



# Multilevel Monte Carlo method for ergodic SDEs without contractivity



Wei Fang\*, Michael B. Giles

Mathematical Institute, University of Oxford, UK

## ARTICLE INFO

### Article history:

Received 15 March 2018  
 Available online 16 January 2019  
 Submitted by A. Jentzen

### Keywords:

SDE  
 Euler–Maruyama  
 Strong convergence  
 Change of measure  
 Invariant measure  
 MLMC

## ABSTRACT

This paper proposes a new multilevel Monte Carlo (MLMC) method for the ergodic SDEs which do not satisfy the contractivity condition. By introducing the change of measure technique, we simulate the path with contractivity and add the Radon–Nikodym derivative to the estimator. We can show the strong error of the path is uniformly bounded with respect to  $T$ . Moreover, the variance of the new level estimators increase linearly in  $T$ , which is a great reduction compared with the exponential increase in standard MLMC. Then the total computational cost is reduced to  $O(\varepsilon^{-2} |\log \varepsilon|^2)$  from  $O(\varepsilon^{-3} |\log \varepsilon|)$  of the standard Monte Carlo method. Numerical experiments support our analysis.

© 2018 Elsevier Inc. All rights reserved.

## 1. Introduction

In this paper we consider an  $m$ -dimensional stochastic differential equation (SDE) driven by an  $m$ -dimensional Brownian motion:

$$dX_t = f(X_t) dt + dW_t, \tag{1}$$

which has a Lipschitz drift  $f : \mathbb{R}^m \rightarrow \mathbb{R}^m$  satisfying the dissipativity condition: for some  $\alpha, \beta > 0$ ,

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

Theorem 6.1 in [23] shows that this class of SDEs is ergodic and solutions converge exponentially to some invariant measure  $\pi$ . Evaluating the expectation of some function  $\varphi(x)$  with respect to that invariant measure  $\pi$  is of great interest in mathematical biology, physics and Bayesian inference in statistics:

$$\pi(\varphi) := \int \varphi(x) d\pi(x) = \lim_{t \rightarrow \infty} \mathbb{E}[\varphi(X_t)], \quad \varphi \in L^1(\pi).$$

\* Corresponding author.

E-mail addresses: wei.fang@maths.ox.ac.uk (W. Fang), mike.giles@maths.ox.ac.uk (M.B. Giles).

Different approaches to computing the expectation include numerical solution of the Fokker–Planck equation, see [27] and the references therein, and estimation of the time average of the ergodic numerical solutions, see [13,19,21–23,25,26,29].

One simple way is to use one of the existing numerical methods for finite time SDEs to simulate the SDE for a sufficiently long time  $T$ , see [7,14,15,18,24] and references therein. The exponential convergence to the invariant measure [23] is given by

$$|\mathbb{E}[\varphi(X_T) - \pi(\varphi)]| \leq \mu^* e^{-\lambda^* T} \quad (3)$$

for some constant  $\mu^*, \lambda^* > 0$ , and bounding this truncation error by  $\varepsilon$  requires

$$T \geq \frac{1}{\lambda^*} \log(\varepsilon^{-1}) + \frac{\log \mu^*}{\lambda^*}, \quad (4)$$

which means the computational cost of each path using uniform time step  $h = O(\varepsilon)$  becomes  $O(\varepsilon^{-1} |\log \varepsilon|)$  for numerical schemes with first order weak convergence. Theorem 1 in [8] shows that under the dissipativity condition (2) the  $p$ -th moments of the numerical solution are bounded uniformly with respect to  $T$ , so the variance of the estimator is bounded by a constant  $V_0$  which does not depend on  $T$ . Therefore, the computational cost to achieve  $\varepsilon^2$  mean square error (MSE) is  $O(\varepsilon^{-3} |\log \varepsilon|)$ .

The multilevel Monte Carlo (MLMC) method, introduced by Giles [10,11], can be applied to reduce the computational cost. If the SDEs further satisfy the contractivity condition: for all  $x, y \in \mathbb{R}^m$ ,

$$\langle x - y, f(x) - f(y) \rangle \leq -\lambda \|x - y\|^2, \quad (5)$$

for some  $\lambda > 0$ , Theorem 3 in [8] has proved first order strong convergence and that the strong error is uniformly bounded with respect to  $T$ . Hence, the variance of the multilevel correction  $V_\ell$  on each level  $\ell$  is bounded by  $Ch_\ell^2$  with  $C > 0$  not depending on  $T$ . The MLMC computational cost to achieve  $\varepsilon^2$  MSE becomes  $O(\varepsilon^{-2} |\log \varepsilon|)$ , where the additional  $O(|\log \varepsilon|)$  comes from the length of simulation time  $T$ . In [8], by simulating different time intervals  $T_\ell$  across different levels  $\ell$ , we further reduce the computational cost to  $O(\varepsilon^{-2})$ .

However, a larger class of SDEs satisfying the dissipativity condition (2) does not satisfy the contractivity condition and instead only satisfies the one-sided Lipschitz condition:

$$\langle x - y, f(x) - f(y) \rangle \leq \lambda \|x - y\|^2, \quad (6)$$

for some  $\lambda > 0$ . The major benefit of the contractivity is that two solutions to the SDE starting from different initial data but driven by the same Brownian motion, will converge exponentially, which means the discretization error from previous time steps will decay exponentially, and then we can prove a uniform bound for the strong error. Without the contractivity, the strong error may increase exponentially with respect to  $T$ . Then multilevel correction variances  $V_\ell$  also increase exponentially, which, as shown in Theorem 5, increases the total computational cost to  $O(\varepsilon^{-2 - \frac{\kappa}{2\lambda^*}} |\log \varepsilon|)$ , where  $\kappa$  is the Lyapunov exponent of the system. For some SDEs with a chaotic property, the Lyapunov exponent  $\kappa$  can be sufficiently large such that  $\frac{\kappa}{2\lambda^*} \geq 1$  and MLMC loses its advantage over the standard Monte Carlo method.

In this paper, a change of measure technique is employed to deal with SDEs satisfying the one-sided Lipschitz condition (6). 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.

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:

$$\begin{aligned} \mathbb{Q}^f : \quad dX_t^f &= f(X_t^f) dt + dW_t^{\mathbb{Q}^f}, \\ \mathbb{Q}^c : \quad dX_t^c &= f(X_t^c) dt + dW_t^{\mathbb{Q}^c}. \end{aligned}$$

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

$$\begin{aligned} dY_t^f &= S(Y_t^c - Y_t^f) dt + f(Y_t^f) dt + dW_t^{\mathbb{P}}, \\ dY_t^c &= S(Y_t^f - Y_t^c) dt + f(Y_t^c) dt + dW_t^{\mathbb{P}}. \end{aligned} \tag{7}$$

The Girsanov theorem gives

$$\mathbb{E}^{\mathbb{Q}^f} [X_t^f] - \mathbb{E}^{\mathbb{Q}^c} [X_t^c] = \mathbb{E}^{\mathbb{P}} \left[ Y_t^f \frac{d\mathbb{Q}^f}{d\mathbb{P}} - Y_t^c \frac{d\mathbb{Q}^c}{d\mathbb{P}} \right], \tag{8}$$

where  $\frac{d\mathbb{Q}^f}{d\mathbb{P}}$  and  $\frac{d\mathbb{Q}^c}{d\mathbb{P}}$  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,

$$\begin{aligned} dW_t^{\mathbb{Q}^f} &= S(Y_t^c - Y_t^f) dt + dW_t^{\mathbb{P}}, \\ dW_t^{\mathbb{Q}^c} &= S(Y_t^f - Y_t^c) dt + dW_t^{\mathbb{P}}, \end{aligned}$$

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

$$d(Y_t^f - Y_t^c) = 2S(Y_t^c - Y_t^f) dt + (f(Y_t^f) - f(Y_t^c)) dt, \tag{9}$$

and provided  $S > \frac{\lambda}{2}$ , Ito’s formula and the one-sided Lipschitz condition (6) give:

$$d\|Y_t^f - Y_t^c\|^2 \leq 2(\lambda - 2S)\|Y_t^f - Y_t^c\|^2 dt,$$

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  $\lambda^*$  of the original SDE and the Lyapunov exponent  $\kappa$ .

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 [11]. To cope with the SDEs with path-dependent jumps, Xia & Giles [30] used a change of measure so that the acceptance probability of the jumps is the same for both fine and coarse paths. Kebaier & Lelong [17] optimize over a class of measures to optimally reduce the variance of MLMC corrections. Andersson & Kohatsu-Higa [1] change the sampling distribution to make the MLMC correction variance finite for unbiased simulation of SDEs using parametrix expansions. Stilger & Poon [28] apply it for MLMC calculation of an interest rate model and Gasparotto [9] for deep out-of-the-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 [3] and the references therein. Importance sampling (change of measure technique) has also been widely used in rare event simulations, see [5,16] 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. [6] proposed a new coupling method to estimate the theoretical convergence rate for Langevin dynamics. See [4] and its references for further exploration.

The rest of the paper is organised as follows. Section 2 introduces the new MLMC method with the change of measure. Section 3 states the main theorems, and the relevant numerical experiments are provided in section 4. Numerical results for SDEs with non-globally Lipschitz drift are given in section 5. The proofs of the main theorems are deferred to section 6, and finally, section 7 has some conclusions and discusses future extensions.

In this paper we consider the infinite time interval  $[0, \infty)$  and let  $(\Omega, \mathcal{F}, \mathbb{P})$  be a probability space with normal filtration  $(\mathcal{F}_t)_{t \in [0, \infty)}$  corresponding to a  $m$ -dimensional standard Brownian motion  $W_t^{\mathbb{P}} = (W^{(1)}, W^{(2)}, \dots, W^{(m)})_t$ . We denote the vector norm by  $\|v\| \triangleq (|v_1|^2 + |v_2|^2 + \dots + |v_m|^2)^{\frac{1}{2}}$ , the inner product of vectors  $v$  and  $w$  by  $\langle v, w \rangle \triangleq v_1 w_1 + v_2 w_2 + \dots + v_m w_m$ , for any  $v, w \in \mathbb{R}^m$  and the Frobenius matrix norm by  $\|A\| \triangleq \sqrt{\sum_{i,j} A_{i,j}^2}$  for any  $A \in \mathbb{R}^{m \times d}$ .

## 2. New MLMC with change of measure

In this paper, we use the standard Euler–Maruyama method to simulate the original SDE (1) using uniform timestep  $h > 0$  under measure  $\mathbb{P}$ :

$$t_{n+1} = t_n + h, \quad \widehat{X}_{t_{n+1}} = \widehat{X}_{t_n} + f(\widehat{X}_{t_n})h + \Delta W_n^{\mathbb{P}}, \quad (10)$$

where  $\Delta W_n^{\mathbb{P}} \triangleq W_{t_{n+1}}^{\mathbb{P}} - W_{t_n}^{\mathbb{P}}$  for  $n = 0, 1, \dots, N-1$  with  $N = T/h$  and there is fixed initial data  $t_0 = 0$ ,  $\widehat{X}_{t_0} = x_0$ . We use the notation  $\underline{t} \triangleq \max\{t_n : t_n \leq t\}$ ,  $n_t \triangleq \max\{n : t_n \leq t\}$  for the nearest time point before time  $t$ , and its index. We define the piece-wise constant interpolant process  $\overline{X}_t = \widehat{X}_{\underline{t}}$  and also define the standard continuous interpolant [18] as

$$\widehat{X}_t = \widehat{X}_{\underline{t}} + f(\widehat{X}_{\underline{t}})(t - \underline{t}) + (W_t - W_{\underline{t}}).$$

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  $h_0 > 0$  and positive integer  $L$ .

$$\widehat{\varphi}_{std} := N_L^{-1} \sum_{n=1}^{N_L} \varphi(\widehat{X}_T^{L,(n)}). \quad (11)$$

Next, we quickly review the standard MLMC scheme introduced in [10,11]. Instead of directly estimating  $\mathbb{E}^{\mathbb{P}}[\varphi(\widehat{X}_T^L)]$ , we have the following telescoping sum in the same probability measure  $\mathbb{P}$ :

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

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

$$\widehat{\varphi}_{mlmc} := N_0^{-1} \sum_{n=1}^{N_0} \varphi(\widehat{X}_T^{0,(n)}) + \sum_{\ell=1}^L N_{\ell}^{-1} \sum_{n=1}^{N_{\ell}} \left( \varphi(\widehat{X}_T^{f,\ell,(n)}) - \varphi(\widehat{X}_T^{c,\ell-1,(n)}) \right). \tag{12}$$

Now we introduce the new MLMC scheme with change of measure using spring coefficient  $S > 0$ .

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

For level  $\ell > 1$ , we simulate the SDE with the additional spring terms using 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  $n \geq 0$ , we update both paths:

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

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

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

Note that the coarse path updates can be combined to give

$$\widehat{Y}_{t_{2n+2}}^c = \widehat{Y}_{t_{2n}}^c + S(\widehat{Y}_{t_{2n}}^f - \widehat{Y}_{t_{2n}}^c)2h + f(\widehat{Y}_{t_{2n}}^c)2h + \Delta W_{2n}^{\mathbb{P}} + \Delta W_{2n+1}^{\mathbb{P}}.$$

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  $n$ th timestep. Under measure  $\mathbb{P}$ , we have

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

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

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

Then the exact Radon–Nikodym derivative for this single step is

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

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

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

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

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

Numerically we obtain two new measures  $\widehat{\mathbb{Q}}^f$  and  $\widehat{\mathbb{Q}}^c$  with  $\Delta W_n^{\widehat{\mathbb{Q}}^f} = \widehat{S}_n^f h + \Delta W_n^{\mathbb{P}}$  and  $\Delta W_n^{\widehat{\mathbb{Q}}^c} = \widehat{S}_n^c h + \Delta W_n^{\mathbb{P}}$  respectively for all steps on fine and coarse paths, where  $\widehat{S}_n^f$  and  $\widehat{S}_n^c$  are the spring terms on  $n$ th 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  $R_{t_0}^f = R_{t_0}^c = 1$ .
- At odd timesteps  $t_{2n+1} = t_{2n} + h$  for  $n \geq 0$ , we only update  $R^f$ :

$$R_{t_{2n+1}}^f = R_{t_{2n}}^f R\left(\widehat{Y}_{t_{2n+1}}^f, \widehat{Y}_{t_{2n}}^f, S(\widehat{Y}_{t_{2n}}^c - \widehat{Y}_{t_{2n}}^f), h\right).$$

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

$$\begin{aligned} R_{t_{2n+2}}^f &= R_{t_{2n+1}}^f R\left(\widehat{Y}_{t_{2n+2}}^f, \widehat{Y}_{t_{2n+1}}^f, S(\widehat{Y}_{t_{2n+1}}^c - \widehat{Y}_{t_{2n+1}}^f), h\right), \\ R_{t_{2n+2}}^c &= R_{t_{2n}}^c R\left(\widehat{Y}_{t_{2n+2}}^c, \widehat{Y}_{t_{2n}}^c, S(\widehat{Y}_{t_{2n}}^f - \widehat{Y}_{t_{2n}}^c), 2h\right). \end{aligned}$$

Then, after  $N$  steps, we obtain the exact Radon–Nikodym derivatives for the whole path:

$$\begin{aligned} \frac{d\widehat{\mathbb{Q}}^f}{d\mathbb{P}} &= R_T^f = \prod_{n=0}^{N-1} R\left(\widehat{Y}_{t_{n+1}}^f, \widehat{Y}_{t_n}^f, S(\widehat{Y}_{t_n}^c - \widehat{Y}_{t_n}^f), h\right), \\ \frac{d\widehat{\mathbb{Q}}^c}{d\mathbb{P}} &= R_T^c = \prod_{n=0}^{N/2-1} R\left(\widehat{Y}_{t_{2n+2}}^c, \widehat{Y}_{t_{2n}}^c, S(\widehat{Y}_{t_{2n}}^f - \widehat{Y}_{t_{2n}}^c), 2h\right). \end{aligned} \tag{13}$$

Finally, the multilevel correction estimator becomes

$$\varphi(\widehat{Y}_T^f) R_T^f - \varphi(\widehat{Y}_T^c) R_T^c, \tag{14}$$

and the identity we use in the new MLMC is

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

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

$$\begin{aligned} \widehat{\varphi}_{new} := & N_0^{-1} \sum_{n=1}^{N_0} \varphi(\widehat{X}_T^{0,(n)}) \\ & + \sum_{\ell=1}^L N_\ell^{-1} \sum_{n=1}^{N_\ell} \left( \varphi(\widehat{X}_T^{f,\ell,(n)}) R_T^{f,\ell,(n)} - \varphi(\widehat{X}_T^{c,\ell-1,(n)}) R_T^{c,\ell,(n)} \right). \end{aligned} \tag{15}$$

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

### 3. 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 (14) and the resulting MLMC complexity.

**Assumption 1** (*Lipschitz and dissipativity*). Assume  $f$  is globally Lipschitz so that there is a constant  $K > 0$  such that

$$\|f(x) - f(y)\| \leq K \|x - y\|, \tag{16}$$

for all  $x, y \in \mathbb{R}^m$ . Furthermore, there exist constants  $\tilde{\alpha}, \tilde{\beta} > 0$  such that for all  $x \in \mathbb{R}^m$ ,  $f$  satisfies the dissipativity condition:

$$\langle x, f(x) \rangle \leq -\tilde{\alpha} \|x\|^2 + \tilde{\beta}. \tag{17}$$

Note that a consequence of the Lipschitz condition is that

$$\|f(x)\| \leq \|f(0)\| + K \|x\| \Rightarrow \|f(x)\|^2 \leq 2 (\|f(0)\|^2 + K^2 \|x\|^2).$$

This assumption ensures the existence and uniqueness of the strong solution to the SDEs [20] and the convergence to the invariant distribution [21]. Note that the Lipschitz assumption is needed for simplicity of the proof but numerical experiments in section 5 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 is uniformly bounded with respect to  $T$ .

**Theorem 1** (*Stability*). If the original SDE satisfies Assumption 1, then using the new change-of-measure algorithm with  $S > 0$ , there exist constants  $C_{(1)}, C_{(2)} > 0$  such that for any  $T > 0$  and  $p \geq 1$ , and for all  $0 < h < C_{(1)}$

$$\sup_{0 \leq n \leq N} \mathbb{E} \left[ \|\widehat{Y}_{t_n}^f\|^p \right]^{1/p} \leq C_{(2)} p^{1/2}, \quad \sup_{0 \leq n \leq N} \mathbb{E} \left[ \|\widehat{Y}_{t_n}^c\|^p \right]^{1/p} \leq C_{(2)} p^{1/2}.$$

**Proof.** The proof is deferred to section 6.1.  $\square$

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 allow us to obtain the uniform bounds. For the first-order strong convergence, we need the following assumption.

**Assumption 2** (One-sided Lipschitz properties). *There exists a constant  $\lambda > 0$  such that for all  $x, y \in \mathbb{R}$ ,  $f$  satisfies the one-sided Lipschitz condition:*

$$\langle x - y, f(x) - f(y) \rangle \leq \lambda \|x - y\|^2, \tag{18}$$

and  $f$  is differentiable and  $\nabla f(x)$  satisfies the Lipschitz condition  $\|\nabla f(x) - \nabla f(y)\| \leq K\|x - y\|$ .

Note that the Lipschitz condition (16) implies this one-sided Lipschitz condition (18). However, the one-sided Lipschitz condition can give a sharper bound for the positive side, which means that  $K$  can be much larger than  $\lambda$ . 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 5 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 spring term  $S$  as shown in the following theorem.

**Theorem 2** (Difference between fine and coarse paths). *If the original SDE satisfies Assumptions 1 and 2, then using the new change-of-measure algorithm with  $S > \lambda/2$ , there exist constants  $C_{(1)}, C_{(2)} > 0$  such that for any  $T > 0$  and  $p \geq 1$ , and for all  $0 < h < C_{(1)}$ ,*

$$\sup_{0 \leq n \leq N} \mathbb{E} \left[ \|\widehat{Y}_{t_n}^f - \widehat{Y}_{t_n}^c\|^p \right]^{1/p} \leq C_{(2)} \min \left( p^{1/2} h^{1/2}, p h \right).$$

**Proof.** The proof is deferred to section 6.2.  $\square$

The  $L_p$  norm of the difference between the fine and coarse paths, as we expected, is uniformly bounded since we add enough spring term to recover the contractivity used in [8]. With this result, we can bound the  $p$ th-moment of the Radon–Nikodym derivatives and then the MLMC estimator (14).

**Theorem 3** (Radon–Nikodym moments). *If the original SDE satisfies Assumptions 1 and 2, then using the new change-of-measure algorithm with  $S > \lambda/2$ , there exist constants  $C_{(1)}, C_{(2)} > 0$  such that, for any  $T > 0$  and  $p \geq 1$ , and for all  $0 < h < \min(C_{(1)}, C_{(2)}/(Tp^2))$ ,*

$$\mathbb{E} \left[ \left| \frac{d\widehat{\mathbb{Q}}^c}{d\mathbb{P}} \right|^p \right] \leq 2, \quad \mathbb{E} \left[ \left| \frac{d\widehat{\mathbb{Q}}^f}{d\mathbb{P}} \right|^p \right] \leq 2.$$

**Proof.** The proof is deferred to section 6.3.  $\square$

**Theorem 4** (MLMC moments). *If the original SDE satisfies Assumptions 1 and 2, and  $\varphi : \mathbb{R}^m \rightarrow \mathbb{R}$  is globally Lipschitz, then using the new change-of-measure algorithm with  $S > \lambda/2$ , for any  $T > 0$  and  $p \geq 1$  there exist constants  $C_{(1)}, C_{(2)}, C_{(3)} > 0$  such that for all  $0 < h < \min(C_{(1)}, C_{(2)}/(Tp^2))$ ,*

$$\mathbb{E} \left[ \left| \varphi(\widehat{Y}_T^f) \frac{d\widehat{\mathbb{Q}}^f}{d\mathbb{P}} - \varphi(\widehat{Y}_T^c) \frac{d\widehat{\mathbb{Q}}^c}{d\mathbb{P}} \right|^p \right]^{1/p} \leq C_{(3)} p^2 \sqrt{T} h.$$

**Proof.** The proof is deferred to section 6.4.  $\square$

Note that this theorem implies that the variance of the estimator (14) is bounded by  $C_2Th^2$  which increases linearly in  $T$ .

We now have everything we require to determine the MLMC complexity.

**Theorem 5** (MLMC for invariant measure). *If  $\varphi$  satisfies the Lipschitz condition and the SDE satisfies Assumption 1 and 2 with convergence rate  $\lambda^*$  and constant  $\mu^*$  in (3), and Lyapunov exponent  $\kappa$ , then by choosing suitable values for  $L$ ,  $S$ ,  $h_0$  and  $N_\ell$  for each level  $\ell$ , there exist constants  $c_1, c_2, c_3 > 0$  such that the estimator  $\widehat{\varphi}$  has a mean square error (MSE) with bound*

$$\mathbb{E} [(\widehat{\varphi} - \pi(\varphi))^2] \leq \varepsilon^2,$$

with  $0 < \varepsilon < 1$  and an expected computational cost  $C_{std}$  for the standard Monte Carlo estimator  $\widehat{\varphi}_{std}$  (11) with bound

$$C_{std} \leq c_1 \varepsilon^{-3} |\log \varepsilon|,$$

and an expected computational cost  $C_{mlmc}$  for the standard MLMC estimator  $\widehat{\varphi}_{mlmc}$  (12) with bound

$$C_{mlmc} \leq c_2 \varepsilon^{-2 - \frac{\kappa}{2\lambda^*}} |\log \varepsilon|,$$

provided  $\kappa/\lambda^* < 2$ , and  $C_{com}$  for the new MLMC estimator with change of measure  $\widehat{\varphi}_{new}$  (15) with bound

$$C_{com} \leq c_3 \varepsilon^{-2} |\log \varepsilon|^2.$$

**Proof.** By Jensen’s inequality, the MSE can be decomposed into three parts:

$$\begin{aligned} \mathbb{E} [(\widehat{\varphi} - \pi(\varphi))^2] &= \mathbb{V} [\widehat{\varphi}] + |\mathbb{E} [\widehat{\varphi}] - \pi(\varphi)|^2 \\ &\leq \mathbb{V} [\widehat{\varphi}] + 2 |\mathbb{E} [\widehat{\varphi}] - \mathbb{E} [\varphi(X_T)]|^2 + 2 |\mathbb{E} [\varphi(X_T)] - \pi(\varphi)|^2, \end{aligned}$$

which enables us to achieve the MSE bound by bounding each part by  $\varepsilon^2/3$ . Similar to (4), we bound the third part by setting

$$T = \frac{1}{\lambda^*} \log(\varepsilon^{-1}) + \frac{\log \sqrt{6}\mu^*}{\lambda^*}, \tag{19}$$

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  $\gamma, \zeta > 0$ .

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

$$C_{std} \leq c_1 \varepsilon^{-3} |\log \varepsilon|,$$

for some constant  $c_1 > 0$ .

The analysis for the two MLMC schemes is similar to the MLMC theorem in [11] and shows the optimal computational cost is bounded by

$$3\varepsilon^{-2} \left( \sum_{\ell=0}^L \sqrt{V_\ell C_\ell} \right)^2 + \sum_{\ell=0}^L C_\ell,$$

where  $C_\ell$  and  $V_\ell$  are the cost and variance for each level.

For standard MLMC, we have first order weak convergence but the variance of  $V_\ell$  for  $\ell \geq 1$  increases exponentially in  $T$ , which gives

$$V_\ell \leq \eta_1 (h_0 2^{-\ell})^2 e^{\kappa T}, \tag{20}$$

for some constant  $\eta_1 > 0$ . A good MLMC coupling requires  $C_0 V_0 > C_1 V_1$ , and given this condition and  $\beta = 2, \gamma = 1$ , the optimal cost is  $O(\varepsilon^{-2} C_0)$ . The condition  $C_0 V_0 > C_1 V_1$  requires

$$h_0 = \vartheta_1 e^{-\kappa T/2} \Rightarrow C_0 = \vartheta_2 \varepsilon^{-\frac{\kappa}{2\lambda^*}} |\log \varepsilon|, \tag{21}$$

for some  $\vartheta_1, \vartheta_2 > 0$ . 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  $c_2$  such that

$$C_{mlmc} \leq c_2 \varepsilon^{-2-\frac{\kappa}{2\lambda^*}} |\log \varepsilon|.$$

For the new MLMC with the change of measure, Theorem 4 gives

$$V_\ell \leq \eta_2 (h_0 2^{-\ell})^2 T, \tag{22}$$

for some  $\eta_2 > 0$ . The condition  $C_0 V_0 > C_1 V_1$  requires

$$h_0 = \vartheta_3 T^{-1/2}, \tag{23}$$

for some  $\vartheta_3 > 0$ , but the bound in Theorem 4 requires the tighter condition

$$h_0 = \vartheta_4 T^{-1} \Rightarrow C_0 = \vartheta_5 |\log \varepsilon|^2,$$

for some  $\vartheta_4, \vartheta_5 > 0$ . Therefore, there exists a constant  $c_3$  such that

$$C_{com} \leq c_3 \varepsilon^{-2} |\log \varepsilon|^2. \quad \square$$

#### 4. Numerical results

In this section, we present the numerical results for a Lipschitz version of the stochastic Lorenz equation with additive noise:

$$f \begin{pmatrix} x_1 \\ x_2 \\ x_3 \end{pmatrix} = \begin{pmatrix} 10(B(x_2) - x_1) \\ (28 - x_3)B(x_1) - x_2 \\ B(x_1)x_2 - \frac{8}{3}x_3 \end{pmatrix},$$

where  $B(x) = 65x/\max(65, |x|)$ . When  $|x_1| > 65$  and  $|x_2| > 65$ , we have

$$f \begin{pmatrix} x_1 \\ x_2 \\ x_3 \end{pmatrix} = \begin{pmatrix} 650 \operatorname{sgn}(x_2) - 10x_1 \\ 65 \operatorname{sgn}(x_1)(28 - x_3) - x_2 \\ 65 \operatorname{sgn}(x_1)x_2 - \frac{8}{3}x_3 \end{pmatrix}.$$

Therefore,  $f$  satisfy the Lipschitz condition (16) and the dissipativity condition (17). In the region of  $|x_1| \leq 65$  and  $|x_2| \leq 65$ , which contains the chaotic attractors, 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 Lipschitz condition. We run 10000 sample paths from  $T = 0$  to 20 to get the following results.

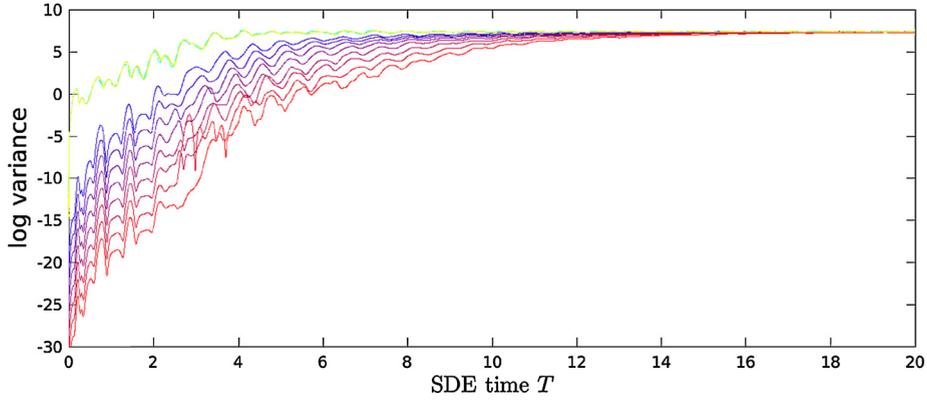
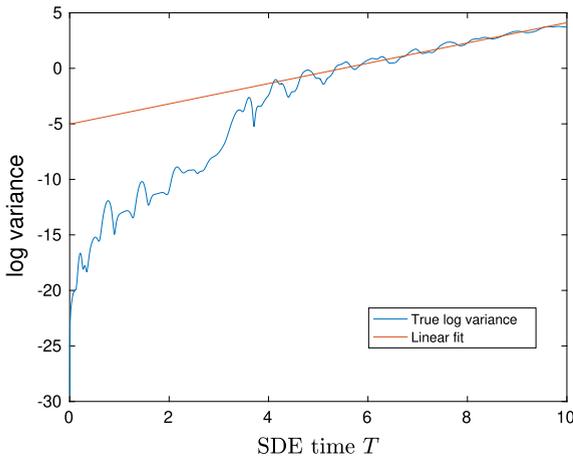
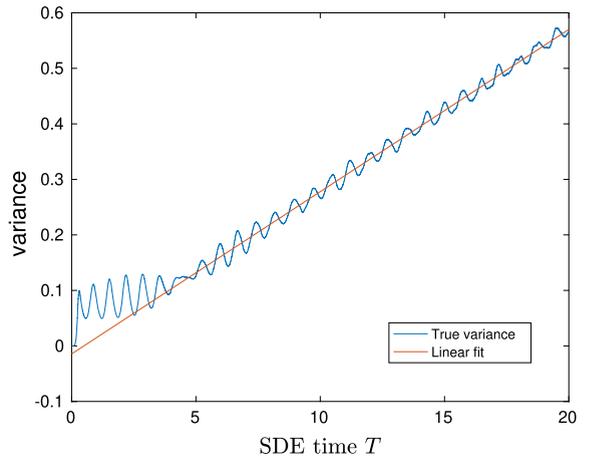


Fig. 1. Variance for each level without change of measure. (For interpretation of the colours in the figure(s), the reader is referred to the web version of this article.)



(a) Linear increase of log variance without change of measure



(b) Linear increase of variance with change of measure

Fig. 2. Variance on level 8 with/without change of measure.

Fig. 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  $h_0 = 2^{-9}$ :

$$\mathbb{V} \left[ \varphi(\widehat{X}_T^c) - \varphi(\widehat{X}_T^f) \right] \sim \eta_1 h_\ell^2 e^{\kappa T},$$

for some  $\eta_1 > 0$ , 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 level increases, the variance decreases at rate 2. For  $T > 10$ , we can see that the standard MLMC on level  $\ell = 8$ , using  $h = 2^{-17}$ , still can not achieve a good coupling. In order to see this exponential increase, we plot the log variance on level 8 using  $h = 2^{-17}$  with respect to  $T$  and the fitted linear function on time interval  $[5, 10]$ , see Fig. 2(a). The  $\kappa$  we fit is 1.36.

Similarly, for the new MLMC with spring term  $S = 10$ , Fig. 3 is the semi-log plot of the variance on each level as a function of  $T$  with change of measure using same  $h_\ell$ :

$$\mathbb{V} \left[ \varphi(\widehat{Y}_T^f) R_T^f - \varphi(\widehat{Y}_T^c) R_T^c \right] \sim \eta_2 h_\ell^2 T,$$

for some  $\eta_2 > 0$ . As the level increases, the variance decreases at a rate 2. In order to see the linear increase

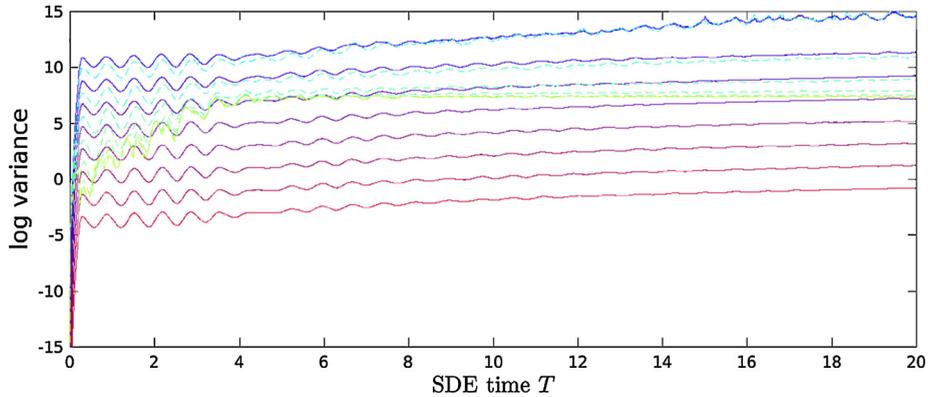


Fig. 3. Variance for each level with change of measure.

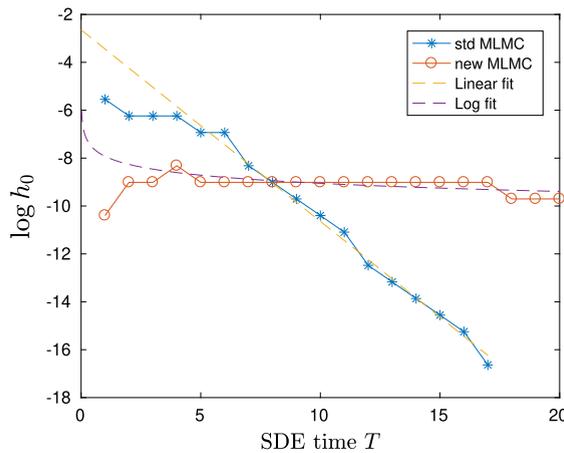


Fig. 4. The required  $h_0$  to achieve a good coupling.

in  $T$ , we plot the variance on level  $\ell = 8$  with respect to  $T$  and the fitted linear function on time interval  $[5, 20]$ , see Fig. 2(b).

We have investigated and illustrated the dependence of  $V_\ell$  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 Fig. 4. The blue line confirms the exponential decrease of  $h_0$  with respect to  $T$  in (21). The coefficient of the log function fit is 0.49 which confirms the relationship (23).

Lastly, we estimate the convergence rate  $\lambda^*$  to the invariant measure. Fig. 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 moving upper bound and moving lower bound) in Fig. 5(b) and the exponential fit. The fitted  $\lambda^*$  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 5.

### 5. 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 [8]. Without any proof, we show some numerical experiments results for the SDE with a double-well potential energy and the stochastic Lorenz equation.

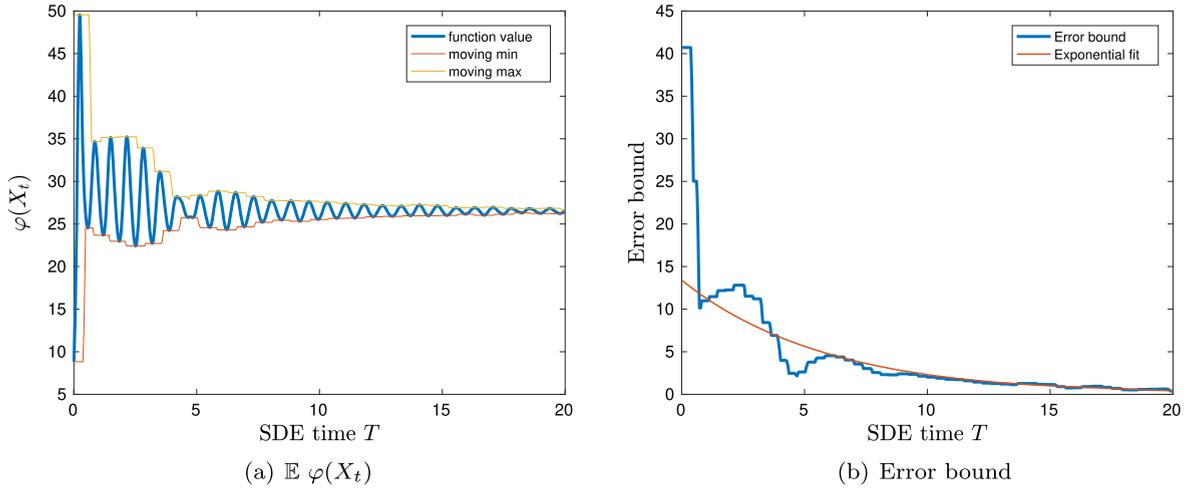


Fig. 5. Estimation of the convergence rate to invariant distribution.

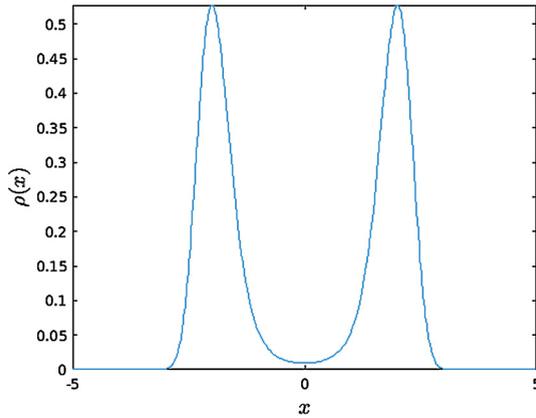


Fig. 6. Probability density function of invariant distribution.

5.1. Double-well potential energy

We consider

$$dX_t = (2X_t - \frac{1}{2}X_t^3) dt + dW_t. \tag{24}$$

The probability density function of its invariant distribution is

$$\frac{\exp(2x^2 - \frac{1}{4}x^4)}{\int_{-\infty}^{\infty} \exp(2x^2 - \frac{1}{4}x^4) dx},$$

and it has two different wells at  $x = \pm 2$ , see Fig. 6. This SDE satisfies the dissipativity condition (17) and one-sided Lipschitz condition (18) 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.

We simulate the SDE with initial value  $x_0 = 0$  to time  $T = 5$ , and use the adaptive function:

$$h^\delta(x) = \frac{\max(1, |x|)}{8 \max(1, |2x - \frac{1}{2}x^3|)} \delta,$$

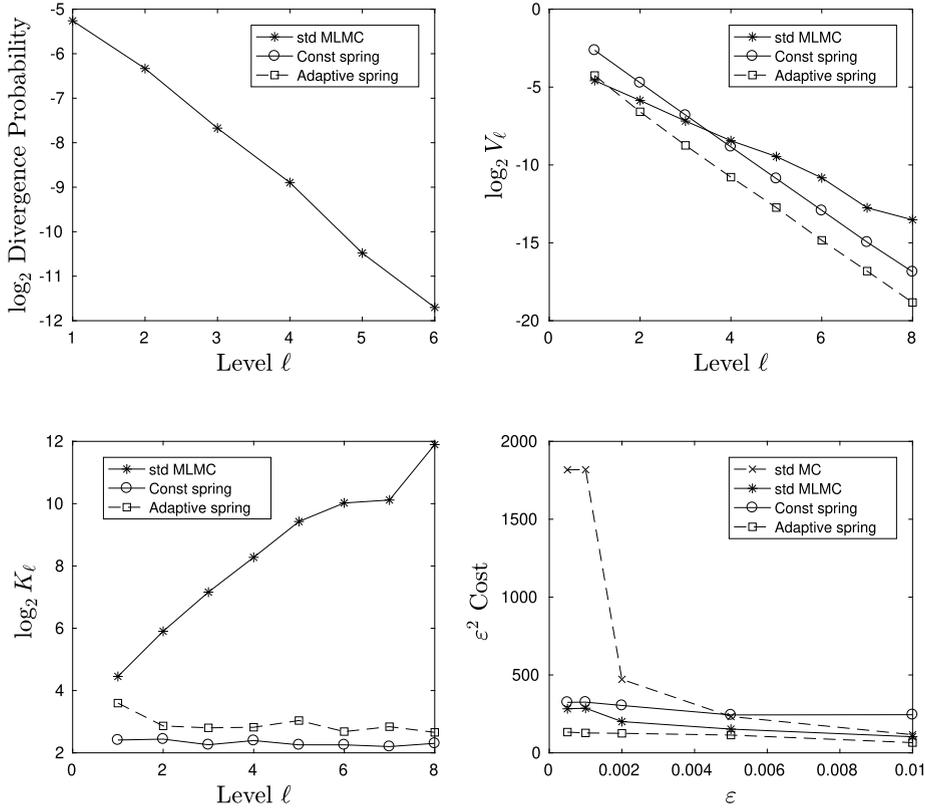


Fig. 7. MLMC convergence test for double-well potential energy.

with  $\delta = 2^{-\ell}$  for each level  $\ell$ . We compare three different schemes:

- standard MLMC with adaptive timestep.
- MLMC with adaptive timestep and change of measure with constant spring coefficient  $S = 1$ .
- MLMC with adaptive timestep and change of measure with adaptive spring coefficient

$$S = \max(0, 2 - 1.5x^2).$$

The second scheme uses  $S = 1$  following the suggestion of Theorem 4. The third scheme improves on the second by choosing adaptive  $S$  and avoiding unnecessary spring term, reducing the variance without losing the control on divergence. By doing first order Taylor expansion on (9), we choose  $S = \max(0, f'(x))$  to deal with the divergence locally.

We run 10000 samples for each level  $\ell$  for the three schemes. The numerical results are shown in Fig. 7.

The top left figure plots the divergence probability with respect to the level  $\ell$ , where the divergence probability is defined as

$$\mathbb{E} \left[ \mathbb{1}_{\|\widehat{X}_T^f - \widehat{X}_T^c\| > 1} \right] = \mathbb{P} \left[ \|\widehat{X}_T^f - \widehat{X}_T^c\| > 1 \right].$$

The probability decreases as  $\ell$  increases since the timestep  $h_\ell$  is smaller and the difference between fine and coarse path decreases. The decrease rate we fit is

$$\mathbb{P} \left[ \|\widehat{X}_T^f - \widehat{X}_T^c\| > 1 \right] \sim O(h_\ell^{1.28}).$$

The two schemes with change of measure have zero divergence on all levels.

The top right figure plots the variance of corrections  $V_\ell$  with respect to level  $\ell$ . The  $V_\ell$  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  $V_\ell$  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  $\ell$ . The kurtosis of standard MLMC will increase exponentially while the kurtosis of the schemes with change of measure will stay constant. Similar intuitive explanation applies here. The divergence samples again dominate the 4th moment and then the kurtosis on each level

$$K_\ell \sim \frac{\mathbb{E} \left[ \|\widehat{X}_T^f - \widehat{X}_T^c\|^4 \right]}{\mathbb{E} \left[ \|\widehat{X}_T^f - \widehat{X}_T^c\|^2 \right]^2} \sim h_\ell^{-1.28}.$$

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  $\varepsilon$ . 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 adaptive spring. They can not only keep the kurtosis constant but also reduce the variance and hence the total computational cost.

### 5.2. Stochastic Lorenz equation

This is a three-dimensional system modelling convection rolls in the atmosphere

$$f \begin{pmatrix} x_1 \\ x_2 \\ x_3 \end{pmatrix} = \begin{pmatrix} 10(x_2 - x_1) \\ x_1(28 - x_3) - x_2 \\ x_1x_2 - \frac{8}{3}x_3 \end{pmatrix}. \tag{25}$$

This SDE does not satisfy the dissipativity condition (17) and one-sided Lipschitz condition (18), and is more chaotic compared with the truncated Lipschitz version in previous section.

We simulate the SDE with initial value  $x_0 = [0, 0, 0]$  to time  $T = 10$ , and use the adaptive function:

$$h^\delta(x) = \frac{\max(100, \|x\|^2)}{2^{11} \max(100, \|f(x)\|^2)} \delta,$$

with  $\delta = 2^{-\ell}$  for each level  $\ell$ . We compare two different schemes:

- standard MLMC with adaptive timestep.
- MLMC with adaptive timestep and change of measure with constant spring coefficient  $S = 10$ .

A possible third scheme is the scheme with adaptive spring which requires us to calculate the largest positive eigenvalue of the Jacobian matrix  $\frac{\partial f}{\partial x}$ .

We run 10000 samples for each level  $\ell$  for two schemes. The numerical results are shown in Fig. 8.

Similarly, the top left figures 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 for all the levels. The rate of decrease for standard MLMC is 0.82.

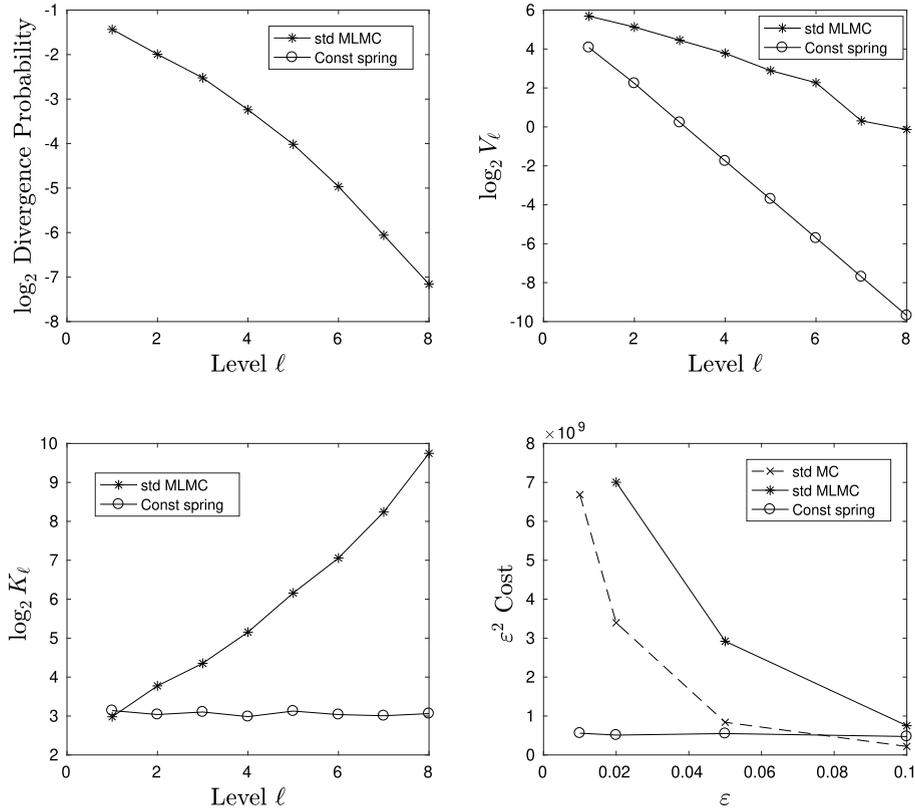


Fig. 8. MLMC convergence test for Lorenz equation.

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  $V_\ell$  is approximately 2 for change of measure and 0.83 for standard MLMC which is similar to the rate of decrease of divergence probability.

The bottom left plot shows that the kurtosis of standard MLMC increases exponentially as level  $\ell$  increases while the kurtosis of change of measure remains constant. The increase rate is 0.96 which is close to the decrease rate of divergence rate.

The last bottom right plot implies that the total computational cost is  $O(\epsilon^{-2})$  for the MLMC with change of measure and  $O(\epsilon^{-3})$  for the standard Monte Carlo method. The  $O(\epsilon^{-3})$  computational cost for standard MLMC is worse than the theoretical results due to the high kurtosis and large variance  $V_\ell$  and it is already quite hard to get the result for  $\epsilon = 0.01$  in a reasonable computational time.

### 6. Proofs

For simplicity of the proof, we introduce the notation  $a(h) \lesssim b(h)$  which means there exists a constant  $\tilde{h}_0 > 0$  such that  $a(h) \leq b(h), \forall 0 < h < \tilde{h}_0$ , 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  $\omega$  or Brownian paths.

Note that for all  $\delta > 0$ , we have  $1/(1 - Sh) \lesssim 1/(1 - 2Sh) \lesssim 1 + 2Sh + \delta h \lesssim 2$ .

#### 6.1. Theorem 1

**Proof.** The proof is given for  $p \geq 4$ ; the result for  $1 \leq p < 4$  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 initial point  $x_0$ , i.e.  $\hat{Y}_{t_0}^f = \hat{Y}_{t_0}^c = x_0$ .

For the odd time point  $t_{2n+1}$  for  $n \geq 0$ ,

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

Squaring both sides gives

$$\|\widehat{Y}_{t_{2n+1}}^c\|^2 = \left\| Sh \widehat{Y}_{t_{2n}}^f + (1 - Sh) \left( \widehat{Y}_{t_{2n}}^c + \frac{f(\widehat{Y}_{t_{2n}}^c)h + \Delta W_{2n}}{1 - Sh} \right) \right\|^2.$$

Due to the convexity of  $x^2$  that, for any  $0 \leq \xi \leq 1$ ,

$$\|\xi A + (1 - \xi)B\|^2 \leq \xi \|A\|^2 + (1 - \xi) \|B\|^2,$$

provided  $h < 1/S$ , we can choose  $\xi = Sh$  to get

$$\begin{aligned} \|\widehat{Y}_{t_{2n+1}}^c\|^2 &\lesssim Sh \|\widehat{Y}_{t_{2n}}^f\|^2 + (1 - Sh) \|\widehat{Y}_{t_{2n}}^c\|^2 + 4 \|\Delta W_{2n}\|^2 + 2 \langle \widehat{Y}_{t_{2n}}^c, f(\widehat{Y}_{t_{2n}}^c) \rangle h \\ &\quad + 4 \|f(\widehat{Y}_{t_{2n}}^c)\|^2 h^2 + 2 \langle \widehat{Y}_{t_{2n}}^c, \Delta W_{2n} \rangle. \end{aligned}$$

Due to the Lipschitz condition (16),

$$\|f(\widehat{Y}_{t_{2n}}^c)\|^2 h^2 \lesssim \gamma h (\|\widehat{Y}_{t_{2n}}^c\|^2 + 1)$$

for any  $\gamma > 0$ . Combining this with dissipativity condition (17), we obtain, for some fixed  $\alpha \in (0, \tilde{\alpha})$  and  $\beta \in (\tilde{\beta}, \infty)$ ,

$$\begin{aligned} \|\widehat{Y}_{t_{2n+1}}^c\|^2 &\lesssim Sh \|\widehat{Y}_{t_{2n}}^f\|^2 + (1 - Sh - 2\alpha h) \|\widehat{Y}_{t_{2n}}^c\|^2 + 4 \|\Delta W_{2n}\|^2 + 2\beta h \\ &\quad + 2 \langle \widehat{Y}_{t_{2n}}^c, \Delta W_{2n} \rangle. \end{aligned} \tag{26}$$

Similarly, we have

$$\begin{aligned} \|\widehat{Y}_{t_{2n+1}}^f\|^2 &\lesssim Sh \|\widehat{Y}_{t_{2n}}^c\|^2 + (1 - Sh - 2\alpha h) \|\widehat{Y}_{t_{2n}}^f\|^2 + 4 \|\Delta W_{2n}\|^2 + 2\beta h \\ &\quad + 2 \langle \widehat{Y}_{t_{2n}}^f, \Delta W_{2n} \rangle. \end{aligned} \tag{27}$$

For the even point  $t_{2n+2}$  for  $n \geq 0$ ,

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

Using the same approach and choosing  $\xi = 2Sh$  provided  $2Sh < 1$ , we get

$$\begin{aligned} \|\widehat{Y}_{t_{2n+2}}^c\|^2 &\lesssim 2Sh \|\widehat{Y}_{t_{2n}}^f\|^2 + (1 - 2Sh - 4\alpha h) \|\widehat{Y}_{t_{2n}}^c\|^2 + 4 \|\Delta W_{2n} + \Delta W_{2n+1}\|^2 \\ &\quad + 4\beta h + 2 \langle \widehat{Y}_{t_{2n}}^c, \Delta W_{2n} + \Delta W_{2n+1} \rangle, \end{aligned}$$

and

$$\begin{aligned} \|\widehat{Y}_{t_{2n+2}}^f\|^2 &\lesssim Sh\|\widehat{Y}_{t_{2n+1}}^c\|^2 + (1 - Sh - 2\alpha h)\|\widehat{Y}_{t_{2n+1}}^f\|^2 + 4\|\Delta W_{2n+1}\|^2 + 2\beta h \\ &\quad + 2\langle \widehat{Y}_{t_{2n+1}}^f, \Delta W_{2n+1} \rangle. \end{aligned}$$

Therefore, for any fixed  $\gamma \in (0, \alpha)$ , we have

$$\begin{aligned} \|\widehat{Y}_{t_{2n+2}}^c\|^2 + \|\widehat{Y}_{t_{2n+2}}^f\|^2 &\lesssim (1 - 4\gamma h)(\|\widehat{Y}_{t_{2n}}^c\|^2 + \|\widehat{Y}_{t_{2n}}^f\|^2) + 12(\|\Delta W_{2n}\|^2 + \|\Delta W_{2n+1}\|^2) \\ &\quad + 8\beta h + 2e^{-4\gamma h}\langle \phi_{t_{2n}}, \Delta W_{2n} \rangle + 2e^{-2\gamma h}\langle \phi_{t_{2n+1}}, \Delta W_{2n+1} \rangle, \end{aligned}$$

where for  $n \in [0, N/2 - 1]$ ,

$$e^{-4\gamma h}\phi_{t_{2n}} = (1 + Sh)\widehat{Y}_{t_{2n}}^c + (1 - Sh - 2\alpha h)\widehat{Y}_{t_{2n}}^f, \quad e^{-2\gamma h}\phi_{t_{2n+1}} = \widehat{Y}_{t_{2n}}^c + \widehat{Y}_{t_{2n+1}}^f.$$

Since  $1 - 4\gamma h \leq e^{-4\gamma h}$  and  $e^{4\gamma h} \lesssim 2$ , we multiply by  $e^{2\gamma t_{2n+2}}$  on both sides to obtain

$$\begin{aligned} e^{2\gamma t_{2n+2}}(\|\widehat{Y}_{t_{2n+2}}^f\|^2 + \|\widehat{Y}_{t_{2n+2}}^c\|^2) &\lesssim e^{2\gamma t_{2n}}(\|\widehat{Y}_{t_{2n}}^f\|^2 + \|\widehat{Y}_{t_{2n}}^c\|^2) + 16\beta e^{2\gamma t_{2n}}h \\ &\quad + 24e^{2\gamma t_{2n}}(\|\Delta W_{2n}\|^2 + \|\Delta W_{2n+1}\|^2) + 2e^{2\gamma t_{2n}}\langle \phi_{t_{2n}}, \Delta W_{2n} \rangle \\ &\quad + 2e^{2\gamma t_{2n+1}}\langle \phi_{t_{2n+1}}, \Delta W_{2n+1} \rangle. \end{aligned}$$

Summing over multiple timesteps gives

$$\begin{aligned} e^{2\gamma t_{2n}}(\|\widehat{Y}_{t_{2n}}^f\|^2 + \|\widehat{Y}_{t_{2n}}^c\|^2) &\lesssim (\|\widehat{Y}_{t_0}^f\|^2 + \|\widehat{Y}_{t_0}^c\|^2) + 24\sum_{k=0}^{2n-1} e^{2\gamma t_k}\|\Delta W_k\|^2 \\ &\quad + 16\beta\sum_{k=0}^{n-1} e^{2\gamma t_{2k}}h + 2\sum_{k=0}^{2n-1} e^{2\gamma t_k}\langle \phi_{t_k}, \Delta W_k \rangle. \end{aligned} \tag{28}$$

For odd time points, combining (26) and (27), by Cauchy–Schwarz inequality and Young’s inequality, there exist constants  $\alpha_1 > 1$ ,  $\beta_1 > \max(1, \alpha_1\beta)$  such that

$$\begin{aligned} \|\widehat{Y}_{t_{2n+1}}^f\|^2 + \|\widehat{Y}_{t_{2n+1}}^c\|^2 &\lesssim (1 - 2\alpha h)(\|\widehat{Y}_{t_{2n}}^f\|^2 + \|\widehat{Y}_{t_{2n}}^c\|^2) + 8\|\Delta W_{2n}\|^2 + 4\beta h \\ &\quad + 2\langle \widehat{Y}_{t_{2n}}^c, \Delta W_{2n} \rangle + 2\langle \widehat{Y}_{t_{2n}}^f, \Delta W_{2n} \rangle \\ &\lesssim (1 - 2\gamma h)\left(\alpha_1(\|\widehat{Y}_{t_{2n}}^f\|^2 + \|\widehat{Y}_{t_{2n}}^c\|^2) + 12\|\Delta W_{2n}\|^2\right) + 4\beta_1 h. \end{aligned}$$

Multiplying by  $e^{2\gamma t_{2n+1}}$  on both sides and using the (28) gives

$$\begin{aligned} e^{2\gamma t_{2n+1}}(\|\widehat{Y}_{t_{2n+1}}^f\|^2 + \|\widehat{Y}_{t_{2n+1}}^c\|^2) &\lesssim \alpha_1(\|\widehat{Y}_{t_0}^f\|^2 + \|\widehat{Y}_{t_0}^c\|^2) + 24\alpha_1\sum_{k=0}^{2n} e^{2\gamma t_k}\|\Delta W_k\|^2 \\ &\quad + 16\beta_1\sum_{k=0}^n e^{2\gamma t_{2k}}h + 2\alpha_1\sum_{k=0}^{2n-1} e^{2\gamma t_k}\langle \phi_{t_k}, \Delta W_k \rangle. \end{aligned} \tag{29}$$

Then, combining (28) and (29), raising both sides to power  $p/2$ , taking the supremum over  $n \in [0, N]$  and taking expectation on the both sides, by using Jensen’s inequality, we have

$$\mathbb{E}\left[\sup_{0 \leq n \leq N} e^{\gamma p t_n} \left(\|\widehat{Y}_{t_n}^f\|^2 + \|\widehat{Y}_{t_n}^c\|^2\right)^{p/2}\right] \lesssim 4^{p/2-1}(24\alpha_1\beta_1)^{p/2}(I_1 + I_2 + I_3 + I_4),$$

where

$$I_1 = \mathbb{E} \left[ (\|\widehat{Y}_{t_0}^f\|^2 + \|\widehat{Y}_{t_0}^c\|^2)^{p/2} \right] = 2^{p/2} \|x_0\|^{p/2}, \quad I_2 = \left| \sum_{k=0}^{N-1} e^{2\gamma t_k} h \right|^{p/2},$$

$$I_3 = \mathbb{E} \left[ \left| \sum_{k=0}^{N-1} e^{2\gamma t_k} \|\Delta W_k\|^2 \right|^{p/2} \right], \quad I_4 = \mathbb{E} \left[ \sup_{1 \leq n \leq N/2} \left| \sum_{k=0}^{2n-1} e^{2\gamma t_k} \langle \phi_{t_k}, \Delta W_k \rangle \right|^{p/2} \right].$$

We will bound these four parts separately.  $I_1$  is a constant. For  $I_2$ , we have

$$I_2 \leq \left| \int_0^T e^{2\gamma t} dt \right|^{p/2} \leq e^{\gamma p T} / (2\gamma)^{p/2}.$$

Next, if  $q_k$ , for  $k = 1, \dots, n$  is an arbitrary discrete probability distribution, and  $b_k$ , for  $k = 1, \dots, n$  is a set of scalar values, then for any  $p > 1$  Jensen’s inequality gives

$$\left| \sum_{k=1}^n q_k b_k \right|^p \leq \sum_{k=1}^n q_k |b_k|^p.$$

If  $a_k, k = 1, \dots, n$  is a set of positive scalar values, then setting  $q_k = a_k / \sum_{k'=1}^n a_{k'}$  gives

$$\left| \sum_{k=1}^n a_k b_k \right|^p \leq \left| \sum_{k=1}^n a_k \right|^{p-1} \sum_{k=1}^n a_k |b_k|^p.$$

For  $I_3$ , using this inequality, we obtain

$$I_3 = \mathbb{E} \left[ \left| \sum_{k=0}^{N-1} e^{2\gamma t_k} h \frac{\|\Delta W_k\|^2}{h} \right|^{p/2} \right] \leq \left| \sum_{k=0}^{N-1} e^{2\gamma t_k} h \right|^{p/2-1} \mathbb{E} \left[ \sum_{k=0}^{N-1} e^{2\gamma t_k} h \frac{\|\Delta W_k\|^p}{h^{p/2}} \right]$$

$$\leq c_p e^{\gamma p T} / (2\gamma)^{p/2},$$

where  $c_p$  is defined by  $c_p = \mathbb{E} [\|\Delta W_k\|^p / h^{p/2}] \leq d^{p/2} p!! \leq d^{p/2} p^{p/2}$ .

For  $I_4$ , we rewrite the summation as an Itô integral and then by the Burkholder–Davis–Gundy inequality in [2], there exists a positive constant  $C_{BDG}$  independent of  $p$  such that

$$I_4 \leq \mathbb{E} \left[ \sup_{0 \leq t \leq T} \left| \int_0^t e^{2\gamma \lfloor s/h \rfloor h} \langle \phi_{\lfloor s/h \rfloor h}, dW_s \rangle \right|^{p/2} \right]$$

$$\leq (C_{BDG} p)^{p/4} \mathbb{E} \left[ \left| \sum_{k=0}^{N-1} e^{4\gamma t_k} \|\phi_{t_k}\|^2 h \right|^{p/4} \right],$$

where, by Young’s inequality,

$$\|\phi_{t_{2k}}\|^2 \lesssim 8(\|\widehat{Y}_{t_{2k}}^c\|^2 + \|\widehat{Y}_{t_{2k}}^f\|^2), \quad \|\phi_{t_{2k+1}}\|^2 \lesssim 4(\|\widehat{Y}_{t_{2k}}^c\|^2 + \|\widehat{Y}_{t_{2k+1}}^f\|^2).$$

Then by Jensen’s inequality and Young’s inequality, for arbitrary  $\zeta > 0$ , we have

$$\begin{aligned}
 I_4 &\leq (12C_{\text{BDG}}p)^{p/4} \left[ \sum_{k=0}^{N-1} e^{2\gamma t_k} h \right]^{p/4-1} \mathbb{E} \left[ \sum_{k=0}^{N-1} e^{2\gamma t_k} h \sup_{0 \leq n \leq N} e^{\gamma p t_n / 2} (\|\widehat{Y}_{t_n}^c\|^2 + \|\widehat{Y}_{t_n}^f\|^2)^{p/4} \right] \\
 &\leq \frac{1}{4\zeta} \mathbb{E} \left[ \sup_{0 \leq n \leq N} e^{\gamma p t_n} \left( \|\widehat{Y}_{t_n}^f\|^2 + \|\widehat{Y}_{t_n}^c\|^2 \right)^{p/2} \right] + \zeta \left( \frac{6C_{\text{BDG}}}{\gamma} \right)^{p/2} p^{p/2} e^{\gamma p T}.
 \end{aligned}$$

Finally, combining all the estimates above and choosing  $\zeta = 4^{p/2-1}(24\alpha_1\beta_1)^{p/2}$ , there exists a constant  $C_{(2)} > 0$  such that

$$\mathbb{E} \left[ \sup_{0 \leq n \leq N} e^{\gamma p t_n} \left( \|\widehat{Y}_{t_n}^f\|^2 + \|\widehat{Y}_{t_n}^c\|^2 \right)^{p/2} \right] \lesssim C_{(2)}^p p^{p/2} e^{\gamma p T},$$

which implies that there exists constant  $C_{(1)} > 0$  such that, for any  $0 < h \leq C_{(1)}$ ,

$$\sup_{0 \leq n \leq N} \mathbb{E} \left[ \|\widehat{Y}_{t_n}^f\|^p \right] \leq C_{(2)}^p p^{p/2}, \quad \sup_{0 \leq n \leq N} \mathbb{E} \left[ \|\widehat{Y}_{t_n}^c\|^p \right] \leq C_{(2)}^p p^{p/2}. \quad \square$$

### 6.2. Theorem 2

**Proof.** The proof is given for  $p \geq 4$ ; the result for  $1 \leq p < 4$  follows from Hölder’s inequality.

The different updates on odd and even time points give

$$\begin{aligned}
 \widehat{Y}_{t_{2n+1}}^f - \widehat{Y}_{t_{2n+1}}^c &= (1 - 2Sh)(\widehat{Y}_{t_{2n}}^f - \widehat{Y}_{t_{2n}}^c) + (f(\widehat{Y}_{t_{2n}}^f) - f(\widehat{Y}_{t_{2n}}^c))h, \\
 \widehat{Y}_{t_{2n+2}}^f - \widehat{Y}_{t_{2n+2}}^c &= (1 - Sh)(\widehat{Y}_{t_{2n+1}}^f - \widehat{Y}_{t_{2n+1}}^c) - Sh(\widehat{Y}_{t_{2n}}^f - \widehat{Y}_{t_{2n}}^c) + (f(\widehat{Y}_{t_{2n+1}}^f) - f(\widehat{Y}_{t_{2n}}^c))h,
 \end{aligned}$$

and then

$$\begin{aligned}
 \widehat{Y}_{t_{2n+2}}^f - \widehat{Y}_{t_{2n+2}}^c &= (1 - 4Sh + 2S^2h^2)(\widehat{Y}_{t_{2n}}^f - \widehat{Y}_{t_{2n}}^c) + (1 - Sh)(f(\widehat{Y}_{t_{2n}}^f) - f(\widehat{Y}_{t_{2n}}^c))h \\
 &\quad + (f(\widehat{Y}_{t_{2n+1}}^f) - f(\widehat{Y}_{t_{2n}}^c))h.
 \end{aligned}$$

Taking the square of both sides gives

$$\begin{aligned}
 \|\widehat{Y}_{t_{2n+2}}^f - \widehat{Y}_{t_{2n+2}}^c\|^2 &= (1 - 4Sh + 2S^2h^2)^2 \|\widehat{Y}_{t_{2n}}^f - \widehat{Y}_{t_{2n}}^c\|^2 \\
 &\quad + (1 - Sh)^2 \|f(\widehat{Y}_{t_{2n}}^f) - f(\widehat{Y}_{t_{2n}}^c)\|^2 h^2 + \|f(\widehat{Y}_{t_{2n+1}}^f) - f(\widehat{Y}_{t_{2n}}^c)\|^2 h^2 \\
 &\quad + 2(1 - 4Sh + 2S^2h^2)(1 - Sh) \langle \widehat{Y}_{t_{2n}}^f - \widehat{Y}_{t_{2n}}^c, f(\widehat{Y}_{t_{2n}}^f) - f(\widehat{Y}_{t_{2n}}^c) \rangle h \\
 &\quad + 2(1 - 4Sh + 2S^2h^2) \langle \widehat{Y}_{t_{2n}}^f - \widehat{Y}_{t_{2n}}^c, f(\widehat{Y}_{t_{2n+1}}^f) - f(\widehat{Y}_{t_{2n}}^c) \rangle h \\
 &\quad + 2(1 - Sh) \langle f(\widehat{Y}_{t_{2n}}^f) - f(\widehat{Y}_{t_{2n}}^c), f(\widehat{Y}_{t_{2n+1}}^f) - f(\widehat{Y}_{t_{2n}}^c) \rangle h^2.
 \end{aligned}$$

Then provided  $S > \lambda/2$ , the Assumption 2, Lipschitz condition (16), Cauchy–Schwarz inequality and Young’s inequality imply, for any fixed  $\gamma \in (0, 2S - \lambda)$ ,

$$\begin{aligned}
 \|\widehat{Y}_{t_{2n+2}}^f - \widehat{Y}_{t_{2n+2}}^c\|^2 &\lesssim (1 - 4\gamma h) \|\widehat{Y}_{t_{2n}}^f - \widehat{Y}_{t_{2n}}^c\|^2 + 2K^2 \|\widehat{Y}_{t_{2n+1}}^f - \widehat{Y}_{t_{2n}}^f\|^2 h^2 \\
 &\quad + 2(1 - 4Sh + 2S^2h^2) \langle \widehat{Y}_{t_{2n}}^f - \widehat{Y}_{t_{2n}}^c, f(\widehat{Y}_{t_{2n+1}}^f) - f(\widehat{Y}_{t_{2n}}^c) \rangle h. \tag{30}
 \end{aligned}$$

Following this estimate, we use two different approach to get different upper bounds.

**First**, we continue to use Young’s inequality and Lipschitz condition (16) to get

$$\|\widehat{Y}_{t_{2n+2}}^f - \widehat{Y}_{t_{2n+2}}^c\|^2 \lesssim (1 - 2\gamma h)\|\widehat{Y}_{t_{2n}}^f - \widehat{Y}_{t_{2n}}^c\|^2 + ((2\gamma)^{-1} + 2) K^2\|\widehat{Y}_{t_{2n+1}}^f - \widehat{Y}_{t_{2n}}^f\|^2 h.$$

Then we multiply by  $e^{\gamma t_{2n+2}}$  on both sides and  $e^{2\gamma h} \lesssim 2$  gives

$$e^{\gamma t_{2n+2}}\|\widehat{Y}_{t_{2n+2}}^f - \widehat{Y}_{t_{2n+2}}^c\|^2 \lesssim e^{\gamma t_{2n}}\|\widehat{Y}_{t_{2n}}^f - \widehat{Y}_{t_{2n}}^c\|^2 + (\gamma^{-1} + 4)K^2 e^{\gamma t_{2n}}\|\widehat{Y}_{t_{2n+1}}^f - \widehat{Y}_{t_{2n}}^f\|^2 h.$$

Summing over multiple timesteps and noting  $\widehat{Y}_{t_0}^f - \widehat{Y}_{t_0}^c = 0$  gives

$$e^{\gamma t_{2n}}\|\widehat{Y}_{t_{2n}}^f - \widehat{Y}_{t_{2n}}^c\|^2 \lesssim (\gamma^{-1} + 4)K^2 \sum_{k=0}^{n-1} e^{\gamma t_{2k}}\|\widehat{Y}_{t_{2k+1}}^f - \widehat{Y}_{t_{2k}}^f\|^2 h.$$

Then raising both sides to the power  $p/2$ , taking supremum over  $n \in [0, N/2]$ , taking expectation and by Jensen’s inequality, we obtain

$$\begin{aligned} \mathbb{E} \left[ \sup_{0 \leq n \leq N/2} e^{\gamma p t_{2n}/2} \|\widehat{Y}_{t_{2n}}^f - \widehat{Y}_{t_{2n}}^c\|^p \right] &\lesssim (\gamma^{-1} + 4)^{p/2} K^p \mathbb{E} \left[ \left| \sum_{k=0}^{N/2-1} e^{\gamma t_{2k}} \|\widehat{Y}_{t_{2k+1}}^f - \widehat{Y}_{t_{2k}}^f\|^2 h \right|^{p/2} \right] \\ &\leq (\gamma^{-1} + 4)^{p/2} K^p \left| \sum_{k=0}^{N/2-1} e^{\gamma t_{2k}} h \right|^{p/2-1} \mathbb{E} \left[ \sum_{k=0}^{N/2-1} e^{\gamma t_{2k}} \|\widehat{Y}_{t_{2k+1}}^f - \widehat{Y}_{t_{2k}}^f\|^p h \right] \\ &\leq (\gamma^{-1} + 4)^{p/2} K^p \left| \sum_{k=0}^{N/2-1} e^{\gamma t_{2k}} h \right|^{p/2-1} \sum_{k=0}^{N/2-1} e^{\gamma t_{2k}} \mathbb{E} \left[ \|\widehat{Y}_{t_{2k+1}}^f - \widehat{Y}_{t_{2k}}^f\|^p \right] h. \end{aligned}$$

By the update on fine path, Lipschitz condition (16), Theorem 1 and Jensen’s inequality, there exists a constant  $C_1 > 0$  such that

$$\begin{aligned} \mathbb{E} \left[ \|\widehat{Y}_{t_{2k+1}}^f - \widehat{Y}_{t_{2k}}^f\|^p \right] &= \mathbb{E} \left[ \|f(\widehat{Y}_{t_{2k}}^f)h + S(\widehat{Y}_{t_{2k}}^c - \widehat{Y}_{t_{2k}}^f)h + \Delta W_{2k}\|^p \right] \\ &\leq 2^{p-1} \mathbb{E} \left[ \|f(\widehat{Y}_{t_{2k}}^f) + S(\widehat{Y}_{t_{2k}}^c - \widehat{Y}_{t_{2k}}^f)\|^p \right] h^p + 2^{p-1} \mathbb{E} \left[ \|\Delta W_{2k}\|^p \right] \\ &\lesssim C_1^p p^{p/2} h^{p/2}, \end{aligned} \tag{31}$$

which implies that

$$\mathbb{E} \left[ \sup_{0 \leq n \leq N/2} e^{\gamma p t_{2n}/2} \|\widehat{Y}_{t_{2n}}^f - \widehat{Y}_{t_{2n}}^c\|^p \right] \lesssim 2^{-p/2} (\gamma^{-1} + 4)^{p/2} K^p C_1^p p^{p/2} e^{\gamma p T/2} h^{p/2}. \tag{32}$$

**Second**, we directly multiply by  $e^{2\gamma t_{2n+2}}$  on both sides of (30) and  $e^{4\gamma h} \lesssim 2$  gives

$$\begin{aligned} e^{2\gamma t_{2n+2}}\|\widehat{Y}_{t_{2n+2}}^f - \widehat{Y}_{t_{2n+2}}^c\|^2 &\lesssim e^{2\gamma t_{2n}}\|\widehat{Y}_{t_{2n}}^f - \widehat{Y}_{t_{2n}}^c\|^2 + 4e^{2\gamma t_{2n}}K^2\|\widehat{Y}_{t_{2n+1}}^f - \widehat{Y}_{t_{2n}}^f\|^2 h^2 \\ &\quad + 2(1 - 4Sh + 2S^2h^2) e^{2\gamma t_{2n+2}} \langle \widehat{Y}_{t_{2n}}^f - \widehat{Y}_{t_{2n}}^c, f(\widehat{Y}_{t_{2n+1}}^f) - f(\widehat{Y}_{t_{2n}}^f) \rangle h. \end{aligned}$$

Summing over multiple timesteps and noting that  $\widehat{Y}_{t_0}^f - \widehat{Y}_{t_0}^c = 0$  gives

$$e^{2\gamma t_{2n}} \|\widehat{Y}_{t_{2n}}^f - \widehat{Y}_{t_{2n}}^c\|^2 \lesssim 4 \sum_{k=0}^{n-1} e^{2\gamma t_{2k}} K^2 \|\widehat{Y}_{t_{2k+1}}^f - \widehat{Y}_{t_{2k}}^f\|^2 h^2 + 2(1 - 4Sh + 2S^2h^2) \sum_{k=0}^{n-1} e^{2\gamma t_{2k+2}} \langle \widehat{Y}_{t_{2k}}^f - \widehat{Y}_{t_{2k}}^c, f(\widehat{Y}_{t_{2k+1}}^f) - f(\widehat{Y}_{t_{2k}}^f) \rangle h.$$

Then raising both sides to the power  $p/2$ , taking supremum over  $n \in [0, N/2]$ , taking expectation and by Jensen’s inequality, we obtain

$$\mathbb{E} \left[ \sup_{0 \leq n \leq N/2} e^{\gamma p t_{2n}} \|\widehat{Y}_{t_{2n}}^f - \widehat{Y}_{t_{2n}}^c\|^p \right] \lesssim 2^{p/2-1} 4^{p/2} (I_1 + I_2),$$

where

$$I_1 = \mathbb{E} \left[ \left| \sum_{k=0}^{N/2-1} e^{2\gamma t_{2k}} K^2 \|\widehat{Y}_{t_{2k+1}}^f - \widehat{Y}_{t_{2k}}^f\|^2 h^2 \right|^{p/2} \right],$$

$$I_2 = \mathbb{E} \left[ \sup_{0 \leq n \leq N/2-1} \left| \sum_{k=0}^n e^{2\gamma t_{2k}} \langle \widehat{Y}_{t_{2k}}^f - \widehat{Y}_{t_{2k}}^c, f(\widehat{Y}_{t_{2k+1}}^f) - f(\widehat{Y}_{t_{2k}}^f) \rangle h \right|^{p/2} \right].$$

For  $I_1$ , Jensen’s inequality and the estimate (31) give

$$I_1 \leq \left| \sum_{k=0}^{N/2-1} e^{2\gamma t_{2k}} h \right|^{p/2-1} \sum_{k=0}^{N/2-1} e^{2\gamma t_{2k}} h K^p \mathbb{E} \left[ \|\widehat{Y}_{t_{2k+1}}^f - \widehat{Y}_{t_{2k}}^f\|^p \right] h^{p/2} \lesssim (2\gamma)^{-p/2} K^p C_1^p p^{p/2} e^{\gamma p T} h^p.$$

For  $I_2$ , we perform a Taylor expansion and by mean value theorem obtain

$$\begin{aligned} \langle \widehat{Y}_{t_{2k}}^f - \widehat{Y}_{t_{2k}}^c, f(\widehat{Y}_{t_{2k+1}}^f) - f(\widehat{Y}_{t_{2k}}^f) \rangle &= \langle \widehat{Y}_{t_{2k}}^f - \widehat{Y}_{t_{2k}}^c, \nabla f(\widehat{Y}_{t_{2k}}^f)(\widehat{Y}_{t_{2k+1}}^f - \widehat{Y}_{t_{2k}}^f) \rangle + R_k \\ &= \langle \widehat{Y}_{t_{2k}}^f - \widehat{Y}_{t_{2k}}^c, \nabla f(\widehat{Y}_{t_{2k}}^f) f(\widehat{Y}_{t_{2k}}^f) \rangle h + \langle \widehat{Y}_{t_{2k}}^f - \widehat{Y}_{t_{2k}}^c, \nabla f(\widehat{Y}_{t_{2k}}^f) \Delta W_{2k} \rangle \\ &\quad + S \langle \widehat{Y}_{t_{2k}}^f - \widehat{Y}_{t_{2k}}^c, \nabla f(\widehat{Y}_{t_{2k}}^f)(\widehat{Y}_{t_{2k}}^c - \widehat{Y}_{t_{2k}}^f) \rangle h + R_k, \end{aligned}$$

where  $|R_k| \leq 2K \|\widehat{Y}_{t_{2k}}^f - \widehat{Y}_{t_{2k}}^c\| \|\widehat{Y}_{t_{2k+1}}^f - \widehat{Y}_{t_{2k}}^f\|^2$  and then by Jensen’s inequality, we have

$$I_2 = 4^{p/2-1} (J_1 + J_2 + J_3 + J_4),$$

where

$$J_1 = \mathbb{E} \left[ \sup_{0 \leq n \leq N/2-1} \left| \sum_{k=0}^n e^{2\gamma t_{2k}} \langle \widehat{Y}_{t_{2k}}^f - \widehat{Y}_{t_{2k}}^c, \nabla f(\widehat{Y}_{t_{2k}}^f) f(\widehat{Y}_{t_{2k}}^f) \rangle h^2 \right|^{p/2} \right],$$

$$J_2 = \mathbb{E} \left[ \sup_{0 \leq n \leq N/2-1} \left| \sum_{k=0}^n e^{2\gamma t_{2k}} S \langle \widehat{Y}_{t_{2k}}^f - \widehat{Y}_{t_{2k}}^c, \nabla f(\widehat{Y}_{t_{2k}}^f)(\widehat{Y}_{t_{2k}}^c - \widehat{Y}_{t_{2k}}^f) \rangle h^2 \right|^{p/2} \right],$$

$$J_3 = \mathbb{E} \left[ \sup_{0 \leq n \leq N/2-1} \left| \sum_{k=0}^n e^{2\gamma t_{2k}} \langle \widehat{Y}_{t_{2k}}^f - \widehat{Y}_{t_{2k}}^c, \nabla f(\widehat{Y}_{t_{2k}}^f) \Delta W_{2k} \rangle h \right|^{p/2} \right],$$

$$J_4 = \mathbb{E} \left[ \sup_{0 \leq n \leq N/2-1} \left| \sum_{k=0}^n e^{2\gamma t_{2k}} R_k h \right|^{p/2} \right].$$

For  $J_1$ , by Cauchy–Schwarz inequality, Jensen’s inequality, Young’s inequality, the Lipschitz property of  $f$  and  $\nabla f$  and Theorem 1, for any  $\zeta > 0$ , there exists a constant  $C_{31} > 0$  such that

$$J_1 \leq \mathbb{E} \left[ \left| \sum_{k=0}^{N/2-1} e^{2\gamma kh} h \right|^{p/2-1} \sum_{k=0}^{N/2-1} e^{2\gamma kh} h e^{\gamma p t_{2k}/2} \|\widehat{Y}_{t_{2k}}^f - \widehat{Y}_{t_{2k}}^c\|^{p/2} \|\nabla f(\widehat{Y}_{t_{2k}}^f) f(\widehat{Y}_{t_{2k}}^f)\|^{p/2} h^{p/2} \right]$$

$$\leq \mathbb{E} \left[ \left| \sup_{0 \leq n \leq N/2} e^{\gamma p t_{2n}/2} \|\widehat{Y}_{t_{2n}}^f - \widehat{Y}_{t_{2n}}^c\|^{p/2} \right| \left| \sum_{k=0}^{N/2-1} e^{2\gamma kh} h \right|^{p/2-1} \sum_{k=0}^{N/2-1} e^{2\gamma kh} h \|\nabla f(\widehat{Y}_{t_{2k}}^f) f(\widehat{Y}_{t_{2k}}^f)\|^{p/2} h^{p/2} \right]$$

$$\leq \frac{1}{4\zeta} \mathbb{E} \left[ \sup_{0 \leq n \leq N/2} e^{\gamma p t_{2n}} \|\widehat{Y}_{t_{2n}}^f - \widehat{Y}_{t_{2n}}^c\|^p \right] + \zeta \mathbb{E} \left[ \left| \sum_{k=0}^{N/2-1} e^{2\gamma kh} h \right|^{p-1} \sum_{k=0}^{N/2-1} e^{2\gamma kh} h \|\nabla f(\widehat{Y}_{t_{2k}}^f) f(\widehat{Y}_{t_{2k}}^f)\|^p h^p \right]$$

$$\leq \frac{1}{4\zeta} \mathbb{E} \left[ \sup_{0 \leq n \leq N/2} e^{\gamma p t_{2n}} \|\widehat{Y}_{t_{2n}}^f - \widehat{Y}_{t_{2n}}^c\|^p \right] + \zeta C_{31}^p p^p e^{\gamma p T} h^p.$$

Similarly, for  $J_2$ , there exists a constant  $C_{32} > 0$  such that, for any  $\zeta > 0$ ,

$$J_2 \leq \frac{1}{4\zeta} \mathbb{E} \left[ \sup_{0 \leq n \leq N/2} e^{\gamma p t_{2n}} \|\widehat{Y}_{t_{2n}}^f - \widehat{Y}_{t_{2n}}^c\|^p \right] + \zeta C_{32}^p p^p e^{\gamma p T} h^p.$$

Then, for  $J_3$ , by the Burkholder–Davis–Gundy inequality in [2], the Lipschitz property of  $\nabla f$  and Theorem 1, there exists a constant  $C_{33} > 0$  such that, for any  $\zeta > 0$ ,

$$J_3 \leq (C_{BDG} p)^{p/4} \mathbb{E} \left[ \left| \sum_{k=0}^{N/2-1} e^{4\gamma t_{2k}} \|\widehat{Y}_{t_{2k}}^f - \widehat{Y}_{t_{2k}}^c\|^2 \|\nabla f(\widehat{Y}_{t_{2k}}^f)\|^2 h^3 \right|^{p/4} \right]$$

$$\leq (C_{BDG} p)^{p/4} \mathbb{E} \left[ \left| \sup_{0 \leq n \leq N/2} e^{\frac{\gamma p t_{2n}}{2}} \|\widehat{Y}_{t_{2n}}^f - \widehat{Y}_{t_{2n}}^c\|^{\frac{p}{2}} \right| \left| \sum_{k=0}^{N/2-1} e^{2\gamma t_{2k}} h \right|^{\frac{p}{4}-1} \sum_{k=0}^{N/2-1} e^{2\gamma t_{2k}} h \|\nabla f(\widehat{Y}_{t_{2k}}^f)\|^{\frac{p}{2}} h^{\frac{p}{2}} \right]$$

$$\leq \frac{1}{4\zeta} \mathbb{E} \left[ \sup_{0 \leq n \leq N/2} e^{\gamma p t_{2n}} \|\widehat{Y}_{t_{2n}}^f - \widehat{Y}_{t_{2n}}^c\|^p \right] + \zeta C_{33}^p p^p e^{\gamma p T} h^p.$$

Similarly, for  $J_4$ , by Jensen’s inequality, Young inequality, Lipschitz property of  $f$  and  $\nabla f$  and Theorem 1, for any  $\zeta > 0$ , there exists a constant  $C_{34} > 0$  such that, for any  $\zeta > 0$ ,

$$J_4 \leq \mathbb{E} \left[ \left| \sum_{k=0}^{N/2-1} e^{2\gamma t_{2k}} h \right|^{p/2-1} \sum_{k=0}^{N/2-1} e^{2\gamma t_{2k}} h e^{\gamma p t_{2k}/2} \|\widehat{Y}_{t_{2k}}^f - \widehat{Y}_{t_{2k}}^c\|^{p/2} K^{p/2} \|(\widehat{Y}_{t_{2k+1}}^f - \widehat{Y}_{t_{2k}}^f)\|^p \right]$$

$$\begin{aligned}
 &\leq \mathbb{E} \left[ \left| \sup_{0 \leq n \leq N/2} e^{\gamma p t_{2n}/2} \|\widehat{Y}_{t_{2n}}^f - \widehat{Y}_{t_{2n}}^c\|^{p/2} \left| \sum_{k=0}^{N/2-1} e^{2\gamma t_{2k} h} \right|^{p/2-1} \sum_{k=0}^{N/2-1} e^{2\gamma t_{2k} h} h K^{p/2} \|\widehat{Y}_{t_{2k+1}}^f - \widehat{Y}_{t_{2k}}^f\|^p \right| \right] \\
 &\leq \frac{1}{4\zeta} \mathbb{E} \left[ \sup_{0 \leq n \leq N/2} e^{\gamma p t_{2n}} \|\widehat{Y}_{t_{2n}}^f - \widehat{Y}_{t_{2n}}^c\|^p \right] \\
 &\quad + \zeta \mathbb{E} \left[ \left| \sum_{k=0}^{N/2-1} e^{2\gamma t_{2k} h} \right|^{p-1} \sum_{k=0}^{N/2-1} e^{2\gamma t_{2k} h} h K^p 2^{2p-1} \left( \|f(\widehat{Y}_{t_{2k}}^f) + S(\widehat{Y}_{t_{2k}}^c - \widehat{Y}_{t_{2k}}^f)\|^2 h^{2p} + \|\Delta W_{2k}\|^2 \right) \right] \\
 &\lesssim \frac{1}{4\zeta} \mathbb{E} \left[ \sup_{0 \leq n \leq N/2} e^{\gamma p t_{2n}} \|\widehat{Y}_{t_{2n}}^f - \widehat{Y}_{t_{2n}}^c\|^p \right] + \zeta C_{34}^p p^p e^{\gamma p T} h^p.
 \end{aligned}$$

Finally, by choosing  $\zeta = 2^{5p/2-2}$ , there exists a constant  $C_4 > 0$  such that

$$\mathbb{E} \left[ \sup_{0 \leq n \leq N/2} e^{\gamma p t_{2n}} \|\widehat{Y}_{t_{2n}}^f - \widehat{Y}_{t_{2n}}^c\|^p \right] \lesssim C_4^p p^p e^{\gamma p T} h^p,$$

which together with (32) implies that there exists a constant  $C_5 > 0$  such that

$$\sup_{0 \leq n \leq N/2} \mathbb{E} \left[ \|\widehat{Y}_{t_{2n}}^f - \widehat{Y}_{t_{2n}}^c\|^p \right] \lesssim C_5^p \min \left( p^{p/2} h^{p/2}, p^p h^p \right).$$

For the odd time points, Lipschitz condition (16) and (18) gives

$$\begin{aligned}
 \|\widehat{Y}_{t_{2n+1}}^f - \widehat{Y}_{t_{2n+1}}^c\|^2 &\leq ((1 - 2Sh)^2 + 2h\lambda(1 - 2Sh) + K^2 h^2) \|\widehat{Y}_{t_{2n}}^f - \widehat{Y}_{t_{2n}}^c\|^2 \\
 &\lesssim 2 \|\widehat{Y}_{t_{2n}}^f - \widehat{Y}_{t_{2n}}^c\|^2,
 \end{aligned} \tag{33}$$

which, by raising both sides to power  $p/2$  and taking expectation, there exist constants  $C_{(1)}, C_{(2)} > 0$  such that,  $\forall 0 < h < C_{(1)}$ ,

$$\sup_{0 \leq n \leq N} \mathbb{E} \left[ \|\widehat{Y}_{t_n}^f - \widehat{Y}_{t_n}^c\|^p \right] \leq C_{(2)}^p \min \left( p^{p/2} h^{p/2}, p^p h^p \right). \quad \square$$

### 6.3. Theorem 3

**Proof.** For simplicity, we only show the proof for  $\frac{d\widehat{\mathbb{Q}}^c}{d\mathbb{P}}$ , and the result for  $\frac{d\widehat{\mathbb{Q}}^f}{d\mathbb{P}}$  follows similarly.

Now we write down the detail of the exact Radon–Nikodym derivative given in (13).

$$\begin{aligned}
 \mathbb{E} \left[ \left| \frac{d\widehat{\mathbb{Q}}^c}{d\mathbb{P}} \right|^p \right] &= \mathbb{E} \left[ \exp \left( -pS \sum_{n=0}^{N/2-1} \left\langle \widehat{Y}_{t_{2n}}^f - \widehat{Y}_{t_{2n}}^c, \Delta W_{2n} + \Delta W_{2n+1} \right\rangle - \frac{p}{2} S^2 \sum_{n=0}^{N/2-1} \left\| \widehat{Y}_{t_{2n}}^f - \widehat{Y}_{t_{2n}}^c \right\|^2 2h \right) \right] \\
 &= \mathbb{E} \left[ \exp \left( \frac{p(2p-1)}{2} S^2 \sum_{n=0}^{N/2-1} \left\| \widehat{Y}_{t_{2n}}^f - \widehat{Y}_{t_{2n}}^c \right\|^2 2h \right) \right. \\
 &\quad \left. \times \exp \left( -pS \sum_{n=0}^{N/2-1} \left\langle \widehat{Y}_{t_{2n}}^f - \widehat{Y}_{t_{2n}}^c, \Delta W_{2n} + \Delta W_{2n+1} \right\rangle - p^2 S^2 \sum_{n=0}^{N/2-1} \left\| \widehat{Y}_{t_{2n}}^f - \widehat{Y}_{t_{2n}}^c \right\|^2 2h \right) \right],
 \end{aligned}$$

and then the Hölder’s inequality gives

$$\mathbb{E} \left[ \left| \frac{d\widehat{Q}^c}{d\mathbb{P}} \right|^p \right] \leq I_1^{1/2} I_2^{1/2},$$

where

$$I_1 = \mathbb{E} \left[ \exp \left( p(2p - 1)S^2 \sum_{n=0}^{N/2-1} \left\| \widehat{Y}_{t_{2n}}^f - \widehat{Y}_{t_{2n}}^c \right\|^2 2h \right) \right],$$

$$I_2 = \mathbb{E} \left[ \exp \left( -2pS \sum_{n=0}^{N/2-1} \left\langle \widehat{Y}_{t_{2n}}^f - \widehat{Y}_{t_{2n}}^c, \Delta W_{2n} + \Delta W_{2n+1} \right\rangle - 2p^2S^2 \sum_{n=0}^{N/2-1} \left\| \widehat{Y}_{t_{2n}}^f - \widehat{Y}_{t_{2n}}^c \right\|^2 2h \right) \right] \leq 1,$$

since the exponential term in  $I_2$  is a super-martingale.

For  $I_1$ , by Fatou’s lemma and Jensen’s inequality, we obtain

$$I_1 \leq \sum_{k=0}^{\infty} \frac{\mathbb{E} \left[ \left| 2(pS)^2 \sum_{n=0}^{N/2-1} \left\| \widehat{Y}_{t_{2n}}^f - \widehat{Y}_{t_{2n}}^c \right\|^2 2h \right|^k \right]}{k!}$$

$$\leq \sum_{k=0}^{\infty} \frac{2^k (pS)^{2k} T^{k-1} \sum_{n=0}^{N/2-1} \mathbb{E} \left[ \left\| \widehat{Y}_{t_{2n}}^f - \widehat{Y}_{t_{2n}}^c \right\|^{2k} \right] 2h}{k!},$$

then by Theorem 2 and the Stirling’s approximation  $k! \geq \sqrt{2\pi} k^{k+1/2} e^{-k}$  for any  $k \geq 1$ , there exist constants  $C_1, C_2 > 0$  such that

$$I_1 \lesssim 1 + \sum_{k=1}^{\infty} \frac{(2pSC_2)^{2k} (Th/e)^k}{\sqrt{2\pi} k} < 2,$$

provided  $(2pSC_2)^2 Th/e < 1/2$ .

Therefore, for all  $T > 0$  and  $p \geq 1$ , there exist constants  $C_{(1)}, C_{(2)} > 0$  such that, for all  $0 < h < \min(C_{(1)}, C_{(2)} / (Tp^2))$ ,

$$\mathbb{E} \left[ \left| \frac{d\widehat{Q}^c}{d\mathbb{P}} \right|^p \right] \leq 2. \quad \square$$

#### 6.4. Theorem 4

**Proof.** By Jensen’s and Hölder’s inequalities, we split the expectation into three parts:

$$\mathbb{E} \left[ \left| \varphi(\widehat{Y}_T^f) \frac{d\widehat{Q}^f}{d\mathbb{P}} - \varphi(\widehat{Y}_T^c) \frac{d\widehat{Q}^c}{d\mathbb{P}} \right|^p \right]$$

$$= \mathbb{E} \left[ \left| \varphi(\widehat{Y}_T^f) \frac{d\widehat{Q}^f}{d\mathbb{P}} - \varphi(\widehat{Y}_T^f) + \varphi(\widehat{Y}_T^f) - \varphi(\widehat{Y}_T^c) + \varphi(\widehat{Y}_T^c) - \varphi(\widehat{Y}_T^c) \frac{d\widehat{Q}^c}{d\mathbb{P}} \right|^p \right]$$

$$\begin{aligned} &\leq 3^{p-1} \mathbb{E} \left[ \left| \varphi(\widehat{Y}_T^f) - \varphi(\widehat{Y}_T^c) \right|^p \right] + 3^{p-1} \mathbb{E} \left[ \left| \varphi(\widehat{Y}_T^f) \right|^{2p} \right]^{1/2} \mathbb{E} \left[ \left| 1 - \frac{d\widehat{Q}^f}{d\mathbb{P}} \right|^{2p} \right]^{1/2} \\ &\quad + 3^{p-1} \mathbb{E} \left[ \left| \varphi(\widehat{Y}_T^c) \right|^{2p} \right]^{1/2} \mathbb{E} \left[ \left| 1 - \frac{d\widehat{Q}^c}{d\mathbb{P}} \right|^{2p} \right]^{1/2}. \end{aligned}$$

By Theorems 1 and 2 and Lipschitz condition, there exists a constant  $C_1 > 0$  such that

$$\begin{aligned} \mathbb{E} \left[ \left| \varphi(\widehat{Y}_T^f) - \varphi(\widehat{Y}_T^c) \right|^p \right] &\leq K^p \mathbb{E} \left[ \left\| \widehat{Y}_T^f - \widehat{Y}_T^c \right\|^p \right] \lesssim C_1^p p^p h^p, \\ \mathbb{E} \left[ \left| \varphi(\widehat{Y}_T^f) \right|^{2p} \right] &\lesssim C_1^p p^p, \quad \mathbb{E} \left[ \left| \varphi(\widehat{Y}_T^c) \right|^{2p} \right] \lesssim C_1^p p^p. \end{aligned}$$

Next we estimate  $\mathbb{E} \left[ \left| 1 - \frac{d\widehat{Q}^c}{d\mathbb{P}} \right|^p \right]$  with  $\frac{d\widehat{Q}^c}{d\mathbb{P}} = \exp(\mathcal{H})$  where

$$\mathcal{H} \triangleq -S \sum_{n=0}^{N/2-1} \left\langle \widehat{Y}_{2n}^f - \widehat{Y}_{2n}^c, \Delta W_{2n} + \Delta W_{2n+1} \right\rangle - \frac{S^2}{2} \sum_{n=0}^{N/2-1} \left\| \widehat{Y}_{2n}^f - \widehat{Y}_{2n}^c \right\|^2 2h.$$

Taylor expansion gives

$$e^x = 1 + e^{\xi(x)}x, \text{ for some } \xi(x) \text{ with } |\xi(x)| < |x|,$$

which, combined with Hölder’s inequality, implies that

$$\mathbb{E} \left[ \left| 1 - \frac{d\widehat{Q}^c}{d\mathbb{P}} \right|^{2p} \right] = \mathbb{E} \left[ (\exp(\xi(\mathcal{H})) |\mathcal{H}|)^{2p} \right] \leq \mathbb{E} \left[ |\exp(\xi(\mathcal{H}))|^{4p} \right]^{1/2} \mathbb{E} \left[ |\mathcal{H}|^{4p} \right]^{1/2}.$$

First, by Theorem 3, we obtain

$$\mathbb{E} \left[ |\exp(\xi(\mathcal{H}))|^{4p} \right] \leq \mathbb{E} [\max(\exp(4p\mathcal{H}), 1)] \leq \mathbb{E} [\exp(4p\mathcal{H})] + 1 \lesssim 3,$$

provided  $h < C_2/(T p^2)$  for some constant  $C_2 > 0$ .

Second, by Jensen’s inequality and the Burkholder–Davis–Gundy inequality in [2], there exists a constant  $C_3 > 0$  such that

$$\begin{aligned} \mathbb{E} \left[ |\mathcal{H}|^{4p} \right] &\leq 2^{4p-1} S^{4p} \mathbb{E} \left[ \left| \sum_{n=0}^{N/2-1} \left\langle \widehat{Y}_{2n}^f - \widehat{Y}_{2n}^c, \Delta W_{2n} + \Delta W_{2n+1} \right\rangle \right|^{4p} \right] \\ &\quad + 2^{4p-1} S^{8p} \mathbb{E} \left[ \left| \sum_{n=0}^{N/2-1} \left\| \widehat{Y}_{2n}^f - \widehat{Y}_{2n}^c \right\|^2 h \right|^{4p} \right] \\ &\leq 2^{4p-1} S^{4p} C_3^p p^{2p} \mathbb{E} \left[ \left| \sum_{n=0}^{N/2-1} \left\| \widehat{Y}_{2n}^f - \widehat{Y}_{2n}^c \right\|^2 2h \right|^{2p} \right] + 2^{4p-1} S^{8p} T^{4p-1} \mathbb{E} \left[ \sum_{n=0}^{N/2-1} \left\| \widehat{Y}_{2n}^f - \widehat{Y}_{2n}^c \right\|^{8p} h \right] \\ &\leq 2^{6p-1} S^{4p} C_3^p p^{2p} T^{2p-1} \mathbb{E} \left[ \sum_{n=0}^{N/2-1} \left\| \widehat{Y}_{2n}^f - \widehat{Y}_{2n}^c \right\|^{4p} h \right] + 2^{4p-1} S^{8p} T^{4p-1} \mathbb{E} \left[ \sum_{n=0}^{N/2-1} \left\| \widehat{Y}_{2n}^f - \widehat{Y}_{2n}^c \right\|^{8p} h \right]. \end{aligned}$$

Then, provided  $h < C_4/\sqrt{Tp}$  for some constant  $C_4 > 0$ , by Theorem 2, there exists a constant  $C_5 > 0$  such that

$$\mathbb{E} \left[ |\mathcal{H}|^{4p} \right] \lesssim C_5^{4p} p^{6p} T^{2p} h^{4p},$$

and then

$$\mathbb{E} \left[ \left| 1 - \frac{d\hat{Q}^c}{d\mathbb{P}} \right|^{2p} \right] \lesssim 3^{1/2} C_5^{2p} p^{3p} T^p h^{2p}.$$

Similarly, we can get the same result for  $\mathbb{E} \left[ \left| 1 - \frac{d\hat{Q}^f}{d\mathbb{P}} \right|^{2p} \right]$ .

Since  $1/\sqrt{Tp} \geq 1/(\sqrt{Tp}) \geq \min(1, 1/(Tp^2))$ , the condition  $h < C_4/\sqrt{Tp}$  can be replaced by  $h \leq C_4 \min(1, 1/(Tp^2))$ . Overall, combining all the estimates above, there exist constants  $C_{(1)}, C_{(2)}, C_{(3)} > 0$  such that, for any  $0 < h < \min(C_{(1)}, C_{(2)}/(Tp^2))$ ,

$$\mathbb{E} \left[ \left| \varphi(\hat{Y}_T^f) \frac{d\hat{Q}^f}{d\mathbb{P}} - \varphi(\hat{Y}_T^c) \frac{d\hat{Q}^c}{d\mathbb{P}} \right|^p \right] \leq C_{(3)}^p p^{2p} T^{p/2} h^p. \quad \square$$

### 7. Conclusions and future work

In this paper, we introduced a change of measure technique for multilevel Monte Carlo estimators. For chaotic ergodic SDEs satisfying the one-sided Lipschitz condition, we reduce 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 results support our analysis.

One direction for extension of the theory is to perform the numerical analysis for ergodic SDEs with a non-globally Lipschitz drift using adaptive timesteps and the change of measure technique, since numerical experiments in section 5 show it works well in these cases. Another possible direction is to follow the idea in [12] to estimate the mean exit time and related path functionals which correspond to the solution of an elliptic PDEs. In this case, we need to estimate the mean exit time in the infinite time interval and the associated path functionals, which again is a long-time simulation problem.

### Acknowledgment

This research was supported by EPSRC through the ICONIC Programme Grant EP/P020720/1.

### References

- [1] P. Andersson, A. Kohatsu-Higa, Unbiased simulation of stochastic differential equations using parametrix expansions, *Bernoulli* 23 (3) (2017) 2028–2057.
- [2] M. Barlow, M. Yor, Semi-martingale inequalities via the Garsia–Rodemich–Rumsey lemma, and applications to local times, *J. Funct. Anal.* 49 (2) (1982) 198–229.
- [3] A. Beskos, G. Roberts, Exact simulation of diffusions, *Ann. Appl. Probab.* 15 (4) (2005) 2422–2444.
- [4] M. Chen, S. Li, Coupling methods for multidimensional diffusion processes, *Ann. Probab.* 17 (1) (1989) 151–177.
- [5] M. Denny, Introduction to importance sampling in rare-event simulations, *Eur. J. Phys.* 22 (4) (2001) 403.
- [6] A. Eberle, A. Guillin, R. Zimmer, Couplings and quantitative contraction rates for Langevin dynamics, *Ann. Probab.* (2019), in press, arXiv preprint, arXiv:1703.01617.
- [7] W. Fang, M. Giles, Adaptive Euler–Maruyama method for SDEs with non-globally Lipschitz drift: Part I, finite time interval, arXiv preprint, arXiv:1609.08101.
- [8] W. Fang, M. Giles, Adaptive Euler–Maruyama method for SDEs with non-globally Lipschitz drift: Part II, infinite time interval, arXiv preprint, arXiv:1703.06743.
- [9] R. Gasparotto, Optimised Importance Sampling in Multilevel Monte Carlo, Master’s thesis, University of Oxford, 2015.

- [10] M. Giles, Multilevel Monte Carlo path simulation, *Oper. Res.* 56 (3) (2008) 607–617.
- [11] M. Giles, Multilevel Monte Carlo methods, *Acta Numer.* 24 (2015) 259.
- [12] M.B. Giles, F. Bernal, Multilevel estimation of expected exit times and other functionals of stopped diffusions, *SIAM/ASA J. Uncertain. Quantificat.* 6 (4) (2018) 1454–1474.
- [13] N. Hansen, Geometric ergodicity of discrete-time approximations to multivariate diffusions, *Bernoulli* 9 (4) (2003) 725–743.
- [14] D. Higham, X. Mao, A. Stuart, Strong convergence of Euler-type methods for nonlinear stochastic differential equations, *SIAM J. Numer. Anal.* 40 (3) (2002) 1041–1063.
- [15] M. Hutzenthaler, A. Jentzen, *Numerical Approximations of Stochastic Differential Equations with Non-Globally Lipschitz Continuous Coefficients*, vol. 236, American Mathematical Society, 2015.
- [16] S. Juneja, P. Shahabuddin, Rare-event simulation techniques: an introduction and recent advances, *Handbooks Oper. Res. Management Sci.* 13 (2006) 291–350.
- [17] A. Kebaier, J. Lelong, Coupling importance sampling and multilevel Monte Carlo using sample average approximation, *Methodol. Comput. Appl. Probab.* (2017) 1–31.
- [18] P. Kloeden, E. Platen, *Numerical Solution of Stochastic Differential Equations*, Springer-Verlag, New York, 1999.
- [19] V. Lemaire, An adaptive scheme for the approximation of dissipative systems, *Stochastic Process. Appl.* 117 (10) (2007) 1491–1518.
- [20] X. Mao, *Stochastic Differential Equations and Applications*, Elsevier, 2007.
- [21] J. Mattingly, A. Stuart, D. Higham, Ergodicity for SDEs and approximations: locally Lipschitz vector fields and degenerate noise, *Stochastic Process. Appl.* 101 (2) (2002) 185–232.
- [22] J. Mattingly, A. Stuart, M. Tretyakov, Convergence of numerical time-averaging and stationary measures via Poisson equations, *SIAM J. Numer. Anal.* 48 (2) (2010) 552–577.
- [23] S. Meyn, R. Tweedie, Stability of Markovian processes III: Foster–Lyapunov criteria for continuous-time processes, *Adv. in Appl. Probab.* 25 (3) (1993) 518–548.
- [24] G. Milstein, M. Tretyakov, Computing ergodic limits for Langevin equations, *Phys. D* 229 (1) (2007) 81–95.
- [25] G. Milstein, M. Tretyakov, *Stochastic Numerics for Mathematical Physics*, Springer Science & Business Media, 2013.
- [26] G. Roberts, R. Tweedie, Exponential convergence of Langevin distributions and their discrete approximations, *Bernoulli* 2 (4) (1996) 341–363.
- [27] C. Soize, *The Fokker–Planck Equation for Stochastic Dynamical Systems and Its Explicit Steady State Solutions*, vol. 17, World Scientific, 1994.
- [28] P. Stilger, S. Poon, Multi-level Monte Carlo simulations with importance sampling, SSRN: 2273215.
- [29] D. Talay, Second-order discretization schemes of stochastic differential systems for the computation of the invariant law, *Stoch. Int. J. Probab. Stoch. Process.* 29 (1) (1990) 13–36.
- [30] Y. Xia, M. Giles, Multilevel path simulation for jump-diffusion SDEs, in: *Monte Carlo and Quasi-Monte Carlo Methods 2010*, Springer, 2012, pp. 695–708.