University of Strathclyde
Department of Mathematics and Statistics

Numerical Approximations of Nonlinear Stochastic Systems

by

Lukasz Szpruch

A thesis presented in fulfilment of the
requirements for the degree of
Doctor of Philosophy

2010

This thesis is the result of the author's original research. It has been composed by the author and has not been previously submitted for examination which has led to the award of a degree.

Signed: Date:

To my parents

Acknowledgements

The postgraduate studies have been a long journey for me, towards my personal and professional development. But I have never been alone in this journey. Therefore, I would like to take this opportunity to thank all who believed in me and accompanied me in this trip.

I would like to thank my supervisors, Professor Xuerong Mao and Professor Des Higham for their support and encouragement during this period of research. Always willing to offer their guidance and useful advices, they showed me the way to become an independent researcher. I would like to thank the whole Department of Mathematics for very friendly and stimulating atmosphere from the first day of my arrival. I am also indebted to the University of Strathclyde for their financial support.

I cannot forget, Dr John Appleby for long discussions on stochastic non-linear problems we had in Chester and Professor Marek Kaluszka who introduced me to theory of probability and stochastic processes and encouraged me to deepen my knowledge into Mathematical Science.

I would like to thank Paulina Tomaszewska for bring back the balance in my life; Mikolaj Roj (known as Miko) for all the help during the last year of my studies; Michal Seweryn for our long conversations about mathematics; Dr Matina (Stamatiki) Rassias for the support given at the beginning of my research journey; Dr Adam Wagner for chats we had in our office.

Very special thanks go to my mum and dad, for many, many years of love, patience, and encouragement. Thank you for believing in me when I didn't believe in myself, supporting any decision I made and for never letting me give up on my dreams.

Abstract

The explicit solution of stochastic differential equations (SDEs) can be found only in a few cases. Therefore, there is a need for accurate numerical approximations that could, for example, enable Monte Carlo Simulations. Convergence and stability of these methods are well understood for SDEs with Lipschitz continuous coefficients. Our research focuses on those situations where the coefficients of the underlying SDEs are non-Lipschitzian. It was demonstrated in the literature, (Hutzenthaler and Jentzen 2009; Higham, Mao, and Yuan 2008) that in this case using the classical methods we may fail to obtain numerically computed paths that are accurate for small step-sizes, or to obtain qualitative information about the behaviour of numerical methods over long time intervals. This work addresses both of these issues, giving a customized analysis of the most widely used numerical methods. Motivated by existing work (Higham, Mao, and Stuart 2003b) and (Hu 1996) we consider implicit schemes. These authors have demonstrated that a backward Euler-Maruyama method strongly converges to the solution of SDEs with one-sided Lipschitz drift and linear growth diffusion coefficients. We extend their work by allowing for a polynomially growing diffusion term. The strong convergence is valuable as it reveals a pathwise error; new efficient Multi-Level Monte Carlo simulations (Giles 2008; Pages 2007) rely on strong convergence and weak convergence. In addition we examine global almost sure asymptotic stability in this nonlinear setting. In particular, we present a stochastic counterpart of the discrete LaSalle principle from which we deduce stability properties of implicit numerical methods.

We also show that an appropriate implicit numerical method preserves positivity. In addition to being a desirable modelling property, in some cases positivity of the numerical approximation is required in order for the scheme to be well defined.

Motivation for our work comes from finance and biology where many widely applied models do not satisfy a Lipschitz condition. We support our theoretical results with relevant examples, such as stochastic interest rate models and stochastic volatility models. Although the considered schemes are implicit we point out that in many practical situations they do not increase computational complexity. We provide numerical results in support of our analysis.

Contents

General Notation

nonnegative : ≥ 0.

a.s. : Almost surely, or with probability 1.

$\emptyset$: The empty set.

$A =: B$: A is defined to be B or A is denoted by B.

$\mathbf{1}_A$: The indicator function of a set A, i.e. $\mathbf{1}_A(x) = 1$ if $x \in A$ or otherwise 0.

A^C : The complement of A in Ω, i.e.$A^C = \Omega - A$.

$A \subset B$: $A \cap B^C = \emptyset$.

$A \subset B$ a.s. : $P(A \cap B^C = \emptyset) = 1$.

$\sigma(C)$: The σ-algebra generated by C.

$a \vee b$: The maximum of a and b.

$a \wedge b$: The minimum of a and b.

$f : A \to B$: The mapping f from A to B.

$\mathbb{R} = \mathbb{R}^1$: The real line.

$\mathbb{R}_+$: The set of all nonnegative real numbers, i.e. $\mathbb{R}_+ = [0, \infty)$.

$\mathcal{B}^d$: The Borel-σ-algebra on $\mathbb{R}^d$.

$|x|$: The Euclidean norm of a vector x and the Frobenius matrix norm.

$C(D; \mathbb{R}^d)$: The family of continuous $\mathbb{R}^d$-valued functions defined on D.

$\langle x, y \rangle$: scalar product of vectors $x, y \in \mathbb{R}^n$.

$C^m(D; \mathbb{R}^d)$: The family of continuously m-times differentiable $\mathbb{R}^d$-valued functions defined on D.

$C^{2,1}(D \times R_+; R)$: The family of all real-valued functions $V(x, t)$ defined on $D \times R_+$

$C^{2,1}(D \times R_+; R)$: which are continuously twice differentiable in $x \in D$ and once differentiable in $t \in R_+$.

$$V_x := \left(\frac{\partial V}{\partial x_1}, \cdots, \frac{\partial V}{\partial x_d} \right).$$

$$V_{xx} := \left(\frac{\partial^2 V}{\partial x_i \partial x_j} \right)_{d \times d}.$$

$\| \xi \|_{L^p} := (E|\xi|^p)^{1/p}$.

$L^p(\Omega; R^n)$: The family of R^n-valued random variables X with $E|X|^p < \infty$.

$L^p_{\mathcal{F}_t}(\Omega; R^n)$: The family of R^n-valued $\mathcal{F}_t$-measurable random variables X with $E|X|^p < \infty$.

$L^p([a, b]; R^n)$: The family of Borel measurable functions $h : [a, b] \to R^n$ such that $\int_a^b |h(t)|^p dt < \infty$.

$\mathcal{L}^p([a, b]; R^n)$: The family of R^n-valued $\mathcal{F}_t$-adapted processes $\{f(t)\}_{a \le t \le b}$ such that $\int_a^b |f(t)|^p dt < \infty$ a.s.

$\mathcal{M}^p([a, b]; R^n)$: The family of R^n-valued $\mathcal{F}_t$-adapted processes $\{f(t)\}_{a \le t \le b}$ in $\mathcal{L}^p([a, b]; R^n)$ such that $E \int_a^b |f(t)|^p dt < \infty$.

Other notations will be explained where they first appear.

List of Figures

Chapter 1

Introduction

> As far as the laws of mathematics refer to reality, they are not certain; and as far as they are certain, they do not refer to reality.
>
> Albert Einstein

In 1827 the botanist Robert Brown noticed that if we look at pollen grains in water through a microscope, they jiggle about. This very lively and irregular state of motion is known nowadays as *Brownian Motion*. However, Brown did not clarify what was causing it. The first of the three papers that Einstein published in 1905 (Einstein 1905) and a work by Marian Smoluchowski (Von Smoluchowski 1906) finally came up with an explanation. A major point of their discoveries was that the motion of water molecules is so complicated that its effect on the pollen grain can best be described probabilistically in terms of exceedingly frequent statistically independent impacts. That was a profound discovery that opened the doors to stochastic differential modeling in science. A mathematically rigorous description of Brownian Motion was given more than twenty years later, in 1923, by Norbert Wiener, (Wiener 1923) and since then sometimes it is called a Wiener process.

Although these discoveries intrigued researchers worldwide, deterministic differential calculus developed by Newton and Leibnitz remained the main tool for modelling a wide range of problems in the natural, social, and biological sciences. Up to the nineteenth century it was commonly thought that if all initial data could only be collected, one would be able to predict the future behaviour of the analyzed system. However, as pointed out by Einstein and Smoluchowki it might happen that the function which we believe describes the change of the system under consideration is not completely known. It might be subject to random environmental effects called nowadays random noise. It was the fact that the Brownian sample paths are (almost surely) nowhere differentiable (Paley, Wiener, and Zygmund 1933), that prevented researchers from including randomness in their models. This also explained why stochastic calculus is far more complex than the deterministic one. Since Brownian Motion is of unbounded variation on any finite time interval, the ordinary Lebesgue-Stieltjes integral cannot be defined. Nevertheless, Brownian Motion has finite quadratic variation and this fact allows one to construct a stochastic integral. The construction is due to the Japanese mathematician Kiyoshi Itô (Itô 1944), and is now known as the Itô stochastic integral. In this thesis we always work with stochastic integrals in the Itô sense. Ordinary differential equations with incorporated noise component are called *Stochastic Differential Equations* (SDEs). Two years later Itô (Itô 1946) proved that once the coefficients of the SDEs are Lipschitz continuous then a system admits a unique solution. Nowadays these equations are essential in modelling various phenomena in mathematical finance (Karatzas and Shreve 1998), physics (Gardiner 1985), molecular biology (Gillespie 1992), epidemiology (Tan and Wai-Yuan 2000), neural networks (Laing and Lord 2009), to mention a few. For these reasons it is extremely important to study the behaviour of the solution to SDEs. Since solutions can be found explicitly only in very few cases, there is a need for development of numerical methods which can give us information about qualitative behaviour of the underlying stochastic systems. The subject of this thesis is to extend current knowledge on approximations for these stochastic systems. We present efficient and accurate approximations for the solutions to a wide family of stochastic processes encountered in mathematical finance and biomathematics, which have not been treated by numerical analysis so far. In the remainder of this chapter we survey the relevant research literature and motivate the research contributions in the thesis.

1.1 Why Are We Interested in Approximations of SDEs?

Let $w(t) = (w_1(t), ..., w_d(t))^T$ be a d-dimensional Brownian motion defined on the probability space, where T denotes the transpose of a vector or a matrix. In this thesis we look at Itô SDEs of the form

$$dx(t) = f(x(t))dt + g(x(t))dw(t). \tag{1.1}$$

Here $x(t) \in \mathbb{R}^n$ for each $t \geq 0$. Thus, $f : \mathbb{R}^n \to \mathbb{R}^n$ and $g : \mathbb{R}^n \to \mathbb{R}^{n\times d}$. For simplicity we assume that initial condition $x_0 \in \mathbb{R}^n$. It can be shown (Mao 2007) that this is not restrictive. In order to approximate SDEs (1.1) numerically, for any step size Δt, we define the partition $\mathcal{P}_{\Delta t} := \{t_k = k\Delta t : k = 0, 1, 2, ...\}$ of the half-line $[0, \infty)$. The most basic and intuitive direct discretization method is the Euler-Maruyama method

$$X_{t_{k+1}} = X_{t_k} + f(X_{t_k})\Delta t + g(X_t)\Delta w_{t_k}, \tag{1.2}$$

where $\Delta w_{t_k} = w(t_{k+1}) - w(t_k)$ are increments of Brownian motion. Maruyama (Maruyama 1955) showed the mean-square convergence of this method, while Gihman and Skorohod (Gihman and Skorohod 1972) proved that the strong order of accuracy of the *Euler-Maruyama* method is 1/2. These results were derived for Lipschitz continuous functions f and g.

Definition 1.1.1. *<u>Global Lipschitz condition</u>. Assume that there exists a positive constant K such that for all $x, y \in \mathbb{R}^n$*

$$|f(x) - f(y)|^2 \vee |g(x) - g(y)| \leq K |x - y|^2 .$$

Here, we recall the theorem from (Kloeden and Platen 1992)

Theorem 1.1.2. *Under **global Lipschitz** condition on functions $\boldsymbol{f}$ and $\boldsymbol{g}$, for any $p \geq 1$ and $T \geq 0$, there exists a positive constant $K = K(p, T)$, independent of Δt such that*

$$\mathbb{E}\Big[\sup_{0\leq t_k\leq T} |x(t_k) - X_{t_k}|^p\Big] \leq K\Delta t^{\boldsymbol{p/2}},$$

where $t_k \in \mathcal{P}_{\Delta t}$.

There are numerous examples in the literature where authors discretize SDEs, typically with an Euler-type or Milstein-type schemes. Within applications, there are three main motivations for such simulations:

- Using a Monte Carlo approach to compute the expected value of a function of $x(t)$, for example to value a bond or the expected payoff of an option, (Andersen, Benzoni, and Lund 2002; Broadie and Kaya 2006; Glasserman 2003);
- Generating time series in order to test parameter estimation algorithms (Duan 2003; Fischer, May, and Walther 2004);
- Approximating the likelihood estimator effectively (Pedersen 1995).

Our research focuses on those situations when the SDEs under consideration are non-linear and non-Lipschitzian. Here standard convergence theory for numerical simulations, as typified by Theorem 9.6.2 (see Theorem 1.1.2 above) in (Kloeden and Platen 1992) or Theorem 1.1 in (Milstein and Tretyakov 2004), cannot be used to deduce that the numerically computed paths are accurate for small step-sizes. Nor can stability analysis, such as in (Higham, Mao, and Stuart 2003a; Higham 2000), be applied to obtain qualitative information about the behaviour of numerical methods over long time intervals. This work addresses both these issues, giving a customized analysis of the most widely used numerical methods. The results obtained justify the type of numerical simulations that are done by researchers and practitioners. We are interested in relaxation of the condition for the diffusion coefficient in order to justify Monte Carlo simulations for highly non-linear systems. Super-linear diffusions

$$|g(x)| \leq \beta \, |x|^{\rho}, \quad \text{for } \rho > 1,$$

arise in financial mathematics, (Ahn and Gao 1999; Campbell, Lo, MacKinlay, and Whitelaw 1998; Ait-Sahalia 1996; Chan, Karolyi, Longstaff, and Sanders 1992; Heston 1997; Lewis 2000) , for example

$$dx(t) = (\mu - \alpha x^{r}(t))dt + \beta x^{\rho}(t)dw(t), \quad r, \rho > 1, \tag{1.3}$$

and stochastic population dynamics (Mao, Marion, and Renshaw 2002; Bahar and Mao 2004; Mao, Sabanis, and Renshaw 2003; Pang, Deng, and Mao 2008; Gard 1988), for example

$$dx(t) = \text{diag}(x_1, x_2, ..., x_n(t))[(b + Ax^2(t))dt + g(x(t))dw(t)]. \tag{1.4}$$

In our research we focus on strong convergence. This form of convergence is valuable as

- weak convergence (Kloeden and Platen 1992) and pathwise convergence (Kloeden and Neuenkirch 2007) follow automatically, and
- efficient Multi-Level Monte Carlo (MLMC) simulations rely on both weak and strong convergence properties (Giles 2008).

Summarizing, direct discretization methods for SDEs (1.1) are important because

- they are widely used in practice;
- even if a transition density is known, it is often computationally faster to simulate with the direct method in cases where the path must be sampled at finely spaced points in order to approximate a path-dependent payoff (Broadie and Kaya 2006; Higham and Mao 2005);
- it is interesting to prove convergence results where there is no global Lipschitz condition for the diffusion term (as mentioned, for example, in (Glasserman 2003));
- it can be regarded as a contribution to the literature on diffusion limits of discrete models, for example ARCH, CEVGARCH(1,1), CEVARCH to mention a few (Nelson 1990; Fornari and Mele 2001).

1.2 Mathematical Finance

In 1900, the mathematician Louis Bachelier in his dissertation “Théorie de la Spéculation” attempted to describe the random nature of stock price fluctuations as a Brownian motion with drift. His intuition was outstanding, since a mathematical definition of a Brownian Motion had not been given by that time. More than sixty years later economist and Nobel prize winner Paul Samuelson, giving full recognition to Bachelier's fundamental contributions, suggested to replace an arithmetic Brownian Motion by a geometric one

$$dS(t) = \alpha S(t)dt + \sigma S(t)dw(t),$$

to account for the fact that stock prices cannot take negative values (Samuelson 1965). Continuing the work of Samuelson in 1973 Black and Scholes (Black and Scholes 1973) and Merton (Merton 1973) derived the price of European call options. This development prompted the massive growth in research into stochastic modelling applied to financial problems, particularly with respect to valuation of contingent claims on underlying assets, so-called financial derivatives. For their contribution Merton and Scholes received the 1997 Nobel Prize in Economics. Black died in 1995, but he was mentioned as a contributor by the Swedish academy. Since the Black-Scholes formula was derived, a number of empirical studies have concluded that the assumption of constant volatility - σ - is inadequate to describe stock returns, based on two findings (1) volatilities of stock returns vary over time, but persist at a certain level (mean-reversion property), these findings can be traced back to the empirical works of (Mandelbrot 1963) and (Fama 1965) with the result that the distributions of stock returns are more leptokurtic than normal; (2) volatilities are correlated with stock returns, and more precisely, they are usually inversely correlated. Furthermore, the volatility smile provides direct evidence for the inconsistent volatility pattern with moneyness in the Black-Scholes model. In order to model the variability of volatility and to capture the volatility smile, several approaches have been suggested. One of the most general, and widely applied approaches is to model volatility by a diffusion process and has been, for example, examined by (Johnson and Shanno 1987), (Wiggins 1987), (Scott 1987), (Hull and White 1987), (Stein and Stein 1991), (Heston 1993), and (Lewis 2000). The models following this approach are the so-called *stochastic volatility models*. Good examples of stochastic volatility models which are treated in this thesis are:

Lewis volatility model (Lewis 2000)

$$\begin{cases} dS(t) = (\mu - \alpha S(t))dt + \sigma_1 V(t) S(t) dw_1(t) \\ dV(t) = (r - \beta V(t)^2)dt + \sigma_2 V(t)^{3/2} dw_2(t); \end{cases}$$

Transformed Heston model (Zhu 2009)

$$\begin{cases} dS(t) = (\mu - \alpha S(t))dt + \sigma_1 V(t)S(t)dw_1(t) \\ dV(t) = (\gamma V(t)^{-1} - \beta V(t))dt + \sigma_2 dw_2(t). \end{cases}$$

Another important class of models are *Stochastic interest models*. For example, a stochastic short rate appears in a risk-neutral stock process as drift. Therefore, volatility of the risk-less rate is a key variable governing the value of contingent claims such as interest rate options. In addition, optimal hedging strategies for risk-averse investors depend critically on the level of term structure volatility. Nowadays stochastic interest rate models form a much larger theoretical field than stochastic volatility models, and even have a longer history because interest rates are the most important factor in economics. Important contributions to the theory of stochastic interest rates models have been made by (Vasicek 1977), (Cox, Ingersoll Jr, and Ross 1985) and (Longstaff 1989). Later Chan, Karolyi and Longstaff (Chan, Karolyi, Longstaff, and Sanders 1992) using the Generalized Method of Moments demonstrated that **highly non-linear models** (with super-linear diffusion coefficient) capture the dynamics of the short-term interest rate better than linear and sub-linear ones. This is because the volatility of the process is highly sensitive to the level of interest rate. This finding was confirmed by Ait-Sahalia (Ait-Sahalia 1996) who investigated several continuous-time interest rate models empirically. He tested parametric models by comparing their implied densities with the density estimated nonparametrically. This study led to a new class of highly non-linear SDEs to model interest rates

$$dr(t) = (\alpha_{-1} r(t)^{-1} - \alpha_0 + \alpha_1 r(t) + \alpha_2 r(t)^2)dt + \sigma r(t)^{\rho} dw(t), \tag{1.5}$$

with $\rho > 1$. Subsequent studies supported the observations made by Ait-Sahalia. Stanton (Stanton 1997), using nonparametric kernel regression, also found significant nonlinearities in spot rate data. Hong and Li (Hong and Li 2005) developed the so called omnibus nonparametric specification test for continuous-time models based on the transition density function, which, unlike the marginal density used by Ait-Sahalia, captures the full dynamics of the continuous process. Their test rejected all but the Ait-Sahalia and CKLS (Chan, Karolyi, Longstaff, and

Sanders 1992) models. It is worthwhile to mention the work of Ahn and Gao (Ahn and Gao 1999) which showed that Inverse Feller Square-Root Process

$$dr(t) = \beta(\mu - r(t))r(t)dt + \sigma r(t)^{3/2}dw(t)$$

outperforms affine models in both time-series as well as cross-sectional tests. Along with Ait-Sahalia, Conley et al. (Conley, Hansen, Luttmer, and Scheinkman 1997) and Gallant et al. (Gallant and Tauchen 2005), have used a variety of empirical techniques to estimate model parameters; and all have suggested that the diffusion term in the SDE grows faster than linearly.

Clearly, studies on stochastic volatility models and stochastic interest rate models have implied that **non-linear** stochastic models have better ability to fit financial data than classical models, therefore they are more realistic. However, such highly nonlinear models are much more difficult to handle for mathematicians. For example, we are not able to find an explicit solution to these stochastic systems and even probability distribution for the solutions can be found only in very few cases. This motivates research on efficient and accurate numerical methods in this **non-linear** setting. These numerical methods could enable efficient Monte Carlo simulations to price various financial instruments as well as shed some light on the complex financial world in order to motivate subsequent studies.

It is also worth mentioning that many **non-linear** stochastic differential financial models may correspond to an econometric counterpart by discretizating them on time points, and some could be referred to as so-called autoregressive random variance models (ARV). Nelson and Foster (Nelson and Foster 1994) and Duan (Duan 1996) showed that some existing stochastic volatility models can be considered as the weak limits of generalized autoregressive conditional heteroscedasticity (GARCH) models. The discrete-time versions of stochastic models also play an important role in empirical tests. For instance, Heston and Nandi (Heston and Nandi 2000) suggested a GARCH option pricing model and derived a closed-form solution which allows for correlation between stock returns and variance and even admits multiple lags in the GARCH process. It is therefore of interest to investigate diffusion limits of such autoregressive processes.

There is a gap in literature on numerical methods for **super-linear stochas-**

tic systems and we believe that this thesis answers some important questions as well as raising new problems in stochastic numerical analysis.

1.3 Overview of Stochastic Numerical Analysis

As we have already mentioned, the first result concerning existence and uniqueness of the solution to the equation (1.1) requires the global Lipschitz condition on both drift and diffusion coefficients (Itô 1946). This result was generalized to the local Lipschitz case by applying the Lyapunov function technique.

Theorem 1.3.1 ((Khasminski 1980)). *Let D be an open subset of $\mathbb{R}^n$. There exists a unique, global solution $x(t) \in D$ to the equation (1.1) on $t \geq 0$ for any given initial value $x(0) = x_0 \in D$ if the following conditions hold:*

i There exists an increasing sequence of bounded domains $\{D_m\}_{m=1}^{\infty}$ with $\bigcup_{m=1}^{\infty} D_m = D$ such that there exists a positive constant $K_m > 0$ for which

$$|f(x) - f(y)| + |g(x) - g(y)| \leq K_m |x - y| \quad \text{for all } x, y \in D_m. \tag{1.6}$$

ii There exists a C^2-function $V : D \to \mathbb{R}_+$ such that

$$LV := V_x f(x) + \frac{1}{2}\text{trace}[g^T(x) V_{xx} g(x)] \leq K(1 + V(x))$$

$$V_m := \inf_{x \in \partial D_m} V(x) \to \infty, \quad as \quad m \to \infty.$$

This powerful theorem allows, for example, a proof of existence of a unique solution on the domains. Very often in applications, it needs to be established that the solution to the SDEs (1.1) stays positive for any $t > 0$. Although, Theorem 1.3.1 allows us to show that for a very wide family of SDEs the solution exists, in general, both the explicit solution and the probability distribution to the solution of (1.1) are not known. We therefore consider computable discrete approximations that, for example, could be used in Monte Carlo simulations.

Recently, some authors reported limitations of classical methods in non-linear settings. It has been shown by Appleby et al. in (Appleby, Kelly, Mao, and Rodkina 2010), that the classical Euler-Maruyama scheme fails to preserve almost sure stability for certain highly non-linear SDEs. More specifically, they showed that Euler-Maruyama explodes to infinity with probability close to one, whereas the corresponding SDE tends to 0 almost surely. Similar example was given by Hutzenthaler. et al. in (Hutzenthaler and Jentzen 2009). Those authors proved that in the case of super-linearly growing coefficients the Euler-Maruyama approximation may not converge in the strong L^p-sense nor in the numerically weak sense to the exact solution.

On the other hand, it has been shown in (Higham, Mao, and Stuart 2003b; Higham, Mao, and Stuart 2003a) that, as in the deterministic case, implicit schemes offer benefits in terms of linear and non-linear stability. What is more, Higham at al. in (Higham, Mao, and Stuart 2003b) using an implicit scheme, presented strong convergence proofs when the drift coefficient is one-sided Lipschitz and the diffusion coefficient is globally Lipschitz. A similar result was derived by Hu in (Hu 1996). These results agree with our intuition from studying numerical approximations for ordinary differential equations where implicit schemes prove to be useful in analysing the so-called stiff problems (Hairer and Wanner 2010). All of these results motivate us to work with implicit schemes. The most basic implicit numerical method is the backward Euler-Maruyama scheme. For the partition $\mathcal{P}_{\Delta t} := \{t_k = k\Delta t : k = 0, 1, 2, ...\}$ of the time interval $[0, \infty)$, we define backward Euler-Maruyama as

$$X_{t_{k+1}} = X_{t_k} + f(X_{t_{k+1}})\Delta t + g(X_{t_k})\Delta w_{t_k}, \tag{1.7}$$

where $X_{t_0} = x_0$. In order to guarantee the existence of a unique global solution for the implicit scheme, we assume that the function f satisfies a one-sided Lipschitz condition

$$\langle x - y, f(x) - f(y) \rangle \leq L\,|x - y|^2 \quad \forall x, y \in \mathbb{R}^n. \tag{1.8}$$

This condition is somehow hard to relax (see discussion in (Jentzen, Kloeden, and Neuenkirch 2009)). But, having financial and bio-mathematical applications in mind it is not very restrictive. These models are often of dissipative nature sometimes called in applied mathematics a *mean-reverting* property.

Once the diffusion coefficient is no longer Lipschitz, the main difficulty is to control its super-linear growth utilizing the dissipative nature of the drift term. Both results of Higham et al. and Hu indicated that once we work only under local Lipschitz conditions, boundedness of moments of the true solution and its approximation is a key property to prove strong convergence. What is more, it was the fact that the Euler-Maruyama scheme does not preserve boundedness of moments of non-linear SDEs which enabled Hutzenthaler et al. (Hutzenthaler and Jentzen 2009) to prove a divergence theorem. The boundedness of moments is also crucial in this thesis. To motivate further the usage of backward schemes we perform a numerical experiment. We consider a non-linear SDE

$$dx(t) = -\alpha x(t)^3 dt + \beta x(t)^2 dw(t). \tag{1.9}$$

the Euler-Maruyama scheme applied to the above equation gives

$$\hat{X}_{t_{k+1}} = \hat{X}_{t_k} - \alpha \hat{X}_{t_k}^3 \Delta t + \beta \hat{X}_{t_k}^2 \Delta w_{t_k}, \tag{1.10}$$

with $\Delta t = \frac{T}{N}$, $N \geq 1$, and has the following property (Hutzenthaler and Jentzen 2009)

$$\lim_{N \to \infty} \mathbb{E} \left| \hat{X}_{t_N} \right|^p = \infty \quad p \geq 1.$$

On the other hand Itô's Lemma (Mao 2007) with function $V(x) = |x|^p$ applied to (1.9) yields

$$\mathbb{E} |x(t)|^p < \infty \quad \text{for} \quad \alpha > \frac{p-1}{2}\beta^2, \quad p \geq 2 \quad t \geq 0.$$

Hence, Hutzenthaler et al. (Hutzenthaler and Jentzen 2009) concluded that for $\alpha > \frac{\beta^2}{2}$

$$\lim_{N \to \infty} \mathbb{E} \left| x(T) - \hat{X}_{t_N} \right|^2 = \infty. \tag{1.11}$$

Therefore, we suggest to approximate (1.9) with

$$X_{t_{k+1}} = X_{t_k} - \alpha X_{t_{k+1}}^3 \Delta t + \beta X_{t_k}^2 \Delta w_{t_k}. \tag{1.12}$$

Let us observe that by solving the appropriate cubic equation we can find $X_{t_{k+1}}$ explicitly in terms of X_{t_k} and Δw_{t_k}.

In Figure 1.1 we compare the behaviour of second moments of the Euler-Maruyama and backward Euler-Maruyama schemes applied to (1.9). We take two different time-steps, $\Delta t = 2^{-8}$ and $\Delta t = 2^{-5}$, and solve (1.10) and (1.12)

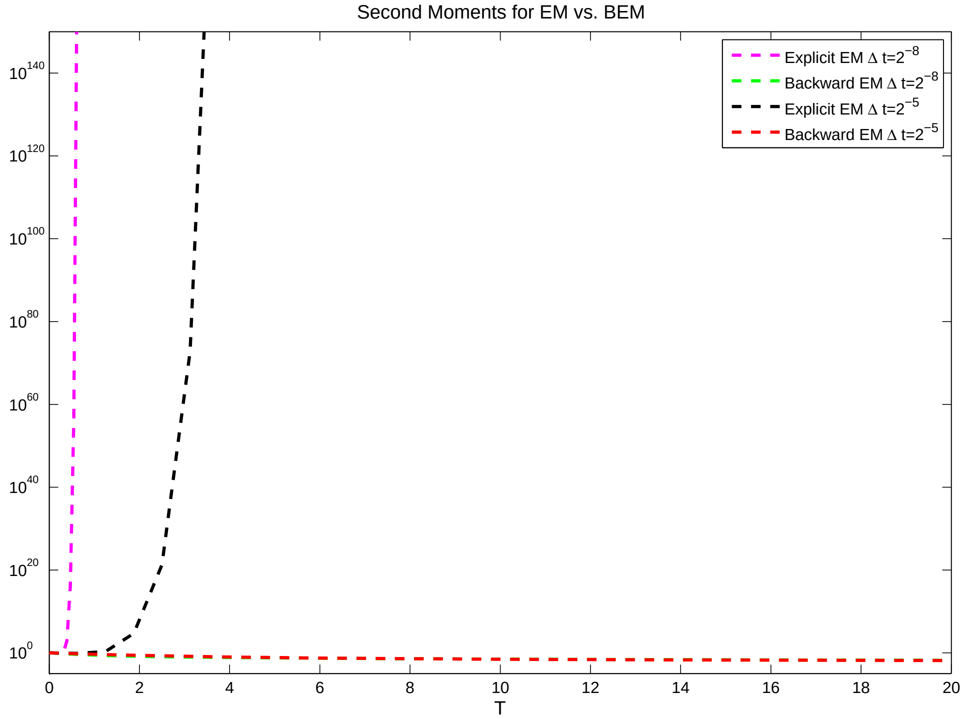

Figure 1.1: Comparison of moments for Explicit and Implicit schemes for highly non-linear SDEs.

respectively. To estimate second moments at time T we average over 10^5 numerically generated paths. We fix the parameters to $\alpha = 6$ and $\beta = \sqrt{6}$ and set the initial condition $x(0) = 1$. It is very interesting to notice that once we decrease the time step second moments of EM scheme explode to infinity even quicker. This is somehow counterintuitive, since it is typical for a smaller step-size to give better accuracy. This shows the importance of an appropriate scheme once we are beyond the global Lipschitz setting.

The above discussion immediately raises the question under what type of conditions we can prove the boundedness of moments for numerical approximations? It is well known that the classical linear growth condition is sufficient to bound the moments for both SDE and Euler-Maruyama (Kloeden and Platen 1992; Mao 2007). Our numerical experiment suggests that EM performs very poorly in super-linear setting. From stochastic analysis we know, that in the case of a continuous solution, a useful first step to relax the linear growth conditions is to apply the Lyapunov function technique, with $V(x) = |x|^2$, (Mao 2007). This leads us to the *monotone condition* (Mao 2007). More precisely, if there exists a constant $K > 0$ such that

$$\langle x, f(x)\rangle + \frac{1}{2}|g(x)|^2 \leq K(1 + |x|^2) \quad \text{for all} \quad x \in \mathbb{R}^n, \tag{1.13}$$

then (1.1) has the following property

$$\sup_{0\leq t\leq T} \mathbb{E}|x(t)|^2 < \infty \quad \forall T > 0.$$

However, to the best of our knowledge, there is no result of this type for numerical approximation of SDEs. Therefore, our goal is to close this gap and to prove strong convergence under the monotone condition (1.13) for BEM. The monotone condition allows us to develop bounds for polynomial coefficients, for example for $f = -x^3$ and $g = x^2$, in (1.9).
Although we are able to prove (3.4.4) strong convergence of implicit Euler-Maruyama scheme to the exact solution under only the general monotone condition (1.13), we have found this condition to be too weak to derive a rate of convergence. This agrees with the work of Higham et al. (Higham, Mao, and Stuart 2003b), where the assumption on boundedness of moments did not lead to a rate of convergence either. By imposing a further, polynomial-like condition on the drift, optimal strong convergence rates were established by Higham et al. in (Higham, Mao, and Stuart 2003b) for backward Euler-Maruyama and split-step backward schemes. By optimal we mean that the same order arises for SDEs under global Lipschitz conditions on f and g (Kloeden and Platen 1992; Müller-Gronbach and Ritter 2008). We will extend their work by changing the linear growth condition on diffusion coefficients into the polynomial type condition. These additional assumptions allowed us to derive the optimal rate of convergence of backward Euler-Maryuama scheme in a non-linear setting. We also would like to comment that our results can be extended to higher order schemes. We will demonstrate that the same assumptions which were required to prove the strong convergence theorem with the optimal rate for BEM, are sufficient to prove the fundamental theorem of Milstein (Milstein 1987; Milstein and

Tretyakov 2004).

So far, to the best of our knowledge, existing results about strong convergence for numerical schemes cover only non-linear SDEs where the diffusion part is of the form σx^ρ with $\rho \in [0.5, 1)$ (Berkaoui, Bossy, and Diop 2007; Mao, Yuan, and Yin 2007; Higham and Mao 2005). Furthermore, the rate of convergence was derived only in (Berkaoui, Bossy, and Diop 2007) under very restrictive assumptions on the parameters. In this thesis we do not treat SDEs with the diffusion term satisfying Hölder continuous condition with $\alpha \in [0.5, 1)$.

In our research we also considered numerical issues arising from SDE models of Ait-Sahalia type. The SDE that we study, which we refer to as the *generalized Ait-Sahalia model*, has the form

$$dx(t) = (\alpha_{-1}x(t)^{-1} - \alpha_0 + \alpha_1 x(t) - \alpha_2 x(t)^r)dt + \sigma x(t)^\rho dw(t), \tag{1.14}$$

where $\alpha_{-1}, \alpha_0, \alpha, \alpha_1, \alpha_2, \sigma$ are positive constants and $r, \rho > 1$. In addition to super-linear diffusion, a further difficulty in (1.14) is that the drift contains a term $\alpha_{-1}x(t)^{-1}$ that does blow up at the origin. It was also indicated in (Zhu 2009) that once we consider the transformed Heston Model

$$\begin{cases} dS(t) = (\mu - \alpha S(t))dt + \sigma_1 V(t)S(t)dw_1(t) \\ dV(t) = (\gamma V(t)^{-1} - \beta V(t))dt + \sigma_2 dw_2(t), \end{cases}$$

classical Euler schemes cannot cope with the term $x(t)^{-1}$. The backward Euler-Maryuma overcomes this difficulty (we will prove its strong convergence) because it preserves positivity of the solution in this case.

Preservation of positivity of the solution to equation (1.1) by numerical approximations is an important issue. It may be required for modelling and for the scheme to be well defined. For example, evaluating the drift coefficient in the 3/2 Heston Volatility model for a negative argument does not make sense. Many fixes have been proposed in literature, but these can lead to substantial bias in simulations, (Lord, Koekkoek, and Van Dijk 2009). For more information about positivity preserving schemes we refer the reader to (Szpruch, Mao, Higham, and Pan 2010; Schurz 2005; Kahl, Gunther, and Rosberg 2008; Appleby, Guzowska, and Rodkina 2010).

After having established a strong convergence result we proceed to stability analysis for nonlinear SDEs (1.1) under the monotone condition. The main problem concerns propagation of errors during the simulation of an approximate path. If the numerical scheme is not stable, then the simulated path may diverge substantially from the exact solution in practical simulations. Similarly, the expectation of the functional estimated by a Monte Carlo simulation may be significantly different from that of the expected functional of the underlying SDE due to numerical instabilities. Our aim here is to investigate almost sure asymptotic properties of numerical schemes for SDE (1.1) via a stochastic version of the LaSalle principle. In (LaSalle 1968) LaSalle improved significantly the Lyapunov stability method for Ordinary Differential Equations. Namely, he developed methods for locating limit sets of nonautonomous systems (Hale and Lunel 1993; LaSalle 1968). The first stochastic counterpart of his great achievement was established by Mao (Mao 1999) under local Lipschitz and linear growth conditions. Recently, this result was generalized by Shen et al. in (Shen, Luo, and Mao 2006) to cover stochastic functional differential equations with local Lipschitz coefficients. Furthermore, it is well known that there exist counterparts of invariant principles for discrete dynamical systems (LaSalle and Artstein 1976). However, there seems to be no discrete counterpart of Mao's version of the LaSalle theorem. In this thesis we investigate a special case of this result with Lyapunov function $V(x) = |x|^2$. We shall show that almost sure global stability can be easily deduced from our results. Our primary objectives in stability analysis are

- Ability to cover highly nonlinear cases;
- Mild assumption on the time step - $A(\alpha)$-stability concept (Higham 2000).

Results which investigate stability analysis for numerical methods can be found in Higham (Higham 2001; Higham 2000) in the scalar linear case, Baker et al. (Baker and Buckwar 2005) for global Lipschitz and Higham et al. (Higham, Mao, and Stuart 2003a) for one-sided Lipschitz drift and linear growth diffusion coefficients.

It is also interesting to investigate how higher order approximations perform once coefficients are not globally Lipschitz. In financial applications the Milstein scheme is usually the method of choice. Recently Giles (Giles 2008; Giles 2006) demonstrated the superiority of MLMC with Milstein in the scalar case to price

Asian, lookback, barrier and digital options. To justify the use of MLMC we need to verify first that the base method converges to the solution of (1.1) in strong sense. In order to achieve a root-mean-square error $O(\varepsilon)$ using a simple Monte Carlo method with a numerical approximation with first order weak convergence, would require computational complexity $O(\varepsilon^{-3})$. If, in addition, we know that the Milstein scheme strongly converges at the optimal rate we can reduce computational complexity to $O(\varepsilon^{-2})$ using MLMC (Giles 2006). Giles's approach is very efficient and pricing options using MLMC with Milstein offers big advantages over the classical approach. This provides an excellent motivation for our work. Typically, in order to prove convergence of the Milstein scheme, stricter assumptions than those for EM are required (Kloeden and Platen 1992). What is more, it was demonstrated by Higham (Higham 2000) that the Milstein scheme applied to a linear scalar SDE has much worse stability properties than Euler-Maruyama, even once we allow for implicitness in the drift. In order to address the issues mentioned above, we will introduce a new double implicit Milstein scheme (Szpruch 2010). We will prove that the scheme has remarkable approximation properties for a rich family of stochastic processes encountered in mathematical finance, because:

- it preserves positivity of the solution;
- the approximation has very good stability properties as opposed to classical Milstein scheme (Higham 2000); the stability properties of double-implicit scheme are as good as backward Euler, that is the double implicit scheme recovers the entire mean-square stability region of its test SDE without severe restrictions on the time step.

So far, convergence of the Milstein scheme, to the best of our knowledge, was analyzed under a global Lipschitz condition only, as in (Kloeden and Platen 1992). By allowing additional implicitness we are able to significantly relax the conditions required for strong convergence and therefore cover many important stochastic differential financial models encountered in the literature, such as the 3/2 Heston volatility model

$$dx(t) = x(t)(\mu - \alpha x(t))dt + \beta x^{3/2} dw(t).$$

An appealing feature in this case is that the solution to the scheme can be found explicitly and, therefore, implicitness does not increase computational complexity.

In order to prove our results new techniques have been developed. We believe that these techniques can be adapted by other researchers to deepen our knowledge on stochastic numerical integration and can be used in the general theory of discrete stochastic processes. The most important techniques are:

- We have utilized a stopping time technique for discrete stochastic processes. It is well known; (Buchmann 2005; Broadie, Glasserman, and Kou 1997; Mannella 1999), that in discrete time approximations for a stochastic process, the problem of overshooting the boundary appears.

- We have introduced a new numerical method, which we have called the Forward-Backward Euler-Maruyama (FBEM). The FBEM scheme enables us to overcome some measurability difficulties and avoids using Malliavin calculus.

1.4 Outline of the Thesis

Our intention is to keep this work relatively self-contained. With this in mind in Chapter 1 we recall some of the fundamental results from the theory of stochastic processes and Itô stochastic calculus. We also define the implicit numerical approximations we use in this thesis. Further we introduce the important and distinct notions of convergence and stability for stochastic processes.

The main body of research is contained in Chapters 3 through 6. In Chapter 3 we present a proof of the strong convergence for an implicit Euler-Maruyama scheme under a general monotone condition. We also consider stability of implicit methods in this non-linear setting via a new discrete stochastic LaSalle principle. This chapter reveals the methodology we have developed to deal with non-linearities. First, utilizing the stopping time technique we prove the boundedness of moments for the numerical method. Then, we introduce a new forward-backward scheme that allows us to employ continuous time stochastic analysis. In Chapter 4 we extend the analysis from Chapter 3. We impose stronger assumptions on the coefficients of underlying SDEs. We introduce a dissipative-type condition on the drift and a polynomial condition on the diffusion coefficients. This enables us to prove boundedness for higher moments of backward Euler-Maruyama and a stronger convergence theorem than in chapter 3. Further by introducing strong monotone and strong polynomial conditions we reveal a rate of convergence for Backward Euler Maruyama which agrees with classical results for EM in global Lipschitz case. We conclude the chapter with a proof of the Fundamental Theorem of Milstein (Milstein 1987) under the assumptions required for our convergence theorem for BEM.

Chapter 5 considers non-linear stochastic differential financial models. We begin with a general SDE which is used in many stochastic volatility and interest rate models. We demonstrate that the assumptions we introduce in Chapter 4 are satisfied in this case and therefore we can conclude that we can successfully approximate the model with backward Euler-Maruyama. The second model we consider in this chapter is an Ait-Sahalia interest rate model. Not only has this model a super-linear growth in the drift and diffusion, but also an additional term blows up at the origin. Nevertheless BEM converges to its solution in the strong sense. What is more we prove that BEM preserves positivity of the solution in this special case. We confirm our theoretical results with appropriate simulations.

The last of our research chapters, Chapter 6, considers strong convergence and stability of a numerical scheme with first order of accuracy. In this initial investigation we restrict ourselves to the scalar case. We introduce a new double implicit Milstein scheme and prove that it has some very desirable properties. Particularly, it preserves positivity for a wide family of SDEs. Adopting methodology from Chapter 3, we show that the additional implicitness in the second order term of the approximation enables us to prove a strong convergence theorem and that the scheme has excellent stability properties.

In the final chapter, we summarize our findings and suggest some possible improvements in further research.

Chapter 2

Mathematical Background

> It is remarkable that a science which began with the consideration of games of chance should have become the most important object of human knowledge.
>
> Théorie Analytique des Probabilités,
> Pierre Simon Laplace

Before presenting our results, in this chapter we recall some fundamental results from the theory of probability and the theory of stochastic processes. We focus only on those facts which are used extensively throughout our study. Although these results are well known we feel the chapter is useful for the clarity of the exposition.

We note here that the proofs of the theorems and lemmas that appear in this section do not comply with the scope of this thesis and therefore are omitted. There are many excellent books to which the reader may refer, for instance:

- (Feller 1968), (Billingsley 1979), (Kallenberg 2002) (Williams 1991) for probability theory and the theory of stochastic processes,
- (Oksendal 1998), (Mao 2007), (Protter 2004), (Karatzas and Shreve 1991), for the theory of stochastic differential equations.

2.1 Stochastic Processes

A *stochastic process* is a mathematical model for the occurrence, at each moment after the initial time, of a random phenomenon. Thus, a stochastic process is a collection of random variables $X = \{X_t; 0 \leq t < \infty\}$ on $(\Omega, \mathcal{F})$, which takes values in a second measurable space $(S, \mathcal{S})$, called the state space. The index $t \in [0, \infty)$ of random variable X_t admits a convenient interpretation as *time*. For a fixed sample point $\omega \in \Omega$, the function $X_t(\omega)$, $t \geq 0$, is a *sample path* (realization, trajectory) of the process X associated with ω. Approximations of these sample paths are a main subject of this thesis. On the other hand, for fixed $t \in [0, \infty)$ the function $X_t(\omega)$, $\omega \in \Omega$, is a random variable.

Definition 2.1.1. *The stochastic process X is called* measurable *if, for every $A \in \mathcal{B}(\mathbb{R}^d)$, the set $\{(t, \omega); X_t(\omega) \in A\}$ belongs to the product σ-field $\mathcal{B}([0, \infty]) \otimes \mathcal{F}$; in other words, if the mapping*

$$(t, \omega) \mapsto X_t(\omega) : ([0, \infty] \times \Omega, \mathcal{B}([0, \infty)) \otimes \mathcal{F}) \mapsto (\mathbb{R}^d, \mathcal{B}(\mathbb{R}^d))$$

is measurable.

The temporal feature of a stochastic process suggests a flow of time, in which, at every moment $t \geq 0$, we can talk about the *past*, *present* and *future* and can ask how much an observer of the process knows about it at the present, as compared to how much he or she knew at some point in the past or will know at some point in the future. In order to keep track of this information we equip our sample space $(\Omega, \mathcal{F})$ with a *filtration*, i.e., a nondecreasing family $\{\mathcal{F}\}_{t \geq 0}$ of sub-σ -fields of $\mathcal{F} : \mathcal{F}_s \subseteq \mathcal{F}_t \subseteq \mathcal{F}$ for $0 \leq s < t < \infty$. The filtration is said to be *right continuous* if $\mathcal{F}_t = \bigcap_{s>t} \mathcal{F}_s$ for all $t \geq 0$. When the probability space is complete, the filtration is said to satisfy the *usual conditions* if it is right continuous and $\mathcal{F}_0$ contains all $\mathbb{P}$ -null sets.

From now on, unless otherwise specified, we shall always work on a given complete probability space $(\Omega, \mathcal{F}, \mathbb{P})$ with a filtration $\{\mathcal{F}\}_{t \geq 0}$ satisfying the usual conditions. We also define $\mathcal{F}_\infty = \sigma(\bigcup_{t \geq 0 \mathcal{F}_t})$.

2.1.1 Stopping Times

Let us keep in mind the interpretation of the parameter t as a time. Let us imagine that we are interested in the occurrence of a certain phenomenon. We are thus forced to pay particular attention to the instant $\tau(\omega)$ at which the phenomenon manifests itself for the *first time*.

Definition 2.1.2. *A random variable* $\tau : \Omega \to [0, \infty]$ *is called an* $\{\mathcal{F}\}_t$ -stopping time *if* $\{\omega : \tau(\omega) \leq t\} \in \mathcal{F}_t$ *for any* $t \geq 0$.

The theory of stopping times is essential for our research. The following two theorems are useful.

Theorem 2.1.3. *If* $\{X_t\}_{t\geq 0}$ *is a progressively measurable process and* τ *is a stopping time, then* $X_\tau \mathbf{1}_{\tau<\infty}$ *is* $\mathcal{F}_\tau$*-measurable.*

Theorem 2.1.4. *Let* $\{X_t\}_{t\geq 0}$ *be an* $\mathbb{R}^d$*-valued cádlág* $\{\mathcal{F}_t\}$ *-adapted process and* D *an open subset of* $\mathbb{R}^d$*. Define*

$$\tau = \inf\{t \geq 0 : X_t \notin D\},$$

where we use convention $\inf\{\emptyset\} = \infty$*. Then* τ *is an* $\{\mathcal{F}_t\}$*- stopping time.*

Lemma 2.1.5 (Fatou)**.** *For any non-negative measurable functions* $\{X_k\}_{k\geq 1}$ *on* $(\Omega, \mathcal{F}, \mathbb{P})$*, we have*

$$\mathbb{E}[\liminf_{k\to\infty} X_k] \leq \liminf_{k\to\infty} \mathbb{E}[X_k].$$

2.1.2 Conditional Expectation

Conditional expectations play a central role in the modern theory of probability. It gives a foundation for the martingales which we introduce below. The concept of conditional expectations was formally introduced by Kolmogorov and it leads to the measure-theoretic definition of conditional probability.

Let $X \in L^1(\Omega; \mathbb{R})$. Let $\mathcal{G} \subset \mathcal{F}$ be a sub -σ -algebra of $\mathcal{F}$. In general, X is not $\mathcal{G}$- measurable. We now seek an integrable $\mathcal{G}$ -measurable random variable Y such that it has the same values as X on the average in the sense that

$$\mathbb{E}(\mathbf{1}_G Y) = E(\mathbf{1}_G X) \quad i.e. \quad \int_G Y(\omega)dP(\omega) = \int_G X(\omega)dP(\omega) \quad \text{for all} \quad G \in \mathcal{G}.$$

By the Radon-Nikodym theorem, there exists unique Y, up to sets of measure 0. It is called the *conditional expectation* of X under the condition $\mathcal{G}$, and we write

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

2.1.3 Martingales

The concept of a martingale has its origins in betting strategies and was popular in 18th century France. Martingales were introduced in probability theory by Paul Pierre Lévy, and the early development of the theory was led by Joseph Leo Doob.

Definition 2.1.6. *An $\mathbb{R}^d$ - valued $\{\mathcal{F}_t\}$ -adapted integrable process $\{M_t\}_{t\geq 0}$ is called a* martingale *with respect to $\{\mathcal{F}_t\}$ if*

$$\mathbb{E}(M_t \mid \mathcal{F}_s) = M_s \quad \text{a.s} \quad \textit{for all} \quad 0 \leq s < t < \infty.$$

Theorem 2.1.7. *Let $\{M_t\}_{t\leq 0}$ be an $\mathbb{R}^d$ -valued martingale with respect to $\{\mathcal{F}_t\}$, and let θ, ρ be two finite stopping times. Then*

$$\mathbb{E}(M_\theta \mid \mathcal{F}_\rho) = M_{\theta\wedge\rho} \quad a.s.$$

Definition 2.1.8. *An $\mathbb{R}^d$- valued $\{\mathcal{F}_t\}$ -adapted integrable process $\{M_t\}_{t\geq 0}$ is called a* local martingale *if there exists a nondecreasing sequence $\{\tau_k\}_{k\geq 1}$ of stopping times with $\tau_k \uparrow \infty$ a.s such that $\{M_{\tau_k\wedge t} - M_0\}_{t\geq 0}$ is martingale.*

2.2 Stochastic Calculus

Before we introduce stochastic differential equations let us recall some basic properties of the Itô integral

$$\int_0^t f(s)dw(s)$$

with respect to an m -dimensional Brownian motion $\{w_t\}$ for a class of $d \times m$ -matrix -valued stochastic processes $\{f(t)\}$.

Definition 2.2.1. *Let $0 \leq a < b < \infty$. Denote by $\mathcal{M}^2([a,b];\mathbb{R})$ the space of all* *real -valued measurable* $\{\mathcal{F}_t\}$ *-adapted processes* $f = \{f(t)\}_a$ *such that*

$$\| f \|_{a,b}^2 = \mathbb{E} \int_a^b |f(t)|^2 \, dt < \infty.$$

Theorem 2.2.2. *Let* $f \in \mathcal{M}^2([a,b]; \mathbb{R}^{d \times m})$*, and let* ρ*,* τ *be two stopping times such that* $0 \leq \rho \leq \tau \leq T$*. Then*

$$\mathbb{E}(\int_\rho^\tau f(t) dw(t) \mid \mathcal{F}_\rho) = 0,$$

$$\mathbb{E}(\left| \int_\rho^\tau f(t) dw(t) \right|^2 \mid \mathcal{F}_\rho) = \mathbb{E} \left(\int_\rho^\tau |f(t)|^2 \, dt \mid \mathcal{F}_a \right).$$

Definition 2.2.3. *A d -dimensional Itô process is an* $\mathbb{R}^d$*-valued continuous adapted process* $x(t) = (x_1(t), \ldots, x_d(t))^T$ *on* $t \geq 0$ *of the form*

$$x(t) = x(0) + \int_0^t f(s) ds + \int_0^t g(s) dw(s),$$

where $f = (f_1, \ldots, f_d)^T \in \mathcal{L}^1(\mathbb{R}_+; \mathbb{R}^d)$ *and* $(g = g_{ij})_{d \times m} \in \mathcal{L}^2(\mathbb{R}_+; \mathbb{R}^{d \times m})$*. We shall say that* $x(t)$ *has stochastic differential dx(t) on* $t \geq 0$ *given by*

$$dx(t) = f(t)dt + g(t)dw(t).$$

Theorem 2.2.4. *Let* $x(t)$ *be a d-dimensional Itô process on* $t \geq 0$ *with a stochastic differential*

$$dx(t) = f(t)dt + g(t)dw(t),$$

where $f \in \mathcal{L}^1(\mathbb{R}_+; \mathbb{R}^d)$ *and* $g \in \mathcal{L}^2(\mathbb{R}_+; \mathbb{R}^{d \times m})$*. Let* $V \in C^{2,1}(\mathbb{R}^d \times \mathbb{R}_+; \mathbb{R})$*. Then* $V(x(t), t)$ *is again an Itô process with stochastic differential given by*

$$\begin{aligned} dV(x(t), t) = {} & [V_t(x(t), t) + V_x(x(t), t) f(t) \\ & + \frac{1}{2} \text{trace}(g^T(t) V_{xx}(x(t), t) g(t))] dt + V_x(x(t), t) g(t) dw(t) \quad a.s, \end{aligned}$$

and we define a diffusion generator L as

$$LV(x(t), t) = V_t(x(t), t) + V_x(x(t), t) f(t) + \frac{1}{2} \text{trace}(g^T(t) V_{xx}(x(t), t) g(t)).$$

2.3 Useful Inequalities

In this section we recall some basic inequalities which are extensively used through our study.

Hölder's inequality

$$\left|\mathbb{E}(X^T Y)\right| \leq (\mathbb{E}\,|X|^p)^{1/p}(\mathbb{E}\,|Y|^q)^{1/q},$$

if $p > 1$, $1/p + 1/q = 1$, $X \in L^p$ and $Y \in L^q$. In case $p = q = 2$ Hölder's inequality is often called Cauchy-Schwarz inequality.

Minkowski's inequality

$$(\mathbb{E}\,|X+Y|^p)^{1/p} \leq (\mathbb{E}\,|X|^p)^{1/p} + (\mathbb{E}\,|Y|^p)^{1/p},$$

if $p \geq 1$, $X, Y \in L^p$.

Young's inequality

$$|a|\,|b| \leq \frac{\varepsilon}{r}\,|a|^r + \frac{1}{q\varepsilon^{q/r}}\,|b|^q\,, \quad \text{where} \quad a, b \in \mathbb{R}^d \quad \text{and} \quad \varepsilon > 0,$$

with $r^{-1} + q^{-1} = 1$, $r, q > 1$.

Theorem 2.3.1 (Burkholder-Davis-Gundy inequality). *Let $g \in \mathcal{L}^2(\mathbb{R}_+; \mathbb{R}^{d\times m})$. Define, for $t \geq 0$,*

$$x(t) = \int_0^t g(s)dw(s) \quad \text{and} \quad A(t) = \int_0^t |g(s)|^2\,ds.$$

Then for every $p > 0$, there exist universal positive constants c_p, C_p, such that

$$c_p\mathbb{E}\,|A(t)|^{\frac{p}{2}} \leq \mathbb{E}\left(\sup_{0\leq s\leq t} |x(t)|^p\right) \leq C_p\mathbb{E}\,|A(t)|^{\frac{p}{2}}$$

for all $t \geq 0$. In particular, one may take

$$\begin{array}{lll} c_p = (p/2)^p, & C_p = (32/p)^{p/2} & \text{if } 0 < p < 2; \\ c_p = 1, & C_p = 4 & \text{if } p = 2; \\ c_p = (2p)^{-p/2}, & C_p = [p^{p+1}/2(p-1)^{p/2}] & \text{if } p > 2. \end{array}$$

Theorem 2.3.2 (Gronwall's inequality)**.** *Let* $T > 0$ *and* $c \geq 0$*. Let* $u(\cdot)$ *be a Borel measurable bounded nonnegative function on* $[0, T]$*, and let* υ *be a nonnegative integrable function on* $[0, T]$*. If*

$$u(t) \leq c + \int_0^t \upsilon(s)u(s)ds \quad \textit{for all } 0 \leq t \leq T,$$

then

$$u(t) \leq c \exp\left(\int_0^t \upsilon(s)ds\right) \quad \textit{for all } 0 \leq t \leq T.$$

Lemma 2.3.3 (Discrete Gronwall's inequality)**.** *Let M be a positive integer. Let* u_k *and* v_k *be non-negative numbers for k=0,1,...,M. If*

$$u_k \leq u_0 + \sum_{j=0}^{k-1} v_j u_j, \qquad \forall k = 1, 2, ..., M,$$

then

$$u_k \leq u_0 \exp\left(\sum_{j=0}^{k-1} v_j\right), \qquad \forall k = 1, 2, ..., M.$$

2.4 Numerical Methods

We study the numerical approximation of the stochastic differential equation

$$dx(t) = f(x(t))dt + g(x(t))dw(t). \tag{2.1}$$

In this section we formalize the definitions of the strong convergence and the stability of the numerical methods. We also demonstrate how implicit schemes can be derived from the Wagner-Platen (Itô-Taylor) expansion in a very natural way.

2.4.1 Implicit Schemes

We recall the definition of the backward Euler-Maruyama scheme.

Definition 2.4.1. *Given any step size* Δt*, we define the partition* $\mathcal{P}_{\Delta t} := \{t_k = k\Delta t : k = 0, 1, 2, ...\}$ *of the time interval* $[0, \infty)$*. The backward Euler-Maruyama* *scheme has the following form:*

$$X_{t_{k+1}} = X_{t_k} + f(X_{t_{k+1}})\Delta t + g(X_{t_k})\Delta w_{t_k}, \tag{2.2}$$

where $\Delta w_{t_k} = w_{t_{k+1}} - w_{t_k}$ *and* $X_{t_0} = x_0$.

A method of first order strong accuracy was first introduced by Milstein in (Milshtein 1975). For simplicity we present the Milstein scheme in the scalar case

$$X_{t_{k+1}} = X_{t_k} + f(X_{t_k})\Delta t + g(X_{t_k})\Delta w_{t_k} \tag{2.3}$$

$$+ \frac{1}{2} L^1 g(X_{t_k}) \int_{t_k}^{t_{k+1}} (w(s) - w(t_k))\, dw(s). \tag{2.4}$$

Originally the Milstein scheme was derived using the theory of Markov operator semigroups. But later W. Wagner and E. Platen (Wagner and Platen 1978), using the Itô Lemma only, showed how higher order schemes can be obtained in a very natural way. Their work is consistent with the deterministic numerical approximation theory, where the Taylor expansion is the main tool. Here we show how using the Itô-Taylor formula one can derive the implicit schemes considered in this thesis. Let us look at the scalar Itô type SDE

$$x(t) = x(0) + \int_0^t f(x(s))ds + \int_0^t g(x(s))dw(s). \tag{2.5}$$

The Itô formula applied to (2.5), with $s > t$, gives

$$f(x(t)) = f(x(s)) - \int_t^s L^0 f(x(z))dz - \int_t^s L^1 f(x(z))dw(z),$$

where we have introduced operators

$$L^0 = f\frac{\partial}{\partial x} + \frac{1}{2} g^2 \frac{\partial^2}{\partial x^2},$$

$$L^1 = g\frac{\partial}{\partial x}.$$

This is the so-called semi-implicit Itô-Taylor expansion. By semi-implicit we mean that only the deterministic terms can be made implicit. We have

$$x(t+\Delta t) = x(t) + \int_t^{t+\Delta t} f(x(s))ds + \int_t^{t+\Delta t} g(x(s))dw(s) \tag{2.6}$$

$$\begin{aligned} &= x(t) + \int_t^{t+\Delta t} \Bigg(f(x(t+\Delta t)) \\ &- \int_s^{t+\Delta t} L^0 f(x(z))dz - \int_s^{t+\Delta t} L^1 f(x(z))dw(z) \Bigg) ds \\ &+ \int_t^{t+\Delta t} \left(g(x(t)) + \int_t^s L^0 g(x(z))dz + \int_t^s L^1 g(x(z))dW(z) \right) dw(s). \end{aligned} \tag{2.7}$$

Truncating the reminder term we obtain the implicit Euler-Maruyama scheme

$$X_{t_{k+1}} = X_{t_k} + f(X_{t_{k+1}})\Delta t + g(X_{t_k})\Delta w_{t_k}.$$

Now applying the Itô formula to

$$L^1 g(x(z)) = L^1 g(x(t)) + \int_t^z L^0 L^1 g(x(h))dh + \int_t^z L^1 L^1 g(x(h))dw(h),$$

yields

$$\begin{aligned} x(t+\Delta t) = x(t) + \int_t^{t+\Delta t} &\Bigg(f(x(t+\Delta t)) \\ &- \int_s^{t+\Delta t} L^0 f(x(z))dz - \int_s^{t+\Delta t} L^1 f(x(z))dw(z) \Bigg) ds \\ &+ \int_t^{t+\Delta t} \left(g(x(t)) + \int_t^s L^0 g(x(z))dz \right) dw(s) \\ &+ \int_t^{t+\Delta t} \Bigg(\int_t^s \big(L^1 g(x(t)) \\ &+ \int_t^z L^0 L^1 g(x(h))dh + \int_t^z L^1 L^1 g(x(h))dw(h) \big) dw(z) \Bigg) dw(s). \end{aligned}$$

It can be rewritten in the following way

$$
\begin{aligned}
x_{t,x}(t+\Delta t) &= x(t) + f(x(t+\Delta t)) \int_t^{t+\Delta t} ds + g(x(t)) \int_t^{t+\Delta t} dw(s) \\
&+ L^1 g(x(t)) \int_t^{t+\Delta t} \int_t^s dw(z) dw(s) \\
&- \int_t^{t+\Delta t} \left(\int_s^{t+\Delta t} L^0 f(x(z)) dz + \int_s^{t+\Delta t} L^1 f(x(z)) dW_z \right) ds \\
&+ \int_t^{t+\Delta t} \left(\int_t^s L^0 g(x(z)) dz \right) dw(s) \\
&+ \int_t^{t+\Delta t} \left(\int_t^s \left(\int_t^z L^0 L^1 g(x(h)) dh + \int_t^z L^1 L^1 g(x(h)) dw(h) \right) dw(z) \right) dw(s).
\end{aligned}
$$

Since we are dealing with a scalar case we can show that (Glasserman 2003; Kloeden and Platen 1992)

$$
\int_t^{t+\Delta t} \int_t^s dw(z) dw(s) = \frac{1}{2}[(w(t+\Delta t) - w(t))^2 - \Delta t].
$$

Hence

$$
\begin{aligned}
x_{t,x}(t+\Delta t) &= x(t) + f(x(t+\Delta t)) \int_t^{t+\Delta t} ds + g(x(t)) \int_t^{t+\Delta t} dw(s) \\
&+ \frac{1}{2} L^1 g(x(t))[(\Delta W_{t+\Delta t})^2 - \Delta t] \\
&- \int_t^{t+\Delta t} \left(\int_s^{t+\Delta t} L^0 f(x(z)) dz + \int_s^{t+\Delta t} L^1 f(x(z)) dW_z \right) ds \\
&+ \int_t^{t+\Delta t} \left(\int_t^s L^0 g(x(z)) dz \right) dw(s) \\
&+ \int_t^{t+\Delta t} \left(\int_t^s \left(\int_t^z L^0 L^1 g(x(h)) dh + \int_t^z L^1 L^1 g(x(h)) dw(h) \right) dw(z) \right) dw(s).
\end{aligned}
$$

We introduce implicitness in the second order term of the expansion

$$
L^1 g(x(t)) = L^1 g(x(t+\Delta t)) - \int_t^{t+\Delta t} L^0 L^1 g(x(h)) dh - \int_t^{t+\Delta t} L^1 L^1 g(x(h)) dw(h),
$$

which leads us to

$$\begin{aligned} x_{t,x}(t+\Delta t) = x(t) + f(x(t+\Delta t)) \int_t^{t+\Delta t} ds + g(x(t)) \int_t^{t+\Delta t} dw(s) + \frac{1}{2} L^1 g(x(t)) \Delta W_{t+\Delta t}^2 \\ - \frac{1}{2} L^1 g(x(t+\Delta t)) \Delta t \\ + \frac{1}{2} \Delta t \Bigg(\int_t^{t+\Delta t} L^0 L^1 g(x(h)) dh + \int_t^{t+\Delta t} L^1 L^1 g(x(h)) dw(h) \Bigg) \\ - \int_t^{t+\Delta t} \left(\int_s^{t+\Delta t} L^0 f(x(z)) dz + \int_s^{t+\Delta t} L^1 f(x(z)) dW_z \right) ds \\ + \int_t^{t+\Delta t} \left(\int_t^s L^0 g(x(z)) dz \right) dw(s) \\ + \int_t^{t+\Delta t} \left(\int_t^s \left(\int_t^z L^0 L^1 g(x(h)) dh + \int_t^z L^1 L^1 g(x(h)) dw(h) \right) dw(z) \right) dw(s). \end{aligned}$$

Definition 2.4.2. *For partition $\mathcal{P}_{\Delta t} := \{t_k = k\Delta t : k = 0, 1, 2, ...\}$ of the time interval $[0, \infty)$. The double implicit Milstein scheme has the following form*

$$X_{t_{k+1}} = X_{t_k} + f(X_{t_{k+1}})\Delta t + g(X_{t_k})\Delta w_{t_k} + \frac{1}{2} L^1 g(X_{t_k}) \Delta w_{t_k}^2 - \frac{1}{2} L^1 g(X_{t_{k+1}}) \Delta t.$$

It is also interesting to note that the scheme may also be obtained from the implicit Milstein scheme for the Stratanovich SDE. In fact, in the scalar case the implicit Milstein scheme for the Stratanovich SDE is given by

$$\begin{aligned} X_{t_{k+1}} = X_{t_k} + \underline{f}(X_{t_{k+1}})\Delta t - \frac{1}{2} L^1 g(X_{t_k}) \Delta t \\ + g(X_{t_k}) \Delta W_{t_k} + \frac{1}{2} L^1 g(X_{t_k}) \Delta w_{t_k}^2, \end{aligned}$$

where

$$\underline{f} = f - \frac{1}{2} L^1 b.$$

2.4.2 Properties of Approximations

Here we formally define the strong measures of error for numerical approximations used in this thesis.

Definition 2.4.3. *We shall say that a general discrete time approximation X_{t_k} with a step-size Δt* converges strongly *to the solution of the SDEs $x(t)$ at time T*

if

$$\lim_{\Delta t \to 0} \mathbb{E}(|x(T) - X_T|) = 0.$$

Definition 2.4.4. *We shall say that a general discrete time approximation X_{t_k}* converges strongly with order δ *at time T if there exists a positive constant C, which does not depend on Δt, such that*

$$\mathbb{E}[|x(T) - X_T|] \leq C\Delta t^{\delta}. \tag{2.8}$$

We say that it converges strongly with order δ uniformly in time *if*

$$\mathbb{E}[\sup_{0 \leq t_k \leq T} |x(t_k) - X(t_k)|] \leq C\Delta t^{\delta}. \tag{2.9}$$

Using the Borel-Cantelli Lemma it is possible to pass from strong convergence to pathwise error. For example, in (Kloeden and Neuenkirch 2007) it is shown that given any $\varepsilon > 0$, there exists a path-dependent random variable $K = K(\varepsilon)$ such that, for all sufficiently small Δt

$$\sup_{0 \leq t_k \leq T} |x(t_k) - X(t_k)| \leq K(\varepsilon)\Delta t^{\delta - \varepsilon} \quad a.s.$$

Once we know that the numerical method designed by us strongly converges to the true solution of a SDE, the second property to be investigated is stability. Convergence and stability complement each other. Convergence gives us information about behaviour of the numerical scheme on a fixed time interval letting the time-step decrease to zero. Stability analysis, on the other hand, allows us to analyse behaviour of the approximation for a fixed step-size when the time interval expands to infinity.

Roughly speaking, we say that the system is stable if the trajectories which are “close” to each other at a specific instant, remain ”close” to each other at all subsequent instants. In many cases, to show that the system is stable it is enough to show that the *trivial solution*, $x(t) = 0$, is stable.

Definition 2.4.5. *The trivial solution of equation* (2.1) *is said to be globally almost surely stable if*

$$\lim_{t \to \infty} |x(t)| = 0 \quad a.s.$$

Definition 2.4.6. *The trivial solution of equation* (2.1) *is said to be globally pth moment stable if*

$$\lim_{t\to\infty} \mathbb{E}\,|x(t)|^p = 0.$$

Once we deal with the stability of the numerical approximation the key question we ask is: For what step-size Δt does the numerical method share the stability property of the underlying test problem? This question is related to the concept of *A-stability*. We say that the method is A-stable if it can reproduce the stability property of its test equation for all $\Delta t > 0$. In case of implicit methods we need to slightly redefine A-stability property. In order to prove the existence of a unique solution to the implicit scheme we typically need the step-size to satisfy $\Delta t < L^{-1}$, where L stands for Lipschitz constant from the one-sided Lipschitz condition. Therefore, in this case we say that the implicit scheme is A-stable if it is stable for all Δt where it is well defined.

Chapter 3

Approximations of Nonlinear SDEs

> Only mathematicians can read "musical scores" containing many numerical formulae, and play that "music" in their hearts. Accordingly, I once believed that without numerical formulae, I could never communicate the sweet melody played in my heart.
>
> ---
>
> My Sixty Years in Studies of Probability Theory
> Kiyosi Itô

In this chapter we are interested in strong convergence and almost sure stability of Backward Euler-Maruyama approximation to the solution of stochastic differential equations with highly nonlinear coefficients. Our goal is to prove convergence under monotone conditions. This work can be read as generalization of the results in (Higham, Mao, and Stuart 2003b), where authors using implicit schemes derive strong convergence theorems under a one-sided Lipschitz condition on a drift and a linear growth condition on the diffusion. They have demonstrated that as in the deterministic setting, drift implicit methods enable analysis of systems with non-linearities in the drift part. In this chapter we show that even if the non-linearities appear in the diffusion part of the underlying SDEs, drift implicit methods still perform very well. In order to prove strong convergence theorems we introduce a new numerical scheme - Forward-Backward

Euler Maruyama. In addition, we examine global almost sure asymptotic stability in this non-linear setting. We present a stochastic counterpart of the discrete LaSalle principle from which we deduce stability properties of numerical methods. The material from this chapter can be found in (Szpruch and Mao 2010).

3.1 Problem Specification

Let $w(t) = (w_1(t), ..., w_d(t))^T$ be a d-dimensional Brownian motion defined on the probability space $(\Omega, \mathcal{F}, \{\mathcal{F}_t\}_{t\geq 0}, \mathbb{P})$. We study the numerical approximation of the stochastic differential equations

$$dx(t) = f(x(t))dt + g(x(t))dw(t). \tag{3.1}$$

Here $x(t) \in \mathbb{R}^n$ for each $t \geq 0$. Thus, $f : \mathbb{R}^n \to \mathbb{R}^n$ and $g : \mathbb{R}^n \to \mathbb{R}^{n\times d}$.

3.2 Existence and Uniqueness of Solution

We require the coefficients f and g to be locally Lipschitz continuous and to satisfy the monotone condition, that is

Assumption 3.2.1. *Both functions f and g in* (3.1) *satisfy the following conditions:*
Local Lipschitz condition. For each integer $m \geq 1$, there is a positive constant K_m such that

$$|f(x) - f(y)| + |g(x) - g(y)| \leq K_m |x - y|, \tag{3.2}$$

for those $x, y \in \mathbb{R}^n$ with $|x| \vee |y| \leq m$.
Monotone condition. There exist positive constants α and β such that

$$\langle x, f(x)\rangle + \frac{1}{2}|g(x)|^2 \leq \alpha + \beta |x|^2, \tag{3.3}$$

for all $x \in \mathbb{R}^n$.

It is a classical result from stochastic analysis that under Assumption 3.2.1, there exists a unique solution for any given initial value $x(0) = x_0 \in \mathbb{R}^n$, (Friedman 1976; Mao 2007). The reason we present the theorem with the proof here is that it reveals the upper bound for the probability that the process $x(t)$ stays on a compact domain for a finite time $T > 0$. The bound will be used to derive the main convergence theorem of this chapter.

Theorem 3.2.2. *Let Assumption 3.2.1 hold. Then for any given initial value $x(0) = x_0 \in \mathbb{R}^n$, there exists a unique, global solution $x(t)$ to the equation (3.1) on $t \geq 0$. Moreover, the solution has the properties that for any $T > 0$,*

$$\mathbb{E}\left|x(T)\right|^2 \leq (\left|x_0\right|^2 + 2\alpha T)\exp(2\beta T), \tag{3.4}$$

and

$$\mathbb{P}(\tau_m \leq T) \leq \frac{(\left|x_0\right|^2 + 2\alpha T)\exp(2\beta T)}{\left|m\right|^2}, \tag{3.5}$$

where

$$\tau_m = \inf\{t \geq 0 : \quad |x(t)| > m\}. \tag{3.6}$$

Proof. It is well known that under Assumption 3.2.1, for any given initial value $x_0 \in \mathbb{R}^n$ there exists a unique solution $x(t)$ to the SDEs (3.1), (Friedman 1976; Mao 2007). Therefore we only need to prove that (3.4) and (3.5) hold. Applying the Itô formula to the function $V(x,t) = \left|x\right|^2$, we obtain

$$x(t) = x(0) + \int_0^t LV(x(s))ds + 2\int_0^t \langle x(s), g(x(s))\rangle dw(s),$$

where the diffusion operator is given by

$$LV(x(s)) = 2\Big(\langle x(s), f(x(s))\rangle + \frac{1}{2}\left|g^2(x(s))\right|\Big). \tag{3.7}$$

By Assumption 3.2.1

$$LV(x,t) \leq 2\alpha + 2\beta\left|x\right|^2. \tag{3.8}$$

Therefore

$$\mathbb{E}\left|x(t\wedge\tau_m)\right|^2 \leq \left|x_0\right|^2 + 2\alpha T + \int_0^t 2\beta\mathbb{E}\left|x(s\wedge\tau_m)\right|^2 ds,$$

and by utilizing Gronwall's inequality we obtain

$$\mathbb{E}\left|x(T\wedge\tau_m)\right|^2 \leq (\left|x_0\right|^2 + 2\alpha T)\exp(2\beta T). \tag{3.9}$$

Hence

$$\mathbb{P}(\tau_m \le T)\,|m|^2 \le [|x_0|^2 + 2\alpha T]\exp(2\beta T). \tag{3.10}$$

Next, by (3.9), letting $m \to \infty$ and applying Fatou's lemma, we obtain

$$\mathbb{E}\,|x(T)|^2 \le [|x_0|^2 + 2\alpha T]\exp(2\beta T), \tag{3.11}$$

which gives the other assertion (3.4) and completes the proof. □

3.3 Backward Euler-Maruyama Scheme

As indicated in Section 1.3, in order to approximate (3.1) we use backward Euler-Maruyama

$$X_{t_{k+1}} = X_{t_k} + f(X_{t_{k+1}})\Delta t + g(X_{t_k})\Delta w_{t_k}, \tag{3.12}$$

where $\Delta w_{t_k} = w_{t_{k+1}} - w_{t_k}$ and $X_{t_0} = x_0$.

3.3.1 Existence and Uniqueness

Since we are dealing with an implicit scheme we need to make sure that it has a unique solution $X_{t_{k+1}}$ given X_{t_k}. The lemma below gives existence and uniqueness conditions for the solution to the equation $F(x) = b$. Based on it we prove existence and uniqueness of the solution to the backward Euler-Maruyama scheme.

Lemma 3.3.1. *Let F, $F : \mathbb{R}^n \to \mathbb{R}^n$, be a vector field on $\mathbb{R}^n$ and consider the equation*

$$F(x) = b, \tag{3.13}$$

for a given $b \in \mathbb{R}^n$. If F is monotone, i.e.,

$$\langle x - y, F(x) - F(y)\rangle > 0,$$

for all $x, y \in \mathbb{R}^n$, $x \ne y$, then equation (3.13) has at most one solution. Furthermore, if F is continuous and it is coercive, i.e.,

$$\lim_{|x|\to\infty} \frac{\langle x, F(x)\rangle}{|x|} = \infty,$$

then for every $b \in \mathbb{R}^n$, the equation (3.13) has a unique solution $x \in \mathbb{R}^n$. Moreover, there exists an inverse operator F^{-1}.

This lemma follows directly from Theorem 26.A in (Zeidler 1985).
To prove the existence and uniqueness of solution to (3.12), in addition to Assumption 3.2.1, we ask for the function f to satisfy a one-sided Lipschitz condition.

Assumption 3.3.2. *One-sided Lipschitz condition. There exists a constant $L > 0$, such that*

$$\langle x - y, f(x) - f(y) \rangle \leq L\,|x - y|^2 \quad \forall x, y \in \mathbb{R}^n. \tag{3.14}$$

Lemma 3.3.3. *Define, for any given $\Delta t \leq \frac{1}{L}$,*

$$F(x) = x - f(x)\Delta t, \quad x \in \mathbb{R}^n.$$

Then under Assumptions 3.2.1 and 3.3.2, for any $b \in \mathbb{R}^n$, there exists a unique $x \in \mathbb{R}^n$ such that

$$F(x) = b.$$

Proof. In view of Lemma 3.3.1 we need to show that the function F is continuous, coercive and strictly monotone. Clearly, $F(x)$ is continuous on $\mathbb{R}^n$ due to Assumption 3.2.1. By Assumption 3.3.2, $F(x)$ is monotone. Indeed

$$\langle x - y, F(x) - F(y) \rangle \geq |x - y|^2 - L\Delta t\,|x - y|^2 = (1 - L\Delta t)\,|x - y|^2 > 0,$$

for $\Delta t < \frac{1}{L}$. Again by Assumption 3.3.2

$$\langle x, f(x) \rangle \leq L\,|x|^2 + \langle x, f(0) \rangle.$$

Hence

$$\langle x, F(x) \rangle = \langle x, x - f(x)\Delta t \rangle \geq (1 - L\Delta t)\,|x|^2 - \langle x, f(0) \rangle \Delta t,$$

which shows that the coercivity condition holds on $\mathbb{R}^n$. The proof of Lemma 3.3.3 is complete. □

From now on we always assume that $\Delta t < \frac{1}{L}$.

3.3.2 Moment Properties of BEM

In this section we will show that the second moment for the solution of BEM is bounded; (Theorem 3.3.6). To obtain a bound we employ a stopping time technique. It is well known, (Buchmann 2005; Broadie, Glasserman, and Kou 1997; Mannella 1999), that in discrete time approximations for a stochastic process, the problem of overshooting the level where we would like to stop our process appears. Nevertheless, Lemma 3.3.4 below shows that the moments for such a stopped processes can be controlled.

Lemma 3.3.4. *Under Assumptions 3.2.1 and 3.3.2 , for any integer $p \geq 2$ and sufficiently large integer m, there exists a constant $K = K(p,m)$, such that*

$$\mathbb{E}\left[|X_{t_k}|^p \mathbf{1}_{[0,\lambda_m]}(k)\right] < K \qquad \text{for any} \quad k \geq 0,$$

where

$$\lambda_m = \inf\{k : |X_{t_k}| > m\}. \tag{3.15}$$

Proof. We observe that when $k \in [0, \lambda_m]$, $\left|X_{t_{k-1}}\right| < m$, but it might be that $|X_{t_k}| > m$, so the lemma is not obvious. By definition of BEM and Young's inequality of the form

$$xy \leq \frac{\delta}{2}x^2 + \frac{1}{2\delta}y^2,$$

where we choose δ such that $\delta(\frac{5p-6}{2p}) = C(\delta) < 1 - \beta L^{-1}$, we have

$$\begin{aligned}
|X_{t_k}|^2 &\leq \delta\,|X_{t_k}|^2 + \frac{1}{2\delta}\left|X_{t_{k-1}}\right|^2 \\
&\quad + \langle X_{t_k}, f(X_{t_k})\Delta t\rangle \\
&\quad + \frac{1}{2\delta}\left|g(X_{t_{k-1}})\Delta w_{t_{k-1}}\right|^2 \\
&\leq \delta\,|X_{t_k}|^2 + \beta\,|X_{t_k}|^2\Delta t + \alpha\Delta t + \frac{1}{2\delta}\left|X_{t_{k-1}}\right|^2 \\
&\quad + \frac{1}{2\delta}\left|g(X_{t_{k-1}})\Delta w_{t_{k-1}}\right|^2,
\end{aligned}$$

where the last inequality follows from Assumption 3.2.1. Multiplying both sides of the above inequality by $|X_{t_k}|^{p-2}$ leads to

$$(1-\beta\Delta t-\delta)\,|X_{t_k}|^p \le \frac{1}{2\delta}\,|X_{t_k}|^{p-2}\left|X_{t_{k-1}}\right|^2+\alpha\Delta t\,|X_{t_k}|^{p-2} + \frac{1}{2\delta}\,|X_{t_k}|^{p-2}\left|g(X_{t_{k-1}})\Delta w_{t_{k-1}}\right|^2.$$

Applying Young's inequality in the form

$$x^{p-2}y^2 \le \delta^2\frac{p-2}{p}x^p+\frac{2}{p\delta^{p-2}}y^p,$$

results in

$$\Big((1-\beta\Delta t)-C(\delta)\Big)\,|X_{t_k}|^p \le \frac{1}{p\delta^{p-1}}\Big(\left|X_{t_{k-1}}\right|^p+(2\delta\alpha\Delta t)^{\frac{p}{2}}+\left|g(X_{t_{k-1}})\right|^p\left|\Delta w_{t_{k-1}}\right|^p\Big).$$

Hence by Hölder's inequality

$$\Big((1-\beta\Delta t)-C(\delta)\Big)\mathbb{E}\left[|X_{t_k}|^p\,\mathbf{1}_{[0,\lambda_m]}(k)\right] \le \frac{1}{p\delta^{p-1}}\Big(|m|^p+(2\delta\alpha\Delta t)^{\frac{p}{2}} + (\mathbb{E}[\left|g(X_{t_{k-1}})\right|^p\mathbf{1}_{[0,\lambda_m]}(k)]^2)^{1/2}(\mathbb{E}\left|\Delta w_{t_{k-1}}\right|^{2p})^{1/2}\Big).$$

By Assumption 3.2.1 and the fact that there exists a positive constant $C(p)$, such that $\mathbb{E}\left|\Delta w_{t_{k-1}}\right|^{2p} < C(p)$, we obtain

$$\mathbb{E}\left[|X_{t_k}|^p\,\mathbf{1}_{[0,\lambda_m]}(k)\right] < C(m,p),$$

as required. □

To prove boundedness of the second moment for (3.12), we need an additional mild assumption on the coefficients f and g.

Assumption 3.3.5. *The coefficients of the equation (3.1) satisfy the polynomial growth condition, that is for some $h \ge 1$ there exists a positive constant $H > 0$,* *such that*

$$|f(x)| \vee |g(x)| \leq H(1+|x|^h), \qquad \forall x \in \mathbb{R}^n. \tag{3.16}$$

Theorem 3.3.6. *Let Assumptions 3.2.1, 3.3.2 and 3.3.5 hold. Let $T > 0$ and $\Delta t^* \in (0, (\max\{L, 4\beta\})^{-1})$. Then, there exists a constant $K > 0$, such that*

$$\sup_{\Delta t \leq \Delta t^*} \sup_{0 \leq t_k \leq T} \mathbb{E} |X_{t_k}|^2 < K.$$

Proof. Let

$$\lambda_m = \inf\{k : |X_{t_k}| > m\}.$$

Then λ_m is a stopping time with respect to $\{\mathcal{F}_{t_k}\}_{k \geq 0}$. From (3.12) we have the following inequality

$$\left|X_{t_{k+1}}\right|^2 - |X_{t_k}|^2 \leq 2\langle X_{t_{k+1}}, f(X_{t_{k+1}})\Delta t\rangle + 2\langle X_{t_k}, g(X_{t_k})\Delta w_{t_k}\rangle + |g(X_{t_k})|^2 |\Delta w_{t_k}|^2.$$

Let N be any nonnegative integer such that $N\Delta t \leq T$. Summing up the sides of the above inequality from $k = 0$ to $N \wedge \lambda_m$, we get

$$
\begin{aligned}
\left|X_{t_{N\wedge\lambda_m+1}}\right|^2 \le & \left(|X_{t_0}|^2 + 2\langle X_{t_0}, g(X_{t_0})\Delta w_{t_0}\rangle + |g(X_{t_0})|^2 |\Delta w_{t_0}|^2\right) \qquad (3.17)\\
& + \sum_{k=1}^{(N\wedge\lambda_m)+1} 2\langle X_{t_k}, f(X_{t_k})\rangle \Delta t \\
& + \sum_{k=1}^{N\wedge\lambda_m} |g(X_{t_k})|^2 \Delta t + \sum_{k=1}^{N\wedge\lambda_m} 2\langle X_{t_k}, g(X_{t_k})\Delta w_{t_k}\rangle \\
& + \sum_{k=1}^{N\wedge\lambda_m} |g(X_{t_k})|^2 \left[|\Delta w_{t_k}|^2 - \Delta t\right] \\
& = \left(|X_{t_0}|^2 + 2\langle X_{t_0}, g(X_{t_0})\Delta w_{t_0}\rangle + |g(X_{t_0})|^2 |\Delta w_{t_0}|^2\right) \\
& + \sum_{k=1}^{N} 2\langle X_{t_k}, f(X_{t_k})\rangle \mathbf{1}_{[0,\lambda_m]}(k)\Delta t \qquad (3.18)\\
& + 2\langle X_{t_{(N\wedge\lambda_m)+1}}, f(X_{t_{(N\wedge\lambda_m)+1}})\rangle \Delta t \\
& + \sum_{k=1}^{N} |g(X_{t_k})|^2 \mathbf{1}_{[0,\lambda_m]}(k)\Delta t \\
& + \sum_{k=1}^{N} 2\langle X_{t_k}, g(X_{t_k})\rangle \mathbf{1}_{[0,\lambda_m]}(k)\Delta w_{t_k} \\
& + \sum_{k=1}^{N} |g(X_{t_k})|^2 \mathbf{1}_{[0,\lambda_m]}(k) \left[|\Delta w_{t_k}|^2 - \Delta t\right].
\end{aligned}
$$

Applying Lemma 3.3.4, Assumption 3.3.5 and noting that X_{t_k} and $\mathbf{1}_{[0,\lambda_m]}(k)$ are $\mathcal{F}_{t_k}$-measurable while Δw_{t_k} is independent of $\mathcal{F}_{t_k}$, we can take the expectation on both sides of (3.17) to get

$$
\begin{aligned}
& \mathbb{E}\left|X_{t_{N\wedge\lambda_m+1}}\right|^2 \\
& \le C_1 + \mathbb{E}\Bigg[\sum_{k=1}^{N}(2\langle X_{t_k}, f(X_{t_k})\rangle \\
& + |g(X_{t_k})|^2)\mathbf{1}_{[0,\lambda_m]}(k)\Delta t + 2\langle X_{t_{(N\wedge\lambda_m)+1}}, f(X_{t_{(N\wedge\lambda_m)+1}})\rangle \Delta t\Bigg],
\end{aligned}
$$

where $C_1 = |X_{t_0}|^2 + |g(X_{t_0})|^2 \Delta t^*$. By Assumption 3.2.1

$$2\langle x, f(x)\rangle + |g(x)|^2 \leq 2\alpha + 2\beta |x|^2, \quad x \in \mathbb{R}^n.$$

We hence obtain that

$$\mathbb{E}\left|X_{t_{(N\wedge\lambda_m)+1}}\right|^2 \leq C_1 + 2\beta\Big[\sum_{k=1}^{N} \mathbb{E}|X_{t_k}|^2 \mathbf{1}_{[0,\lambda_m]}(k)\Delta t + \mathbb{E}\left|X_{t_{(N\wedge\lambda_m)+1}}\right|^2 \Delta t\Big] \\ + 2\alpha(T+\Delta t).$$

Then

$$\mathbb{E}\left|X_{t_{(N\wedge\lambda_m)+1}}\right|^2 \leq (C_1 + 2\alpha(T+\Delta t))(1-2\beta\Delta t)^{-1} \\ + (1-2\beta\Delta t)^{-1}2\beta\Big[\sum_{k=1}^{N} \mathbb{E}|X_{t_k}|^2 \mathbf{1}_{[0,\lambda_m]}(k)\Delta t\Big].$$

Now we can observe that

$$\mathbb{E}\left[\left|X_{t_{N+1}}\right|^2 \mathbf{1}_{[0,\lambda_m]}(N)\right] \leq \mathbb{E}\left|X_{t_{N\wedge\lambda_m+1}}\right|^2.$$

By discrete Gronwall's inequality and the fact that $(1-2\beta\Delta t)^{-1} \leq 2$ for $\Delta t \leq \Delta t^*$

$$\mathbb{E}\left[\left|X_{t_{N+1}}\right|^2 \mathbf{1}_{[0,\lambda_m]}(N)\right] \leq (C_1 + 2\alpha_0(T+\Delta t))2\exp\left((4\beta)T\right), \tag{3.19}$$

where we use the fact that $N\Delta t \leq T$. Thus, letting $m \to \infty$ in (3.19) and applying Fatou's lemma, we get

$$\mathbb{E}\left|X_{t_{N+1}}\right|^2 \leq 2(C_1 + 2\alpha_0(T+\Delta t))\exp\left(4\beta T\right).$$

The proof is complete. □

3.4 Forward-Backward Euler-Maruyama Scheme

We find in our analysis it is convenient to work with a continuous extension of a numerical method. This continuous extension enables us to use a powerful continuous-time stochastic analysis in order to formulate theorems on numerical approximations. Let us define

$$\eta(t) := t_k, \quad \text{for} \quad t \in [t_k, t_{k+1}),\ k \geq 0,$$

$$\eta_+(t) := t_{k+1}, \quad \text{for} \quad t \in [t_k, t_{k+1}),\ k \geq 0.$$

The continuous version of the BEM is given by

$$X(t) = X_{t_0} + \int_0^t f(X_{\eta_+(s)})ds + \int_0^t g(X_{\eta(s)})dw(s), \quad t \geq 0. \tag{3.20}$$

However, in case of the BEM standard measurability conditions are not satisfied. Indeed, we may notice that $X(t)$, defined by (3.20), is not $\mathcal{F}_t$-measurable. For this reason we introduce a new numerical scheme. In terms of the general drift and diffusion coefficients, f and g, once we compute the value X_{t_k} from BEM, that is

$$X_{t_k} = X_{t_{k-1}} + f(X_{t_k})\Delta t + g(X_{t_{k-1}})\Delta w_{t_{k-1}}, \tag{3.21}$$

we define the Forward-Backward Euler-Maruyama (FBEM) scheme as follows

$$\hat{X}_{t_{k+1}} = \hat{X}_{t_k} + f(X_{t_k})\Delta t + g(X_{t_k})\Delta w_{t_k}, \tag{3.22}$$

where $\hat{X}_{t_0} = X_{t_0} = x_0$. The continuous version of the FBEM is given by

$$\hat{X}(t) = \hat{X}_{t_0} + \int_0^t f(X_{\eta(s)})ds + \int_0^t g(X_{\eta(s)})dw(s), \quad t \geq 0. \tag{3.23}$$

Note that the continuous and discrete FBEM coincide at the gridpoints; that is, $\hat{X}(t_k) = \hat{X}_{t_k}$.

3.4.1 Strong Convergence On Compact Domain

It this subsection we prove the strong convergence theorem. We begin by showing that the FBEM and the BEM schemes stay close to each other on a compact domain. Then we estimate the probability that BEM (3.12) will not explode on a finite time interval.

Lemma 3.4.1. *Under Assumptions 3.2.1, 3.3.2 and 3.3.5 for any integer $p \geq 2$, $T \geq 0$, and $\Delta t^* \in (0, (\max\{L, 4\beta\})^{-1})$ there exists a constant $K = K(m, p, T)$*

such that, if $\Delta t \leq \Delta t^*$,

$$\mathbb{E}\left[\left|\hat{X}_{t_k} - X_{t_k}\right|^p \mathbf{1}_{[0,\lambda_m]}(k)\right] \leq K\Delta t^p, \qquad \forall t_k \in [0,T].$$

Proof. Summing up forward-backward and backward schemes, respectively, we obtain

$$\hat{X}_{t_N} = \hat{X}_{t_0} + \sum_{k=0}^{N-1} f(X_{t_k})\Delta t + \sum_{k=0}^{N-1} g(X_{t_k})\Delta w_{t_k},$$

$$X_{t_N} = X_{t_0} + \sum_{k=0}^{N-1} f(X_{t_{k+1}})\Delta t + \sum_{k=0}^{N-1} g(X_{t_k})\Delta w_{t_k}.$$

Now by Hölder's inequality, Lemma 3.3.4 and Assumption 3.3.5, there exists a constant $C > 0$, such that

$$\mathbb{E}\left[\left|\hat{X}_{t_N} - X_{t_N}\right|^p \mathbf{1}_{[0,\lambda_m]}(N)\right] \tag{3.24}$$

$$= \mathbb{E}\left[\left|f(X_{t_0})\Delta t - f(X_{t_N})\Delta t\right|^p \mathbf{1}_{[0,\lambda_m]}(N)\right] \leq C\Delta t^p, \tag{3.25}$$

as required. □

Theorem 3.4.2. *Let Assumptions 3.2.1, 3.3.2, 3.3.5 hold and* $T > 0$ *be arbitrary. Then, for any given* $\epsilon > 0$*, there exists an* N_0 *such that for every* $m \geq N_0$*, we can find a* $\Delta t_0 = \Delta t_0(m)$ *so that whenever* $\Delta t \leq \Delta t_0$*,*

$$\mathbb{P}(\vartheta_m < T) \leq \epsilon,$$

where $\vartheta_m = \inf\{t > 0 : \left|\hat{X}(t)\right| \geq m \quad \text{or} \quad \left|X_{\eta(t)}\right| \geq m\}$.

Proof. Let $s \in [0, T \wedge \vartheta_m)$. Then by the Itô formula with $V(x) = |x|^2$,

$$\begin{aligned} dV(\hat{X}(s)) &= 2\langle \hat{X}(s), f(X_{\eta(s)})\rangle ds + \text{trace}[g^T(X_{\eta(s)})I_{n\times n}g(X_{\eta(s)})]ds \\ &+ 2\langle \hat{X}(s), g(X_{\eta(s)})\rangle dw(s) \\ &\leq 2\langle \hat{X}(s) - X_{\eta(s)} + X_{\eta(s)}, f(X_{\eta(s)})\rangle ds \\ &+ \left|g(X_{\eta(s)})\right|^2 ds \\ &+ 2\langle \hat{X}(s), g(X_{\eta(s)})dw(s) \\ &= LV(X_{\eta(s)}))ds + 2\langle \hat{X}(s) - X_{\eta(s)}, f(X_{\eta(s)})\rangle ds \\ &+ 2\langle \hat{X}(s), g(X_{\eta(s)})\rangle dw(s) \\ &\leq LV(X_{\eta(s)}))ds + 2\left|\hat{X}(s) - X_{\eta(s)}\right| \left|f(X_{\eta(s)})\right| ds \\ &+ 2\langle \hat{X}(s), g(X_{\eta(s)})\rangle dw(s), \end{aligned}$$

where the diffusion operator is defined by (3.7). By Assumption 3.2.1, for $|x| \leq m$

$$|f(x)|^2 \leq 2(|f(y) - f(0)|^2 + |f(0)|^2) \leq 2(K_m |x|^2 + |f(0)|^2),$$

and

$$|g(x)|^2 \leq 2(|g(y) - g(0)|^2 + |g(0)|^2) \leq 2(K_m |x|^2 + |g(0)|^2).$$

Recalling that $LV(x) < 2(\alpha + \beta |x|^2)$, we then have

$$\begin{aligned} \mathbb{E}\left|\hat{X}(T\wedge\vartheta_m)\right|^2 \leq \left|\hat{X}(0)\right|^2 + 2\alpha T + 4\beta \int_0^T \mathbb{E}\left|\hat{X}(s\wedge\vartheta_m)\right|^2 ds \\ + C(m)\mathbb{E}\int_0^{T\wedge\vartheta_m} \left|X_{\eta(s)} - \hat{X}(s)\right| ds. \end{aligned}$$

By Lemma 3.4.1, we obtain

$$\mathbb{E}\int_0^{T\wedge\vartheta_m} \left|X_{\eta(s)} - \hat{X}_{\eta(s)}\right| ds \leq C(m, T)\Delta t. \tag{3.26}$$

To bound the term $\mathbb{E}\int_0^{T\wedge\vartheta_m} \left|\hat{X}_{\eta(s)} - \hat{X}(s)\right| ds$, given $s \in [0, T\wedge\vartheta_m)$, let k be an integer for which $s \in [t_k, t_{k+1})$. Then

$$\left|\hat{X}_{\eta(s)} - \hat{X}(s)\right| = \left|\int_{t_k}^s f(X_{t_k})ds + \int_{t_k}^s g(X_{t_k})dw(s)\right|.$$

By Hölder's inequality

$$\mathbb{E}\int_0^{T\wedge\vartheta_m}\left|\hat{X}_{\eta(s)}-\hat{X}(s)\right|ds \le C(m,T)\Delta t^{\frac{1}{2}},$$

where $C(m,T) > 0$ is constant. This leads us to

$$\begin{aligned}
E\int_0^{T\wedge\vartheta_m}\left|X_{\eta(s)}-\hat{X}(s)\right|ds &\le \mathbb{E}\int_0^{T\wedge\vartheta_m}\left|\hat{X}_{\eta(s)}-\hat{X}(s)\right|ds \\
&\quad + \mathbb{E}\int_0^{T\wedge\vartheta_m}\left|X_{\eta(s)}-\hat{X}_{\eta(s)}\right|ds \\
&\le C(m,T)\Delta t^{\frac{1}{2}}. \qquad (3.27)
\end{aligned}$$

Therefore

$$\mathbb{E}\left|\hat{X}(t\wedge\vartheta_m)\right|^2 \le \left|\hat{X}(0)\right|^2 + 2\alpha T + C(m,T)\Delta t^{\frac{1}{2}} + 4\beta\int_0^T \mathbb{E}\left|\hat{X}(s\wedge\vartheta_m)\right|^2 ds.$$

By Gronwall's inequality

$$\mathbb{E}\left|\hat{X}(t\wedge\vartheta_m)\right|^2 \le [\left|\hat{X}(0)\right|^2 + 2\alpha T + C(m,T)\Delta t^{\frac{1}{2}}]\exp(4\beta T), \qquad (3.28)$$

which implies that

$$\mathbb{P}(\vartheta_m < T) \le \frac{[\left|\hat{X}(0)\right|^2 + 2\alpha T + C(m,T)\Delta t^{1/2}]\exp(4\beta T)}{|m|^2}.$$

Now, for any given $\epsilon > 0$, we choose N_0 such that for any $m \ge N_0$

$$\frac{[\left|\hat{X}(0)\right|^2 + 2\alpha T]\exp(4\beta T)}{|m|^2} \le \frac{\epsilon}{2}.$$

Then, we can choose $\Delta t_0 = \Delta t_0(m)$, such that for any $\Delta t \le \Delta t_0$

$$\frac{\exp(4\beta T)C(m,T)\Delta t^{1/2}}{|m|^2} \le \frac{\epsilon}{2},$$

whence $\mathbb{P}(\vartheta_m < T) \le \epsilon$ as required. □

3.4.2 Strong Convergence on the Whole Domain

In this section we study the strong convergence of BEM (3.12) to the solution of (3.1). First, we will show that the continuous extension of FBEM (3.23) converges to the true solution on a compact domain. This, together with Theorem 3.4.2, will enable us to extend convergence to the whole domain.
Let us define the stopping time θ_m as follows

$$\theta_m = \tau_m \wedge \vartheta_m,$$

where τ_m and ϑ_m are defined in Theorem 3.2.2 and Theorem 3.4.2, respectively.

Lemma 3.4.3. *Under Assumptions 3.2.1, 3.3.2 and 3.3.5 for any $p \geq 2$, $T > 0$ and sufficiently large m, there exists a constant $K = K(p, T, m)$, such that*

$$\mathbb{E}\left[\sup_{0\leq t\leq T}\left|\hat{X}(t\wedge\theta_m) - x(t\wedge\theta_m)\right|^p\right] \leq K\Delta t^{\frac{p}{2}}.$$

Proof. For any $T_1 \in [0, T]$, by Hölder's and BDG inequalities

$$\begin{aligned}
&\mathbb{E}\left[\sup_{0\leq t\leq T_1}\left|\hat{X}(t\wedge\theta_m) - x(t\wedge\theta_m)\right|^p\right] \\
&\leq 2^{p-1}\Bigg(T^{p-1}\mathbb{E}\int_0^{T_1\wedge\theta_m}\left[f(X_{\eta(s)}) - f(x(s))\right]^p ds \\
&+ C(p)\mathbb{E}\int_0^{T_1\wedge\theta_m}\left[g(X_{\eta(s)}) - g(x(s))\right]^p ds\Bigg),
\end{aligned}$$

where $C(p)$ is a constant. Let $s \in [0, T_1\wedge\theta_m)$. Then, the local Lipschitz condition on f and g implies that

$$\left|f(X_{\eta(s)}) - f(x(s))\right|^p + \left|g(X_{\eta(s)}) - g(x(s))\right|^p \leq C(m,p)\left|X_{\eta(s)} - x(s)\right|^p.$$

Hence

$$
\begin{aligned}
&\mathbb{E}\left[\sup_{0\le t\le T_1}\left|\hat{X}(t\wedge\theta_m)-x(t\wedge\theta_m)\right|^p\right]\\
&\le 2^{(p-1)}C(m,p)\Bigg(T^{p-1}\mathbb{E}\int_0^{T_1\wedge\theta_m}\left|X_{\eta(s)}-x(s)\right|^p ds\\
&+C(p)\mathbb{E}\int_0^{T_1\wedge\theta_m}\left|X_{\eta(s)}-x(s)\right|^p ds\Bigg)\\
&\le 4^{(p-1)}C(m,p)\Bigg(T^{p-1}\mathbb{E}\int_0^{T_1\wedge\theta_m}\left|\hat{X}(s)-x(s)\right|^p+\left|X_\eta(s)-\hat{X}(s)\right|^p ds\\
&+C(p)\mathbb{E}\int_0^{T_1\wedge\theta_m}[\left|\hat{X}(s)-x(s)\right|^p+\left|X_{\eta(s)}-\hat{X}(s)\right|^p]ds\Bigg)\\
&\le 4^{(p-1)}C(m,p)(T^{p-1}+C(p))\mathbb{E}\int_0^{T_1}\left|\hat{X}(s\wedge\theta_m)-x(s\wedge\theta_m)\right|^p ds\\
&+4^{(p-1)}C(m,p)(T^{p-1}+C(p))\mathbb{E}\int_0^{T_1\wedge\theta_m}\left|X_{\eta(s)}-\hat{X}(s)\right|^p ds.
\end{aligned}
$$

By the same reasoning which gave us (3.27), we can deduce that

$$\mathbb{E}\int_0^{T_1\wedge\theta_m}\left|X_\eta(s)-\hat{X}(s)\right|^p ds\le C(m,T,p)\Delta t^{\frac{p}{2}}.$$

Hence

$$
\begin{aligned}
&\mathbb{E}\left[\sup_{0\le t\le T_1}\left|\hat{X}(t\wedge\theta_m)-x(t\wedge\theta_m)\right|^p\right]\\
&\le 4^{(p-1)}C(m,p)(T^{p-1}+C(p))\\
&\times\left[C(m,T,p)\Delta t^{\frac{p}{2}}+\int_0^{T_1}\mathbb{E}\left[\sup_{0\le t\le s}\left|\hat{X}(t\wedge\theta_m)-x(t\wedge\theta_m)\right|^p\right]ds\right].
\end{aligned}
$$

By Gronwall's inequality

$$\mathbb{E}\left[\sup_{0\le t\le T}|X(t\wedge\theta_m)-x(t\wedge\theta_m)|^p\right]\le C_2\Delta t^{\frac{p}{2}}e^{C_1T},$$

where $C_1=4^{(p-1)}C(m,p)(T^{p-1}+C(p))$ and $C_2=C_1C(m,T,p)$. □

Now we are ready to prove the strong convergence theorem.

Theorem 3.4.4. *Under Assumptions 3.2.1, 3.3.2 and 3.3.5 for any given $T > 0$ and $s \in [1, 2)$, we have*

$$\lim_{\Delta t \to 0} \mathbb{E} \left| X(T) - x(T) \right|^s = 0. \tag{3.29}$$

Proof. Let

$$e(T) = X(T) - x(T).$$

Applying Young's inequality

$$x^s y \le \frac{\delta s}{2} x^2 + \frac{2-s}{2\delta^{\frac{s}{2-s}}} y^{\frac{2}{2-s}}, \quad \forall x, y, \delta > 0,$$

leads us to

$$\begin{aligned} \mathbb{E} \left| e(T) \right|^s &= \mathbb{E} \left[\left| e(T) \right|^s \mathbf{1}_{\{\tau_m > T, \vartheta_m > T\}} \right] + \mathbb{E} \left[\left| e(T) \right|^s \mathbf{1}_{\{\tau_m \le T \quad or \quad \vartheta_m \le T\}} \right] \\ &\le 2^{s-1} \Bigg[\mathbb{E} [\left| \hat{X}(T) - x(T) \right|^s \mathbf{1}_{\{\tau_m > T, \vartheta_m > T\}}] \\ &+ \mathbb{E} [\left| X(T) - \hat{X}(T) \right|^s \mathbf{1}_{\{\tau_m > T, \vartheta_m > T\}}] \Bigg] \\ &+ \frac{\delta s}{2} \mathbb{E} \left[\left| e(T) \right|^2 \right] + \frac{2-s}{2\delta^{\frac{s}{2-s}}} \mathbb{P}(\tau_m \le T \quad \text{or} \quad \vartheta_m \le T). \end{aligned} \tag{3.30}$$

To complete the proof we need to estimate the expressions on the right hand side of this inequality. First, let us observe that by Lemma 3.4.1 we obtain

$$\mathbb{E} [\left| X(T) - \hat{X}(T) \right|^s \mathbf{1}_{\{\tau_m > T, \vartheta_m > T\}}] \le C(m, s, T) \Delta t^s.$$

Given any $\epsilon > 0$, by Hölder's inequality and Theorems 3.2.2 and 3.3.6 , we choose δ such that

$$\frac{\delta s}{2} \mathbb{E} \left[\left| e(T) \right|^2 \right] \le 4 \frac{\delta s}{2} \mathbb{E} \left[\left| x(T) \right|^2 + \left| X(T) \right|^2 \right] \le \frac{\epsilon}{3}.$$

Now by (3.5) there exists N_0 such that for $m \ge N_0$

$$\frac{2-s}{2\delta^{\frac{s}{2-s}}} \mathbb{P}(\tau_m \le T) \le \frac{\epsilon}{3},$$

and finally by Lemma 3.4.3 and Theorem 3.4.2 we choose Δt sufficiently small, such that

$$2^{s-1}\left[\mathbb{E}[\left|\hat{X}(T)-x(T)\right|^s \mathbf{1}_{\{\tau_m>T,\vartheta_m>T\}}]+\mathbb{E}[\left|X(T)-\hat{X}(T)\right|^s \mathbf{1}_{\{\tau_m>T,\vartheta_m>T\}}]\right]$$
$$+\frac{2-s}{2\delta^{\frac{s}{2-s}}}\mathbb{P}(\vartheta_m\le T)\le\frac{\epsilon}{3},$$

which completes the proof. □

3.4.3 θ-Euler-Maryuama Scheme

In the stability analysis which will be presented in Section 3.5 we consider a θ-Euler-Maryuama scheme of the following form

$$X_{t_{k+1}}=X_{t_k}+\theta f(X_{t_{k+1}})\Delta t+(1-\theta)f(X_{t_k})\Delta t+g(X_{t_k})\Delta w_{t_k}. \tag{3.31}$$

For $\Delta t<(\max\{L,4\beta\}\theta)^{-1}$ the scheme is well defined. What is more, all the previous results hold once we replace condition (3.3) in Assumption 3.2.1 by the following one

$$\langle x,f(x)\rangle+\frac{1}{2}\left|g(x)\right|^2$$
$$+\frac{(1-2\theta)}{2}\left|f(x)\right|^2\Delta t\le\alpha+\beta\left|x\right|^2\quad\forall x\in\mathbb{R}^n,\quad\forall\Delta t\in(0,(\max\{L,4\beta\}\theta)^{-1}].$$

Clearly for $\theta\ge 0.5$ the above condition does not add any additional restrictions on the coefficients of SDEs (3.1).

Since Theorem 3.4.4 covers highly nonlinear SDEs it might be computationally expensive to find the inverse function F^{-1} to function F defined in the following way

$$F(x)=x-\theta f(x)\Delta t. \tag{3.32}$$

In this case, we suggest splitting the drift coefficient in SDEs (3.1) into the sum of two functions $f(x)=f_1(x)+f_2(x)$. Due to linearity of the inner product this does not affect any results for SDEs (3.1), but allows us to introduce partial implicitness in the numerical scheme. This partially implicit θ-Euler-Maruyama scheme has the following form

$$X_{t_{k+1}}=X_{t_k}+\theta f_1(X_{t_{k+1}})\Delta t+(1-\theta)f_1(X_{t_k})\Delta t+f_2(X_{t_k})\Delta t+g(X_{t_k})\Delta w_{t_k}.$$

Again, all results from previous sections hold, once we replace condition (3.3) in Assumption 3.2.1 by the following one

$$\begin{aligned}&\langle x, f(x)\rangle + \frac{1}{2}|g(x)|^2 + \big[(1-\theta)\langle f_1(x), f_2(x)\rangle \\ &+ \frac{1}{2}|f_2(x)|^2 + \frac{1}{2}(1-2\theta)|f_1(x)|^2\big]\Delta t \\ &\leq \alpha + \beta|x|^2 \quad \forall x \in \mathbb{R}^n, \quad \forall \Delta t \in (0, (\max\{L, 4\beta\}\theta)^{-1}].\end{aligned}$$

3.5 Stability Analysis

In this section we examine the global almost sure stability of the θ-EM scheme (3.31). The stability conditions we derive are related to mean-square stability as we are interested in results that do not put severe restrictions on the time step. First, we give some preliminary analysis for SDEs (3.1). We give conditions on the coefficients of the SDEs (3.1) that are sufficient for a globally almost sure stable system. Later we prove that the θ-EM scheme (3.31) reproduces this asymptotic behaviour very well.

3.5.1 Continuous Case

In (Shen, Luo, and Mao 2006), the authors proved a very general Stochastic LaSalle Theorem. Here we present a simplified version of their theorem, with a fixed Lyapunov function $V(x) = |x|^2$.

Theorem 3.5.1 (Mao et al. (Shen, Luo, and Mao 2006))**.** *Let Assumption 3.2.1 hold. Assume further that there exists a function $z \in C(\mathbb{R}^n; \mathbb{R}_+)$ such that*

$$\langle x, f(x)\rangle + \frac{1}{2}|g(x)|^2 \leq -z(x) \tag{3.33}$$

for all $x \in \mathbb{R}^n$. We then have the following assertions:

- $D_z := \{x \in \mathbb{R}^n : z(x) = 0\} \neq \emptyset$.
- *For any $x_0 \in \mathbb{R}^n$, the solution $x(t, x_0)$ of (3.1) has the properties that*

$$\int_0^\infty \mathbb{E}[z(x(t, x_0))]dt < \infty \quad a.s.,$$

$$\limsup_{t\to\infty} |x(t,x_0)|^2 < \infty \quad a.s. \quad and$$
$$\lim_{t\to\infty} d(x(t,x_0); D_z) = 0 \quad a.s.,$$

where $d(x;A) = \inf_{y\in A} |x-y|$.

What is more, if the following condition holds

$$z(x) = 0 \quad iff \quad x = 0,$$

then

$$\lim_{t\to\infty} x(t,x_0) = 0 \quad a.s. \quad \forall\ x_0 \in \mathbb{R}^n.$$

3.5.2 Discrete Case

Recently, it was shown by Appleby et al. (Appleby, Kelly, Mao, and Rodkina 2010), that the classical Euler-Maruyama scheme may fail to preserve almost sure stability of a test equation. In fact, they considered the following equation

$$X_{k+1} = X_k - \beta X_k |X_k|^p \Delta t + \sigma_k |X_k|^\rho \sqrt{\Delta t}\xi_{n+1}. \tag{3.34}$$

They concluded that for arbitrary initial data X_0 under Assumption $p+1 > 2\rho$, the solution to the equation (3.34) explodes to infinity with positive probability. This obviously violates the almost sure stability property. On the other hand, Appleby et al. (Appleby, Mao, and Rodkina 2008) showed that the continuous counterpart of the equation (3.34) does converge to zero almost surely. This motivates stability analysis for numerical approximations for highly non-linear SDEs of this type.

3.5.3 Almost Sure Stability

We begin this section with the following Lemma.

Lemma 3.5.2. *Let* $Z = \{Z_n\}_{n\in\mathbb{N}}$ *be a nonnegative stochastic process with Doob decomposition* $Z_n = Z_0 + A_n^1 - A_n^2 + M_n$, *where* $A^1 = \{A_n^1\}_{n\in\mathbb{N}}$ *and* $A^1 = \{A_n^1\}_{n\in\mathbb{N}}$ *are a.s. nondecreasing, predictable processes with* $A_0^1 = A_0^2 = 0$, *and*

$M = \{M_n\}_{n\in\mathbb{N}}$ is local $\{\mathcal{F}_n\}_{n\in\mathbb{N}}$-martingale with $M_0 = 0$. Then

$$\left\{\omega : \lim_{n\to\infty} A_n^1 < \infty\right\} \subseteq \left\{\omega : \lim_{n\to\infty} A_n^2 < \infty\right\} \cap \left\{\lim_{n\to\infty} Z_n \textit{exists and its finite}\right\} \quad a.s.$$

The original lemma can be found in Shiryaev (Liptser and Shiryayev 1989). This result combines Doob's Decomposition and Martingale Convergence Theorems.
Since we use a Lyapunov function $V(x) = |x|^2$, our results extend mean-square stability for linear systems, (Higham 2001; Higham 2000), to a highly nonlinear setting.

Theorem 3.5.3. *Let Assumptions 3.2.1, 3.3.2 and 3.3.5 hold. Assume that there exists a function $z \in C(\mathbb{R}^n; \mathbb{R}_+)$ such that for all $x \in \mathbb{R}^n$ and for all $\Delta t \in (0, (\max\{L, 2\beta\}\theta)^{-1})$,*

$$\langle x, f(x)\rangle + \frac{1}{2}|g(x)|^2 + \frac{(1-2\theta)}{2} f^2(x)\Delta t \le -z(x). \tag{3.35}$$

Then the θ-EM solution defined by (3.31), obeys

$$\limsup_{k\to\infty} |X(t_k)|^2 < \infty \quad a.s., \tag{3.36}$$

$$\lim_{k\to\infty} z\left(X(t_k)\right) = 0 \;\; a.s. \tag{3.37}$$

and

$$\sum_{k=0}^{\infty} \mathbb{E}[z(X_{t_k})]\Delta t < \infty. \tag{3.38}$$

If additionally $z(x) = 0$ iff $x = 0$

$$\lim_{k\to\infty} X(t_k) = 0 \;\; a.s.$$

Proof. By definition of operator F in (3.32), we can represent the θ-EM scheme (3.31) as

$$F(X_{t_{k+1}}) = F(X(t_k)) + f(X(t_k))\Delta t + g(X(t_k))\Delta w_{k+1}.$$

Consequently

$$\begin{aligned}\left|F(X_{t_{k+1}})\right|^2 &= |F(X(t_k))|^2 + |f(X(t_k))|^2 \Delta t^2 + |g(X(t_k))|^2 \Delta t \\ &+ 2\langle F(X(t_k)), f(X(t_k))\rangle \Delta t + \Delta M_{k+1} \\ &= |F(X(t_k)|^2 \\ &+ \left(2\langle X(t_k), f(X(t_k))\rangle + |g(X(t_k))|^2\right) \Delta t \\ &+ (1-2\theta)\, |f(X(t_k))|^2 \Delta t^2 + \Delta M_{k+1},\end{aligned} \tag{3.39}$$

where

$$\begin{aligned}\Delta M_{k+1} &= |g(X(t_k))|^2 (\Delta w_{k+1}^2 - \Delta t) + 2\langle F(X(t_k)), g(X(t_k))\rangle \Delta w_{k+1} \\ &+ 2\langle f(X(t_k))\Delta t, g(X(t_k))\rangle \Delta w_{k+1},\end{aligned}$$

so that $\sum_{k=1}^{N} \Delta M_k$ is a local martingale due to Assumption 3.3.5 and Lemma 3.3.4. Hence, we have obtained the decomposition required to apply Lemma 3.5.2, that is

$$\left|F(X_{t_{k+1}})\right|^2 = |F(X_{t_k})|^2 - A(X_{t_k})\Delta t + \Delta M_{k+1},$$

where

$$\begin{aligned}A(X_{t_k}) = -\Bigg(&\left(2\langle X(t_k), f(X(t_k))\rangle + |g(X(t_k))|^2\right) \\ &+ (1-2\theta)\, |f(X(t_k))|^2 \Delta t\Bigg).\end{aligned} \tag{3.40}$$

Therefore

$$\left|F(X_{t_{N+1}})\right|^2 = |F(X_{t_0})|^2 - \sum_{k=0}^{N} A(X_{t_k})\Delta t + \sum_{k=0}^{N} \Delta M_k.$$

Now we are in a position to apply Lemma 3.5.2 to get

$$\lim_{k\to\infty} |F(X_{t_k})|^2 < \infty, \tag{3.41}$$

from where, by Assumption 3.2.1, it is easy to show that

$$\limsup_{k\to\infty} |X_{t_k}|^2 < \infty \quad \text{a.s.}.$$

By Lemma 3.5.2,

$$\sum_{k=0}^{\infty} z(X_{t_k})\Delta t \leq \sum_{k=0}^{\infty} A(X_{t_k})\Delta t < \infty \quad \text{a.s.},$$

which implies

$$\lim_{k\to\infty} z(X_{t_k}) = 0 \quad \text{a.s.}.$$

Summing up the both sides of (3.39) gives us

$$\sum_{k=0}^{N} z(X_{t_k})\Delta t \leq |F(X_{t_0})|^2 + \sum_{k=0}^{N} \Delta M_{k+1},$$

taking expectation of both sides of above inequality and application of Fatou Lemma proves (3.37) and completes the proof of the theorem. □

Chapter 4

Rate of Convergence

> A mathematician is a device for turning coffee into theorems.
>
> Paul Erdos

In this chapter we extend the analysis from Chapter 3. We demonstrate that once we replace the monotone condition by a dissipative-type condition on the drift and the polynomial condition on the diffusion terms we are able to prove stronger versions of Theorem 3.4.4. These stronger assumptions will enable us to generalize the current theory of strong convergence rates for the backward Euler-Maruyama scheme for super-linear SDEs. First we obtain an upper-bound for the pth moments of the solution to (3.1), which is uniform in time. Later on we show that BEM has an ability to reproduce this feature. Then by imposing a stronger polynomial condition we reveal the rate of convergence for BEM, which equals a half. We conclude the chapter with a generalization of the Fundamental Theorem by Milstein (Milshtein 1975; Milstein and Tretyakov 2004). Indeed, we show that with the same assumptions required to prove our strong convergence theorem we can extend the Fundamental Theorem to a non-Lipschitz case.

4.1 Preliminary Analysis

Similarly, like in Assumption 3.2.1, we require the coefficients f and g to be local *Lipschitz continuous.* However, in place of the monotone condition we introduce the following *dissipative-type* and *polynomial growth* conditions on the drift and diffusion coefficients, respectively.

Assumption 4.1.1. *Both functions f and g satisfy the following conditions:*
Local Lipschitz condition. For each integer $m \geq 1$, there is a positive constant K_m such that

$$|f(x) - f(y)| + |g(x) - g(y)| \leq K_m |x - y|, \tag{4.1}$$

for those $x, y \in \mathbb{R}^n$ with $|x| \vee |y| \leq m$.
Dissipative condition. For some $\rho \geq 1$, and $r \in \mathbb{N}$, $r \geq 1$, there exist positive constants $\alpha_0, \alpha_1, \beta, \beta_0, \beta_1 > 0$, such that

$$-\beta_1 |x|^{r+1} - \beta_0 \leq \langle x, f(x) \rangle \leq \alpha_0 - \alpha_1 |x|^{r+1}, \tag{4.2}$$

$$|g(x)| \leq \beta |x|^{\rho}, \tag{4.3}$$

for all $x \in \mathbb{R}^n$.

Further on in this thesis, we demonstrate that Assumption 4.1.1 covers both a wide family of SDEs applied in Mathematical Finance and Bio-mathematics. In order to proceed with our analysis, we make an assumption about values of the parameters. As will become clear from the proofs, this type of assumption allows us to control the potential growth coming from the diffusion term using the dissipative nature of the drift.

Assumption 4.1.2. *The parameters in Assumption 4.1.1 obey*

$$r + 1 > 2\rho.$$

From now on, without loss of generality, we assume that through the rest of this chapter $p \in \mathbb{N}$ is always an even number. Clearly, Assumptions 4.1.1 and 4.1.2 imply Assumption 3.2.1. Therefore, the statement of Theorem 3.2.2 holds. However, with these stronger assumptions we are able to obtain sharper bounds.

Theorem 4.1.3. *Let Assumptions 4.1.1 and 4.1.2 hold. Then for any given initial value $x(0) = x_0 \in \mathbb{R}^n$, there exists a unique, global solution $x(t)$ to the equation (3.1) for $t \geq 0$. Moreover, for any $T \geq 0$ and for every $p \geq 2$ there* *exists a constant $K = K(r,\rho)$ independent of T, such that*

$$\mathbb{E}\,|x(t)|^p \leq e^{-T}\,|x_0|^p + K, \quad \forall t \in [0,T], \tag{4.4}$$

and for every $m > m_0$, where $|x_0| < m_0$,

$$\mathbb{P}(\tau_m \leq T) \leq |m|^{-p}\Big(|x_0|^p + KT\Big), \tag{4.5}$$

where

$$\tau_m = \inf\{t \geq 0 : \quad |x(t)| > m\}. \tag{4.6}$$

Proof. Existence and uniqueness follow from Theorem 3.2.2. Therefore, we only need to prove that (4.4) and (4.5) hold. For any $p \geq 2$, applying the Itô formula to the function $V(x,t) = e^t\,|x|^p$, we compute the diffusion operator

$$LV(x,t) \leq e^t\Big(\,|x|^p + p\,|x|^{p-2}\,\langle x, f(x)\rangle + \frac{1}{2}p(p-1)\,|x|^{p-2}\,\big|g^2(x)\big|\,\Big). \tag{4.7}$$

By Assumptions 4.1.1 and 4.1.2, there exists a constant $C > 0$ such that

$$|x|^p + p\,|x|^{p-2}\,(\alpha_0 - \alpha_1\,|x|^{r+1}) + \frac{1}{2}\beta p(p-1)\,|x|^{p-2}\,|x|^{2\rho} \leq C, \tag{4.8}$$

and as consequence

$$LV(x,t) \leq Ce^t.$$

Therefore

$$\mathbb{E}\Big[e^{t\wedge\tau_m}\,|x(t\wedge\tau_m)|^p\Big] \leq |x_0|^p + Ce^t.$$

Next, letting $m \to \infty$ and applying Fatou's lemma, we obtain

$$\mathbb{E}\,|x(t)|^p \leq e^{-t}\,|x_0|^p + C.$$

Now by similar analysis with a function $V(x,t) = |x|^p$ we obtain

$$\mathbb{P}(\tau_m \leq T)[|m|^p] \leq |x_0|^p + CT.$$

This implies that $\lim_{m\to\infty}\mathbb{P}(\tau_m \leq T) = 0$ as desired and finishes the proof of Theorem 4.1.3. □

Now we are also able to show that pth-moments are bounded uniformly within the whole time interval $[0, T]$.

Lemma 4.1.4. *Under Assumptions 4.1.1 and 4.1.2, for any $T \geq 0$, we have*

$$\mathbb{E}(\sup_{0 \leq t \leq T} |x(t)|^p) < \infty \quad p \geq 2. \tag{4.9}$$

Proof. By the Itô formula applied to a function $V(x) = |x|^p$ and Theorem 4.1.3, we can show that

$$\begin{aligned}\mathbb{E}[\sup_{0 \leq t \leq T} |x(t)|^p] \leq |x_0|^p + CT \\ + \mathbb{E}\Big[\sup_{0 \leq t \leq T} \int_0^t p\,|x(u)|^{p-2}\, x^T(u) g(x(u)) dw(u)\Big],\end{aligned} \tag{4.10}$$

where C is a positive constant. By Burkholder-Davis-Gundy's and Jensen's inequality, we can show that

$$\begin{aligned}&\mathbb{E}\Big[\sup_{0 \leq t \leq T} \int_0^t p\,|x(u)|^{p-2}\, x^T(u) g(x(u)) dw(u)\Big] \\ &\leq C\mathbb{E}\Big(\int_0^T \left[p\,|x(u)|^{p-2}\, x^T(u) g(x(u))\right]^2 du\Big)^{\frac{1}{2}} \\ &\leq C\Big(\int_0^T \mathbb{E}\left[p\,|x(u)|^{p-2}\, x^T(u) g(x(u))\right]^2 du\Big)^{\frac{1}{2}},\end{aligned} \tag{4.11}$$

where C stands for a constant which may vary from line to line. By Theorem 4.1.3, the conclusion follows. □

4.2 Backward Euler-Maruyama Scheme

We work here with backward Euler-Maruyama scheme (2.2). From Lemma 3.3.3 we know that it possesses a unique solution as long as $\Delta t < \frac{1}{L}$. In this section we prove a much stronger version of the Theorem 3.3.6, namely we prove uniform boundedness of the pth-moments of (2.2).

4.2.1 Moment Estimates

Before we state and prove the boundedness of moments for BEM (2.2) we prove the following lemma, which gives us an estimate required to prove Theorem 4.2.2.

Lemma 4.2.1. *Under Assumption 4.1.1, for any integer $s \geq 1$ there exist positive constants K_1 and K_2 such that*

$$(-1)^{s+1}(\langle x, f(x)\rangle)^s \leq K_1 - K_2 \left|x\right|^{s(r+1)}. \tag{4.12}$$

Proof. We prove the lemma by induction. We claim that the following inequalities hold

$$-\tilde{C}_1 \left|x\right|^{s(r+1)} - \tilde{C}_0 \leq (-1)^{s+1}\langle x, f(x)\rangle^s \leq C_0 - C_1 \left|x\right|^{s(r+1)}, \tag{4.13}$$

where C_0, C_1, $\tilde{C}_0$ and $\tilde{C}_1$ are positive constants, which may depend on α_0, α_1, β_0, β_1, but not on x. From Assumption 4.1.1 the statement (4.13) holds for $s = 1$. Now, given that it holds at level s we show that (4.13) holds true for $s+1$. From Assumption 4.1.1 suppose first that $\langle x, f(x)\rangle \geq 0$

$$\begin{aligned}&(-1)\langle x, f(x)\rangle(-\tilde{C}_1 \left|x\right|^{s(r+1)} - \tilde{C}_0)\\ &\geq (-1)^{s+2}\langle x, f(x)\rangle^{s+1} \geq (C_0 - C_1 \left|x\right|^{s(r+1)})(-1)\langle x, f(x)\rangle.\end{aligned}$$

In the case $C_0 - C_1 \left|x\right|^{s(r+1)} > 0$

$$\begin{aligned}&(-1)(\alpha_0 - \alpha_1 \left|x\right|^{r+1})(-\tilde{C}_1 \left|x\right|^{s(r+1)} - \tilde{C}_0)\\ &\geq (-1)^{s+2}\langle x, f(x)\rangle^{s+1} \geq (C_0 - C_1 \left|x\right|^{s(r+1)})(-1)(\alpha_0 - \alpha_1 \left|x\right|^{r+1}),\end{aligned}$$

and in the case $C_0 - C_1 \left|x\right|^{s(r+1)} \leq 0$

$$\begin{aligned}&(-1)(\alpha_0 - \alpha_1 \left|x\right|^{r+1})(-\tilde{C}_1 \left|x\right|^{s(r+1)} - \tilde{C}_0)\\ &\geq (-1)^{s+2}\langle x, f(x)\rangle^{s+1} \geq (C_0 - C_1 \left|x\right|^{s(r+1)})(\beta_1 \left|x\right|^{r+1} + \beta_0).\end{aligned}$$

Alternatively, suppose $\langle x, f(x)\rangle < 0$. By Assumption 4.1.1

$$\begin{aligned}&(-1)\langle x, f(x)\rangle(-\tilde{C}_1 |x|^{s(r+1)} - \tilde{C}_0)\\ &\leq (-1)^{s+2}\langle x, f(x)\rangle^{s+1} \leq (C_0 - C_1 |x|^{s(r+1)})(-1)\langle x, f(x)\rangle.\end{aligned}$$

In the case $C_0 - C_1 |x|^{s(r+1)} > 0$ we obtain

$$\begin{aligned}&(-1)(-\beta_0 - \beta_1 |x|^{r+1})(-\tilde{C}_1 |x|^{s(r+1)} - \tilde{C}_0)\\ &\leq (-1)^{s+2}\langle x, f(x)\rangle^{s+1} \leq (C_0 - C_1 |x|^{s(r+1)})(\beta_0 + \beta_1 |x|^{r+1}),\end{aligned}$$

and in the case $C_0 - C_1 |x|^{s(r+1)} \leq 0$

$$\begin{aligned}&(-1)(-\beta_0 - \beta_1 |x|^{r+1})(-\tilde{C}_1 |x|^{s(r+1)} - \tilde{C}_0)\\ &\leq (-1)^{s+2}\langle x, f(x)\rangle^{s+1} \leq (C_0 - C_1 |x|^{s(r+1)})(-1)(\alpha_0 - \alpha_1 |x|^{r+1}).\end{aligned}$$

Now we observe that there exist positive constants $\hat{C}_1, \hat{C}_2$ and $\hat{C}_3$, such that

$$\begin{aligned}&(-1)(\alpha_0 - \alpha_1 |x|^{r+1})(-\tilde{C}_1 |x|^{s(r+1)} - \tilde{C}_0)\\ &\leq \hat{C}_1 - \frac{\alpha_1 \tilde{C}_1}{2} |x|^{(s+1)(r+1)},\end{aligned}$$

$$\begin{aligned}&(C_0 - C_1 |x|^{s(r+1)})(\beta_0 + \beta_1 |x|^{r+1})\\ &\leq \hat{C}_2 - \frac{C_1 \beta_1}{2} |x|^{(s+1)(r+1)},\end{aligned}$$

$$\begin{aligned}&(C_0 - C_1 |x|^{s(r+1)})(-1)(\alpha_0 - \alpha_1 |x|^{r+1})\\ &\leq \hat{C}_3 - \frac{C_1 \alpha_1}{2} |x|^{(s+1)(r+1)}.\end{aligned}$$

Now we define $C_0 = \max\{\hat{C}_1, \hat{C}_2, \hat{C}_3\}$ and $C_1 = \min\{\frac{\alpha_1 \tilde{C}_1}{2}, \frac{C_1\beta_1}{2}, \frac{C_1\alpha_1}{2}\}$, and the proof of the upper bound in (4.13) is complete. Similarly, we can find the lower bound in (4.13). The proof by induction of Lemma 4.2.1 is complete. □

Theorem 4.2.2. *Let Assumptions 4.1.1, 4.1.2, 3.3.2 and 3.3.5 hold. Let* $\Delta t^* \in$

$(0, 1/L)$ *be sufficiently small so that whenever* $\Delta t \leq \Delta t^*$,

$$1 \leq \frac{\left[e^{\Delta t} - 1\right]}{\Delta t} \leq 2 \quad \text{and} \quad e^{\Delta t} \leq 2. \tag{4.14}$$

Then for any integer $p \geq 2$

$$\sup_{\Delta t \leq \Delta t^*} \sup_{k \geq 0} \mathbb{E} \left|X_{t_k}\right|^p < \infty.$$

Proof. We begin with the following inequality

$$\begin{aligned} \left|X_{t_{k+1}}\right|^2 - \left|X_{t_k}\right|^2 \leq{}& 2\langle X_{t_{k+1}}, f(X_{t_{k+1}})\rangle \Delta t \\ &+ 2\langle X_{t_k}, g(X_{t_k})\rangle \Delta W_{t_k} + \left|g(X_{t_k})\right|^2 \left|\Delta W_{t_k}\right|^2. \end{aligned} \tag{4.15}$$

Recalling Assumption 4.1.1, we note that

$$\begin{aligned} 0 &< \left|X_{t_{k+1}}\right|^2 - 2\langle X_{t_{k+1}}, f(X_{t_{k+1}})\rangle \Delta t + 2\alpha_0 \Delta t \\ &\leq \left|X_{t_k}\right|^2 + 2\langle X_{t_k}, g(X_{t_k})\rangle \Delta W_{t_k} + \left|g(X_{t_k})\right|^2 \left|\Delta W_{t_k}\right|^2 + 2\alpha_0 \Delta t. \end{aligned}$$

Raising the both sides of above inequality to the power $\frac{p}{2}$ leads us to

$$\begin{aligned} &\left(\left|X_{t_{k+1}}\right|^2 - 2\langle X_{t_{k+1}}, f(X_{t_{k+1}}) \Delta t\rangle + 2\alpha_0 \Delta t\right)^{p/2} \\ &\leq \left(\left|X_{t_k}\right|^2 + 2\langle X_{t_k}, g(X_{t_k}) \Delta W_{t_k}\rangle + \left|g(X_{t_k})\right|^2 \left|\Delta W_{t_k}\right|^2 + 2\alpha_0 \Delta t\right)^{p/2}. \end{aligned}$$

Now by the binomial theorem we have

$$\begin{aligned} &\sum_{l=0}^{p/2} \sum_{s=0}^{l} \binom{l}{s} \binom{p/2}{l} (-1)^s \left|X_{t_{k+1}}\right|^{2(\frac{p}{2}-l)} \left(2\langle X_{t_{k+1}}, f(X_{t_{k+1}}) \Delta t\rangle\right)^s (2\alpha_0 \Delta t)^{l-s} \\ &\leq \sum_{l=0}^{p/2} \sum_{s=0}^{l} \sum_{i=0}^{l-s} \binom{l}{s} \binom{p/2}{l} \binom{l-s}{i} \left|X_{t_k}\right|^{2(\frac{p}{2}-l)} \\ &\times \left(2\langle X_{t_k}, g(X_{t_k}) \Delta W_{t_k}\rangle\right)^s \left(\left|g(X_{t_k})\right|^2 \left|\Delta W_{t_k}\right|^2\right)^{l-s-i} (2\alpha_0 \Delta t)^i. \end{aligned}$$

Hence

$$
\begin{aligned}
&\left|X_{t_{k+1}}\right|^p - |X_{t_k}|^p \\
&\leq \sum_{l=1}^{p/2}\sum_{s=0}^{l}\binom{l}{s}\binom{p/2}{l}(-1)^{s+1}\left|X_{t_{k+1}}\right|^{2(\frac{p}{2}-l)}\left(2\langle X_{t_{k+1}}, f(X_{t_{k+1}})\Delta t\rangle\right)^s(2\alpha_0\Delta t)^{l-s} \\
&+\sum_{l=1}^{p/2}\sum_{s=0}^{l}\sum_{i=0}^{l-s}\binom{l}{s}\binom{p/2}{l}\binom{l-s}{i}|X_{t_k}|^{2(\frac{p}{2}-l)} \\
&\times(2\langle X_{t_k}, g(X_{t_k})\Delta W_{t_k}\rangle)^s\left(|g(X_{t_k})|^2|\Delta W_{t_k}|^2\right)^{l-s-i}(2\alpha_0\Delta t)^i \\
&=\sum_{l=1}^{p/2}\sum_{s=0}^{l}\binom{l}{s}\binom{p/2}{l}(-1)^{s+1}\left|X_{t_{k+1}}\right|^{2(\frac{p}{2}-l)}\left(2\langle X_{t_{k+1}}, f(X_{t_{k+1}})\Delta t\rangle\right)^s(2\alpha_0\Delta t)^{l-s} \\
&+\sum_{l=1}^{p/2}\sum_{\substack{s=0\\ s-even}}^{l}\sum_{i=0}^{l-s}\binom{l}{s}\binom{p/2}{l}\binom{l-s}{i}|X_{t_k}|^{2(\frac{p}{2}-l)} \\
&\times(2\langle X_{t_k}, g(X_{t_k})\Delta W_{t_k}\rangle)^s\left(|g(X_{t_k})|^2|\Delta W_{t_k}|^2\right)^{l-s-i}(2\alpha_0\Delta t)^i \\
&+\sum_{l=1}^{p/2}\sum_{\substack{s=0\\ s-odd}}^{l}\sum_{i=0}^{l-s}\binom{l}{s}\binom{p/2}{l}\binom{l-s}{i}|X_{t_k}|^{2(\frac{p}{2}-l)} \\
&\times(2\langle X_{t_k}, g(X_{t_k})\Delta W_{t_k}\rangle)^s\left(|g(X_{t_k})|^2|\Delta W_{t_k}|^2\right)^{l-s-i}(2\alpha_0\Delta t)^i .
\end{aligned}
$$

Note that

$$
\begin{aligned}
e^{t_{k+1}}\left|X_{t_{k+1}}\right|^p - e^{t_k}|X_{t_k}|^p &= e^{t_k}\left[\left|X_{t_{k+1}}\right|^p - |X_{t_k}|^p\right] + \left[e^{t_{k+1}} - e^{t_k}\right]\left|X_{t_{k+1}}\right|^p \\
&= e^{t_k}\left[\left|X_{t_{k+1}}\right|^p - |X_{t_k}|^p\right] + e^{t_k}\frac{\left[e^{\Delta t}-1\right]}{\Delta t}\Delta t\left|X_{t_{k+1}}\right|^p,
\end{aligned}
$$

and therefore

$$
e^{t_{k+1}}\left|X_{t_{k+1}}\right|^p - e^{t_k}|X_{t_k}|^p \leq e^{t_k}\left[\left|X_{t_{k+1}}\right|^p - |X_{t_k}|^p\right] + 2e^{t_k}\Delta t\left|X_{t_{k+1}}\right|^p .
$$

As a consequence we have

$$
\begin{aligned}
& e^{t_{k+1}} \left|X_{t_{k+1}}\right|^p - e^{t_k} \left|X_{t_k}\right|^p \\
& \leq e^{t_k} \Bigg(\frac{p}{2}(-1) \left|X_{t_{k+1}}\right|^{2(\frac{p}{2}-1)} (2\alpha_0 \Delta t) \\
& + \frac{p}{2} \left|X_{t_{k+1}}\right|^{2(\frac{p}{2}-1)} \left(2\langle X_{t_{k+1}}, f(X_{t_{k+1}})\Delta t\rangle\right) + 2\Delta t \left|X_{t_{k+}}\right|^p \\
& + \sum_{l=2}^{p/2} \sum_{s=0}^{l} \binom{l}{s} \binom{p/2}{l} (-1)^{s+1} \left|X_{t_{k+1}}\right|^{2(\frac{p}{2}-l)} \left(2\langle X_{t_{k+1}}, f(X_{t_{k+1}})\Delta t\rangle\right)^s (2\alpha_0 \Delta t)^{l-s} \\
& + \sum_{l=1}^{p/2} \sum_{\substack{s=0 \\ s-even}}^{l} \sum_{i=0}^{l-s} \binom{l}{s} \binom{p/2}{l} \binom{l-s}{i} \left|X_{t_k}\right|^{2(\frac{p}{2}-l)} \\
& \times \left(2\langle X_{t_k}, g(X_{t_k})\Delta W_{t_k}\rangle\right)^s \left(\left|g(X_{t_k})\right|^2 \left|\Delta W_{t_k}\right|^2\right)^{l-s-i} (2\alpha_0 \Delta t)^i \\
& + \sum_{l=1}^{p/2} \sum_{\substack{s=0 \\ s-odd}}^{l} \sum_{i=0}^{l-s} \binom{l}{s} \binom{p/2}{l} \binom{l-s}{i} \left|X_{t_k}\right|^{2(\frac{p}{2}-l)} \\
& \times \left(2\langle X_{t_k}, g(X_{t_k})\Delta W_{t_k}\rangle\right)^s \left(\left|g(X_{t_k})\right|^2 \left|\Delta W_{t_k}\right|^2\right)^{l-s-i} (2\alpha_0 \Delta t)^i \Bigg).
\end{aligned}
$$

For every sufficiently large integer m, we define the stopping time

$$\lambda_m = \inf\{k : |X_{t_k}| > m\}.$$

Let N be any nonnegative integer such that $N\Delta t \leq T$. Summing up the both sides of the inequality from $k = 0$ to $N \wedge \lambda_m$, we get

$$e^{t_{N\wedge\lambda_m+1}} \left|X_{t_{N\wedge\lambda_m+1}}\right|^p \leq \Bigg(|X_{t_0}|^p + \sum_{l=1}^{p/2} \sum_{s=0}^{l} \sum_{i=0}^{l-s} \binom{l}{s} \binom{p/2}{l} \binom{l-s}{i} |X_{t_0}|^{2(\frac{p}{2}-l)}$$

$$\times \left(2\langle X_{t_0}, g(X_{t_0})\Delta W_{t_0}\rangle\right)^s \left(|g(X_{t_0})|^2 |\Delta W_{t_0}|^2\right)^{l-s-i} (2\alpha_0 \Delta t)^i \Bigg)$$

$$+ \sum_{k=1}^{N\wedge\lambda_m+1} e^{t_{k-1}} \left(-\frac{p}{2} |X_{t_k}|^{2(\frac{p}{2}-1)} (2\alpha_0 \Delta t) \right)$$

$$+ \sum_{k=1}^{N\wedge\lambda_m+1} e^{t_{k-1}} \left(\frac{p}{2} |X_{t_k}|^{2(\frac{p}{2}-1)} 2\langle X_{t_k}, f(X_{t_k})\Delta t\rangle + 2\Delta t \, |X_{t_k}|^p \right)$$

$$+ \sum_{k=1}^{N\wedge\lambda_m+1} e^{t_{k-1}} \Bigg(\sum_{l=2}^{p/2} \sum_{s=0}^{l} \binom{l}{s} \binom{p/2}{l} (-1)^{s+1} |X_{t_k}|^{2(\frac{p}{2}-l)}$$

$$\times \left(2\langle X_{t_k}, f(X_{t_k})\Delta t\rangle\right)^s (2\alpha_0 \Delta t)^{l-s} \Bigg)$$

$$+ \sum_{k=1}^{N\wedge\lambda_m} e^{t_k} \Bigg(\sum_{l=1}^{p/2} \sum_{\substack{s=0 \\ s-even}}^{l} \sum_{i=0}^{l-s} \binom{l}{s} \binom{p/2}{l} \binom{l-s}{i} |X_{t_k}|^{2(\frac{p}{2}-l)}$$

$$\times \left(2\langle X_{t_k}, g(X_{t_k})\Delta W_{t_k}\rangle\right)^s \left(|g(X_{t_k})|^2 |\Delta W_{t_k}|^2\right)^{l-s-i} (2\alpha_0 \Delta t)^i \Bigg)$$

$$+ \sum_{k=1}^{N\wedge\lambda_m} e^{t_k} \Bigg(\sum_{l=1}^{p/2} \sum_{\substack{s=0 \\ s-odd}}^{l} \sum_{i=0}^{l-s} \binom{l}{s} \binom{p/2}{l} \binom{l-s}{i} |X_{t_k}|^{2(\frac{p}{2}-l)}$$

$$\times \left(2\langle X_{t_k}, g(X_{t_k})\Delta W_{t_k}\rangle\right)^s \left(|g(X_{t_k})|^2 |\Delta W_{t_k}|^2\right)^{l-s-i} (2\alpha_0 \Delta t)^i \Bigg).$$

By Lemma 3.3.4, taking expectation of the above inequality leads to

$$
\begin{aligned}
&\mathbb{E}e^{t_{N\wedge\lambda_m+1}}\left|X_{t_{N\wedge\lambda_m+1}}\right|^p \\
&\leq C_1 + \mathbb{E}\sum_{k=1}^{N} e^{t_{k-1}}\left(\frac{p}{2}\left|X_{t_k}\right|^{2(\frac{p}{2}-1)} 2\langle X_{t_k}, f(X_{t_k})\Delta t\rangle + 2\Delta t\left|X_{t_k}\right|^p\right)\mathbf{1}_{[0,\lambda_m]}(k) \\
&+ \mathbb{E}e^{t_{N\wedge\lambda_m}}\left(\frac{p}{2}\left|X_{t_k}\right|^{2(\frac{p}{2}-1)} 2\langle X_{t_{N\wedge\lambda_m+1}}, f(X_{t_{N\wedge\lambda_m+1}})\Delta t\rangle + 2\Delta t\left|X_{t_{N\wedge\lambda_m+1}}\right|^p\right) \\
&+ \mathbb{E}\sum_{k=1}^{N} e^{t_{k-1}}\left(\sum_{l=2}^{p/2}\sum_{s=0}^{l}\binom{l}{s}\binom{p/2}{l}(-1)^{s+1}\left|X_{t_k}\right|^{2(\frac{p}{2}-l)}\right. \\
&\left.\times\left(2\langle X_{t_k}, f(X_{t_k})\rangle\right)^s\left(2\alpha_0\right)^{l-s}\Delta t^l\right)\mathbf{1}_{[0,\lambda_m]}(k) \\
&+ \mathbb{E}e^{t_N}\left(\sum_{l=2}^{p/2}\sum_{s=0}^{l}\binom{l}{s}\binom{p/2}{l}(-1)^{s+1}\left|X_{t_{N\wedge\lambda_m+1}}\right|^{2(\frac{p}{2}-l)}\right. \\
&\left.\times\left(2\langle X_{t_{N\wedge\lambda_m+1}}, f(X_{t_{N\wedge\lambda_m+1}})\rangle\right)^s\left(2\alpha_0\right)^{l-s}\Delta t^l\right)\mathbf{1}_{[0,\lambda_m]}(k) \\
&+ \mathbb{E}\sum_{k=1}^{N} e^{t_k}\left(\sum_{l=1}^{p/2}\sum_{s=0}^{l/2}\sum_{i=0}^{l-s}\binom{l}{2s}\binom{p/2}{l}\binom{l-2s}{i}\left|X_{t_k}\right|^{2(\frac{p}{2}-l)}\right. \\
&\left.\times\left(2\langle X_{t_k}, g(X_{t_k})\rangle\right)^{2s}\left(\left|g(X_{t_k})\right|^2\right)^{l-2s-i}\left(2\alpha_0\right)^i\Delta t^{l-2s}\right)\mathbf{1}_{[0,\lambda_m]}(k),
\end{aligned}
$$

where

$$
\begin{aligned}
C_1 = \left|X_{t_0}\right|^p &+ \sum_{l=1}^{p/2}\sum_{s=0}^{l/2}\sum_{i=0}^{l-2s}\binom{l}{2s}\binom{p/2}{l}\binom{l-2s}{i}\left|X_{t_0}\right|^{2(\frac{p}{2}-l)} \\
&\times\left(2\langle X_{t_0}, g(X_{t_0})\rangle\right)^{2s}\left(\left|g(X_{t_0})\right|^2\left|\Delta\right|^2\right)^{l-2s-i}\left(2\alpha_0\right)^i\Delta t^{l-s}.
\end{aligned}
$$

is a constant. In order to complete the proof of the theorem we need to show that the dissipative nature of the drift can prevent potential growth of the diffusion term. To this end, for any given z, $1 < z < p$, we need to find the highest power q_1 of $-X_{t_k}^{q_1}\Delta t^z$, from the drift term. Next we need to obtain the highest power q_2 of $X_{t_k}^{q_2}\Delta t^z$ from the diffusion term. It is useful to notice that we obtain the highest power of drift and diffusion terms when $s = l$ and $i = 0$, respectively.

That is to say, for any given z by Assumption 4.1.1 and Lemma 4.2.1, we have

$$q_1 = 2(\frac{p}{2} - z) + (r+1)z \quad \text{and} \quad q_2 = 2(\frac{p}{2} - z) + 2\rho z.$$

This together with Assumption 4.1.2 clearly implies that

$$\mathbb{E}e^{t_{N+1}} \left| X_{t_{N\wedge\lambda_n}+1} \right|^p \leq C_1 + C \sum_{k=1}^{N+1} e^{t_{k-1}} \Delta t,$$

which implies the assertion easily. The proof is complete. □

By analogy with Lemma 4.1.4 we can extend the statement of Theorem 4.2.2 to the following one.

Theorem 4.2.3. *Under Assumptions 4.1.1, 4.1.2, 3.3.2 and 3.3.5, for any integer $p \geq 2$ and $T > 0$, there is $\Delta t^* \in (0, 1/L)$ and a constant $K = K(T, p)$ such that*

$$\sup_{\Delta t \leq \Delta t^*} \mathbb{E} \left[\sup_{0 \leq t_k \leq T} |X_{t_k}|^p \right] < K.$$

Proof. In the following we assume that N and M are positive integers such that

$$N\Delta t \leq M\Delta t \leq T. \tag{4.16}$$

From (4.15) we have

$$\begin{aligned}
|X_{t_N}|^2 &\leq |X_{t_0}|^2 + \sum_{k=0}^{N} 2\langle X_{t_k}, f(X_{t_k})\rangle \Delta t \\
&\quad + \sum_{k=0}^{N-1} |g(X_{t_k})\Delta W_{t_k}|^2 + \sum_{k=0}^{N-1} 2\langle X_{t_k}, g(X_{t_k})\Delta W_{t_k}\rangle \\
&\leq |X_{t_0}|^2 + 2\alpha_0(T + \Delta t) \\
&\quad + \sum_{k=0}^{N-1} |g(X_{t_k})\Delta W_{t_k}|^2 + \sum_{k=0}^{N-1} 2\langle X_{t_k}, g(X_{t_k})\Delta W_{t_k}\rangle.
\end{aligned}$$

Raising both sides to the power of $p/2$ we have

$$
\begin{aligned}
|X_{t_N}|^p \leq 4^{\frac{p-2}{2}} \Bigg(|X_{t_0}|^p + \Big(2\alpha_0 (T + \Delta t) \Big)^{p/2} \\
+ \Big(\sum_{k=0}^{N-1} |g(X_{t_k}) \Delta W_{t_k}|^2 \Big)^{p/2} + 2^{p/2} \Big(\sum_{k=0}^{N-1} \langle X_{t_k}, g(X_{t_k}) \Delta W_{t_k} \rangle \Big)^{p/2} \Bigg).
\end{aligned}
$$

Thanks to Theorem 4.2.2, there exists a positive constant $C = C(p)$, such that

$$
\begin{aligned}
\mathbb{E}\left[\sup_{0 \leq N \leq M} \Big(\sum_{k=0}^{N-1} |g(X_{t_k}) \Delta W_{t_k}|^2 \Big)^{p/2} \right] &\leq \mathbb{E}\left[M^{\frac{p-2}{2}} \sum_{k=0}^{M-1} |g(X_{t_k}) \Delta W_{t_k}|^p \right] \\
&\leq C \left[M^{\frac{p-2}{2}} \sum_{k=0}^{M-1} \Delta t^{p/2} \right] \\
&\leq C T^{p/2}.
\end{aligned}
$$

Finally, using the BDG inequality

$$
\begin{aligned}
\mathbb{E}\left[\sup_{0 \leq N \leq M} \left| \sum_{k=0}^{N-1} \langle X_{t_k}, g(X_{t_k}) \Delta W_{t_k} \rangle \right|^{p/2} \right] &\leq C \mathbb{E}\left[\sum_{k=0}^{M-1} |g(X_{t_k})|^2 \Delta t \right]^{p/4} \\
&\leq C M^{p/4-1} \Delta t^{p/4} \sum_{k=0}^{M-1} \mathbb{E} |g(X_{t_k})|^{p/2} \\
&\leq C T^{p/4}.
\end{aligned}
$$

Therefore we obtain

$$
\mathbb{E}\left[\sup_{0 \leq N \leq M} |X_{t_N}|^p \right] \leq C(p, T), \tag{4.17}
$$

and the desired result follows. □

4.3 Forward-Backward Euler-Maryuama Scheme

Here by analogy with Section 3.4 we work with the Forward-Backward Euler-Maryuama scheme. Under Assumptions 4.1.1, 4.1.2, 3.3.2, 3.3.5 the statement of Theorem 3.4.2 holds. We also have the following theorem which shows that both (2.2) and (3.22) stay close in L^p.

Theorem 4.3.1. *Under Assumptions 4.1.1, 4.1.2, 3.3.2 and 3.3.5, for any* $p \geq 2$, *there is* $\Delta t^* \in (0, 1/L)$ *and a constant* $K = K(p, T)$ *such that*

$$\mathbb{E}\left[\sup_{0 \leq t_k \leq T} \left|\hat{X}_{t_k} - X_{t_k}\right|^p\right] < K\Delta t^p,$$

and

$$\sup_{\Delta t \leq \Delta t^*} \mathbb{E}\left[\sup_{0 \leq t_k \leq T} \left|\hat{X}_{t_k}\right|^p\right] < \infty.$$

Proof. Let N and M be defined in exactly the same way as in (4.16). Summing up forward-backward and backward schemes, respectively we obtain

$$\hat{X}_{t_N} = \hat{X}_{t_0} + \sum_{k=0}^{N-1} f(X_{t_k})\Delta t + \sum_{k=0}^{N-1} g(X_{t_k})\Delta W_{t_k}, \tag{4.18}$$

$$X_{t_N} = X_{t_0} + \sum_{k=1}^{N-1} f(X_{t_{k+1}})\Delta t + \sum_{k=0}^{N-1} g(X_{t_k})\Delta W_{t_k}. \tag{4.19}$$

Now by the Hölder's inequality, Assumption 3.3.5 and Theorem 4.2.3 there exists a constant $C > 0$, such that

$$\mathbb{E}\left[\sup_{0 \leq N \leq M} \left|\hat{X}_{t_N} - X_{t_N}\right|^p\right] = \mathbb{E}\left[\sup_{0 \leq N \leq M} |f(X_{t_0})\Delta t - f(X_{t_N})\Delta t|^p\right] \leq C\Delta t^p, \tag{4.20}$$

and

$$\mathbb{E}\left[\sup_{0 \leq N \leq M} \left|\hat{X}_{t_N}\right|^p\right] \leq 2^{p-1}\mathbb{E}\left[\sup_{0 \leq N \leq M} \left(\left|\hat{X}_{t_{N+1}} - X_{t_N}\right|^p + |X_{t_N}|^p\right)\right] \leq K(1 + \Delta t^p), \tag{4.21}$$

as required. □

Having bounded the moments for the discrete FBEM, we can bound the continuous FBEM in the following sense.

Lemma 4.3.2. *Under Assumptions 4.1.1, 4.1.2, 3.3.2 and 3.3.5, for any integer* $p \geq 2$,

$$\sup_{\Delta t \leq \Delta t^*} \mathbb{E}\left(\sup_{0 \leq t \leq T} \left|\hat{X}(t)\right|^p\right) < \infty, \quad \forall T > 0.$$

Proof. It follows from (3.23) that

$$\mathbb{E}\Big(\sup_{0\le t\le T}\Big|\hat{X}(t)\Big|^p\Big) \le 3^{p-1}\Bigg[\mathbb{E}\Big|\hat{X}(0)\Big|^p + \mathbb{E}\left|\int_0^T f(X_{\eta(s)})ds\right|^p + \mathbb{E}\left(\sup_{0\le t\le T}\left|\int_0^t g(X_{\eta(s)})dw(s)\right|^p\right)\Bigg].$$

This, together with Theorem 4.2.3 and Assumption 3.3.5, implies the assertion. ☐

Lemma 4.3.3. *Under Assumptions 4.1.1, 4.1.2, 3.3.2 and 3.3.5 there exists a constant $C = C(T, p)$, such that*

$$\mathbb{E}\left|X_{\eta(s)} - \hat{X}(s)\right|^p \le K\Delta t^{p/2}, \quad s \in [0, T]. \tag{4.22}$$

Proof. By Hölder's inequality

$$\left|X_{\eta(s)} - \hat{X}(s)\right|^p \le 2^{p-1}(\left|\hat{X}_{\eta(s)} - \hat{X}(s)\right|^p + \left|X_{\eta(s)} - \hat{X}_{\eta(s)}\right|^p). \tag{4.23}$$

But by Theorem 4.2.3 we obtain

$$\mathbb{E}\left|\hat{X}_{\eta(s)} - \hat{X}(s)\right|^p = \mathbb{E}\left|\int_{t_k}^s f(X_{t_k})ds + \int_{t_k}^s g(X_{t_k})dw(s)\right|^p \le K\Delta t^{\frac{p}{2}}, \tag{4.24}$$

and by Theorem 4.3.1 we can show that

$$\mathbb{E}\left|X_{\eta(s)} - \hat{X}_{\eta(s)}\right|^p \le K\Delta t^p. \tag{4.25}$$

☐

4.4 Strong Convergence

It this section we prove the strong convergence theorems. However, the rate of convergence is still not revealed. Later, imposing additional conditions we prove the strong convergence theorem with an optimal rate.

Theorem 4.4.1. *Under Assumptions 4.1.1, 4.1.2, 3.3.2, 3.3.5 for any given $T > 0$ and $p \geq 2$, we have*

$$\lim_{\Delta t \to 0} \mathbb{E}\left[\sup_{0 \leq t \leq T} |X(t) - x(t)|^p\right] = 0. \tag{4.26}$$

The proof of Theorem 4.4.1 is analogous to the proof of Theorem 3.4.4. However, the measure of error of the approximation (2.2) to the solution of (3.1) is stronger. This improvement is possible due to Theorem 4.2.3.

4.4.1 Stochastic Lotka-Voltera System

Applications of stochastic differential equations are a growing interest in biomathematics. We focus here on a Stochastic Lotka-Voltera model, which is well established in the academic literature (Mao, Marion, and Renshaw 2002; Bahar and Mao 2004; Mao, Sabanis, and Renshaw 2003; Pang, Deng, and Mao 2008). However, to the best of our knowledge, there is no theoretical support for numerical methods which could enable further insight into the problem. We demonstrate that the coefficients of a fairly general stochastic population dynamics model satisfy the assumptions required in Theorem 4.4.1.
We look at a stochastic extension of the following model

$$\frac{dx(t)}{dt} = \text{diag}(x_1(t), x_2(t), ..., x_n(t))[b + Ax^2(t)], \tag{4.27}$$

where $x^s = (x_1^s, x_2^s, ..., x_n^s)^T$ for any $s \geq 1$, $b = (b_1, b_2, ..., b_n)^T$ and $A = (A_{ij})_{n \times n}$.
We consider state dependent perturbations of (4.27)

$$dx(t) = \text{diag}(x_1(t), x_2(t), ..., x_n(t))[(b + Ax^2(t))dt + g(x(t))dB(t)]. \tag{4.28}$$

The existence and uniqueness of the solution for 4.28 can be found in (Mao 2007).

Assumption 4.4.2. *We assume that for the square matrix A in (4.28) we have*

$$\lambda_{max}(A + A^T) < 0,$$

where

$$\lambda_{max}(A) = \sup_{x \in \mathbb{R}^n, |x|=1} x^T Ax.$$

In order to demonstrate that BEM (2.2) strongly converges to the solution of (4.28) we need to check that Assumption 4.1.1 holds.

$$\begin{aligned}\langle x, f(x)\rangle &= x^T \text{diag}(x_1, x_2, ..., x_n)[b + Ax^2] \\ &= (x_1^2, x_2^2, ..., x_n^2)[b + Ax^2] \\ &\leq |b| \, |x|^2 + (x^2)^T A x^2 \\ &= |b| \, |x|^2 + \frac{1}{2}(x^2)^T (A + A^T) x^2 \\ &\leq |b| \, |x|^2 + \frac{1}{2}\lambda_{max}(A + A^T) \sum_{i=1}^{n} x_i^4.\end{aligned}$$

Next we can observe that

$$\frac{1}{n^2} |x|^4 \leq \sum_{i=1}^{n} x_i^4 \leq |x|^4 . \tag{4.29}$$

We obtain the lower bound in (4.29) by simple manipulation, i.e,

$$\begin{aligned}|x|^4 = (|x|^2)^2 &\leq (n \max_{1 \leq i \leq n} x_i^2)^2 \\ &\leq n^2 \max_{1 \leq i \leq n} x_i^4 \leq n^2 \sum_{i=1}^{n} x_i^4,\end{aligned}$$

and the upper bound is obvious. Therefore

$$\langle x, f(x)\rangle \leq |b| \, |x|^2 + \frac{1}{2}\lambda_{max}(A + A^T) n^{-2} |x|^4 ,$$

and we can deduce that

$$\langle x, f(x)\rangle \geq - |x|^2 \, |b| + \frac{1}{2}\lambda_{min}(A + A^T) \, |x|^4 . \tag{4.30}$$

It is clear that all other assumptions required to prove Theorem 4.4.1 hold. This shows that BEM (2.2) is indeed a good approximation to the solution of (4.28).

4.5 Rate of Convergence

So far we have proved the strong convergence, but the rate of convergence has not been revealed. In this section we propose additional assumptions which enable us to derive the optimal rate for Euler-Maruyama type method, (Hofmann, Muller-Gronbach, and Ritter 2001). Below, we impose a stronger version of Assumptions 3.3.2 and 3.3.5.

Assumption 4.5.1. *Strong Monotone condition. For any constant $K_1 > 0$, there exists a constant $K = K(K_1) > 0$ such that*

$$\langle x - y, f(x) - f(y)\rangle + K_1 |g(x) - g(y)|^2 \leq K |x - y|^2 . \tag{4.31}$$

Assumption 4.5.2. *Strong Polynomial condition. For some $h \geq 1$, coefficients of the equation (3.1) satisfy a polynomial growth condition of the following form*

$$|f(x) - f(y)| \vee |g(x) - g(y)| \leq K(1 + |x|^h + |y|^h) |x - y| . \tag{4.32}$$

Theorem 4.5.3. *Under Assumptions 4.1.1, 4.1.2, 4.5.1 and 4.5.2, there exist a constant $K = K(p, T) > 0$, independent of Δt, such that, for any $p > 2$,*

$$\mathbb{E}\left|\hat{X}(t) - x(t)\right|^p \leq K\Delta t^{p/2} \quad \textit{for any } t \in [0, T]. \tag{4.33}$$

Proof. Let $t \in [0, T]$ and let us denote the error by $e(t)$, that is,

$$e(t) = \hat{X}(t) - x(t).$$

Hence

$$e(t) = \int_0^t [f(X(\eta(s))) - f(x(s))]\, ds + \int_0^t [g(X(\eta(s))) - g(x(s))]\, dB_s. \tag{4.34}$$

The Itô formula yields

$$
\begin{aligned}
|e(t)|^p &= \int_0^t p\,|e(s)|^{p-2}\left[\langle (f(X(\eta(s))) - f(x(s))), e(s)\rangle + \frac{1}{2}\,|g(X(\eta(s))) - g(x(s))|^2\right] ds \\
&+ \int_0^t p(p-2)\,|e(s)|^{p-4} \\
&\times \mathrm{trace}[(g(X(\eta(s))) - g(x(s)))^T e(s)(e(s))^T (g(X(\eta(s))) - g(x(s)))]ds \\
&+ M(t),
\end{aligned}
$$

where

$$
M(t) = \int_0^t p\,|e(s)|^{p-1}\,\langle e(s), g(X(\eta(s))) - g(x(s))\rangle dB_s.
$$

But

$$
\begin{aligned}
&\mathrm{trace}[(g(X(\eta(s))) - g(x(s)))^T e(s)(e(s))^T (g(X(\eta(s))) - g(x(s)))] \\
&= [e(s)]^T (g(X(\eta(s))) - g(x(s)))(g(X(\eta(s))) - g(x(s)))^T e(s) \\
&\le |e(s)|^2\,|(g(X(\eta(s))) - g(x(s)))|^2\,.
\end{aligned}
$$

Therefore

$$
\begin{aligned}
|e(t)|^p &\le \int_0^t p\,|e(s)|^{p-2}\left[\langle (f(X(\eta(s))) - f(\hat{X}(s)) + f(\hat{X}(s)) - f(x(s))), e(s)\rangle\right] ds \\
&+ \int_0^t |e(s)|^{p-2}\,\frac{p(p-1)}{2}\left|g(X(\eta(s))) - g(\hat{X}(s)) + g(\hat{X}(s)) - g(x(s))\right|^2 ds \\
&+ M(t) \\
&\le \int_0^t p\,|e(s)|^{p-2}\left[\langle f(X(\eta(s))) - f(\hat{X}(s), e(s)\rangle + \langle f(\hat{X}(s)) - f(x(s)), e(s)\rangle\right] ds \\
&+ \int_0^t |e(s)|^{p-2}\,p(p-1)[\left|g(X(\eta(s))) - g(\hat{X}(s))\right|^2 + \left|g(\hat{X}(s)) - g(x(s))\right|^2]ds \\
&+ M(t).
\end{aligned}
$$

By the Cauchy-Schwarz inequality

$$|e(t)|^p \leq \int_0^t p\,|e(s)|^{p-2} \left[\langle f(\hat{X}(s)) - f(x(s)), e(s) \rangle + (p-1) \left| g(\hat{X}(s)) - g(x(s)) \right|^2 \right] ds$$
$$+ \int_0^t |e(s)|^{p-2}\, p(p-1) \left[(p-1)^{-1} \left| \langle f(X(\eta(s))) - f(\hat{X}(s), e(s) \rangle \right| \right.$$
$$\left. + \left| g(X(\eta(s))) - g(\hat{X}(s)) \right|^2 \right] ds + M(t)$$
$$\leq \int_0^t p\,|e(s)|^{p-2} \left[\langle f(\hat{X}(s)) - f(x(s)), e(s) \rangle + (p-1) \left| g(\hat{X}(s)) - g(x(s)) \right|^2 \right] ds$$
$$+ \int_0^t |e(s)|^{p-2}\, p(p-1) \left[(p-1)^{-1} \left| f(X(\eta(s))) - f(\hat{X}(s) \right| |e(s)| \right.$$
$$\left. + \left| g(X(\eta(s))) - g(\hat{X}(s)) \right|^2 \right] ds + M(t).$$

By Young's inequality

$$|e(s)|^{p-1} \left| f(X(\eta(s))) - f(\hat{X}(s)) \right| < \frac{p-1}{p} |e(s)|^p + \frac{1}{p} \left| f(X(\eta(s))) - f(\hat{X}(s)) \right|^p,$$

and

$$|e(s)|^{p-2} \left| g(X(\eta(s))) - g(\hat{X}(s)) \right|^2 < \frac{p-2}{p} |e(s)|^p + \frac{2}{p} \left| g(X(\eta(s))) - g(\hat{X}(s)) \right|^p.$$

Hence

$$|e(t)|^p \leq \int_0^t p\,|e(s)|^{p-2} \left[\langle (f(\hat{X}(s)) - f(x(s))), e(s) \rangle \right.$$
$$\left. + (p-1) \left| g(\hat{X}(s)) - g(x(s)) \right|^2 \right] ds$$
$$+ \int_0^t \left[(p-1)\,|e(s)|^p + \left| f(X(\eta(s))) - f(\hat{X}(s)) \right|^p \right] ds$$
$$+ \int_0^t \left[(p-2)(p-1)\,|e(s)|^p + 2(p-1) \left| g(X(\eta(s))) - g(\hat{X}(s)) \right|^p \right] ds + M(t).$$

Now by Assumptions 4.5.1 and 4.5.2, there is a positive constant $C = C(p)$, such that

$$\begin{aligned} |e(t)|^p \le & \int_0^t C(p) \, |e(s)|^p \, ds \\ & + \int_0^t C(p)(1 + |X(\eta(s))|^{ph} + \left|\hat{X}(s)\right|^{ph}) \left|X(\eta(s)) - \hat{X}(s)\right|^p ds \\ & + M(t). \end{aligned} \tag{4.35}$$

By Hölder's inequality and Fubini's Theorem,

$$\begin{aligned} \mathbb{E}\,|e(t)|^p \le & \int_0^t C(p)\mathbb{E}\,|e(s)|^p \, ds \\ & + p \int_0^t K \left[\mathbb{E}(1 + |X(\eta(s))|^{ph} + \left|\hat{X}(s)\right|^{ph})^2 \mathbb{E} \left|X(\eta(s)) - \hat{X}(s)\right|^{2p} \right]^{1/2} ds. \end{aligned}$$

Now by Theorem 4.2.2, Lemma 4.3.3 and Gronwall's inequality, there exists a positive constant $C = C(p, T)$, such that

$$\mathbb{E}\,|e(t)|^p \le C\Delta t^{p/2} \quad \text{for any } t \in [0, T]. \tag{4.36}$$

□

Now we extend above result to convergence uniformly in time.

Theorem 4.5.4. *Under Assumptions 4.1.1, 4.1.2, 4.5.1 and 4.5.2, there exist a constant $C = C(p, T) > 0$, independent of Δt, such that for any $p > 2$, Forward-Backward Euler-Maruyama scheme has the property*

$$\mathbb{E}\left[\sup_{0 \le t \le T} \left| \hat{X}(t) - x(t) \right|^p \right] \le C\Delta t^{p/2}. \tag{4.37}$$

Proof. From (4.35) we have

$$\begin{aligned}\sup_{0\le t\le T} |e(t)|^p \le& \int_0^T C(p)\,|e(s)|^p\,ds \\ &+ \int_0^T C(p)(1+|X(\eta(s))|^{ph} + \left|\hat{X}(s)\right|^{ph})\left|X(\eta(s)) - \hat{X}(s)\right|^p ds \\ &+ \sup_{0\le t\le T} M(t). \end{aligned} \tag{4.38}$$

It is clear that we need to consider the last term on the right hand side of the above inequality, that is

$$\mathbb{E}\left[\sup_{0\le t\le T} M(t)\right] = \mathbb{E}\left[\sup_{0\le t\le T}\int_0^t p\,|e(s)|^{p-1}\,\langle e(s), (g(X(\eta(s))) - g(x(s)))\rangle dB_s\right]. \tag{4.39}$$

From Young's and BDG inequalities,

$$\begin{aligned}&\mathbb{E}\left[\sup_{0\le t\le T} M(t)\right] \le C(p)\mathbb{E}\left[\int_0^T |e(s)|^{2(p-1)}\,|g(X(\eta(s))) - g(x(s))|^2\,ds\right]^{1/2} \\ &\le C(p)\mathbb{E}\left[\sup_{0\le t\le T}|e(t)|^p\int_0^T |e(s)|^{p-2}\,|g(X(\eta(s))) - g(x(s))|^2\,ds\right]^{1/2} \\ &\le \frac{1}{2}\mathbb{E}\left[\sup_{0\le t\le T}|e(t)|^p\right] + C(p)\mathbb{E}\left[\int_0^T |e(s)|^{p-2}\,|g(X(\eta(s))) - g(x(s))|^2\,ds\right] \\ &\le \frac{1}{2}\mathbb{E}\left[\sup_{0\le t\le T}|e(t)|^p\right] + C(p)\mathbb{E}\left[\int_0^T |e(s)|^{p-2}\left|g(\hat{X}(s)) - g(x(s))\right|^2 ds\right] \\ &+ C(p)\mathbb{E}\left[\int_0^T |e(s)|^{p-2}\left|g(X(\eta(s))) - g(\hat{X}(s))\right|^2 ds\right] \\ &\le \frac{1}{2}\mathbb{E}\left[\sup_{0\le t\le T}|e(t)|^p\right] + C(p)\mathbb{E}\left[\int_0^T |e(s)|^{p-2}\left|g(\hat{X}(s)) - g(x(s))\right|^2 ds\right] \\ &+ C(p)\mathbb{E}\left[\int_0^T |e(s)|^p + \left|g(X(\eta(s))) - g(\hat{X}(s))\right|^p ds\right].\end{aligned}$$

Now by Assumption 4.5.2

$$
\begin{aligned}
\mathbb{E}\left[\sup_{0\leq t\leq T} M(t)\right] \leq \frac{1}{2}\mathbb{E}\left[\sup_{0\leq t\leq T}|e(t)|^p\right] \\
&+ C(p)\mathbb{E}\left[\int_0^T K(1+|x(s)|^{2h}+\left|\hat{X}(s)\right|^{2h})\left|x(s)-\hat{X}(s)\right|^p ds\right] \\
&+ C(p)\mathbb{E}\left[\int_0^T [e(s)]^p ds\right] \\
&+ C(p)\mathbb{E}\left[\int_0^T K(1+|X(\eta(s))|^{ph}+\left|\hat{X}(s)\right|^{ph})\left|X(\eta(s))-\hat{X}(s)\right|^p ds\right].
\end{aligned}
$$

By (4.38), Hölder's inequality and Fubini's Theorem, we obtain

$$
\begin{aligned}
&\frac{1}{2}\mathbb{E}\left[\sup_{0\leq t\leq T}|e(t)|^p\right] \leq \int_0^T C(p)\mathbb{E}\left|e(s)\right|^p ds \\
&+ C(p)\int_0^T \left[\mathbb{E}K(1+|X(\eta(s))|^{ph}+\left|\hat{X}(s)\right|^{ph})^2\mathbb{E}\left|X(\eta(s))-\hat{X}(s)\right|^{2p}\right]^{\frac{1}{2}} ds \\
&+ C(p)\int_0^T \left[\mathbb{E}K(1+|x(s)|^{ph}+\left|\hat{X}(s)\right|^{ph})^2\mathbb{E}\left|x(s)-\hat{X}(s)\right|^{2p}\right]^{\frac{1}{2}} ds.
\end{aligned}
$$

By (4.36), Theorem 4.2.2, Lemma 4.3.3, we get

$$
\begin{aligned}
\mathbb{E}\left[\sup_{0\leq t\leq T}|e(t)|^p\right] &\leq \int_0^T C(p,T)\Delta t^{p/2} ds \\
&+ C(p,T)\Delta t^{p/2}.
\end{aligned}
$$

Hence

$$
\mathbb{E}\left[\sup_{0\leq t\leq T}|e(t)|^p\right] < C(p,T)\Delta t^{p/2}. \tag{4.40}
$$

The proof is finished. □

Now we are ready to formulate the theorem on strong convergence of the Backward Euler-Maruyama (2.2) scheme to the solution of SDE (3.1).

Theorem 4.5.5. *Under Assumptions 4.1.1, 4.1.2, 4.5.1 and 4.5.2, for arbitrary* $T > 0$ *and* $p > 2$*, we have*

$$
\mathbb{E}\left[\sup_{0\leq t\leq T}\left|X_{\eta(t)}-x(\eta(t))\right|^p\right] \leq K\Delta^{p/2}, \tag{4.41}
$$

Proof. By Hölder's inequality

$$\mathbb{E}\left[\sup_{0\le t\le T}\left|X_{\eta(t)}-x(\eta(t))\right|^p\right]\le 2^{p-1}\left[\mathbb{E}\left[\sup_{0\le t\le T}\left|X_{\eta(t)}-\hat{X}_{\eta(t)}\right|^p\right]\right.$$
$$\left.+\,\mathbb{E}\left[\sup_{0\le t\le T}\left|\hat{X}_{\eta(t)}-x(\eta(t))\right|^p\right]\right].$$

Now by Lemma 4.3.1

$$\mathbb{E}\left[\sup_{0\le t\le T}\left|X_{\eta(t)}-\hat{X}_{\eta(t)}\right|^p\right]\le K\Delta t^p.$$

By Theorem 4.5.4

$$\mathbb{E}\left[\sup_{0\le t\le T}\left|\hat{X}_{\eta(t)}-x(\eta(t))\right|^p\right]\le K\Delta^{p/2},$$

as required. □

4.6 A Note on Fundamental Theorem

Similarly to the deterministic theory of numerical methods, Milstein (Milstein 1987) showed that it is enough to investigate one-step approximation to the SDEs (3.1) to state the convergence on the whole time interval. This theorem, often called the Fundamental Theorem, is very useful in situations where we would like to verify convergence of less standard schemes, for example (Milstein, Platen, and Schurz 1998). Nevertheless, to the best of our knowledge, there is no generalization of the result to a non-global Lipschitz case. Here, we demonstrate that under the assumptions we require to derive the convergence rate for BEM we can prove Fundamental Theorem as well.

For clarity of the exposition we adopt notation from (Milstein and Tretyakov 2004). By $x_{s,z}(t)$ we denote the process x at the time t with an initial condition $x(s)=z$. Similarly for the approximation, by $X_{s,z}(t_k)$ we denote a process X at a time t_k with an initial condition $X(s)=z$. Clearly

$$X_{t_{k+1}}=X_{t_k,X_{t_k}}(t_{k+1})=X_{t_0,X_{t_0}}(t_{k+1}).$$

Theorem 4.6.1 (Milstein (Milstein 1987)). *Suppose the one-step approximation $X_{t,x}(t+\Delta t)$ has order of accuracy p_1 for the mathematical expectation of deviation and order of accuracy p_2 for the mean-square deviation; more precisely, for arbitrary $t_0 \leq t \leq T-\Delta t$, $x \in \mathbb{R}^d$ the following inequalities hold:*

$$|\mathbb{E}(x_{t,z}(t+\Delta t) - X_{t,z}(t+\Delta t))| \leq K(1+|z|^2)^{1/2}\Delta t^{p_1},$$

$$(\mathbb{E}\,|x_{t,z}(t+\Delta t) - X_{t,z}(t+\Delta t)|^2)^{1/2} \leq K(1+|z|^2)^{1/2}\Delta t^{p_2}.$$

Also let

$$p_2 \geq \frac{1}{2}, \quad p_1 \geq p_2 + \frac{1}{2}.$$

Then for any N and $k = 0, 1, ..., N$ the following inequality holds:

$$(\mathbb{E}\,|x_{t_0,x_0}(t_k) - X_{t_0,x_0}(t_k)|^2)^{1/2} \leq K(1+|x_0|^2)^{1/2}\Delta t^{p_2-1/2},$$

i.e., the order of accuracy of the method constructed using the one-step approximation $X_{t,x}(t+\Delta t)$ is $p = p_2 - 1/2$.

The theorem was proved under the global Lipschitz condition in (Milstein and Tretyakov 2004). The authors demonstrated that the proof of the theorem follows from the following lemmas, which we derive for Backward Euler-Maruyama.

Lemma 4.6.2. *Let Assumptions 4.1.1, 4.1.2, 3.3.2 and 3.3.5 hold. For all natural N and all $k = 0, ..., N$ there exists a positive constant $K > 0$ such that Backward Euler-Maruyama scheme has property*

$$\mathbb{E}\,|X_k|^p \leq K(1+|X_0|^p) \quad \text{for} \quad p \geq 2. \tag{4.42}$$

In (Milstein and Tretyakov 2004) the lemma was proved with $p = 2$. Lemma 4.6.2 clearly corresponds to our Theorem 4.2.2.

Lemma 4.6.3. *Let Assumptions 4.1.1, 4.1.2, 4.5.1 and 4.5.2 hold. For the solution to the 3.1 there is a representation*

$$x_{t,x}(t+\Delta t) - x_{t,y}(t+\Delta t) = x - y + Z \tag{4.43}$$

for which

$$\mathbb{E}\,|x_{t,x}(t+\Delta t) - x_{t,y}(t+\Delta t)|^2 \leq |x-y|^2\,(1+K\Delta t), \tag{4.44}$$

$$\mathbb{E}Z^2 \leq K\left|x-y\right|^2 \Delta t. \tag{4.45}$$

Proof. Let $p \geq 2$. By the Itô formula for $0 \leq r \leq \Delta t$,

$$\begin{aligned}
&\mathbb{E}\left|x(t+r;z,t)-x(t+r;y,t)\right|^p \leq |z-y|^p \\
&+p\mathbb{E}\int_t^{t+r}\left|x(s;z,t)-x(s;y,t)\right|^{p-2}\langle x(s;z,t)-x(s;y,t), f(x(s;z,t))-f(x(s;y,t))\rangle ds \\
&+\frac{p(p-1)}{2}\mathbb{E}\int_t^{t+r}\left|x(s;z,t)-x(s;y,t)\right|^{p-2}\left|g(x(s;z,t))-g(x(s;y,t))\right|^2 ds.
\end{aligned}$$

By the strong monotone condition 4.5.1 we have

$$\mathbb{E}\left|x(t+r;z,t)-x(t+r;y,t)\right|^p \leq |z-y|^p + K\int_t^{t+r}\mathbb{E}\left|x(s;z,t)-x(s;y,t)\right|^p ds. \tag{4.46}$$

Gronwall's inequality implies

$$\mathbb{E}\left|x(t+r;z,t)-x(t+r;y,t)\right|^p \leq |z-y|^p e^{K\Delta t}, \quad 0 \leq r \leq \Delta t, \tag{4.47}$$

from which (4.44) follows. Now, writing Z explicitly

$$Z = \int_t^{t+\Delta t} f(x(s;z,t))-f(x(s;y,t))ds + \int_t^{t+\Delta t} g(x(s;z,t))-g(x(s;y,t))dw(s),$$

squaring and taking expectation, we obtain

$$\mathbb{E}\left|Z\right|^2 \leq K\int_t^{t+\Delta t}\mathbb{E}\left|f(x(s;z,t))-f(x(s;y,t))\right|^2 ds \tag{4.48}$$

$$+K\int_t^{t+\Delta t}\mathbb{E}\left|g(x(s;z,t))-g(x(s;y,t))\right|^2 ds. \tag{4.49}$$

Using the strong polynomial growth assumption 4.5.2 along with Lemma 4.6.2 and the Cauchy-Schwarz inequality lead us to

$$\mathbb{E}\left|Z\right|^2 = K\int_t^{t+\Delta t}\left(\mathbb{E}\left|x(s;z,t)-x(s;y,t)\right|^4\right)^{\frac{1}{2}} ds, \tag{4.50}$$

which together with (4.47) gives us (4.45). □

Chapter 5

Financial models

> The purpose of models is not to fit the data but to sharpen the questions.
>
> ---
>
> 11th R A Fisher Memorial Lecture, Royal Society.
>
> Samuel Karlin

In this chapter we demonstrate that the methodology developed here enables us to approximate many non-linear stochastic differential models. First we consider a fairly general SDE which appears in many interest rate and stochastic volatility models.
Further on, we take a closer look at the Ait-Sahalia interest rate model. The reason we treat this model separately is twofold. First, it exhibits a very special case of nonlinearity. A second feature that distinguishes analysis of Ait-Sahalia model is that in this case the Backward-Euler Maruyama method (2.2) preserves positivity.

5.1 General Stochastic Differential Financial Model

Consider the stochastic differential equation

$$dx(t) = f(x(t))dt + g(x(t))dw(t), \tag{5.1}$$

where

$$f(x) = \mu - \alpha x^r \quad \text{and} \quad g(x) = \beta x^\rho, \tag{5.2}$$

with $\rho \geq 1$, $r > 1$, and $\mu, \alpha, \beta > 0$. Further, assume that r is an odd number and that $r + 1 > 2\rho$.
The SDE (5.1) covers the family of highly nonlinear mean reverting models which play an essential role in the modern theory of interest rates, (Ahn and Gao 1999; Campbell, Lo, MacKinlay, and Whitelaw 1998; Ait-Sahalia 1996; Chan, Karolyi, Longstaff, and Sanders 1992).
Although Theorem 4.1.3 shows that the equation (5.1) admits a unique solution, having financial applications in view, we need to show that the solution of equation (5.1) stays non-negative.

Theorem 5.1.1. *For any given initial value $x(0) = x_0 > 0$, there exists a unique, non-negative global solution $x(t)$ to the equation (5.1) for $t \geq 0$.*

Proof. Clearly the coefficients (5.2) are locally Lipschitz continuous in $(0, \infty)$. Following the standard truncation method (see e.g. (Friedman 1976; Mao 2007) we can show that for any given initial value $x_0 > 0$ there exists a unique maximal local solution $x(t)$, $t \in [0, \tau_e)$, where τ_e is the stopping time of the explosion or first zero time. To prove our theorem, we need to show that $\tau_e = \infty$ a.s.
For every sufficiently large integer $m > 0$, such that $1/m < x(0) < m$, define the stopping time

$$\tau_m = \inf\{t \in [0, \tau_e) : \quad x(t) \notin (1/m, m)\}, \tag{5.3}$$

where throughout this thesis we set $\inf(\emptyset) = \infty$. Obviously τ_m is increasing as $m \to \infty$. Set $\tau_\infty = \lim_{m\to\infty} \tau_m$, whence $\tau_\infty \leq \tau_e$ a.s. If we can prove $\tau_m \to \infty$ a.s. as $m \to \infty$, then $\tau_e = \infty$ a.s. and $x(t) \geq 0$ a.s. for all $t \geq 0$. In other words, to complete the proof we may show that $\tau_\infty = \infty$ a.s. To prove this, it is enough to show that $\mathbb{P}\{\tau_m \leq T\} \to 0$ as $m \to \infty$ for any given constant $T > 0$. This immediately implies that $\mathbb{P}\{\tau_\infty = \infty\} = 1$ as required.

Let us define a function $V \in C^2(\mathbb{R}_+, \mathbb{R}_+)$ by

$$V(x) = x^{0.5} - 1 - 0.5 \log x. \tag{5.4}$$

It is easy to see that $V(x) \to \infty$ as $x \to \infty$ or $x \to 0$. Then, let us compute the diffusion operator

$$\begin{aligned} LV(x) &= V_x(x)f(x) + \frac{1}{2}V_{xx}(x)g(x)^2 \\ &\leq 0.5\left[(x^{-0.5} - x^{-1})(\mu - \alpha x^r) + \frac{1}{2}(-0.5x^{-1.5} + x^{-2})(\beta x^\rho)^2\right]. \end{aligned} \tag{5.5}$$

If $x \in (0, \infty)$, by similar argument to (4.8) we can find a constant K such that

$$LV(x) \leq K. \tag{5.6}$$

By the Itô formula,

$$\mathbb{E}V(x(T \wedge \tau_m)) \leq V(x(0)) + KT. \tag{5.7}$$

Therefore

$$\mathbb{P}(\tau_m \leq T)[V(1/m) \wedge V(m)] \leq \mathbb{E}V(x(T \wedge \tau_m)) \leq V(x(0)) + KT. \tag{5.8}$$

This implies that $\lim_{m\to\infty} \mathbb{P}(\tau_m \leq T) = 0$ as desired. $\square$

Since BEM scheme does not preserve positivity of the solution we will approximate SDE (5.1) by

$$X_{t_{k+1}} = X_{t_k} + (\mu - \alpha X^r_{t_{k+1}})\Delta t + \beta X^\rho_{t_k} \mathbf{1}_{\{X_{t_k} \geq 0\}} \Delta w_{t_k}. \tag{5.9}$$

This fix has been found to be the most efficient among several alternatives (Lord, Koekkoek, and Van Dijk 2009). Now we show that the general mean reverting type SDE (5.1) satisfies all the required assumptions of Theorem 4.5.5.

Lemma 5.1.2. *Coefficients (5.2) satisfy the strong polynomial growth condition 4.5.2, that is*

$$|f(x) - f(y)| \leq K(1 + |x|^{r-1} + |y|^{r-1})\,|x - y|\,, \tag{5.10}$$

$$|g(x) - g(y)| \leq K(1 + |x|^{\rho-1} + |y|^{\rho-1})\,|x - y|\,. \tag{5.11}$$

Proof. First let us consider the function f in (5.2), and assume that $x < y$. By the mean-value theorem there exists $s \in [x, y]$

$$f(x) - f(y) = f'(s)(x - y),$$

and by symmetry

$$|f(x) - f(y)| = |f'(s)| \, |x - y| \, .$$

Clearly, there exists a constant K such that

$$|f(x) - f(y)| \leq K \left(1 + |x|^{r-1} + |y|^{r-1}\right) |x - y| \, .$$

By the same argument we can show that condition (5.11) holds for the function g defined in (5.2). □

Lemma 5.1.3. *For every constant $K_1 > 0$, there exists a constant $K = K(K_1) > 0$, such that coefficients* (5.2) *satisfy the strong monotone condition in Assumption 4.5.1, i.e.,*

$$(x - y)(f(x) - f(y)) + K_1(g(x) - g(y))^2 \leq K(x - y)^2. \tag{5.12}$$

Proof. We present the proof for $x > y$, since $y < x$ is analogous. Binomial numbers can be factored algebraically as

$$x^r - y^r = (x - y)(x^{r-1} + x^{r-2}y + ... + xy^{r-2} + y^{r-1}). \tag{5.13}$$

By (5.2) and (5.11)

$$\begin{aligned}
&(x - y)(f(x) - f(y)) + K_1(g(x) - g(y))^2 \leq \\
&\alpha(x - y)^2(-x^{r-1} ... - y^{r-1}) + K_1\beta^2(x - y)^2(x^{2(\rho-1)} + y^{2(\rho-1)}) = \\
&\left[-\alpha x^{r-1} ... - \alpha y^{r-1} + K_1\beta^2(x^{2(\rho-1)} + y^{2(\rho-1)})\right](x - y)^2.
\end{aligned}$$

Now under the assumptions on ρ and r in (5.2) for any K_1 there exists a constant $K = K(K_1) > 0$, such that

$$-\alpha x^{r-1} - \alpha y^{r-1} + K_1\beta^2(x^{2(\rho-1)} + y^{2(\rho-1)}) \leq K, \quad x, y \in \mathbb{R}, \tag{5.14}$$

which completes the proof of the Lemma. □

5.1.1 Numerical Example

We consider the model (5.1) with $r = 3$ and $\rho = 2$, that is

$$dx(t) = (\mu - \alpha x(t)^3)dt + \beta x^2(t)dw(t), \tag{5.15}$$

where $(\mu, \alpha, \beta) = (0.5, 0.2, \sqrt{0.2})$. The assumptions of Theorem 4.5.5 hold apart from condition $r + 1 > 2\rho$. Nevertheless simulation suggest that strong convergence holds.
By Lemma 3.3.3 there exists a unique solution to (5.9). Since we employ BEM to approximate (5.15) on each step of the numerical simulation we need to find the inverse of the function $F(x) = \alpha x^3 \Delta t + x$. In this case we can find the inverse function explicitly and therefore computational complexity does not increase. Indeed, we observe that it is enough to find the real root of the cubic equation

$$\alpha X^3_{t_{k+1}} \Delta t + X_{t_{k+1}} - (X_{t_k} + \mu \Delta t + \beta X^2_{t_k} \Delta w_{t_k}) = 0. \tag{5.16}$$

In our numerical experiment, we focus on the error at the endpoint $T = 1$, so we let

$$e^{strong}_{\Delta t} = \mathbb{E}\left|x(T) - X_T\right|.$$

In Figure 5.1 we plot $e^{strong}_{\Delta t}$ against Δt on log-log scale. Error bars representing 95% confidence intervals are shown by circles.

Although we do not know the explicit form of the solution to (5.15), Theorem 4.5.5 guarantees that BEM (5.9) strongly converges to the true solution. Therefore, it is reasonable to take BEM with very small time step, we choose $\Delta t = 2^{-15}$, as a reference solution. We then compare it to BEM evaluated with $(2^4\Delta t, 2^6\Delta t, 2^8\Delta t, 2^{10}\Delta t)$ in order to estimate the rate of convergence. Since we are using Monte Carlo method, the sampling error decays like $1/\sqrt{M}$, M- is a number of sample paths. We set $M = 1000$. From the Figure 5.1 we see that there appears to exist a positive constant such that

$$e^{strong}_{\Delta t} \leq C\Delta t^{\frac{1}{2}} \quad \text{for sufficiently small } \Delta t.$$

A least squares fit for $\log C$ and q produced the value 0.5696 for q with a least square residual of 0.0861. Hence, our results are consistent with strong order of

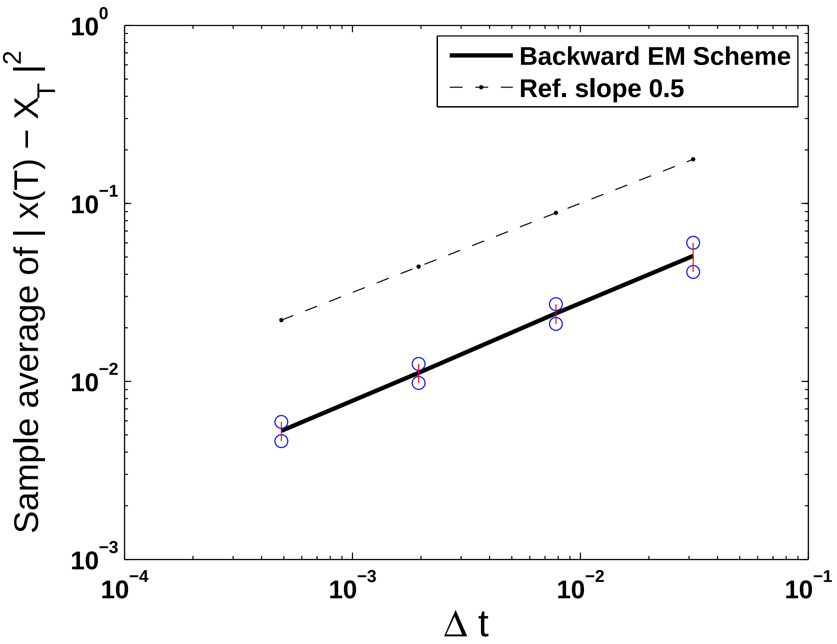

Figure 5.1: Strong error plot: dashed black is the reference slope. The straight black line is the extrapolation of error estimates for BEM.

convergence equal to one-half.

5.2 Ait-Sahalia Model

The SDE that we refer to as the *generalized Ait-Sahalia model* has the form

$$dx(t) = (\alpha_{-1}x(t)^{-1} - \alpha_0 + \alpha_1 x(t) - \alpha_2 x(t)^r)dt + \sigma x(t)^\rho dw(t), \tag{5.17}$$

where $\alpha_{-1}, \alpha_0, \alpha, \alpha_1, \alpha_2, \sigma$ are positive constants and $\rho > 1$. The model was introduce in (Ait-Sahalia 1996) with $r = 2$. In order to show that the model (5.17) is meaningful, the next theorem guarantees that a unique solution exists, and remains in $\mathbb{R}_+ := (0, \infty)$.

Theorem 5.2.1. *Given any initial value $x(0) = x_0 > 0$, there exists a unique, positive global solution $x(t)$ to the equation (5.17) on $t \geq 0$.*

Proof. Define the coefficients of the equation (5.17) using

$$f(x) = \alpha_{-1}x^{-1} - \alpha_0 + \alpha_1 x - \alpha_2 x^r \quad \text{and} \quad g(x) = \sigma x^\rho \quad \text{for } x > 0. \tag{5.18}$$

Clearly, f and g are locally Lipschitz continuous in $(0, \infty)$. Following the standard truncation method (see e.g. (Mao 2007; Friedman 1976)), we can show that for any given initial value $x_0 > 0$ there exists a unique maximal local solution $x(t)$, $t \in [0, \tau_e)$, where τ_e is the stopping time of the explosion or first zero time. To prove our theorem, we need to show that $\tau_e = \infty$ a.s.

For every sufficiently large integer $k > 0$, such that $1/k < x(0) < k$, define the stopping time

$$\tau_k = \inf\{t \in [0, \tau_e) : \quad x(t) \notin (1/k, k)\},$$

where we set $\inf(\emptyset) = \infty$. Obviously τ_k is increasing as $k \to \infty$. Set $\tau_\infty = \lim_{k\to\infty} \tau_k$, whence $\tau_\infty \leq \tau_e$ a.s. If we can prove $\tau_k \to \infty$ a.s as $k \to \infty$, then $\tau_e = \infty$ a.s and $x(t) \geq 0$ a.s. for all $t \geq 0$. In other words, to complete the proof what we need to show is that $\tau_\infty = \infty$ a.s. To prove this, it is enough to show that $P\{\tau_k \leq T\} \to 0$ as $k \to \infty$ for any given constant $T > 0$, for this immediately implies that $P\{\tau_\infty = \infty\} = 1$ as required.

Fix two constants $\gamma_1 \in (0, 1)$ and $\gamma_2 > 1$. Let us define a function $V \in C^2(\mathbb{R}_+, \mathbb{R}_+)$ by

$$V(x) = x^{\gamma_1} + x^{-\gamma_2}. \tag{5.19}$$

It is easy to see that $V(x) \to \infty$ as $x \to \infty$ or $x \to 0$. Compute the diffusion operator

$$\begin{aligned}
LV(x) &= V_x(x)f(x) + \frac{1}{2}V_{xx}(x)g(x)^2 \\
&= (\gamma_1 x^{\gamma_1 - 1} - \gamma_2 x^{-(\gamma_2+1)})f(x) \\
&+ \frac{1}{2}(\gamma_1(\gamma_1 - 1)x^{\gamma_1-2} + \gamma_2(\gamma_2+1)x^{-(\gamma_2+2)})g(x)^2 \\
&= \gamma_1\alpha_{-1}x^{\gamma_1-2} - \alpha_0\gamma_1 x^{\gamma_1-1} + \alpha_1\gamma_1 x^{\gamma_1} - \alpha_2\gamma_1 x^{\gamma_1-1+r} \\
&- \alpha_{-1}\gamma_2 x^{-(\gamma_2+2)} + \alpha_0\gamma_2 x^{-(\gamma_2+1)} - \alpha_1\gamma_2 x^{-\gamma_2} + \gamma_2 x^{-(\gamma_2+1)+r} \\
&+ \frac{\sigma^2}{2}(\gamma_1(\gamma_1-1)x^{\gamma_1-2+2\rho} + \gamma_2(\gamma_2+1)x^{-(\gamma_2+2)+2\rho}).
\end{aligned}$$

Recalling that $\gamma_1 \in (0,1)$ and $\gamma_2 > 1$, we can find a constant K such that

$$LV(x) \le K. \tag{5.20}$$

By the Itô formula,

$$\mathbb{E}V(x(T \wedge \tau_k)) \le V(x_0) + KT. \tag{5.21}$$

Therefore

$$\mathbb{P}(\tau_k \le T)[V(1/k) \wedge V(k)] \le \mathbb{E}V(x(T \wedge \tau_k)) \le V(x(0)) + KT.$$

This implies that $\lim_{k\to\infty} \mathbb{P}(\tau_k \le T) = 0$ as desired. The proof is complete. □

In order to proceed with our analysis, we make an assumption about the values of the parameters.

Assumption 5.2.2. *The parameters in equation (5.17) obey $r > 1$ and*

$$r + 1 > 2\rho. \tag{5.22}$$

The following lemma gives moment bounds for the solution of the SDE.

Lemma 5.2.3. *Under Assumption 5.2.2, for any $p \ge 2$,*

$$\sup_{0\le t<\infty} \mathbb{E}\,|x(t)|^p < \infty \tag{5.23}$$

and

$$\sup_{0\le t<\infty} \mathbb{E}\Big(\frac{1}{|x(t)|^p}\Big) < \infty. \tag{5.24}$$

Proof. For every sufficiently large integer n, define the stopping time

$$\tau_n = \inf\{t > 0 : x(t) \notin (\frac{1}{n}, n)\}.$$

Applying the Itô formula to the function $V(x,t) = e^t x^p$, we compute the diffusion operator

$$\begin{aligned} LV(x,t) = e^t\Big(&x^p + px^{p-1}[\alpha_{-1}x^{-1} - \alpha_0 + \alpha_1 x - \alpha_2 x^r] \\ &+ \frac{\sigma^2}{2}p(p-1)x^{p-2+2\rho}\Big). \end{aligned}$$

By Assumption 5.2.2, there exists a constant $K > 0$ such that

$$LV(x,t) \leq Ke^t.$$

Therefore

$$\mathbb{E}\Big[e^{t\wedge\tau_n}x(t\wedge\tau_n)^p\Big] \leq x_0^p + Ke^t.$$

Letting $n \to \infty$ and applying the Fatou lemma, we have

$$\mathbb{E}\,|x(t)|^p \leq \frac{x_0^p}{e^t} + K,$$

which gives assertion (5.23). In the same way, we can apply the Itô formula to the function $V(x,t) = e^t x^{-p}$ to show (5.24). □

Lemma 5.2.4. *Under Assumption 5.2.2, for any $p \geq 2$,*

$$\mathbb{E}(\sup_{0\leq t\leq T}|x(t)|^p) < \infty, \quad \forall T > 0.$$

Proof. By the Itô formula, we can show that

$$\begin{aligned}
&\mathbb{E}[\sup_{0\leq t\leq T}|x(t)|^p] \leq x_0^p \\
&+\mathbb{E}\int_0^T p\left|x(t)^{p-1}(\alpha_{-1}x(t)^{-1} - \alpha_0 + \alpha_1 x(t) - \alpha_2 x(t)^r) + 0.5(p-1)\sigma^2 x(t)^{2(\rho-1)+p}\right| dt \\
&+\mathbb{E}\Big[\sup_{0\leq t\leq T}\int_0^t \sigma p x(u)^{\rho+p-1}dw(u)\Big] \\
&\leq x(0)^p + KT + \mathbb{E}\Big[\sup_{0\leq t\leq T}\int_0^t \sigma p x(u)^{\rho+p-1}dw(u)\Big],
\end{aligned}$$

where K is a constant. By the Hölder and (BDG) inequalities, we can show that

$$\begin{aligned}
\mathbb{E}\Big[\sup_{0\leq t\leq T}\int_0^t \sigma p x(u)^{\rho+p-1}dw(u)\Big] &\leq C\mathbb{E}\Big(\int_0^T x(t)^{2(\rho+p-1)}dt\Big)^{\frac{1}{2}} \\
&\leq C\Big(\int_0^T \mathbb{E}[x(t)^{2(\rho+p-1)}]dt\Big)^{\frac{1}{2}},
\end{aligned}$$

where C stands for a constant which may vary from line to line. By Lemma 5.2.3, the conclusion follows. □

5.2.1 Backward Euler-Maruyama Scheme

In the previous section we showed the existence of a unique global solution to the SDE (5.17), but we are not aware of an explicit expression for the solution or its transition density. We therefore consider computable discrete time approximations that could be used in Monte Carlo simulations.
Given any step size Δt, we define the partition $\mathcal{P}_{\Delta t} := \{t_k = k\Delta t : k = 0, 1, 2, ...\}$ of the time interval $[0, \infty)$, and introduce the backward Euler-Maruyama scheme

$$X_{t_{k+1}} = X_{t_k} + \left(\alpha_{-1}X_{t_{k+1}}^{-1} - \alpha_0 + \alpha_1 X_{t_{k+1}} - \alpha_2 X_{t_{k+1}}^r\right)\Delta t + \sigma X_{t_k}^{\rho}\Delta w_{t_k}, \quad (5.25)$$

where $\Delta w_{t_k} = w_{t_{k+1}} - w_{t_k}$ and $X_{t_0} = x(0)$. The following lemma shows that this implicit method is well defined and preserves positivity of the solution.

Lemma 5.2.5. *Define, for any given* $\Delta t \leq 1/\alpha_1$,

$$F(x) = x - \alpha_{-1}x^{-1}\Delta t + \alpha_0\Delta t - \alpha_1 x\Delta t + \alpha_2 x^r\Delta t, \quad x \in \mathbb{R}_+.$$

Then for any $b \in \mathbb{R}$ *there exists a unique* $x \in \mathbb{R}_+$ *such that* $F(x) = b$.

Proof. The lemma follows if we can show that the function F is continuous, coercive and strictly monotone 3.3.1. Clearly, $F(x)$ is continuous on $\mathbb{R}_+$ with $\lim_{x\to\infty} F(x) = \infty$ and $\lim_{x\to 0^+} F(x) = -\infty$, so the function F is coercive on $\mathbb{R}_+$. Since $F'(x) = 1 + (\alpha_{-1}x^{-2} - \alpha_1 + r\alpha_2 x^{r-1})\Delta t > 1 - \alpha_1\Delta t$, we see that $\dot{F}(x) > 0$ whenever $\Delta t \leq 1/\alpha_1$, showing strict monotonicity. □

From now on we always let $\Delta t \leq 1/\alpha_1$ so that the BEM is well defined and preserves positivity. In contrast, let us point out that the (standard) Euler-Maruyama scheme does not preserve the positivity of the solution to equation (5.17). In fact, recall that the Euler-Maruyama scheme applied to equation (5.17) has the form

$$X_{t_{k+1}} = X_{t_k} + \left(\alpha_{-1}X_{t_k}^{-1} - \alpha_0 + \alpha_1 X_{t_k} - \alpha_2 X_{t_k}^r\right)\Delta t + \sigma X_{t_k}^{\rho}\Delta w_{t_k}.$$

Without loss of generality, we assume that $X_{t_k} > 0$ is given. Note that $X_{t_{k+1}} < 0$ is equivalent to $\Delta w_{t_k} < -\left(X_{t_k} + \left(\alpha_{-1}X_{t_k}^{-1} - \alpha_0 + \alpha_1 X_{t_k} - \alpha_2 X_{t_k}^r\right)\Delta t\right)/\sigma X_{t_k}^{\rho} := K(X_{t_k})$, but clearly $\mathbb{P}(\Delta w_{t_k} < K(X_{t_k})) > 0$.

5.2.2 Moment Properties of BEM

We will work on the discrete filtration $\{\mathcal{F}_{t_k}\}_{k\geq 0}$. By Lemma 5.2.5, X_{t_k} is $\mathcal{F}_{t_k}$-measurable.

Lemma 5.2.6. *Let $r > 1$, then for any $p \geq 2$ and sufficiently large integer n, there exists a constant $K(p,n)$, such that*

$$\sup_{\Delta t \leq 1/2\alpha_1} \mathbb{E}\,|X_{t_k}|^p\,\mathbf{1}_{[0,\lambda_n]}(k) < K(p,n) \qquad \textit{for any } k \geq 0,$$

where

$$\lambda_n = \inf\{k : X_{t_k} \notin (\frac{1}{n}, n)\}. \tag{5.26}$$

Lemma 5.2.6 follows from Lemma 3.3.4. Similarly to Theorem 4.2.2 we can demonstrate that the following theorem holds.

Theorem 5.2.7. *Under Assumption 5.2.2, for any $p > 2$, there is a $\Delta t^* \in (0, 1/2\alpha_1)$ such that*

$$\sup_{\Delta t \leq \Delta t^*} \sup_{k\geq 0} \mathbb{E}\,|X_{t_k}|^p < \infty.$$

5.2.3 Forward-Backward Euler-Maruyama Scheme

By analogy to Theorem 4.3.1 we have:

Theorem 5.2.8. *Under Assumption 5.2.2, for any $p > 2$, there is a $\Delta t^* \in (0, 1/2\alpha_1)$ and a constant $K = K(p,T)$ such that*

$$\mathbb{E}\left[\sup_{0\leq t_k \leq T} \left|\hat{X}_{t_{k+1}} - X_{t_k}\right|^p\right] < K\Delta t^{p/2},$$

and

$$\sup_{\Delta t \leq \Delta t^*} \mathbb{E}\left[\sup_{0\leq t_k\leq T} \left|\hat{X}_{t_k}\right|^p\right] < \infty.$$

Now let us recall that a continuous version of the FBEM is given by

$$\hat{X}(t) = \hat{X}_{t_0} + \int_0^t f(X_{\eta(s)})ds + \int_0^t g(X_{\eta(s)})dw(s), \quad t \geq 0. \tag{5.27}$$

In order to proceed with our analysis we need to prove the following lemmas.

Lemma 5.2.9. *Under Assumption 5.2.2, there is a* $\Delta t^* > 0$ *such that for any* $p \geq 2$,

$$\sup_{\Delta t \leq \Delta t^*} \mathbb{E}\left[\int_0^t \frac{1}{X_{\eta(s)}} ds\right]^p < \infty, \quad \forall t > 0.$$

Proof. We only need to prove the lemma for $t = t_N$ for any $N \geq 1$. It follows from (5.27) that

$$\alpha_{-1} \int_0^{t_N} \frac{1}{X_{\eta(s)}} ds = \hat{X}_{t_N} - X_{t_0} + \alpha_0 t_N - \alpha_1 \int_0^{t_N} X_{\eta(s)} ds + \alpha_2 \int_0^{t_N} X^r_{\eta(s)} ds - \sigma \int_0^{t_N} X^\rho_{\eta(s)} dw(s).$$

It is then straightforward to show the assertion by Theorem 5.2.7. □

Lemma 5.2.10. *Under Assumption 5.2.2, there is a* $\Delta t^* > 0$ *such that for any* $p \geq 2$,

$$\sup_{\Delta t \leq \Delta t^*} \mathbb{E}\Big(\sup_{0 \leq t \leq T} \left|\hat{X}(t)\right|^p\Big) < \infty, \quad \forall T > 0. \tag{5.28}$$

Proof. It follows from (5.27) that

$$\mathbb{E}\Big(\sup_{0 \leq t \leq T} \left|\hat{X}(t)\right|^p\Big) \leq 3^{p-1}\Bigg[\mathbb{E}\left(\hat{X}(0)^p\right) + \mathbb{E}\left(\int_0^T (\alpha_{-1} X^{-1}_{\eta(s)} + \alpha_0 + \alpha_1 X_{\eta(s)} + \alpha_2 X^r_{\eta(s)}) ds\right)^p + \mathbb{E}\left(\sup_{0 \leq t \leq T} \left|\int_0^t \sigma X^\rho_{\eta(s)} dw(s)\right|^p\right)\Bigg].$$

This, together with Theorem 5.2.7 and Lemma 5.2.9, implies the assertion. □

Theorem 5.2.11. *Let Assumption 5.2.2 hold and* $T > 0$ *be fixed. Then, for any given* $\epsilon > 0$, *there exists an* N_0 *such that for every* $n \geq N_0$, *we can find a* $\Delta t_0 = \Delta t_0(n)$ *so that whenever* $\Delta t \leq \Delta t_0$,

$$\mathbb{P}(\vartheta_n < T) \leq \epsilon,$$

where $\vartheta_n = \inf\{t > 0 : \hat{X}(t) \notin (\frac{1}{n}, n) \quad \text{or} \quad X_{\eta(s)} \notin (\frac{1}{n}, n)\}$.

Proof. Let $s \in [0, T \wedge \vartheta_n)$. Then for the function $V(x)$ defined by (5.19) we have

$$V_x(\hat{X}(s))\left(f(X_{\eta(s)}) - f(\hat{X}(s))\right)$$
$$+ V_{xx}(\hat{X}(s))\left(g^2(X_{\eta(s)}) - g^2(\hat{X}(s))\right) \leq L(n)\left|X_{\eta(s)} - \hat{X}(s)\right|,$$

where $L(n)$ are local Lipschitz constants. By the Itô formula, we can show that

$$\begin{aligned} dV(\hat{X}(s)) = &\Big[LV(\hat{X}(s)) + V_x(\hat{X}(s))\left(f(X_{\eta(s)}) - f(\hat{X}(s))\right) \\ &+ \frac{1}{2}V_{xx}(\hat{X}(s))\left(g^2(X_{\eta(s)}) - g^2(\hat{X}(s)\right)\Big]ds \\ &+ V_x(\hat{X}(s))g(X_{\eta(s)})dw(s), \end{aligned}$$

where LV has been defined in the proof of Theorem 5.2.1. Recalling (5.20), we then have

$$\begin{aligned} &\mathbb{E}V(\hat{X}(T \wedge \vartheta_n)) \leq \\ &V(\hat{X}(0)) + KT + \mathbb{E}\int_0^{T\wedge\vartheta_n} V_x(\hat{X}(s))\left(f(X_{\eta(s)}) - f(\hat{X}(s))\right)ds \\ &+ \mathbb{E}\int_0^{T\wedge\vartheta_n} V_{xx}(\hat{X}(s))\left(g^2(X_{\eta(s)}) - g^2(\hat{X}(s))\right)ds \\ &\leq V(\hat{X}(0)) + KT + L(n)\mathbb{E}\int_0^{T\wedge\vartheta_n}\left|X_{\eta(s)} - \hat{X}(s)\right|ds \\ &\leq V(\hat{X}(0)) + KT + L(n)\mathbb{E}\int_0^{T\wedge\vartheta_n}\left|\hat{X}_{\eta(s)+\Delta t} - \hat{X}(s)\right|ds \\ &+ L(n)\int_0^T \mathbb{E}\left|X_{\eta(s)} - \hat{X}_{\eta(s)+\Delta t}\right|ds. \end{aligned}$$

By Theorem 5.2.8

$$\mathbb{E}\left|X_{\eta(s)} - \hat{X}_{\eta(s)+\Delta t}\right| < K\Delta t^{\frac{1}{2}}. \tag{5.29}$$

To bound the term $\mathbb{E}\int_0^{T\wedge\vartheta_n}\left|\hat{X}_{\eta(s)+\Delta t} - \hat{X}(s)\right|ds$, given $s \in [0, T \wedge \vartheta_n)$, let k be an integer for which $s \in [t_k, t_{k+1})$. Then

$$\left|\hat{X}_{\eta(s)+\Delta t} - \hat{X}(s)\right| = \int_s^{t_{k+1}} f(X_{t_k})ds + \int_s^{t_{k+1}} g(X_{t_k})dw(s).$$

By Hölder's inequality

$$\mathbb{E}\int_0^{T\wedge\vartheta_n}\left|\hat{X}_{\eta(s)+\Delta t}-\hat{X}(s)\right|ds\leq C(n,T)\Delta t^{\frac{1}{2}}, \tag{5.30}$$

where $C(n,T)>0$ is constant. Therefore

$$\mathbb{E}V(\hat{X}(t\wedge\vartheta_n))\leq V(\hat{X}(0))+KT+(L(n))(K+C(T,n))\Delta t^{\frac{1}{2}},$$

which implies that

$$\mathbb{P}(\vartheta_n<T)\leq\frac{V(\hat{X}(0))+KT+(L(n))(K+C(T,n))\Delta t^{\frac{1}{2}}}{V(1/n)\wedge V(n)}.$$

Now for any given $\epsilon>0$ we choose N_0 such that for any $n\geq N_0$

$$\frac{V(\hat{X}(0))+KT}{V(1/n)\wedge V(n)}\leq\frac{\epsilon}{2}.$$

Then we can choose $\Delta t_0=\Delta t_0(n)$, such that for any $\Delta t\leq\Delta t_0$

$$\frac{(L(n))(K+C(T,n))\Delta t^{\frac{1}{2}}}{V(1/n)\wedge V(n)}\leq\frac{\epsilon}{2},$$

whence $\mathbb{P}(\vartheta_n<T)\leq\epsilon$ as required. □

5.2.4 Strong Convergence

In this section, we use the previous results to establish the strong convergence of BEM.

Theorem 5.2.12. *Let $p\leq 1$ Under Assumption (5.2.2), we have*

$$\lim_{\Delta t\to 0}\mathbb{E}\left[\sup_{0\leq t\leq T}\left|\hat{X}(t)-x(t)\right|^p\right]=0.$$

Proof. Let

$$e(t)=\hat{X}(t)-x(t).$$

We have

$$\begin{aligned}\mathbb{E}\left[\sup_{0\le t\le T}|e(t)|^p\right] &= \mathbb{E}\left[\sup_{0\le t\le T}|e(t)|^p\,\mathbf{1}_{\{\tau_n>T,\vartheta_n>T\}}\right] \\ &+\mathbb{E}\left[\sup_{0\le t\le T}|e(t)|^p\,\mathbf{1}_{\{\tau_n\le T\quad or\quad \vartheta_n\le T\}}\right] \\ &\le \mathbb{E}\left[\sup_{0\le t\le T}|e(t)|^p\,\mathbf{1}_{\{\theta_n>T\}}\right]+\frac{\delta}{2}\mathbb{E}\left[\sup_{0\le t\le T}|e(t)|^{2p}\right] \\ &+\frac{1}{2\delta}\mathbb{P}(\tau_n\le T\quad \text{or}\quad \vartheta_n\le T).\end{aligned} \tag{5.31}$$

To finish the proof we need to estimate the expressions on the right hand side of this inequality. By Hölder's inequality and Lemmas 5.2.10 and 5.2.4, we choose δ such that

$$\frac{\delta}{2}\mathbb{E}\left[\sup_{0\le t\le T}|e(t)|^{2p}\right]\le 2^{2p-1}\frac{\delta}{2}\mathbb{E}\left[\sup_{0\le t\le T}|x(t)|^{2p}+\sup_{0\le t\le T}\left|\hat{X}(t)\right|^{2p}\right]\le\frac{\epsilon}{3}$$

Next, by Theorem 5.2.1 there exists N_0 such that for $n\ge N_0$

$$\frac{1}{2\delta}\mathbb{P}(\tau_n\le T)\le\frac{\epsilon}{3},$$

and finally by Theorem 5.2.11 and Lemma 3.4.3- we may choose Δt sufficiently small such that

$$\mathbb{E}\left[\sup_{0\le t\le T}|e(t)|^p\,\mathbf{1}_{\{\theta_n>T\}}\right]+\mathbb{P}(\vartheta_n\le T)\le\frac{\epsilon}{3}.$$

□

Now, we will show that the Backward-Euler scheme (5.25) strongly converges to the solution of (5.17).

Theorem 5.2.13. *Under Assumption (5.2.2), we have*

$$\lim_{\Delta t\to 0}\mathbb{E}\left[\sup_{0\le t\le T}\left|X_{\eta(t)}-x(\eta(t))\right|^p\right]=0.$$

Proof. By Hölder's inequality

$$\mathbb{E}\left|X_{\eta(t)} - x(\eta(t))\right|^p \leq 3^{p-1}\big[\mathbb{E}\left[\sup_{0\leq t\leq T}\left|X_{\eta(t)} - \hat{X}_{\eta(t)+1}\right|\right]^p$$
$$+\,\mathbb{E}\left[\sup_{0\leq t\leq T}\left|\hat{X}_{\eta(t)} - x(\eta(t))\right|\right]^p + \mathbb{E}\left[\sup_{0\leq t\leq T}\left|\hat{X}_{\eta(t)+1} - \hat{X}_{\eta(t)}\right|^p\right]\big].$$

Now from Theorem 5.2.8 and Theorem 5.2.8

$$\mathbb{E}\left[\sup_{0\leq t\leq T}\left|X_{\eta(t)} - \hat{X}_{\eta(t)+1}\right|\right]^p \leq K\Delta t^{\frac{p}{2}}.$$

By Theorem 5.2.12

$$\lim_{\Delta t\to 0}\mathbb{E}\left[\sup_{0\leq t\leq T}\left|\hat{X}_{\eta(t)} - x(\eta(t))\right|^p\right] = 0.$$

To finish the proof it is enough to show that

$$\lim_{\Delta t\to 0}\mathbb{E}\left[\sup_{0\leq t\leq T}\left|\hat{X}_{\eta(t)+1} - \hat{X}_{\eta(t)}\right|^p\right] = 0.$$

By analogy to the proof of Theorem 5.2.12, we have

$$\begin{aligned}
\mathbb{E}\left[\sup_{0\leq t\leq T}\left|\hat{X}_{\eta(t)+1} - \hat{X}_{\eta(t)}\right|^p\right] &= \mathbb{E}\left[\sup_{0\leq t\leq T}\left|\hat{X}_{\eta(t)+1} - \hat{X}_{\eta(t)}\right|^p \mathbf{1}_{\{\vartheta_n > T\}}\right] \\
&+ \mathbb{E}\left[\sup_{0\leq t\leq T}\left|\hat{X}_{\eta(t)+1} - \hat{X}_{\eta(t)}\right|^p \mathbf{1}_{\{\vartheta_n \leq T\}}\right] \\
&\leq \mathbb{E}\left[\sup_{0\leq t\leq T}\left|\hat{X}_{\eta(t)+1} - \hat{X}_{\eta(t)}\right|^p \mathbf{1}_{\{\vartheta_n > T\}}\right] \\
&+ \frac{\delta}{2}\mathbb{E}\left[\sup_{0\leq t\leq T}\left|\hat{X}_{\eta(t)+1} - \hat{X}_{\eta(t)}\right|^{2p}\right] \\
&+ \frac{1}{2\delta}\mathbb{P}(\vartheta_n \leq T).
\end{aligned}$$

By Lemma 5.2.10 and Theorem 5.2.11 it is straightforward to finish the proof. □

5.3 Numerical Example

We consider the original Ait-Sahalia model, that is (1.14) with $r = 2$ and $\rho = 1.5$

$$dx(t) = (\alpha_{-1}x(t)^{-1} - \alpha_0 + \alpha_1 x(t) - \alpha_2 x(t)^2)dt + \sigma x(t)^{1.5}dw(t), \tag{5.32}$$

where $(\alpha_{-1}, \alpha_0, \alpha_1, \alpha_2, \sigma) = (0.00107, 0.0517, 0.877, 4.604, 0.64754)$ are taken from (Ait-Sahalia 1999). The Assumption 5.2.2 is slightly violated here, but as the numerical experiment demonstrates strong convergence appears to hold.
According to Lemma 5.2.5, BEM

$$X_{t_{k+1}} = X_{t_k} + \left(\alpha_{-1}X_{t_{k+1}}^{-1} - \alpha_0 + \alpha_1 X_{t_{k+1}} - \alpha_2 X_{t_{k+1}}^2\right)\Delta t + \sigma X_{t_k}^{1.5}\Delta w_{t_k}, \tag{5.33}$$

admits a unique positive solution. In order to do computer simulations, on each step of the recurrence (5.33) we need to find the positive root

$$X_{t_{k+1}} - \alpha_{-1}X_{t_{k+1}}^{-1}\Delta t + \alpha_0\Delta t - \alpha_1 X_{t_{k+1}}\Delta t + \alpha_2 X_{t_{k+1}}^2\Delta t - B = 0, \tag{5.34}$$

where $B = \sigma X_{t_k}^{1.5}\Delta w_{t_k}$. In this case, we can find the inverse of the function (3.32) explicitly. Indeed, we can rewrite (5.34) in the following form

$$\alpha_2 X_{t_{k+1}}^3\Delta t + (1 - \alpha_1\Delta t)X_{t_{k+1}}^2 + (\alpha_0\Delta t - B)X_{t_{k+1}} - \alpha_{-1}\Delta t = 0.$$

Due to Lemma 5.2.5 we choose the real positive solution of the above cubic equation. This observation demonstrates that implicit schemes do not necessary increase computational complexity in comparison to classical explicit procedures. In our numerical experiment, we focus on the error at the endpoint $T = 1$, so we let

$$e_{\Delta t}^{strong} = \mathbb{E}\left|x(T) - X_T\right|.$$

We plot $e_{\Delta t}^{strong}$ against Δt on log-log scale. Error bars representing 95% confidence intervals are shown by circles, and a reference line of slop 1/2 is also given.

Although we do not know the explicit form of the solution to (5.32), Theorem 5.2.13 guarantees that BEM (5.33) strongly converges to the true solution. Therefore, it is reasonable to take BEM with very small time step, we choose $\Delta t = 2^{-15}$, as a reference solution. We then compare it with BEM evaluated

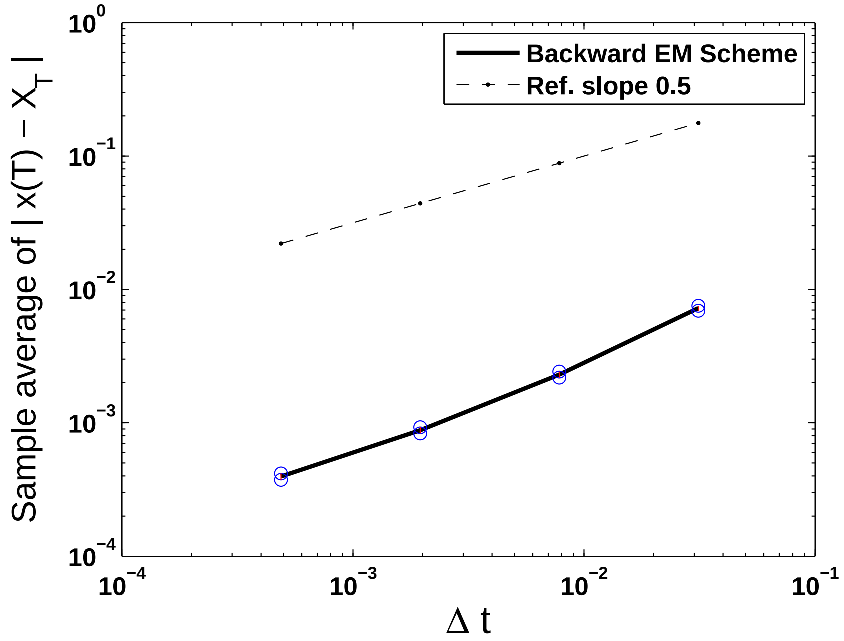

Figure 5.2: Strong error plot backward Euler-Maruyama Scheme applied to Ait-Sahalia interest rate model.

with $(2^4\Delta t, 2^6\Delta t, 2^8\Delta t, 2^{10}\Delta t)$ in order to estimate the rate of convergence. Since we are using Monte Carlo method, the sampling error decays like $1/\sqrt{M}$, M- is a number of sample paths. We set $M = 1000$. From the Figure 6.1 we see that there appears to exist a positive constant such that

$$e_{\Delta t}^{strong} \leq C\Delta t^{\frac{1}{2}} \quad \text{for sufficiently small } \Delta t.$$

A least squares fit for $\log C$ and q produced a value of 0.6984 for q with a least square residual of 0.1732.

5.4 Corollary on Option Valuation

Theorem 5.2.13 is relevant in any context where the SDE (5.17) is to be simulated numerically. For example, sample paths may be needed within a model calibration exercise. Furthermore, the SDE may represent an asset on which an option is to be valued. It is shown in (Higham and Mao 2005; Giles, Higham, and Mao

2009) that for many path dependent options, strong convergence of the SDE asset simulation guarantees convergent Monte Carlo simulations for the option value. For example, an up-and-out call gives a European payoff if the asset never exceeds the fixed barrier, B, where $B > K$ and K is the exercise price; otherwise it pays zero. The payoff at the expiry date T thus has the form

$$P = \mathbb{E}\left[(x(T) - K)^{+}\mathbf{1}_{\{\sup_{0\le t\le T} x(t)<B\}}\right].$$

Accordingly, we may define the approximate payoff based on the numerical method (5.25), to be

$$P_{\Delta t} = \mathbb{E}\left[(X_{\eta(T)} - K)^{+}\mathbf{1}_{\{\sup_{0\le t\le T} X_{\eta(t)}<B\}}\right].$$

It then follows from Theorem 5.2.13 that

$$\lim_{\Delta t\to 0}|P - P_{\Delta t}| = 0.$$

Chapter 6

Double Implicit Milstein Scheme

> What we know is not much. What we do not know is immense.
>
> Pierre Simon Laplace

So far we have only considered numerical approximations with an order of accuracy one-half. It is very interesting to investigate how higher order approximations perform in a super-linear setting. In this chapter we consider the strong convergence and stability of a Milstein type approximation to the solution of stochastic differential equations with highly nonlinear coefficients. Typically, in order to prove convergence of the Milstein scheme, stricter assumptions than those for EM are required (Kloeden and Platen 1992). What is more, it was demonstrated by Higham (Higham 2000) that the Milstein scheme applied to a linear scalar SDE has much worse stability properties than Euler-Maruyama, even once we allow for implicitness in the drift. Similarly to EM, the classical Milstein scheme does not preserve positivity. In order to address the issues mentioned above, we introduce a double implicit Milstein scheme and show that it possesses some desirable properties. It preserves positivity for a rich family of financial stochastic differential models and it can reproduce stability behaviour of the underlying SDEs without severe restriction on the time step. Although drift implicit Milstein was studied in (Kahl, Gunther, and Rosberg 2008), questions of convergence and stability remain unanswered. So far, convergence of the Milstein scheme, to the best of our knowledge, has been analyzed only under a global Lipschitz condition (Kloeden and Platen 1992). By allowing additional implicitness we are able to significantly relax the conditions required for strong convergence and therefore cover many important stochastic differential financial models encountered in the literature. Although the scheme is implicit in general, we point out examples of financial models where an explicit formula for the solution to the scheme can be found. Our results apply directly to the case of Multi-level Monte Carlo simulations for nonlinear SDEs.

6.1 Problem Specification

In contrast to previous chapters of this thesis, here we consider a scalar stochastic differential equation

$$dx(t) = f(x(t))dt + g(x(t))dw(t). \tag{6.1}$$

Here $x(t) \in \mathbb{R}$ for each $t \geq 0$. We assume that $f \in C^1(\mathbb{R}, \mathbb{R})$ and $g \in C^3(\mathbb{R}, \mathbb{R})$. The reason we restrict ourselves to the scalar case is that higher order methods for general SDEs, especially in the strong sense, carry additional difficulties. It is well known that in the general multidimensional case, such as stochastic volatility models and correlated multidimensional SDEs, there is no exact solution for iterated integrals of second order (Lévy Areas) which appear in Itô-Taylor expansions (Kloeden and Platen 1992; Glasserman 2003).

We introduce a new (θ, σ)-Milstein-scheme for a general scalar SDE. Given any step size Δt, we define the partition $\mathcal{P}_{\Delta t} := \{t_k = k\Delta t : k = 0, 1, 2, ...\}$ of the half line $[0, \infty)$. The (θ, σ)-Milstein-scheme then has the following form

$$\begin{aligned} X_{t_{k+1}} &= X_{t_k} + \theta f(X_{t_{k+1}})\Delta t + (1-\theta) f(X_{t_k})\Delta t + g(X_{t_k})\Delta W_{t_k} + \frac{1}{2}L^1 g(X_{t_k})\Delta W_{t_k}^2 \\ &\quad - \frac{(1-\sigma)}{2}L^1 g(X_{t_k})\Delta t - \frac{\sigma}{2}L^1 g(X_{t_{k+1}})\Delta t, \end{aligned} \tag{6.2}$$

where $0 \leq \theta, \sigma \leq 1$ are free paraments and $L^1 = g\frac{\partial}{\partial x}$. We can notice that the $(0, 0)$-Milstein scheme reduces to classical Milstein (Milstein and Tretyakov 2004). We sometimes refer to $(1, 1)$-Milstein as a double implicit scheme.

6.2 Existence of a Solution for the Implicit Schemes

Before we prove the existence of a unique positive solution to (6.2), we demonstrate how we can prove the existence of a solution (not necessarily positive) to (6.2). It will be then clear what assumptions will guarantee positivity. In order to prove that the (θ, σ)-Milstein (6.2) scheme is well defined we impose the following conditions.

Assumption 6.2.1. *Coefficients f and g of the equation* (6.1) *satisfy the following two conditions:*
One-sided Lipschitz condition. There exists a constant $K > 0$, such that

$$(x-y)(f(x)-f(y)) \leq K\,|x-y|^2 \quad \textit{for all} \quad x, y \in \mathbb{R}. \tag{6.3}$$

Monotone condition. Operator L^1 acting on g satisfies the following property

$$(x-y)(L^1 g(x) - L^1 g(y)) \geq 0 \quad \textit{for all} \quad x, y \in \mathbb{R}. \tag{6.4}$$

Remark 6.2.2. *From Assumption 6.2.1 we immediately have that*

$$xf(x) \leq K\left|x^2\right| + xf(0) \leq \frac{2K+1}{2}\,|x|^2 + \frac{1}{2}\,|f(0)|^2$$

and

$$xL^1 g(x) \geq xL^1 g(0).$$

Lemma 6.2.3. *Define, for any given $\Delta t < \frac{2}{\theta(2K+1)}$,*

$$F(x) = x - \theta f(x)\Delta t + \frac{\sigma}{2} L^1 g(x), \quad x \in \mathbb{R}.$$

Then under Assumption 6.2.1, for any $b \in \mathbb{R}$, there exists a unique $x \in \mathbb{R}$ such that

$$F(x) = b.$$

Proof. In view of Lemma 3.3.1 it is enough to show that the function F is continuous, coercive and strictly monotone. Clearly, $F(x)$ is continuous on $\mathbb{R}$. By Assumption 6.2.1,

$$(x-y)(F(x)-F(y)) \geq |x-y|^2 - \theta K \Delta t\,|x-y|^2 = (1-\theta K\Delta t)\,|x-y|^2 > 0,$$

for $\Delta t < \frac{2}{\theta(2K+1)}$. Also, by Assumption 6.2.1

$$\begin{aligned} xF(x) &= x(x - \theta f(x)\Delta t + \frac{\sigma}{2} L^1 g(x)\Delta t) \\ &\geq |x|^2 (1 - \theta \frac{2K+1}{2}\Delta t) - \frac{\theta}{2} x f(0)\Delta t + \frac{\sigma}{2} x L^1 g(0)\Delta t \end{aligned} \tag{6.5}$$

and coercivity follows. The proof is therefore complete. □

From now on we always assume that $\Delta t < \frac{2}{\theta(2K+1)}$.

6.2.1 Existence of a Positive Solution for (θ, σ)-Milstein Scheme.

In this subsection we introduce assumptions on coefficients f and g of the equation (6.1) that allow us to prove the existence of a positive solution to (6.2).

Definition 6.2.4. *Given $x(0) > 0$, if the solution of* (6.1) *satisfies*

$$P(\{x(t) > 0 : t > 0\}) = 1, \tag{6.6}$$

then a stochastic integration scheme to compute approximations $X_{t_k} = x(t_k)$ preserves positivity if

$$P(\{X_{t_{k+1}} > 0 | X_{t_k} > 0\}) = 1. \tag{6.7}$$

Let us notice that to prove the existence of positive solution the implicit scheme we need to assume that the one-sided Lipschitz condition on f and the monotone condition on $L^1 g$ hold on positive domain only. This significantly relaxes the conditions required for the existence and uniqueness of the solution to the implicit scheme (6.2).

Assumption 6.2.5. *Coefficients f and g of the equation* (6.1) *satisfy the following two conditions:*
One-sided Lipschitz condition. *There exists a constant $K > 0$, such that*

$$(x - y)(f(x) - f(y)) \leq K |x - y|^2 \quad \textit{for all} \quad x, y \in \mathbb{R}_+. \tag{6.8}$$

Monotone condition. *Operator L^1 acting on g satisfies the following property*

$$(x - y)(L^1 g(x) - L^1 g(y)) \geq 0 \quad \textit{for all} \quad x, y \in \mathbb{R}_+. \tag{6.9}$$

Many mean-reverting models with super- and sup-linear diffusion coefficients satisfy these conditions. For example, the mean-reverting function

$$f(x) = (\mu - x^p) \quad \text{for} \quad p > 0, \quad \mu \in \mathbb{R},$$

satisfies (6.8), but not (6.3). Condition (6.3) holds only in case when p is an odd number. The good example of function that satisfies (6.9) is polynomial function

$$g(x) = x^p \quad \text{for} \quad p \geq 0.5, \quad x \geq 0.$$

Assumption 6.2.6. *The coefficients f and g satisfy the following conditions:*

$$L^1 g(x) > 0 \qquad x > 0, \tag{6.10}$$

$$-\theta f_0(0) + L^1 g(0) \leq 0. \tag{6.11}$$

Theorem 6.2.7. *Let Assumptions 6.2.5 and 6.2.6 hold. Then there exists a unique positive solution to (θ, σ)-Milstein scheme* (6.2) *if*

$$x - \frac{g(x)}{2g'(x)} + (1-\theta)f(x)\Delta t - \frac{(1-\sigma)}{2} L^1 g(x)\Delta t > 0, \qquad x > 0. \tag{6.12}$$

Proof. In view of Lemma 6.2.3 we define an operator F as

$$F(x) = x - \theta f(x)\Delta t + \frac{1}{2}\sigma L^1 g(x)\Delta t. \tag{6.13}$$

From (6.5) and Assumption 6.2.6 operator F is monotone on $(0, \infty)$, and has a property

$$\lim_{x\to\infty} F(x) = \infty \quad \text{and} \quad \lim_{x\to 0} F(x) \leq 0. \tag{6.14}$$

Due to Lemma 3.3.1, to finish the proof we need to show that

$$b(x) = x + (1-\theta)f(x)\Delta t + g(x)\Delta W_{t_{k+1}} + \frac{1}{2}L^1 g(x)\Delta W^2_{t_{k+1}} - \frac{(1-\sigma)}{2}L^1 g(x)\Delta t > 0, \tag{6.15}$$

from which it follows that there exists a positive solution to

$$F(X_{t_{k+1}}) = b(X_{t_k}). \tag{6.16}$$

First, we find the minimum of the function

$$H(y) = g(x)y + \frac{1}{2}L^1 g(x)y^2. \tag{6.17}$$

Under assumption $L^1 g(x) > 0$ this function possesses a global minimum:

$$\begin{aligned} H(y) &= \frac{1}{2}L^1 g(x)\left[y^2 + 2y/g^{'}(x)\right] \\ &= \frac{1}{2}L^1 g(x)\left[(y + 1/g^{'}(x))^2 - (1/g^{'}(x))^2\right] \geq \frac{-g(x)}{2g^{'}(x)}. \end{aligned}$$

Hence

$$b(x) \geq x + (1-\theta)f(x)\Delta t - \frac{(1-\sigma)}{2}L^1 g(x)\Delta t - \frac{g(x)}{2g^{'}(x)} > 0, \tag{6.18}$$

as required. □

6.2.2 Positivity Preserving Approximation of Heston Volatility Model

Here we demonstrate that approximation of the well-known 3/2-Heston volatility model with the double implicit Milstein scheme preserves positivity. We would like to point out that implicitness in the numerical approximation does not increase computational cost in this case, since we are able to find an explicit solution. For the 3/2 model

$$dx(t) = x(t)(\mu - \alpha x(t))dt + \beta x^{3/2} dw(t),$$

the $(1, 1)$-Milstein scheme has the form

$$\begin{aligned} X_{t_{k+1}} = X_{t_k} &+ f(X_{t_{k+1}})\Delta t + g(X_{t_k})\Delta W_{t_k} \\ &+ \frac{1}{2}L^1 g(X_{t_k})\Delta W_{t_k}^2 - \frac{1}{2}L^1 g(X_{t_{k+1}})\Delta t, \end{aligned} \tag{6.19}$$

with $f(x) = \mu x - \alpha x^2$ and $g(x) = \beta x^{3/2}$. In order to show that the solution to (6.19) is always positive we need to analyse the functions $F(x)$ and $b(x)$, where

$$F(x) = (1 - \mu\Delta t)x + (\alpha + \frac{3}{4}\beta^2)x^2\Delta t,$$

and

$$b(x) = x\big(x + \frac{1}{2}\beta x^{1/2} w_{t_k}\big)^2 + 1/4\beta^2 x^2 w_{t_k}^2 > 0.$$

We can easily verify that the assumptions of Theorem 6.4.2 hold. An explicit formula for $X_{t_{k+1}}$ can found by solving the relevant quadratic equation and choosing the positive solution.

6.3 Moment Bounds for Double Implicit Milstein Scheme

In this section we prove the boundedness of second moments for the double implicit Milstein

$$\begin{aligned} X_{t_{k+1}} &= X_{t_k} + f(X_{t_{k+1}})\Delta t + g(X_{t_k})\Delta w_{t_k} \\ &\quad + \frac{1}{2}L^1 g(X_{t_k})\Delta w_{t_k}^2 - \frac{\mathbf{1}}{2}L^1 g(X_{t_{k+1}})\Delta t. \end{aligned} \tag{6.20}$$

As in previous chapters such bounds are necessary for our proof of strong convergence. Having financial applications in mind we derive the strong convergence result under conditions which guarantee positivity of the process X_{t_k} for all $k \geq 1$. In addition, we assume that the following polynomial growth condition holds.

Assumption 6.3.1. *The coefficients of equation (6.1) satisfy a polynomial growth condition, that is, for some $h > 0$ there exists a positive constant $H > 0$ such that*

$$|f(x)| \vee |g(x)| \vee \left|L^1 g(x)\right| \leq H(1 + |x|^h), \qquad \forall x \in \mathbb{R}. \tag{6.21}$$

Lemma 6.3.2. *Under Assumptions 6.2.5, 6.2.6 and 6.3.1, for any integer $p \geq 2$ and sufficiently large integer m, there exists a constant $C(p, m)$, such that the $(1, 1)$-Milstein scheme has the property*

$$\mathbb{E}\left[|X_{t_k}|^p \mathbf{1}_{[0,\lambda_m]}(k)\right] < C(p, m) \qquad \text{for any} \quad k \geq 0,$$

where

$$\lambda_m = \inf\{k : |X_{t_k}| > m\}. \tag{6.22}$$

Proof. We observe that when $k \in [0, \lambda_m]$, $X_{t_{k-1}} \in (-m, m)$, but X_{t_k} may not stay in $(-m, m)$, so the lemma is not obvious. By definition of the (1,1)-Milstein scheme

$$\begin{aligned}|X_{t_k}|^2 &= X_{t_k} X_{t_{k-1}} + X_{t_k} f(X_{t_k})\Delta t + X_{t_k} g(X_{t_{k-1}})\Delta w_{t_{k-1}} \\ &+ \frac{1}{2} X_{t_k} L^1 g(X_{t_{k-1}})\Delta w^2_{t_{k-1}} - \frac{\mathbf{1}}{2} L^1 X_{t_k} g(X_{t_k})\Delta t.\end{aligned}$$

Utilizing Young's inequality in the form

$$xy \leq \frac{\delta}{2}x^2 + \frac{1}{2\delta}y^2,$$

where $\delta > 0$ is small enough, such that $\delta(\frac{5}{4} + \frac{7(p-2)}{4p}) + \frac{2K+1}{2}\Delta t = C(\delta, \Delta t) < 1$, we have

$$\begin{aligned}|X_{t_k}|^2 &\leq \frac{\delta}{2}|X_{t_k}|^2 + \frac{1}{2\delta}\left|X_{t_{k-1}}\right|^2 \\ &+ X_{t_k} f(X_{t_k})\Delta t + \frac{\delta}{2}|X_{t_k}|^2 + \frac{1}{2\delta}\left|g(X_{t_{k-1}}, t_{k-1})\Delta w_{t_{k-1}}\right|^2 \\ &+ \frac{\delta}{4}|X_{t_k}|^2 + \frac{1}{4\delta}\left|L^1 g(X_{t_{k-1}}, t_{k-1})\Delta w^2_{t_{k-1}}\right|^2 - \frac{\mathbf{1}}{2} X_{t_k} L^1 g(X_{t_k})\Delta t.\end{aligned}$$

Due to Remark 6.2.2 and Assumption 6.2.6

$$\begin{aligned}|X_{t_k}|^2 &\leq (\frac{5\delta}{4} + \frac{2K+1}{2}\Delta t)\,|X_{t_k}|^2 + \frac{1}{2\delta}\left|X_{t_{k-1}}\right|^2 \\ &+ \frac{1}{2}|f(0)|^2\Delta t + \frac{1}{2\delta}\left|g(X_{t_{k-1}})\Delta w_{t_{k-1}}\right|^2 \\ &+ \frac{1}{4\delta}\left|L^1 g(X_{t_{k-1}})\Delta w^2_{t_{k-1}}\right|^2.\end{aligned}$$

Multiplying both sides of the above inequality by $|X_{t_k}|^{p-2}$ leads to

$$\begin{aligned}|X_{t_k}|^p \leq (\frac{5\delta}{4} + \frac{2K+1}{2}\Delta t)\,|X_{t_k}|^p + \frac{1}{2\delta}\left|X_{t_{k-1}}\right|^2 |X_{t_k}|^{p-2} \\ + \frac{1}{2}\,|f(0)|^2\,\Delta t\,|X_{t_k}|^{p-2} + \frac{1}{2\delta}\left|g(X_{t_{k-1}})\Delta w_{t_{k-1}}\right|^2 |X_{t_k}|^{p-2} \\ + \frac{1}{4\delta}\left|L^1 g(X_{t_{k-1}})\Delta w_{t_{k-1}}^2\right|^2 |X_{t_k}|^{p-2}.\end{aligned}$$

Applying Young's inequality in the form

$$x^{p-2}y^2 < \delta^2\frac{p-2}{p}x^p + \frac{2}{p\delta^{p-2}}y^p,$$

results in

$$\begin{aligned}|X_{t_k}|^p \leq (\frac{5\delta}{4} + \frac{2K+1}{2}\Delta t)\,|X_{t_k}|^p + \frac{1}{2\delta}\left(\frac{2}{p\delta^{p-2}}\left|X_{t_{k-1}}\right|^p + \delta^2\frac{p-2}{p}|X_{t_k}|^p\right) \\ + \frac{1}{2\delta}\left(\frac{2}{p\delta^{p-2}}\left|(\delta\Delta t)^{1/2}f(0)\right|^p + \delta^2\frac{p-2}{p}|X_{t_k}|^p\right) \\ + \frac{1}{2\delta}\left(\frac{2}{p\delta^{p-2}}\left|g(X_{t_{k-1}})\Delta w_{t_{k-1}}\right|^p + \delta^2\frac{p-2}{p}|X_{t_k}|^p\right) \\ + \frac{1}{4\delta}\left(\frac{2}{p\delta^{p-2}}\left|L^1 g(X_{t_{k-1}})\Delta w_{t_{k-1}}^2\right|^p + \delta^2\frac{p-2}{p}|X_{t_k}|^p\right).\end{aligned}$$

Rearranging, we have

$$\begin{aligned}|X_{t_k}|^p \leq (\frac{5\delta}{4} + \frac{2K+1}{2}\Delta t + \delta\frac{7(p-2)}{4p})\,|X_{t_k}|^p + \frac{1}{p\delta^{p-1}}\Bigg(\left|X_{t_{k-1}}\right|^p + \left|(\delta\Delta t)^{1/2}f(0)\right|^p \\ + \left|g(X_{t_{k-1}})\Delta w_{t_{k-1}}\right|^p + \frac{1}{2}\left|L^1 g(X_{t_{k-1}})\Delta w_{t_{k-1}}^2\right|^p\Bigg).\end{aligned}$$

Hence

$$\begin{aligned}(1 - C(\delta,\Delta t))\mathbb{E}\left[|X_{t_k}|^p\,\mathbf{1}_{[0,\lambda_m]}(k)\right] \leq \frac{1}{p\delta^{p-1}}\mathbb{E}\Bigg(\left|X_{t_{k-1}}\right|^p + \left|(\delta\Delta t)^{1/2}f(0,t)\right|^p \\ + \left|g(X_{t_{k-1}})\Delta w_{t_{k-1}}\right|^p + \frac{1}{2}\left|L^1 g(X_{t_{k-1}})\Delta w_{t_{k-1}}^2\right|^p\Bigg)\mathbf{1}_{[0,\lambda_m]}(k),\end{aligned}$$

where $C(\delta, \Delta t) = \frac{5\delta}{4} + \frac{2K+1}{2}\Delta t + \delta\frac{7(p-2)}{4p}$. By Hölder's inequality,

$$(1 - C(\delta, \Delta t))\mathbb{E}\left[|X_{t_k}|^p \mathbf{1}_{[0,\lambda_m]}(k)\right] \leq \frac{1}{p\delta^{p-1}}\left(|m|^p + \left|(\delta\Delta t)^{1/2} f(0)\right|^p\right)$$

$$+ \frac{1}{p\delta^{p-1}}\left((\mathbb{E}(\left|g(X_{t_{k-1}})\right|^p \mathbf{1}_{[0,\lambda_m]}(k))^2)^{1/2}(\mathbb{E}\left|\Delta w_{t_{k-1}}\right|^{2p})^{1/2}\right.$$

$$\left.+ (\mathbb{E}(\left|L^1 g(X_{t_{k-1}})\right|^p \mathbf{1}_{[0,\lambda_m]}(k))^2)^{1/2}(\mathbb{E}\left|\Delta w_{t_k}\right|^{4p})^{1/2}\right).$$

By Assumption 6.3.1 and the fact that there exists a positive constant $C(p)$, such that $\mathbb{E}\left|\Delta w_{t_{k-1}}\right|^{4p} < C(p)$, we obtain

$$\mathbb{E}\left[|X_{t_k}|^p \mathbf{1}_{[0,\lambda_m]}(k)\right] < C(m, p),$$

as required. □

Assumption 6.3.3.

$$2xf(x) + |g(x)|^2 \leq \alpha + \beta |x|^2, \quad x \in \mathbb{R}. \tag{6.23}$$

$$\frac{1}{2}L^1 g(x)(2f(x) + L^1 g(x)) - |f(x)|^2 \leq \alpha + \beta |x|^2, \quad x \in \mathbb{R}. \tag{6.24}$$

Remark 6.3.4. *Within financial applications very often we deal with polynomial coefficients. As an example we consider a general mean-reverting SDE of the form*

$$dx(t) = (\mu - x(t)^p)dt + x(t)^\rho dw(t), \quad p, \rho > 1.$$

Once $p + 1 > 2\rho$, condition (6.23) *holds. But $p + 1 > 2\rho$ also implies* (6.24).

Theorem 6.3.5. *Let Assumptions 6.2.5, 6.3.1 and 6.3.3 hold. Let also $\Delta t^* \in (0, \frac{2K+1}{2} \wedge (2\beta)^{-1})$. Then there exists a constant $C = C(T)$ such that for all $\Delta t < \Delta t^*$ in the* (1, 1)*-Milstein scheme we have*

$$\sup_{0 \leq t_k \leq T} \mathbb{E}\,|X_{t_k}|^2 < K, \quad T \geq 0.$$

Proof. Let

$$\lambda_m(k) = \inf\{k : |X_{t_k}| > m\}.$$

Then λ_m is a stopping time with respect to $\{\mathcal{F}_{t_k}\}_{k\geq 0}$. We have the following equality

$$\begin{aligned} X_{t_{k+1}} &= X_{t_k} + f(X_{t_{k+1}})\Delta t + g(X_{t_k})\Delta w_{t_k} \\ &+ \frac{1}{2}L^1 g(X_{t_k})\Delta w_{t_k}^2 - \frac{\mathbf{1}}{2}L^1 g(X_{t_{k+1}})\Delta t. \end{aligned}$$

Squaring both sides leads us to

$$\begin{aligned} &\left|X_{t_{k+1}}\right|^2 - \left|X_{t_k}\right|^2 = 2X_{t_{k+1}} f(X_{t_{k+1}})\Delta t + \left|g(X_{t_k})\right|^2 \left|\Delta w_{t_k}\right|^2 \\ &+ X_{t_k} L^1 g(X_{t_k})\Delta w_{t_k}^2 - X_{t_{k+1}} L^1 g(X_{t_{k+1}})\Delta t \\ &- \frac{1}{4}\left|L^1 g(X_{t_{k+1}})\right|^2 \Delta t^2 + \frac{1}{4}\left|L^1 g(X_{t_k})\Delta w_{t_k}^2\right|^2 \\ &+ f(X_{t_{k+1}})L^1 g(X_{t_{k+1}})\Delta t^2 - \left|f(X_{t_{k+1}})\right|^2 \Delta t^2 \\ &+ 2X_{t_k} g(X_{t_k})\Delta t w_{t_k} + g(X_{t_k})L^1 g(X_{t_k})\Delta w_{t_k}^3. \end{aligned}$$

Let N be any nonnegative integer, such that $N\Delta t \leq T$. Summing up both sides of the above equality from $k = 0$ to $N \wedge \lambda_m$, we get

$$\begin{aligned}
\left|X_{t_{N\wedge\lambda_m+1}}\right|^2 = &\Bigg(|X_{t_0}|^2 + |g(X_{t_0})|^2\,|\Delta w_{t_0}|^2 + X_{t_0}L^1g(X_{t_0})\Delta w_{t_0}^2 \\
&+ \frac{1}{4}\left|L^1g(X_{t_0})\Delta w_{t_0}\right|^2 + 2X_{t_0}g(X_{t_0})\Delta t w_{t_0} + g(X_{t_0})L^1g(X_{t_0})\Delta w_{t_0}^3\Bigg) \\
&+ \sum_{k=1}^{(N\wedge\lambda_m)+1} 2X_{t_k}f(X_{t_k})\Delta t + \sum_{k=1}^{N\wedge\lambda_m} |g(X_{t_k})|^2\,\Delta t \\
&+ \sum_{k=1}^{N\wedge\lambda_m} X_{t_k}L^1g(X_{t_k})\Delta t - \sum_{k=1}^{(N\wedge\lambda_m)+1} X_{t_k}L^1g(X_{t_k})\Delta t \\
&- \sum_{k=1}^{(N\wedge\lambda_m)+1} \frac{1}{4}\left|L^1g(X_{t_k})\right|^2\Delta t^2 + \sum_{k=1}^{N\wedge\lambda_m}\frac{3}{4}\left|L^1g(X_{t_k})\right|^2\Delta t^2 \\
&+ \sum_{k=1}^{(N\wedge\lambda_m)+1} f(X_{t_k})L^1g(X_{t_k})\Delta t^2 - \sum_{k=1}^{(N\wedge\lambda_m)+1} |f(X_{t_k})|^2\,\Delta t^2 \\
&+ 2\sum_{k=1}^{N\wedge\lambda_m} X_{t_k}g(X_{t_k})\Delta t w_{t_k} + \sum_{k=1}^{N\wedge\lambda_m} g(X_{t_k})L^1g(X_{t_k})\Delta w_{t_k}^3 \\
&+ \sum_{k=1}^{N\wedge\lambda_m}\Big(|g(X_{t_k})|^2 + X_{t_k}, L^1g(X_{t_k})\Big)\left[|\Delta w_{t_k}|^2 - \Delta t\right] \\
&+ \sum_{k=1}^{N\wedge\lambda_m}\frac{1}{4}\left|L^1g(X_{t_k})\right|^2\left[|\Delta w_{t_k}|^4 - 3\Delta t^2\right].
\end{aligned}$$

Hence

$$
\begin{aligned}
\left|X_{t_{N\wedge\lambda_m+1}}\right|^2 \leq & \Bigg(|X_{t_0}|^2 + |g(X_{t_0})|^2 |\Delta w_{t_0}|^2 + X_{t_0} L^1 g(X_{t_0}) \Delta w_{t_0}^2 \\
& + \frac{1}{4}\left|L^1 g(X_{t_0})\Delta w_{t_0}\right|^2 + 2X_{t_0} g(X_{t_0}) \Delta t w_{t_0} + g(X_{t_0}) L^1 g(X_{t_0}) \Delta w_{t_0}^3 \Bigg) \\
& + \sum_{k=1}^{N} 2X_{t_k} f(X_{t_k}) \mathbf{1}_{[0,\lambda_m]}(k)\Delta t + \sum_{k=1}^{N} |g(X_{t_k})|^2 \mathbf{1}_{[0,\lambda_m]}(k)\Delta t \\
& + 2X_{t_{N\wedge\lambda_m+1}} f(X_{t_{N\wedge\lambda_m+1}})\Delta t \\
& - X_{t_{N\wedge\lambda_m+1}} L^1 g(X_{t_{N\wedge\lambda_m+1}})(k)\Delta t \\
& + \sum_{k=1}^{N} \frac{1}{2}\left|L^1 g(X_{t_k})\right|^2 \mathbf{1}_{[0,\lambda_m]}(k)\Delta t^2 \\
& - \frac{1}{4}\left|L^1 g(X_{t_{N\wedge\lambda_m+1}})\right|^2 \Delta t^2 \\
& + \sum_{k=1}^{N} f(X_{t_k}) L^1 g(X_{t_k}) \mathbf{1}_{[0,\lambda_m]}(k)\Delta t^2 - \sum_{k=1}^{N} |f(X_{t_k})|^2 \mathbf{1}_{[0,\lambda_m]}(k)\Delta t^2 \\
& + f(X_{t_{N\wedge\lambda_m+1}}) L^1 g(X_{t_{N\wedge\lambda_m+1}})\Delta t^2 - \left|f(X_{t_{N\wedge\lambda_m+1}})\right|^2 \Delta t^2 \\
& + 2\sum_{k=1}^{N} \langle X_{t_k} g(X_{t_k}) \mathbf{1}_{[0,\lambda_m]}(k)\Delta t w_{t_k} + \sum_{k=1}^{N} g(X_{t_k}) L^1 g(X_{t_k}) \mathbf{1}_{[0,\lambda_m]}(k)\Delta w_{t_k}^3 \\
& + \sum_{k=1}^{N} \Bigg(|g(X_{t_k})|^2 + X_{t_k} L^1 g(X_{t_k}) \Bigg) \mathbf{1}_{[0,\lambda_m]}(k) \left[|\Delta w_{t_k}|^2 - \Delta t \right] \\
& + \sum_{k=1}^{N} \frac{1}{4}\left|L^1 g(X_{t_k})\right|^2 \mathbf{1}_{[0,\lambda_m]}(k) \left[|\Delta w_{t_k}|^4 - 3\Delta t^2 \right].
\end{aligned}
$$

By Assumptions 6.3.3, 6.2.6 and Lemma 6.3.2, noting that X_{t_k} and $\mathbf{1}_{[0,\lambda_m]}(k)$ are $\mathcal{F}_{t_k}$-measurable while Δw_{t_k} is independent of $\mathcal{F}_{t_k}$, we can take expectation on both sides of the above inequality to get

$$\begin{aligned}\mathbb{E}\Big[\big|X_{t_{N\wedge\lambda_m+1}}\big|^2\Big] \le C_1 &+ \mathbb{E}\Big[\sum_{k=1}^{N}(2X_{t_k}f(X_{t_k}) + |g(X_{t_k})|^2)\mathbf{1}_{[0,\lambda_m]}(k)\Delta t\Big] \\ &+ \mathbb{E}\Big[2X_{t_{(N\wedge\lambda_m)+1}}f(X_{t_{(N\wedge\lambda_m)+1}})\Delta t\Big] \\ &+ \mathbb{E}\Big[\sum_{k=1}^{N}(\frac{1}{2}L^1g(X_{t_k})2f(X_{t_k}) + L^1g(X_{t_k})\mathbf{1}_{[0,\lambda_m]}(k)\Delta t^2 \\ &- |f(X_{t_k})|^2\,\mathbf{1}_{[0,\lambda_m]}(k)\Delta t^2\big) + f(X_{t_{N\wedge\lambda_m+1}})L^1g(X_{t_{N\wedge\lambda_m+1}})\Delta t^2 \\ &- \big|f(X_{t_{N\wedge\lambda_m+1}})\big|^2\,\Delta t^2\Big],\end{aligned}$$

where

$$C_1 = |X_{t_0}|^2 + \left(|g(X_{t_0})|^2 + X_{t_0}L^1g(X_{t_0}) + \frac{1}{4}\left|L^1g(X_{t_0})\right|^2\right)\Delta t^*.$$

By Assumption 6.3.3,

$$\mathbb{E}\left[\left|X_{t_{(N\wedge\lambda_m)+1}}\right|^2\right] \le C_1+\beta\Big[\sum_{k=1}^{N}E\left|X_{t_k}\right|^2\mathbf{1}_{[0,\lambda_m]}(k)\Delta t+\mathbb{E}\left|X_{t_{(N\wedge\lambda_m)+1}}\right|^2\Delta t\Big]+\alpha(T+\Delta t). \tag{6.25}$$

Then

$$\begin{aligned}\mathbb{E}\left[\left|X_{t_{(N\wedge\lambda_m)+1}}\right|^2\right] &\le (C_1 + \alpha(T+\Delta t))(1-\beta\Delta t)^{-1} \\ &+ (1-\beta\Delta t)^{-1}\beta\Big[\sum_{k=1}^{N}E\left|X_{t_k}\right|^2\mathbf{1}_{[0,\lambda_m]}(k)\Delta t\Big].\end{aligned} \tag{6.26}$$

Now we can observe that

$$\mathbb{E}\left[\left|X_{t_{N+1}}\right|^2\mathbf{1}_{[0,\lambda_m]}(N)\right] \le \mathbb{E}\left[\left|X_{t_{N\wedge\lambda_m+1}}\right|^2\right].$$

By the discrete Gronwall inequality

$$\mathbb{E}\left[\left|X_{t_{N+1}}\right|^2\mathbf{1}_{[0,\lambda_m]}(N)\right] \le (C_1 + \alpha(T+\Delta t))(1-\beta\Delta t)^{-1}\exp\left(((1-\beta\Delta t)^{-1}\beta)T\right),$$

where we use the fact that $N\Delta t \leq T$. Thus, letting $m \to \infty$ in (6.26) and applying Fatou's lemma, we get

$$\mathbb{E}\left|X_{t_{N+1}}\right|^2 \leq (C_1 + \alpha(T + \Delta t))(1 - \beta\Delta t)^{-1} \exp\left(((1 - \beta\Delta t)^{-1}\beta)T\right).$$

The proof is complete. □

6.4 Strong Convergence

In this section we prove that the double implicit Milstein scheme (6.20) strongly converges to the solution of the SDE (6.1).

6.4.1 Forward-Backward Milstein Scheme

In our analysis we found it convenient to extend the discrete time Milstein scheme to a continuous time stochastic process. This extension allows to use stochastic calculus.

In terms of the general drift and diffusion coefficients f and g, we first compute the value X_{t_k} from the Milstein scheme (6.20) , that is

$$\begin{aligned} X_{t_k} &= X_{t_{k-1}} + f(X_{t_k})\Delta t + g(X_{t_{k-1}})\Delta w_{t_{k-1}} \\ &\quad + \frac{1}{2}L^1 g(X_{t_{k-1}})\Delta w_{t_{k-1}}^2 - \frac{\mathbf{1}}{2}L^1 g(X_{t_k})\Delta t. \end{aligned}$$

Then we define the Forward-Backward Milstein scheme (FBM) as follows

$$\begin{aligned} \hat{X}_{t_{k+1}} &= \hat{X}_{t_k} + f(X_{t_k})\Delta t + g(X_{t_k})\Delta w_{t_k} \\ &\quad + \frac{1}{2}L^1 g(X_{t_k})\Delta w_{t_k}^2 - \frac{\mathbf{1}}{2}L^1 g(X_{t_k})\Delta t \end{aligned} \tag{6.27}$$

where $\hat{X}_{t_0} = X_{t_0} = x_0$.

6.4.2 Continuous Milstein Scheme

In this subsection we define a continuous extension of FBM scheme (6.27). First we observe that

$$\int_{t_k}^{t_{k+1}} \int_{t_k}^{s} dw(z)dw(s) = \frac{1}{2}[\Delta w_{t_k}^2 - \Delta t].$$

Let us introduce the notation

$$\eta(t) := t_k, \quad \text{for} \quad t \in [t_k, t_{k+1}),\ k \geq 0,$$

and define the process $m(t)$

$$m(t) = \int_0^t z(s)dw(s),$$

with

$$z(s) = 2(w(s) - w(\eta(s))).$$

By the martingale representation theorem $m(t)$ is clearly a martingale. We also observe that

$$m(t_{k+1}) - m(t_k) = \Delta w_{t_k}^2 - \Delta t,$$

and

$$\mathbb{E}\left[\int_0^t z(s)dw(s)\right] = 0.$$

Therefore we can define a continuous extension of FBM scheme (6.27) as follows

$$\hat{X}(t) = \hat{X}(0) + \int_0^t f(X_{\eta(s)})ds + \int_0^t g(X_{\eta(s)})dw(s) + \frac{1}{2}\int_0^t L^1 g(X_{\eta(s)})z(s)dw(s). \tag{6.28}$$

Note that the continuous and discrete FBM coincide at the grid-points; that is, $\hat{X}(t_k) = \hat{X}_{t_k}$.

6.4.3 Strong Convergence on a Compact Domain

We begin by showing that the FBM (6.27) and Milstein scheme (6.20) stay close on a compact domain. Then we estimate the probability that CFBM (6.28) will not explode on a finite time interval.

Lemma 6.4.1. *Let $\Delta t^* \in (0, \frac{2K+1}{2} \wedge (2\beta)^{-1})$. Under Assumptions 6.2.5 and 6.3.1, for any integer $p \geq 2$ and $T \geq 0$, there exists a constant $C = C(m,p,T)$ such that for all $\Delta t \leq \Delta t^*$,*

$$\mathbb{E}\left[\left|\hat{X}_{t_k} - X_{t_k}\right|^p \mathbf{1}_{[0,\lambda_m]}(k)\right] \leq C\Delta t^p, \qquad \forall t_k \in [0,T].$$

Proof. Summing up (1,1)-Milstein and FBM schemes, respectively, we obtain

$$\begin{aligned} X_{t_N} = X_{t_0} &+ \sum_{k=0}^{N-1} f(X_{t_{k+1}})\Delta t + \sum_{k=0}^{N-1} g(X_{t_k})\Delta w_{t_k} \\ &+ \sum_{k=0}^{N-1} \frac{1}{2} L^1 g(X_{t_k})\Delta w_{t_k}^2 - \sum_{k=0}^{N-1} \frac{\mathbf{1}}{2} L^1 g(X_{t_{k+1}})\Delta t \end{aligned}$$

and

$$\begin{aligned} \hat{X}_{t_N} = \hat{X}_{t_0} &+ \sum_{k=0}^{N-1} f(X_{t_k})\Delta t + \sum_{k=0}^{N-1} g(X_{t_k})\Delta w_{t_k} \\ &+ \sum_{k=0}^{N-1} \frac{1}{2} L^1 g(X_{t_k})\Delta w_{t_k}^2 - \sum_{k=0}^{N-1} \frac{\mathbf{1}}{2} L^1 g(X_{t_k})\Delta t. \end{aligned}$$

Now by Hölder's inequality, Lemma 6.3.2 and Assumption 6.3.1, there exists a constant $C = C(m,p,T)$ such that

$$\begin{aligned} &\mathbb{E}\left[\left|\hat{X}_{t_N} - X_{t_N}\right|^p \mathbf{1}_{[0,\lambda_m]}(N)\right] \qquad (6.29) \\ &= \mathbb{E}\left[\left|f(X_{t_0}) - f(X_{t_N}) + L^1 g(X_{t_0}) - L^1 g(X_{t_N})\right|^p \mathbf{1}_{[0,\lambda_m]}(N)\right] \Delta t^p \leq C\Delta t^p, \end{aligned}$$

as required. ☐

Theorem 6.4.2. *Let Assumptions of Theorem and 6.3.1, 6.3.3 hold and $T > 0$ be arbitrary. Then, for any given $\epsilon > 0$, there exists an N_0 such that for every $m \geq N_0$, we can find a $\Delta t_0 = \Delta t_0(m)$ so that whenever $\Delta t \leq \Delta t_0$,*

$$\mathbb{P}(\vartheta_m < T) \leq \epsilon,$$

where $\vartheta_m = \inf\{t > 0 : \left|\hat{X}(t)\right| \geq m \quad \text{or} \quad \left|X_{\eta(t)}\right| > m\}.$

Proof. Let $s \in [0, T \wedge \vartheta_m)$. Then by the Itô formula with $V(x) = |x|^2$,

$$
\begin{aligned}
d\left|\hat{X}(s)\right|^2 &\le 2\hat{X}(s)f(X_{\eta(s)})ds + \left|g(X_{\eta(s)})\right|^2 ds + \frac{1}{4}\left|L^1 g(X_{\eta(s)})z(s)\right|^2 ds \\
&+ 2\hat{X}(s)g(X_{\eta(s)})dw(s) \\
&= LV(X_{\eta(s)}))ds + 2(\hat{X}(s) - X_{\eta(s)})f(X_{\eta(s)})ds \\
&+ \frac{1}{4}\left|L^1 g(X_{\eta(s)})z(s)\right|^2 ds + 2\hat{X}(s)g(X_{\eta(s)})dw(s) \\
&\le LV(X_{\eta(s)}))ds + 2\left|\hat{X}(s) - X_{\eta(s)}\right|\left|f(X_{\eta(s)})\right| ds \\
&+ \frac{1}{4}\left|L^1 g(X_{\eta(s)})\right|^2 |z(s)|^2 ds + 2\hat{X}(s)g(X_{\eta(s)})dw(s),
\end{aligned}
$$

where the diffusion operator is defined by $LV(x) = 2xf(x) + |g(x)|^2$. By Local Lipschitz continuity and Assumptions 6.2.5 and 6.2.6, for $|x| \le m$ there exists a positive constant K_m, such that

$$
|f(x)|^2 \le 2(|f(x) - f(0)|^2 + |f(0)|^2) \le 2K_m |x|^2
$$

$$
|g(x)|^2 \le 2(|g(x) - g(0)|^2 + |g(0)|^2) \le 2(K_m |x|^2 + |g(0)|^2)
$$

and

$$
\left|L^1 g(x)\right|^2 \le 2(\left|L^1 g(x) - L^1 g(0)\right|^2 + \left|L^1 g(0)\right|^2) \le 2K_m |x|^2 .
$$

By Assumption 6.3.1 and the fact that $w(s) - w(\eta(s))$ is a normally distributed random variable

$$
\begin{aligned}
\mathbb{E}\int_0^{T\wedge\vartheta_m} \left|L^1 g(X_{\eta(s)})\right|^2 |z(s)|^2 ds &= \mathbb{E}\int_0^{T\wedge\vartheta_m} \left|L^1 g(X_{\eta(s)})\right|^2 |w(s) - w(\eta(s))|^2 ds \\
&\le \mathbb{E}\int_0^T \mathbf{1}_{\{s<\vartheta_m\}} H(1 + \left|X_{\eta(s)}\right|^h) |w(s) - w(\eta(s))|^2 ds \\
&\le C(m)\int_0^T \mathbb{E}\,|w(s) - w(\eta(s))|^2 ds \\
&\le C(m)\int_0^T \Delta t ds.
\end{aligned}
$$

Recalling that $LV(x) < (\alpha + \beta |x|^2)$, we then have

$$\mathbb{E}\left|\hat{X}(T \wedge \vartheta_m)\right|^2 \le \left|\hat{X}(0)\right|^2 + \alpha T + 2\beta \int_0^T \mathbb{E}\left|\hat{X}(s \wedge \vartheta_m)\right|^2 ds$$
$$+ C(m,T)\mathbb{E}\int_0^{T\wedge\vartheta_m} \left|X_{\eta(s)} - \hat{X}(s)\right| ds + C(m,T)\Delta t ds,$$

where we use the fact that

$$2\int_0^{T\wedge\vartheta_m} \left|X_{\eta(s)} - \hat{X}(s)\right|^2 ds \le 2\int_0^{T\wedge\vartheta_m} \left|X_{\eta(s)} + \hat{X}(s)\right| \left|X_{\eta(s)} - \hat{X}(s)\right| ds$$
$$\le C(m)\int_0^{T\wedge\vartheta_m} \left|X_{\eta(s)} - \hat{X}(s)\right| ds.$$

By Lemma 6.4.1, we obtain

$$\mathbb{E}\int_0^{T\wedge\vartheta_m} \left|X_{\eta(s)} - \hat{X}_{\eta(s)}\right| ds \le C(m,T)\Delta t. \tag{6.30}$$

To bound the term $\mathbb{E}\int_0^{T\wedge\vartheta_m} \left|\hat{X}_{\eta(s)} - \hat{X}(s)\right| ds$, given $s \in [0, T\wedge\vartheta_m)$, let k be an integer for which $s \in [t_k, t_{k+1})$. Then

$$\left|\hat{X}_{\eta(s)} - \hat{X}(s)\right| = \left|\int_s^{t_{k+1}} f(X_{t_k}) ds + \int_s^{t_{k+1}} g(X_{t_k}) dw(s)\right|.$$

By Hölder's inequality

$$\mathbb{E}\int_0^{T\wedge\vartheta_m} \left|\hat{X}_{\eta(s)} - \hat{X}(s)\right| ds \le C(m,T)\Delta t^{\frac{1}{2}},$$

where $C(m,T) > 0$ is constant. This leads us to

$$E\int_0^{T\wedge\vartheta_m} \left|X_{\eta(s)} - \hat{X}(s)\right| ds \le \mathbb{E}\int_0^{T\wedge\vartheta_m} \left|\hat{X}_{\eta(s)} - \hat{X}(s)\right| ds$$
$$+ \mathbb{E}\int_0^{T\wedge\vartheta_m} \left|X_{\eta(s)} - \hat{X}_{\eta(s)}\right| ds$$
$$\le C(m,T)\Delta t^{\frac{1}{2}}. \tag{6.31}$$

Therefore

$$\mathbb{E}\left|\hat{X}(t\wedge\vartheta_m)\right|^2 \leq \left|\hat{X}(0)\right|^2+\alpha T+C(m,T)\Delta t^{\frac{1}{2}}+C(m,T)\Delta t+2\beta\int_0^T \mathbb{E}\left|\hat{X}(s\wedge\vartheta_m)\right|^2 ds.$$

By Gronwall's inequality

$$\mathbb{E}\left|\hat{X}(t\wedge\vartheta_m)\right|^2 \leq [\left|\hat{X}(0)\right|^2 + \alpha T + C(m,T)\Delta t^{\frac{1}{2}} + C(m,T)\Delta t]\exp(2\beta T), \quad (6.32)$$

which implies that

$$\mathbb{P}(\vartheta_m < T) \leq \frac{[\left|\hat{X}(0)\right|^2 + \alpha T + C(m,T)\Delta t^{1/2} + C(m,T)\Delta t]\exp(2\beta T)}{|m|^2}.$$

Now, for any given $\epsilon > 0$, we choose N_0 such that for any $m \geq N_0$

$$\frac{[\left|\hat{X}(0)\right|^2 + \alpha T]\exp(2\beta T)}{|m|^2} \leq \frac{\epsilon}{2}.$$

Then, we can choose $\Delta t_0 = \Delta t_0(m)$ such that for any $\Delta t \leq \Delta t_0$

$$\frac{\exp(\beta T)\Big(C(m,T)\Delta t^{1/2} + C(m,T)\Delta t\Big)}{|m|^2} \leq \frac{\epsilon}{2},$$

whence $\mathbb{P}(\vartheta_m < T) \leq \epsilon$ as required. □

6.4.4 Strong Convergence on the Whole Domain

In this section we present the strong convergence of the $(1,1)$-Milstein scheme (6.20) to the solution of (6.1). First, we will show that CFBM (6.28) converges to the true solution on a compact domain. This, together with Theorem 6.4.2, will enable us to extend convergence to the whole domain.

Let us define the stoping time θ_m as

$$\theta_m = \tau_m \wedge \vartheta_m.$$

Lemma 6.4.3. *Under Assumptions 6.2.5, 6.2.6 and 6.3.3 for any $p \geq 2$, $T > 0$*

and sufficiently large m, there exists a constant $C = C(p, T, m)$, such that

$$\mathbb{E}\left[\sup_{0\le t\le T}\left|\hat{X}(t\wedge\theta_m) - x(t\wedge\theta_m)\right|^p\right] \le C\Delta t^p.$$

Let us observe that due to the stopping time technique we have moved our problem from a local to global Lipschitz setting and therefore the lemma can be proved exactly in the same manner as in (Milstein and Tretyakov 2004; Kloeden and Platen 1992). We just need to bear in mind that in our case a constant C depends on m. Now we are ready to prove a strong convergence theorem.

Theorem 6.4.4. *Under Assumptions 6.2.5, 6.2.6 and 6.3.3, for any given $T > 0$ and $s \in [1, 2)$, we have*

$$\lim_{\Delta t\to 0}\mathbb{E}\left|X(T) - x(T)\right|^s = 0. \tag{6.33}$$

Proof. Let

$$e(T) = X(T) - x(T).$$

Applying Young's inequality in the form

$$x^s y \le \frac{\delta s}{2}x^2 + \frac{2-s}{2\delta^{\frac{s}{2-s}}}y^{\frac{2}{2-s}}, \quad \forall x, y, \delta > 0,$$

leads us to

$$\begin{aligned}\mathbb{E}\left|e(T)\right|^s &= \mathbb{E}\left[\left|e(T)\right|^s \mathbf{1}_{\{\tau_m>T,\vartheta_m>T\}}\right] + \mathbb{E}\left[\left|e(T)\right|^s \mathbf{1}_{\{\tau_m\le T \quad or \quad \vartheta_m\le T\}}\right]\\ &\le 2^{s-1}\left[\mathbb{E}[\left|\hat{X}(T) - x(T)\right|^s \mathbf{1}_{\{\tau_m>T,\vartheta_m>T\}}] + \mathbb{E}[\left|X(T) - \hat{X}(T)\right|^s \mathbf{1}_{\{\tau_m>T,\vartheta_m>T\}}]\right]\\ &+ \frac{\delta s}{2}\mathbb{E}\left[\left|e(T)\right|^2\right] + \frac{2-s}{2\delta^{\frac{s}{2-s}}}\mathbb{P}(\tau_m\le T \quad \text{or} \quad \vartheta_m\le T).\end{aligned}$$

To finish the proof we need to estimate the expressions on the right hand side of this inequality. First, let us observe that by Lemma 6.4.1 we obtain

$$\mathbb{E}[\left|X(T) - \hat{X}(T)\right|^s \mathbf{1}_{\{\tau_m>T,\vartheta_m>T\}}] \le C(m, s, T)\Delta t^s.$$

Given any $\epsilon > 0$, by Hölder's inequality and Theorem 2.2 in (Szpruch and Mao

2010) and Theorem 6.3.5, we choose δ such that

$$\frac{\delta s}{2}\mathbb{E}\left[|e(T)|^2\right] \leq 4^{p-1}\frac{\delta s}{2}\mathbb{E}\left[|x(T)|^2 + |X(T)|^2\right] \leq \frac{\epsilon}{3}.$$

Now, again by Theorem 2.2 in (Szpruch and Mao 2010) there exists N_0 such that for $m \geq N_0$

$$\frac{2-s}{2\delta^{\frac{s}{2-s}}}\mathbb{P}(\tau_m \leq T) \leq \frac{\epsilon}{3},$$

and finally by Lemma 6.4.3 and Theorem 6.4.2 we may choose Δt sufficiently small such that

$$2^{s-1}\left[\mathbb{E}[\left|\hat{X}(T) - x(T)\right|^s \mathbf{1}_{\{\tau_m > T, \vartheta_m > T\}}] + \mathbb{E}[\left|X(T) - \hat{X}(T)\right|^s \mathbf{1}_{\{\tau_m > T, \vartheta_m > T\}}]\right]$$
$$+ \frac{2-s}{2\delta^{\frac{s}{2-s}}}\mathbb{P}(\vartheta_m \leq T) \leq \frac{\epsilon}{3},$$

which completes the proof. □

6.4.5 Numerical Experiment

In order to estimate the rate of convergence we proceed with numerical experiments for

$$dx(t) = x(t)(\mu - \alpha x(t))dt + \beta x^{3/2}(t)dw(t).$$

We focus on the strong error at the endpoint T,

$$e_{\Delta t}^{strong} = \mathbb{E}\left|x(T) - X_T\right|,$$

with T=1. We plot $e_{\Delta t}^{strong}$ against Δt on a log-log scale. Error bars representing 95% confidence intervals are shown by circles, and a reference line of slop 1 is also given. Although we do not know the explicit form of the solution, Theorem 6.4.4 guarantees that the (1,1)-Milstein scheme (6.19) strongly converges to the true solution. Therefore, it is reasonable to take the (1,1)-Milstein scheme with a very small time step. We choose $\Delta t = 2^{-15}$ as a reference solution. We compare this to the (1,1)-Milstein scheme evaluated with $(2^4\Delta t, 2^6\Delta t, 2^8\Delta t, 2^{10}\Delta t)$ in order to estimate the rate of convergence. Since we are using a Monte Carlo method, the sampling error decays like $1/\sqrt{M}$, M- is the number of sample paths. We set $M = 1000$. From Figure (6.1) we see that there appears to exist a positive

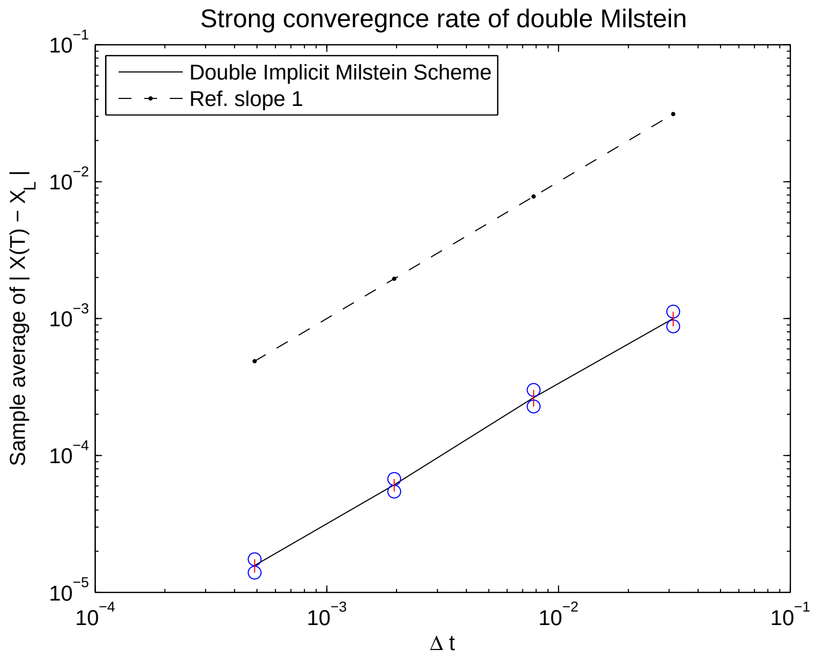

Figure 6.1: Strong error of double-implicit Milstein scheme applied to Heston 3/2 Stochastic volatility model.

constant C such that

$$e_{\Delta t}^{strong} \leq C\Delta t \quad \text{for sufficiently small } \Delta t.$$

If this inequality holds with approximate equality then taking logs gives

$$\log e_{\Delta t}^{strong} \approx \log C + \frac{1}{2} \log \Delta t. \tag{6.34}$$

A least squares fit for $\log C$ and q produced the value 1.0052 for q with a least square residual of 0.0565. Hence, our results are consistent with strong order of convergence equal to one.

Remark 6.4.5. *To extend our strong convergence Theorem 6.4.4 to the more* *general (θ, σ)-Milstein Scheme (6.37) we need to replace Assumption 6.3.3 by*

$$\begin{aligned} &2xf(x) + |g(x)|^2 + (1-2\theta)\,|f(x)|^2\,\Delta t \\ &+ \frac{\Delta t}{2} L^1 g(x)(2\sigma f(x) + L^1 g(x)) \le \alpha + \beta\,|x|^2 \quad x \in \mathbb{R}. \end{aligned}$$

Clearly for $\theta \ge 0.5$ and $\sigma = 1$ the above condition does not add any additional restrictions on coefficients of the SDE (6.1).

6.5 Stability Analysis

In this section we examine the global stability of the (σ, θ)-Milstein scheme (6.2). We prove that the (σ, θ)-Milstein scheme (6.2) reproduces the asymptotic behaviour of (6.1) very well.

6.5.1 Mean-Square Stability for Milstein-Type Scheme

Since we use Lyapunov function $V(x) = |x|^2$, our results extend mean-square stability for linear systems, (Higham 2001; Higham 2000) to a highly nonlinear setting. It is well known (Kloeden and Platen 1992) that adding a second order term in a Stochastic Taylor expansion increases the rate of strong convergence from 0.5 to 1 but affects stability. Higham (Higham 2000), considered the linear test SDE

$$dx(t) = \alpha x(t)dt + \mu x(t)dw(t), \tag{6.35}$$

with an initial condition $x(0) = x_0$. He showed that for the θ-Milstein scheme

$$\begin{aligned} X_{t_{k+1}} = X_{t_k} + \theta\alpha X_{t_{k+1}}\Delta t + (1-\theta)\alpha X_{t_k}\Delta t + \mu X_{t_k}\Delta w_{t_{k+1}} \\ + \frac{1}{2}\mu^2 X_{t_k}[\Delta w_{t_{k+1}}^2 - \Delta t], \end{aligned}$$

the linear stability region, i.e.,

$$R_{MS} := \{\Delta t\alpha, \Delta t\mu^2 \in \mathbb{R} : \text{method mean-square stable on 6.35}\} \tag{6.36}$$

is significantly smaller than for the θ-EM scheme. Thus it is natural to ask if there exists a second order scheme which can preserve stability of its test equation for reasonably big step size? The aim of this section is to give a positive answer to this question. As a motivational example, first we consider the scalar linear SDE (6.35). In the case of the θ-EM method we are not able to introduce additional implicitness, since $\mathbb{E}\left|(1-\mu\Delta w_{t_{k+1}})^{-1}\right| = \infty$, (Milstein and Tretyakov 2004). However, having a second order term from the stochastic Taylor expansion

$$\frac{1}{2}\mu^2 X_{t_k}[\Delta w_{t_{k+1}}^2 - \Delta t]$$

and introducing partial implicitness leads to the (θ,σ)-Milstein scheme, i.e.,

$$\begin{aligned} X_{t_{k+1}} = X_{t_k} &+ \theta\alpha X_{t_{k+1}}\Delta t + (1-\theta)\alpha X_{t_k}\Delta t + \mu X_{t_k}\Delta w_{t_{k+1}} \\ &+ \frac{1}{2}\mu^2 X_{t_k}\Delta w_{t_{k+1}}^2 - \frac{(1-\sigma)}{2}\mu^2 X_{t_k}\Delta t - \frac{\sigma}{2}\mu^2 X_{t_{k+1}}\Delta t. \end{aligned} \tag{6.37}$$

Theorem 6.5.1. *The* (θ,σ)*-Milstein Scheme* (6.37) *is globally mean square stable, i.e.*

$$\lim_{k\to\infty}\mathbb{E}\left|X_{t_k}\right|^2 = 0, \tag{6.38}$$

if and only if,

$$(2\alpha+\mu^2) + \Delta t\alpha^2(1-2\theta) + \frac{\Delta t\mu^2}{2}(2\sigma\alpha+\mu^2) < 0. \tag{6.39}$$

Proof. We rewrite (θ,σ)-Milstein Scheme (6.37) as a recurrence of the form

$$X_{t_{k+1}} = X_{t_k}\left(p + q\xi_{t_{k+1}} + r\xi_{t_{k+1}}^2\right),$$

where $\xi \sim \mathbb{N}(0,1)$,

$$p = \frac{1+(1-\theta)\alpha\Delta t - \frac{(1-\sigma)}{2}\mu^2\Delta t}{1-\theta\alpha\Delta t + \frac{\sigma}{2}\mu^2\Delta t},$$

$$q = \frac{\mu\sqrt{\Delta t}}{1-\theta\alpha\Delta t + \frac{\sigma}{2}\mu^2\Delta t},$$

$$r = \frac{\frac{1}{2}\mu^2\Delta t}{1-\theta\alpha\Delta t + \frac{\sigma}{2}\mu^2\Delta t}.$$

Then

$$\left|X_{t_{k+1}}\right|^2 = \left|X_{t_k}\right|^2\left(p^2 + q^2\xi_{t_{k+1}}^2 + r^2\xi_{t_{k+1}}^4 + 2pq\xi_{t_{k+1}} + 2pr\xi_{t_{k+1}}^2 + 2qr\xi_{t_{k+1}}^3\right).$$

Taking conditional expectation of both sides leads us to

$$\mathbb{E}[\left|X_{t_{k+1}}\right|^2 |\mathcal{F}_{t_k}] = |X_{t_k}| \left(p^2 + q^2 + 3r^2 + 2pr\right).$$

Taking conditional expectation of both sides again we obtain

$$\mathbb{E}\left|X_{t_{k+1}}\right|^2 = \mathbb{E}\left|X_{t_k}\right|^2 \left(p^2 + q^2 + 3r^2 + 2pr\right). \tag{6.40}$$

Therefore the condition which characterizes the stability has the form

$$(p+r)^2 + q^2 + 2r^2 < 1,$$

This is equivalent to (6.39), as required. □

Remark 6.5.2. *Let us observe that for $\theta = 0.5$ and $\sigma = 1$ we have recovered exactly the same condition as in the SDE setting*

$$(2\alpha + \mu^2) < 0, \tag{6.41}$$

so the method perfectly reproduces stability for any step-size.

Motivated by (Higham 2001) we will draw stability regions for (6.37) in $x - y$ plane, where $x = \alpha\Delta t$ and $y = \mu^2\Delta t$. In Figure 6.2 the stability region of the underlying SDE (6.35) is shown in light grey color. The upper pictures in Figure 6.2 superimpose the stability region of the $(\theta, 0)$-Milstein scheme with $\theta = 0, 0.5, 1$ respectively. We see that even in the case of a linear scalar equation we are not able to reproduce the stability region of the underling test equation (6.35). However, by introducing additional implicitness we can overcome this poor performance. The lower pictures in Figure 6.2 superimpose the stability region of the (θ, σ)-Milstein scheme with $(0, 1), (0.5, 1), (1, 1)$, respectively. As stated in Remark 6.5.2, we recover exactly the stability region of the underlying test SDE (6.35) for $(0.5, 1)$.

6.5.2 Almost Sure Stability for Milstein

Here we demonstrate that linear mean-square stability analysis can be extended to non-linear case. Similarly to Theorem 3.5.3 we prove the stochastic LaSalle

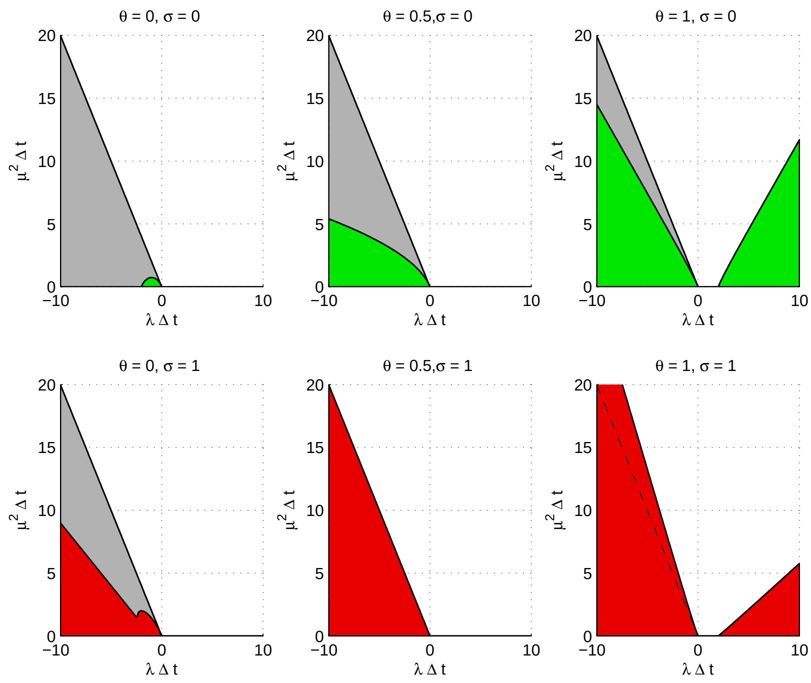

Figure 6.2: Mean square stability of Implicit Milstein scheme vs. double-implicit Milstein scheme for linear a SDE.

theorem for (θ, σ)-Milstein scheme 6.2.

Theorem 6.5.3. *Let Assumptions 6.2.5 and 6.3.1 hold. Assume that for the (θ, σ)-Milstein Scheme (6.37) there exists a function $z \in C(\mathbb{R}^n; \mathbb{R}_+)$ such that*

$$\begin{aligned} &2xf(x) + |g(x)|^2 + (1 - 2\theta)\,|f(x)|^2\,\Delta t \\ &+ \frac{\Delta t}{2} L^1 g(x)(2\sigma f(x) + L^1 g(x)) \le -z(x) \qquad \textit{for all } (x,t) \in \mathbb{R}. \end{aligned}$$

Then

$$\limsup_{k\to\infty} |X(t_k)|^2 < \infty \quad a.s., \tag{6.42}$$

and

$$\lim_{k\to\infty} z(X(t_k)) = 0 \ \ a.s. \tag{6.43}$$

If additionally $z(x) = 0$ *iff* $x = 0$

$$\lim_{k\to\infty} X(t_k) = 0 \ a.s.$$

Proof.

$$F(x) = x - \theta f(x)\Delta t + \frac{1}{2}\sigma L^1 g(x)\Delta t \tag{6.44}$$

Using the definition of the function F we can rewrite the scheme

$$\begin{aligned} X_{t_{k+1}} &= X_{t_k} + \theta f(X_{t_{k+1}})\Delta t + (1-\theta) f(X_{t_k})\Delta t + g(X_{t_k})\Delta w_{t_{k+1}} \\ &+ \frac{1}{2}L^1 g(X_{t_k}))\Delta w_{t_{k+1}}^2 - \frac{(1-\sigma)}{2}L^1 g(X_{t_k})\Delta t - \frac{\sigma}{2}L^1 g(X_{t_{k+1}})\Delta t \end{aligned}$$

as

$$\begin{aligned} F(X_{t_{k+1}}) &= F(X_{t_k}) + f(X_k)\Delta t + g(X_k)\Delta w_{t_{k+1}} \\ &+ \frac{1}{2}L^1 g(X_{t_k})\Delta w_{t_{k+1}}^2 - \frac{1}{2}L^1 g(X_{t_k})\Delta t. \end{aligned}$$

Squaring both sides we have

$$\begin{aligned} \left|F(X_{t_{k+1}})\right|^2 &= |F(X_{t_k})|^2 + |f(X_{t_k})|^2 \Delta t^2 + |g(X_{t_k})|^2 \Delta t + 2F(X_{t_k}) f(X_{t_k})\Delta t \\ &+ \frac{3}{4}\left|L^1 g(X_{t_k})\right|^2 \Delta t^2 + \frac{1}{4}L^1 \left|g(X_{t_k})\right|^2 \Delta t^2 \\ &+ F(X_{t_k})L^1 g(X_{t_k})\Delta t - F(X_{t_k})L^1 g(X_{t_k})\Delta t \\ &+ f(X_{t_k})L^1 g(X_{t_k})\Delta t^2 - f(X_{t_k})L^1 g(X_{t_k})\Delta t^2 \\ &- \frac{1}{2}\left|L^1 g(X_{t_k})\right|^2 \Delta t^2 + r_{k+1}, \end{aligned}$$

where

$$
\begin{aligned}
r_{k+1} &= |g(X_{t_k})|^2 [\Delta W_{t_{k+1}}^2 - \Delta t] + 2F(X_{t_k})g(X_k)\Delta w_{t_{k+1}} \\
&+ \frac{1}{4}\left|L^1 g(X_{t_k})\right|^2 [\Delta W_{t_{k+1}}^4 - 3\Delta t^2] \\
&+ F(X_{t_k})L^1 g(X_{t_k})[\Delta W_{t_{k+1}}^2 - \Delta t] \\
&+ f(X_k)\Delta t g(X_k, t_k)\Delta w_{t_{k+1}} + f(X_{t_k})L^1 g(X_{t_k})[\Delta W_{t_{k+1}}^2 - \Delta t] \\
&+ g(X_k)\Delta w_{t_{k+1}} L^1 g(X_{t_k})\Delta w_{t_{k+1}}^2 - L^1 g(X_{t_k}) \\
&+ \frac{1}{2}\left|L^1 g(X_{t_k})\right|^2 [\Delta W_{t_{k+1}}^2 - \Delta t]\Delta t.
\end{aligned}
$$

This implies that

$$
\begin{aligned}
\left|F(X_{t_{k+1}})\right|^2 &= |F(X_{t_k})|^2 + |f(X_{t_k})|^2 \Delta t^2 \\
&+ |g(X_{t_k})|^2 \Delta t + 2F(X_{t_k})f(X_{t_k})\Delta t + \frac{1}{2}\left|L^1 g(X_{t_k})\right|^2 \Delta t^2 + r_{k+1} \\
&= |F(X_{t_k})|^2 + 2X_{t_k} f(X_{t_k})\Delta t + |g(X_{t_k})|^2 \Delta t + (1-2\theta)\, |f(X_{t_k})|^2 \Delta t^2 \\
&+ \frac{\sigma}{2} L^1 g(X_{t_k}) f(X_{t_k})\Delta t^2 + \frac{1}{2}\left|L^1 g(X_{t_k})\right|^2 \Delta t^2 + r_{k+1} \\
&= |F(X_{t_k})|^2 + 2X_{t_k} f(X_{t_k})\Delta t + |g(X_{t_k})|^2 \Delta t + (1-2\theta)\, |f(X_{t_k})|^2 \Delta t^2 \\
&+ \frac{\Delta t^2}{2} L^1 g(X_{t_k})(2\sigma f(X_{t_k}) + g(X_{t_k})) + r_{k+1}.
\end{aligned}
$$

Hence, we have obtained a decomposition that allow us to apply Theorem 3.5.2, i.e.,

$$
\left|F(X_{t_{k+1}})\right|^2 = |F(X_{t_k})|^2 - A_{t_k}\Delta t + r_{k+1},
$$

where

$$
\begin{aligned}
A_{t_k} = -\Bigg(& 2\langle x, f(x)\rangle + |g(x)|^2 + (1-2\theta)\, |f(x)|^2 && (6.45) \\
& + \frac{\Delta t}{2} L^1 g(x)(\sigma f(x) + L^1 g(x))\Bigg). && (6.46)
\end{aligned}
$$

Therefore

$$\left|F(X_{t_{N+1}}, t_{N+1})\right|^2 = |F(X_{t_0}, t_0)|^2 - \sum_{k=0}^{N} A_{t_k}\Delta t + \sum_{k=0}^{N} r_{k+1}. \tag{6.47}$$

Now we are in position to apply Theorem 3.5.2 to get

$$\lim_{k\to\infty} |F(X_{t_k})|^2 < \infty, \tag{6.48}$$

from which it follows that $\limsup_{k\to\infty} |X(t_k)|^2$ exist and is finite almost surely. Another implication of Theorem 3.5.2 is

$$\sum_{k=0}^{\infty} z(X_{t_k})\Delta t \leq \sum_{k=0}^{\infty} A_{t_k}\Delta t < \infty \quad \text{a.s,}$$

which implies

$$\lim_{k\to\infty} z(X_{t_k}) = 0 \quad \text{a.s.} \tag{6.49}$$

and the final part of the theorem follows immediately. □

Chapter 7

Future Research

In this thesis we answered some important questions on stochastic numerical analysis. However, there is still an ocean of problems that need to be considered in future research. Here we point out some of them.

- In this thesis we investigated numerical approximations for SDEs with non-Lipschitz coefficients. We have demonstrated that implicit methods have very desirable properties. Nevertheless, having in mind Theorem 1.3.1, numerical methods are still far behind. Lyapunov function techniques enable us to prove existence and uniqueness of the solution for a very rich family of SDEs. We demonstrated that employing implicit schemes we are able to accurately approximate these solutions for which existence follows from Theorem 1.3.1 once we choose Lyapunov function $V(x) = |x|^2$. Therefore additional research is needed. Similarly we investigated the very special case of stability analysis conducted in (Shen, Luo, and Mao 2006). Additional insight into general almost sure stability properties for numerical schemes also would be of relevance. It would be also interesting to investigate if BEM is sufficient to reproduce almost sure stability of the underlying SDE.

- Our analysis of higher order schemes is very promising. In the case of scalar equations we have developed a second order scheme with superior properties: strong convergence, stability and preservation of positivity (Chapter: 6). Further research on this topic needs to be done in order to cover stochastic volatility models and correlated multidimensional SDEs. It is well known that in the general multidimensional case, there is no exact solution for iterated integrals of second order (Lévy Area) (Kloeden and Platen 1992; Glasserman 2003). Simulations of these are computationally expensive. Recently it has been demonstrated that with the use of some sophisticated orthogonal transformations we may avoid simulation of Lévy Areas. This technique allows maintenance of the higher order of strong convergence, equal to that of the scalar case, (Cruzeiro, Malliavin, and Thalmaier 2004; Cruzeiro and Malliavin 2006; Alves and Cruzeiro 2008; Malliavin and Thalmaier 2003). This is a very promising stream of research of great practical importance. We are planning to extend current strong convergence theorems for non-linear stochastic differential financial models using second order schemes by utilising the orthogonal transformations.

- Further improvement of stochastic numerical integration can be achieved by implementing a step-adaptive technique. The superiority of the adaptive method compared to equidistant discretization has been demonstrated in (Hofmann, Muller-Gronbach, and Ritter 2000a; Hofmann, Muller-Gronbach, and Ritter 2000b; Hofmann, Muller-Gronbach, and Ritter 2001; Lamba 2003; Lehn, Rosler, and Schein 2002; Lamba, Mattingly, and Stuart 2007; Müller-Gronbach and Ritter 2008; Muller-Gronbach and Ritter 2007). The above papers proved that adaptive methods can achieve the required accuracy with significantly lower cost than standard techniques. However, the developed theory so far has not been justified by simulations for many non-linear financial models. The step-adapted techniques have also been proved to overcome some difficulties when simulating stopped processes (Dzougoutov, Moon, von Schwerin, Szepessy, and Tempone). Techniques we have developed so far to deal with non-linear systems can be applied in these models. We believe that this approach would lead to fruitful results. It could be very beneficial to combine the step-adapted methods with the Multi-Level Monte Carlo technique in order to reduce computational complexity even further.

- In this thesis we looked at SDEs driven by Brownian motion. Our results can be extended to incorporate jump processes in the similar fashion as in (Higham and Kloeden 2005; Higham and Kloeden 2004).

  The author is also currently working on extensions of the results in this thesis for stochastic delay differential equations. Having this in mind it is worth mentioning work on delay stochastic models (Kazmerchuk, Swishchuk, and Wu 2005; Swishchuk 2005; Kazmerchuk, Swishchuk, and Wu 2007). These authors have considered the continuous time analog of GARCH(1,1) with incorporated delays. They idea is related to the famous Hobson and Rogers model, (Hobson and Rogers 1998), where volatility is expressed in terms of exponentially weighted moments of historic log-price. The same type of consideration has led to the derivation of a Delayed Black and Scholes formula (Arriojas, Hu, Mohammed, and Pap 2007). This motivates us to research numerical methods for stochastic systems that are not Markovian.

References

Ahn, D. and B. Gao (1999). A parametric nonlinear model of term structure dynamics. *Review of Financial Studies 12*(4), 721.

Ait-Sahalia, Y. (1996). Testing continuous-time models of the spot interest rate. *Review of Financial Studies 9*(2), 385–426.

Ait-Sahalia, Y. (1999). Transition densities for interest rate and other nonlinear diffusions. *Journal of Finance 54*(4), 1361–1395.

Alves, C. and A. Cruzeiro (2008). Monte-Carlo simulation of stochastic differential systemsa geometrical approach. *Stochastic Processes and their Applications 118*(3), 346–367.

Andersen, T., L. Benzoni, and J. Lund (2002). An empirical investigation of continuous-time equity return models. *The Journal of Finance 57*(3), 1239–1284.

Appleby, J., M. Guzowska, and A. Rodkina (2010). Preservation of positivity in the solution of discretised stochastic differential equation. *working paper*.

Appleby, J., C. Kelly, X. Mao, and A. Rodkina (2010). On the local dynamics of polynomial differene equations with fading stochastic perturbations. *Dynamics of Continuous, Discrete and Impulsive Systems 17*, 401–430.

Appleby, J., X. Mao, and A. Rodkina (2008). Stabilization and destabilization of nonlinear differential equations by noise. *IEEE Transactions on Automatic Control 53*(3), 683–691.

Arriojas, M., Y. Hu, S. Mohammed, and G. Pap (2007). A delayed Black and Scholes formula. *Stochastic Analysis and Applications 25*(2), 471–492.

Bahar, A. and X. Mao (2004). Stochastic delay population dynamics. *International Journal of Pure and Applied Mathematics . 11*, 377–400.

Baker, C. and E. Buckwar (2005). Exponential stability in p-th mean of solutions, and of convergent Euler-type solutions, of stochastic delay differential equations. *Journal of Computational and Applied Mathematics 184*(2), 404–427.

Berkaoui, A., M. Bossy, and A. Diop (2007). Euler scheme for SDEs with non-Lipschitz diffusion coefficient: strong convergence. *ESAIM: Probability and Statistics 12*, 1–11.

Billingsley, P. (1979). *Probability and Measure*. John Wiley & Sons.

Black, F. and M. Scholes (1973). The pricing of options and corporate liabilities. *The Journal of Political Economy 81*(3), 637–654.

Broadie, M., P. Glasserman, and S. Kou (1997). A continuity correction for discrete barrier options. *Mathematical Finance 7*(4), 325–349.

Broadie, M. and O. Kaya (2006). Exact simulation of stochastic volatility and other affine jump diffusion processes. *Operations Research 54*(2), 217–231.

Buchmann, F. (2005). Simulation of stopped diffusions. *Journal of Computational Physics 202*(2), 446–462.

Campbell, J., A. Lo, A. MacKinlay, and R. Whitelaw (1998). The econometrics of financial markets. *Macroeconomic Dynamics 2*(04), 559–562.

Chan, K., G. Karolyi, F. Longstaff, and A. Sanders (1992). An empirical comparison of alternative models of the short-term interest rate. *The journal of finance 47*(3), 1209–1227.

Conley, T., L. Hansen, E. Luttmer, and J. Scheinkman (1997). Short-term interest rates as subordinated diffusions. *The Review of Financial Studies 10*(3), 525–577.

Cox, J., J. Ingersoll Jr, and S. Ross (1985). A theory of the term structure of interest rates. *Econometrica 53*(2), 385–407.

Cruzeiro, A. and P. Malliavin (2006). Numerical approximation of diffusions in Rd using normal charts of a Riemannian manifold. *Stochastic Processes and their Applications 116*(7), 1088–1095.

Cruzeiro, A., P. Malliavin, and A. Thalmaier (2004). Geometrization of Monte-Carlo numerical analysis of an elliptic operator: strong approximation.

Comptes rendus-Mathématique 338(6), 481–486.

Duan, J. (1996). A unified theory of option pricing under stochastic volatility-from garch to diffusion. *Hong Kong University of Science and Technology, Working Paper*.

Duan, J. (2003). A specification test for time series models by a normality transformation. *workingpaper, University of Toronto*.

Dzougoutov, A., K. Moon, E. von Schwerin, A. Szepessy, and R. Tempone. Adaptive Monte Carlo algorithms for stopped diffusion. *Multiscale methods in science and engineering, Lecture notes in computational science and engineering 44*, 59–88.

Einstein, A. (1905). On the movement of small particles suspended in stationary liquid demanded by the molecular-kinetic theory of heat, Ann. d.

Fama, E. (1965). The behavior of stock-market prices. *Journal of business 38*(1).

Feller, W. (1968). An introduction to probability theory and its applications. Vol. 1 and 2.

Fischer, T., A. May, and B. Walther (2004). Simulation of the yield curve: Checking a Cox-Ingersoll-Ross model. *Working Paper, TU Darmstadt*.

Fornari, F. and A. Mele (2001). Recovering the probability density function of asset prices using GARCH as diffusion approximations. *Journal of Empirical Finance 8*(1), 83–110.

Friedman, A. (1976). Stochastic differential equations and applications.

Gallant, A. and G. Tauchen (2005). Estimation of continuous-time models for stock returns and interest rates. *Macroeconomic Dynamics 1*(01), 135–168.

Gard, T. (1988). *Introduction to Stochastic Differential Equations*. Marcel Dekker, New York.

Gardiner, C. (1985). *Handbook of Stochastic Methods*. Springer Berlin.

Gihman, I. and A. Skorohod (1972). *Stochastic Differential Equations*. Springer-Verlag.

Giles, M. (2006). Improved multilevel Monte Carlo convergence using the Milstein scheme. *Monte Carlo and Quasi-Monte Carlo Methods 2006*, 343–358.

Giles, M. (2008). Multilevel monte carlo path simulation. *Operations Research-Baltimore 56*(3), 607–617.

Giles, M., D. Higham, and X. Mao (2009). Analysing multi-level Monte Carlo for options with non-globally Lipschitz payoff. *Finance and Stochastics 13*(3), 403–413.

Gillespie, D. (1992). *Markov Processes: An Introduction for Physical scientists.* Academic Press.

Glasserman, P. (2003). *Monte Carlo Methods in Financial Engineering.* Springer Verlag.

Hairer, E. and G. Wanner (2010). *Solving Ordinary Differential Equations II: Stiff and Differential-Algebraic Problems.* Springer.

Hale, J. and S. Lunel (1993). *Introduction to Functional Differential Equations.* Springer Verlag.

Heston, S. (1993). A closed-form solution for options with stochastic volatility with applications to bond and currency options. *Review of Financial Studies 6*(2), 327–343.

Heston, S. (1997). A simple new formula for options with stochastic volatility. *Course notes of Washington University in St. Louis, Missouri*.

Heston, S. and S. Nandi (2000). A closed-form GARCH option valuation model. *Review of Financial Studies 13*(3), 585.

Higham, D. (2000). A-stability and stochastic mean-square stability. *BIT Numerical Mathematics 40*(2), 404–409.

Higham, D. (2001). Mean-square and asymptotic stability of the stochastic theta method. *SIAM Journal on Numerical Analysis*, 753–769.

Higham, D. and P. Kloeden (2004). Convergence and stability of implicit methods for jump-diffusion systems. *International Journal of Numerical Analysis*, 1–18.

Higham, D. and P. Kloeden (2005). Numerical methods for nonlinear stochastic differential equations with jumps. *Numerische Mathematik 101*(1), 101–119.

Higham, D. and X. Mao (2005). Convergence of Monte Carlo simulations involving the mean-reverting square root process. *Journal of Computational Finance 8*(3), 35.

Higham, D., X. Mao, and A. Stuart (2003a). Exponential mean-square stability of numerical solutions to stochastic differential equations. *LMS J. Comput. Math 6*, 297–313.

Higham, D., X. Mao, and A. Stuart (2003b). Strong convergence of Euler-type methods for nonlinear stochastic differential equations. *SIAM Journal on Numerical Analysis 40*(3), 1041–1063.

Higham, D., X. Mao, and C. Yuan (2008). Almost sure and moment exponential stability in the numerical simulation of stochastic differential equations. *SIAM Journal on Numerical Analysis 45*(2), 592–609.

Hobson, D. and L. Rogers (1998). Complete models with stochastic volatility. *Mathematical Finance 8*(1), 27–48.

Hofmann, N., T. Muller-Gronbach, and K. Ritter (2000a). Optimal approximation of stochastic differential equations by adaptive step-size control. *Mathematics of Computation 69*(231), 1017–1034.

Hofmann, N., T. Muller-Gronbach, and K. Ritter (2000b). Step size control for the uniform approximation of systems of stochastic differential equations with additive noise. *Annals of Applied Probability*, 616–633.

Hofmann, N., T. Muller-Gronbach, and K. Ritter (2001). The optimal discretization of stochastic differential equations. *Journal of Complexity 17*(1), 117–153.

Hong, Y. and H. Li (2005). Nonparametric specification testing for continuous-time models with applications to term structure of interest rates. *Review of Financial Studies 18*(1), 37.

Hu, Y. (1996). Semi-implicit Euler-Maruyama scheme for stiff stochastic equations. *Progress in Probability*, 183–202.

Hull, J. and A. White (1987). The pricing of options on assets with stochastic volatilities. *Journal of Finance 42*(2), 281–300.

Hutzenthaler, M. and A. Jentzen (2009). Non-globally Lipschitz counterexamples for the stochastic Euler scheme. *Arxiv preprint arXiv:0905.0273*.

Itô, K. (1944). Stochastic integral. *Proceedings of the Imperial Academy (Tokyo) 20*, 519–524.

Itô, K. (1946). On a stochastic integral equation. *Proceedings of the Japan Academy 22*(1-4), 32–35.

Jentzen, A., P. Kloeden, and A. Neuenkirch (2009). Pathwise approximation of stochastic differential equations on domains: higher order convergence rates without global Lipschitz coefficients. *Numerische Mathematik 112*(1), 41–64.

Johnson, H. and D. Shanno (1987). Option pricing when the variance is changing. *Journal of Financial and Quantitative Analysis 22*(2), 143–151.

Kahl, C., M. Gunther, and T. Rosberg (2008). Structure preserving stochastic integration schemes in interest rate derivative modeling. *Applied Numerical Mathematics 58*(3), 284–295.

Kallenberg, O. (2002). *Foundations of Modern Probability*. Springer Verlag.

Karatzas, I. and S. Shreve (1991). *Brownian Motion and Stochastic Calculus*. Springer.

Karatzas, I. and S. Shreve (1998). *Methods of Mathematical Finance*. Springer Verlag.

Kazmerchuk, Y., A. Swishchuk, and J. Wu (2005). A continuous-time GARCH model for stochastic volatility with delay. *Canadian Applied Mathem. Quarterly 13*(2).

Kazmerchuk, Y., A. Swishchuk, and J. Wu (2007). The pricing of options for securities markets with delayed response. *Mathematics and Computers in Simulation 75*(3-4), 69–79.

Khasminski, R. (1980). *Stochastic Stability of Differential Equations*. Kluwer Academic Pub.

Kloeden, P. and A. Neuenkirch (2007). The pathwise convergence of approximation schemes for stochastic differential equations. *Journal of Computation and Mathematics 10*, 235–253.

Kloeden, P. and E. Platen (1992). *Numerical Solution of Stochastic Differential Equations*. Springer.

Laing, C. and G. Lord (2009). *Stochastic Methods in Neuroscience*. Oxford Univ Pr.

Lamba, H. (2003). An adaptive timestepping algorithm for stochastic differential equations. *Journal of Computational and Applied Mathematics 161*(2), 417–430.

Lamba, H., J. Mattingly, and A. Stuart (2007). An adaptive Euler-Maruyama scheme for SDEs: convergence and stability. *IMA Journal of Numerical Analysis 27*(3), 479.

LaSalle, J. (1968). Stability theory for ordinary differential equations. *J. Differential Equations 4*(1), 57–65.

LaSalle, J. and Z. Artstein (1976). *The Stability of Dynamical Systems*. Society for Industrial Mathematics.

Lehn, J., A. Rosler, and O. Schein (2002). Adaptive schemes for the numerical solution of SDEsa comparison. *Journal of Computational and Applied Mathematics 138*(2), 297–308.

Lewis, A. (2000). *Option Valuation Under Stochastic Volatility*. Finance Press.

Liptser, R. and A. Shiryayev (1989). *Theory of Martingales*. Kluwer Academic Publishers.

Longstaff, F. (1989). A nonlinear general equilibrium model of the term structure of interest rates. *Journal of Financial Economics 23*(2), 195–224.

Lord, R., R. Koekkoek, and D. Van Dijk (2009). A comparison of biased simulation schemes for stochastic volatility models. *Quantitative Finance 10*(2), 177–194.

Malliavin, P. and A. Thalmaier (2003). Numerical error for SDE: Asymptotic expansion and hyperdistributions. *Comptes rendus-Mathématique 336*(10), 851–856.

Mandelbrot, B. (1963). The variation of certain speculative prices. *Journal of Business 36*(4).

Mannella, R. (1999). Absorbing boundaries and optimal stopping in a stochastic differential equation. *Physics Letters A 254*(5), 257–262.

Mao, X. (1999). Stochastic versions of the LaSalle theorem. *Journal of Differential Equations 153*(1), 175–195.

Mao, X. (2007). *Stochastic Differential Equations and Applications*. Horwood Pub Ltd.

Mao, X., G. Marion, and E. Renshaw (2002). Environmental Brownian noise suppresses explosions in population dynamics. *Stochastic Process. Appl 97*(1), 95–110.

Mao, X., S. Sabanis, and E. Renshaw (2003). Asymptotic behaviour of the stochastic Lotka–Volterra model. *Journal of Mathematical Analysis and Applications 287*(1), 141–156.

Mao, X., C. Yuan, and G. Yin (2007). Approximations of Euler–Maruyama type for stochastic differential equations with Markovian switching, under non-Lipschitz conditions. *Journal of Computational and Applied Mathematics 205*(2), 936–948.

Maruyama, G. (1955). Continuous Markov processes and stochastic equations. *Rendiconti del Circolo Matematico di Palermo 4*(1), 48–90.

Merton, R. (1973). Theory of rational option pricing. *The Bell Journal of Economics and Management Science 4*(1), 141–183.

Milshtein, G. (1975). Approximate integration of stochastic differential equations. *Theory of Probability and its Applications 19*, 557.

Milstein, G. (1987). A theorem on the order of convergence of mean square approximations of solutions of systems of stochastic differential equations. *Theory of Probability and its Applications 32*, 809–811.

Milstein, G., E. Platen, and H. Schurz (1998). Balanced implicit methods for stiff stochastic systems. *SIAM Journal on Numerical Analysis 35*(3), 1010–1019.

Milstein, G. and M. Tretyakov (2004). *Stochastic Numerics for Mathematical Physics. Scientific Computation.* Springer-Verlag, Berlin.

Muller-Gronbach, T. and K. Ritter (2007). Lower bounds and nonuniform time discretization for approximation of stochastic heat equations. *Foundations of Computational Mathematics 7*(2), 135–181.

Müller-Gronbach, T. and K. Ritter (2008). Minimal errors for strong and weak approximation of stochastic differential equations. *Monte Carlo and Quasi-Monte Carlo Methods 2006*, 53.

Nelson, D. (1990). ARCH models as diffusion approximations. *Journal of Econometrics 45*(1-2), 7–38.

Nelson, D. and D. Foster (1994). Asymptotic filtering theory for univariate ARCH models. *Econometrica: Journal of the Econometric Society 62*(1), 1–41.

Oksendal, B. (1998). *Stochastic Differential Equations: An Introduction with Applications*. Springer-Verlag Berlin Heidelberg.

Pages, G. (2007). Multi-step Richardson-Romberg Extrapolation: Remarks on variance control and complexity. *Monte Carlo Methods and Applications 13*(1), 37–70.

Paley, R., N. Wiener, and A. Zygmund (1933). Notes on random functions. *Mathematische Zeitschrift 37*(1), 647–668.

Pang, S., F. Deng, and X. Mao (2008). Asymptotic properties of stochastic population dynamics. *Dynamics of Continuous, Discrete and Impulsive Systems Series A 15*, 603–620.

Pedersen, A. (1995). Consistency and asymptotic normality of an approximate maximum likelihood estimator for discretely observed diffusion processes. *Bernoulli 1*(3), 257–279.

Protter, P. (2004). *Stochastic Integration and Differential Equations*. Springer Verlag.

Samuelson, P. (1965). Rational theory of warrant prices. *Industrial Management Review 6*(2), 13–31.

Schurz, H. (2005). Convergence and stability of balanced implicit methods for systems of SDEs. *Int. J. Numer. Anal. Model 2*(2), 197–220.

Scott, L. (1987). Option pricing when the variance changes randomly: Theory, estimation, and an application. *Journal of Financial and Quantitative analysis 22*(4), 419–438.

Shen, Y., Q. Luo, and X. Mao (2006). The improved LaSalle-type theorems for stochastic functional differential equations. *Journal of Mathematical Analysis and Applications 318*(1), 134–154.

Stanton, R. (1997). A nonparametric model of term structure dynamics and the market price of interest rate risk. *The Journal of Finance 52*(5), 1973–2002.

Stein, E. and J. Stein (1991). Stock price distributions with stochastic volatility: an analytic approach. *Review of Financial Studies 4*(4), 727–752.

Swishchuk, A. (2005). Modeling and pricing of variance swaps for stochastic volatilities with delay. *Wilmott Magazine 19*, 63–73.

Szpruch, L. (2010). Positivity Preserving Milstein-type scheme. Convergence and Stability . *working paper*.

Szpruch, L. and X. Mao (2010). Strong convergence of numerical methods for nonlinear stochastic differential equations under monotone conditions. *working paper*.

Szpruch, L., X. Mao, D. Higham, and J. Pan (2010). Strongly Nonlinear Ait-Sahalia-Type Interest Rate Model and its Numerical Approximation. *accepted to BIT Numerical Mathematics*.

Tan, W. and T. Wai-Yuan (2000). *Stochastic Modeling of AIDS Epidemiology and HIV Pathogenesis*. World Scientific Pub Co Inc.

Vasicek, O. (1977). An equilibrium characterization of the term structure. *Journal of Financial Economics 5*(2), 177–188.

Von Smoluchowski, M. (1906). Zur kinetischen theorie der brownschen molekularbewegung und der suspensionen. *Annalen der physik 326*, 756–780.

Wagner, W. and E. Platen (1978). Approximation of Ito integral equations. *Preprint ZIMM, Akad. der. Wiss der DDR, Berlin*.

Wiener, N. (1923). Differential space. *J. Math. Phys 2*(13), 174.

Wiggins, J. (1987). Option values under stochastic volatility: Theory and empirical estimates. *Journal of Financial Economics 19*(2), 351–372.

Williams, D. (1991). *Probability with Martingales*. Cambridge Univ Press.

Zeidler, E. (1985). *Nonlinear Functional Analysis and its Applications*. Springer Verlag.

Zhu, J. (2009). *Modular Pricing of Options: An Application of Fourier Analysis*. Springer Verlag.