8-2016

Numerical Solutions of Stochastic Differential Equations

Liguo Wang
University of Tennessee, Knoxville, lwang43@vols.utk.edu

To the Graduate Council:

I am submitting herewith a dissertation written by Liguo Wang entitled "Numerical Solutions of Stochastic Differential Equations." I have examined the final electronic copy of this dissertation for form and content and recommend that it be accepted in partial fulfillment of the requirements for the degree of Doctor of Philosophy, with a major in Mathematics.

Jan Rosinski, Major Professor

We have read this dissertation and recommend its acceptance:

Xia Chen, Xiaobing Feng, Wenjun Zhou

Accepted for the Council:
Dixie L. Thompson

Vice Provost and Dean of the Graduate School

(Original signatures are on file with official student records.)

Numerical Solutions of Stochastic Differential Equations

A Dissertation Presented for the

Doctor of Philosophy

Degree

The University of Tennessee, Knoxville

Liguo Wang

August 2016

To my beloved parents Liangqin Wang and Quanfeng Xu, for their endless love, encouragement and support.

Acknowledgements

First of all, I am deeply grateful and indebted to my advisor Prof. Jan Rosinski for his support and encouragement during my study at the University of Tennessee, Knoxville. Without his patient guidance and constant inspiration, it would be impossible for me to finish this work.

I would like to thank Professors Xia Chen, Xiaobing Feng and Wenjun Zhou for taking their precious time to read my dissertation and serve on my committee. Special thanks go to Professors Vasileios Maroulas and Jie Xiong for their interesting probability courses. I also want to show my appreciation to the mathematics department at the University of Tennessee, Knoxville for providing such a cordial environment of studying mathematics. Special thanks go to Ms. Pam Amentrout for her endless patience and help to me.

I am thankful to my fellow graduate students and friends Fei Xing, Jin Dong, Ernest Jum, Eddie Tu, Kei Kobayashi, Yukun Li, Kai Kang, Wenqiang Feng, Bo Gao, Xiaoyang Pan and Le Yin for their helpful discussions and encouragement. I would also like to thank my officemates through the years, Kokum Rekha De Silva, Joshua Mike and Mustafa Elmas, for their help and the many happy conversations. Special thanks also go to Dr. Shiying Si for his kindly help when I did my intern job.

Finally, I would like to thank my beloved parents, my two elder sisters and my aunt for their constant love and support.

Knowledge as action. — Yangming Wang

Abstract

In this dissertation, we consider the problem of simulation of stochastic differential equations driven by Brownian motions or the general Lévy processes. There are two types of convergence for a numerical solution of a stochastic differential equation, the strong convergence and the weak convergence. We first introduce the strong convergence of the tamed Euler-Maruyama scheme under non-globally Lipschitz conditions, which allow the polynomial growth for the drift and diffusion coefficients. Then we prove a new weak convergence theorem given that the drift and diffusion coefficients of the stochastic differential equation are only twice continuously differentiable with bounded derivatives up to order 2 and the test function are third order continuously differentiable with all of its derivatives up to order 3 satisfying a polynomial growth condition. We also introduce the multilevel Monte Carlo method, which is efficient in reducing the total computational complexity of computing the expectation of a functional of the solution of a stochastic differential equation. This method combines the three sides of the simulation of stochastic differential equations: the strong convergence, the weak convergence and the Monte Carlo method. At last, a recent progress of the strong convergence of the numerical solutions of stochastic differential equations driven by Lévy processes under non-globally Lipschitz conditions is also presented.

Table of Contents

List of Figures

Chapter 1

Introduction

Stochastic differential equations (SDEs) driven by Brownian motions or Lévy processes are important tools in a wide range of applications, including biology, chemistry, mechanics, economics, physics and finance [2, 31, 33, 45, 58]. Those equations are interpreted in the framework of Itô calculus [2, 45] and examples are like, the geometric Brownian motion,

$$dX(t) = \mu X(t)dt + \sigma X(t)dW(t), \quad X(0) = X_0, \tag{1.1}$$

which plays a very important role in the Black-Sholes-Merton option pricing model, or, the Feller's branching diffusion in biology,

$$dX(t) = \alpha X(t)dt + \sigma\sqrt{X(t)}dW(t), \quad X(0) = X_0 > 0,$$

where $W(t)$ is the Brownian motion in both examples. Another example of SDE driven by a Lévy process is the following jump-diffusion process [40]:

$$dS(t) = a(t, S(t-))dt + b(t, S(t-))dW(t) + c(t, S(t-))dJ(t), \quad 0 \leq t \leq T,$$

where the jump term $J(t)$ is a compound Poisson process $\sum_{i=1}^{N(t)} Y_i$, the jump magnitude Y_i has a prescribed distribution, and $N(t)$ is a Poisson process with intensity λ, independent of the Brownian motion $W(t)$. This equation is used to model the stock price which may be discontinuous and is a generalization of equation (1.1).

Usually, the SDEs we encounter do not have analytical solutions and developing efficient numerical methods to simulate those SDEs is an important research topic. The goal of this thesis is to introduce the recent development of those numerical methods, including our own work on the weak convergence of the Euler-Maruyama scheme using Malliavin Calculus. Unlike the deterministic differential equations, there are two kinds of convergence measuring the approximation performance of a numerical scheme and they are used in different scenarios [33, 49].

Definition 1.1 (Strong convergence)**.** *Suppose Y is a discrete-time approximation of the solution $X(t)$ of a given SDE with maximum step size $\Delta > 0$. We say that Y converges to $X(t)$ in the strong sense with order $\gamma \in (0, \infty]$ if there exists a finite constant $C > 0$ and a positive constant Δ_0 such that*

$$E[\|X(T) - Y(T)\|] \leq C\Delta^{\gamma} \tag{1.2}$$

for any time discretization with maximum step size $\Delta \in (0, \Delta_0)$.

Definition 1.2 (Weak convergence)**.** *Suppose Y is a discrete-time approximation of the solution $X(t)$ of a given SDE with maximum step size $\Delta > 0$. We say that Y converges to $X(t)$ in the weak sense with order $\beta \in (0, \infty]$ if for any function g in a suitable function space there exists a finite constant $C > 0$ and a positive constant Δ_0 such that*

$$|E[g(X(T))] - E[g(Y(T))]| \leq C\Delta^{\beta} \tag{1.3}$$

for any time discretization with maximum step size $\Delta \in (0, \Delta_0)$.

Usually, the weak convergence order of a numerical scheme is higher than the strong convergence order of the same scheme, due to that the weak convergence is in the distributional sense.

Now consider a general SDE driven by a Brownian motion,

$$dX(t) = \mu(t, X(t))dt + \sigma(t, X(t))dW(t), \quad t \in (0, T], \quad X(0) = X_0, . \tag{1.4}$$

The most commonly used numerical scheme to solve the above SDE is the Euler-Maruyama (EM) scheme. It takes $Y_0 = X_0$ and

$$Y_{k+1} = Y_k + \mu(t_k, Y_k)\Delta + \sigma(t_k, Y_k)\Delta W_k,$$

where $t_k = k\frac{T}{N}$, $\Delta W_k = W(t_{k+1}) - W(t_k)$. It is well known that the EM scheme converges strongly with order $\frac{1}{2}$ if the coefficients $\mu(t,x)$ and $\sigma(t,x)$ satisfy the global Lipschitz condition and the linear growth condition (see Section 2.3 for more details). But these two conditions are so strict that many SDEs do not have such nice properties. In fact, a very large number of SDEs have C^1 functions as their coefficients and only satisfy the local Lipschitz condition. For example, the following stochastic Ginzburg-Landau equation:

$$dX(t) = (X(t) - X^3(t))dt + X(t)dW(t).$$

Therefore, the development of efficient numerical schemes for such SDEs has been and will continue to be an important research topic in the area of SDEs. To the author's knowledge, Hu [24] and Higham, Man and Stuart [23] were the pioneers of studying the strong convergence problem of the EM scheme under local Lipschitz conditions. In [23], although σ is still assumed to be globally Lipschitz continuous, μ only need to satisfy a one-sided Lipschitz condition and a polynomial growth condition, which is a substantial progress compared with the previous results. They proposed the following implicit (backward) Euler scheme

$$Y_{k+1} = Y_k + \mu(Y_{k+1})\Delta t + \sigma(Y_k)\Delta W_k$$

and proved that the scheme also achieves order $\frac{1}{2}$ strong convergence. But the shortcoming of this method is that it is an implicit scheme, which requires much more computational effort due to the need of solving a nonlinear equation in each time step. To overcome this problem, Hutzenthaler, Jentzen and Kloeden [28] proposed an explicit (tamed) Euler scheme,

$$Y_{k+1} = Y_k + \frac{\mu(Y_k)\Delta t}{1 + \|\mu(Y_k)\|\Delta t} + \sigma(Y_k)\Delta W_k,$$

assuming the same conditions as in [23] and still achieving the strong convergence order $\frac{1}{2}$. Then it was Sabanis [55] with another $\frac{1}{2}$ order strong convergent scheme,

$$Y_{k+1} = Y_k + \frac{\mu(k/N, Y_k)/N}{1 + N^{-1/2}\|Y_k\|^{3l/2+2}} + \frac{\sigma(k/N, Y_k)(W(\frac{k+1}{N}) - W(\frac{k}{N}))}{1 + N^{-1/2}\|Y_k\|^{3l/2+2}},$$

allowing also a polynomial growth condition on σ (see more details in Section 3.3). All of these have made the EM family a useful and prosperous computing toolbox for solving SDEs numerically.

The idea of taming can also be applied to the following SDEs driven by Lévy noise under local Lipschitz conditions:

$$\begin{cases} dX(t) = a(X(t-))dt + b(X(t-))dW(t) + \int_{\mathbb{R}^d} f(X(t-), y)\tilde{N}(dt, dy), \\ X(0) = x_0, \end{cases}$$

where $a(x)$ may have a polynomial growth. The tamed Euler scheme is quite similar to those we discussed above. We refer to Chapter 6 for more details.

Another problem of the EM scheme occurs in the context of the weak convergence. Unlike the strong convergence, the weak convergence depends largely on the regularity of the coefficients μ, σ and the test function g. Let $f(t,x) = E[g(X(T))|X(t) = x]$. As long as $\mu(t,x)$, $\sigma(t,x)$ and $g(x)$ satisfy some regularity conditions (see Section 2.2 for more details), we can rewrite the difference in (1.3) as

$$\begin{aligned}
&E[g(X(T))] - E[g(Y(T))] \\
&= -E\sum_{i=0}^{N-1}\Big[f\Big(\frac{(i+1)T}{N}, Y_{i+1}\Big) - f\Big(\frac{iT}{N}, Y_i\Big)\Big] \\
&= E\sum_{i=0}^{N-1}\int_{t_i}^{t_{i+1}}\Big[(\mu(s,Y(s)) - \mu(t_i,Y_i))\frac{\partial f}{\partial x}(s,Y(s))\Big]ds \\
&\quad + E\sum_{i=0}^{N-1}\int_{t_i}^{t_{i+1}}\Big[\frac{1}{2}(\sigma^2(s,Y(s)) - \sigma^2(t_i,Y_i))\frac{\partial^2 f}{\partial x^2}(s,Y(s))\Big]ds \qquad (1.5)
\end{aligned}$$

Most of the research which deals with analysis of the weak convergence error is based on the above decomposition. For example, in [60], each difference in the first equality of (1.5) is expanded further using the Taylor expansion. While in [33], the analysis is mainly based on the second equality of (1.5). It is well known that the EM scheme converges weakly with order 1 if, among other conditions, μ, σ and g are fourth order continuously differentiable with all of their derivatives up to order 4 satisfying a polynomial growth condition (i.e. $\mu(x), \sigma(x), g(x) \in C_p^4(\mathbb{R}^m)$) [33], or μ and σ are infinitely differentiable with all of their derivatives of any order bounded (i.e. $\mu(x), \sigma(x) \in C_b^\infty(\mathbb{R}^m)$) and g are only measurable and bounded (or with the polynoimal growth) [5]. In Chapter 4, we prove that the weak convergence also holds with order 1 in the 1 dimensional case if μ and σ^2 are only twice continuously differentiable with bounded derivatives up to order 2 (i.e. $\mu(x), \sigma^2(x) \in C_b^2(\mathbb{R})$) and g are third order continuously differentiable with all of its derivatives up to order 3 satisfying a polynomial growth condition (i.e. $g(x) \in C_p^3(\mathbb{R})$). In our proof, we apply the integration by parts technique from Malliavin calculus to the decomposition (1.5). By this method, we can decrease the smoothness conditions on μ, σ and g as compared with [5] or [33]. To the best of the author's knowledge, this result has not been provided before. It is also worthwhile mentioning that the analytical methods we use in this section are largely numerical scheme-independent and can also be generalized to other numerical schemes like the Milstein scheme or the schemes we introduce in Chapter 3.

Unlike the deterministic differential equations, the solution of a given SDE is a stochastic process. Usually, in practical applications we need to find the expectation $E[g(X(T))]$, where $X(T)$ is the terminal value of the solution and g is a function of $X(T)$. Typically, the distribution of $g(X(T))$ is unknown and $E[g(X(T))]$ can not be computed directly. The most commonly used method to address this issue is the Monte Carlo method. We first generate N independent discretized Brownian paths, and then use these Brownian paths and the numerical scheme to generate N independent sample paths of the solution. Denote by $Y^{(i)}(T)$ the approximate value of $X(T)$ at the ith sample path, then the expectation $E[g(X(T))]$ can be computed as

$$E[g(X(T))] \approx \frac{1}{N}\sum_{i=1}^{N} g(Y^{(i)}(T)). \tag{1.6}$$

The total computational complexity of finding $E[g(X(T))]$ depends both on the number of sample paths and the number of steps in the time discretisation. In fact, the mean square error (MSE) of the Monte Carlo estimation is asymptotically

$$MSE \approx O(N^{-1}) + O(\Delta^{2k}), \tag{1.7}$$

where Δ is the uniform step size of the time discretisation and k is the weak convergence order of our numerical scheme. Therefore, to reach the RMSE (RMSE=$\sqrt{MSE}$) $O(\varepsilon)$, the total computational cost of computing $E[f(X(T))]$ is $O(\varepsilon^{-(2+1/k)})$, which is very computationally expensive. It is well known that we can manage to reduce the total computational complexity considerably if proper variance reduction method is used [33]. To do so, Giles [16] proposed a multilevel Monte Carlo method, dealing with the problem from the perspective of variance reduction. The new method adopts different levels of time steps and uses the numerical solution from one level of the discretisation as a control variate of the numerical solution from the next level. Suppose we use L levels in total. In each level l, the time step h_l is equal to $h_l = M^{-l}T$, where $l = 0, 1, \cdots, L$ and $M \geq 2$ is an integer. Denote by $\widehat{P}_l$ the approximation to $f(X(T))$ using a numerical scheme with time step h_l. Then we can write

$$E[\widehat{P}_L] = E[\widehat{P}_0] + \sum_{l=1}^{L} E[\widehat{P}_l - \widehat{P}_{l-1}].$$

Therefore, to give $E[f(X)]$ an estimate, the simplest way is to estimate the expectations on the right hand side of the above equality using a standard Monte Carlo estimator. For $l = 0$, we use the following estimator

$$E[\widehat{P}_0] \approx \frac{1}{N_0} \sum_{i=1}^{N_0} \widehat{P}_0^{(i)}.$$

For $l \geq 1$,

$$E[\widehat{P}_l - \widehat{P}_{l-1}] \approx \sum_{i=1}^{N_l} (\widehat{P}_l^{(i)} - \widehat{P}_{l-1}^{(i)}),$$

where both $\widehat{P}_l^{(i)}$ and $\widehat{P}_{l-1}^{(i)}$ are obtained from the ith Brownian path. We can see that, in this procedure, we need to determine the number of levels L and the number of sample paths N_l in each level l. Once those values are determined, we can give an estimate of $E[f(X)]$ and the total computing complexity of obtaining it (see Section 5.3 for the details of how to determine L and N_l). It turns out that the multilevel Monte Carlo method is very efficient and can reduce the total computational complexity by a large extent. For example, to reach RMSE $O(\varepsilon)$, the total computational cost of the Monte Carlo estimation with EM scheme needs to be $O(\varepsilon^{-3})$, while it is only $O(\varepsilon^{-2}(\log \varepsilon)^2)$ for the multilevel Monte Carlo method (see Theorem 5.1 for more details). Another interesting point about the multilevel Monte Carlo method is that it depends heavily on the strong convergence of a numerical scheme to estimate the variance of the Monte Carlo estimation at each level if the test function is Lipschitz continuous, although our target (find $E[g(X(T))]$) is a weak convergence type problem. What is even more interesting is that we can combine the strong convergence and the multilevel Monte Carlo method to estimate $E[g(X(T))]$ without knowing the weak convergence of the numerical scheme and without excessively increasing the total computational cost, if the test function is Lipschitz continuous. This idea is extremely attractive considering the weak convergence of a numerical scheme requires too much on the smoothness of the coefficients of the stochastic differential equation and the test function, while the strong convergence does not (see more details in Section 5.3).

Organization of the Dissertation

In Chapter 2 we give preliminaries of the theory of SDEs that are needed in our dissertation. Chapter 3 is mainly focused on the strong convergence of the numerical solutions of SDEs driven by Brownian motions under non-globally Lipschitz conditions. Some numerical experiments are also presented. In Chapter 4 we state and prove a new weak convergence theorem under some mild conditions mentioned above. In Chapter 5, we introduce the multilevel Monte Carlo method. Finally, in Chapter 6, we present a new result on the strong convergence of the numerical solutions of SDEs driven by Lévy processes under non-globally Lipschitz conditions. Some important and often used inequalities as well as their proofs are included in Appendix A. Simulations and figures were obtained using Matlab. The Matlab codes are provided in Appendix B.

Chapter 2

Preliminaries on Stochastic Differential Equations

This chapter provides the preliminaries for the whole dissertation. We give an overview of stochastic differential equations driven by Brownian motion or Lévy motion. We shall introduce the existence and uniqueness theorems of such equations. We shall also introduce the connection between stochastic differential equations driven by Brownian motion and partial differential equations, which is indispensable in the analysis of weak approximations of such stochastic differential equations. For a thorough introduction of the theory of stochastic differential equations, we refer to [2, 31, 39, 45].

2.1 Existence and Uniqueness

Throughout this dissertation, $(\Omega, \mathcal{F}, \mathbb{F} = \{\mathcal{F}_t\}_{t\geq 0}, P)$ will always denote a filtered probability space satisfying the usual hypothesis of right-continuity and completeness.

We first consider the stochastic differential equations driven by Brownian motion:

$$\begin{cases} dX(t) = \mu(t, X(t))dt + \sigma(t, X(t))dW(t), & t \in (0, T] \\ X(0) = x_0, \end{cases} \tag{2.1}$$

where $X(t) \in \mathbb{R}^m$ for all $t \in [0, T]$, $W(t)$ is a d-dimensional Brownian motion (Wiener process) starting at 0, $\mu : [0, T] \times \mathbb{R}^m \to \mathbb{R}^m$ and $\sigma : [0, T] \times \mathbb{R}^m \to \mathbb{R}^{m\times d}$. We also assume that x_0 is $\mathcal{F}_0$-measurable and independent of $(W(t), 0 \le t \le T)$. The boundedness condition on x_0 can be flexible. For now we only assume that $E[x_0^2]$ is finite.

The following Lipschitz and linear growth condition are standard in the theory of stochastic differential equations.

- (Lipschitz condition) For all $x, y \in \mathbb{R}^m$ and all $t \in [0, T]$,

$$\|\mu(t, x) - \mu(t, y)\| + \|\sigma(t, x) - \sigma(t, y)|| \le K(T)\|x - y\|. \tag{2.2}$$

- (Linear growth condition) For all $(t, x) \in [0, T] \times \mathbb{R}^m$,

$$\|\mu(t, x)\| + \|\sigma(t, x)\| \le K(T)(1 + \|x\|). \tag{2.3}$$

In (2.2) and (2.3), the constant K is positive and only depends on T. Sometimes, we also use the following Lipschitz and linear growth condition interchangeably.

- (Lipschitz condition) For all $x, y \in \mathbb{R}^m$ and all $t \in [0, T]$,

$$\|\mu(t, x) - \mu(t, y)\|^2 \vee \|\sigma(t, x) - \sigma(t, y)\|^2 \le K(T)\|x - y\|^2. \tag{2.4}$$

- (Linear growth condition) For all $(t, x) \in [0, T] \times \mathbb{R}^m$,

$$\|\mu(t, x)\|^2 \vee \|\sigma(t, x)\|^2 \le K(T)(1 + \|x\|^2). \tag{2.5}$$

Here, $a \vee b := \max(a, b)$ for any $a, b \in \mathbb{R}$.

Theorem 2.1 (existence and uniqueness, [31], Theorem 5.4)**.** *Suppose μ and σ satisfy the Lipschitz condition (2.2) and linear growth condition (2.3), and x_0 is independent of $(W(t), 0 \le t \le T)$ with $E[\|x_0\|^2] < \infty$, then the equation (2.1) has a unique solution and satisfies*

$$E\Big[\sup_{0 \le t \le T} \|X(t)\|^2\Big] < C(1 + E[\|x_0\|^2]),$$

where C depends only on K and T.

Throughout this dissertation, we use $C > 0$ to denote a generic constant which varies at different occurrences. If needed, the parameters on which C depends will also be specified in the parentheses after it.

Actually, the Lipschitz condition can be replaced by the local Lipschitz condition:

- (Local Lipschitz condition) For every real number $R > 0$ and $T > 0$, there exists a positive constant K, depending on T and R, such that for all $t \in [0, T]$ and all $x, y \in \mathbb{R}^m$ with $\|x\|, \|y\| \le R$,

$$\|\mu(t, x) - \mu(t, y)\| + \|\sigma(t, x) - \sigma(t, y)\| \le K(T, R)\|x - y\|. \tag{2.6}$$

This condition is locally Lipschitz in x uniformly in t.

The existence and uniqueness theorem still holds under the local Lipschitz condition.

Theorem 2.2 (existence and uniqueness, [31], Theorem 5.4)**.** *Suppose μ and σ satisfy the local Lipschitz condition (2.6) and linear growth condition (2.3), and x_0 is independent of $(W(t), 0 \le t \le T)$ with $E[\|x_0\|^2] < \infty$, then the equation (2.1) has a unique solution and satisfies*

$$E\Big[\sup_{0 \le t \le T} \|X(t)\|^2\Big] < C(1 + E[\|x_0\|^2]),$$

where C depends only on K and T.

Having the local Lipschitz condition, many functions such as functions having continuous partial derivatives of first order with respect to x on $[0, T] \times \mathbb{R}^m$ can serve as the drift and diffusion coefficients. But it still excludes some common functions like $-|x|^2 x$ as the coefficients. The following theorem relaxes the linear growth condition.

Theorem 2.3 (existence and uniqueness, [39], Theorem 2.3.5)**.** *Assume that the local Lipschitz condition (2.6) holds, but the linear growth condition (2.5) is replaced with the following monotone condition: there exists a positive constant C such that for all $(t, x) \in [0, T] \times \mathbb{R}^m$,*

$$x^T \mu(t, x) + \frac{1}{2}\|\sigma(t, x)\|^2 \leq C(1 + \|x\|^2). \tag{2.7}$$

Then there exists a unique solution $X(t)$ to equation (2.1) and satisfies

$$E \int_0^T \|X(t)\|^2 dt < \infty.$$

For example, consider the following SDE:

$$dX(t) = [X(t) - X^3(t)]dt + X^2(t)dW(t), \quad t \in [0, T].$$

Although the coefficients are local Lipschitz continuous, they do not satisfy the linear growth condition. Nevertheless, the monotone condition is satisfied:

$$x(x - x^3) + \frac{1}{2}x^4 \leq x^2 \leq 1 + x^2.$$

Therefore by Theorem 2.3, it admits a unique solution.

We conclude this section by giving the L^p-estimates of the solution of (2.1).

Theorem 2.4 ([39], Theorem 2.4.1)**.** *Assume $X(t)$ is the unique solution of the equation* (2.1)*. Let $p \geq 2$ and $x_0 \in L^p(\Omega; \mathbb{R}^m)$. Assume that there exists a constant*

$\alpha > 0$ *such that for all* $(t, x) \in [0, T] \times \mathbb{R}^m$,

$$x^T \mu(t, x) + \frac{p-1}{2} \|\sigma(t, x)\|^2 \leq \alpha(1 + \|x\|^2). \tag{2.8}$$

Then

$$E[\|X(t)\|^p] \leq C := 2^{\frac{p-2}{2}}(1 + E[\|x_0\|^p])e^{p\alpha t} \tag{2.9}$$

for all $t \in [0, T]$.

Note that the linear growth condition (2.3) is just a special case of (2.8). So the above L^p-estimate is also true if the linear growth condition is fulfilled.

Corollary 2.4.1 ([39], Corollary 2.4.2)**.** *Let* $p \geq 2$ *and* $x_0 \in L^p(\Omega; \mathbb{R}^m)$. *Assume that the linear growth condition* (2.3) *holds. Then inequality* (2.9) *holds with* $\alpha = \sqrt{K} + K(p-1)/2$.

2.2 Stochastic Differential Equations and Partial Differential Equations

There is a close relation between stochastic differential equations and partial differential equations (PDE). The Kolmogorov backward equation is one of the most important and useful relations between the two. This PDE will play an important role in our analysis of weak approximations of the solutions of SDEs in Chapter 4.

Theorem 2.5 (Kolmogorov's Equation, [31], Theorem 6.9)**.** *Let* $X(t)$ *be the solution of the equation (2.1) with* $m = 1$. *Assume that the coefficients* $\mu(t, x)$ *and* $\sigma(t, x)$ *are locally Lipschitz and satisfy the linear growth condition. Assume in addition that they possess continuous partial derivatives with respect to* x *up to order two, and that they have at most polynomial growth. If* $g(x)$ *is twice continuously differentiable and satisfies together with its derivatives a polynomial growth condition, then the function*

$f(t,x) = E[g(X(T))|X(t) = x]$ satisfies

$$\begin{cases} \frac{\partial f}{\partial t}(t,x) + \mu(t,x)\frac{\partial f}{\partial x}(t,x) + \frac{1}{2}\sigma^2(t,x)\frac{\partial^2 f}{\partial x^2}(t,x) = 0, & t \in [0,T), x \in \mathbb{R}, \\ f(T,x) = g(x). \end{cases} \tag{2.10}$$

The multi-dimensional version can be found in, e.g. [45]. For convenience, we often use the following second order differential operator

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

and write equation (2.10) as

$$\begin{cases} \frac{\partial f}{\partial t}(t,x) + L_t f(t,x) = 0, & t \in [0,T), x \in \mathbb{R}, \\ f(T,x) = g(x). \end{cases} \tag{2.11}$$

To discuss the differential properties of the function $f(t,x)$, we use a more general version of equation (2.1):

$$\begin{cases} dX(\theta) = \mu(\theta, X(\theta))d\theta + \sigma(\theta, X(\theta))dW(\theta), & \theta \in (t,T] \\ X(t) = x. \end{cases} \tag{2.12}$$

Note that here $X(\theta)$ is still $\mathbb{R}^m$-valued. We write the solution in the form $X^{t,x}(\theta)$ to represent the dependence of $X(\theta)$ on the initial data (t,x). Therefore, $f(t,x)$ can be written as $f(t,x) = E[g(X^{t,x}(T))]$.

The differentiability of $X^{t,x}(\theta)$ with respect to x depends on the smoothness of the coefficients μ and σ.

Proposition 2.6 ([38], Theorem 2.3.3)**.** *Let k be a positive integer and $0 < \alpha \leq 1$. Suppose that coefficients μ and σ are $C^{k,\alpha}$ functions of x for some α and their derivatives up to k-th order are bounded. Then the solution $X^{t,x}(\theta)$ is a $C^{k,\beta}$ function of x for any β less than α.*

With this theorem, we can discuss the differentiability of $f(t,x)$ due to its definition $f(t,x) = E[X^{t,x}(T)]$. More details can be found in Chapter 4.

2.3 Numerical Solutions of Stochastic Differential Equations Driven by Brownian Motion

It is common that the equation (2.1) does not have a closed-form solution in many cases, where a numerical solution becomes a necessity. But unlike in the deterministic differential equations, there are two kinds of convergence of the numerical solutions of SDEs. The first kind of convergence is the *strong convergence*.

Definition 2.7. *Suppose Y is a discrete-time approximation of the solution $X(t)$ of (2.1) with maximum step size $\Delta > 0$. We say that Y converges to $X(t)$ in the strong sense with order $\gamma \in (0, \infty]$ if there exists a finite constant $C > 0$ and a positive constant Δ_0 such that*

$$E[\|X(T) - Y(T)\|] \leq C\Delta^{\gamma} \tag{2.13}$$

for any time discretization with maximum step size $\Delta \in (0, \Delta_0)$.

The other kind of convergence is the *weak convergence*.

Definition 2.8. *Suppose Y is a discrete-time approximation of the solution $X(t)$ of (2.1) with maximum step size $\Delta > 0$. We say that Y converges to $X(t)$ in the weak sense with order $\beta \in (0, \infty]$ if for any function $g : \mathbb{R}^m \to \mathbb{R}$ in a suitable function space there exists a finite constant $C > 0$ and a positive constant Δ_0 such that*

$$|E[g(X(T))] - E[g(Y(T))]| \leq C\Delta^{\beta} \tag{2.14}$$

for any time discretization with maximum step size $\Delta \in (0, \Delta_0)$.

The function space in the Definition 2.8 can be flexible. For example, it can be the space of all polynomial functions. It can also be the space $C_p^k(\mathbb{R}^m)$ in which all the functions are k-th continuously differentiable and all their partial derivatives up to order k have polynomial growth.

Complete reviews of the numerical solutions of SDEs driven by Brownian motion can be found in, e.g. [22, 33, 48].

In the following we first introduce the most commonly used numerical scheme to solve (2.1), the Euler-Maruyama (EM) scheme. Given a fixed integer $N > 0$, set the time step $\Delta t = T/N$. For any integer k satisfying $0 \leq k \leq N$, set $t_k = k\Delta t$. We define at each node in $[0, T]$: $Y_0 := x_0$ and

$$Y_{k+1} := Y_k + \mu(t_k, Y_k)\Delta t + \sigma(t_k, Y_k)\Delta W_k, \qquad 0 \leq k \leq N-1, \tag{2.15}$$

where $\Delta W_k = W(t_{k+1}) - W(t_k)$. Furthermore, we define the continuous-time approximation of the solution of (2.1) as:

$$\begin{aligned}\overline{Y}(t) &:= Y_k + \mu(t_k, Y_k)(t - t_k) + \sigma(t_k, Y_k)(W(t) - W(t_k)) \\ &= Y_k + \int_{t_k}^{t} \mu(t_k, Y_k)ds + \int_{t_k}^{t} \sigma(t_k, Y_k)dW(s) \quad \text{for } t \in [t_k, t_{k+1}).\end{aligned} \tag{2.16}$$

It is obvious that $\overline{Y}(t_k) = Y_k$. If we define the shift operator in the following way:

$$\eta(t) = t_k, \quad t \in [t_k, t_{k+1}),$$

then scheme (2.16) can be written as

$$\overline{Y}(t) = Y_0 + \int_0^t \mu(\eta(t), \overline{Y}(\eta(t)))ds + \int_0^t \sigma(\eta(t), \overline{Y}(\eta(t)))dW(s), \quad t \in [0, T]. \tag{2.17}$$

Similar to getting Corollary 2.4.1, if $\mu(t, x)$ and $\sigma(t, x)$ satisfy the linear growth condition and $x_0 \in L^p(\Omega),\ p \geq 2$, it then follows that,

$$\sup_{0 \leq t \leq T} \|\overline{Y}(t)\| \in L^p(\Omega). \tag{2.18}$$

There are also many other numerical schemes to solve (2.1), such as the Milstein scheme, the Runge-Kutta type scheme, the Itô-Taylor expansion scheme, etc.. We refer to [33] for a thorough treatment of the common numerical schemes we encounter in the field of numerical solutions of SDEs.

For the Euler-Maruyama approximate solutions, the strong convergence order is $\frac{1}{2}$. This is the following theorem.

Theorem 2.9 ([39], Theorem 2.7.3)**.** *Assume that the Lipschitz condition (2.2) and the linear growth condition (2.3) hold. Let $X(t)$ be the unique solution of equation (2.1), and $\overline{Y}(t)$ be its Euler-Maruyama approximate solution. Then*

$$E\Big[\sup_{0\le t\le T}\|\overline{Y}(t)-X(t)\|^2\Big]\le\frac{C}{N},\tag{2.19}$$

where the constant C depends on K, T and $E[\|x_0\|^2]$.

If μ and σ do not depend on the time variable t and satisfy the global Lipschitz condition, then for any $p\ge 1$, we also have

$$E\Big[\sup_{0\le t\le T}\|X(t)-\overline{Y}(t)\|^p\Big]\le\frac{C(p,T)}{N^{p/2}}.\tag{2.20}$$

See e.g. Bouleau and Lepingle [8].

In fact, the above convergence is stronger than the strong convergence we defined in Definition 2.7. The error estimation is uniform with respect to the whole sample path rather than just the terminal value of it. This advantage can be very useful in many cases. For example, in the context of the Asian options, the payoff of the option at the expiration time depends on the average of the whole sample path

$$\frac{1}{T}\int_0^T S(t)dt,$$

where $S(t)$ is the price of underlying stock.

We also remark that the assumed globally Lipschitz condition and the linear growth condition in Theorem 2.9 are very strong conditions and may fail to hold in many situations. For example, the one-dimensional stochastic Ginzburg-Landau equation takes the form

$$\begin{cases} dX(t) = (X(t) - X^3(t))dt + X(t)dW(t), \quad t \in (0,1], \\ X(0) = 1, \end{cases}$$

The drift coefficient takes the form $\mu(x) = x - x^3$, which is clearly not globally Lipschitz continuous. But it is continuously differentiable and thus locally Lipschitz continuous. In fact, the family of SDEs with C^1 drift and diffusion coefficients consist of a very large part of SDEs we encounter in research. Hu [23] and Higham, Mao and Stuart [24] were the first to study the strong convergence problem of EM approximate solutions under the non-globally Lipschitz continuous conditions. After that, the study of numerical solutions of SDEs with local Lipschitz continuous coefficients has been a very active area. We will give a thorough introduction of such problems in Chapter 3.

As for the weak convergence of the Euler-Maruyama scheme, the convergence order is typically 1, but it can be under different conditions. For example, in Theorem 14.1.5 of [33], to achieve weak convergence order 1, it assumes $\mu(t,x)$ and $\sigma(t,x)$ being twice continuously differentiable and the test function $g(x)$ being fourth continuously differentiable, together with some other conditions (Hölder continuity, etc.). While in Theorem 14.5.1 of [33], it assumes a homogeneous equation and functions $\mu(x)$, $\sigma(x)$ and $g(x)$ being in $C_p^4(\mathbb{R}^m)$, among some other conditions. Both of the two theorems require strongly smooth conditions on μ, σ and g. In [5], also for the homogeneous equation, the test function is assumed to be only measurable and bounded (or with polynomial growth), but $\mu(x)$ and $\sigma(x)$ are required to be C^∞ with bounded derivatives of any order. In Chapter 4, we will give some improvements of the conditions assumed on μ, σ and g.

2.4 Stochastic Differential Equations Driven by Lévy Processes

Unlike Brownian motion, Lévy processes are stochastic processes allowing for jumps in their sample paths. In the following, we will give a short introduction to such processes. For an excellent and intuitive introduction to Lévy processes, we refer to [46]. We also refer to [56] for a thorough introduction to the infinitely divisible distributions and [2] for stochastic calculus with respect to Lévy processes. For Lévy processes in finance, see e.g. [10, 49, 57, 58].

In general, we say a càdlàg (right continuous with left limits) and adapted stochastic process $L = L(t), 0 \leq t \leq T$ defined on a filtered probability space $(\Omega, \mathcal{F}, \mathbb{F} = (\mathcal{F}_t)_{t \leq 0}, P)$ is a Lévy process if the following conditions are satisfied:

(L1) $L(0) = 0$ a.s.;

(L2) L has independent and stationary increments;

(L3) L is stochastically continuous, i.e. for all $a > 0$ and for all $s \geq 0$

$$\lim_{t \to s} P(\|L(t) - L(s)\| > a) = 0.$$

By definition, Brownian motion is a special Lévy process. Other examples of Lévy processes are like Poisson process, compound Poisson process, α-stable process, etc. [2, 7, 10, 56].

It is often convenient to use *Poisson random measure* to analyze the jumps of a Lévy process. Consider a set $A \in \mathcal{B}(\mathbb{R}^d \backslash \{0\})$ such that $0 \notin \overline{A}$ and let $0 \leq t \leq T$. Define the random measure of the jumps of the process L by

$$\begin{aligned} N(\omega, t, A) &= \#\{0 \leq s \leq t; \Delta L(s, \omega) \in A\} \\ &= \sum_{s \leq t} \mathbb{1}_A(\Delta L(s, \omega)). \end{aligned} \tag{2.21}$$

Therefore, $N(\omega, t, A)$ counts the jumps of the process L of size in A up to time t. It can be verified that for fixed A, $N(\omega, t, A)$ is a Poisson process with intensity $\nu(A) = E[N(\omega, 1, A)]$ and for fixed t, N is a Poisson random measure. The compensated Poisson measure is then defined as

$$\tilde{N}(t, A) := N(t, A) - t\nu(A).$$

Definition 2.10. *The measure ν defined by*

$$\nu(A) = E[N(\omega, 1, A)] = E\Big[\sum_{s \leq 1} \mathbb{1}_A(\Delta L(s, \omega))\Big]$$

is called the Lévy measure of the Lévy process L.

In general, the Lévy measure describes the expected number of jumps of a certain size in a time interval of length 1 and satisfies

$$\nu(\{0\}) = 0, \quad \text{and} \quad \int_{\mathbb{R}^d} (1 \wedge \|x\|^2)\nu(dx) < \infty.$$

It can be proved that if $\nu(\mathbb{R}^d) < \infty$, then almost all paths of L have a finite number of jumps on every compact interval. In this case, the Lévy process has finite activity. If $\nu(\mathbb{R}^d) = \infty$, then almost all paths of L have an infinite number of jumps on every compact interval. In this case, the Lévy process has infinite activity. See e.g. Theorem 21.3 in Sato [56] for the proof.

We can also define an integral with respect to the Poisson random measure N. Consider a set $A \in \mathcal{B}(\mathbb{R}^d \backslash \{0\})$ such that $0 \notin \overline{A}$ and a function $f : \mathbb{R}^d \to \mathbb{R}^m$, Borel measurable and finite on A. We define the integral with respect to N as follows:

$$\int_A f(x)N(t, dx) = \sum_{s \leq t} f(\Delta L(s))\mathbb{1}_A(\Delta L(s)).$$

Note that the above integral is a $\mathbb{R}^m$-valued stochastic process. In the following, we use

$$\int_0^t \int_A f(x) N(ds, dx)$$

to denote this process. Similarly, for $f \in L^1(A)$, we define

$$\int_0^t \int_A f(x) \tilde{N}(ds, dx) = \int_0^t \int_A f(x) N(ds, dx) - t \int_A f(x) \nu(dx).$$

With the help of Poisson random measure, we have the following decomposition of a Lévy process.

Theorem 2.11 (Lévy-Itô Decomposition, [2], Theorem 2.4.16)**.** *Let L be a $\mathbb{R}^d$-valued Lévy process, then there exists $b \in \mathbb{R}^d$, a Brownian motion $W_A(t)$ with covariance matrix A and an independent Poisson random measure N on $\mathbb{R}^+ \times (\mathbb{R}^d - \{0\})$ such that, for each $t \geq 0$,*

$$L(t) = bt + W_A(t) + \int_0^t \int_{\|x\|<1} x \tilde{N}(ds, dx) + \int_0^t \int_{\|x\| \geq 1} x N(ds, dx). \tag{2.22}$$

Sometimes it is convenient to write

$$W_A(t) = (W_A^1(t), \cdots, W_A^d(t))$$

in the form

$$W_A^i(t) = \sum_{j=1}^m \sigma_j^i W^j(t),$$

where $W^1, \cdots, W^m$ are standard one-dimensional Brownian motions and σ is a $d \times m$ real-valued matrix for which $\sigma\sigma^T = A$. If L is only a real-valued Lévy process, the term $W_A(t)$ can be replaced by $\sigma W(t)$, where $\sigma \geq 0$ and $W(t)$ is a standard one-dimensional Brownian motion.

It can be proved that $L(t) \in L^p(\Omega), p \geq 1$ if and only if $\int_{\|x\| \geq 1} \|x\|^p \nu(dx) < \infty$. In particular, $L(t) \in L^1(\Omega)$ if and only if $\int_{\|x\| \geq 1} \|x\| \nu(dx) < \infty$. Therefore, if we assume

$E[\|L(t)\|] < \infty$, we can rewrite (2.22) as

$$L(t) = b_1 t + W_A(t) + \int_0^t \int_{\mathbb{R}^d} x\tilde{N}(ds, dx), \tag{2.23}$$

where $b_1 = b + \int_{\|x\|\geq 1} x\nu(dx)$.

In general, the simulation of a Lévy process is more complex than a Brownian motion. The simulation method varies from one kind of Lévy process to another. We refer to Cont and Tankov [10], Platen and Bruti-Liberati [49], Asmussen and Rosinski [3], Rosinski [52, 53] for the details.

In view of (2.22), we consider the following stochastic differential equation driven by a stochastic process with jumps,

$$\begin{aligned} X(t) = X(0) &+ \int_0^t a(X(s-))ds + \int_0^t b(X(s-))dW(s) \\ &+ \int_0^t \int_{\|y\|<1} f(X(s-), y)\tilde{N}(ds, dy) + \int_0^t \int_{\|y\|\geq 1} g(X(s-), y)N(ds, dy), \end{aligned} \tag{2.24}$$

where $X(0)$ is $\mathcal{F}_0$-measurable, $X(t)$ is a $\mathbb{R}^m$-valued stochastic process, W and N are independent of $\mathcal{F}_0$, $a : \mathbb{R}^m \to \mathbb{R}^m$, $b : \mathbb{R}^m \to \mathbb{R}^{m\times d}$, $f : \mathbb{R}^m \times \mathbb{R}^d \to \mathbb{R}^m$ and $g : \mathbb{R}^m \times \mathbb{R}^d \to \mathbb{R}^m$. There exists a unique solution to the equation if the following conditions are satisfied [2, 50]:

- Lipschitz condition: there exists a constant $C > 0$ such that for all $x_1, x_2 \in \mathbb{R}^m$,

$$\begin{aligned} &\|a(x_1) - a(x_2)\|^2 + \|B(x_1, x_1) - 2B(x_1, x_2) + B(x_2, x_2)\| \\ &\quad + \int_{\|y\|<1} \|f(x_1, y) - f(x_2, y)\|^2 \nu(dy) \leq C\|x_1 - x_2\|^2; \end{aligned}$$

- Growth condition: there exists a constant $C > 0$ such that for all $x \in \mathbb{R}^m$,

$$\|a(x)\|^2 + \|B(x, x)\| + \int_{\|y\|<1} \|f(x, y)\|^2 \nu(dy) \leq C(1 + \|x\|^2);$$

- Big jump condition: g is jointly measurable and $x \mapsto g(x,y)$ is continuous for any $y \in \{y : \|y\| \geq 1\}$.

Here $B(x_1, x_2) = b(x_1)b(x_2)^T$ and we use the seminorm on the matrix B:

$$\|B\| = \sum_{i=1}^{m} |B_{ii}|.$$

In view of (2.23), we sometimes consider the following SDE:

$$X(t) = X(0) + \int_0^t a(X(s-))ds + \int_0^t b(X(s-))dW(s) + \int_0^t \int_{\mathbb{R}^d} f(X(s-), y)\tilde{N}(ds, dy). \tag{2.25}$$

It can be proved that there exists a unique solution to (2.25) if the following conditions are satisfied [21]:

A-1. There exists a constant C such that for any $x \in \mathbb{R}^m$,

$$\langle x, a(x)\rangle + \|\sigma(x)\|^2 + \int_{\mathbb{R}^d} \|f(x,y)\|^2 \nu(dy) \leq C(1 + \|x\|^2).$$

A-2. For every $R > 0$, there exists a constant $C(R)$, depending on R, such that for any $\|x_1\|, \|x_2\| \leq R$,

$$\langle x_1 - x_2, a(x_1) - a(x_2)\rangle + \|b(x_1) - b(x_2)\|^2 + \int_{\mathbb{R}^d} \|f(x_1, y) - f(x_2, y)\|^2 \nu(dy)$$
$$\leq C(R)\|x_1 - x_2\|^2.$$

A-3. The function $a(x)$ is continuous in $x \in \mathbb{R}^m$.

Condition A-1 is a monotone condition. Condition A-2 states that a satisfies the one-sided local Lipschitz condition and b and f satisfy the local Lipschitz condition. Furthermore, if $E[\|X(0)\|^p] < \infty$ for some $p \geq 2$ and if there exists a constant $C_1 > 0$, such that

$$\int_{\mathbb{R}^d} \|f(x,y)\|^p \nu(dy) \leq C_1(1 + \|x\|^p)$$

for any $x \in \mathbb{R}^m$, then we have

$$E\Big[\sup_{0\le t\le T}\|X(t)\|^p\Big] \le C, \tag{2.26}$$

with $C := C(T, p, C_1, E[\|X(0)\|^p])$. See e.g. Lemma 2.2 in [11] for the proof.

Chapter 3

Strong Convergence of Numerical Approximations of SDEs Driven by Brownian Motion under Local Lipschitz Conditions

3.1 Strong Convergence of Euler-Maruyama Approximations of SDEs under Local Lipschitz Conditions

We first consider the following stochastic differential equation with homogeneous coefficients:

$$\begin{cases} dX(t) = \mu(X(t))dt + \sigma(X(t))dW(t), \quad t \in (0,T] \\ X(0) = x_0. \end{cases} \tag{3.1}$$

Like in the assumptions of (2.1), $X(t) \in \mathbb{R}^m$ for all $t \in [0,T]$, $W(t)$ is a d-dimensional Brownian motion starting at 0, $\mu : \mathbb{R}^m \to \mathbb{R}^m$ and $\sigma : \mathbb{R}^m \to \mathbb{R}^{m\times d}$, x_0 is $\mathcal{F}_0$-measurable and independent of $(W(t), 0 \le t \le T)$. But in this section, we assume that all the pth moments of x_0, $p > 0$ are finite.

In this case, the local Lipschitz condition is

- (Local Lipschitz Condition) For every real number $R > 0$, there exists a positive constant C, depending only on R, such that for all $x, y \in \mathbb{R}^m$ with $\|x\|, \|y\| \le R$,

$$\|\mu(x) - \mu(y)\| + \|\sigma(x) - \sigma(y)\| \le C(R)\|x - y\|, \tag{3.2}$$

or

$$\|\mu(x) - \mu(y)\|^2 \vee \|\sigma(x) - \sigma(y)\|^2 \le C(R)\|x - y\|^2. \tag{3.3}$$

We still use the notations from section (2.3) to express the approximate solution of (3.1). Given the homogeneous coefficients, the discrete approximation in this section takes the form:

$$Y_{k+1} := Y_k + \mu(Y_k)\Delta t + \sigma(Y_k)\Delta W_k, \qquad 0 \le k \le N-1. \tag{3.4}$$

Furthermore, the continuous-time approximation of the solution of (3.1) is

$$\overline{Y}(t) := Y_k + \mu(Y_k)(t - t_k) + \sigma(Y_k)(W(t) - W(t_k)) \quad \text{for } t \in [t_k, t_{k+1}). \tag{3.5}$$

Sometimes, it is more convenient to work with the equivalent definition

$$\overline{Y}(t) := Y_0 + \int_0^t \mu(Y(s))ds + \int_0^t \sigma(Y(s))dW(s), \tag{3.6}$$

where $Y(t)$ is defined by

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

It is obvious that $\overline{Y}(t_k) = Y(t_k) = Y_k$. The following theorem is about the strong convergence of the Euler-Maruyama approximate solution of equation (3.1) under the local Lipschitz condition.

Theorem 3.1 ([23], Theorem 2.2)**.** *Suppose the coefficients μ and σ in equation (3.1) satisfy the local Lipschitz condition and for some $p > 2$ there is a constant A such that*

$$E\Big[\sup_{0\le t\le T} \|\overline{Y}(t)\|^p\Big] \vee E\Big[\sup_{0\le t\le T} \|X(t)\|^p\Big] \le A. \tag{3.7}$$

Then the Euler-Maruyama solution (3.5) satisfies

$$\lim_{\Delta t\to 0} E\Big[\sup_{0\le t\le T} \|\overline{Y}(t) - X(t)\|^2\Big] = 0. \tag{3.8}$$

Proof. First, we define

$$\begin{aligned}
\tau_R &:= \inf\{t \ge 0 : \|\overline{Y}(t)\| \ge R\},\\
\rho_R &:= \inf\{t \ge 0 : \|X(t)\| \ge R\},\\
\theta_R &:= \tau_R \wedge \rho_R,
\end{aligned}$$

and

$$e(t) := \overline{Y}(t) - X(t).$$

Recall the Young inequality: for $r^{-1} + q^{-1} = 1$,

$$ab \le \frac{\delta}{r} a^r + \frac{1}{q\delta^{q/r}} b^q, \quad \forall a, b, \delta > 0.$$

We thus have for any $\delta > 0$

$$\begin{aligned}
E\Big[\sup_{0\le t\le T} \|e(t)\|^2\Big] &= E\Big[\sup_{0\le t\le T} \|e(t)\|^2 \mathbb{1}_{\{\tau_R > T, \rho_R > T\}}\Big] + E\Big[\sup_{0\le t\le T} \|e(t)\|^2 \mathbb{1}_{\{\tau_R \le T \text{ or } \rho_R \le T\}}\Big]\\
&\le E\Big[\sup_{0\le t\le T} \|e(t\wedge\theta_R)\|^2 \mathbb{1}_{\{\theta_R > T\}}\Big] + \frac{2\delta}{p} E\Big[\sup_{0\le t\le T} \|e(t)\|^p\Big]
\end{aligned}$$

$$+\frac{1-2/p}{\delta^{2/(p-2)}}P(\tau_R \leq T \text{ or } \rho_R \leq T) \tag{3.9}$$

Now

$$P(\tau_R \leq T) = E\Big[\mathbb{1}_{\{\tau_R \leq T\}}\frac{\|\overline{Y}(\tau_R)\|^p}{R^p}\Big] \leq \frac{1}{R^p}E\Big[\sup_{0\leq t\leq T}\|\overline{Y}(t)\|^p\Big] \leq \frac{A}{R^p},$$

using (3.7). A similar result can be derived for ρ_R such that

$$P(\tau_R \leq T \text{ or } \rho_R \leq T) \leq P(\tau_R \leq T) + P(\rho_R \leq T) \leq \frac{2A}{R^p}.$$

Using these bounds along with

$$E\Big[\sup_{0\leq t\leq T}\|e(t)\|^p\Big] \leq 2^{p-1}E\Big[\sup_{0\leq t\leq T}(\|\overline{Y}(t)\|^p + \|X(t)\|^p)\Big] \leq 2^pA$$

in (3.9) gives

$$\begin{aligned} E\Big[\sup_{0\leq t\leq T}\|e(t)\|^2\Big] \leq E\Big[&\sup_{0\leq t\leq T}\|\overline{Y}(t\wedge\theta_R) - X(t\wedge\theta_R)\|^2\Big] \\ &+ \frac{2^{p+1}\delta A}{p} + \frac{(p-2)2A}{p\delta^{2/(p-2)}R^p}. \end{aligned} \tag{3.10}$$

Using

$$X(t\wedge\theta_R) := x_0 + \int_0^{t\wedge\theta_R}\mu(X(s))ds + \int_0^{t\wedge\theta_R}\sigma(X(s))dW(s),$$

(3.6) and Cauchy-Schwarz, we have

$$\begin{aligned} &\|\overline{Y}(t\wedge\theta_R) - X(t\wedge\theta_R)\|^2 \\ &= \Big\|\int_0^{t\wedge\theta_R}\mu(Y(s)) - \mu(X(s))ds + \int_0^{t\wedge\theta_R}\sigma(Y(s)) - \sigma(X(s))dW(s)\Big\|^2 \\ &\leq 2\Big[T\int_0^{t\wedge\theta_R}\|\mu(Y(s)) - \mu(X(s))\|^2ds + \Big\|\int_0^{t\wedge\theta_R}\sigma(Y(s)) - \sigma(X(s))dW(s)\Big\|^2\Big] \end{aligned}$$

From the local Lipschitz condition (3.3) and Doob's martingale inequality (A.9) we have for any $\tau \le T$,

$$
\begin{aligned}
&E\Big[\sup_{0\le t\le\tau}\|\overline{Y}(t\wedge\theta_R)-X(t\wedge\theta_R)\|^2\Big]\\
&\le 2C(R)(T+4)E\int_0^{\tau\wedge\theta_R}\|Y(s)-X(s)\|^2ds\\
&\le 4C(R)(T+4)E\int_0^{\tau\wedge\theta_R}\big(\|Y(s)-\overline{Y}(s)\|^2+\|\overline{Y}(s)-X(s)\|^2\big)ds\\
&\le 4C(R)(T+4)\Big(E\int_0^{\tau\wedge\theta_R}\|Y(s)-\overline{Y}(s)\|ds+E\int_0^{\tau}\|\overline{Y}(s-\theta_R)-X(s-\theta_R)\|^2ds\Big)\\
&\le 4C(R)(T+4)\Big(E\int_0^{\tau\wedge\theta_R}\|Y(s)-\overline{Y}(s)\|^2ds\\
&\qquad\qquad+\int_0^{\tau}E\Big[\sup_{0\le r\le s}\|\overline{Y}(r\wedge\theta_R)-X(r\wedge\theta_R)\|^2\Big]ds\Big).
\end{aligned}
\tag{3.11}
$$

To bound the first term in the parentheses on the right-hand side of (3.11), given $s\in[0,T\wedge\theta_R)$, let k_s be the integer for which $s\in[t_{k_s},t_{k_s+1})$. Then

$$
\begin{aligned}
Y(s)-\overline{Y}(s) &= Y_{k_s}-\Big(Y_{k_s}+\int_{t_{k_s}}^{s}\mu(Y(s))ds+\int_{t_{k_s}}^{s}\sigma(Y(s))dW(s)\Big)\\
&= -\mu(Y_{k_s})(s-t_{k_s})-\sigma(Y_{k_s})(W(s)-W(t_{k_s})).
\end{aligned}
$$

Therefore,

$$
\|Y(s)-\overline{Y}(s)\|^2\le 2\Big(\|\mu(Y_{k_s})\|^2\Delta t^2+\|\sigma(Y_{k_s})\|^2\|W(s)-W(t_{k_s})\|^2\Big). \tag{3.12}
$$

By the local Lipschitz condition (3.3), for $\|x\|\le R$, we have

$$
\|\mu(x)\|^2\le 2(\|\mu(x)-\mu(0)\|^2+\|\mu(0)\|^2)\le 2\Big(C(R)\|x\|^2+\|\mu(0)\|^2\Big),
$$

and, similarly,

$$
\|\sigma(x)\|^2\le 2\Big(C(R)\|x\|^2+\|\sigma(0)\|^2\Big).
$$

Hence, in (3.12),

$$\|Y(s)-\overline{Y}(s)\|^2 \le 4\Big(C(R)\|Y_{k_s}\|^2+\|\mu(0)\|^2\vee\|\sigma(0)\|^2)(\Delta t^2+\|W(s)-W(t_{k_s})\|^2\Big).$$

Using (3.7) and the Lyapunov inequality (A.8)

$$\begin{aligned}
&E\int_0^{\tau\wedge\theta_R}\|Y(s)-\overline{Y}(s)\|^2ds\\
&\le E\int_0^{\tau\wedge\theta_R}4\Big(C(R)\|Y_{k_s}\|^2+\|\mu(0)\|^2\vee\|\sigma(0)\|^2\Big)\Big((\Delta t)^2+\|W(s)-W(t_{k_s})\|^2\Big)ds\\
&\le \int_0^{\tau}4E\Big[\Big(C(R)\|Y_{k_s}\|^2+\|\mu(0)\|^2\vee\|\sigma(0)\|^2\Big)\Big((\Delta t)^2+\|W(s)-W(t_{k_s})\|^2\Big)\Big]ds\\
&\le \int_0^{T}4\Big(C(R)E[\|Y_{k_s}\|^2]+\|\mu(0)\|^2\vee\|\sigma(0)\|^2\Big)\Big((\Delta t)^2+d\Delta t\Big)ds\\
&\le 4T\Big(C(R)A^{2/p}+\|\mu(0)\|^2\vee\|\sigma(0)\|^2\Big)\Delta t(\Delta t+d).
\end{aligned}$$

In (3.11) we then have

$$\begin{aligned}
&E\Big[\sup_{0\le t\le\tau}\|\overline{Y}(t\wedge\theta_R)-X(t\wedge\theta_R)\|^2\Big]\\
&\le 16C(R)(T+4)T\Delta t(\Delta t+d)\Big(C(R)A^{2/p}+\|\mu(0)\|^2\vee\|\sigma(0)\|^2\Big)\\
&\quad+4C(R)(T+4)\int_0^{\tau}E\sup_{0\le r\le s}\big[\|\overline{Y}(r\wedge\theta_R)-X(r\wedge\theta_R)\|^2\big]ds.
\end{aligned}$$

Applying the Gronwall inequality (A.7) we obtain

$$E\Big[\sup_{0\le t\le T}\|\overline{Y}(t\wedge\theta_R)-X(t\wedge\theta_R)\|^2\Big]\le C\Delta t(C(R)^2+1)e^{4C(R)(T+4)},$$

where C is a universal constant independent of Δt, R and δ. Inserting this into (3.10) gives

$$E\Big[\sup_{0\le t\le T}\|e(t)\|^2\Big]\le C\Delta t(C(R)^2+1)e^{4C(R)(T+4)}+\frac{2^{p+1}\delta A}{p}+\frac{(1-2/p)2A}{\delta^{2/(p-2)}R^p}. \quad (3.13)$$

Given any $\epsilon > 0$, we can choose δ so that $2^{p+1}\delta A/p < \epsilon/3$, then choose R so that

$$\frac{(1-2/p)2A}{\delta^{2/(p-2)}R^p} < \epsilon/3,$$

and then choose Δt sufficiently small for

$$C\Delta t(C(R)^2+1)e^{4C(R)(T+4)} < \epsilon/3$$

so that in (3.13),

$$E\Big[\sup_{0\le t\le T}\|e(t)\|^2\Big] \le \epsilon,$$

as required. □

This theorem establishes the strong convergence of Euler-Maruyama approximate solutions of (3.1). But the bounded condition it assumes for the pth moment of $X(t)$ and $\overline{Y}(t)$ is not satisfying. Although it may be possible to verify the bound of the pth moment of $X(t)$, as many textbooks have done it, the bound of the pth moment of $\overline{Y}(t)$ is often very difficult to verify and sometimes may fail to hold. Besides, the convergence (3.8) is a very general one and does not involve the explicit convergence rate.

To remove the bound restriction on the pth moment of $\overline{Y}(t)$, the same author in [23] proposed a new numerical scheme called the *split-step backward Euler (SSBE)* method, which is defined by taking $Z_0 = x_0$ and

$$Z_k^\star = Z_k + \Delta t\mu(Z_k^\star), \tag{3.14}$$

$$Z_{k+1} = Z_k^\star + \sigma(Z_k^\star)\Delta W_k. \tag{3.15}$$

They proved that the new SSBE method converges strongly without assuming any bound of the pth moment of the approximate solution. But more restrictions on the drift and diffusion coefficients are needed. This is the following theorem.

Theorem 3.2 ([23], Theorem 3.3)**.** *Suppose the functions μ and σ in (3.1) are C^1, and there exist a constant $C > 0$ such that*

$$\langle x - y, \mu(x) - \mu(y)\rangle \leq C\|x - y\|^2, \quad \forall x, y \in \mathbb{R}^m, \tag{3.16}$$

$$\|\sigma(x) - \sigma(y)\|^2 \leq C\|x - y\|^2, \quad \forall x, y \in \mathbb{R}^m. \tag{3.17}$$

Consider the SSBE (3.14)-(3.15) applied to the SDE (3.1) under the above assumption. There exists a continuous-time extension $\overline{Z}(t)$ of the numerical solution (so that $\overline{Z}(t_k) = Z_k$) for which

$$\lim_{\Delta t \to 0} E\Big[\sup_{0 \leq t \leq T} \|\overline{Z}(t) - X(t)\|^2\Big] = 0.$$

Proof. See [23] for the details. □

The condition this theorem assumes on the drift coefficient μ is called a *one-sided Lipschitz condition*. A good example is the following polynomial function

$$f(x) = -x^p + x, \quad \text{where } p \geq 3 \text{ is an odd integer.}$$

It can be easily verified that it satisfies the condition (3.16). By taking $y = 0$, (3.16) and (3.17) also imply

$$\langle \mu(x), x\rangle \vee \|\sigma(x)\|^2 \leq \alpha + \beta\|x\|^2, \quad \forall x \in \mathbb{R}^m, \tag{3.18}$$

where $\alpha := \frac{1}{2}\|\mu(0)\|^2 \vee 2\|\sigma(0)\|^2$ and $\beta := (C + \frac{1}{2}) \vee 2C$. Condition (3.18) is actually a monotone-type condition (see condition (2.7)). Note that μ and σ are also locally Lipschitz continuous (μ, σ are C^1). Therefore, by Theorem 2.3, under the assumptions of Theorem (3.2), there exists a unique solution of (3.1).

To remove the bound restriction on the approximate solution and give an explicit convergence rate at the same time, [23] also proposed the *Backward Euler* scheme by setting $U_0 = x_0$ and

$$U_{k+1} = U_k + \mu(U_{k+1})\Delta t + \sigma(U_k)\Delta W_k. \tag{3.19}$$

Theorem 3.3 ([23], Theorem 5.3)**.** *Suppose the conditions in Theorem (3.2) hold. Moreover, assume that there exist constants $C > 0$ and $q \in \mathbb{Z}^+$ such that all $x, y \in \mathbb{R}^m$,*

$$\|\mu(x) - \mu(y)\|^2 \leq C(1 + \|x\|^q + \|y\|^q)\|x - y\|^2. \tag{3.20}$$

Consider the backward Euler method (3.19) applied to SDE (3.1). There exists a continuous-time extension $\overline{U}(t)$ of the numerical solution (so that $\overline{U}(t_k) = U_k$) for which

$$E\Big[\sup_{0\leq t\leq T} \|\overline{U}(t) - X(t)\|^2\Big] = O(\Delta t). \tag{3.21}$$

Note that if μ satisfies condition (3.20), μ behaves polynomially or is superlinearly growing. Obviously, by the mean value theorem for derivatives, if the derivative of a function grows at most polynomially, this function must satisfy the condition (3.20). So sometimes we also define the polynomial growth of a function from the derivative perspective (see condition (3.23)).

This theorem provides a possibility of the computation of the numerical solutions of many SDEs which take nonlinear functions as the drift coefficients satisfying (3.16) and (3.20).

However, the backward Euler scheme (3.19) is an implicit method and its implementation requires too much computational effort. On the other hand, the explicit Euler-Maruyama scheme may not converge in the strong sense to the exact solution of an SDE with the one-sided Lipschitz continuous (inequality (3.16)) and superlinearly growing (inequality (3.20)) drift coefficients. Even worse, Theorem 1 in [27] also shows for such an SDE that the absolute moments of the explicit Euler approximations at a finite time point $T \in (0, \infty)$ diverge to infinity. To address this issue, Hutzenthaler, Jentzen and Kloeden [28] proposed a "tamed" version of the explicit Euler scheme which is strongly convergent for SDEs with superlinearly growing drift coefficients. This is our next section.

3.2 A Tamed Euler Scheme

In this section, we still consider the SDE (2.1), but in a different form. Let $\sigma = (\sigma_1, \sigma_2, \cdots, \sigma_d)$, where the σ_i's, $1 \le i \le d$, are the column vectors of the matrix σ. Let $W = (W^{(1)}, W^{(2)}, \cdots, W^{(d)})$ be the d-dimensional Brownian motion. The SDE is then expressed as

$$X(t) = x_0 + \int_0^t \mu(X(s))ds + \sum_{i=1}^{d} \int_0^t \sigma_i(X(s))dW^{(i)}(s) \tag{3.22}$$

for all $t \in [0, T]$.

In this section, we still assume that σ is globally Lipschitz continuous. We also assume that the drift coefficient $\mu : \mathbb{R}^m \to \mathbb{R}^m$ is a continuously differentiable (i.e. C^1) and globally one-sided Lipschitz continuous function whose derivative grows at most polynomially. That is, there exists a positive real number C such that

$$\|\mu'(x)\| \le C(1 + \|x\|^C), \tag{3.23}$$

$$\|\sigma(x) - \sigma(y)\| \le C\|x - y\|, \tag{3.24}$$

$$\langle x - y, \mu(x) - \mu(y)\rangle \le C\|x - y\|^2. \tag{3.25}$$

Note that if μ satisfies (3.23), then by the mean value theorem for derivatives, it also satisfies (3.20). As we discussed in Theorem 3.3, the numerical scheme (3.19) converges strongly to the real solution of SDE (3.22) as long as (3.23), (3.24) and (3.25) are satisfied. However, in each time step of (3.19), the zero of a nonlinear equation has to be determined, which requires more computational effort. To solve this problem, in their seminal paper [27], Hutzenthaler, Jentzen and Kloeden proposed a new explicit Euler scheme by defining $V_0 = x_0$ and

$$V_{k+1} = V_k + \frac{\mu(V_k)\Delta t}{1 + \|\mu(V_k)\|\Delta t} + \sigma(V_k)\Delta W_k, \qquad 0 \le k \le N-1. \tag{3.26}$$

This scheme is called the *tamed Euler scheme*. Note that the drift term is "tamed" by the factor $1 + \|\mu(V_k)\|\Delta t$ and thus bounded by 1. And this prevents the large excursions generated by the drift term of the numerical scheme. Since the diffusion term σ is still required to be globally Lipschitz continuous, we can expect that the numerical scheme (3.26) behaves nicely and does not have a possibility of blowing up.

After a small transformation, the numerical scheme (3.26) becomes

$$V_{k+1} = V_k + \mu(V_k)\Delta t + \sigma(V_k)\Delta W_k - (\Delta t)^2 \frac{\mu(V_k)\|\mu(V_k)\|}{1 + \|\mu(V_k)\|\Delta t}. \tag{3.27}$$

We can see that this is the Euler-Maruyama scheme added by a second-order term. As usual, we also define the continuous-time approximation of the tamed Euler scheme. It is natural to have

$$\overline{V}(t) := V_k + \frac{(t - t_k)\mu(V_k)}{1 + \|\mu(V_k)\|\Delta t} + \sigma(V_k)(W(t) - W(t_k)) \tag{3.28}$$

for all $t \in [t_k, t_{k+1}), k = 0, 1, \cdots, N-1$.

Theorem 3.4 ([28], Theorem 1.1)**.** *Suppose that the drift coefficient $\mu(x)$ is a continuously differentiable and globally one-sided Lipschitz continuous function whose derivative grows at most polynomially. Suppose also that the diffusion coefficient $\sigma(x)$ is globally Lipschitz continuous and $E[\|x_0\|^p] < \infty$ for all $p \ge 1$. Then there exists a family $C_p \in [0, \infty)$, $p \in [1, \infty)$, of real numbers such that*

$$\Big(E\Big[\sup_{t \in [0,T]} \|X(t) - \overline{V}(t)\|^p \Big] \Big)^{1/p} \le C_p \cdot N^{-1/2} \tag{3.29}$$

for all $N \in \mathbb{N}$ and all $p \in [1, \infty)$.

To prove this theorem, we need the following two lemmas.

Lemma 3.4.1 ([28], Lemma 3.9)**.** *Let V_k be given by (3.26). Then we have that*

$$\sup_{N\in\mathbb{N}}\sup_{k\in\{0,1,\cdots,N\}} E[\|V_k\|^p] < \infty \tag{3.30}$$

for all $p \in [1,\infty)$.

This lemma is very crucial in the proof of Theorem 3.4. See Lemma 3.1–3.9 for the proof.

Lemma 3.4.2 ([28], Lemma 3.10)**.** *Let V_k be given by (3.26). Then we have that*

$$\sup_{N\in\mathbb{N}}\sup_{k\in\{0,1,\cdots,N\}} E[\|\mu(V_k)\|^p] < \infty, \tag{3.31}$$

$$\sup_{N\in\mathbb{N}}\sup_{k\in\{0,1,\cdots,N\}} E[\|\sigma(V_k)\|^p] < \infty \tag{3.32}$$

for all $p \in [1,\infty)$.

Proof. First of all, by the polynomial growth property of $\mu(x)$, we have

$$\begin{aligned}\|\mu(x)-\mu(0)\| &\le C(1+\|x\|^C)\|x\| \\ &= C\|x\| + C\|x\|^{C+1} \\ &\le C(1+\|x\|^{C+1}) + C(1+\|x\|^{C+1}) \\ &\le 2C(1+\|x\|^{C+1}).\end{aligned}$$

Therefore, we have

$$\|\mu(x)\| \le (2C+\|\mu(0)\|)(1+\|x\|^{C+1}). \tag{3.33}$$

Combining (3.33) and Lemma 3.4.1, we have

$$\begin{aligned}&\sup_{N\in\mathbb{N}}\sup_{k\in\{0,1,\cdots,N\}} \|\mu(V_k)\|_{L^p(\Omega;\mathbb{R}^m)} \\ &\le (2C+\|\mu(0)\|)\Big(1+\sup_{N\in\mathbb{N}}\sup_{k\in\{0,1,\cdots,N\}}\|V_k\|^{(C+1)}_{L^{p(C+1)}(\Omega;\mathbb{R}^m)}\Big)\end{aligned}$$

$$< \infty$$

Additionally, the inequality $\|\sigma(x)\| \leq C\|x\| + \|\sigma(0)\|$ for all $x \in \mathbb{R}^m$ and again Lemma 3.4.1 show that

$$\begin{aligned}
&\sup_{N\in\mathbb{N}} \sup_{k\in\{0,1,\cdots,N\}} \|\sigma(V_k)\|_{L^p(\Omega;\mathbb{R}^m)} \\
&\leq C\Big(\sup_{N\in\mathbb{N}} \sup_{k\in\{0,1,\cdots,N\}} \|V_k\|_{L^p(\Omega;\mathbb{R}^m)}\Big) + \|\sigma(0)\| \\
&< \infty
\end{aligned}$$

for all $p \in [1, \infty)$. □

Next, we give the proof of Theorem 3.4.

Proof. We first define the shift operator

$$\eta(t) = t_i, \qquad \text{if } t_i \leq t < t_{i+1}, \quad i = 0, 1, \cdots, N-1.$$

In this notation, equation (3.28) reads as

$$\overline{V}(s) = x_0 + \int_0^s \frac{\mu(\overline{V}(\eta(u)))}{1 + T/N\|\mu(\overline{V}(\eta(u)))\|} du + \int_0^s \sigma(\overline{V}(\eta(u))) dW(u) \tag{3.34}$$

for all $s \in [0, T]$ P-a.s.. Our goal is then to estimate the quantity $E[\sup_{s\in[0,t]} \|X(s) - \overline{V}(s)\|^p]$ for $t \in [0, T]$ and $p \in [1, \infty)$. Using (3.22) and (3.34), we get

$$\begin{aligned}
X(s) - \overline{V}(s) = &\int_0^s \Big(\mu(X(u)) - \frac{\mu(\overline{V}(\eta(u)))}{1 + T/N\|\mu(\overline{V}(\eta(u)))\|}\Big) du \\
&+ \sum_{i=1}^{d} \int_0^s \Big(\sigma_i(X(u)) - \sigma_i(\overline{V}(\eta(u)))\Big) dW^{(i)}(u)
\end{aligned}$$

for all $s \in [0, T]$ P-a.s.. Itô's formula hence gives that

$$
\begin{aligned}
\|X(s) - \overline{V}(s)\|^2 = 2 \int_0^s & \langle X(u) - \overline{V}(u), \mu(X(u)) - \mu(\overline{V}(u)) \rangle du \\
&+ 2 \int_0^s \langle X(u) - \overline{V}(u), \mu(\overline{V}(u)) - \mu(\overline{V}(\eta(u))) \rangle du \\
&+ \frac{2T}{N} \int_0^s \Big\langle X(u) - \overline{V}(u), \frac{\mu(\overline{V}(\eta(u))) \|\mu(\overline{V}(\eta(u)))\|}{1 + T/N \|\mu(\overline{V}(\eta(u)))\|} \Big\rangle du \\
&+ 2 \sum_{i=1}^d \int_0^s \langle X(u) - \overline{V}(u), \sigma_i(X(u)) - \sigma_i(\overline{V}(\eta(u))) \rangle dW^{(i)}(u) \\
&+ \sum_{i=1}^d \int_0^s \|\sigma_i(X(u)) - \sigma_i(\overline{V}(\eta(u)))\|^2 du
\end{aligned}
$$

By the one-sided Lipschitz continuity of μ, the global Lipschitz continuity of σ and Cauchy-Schwarz inequality, we have

$$
\begin{aligned}
\|X(s) - \overline{V}(s)\|^2 \leq (2C + 2C^2 d) & \int_0^s \|X(u) - \overline{V}(u)\|^2 du \\
&+ 2 \int_0^s \|X(u) - \overline{V}(u)\| \cdot \|\mu(\overline{V}(u)) - \mu(\overline{V}(\eta(u)))\| du \\
&+ \frac{2T}{N} \int_0^s \|X(u) - \overline{V}(u)\| \cdot \|\mu(\overline{V}(\eta(u)))\|^2 du \\
&+ 2 \Big| \sum_{i=1}^d \int_0^s \langle X(u) - \overline{V}(u), \sigma_i(X(u)) - \sigma(\overline{V}(\eta(u))) \rangle dW^{(i)}(u) \Big| \\
&+ 2C^2 d \int_0^s \|\overline{V}(u) - \overline{V}(\eta(u))\|^2 du
\end{aligned}
$$

for all $s \in [0, T]$ P-a.s.. Therefore,

$$
\begin{aligned}
&\sup_{s \in [0,t]} \|X(s) - \overline{V}(s)\|^2 \\
&\leq 2(C + C^2 d + 1) \int_0^t \|X(s) - \overline{V}(s)\|^2 ds + \int_0^T \|\mu(\overline{V}(s)) - \mu(\overline{V}(\eta(s)))\|^2 ds \\
&\quad + \frac{T^2}{N^2} \int_0^T \|\mu(\overline{V}(\eta(s)))\|^4 ds + 2C^2 d \int_0^T \|\overline{V}(s) - \overline{V}(\eta(s))\|^2 ds
\end{aligned}
$$

$$+ 2 \sup_{s\in[0,t]} \Big| \sum_{i=1}^{d} \int_0^s \langle X(u) - \overline{V}(u), \sigma_i(X(u)) - \sigma_i(\overline{V}(\eta(s)))\rangle dW^{(i)}(u) \Big|$$

P-a.s.. for all $t \in [0, T]$. The Minkowski's inequality (A.5), Minkowski's integral inequality (A.6) and Burkholder-Davis-Gundy type inequality (A.11) yield that

$$\begin{aligned}
&\Big\| \sup_{s\in[0,t]} \|X(s) - \overline{V}(s)\|^2 \Big\|_{L^{p/2}(\Omega;\mathbb{R})} \\
&\leq 2(C + C^2 d + 1) \int_0^t \|X(s) - \overline{V}(s)\|^2_{L^p(\Omega;\mathbb{R}^m)} ds \\
&\quad + \int_0^T \|\mu(\overline{V}(s)) - \mu(\overline{V}(\eta(s)))\|^2_{L^p(\Omega;\mathbb{R}^m)} ds \\
&\quad + \frac{T^2}{N^2} \int_0^T \|\mu(\overline{V}(\eta(s)))\|^4_{L^{2p}(\Omega;\mathbb{R}^m)} ds \\
&\quad + 2C^2 d \int_0^T \|\overline{V}(s) - \overline{V}(\eta(s))\|^2_{L^p(\Omega;\mathbb{R}^m)} ds \\
&\quad + p\Big(\sum_{i=1}^{d} \int_0^t \|\langle X(s) - \overline{V}(s), \sigma_i(X(s)) - \sigma_i(\overline{V}(\eta(s)))\rangle\|^2_{L^{p/2}(\Omega;\mathbb{R})} ds \Big)^{1/2}
\end{aligned} \tag{3.35}$$

for all $t \in [0, T]$ and all $p \in [4, \infty)$. Next the Cauchy-Schwarz inequality, the Hölder inequality and $ab \leq \frac{a^2}{2} + \frac{b^2}{2}$ imply that

$$\begin{aligned}
&p\Big(\sum_{i=1}^{d} \int_0^t \|\langle X(s) - \overline{V}(s), \sigma_i(X(s)) - \sigma_i(\overline{V}(\eta(s)))\rangle\|^2_{L^{p/2}(\Omega;\mathbb{R})} ds \Big)^{1/2} \\
&\leq p\Big(\sum_{i=1}^{d} \int_0^t \|X(s) - \overline{V}(s)\|^2_{L^p(\Omega;\mathbb{R}^m)} \|\sigma_i(X(s)) - \sigma_i(\overline{V}(\eta(s)))\|^2_{L^p(\Omega;\mathbb{R}^m)} ds \Big)^{1/2} \\
&\leq p\Big(\sup_{s\in[0,t]} \|X(s) - \overline{V}(s)\|_{L^p(\Omega;\mathbb{R}^m)} \Big) \Big(C^2 d \int_0^t \|X(s) - \overline{V}(\eta(s))\|^2_{L^p(\Omega;\mathbb{R}^m)} ds \Big)^{1/2} \\
&\leq \frac{1}{2} \sup_{s\in[0,t]} \|X(s) - \overline{V}(s)\|^2_{L^p(\Omega;\mathbb{R})} + \frac{p^2 C^2 d}{2} \int_0^t \|X(s) - \overline{V}(\eta(s))\|^2_{L^p(\Omega;\mathbb{R}^m)} ds \\
&\leq \frac{1}{2} \Big\| \sup_{s\in[0,t]} \|X(s) - \overline{V}(s)\| \Big\|^2_{L^p(\Omega;\mathbb{R})} + \frac{p^2 C^2 d}{2} \int_0^t \|X(s) - \overline{V}(\eta(s))\|^2_{L^p(\Omega;\mathbb{R}^m)} ds
\end{aligned} \tag{3.36}$$

for all $t \in [0,T]$ and all $p \in [4,\infty)$. Inserting inequality (3.36) into (3.35) and applying the estimate $(a+b)^2 \leq 2a^2 + 2b^2$ then yields that

$$
\begin{aligned}
&\Big\| \sup_{s\in[0,t]} \|X(s) - \overline{V}(s)\| \Big\|^2_{L^p(\Omega;\mathbb{R})} \\
&= \Big\| \sup_{s\in[0,t]} \|X(s) - \overline{V}(s)\|^2 \Big\|_{L^{p/2}(\Omega;\mathbb{R})} \\
&\leq 2\Big(C + C^2 d + 1 + \frac{p^2 C^2 d}{2}\Big) \int_0^t \|X(s) - \overline{V}(s)\|^2_{L^p(\Omega;\mathbb{R}^m)} ds \\
&\quad + \int_0^T \|\mu(\overline{V}(s)) - \mu(\overline{V}(\eta(s)))\|^2_{L^p(\Omega;\mathbb{R}^m)} ds \\
&\quad + \frac{T^2}{N^2} \int_0^T \|\mu(\overline{V}(\eta(s)))\|^4_{L^{2p}(\Omega;\mathbb{R}^m)} ds \\
&\quad + (2C^2 d + p^2 C^2 d) \int_0^T \|\overline{V}(s) - \overline{V}(\eta(s))\|^2_{L^p(\Omega;\mathbb{R}^m)} ds \\
&\quad + \frac{1}{2} \Big\| \sup_{s\in[0,t]} \|X(s) - \overline{V}(s)\| \Big\|^2_{L^p(\Omega;\mathbb{R})}
\end{aligned}
$$

and therefore, we obtain that

$$
\begin{aligned}
&\frac{1}{2} \Big\| \sup_{s\in[0,t]} \|X(s) - \overline{V}(s)\| \Big\|^2_{L^p(\Omega;\mathbb{R})} \\
&\leq 2\Big(C + C^2 d + 1 + p^2 C^2 d\Big) \int_0^t \|X(s) - \overline{V}(s)\|^2_{L^p(\Omega;\mathbb{R}^m)} ds \\
&\quad + \int_0^T \|\mu(\overline{V}(s)) - \mu(\overline{V}(\eta(s)))\|^2_{L^p(\Omega;\mathbb{R}^m)} ds \\
&\quad + \frac{T^2}{N^2} \int_0^T \|\mu(\overline{V}(\eta(s)))\|^4_{L^{2p}(\Omega;\mathbb{R}^m)} ds \\
&\quad + (2C^2 d + p^2 C^2 d) \int_0^T \|\overline{V}(s) - \overline{V}(\eta(s))\|^2_{L^p(\Omega;\mathbb{R}^m)} ds
\end{aligned}
$$

for all $t \in [0,T]$ and all $p \in [4,\infty)$. By Gronwall's lemma, we have

$$
\begin{aligned}
&\Big\| \sup_{t\in[0,T]} \|X(t) - \overline{V}(t)\| \Big\|^2_{L^p(\Omega;\mathbb{R})} \\
&\leq 2e^{4T(p^2 C^2 d + C + 1)} \Big(\int_0^T \|\mu(\overline{V}(s)) - \mu(\overline{V}(\eta(s)))\|^2_{L^p(\Omega;\mathbb{R}^m)} ds
\end{aligned}
$$

$$
\begin{aligned}
&+ \frac{T^2}{N^2}\int_0^T \|\mu(\overline{V}(\eta(s)))\|^4_{L^{2p}(\Omega;\mathbb{R}^m)}ds \\
&+ 2p^2C^2d\int_0^T \|\overline{V}(s)-\overline{V}(\eta(s))\|^2_{L^p(\Omega;\mathbb{R}^m)}ds\Big)
\end{aligned}
$$

and hence, the inequality $\sqrt{a+b+c} \leq \sqrt{a}+\sqrt{b}+\sqrt{c}$ for all $a,b,c \in [0,\infty)$ gives that

$$
\begin{aligned}
&\Big\|\sup_{t\in[0,T]}\|X(t)-\overline{V}(t)\|\Big\|^2_{L^p(\Omega;\mathbb{R})} \\
&\leq \sqrt{2T}e^{2T(p^2C^2d+C+1)}\Big(\sup_{t\in[0,T]}\|\mu(\overline{V}(t)))-\mu(\overline{V}(\eta(s)))\|_{L^p(\Omega;\mathbb{R}^m)} \\
&\quad + \frac{T}{N}\Big[\sup_{k\in\{0,1,\cdots,N\}}\|\mu(V_k)\|^2_{L^{2p}(\Omega;\mathbb{R}^m)}\Big] \\
&\quad + pC\sqrt{2d}\Big[\sup_{t\in[0,T]}\|\overline{V}(t)-\overline{V}(\eta(s))\|_{L^p(\Omega;\mathbb{R}^m)}\Big]\Big)
\end{aligned} \tag{3.37}
$$

for all $p \in [4,\infty)$. Additionally, the Burkholder-Davis-Dundy type inequality (A.11) shows that

$$
\begin{aligned}
&\sup_{t\in[0,T]}\|\overline{V}(t)-\overline{V}(\eta(s))\|_{L^p(\Omega;\mathbb{R}^m)} \\
&\leq \frac{T}{N}\Big(\sup_{t\in[0,T]}\Big\|\frac{\mu(\overline{V}(\eta(s)))}{1+T/N\|\mu(\overline{V}(\eta(s)))\|}\Big\|\Big) \\
&\quad + \sup_{t\in[0,T]}\Big\|\int_{\eta(t)}^t \sigma(\overline{V}(\eta(t)))dW(s)\Big\|_{L^p(\Omega;\mathbb{R}^m)} \\
&\leq \frac{T}{\sqrt{N}}\Big(\sup_{k\in\{0,1,\cdots,N\}}\|\mu(V_k)\|_{L^p(\Omega;\mathbb{R}^m)}\Big) \\
&\quad + \frac{p\sqrt{Td}}{\sqrt{N}}\Big(\sup_{i\in\{1,2,\cdots,d\}}\sup_{k\in\{0,1,\cdots,N\}}\|\sigma_i(V_k)\|_{L^p(\Omega;\mathbb{R}^m)}\Big)
\end{aligned}
$$

for all $p \in [2,\infty)$. Lemma 3.4.2 hence implies that

$$
\sup_{N\in\mathbb{N}}\Big(\sqrt{N}\Big[\sup_{t\in[0,T]}\|\overline{V}(t)-\overline{V}(\eta(s))\|_{L^p(\Omega;\mathbb{R}^m)}\Big]\Big) < \infty \tag{3.38}
$$

for all $p \in [1, \infty)$. In particular, we obtain that

$$\sup_{N \in \mathbb{N}} \sup_{t \in [0,T]} \|\overline{V}(t)\|_{L^p(\Omega;\mathbb{R}^m)} < \infty \tag{3.39}$$

for all $p \in [1, \infty)$ due to Lemma 3.4.1. Moreover, the estimate

$$\|\mu(x) - \mu(y)\| \leq C(1 + \|x\|^C + \|y\|^C)\|x - y\|, \quad x, y \in \mathbb{R}^m$$

gives that

$$\begin{aligned} &\sup_{t \in [0,T]} \|\mu(\overline{V}(t)) - \mu(\overline{V}(\eta(s)))\|_{L^p(\Omega;\mathbb{R}^m)} \\ &\leq C\Big(1 + 2 \sup_{t \in [0,T]} \|\overline{V}(t)\|^C_{L^{2pC}(\Omega;\mathbb{R}^m)}\Big)\Big(\sup_{t \in [0,T]} \|\overline{V}(t) - \overline{V}(\eta(s))\|_{L^{2p}(\Omega;\mathbb{R}^m)}\Big) \end{aligned} \tag{3.40}$$

for all $p \in [1, \infty)$. Inequalities (3.38) and (3.39) hence show that

$$\sup_{N \in \mathbb{N}} \Big(\sqrt{N}\Big[\sup_{t \in [0,T]} \|\mu(\overline{V}(t)) - \mu(\overline{V}(\eta(t)))\|_{L^p(\Omega;\mathbb{R}^m)}\Big]\Big) < \infty \tag{3.41}$$

for all $p \in [1, \infty)$. Combining (3.37), (3.38), (3.41) and Lemma 3.4.2 finally shows (3.29). □

Numerical Experiments

The example we use for our numerical experiment in this section is the 1-dimensional stochastic Ginzburg-Landau equation,

$$dX(t) = (X(t) - X^3(t))dt + X(t)dW(t), \quad X(0) = 1.$$

Here, $\mu(x) = x - x^3$, $\sigma(x) = x$ and $t \in [0, 1]$. Clearly, $\mu(x)$ satisfies a one-sided Lipschitz condition and grows superlinearly. We use 5 different time steps: $\Delta t = 2^{-12}, 2^{-11}, 2^{-10}, 2^{-9}, 2^{-8}$ and 1000 realizations for each discretisation. The following

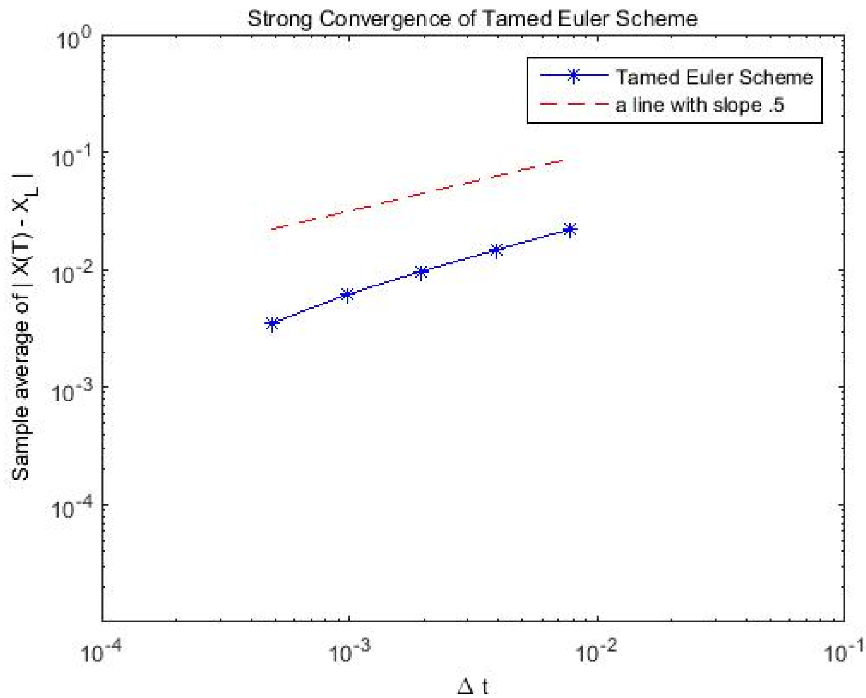

Figure 3.1: Log-log plot of the strong error from the tamed Euler approximation versus the time step Δt with the drift coefficients superlinearly growing.

figure is the loglog plot of the experimental error with respect to the 5 different time steps. We can see that the numerical scheme converges strongly with order $\frac{1}{2}$.

3.3 Euler Approximations with Superlinearly Growing Diffusion Coefficients

The tamed Euler scheme we introduced in section 3.2 is a significant progress in the computation of numerical solutions of stochastic differential equations. It is an explicit numerical scheme and only assumes that the drift coefficient is one-sided Lipschitz and its derivative grows at most polynomially. However, it still requires the global Lipschitz continuity of the diffusion coefficient. In Sabanis [55], the author introduces a new explicit Euler-type numerical scheme to compute the numerical solutions of SDEs whose diffusion coefficients can be superlinearly growing. We introduce this new scheme in this section based on [55]. For the Milstein-type numerical scheme to address this issue, see e.g. [37].

In this section, we consider the following stochastic differential equation:

$$\begin{cases} dX(t) = \mu(t, X(t))dt + \sigma(t, X(t))dW(t), \quad t \in (0, T] \\ X(0) = x_0, \end{cases} \tag{3.42}$$

where $X(t) \in \mathbb{R}^m$ for all $t \in [0, T]$, $W(t)$ is a d-dimensional Brownian motion starting at 0, $\mu : [0, T] \times \mathbb{R}^m \to \mathbb{R}^m$ and $\sigma : [0, T] \times \mathbb{R}^m \to \mathbb{R}^{m \times d}$. We also assume that x_0 is $\mathcal{F}_0$-measurable, almost surely finite and independent of $(W(t), 0 \leq t \leq T)$. Let $p_0, p_1 \in [2, \infty)$ be positive constants. We consider the following conditions.

A-1. $E[\|x_0\|^{p_0}] < \infty$.

A-2. $\mu(t, x)$ and $\sigma(t, x)$ are locally Lipschitz continuous in x for any $t \in [0, T]$ (see (2.6)).

A-3. There exist positive constants l and L such that, for any $t \in [0, T]$,

$$2\langle x - y, \mu(t, x) - \mu(t, y)\rangle + (p_1 - 1)\|\sigma(t, x) - \sigma(t, y)\|^2 \leq L\|x - y\|^2$$

and

$$\|\mu(t, x) - \mu(t, y)\| \leq L(1 + \|x\|^l + \|y\|^l)\|x - y\|$$

for all $x, y \in \mathbb{R}^m$.

A-4. There exists a positive constant K such that,

$$2x^T\mu(t, x) + (p_0 - 1)\|\sigma(t, x)\|^2 \leq K(1 + \|x\|^2)$$

for any $t \in [0, T]$ and $x \in \mathbb{R}^d$.

Note that, due to A-2, $\mu(t,x)$ and $\sigma(t,x)$ are locally bounded in x for any $t \in [0,T]$. That is, for every $R \geq 0$, there exists a positive constant N_R such that

$$\sup_{\|x\|\leq R} \|\mu(t,x)\| \leq N_R$$

$$\sup_{\|x\|\leq R} \|\sigma(t,x)\| \leq N_R$$

for any $t \in [0,T]$. We also observe that if A-2, A-3, and A-4 hold, then

$$\|\mu(t,x)\| \leq \|\mu(t,x) - \mu(t,0)\| + \|\mu(t,0)\| \leq L(1+\|x\|^l)\|x\| + N_0 \leq C(1+\|x\|^{l+1}) \tag{3.43}$$

for any $t \in [0,T]$ and $x \in \mathbb{R}^m$, where C is a positive constant. Similarly, by A-4, we obtain

$$\|\sigma(t,x)\|^2 \leq K(1+\|x\|^2) + 2C(1+\|x\|^{l+1})\|x\| \leq C(1+\|x\|^{l+2}) \tag{3.44}$$

for any $t \in [0,T]$ and $x \in \mathbb{R}^m$. A-1, A-2 and A-4, by Theorem 2.3, guarantee that there exists a unique solution of equation (3.42).

We now consider the numerical scheme. To be consistent with the two numerical schemes we are going to introduce in this section, we will use the following unified notation. For every $N \geq 1$, the following numerical scheme is defined

$$dX_N(t) = \mu_N(t, X_N(\kappa_N(t)))dt + \sigma_N(t, X_N(\kappa_N(t)))dW(t), \qquad \forall\, t \in [0,T], \tag{3.45}$$

with the same initial value x_0 as equation (3.42), where $\mu_N(t,x)$ and $\sigma_N(t,x)$ are $\mathcal{B}(\mathbb{R}_+) \otimes \mathcal{B}(\mathbb{R}^d)$-measurable functions which take values in $\mathbb{R}^m$ and $\mathbb{R}^{m\times d}$ respectively and $\kappa_N(t) := [Nt]/N$. Note that the function $\kappa_N(t)$ jumps with size $1/N$, while the shift operator function $\eta(t)$ we defined in the last section jumps with size T/N. In other words, there are $N+1$ nodes within the interval $[0,1]$ in the numerical scheme (3.45), while there are $N+1$ nodes within the interval $[0,T]$ in the numerical schemes we introduced in the previous sections. We can see that, by defining the numerical scheme as in (3.45), we already have a continuous-time approximation of the solution of equation (3.42).

The following condition we assume is very important for our arguments.

B. There exists an $\alpha \in (0, 1/2]$ and a constant C such that, for every $N \geq 1$,

$$\|\mu_N(t,x)\| \leq \min(CN^\alpha, \|\mu(t,x)\|) \quad \text{and} \quad \|\sigma_N(t,x)\| \leq \min(CN^\alpha, \|\sigma(t,x)\|) \tag{3.46}$$

for any $t \in [0,T]$ and $x \in \mathbb{R}^m$.

Let $\alpha \in (0, 1/2]$, we now define

- **Model 1:**

$$\mu_N(t,x) := \frac{1}{1 + N^{-\alpha}\|\mu(t,x)\| + N^{-\alpha}\|\sigma(t,x)\|}\mu(t,x) \tag{3.47}$$

and

$$\sigma_N(t,x) := \frac{1}{1 + N^{-\alpha}\|\mu(t,x)\| + N^{-\alpha}\|\sigma(t,x)\|}\sigma(t,x) \tag{3.48}$$

for any $t \in [0,T]$, $x \in \mathbb{R}^m$ and $N \geq 1$.

- **Model 2:**

$$\mu_N(t,x) := \frac{1}{1 + N^{-\alpha}\|x\|^{3l/2+2}}\mu(t,x) \tag{3.49}$$

and

$$\sigma_N(t,x) := \frac{1}{1 + N^{-\alpha}\|x\|^{3l/2+2}}\sigma(t,x) \tag{3.50}$$

for any $t \in [0,T]$, $x \in \mathbb{R}^m$ and $N \geq 1$.

It can be verified easily that both Model 1 and Model 2 satisfy the condition B for any $t \in [0,T]$ and $x \in \mathbb{R}^d$. Let p_0^* be the largest even number which is smaller than or equal to p_0. In order to ease the notation, we say that the $\mathfrak{p}$-condition is satisfied if one of the following two cases hold true:

- **(Model 1)** The coefficients μ_N and σ_N are given by equations (3.47) and (3.48) with $\alpha = 1/2$, $p < p_1$ and either $p \leq \frac{p_0}{5l/2+3}$ if $l \in (0,2) \cap (0, \frac{p_0}{4} - 1]$ or $p \leq \frac{p_0^*}{2(l+1)}$ if $l \in (0, \frac{p_0^*}{4} - 1]$ and $m = d = 1$.

- **(Model 2)** The coefficients μ_N and σ_N are given by equations (3.49) and (3.50) with $\alpha = 1/2$, $p < p_1$, $p \leq \frac{p_0}{5l/2+3}$ and $l \leq p_0 - 2$.

We then can recover the optimal rate of strong convergence for Euler approximations.

Theorem 3.5 ([55], Theorem 2)**.** *Suppose A-1–A-4 and the $\mathfrak{p}$-condition hold, then the numerical scheme (3.45) converges to the true solution of SDE (3.42) in L^p-sense with order $1/2$, i.e.*

$$\sup_{0 \leq t \leq T} E\Big[\|X(t) - X_N(t)\|^p\Big] \leq CN^{-p/2} \tag{3.51}$$

where C is a constant independent of N.

The uniform L^p convergence for smaller values of p is given below.

Theorem 3.6 ([55], Theorem 3)**.** *Suppose A-1–A-4 and the $\mathfrak{p}$-condition hold, then the numerical scheme (3.45) converges to the true solution of SDE (3.42) in uniform L^q-sense with order $1/2$, i.e.*

$$E\Big[\sup_{0 \leq t \leq T} \|X(t) - X_N(t)\|^q\Big] \leq CN^{-q/2} \tag{3.52}$$

where C is a constant independent of N, for all $q < p$.

To prove the above two theorems, we need the following five estimates.

Lemma 3.6.1 ([55], Lemma 1)**.** *Consider the numerical scheme (3.45) and let A-1–A-4 and B hold and*

$$\sup_{N \geq 1} \sup_{0 \leq t \leq T} E[\|X_N(t)\|^q] < \infty, \tag{3.53}$$

for some $q \geq 2$, then for any $p \leq \frac{2}{l+2}q$ and $l \in (0, q-2]$,

$$\sup_{0 \leq t \leq T} E\|X_N(t) - X_N(\kappa_N(t))\|^p \leq CN^{-p/2}, \tag{3.54}$$

where C is a positive constant independent of N.

As in Section 3.1 and Section 3.2, the following three bound estimates of the numerical solution is crucial in our proof of strong convergence.

Lemma 3.6.2 ([55], Lemma 7)**.** *Consider the numerical scheme (3.45) with coefficients given by (3.47) and (3.48) with $\alpha = 1/2$. Suppose that A-1–A-4 with $l \in (0, 2)$ hold, then for some $C := C(p, T, K, E[\|x_0\|^p])$,*

$$\sup_{N \geq 1} \sup_{0 \leq t \leq T} E\Big[\|X_N(t)\|^p\Big] < C \tag{3.55}$$

for every $p \leq p_0$.

Lemma 3.6.3 ([55], Lemma 5)**.** *Consider the numerical scheme (3.45) with coefficients given by (3.49) and (3.50) with $\alpha = 1/2$. Let also A-1–A-4 hold true. Then, for every $p \leq p_0$*

$$\sup_{N \geq 1} \sup_{0 \leq t \leq T} E\Big[\|X_N(t)\|^p\Big] < C \tag{3.56}$$

where the constant $C := C(p, T, K, E[\|x_0\|^p])$.

Lemma 3.6.4 ([55], Lemma 9)**.** *Consider the numerical scheme (3.45) with coefficients given by (3.47) and (3.48) with $\alpha = 1/2$ when $m = d = 1$. Suppose that A-1–A-4 hold, then for some $C := C(p, T, K, E[\|x_0\|^p])$,*

$$\sup_{N \geq 1} \sup_{0 \leq t \leq T} E\Big[\|X_N(t)\|^p\Big] < C \tag{3.57}$$

for every $p \leq p_0^*$.

Lemma 3.6.5 ([55], Lemma 10)**.** *Consider the numerical scheme (3.45). Suppose A-1–A-4 and the* $\mathfrak{p}$*-condition hold. Then,*

$$E\Big[\int_0^T \|\mu(s, X_N(\kappa_N(s))) - \mu_N(s, X_N(\kappa_N(s)))\|^p ds\Big] \leq CN^{-\alpha p} \tag{3.58}$$

and

$$E\Big[\int_0^T \|\sigma(s, X_N(\kappa_N(s))) - \sigma_N(s, X_N(\kappa_N(s)))\|^p ds\Big] \leq CN^{-\alpha p} \tag{3.59}$$

We refer to [55] for the proof of the above four lemmas. We now give the proof of the two main theorems of this section.

Proof of Theorem 3.5. We first consider, for every $N \geq 1$ and $t \in [0, T]$, the difference processes,

$$\chi_N(t) := X(t) - X_N(t),$$

$$\beta_N(t) := \mu(t, X(t)) - \mu_N(t, X_N(\kappa_N(t))),$$

$$\alpha_N(t) := \sigma(t, X(t)) - \sigma_N(t, X_N(\kappa_N(t))).$$

Therefore, by (3.42) and (3.45), we have

$$d\chi_N(t) = \beta_N(t)dt + \alpha_N(t)dW(t).$$

Note that $(|x|^p)' = \Big((x^2)^{p/2}\Big)' = px|x|^{p-2}$ and $(|x|^p)'' = p(p-1)|x|^{p-2}$. By Itô's formula,

$$\begin{aligned} d\|\chi_N(t)\|^p = p\|\chi_N(t)\|^{p-2}\Big(\langle\chi_N(t), \beta_N(t)\rangle + \frac{p-1}{2}\|\alpha_N(t)\|^2\Big)dt \\ + p\|\chi_N(t)\|^{p-2}\langle\chi_N(t), \alpha_N(t)dW(t)\rangle. \end{aligned} \tag{3.60}$$

Therefore,

$$\|\chi_N(t)\|^p \leq \frac{p}{2}\int_0^t \|\chi_N(s)\|^{p-2}\Big(2\langle\chi_N(s), \beta_N(s)\rangle + (p-1)\|\alpha_N(s)\|^2\Big)ds$$

$$+ p\int_0^t \|\chi_N(s)\|^{p-2}\langle \chi_N(s), \alpha_N(s)dW(s)\rangle. \tag{3.61}$$

Note that in the above equality, for any $\epsilon > 0$,

$$\begin{aligned}
&2\langle \chi_N(s), \beta_N(s)\rangle + (p-1)\|\alpha_N(s)\|^2 \\
&= 2\langle X(s) - X_N(s), \mu(s, X(s)) - \mu(s, X_N(s))\rangle \\
&\quad + 2\langle X(s) - X_N(s), \mu(s, X_N(s)) - \mu(s, X_N(\kappa_N(s)))\rangle \\
&\quad + 2\langle X(s) - X_N(s), \mu(s, X_N(\kappa_N(s))) - \mu_N(s, X_N(\kappa_N(s)))\rangle \\
&\quad + (p-1)\Big(\|\sigma(s, X(s)) - \sigma(s, X_N(s))\|^2 \\
&\quad + \|\sigma(s, X_N(s)) - \sigma(s, X_N(\kappa_N(s)))\|^2 \\
&\quad + \|\sigma(s, X_N(\kappa_N(s))) - \sigma_N(s, X_N(\kappa_N(s)))\|^2 \\
&\quad + 2\big(\sqrt{\epsilon/2}\|\sigma(s, X(s)) - \sigma(s, X_N(s))\|\big)\big(\sqrt{2/\epsilon}\|\sigma(s, X_N(s)) - \sigma(s, X_N(\kappa_N(s)))\|\big) \\
&\quad + 2\big(\sqrt{\epsilon/2}\|\sigma(s, X(s)) - \sigma(s, X_N(s))\|\big)\big(\sqrt{2/\epsilon}\|\sigma(s, X_N(\kappa_N(s))) - \sigma_N(s, X_N(\kappa_N(s)))\|\big) \\
&\quad + 2\|\sigma(s, X_N(s)) - \sigma(s, X_N(\kappa_N(s)))\|\|\sigma(s, X_N(\kappa_N(s))) - \sigma_N(s, X_N(\kappa_N(s)))\|\Big) \\
&\leq 2\langle X(s) - X_N(s), \mu(s, X(s)) - \mu(s, X_N(s))\rangle \\
&\quad + 2\langle X(s) - X_N(s), \mu(s, X_N(s)) - \mu(s, X_N(\kappa_N(s)))\rangle \\
&\quad + 2\langle X(s) - X_N(s), \mu(s, X_N(\kappa_N(s))) - \mu_N(s, X_N(\kappa_N(s)))\rangle \\
&\quad + (p-1)\Big((1+\epsilon)\|\sigma(s, X(s)) - \sigma(s, X_N(s))\|^2 \\
&\quad + 2(1+\frac{1}{\epsilon})\|\sigma(s, X_N(s)) - \sigma(s, X_N(\kappa_N(s)))\|^2 \\
&\quad + 2(1+\frac{1}{\epsilon})\|\sigma(s, X_N(\kappa_N(s))) - \sigma_N(s, X_N(\kappa_N(s)))\|^2\Big).
\end{aligned} \tag{3.62}$$

Due to A-3, we have that

$$\begin{aligned}
(p_1 - 1)\|\sigma(t,x) - \sigma(t,y)\|^2 &\leq L\|x-y\|^2 - 2\langle x-y, \mu(t,x) - \mu(t,y)\rangle \\
&\leq C(1 + \|x\|^l + \|y\|^l)\|x-y\|^2.
\end{aligned} \tag{3.63}$$

If ϵ is small enough, one can get that $(1+\epsilon)(p-1) \le p_1 - 1$ based on the condition that $p < p_1$. By A-3, (3.63) and Cauchy-Schwarz inequality, we have

$$\begin{aligned}
&2\langle \chi_N(s), \beta_N(s)\rangle + (p-1)\|\alpha_N(s)\|^2 \\
&\le C\|\chi_N(s)\|^2 + C(1 + \|X_N(s)\|^{2l} + \|X_N(\kappa_N(s))\|^{2l})\|X_N(s) - X_N(\kappa_N(s))\|^2 \\
&\quad + \|\mu(s, X_N(\kappa_N(s))) - \mu_N(s, X_N(\kappa_N(s)))\|^2 \\
&\quad + C\|\sigma(s, X_N(\kappa_N(s))) - \sigma_N(s, X_N(\kappa_N(s)))\|^2. \qquad (3.64)
\end{aligned}$$

Due to (3.46), Hölder's inequality, (3.44), Lemma 3.6.2, Lemma 3.6.3 and that $2p < p_0$ and $(l+2)p < p_0$ (or $2p < p_0^*$ and $(l+2)p < p_0^*$ when we consider the Model 1 with $m = d = 1$) due to the $\mathfrak{p}$-condition, we have

$$\begin{aligned}
&E\Big[\int_0^T \|\chi_N(s)\|^{2(p-2)}\|\alpha_N^T(s)\chi_N(s)\|^2 ds\Big] \\
&\le 4E\Big[\int_0^T \|\chi_N(s)\|^{2(p-1)}\|\sigma(s, X(s))\|^2 ds\Big] \\
&\le 4\int_0^T \Big(E\Big[\|\chi_N(s)\|^{2p}\Big]\Big)^{(p-1)/p}\Big(E\Big[\|\sigma(s, X(s))\|^{2p}\Big]\Big)^{1/p} ds \\
&\le C\int_0^T \Big(E\Big[\|X(s)\|^{2p} + \|X_N(s)\|^{2p}\Big]\Big)^{(p-1)/p}\Big(E\Big[1 + \|X(s)\|^{(l+2)p}\Big]\Big)^{1/p} ds \\
&\le C. \qquad (3.65)
\end{aligned}$$

Therefore we get

$$E\Big[\int_0^T \|\chi_N(t)\|^{p-2}\langle \chi_N(t), \alpha_N(t)dW(t)\rangle\Big] = 0.$$

Note also that

$$\frac{1}{p/(p-2)} + \frac{1}{p/2} = 1.$$

Therefore, taking expectation on both sides of (3.61) and using (3.62), (3.64) and Young's inequality, we can get

$$
\begin{aligned}
& E[\|\chi_N(t)\|^p] \\
& \leq CE\Big[\int_0^t \big\{\|\chi_N(s)\|^p + (1+\|X_N(s)\|^{2l} + \|X_N(\kappa_N(s))\|^{2l})^{p/2}\|X_N(s) - X_N(\kappa_N(s))\|^p \\
& \quad + \|\mu(s, X_N(\kappa_N(s))) - \mu_N(s, X_N(\kappa_N(s)))\|^p \\
& \quad + \|\sigma(s, X(\kappa_N(s))) - \sigma_N(s, X_N(\kappa_N(s)))\|^p\big\}ds\Big]. \qquad (3.66)
\end{aligned}
$$

Moreover,

$$
\begin{aligned}
\mathcal{E}(t) &:= E\Big[\int_0^t C(1+\|X_N(s)\|^{lp} + \|X_N(\kappa_N(s))\|^{lp})\|X_N(s) - X_N(\kappa_N(s))\|^p ds\Big] \\
&\leq C\int_0^t \Big(\sqrt{E[\|X_N(s) - X_N(\kappa_N(s))\|^{2p}]}\Big)ds
\end{aligned}
$$

due to Hölder's inequality, (3.56), (3.57) and the fact that $2lp < p_0$ (or $2lp < p_0^*$ when we consider the Model 1 with $m = d = 1$) due to the $\mathfrak{p}$-condition. By (3.54), we have

$$
\sup_{0\leq t\leq T} \mathcal{E}(t) \leq CN^{-p/2}. \qquad (3.67)
$$

By (3.66), (3.67), Lemma 3.6.1 (taking $\alpha = 1/2$), we finally have

$$
\sup_{0\leq t\leq T} E[\|\chi_N(t)\|^p] \leq CN^{-p/2}.
$$

□

To prove Theorem 3.6, we need another lemma.

Lemma 3.6.6 ([55], Lemma 11). *Let $T \in [0,\infty)$ and let $f := \{f_t\}_{t\in[0,T]}$ and $g := \{g_t\}_{t\in[0,T]}$ be non-negative continuous $\mathcal{F}_t$-adapted processes such that, for any constant $c > 0$,*

$$
E[f_\tau \mathbb{1}_{\{g_0\leq c\}}] \leq E[g_\tau \mathbb{1}_{\{g_0\leq c\}}]
$$

for any stopping time $\tau \leq T$. Then, for any stopping time $\tau \leq T$ and $\gamma \in (0,1)$,

$$E\Big[\sup_{t\leq\tau} f_t^\gamma\Big] \leq \frac{2-\gamma}{1-\gamma} E\Big[\sup_{t\leq\tau} g_t^\gamma\Big].$$

Proof of Theorem 3.6. Let p satisfy the $\mathfrak{p}$-condition and let χ_N, β_N and α_N be defined as in the proof of Theorem 3.5. Define $\phi(t) : [0,T] \to \mathbb{R}$ by

$$\phi(t) := \exp(-(L+2)t).$$

Then, by Itô's formula,

$$\begin{aligned}
&d(\phi(t)\|\chi_N\|^2)^{p/2} \\
&= \frac{p}{2}\phi(t)^{p/2}\|\chi_N(t)\|^{p-2}(2\chi_N(t)d\chi_N(t) + (p-1)\|\alpha_N(t)\|^2 dt) - \frac{p(L+2)}{2}\phi(t)^{p/2}\|\chi_N(t)\|^p dt \\
&= \frac{p}{2}\phi(t)^{p/2}\|\chi_N(t)\|^{p-2}(2\chi_N(t)\beta_N(s) + (p-1)\|\alpha_N(t)\|^2)dt - \frac{p(L+2)}{2}\phi(t)^{p/2}\|\chi_N(t)\|^p dt \\
&\quad + p\phi(t)^{p/2}\|\chi_N(t)\|^{p-2}\chi_N(s)\alpha_N(s)dW(t).
\end{aligned}$$

By (3.64), we have

$$\begin{aligned}
&(\phi(t)\|\chi_N\|^2)^{p/2} \\
&\leq \int_0^t \Big[\frac{p}{2}\phi(t)^{p/2}\|\chi_N(t)\|^{p-2}\Big((L+2)\|\chi_N(t)\|^2 + \zeta_N(t)\Big) - \frac{p(L+2)}{2}\phi(t)^{p/2}\|\chi_N(t)\|^p\Big]dt \\
&\quad + \int_0^t p\phi(t)^{p/2}\|\chi_N(t)\|^{p-2}\chi_N(s)\alpha_N(s)dW(t) \\
&= \int_0^t \Big(\frac{p}{2}\phi(t)^{p/2}\|\chi_N(t)\|^{p-2}\zeta_N(t)\Big)dt + \int_0^t p\phi(t)^{p/2}\|\chi_N(t)\|^{p-2}\chi_N(s)\alpha_N(s)dW(t)
\end{aligned} \tag{3.68}$$

where

$$\begin{aligned}
\zeta_N(t) := C\Big(&(1+\|X_N(s)\|^{2l} + \|X_N(\kappa_N(s))\|^{2l})\|X_N(s) - X_N(\kappa_N(s))\|^2 \\
&+ \|\mu(s, X_N(\kappa_N(s))) - \mu_N(s, X_N(\kappa_N(s)))\|^2
\end{aligned}$$

$$+ \|\sigma(s, X_N(\kappa_N(s))) - \sigma_N(s, X_N(\kappa_N(s)))\|^2\Big),$$

where $C > 0$ is independent of N. Since the equality (3.65) holds, it is immediately that

$$E\int_0^t p\phi(t)^{p/2}\|\chi_N(t)\|^{p-2}\chi_N(s)\alpha_N(s)dW(t) = 0.$$

Note also that $\phi(t)^{p/2} \le \phi(t)^{(p-2)/2}$. Then by (3.68), for any stopping time $\tau \le T$,

$$E[(\phi(\tau)\|\chi_N(\tau)\|)^{p/2}] \le \frac{p}{2}E\Big[\int_0^\tau \Big(\phi(t)\|\chi_N(t)\|^2\Big)^{(p-2)/2}\zeta_N(t)dt\Big].$$

Therefore, by Lemma 3.6.5,

$$E\Big[\sup_{t\le T}(\phi(t)\|\chi_N(t)\|^2)^{p\gamma/2}\Big] \le CE\Big[\Big(\int_0^T (\phi(t)\|\chi_N(t)\|^2)^{(p-2)/2}\zeta_N(t)dt\Big)^\gamma\Big]$$

for any $\gamma \in (0, 1)$. Then, for $p > 2$, by Young's inequality ($\frac{1}{p/(p-2)} + \frac{1}{p} = 1$),

$$E\Big[\sup_{t\le T}(\phi(t)\|\chi_N(t)\|^2)^{p\gamma/2}\Big] \le \frac{1}{2}E\Big[\sup_{t\le T}(\phi(t)\|\chi_N(t)\|^2)^{p\gamma/2}\Big] + CE\Big[\Big(\int_0^T \zeta_N(t)dt\Big)^{p\gamma/2}\Big].$$

It implies that

$$\begin{aligned} E\Big[\sup_{t\le T}(\phi(t)\|\chi_N(t)\|^2)^{p\gamma/2}\Big] &\le CE\Big[\Big(\int_0^T \zeta_N(t)^{p/2}dt\Big)^\gamma\Big] \\ &\le C\Big(E\Big[\int_0^T \zeta_N(t)^{p/2}\Big]\Big)^\gamma, \end{aligned}$$

where the last inequality is due to the concavity of the function x^γ when $\gamma \in (0, 1)$. By (3.68), it is very easy to see that the above inequality is also true if $p = 2$. By the definition of ζ_N, (3.67) and Lemma 3.6.5, we have

$$\begin{aligned} E\Big[\int_0^T \zeta_N(t)^{p/2}dt\Big] \le C\Big\{\mathcal{E}(t) &+ E\Big[\int_0^T \|\mu(s, X_N(\kappa_N(s))) - \mu_N(s, X_N(\kappa_N(s)))\|^p dt\Big] \\ &+ E\Big[\int_0^T \|\sigma(s, X_N(\kappa_N(s))) - \sigma_N(s, X_N(\kappa_N(s)))\|^p dt\Big]\Big\} \end{aligned}$$

$$\leq CN^{-\alpha p}.$$

Thus,

$$E\Big[\sup_{t\leq T}(\phi(t)\|\chi_N(t)\|^2)^{p\gamma/2}\Big] \leq CN^{-\alpha p\gamma},$$

which leads to

$$E\Big[\sup_{t\leq T}\|\chi_N(t)\|^{p\gamma}\Big] \leq \exp((L+2)T)E\Big[\sup_{t\leq T}(\phi(t)\|\chi_N(t)\|^2)^{p\gamma/2}\Big]$$
$$\leq CN^{-\alpha p\gamma}.$$

Since $\gamma \in (0,1)$, we are done. □

Numerical Experiments

The example we use for our numerical experiment in this section is a 1-dimensional stochastic differential equation,

$$dX(t) = X(t)(1-|X(t)|)dt + |X(t)|^{3/2}dW(t), \quad X(0) = 1.$$

Here, $\mu(x) = x(1-|x|)$, $\sigma(x) = |x|^{3/2}$ and $t \in [0,1]$. Clearly, $\mu(x)$ and $\sigma(x)$ satisfy the monotone conditon and the polynomial growth condition with $l = 1$. We use Model 2 as our numerical scheme with $\alpha = 1/2$. We use 5 different time steps: $\Delta t = 2^{-12}, 2^{-11}, 2^{-10}, 2^{-9}, 2^{-8}$ and 1000 realizations for each discretisation. The following figure is the loglog plot of the experimental error with respect to the 5 different time steps. We can see that the numerical scheme converges strongly with order $\frac{1}{2}$.

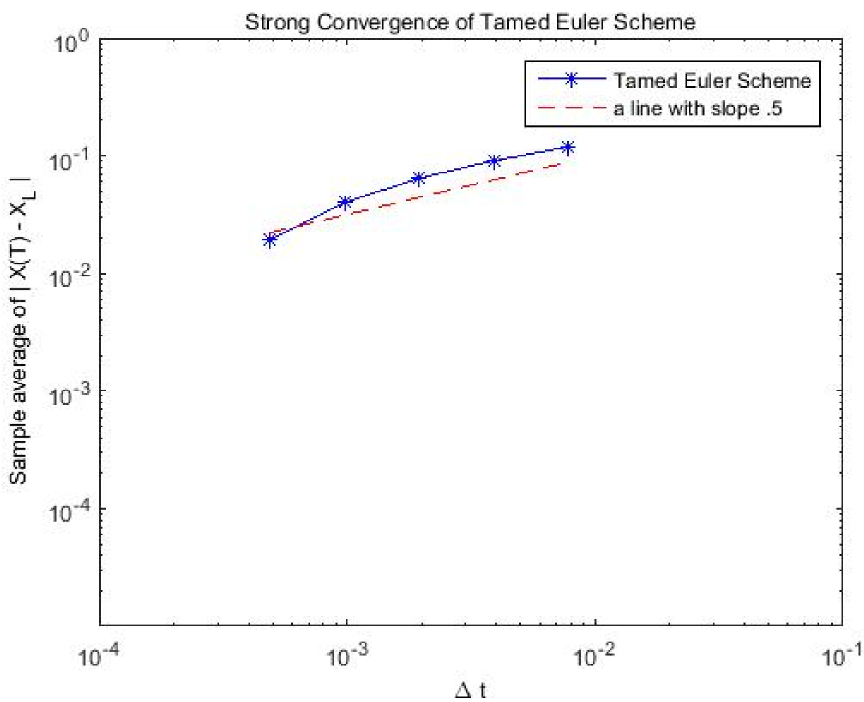

Figure 3.2: Log-log plot of the strong error from the numerical approximation versus the time step Δt with the drift and diffusion coefficients superlinearly growing.

Chapter 4

Weak Convergence of Euler-Maruyama Approximation of SDEs Driven by Brownian Motion

4.1 Introduction

Let us consider the following stochastic differential equation:

$$\begin{cases} dX(t) = \mu(t, X(t))dt + \sigma(t, X(t))dW(t), & t \in (0, T] \\ X(0) = x_0, \end{cases} \tag{4.1}$$

where $W(t)$ is a one-dimensional Wiener process starting at 0, $X(t)$ is a one-dimensional stochastic process and $\mu(t, x), \sigma(t, x)$ satisfy the following Lipschitz and linear growth condition

$$|\mu(t, x) - \mu(t, y)| + |\sigma(t, x) - \sigma(t, y)| < K(T)|x - y|, \quad t \in [0, T] \tag{4.2}$$

$$|\mu(t, x)| + |\sigma(t, x)| \leq K(T)(1 + |x|), \quad t \in [0, T] \tag{4.3}$$

such that the solution of (4.1) exists and is unique. Since we will use the second moment of the solution in our proof, we also assume that x_0 is independent of $\{W(t), 0 \le t \le T\}$ and $E[x_0^2] < \infty$ such that

$$E\Big[\sup_{0\le t\le T} X^2(t)\Big] < C(1 + E[x_0^2]), \tag{4.4}$$

where the constant C depends only on K and T.

We now give the Euler-Maruyama scheme. In this section, the time step is denoted by $\Delta = T/N$. For any integer i satisfying $0 \le i \le N$, set $t_i = i\Delta$. We define at each node in $[0, T]$: $Y_0 := x_0$ and

$$Y_{i+1} := Y_i + \mu(t_i, Y_i)\Delta + \sigma(t_i, Y_i)\Delta W_i, \qquad 0 \le i \le N-1, \tag{4.5}$$

where $\Delta W_i = W(t_{i+1}) - W(t_i)$. The continuous-time approximation is defined as:

$$Y(t) := Y_i + \mu(t_i, Y_i)(t - t_i) + \sigma(t_i, Y_i)(W(t) - W(t_i)) \tag{4.6}$$

$$= Y_i + \int_{t_i}^{t} \mu(t_i, Y_i)ds + \int_{t_i}^{t} \sigma(t_i, Y_i)dW(s) \quad \text{for } t \in [t_i, t_{i+1}). \tag{4.7}$$

Let us also recall the definition of the weak convergence of a numerical scheme. We say that a time discrete approximation Y converges in the weak sense with order $\beta \in (0, \infty]$ if for any function g in a suitable function space there exists a finite constant C and a positive constant δ_0 such that

$$|E[g(X(T))] - E[g(Y_N)]| \le C\delta^{\beta} \tag{4.8}$$

for any time discretization with maximum step size $\delta \in (0, \delta_0)$.

Before we continue, we first define some notations of function spaces. We denote by $C_b^l([0,T]\times\mathbb{R})$ the space of l times continuously differentiable functions $f : [0,T]\times\mathbb{R} \to \mathbb{R}$ for which all its partial derivatives up to order l are bounded uniformly in t (f may not be bounded). $C_b^l(\mathbb{R})$ is defined in a similar way. We also denote by $C_p^l([0,T]\times\mathbb{R})$

the space of l times continuously differentiable functions $f : [0,T] \times \mathbb{R} \to \mathbb{R}$ for which all its partial derivatives up to order l have polynomial growth uniformly in t.

It is well established that, provided μ, σ and g satisfy certain conditions, the Euler-Maruyama scheme has weak convergence rate 1 [5, 33, 60]. For example, in [33], if both $\mu(t,x)$ and $\sigma(t,x)$ are homogeneous, it is required that $\mu(x)$, $\sigma(x)^2$ and $g(x)$ are all in the function space $C_p^4(\mathbb{R}^m)$, together with some other conditions. While in [5], although g is only required to be measurable and bounded (or has a polynomial growth), μ and σ are assumed to be homogeneous and to be C^∞ functions with bounded derivatives of any order. See also [25, 34] for other related results.

Due to the close relation between the weak approximation of the solution of (4.1) and the Kolmogorov backward partial differential equation, Malliavin calculus, which is powerful to deal with the derivatives of functions of random variables, can serve as an efficient tool to analyze the approximation error. For example, in [5, 25, 34, 35], techniques from Malliavin calculus, like integration by parts, are used very often to assist to get the expressions of the approximation errors.

Another advantage of using Malliavin calculus to deal with the weak convergence problems is that it can also handle stochastic integrals with anticipating integrand. Therefore, the weak approximation problem of stochastic differential equations with terminal conditions can also be dealt with in the frame of Malliavin calculus, see e.g. [35]. In history, it had been believed for a long time that such equations with terminal conditions were not amenable to the analysis of approximation errors, due to the inability of Itô integral for anticipating integrands.

In this section, we do not assume the drift and diffusion terms are homogeneous or C^∞ functions. We only need $\mu(t,x) \in C_b^2([0,T] \times \mathbb{R})$, $\sigma^2(t,x) \in C_b^2([0,T] \times \mathbb{R})$ and $g(x) \in C_p^3(\mathbb{R})$. As we introduced in Section 2.2, if $\mu(t,x)$, $\sigma(t,x)$ and $g(x)$ satisfy such conditions and the linear growth condition, then $f(t,x) := E[g(X(T))|X(t) = x]$

satisfies the following **Kolmogorov backward equation:**

$$\begin{cases} \frac{\partial f}{\partial t}(t,x) + L_t f(t,x) = 0, \quad 0 \le t < T, x \in \mathbb{R} \\ f(T,x) = g(x), \end{cases} \tag{4.9}$$

where L_t is the second order differential operator defined by

$$L_t f(t,x) = \mu(t,x)\frac{\partial f}{\partial x}(t,x) + \frac{1}{2}\sigma^2(t,x)\frac{\partial^2 f}{\partial x^2}(t,x). \tag{4.10}$$

By the definition of f, we have

$$E[f(0,x_0)] = E[E[g(X(T)|X(0)=x_0)]] = E[g(X(T))]. \tag{4.11}$$

By the boundary condition, we have

$$E[f(T,Y(T))] = E[g(Y(T))]. \tag{4.12}$$

The traditional technique in the proof of the weak convergence of the Euler scheme is to write

$$\begin{aligned} &E[g(X(T))] - E[g(Y(T))] \\ &= -\big(E[f(T,Y(T))] - E[f(0,Y_0)]\big) \qquad \text{(by (4.11) and (4.12))} \\ &= -E\sum_{i=0}^{N-1}\Big[f\Big(\frac{(i+1)T}{N}, Y_{i+1}\Big) - f\Big(\frac{iT}{N}, Y_i\Big)\Big] \end{aligned} \tag{4.13}$$

and apply Taylor's formula on each difference of the above equality [59, 60]. In addition, equation (4.9) may also be used in the computations. In this section, we apply the techniques from Malliavin Calculus, such as the chain rule and integration by parts, in the computations, which results in less need of the smoothness of the drift, diffusion and test functions.

4.2 Preliminaries of Malliavin Calculus

There are many good monographs on Malliavin Calculus, see e.g. [4, 42, 43, 44]. In this section, we only introduce the materials that are necessary to our computations.

Suppose $W(t)$ is a one dimensional Wiener process on the filtered probability space $(\Omega, \mathcal{F}, \mathbb{F} = \{\mathcal{F}_t\}_{0\le t\le T}, P)$. For $h(t) \in L^2([0,T])$, we denote by $W(h)$ the Itô stochastic integral $\int_0^T h(t)dW(t)$.

Let $\mathcal{S}$ denote the set of all random variables of the form

$$f(W(h_1), \cdots, W(h_m)),$$

where m is a positive integer, $f : \mathbb{R}^m \to \mathbb{R}$ is a C^∞ function such that f and its partial derivatives have at most polynomial growth, and $h_i \in L^2([0,T]), i = 1, \cdots, m$. Before we continue, we point out a fact that the space $\mathcal{S}$ is dense in $L^p(\Omega)$ for every $p \ge 1$ [43].

Definition 4.1. *Let $F \in \mathcal{S}$, the Malliavin derivative of F is a stochastic process defined by*

$$D_tF = \sum_{i=1}^{m} \frac{\partial f}{\partial x_i}(W(h_1), \cdots, W(h_m))h_i(t).$$

We often write D_tF as DF if there is no confusion. Specifically, if $F = \int_0^T h(t)dW(t)$, then $DF = D_tF = h(t)$.

The operator $D : \mathcal{S} \subset L^p(\Omega) \to L^p(\Omega, L^2([0,T]))$ is closable for any $p \in [1, \infty)$. We denote by $\mathbb{D}^{1,p}$ the closure of $\mathcal{S}$ with respect to the norm

$$||F||_{\mathbb{D}^{1,p}} = \left(E[|F|^p] + E[||DF||^p_{L^2}]\right)^{1/p},$$

and

$$\mathbb{D}^{1,\infty} = \cap_{p\in\mathbb{N}}\mathbb{D}^{1,p}.$$

It is immediate, using the definition of D, that the product rule holds. That is, if $F, G \in \mathbb{D}^{1,p}$, then $FG \in \mathbb{D}^{1,p}$ and $D(FG) = FDG + GDF$.

Proposition 4.2 (Chain rule, [4], Proposition 10)**.** *Let $\phi : \mathbb{R} \to \mathbb{R}$ be a continuously differentiable function with bounded derivative. Suppose $F \in \mathbb{D}^{1,p}$ for some $p \geq 1$. Then $\phi(F) \in \mathbb{D}^{1,p}$ and we have*

$$D\phi(F) = \phi'(F)DF.$$

If $F \in D^{1,\infty}$, then the conclusion is true for ϕ which is a continuously differentiable function with its derivative having a polynomial growth.

The following proposition is very useful in our proof.

Proposition 4.3 ([44], Corollary 3.13)**.** *Let $u = u(s), s \in [0,T]$, be an $\mathcal{F}_t$-adapted stochastic process and assume that $u(s) \in \mathbb{D}^{1,2}$ for all s. Then*

1. *$D_t(s), s \in [0,T]$, is $\mathcal{F}_t$-adapted for all t;*
2. *$D_t u(s) = 0$ for $t > s$.*

We now introduce the adjoint operator of D.

Definition 4.4. *We denote by Dom(δ) the subset of $L^2(\Omega, L^2([0,T]))$ composed of those elements u such that there exists a constant $c > 0$ satisfying*

$$\left|E[\langle DF, u\rangle_{L^2}]\right| \leq c\sqrt{E[F^2]} \quad \textit{for all } F \in \mathbb{D}^{1,2}. \tag{4.14}$$

Fix $u \in Dom(\delta)$. By (4.14), the linear operator $F \mapsto E[\langle DF, u\rangle_{L^2}]$ is continuous from $\mathcal{S}$, equipped with the $L^2(\Omega)$ norm, into $\mathbb{R}$. So we can extend it to a linear operator from $L^2(\Omega)$ into $\mathbb{R}$. By the Riesz representation theorem, there exists a unique element in $L^2(\Omega)$, noted $\delta(u)$, such that $E[\langle DF, u\rangle_{L^2}] = E[F\delta(u)]$. This is our next definition.

Definition 4.5. *If $u \in Dom(\delta)$, then $\delta(u)$ is the unique element of $L^2(\Omega)$ characterized by the following duality formula:*

$$E[F\delta(u)] = E[\langle DF, u\rangle_{L^2}] \tag{4.15}$$

for all $F \in \mathbb{D}^{1,2}$.

Formula (4.15) is often called an **integration by parts formula**. Usually, if u is a $\mathcal{F}_T$-measurable stochastic process and is such that $E[\int_0^T u^2 dt] < \infty$, $\delta(u)$ is often written as $\int_0^T u\delta W(t)$ and we call it the **Skorohod integral**. We also point out that if u, in addition, is adapted to the filtration $\mathcal{F}_t$, the Skorohod integral $\int_0^T u\delta W(t)$ is nothing but the Itô integral $\int_0^T udW(t)$. Therefore, if $s_1 < s_2$ and u is a fixed $\mathcal{F}_{s_1}$-measurable random variable , it is straightforward that $\int_{s_1}^{s_2} u\delta W(t) = \int_{s_1}^{s_2} udW(t) = u \cdot (W(s_2) - W(s_1))$.

The following proposition is useful in many situations.

Proposition 4.6 ([42], Proposition 2.5.4)**.** *Let* $F \in \mathbb{D}^{1,2}$ *and* $u \in Dom(\delta)$ *be such that the three expectations* $E[F^2||u||^2_{L^2}]$, $E[F^2\delta(u)^2]$ *and* $E[\langle DF, u\rangle^2_{L^2}]$ *are finite. Then* $Fu \in Dom(\delta)$ *and*

$$\delta(Fu) = F\delta(u) - \langle DF, u\rangle_{L^2}. \tag{4.16}$$

For example, if $F \in \mathbb{D}^{1,2}$ and $u = 1$, then we have

$$\int_0^T F\delta W(t) = F\int_0^T 1\delta W(t) - \int_0^T D_t F \cdot 1dt = F(W(T) - W(0)) - \int_0^T D_t F dt. \tag{4.17}$$

We will use this trick very often in our proof to find $\int_0^T F\delta W(t)$.

4.3 Weak Convergence of the EM scheme using Malliavin Calculus

We now state the main theorem of this chapter, which assumes weaker conditions on the drift and diffusion coefficients. This new theorem is included in Rosiński and Wang [54].

Theorem 4.7 ([54], Theorem 3.1)**.** *Suppose the following conditions hold:*

1. $\mu(t, x) \in C_b^2([0, T] \times \mathbb{R})$, $\sigma^2(t, x) \in C_b^2([0, T] \times \mathbb{R})$ *and* $g(x) \in C_p^3(\mathbb{R})$;

2. the linear growth condition for $\mu(t,x)$ and $\sigma(t,x)$ hold;

3. all the partial derivatives of $\mu(t,x)$ and $\sigma(t,x)$ with respect to x up to order 2 are bounded by a constant $M > 0$ for any t;

4. there exists a positive number L such that $|\sigma(t,x)| \geq L$ for any $(t,x) \in [0,T]\times\mathbb{R}$.

Then the Euler-Maruyama scheme (4.5) has weak convergence order 1. That is, there exists a positive number C, which depends on M, T and L, such that

$$\big|E[g(X(T))] - E[g(Y(T))]\big| \leq CN^{-1}. \tag{4.18}$$

Before we prove this theorem, we first give two lemmas that are needed in our proof.

Lemma 4.7.1. *Suppose $F, G \in \mathbb{D}^{1,2}$ and $\int_0^T D_t G dt \neq 0$ a.e.. Let $\phi : \mathbb{R} \to \mathbb{R}$ be a continuously differentiable function with bounded derivative, then*

$$E[F\phi'(G)] = E\Big[\phi(G)\delta\Big(\frac{F}{\int_0^T D_t G dt}\Big)\Big]. \tag{4.19}$$

If $\phi(x)$ is continuously differentiable with polynomial growth and $G \in \mathbb{D}^{1,\infty}$, the above conclusion is also true.

Proof. By chain rule, we have

$$\int_0^T F D_t \phi(G) dt = \int_0^T F\phi'(G) D_t G dt = F\phi'(G)\int_0^T D_t G dt.$$

Observing that $\int_0^T D_t G dt$ is nonzero, by duality, we have

$$\begin{aligned} E[F\phi'(G)] &= E\Big[\frac{\int_0^T F D_t\phi(G)dt}{\int_0^T D_t G dt}\Big] \\ &= E\Big[\Big\langle D_t\phi(G), \frac{F}{\int_0^T D_t G dt}\Big\rangle_{L^2}\Big] \end{aligned}$$

$$= E\Big[\phi(G)\delta\Big(\frac{F}{\int_0^T D_t G dt}\Big)\Big].$$

If $\phi(x)$ is continuously differentiable with polynomial growth and $F \in \mathbb{D}^{1,\infty}$, by the same argument above and Proposition 4.2, we can easily get (4.19). □

See [5, 29] for a more general result with ϕ defined on $\mathbb{R}^m$.

It is well known that, as long as $\mu(t, x)$ and $\sigma(t, x)$ satisfy the linear growth condition, one has [8], for any $p \geq 1$,

$$\sup_{0 \leq t \leq T} \|Y(t)\| \in L^p(\Omega). \tag{4.20}$$

Note also that

$$Y(s) = Y_i + \mu(t_i, Y_i)(s - t_i) + \sigma(t_i, Y_i)(W(s) - W(t_i)).$$

By the chain rule (Proposition 4.2), one has the following:

Lemma 4.7.2. *Suppose $\mu(t, x)$, $\sigma(t, x)$ and $g(x)$ are assumed as in Theorem 4.7, then $Y(s) \in \mathbb{D}^{1,2}$ and*

$$D_\tau Y(s) = \sigma(t_i, Y_i)\mathbb{1}_{(t_i,s]}(\tau), \quad s \in (t_i, t_{i+1}], \tau \in (t_i, T]. \tag{4.21}$$

Furthermore, $F(s) := f(s, Y(s)) \in \mathbb{D}^{1,2}$ and

$$D_\tau F(s) = \frac{\partial f}{\partial x}(s, Y(s)) D_\tau Y(s). \tag{4.22}$$

We now give the proof of Theorem 4.7.

Proof. First of all, by (4.10) and (4.13),

$$E[g(X(T))] - E[g(Y(T))]$$

$$
\begin{aligned}
&= -E\sum_{i=0}^{N-1}\Big[f\Big(\frac{(i+1)T}{n}, Y_{i+1}\Big) - f\Big(\frac{iT}{n}, Y_i\Big)\Big] \\
&= -E\sum_{i=0}^{N-1}\int_{t_i}^{t_{i+1}}\Big[\frac{\partial f}{\partial t}(s, Y(s)) + \mu(t_i, Y_i)\frac{\partial f}{\partial x}(s, Y(s)) + \frac{1}{2}\sigma^2(t_i, Y_i)\frac{\partial^2 f}{\partial x^2}(s, Y(s))\Big]ds \\
&= -E\sum_{i=0}^{N-1}\int_{t_i}^{t_{i+1}}\Big[\mu(t_i, Y_i)\frac{\partial f}{\partial x}(s, Y(s)) + \frac{1}{2}\sigma^2(t_i, Y_i)\frac{\partial^2 f}{\partial x^2}(s, Y(s)) - L_t f(s, Y(s))\Big]ds \\
&= -E\sum_{i=0}^{N-1}\int_{t_i}^{t_{i+1}}\Big[(\mu(t_i, Y_i) - \mu(s, Y(s)))\frac{\partial f}{\partial x}(s, Y(s)) \\
&\quad + \frac{1}{2}(\sigma^2(t_i, Y_i) - \frac{1}{2}\sigma^2(s, Y(s)))\frac{\partial^2 f}{\partial x^2}(s, Y(s))\Big]ds \\
&= E\sum_{i=0}^{N-1}\int_{t_i}^{t_{i+1}}\Big[(\mu(s, Y(s)) - \mu(t_i, Y_i))\frac{\partial f}{\partial x}(s, Y(s))\Big]ds \\
&\quad + E\sum_{i=0}^{N-1}\int_{t_i}^{t_{i+1}}\Big[\frac{1}{2}(\sigma^2(s, Y(s)) - \sigma^2(t_i, Y_i))\frac{\partial^2 f}{\partial x^2}(s, Y(s))\Big]ds \\
&=: I_N + J_N
\end{aligned} \tag{4.23}
$$

In the following, we consider the individual differences in I_N and J_N.

By (4.17), Lemma 4.7.1 and Lemma 4.7.2,

$$
\begin{aligned}
&\int_{t_i}^{t_{i+1}} E\Big[\mu(t_i, Y_i)\frac{\partial f}{\partial x}(s, Y(s))\Big]ds \\
&= \int_{t_i}^{t_{i+1}} E\Big[F(s)\delta\Big(\frac{\mu(t_i, Y_i)\mathbb{1}_{(t_i, t_{i+1}]}}{\sigma(t_i, Y_i)(s - t_i)}\Big)\Big]ds \\
&= \int_{t_i}^{t_{i+1}} E\Big[F(s)\int_{t_i}^{t_{i+1}}\frac{\mu(t_i, Y_i)}{\sigma(t_i, Y_i)(s - t_i)}\delta W(\tau)\Big]ds \\
&= \int_{t_i}^{t_{i+1}} E\Big[F(s)\frac{\mu(t_i, Y_i)\Delta W_i}{\sigma(t_i, Y_i)(s - t_i)}\Big]ds.
\end{aligned} \tag{4.24}
$$

Similarly,

$$
\begin{aligned}
&\int_{t_i}^{t_{i+1}} E\Big[\mu(s, Y(s))\frac{\partial f}{\partial x}(s, Y(s))\Big]ds \\
&= \int_{t_i}^{t_{i+1}} E\Big[F(s)\delta\Big(\frac{\mu(s, Y(s))\mathbb{1}_{(t_i, t_{i+1}]}}{\sigma(t_i, Y_i)(s - t_i)}\Big)\Big]ds
\end{aligned}
$$

$$
\begin{aligned}
&= \int_{t_i}^{t_{i+1}} E\Big[F(s)\frac{\delta(\mu(s,Y(s))\mathbb{1}_{(t_i,t_{i+1}]})}{\sigma(t_i,Y_i)(s-t_i)}\Big]ds \\
&= \int_{t_i}^{t_{i+1}} E\Big[F(s)\frac{\mu(s,Y(s))\Delta W_i - \int_{t_i}^{t_{i+1}} D\mu(s,Y(s))d\tau}{\sigma(t_i,Y_i)(s-t_i)}\Big]ds \\
&= E\int_{t_i}^{t_{i+1}} F(s)\frac{\mu(s,Y(s))\Delta W_i - \int_{t_i}^{t_{i+1}} \frac{\partial\mu}{\partial x}(s,Y(s))D_\tau Y(s)d\tau}{\sigma(t_i,Y_i)(s-t_i)}ds \\
&= E\int_{t_i}^{t_{i+1}} F(s)\frac{\mu(s,Y(s))\Delta W_i}{\sigma(t_i,Y_i)(s-t_i)} - F(s)\frac{\partial\mu}{\partial x}(s,Y(s))ds \qquad (4.25)
\end{aligned}
$$

By (4.24) and (4.25), we have

$$
\begin{aligned}
I_N &= E\sum_{i=0}^{N-1}\int_{t_i}^{t_{i+1}} [(\mu(s,Y(s)) - \mu(t_i,Y_i))\frac{\partial f}{\partial x}(s,Y(s))]ds \\
&= E\sum_{i=0}^{N-1}\int_{t_i}^{t_{i+1}} F(s)\Big[\frac{(\mu(s,Y(s)) - \mu(t_i,Y_i))\Delta W_i}{\sigma(t_i,Y_i)(s-t_i)} - \frac{\partial\mu}{\partial x}(s,Y(s))\Big]ds \qquad (4.26)
\end{aligned}
$$

Similarly, for $F(t_i) = f(t_i, Y(t_i)) = f(t_i, Y_i)$, we have $D_\tau F(t_i) = \frac{\partial f}{\partial x}(t_i, Y_i)D_\tau Y(t_i) = 0$, $\tau \in (t_i, t_{i+1}]$. We have, by duality,

$$
\begin{aligned}
0 &= E\int_{t_i}^{t_{i+1}} \frac{\mu(s,Y(s)) - \mu(t_i,Y_i)}{\sigma(t_i,Y_i)(s-t_i)} D_\tau F(t_i,Y_i)d\tau \\
&= E\Big[F(t_i)\delta\Big(\frac{\mu(s,Y(s)) - \mu(t_i,Y_i)}{\sigma(t_i,Y_i)(s-t_i)}\Big)\Big] \\
&= E\Big[F(t_i)\Big(\frac{(\mu(s,Y(s)) - \mu(t_i,Y_i))\Delta W_i}{\sigma(t_i,Y_i)(s-t_i)} - \frac{\partial\mu}{\partial x}(s,Y(s))\Big)\Big] \qquad (4.27)
\end{aligned}
$$

Adding (4.27) to (4.26), and observing $\Delta W_i = W(s) - W(t_i) + W(t_{i+1}) - W(s)$, we have

$$
\begin{aligned}
I_N &= E\sum_{i=0}^{N-1}\int_{t_i}^{t_{i+1}} [(\mu(s,Y(s)) - \mu(t_i,Y_i))\frac{\partial f}{\partial x}(s,Y(s))]ds \\
&= E\sum_{i=0}^{N-1}\int_{t_i}^{t_{i+1}} (F(s) - F(t_i))\Big[\frac{(\mu(s,Y(s)) - \mu(t_i,Y_i))\Delta W_i}{\sigma(t_i,Y_i)(s-t_i)} - \frac{\partial\mu}{\partial x}(s,Y(s))\Big]ds \\
&= \sum_{i=0}^{N-1}\int_{t_i}^{t_{i+1}} E\Big\{(F(s) - F(t_i))\Big[\frac{(\mu(s,Y(s)) - \mu(t_i,Y_i))(W(s) - W(t_i))}{\sigma(t_i,Y_i)(s-t_i)}
\end{aligned}
$$

$$- \frac{\partial \mu}{\partial x}(s, Y(s))\Big]\Big\}ds, \tag{4.28}$$

where in the last equality we used the fact that $W(t_{i+1}) - W(s)$ is independent of $\{\mathcal{F}_t, 0 \le t \le s\}$ and $E[W(t_{i+1}) - W(s)] = 0$.

To reduce the notation burden in the following, we fix an interval $(t_i, t_{i+1}]$ and denote by $\Delta_s = s - t_i$ and $\Delta_{\mu\sigma} = \mu(t_i, Y_i)(s-t_i) + \sigma(t_i, Y_i)(W(s) - W(t_i)), s \in (t_i, t_{i+1}]$.

In view of (4.20), it follows directly that, for any positive integer k,

$$E[|\mu(t_i, Y_i)|^k] \le E[(1 + |Y_i|)^k] \le C(k, T), \tag{4.29}$$

$$E[|\sigma(t_i, Y_i)|^k] \le E[(1 + |Y_i|)^k] \le C(k, T). \tag{4.30}$$

The following lemma gives the bound of the increment of $\Delta_{\mu\sigma} = \mu(t_i, Y_i)(s - t_i) + \sigma(t_i, Y_i)(W(s) - W(t_i))$.

Lemma 4.7.3. *Suppose $\mu(t, x)$ and $\sigma(t, x)$ satisfy the linear growth condition, then there exists a positive constant $C(T)$ such that:*

$$E[|\Delta_{\mu\sigma}|] \le C(T)\Delta_s^{1/2}, \tag{4.31}$$

$$E[|\Delta_{\mu\sigma}|^2] \le C(T)\Delta_s. \tag{4.32}$$

Proof. This is straightforward by combining the linear growth condition, the bound of Y_i and Cauchy-Schwarz.

$$\begin{aligned}
&E[|\mu(t_i, Y_i)(s - t_i) + \sigma(t_i, Y_i)(W(s) - W(t_i))|] \\
&\le E[|\mu(t_i, Y_i)\Delta_s|] + E[|\sigma(t_i, Y_i)(W(s) - W(t_i))|] \\
&= \Delta_s E[|\mu(t_i, Y_i)|] + E[|W(s) - W(t_i)|]E[|\sigma(t_i, Y_i)|] \\
&\le \Delta_s E[1 + |Y_i|] + \Delta_s^{1/2} E[1 + |Y_i|] \\
&= \Delta_s^{1/2}\Big(\Delta_s^{1/2} E[1 + |Y_i|] + E[1 + |Y_i|\Big) \\
&\le C(T)\Delta_s^{1/2}.
\end{aligned}$$

The L^2 bound of $\Delta_{\mu\sigma}$ can be proved in a similar way. □

In fact, applying the same argument as in Lemma 4.7.3, one obtains the following estimate with no difficulty:

$$E[|\Delta_{\mu\sigma}|^k] \leq C(k,T)\Delta_s^{k/2}.$$

We now look at the terms in the last equality of (4.28). By Taylor expansion,

$$\begin{aligned} F(s) - F(t_i) &= f(t_i + \Delta_s, Y_i + \Delta_{\mu\sigma}) - f(t_i, Y_i) \\ &= \Delta_s \frac{\partial f}{\partial t}(t_i, Y_i) + \Delta_{\mu\sigma}\frac{\partial f}{\partial x}(t_i, Y_i) + R_1, \end{aligned} \tag{4.33}$$

where R_1 is the Lagrange-type residual of the expansion and takes the form

$$\begin{aligned} R_1 = &\frac{1}{2}\Delta_s^2 \frac{\partial^2 f}{\partial t^2}(t_i + c_1\Delta_s, Y_i + c_1\Delta_{\mu\sigma}) + \frac{1}{2}\Delta_{\mu\sigma}^2 \frac{\partial^2 f}{\partial x^2}(t_i + c_1\Delta_s, Y_i + c_1\Delta_{\mu\sigma}) \\ &+ \Delta_s\Delta_{\mu\sigma}\frac{\partial^2 f}{\partial t \partial x}(t_i + c_1\Delta_s, Y_i + c_1\Delta_{\mu\sigma}), \end{aligned} \tag{4.34}$$

where $c_1 \in (0,1)$. Since $f \in C_p^3(\mathbb{R})$ and the bound of $\Delta_{\mu\sigma}$ holds (Lemma 4.31), it is obvious that

$$E[|R_1|] \leq \sqrt{E[|R_1|^2]} \leq C(T,M)\Delta_s. \tag{4.35}$$

Similarly,

$$\begin{aligned} &\Big(\mu(s, Y(s)) - \mu(t_i, Y_i)\Big)\Big(W(s) - W(t_i)\Big) \\ = &\Big(\Delta_s \frac{\partial \mu}{\partial t}(t_i, Y_i) + \Delta_{\mu\sigma}\frac{\partial \mu}{\partial x}(t_i, Y_i) + R_2\Big) \cdot \Big(W(s) - W(t_i)\Big), \end{aligned} \tag{4.36}$$

where R_2, similar to R_1, takes the form

$$\begin{aligned} R_2 = &\frac{1}{2}\Delta_s^2 \frac{\partial^2 \mu}{\partial t^2}(t_i + c_2\Delta_s, Y_i + c_2\Delta_{\mu\sigma}) + \frac{1}{2}\Delta_{\mu\sigma}^2 \frac{\partial^2 \mu}{\partial x^2}(t_i + c_2\Delta_s, Y_i + c_2\Delta_{\mu\sigma}) \\ &+ \Delta_s\Delta_{\mu\sigma}\frac{\partial^2 \mu}{\partial t \partial x}(t_i + c_2\Delta_s, Y_i + c_2\Delta_{\mu\sigma}), \end{aligned} \tag{4.37}$$

where $c_2 \in (0,1)$, and has the following bound

$$E[|R_2|] \leq \sqrt{E[|R_2|^2]} \leq C(T,M)\Delta_s. \tag{4.38}$$

By the same arguments of getting (4.35) and (4.38), we also have

$$E[|R_1 R_2|^2] \leq C(T,M)\Delta_s^4, \tag{4.39}$$

$$E\Big[\Big|R_2\Big(\Delta_s \frac{\partial f}{\partial t}(t_i, Y_i) + \Delta_{\mu\sigma}\frac{\partial f}{\partial x}(t_i, Y_i)\Big)\Big|^2\Big] \leq C(T,M)\Delta_s^3, \tag{4.40}$$

$$E\Big[\Big|R_1\Big(\Delta_s \frac{\partial \mu}{\partial t}(t_i, Y_i) + \Delta_{\mu\sigma}\frac{\partial \mu}{\partial x}(t_i, Y_i)\Big)\Big|^2\Big] \leq C(T,M)\Delta_s^3. \tag{4.41}$$

Lastly,

$$\frac{\partial \mu}{\partial x}(s, Y(s)) = \frac{\partial \mu}{\partial x}(t_i, Y_i) + R_3 \tag{4.42}$$

where R_3 takes the form

$$R_3 = \Delta_s \frac{\partial^2 \mu}{\partial t \partial x}(t_i + c_3\Delta_s, Y_i + c_3\Delta_{\mu\sigma}) + \Delta_{\mu\sigma}\frac{\partial^2 \mu}{\partial x^2}(t_i + c_3\Delta_s, Y_i + c_3\Delta_{\mu\sigma}), \tag{4.43}$$

where $c_3 \in (0,1)$. Similar to getting (4.35), R_3 also satisfies

$$E[|R_3|] \leq \sqrt{E[|R_3|^2]} \leq C(T,M)\Delta_s^{1/2}. \tag{4.44}$$

Similar to the arguments of getting (4.39) and (4.40), we can also have

$$E[|R_1 R_3|] \leq C(T,M)\Delta_s^{3/2}, \tag{4.45}$$

$$E\Big[\Big|R_3\Big(\Delta_s \frac{\partial f}{\partial t}(t_i, Y_i) + \Delta_{\mu\sigma}\frac{\partial f}{\partial x}(t_i, Y_i)\Big)\Big|\Big] \leq C(T,M)\Delta_s, \tag{4.46}$$

$$E\Big[\Big|R_1 \frac{\partial \mu}{\partial x}(t_i, Y_i)\Big|\Big] \leq C(T,M)\Delta_s. \tag{4.47}$$

Combining (4.33) and (4.36), we get that

$$\begin{aligned}
&(F(s)-F(t_i))\frac{(\mu(s,Y(s))-\mu(t_i,Y_i))(W(s)-W(t_i))}{\sigma(t_i,Y_i)(s-t_i)}\\
&=\frac{W(s)-W(t_i)}{\sigma(t_i,Y_i)\Delta_s}\Big[\Delta_s^2\frac{\partial f}{\partial t}(t_i,Y_i)\frac{\partial \mu}{\partial t}(t_i,Y_i)\\
&\qquad+\Delta_s\Delta_{\mu\sigma}\Big(\frac{\partial f}{\partial t}(t_i,Y_i)\frac{\partial \mu}{\partial x}(t_i,Y_i)+\frac{\partial f}{\partial x}(t_i,Y_i)\frac{\partial \mu}{\partial t}(t_i,Y_i)\Big)\\
&\qquad+\Delta_{\mu\sigma}^2\frac{\partial f}{\partial x}(t_i,Y_i)\frac{\partial \mu}{\partial x}(t_i,Y_i)+R_2\Big(\Delta_s\frac{\partial f}{\partial t}(t_i,Y_i)+\Delta_{\mu\sigma}\frac{\partial f}{\partial x}(t_i,Y_i)\Big)\\
&\qquad+R_1\Big(\Delta_s\frac{\partial \mu}{\partial t}(t_i,Y_i)+\Delta_{\mu\sigma}\frac{\partial \mu}{\partial x}(t_i,Y_i)\Big)+R_1R_2\Big].
\end{aligned} \tag{4.48}$$

Since $W(s)-W(t_i)$ is independent of $\{\mathcal{F}_t, 0\le t\le t_i\}$, the first term on the right hand side of (4.48) satisfies

$$E\Big[\frac{W(s)-W(t_i)}{\sigma(t_i,Y_i)\Delta_s}\Delta_s^2\frac{\partial f}{\partial t}(t_i,Y_i)\frac{\partial \mu}{\partial t}(t_i,Y_i)\Big]=0. \tag{4.49}$$

Therefore, the second term on the right hand side of (4.48) satisfies

$$\begin{aligned}
&\Big|E\Big[\frac{W(s)-W(t_i)}{\sigma(t_i,Y_i)\Delta_s}\Delta_s\Delta_{\mu\sigma}\Big(\frac{\partial f}{\partial t}(t_i,Y_i)\frac{\partial \mu}{\partial x}(t_i,Y_i)+\frac{\partial f}{\partial x}(t_i,Y_i)\frac{\partial \mu}{\partial t}(t_i,Y_i)\Big)\Big]\Big|\\
&\le E\Big[\Big|\frac{W(s)-W(t_i)}{\sigma(t_i,Y_i)}\Delta_{\mu\sigma}\Big(\frac{\partial f}{\partial t}(t_i,Y_i)\frac{\partial \mu}{\partial x}(t_i,Y_i)+\frac{\partial f}{\partial x}(t_i,Y_i)\frac{\partial \mu}{\partial t}(t_i,Y_i)\Big)\Big|\Big]\\
&\le C(L,M)E[|(W(s)-W(t_i))\Delta_{\mu\sigma}|]\\
&\le C(L,M)\sqrt{E[(W(s)-W(t_i))^2]}\sqrt{E[\Delta_{\mu\sigma}^2]}\\
&\le C(L,M,T)\Delta_s.
\end{aligned} \tag{4.50}$$

Similarly, observing $E[W(s)-W(t_i)]=0, E[(W(s)-W(t_i))^3]=0$, the third term in (4.48) satisfies

$$\begin{aligned}
&\Big|E\Big[\frac{W(s)-W(t_i)}{\sigma(t_i,Y_i)\Delta_s}\Delta_{\mu\sigma}^2\frac{\partial f}{\partial x}(t_i,Y_i)\frac{\partial \mu}{\partial x}(t_i,Y_i)\Big]\Big|\\
&=\Big|E\Big[\frac{W(s)-W(t_i)}{\sigma(t_i,Y_i)\Delta_s}\Big(\Delta_s^2\mu^2(t_i,Y_i)+2\Delta_s(W(s)-W(t_i))\mu(t_i,Y_i)\sigma(t_i,Y_i)
\end{aligned}$$

$$
\begin{aligned}
&\quad + (W(s) - W(t_i))^2\sigma^2(t_i, Y_i)\Big)\frac{\partial f}{\partial x}(t_i, Y_i)\frac{\partial \mu}{\partial x}(t_i, Y_i)\Big]\Big| \\
&= \Big|E\Big[\frac{(W(s) - W(t_i))\Delta_s\mu^2(t_i, Y_i)}{\sigma(t_i, Y_i)}\Big] + 2E[(W(s) - W(t_i))^2\mu(t_i, Y_i)] \\
&\quad + E\Big[\frac{(W(s) - W(t_i))^3\sigma(t_i, Y_i)}{\Delta_s}\frac{\partial f}{\partial x}(t_i, Y_i)\frac{\partial \mu}{\partial x}(t_i, Y_i)\Big]\Big| \\
&= |0 + 2E[(W(s) - W(t_i))^2\mu(t_i, Y_i)] + 0| \\
&= |2E[(W(s) - W(t_i))^2]E[\mu(t_i, Y_i)]| \\
&\leq C(T)\Delta_s. \qquad (4.51)
\end{aligned}
$$

By (4.40)), the fourth term on the right hand side of (4.48) satisfies

$$
\begin{aligned}
&\Big|E\Big[\frac{W(s) - W(t_i)}{\sigma(t_i, Y_i)\Delta_s}R_2\Big(\Delta_s\frac{\partial f}{\partial t}(t_i, Y_i) + \Delta_{\mu\sigma}\frac{\partial f}{\partial x}(t_i, Y_i)\Big)\Big]\Big| \\
&\leq \frac{C(L)}{\Delta_s}E\Big[\Big|(W(s) - W(t_i))R_2\Big(\Delta_s\frac{\partial f}{\partial t}(t_i, Y_i) + \Delta_{\mu\sigma}\frac{\partial f}{\partial x}(t_i, Y_i)\Big)\Big|\Big] \\
&\leq \frac{C(L)}{\Delta_s}\sqrt{E[(W(s) - W(t_i))^2]}\sqrt{E\Big[\Big|R_2\Big(\Delta_s\frac{\partial f}{\partial t}(t_i, Y_i) + \Delta_{\mu\sigma}\frac{\partial f}{\partial x}(t_i, Y_i)\Big)\Big|^2\Big]} \\
&\leq \frac{C(L, M, T)}{\Delta_s}\Delta_s^{1/2}\Delta_s^{3/2} \\
&\leq C(L, M, T)\Delta_s. \qquad (4.52)
\end{aligned}
$$

Similarly, the fifth term on the right hand side of (4.48) satisfies

$$
\Big|E\Big[\frac{W(s) - W(t_i)}{\sigma(t_i, Y_i)\Delta_s}R_1\Big(\Delta_s\frac{\partial \mu}{\partial t}(t_i, Y_i) + \Delta_{\mu\sigma}\frac{\partial \mu}{\partial x}(t_i, Y_i)\Big)\Big]\Big| \leq C(L, M, T)\Delta_s. \qquad (4.53)
$$

Finally, by (4.39), the last term on the right hand side of (4.48) satisfies

$$
\begin{aligned}
\Big|E\Big[\frac{W(s) - W(t_i)}{\sigma(t_i, Y_i)\Delta_s}R_1R_2\Big]\Big| &\leq \frac{C(L)}{\Delta_s}E[|(W(s) - W(t_i))R_1R_2|] \\
&\leq \frac{C(L)}{\Delta_s}\sqrt{E[(W_s - W_{t_i})^2]}\sqrt{E[|R_1R_2|^2]} \\
&\leq \frac{C(L, M, T)}{\Delta_s}\Delta_s^{1/2}\Delta_s^2 \\
&= C(L, M, T)\Delta_s^{3/2}. \qquad (4.54)
\end{aligned}
$$

Combining (4.48), (4.49), (4.50), (4.51), (4.52), (4.53) and (4.54), we have

$$\Big|E\Big[(F(s)-F(t_i))\frac{(\mu(s,Y(s))-\mu(t_i,Y_i))(W(s)-W(t_i))}{\sigma(t_i,Y_i)(s-t_i)}\Big]\Big| \le C(L,M,T)\Delta_s. \quad (4.55)$$

On the other hand, by (4.33) and (4.42), the other term in (4.28) is

$$\begin{aligned}
&\Big(F(s)-F(t_i)\Big)\frac{\partial\mu}{\partial x}(s,Y(s))\\
&=\Big(\Delta_s\frac{\partial f}{\partial t}(t_i,Y_i)+\Delta_{\mu\sigma}\frac{\partial f}{\partial x}(t_i,Y_i)+R_1\Big)\Big(\frac{\partial\mu}{\partial x}(t_i,Y_i)+R_3\Big)\\
&=\Delta_s\frac{\partial f}{\partial t}(t_i,Y_i)\frac{\partial\mu}{\partial x}(t_i,Y_i)+\Delta_{\mu\sigma}\frac{\partial f}{\partial x}(t_i,Y_i)\frac{\partial\mu}{\partial x}(t_i,Y_i)+R_1\frac{\partial\mu}{\partial x}(t_i,Y_i)\\
&\quad+R_3\Big(\Delta_s\frac{\partial f}{\partial t}(t_i,Y_i)+\Delta_{\mu\sigma}\frac{\partial f}{\partial x}(t_i,Y_i)\Big)+R_1R_3 \qquad (4.56)
\end{aligned}$$

Due to the assumption that $\mu(t,x)\in C_b^2(\mathbb{R})$ and $f\in C_p^3(\mathbb{R})$, the first term of the right hand side of (4.56) satisfies

$$\Big|E\Big[\Delta_s\frac{\partial f}{\partial t}(t_i,Y_i)\frac{\partial\mu}{\partial x}(t_i,Y_i)\Big]\Big|\le C(M)\Delta_s. \quad (4.57)$$

Similarly, the second term of the right hand side of (4.56) satisfies

$$\begin{aligned}
&\Big|E\Big[\Delta_{\mu\sigma}\frac{\partial f}{\partial x}(t_i,Y_i)\frac{\partial\mu}{\partial x}(t_i,Y_i)\Big]\Big|\\
&=\Big|E\Big[\Delta_s\mu(t_i,Y_i)\frac{\partial f}{\partial x}(t_i,Y_i)\frac{\partial\mu}{\partial x}(t_i,Y_i)+(W(s)-W(t_i))\sigma(t_i,Y_i)\frac{\partial f}{\partial x}(t_i,Y_i)\frac{\partial\mu}{\partial x}(t_i,Y_i)\Big]\Big|\\
&=\Big|E\Big[\Delta_s\mu(t_i,Y_i)\frac{\partial f}{\partial x}(t_i,Y_i)\frac{\partial\mu}{\partial x}(t_i,Y_i)+0\Big]\Big|\\
&\le C(T,M)\Delta_s. \qquad (4.58)
\end{aligned}$$

The estimates of the remaining terms on the right hand side of (4.56) are exactly (4.45), (4.46) and (4.47). Therefore, by (4.56), (4.57), (4.58), (4.45), (4.46) and (4.47), we have

$$\Big|E\Big[\Big(F(s)-F(t_i)\Big)\frac{\partial\mu}{\partial x}(s,Y(s))\Big]\Big|\le C(T,M)\Delta_s. \quad (4.59)$$

Finally, by (4.28), (4.55) and (4.59), we have

$$
\begin{aligned}
|I_N| &\leq \sum_{i=0}^{N-1} \int_{t_i}^{t_{i+1}} \Big| E\Big[(F(s) - F(t_i)) \frac{(\mu(s, Y(s)) - \mu(t_i, Y_i))(W_s - W_{t_i})}{\sigma(t_i, Y_i)(s - t_i)} \Big] \Big| \\
&\qquad + \Big| E\Big[\Big(F(s) - F(t_i) \Big) \frac{\partial \mu}{\partial x}(s, Y(s)) \Big] \Big| ds \\
&\leq \sum_{i=0}^{N-1} \int_{t_i}^{t_{i+1}} C(L, M, T)\Delta_s + C(T, M)\Delta_s ds \\
&\leq C(L, M, T) \sum_{i=0}^{N-1} \frac{1}{2}(t_{i+1} - t_i)^2 \\
&= C(L, M, T) \sum_{i=0}^{N-1} \frac{1}{2}\Delta^2 \leq C(L, M, T)\Delta.
\end{aligned} \tag{4.60}
$$

As for J_N, let us compare J_N to I_N first. I_N takes the form

$$
I_N = E \sum_{i=0}^{N-1} \int_{t_i}^{t_{i+1}} \Big[(\mu(s, Y(s)) - \mu(t_i, Y_i)) \frac{\partial f}{\partial x}(s, Y(s)) \Big] ds, \tag{4.61}
$$

and J_N takes the form

$$
\begin{aligned}
J_N &= E \sum_{i=0}^{N-1} \int_{t_i}^{t_{i+1}} \Big[\frac{1}{2}(\sigma^2(s, Y(s)) - \sigma^2(t_i, Y_i)) \frac{\partial^2 f}{\partial x^2}(s, Y(s)) \Big] ds \\
&= E \sum_{i=0}^{N-1} \int_{t_i}^{t_{i+1}} \Big[\frac{1}{2}(\sigma^2(s, Y(s)) - \sigma^2(t_i, Y_i)) \frac{\partial}{\partial x}\frac{\partial f}{\partial x}(s, Y(s)) \Big] ds.
\end{aligned} \tag{4.62}
$$

Therefore, if $\sigma^2(t, x)$ satisfies the same conditions as $\mu(t, x)$ does and $\frac{\partial f}{\partial x}(t, x)$ satisfies the same conditions as $f(t, x)$ does, then J_N should have the similar estimates as I_N has, as shown in (4.60). Recall that our assumptions in Theorem 4.7 state that $\mu(t, x) \in C_b^2([0, T] \times \mathbb{R})$, $\sigma^2(t, x) \in C_b^2([0, T] \times \mathbb{R})$ and $g(x) \in C_p^3(\mathbb{R})$. Therefore, similar to (4.60), J_N also has the following estimate

$$
|J_N| = \Big| E \sum_{i=0}^{N-1} \int_{t_i}^{t_{i+1}} \Big[\frac{1}{2}(\sigma^2(s, Y(s)) - \sigma^2(t_i, Y_i)) \frac{\partial^2 f}{\partial x^2}(s, Y(s)) \Big] ds \Big|
$$

$$\leq C(L, M, T)\Delta. \tag{4.63}$$

Then, by (4.23), (4.60) and (4.63), it is straightforward that

$$\begin{aligned}
&|E[g(X(T))] - E[g(Y(T))]| \\
&= \Big| E \sum_{i=0}^{N-1} \int_{t_i}^{t_{i+1}} \Big[(\mu(s, Y(s)) - \mu(t_i, Y_i)) \frac{\partial f}{\partial x}(s, Y(s)) \Big] ds \\
&\quad + E \sum_{i=0}^{N-1} \int_{t_i}^{t_{i+1}} \Big[\frac{1}{2}(\sigma^2(s, Y(s)) - \sigma^2(t_i, Y_i)) \frac{\partial^2 f}{\partial x^2}(s, Y(s)) \Big] ds \Big| \\
&= |I_N + J_N| \leq |I_N| + |J_N| \\
&\leq C(L, T, M)\Delta. && (4.64)
\end{aligned}$$

□

Chapter 5

Variance Reduction and Multilevel Monte Carlo Methods

In this chapter, we introduce the variance reduction techniques in Monte Carlo (MC) sampling methods. The standard Monte Carlo method converges very slowly. By performing variance reduction, we can reduce the number of operations or computing complexity and keep the desired level of accuracy at the same time. In 2008, Giles [16] introduced a new variance reduction method to the computing of the expectation of functionals of solutions of stochastic differential equations, called Multilevel Monte Carlo (MLMC) , so that the computing complexity can be reduced considerably. After that, this method has been applied extensively in many areas of stochastic simulation involving the Monte Carlo sampling. For a comprehensive introduction of the recent progress of Multilevel Monte Carlo methods, we refer to the webpage edited by Giles himself: https://people.maths.ox.ac.uk/gilesm/mlmc_community.html.

5.1 Monte Carlo Methods

In general, the purpose of Monte Carlo simulation is to estimate the quantity $E[g(X)]$, where X is a random variable and $g(x)$ is a deterministic function. The common examples are like the statistical moments of X, or the expected payoff of a financial derivative such as option. The MC estimator takes the form

$$E[g(X)] \approx \frac{1}{N}\sum_{i=1}^{N} g(X(\omega_i)), \tag{5.1}$$

where $\omega_i, i = 1, 2, \cdots, N$ are independent samples of the random variable X, and N is the total number of samples. Note that the MC estimator $\frac{1}{N}\sum_{i=1}^{N} g(X(\omega_i))$ is a random variable itself and that

$$E[g(X)] = E\Big[\frac{1}{N}\sum_{i=1}^{N} g(X(\omega_i))\Big].$$

The error of the MC estimation is defined as

$$\epsilon(g, N) := E[g(X)] - \frac{1}{N}\sum_{i=1}^{N} g(X(\omega_i)).$$

Again, $\epsilon(g, N)$ is also a random variable and obviously $E[\epsilon(g, N)] = 0$. In the following, we denote by $V[X]$ the variance of the random variable X. By the Law of Large Numbers and the Central Limit Theorem, the MC estimator (5.1) converges to the expectation $E[g(X)]$ when $N \to \infty$ and the estimation error decays as

$$\sqrt{V[\epsilon(g, N)]} \approx \sqrt{\frac{V[g(X)]}{N}}.$$

There are two parts in the estimation error: the variance of $g(X)$ and the total number of samples N. Suppose $V[g(X)] < \infty$ is a fixed constant, then the MC estimation error decays in the order $O(N^{-1/2})$. This is a rather slow decay. For example, given $V[g(X)] < \infty$ fixed, if we want to reduce the estimation error by a factor of 10, we need to increase the number of samples by a factor of 100. Therefore, to accelerate the convergence of MC estimation by increasing the number of samples is very computationally expensive. But we can also look at this problem from the

$V[g(X)]$ perspective. Suppose the desired error of the MC estimation is at most ε, then the number of samples to reach such an accuracy should be at least

$$N \approx \frac{V[g(X)]}{\varepsilon^2}.$$

To achieve a desired error ε, if we can reduce the standard deviation of $g(X)$ by a factor of 10 (or reduce the variance of $g(X)$ by a factor of 100), the required number of samples can be reduced by a factor of 100. Therefore, variance reduction is a very effective way to accelerate the convergence of MC estimation. There are many well-established variance reduction techniques, such as antithetic variables, importance sampling, control variates, etc.. We refer to [20, 47, 51] for a thorough introduction of these methods. In the following, we only introduce the control variates method, due to its close relation to the Multilevel Monte Carlo method.

Control Variates. Suppose $h(X)$ is another function of X with $E[h(X)]$ known. Then $E[g(X)]$ can be written as

$$E[g(X)] = E[h(X)] + E[g(X) - h(X)].$$

Since $E[h(X)]$ is known, in order to estimate $E[g(X)]$, we only need to estimate the expectation $E[g(X) - h(X)]$. The corresponding estimator takes the form

$$E[g(X) - h(X)] \approx \frac{1}{N}\sum_{i=1}^{N}[g(X(\omega_i)) - h(X(\omega_i))].$$

By our analysis above, this estimation error decays as

$$\sqrt{\frac{V[g(X) - h(X)]}{N}}.$$

If we can choose such a function $h(x)$ that $h(X)$ fluctuates around its mean in a similar way as $g(X)$ does, we can expect the variance of $g(X) - h(X)$ is much smaller than the variance of $g(X)$. By doing so, the required number of samples can be reduced considerably given a certain level of error tolerance. For example, for a fixed error tolerance ε, if the standard deviation of $g(X) - h(X)$ is only $\frac{1}{10}$ of the standard deviation of $g(X)$, then the required number of samples to estimate $E[g(X) - h(X)]$ is only $\frac{1}{100}$ of the required number of samples to estimate $E[g(X)]$. We conclude this part by remarking that $E[h(X)]$ does not need to be close to $E[g(X)]$. All we need is that $h(X)$ has a similar variability as $g(X)$.

5.2 Monte Carlo Methods and the Simulation of Stochastic Differential Equations

In this section, we still consider the m-dimensional SDE

$$\begin{cases} dX(t) = \mu(t, X(t))dt + \sigma(t, X(t))dW(t), \quad t \in (0, T] \\ X(0) = x_0. \end{cases} \tag{5.2}$$

The setting of this equation is as usual. For now, we only assume there exists a unique solution to (5.2). As we discussed before, people are often motivated to compute the expected value of $f(X(T))$, where f is a scalar function satisfying some smooth conditions. For example, for an European call option, its payoff is equal to $f(X(T)) = \max(X(T) - K, 0)$, where $X(T)$ is the terminal price of the underlying stock and K is the exercise price of the option at expiration.

Suppose a numerical scheme with time step h solves the equation (5.2) numerically. We denote by $\widehat{X}_{T/h}$ the value of the numerical solution at time T. Suppose we have generated N independent path simulations by the numerical scheme. Then the Monte Carlo estimate of the expectation $E[f(X(T))]$ is

$$E[f(X(T))] \approx \frac{1}{N} \sum_{i=1}^{N} f(\widehat{X}^{(i)}_{T/h}),$$

where $\widehat{X}^{(i)}_{T/h}$ is the terminal value of the numerical solution on the ith sample path. Denote by $Y := E[f(X(T))]$ and $\widehat{Y} := \frac{1}{N}\sum_{i=1}^{N} f(\widehat{X}^{(i)}_{T/h})$. Then the mean square error of the MC estimate is

$$MSE \equiv E\big[(\widehat{Y} - E[Y])^2\big].$$

We decompose the MSE in the following way:

$$\begin{aligned}
&E\big[(\widehat{Y} - E[Y])^2\big] \\
&= E\big[(\widehat{Y} - E[\widehat{Y}] + E[\widehat{Y}] - E[Y])^2\big] \\
&= E\big[(\widehat{Y} - E[\widehat{Y}])^2\big] + \big(E[\widehat{Y}] - E[Y]\big)^2 + 2E\big[(\widehat{Y} - E[\widehat{Y}])(E[\widehat{Y}] - E[Y])\big] \\
&= E\big[(\widehat{Y} - E[\widehat{Y}])^2\big] + \big(E[\widehat{Y}] - E[Y]\big)^2
\end{aligned}$$

The first part is the variance of the MC estimate and the other part is the bias introduced by the approximation. Due to the mutual independence of the sample paths, the variance of $\widehat{Y}$ is

$$\begin{aligned}
V[\widehat{Y}] &= V\Big[\frac{1}{N}\sum_{i=1}^{N} f(\widehat{X}^{(i)}_{T/h})\Big] \\
&= \frac{1}{N^2} V\Big[\sum_{i=1}^{N} f(\widehat{X}^{(i)}_{T/h})\Big] \\
&= \frac{1}{N} V[f(X(T))].
\end{aligned}$$

As we discussed in Section 2.3, if we use Euler-Maruyama scheme to solve equation (5.2), as long as μ, σ and f satisfy certain smooth conditions, the weak convergence has order 1. In this case,

$$\big(E[\widehat{Y}] - E[Y]\big)^2 = O(h^2).$$

Therefore, with Euler-Maruyama scheme, the MSE of the MC estimate $\widehat{Y}$ is asymptotically

$$MSE \approx O(N^{-1}) + O(h^2).$$

To make the root mean square error (RMSE=$\sqrt{\text{MSE}}$) $O(\varepsilon)$, the MSE should be $O(\varepsilon^2)$. So it requires that $N = O(\varepsilon^{-2})$ and $h = O(\varepsilon)$. The total computational cost is then $O(\varepsilon^{-3})$.

In general, if a numerical solution of equation (5.2) has week convergence order k, then to achieve the RSME $O(\varepsilon)$, the total computational cost of computing $E[f(X(T))]$ is $O(\varepsilon^{-(2+1/k)})$.

For a general introduction to the variance reduction methods which can be applied to the computation of functionals of solutions of (5.2), we refer to Chapter 16 of [33].

5.3 Multilevel Monte Carlo Methods

Multilevel Monte Carlo (MLMC) method is a control variate type variance reduction approach. As we discussed earlier, to compute $E[f(X)]$ and use the control variate method, we need to choose another function g such that $E[g(X)]$ is known and $g(X)$ has a similar variability as $f(X)$. We know that the solution of a SDE usually does not have a closed-form expression, let alone the existence of an available control variate of it. However, if we do not ask too much about the availability of the value of $E[g(X(T))]$, a control variate is not so difficult to find. For example, for a given Brownian path $W(t)$, let P denote the random variable $f(X(T))$ and $\widehat{P}_h$ denote the approximation to P using a numerical discretisation with time step h. Using the same Brownian path and a larger step size $2h$, we can expect that the new approximation $\widehat{P}_{2h}$ would not fluctuate too differently from $\widehat{P}_h$. Therefore, we can use $\widehat{P}_{2h}$ as a control variate of $\widehat{P}_h$. If we use more levels of time steps, it will give us a multilevel Monte Carlo method.

Consider a SDE like (5.2) and the Monte Carlo path simulations with a geometric series of time steps $h_l = M^{-l}T$, where $l = 0, 1, \cdots, L$ and $M \geq 2$ is an integer. In level 0, there is only one time step for the whole time interval $[0, T]$, whereas level $l \geq 1$ has M^l uniform time steps. We denote by P the random variable $f(X(T))$.

Let $W(t)$ be a given Brownian path, and let $\widehat{X}_l$ and $\widehat{P}_l$ denote the approximations to $X(T)$ and P using a numerical scheme with time step h_l.

Our goal is to estimate the quantity $E[f(X)]$. In general, the weak approximation error decreases as the step size of the discretisation decreases [5, 33, 60]. Then it is natural to write

$$E[\widehat{P}_L] = E[\widehat{P}_0] + \sum_{l=1}^{L} E[\widehat{P}_l - \widehat{P}_{l-1}].$$

Therefore, to give $E[f(X)]$ an estimate, the simplest way is to estimate the expectations on the right hand side of the above equality using a standard Monte Carlo estimator. Note that we may need to use different numbers of sample paths to give the Monte Carlo estimates of those expectations. Denote by $\widehat{Y}_0$ the estimator of $E[\widehat{P}_0]$ using N_0 sample paths and $\widehat{Y}_l$ the estimator of $E[\widehat{P}_l - \widehat{P}_{l-1}]$ using N_l sample paths. For $l = 0$,

$$\widehat{Y}_0 = \frac{1}{N_0} \sum_{i=1}^{N_0} \widehat{P}_0^{(i)}.$$

For $l \geq 1$,

$$\widehat{Y}_l = \frac{1}{N_l} \sum_{i=1}^{N_l} (\widehat{P}_l^{(i)} - \widehat{P}_{l-1}^{(i)}),$$

where both $\widehat{P}_l^{(i)}$ and $\widehat{P}_{l-1}^{(i)}$ are obtained from the ith Brownian path. Note that $\widehat{P}_l^{(i)}$ and $\widehat{P}_{l-1}^{(i)}$ use the same Brownian path, although they come from two approximations with different time steps. In our setting at the beginning, we initiated a geometric series of time steps. The reason of doing this is now clear. For two approximations with consecutive levels of time steps, $\widehat{P}_l^{(i)}$ and $\widehat{P}_{l-1}^{(i)}$, we first evaluate $\widehat{P}_l^{(i)}$ using M^l Brownian increments. Then we sum those increments in groups of size M to generate M^{l-1} Brownian increments in preparation for the computation of $\widehat{P}_{l-1}^{(i)}$. Obviously,

$$E[\widehat{Y}_0] = E[\widehat{P}_0], \quad E[\widehat{Y}_l] = E[\widehat{P}_l - \widehat{P}_{l-1}], \ l \geq 1,$$

and

$$V[\widehat{Y}_0] = N_0^{-1} V_0, \quad V[\widehat{Y}_l] = N_l^{-1} V_l, \ l \geq 1,$$

where V_0 and V_l are the variance of $\widehat{P}_0$ and $\widehat{P}_l - \widehat{P}_{l-1}$, respectively.

Finally, the estimator of $E[f(X)]$ is

$$\widehat{Y} = \sum_{l=0}^{L} \widehat{Y}_l.$$

By our previous analysis,

$$E[\widehat{Y}] = E[\widehat{P}_L], \quad V[\widehat{Y}] = \sum_{l=0}^{L} N_l^{-1} V_l,$$

and the total computational cost is asymptotically proportional to

$$\sum_{l=1}^{L} N_l h_l^{-1}.$$

As usual, the MSE of the MLMC estimator can be decomposed as

$$\begin{aligned} MSE &= E\big[(\widehat{Y} - E[P])^2\big] \\ &= E\big[(\widehat{Y} - E[\widehat{Y}] + E[\widehat{Y}] - E[P])^2\big] \\ &= E\big[(\widehat{Y} - E[\widehat{Y}])^2\big] + (E[\widehat{Y}] - E[P])^2 \\ &= E\big[(\widehat{Y} - E[\widehat{Y}])^2\big] + (E[\widehat{P}_L] - E[P])^2. \end{aligned}$$

The first part is the variance of the MLMC estimator and

$$E\big[(\widehat{Y} - E[\widehat{Y}])^2\big] = V[\widehat{Y}] = \sum_{l=0}^{L} N_l^{-1} V_l.$$

The second part is the square of the weak approximation error of the scheme with smallest step size $M^{-l}T$. Given an error tolerance ε^2 on MSE, we need to determine the values of N_l and L such that

$$E\big[(\widehat{Y} - E[\widehat{Y}])^2\big] \leq \frac{1}{2}\varepsilon^2 \tag{5.3}$$

and

$$(E[\widehat{P}_L] - E[P])^2 \leq \frac{1}{2}\varepsilon^2. \tag{5.4}$$

Hopefully, after N_l and L are determined, the total computational cost can be reduced considerably compared with using the standard MC method.

Inequality (5.4) is typically a weak convergence result. Suppose for a well-established numerical scheme, there exist a positive real number k and a positive constant C independent of the time step h_L, such that

$$|E[\widehat{P}_L] - E[P]| \leq Ch_L^k.$$

Since $h_L = M^{-L}T$, we must have

$$C^2T^{2k}M^{-2kL} \leq \frac{1}{2}\varepsilon^2,$$

which leads to

$$L \geq \frac{1}{k}\log_M(\sqrt{2}CT^k\varepsilon^{-1}) = \frac{\log(\sqrt{2}CT^k\varepsilon^{-1})}{k\log M}.$$

Since L is also an integer, we can take

$$L_{max} = \left\lceil \frac{\log(\sqrt{2}CT^k\varepsilon^{-1})}{k\log M} \right\rceil, \tag{5.5}$$

where $\lceil x \rceil$ is the smallest integer not less than x.

Usually, like in our previous chapters, we often use $\Delta t = T/N$ as our step size, where N is the total number of steps. In this case, the number of L has a very simple expression. In fact, we have $M^{-L}T = T/N$, which leads to

$$L = \frac{\log N}{\log M}.$$

Our next step is to choose N_l to make (5.3) hold and to minimize the variance $V[\widehat{Y}]$ for the fixed computational cost $C = \sum_{l=0}^{L} N_l h_l^{-1}$ at the same time. Actually, we can treat N_l as continuous variables and use the Lagrange multiplier to find the minimum of

$$\sum_{l=0}^{L} N_l^{-1} V_l + \lambda\Big(\sum_{l=1}^{L} N_l h_l^{-1} - C \Big).$$

Taking partial derivative of the above expression with respect to each N_l and setting them to be 0, we get

$$-N_l^{-2} V_l + \lambda h_l^{-1} = 0,$$

which gives

$$N_l = \lambda^{-1/2} \sqrt{V_l h_l}.$$

Therefore,

$$V[\widehat{Y}] = \sum_{l=0}^{L} N_l^{-1} V_l = \sum_{l=0}^{L} \frac{\sqrt{\lambda} V_l}{\sqrt{V_l h_l}} = \sqrt{\lambda} \sum_{l=0}^{L} \sqrt{V_l / h_l}.$$

Note that we also need $V[\widehat{Y}] \leq \frac{\varepsilon^2}{2}$. Therefore,

$$\lambda^{-1/2} \geq 2\varepsilon^{-2} \sum_{l=0}^{L} \sqrt{V_l / h_l}$$

and the optimal number of sample paths for level l is

$$N_l = \Big\lceil 2\varepsilon^{-2} \sqrt{V_l h_l} \sum_{l=0}^{L} \sqrt{V_l / h_l} \Big\rceil.$$

So the total computational cost will be approximately

$$\sum_{l=0}^{L} N_l h_l^{-1} \approx 2\varepsilon^{-2} \Big(\sum_{l=0}^{L} \sqrt{V_l / h_l} \Big)^2. \tag{5.6}$$

To have a rough idea what the computational complexity in (5.6) looks like, we take the Euler-Maruyama scheme for example. Here we assume both μ and σ depend only on x and they are infinitely many times differentiable with all of their derivatives bounded. We also assume that the function $f(x)$ is Lipschitz continuous. Assuming also some other conditions, Bally and Tally [5] showed that the Euler-Maruyama approximate solution converges to the real solution weakly with order 1. Meanwhile, by Theorem 2.9, the Euler-Maruyama scheme also has strong convergence with order 1/2. Hence, when l is large enough,

$$E[\widehat{P}_l - P] = E[f(\widehat{X}_l) - f(X(T))] = O(h_l), \tag{5.7}$$

and

$$E[\|\widehat{X}_l - X(T)\|^2] = O(h_l). \tag{5.8}$$

Since $f(x)$ is Lipschitz continuous, we also have

$$\begin{aligned} V[\widehat{P}_l - P] &\leq E[(\widehat{P}_l - P)^2] \\ &= E[(f(\widehat{X}_l) - f(X(T)))^2] \\ &\leq C^2 E[\|\widehat{X}_l - X(T)\|^2], \end{aligned}$$

where C is the Lipschitz constant. Hence, by (5.8),

$$V[\widehat{P}_l - P] = O(h_l). \tag{5.9}$$

What is more,

$$\begin{aligned} V_l &= V[\widehat{P}_l - \widehat{P}_{l-1}] \\ &= V[(\widehat{P}_l - P) - (\widehat{P}_{l-1} - P)] \\ &\leq 2(V[\widehat{P}_l - P] + V[\widehat{P}_{l-1} - P]). \end{aligned}$$

Combining this with (5.9) gives that

$$V_l = O(h_l).$$

Getting V_l back to the computational cost (5.6), we have

$$2\varepsilon^{-2}\Big(\sum_{l=0}^{L}\sqrt{V_l/h_l}\Big)^2 = O(\varepsilon^{-2}L^2).$$

Note that, by (5.5), for Euler-Maruyama scheme ($k=1$), $L = O(\log(\varepsilon^{-1}))$ and thus the total computational cost is

$$O\big(\varepsilon^{-2}(\log\varepsilon)^2\big).$$

To compare this computational cost by MLMC method with the standard MC method whose cost is $O(\varepsilon^{-3})$, we take $\varepsilon = 0.001$. Then $(\log\varepsilon)^2 \approx 48$, which is much less than $\varepsilon^{-1} = 1000$.

From the above computation, we can see that it is very crucial for us to figure out what $V_l = V[\widehat{P}_l - \widehat{P}_{l-1}]$ is. To achieve this goal, in the Euler-Maruyama case, we used the fact that it has $O(h^{1/2})$ strong convergence and that $f(x)$ is Lipschitz continuous. It is very interesting to point out that, even if we are looking at a weak approximation problem (computing $E[f(X(T))]$) here, the key point is to use the strong convergence of the underlying numerical scheme. Furthermore, the only place where the weak convergence is used is when we use it to determine the value of L. Recall that

$$L \geq \frac{\log(\sqrt{2}CT^k\varepsilon^{-1})}{k\log M},$$

where k is the weak convergence order. It is obvious that the weak convergence order k is not so important in determining the value of L, compared to $\log(\varepsilon^{-1})$. To be more specific, in the Euler-Maruyama scheme, suppose for now we do not know the weak convergence order of the scheme but know the strong convergence order 1/2 instead. If the test function $f(x)$ is Lipschitz continuous, it is immediate that

$$\big|E[f(\widehat{X}_l) - f(X(T))]\big| \leq CE[\|\widehat{X}_l - X(T)\|] = O(h_l^{1/2}).$$

Taking $k = 1/2$, it follows that

$$L \geq \frac{2\log(\sqrt{2}CT^{1/2}\varepsilon^{-1})}{\log M}, \quad k = 1/2$$

which is not so far away from it is when $k = 1$:

$$L \geq \frac{\log(\sqrt{2}CT\varepsilon^{-1})}{\log M}, \quad k = 1.$$

See also [19] for examples in which the non-Lipschitz continuous test functions are involved.

Another disadvantage of using the weak convergence is that it requires too much on the smoothness of μ, σ and $f(x)$ (see our remarks in Section 2.3 for more details). While the strong convergence often does not ask too much on those things (see e.g. Theorem 2.9 or our main theorems in Chapter 3). Because of these two points, there has been a great motivation of developing strongly convergent numerical methods for SDE (5.2) [23, 26, 28, 32, 55, 61].

Next we give the complexity theorem of the MLMC method. It is stated in a very general way such that it can be applied to various cases. This theorem does not specify which numerical scheme is used either.

Theorem 5.1 ([16], Theorem 3.1)**.** *Let P denote a functional of the solution of stochastic differential equations (5.2) for a given Brownian path $W(t)$, and let $\widehat{P}_l$ denote the corresponding approximation using a numerical discretisation with time step $h_l = M^{-l}T$. If there exist independent estimators $\widehat{Y}_l$ based on N_l Monte Carlo samples, and positive constants $\alpha \geq \frac{1}{2}, \beta, c_1, c_2, c_3$ such that*

i) $|E[\widehat{P}_l - P]| \leq c_1 h_l^{\alpha}$;

ii) $E[\widehat{Y}_l] = \begin{cases} E[\widehat{P}_0], & l = 0, \\ E[\widehat{P}_l - \widehat{P}_{l-1}], & l \geq 0; \end{cases}$

iii) $V[\widehat{Y}_l] \leq c_2 N_l^{-1} h_l^{\beta}$;

iv) C_l*, the computational complexity of* $\widehat{Y}_l$*, is bounded by*

$$C_l \leq c_3 N_l h_l^{-1},$$

then there exists a positive constant c_4 *such that for any* $\varepsilon < e^{-1}$ *there are values* L *and* N_l *for which the multilevel estimator*

$$\widehat{Y} = \sum_{l=0}^{L} \widehat{Y}_l,$$

has a mean-square-error with bound

$$MSE \equiv E\left[(\widehat{Y} - E[P])^2\right] < \varepsilon^2$$

with a computational complexity C *with bound*

$$C \leq \begin{cases} c_4 \varepsilon^{-2}, & \beta > 1, \\ c_4 \varepsilon^{-2} (\log \varepsilon)^2, & \beta = 1, \\ c_4 \varepsilon^{-2-(1-\beta)/\alpha}, & 0 < \beta < 1. \end{cases}$$

Proof. We start by choosing

$$L = \left\lceil \frac{\log(\sqrt{2} c_1 T^\alpha \varepsilon^{-1})}{\alpha \log M} \right\rceil,$$

so that

$$\frac{1}{\sqrt{2}} M^{-\alpha} \varepsilon < c_1 h_L^\alpha \leq \frac{1}{\sqrt{2}} \varepsilon, \tag{5.10}$$

and hence, because of properties *i)* and *ii)*,

$$\left(E[\widehat{Y}] - E[P]\right)^2 \leq \frac{1}{2} \varepsilon^2.$$

This $\frac{1}{2}\varepsilon^2$ upper bound on the square of the bias error, together with the $\frac{1}{2}\varepsilon^2$ upper bound on the variance of the estimator to be proved later, gives an ε^2 upper bound on the estimator MSE.

Also,

$$\sum_{l=0}^{L} h_l^{-1} = h_L^{-1} \sum_{l=0}^{L} M^{-l} < \frac{M}{M-1} h_L^{-1}.$$

By (5.10), we also have

$$h_L^{-1} < M\left(\frac{\varepsilon}{\sqrt{2}c_1}\right)^{-1/\alpha}.$$

Given $\alpha \geq \frac{1}{2}$ and $\varepsilon < e^{-1}$, it follows that

$$\varepsilon^{-1/\alpha} \leq \varepsilon^{-2}.$$

Then the above three inequalities give that

$$\sum_{l=0}^{L} h_l^{-1} < \frac{M^2}{M-1}\left(\sqrt{2}c_1\right)^{1/\alpha} \varepsilon^{-2}. \tag{5.11}$$

Now we consider the different possible values for β.

a) If $\beta = 1$, we set $N_l = \lceil 2\varepsilon^{-2}(L+1)c_2 h_l \rceil$ so that

$$V[\widehat{Y}] = \sum_{l=0}^{L} V[\widehat{Y}_l] \leq \sum_{l=0}^{L} c_2 N_l^{-1} h_l \leq \frac{1}{2}\varepsilon^2,$$

which is the required upper bound on the variance of the estimator.

To bound the computational complexity C we begin with an upper bound on L given by

$$L \leq \frac{\log \varepsilon^{-1}}{\alpha \log M} + \frac{\log(\sqrt{2}c_1 T^{\alpha})}{\alpha \log M} + 1.$$

Given that $1 < \log \varepsilon^{-1}$ for $\varepsilon < e^{-1}$, it follows that

$$L + 1 \leq c_5 \log \varepsilon^{-1},$$

where

$$c_5 = \frac{1}{\alpha \log M} + \max\Big(0, \frac{\log(\sqrt{2}c_1 T^\alpha)}{\alpha \log M}\Big) + 2.$$

Upper bounds for N_l are given by

$$N_l \leq 2\varepsilon^{-2}(L+1)c_2 h_l + 1.$$

Hence the computational complexity is bounded by

$$C \leq c_3 \sum_{l=1}^{L} N_l h_l^{-1} \leq c_3 \Big(2\varepsilon^{-2}(L+1)^2 c_2 + \sum_{l=0}^{L} h_l^{-1}\Big).$$

Using the upper bound for $L+1$ and inequality (5.11), and the fact that $1 < \log \varepsilon^{-1}$ for $\varepsilon < e^{-1}$, it follows that

$$C \leq c_4 \varepsilon^{-2} (\log \varepsilon)^2,$$

where

$$c_4 = 2c_3 c_5^2 c_2 + c_3 \frac{M^2}{M-1} \big(\sqrt{2}c_1\big)^{1/\alpha}.$$

b) For $\beta > 1$, setting

$$N_l = \Big\lceil 2\varepsilon^{-2} c_2 T^{(\beta-1)/2} \big(1 - M^{-(\beta-1)/2}\big)^{-1} h_l^{(\beta+1)/2} \Big\rceil,$$

then

$$V[\widehat{Y}] = \sum_{l=0}^{L} V[\widehat{Y}_l] \leq \frac{1}{2} \varepsilon^2 T^{-(\beta-1)/2} \big(1 - M^{-(\beta-1)/2}\big) \sum_{l=0}^{L} h_l^{(\beta-1)/2}.$$

Using the standard result for a geometric series,

$$\sum_{l=0}^{L} h_l^{(\beta-1)/2} = T^{(\beta-1)/2} \sum_{l=0}^{L} \Big(M^{-(\beta-1)/2}\Big)^l$$

$$< T^{(\beta-1)/2}\Big(1 - M^{-(\beta-1)/2}\Big)^{-1},$$

and hence we obtain an $\frac{1}{2}\varepsilon^2$ upper bound on the variance of the estimator. Using the N_l upper bound

$$N_l < 2\varepsilon^{-2} c_2 T^{(\beta-1)/2}\big(1 - M^{-(\beta-1)/2}\big)^{-1} h_l^{(\beta+1)/2} + 1,$$

the computational complexity is bounded by

$$C \le c_3\Big(2\varepsilon^{-2} c_2 T^{(\beta-1)/2}\big(1 - M^{-(\beta-1)/2}\big)^{-1} \sum_{l=0}^{L} h_l^{(\beta-1)/2} + \sum_{l=0}^{L} h_l^{-1}\Big).$$

Using inequalities (5.10) and (5.11) gives

$$C \le c_4 \varepsilon^{-2},$$

where

$$c_4 = 2c_3 c_2 T^{\beta-1}\big(1 - M^{-(\beta-1)/2}\big)^{-2} + c_3 \frac{M^2}{M-1}\big(\sqrt{2}c_1\big)^{1/\alpha}.$$

c) For $\beta < 1$, setting

$$N_l = \Big\lceil 2\varepsilon^{-2} c_2 h_L^{-(1-\beta)/2}\big(1 - M^{-(1-\beta)/2}\big)^{-1} h_l^{(\beta+1)/2}\Big\rceil,$$

then

$$V[\widehat{Y}] = \sum_{l=0}^{L} V[\widehat{Y}_l] < \frac{1}{2}\varepsilon^2 h_L^{(1-\beta)/2}\big(1 - M^{-(1-\beta)/2}\big) \sum_{l=0}^{L} h_l^{-(1-\beta)/2}.$$

Since

$$\begin{aligned} \sum_{l=0}^{L} h_l^{-(1-\beta)/2} &= h_L^{-(1-\beta)/2} \sum_{l=0}^{L} \Big(M^{-(1-\beta)/2}\Big)^l \\ &< h_L^{-(1-\beta)/2}\Big(1 - M^{-(1-\beta)/2}\Big)^{-1}, \end{aligned} \tag{5.12}$$

we obtain an $\frac{1}{2}\varepsilon^2$ upper bound on the variance of the estimator. Using the N_l upper bound

$$N_l < 2\varepsilon^{-2} c_2 h_L^{-(1-\beta)/2} \left(1 - M^{-(1-\beta)/2}\right)^{-1} h_l^{(\beta+1)/2} + 1,$$

the computational complexity is bounded by

$$C \leq c_3 \Big(2\varepsilon^{-2} c_2 h_L^{-(1-\beta)/2} \left(1 - M^{-(1-\beta)/2}\right)^{-1} \sum_{l=0}^{L} h_l^{-(1-\beta)/2} + \sum_{l=0}^{L} h_l^{-1} \Big).$$

Using inequality (5.12) gives

$$h_L^{-(1-\beta)/2} \left(1 - M^{-(1-\beta)/2}\right)^{-1} \sum_{l=0}^{L} h_l^{-(1-\beta)/2} < h_L^{-(1-\beta)} \Big(1 - M^{-(1-\beta)/2}\Big)^{-2}.$$

The first inequality in (5.10) gives

$$h_L^{-(1-\beta)} < \left(\sqrt{2} c_1\right)^{(1-\beta)/\alpha} M^{1-\beta} \varepsilon^{-(1-\beta)/\alpha}.$$

Combining the above two inequalities, and also using inequality (5.11) and the fact that $\varepsilon^{-2} < \varepsilon^{-2-(1-\beta)/\alpha}$ for $\varepsilon < e^{-1}$, gives

$$C \leq c_4 \varepsilon^{-2-(1-\beta)/\alpha},$$

where

$$c_4 = 2 c_3 c_2 \left(\sqrt{2} c_1\right)^{(1-\beta)/\alpha} M^{1-\beta} \left(1 - M^{-(1-\beta)/2}\right)^{-2} + c_3 \frac{M^2}{M-1} \left(\sqrt{2} c_1\right)^{1/\alpha}.$$

□

Remark 5.2. *Condition i) often depends on the weak convergence results of numerical approximations of SDEs. We refer to Section 2 and 4 for more details of such things. The main challenge is condition iii). We have shown that, with a Lipschitz test* *function, we can use the strong convergence results to obtain the variance estimation. But, in general, this task may not be so trivial. We also refer to [16, 17, 18] for more remarks.*

The following theorem is given in Giles [18], and is a slight generalization of Theorem 5.1.

Theorem 5.3 ([18], Theorem 2.1)**.** *Let P denote a functional of the solution of a stochastic differential equation, and let $\widehat{P}_l$ denote the corresponding level l numerical approximation. If there exist independent estimators $\widehat{Y}_l$ based on N_l Monte Carlo samples, and positive constants $\alpha, \beta, \gamma, c_1, c_2, c_3$ such that $\alpha \geq \frac{1}{2}\min(\beta, \gamma)$ and*

i) $|E[\widehat{P}_l - P]| \leq c_1 2^{-\alpha l}$;

ii) $E[\widehat{Y}_l] = \begin{cases} E[\widehat{P}_0], & l = 0, \\ E[\widehat{P}_l - \widehat{P}_{l-1}], & l > 0; \end{cases}$

iii) $V[\widehat{Y}_l] \leq c_2 N_l^{-1} 2^{-\beta l}$;

iv) $C_l \leq c_3 N_l 2^{\gamma l}$, *where C_l is the computational complexity of Y_l,*

then there exists a positive constant c_4 such that for any $\varepsilon < e^{-1}$ there are values of L and N_l for which the multilevel estimator

$$\widehat{Y} = \sum_{l=0}^{L} \widehat{Y}_l,$$

has a mean-square-error with bound

$$MSE \equiv E[(\widehat{Y} - E[P])^2] < \varepsilon^2$$

with a computational complexity C with bound

$$C \leq \begin{cases} c_4 \varepsilon^{-2}, & \beta > \gamma, \\ c_4 \varepsilon^{-2} (\log \varepsilon)^2, & \beta = \gamma, \\ c_4 \varepsilon^{-2-(\gamma-\beta)/\alpha}, & 0 < \beta < \gamma. \end{cases}$$

This theorem specifies the value of M to be 2. Of course, M can also be other natural numbers. See e.g. [16] on how to choose an optimal value for M. Since the computational complexity on level l is changed to be dependent on γ, the new classification of the total computational cost is based on the comparison between β and γ accordingly, rather than between β and 1.

In [17], Giles provided another generalization of the complexity theorem based on Cliffe, Giles, Scheichl, and Teckentrup [9], which allows for applications in which the simulation cost of individual samples is itself random. See [17] for more details.

For MLMC for SDEs driven by jump processes, we refer to Dereich [12] and Xia [62] and references therein. For MLMC for stochastic partial differential equations, see e.g. [6] and [9].

As in the classic Monte Carlo simulation, there is also a central limit theorem corresponding the multilevel Monte Carlo Euler simulation. We refer to [1] for the details. We also refer to [13] for the central limit theorem of the multilevel Monte Carlo Euler simulation for SDEs driven by Lévy noise.

Chapter 6

Strong Convergence of Numerical Approximations of SDEs Driven by Lévy Noise under Local Lipschitz Conditions

In this chapter, we introduce the strong convergence of numerical approximations of SDEs driven by Lévy noise under local Lipschitz condtions. As far as the author knows, the literature on this topic is very sparse. [11, 36] might be among the first few papers addressing this issue and there is still much space to be explored.

6.1 Strong Convergence of Tamed Euler Approximations of SDEs Driven by Lévy Noise with Superlinearly Growing Drift Coefficients

In this section, we mainly consider the following SDE driven by a Lévy noise:

$$\begin{cases} dX(t) = a(X(t-))dt + b(X(t-))dW(t) + \int_{\mathbb{R}^d} f(X(t-), y)\tilde{N}(dt, dy), \\ X(0) = x_0, \end{cases} \tag{6.1}$$

or, equivalently, the following integral form:

$$X(t) = X(0) + \int_0^t a(X(s-))ds + \int_0^t b(X(s-))dW(s) + \int_0^t \int_{\mathbb{R}^d} f(X(s-), y)\tilde{N}(ds, dy), \tag{6.2}$$

where x_0 is $\mathcal{F}_0$-measurable, $X(t)$ is an $\mathbb{R}^m$-valued stochastic process, $W(t)$ is a standard d-dimensional Brownian motion and N an independent Poisson random measure on $\mathbb{R}^+ \times (\mathbb{R}^d - \{0\})$ with associated compensator $\tilde{N}$ and intensity measure ν, where we assume that ν is a Lévy measure. Besides, W and N are independent of $\mathcal{F}_0$. We also assume that $a : \mathbb{R}^m \to \mathbb{R}^m$, $b : \mathbb{R}^m \to \mathbb{R}^{m\times d}$ and $f : \mathbb{R}^m \times \mathbb{R}^d \to \mathbb{R}^m$. In the following, we will write $X(t)$ instead of $X(t-)$ on the right hand side of the above equation. This will not cause any problem since the compensators of the martingales driving the equation are continuous.

We assume the following conditions for the above SDE:

A-1. There exists a constant $C > 0$ such that for any $x \in \mathbb{R}^m$,

$$\langle x, a(x)\rangle + \|b(x)\|^2 + \int_{\mathbb{R}^d} \|f(x, y)\|^2 \nu(dy) \le C(1 + \|x\|^2).$$

A-2. There exists a constant $C > 0$ such that for $x_1, x_2 \in \mathbb{R}^m$,

$$\langle x_1 - x_2, a(x_1) - a(x_2)\rangle + \|b(x_1) - b(x_2)\|^2 + \int_{\mathbb{R}^d} \|f(x_1, y) - f(x_2, y)\|^2 \nu(dy)$$
$$\leq C\|x_1 - x_2\|^2.$$

A-3. The function $a(x)$ is continuous in $x \in \mathbb{R}^m$.

A-4. $E[\|x_0\|^p] < \infty$ for some fixed $p \geq 2$.

A-5. There exists a constant $C > 0$ such that for any $x \in \mathbb{R}^m$,

$$\int_{\mathbb{R}^d} \|f(x, y)\|^p \nu(dy) \leq C(1 + \|x\|^p).$$

Note that if condition A-3 is fulfilled, then it is obvious that $a(x)$ is locally bounded. That is, for any $R > 0$, there exits a constant $C(R) > 0$, such that

$$a(x) \leq C(R) \tag{6.3}$$

for any $\|x\| \leq C(R)$. As we introduced in Section 2.4, if SDE (6.2) satisfies conditions A-1–A-3, there exists a unique solution to (6.2). If, in addition, conditions A-4 and A-5 are also satisfied, the following p-th moment bound holds for the solution:

$$E\Big[\sup_{0\leq t\leq T} \|X(t)\|^p\Big] \leq C.$$

In the following, we also assume

A-6. There exist constants $C > 0$, $q \geq 2$ and $\chi > 0$ such that

$$\int_{\mathbb{R}^d} \|f(x_1, y) - f(x_2, y)\|^q \nu(dy) \leq C\|x_1 - x_2\|^q,$$

$$\|a(x_1) - a(x_2)\|^2 \leq C(1 + \|x_1\|^\chi + \|x_2\|^\chi)\|x_1 - x_2\|^2, \tag{6.4}$$

for any $x_1, x_2 \in \mathbb{R}^m$ and a $\delta \in (0,1)$ such that $\max\{(\chi+2)q, \frac{q\chi}{2}\frac{q+\delta}{\delta}\} \leq p$. Note that for now $p \geq (\chi+2)q > 4$. By (6.3) and (6.4), it is immediate that

$$\|a(x)\|^2 \leq C(1+\|x\|^{\chi+2}) \tag{6.5}$$

for any $x \in \mathbb{R}^m$.

To solve (6.2) numerically, Dareiotis, Kumar and Sabanis [11] proposed a tamed Euler scheme defined by

$$dX_N(t) = a_N(X_N(\kappa_N(t)))dt + b(X_N(\kappa_N(t)))dW(t) + \int_{\mathbb{R}^d} f(X_N(\kappa_N(t)), y)\tilde{N}(dt, dy), \tag{6.6}$$

where $X_N(0) = x_0$,

$$\kappa_N(t) := \frac{[Nt]}{N}$$

for any $t \in [0,T]$, and

$$a_N(x) = \frac{a(x)}{1 + N^{-1/2}\|a(x)\|}.$$

Its integral form is

$$\begin{aligned} X_N(t) = x_0 &+ \int_0^t a_N(X_N(\kappa_N(s)))ds + \int_0^t b(X_N(\kappa_N(s)))dW(s) \\ &+ \int_0^t \int_{\mathbb{R}^d} f(X_N(\kappa_N(s)), y)\tilde{N}(ds, dy). \end{aligned} \tag{6.7}$$

It is obvious that

$$\|a_N(x)\| \leq \|a(x)\|,$$

and

$$\|a_N(x)\| = \left\|\frac{a(x)}{1 + N^{-1/2}\|a(x)\|}\right\| \leq \left\|\frac{a(x)}{N^{-1/2}\|a(x)\|}\right\| = N^{1/2}. \tag{6.8}$$

Furthermore, for any $x \in \mathbb{R}^d$, we have

$$\|a_N(x) - a(x)\| = \left\|\frac{a(x)}{1 + N^{-1/2}\|a(x)\|} - a(x)\right\|$$

$$= \Big\| \frac{N^{-1/2} a(x) \|a(x)\|}{1 + N^{-1/2} \|a(x)\|} \Big\|$$
$$\leq N^{-1/2} \|a(x)\|^2. \tag{6.9}$$

Before we give the main theorem of this chapter, we state several lemmas that will be useful in its proof.

Lemma 6.0.1. *Let $r \geq 2$. There exists a constant C, depending only on r, such that for every real-valued, $\mathcal{B}([0,T]) \times \mathcal{B}(\mathbb{R}^d)$-measurable function g satisfying*

$$\int_0^T \int_{\mathbb{R}^d} \|g(t,y)\|^2 \nu(dy) dt < \infty,$$

the following estimate holds,

$$E\Big[\sup_{0 \leq t \leq T} \Big\| \int_0^t \int_{\mathbb{R}^d} g(s,y) \tilde{N}(ds,dy) \Big\|^r \Big]$$
$$\leq CE\Big(\int_0^T \int_{\mathbb{R}^d} \|g(t,y)\|^2 \nu(dy) dt \Big)^{r/2} + CE \int_0^T \int_{\mathbb{R}^d} \|g(t,y)\|^r \nu(dy) dt. \tag{6.10}$$

It is known that if $1 \leq r \leq 2$, then the second term in (6.10) can be dropped.

This lemma is a simplified version of Lemma 2.1 in [11]. The proof can be found in [41].

The following estimation on the p-th moment of the numerical solution $X_N(t)$ is crucial in the whole section.

Lemma 6.0.2 ([11], Lemma 3.2)**.** *Let A-1–A-5 be satisfied. Then,*

$$\sup_{N \in \mathbb{N}} E\Big[\sup_{0 \leq t \leq T} \|X_N(t)\|^p \Big] \leq C,$$

with $C := C(T, p, E[\|x_0\|^p])$ independent of N.

Proof. See Lemma 3.2 in [11]. □

Lemma 6.0.3 ([11], Lemma 3.3)**.** *Let assumptions A-1–A-5 be satisfied. Then the numerical scheme* (6.6) *satisfies*

$$\sup_{0\le t\le T} E[\|X_N(t)-X_N(\kappa_N(t))\|^r]\le CN^{-1}$$

for any $2\le r\le p$ *with* $C:=C(T,p,E[\|x_0\|^p])$ *which does not depend on* N*.*

Now we introduce the main theorem in this section, the convergence rate of numerical scheme (6.6).

Theorem 6.1 ([11], Corollary 3.2)**.** *Let Assumptions A-1–A-6 be satisfied. Then the numerical scheme* (6.6) *converges in the* L^q *sense, i.e.*

$$E\Big[\sup_{0\le t\le T}\|X(t)-X_N(t)\|^q\Big]\le CN^{-\frac{q}{q+\delta}}, \tag{6.11}$$

where constant $C>0$ *does not depend on* N*.*

Proof. Let $e(t):=X(t)-X_N(t)$ and define

$$\bar{a}(t):=a(X(t))-a_N(X_N(\kappa_N(t))), \tag{6.12}$$

$$\bar{b}(t):=b(X(t))-b(X_N(\kappa_N(t))), \tag{6.13}$$

$$\bar{f}(t,y):=f(X(t),y)-f(X_N(\kappa_N(t)),y). \tag{6.14}$$

By (6.2) and (6.6), the error process $e(t)$ satisfies

$$e(t)=\int_0^t\bar{a}(s)ds+\int_0^t\bar{b}(s)dW(s)+\int_0^t\int_{\mathbb{R}^d}\bar{f}(s,y)\tilde{N}(ds,dy), \tag{6.15}$$

for any $t\in[0,T]$. By Itô's formula, one obtains

$$\begin{aligned}\|e(t)\|^q&=q\int_0^t\|e(s)\|^{q-2}e(s)\bar{a}(s)ds+q\int_0^t\|e(s)\|^{q-2}e(s)\bar{b}(s)dW(s)\\&\quad+\frac{q(q-2)}{2}\int_0^t\|e(s)\|^{q-4}\|\bar{b}(s)^Te(s)\|^2ds+\frac{q}{2}\int_0^t\|e(t)\|^{q-2}\|\bar{b}\|^2ds\end{aligned}$$

$$+ q\int_0^t \int_{\mathbb{R}^d} \|e(s)\|^{q-2} e(s)\bar{f}(s,y)\tilde{N}(ds,dy)$$
$$+ \int_0^t \int_{\mathbb{R}^d} \Big(\|e(s)+\bar{f}(s,y)\|^q - \|e(s)\|^q - q\|e(s)\|^{q-2}e(s)\bar{f}(s,y)\Big)N(ds,dy) \tag{6.16}$$

for all $t \in [0,T]$. We write

$$\begin{aligned} e(s)\bar{a}(s) = (X(s)-X_N(s))&\Big(a(X(s)) - a(X_N(s))\Big) \\ &+ (X(s)-X_N(s))\Big(a(X_N(s)) - a(X_N(\kappa_N(s)))\Big) \\ &+ (X(s)-X_N(s))\Big(a(X_N(\kappa_N(s))) - a_N(X_N(\kappa_N(s)))\Big). \end{aligned} \tag{6.17}$$

Hence, by condition A-2, Cauchy-Schwarz inequality and Young's inequality,

$$\begin{aligned} &\|e(s)\|^{q-2}e(s)\bar{a}(s) \\ &\le \|e(s)\|^{q-2}\Big(C\|e(s)\|^2 + \|e(s)\|^2 + \|a(X_N(s)) - a(X_N(\kappa_N(s)))\|^2 \\ &\quad + \|a(X_N(\kappa_N(s))) - a_N(X_N(\kappa_N(s)))\|^2\Big) \\ &= C\|e(s)\|^q + \|e(s)\|^{q-2}\|a(X_N(s)) - a(X_N(\kappa_N(s)))\|^2 \\ &\quad + \|e(s)\|^{q-2}\|a(X_N(\kappa_N(s))) - a_N(X_N(\kappa_N(s)))\|^2 \\ &\le C\|e(s)\|^q + C\|a(X_N(s)) - a(X_N(\kappa_N(s)))\|^q \\ &\quad + C\|a(X_N(\kappa_N(s))) - a_N(X_N(\kappa_N(s)))\|^q, \end{aligned} \tag{6.18}$$

where we used the conjugate equality

$$\frac{1}{q/(q-2)} + \frac{1}{q/2} = 1$$

for the last inequality. Therefore, taking suprema over $[0,u]$ for any $u \in [0,T]$ and expectations on both sides of (6.16), we obtain

$$\begin{aligned}
&E\Big[\sup_{0\le t\le u}\|e(t)\|^q\Big] \\
&\le CE\int_0^u \|e(s)\|^q ds + CE\int_0^u \|a(X_N(s)) - a(X_N(\kappa_N(s)))\|^q ds \\
&\quad + CE\int_0^u \|a(X_N(\kappa_N(s))) - a_N(X_N(\kappa_N(s)))\|^q ds \\
&\quad + qE\Big[\sup_{0\le t\le u}\Big\|\int_0^t \|e(s)\|^{q-2}e(s)\bar{b}(s)dW(s)\Big\|\Big] \\
&\quad + \frac{q(q-2)}{2}E\int_0^u \|e(s)\|^{q-4}\|\bar{b}(s)^T e(s)\|^2 ds + \frac{q}{2}E\int_0^u \|e(t)\|^{q-2}\|\bar{b}\|^2 ds \\
&\quad + qE\Big[\sup_{0\le t\le u}\Big\|\int_0^t\int_{\mathbb{R}^d} \|e(s)\|^{q-2}e(s)\bar{f}(s,y)\tilde{N}(ds,dy)\Big\|\Big] \\
&\quad + E\Big[\sup_{0\le t\le u}\int_0^t\int_{\mathbb{R}^d}\Big(\|e(s)\|^{q-2}\|\bar{f}(s,y)\|^2 + \|\bar{f}(s,y)\|^q\Big)N(ds,dy)\Big] \\
&= G_1 + G_2 + G_3 + G_4 + G_5 + G_6 + G_7 + G_8. \qquad (6.19)
\end{aligned}$$

In the following we estimate $G_1 - G_8$ one by one. Firstly,

$$G_1 := CE\int_0^u \|e(s)\|^q ds \le C\int_0^u E\Big[\sup_{0\le r\le s}\|e(r)\|^q\Big]ds \qquad (6.20)$$

for any $u \in [0,T]$. By the second inequality of assumption A-6, Hölder's inequality and Lemma 6.0.2, G_2 can be estimated by

$$\begin{aligned}
G_2 &:= CE\int_0^u \|a(X_N(s)) - a(X_N(\kappa_N(s)))\|^q ds \\
&\le CE\int_0^u (1 + \|X_N(s)\|^\chi + \|X_N(\kappa_N(s))\|^\chi)^{q/2}\|X_N(s) - X_N(\kappa_N(s))\|^q ds \\
&\le C\int_0^u \Big(1 + E[\|X_N(s)\|^{\chi\frac{q}{2}\frac{q+\delta}{\delta}}] + E[\|X_N(\kappa_N(s))\|^{\chi\frac{q}{2}\frac{q+\delta}{\delta}}]\Big)^{\frac{\delta}{q+\delta}} \\
&\qquad \cdot \Big(E[\|X_N(s) - X_N(\kappa_N(s))\|^{q+\delta}]\Big)^{\frac{q}{q+\delta}} ds \\
&\le C\int_0^T \Big(E[\|X_N(s) - X_N(\kappa_N(s))\|^{q+\delta}]\Big)^{\frac{q}{q+\delta}} ds. \qquad (6.21)
\end{aligned}$$

The estimate of G_3 is trivial:

$$\begin{aligned} G_3 &:= CE\int_0^u \|a(X_N(\kappa_N(s))) - a_N(X_N(\kappa_N(s)))\|^q ds \\ &\le CE\int_0^T \|a(X_N(\kappa_N(s))) - a_N(X_N(\kappa_N(s)))\|^q ds. \end{aligned} \tag{6.22}$$

Due to Burkholder-Davis-Gundy inequality, we have

$$\begin{aligned} G_4 &:= qE\Big[\sup_{0\le t\le u}\Big\|\int_0^t \|e(s)\|^{q-2}e(s)\bar{b}(s)dW(s)\Big\|\Big] \\ &\le CE\Big[\int_0^u \|e(s)\|^{2q-2}\|\bar{b}(s)\|^2 ds\Big]^{1/2} \\ &\le E\Big[\sup_{0\le s\le u}\|e(s)\|^{q-1}\Big(\int_0^u \|\bar{b}(s)\|^2 ds\Big)^{1/2}\Big]. \end{aligned} \tag{6.23}$$

Using Young's inequality and Hölder's inequality, we get

$$\begin{aligned} G_4 &\le \frac{1}{8}E\Big[\sup_{0\le s\le u}\|e(s)\|^q\Big] + CE\Big[\Big(\int_0^u \|\bar{b}(s)\|^2 ds\Big)^{q/2}\Big] \\ &\le \frac{1}{8}E\Big[\sup_{0\le s\le u}\|e(s)\|^q\Big] + CE\int_0^u \|\bar{b}(s)\|^q ds \end{aligned} \tag{6.24}$$

for any $u \in [0,T]$. Due to Cauchy-Schwarz inequality and Young's inequality, G_5 and G_6 can be estimated together by

$$\begin{aligned} &G_5 + G_6 \\ &:= \frac{q(q-2)}{2}E\int_0^u \|e(s)\|^{q-4}\|\bar{b}(s)^T e(s)\|^2 ds + \frac{q}{2}E\int_0^u \|e(t)\|^{q-2}\|\bar{b}\|^2 ds \\ &\le CE\int_0^u \|e(t)\|^{q-2}\|\bar{b}\|^2 ds \\ &\le C\int_0^u E\Big[\sup_{0\le r\le s}\|e(r)\|^q\Big]ds + CE\int_0^u \|\bar{b}(s)\|^q ds \end{aligned} \tag{6.25}$$

for any $u \in [0,T]$. Combining (6.24) and (6.25), we have

$$G_4 + G_5 + G_6$$

$$\leq \frac{1}{8}E\Big[\sup_{0\leq s\leq u}\|e(s)\|^q\Big] + C\int_0^u E\Big[\sup_{0\leq r\leq s}\|e(r)\|^q\Big]ds + CE\int_0^u \|\bar{b}(s)\|^q ds. \tag{6.26}$$

We now split $\bar{b}(s)$ as

$$\bar{b}(s) = \Big(b(X(s)) - b(X_N(s))\Big) + \Big(b(X_N(s)) - b(X_N(\kappa_N(s)))\Big). \tag{6.27}$$

Then by assumption A-2 and (6.26), we obtain

$$\begin{aligned} G_4 + G_5 + G_6 \leq \frac{1}{8}E\Big[\sup_{0\leq s\leq u}\|e(s)\|^q\Big] + C\int_0^u E\Big[\sup_{0\leq r\leq s}\|e(r)\|^q\Big]ds \\ + C\int_0^T E\|X_N(s) - X_N(\kappa_N(s))\|^q ds \end{aligned} \tag{6.28}$$

for any $u \in [0, T]$. To estimate G_7 and G_8, we split $\bar{f}(s, y)$ as

$$\bar{f}(s,y) = \Big(f(X(s),y) - f(X_N(s),y)\Big) + \Big(f(X_N(s),y) - f(X_N(\kappa_N(s)),y)\Big). \tag{6.29}$$

Therefore, G_7 can be estimated by

$$\begin{aligned} G_7 &:= qE\Big[\sup_{0\leq t\leq u}\Big\|\int_0^t\int_{\mathbb{R}^d}\|e(s)\|^{q-2}e(s)\bar{f}(s,y)\tilde{N}(ds,dy)\Big\|\Big] \\ &\leq CE\Big[\sup_{0\leq t\leq u}\Big\|\int_0^t\int_{\mathbb{R}^d}\|e(s)\|^{q-2}e(s)(f(X(s),y) - f(X_N(s),y))\tilde{N}(ds,dy)\Big\|\Big] \\ &\quad + CE\Big[\sup_{0\leq t\leq u}\Big\|\int_0^t\int_{\mathbb{R}^d}\|e(s)\|^{q-2}e(s)(f(X_N(s),y) - f(X_N(\kappa_N(s)),y))\tilde{N}(ds,dy)\Big\|\Big]. \end{aligned} \tag{6.30}$$

By Lemma 6.0.1, it follows that

$$\begin{aligned} G_7 &\leq E\Big[\Big(\int_0^u\int_{\mathbb{R}^d}\|e(s)\|^{2q-2}\|f(X(s),y) - f(X_N(s),y)\|^2\nu(dy)ds\Big)^{1/2}\Big] \\ &\quad + E\Big[\Big(\int_0^u\int_{\mathbb{R}^d}\|e(s)\|^{2q-2}\|f(X_N(s),y) - f(X_N(\kappa_N(s)),y)\|^2\nu(dy)ds\Big)^{1/2}\Big] \end{aligned} \tag{6.31}$$

for any $u \in [0,T]$. As obtaining (6.24), we use Young's inequality and Hölder's inequality and obtain,

$$G_7 \leq \frac{1}{8}E\Big[\sup_{0\leq s\leq u}\|e(s)\|^q\Big] + E\Big[\int_0^u \Big(\int_{\mathbb{R}^d}\|f(X(s),y) - f(X_N(s),y)\|^2\nu(dy)\Big)^{q/2}ds\Big]$$
$$+ E\Big[\int_0^u \Big(\int_{\mathbb{R}^d}\|f(X_N(s),y) - f(X_N(\kappa_N(s)),y)\|^2\nu(dy)\Big)^{q/2}ds\Big]. \tag{6.32}$$

By Assumption A-2, it then follows that

$$G_7 \leq \frac{1}{8}E\Big[\sup_{0\leq s\leq u}\|e(s)\|^q\Big] + \int_0^u E\Big[\sup_{0\leq r\leq s}\|e(r)\|^q ds\Big]$$
$$+ \int_0^T E[\|X_N(s) - X_N(\kappa_N(s))\|^q]ds \tag{6.33}$$

for any $u \in [0,T]$. Finally, we can write G_8 as

$$G_8 := E\Big[\sup_{0\leq t\leq u}\int_0^t\int_{\mathbb{R}^d}\Big(\|e(s)\|^{q-2}\|\bar{f}(s,y)\|^2 + \|\bar{f}(s,y)\|^q\Big)N(ds,dy)\Big]$$
$$= E\int_0^u\int_{\mathbb{R}^d}\|e(s)\|^{q-2}\|\bar{f}(s,y)\|^2\nu(dy)ds + E\int_0^u\int_{\mathbb{R}^d}\|\bar{f}(s,y)\|^q\nu(dy)ds$$
$$=: H_1 + H_2 \tag{6.34}$$

for any $u \in [0,T]$. By the splitting (6.29), Cauchy-Schwarz inequality and assumption A-2, we have

$$H_1 \leq E\int_0^u\int_{\mathbb{R}^d}\|e(s)\|^{q-2}\|f(X(s),y) - f(X_N(s),y)\|^2\nu(dy)ds$$
$$+ E\int_0^u\int_{\mathbb{R}^d}\|e(s)\|^{q-2}\|f(X_N(s),y) - f(X_N(\kappa_N(s)),y)\|^2\nu(dy)ds$$
$$\leq CE\int_0^u\|e(s)\|^q ds + CE\int_0^u\|e(s)\|^{q-2}\|X_N(s) - X_N(\kappa_N(s))\|^2ds \tag{6.35}$$

for any $u \in [0,T]$. Using Young's inequality again, we obtain

$$H_1 \leq C\int_0^u E\Big[\sup_{0\leq r\leq s}\|e(s)\|^q\Big]ds + C\int_0^T E[\|X_N(s) - X_N(\kappa_N(s))\|^q]ds \tag{6.36}$$

for any $u \in [0, T]$. Using again the splitting (6.29), Cauchy-Schwarz inequality and assumption A-2, we get

$$H_2 \leq C \int_0^u E\Big[\sup_{0 \leq r \leq s} \|e(s)\|^q\Big] ds + C \int_0^T E[\|X_N(s) - X_N(\kappa_N(s))\|^q] ds \tag{6.37}$$

for any $u \in [0, T]$. Combining (6.36) and (6.37), we obtain

$$G_8 \leq C \int_0^u E\Big[\sup_{0 \leq r \leq s} \|e(s)\|^q\Big] ds + C \int_0^T E[\|X_N(s) - X_N(\kappa_N(s))\|^q] ds. \tag{6.38}$$

Combining (6.19), (6.20), (6.21), (6.22), (6.26), (6.33) and (6.38), we obtain

$$\begin{aligned} \frac{3}{4} E\Big[\sup_{0 \leq t \leq u} \|e(t)\|^q\Big] \leq C \int_0^u E\Big[\sup_{0 \leq r \leq s} \|e(r)\|^q\Big] ds & \\ + C \int_0^T \Big(E[\|X_N(s) - X_N(\kappa_N(s))\|^{q+\delta}]\Big)^{\frac{q}{q+\delta}} ds & \\ + CE \int_0^T \|a(X_N(\kappa_N(s))) - a_N(X_N(\kappa_N(s)))\|^q ds & \\ + C \int_0^T E[\|X_N(s) - X_N(\kappa_N(s))\|^q] ds. & \end{aligned} \tag{6.39}$$

Applying Gronwall's inequality (A.7), we have

$$\begin{aligned} E\Big[\sup_{0 \leq t \leq T} \|e(t)\|^q\Big] \leq C \int_0^T \Big(E\|X_N(s) - X_N(\kappa_N(s))\|^{q+\delta}\Big)^{\frac{q}{q+\delta}} ds & \\ + CE \int_0^T \|a(X_N(\kappa_N(s))) - a_N(X_N(\kappa_N(s)))\|^q ds & \\ + C \int_0^T E\|X_N(s) - X_N(\kappa_N(s))\|^q ds. & \end{aligned} \tag{6.40}$$

By (6.5) and (6.9), we can get

$$\begin{aligned} \|a(X_N(\kappa_N(s))) - a_N(X_N(\kappa_N(s)))\| &\leq N^{-1/2} \|a(X_N(\kappa_N(s)))\|^2 \\ &\leq CN^{-1/2}(1 + \|X_N(\kappa_N(s))\|^{\chi+2}). \end{aligned} \tag{6.41}$$

By Lemma (6.0.2), it then follows that

$$E\Big[\sup_{0\le t\le T}\|e(t)\|^q\Big] \le CN^{-\frac{q}{q+\delta}},$$

which completes the proof. □

Numerical Experiments

The example we use for our numerical experiment in this section is a 1-dimensional stochastic differential equation driven by Lévy motion,

$$dX(t) = -X^5(t)dt + X(t)dW(t) + \int_{\mathbb{R}} X(t)y\tilde{N}(dt,dy), \quad X(0) = 1.$$

Here, $t \in [0,1]$. The jump size follows standard normal distribution and the jump intensity is 2. We use 5 different time steps: $\Delta t = 2^{-12}, 2^{-11}, 2^{-10}, 2^{-9}, 2^{-8}$ and 1000 realizations for each discretisation. The following figure is the loglog plot of the experimental error with respect to the 5 different time steps.

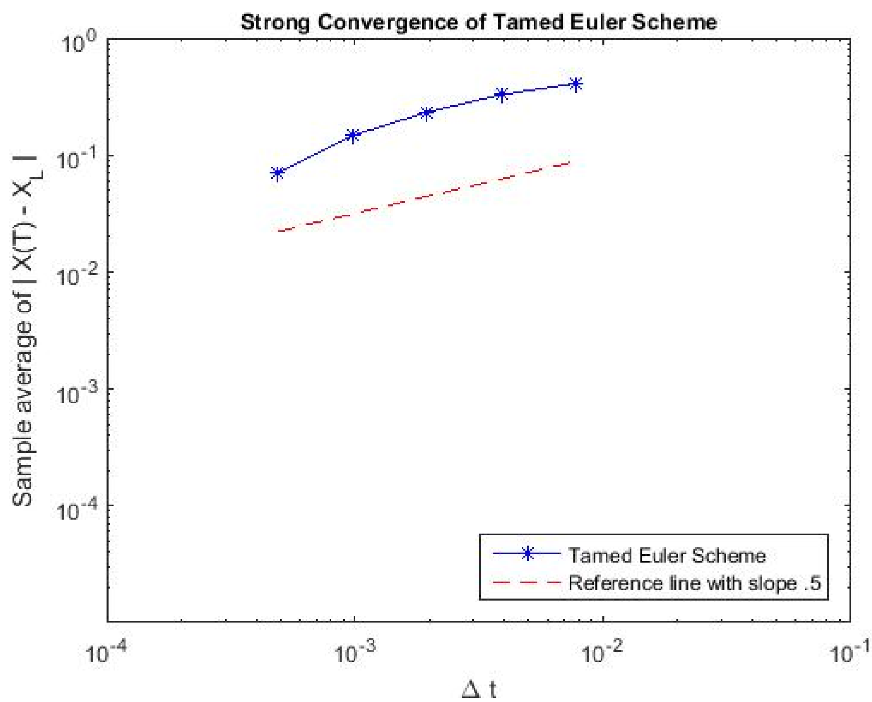

Figure 6.1: Log-log plot of the strong error from the numerical approximation for a SDE driven by Lévy motion with superlinearly growing drift

Bibliography

[1] M. B. Alaya and A. Kebaier. Central limit theorem for the multilevel Monte Carlo Euler method. *The Annals of Applied Probability*, 25(1):211–234, 2015. 95

[2] D. Applebaum. *Lévy Processes and Stochastic Calculus, 2nd edition*. Cambridge University Press, 2009. 1, 9, 19, 21, 22

[3] S. Asmussen and J. Rosiński. Approximations of small jumps of Lévy processes with a view towards simulation. *Journal of Applied Probability*, 38:482–493, 2001. 22

[4] V. Bally. Introduction to Malliavin Calculus. *Lecture Notes, available at http://perso-math.univ-mlv.fr/users/bally.vlad*, 2007. 61, 62

[5] V. Bally and D. Talay. The law of the Euler scheme for stochastic differential equations I: convergence rate of the distribution function. *Probability Theory and Related Fields*, 104(1):43–60, 1995. 5, 18, 59, 65, 82, 86

[6] A. Barth and A. Lang. Multilevel Monte Carlo method with applications to stochastic partial differential equations. *International Journal of Computer Mathematics*, 89(18):2479–2498, 2012. 95

[7] J. Bertoin. *Lévy Processes*. Cambridge University Press, 1998. 19

[8] N. Bouleau and D. Lepingle. *Numerical Methods for Stochastic Processes*. John Wiley & Sons, 1994. 17, 65

[9] K. A. Cliffe, M. B. Giles, R. Scheichl, and A. L. Teckentrup. Multilevel Monte Carlo methods and applications to elliptic PDEs with random coefficients. *Computing and Visualization in Science*, 14(1):3–15, 2011. 95

[10] R. Cont and P. Tankov. *Financial Modelling with Jump Processes*. Chapman & Hall/CRC, 2004. 19, 22

[11] K. Dareiotis, C. Kumar, and S. Sabanis. On tamed euler approximations of sdes driven by Lévy noise with applications to delay equations. *arXiv preprint arXiv:1403.0498*, 2014. 24, 96, 99, 100, 101

[12] S. Dereich. Multilevel Monte Carlo algorithms for Lévy-driven SDEs with Gaussian correction. *The Annals of Applied Probability*, pages 283–311, 2011. 95

[13] S. Dereich and S. Li. Multilevel Monte Carlo for Lévy-driven SDEs: central limit theorems for adaptive Euler schemes. *The Annals of Applied Probability*, 26(1): 136–185, 2016. 95

[14] L. C. Evans. *Partial Differential Equations, 2nd edition*. American Mathematical Society, 2010. 119, 122

[15] G. B. Folland. *Real Analysis: Modern Techniques and Their Applications*. John Wiley & Sons, 2013. 121

[16] M. B. Giles. Multilevel Monte Carlo path simulation. *Operations Research*, 56 (3):607–617, 2008. 6, 76, 88, 94, 95

[17] M. B. Giles. Multilevel Monte Carlo methods. *Acta Numerica*, 24:259–328, 2015. 94, 95

[18] M. B. Giles and L. Szpruch. Multilevel Monte Carlo methods for applications in finance. *arXiv preprint arXiv:1212.1377*, 2012. 94

[19] M. B. Giles, D. J. Higham, and X. Mao. Analysing multi-level Monte Carlo for options with non-globally Lipschitz payoff. *Finance and Stochastics*, 13(3): 403–413, 2009. 88

[20] P. Glasserman. *Monte Carlo Methods in Financial Engineering*, volume 53. Springer Science & Business Media, 2003. 78

[21] I. Gyöngy and N. V. Krylov. On stochastic equations with respect to semimartingales I. *Stochastics: An International Journal of Probability and Stochastic Processes*, 4(1):1–21, 1980. 23

[22] D. J. Higham. An algorithmic introduction to numerical simulation of stochastic differential equationns. *SIAM Reviews*, 43(3):525–546, 2001. 16

[23] D. J. Higham, X. Mao, and A. M. Stuart. Strong convergence of Euler-type methods for nonlinear stochastic differential equations. *SIAM Journal of Numerical Analysis*, 40(3):1041–1063, 2002. 3, 4, 18, 27, 31, 32, 33, 88

[24] Y. Hu. Semi-implicit Euler-Maruyama scheme for stiff stochastic equations. In *Stochastic Analysis and Related Topics V*, pages 183–202. Springer, 1996. 3, 18

[25] Y. Hu and S. Watanabe. Donsker delta functions and approximations of heat kernels by the time discretization method. *J. Math. Kyoto Univ.*, 36:499–518, 1996. 59

[26] M. Hutzenthaler and A. Jentzen. *Numerical approximations of stochastic differential equations with non-globally Lipschitz continuous coefficients*, volume 236. American Mathematical Society, 2015. 88

[27] M. Hutzenthaler, A. Jentzen, and P. E. Kloeden. Strong and weak divergence in finite time of Euler's method for stochastic differential equations with non-globally Lipschitz continuous coefficients. *Proceedings of the Royal Society of London A: Mathematical, Physical and Engineering Sciences*, 467(2130):1563–1576, 2011. 33, 34

[28] M. Hutzenthaler, A. Jentzen, and P. E. Kloeden. Strong convergence of an explicit numerical method for SDEs with nonglobally Lipschitz continuous

coefficients. *The Annals of Applied Probability*, 22(4):1611–1641, 2012. 4, 33, 35, 36, 88, 123

[29] N. Ikeda and S. Watanabe. *Stochastic Differential Equations and Diffusion Processes*. Elsevier, 2014. 65

[30] I. Karatzas and S. Shreve. *Brownian Motion and Stochastic Calculus, 2nd edition*. Springer Science & Business Media, 2012. 122

[31] F. Klebaner. *Introduction to Stochastic Calculus with Applicaitons, 3rd edition*. Imperial College Press, 2012. 1, 9, 11, 13

[32] P. Kloeden and A. Neuenkirch. Convergence of numerical methods for stochastic differential equations in mathematical finance. *arXiv preprint arXiv:1204.6620*, 2012. 88

[33] P. E. Kloeden and E. Platen. *Numerical Solution of Stochastic Differential Equations*. Springer, 1992. 1, 2, 5, 6, 16, 17, 18, 59, 81, 82

[34] A. Kohatsu-Higa. High order Itô-Taylor approximation to heat kernels. *J. Math. Kyoto Univ*, 37:129–150, 1997. 59

[35] A. Kohatsu-Higa. Weak approximations. a Malliavin Calculus approach. *Math. Comp.*, 70(233):135–172, 2001. 59

[36] C. Kumar and S. Sabanis. On tamed Milstein schemes of SDEs driven by Lévy noise. *arXiv preprint arXiv:1407.5347*, 2014. 96

[37] C. Kumar and S. Sabanis. On Milstein approximations with varying coefficients: the case of super-linear diffusion coefficients. *arXiv preprint arXiv:1601.02695*, 2016. 44

[38] H. Kunita. Stochastic differential equations and stochastic flows of diffeomorphisms. In *École d'Été de Probabilités de Saint-Flour XII-1982*, pages 143–303. Springer, 1984. 14

[39] X. Mao. *Stochastic Differential Equations and Applicatios, 2nd edition*. Elsevier, 2007. 9, 12, 13, 17

[40] R. C. Merton. Option pricing when underlying stock returns are discontinuous. *Journal of Financial Economics*, 3(1-2):125–144, 1976. 1

[41] R. Mikulevicius and H. Pragarauskas. On L_p-estimates of some singular integrals related to jump processes. *SIAM Journal on Mathematical Analysis*, 44(4):2305–2328, 2012. 100

[42] I. Nourdin and G. Peccati. *Normal Approximations with Malliavin Calculus, from Stein's Method to Universality*. Cambridge University Press, 2012. 61, 63

[43] D. Nualart. *The Malliavin Calculus and Related Topics, 2nd edition*. Springer, 2006. 61

[44] G. Di Nunno, B. Øksendal, and F. Proske. *Malliavin Calculus for Lévy Processes with Applications to Finance*. Springer, 2009. 61, 62

[45] B. Øksendal. *Stochastic Differential Equations: An Introduction with Applications, 6th edition*. Springer, 2003. 1, 9, 14

[46] A. Papapantoleon. An introduction to Lévy processes with applications in finance. *arXiv preprint arXiv:0804.0482*, 2008. 19

[47] E. Pardoux. *Markov Processes and Applications: Algorithms, Networks, Genome and Finance*. John Wiley & Sons, 2008. 78

[48] E. Platen. An introduction to numerical methods for stochastic differential equations. *Acta Numerica*, 8:197–246, 1999. 16

[49] E. Platen and N. Bruti-Liberati. *Numerical Solution of Stochastic Differential Equations with Jumps in Finance*. Springer Science & Business Media, 2010. 2, 19, 22

[50] P. E. Protter. *Stochastic Integration and Differential Equations, 2nd edition*. Springer, 2003. 22

[51] C. Robert and G. Casella. *Monte Carlo Statistical Methods*. Springer Science & Business Media, 2013. 78

[52] J. Rosiński. Series representations of Lévy processes from the perspective of point processes. In *Lévy Processes*, pages 401–415. Springer, 2001. 22

[53] J. Rosiński. Simulation of Lévy processes, encyclopedia of statistics in quality and reliability: Computationally intensive methods and simulation, 2008. 22

[54] J. Rosiński and L. Wang. Weak convergence of the Euler scheme for SDEs via Malliavin calculus. *in preparation*, 2016. 63

[55] S. Sabanis. Euler approximations with varying coefficients: the case of superlinearly growing diffusion coefficients. *to apprear in Annals of Applied Probability*, 2016. 4, 43, 44, 47, 48, 49, 52, 88

[56] K. Sato. *Lévy Processes and Infinitely Divisible Distributions, 2nd edition*. Cambridge University Press, 2013. 19, 20

[57] W. Shoutens. *Lévy Processes in Finance: Pricing Finanicial Derivatives*. John Wiley & Sons, 2003. 19

[58] S. E. Shreve. *Stochastic Calculus for Finance II: Continuous-Time Models*. Springer, 2004. 1, 19

[59] D. Talay. Efficient numerical schemes for the approximation of expectations of functionals of the solution of a SDE, and applications. In *Filtering and Control of Random Processes*, pages 294–313. Springer, 1984. 60

[60] D. Talay and L. Tubaro. Expansion of the global error for numerical schemes solving stochastic differential equations. *Stochastic Analysis and Applications*, 8: 483–509, 1990. 5, 59, 60, 82

[61] M. Tretyakov and Z. Zhang. A fundamental mean-square convergence theorem for SDEs with locally Lipschitz coefficients and its applications. *SIAM Journal on Numerical Analysis*, 51(6):3135–3162, 2013. 88

[62] Y. Xia. *Multilevel Monte Carlo for jump processes.* PhD thesis, University of Oxford, 2013. 95

Appendices

Appendix A

Inequalities

A.1 Elementary Inequalities

The following elementary inequalities are taken from the appendices of Evans [14].

1. Cauchy's inequality.
$$ab \leq a^2 + b^2, \quad a, b \in \mathbb{R}. \tag{A.1}$$

   Proof. $0 \leq (a-b)^2 = a^2 - 2ab + b^2$. □

2. Young's inequality. Let $1 < p, q < \infty$, $\frac{1}{p} + \frac{1}{q} = 1$. Then
$$ab \leq \frac{a^p}{p} + \frac{b^q}{q}, \quad a, b > 0. \tag{A.2}$$

   Proof. Since the mapping $x \mapsto e^x$ is convex,
$$\begin{aligned} ab = e^{\ln a + \ln b} &= e^{\frac{1}{p}\ln a^p + \frac{1}{q}\ln b^q} \\ &\leq \frac{1}{p}e^{\ln a^p} + \frac{1}{q}e^{\ln b^q} \\ &= \frac{a^p}{p} + \frac{b^q}{q}. \end{aligned}$$
   □

3. Cauchy-Schwarz inequality.

$$\|\langle x, y\rangle\| \leq \|x\|\|y\|, \quad x, y \in \mathbb{R}^m. \tag{A.3}$$

Proof. Let $\epsilon > 0$ and note

$$0 \leq \|x \pm \epsilon y\|^2 = \|x\|^2 \pm 2\epsilon\langle x, y\rangle + \epsilon^2\|y\|^2.$$

Consequently,

$$\pm\langle x, y\rangle \leq \frac{1}{2\epsilon}\|x\|^2 + \frac{\epsilon}{2}\|y\|^2.$$

Minimizing the right hand side by setting $\epsilon = \frac{\|x\|}{\|y\|}$, provided $y \neq 0$. □

4. Hölder's inequality. Assume $1 \leq p, q \leq \infty$, $\frac{1}{p} + \frac{1}{q} = 1$. Then if $u \in L^p(\Omega; \mathbb{R}^m)$, $v \in L^q(\Omega; \mathbb{R}^m)$, we have

$$\int_\Omega |\langle u, v\rangle| P(d\omega) \leq \|u\|_{L^p(\Omega;\mathbb{R}^m)}\|v\|_{L^q(\Omega;\mathbb{R}^m)}. \tag{A.4}$$

Proof. By homogeneity, we may assume $\|u\|_{L^p} = \|v\|_{L^q} = 1$. Then Young's inequality implies for $1 < p, q < \infty$ that

$$\int_\Omega |\langle u, v\rangle| P(d\omega) \leq \frac{1}{p}\int_\Omega \|u\|^p P(d\omega) + \frac{1}{q}\int_\Omega \|v\|^q P(d\omega) = 1 = \|u\|_{L^p}\|v\|_{L^q}.$$

□

5. Minkowski's inequality. Assume $1 \leq p \leq \infty$ and $u, v \in L^p(\Omega; \mathbb{R}^m)$. Then

$$\|u + v\|_{L^p} \leq \|u\|_{L^p} + \|v\|_{L^p}. \tag{A.5}$$

Proof.

$$\begin{aligned}
&\|u+v\|_{L^p}^p \\
&= \int_\Omega \|u+v\|^p P(d\omega) \le \int_\Omega \|u+v\|^{p-1}(\|u\|+\|v\|)P(d\omega) \\
&\le \Big(\int_\Omega \|u+v\|^p P(d\omega)\Big)^{\frac{p-1}{p}} \Big(\Big(\int_\Omega \|u\|^p P(d\omega)\Big)^{1/p} + \Big(\int_\Omega \|v\|^p P(d\omega)\Big)^{1/p}\Big) \\
&= \|u+v\|_{L^p}^{p-1}\big(\|u\|_{L^p} + \|v\|_{L^p}\big).
\end{aligned}$$

□

6. Minkowski's inequality for integrals. Suppose that $(X, \mathcal{M}, \mu)$ and $(Y, \mathcal{N}, \nu)$ are σ-finite measure spaces, and let f be an $(\mathcal{M}\otimes\mathcal{N})$-measurable function on $X\times Y$. If $f \ge 0$ and $1 \le p < \infty$, then

$$\Big[\int_X \Big(\int_Y f(x,y)d\nu(y)\Big)^p d\mu(x)\Big]^{1/p} \le \int_Y \Big[\int_X f(x,y)^p d\mu(x)\Big]^{1/p} d\nu(y). \tag{A.6}$$

Proof. The proof can be found on page 194 in Folland [15]. See also a generalized version of this inequality on that page. □

A.2 Gronwall's Inequality

Let $f(t)$ be a nonnegative, integrable function on $[0, T]$ which satisfies for a.e. t the integral inequality

$$f(t) \le C_1 \int_0^t f(s)ds + C_2$$

for constants $C_1, C_2 \ge 0$. Then

$$f(t) \le C_2(1 + C_1 t e^{C_1 t}) \tag{A.7}$$

for a.e. $0 \le t \le T$.

Proof. The proof can be found on page 625 in Evans [14]. □

A.3 Probability Inequalities

All the inequalities in this section occur in a general probability space $(\Omega, \mathcal{F}, P)$.

1. Lyapunov inequality. Let X be a $\mathbb{R}^m$-valued random variable and not concentrated on a single point. If $E[\|X\|^s]$ exists for some $s > 0$, then for all $0 < r < s$ and $a \in \mathbb{R}^m$,

$$(E[\|X - a\|^r])^{1/r} \le (E[\|X - a\|^s])^{1/s}. \tag{A.8}$$

Proof. Note that $\frac{1}{s/r} + \frac{1}{s/(s-r)} = 1$. Then this inequality is just a simple application of the Hölder's inequality (A.4). □

2. Doob's martingale inequality. Suppose $M(t)$ is a martingale (or a positive submartingale) on the interval $[0, \infty)$, then for any $p > 1$ and $t \in [0, \infty)$,

$$E\Big[\sup_{0 \le s \le t} |M(s)|^p\Big] \le \Big(\frac{p}{p-1}\Big)^p E[|M(t)|^p]. \tag{A.9}$$

Proof. See e.g. pages 13-14 in Karatzas and Shreve [30] for a proof. □

3. Burkholder-Davis-Gundy inequality. Suppose $M(t)$ is a local martingale on the interval $[0, T]$, null at 0. There are constants c_p and C_p depending only on p, such that

$$c_p E\Big[\Big([M, M](T)\Big)^{p/2}\Big] \le E\Big[\Big(\sup_{0 \le t \le T} |M(t)|\Big)^p\Big] \le C_p E\Big[\Big([M, M](T)\Big)^{p/2}\Big], \tag{A.10}$$

for $1 \le p < \infty$. Here $[M, M](t)$ is the quadratic variation process of M. If, moreover, $M(t)$ is continuous, then the result holds also for $0 < p < 1$.

Proof. The proof can be found on e.g. page 166 in Karatzas and Shreve [30]. □

4. Burkholder-Davis-Gundy inequality for stochastic integrals. Let $k \in \mathbb{N}$ and let $Z : [0,T] \times \Omega \to \mathbb{R}^{k \times d}$ be a predictable stochastic process satisfying $P(\int_0^T \|Z(s)\|^2 ds < \infty) = 1$. Then we obtain that

$$\Big\| \sup_{s \in [0,t]} \| \int_0^s Z(u) dW(u) \| \Big\|_{L^p(\Omega;\mathbb{R})} \leq p \Big(\int_0^t \sum_{i=1}^{d} \|Z(s)\vec{e_i}\|^2_{L^p(\Omega;\mathbb{R}^k)} ds \Big)^{1/2} \tag{A.11}$$

for all $t \in [0,T]$ and all $p \in [2,\infty)$, where $\vec{e_i}$ is the ith standard unit vector in the space $\mathbb{R}^d$.

Proof. This inequality is a straightforward corollary of Doob's martingale inequality and Burkholder-Davis-Gundy inequality. See Lemma 3.7 in [28] for the detailed proof. □

Appendix B

MATLAB Codes

B.1 MATLAB Codes for Generating the Graph in Section 3.2

```
% Author: Liguo Wang
% Department of Mathematics
% University of Tennessee, Knoxville
% April, 2016

% Tamed Euler scheme for SDEs
% Solves  1-dim  dX(t) = (X(t)-X^3(t))dt + X(t)dW(t),   X(0) = 1
% Discretized Brownian path over [0,1] has dt = 2^(-12)
% Find the exact solution by using the timestep dt
% E-M uses 5 different timesteps:  16dt, 8dt, 4dt, 2dt, dt
% Examine strong convergence at T=1:  E | X_L - X(T) | using Monte-Carlo

clc;
clear all;
close all;

randn('state',100)
T = 1; N = 2^12; dt = T/N;         %
```

```
M = 1000;                         % number of paths sampled


Xerr = zeros(M,5);                % preallocate array
for s = 1:M
    Xtrue = 1;
    dW = sqrt(dt)*randn(1,N);     % Brownian increments
    for i=1:N
        Xtrue = Xtrue + (Xtrue-Xtrue^3)*dt/(1+(Xtrue-Xtrue^3)*dt)...
         + Xtrue*dW(1,i);
    end
    for p = 1:5
        R = 2^(p); Dt = R*dt; L = N/R;     % L Euler steps of size Dt = R*dt
        Xtemp = 1;
        for j = 1:L
             Winc = sum(dW(R*(j-1)+1:R*j));
             Xtemp = Xtemp + (Xtemp-Xtemp^3)*Dt/(1+(Xtemp-Xtemp^3)*Dt)...
              + Xtemp*Winc;
        end
        Xerr(s,p) = abs(Xtemp - Xtrue);      % store the error at t = 1
    end
end

Dtvals = dt*(2.^([1:5]));
% subplot(221)                                 % top LH picture
loglog(Dtvals,mean(Xerr),'b*-'), hold on
loglog(Dtvals,(Dtvals.^(.5)),'r--'), hold off % reference slope of 1/2
axis([1e-4 1e-1 1e-5 1])
xlabel('\Delta t'), ylabel('Sample average of | X(T) - X_L |')
title('Strong Convergence of Tamed Euler Scheme','FontSize',10)
legend('Tamed Euler Scheme','Reference line with slope .5')
```

B.2 MATLAB Codes for Generating the Graph in Section 3.3

```
% Tamed Euler scheme for SDEs with superlinealy growing diffusion
% Solves  1-dim  dX(t) = X(t)(1-|X(t)|)dt + |X(t)|^{3/2}dW(t), X(0) = 1
% Discretized Brownian path over [0,1] has dt = 2^(-12)
% Find the exact solution by using the timestep dt
% E-M uses 5 different timesteps:  16dt, 8dt, 4dt, 2dt, dt
% Examine strong convergence at T=1:  E | X_L - X(T) | using Monte-Carlo

clc;
clear all;
close all;

randn('state',100)
T = 1; N = 2^12; dt = T/N;          %
M = 1000;                          % number of paths sampled



Xerr = zeros(M,5);                 % preallocate array
for s = 1:M
    Xtrue = 1;
    dW = sqrt(dt)*randn(1,N);      % Brownian increments
    for i=1:N
        Xtrue = Xtrue + Xtrue*(1-abs(Xtrue))*dt/(1+dt^(0.5)*abs(Xtrue)...
        ^(3.5)) + abs(Xtrue)^(1.5)*dW(1,i)/(1+dt^(0.5)*abs(Xtrue)^(3.5));
    end
    for p = 1:5
        R = 2^(p); Dt = R*dt; L = N/R;     % L Euler steps of size Dt = R*dt
        Xtemp = 1;
        for j = 1:L
             Winc = sum(dW(R*(j-1)+1:R*j));
             Xtemp = Xtemp + Xtemp*(1-abs(Xtemp))*Dt/(1+Dt^(0.5)*abs(Xtemp)...
             ^(3.5)) + abs(Xtemp)^(1.5)*Winc/(1+Dt^(0.5)*abs(Xtemp)^(3.5));
```

```
        end
        Xerr(s,p) = abs(Xtemp - Xtrue);      % store the error at t = 1
    end
end

Dtvals = dt*(2.^([1:5]));
% subplot(221)                                  % top LH picture
loglog(Dtvals,mean(Xerr),'b*-'), hold on
loglog(Dtvals,(Dtvals.^(.5)),'r--'), hold off % reference slope of 1/2
axis([1e-4 1e-1 1e-5 1])
xlabel('\Delta t'), ylabel('Sample average of | X(T) - X_L |')
title('Strong Convergence of Tamed Euler Scheme','FontSize',10)
legend('Tamed Euler Scheme','Reference line with slope .5')
```

B.3 MATLAB Codes for Generating the Graph in Section 6.1

```
% Author: Liguo Wang
% Department of Mathematics
% University of Tennessee, Knoxville
% April, 2016

% Tamed Euler scheme for SDEs with superlinealy growing diffusion
% Solves  1-dim  dX(t) = -X^5(t)dt+X(t)dW(t)+\int_{R}X(t)y\tilde{N}(dt,dy)
% The jump size follows standard normal distribution
% Discretized Brownian path over [0,1] has dt = 2^(-12)
% Find the exact solution by using the timestep dt
% E-M uses 5 different timesteps:  16dt, 8dt, 4dt, 2dt, dt
% Examine strong convergence at T=1:  E | X_L - X(T) | using Monte-Carlo

clc;
clear all;
close all;
```

```
randn('state',100)
T = 1; N = 2^12; dt = T/N;
M = 1000;                         % number of paths sampled
lambda = 2;                       % intensity of the Poisson random var.


Xerr = zeros(M,5);                % preallocate array
for s = 1:M
    Xtrue = 1;
    dW = sqrt(dt)*randn(1,N);     % Brownian increments
    dN = poissrnd(lambda*dt, 1, N);
    for i=1:N
        Xtrue = Xtrue - Xtrue^5/(N+N^(0.5)*abs(Xtrue)^5)...
         + Xtrue*dW(1,i) + Xtrue*dt*sum(randn(1,dN(1,i)));
    end
    for p = 1:5
        R = 2^(p); Dt = R*dt; L = N/R;     % L Euler steps of size Dt = R*dt
        Xtemp = 1;
        for j = 1:L
             Winc = sum(dW(R*(j-1)+1:R*j));
             Ninc = sum(dN(R*(j-1)+1:R*j));
             Xtemp = Xtemp - Xtemp^5/(N+N^(0.5)*abs(Xtemp)^5) + Xtemp*Winc...
            + Xtemp*Dt*sum(randn(1,Ninc));
        end
        Xerr(s,p) = abs(Xtemp - Xtrue);      % store the error at t = 1
    end
end

Dtvals = dt*(2.^([1:5]));
% subplot(221)                                      % top LH picture
loglog(Dtvals,mean(Xerr),'b*-'), hold on
loglog(Dtvals,(Dtvals.^(.5)),'r--'), hold off % reference slope of 1/2
axis([1e-4 1e-1 1e-5 1])
```

```
xlabel('\Delta t'), ylabel('Sample average of | X(T) - X_L |')
title('Strong Convergence of Tamed Euler Scheme','FontSize',10)
legend('Tamed Euler Scheme','Reference line with slope .5', 'Location', ...
    'southeast')
```

Vita

Liguo Wang was born in Nanyang, China in 1985. He attended Henan Normal University to study mathematics in 2004 and received a Bachelor of Science in 2008. He then attended the graduate school of University of Science and Technology of China in September 2008 and graduated with a Master of Science in Mathematics three years later.

In August 2011, he joined the Department of Mathematics at the University of Tennessee, Knoxville as a graduate student, concentrating on stochastic differential equations. He received his Doctorate of Philosophy in Mathematics and a Master of Science in Statistics in August 2016.

Starting from September 2016, he will work as a senior quantitative model developer at the Bank of New York Mellon in Pittsburgh, Pennsylvania.