Adaptive Timestepping for SDEs with Non-globally Lipschitz Drift

Wei Fang

St Hugh's College

University of Oxford

A thesis submitted for the degree of

Doctor of Philosophy

Hilary 2019

To my Parents:
Chanlong Fang & *Xiaoli Tian*
for their unending support and encouragement.

Acknowledgements

I would like to express my gratitude to Prof. Mike Giles for being my supervisor, mentor and friend, and also for the never-ending and illuminating discussions and debates on SDEs, MLMC and in general on maths and life.

Among my fellow DPhil students, I owe many thanks to Zhenru Wang, Dunhong Jin, Guangyu Xi, and Yixuan Wang for numerous discussions about maths and life, and reading and commenting on this thesis. I cannot mention here all the students and other researchers I have had the chance to talk to and learn about maths. I am grateful to all of them.

I am thankful to Prof. Endre Suli and Prof. Ben Hambly to be the examiners of my transfer and confirmation at the Mathematical Institute, and to Prof. Endre Suli and Prof. Arnulf Jentzen from ETH Zurich to be the examiners of my viva. I am also grateful to Dr. Howard Thom from University of Bristol for the cooperation on the EVPPI project which is my first time to be an independent researcher.

I thankfully acknowledge the partial financial support from the Hubs for Trials Methodology Research (HTMR) through EVPPI project and the Engineering and Physical Sciences Research Council (EPSRC) through the ICONIC project.

Finally, I am deeply grateful to my families for their many forms of devoted support and in particular to my parents for their love and unlimited patience.

Statement of Originality

I hereby declare that this thesis contains no material which has been accepted or is currently being submitted for any other degree, diploma, certificate or other qualifications at the University of Oxford or elsewhere. To the best of my knowledge, this thesis contains no material previously published and precise reference is made when a previously published result is used or discussed.

This thesis includes three papers published and two Arxiv papers, the first one is in Arxiv [22] (Chapter 2); the second one is in Arxiv [23] (Chapter 3); the summary of the theoretical results of the two Arxiv papers is in the conference paper [21] and their published version is accepted in The Annals of Applied Probability [24]; the third paper is accepted in the Journal of Mathematical Analysis and Applications [25] (Chapter 4), all of which are co-authored with my supervisor Prof. Mike Giles.

The remainder of the thesis (Chapter 5) contains one additional unpublished papers, also co-authored with my supervisor Prof. Mike Giles.

The ideas, development and writing up of the thesis, and the papers included in it, were the principal responsibility of myself, Wei Fang, working under the supervision of Prof. Mike Giles.

Abstract

In this thesis, we focus on the numerical approximation of SDEs with a drift which is not globally Lipschitz, and corresponding sensitivity calculations.

First, we propose an adaptive timestep construction for an Euler-Maruyama approximation of these SDEs over a finite time interval. It is proved that if the timestep is bounded appropriately, then over a finite time interval the numerical approximation is stable, and the expected number of timesteps is finite. Moreover, we extend this scheme to ergodic SDEs with contractivity over an infinite time interval and prove that the bound for moments and the strong error of the numerical solution are uniform in T, which allow us to introduce the adaptive multilevel Monte Carlo (MLMC) method to further improve the efficiency.

Next, we apply a new MLMC method for the ergodic SDEs without contractivity. By introducing a change of measure technique, we simulate the paths with contractivity and add the Radon-Nikodym derivative to the estimator. It is shown that the variance of the new level estimators increases linearly in T, which is a great reduction compared with the exponential increase in the standard MLMC.

Lastly, we derive a new pathwise sensitivity estimator for chaotic SDEs by introducing a spring term between the original and perturbed SDEs. The variance of the new estimator increases only linearly in time T, compared with the exponential increase of the standard pathwise estimator. We also consider using this estimator for the SDE with Richardson-Romberg extrapolation on the volatility parameter to approximate the sensitivities of the invariant measure of chaotic ODEs.

All of the analysis is supported by numerical results.

Contents

List of Figures

Chapter 1

Introduction

Throughout this thesis, we consider an m-dimensional stochastic differential equation (SDE) driven by a d-dimensional Brownian motion:

$$\mathrm{d}X_t = f(X_t)\,\mathrm{d}t + g(X_t)\,\mathrm{d}W_t, \tag{1.1}$$

for $t \in (0,\infty)$ with a fixed initial value x_0, the drift coefficient $f : \mathbb{R}^m \to \mathbb{R}^m$ and the diffusion coefficient $g : \mathbb{R}^m \to \mathbb{R}^{m\times d}$.

We assume that the following settings are satisfied. Let $T \in (0,\infty)$ be a fixed positive real number, and let $(\Omega, \mathcal{F}, \mathbb{P})$ be a probability space with normal filtration $(\mathcal{F}_t)_{t\in[0,T]}$ for chapter 2 and $(\mathcal{F}_t)_{t\in[0,\infty)}$ for chapters 3, 4 and 5 corresponding to a d-dimensional standard Brownian motion $W_t = (W^{(1)}, W^{(2)}, ..., W^{(d)})_t$.

We denote the vector norm by $\|v\| = (|v_1|^2 + |v_2|^2 + ... + |v_m|^2)^{\frac{1}{2}}$, the inner product of vectors v and w by $\langle v, w\rangle = v_1w_1 + v_2w_2 + ... + v_mw_m$, for any $v, w \in \mathbb{R}^m$ and the Frobenuius matrix norm by $\|A\| = \sqrt{\sum_{i,j} A_{i,j}^2}$ for all $A \in \mathbb{R}^{m\times d}$.

In the following sections, we first introduce the standard theory on globally Lipschitz SDEs and their numerical methods, and then consider the SDEs with non-globally Lipschitz coefficients and their finite time approximations and computations of ergodic limits, and finally give an overview of this thesis.

1.1 Standard theory

1.1.1 Globally Lipschitz SDEs & Euler-Maruyama method

The following theorem, which can be found in many textbooks, for example [53], gives a simple criterion for the existence and uniqueness of strong solutions of SDEs.

Theorem 1.1.1. *Assume that there exist constants* $K_1, K_2 \in (0,\infty)$ *such that*

1. (Lipschitz condition) *for all* $x,y \in \mathbb{R}^m$ *it holds that*

$$\|f(x)-f(y)\|^2 \leq K_1\|x-y\|^2, \quad \|g(x)-g(y)\|^2 \leq K_1\|x-y\|^2; \tag{1.2}$$

2. (Linear growth condition) *for all* $x \in \mathbb{R}^m$ *it holds that*

$$\|f(x)\|^2 + \|g(x)\|^2 \leq K_2(1+\|x\|^2). \tag{1.3}$$

Then, for each real number $T \in (0,\infty)$, *there exists a unique strong solution* $(X_t)_{0\leq t\leq T}$ *to equation* (1.1).

Under the globally Lipschitz assumption, one simple and straightforward numerical method is the explicit Euler-Maruyama method [72] using $N \in \mathbb{N}$ uniform timesteps. It is given by mappings $\widehat{X}_{nh}: \ \Omega \to \mathbb{R}^m$, for any $n \in \{0,1,...,N\}, \ h = T/N$, satisfying $\widehat{X}_0 = x_0$ and

$$\widehat{X}_{(n+1)h} = \widehat{X}_{nh} + f(\widehat{X}_{nh})h + g(\widehat{X}_{nh})\Delta W_n, \tag{1.4}$$

where $\Delta W_n = W_{(n+1)h} - W_{nh}$. For any $t \in (nh,(n+1)h)$, we use the standard continuous interpolation for $\widehat{X}_t$ that

$$\widehat{X}_t = \widehat{X}_{nh} + f(\widehat{X}_{nh})(t-nh) + g(\widehat{X}_{nh})(W_t - W_{nh}). \tag{1.5}$$

Kloeden & Platen [58] show the mean square stability of this numerical solution, that is,

$$\mathbb{E}\left[\|\widehat{X}_T\|^2\right] < \infty, \tag{1.6}$$

first-order weak convergence, that is,

$$\mathbb{E}\left[X_T - \widehat{X}_T\right] = O(h), \tag{1.7}$$

and $\frac{1}{2}$-order strong convergence, that is,

$$\mathbb{E}\left[\|X_T - \widehat{X}_T\|^2\right]^{1/2} = O(h^{1/2}), \tag{1.8}$$

which can be used to estimate the error in the simulated terminal value. However, when considering the estimation of path-dependent functions of $(X_t)_{0\le t\le T}$, we show these results in a stronger sense than (1.8), that is the strong mean square stability of the scheme,

$$\mathbb{E}\left[\sup_{0\le t\le T} \|\widehat{X}_t\|^2\right] < \infty, \tag{1.9}$$

and the strong convergence that

$$\mathbb{E}\left[\sup_{0\le t\le T} \|X_t - \widehat{X}_t\|^2\right]^{1/2} = O(h^{1/2}). \tag{1.10}$$

By the Itô-Taylor expansion, this can be easily extended to higher convergence order schemes, such as the Milstein method [76].

To estimate $\mathbb{E}[\varphi(X_T)]$, for some function φ, numerically we use an average of N independent samples $\varphi(\widehat{X}_T(\omega^{(n)}))$ with $\omega^{(n)}$ coming from the probability space $(\Omega, \mathcal{F}, \mathbb{P})$

$$\frac{1}{N}\sum_{n=1}^{N} \varphi(\widehat{X}_T(\omega^{(n)})). \tag{1.11}$$

The variance of this standard Monte Carlo estimate is $N^{-1}\mathbb{V}[\varphi(\widehat{X}_T)]$. In order to achieve ε^2 mean square error (MSE), we need $O(\varepsilon^{-2})$ samples. For the explicit Euler-Maruyama method, the first-order weak convergence requires $h = O(\varepsilon)$. Therefore, the total computational cost of the standard Monte Carlo Euler method is $O(\varepsilon^{-3})$.

1.1.2 Multilevel Monte Carlo

One approach to reduce the computational cost is the Multilevel Monte Carlo (MLMC) method, proposed by Giles in [30, 32]. Instead of direct estimation, a telescope summation is estimated:

$$\mathbb{E}[\varphi(\widehat{X}_T^L)] = \mathbb{E}[\varphi(\widehat{X}_T^0)] + \sum_{\ell=1}^{L} \mathbb{E}[\varphi(\widehat{X}_T^\ell) - \varphi(\widehat{X}_T^{\ell-1})], \tag{1.12}$$

where $\widehat{X}_T^\ell$ is the estimator using step size $M^{-\ell}h_0$ for some constant $h_0 \in (0,\infty)$ and $M \in \{2,3,...\}$. The benefit of MLMC is that it can greatly reduce the computational cost by the following theorem.

Theorem 1.1.2 (Giles [32]). *If there exist independent estimators Y_ℓ based on N_ℓ Monte Carlo samples, each with expected cost C_ℓ and variance V_ℓ, and positive constants α, β, γ, c_1, c_2, c_3 such that $\alpha \geq \frac{1}{2}\min(\beta,\gamma)$ and*

i) $|\mathbb{E}[\varphi(\widehat{X}^\ell) - \varphi(X)]| \;\leq\; c_1\, 2^{-\alpha\ell}$,

ii) $\mathbb{E}[Y_\ell] = \begin{cases} \mathbb{E}[\varphi(\widehat{X}^0)], & \ell = 0, \\ \mathbb{E}[\varphi(\widehat{X}^\ell) - \varphi(\widehat{X}^{\ell-1})], & \ell \in \{1,2,...,L\}, \end{cases}$

iii) $V_\ell = \mathbb{V}[Y_\ell] \leq c_2\, 2^{-\beta\ell}$,

iv) $C_\ell \leq c_3\, 2^{\gamma\ell}$,

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

$$Y = \sum_{\ell=0}^{L} Y_\ell \tag{1.13}$$

has a mean square error (MSE) with bound

$$\mathbb{E}\left[(Y - \mathbb{E}[\varphi(X)])^2\right] \leq \varepsilon^2 \tag{1.14}$$

with a computational cost C_{mlmc} *with bound*

$$\mathbb{E}[\mathrm{C}_{mlmc}] \leq \begin{cases} c_4\,\varepsilon^{-2}, & \beta > \gamma, \\ c_4\,\varepsilon^{-2}(\log\varepsilon)^2, & \beta = \gamma, \\ c_4\,\varepsilon^{-2-(\gamma-\beta)/\alpha}, & \beta < \gamma. \end{cases} \tag{1.15}$$

We should remark here that MLMC can be applied to any stable numerical scheme, where α and β can be determined by the weak convergence order and strong convergence order separately. Besides, strong convergence implies weak convergence in most cases. Therefore, it is crucial to analyse the strong convergence of the scheme to determine the MLMC complexity.

1.2 Non-globally Lipschitz SDEs

1.2.1 Strong solutions of non-globally Lipschitz SDEs

Although a relatively complete theory for globally Lipschitz SDEs and their approximations has been established, it is still far from satisfactory since many practical examples have only locally Lipschitz coefficients. The following theorem, see Theorem 3.6 in [67], shows an extended criterion for strong existence and uniqueness of SDEs.

Theorem 1.2.1. *Assume the following conditions are satisfied:*

(i) (Locally Lipschitz condition) *for any* $R \in (0,\infty)$, *there exists a positive constant* C_R *such that for all* $x,y \in \mathbb{R}^m$ *with* $\|x\|, \|y\| \leq R$ *it holds that*

$$\|f(x) - f(y)\| + \|g(x) - g(y)\| \leq C_R \|x - y\|; \tag{1.16}$$

(ii) (One-sided linear growth condition) *there exist constants* $\alpha, \beta \in (0,\infty)$ *such that for all* $x \in \mathbb{R}^m$ *it holds that*

$$\langle x, f(x)\rangle + \frac{1}{2}\,\|g(x)\|^2 \leq \alpha\|x\|^2 + \beta. \tag{1.17}$$

Then, for each real number $T \in (0,\infty)$, there exists a unique strong solution $(X_t)_{0\le t\le T}$ to equation (1.1)*.*

In addition, we introduce the *one-sided Lipschitz condition*, that is there exists a constant $K \in (0,\infty)$ such that for all $x, y \in \mathbb{R}^m$ it holds that

$$\langle x - y, f(x) - f(y)\rangle \le K\|x - y\|^2, \tag{1.18}$$

which facilitates the analysis of the strong convergence order of the numerical scheme. By taking $y = 0$ and using the Cauchy-Schwartz inequality, we obtain the one-sided linear growth condition for f :

$$\langle x, f(x)\rangle \le K\|x\|^2 + \langle x, f(0)\rangle \le (K + \frac{1}{2})\|x\|^2 + \frac{1}{2}\|f(0)\|^2. \tag{1.19}$$

The one-sided Lipschitz condition for the drift coefficient and the globally Lipschitz diffusion coefficient together imply the one-sided linear growth condition and the locally Lipschitz condition, and therefore ensure the strong existence and uniqueness of SDEs.

1.2.2 Ergodicity of SDEs

The stochastic process $(X_t)_{t\ge 0}$ satisfying the SDE (1.1) is also a Markov process, and ergodicity is an important property to investigate. We will introduce some basic theory about a specific class of ergodic SDEs in [73]: dissipative SDEs.

Assumption 1.2.1 (Dissipativity Condition)**.** *There exist constants $\alpha, \beta \in (0,\infty)$ such that for all $x \in \mathbb{R}^m$ it holds that*

$$\langle x, f(x)\rangle + \frac{1}{2}\|g(x)\|^2 \le -\alpha\|x\|^2 + \beta. \tag{1.20}$$

Assumption 1.2.2 (Minorization Condition [73])**.** *The Markov process $(X_t)_{t\ge 0}$ with transition kernel $P_t(x; A) = \mathbb{P}[X_t \in A | X_0 = x]$ for any $A \in \mathcal{B}(\mathbb{R}^m)$ the Borel σ-algebra on $\mathbb{R}^m$, satisfies that for some fixed compact set C,*

1. *for some* $y^* \in int(C)$ *there is, for any* $\delta \in (0,\infty)$, *a* $t_1 = t_1(\delta)$ *such that for all* $x \in C$ *it holds that*

$$P_{t_1}(x, \mathcal{B}_\delta(y^*)) \in (0,\infty), \tag{1.21}$$

where $\mathcal{B}_\delta(y^*)$ *denotes the open ball of radius* δ *centered at* y^*.

2. *for any* $t \in (0,\infty)$, *the transition kernel* $P_t(x;A)$ *possesses a continuous density* $p_t(x,y)$, *precisely*

$$P_t(x,A) = \int_A p_t(x,y)\,\mathrm{d}y. \tag{1.22}$$

Theorem 1.2.2 (Mattingly et al. [73]). *Let* f *is locally Lipschitz,* g *is bounded, and Assumptions 1.2.1 and 1.2.2 hold, then the solution to the SDE* (1.1) *has a unique invariant measure* π, *and there exist constants* $l \in [1,\infty), \lambda, \kappa \in (0,\infty)$ *such that for any* $\varphi(x) \le \alpha\|x\|^{2l} + \beta$, *and for all* $t \in (0,\infty)$ *it holds that*

$$|\mathbb{E}[\varphi(X_t)] - \pi(\varphi)| \le \kappa(1 + \|X_0\|^l)e^{-\lambda t}, \tag{1.23}$$

where $\pi(\varphi) = \int_{-\infty}^{\infty} \varphi(x)\,\mathrm{d}\pi(x)$.

We should remark here that if the diffusion coefficient g is bounded and non-degenerate, the work of Has'minskii (1980) [56] implies that the dissipativity condition (1.20) ensures the ergodicity of the SDE (1.1). However, when $m \neq d$ or g is degenerate, we should check the minorization condition carefully [73].

1.2.3 Examples of non-globally Lipschitz SDEs

In this subsection, following the format of Hutzenthaler et al. [47], we give some important examples of non-globally Lipschitz SDEs arising from different research areas.

1. **Stochastic Ginzburg-Landau equation**

   This describes a phase transition from the theory of superconductivity [47]:

$$\mathrm{d}X_t = \left(\left(\eta + \frac{1}{2}\sigma^2\right)X_t - \lambda X_t^3\right)\mathrm{d}t + \sigma X_t\,\mathrm{d}W_t, \quad X_0 = x_0 \in (0,\infty), \tag{1.24}$$

   where $\eta \in [0,\infty)$, $\lambda, \sigma \in (0,\infty)$. It has analytical solution:

$$X_t = \frac{x_0 \exp(\eta t + \sigma W_t)}{\sqrt{1 + 2x_0^2\lambda \int_0^t \exp(2\eta s + 2\sigma W_s)ds}}. \tag{1.25}$$

   The drift coefficient is one-sided Lipschitz continuous and satisfies the one-sided linear growth and dissipativity conditions. The diffusion coefficient is globally Lipschitz continuous.

2. **Stochastic Verhulst equation**

   This is a model for a population with competition between individuals [47]:

$$\mathrm{d}X_t = \left(\left(\eta + \frac{1}{2}\sigma^2\right)X_t - \lambda X_t^2\right)\mathrm{d}t + \sigma X_t\,\mathrm{d}W_t, \quad X_0 = x_0 \in (0,\infty), \tag{1.26}$$

   where $\eta, \lambda, \sigma \in (0,\infty)$. It has analytical solution:

$$X_t = \frac{x_0 \exp(\eta t + \sigma W_t)}{1 + x_0\lambda \int_0^t \exp(\eta s + \sigma W_s)ds}. \tag{1.27}$$

   The drift coefficient is locally Lipschitz continuous and the diffusion coefficient is globally Lipschitz continuous. The drift coefficient satisfies the one-sided linear growth and dissipativity conditions since X_t is positive.

3. **Feller diffusion with logistic growth**

   The branching process with logistic growth is a stochastic Verhulst equation with Feller noise [47]:

$$\mathrm{d}X_t = \lambda X_t(K - X_t)\,\mathrm{d}t + \sigma\sqrt{X_t}\,\mathrm{d}W_t, \quad X_0 = x_0 \in (0,\infty), \tag{1.28}$$

   where $\lambda, K, \sigma \in (0,\infty)$. The drift coefficient is locally Lipschitz continuous and satisfies a one-sided linear growth condition and dissipativity condition provided X_t is positive, but the diffusion coefficient is only Hölder continuous.

4. **Ohta-Kimura model**

This is a population model with two types, fixed total population and random selection [45]:

$$\mathrm{d}X_t = \sigma X_t(1 - X_t)\,\mathrm{d}W_t, \quad X_0 = x_0 \in (0, 1), \tag{1.29}$$

where $\sigma \in (0, \infty)$. The diffusion coefficient is locally Lipschitz continuous but does not satisfy the one-sided linear growth condition. However, we can perform a transformation by taking $Y_t = \log X_t - \log(1 - X_t)$, which yields that

$$\mathrm{d}Y_t = \frac{\sigma^2}{2}\frac{e^{Y_t} - 1}{e^{Y_t} + 1}\,\mathrm{d}t + \sigma\,\mathrm{d}W_t, \tag{1.30}$$

and this SDE has a globally Lipschitz drift and a constant diffusion coefficient.

5. **Volatility Process**

This is a general model for a stochastic volatility process [46]:

$$\mathrm{d}X_t = [\delta + \gamma X_t - \alpha(X_t)^a]\ \mathrm{d}t + \beta(X_t)^b\,\mathrm{d}W_t, \tag{1.31}$$

where $a \in [1, \infty)$, $b \in [\frac{1}{2}, \infty)$, $\alpha, \beta \in (0, \infty)$, $\gamma \in \mathbb{R}$, $\delta \in [0, \infty)$, satisfying

$$a + 1 \geq 2b; \text{ if } b = \frac{1}{2},\ \delta \geq \frac{\beta^2}{2}. \tag{1.32}$$

The drift coefficient is locally Lipschitz continuous, but the diffusion coefficient is Hölder continuous and it does not usually satisfy the one-sided linear growth condition for general a and b.

6. **Stochastic van der Pol oscillator**

This describes state oscillation [46]:

$$\begin{aligned} \mathrm{d}X_t^{(1)} &= X_t^{(2)}\,\mathrm{d}t, \\ \mathrm{d}X_t^{(2)} &= \left[\alpha\left(\gamma - (X_t^{(1)})^2\right)X_t^{(2)} - \delta X_t^{(1)}\right]\,\mathrm{d}t + \beta\,\mathrm{d}W_t, \end{aligned} \tag{1.33}$$

where $\alpha, \gamma, \delta, \beta \in (0, \infty)$. The drift coefficient is not one-sided Lipschitz continuous but satisfies the locally Lipschitz condition and the one-sided linear growth condition. The diffusion coefficient is constant.

7. **Stochastic Duffing-van der Pol oscillator**

   The Duffing equation is a further model for an oscillator. The Duffing-van der Pol equation unifying both the Duffing equation and the van der Pol equation has been used in certain aeroelasticity problems [46]:

$$\begin{aligned} \mathrm{d}X_t^{(1)} &= X_t^{(2)}\,\mathrm{d}t, \\ \mathrm{d}X_t^{(2)} &= \Big[\alpha_1 X_t^{(1)} - \alpha_2 X_t^{(2)} - \alpha_3 X_t^{(2)} (X_t^{(1)})^2 - (X_t^{(1)})^3\Big]\,\mathrm{d}t \\ &\qquad + \beta_1 X_t^{(1)}\,\mathrm{d}W_t^{(1)} + \beta_2 X_t^{(2)}\,\mathrm{d}W_t^{(2)} + \beta_3\,\mathrm{d}W_t^{(3)}, \end{aligned} \tag{1.34}$$

   where $\alpha_1, \alpha_2, \alpha_3, \beta_1, \beta_2, \beta_3 \in \mathbb{R}$. The drift coefficient is not one-sided Lipschitz continuous and even fails to satisfy the one-sided linear growth condition. The diffusion coefficient is globally Lipschitz continuous.

8. **Stochastic Lorenz equation**

   This is a three-dimensional system modelling convection rolls in the atmosphere [46]:

$$\begin{aligned} \mathrm{d}X_t^{(1)} &= \Big[\alpha_1 X_t^{(2)} - \alpha_1 X_t^{(1)}\Big]\,\mathrm{d}t + \beta_1 X_t^{(1)}\,\mathrm{d}W_t^{(1)}, \\ \mathrm{d}X_t^{(2)} &= \Big[\alpha_2 X_t^{(1)} - X_t^{(2)} - X_t^{(1)} X_t^{(3)}\Big]\,\mathrm{d}t + \beta_2 X_t^{(2)}\,\mathrm{d}W_t^{(2)}, \\ \mathrm{d}X_t^{(3)} &= \Big[X_t^{(1)} X_t^{(2)} - \alpha_3 X_t^{(3)}\Big]\,\mathrm{d}t + \beta_3 X_t^{(3)}\,\mathrm{d}W_t^{(3)}, \end{aligned} \tag{1.35}$$

   where $\alpha_1, \alpha_2, \alpha_3, \beta_1, \beta_2, \beta_3 \in \mathbb{R}$. The drift coefficient is not one-sided Lipschitz continuous but satisfies the global one-sided linear growth condition. The diffusion coefficient is globally Lipschitz continuous.

9. **Stochastic Brusselator in the well-stirred case**

   This is a model for a trimolecular chemical reaction [46]:

$$\begin{aligned} \mathrm{d}X_t^{(1)} &= \Big[\delta - (\alpha+1) X_t^{(1)} + X_t^{(2)} (X_t^{(1)})^2\Big]\,\mathrm{d}t + g_1(X_t^{(1)})\,\mathrm{d}W_t^{(1)}, \\ \mathrm{d}X_t^{(2)} &= \Big[\alpha X_t^{(1)} - X_t^{(2)} (X_t^{(1)})^2\Big]\,\mathrm{d}t + g_2(X_t^{(2)})\,\mathrm{d}W_t^{(2)}, \end{aligned} \tag{1.36}$$

   where $\delta, \alpha \in \mathbb{R}$. The drift coefficient is not one-sided Lipschitz continuous and even fails to satisfy the one-sided linear growth condition. The diffusion coefficient g is usually globally Lipschitz continuous.

10. **Wright Fisher diffusion**

The Wright-Fisher family of diffusion processes is a class of evolutionary models widely used in population genetics, with applications also in finance and Bayesian statistics:

$$\mathrm{d}X_t = (a - bX_t)\,\mathrm{d}t + \gamma\sqrt{|X_t(1-X_t)|}\,\mathrm{d}W_t, \quad X_0 = x_0 \in (0,1), \tag{1.37}$$

where $a, b, \gamma \in (0, \infty)$ and if $2a/\gamma^2 \in [1, \infty)$, $2(b-a)/\gamma^2 \in [1, \infty)$. This SDE has a unique strong solution with

$$\mathbb{P}\left[X_t \in (0,1),\, t > 0\right] = 1. \tag{1.38}$$

The drift coefficient is globally Lipschitz continuous but the diffusion coefficient is not even locally Lipschitz continuous, though it satisfies the one-sided linear growth condition. However, we can perform a transformation by taking $Y_t = \log X_t - \log(1 - X_t)$, which yields that

$$\begin{aligned} \mathrm{d}Y_t = &\left(2a - b + (a - \frac{1}{2}\gamma^2)e^{-Y_t} + (a - b + \frac{1}{2}\gamma^2)e^{Y_t}\right)\mathrm{d}t \\ &+ \gamma\left(e^{\frac{Y_t}{2}} + e^{-\frac{Y_t}{2}}\right)\mathrm{d}W_t. \end{aligned} \tag{1.39}$$

By this transformation, we can see that the condition for a, b, γ can ensure that the coefficient of e^{Y_t} is negative and the coefficient of e^{-Y_t} is positive and satisfies the one-sided linear growth condition.

11. **Constant Elasticity of Variance (CEV)**

The CEV model is a stochastic volatility model, which attempts to capture stochastic volatility and the leverage effect. The model is widely used by practitioners in the financial industry, especially for modelling equities and commodities [1]:

$$\mathrm{d}X_t = \mu X_t\,\mathrm{d}t + \sigma(X_t^+)^p\,\mathrm{d}W_t, \tag{1.40}$$

where $p \in [\frac{1}{2}, 1), \mu \in \mathbb{R}$, $\sigma \in (0, \infty)$ and X_t^+ is the positive part of X_t. It features a non-Lipschitz diffusion coefficient and gets absorbed at zero with a positive probability.

12. **Stochastic Langevin equation**

The stochastic Langevin SDEs describe the position q and momentum p of a particle of unit mass in a damped-driven Hamiltonian system [78], namely

$$\begin{aligned} \mathrm{d}q &= p\,\mathrm{d}t, \\ \mathrm{d}p &= -\gamma p\,\mathrm{d}t - \nabla F(q)\,\mathrm{d}t + \sigma\,\mathrm{d}W_t, \end{aligned} \tag{1.41}$$

where $\gamma, \sigma \in (0, \infty)$. There is another special class of Langevin equations described by its invariant measure:

$$\mathrm{d}X_t = \frac{1}{2}\nabla \log \pi(X_t)\,\mathrm{d}t + \mathrm{d}W_t, \tag{1.42}$$

where π is the probability density function of the invariant measure.

13. **FENE model**

FENE stands for the Finitely Extensible Nonlinear Elastic model of a long-chained polymer. It simplifies the chain of monomers by connecting a sequence of beads with nonlinear springs. The spring force law is governed by an inverse Langevin function or approximated by the Warner's relationship [35]:

$$\mathrm{d}X_t = -\frac{4\mu}{1 - \|X_t\|^2}\,X_t\,\mathrm{d}t + 2\,\mathrm{d}W_t, \quad \|X_0\| < 1, \tag{1.43}$$

where $\mu \in (0, \infty)$ is a constant and W_t is a 3-dimensional Brownian motion. The drift term ensures that for all $t \in (0, \infty)$ it holds that $\|X_t\| < 1$ as is shown in [6] and satisfies $\langle x, f(x)\rangle \leq 0$.

1.3 Numerical methods

Globally Lipschitz continuity and linear growth on both the drift coefficient f and the diffusion coefficient g imply the stability and strong convergence of the explicit Euler-Maruyama scheme (1.4). It has been shown to be divergent in the strong L^p-sense and the weak sense to the exact solution for super-linearly growing coefficients by Hutzenthaler, Jentzen & Kloeden in [47]. For example, for the scalar SDE

$$\mathrm{d}X_t = -X_t^3\,\mathrm{d}t + \mathrm{d}W_t, \tag{1.44}$$

it has been shown that for any $p \in [1, \infty)$ it holds that

$$\lim_{h\to 0} \mathbb{E}\left[\|X_T - \widehat{X}_T\|^p\right] = \infty. \tag{1.45}$$

In the following subsections, we review the existing numerical methods for non-globally Lipschitz SDEs with uniform timesteps, the variable timestep schemes for SDEs and the numerical methods for ergodic SDEs.

1.3.1 Uniform timestep

A good fundamental work on the numerical solutions for non-globally Lipschitz SDEs is an instructive conditional result by Higham, Mao & Stuart in [42]: by assuming the boundedness of the p-th moment of the exact solution and the numerical solution, strong convergence of Euler-type schemes for locally Lipschitz coefficients has been proved. With the one-sided Lipschitz condition and polynomial growth condition on drift, and Lipschitz condition on diffusion coefficient, they also show the stability and the standard strong convergence order for two implicit schemes:

The split-step backward Euler method (SSBE)is

$$\begin{aligned} \widehat{X}^*_{nh} &= \widehat{X}_{nh} + f(\widehat{X}^*_{nh})\,h, \\ \widehat{X}_{(n+1)h} &= \widehat{X}^*_{nh} + g(\widehat{X}^*_{nh})\,\Delta W_n, \end{aligned} \tag{1.46}$$

and the drift-implicit backward Euler method (BEM) is

$$\widehat{X}_{(n+1)h} = \widehat{X}_{nh} + f(\widehat{X}_{(n+1)h})\,h + g(\widehat{X}_{nh})\,\Delta W_n. \tag{1.47}$$

Mao & Szpruch [71] proved the strong convergence of BEM for the SDEs with dissipative drift and diffusion coefficient with polynomial growth. In [70], they also proved that the implicit θ-Euler method

$$\widehat{X}_{(n+1)h} = \widehat{X}_{nh} + \theta f(\widehat{X}_{(n+1)h})\,h + (1-\theta) f(\widehat{X}_{nh})\,h + g(\widehat{X}_{nh})\,\Delta W_n, \tag{1.48}$$

converges strongly for $\frac{1}{2} \leq \theta \leq 1$ under more general conditions, which permit a non-globally Lipschitz diffusion coefficient.

However, except for some special cases, implicit methods can require significant additional computational costs, especially for multi-dimensional SDEs; therefore, a stable explicit method is desired. Milstein & Tretyakov proposed a general approach, which discards approximate paths that cross a sphere with a sufficiently large radius R [77]. However, it is not easy to quantify the errors due to R. The explicit tamed Euler method proposed by Hutzenthaler, Jentzen & Kloeden [48] is

$$\widehat{X}_{(n+1)h} = \widehat{X}_{nh} + \frac{f(\widehat{X}_{nh})}{1+C\,h\|f(\widehat{X}_{nh})\|}\,h + g(\widehat{X}_{nh})\,\Delta W_n, \tag{1.49}$$

for some fixed constant $C \in (0, \infty)$. They prove both stability and the standard order $\frac{1}{2}$ strong convergence. This approach has been extended to the tamed Milstein method by Wang & Gan [105], proving order 1 strong convergence for SDEs with commutative noise, and by Sabanis [90] to the case of superlinearly growing diffusion coefficient satisfying a Khasminskii-type condition. Next, Mao [68] proposes a truncated Euler method which has the form

$$\begin{aligned}\widehat{X}_{(n+1)h} \;&= \widehat{X}_{nh} \;+\; f\left(\min(K\|\widehat{X}_{nh}\|^{-1},1)\,\widehat{X}_{nh}\right)h \\ &\quad+\; g\left(\min(K\|\widehat{X}_{nh}\|^{-1},1)\,\widehat{X}_{nh}\right)\Delta W_n.\end{aligned} \tag{1.50}$$

By making K a function of h, strong convergence is proved for SDEs satisfying the Khasminskii-type condition; in [69] it is proved the order of convergence is arbitrarily close to $\frac{1}{2}$. This approach has also been extended to the truncated Milstein method by Guo et al. [40] for SDEs with commutative noise.

Finally, we mention the projected schemes by Beyn, Isaak & Kruse [9, 10], the balanced schemes by Tretyakov & Zhang [99], the semi-tamed Euler method by Zong, Wu & Huang [107], and the stopped Euler method by Liu & Mao [66]. We refer to [7] for a comprehensive summary of the numerical methods for the SDEs and SPDEs with superlinearly growing nonlinearity.

1.3.2 Adaptive timestep

There have been two main kinds of variable timestep schemes for SDEs.

One, following the adaptive idea in ODE discretizations, is to choose the step size by estimating the local error for each time step. Gaines & Lyons [28] show that to guarantee the convergence of this kind of variable timestep scheme, a strong order of at least one is needed. They estimate the error propagation by running a sample path with a timestep which is small enough, and then do the approximation by halving and doubling the timestep based on the estimated local error. Mauthner [74] uses the stochastic Runge-Kutta method (SRK) and estimates the local error by the difference between the approximations of the fine and coarse paths. Burrage & Burrage [12] extend the halving and doubling strategy to general variable timestep and Burrage, Herdiana & Burrage [13] introduce proportional integral control to estimate the stepsize. Valinejad & Hosseini [101] propose two local error estimates based on drift and diffusion coefficients respectively.

All the schemes above show point-wise convergence in the sense of the final point value $\|\widehat{X}_T - X_T\| \to 0$, which is relatively stronger than strong convergence we defined before. The reason is that they are mainly concerned with the Stratonovich SDEs to describe the evolution of a physical system and need to investigate the approximation for a single path with given Brownian path.

Lamba [59], Lamba & Seaman [61], and Lamba, Mattingly & Stuart [60] continue to use the local error estimation but analyse the scheme in the mean-square sense. Römisch & Winkler [87] estimate the local error for all the paths simultaneously to choose the timestep but use a semi-implicit method to treat the general drift and small diffusion. Ilie, Jackson & Enright [50] develop a cheaper way to estimate the local error for globally Lipschitz coefficients.

All of this kind of schemes based on local error estimation are hard to extend to MLMC. Besides, the main difficulty for this kind of scheme is that, when rejection of step size is allowed, the solution should remain on the same Brownian path, otherwise a bias will be introduced. One approach to solving it is to introduce the a Brownian bridge, which increases the complexity. However, from the view point of the Itô integral, it is not reasonable to choose the current timestep based on future information.

Another kind of variable timestep scheme is setting the timestep in advance using all the available information and avoiding the rejection. Hofmann, Müller–Gronbach & Ritter [43, 44] introduce a conditional Hölder constant and set the timestep to be proportional to the inverse of the constant using the Euler and Milstein scheme respectively. Müller–Gronbach [79] achieves the asymptotically optimal pointwise approximation of SDEs. However, all three schemes assume the globally Lipschitz continuity of the coefficients. Kelly & Lord [55] propose an adaptive strategy for the timestep function, but the scheme is not a pure adaptive scheme but an adaptive tamed scheme, which imposes a lower bound for the timestep and uses a tamed Euler scheme in extreme cases.

1.3.3 Ergodic methods

For ergodic SDEs, evaluating the expectation of some function $\varphi(x)$ with respect to that invariant measure π is of great interest in mathematical biology, physics and Bayesian inference in statistics. For any $\varphi \in L^1(\pi)$, it holds that

$$\pi(\varphi) = \int \varphi(x) \, \mathrm{d}\pi(x) = \lim_{t \to \infty} \mathbb{E}\left[\varphi(X_t)\right]. \tag{1.51}$$

Several different methodologies have been developed to estimate it.

First Approach

First, we can compute the probability density function $\rho(x)$ of π by solving the corresponding stationary Fokker-Planck equation, see [91]. However, the stationary Fokker-Planck equation is a partial differential equation (PDE) and its numerical solution becomes extremely expensive when the dimension of the PDE becomes large.

Second Approach

The second approach is based on the ergodicity of the SDEs, that is

$$\lim_{T\to\infty} \frac{1}{T}\int_0^T \varphi(X_t)\,\mathrm{d}t \;=\; \pi(\varphi),\;\; a.s., \tag{1.52}$$

where the limit does not depend on initial value x_0. This approach uses discretized numerical schemes to approximate the SDEs and requires the numerical solution $\widehat{X}_t$ to preserve the ergodicity. In practice, we can choose a sufficiently large N and compute

$$\frac{1}{N}\sum_{n=1}^{N}\varphi(\widehat{X}_{nh}), \tag{1.53}$$

where $\widehat{X}_{nh}$ is the numerical solution at the nth discretized time point using an ergodic method with a uniform timestep h. Under the dissipativity condition (1.20) together with the globally Lipschitz condition for f, Talay [93] shows the standard weak convergence order for the Milstein method

$$\lim_{N\to\infty}\frac{1}{N}\sum_{n=1}^{N}\varphi(\widehat{X}_{nh}) = \int \varphi(x)\,\mathrm{d}\pi(x) + O(h). \tag{1.54}$$

Roberts & Tweedie in [86] analyse the ergodicity of the unadjusted Langevin algorithm for the Langevin equation, which has a uniform diffusion coefficient and satisfies the dissipativity condition (1.20). This scheme corresponds to the standard Euler-Maruyama method

$$\widehat{X}_{(n+1)h} = \widehat{X}_{nh} + f(\widehat{X}_{nh})\,h + \Delta W_n, \tag{1.55}$$

using a uniform timestep of size h with Brownian increments ΔW_n. The paper shows that the numerical solution is not ergodic when f has a polynomial degree larger than 1. A Metropolis-adjusted Langevin algorithm (MALA) is introduced by adding an accept/reject step to correct the empirical distribution to the target distribution, but the numerical solutions are still not exponentially ergodic for non-linear drift f. For non-globally Lipschitz drift, they propose without proof the Metropolis-adjusted Langevin truncated algorithm (MALTA) by bounding the drift with a large positive constant D.

For globally Lipschitz SDEs satisfying the dissipativity condition (1.20), the standard Euler-Maruyama method is shown in [73] to inherit ergodicity provided the timesteps are sufficiently small. However, the standard Euler-Maruyama method and the Milstein method fail to be stable for non-globally Lipschitz ergodic SDEs. The split-step backward Euler method (1.46) and the drift-implicit backward Euler method (1.47) are proved to be ergodic in [73]. Talay [94] showed the ergodicity of the implicit Euler method for non-globally Lipschitz stochastic Hamiltonian systems.

Under the same conditions, Hansen in [41] considers the local linearization of the drift coefficient, that is the first-order Taylor approximation of $\tilde{X}_t$: given $\tilde{X}_{nh} = x$, for any $t \in [nh, (n+1)h)$,

$$\mathrm{d}\tilde{X}_t = (f(x) + \nabla f(x)(\tilde{X}_t - x))\,\mathrm{d}t + \mathrm{d}W_t, \tag{1.56}$$

which is an Ornstein-Uhlenbeck diffusion and we can calculate the analytical conditional distribution of $\tilde{X}_{(n+1)h}$. However, this analytical treatment only applies when the diffusion coefficient is uniform.

An adaptive timestepping algorithm proposed by Lamba, Mattingly & Stuart in [60] chooses the step size by halving or doubling based on local error estimation and a user-input tolerance τ. More precisely,

$$\begin{aligned} \widehat{X}^*_{t_n} &= \widehat{X}_{t_n} + f(\widehat{X}_{t_n})\,h_n \\ \widehat{X}_{t_{n+1}} &= \widehat{X}^*_{t_n} + g(\widehat{X}_{t_n})\,\Delta W_n \end{aligned} \tag{1.57}$$

where $h_n = 2^{-k_n} h_{\max}$ satisfies

$$h_n \leq \min(2h_{n-1}, h_{\max}), \quad k_n = \min(k \in \mathbb{Z} : \ |f(\widehat{X}^*_{t_n}) - f(\widehat{X}_{t_n})| \leq \tau). \tag{1.58}$$

This scheme is proved to preserve the ergodicity of the original SDE under the dissipativity condition (1.20) and boundedness and invertibility of the diffusion coefficient g. Lemaire [64] considers an infinite time interval under the dissipativity condition generated by a general Lyapunov function using a timestep with an upper bound which decreases towards zero over time, and proves convergence of the empirical distribution to the invariant distribution of the SDE. Pagès &

Panloup [82] extend this scheme to a weighted multilevel estimator using the idea of multilevel Richardson-Romberg extrapolation introduced in [65].

Brosse et al. [11] introduce the tamed unadjusted Langevin method, which extends the finite time tamed Euler method to an infinite time interval. With the dissipativity condition, the convergence to the invariant measure in V-total variation norm and Wasserstein distance are proved.

Third Approach

Finally, without requiring the ergodicity of the scheme, for exponentially ergodic SDEs, we can choose a sufficiently large T such that

$$|\mathbb{E}[\varphi(X_T)] - \pi(\varphi)| \leq \varepsilon. \tag{1.59}$$

Then, for this fixed T, we can use of all the methods mentioned in the previous subsections to estimate $\mathbb{E}[\varphi(X_T)]$. Milstein & Tretyakov [78] analyse the error of this kind of approach based on their quasi-symplectic method. In practice, a suitable choice of initial data is important because the transition period to a sufficient proximity of the equilibrium can be rather long. Therefore, running a small number of pioneer paths can be employed to obtain a good initial distribution for the overwhelming majority of simulations.

We should remark here that the first PDE approach is far too expensive in high dimensions. The second time-averaging approach (1.52) requires the numerical methods to preserve the ergodicity but the third approach does not. The length of the time interval $[0, T]$ used in the time-averaging approach is much longer than the third approach, for it needs not only to ensure that the distribution of $\widehat{X}_T$ is sufficiently close to the invariant measure π, but also to guarantee a small variance for the average. Therefore, the second approach needs to simulate a single long path but the third one needs to simulate a lot of relatively short paths, which allows multilevel Monte Carlo (MLMC) and parallel computing techniques to be employed. One important concern about the third approach is that in the numerical analysis of existing algorithms in the case of a finite time interval $[0, T]$, the strong error increases exponentially as T increases. This is not acceptable when we need to simulate for a much larger T and it will be a key concern in this thesis. The final issue about the second and third approaches is how to choose a good T such that the weak error is bounded appropriately.

1.4 Overview of the thesis

In this thesis, we propose instead to use the standard explicit Euler-Maruyama method, but with an adaptive timestep h_n which is a function of the current approximate solution $\widehat{X}_{t_n}$. The idea of using an adaptive timestep comes from considering the divergence of the uniform timestep method for the SDE (1.44). When there is no noise, the requirement for the explicit Euler approximation of the corresponding ODE to have a stable monotonic decay is that its timestep satisfies $h < \widehat{X}_{t_n}^{-2}$. An intuitive explanation for the instability of the uniform timestep Euler-Maruyama approximation of the SDE is that there is always a very small probability of a large Brownian increment ΔW_n which pushes the approximation $\widehat{X}_{t_{n+1}}$ into the region $h > 2\, \widehat{X}_{t_{n+1}}^{-2}$ leading to an oscillatory super-exponential growth. Using an adaptive timestep avoids this problem.

In addition, we are concerned with strong convergence, not weak convergence, because our interest is in using the numerical approximation as part of a multi-level Monte Carlo (MLMC) computation [30, 32] for which the strong convergence properties are key in establishing the rate of decay of the variance of the multilevel correction. Usually, MLMC is used with a geometric sequence of time grids, with each coarse timestep corresponding to a fixed number of fine timesteps. However, it has been shown that it is not difficult to implement MLMC using the same driving Brownian path for the coarse and fine paths, even when they have no time points in common [35].

In chapter 2, it is proved that if the timestep is bounded appropriately, then over a finite time interval the numerical approximation is stable, and the expected number of timesteps is finite. Furthermore, the order of strong convergence is the same as usual, i.e. order 1/2 for SDEs with a non-uniform globally Lipschitz diffusion coefficient, and order 1 for Langevin SDEs with unit diffusion coefficient and a drift with sufficient smoothness.

Paper [35] also provides another motivation for this thesis, the analysis of the invariant distribution of Langevin equations with a drift $-\nabla V(X_t)$ where $V(x)$ is a potential function which comes from the modelling of molecular dynamics. [35] considers the FENE model (13) which in the case of a molecule with a single bond has a 3D potential $-\mu \log(1-\|x\|^2)$. Considerations of stability and accuracy lead to the use of a timestep of the form $\delta\,(1-\|\widehat{X}_n\|)^2/\max(2\mu, 36)$, for some $\delta \in (0,1]$. Because of this, we pay particular attention to the case of ergodic SDEs satisfying the dissipativity condition.

In chapter 3, we extend the adaptive scheme proposed to the ergodic SDEs with a drift which is not globally Lipschitz but contractive over an infinite time interval. The benefit of contractivity is that it ensures that two solutions to the SDE starting from different initial data but driven by the same Brownian motion, will come together exponentially. For numerical schemes, this means the error made on previous time steps will decay expoentially. By setting a suitable condition for h, we can show that, instead of an exponential bound, the numerical solution has a uniform bound with respect to T for both moments and the strong error. Then, MLMC methodology [30, 32] is employed and non-nested timestepping is used to construct an adaptive MLMC [35]. Following the idea of Glynn and Rhee [38] to estimate the invariant measure of some Markov chains, we introduce an adaptive MLMC algorithm for the infinite interval, in which each level ℓ has a different time interval length T_ℓ, to achieve a better computational performance. Note that using different time interval lengths allows us not to worry how to choose an appropriate T before simulation. The MLMC algorithm will automatically terminate at a level L with a sufficiently large T_L. With the contractivity, the optimal computational cost to achieve mean square error $O(\varepsilon^2)$ for the expectation with respect to the invariant measure of the ergodic SDEs is $O(\varepsilon^{-2})$, compared with $O(\varepsilon^{-3}|\log\varepsilon|)$ for the standard Monte Carlo method.

However, a larger class of SDEs satisfying the dissipativity condition does not satisfy the contractivity condition and instead only satisfies the one-sided Lipschitz condition. Without the contractivity, the strong error may increase exponentially with respect to T. Then the multilevel correction variances V_ℓ also increase exponentially, which, as shown in Theorem 4.2.5, increases the total computational cost to $O(\varepsilon^{-2-\frac{\kappa}{2\lambda^*}}|\log\varepsilon|)$, where κ is the Lyapunov exponent of the system and λ^* is the exponential convergence rate to the invariant measure. For some SDEs with a chaotic property, the Lyapunov exponent κ can be sufficiently large such that $\frac{\kappa}{2\lambda^*} \geq 1$ and MLMC loses its advantage over the standard Monte Carlo method.

In chapter 4, a change of measure technique is employed to deal with SDEs with f satisfying the one-sided Lipschitz condition : there exists a constant $\lambda \in (0,\infty)$ such that for any $x, y \in \mathbb{R}^m$ it holds that

$$\langle x-y, f(x)-f(y)\rangle \;\leq\; \lambda\, \|x-y\|^2. \tag{1.60}$$

and g is an identity matrix. The key feature of this class of SDEs, especially the chaotic SDEs, is that the behaviour of solutions is highly sensitive to initial conditions and the difference between the fine path and coarse path will increase exponentially. An intuitive way to avoid this kind of divergence is by adding a spring between the fine path and coarse path to draw them closer to each other.

Mathematically, instead of simulating the fine path and coarse path of the original SDEs, that is, in their separated path spaces with different measures

$$\begin{aligned} \mathbb{Q}^f: \quad \mathrm{d}X_t^f &= f(X_t^f)\,\mathrm{d}t + \mathrm{d}W_t^{\mathbb{Q}^f}, \\ \mathbb{Q}^c: \quad \mathrm{d}X_t^c &= f(X_t^c)\,\mathrm{d}t + \mathrm{d}W_t^{\mathbb{Q}^c}, \end{aligned} \tag{1.61}$$

we add a spring term with spring coefficient $S > \frac{\lambda}{2}$ for both fine path and coarse path for all $\ell \in \{1, 2, ..., L\}$, and simulate the fine path and coarse path in the same probability measure $\mathbb{P}$, that is,

$$\begin{aligned} \mathrm{d}Y_t^f &= S(Y_t^c - Y_t^f)\,\mathrm{d}t + f(Y_t^f)\,\mathrm{d}t + \mathrm{d}W_t^{\mathbb{P}}, \\ \mathrm{d}Y_t^c &= S(Y_t^f - Y_t^c)\,\mathrm{d}t + f(Y_t^c)\,\mathrm{d}t + \mathrm{d}W_t^{\mathbb{P}}. \end{aligned} \tag{1.62}$$

Girsanov's theorem implies that

$$\mathbb{E}^{\mathbb{Q}^f}[X_t^f] - \mathbb{E}^{\mathbb{Q}^c}[X_t^c] = \mathbb{E}^{\mathbb{P}}\left[Y_t^f \frac{\mathrm{d}\mathbb{Q}^f}{\mathrm{d}\mathbb{P}} - Y_t^c \frac{\mathrm{d}\mathbb{Q}^c}{\mathrm{d}\mathbb{P}}\right], \tag{1.63}$$

where $\frac{\mathrm{d}\mathbb{Q}^f}{\mathrm{d}\mathbb{P}}$ and $\frac{\mathrm{d}\mathbb{Q}^c}{\mathrm{d}\mathbb{P}}$ are the corresponding Radon-Nikodym derivatives of the measure $\mathbb{Q}^f$ on the fine path space and measure $\mathbb{Q}^c$ on the coarse path space with respect to the $\mathbb{P}$ measure in which we are simulating both paths. In practice, we will derive the Radon-Nikodym derivative exactly for the numerical solution instead of numerically approximating the derivatives above. In the new MLMC scheme, essentially, the fine path Y_t^f and coarse path Y_t^c share the same driving Brownian motion W_t in measure $\mathbb{P}$. Correspondingly, the Brownian motions for the original SDEs

$$\begin{aligned} \mathrm{d}W_t^{\mathbb{Q}^f} &= S(Y_t^c - Y_t^f)\,\mathrm{d}t + \mathrm{d}W_t^{\mathbb{P}}, \\ \mathrm{d}W_t^{\mathbb{Q}^c} &= S(Y_t^f - Y_t^c)\,\mathrm{d}t + \mathrm{d}W_t^{\mathbb{P}}, \end{aligned} \tag{1.64}$$

are slightly different in measure $\mathbb{P}$, which is different from the standard MLMC. The benefit of this change is that the difference between the new simulated SDEs satisfies

$$\mathrm{d}(Y_t^f - Y_t^c) = 2S(Y_t^c - Y_t^f)\,\mathrm{d}t + (f(Y_t^f) - f(Y_t^c))\,\mathrm{d}t, \tag{1.65}$$

and provided $S > \frac{\lambda}{2}$, Ito's formula and the one-sided Lipschitz condition (1.60) ensures that

$$\mathrm{d}\,\|Y_t^f - Y_t^c\|^2 \le 2(\lambda - 2S)\|Y_t^f - Y_t^c\|^2\,\mathrm{d}t, \tag{1.66}$$

which will recover the contractivity between the fine and coarse paths. Note that the choice of the simple form of the spring term is motivated by this intuitive explanation and makes it easy to prove that contractivity is recovered. It also works well in practice, but we do not claim it is optimal and further research is required to investigate and analyse possible improvements. Due to the contractivity, we can prove that the strong difference between the coarse and fine paths is uniformly bounded with respect to T. More importantly, we can show that, together with the Radon-Nikodym derivatives, the variance of the new MLMC correction estimator increases only linearly in T, which is a great improvement compared with the exponential increase without the change of measure. The total computational cost can be reduced to $O(\varepsilon^{-2}|\log \varepsilon|^2)$, where the order is independent of the convergence rate λ^* of the original SDE and the Lyapunov exponent κ. We provide the numerical analysis only for the case of a globally Lipschitz drift but this scheme works well for SDEs with non-globally Lipschitz drift such as the stochastic Lorenz equation which is only locally one-sided Lipschitz. Furthermore, we extend this scheme to the computation of mean exit times for multi-dimensional SDEs and associated functionals corresponding to solutions to high-dimensional parabolic PDEs.

In chapter 5, continuing with the estimation for the expectation with respect to the invariant measure, we consider the computation of the sensitivities of the quantity with respect to the parameters of the chaotic system, which are of great interest and are helpful for calibrations and further considerations in real-world applications, such as aerodynamic shape optimization [51], statistics and weather forecast [96]. However, the computations of the sensitivities for both ODEs and SDEs are difficult due to the chaotic property. A small perturbation of the coefficient results in a large divergence of two solutions. Lea, Allen & Haine [62] use the Lorenz equation (ODE) as an example to illustrate the failure of the pathwise sensitivity method due to the blow-up of the variation process. A similar problem remains for the stochastic Lorenz equation. The variation process may become not ergodic and blow up exponentially, even though the chaotic system itself is ergodic, see section 5.4 for detailed numerical results.

Similarly to chapter 4, by introducing a spring term between the original and perturbated SDEs, we derive a new pathwise sensitivity estimator by importance sampling. The variance of the new estimator increases only linearly in time T, compared with the exponential increase of the standard pathwise estimator. We compare our estimator with the Malliavin estimator and extend both of them to the Multilevel Monte Carlo method with the change of measure technique introduced in the previous chapter, which further improves the computational efficiency. Lastly, we also consider using this estimator for the SDE with small uniform diffusion coefficient to approximate the sensitivities of the invariant measure of chaotic ODEs using Richardson-Romberg Extrapolation.

Finally, in chapter 6, we summarize the thesis and discuss possible directions for future research.

Chapter 2

Finite time interval

This chapter is an extended version of the Arxiv paper [22]. In this chapter, we introduce our adaptive scheme on a finite time interval. The rest of this chapter is organised as follows. Section 1 states the main theorems and proves some minor lemmas. Section 2 has a number of example applications, many from [46], illustrating how suitable adaptive timestep functions can be determined. It also presents some numerical results comparing the performance of the adaptive Euler-Maruyama method to other methods. Section 3 contains the proofs of the three main theorems. Finally, section 4 concludes this chapter.

2.1 Adaptive algorithm and theoretical results

2.1.1 Adaptive Euler-Maruyama method

The adaptive Euler-Maruyama discretisation is

$$t_{n+1} = t_n + h_n, \qquad \widehat{X}_{t_{n+1}} = \widehat{X}_{t_n} + f(\widehat{X}_{t_n})\, h_n + g(\widehat{X}_{t_n})\, \Delta W_n, \tag{2.1}$$

where $h_n = h(\widehat{X}_{t_n})$ and $\Delta W_n = W_{t_{n+1}} - W_{t_n}$, and there is fixed initial datum $t_0 = 0$, $\widehat{X}_0 = x_0$.

One key point in the analysis is to prove that t_n strictly increases without bound as n increases. More specifically, the analysis proves that for any $T \in (0, \infty)$, almost surely for each path there exists an integer $N \in \{0, 1, 2, ...\}$ such that $t_N \geq T$.

We use the notation $\underline{t} = \max\{t_n : t_n \leq t\}$, $n_t = \max\{n : t_n \leq t\}$ for the nearest time point before time t, and its index.

We denote the piecewise constant interpolant process $\overline{X}_t = \widehat{X}_{\underline{t}}$ and also use the standard continuous interpolant [58] as that satisfying

$$\widehat{X}_t = \widehat{X}_{\underline{t}} + f(\widehat{X}_{\underline{t}})(t - \underline{t}) + g(\widehat{X}_{\underline{t}})(W_t - W_{\underline{t}}), \tag{2.2}$$

so that $\widehat{X}_t$ is the solution of the SDE

$$\mathrm{d}\widehat{X}_t = f(\widehat{X}_{\underline{t}})\,\mathrm{d}t + g(\widehat{X}_{\underline{t}})\,\mathrm{d}W_t = f(\overline{X}_t)\,\mathrm{d}t + g(\overline{X}_t)\,\mathrm{d}W_t. \tag{2.3}$$

In the following subsections, we state the key results on stability and strong convergence, and related results on the number of timesteps, introducing various assumptions as required for each. The main proofs are deferred to Section 2.3.

2.1.2 Stability

Assumption 2.1.1 (Locally Lipschitz and linear growth)**.** *Assume that f and g are both locally Lipschitz. Furthermore, assume that there exist constants $\alpha, \beta \in [0, \infty)$ such that for all $x \in \mathbb{R}^m$, f satisfies the one-sided linear growth condition*

$$\langle x, f(x) \rangle \leq \alpha \|x\|^2 + \beta, \tag{2.4}$$

and g satisfies the linear growth condition

$$\|g(x)\|^2 \leq \alpha \|x\|^2 + \beta. \tag{2.5}$$

Together, (2.4) and (2.5) imply the one-sided linear growth condition for f and g :

$$\langle x, f(x)\rangle + \tfrac{1}{2}\|g(x)\|^2 \leq \tfrac{3}{2}(\alpha\|x\|^2 + \beta), \tag{2.6}$$

which is a key assumption in the analysis of Mao & Szpruch [70] and Mao [68] for SDEs with volatilities which are not globally Lipschitz. However, in our analysis we choose to use this slightly stronger assumption than (2.6), which provides the basis for the following lemma on the stability of the SDE solution.

Lemma 2.1.1 (SDE stability)**.** *If the SDE satisfies Assumption 2.1.1, then for all $p \in (0, \infty)$ it holds that*

$$\mathbb{E}\left[\sup_{0\leq t\leq T} \|X_t\|^p\right] < \infty. \tag{2.7}$$

Proof of Lemma 2.1.1. The proof is given in Lemma 3.2 in [42]; the statement of that lemma makes stronger assumptions on f and g, corresponding to (2.17) and (2.18), but the proof only uses the conditions in Assumption 2.1.1. This completes the proof of Lemma 2.1.1. □

We now specify the critical assumption about the adaptive timestep.

Assumption 2.1.2 (Adaptive timestep)**.** *The adaptive timestep function $h : \mathbb{R}^m \to (0, \infty)$ is continuous and strictly positive, and there exist constants $\alpha, \beta \in (0, \infty)$ such that for all $x \in \mathbb{R}^m$, h satisfies the inequality that*

$$\langle x, f(x)\rangle + \tfrac{1}{2}\, h(x)\, \|f(x)\|^2 \leq \alpha\|x\|^2 + \beta. \tag{2.8}$$

Note that if another timestep function $h^\delta(x)$ is smaller than $h(x)$, then $h^\delta(x)$ also satisfies the Assumption 2.1.2. Note also that the form of (2.8), which is motivated by the requirements of the proof of the next theorem, is very similar to (2.4). Indeed, if (2.8) is satisfied then (2.4) is also true for the same values of α and β.

Theorem 2.1.1 (Finite time stability)**.** *If the SDE satisfies Assumption 2.1.1, and the timestep function h satisfies Assumption 2.1.2, then T is almost surely* *attainable (i.e. for $\omega \in \Omega$, $\mathbb{P}(\exists N(\omega) < \infty$ s.t. $t_{N(\omega)} \geq T) = 1$) and for all $p \in (0, \infty)$ there exists a constant $C_{p,T} \in (0, \infty)$ which depends solely on p, T and the constants α, β in Assumption 2.1.2, such that*

$$\mathbb{E}\left[\sup_{0 \leq t \leq T} \|\widehat{X}_t\|^p\right] < C_{p,T}. \tag{2.9}$$

Proof of Theorem 2.1.1. The proof is deferred to Section 2.3. □

To bound the expected number of timesteps, we require an assumption on how quickly $h(x)$ can approach zero as $\|x\| \to \infty$.

Assumption 2.1.3 (Timestep lower bound)**.** *There exist constants $\xi, \zeta, q \in (0, \infty)$ such that the adaptive timestep function satisfies the inequality*

$$h(x) \geq (\xi \|x\|^q + \zeta)^{-1}. \tag{2.10}$$

Given this assumption, we obtain the following lemma.

Lemma 2.1.2 (Bounded timestep moments)**.** *If the SDE satisfies Assumption 2.1.1, and the timestep function h satisfies Assumptions 2.1.2 and 2.1.3, then for all $p \in (0, \infty)$ it holds that*

$$\mathbb{E}\left[N_T^p\right] < \infty, \tag{2.11}$$

where $N_T = \min\{n : t_n \geq T\}$ is the number of timesteps required by a path approximation.

Proof of Lemma 2.1.2. Assumption 2.1.3 implies that

$$N_T \leq 1 + T \sup_{0 \leq t \leq T} \frac{1}{h(\widehat{X}_t)} \leq 1 + T \left(\xi \sup_{0 \leq t \leq T} \|\widehat{X}_t\|^q + \zeta\right). \tag{2.12}$$

Combining this and Theorem 2.1.1 completes the proof of Lemma 2.1.2. □

2.1.3 Strong convergence

Standard strong convergence analysis for an numerical approximation with a uniform timestep h considers the limit $h \to 0$. This clearly needs to be modified when using an adaptive timestep, and we will instead consider a timestep function $h^\delta(x)$ controlled by a scalar parameter $\delta \in [0,1]$, and consider the limit $\delta \to 0$.

Given a timestep function $h(x)$ which satisfies Assumptions 2.1.2 and 2.1.3, ensuring stability as analysed in the previous section, there are two quite natural ways in which we might introduce δ to define $h^\delta(x)$:

$$\begin{aligned} h^\delta(x) &= \delta \min(T, h(x)), \\ h^\delta(x) &= \min(\delta T, h(x)). \end{aligned} \tag{2.13}$$

The first refines the timestep everywhere, while the latter concentrates the computational effort on reducing the maximum timestep, with $h(x)$ introduced to ensure stability when $\|\widehat{X}_t\|$ is large.

In our analysis, we will cover both possibilities by making the following assumption.

Assumption 2.1.4. *The timestep function h^δ satisfies the inequality*

$$\delta \min(T, h(x)) \le h^\delta(x) \le \min(\delta T, h(x)), \tag{2.14}$$

with h satisfying Assumption 2.1.3.

Given this assumption, we obtain the following theorem:

Theorem 2.1.2 (Strong convergence)**.** *If the SDE satisfies Assumption 2.1.1, and the timestep function h^δ satisfies Assumption 2.1.4 with h satisfying Assumption 2.1.2, then for all $p \in (0, \infty)$ it holds that*

$$\lim_{\delta \to 0} \mathbb{E}\left[\sup_{0 \le t \le T} \|\widehat{X}_t - X_t\|^p \right] = 0. \tag{2.15}$$

Proof of Theorem 2.1.2. The proof is essentially identical to the uniform timestep Euler-Maruyama analysis in Theorem 2.2 in [42] by Higham, Mao & Stuart.

The only change required by the use of an adaptive timestep is to note that

$$\widehat{X}_s - \overline{X}_s = f(\overline{X}_s)\,(s-\underline{s}) + g(\overline{X}_s)\,(W_s - W_{\underline{s}}) \tag{2.16}$$

and $s-\underline{s} < \delta T$ and $\mathbb{E}\left[\,\|W_s - W_{\underline{s}}\|^2 \mid \mathcal{F}_{\underline{s}}\,\right] = d\,(s-\underline{s})$. This completes the proof of Theorem 2.1.2. □

To prove an order of strong convergence requires new assumptions on f and g:

Assumption 2.1.5 (Lipschitz properties)**.** *There exists a constant* $\alpha \in (0,\infty)$ *such that* f *satisfies the one-sided Lipschitz condition: for any* $x, y \in \mathbb{R}^m$ *it holds that*

$$\langle x-y, f(x)-f(y)\rangle \leq \tfrac{1}{2}\alpha\|x-y\|^2, \tag{2.17}$$

and g *satisfies the globally Lipschitz condition: for any* $x, y \in \mathbb{R}^m$ *it holds that*

$$\|g(x)-g(y)\|^2 \leq \tfrac{1}{2}\alpha\|x-y\|^2. \tag{2.18}$$

In addition, f *satisfies the locally polynomial growth Lipschitz condition: there exist constants* $\gamma, \mu, q \in (0,\infty)$ *such that for any* $x, y \in \mathbb{R}^m$ *it holds that*

$$\|f(x)-f(y)\| \leq \left(\gamma\left(\|x\|^q + \|y\|^q\right) + \mu\right)\,\|x-y\|. \tag{2.19}$$

Note that setting $y=0$ ensures that

$$\begin{aligned} \langle x, f(x)\rangle &\leq \tfrac{1}{2}\alpha\|x\|^2 + \langle x, f(0)\rangle \leq \alpha\|x\|^2 + \tfrac{1}{2}\alpha^{-1}\|f(0)\|^2, \\ \|g(x)\|^2 &\leq 2\|g(x)-g(0)\|^2 + 2\|g(0)\|^2 \leq \alpha\|x\|^2 + 2\|g(0)\|^2. \end{aligned} \tag{2.20}$$

Hence, Assumption 2.1.5 implies Assumption 2.1.1, with the same α and an appropriate β.

Also, if the drift and volatility is differentiable, the following assumption is equivalent to Assumption 2.1.5, and usually easier to check in practice.

Assumption 2.1.6 (Lipschitz properties). *There exists a constant* $\alpha \in (0,\infty)$ *such that for all* $x, e \in \mathbb{R}^m$ *with* $\|e\|=1$*,* f *satisfies the one-sided Lipschitz condition*

$$\langle e, \nabla f(x)\, e\rangle \leq \tfrac{1}{2}\alpha, \tag{2.21}$$

and g *satisfies the globally Lipschitz condition*

$$\|\nabla g(x)\|^2 \leq \tfrac{1}{2}\alpha, \tag{2.22}$$

and in addition f *satisfies the locally polynomial growth Lipschitz condition* (2.19)*.*

Theorem 2.1.3 (Strong convergence order). *If the SDE satisfies Assumption 2.1.5, and the timestep function* h^δ *satisfies Assumption 2.1.4 with* h *satisfying Assumption 2.1.2, then for all* $p>0$ *there exists a constant* $C_{p,T} \in (0,\infty)$ *such that*

$$\mathbb{E}\left[\sup_{0\leq t\leq T} \|\widehat{X}_t - X_t\|^p\right] \leq C_{p,T}\, \delta^{p/2}. \tag{2.23}$$

Proof of Theorem 2.1.3. The proof is deferred to Section 2.3. □

Lemma 2.1.3 (Number of timesteps). *If the SDE satisfies Assumption 2.1.5, and the timestep function* $h^\delta(x)$ *satisfies Assumption 2.1.4, with* h *satisfying Assumption 2.1.2, then for all* $p \in (0,\infty)$ *there exists a constant* $c_{p,T} \in (0,\infty)$ *such that*

$$\mathbb{E}\,[N_T^p] \leq c_{p,T}\, \delta^{-p}. \tag{2.24}$$

Proof of Lemma 2.1.3. Equation (2.14) and Assumption 2.1.3 implies that

$$\begin{aligned} N_T &\leq 1 + T \sup_{0\leq t\leq T} \frac{1}{h^\delta(\widehat{X}_t)} \\ &\leq 1 + \delta^{-1}\, T \sup_{0\leq t\leq T} \max(h^{-1}(x), T^{-1}) \\ &\leq \delta^{-1}\, T \left(\xi \sup_{0\leq t\leq T} \|\widehat{X}_t\|^q + \zeta + (1+\delta)\, T^{-1}\right), \end{aligned} \tag{2.25}$$

since $h^\delta(x) \leq h(x)$ and $h^\delta(x)$ satisfies the requirements for stability. Combining this and Theorem 2.1.1 completes the proof of Lemma 2.1.3. □

The conclusion from Theorem 2.1.3 and Lemma 2.1.3 is that

$$\mathbb{E}\left[\sup_{0\leq t\leq T}\|\widehat{X}_t-X_t\|^p\right]^{1/p} \leq C_{p,T}^{1/p}\, c_{1,T}^{1/2}\, (\mathbb{E}\,[N_T])^{-1/2}, \tag{2.26}$$

which corresponds to order $\frac{1}{2}$ strong convergence when comparing the accuracy to the expected cost.

First order strong convergence is achievable for Langevin SDEs in which $m=d$ and g is the identity matrix I_m, but this requires stronger assumptions on the drift f.

Assumption 2.1.7 (Enhanced Lipschitz properties)**.** *There exists a constant* $\alpha \in (0,\infty)$ *such that for any* $x,y \in \mathbb{R}^m$, f *satisfies the one-sided Lipschitz condition*

$$\langle x-y, f(x)-f(y)\rangle \leq \tfrac{1}{2}\alpha\|x-y\|^2. \tag{2.27}$$

In addition, f *is differentiable, and* f *and* ∇f *satisfy the locally polynomial growth Lipschitz condition* (2.19)*.*

We now state the theorem on improved strong convergence.

Theorem 2.1.4 (Strong convergence for Langevin SDEs)**.** *If* $m = d$, $g \equiv I_m$, f *satisfies Assumption 2.1.7, and the timestep function* h^δ *satisfies Assumption 2.1.4 with* h *satisfying Assumption 2.1.2, then for all* $T,p \in (0,\infty)$ *there exists a constant* $C_{p,T} \in (0,\infty)$ *such that*

$$\mathbb{E}\left[\sup_{0\leq t\leq T}\|\widehat{X}_t-X_t\|^p\right] \leq C_{p,T}\,\delta^p. \tag{2.28}$$

Proof of Theorem 2.1.4. The proof is deferred to Section 2.3. □

Comment: first order strong convergence can also be achieved for a general $g(x)$ by using an adaptive timestep Milstein discretisation, provided ∇g satisfies an additional globally Lipschitz condition. A formal statement and proof of this is omitted as it requires a lengthy extension to the stability analysis. In addition, this numerical approach is only practical in cases in which a commutativity condition is satisfied and therefore there is no need to simulate the Lévy areas which the Milstein method otherwise requires [58].

2.2 Examples and numerical results

In this section we discuss a number of example SDEs with non-globally Lipschitz drift. In each case we comment on the applicability of the theory and a suitable choice for the adaptive timestep. We then present numerical results for three test cases which illustrate some key aspects.

2.2.1 Scalar SDEs

In each of the cases to be presented, the drift is of the form

$$f(x) \approx -\, c\, \mathrm{sign}(x)\, |x|^q, \quad \text{as } |x| \to \infty, \tag{2.29}$$

for some constants $c \in (0, \infty)$, $q \in (1, \infty)$. Therefore, as $|x| \to \infty$, the maximum stable timestep satisfying Assumption 2.1.2 corresponds to $\langle x, f(x)\rangle + \frac{1}{2} h(x)\, |f(x)|^2 \approx 0$ and hence $h(x) \approx 2|x|/|f(x)| \approx 2\, c^{-1} |x|^{1-q}$. A suitable choice for $h(x)$ and $h^\delta(x)$ is therefore

$$h(x) = \min\left(T, c^{-1}|x|^{1-q}\right), \quad h^\delta(x) = \delta\, h(x). \tag{2.30}$$

Stochastic Ginzburg-Landau equation (1.24)

This SDE is usually defined on the domain $(0, \infty)$, since if $X_0 \in (0, \infty)$, for all $t \in (0, \infty)$, $X_t \in (0, \infty)$. However, the numerical approximation is not guaranteed to remain strictly positive and the domain can be extended to $\mathbb{R}$ without any change to the SDE.

The drift and volatility satisfy Assumptions 2.1.1 and 2.1.5, and therefore all of the theory is applicable, with a suitable choice for $h^\delta(x)$, based on (2.29) and (2.30), being that

$$h^\delta(x) = \delta\, \min\left(T, \lambda^{-1} x^{-2}\right). \tag{2.31}$$

Stochastic Verhulst equation (1.26)

This SDE is defined on the domain $(0,\infty)$, but can be extended to $\mathbb{R}$ by modifying it to

$$\mathrm{d}X_t = \left(\left(\eta+\tfrac{1}{2}\sigma^2\right)X_t - \lambda\,|X_t|X_t\right)\mathrm{d}t + \sigma X_t\,\mathrm{d}W_t, \tag{2.32}$$

so that the drift is positive in the limit $x\to-\infty$.

The drift and volatility then satisfy Assumptions 2.1.1 and 2.1.5, and therefore all of the theory is applicable, with a suitable choice for $h^\delta(x)$, based on (2.29) and (2.30), being that

$$h^\delta(x) = \delta\,\min\left(T, \lambda^{-1}|x|^{-1}\right). \tag{2.33}$$

2.2.2 Multi-dimensional SDEs

With multi-dimensional SDEs there are two cases of particular interest. For SDEs with a drift which, for some $\beta\in(0,\infty)$ and sufficiently large $\|x\|$, satisfies the condition

$$\langle x, f(x)\rangle \le -\beta\,\|x\|\,\|f(x)\|, \tag{2.34}$$

one can take $\langle x, f(x)\rangle + \frac{1}{2}h(x)\,|f(x)|^2 \approx 0$ and therefore a suitable definition of $h(x)$ for large $\|x\|$ is that

$$h(x) = \min(T, \|x\|/\|f(x)\|). \tag{2.35}$$

For SDEs with a drift which does not satisfy the condition, but for which $\|f(x)\|\to\infty$ as $\|x\|\to\infty$, an alternative choice for large $\|x\|$ is to use

$$h(x) = \min(T, \gamma\,\|x\|^2/\|f(x)\|^2), \tag{2.36}$$

for some $\gamma\in(0,\infty)$. The difficulty in this case is choosing the best value for γ, taking into account both accuracy and cost.

Stochastic van der Pol oscillator (1.33)

This SDE can be put in the standard form by setting

$$f(x) = \begin{pmatrix} x_2 \\ \alpha\,(\mu - x_1^2)\, x_2 - \delta x_1 \end{pmatrix}, \quad g(x) = \begin{pmatrix} 0 \\ \beta \end{pmatrix}. \tag{2.37}$$

It follows that

$$\langle x, f(x)\rangle \;=\; -\alpha\, x_1^2\, x_2^2 + \alpha\,\mu\, x_2^2 + (1-\delta)\, x_1\, x_2 \;\le\; \left(\alpha\mu + \tfrac{1}{2}(1-\delta)\right) \|x\|^2. \tag{2.38}$$

Therefore the drift and volatility satisfy Assumption 2.1.1 and the numerical approximations will be stable if the maximum timestep is defined by (2.36).

However, it can be verified that $\langle e, \nabla f(x)\, e\rangle$ is not uniformly bounded for an arbitrary e such that $\|e\| = 1$, and therefore the drift does not satisfy the one-sided Lipschitz condition. Hence the stability and strong convergence theory in this chapter is applicable, but not the theorems on the order of convergence. Nevertheless, numerical experiments exhibit first order strong convergence, which is consistent with the fact that the volatility in uniform, so it seems there remains a gap here in the theory.

Stochastic Lorenz equation (1.35)

This SDE can be put in the standard form by setting

$$f(x) = \begin{pmatrix} \alpha_1(x_2 - x_1) \\ \alpha_2 x_1 - x_2 - x_1 x_3 \\ x_1 x_2 - \alpha_3 x_3 \end{pmatrix}, \quad g(x) = \begin{pmatrix} \beta_1 x_1 & 0 & 0 \\ 0 & \beta_2 x_2 & 0 \\ 0 & 0 & \beta_3 x_3 \end{pmatrix}. \tag{2.39}$$

The diffusion coefficient is globally Lipschitz, and since $\langle x, f(x)\rangle$ consists solely of quadratic terms, the drift satisfies the one-sided linear growth condition. Noting that $\|f\|^2 \approx x_1^2(x_2^2 + x_3^2) < \|x\|^4$ as $\|x\| \to \infty$, an appropriate maximum timestep is

$$h(x) = \min(T, \gamma\|x\|^{-2}), \tag{2.40}$$

for any $\gamma \in (0,\infty)$. However, the drift does not satisfy the one-sided Lipschitz condition, and therefore the theory on the order of strong convergence is not applicable.

FENE model (1.43)

The unusual feature of the FENE model is that the potential $V(x)$ becomes infinite for finite values of x. In the simplest case of a molecule with a single bond, x is three-dimensional and $V(x)$ takes the form $V(x) = -\log(1-\|x\|^2)$. The SDE is defined on $\|x\| < 1$, with the drift term ensure that $\|X_t\| < 1$ for all $t \in (0,\infty)$. Also, it can be verified that $\langle x, f(x)\rangle \leq 0$.

Because the SDE is not defined on all of $\mathbb{R}^3$, the theory in this chapter is not applicable. However, it was one of the original motivations for the analysis in this thesis, since it seems natural to use an adaptive timestep, taking smaller timestep as $\|\widehat{X}_t\|$ approaches 1, to maintain good accuracy, as the drift varies so rapidly near the boundary, and to greatly reduce the possibility of needing to clamp the computed solution to prevent it from crossing a numerical boundary at radius $1-\delta$ for some $\delta \ll 1$ [35]. Numerical results indicate that the order of strong convergence is very close to 1.

2.2.3 Numerical results

The numerical tests include three test cases from [48] plus one new test which provides some motivation for the research reported in this chapter.

Test case 1

The first scalar test case taken from [48] is that

$$\mathrm{d}X_t = -X_t^5\,\mathrm{d}t + X_t\,\mathrm{d}W_t, \qquad X_0 = 1, \tag{2.41}$$

with $T=1$. The three methods tested are the Tamed Euler scheme, with $C=1$, the implicit Euler scheme, and the new Euler scheme with adaptive timestep

$$h^\delta(x) = \delta \, \frac{\max(1, |x|)}{\max(1, |f(x)|}. \tag{2.42}$$

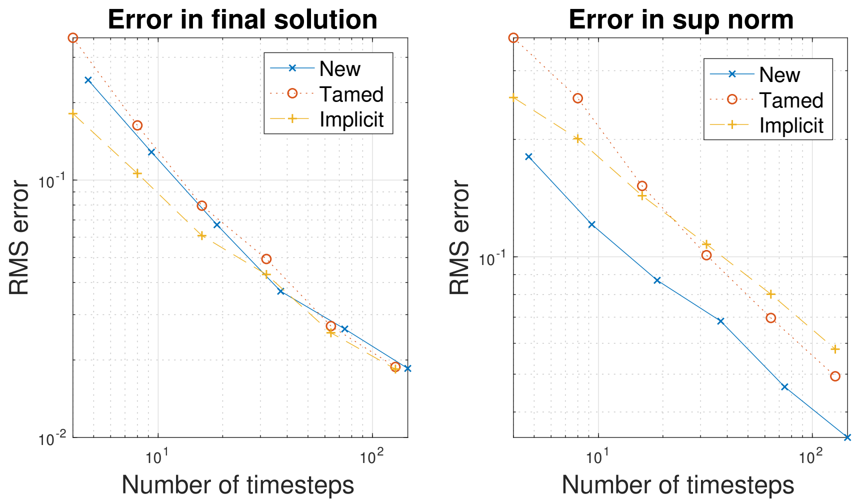


Figure 2.1: Numerical results for test case 1

Figure 2.1 shows the the root-mean-square error plotted against the average number of timesteps. The plot on the left shows the error in the terminal time, while the plot on the right shows the error in the maximum magnitude of the solution. The error in each case is computed by comparing the numerical solution to a second solution with a timestep, or δ, which is 4 times smaller.

When looking at the error in the final solution, all 3 methods have similar accuracy with $\frac{1}{2}$ order strong convergence. However, as reported in [48], the cost of the implicit method per timestep is much higher. The plot of the error in the maximum magnitude shows that the new method is slightly more accurate, presumably because it uses smaller timesteps when the solution is large. The plot was included to show that comparisons between numerical methods depend on the choice of accuracy measure being used.

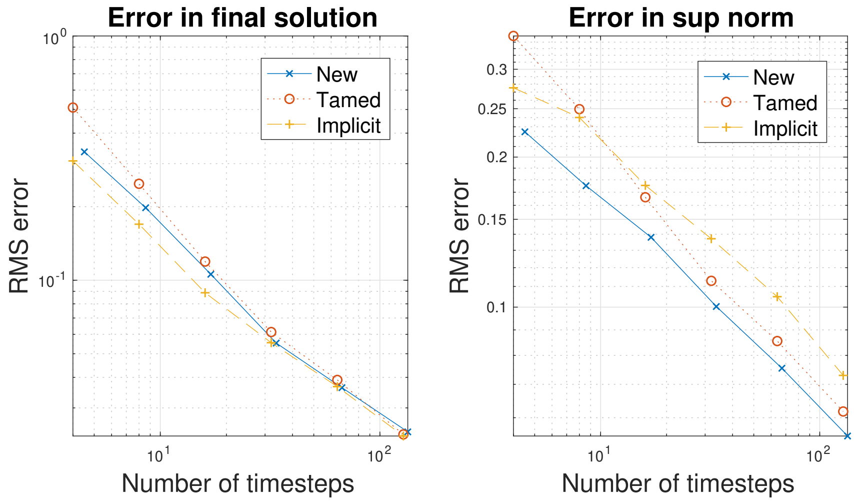


Figure 2.2: Numerical results for test case 2

Test case 2

The second scalar test case taken from [48] is that

$$\mathrm{d}X_t = (X_t - X_t^3)\,\mathrm{d}t + X_t\,\mathrm{d}W_t, \qquad X_0 = 1, \tag{2.43}$$

with $T = 1$. The results in Figure 2.2 are similar to the first test case.

Test case 3

The third test case taken from [48] is 10-dimensional SDE that

$$\mathrm{d}X_t = (X_t - \|X_t\|^2\,X_t)\,\mathrm{d}t + \mathrm{d}W_t, \qquad X_0 = 0, \tag{2.44}$$

with $T = 1$. The results in the left-hand plot in Figure 2.3 show that the error in the final value exhibits order 1 strong convergence using all 3 methods, as expected.

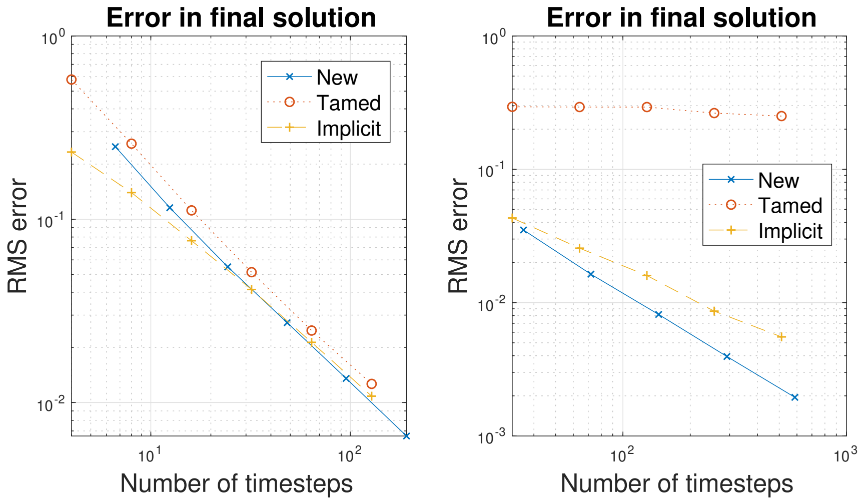


Figure 2.3: Numerical results for test cases 3 (on left) and 4 (on right)

Test case 4

The final test case is for the 3-dimensional FENE SDE discussed previously,

$$\mathrm{d}X_t = -\frac{X_t}{1-\|X_t\|^2}\,\mathrm{d}t + \mathrm{d}W_t, \qquad X_0 = 0, \tag{2.45}$$

with $T=1$. As commented on previously, this SDE is not covered by the theory in this chapter, but it was a motivation for the research because it is natural to use an adaptive timestep of the form

$$h^\delta(x) = \frac{\delta}{4}\,(1-\|x\|^2) \tag{2.46}$$

to reduce the timestep when $\|\widehat{X}_t\|$ approaches the maximum radius.

All three methods are clamped so that they do not exceed a radius of $r_{max} = 1-10^{-10}$; if the new computed value $\widehat{X}_{t_{n+1}}$ exceeds this radius then it is replaced by $(r_{max}/\|\widehat{X}_{t_{n+1}}\|)\widehat{X}_{t_{n+1}}$.

The numerical results in the right-hand plot in Figure 2.3 show that the new scheme is considerably more accurate than either of the others, confirming that an adaptive timestep is desirable in this situation in which the drift varies enormously as $\|\widehat{X}_t\|$ approaches the maximum radius.

2.3 Proofs

This section has the proofs of the three main theorems in this chapter, one on stability, and two on the order of strong convergence.

2.3.1 Theorem 2.1.1

Proof of Theorem 2.1.1. The proof proceeds in four steps. First, we introduce a constant K to modify our discretisation scheme. Second, we derive an upper bound for $\|\widehat{X}_t^K\|^p$. Third, we show that the moments $\mathbb{E}[\sup_{0\leq t\leq T}\|\widehat{X}_t^K\|^p]$ are each bounded by a constant $C_{p,T}$ which depends on p and T but is independent of K. Finally, we reach the desired conclusion by taking the limit $K\to\infty$ and using the Monotone Convergence Theorem.

The proof is given for $p\in[4,\infty)$; the result for $p\in(0,4)$ follows from Hölder's inequality.

Step 1: K-Scheme definition

For any $K>\|X_0\|$, we modify our discretisation scheme to

$$\widehat{X}_{t_{n+1}}^K = P_K\left(\widehat{X}_{t_n}^K + f(\widehat{X}_{t_n}^K)\,h_n + g(\widehat{X}_{t_n}^K)\,\Delta W_n\right), \tag{2.47}$$

where $P_K(Y)\triangleq\min(1,K/\|Y\|)\,Y$ and therefore for all $n\in\{0,1,2,...\}$ it holds that $\|\widehat{X}_{t_n}^K\|\leq K$. The piecewise constant approximation for intermediate times is again $\overline{X}_t^K=\widehat{X}_{\underline{t}}^K$, and the continuous approximation is

$$\widehat{X}_t^K = P_K\left(\widehat{X}_{\underline{t}}^K + f(\widehat{X}_{\underline{t}}^K)\,(t-\underline{t}) + g(\widehat{X}_{\underline{t}}^K)\,(W_t-W_{\underline{t}})\right). \tag{2.48}$$

Since $h(x)$ is continuous and strictly positive, it follows that

$$h_{min}^K \triangleq \inf_{\|x\|\leq K} h(x)\in(0,\infty). \tag{2.49}$$

This strictly positive lower bound for the timesteps implies that T is attainable.

Step 2: pth-moment of K-Scheme solution

$\|P_K(Y)\| \leq \|Y\|$, so let ϕ be a function given by $\phi(x) \triangleq x + h(x)f(x)$. Then (2.47) implies that

$$\begin{aligned}\|\widehat{X}^K_{t_{n+1}}\|^2 \;\leq\; & \|\widehat{X}^K_{t_n}\|^2 + 2\,h_n\left(\langle \widehat{X}^K_{t_n}, f(\widehat{X}^K_{t_n})\rangle + \tfrac{1}{2}h_n\|f(\widehat{X}^K_{t_n})\|^2\right)\\ & + 2\,\langle \phi(\widehat{X}^K_{t_n}), g(\widehat{X}^K_{t_n})\,\Delta W_n\rangle + \|g(\widehat{X}^K_{t_n})\,\Delta W_n\|^2. \qquad (2.50)\end{aligned}$$

Using condition (2.8) for $h(x)$ then ensures that

$$\begin{aligned}\|\widehat{X}^K_{t_{n+1}}\|^2 \;\leq\; & \|\widehat{X}^K_{t_n}\|^2 \;+\; 2\,\alpha\|\,\widehat{X}^K_{t_n}\|^2 h_n \;+\; 2\,\beta\,h_n\\ & + 2\,\langle \phi(\widehat{X}^K_{t_n}), g(\widehat{X}^K_{t_n})\,\Delta W_n\rangle + \|g(\widehat{X}^K_{t_n})\,\Delta W_n\|^2. \qquad (2.51)\end{aligned}$$

Similarly, for the partial timestep from $\underline{t}$ to t, since $(t-\underline{t}) \leq h_{n_t}$, it holds that

$$\langle \widehat{X}^K_{\underline{t}}, f(\widehat{X}^K_{\underline{t}})\rangle + \tfrac{1}{2}\,(t-\underline{t})\,\|f(\widehat{X}^K_{\underline{t}})\|^2 \leq \alpha\,\|\widehat{X}^K_{\underline{t}}\|^2 + \beta, \qquad (2.52)$$

and therefore we obtain that

$$\begin{aligned}\|\widehat{X}^K_t\|^2 \;\leq\; & \|\widehat{X}^K_{\underline{t}}\|^2 \;+\; 2\,\alpha\|\,\widehat{X}^K_{\underline{t}}\|^2(t-\underline{t}) \;+\; 2\,\beta\,(t-\underline{t})\\ & + 2\,\langle \widehat{X}^K_{\underline{t}} + f(\widehat{X}^K_{\underline{t}})\,(t-\underline{t}), g(\widehat{X}^K_{\underline{t}})\,(W_t - W_{\underline{t}})\rangle\\ & + \|g(\widehat{X}^K_{\underline{t}})\,(W_t - W_{\underline{t}})\|^2. \qquad (2.53)\end{aligned}$$

Summing (2.51) over multiple timesteps and then adding (2.53) implies

$$\begin{aligned}\|\widehat{X}^K_t\|^2 \;\leq\; & \|X_0\|^2 \;+\; 2\,\alpha\left(\sum_{k=0}^{n_t-1}\|\widehat{X}^K_{t_k}\|^2 h_k + \|\widehat{X}^K_{\underline{t}}\|^2(t-\underline{t})\right) \;+\; 2\,\beta\,t\\ & + 2\sum_{k=0}^{n_t-1}\langle \phi(\widehat{X}^K_{t_k}), g(\widehat{X}^K_{t_k})\Delta W_k\rangle + \sum_{k=0}^{n_t-1}\|g(\widehat{X}^K_{t_k})\,\Delta W_k\|^2 \qquad (2.54)\\ & + 2\langle \widehat{X}^K_{\underline{t}} + f(\widehat{X}^K_{\underline{t}})\,(t-\underline{t}),\; g(\widehat{X}^K_{\underline{t}})(W_t - W_{\underline{t}})\rangle + \|g(\widehat{X}^K_{\underline{t}})\,(W_t - W_{\underline{t}})\|^2.\end{aligned}$$

Re-writing the first summation as a Riemann integral, and the second as an Itô integral, raising both sides to the power $p/2$ and using Jensen's inequality, we obtain that

$$
\begin{aligned}
\|\widehat{X}_t^K\|^p \;\le\; 7^{p/2-1}\Bigg\{ & \|X_0\|^p \;+\; \left(2\,\alpha\int_0^t \|\overline{X}_s^K\|^2\,\mathrm{d}s\right)^{p/2} \;+\; (2\,\beta\,t)^{p/2} \\
& + \left|2\int_0^{\underline{t}}\langle\phi(\overline{X}_s^K), g(\overline{X}_s^K)\,\mathrm{d}W_s\rangle\right|^{p/2} + \left(\sum_{k=0}^{n_t-1}\|g(\overline{X}_{t_k}^K)\,\Delta W_k\|^2\right)^{p/2} \\
& + \left|2\,\langle\overline{X}_t^K + f(\overline{X}_t^K)\,(t-\underline{t}), g(\overline{X}_t^K)(W_t - W_{\underline{t}})\rangle\right|^{p/2} \\
& + \|g(\overline{X}_t^K)(W_t - W_{\underline{t}})\|^p\Bigg\}\,.
\end{aligned} \tag{2.55}
$$

Step 3: Expected supremum of pth-moment of K-Scheme

For any $t \in [0,T]$ we take the supremum on both sides of inequality (2.55) and then take the expectation to obtain that

$$
\mathbb{E}\left[\sup_{0\le s\le t}\|\widehat{X}_s^K\|^p\right] \;\le\; 7^{p/2-1}\,(I_1 + I_2 + I_3 + I_4 + I_5)\,, \tag{2.56}
$$

where

$$
\begin{aligned}
I_1 &= \|X_0\|^p + \mathbb{E}\left[\left(2\alpha\int_0^t \|\overline{X}_s^K\|^2\,\mathrm{d}s\right)^{p/2}\right] + (2\,\beta\,t)^{p/2}, \\
I_2 &= \mathbb{E}\left[\sup_{0\le s\le \underline{t}}\left|2\int_0^s\langle\phi(\overline{X}_u^K), g(\overline{X}_u^K)\,\mathrm{d}W_u\rangle\right|^{p/2}\right], \\
I_3 &= \mathbb{E}\left[\left(\sum_{k=0}^{n_t-1}\|g(\overline{X}_{t_k}^K)\,\Delta W_k\|^2\right)^{p/2}\right], \\
I_4 &= \mathbb{E}\left[\sup_{0\le s\le t}\left|2\langle\overline{X}_s^K + f(\overline{X}_s^K)\,(s-\underline{s}), g(\overline{X}_s^K)(W_s - W_{\underline{s}})\rangle\right|^{p/2}\right], \\
I_5 &= \mathbb{E}\left[\sup_{0\le s\le t}\|g(\overline{X}_s^K)\,(W_s - W_{\underline{s}})\|^p\right].
\end{aligned} \tag{2.57}
$$

We now consider I_1, I_2, I_3, I_4, I_5 in turn. Using Jensen's inequality, we obtain that

$$
I_1 \le \|x_0\|^p + (2\alpha)^{p/2}T^{p/2-1}\int_0^t \mathbb{E}\left[\sup_{0\le u\le s}\|\widehat{X}_u^K\|^p\right]\,\mathrm{d}s + (2\,\beta\,T)^{p/2}. \tag{2.58}
$$

For I_2, we begin by noting that due to condition (2.8), for $u < \underline{t}$ it holds that

$$\begin{aligned} \|\phi(\overline{X}_u^K)\|^2 &= \|\overline{X}_u^K\|^2 + 2\,h(\overline{X}_u^K)\,\left(\langle \overline{X}_u^K, f(\overline{X}_u^K)\rangle + \tfrac{1}{2}\,h(\overline{X}_u^K)\|f(\overline{X}_u^K)\|^2\right) \\ &\leq \|\overline{X}_u^K\|^2 + 2\,h(\overline{X}_u^K)\,(\alpha\|\overline{X}_u^K\|^2 + \beta) \\ &\leq (1 + 2\,\alpha\,T)\|\overline{X}_u^K\|^2 + 2\,\beta\,T, \end{aligned} \tag{2.59}$$

and hence by Jensen's inequality, we obtain that

$$\|\phi(\overline{X}_u^K)\|^{p/2} \leq 2^{p/4-1}\left((1 + 2\,\alpha\,T)^{p/4}\|\overline{X}_u^K\|^{p/2} + (2\,\beta\,T)^{p/4}\right). \tag{2.60}$$

In addition, the linear growth condition (2.5) ensures that

$$\|g(\overline{X}_u^K)\|^{p/2} \leq 2^{p/4-1}\left(\alpha^{p/4}\|\overline{X}_u^K\|^{p/2} + \beta^{p/4}\right), \tag{2.61}$$

and combining the last two inequalities, there exists a constant $c_{p,T} \in (0,\infty)$ depending on p and T, in addition to α, β, such that

$$\|\phi(\overline{X}_u^K)^T g(\overline{X}_u^K)\|^{p/2} \;\leq\; c_{p,T}\left(\|\overline{X}_u^K\|^p + 1\right). \tag{2.62}$$

Then, by the Burkholder-Davis-Gundy inequality, there is a constant $C_p \in (0,\infty)$ such that

$$\begin{aligned} I_2 &\leq C_p\,2^{p/2}\,\mathbb{E}\left[\left(\int_0^t \|\phi(\overline{X}_u^K)^T g(\overline{X}_u^K)\|^2\,\mathrm{d}u\right)^{p/4}\right] \\ &\leq C_p\,2^{p/2}\,T^{p/4-1}\,\mathbb{E}\left[\int_0^t \|\phi(\overline{X}_u^K)^T g(\overline{X}_u^K)\|^{p/2}\,\mathrm{d}u\right] \\ &\leq c_{p,T}\,C_p\,2^{p/2}\,T^{p/4-1}\left(\int_0^t \mathbb{E}\left[\sup_{0\leq u\leq s}\|\widehat{X}_u^K\|^p\right]\mathrm{d}s \;+\; T\right). \end{aligned} \tag{2.63}$$

For I_3, we start by observing that by standard results there exists a constant $c_p \in (0,\infty)$ which depends solely on p such that for any $t_k \leq s < t_{k+1}$ it holds that

$$\mathbb{E}[\sup_{t_k\leq u\leq s}\|\,W_u - W_{t_k}\,\|^p \mid \mathcal{F}_{t_k}] \;=\; c_p\,(s - \underline{s})^{p/2}. \tag{2.64}$$

One variant of Jensen's inequality, when h_k, u_k are both positive and $p \in [1,\infty)$, is that

$$\left(\sum_k h_k\, u_k\right)^p \le \left(\sum_k h_k\right)^{p-1} \sum_k h_k u_k^p. \tag{2.65}$$

Using this, and (2.64) with $s = t_{k+1}$ so that $s - \underline{s} = h_k$, it holds that

$$\begin{aligned} I_3 &\le T^{p/2-1}\,\mathbb{E}\left[\sum_{k=0}^{n_t-1} h_k\, \|g(\overline{X}^K_{t_k})\|^p\ \frac{\|\Delta W_k\|^p}{h_k^{p/2}}\right] \\ &\le T^{p/2-1}\, c_p\ \mathbb{E}\left[\int_0^t \|g(\overline{X}^K_s)\|^p\,\mathrm{d}s\right]. \end{aligned} \tag{2.66}$$

Using condition (2.5), and Jensen's inequality, we then obtain that

$$I_3 \le (2\,T)^{p/2-1}\, c_p \left(\alpha^{p/2}\int_0^t \mathbb{E}\left[\sup_{0\le u\le s}\|\widehat{X}^K_u\|^p\right]\mathrm{d}s\ +\ \beta^{p/2}\,T\right). \tag{2.67}$$

For I_4, using (2.52) and following the same argument as for I_2, there exists a constant $c_{p,T}\in(0,\infty)$ depending on both p and T such that

$$\|\overline{X}^K_s + f(\overline{X}^K_s)(s-\underline{s})\|^{p/2}\|g(\overline{X}^K_s)\|^{p/2} \le c_{p,T}\ \left(\|\overline{X}^K_s\|^p + 1\right). \tag{2.68}$$

Therefore, again using (2.64), we obtain that

$$\begin{aligned} I_4 &\le 2^{p/2}\,\mathbb{E}\left[\sup_{0\le s\le t}\left|\langle \overline{X}^K_s + f(\overline{X}^K_s)(s-\underline{s}),\, g(\overline{X}^K_s)\,(W_s - W_{\underline{s}})\rangle\right|^{p/2}\right] \\ &\le c_{p,T}\,2^{p/2}\,\mathbb{E}\left[\sum_{k=0}^{n_t-1}\left(\|\overline{X}^K_{t_k}\|^p + 1\right)\sup_{t_k\le s<t_{k+1}}\|(W_s - W_{\underline{s}})\|^{p/2}\right. \\ &\qquad\qquad \left. +\ \left(\|\overline{X}^K_t\|^p + 1\right)\sup_{\underline{t}\le s\le t}\|(W_s - W_{\underline{s}})\|^{p/2}\right] \\ &\le c_{p/2}\,c_{p,T}\,2^{p/2}\,T^{p/4-1}\ \mathbb{E}\left[\sum_{k=0}^{n_t-1}\left(\|\overline{X}^K_{t_k}\|^p + 1\right)h_k + \left(\|\overline{X}^K_t\|^p + 1\right)(t - \underline{t})\right] \\ &\le c_{p/2}\,c_{p,T}\,2^{p/2}\,T^{p/4-1}\left(\int_0^t \mathbb{E}\left[\sup_{0\le u\le s}\|\widehat{X}^K_u\|^p\right]\mathrm{d}s\ +\ T\right). \end{aligned} \tag{2.69}$$

Similarly, using the same definition for c_p, we obtain that

$$I_5 \ \le\ c_p\,(2\,T)^{p/2-1}\left(\alpha^{p/2}\int_0^t \mathbb{E}\left[\sup_{0\le u\le s}\|\widehat{X}^K_u\|^p\right]\mathrm{d}s\ +\ \beta^{p/2}\,T\right). \tag{2.70}$$

Collecting together the bounds for I_1, I_2, I_3, I_4, I_5, we conclude that there exist constants $C^1_{p,T}, C^2_{p,T} \in (0,\infty)$ such that

$$\mathbb{E}\left[\sup_{0\le s\le t} \|\widehat{X}^K_s\|^p\right] \le C^1_{p,T} + C^2_{p,T}\int_0^t \mathbb{E}\left[\sup_{0\le u\le s} \|\widehat{X}^K_u\|^p\right] \mathrm{d}s, \tag{2.71}$$

and Grönwall's inequality gives the result that there exists a constant $C_{p,T} \in (0,\infty)$ such that

$$\mathbb{E}\left[\sup_{0\le t\le T} \|\widehat{X}^K_t\|^p\right] \le C^1_{p,T}\, \exp(C^2_{p,T}\, T) = C_{p,T}. \tag{2.72}$$

Step 4: Expected supremum of pth-moment of $\widehat{X}_t$

For any $\omega \in \Omega$, $\widehat{X}_t = \widehat{X}^K_t$ for all $0 \le t \le T$ if, and only if, $\sup_{0\le t\le T} \|\widehat{X}_t\| \le K$. Therefore, Markov's inequality ensures that

$$\begin{aligned}\mathbb{P}(\sup_{0\le t\le T} \|\widehat{X}_t\| < K) &= \mathbb{P}(\sup_{0\le t\le T} \|\widehat{X}^K_t\| < K) \\ &\ge 1 - \mathbb{E}[\sup_{0\le t\le T} \|\widehat{X}^K_t\|^4]/K^4 \to 1 \end{aligned}\tag{2.73}$$

as $K \to \infty$. Hence, almost surely, $\sup\limits_{0\le t\le T} \|\widehat{X}_t\| < \infty$ and T is attainable. Also, we obtain that

$$\lim_{K\to\infty} \sup_{0\le t\le T} \|\widehat{X}^K_t(\omega)\| = \sup_{0\le t\le T} \|\widehat{X}_t(\omega)\| \tag{2.74}$$

and for $0 < K_1 \le K_2$, it holds that

$$\sup_{0\le t\le T} \|\widehat{X}^{K_1}_t(\omega)\| \le \sup_{0\le t\le T} \|\widehat{X}^{K_2}_t(\omega)\| \le \sup_{0\le t\le T} \|\widehat{X}_t(\omega)\|. \tag{2.75}$$

Therefore, the Monotone Convergence Theorem ensures that

$$\mathbb{E}\left[\sup_{0\le t\le T} \|\widehat{X}_t\|^p\right] = \lim_{K\to\infty} \mathbb{E}\left[\sup_{0\le t\le T} \|\widehat{X}^K_t\|^p\right] \le C_{p,T}. \tag{2.76}$$

This completes the proof of Theorem 2.1.1. □

2.3.2 Theorem 2.1.3

Proof of Theorem 2.1.3. The approach which is followed is to bound the approximation error $e_t \triangleq \widehat{X}_t - X_t$ by terms which depend on either $\widehat{X}_s - \overline{X}_s$ or e_s, and then use local analysis within each timestep to bound the former, and Grönwall's inequality to handle the latter.

The proof is again given for $p \in [4, \infty)$; the result for $p \in (0, 4)$ follows from Hölder's inequality.

We start by combining the original SDE with (2.3) to obtain that

$$\mathrm{d}e_t \;=\; \big(f(\overline{X}_t) - f(X_t)\big)\,\mathrm{d}t + \big(g(\overline{X}_t) - g(X_t)\big)\,\mathrm{d}W_t, \tag{2.77}$$

and then the Itô's formula, together with $e_0 = 0$, implies that

$$\begin{aligned}\|e_t\|^2 \;&\leq\; 2\int_0^t \langle e_s, f(\widehat{X}_s) - f(X_s)\rangle\,\mathrm{d}s - 2\int_0^t \langle e_s, f(\widehat{X}_s) - f(\overline{X}_s)\rangle\,\mathrm{d}s \\ &\quad + \int_0^t \|g(\overline{X}_s) - g(X_s)\|^2\,\mathrm{d}s + 2\int_0^t \langle e_s, (g(\overline{X}_s) - g(X_s))\,\mathrm{d}W_s\rangle. \end{aligned} \tag{2.78}$$

Using the conditions in Assumption 2.1.5, (2.17) implies that

$$\langle e_s, f(\widehat{X}_s) - f(X_s)\rangle \leq \tfrac{1}{2}\alpha\,\|e_s\|^2, \tag{2.79}$$

(2.19) ensures that

$$\begin{aligned}\Big|\langle e_s, f(\widehat{X}_s) - f(\overline{X}_s)\rangle\Big| \;&\leq\; \|e_s\|\, L(\widehat{X}_s, \overline{X}_s)\, \|\widehat{X}_s - \overline{X}_s\| \\ &\leq\; \tfrac{1}{2}\|e_s\|^2 + \tfrac{1}{2}L(\widehat{X}_s, \overline{X}_s)^2 \|\widehat{X}_s - \overline{X}_s\|^2, \end{aligned} \tag{2.80}$$

where $L(x, y) \triangleq \gamma(\|x\|^q + \|y\|^q) + \mu$, and (2.18) assures that

$$\|g(\overline{X}_s) - g(X_s)\|^2 \;\leq\; \tfrac{1}{2}\,\alpha\,\|\overline{X}_s - X_s\|^2 \;\leq\; \alpha\,\|e_s\|^2 + \alpha\,\|\widehat{X}_s - \overline{X}_s\|^2. \tag{2.81}$$

Hence, we obtain that

$$\begin{aligned}\|e_t\|^2 \;&\leq\; (2\alpha + 1)\int_0^t \|e_s\|^2\,\mathrm{d}s + \int_0^t \Big(L(\widehat{X}_s, \overline{X}_s)^2 + \alpha\Big)\,\|\widehat{X}_s - \overline{X}_s\|^2\,\mathrm{d}s \\ &\quad + 2\int_0^t \langle e_s, (g(\overline{X}_s) - g(X_s))\,\mathrm{d}W_s\rangle, \end{aligned} \tag{2.82}$$

and then the Jensen's inequality ensures that

$$\begin{aligned}\|e_t\|^p &\leq (3T)^{p/2-1}(2\alpha+1)^{p/2}\int_0^t \|e_s\|^p \,\mathrm{d}s \\ &+ (3T)^{p/2-1}\int_0^t \left(L(\widehat{X}_s,\overline{X}_s)^2+\alpha\right)^{p/2}\|\widehat{X}_s-\overline{X}_s\|^p \,\mathrm{d}s \\ &+ 3^{p/2-1}2^{p/2}\left|\int_0^t \langle e_s,(g(\overline{X}_s)-g(X_s))\,\mathrm{d}W_s\rangle\right|^{p/2}. \end{aligned} \tag{2.83}$$

Taking the supremum of each side, and then the expectation ensures that

$$\begin{aligned}\mathbb{E}\left[\sup_{0\leq s\leq t}\|e_s\|^p\right] &\leq (3T)^{p/2-1}(2\alpha+1)^{p/2}\int_0^t \mathbb{E}\left[\sup_{0\leq u\leq s}\|e_u\|^p\right]\mathrm{d}s \\ &+ (3T)^{p/2-1}\int_0^t \mathbb{E}\left[\left(L(\widehat{X}_s,\overline{X}_s)^2+\alpha\right)^{p/2}\|\widehat{X}_s-\overline{X}_s\|^p\right]\mathrm{d}s \quad (2.84) \\ &+ 3^{p/2-1}2^{p/2}\mathbb{E}\left[\sup_{0\leq s\leq t}\left|\int_0^s \langle e_u,(g(\overline{X}_u)-g(X_u))\,\mathrm{d}W_u\rangle\right|^{p/2}\right]. \end{aligned}$$

By Hölder's inequality, it holds that

$$\begin{aligned}&\mathbb{E}\left[\left(L(\widehat{X}_s,\overline{X}_s)^2+\alpha\right)^{p/2}\|\widehat{X}_s-\overline{X}_s\|^p\right] \\ &\leq \left(\mathbb{E}\left[\left(L(\widehat{X}_s,\overline{X}_s)^2+\alpha\right)^{p}\right]\mathbb{E}\left[\|\widehat{X}_s-\overline{X}_s\|^{2p}\right]\right)^{1/2}, \end{aligned} \tag{2.85}$$

and $\mathbb{E}\left[\left(L(\widehat{X}_s,\overline{X}_s)^2+\alpha\right)^{p}\right]$ is uniformly bounded on $[0,T]$ due to the stability property in Theorem 2.1.1.

In addition, by the Burkholder-Davis-Gundy inequality (which gives the constant C_p which depends only on p) followed by Jensen's inequality plus the globally

Lipschitz condition for g, we obtain that

$$
\begin{aligned}
\mathbb{E}\left[\sup_{0\le s\le t}\left|\int_0^s \langle e_u, (g(\overline{X}_u)-g(X_u))\,\mathrm{d}W_u\rangle\right|^{p/2}\right] & \\
\le\ C_p\,\mathbb{E}\left[\left(\int_0^t \|e_s\|^2\|g(\overline{X}_s)-g(X_s)\|^2\,\mathrm{d}s\right)^{p/4}\right]. & \\
\le\ C_p\,T^{p/4-1}\,(\tfrac{1}{2}\alpha)^{p/4}\mathbb{E}\left[\int_0^t \|e_s\|^{p/2}\|\overline{X}_s-X_s\|^{p/2}\,\mathrm{d}s\right] & \quad (2.86)\\
\le\ C_p\,T^{p/4-1}\,(\tfrac{1}{2}\alpha)^{p/4}\mathbb{E}\left[\int_0^t \tfrac{1}{2}\,\|e_s\|^p+\tfrac{1}{2}\,\|\overline{X}_s-X_s\|^p\,\mathrm{d}s\right] & \\
\le\ C_p\,T^{p/4-1}\,(\tfrac{1}{2}\alpha)^{p/4}\mathbb{E}\left[\int_0^t (\tfrac{1}{2}+2^{p-2})\|e_s\|^p+2^{p-2}\|\widehat{X}_s-\overline{X}_s\|^p\,\mathrm{d}s\right]. &
\end{aligned}
$$

Hence, using $\mathbb{E}[\|\widehat{X}_s-\overline{X}_s\|^p]\le(\mathbb{E}[\|\widehat{X}_s-\overline{X}_s\|^{2p}])^{1/2}$, there are constants $C^1_{p,T}, C^2_{p,T}\in(0,\infty)$ such that

$$
\begin{aligned}
\mathbb{E}\left[\sup_{0\le s\le t}\|e_s\|^p\right] \ \le\ & C^1_{p,T}\int_0^t \mathbb{E}\left[\sup_{0\le u\le s}\|e_u\|^p\right]\mathrm{d}s \\
& +C^2_{p,T}\int_0^t\left(\mathbb{E}\left[\|\widehat{X}_s-\overline{X}_s\|^{2p}\right]\right)^{1/2}\mathrm{d}s. \qquad (2.87)
\end{aligned}
$$

For any $s\in[0,T]$, $\widehat{X}_s-\overline{X}_s = f(\widehat{X}_{\underline{s}})(s-\underline{s})+g(\widehat{X}_{\underline{s}})(W_s-W_{\underline{s}})$, and hence, by a combination of Jensen's and Hölder's inequalities, we obtain that

$$
\begin{aligned}
\mathbb{E}\left[\|\widehat{X}_s-\overline{X}_s\|^{2p}\right] \ \le\ & 2^{2p-1}\left(\mathbb{E}\left[\|f(\widehat{X}_{\underline{s}})\|^{4p}\right]\mathbb{E}\left[(s-\underline{s})^{4p}\right]\right)^{1/2} \\
+\ & 2^{2p-1}\left(\mathbb{E}\left[\|g(\widehat{X}_{\underline{s}})\|^{4p}\right]\mathbb{E}\left[\|W_s-W_{\underline{s}}\|^{4p}\right]\right)^{1/2}. \qquad (2.88)
\end{aligned}
$$

$\mathbb{E}[\|f(\widehat{X}_{\underline{s}})\|^{4p}]$ and $\mathbb{E}[\|g(\widehat{X}_{\underline{s}})\|^{4p}]$ are both uniformly bounded on $[0,T]$ due to stability and the polynomial bounds on the growth of f and g. Furthermore, it holds that $\mathbb{E}[(s-\underline{s})^{4p}]\le(\delta T)^{4p}\le\delta^{2p}T^{4p}$, and by standard results there is a constant c_p such that $\mathbb{E}[\|W_s-W_{\underline{s}}\|^{4p}]=\mathbb{E}[\,\mathbb{E}[\|W_s-W_{\underline{s}}\|^{4p}\,|\,\mathcal{F}_{\underline{s}}]\,]\le c_p(\delta T)^{2p}$. Hence, there exists a constant $C^3_{p,T}\in(0,\infty)$ such that $\mathbb{E}[\,\|\widehat{X}_s-\overline{X}_s\|^{2p}]\le C^3_{p,T}\,\delta^p$, and therefore equation (2.87) implies that

$$
\mathbb{E}\left[\sup_{0\le s\le t}\|e_s\|^p\right]\le C^1_{p,T}\int_0^t\mathbb{E}\left[\sup_{0\le u\le s}\|e_u\|^p\right]\mathrm{d}s\ +\ C^2_{p,T}\sqrt{C^3_{p,T}}\;T\,\delta^{p/2}, \qquad (2.89)
$$

and Grönwall's inequality then completes the proof of Theorem 2.1.3. $\square$

2.3.3 Theorem 2.1.4

Proof of Theorem 2.1.4. The proof is again given for $p \in [4, \infty)$; the result for $p \in (0, 4)$ follows from Hölder's inequality.

The error $e_t \triangleq \widehat{X}_t - X_t$ satisfies the SDE $\mathrm{d}e_t = \big(f(\overline{X}_t) - f(X_t)\big)\,\mathrm{d}t$ and hence we obtain that

$$\begin{aligned}\|e_t\|^2 &= 2\int_0^t \langle e_s, f(\widehat{X}_s) - f(X_s)\rangle\,\mathrm{d}s \;-2\int_0^t \langle e_s, f(\widehat{X}_s) - f(\overline{X}_s)\rangle\,\mathrm{d}s \\ &\leq \alpha\int_0^t \|e_s\|^2\,\mathrm{d}s - 2\int_0^t \langle e_s, f(\widehat{X}_s) - f(\overline{X}_s)\rangle\,\mathrm{d}s, \end{aligned} \tag{2.90}$$

due to the one-sided Lipschitz condition (2.17), so therefore it holds that

$$\begin{aligned}\mathbb{E}\left[\sup_{0\leq s\leq t}\|e_s\|^p\right] \;\leq\;& \alpha^{p/2}(2T)^{p/2-1}\int_0^t \mathbb{E}\left[\sup_{0\leq u\leq s}\|e_u\|^p\right]\mathrm{d}s \\ &+2^{p-1}\mathbb{E}\left[\sup_{0\leq s\leq t}\left|\int_0^s \langle e_u, f(\widehat{X}_u) - f(\overline{X}_u)\rangle\,\mathrm{d}u\right|^{p/2}\right]. \end{aligned} \tag{2.91}$$

Within a single timestep, it holds that $\widehat{X}_s - \overline{X}_s = f(\overline{X}_s)(s - \underline{s}) + (W_s - W_{\underline{s}})$, and therefore Lemma C.1 implies that

$$\begin{aligned}\langle e_s, f(\widehat{X}_s) - f(\overline{X}_s)\rangle &= \langle e_s, \nabla f(\overline{X}_s)(\widehat{X}_s - \overline{X}_s)\rangle \;+\; R_s \\ &= \langle e_s, (s-\underline{s})\nabla f(\overline{X}_s) f(\overline{X}_s)\rangle + \langle (e_s - e_{\underline{s}}), \nabla f(\overline{X}_s)(W_s - W_{\underline{s}})\rangle \\ &\quad +\langle e_{\underline{s}}, \nabla f(\overline{X}_s)(W_s - W_{\underline{s}})\rangle + \; R_s, \end{aligned} \tag{2.92}$$

where $|R_s| \leq \Big(\gamma\,(\|\widehat{X}_s\|^q + \|\overline{X}_s\|^q) + \mu\Big)\;\|e_s\|\;\|\widehat{X}_s - \overline{X}_s\|^2$. Hence, it holds that

$$\mathbb{E}\left[\sup_{0\leq s\leq t}\left|\int_0^s \langle e_u, f(\widehat{X}_u) - f(\overline{X}_u)\rangle\,\mathrm{d}u\right|^{p/2}\right] \leq 4^{p/2-1}(I_1 + I_2 + I_3 + I_4), \tag{2.93}$$

where

$$\begin{aligned}
I_1 &= \mathbb{E}\left[\sup_{0\le s\le t}\left|\int_0^s \langle e_u,(u-\underline{u})\nabla f(\overline{X}_u)f(\overline{X}_u)\rangle\,\mathrm{d}u\right|^{p/2}\right],\\
I_2 &= \mathbb{E}\left[\sup_{0\le s\le t}\left|\int_0^s R_u\,\mathrm{d}u\right|^{p/2}\right],\\
I_3 &= \mathbb{E}\left[\sup_{0\le s\le t}\left|\int_0^s \langle (e_u-e_{\underline{u}}),\nabla f(\overline{X}_u)(W_u-W_{\underline{u}})\rangle\,\mathrm{d}u\right|^{p/2}\right],\\
I_4 &= \mathbb{E}\left[\sup_{0\le s\le t}\left|\int_0^s \langle e_{\underline{u}},\nabla f(\overline{X}_u)(W_u-W_{\underline{u}})\rangle\,\mathrm{d}u\right|^{p/2}\right]. \qquad (2.94)
\end{aligned}$$

We now bound I_1, I_2, I_3, I_4 in turn. Noting that $s-\underline{s}\le\delta T$, we obtain that

$$\begin{aligned}
I_1 &\le T^{p/2-1}\int_0^t \mathbb{E}\left[\|e_s\|^{p/2}(\delta T)^{p/2}\|f(\overline{X}_s)\|^{p/2}\|\nabla f(\overline{X}_s)\|^{p/2}\right]\,\mathrm{d}s\\
&\le \tfrac{1}{2}T^{p/2-1}\int_0^t \mathbb{E}\left[\sup_{0\le u\le s}\|e_u\|^p\right]\,\mathrm{d}s\\
&\quad+\tfrac{1}{2}T^{p/2-1}(\delta T)^p\int_0^t \mathbb{E}\left[\|f(\overline{X}_s)\|^p\|\nabla f(\overline{X}_s)\|^p\right]\,\mathrm{d}s. \qquad (2.95)
\end{aligned}$$

The last integral is finite because of stability and the polynomial bounds on the growth of both f and ∇f, and hence there is a constant $C^1_{p,T}\in(0,\infty)$ such that

$$I_1\le \tfrac{1}{2}T^{p/2-1}\int_0^t \mathbb{E}\left[\sup_{0\le u\le s}\|e_u\|^p\right]\,\mathrm{d}s + C^1_{p,T}\,\delta^p. \qquad (2.96)$$

Similarly, Hölder's inequality ensures that

$$\begin{aligned}
I_2 &\le T^{p/2-1}\int_0^t \mathbb{E}\left[\|e_s\|^{p/2}\left(\gamma\,(\|\widehat{X}_s\|^q+\|\overline{X}_s\|^q)+\mu\right)^{p/2}\|\widehat{X}_s-\overline{X}_s\|^p\right]\mathrm{d}s\\
&\le \tfrac{1}{2}T^{p/2-1}\int_0^t \mathbb{E}\left[\sup_{0\le u\le s}\|e_u\|^p\right]\,\mathrm{d}s \qquad (2.97)\\
&\quad+\tfrac{1}{2}T^{p/2-1}\int_0^t\left(\mathbb{E}\left[\left(\gamma\,(\|\widehat{X}_s\|^q+\|\overline{X}_s\|^q)+\mu\right)^{2p}\right]\mathbb{E}\left[\|\widehat{X}_s-\overline{X}_s\|^{4p}\right]\right)^{1/2}\mathrm{d}s,
\end{aligned}$$

and hence, using stability and bounds on $\mathbb{E}\left[\|\widehat{X}_s-\overline{X}_s\|^{4p}\right]$ from the proof of Theorem 2.1.3, there is a constant $C^2_{p,T}\in(0,\infty)$ such that

$$I_2\le \tfrac{1}{2}T^{p/2-1}\int_0^t \mathbb{E}\left[\sup_{0\le u\le s}\|e_u\|^p\right]\,\mathrm{d}s + C^2_{p,T}\,\delta^p. \qquad (2.98)$$

For the next term, I_3, we start by bounding $\|e_s - e_{\underline{s}}\|$. Since it holds that

$$e_s - e_{\underline{s}} = \int_{\underline{s}}^{s} \left(f(\overline{X}_u) - f(X_u) \right) \, \mathrm{d}u, \tag{2.99}$$

by Jensen's inequality and Assumption 2.1.7 it follows that

$$\begin{aligned} \|e_s - e_{\underline{s}}\|^p &\leq (\delta\, T)^{p-1} \int_{\underline{s}}^{s} \|f(\overline{X}_u) - f(X_u)\|^p \, \mathrm{d}u \\ &\leq (2\,\delta\, T)^{p-1} \int_{\underline{s}}^{s} L^p(\overline{X}_u, X_u) \left(\|e_u\|^p + \|\widehat{X}_u - \overline{X}_u\|^p \right) \mathrm{d}u, \end{aligned} \tag{2.100}$$

where $L(\overline{X}_u, X_u) \equiv \gamma(\|\overline{X}_u\|^k + \|X_u\|^k) + \mu$. We again have an $O(\delta^{p/2})$ bound for $\mathbb{E}[\|\widehat{X}_s - \overline{X}_s\|^p]$, while Theorem 2.1.3 proves that there is a constant $c_{p,T} \in (0, \infty)$ such that

$$\mathbb{E}[\|e_s\|^p] \leq c_{p,T}\, \delta^{p/2}. \tag{2.101}$$

Combining these, and using Hölder's inequality and the bound for $\mathbb{E}[L^p(\overline{X}_u, X_u)]$ for all $p \in [2, \infty)$, due to the usual stability results, we find that there is a different constant $c_{p,T}$ such that

$$\mathbb{E}[\|e_s - e_{\underline{s}}\|^p] \leq c_{p,T}\, \delta^{3p/2}. \tag{2.102}$$

Now, we obtain that

$$I_3 \leq T^{p/2-1} \int_0^t \mathbb{E}\left[\|e_s - e_{\underline{s}}\|^{p/2} \|\nabla f(\overline{X}_s)\|^{p/2} \|W_s - W_{\underline{s}}\|^{p/2} \right] \mathrm{d}s, \tag{2.103}$$

so using Hölder's inequality and the usual stability bounds, we conclude that there is a constant $C^3_{p,T} \in (0, \infty)$ such that

$$I_3 \leq C^3_{p,T}\, \delta^p. \tag{2.104}$$

Lastly, we consider I_4. For the timestep $[t_n, t_{n+1}]$, we obtain that

$$\mathrm{d}\left((t - t_{n+1})(W_t - W_{t_n}) \right) = (W_t - W_{t_n})\, \mathrm{d}t + (t - t_{n+1})\, \mathrm{d}W_t \tag{2.105}$$

and therefore, integrating by parts within each timestep implies that

$$\begin{aligned}&\int_0^s \langle e_{\underline{u}}, \nabla f(\overline{X}_u)(W_u - W_{\underline{u}})\rangle \,\mathrm{d}u \\ &\quad = \int_0^s (\overline{u}-u)\langle e_{\underline{u}}, \nabla f(\overline{X}_u)\mathrm{d}W_u\rangle - (\overline{s}-s)\langle e_{\underline{s}}, \nabla f(\overline{X}_s)(W_s - W_{\underline{s}})\rangle, \quad (2.106)\end{aligned}$$

where $\overline{u} = \min\{t_n : t_n > u\} = t_{n_u+1}$. Hence, it holds that $I_4 \leq 2^{p/2-1}(I_{41} + I_{42})$ where

$$\begin{aligned} I_{41} &= \mathbb{E}\left[\sup_{0\leq s\leq t}\left|\int_0^s (\overline{u}-u)\,\langle e_{\underline{u}}, \nabla f(\overline{X}_u)\,\mathrm{d}W_u\rangle\right|^{p/2}\right], \\ I_{42} &= \mathbb{E}\left[\sup_{0\leq s\leq t}\left|(\overline{s}-s)\,\langle e_{\underline{s}}, \nabla f(\overline{X}_s)(W_s - W_{\underline{s}})\rangle\right|^{p/2}\right]. \qquad (2.107)\end{aligned}$$

The Burkholder-Davis-Gundy inequality ensures that

$$\begin{aligned} I_{41} &\leq C_p\,\mathbb{E}\left[\left(\int_0^t (\overline{s}-s)^2 \|e_{\underline{s}}\|^2 \|\nabla f(\overline{X}_s)\|^2\,\mathrm{d}s\right)^{p/4}\right] \\ &\leq C_p\,T^{3p/4-1}\,\mathbb{E}\left[\int_0^t \|e_{\underline{s}}\|^{p/2}\,\delta^{p/2}\,\|\nabla f(\overline{X}_s)\|^{p/2}\,\mathrm{d}s\right] \\ &\leq \tfrac{1}{2}C_p\,T^{3p/4-1}\,\mathbb{E}\left[\int_0^t \left(\sup_{0\leq u\leq s}\|e_u\|^p + \delta^p\,\|\nabla f(\overline{X}_s)\|^p\right)\mathrm{d}s\right], \qquad (2.108)\end{aligned}$$

with $\mathbb{E}[\|\nabla f(\overline{X}_s)\|^p]$ uniformly bounded on $[0,T]$ so that there is a constant $C^{41}_{p,T} \in (0,\infty)$ such that

$$I_{41} \leq \tfrac{1}{2}\,C_p\,T^{3p/4-1}\int_0^t \mathbb{E}\left[\sup_{0\leq u\leq s}\|e_u\|^p\right]\mathrm{d}s + C^{41}_{p,T}\,\delta^p. \qquad (2.109)$$

Turning to I_{42}, Young's inequality and Hölder's inequality establish that

$$\begin{aligned} I_{42} &\leq \frac{1}{2\xi}\mathbb{E}\left[\sup_{0\leq s\leq t}\|e_s\|^p\right] \\ &\quad + \frac{\xi}{2}(2\delta T)^p\left(\mathbb{E}\left[\sup_{0\leq s\leq t}\|\nabla f\|^{2p}\right]\mathbb{E}\left[\sup_{0\leq s\leq t}\|W_s\|^{2p}\right]\right)^{1/2}, \qquad (2.110)\end{aligned}$$

for any $\xi \in (0,\infty)$, and hence there is a constant $C^{42}_{p,T} \in (0,\infty)$ such that

$$I_{42} \leq \frac{1}{2\xi}\,\mathbb{E}\left[\sup_{0\leq s\leq t}\|e_s\|^p\right] + \xi\,C^{42}_{p,T}\,\delta^p. \qquad (2.111)$$

Returning to (2.91), and inserting the bounds for I_1, I_2, I_3, I_4, I_{41}, and I_{42}, with $\xi = 2^{5p/2-4}$, implies

$$\mathbb{E}\left[\sup_{0\le s\le t} \|e_s\|^p\right] \le \tfrac{1}{2}\mathbb{E}\left[\sup_{0\le s\le t} \|e_s\|^p\right] + C^5_{p,T}\int_0^t \mathbb{E}\left[\sup_{0\le u\le s} \|e_u\|^p\right] \mathrm{d}s + C^6_{p,T}\,\delta^p, \quad (2.112)$$

for certain constants $C^5_{p,T}, C^6_{p,T} \in (0,\infty)$. Rearranging and using Grönwall's inequality we obtain the final conclusion that there exists a constant $C_{p,T} \in (0,\infty)$ such that

$$\mathbb{E}\left[\sup_{0\le t\le T} \|e_t\|^p\right] \le C_{p,T}\,\delta^p. \quad (2.113)$$

This completes the proof of Theorem 2.1.4. □

2.4 Conclusion

In this chapter, we introduced an adaptive timestepping method for SDEs with non-globally Lipschitz drift. We established the stability of the adaptive scheme for a finite time interval as shown in Theorem 2.1.1. For the strong convergence of the adaptive scheme, we introduced a scalar factor δ to control the average size of the time step and established the strong convergence with respect to δ, as shown in Theorem 2.1.2. When looking at the accuracy versus the expected cost of each path, if the drift f also satisfies a one-sided Lipschitz condition, then the order of strong convergence is $\frac{1}{2}$, as shown in Theorem 2.1.3. For the important class of Langevin equations with unit volatility, we showed in Theorem 2.1.4 that the order of strong convergence is 1. The analysis is supported by numerical experiments for a variety of SDEs in section 2.2.

Chapter 3

Infinite time interval

This chapter is an extended version of the Arxiv paper [23]. In this chapter, we extend our adaptive scheme to an infinite time interval. The rest of the chapter is organised as follows. Section 1 states the main theorems and proves some minor lemmas. Section 2 introduces the MLMC schemes, and the relevant numerical experiments are provided in section 3. We extend our MLMC schemes to a larger class of ergodic SDEs in section 4. The proofs of the three main theorems are deferred to section 5. Finally, section 6 concludes this chapter.

3.1 Theoretical results for infinite time interval

The adaptive algorithm considered in this chapter is the same as the one introduced in section 2.1.1. In the following subsections, we state some preliminary lemmas, the key results on stability and strong convergence, and related results on an number of timesteps, introducing various assumptions as required for each over infinite time interval.

3.1.1 Stability

Assumption 3.1.1 (Locally Lipschitz and linear growth)**.** *Assume that f and g are both locally Lipschitz. Furthermore, assume that there exist constants $\alpha, \beta \in (0, \infty)$ such that for all $x \in \mathbb{R}^m$, f satisfies the dissipativity condition that*

$$\langle x, f(x) \rangle \leq -\alpha \|x\|^2 + \beta, \tag{3.1}$$

and g is globally bounded satisfying that

$$\|g(x)\|^2 \leq \beta, \tag{3.2}$$

and non-degenerate with full rank m for all $x \in \mathbb{R}^m$.

Theorem 4.4 in [73] and Theorem 6.1 in [75] show that this Assumption ensures the existence and uniqueness of the invariant measure. We can also prove the following uniform moment bound for the SDE solution.

Lemma 3.1.1 (SDE stability)**.** *If the SDE satisfies Assumption 3.1.1 with $X_0 = x_0$, then for all $p \in (0, \infty)$, there is a constant $C_p \in (0, \infty)$ which only depends on x_0 and p such that for all $t \in [0, \infty)$ it holds that*

$$\mathbb{E}\left[\|X_t\|^p\right] \leq C_p. \tag{3.3}$$

Proof of Lemma 3.1.1. The result follows Theorem 1.8 in [56]. This completes the proof of Lemma 3.1.1. □

We now specify the critical assumption about the adaptive timestep.

Assumption 3.1.2 (Adaptive timestep)**.** *The adaptive timestep function $h : \mathbb{R}^m \to (0, h_{\max}]$ is continuous and bounded, with $h_{\max} \in (0, \infty)$, and there exist constants $\alpha, \beta \in (0, \infty)$ such that for all $x \in \mathbb{R}^m$, h satisfies the inequality that*

$$\langle x, f(x) \rangle + \tfrac{1}{2}\, h(x) \, \|f(x)\|^2 \leq -\alpha \|x\|^2 + \beta. \tag{3.4}$$

Compared with the Assumption 2.1.2 in the finite time analysis, this assumption additionally bound h to achieve the uniform bound.

Theorem 3.1.1 (Infinite time stability)**.** *If the SDE satisfies Assumption 3.1.1, and the timestep function h satisfies Assumption 3.1.2, then for all $p \in (0, \infty)$ there exists a constant $C_p \in (0, \infty)$ which depends solely on p, x_0, $h_{\max}$ and the constants α, β in Assumption 3.1.2 such that for all $t \in [0, \infty)$ it holds that*

$$\mathbb{E}\left[\|\widehat{X}_t\|^p\right] < C_p, \ \mathbb{E}\left[\|\overline{X}_t\|^p\right] < C_p. \tag{3.5}$$

Proof of Theorem 3.1.1. The proof is deferred to Section 3.5. □

Then, we obtain the following lemma.

Lemma 3.1.2 (Bounded timestep moments)**.** *If the SDE satisfies Assumption 3.1.1 , and the timestep function h satisfies Assumptions 3.1.2 and 2.1.3, then for all $T \in [1, \infty)$, $p \in (0, \infty)$ there exists a constant $C_{h,p} \in (0, \infty)$ which depends on p and C_p in Theorem 3.1.1 such that it holds that*

$$\mathbb{E}\left[(N_T - 1)^p\right] < C_{h,p} T^p. \tag{3.6}$$

Proof of Lemma 3.1.2. By Assumption 2.1.3, we obtain that

$$N_T = \sum_{k=1}^{N_T} 1 = \sum_{k=1}^{N_T} \frac{h(\widehat{X}_{t_k})}{h(\widehat{X}_{t_k})} = \int_0^T \frac{1}{h(\overline{X}_t)} \mathrm{d}t + 1 \leq \int_0^T (\xi \|\overline{X}_t\|^q + \zeta + \frac{1}{T}) \mathrm{d}t + 1. \tag{3.7}$$

Therefore, the Jensen's inequality ensures that

$$\mathbb{E}\left[(N_T - 1)^p\right] \leq 2^{p-1} \left(T^{p-1} \int_0^T \mathbb{E}\left[\left(\xi \|\overline{X}_t\|^q + \zeta\right)^p \right] \mathrm{d}t + 1 \right). \tag{3.8}$$

Combining this and Theorem 3.1.1 completes the proof of Lemma 3.1.2. □

3.1.2 Strong convergence

To prove an order of strong convergence requires new assumptions on f and g:

Assumption 3.1.3 (Contractive Lipschitz properties). *For some fixed* $p^* \in [2, \infty)$, *there exist constants* $\lambda, \eta \in (0, \infty)$ *such that for all* $x, y \in \mathbb{R}^m$, f *and* g *satisfy the contractive Lipschitz condition*

$$\langle x-y, f(x)-f(y)\rangle + \frac{p^*-1}{2}\,\|g(x)-g(y)\|^2 \leq -\lambda\,\|x-y\|^2, \tag{3.9}$$

and g *satisfies the Lipschitz condition*

$$\|g(x)-g(y)\|^2 \leq \eta\,\|x-y\|^2. \tag{3.10}$$

In addition, f *satisfies the local polynomial growth Lipschitz condition* (2.19).

This Assumption ensures that two solutions to this SDE starting from different places but driven by the same Brownian increment, will come together exponentially, as shown in the following lemma.

Lemma 3.1.3 (SDE contractivity). *If the SDE satisfies Assumption 3.1.3, then for any* $p \in [2, p^*]$, *any two solutions to the SDE:* X_t *and* Y_t, *driven by the same Brownian motion but starting from* x_0 *and* y_0, *where* $x_0 \neq y_0$, *satisfy, for all* $t \in (0, \infty)$, *the bound*

$$\mathbb{E}\left[\|X_t - Y_t\|^p\right] \leq \mathrm{e}^{-\lambda p t}\,\|x_0 - y_0\|^p. \tag{3.11}$$

Proof of Lemma 3.1.3. First, we define that $e_t \triangleq X_t - Y_t$, and since X_t and Y_t are driven by the same Brownian motion, we obtain that

$$\mathrm{d}e_t = (f(X_t) - f(Y_t))\,\mathrm{d}t + (g(X_t) - g(Y_t))\,\mathrm{d}W_t. \tag{3.12}$$

By Itô's formula, we have for any $t \in (0, T]$ that

$$\begin{aligned}
\mathrm{e}^{\lambda p t}\|e_t\|^p - \|e_0\|^p \leq & \int_0^t \lambda p\,\mathrm{e}^{\lambda p s}\|e_s\|^p\,\mathrm{d}s + \int_0^t p\langle e_s, f(X_s) - f(Y_s)\rangle \mathrm{e}^{\lambda p s}\|e_s\|^{p-2}\,\mathrm{d}s \\
& + \int_0^t \frac{p(p-1)}{2}\|g(X_s) - g(Y_s)\|^2 \mathrm{e}^{\lambda p s}\|e_s\|^{p-2}\mathrm{d}s \\
& + \int_0^t p\,\mathrm{e}^{\lambda p s}\|e_s\|^{p-2}\langle e_s, (g(X_s) - g(Y_s))\,\mathrm{d}W_s\rangle.
\end{aligned} \tag{3.13}$$

Therefore, by taking expectations on both sides and using the contractive Lipschitz property (3.9), we obtain that

$$\mathbb{E}\left[\mathrm{e}^{\lambda pt}\|e_t\|^p\right] \le \|e_0\|^p = \|x_0 - y_0\|^p. \tag{3.14}$$

This completes the proof of Lemma 3.1.3. □

This lemma means that the error made in previous time steps will decay exponentially and then we can prove a uniform bound for the strong error.

Theorem 3.1.2 (Strong convergence order)**.** *If the SDE satisfies Assumption 3.1.3, and the timestep function h^δ satisfies Assumption 2.1.4 with h satisfying Assumption 3.1.2, then for all $p \in (0, p^*]$, there exists a constant $C_p \in (0, \infty)$ such that for all $t \in [0, \infty)$ it holds that*

$$\mathbb{E}\left[\,\|\widehat{X}_t - X_t\|^p\right] \le C_p\, \delta^{p/2}. \tag{3.15}$$

Proof of Theorem 3.1.2. The proof is deferred to Section 3.5. □

For the finite time interval $[0, T]$, we can show that the expected number of timesteps per path increases linearly in T which is the same as for the case of uniform timesteps.

Lemma 3.1.4 (Number of timesteps)**.** *If the SDE satisfies Assumption 3.1.1 , and the timestep function h^δ satisfies Assumption 2.1.4 with h satisfying Assumptions 3.1.2 and 2.1.3, then for all $T, p \in (0, \infty)$ there exists a constant $C_{h,p} \in (0, \infty)$ same as in Lemma 3.1.2 such that*

$$\mathbb{E}\left[(N_T - 1)^p\right] \le C_{h,p}\, T^p\, \delta^{-p}. \tag{3.16}$$

Proof of Lemma 3.1.4. The proof is very similar to the proof of Lemma 3.1.2, noting that

$$h^\delta(x) \ge \delta\, h(x) \ge \delta\left(\xi\|x\|^q + \zeta\right)^{-1}, \tag{3.17}$$

due to Assumptions 2.1.3 and 2.1.4. Combing the uniform moment bound from Theorem 3.1.1 and the inequality (3.8) completes the proof of Lemma 3.1.4. □

Again, we can prove first order strong convergence for SDEs with a uniform diffusion coefficient.

Assumption 3.1.4 (Enhanced contractive Lipschitz properties)**.** *There exists a constant* $\lambda \in (0,\infty)$ *such that for all* $x, y \in \mathbb{R}^m$*,* f *satisfies the contractive one-sided Lipschitz condition that*

$$\langle x-y, f(x)-f(y)\rangle \leq -\lambda \|x-y\|^2. \tag{3.18}$$

In addition, f *is differentiable, and* f *and* ∇f *satisfy the local polynomial growth Lipschitz condition* (2.19)*.*

We now state the theorem on improved strong convergence.

Theorem 3.1.3 (Strong convergence for Langevin SDEs)**.** *If* $m=d$*,* $g \equiv I_m$*,* f *satisfies Assumption 3.1.4, and the timestep function* h^δ *satisfies Assumption 2.1.4 with* h *satisfying Assumption 2.1.3, then for all* $p \in (0,\infty)$ *there exists a constant* $C_p \in (0,\infty)$ *such that for all* $t \in [0,\infty)$ *it holds that*

$$\mathbb{E}\left[\|\widehat{X}_t - X_t\|^p\right] \leq C_p\, \delta^p. \tag{3.19}$$

Proof of Theorem 3.1.3. The proof is deferred to Section 3.5. □

3.2 Adaptive MLMC for invariant distributions

We are interested in the problem of approximating

$$\pi(\varphi) := \mathbb{E}_\pi \varphi = \int_{\mathbb{R}^m} \varphi(x)\pi(\mathrm{d}x), \tag{3.20}$$

where π is the invariant measure of the SDE (1.1). Numerically, we can approximate this quantity by simulating $\mathbb{E}\left[\varphi(X_T)\right]$ for a sufficiently large T. In the following subsections, we will introduce our adaptive multilevel Monte Carlo algorithm and its numerical analysis.

3.2.1 Algorithm

Consider the identity of the standard MLMC (1.12) that

$$\mathbb{E}\left[\varphi_L\right] = \mathbb{E}\left[\varphi_0\right] + \sum_{\ell=1}^{L} \mathbb{E}\left[\varphi_\ell - \varphi_{\ell-1}\right], \tag{3.21}$$

where $\varphi_\ell := \varphi(\widehat{X}_T^\ell)$ with $\widehat{X}_T^\ell$ being the numerical estimator of X_T, which uses the adaptive function h^δ with $\delta = M^{-\ell}$ for some fixed $M \in [1, \infty)$. Then the standard MLMC estimator is the following telescoping sum

$$\frac{1}{N_0}\sum_{n=1}^{N_0} \varphi(\widehat{X}_T^{(n,0)}) + \sum_{\ell=1}^{L}\left\{\frac{1}{N_\ell}\sum_{n=1}^{N_\ell}\left(\varphi(\widehat{X}_T^{(n,\ell)}) - \varphi(\widehat{X}_T^{(n,\ell-1)})\right)\right\}, \tag{3.22}$$

where $\widehat{X}_T^{(n,\ell)}$ is the terminal value of the nth numerical path in the time interval $[0, T]$ using a suitable adaptive function h^δ with $\delta = M^{-\ell}$.

In contrast with the standard MLMC with a fixed time interval $[0, T]$, we now allow different levels to have different lengths of the time interval T_ℓ, satisfying $0 < T_0 < T_1 < \cdots < T_\ell < \cdots < T_L = T$, which means that as the level ℓ increases, we obtain a better approximation not only by using smaller timesteps but also by simulating a longer time interval. However, the difficulty is how to construct a good coupling on each level ℓ since the fine path and coarse path have different lengths of time interval T_ℓ and $T_{\ell-1}$.

Following the idea of Glynn and Rhee [38] to estimate the invariant measure of some Markov chains, we perform the coupling by starting a level ℓ fine path simulation at time $t_0^f = -T_\ell$ and a coarse path simulation at time $t_0^c = -T_{\ell-1}$ and terminate both paths at $t = 0$. Since the drift f and volatility g do not depend explicitly on time t, the distribution of the numerical solution simulated on the time interval $[-T_\ell, 0]$ is same as one simulated on $[0, T_\ell]$. The key point here is that the fine path and coarse path share the same driving Brownian motion during the overlap time interval $[-T_{\ell-1}, 0]$. Owing to the result of Lemma 3.1.3, two solutions to the SDE satisfying Assumption 3.1.3, starting from different initial points and driven by the same Brownian motion will converge exponentially. Therefore, the fact that different levels terminate at the same time is crucial to the variance reduction of the multilevel scheme.

Our new multilevel scheme still satisfies the identity (3.21) but with $\varphi_\ell = \varphi(\widehat{X}_0^\ell)$ with $\widehat{X}_0^\ell$ being the terminal value of the numerical path approximation on the time interval $[-T_\ell, 0]$ using adaptive function h^δ with $\delta = M^{-\ell}$. The corresponding new MLMC estimator is

$$\widehat{Y} \coloneqq \frac{1}{N_0}\sum_{n=1}^{N_0}\varphi(\widehat{X}_0^{(n,0)}) + \sum_{\ell=1}^{L}\left\{\frac{1}{N_\ell}\sum_{n=1}^{N_\ell}\left(\varphi(\widehat{X}_0^{(n,\ell)}) - \varphi(\widehat{X}_0^{(n,\ell-1)})\right)\right\}, \tag{3.23}$$

where $\widehat{X}_0^{(n,\ell)}$ is the terminal value of the nth numerical path through the time interval $[-T_\ell, 0]$ using the adaptive function h^δ with $\delta = M^{-\ell}$. Figure 3.1 and Algorithm 1 illustrate the detailed implementation of a single adaptive MLMC sample using a non-nested adaptive timestep on level ℓ with $M = 2$.

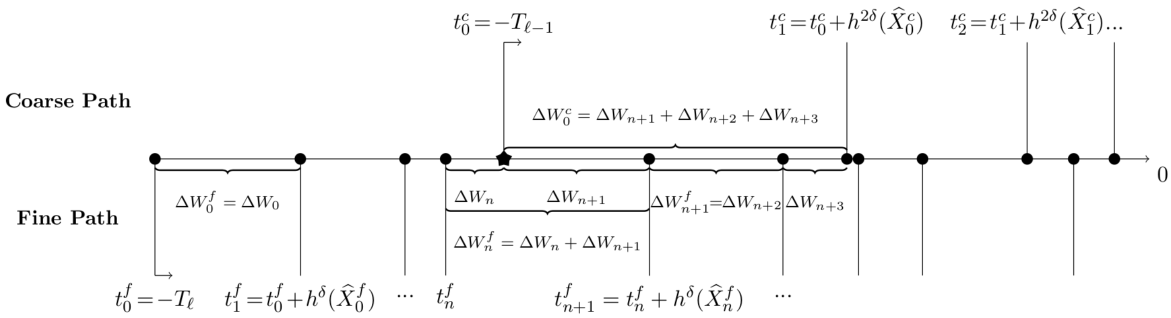


Figure 3.1: Algorithm of the adaptive MLMC for an infinite interval

3.2.2 Numerical analysis

First, we state the exponential convergence to the invariant measure of the original SDE, which can help us to measure the approximation error caused by truncating the infinite time interval.

Lemma 3.2.1 (Exponential convergence). *If the SDE satisfies Assumptions 3.1.1 and 3.1.3 and φ satisfies the polynomial growth Lipschitz condition* (2.19), *then*

Algorithm 1: Outline of the algorithm for a single adaptive MLMC sample for scalar SDE on level ℓ in time interval $[-T_\ell, 0]$.

p $t := -T_\ell$; $t^c := -T_{\ell-1}$; $t^f := -T_\ell$;
$h^c := 0$; $h^f := 0$;
$\Delta W^c := 0$; $\Delta W^f := 0$;
$\widehat{X}^c = x_0$; $\widehat{X}^f = x_0$;
while $t < 0$ **do**
  $t_{old} := t$;
  $t := \min(t^c, t^f)$;
  $\Delta W := N(0, t - t_{old})$;
  $\Delta W^c := \Delta W^c + \Delta W$;
  if $t = -T_{\ell-1}$ **then**
    $\Delta W^c := 0$;
  end
  $\Delta W^f := \Delta W^f + \Delta W$;
  if $t = t^c$ **then**
    update coarse path $\widehat{X}^c$ using h^c and ΔW^c;
    compute new adapted coarse path timestep $h^c = h^{2\delta}(\widehat{X}^c)$;
    $h^c := \min(h^c, -t^c)$;
    $t^c := t^c + h^c$;
    $\Delta W^c := 0$;
  end
  if $t = t^f$ **then**
    update fine path $\widehat{X}^f$ using h^f and ΔW^f;
    compute new adapted fine path timestep $h^f = h^{\delta}(\widehat{X}^f)$;
    $h^f := \min(h^f, -t^f)$;
    $t^f := t^f + h^f$;
    $\Delta W^f := 0$;
  end
end
Result: $\widehat{X}^f - \widehat{X}^c$

there exists a constant $\mu_0 \in (0, \infty)$ *depending on* x_0 *and the constants in Lemma 3.1.1 and 3.1.3 such that*

$$|\mathbb{E}\left[\varphi(X_t) - \pi(\varphi)\right]| \leq \mu_0 \, \mathrm{e}^{-\lambda t}. \tag{3.24}$$

Proof of Lemma 3.2.1. Let Y_0 be a new random variable which follows the invariant measure π. Then the solution Y_t to the SDE with the initial value Y_0 will also follows the invariant measure for any $t \in (0, \infty)$. Therefore, by the polynomial growth Lipschitz property of φ and Lemmas 3.1.1 and 3.1.3 and Hölder's inequality, there exist constants $\mu_0, \mu_1 \in (0, \infty)$ such that

$$\begin{aligned} |\mathbb{E}\left[\varphi(X_t) - \pi(\varphi)\right]| &= |\mathbb{E}\left[\varphi(X_t) - \varphi(Y_t)\right]| \leq \mathbb{E}\left[(\gamma(\|X_t\|^q + \|Y_t\|^q) + \mu) \, \|X_t - Y_t\|\right] \\ &\leq \mathbb{E}\left[|\gamma(\|X_t\|^q + \|Y_t\|^q) + \mu|^2\right]^{1/2} \mathbb{E}\left[\|X_t - Y_t\|^2\right]^{1/2} \qquad (3.25) \\ &\leq \mu_1 \mathbb{E}\left[\|X_0 - Y_0\|^2\right]^{1/2} \mathrm{e}^{-\lambda t} \leq 2\mu_1 \left[\|x_0\| + C_1\right] \mathrm{e}^{-\lambda t} := \mu_0 \mathrm{e}^{-\lambda t}. \end{aligned}$$

This completes the proof of Lemma 3.2.1. □

Note that Assumption 3.1.3 is a sufficient condition for this Lemma. We use it here to show that the contractivity rate λ is a lower bound for the true convergence rate λ^* and it is λ that determines the choice of T_ℓ shown in the following results.

Now, we first bound the variance of the MLMC correction for each level.

Lemma 3.2.2 (Variance of MLMC corrections for a bounded diffusion coefficient)**.** *If* φ *satisfies the polynomial growth Lipschitz condition* (2.19)*, the SDE satisfies Assumption 3.1.3 and the timestep function* h^δ *satisfies Assumption 2.1.4 with* h *satisfying Assumption 3.1.2, and* $\delta = M^{-\ell}$ *for each level, then for each level* ℓ*, there exist constants* c_1 *and* c_2 *such that the variance of correction* $V_\ell := \mathbb{V}\left[\varphi(\widehat{X}_0^\ell) - \varphi(\widehat{X}_0^{\ell-1})\right]$ *satisfies that*

$$V_\ell \leq c_1 \, M^{-\ell} + c_2 \, \mathrm{e}^{-2\lambda T_{\ell-1}}. \tag{3.26}$$

Proof of Lemma 3.2.2. By the polynomial growth Lipschitz condition (2.19) of φ, Hölder's inequality and Theorem 3.1.1, there exists a constant $\kappa \in (0, \infty)$ such that

$$V_\ell \le \mathbb{E}\left[\left|\varphi(\widehat{X}_0^\ell) - \varphi(\widehat{X}_0^{\ell-1})\right|^2\right] \le \kappa\, \mathbb{E}\left[\left\|\widehat{X}_0^\ell - \widehat{X}_0^{\ell-1}\right\|^{p^*}\right]^{2/p^*}. \tag{3.27}$$

$\widehat{X}_t^\ell$ and $\widehat{X}_t^{\ell-1}$ share the same driving Brownian motion from $-T_{\ell-1}$ to 0. We can denote the corresponding solution to the SDE (1.1) starting from x_0 at time $-T_{\ell-1}$ and driven by the same Brownian motion as $\widehat{X}_t^{\ell-1}$ through time interval $[-T_{\ell-1}, 0]$ by X_t^c, and the solution starting from x_0 at time $-T_\ell$ driven by the same Brownian motion as $\widehat{X}_t^\ell$ through the time interval $[-T_\ell, 0]$ by X_t^f.

Then, by Jensen's inequality, we obtain that

$$\mathbb{E}\left[\left\|\widehat{X}_0^\ell - \widehat{X}_0^{\ell-1}\right\|^{p^*}\right] \le 3^{p^*-1}\,(E_1 + E_2 + E_3), \tag{3.28}$$

where

$$\begin{aligned} E_1 &= \mathbb{E}\left[\left\|X_0^c - \widehat{X}_0^{\ell-1}\right\|^{p^*}\right], \\ E_2 &= \mathbb{E}\left[\left\|\widehat{X}_0^\ell - X_0^f\right\|^{p^*}\right], \\ E_3 &= \mathbb{E}\left[\left\|X_0^f - X_0^c\right\|^{p^*}\right]. \end{aligned} \tag{3.29}$$

Theorem 3.1.2 implies that there exists a constant $C_{p^*} \in (0, \infty)$ which does not depend on T_ℓ such that

$$E_1 \le C_{p^*} M^{-p^*(\ell-1)/2},\ E_2 \le C_{p^*} M^{-p^*\ell/2}, \tag{3.30}$$

and Lemma 3.1.1 and Lemma 3.1.3 imply that there exists a constant $C \in (0, \infty)$ depending on x_0 and C_4 in Lemma 3.1.1 such that

$$\begin{aligned} E_3 &\le \mathbb{E}\left[\|X_{-T_{\ell-1}}^f - x_0\|^{p^*}\right] \mathrm{e}^{-p^*\lambda T_{\ell-1}} \\ &\le 2^{p^*-1}\left(\|x_0\|^{p^*} + \mathbb{E}\left[\|X_{-T_{\ell-1}}^f\|^{p^*}\right]\right) \mathrm{e}^{-p^*\lambda T_{\ell-1}} \le C\,\mathrm{e}^{-p^*\lambda T_{\ell-1}}. \end{aligned} \tag{3.31}$$

Finally, by the fact that $a^v + b^v \geq (a+b)^v$ for any $a, b \in (0,\infty)$ and $v \in (0,1)$, there exist constants $c_1, c_2 \in (0,\infty)$ such that

$$\begin{aligned} V_\ell &\leq \kappa \left[3^{p^*-1}\left(C_{p^*}M^{-p^*(\ell-1)/2} + C_{p^*}M^{-p^*\ell/2} + C\mathrm{e}^{-p^*\lambda T_{\ell-1}}\right)\right]^{2/p^*} \\ &\leq c_1\, M^{-\ell} + c_2\, \mathrm{e}^{-2\lambda T_{\ell-1}}. \end{aligned} \tag{3.32}$$

This completes the proof of Lemma 3.2.2. □

Given this, we obtain the following theorem for the complexity of the MLMC algorithm to achieve a specified MSE accuracy.

Theorem 3.2.1 (MLMC for invariant measure)**.** *If φ satisfies the polynomial growth Lipschitz condition* (2.19)*, the SDE satisfies Assumption 3.1.3 and the timestep function h^δ satisfies Assumption 2.1.4 with h satisfying Assumption 3.1.2 and $\delta = M^{-\ell}$ for each level, then by choosing suitable values for L and T_ℓ, N_ℓ for each level ℓ, there exists a constant c_3 such that the MLMC estimator* (3.23) *has a MSE satisfying the bound*

$$\mathbb{E}\left[(\widehat{Y} - \pi(\varphi))^2\right] \leq \varepsilon^2, \tag{3.33}$$

and an expected computational cost C_{MLMC} satisfying the bound

$$\mathrm{C}_{MLMC} \leq c_3\, \varepsilon^{-2} |\log \varepsilon|^3. \tag{3.34}$$

Proof of Theorem 3.2.1. By Jensen's inequality, the mean square error can be decomposed into three parts:

$$\begin{aligned} \mathbb{E}\left[(\widehat{Y} - \pi(\varphi))^2\right] &= \mathbb{V}\left[\widehat{Y}\right] + \left|\mathbb{E}\left[\widehat{Y}\right] - \pi(\varphi)\right|^2 \qquad (3.35) \\ &\leq \mathbb{V}\left[\widehat{Y}\right] + 2\left|\mathbb{E}\left[\widehat{Y}\right] - \mathbb{E}\left[\varphi(X_{T_L})\right]\right|^2 + 2\left|\mathbb{E}\left[\varphi(X_{T_L})\right] - \pi(\varphi)\right|^2, \end{aligned}$$

which enables us to achieve the MSE bound by bounding each part by $\varepsilon^2/3$.

If we set that

$$T_\ell = (\ell+1)\log M/2\lambda, \tag{3.36}$$

then it holds that $V_\ell \le (c_1 + c_2)M^{-\ell}$, which has the same order of magnitude as the variance bound for the standard MLMC theorem. Lemma 3.2.1 implies that

$$2\left|\mathbb{E}\left[\varphi(X_{T_L})\right] - \pi(\varphi)\right|^2 \le 2\,\mu_0^2\,\mathrm{e}^{-2\lambda T_L} \le \frac{\varepsilon^2}{3}, \tag{3.37}$$

provided that

$$L \ge \left\lceil \frac{2|\log\varepsilon|}{\log M} + \frac{\log(6\mu_0^2)}{\log M} \right\rceil. \tag{3.38}$$

By Theorems 3.1.1 and 3.1.2, the polynomial growth Lipschitz condition (2.19) of φ and Hölder's inequality, there exist constants $\kappa_1, \kappa_2 \in (0,\infty)$ such that

$$\begin{aligned} 2\left|\mathbb{E}\left[\widehat{Y}\right] - \mathbb{E}\left[\varphi(X_{T_L})\right]\right|^2 \;&=\; 2\left|\mathbb{E}\left[\varphi(\widehat{X}^L_{T_L}) - \varphi(X_{T_L})\right]\right|^2 \\ &\le\; 2\kappa_1\,\mathbb{E}\left[\|\widehat{X}^L_{T_L} - X_{T_L}\|^4\right]^{1/2} \;\le\; \kappa_2 M^{-L} \;\le\; \frac{\varepsilon^2}{3}, \end{aligned} \tag{3.39}$$

provided that

$$L \ge \left\lceil \frac{2|\log\varepsilon|}{\log M} + \frac{\log(3\kappa_2)}{\log M} \right\rceil. \tag{3.40}$$

Combining the requirements (3.38) and (3.40), we choose

$$L = \left\lceil \frac{2|\log\varepsilon|}{\log M} + \frac{\log\left(\max(6\mu_0^2, 3\kappa_2)\right)}{\log M} \right\rceil, \tag{3.41}$$

giving $L = O(|\log\varepsilon|)$ as $\varepsilon \to 0$. Therefore, we have $V_\ell = O(M^{-\ell})$ and $C_\ell = O(\ell M^\ell)$, where C_ℓ is the expected cost of a sample on level ℓ. Following the analysis in [32], by choosing

$$N_\ell = \left\lceil 3\,(c_1 + c_2)\,\frac{M^{-\ell}}{\sqrt{\ell+1}}\;\varepsilon^{-2}\sum_{\ell'=0}^{L}\sqrt{\ell'+1} \right\rceil, \tag{3.42}$$

to ensure that the overall variance is less than $\frac{\varepsilon^2}{3}$, then the expected total cost is bounded by, for some constant $C_0 \in (0,\infty)$,

$$\mathrm{C}_{MLMC} \;\le\; 3\,C_0\,(c_1 + c_2)\,\varepsilon^{-2}\left(\sum_{\ell=0}^{L}\sqrt{\ell+1}\right)^2 + C_0\sum_{\ell=0}^{L}(\ell+1)M^\ell. \tag{3.43}$$

Since

$$\sum_{\ell=0}^{L} \sqrt{\ell+1} \;\leq\; \int_0^{L+1} \sqrt{x+1}\, \mathrm{d}x \;\leq\; \tfrac{2}{3}(L+2)^{3/2} \;=\; O(|\log\varepsilon|^{3/2}), \tag{3.44}$$

and

$$\sum_{\ell=0}^{L} (\ell+1) M^{\ell} \;\leq\; (L+1)^2 M^L \;=\; O(\varepsilon^{-2}|\log\varepsilon|^2), \tag{3.45}$$

we obtain the desired final result that there exists a constant c_3 such that

$$\mathrm{C}_{MLMC} \leq c_3\, \varepsilon^{-2} |\log\varepsilon|^3. \tag{3.46}$$

This completes the proof of Theorem 3.2.1. □

For the SDEs with uniform diffusion coefficient, the computational cost can be reduced to $O(\varepsilon^{-2})$.

Theorem 3.2.2 (SDEs with uniform diffusion coefficient)**.** *If φ satisfies the polynomial growth Lipschitz condition* (2.19)*, and for the SDE, $m = d$, $g \equiv I_m$, f satisfies Assumption 3.1.4, and the timestep function h^δ satisfies Assumption 2.1.4 with h satisfying Assumption 3.1.2 and $\delta = M^{-\ell}$ for each level, then for each level ℓ, there exist constants $c_1, c_2 \in (0, \infty)$ such that*

$$V_\ell \leq c_1\, M^{-2\ell} + c_2\, \mathrm{e}^{-2\lambda T_{\ell-1}}. \tag{3.47}$$

Furthermore, by choosing suitable L, T_ℓ and N_ℓ for each level ℓ in the MLMC estimator (3.23)*, one can achieve the MSE bound ε^2 at an expected computational cost bounded by*

$$\mathrm{C}_{MLMC} \leq c_3\, \varepsilon^{-2}, \tag{3.48}$$

for some constant $c_3 \in (0, \infty)$.

Proof of Theorem 3.2.2. Following a similar argument to the proof of Lemma 3.2.2, Theorem 3.1.3 implies $V_\ell \leq c_1\, M^{-2\ell} + c_2\, \mathrm{e}^{-2\lambda T_{\ell-1}}$, and by choosing T_ℓ to be

$$T_\ell = (\ell+1) \log M / \lambda, \tag{3.49}$$

we obtain $V_\ell \le (c_1 + c_2)M^{-2\ell}$. The computational cost of a single MLMC sample on level ℓ satisfies

$$C_\ell \;\le\; C_0(\ell{+}1)M^\ell \;\le\; C\,M^{(1+\epsilon)\ell}, \tag{3.50}$$

for any $0{<}\epsilon{\ll}1$ and some $C \in (0,\infty)$. Therefore, the standard MLMC Theorem 1.1.2 is applicable with $\gamma < \beta$, giving an $O(\varepsilon^{-2})$ complexity. This completes the proof of Theorem 3.2.2. □

Note that the choice of T_ℓ (3.49) for the equation with a uniform diffusion coefficient is different from (3.36) for SDEs with bounded diffusion coefficient. In other words, the strong convergence result and the contractive convergence rate λ together determine T_ℓ. In some cases, λ needs to be estimated numerically through Lemma 3.1.3. The difference in the variance convergence rate also affects the choice of M. Based on the analysis in [30], the optimal M for SDEs with general g is in the range $4-8$, while in the uniform diffusion coefficient case the optimal M is around 2.

3.3 Numerical experiment

In this section we present numerical results for the following scalar SDE

$$\mathrm{d}X_t = \left(-X_t - X_t^3\right)\,\mathrm{d}t + \mathrm{d}W_t, \tag{3.51}$$

which satisfies both the dissipativity condition (3.1) and the contractive condition (3.18). Our interest is to compute $\pi(\varphi)$ where $\varphi(x) = (x+1)^2$ satisfying the polynomial growth Lipschitz condition.

Since the probability density function π is

$$\frac{\exp(-x^2 - \frac{1}{2}x^4)}{\int_{-\infty}^{\infty} \exp(-x^2 - \frac{1}{2}x^4)\,\mathrm{d}x}, \tag{3.52}$$

we can use numerical integration to calculate an approximate value: $\varphi(\pi) \approx 1.2896$ with accuracy 10^{-5}, and use this value as a benchmark for our numerical tests.

Next we need to determine T_ℓ for each level. Linear perturbations to the SDE satisfy the ODE

$$\mathrm{d}Y_t = -\left(1 + 3\,X_t^2\right) Y_t\,\mathrm{d}t, \tag{3.53}$$

and therefore $\lambda \geq 1$. Hence we choose to use $T_\ell = (\ell+1)\log 2$ to ensure that the truncation error is acceptably small.

Figure 3.2 displays the variance of the multilevel correction on each level as a function of T; this is to be compared to the bound in result (3.47). The exponential part dominates the variance at the beginning, so the variance decays exponentially. As time increases, the $M^{-2\ell}$ term becomes the major part of the variance, and the variance stops decreasing.

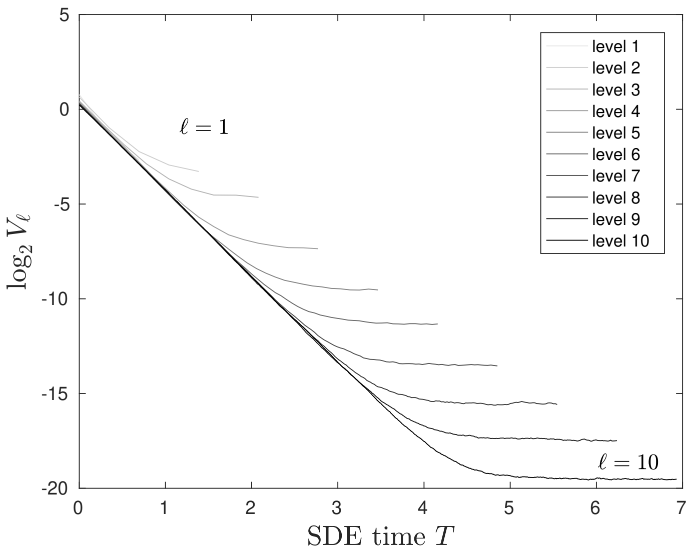


Figure 3.2: Variance of corrections on each level ℓ (test case)

Figure 3.3 presents the MLMC results. The top right plot shows first order convergence for the weak error and the top left plot shows second order convergence for the multilevel correction variance. Hence the computational cost for RMS accuracy ε is $O(\varepsilon^{-2})$ which is verified in the bottom right plot, while the bottom left plot shows the number of MLMC samples on each level as a function of the target accuracy. Here, we also compared our MLMC scheme with standard Monte Carlo

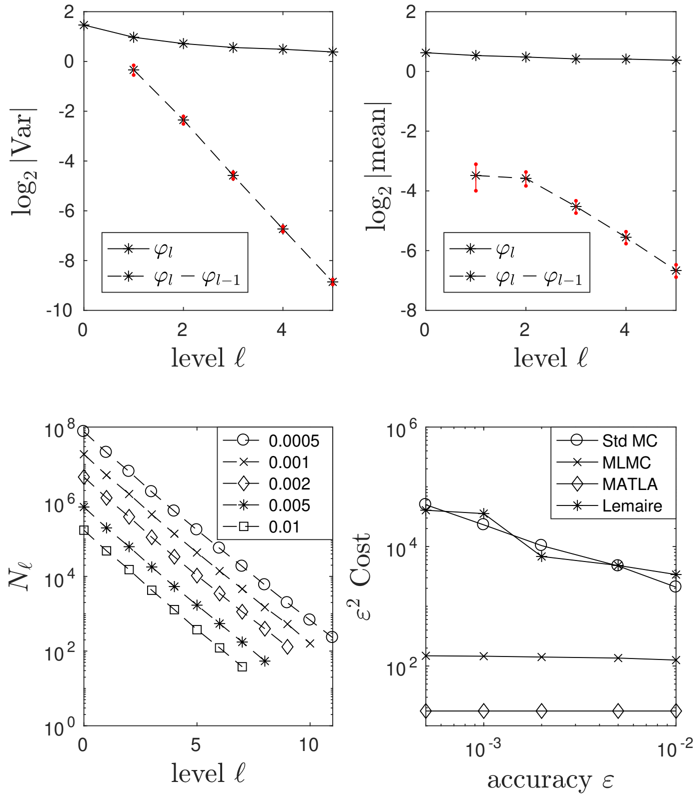


Figure 3.3: Adaptive MLMC for invariant distribution (test case)

(Standard MC) method directly simulating $\widehat{X}_{T_L}$, the adaptive scheme proposed by Lemaire using same step sequence as in Example 7.1 in [64], and the MATLA algorithm in [86] with timestep $h = 0.1$. Both standard MC and the adaptive scheme by Lemaire have the order $O(\varepsilon^{-3})$. MLMC and MATLA have the optimal complexity $O(\varepsilon^{-2})$. In this case, MATLA performs better due to the relatively short mixing time and low correlations. However, MATLA and the adaptive scheme by Lemaire only simulate one path and are difficult to perform by parallel computing. In addition, MATLA can only be applied when the density of the invariant measure is known up to a constant.

3.4 Extension to a larger class of ergodic SDEs

In practice, the condition (3.9) in Assumption 3.1.3 is restrictive and means that the SDE is contractive everywhere in the whole space. However, this condition is only a sufficient condition to make the adaptive MLMC work. Numerically, our adaptive MLMC does not need contractivity everywhere but do require contractivity in a global sense. For example, the double-well potential SDEs which only lose the contractivity near the connection area of the two wells. Therefore, intuitively, our scheme works well for the systems with negative Lyapunov exponent.

3.4.1 Double-well potential SDE

First, we apply adaptive MLMC for a 100-dimensional SDEs

$$\mathrm{d}X_t \;=\; \left(X_t - \frac{1}{100}\|X_t\|^2 X_t\right)\mathrm{d}t + \mathrm{d}W_t,\;\; x_0 = 0. \tag{3.54}$$

Our interest is to compute $\pi(\varphi)$ with $\varphi(x) = \|x\|^2$ satisfying the polynomial growth Lipschitz condition. Although this SDE does not satisfy the Assumption 3.1.3, it has negative Lyapunov exponent and the numerical estimation of the contractivity rate λ in Lemma 3.1.3 is 0.15. We choose T_ℓ based on equation (3.49).

Figure 3.4 shows the variance decays due to the contractivity and the first order strong convergence. The convergence results for the adaptive MLMC and comparisons with other schemes are shown in Figure 3.5. We use MATLA with timestep $h = 0.02$ as the optimal scaling suggested in [85] and the adaptive scheme proposed by Lemaire using same step sequence as in Example 7.1 in [64].

3.4.2 FENE model

We extend our adaptive MLMC scheme to the FENE model from test case 4 in the previous chapter. Figure 3.6 shows that the adaptive MLMC also works well

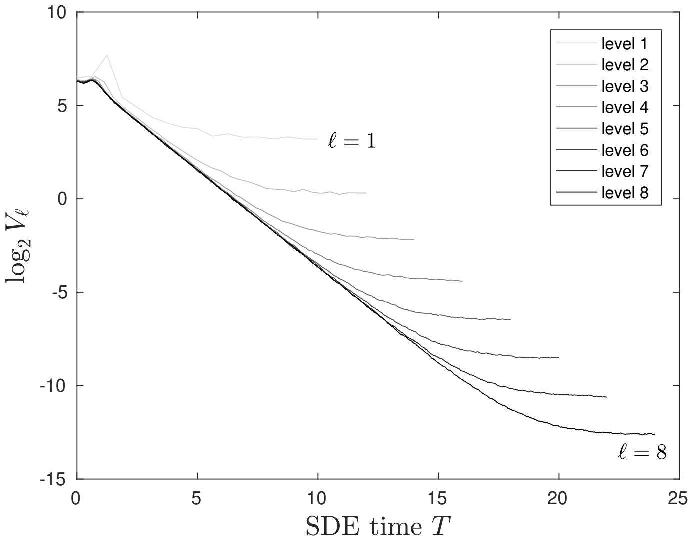


Figure 3.4: Variance of corrections on each level ℓ (double-well)

and achieves the optimal computational cost $O(\varepsilon^{-2})$ for the invariant measure computation. None of the other methods are applicable here.

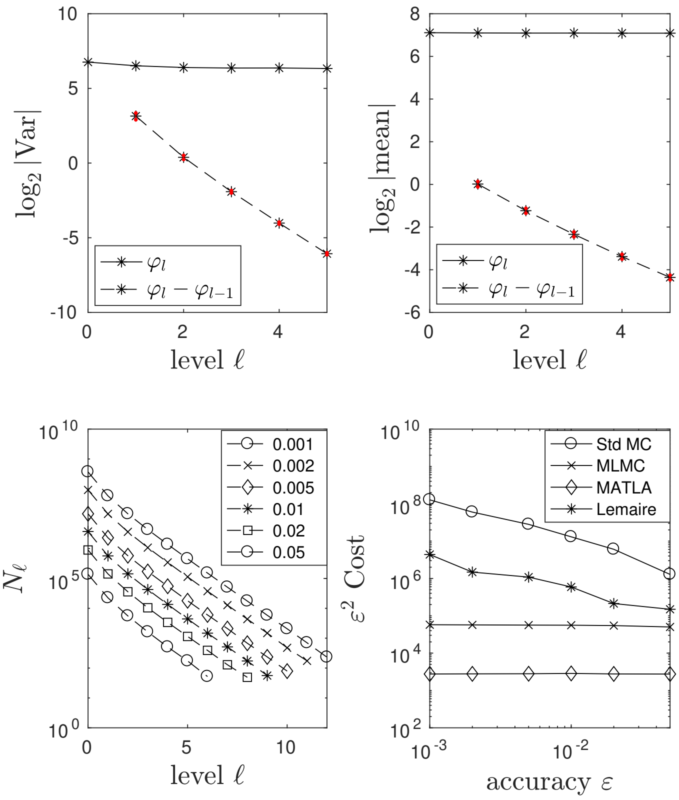


Figure 3.5: Adaptive MLMC for the invariant distribution (double-well)

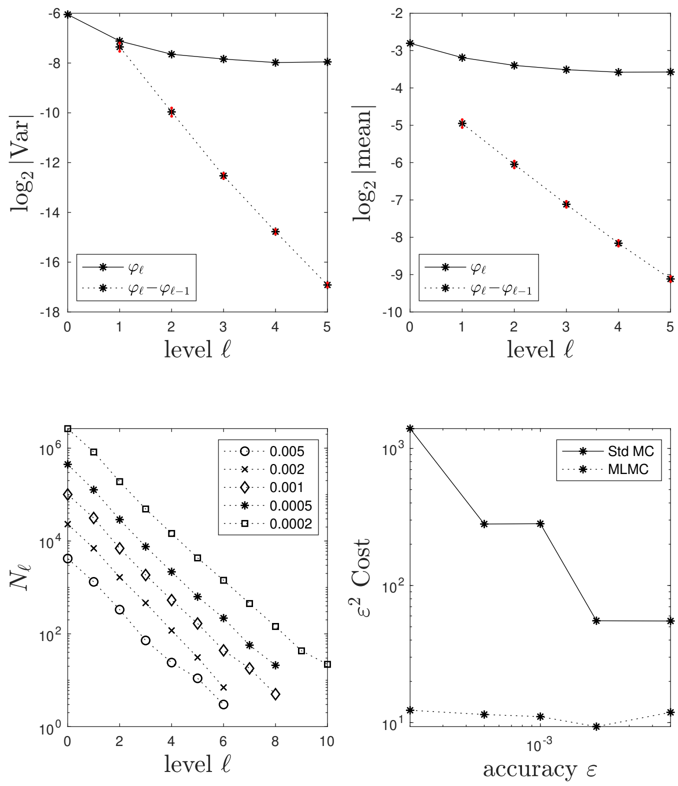


Figure 3.6: Adaptive MLMC for the invariant distribution (FENE)

3.5 Proofs

This section contains the proofs of the three main theorems in this chapter, one on stability, and two on the order of strong convergence.

3.5.1 Preliminaries

In this subsection, we introduce some inequalities and results we use frequently in the following sections.

Young's inequality

For any $\zeta \in (0,\infty)$ and $P, Q \in (1,\infty)$ satisfying $\frac{1}{P}+\frac{1}{Q} = 1$, and for any $A, B \in (0,\infty)$, the following inequality holds

$$A\,B \;=\; A\zeta\,\frac{B}{\zeta} \;\leq\; \frac{A^P\zeta^P}{P} + \frac{B^Q}{Q\,\zeta^Q}. \tag{3.55}$$

In this chapter, we use two particular cases. First, we take $P = Q = 2$, and $\zeta^2 = 2\xi \in (0,\infty)$ and obtain that

$$A\,B \leq \xi A^2 + \frac{B^2}{4\,\xi}. \tag{3.56}$$

Second, for any $p \in [2,\infty)$ and $\xi \in (0,\infty)$, we take $P = \frac{p}{p-2}$, $Q = \frac{p}{2}$, $A = a^{p-2}$, $B = b^2$ and $\zeta = \xi^{(p-2)/p}$, and obtain that

$$a^{p-2}b^2 \leq \frac{(p-2)\,\xi}{p}\,a^p + \frac{2}{p\,\xi^{(p-2)/2}}\,b^p. \tag{3.57}$$

When using these in proofs, we often keep ξ arbitrary initially and choose it later to make one term sufficiently small, as needed.

Jensen's inequality

One variant of the Jensen's inequality is

$$\left(\sum_{k=0}^{N} \mathrm{e}^{\lambda t_k} h_k\, u_k\right)^p \leq \left(\sum_{k=0}^{N} \mathrm{e}^{\lambda t_k} h_k\right)^{p-1} \sum_{k=0}^{N} \mathrm{e}^{\lambda t_k} h_k u_k^p, \tag{3.58}$$

where $p \in [1, \infty)$ and $h_k,\ u_k \in [0, \infty)$. Its continuous version is

$$\left|\int_0^t \phi(s)\, \mathrm{e}^{\gamma s}\, \mathrm{d}s\right|^p \ \leq\ \left(\int_0^t \mathrm{e}^{\gamma s}\, \mathrm{d}s\right)^{p-1} \int_0^t |\phi(s)|^p \mathrm{e}^{\gamma s}\, \mathrm{d}s, \tag{3.59}$$

where $\gamma \in (0, \infty)$ and ϕ is a real-valued function with finite p-th moment with respect to the distribution with density proportional to $e^{\gamma s}$ in a finite time interval $[0, t]$.

Exponentially weighted supremum

For simplicity, for $\alpha > 0$, we can denote $\widehat{M}_t^{\alpha,p} = \sup_{0\leq s\leq t} \mathrm{e}^{\alpha p s} \|\widehat{X}_s\|^p$, and then by Young's inequality (3.56), for any $\xi > 0$, it holds that

$$\widehat{M}_t^{\alpha,p/2} \leq \xi\, \widehat{M}_t^{\alpha,p} + \frac{1}{4\,\xi}. \tag{3.60}$$

We can also denote $\overline{M}_t^{\alpha,p} = \sup_{0\leq s\leq t} \mathrm{e}^{\alpha p s} \|\overline{X}_s\|^p$, which implies that

$$\overline{M}_t^{\alpha,p} \leq \mathrm{e}^{\alpha p h_{\max}}\, \widehat{M}_t^{\alpha,p}, \tag{3.61}$$

since $\overline{X}_s = \widehat{X}_{\underline{s}}$ and $|s - \underline{s}| \leq h_{\max}$, and

$$\begin{aligned} \int_0^t \mathrm{e}^{\gamma p s/2} \|\overline{X}_s\|^{p/2} \mathrm{d}s \ &\leq\ \overline{M}_t^{\alpha,p/2} \int_0^t \mathrm{e}^{(\gamma-\alpha) p s/2} \mathrm{d}s \\ &\leq\ \frac{2\mathrm{e}^{(\gamma-\alpha)pt/2}}{p\,(\gamma - \alpha)}\, \mathrm{e}^{\alpha p h_{\max}/2} \widehat{M}_t^{\alpha,p/2}. \end{aligned} \tag{3.62}$$

provided $\gamma > \alpha > 0$.

3.5.2 Theorem 3.1.1

Proof of Theorem 3.1.1. By Theorem 2.1.1, we know that T is almost surely attainable. Therefore we can directly analyse our discretization scheme without the K truncation which was used in that proof. The proof proceeds in three steps. First, we derive an upper bound for $e^{\alpha pt}\|\widehat{X}_t\|^p$. Second, we show that the moments $\mathbb{E}[\widehat{M}_t^{\alpha,p}]$ and $\mathbb{E}[\overline{M}_t^{\alpha,p}]$ are each bounded by $C_p e^{\alpha pt}$ where C_p is a constant which only depends on p, x_0, $h_{\max}$ and the constants α, β in Assumption 3.1.2. Finally, we get the uniform bound for $\mathbb{E}[\|\widehat{X}_t\|^p]$ and $\mathbb{E}[\|\overline{X}_t\|^p]$.

The proof is given for $p \in [4, \infty)$; the result for $p \in (0, 4)$ follows from Hölder's inequality.

Step 1: If we denote $\phi(x) = x + h(x) f(x)$, then we obtain that

$$\begin{aligned}\|\widehat{X}_{t_{n+1}}\|^2 \;=\;& \|\widehat{X}_{t_n}\|^2 + 2\, h_n \left(\langle \widehat{X}_{t_n}, f(\widehat{X}_{t_n})\rangle + \tfrac{1}{2} h_n \|f(\widehat{X}_{t_n})\|^2\right) \\ & + 2\, \langle \phi(\widehat{X}_{t_n}), g(\widehat{X}_{t_n})\, \Delta W_n\rangle + \|g(\widehat{X}_{t_n})\, \Delta W_n\|^2. \qquad (3.63)\end{aligned}$$

Using condition (3.4) for h then implies that

$$\begin{aligned}\|\widehat{X}_{t_{n+1}}\|^2 \;\le\;& \|\widehat{X}_{t_n}\|^2 \;-\; 2\,\alpha \|\,\widehat{X}_{t_n}\|^2 h_n \;+\; 2\,\beta\, h_n \\ & + 2\, \langle \phi(\widehat{X}_{t_n}), g(\widehat{X}_{t_n})\, \Delta W_n\rangle + \|g(\widehat{X}_{t_n})\, \Delta W_n\|^2. \qquad (3.64)\end{aligned}$$

Since $1 - 2\alpha h_n \le e^{-2\alpha h_n}$ and g and h are both bounded, we multiply by $e^{2\alpha t_{n+1}}$ on both sides to obtain that

$$\begin{aligned}e^{2\alpha t_{n+1}}\|\widehat{X}_{t_{n+1}}\|^2 \;\le\;& e^{2\alpha t_n}\|\widehat{X}_{t_n}\|^2 + 2e^{2\alpha(t_n + h_{\max})}\,\beta\, h_n + e^{2\alpha(t_n+h_{\max})}\beta \|\,\Delta W_n\|^2 \\ & + 2\, e^{2\alpha t_{n+1}} \langle \phi(\widehat{X}_{t_n}), g(\widehat{X}_{t_n})\, \Delta W_n\rangle. \qquad (3.65)\end{aligned}$$

Similarly, for the partial timestep from $\underline{t}$ to t, since $(t - \underline{t}) \le h_{n_t}$, it holds that

$$\langle \widehat{X}_{\underline{t}}, f(\widehat{X}_{\underline{t}})\rangle + \tfrac{1}{2}(t - \underline{t})\|f(\widehat{X}_{\underline{t}})\|^2 \le -\alpha \|\widehat{X}_{\underline{t}}\|^2 + \beta, \qquad (3.66)$$

and therefore we obtain that

$$\mathrm{e}^{2\alpha t}\|\widehat{X}_t\|^2 \le \mathrm{e}^{2\alpha \underline{t}}\|\widehat{X}_{\underline{t}}\|^2 \; + \; 2\mathrm{e}^{2\alpha(\underline{t}+h_{\max})}\,\beta\,(t-\underline{t}) + \mathrm{e}^{2\alpha(\underline{t}+h_{\max})}\beta\|\,W_t - W_{\underline{t}}\|^2 \\ + 2\,\mathrm{e}^{2\alpha t}\langle \phi(\widehat{X}_{\underline{t}}), g(\widehat{X}_{\underline{t}})\,(W_t - W_{\underline{t}})\rangle. \tag{3.67}$$

Summing (3.65) over multiple timesteps and then adding (3.67) ensures that

$$\mathrm{e}^{2\alpha t}\|\widehat{X}_t\|^2 \;\le\; \|x_0\|^2 \;+\; 2\beta\mathrm{e}^{2\alpha h_{\max}}\left(\sum_{k=0}^{n_t-1}\mathrm{e}^{2\alpha t_k}h_k + \mathrm{e}^{2\alpha\underline{t}}(t-\underline{t})\right) \tag{3.68}$$
$$+ 2\sum_{k=0}^{n_t-1}\mathrm{e}^{2\alpha t_{k+1}}\langle\phi(\widehat{X}_{t_k}), g(\widehat{X}_{t_k})\Delta W_k\rangle) + \beta\mathrm{e}^{2\alpha h_{\max}}\sum_{k=0}^{n_t-1}\mathrm{e}^{2\alpha t_k}\|\Delta W_k\|^2$$
$$+ 2\mathrm{e}^{2\alpha t}\langle\widehat{X}_{\underline{t}} + f(\widehat{X}_{\underline{t}})\,(t-\underline{t}),\; g(\widehat{X}_{\underline{t}})(W_t - W_{\underline{t}})\rangle + \;\beta\mathrm{e}^{2\alpha(\underline{t}+h_{\max})}\|W_t - W_{\underline{t}}\|^2.$$

Bounding the first summation using a Riemann integral, and re-writing the second as an Itô integral, raising both sides to the power $p/2$ and using Jensen's inequality, we obtain that

$$\mathrm{e}^{\alpha p t}\|\widehat{X}_t\|^p \;\le\; 6^{p/2-1}\mathrm{e}^{\alpha p h_{\max}}\left\{\|x_0\|^p \;+\; \left(2\,\beta\int_0^t \mathrm{e}^{2\alpha s}\,\mathrm{d}s\right)^{p/2}\right. \tag{3.69}$$
$$+ \left|2\int_0^{\underline{t}}\mathrm{e}^{2\alpha(\underline{s}+h(\overline{X}_s))}\langle\phi(\overline{X}_s), g(\overline{X}_s)\,\mathrm{d}W_s\rangle\right|^{p/2} + \left(\beta\sum_{k=0}^{n_t-1}\mathrm{e}^{2\alpha t_k}\|\Delta W_k\|^2\right)^{p/2}$$
$$\left. + \left|2\mathrm{e}^{2\alpha\underline{t}}\langle\overline{X}_t + f(\overline{X}_t)\,(t-\underline{t}), g(\overline{X}_t)(W_t - W_{\underline{t}})\rangle\right|^{p/2} + \;\beta^{p/2}\mathrm{e}^{\alpha p\underline{t}}\|W_t - W_{\underline{t}}\|^p\right\}.$$

Step 2: For any $t \in [0, T]$, we take the supremum on both sides of inequality (3.69) and then take the expectation to obtain that

$$\mathbb{E}\left[\widehat{M}_t^{\alpha,p}\right] = \mathbb{E}\left[\sup_{0\le s\le t}\mathrm{e}^{\alpha p s}\|\widehat{X}_s\|^p\right] \;\le\; 6^{p/2-1}\mathrm{e}^{\alpha p h_{\max}}\,(I_1 + I_2 + I_3 + I_4 + I_5)\,, \tag{3.70}$$

where

$$\begin{aligned}
I_1 &= \|x_0\|^p + \left(2\,\beta \int_0^t \mathrm{e}^{2\alpha s}\,\mathrm{d}s\right)^{p/2}, \\
I_2 &= \mathbb{E}\left[\sup_{0\le s\le \underline{t}} \left| 2\int_0^s \mathrm{e}^{2\alpha(\underline{u}+h(\overline{X}_u))} \langle \phi(\overline{X}_u), g(\overline{X}_u)\,\mathrm{d}W_u\rangle \right|^{p/2}\right], \\
I_3 &= \mathbb{E}\left[\left(\beta \sum_{k=0}^{n_t-1} \mathrm{e}^{2\alpha t_k} \|\Delta W_k\|^2\right)^{p/2}\right], \qquad (3.71)\\
I_4 &= \mathbb{E}\left[\sup_{0\le s\le t} \left|2\mathrm{e}^{2\alpha \underline{s}} \langle \overline{X}_s + f(\overline{X}_s)\,(s-\underline{s}), g(\overline{X}_s)(W_s - W_{\underline{s}})\rangle\right|^{p/2}\right], \\
I_5 &= \mathbb{E}\left[\sup_{0\le s\le t} \beta^{p/2} \mathrm{e}^{\alpha p \underline{s}} \|W_s - W_{\underline{s}}\|^p\right].
\end{aligned}$$

We now consider I_1, I_2, I_3, I_4, I_5 in turn. It holds that

$$I_1 = \|x_0\|^p + (2\beta)^{p/2} \left(\frac{\mathrm{e}^{2\alpha t}-1}{2\alpha}\right)^{p/2} \le \|x_0\|^p + (\beta/\alpha)^{p/2} \mathrm{e}^{\alpha p t}. \qquad (3.72)$$

By the Burkholder-Davis-Gundy inequality, there exist a constant $C_p^1 \in (0,\infty)$ such that

$$\begin{aligned}
I_2 &= \mathbb{E}\left[\sup_{0\le s\le \underline{t}} \left| 2\int_0^s \mathrm{e}^{2\alpha(\underline{u}+h(\overline{X}_u))} \langle \phi(\overline{X}_u), g(\overline{X}_u)\,\mathrm{d}W_u\rangle \right|^{p/2}\right] \\
&\le \mathbb{E}\left[C_p^1 \left(\int_0^{\underline{t}} \mathrm{e}^{4\alpha u} \|\phi(\overline{X}_u)^T g(\overline{X}_u)\|^2\,\mathrm{d}u\right)^{p/4}\right]. \qquad (3.73)
\end{aligned}$$

Due to condition (3.4), for $u<\underline{t}$ we obtain that

$$\begin{aligned}
\|\phi(\overline{X}_u)\|^2 &= \|\overline{X}_u\|^2 + 2\,h(\overline{X}_u)\,\left(\langle \overline{X}_u, f(\overline{X}_u)\rangle + \tfrac{1}{2}\,h(\overline{X}_u)\|f(\overline{X}_u)\|^2\right) \\
&\le \|\overline{X}_u\|^2 + 2\,h(\overline{X}_u)\,(-\alpha\|\overline{X}_u\|^2 + \beta) \\
&\le \|\overline{X}_u\|^2 + 2\,\beta h_{\max}, \qquad (3.74)
\end{aligned}$$

and hence by Jensen's inequality and the boundedness condition (3.2) of g, we obtain that

$$\|\phi(\overline{X}_u)^T g(\overline{X}_u)\|^{p/2} \le 2^{p/4-1} \beta^{p/4} \left(\|\overline{X}_u\|^{p/2} + (2\,\beta h_{\max})^{p/4}\right). \qquad (3.75)$$

Therefore, using Jensen's inequality (3.59) with $\gamma = 2\alpha$, followed by (3.62) with $\gamma = (1+4/p)\alpha$ and then (3.60) with $\xi = \mathrm{e}^{-\alpha p t/2}\zeta$, there exists a constant $C_p^2 \in (0,\infty)$ which is linearly dependent on ζ^{-1} such that

$$\begin{aligned} I_2 &\leq \mathbb{E}\left[C_p^1(\mathrm{e}^{2\alpha t}/(2\alpha))^{p/4-1}\int_0^t \mathrm{e}^{\alpha(p/2+2)u}\|\phi(\overline{X}_u)^T g(\overline{X}_u)\|^{p/2}\,\mathrm{d}u\right] \\ &\leq \mathbb{E}\left[C_p^1(\mathrm{e}^{2\alpha t}/\alpha)^{p/4-1}\beta^{p/4}\int_0^t \mathrm{e}^{\alpha(p/2+2)u}\,\left(\|\overline{X}_u\|^{p/2} + (2\,\beta h_{\max})^{p/4}\right)\mathrm{d}u\right] \\ &\leq \mathbb{E}\left[\frac{C_p^1}{2}\left(\frac{\beta}{\alpha}\right)^{p/4}\mathrm{e}^{\alpha p(t+h_{\max})/2}\widehat{M}_t^{\alpha,p/2}\right] + C_p^1\beta^{p/2}\left(\frac{2h_{\max}}{\alpha}\right)^{p/4}\frac{2\mathrm{e}^{\alpha p t}}{p+4} \\ &\leq \frac{C_p^1}{2}\left(\frac{\beta}{\alpha}\right)^{p/4}\mathrm{e}^{\alpha p h_{\max}/2}\,\zeta\,\mathbb{E}\left[\widehat{M}_t^{\alpha,p}\right] + C_p^2\,\mathrm{e}^{\alpha p t}. \end{aligned} \tag{3.76}$$

Using Jensen's inequality (3.58), we obtain that

$$\begin{aligned} I_3 &\leq \beta^{p/2}\left(\int_0^t \mathrm{e}^{2\alpha s}\mathrm{d}s\right)^{p/2-1}\mathbb{E}\left[\sum_{k=0}^{n_t-1} h_k\,\mathrm{e}^{2\alpha t_k}\,\frac{\|\Delta W_k\|^p}{h_k^{p/2}}\right] \\ &\leq c_p\left(\beta\int_0^t \mathrm{e}^{2\alpha s}\mathrm{d}s\right)^{p/2} \leq c_p(\beta/2\alpha)^{p/2}\mathrm{e}^{\alpha p t}, \end{aligned} \tag{3.77}$$

where $c_p = \mathbb{E}[\sup_{0\leq t\leq 1}\|W_t - W_0\|^p] < \infty$. so that $\mathbb{E}[\|\Delta W_k\|^p] \leq c_p h_k^{p/2}$.

In considering I_4, we start by observing that for $s \in [t_k, t_{k+1})$ it holds that

$$\mathbb{E}\left[\sup_{t_k\leq u\leq s}\|(W_u - W_{t_k})\|^p \mid \mathcal{F}_{\underline{s}}\right] = c_p\,(s-\underline{s})^{p/2} \leq c_p {h_{\max}}^{p/2-1}(s-\underline{s}). \tag{3.78}$$

In addition, using (3.66) and following the same argument as for I_2 ensure that

$$\|\overline{X}_s + f(\overline{X}_s)(s-\underline{s})\|^{p/2}\|g(\overline{X}_s)\|^{p/2} \leq 2^{p/4-1}\beta^{p/4}\left(\|\overline{X}_s\|^{p/2} + (2\,\beta h_{\max})^{p/4}\right). \tag{3.79}$$

Therefore, combining the estimates (3.78), (3.62) with $\gamma = 2\alpha$ and (3.60) with $\xi = \mathrm{e}^{-\alpha p t/2}\zeta$, there exists a $C_p^3 \in (0,\infty)$ which is linearly dependent on ζ^{-1} such that

$$\begin{aligned}
I_4 &\le 2^{p/2}\,\mathbb{E}\left[\sup_{0\le s\le t} \mathrm{e}^{\alpha p\underline{s}}\left|\langle \overline{X}_s + f(\overline{X}_s)(s-\underline{s}), g(\overline{X}_s)\,(W_s - W_{\underline{s}})\rangle\right|^{p/2}\right] \\
&\le 2^{p/2}\,\mathbb{E}\left[\sup_{0\le s\le t} \mathrm{e}^{\alpha p\underline{s}}\|\overline{X}_s + f(\overline{X}_s)(s-\underline{s})\|^{p/2}\|g(\overline{X}_s)\|^{p/2}\|(W_s - W_{\underline{s}})\|^{p/2}\right] \\
&\le 2^{3p/4-1}\beta^{p/4}\,\mathbb{E}\left[\sum_{k=0}^{n_t-1}\mathrm{e}^{\alpha p t_k}\left(\|\overline{X}_{t_k}\|^{p/2} + (2\beta h_{\max})^{p/4}\right)\sup_{t_k\le s<t_{k+1}}\|W_s - W_{\underline{s}}\|^{p/2}\right. \\
&\qquad\left. + \mathrm{e}^{\alpha p\underline{t}}\left(\|\overline{X}_{\underline{t}}\|^{p/2} + (2\beta h_{\max})^{p/4}\right)\sup_{\underline{t}\le s<t}\|W_s - W_{\underline{s}}\|^{p/2}\right] \\
&\le 2^{3p/4-1}\beta^{p/4}c_{p/2}{h_{\max}}^{p/4-1}\,\mathbb{E}\left[\int_0^t \mathrm{e}^{\alpha p s}\left(\|\overline{X}_s\|^{p/2} + (2\beta h_{\max})^{p/4}\right)\mathrm{d}s\right] \\
&\le 2^{3p/4}\beta^{p/4}c_{p/2}{h_{\max}}^{p/4-1}\,(p\alpha)^{-1}\mathrm{e}^{\alpha p h_{\max}/2}\,\zeta\,\mathbb{E}\left[\widehat{M}_t^{\alpha,p}\right] + C_p^3\mathrm{e}^{\alpha p t}. \qquad (3.80)
\end{aligned}$$

Similarly, again using the same definition for c_p, we obtain that

$$I_5 \le c_p\,\beta^{p/2}{h_{\max}}^{p/2-1}\mathrm{e}^{\alpha p t}/(\alpha p). \qquad (3.81)$$

Collecting together the bounds for I_1, I_2, I_3, I_4, I_5, we conclude that we can choose $\zeta \in (0,\infty)$ sufficiently small so that there exist constants $C_p^4, C_p^5 \in (0,\infty)$ such that

$$\mathbb{E}\left[\widehat{M}_t^{\alpha,p}\right] \le \tfrac{1}{2}\,\mathbb{E}\left[\widehat{M}_t^{\alpha,p}\right] + C_p^4\|x_0\|^p + C_p^5\,\mathrm{e}^{\alpha p t}, \qquad (3.82)$$

and hence

$$\mathbb{E}\left[\widehat{M}_t^{\alpha,p}\right] \le 2C_p^4\,\|x_0\|^p + 2C_p^5\,\mathrm{e}^{\alpha p t}. \qquad (3.83)$$

Step 3: Due to the definition of $\overline{M}_t^{\alpha,p}$ and inequality (3.61), for any $t \in [0,\infty)$ it holds that

$$\begin{aligned}
\mathbb{E}\left[\|\overline{X}_t\|^p\right] \;\le\; \mathrm{e}^{-\alpha p t}\,\mathbb{E}\left[\overline{M}_t^{\alpha,p}\right] &\;\le\; \mathrm{e}^{-\alpha p t}\,\mathrm{e}^{\alpha p h_{\max}}\,\mathbb{E}\left[\widehat{M}_t^{\alpha,p}\right] \\
&\;\le\; \mathrm{e}^{\alpha p h_{\max}}(2\,C_p^4\|x_0\|^p + 2C_p^5) \;\triangleq\; C_p, \qquad (3.84)
\end{aligned}$$

and similarly

$$\mathbb{E}\left[\|\widehat{X}_t\|^p\right] \;\le\; \mathrm{e}^{-\alpha p t}\,\mathbb{E}\left[\widehat{M}_t^{\alpha,p}\right] \;\le\; 2\,C_p^4\,\|x_0\|^p + 2C_p^5 \;<\; C_p. \qquad (3.85)$$

This completes the proof of Theorem 3.1.1. □

3.5.3 Theorem 3.1.2

Proof of Theorem 3.1.2. The proof is for $p \in [2, p^*]$; the result for $p \in (0, 2)$ follows from Hölder's inequality.

We start by combining the original SDE (1.1) with (2.3) that to obtain

$$e_t \; = \; \int_0^t \big(f(\overline{X}_s) - f(X_s)\big) \, \mathrm{d}s + \int_0^t \big(g(\overline{X}_s) - g(X_s)\big) \, \mathrm{d}W_s, \tag{3.86}$$

and then by Itô's formula and Young's inequality (3.56), together with $e_0 = 0$, and $\lambda, \ \eta$ as defined in Assumption 3.1.3, we obtain that

$$\begin{aligned}
\mathrm{e}^{\lambda pt/2}\|e_t\|^p \; \le \; & \int_0^t \frac{p\lambda}{2} \mathrm{e}^{\lambda ps/2}\|e_s\|^p \, \mathrm{d}s + \int_0^t p\langle e_s, f(\overline{X}_s) - f(X_s)\rangle \mathrm{e}^{\lambda ps/2}\|e_s\|^{p-2} \, \mathrm{d}s \\
& + \int_0^t \frac{p(p-1)}{2}\|g(\overline{X}_s) - g(X_s)\|^2 \mathrm{e}^{\lambda ps/2}\|e_s\|^{p-2} \, \mathrm{d}s \\
& + \int_0^t p \, \langle e_s, (g(\overline{X}_s) - g(X_s))\mathrm{e}^{\lambda ps/2} \, \|e_s\|^{p-2} \, \mathrm{d}W_s\rangle \\
\le \; & \int_0^t \frac{p\lambda}{2} \mathrm{e}^{\lambda ps/2}\|e_s\|^p \, \mathrm{d}s + \int_0^t p\langle e_s, f(\widehat{X}_s) - f(X_s)\rangle \mathrm{e}^{\lambda ps/2} \, \|e_s\|^{p-2} \, \mathrm{d}s \\
& - \int_0^t p\langle e_s, f(\widehat{X}_s) - f(\overline{X}_s)\rangle \mathrm{e}^{\lambda ps/2} \, \|e_s\|^{p-2} \, \mathrm{d}s \\
& + p \int_0^t \left(\frac{p-1}{2} + \frac{\lambda}{4\eta}\right) \|g(\widehat{X}_s) - g(X_s)\|^2 \mathrm{e}^{\lambda ps/2} \, \|e_s\|^{p-2} \mathrm{d}s \\
& + p \int_0^t \left(\frac{p-1}{2} + \frac{\eta(p-1)^2}{\lambda}\right) \|g(\widehat{X}_s) - g(\overline{X}_s)\|^2 \mathrm{e}^{\lambda ps/2} \, \|e_s\|^{p-2} \mathrm{d}s \\
& + \int_0^t p \, \langle e_s, (g(\overline{X}_s) - g(X_s))\mathrm{e}^{\lambda ps/2} \, \|e_s\|^{p-2} \, \mathrm{d}W_s\rangle
\end{aligned} \tag{3.87}$$

Using the conditions in Assumption 3.1.3, (3.10) implies that

$$\|g(\widehat{X}_s) - g(\overline{X}_s)\|^2 \le \eta \|\widehat{X}_s - \overline{X}_s\|^2. \tag{3.88}$$

(3.9) and (3.10) imply that

$$\langle e_s, f(\widehat{X}_s)-f(X_s)\rangle + \left(\frac{p-1}{2}+\frac{\lambda}{4\eta}\right)\|g(\widehat{X}_s)-g(X_s)\|^2 \le -\frac{3\lambda}{4}\|e_s\|^2. \quad (3.89)$$

(2.19) and Young's inequality (3.56) imply that

$$\begin{aligned}\left|\langle e_s, f(\widehat{X}_s)-f(\overline{X}_s)\rangle\right| &\le \|e_s\|\, L(\widehat{X}_s,\overline{X}_s)\,\|\widehat{X}_s-\overline{X}_s\| \\ &\le \frac{\lambda}{8}\|e_s\|^2+\frac{2}{\lambda}L(\widehat{X}_s,\overline{X}_s)^2\|\widehat{X}_s-\overline{X}_s\|^2, \end{aligned}\quad (3.90)$$

where $L(x,y)\triangleq\gamma(\|x\|^q+\|y\|^q)+\mu$. Hence, it holds that

$$\begin{aligned}\mathrm{e}^{\lambda pt/2}\|e_t\|^p \le & \int_0^t -\frac{p\lambda}{8}\,\mathrm{e}^{\lambda ps/2}\|e_s\|^p\,\mathrm{d}s+\int_0^t p\,\hat{L}(\widehat{X}_s,\overline{X}_s)\|\widehat{X}_s-\overline{X}_s\|^2\mathrm{e}^{\lambda ps/2}\|e_s\|^{p-2}\,\mathrm{d}s \\ &+\int_0^t p\langle e_s,(g(\overline{X}_s)-g(X_s))\mathrm{e}^{\lambda ps/2}\|e_s\|^{p-2}\,\mathrm{d}W_s\rangle, \end{aligned}\quad (3.91)$$

where $\hat{L}(x,y)=\frac{2}{\lambda}L(x,y)^2+\frac{(p-1)\eta}{2}+\frac{\eta^2(p-1)^2}{\lambda}$. Young's inequality (3.57) implies that

$$\begin{aligned}\mathrm{e}^{\lambda pt/2}\|e_t\|^p \le & \int_0^t 2\left(\frac{8(p-2)}{p\lambda}\right)^{p/2-1}\hat{L}(\widehat{X}_s,\overline{X}_s)^{p/2}\mathrm{e}^{\lambda ps/2}\|\widehat{X}_s-\overline{X}_s\|^p\,\mathrm{d}s \\ &+\int_0^t p\langle e_s,(g(\overline{X}_s)-g(X_s))\mathrm{e}^{\lambda pt/2}\|e_s\|^{p-2}\,\mathrm{d}W_s\rangle. \end{aligned}\quad (3.92)$$

Taking the expectation of each side ensures that

$$\mathbb{E}\left[\mathrm{e}^{\lambda pt/2}\|e_t\|^p\right]\le 2\left(\frac{8(p-2)}{p\lambda}\right)^{p/2-1}\int_0^t\mathbb{E}\left[\hat{L}(\widehat{X}_s,\overline{X}_s)^{p/2}\|\widehat{X}_s-\overline{X}_s\|^p\right]\mathrm{e}^{\lambda ps/2}\mathrm{d}s. \quad (3.93)$$

By Hölder's inequality, it holds that

$$\mathbb{E}\left[\hat{L}(\widehat{X}_s,\overline{X}_s)^{p/2}\|\widehat{X}_s-\overline{X}_s\|^p\right]\le\left(\mathbb{E}\left[\hat{L}(\widehat{X}_s,\overline{X}_s)^p\right]\mathbb{E}\left[\|\widehat{X}_s-\overline{X}_s\|^{2p}\right]\right)^{1/2}, \quad (3.94)$$

and $\mathbb{E}\left[\hat{L}(\widehat{X}_s,\overline{X}_s)^p\right]$ can be bounded by a constant C_p^1 due to the stability property in Theorem 3.1.1. For any $s\in[0,T]$, $\widehat{X}_s-\overline{X}_s=f(\widehat{X}_{\underline{s}})(s-\underline{s})+g(\widehat{X}_{\underline{s}})(W_s-W_{\underline{s}})$, and hence, by a combination of Jensen's and Hölder's inequalities, we obtain that

$$\begin{aligned}\mathbb{E}\left[\|\widehat{X}_s-\overline{X}_s\|^{2p}\right] \le\ & 2^{2p-1}\left(\mathbb{E}\left[\|f(\widehat{X}_{\underline{s}})\|^{4p}\right]\mathbb{E}\left[(s-\underline{s})^{4p}\right]\right)^{1/2} \\ +\ & 2^{2p-1}\left(\mathbb{E}\left[\|g(\widehat{X}_{\underline{s}})\|^{4p}\right]\mathbb{E}\left[\|W_s-W_{\underline{s}}\|^{4p}\right]\right)^{1/2}. \end{aligned}\quad (3.95)$$

$\mathbb{E}[\|f(\widehat{X}_{\underline{s}})\|^{4p}]$ and $\mathbb{E}[\|g(\widehat{X}_{\underline{s}})\|^{4p}]$ are both finite, due to stability and the polynomial bounds on the growth of $f(x)$ and $g(x)$. Furthermore, we have $\mathbb{E}[(s-\underline{s})^{4p}] \leq (\delta h_{\max})^{4p} \leq {h_{\max}}^{4p}\delta^{2p}$, and by standard results there is a constant c_p such that $\mathbb{E}[\|W_s - W_{\underline{s}}\|^{4p}] = \mathbb{E}[\ \mathbb{E}[\|W_s - W_{\underline{s}}\|^{4p} \,|\, \mathcal{F}_{\underline{s}}]\ ] \leq c_p(\delta h_{\max})^{2p}$. Hence, there exists a constant $C_p^2 \in (0,\infty)$ such that $\mathbb{E}[\,\|\widehat{X}_s - \overline{X}_s\|^{2p}] \leq C_p^2\,\delta^p$, and therefore equation (3.93) implies that

$$\mathbb{E}\left[\mathrm{e}^{\lambda p t/2}\|e_t\|^p\right] \leq 2\left(\frac{8(p-2)}{p\lambda}\right)^{p/2-1}\int_0^t \sqrt{C_p^1 C_p^2}\ \delta^{p/2}\mathrm{e}^{\lambda p s/2}\,\mathrm{d}s, \tag{3.96}$$

which provides the final result: there exists a constant $C_p \in (0,\infty)$ such that for all $t \in [0,\infty)$ it holds that

$$\mathbb{E}\left[\|e_t\|^p\right] \leq \frac{4}{\lambda p}\left(\frac{8(p-2)}{p\lambda}\right)^{p/2-1}\sqrt{C_p^1 C_p^2}\ \delta^{p/2} = C_p\,\delta^{p/2}. \tag{3.97}$$

This completes the proof of Theorem 3.1.2. □

3.5.4 Theorem 3.1.3

Proof of Theorem 3.1.3. The proof is given for $p \in [4,\infty)$; the result for $p \in (0,4)$ follows from Hölder's inequality. The error $e_t \triangleq \widehat{X}_t - X_t$ satisfies the SDE $\mathrm{d}e_t = \big(f(\overline{X}_t) - f(X_t)\big)\,\mathrm{d}t$ and hence by Itô's formula, it holds that

$$\begin{aligned} \mathrm{e}^{2\lambda t}\|e_t\|^2 &= \int_0^t 2\lambda\,\mathrm{e}^{2\lambda s}\|e_s\|^2\,\mathrm{d}s + \int_0^t 2\,\mathrm{e}^{2\lambda s}\langle e_s, f(\widehat{X}_s) - f(X_s)\rangle\,\mathrm{d}s \\ &\quad - \int_0^t 2\,\mathrm{e}^{2\lambda s}\langle e_s, f(\widehat{X}_s) - f(\overline{X}_s)\rangle\,\mathrm{d}s \\ &\leq -\int_0^t 2\,\mathrm{e}^{2\lambda s}\langle e_s, f(\widehat{X}_s) - f(\overline{X}_s)\rangle\,\mathrm{d}s \end{aligned} \tag{3.98}$$

due to the one-sided Lipschitz condition (3.18), so therefore we obtain that

$$\mathbb{E}\left[\sup_{0\leq s\leq t}\mathrm{e}^{\lambda p s}\|e_s\|^p\right] \leq 2^{p/2}\,\mathbb{E}\left[\sup_{0\leq s\leq t}\left|\int_0^s \mathrm{e}^{2\lambda u}\langle e_u, f(\widehat{X}_u) - f(\overline{X}_u)\rangle\,\mathrm{d}u\right|^{p/2}\right]. \tag{3.99}$$

Within a single timestep, it holds that $\widehat{X}_u - \overline{X}_u = f(\overline{X}_u)(u - \underline{u}) + (W_u - W_{\underline{u}})$, and therefore, following a similar approach to the proof of Theorem 2.1.4, Lemma C.1 implies that

$$\begin{aligned} e^{2\lambda u}\langle e_u, f(\widehat{X}_u) - f(\overline{X}_u)\rangle &= e^{2\lambda u}\langle e_u, \nabla f(\overline{X}_u)(\widehat{X}_u - \overline{X}_u)\rangle \; + \; e^{2\lambda u} R_u \\ &= e^{2\lambda u}\langle e_u, (u-\underline{u})\nabla f(\overline{X}_u) f(\overline{X}_u)\rangle \; + \; e^{2\lambda u} R_u \\ &\quad + (e^{2\lambda u} - e^{2\lambda \underline{u}})\langle e_u, \nabla f(\overline{X}_u)(W_u - W_{\underline{u}})\rangle \\ &\quad + e^{2\lambda \underline{u}}\langle (e_u - e_{\underline{u}}), \nabla f(\overline{X}_u)(W_u - W_{\underline{u}})\rangle \\ &\quad + e^{2\lambda \underline{u}}\langle e_{\underline{u}}, \nabla f(\overline{X}_u)(W_u - W_{\underline{u}})\rangle \end{aligned} \tag{3.100}$$

where $|R_u| \le \Big(\gamma\left(\|\widehat{X}_u\|^q + \|\overline{X}_u\|^q\right) + \mu\Big) \,\|e_u\|\,\|\widehat{X}_u - \overline{X}_u\|^2$, and hence it holds that

$$\mathbb{E}\left[\sup_{0\le s\le t} e^{\lambda p s}\|e_s\|^p\right] \le \frac{10^{p/2}}{5}(I_1 + I_2 + I_3 + I_4 + I_5), \tag{3.101}$$

where

$$\begin{aligned} I_1 &= \mathbb{E}\left[\sup_{0\le s\le t}\left|\int_0^s e^{2\lambda u}\langle e_u, (u-\underline{u})\nabla f(\overline{X}_u) f(\overline{X}_u)\rangle \,\mathrm{d}u\right|^{p/2}\right], \\ I_2 &= \mathbb{E}\left[\sup_{0\le s\le t}\left|\int_0^s e^{2\lambda u} R_u \;\mathrm{d}u\right|^{p/2}\right], \\ I_3 &= \mathbb{E}\left[\sup_{0\le s\le t}\left|\int_0^s (e^{2\lambda u} - e^{2\lambda \underline{u}})\langle e_u, \nabla f(\overline{X}_u)(W_u - W_{\underline{u}})\rangle \,\mathrm{d}u\right|^{p/2}\right], \\ I_4 &= \mathbb{E}\left[\sup_{0\le s\le t}\left|\int_0^s e^{2\lambda \underline{u}}\langle (e_u - e_{\underline{u}}), \nabla f(\overline{X}_u)(W_u - W_{\underline{u}})\rangle \,\mathrm{d}u\right|^{p/2}\right], \\ I_5 &= \mathbb{E}\left[\sup_{0\le s\le t}\left|\int_0^s e^{2\lambda \underline{u}}\langle e_{\underline{u}}, \nabla f(\overline{X}_u)(W_u - W_{\underline{u}})\rangle \,\mathrm{d}u\right|^{p/2}\right]. \end{aligned} \tag{3.102}$$

We now bound I_1, I_2, I_3, I_4, I_5 in turn. Noting that $u - \underline{u} \le \delta h_{\max}$, by Young's inequality (3.56) and Jensen's inequality (3.59), we obtain that

$$\begin{aligned} I_1 &\leq \frac{\mathrm{e}^{\lambda(p/2-1)t}}{\lambda^{p/2-1}} \mathbb{E}\left[\int_0^t \mathrm{e}^{\lambda pu/2}\left(\delta h_{\max}\|e_u\|\|f(\overline{X}_u)\|\|\nabla f(\overline{X}_u)\|\right)^{p/2} \mathrm{e}^{\lambda u} \mathrm{d}u\right] \\ &\leq \frac{\mathrm{e}^{\lambda(p/2-1)t}}{\lambda^{p/2-1}} (\delta h_{\max})^{p/2} \mathbb{E}\left[\left(\sup_{0\leq s\leq t} \mathrm{e}^{\lambda ps/2}\|e_s\|^{p/2}\right)\int_0^t \|f(\overline{X}_u)\|^{p/2}\|\nabla f(\overline{X}_u)\|^{p/2} \mathrm{e}^{\lambda u} \mathrm{d}u\right] \\ &\leq \xi\, \mathbb{E}\left[\sup_{0\leq s\leq t} \mathrm{e}^{\lambda ps}\|e_s\|^p\right] + \frac{\mathrm{e}^{\lambda(p-1)t}}{4\xi\, \lambda^{p-1}} (\delta h_{\max})^p \int_0^t \mathbb{E}\left[\|f(\overline{X}_u)\|^p\|\nabla f(\overline{X}_u)\|^p\right] \mathrm{e}^{\lambda u}\, \mathrm{d}u. \end{aligned} \tag{3.103}$$

The last integral is finite because of stability and the polynomial bounds on the growth of both f and ∇f, and hence there is a constant $C_p^1 \in (0, \infty)$ such that

$$I_1 \leq \xi\, \mathbb{E}\left[\sup_{0\leq s\leq t} \mathrm{e}^{\lambda ps}\|e_s\|^p\right] + (C_p^1/\xi)\, \mathrm{e}^{\lambda pt}\, \delta^p. \tag{3.104}$$

Similarly, using Young's inequality (3.56), Jensen's inequality (3.59) and the Hölder inequality, we obtain that

$$\begin{aligned} I_2 &\leq \frac{\mathrm{e}^{\lambda(p/2-1)t}}{\lambda^{p/2-1}} \int_0^t \mathbb{E}\left[\mathrm{e}^{\lambda pu/2}\|e_u\|^{p/2} L^{p/2}(\widehat{X}_u, \overline{X}_u)\|\widehat{X}_u - \overline{X}_u\|^p\right] \mathrm{e}^{\lambda u} \mathrm{d}u \\ &\leq \xi \mathbb{E}\left[\sup_{0\leq s\leq t} \mathrm{e}^{\lambda ps}\|e_s\|^p\right] + \frac{\mathrm{e}^{\lambda(p-1)t}}{4\xi\, \lambda^{p-1}} \int_0^t \left(\mathbb{E}\left[L^{2p}(\widehat{X}_u, \overline{X}_u)\right] \mathbb{E}\left[\|\widehat{X}_u - \overline{X}_u\|^{4p}\right]\right)^{1/2} \mathrm{e}^{\lambda u} \mathrm{d}u, \end{aligned} \tag{3.105}$$

where $L(\widehat{X}_u, \overline{X}_u) \equiv \gamma\,(\|\widehat{X}_u\|^q + \|\overline{X}_u\|^q) + \mu$. Hence, using stability and bounds on $\mathbb{E}\left[\|\widehat{X}_u - \overline{X}_u\|^{4p}\right]$ from the proof of Theorem 3.1.2, there is a constant $C_p^2 \in (0, \infty)$ such that

$$I_2 \leq \xi\, \mathbb{E}\left[\sup_{0\leq s\leq t} \mathrm{e}^{\lambda ps}\|e_s\|^p\right] + (C_p^2/\xi)\, \mathrm{e}^{\lambda pt}\, \delta^p. \tag{3.106}$$

The fact that $\mathrm{e}^{2\lambda u} - \mathrm{e}^{2\lambda \underline{u}} \leq 2\lambda\, \mathrm{e}^{2\lambda u}(u - \underline{u})$, and Jensen's inequality (3.59) imply that

$$I_3 \leq \frac{\mathrm{e}^{\lambda(p/2-1)t}}{\lambda^{p/2-1}} \int_0^t \mathbb{E}\left[\mathrm{e}^{\lambda pu/2}\left(2\lambda \delta h_{\max}\|e_u\|\|\nabla f(\overline{X}_u)\|\, \|W_u - W_{\underline{u}}\|\right)^{p/2}\right] \mathrm{e}^{\lambda u} \mathrm{d}u, \tag{3.107}$$

and using the approach to estimating I_1, there is similarly a constant $C_p^3 \in (0, \infty)$ such that

$$I_3 \leq \xi\, \mathbb{E}\left[\sup_{0\leq s\leq t} \mathrm{e}^{\lambda ps}\|e_s\|^p\right] + (C_p^3/\xi)\, \mathrm{e}^{\lambda pt}\, \delta^p. \tag{3.108}$$

For the next term, I_4, similarly to the estimation of the bound of $\|e_s - e_{\underline{s}}\|$ in the proof of Theorem 2.1.4, by Theorem 3.1.2, we find that there is a constant $c_p \in (0,\infty)$ such that

$$\mathbb{E}[\|e_s - e_{\underline{s}}\|^p] \leq c_p\, \delta^{3p/2}. \tag{3.109}$$

Now, by Jensen's inequality (3.59), we obtain that

$$I_4 \leq \frac{\mathrm{e}^{\lambda(p-2)t}}{(2\lambda)^{p/2-1}} \int_0^t \mathbb{E}\left[\|e_s - e_{\underline{s}}\|^{p/2} \|\nabla f(\overline{X}_s)\|^{p/2} \|W_s - W_{\underline{s}}\|^{p/2}\right] \mathrm{e}^{2\lambda s} \mathrm{d}s, \tag{3.110}$$

so using Hölder's inequality and the usual stability bounds, we conclude that there is a constant $C_p^4 \in (0,\infty)$ such that

$$I_4 \leq C_p^4\, \mathrm{e}^{\lambda p t}\, \delta^p. \tag{3.111}$$

Lastly, considering that

$$\mathrm{d}\left((t - t_{n+1})(W_t - W_{t_n})\right) = (W_t - W_{t_n})\, \mathrm{d}t + (t - t_{n+1})\, \mathrm{d}W_t, \tag{3.112}$$

in the time interval $[t_n, t_{n+1}]$ conditioned on $\mathcal{F}_{t_n}$, we obtain that

$$\int_{t_n}^{t_{n+1}} (W_u - W_{t_n})\, \mathrm{d}u = -\int_{t_n}^{t_{n+1}} (u - t_{n+1})\, \mathrm{d}W_u, \tag{3.113}$$

and therefore we can split I_5 into two parts

$$\begin{aligned} I_5 &\leq 2^{p/2-1} \left\{ \mathbb{E}\left[\sup_{0\leq s\leq t} \left| \sum_{k=0}^{n_s-1} \int_{t_k}^{t_{k+1}} \mathrm{e}^{2\lambda \underline{u}} (u - t_{k+1}) \langle e_{t_k}, \nabla f(\overline{X}_{t_k}) \mathrm{d}W_u \rangle \right|^{p/2}\right] \right. \\ &\qquad \left. + \mathbb{E}\left[\sup_{0\leq s\leq t} \left| \int_{\underline{s}}^{s} \mathrm{e}^{2\lambda \underline{u}} \langle e_{\underline{u}}, \nabla f(\overline{X}_{\underline{u}})(W_u - W_{\underline{u}}) \rangle\, \mathrm{d}u \right|^{p/2}\right] \right\} \\ &= 2^{p/2-1}\, (I_{51} + I_{52}). \end{aligned} \tag{3.114}$$

By the Burkholder-Davis-Gundy inequality, Young's inequality (3.56) and Jensen's inequality (3.59), there exists a constant $c_p^1 \in (0,\infty)$ such that

$$\begin{aligned}
I_{51} \;\leq\;& \mathbb{E}\left[\sup_{0\leq s\leq t}\left|\int_0^s \mathrm{e}^{2\lambda \underline{u}}(u-\underline{u}-h(\overline{X}_{\underline{u}}))\langle e_{\underline{u}}, \nabla f(\overline{X}_{\underline{u}})\mathrm{d}W_u\rangle\right|^{p/2}\right]\\
\leq\;& \mathbb{E}\left[c_p^1\left(\int_0^t \mathrm{e}^{4\lambda\underline{u}}(\delta h_{\max})^2\|e_{\underline{u}}\|^2\|\nabla f(\overline{X}_{\underline{u}})\|^2\mathrm{d}u\right)^{p/4}\right] \qquad (3.115)\\
\leq\;& \mathbb{E}\left[c_p^1\sup_{0\leq s\leq t}\mathrm{e}^{\lambda ps/2}\|e_s\|^{p/2}\left(\int_0^t(\delta h_{\max})^2\|\nabla f(\overline{X}_{\underline{u}})\|^2\mathrm{e}^{2\lambda u}\mathrm{d}u\right)^{p/4}\right]\\
\leq\;& \xi\mathbb{E}\left[\sup_{0\leq s\leq t}\mathrm{e}^{\lambda ps}\|e_s\|^p\right]+\frac{1}{4\xi}\mathbb{E}\left[\left(c_p^1\right)^2\frac{\mathrm{e}^{\lambda(p-2)t}}{(2\lambda)^{p/2-1}}\int_0^t(\delta h_{\max})^p\|\nabla f(\overline{X}_{\underline{u}})\|^p\mathrm{e}^{2\lambda u}\mathrm{d}u\right].
\end{aligned}$$

Hence, there exists a constant $C_p^{51} \in (0,\infty)$ such that

$$I_{51} \leq \xi\,\mathbb{E}\left[\sup_{0\leq s\leq t}\mathrm{e}^{\lambda ps}\|e_s\|^p\right]+(C_p^{51}/\xi)\,\mathrm{e}^{\lambda pt}\,\delta^p. \qquad (3.116)$$

Finally, for I_{52}, the Young's inequality (3.56) and Jensen's inequality (3.59) imply

$$\begin{aligned}
I_{52} \;=\;& \mathbb{E}\left[\sup_{0\leq s\leq t}\left|\int_{\underline{s}}^s\langle e_{\underline{u}}, \nabla f(\overline{X}_u)(W_u-W_{\underline{u}})\rangle\mathrm{e}^{2\lambda\underline{u}}\,\mathrm{d}u\right|^{p/2}\right] \qquad (3.117)\\
\leq\;& (\delta h_{\max})^{p/2-1}\mathbb{E}\left[\sup_{0\leq s\leq t}\int_{\underline{s}}^s\|e_{\underline{u}}\|^{p/2}\|\nabla f(\overline{X}_u)\|^{p/2}\|W_u-W_{\underline{u}}\|^{p/2}\mathrm{e}^{\lambda p\underline{u}}\,\mathrm{d}u\right]\\
\leq\;& (\delta h_{\max})^{p/2-1}\mathbb{E}\left[\left(\sup_{0\leq s\leq t}\mathrm{e}^{\lambda ps/2}\|e_s\|^{p/2}\right)\int_0^t\|\nabla f(\overline{X}_u)\|^{p/2}\|W_u-W_{\underline{u}}\|^{p/2}\mathrm{e}^{\lambda pu/2}\,\mathrm{d}u\right]\\
\leq\;& \xi\,\mathbb{E}\left[\sup_{0\leq s\leq t}\mathrm{e}^{\lambda ps}\|e_s\|^p\right]+\frac{\mathrm{e}^{\lambda pt/2}}{2\xi\,\lambda p}(\delta h_{\max})^{3p/2-2}\int_0^t\mathbb{E}\left[\|\nabla f(\overline{X}_u)\|^p\right]\mathrm{e}^{\lambda pu/2}\mathrm{d}u.
\end{aligned}$$

Thus, we find that there exists a constant $C_p^{52} > 0$ such that

$$I_{52} \leq \xi\,\mathbb{E}\left[\sup_{0\leq s\leq t}\mathrm{e}^{\lambda ps}\|e_s\|^p\right]+(C_p^{52}/\xi)\,\mathrm{e}^{\lambda pt}\,\delta^p. \qquad (3.118)$$

Combining the five bounds, and choosing ξ to be sufficiently small, we conclude that there is a constant $C_p \in (0,\infty)$ such that

$$\mathbb{E}\left[\sup_{0\leq s\leq t}\mathrm{e}^{\lambda ps}\|e_s\|^p\right] \leq C_p\,\mathrm{e}^{\lambda pt}\,\delta^p. \qquad (3.119)$$

and therefore for all $t \in [0,\infty)$ it holds that

$$\mathbb{E}\left[\|e_t\|^p\right] \leq C_p\,\delta^p. \qquad (3.120)$$

This completes the proof of Theorem 3.1.3. □

3.6 Conclusion

In this chapter, we considered the adaptive approximations of the ergodic SDEs in an infinite interval. The uniform bound with respect to T are established for the moments, as shown in Theorem 3.1.1, and strong error of the numerical solutions when the contractivity condition is satisfied, as shown in Theorem 3.1.2. In addition, we extended this adaptive scheme to the MLMC method for the ergodic limits by using different time intervals on different levels, as shown in Algorithm 1. By shifting the time intervals we constructed an efficient coupling of the fine path and the coarse path with different simulation times. The numerical results contained in section 3.3 support our analysis. Last but not least, we extended without proof our scheme to a larger class of SDEs with contractivity on average, for example, the FENE model and SDEs with a double-well potential, in section 3.4. Our adaptive scheme still works well and achieves an optimal complexity.

Chapter 4

SDEs without contractivity

This chapter is an extended version of the published paper [25]. In this chapter, we construct a new MLMC scheme for ergodic SDEs without contractivity by using change of measure techniques, as we outlined in the introduction. However, due to loss of contractivity, adaptively choosing T_ℓ does not work and we simply simulate the SDE for a sufficiently long time T. The exponential convergence to the invariant measure [75] is given by

$$|\mathbb{E}\left[\varphi(X_T) - \pi(\varphi)\right]| \leq \mu^* \, \mathrm{e}^{-\lambda^* T} \tag{4.1}$$

for some constant $\mu^*, \lambda^* \in (0, \infty)$, and bounding this truncation error by ε requires

$$T \, \geq \, \frac{1}{\lambda^*} \log(\varepsilon^{-1}) + \frac{\log \mu^*}{\lambda^*}. \tag{4.2}$$

Change of measure techniques have been used in previous research to reduce the variance of corrections in MLMC. Giles proposed to use the same Gaussian samples for the final step of both fine and coarse paths with a change of measure for the pricing of the digital option on page 38 in [32]. To cope with SDEs with path-dependent jumps, Xia & Giles [106] used a change of measure so that the acceptance probability of the jumps is the same for both fine and coarse paths. Kebaier & Lelong [54] optimize over a class of measures to optimally reduce the variance of

MLMC corrections. Andersson & Kohatsu-Higa [2] change the sampling distribution to make the MLMC correction variance finite for unbiased simulation of SDEs using parametrix expansions. Stilger & Poon [92] apply it for MLMC calculation of an interest rate model and Gasparotto [29] for deep out-of-money options to reduce the variance.

The change of measure technique together with the Lamperti transform is also the core part in the exact simulation of SDEs, see [8] and the references therein. Importance sampling (change of measure technique) has also been widely used in rare event simulations, see [19, 52] for a good introduction and review and the references therein.

Lastly, the construction of good coupling between paths is also useful for theoretical results. Eberle et al. [20] proposed a new coupling method to estimate the theoretical convergence rate for Langevin dynamics. See [16] and its references for further exploration.

The rest of this chapter is organised as follows. Section 1 introduces the new MLMC method with the change of measure. Section 2 states the main theorems, and the relevant numerical experiments are provided in section 3. Numerical results for SDEs with non-globally Lipschitz drift are given in section 4. We apply this technique to the numerical method for some elliptic PDEs in section 5. The proofs of the main theorems are deferred to section 6. Finally, section 7 concludes this chapter.

4.1 New MLMC with change of measure

In this chapter, we use the standard Euler-Maruyama method to simulate the original SDE (1.1) using N uniform timesteps under the measure $\mathbb{P}$, that is

$$t_{n+1} = t_n + h, \quad \widehat{X}_{t_{n+1}} = \widehat{X}_{t_n} + f(\widehat{X}_{t_n})h + \Delta W_n^{\mathbb{P}}, \tag{4.3}$$

where $\Delta W_n^{\mathbb{P}} \triangleq W_{t_{n+1}}^{\mathbb{P}} - W_{t_n}^{\mathbb{P}}$ for $n \in \{0, 1, ..., N-1\}$ with $h = T/N$ and there is fixed initial data $t_0 = 0$, $\widehat{X}_{t_0} = x_0$. Then, the standard Monte Carlo estimator for

$\mathbb{E}^{\mathbb{P}}[\varphi(X_T)]$ is the mean of the values $\varphi(\widehat{X}_T^L)$, from N_L independent path simulations using $h = 2^{-L}h_0$ for some suitable constant $h_0 \in (0, \infty)$ and positive integer L is

$$\widehat{\varphi}_{std} \coloneqq N_L^{-1} \sum_{n=1}^{N_L} \varphi(\widehat{X}_T^{L,(n)}). \tag{4.4}$$

For standard MLMC, instead of directly estimating $\mathbb{E}^{\mathbb{P}}\left[\varphi(\widehat{X}_T^L)\right]$, we have the following telescoping sum in the same probability measure $\mathbb{P}$:

$$\mathbb{E}^{\mathbb{P}}\left[\varphi(\widehat{X}_T^L)\right] = \mathbb{E}^{\mathbb{P}}\left[\varphi(\widehat{X}_T^0)\right] + \sum_{\ell=1}^{L} \mathbb{E}^{\mathbb{P}}\left[\varphi(\widehat{X}_T^{f,\ell}) - \varphi(\widehat{X}_T^{c,\ell-1})\right], \tag{4.5}$$

where $\widehat{X}_T^{f,\ell}$ and $\widehat{X}_T^{c,\ell-1}$ share the same driving Brownian motion. Then, the standard MLMC estimator becomes

$$\widehat{\varphi}_{mlmc} \coloneqq N_0^{-1} \sum_{n=1}^{N_0} \varphi(\widehat{X}_T^{0,(n)}) + \sum_{\ell=1}^{L} N_\ell^{-1} \sum_{n=1}^{N_\ell} \left(\varphi(\widehat{X}_T^{f,\ell,(n)}) - \varphi(\widehat{X}_T^{c,\ell-1,(n)})\right). \tag{4.6}$$

Now we introduce the new MLMC scheme with change of measure using a spring coefficient $S \in (0, \infty)$.

For level 0, the numerical estimator is the same as the standard MLMC $\varphi(\widehat{X}_T^0)$.

For level $\ell \in \{1, 2, ..., L\}$, we simulate the SDE with the additional spring terms using the timestep $h = 2^{-\ell}\, h_0$ for the fine path and $2h$ for the coarse path.

- At t_0, we set $\widehat{Y}_{t_0}^f = \widehat{Y}_{t_0}^c = x_0$.
- At odd timesteps $t_{2n+1} = t_{2n} + h$ for $n \in \{0, 1, 2, ..., N/2{-}1\}$, we update both paths

$$\begin{aligned} \widehat{Y}_{t_{2n+1}}^c &= \widehat{Y}_{t_{2n}}^c + S(\widehat{Y}_{t_{2n}}^f - \widehat{Y}_{t_{2n}}^c)h + f(\widehat{Y}_{t_{2n}}^c)h + \Delta W_{2n}^{\mathbb{P}}, \\ \widehat{Y}_{t_{2n+1}}^f &= \widehat{Y}_{t_{2n}}^f + S(\widehat{Y}_{t_{2n}}^c - \widehat{Y}_{t_{2n}}^f)h + f(\widehat{Y}_{t_{2n}}^f)h + \Delta W_{2n}^{\mathbb{P}}. \end{aligned} \tag{4.7}$$

- At even timesteps $t_{2n+2} = t_{2n+1} + h$ for $n \in \{0, 1, 2, ..., N/2-1\}$, we update the spring term and drift term of the fine path, but keep both the same for the coarse path

$$\begin{aligned} \widehat{Y}_{t_{2n+2}}^c &= \widehat{Y}_{t_{2n+1}}^c + S(\widehat{Y}_{t_{2n}}^f - \widehat{Y}_{t_{2n}}^c)h + f(\widehat{Y}_{t_{2n}}^c)h + \Delta W_{2n+1}^{\mathbb{P}}, \\ \widehat{Y}_{t_{2n+2}}^f &= \widehat{Y}_{t_{2n+1}}^f + S(\widehat{Y}_{t_{2n+1}}^c - \widehat{Y}_{t_{2n+1}}^f)h + f(\widehat{Y}_{t_{2n+1}}^f)h + \Delta W_{2n+1}^{\mathbb{P}}. \end{aligned} \tag{4.8}$$

Note that the coarse path updates can be combined to obtain that

$$\widehat{Y}^c_{t_{2n+2}} = \widehat{Y}^c_{t_{2n}} + S(\widehat{Y}^f_{t_{2n}} - \widehat{Y}^c_{t_{2n}})2h + f(\widehat{Y}^c_{t_{2n}})2h + \Delta W^{\mathbb{P}}_{2n} + \Delta W^{\mathbb{P}}_{2n+1}. \tag{4.9}$$

Next, we derive the exact Radon-Nikodym derivatives for both fine and coarse paths. To begin with, suppose we only apply the change of measure to the nth timestep. Under the measure $\mathbb{P}$, we obtain that

$$\widehat{Y}_{t_{n+1}} = \widehat{X}_{t_n} + \widehat{S}h + f(\widehat{X}_{t_n})h + \Delta W^{\mathbb{P}}_n \Rightarrow \widehat{Y}_{t_{n+1}} \sim N^{\mathbb{P}}(\widehat{X}_{t_n} + f(\widehat{X}_{t_n})h + \widehat{S}h, hI), \tag{4.10}$$

where I is the identity matrix, $N^{\mathbb{P}}(\mu, \Sigma)$ is a normal distribution under the measure $\mathbb{P}$ and $\widehat{S}$ is the spring term. Under a new measure $\widehat{\mathbb{Q}}_n$ with $\Delta W^{\widehat{\mathbb{Q}}_n}_n = \widehat{S}h + \Delta W^{\mathbb{P}}_n$, we get

$$\widehat{Y}_{t_{n+1}} = \widehat{Y}_{t_n} + f(\widehat{Y}_{t_n})h + \Delta W^{\widehat{\mathbb{Q}}_n}_n \Rightarrow \widehat{Y}_{t_{n+1}} \sim N^{\widehat{\mathbb{Q}}_n}(\widehat{Y}_{t_n} + f(\widehat{Y}_{t_n})h, hI). \tag{4.11}$$

Then the exact Radon-Nikodym derivative for this single step is that

$$\frac{\mathrm{d}\widehat{\mathbb{Q}}_n}{\mathrm{d}\mathbb{P}} = \frac{\rho(\widehat{Y}_{t_{n+1}} | \widehat{Y}_{t_n} + f(\widehat{Y}_{t_n})h, hI)}{\rho(\widehat{Y}_{t_{n+1}} | \widehat{Y}_{t_n} + f(\widehat{Y}_{t_n})h + \widehat{S}h, hI)} := \mathrm{R}(\widehat{Y}_{t_{n+1}}, \widehat{Y}_{t_n}, \widehat{S}, h), \tag{4.12}$$

where $\rho(x|\mu, \Sigma)$ is the probability density function of $N(\mu, \Sigma)$ and

$$\begin{aligned} \mathrm{R}(\widehat{Y}_{t_{n+1}}, \widehat{Y}_{t_n}, \widehat{S}, h) &= \exp\left(-\left\langle \widehat{Y}_{t_{n+1}} - \widehat{Y}_{t_n} - f(\widehat{Y}_{t_n})h, \widehat{S}\right\rangle + \|\widehat{S}\|^2 h/2\right) \\ &= \exp\left(-\left\langle \Delta W^{\mathbb{P}}_n,\ \widehat{S}\right\rangle - \|\widehat{S}\|^2 h/2\right). \end{aligned} \tag{4.13}$$

Now, suppose that we introduce such changes on each timestep of the whole path, so under a new measure $\widehat{\mathbb{Q}}$, we have $\Delta W^{\widehat{\mathbb{Q}}}_n = \widehat{S}h + \Delta W^{\mathbb{P}}_n$, for all $n \in \{0, 1, 2, ..., N-1\}$. Since $\Delta W^{\mathbb{P}}_n$ and $\Delta W^{\widehat{\mathbb{Q}}}_n$, for all $n \in \{0, 1, 2, ..., N-1\}$, are sets of independent Brownian increments under the measure $\mathbb{P}$ and $\widehat{\mathbb{Q}}$ respectively, the exact Radon-Nikodym derivative becomes

$$\frac{\mathrm{d}\widehat{\mathbb{Q}}}{\mathrm{d}\mathbb{P}} := \prod_{n=0}^{N-1} \mathrm{R}(\widehat{Y}_{t_{n+1}}, \widehat{Y}_{t_n}, \widehat{S}, h). \tag{4.14}$$

Numerically we obtain two new measures $\widehat{\mathbb{Q}}^f$ and $\widehat{\mathbb{Q}}^c$ with $\Delta W^{\widehat{\mathbb{Q}}^f}_n = \widehat{S}^f_n h + \Delta W^{\mathbb{P}}_n$ and $\Delta W^{\widehat{\mathbb{Q}}^c}_n = \widehat{S}^c_n h + \Delta W^{\mathbb{P}}_n$ respectively for all steps on the fine and coarse paths, where $\widehat{S}^f_n$ and $\widehat{S}^c_n$ are the spring terms on nth step for fine and coarse paths. Then we can calculate the exact Radon-Nikodym derivatives step by step at the same time as updating the paths.

- At t_0, we set $\mathrm{R}^f_{t_0} = \mathrm{R}^c_{t_0} = 1$.
- At odd timesteps $t_{2n+1} = t_{2n} + h$ for $n \in \{0, 1, 2, ..., N/2-1\}$, we only update R^f

$$\mathrm{R}^f_{t_{2n+1}} \quad = \quad \mathrm{R}^f_{t_{2n}}\, \mathrm{R}\left(\widehat{Y}^f_{t_{2n+1}},\ \widehat{Y}^f_{t_{2n}},\ S(\widehat{Y}^c_{t_{2n}} - \widehat{Y}^f_{t_{2n}}),\ h\right). \tag{4.15}$$

- At even timesteps $t_{2n+2} = t_{2n+1} + h$ for $n \in \{0, 1, 2, ..., N/2-1\}$, we update both R^f and R^c

$$\begin{aligned} \mathrm{R}^f_{t_{2n+2}} \quad &= \quad \mathrm{R}^f_{t_{2n+1}} \mathrm{R}\left(\widehat{Y}^f_{t_{2n+2}}, \widehat{Y}^f_{t_{2n+1}}, S(\widehat{Y}^c_{t_{2n+1}} - \widehat{Y}^f_{t_{2n+1}}), h\right), \\ \mathrm{R}^c_{t_{2n+2}} \quad &= \quad \mathrm{R}^c_{t_{2n}}\ \ \mathrm{R}\left(\widehat{Y}^c_{t_{2n+2}},\ \widehat{Y}^c_{t_{2n}},\ S(\widehat{Y}^f_{t_{2n}} - \widehat{Y}^c_{t_{2n}}),\ 2h\right). \end{aligned} \tag{4.16}$$

Then, after N steps, we obtain the exact Radon-Nikodym derivatives for the whole path

$$\begin{aligned} \frac{\mathrm{d}\widehat{\mathbb{Q}}^f}{\mathrm{d}\mathbb{P}} \quad &= \quad \mathrm{R}^f_T = \prod_{n=0}^{N-1}\ \mathrm{R}\left(\widehat{Y}^f_{t_{n+1}},\ \widehat{Y}^f_{t_n},\ S(\widehat{Y}^c_{t_n} - \widehat{Y}^f_{t_n}),\ h\right), \\ \frac{\mathrm{d}\widehat{\mathbb{Q}}^c}{\mathrm{d}\mathbb{P}} \quad &= \quad \mathrm{R}^c_T = \prod_{n=0}^{N/2-1}\ \mathrm{R}\left(\widehat{Y}^c_{t_{2n+2}},\ \widehat{Y}^c_{t_{2n}},\ S(\widehat{Y}^f_{t_{2n}} - \widehat{Y}^c_{t_{2n}}),\ 2h\right). \end{aligned} \tag{4.17}$$

Finally, the multilevel correction estimator becomes

$$\varphi(\widehat{Y}^f_T)\, \mathrm{R}^f_T - \varphi(\widehat{Y}^c_T)\, \mathrm{R}^c_T, \tag{4.18}$$

and the identity we use in the new MLMC is that

$$\mathbb{E}^{\mathbb{P}}\left[\varphi(\widehat{X}^L_T)\right] = \mathbb{E}^{\mathbb{P}}\left[\varphi(\widehat{X}^0_T)\right] + \sum_{\ell=1}^{L} \mathbb{E}^{\mathbb{P}}\left[\varphi(\widehat{Y}^{f,\ell}_T)\mathrm{R}^{f,\ell}_T - \varphi(\widehat{Y}^{c,\ell-1}_T \mathrm{R}^{c,\ell}_T)\right], \tag{4.19}$$

where $\mathrm{R}^{f,\ell}_T$ and $\mathrm{R}^{c,\ell}_T$ are the exact Radon-Nikodym derivatives for the fine and coarse paths on level ℓ. The new MLMC estimator becomes

$$\begin{aligned} \widehat{\varphi}_{new} \quad &:= \quad N_0^{-1} \sum_{n=1}^{N_0} \varphi(\widehat{X}^{0,(n)}_T) \\ &\quad + \sum_{\ell=1}^{L} N_\ell^{-1} \sum_{n=1}^{N_\ell} \left(\varphi(\widehat{X}^{f,\ell,(n)}_T)\mathrm{R}^{f,\ell,(n)}_T - \varphi(\widehat{X}^{c,\ell-1,(n)}_T)\mathrm{R}^{c,\ell,(n)}_T\right). \end{aligned} \tag{4.20}$$

In the following sections, we only work under measure $\mathbb{P}$, so we use W_t to denote $W^{\mathbb{P}}_t$ for simplicity.

4.2 Theoretical results

In this section, we state the key results on the stability and strong error of the path after the change of measure, and then the variance of the estimator (4.18) and the resulting MLMC complexity.

Assumption 4.2.1 (Lipschitz continuity and dissipativity)**.** *Assume f is globally Lipschitz so that there is a constant $K \in (0,\infty)$ such that for any $x, y \in \mathbb{R}^m$ it holds that*

$$\|f(x)-f(y)\| \leq K\,\|x-y\|. \tag{4.21}$$

Furthermore, assume that there exist constants $\tilde{\alpha}, \tilde{\beta} \in (0,\infty)$ such that for all $x \in \mathbb{R}^m$, f satisfies the dissipativity condition

$$\langle x, f(x)\rangle \leq -\tilde{\alpha}\|x\|^2 + \tilde{\beta}. \tag{4.22}$$

Note that a consequence of the globally Lipschitz condition is that

$$\|f(x)\| \leq \|f(0)\| + K\|x\| \;\Rightarrow\; \|f(x)\|^2 \leq 2\left(\|f(0)\|^2 + K^2\|x\|^2\right). \tag{4.23}$$

This assumption ensures the existence and uniqueness of the strong solution to the SDEs [67] and the convergence to the invariant distribution [73]. Note that the globally Lipschitz assumption is needed for simplicity of the proof but numerical experiments in section 4.4 show that the change of measure technique also works well for SDEs with non-globally Lipschitz drift. The following theorem, based on this assumption, shows that our numerical scheme with sufficiently small h is stable and the moments of the numerical solution are uniformly bounded with respect to T.

Theorem 4.2.1 (Stability)**.** *If the original SDE satisfies Assumption 4.2.1 using the new change-of-measure algorithm with $S \in (0,\infty)$, then for any $T \in (0,\infty)$, $p \in [1,\infty)$, there exist constants $C_{(1)}, C_{(2)} \in (0,\infty)$ independent of T and p such that for all $h \in (0, C_{(1)})$, it holds that*

$$\sup_{n\in\{0,1,2,\ldots,N\}} \mathbb{E}\left[\|\widehat{Y}^f_{t_n}\|^p\right]^{1/p} \leq C_{(2)}\, p^{1/2}, \qquad \sup_{n\in\{0,1,2,\ldots,N\}} \mathbb{E}\left[\|\widehat{Y}^c_{t_n}\|^p\right]^{1/p} \leq C_{(2)}\, p^{1/2}. \tag{4.24}$$

Proof of Theorem 4.2.1. The proof is deferred to section 4.6.1. □

It is important to note that the constants $C_{(1)}, C_{(2)}$ depend on the specifics of the original SDE and the value of S, but not on T, h or the moment power p. This result is expected since the spring term is only a linear function of the numerical solution and the magnitude is small which does not destroy the dissipativity condition and allows us to obtain the uniform bounds. For the first-order strong convergence, we need the following assumption.

Assumption 4.2.2 (One-sided Lipschitz property). *There exists a constant $\lambda \in (0,\infty)$ such that for all $x, y \in \mathbb{R}$, f satisfies the one-sided Lipschitz condition*

$$\langle x-y, f(x)-f(y)\rangle \leq \lambda \, \|x-y\|^2, \tag{4.25}$$

and f is differentiable and $\nabla f(x)$ satisfies the globally Lipschitz condition (4.21)*.*

Note that the globally Lipschitz condition (4.21) implies this one-sided Lipschitz condition (4.25). However, the one-sided Lipschitz condition can give a sharper bound for the positive side, which means that K can be much larger than λ. The spring term in our algorithm is only needed when the inner product $\langle x-y, f(x)-f(y)\rangle$ is positive, to prevent the exponential divergence of the fine and coarse paths. See the adaptive spring for double-well potential energy SDE in section 4.4 where we choose S to be a function of the current state to minimize the spring term and thereby reduce the size of the Radon-Nikodym derivative. The other consideration is that possibly we can extend this scheme to SDEs with locally one-sided Lipschitz drift, for example the stochastic Lorenz equation. Therefore, this condition helps us to obtain an accurate choice of the spring term S as shown in the following theorem.

Theorem 4.2.2 (Difference between fine and coarse paths). *If the original SDE satisfies Assumptions 4.2.1 and 4.2.2 using the new change-of-measure algorithm with $S > \lambda/2$, then for any $T \in (0,\infty)$ and $p \in [1,\infty)$, there exist constants $C_{(1)}, C_{(2)} \in (0,\infty)$ independent of T and p such that for all $h \in (0, C_{(1)})$ it holds* *that*

$$\sup_{n\in\{0,1,2,\dots,N\}} \mathbb{E}\left[\|\widehat{Y}^f_{t_n} - \widehat{Y}^c_{t_n}\|^p\right]^{1/p} \le C_{(2)} \min\left(p^{1/2}\, h^{1/2},\ p\, h\right). \tag{4.26}$$

Proof of Theorem 4.2.2. The proof is deferred to section 4.6.2. □

The L_p norm of the difference between the fine and coarse paths, as we expected, is uniformly bounded since we add a sufficiently strong spring term to recover the contractivity. With this result, we can bound the pth-moment of the Radon-Nikodym derivatives and then the MLMC estimator (4.18).

Theorem 4.2.3 (Radon-Nikodym moments)**.** *If the original SDE satisfies Assumptions 4.2.1 and 4.2.2 using the new change-of-measure algorithm with $S > \lambda/2$, then for any $T \in (0,\infty)$ and $p \in [1,\infty)$, there exist constants $C_{(1)}, C_{(2)} \in (0,\infty)$ independent of T and p such that for all $h \in (0, \min(C_{(1)}, C_{(2)}/(Tp^2)))$ it holds that*

$$\mathbb{E}\left[\left|\frac{\mathrm{d}\widehat{\mathbb{Q}}^c}{\mathrm{d}\mathbb{P}}\right|^p\right] \le 2, \quad \mathbb{E}\left[\left|\frac{\mathrm{d}\widehat{\mathbb{Q}}^f}{\mathrm{d}\mathbb{P}}\right|^p\right] \le 2. \tag{4.27}$$

Proof of Theorem 4.2.3. The proof is deferred to section 4.6.3. □

Theorem 4.2.4 (MLMC moments)**.** *If the original SDE satisfies Assumptions 4.2.1 and 4.2.2, and $\varphi : \mathbb{R}^m \to \mathbb{R}$ is globally Lipschitz using the new change-of-measure algorithm with $S > \lambda/2$,, then for any $T \in (0,\infty)$ and $p \in [1,\infty)$ there exist constants $C_{(1)}, C_{(2)}, C_{(3)} \in (0,\infty)$ such that for all $h \in (0, \min(C_{(1)}, C_{(2)}/(Tp^2))$ it holds that*

$$\mathbb{E}\left[\left|\varphi(\widehat{Y}^f_T)\,\frac{\mathrm{d}\widehat{\mathbb{Q}}^f}{\mathrm{d}\mathbb{P}} - \varphi(\widehat{Y}^c_T)\,\frac{\mathrm{d}\widehat{\mathbb{Q}}^c}{\mathrm{d}\mathbb{P}}\right|^p\right]^{1/p} \le C_{(3)}\, p^2 \sqrt{T}\, h. \tag{4.28}$$

Proof of Theorem 4.2.4. The proof is deferred to section 4.6.4. □

Note that this theorem implies that the variance of the estimator (4.18) is bounded by $C_2 T h^2$, which increases linearly in T.

We now have everything we require to determine the MLMC complexity.

Theorem 4.2.5 (MLMC for invariant measure)**.** *If φ satisfies the globally Lipschitz condition and the SDE satisfies Assumption 4.2.1 and 4.2.2 with convergence rate λ^* and constant μ^* in* (4.1)*, and Lyapunov exponent κ, then by choosing suitable values for L, S, h_0 and N_ℓ for each level ℓ, there exist constants c_1, c_2, $c_3 \in (0,\infty)$ independent of T and p such that the estimator $\widehat{\varphi}$ has a mean square error (MSE) satisfying the bound*

$$\mathbb{E}\left[(\widehat{\varphi} - \pi(\varphi))^2\right] \leq \varepsilon^2,$$

with $\varepsilon \in (0,1)$ and an expected computational cost C_{std} for the standard Monte Carlo estimator $\widehat{\varphi}_{std}$ (4.4) *satisfying the bound*

$$\mathrm{C}_{std} \leq \; c_1\, \varepsilon^{-3} |\log \varepsilon|, \tag{4.29}$$

and an expected computational cost C_{mlmc} for the standard MLMC estimator $\widehat{\varphi}_{mlmc}$ (4.6) *satisfying the bound*

$$\mathrm{C}_{mlmc} \leq \; c_2\, \varepsilon^{-2-\frac{\kappa}{2\lambda^*}} |\log \varepsilon|, \tag{4.30}$$

provided $\kappa/\lambda^* < 2$, and C_{com} for the new MLMC estimator with change of measure $\widehat{\varphi}_{new}$ (4.20) *satisfying the bound*

$$\mathrm{C}_{com} \leq \; c_3\, \varepsilon^{-2} |\log \varepsilon|^2. \tag{4.31}$$

Proof of Theorem 4.2.5. By Jensen's inequality, the MSE can be decomposed into three parts:

$$\begin{aligned} \mathbb{E}\left[(\widehat{\varphi} - \pi(\varphi))^2\right] \;&=\; \mathbb{V}\left[\widehat{\varphi}\right] + \left|\mathbb{E}\left[\widehat{\varphi}\right] - \pi(\varphi)\right|^2 \qquad (4.32)\\ &\leq\; \mathbb{V}\left[\widehat{\varphi}\right] + 2\left|\mathbb{E}\left[\widehat{\varphi}\right] - \mathbb{E}\left[\varphi(X_T)\right]\right|^2 + 2\left|\mathbb{E}\left[\varphi(X_T)\right] - \pi(\varphi)\right|^2, \end{aligned}$$

which enables us to achieve the MSE bound by bounding each part by $\varepsilon^2/3$. Due to the exponential convergence to the invariant measure (4.1), we bound the third part by setting

$$T \;=\; \frac{1}{\lambda^*}\log(\varepsilon^{-1}) + \frac{\log\sqrt{6}\mu^*}{\lambda^*}, \tag{4.33}$$

to bound the truncation error. The first order weak convergence requires $h_L = O(\varepsilon)$ and $L \geq \lceil \gamma \log_2(\varepsilon^{-1}) + \zeta \rceil$ for some $\gamma, \zeta \in (0, \infty)$.

For the standard Monte Carlo method using h_L, the computational cost for each path is $O(\varepsilon^{-1}|\log \varepsilon|)$ and the bound on variance requires $O(\varepsilon^{-2})$ samples, which gives a total computational cost

$$\mathrm{C}_{std} \leq \ c_1 \, \varepsilon^{-3} |\log \varepsilon|, \tag{4.34}$$

for some constant $c_1 \in (0, \infty)$.

The analysis for the two MLMC schemes is similar to the MLMC theorem in [32] and shows that the optimal computational cost is bounded by

$$3\varepsilon^{-2} \left(\sum_{\ell=0}^{L} \sqrt{V_\ell \, \mathrm{C}_\ell} \right)^2 \ + \ \sum_{\ell=0}^{L} \mathrm{C}_\ell \ , \tag{4.35}$$

where C_ℓ and V_ℓ are the cost and variance for each level.

For standard MLMC, we have first order weak convergence but the variance of V_ℓ for $\ell \in \{1, 2, ..., L\}$ increases exponentially in T, which implies that

$$V_\ell \leq \ \eta_1 \, (h_0 \, 2^{-\ell})^2 \, \mathrm{e}^{\kappa T}, \tag{4.36}$$

for some constant $\eta_1 \in (0, \infty)$. A good MLMC coupling requires $\mathrm{C}_0 V_0 > \mathrm{C}_1 V_1$, and given this condition and $\beta = 2, \ \gamma = 1$, the optimal cost is $O(\varepsilon^{-2}\mathrm{C}_0)$. The condition $\mathrm{C}_0 V_0 > \mathrm{C}_1 V_1$ requires

$$h_0 = \vartheta_1 \, \mathrm{e}^{-\kappa T/2} \quad \Rightarrow \quad \mathrm{C}_0 = \vartheta_2 \, \varepsilon^{-\frac{\kappa}{2\lambda^*}} \, |\log \varepsilon|, \tag{4.37}$$

for some $\vartheta_1, \vartheta_2 \in (0, \infty)$. The condition $\kappa/\lambda^* < 2$ ensures that h_0 is greater than the timestep required by the standard Monte Carlo method so additional MLMC levels are required to achieve the desired weak convergence. Therefore, there exists a constant $c_2 \in (0, \infty)$ such that

$$\mathrm{C}_{mlmc} \leq \ c_2 \, \varepsilon^{-2-\frac{\kappa}{2\lambda^*}} \, |\log \varepsilon|. \tag{4.38}$$

For the new MLMC with the change of measure, Theorem 4.2.4 ensures that

$$V_\ell \leq \ \eta_2\, (h_0\, 2^{-\ell})^2\, T, \tag{4.39}$$

for some $\eta_2 \in (0, \infty)$. The condition $\mathrm{C}_0 V_0 > \mathrm{C}_1 V_1$ requires

$$h_0 = \vartheta_3\, T^{-1/2}, \tag{4.40}$$

for some $\vartheta_3 \in (0, \infty)$, but the bound in Theorem 4.2.4 requires the tighter condition

$$h_0 = \vartheta_4\, T^{-1} \quad \Rightarrow \quad \mathrm{C}_0 = \vartheta_5\, |\log \varepsilon|^2, \tag{4.41}$$

for some $\vartheta_4, \vartheta_5 \in (0, \infty)$. Therefore, there exists a constant $c_3 \in (0, \infty)$ such that

$$\mathrm{C}_{com} \leq \ c_3\, \varepsilon^{-2} |\log \varepsilon|^2. \tag{4.42}$$

This completes the proof of Theorem 4.2.5. □

4.3 Numerical results

In this section, we present the numerical results for a globally Lipschitz version of the stochastic Lorenz equation with additive noise:

$$f\begin{pmatrix} x_1 \\ x_2 \\ x_3 \end{pmatrix} = \begin{pmatrix} 10(B(x_2) - x_1) \\ (28 - x_3)B(x_1) - x_2 \\ B(x_1)x_2 - \frac{8}{3}x_3 \end{pmatrix}. \tag{4.43}$$

where $B(x) = 65x/\max(65, |x|)$. When $|x_1| \in (65, \infty)$ and $|x_2| \in (65, \infty)$, we have

$$f\begin{pmatrix} x_1 \\ x_2 \\ x_3 \end{pmatrix} = \begin{pmatrix} 650\,\mathrm{sgn}(x_2) - 10x_1 \\ 65\,\mathrm{sgn}(x_1)(28 - x_3) - x_2 \\ 65\,\mathrm{sgn}(x_1)x_2 - \frac{8}{3}x_3 \end{pmatrix}. \tag{4.44}$$

Therefore, f satisfies the globally Lipschitz condition (4.21) and the dissipativity condition (4.22). In the region of $|x_1| \in [0, 65]$ and $|x_2| \in [0, 65]$, which contains the chaotic attractor, this function will retain the chaotic property of the original

Lorenz equation. Our interest is to compute $\pi(\varphi)$, where $\varphi(x) = \|x\|$ satisfying the globally Lipschitz condition. We ran 10000 sample paths from $T = 0$ to 20 to get the following results.

Figure 4.1 is a semi-log plot of the variance on each level as a function of T without the change of measure. The blue to red lines correspond to the variance on each level $\ell = 1$ to 8 with $h_\ell = 2^{-\ell}h_0$ and $h_0 = 2^{-9}$:

$$\mathbb{V}\left[\varphi(\widehat{X}_T^c) - \varphi(\widehat{X}_T^f)\right] \sim \eta_1\, h_\ell^2\, \mathrm{e}^{\kappa T}, \tag{4.45}$$

for some $\eta_1 \in (0, \infty)$, which increases exponentially with respect to T and stops increasing when it reaches the decoupling upper bound $\mathbb{V}\left[\varphi(\widehat{X}_t^c)\right] + \mathbb{V}\left[\varphi(\widehat{X}_t^f)\right]$ shown in yellow to green lines. In addition, as the level increases, the variance decreases at rate 2. For $T \in [10, 20]$, we can see that the standard MLMC on level $\ell = 8$, using $h = 2^{-17}$, still can not achieve a good coupling, that is, the variance of the level estimator is approximately the sum of the variances of the fine and course estimators. In order to see this exponential increase, we plot the log variance on

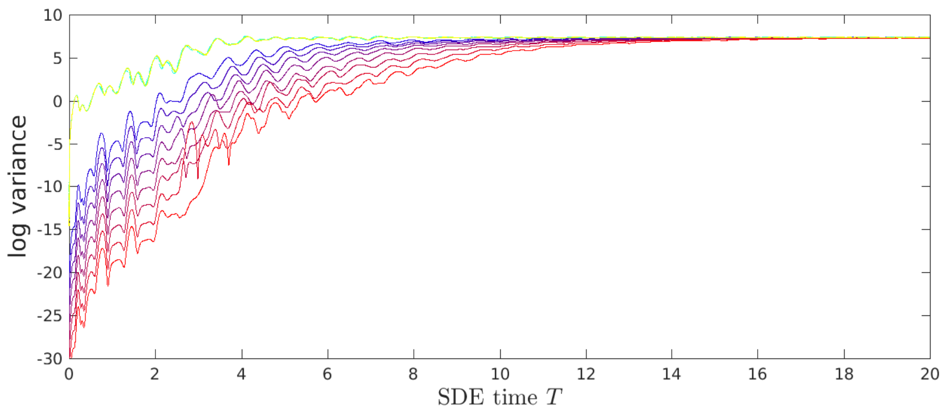
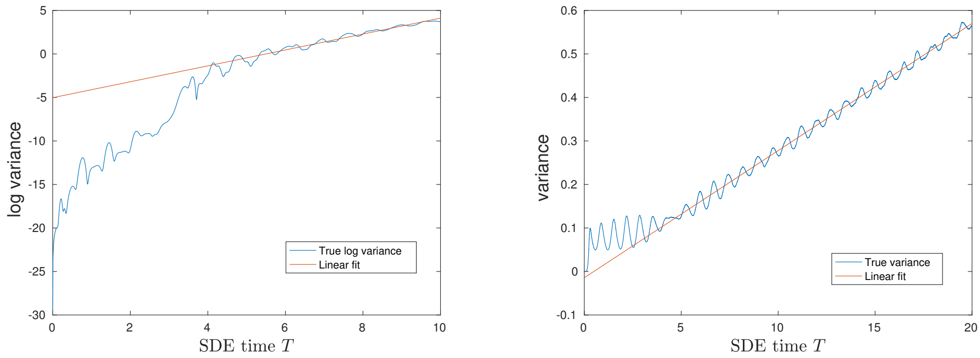


Figure 4.1: Variance for each level without change of measure

level 8 using $h = 2^{-17}$ with respect to T and the fitted linear function on time interval $[5, 10]$, see Figure 4.2(a). The κ we fit is 1.36.



(a) Linear increase of log variance without change of measure

(b) Linear increase of variance with change of measure

Figure 4.2: Variance on level 8 with/without change of measure

Similarly, for the new MLMC with spring term $S = 10$, Figure 4.3 is the semi-log plot of the variance on each level as a function of T with change of measure using same h_ℓ:

$$\mathbb{V}\left[\varphi(\widehat{Y}_T^f)\mathrm{R}_T^f - \varphi(\widehat{Y}_T^c)\mathrm{R}_T^c\right] \sim \eta_2\, h_\ell^2\, T, \tag{4.46}$$

for some $\eta_2 \in (0,\infty)$. As the level increases, the variance decreases at a rate 2. In order to see the linear increase in T, we plot the variance on level $\ell = 8$ with

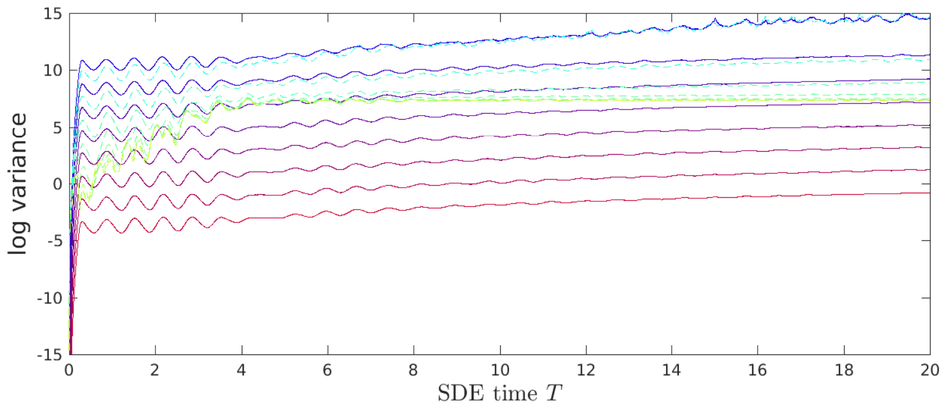


Figure 4.3: Variance for each level with change of measure

respect to T and the fitted linear function on time interval $[5, 20]$, see Figure 4.2(b).

We have investigated and illustrated the dependence of V_ℓ on T for both schemes. Next, we investigate the impact of this increase on MLMC schemes, that is the requirement of h_0 to achieve a good coupling, that is $V_0 > 2V_1$. We plot $\log h_0$ with respect to T in Figure 4.4. The blue line confirms the exponential decrease of h_0 with respect to T in (4.37). The coefficient of the log function fit is 0.49 which confirms the relationship (4.40).

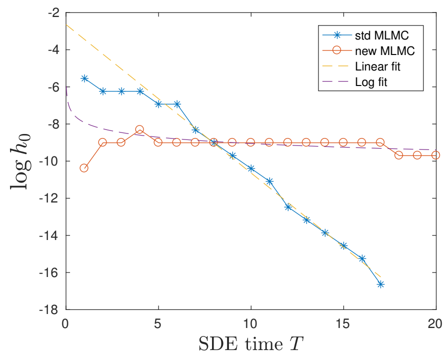
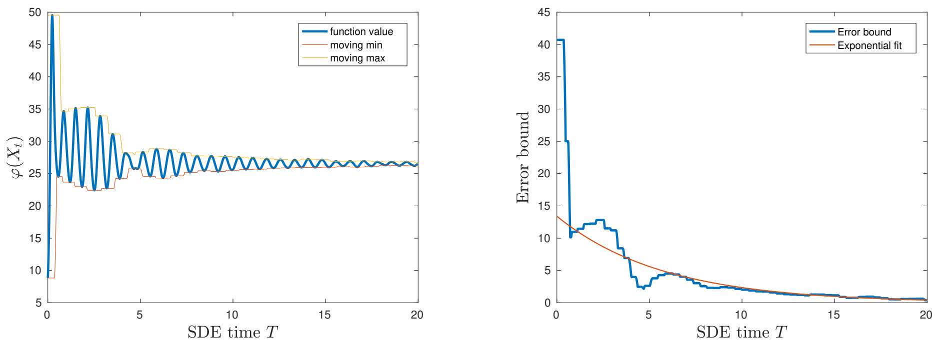


Figure 4.4: The required h_0 to achieve a good coupling

Lastly, we estimate the convergence rate λ^* to the invariant measure. Figure 4.5(a) plots the function value $\varphi(X_t)$ with respect to time t and its moving upper bound and lower bound. We plot the error bound (the difference between the moving upper bound and the moving lower bound) in Figure 4.5(b) and the exponential fit. The fitted λ^* is 0.1741. Therefore, in this case with $\lambda^* = 0.1741$ and $\kappa = 1.3601$, the standard MLMC fails to achieve any computational savings by Theorem 4.2.5.

4.4 Extension to non-Lipschitz SDEs

In this section, we extend this change of measure technique to ergodic SDEs with non-Lipschitz drift using the adaptive timestepping method proposed in subsection



(a) $\mathbb{E}\,\varphi(X_t)$

(b) Error bound

Figure 4.5: Estimation of the convergence rate to the invariant distribution

2.1.1. Without any proof, we show some numerical results for the SDE with a double-well potential energy and the stochastic Lorenz equation.

4.4.1 Double-well potential energy

We consider

$$\mathrm{d}X_t = (2X_t - \frac{1}{2}X_t^3)\,\mathrm{d}t + \mathrm{d}W_t. \tag{4.47}$$

The probability density function of its invariant distribution is

$$\frac{\exp(2x^2 - \frac{1}{4}x^4)}{\int_{-\infty}^{\infty}\exp(2x^2 - \frac{1}{4}x^4)\mathrm{d}x}, \tag{4.48}$$

and it has two different wells, at $x = \pm 2$, see Figure 4.4.1.

This SDE satisfies the dissipativity condition (4.22) and one-sided Lipschitz condition (4.25) with $\lambda = 2$ but the drift is non-globally Lipschitz. For the standard MLMC scheme, the issue is that the fine and coarse paths may diverge to different wells, which can result in a large variance and high kurtosis. Using the change of measure technique can reduce the divergence and then improve the efficiency.

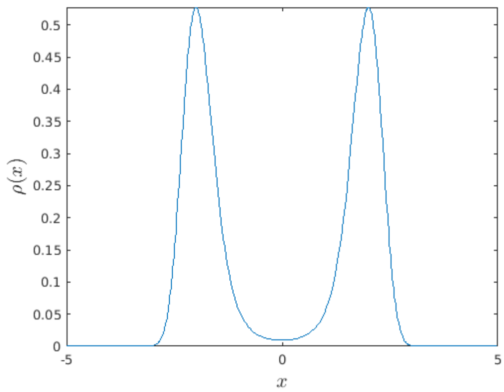


Figure 4.6: Probability density function of the invariant distribution

We simulate the SDE with initial value $x_0 = 0$ to time $T = 5$, and use the adaptive function

$$h^\delta(x) = \frac{\max(1, |x|)}{8 \max(1, |2x - \frac{1}{2}x^3|)} \, \delta, \tag{4.49}$$

with $\delta = 2^{-\ell}$ for each level ℓ. We compare three different schemes:

- standard MLMC with adaptive timestepping.
- MLMC with adaptive timestepping and change of measure with a constant spring coefficient $S = 1$.
- MLMC with adaptive timestepping and change of measure with an adaptive spring coefficient

$$S(x) = \max(0, 2 - 1.5x^2). \tag{4.50}$$

The second scheme uses $S = 1$ following the suggestion of Theorem 4.2.4. The third scheme improves on the second by choosing an adaptive S and avoiding unnecessary spring term, reducing the variance without losing the control on the divergence. By performing a first order Taylor expansion on (1.65), we choose $S(x) = \max(0, f'(x))$ to deal with the divergence locally.

We ran 10000 samples for each level ℓ for the three schemes. The numerical results are shown in Figure 4.7.

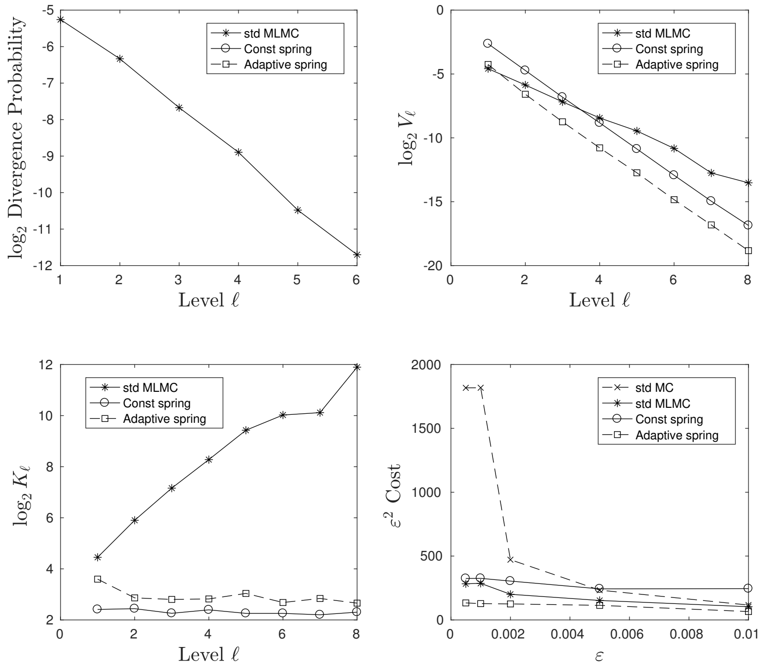


Figure 4.7: MLMC convergence test for the double-well potential energy

The top left figure plots the divergence probability with respect to the level ℓ, where the divergence probability is defined as

$$\mathbb{E}\left[\mathbb{1}_{\|\widehat{X}_T^f - \widehat{X}_T^c\|>1}\right] \;=\; \mathbb{P}\left[\|\widehat{X}_T^f - \widehat{X}_T^c\| > 1\right]. \tag{4.51}$$

The probability decreases as ℓ increases since the timestep h_ℓ is smaller and the difference between the fine and coarse path decreases. The decrease rate we fit is

$$\mathbb{P}\left[\|\widehat{X}_T^f - \widehat{X}_T^c\| > 1\right] \sim O(h_\ell^{1.28}). \tag{4.52}$$

The two schemes with change of measure have zero divergence on all levels.

The top right figure plots the variance of corrections V_ℓ with respect to level ℓ. The V_ℓ of the two schemes with change of measure decrease at the similar rate 2 while the standard MLMC has a slower rate of approximately 1.28 since the divergence of the fine and coarse paths dominated the variance. The scheme with the adaptive spring coefficient has lower V_ℓ than the scheme with constant spring coefficient since the unnecessary spring will increase the variance of the Radon-Nikodym derivative.

The bottom left figure shows the log kurtosis with respect to level ℓ. The kurtosis of standard MLMC will increase exponentially while the kurtosis of the schemes with change of measure will stay constant. A similar intuitive explanation applies here. The divergence samples again dominate the 4th moment and then the kurtosis on each level is

$$K_\ell \sim \frac{\mathbb{E}\left[\|\widehat{X}_T^f - \widehat{X}_T^c\|^4\right]}{\mathbb{E}\left[\|\widehat{X}_T^f - \widehat{X}_T^c\|^2\right]^2} \sim h_\ell^{-1.28}. \tag{4.53}$$

The rate of increase in the figure is 1.06 which is quite close to the rate of decrease of the divergence probability.

The bottom right figure plots the costs of the three schemes together with the standard Monte Carlo method with respect to ε. The costs of all the MLMC schemes are $O(\varepsilon^{-2})$ while the standard MC is $O(\varepsilon^{-3})$ and the scheme with adaptive spring has the lowest cost.

Overall, the new MLMC schemes with change of measure perform better especially the one with an adaptive spring. They can not only keep the kurtosis constant but also reduce the variance and hence the total computational cost.

4.4.2 Stochastic Lorenz equation

This is a three-dimensional system modelling convection rolls in the atmosphere: $\mathrm{d}x_t = f(x_t)\,\mathrm{d}t + \mathrm{d}W_t$, with

$$f\begin{pmatrix} x_1 \\ x_2 \\ x_3 \end{pmatrix} = \begin{pmatrix} 10(x_2 - x_1) \\ x_1(28 - x_3) - x_2 \\ x_1 x_2 - \frac{8}{3}x_3 \end{pmatrix}. \tag{4.54}$$

This SDE does not satisfy the dissipativity condition (4.22) and one-sided Lipschitz condition (4.25), and is more chaotic compared with the truncated Lipschitz version in previous section.

We simulate the SDE with initial value $x_0 = 0$ to time $T = 10$, and use the adaptive function:

$$h^\delta(x) = \frac{\max(100, \|x\|^2)}{2^{11}\max(100, \|f(x)\|^2)}\delta, \tag{4.55}$$

with $\delta = 2^{-\ell}$ for each level ℓ. We compare two different schemes:

- standard MLMC with adaptive timestepping.
- MLMC with adaptive timestepping and change of measure with a constant spring coefficient $S = 10$.

A possible third scheme is the scheme with an adaptive spring, which requires us to calculate the largest positive eigenvalue of the Jacobian matrix $\frac{\partial f}{\partial x}$.

We ran 10000 samples for each level ℓ for the two schemes. The numerical results are shown in the Figure 4.4.2.

The top left figure shows that the change of measure technique can greatly reduce the ratio of divergence $\mathbb{E}\left[\mathbb{1}_{\|\widehat{X}_T^f - \widehat{X}_T^c\| > 10}\right]$ and actually no divergence occurs in this numerical experiment at any of the levels. The rate of decrease for standard MLMC is 0.82.

The top right figure illustrates the variance reduction of the change of measure technique, and the rate of decrease of the variance for level corrections V_ℓ is approximately 2 for change of measure and 0.83 for standard MLMC, which is similar to the rate of decrease of the divergence probability.

The bottom left plot shows that the kurtosis of standard MLMC increases exponentially as the level ℓ increases while the kurtosis of change of measure remains constant. The increase rate is 0.96 which is close to the decrease of the divergence rate.

The bottom right plot implies that the total computational cost is $O(\varepsilon^{-2})$ for the MLMC with change of measure and $O(\varepsilon^{-3})$ for the standard Monte Carlo method.

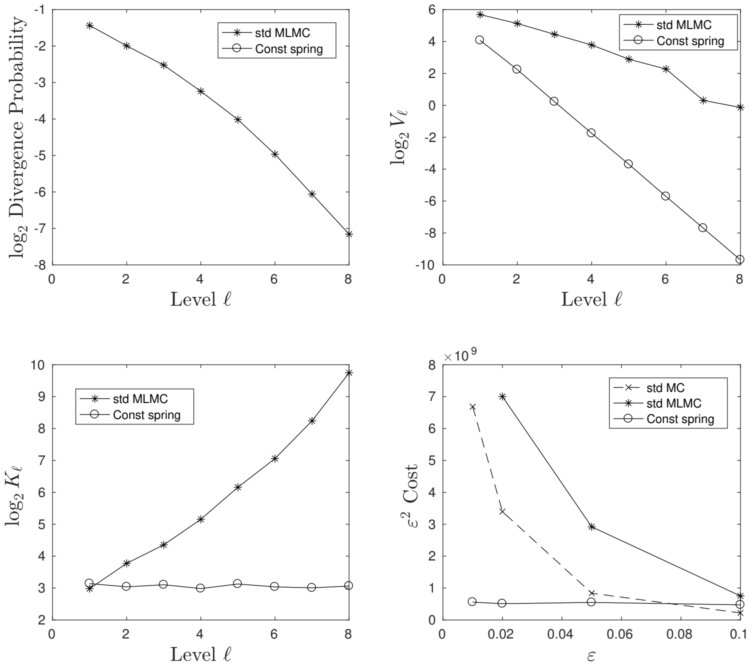


Figure 4.8: MLMC convergence test for the Lorenz equation

The computational cost for standard MLMC is worse than the standard Monte Carlo method due to the high kurtosis and large variance V_ℓ and it is already quite hard to get the result for $\varepsilon = 0.01$ in a reasonable computational time.

4.5 Extension to elliptic PDEs

In [33], Giles & Bernal proposed a new splitting method for MLMC to estimate the mean exit times for multi-dimensional SDEs and associated functionals corresponding to the solutions to high-dimensional parabolic PDEs. When the time horizon goes to infinity, the mean exit times with its functionals correspond to the solutions to elliptic PDEs. In this section, we will explore the possible application of the change of measure technique to this research direction.

The benefit of performing a change of measure for MLMC is similar as in the previous sections, since in some cases, like for a double-well potential SDE and the Lorenz equation, the fine path and coarse path may diverge to different wells or diverge exponentially. Adding a spring term to the two paths can avoid the divergence and reduce the distance when one path exits the domain, which can reduce the variance and kurtosis.

We add the spring term to both paths until one of them leaves the domain and after that, we use the splitting method proposed in [33] to simulate the remaining path without the spring term and use the boundary correction technique proposed in [39]. Now, we present some numerical results for the double-well potential energy SDE. We consider the SDE (4.47) and estimate the expected exit time $\mathbb{E}[\tau]$ from the domain $\{x \in \mathbb{R} : \ |x| < 2\}$, which correspond to the solution to the following one-dimensional elliptic PDE (actually an ODE):

$$\begin{aligned} \frac{1}{2}\frac{\partial^2 V}{\partial x^2} + (2x - \frac{1}{2}x^3)\frac{\partial V}{\partial x} + 1 &= 0, \\ V(2) &= 0, \\ V(-2) &= 0. \end{aligned} \tag{4.56}$$

We use the finite difference method:

$$\begin{aligned} \frac{V_{n+1} - 2V_n + V_{n-1}}{2\Delta x^2} + (2x_n - \frac{1}{2}x_n^3)\frac{V_{n+1} - V_{n-1}}{2\Delta x} + 1 = 0, \quad n = 1, \cdots, N-1, \\ V_0 = 0, \quad V_N = 0, \end{aligned}$$

where $N = \frac{4}{\Delta x}$ and solve it by solving a linear system. For the initial point of SDE $X_0 = 0$, we get

$$\mathbb{E}[\tau] = V(0) = 1.3722 \tag{4.57}$$

with $\Delta x = 2 * 10^{-4}$.

We simulate the SDE with initial value $x_0 = 0$ to a sufficiently long time $T = 10$ to control the truncation error of the infinite time interval. Although (4.47) is non-globally Lipschitz, we simulate it in a bounded region and in this region it is Lipschitz and we can use the Euler-Maruyama method with a uniform timestep $2^{-\ell-2}$ for each level. We compare the following three schemes:

- standard MLMC;
- MLMC with splitting proposed in [33];
- MLMC with splitting and change of measure with adaptive spring coefficient

$$S = \max\left(0, 2 - 1.5x^2\right). \tag{4.58}$$

We ran 10000 samples for each level ℓ for the three schemes. The numerical results are shown in Figure 4.9. The top left figure shows the weak convergence rate; both algorithms have a rate 1. The top right figure demonstrates that the change of measure technique can reduce the variance and result in a higher decrease rate than the standard MLMC method. The bottom left figure plots the kurtosis with respect to the level ℓ and illustrates that the change of measure technique does reduce the kurtosis significantly and the kurtosis remains bounded while the kurtosis for the standard MLMC increases exponentially. The bottom right figure shows that the computational cost is reduced by a constant factor and both MLMC schemes achieve $O(\varepsilon^{-2})$ cost while standard Monte Carlo method has $O(\varepsilon^{-3})$ computational cost.

4.6 Proofs

For simplicity of the proof, we introduce the notation $a(h) \lesssim b(h)$ which means that there exists a constant $\tilde{h}_0 \in (0, \infty)$ such that for all $h \in (0, \tilde{h}_0)$ it holds that $a(h) \leq b(h)$, where $\tilde{h}_0$ is allowed to depend on constants such as S, K, $\tilde{\alpha}$, $\tilde{\beta}$, $f(0)$ but not on stochastic samples ω or Brownian paths.

Note that for all $\delta \in (0, \infty)$, it holds that

$$1/(1 - Sh) \lesssim 1/(1 - 2Sh) \lesssim 1 + 2Sh + \delta h \lesssim 2. \tag{4.59}$$

4.6.1 Theorem 4.2.1

Proof of Theorem 4.2.1. The proof is given for $p \in [4, \infty)$; the result for $p \in [1, 4)$ follows from Hölder's inequality. We start our proof by analyzing the numerical paths step by step. When $t = t_0 = 0$, the two numerical paths are both at the initial point x_0, i.e. $\widehat{Y}^f_{t_0} = \widehat{Y}^c_{t_0} = x_0$.

For the odd time point t_{2n+1} for $n \in \{0, 1, 2, ..., N/2 - 1\}$, we obtain that

$$\begin{aligned} \widehat{Y}^c_{t_{2n+1}} &= \widehat{Y}^c_{t_{2n}} + S(\widehat{Y}^f_{t_{2n}} - \widehat{Y}^c_{t_{2n}})h + f(\widehat{Y}^c_{t_{2n}})h + \Delta W_{2n}, \\ \widehat{Y}^f_{t_{2n+1}} &= \widehat{Y}^f_{t_{2n}} + S(\widehat{Y}^c_{t_{2n}} - \widehat{Y}^f_{t_{2n}})h + f(\widehat{Y}^f_{t_{2n}})h + \Delta W_{2n}. \end{aligned} \tag{4.60}$$

Squaring both sides of the first equality in (4.60) implies that

$$\|\widehat{Y}^c_{t_{2n+1}}\|^2 = \left\| Sh\,\widehat{Y}^f_{t_{2n}} + (1 - Sh)\left(\widehat{Y}^c_{t_{2n}} + \frac{f(\widehat{Y}^c_{t_{2n}})h + \Delta W_{2n}}{1 - Sh}\right)\right\|^2. \tag{4.61}$$

Due to the convexity of x^2, for any $\xi \in [0, 1]$, it holds that

$$\|\xi A + (1 - \xi)B\|^2 \le \xi\|A\|^2 + (1 - \xi)\|B\|^2, \tag{4.62}$$

provided $h \in (0, 1/S)$, so we can choose $\xi = Sh$ to obtain that

$$\begin{aligned} \|\widehat{Y}^c_{t_{2n+1}}\|^2 \;\lesssim\;& Sh\|\widehat{Y}^f_{t_{2n}}\|^2 + (1 - Sh)\|\widehat{Y}^c_{t_{2n}}\|^2 + 4\,\|\Delta W_{2n}\|^2 + 2\,\langle \widehat{Y}^c_{t_{2n}}, f(\widehat{Y}^c_{t_{2n}})\rangle h \\ & + 4\,\|f(\widehat{Y}^c_{t_{2n}})\|^2 h^2 + 2\,\langle \widehat{Y}^c_{t_{2n}}, \Delta W_{2n}\rangle. \end{aligned} \tag{4.63}$$

The globally Lipschitz condition (4.21) ensures that

$$\|f(\widehat{Y}^c_{t_{2n}})\|^2 h^2 \lesssim \gamma\, h(\|\widehat{Y}^c_{t_{2n}}\|^2 + 1) \tag{4.64}$$

for any $\gamma \in (0, \infty)$. Combining this with the dissipativity condition (4.22), we obtain, for some fixed $\alpha \in (0, \tilde{\alpha})$ and $\beta \in (\tilde{\beta}, \infty)$, that

$$\begin{aligned} \|\widehat{Y}^c_{t_{2n+1}}\|^2 \;\lesssim\;& Sh\|\widehat{Y}^f_{t_{2n}}\|^2 + (1 - Sh - 2\alpha h)\|\widehat{Y}^c_{t_{2n}}\|^2 + 4\,\|\Delta W_{2n}\|^2 + 2\beta h \\ & + 2\,\langle \widehat{Y}^c_{t_{2n}}, \Delta W_{2n}\rangle. \end{aligned} \tag{4.65}$$

Similarly, we obtain that

$$\begin{aligned}\|\widehat{Y}^f_{t_{2n+1}}\|^2 &\lesssim Sh\|\widehat{Y}^c_{t_{2n}}\|^2+(1-Sh-2\alpha h)\|\widehat{Y}^f_{t_{2n}}\|^2+4\,\|\Delta W_{2n}\|^2+2\beta h\\ &\quad+2\,\langle \widehat{Y}^f_{t_{2n}},\Delta W_{2n}\rangle. \end{aligned}\tag{4.66}$$

For an even point t_{2n+2} for $n\in\{0,1,2,...,N/2-1\}$, we obtain that

$$\begin{aligned}\widehat{Y}^c_{t_{2n+2}} &= \widehat{Y}^c_{t_{2n}}+S(\widehat{Y}^f_{t_{2n}}-\widehat{Y}^c_{t_{2n}})2h+f(\widehat{Y}^c_{t_{2n}})2h+\Delta W_{2n}+\Delta W_{2n+1},\\ \widehat{Y}^f_{t_{2n+2}} &= \widehat{Y}^f_{t_{2n+1}}+S(\widehat{Y}^c_{t_{2n+1}}-\widehat{Y}^f_{t_{2n+1}})h+f(\widehat{Y}^f_{t_{2n+1}})h+\Delta W_{2n+1}.\end{aligned}\tag{4.67}$$

Using the same approach and choosing $\xi=2Sh$ provided $2Sh\in(0,1)$, we obtain that

$$\begin{aligned}\|\widehat{Y}^c_{t_{2n+2}}\|^2 &\lesssim 2Sh\|\widehat{Y}^f_{t_{2n}}\|^2+(1-2Sh-4\alpha h)\|\widehat{Y}^c_{t_{2n}}\|^2+4\|\Delta W_{2n}+\Delta W_{2n+1}\|^2\\ &\quad+4\beta h+\,2\,\langle \widehat{Y}^c_{t_{2n}},\Delta W_{2n}+\Delta W_{2n+1}\rangle,\end{aligned}\tag{4.68}$$

and

$$\begin{aligned}\|\widehat{Y}^f_{t_{2n+2}}\|^2 &\lesssim Sh\|\widehat{Y}^c_{t_{2n+1}}\|^2+(1-Sh-2\alpha h)\|\widehat{Y}^f_{t_{2n+1}}\|^2+4\,\|\Delta W_{2n+1}\|^2+2\beta h\\ &\quad+2\,\langle \widehat{Y}^f_{t_{2n+1}},\Delta W_{2n+1}\rangle.\end{aligned}\tag{4.69}$$

Therefore, for any fixed $\gamma\in(0,\alpha)$, we obtain that

$$\begin{aligned}\|\widehat{Y}^c_{t_{2n+2}}\|^2+\|\widehat{Y}^f_{t_{2n+2}}\|^2 &\lesssim (1-4\gamma h)(\|\widehat{Y}^c_{t_{2n}}\|^2+\|\widehat{Y}^f_{t_{2n}}\|^2)+12(\|\Delta W_{2n}\|^2+\|\Delta W_{2n+1}\|^2)\\ &\quad+8\beta h+\,2\,\mathrm{e}^{-4\gamma h}\langle\phi_{t_{2n}},\Delta W_{2n}\rangle+2\,\mathrm{e}^{-2\gamma h}\langle\phi_{t_{2n+1}},\Delta W_{2n+1}\rangle,\end{aligned}\tag{4.70}$$

where for $n\in\{0,1,2,...,N/2-1\}$, it holds that

$$\mathrm{e}^{-4\gamma h}\phi_{t_{2n}}=(1+Sh)\widehat{Y}^c_{t_{2n}}+(1-Sh-2\alpha h)\widehat{Y}^f_{t_{2n}},\quad \mathrm{e}^{-2\gamma h}\phi_{t_{2n+1}}=\widehat{Y}^c_{t_{2n}}+\widehat{Y}^f_{t_{2n+1}}.\tag{4.71}$$

Since $1-4\gamma h\le \mathrm{e}^{-4\gamma h}$ and $\mathrm{e}^{4\gamma h}\lesssim 2$, we multiply by $\mathrm{e}^{2\gamma t_{2n+2}}$ on both sides to obtain that

$$\begin{aligned}\mathrm{e}^{2\gamma t_{2n+2}}(\|\widehat{Y}^f_{t_{2n+2}}\|^2+\|\widehat{Y}^c_{t_{2n+2}}\|^2) &\lesssim \mathrm{e}^{2\gamma t_{2n}}(\|\widehat{Y}^f_{t_{2n}}\|^2+\|\widehat{Y}^c_{t_{2n}}\|^2)+\,16\beta\mathrm{e}^{2\gamma t_{2n}}h\\ &\quad+24\,\mathrm{e}^{2\gamma t_{2n}}(\|\Delta W_{2n}\|^2+\|\Delta W_{2n+1}\|^2)+2\,\mathrm{e}^{2\gamma t_{2n}}\langle\phi_{t_{2n}},\Delta W_{2n}\rangle\\ &\quad+2\,\mathrm{e}^{2\gamma t_{2n+1}}\langle\phi_{t_{2n+1}},\Delta W_{2n+1}\rangle.\end{aligned}\tag{4.72}$$

Summing over multiple timesteps implies that

$$\begin{aligned}
\mathrm{e}^{2\gamma t_{2n}}(\|\widehat{Y}^f_{t_{2n}}\|^2+\|\widehat{Y}^c_{t_{2n}}\|^2) &\lesssim (\|\widehat{Y}^f_{t_0}\|^2+\|\widehat{Y}^c_{t_0}\|^2)+24\sum_{k=0}^{2n-1}\mathrm{e}^{2\gamma t_k}\|\Delta W_k\|^2\\
&\quad +16\beta\sum_{k=0}^{n-1}\mathrm{e}^{2\gamma t_{2k}}h+\ 2\sum_{k=0}^{2n-1}\mathrm{e}^{2\gamma t_k}\langle\phi_{t_k},\Delta W_k\rangle. \qquad (4.73)
\end{aligned}$$

For odd time points, combining (4.65) and (4.66), by the Cauchy-Schwarz inequality and Young's inequality, there exist constants $\alpha_1\in(1,\infty)$, $\beta_1\in(\max(1,\alpha_1\beta),\infty)$ such that

$$\begin{aligned}
\|\widehat{Y}^f_{t_{2n+1}}\|^2+\|\widehat{Y}^c_{t_{2n+1}}\|^2 &\lesssim (1-2\alpha h)(\|\widehat{Y}^f_{t_{2n}}\|^2+\|\widehat{Y}^c_{t_{2n}}\|^2)+8\,\|\Delta W_{2n}\|^2+4\beta h\\
&\quad +2\langle\widehat{Y}^c_{t_{2n}},\Delta W_{2n}\rangle+2\langle\widehat{Y}^f_{t_{2n}},\Delta W_{2n}\rangle \qquad (4.74)\\
&\lesssim (1-2\gamma h)\Big(\alpha_1(\|\widehat{Y}^f_{t_{2n}}\|^2+\|\widehat{Y}^c_{t_{2n}}\|^2+12\,\|\Delta W_{2n}\|^2)+4\,\beta_1 h\Big).
\end{aligned}$$

Multiplying by $\mathrm{e}^{2\gamma t_{2n+1}}$ on both sides and using the (4.73) ensures that

$$\begin{aligned}
\mathrm{e}^{2\gamma t_{2n+1}}(\|\widehat{Y}^f_{t_{2n+1}}\|^2+\|\widehat{Y}^c_{t_{2n+1}}\|^2) &\lesssim \alpha_1(\|\widehat{Y}^f_{t_0}\|^2+\|\widehat{Y}^c_{t_0}\|^2)+24\,\alpha_1\sum_{k=0}^{2n}\mathrm{e}^{2\gamma t_k}\|\Delta W_k\|^2\\
&\quad +16\,\beta_1\sum_{k=0}^{n}\mathrm{e}^{2\gamma t_{2k}}h+2\,\alpha_1\sum_{k=0}^{2n-1}\mathrm{e}^{2\gamma t_k}\langle\phi_{t_k},\Delta W_k\rangle. \quad (4.75)
\end{aligned}$$

Then, combing (4.73) and (4.75), raising both sides to power $p/2$, taking the supremum over $n\in\{0,1,2,...,N\}$ and taking the expectation on the both sides, by using Jensen's inequality, we obtain that

$$\mathbb{E}\left[\sup_{n\in\{0,1,2,\ldots,N\}}\mathrm{e}^{\gamma p t_n}\left(\|\widehat{Y}^f_{t_n}\|^2+\|\widehat{Y}^c_{t_n}\|^2\right)^{p/2}\right]\lesssim 4^{p/2-1}(24\,\alpha_1\beta_1)^{p/2}(I_1+I_2+I_3+I_4), \qquad (4.76)$$

where

$$
\begin{aligned}
I_1 &= \mathbb{E}\left[(\|\widehat{Y}^f_{t_0}\|^2 + \|\widehat{Y}^c_{t_0}\|^2)^{p/2}\right] = 2^{p/2}\|x_0\|^{p/2}, \\
I_2 &= \left|\sum_{k=0}^{N-1} \mathrm{e}^{2\gamma t_k} h\right|^{p/2}, \\
I_3 &= \mathbb{E}\left[\left|\sum_{k=0}^{N-1} \mathrm{e}^{2\gamma t_k}\|\Delta W_k\|^2\right|^{p/2}\right], \\
I_4 &= \mathbb{E}\left[\sup_{n\in\{0,1,2,\dots,N/2\}}\left|\sum_{k=0}^{2n-1} \mathrm{e}^{2\gamma t_k}\langle \phi_{t_k}, \Delta W_k\rangle\right|^{p/2}\right]. \qquad (4.77)
\end{aligned}
$$

We will bound these four parts separately. I_1 is a constant. For I_2, we obtain that

$$
I_2 \le \left|\int_0^T \mathrm{e}^{2\gamma t}\,\mathrm{d}t\right|^{p/2} \le \mathrm{e}^{\gamma p T}/(2\gamma)^{p/2}. \qquad (4.78)
$$

Next, if $q_k \ge 0$, for any $k \in \{1, 2, \dots, n\}$, is an arbitrary discrete probability distribution, and b_k, for $k \in \{1, 2, \dots, n\}$, is a set of scalar values, then for any $p \in (1, \infty)$ Jensen's inequality implies that

$$
\left|\sum_{k=1}^{n} q_k b_k\right|^p \le \sum_{k=1}^{n} q_k |b_k|^p. \qquad (4.79)
$$

If a_k, for any $k \in \{1, 2, \dots, n\}$, is a set of positive scalar values, then setting $q_k = a_k / \sum_{k'=1}^{n} a_{k'}$ ensures that

$$
\left|\sum_{k=1}^{n} a_k b_k\right|^p \le \left|\sum_{k=1}^{n} a_k\right|^{p-1} \sum_{k=1}^{n} a_k |b_k|^p. \qquad (4.80)
$$

For I_3, using this inequality, we obtain that

$$
\begin{aligned}
I_3 &= \mathbb{E}\left[\left|\sum_{k=0}^{N-1} \mathrm{e}^{2\gamma t_k} h \frac{\|\Delta W_k\|^2}{h}\right|^{p/2}\right] \le \left|\sum_{k=0}^{N-1} \mathrm{e}^{2\gamma t_k} h\right|^{p/2-1} \mathbb{E}\left[\sum_{k=0}^{N-1} \mathrm{e}^{2\gamma t_k} h \frac{\|\Delta W_k\|^p}{h^{p/2}}\right] \\
&\le c_p\, \mathrm{e}^{\gamma p T}/(2\gamma)^{p/2}, \qquad (4.81)
\end{aligned}
$$

where c_p is that $c_p = \mathbb{E}\left[\|\Delta W_k\|^p / h^{p/2}\right] \le d^{p/2}\, p!! \le d^{p/2} p^{p/2}$.

For I_4, we rewrite the summation as an Itô integral and then deduce by the

Burkholder-Davis-Gundy inequality in [5] that there exists a positive constant $C_{\text{BDG}} \in (0,\infty)$ independent of p such that

$$\begin{aligned} I_4 &\le \mathbb{E}\left[\sup_{0\le t\le T}\left|\int_0^t \mathrm{e}^{2\gamma\lfloor s/h\rfloor h}\langle \phi_{\lfloor s/h\rfloor h}, \mathrm{d}W_s\rangle\right|^{p/2}\right] \\ &\le (C_{\text{BDG}}\, p)^{p/4}\, \mathbb{E}\left[\left|\sum_{k=0}^{N-1}\mathrm{e}^{4\gamma t_k}\|\phi_{t_k}\|^2 h\right|^{p/4}\right], \end{aligned} \tag{4.82}$$

where, by Young's inequality, it holds that

$$\|\phi_{t_{2k}}\|^2 \lesssim 8(\|\widehat{Y}^c_{t_{2k}}\|^2 + \|\widehat{Y}^f_{t_{2k}}\|^2), \qquad \|\phi_{t_{2k+1}}\|^2 \lesssim 4\,(\|\widehat{Y}^c_{t_{2k}}\|^2 + \|\widehat{Y}^f_{t_{2k+1}}\|^2). \tag{4.83}$$

Then by Jensen's inequality and Young's inequality, for arbitrary $\zeta \in (0,\infty)$, we obtain that

$$\begin{aligned} I_4 &\le (12C_{\text{BDG}}p)^{p/4}\left|\sum_{k=0}^{N-1}\mathrm{e}^{2\gamma t_k}h\right|^{p/4-1}\mathbb{E}\left[\sum_{k=0}^{N-1}\mathrm{e}^{2\gamma t_k}h\sup_{n\in\{0,1,2,\dots,N\}}\mathrm{e}^{\gamma p t_n/2}(\|\widehat{Y}^c_{t_n}\|^2+\|\widehat{Y}^f_{t_n}\|^2)^{p/4}\right] \\ &\le \frac{1}{4\zeta}\,\mathbb{E}\left[\sup_{n\in\{0,1,2,\dots,N\}}\mathrm{e}^{\gamma p t_n}\left(\|\widehat{Y}^f_{t_n}\|^2+\|\widehat{Y}^c_{t_n}\|^2\right)^{p/2}\right] + \zeta\left(\frac{6\,C_{\text{BDG}}}{\gamma}\right)^{p/2} p^{p/2}\mathrm{e}^{\gamma p T}. \end{aligned} \tag{4.84}$$

Finally, combining all the estimates above and choosing $\zeta = 4^{p/2-1}(24\alpha_1\beta_1)^{p/2}$, there exists a constant $C_{(2)} \in (0,\infty)$ such that

$$\mathbb{E}\left[\sup_{n\in\{0,1,2,\dots,N\}}\mathrm{e}^{\gamma p t_n}\left(\|\widehat{Y}^f_{t_n}\|^2+\|\widehat{Y}^c_{t_n}\|^2\right)^{p/2}\right] \lesssim C_{(2)}^p\, p^{p/2}\mathrm{e}^{\gamma p T}, \tag{4.85}$$

which implies that there exists a constant $C_{(1)} \in (0,\infty)$ such that for all $h \in (0, C_{(1)})$ it holds that

$$\sup_{n\in\{0,1,2,\dots,N\}}\mathbb{E}\left[\|\widehat{Y}^f_{t_n}\|^p\right] \le C_{(2)}^p\, p^{p/2}, \qquad \sup_{n\in\{0,1,2,\dots,N\}}\mathbb{E}\left[\|\widehat{Y}^c_{t_n}\|^p\right] \le C_{(2)}^p\, p^{p/2}. \tag{4.86}$$

This completes the proof of Theorem 4.2.1. □

4.6.2 Theorem 4.2.2

Proof of Theorem 4.2.2. The proof is given for $p \in [4, \infty)$; the result for $p \in [1, 4)$ follows from Hölder's inequality.

The different updates on odd and even time points imply that

$$\widehat{Y}^f_{t_{2n+1}} - \widehat{Y}^c_{t_{2n+1}} = (1 - 2Sh)(\widehat{Y}^f_{t_{2n}} - \widehat{Y}^c_{t_{2n}}) + (f(\widehat{Y}^f_{t_{2n}}) - f(\widehat{Y}^c_{t_{2n}}))h, \tag{4.87}$$
$$\widehat{Y}^f_{t_{2n+2}} - \widehat{Y}^c_{t_{2n+2}} = (1 - Sh)(\widehat{Y}^f_{t_{2n+1}} - \widehat{Y}^c_{t_{2n+1}}) - Sh(\widehat{Y}^f_{t_{2n}} - \widehat{Y}^c_{t_{2n}}) + (f(\widehat{Y}^f_{t_{2n+1}}) - f(\widehat{Y}^c_{t_{2n}}))h,$$

and then

$$\begin{aligned}\widehat{Y}^f_{t_{2n+2}} - \widehat{Y}^c_{t_{2n+2}} = {} & (1 - 4Sh + 2S^2h^2)(\widehat{Y}^f_{t_{2n}} - \widehat{Y}^c_{t_{2n}}) + (1 - Sh)(f(\widehat{Y}^f_{t_{2n}}) - f(\widehat{Y}^c_{t_{2n}}))h \\ & + (f(\widehat{Y}^f_{t_{2n+1}}) - f(\widehat{Y}^f_{t_{2n}}))h. \end{aligned} \tag{4.88}$$

Taking the square of both sides ensures that

$$\begin{aligned}\|\widehat{Y}^f_{t_{2n+2}} - \widehat{Y}^c_{t_{2n+2}}\|^2 = {} & (1 - 4Sh + 2S^2h^2)^2 \|\widehat{Y}^f_{t_{2n}} - \widehat{Y}^c_{t_{2n}}\|^2 \\ & + (1 - Sh)^2 \|f(\widehat{Y}^f_{t_{2n}}) - f(\widehat{Y}^c_{t_{2n}})\|^2 h^2 + \|f(\widehat{Y}^f_{t_{2n+1}}) - f(\widehat{Y}^f_{t_{2n}})\|^2 h^2 \\ & + 2(1 - 4Sh + 2S^2h^2)(1 - Sh)\langle \widehat{Y}^f_{t_{2n}} - \widehat{Y}^c_{t_{2n}}, f(\widehat{Y}^f_{t_{2n}}) - f(\widehat{Y}^c_{t_{2n}})\rangle h \\ & + 2(1 - 4Sh + 2S^2h^2)\langle \widehat{Y}^f_{t_{2n}} - \widehat{Y}^c_{t_{2n}}, f(\widehat{Y}^f_{t_{2n+1}}) - f(\widehat{Y}^f_{t_{2n}})\rangle h \\ & + 2(1 - Sh)\langle f(\widehat{Y}^f_{t_{2n}}) - f(\widehat{Y}^c_{t_{2n}}), f(\widehat{Y}^f_{t_{2n+1}}) - f(\widehat{Y}^f_{t_{2n}})\rangle h^2. \end{aligned} \tag{4.89}$$

Then provided $S > \lambda/2$, Assumption 4.2.2, the globally Lipschitz condition (4.21), the Cauchy-Schwarz inequality and Young's inequality imply, for any fixed $\gamma \in (0, 2S - \lambda)$, that

$$\begin{aligned}\|\widehat{Y}^f_{t_{2n+2}} - \widehat{Y}^c_{t_{2n+2}}\|^2 \lesssim {} & (1 - 4\gamma h)\|\widehat{Y}^f_{t_{2n}} - \widehat{Y}^c_{t_{2n}}\|^2 + 2K^2 \|\widehat{Y}^f_{t_{2n+1}} - \widehat{Y}^f_{t_{2n}}\|^2 h^2 \\ & + 2(1 - 4Sh + 2S^2h^2)\langle \widehat{Y}^f_{t_{2n}} - \widehat{Y}^c_{t_{2n}}, f(\widehat{Y}^f_{t_{2n+1}}) - f(\widehat{Y}^f_{t_{2n}})\rangle h. \end{aligned} \tag{4.90}$$

Following this estimate, we use two different approaches to get different upper bounds.

FIRST APPROACH. We continue to use Young's inequality and the globally Lipschitz condition (4.21) to obtain that

$$\begin{aligned}\|\widehat{Y}^f_{t_{2n+2}} - \widehat{Y}^c_{t_{2n+2}}\|^2 \lesssim &(1-2\gamma h)\|\widehat{Y}^f_{t_{2n}} - \widehat{Y}^c_{t_{2n}}\|^2 \\ &+ \left((2\gamma)^{-1}+2\right) K^2 \|\widehat{Y}^f_{t_{2n+1}} - \widehat{Y}^f_{t_{2n}}\|^2 h.\end{aligned} \tag{4.91}$$

Then we multiply by $\mathrm{e}^{\gamma t_{2n+2}}$ on both sides and $\mathrm{e}^{2\gamma h} \lesssim 2$ to deduce that

$$\begin{aligned}\mathrm{e}^{\gamma t_{2n+2}}\|\widehat{Y}^f_{t_{2n+2}} - \widehat{Y}^c_{t_{2n+2}}\|^2 \lesssim &\mathrm{e}^{\gamma t_{2n}}\|\widehat{Y}^f_{t_{2n}} - \widehat{Y}^c_{t_{2n}}\|^2 \\ &+ (\gamma^{-1}+4)K^2\,\mathrm{e}^{\gamma t_{2n}}\|\widehat{Y}^f_{t_{2n+1}} - \widehat{Y}^f_{t_{2n}}\|^2 h.\end{aligned} \tag{4.92}$$

Summing over multiple timesteps and noting that $\widehat{Y}^f_{t_0} - \widehat{Y}^c_{t_0} = 0$ ensures that

$$\mathrm{e}^{\gamma t_{2n}}\|\widehat{Y}^f_{t_{2n}} - \widehat{Y}^c_{t_{2n}}\|^2 \lesssim (\gamma^{-1}+4)K^2 \sum_{k=0}^{n-1} \mathrm{e}^{\gamma t_{2k}}\|\widehat{Y}^f_{t_{2k+1}} - \widehat{Y}^f_{t_{2k}}\|^2 h. \tag{4.93}$$

Raising both sides to the power $p/2$, taking the supremum over $n\in\{0,1,2,...,N/2\}$, taking the expectation and by applying Jensen's inequality, we obtain that

$$\begin{aligned}\mathbb{E}\left[\sup_{n\in\{0,1,2,\dots,N/2\}} \mathrm{e}^{\gamma p t_{2n}/2}\|\widehat{Y}^f_{t_{2n}} - \widehat{Y}^c_{t_{2n}}\|^p\right] &\lesssim (\gamma^{-1}+4)^{p/2}K^p\,\mathbb{E}\left[\left|\sum_{k=0}^{N/2-1}\mathrm{e}^{\gamma t_{2k}}\|\widehat{Y}^f_{t_{2k+1}} - \widehat{Y}^f_{t_{2k}}\|^2 h\right|^{p/2}\right] \\ &\leq (\gamma^{-1}+4)^{p/2}K^p\left|\sum_{k=0}^{N/2-1}\mathrm{e}^{\gamma t_{2k}}h\right|^{p/2-1}\mathbb{E}\left[\sum_{k=0}^{N/2-1}\mathrm{e}^{\gamma t_{2k}}\|\widehat{Y}^f_{t_{2k+1}} - \widehat{Y}^f_{t_{2k}}\|^p h\right] \\ &\leq (\gamma^{-1}+4)^{p/2}K^p\left|\sum_{k=0}^{N/2-1}\mathrm{e}^{\gamma t_{2k}}h\right|^{p/2-1}\sum_{k=0}^{N/2-1}\mathrm{e}^{\gamma t_{2k}}\mathbb{E}\left[\|\widehat{Y}^f_{t_{2k+1}} - \widehat{Y}^f_{t_{2k}}\|^p\right]h.\end{aligned} \tag{4.94}$$

By the update on the fine path, the globally Lipschitz condition (4.21), Theorem 4.2.1 and Jensen's inequality, there exists a constant $C_1 \in (0,\infty)$ such that

$$\begin{aligned}\mathbb{E}\left[\|\widehat{Y}^f_{t_{2k+1}} - \widehat{Y}^f_{t_{2k}}\|^p\right] &= \mathbb{E}\left[\|f(\widehat{Y}^f_{t_{2k}})h + S(\widehat{Y}^c_{t_{2k}} - \widehat{Y}^f_{t_{2k}})h + \Delta W_{2k}\|^p\right] \\ &\leq 2^{p-1}\mathbb{E}\left[\|f(\widehat{Y}^f_{t_{2k}}) + S(\widehat{Y}^c_{t_{2k}} - \widehat{Y}^f_{t_{2k}})\|^p\right]h^p + 2^{p-1}\mathbb{E}\left[\|\Delta W_{2k}\|^p\right] \\ &\lesssim C_1^p p^{p/2} h^{p/2},\end{aligned} \tag{4.95}$$

which implies that

$$\mathbb{E}\left[\sup_{n\in\{0,1,2,\ldots,N/2\}} \mathrm{e}^{\gamma p t_{2n}/2}\|\widehat{Y}^f_{t_{2n}}-\widehat{Y}^c_{t_{2n}}\|^p\right] \lesssim 2^{-p/2}(\gamma^{-1}+4)^{p/2}K^pC_1^p\,p^{p/2}\mathrm{e}^{\gamma pT/2}h^{p/2}. \quad (4.96)$$

SECOND APPOACH. We directly multiply by $\mathrm{e}^{2\gamma t_{2n+2}}$ on both sides of (4.90) and note that $\mathrm{e}^{4\gamma h} \lesssim 2$ ensures that

$$\begin{aligned}\mathrm{e}^{2\gamma t_{2n+2}}\|\widehat{Y}^f_{t_{2n+2}} - \widehat{Y}^c_{t_{2n+2}}\|^2 \;&\lesssim\; \mathrm{e}^{2\gamma t_{2n}}\|\widehat{Y}^f_{t_{2n}} - \widehat{Y}^c_{t_{2n}}\|^2 + 4\,\mathrm{e}^{2\gamma t_{2n}}K^2\|\widehat{Y}^f_{t_{2n+1}} - \widehat{Y}^f_{t_{2n}}\|^2h^2 \\ &+ 2(1-4Sh+2S^2h^2)\,\mathrm{e}^{2\gamma t_{2n+2}}\langle \widehat{Y}^f_{t_{2n}} - \widehat{Y}^c_{t_{2n}}, f(\widehat{Y}^f_{t_{2n+1}}) - f(\widehat{Y}^f_{t_{2n}})\rangle h. \quad (4.97)\end{aligned}$$

Summing over multiple timesteps and noting that $\widehat{Y}^f_{t_0} - \widehat{Y}^c_{t_0} = 0$ implies that

$$\begin{aligned}\mathrm{e}^{2\gamma t_{2n}}\|\widehat{Y}^f_{t_{2n}} - \widehat{Y}^c_{t_{2n}}\|^2 \;&\lesssim\; 4\sum_{k=0}^{n-1}\mathrm{e}^{2\gamma t_{2k}}K^2\|\widehat{Y}^f_{t_{2k+1}} - \widehat{Y}^f_{t_{2k}}\|^2h^2 \quad (4.98)\\ &+ 2(1-4Sh+2S^2h^2)\sum_{k=0}^{n-1}\mathrm{e}^{2\gamma t_{2k+2}}\langle \widehat{Y}^f_{t_{2k}} - \widehat{Y}^c_{t_{2k}}, f(\widehat{Y}^f_{t_{2k+1}}) - f(\widehat{Y}^f_{t_{2k}})\rangle h.\end{aligned}$$

Raising both sides to the power $p/2$, taking supremum over $n \in \{0,1,2,...,N/2\}$, taking expectation and by Jensen's inequality, we obtain that

$$\mathbb{E}\left[\sup_{n\in\{0,1,2,\ldots,N/2\}} \mathrm{e}^{\gamma p t_{2n}}\|\widehat{Y}^f_{t_{2n}} - \widehat{Y}^c_{t_{2n}}\|^p\right] \lesssim 2^{p/2-1}4^{p/2}(I_1+I_2), \quad (4.99)$$

where

$$\begin{aligned} I_1 \;&=\; \mathbb{E}\left[\left|\sum_{k=0}^{N/2-1}\mathrm{e}^{2\gamma t_{2k}}K^2\|\widehat{Y}^f_{t_{2k+1}} - \widehat{Y}^f_{t_{2k}}\|^2h^2\right|^{p/2}\right], \quad (4.100)\\ I_2 \;&=\; \mathbb{E}\left[\sup_{n\in\{0,1,2,\ldots,N/2-1\}}\left|\sum_{k=0}^{n}\mathrm{e}^{2\gamma t_{2k}}\langle \widehat{Y}^f_{t_{2k}} - \widehat{Y}^c_{t_{2k}}, f(\widehat{Y}^f_{t_{2k+1}}) - f(\widehat{Y}^f_{t_{2k}})\rangle h\right|^{p/2}\right].\end{aligned}$$

For I_1, Jensen's inequality and the estimate (4.95) implies that

$$\begin{aligned} I_1 \;&\le\; \left|\sum_{k=0}^{N/2-1}\mathrm{e}^{2\gamma t_{2k}}h\right|^{p/2-1}\sum_{k=0}^{N/2-1}\mathrm{e}^{2\gamma t_{2k}}hK^p\mathbb{E}\left[\|\widehat{Y}^f_{t_{2k+1}} - \widehat{Y}^f_{t_{2k}}\|^p\right]h^{p/2}\\ &\lesssim\; (2\gamma)^{-p/2}K^pC_1^p\,p^{p/2}\,\mathrm{e}^{\gamma pT}h^p. \quad (4.101)\end{aligned}$$

For I_2, we perform a Taylor expansion and by the mean value theorem obtain that

$$\begin{aligned}\langle \widehat{Y}^f_{t_{2k}} - \widehat{Y}^c_{t_{2k}}, f(\widehat{Y}^f_{t_{2k+1}}) - f(\widehat{Y}^f_{t_{2k}})\rangle &= \langle \widehat{Y}^f_{t_{2k}} - \widehat{Y}^c_{t_{2k}}, \nabla f(\widehat{Y}^f_{t_{2k}})(\widehat{Y}^f_{t_{2k+1}} - \widehat{Y}^f_{t_{2k}})\rangle + R_k \\ &= \langle \widehat{Y}^f_{t_{2k}} - \widehat{Y}^c_{t_{2k}}, \nabla f(\widehat{Y}^f_{t_{2k}}) f(\widehat{Y}^f_{t_{2k}})\rangle h + \langle \widehat{Y}^f_{t_{2k}} - \widehat{Y}^c_{t_{2k}}, \nabla f(\widehat{Y}^f_{t_{2k}})\Delta W_{2k}\rangle \\ &\quad + S\langle \widehat{Y}^f_{t_{2k}} - \widehat{Y}^c_{t_{2k}}, \nabla f(\widehat{Y}^f_{t_{2k}})(\widehat{Y}^c_{t_{2k}} - \widehat{Y}^f_{t_{2k}})\rangle h + R_k, \end{aligned} \tag{4.102}$$

where $|R_k| \leq 2K\|\widehat{Y}^f_{t_{2k}} - \widehat{Y}^c_{t_{2k}}\|\|\widehat{Y}^f_{t_{2k+1}} - \widehat{Y}^f_{t_{2k}}\|^2$ and then Jensen's inequality implies that

$$I_2 = 4^{p/2-1}(J_1 + J_2 + J_3 + J_4), \tag{4.103}$$

where

$$\begin{aligned} J_1 &= \mathbb{E}\left[\sup_{n\in\{0,1,2,\ldots,N/2-1\}} \left|\sum_{k=0}^{n} \mathrm{e}^{2\gamma t_{2k}} \langle \widehat{Y}^f_{t_{2k}} - \widehat{Y}^c_{t_{2k}}, \nabla f(\widehat{Y}^f_{t_{2k}}) f(\widehat{Y}^f_{t_{2k}})\rangle h^2\right|^{p/2}\right], \\ J_2 &= \mathbb{E}\left[\sup_{n\in\{0,1,2,\ldots,N/2-1\}} \left|\sum_{k=0}^{n} \mathrm{e}^{2\gamma t_{2k}} S\langle \widehat{Y}^f_{t_{2k}} - \widehat{Y}^c_{t_{2k}}, \nabla f(\widehat{Y}^f_{t_{2k}})(\widehat{Y}^c_{t_{2k}} - \widehat{Y}^f_{t_{2k}})\rangle h^2\right|^{p/2}\right], \\ J_3 &= \mathbb{E}\left[\sup_{n\in\{0,1,2,\ldots,N/2-1\}} \left|\sum_{k=0}^{n} \mathrm{e}^{2\gamma t_{2k}} \langle \widehat{Y}^f_{t_{2k}} - \widehat{Y}^c_{t_{2k}}, \nabla f(\widehat{Y}^f_{t_{2k}})\Delta W_{2k}\rangle h\right|^{p/2}\right], \\ J_4 &= \mathbb{E}\left[\sup_{n\in\{0,1,2,\ldots,N/2-1\}} \left|\sum_{k=0}^{n} \mathrm{e}^{2\gamma t_{2k}} R_k\, h\right|^{p/2}\right]. \end{aligned} \tag{4.104}$$

For J_1, by the Cauchy-Schwarz inequality, Jensen's inequality, Young's inequality, the globally Lipschitz property of f and ∇f and Theorem 4.2.1, for any $\zeta \in (0,\infty)$, there exists a constant $C_{31} \in (0,\infty)$ such that

$$
\begin{aligned}
J_1 &\le \mathbb{E}\left[\left|\sup_{n\in\{0,1,2,\ldots,N/2\}} \mathrm{e}^{\gamma p t_{2n}/2}\|\widehat{Y}^f_{t_{2n}}-\widehat{Y}^c_{t_{2n}}\|^{p/2}\right|\left|\sum_{k=0}^{N/2-1}\mathrm{e}^{2\gamma kh}h\right|^{p/2-1}\right.\\
&\qquad\qquad\left.\times\sum_{k=0}^{N/2-1}\mathrm{e}^{2\gamma kh}h\|\nabla f(\widehat{Y}^f_{t_{2k}})f(\widehat{Y}^f_{t_{2k}})\|^{p/2}h^{p/2}\right]\\
&\le \frac{1}{4\zeta}\mathbb{E}\left[\sup_{n\in\{0,1,2,\ldots,N/2\}}\mathrm{e}^{\gamma p t_{2n}}\|\widehat{Y}^f_{t_{2n}}-\widehat{Y}^c_{t_{2n}}\|^{p}\right]\\
&\qquad+\zeta\ \mathbb{E}\left[\left|\sum_{k=0}^{N/2-1}\mathrm{e}^{2\gamma kh}h\right|^{p-1}\sum_{k=0}^{N/2-1}\mathrm{e}^{2\gamma kh}h\|\nabla f(\widehat{Y}^f_{t_{2k}})f(\widehat{Y}^f_{t_{2k}})\|^{p}h^{p}\right]\\
&\le \frac{1}{4\zeta}\mathbb{E}\left[\sup_{n\in\{0,1,2,\ldots,N/2\}}\mathrm{e}^{\gamma p t_{2n}}\|\widehat{Y}^f_{t_{2n}}-\widehat{Y}^c_{t_{2n}}\|^{p}\right]+\zeta\,C_{31}^p\,p^p\,\mathrm{e}^{\gamma pT}\,h^p. \qquad (4.105)
\end{aligned}
$$

Similarly, for J_2, there exists a constant $C_{32}\in(0,\infty)$ such that for any $\zeta\in(0,\infty)$ it holds that

$$
J_2 \le \frac{1}{4\zeta}\mathbb{E}\left[\sup_{0\le n\le N/2}\mathrm{e}^{\gamma p t_{2n}}\|\widehat{Y}^f_{t_{2n}}-\widehat{Y}^c_{t_{2n}}\|^{p}\right]+\zeta\,C_{32}^p\,p^p\,\mathrm{e}^{\gamma pT}\,h^p. \qquad (4.106)
$$

Then, for J_3, by the Burkholder-Davis-Gundy inequality in [5], the globally Lipschitz property of ∇f and Theorem 4.2.1, there exists a constant $C_{33}\in(0,\infty)$ such that for any $\zeta\in(0,\infty)$ it holds that

$$
\begin{aligned}
J_3 &\le (C_{\mathrm{BDG}}\,p)^{p/4}\ \mathbb{E}\left[\left|\sum_{k=0}^{N/2-1}\mathrm{e}^{4\gamma t_{2k}}\|\widehat{Y}^f_{t_{2k}}-\widehat{Y}^c_{t_{2k}}\|^2\|\nabla f(\widehat{Y}^f_{t_{2k}})\|^2h^3\right|^{p/4}\right]\\
&\le (C_{\mathrm{BDG}}\,p)^{p/4}\mathbb{E}\left[\left|\sup_{n\in\{0,1,2,\ldots,N/2\}}\mathrm{e}^{\frac{\gamma p t_{2n}}{2}}\|\widehat{Y}^f_{t_{2n}}-\widehat{Y}^c_{t_{2n}}\|^{\frac{p}{2}}\right|\left|\sum_{k=0}^{N/2-1}\mathrm{e}^{2\gamma t_{2k}}h\right|^{\frac{p}{4}-1}\right.\\
&\qquad\qquad\left.\times\sum_{k=0}^{N/2-1}\mathrm{e}^{2\gamma t_{2k}}h\|\nabla f(\widehat{Y}^f_{t_{2k}})\|^{\frac{p}{2}}h^{\frac{p}{2}}\right]\\
&\le \frac{1}{4\zeta}\ \mathbb{E}\left[\sup_{n\in\{0,1,2,\ldots,N/2\}}\mathrm{e}^{\gamma p t_{2n}}\|\widehat{Y}^f_{t_{2n}}-\widehat{Y}^c_{t_{2n}}\|^{p}\right]+\zeta\,C_{33}^p\,p^p\,\mathrm{e}^{\gamma pT}\,h^p. \qquad (4.107)
\end{aligned}
$$

Similarly, for J_4, by Jensen's inequality, Young's inequality, the globally Lipschitz property of f and ∇f and Theorem 4.2.1, there exists a constant $C_{34} \in (0,\infty)$ such that for any $\zeta \in (0,\infty)$ it holds that

$$\begin{aligned}
J_4 \le\ & \mathbb{E}\left[\left|\sum_{k=0}^{N/2-1} \mathrm{e}^{2\gamma t_{2k}} h\right|^{p/2-1} \sum_{k=0}^{N/2-1} \mathrm{e}^{2\gamma t_{2k}} h \mathrm{e}^{\gamma p t_{2k}/2} \|\widehat{Y}^f_{t_{2k}} - \widehat{Y}^c_{t_{2k}}\|^{p/2} K^{p/2} \|(\widehat{Y}^f_{t_{2k+1}} - \widehat{Y}^f_{t_{2k}})\|^p\right] \\
\le\ & \mathbb{E}\left[\left|\sup_{n\in\{0,1,2,\ldots,N/2\}} \mathrm{e}^{\gamma p t_{2n}/2} \|\widehat{Y}^f_{t_{2n}} - \widehat{Y}^c_{t_{2n}}\|^{p/2}\right| \left|\sum_{k=0}^{N/2-1} \mathrm{e}^{2\gamma t_{2k}} h\right|^{p/2-1} \right. \\
& \left. \times \sum_{k=0}^{N/2-1} \mathrm{e}^{2\gamma t_{2k}} h K^{p/2} \|(\widehat{Y}^f_{t_{2k+1}} - \widehat{Y}^f_{t_{2k}})\|^p\right] \\
\le\ & \frac{1}{4\zeta}\, \mathbb{E}\left[\sup_{n\in\{0,1,2,\ldots,N/2\}} \mathrm{e}^{\gamma p t_{2n}} \|\widehat{Y}^f_{t_{2n}} - \widehat{Y}^c_{t_{2n}}\|^p\right] + \zeta\, \mathbb{E}\left[\left|\sum_{k=0}^{N/2-1} \mathrm{e}^{2\gamma t_{2k}} h\right|^{p-1} \right. \\
& \left. \times \sum_{k=0}^{N/2-1} \mathrm{e}^{2\gamma t_{2k}} h K^p 2^{2p-1} \left(\|f(\widehat{Y}^f_{t_{2k}}) + S(\widehat{Y}^c_{t_{2k}} - \widehat{Y}^f_{t_{2k}})\|^{2p} h^{2p} + \|\Delta W_{2k}\|^{2p}\right)\right] \\
\lesssim\ & \frac{1}{4\zeta} \mathbb{E}\left[\sup_{n\in\{0,1,2,\ldots,N/2\}} \mathrm{e}^{\gamma p t_{2n}} \|\widehat{Y}^f_{t_{2n}} - \widehat{Y}^c_{t_{2n}}\|^p\right] + \zeta\, C_{34}^p\, p^p \mathrm{e}^{\gamma p T}\, h^p.
\end{aligned} \tag{4.108}$$

Finally, by choosing $\zeta = 2^{5p/2-2}$, there exists a constant $C_4 \in (0,\infty)$ such that

$$\mathbb{E}\left[\sup_{n\in\{0,1,2,\ldots,N/2\}} \mathrm{e}^{\gamma p t_{2n}} \|\widehat{Y}^f_{t_{2n}} - \widehat{Y}^c_{t_{2n}}\|^p\right] \lesssim C_4^p\, p^p\, \mathrm{e}^{\gamma p T} h^p, \tag{4.109}$$

which together with (4.96) implies that there exists a constant $C_5 \in (0,\infty)$ such that

$$\sup_{n\in\{0,1,2,\ldots,N/2\}} \mathbb{E}\left[\|\widehat{Y}^f_{t_{2n}} - \widehat{Y}^c_{t_{2n}}\|^p\right] \lesssim C_5^p \min\left(p^{p/2}\, h^{p/2},\ p^p\, h^p\right). \tag{4.110}$$

For the odd time points, the globally Lipschitz condition (4.21) and (4.25) imply that

$$\begin{aligned}
\|\widehat{Y}^f_{t_{2n+1}} - \widehat{Y}^c_{t_{2n+1}}\|^2 &\le \left((1-2Sh)^2 + 2h\lambda(1-2Sh) + K^2h^2\right) \|\widehat{Y}^f_{t_{2n}} - \widehat{Y}^c_{t_{2n}}\|^2 \\
&\lesssim 2\, \|\widehat{Y}^f_{t_{2n}} - \widehat{Y}^c_{t_{2n}}\|^2,
\end{aligned} \tag{4.111}$$

which, by raising both sides to power $p/2$ and taking the expectation, there exist constants $C_{(1)},\ C_{(2)} \in (0,\infty)$ such that for all $h \in (0, C_{(1)})$ it implies that

$$\sup_{n\in\{0,1,2,\ldots,N\}} \mathbb{E}\left[\|\widehat{Y}^f_{t_n} - \widehat{Y}^c_{t_n}\|^p\right] \le C^p_{(2)} \min\left(p^{p/2}\, h^{p/2},\ p^p\, h^p\right). \tag{4.112}$$

This completes the proof of Theorem 4.2.2. □

4.6.3 Theorem 4.2.3

Proof of Theorem 4.2.3. For simplicity, we only demonstrate the proof for $\frac{\mathrm{d}\widehat{\mathbb{Q}}^c}{\mathrm{d}\mathbb{P}}$, and the result for $\frac{\mathrm{d}\mathbb{Q}^f}{\mathrm{d}\mathbb{P}}$ follows similarly.

Now we write down the details of the exact Radon-Nikodym derivative (4.17). Note that

$$\begin{aligned}
\mathbb{E}\left[\left|\frac{\mathrm{d}\widehat{\mathbb{Q}}^c}{\mathrm{d}\mathbb{P}}\right|^p\right] &= \mathbb{E}\Bigg[\exp\Bigg(-pS\sum_{n=0}^{N/2-1}\left\langle \widehat{Y}^f_{t_{2n}} - \widehat{Y}^c_{t_{2n}},\ \Delta W_{2n} + \Delta W_{2n+1}\right\rangle \\
&\qquad\qquad -\frac{p}{2}S^2\sum_{n=0}^{N/2-1}\left\|\widehat{Y}^f_{t_{2n}} - \widehat{Y}^c_{t_{2n}}\right\|^2 2h\Bigg)\Bigg] \\
&= \mathbb{E}\Bigg[\exp\left(\frac{p(2p-1)}{2}S^2\sum_{n=0}^{N/2-1}\left\|\widehat{Y}^f_{t_{2n}} - \widehat{Y}^c_{t_{2n}}\right\|^2 2h\right) \qquad (4.113) \\
&\times\exp\left(-pS\sum_{n=0}^{N/2-1}\left\langle \widehat{Y}^f_{t_{2n}} - \widehat{Y}^c_{t_{2n}}, \Delta W_{2n} + \Delta W_{2n+1}\right\rangle - p^2S^2\sum_{n=0}^{N/2-1}\left\|\widehat{Y}^f_{t_{2n}} - \widehat{Y}^c_{t_{2n}}\right\|^2 2h\right)\Bigg].
\end{aligned}$$

Hölder's inequality implies that

$$\mathbb{E}\left[\left|\frac{\mathrm{d}\widehat{\mathbb{Q}}^c}{\mathrm{d}\mathbb{P}}\right|^p\right] \le I_1^{1/2}\, I_2^{1/2}, \tag{4.114}$$

where

$$I_1 = \mathbb{E}\left[\exp\left(p(2p-1)S^2\sum_{n=0}^{N/2-1}\left\|\widehat{Y}^f_{t_{2n}} - \widehat{Y}^c_{t_{2n}}\right\|^2 2h\right)\right], \tag{4.115}$$

$$I_2 = \mathbb{E}\left[\exp\left(-2pS\sum_{n=0}^{N/2-1}\left\langle \widehat{Y}^f_{t_{2n}} - \widehat{Y}^c_{t_{2n}}, \Delta W_{2n} + \Delta W_{2n+1}\right\rangle - 2p^2S^2\sum_{n=0}^{N/2-1}\left\|\widehat{Y}^f_{t_{2n}} - \widehat{Y}^c_{t_{2n}}\right\|^2 2h\right)\right]$$
$$\leq \; 1,$$

since the exponential term in I_2 is a super-martingale.

For I_1, by Fatou's lemma and Jensen's inequality, we obtain that

$$\begin{aligned} I_1 &\leq \sum_{k=0}^{\infty} \frac{\mathbb{E}\left[\left|2(pS)^2\sum_{n=0}^{N/2-1}\|\widehat{Y}^f_{t_{2n}} - \widehat{Y}^c_{t_{2n}}\|^2 2h\right|^k\right]}{k!} \\ &\leq \sum_{k=0}^{\infty} \frac{2^k(pS)^{2k}T^{k-1}\sum_{n=0}^{N/2-1}\mathbb{E}\left[\left\|\widehat{Y}^f_{t_{2n}} - \widehat{Y}^c_{t_{2n}}\right\|^{2k}\right]2h}{k!}. \end{aligned} \tag{4.116}$$

Then by Theorem 4.2.2 and Stirling's approximation $k! \geq \sqrt{2\pi}k^{k+1/2}\mathrm{e}^{-k}$ for any $k \in \{1, 2, ...\}$, there exist constants $C_1,\ C_2 \in (0, \infty)$ such that

$$I_1 \;\lesssim\; 1 + \sum_{k=1}^{\infty}\frac{(2pSC_2)^{2k}(Th/\mathrm{e})^k}{\sqrt{2\pi k}} < 2, \tag{4.117}$$

provided $(2pSC_2)^2Th/\mathrm{e} \in (0, 1/2)$.

Therefore, for all $T \in (0, \infty)$ and $p \in [1, \infty)$, there exist constants $C_{(1)},\ C_{(2)} \in (0, \infty)$ such that for all $h \in (0, \min(C_{(1)}, C_{(2)}/(Tp^2)))$ it holds that

$$\mathbb{E}\left[\left|\frac{\mathrm{d}\widehat{\mathbb{Q}}^c}{\mathrm{d}\mathbb{P}}\right|^p\right] \leq 2. \tag{4.118}$$

This completes the proof of Theorem 4.2.3. □

4.6.4 Theorem 4.2.4

Proof of Theorem 4.2.4. By Jensen's and Hölder's inequalities, we split the expectation into three parts

$$\begin{aligned}
&\mathbb{E}\left[\left|\varphi(\widehat{Y}_T^f)\tfrac{\mathrm{d}\widehat{\mathbb{Q}}^f}{\mathrm{d}\mathbb{P}} - \varphi(\widehat{Y}_T^c)\tfrac{\mathrm{d}\widehat{\mathbb{Q}}^c}{\mathrm{d}\mathbb{P}}\right|^p\right] \\
&\quad = \mathbb{E}\left[\left|\varphi(\widehat{Y}_T^f)\tfrac{\mathrm{d}\widehat{\mathbb{Q}}^f}{\mathrm{d}\mathbb{P}} - \varphi(\widehat{Y}_T^f) + \varphi(\widehat{Y}_T^f) - \varphi(\widehat{Y}_T^c) + \varphi(\widehat{Y}_T^c) - \varphi(\widehat{Y}_T^c)\tfrac{\mathrm{d}\widehat{\mathbb{Q}}^c}{\mathrm{d}\mathbb{P}}\right|^p\right] \\
&\quad \le 3^{p-1}\mathbb{E}\left[\left|\varphi(\widehat{Y}_T^f) - \varphi(\widehat{Y}_T^c)\right|^p\right] + 3^{p-1}\mathbb{E}\left[\left|\varphi(\widehat{Y}_T^f)\right|^{2p}\right]^{1/2}\mathbb{E}\left[\left|1 - \tfrac{\mathrm{d}\widehat{\mathbb{Q}}^f}{\mathrm{d}\mathbb{P}}\right|^{2p}\right]^{1/2} \\
&\qquad + 3^{p-1}\mathbb{E}\left[\left|\varphi(\widehat{Y}_T^c)\right|^{2p}\right]^{1/2}\mathbb{E}\left[\left|1 - \tfrac{\mathrm{d}\widehat{\mathbb{Q}}^c}{\mathrm{d}\mathbb{P}}\right|^{2p}\right]^{1/2}.
\end{aligned} \tag{4.119}$$

By Theorems 4.2.1 and 4.2.2 and the globally Lipschitz condition, there exists a constant $C_1 \in (0, \infty)$ such that

$$\begin{aligned}
\mathbb{E}\left[\left|\varphi(\widehat{Y}_T^f) - \varphi(\widehat{Y}_T^c)\right|^p\right] \le K^p\,\mathbb{E}\left[\left\|\widehat{Y}_T^f - \widehat{Y}_T^c\right\|^p\right] \lesssim C_1^p\, p^p\, h^p, \\
\mathbb{E}\left[\left|\varphi(\widehat{Y}_T^f)\right|^{2p}\right] \lesssim C_1^p\, p^p, \qquad \mathbb{E}\left[\left|\varphi(\widehat{Y}_T^c)\right|^{2p}\right] \lesssim C_1^p\, p^p.
\end{aligned} \tag{4.120}$$

Next we estimate $\mathbb{E}\left[\left|1 - \frac{\mathrm{d}\widehat{\mathbb{Q}}^c}{\mathrm{d}\mathbb{P}}\right|^p\right]$ with $\frac{\mathrm{d}\mathbb{Q}^c}{\mathrm{d}\mathbb{P}} = \exp(\mathcal{H})$ where

$$\mathcal{H} = -S\sum_{n=0}^{N/2-1}\left\langle \widehat{Y}_{2n}^f - \widehat{Y}_{2n}^c,\ \Delta W_{2n} + \Delta W_{2n+1}\right\rangle - \frac{S^2}{2}\sum_{n=0}^{N/2-1}\left\|\widehat{Y}_{2n}^f - \widehat{Y}_{2n}^c\right\|^2 2h. \tag{4.121}$$

Taylor expansion implies that

$$\mathrm{e}^x = 1 + \mathrm{e}^{\xi(x)}x, \text{ for some } \xi(x) \text{ with } |\xi(x)| < |x|, \tag{4.122}$$

which, combined with Hölder's inequality, implies that

$$\mathbb{E}\left[\left|1 - \frac{\mathrm{d}\widehat{\mathbb{Q}}^c}{\mathrm{d}\mathbb{P}}\right|^{2p}\right] = \mathbb{E}\left[\left(\exp(\xi(\mathcal{H}))\,|\mathcal{H}|\right)^{2p}\right] \le \mathbb{E}\left[\left|\exp(\xi(\mathcal{H}))\right|^{4p}\right]^{1/2}\mathbb{E}\left[|\mathcal{H}|^{4p}\right]^{1/2}. \tag{4.123}$$

First, by Theorem 4.2.3, we obtain that

$$\mathbb{E}\left[|\exp(\xi(\mathcal{H}))|^{4p}\right] \leq \mathbb{E}\left[\max(\exp(4p\mathcal{H}),1)\right] \leq \mathbb{E}\left[\exp(4p\mathcal{H})\right] + 1 \lesssim 3, \quad (4.124)$$

provided $h \in (0, C_2/(T\,p^2))$ for some constant $C_2 \in (0,\infty)$.

Second, by Jensen's inequality and the Burkholder-Davis-Gundy inequality in [5], there exists a constant $C_3 \in (0,\infty)$ such that

$$\begin{aligned}
\mathbb{E}\left[|\mathcal{H}|^{4p}\right] \;&\leq\; 2^{4p-1}S^{4p}\,\mathbb{E}\left[\left|\sum_{n=0}^{N/2-1}\left\langle \widehat{Y}^f_{2n}-\widehat{Y}^c_{2n},\;\Delta W_{2n}+\Delta W_{2n+1}\right\rangle\right|^{4p}\right] \\
&\qquad + 2^{4p-1}S^{8p}\,\mathbb{E}\left[\left|\sum_{n=0}^{N/2-1}\left\|\widehat{Y}^f_{2n}-\widehat{Y}^c_{2n}\right\|^2 h\right|^{4p}\right] \\
&\leq\; 2^{4p-1}S^{4p}C_3^p p^{2p}\mathbb{E}\left[\left|\sum_{n=0}^{N/2-1}\left\|\widehat{Y}^f_{2n}-\widehat{Y}^c_{2n}\right\|^2 2h\right|^{2p}\right] \\
&\qquad + 2^{4p-1}S^{8p}T^{4p-1}\mathbb{E}\left[\sum_{n=0}^{N/2-1}\left\|\widehat{Y}^f_{2n}-\widehat{Y}^c_{2n}\right\|^{8p} h\right] \\
&\leq\; 2^{6p-1}S^{4p}C_3^p p^{2p}T^{2p-1}\mathbb{E}\left[\sum_{n=0}^{N/2-1}\left\|\widehat{Y}^f_{2n}-\widehat{Y}^c_{2n}\right\|^{4p} h\right] \qquad (4.125) \\
&\qquad + 2^{4p-1}S^{8p}T^{4p-1}\mathbb{E}\left[\sum_{n=0}^{N/2-1}\left\|\widehat{Y}^f_{2n}-\widehat{Y}^c_{2n}\right\|^{8p} h\right].
\end{aligned}$$

Then, provided $h \in (0, C_4/\sqrt{Tp})$ for some constant $C_4 \in (0,\infty)$, by Theorem 4.2.2, there exists a constant $C_5 \in (0,\infty)$ such that

$$\mathbb{E}\left[|\mathcal{H}|^{4p}\right] \lesssim C_5^{4p}\,p^{6p}\,T^{2p}\,h^{4p}, \quad (4.126)$$

and then

$$\mathbb{E}\left[\left|1-\frac{\mathrm{d}\widehat{\mathbb{Q}}^c}{\mathrm{d}\mathbb{P}}\right|^{2p}\right] \lesssim 3^{1/2}\,C_5^{2p}\,p^{3p}\,T^p\,h^{2p}. \quad (4.127)$$

Similarly, we can get the same result for $\mathbb{E}\left[\left|1-\frac{\mathrm{d}\widehat{\mathbb{Q}}^f}{\mathrm{d}\mathbb{P}}\right|^{2p}\right]$.

Since $1/\sqrt{Tp} \geq 1/(\sqrt{T}p) \geq \min(1, 1/(Tp^2))$, the condition $h \in (0, C_4/\sqrt{Tp})$ can be replaced by $h(0, C_4 \min(1, 1/(Tp^2)))$.

Overall, combining all the estimates above, there exist constants $C_{(1)}, C_{(2)}, C_{(3)} \in (0, \infty)$ such that for any $h \in (0, \min(C_{(1)}, C_{(2)}/(Tp^2)))$ it holds that

$$\mathbb{E}\left[\left|\varphi(\widehat{Y}_T^f)\frac{\mathrm{d}\widehat{\mathbb{Q}}^f}{\mathrm{d}\mathbb{P}} - \varphi(\widehat{Y}_T^c)\frac{\mathrm{d}\widehat{\mathbb{Q}}^c}{\mathrm{d}\mathbb{P}}\right|^p\right] \leq C_{(3)}^p\, p^{2p}\, T^{p/2}\, h^p. \tag{4.128}$$

This completes the proof of Theorem 4.2.4. □

4.7 Conclusion

In this chapter, we considered a larger class of ergodic SDEs without contractivity and introduced a new MLMC scheme by adding a spring term to link the coarse and fine path approximations as is shown in section 4.1. We obtained the new MLMC identity with Radon-Nikodym derivatives. When computing the expectations with respect to the invariant measures of chaotic ergodic SDEs, which satisfy a one-sided Lipschitz condition, we reduced the exponential increase of the variance of the MLMC level estimator to a linear increase with respect to T.

Our numerical analysis only works for SDEs with globally Lipschitz drift. Theorem 4.2.1 shows the stability of the paths after adding a spring term. Uniform bounds for the strong error and the Radon-Nikodym derivative are established in Theorems 4.2.2 and 4.2.3, which implies the linear growth of the variances of the level estimators in Theorem 4.2.4. Then, the MLMC Theorem 4.2.5 shows that our new MLMC scheme does improve the accuracy and reduce the computational cost. The numerical experiments in section 4.4 demonstrate that the new MLMC with the proposed change of measure technique works well for the case with non-globally Lipschitz drift using adaptive timestepping.

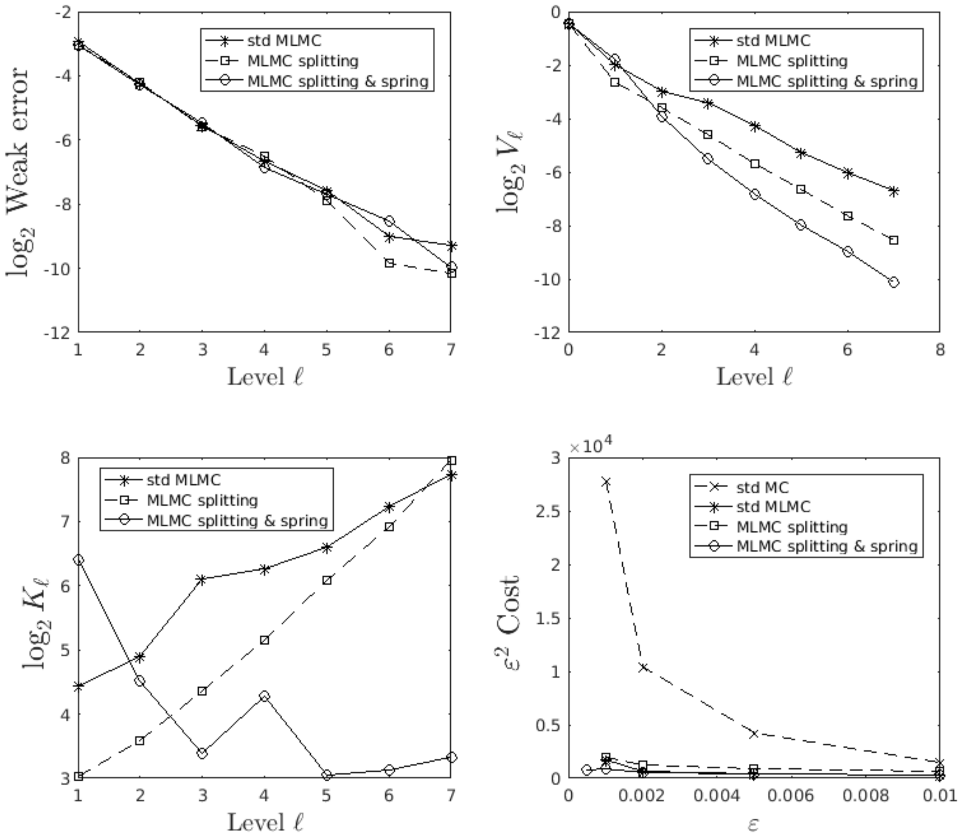


Figure 4.9: First exit time for the double-well potential energy SDE

Chapter 5

Sensitivity of chaotic system

This chapter is an extended version of an unpublished paper co-authered with my supervisor Prof. Mike Giles. In this chapter, we consider an m-dimensional chaotic SDE driven by an m-dimensional Brownian motion with parameter θ

$$\mathrm{d}X_t^{(\theta)} \;=\; f(\theta; X_t^{(\theta)})\,\mathrm{d}t \;+\sigma\,\mathrm{d}W_t, \tag{5.1}$$

which has a fixed initial datum ξ_0 and a locally Lipschitz drift $f:\mathbb{R}^m\to\mathbb{R}^m$ such that the strong solution exists uniquely and the SDE is geometric ergodic, that is,

$$\left|\mathbb{E}\left[\varphi(X_T^{(\theta)}) - \pi^{(\theta)}(\varphi)\right]\right| \le \mu\,\mathrm{e}^{-\lambda T} \tag{5.2}$$

for some constants $\mu,\lambda\in(0,\infty)$, where φ is a differentiable function with polynomial growth and $\pi(\varphi)$ is the expectation of φ with respect to the invariant measure of the SDE. For the computation of $\pi^{(\theta)}(\varphi)$, one common approach is to truncate the infinite time interval at a sufficiently large finite time T and estimate

$$F(\theta) \;=\; \mathbb{E}\left[\varphi(X_T^{(\theta)})\right]. \tag{5.3}$$

For the computation of the sensitivity of $\pi^{(\theta)}(\varphi)$ with respect to θ, Assaraf et al. [4] showed that

$$\frac{\partial\,\pi^{(\theta)}(\varphi)}{\partial\theta} = \frac{\partial}{\partial\theta}\left(\lim_{T\to\infty}\mathbb{E}\left[\varphi(X_T^{(\theta)})\right]\right) = \lim_{T\to\infty}\left(\frac{\partial}{\partial\theta}\mathbb{E}\left[\varphi(X_T^{(\theta)})\right]\right) \tag{5.4}$$

under some restrictive conditions. The exchangeability of the limit and the differentiation for general ergodic SDEs is unknown, and in this chapter we assume that this holds and the convergence speed is still exponentially fast, i.e.,

$$\left| \frac{\partial}{\partial \theta} \mathbb{E}\left[\varphi(X_T^{(\theta)})\right] - \frac{\partial\, \pi^{(\theta)}(\varphi)}{\partial \theta} \right| \leq \mu^* \, \mathrm{e}^{-\lambda^* T} \tag{5.5}$$

for some constants $\mu^*, \lambda^* \in (0, \infty)$. Therefore we approximate the sensitivity of the $\pi^{(\theta)}(\varphi)$ by $\partial F / \partial \theta$ with a sufficiently large T.

There are several approaches to computing sensitivities of SDEs [36]: finite difference (FD) method, likelihood ratio method (LR [37]), pathwise sensitivity calculation (PS), the weak derivative method (WD [83]) and Malliavin calculus (MA [26]). The FD estimator is straightforward to implement. For example, we can use the central difference $(F(\theta + \varepsilon/2) - F(\theta - \varepsilon/2))/\varepsilon$ to approximate the first-order derivative. Using the same Brownian paths for both $F(\theta + \varepsilon/2)$ and $F(\theta - \varepsilon/2))$ greatly reduces the variance, but the estimator is biased. Both LR and PS estimators are unbiased. The difference is in the choice of the sample space Ω to perform the differentiation [63]. By choosing the discrete numerical solution paths as ω, differentiating their log joint distribution gives the LR estimator. However, the variance of the LR estimator in most cases is $O(h^{-1})$ as $h \to 0$, where h is the timestep size of the numerical approximation. Again, a tradeoff between bias and variance is needed. The PS estimator, based on choosing the independent $U(0, 1)$ variates as the sample space, is more popular due to its low variance and applicability of the adjoint method, which improves the efficiency when calculating many sensitivities [34]. It requires differentiating the original SDE (5.1) to obtain the variation process. In our case, the plain PS estimator suffers from the problem that the variation process blows up exponentially. Similarly to LR, the WD method differentiates the probability measure but it may require resimulations, which increases the computational cost greatly. Lastly, Malliavin calculus estimators can be viewed as a combination of the LR and PS methods [17], but are less popular in practice due to the heavy machinery of Malliavin calculus. However, we will show later that in our case, the Malliavin calculus provides a comparable estimator.

Another issue for the computation of sensitivities is the extension to functions φ with discontinuity, which is out of our scope in this chapter but may provide possible future research directions. LR, WD and MA all cope with this difficulty but FD and PS suffer. Giles [31] proposed the vibrato Monte Carlo method, which can be viewed as a combination of PS and LR for the final timestep. Chan & Joshi [14] proposed partial proxy schemes by approximately shifting the means of the standard normal variables that govern the system. Tong & Liu [98] derived a new estimator for discontinuous φ and used importance sampling to make all the samples fall into the same set to avoid the discontinuity.

In this chapter, we first derive the Malliavin estimator for the chaotic SDEs following the idea in [26], since it expresses the sensitivity by the process itself and avoids the variation process. However, it only works for the sensitivity of the drift parameter and fails for others since the estimators again involve the variation process. The other drawback is that it is difficult to apply the adjoint method. Therefore, we propose a new PS estimator with importance sampling. Similar to the idea of introducing a spring term between fine and coarse paths in previous chapter, we add a spring term between $X_t^{(\theta)}$ and $X_t^{(\theta+\varepsilon)}$ and derive the new PS estimator with the Radon-Nikodym derivative and take the differentiation limit $\varepsilon \to 0$. The benefit of this change is that the variation process with spring term becomes ergodic again, which allows the computation of sensitivities of the volatility parameter and initial condition. The variances of both Malliavin and the new PS estimators increase only linearly in time T which is a great improvement compared with the exponential increase of the plain PS estimator. Next, we apply the MLMC method [30, 32] to improve the efficiency further. Since the estimator involves the original ergodic SDE which is not contractive, we need to employ the change of measure technique from the previous chapter to add another spring term between the fine and coarse paths for level estimators. Then, we can use the same MLMC scheme with the same random samples and Radon-Nikodym derivatives to calculate the original value $F(\theta)$ and its derivative simultaneously. Finally, we consider the approximation of the sensitivities of chaotic ODEs by the stochastic version with small volatility σ.

The rest of the chapter is organised as follows. Section 2 reviews the pathwise sensitivity method and explains the problem of the plain PS estimator. Section 3 derives the Malliavin estimator and the new pathwise sensitivity method with importance sampling is introduced in Section 4. Numerical results are shown in section 5. The applications of MLMC schemes are presented in section 6. Finally, section 7 concludes this chapter.

To focus on the computation of the sensitivity, we assume the existence and strong convergence of a suitable numerical scheme.

Assumption 5.0.1 (Numerical Solution)**.** *For the ergodic SDE* (5.1)*, we assume that there exists a numerical solution* $\widehat{X}_t$*, whose moments are uniformly bounded with respect to time* T*, that is, there exist constants* $C_p, c_p \in (0,\infty)$ *independent of* T *such that for any* $p \in [1,\infty), T \in (0,\infty)$*, it holds that*

$$\sup_{0\le t\le T} \mathbb{E}\left[\|\widehat{X}_t\|^p\right] \le C_p, \tag{5.6}$$

and the strong error is also uniformly bounded

$$\sup_{0\le t\le T} \mathbb{E}\left[\|\widehat{X}_t - X_t\|^p\right] \le c_p\, h^p, \tag{5.7}$$

where h *is the average timestep size.*

For simplicity, for rest of the chapter, we denote the discrete points of a numerical solution by $\widehat{X}_{t_n}$ for $n \in \{0,1,2...,N\}$, and $t_{n+1} = t_n + h_n$, which is suitable for both uniform timestep and adaptive schemes. For the numerical experiments on the stochastic Lorenz equation, we use the adaptive Euler-Maruyama method proposed in previous chapters, in which the uniform bounds for moments and the strong error are achieved for a class of ergodic SDEs.

5.1 Pathwise sensitivity

Pathwise sensitivity is a standard approach also known as Infinitesimal Perturbation Analysis (IPA) [63]. Assuming the exchangeability of the integral and the differentiation. (Theorem 1 in [63] provides a practical and general sufficient validity condition for the exchange of derivative and integral. Fortunately, these conditions are satisfied in most cases.) By Leibniz's integral rule and chain rule, we obtain from (5.3) that

$$\frac{\partial F(\theta)}{\partial \theta} \;=\; \mathbb{E}\left[\frac{\partial \varphi(X_T^{(\theta)})}{\partial \theta}\right] \;=\; \mathbb{E}\left[\left\langle \nabla\varphi(X_T^{(\theta)}),\; \frac{\partial X_T^{(\theta)}}{\partial \theta}\right\rangle\right]. \tag{5.8}$$

For $x_T^{(\theta)} = \frac{\partial X_T^{(\theta)}}{\partial \theta}$, again assuming the exchangeability of the integral and the differentiation (guaranteed by Theorem 39 in chapter 7 in [84]), by Leibniz's integral rule, we have the variation process

$$\mathrm{d}\, x_t^{(\theta)} = \left(\frac{\partial f(\theta; X_t^{(\theta)})}{\partial \theta} + \frac{\partial f(\theta; X_t^{(\theta)})}{\partial X_t^{(\theta)}} x_t^{(\theta)}\right) \mathrm{d}t. \tag{5.9}$$

Numerically, we can simulate $X_T^{(\theta)}$ and $x_T^{(\theta)}$ simultaneously:

$$\begin{aligned} \widehat{X}_{t_{n+1}}^{(\theta)} &= \widehat{X}_{t_n}^{(\theta)} + f(\theta; \widehat{X}_{t_n}^{(\theta)})\, h_n + \sigma\, \Delta W_n, \\ \widehat{x}_{t_{n+1}}^{(\theta)} &= \widehat{x}_{t_n}^{(\theta)} + \left(\frac{\partial f(\theta; \widehat{X}_{t_n}^{(\theta)})}{\partial \theta} + \frac{\partial f(\theta; \widehat{X}_{t_n}^{(\theta)})}{\partial \widehat{X}_{t_n}^{(\theta)}}\, \widehat{x}_{t_n}^{(\theta)}\right) h_n. \end{aligned} \tag{5.10}$$

with initial conditions $\widehat{X}_0^{(\theta)} = \xi_0$ and $\widehat{x}_0 = 0$. We then have the standard PS estimator

$$\frac{\partial \widehat{F}}{\partial \theta} \;=\; \mathbb{E}\left[\left\langle \nabla\varphi(\widehat{X}_T^{(\theta)}),\; \widehat{x}_T^{(\theta)}\right\rangle\right], \tag{5.11}$$

which is the unbiased estimator of the sensitivity of $\mathbb{E}\left[\varphi(\widehat{X}_T^{(\theta)})\right]$ with respect to θ.

The original process $X_t^{(\theta)}$ is ergodic and under suitable conditions, the moments of the numerical solution $X_{t_n}^{(\theta)}$ are uniformly bounded. However, for the variation process x_t^{θ} of the chaotic system, it may be no longer ergodic and the moments may increase exponentially in T. As a result, we have the following theorem.

Theorem 5.1.1 (Standard Pathwise)**.** *If the moments of the original variation process x_t increase exponentially with respect to time t, then for any numerical solution $\widehat{x}_t$ with first order strong convergence, the variance of the standard PS* *estimator $\frac{\partial \widehat{F}}{\partial \theta}$ also increases at most exponentially, that is*

$$\mathbb{V}\left[\frac{\partial \widehat{F}}{\partial \theta}\right] \leq \eta \, \mathrm{e}^{\kappa' T}, \tag{5.12}$$

for some constant $\eta, \kappa' \in (0, \infty)$. The constant κ' is the Lyapunov exponent of the variation process.

Proof of Theorem 5.1.1. Hölder's inequality implies that

$$\mathbb{V}\left[\frac{\partial \widehat{F}}{\partial \theta}\right] \leq \mathbb{E}\left[\left|\frac{\partial \widehat{F}}{\partial \theta}\right|^2\right] \leq \mathbb{E}\left[\left\|\nabla \varphi(\widehat{X}_T^{(\theta)})\right\|^4\right]^{1/2} \mathbb{E}\left[\left\|\widehat{x}_T^{(\theta)}\right\|^4\right]^{1/2}. \tag{5.13}$$

The first expectation on the right hand side is bounded uniformly in T due to the Assumption 5.0.1 and polynomial growth of φ. For the second one, the strong convergence of $\widehat{x}_t$ implies

$$\mathbb{E}\left[\left\|\widehat{x}_T^{(\theta)}\right\|^4\right] \leq 8\,\mathbb{E}\left[\left\|x_T^{(\theta)}\right\|^4\right] + \lambda, \tag{5.14}$$

for some constant $\lambda \in (0, \infty)$. Then combining this and the exponential increase of the $\mathbb{E}\left[\left\|x_T^{(\theta)}\right\|^4\right]$ completes the proof of Theorem 5.1.1. □

If we consider the sensitivity of the outputs related to the invariant measure, we need to choose $T \sim \frac{1}{\lambda^*} |\log \varepsilon|$ due to (5.5) to bound the truncation error, which together with Theorem 5.1.1 implies that

$$\mathbb{V}\left[\frac{\partial \widehat{F}}{\partial \theta}\right] \sim \eta \, \varepsilon^{-\kappa'/\lambda^*}. \tag{5.15}$$

Therefore, in order to achieve the ε^2 MSE, we need to simulate $O(\varepsilon^{-2-\kappa'/\lambda^*})$ paths and for each path, we need $h_n = O(\varepsilon)$ to bound the weak error and $T = O(|\log \varepsilon|)$. The total computational cost of the standard Monte Carlo method becomes $O(\varepsilon^{-3-\kappa'/\lambda^*} |\log \varepsilon|)$.

5.2 Malliavin estimator

Following the idea in [26], we derive the sensitivity estimator using Malliavin calculus. (The derivation is similar to the proof in Proposition 3.1 in [26], we include it here for clarity and simplicity of this case and it is also similar to our approach to derive the new PS estimator.) Without loss of generality, we assume we are calculating the sensitivity at $\theta = 0$. Under the measure $\mathbb{P}$, the original SDE with $\theta = 0$ is

$$\mathrm{d}X_t^{(0)} = f(X_t^{(0)})\,\mathrm{d}t + \sigma\,\mathrm{d}W_t^{\mathbb{P}} \tag{5.16}$$

and the perturbed SDE is

$$\mathrm{d}X_t^{(\theta)} = \left[f(X_t^{(\theta)}) + \theta\,\gamma(X_t^{(\theta)})\right]\,\mathrm{d}t + \sigma\,\mathrm{d}W_t^{\mathbb{P}}, \tag{5.17}$$

where $\gamma :\ \mathbb{R}^m \to \mathbb{R}^m$ is the perturbation term. We then consider the expectations

$$F(0) = \mathbb{E}^{\mathbb{P}}\left[\varphi(X_T^{(0)})\right],\quad F(\theta) = \mathbb{E}^{\mathbb{P}}\left[\varphi(X_T^{(\theta)})\right], \tag{5.18}$$

and the derivative by taking the limit

$$\left.\frac{\partial F(\theta)}{\partial\theta}\right|_{\theta=0} = \lim_{\theta\to 0}\frac{F(\theta)-F(0)}{\theta}. \tag{5.19}$$

Considering a new measure $\mathbb{Q}$ with

$$\mathrm{d}W_t^{\mathbb{Q}} = \theta\,\gamma(X_t^{(\theta)})/\sigma\,\mathrm{d}t + \mathrm{d}W_t^{\mathbb{P}} \tag{5.20}$$

being the standard Brownian motion, we obtain that

$$\mathrm{d}X_t^{(\theta)} = f(X_t^{(\theta)})\,\mathrm{d}t + \sigma\,\mathrm{d}W_t^{\mathbb{Q}}, \tag{5.21}$$

under the measure $\mathbb{Q}$. Therefore, Girsanov's theorem ensures that

$$F(\theta) = \mathbb{E}^{\mathbb{Q}}\left[\varphi(X_T^{(\theta)})R_T(\theta)\right], \tag{5.22}$$

where $R_T(\theta)$ is the Radon-Nikodym derivative

$$R_t(\theta) = \frac{\mathrm{d}\mathbb{P}}{\mathrm{d}\mathbb{Q}} = \exp\left(\int_0^t \frac{\theta}{\sigma}\left\langle\gamma(X_u^{(\theta)}),\,\mathrm{d}W_u^{\mathbb{Q}}\right\rangle - \frac{1}{2}\int_0^t \frac{\theta^2}{\sigma^2}\left\|\gamma(X_u^{(\theta)})\right\|^2\,\mathrm{d}u\right). \tag{5.23}$$

Since $(X_t^{(\theta)}, W_t^{\mathbb{Q}})$ under the measure $\mathbb{Q}$ has the same joint distribution as $(X_t^{(0)}, W_t^{\mathbb{P}})$ under the measure $\mathbb{P}$, we obtain that

$$F(\theta) = \mathbb{E}^{\mathbb{P}}\left[\varphi(X_T^{(0)}) R_T(\theta)\right], \tag{5.24}$$

with

$$R_t(\theta) = \exp\left(\int_0^t \frac{\theta}{\sigma}\left\langle\gamma(X_u^{(0)}),\, \mathrm{d}W_u^{\mathbb{P}}\right\rangle - \frac{1}{2}\int_0^t \frac{\theta^2}{\sigma^2}\left\|\gamma(X_u^{(0)})\right\|^2 \mathrm{d}u\right). \tag{5.25}$$

By the definition of the exponential martingale and Itô formula, we obtain that

$$\frac{R_T(\theta)-1}{\theta} = \int_0^T \frac{1}{\sigma} R_t(\theta)\left\langle\gamma(X_t^{(0)}),\, \mathrm{d}W_t^{\mathbb{P}}\right\rangle, \tag{5.26}$$

and as $\theta \to 0$, it holds that

$$\frac{R_T(\theta)-1}{\theta} \to \int_0^T \frac{1}{\sigma}\left\langle\gamma(X_t^{(0)}),\, \mathrm{d}W_t^{\mathbb{P}}\right\rangle \quad \textbf{in } L^2. \tag{5.27}$$

Finally, we take the limit $\theta \to 0$ to obtain that

$$\begin{aligned} \frac{F(\theta)-F(0)}{\theta} &= \mathbb{E}\left[\varphi(X_T^{(0)})\frac{R_T(\theta)-1}{\theta}\right] \\ &\to \mathbb{E}\left[\varphi(X_T^{(0)})\int_0^T\left\langle\frac{\gamma(X_t^{(0)})}{\sigma},\ \mathrm{d}W_t^{\mathbb{P}}\right\rangle\right], \end{aligned} \tag{5.28}$$

and obtain the Malliavin estimator

$$\left.\frac{\partial F(\theta)}{\partial\theta}\right|_{\theta=0} = \mathbb{E}\left[\varphi(X_T^{(0)})\int_0^T\left\langle\frac{\gamma(X_t^{(0)})}{\sigma},\ \mathrm{d}W_t^{\mathbb{P}}\right\rangle\right], \tag{5.29}$$

and numerically

$$\frac{\partial\widetilde{F}}{\partial\theta} = \varphi(\widehat{X}_T)\sum_{n=0}^{N-1}\left\langle\frac{\gamma(\widehat{X}_{t_n})}{\sigma},\ \Delta W_n^{\mathbb{P}}\right\rangle. \tag{5.30}$$

The benefit of this approach is the avoidance of the variation process and then the variance of the Malliavin estimator increases only linearly in T, as shown in the following theorem.

Theorem 5.2.1 (Malliavin Estimator)**.** *If $\widehat{X}_t$ satisfies Assumption 5.0.1 and the perturbation term $\|\gamma(x)\|$ increases at most polynomially when $\|x\|$ is large, the variance of the Malliavin estimator increases at most linearly with respect to T, that is*

$$\mathbb{V}\left[\frac{\partial \widetilde{F}}{\partial \theta}\right] \;\leq\; \frac{\eta'}{\sigma^2} T, \tag{5.31}$$

for some constant $\eta \in (0, \infty)$.

Proof of Theorem 5.2.1. By Hölder's , the Burkholder–Davis–Gundy and Jensen's inequalities and the uniform moments in Assumption 5.0.1, we obtain that

$$\begin{aligned}\mathbb{V}\left[\frac{\partial \widetilde{F}}{\partial \theta}\right] \;&\leq\; \mathbb{E}\left[\left|\frac{\partial \widetilde{F}}{\partial \theta}\right|^2\right] \leq \mathbb{E}\left[\left|\varphi(\widehat{X}_T)\right|^4\right]^{1/2} \mathbb{E}\left[\left|\sum_{n=0}^{N-1}\left\langle \frac{\gamma(\widehat{X}_{t_n})}{\sigma},\; \Delta W_n^{\mathbb{P}}\right\rangle\right|^4\right]^{1/2} \\ &\leq\; \lambda' \, \mathbb{E}\left[\left|\varphi(\widehat{X}_T)\right|^4\right]^{1/2} \mathbb{E}\left[\left|\sum_{n=0}^{N-1} \frac{\|\gamma(\widehat{X}_{t_n})\|^2}{\sigma^2} h_n\right|^2\right]^{1/2} \leq \frac{\eta'}{\sigma^2} T,\end{aligned} \tag{5.32}$$

for some constant λ', $\eta' \in (0, \infty)$. This completes the proof of Theorem 5.2.1. □

Choosing $T \sim \frac{1}{\lambda^*} |\log \varepsilon|$ implies that

$$\mathbb{V}\left[\frac{\partial \widetilde{F}}{\partial \theta}\right] \sim \frac{\eta'}{\lambda^* \sigma^2} \; |\log \varepsilon| \, . \tag{5.33}$$

Therefore, in order to achieve the ε^2 MSE, we need to run $O(\varepsilon^{-2}|\log \varepsilon|)$ paths and the total computational cost of the standard Monte Carlo method is $O(\varepsilon^{-3}|\log \varepsilon|^2)$, which is a great improvement over the standard PS method. However, this benefit is limited to the sensitivity with respect to the drift parameters. For the sensitivity with respect to the volatility parameters and the initial condition, the Malliavin estimators again involve the variation process so that the variance of the estimator increases exponentially. Another consideration is the application of the adjoint technique. Since we have to derive different Malliavin estimators for different parameters, we are unable to apply the adjoint technique and then the computational cost increases linearly with the number of sensitivities.

5.3 Importance sampling for sensitivity

In this section, we consider how to use the importance sampling technique to derive the new PS estimator. The main idea is that by introducing a spring term between the original and perturbed SDEs, we obtain a new variation process which is ergodic again. Without loss of generality, we assume we are calculating the sensitivity at $\theta = 0$.

Let $X_t^{(0)}$ be the solution to equation (5.16) when $\theta = 0$, and consider the new perturbed SDE under the measure $\tilde{\mathbb{Q}}$,

$$\mathrm{d}Y_t^{(\theta)} = \left(f(\theta; Y_t^{(\theta)}) + S\left(Y_t^{(0)} - Y_t^{(\theta)} \right) \right) \mathrm{d}t + \sigma\, \mathrm{d}W_t^{\tilde{\mathbb{Q}}}, \tag{5.34}$$

with

$$\mathrm{d}W_t^{\mathbb{P}} = S\left(Y_t^{(0)} - Y_t^{(\theta)} \right) / \sigma\, \mathrm{d}t + \mathrm{d}W_t^{\tilde{\mathbb{Q}}}, \tag{5.35}$$

which introduces a spring term with sufficiently large $S \in (0, \infty)$ to prevent $\left\| Y_t^{(0)} - Y_t^{(\theta)} \right\|$ from increasing exponentially. Note that $Y_t^{(\theta)} = X_t^{(0)}$ and $W_t^{\tilde{\mathbb{Q}}} = W_t^{\mathbb{P}}$ when $\theta = 0$. Girsanov's theorem implies that

$$F(\theta) = \mathbb{E}^{\tilde{\mathbb{Q}}} \left[\varphi(Y_T^{(\theta)}) R_T(\theta) \right], \tag{5.36}$$

where $R_t(\theta)$ is corresponding Radon-Nikodym derivative

$$\begin{aligned} R_t(\theta) &= \frac{\mathrm{d}\mathbb{P}}{\mathrm{d}\tilde{\mathbb{Q}}} \\ &= \exp\left(-\int_0^t \frac{S}{\sigma} \left\langle Y_u^{(0)} - Y_u^{(\theta)}, \mathrm{d}W_u^{\tilde{\mathbb{Q}}} \right\rangle - \frac{1}{2} \int_0^t \frac{S^2}{\sigma^2} \left\| Y_u^{(0)} - Y_u^{(\theta)} \right\|^2 \mathrm{d}u \right). \end{aligned} \tag{5.37}$$

By Leibniz's integral rule, we obtain that

$$\frac{\partial F(\theta)}{\partial \theta} = \mathbb{E}^{\tilde{\mathbb{Q}}} \left[\frac{\partial \varphi(Y_T^{(\theta)})}{\partial \theta} R_T(\theta) + \varphi(Y_T^{(\theta)}) \frac{\partial R_T(\theta)}{\partial \theta} \right], \tag{5.38}$$

and the variation process $y_t^{(\theta)} = \partial Y_t^{(\theta)} / \partial \theta$ satisfies that

$$\mathrm{d}\, y_t^{(\theta)} = \left(\frac{\partial f(\theta; Y_t^{(\theta)})}{\partial \theta} + \left(\frac{\partial f(\theta; Y_t^{(\theta)})}{\partial Y_t^{(\theta)}} - SI \right) y_t^{(\theta)} \right) \mathrm{d}t, \tag{5.39}$$

where I is the m-dimensional identity matrix, and

$$\frac{\partial R_t(\theta)}{\partial\theta} = R_t(\theta)\left(\int_0^t \frac{S}{\sigma}\left\langle y_u^{(\theta)},\, \mathrm{d}W_u^{\tilde{\mathbb{Q}}}\right\rangle + \int_0^t \frac{S^2}{\sigma^2}\left\langle Y_u^{(0)} - Y_u^{(\theta)}, y_u^{(\theta)}\right\rangle \mathrm{d}u\right). \quad (5.40)$$

Therefore, taking $\theta = 0$ and again under the measure $\mathbb{P}$, it holds that

$$\left.\frac{\partial F(\theta)}{\partial\theta}\right|_{\theta=0} = \mathbb{E}^{\mathbb{P}}\left[\langle\nabla\varphi(Y_T^{(0)}),\, y_T^{(0)}\rangle + \varphi(Y_T^{(0)})\int_0^T \frac{S}{\sigma}\left\langle y_t^{(0)},\, \mathrm{d}W_t^{\mathbb{P}}\right\rangle\right]. \quad (5.41)$$

Numerically, we only need to simulate $Y_t^{(0)}$ and $y_t^{(0)}$:

$$\begin{aligned} \mathrm{d}\,Y_t^{(0)} &= f(0; Y_t^{(0)})\ \mathrm{d}t\ + \sigma\ \mathrm{d}W_t^{\mathbb{P}}, \\ \mathrm{d}\,y_t^{(0)} &= \left(\left.\frac{\partial f(\theta; Y_t^{(0)})}{\partial\theta}\right|_{\theta=0} + \left(\left.\frac{\partial f(\theta; Y_t^{(\theta)})}{\partial Y_t^{(\theta)}}\right|_{\theta=0} - SI\right) y_t^{(0)}\right)\mathrm{d}t, \end{aligned} \quad (5.42)$$

and the new PS estimator is

$$\frac{\partial\breve{F}}{\partial\theta} = \left\langle\nabla\varphi(\widehat{Y}_T^{(0)}),\, \widehat{y}_T^{(0)}\right\rangle + \varphi(\widehat{Y}_T^{(0)})\sum_{n=0}^{N-1}\frac{S}{\sigma}\left\langle\widehat{y}_{t_n}^{(0)},\, \Delta W_n^{\mathbb{P}}\right\rangle. \quad (5.43)$$

Note that we still simulate the same original process under the measure $\mathbb{P}$ with $Y_t^{(0)} = X_t$ but with a new variation process $y_t^{(0)} \neq x_t$. Comparing the two variation processes (5.9) and (5.42), the new one has an additional linear term, which makes the process become ergodic again for a sufficiently large $S > 0$. By Assumption 5.0.1, it is possible to find a suitable numerical scheme such that the moments and the strong error of $\widehat{y}_t$ are uniformly bounded. Then, similarly to the Malliavin estimator, the variance of the new estimator $\frac{\partial\breve{F}}{\partial\theta}$ increases linearly in T, as is shown in the following theorem.

Theorem 5.3.1 (Importance Sampling)**.** *Suppose $\widehat{Y}_t$ and $\widehat{y}_t$ satisfy the Assumption 5.0.1, the variance of the new pathwise estimator with importance sampling increases at most linearly with respect to T, that is,*

$$\mathbb{V}\left[\frac{\partial\breve{F}}{\partial\theta}\right] \leq \frac{\eta''}{\sigma^2}T, \quad (5.44)$$

for some constant $\eta'' \in (0,\infty)$.

Proof of Theorem 5.3.1. Using the same approach as in the proof of Theorem 5.2.1 completes the proof of Theorem 5.3.1. □

Choosing $T \sim \frac{1}{\lambda^*}|\log \varepsilon|$ implies that

$$\mathbb{V}\left[\frac{\partial \widetilde{F}}{\partial \theta}\right] \sim \frac{\eta''}{\lambda^* \sigma}\,|\log \varepsilon|. \tag{5.45}$$

Therefore, in order to achieve the ε^2 MSE, we need to run $O(\varepsilon^{-2}|\log \varepsilon|)$ paths and the total computational cost of the standard Monte Carlo method is $O(\varepsilon^{-3}|\log \varepsilon|^2)$, which is of the same order as for the Malliavin estimators.

However, one of the additional benefits of this approach is that fundamentally it changes the property of the variation process and we can easily extend it to the computation of other sensitivities with respect to volatility parameters and the initial condition.

The other benefit is that the pathwise sensitivities fit into the structure of the adjoint technique naturally which calculates all the sensitivities with a fixed upper bound for the computational cost (up to a factor 4 of the cost for the original SDE) [34].

5.4 Numerical results

In this section, we present some numerical results for the stochastic Lorenz equation

$$\mathrm{d}\,X_t \;=\; \begin{pmatrix} 10(X_t^2 - X_t^1) \\ X_t^1(\theta - X_t^3) - X_t^2 \\ X_t^1 X_t^2 - \frac{8}{3}X_t^3 \end{pmatrix} \mathrm{d}t + \sigma\ \mathrm{d}W_t^{\mathbb{P}}, \tag{5.46}$$

with initial condition $x_0 = (-2.4, -3.7, 14.98)^T$ and $\sigma = 6$. We estimate the sensitivity of $\mathbb{E}\left[X_T^3\right]$ with respect to θ at $\theta = 28$.

For the standard pathwise sensitivity method, we directly derive the variation process following (5.9):

$$\mathrm{d}x_t = \left(\begin{pmatrix} 0 \\ X_t^1 \\ 0 \end{pmatrix} + \begin{pmatrix} -10 & 10 & \\ 28 - X_t^3 & -1 & -X_t^1 \\ X_t^2 & X_t^1 & -\frac{8}{3} \end{pmatrix} x_t \right) \mathrm{d}t \tag{5.47}$$

and the standard PS estimator

$$\frac{\partial \widehat{F}}{\partial \theta} = \widehat{x}_T^3. \tag{5.48}$$

Numerically, we set $T = 20$ and $\sigma = 6$ using the same adaptive timestep function as in previous chapter with $\delta = 2^{-9}$ and simulate $N = 4 \times 10^7$ paths. Then we plot the log variance of the estimator $\frac{\partial \widehat{F}}{\partial \theta}$ with respect to T in Figure 5.1. It shows clearly that the log variance increases linearly in T, which implies that the variance itself increases exponentially. The slope we fitted is 3.37 which is an approximation of κ'/λ^* and indicates that the computational cost the standard Monte Carlo method using standard PS estimator to achieve $O(\varepsilon^2)$ MSE becomes $O(\varepsilon^{-5.37}|\log \varepsilon|)$.

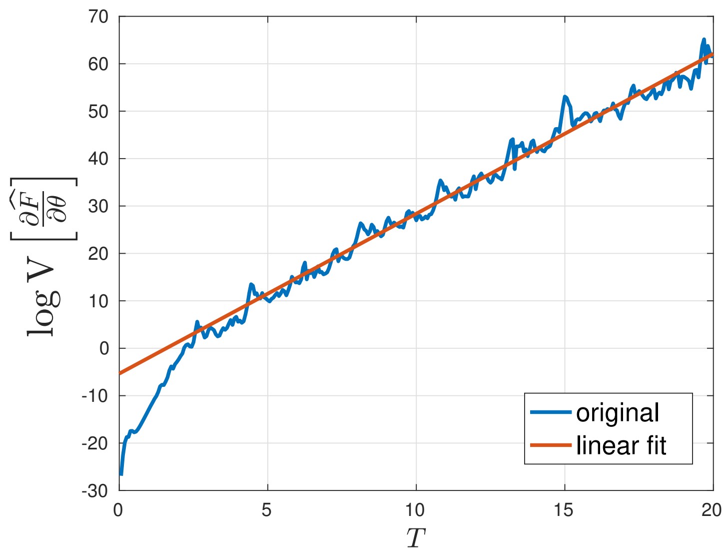


Figure 5.1: Exponential increase of the variance of the standard PS estimator

Next, following (5.29) and (5.30), we have $\gamma(X_t) = (0, X_t^1, 0)^T$ and get the Malliavin estimation

$$\frac{\partial F(\theta)}{\partial \theta} = \mathbb{E}^{\mathbb{P}}\left[X_T^3 \int_0^T \frac{1}{\sigma} X_t^1 \, \mathrm{d}W_t^{\mathbb{P},2}\right], \tag{5.49}$$

and the Malliavin estimator

$$\frac{\partial \widetilde{F}}{\partial \theta} = \widehat{X}_T^3 \sum_{n=0}^{N-1} \widehat{X}_{t_n}^1 \Delta W_n^{\mathbb{P},2}. \tag{5.50}$$

Finally, following (5.42), we obtain the new SDEs with importance sampling,

$$\mathrm{d}Y_t \;=\; \begin{pmatrix} 10(Y_t^2 - Y_t^1) \\ Y_t^1(28 - Y_t^3) - Y_t^2 \\ Y_t^1 Y_t^2 - \frac{8}{3} Y_t^3 \end{pmatrix} \mathrm{d}t + \sigma \, \mathrm{d}W_t^{\mathbb{P}}, \tag{5.51}$$

and

$$\mathrm{d}y_t \;=\; \left(\begin{pmatrix} 0 \\ Y_t^1 \\ 0 \end{pmatrix} + \begin{pmatrix} -10 - S & 10 & \\ 28 - Y_t^3 & -1 - S & -Y_t^1 \\ Y_t^2 & Y_t^1 & -\frac{8}{3} - S \end{pmatrix} y_t \right) \mathrm{d}t, \tag{5.52}$$

for some $S > 0$. We get the new PS expression

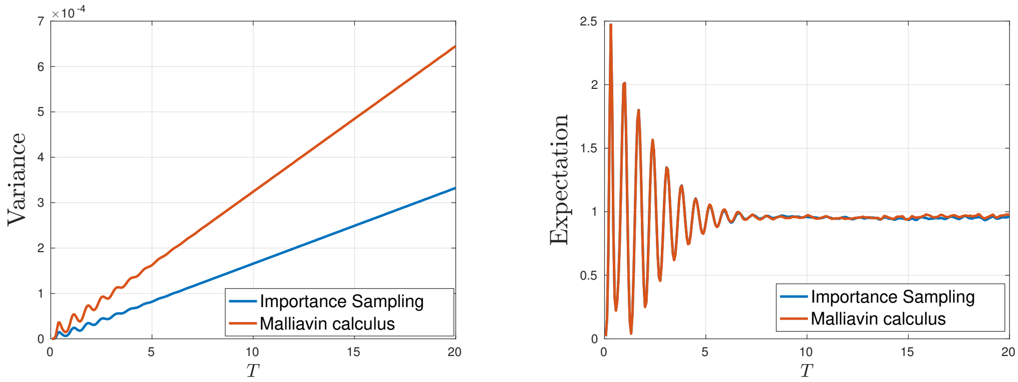
$$\frac{\partial F(\theta)}{\partial \theta} \;=\; \mathbb{E}^{\mathbb{P}}\left[y_T^3 + Y_T^3 \int_0^T \frac{S}{\sigma} \left\langle y_t, \, \mathrm{d}W_t^{\mathbb{P}} \right\rangle \right], \tag{5.53}$$

and the new PS estimator

$$\frac{\partial \breve{F}}{\partial \theta} = \widehat{y}_T^3 + \widehat{Y}_T^3 \sum_{n=0}^{N-1} \frac{S}{\sigma} \left\langle \widehat{y}_{t_n}, \, \Delta W_n^{\mathbb{P}} \right\rangle. \tag{5.54}$$

Numerically, we again set the same T, h, σ and N as above and choose $S = 10$. Then we plot the variances of the Mallianvin estimator and the new PS estimator with respect to T in Figure 5.2(a).

Figure 5.2(a) shows clearly the variances of both the Malliavin and the new PS estimator increase linearly in T, but the variance of the new PS estimator has a smaller constant factor and is slightly more efficient. Figure 5.2(b) shows that both methods give similar estimates and confirms that the estimators are ergodic and converge to the sensitivity under the invariant measure.



(a) Comparison of Variances (b) Comparison of Expectations

Figure 5.2: Comparison of Malliavin and new PS estimators

5.5 MLMC for sensitivity

In this section, we extend both the Malliavin and the new PS estimators to MLMC [30, 32]. For ergodic SDEs without contractivity, we need to employ the change of measure technique introduced in the previous chapter to add a spring term between the fine and coarse paths, since otherwise the fine and coarse paths may diverge exponentially due to the chaotic property. We first quickly review this technique and then present the numerical results.

5.5.1 MLMC with change of measure

Instead of considering the fine and coarse paths of the original SDEs under the same measure $\mathbb{P}$,

$$\begin{aligned} \mathrm{d}X_t^f &= f(X_t^f)\,\mathrm{d}t + \sigma\,\mathrm{d}W_t^{\mathbb{P}}, \\ \mathrm{d}X_t^c &= f(X_t^c)\,\mathrm{d}t + \sigma\,\mathrm{d}W_t^{\mathbb{P}}, \end{aligned} \tag{5.55}$$

we add a spring term with spring coefficient $S > 0$ for both the fine path and the coarse path and consider both paths under different measures

$$\begin{aligned}\mathbb{Q}^f: \quad \mathrm{d}Y_t^f &= f(Y_t^f)\,\mathrm{d}t + \sigma\,\mathrm{d}W_t^{\mathbb{Q}^f},\\ \mathbb{Q}^c: \quad \mathrm{d}Y_t^c &= f(Y_t^c)\,\mathrm{d}t + \sigma\,\mathrm{d}W_t^{\mathbb{Q}^c},\end{aligned} \tag{5.56}$$

with

$$\begin{aligned}\mathrm{d}W_t^{\mathbb{Q}^f} &= \frac{S}{\sigma}(Y_t^c - Y_t^f)\,\mathrm{d}t + \mathrm{d}W_t^{\mathbb{P}},\\ \mathrm{d}W_t^{\mathbb{Q}^c} &= \frac{S}{\sigma}(Y_t^f - Y_t^c)\,\mathrm{d}t + \mathrm{d}W_t^{\mathbb{P}}.\end{aligned} \tag{5.57}$$

Therefore, under the simulation measure $\mathbb{P}$, we obtain that

$$\begin{aligned}\mathrm{d}Y_t^f &= S(Y_t^c - Y_t^f)\,\mathrm{d}t + f(Y_t^f)\,\mathrm{d}t + \sigma\,\mathrm{d}W_t^{\mathbb{P}},\\ \mathrm{d}Y_t^c &= S(Y_t^f - Y_t^c)\,\mathrm{d}t + f(Y_t^c)\,\mathrm{d}t + \sigma\,\mathrm{d}W_t^{\mathbb{P}}.\end{aligned} \tag{5.58}$$

Girsanov's theorem ensures that

$$\begin{aligned}\mathbb{E}^{\mathbb{P}}[\varphi(X_T^f)] - \mathbb{E}^{\mathbb{P}}[\varphi(X_T^c)] &= \mathbb{E}^{\mathbb{Q}^f}[\varphi(Y_T^f)] - \mathbb{E}^{\mathbb{Q}^c}[\varphi(Y_T^c)]\\ &= \mathbb{E}^{\mathbb{P}}\left[\varphi(Y_T^f)\frac{\mathrm{d}\mathbb{Q}^f}{\mathrm{d}\mathbb{P}_T} - \varphi(Y_T^c)\frac{\mathrm{d}\mathbb{Q}^c}{\mathrm{d}\mathbb{P}_T}\right],\end{aligned} \tag{5.59}$$

where $\frac{\mathrm{d}\mathbb{Q}^f}{\mathrm{d}\mathbb{P}_T}$ is the corresponding Radon-Nikodym derivative with the following form

$$\frac{\mathrm{d}\mathbb{Q}^f}{\mathrm{d}\mathbb{P}_T} = \exp\left(-\int_0^T \left\langle \frac{S}{\sigma}(Y_t^f - Y_t^c),\, \mathrm{d}W_t^{\mathbb{P}} \right\rangle - \frac{1}{2}\int_0^T \frac{S^2}{\sigma^2}\left\|Y_t^f - Y_t^c\right\|^2 \mathrm{d}t\right) \tag{5.60}$$

and $\frac{\mathrm{d}\mathbb{Q}^c}{\mathrm{d}\mathbb{P}_T}$ is similar. The benefit of this technique is that under the measure $\mathbb{P}$, we recover the contractivity between Y_t^c and Y_t^f and the variance of the level estimator increases linearly in T instead of exponentially. Note that for the numerical implementation we derive the exact Radon-Nikodym derivative for the numerical solution instead of the numerical approximation of (5.60). For the detailed numerical scheme, see section 2 in previous chapter.

However, in this chapter, we observe that both the Malliavin and the new PS estimations (5.29) and (5.41) involve Itô integrals and we need to be careful when performing the change of measure and respect the following identity:

$$\mathbb{E}^{\mathbb{Q}^f}\left[\int_0^T \left\langle \psi(Y_t^f), \mathrm{d}W_t^{\mathbb{Q}^f} \right\rangle\right] = \mathbb{E}^{\mathbb{P}}\left[\int_0^T \left\langle \psi(Y_t^f), \frac{S}{\sigma}(Y_t^c - Y_t^f)\mathrm{d}t + \mathrm{d}W_t^{\mathbb{P}} \right\rangle \frac{\mathrm{d}\mathbb{Q}^f}{\mathrm{d}\mathbb{P}_T}\right], \tag{5.61}$$

where the Itô integral is viewed as a functional of the whole path. The identity for the coarse path is similar. Therefore, for numerical experiments, we need to use $W_t^{\mathbb{P}}$ to reconstruct $W_t^{\mathbb{Q}^f}$ and $W_t^{\mathbb{Q}^c}$ to calculate the integral following (5.57).

Theorem 5.5.1 (Level Estimators with Change of Measure)**.** *Suppose $\widehat{Y}_t$ and $\widehat{y}_t$ satisfy the Assumption 5.0.1 under the measure $\mathbb{P}$ and f and γ are globally Lipschitz, the variance of the new MLMC level estimator with change of measure* (5.59) *and timestep h increases at most quadratically with respect to T, that is*

$$\mathbb{V}\left[\tilde{\varphi}(\widehat{Y}_T^f)\,\frac{\mathrm{d}\widehat{\mathbb{Q}}^f}{\mathrm{d}\mathbb{P}_T} - \tilde{\varphi}(\widehat{Y}_T^c)\,\frac{\mathrm{d}\widehat{\mathbb{Q}}^c}{\mathrm{d}\mathbb{P}_T}\right] \leq \frac{\eta'''}{\sigma^4}\,T^2\,h^2. \tag{5.62}$$

for some constant $\eta''' \in (0,\infty)$, where $\tilde{\varphi}$ is the functional of the Malliavin or new PS estimator.

Proof. Using the same approach as in the proof of Theorem 4.2.4 in the previous chapter and the fact that $\mathbb{E}\left[|\tilde{\varphi}(\widehat{Y}_T^f)|^p\right] \leq \lambda''/\sigma^p\,T^{p/2}$ gives the final result. Intuitively, the first order strong convergence gives h^2, the Itô integral gives a factor T/σ^2 by Theorem 5.3.1 and another factor of T/σ^2 comes from the Radon-Nikodym derivative (5.60). □

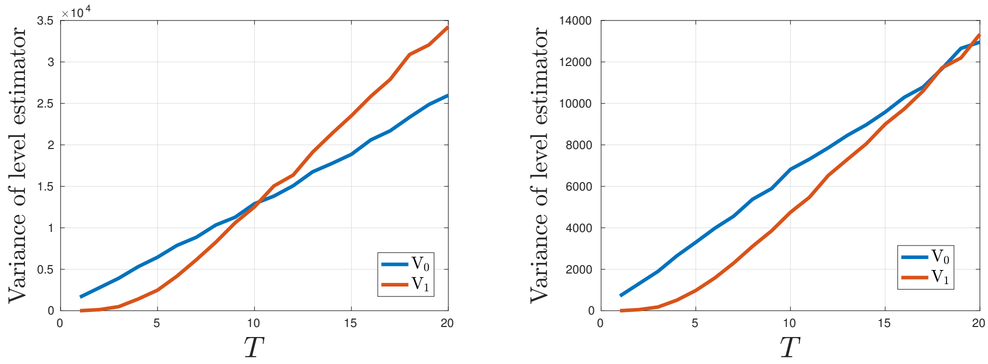
The MLMC theorem in the previous chapter applies here except that $V_0 = O(T/\sigma^2)$. By choosing $T = O(|\log\varepsilon|)$ we have $V_0 = O(|\log\varepsilon|)$, and the optimal computational cost of the MLMC with change of measure to achieve $O(\varepsilon^2)$ MSE becomes $O(\varepsilon^{-2}|\log\varepsilon|^3)$. Compared with the results in the previous chapter, the additional log term comes from the order of V_0.

5.5.2 Numerical results

In this subsection, we discuss the applications of standard MLMC and MLMC with change of measure to both the Malliavin estimator (5.29) and the new PS estimator (5.41). Numerically, we use adaptive timestepping with $h_0 = 2^{-7}$ and

$N = 10^5$ to compute the variances V_0 on level 0 and V_1 on level 1 with respect to T for the stochastic Lorenz equation with $\sigma = 6$.

First, we directly use the standard MLMC scheme for both estimators without change of measure. The variance on level 0 increases linearly in T, which is the same as the standard Monte Carlo method. However, for level $\ell \in \{1, 2, ..., L\}$, the variance of the level estimator still increases exponentially due to the chaotic property of the original SDEs.



(a) Malliavin estimator

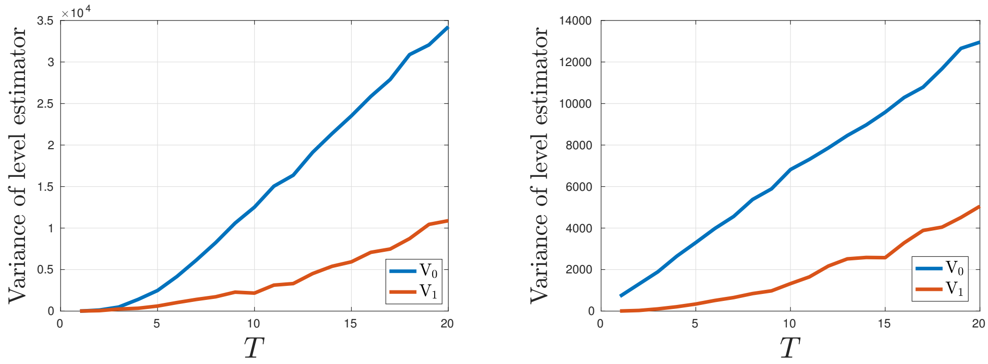
(b) New PS estimator

Figure 5.3: MLMC variances on levels 0 and 1 without change of measure

In Figure 5.3, we plot the V_0 and V_1 with respect to T and it confirms the linear increase of V_0 and the initial exponential increase of V_1. For both estimators, V_1 becomes larger than V_0 as T increases due to the bad coupling between fine and coarse paths.

Second, we add the spring term between the fine and coarse paths and perform the change of measure on the level estimator following (5.59) with $S = 10$. The variance on level 0 increases linearly. For level $\ell \in \{1, 2, ..., L\}$, the variance of the level estimator increases only quadratically as shown in Theorem 5.5.1.

In Figure 5.4, we plot V_0 and V_1 with respect to T and it confirms the linear increase of V_0 and the good coupling on level 1. The quadratic increases with respect to T are shown in Figure 5.5 by plotting V_1 with respect to T^2.



(a) Malliavin estimator

(b) New PS estimator

Figure 5.4: MLMC variances on levels 0 and 1 with change of measure

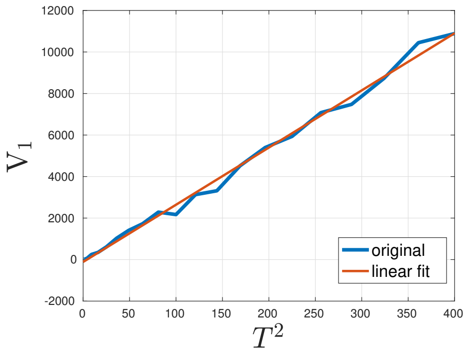
Finally, we ran the MLMC scheme for both estimators and $T = 10$, which is sufficiently large for acceptable weak convergence by Figure 5.2(b).

Figures 5.6 and 5.7 show that MLMC for both the Malliavin and the new PS estimator works well. The top left plot shows that the variance of the level estimator decreases at rate 2, i.e. the variance is proportional to $2^{-2\ell}$ corresponding to first order strong convergence. The top right plot shows first order weak convergence. The bottom left plot shows the different number of paths on different levels for different accuracies. The bottom right plot shows that the computational cost for MLMC is $O(\varepsilon^{-2})$ and for standard Monte Carlo is $O(\varepsilon^{-3})$.

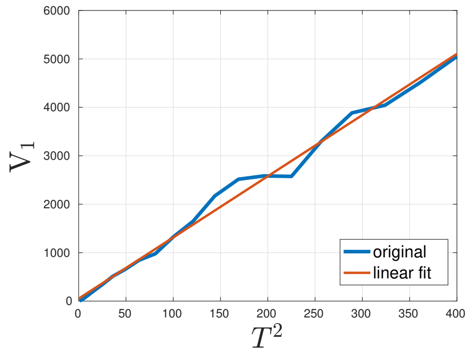
5.6 Other sensitivities

In this section, we consider the sensitivities of the invariant measure with respect to the volatility parameter σ and the initial datum ξ_0; the latter should be 0. Note that the Malliavin estimators fail here since they involve the variation process shown in Proposition 3.3 in [26].

Assume that we are calculating the sensitivity at σ_0 and consider the perturbed volatility σ. The derivation is similar to Section 5.3 except that we have the new



(a) Malliavin estimator



(b) New PS estimator

Figure 5.5: Quadratic increase with respect to T

variation process with respect to σ :

$$\mathrm{d}z_t^{(\sigma)} = \left(\frac{\partial f(Y_t^{(\sigma)})}{\partial Y_t^{(\sigma)}} - SI \right) z_t^{(\sigma)}\, \mathrm{d}t + \mathrm{d}W_t^{\tilde{\mathbb{Q}}}, \tag{5.63}$$

and the derivative of the Radon-Nikodym derivative with respect to σ,

$$\begin{aligned} \frac{\partial R_t^{(\sigma)}}{\partial \sigma} \;=\; & R_t^{(\sigma)} \left(\int_0^t \frac{S}{\sigma} \left\langle z_u^{(\sigma)},\, \mathrm{d}W_u^{\tilde{\mathbb{Q}}} \right\rangle + \int_0^t \frac{S}{\sigma^2} \left\langle Y_u^{(\sigma_0)} - Y_u^{(\sigma)},\, \mathrm{d}W_u^{\tilde{\mathbb{Q}}} \right\rangle \right. \qquad (5.64) \\ & \left. + \int_0^t \frac{S^2}{\sigma^2} \langle Y_u^{(\sigma_0)} - Y_u^{(\sigma)}, z_u^{(\sigma)} \rangle \,\mathrm{d}u + \int_0^t \frac{S^2}{\sigma^3} \left\| Y_u^{(\sigma_0)} - Y_u^{(\sigma)} \right\|^2 \mathrm{d}u \right). \end{aligned}$$

Therefore, taking $\sigma = \sigma_0$ and again under the measure $\mathbb{P}$, it holds that

$$\left. \frac{\partial F(\sigma)}{\partial \sigma} \right|_{\sigma=\sigma_0} = \mathbb{E}^{\mathbb{P}} \left[\langle \nabla \varphi(Y_T^{(\sigma_0)}),\, z_T^{(\sigma_0)} \rangle + \varphi(Y_T^{(\sigma_0)}) \int_0^T \frac{S}{\sigma} \left\langle z_t^{(\sigma_0)},\, \mathrm{d}W_t^{\mathbb{P}} \right\rangle \right]. \tag{5.65}$$

Numerically, we only need to simulate $Y_t^{(\sigma_0)}$ and $y_t^{(\sigma_0)}$:

$$\begin{aligned} \mathrm{d}\, Y_t^{(\sigma_0)} &= f(Y_t^{(\sigma_0)})\, \mathrm{d}t \; + \sigma_0\, \mathrm{d}W_t^{\mathbb{P}}, \\ \mathrm{d}\, z_t^{(\sigma_0)} &= \left(\frac{\partial f(Y_t^{(\sigma_0)})}{\partial Y_t^{(\sigma_0)}} - SI \right) z_t^{(\sigma_0)}\, \mathrm{d}t \; + \; \mathrm{d}W_t^{\mathbb{P}}, \end{aligned} \tag{5.66}$$

and the new PS estimator is

$$\frac{\partial \breve{F}}{\partial \sigma} = \left\langle \nabla \varphi(\widehat{Y}_T^{(\sigma)})\, , \widehat{z}_T^{(\sigma)} \right\rangle + \varphi(\widehat{Y}_T^{(\sigma)}) \sum_{n=0}^{N-1} \frac{S}{\sigma} \left\langle \widehat{z}_{t_n}^{(\sigma)},\, \Delta W_n^{\mathbb{P}} \right\rangle. \tag{5.67}$$

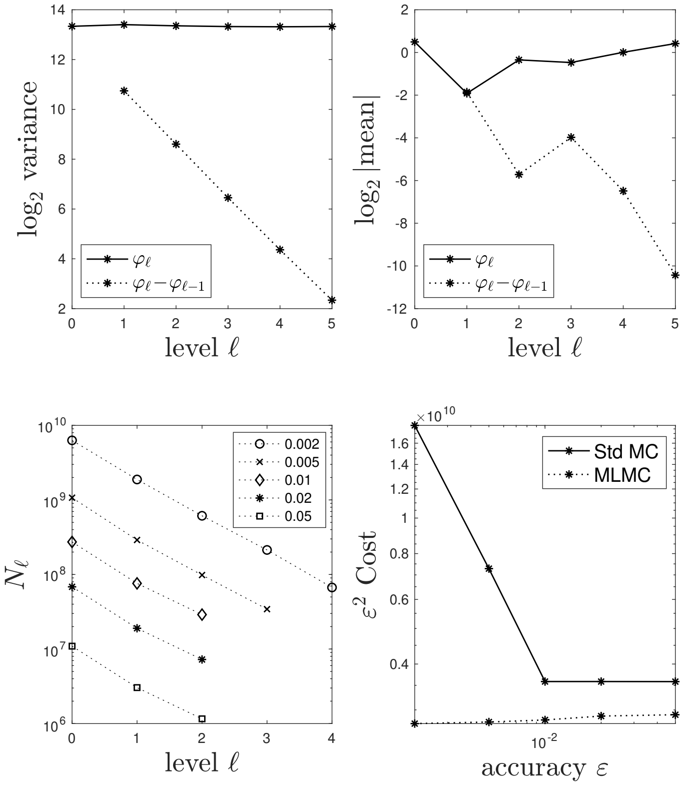


Figure 5.6: MLMC for the Malliavin estimator

We perform the numerical experiment on the stochastic Lorenz equation with $\sigma = 6$. Figure 5.8(a) shows the estimator is ergodic and the expectation converges to the sensitivity of the invariant measure 0.1274, which is close to the finite difference value 0.1268 using the data in Figure 5.11. Figure 5.8(b) confirms the linear increase of the variance.

For the initial condition ξ_0, we get the new variation process

$$\mathrm{d}\, z_t = \left(\frac{\partial f(Y_t)}{\partial Y_t} - SI \right) z_t \, \mathrm{d}t, \tag{5.68}$$

with $z_0 = 1$ and the new PS estimator

$$\frac{\partial \breve{F}}{\partial \xi_0} = \left\langle \nabla \varphi(\widehat{Y}_T) \, , \widehat{z}_T \right\rangle + \varphi(\widehat{Y}_T) \sum_{n=0}^{N-1} \frac{S}{\sigma} \left\langle \widehat{z}_{t_n}, \, \Delta W_n^{\mathbb{P}} \right\rangle . \tag{5.69}$$

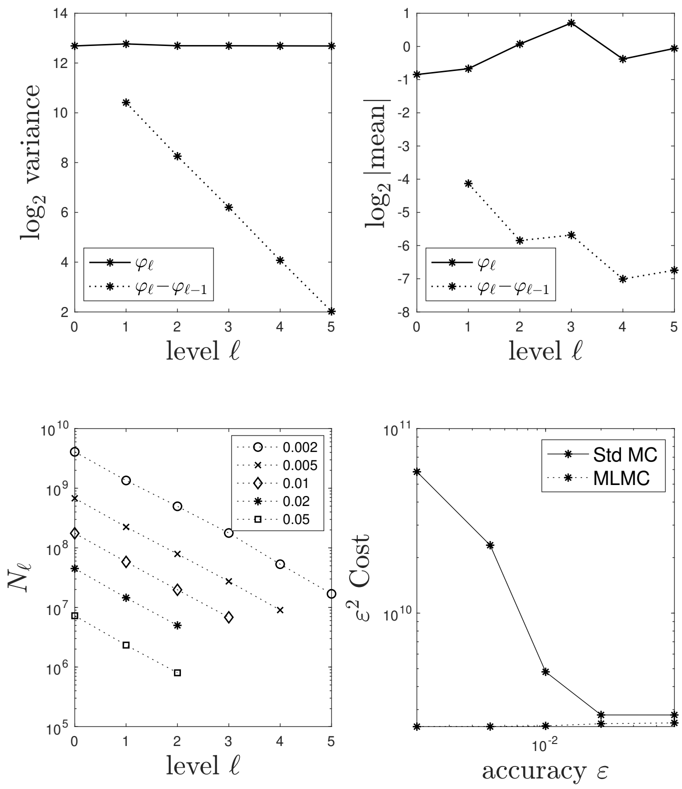
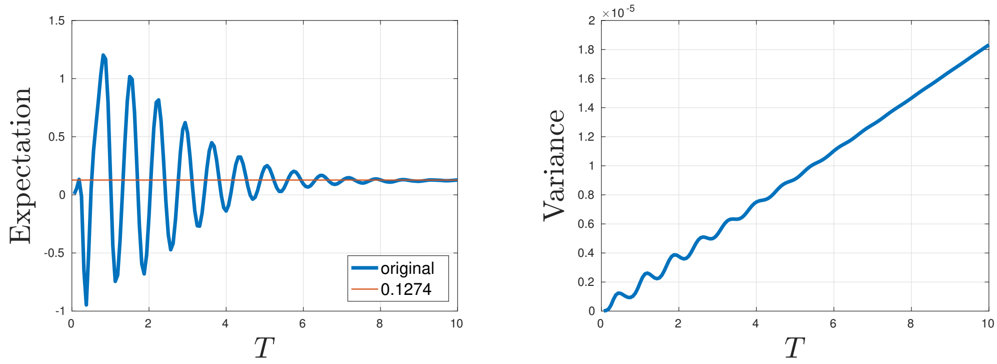


Figure 5.7: MLMC for the new PS estimator

Figure 5.9(a) shows that the estimator is ergodic and the expectation converges to 0 since the initial condition won't affect the invariant measure. Figure 5.9(b) shows the variance is uniformly bounded.

5.7 Extension to the Lorenz problem

In this section, we consider using the new PS estimator for stochastic SDEs to approximate the sensitivities of invariant measures of chaotic ODEs, which is a long-standing problem. For the invariant measure of a chaotic ODE, the main



(a) Expectation with respect to T

(b) Linear increase of variance

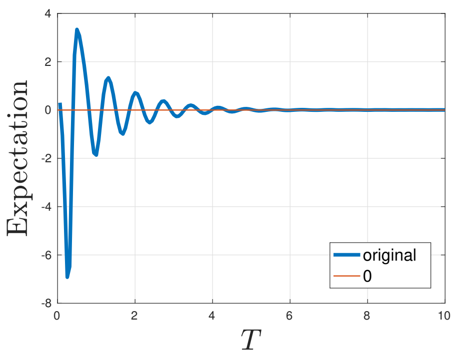
Figure 5.8: The sensitivity with respect to volatility σ

concept is the Sinai-Ruelle-Bowen (SRB) measure. An invariant probability measure π^0 for a flow Z_t is a physical probability measure if the subset of z satisfying for all continuous function φ

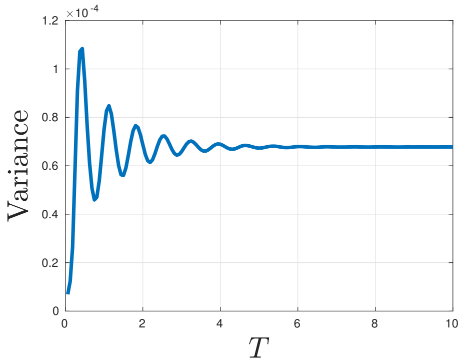
$$\lim_{T\to+\infty}\frac{1}{T}\int_0^T\varphi(Z_t(z))\mathrm{d}t=\pi^0(\varphi). \tag{5.70}$$

has positive volume in space. [3] and [100] together show that the Lorenz equation has a unique SRB measure. For the SRB measure of the system, we are also concerned about the the sensitivity with respect to some parameters of the system. It is shown in [88, 89] that statistical quantities are differentiable with respect to small perturbations of parameters for quasi-hyperbolic systems, for example, the Lorenz system.

However, the sensitivity computations of chaotic systems like the Lorenz system are difficult due to the ill-conditioned initial value problem. Lea, Allen & Haine [62] used the Lorenz equation as an example to illustrate the failure of adjoint methods as the computed sensitivity blows up exponentially. They also proposed an ensemble-adjoint approach, which uses the average of the computed sensitivities in a intermediate time scale. Choosing the time scale is a hard problem and this approach is unlikely to be stable for more complex systems. Wang [102] introduced a shadow operator to the original system and found the shadow trajectory



(a) Expectation with respect to T



(b) Uniformly bounded variance

Figure 5.9: The sensitivity with respect to the initial value

by inverting the shadow operator and then calculated the sensitivity. Further, Wang, Hu & Blonigan [104] found the shadow trajectory by solving a constrained minimization problem and the corresponding convergence result is shown in [103].

Our approach is to estimate the sensitivity of the ODE by estimating the sensitivity of the SDE with small volatility $\sigma \in (0, \infty)$. The consideration is that, without the random force, the evolution of the invariant measure of the ODE is governed by the corresponding Liouville equation which may not have a well-behaved steady solution. By adding the random noise, the evolution of the invariant measure of the SDE is governed by the corresponding steady Fokker-Planck equation which has a smooth, well-behaved solution. Thuburn [97] appreciated this point and introduced the small diffusion term and calculated the sensitivities by considering the Fokker-Planck equation, but it required the solution of the high-dimensional elliptic PDE numerically, which is computationally expensive. In addition, as shown in Figure 5.10(a) later this section, the random noise also accelerates the convergence speed to the equilibrium of the system. Next, we first state some theoretical results about the convergence of the solution and invariant measure from the SDE to the ODE, and then show our numerical results for the ordinary Lorenz system.

5.7.1 Convergence of SDEs to ODEs

Freidlin & Wentzell [27] provide a systematic framework for the convergence of SDEs to ODEs. Theorem 1.1 in chapter 2 in [27] shows that if f is continuous and the ODE has a unique solution x_t, then for sufficiently small $\sigma \in (0, \infty)$, we have

$$\mathbb{P}\left(\lim_{\sigma\to 0}\max_{0\le t\le T}\|X_t^\sigma - x_t\| = 0\right) = 1, \tag{5.71}$$

where X_t^σ is the solution to the SDE with volatility σ and Theorem 1.2 shows that if the drift is globally Lipschitz and increases no faster than linearly in x, then for all $t \in (0, \infty)$, we obtain that

$$\mathbb{E}\left[\|X_t^\sigma - x_t\|^2\right] \le a(t)\,\sigma^2, \tag{5.72}$$

where $a(t)$ is a monotonic increasing function depending on the initial data and the Lipschitz constant.

For the invariant measure, Theorem 4.2 in chapter 6 in [27] shows that if the drift $f(x)$ is bounded and uniformly continuous, and the deterministic system has a unique invariant measure π^0, then π^σ converges weakly to π^0 as $\sigma \to 0$, where π^σ is the invariant measure of the SDE with volatility σ.

5.7.2 Numerical investigation

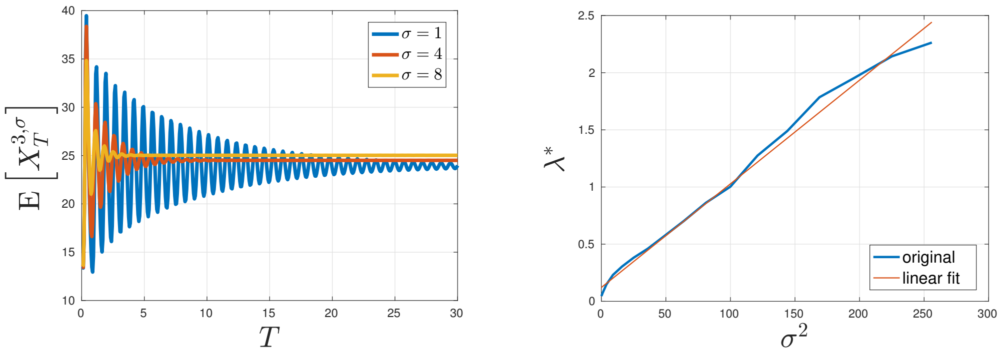
Now we investigate the ordinary Lorenz equation

$$\frac{\mathrm{d}x_t}{\mathrm{d}t} = \begin{pmatrix} 10(x_t^2 - x_t^1) \\ x_t^1(\theta - x_t^3) - x_t^2 \\ x_t^1 x_t^2 - \frac{8}{3}x_t^3 \end{pmatrix}. \tag{5.73}$$

We estimate the sensitivity of $\bar{x}_3 = \lim_{T\to\infty}\frac{1}{T}\int_0^T x_t^3\,\mathrm{d}t$ with respect to θ at $\theta = 28$, which is approximately 0.981, by an average of 50 finite difference estimates using different initial conditions as in [15].

Theorem 3.1 in chapter IV in [57] shows that the invariant measure π^σ of the stochastic Lorenz equation weakly converges to π^0 as $\sigma \to 0$ in discrete version, but the convergence speed is not known. Before we numerically investigate the convergence speed, we first notice that the magnitude of the volatility σ affects the speed of convergence starting from the same initial point.

First, we plot the expectation $\mathbb{E}\left[X_T^{3,\sigma}\right]$ with respect to time T for different σ from the same initial point $[-2.4, -3.7, 14.98]$. See Figure 5.10(a).



(a) Evolutions of different σ (b) Convergence speed for different σ^2

Figure 5.10: The effect of σ on convergence speed λ^*

We see that the time period needed to approach equilibrium decreases and the equilibrium value increases as σ increases, which indicates that σ affects the convergence speed λ^*. Next, to quantify this effect, we calculate the error bound using the moving maximum and minimum and then perform a linear regression of log error bound on T to get the estimated convergence speed λ^*. We plot the estimated λ^* with respect to σ^2 in Figure 5.10(b), which shows clearly that the estimated convergence rate λ^* is approximately proportional to σ^2. Therefore, we need to use a longer T to simulate the SDE with a smaller σ to achieve the same truncation error in T. It also indicates that the convergence speed of the ODE is far slower than the SDE since ODE has $\sigma = 0$.

Next, we investigate the weak convergence order of the invariant measures with respect to σ. First, we estimate $\bar{x}_3$ for different values of θ ranging from 27.50 to 28.50 using the 4th-order Runge-Kutta method with $h = 0.001$, $T = 300$ and $x_0 = (-2.4, -3.7, 14.98)^T$. See the black line in Figure 5.11. It oscillates wildly which adversely affects the accuracy of the finite difference method.

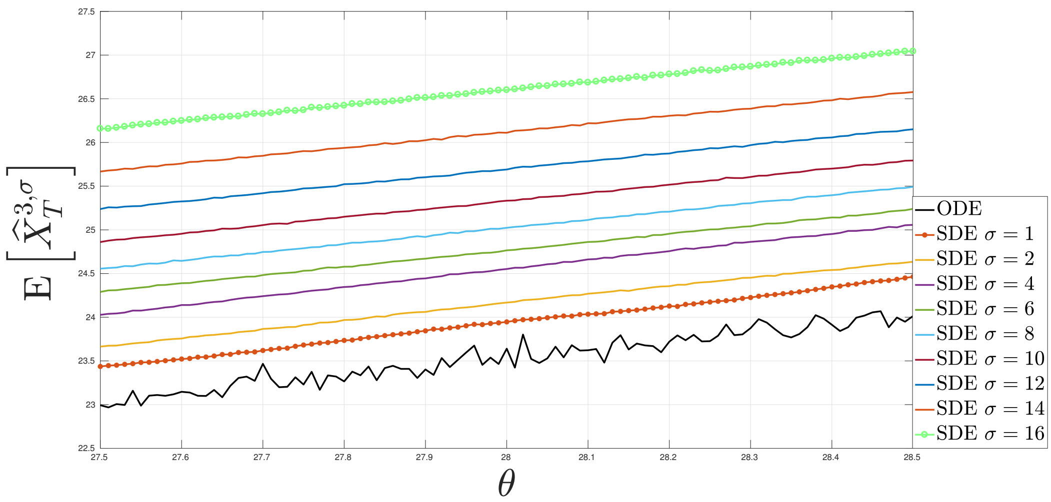


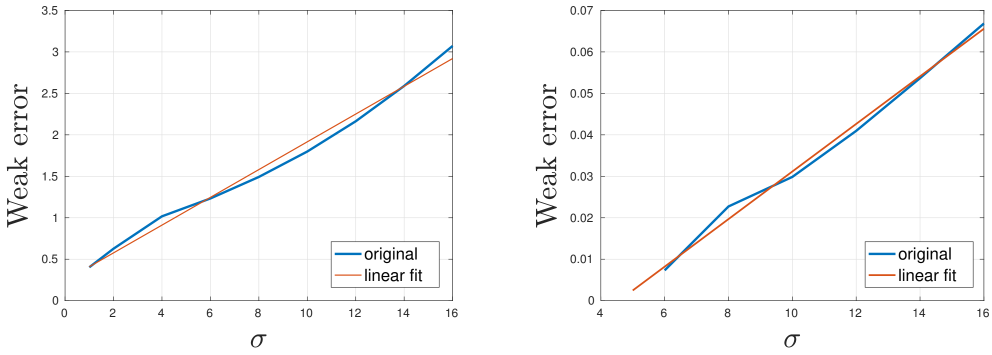
Figure 5.11: ODE and SDEs with different σ

For SDEs, we know from the previous estimates of λ^* that we need to choose different times T for different σ to ensure that the equilibrium is achieved. Figure 5.11 shows that, by introducing a small random noise term, the solutions of the SDE become smoother with respect to different θ and they converge to the solution of the ODE as σ decreases. Simultaneously the sensitivity problem becomes better-posed and the finite difference method already can provide a good estimate of the sensitivity.

Next, based on Figure 5.11, we estimate the weak convergence order with respect to σ, see Figure 5.12(a). It shows first order weak convergence.

Similarly, we plot the evolutions of $\mathrm{E}\left[\frac{\partial \breve{F}}{\partial \theta}(\sigma)\right]$ with respect to T for different σ and observe the same results as the original value. Compared with Figure 5.10, the time period needed to achieve the equilibrium of the sensitivities is longer since the estimator is a path integral of the original value.

We also plot the sensitivity with respect to θ for the SDE with different values of σ in Figure 5.14. The black line shows the sensitivity of the ODE, a constant 0.981 shown in [15]. For the SDE, we use the new PS estimator and similarly simulate



(a) Original value (b) Sensitivity

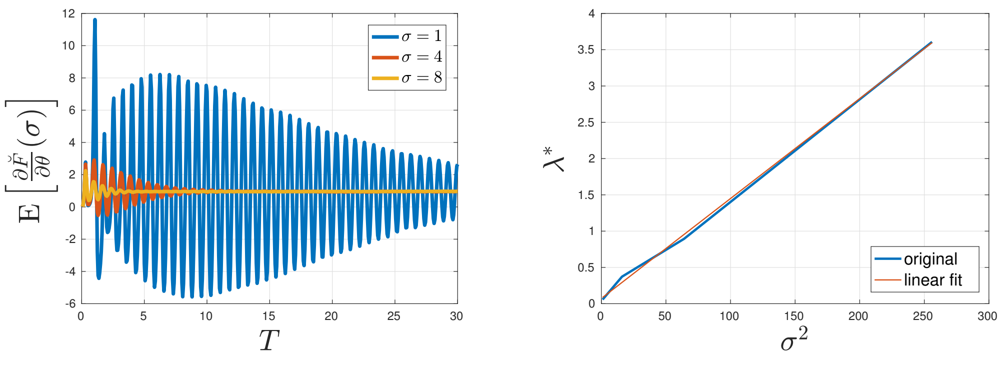
Figure 5.12: Weak convergence with respect to σ

longer T for smaller σ. As shown in Figure 5.14, the sensitivities of the SDE are less smooth with respect to θ due to the large variance. We plot the average value in the same color and observe that the sensitivity $\mathrm{E}\left[\frac{\partial \breve{F}}{\partial \theta}(\sigma)\right]$ increases as σ decreases and converges to the one for the ODE. The first order weak convergence is shown in Figure 5.12(b). Now, we analyze the computational cost first by stating the following numerical results obtained from the previous experiments.

- The convergence speed λ^* is proportional to σ^2, that is $\lambda^* \sim O(\sigma^2)$. See Figure 5.10(b).
- The weak convergence order of invariant measure from SDE to ODE with respect to σ is 1, see Figures 5.12(a) and 5.12(b).
- The variance of the new PS estimator linearly increases as T increases see Figure 5.2(a) and as σ^{-2} increases by Theorem 5.3.1, that is $\mathbb{V} \sim O(T/\sigma^2)$.

Theorem 5.7.1 (ODE Approximation)**.** *Suppose that the conclusions above hold. Then, to achieve $O(\varepsilon^2)$ MSE, the optimal complexity of the standard Monte Carlo method to approximate the sensitivity of the ODE is $O(\varepsilon^{-9}|\log \varepsilon|^2)$*

Proof of Theorem 5.7.1. Denoting the sensitivity of the ODE by π' and that of



(a) Evolutions of different σ (b) Convergence speed for different σ^2

Figure 5.13: The effect of σ on convergence speed λ^*

the SDE by Π', the weak error can be decomposited into

$$\begin{aligned}\left|\mathbb{E}\left[\frac{\partial \breve{F}}{\partial \theta}(\sigma)\right]-\pi'\right| &= \left|\mathbb{E}\left[\frac{\partial \breve{F}}{\partial \theta}(\sigma)-\frac{\partial F}{\partial \theta}(\sigma)+\frac{\partial F}{\partial \theta}(\sigma)-\Pi'+\Pi'-\pi'\right]\right| \\ &\leq \left|\mathbb{E}\left[\frac{\partial \breve{F}}{\partial \theta}(\sigma)-\frac{\partial F}{\partial \theta}(\sigma)\right]\right|+\left|\mathbb{E}\left[\frac{\partial F}{\partial \theta}(\sigma)-\Pi'\right]\right|+\left|\mathbb{E}\left[\Pi'-\pi'\right]\right| \\ &= O(h)+O(e^{-\lambda^* T})+O(\sigma). \end{aligned} \tag{5.74}$$

Bounding the weak error by $O(\varepsilon)$, we obtain that

$$h \sim O(\varepsilon), \quad \sigma \sim O(\varepsilon), \tag{5.75}$$

and

$$e^{-\lambda^* T} \sim O(\varepsilon) \Rightarrow \lambda^* T \sim O(|\log \varepsilon|) \Rightarrow T \sim O(\varepsilon^{-2}|\log \varepsilon|), \tag{5.76}$$

which implies that the computational cost per path is $T/h \sim O(\varepsilon^{-3}|\log \varepsilon|)$.

Next, the variance of the MC estimator also needs to be bounded by $O(\varepsilon^2)$, which requires that the number of paths N satisfies

$$T/(\sigma^2 N) \sim O(\varepsilon^2) \Rightarrow N \sim O(\varepsilon^{-6}|\log \varepsilon|). \tag{5.77}$$

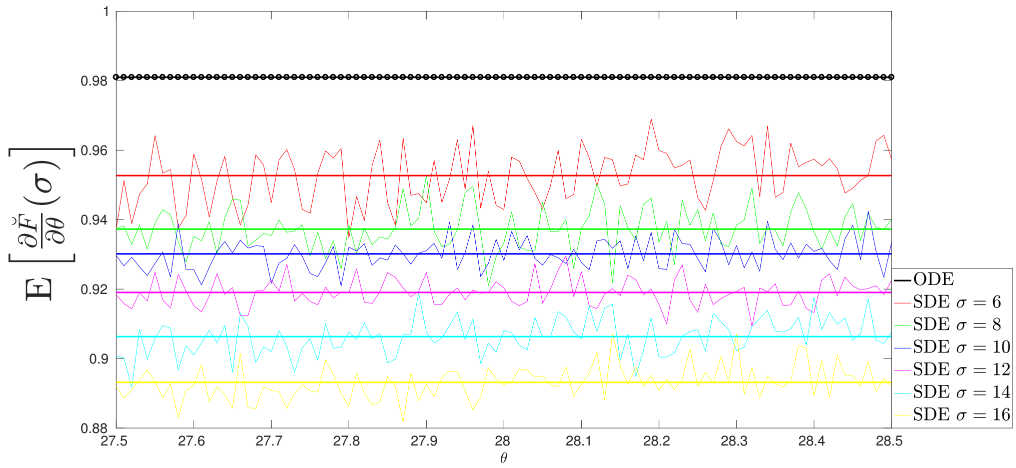


Figure 5.14: Sensitivities of ODE and SDEs with different σ

Therefore, the total computational cost is $O(\varepsilon^{-9}|\log \varepsilon|^2)$. This completes the proof of Theorem 5.7.1. □

For MLMC, we still require $T \sim O(\varepsilon^{-2}|\log \varepsilon|)$ and the timestep in the finest level is $O(\varepsilon)$. The key point here is efficient coupling, that is the choice of the timestep size h_0 on level 0, which requires $C_0V_0 > C_1V_1$, that is, following Theorem 5.3.1 and 5.5.1, we have that

$$V_1 = O(h_0^2 T^2/\sigma^4) \sim V_0 = O(T/\sigma^2), \tag{5.78}$$

and then

$$h_0 \sim O(\sigma/\sqrt{T}) \Rightarrow h_0 \sim O(\varepsilon^2|\log \varepsilon|^{-1/2}), \tag{5.79}$$

which means that the required h_0 is far smaller than the size we need and as a result, MLMC with change of measure doesn't work. The main issue here is that the small σ destroys the variance of the Radon-Nikodym derivatives. Therefore, constructing an estimator with higher order convergence in σ to the ODE sensitivity is the key, and we deal with that in the next subsection.

5.7.3 Richardson-Romberg extrapolation

In this subsection, we construct new estimators by applying Richardson-Romberg (R-R) extrapolation to σ to improve the weak convergence order of σ and then reduce the complexity. R-R extrapolation was first introduced in [95] and extended to multi-step R-R extrapolation in [81] to achieve a higher order weak convergence rate, and further extended to multilevel R-R extrapolation in [65].

We assume that under suitable conditions the sensitivity $\Pi'(\sigma)$ of the invariant measure of SDEs with volatility σ has the following expansion at $\sigma = 0$ with $\Pi'(0) = \pi'$ which is the sensitivity of the ODEs,

$$\Pi'(\sigma) = \pi' + \sum_{k=1}^{R-1} c_k \, \sigma^k + O(\sigma^R) \tag{5.80}$$

where the real constants c_k, for $k \in \{1, 2, ..., R-1\}$, do not depend on σ. Denoting $\tilde{\Pi}'(\sigma) = (\Pi'(\sigma/1), \Pi'(\sigma/2), ..., \Pi'(\sigma/R))^T$, and following the multi-step R-R extrapolation procedure, there exists a weight vector $\tilde{w} = (w_1, ..., w_R) \in \mathbb{R}^R$ such that

$$\tilde{w} \, \tilde{\Pi}'(\sigma) = \sum_{k=1}^{R} w_k \, \Pi'(\sigma/k) = \pi' + O(\sigma^R), \tag{5.81}$$

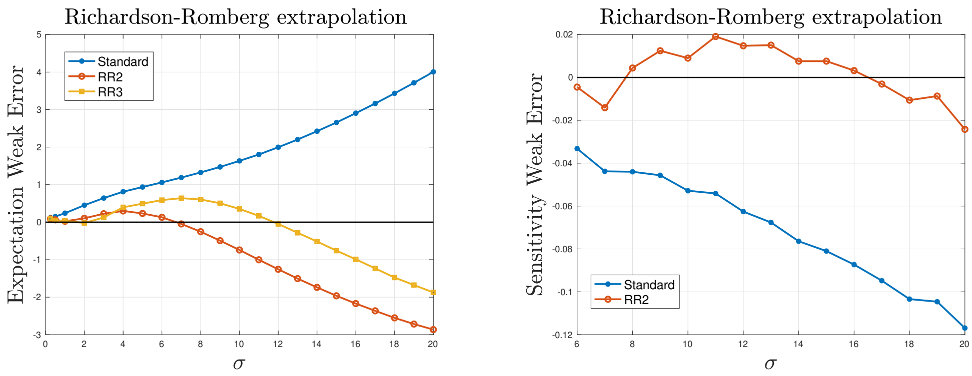
where $w_k = \frac{(-1)^{R-k} k^R}{k!(R-k)!}$ for $k = 1, ..., R$. Specifically, we obtain that

$$\begin{aligned} &R = 1 \qquad w_1 = 1; \\ &R = 2 \qquad w_1 = -1, \ w_2 = 2; \\ &R = 3 \qquad w_1 = 1/2, \ w_2 = -4, \ w_3 = 9/2. \end{aligned} \tag{5.82}$$

Similarly, we construct new estimators with higher order weak convergence rate for σ. For the original expectation $F(\theta, \sigma)$, we obtain that

$$\begin{aligned} &R = 1 \qquad F(\theta, \sigma) \\ &R = 2 \qquad -F(\theta, 2\sigma) + 2F(\theta, \sigma); \\ &R = 3 \qquad \frac{1}{2} F(\theta, 3\sigma) - 4F(\theta, 1.5\sigma) + \frac{9}{2} F(\theta, \sigma); \end{aligned} \tag{5.83}$$

where we use σ as the smallest value for each estimator since the estimator with smallest σ converges slowest, which determines the simulation time T. The construction of the estimators for the sensitivities is exactly the same. Numerically, we plot the weak errors for these three estimators with the same smallest $\sigma = 1/4, 1/2, 1, 2, ..., 20$ for the Lorenz problem in Figure 5.15(a) for the original expectation and Figure 5.15(b) for the sensitivity.



(a) Original value (b) Sensitivity

Figure 5.15: Weak convergence of different R-R estimators

We see that the R-R extrapolation improves the weak convergence rate significantly. Especially for the sensitivity, the 2nd order R-R estimator already gives an accurate estimate, which may be due to the linearity of the sensitivity in σ. The computational cost increases linearly in R since we need to simulate R paths with different σ but with the same driving Brownian motion to reduce the variance.

Therefore, following the same approach as in Theorem 5.7.1, for the estimator with order R weak convergence rate for σ, to achieve ε^2 MSE, we only need $\sigma \sim O(\varepsilon^{1/R})$ and $T \sim O(\varepsilon^{-2/R}|\log \varepsilon|)$, and reduce the computational cost of the standard Monte Carlo method to $O(R\varepsilon^{-3-6/R}|\log \varepsilon|^2)$. The optimal R to minimize the cost is $\lceil 6|\log \varepsilon| \rceil$ which gives a total cost $O(\varepsilon^{-3}|\log \varepsilon|^3)$.

Similarly, for the MLMC with change of measure, to achieve a good coupling, we require $h_0 \sim O(\varepsilon^{2/R}|\log \varepsilon|^{-1/2})$, and then the optimal computational cost of

MLMC is $O(\varepsilon^{-2-8/R}|\log\varepsilon|^3)$. The optimal R to minimize the cost is $\lceil 8|\log\varepsilon|\rceil$, which gives a total cost $O(\varepsilon^{-2}|\log\varepsilon|^4)$.

For the Lorenz problem, Figure 5.15(b) shows that the 2nd order R-R estimator with $\sigma = 15$ is accurate enough. The convergence rate λ^* is much larger and we only need to simulate the SDEs to time $T = 2$, which is already sufficiently large to achieve equilibrium. Another benefit is that the h_0 used on level 0 is much larger.

Compared with Figure 5.7, the new R-R estimator has a much better accuracy and much smaller computational cost. The following table gives the details of the comparison. Note that the standard estimator still have a large bias due to σ and reducing σ to a sufficiently small value to achieve higher accuracy becomes computationally infeasible within a reasonable time period, while the 2nd order R-R estimator has achieved good accuracy with a much smaller computational cost.

Table 5.1: Computation cost comparison

	σ	T	h_0	Computational cost	Estimated value
Standard	6	10	2^{-7}	$2.450e+13$	0.93264
RR2	15	2	2^{-4}	$2.668e+11$	0.97793

5.8 Conclusion

In this chapter, following the idea of introducing a spring proposed in the previous chapter, a new pathwise sensitivity estimator using importance sampling for the sensitivities of chaotic SDEs is derived in section 5.3. Numerical results in section 5.4 show that both the new estimator and the Malliavin estimator perform well. The variance of the new pathwise estimator increases linearly in T, which is a great improvement compared with the exponential increase of the standard pathwise estimator. Next we extended them to MLMC scheme with change of measure successfully in section 5.5. Similarly, the estimators for the sensitivities with respect to the initial value and the volatility parameters are derived in section 5.6. Last, we used the new estimator for SDEs with small volatility to approximate the sensitivity of ODEs. By combining this with a Richardson-Romberg extrapolation, we also obtained a new efficient estimator to avoid the effect of small volatility on the variance of the estimator.

Chapter 6

Conclusions

6.1 Summary of results

The central conclusion of this thesis is that by using an adaptive timestep it is possible to make the Euler-Maruyama approximation stable for SDEs with globally Lipschitz volatility and a drift which is not globally Lipschitz but is locally Lipschitz and satisfies a one-sided linear growth condition.

In chapter 2, we proved the stability and strong convergence of the adaptive scheme for a finite time interval. If the drift also satisfies a one-sided Lipschitz condition then the order of strong convergence is $\frac{1}{2}$, when looking at the accuracy versus the expected cost of each path. For the important class of Langevin equations with unit volatility, the order of strong convergence is 1. The orders of strong convergence are important because they lead to MLMC results for the computational cost to achieve a desired MSE for a certain class of functionals. The numerical experiments suggest that in some applications the new method may not be significantly better than the tamed Euler-Maruyama method proposed and analysed by Hutzenthaler, Jentzen & Kloeden [48], but in others it is shown to be superior. For the FENE model, although the theory does not apply, the adaptive scheme is natural and works best compared with others.

In chapter 3, we extended the analysis to the ergodic SDEs and proved that the moments and strong error of the numerical solutions are uniformly bounded in time T. Moreover, we extended this adaptive scheme to the MLMC method for expectations with respect to the invariant measure by using different time intervals on different levels, and constructed an efficient coupling of the fine path and the coarse path with different simulation times. Numerical results support our theory. For SDEs with contractivity on average, for example, the FENE model and SDEs with a double-well potential, our scheme with adaptive T works well and achieves an optimal complexity.

In chapter 4, we introduced a new change of measure technique for multilevel Monte Carlo estimators by adding a spring term to link the coarse and fine path approximations. For chaotic ergodic SDEs satisfying the one-sided Lipschitz condition, we reduced the exponential increase of the variance to a linear increase, which greatly reduces the computational cost when our interest is in the expectation with respect to the invariant measure. The numerical analysis only works for globally Lipschitz drift, but numerical experiments in section 4.4 show that the change of measure technique works well for the case with non-globally Lipschitz drift using adaptive timestepping.

In chapter 5, we proposed a new pathwise sensitivity estimator using importance sampling for the sensitivities of chaotic SDEs. Both the new estimator and Malliavin estimator perform well and have been extended to MLMC successfully. The benefit of the new pathwise estimator fundamentally changes the variation process and is easily extended to sensitivity with respect to the initial value and the volatility parameters. In addition, we also considered using this estimator for SDE with small volatility to approximate the sensitivity of ODEs. Together with the Richardson-Romberg extrapolation, we constructed a new efficient estimator overcoming the effect of the small volatility.

6.2 Future work

A number of areas for further research arise this work. One area for research concerns the restriction of the current assumptions. The first direction is to extend the numerical analysis to ergodic SDEs with non-globally Lipschitz drifts using adaptive timestepping and the change of measure technique, since numerical experiments in section 4.4 show it works well. Another direction for extension of the theory is to SDEs with a volatility which is not globally Lipschitz, but instead satisfies the Khasminskii-type condition used by Mao & Szpruch [68, 70]. A third option is to use a Lyapunov function $V(x)$ in place of $\|x\|^2$ in the stability analysis; this might enable one to prove stability and convergence for a larger set of SDEs. For SDEs such as the stochastic van der Pol oscillator and the stochastic Lorenz equation, if we could prove exponential integrability using the approach of Hutzenthaler, Jentzen & Wang [49] then it may be possible to establish an order of strong convergence using a locally one-sided Lipschitz condition. The other option is to apply adaptive methods to SDEs with a discontinuous drift. Neuenkirch et al. [80] have done some pioneering work in this direction.

Next, we can explore other applications of spring couplings between coarse and fine paths. For the simulation of the invariant measure of some Markov chains arising from chemical reactions, we can extend the results in [38] by Glynn and Rhee to non-contractive Markov chains by introducing a spring term to couple coarse and fine Markov chains. Another possible application may be to computing the invariant measure of SPDEs [18].

Lastly, we can continue to improve our algorithms. One possibility is to extend the analysis to Milstein approximations, which are particularly important when the SDE is scalar or satisfies the commutativity condition, which means that the Milstein approximation does not require the simulation of Lévy areas. Another direction is to develop a more efficient scheme for the sensitivity of the expectation of a discontinuous functional φ with respect to the invariant measure. For example, we could consider the combination of our technique with vibrato Monte Carlo method [31].

Appendix A

Multi-dimensional Itô formula

For a SDE:

$$\mathrm{d}\widehat{X}_t = f(\bar{X}_t)\,\mathrm{d}t + g(\bar{X}_t)\,\mathrm{d}W_t, \tag{A.1}$$

we assume that f and g are locally Lipschitz and satisfy the one-sided linear growth condition. Note that

$$\frac{\partial \|X\|^p}{\partial X_i} = p\,\|X\|^{p-1}\frac{\partial \|X\|}{\partial X_i} = p\,\|X\|^{p-1}\frac{X_i}{\|X\|} = p\,\|X\|^{p-2}X_i$$

$$\begin{aligned} \frac{\partial \|X\|^p}{\partial X} &= p\,\|X\|^{p-2}X, \\ \frac{\partial^2 \|X\|^p}{\partial X_i^2} &= p\,\|X\|^{p-2} + p(p-2)\,\|X\|^{p-4}X_i^2, \\ \frac{\partial^2 \|X\|^p}{\partial X_i \partial X_j} &= p(p-2)\|X\|^{p-4}X_iX_j, \end{aligned} \tag{A.2}$$

and then by Itô's formula, we obtain that

$$
\begin{aligned}
\mathrm{d}\|\widehat{X}_t\|^p =& p\,\langle \widehat{X}_t, f(\bar{X}_t)\rangle\, \|\widehat{X}_t\|^{p-2}\mathrm{d}t + p\,\widehat{X}_t^T\, g(\bar{X}_t)\, \|\widehat{X}_t\|^{p-2}\mathrm{d}W_t \\
&+ \frac{1}{2}\sum_{i=1}^{m}\sum_{j=1}^{m} \frac{\partial^2 \|\widehat{X}_t\|^p}{\partial \widehat{X}_i \partial \widehat{X}_j} g_i(\bar{X}_t) g_j(\bar{X}_t)^T \,\mathrm{d}t \\
=& p\,\langle \widehat{X}_t, f(\bar{X}_t)\rangle\, \|\widehat{X}_t\|^{p-2}\mathrm{d}t + p\,\widehat{X}_t^T\, g(\bar{X}_t)\, \|\widehat{X}_t\|^{p-2}\mathrm{d}W_t \\
&+ \frac{1}{2}\sum_{i=1}^{m} p\|\widehat{X}_t\|^{p-2} g_i(\bar{X}_t) g_i(\bar{X}_t)^T \,\mathrm{d}t \\
&+ \frac{1}{2}\sum_{i=1}^{m}\sum_{j=1}^{m} p(p-2)\|\widehat{X}_t\|^{p-4}\widehat{X}_i\widehat{X}_j g_i(\bar{X}_t) g_j(\bar{X}_t)^T \,\mathrm{d}t,
\end{aligned}
\tag{A.3}
$$

where g_i is the i-th row vector of g. By the definition of the Frobenius norm of a matrix, we obtain that

$$
\frac{1}{2}\sum_{i=1}^{m} p\|\widehat{X}_t\|^{p-2} g_i(\bar{X}_t) g_i(\bar{X}_t)^T = \frac{1}{2} p\|\widehat{X}_t\|^{p-2}\|g(\bar{X}_t)\|^2, \tag{A.4}
$$

and

$$
\begin{aligned}
&\frac{1}{2}\sum_{i=1}^{m}\sum_{j=1}^{m} p(p-2)\|\widehat{X}_t\|^{p-4}\widehat{X}_i\widehat{X}_j g_i(\bar{X}_t) g_j(\bar{X}_t)^T \\
&\le \frac{1}{4}\sum_{i=1}^{m}\sum_{j=1}^{m} p(p-2)\|\widehat{X}_t\|^{p-4}\widehat{X}_i^2 g_j(\bar{X}_t) g_j(\bar{X}_t)^T \\
&\qquad\qquad + \frac{1}{4}\sum_{i=1}^{m}\sum_{j=1}^{m} p(p-2)\|\widehat{X}_t\|^{p-4}\widehat{X}_j^2 g_i(\bar{X}_t) g_i(\bar{X}_t)^T \\
&= \frac{1}{2} p(p-2)\|\widehat{X}_t\|^{p-2}\|g(\bar{X}_t)\|^2.
\end{aligned}
\tag{A.5}
$$

Therefore, the definition implies that

$$
\begin{aligned}
\mathrm{d}\|\widehat{X}_t\|^p \le& p\left(\langle \widehat{X}_t, f(\bar{X}_t)\rangle + \frac{(p-1)}{2}\|g(\bar{X}_t)\|^2\right)\|\widehat{X}_t\|^{p-2}\mathrm{d}t \\
&+ p\,\widehat{X}_t^T\, g(\bar{X}_t)\, \|\widehat{X}_t\|^{p-2}\mathrm{d}W_t,
\end{aligned}
\tag{A.6}
$$

from which we can derive that

$$
\begin{aligned}
\|\widehat{X}_t\|^p \le& \|\widehat{X}_0\|^p + \int_0^t p\left(\langle \widehat{X}_s, f(\bar{X}_s)\rangle + \frac{(p-1)}{2}\|g(\bar{X}_s)\|^2\right)\|\widehat{X}_s\|^{p-2}\mathrm{d}s \\
&+ \int_0^t p\,\widehat{X}_s^T\, g(\bar{X}_s)\, \|\widehat{X}_s\|^{p-2}\mathrm{d}W_s.
\end{aligned}
\tag{A.7}
$$

Appendix B

Inequalities

Lemma B.1 (Markov's inequality)**.** *If X is a nonnegative random variable and $a > 0$, then it holds that*

$$\mathbb{P}(X \geq a) \leq \frac{\mathbb{E}[X]}{a}. \tag{B.1}$$

If ϕ is a monotonically increasing function from the nonnegative reals to the nonnegative reals, X is a random variable, $a \geq 0$, and $\phi(a) > 0$, then it holds that

$$\mathbb{P}(|X| \geq a) \leq \frac{\mathbb{E}[\phi(|X|)]}{\phi(a)}. \tag{B.2}$$

In this thesis, we usually take $\phi(x) = x^p$.

Lemma B.2 (Jensen's inequality)**.** *For a real convex function ϕ, real numbers $X_1, X_2, ..., X_n$ in its domain, and positive weights a_i, Jensen's inequality can be stated as*

$$\phi\left(\frac{\sum a_i X_i}{\sum a_i}\right) \leq \frac{\sum a_i \phi(X_i)}{\sum a_i}. \tag{B.3}$$

In this thesis, we have two different forms: $a_i = \frac{1}{N}$, $\sum a_i = 1$ and $\phi(x) = x^p$:

$$\left(\sum_{i=1}^{N} \frac{1}{N} X_i\right)^p \leq \sum_{i=1}^{N} \frac{1}{N} (X_i)^p; \tag{B.4}$$

$a_i = h_i$, $\sum h_i = T$ and $\phi(x) = x^p$:

$$\left(\frac{\sum h_i X_i}{T}\right)^p \leq \frac{\sum h_i (X_i)^p}{T} \Rightarrow \left(\sum h_i X_i\right)^p \leq T^{p-1} \sum h_i (X_i)^p. \tag{B.5}$$

Lemma B.3 (Generalized Jensen's inequality)**.** *Let $(\Omega, \mathcal{F}, \mu)$ be a probability space, such that $\mu(\Omega) = 1$. If g is a real-valued function that is μ-integrable, and if ϕ is a convex function on the real line, then it holds that*

$$\phi\left(\int_\Omega g \, \mathrm{d}\mu\right) \leq \int_\Omega \phi \circ g \, \mathrm{d}\mu, \tag{B.6}$$

that is

$$\phi(\mathbb{E}[X]) \leq \mathbb{E}[\phi(X)]. \tag{B.7}$$

In this thesis, we have two different forms: assume μ to be uniform distribution on $[0, T]$; then,

$$\phi\left(\frac{1}{T} \int_0^T g(t) \, \mathrm{d}t\right) \leq \frac{1}{T} \int_0^T \phi(g(t)) \, \mathrm{d}t; \tag{B.8}$$

Assume μ to be exponential distribution on $[0, T]$,

$$\phi\left(\frac{1}{\int_0^T e^{\gamma t} \, \mathrm{d}t} \int_0^T g(t) e^{\gamma t} \, \mathrm{d}t\right) \leq \frac{1}{\int_0^T e^{\gamma t} \, \mathrm{d}t} \int_0^T \phi(g(t)) e^{\gamma t} \, \mathrm{d}t. \tag{B.9}$$

Lemma B.4 (Hölder's inequality)**.** *For any scalar random variables X, Y,, and any $p, q > 1$ such that $1/p + 1/q = 1$, it holds that*

$$\mathbb{E}[|XY|] \leq \mathbb{E}[|X|^p]^{1/p} \, \mathbb{E}[|Y|^q]^{1/q}. \tag{B.10}$$

Lemma B.5 (Young's inequality)**.** *For $r^{-1} + q^{-1} = 1$, $r > 1$, $q > 1$,*

$$ab \leq \frac{\sigma}{r} a^r + \frac{1}{q \sigma^{\frac{q}{r}}} b^q \quad \forall \, a, b, \sigma > 0. \tag{B.11}$$

In this thesis, we also use three special cases: $r = q = 2$,

$$ab \leq \sigma a^2 + \frac{1}{\sigma} b^2; \tag{B.12}$$

$r = \frac{p}{p-2}$, $q = \frac{p}{2}$,

$$a^{p-2} b^2 \leq \frac{(p-2)\sigma}{p} a^p + \frac{2}{p\sigma^{\frac{p-2}{2}}} b^p, \tag{B.13}$$

and $r = \frac{p}{p-1}$, $q = p$,

$$a^{p-1} b \ \leq \ \frac{(p-1)\sigma}{p} a^p \ + \ \frac{1}{p\sigma^{p-1}} b^p. \tag{B.14}$$

Lemma B.6 (Cauchy-Schwartz inequality)**.** *For all vectors X and Y of an inner product space it holds that*

$$\langle X, Y \rangle \leq \|X\| \|Y\|. \tag{B.15}$$

Lemma B.7 (Doob's inequality)**.** *If M_t is a martingale with $M_0 = 0$, then for any $p > 1$ it holds that*

$$\mathbb{E}\left[|M_T|^p\right] \leq \mathbb{E}\left[\sup_{0 \leq s \leq t} |M_s|^p\right] \leq \left(\frac{p}{p-1}\right)^p \mathbb{E}\left[|M_T|^p\right]. \tag{B.16}$$

Lemma B.8 (Burkholder-Davis-Gundy (BDG) inequality)**.** *If M_t is a martingale with $M_0 = 0$, then for any $p \geq 1$, there are constants c_p, C_p such that*

$$c_p \mathbb{E}\left[|[M]_T|^{\frac{p}{2}}\right] \leq \mathbb{E}\left[\sup_{0 \leq s \leq t} |M_s|^p\right] \leq C_p \mathbb{E}\left[|[M]_T|^{\frac{p}{2}}\right], \tag{B.17}$$

where $[M]_t$ is the quadratic variation process of M_t.

Lemma B.9 (BDG inequality for multi-dimensional Ito integral)**.** *If there is an m-dimensional Ito integral $\int_0^s \varphi(X_u) \mathrm{d}W_u$ satisfying $\int_0^t \mathbb{E}\left[\|\varphi(X_u)\|^p\right] \mathrm{d}u < \infty$, then for $p \geq 1$, there exists a $C_p > 0$ such that*

$$\mathbb{E}\left[\sup_{0 \leq s \leq t} \left| \int_0^s \varphi(X_u) \mathrm{d}W_u \right|^p\right] \ \leq C_p \mathbb{E}\left[\left| \int_0^t \|\varphi(X_u)\|^2 \, \mathrm{d}u \right|^{\frac{p}{2}}\right]. \tag{B.18}$$

Proof of Lemma B.9. By Jensen inequality and Lemma B.8, we obtain that

$$\begin{aligned}
&\mathbb{E}\left[\sup_{0\le s\le t}\left|\int_0^s \varphi(X_u)\mathrm{d}W_u\right|^p\right]\\
\le&\mathbb{E}\left[\sup_{0\le s\le t}\left|\sum_{i=1}^m\int_0^s \varphi_i(X_u)\mathrm{d}W_u^i\right|^p\right] \le m^{p-1}\mathbb{E}\left[\sup_{0\le s\le t}\sum_{i=1}^m\left|\int_0^s \varphi_i(X_u)\mathrm{d}W_u^i\right|^p\right]\\
\le&m^{p-1}\sum_{i=1}^m\mathbb{E}\left[\sup_{0\le s\le t}\left|\int_0^s \varphi_i(X_u)\mathrm{d}W_u^i\right|^p\right] \le m^{p-1}C_p\sum_{i=1}^m\mathbb{E}\left[\left|\int_0^t |\varphi_i(X_u)|^2\,\mathrm{d}u\right|^{\frac{p}{2}}\right]\\
\le&m^{p-1}C_p\sum_{i=1}^m\mathbb{E}\left[\left|\int_0^t \|\varphi(X_u)\|^2\,\mathrm{d}u\right|^{\frac{p}{2}}\right] \le m^pC_p\mathbb{E}\left[\left|\int_0^t \|\varphi(X_u)\|^2\,\mathrm{d}u\right|^{\frac{p}{2}}\right].
\end{aligned} \tag{B.19}$$

This completes the proof of Lemma B.9. □

Lemma B.10 (Grönwall's inequality)**.** *Assume that the continuous functions* $u, \kappa : [0,T] \to [0,\infty)$ *and* $K > 0$ *satisfy that*

$$u(t) \le K + \int_0^t \kappa(s)u(s)\,\mathrm{d}s, \tag{B.20}$$

for all $t \in [0,T]$*. Then the Gronwall's inequality is that*

$$u(t) \le K\exp\left(\int_0^t \kappa(s)\,\mathrm{d}s\right). \tag{B.21}$$

Appendix C

Useful lemmas

Lemma C.1. *If f satisfies Assumption 4.2.2, then for any $x, y, v \in \mathbb{R}^m$, it holds that*

$$\langle v, f(x) - f(y)\rangle = \langle v, \nabla f(x)(x-y)\rangle + R(x, y, v), \tag{C.1}$$

where the remainder term has the bound

$$|R(x, y, v)| \leq L\|v\| \, \|x-y\|^2. \tag{C.2}$$

Proof of Lemma C.1. If we define the scalar function $u(\lambda)$ for $0 \leq \lambda \leq 1$ by

$$u(\lambda) = \langle v, f(y + \lambda(x-y))\rangle, \tag{C.3}$$

then $u(\lambda)$ is continuously differentiable, and by the Mean Value Theorem $u(1) - u(0) = u'(\lambda^*)$ for some $0 < \lambda^* < 1$, which implies that

$$\langle v, f(x) - f(y)\rangle = \langle v, \nabla f(y + \lambda^*(x-y)) \, (x-y)\rangle. \tag{C.4}$$

The final result then follows from the Lipschitz property of ∇f. This completes the proof of Lemma C.1. □

Bibliography

[1] Abramov, V., Klebaner, F., and Liptser, R. The Euler-Maruyama approximations for the CEV model. *ArXiv preprint arXiv:1005.0728* (2010).

[2] Andersson, P., and Kohatsu-Higa, A. Unbiased simulation of stochastic differential equations using parametrix expansions. *Bernoulli 23*, 3 (2017), 2028–2057.

[3] Araujo, V., Pacifico, M., Pujals, E., and Viana, M. Singular-hyperbolic attractors are chaotic. *Transactions of the American Mathematical Society 361*, 5 (2009), 2431–2485.

[4] Assaraf, R., Jourdain, B., Lelièvre, T., and Roux, R. Computation of sensitivities for the invariant measure of a parameter dependent diffusion. *Stochastics and Partial Differential Equations: Analysis and Computations 6*, 2 (2018), 125–183.

[5] Barlow, M. T., and Yor, M. Semi-martingale inequalities via the Garsia-Rodemich-Rumsey lemma, and applications to local times. *Journal of Functional Analysis 49*, 2 (1982), 198–229.

[6] Barrett, J. W., and Süli, E. Existence of global weak solutions to some regularized kinetic models for dilute polymers. *SIAM Multiscale Modelling and Simulation 6*, 2 (2007), 506–546.

[7] Beccari, M., Hutzenthaler, M., Jentzen, A., Kurniawan, R., Lindner, F., and Salimova, D. Strong and weak divergence of expo-

nential and linear-implicit Euler approximations for stochastic partial differential equations with superlinearly growing nonlinearities. *arXiv preprint arXiv:1903.06066* (2019).

[8] Beskos, A., and Roberts, G. O. Exact simulation of diffusions. *The Annals of Applied Probability 15*, 4 (2005), 2422–2444.

[9] Beyn, W.-J., Isaak, E., and Kruse, R. Stochastic C-stability and B-consistency of explicit and implicit Euler-type schemes. *Journal of Scientific Computing 67*, 3 (2016), 955–987.

[10] Beyn, W.-J., Isaak, E., and Kruse, R. Stochastic C-stability and B-consistency of explicit and implicit Milstein-type schemes. *Journal of Scientific Computing 70*, 3 (2017), 1042–1077.

[11] Brosse, N., Durmus, A., Moulines, É., and Sabanis, S. The tamed unadjusted Langevin algorithm. *Stochastic Processes and their Applications* (2018).

[12] Burrage, P. M., and Burrage, K. A variable stepsize implementation for stochastic differential equations. *SIAM Journal on Scientific Computing 24*, 3 (2003), 848–864.

[13] Burrage, P. M., Herdiana, R., and Burrage, K. Adaptive stepsize based on control theory for stochastic differential equations. *Journal of Computational and Applied Mathematics 170*, 2 (2004), 317–336.

[14] Chan, J. H., and Joshi, M. Fast Monte Carlo Greeks for financial products with discontinuous pay-offs. *Mathematical Finance 23*, 3 (2013), 459–495.

[15] Chater, M., Ni, A., and Wang, Q. Simplified least squares shadowing sensitivity analysis for chaotic ODEs and PDEs. *Journal of Computational Physics 329* (2017), 126–140.

[16] Chen, M., and Li, S. Coupling methods for multidimensional diffusion processes. *The Annals of Probability 17*, 1 (1989), 151–177.

[17] Chen, N., and Glasserman, P. Malliavin greeks without Malliavin calculus. *Stochastic Processes and their Applications 117*, 11 (2007), 1689–1723.

[18] Da Prato, G., and Zabczyk, J. *Stochastic equations in infinite dimensions*. Cambridge university press, 2014.

[19] Denny, M. Introduction to importance sampling in rare-event simulations. *European Journal of Physics 22*, 4 (2001), 403.

[20] Eberle, A., Guillin, A., and Zimmer, R. Couplings and quantitative contraction rates for Langevin dynamics. *The Annals of Probability* (2017).

[21] Fang, W., and Giles, M. Adaptive Euler–Maruyama method for SDEs with non-globally Lipschitz drift. In *International Conference on Monte Carlo and Quasi-Monte Carlo Methods in Scientific Computing* (2016), Springer, pp. 217–234.

[22] Fang, W., and Giles, M. Adaptive Euler-Maruyama method for SDEs with non-globally Lipschitz drift: Part I, finite time interval. *arXiv preprint arXiv:1609.08101* (2016).

[23] Fang, W., and Giles, M. Adaptive Euler-Maruyama method for SDEs with non-globally Lipschitz drift: Part II, infinite time interval. *arXiv preprint arXiv:1703.06743* (2017).

[24] Fang, W., and Giles, M. Adaptive Euler-Maruyama method for SDEs with non-globally Lipschitz drift. *The Annals of Applied Probability* (2019).

[25] Fang, W., and Giles, M. Multilevel Monte Carlo method for ergodic SDEs without contractivity. *Journal of Mathematical Analysis and Applications* (2019).

[26] Fournié, E., Lasry, J., Lebuchoux, J., Lions, P., and Touzi, N. Applications of Malliavin calculus to Monte Carlo methods in finance. *Finance and Stochastics 3*, 4 (1999), 391–412.

[27] Freidlin, M. I., and Wentzell, A. D. Random perturbations. In *Random Perturbations of Dynamical Systems*. Springer, 1998, pp. 15–43.

[28] Gaines, J. G., and Lyons, T. J. Variable step size control in the numerical solution of stochastic differential equations. *SIAM Journal on Applied Mathematics 57*, 5 (1997), 1455–1484.

[29] Gasparotto, R. Optimised importance sampling in multilevel Monte Carlo. Master's thesis, University of Oxford, 2015.

[30] Giles, M. B. Multilevel Monte Carlo path simulation. *Operations Research 56*, 3 (2008), 607–617.

[31] Giles, M. B. Vibrato Monte Carlo sensitivities. In *Monte Carlo and Quasi-Monte Carlo Methods 2008*. Springer, 2009, pp. 369–382.

[32] Giles, M. B. Multilevel Monte Carlo methods. *Acta Numerica 24* (2015), 259–328.

[33] Giles, M. B., and Bernal, F. Multilevel estimation of expected exit times and other functionals of stopped diffusions. *SIAM/ASA Journal on Uncertainty Quantification 6*, 4 (2018), 1454–1474.

[34] Giles, M. B., and Glasserman, P. Smoking adjoints: Fast Monte Carlo greeks. *Risk 19*, 1 (2006), 88–92.

[35] Giles, M. B., Lester, C., and Whittle, J. Non-nested adaptive timesteps in multilevel Monte Carlo computations. In *Monte Carlo and Quasi-Monte Carlo Methods 2014*. Springer, 2016, pp. 303–314.

[36] Glasserman, P. *Monte Carlo Methods in Financial Engineering*, vol. 53. Springer Science & Business Media, 2013.

[37] Glynn, P. W. Likelilood ratio gradient estimation: an overview. In *Proceedings of the 19th conference on Winter simulation* (1987), ACM, pp. 366–375.

[38] Glynn, P. W., and Rhee, C. Exact estimation for Markov chain equilibrium expectations. *Journal of Applied Probability 51* (2014), 377–389.

[39] Gobet, E., and Menozzi, S. Stopped diffusion processes: boundary corrections and overshoot. *Stochastic Processes and their Applications 120*, 2 (2010), 130–162.

[40] Guo, Q., Liu, W., Mao, X., and Yue, R. The truncated Milstein method for stochastic differential equations with commutative noise. *Journal of Computational and Applied Mathematics 338* (2018), 298–310.

[41] Hansen, N. R. Geometric ergodicity of discrete-time approximations to multivariate diffusions. *Bernoulli 9*, 4 (2003), 725–743.

[42] Higham, D. J., Mao, X., and Stuart, A. M. Strong convergence of Euler-type methods for nonlinear stochastic differential equations. *SIAM Journal on Numerical Analysis 40*, 3 (2002), 1041–1063.

[43] Hofmann, N., Müller-Gronbach, T., and Ritter, K. Optimal approximation of stochastic differential equations by adaptive step-size control. *Mathematics of Computation 69*, 231 (2000), 1017–1034.

[44] Hofmann, N., Müller-Gronbach, T., and Ritter, K. The optimal discretization of stochastic differential equations. *Journal of Complexity 17*, 1 (2001), 117–153.

[45] Hutzenthaler, M., and Jentzen, A. Non-globally Lipschitz counterexamples for the stochastic Euler scheme. *ArXiv preprint arXiv:0905.0273* (2009).

[46] Hutzenthaler, M., and Jentzen, A. Numerical approximations of stochastic differential equations with non-globally Lipschitz continuous coefficients. *MEMOIRS of the American Mathematical Society 236*, 1112 (2015).

[47] Hutzenthaler, M., Jentzen, A., and Kloeden, P. E. 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 (2011), 1563–1576.

[48] Hutzenthaler, M., Jentzen, A., and Kloeden, P. E. Strong convergence of an explicit numerical method for SDEs with nonglobally Lipschitz continuous coefficients. *The Annals of Applied Probability 22*, 4 (2012), 1611–1641.

[49] Hutzenthaler, M., Jentzen, A., and Wang, X. Exponential integrability properties of numerical approximation processes for nonlinear stochastic differential equations. *Mathematics of Computation 87*, 311 (2018), 1353–1413.

[50] Ilie, S., Jackson, K. R., and Enright, W. H. Adaptive time-stepping for the strong numerical solution of stochastic differential equations. *Numerical Algorithms 68*, 4 (2015), 791–812.

[51] Jameson, A. Aerodynamic design via control theory. *Journal of scientific computing 3*, 3 (1988), 233–260.

[52] Juneja, S., and Shahabuddin, P. Rare-event simulation techniques: an introduction and recent advances. *Handbooks in Operations Research and Management Science 13* (2006), 291–350.

[53] Karatzas, I., and Shreve, S. *Brownian Motion and Stochastic Calculus*, vol. 113. Springer Science & Business Media, 2012.

[54] Kebaier, A., and Lelong, J. Coupling importance sampling and multilevel Monte Carlo using sample average approximation. *Methodology and Computing in Applied Probability 20*, 2 (2017), 1–31.

[55] Kelly, C., and Lord, G. J. Adaptive time-stepping strategies for nonlinear stochastic systems. *IMA Journal of Numerical Analysis 38*, 3 (2017), 1523–1549.

[56] Khasminskii, R. *Stochastic Stability of Differential Equations*, vol. 66. Springer Science & Business Media, 2011.

[57] Kifer, Y. Random perturbations of dynamical systems. In *Nonlinear Problems in Future Particle Accelerators*. World Scientific, 1988, p. 189.

[58] Kloeden, P. E., and Platen, E. *Numerical Solution of Stochastic Differential Equations*. Springer, Berlin, 1992.

[59] Lamba, H. An adaptive timestepping algorithm for stochastic differential equations. *Journal of Computational and Applied Mathematics 161*, 2 (2003), 417–430.

[60] Lamba, H., Mattingly, J. C., and Stuart, A. M. An adaptive Euler-Maruyama scheme for SDEs: convergence and stability. *IMA Journal of Numerical Analysis 27*, 3 (2007), 479–506.

[61] Lamba, H., and Seaman, T. Mean-square stability properties of an adaptive time-stepping SDE solver. *Journal of Computational and Applied Mathematics 194*, 2 (2006), 245–254.

[62] Lea, D. J., Allen, M. R., and Haine, T. W. N. Sensitivity analysis of the climate of a chaotic system. *Tellus A: Dynamic Meteorology and Oceanography 52*, 5 (2000), 523–532.

[63] L'Ecuyer, P. A unified view of the IPA, SF, and LR gradient estimation techniques. *Management Science 36*, 11 (1990), 1364–1383.

[64] Lemaire, V. An adaptive scheme for the approximation of dissipative systems. *Stochastic Processes and their Applications 117*, 10 (2007), 1491–1518.

[65] Lemaire, V., and Pagès, G. Multilevel Richardson–Romberg extrapolation. *Bernoulli 23*, 4A (2017), 2643–2692.

[66] Liu, W., and Mao, X. Strong convergence of the stopped Euler–Maruyama method for nonlinear stochastic differential equations. *Applied Mathematics and Computation 223* (2013), 389–400.

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

[68] Mao, X. The truncated Euler-Maruyama method for stochastic differential equations. *Journal of Computational and Applied Mathematics 290* (2015), 370–384.

[69] Mao, X. Convergence rates of the truncated Euler-Maruyama method for stochastic differential equations. *Journal of Computational and Applied Mathematics 296* (2016), 362–375.

[70] Mao, X., and Szpruch, L. Strong convergence and stability of implicit numerical methods for stochastic differential equations with non-globally Lipschitz continuous coefficients. *Journal of Computational and Applied Mathematics 238* (2013), 14–28.

[71] Mao, X., and Szpruch, L. Strong convergence rates for backward Euler–Maruyama method for non-linear dissipative-type stochastic differential equations with super-linear diffusion coefficients. *Stochastics: An International Journal of Probability and Stochastic Processes 85*, 1 (2013), 144–171.

[72] Maruyama, G. Continuous Markov processes and stochastic equations. *Rendiconti del Circolo Matematico di Palermo 4*, 1 (1955), 48.

[73] Mattingly, J. C., Stuart, A. M., and Higham, D. J. Ergodicity for SDEs and approximations: locally Lipschitz vector fields and degenerate noise. *Stochastic Processes and their Applications 101*, 2 (2002), 185–232.

[74] Mauthner, S. Step size control in the numerical solution of stochastic differential equations. *Journal of Computational and Applied Mathematics 100*, 1 (1998), 93–109.

[75] Meyn, S. P., and Tweedie, R. L. Stability of Markovian processes III: Foster-Lyapunov criteria for continuous-time processes. *Advances in Applied Probability 25*, 3 (1993), 518–548.

[76] Milstein, G. N. *Numerical Integration of Stochastic Differential Equations*, vol. 313. Springer Science & Business Media, 1994.

[77] Milstein, G. N., and Tretyakov, M. V. Numerical integration of stochastic differential equations with nonglobally Lipschitz coefficients. *SIAM Journal on Numerical Analysis 43*, 3 (2005), 1139–1154.

[78] Milstein, G. N., and Tretyakov, M. V. Computing ergodic limits for Langevin equations. *Physica D: Nonlinear Phenomena 229*, 1 (2007), 81–95.

[79] Müller-Gronbach, T. Optimal pointwise approximation of SDEs based on Brownian motion at discrete points. *The Annals of Applied Probability 14*, 4 (2004), 1605–1642.

[80] Neuenkirch, A., Szölgyenyi, M., and Szpruch, L. An adaptive Euler-Maruyama scheme for stochastic differential equations with discontinuous drift and its convergence analysis. *ArXiv preprint arXiv:1802.04521* (2018).

[81] Pagès, G. Multi-step Richardson-Romberg extrapolation: remarks on variance control and complexity. *Monte Carlo Methods and Applications 13*, 1 (2007), 37–70.

[82] Pagès, G., and Panloup, F. Weighted multilevel Langevin simulation of invariant measures. *The Annals of Applied Probability 28*, 6 (2018), 3358–3417.

[83] Pflug, G. C. Derivatives of probability measures-concepts and applications to the optimization of stochastic systems. In *Discrete Event Systems: Models and Applications*. Springer, 1988, pp. 252–274.

[84] Protter, P. E. *Stochastic Integration and Differential equations*. Springer-Verlag, Berlin, 2004.

[85] Roberts, G. O., and Rosenthal, J. S. Optimal scaling of discrete approximations to langevin diffusions. *Journal of the Royal Statistical Society: Series B (Statistical Methodology) 60*, 1 (1998), 255–268.

[86] Roberts, G. O., and Tweedie, R. L. Exponential convergence of Langevin distributions and their discrete approximations. *Bernoulli 2*, 4 (1996), 341–363.

[87] Römisch, W., and Winkler, R. Stepsize control for mean-square numerical methods for stochastic differential equations with small noise. *SIAM Journal on Scientific Computing 28*, 2 (2006), 604–625.

[88] Ruelle, D. Differentiation of SRB states. *Communications in Mathematical Physics 187*, 1 (1997), 227–241.

[89] Ruelle, D. A review of linear response theory for general differentiable dynamical systems. *Nonlinearity 22*, 4 (2009), 855.

[90] Sabanis, S. Euler approximations with varying coefficients: the case of superlinearly growing diffusion coefficients. *The Annals of Applied Probability 26*, 4 (2016), 2083–2105.

[91] Soize, C. *The Fokker-Planck equation for stochastic dynamical systems and its explicit steady state solutions*, vol. 17. World Scientific, 1994.

[92] Stilger, P. S., and Poon, S. Multi-level Monte Carlo simulations with importance sampling. *SSRN: 2273215* (2014).

[93] Talay, D. Second-order discretization schemes of stochastic differential systems for the computation of the invariant law. *Stochastics and Stochastic Reports 29*, 1 (1990), 13–36.

[94] Talay, D. Stochastic Hamiltonian systems: exponential convergence to the invariant measure, and discretization by the implicit Euler scheme. *Markov Processes and Related Fields 8*, 2 (2002), 163–198.

[95] Talay, D., and Tubaro, L. Expansion of the global error for numerical schemes solving stochastic differential equations. *Stochastic Analysis and Applications 8*, 4 (1990), 483–509.

[96] Thepaut, J.-N., and Courtier, P. Four-dimensional variational data assimilation using the adjoint of a multilevel primitive-equation model. *Quarterly Journal of the Royal Meteorological Society 117*, 502 (1991), 1225–1254.

[97] Thuburn, J. Climate sensitivities via a Fokker–Planck adjoint approach. *Quarterly Journal of the Royal Meteorological Society 131*, 605 (2005), 73–92.

[98] Tong, S., and Liu, G. Importance sampling for option Greeks with discontinuous payoffs. *INFORMS Journal on Computing 28*, 2 (2016), 223–235.

[99] Tretyakov, M. V., and Zhang, Z. A fundamental mean-square convergence theorem for SDEs with locally Lipschitz coefficients and its applications. *SIAM Journal on Numerical Analysis 51*, 6 (2013), 3135–3162.

[100] Tucker, W. A rigorous ODE solver and Smale's 14th problem. *Foundations of Computational Mathematics 2*, 1 (2002), 53–117.

[101] Valinejad, A., and Hosseini, S. M. A variable step-size control algorithm for the weak approximation of stochastic differential equations. *Numerical Algorithms 55*, 4 (2010), 429–446.

[102] Wang, Q. Forward and adjoint sensitivity computation of chaotic dynamical systems. *Journal of Computational Physics 235* (2013), 1–13.

[103] Wang, Q. Convergence of the least squares shadowing method for computing derivative of ergodic averages. *SIAM Journal on Numerical Analysis 52*, 1 (2014), 156–170.

[104] Wang, Q., Hu, R., and Blonigan, P. Least squares shadowing sensitivity analysis of chaotic limit cycle oscillations. *Journal of Computational Physics 267* (2014), 210–224.

[105] Wang, X., and Gan, S. The tamed Milstein method for commutative stochastic differential equations with non-globally Lipschitz continuous coefficients. *Journal of Difference Equations and Applications 19*, 3 (2013), 466–490.

[106] Xia, Y., and Giles, M. B. Multilevel path simulation for jump-diffusion SDEs. In *Monte Carlo and Quasi-Monte Carlo Methods 2010*. Springer, 2012, pp. 695–708.

[107] Zong, X., Wu, F., and Huang, C. Convergence and stability of the semi-tamed Euler scheme for stochastic differential equations with non-Lipschitz continuous coefficients. *Applied Mathematics and Computation 228* (2014), 240–250.