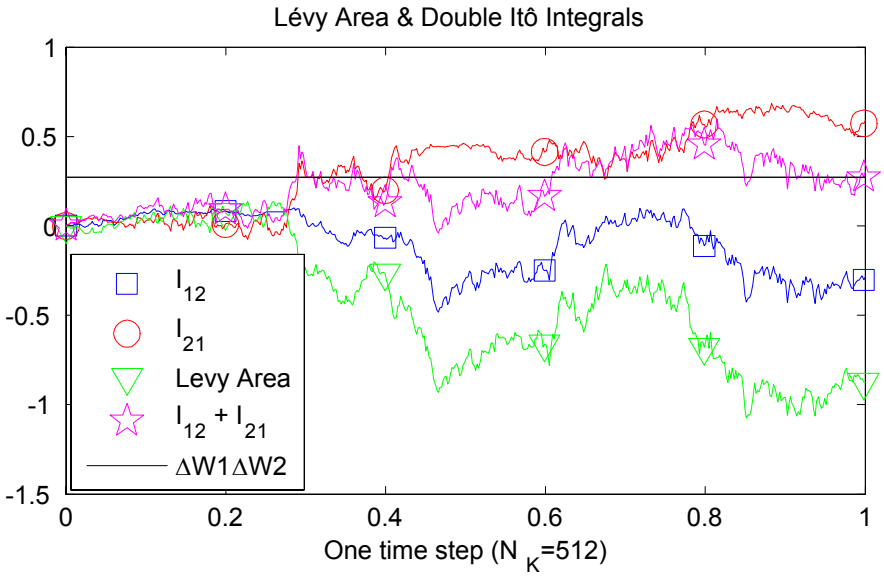

Figure 3.2: Simulation of the Lévy Area (3.26) and double Itô integrals (3.25).

The Jacobian matrix ∂_{A_j} is defined by:

$$\partial_{A_j} = \begin{bmatrix} F_1 b_{1,j} & F_2 b_{1,j} & \dots & F_M b_{1,j} \\ F_1 b_{2,j} & F_2 b_{2,j} & \dots & F_M b_{2,j} \\ \dots & \dots & \dots & \dots \\ F_1 b_{d,j} & F_2 b_{d,j} & \dots & F_M b_{d,j} \end{bmatrix} ,$$

$$F_k = \frac{\partial}{\partial X_k} .$$

The relationship between the Milstein scheme using Itô operators and Vector Form is:

$$L_{j_1} b_{i,j_2} = \left(\partial_{A_{j_2}} A_{j_1}\right)_i = \sum_{k=1}^{d} b_{k,j_1} F_k b_{i,j_2} = \sum_{k=1}^{d} b_{k,j_1} \frac{\partial b_{i,j_2}}{\partial X_k} .$$

3.4 Approximations of the Double Integral

In this section, we present some methods for simulating the double integrals in the Milstein approximation.

$$\left[I_{(1,2)}\right]_t^{t+\Delta t} = \int_t^{t+\Delta t} \int_t^s dW_{1,u} dW_{2,s} , \tag{3.25}$$

$$\left[I_{(2,1)}\right]_t^{t+\Delta t} = \int_t^{t+\Delta t} \int_t^s dW_{2,u} dW_{1,s} .$$

These double integrals, as one can see in Figure 3.2, start at zero, and then each one follows its own random path. At the end of the time step Δt, the sum of both integrals is $\Delta W_1 \Delta W_2$ and the difference between both is what is called the **Lévy Area.** As one has already seen, this is a very important concept in stochastic calculus:

$$\left[\overline{\underline{L}}_{(1,2)}\right]_t^{t+\Delta t} = \left[I_{(1,2)} - I_{(2,1)}\right]_t^{t+\Delta t} \ . \tag{3.26}$$

It is well known [24] that the double Itô integral has the following mean and variance:

$$E\left[I_{(1,2)}\right] = E\left[I_{(2,1)}\right] = \frac{\Delta W_{1,t} \Delta W_{2,t}}{2} \ ,$$
$$Var\left[I_{(1,2)}\right] = Var\left[I_{(2,1)}\right] = \frac{\Delta t}{12}\left(\Delta t + R^2\right) \ ,$$

where:

$$R^2 = \left(\Delta W_1\right)^2 + \left(\Delta W_2\right)^2 \ ,$$

and, for the Lévy Area:

$$E\left[\overline{\underline{L}}_{(1,2)}\right] = E\left[I_{(1,2)} - I_{(2,1)}\right] = 0 \ , \tag{3.27}$$
$$Var\left[\overline{\underline{L}}_{(1,2)}\right] = \frac{\Delta t}{3}\left(\Delta t + R^2\right) \ .$$

The numerical difficulty is how to calculate the double Itô integral $I_{(1,2)}$ or the Lévy Area $\overline{\underline{L}}_{(j_1,j_2)}$. The technique of Gaines and Lyons [8] can be used to sample the distribution for the Lévy Area conditional on $\Delta W_{1,t}$, $\Delta W_{2,t}$. However there is no generalization of this to higher dimensions apart from the approximation of [37], which has a significant computational cost.

In this section we shall present at the beginning the subdivision method proposed by Kloeden [23] to simulate the double integral (3.25). We follow with the problems of using this method and propose a solution. The big disadvantage with this method is that it takes a long time (computationally expensive) if one wants to obtain a good approximation. At the end of the section we present an explicit formula obtained by inverting the Fourier transformation of the cumulative distribution function of the Lévy Area (3.26). Unfortunately, this method is only valid for small values of $\triangle t$ [29].

3.4.1 Subdivision (Kloeden)

Kloeden in [23] says that the double integral (3.25) can be approximated by applying the Euler scheme to the following 2−Dimensional Itô SDE:

$$dX_{1,t_n} = X_{2,t_n} dW_{1,t} \ , \tag{3.28}$$
$$dX_{2,t_n} = dW_{2,t} \ ,$$

using the initial conditions:

$$\begin{aligned} X_{1,t_n} &= 0 \ , \\ X_{2,t_n} &= W_{2,t_n} \ . \end{aligned} \tag{3.29}$$

Over the discretization subinterval $[t, t + \Delta t]$ with a suitable size:

$$\delta t = \frac{\Delta t}{N_K} \ ,$$

the stochastic Euler scheme for (3.28) is:

$$\begin{aligned} Y_{1,k+1} &= Y_{1,k} + Y_{2,k}\delta W_{1,n,k} \ , \\ Y_{2,k+1} &= Y_{2,k} + \delta W_{2,n,k} \ , \\ Y_{1,0} &= 0 \ , \quad Y_{2,0} = W_{2,t_n} \ . \end{aligned} \tag{3.30}$$

When $k = N_K - 1$, one obtains the approximation of the double integral (3.25):

$$\begin{aligned} Y_{1,N_K} &\approx \left[I_{(2,1)}\right]_{t_n}^{t_{n+1}} \ , \\ Y_{2,N_K} &= W_{2,n} \ . \end{aligned}$$

The new Wiener processes $\delta W_{1,n,k}$ and $\delta W_{2,n,k}$ can be obtained using a Brownian Bridge (B.1) and have to be equal to:

$$dW_{j,t} = \textstyle\sum_{k=1}^{N_K} \delta W_{j,n,k} \ .$$

Kloeden in [23] says that the strong order of convergence for the stochastic Euler scheme (3.30) ensures that:

$$E\left(\left|Y_{1,N_K} - \left[I_{(2,1)}\right]_{t_n}^{t_{n+1}}\right|\right) \leq C\sqrt{\delta t} \ .$$

Therefore, $\left[I_{(2,1)}\right]_{t_n}^{t_{n+1}}$ can be approximated in the Milstein scheme by Y_{1,N_K} with $\delta t = (\Delta t)^2$ without affecting the overall order of convergence. Other higher order multiple stochastic integrals can be simulated in a similar way.

3.4.2 Subdivision (IC = 0)

If one simulates the two integrals using the subdivision method from Kloeden described above, Figure 3.3 shows that the expectation of the Lévy Area (3.27) is equal to zero only in the first time step ($n = 1$) and changes as n tends to infinity. However, changing the initial conditions (3.29) to zero, the expectation of the Lévy Area for all

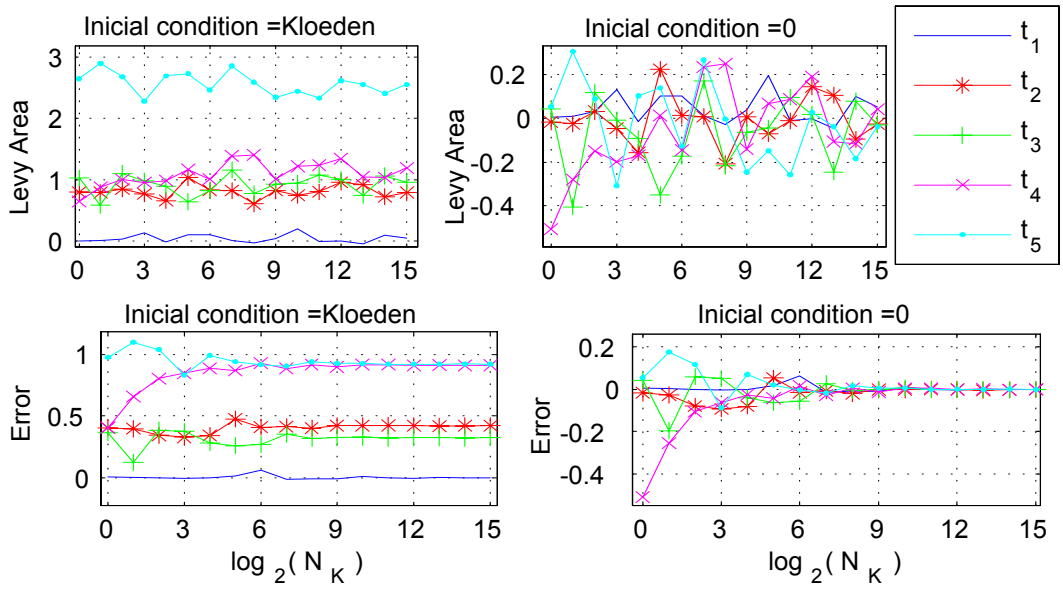

Figure 3.3: Comparison between the two subdivision methods ($n = 5$).

n time steps is equal to zero. Furthermore, simulating both methods with different integration steps N_K, and if the error is:

$$error = \Delta W_{1,n}\Delta W_{2,n} - \left(\left[I_{(1,2)} \right]_{t_n}^{t_{n+1}} + \left[I_{(2,1)} \right]_{t_n}^{t_{n+1}} \right) \ , \tag{3.31}$$

one has to see that this error tends to zero as N_K tends to infinity. This behavior is not true when the initial conditions are not equal to zero (Figure 3.3). We conclude that the double Itô integral (3.25) can be simulated using the subdivision method given by Kloeden by changing the initial conditions (3.29) to zero. The accuracy or error in the calculation of (3.28) depends directly on the value of N_K. For more information, see [29].

3.4.3 Subdivision (Lévy Area)

Another way to simulate the double integral is by using the properties (3.23) and simulating the Lévy Area:

$$\begin{aligned} dX_{1,t_n} &= dW_{1,t} \ , \quad dX_{2,t_n} = dW_{2,t} \ , \\ dX_{3,t_n} &= X_{2,t_n} dW_{1,t} - X_{1,t_n} dW_{2,t} \ . \end{aligned} \tag{3.32}$$

The stochastic Euler scheme for (3.32) is:

$$\begin{aligned} Y_{1,k+1} &= Y_{1,k} + \delta W_{1,n,k} \ , \quad Y_{2,k+1} = Y_{2,k} + \delta W_{2,n,k} \ , \\ Y_{3,k+1} &= Y_{3,k} + (Y_{2,k}\delta W_{2,n,k} - Y_{1,k}\delta W_{1,n,k}) \ , \end{aligned}$$

using the initial conditions:

$$Y_{i,k} = 0 \ .$$

When $k = N_K - 1$, one obtains the approximation of the Lévy Area (3.26):

$$Y_{3,N_K} \approx \left[\overline{\underline{L}}_{(1,2)}\right]_t^{t+\Delta t} \ .$$

Therefore, $\overline{\underline{L}}_{(1,2)}$ can be approximated in the Milstein scheme by Y_{3,N_K} with $\delta t = (\Delta t)^2$ without affecting the overall order of convergence.

3.4.4 Fourier Lévy formula

We start with the integrated form of the Lévy Area (3.26):

$$\overline{\underline{L}}_{(1,2)}(\Delta t) = \int_0^{\Delta t} \left(W_1(t)dW_2(t) - W_2(t)dW_1(t)\right) \ .$$

The Fourier transformation of the density of $\overline{\underline{L}}_{(1,2)}$[4] conditional on $\Delta W_1, \Delta W_2$ is given by:

$$\widehat{\overline{\underline{L}}}_{(1,2)}(w) = E\left[\exp\left(iw\overline{\underline{L}}_{(1,2)}(t)\right)|W_1(t) = \Delta W_1, W_2(t) = \Delta W_2\right] \ ,$$

and is explicitly known by [9] (it is also given in Lévy's original paper [24]) as:

$$\widehat{\overline{\underline{L}}}_{(1,2)}(w) = f_X(w)f_Y(w) \ ,$$

where, given $R^2 = (\Delta W_1)^2 + (\Delta W_2)^2$:

$$f_X(w) = \frac{w\Delta t}{\sinh(w\Delta t)} \ ,$$

$$f_Y(w) = \exp\left[-\frac{R^2}{2\Delta t}\left(w\Delta t \ \coth\left(w\Delta t\right) - 1\right)\right] \ .$$

The probability density function (*pdf*) for X can be obtained exactly by inverting the Fourier transform $f_X(w)$:

$$p(x) = \frac{\pi}{4\Delta t}\,\mathrm{sech}^2\left(\frac{\pi x}{2\Delta t}\right) \ ,$$

and then the cumulative distribution function (*cdf*) is:

$$P(x) = \int_{-\infty}^{x} p(x)dx = 1 - \frac{1}{1 + e^{\pi x/\Delta t}} \ ,$$

[4]The Fourier transformation of the density of the Lévy Area L is:

$$\widehat{p}\,(w) = \int_{-\infty}^{\infty} e^{iwL} p\,(L)\,dL \ , \quad p\,(L) = \frac{1}{2\pi}\int_{-\infty}^{\infty} e^{-iwL}\widehat{p}\,(w)\,dw \ .$$

leading, via its inverse, to the sample rule:

$$X = \frac{1}{P(Q)} = \frac{\Delta t}{\pi} \ln\left(\frac{Q}{1-Q}\right) \ , \qquad Q \sim U(0,1) \ .$$

Note that the variance of X is:

$$Var\left[X\right] = -\lim_{w\to 0}\left[\frac{d^2 f_X}{dw^2}\right] = \frac{\Delta t^2}{3} \ .$$

So far as we know, the *pdf* for Y cannot be written down in exact form, but for small Δt, we have:

$$f_Y(w) = \exp\left[-\frac{R^2}{2\Delta t}\left(w\Delta t \ \coth\left(w\Delta t\right) - 1\right)\right] = \exp\left[-\frac{R^2 w^2 \Delta t}{6} + O\left(\Delta t^3\right)\right] \ ,$$

$$f_Y(w) \sim \exp\left[-\frac{R^2 w^2 \Delta t}{6}\right] \ ,$$

which is the Fourier transform of another normal distribution with density:

$$p(y) = \frac{1}{\sqrt{2\pi\sigma}} \exp\left(-\frac{y^2}{2\sigma^2}\right) \ ,$$

$$\sigma^2 = \frac{R^2 \Delta t}{3} \ .$$

Samples of Y can then be made in the usual way.

$$Y = \sqrt{\frac{R^2 \Delta t}{3}} Z \ , \qquad Z \sim N(0,1) \ .$$

So the double integral (3.25) can be approximated using the formula:

$$\left[I_{(2,1)}\right]_{t_n}^{t_{n+1}} = \frac{1}{2}\left(\Delta W_{1,n} \Delta W_{2,n}\right) - \frac{1}{2}\left(X + Y\right) \ . \tag{3.33}$$

So, although (3.33) is an approximation, one can see that we recover the exact total variance required.

$$Var\left[\underline{\overline{L}}_{(1,2)}\right] = Var\left[X\right] + Var\left[Y\right] = \frac{\Delta t}{3}\left(\Delta t + R^2\right) \ .$$

Because the *pdf* for Y cannot be written down in exact form this approximation is only valid for small values of $\triangle t$. For more information, see [29].

3.5 Convergence

Convergence for numerical schemes can be defined in various ways. It is common to distinguish between strong and weak convergence, depending on whether the realizations or only their probability distributions are required to be close, respectively. In this section we shall define strong and weak convergence and how to measure it, even if we do not have an exact solution.

3.5.1 Strong Convergence

When we talk about strong convergence, we are referring to how fast our discrete time approximation converges to the exact solution as it is refined.

Definition: We shall state that a discrete time approximation $\widehat{X}(T)$ *converges strongly with order* $\gamma > 0$ at time T as $\triangle t \downarrow 0$ to the exact solution $X(T)$ if there exists a positive constant C, that does not depend on γ, such that:

$$\epsilon_{Strong}(\triangle t) \equiv E\left[\ \left|X(T) - \widehat{X}(T, \triangle t)\right|\ \right] \leq C\triangle t^{\gamma}\ . \tag{3.34}$$

Refer to the theorem in ([22], page 362), which proves that if (3.34) is true, it can be implied that the order of strong convergence is not only in the last point T, but also uniformly over all time steps $\triangle t$ within the whole time interval $t_n \in [t_0, T]$.

$$E\left[\sup_{t_0 \leq t_n \leq T}\left|X(t_n) - \widehat{X}(t_n, \triangle t)\right|\ \right] \leq C_1\triangle t^{\gamma}\ .$$

It should be noted that γ is the largest such power for (3.34). Note that using the method of least squares, one can calculate the constant C and the order of convergence γ for (3.34).

$$\log\left(\epsilon_{Strong}(\triangle t)\right) = \log(C) + \gamma\log(\triangle t)\ .$$

3.5.2 Weak Convergence

When we talk about weak convergence, we want to know how fast the expectation of our discrete time approximation converges to the exact expectation of our system.

Definition: We shall state that a discrete time approximation $\widehat{X}(T)$ *converges weakly with order* $\beta > 0$ at time T as $\triangle t \downarrow 0$ to the exact solution $X(T)$ if there exists a positive constant K, that does not depend on β, such that:

$$\epsilon_{Weak}(\triangle t, M) \equiv \left|\ E\left[g\left(X(T)\right)\right] - E\left[g\left(\widehat{X}(T, \triangle t)\right)\right]\ \right| \leq K\triangle t^{\beta}\ , \tag{3.35}$$

for any function $g \in C_p^{2(\beta+1)}\left(\mathbb{R}^d, \mathbb{R}\right)$. It should be noted that β is the largest such power for (3.35). Even though a realization of $\widehat{X}(T, \triangle t)$ is computable using a stochastic scheme, the expectation $E\left[g\left(\widehat{X}(T, \triangle t)\right)\right]$ is, in general, not. However, it can be approximated by a sample average of M independent realizations, which is the basis of Monte Carlo methods. The exact computational error, ϵ_{Weak}, naturally separates

into two parts:

$$\begin{aligned}
\epsilon_{Weak}(\triangle t, M) &\equiv \left| E\left[g\left(X(T)\right)\right] - \frac{1}{M}\sum_{j=1}^{M}\left[g\left(\widehat{X}_j(T,\triangle t)\right)\right]\right| \\
&\leq \left| E\left[g\left(X(T)\right)\right] - E\left[g\left(\widehat{X}(T,\triangle t)\right)\right]\right| \\
&\quad + \left| E\left[g\left(\widehat{X}(T,\triangle t)\right)\right] - \frac{1}{M}\sum_{j=1}^{M}\left[g\left(\widehat{X}_j(T,\triangle t)\right)\right]\right| \ , \\
\epsilon_{Weak}(\triangle t, M) &\leq \epsilon_{Weak}\left(\triangle t\right) + \epsilon_{Stat}(\triangle t, M) \ .
\end{aligned}$$

The time step $\triangle t$ determines the time discretization or weak error ϵ_{Weak}, and the number of paths or realizations M mainly determines the statistical error ϵ_{Stat}. If a scheme is strong to a certain order, it will be weak to at least that order, and possibly more, but not vice versa. Aspects of the use of Euler and Milstein schemes for the weak approximation of SDE's have been addressed before and they have shown that both schemes are 1.0 order of weak convergence:

$$\left| E\left[g\left(X(T)\right)\right] - E\left[g\left(\widehat{X}(T,\triangle t)\right)\right]\right| = O\left(\triangle t\right) \ .$$

A comprehensive review of the construction and the analysis of the strong and weak convergence order for higher order methods can be found in the inspiring book by Kloeden and Platen [22].

3.5.3 Convergence without an Exact Solution

If one applies any discrete approximation scheme to a stochastic process (3.36) and wants to numerically evaluate the strong or weak convergence order of our approximation $\widehat{X}(T)$, an exact solution $X(T)$ is normally required. However, at present, there are no solutions available for many SDEs. Instead, the next theorems allow us to determine the order of convergence for our discrete time approximation without an exact solution. We have published these results in [29] and [30].

All models can be described through a SDE of the form:

$$dX_t = \mu\left(X_t, t\right) dt + \sigma\left(X_t, t\right) dW_t \ , \qquad X(t_0) = X_0 \ , \tag{3.36}$$

where:

$$\begin{aligned}
X_t &= X(t) \in \mathbb{R}^d \ , \ W_t \in \mathbb{R}^M \ , \ t \in [t_0, ...T] \in \mathbb{R} \ , \\
\sigma\left(X_t, t\right) &= \sigma\left(b_{i,k}\left(X_t, t\right)\right) \in \mathbb{R}^{d \times M} \ , \quad \mu\left(X_t, t\right) = \mu\left(a_i\left(X_t, t\right)\right) \in \mathbb{R}^d \ ,
\end{aligned}$$

$$E\left[dW_{j,t}dW_{k,t}\right] = 0 \ , \qquad for \ j \neq k \ .$$

Theorem 1: Existence and Uniqueness of Strong Solutions

Suppose $E\left[||X_0||^2\right]$ is finite and that there is a constant K for which for all $t \in [t_0, T]$ and all $x, y \in \mathbb{R}^d$ the following condition is satisfied:

$$\|\mu(x,t) - \mu(y,t)\| + \|\sigma(x,t) - \sigma(y,t)\| \leq K\|x - y\| \ , \text{ (Global Lipschitz condition)}^5. \tag{3.37}$$

Then the SDE (3.36) admits a strong solution X and satisfies:

$$E\left[||X_t||^2\right] < \infty \ .$$

This solution is unique in the sense that if $\widehat{X}$ is also a solution, then:

$$P\left(X_t = \widehat{X}_t, \ \forall t \in [t_0, T]\right) = 1 \ .$$

Note that if the SDE (3.36) satisfies the global Lipschitz condition (3.37), then the linear growth bounds is satisfied:

$$\|\mu(x,t)\| + \|\sigma(x,t)\| \leq K(1 + \|x\|) \ .$$

Proofs and additional explanation can be found in [12] and [22].

Theorem 2: Strong Convergence Order without an Exact Solution

A) If a discrete approximation $\widehat{X}$ of (3.36) with time step Δt has strong convergence order γ, i.e. there exist a constant C_1 such that:

$$E\left[\ \left|X(T) - \widehat{X}(T, \triangle t)\right| \ \right] \leq C_1 \triangle t^{\gamma} \ . \tag{3.38}$$

Then, there exists a positive constant, C_2, such that:

$$E\left[\left|\widehat{X}(T, \triangle t) - \widehat{X}\left(T, \frac{\triangle t}{2}\right)\right| \ \right] \leq C_2 \triangle t^{\gamma} \ . \tag{3.39}$$

B) Conversely, if it is known that the discretization is strongly convergent and (3.39) holds for some positive constant C_2, then the strong convergence order is γ.

Proof A):

If (3.38) is true for all Δt, then:

$$E\left[\left|X(T) - \widehat{X}\left(T, \frac{\triangle t}{2}\right)\right|\right] \leq C_1 \left(\frac{\triangle t}{2}\right)^{\gamma} \ . \tag{3.40}$$

[5] where $\|\circ\|$ is the Euclidean norm defined by: $\|x\| := \sqrt{\sum_{i=1}^{n} x_i^2}$.

Using the triangle law ($|A-B| \leq |A| + |B|$) and adding (3.38) and (3.40), one gets:

$$E\left[\left|\widehat{X}(T,\triangle t) - \widehat{X}\left(T,\frac{\triangle t}{2}\right)\right|\right] \leq C_1\left(1+\left(\frac{1}{2}\right)^{\gamma}\right)\triangle t^{\gamma} = C_2\triangle t^{\gamma}\ .$$

Proof B):

Using the triangle law:

$$E\left[\ \left|X(T)-\widehat{X}(T,\triangle t)\right|\ \right] \leq E\left[\left|X(T)-\widehat{X}\left(T,\left(\frac{1}{2}\right)^{M}\triangle t\right)\right|\right] + \sum_{m=0}^{M-1} E\left[\left|\widehat{X}\left(T,\left(\frac{1}{2}\right)^{m+1}\triangle t\right)-\widehat{X}\left(T,\left(\frac{1}{2}\right)^{m}\triangle t\right)\right|\right]\ .$$

Due to strong convergence:

$$\lim_{M\to\infty} E\left[\left|X(T)-\widehat{X}\left(T,\left(\frac{1}{2}\right)^{M}\triangle t\right)\right|\right] = 0\ .$$

Hence, using (3.39):

$$\begin{aligned} E\left[\left|X(T)-\widehat{X}(T,\triangle t)\right|\right] &\leq \sum_{m=0}^{\infty} C_2\left(\frac{1}{2}\right)^{m\gamma}\triangle t^{\gamma} \\ &= \frac{C_2}{1-\left(\frac{1}{2}\right)^{\gamma}}\triangle t^{\gamma} = C_1\triangle t^{\gamma} \end{aligned} \qquad \square$$

Theorem 3: Weak Convergence Order without an Exact Solution

A) If a discrete approximation $\widehat{X}$ of (3.36) with time step Δt has weak convergence order β for some positive constant K_1, i.e.:

$$\left|\ E\left[g\left(X(T)\right)\right] - E\left[g\left(\widehat{X}(T,\triangle t)\right)\right]\right| \leq K_1\triangle t^{\beta}\ . \tag{3.41}$$

Then, there exists a positive constant, K_2, such that:

$$\left|\ E\left[g\left(\widehat{X}(T,\triangle t)\right)\right] - E\left[g\left(\widehat{X}\left(T,\frac{\triangle t}{2}\right)\right)\right]\ \right| \leq K_2\triangle t^{\beta}\ . \tag{3.42}$$

B) Conversely, if it is known that the discretization is weakly convergent and (3.42) holds for some positive constant K_2, then the weak convergence order is β.

Proof: The proof is very similar to Theorem 2. Additional explanation can be found in [29].

3.6 Examples and Simulations.

In this section, we present two financial examples where we measure the strong and weak convergence for Euler and Milstein schemes. The first example is a Portfolio with N assets that follows an exponential Brownian motion. To measure the order of convergence we use both Theorems 2 and 3 (3.38, 3.41) presented in the last section and the exact solutions of our system. We show numerically that both theorems convergence to the right solution (order of convergence). The second example is a European Option assuming that the asset price follows a mean reverting stochastic volatility model (SVM). Because there is not an exact solution for this SVM, Theorems 2 and 3 are used to obtain the order of convergence.

3.6.1 Example 1 (Portfolio with N assets)

If $\Pi(t)$ is the total value of a portfolio at time t that contains N_P assets $S_i(t)$ and they follow an Exponential Brownian Motion process (EBM), then the portfolio Π is described by:

$$\begin{aligned} dS_i(t) &= S_i(t)\left((r(t) - D_i(t))\,dt + \sigma_i(t)d\widehat{W}_i(t)\right)\ , \\ \Pi(t) &= \sum_{i=1}^{N_P} S_i(t)\ , \end{aligned} \tag{3.43}$$

where $r(t)$ and $D_i(t)$ are the interest rate and continuous dividend at time t for the asset S_i respectively, and $\widehat{W}_i(t)$ and $\widehat{W}_j(t)$ are N_P Wiener processes with correlation coefficient $\rho_{i,j}$. The exact expectation for our portfolio at time t is:

$$E\left[\ \Pi(t)\ \right] = \sum_{i=1}^{N_P} S_i(t_0)\exp\left(\ \int_{t_0}^{t}(r(s) - D_i(s))\,ds\right)\ , \tag{3.44}$$

and for every realization or simulation, the exact solution is:

$$\Pi^{(j)}(t) = \sum_{i=1}^{N_P} S_i(t_0)\exp\left(\int_{t_0}^{t}\left(r(s) - D_i(s) - \frac{1}{2}\sigma_i^2(s)\right)ds + \int_{t_0}^{t}\sigma_i(s)d\widehat{W}_i(s)\right)\ . \tag{3.45}$$

If the time $t \in [t_0, T]$ is subdivided into equal time steps N_{steps}:

$$\Delta t = \frac{(T - t_0)}{N_{steps}}\ ,$$

the first strong Taylor approximation of order 0.5 (Euler scheme) is:

$$\widehat{S}_i(t_n + \Delta t) = \widehat{S}_i(t_n)\left(1 + (r(t_n) - D_i(t_n))\,\Delta t + \sigma_i(t_n)\Delta\widehat{W}_i(t_n)\right)\ . \tag{3.46}$$

The second strong Taylor approximation of order 1.0 is usually called the already introduced Milstein scheme. The same results are obtained if the definition of the Milstein scheme is applied directly to each equation or if the scheme using the vector form of (3.43) is applied with independent noise.

$$\begin{aligned} \widehat{S}_i(t_n+\Delta t) &= \widehat{S}_i(t_n)\left(1+\left(r(t_n)-D_i(t_n)\right)\Delta t+\sigma_i(t_n)\Delta\widehat{W}_i(t_n)\right) \\ &\quad +\frac{1}{2}\widehat{S}_i(t_n)\sigma_i^2(t_n)\left(\left(\Delta\widehat{W}_i(t_n)\right)^2-\Delta t\right) . \end{aligned} \tag{3.47}$$

Consider the following parameters and initial conditions for our portfolio $\Pi\left(t_0\right)$:

$$N=4\ ,\ S_i(t_0)=1\ ,\ t_0=0.1\ ,\ T=.9\ , \tag{3.48}$$

$$\begin{aligned} r\left(t\right)-D\left(t\right) &= \left[\begin{array}{cccc} 0.02 & 0.04 & 0.06 & 0.08 \end{array}\right]^T\ , \\ \sigma\left(t\right) &= \left[\begin{array}{cccc} 0.1 & 0.2 & 0.3 & 0.4 \end{array}\right]^T\ , \end{aligned}$$

and the correlation matrix for our Wiener process:

$$\rho=\left[\begin{array}{cccc} 1 & \overline{\rho}_1 & 0 & 0 \\ \overline{\rho}_1 & 1 & \overline{\rho}_2 & 0 \\ 0 & \overline{\rho}_2 & 1 & \overline{\rho}_3 \\ 0 & 0 & \overline{\rho}_3 & 1 \end{array}\right]\ ,$$

$$\overline{\rho}_1=0.25\ ,\ \overline{\rho}_2=-0.5\ ,\ \overline{\rho}_3=0.75\ .$$

Using (3.48), the exact solution (3.45) of (3.43) and running enough simulations ($M=10^4$ paths) to calculate the order of the strong convergence (3.34), we obtain as expected in Table 3.1 and Figure 3.4, magnitude 0.5 and 1.0 strong orders of convergence with respect to $\triangle t$ for Euler and Milstein schemes respectively. On the other hand, if we use Theorem 2 (3.38), ignoring the existence of the exact solution (3.45) for our system (3.43) and compute the same simulations ($M=10^4$), we obtain as expected in Table 3.1 and Figure 3.5, the same strong order of convergence γ as if one uses the exact solution.

As one can see, the use of Theorem 2 gives a good estimate of the strong order of convergence γ_i for our system (3.49). The only difference between the results, the exact solution (Figure 3.4) and Theorem 3 (Figure 3.5), are the value of the constants C_1 and C_2. Nevertheless they are related as:

$$C_1=\frac{2C_2}{1+\left(\frac{1}{2}\right)^{\gamma}}\ .$$

Scheme	C_1	C_2	γ_1	γ_2
$Euler - S_1$	0.004	0.003	0.49	0.48
$Euler - S_2$	0.019	0.015	0.48	0.48
$Euler - S_3$	0.044	0.034	0.48	0.48
$Euler - S_4$	0.080	0.063	0.48	0.48
$Euler - \Pi$	0.116	0.092	**0.48**	**0.48**
$Milstein - S_1$	0.001	0.000	0.99	0.98
$Milstein - S_2$	0.005	0.003	0.97	0.95
$Milstein - S_3$	0.012	0.009	0.95	0.93
$Milstein - S_4$	0.022	0.018	0.93	0.92
$Milstein - \Pi$	0.032	0.025	**0.95**	**0.93**

Table 3.1: Order of strong convergence test of (3.43) using the exact solution (C_1,γ_1) and Theorem 2 (C_2,γ_2).

Using (3.48), the exact expectation (3.44) of (3.43) and running $M = 10^9$ Monte Carlo paths to calculate the order of the weak convergence (3.35), we obtain as expected in Table 3.2 and Figure 3.6, a 1.0 weak order of convergence for both the Euler and Milstein schemes with respect to $\triangle t$. On the other hand, if we use Theorem 3, ignoring the existence of the exact expectation (3.44) for our system (3.43) and compute enough simulations ($M = 10^8$), we obtain as expected in Table 3.2 and Figure 3.7, the same weak order of convergence β as if one uses the exact expectation. Even though we have used 10 times more Monte Carlo paths to calculate the weak convergence order without using Theorem 3, one can see in the results (Table 3.1 and Figure 3.6) that there were not enough simulations.

Scheme	K_1	K_2	β_1	β_2
$Euler - S_1$	X[6]	.0001	X	0.99
$Euler - S_2$	X	.0004	X	0.99
$Euler - S_3$	.0013	.0010	0.93	0.98
$Euler - S_4$	.0030	.0017	1.06	0.98
$Euler - \Pi$	.0049	.0033	**0.98**	**0.98**
$Milstein - S_1$	X	.0001	X	0.99
$Milstein - S_2$	X	.0004	X	0.99
$Milstein - S_3$	.0013	.0010	0.93	0.99
$Milstein - S_4$	.0030	.0018	1.06	0.99
$Milstein - \Pi$	.0049	.0033	**0.98**	**0.99**

Table 3.2: Order of weak convergence test for (3.43) using the exact solution (K_1,β_1) and Theorem 3 (K_2,β_2).

[6]The simulation requires more Monte Carlo paths to correctly calculate the constant.

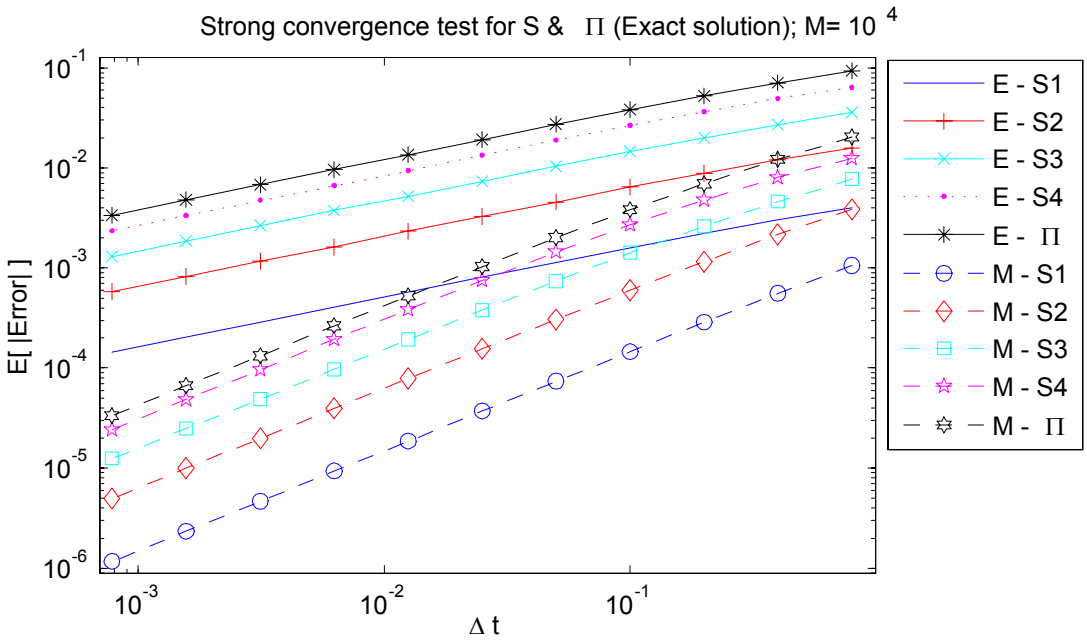

Figure 3.4: Strong convergence test of (3.43) using the exact solution (3.45).

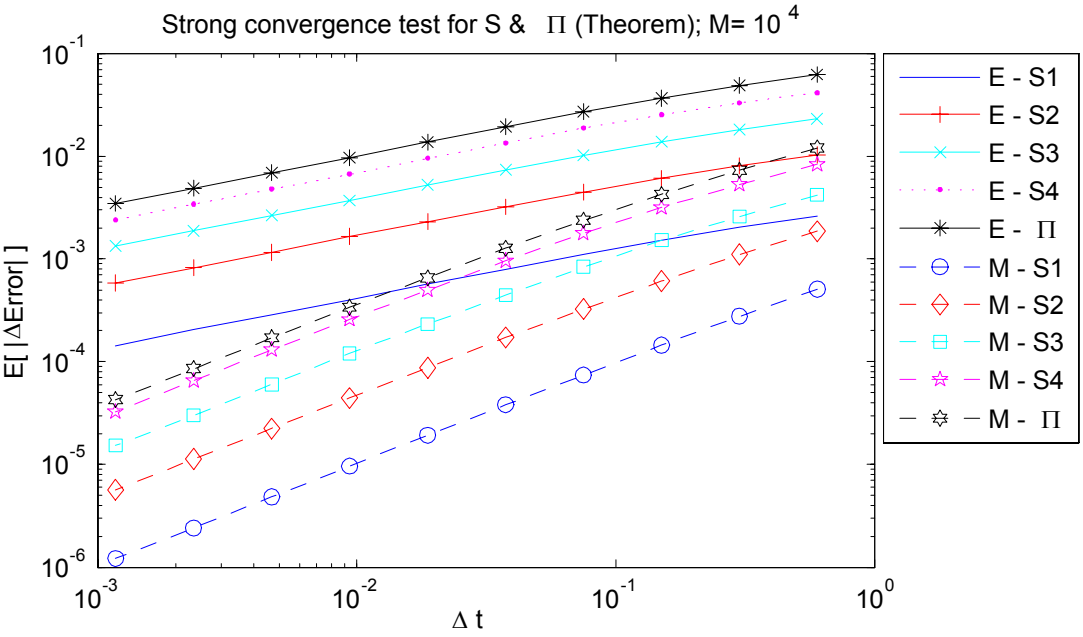

Figure 3.5: Strong convergence test of (3.43) using Theorem 2 (3.38).

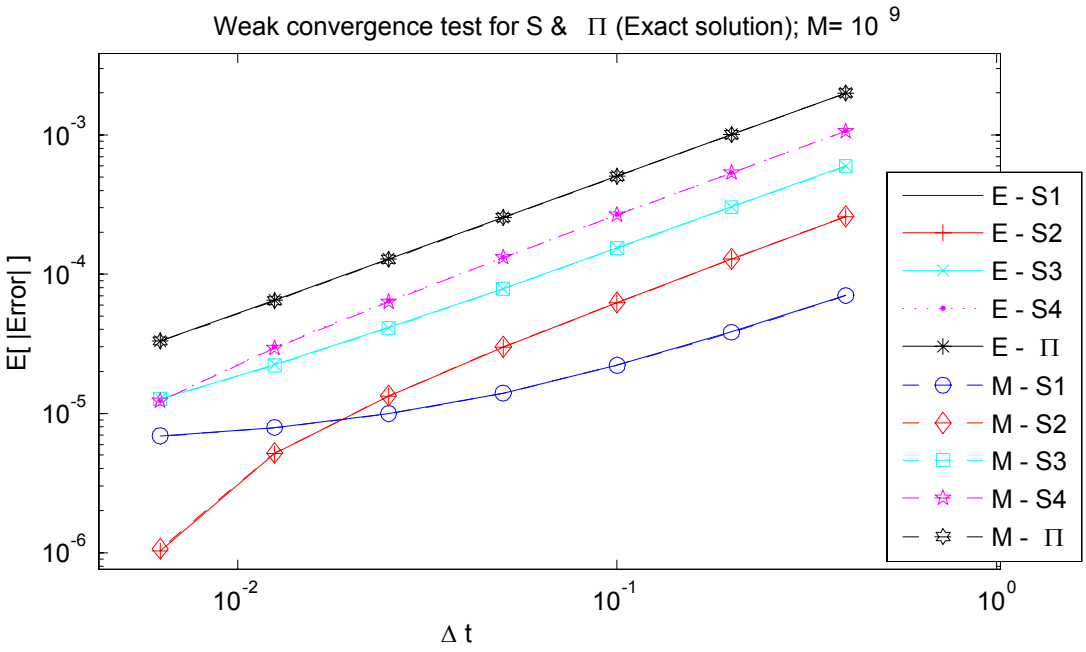

Figure 3.6: Weak convergence test of (3.43) using the exact expectation (3.44).

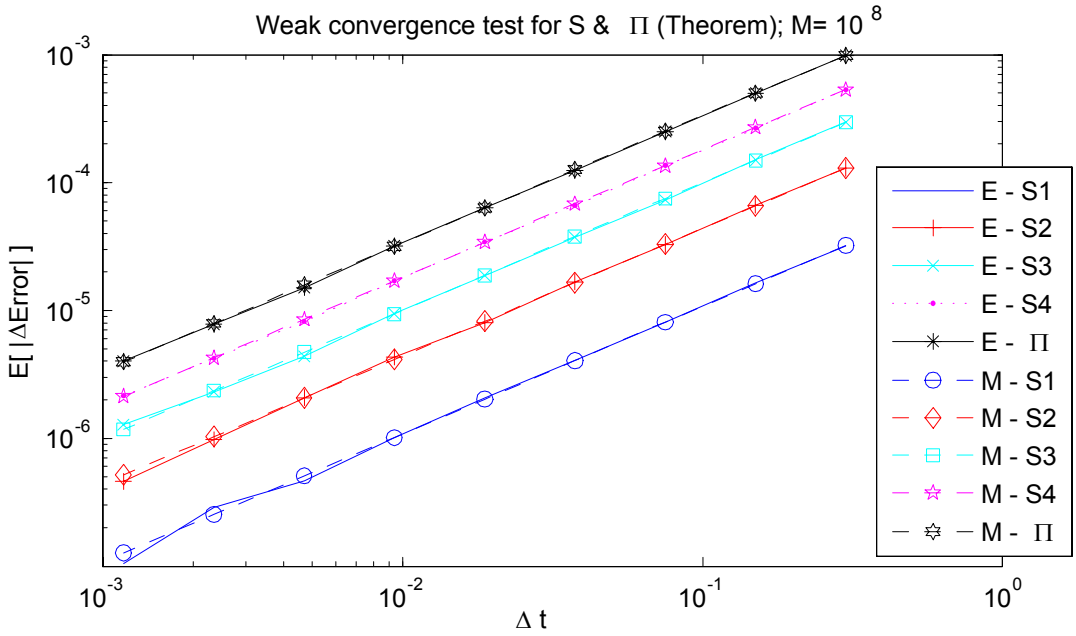

Figure 3.7: Weak convergence test of (3.43) using Theorem 3 (3.41).

As one can see, the use of the Euler and Milstein schemes give almost the same weak order of convergence β. The reason for this is that one is calculating the expectation instead of the corresponding path or realization. When calculating the expectation, the use of either the Euler or Milstein schemes has negligible difference on the outcome. The only difference is, when one uses Theorem 3 (Figure 3.7), Milstein scheme requires less paths in the Monte Carlo integration ($M = 10^5$ paths) than Euler scheme ($M = 10^8$ paths) to calculate the expectation (3.42). This is because of its lower variance (see Figure 3.13).

The use of Theorem 3 gives a good estimate of the weak order of convergence β for our system (3.43). The only difference between the results, the exact expectation (Figure 3.6) and Theorem 3 (Figure 3.7), are the value of the constants K_1 and K_2. However they are related as:

$$K_1 = \frac{K_2}{1 - \left(\frac{1}{2}\right)^{\beta}} \ .$$

It should be noted that if one uses Theorem 3 to calculate the weak order of convergence, then the subroutine takes much less time. This is because instead of calculating the expectation of the solution of our system (3.44), one is computing the expectation of the difference between two solutions with different time steps. In consequence, one needs less paths in the Monte Carlo integration. Also it should be noted that the standard Monte Carlo method used to calculate the expectations is very slow (computationally expensive). For this example alone, it took 495 hours for Figure 3.6 comparing with 49.5 hours for Figure 3.7 and 0.11 hours for Figure 3.13 (approximately, 48% for Euler and 52% for Milstein scheme).

3.6.2 Example 2 (European Options using Stochastic Volatility Models)

We begin with the usual Exponential Brownian Motion where the volatility σ is written as the square root of a variance ν and is assumed to follow a mean reverting SDE:

$$\begin{aligned} dS_t &= S_t\left(\mu_t dt + \sqrt{\nu_t} d\widehat{W}_{1,t}\right) \ , \qquad E\left[d\widehat{W}_{1,t}, d\widehat{W}_{2,t}\right] = \rho dt \ , \\ d\nu_t &= \nu_t^{\lambda_1}\left(k(\varpi - \nu_t) - \Lambda\nu_t\right) dt + \xi\nu_t^{\lambda_2} d\widehat{W}_{2,t} \ , \end{aligned} \tag{3.49}$$

where μ is the instantaneous drift or return of the stock price at time t and the seven parameters, κ, ϖ, Λ, ξ, λ_1, λ_2, ρ are constants and determine the evolution of the

asset price S and variance ν. They are defined as:

S_{t_0} = initial price	ν_{t_0} = initial volatility
t_0 = initial time	T = maturity
μ = drift	ϖ = long-run mean
κ = mean-reverting speed	ξ = volatility of volatility
Λ = market price of risk	ρ = correlation coefficient
λ_1 = random mean parameter	λ_2 = radial distance from OU

.

The value of a European Option at time T with strike K is equal to:

$$V(T) = \left\{ \begin{array}{ll} \max(S_T - K, 0) & \text{for call options} \\ \max(K - S_T, 0) & \text{for put options} \end{array} \right\} . \tag{3.50}$$

Unfortunately, there is no exact solution and expectation in the literature for the option value (3.50) using the SVM (3.49). However the expectation of S is:

$$E\left[\ S_T\ \right] = S_{t_0} \exp\left(\int_{t_0}^{T} \mu(s) ds \right) . \tag{3.51}$$

The first strong Taylor approximation of order 0.5, Euler scheme, is:

$$\begin{aligned} \widehat{S}_{t+\Delta t} &= \widehat{S}_t \left(1 + \mu_t \Delta t + \sqrt{\widehat{\nu}_t} \Delta \widehat{W}_{1,t}\right) , \\ \widehat{\nu}_{t+\Delta t} &= \widehat{\nu}_t + \widehat{\nu}_t^{\lambda_1} \left(k(\varpi - \widehat{\nu}_t) - \Lambda \widehat{\nu}_t\right) \Delta t + \xi \widehat{\nu}_t^{\lambda_2} \Delta \widehat{W}_{2,t} . \end{aligned} \tag{3.52}$$

If one applies Milstein scheme to each equation of (3.49), ignoring stochastic variation ν_t in S_t equation, one obtains (Milstein $1D$):

$$\begin{aligned} \widehat{S}_{t+\Delta t} &= \widehat{S}_t \left(1 + \mu_t \Delta t + \sqrt{\widehat{\nu}_t} \Delta \widehat{W}_{1,t} + \frac{1}{2} \widehat{\nu}_t \left(\Delta \widehat{W}_{1,t}^2 - \Delta t\right)\right) , \\ \widehat{\nu}_{t+\Delta t} &= \widehat{\nu}_t + \widehat{\nu}_t^{\lambda_1} \left(k(\varpi - \widehat{\nu}_t) - \Lambda \widehat{\nu}_t\right) \Delta t + \xi \widehat{\nu}_t^{\lambda_2} \Delta \widehat{W}_{2,t} + \frac{1}{2} \lambda_2 \xi^2 \widehat{\nu}_t^{2\lambda_2 - 1} \left(\Delta \widehat{W}_{2,t}^2 - \Delta t\right) . \end{aligned} \tag{3.53}$$

If one applies Milstein scheme using both the vector form of (3.49) with independent noise and the double Itô integrals (Milstein $2D - I$):

$$\begin{aligned} \widehat{S}_{t+\Delta t} &= \widehat{S}_t \left(1 + \mu_t \Delta t + \sqrt{\widehat{\nu}_t} \Delta \widehat{W}_{1,t} + \frac{1}{2} \hat{\rho} \xi \widehat{\nu}_t^{\lambda_2 - \frac{1}{2}} \left[I_{(2,1)}\right]_t^{t+\Delta t}\right) \\ &\quad + \widehat{S}_t \left(\frac{1}{2} \widehat{\nu}_t + \frac{1}{4} \rho \xi \widehat{\nu}_t^{\lambda_2 - \frac{1}{2}}\right) \left(\Delta \widehat{W}_{1,t}^2 - \Delta t\right) , \end{aligned} \tag{3.54}$$

$$\widehat{\nu}_{t+\Delta t} = \widehat{\nu}_t + \widehat{\nu}_t^{\lambda_1} \left(k(\varpi - \widehat{\nu}_t) - \Lambda \widehat{\nu}_t\right) \Delta t + \xi \widehat{\nu}_t^{\lambda_2} \Delta \widehat{W}_{2,t} + \frac{1}{2} \lambda_2 \xi^2 \widehat{\nu}_t^{2\lambda_2 - 1} \left(\Delta \widehat{W}_{2,t}^2 - \Delta t\right) ,$$

$$\left[I_{(2,1)}\right]_t^{t+\Delta t} = \int_{t_n}^{t_{n+1}} \int_{t_n}^{s_1} dW_{2,s_2} dW_{1,s_1} . \tag{3.55}$$

If one applies the concept of the Lévy Area (Milstein $2D - L$), one only needs to substitute (3.56) in (3.54) by:

$$\left[I_{(2,1)}\right]_t^{t+\Delta t} = \frac{1}{2}\left(dW_{1,t}dW_{2,t} - \left[\overline{L}_{(1,2)}\right]_t^{t+\Delta t}\right) \ . \tag{3.56}$$

One can note that when one applies the Milstein scheme to each equation or apply the scheme with independent noise and the vector form of (3.49), one obtains different equations. The double integral (3.55) and the Lévy Area (3.56) can be calculated using the methods proposed in the section below (page 38). They are the key in solving the Milstein scheme for the stochastic volatility model (3.49).

Consider the following parameters and initial conditions for our portfolio $\Pi\left(t_0\right)$:

$$S_0 = 100,\ \nu_0 = 0.2^2,\ t_0 = 0.1,\ T = 0.6,\ \mu = 0.05,\ \rho = -0.5, \tag{3.57}$$

$$\kappa = 1.5,\ \varpi = 0.15^2,\ \Lambda = 0,\ \xi = 0.2,\ \lambda_1 = 0,\ \lambda_2 = 0.7,$$

$$K_{Call} = K_{Put} = 105 \ .$$

Using Theorem 2 from page 46 (because there does not exist an exact solution for the system (3.49)), the parameters and initial conditions (3.57) and running enough simulations ($M = 5000$ paths) to calculate the order of strong convergence (3.34), we obtain as expected in Table 3.3 and Figures 3.8 and 3.9, magnitude 0.5 and 1.0 strong orders of convergence with respect to $\triangle t$ for Euler and Milstein schemes. As one can see, the Milstein $1D$ scheme (3.53) has the same order of convergence 0.5 as the Euler scheme (3.52) without having to calculate the double integral (3.55). The reason for this is that the scheme was applied without taking into account the correlation between the two systems. To obtain a 1.0 order of convergence with the Milstein scheme (3.54 or 3.56), one needs to apply the scheme to the vector form of (3.49), use independent Wiener processes and compute correctly the double integral or Lévy Area.

Scheme	C_3	C_4	C_5	C_6	γ_3	γ_4	γ_5	γ_6
Euler	3.46	.006	1.85	2.05	0.51	**0.58**	0.51	0.51
Milstein $1D$	3.44	.011	2.08	1.84	0.50	1.07	**0.50**	**0.49**
Milstein $2D$ $(I{=}0)$	3.10	.011	1.77	1.73	0.50	1.07	0.50	0.50
Milstein $2D$ $(L{=}0)$	2.52	.011	1.43	1.40	0.53	1.07	0.52	0.52
Milstein $2D$	5.16	.011	3.00	2.87	**0.99**	1.07	**0.99**	**0.98**

Table 3.3: Order of strong convergence test for S (C_3,γ_3), for ν (C_4,γ_4), for Call option (C_5,γ_5) and for Put option (C_6,γ_6) using Theorem 2 (3.38) and (3.49).

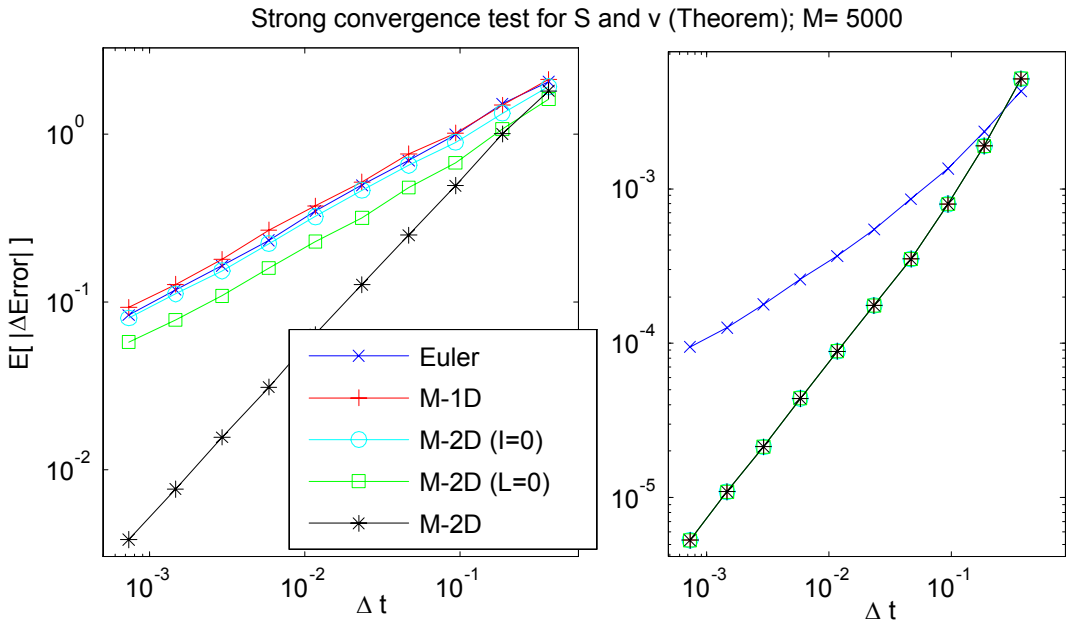

Figure 3.8: Strong convergence test for the SVM (3.49) using Theorem 2 (3.38).

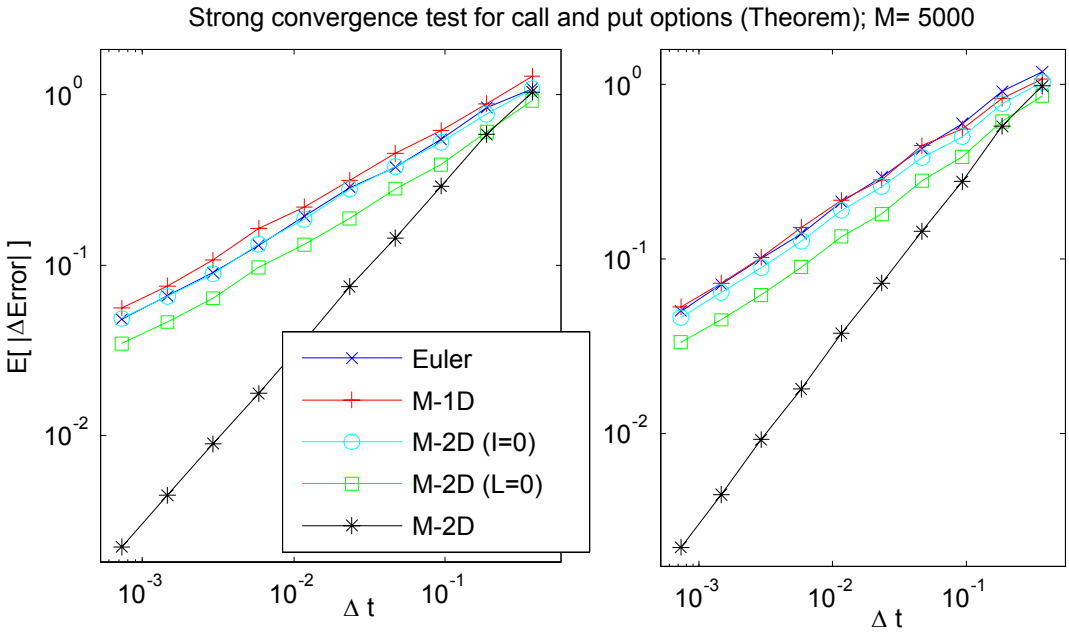

Figure 3.9: Strong convergence test for the option value (3.50) using Theorem 2.

To measure the weak convergence for the stochastic volatility model (3.49), one only needs the expectation of $S(t)$ in order to compare it with the expectation of our approximation $\widehat{S}(t)$. Running simulations using a sufficient number of paths ($M = 7 \times 10^8$) to approximate the expectation (3.51), we obtain as expected in Figure 3.10, a 1.0 weak order of convergence for all schemes with respect to $\triangle t$. Doing the same but now using Theorem 3, we obtain as expected in Figure 3.11, the same order of weak convergence as if one uses the exact expectation. We have excluded in the simulation the scheme that include the calculation of the Lévy Area (3.56) because of the high computational cost required to approximate correctly the expectation of S. Because we do not have an exact expectation of the option price (3.50), we have used Theorem 3 to calculate the weak order of convergence (Figure (3.12) and Table 3.4).

Scheme	K_3	K_4	K_5	K_6	β_3	β_4	β_5	β_6
Euler	.07	.042	.006	.72	$1.15^{(5)}$	0.99	1.08	1.08
Milstein $1D$	.07	.042	.006	.70	X^7	0.99	1.08	1.08
Milstein $2D$ $(I = 0)$	.07	.042	.006	.69	$1.15^{(5)}$	0.99	1.08	1.08
Milstein $2D$ $(L = 0)$	.07	.041	.006	.69	$1.14^{(5)}$	0.99	1.08	1.08

Table 3.4: Weak convergence test of S (K_3,β_3, exact solution), for S (K_4,β_4, Theorem 3). for v (K_5,β_5), for the Call option (K_6,β_6).

Again, when the expectation is calculated using the Euler or Milstein schemes there is negligible difference in the outcomes. The Monte Carlo method used to calculate the expectation was, again, very slow. For this example alone, it took 510 hours for Figures 3.10 and 3.11 (24% for Euler, 25% for Milstein $1D$, 25% for Milstein $2D-I$ and 26% Milstein $2D-L$). Even though we have used the same Monte Carlo paths to calculate the weak convergence, one can see in the results (Table 3.4 and Figure 3.10) that there were not enough simulations to calculate the convergence order correctly when one does not use Theorem 3. Comparing with Example 1 (Figure 3.13), the Euler or Milstein schemes does not have any difference in the number of paths required for Monte Carlo integration to calculate the expectation (3.51). This is because of its stochastic variance (Figure 3.14).

[7] The simulation requires more Monte Carlo paths to correctly calculate the constant.

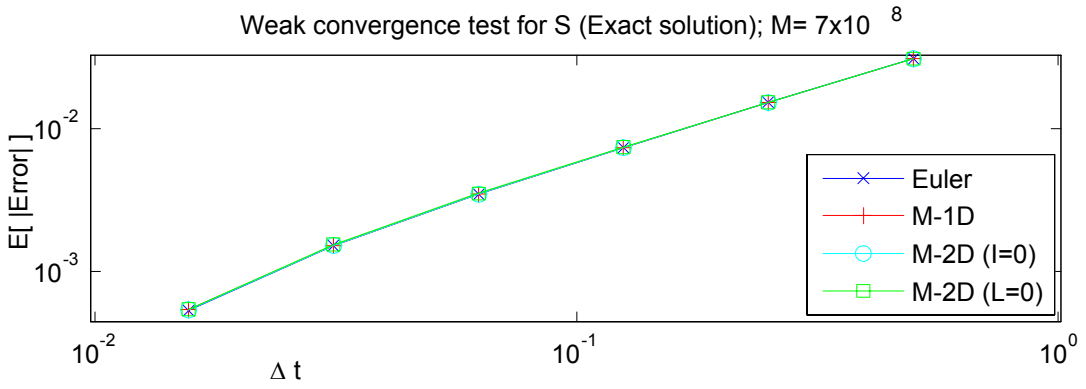

Figure 3.10: Weak convergence test of (3.49) using the exact expectation (3.51).

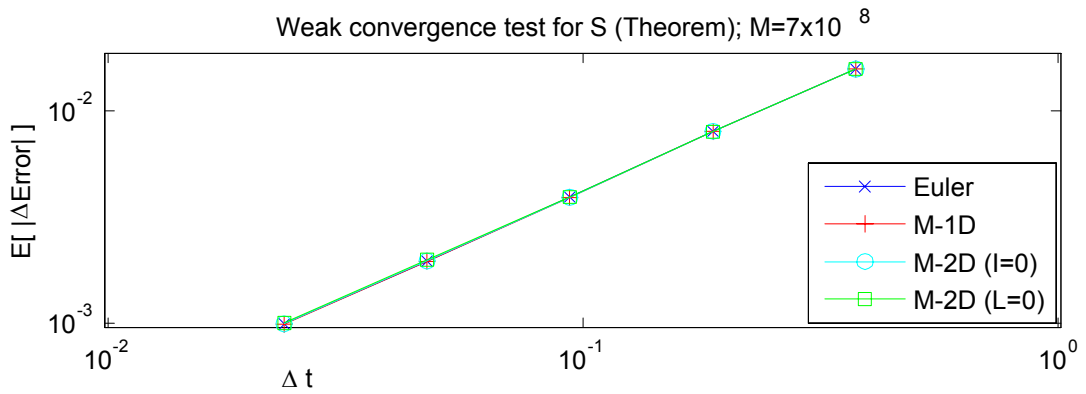

Figure 3.11: Weak convergence test of (3.49) using Theorem 3 (3.41).

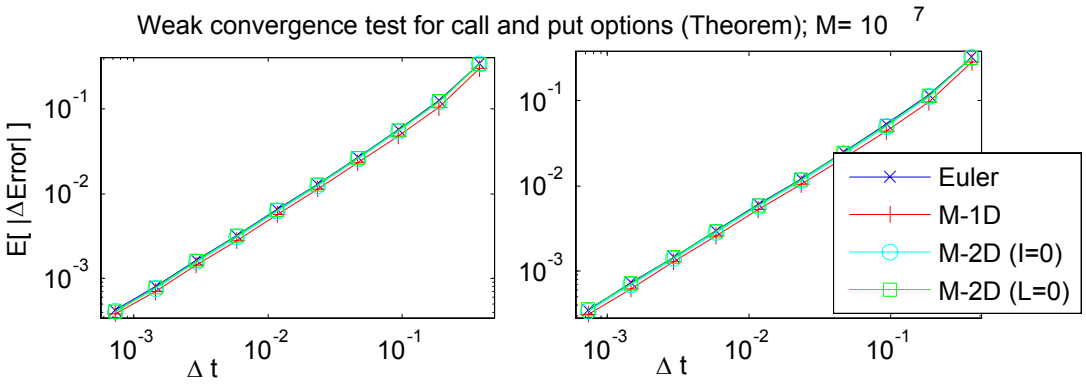

Figure 3.12: Weak convergence test for European options using Theorem 3 (3.41).

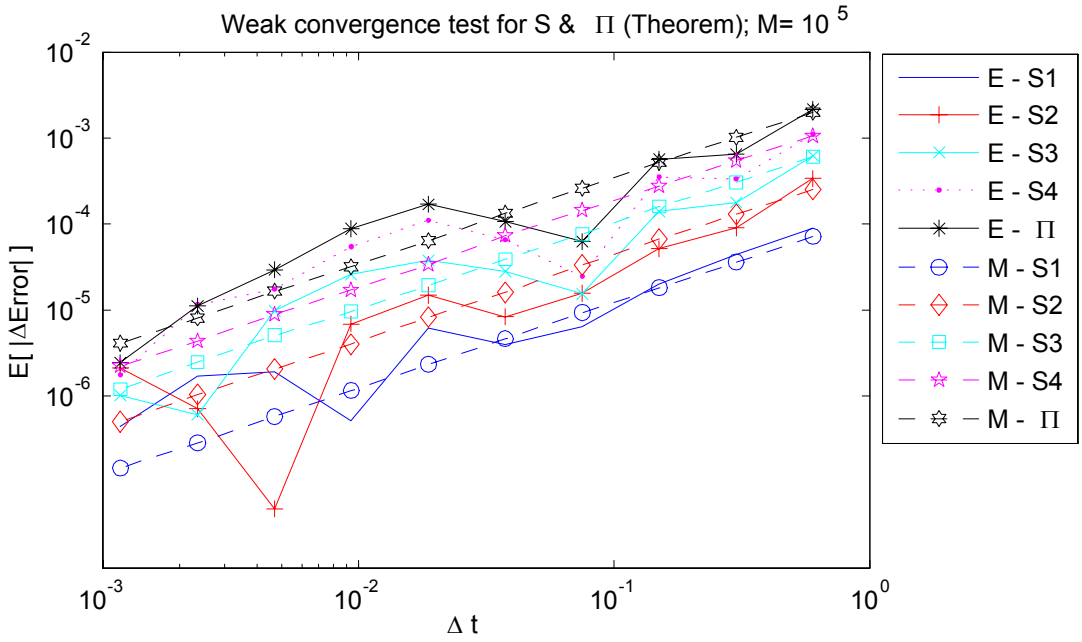

Figure 3.13: Weak convergence test of (3.43) using Theorem 3 ($MC = 10^5$).

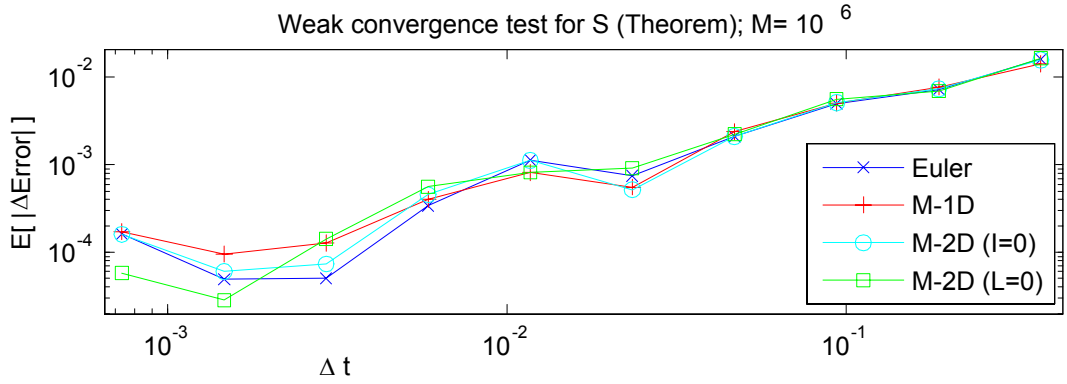

Figure 3.14: Weak convergence test of (3.49) using Theorem 3 ($MC = 10^6$).

3.7 Conclusions

For the $N-$Dimensional Exponential Brownian Motion process (Portfolio with N assets (3.43)) we obtain, as expected, a 0.5 and 1.0 strong order of convergence for Euler (3.46) and Milstein (3.47) schemes respectively. However, for the stochastic volatility model (3.49), we obtain a 0.5 strong order of convergence for both Euler (3.52) and Milstein $1D$ (3.53) schemes. The reason for this is that the Milstein $1D$ scheme was applied without taking into account the correlation between the two systems. To obtain a 1.0 strong order of convergence with the Milstein scheme (3.54) or (3.56), one has to apply the scheme to the vector form of (3.49), use independent Wiener processes and compute correctly the double integral (3.55) or Lévy Area (3.56).

Both the Exponential Brownian Motion (3.43) and the stochastic volatility model (3.49), give a 1.0 weak order of convergence and almost the same constant C for all schemes with respect to $\triangle t$ (the same expectation error). The application of either the Euler or Milstein schemes to calculate an expectation in the standard way has negligible difference at all in the outcome.

The use of Theorems 2 and 3 successfully determines the strong and weak orders of convergence. Each theorem was tested using exact solutions or expectations to verify the results. Examples in the chapter demonstrate that the use of both theorems require at least 100 times fewer Monte Carlo paths than the standard method to correctly calculate the order of convergence. Because there are no exact solutions for the stochastic volatility model (3.49), the use of the theorems was fundamental to establish its convergence order.

Chapter 4

θ Scheme (Orthogonal Milstein Scheme)

Strong convergence properties of discretizations of stochastic differential equations (SDEs) are very important in stochastic calculus. The Milstein scheme gives first order strong convergence for all $1-$dimensional systems (one Wiener process). However, for two or more Wiener processes, such as correlated portfolios and stochastic volatility models, there is no exact solution for the iterated integrals of second order (Lévy area) and the Milstein scheme neglecting the Lévy area usually gives the same order of convergence as the Euler-Maruyama scheme.

The purpose of the chapter is to show that if certain conditions are satisfied, one can avoid the calculation of the Lévy area and obtain first convergence order by applying an orthogonal transformation. We demonstrate when the conditions of the $2-$Dimensional problem permit this and give an exact solution for the orthogonal transformation (θ Scheme or Orthogonal Milstein Scheme).

4.1 Orthogonal Transformation $2D$

We begin with a $2-$Dimensional Itô stochastic differential equation (SDE) with a $2-$Dimensional Wiener process:

$$\begin{aligned} dx &= \mu^{(x)}(x,y,t)\,dt + \sigma(x,y,t)\,d\widehat{W}_{1,t}\ , \\ dy &= \mu^{(y)}(x,y,t)\,dt + \xi(x,y,t)\,d\widehat{W}_{2,t}\ , \qquad E\left[d\widehat{W}_{1,t}, d\widehat{W}_{2,t}\right] = \rho\,dt\ . \end{aligned} \tag{4.1}$$

Alternatively, in vector form:

$$dZ(t) = A_0\left(t,Z\right)dt + \sum_{k=1}^{2} A_k\left(t,Z\right)d\widehat{W}_{k,t}\ , \qquad Z \in \mathbb{R}^2\ .$$

This is in fact, only a symbolic representation for the stochastic integral equation:

$$Z(t) = Z(t_0) + \int_{t_0}^{t} A_0(s, Z)\, ds + \sum_{k=1}^{2} \int_{t_0}^{t} A_k(s, Z)\, d\widehat{W}_{k,s} \ .$$

The first integral is a deterministic Riemann integral and the second is a stochastic integral. Using the standard definition of constant correlation:

$$\left[\begin{array}{c} d\widehat{W}_{1,t} \\ d\widehat{W}_{2,t} \end{array} \right]_{Std} = \left[\begin{array}{cc} 1 & 0 \\ \rho & \widehat{\rho} \end{array} \right] \left[\begin{array}{c} dW_{1,t} \\ dW_{2,t} \end{array} \right] \ , \qquad \widehat{\rho} = \sqrt{1-\rho^2} \ , \tag{4.2}$$

one can represent the system (4.1) in vector form with independent noise as:

$$d \left[\begin{array}{c} x \\ y \end{array} \right] = \left[\begin{array}{c} \mu^{(x)} \\ \mu^{(y)} \end{array} \right] dt + \left[\begin{array}{c} \sigma \\ \rho\, \xi \end{array} \right] dW_{1,t} + \left[\begin{array}{c} 0 \\ \widehat{\rho}\, \xi \end{array} \right] dW_{2,t} \ , \tag{4.3}$$

$$\langle dW_{1,t}, dW_{2,t} \rangle = 0 \ .$$

The 1 strong order Milstein scheme for (4.3) with time step Δt is (Appendix (B.7)):

$$\left[\begin{array}{c} x_{t+\Delta t} \\ y_{t+\Delta t} \end{array} \right] = \left[\begin{array}{c} x_t \\ y_t \end{array} \right] + \left[\begin{array}{c} \mu^{(x_t)} \\ \mu^{(y_t)} \end{array} \right] \Delta t + \left[\begin{array}{c} \sigma \\ \rho\xi \end{array} \right] \Delta W_{1,t} + \left[\begin{array}{c} 0 \\ \widehat{\rho}\xi \end{array} \right] \Delta W_{2,t} \tag{4.4}$$

$$+\frac{1}{2} \left[\begin{array}{c} \sigma\sigma_x + \rho\xi\sigma_y \\ \rho\sigma\xi_x + \rho^2\xi\xi_y \end{array} \right] \left(\Delta W_{1,t}^2 - \Delta t \right) + \frac{1}{2} \left[\begin{array}{c} 0 \\ \widehat{\rho}^2\xi\xi_y \end{array} \right] \left(\Delta W_{2,t}^2 - \Delta t \right)$$

$$+\frac{1}{2} \left[\begin{array}{c} \widehat{\rho}\xi\sigma_y \\ \widehat{\rho}\sigma\xi_x + 2\rho\widehat{\rho}\xi\xi_y \end{array} \right] \left(\Delta W_{1,t} \Delta W_{2,t} \right) + \frac{1}{2} \left[A_1, A_2 \right] \left[\underline{\overline{L}}_{(1,2)} \right]_t^{t+\Delta t} \ ,$$

where subscript x and y denote partial derivatives, $\underline{\overline{L}}_{(1,2)}$ is the Lévy area defined by:

$$\left[\underline{\overline{L}}_{(1,2)} \right]_t^{t+\Delta t} = \int_t^{t+\Delta t} \int_t^S dW_{1,U} dW_{2,S} - \int_t^{t+\Delta t} \int_t^S dW_{2,U} dW_{1,S} \ .$$

$[A_1, A_2]$ is the Lie bracket defined by (∂_{A_i} is the Jacobian matrix of A_i):

$$[A_1, A_2] = (\partial_{A_2} A_1 - \partial_{A_1} A_2) = \left[\begin{array}{c} -\widehat{\rho}\xi\sigma_y \\ \widehat{\rho}\sigma\xi_x \end{array} \right] \ . \tag{4.5}$$

The numerical difficulty with the Milstein scheme is how to simulate efficiently the Lévy area $\underline{\overline{L}}_{(1,2)}$ (computationally very expensive). On the other hand, if one makes an orthogonal transformation of the uncorrelated process (4.3), one does not change the distribution (see Theorem 6 (B.23)) and gets:

$$\begin{aligned} d\widetilde{x} &= \mu^{(\widetilde{x})}(\widetilde{x}, \widetilde{y}, t)\, dt + \sigma(\widetilde{x}, \widetilde{y}, t)\, d\widetilde{\widehat{W}}_{1,t} \ , \\ d\widetilde{y} &= \mu^{(\widetilde{y})}(\widetilde{x}, \widetilde{y}, t)\, dt + \xi(\widetilde{x}, \widetilde{y}, t)\, d\widetilde{\widehat{W}}_{2,t} \ , \end{aligned} \tag{4.6}$$

where:

$$\begin{bmatrix} \widetilde{\widehat{dW}}_{1,t} \\ \widetilde{\widehat{dW}}_{2,t} \end{bmatrix} = \begin{bmatrix} 1 & 0 \\ \rho & \widehat{\rho} \end{bmatrix} \begin{bmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{bmatrix} \begin{bmatrix} dW_{1,t} \\ dW_{2,t} \end{bmatrix} . \tag{4.7}$$

If one computes the coefficients of the Lévy area (Lie bracket) for the new orthogonal process using independent Brownian paths $W_{1,t}$,$W_{2,t}$:

$$[A_1, A_2] = \begin{bmatrix} -\widehat{\rho}\xi\sigma_{\widetilde{y}} - \sigma^2\theta_{\widetilde{x}} - \rho\sigma\xi\theta_{\widetilde{y}} \\ \widehat{\rho}\sigma\xi_{\widetilde{x}} - \rho\sigma\xi\theta_{\widetilde{x}} - \xi^2\theta_{\widetilde{y}} \end{bmatrix} .$$

To avoid having to simulate the Lévy area $\overline{\underline{L}}_{(1,2)}$, one needs the Lie brackets to be identically zero, i.e., you need to impose the following conditions:

$$\begin{aligned} -\widehat{\rho}\xi\sigma_{\widetilde{y}} - \sigma^2\theta_{\widetilde{x}} - \rho\sigma\xi\theta_{\widetilde{y}} &= 0 , \\ +\widehat{\rho}\sigma\xi_{\widetilde{x}} - \rho\sigma\xi\theta_{\widetilde{x}} - \xi^2\theta_{\widetilde{y}} &= 0 . \end{aligned}$$

Simplifying one gets:

$$\begin{aligned} \Phi &\doteq \frac{\partial\theta}{\partial\widetilde{x}} = \frac{-1}{\widehat{\rho}}\left(\frac{\xi\sigma_{\widetilde{y}}}{\sigma^2} + \frac{\rho\xi_{\widetilde{x}}}{\xi}\right) , \\ \Psi &\doteq \frac{\partial\theta}{\partial\widetilde{y}} = \frac{1}{\widehat{\rho}}\left(\frac{\sigma\xi_{\widetilde{x}}}{\xi^2} + \frac{\rho\sigma_{\widetilde{y}}}{\sigma}\right) . \end{aligned} \tag{4.8}$$

If one wants to find a solution for θ, one must first determine when the system is consistent, or integrable. This requires that:

$$\frac{\partial\Phi}{\partial\widetilde{y}} = \frac{\partial^2\theta}{\partial\widetilde{x}\,\partial\widetilde{y}} = \frac{\partial\Psi}{\partial\widetilde{x}} , \tag{4.9}$$

and the solution for θ is:

$$\theta(\widetilde{x},\widetilde{y}) = \int^{(\widetilde{x},\widetilde{y})} (\Phi\, d\widetilde{x} + \Psi\, d\widetilde{y}) . \tag{4.10}$$

However, if one applies Itô's lemma (Appendix (B.6)), one also obtains the following SDE for θ:

$$\begin{aligned} d\theta &= \mu^{(\theta)}dt + \sigma\Phi d\widetilde{\widehat{W}}_{1,t} + \xi\Psi d\widetilde{\widehat{W}}_{2,t} , \\ \mu^{(\theta)} &= \mu^{(\widetilde{x})}\Phi + \mu^{(\widetilde{y})}\Psi + \frac{1}{2}\sigma^2\frac{\partial\Phi}{\partial\widetilde{x}} + \rho\sigma\xi\frac{\partial^2\theta}{\partial\widetilde{x}\partial\widetilde{y}} + \frac{1}{2}\xi^2\frac{\partial\Psi}{\partial\widetilde{y}} . \end{aligned} \tag{4.11}$$

If one chooses to define θ in this way our system becomes a 3−Dimensional Itô process with two Wiener process inputs (θ−scheme):

$$\begin{bmatrix} d\widetilde{x} \\ d\widetilde{y} \\ d\theta \end{bmatrix} = \begin{bmatrix} \mu^{(\widetilde{x})} \\ \mu^{(\widetilde{y})} \\ \mu^{(\theta)} \end{bmatrix} dt + \begin{bmatrix} \sigma \\ 0 \\ \sigma\Phi \end{bmatrix} d\widetilde{\widehat{W}}_{1,t} + \begin{bmatrix} 0 \\ \xi \\ \xi\Psi \end{bmatrix} d\widetilde{\widehat{W}}_{2,t} . \tag{4.12}$$

If one computes again the Lie brackets with independent noise, one obtains (see Appendix (B.22)):

$$[A_1, A_2] = \begin{bmatrix} 0 \\ 0 \\ \widehat{\rho}\sigma\xi \left(\dfrac{\partial\ \Psi}{\partial \widetilde{x}} - \dfrac{\partial \Phi}{\partial \widetilde{y}} \right) \end{bmatrix} . \tag{4.13}$$

Note that when condition (4.9) is satisfied this Lie bracket (4.13) is identically zero. However, because not all SDEs satisfy condition (4.9) and the value of Lie brackets (4.13) does not depend on the drift for θ, one can always put:

$$\mu^{(\theta)} = 0 \ .$$

In the remainder of the chapter we shall explain why using the drift for θ equal to zero is the best approach. We also shall investigate when particular applications satisfy condition (4.9), in which case one can discretise either (4.6) or (4.12) and when they do not, in which case one can only discretise (4.12) or the original untransformed SDE (4.1). Our objective is to try to achieve higher order strong convergence without the simulation of the Lévy areas.

When the Lie bracket is not equal to zero, the important question to be considered is how precisely does θ need to be calculated to obtain first strong order convergence in $\widetilde{x}$ and $\widetilde{y}$? For example, does neglecting the Lie bracket affect the accuracy of θ but not in $\widetilde{x}$ and $\widetilde{y}$?

One approach of $\theta-$scheme results is given by Ana-Bela Cruzeiro, Paul Malliavin and T. Thalmaier in [6]. Because dW and $d\widetilde{W}$ have the same distribution (see Theorem 6 (B.23)), they ignore the calculation of θ. For example, the 1 strong order Milstein scheme for (4.6) with time step Δt using (4.8) is (see Appendix (B.17)):

$$\begin{bmatrix} \widetilde{x}_{t+\Delta t} \\ \widetilde{y}_{t+\Delta t} \end{bmatrix} = \begin{bmatrix} \widetilde{x}_t \\ \widetilde{y}_t \end{bmatrix} + \begin{bmatrix} \mu^{(\widetilde{x})} \\ \mu^{(\widetilde{y})} \end{bmatrix} \Delta t + \begin{bmatrix} \sigma & 0 \\ \rho\,\xi & \widehat{\rho}\,\xi \end{bmatrix} \begin{bmatrix} \Delta\widetilde{W}_{1,t} \\ \Delta\widetilde{W}_{2,t} \end{bmatrix} + \frac{1}{2} R_M \ , \tag{4.14}$$

where R_M is equal to:

$$\begin{aligned} R_M = & \begin{bmatrix} \sigma\sigma_x + \rho\xi\sigma_y \\ \rho\sigma\xi_x + \rho^2\xi\xi_y - \frac{\widehat{\rho}^{\,2}\xi^2\sigma_y}{\sigma} \end{bmatrix} \Big(\Delta\widetilde{W}_{1,t}^2 - \Delta t \Big) \\ & + \begin{bmatrix} -\rho\xi\sigma_y - \frac{\sigma^2\xi_x}{\xi} \\ \widehat{\rho}^{\,2}\xi\xi_y - \rho\sigma\xi_x - \frac{\rho^2\xi^2\sigma_y}{\sigma} \end{bmatrix} \Big(\Delta\widetilde{W}_{2,t}^2 - \Delta t \Big) \\ & + \begin{bmatrix} 2\widehat{\rho}\,\xi\sigma_y \\ 2\widehat{\rho}\sigma\xi_x + 2\rho\widehat{\rho}\Big(\xi\xi_y + \frac{\xi^2\sigma_y}{\sigma} \Big) \end{bmatrix} \Delta\widetilde{W}_{1,t}\Delta\widetilde{W}_{2,t} \ , \end{aligned}$$

and

$$\begin{bmatrix} \Delta\widetilde{W}_{1,t} \\ \Delta\widetilde{W}_{2,t} \end{bmatrix} = \begin{bmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{bmatrix} \begin{bmatrix} \Delta W_{1,t} \\ \Delta W_{2,t} \end{bmatrix} .$$

Replacing $\Delta\widetilde{W}$ by ΔW one obtains the Malliavin scheme published in [6] and in book [26]. Note that the advantage of this scheme is that one does not need to simulate the Lévy Area or worry about the value of θ every time step. For weak solutions the Malliavin scheme is a good approach. However for strong solutions, it has the same or worse strong convergence constant than both the scheme that includes the simulation of θ and the Milstein scheme that does not include the orthogonal transformation (4.4). For illustration, see the example in the next section with simulation plots (figures (4.1) to (4.4)).

4.2 Orthogonal Stochastic Volatility Models

In this section we consider four mean reverting stochastic volatility models. All four have the following generic form:

$$\begin{aligned} dx &= \mu^{(x)}dt + \alpha\, x^{\gamma_1} y^{\lambda_1} d\widehat{W}_{1,t} , \\ dy &= \mu^{(\tilde{y})}dt + \beta\, x^{\gamma_2} y^{\lambda_2} d\widehat{W}_{2,t} , \qquad E\left[d\widehat{W}_{1,t}, d\widehat{W}_{2,t}\right] = \rho\, dt . \end{aligned} \tag{4.15}$$

If one applies an orthogonal transformation, (4.15) changes to:

$$\begin{aligned} dx &= \mu^{(x)}dt + \alpha\, x^{\gamma_1} y^{\lambda_1} d\widetilde{\widehat{W}}_{1,t} , \\ dy &= \mu^{(\tilde{y})}dt + \beta\, x^{\gamma_2} y^{\lambda_2} d\widetilde{\widehat{W}}_{2,t} , \end{aligned} \tag{4.16}$$

where $d\widetilde{\widehat{W}}_{i,t}$ are the orthogonal correlated Wiener processes defined in (4.7). If one would like to obtain an exact solution of θ (4.10), the integrability condition (4.9) becomes (see Appendix (B.18)):

$$\frac{\partial\Phi}{\partial\widetilde{y}} = \frac{\lambda_C\, \lambda_1\, \beta\, y^{\lambda_C - 1}}{-\widehat{\rho}\alpha\, x^{\gamma_C+1}} = \frac{\gamma_C\gamma_2\alpha\; x^{\gamma_C-1}}{\widehat{\rho}\beta\, y^{\lambda_C+1}} = \frac{\partial\Psi}{\partial\widetilde{x}} , \tag{4.17}$$

$$\gamma_C = \gamma_1 - \gamma_2 - 1 , \quad \lambda_C = \lambda_2 - \lambda_1 - 1 ,$$

so then, for $\alpha, \beta, \lambda_1, \gamma_2 \neq 0$, one can conclude that θ is integrable if, and only if, $\lambda_C = \gamma_C = 0$, in which case the solution is:

$$\theta = \left(\frac{\rho\lambda_1\beta + \gamma_2\alpha}{\widehat{\rho}\beta}\right)\log y - \left(\frac{\rho\gamma_2\alpha + \lambda_1\beta}{\widehat{\rho}\alpha}\right)\log x . \tag{4.18}$$

Using the θ scheme (4.12), the 3−Dimensional Itô process for (4.16) is:

$$\begin{bmatrix} d\widetilde{x} \\ d\widetilde{y} \\ d\theta \end{bmatrix} = \begin{bmatrix} \mu^{(\widetilde{x})} \\ \mu^{(\widetilde{y})} \\ \mu^{(\theta)} \end{bmatrix} dt + \begin{bmatrix} \alpha\, x^{\gamma_1} y^{\lambda_1} \\ 0 \\ \alpha\, x^{\gamma_1} y^{\lambda_1} \Phi \end{bmatrix} d\widetilde{\widehat{W}}_{1,t} + \begin{bmatrix} 0 \\ \beta\, x^{\gamma_2} y^{\lambda_2} \\ \beta\, x^{\gamma_2} y^{\lambda_2} \Psi \end{bmatrix} d\widetilde{\widehat{W}}_{2,t} \ , \quad (4.19)$$

where:

$$\mu^{(\theta)} = 0 \ , \quad (4.20)$$

$$\Phi = \frac{\lambda_1 \beta\ y^{\lambda_C} + \rho\gamma_2\ \alpha\, x^{\gamma_C}}{-\widehat{\rho}\alpha\, x^{\gamma_C+1}} \ , \qquad \Psi = \frac{\rho\lambda_1\, \beta\, y^{\lambda_C} + \gamma_2 \alpha\ x^{\gamma_C}}{\widehat{\rho}\beta\, y^{\lambda_C+1}} \ .$$

If one computes the Lie brackets:

$$[A_1, A_2] = \begin{bmatrix} 0 \\ 0 \\ x^{2\gamma_2} y^{2\lambda_1} \left(\gamma_C \gamma_2 \alpha^2 x^{2\gamma_C} + \lambda_C \lambda_1 \beta^2 y^{2\lambda_C} \right) \end{bmatrix} . \quad (4.21)$$

Even without the condition (4.17) being satisfied one can perhaps improve the convergence using the θ−scheme without the simulation of the Lévy areas. However this depends on the parameters of our system. In other words, the accuracy is dependent on the value of the Lie bracket of the scheme (4.21). It will give us the bias in the calculation of the value of θ and hence in x and y. Note that when condition (4.17) is satisfied this Lie bracket (4.21) is identically zero.

4.2.1 The Quadratic Volatility Model (Case 1)

The first case we consider is the Quadratic Volatility Model:

$$\begin{aligned} dx &= x\,\overline{\mu}\,dt + x\,y\,d\widehat{W}_{1,t} \ , \quad (4.22) \\ dy &= k\,(\varpi_2 - y)\,dt + \beta_2\, y^2 d\widehat{W}_{2,t} \ . \end{aligned}$$

Because $\lambda_C = 0$, one can use either equation (4.16) together with (4.18), or the 3−Dimensional θ scheme (4.19). Because of the orthogonal transformation, neither requires the calculation of the Lévy area. Figure 4.1 and Table 4.1 show that, as expected, the Euler scheme and the Milstein scheme with zero Lévy areas (setting $\underline{\overline{L}}_{(1,2)} = 0$ in (4.4)) give strong convergence order 0.5. On the other hand, the Milstein scheme (4.4) with a proper value for the distribution of the Lévy area (through simulating the Lévy area using N subintervals within each time step) gives 1.0 order strong convergence, as do the two orthogonal θ−schemes. We have used the following parameters: $t_o = 0$; $T = 1$; $\rho = -0.50$; $\overline{\mu} = 0.1$; $k = 1.4$; $\varpi_2 = 0.32$; $\beta = 1.22$ and initial conditions $x(t_o) = 1$; $y(t_o) = \varpi_2$.

4.2.2 The 3/2 Model (Case 2)

The second case we consider is the following stochastic variance model, usually called the 3/2 Model [25]:

$$\begin{aligned} dx &= x\,\overline{\mu}\,dt + x\sqrt{y}d\widehat{W}_{1,t}\ , \\ dy &= k\,y(\varpi_{3/2}-y)\,dt + \beta_{3/2}\,y^{3/2}d\widehat{W}_{2,t}\ . \end{aligned} \tag{4.23}$$

Because $\lambda_C = 0$, we obtain almost the same results as Case 1 (Figure 4.2 and Table 4.1). The parameters and initial conditions are the same as in Case 1 except for $\varpi_{3/2} = \varpi_2^2$; $\beta = 2.44$; $y(t_o) = \varpi_2^2$; which are chosen so that x and y will have approximately the same relative volatility.

4.2.3 The GARCH Diffusion Model (Case 3)

The third case we consider is the following stochastic variance model, usually called GARCH Diffusion Model:

$$\begin{aligned} dx &= x\,\overline{\mu}\,dt + x\sqrt{y}d\widehat{W}_{1,t}\ , \\ dy &= k\,(\varpi_1-y)\,dt + \beta_1\,yd\widehat{W}_{2,t}\ . \end{aligned} \tag{4.24}$$

In this case $\lambda_C = 0.5$, and since the integrability condition is not satisfied it is not possible to use the $2D-\ \theta$ scheme. Figure 4.3 and Table 4.1 show that the only schemes that achieved first order convergence are the Milstein and θ schemes which simulate the Lévy area. However, the simulation results also show there is a remarkable difference between the original and the orthogonal scheme without the simulation of the Lévy area, not the improved order of convergence achieved in the first case but a much improved constant of proportionality. The parameters and initial conditions are the same as in Case 2 except for $\varpi_1 = \varpi_2^2$; $\beta = 0.78$; this is again chosen so ensure that x and y will have approximately the same relative volatility as in the first two cases.

4.2.4 The Square Root Model (Case 4)

The worst case for this example using the orthogonal transformation is Heston's Square Root Model [14]:

$$\begin{aligned} dx &= x\,\mu\,dt + x\sqrt{y}d\widehat{W}_{1,t}\ , \\ dy &= k\,(\varpi_{1/2}-y)\,dt + \beta_{1/2}\sqrt{y}d\widehat{W}_{2,t}\ . \end{aligned} \tag{4.25}$$

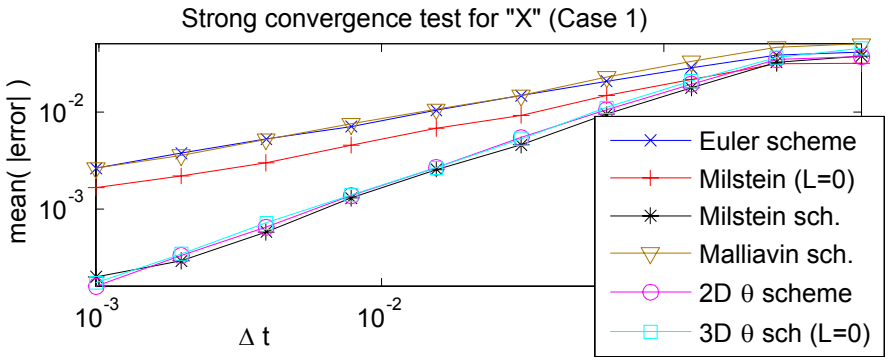

Figure 4.1: Strong convergence test for x (Case 1).

In this case, $\lambda_c = 1$. Figure 4.4 and Table 4.1 show that neither of the Milstein schemes in which the Lévy areas are set to zero performs very well. Both have order 0.5 strong convergence, and the constant of proportionality is not much better than for the Euler scheme. When the Lévy areas are simulated correctly, the Milstein and θ schemes do exhibit the expected first order strong convergence. This demonstrates the importance of the Lévy areas in this case. The parameters and initial conditions are the same as in Case 2 except for $\varpi_{1/2} = \varpi_2^2$; β=0.25.

Scheme	Description	C-1	C-2	C-3	C-4
Euler scheme	set $\triangle t$=dt, ΔW_i=dW_i in (4.3)	0.49	0.50	0.51	0.50
Milstein (L=0)	Milstein (4.4), set $\overline{L}_{(1,2)}$=0	0.52	0.54	0.53	0.53
Milstein sch.	Milstein (4.4), simulate $\overline{L}_{(1,2)}$	**0.94**	**0.95**	**0.96**	**0.96**
Malliavin sch.	Milstein (4.16), set $\Delta\widetilde{W}_i$=dW_i	0.50	0.52	0.50	0.49
$2D$-θ scheme	Milstein (4.16) with (4.18)	**0.96**	**0.95**	n/a[1]	n/a
$3D$-θ sch. (L=0)	Milstein (4.19), set $\overline{L}_{(1,2)}$=0	**0.96**	**0.95**	0.78	0.63
$3D$-θ scheme	Milstein (4.19), simulate $\overline{L}_{(1,2)}$	**0.96**	**0.95**	**0.95**	**0.94**

Table 4.1: Convergence orders γ for SVMs (all cases (4.22-4.25)).

4.2.5 Drift for θ Scheme

This section explains how the strong convergence order for x and y can change when applying a different drift in the θ equation in the $3D-\theta$ scheme (4.19). Lets start by assuming that θ is a function of x and y, two stochastic processes. Applying Itô's lemma, one obtains the following SDE (Appendix (B.6)):

$$d\theta = \mu^{(\theta)}dt + xy^{\lambda_1}\Phi d\widetilde{\widetilde{W}}_{1,t} + \beta\, y^{\lambda_2}\Psi d\widetilde{\widetilde{W}}_{2,t}\ ,$$

[1] n/a = not applicable.

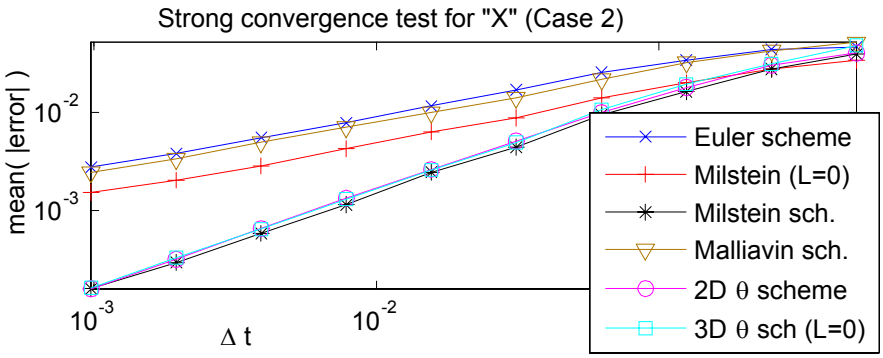

Figure 4.2: Strong convergence test for x (Case 2).

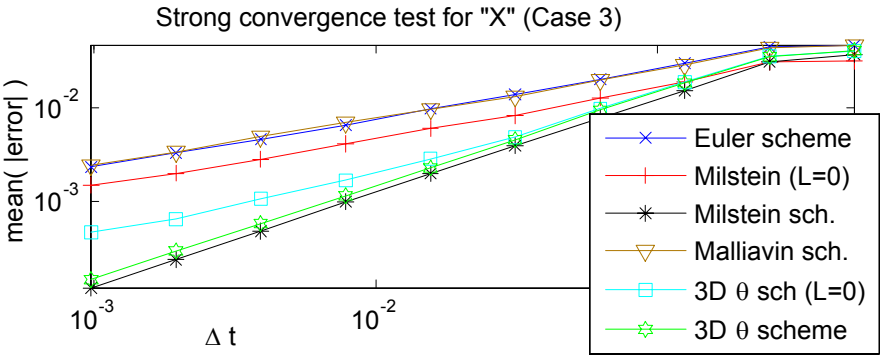

Figure 4.3: Strong convergence test for x (Case 3).

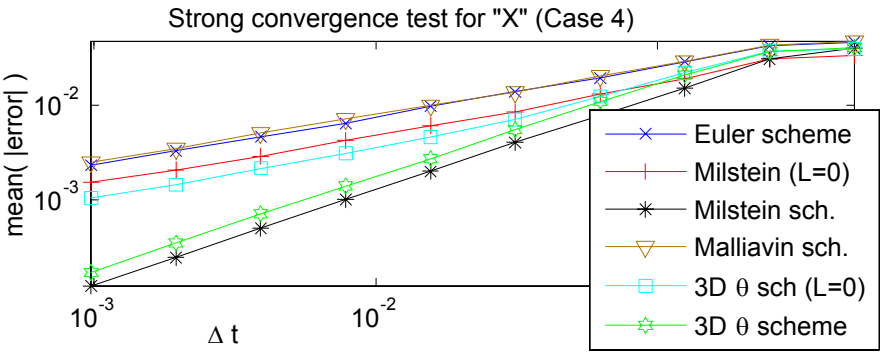

Figure 4.4: Strong convergence test for x (Case 4).

$$\mu^{(\theta)} = \mu^{(\widetilde{x})}\Phi + \mu^{(\widetilde{y})}\Psi + \frac{1}{2}\left(xy^{\lambda_1}\right)^2 \frac{\partial \Phi}{\partial \widetilde{x}} + \rho\beta\, xy^{\lambda_1+\lambda_2}\frac{\partial^2\theta}{\partial\widetilde{x}\partial\widetilde{y}} + \frac{1}{2}\left(\beta\, y^{\lambda_2}\right)^2 \frac{\partial \Psi}{\partial \widetilde{y}}\ , \quad (4.26)$$

where:

$\frac{\partial\theta}{\partial\widetilde{x}} = \Phi = \frac{-\lambda_1\beta\ y^{\lambda_C}}{\widehat{\rho}x}$	$\frac{\partial\theta}{\partial\widetilde{y}} = \Psi = \frac{\rho\lambda_1}{\widehat{\rho}\,y}$
$\frac{\partial^2\theta}{\partial\widetilde{x}^2} = \frac{\partial\Phi}{\partial\widetilde{x}} = \frac{\lambda_1\beta\ y^{\lambda_C}}{\widehat{\rho}x^2}$	$\frac{\partial^2\theta}{\partial\widetilde{y}^2} = \frac{\partial\Psi}{\partial\widetilde{y}} = \frac{-\rho\lambda_1}{\widehat{\rho}\,y^2}$

.

The problem comes when calculating the cross derivatives:

$\frac{\partial^2\theta}{\partial\widetilde{x}\partial\widetilde{y}} = \frac{\partial\Phi}{\partial\widetilde{y}} = \frac{-\lambda_C\lambda_1\beta\ y^{\lambda_C-1}}{\widehat{\rho}x}$	$\frac{\partial^2\theta}{\partial\widetilde{y}\partial\widetilde{x}} = \frac{\partial\Psi}{\partial\widetilde{x}} = 0$

. (4.27)

Only when condition (4.17) is satisfy ($\lambda_C = 0$), the cross derivatives are equal and the correct drift (4.26) can be applied in the $3D-\theta$ scheme (4.19). To understand how the drift of θ changes the convergence in x and y, three examples are presented using Case 2 and the same initials conditions:

Case 2	$\mu^{(\theta)} = 0$
Case 2a	$\mu^{(\theta)}$ =Itô (4.26)
Case 2b	$\mu^{(\theta)} = \pi/2$

.

In the first plot (Figure (4.5)) we calculate the strong order of convergence of θ at time T between the formula (4.18) and 3−Dimensional θ scheme (4.19) using different time steps Δt:

$$E\left[\ \left|\theta_{3D}\left(T\right) - \theta_{2D}\left(T\right)\right|\ \right] = C_\mu \Delta t^{\gamma_\mu}\ .$$

In Figure (4.5) we obtain as predicted, one strong order convergence when using the correct drift for θ (4.26), and a constant error when using the other drifts. However, when applying a strong convergence test to θ and x, Figures 4.6 and 4.7 display some differences. This is because the value θ changes every time step depending on its drift. The greater the value of θ, the poorer the strong convergence constant for θ and therefore with x and y. In addition, because not all SDEs satisfy the conditions (4.17) to calculate the cross derivatives (4.27) and the only goal when using the θ scheme is to obtain zero in the coefficients of the Lie Brackets (4.21) for x and y, we conclude that the best approach for the θ scheme is when using:

$$\mu^{(\theta)} = 0\ .$$

4.3 $2D$ Orthogonal Milstein Scheme (θ Scheme)

This section presents the definition of the θ scheme that generalizes the application of an orthogonal transformation to a 2−Dimensional SDE [30].

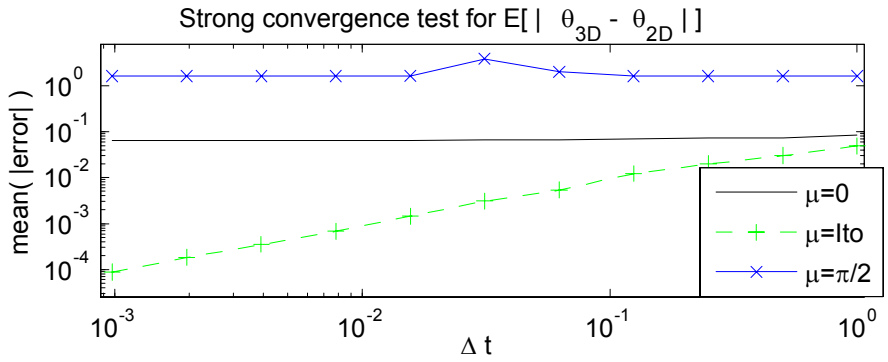

Figure 4.5: Expectation of the absolute error of θ at time T.

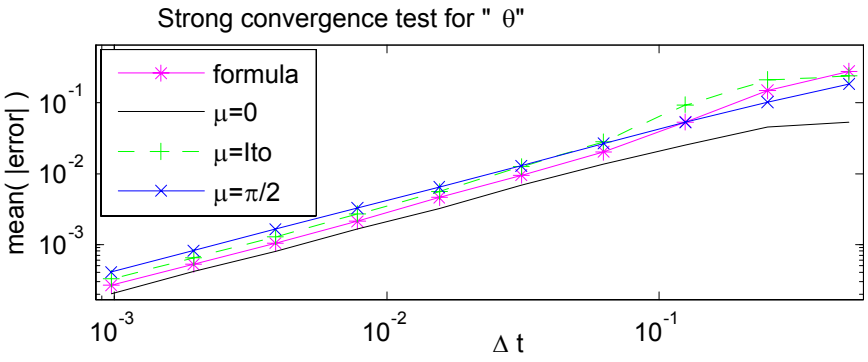

Figure 4.6: Strong convergence test for θ (Case 2).

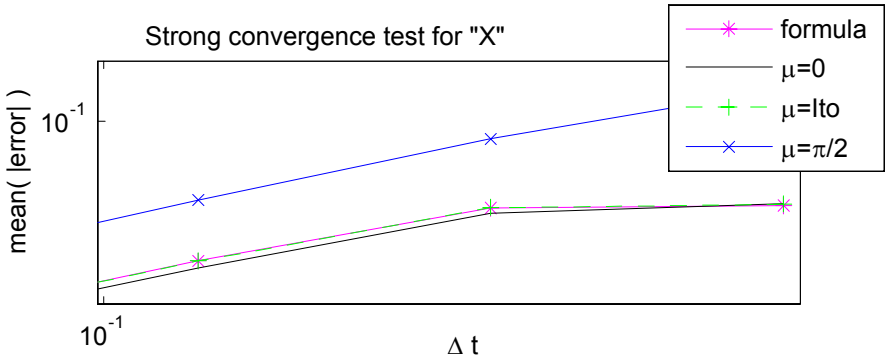

Figure 4.7: Strong convergence test for x (zoom Case 2).

4.3.1 $2D-\theta$ Scheme

Theorem 4: $2D-\theta$ scheme (Exact solution)

If one has a 2−Dimensional Itô stochastic differential equation with two independent Wiener process:

$$d\begin{bmatrix} X_{1,t} \\ X_{2,t} \end{bmatrix} = \begin{bmatrix} a_1 \\ a_2 \end{bmatrix} dt + \begin{bmatrix} b_{1,1} & b_{1,2} \\ b_{2,1} & b_{2,2} \end{bmatrix} \begin{bmatrix} dW_{1,t} \\ dW_{2,t} \end{bmatrix} , \tag{4.28}$$

where a_i,$b_{i,k}$ are smooth functions of t, X_t and Y_t, satisfying the global Lipschitz conditions (3.37) and:

$$\frac{\partial \Psi}{\partial X_1} = \frac{\partial \Phi}{\partial X_2} , \tag{4.29}$$

where:

$$\Phi \doteq \frac{H_1^- \left(b_{2,1}^2 + b_{2,2}^2\right) - H_2^- \left(b_{1,1}b_{2,1} + b_{1,2}b_{2,2}\right)}{\left(b_{1,1}b_{2,2} - b_{1,2}b_{2,1}\right)^2} ,$$

$$\Psi \doteq \frac{H_2^- \left(b_{1,1}^2 + b_{1,2}^2\right) - H_1^- \left(b_{1,1}b_{2,1} + b_{1,2}b_{2,2}\right)}{\left(b_{1,1}b_{2,2} - b_{1,2}b_{2,1}\right)^2} ,$$

and H_i^- are the coefficients of the Lévy area (Lie bracket) of (4.28) and are defined by:

$$H_j^- = L_1 b_{j,2} - L_2 b_{j,1} ,$$

$$L_j := \sum_{k=1}^{d} b_{k,j} \frac{\partial}{\partial X_k} .$$

If one applies an orthogonal transformation to (4.28) described by:

$$\begin{bmatrix} d\widetilde{W}_{1,t} \\ d\widetilde{W}_{2,t} \end{bmatrix} = \begin{bmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{bmatrix} \begin{bmatrix} dW_{1,t} \\ dW_{2,t} \end{bmatrix} , \tag{4.30}$$

where:

$$\theta_t\left(X_1, X_2\right) = \int^{X_1} \Phi dX_1 + \int^{X_2} \Psi dX_2 , \tag{4.31}$$

then the new orthogonal process has 1 strong order convergence using the Milstein scheme neglecting the simulation of the Lévy Area. Conversely, for $H_i^- \neq 0$ (the commutativity condition is not satisfied), the Milstein scheme of (4.28) with zero Lévy Area has 0.5 strong order convergence.

Proof:

The 1 strong order Milstein scheme for (4.28) with time step Δt is (Appendix B.7):

$$\begin{bmatrix} \widehat{X}_{1,t+\Delta t} \\ \widehat{X}_{2,t+\Delta t} \end{bmatrix} = \begin{bmatrix} \widehat{X}_{1,t} \\ \widehat{X}_{2,t} \end{bmatrix} + \begin{bmatrix} a_1 \\ a_2 \end{bmatrix} \Delta t + \begin{bmatrix} b_{1,1} & b_{1,2} \\ b_{2,1} & b_{2,2} \end{bmatrix} \begin{bmatrix} \Delta W_{1,t} \\ \Delta W_{2,t} \end{bmatrix} + \frac{1}{2} R_M ,$$

$$R_M = \sum_{j=1}^{2} \begin{bmatrix} L_j b_{1,j} \\ L_j b_{2,j} \end{bmatrix} \left(\Delta W_{j,t}^2 - \Delta t\right) + \begin{bmatrix} H_1^+ \\ H_1^+ \end{bmatrix} \Delta W_{1,t} \Delta W_{2,t} + \begin{bmatrix} H_1^- \\ H_2^- \end{bmatrix} \left[\underline{L}_{(1,2)}\right]_t^{t+\Delta t} ,$$

where:

$$H_j^{\pm} = L_1 b_{j,2} \pm L_2 b_{j,1} .$$

For $H_j^{\pm} \neq 0$, the Milstein scheme is 1 strong order convergent when one includes all terms in the equation (see Theorem 10.3.5, page 350 from [22]), otherwise it becomes 0.5 strong order convergence. In general, if X_T is the solution of the SDE (4.28) and $\widehat{X}_T$ is the numerical approximation using the Milstein scheme, for $H_j^- \neq 0$ and neglecting the simulation of the Lévy Area, one can say:

$$E\left[X_T - \widehat{X}_T\right] \leq \widehat{C}_1 \left(\Delta t\right)^{0.5} .$$

On the other hand, if one makes an orthogonal transformation (4.30) to (4.28), one obtains:

$$d\begin{bmatrix} \widetilde{X}_{1,t} \\ \widetilde{X}_{2,t} \end{bmatrix} = \begin{bmatrix} a_1 \\ a_2 \end{bmatrix} dt + \begin{bmatrix} b_{1,1} & b_{1,2} \\ b_{2,1} & b_{2,2} \end{bmatrix} \begin{bmatrix} d\widetilde{W}_{1,t} \\ d\widetilde{W}_{2,t} \end{bmatrix} . \tag{4.32}$$

The system (4.32) with independent noise can be represented as:

$$d\begin{bmatrix} \widetilde{X}_{1,t} \\ \widetilde{X}_{2,t} \end{bmatrix} = \begin{bmatrix} a_1 \\ a_2 \end{bmatrix} dt + \begin{bmatrix} \widetilde{b}_{1,1} & \widetilde{b}_{1,2} \\ \widetilde{b}_{2,1} & \widetilde{b}_{2,2} \end{bmatrix} \begin{bmatrix} dW_{1,t} \\ dW_{2,t} \end{bmatrix} , \tag{4.33}$$

where:

$$\begin{bmatrix} \widetilde{b}_{1,1} & \widetilde{b}_{1,2} \\ \widetilde{b}_{2,1} & \widetilde{b}_{2,2} \end{bmatrix} = \begin{bmatrix} b_{1,1} & b_{1,2} \\ b_{2,1} & b_{2,2} \end{bmatrix} \begin{bmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{bmatrix} .$$

The 1 strong order Milstein scheme for (4.33) with time step Δt is (Appendix (B.12)):

$$\begin{bmatrix} \widehat{\widetilde{X}}_{1,t+\Delta t} \\ \widehat{\widetilde{X}}_{2,t+\Delta t} \end{bmatrix} = \begin{bmatrix} \widehat{\widetilde{X}}_{1,t} \\ \widehat{\widetilde{X}}_{2,t} \end{bmatrix} + \begin{bmatrix} a_1 \\ a_2 \end{bmatrix} \Delta t + \begin{bmatrix} \widetilde{b}_{1,1} & \widetilde{b}_{1,2} \\ \widetilde{b}_{2,1} & \widetilde{b}_{2,2} \end{bmatrix} \begin{bmatrix} \Delta W_{1,t} \\ \Delta W_{2,t} \end{bmatrix} + \frac{1}{2} R_M ,$$

$$R_M = \sum_{j=1}^{2} \begin{bmatrix} \widetilde{L}_j \widetilde{b}_{1,j} \\ \widetilde{L}_j \widetilde{b}_{2,j} \end{bmatrix} \left(\Delta W_{j,t}^2 - \Delta t\right) + \begin{bmatrix} \widetilde{H}_1^+ \\ \widetilde{H}_2^+ \end{bmatrix} \Delta W_{1,t} \Delta W_{2,t} + \begin{bmatrix} \widetilde{H}_1^- \\ \widetilde{H}_2^- \end{bmatrix} \left[\underline{L}_{(1,2)}\right]_t^{t+\Delta t} ,$$

where:

$$\widetilde{H}_j^{\pm} = \widetilde{L}_1 \widetilde{b}_{j,2} \pm \widetilde{L}_2 \widetilde{b}_{j,1} ,$$

$$\widetilde{L}_j := \sum_{k=1}^{d} \widetilde{b}_{k,j} \frac{\partial}{\partial \widetilde{X}_k} \ .$$

If one computes the coefficients of the Lévy Area using independent Wiener processes (Appendix (B.16)), one gets:

$$\begin{bmatrix} \widetilde{H}_1^- \\ \widetilde{H}_2^- \end{bmatrix} = \begin{bmatrix} H_1^- - \frac{\partial\theta}{\partial X_1}\left(b_{1,1}^2 + b_{1,2}^2\right) - \frac{\partial\theta}{\partial X_2}\left(b_{1,1}b_{2,1} + b_{1,2}b_{2,2}\right) \\ H_2^- - \frac{\partial\theta}{\partial X_2}\left(b_{2,1}^2 + b_{2,2}^2\right) - \frac{\partial\theta}{\partial X_1}\left(b_{1,1}b_{2,1} + b_{1,2}b_{2,2}\right) \end{bmatrix} \ . \tag{4.34}$$

To avoid having to simulate the Lévy Area, one needs (4.34) to be identically zero, i.e., you need to impose the following conditions:

$$\begin{bmatrix} \widetilde{H}_1^- & \widetilde{H}_2^- \end{bmatrix} = 0 \ .$$

Simplifying one gets:

$$\Phi \doteq \frac{\partial\theta}{\partial X_1} = \frac{H_1^-\left(b_{2,1}^2 + b_{2,2}^2\right) - H_2^-\left(b_{1,1}b_{2,1} + b_{1,2}b_{2,2}\right)}{\left(b_{1,1}b_{2,2} - b_{1,2}b_{2,1}\right)^2} \ ,$$

$$\Psi \doteq \frac{\partial\theta}{\partial X_2} = \frac{H_2^-\left(b_{1,1}^2 + b_{1,2}^2\right) - H_1^-\left(b_{1,1}b_{2,1} + b_{1,2}b_{2,2}\right)}{\left(b_{1,1}b_{2,2} - b_{1,2}b_{2,1}\right)^2} \ .$$

To find a solution for θ, one must first determine when the system is consistent, or integrable; this requires condition (4.29) and the solution for θ is (4.31). □

4.3.2 $3D - \theta$ Scheme

If one has a 2−Dimensional Itô process (4.28) and applies an orthogonal transformation (4.30) to it, where the rotation angle θ_t is described using a third SDE:

$$d\begin{bmatrix} \widetilde{X}_{1,t} \\ \widetilde{X}_{2,t} \\ \theta_t \end{bmatrix} = \begin{bmatrix} a_1 \\ a_2 \\ 0 \end{bmatrix} dt + \begin{bmatrix} b_{1,1} & b_{1,2} \\ b_{2,1} & b_{2,2} \\ (\Phi b_{1,1} + \Psi b_{2,1}) & (\Phi b_{1,2} + \Psi\, b_{2,2}) \end{bmatrix} \begin{bmatrix} d\widetilde{W}_{1,t} \\ d\widetilde{W}_{2,t} \end{bmatrix} \ , \tag{4.35}$$

then, for sufficiently smooth functions $b_{i,k}$, the Milstein scheme for the 3−Dimensional SDE (4.35) can have better strong convergence than (4.28) using the Milstein scheme neglecting the simulation of the Lévy Area. The accuracy of θ_t and hence in $\widetilde{X}_{i,t}$ depends on the value of the Lie bracket (4.36) of the process (4.35):

$$R_{\underline{\widetilde{L}}} = \begin{bmatrix} 0 \\ 0 \\ (b_{1,1}b_{2,2} - b_{1,2}b_{2,1})\left(\frac{\partial\Psi}{\partial X_1} - \frac{\partial\Phi}{\partial X_2}\right) \end{bmatrix} \ . \tag{4.36}$$

The 1 strong order Milstein scheme for (4.35) with time step Δt is (Appendix (B.19)):

$$\begin{bmatrix} \widehat{\widetilde{X}}_{1,t+\Delta t} \\ \widehat{\widetilde{X}}_{2,t+\Delta t} \\ \widehat{\theta}_{t+\Delta t} \end{bmatrix} = \begin{bmatrix} \widehat{\widetilde{X}}_{1,t} \\ \widehat{\widetilde{X}}_{2,t} \\ \widehat{\theta}_{t} \end{bmatrix} + \begin{bmatrix} a_1 \\ a_2 \\ 0 \end{bmatrix} \Delta t + \begin{bmatrix} \widetilde{b}_{1,1} & \widetilde{b}_{1,2} \\ \widetilde{b}_{2,1} & \widetilde{b}_{2,2} \\ b_{\theta,1} & b_{\theta,2} \end{bmatrix} \begin{bmatrix} \Delta W_{1,t} \\ \Delta W_{2,t} \end{bmatrix} + \frac{1}{2} R_M \ ,$$

$$\begin{aligned} R_M &= \sum_{j=1}^{2} \begin{bmatrix} \widetilde{L}_j \widetilde{b}_{1,j} \\ \widetilde{L}_j \widetilde{b}_{2,j} \\ \widetilde{L}_j \widetilde{b}_{3,j} \end{bmatrix} \left(\Delta W_{j,t}^2 - \Delta t \right) + \begin{bmatrix} \widetilde{H}_1^+ \\ \widetilde{H}_2^+ \\ \widetilde{H}_3^+ \end{bmatrix} \Delta W_{1,t} \Delta W_{2,t} \\ &\quad + \begin{bmatrix} \widetilde{H}_1^- \\ \widetilde{H}_2^- \\ \widetilde{H}_3^- \end{bmatrix} \left[\underline{\overline{L}}_{(1,2)} \right]_t^{t+\Delta t} \ , \end{aligned}$$

where:

$$\widetilde{H}_j^{\pm} = \widetilde{L}_1 \widetilde{b}_{j,2} \pm \widetilde{L}_2 \widetilde{b}_{j,1} \ .$$

If one computes the coefficients of the Lévy Area of the last equation (Appendix (B.22)), one obtains:

$$R_{\underline{\overline{L}}} = \begin{bmatrix} 0 & 0 & \widetilde{H}_3^- \end{bmatrix}^T \ .$$

If the value of $\widetilde{H}_3^-$ in the Lie bracket $R_{\underline{\overline{L}}}$ is small enough, the accuracy of θ_t is not affected by neglecting this term in the equation and hence, the $3D$ Itô process (4.35) will have better strong convergence than (4.28) using Milstein scheme neglecting the simulation of the Lévy Area. Note that when condition (4.29) is satisfied the Lie bracket (4.36) is identically zero ($\widetilde{H}_3^- = 0$).

4.3.3 Example of θ Scheme

Consider the following $2D$ SDEs:

$$\begin{aligned} dx &= x\,\mu_x\,dt + 0.5 x^{\gamma} \sqrt{y}\, d\widehat{W}_{1,t} \ , \\ dy &= x\,\mu_y\ dt + 0.5 \sqrt{x} y^{\lambda} d\widehat{W}_{2,t} \ , \qquad E\left[d\widehat{W}_{1,t}, d\widehat{W}_{2,t} \right] = \rho\, dt \ , \end{aligned} \tag{4.37}$$

where:

$$\mu_x = \mu_y = 0.05,\ \rho = -0.2,\ x\left(t_o\right) = 1,\ y\left(t_o\right) = 0.3^2 \ .$$

If $\gamma = \lambda = 1.5$, then we have the integrability condition (4.29) or (4.17) and either Theorem 4 ($2D-\theta$ scheme) or $3D-\theta$ scheme can be applied. Figures 4.8 and 4.9 show that the new orthogonal process of (4.37) has 1 strong order convergence in x and y using the Milstein scheme neglecting the simulation of the Lévy Area. Conversely,

Euler, Malliavin and the Milstein schemes with zero Lévy Area have 0.5 strong order convergence in x and y.

If $\gamma = \lambda = 1$, then the integrability condition is not (4.29) or (4.17) and only the $3D-\theta$ scheme can be applied. Figure 4.10 shows that the only schemes that achieved first order convergence are the Milstein and θ schemes which simulate the Lévy area. However, Figure 4.10 shows there is a remarkable difference between the original and the orthogonal scheme without the simulation of the Lévy area, not the improved order of convergence achieved in the first case ($\gamma = \lambda = 1.5$) but a much improved constant of proportionality.

The numerical results do not simply confirm an outcome that has been rigorously derived in Theorem 4 under global Lipschitz conditions, but are indicative that under less restrictive assumptions such results are possible; this is however not within the scope of the thesis.

4.4 θ Scheme (N-Dimension)

In this section we shall present a summary when one deals with an $N-$Dimensional SDE and would like to apply an orthogonal transformation to avoid the calculation of the Lévy Area. All models can be described through a SDE of the form:

$$dX_t = \mu\left(X_t, t\right) dt + \sigma\left(X_t, t\right) dW_t \ , \qquad X(t_0) = X_0 \ , \tag{4.38}$$

where:

$$\begin{aligned} X_t &= X(t) \in \mathbb{R}^d \ , \ W_t \in \mathbb{R}^M \ , \ t \in [t_0, ...T] \in \mathbb{R} \ , \\ \sigma\left(X_t, t\right) &= \sigma\left(b_{i,k}\left(X_t, t\right)\right) \in \mathbb{R}^{d \times M} \ , \quad \mu\left(X_t, t\right) = \mu\left(a_i\left(X_t, t\right)\right) \in \mathbb{R}^d \ , \\ & E\left[dW_{j,t} dW_{k,t}\right] = 0 \ , \qquad for \ i \neq k \ , \end{aligned}$$

or in matrix form by:

$$d \begin{bmatrix} X_{1,t} \\ X_{2,t} \\ ... \\ X_{d,t} \end{bmatrix} = \begin{bmatrix} a_1\left(X_t, t\right) \\ a_2\left(X_t, t\right) \\ ..... \\ a_d\left(X_t, t\right) \end{bmatrix} dt + \begin{bmatrix} b_{1,1} & b_{1,2} & ... & b_{1,M} \\ b_{2,1} & b_{2,2} & ... & b_{2,M} \\ ... & ... & ... & ... \\ b_{d,1} & b_{d,2} & ... & b_{d,M} \end{bmatrix} \begin{bmatrix} dW_{1,t} \\ dW_{2,t} \\ ... \\ dW_{M,t} \end{bmatrix} .$$

If one replaces the Wiener process W_t by an orthogonal transform $\widetilde{W_t}$, the probability distribution does not change and we obtain the set of all orthogonal transformations of our system (4.38):

$$d\widetilde{X}_t = \mu\left(\widetilde{X}_t, t\right) dt + \sigma\left(\widetilde{X}_t, t\right) d\widetilde{W}_t \ , \tag{4.39}$$

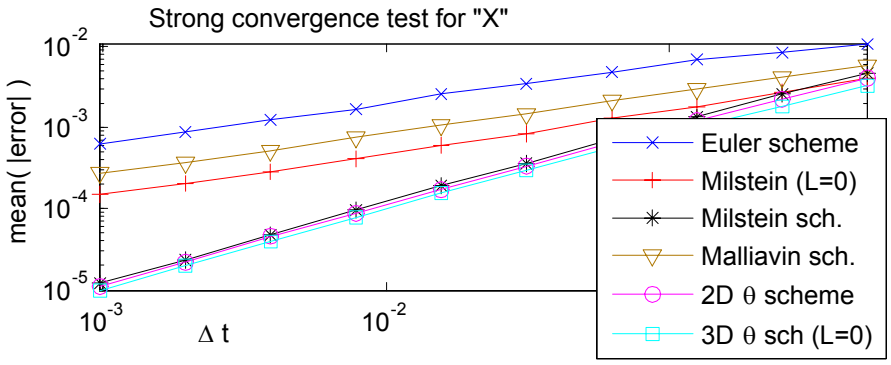

Figure 4.8: Strong convergence test for x ($2D$ & $3D-\theta$ scheme).

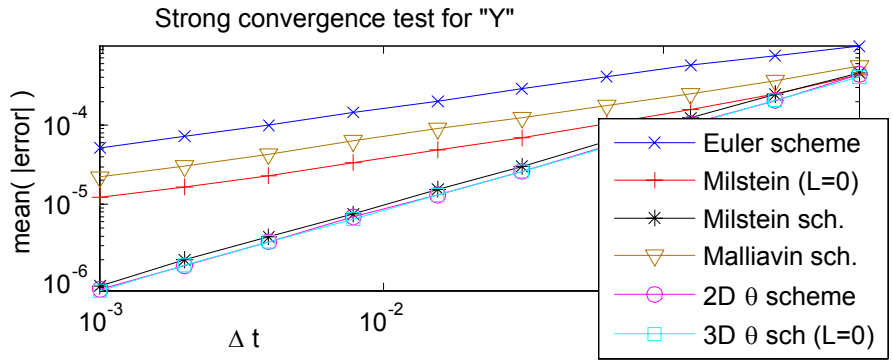

Figure 4.9: Strong convergence test for y ($2D$ & $3D-\theta$ scheme).

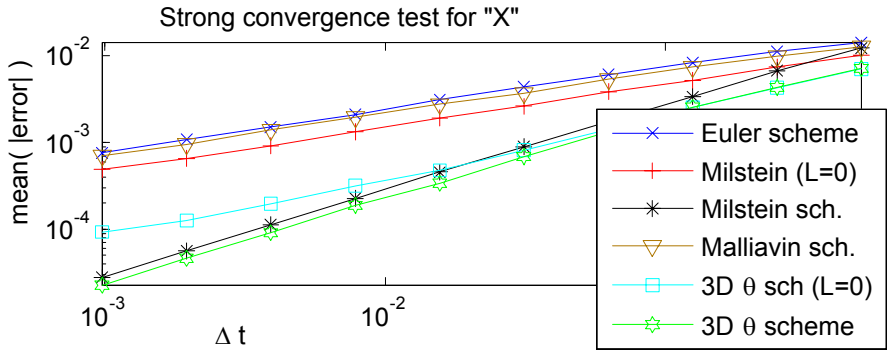

Figure 4.10: Strong convergence test for x ($3D-\theta$ scheme).

where:

$$d\widetilde{W}_t = \Gamma(\theta_t)\, dW_t \qquad \text{and} \qquad \Gamma(\theta_t) = \Gamma\left(\Theta_{i,k}\left(\widetilde{X}_t, t\right)\right) \in \mathbb{R}^{M \times M} \ .$$

Using independent Wiener process, (4.39) can be represented by:

$$d\widetilde{X}_t = \mu\left(\widetilde{X}_t, t\right) dt + \sigma\left(\widetilde{X}_t, t\right) \Gamma(\theta_t)\, dW_t \ ,$$

or

$$d\widetilde{X}_t = \mu\left(\widetilde{X}_t, t\right) dt + \widetilde{\sigma}\left(\widetilde{X}_t, t : \theta_t\right) dW_t \ , \tag{4.40}$$

where:

$$\widetilde{\sigma}\left(\widetilde{X}_t, t : \theta_t\right) = \begin{bmatrix} b_{1,1} & b_{1,2} & ... & b_{1,M} \\ b_{2,1} & b_{2,2} & ... & b_{2,M} \\ ... & ... & ... & ... \\ b_{d,1} & b_{d,2} & ... & b_{d,M} \end{bmatrix} \begin{bmatrix} \Theta_{1,1} & \Theta_{1,2} & ... & \Theta_{1,M} \\ \Theta_{2,1} & \Theta_{2,2} & ... & \Theta_{2,M} \\ ... & ... & ... & ... \\ \Theta_{M,1} & \Theta_{M,2} & ... & \Theta_{M,M} \end{bmatrix} .$$

Note that, if one replaces again the Wiener process dW in (4.40) by:

$$d\widetilde{W}'_t = \Gamma^T dW_t \ ,$$

then, one recovers the original process (4.38):

$$\widetilde{X}'_t = X_t \ .$$

This is easy to see if one knows the main property of an orthogonal transformation:

$$\begin{aligned} dX_t &= \mu(X_t, t)\, dt + \sigma(X_t, t)\, dW_t \\ &= \mu(X_t, t)\, dt + \sigma(X_t, t)\, \Gamma(\theta_t)\, \Gamma(\theta_t)^{-1} dW_t \\ &= \mu(X_t, t)\, dt + \widetilde{\sigma}(X_t, t : \theta_t)\, \Gamma(\theta_t)^T dW_t \\ &= \mu(X_t, t)\, dt + \widetilde{\sigma}(X_t, t : \theta_t)\, d\widetilde{W}'_t \\ &= d\widetilde{X}'_t \ . \end{aligned}$$

The 1 strong order Milstein scheme for (4.40) with time step Δt using Itô operators is:

$$Z_{i,t+\Delta_t} = Z_{i,t} + \mu_i \Delta t + \sum_{j=1}^{M} \widetilde{b}_{i,j} \Delta W_{j,t} + \frac{1}{2} R_M \ ,$$

where

$$\widetilde{b}_{i,j}\left(\widetilde{X}_t, t : \theta_t\right) = \sum_{s}^{M} b_{i,s} \Theta_{s,j} \ .$$

If one uses the Lévy Areas, R_M is equal to:

$$R_M = \sum_{j_1,j_2=1}^{M} \widetilde{L}_{j_1}\widetilde{b}_{i,j_2}\left(\Delta W_{j_1,t}\Delta W_{j_2,t} - \widetilde{\delta}_{j_1,j_2}\Delta t\right) + \sum_{j_1<j_2=1}^{M}\left(R^{\overline{L}}_{(j_1,j_2)}\right)_i \left[\overline{L}_{(j_1,j_2)}\right]_t^{t+\Delta t} ,$$

$$\left(R^{\overline{L}}_{(j_1,j_2)}\right)_i = \left(\widetilde{L}_{j_1}\widetilde{b}_{i,j_2} - \widetilde{L}_{j_2}\widetilde{b}_{i,j_1}\right) . \tag{4.41}$$

$\widetilde{\delta}_{j_1,j_2}$ is the Kronecker symbol ($\widetilde{\delta}_{j_1,j_2} = 1$ if $j_1 = j_2$ and zero otherwise) and the Itô operators are defined by:

$$\widetilde{L}_j := \sum_{k=1}^{d} \widetilde{b}_{k,j}\frac{\partial}{\partial Z_k} .$$

Using the definition of the variables and considering the vector fields to be independent of time, the coefficients for the Lévy Area (4.41) are equal to:

$$\left(R^{\overline{L}}_{(j_1,j_2)}\right)_i = \sum_{k=1}^{d}\sum_{s_1=1}^{M}\sum_{s_2=1}^{M} b_{k,s_2}\left(\begin{array}{c} \frac{\partial b_{i,s_1}}{\partial Z_k}\left(\Theta_{s_1,j_2}\Theta_{s_2,j_1} - \Theta_{s_1,j_1}\Theta_{s_2,j_2}\right) \\ +b_{i,s_1}\left(\Theta_{s_2,j_1}\frac{\partial \Theta_{s_1,j_2}}{\partial Z_k} - \Theta_{s_2,j_2}\frac{\partial \Theta_{s_1,j_1}}{\partial Z_k}\right)\end{array}\right) . \tag{4.42}$$

Using orthogonal properties, (4.42) can be reduced to:

$$\left(R^{\overline{L}}_{(j_1,j_2)}\right)_i = \sum_{k=1}^{d}\left((-1)^{k+1}\sum_{s=1}^{M} b_{i,s}b_{k,s} + \theta_k \sum_{s_1<s_2=1}^{M}\left(b_{k,s_1}\frac{\partial b_{i,s_2}}{\partial Z_k} - b_{k,s_2}\frac{\partial b_{i,s_1}}{\partial Z_k}\right)\right) ,$$

where θ_k are the orthogonal functions defined by:

$$\theta_k = (-1)^{k+1}\left(\Theta_{k,k+1_{\bar{X}_k}}\Theta_{k,k} - \Theta_{k,k_{\bar{X}_k}}\Theta_{k,k+1}\right) .$$

To avoid having to simulate the Lévy areas $\overline{L}_{(j_1,j_2)}$, one needs to impose the following conditions:

$$\left(R^{\overline{L}}_{(j_1,j_2)}\right)_i = 0 .$$

4.5 Conclusions

Strong convergence properties of discretizations of stochastic differential equations (SDEs) are very important in stochastic calculus. We have shown that under certain conditions the use of the orthogonal θ scheme can achieve the first order strong convergence properties of the Milstein numerical discretization without the expensive simulation of Lévy areas. Conversely, the Milstein scheme with zero Lévy Area has a 0.5 strong order convergence.

The bias or error in the computation of the rotation angle θ that makes the Lie bracket equal to zero in the orthogonal scheme is crucial to obtain a better convergence

order. When the conditions for integrability are satisfied, one can use the formula for θ to obtain the value of the rotation angle and obtain first order strong convergence. Otherwise, one has to use the $3-$Dimensional transformation and check the magnitude of the Lie brackets to decide if it is likely to give computational savings in the solution of our system.

Standard convergence theory for numerical methods for SDEs (e.g. as in Kloeden and Platen [22]) makes a global Lipschitz assumption on the coefficients. However, most of the SDE models that are mentioned in the chapter, and used in the computational experiments, do not satisfy such global Lipschitz conditions (e.g. example (4.37)). The numerical results are not simply confirming a theory that has been proved; they are giving numerical evidence that the conclusions about strong order remain true in circumstances where no theory currently exists.

The numerical results show a better strong order of convergence than the standard Milstein scheme (4.4 without the simulation of the Lévy Area) when an orthogonal transformation is applied to the quadratic volatility model (4.22), or the 3/2 Model (4.23) or the GARCH diffusion Model (4.24). Unfortunately, similar results are not achieved with the Heston model (4.25), and so the orthogonal transformation without the simulation of the Lévy Area is not recommended in this case.

Chapter 5

Pricing Exotic Options using MSL-MC

In finance, the convergence properties of discretizations of stochastic differential equations (SDEs) are very important for hedging and the valuation of exotic options. The last chapter shows that if certain conditions are satisfied, one can avoid the calculation of the Lévy area and obtain first order convergence by applying an orthogonal transformation. We have demonstrated when the conditions of the $2-$Dimensional problem permit it and give an exact solution for the orthogonal transformation.

This chapter demonstrates how the use of stochastic volatility models and the θ scheme can improve the convergence of the multi-level Monte Carlo method (ML-MC [10]), so that the computational cost to achieve an accuracy of $O(\epsilon)$ is reduced from $O(\epsilon^{-3})$ to $O(\epsilon^{-2})$ for a Lipschitz payoff. We present a modification to the ML-MC algorithm that can be used to achieve better savings in some cases. To illustrate these, various examples of exotic options using a wide variety of payoffs and the new Multischeme Multilevel Monte Carlo method (MSL-MC) are given. For standard payoffs, both European and Digital options are presented. For complex payoffs, such as combinations of European options, examples are also given (Butterfly Spread, Strip and Strap options). Finally, for path dependent payoffs, both Asian and Variance Swap options are demonstrated.

5.1 Multilevel Monte Carlo Path Simulation Method (ML-MC)

Usually, it is the weak convergence property of numerical discretizations which is most important, because in financial applications one is mostly concerned with the accurate estimation of expected payoffs. However, in the recently developed Multilevel Monte

Carlo path simulation method (ML-MC [10]), the strong convergence property plays a crucial role.

The key idea in the ML-MC approach is the use of a multilevel algorithm with different time steps Δt on each level. Suppose level L uses 2^L time steps of size $\Delta t_L = 2^{-L}\,T$, and define P_L to be the numerical approximation to the payoff on this level. Let L_F represent the finest level, with time steps so small that the bias due to the numerical discretization is smaller than the accuracy ϵ which is desired. Due to the linearity of the expectation operator, the expectation on the finest grid can be expressed as:

$$E\left[P_{L_F}\right] = E\left[P_0\right] + \sum_{L=1}^{L_F} E\left[P_L - P_{L-1}\right] \ . \tag{5.1}$$

The quantity $E\left[P_L - P_{L-1}\right]$ represents the expected difference in the payoff approximation on levels L and $L-1$. This is estimated using a set of Brownian paths, with the same Brownian paths being used on both levels. This is where the strong convergence properties are crucial. The small difference between the terminal values for the paths computed on levels L and $L-1$ leads to a small value for the payoff difference. Consequently, the variance:

$$V_L = V[P_L - P_{L-1}] \ ,$$

decreases rapidly with level L. In particular, for a European option with a Lipschitz payoff, the order with which the variance converges to zero is double the strong order of convergence. Using M_L independent paths to estimate $E\left[P_L - P_{L-1}\right]$, if one defines the level 0 variance to be $V_0 = V[P_0]$ then the variance of the combined multilevel estimator is $\sum_{L=0}^{L_F} M_L^{-1} V_L$. The computational cost is proportional to the total number of time steps: $\sum_{L=0}^{L_F} M_L \Delta t_L^{-1}$. Varying M_L to minimize the variance for a given computational cost gives a constrained optimization problem whose solution is $M_L = C_M \sqrt{V_L \Delta t_L}$. The value for the constant of proportionality, C_M, is chosen to make the overall variance less than the ϵ^2, so that the r.m.s. error is less than ϵ.

The analysis in [10] shows that in the case of an Euler discretization with a Lipschitz payoff, the computational cost of the ML-MC algorithm is $O(\epsilon^{-2}\,(\log\epsilon)^2)$, which is significantly better than the $O(\epsilon^{-3})$ cost of the standard Monte Carlo method. Furthermore, the analysis shows that first order strong convergence should lead to $O(\epsilon^{-2})$ cost for Lipschitz payoffs; this will be demonstrated in the results to come which have been published in [30].

5.1.1 Pricing European Options using ML-MC

Consider the following four stochastic volatility and variance models presented in the first chapters:

$$dS = S\left(\overline{\mu}\,dt + \sigma\,d\widehat{W}_{1,t}\right)\ , \qquad \sigma = \nu^2\ .$$

- The Quadratic Volatility Model (Case 1)

$$d\sigma = k\,(\varpi_2 - \sigma)\,dt + \beta_2\,\sigma^2 d\widehat{W}_{2,t}\ . \tag{5.2}$$

- The 3/2 Model (Case 2, [25]);

$$d\nu = k\nu(\varpi_{3/2} - \nu)\,dt + \beta_{3/2}\,\nu^{3/2} d\widehat{W}_{2,t}\ . \tag{5.3}$$

- The GARCH Diffusion Model (Case 3)

$$d\nu = k\,(\varpi_1 - \nu)\,dt + \beta_1\,\nu d\widehat{W}_{2,t}\ . \tag{5.4}$$

- The Square Root Model (Case 4, [14])

$$d\nu = k\,(\varpi_{1/2} - \nu)\,dt + \beta_{1/2}\sqrt{\nu} d\widehat{W}_{2,t}\ . \tag{5.5}$$

The first set of numerical results are for a European option with strike K and maturity T, for which the payoff is given by:

$$P = \left\{ \begin{array}{ll} \max\left(S(T) - K, 0\right) & \text{for call options} \\ \max\left(K - S(T), 0\right) & \text{for put options} \end{array} \right\}\ . \tag{5.6}$$

Using the Case 2 volatility model (5.3) and a put option with strike $K = 1.1$, the ML-MC results in Figure 5.1 are obtained. The top left plot shows the weak convergence in the estimated value of the payoff as the finest grid level L is increased. All of the methods tend asymptotically to the same value. The bottom left plot shows the convergence of the quantity $V_L = V[P_L - P_{L-1}]$. The $3D-\ \theta$ scheme, defined in (4.35), exhibits second order convergence due to the first order strong convergence. The Milstein approximation with the Lévy areas set equal to zero (setting $\underline{\overline{L}}_{(1,2)} = 0$ in (4.4)) and the Euler discretization both give first order convergence, which is consistent with their 0.5 order strong convergence properties. We have used the following parameters: $t_o = 0$; $T = 1$; $\rho = -0.50$; $\overline{\mu} = 0.1$; $k = 1.4$; $\varpi_2 = 0.32^2$; $\beta = 2.44$, initial conditions: $S(t_o) = 1$; $\nu(t_o) = \varpi_2^2$.

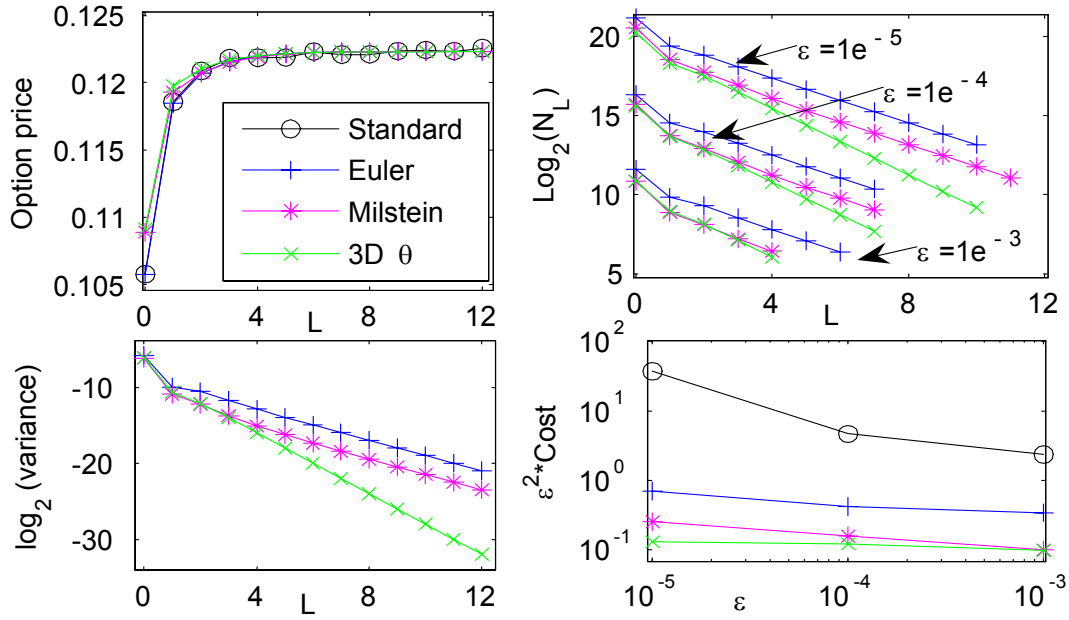

Figure 5.1: European put option, Case 2. Top left: convergence in option value with grid level. Bottom left: convergence in the ML-MC variance. Top right: number of Monte Carlo paths N_l required on each level, depending on the desired accuracy. Bottom right: overall computational cost as a function of accuracy ϵ.

The top right plot shows three sets of results for different values of the desired r.m.s. accuracy ϵ. The ML-MC algorithm uses the correction obtained at each level of time step refinement to estimate the remaining bias due to the discretization, and therefore determine the number of levels of refinement required [10]. The results illustrate the aforementioned, with the smaller values for ϵ leading to more levels of refinement. To achieve the desired accuracy, it is also necessary to reduce the variance in the combined estimator to the required level, so many more paths (roughly proportional to ϵ^{-2}) are required for smaller values of ϵ. The final point to observe in this plot is how many fewer paths are required on the fine grid levels compared to the coarsest grid level for which there is just one time step covering the entire time interval to maturity. This is a consequence of the variance convergence in the previous plot, together with the optimal choice for M_L described earlier.

The final bottom right plot shows the overall computational cost as a function of ϵ. The cost C_ϵ is defined as the total number of time steps, summed over all paths and all grid levels. It is expected that C_ϵ will be $O(\epsilon^{-2})$ for the best ML-MC methods and so the quantity which is plotted is $\epsilon^2 C_\epsilon$ versus ϵ. The results show that $\epsilon^2 C_\epsilon$ is almost perfectly independent of ϵ for the $3D-\theta$ scheme and varies only slightly with ϵ for the Milstein scheme. The Euler ML-MC scheme shows a bit more

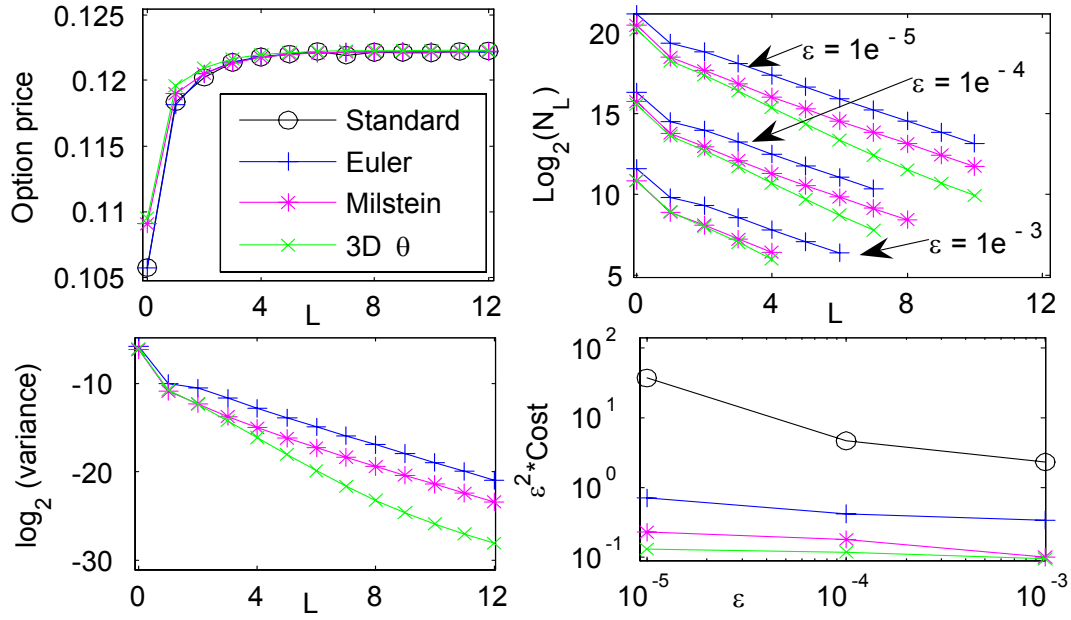

Figure 5.2: European put option, Case 3. Top left: convergence in option value. Bottom left: convergence in ML-MC variance. Top right: number of Monte Carlo paths N_l required on each level. Bottom right: overall computational cost.

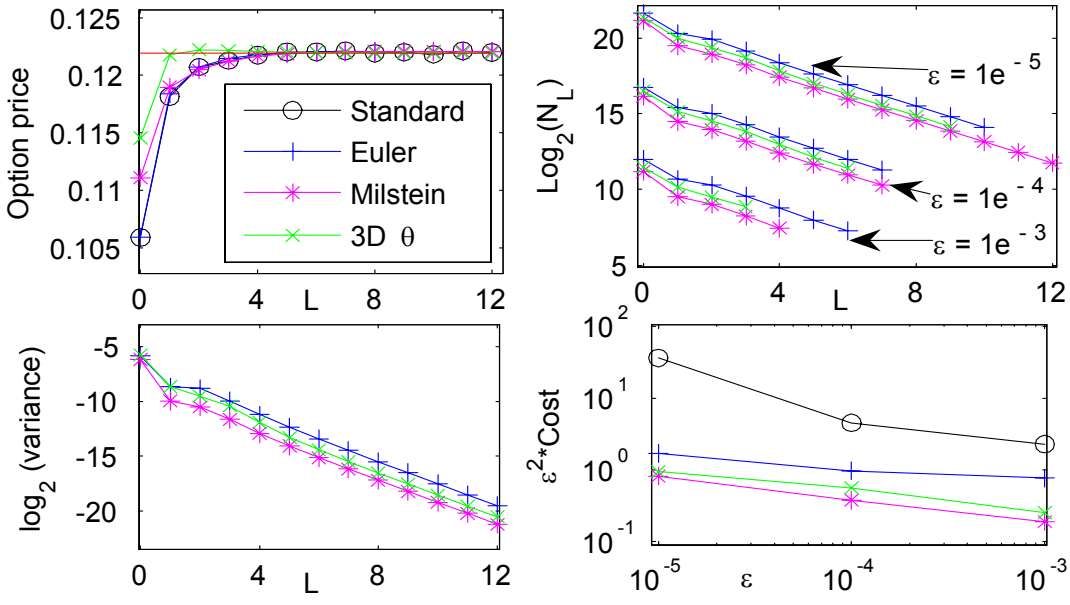

Figure 5.3: European put option, Case 4. Top left: convergence in option value (red line is analytic value). Bottom left: convergence in ML-MC variance. Top right: number of Monte Carlo paths N_l required on each level. Bottom right: computational cost.

growth as $\epsilon \to 0$, which is consistent with the analysis in [10], which predicts that $C_\epsilon = O(\epsilon^{-2}(\log \epsilon)^2)$. The final comparison line is the standard Monte Carlo method using the Euler discretization, for which $C_\epsilon = O(\epsilon^{-3})$.

The use of fewer Monte Carlo paths M_L is reflected directly in the computational cost of the process. For the most accurate case, $\epsilon = 10^{-5}$, the Euler, the Milstein and $3D-\ \theta$ schemes using the ML-MC algorithm are respectively approximately 50, 150 and 300 times more efficient (5.8) than the standard Monte Carlo method using the Euler discretization.

$$C_\epsilon^{\text{ML-MC}} = \sum_{L=0}^{L_F} \left(M_L 2^L\right) \ , \qquad C_\epsilon^{\text{Std Euler}} = 2\left(\frac{V[P_L]}{\epsilon^2}\right) 2^L \ , \tag{5.7}$$

$$\text{Savings}\,(\epsilon) = \frac{C_\epsilon^{\text{Std Euler}}}{C_\epsilon^{\text{ML-MC}}} \ . \tag{5.8}$$

Figures 5.2 and 5.3 show the corresponding results for Cases 3 and 4, corresponding to the GARCH Diffusion Model (5.4) and the Heston Model (5.5) respectively. For Case 3 , the computational savings (5.8) from using the ML-MC method are similar to Case 2 (the 3/2 Model), while for Case 4 the savings (5.8) from the Euler, Milstein and $3D-\ \theta$ scheme versions of the ML-MC scheme are roughly 20, 40 and 40, in the most accurate case. The parameters and initial conditions for Cases 3 and Case 4 are the same as in Case 2 except for $\beta = 0.78$ and $\beta = 0.25$; which are chosen so that x and y will have approximately the same relative volatility (see the Steady-State Probability Distribution section for more information (Chapter 2, Section 2.2.4)).

5.2 Multischeme Multilevel Monte Carlo Method (MSL-MC)

Strong convergence properties play a crucial role in Multilevel Monte Carlo path simulation method (ML-MC [10]). The better the strong convergence order γ and constant of proportionality $C_{\triangle t}$, the more efficient the ML-MC:

$$E\left[\ \left|S(T) - \widehat{S}(T, \triangle t)\right|\ \right] \leq C \triangle t^\gamma \ . \tag{5.9}$$

Chapters 3 and 4 demonstrate that using the SVMs (5.2-5.5), the Euler, the Malliavin[1] and the Milstein schemes with zero Lévy areas (setting $\overline{L}_{(1,2)} = 0$ in (4.4)) give a strong convergence order of 0.5. On the other hand, the use of $\theta-$scheme (orthogonal Milstein scheme) with zero Lévy areas can give either 0.5 or 1.0 strong convergence

[1]Scheme defined in chapter 4 (4.14) or it has been published in [6] and [26].

orders depending on the model parameters. When a proper value for the distribution of the Lévy area is simulated (through simulating the Lévy area using N subintervals within each time step) the Milstein scheme and the $3D-\theta$ scheme both give 1.0 order strong convergence. However, the constant of proportionality $C_{\triangle t}$ (5.9) changes depending on the parameters of the system and for some cases the Euler or the Malliavin scheme can give a better strong convergence error than the Milstein or θ scheme. Everything depends on the parameters and initial conditions of the problem. This is demonstrated more clearly in Figure (5.4) where the strong convergence tests for a European Call option price with various parameters (5.10) and SVMs (5.2-5.5) are presented.

Example	1	2	3	4	5	6	7
$T=$	10	1	1	1	0.2	0.2	1
$k=$	1	10	0.2	0.2	1	1	1
$\varpi_i=$	0.3^2	0.3^2	0.1^2	0.1^2	0.3^2	0.3^2	0.03^2
$\beta_i=$	0.2	0.2	0.2	3	1	3	0.2
Case $=$	4	4	3	3	2	2	4

$$t_o=0,\ S(t_o)=1,\ \rho=-0.50,\ \overline{\mu}=0.05,\ \nu(t_o)=\varpi_i,$$
$$K_{Call}=0.95S(t_o)e^{\overline{\mu}(T-t_o)},\ K_{Put}=1.05S(t_o)e^{\overline{\mu}(T-t_o)}\ . \tag{5.10}$$

In the top of Figure (5.4), example 1 ($EX1$) and example 2 ($EX2$), are the strong convergence tests using the Square Root Model (Case 4) and maturity T or the mean reverting speed k equal to 10. The graphics show a "lump" for big Δt. Pricing a European Call option using these parameters and an estimated error $\epsilon=10^{-2}$, the Euler scheme is the optimal scheme to use. By contrast, for $\epsilon\leq 10^{-4}$, the $3D-\theta$ scheme gives the best results. Using case 3 (SVM (5.4)) and small mean reverting speed k ($EX3$ & $EX4$ in (5.4)), the optimal scheme depends on the value of β. Using case 2 (SVM (5.3)) and small maturity T ($EX5$ & $EX6$), the θ scheme is first in computational time. For small mean ϖ ($EX7$), all schemes have poor behavior. For $\epsilon=10^{-2}$, the Euler scheme is the optimal scheme to use. However, for $\epsilon\leq 10^{-3}$, the Malliavin scheme gives the best results. In the Appendix of the thesis are the corresponding strong convergence tests for the asset S (Figure C.1), the variance υ (Figure C.2), the rotation or angle θ (Figure C.3) and the European Put option price (Figure C.4) using (5.10). It is no surprise that all strong convergence plots are almost the same, having the same order of convergence as the European Call option plot (Figure 5.4) presented in this example.

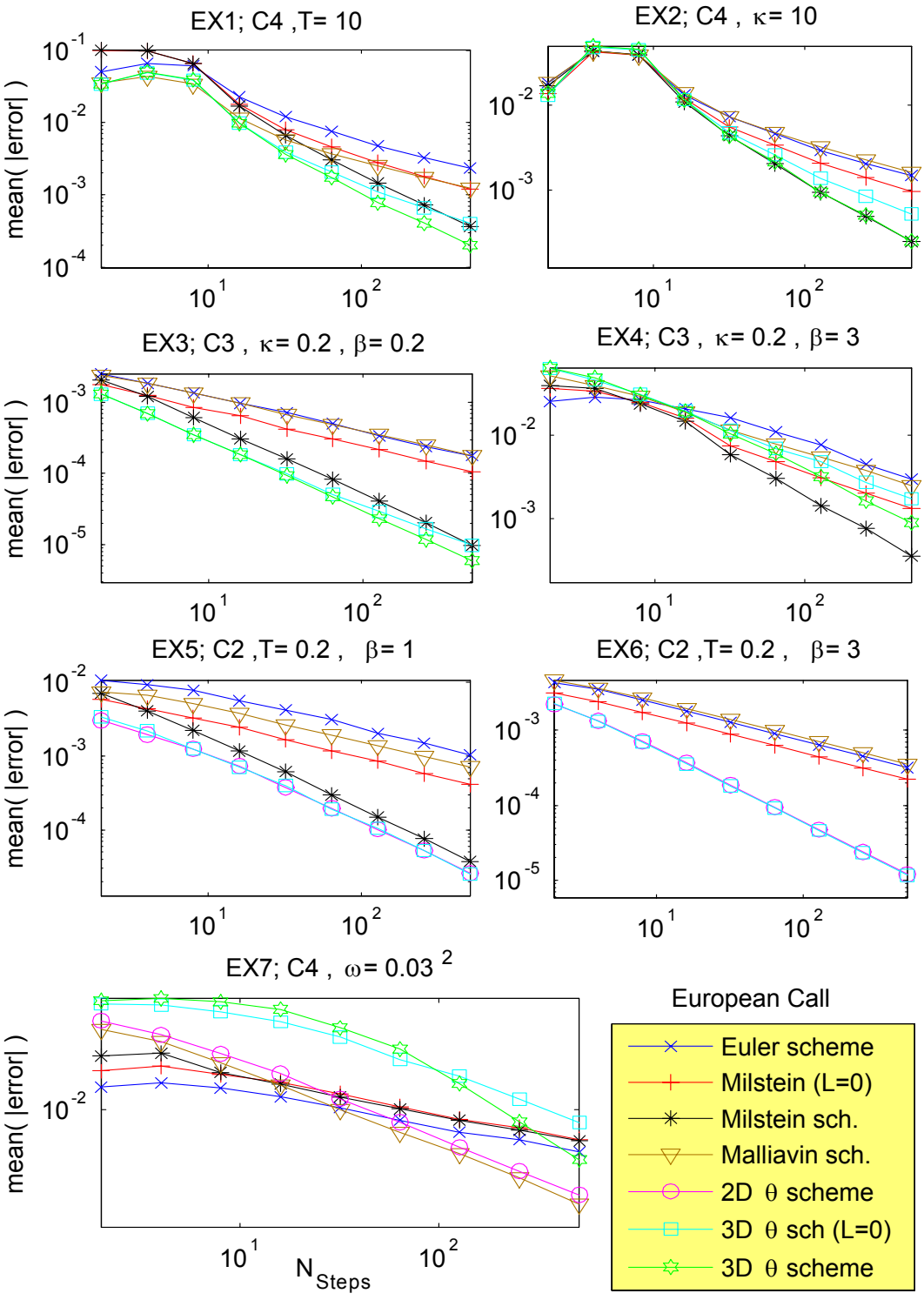

Figure 5.4: Strong convergence tests for a European Call option using (5.10).

When analyzing the option pricing problem in depth, the accuracy or error ϵ between the price option and the estimated price depends mainly on the characteristics or importance of the problem. The stochastic volatility model (SVM) and its parameters depend on the stock market data for the asset S. However, the scheme, the number of time steps and how many Monte Carlo paths are used to estimate the option price depends only on the method or algorithm applied. On the other hand, Figure (5.4) proves (as is well known in practice) that a single optimal scheme does not exist for general purposes. The selection of the scheme and the number of time steps depends totally on both the required accuracy of the problem and the parameters of the SVM. Therefore, the construction of an intelligent algorithm that can use different time approximations for different inputs will be found to be helpful.

5.2.1 Definition of the MSL-MC

Proposed algorithms when using different schemes:

A) Use the ML-MC method with an intelligent algorithm that, depending on the parameters of the SVM, can select both the optimal starting level L_O and the optimal scheme to calculate (5.1):

$$E\left[P_{L_F}\right] = E\left[P_{L_O}\right] + \sum_{L=L_O+1}^{L_F} E\left[P_L - P_{L-1}\right] \ . \tag{5.11}$$

B) Use the ML-MC method with an intelligent algorithm that, depending on the parameters of the SVM, can select both the optimal starting level L_O and the optimal scheme it uses in each level L to calculate (5.1). Because of the use of different schemes, (5.1) has to change to:

$$E\left[P_{L_F}\right] = E\left[P_{S_{L_O}}\right] + \sum_{L=L_O+1}^{L_F} E\left[P_{S_L} - P_{S_{L-1}}\right] \ , \tag{5.12}$$

where P_{S_L} is the payoff value using the optimal scheme for level L. In the Appendix of the thesis (page 134) a formal definition of the MSL-MC algorithm is presented.

5.2.2 Pricing European Options using MSL-MC

Consider the GARCH Diffusion Model (Case 3 (5.4)) and the proposed solution *A* (5.11) using different starting levels $L_O = 2, 3, 4, 5$ for the Milstein scheme and the $3D-\theta$ scheme (setting the Lévy Area equal to zero). When simulating the strong convergence test for the call option price (Figure (5.5)), the convergence in the ML-MC

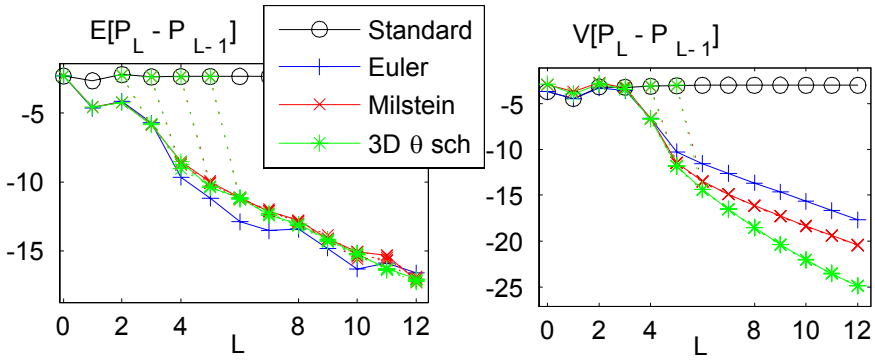

Figure 5.5: European option: Convergence in the MSL-MC mean and variance with grid level.

mean ($E[P_L - P_{L-1}]$) and the ML-MC variance ($V[P_L - P_{L-1}]$) with grid level does not change if one uses a non-zero starting level L_O. Because the Lévy Area is not simulated, all schemes give 1 order of strong convergence (which is consistent with their 0.5 order strong convergence properties) with different constant of proportionality. This example has the following parameters and initial conditions:

$$\begin{aligned} S(t_o) &= 1,\ \nu(t_o)=0.2^2,\ \varpi_2=0.3^2,\ t_o=0,\ \rho=-0.5,\ \overline{\mu}=0.05, \\ T &= 3,\ k=5,\ \beta=03,\ K_{Call}=0.95e^{\overline{\mu}(T=t_o)}. \end{aligned} \tag{5.13}$$

Calculating a call option using different accuracy or error ϵ demonstrates in Figure (5.6) that the number of Monte Carlo paths M_L changes when (5.11) has a non-zero starting level L_O. The use of fewer Monte Carlo paths M_L is reflected directly in the computational cost of the process (simulation time). The computational cost of the process C_ϵ is defined as the total number of time steps, summed over all paths and all grid levels (5.7). For the most accurate case, $\epsilon = 10^{-4}$, the Euler, the Milstein and $3D-\ \theta$ schemes ($L_O = 0$) are roughly 3.4, 3 and 3.6 times more efficient (5.8) than the standard Monte Carlo method using the Euler discretization. On the other hand, using a starting level ($L_O = 3$), the Milstein and $3D-\ \theta$ schemes are respectively approximately 10.5 and 12.5 more efficient (5.8) in the most accurate case. It is important to note that if you start on level 0, for $\epsilon = 10^{-2}$ and $\epsilon = 10^{-3}$, the ML-MC gives you equal or poorer computational cost than the standard Euler method. This is because of the strong convergence properties the example gives for big Δt (Figure 5.5). These results show the importance of starting at the right level in (5.11).

Another important result to mention for Figure 5.6 is the computational European option price for different accuracy or error ϵ. For $\epsilon \leq 10^{-2}$, one can see that

all schemes give different estimated prices, however they are inside the boundaries or limits required $\left(P = \widehat{P} \pm \epsilon\right)$. As $\epsilon \to 0$, all schemes converge to the same computational price. The MSL-MC algorithm[2] stops when the estimated option price is inside the boundaries.

The computation time for each scheme to complete one subroutine is another important factor to consider in the selection of the optimal scheme. Each scheme takes different computational time to complete the simulation. This is because they have extra terms or more equations to calculate in one subroutine. Table 5.1 shows that when $L_O = 3$ the MSL-MC gives the best computation time for all ϵ (2.7 times faster than when $L_O = 0$). Because the $3D-\ \theta$ scheme takes roughly 1.9 times more to complete one subroutine, the Milstein scheme is the optimal scheme to use for this example. If one wants better accuracy for the option price, e.g. $\epsilon = 10^{-5}$, the $3D-\ \theta$ scheme will be the optimal scheme.

Scheme	L_O	$\epsilon = 10^{-2}$	$\epsilon = 10^{-3}$	$\epsilon = 10^{-4}$
Euler (Standard)	n/a	0.007	1.195	977.5
Milstein (Standard)	n/a	0.007	1.216	987.6
$3D-\ \theta$ scheme (Standard)	n/a	0.013	2.200	1788.7
Euler scheme	$L_O = 0$	0.011	1.036	126.0
Milstein scheme	$L_O = 0$	0.014	1.299	145.0
Milstein scheme	$L_O = 2$	0.007	0.808	87.3
Milstein scheme	$L_O = 3$	**0.006**	**0.379**	**46.4**
Milstein scheme	$L_O = 4$	0.010	0.425	49.8
Milstein scheme	$L_O = 5$	0.020	0.751	82.6
$3D-\ \theta$ scheme	$L_O = 0$	0.024	2.214	231.3
$3D-\ \theta$ scheme	$L_O = 2$	0.013	1.342	142.6
$3D-\ \theta$ scheme	$L_O = 3$	**0.011**	**0.703**	**77.1**
$3D-\ \theta$ scheme	$L_O = 4$	0.020	0.774	85.3
$3D-\ \theta$ scheme	$L_O = 5$	0.039	1.417	146.0

Table 5.1: Computation time for a European option using the MSL-MC (minutes).

The computational time using the scheme S_j and the standard Monte Carlo method can be calculated roughly by:

$$Time_{Std}^{S_j} \approx \left(\frac{V[P_L]}{\epsilon^2/2}\right) Time_L^{S_j} \ ,$$

where $Time_L^{S_j}$ is the simulation time for one Monte Carlo subroutine using the the scheme S_j and $\Delta t = 2^L$.

[2] In the Appendix of the thesis (page 134) a formal definition of the MSL-MC algorithm is presented.

5.2.3 Digital Option

The payoff for a digital option is given by:

$$P = \left\{ \begin{array}{ll} H\left(S(T) - K\right) & \text{for call options} \\ H\left(K - S(T)\right) & \text{for put options} \end{array} \right\},$$

where $H(x)$ is the Heaviside function ($H(x) = 1$ if $x > 0$, else $H(x) = 0$). Figure 5.7 shows the results of pricing a Digital option using the 3/2 Model model (Case 2 (5.3)) and the MSL-MC. The parameters and initial conditions are the same as the European example (5.13) except for $T = 0.2$, $\kappa = 1$ and $\beta = 3$. Because this payoff is not Lipschitz continuous, it shows the poorest benefits from the MSL-MC approach. For the most accurate case, $\epsilon = 10^{-4}$, the Euler, the Milstein, $2D-\ \theta$ scheme and $3D-\ \theta$ scheme (Lévy area equal to zero) using the MSL-MC algorithm are respectively approximately 2.5, 10, 18 and 25 times more efficient (5.8) than the standard Monte Carlo method using the Euler discretization. The difference in savings in θ schemes is because we are using $\mu^{(\theta)} = 0$ in the $3D-\ \theta$ scheme. It would be as efficient if we use (4.26).

Because these parameters give a linear variance reduction (Figure 5.7), applying a non-zero starting level ($L_O = 1$) to calculate (5.11) does not provide any improvement in the option price simulation and in some cases it can be less efficient. The computational cost is reflected directly in the simulation time one requires to calculate the option price with a certain accuracy ϵ (Table 5.2). This example shows the importance of the $2D-\ \theta$ scheme using the MSL-MC which is 6 or 2 times faster than the Euler or Milstein schemes respectively.

Scheme	$\epsilon = 10^{-3}$	$\epsilon = 10^{-4}$
Euler (Standard simulation)	0.28	226.9
Euler scheme ($L_O = 0$)	0.29	174
Milstein scheme ($L_O = 0$)	0.15	53.5
$2D-\ \theta$ scheme ($L_O = 0$)	**0.08**	**29.1**
$3D-\ \theta$ scheme ($L_O = 0$)	0.10	47.2
$2D-\ \theta$ scheme ($L_O = 1$)	0.13	35.5

Table 5.2: Computation time for a Digital option using the MSL-MC (minutes).

5.2.4 Multi-Options

Combinations of options are frequently used in the market. Using the appropriate portfolio allows the buyer to fix a strategy depending on his expectation of the market.

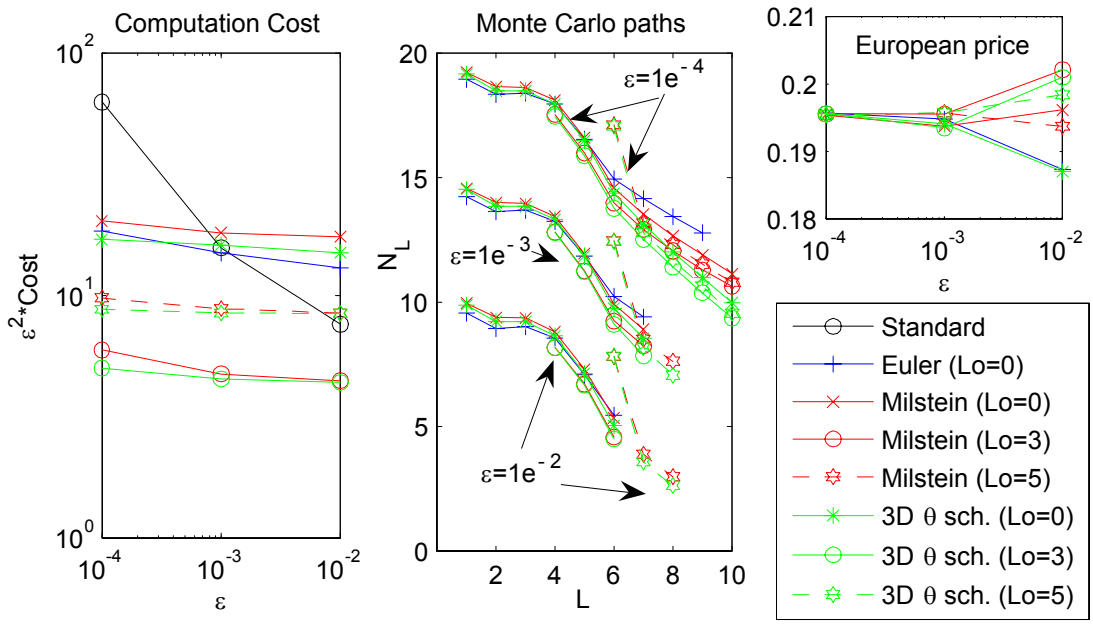

Figure 5.6: European option: Left: overall computational cost. Middle: number of Monte Carlo paths N_l required on each level. Right: convergence in computational option value for different ϵ.

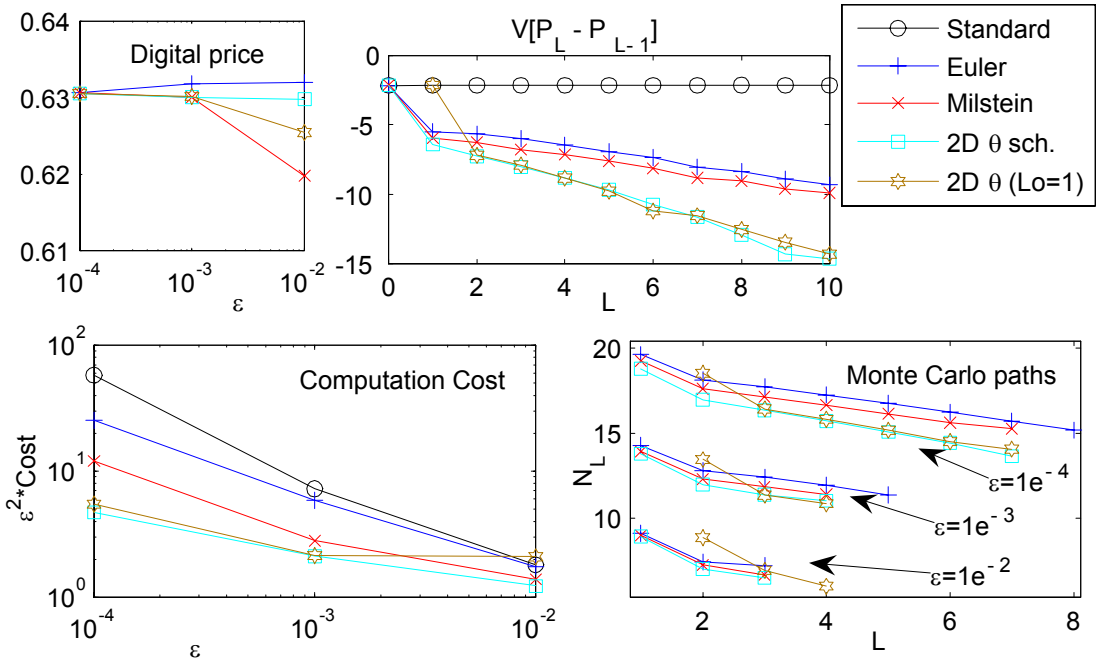

Figure 5.7: Digital option. Top left: convergence in computational option value for different ϵ. Bottom left: overall computational cost. Top right: convergence in MSL-MC variance. Bottom right: number of Monte Carlo paths N_l required on each level.

The payoff for a multi-European option is given by:

$$P = \#_C P_C(T,K) + \#_P P_P(T,K) \ ,$$

where P_C and P_P are Call and Put European options (5.6) with strike price K and maturity T . $\#_C$ and $\#_P$ are the number of call and put options in the portfolio. The most frequent and simple combination when $K_C = K_P$ is called "**Strips**" or "**Straps**". Strip derivatives use 1 call and 2 put options and the point of view of the market is that the stock price at maturity will finish below or above the strike price, below more likely. Strap derivatives use 1 put and 2 call options and the point of view of the market is the same as a Strip option, however above the strike price is more likely. When $K_C > K_P$ a "**Strangle**" can be obtained. Another famous combination of Vanilla options is called a "**Butterfly spread**" which has a payoff equal to:

$$P = P_C(T,K_C) + P_P(T,K_P) - P_C(T,K_A) - P_P(T,K_A) \ ,$$

$$\text{where:} \quad K_A = \frac{K_C + K_P}{2} \ .$$

Figure 5.8 shows the option price for a Strip derivative using the MSL-MC algorithm, the Quadratic Volatility Model (Case 1 (5.4)) and the same parameters as the examples above (5.13), except for $T=1$, $\kappa = 10$, $\beta=0.5$ and $K_{Call} = K_{Put}$. For the most accurate case, $\epsilon = 10^{-4}$, the Euler scheme, the Milstein scheme, $2D-\ \theta$ scheme and $3D-\ \theta$ scheme are only roughly 3 times more efficient (5.8) than the standard method using the Euler scheme. However, using a non-zero starting level ($L_O = 3$), the $2D-\ \theta$ scheme and $3D-\ \theta$ scheme are 19 and 21 times more efficient than the standard method. This example shows again that the use of $\mu^{(\theta)} = 0$ instead of $\mu^{(\theta)} =$(4.26) in the $3D-\ \theta$ scheme is the optimal solution.

On the other hand, using the same parameters as before (5.13) except for $\kappa = 0.4$ and $K_{Put} = 1.1e^{\overline{\mu}(T-t_o)}$, Figure 5.9 shows the option price for a Butterfly derivative. For the most accurate case, $\epsilon = 10^{-5}$, the Euler scheme, the Milstein scheme, $2D-\ \theta$ scheme and $3D-\ \theta$ scheme are 10, 46, 80, 112 times more efficient than the standard method using the Euler scheme. However, using a non-zero starting level ($L_O = 3$), the $2D-\ \theta$ scheme and $3D-\ \theta$ scheme are 48 and 52 times more efficient than the standard method. In contrast to the Strip option price (Figure 5.8), this example does not give any improvement through the use of a non-zero starting level. The simulation times to calculate the option prices with a certain accuracy ϵ are presented in Table 5.3. Both examples show the importance of analyzing the parameters of the model

before simulation to make the right decision when selecting the scheme and starting level in (5.11).

Scheme	Strip $\epsilon = 10^{-3}$	Strip $\epsilon = 10^{-4}$	Butterfly $\epsilon = 10^{-5}$
Euler (Standard)	13.8	104.7	28.39
Euler scheme ($L_O = 0$)	0.56	64.1	13.23
Milstein scheme ($L_O = 0$)	0.75	79.3	3.36
$2D-\ \theta$ scheme ($L_O = 0$)	1.31	135.1	2.85
$3D-\ \theta$ scheme ($L_O = 0$)	1.20	122.6	**2.28**
$2D-\ \theta$ scheme ($L_O = 3$)	**0.18**	**20.2**	4.98
$3D-\ \theta$ scheme ($L_O = 3$)	0.20	21.6	5.12

Table 5.3: Computation time for Multi-options using the MSL-MC (minutes).

5.2.5 Asian Option

Asian options are another type of exotic option. They have a payoff that depends on some average property of the asset price over life, or part of the life, of the option and is given by:

$$P = \left\{ \begin{array}{ll} \max\left(\overline{S}(T) - K,\ 0\right) & \text{for call options} \\ \max\left(K - \overline{S}(T),\ 0\right) & \text{for put options} \end{array} \right\} .$$

Where $\overline{S}$ is either, the arithmetic average which can be approximated numerically as:

$$\overline{S}(T) = \frac{1}{T}\int_0^T S(t)dt \approx \frac{\Delta t}{2\,T}\sum_{n=1}^{N_{\Delta t}} \left(\widehat{S}_n + \widehat{S}_{n-1}\right) , \tag{5.14}$$

or the geometric average which can be calculated numerically as:

$$\overline{S}(T) = \left(\prod_{n=1}^{N_{\Delta t}} \widehat{S}_n\right)^{1/N_{\Delta t}} . \tag{5.15}$$

The average is less volatile than the asset itself, so options may be cheaper and less subject to manipulation. Asian options may be found embedded in structured products. Using the GARCH Diffusion Model (Case 3 (5.4)) and the arithmetic average of $\overline{S}$ (5.14), Figure 5.10 shows that for the most accurate case, $\epsilon = 10^{-5}$, the Euler, the Milstein and $3D-\ \theta$ schemes using the ML-MC algorithm are respectively approximately 67, 90 and 115 times more efficient (5.8) than the standard method using the Euler scheme. On the other hand, taking the MSL-MC approach B (5.12) with zero starting level ($L_O = 0$), the results are disappointing. Unfortunately, when

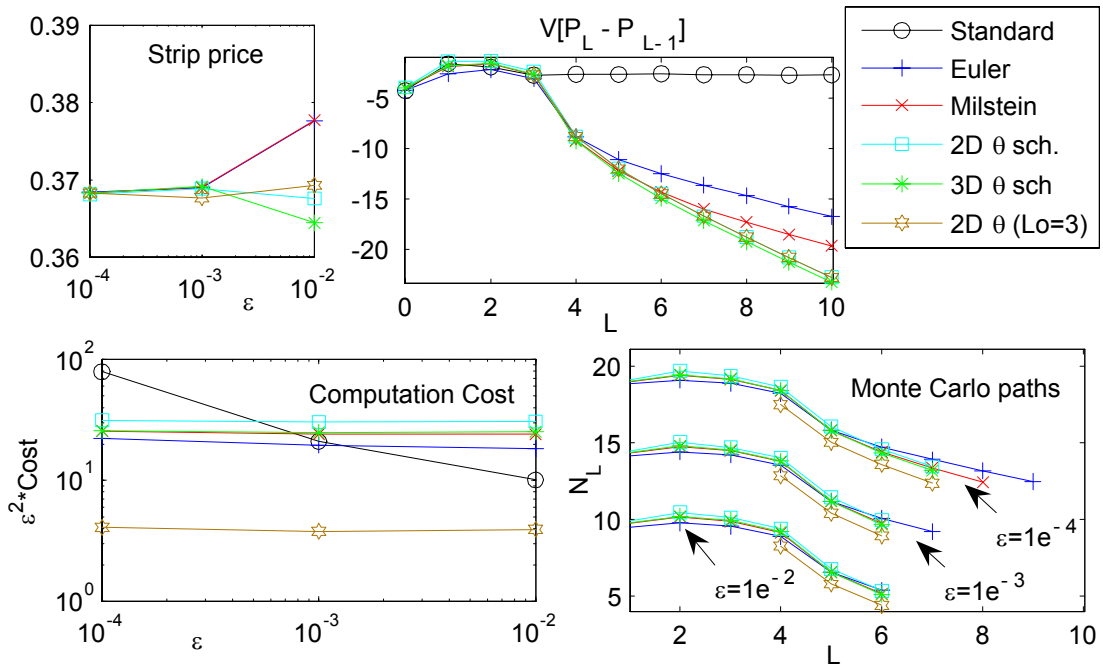

Figure 5.8: Strip Option. Top left: convergence in computational option value for different ϵ. Bottom left: overall computational cost. Top right: convergence in MSL-MC variance. Bottom right: number of Monte Carlo paths N_l required on each level.

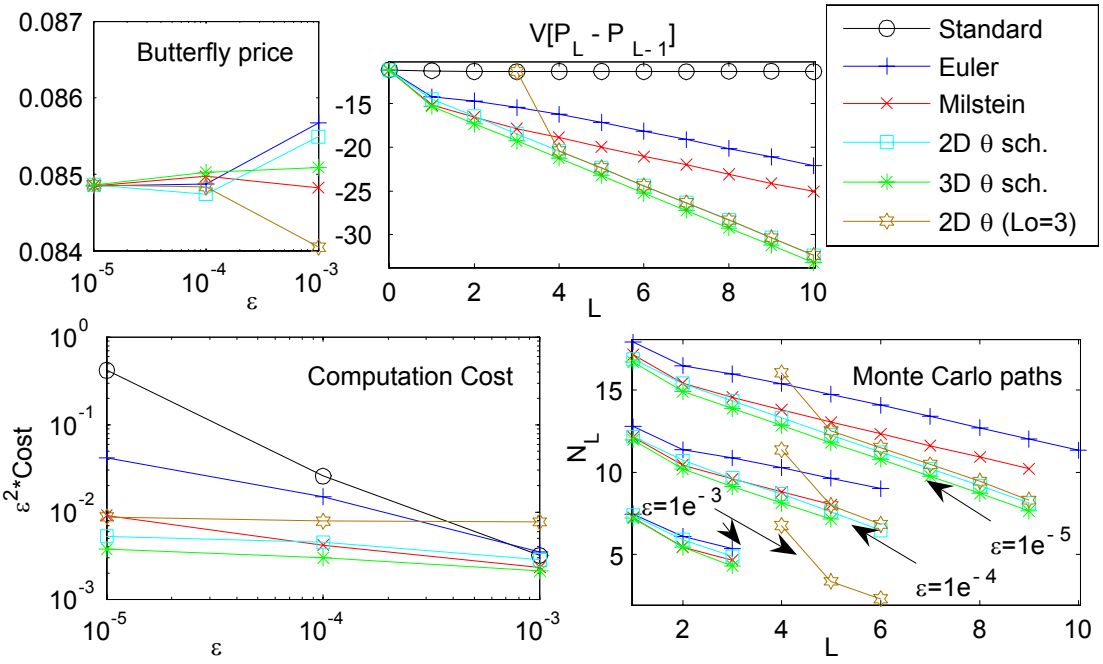

Figure 5.9: Butterfly Option. Top left: convergence in computational option value for different ϵ. Bottom left: overall computational cost. Top right: convergence in MSL-MC variance. Bottom right: number of Monte Carlo paths N_l required on each level.

one makes the change of scheme at level L in (5.12), the difference between the payoffs using different schemes:

$$\left[P_{S_L} - P_{S_{L-1}}\right] \tag{5.16}$$

is bigger than if one uses the same scheme. As a result of this, the MSL-MC algorithm require more Monte Carlo paths to calculate (5.16) and increase the simulation cost in the option price (Figure 5.10 and Table 5.4). We have used the Milstein scheme if $0 < L \leq 6$ and the $3D-\theta$ scheme for the rest of the levels ($L > 6$). The parameters and initial conditions for this example are the same as all examples presented above (5.13), except for $T=1$, $\kappa = 1$, $\beta=0.5$ and $K_{Put} = E\left[S(t_o)\right]$. The simulation times to calculate the option price with a certain accuracy ϵ are presented in Table 5.4. Using the MSL-MC method (5.11), this example shows the importance of considering the use of different schemes depending on the accuracy or error ϵ between the price option and the estimated price. Unfortunately, these results also demonstrate that the MSL-MC method (5.12) that uses different schemes at different level L converges to the right price but does not help to improve the computation cost of the process.

5.2.6 Variance Swap Option

A variance swap on an interval $[0, T]$ is a derivative contract on an underlying asset that has payoff given by:

$$P = N\left(\overline{\nu}(T) - K_{var}\right) \ ,$$

where $\overline{\nu}(T)$ is the average of the variance in the time interval $[0, T]$, K_{var} is a fair price of variance of the underlying over the period $[0, T]$ and N is the notional amount or nominal price of the swap. The definition of the realized variance is specified in the contract but in general, it can be approximated numerically in the same way as $\overline{S}(T)$ in the previous example (5.14-5.15).

Using the GARCH Diffusion Model (Case 3 (5.4)) and the arithmetic average of $\overline{v}$ (5.14), Figure 5.11 and Table 5.4 show that for the most accurate case, $\epsilon = 10^{-4}$, the Euler, the Milstein and $3D-\theta$ schemes using the ML-MC algorithm are respectively approximately 7, 9 and 2 times more efficient (5.8) than the standard method using the Euler scheme. As mention in examples above, to improve the computational cost, one needs to use a non-zero starting level in (5.11). Using the Milstein scheme and $Lo = 4$, the MSL-MC is 55 times more efficient than the standard method. The simulation times to calculate the option price with a certain accuracy ϵ are presented in Table 5.4. The parameters and initial conditions are the same as the

Asian example (5.13) except for $T=10$, $\beta=0.2$, $K_{var}=\varpi_2$ and $N=1$. As expected, in this case the Milstein method gives first order strong convergence for ν, whereas the $3-$Dimensional θ scheme gives similar accuracy initially but tails off towards order 0.5 strong convergence on the finest grids.

	Asian	Asian	Swaps	Swaps
Scheme	$\epsilon = 10^{-4}$	$\epsilon = 10^{-5}$	$\epsilon = 10^{-3}$	$\epsilon = 10^{-4}$
Euler (Standard)	6.12	478	0.13	82.12
Euler scheme ($L_O = 0$)	0.52	64.16	0.10	12.41
Milstein scheme ($L_O = 0$)	**0.44**	51.36	0.12	10.52
$3D-$ θ scheme ($L_O = 0$)	0.67	**41.23**	0.45	75.32
Milstein scheme ($L_O = 4$)	n/a	n/a	**0.03**	**2.10**
Multi–scheme ($L_O = 0$)	0.90	95.93	n/a	n/a

Table 5.4: Computation time for Asian and Variance Swap options using the MSL-MC (minutes).

5.3 Conclusions

In finance, stochastic variance and volatility models are very important for the valuation of exotic options. The Multilevel Monte Carlo path simulation method (ML-MC [10]) works without any problems with all schemes and calculates the right price for all exotic options presented in this chapter. It is a powerful tool, and in combination with the new θ scheme, can substantially reduce the computational cost in pricing options, lowering the cost required to achieve an r.m.s. error of size ϵ from $O(\epsilon^{-3})$ to $O(\epsilon^{-2})$) for some cases. The Multischeme Multilevel Monte Carlo (MSL-MC) is an improved/updated version of the ML-MC algorithm that, depending on the parameters of the stochastic volatility models (SVM) and accuracy or error ϵ between the price option and the estimated price, can select both the optimal starting level and the optimal scheme. Unfortunately, the use of different schemes at different levels does converge to the right price but does not help improve the computation cost of the process. Pricing exotic option examples demonstrate considerable computational savings when both the θ scheme and the MSL-MC are applied to stochastic volatility models in order to price exotic options.

When one reviews all the exotic option pricing examples presented in this chapter, we can conclude that the ML-MC have to be improved to obtain even better savings in the computation time. It is important to analyze the parameters of the model before simulation to make the right decision in the selection of the scheme and starting

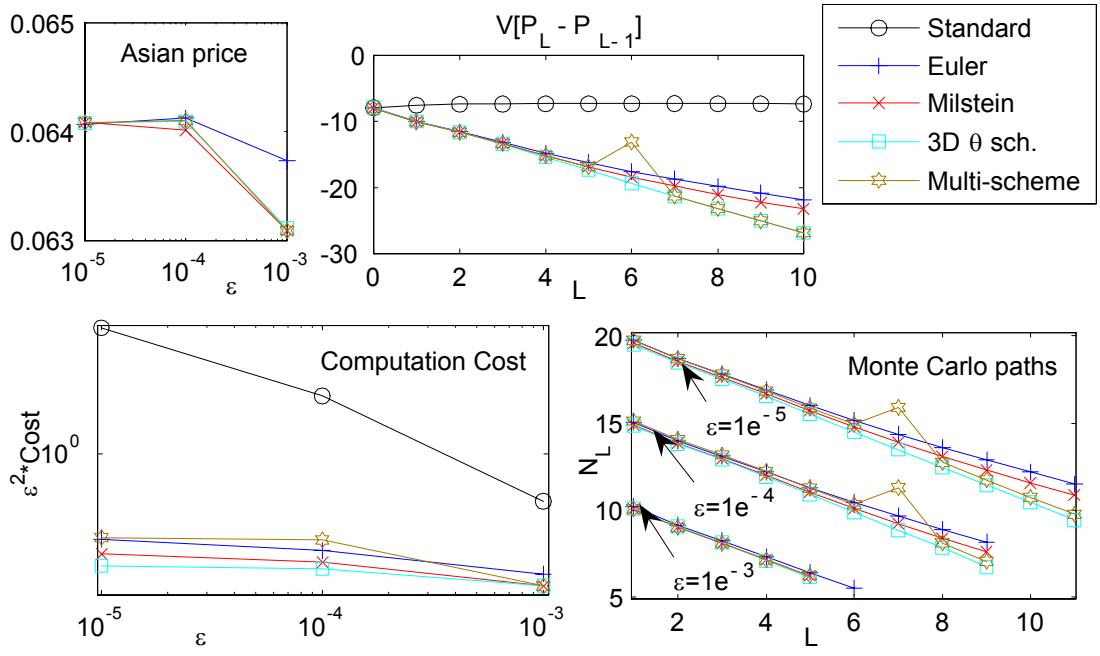

Figure 5.10: Asian option. Top left: convergence in computational option value for different ϵ. Bottom left: overall computational cost. Top right: convergence in MSL-MC variance. Bottom right: number of Monte Carlo paths N_l required on each level.

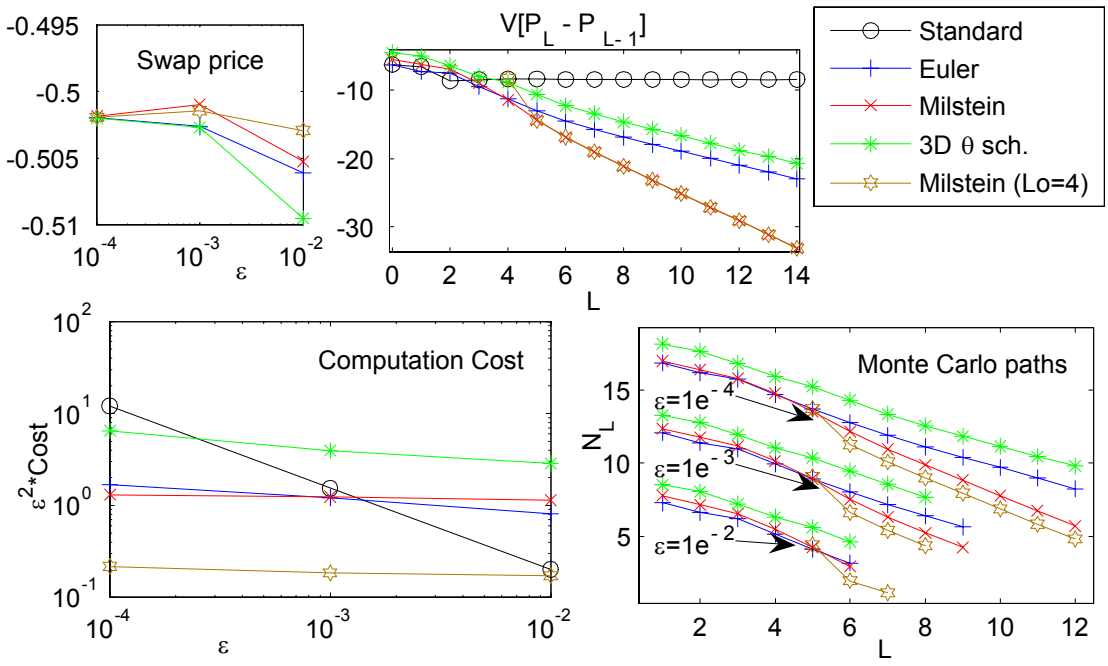

Figure 5.11: Variance swap option. Top left: convergence in computational option value for different ϵ. Bottom left: overall computational cost. Top right: convergence in MSL-MC variance. Bottom right: number of Monte Carlo paths N_l required on each level.

level. Figures 5.6 and 5.8 show the importance of starting at the right level in (5.11). Figures 5.7 and 5.9 demonstrate the importance of the θ scheme using the MSL-MC method (5.11) which is 6 or 2 times faster than the Euler or Milstein schemes respectively. Figure 5.10 and 5.11 shows the importance of considering the use of different schemes depending on the accuracy or error ϵ between the price option and the estimated price. In conclusion, the MSL-MC method provides better or equal savings or computational cost that the ML-MC if you use the correct scheme and starting level in the algorithm.

Chapter 6

Outlook and Extensions

The prices of exotic options given by models based on Black-Scholes assumptions can be wildly inaccurate because they are frequently even more sensitive to levels of volatility than standard European calls and puts. Therefore currently traders or dealers of these financial instruments are motivated to find models to price options which take the volatility smile and skew into account. To this extent, stochastic volatility models are partially successful because they can capture, and potentially explain, the smiles, skews and other structures which have been observed in market prices for options. Indeed, they are widely used in the financial community as a refinement of the Black-Scholes model. A strong example of the existence of random correlated volatility is when the historic volatility of the Stock Exchange index is plotted (Figure 2.4). This evidence shows that stock volatility is not constant at all and moreover that volatility shocks persist through time. This conclusion was reached by many authors in the literature; stochastic volatility models are needed to describe and explain volatility patterns.

When one analyses the steady-state probability distribution of the stochastic volatility models that are outlined in the literature, you can conclude that despite some similarities, all SVMs are important and have different properties. The definition of a more general stochastic volatility model (2.18) that represents all of them is necessary for the study and understanding of the option price properties. The selection of the parameters in (2.18) will depend on the properties of the real data one wants to match or simulate.

Strong convergence properties of discretizations of stochastic differential equations (SDEs) are very important in stochastic calculus. If one applies any discrete approximation scheme to a stochastic process and wants to numerically evaluate the strong or weak convergence order of our approximation $\widehat{X}(T)$, an exact solution $X(T)$ is normally required. However, at present, there are no solutions available for many SDEs.

The use of Theorems 2 and 3, "Strong and Weak Convergence Order without an Exact Solution", successfully determines the strong and weak orders of convergence. Each theorem was tested using exact solutions or expectations to verify the results. Examples in Chapter 2 demonstrate that the use of both theorems require at least 100 times fewer Monte Carlo paths than the standard method to correctly calculate the order of convergence. Because there are no exact solutions for the stochastic volatility model (2.18), the use of the theorems was fundamental to establish its convergence order.

Numerical examples in the thesis demonstrate, as expected, a 0.5 and 1.0 strong order of convergence for Euler and Milstein schemes respectively. Conversely, a 1.0 weak order of convergence for all schemes is obtained (the same expectation error). The application of either the Euler or Milstein schemes to calculate an expectation in the standard way has negligible difference at all in the outcome. To obtain a 1.0 strong order of convergence with the Milstein scheme, one has to apply the scheme to the vector form of the SDE, use independent Wiener processes and compute correctly the double integral or Lévy Area.

We have shown that under certain conditions the use of the orthogonal θ scheme can achieve the first order strong convergence properties of the Milstein numerical discretization without the expensive simulation of Lévy areas. Conversely, the Milstein scheme with zero Lévy Area has a 0.5 strong order convergence. The bias or error in the computation of the rotation angle θ that makes the Lie bracket equal to zero in the orthogonal scheme is crucial to obtain a better convergence order. When the conditions for integrability are satisfied, one can use the formula for θ to obtain the value of the rotation angle and obtain first order strong convergence. Otherwise, one has to use the $3-$Dimensional transformation and check the magnitude of the Lie brackets to decide if it is likely to give computational savings in the solution of our system. The numerical results in chapter 4 and 5 demonstrate a better strong order of convergence than the standard Milstein scheme when an orthogonal transformation is applied.

In finance, stochastic variance and volatility models are very important for the valuation of exotic options. The Multilevel Monte Carlo path simulation method (ML-MC [10]) works without any problems with all schemes and calculates the right price for all exotic options presented in this chapter. It is a powerful tool, and in combination with the new θ scheme, can substantially reduce the computational cost in pricing options, lowering the cost required to achieve an r.m.s. error of size ϵ from

$O(\epsilon^{-3})$ to $O(\epsilon^{-2})$) for some cases. The Multischeme Multilevel Monte Carlo (MSL-MC) is an improved/updated version of the ML-MC algorithm that, depending on the parameters of the stochastic volatility models (SVM) and accuracy or error ϵ between the price option and the estimated price, can select both the optimal starting level and the optimal scheme. Unfortunately, the use of different schemes at different levels does converge to the right price but does not help improve the computation cost of the process. Pricing exotic option examples demonstrate considerable computational savings when both the θ scheme and the MSL-MC are applied to stochastic volatility models in order to price exotic options. The MSL-MC method provides equal or better computational cost than the ML-MC.

In summary, this thesis proposes a better numerical approximation to calculate solutions for multi dimensional SDE's than the standard Monte Carlo integration. We introduce a new scheme or discrete time approximation where, for some conditions, a better strong convergence order is obtained than that using the standard Milstein scheme without the simulation of the expensive Lévy Area. We demonstrate when the conditions of the $2-$Dimensional problem permit this and give an exact solution for the orthogonal transformation (θ Scheme or Orthogonal Milstein Scheme).

Standard convergence theory for numerical methods for SDEs (e.g. as in Kloeden and Platen [22]) makes a global Lipschitz assumption on the coefficients. However, most of the SDE models that are mentioned in the thesis, and used in the computational experiments, do not satisfy such global Lipschitz conditions. The numerical results presented in Chapter 4 and 5 are not simply confirming a theory that has been proved, they are giving numerical evidence that the conclusions about strong order remain true in circumstances where no theory currently exists.

Using a wide variety of pricing exotic option examples we demonstrate that considerable computational savings can be made by using the new θ Scheme [30] and the improved Multischeme Multilevel Monte Carlo method (MSL-MC). The computational cost to achieve an accuracy of $O(\epsilon)$ is reduced from $O(\epsilon^{-3})$ to $O(\epsilon^{-2})$ for some parameters or Lipschitz conditions. A general stochastic volatility model that represents most of the stochastic volatility models that are outlined in the literature is proposed. Because it does not have a closed-form solution for option prices (as usual), we have demonstrated and tested with numerical examples two theorems that measure with confidence the order of strong and weak convergence of schemes without an exact solution or expectation of the system. We have focused our research on

continuous time diffusion models for the volatility and variance with their discrete time approximations (ARV).

For future work, we recommend that for multi dimensional SDE's such as portfolios with a large variety of exotic options, the investigation and test of the multi dimensional θ scheme would be very interesting. For some parameters, it will be obvious that the new orthogonal scheme will provide considerable computational time savings when calculating the strong and weak solutions and therefore find use in the calculation of the expectation price of the portfolio.

The future work in the pricing of exotic options could take many paths. Primarily, a further study should be undertaken into the improved MSL-MC algorithm using different schemes to improve the computational cost. The use of quasi-Monte Carlo methods will definitely help improve the computational cost in the MSL-MC. For non Lipschitz payoffs or payoffs dependent on the asset price or volatility of the process, the use of advanced tools for their calculation, as suggested by Giles [11] or Glasserman [12], is highly recommended.

Appendix A

Stochastic Volatility

A.1 Mathematical Definitions

A.1.1 Ornstein-Uhlenbeck or Gauss-Markov Process

By definition, a stochastic process $\{Y_T : t \succeq 0\}$ is:

- **stationary** if, for all $t_1 < t_2 < ... < t_n$ and $h > 0$, the random $n-$vectors $(Y_{t_1}, Y_{t_2}, ..., Y_{t_n})$ and $\left(Y_{t_{1+h}}, Y_{t_{2+h}}, ..., Y_{t_{n+h}}\right)$ are identically distributed. That is, changes in time do not modify the probability or distribution.

- **Gaussian** if, for all $t_1 < t_2 < ... < t_n$, the n vector $(Y_{t_1}, Y_{t_2}, ..., Y_{t_n})$ is multivariate normally distributed.

- **Markovian** if, for all $t_1 < t_2 < ... < t_n$, the $P(Y_{t_n} \leq y | Y_{t_1}, Y_{t_2}, ..., Y_{t_{n-1}}) = P(Y_{t_n} \leq y | Y_{t_{n-1}})$. That is, the future is determined only by the present and not the past.

Also, a process $\{Y_T : t \succeq 0\}$ is said to have **independent increments** if, for all $t_0 < t_1 < t_2 < ... < t_n$, the random variables $Y_{t_1} - Y_{t_0}$, $Y_{t_2} - Y_{t_1}, ..., Y_{t_n} - Y_{t_{n-1}}$ are independent. This condition implies that $\{Y_T : t \succeq 0\}$ is Markovian, but not conversely. Furthermore, the increments are said to be stationary if, for any $t > s$ and $h > 0$, the distribution of $(Y_{t+h} - Y_{s+h})$ is the same as the distribution of $(Y_t - Y_s)$. This additional provision is needed for the following definition.

A stochastic process $\{W_T : t \succeq 0\}$ is a **Wiener-Lévy process or Brownian motion** if it has stationary independent increments, if W_T is normally distributed, the $E(W_t) = 0$ for each $t > 0$, and if $W_0 = 0$. It then follows that $\{W_T : t \succeq 0\}$ is Gaussian and that $Cov(W_t, W_s) = \sigma^2 \min\{t, s\}$, where the variance parameter σ^2 is a

positive constant. Almost all paths of Brownian motion are continuous but nowhere differentiable.

One technical stipulation is required for the following. A stochastic process $\{Y_T : t \succeq 0\}$ is **continuous in probability** if, for all $u \in \Re^+$ and $\varepsilon > 0$,

$$P(|Y_v - Y_u| \geq \varepsilon) \to 0 \text{ as } v \to u \ .$$

This holds true if $Cov(Y_t, Y_s)$ is continuous over $\Re^+ \times \Re^+$. Note that this is a statement about distributions, not sample paths.

Using these definitions, we can now define our intended topic. A stochastic process $\{X_T : t \succeq 0\}$ is an **Ornstein-Uhlenbeck Process or a Gauss-Markov process** if it is stationary, Gaussian, Markovian, and continuous in probability.

A.1.2 Itô's Lemma ($1D$)

Itô's lemma is the most important result about the manipulation of random variables that one requires. It is to functions of random variables what Taylor's theorem is to functions of deterministic variables. It relates the small change in a function of a random variable to the small change in the random variable itself. The lemma is, of course, more general than this and can be applied to functions of any random variable. If X satisfy the following SDE:

$$dX = A(X,t)dt + B(X,t)dW \ ,$$

where A usually called drift term, B noise intensity term or volatility function and W is a Wiener-Lévy process or Brownian motion[1]. Thus given $f(X,t)$, Itô's lemma says that:

$$df = B\frac{\partial f}{\partial X}dW + \left(\frac{\partial f}{\partial t} + A\frac{\partial f}{\partial X} + \frac{B^2}{2}\frac{\partial^2 f}{\partial X^2}\right)dt \ . \tag{A.1}$$

A.1.3 Fokker-Planck Equation

The Fokker-Planck equation (named after Adrian Fokker and Max Planck; also known as the Kolmogorov forward equation) describes the time evolution of the probability density function $p(X,t)$ of the position and velocity of a particle, but it can be

[1] A stochastic process $\{W_T : t \succeq 0\}$ is a **Wiener-Lévy process or Brownian motion** if it has stationary independent increments, $W_0 = 0$ and W_t is normally distributed ($E[W_t] = 0$).

generalized to any other observable, too. The general form of the $N-$dimensional Fokker-Planck equation is then:

$$\frac{\partial p}{\partial t} = -\sum_{i=1}^{N} \frac{\partial}{\partial x_i} \left(p A_i \right) + \sum_{i,j=1}^{N} \frac{\partial^2}{\partial x_i \partial x_j} \left(p B_{i,j} \right) \ , \tag{A.2}$$

where $A_i(X,t)$ is the drift vector and $B_{i,j}(X,t)$ the diffusion tensor, which results from the presence of the stochastic force. The Fokker-Planck equation can be used for computing the probability densities of stochastic differential equations. Consider the Itô stochastic differential equation:

$$dX_t = f(X_t,t)dt + g(X_t,t)dW_t \ , \qquad f(X_{t_0}) = X_0 \ ,$$

where $X_t \in \mathbb{R}^N$ is the sate variable, $W_t \in \mathbb{R}^{M<N}$ is a standard $M-$dimensional Wiener process. The probability density $p(X,t)$ of X_t is given by the Fokker-Planck equation with the drift and diffusion terms equal to:

$$A_i = f_i \left(X_t, t \right) \ , \qquad B_{i,j} = \frac{1}{2} \sum_{k=1}^{N} g_{i,k} \left(X_t, t \right) g_{k,j} \left(X_t, t \right) \ .$$

Under certain conditions, this evolves towards a steady-state distribution in which $\partial p / \partial t = 0$ and hence:

$$\sum_{i=1}^{N} \frac{\partial}{\partial x_i} \left(p A_i \right) = \sum_{i,j=1}^{N} \frac{\partial^2}{\partial x_i \partial x_j} \left(p B_{i,j} \right) \ . \tag{A.3}$$

A.2 Financial Definitions

A.2.1 Arbitrage Possibility

An **arbitrage possibility**[2] on a financial market is a self-financed portfolio h such that its value V has the following behaviour during a period of time:

$$\begin{aligned} V^h(0) &= C > 0 \ , \\ P\left(V^h(T) \geq C \right) &= 1 \ , \\ P\left(V^h(T) > C \right) &> 0 \ . \end{aligned} \tag{A.4}$$

We say that the market is arbitrage free if there are no arbitrage possibilities. An arbitrage possibility is thus equivalent to the possibility of making a positive amount of money out of nothing with probability 1 or a.s. (almost sure). It is thus a riskless money making machine or, if you will, a free lunch on the financial market.

[2] Definition by T. Björk [2].

A.2.2 In-Out-At the Money

Out of the money[3]: An option with no intrinsic value. A call option when the asset price is below the strike, a put option when the asset price is above the strike.

In the money: An option with positive intrinsic value. A call option when the asset price is above the strike, a put option when the asset price is below the strike.

At the money: A call or put with a strike that is close to the current asset level.

A.2.3 Risk-Neutral Valuation ($1D$)

A.2.3.1 Market Price of Risk

The term "traded security" is mainly described as a traded asset that is held solely for investment by a significant number of individuals. Stocks, bonds, gold and silver are all traded securities. However, interest rates, inflation rates, volatility and most of the commodities are not. Consider the following stochastic process:

$$\frac{d\theta}{\theta} = \alpha dt + \beta dW \ , \tag{A.5}$$

where α and β are the expected growth rate and the volatility of θ, respectively and they are only functions of time t and θ. dW is a Wiener processes. We do not assume that θ is the price of a trade security. It could be something as far removed from financial markets as the temperature in the center of Mexico.

Suppose that F_1 and F_2 are two derivative securities dependent only on θ and t. These could be options or other securities that are defined so that they provide a payoff equal to some function of θ. Using Itô, one can show that these contracts follow a SDE in the form:

$$dF_j = F_j \left(\mu_j dt + \sigma_j dW\right) \quad \text{for } j = 1, 2 \ ,$$

where μ_j, σ_j are only functions of θ and t, and dW is the same Wiener process as in equation (A.5). If one constructs a portfolio Π as:

$$\Pi = \phi_1 F_1 + \phi_2 F_2 \ ,$$

then one can make Π an instantaneously riskless portfolio if:

$$\phi_1 = \sigma_2 F_2 \ , \quad \phi_2 = -\sigma_1 F_1 \ .$$

[3]Definition by P. Wilmott [39].

Since Π is instantaneously riskless, it must earn the risk-free rate. Hence

$$d\Pi = r\Pi dt \ .$$

Doing some substitutions:

$$\frac{\mu_1 - r}{\sigma_1} = \frac{\mu_2 - r}{\sigma_2} = \lambda \ .$$

The parameter λ, as the value of each side, is in general depended only on both θ and t, however it is not dependent on the nature of the derivative security F_j. It is known as the "**market price of risk**[4]" and it is a measure of the level of the market's risk aversion. The higher it is, the more compensation (measured in terms of excess expected rate of return) the market requires for taking the risk (measured in terms of standard deviation of return). In general, for assets like stocks, the market price of risk λ is defined as follows:

$$\lambda = \frac{\mu - r}{\sigma} \ , \tag{A.6}$$

where μ is the expected return of a financial security F, r is the risk-less interest rate and σ is the volatility of F. The market price of risk is more difficult to define in cases of fixed income securities since it depends on assumptions made about the interest rates term structure. In the Cox–Ingersoll–Ross (CIR) model, [4] and [5] that is based on the following stochastic differential equation:

$$dr_t = \kappa_r \left(w_r - r\right) dt + \sigma\sqrt{r} dW_t \ .$$

W_t is a standard Brownian motion, κ_r is the speed of adjustment and w_r is the long-term average rate.

A.2.3.2 Risk-Neutral Valuation

Suppose that a derivative security F which depended on θ (A.5) provides a payoff at time T. This security F can always be valued as if the world were risk neutral. To make this true one needs to make the "**Risk adjustment**[5]" and it is necessary to set the expected growth rate of each underlying variable θ equal to $(\alpha - \lambda\beta)$ rather than α. The volatility β remains the same. λ is the market price of risk of the underlying variable θ. For convenience, we will refer to a world where expected growth rates are changed to a "**risk-neutral world**" if:

$$F\left(t\right) = e^{-(T-t)r} E_Q \left[F\left(\theta_k, T\right)\right] \ .$$

[4]Definition by J. Hull [21].
[5]Definition by J. Hull [21].

Where E_Q denotes the expected value in a risk-neutral world (Risk adjustment), then F is the correct value of the financial security. The probability measure in the risk-neutral world is thus called an "equivalent martingale measure". This result was first developed by Cox, Ingersoll and Ross [4] and represents an important extension to the Risk-neutral valuation argument.

Example [21]: The current price of θ, say copper, is 0.8 pounds. The risk-free interest rate r is 5% per annum. The expected growth rate of copper α is 20% per annum and its volatility β is 20% per annum. The market price of risk λ associated with copper is 0.5. A contract is traded which allows the holder to receive 1000 pounds of copper at no cost in 6 months time. The expected growth rate of the price of copper in a risk-neutral world is:

$$\alpha - \lambda\beta = -0.08 \ ,$$

or -8% per annum. The discounting expected payoff from the contract in a risk-neutral world is therefore:

$$E_Q \left[F \left(\theta_k, T \right) \right] = 800 e^{-0.08*0.5} = 768.63 \ .$$

So then, we estimate the current value of the contract $F\left(t\right)$ as:

$$F\left(t\right) = e^{-(T-t)r} E_Q \left[F \left(\theta_k, T \right) \right] = 749.65 \ .$$

A.2.4 Risk-Neutral Valuation (Stochastic Volatility Models)

Generally speaking, stochastic volatility models are not complete, hence typical contingent claims (such as European options) cannot be priced by arbitrage. Still, it is possible to derive, under additional hypothesis, the partial differential equation satisfied by the value of a contingent claim. To derive this PDE - which generalizes the Black-Scholes PDE - one needs first to specify the so called market price for risk, which reflects the expected excess return per unit risk over the risk-free rate. Intuitively, the market price for risk represents the return-to-risk trade-off demanded by investors for bearing the volatility risk of the stock. For some specifications of the dynamics of stochastic volatility and the market price of risk, a few closed form expressions are available, otherwise, numerical procedures need to be employed.

In this section we consider the option pricing implications of diffusion models for stochastic volatility. In particular it is no longer true that there are unique preference independent option prices. Instead the model is incomplete and economic considerations (such as risk aversion) must be introduced to obtain pricing formula. If volatility

were a traded asset then it would be possible to invest in volatility and the stock to form a riskless hedge portfolio for the option. However this is not the case so there is no riskless hedge and the price of options will depend on the risk preferences of investors. These preferences may be expressed via a utility function (see reference [25],[18] for example), or via a local-risk minimization criterion (see Hofmann, Platen and Schweizer [19]).

Consider the following volatility model or probability measure P under which dW_i are Brownian motions:

$$P : \left\{ \begin{array}{l} dS = S(\mu - D)dt + S\sqrt{\nu}d\widehat{W}_1 \\ d\nu = f(\nu)dt + g(\nu)d\widehat{W}_2 \end{array} \right\} . \tag{A.7}$$

To value an option or financial security F, do not use (A.7), but a closely related process which is often call the risk-adjusted process $\widehat{P}$ (replace the expected return μ by the interest rate r, and use the risk-adjusted volatility drift φ). This procedure is carried out explicitly for a class of equilibrium models. The risk-adjusted process $\widehat{P}$ will be in the risk-neutral world or equivalent martingale measure and will produce the theoretical fair price of the financial security F.

$$\widehat{P} : \left\{ \begin{array}{c} dS = S(r - D)dt + S\sqrt{\nu}d\widehat{W}_1 \\ d\nu = \left(f(\nu) - \varphi\left(\nu\right)\right)dt + g(\nu)d\widehat{W}_2 \end{array} \right\} . \tag{A.8}$$

Where $\varphi\left(\nu\right)$ can be defined as [18]:

$$\varphi\left(\nu\right) = g(\nu)\left(\lambda\rho - \overline{\Lambda}\sqrt{1-\rho^2}\right) .$$

λ is the **market price of asset risk** (A.6) associated with dW_1 and $\overline{\Lambda}$ is the **market price of volatility risk** associated with dW_2. The latter shows how much of the expected return of V is explained by the risk (standard deviation) of ν in the Capital Asset Pricing Model framework. The option pricing equation has an analogue in expressions given by Wiggins [36], Scott [31] and Stein and Stein [35].

A. Lewis [25] explains how to obtain the volatility drift adjustment $\varphi\left(\nu\right)$ as a function of the risk-aversion parameter γ also called constant proportional risk aversion (CPRA), which means:

$$\overline{U}\left(t\right) = e^{-Rt}\frac{\left(C_t\right)^{\gamma}}{\gamma} , \qquad \overline{B}\left(T\right) = e^{-RT}\frac{\left(W_T\right)^{\gamma}}{\gamma} ,$$

where γ is the CPRA parameter and $R \geq 0$ is an impatience parameter, both constants. Assuming that the representative agent is either risk neutral ($\gamma = 1$) or risk

averse ($\gamma < 1$). For any γ, $\overline{U}(t)$ is an increasing function of the rate of consumption utility C_t at time t. With ($\gamma < 1$), $\overline{U}$ has a negative second derivative $\left(\overline{U}_{CC} \leq 0\right)$ and means the agent is risk-averse. Clearly, $\overline{B}$ has the same properties where W_T is the society's wealth. The CPRA parameter is related to Pratt [28] in 1964 with the following identity:

$$\gamma = \frac{-\partial^2 \overline{U}}{\partial C^2} \left(\frac{\partial \overline{U}}{\partial C} \right)^{-1} = (1 - \gamma) \ .$$

To obtain the volatility drift adjustment $\varphi(\nu, t)$ one needs to solve the **risk premium coefficient** $A(\nu, t)$ which satisfies the non-linear PDE[6]:

$$\frac{\partial A}{\partial t} = -(1-\gamma) A^{-\gamma(1-\gamma)} + A \left[R - r\gamma - \frac{\gamma\nu(1-\gamma)}{2} \right] - f(\nu) \frac{\partial A}{\partial \nu} - \frac{g(\nu)^2}{2} \frac{\partial^2 A}{\partial \nu^2} \ , \quad \text{(A.9)}$$

where the boundary conditions $A(\nu, T) = 1$. Then, compute ψ:

$$\psi(\nu, t) = \frac{\partial A(\nu, t)}{\partial \nu} / A(\nu, t) \ , \quad \text{(A.10)}$$

to obtain:

$$\varphi(\nu, t) = (1-\gamma) \rho \sqrt{\nu} g(\nu) - g(\nu)^2 \psi(\nu, t) \ . \quad \text{(A.11)}$$

If $T \to \infty$ (infinite horizon problems like pension funds), then the risk coefficient becomes independent of time $[A(\nu, t) = A(\nu)]$ and the PDE (A.9) changes to:

$$(1-\gamma) A^{-\gamma(1-\gamma)} = A \left[R - r\gamma - \frac{\gamma\nu(1-\gamma)}{2} \right] - f(\nu) \frac{\partial A}{\partial \nu} - \frac{g(\nu)^2}{2} \frac{\partial^2 A}{\partial \nu^2} \ . \quad \text{(A.12)}$$

If $T < \infty$ (pure investment problem, any option prices), then the PDE (A.9) yields the form:

$$\frac{\partial A}{\partial t} = A \left[R - r\gamma - \frac{\gamma\nu(1-\gamma)}{2} \right] - f(\nu) \frac{\partial A}{\partial \nu} - \frac{g(\nu)^2}{2} \frac{\partial^2 A}{\partial \nu^2} \ . \quad \text{(A.13)}$$

If $\gamma = 0$ (Log-utility), regardless of the horizon (time to maturity), $A(\nu, t)$ is independent of ν, and so $\psi = 0$. Then the martingale pricing process for any option is (A.8) with:

$$\varphi(\nu, t) = \rho \sqrt{\nu} g(\nu) \ .$$

If $\gamma = 1$ (Risk-neutrality) is a degenerate one and technically (A.9) is ill-defined, especially if $R \neq r$. With $R = r$, $A(\nu, t)$ sticks at its boundary value $A(\nu, t) = 1$, and there is no adjustment to volatility drift:

$$\varphi(\nu, t) = 0 \ .$$

[6] Definition by A. Lewis [25].

Fortunately for our purpose, when it is only considering a pure investment problem (A.13), [25] gives some exact solutions for the risk-adjusted volatility drift $\varphi(\nu, t)$ in function of the risk-aversion parameter γ. For the square root model (2.9):

$$P : \left\{ \begin{array}{c} dS = S(\mu - D)dt + S\sqrt{\nu}d\widehat{W}_1 \\ d\nu = (\kappa(\varpi - \nu) - \Lambda\nu)\, dt + \xi\sqrt{\nu}d\widehat{W}_2 \end{array} \right\} ,$$

where:

$$\Lambda = -\kappa + (1-\gamma)\,\rho\xi + \sqrt{\kappa^2 - \gamma\,(1-\gamma)\,\xi^2} ,$$

with the restriction on the parameters:

$$\gamma \leq 1 \text{ and } \kappa^2 \leq \gamma\,(1-\gamma)\,\xi^2 .$$

For the the 3/2 Model (2.11):

$$P : \left\{ \begin{array}{c} dS = S(\mu - D)dt + S\sqrt{\nu}d\widehat{W}_1 \\ d\nu = \nu\,(\kappa(\varpi - \nu) - \Lambda\nu)\, dt + \xi v^{3/2}d\widehat{W}_2 \end{array} \right\} ,$$

where:

$$\Lambda = -\left(\kappa + \frac{\xi^2}{2}\right) + (1-\gamma)\,\rho\xi + \sqrt{\left(\kappa + \frac{\xi^2}{2}\right)^2 - \gamma\,(1-\gamma)\,\xi^2} ,$$

with the restriction on the parameters:

$$\gamma \leq 1 \text{ and } \gamma\,(1-\gamma)\,\xi^2 \leq \left(\kappa + \frac{\xi^2}{2}\right)^2 .$$

A.3 Formula derivation for Heston Volatility

Starting with the general stochastic volatility model (2.18)[7]:

$$\begin{aligned} dS &= S(\mu - D)dt + S\sqrt{\nu}d\widehat{W}_1 , \\ d\nu &= \nu^{\lambda_1}\left(\kappa(\varpi - \nu)\right)dt + \xi\nu^{\lambda_2}d\widehat{W}_2 , \end{aligned}$$

or the second equation can be represented in the form:

$$d\nu = f(\nu)dt + g(\nu)d\widehat{W}_2 .$$

In doing so, one arrives at the General PDE for stochastic volatility (2.16):

$$\frac{\partial V}{\partial t} + \frac{1}{2}S^2\nu\frac{\partial^2 V}{\partial S^2} + \rho S\sqrt{\nu}g(\nu)\frac{\partial^2 V}{\partial S\partial\nu} + \frac{1}{2}g\,(\nu)^2\,\frac{\partial^2 V}{\partial\nu^2} + S(r-D)\frac{\partial V}{\partial S} + f\,(\nu)\,\frac{\partial V}{\partial\nu} = rV . \tag{A.14}$$

[7] Information referenced from W. Shaw [32].

The last PDE (A.14) and the original Black-Scholes PDE (2.2) have a lot of similarities. Applying a similar set of transformations and using T equal to the maturity of the contract:

$$\begin{aligned}\tau &= T - t \ , \\ x &= \log(S) + (r - D)(T - t) \ , \\ V &= U(x, \nu, \tau) e^{-r(T-t)} \ .\end{aligned}$$

After some routine calculus using the chain rule, leads to a PDE for U in the form of:

$$\frac{1}{2}\nu \left(\frac{\partial^2 U}{\partial x^2} - \frac{\partial U}{\partial x} \right) + \rho\sqrt{\nu} g(\nu) \frac{\partial^2 U}{\partial x \partial \nu} + \frac{1}{2} g\left(\nu\right)^2 \frac{\partial^2 U}{\partial \nu^2} + f\left(\nu\right) \frac{\partial U}{\partial \nu} = \frac{\partial U}{\partial \tau} \ .$$

Now, we introduce the Fourier transform in the form:

$$\begin{aligned} U(x, \nu, \tau) &= \frac{1}{2\pi} \int_{-\infty}^{\infty} e^{-iwx} \widehat{U}(w, \nu, \tau) dw \ , \qquad \text{(A.15)} \\ \widehat{U}(w, \nu, \tau) &= \int_{-\infty}^{\infty} e^{iwx} U(x, \nu, \tau) dx \ . \end{aligned}$$

At maturity, where $\tau = T - t = 0$, you have:

$$\widehat{U}(w, \nu, 0) = \int_{-\infty}^{\infty} e^{iwx} U(x, \nu, 0) dx = \int_{-\infty}^{\infty} e^{iwx} V(x, \nu, 0) dx \ ,$$

which is the Fourier Transform of the payoff expressed in terms of the logarithm of the asset price. So the differentiation w.r.t.[8] x is equal to the multiplication by $(-iw)$ in the transform:

$$\frac{1}{2} g^2(\nu) \frac{\partial^2 \widehat{U}}{\partial \nu^2} + \left(f\left(\nu\right) - iw\rho\sqrt{\nu} g(\nu) \right) \frac{\partial \widehat{U}}{\partial \nu} - \frac{1}{2}\nu \left(w^2 - iw \right) \widehat{U} = \frac{\partial \widehat{U}}{\partial \tau} \ . \qquad \text{(A.16)}$$

- **The Vanilla Call**

The payoff of a call European option is $\max[S - K, 0]$ in terms of our original variables. In terms of logarithmic variables:

$$V\left(x, \nu, 0\right) = \max\left[e^x - K, 0\right] \ ,$$

[8] **w.r.t.** is a mathematical abbreviation of "with respect to".

so the Fourier transform of the payoff is:

$$\widehat{U_{V-Call}}(w,\nu,0) = \int_{-\infty}^{\infty} e^{iwx} V(x,\nu,0)dx = \int_{-\infty}^{\infty} e^{iwx} \max\left[e^x - K, 0\right] dx$$
$$= \int_{\log[K]}^{\infty} e^{iwx} \left(e^x - K\right) dx = \int_{\log[K]}^{\infty} \left(e^{(1+iw)x} - Ke^{iwx}\right) dx \ ,$$

so for $Im(w) > 1$:

$$\widehat{U_{V-Call}}(w,\nu,0) = \left[\frac{e^{(1+iw)x}}{(1+iw)} - K\frac{e^{iwx}}{iw}\right]_{x=\ln K}^{x=\infty} = \frac{K^{(1+iw)}}{iw - w^2} \ ,$$

one needs to check when this integral actually converges and bear in mind that w can be any complex number. The exponential needs to decay as x becomes large, so that the integral converges. This ONLY happens if $Im(w) > 1$ and, when this is true, one can evaluate the integral with some simplifications:

$$\widehat{U_{V-Call}}(w,\nu,0) = \frac{K^{(1+iw)}}{iw - w^2} \ ,$$
$$\text{subject to: } \mathrm{Im}(w) > 1.$$

- **The Vanilla Put**

Here conditions remain the same, except that this time the integral converges only if $Im(w) < 0$. When this is true, one obtains an identical transform:

$$\widehat{U_{V-Put}}(w,\nu,0) = \frac{K^{(1+iw)}}{iw - w^2} \ ,$$
$$\text{subject to: } \mathrm{Im}(w) < 0 \ .$$

The difference in this approach between the Call and the Put option is where the transform is defined, and hence where the inversion contour lies.

- **Digital Calls and Puts**

For a digital call, the transformed payoff is:

$$\widehat{U_{D-Call}}(w,\nu,0) = \frac{-K^{iw}}{iw} \ ,$$
$$\text{subject to: } \mathrm{Im}(w) > 0 \ .$$

For a digital put:

$$\widehat{U_{D-Put}}(w,\nu,0) = \frac{+K^{iw}}{iw} \ ,$$
$$\text{subject to: } \mathrm{Im}(w) < 0 \ .$$

The fundamental solution

Suppose we find the solution of the PDE, say $G(w,\nu,\tau)$, with the property that at $t = T$, $G(w,\nu,0) = 1$. Then the solution to the transformed PDE (A.16) with payoff condition $\widehat{U}(w,\nu,0)$ (which does not depend on v) is just the product of this with G and the solution to our original PDE is the discounted value of this with our various coordinate changes unwound:

$$V = \frac{1}{2\pi} e^{-r(T-t_0)} \int_{ic-\infty}^{ic+\infty} e^{-iwx} \widehat{U}(w,\nu,0) G(w,\nu,\tau) dw \ , \qquad \text{(A.17)}$$

$$\text{where} \quad x = \log(S) + (r-D)(T-t), \quad \tau = T - t_0 \ .$$

Lewis [25] discusses how to solve (A.17) for the general case, here we will solve only for the Heston model ($\gamma = \frac{1}{2}$).

- **Greeks for free**

Before figuring out G, we should point out that (A.17) is a remarkably useful representation. If one wants to differentiate V with respect to S to obtain Δ, one merely multiplies the integral by:

$$-\frac{iw}{S} \ ,$$

and for Γ, the integral is multiplied by:

$$-\frac{w^2}{S^2} \ .$$

This representation also makes obvious the link between ρ and Δ.

Finding the fundamental solution for the Heston Model

For the Heston Model, $\left(\lambda_1 = 0,\ \lambda_2 = \frac{1}{2}\right)$, the PDE (A.16) yield the form:

$$\frac{\partial G}{\partial \tau} = \frac{1}{2}\xi^2 \nu \frac{\partial^2 G}{\partial \nu^2} + \left((k(\varpi - \nu) - \Lambda\nu) - iw\rho\xi\nu\right) \frac{\partial G}{\partial \nu} - \frac{1}{2}\nu\left(w^2 - iw\right) G \ . \qquad \text{(A.18)}$$

What Heston did in [14] was to try to find a solution in the form:

$$G = e^{A[\tau,w] + \nu B[\tau,w]} \ ,$$

with:

$$A\left[0,w\right] = 0 = B\left[0,w\right] \ ,$$

in order to satisfy the condition that $G[0,w] = 1$ (at maturity). Substituting this assumption for the form of G into the PDE (A.18), one obtains the following condition:

$$A' + B'\nu = \frac{1}{2}\xi^2\nu B^2 + B\left((k(\varpi - \nu) - \Lambda\nu) - iw\rho\xi\nu\right) - \frac{1}{2}\nu\left(w^2 - iw\right) \ .$$

The A' and B' denote the $\tau-$ derivative. This must be true for all ν so we separately equate the terms that are independent of ν and linear in ν to obtain the pair of ordinary differential equations:

$$\begin{aligned} A' &= k\varpi B \ , \\ B' &= \frac{1}{2}\xi^2 B^2 - B\left(\kappa + \Lambda + iw\rho\xi\right) - \frac{1}{2}\left(w^2 - iw\right) \ . \end{aligned}$$

Solving this, one gets:

$$A = \frac{k\varpi}{\xi^2}\left(\left(\kappa + \Lambda + iw\rho\xi + c_1\right)\tau - 2\log\left(\frac{1 - c_2 e^{c_1\tau}}{1 - c_2}\right)\right) \ ,$$

$$B = \frac{\left(\kappa + \Lambda + iw\rho\xi + c_1\right)}{\xi^2}\left(\frac{1 - e^{c_1\tau}}{1 - c_2 e^{c_1\tau}}\right) \ ,$$

where:

$$c_1 = \sqrt{\left(w^2 - iw\right)\xi^2 + \left(\kappa + \Lambda + iw\rho\xi\right)^2} \ , \qquad c_2 = \frac{\kappa + \Lambda + iw\rho\xi + c_1}{\kappa + \Lambda + iw\rho\xi - c_1} \ .$$

It is however, better to do direct numerical integration of the ODE for A as you avoid the branch cut difficulties arising from the choice of the branch of the complex log (see [32]).

In conclusion, the exact solution of the option price using Heston volatility is:

$$V(S,T) = \frac{1}{2\pi}e^{-r(T-t_0)}\int_{ic-\infty}^{ic+\infty}\widehat{U}(w,\nu,0)\left(e^{A+\nu B}\right)e^{-iwx}dw \ ,$$

$$x = \log(S) + (r - D)(T - t) \ , \quad \tau = T - t_0 \ ,$$

using:

Type of option	Payoff	$\widehat{U}(w,\nu,0)$	Conditions
European Call	$\max(S-K,0)$	$\frac{K^{(1+iw)}}{iw-w^2}$	$\mathrm{Im}(w) > 1$
European Put	$\max(K-S,0)$	$\frac{K^{(1+iw)}}{iw-w^2}$	$\mathrm{Im}(w) < 0$
Digital Call	$H(S-K,0)$	$\frac{-K^{iw}}{iw}$	$\mathrm{Im}(w) > 0$
Digital Put	$H(K-S,0)$	$\frac{K^{iw}}{iw}$	$\mathrm{Im}(w) < 0$

.

If one wants to differentiate V with respect to S to obtain Δ and/or Γ, one merely multiplies the integral by:

$\Delta = \frac{\partial V}{\partial S}$	$-\frac{iw}{S}$
$\Gamma = \frac{\partial^2 V}{\partial S^2}$	$-\frac{w^2}{S^2}$

.

For further information or more details, see [32], or [25].

A.4 Equilibrium between all SVMs

The SVM (2.18) can be represented as:

$$\begin{aligned} \frac{dx}{x} &= \mu dt + \max\left(\overline{\sigma}, 0\right) d\widehat{W}_{1,t} \ , \\ dy &= k_j y^{\lambda_3}\left(\varpi_j^{\lambda_0} - y\right) dt + \beta_j y^{\lambda_2} d\widehat{W}_{2,t} \ , \\ \overline{\sigma} &= y^{\lambda_1} \ , \qquad j = case. \end{aligned}$$

Using Itô's lemma:

$$d\overline{\sigma} = f_j\left(\overline{\sigma}\right) dt + g_j\left(\overline{\sigma}\right) d\widehat{W}_{2,t} \ ,$$

where:

$$\begin{aligned} f_j\left(\overline{\sigma}\right) &= \lambda_1 k_j\left(\overline{\sigma}^{\frac{\lambda_1+\lambda_3-1}{\lambda_1}}\varpi_j^{\lambda_0} - \overline{\sigma}^{\frac{\lambda_1+\lambda_3}{\lambda_1}}\right) + \frac{\lambda_1\left(\lambda_1-1\right)\beta_j^2}{2}\overline{\sigma}^{\frac{\lambda_1+2\lambda_2-2}{\lambda_1}} \ , \\ g_j\left(\overline{\sigma}\right) &= \lambda_1\beta_j\overline{\sigma}^{\frac{\lambda_1+\lambda_2-1}{\lambda_1}} \ . \end{aligned}$$

To make a comparison between the steady-state distribution for different cases, one needs to set the following equilibrium. For any choice of $\overline{\overline{\sigma}}$, using the asymptotic approximation (2.21) and taking the square root Model (Heston model, $j = H$) as the master model, we have:

$$\begin{aligned} A) \qquad f_j\left(\overline{\sigma}\right) &= f_H\left(\overline{\sigma}\right) \ . \\ B) \qquad g_j\left(\overline{\sigma}\right) &= g_H\left(\overline{\sigma}\right) \ . \\ C) \qquad \frac{\partial f_j\left(\overline{\sigma}\right)}{\partial\overline{\sigma}} &= \frac{\partial f_H\left(\overline{\sigma}\right)}{\partial\overline{\sigma}} \ . \end{aligned} \tag{A.19}$$

where:

$$\begin{aligned} \frac{\partial f_j\left(\overline{\sigma}\right)}{\partial\overline{\sigma}} &= k_j\left(\left(\lambda_1+\lambda_3-1\right)\overline{\sigma}^{\frac{\lambda_3-1}{\lambda_1}}\varpi_j^{\lambda_0} - \left(\lambda_1+\lambda_3\right)\overline{\sigma}^{\frac{\lambda_3}{\lambda_1}}\right) \\ &\quad + \frac{\left(\lambda_1-1\right)\left(\lambda_1+2\lambda_2-2\right)\beta_j^2}{2}\overline{\sigma}^{\frac{2\lambda_2-2}{\lambda_1}} \ . \end{aligned}$$

Taking the square root Model (Heston model) as the master model:

λ_0	λ_1	λ_2	λ_3
2	0.5	1/2	0

,

$$f_H\left(\overline{\sigma}\right) = \frac{1}{\overline{\sigma}}\left(\frac{k_H}{2}\left(\varpi_H^2 - \overline{\sigma}^2\right) - \frac{\beta_H^2}{8}\right) \ ,$$

$$g_H\left(\overline{\sigma}\right) = \frac{\beta_H}{2} \ ,$$

$$\frac{\partial f_H(\overline{\sigma})}{\partial \overline{\sigma}} = \frac{1}{\overline{\sigma}^2}\left(\frac{-k_H}{2}\left(\varpi_H^2 + \overline{\sigma}^2\right) + \frac{\beta_H^2}{8}\right) .$$

Solving (A.19), one gets:

$$\beta_j = \frac{\beta_H}{2\lambda_1}\overline{\sigma}^{\frac{1-\lambda_1-\lambda_2}{\lambda_1}} \quad \text{then} \quad \beta_j^2 = \frac{\beta_H^2}{4\lambda_1^2}\overline{\sigma}^{\frac{2-2\lambda_1-2\lambda_2}{\lambda_1}} , \tag{A.20}$$

$$k_j = \overline{\sigma}^{\frac{-2\lambda_1-\lambda_3}{\lambda_1}}\left(\frac{k_H}{2\lambda_1}\left(\varpi_H^2\left(2\lambda_1+\lambda_3-1\right)+\overline{\sigma}^2\left(1-\lambda_3\right)\right) - \frac{C_{\lambda_j^1}\beta_H^2}{4}\right) , \tag{A.21}$$

$$\varpi_j^{\lambda_0} = \overline{\sigma}^{\frac{1}{\lambda_1}}\left(1 + \frac{\left(k_H\left(\varpi_H^2-\overline{\sigma}^2\right) - \frac{\beta_H^2}{4}\left(\frac{(2\lambda_1-1)}{\lambda_1}\right)\right)}{k_H\left(\varpi_H^2\left(2\lambda_1+\lambda_3-1\right)+\overline{\sigma}^2\left(1-\lambda_3\right)\right) - \frac{\beta_H^2}{4}\left(2\lambda_1 C_{\lambda_j^1}\right)}\right) , \tag{A.22}$$

$$C_{\lambda_j^1} = 1 - \frac{2\lambda_1\left(\lambda_2-\lambda_3\right)+\left(\lambda_3-2\lambda_2+1\right)}{2\lambda_1^2} .$$

Using the asymptotic approximation (2.21), $f_H\left(\overline{\sigma}_H^*\right)=0$:

$$\overline{\sigma}_H^* = \sqrt{\left(\varpi_H^2 - \frac{\beta_H^2}{4k_H}\right)} . \tag{A.23}$$

Solving equations (A.20-A.22) using (A.23), one finally obtains:

$$\beta_j = \frac{\beta_H}{2\lambda_1}\left(\frac{k_H}{k_H\varpi_H^2 - \frac{1}{4}\beta_H^2}\right)^{\frac{\lambda_1+\lambda_2-1}{2\lambda_1}} , \tag{A.24}$$

$$k_j = \left(\frac{k_H}{k_H\varpi_H^2 - \frac{1}{4}\beta_H^2}\right)^{\frac{2\lambda_1+\lambda_3}{2\lambda_1}}\left(k_H\varpi_H^2 - \frac{C_{\lambda_j}\beta_H^2}{8}\right) , \tag{A.25}$$

$$\varpi_j = \left(\left(\varpi_H^2 - \frac{\beta_H^2}{4k_H}\right)^{\frac{1}{2\lambda_1}}\left(1 - \frac{\left(\lambda_1-1\right)\beta_H^2}{\lambda_1^2\left(8k_H\varpi_H^2 - C_{\lambda_j}\beta_H^2\right)}\right)\right)^{\frac{1}{\lambda_0}} , \tag{A.26}$$

$$C_{\lambda_j} = 2 + \frac{\left(\lambda_1-1\right)\left(\lambda_3-2\lambda_2+1\right)}{\lambda_1^2} .$$

For $\overline{\sigma} = y$ ($\lambda_1 = 1$):

Case	λ_0	λ_1	λ_2	λ_3	–	$dy =$
$j = 1a$	1	1	1	0	–	$dy = k_{1a}\left(\varpi_{1a} - y\right)dt + \beta_{1a} y d\widehat{W}_{2,t}$
$j = 2a$	1	1	2	0	–	$dy = k_{2a}\left(\varpi_{2a} - y\right)dt + \beta_{2a} y^2 d\widehat{W}_{2,t}$
$j = 3a$	1	1	2	1	–	$dy = k_{3a}\left(y\varpi_{3a} - y^2\right)dt + \beta_{3a} y^2 d\widehat{W}_{2,t}$

.

Solving equations (A.24-A.26):

$k_{1a} = k_H$	$\varpi_{1a} = \sqrt{\varpi_H^2 - \frac{\beta_H^2}{4k_H}}$	$\beta_{1a} = \sqrt{\frac{k_H \beta_H^2}{4k_H \varpi_H^2 - \beta_H^2}}$
$k_{2a} = k_H$	$\varpi_{2a} = \sqrt{\varpi_H^2 - \frac{\beta_H^2}{4k_H}}$	$\beta_{2a} = \frac{k_H \beta_H}{2\left(k_H \varpi_H^2 - \frac{1}{4}\beta_H^2\right)}$
$k_{3a} = \frac{2k_H^{3/2}}{\sqrt{4k_H \varpi_H^2 - \beta_H^2}}$	$\varpi_{3a} = \sqrt{\varpi_H^2 - \frac{\beta_H^2}{4k_H}}$	$\beta_{3a} = \frac{k_H \beta_H}{2\left(k_H \varpi_H^2 - \frac{1}{4}\beta_H^2\right)}$

.

For $\overline{\sigma} = y^{0.5}$ ($\lambda_1 = 0.5$):

Case	λ_0	λ_1	λ_2	λ_3	–	$dy =$
$j = 1b$	2	0.5	1/2	0	–	$dy = k_{1b}\left(\varpi_{1b}^2 - y\right)dt + \beta_{1b} y^{0.5} d\widehat{W}_{2,t}$
$j = 2b$	2	0.5	1	0	–	$dy = k_{2b}\left(\varpi_{2b}^2 - y\right)dt + \beta_{2b} y d\widehat{W}_{2,t}$
$j = 3b$	2	0.5	3/2	0	–	$dy = k_{3b}\left(\varpi_{3b}^2 - y\right)dt + \beta_{3b} y^{3/2} d\widehat{W}_{2,t}$
$j = 4b$	2	0.5	3/2	1	–	$dy = k_{4b}\left(y\varpi_{4b}^2 - y^2\right)dt + \beta_{4b} y^{3/2} d\widehat{W}_{2,t}$

.

Solving equations (A.24-A.26):

$k_{1b} = k_H$	$\varpi_{1b} = \varpi_H$	$\beta_{1b} = \beta_H$
$k_{2b} = k_H - \frac{k_H \beta_H^2}{4k_H \varpi_H^2 - \beta_H^2}$	$\varpi_{2b} = \sqrt{\varpi_H^2 + \frac{\beta_H^4}{4k_H\left(4k_H \varpi_H^2 - 2\beta_H^2\right)}}$	$\beta_{2b} = \sqrt{\frac{k_H \beta_H^2}{k_H \varpi_H^2 - \frac{1}{4}\beta_H^2}}$
$k_{3b} = k_H - 2\frac{k_H \beta_H^2}{4k_H \varpi_H^2 - \beta_H^2}$	$\varpi_{3b} = \sqrt{\varpi_H^2 + \frac{\beta_H^4}{2k_H\left(4k_H \varpi_H^2 - 3\beta_H^2\right)}}$	$\beta_{3b} = \frac{k_H \beta_H}{k_H \varpi_H^2 - \frac{1}{4}\beta_H^2}$
$k_{4b} = \frac{8k_H^2\left(2k_H \varpi_H^2 - \beta_H^2\right)}{\left(4k_H \varpi_H^2 - \beta_H^2\right)^2}$	$\varpi_{4b} = \sqrt{\varpi_H^2 + \frac{\beta_H^4 - 16k_H \varpi_H^2}{8k_H\left(2k_H \varpi_H^2 - \beta_H^2\right)}}$	$\beta_{4b} = \frac{k_H \beta_H}{k_H \varpi_H^2 - \frac{1}{4}\beta_H^2}$

.

Appendix B

Discrete Time Approximations

B.1 Brownian Bridge

Conditioning a Brownian motion on its endpoints produces a Brownian bridge. For example, if one has a Wiener process ΔW_t, one can construct a Brownian bridge $\Delta W_t^{(B)}$ that gives intermediate points between the time interval $[t - \Delta t, \Delta t]$

$$\Delta W_t^{(B)} = \left[\Delta W_{t,1}^{(B)}, \Delta W_{t,2}^{(B)} .... \Delta W_{t,N_P}^{(B)}\right] .$$

If $Z_{t,j}$ are independent normal distributed random numbers, and using :

$$\begin{aligned} \Delta W_{t,1}^{(B)} &= \frac{1}{2}\left(\Delta W_t + \sqrt{\Delta t} Z_{t,1}\right) , \\ \Delta W_{t,2}^{(B)} &= \frac{1}{2}\left(\Delta W_t - \sqrt{\Delta t} Z_{t,1}\right) , \end{aligned} \tag{B.1}$$

one can obtain the first two intermediate points and then progressively obtain $4, 8, 16, ..$ points. The number of points or divisions N_K in the Wiener path depends directly on the number of times N_{Sub} is repeated the subroutine (B.1) and it can be calculated by:

$$N_K = 2^{N_{Sub}-1} .$$

The main point here is to maintain the original properties:

$$\sum_{i=1}^{N_K} \Delta W_{t,i} = \Delta W_t ,$$

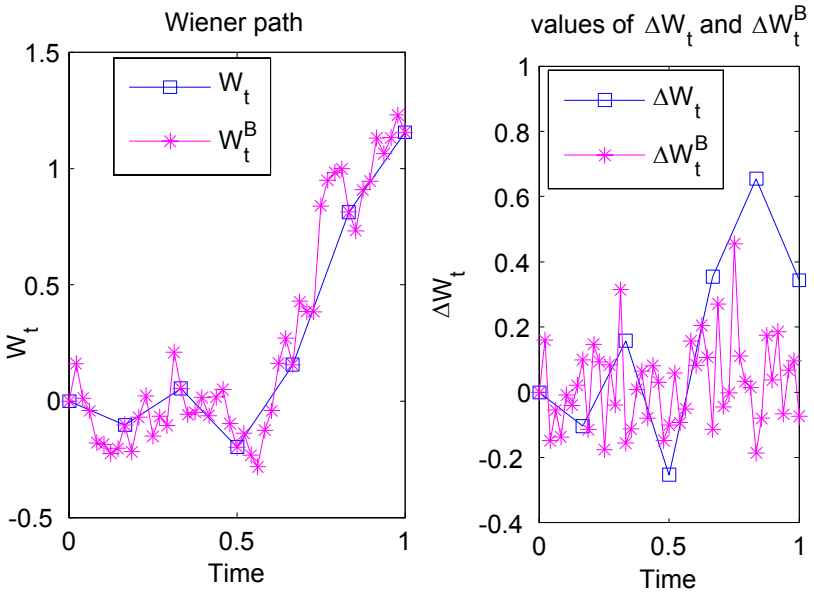

Figure B.1: Brownian bridge with $dt = 1/6$ and $N_P = 8$

$$
\begin{aligned}
E\left[\Delta W_t\right] &= E\left[\Delta W_{t,i}^{(B)}\right] = 0 \ , \\
Var\left[\Delta W_t\right] &= \Delta t \ , \qquad Var\left[\Delta W_{t,i}^{(B)}\right] = \frac{\Delta t}{N_P} \ .
\end{aligned}
$$

B.2 Itô's Lemma ($2D$):

Itô's lemma is the most important result about the manipulation of random variables that one requires. It relates the small change in a function of a random variable to the small change in the random variable itself.

A) If X and Y satisfy the following SDEs:

$$
\begin{aligned}
dX &= A(X,Y,t)dt + B(X,Y,t)d\widehat{W}_1 \ , \\
dY &= C(X,Y,t)dt + D(X,Y,t)d\widehat{W}_2 \ , \\
E\left[dW_1 dW_2\right] &= \rho dt \ ,
\end{aligned}
$$

where A and C are functions of X,Y and t, usually called drift terms, B and C are the noise intensity terms or volatility functions and $\widehat{W}_1, \widehat{W}_2$ are two Wiener-Lévy processes or Brownian motions with correlation ρ. Thus given $f(X,Y,t)$, Itô's lemma says that:

$$
df = B\frac{\partial f}{\partial X}d\widehat{W}_1 + D\frac{\partial f}{\partial Y}d\widehat{W}_2 \tag{B.2}
$$

$$+\left(A\frac{\partial f}{\partial X}+C\frac{\partial f}{\partial Y}+\frac{\partial f}{\partial t}+\frac{1}{2}B^2\frac{\partial^2 f}{\partial X^2}+\rho BD\frac{\partial^2 f}{\partial X\partial Y}+\frac{1}{2}D^2\frac{\partial^2 f}{\partial Y^2}\right)dt\ .$$

B) If X and Y satisfy the following SDEs:

$$dX=A(X,Y,t)dt+B_1(X,Y,t)dW_1+B_2(X,Y,t)dW_2\ ,$$

$$dY=C(X,Y,t)dt+D_1(X,Y,t)dW_1+D_2(X,Y,t)dW_2\ ,$$

$$E\left[dW_1dW_2\right]=0\ ,$$

and thus given $f(X,Y,t)$, Itô's lemma says that:

$$df=B_1\frac{\partial f}{\partial X}dW_1+B_2\frac{\partial f}{\partial X}dW_2+D_1\frac{\partial f}{\partial Y}dW_1+D_2\frac{\partial f}{\partial Y}dW_2 \tag{B.3}$$

$$+\left(A\frac{\partial f}{\partial X}+C\frac{\partial f}{\partial Y}+\frac{\partial f}{\partial t}\right)dt$$

$$+\left(\frac{1}{2}\left(B_1^2+B_2^2\right)\frac{\partial^2 f}{\partial X^2}+\left(B_1D_1+B_2D_2\right)\frac{\partial^2 f}{\partial X\partial Y}+\frac{1}{2}\left(D_1^2+D_2^2\right)\frac{\partial^2 f}{\partial Y^2}\right)dt\ .$$

Example:

$$d\begin{bmatrix}\widetilde{x}\\ \widetilde{y}\end{bmatrix}=\begin{bmatrix}\mu^{(\widetilde{x})}\\ \mu^{(\widetilde{y})}\end{bmatrix}dt+\begin{bmatrix}\sigma\\ 0\end{bmatrix}d\widetilde{\widehat{W}}_{1,t}+\begin{bmatrix}0\\ \xi\end{bmatrix}d\widetilde{\widehat{W}}_{2,t}\ , \tag{B.4}$$

where:

$$\begin{bmatrix}d\widetilde{\widehat{W}}_{1,t}\\ d\widetilde{\widehat{W}}_{2,t}\end{bmatrix}=\begin{bmatrix}1 & 0\\ \rho & \widehat{\rho}\end{bmatrix}\begin{bmatrix}\cos\theta & -\sin\theta\\ \sin\theta & \cos\theta\end{bmatrix}\begin{bmatrix}dW_{1,t}\\ dW_{2,t}\end{bmatrix}\ ,$$

and the correlation for $d\widetilde{\widehat{W}}_{1,t}$ is:

$$E\left[d\widetilde{\widehat{W}}_{1,t}d\widetilde{\widehat{W}}_{2,t}\right]=\rho dt\ .$$

Using independent Wiener processes, (B.4) can be represented as:

$$d\begin{bmatrix}\widetilde{x}\\ \widetilde{y}\end{bmatrix}=\begin{bmatrix}\mu^{(\widetilde{x})}\\ \mu^{(\widetilde{y})}\end{bmatrix}dt+\begin{bmatrix}\sigma\cos\theta\\ \xi\left(\rho\cos\theta+\widehat{\rho}\sin\theta\right)\end{bmatrix}dW_{1,t} \tag{B.5}$$

$$+\begin{bmatrix}-\sigma\sin\theta\\ \xi\left(-\rho\sin\theta+\widehat{\rho}\cos\theta\right)\end{bmatrix}dW_{2,t}\ .$$

If one applies Itô's lemma (B.2) to (B.4) or Itô's lemma (B.3) to (B.5), then one obtains the same expression for $d\theta$:

$$d\theta=\sigma\frac{\partial\theta}{\partial\widetilde{x}}d\widetilde{\widehat{W}}_{1,t}+\xi\frac{\partial\theta}{\partial\widetilde{y}}d\widetilde{\widehat{W}}_{2,t} \tag{B.6}$$

$$+\left(\mu^{(\widetilde{x})}\frac{\partial\theta}{\partial\widetilde{x}}+\mu^{(\widetilde{y})}\frac{\partial\theta}{\partial\widetilde{y}}+\frac{1}{2}\sigma^2\frac{\partial^2\theta}{\partial\widetilde{x}^2}+\rho\sigma\xi\frac{\partial^2\theta}{\partial\widetilde{x}\partial\widetilde{y}}+\frac{1}{2}\xi^2\frac{\partial^2\theta}{\partial\widetilde{y}^2}\right)dt\ .$$

B.3 Orthogonal Milstein Scheme (Operations)

B.3.1 Milstein Scheme (Itô Operators)

We start the following $2D$ Itô SDE with a $2D$ independent Wiener process:

$$d\begin{bmatrix} X_{1,t} \\ X_{2,t} \end{bmatrix} = \begin{bmatrix} a_1 \\ a_2 \end{bmatrix} dt + \begin{bmatrix} b_{1,1} & b_{1,2} \\ b_{2,1} & b_{2,2} \end{bmatrix} \begin{bmatrix} dW_{1,t} \\ dW_{2,t} \end{bmatrix} , \tag{B.7}$$

where we suppose a_i,$b_{i,k}$ are sufficiently smooth functions of $X_{i,t}$ in $[t_0...T]$. The 1 strong order Milstein scheme for (B.7) with time step Δt using Itô operators is:

$$\begin{bmatrix} \widehat{X}_{1,t+\Delta t} \\ \widehat{X}_{2,t+\Delta t} \end{bmatrix} = \begin{bmatrix} \widehat{X}_{1,t} \\ \widehat{X}_{2,t} \end{bmatrix} + \begin{bmatrix} a_1 \\ a_2 \end{bmatrix} \Delta t + \begin{bmatrix} b_{1,1} & b_{1,2} \\ b_{2,1} & b_{2,2} \end{bmatrix} \begin{bmatrix} \Delta W_{1,t} \\ \Delta W_{2,t} \end{bmatrix} + \frac{1}{2} R_M ,$$

where using Lévy Areas, R_M is equal to:

$$R_M = \begin{bmatrix} L_1 b_{1,1} \\ L_1 b_{2,1} \end{bmatrix} \left(\Delta W_{1,t}^2 - \Delta t\right) + \begin{bmatrix} L_2 b_{1,2} \\ L_2 b_{2,2} \end{bmatrix} \left(\Delta W_{2,t}^2 - \Delta t\right) \tag{B.8}$$

$$+ \begin{bmatrix} L_1 b_{1,2} + L_2 b_{1,1} \\ L_1 b_{2,2} + L_2 b_{2,1} \end{bmatrix} \Delta W_{1,t} \Delta W_{2,t} + \begin{bmatrix} (L_1 b_{1,2} - L_2 b_{1,1}) \\ (L_1 b_{2,2} - L_2 b_{2,1}) \end{bmatrix} \left[\underline{\overline{L}}_{(1,2)}\right]_t^{t+\Delta t} .$$

The Itô operators are defined by:

$$L_j := \sum_{k=1}^{d} b_{k,j} \frac{\partial}{\partial X_k} .$$

(B.8) can be simplified by:

$$\begin{aligned} R_M &= \sum_{j=1}^{2} \begin{bmatrix} L_j b_{1,j} \\ L_j b_{2,j} \end{bmatrix} \left(\Delta W_{j,t}^2 - \Delta t\right) + \begin{bmatrix} H_1^+ \\ H_2^+ \end{bmatrix} \Delta W_{1,t} \Delta W_{2,t} \\ &\quad + \begin{bmatrix} H_1^- \\ H_2^- \end{bmatrix} \left[\underline{\overline{L}}_{(1,2)}\right]_t^{t+\Delta t} , \end{aligned}$$

where:

$$H_j^{\pm} = L_1 b_{j,2} \pm L_2 b_{j,1} .$$

Doing some computations, (B.8) is equal to:

$$\begin{aligned} R_M &= \begin{bmatrix} C_{1,X} \\ C_{1,Y} \end{bmatrix} \left(\Delta W_{1,t}^2 - \Delta t\right) + \begin{bmatrix} C_{2,X} \\ C_{2,Y} \end{bmatrix} \left(\Delta W_{2,t}^2 - \Delta t\right) \\ &\quad + \begin{bmatrix} C_{3,X} + C_{4,X} \\ C_{3,Y} + C_{4,Y} \end{bmatrix} \Delta W_{1,t} \Delta W_{2,t} + \begin{bmatrix} C_{3,X} - C_{4,X} \\ C_{3,Y} - C_{4,Y} \end{bmatrix} \left[\underline{\overline{L}}_{(1,2)}\right]_t^{t+\Delta t} , \end{aligned}$$

where:

$C_{1,X} = b_{1,1}b_{1,1_{X_1}} + b_{2,1}b_{1,1_{X_2}}$	$C_{1,Y} = b_{1,1}b_{2,1_{X_1}} + b_{2,1}b_{2,1_{X_2}}$
$C_{2,X} = b_{1,2}b_{1,2_{X_1}} + b_{2,2}b_{1,2_{X_2}}$	$C_{2,Y} = b_{1,2}b_{2,2_{X_1}} + b_{2,2}b_{2,2_{X_2}}$
$C_{3,X} = b_{1,1}b_{1,2_{X_1}} + b_{2,1}b_{1,2_{X_2}}$	$C_{3,Y} = b_{1,1}b_{2,2_{X_1}} + b_{2,1}b_{2,2_{X_2}}$
$C_{4,X} = b_{1,2}b_{1,1_{X_1}} + b_{2,2}b_{1,1_{X_2}}$	$C_{4,Y} = b_{1,2}b_{2,1_{X_1}} + b_{2,2}b_{2,1_{X_2}}$

,

and the Lie bracket is equal to:

$$\begin{bmatrix} H_1^- \\ H_2^- \end{bmatrix} = \begin{bmatrix} b_{1,1}b_{1,2_{X_1}} + b_{2,1}b_{1,2_{X_2}} - b_{1,2}b_{1,1_{X_1}} - b_{2,2}b_{1,1_{X_2}} \\ b_{1,1}b_{2,2_{X_1}} + b_{2,1}b_{2,2_{X_2}} - b_{1,2}b_{2,1_{X_1}} - b_{2,2}b_{2,1_{X_2}} \end{bmatrix} .$$

Having example (4.3), you get:

$$\overline{\sigma}(x,y) = \begin{bmatrix} \sigma & 0 \\ \rho\,\xi & \widehat{\rho}\,\xi \end{bmatrix} ,$$

then:

$C_{1,X} = \sigma\sigma_x + \rho\xi\sigma_y$	$C_{1,Y} = \rho\sigma\xi_x + \rho^2\xi\xi_y$
$C_{2,X} = 0$	$C_{2,Y} = \widehat{\rho}^2\xi\xi_y$
$C_{3,X} = 0$	$C_{3,Y} = \widehat{\rho}\sigma\xi_x + \rho\widehat{\rho}\xi\xi_y$
$C_{4,X} = \widehat{\rho}\xi\sigma_y$	$C_{4,Y} = \rho\widehat{\rho}\xi\xi_y$

,

and the Lie bracket is equal to:

$$\begin{bmatrix} H_1^- \\ H_2^- \end{bmatrix} = \begin{bmatrix} -\widehat{\rho}\xi\sigma_y \\ \widehat{\rho}\sigma\xi_x \end{bmatrix} .$$

B.3.2 Orthogonal Milstein Scheme

If one makes an orthogonal transformation to (B.7), one gets:

$$d\begin{bmatrix} \widetilde{X}_t \\ \widetilde{Y}_t \end{bmatrix} = \begin{bmatrix} a_1 \\ a_2 \end{bmatrix} dt + \begin{bmatrix} b_{1,1} & b_{1,2} \\ b_{2,1} & b_{2,2} \end{bmatrix} \begin{bmatrix} d\widetilde{W}_{1,t} \\ d\widetilde{W}_{2,t} \end{bmatrix} , \tag{B.9}$$

where:

$$\begin{bmatrix} d\widetilde{W}_{1,t} \\ d\widetilde{W}_{2,t} \end{bmatrix} = \begin{bmatrix} \cos\theta & \mp\sin\theta \\ \sin\theta & \pm\cos\theta \end{bmatrix} \begin{bmatrix} dW_{1,t} \\ dW_{2,t} \end{bmatrix} . \tag{B.10}$$

One can represent the system (B.9) with independent noise as:

$$d\begin{bmatrix} \widetilde{X}_t \\ \widetilde{Y}_t \end{bmatrix} = \begin{bmatrix} a_1 \\ a_2 \end{bmatrix} dt + \begin{bmatrix} \widetilde{b}_{1,1} & \widetilde{b}_{1,2} \\ \widetilde{b}_{2,1} & \widetilde{b}_{2,2} \end{bmatrix} \begin{bmatrix} dW_{1,t} \\ dW_{2,t} \end{bmatrix} , \tag{B.11}$$

where:

$$\begin{bmatrix} \widetilde{b}_{1,1} & \widetilde{b}_{1,2} \\ \widetilde{b}_{2,1} & \widetilde{b}_{2,2} \end{bmatrix} = \begin{bmatrix} b_{1,1} & b_{1,2} \\ b_{2,1} & b_{2,2} \end{bmatrix} \begin{bmatrix} \cos\theta & \mp\sin\theta \\ \sin\theta & \pm\cos\theta \end{bmatrix} .$$

The 1 strong order Milstein scheme for (B.11) with time step Δt using Itô operators is:

$$\begin{bmatrix} \widehat{\widetilde{X}}_{t+\Delta t} \\ \widehat{\widetilde{Y}}_{t+\Delta t} \end{bmatrix} = \begin{bmatrix} \widehat{\widetilde{X}}_t \\ \widehat{\widetilde{Y}}_t \end{bmatrix} + \begin{bmatrix} a_1 \\ a_2 \end{bmatrix} \Delta t + \begin{bmatrix} \widetilde{b}_{1,1} & \widetilde{b}_{1,2} \\ \widetilde{b}_{2,1} & \widetilde{b}_{2,2} \end{bmatrix} \begin{bmatrix} \Delta W_{1,t} \\ \Delta W_{2,t} \end{bmatrix} + \frac{1}{2} R_M \ ,$$

where using Lévy Areas, R_M is equal to:

$$R_M = \begin{bmatrix} \widetilde{L}_1\widetilde{b}_{1,1} \\ \widetilde{L}_1\widetilde{b}_{2,1} \end{bmatrix} \left(\Delta W_{1,t}^2 - \Delta t\right) + \begin{bmatrix} \widetilde{L}_2\widetilde{b}_{1,2} \\ \widetilde{L}_2\widetilde{b}_{2,2} \end{bmatrix} \left(\Delta W_{2,t}^2 - \Delta t\right) \tag{B.12}$$

$$+ \begin{bmatrix} \widetilde{L}_1\widetilde{b}_{1,2} + \widetilde{L}_2\widetilde{b}_{1,1} \\ \widetilde{L}_1\widetilde{b}_{2,2} + \widetilde{L}_2\widetilde{b}_{2,1} \end{bmatrix} \Delta W_{1,t}\Delta W_{2,t} + \begin{bmatrix} \left(\widetilde{L}_1\widetilde{b}_{1,2} - \widetilde{L}_2\widetilde{b}_{1,1}\right) \\ \left(\widetilde{L}_1\widetilde{b}_{2,2} - \widetilde{L}_2\widetilde{b}_{2,1}\right) \end{bmatrix} \left[\underline{L}_{(1,2)}\right]_t^{t+\Delta t} \ .$$

The Itô operators are defined by:

$$\widetilde{L}_j := \sum_{k=1}^{d} \widetilde{b}_{k,j} \frac{\partial}{\partial \widetilde{X}_k} \ .$$

(B.12) can be simplified by:

$$\begin{aligned} R_M \ = \ & \sum_{j=1}^{2} \begin{bmatrix} \widetilde{L}_j\widetilde{b}_{1,j} \\ \widetilde{L}_j\widetilde{b}_{2,j} \end{bmatrix} \left(\Delta W_{j,t}^2 - \Delta t\right) + \begin{bmatrix} \widetilde{H}_1^+ \\ \widetilde{H}_2^+ \end{bmatrix} \Delta W_{1,t}\Delta W_{2,t} \\ & + \begin{bmatrix} \widetilde{H}_1^- \\ \widetilde{H}_2^- \end{bmatrix} \left[\underline{L}_{(1,2)}\right]_t^{t+\Delta t} \ , \end{aligned} \tag{B.13}$$

where:

$$\widetilde{H}_j^{\pm} = \widetilde{L}_1\widetilde{b}_{j,2} \pm \widetilde{L}_2\widetilde{b}_{j,1} \ .$$

Doing some computation, (B.13) is equal to:

$$R_M = \begin{bmatrix} \widetilde{C}_{1,X}\cos^2\theta + \widetilde{C}_{2,X}\sin^2\theta + \left(\widetilde{C}_{3,X} + \widetilde{C}_{4,X}\right)\sin\theta\cos\theta \\ \widetilde{C}_{1,Y}\cos^2\theta + \widetilde{C}_{2,Y}\sin^2\theta + \left(\widetilde{C}_{3,Y} + \widetilde{C}_{4,Y}\right)\sin\theta\cos\theta \end{bmatrix} \left(\Delta W_{1,t}^2 - \Delta t\right)$$

$$+ \begin{bmatrix} \widetilde{C}_{2,X}\cos^2\theta + \widetilde{C}_{1,X}\sin^2\theta - \left(\widetilde{C}_{3,X} + \widetilde{C}_{4,X}\right)\sin\theta\cos\theta \\ \widetilde{C}_{2,Y}\cos^2\theta + \widetilde{C}_{1,Y}\sin^2\theta - \left(\widetilde{C}_{3,Y} + \widetilde{C}_{4,Y}\right)\sin\theta\cos\theta \end{bmatrix} \left(\Delta W_{2,t}^2 - \Delta t\right) \tag{B.14}$$

$$\begin{bmatrix} \left(\widetilde{C}_{3,X} + \widetilde{C}_{4,X}\right)\left(\cos^2\theta - \sin^2\theta\right) + 2\left(\widetilde{C}_{2,X} - \widetilde{C}_{1,X}\right)\sin\theta\cos\theta \\ \left(\widetilde{C}_{3,Y} + \widetilde{C}_{4,Y}\right)\left(\cos^2\theta - \sin^2\theta\right) + 2\left(\widetilde{C}_{2,Y} - \widetilde{C}_{1,Y}\right)\sin\theta\cos\theta \end{bmatrix} \Delta W_{1,t}\Delta W_{2,t}$$

$$\pm \begin{bmatrix} \widetilde{C}_{3,X} - \widetilde{C}_{4,X} \\ \widetilde{C}_{3,Y} - \widetilde{C}_{4,Y} \end{bmatrix} \left[\underline{L}_{(1,2)}\right]_t^{t+\Delta t} \ ,$$

where:

$\widetilde{C}_{1,X}$=$C_{1,X}$+$\theta_{X_1} b_{1,1} b_{1,2}$+$\theta_{X_2} b_{1,2} b_{2,1}$	$\widetilde{C}_{1,Y}$=$C_{1,Y}$+$\theta_{X_1} b_{1,1} b_{2,2}$+$\theta_{X_2} b_{2,1} b_{2,2}$
$\widetilde{C}_{2,X}$=$C_{2,X}$-$\theta_{X_1} b_{1,1} b_{1,2}$-$\theta_{X_2} b_{1,1} b_{2,2}$	$\widetilde{C}_{2,Y}$=$C_{2,Y} - \theta_{X_1} b_{2,1} b_{1,2} - \theta_{X_2} b_{2,1} b_{2,2}$
$\widetilde{C}_{3,X}$=$C_{3,X} - \theta_{X_1} b_{1,1}^2 - \theta_{X_2} b_{1,1} b_{2,1}$	$\widetilde{C}_{3,Y}$=$C_{3,Y} - \theta_{X_1} b_{1,1} b_{2,1} - \theta_{X_2} b_{2,1}^2$
$\widetilde{C}_{4,X}$=$C_{4,X} + \theta_{X_1} b_{1,2}^2 + \theta_{X_2} b_{1,2} b_{2,2}$	$\widetilde{C}_{4,Y}$=$C_{4,Y} + \theta_{X_1} b_{1,2} b_{2,2} + \theta_{X_2} b_{2,2}^2$

.

Using the definition of the orthogonal Wiener process (B.10), one can deduce the following expressions:

$$\begin{aligned}
\Delta\widetilde{W}_{1,t}^2 &= \cos^2\theta\ \Delta W_{1,t}^2 \mp 2\sin\theta\cos\theta\ \Delta W_{1,t}\Delta W_{2,t} + \sin^2\theta\ \Delta W_{2,t}^2\ , \\
\Delta\widetilde{W}_{2,t}^2 &= \sin^2\theta\ \Delta W_{1,t}^2 \pm 2\sin\theta\cos\theta\ \Delta W_{1,t}\Delta W_{2,t} + \cos^2\theta\ \Delta W_{2,t}^2\ , \\
\Delta\widetilde{W}_{1,t}\Delta\widetilde{W}_{2,t} &= \sin\theta\cos\theta\left(\Delta W_{1,t}^2 - \Delta W_{2,t}^2\right) \pm \left(\cos^2\theta - \sin^2\theta\right)\Delta W_{1,t}\Delta W_{2,t}\ .
\end{aligned}$$

Using the last equations, (B.14) can be simplified by:

$$R_M = \begin{bmatrix} \widetilde{C}_{1,X} \\ \widetilde{C}_{1,Y} \end{bmatrix}\left(\Delta\widetilde{W}_{1,t}^2 - \Delta t\right) + \begin{bmatrix} \widetilde{C}_{2,X} \\ \widetilde{C}_{2,Y} \end{bmatrix}\left(\Delta\widetilde{W}_{2,t}^2 - \Delta t\right) \tag{B.15}$$

$$+ \begin{bmatrix} \widetilde{C}_{3,X} + \widetilde{C}_{4,X} \\ \widetilde{C}_{3,Y} + \widetilde{C}_{4,Y} \end{bmatrix}\Delta\widetilde{W}_{1,t}\Delta\widetilde{W}_{2,t} + \begin{bmatrix} \pm\left(\widetilde{C}_{3,X} - \widetilde{C}_{4,X}\right) \\ \pm\left(\widetilde{C}_{3,Y} - \widetilde{C}_{4,Y}\right) \end{bmatrix}\left[\underline{L}_{(1,2)}\right]_t^{t+\Delta t}\ ,$$

and the Lie bracket is equal to:

$$R_L = \begin{bmatrix} \pm\left(C_{3,X} - C_{4,X} - \theta_{X_1}\left(b_{1,1}^2 + b_{1,2}^2\right) - \theta_{X_2}\left(b_{1,1}b_{2,1} + b_{1,2}b_{2,2}\right)\right) \\ \pm\left(C_{3,Y} - C_{4,Y} - \theta_{X_1}\left(b_{1,1}b_{2,1} + b_{1,2}b_{2,2}\right) - \theta_{X_2}\left(b_{2,1}^2 + b_{2,2}^2\right)\right) \end{bmatrix}\ . \tag{B.16}$$

To make zero (B.16), one needs:

$$\begin{aligned}
\left(C_{3,X} - C_{4,X}\right) &= \theta_{X_1}\left(b_{1,1}^2 + b_{1,2}^2\right) + \theta_{X_2}\left(b_{1,1}b_{2,1} + b_{1,2}b_{2,2}\right)\ , \\
\left(C_{3,Y} - C_{4,Y}\right) &= \theta_{X_1}\left(b_{1,1}b_{2,1} + b_{1,2}b_{2,2}\right) + \theta_{X_2}\left(b_{2,1}^2 + b_{2,2}^2\right)\ ,
\end{aligned}$$

so then, for θ_{X_1}:

$$\theta_{X_1} = \frac{\left(C_{3,X} - C_{4,X}\right)\left(b_{2,1}^2 + b_{2,2}^2\right) - \left(C_{3,Y} - C_{4,Y}\right)\left(b_{1,1}b_{2,1} + b_{1,2}b_{2,2}\right)}{\left(b_{1,1}^2 + b_{1,2}^2\right)\left(b_{2,1}^2 + b_{2,2}^2\right) - \left(b_{1,1}b_{2,1} + b_{1,2}b_{2,2}\right)^2}\ ,$$

and for θ_{X_2}:

$$\theta_{X_2} = \frac{\left(C_{3,X} - C_{4,X}\right)\left(b_{1,1}b_{2,1} + b_{1,2}b_{2,2}\right) - \left(C_{3,Y} - C_{4,Y}\right)\left(b_{1,1}^2 + b_{1,2}^2\right)}{\left(b_{1,1}b_{2,1} + b_{1,2}b_{2,2}\right)^2 - \left(b_{1,1}^2 + b_{1,2}^2\right)\left(b_{2,1}^2 + b_{2,2}^2\right)}\ .$$

If:

$$\left(b_{1,1}b_{2,2} - b_{1,2}b_{2,1}\right)^2 = \left(b_{1,1}^2 + b_{1,2}^2\right)\left(b_{2,1}^2 + b_{2,2}^2\right) - \left(b_{1,1}b_{2,1} + b_{1,2}b_{2,2}\right)^2\ ,$$

$$\begin{bmatrix} H_1^- \\ H_2^- \end{bmatrix} = \begin{bmatrix} C_{3,X} - C_{4,X} \\ C_{3,Y} - C_{4,Y} \end{bmatrix} ,$$

then:

$$\Phi \doteq \frac{\partial\theta}{\partial X_1} = \frac{H_1^- \left(b_{2,1}^2 + b_{2,2}^2\right) - H_2^- \left(b_{1,1}b_{2,1} + b_{1,2}b_{2,2}\right)}{\left(b_{1,1}b_{2,2} - b_{1,2}b_{2,1}\right)^2} ,$$

$$\Psi \doteq \frac{\partial\theta}{\partial X_2} = \frac{H_2^- \left(b_{1,1}^2 + b_{1,2}^2\right) - H_1^- \left(b_{1,1}b_{2,1} + b_{1,2}b_{2,2}\right)}{\left(b_{1,1}b_{2,2} - b_{1,2}b_{2,1}\right)^2} .$$

Having example (4.6), you get:

$$\overline{\sigma}(x,y) = \begin{bmatrix} \sigma & 0 \\ \rho\,\xi & \widehat{\rho}\,\xi \end{bmatrix} \begin{bmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{bmatrix} ,$$

then:

$\widetilde{C}_{1,X} = \sigma\sigma_x + \rho\xi\sigma_y$	$\widetilde{C}_{1,Y} = \rho\sigma\xi_x + \rho^2\xi\xi_y + \theta_{X_1}\widehat{\rho}\sigma\xi + \theta_{X_2}\rho\widehat{\rho}\xi^2$
$\widetilde{C}_{2,X} = -\theta_{X_2}\widehat{\rho}\sigma\xi$	$\widetilde{C}_{2,Y} = \widehat{\rho}^2\xi\xi_y - \theta_{X_2}\rho\widehat{\rho}\xi^2$
$\widetilde{C}_{3,X} = -\theta_{X_1}\sigma^2 - \theta_{X_2}\rho\sigma\xi$	$\widetilde{C}_{3,Y} = \widehat{\rho}\sigma\xi_x + \rho\widehat{\rho}\xi\xi_y - \theta_{X_1}\rho\sigma\xi - \theta_{X_2}\rho^2\xi^2$
$\widetilde{C}_{4,X} = \widehat{\rho}\xi\sigma_y$	$\widetilde{C}_{4,Y} = \rho\widehat{\rho}\xi\xi_y + \theta_{X_2}\widehat{\rho}^2\xi^2$

,

and if one uses:

$$\Phi \doteq \frac{\partial\theta}{\partial X_1} = \frac{\xi^2\sigma_y + \rho\sigma^2\xi_x}{-\widehat{\rho}\,\sigma^2\xi} = \frac{-1}{\widehat{\rho}}\left(\frac{\xi\sigma_y}{\sigma^2} + \frac{\rho\xi_x}{\xi}\right) ,$$

$$\Psi \doteq \frac{\partial\theta}{\partial X_2} = \frac{\sigma^2\xi_x + \rho\xi^2\sigma_y}{\widehat{\rho}\,\sigma\xi^2} = \frac{1}{\widehat{\rho}}\left(\frac{\sigma\xi_x}{\xi^2} + \frac{\rho\sigma_y}{\sigma}\right) ,$$

then:

$\widetilde{C}_{1,X} = \sigma\sigma_x + \rho\xi\sigma_y$	$\widetilde{C}_{1,Y} = \rho\sigma\xi_x + \rho^2\xi\xi_y - \frac{\widehat{\rho}^2\xi^2\sigma_y}{\sigma}$
$\widetilde{C}_{2,X} = -\rho\xi\sigma_y - \frac{\sigma^2\xi_x}{\xi}$	$\widetilde{C}_{2,Y} = -\rho\sigma\xi_x + \widehat{\rho}^2\xi\xi_y - \frac{\rho^2\xi^2\sigma_y}{\sigma}$
$\widetilde{C}_{3,X} = \widehat{\rho}\,\xi\sigma_y$	$\widetilde{C}_{3,Y} = \widehat{\rho}\sigma\xi_x + \rho\widehat{\rho}\left(\xi\xi_y + \frac{\xi^2\sigma_y}{\sigma}\right)$
$\widetilde{C}_{4,X} = \widetilde{C}_{3,X}$	$\widetilde{C}_{4,Y} = \widetilde{C}_{3,Y}$

. (B.17)

Having example (4.16), you get:

$$\widehat{\sigma}\,(x,y) = \begin{bmatrix} \alpha\,x^{\gamma_1}y^{\lambda_1} & 0 \\ \rho\beta\,x^{\gamma_2}y^{\lambda_2} & \widehat{\rho}\beta\,x^{\gamma_2}y^{\lambda_2} \end{bmatrix} \begin{bmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{bmatrix} ,$$

then:

$$\begin{bmatrix} \widetilde{H}_1^- \\ \widetilde{H}_2^- \end{bmatrix} = \begin{bmatrix} -\widehat{\rho}\lambda_1\alpha\beta\,x^{\gamma_1+\gamma_2}y^{\lambda_2+\lambda_1-1} \\ \widehat{\rho}\gamma_2\alpha\,\beta x^{\gamma_1+\gamma_2-1}y^{\lambda_1+\lambda_2} \end{bmatrix} ,$$

and

$$\begin{array}{ll} \Phi = \frac{\lambda_1\beta\,y^{\lambda_C}}{-\widehat{\rho}\alpha\,x^{\gamma_C+1}} + \frac{\rho\gamma_2}{-\widehat{\rho}\,x} , & \Psi = \frac{\gamma_2\alpha\,x^{\gamma_C}}{\widehat{\rho}\beta\,y^{\lambda_C+1}} + \frac{\rho\lambda_1}{\widehat{\rho}\,y} , \\ \Phi_X = \frac{(\gamma_C+1)\lambda_1\beta\,y^{\lambda_C}}{\widehat{\rho}\alpha\,x^{\gamma_C}} + \frac{\rho\gamma_2}{\widehat{\rho}\,x^2} , & \Psi_X = \frac{\gamma_C\gamma_2\alpha\,x^{\gamma_C-1}}{\widehat{\rho}\beta\,y^{\lambda_C+1}} , \\ \Phi_Y = \frac{\lambda_C\lambda_1\beta\,y^{\lambda_C-1}}{-\widehat{\rho}\alpha\,x^{\gamma_C+1}} , & \Psi_Y = \frac{(\lambda_C+1)\gamma_2\alpha\,x^{\gamma_C}}{-\widehat{\rho}\beta\,y^{\lambda_C}} + \frac{\rho\lambda_1}{-\widehat{\rho}\,y^2} . \end{array} \quad \text{(B.18)}$$

B.3.3 θ Scheme

We start with the following 3−Dimensional Itô SDE with a 2−Dimensional Wiener process:

$$d\begin{bmatrix} \widetilde{X}_{1,t} \\ \widetilde{X}_{2,t} \\ \theta_t \end{bmatrix} = \begin{bmatrix} a_1 \\ a_2 \\ a_\theta \end{bmatrix} dt + \begin{bmatrix} b_{1,1} & b_{1,2} \\ b_{2,1} & b_{2,2} \\ b_{\theta,1} & b_{\theta,2} \end{bmatrix} \begin{bmatrix} d\widetilde{W}_{1,t} \\ d\widetilde{W}_{2,t} \end{bmatrix} , \tag{B.19}$$

where:

$$a_\theta = \Phi a_1 + \Psi a_2 ,$$

$$\begin{bmatrix} b_{\theta,1} & b_{\theta,2} \end{bmatrix} = \begin{bmatrix} \Phi b_{1,1} + \Psi b_{2,1} & \Phi b_{1,2} + \Psi\, b_{2,2} \end{bmatrix} ,$$

$$\Phi \doteq \frac{H_1^- \left(b_{2,1}^2 + b_{2,2}^2\right) - H_2^- \left(b_{1,1}b_{2,1} + b_{1,2}b_{2,2}\right)}{\left(b_{1,1}b_{2,2} - b_{1,2}b_{2,1}\right)^2} ,$$

$$\Psi \doteq \frac{H_2^- \left(b_{1,1}^2 + b_{1,2}^2\right) - H_1^- \left(b_{1,1}b_{2,1} + b_{1,2}b_{2,2}\right)}{\left(b_{1,1}b_{2,2} - b_{1,2}b_{2,1}\right)^2} .$$

The 1 strong order Milstein scheme for (B.19) with time step Δ_t is:

$$\begin{bmatrix} \widehat{\widetilde{X}}_{1,t+\Delta t} \\ \widehat{\widetilde{X}}_{2,t+\Delta t} \\ \widehat{\theta}_{t+\Delta t} \end{bmatrix} = \begin{bmatrix} \widehat{\widetilde{X}}_{1,t} \\ \widehat{\widetilde{X}}_{2,t} \\ \widehat{\theta}_t \end{bmatrix} + \begin{bmatrix} a_1 \\ a_2 \\ a_\theta \end{bmatrix} \Delta t + \begin{bmatrix} \widetilde{b}_{1,1} & \widetilde{b}_{1,2} \\ \widetilde{b}_{2,1} & \widetilde{b}_{2,2} \\ b_{\theta,1} & b_{\theta,2} \end{bmatrix} \begin{bmatrix} \Delta W_{1,t} \\ \Delta W_{2,t} \end{bmatrix} + \frac{1}{2} R_M , \tag{B.20}$$

$$R_M = \sum_{j=1}^{2} \begin{bmatrix} \widetilde{L}_j \widetilde{b}_{1,j} \\ \widetilde{L}_j \widetilde{b}_{2,j} \\ \widetilde{L}_j \widetilde{b}_{3,j} \end{bmatrix} \left(\Delta W_{j,t}^2 - \Delta t\right)$$

$$+ \begin{bmatrix} \widetilde{H}_1^+ \\ \widetilde{H}_2^+ \\ \widetilde{H}_3^+ \end{bmatrix} \Delta W_{1,t} \Delta W_{2,t} + \begin{bmatrix} \widetilde{H}_1^- \\ \widetilde{H}_2^- \\ \widetilde{H}_3^- \end{bmatrix} \left[\underline{\overline{L}}_{(1,2)}\right]_t^{t+\Delta t} ,$$

where:

$$\widetilde{H}_j^\pm = \widetilde{L}_1 \widetilde{b}_{j,2} \pm \widetilde{L}_2 \widetilde{b}_{j,1} .$$

Doing some computation, (B.20) is equal to:

$$R_M = \begin{bmatrix} \widetilde{C}_{1,X} \\ \widetilde{C}_{1,Y} \\ \widetilde{C}_{1,\theta} \end{bmatrix} \left(\Delta \widetilde{W}_{1,t}^2 - \Delta t\right) + \begin{bmatrix} \widetilde{C}_{2,X} \\ \widetilde{C}_{2,Y} \\ \widetilde{C}_{2,\theta} \end{bmatrix} \left(\Delta \widetilde{W}_{2,t}^2 - \Delta t\right) \tag{B.21}$$

$$+ \begin{bmatrix} \widetilde{C}_{3,X} + \widetilde{C}_{4,X} \\ \widetilde{C}_{3,Y} + \widetilde{C}_{4,Y} \\ \widetilde{C}_{3,\theta} + \widetilde{C}_{4,\theta} \end{bmatrix} \Delta \widetilde{W}_{1,t} \Delta \widetilde{W}_{2,t} + \begin{bmatrix} \pm\left(\widetilde{C}_{3,X} - \widetilde{C}_{4,X}\right) \\ \pm\left(\widetilde{C}_{3,Y} - \widetilde{C}_{4,Y}\right) \\ \pm\left(\widetilde{C}_{3,\theta} - \widetilde{C}_{4,\theta}\right) \end{bmatrix} \left[\underline{\overline{L}}_{(1,2)}\right]_t^{t+\Delta t} ,$$

where:

$\widetilde{C}_{1,X} = C_{1,X} + \Phi b_{1,1}b_{1,2} + \Psi b_{1,2}b_{2,1}$	$\widetilde{C}_{1,Y} = C_{1,Y} + \Phi b_{1,1}b_{2,2} + \Psi b_{2,1}b_{2,2}$
$\widetilde{C}_{2,X} = C_{2,X} - \Phi b_{1,1}b_{1,2} - \Psi b_{1,1}b_{2,2}$	$\widetilde{C}_{2,Y} = C_{2,Y} - \Phi b_{2,1}b_{1,2} - \Psi b_{2,1}b_{2,2}$
$\widetilde{C}_{3,X} = C_{3,X} - \Phi b_{1,1}^2 - \Psi b_{1,1}b_{2,1}$	$\widetilde{C}_{3,Y} = C_{3,Y} - \Phi b_{1,1}b_{2,1} - \Psi b_{2,1}^2$
$\widetilde{C}_{4,X} = C_{4,X} + \Phi b_{1,2}^2 + \Psi b_{1,2}b_{2,2}$	$\widetilde{C}_{4,Y} = C_{4,Y} + \Phi b_{1,2}b_{2,2} + \Psi b_{2,2}^2$

,

$\widetilde{C}_{1,\theta} = b_{1,1}b_{\theta,1_{\tilde{X}_1}} + b_{2,1}b_{\theta,1_{\tilde{X}_2}} + b_{\theta,1}b_{\theta,2}$
$\widetilde{C}_{2,\theta} = b_{1,2}b_{\theta,2_{\tilde{X}_1}} + b_{2,2}b_{\theta,2_{\tilde{X}_2}} - b_{\theta,1}b_{\theta,2}$
$\widetilde{C}_{3,\theta} = b_{1,1}b_{\theta,2_{\tilde{X}_1}} + b_{2,1}b_{\theta,2_{\tilde{X}_2}} - b_{\theta,1}^2$
$\widetilde{C}_{4,\theta} = b_{1,2}b_{\theta,1_{\tilde{X}_1}} + b_{2,2}b_{\theta,1_{\tilde{X}_2}} + b_{\theta,2}^2$

.

Doing some operations, the Lie bracket is equal to:

$$\begin{bmatrix} \widetilde{H}_1^- \\ \widetilde{H}_2^- \\ \widetilde{H}_3^- \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ \pm\left(b_{1,1}b_{2,2} - b_{1,2}b_{2,1}\right)\left(\Psi_{\tilde{X}_1} - \Phi_{\tilde{X}_2}\right) \end{bmatrix}. \tag{B.22}$$

Knowing that:

$$\begin{bmatrix} b_{\theta,1} & b_{\theta,2} \end{bmatrix} = \begin{bmatrix} \Phi b_{1,1} + \Psi b_{2,1} & \Phi b_{1,2} + \Psi\, b_{2,2} \end{bmatrix},$$

then:

$\widetilde{C}_{1,\theta} = \Phi\left(C_{1,X} + \Phi b_{1,1}b_{1,2}\right) + \Psi\left(C_{1,Y} + \Psi b_{2,1}\, b_{2,2}\right) + \Phi\Psi\left(b_{1,1}\, b_{2,2} + b_{1,2}b_{2,1}\right)$
$+\Phi_{\tilde{X}_1} b_{1,1}^2 + \Psi_{\tilde{X}_2} b_{2,1}^2 + \left(\Phi_{\tilde{X}_2} + \Psi_{\tilde{X}_1}\right) b_{1,1}b_{2,1}$
$\widetilde{C}_{2,\theta} = \Phi\left(C_{2,X} - \Phi b_{1,1}b_{1,2}\right) + \Psi\left(C_{2,Y} - \Psi b_{2,1}\, b_{2,2}\right) - \Phi\Psi\left(b_{1,1}\, b_{2,2} + b_{1,2}b_{2,1}\right)$
$+\Phi_{\tilde{X}_1} b_{1,2}^2 + \Psi_{\tilde{X}_2} b_{2,2}^2 + \left(\Phi_{\tilde{X}_2} + \Psi_{\tilde{X}_1}\right) b_{1,2}b_{2,2}$
$\widetilde{C}_{3,\theta} = \Phi\left(C_{3,X} - \Phi b_{1,1}^2\right) + \Psi\left(C_{3,Y} - \Psi b_{2,1}^2\right) - 2\Phi\Psi b_{1,1}b_{2,1}$
$+\Phi_{\tilde{X}_1} b_{1,1}b_{1,2} + \Phi_{\tilde{X}_2} b_{1,2}b_{2,1} + \Psi_{\tilde{X}_1}\, b_{1,1}b_{2,2} + \Psi_{\tilde{X}_2}\, b_{2,1}b_{2,2}$
$\widetilde{C}_{4,\theta} = \Phi\left(C_{4,X} + \Phi b_{1,2}^2\right) + \Psi\left(C_{4,Y} + \Psi\, b_{2,2}^2\right) + 2\Phi\Psi b_{1,2}\, b_{2,2}$
$+\Phi_{\tilde{X}_1} b_{1,1}b_{1,2} + \Psi_{\tilde{X}_1} b_{1,2}b_{2,1} + \Phi_{\tilde{X}_2} b_{1,1}b_{2,2} + \Psi_{\tilde{X}_2} b_{2,1}b_{2,2}$

.

Having example (4.12), you get:

$$\overline{\sigma}(x,y) = \begin{bmatrix} \sigma & 0 \\ \rho\,\xi & \widehat{\rho}\,\xi \\ \sigma\Phi & \xi\Psi \end{bmatrix} \begin{bmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{bmatrix},$$

then:

$\widetilde{C}_{1,\theta} = -\widehat{\rho}\left(2\sigma_y\xi_x + \xi\sigma_{xy} + \frac{\xi}{\sigma}\left(-\sigma_x\sigma_y + \rho\left(\sigma_y\xi_y + \xi\sigma_{yy}\right)\right)\right)$
$\widetilde{C}_{2,\theta} = \widehat{\rho}\left(2\sigma_y\xi_x + \sigma\xi_{xy} + \frac{\xi}{\sigma}\rho\left(\sigma_y\xi_y + \xi\sigma_{yy}\right) - \frac{\sigma}{\xi}\xi_x\xi_y\right)$
$\widetilde{C}_{3,\theta} = \rho\left(2\sigma_y\xi_x + \sigma\xi_{xy} + \xi\sigma_{xy} + \frac{\xi}{\sigma}\left(-\sigma_x\sigma_y + \rho\left(\sigma_y\xi_y + \xi\sigma_{yy}\right)\right)\right)$
$+\frac{\sigma}{\xi}\left(\sigma\xi_{xx} + \xi_x\left(\sigma_x - \rho\xi_y\right)\right) + \frac{-\sigma^2\xi_x^2}{\xi^2} + \frac{-\xi^2\sigma_y^2}{\sigma^2}$
$\widetilde{C}_{4,\theta} = \rho\left(2\sigma_y\xi_x + \frac{\xi}{\sigma}\rho\left(\sigma_y\xi_y + \xi\sigma_{yy}\right)\right) + \frac{\xi^2\sigma_y^2}{\sigma^2} + \frac{\sigma^2\xi_x^2}{\xi^2} - \frac{\xi}{\sigma}\left(\sigma_y\xi_y + \xi\sigma_{yy}\right)$

,

and $\widetilde{H}_3^-$ from the Lie bracket (B.22) is equal to:

$$\begin{aligned}\widetilde{H}_3^- &= \rho\left(\sigma\xi_{xy}+\xi\sigma_{xy}-\frac{\xi}{\sigma}\sigma_x\sigma_y\right)+\frac{\sigma}{\xi}\left(\sigma\xi_{xx}+\xi_x\left(\sigma_x-\rho\xi_y\right)\right)\\ &\quad+\frac{-2\xi^2\sigma_y^2}{\sigma^2}+\frac{-2\sigma^2\xi_x^2}{\xi^2}+\frac{\xi}{\sigma}\left(\sigma_y\xi_y+\xi\sigma_{yy}\right)\ .\end{aligned}$$

B.4 Orthogonal Transformation Theorems

In this section, we shall present two theorems where we prove that if an orthogonal transformation is applied to a Wiener process dW, then the new orthogonal Wiener process $d\widetilde{W}$ is independent and have the same distribution as the original process dW.

Theorem 6: Distribution of an orthogonal Wiener processes

If $dW_{1,t}$, $dW_{2,t}$ are two independent Wiener process and you apply an orthogonal transformation to them:

$$\left[\begin{array}{c} d\widetilde{W}_{1,t} \\ d\widetilde{W}_{2,t}\end{array}\right]=\left[\begin{array}{cc}\cos\theta_t & \sin\theta_t \\ \mp\sin\theta_t & \pm\cos\theta_t\end{array}\right]\left[\begin{array}{c} dW_{1,t} \\ dW_{2,t}\end{array}\right]\ , \tag{B.23}$$

then:

- A) the new orthogonal Wiener processes, $d\widetilde{W}_{1,t}$ and $d\widetilde{W}_{2,t}$ are independent.
- B) $dW_{i,t}$ and $d\widetilde{W}_{i,t}$ have the same distribution.

Proof:

A) If $dW_{1,t}$ and $dW_{2,t}$ are independent, then:

$$E\left[dW_{1,t}dW_{2,t}\right]=0\ .$$

Doing the same for the orthogonal Wiener process:

$$E\left[d\widetilde{W}_{1,t}d\widetilde{W}_{2,t}\right]=E\left[\begin{array}{c}\mp\sin\theta\cos\theta\left(dW_{1,t}^2-dW_{2,t}^2\right) \\ \pm dW_{1,t}dW_{2,t}\left(\cos^2\theta-\sin^2\theta\right)\end{array}\right]=0\ .$$

B) The probability density function (PDF) of an $N-$dimensional multivariate normal is [13]:

$$f(x)=f(x_1,x_2,...x_N)=\frac{1}{(2\pi)^{\frac{N}{2}}\sqrt{\det\left(\sum\right)}}\exp\left(-\frac{1}{2}\left(x-\mu\right)^T\sum^{-1}\left(x-\mu\right)\right)\ ,$$

where $\mu = [\mu_1, \mu_2, ..., \mu_N]^T$ is the mean and $\sum$ is the covariance matrix (positive-definite real $N\ x\ N$ matrix). If one calculates them for dW_t:

$$\mu_{dW_t} = [E[dW_{1,t}]\ \ E[dW_{2,t}]] = 0\ ,$$

$$\sum\nolimits_{dW_t} = \begin{bmatrix} V[dW_{1,t}] & \rho\sqrt{V[dW_{1,t}]}\sqrt{V[dW_{2,t}]} \\ \rho\sqrt{V[dW_{1,t}]}\sqrt{V[dW_{2,t}]} & V[dW_{2,t}] \end{bmatrix} = \begin{bmatrix} dt & 0 \\ 0 & dt \end{bmatrix}\ .$$

If dW_t and $d\widetilde{W}_t$ have the same distribution, then they have the same mean and covariance matrix.

$$\mu_{d\widetilde{W}_t} = \left[E\left[d\widetilde{W}_{1,t}\right]\ \ E\left[d\widetilde{W}_{2,t}\right]\right] = 0 = \mu_{dW_t}\ ,$$

$$\sum\nolimits_{d\widetilde{W}_t} = \begin{bmatrix} dt & 0 \\ 0 & dt \end{bmatrix} = \sum\nolimits_{dW_t} \qquad \square$$

Theorem 7: General representation for an orthogonal matrix

If $\Gamma(\Theta_t)$ is an orthogonal matrix:

$$\Gamma(\Theta_t) := \begin{bmatrix} \Theta_{1,1} & \Theta_{1,2} \\ \Theta_{2,1} & \Theta_{2,2} \end{bmatrix}\ ,$$

$$\Theta_{1,1}\Theta_{2,2} - \Theta_{1,2}\Theta_{2,1} = \pm 1 = \psi\ ,$$

then:

- *A)* $\Gamma(\Theta_t)$ can have only two families of representations:

$$\Gamma(\Theta_t) = \begin{bmatrix} \Theta_{1,1} & \Theta_{1,2} \\ \mp\Theta_{1,2} & \pm\Theta_{1,1} \end{bmatrix}\ , \qquad \left\{ \begin{array}{l} \Theta_{1,1} = \pm\Theta_{2,2} \\ \Theta_{2,1} = \mp\Theta_{1,2} \end{array} \right\}.$$

- *B)* If $\Gamma(\Theta_t)$ and $\Xi(\phi_t)$ are independent orthogonal matrices, the multiplication of both still has the same two representation as point *A)*.

Proof:

A) Using the definition of an orthogonal matrix:

$$\Gamma^{-1} = \Gamma^T\ ,$$

then:

$$\Gamma\Gamma^{-1} = \Gamma\Gamma^T = \psi \qquad \text{and} \qquad \Gamma^{-1}\Gamma = \Gamma^T\Gamma = \psi\ ,$$

computing operations:

$$\begin{aligned}
\Gamma\Gamma^T &= \begin{bmatrix} \Theta_{1,1}^2+\Theta_{1,2}^2 & \Theta_{1,1}\Theta_{2,1}+\Theta_{1,2}\Theta_{2,2} \\ \Theta_{1,1}\Theta_{2,1}+\Theta_{1,2}\Theta_{2,2} & \Theta_{2,1}^2+\Theta_{2,2}^2 \end{bmatrix} = \begin{bmatrix} \psi & 0 \\ 0 & \psi \end{bmatrix} , \\
\Gamma^T\Gamma &= \begin{bmatrix} \Theta_{1,1}^2+\Theta_{2,1}^2 & \Theta_{1,1}\Theta_{1,2}+\Theta_{2,1}\Theta_{2,2} \\ \Theta_{1,1}\Theta_{1,2}+\Theta_{2,1}\Theta_{2,2} & \Theta_{1,2}^2+\Theta_{2,2}^2 \end{bmatrix} = \begin{bmatrix} \psi & 0 \\ 0 & \psi \end{bmatrix} ,
\end{aligned}$$

then:

$$\begin{aligned}
\Theta_{1,1}\Theta_{2,1}+\Theta_{1,2}\Theta_{2,2} &= 0 , \\
\Theta_{1,1}\Theta_{1,2}+\Theta_{2,1}\Theta_{2,2} &= 0 \qquad \text{then} \quad \Theta_{1,2} = \pm\Theta_{2,1} .
\end{aligned}$$

Computing for the two solutions:

$$\begin{aligned}
if \quad \Theta_{1,2} &= -\Theta_{2,1} \quad \text{then} \quad \Theta_{1,1} = \Theta_{2,2} , \\
if \quad \Theta_{1,2} &= +\Theta_{2,1} \quad \text{then} \quad \Theta_{1,1} = -\Theta_{2,2} .
\end{aligned}$$

B): Using point A):

$$\Gamma\left(\Theta_t\right) := \begin{bmatrix} \Theta_{1,1} & \Theta_{1,2} \\ \mp\Theta_{1,2} & \pm\Theta_{1,1} \end{bmatrix} , \qquad \Xi\left(\phi_t\right) := \begin{bmatrix} \phi_{1,1} & \phi_{1,2} \\ \mp\phi_{1,2} & \pm\phi_{1,1} \end{bmatrix} .$$

Doing the multiplication of both matrices:

$$\begin{aligned}
\text{solution 1} \quad : \qquad \Gamma\left(\Theta_t\right)\Xi\left(\phi_t\right) &= \begin{bmatrix} \Theta_{1,1} & \Theta_{1,2} \\ \mp\Theta_{1,2} & \pm\Theta_{1,1} \end{bmatrix} \begin{bmatrix} \phi_{1,1} & \phi_{1,2} \\ -\phi_{1,2} & \phi_{1,1} \end{bmatrix} \\
&= \begin{bmatrix} \Theta_{1,1}\phi_{1,1}-\Theta_{1,2}\phi_{1,2} & \Theta_{1,1}\phi_{1,2}+\Theta_{1,2}\phi_{1,1} \\ \mp\left(\Theta_{1,1}\phi_{1,2}+\Theta_{1,2}\phi_{1,1}\right) & \pm\left(\Theta_{1,1}\phi_{1,1}-\Theta_{1,2}\phi_{1,2}\right) \end{bmatrix} \\
&= \begin{bmatrix} \Phi_{1,1} & \Phi_{1,2} \\ \mp\Phi_{1,2} & \pm\Phi_{1,1} \end{bmatrix} ,
\end{aligned}$$

$$\begin{aligned}
\text{solution 2} \quad : \qquad \Gamma\left(\Theta_t\right)\Xi\left(\phi_t\right) &= \begin{bmatrix} \Theta_{1,1} & \Theta_{1,2} \\ \mp\Theta_{1,2} & \pm\Theta_{1,1} \end{bmatrix} \begin{bmatrix} \phi_{1,1} & \phi_{1,2} \\ \phi_{1,2} & -\phi_{1,1} \end{bmatrix} \\
&= \begin{bmatrix} \Theta_{1,1}\phi_{1,1}+\Theta_{1,2}\phi_{1,2} & \Theta_{1,1}\phi_{1,2}-\Theta_{1,2}\phi_{1,1} \\ \pm\left(\Theta_{1,1}\phi_{1,2}-\Theta_{1,2}\phi_{1,1}\right) & \mp\left(\Theta_{1,1}\phi_{1,1}+\Theta_{1,2}\phi_{1,2}\right) \end{bmatrix} \\
&= \begin{bmatrix} \widehat{\Phi}_{1,1} & \widehat{\Phi}_{1,2} \\ \pm\widehat{\Phi}_{1,2} & \mp\widehat{\Phi}_{1,1} \end{bmatrix}
\end{aligned}$$

□

Appendix C

MSL-MC

C.1 MSL-MC Algorithm

This section defines the algorithm one has to use to estimate an exotic option price using the Multischeme Multilevel Montecarlo method (MSL-MC). The MSL-MC is an updated algorithm of the ML-MC [10].

The expectation of a payoff P with maturity T is calculated by:

$$E\left[P_{L_F}\right] = E\left[P_{L_O}\right] + \sum_{L=L_O+1}^{L_F} E\left[P_L - P_{L-1}\right] \ . \tag{C.1}$$

Or using multi-schemes in the simulation:

$$E\left[P_{L_F}\right] = E\left[P_{S_{L_O}}\right] + \sum_{L=L_O+1}^{L_F} E\left[P_{S_L} - P_{S_{L-1}}\right] \ ,$$

where L is the level of the algorithm that simulates the scheme or time approximation with different time steps Δt:

$$\Delta t = \frac{T - t_0}{m^L} \qquad \text{for } m = 2, 4, 6...$$

L_O is the optimal starting level and P_{S_L} is the payoff value using the optimal scheme for level L. For a given ϵ:

$$\epsilon = |\text{ exact solution } - \text{ approximation }| \ ,$$

the algorithm has to simulate a repeated cycle for each level L, where one calculates the option price using M_L paths:

$$M_L = \left\lceil 2\epsilon^{-2}\sqrt{V_L \Delta t_L}\sum_{l=1}^{L}\sqrt{\frac{V_l}{\Delta t_l}}\right\rceil \ .$$

When $L = L_O$, use $M_L = 1000$. For $L > L_O$, simulate extra samples at each level as needed for new M_L. The algorithm will stop when it converges:

$$\max\left\{\frac{|P_{L-1}|}{m}, |P_L|\right\} < \frac{\epsilon}{\sqrt{2}}(m-1) \ .$$

Compared with the standard method and setting the Lévy Area equal to zero, the mean square error MSE,

$$MSE \approx c_1 dt + c_2 dt^2 = O\left(\epsilon^{-3}\right) \ ,$$

is reduced in some cases to:

scheme	Standard Method	ML-MC
Euler or Milstein scheme ($L = 0$)	$O\left(\epsilon^{-3}\right)$	$O\left(\epsilon^{-2}\left(\log\epsilon\right)^2\right)$
θ scheme ($L = 0$)	$O\left(\epsilon^{-3}\right)$	$O\left(\epsilon^{-2}\right)$

.

[10] and results in Chapter 5 demonstrate that $m = 2$ is the optimal to use for all schemes. Only for specific examples, $m = 4$ is used.

C.2 Strong Convergence Plots

This section presents the corresponding strong convergence tests for the asset S (Figure C.1), the variance v (Figure C.2), the rotation or angle θ (Figure C.3) and a European Put option price (Figure C.4). It is no surprise that all strong convergence plots are almost the same, having the same order of convergence as the European Call option plot (Figure 5.4) presented in Chapter 5. We have used the parameters and initial conditions (5.10) with the stochastic volatility models (5.2-5.5).

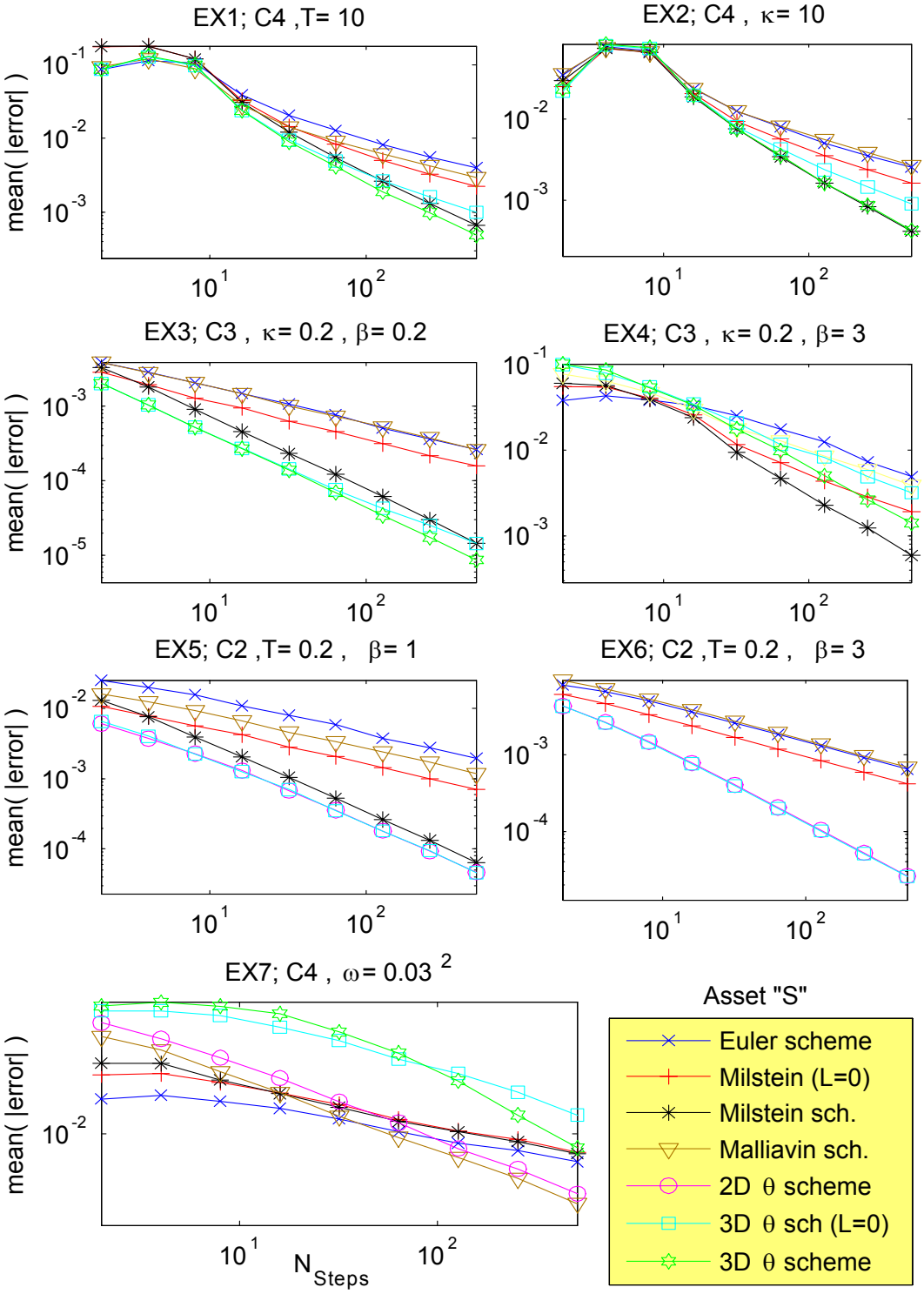

Figure C.1: Strong convergence tests for S using (5.10).

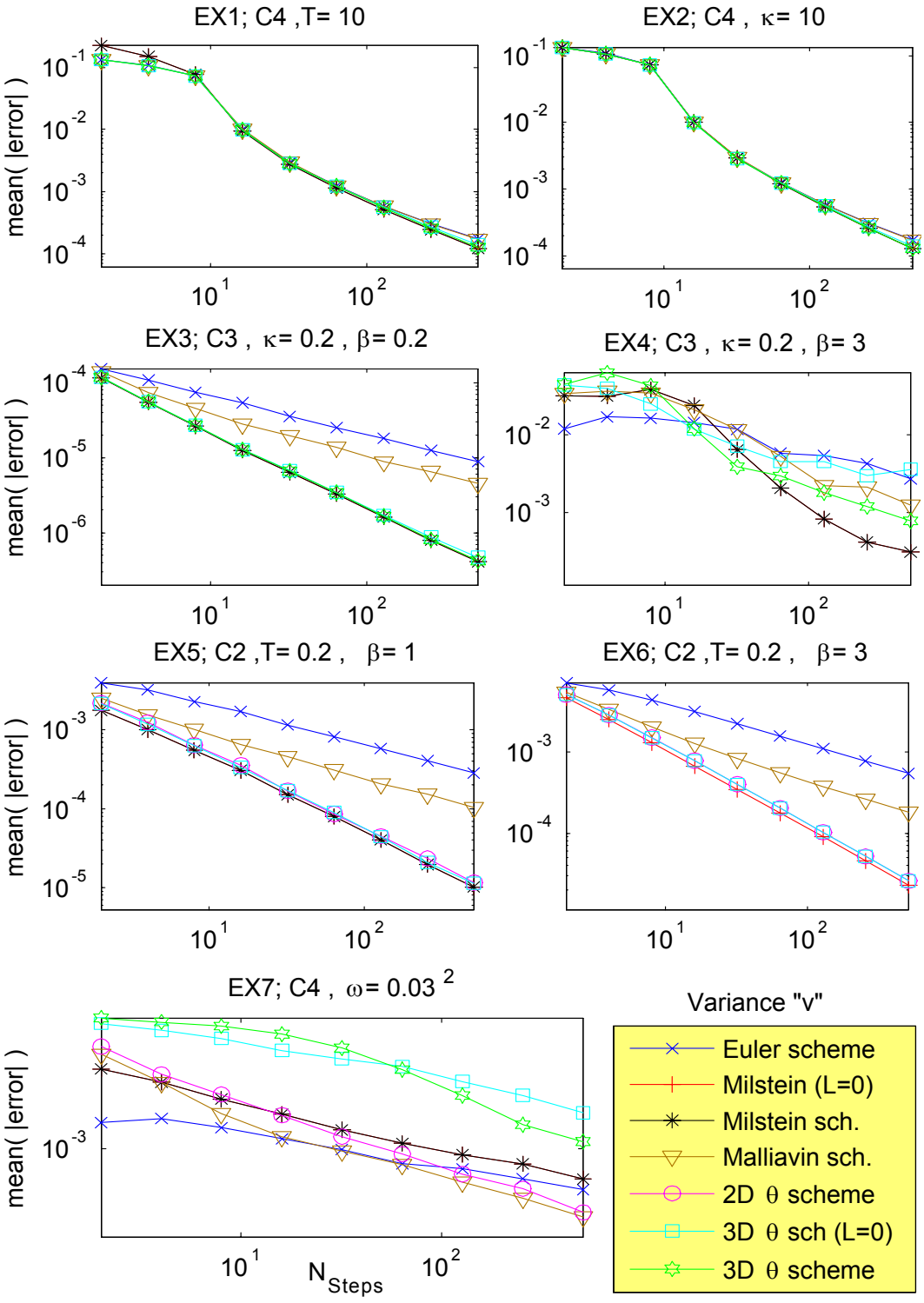

Figure C.2: Strong convergence tests for the variance ν using (5.10).

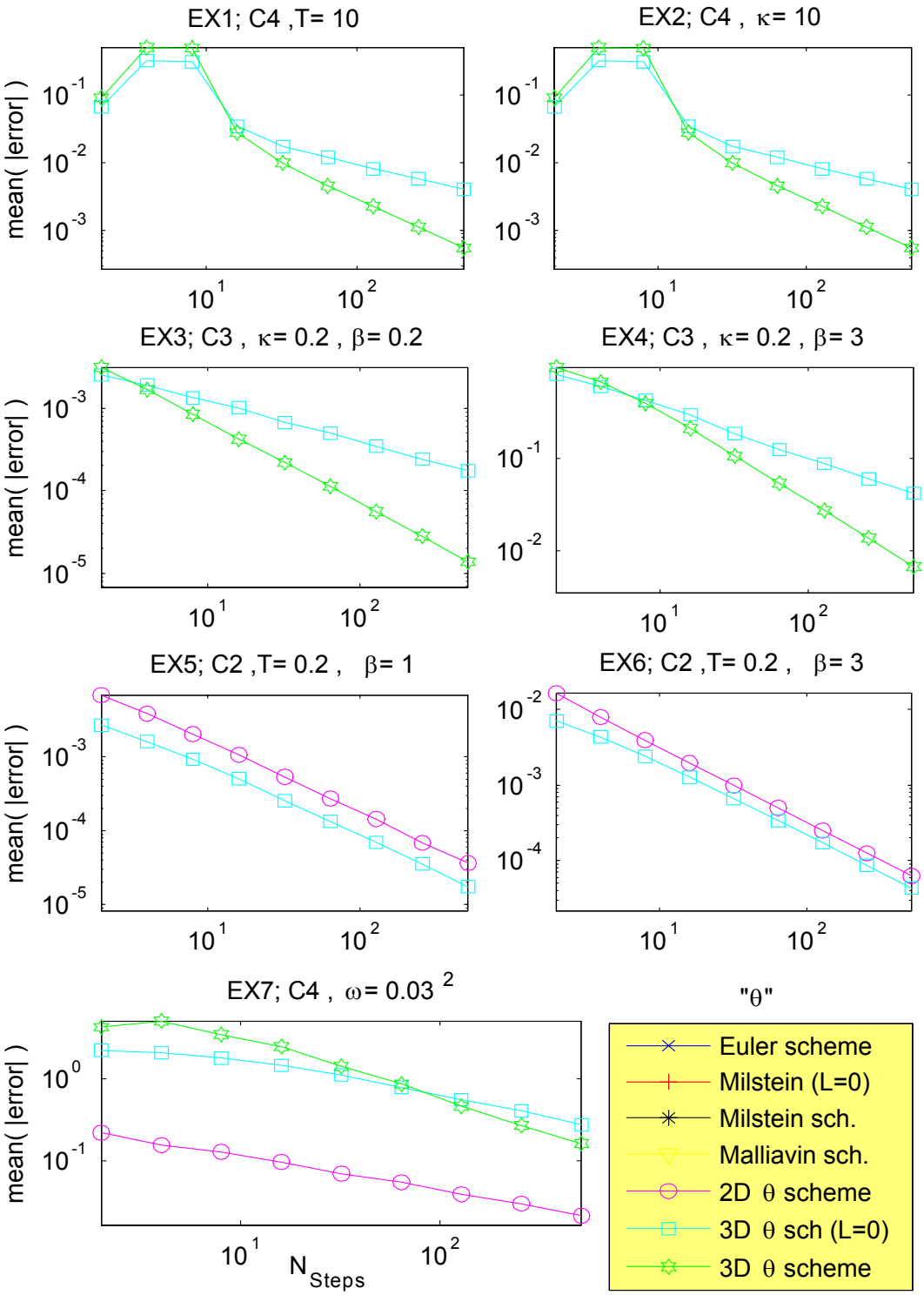

Figure C.3: Strong convergence tests for θ using (5.10).

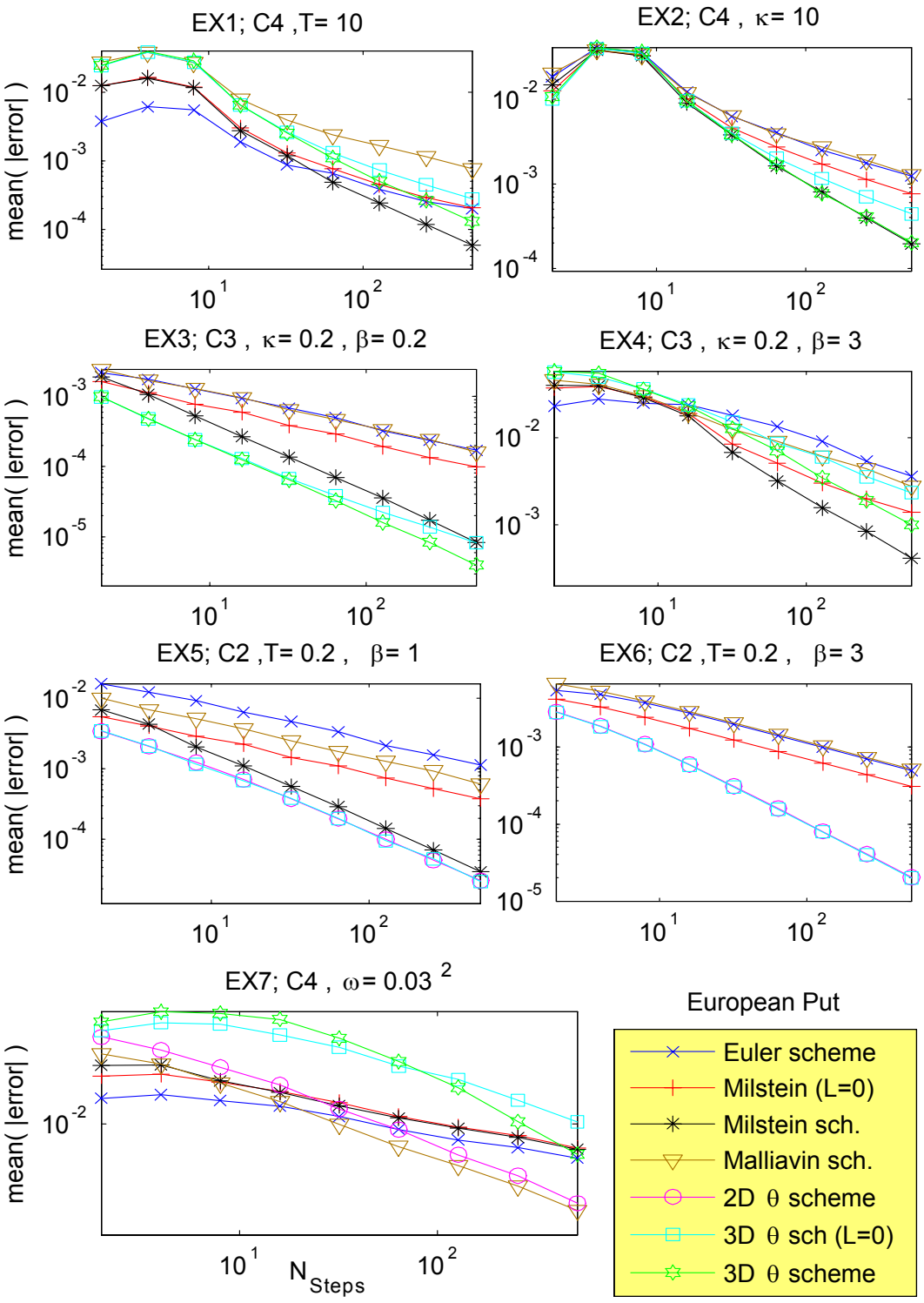

Figure C.4: Strong convergence tests for a European Put option using (5.10).

Bibliography

[1] **Bera, K., and Higgins, L.** (1998): "A Survey of Arch Models: Properties, Estimation and Testing". *Risk Books*, Volatility, June−98, 23-59.

[2] **Björk, T.** (1998): "Arbitrage Theory in Continuous Time". *Oxford University Press Inc., New York.*

[3] **Black, F., and Scholes, M.** (1973): "The pricing of Options and Corporate Liabilities". *Journal of Political Economy,* 81, May, 637 − 659.

[4] **Cox, J., Ingersoll, J., and Ross, S.** (1985): "An Intertemporal General Equilibrium Model of Asset Prices". *Econometrica,* 53, 363 − 384.

[5] **Cox, J., Ingersoll, J., and Ross, S.** (1985): "A Theory of the Term Structure of Interest Rates". *Econometrica*, 53, 385 − 408.

[6] **Cruzeiro, A. B., Malliavin, P., and Thalmaier, A.** (2004): "Geometrization of Monte-Carlo numerical analysis of an elliptic operator: strong approximation". *C. R. Acad. Sci. Paris, Ser. I*, 338, 481 − 486.

[7] **Dupire, B.** (1994): "Pricing with a Smile". *Risk Magazine*, 7, 18-20.

[8] **Gaines, J. G., and Lyons, T. J.** (1994): "Random Generation of Stochastic Area Integrals". *SIAM Journal on Applied Mathematics*, 54, No. 4, 1132 − 1146.

[9] **Ghomrasni, R.** (2004): "On Distributions Associated with the Generalized Lévy's Stochastic Area Formula". *University of Aarhus, Centre for Mathematical Physics and Stochastics* (MaPhySto) [MPS]; (*RR* 2003/4).

[10] **Giles, M.** (2006): "Multi-level Monte Carlo path simulation". Technical Report No. **NA06/03**, Oxford University Computing Laboratory, Parks Road, Oxford, U.K., to appear in Operations Research.

[11] **Giles, M.** (2006): "Improved multilevel Monte Carlo convergence using the Milstein scheme". Technical Report No. **NA06/02**, Oxford University Computing Laboratory, Parks Road, Oxford, U.K.

[12] **Glasserman, P.** (2004): "Monte Carlo Methods in Financial Engineering". *Springer*.

[13] **Grimmett, G., and Stirzaker, D.** (2004): "Probability and Random Processes". *Oxford University Press*, Third Edition.

[14] **Heston, S. L.** (1993): "A Closed-Form Solution for Options with Stochastic Volatility with Applications to Bond and Currency Options". *The Review of Financial Studies*, Volume 6, Issue 2, 327-343.

[15] **Heston, S. L.** (1997): "A Simple New Formula for Options With Stochastic Volatility". Course notes of Washington University in St. Louis, Missouri.

[16] **Higham, D.** (2001): "An Algorithmic Introduction to Numerical Simulation of Stochastic Differential Equations". *SIAM*, Vol. 43, number. 3, pp. $525 - 546$.

[17] **Higham, D.** (2004): "An Introduction to Financial Option Valuation". *Mathematics, Stochastic and Computation*, Cambridge University Press, Cambridge.

[18] **Hobson, D.** (1996): "Stochastic Volatility". Course Notes of School of Mathematical Sciences, University of Bath.

[19] **Hofmann, N., Platen, E., and Schweizer, M.** (1992): "Option pricing under incompleteness and stochastic volatility". *Mathematical Finance*, 2, $153 - 187$.

[20] **Hull, J., and White, A.** (1987): "The Pricing of Options on Assets with Stochastic Volatilities". *The Journal of Finance*, Vol. $XLII$, No 2, June.

[21] **Hull, J.** (1993): "Options, Futures, and other Derivation Securities". *Prentice Hall, Inc.*

[22] **Kloeden, P. E., and Platen, E.** (1999): "Numerical Solution of Stochastic Differential Equations". *Springer*.

[23] **Kloeden, P. E.** (2002): "The Systematic Derivation of Higher Order Numerical Schemes for Stochastic Differential Equations". *Milan Journal of Mathematics*, 70, 187-207.

[24] **Lévy, P.** (1950): "Wiener's Random Function, and other Laplacian Random Functions". *Proceedings of the Second Berkeley Symposium on Mathematical Statistics and Probability*, 171-187.

[25] **Lewis, A. L.** (2000): "Option Valuation under Stochastic Volatility: with Mathematica Code". *Finance Press.*

[26] **Malliavin, P., and Thalmaier, A.** (2005): "Stochastic Calculus of Variations in Mathematical Finance". *Springer.*

[27] **Milstein, G. N.** (1995): "Numerical Integration of Stochastic Differential Equations", *Math. Appl.*, Vol. 313, Kluwer Academic, Dordrecht. Translated and revised from the Russian original.

[28] **Pratt, J. W.** (1964): "Risk Aversion in the Small and in the Large", *Econometrica*, 32, Jan. $122 - 136$.

[29] **Schmitz-Abe, K., and Shaw, W.** (2005): "Measure Order of Convergence without an Exact Solution, Euler vs Milstein Scheme". *International Journal of Pure and Applied Mathematics*, Vol. 24, No 3, $365 - 381$.

[30] **Schmitz-Abe, K., and Giles, M.** (2006): "Pricing Exotic Options using Strong Convergence Properties". *Springer, Progress in Industrial Mathematics at ECMI 2006.*

[31] **Scott, L. O.** (1987): "Option pricing when the variance changes randomly: theory, estimation and an application". *Journal of Financial and Quantitative Analysis*, 22, $419 - 438$.

[32] **Shaw, W.** (2006): "Stochastic Volatility, Models of Heston Type". Course Notes.

[33] **Shaw, W.** (2006): "A note on the Discontinuity Problem in Heston's Stochastic Volatility Model". Course Notes.

[34] **Sin, C.** (1998): "Complications with Stochastic Volatility Models", *Advances in Applied Probability*, 30, $256 - 268$.

[35] **Stein, E. M., and Stein, J. C.** (1991): "Stock price distributions with stochastic volatility: an analytic approach". *Review of Financial Studies*, 4, $727 - 752$.

[36] **Wiggins, J.** (1987): "Options Vales under Stochastic Volatility", *The Journal of Financial Economics*, Vol. 19 No 2, (December), pp. 351 – 372.

[37] **Wiktorsson, M.** (2001): "Joint characteristic function and simultaneous simulation of iterated Itô integrals for multiple independent Brownian motions". *The Annals of Applied Probability*, Vol. 11, No 2, 470 – 487.

[38] **Wilmott, P., Howison, S., and Dewynne, J.** (1995): "The Mathematics of Financial Derivatives". *Cambridge University Press.*

[39] **Wilmott, P.** (1998): "Derivatives: The Theory and Practice of Financial Engineering". *John Wiley and Sons.*

[40] **Xuerong M., Chenggui Y. and Yin G.** (2007): "Approximations of Euler-Maruyama type for stochastic differential equations with Markovian switching, under non-Lipschitz conditions", *Journal of Computational and Applied Mathematics*, Volume 205, Issue 2, 15, pp. 936 – 948.