Adaptive Euler-Maruyama method for SDEs with non-globally Lipschitz drift

Wei Fang and Michael B. Giles

Abstract This paper, based on two main papers [2, 3] which contains the full details of the literature review, numerical analysis and numerical experiments, aims to give an overview of the adaptive Euler-Maruyama method for SDEs with non-globally Lipschitz drift in a concise structure without any proof. It shows 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 $\frac{1}{2}$ for SDEs with a non-uniform globally Lipschitz volatility, and order 1 for Langevin SDEs with unit volatility and a drift with sufficient smoothness. For a class of ergodic SDEs, we also show that the bound for the moments and the strong error of the numerical solution are uniform in T, which allow us to introduce the adaptive multilevel Monte Carlo method to compute the expectations with respect to the invariant measure. The analysis is supported by numerical experiments.

1 Introduction

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

$$\mathrm{d}X_t = f(X_t)\,\mathrm{d}t + g(X_t)\,\mathrm{d}W_t, \tag{1}$$

with a fixed initial value x_0. The standard theory assumes the drift $f:\mathbb{R}^m \to \mathbb{R}^m$ and the volatility $g:\mathbb{R}^m \to \mathbb{R}^{m\times d}$ are both globally Lipschitz. Under this assumption, there is well-established theory on the existence and uniqueness of strong solutions, and

Wei Fang
Mathematical Institute, University of Oxford, UK, OX2 6GG. e-mail: wei.fang@maths.ox.ac.uk

Michael B. Giles
Mathematical Institute, University of Oxford, UK, OX2 6GG. e-mail: mike.giles@maths.ox.ac.uk

the numerical approximation $\widehat{X}_t$ obtained from the Euler-Maruyama discretization

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

using a uniform timestep of size h with Brownian increments ΔW_n, plus a suitable interpolation within each timestep, is known [19] to have a strong error which is $O(h^{1/2})$ so that for any $T, p > 0$,

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

The interest in this paper is in other cases in which g is again globally Lipschitz, but f is only locally Lipschitz. If, for some $\alpha, \beta \geq 0$, f also satisfies the one-sided growth condition

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

where $\langle\cdot,\cdot\rangle$ denotes an inner product, then it is again possible to prove the existence and uniqueness of strong solutions (see Theorems 2.3.5 and 2.4.1 in [20]). Furthermore (see Lemma 3.2 in [10]), these solutions are stable in the sense that for any $T, p > 0$, $\mathbb{E}\left[\sup_{0\leq t\leq T} \|X_t\|^p\right] < \infty$. The problem is that the numerical approximation given by the uniform timestep Euler-Maruyama discretization may not be stable. Indeed, for the SDE

$$\mathrm{d}X_t = -X_t^3\,\mathrm{d}t + \mathrm{d}W_t, \tag{2}$$

it has been proved [13] that for any $T > 0$ and $p \geq 2$, $\lim_{h\to 0} \mathbb{E}\left[\|\widehat{X}_T\|^p\right] = \infty$.

This behaviour has led to research on numerical methods which achieve strong convergence for these SDEs with a non-globally Lipschitz drift, see [2, 10, 12, 14, 21, 22, 26, 32] and the references therein.

The other motivation for this paper is the analysis of a class of ergodic SDEs which exponentially converge to some invariant measure π, for example, the FENE model in [1]. 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:

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

which drives us to consider the stability and strong convergence of the algorithm in the infinite time interval. Different approaches to computing the expectation include numerical solution of the Fokker-Planck equation, see [30] and the reference therein, and ergodic numerical solutions, see [9, 17, 23, 25, 27, 29, 31]. We assume that the SDEs have a locally Lipschitz drift $f: \mathbb{R}^m \to \mathbb{R}^m$ satisfying the dissipative condition: for some $\alpha, \beta > 0$,

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

and a bounded and non-degenerate volatility $g: \mathbb{R}^m \to \mathbb{R}^{m\times d}$.

In this paper, 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 $\widehat{X}_{t_n}$. Adaptive timesteps have been used in previous research to improve the accuracy of numerical approximations, see [4, 11, 15, 16, 17, 18, 24, 28] and the references therein. The idea of using an adaptive timestep in this paper comes from considering the divergence of the uniform timestep method for the SDE (2). 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.

For the ergodic SDEs, 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, multi-level Monte Carlo (MLMC) methodology [5, 6] is employed and non-nested timestepping is used to construct an adaptive MLMC [7]. Following the idea of Glynn and Rhee [8] to estimate the invariant measure of some Markov chains, we introduce an adaptive MLMC algorithm for the infinite time interval, in which each level ℓ has a different time interval length T_ℓ, to achieve a better computational performance.

The rest of the paper is organized as follows. The adaptive algorithm is presented and the main theorems both in finite time interval and infinite time interval are stated in Section 2. Section 3 introduces the MLMC schemes, and the relevant numerical experiments are provided in section 4. Finally, section 5 concludes.

In this paper we consider both the finite time interval $[0,T]$ with $T>0$ be a fixed positive real number and the infinite time interval $[0,\infty)$. Let $(\Omega,\mathscr{F},\mathbb{P})$ be a probability space with normal filtration $(\mathscr{F}_t)_{t\in[0,\infty)}$ for section 2 and $(\mathscr{F}_t)_{t\in(-\infty,0]}$ for section 3 corresponding to a d-dimensional standard Brownian motion $W_t = (W^{(1)},W^{(2)},\ldots,W^{(d)})_t^T$. We denote the vector norm by $\|v\| \triangleq (|v_1|^2+|v_2|^2+\ldots+|v_m|^2)^{\frac{1}{2}}$, the inner product of vectors v and w by $\langle v,w\rangle \triangleq v_1w_1+v_2w_2+\ldots+v_mw_m$, 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 all $A\in\mathbb{R}^{m\times d}$.

2 Adaptive algorithm and theoretical results

2.1 Adaptive Euler-Maruyama method

The adaptive Euler-Maruyama discretization is

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

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

One key point in the analysis is to prove that t_n increases without bound as n increases. More specifically, the analysis proves that for any $T>0$, almost surely for each path there is an N such that $t_N\geq T$.

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

We define the piecewise constant interpolant process $\bar{X}_t = \widehat{X}_{\underline{t}}$ and also define the standard continuous interpolant [19] as

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

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

$$\mathrm{d}\widehat{X}_t = f(\widehat{X}_{\underline{t}})\,\mathrm{d}t + g(\widehat{X}_{\underline{t}})\,\mathrm{d}W_t = f(\bar{X}_t)\,\mathrm{d}t + g(\bar{X}_t)\,\mathrm{d}W_t. \tag{4}$$

In the following two subsections, we state the key results on stability and strong convergence in both finite and infinite time intervals, and related results on the number of timesteps, introducing various assumptions as required for each. All the proofs are in [2] and [3].

2.2 Finite Time Interval

2.2.1 Stability

Assumption 1 (Local Lipschitz and linear growth) *f and g are both locally Lipschitz, so that for any $R>0$ there is a constant C_R such that*

$$\|f(x)-f(y)\| + \|g(x)-g(y)\| \le C_R\,\|x-y\|$$

for all $x,y \in \mathbb{R}^m$ with $\|x\|,\|y\| \le R$. Furthermore, there exist constants $\alpha,\beta \ge 0$ such that for all $x \in \mathbb{R}^m$, f satisfies the one-sided linear growth condition:

$$\langle x, f(x)\rangle \le \alpha\|x\|^2 + \beta, \tag{5}$$

and g satisfies the linear growth condition:

$$\|g(x)\|^2 \le \alpha\|x\|^2 + \beta. \tag{6}$$

Together, (5) and (6) imply the monotone condition $\langle x, f(x)\rangle + \frac{1}{2}\|g(x)\|^2 \le \frac{3}{2}(\alpha\|x\|^2+\beta)$, which is a key assumption in the analysis of Mao & Szpruch [22] and Mao [21] for SDEs with volatilities which are not globally Lipschitz. However, in our analysis we choose to use this slightly stronger assumption, which provides the basis for the following lemma on the stability of the SDE solution.

Lemma 1 (SDE stability). *If the SDE satisfies Assumption 1, then for all $p>0$*

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

We now specify the critical assumption about the adaptive timestep.

Assumption 2 (Adaptive timestep) *The adaptive timestep function $h:\mathbb{R}^m \to \mathbb{R}^+$ is continuous and strictly positive, and there exist constants $\alpha,\beta>0$ such that for all $x\in\mathbb{R}^m$, $h(x)$ satisfies the inequality*

$$\langle x, f(x)\rangle + \tfrac{1}{2}\, h(x)\, \|f(x)\|^2 \le \alpha \|x\|^2 + \beta. \tag{7}$$

Note that if another timestep function $h^\delta(x)$ is smaller than $h(x)$, then $h^\delta(x)$ also satisfies the Assumption 2. Note also that the form of (7), which is motivated by the requirements of the proof of the next theorem, is very similar to (5). Indeed, if (7) is satisfied then (5) is also true for the same values of α and β.

Theorem 1 (Finite time stability). *If the SDE satisfies Assumption 1, and the timestep function h satisfies Assumption 2, then T is almost surely attainable (i.e. for $\omega\in\Omega$, $\mathbb{P}(\exists N(\omega)<\infty$ s.t. $t_{N(\omega)}\ge T)=1$) and for all $p>0$ there exists a constant $C_{p,T}$ which depends solely on p, T and the constants α,β in Assumption 2, such that*

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

2.2.2 Strong convergence

Standard strong convergence analysis for an 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 $0<\delta\le 1$, and consider the limit $\delta\to 0$.

Given a timestep function $h(x)$ which satisfies Assumption 2, ensuring stability as analysed in the previous section, there are two quite natural ways in which we might introduce δ to define $h^\delta(x)$:

$$h^\delta(x) = \delta\, \min(T, h(x)), \;\; h^\delta(x) = \min(\delta\, T, h(x)).$$

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 3 *The timestep function h^δ, satisfies the inequality*

$$\delta\, \min(T, h(x)) \le h^\delta(x) \le \min(\delta\, T, h(x)), \tag{8}$$

and h satisfies Assumption 2.

Given this assumption, we obtain the following theorem:

Theorem 2 (Strong convergence). *If the SDE satisfies Assumption 1, and the timestep function h^δ satisfies Assumption 3, then for all $p>0$*

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

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

Assumption 4 (Lipschitz properties) *There exists a constant $\alpha>0$ such that for all $x,y\in\mathbb{R}^m$, f satisfies the one-sided Lipschitz condition:*

$$\langle x-y, f(x)-f(y)\rangle \leq \tfrac{1}{2}\alpha\|x-y\|^2, \tag{9}$$

and g satisfies the Lipschitz condition:

$$\|g(x)-g(y)\|^2 \leq \tfrac{1}{2}\alpha\|x-y\|^2. \tag{10}$$

In addition, f satisfies the polynomial growth Lipschitz condition

$$\|f(x)-f(y)\| \leq (\gamma(\|x\|^q+\|y\|^q)+\mu)\ \|x-y\|, \tag{11}$$

for some $\gamma,\mu,q>0$.

Note that setting $y=0$ gives

$$\langle x, f(x)\rangle \ \leq\ \tfrac{1}{2}\alpha\|x\|^2 + \langle x, f(0)\rangle \ \leq\ \alpha\|x\|^2 + \tfrac{1}{2}\alpha^{-1}\|f(0)\|^2,$$

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

Hence, Assumption 4 implies Assumption 1, with the same α and an appropriate β.

Theorem 3 (Strong convergence order). *If the SDE satisfies Assumption 4, and the timestep function h^δ satisfies Assumption 3, then for all $p>0$ there exists a constant $C_{p,T}$ such that*

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

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

Assumption 5 (Timestep lower bound) *There exist constants $\xi,\zeta,q>0$, such that the adaptive timestep function satisfies the inequality*

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

Lemma 2 (Number of timesteps). *If the SDE satisfies Assumption 1, and the timestep function $h^\delta(x)$ satisfies Assumption 3, with $h(x)$ satisfying Assumptions 2 and Assumption 5, then for all $p>0$ there exists a constant $c_{p,T}$ such that*

$$\mathbb{E}\left[(N_T-1)^p\right] \leq c_{p,T}\, \delta^{-p}.$$

where N_T is again the number of timesteps required by a path approximation.

The conclusion from Theorem 3 and Lemma 2 is that

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

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 6 (Enhanced Lipschitz properties) *f satisfies the Assumption 4 and in addition, f is differentiable, and f and ∇f satisfy the polynomial growth Lipschitz condition*

$$\|f(x)-f(y)\| + \|\nabla f(x)-\nabla f(y)\| \le (\gamma(\|x\|^q+\|y\|^q)+\mu)\,\|x-y\|, \qquad (12)$$

for some $\gamma, \mu, q > 0$.

We now state the theorem on improved strong convergence.

Theorem 4 (Strong convergence for Langevin SDEs). *If $m=d$, $g \equiv I_m$, f satisfies Assumption 6, and the timestep function h^δ satisfies Assumption 3, then for all $T, p \in (0,\infty)$ there exists a constant $C_{p,T}$ such that*

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

Comment: first order strong convergence can also be achieved for a general $g(x)$ by using an adaptive timestep Milstein discretization, provided ∇g satisfies an additional Lipschitz condition. However, this numerical approach is only practical in cases in which the commutativity condition is satisfied and therefore there is no need to simulate the Lévy areas which the Milstein method otherwise requires [19].

2.3 Infinite Time Interval

Now, we focus on a class of ergodic SDEs and show that the moment bounds and strong error bound is uniform in T which is a stronger result than for the finite time interval.

2.3.1 Stability

Assumption 7 (Dissipative condition) *f and g satisfy the Assumption 1 and there exist constants $\alpha, \beta > 0$ such that for all $x \in \mathbb{R}^m$, f satisfies the dissipative one-sided linear growth condition:*

$$\langle x, f(x)\rangle \le -\alpha\|x\|^2 + \beta, \qquad (13)$$

and g is globally bounded and non-degenerate:

$$\|g(x)\|^2 \le \beta. \qquad (14)$$

Theorem 4.4 in [23] and Theorem 6.1 in [25] 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 (SDE stability in infinite time interval). *If the SDE satisfies Assumption 7 with $X_0 = x_0$, then for all $p \in (0,\infty)$, there is a constant C_p which only depends on x_0 and p such that, $\forall t \geq 0$,*

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

We now specify the critical assumption about the adaptive timestep for infinite time interval.

Assumption 8 (Adaptive timestep for infinite time interval) *The adaptive timestep function $h : \mathbb{R}^m \to (0, h_{max}]$ is continuous and bounded, with $0 < h_{max} < \infty$, and there exist constants $\alpha, \beta > 0$ such that for all $x \in \mathbb{R}^m$, h satisfies the inequality*

$$\langle x, f(x)\rangle + \tfrac{1}{2}\, h(x)\, \|f(x)\|^2 \leq -\alpha \|x\|^2 + \beta. \tag{15}$$

Note that if another timestep function $h^\delta(x)$ is smaller than $h(x)$, then $h^\delta(x)$ also satisfies this Assumption. Note also that the form of (15), which is motivated by the requirements of the proof of the next theorem, is very similar to (13). Indeed, if (15) is satisfied then (13) is also true for the same values of α and β. Compared with the condition in the finite time analysis, we need additionally to bound h properly to achieve the uniform bound.

Theorem 5 (Stability in infinite interval). *If the SDE satisfies Assumption 7, and the timestep function h satisfies Assumption 8, then for all $p \in (0,\infty)$ there exists a constant C_p which depends solely on p, x_0, h_{max} and the constants α, β in Assumption 8 such that, $\forall t \geq 0$,*

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

2.3.2 Strong convergence

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

Assumption 9 (Contractive Lipschitz properties) *f and g satisfy Assumption 4 and for some fixed $p^* \in (1,\infty)$, there exist constants $\lambda > 0$ such that for all $x, y \in \mathbb{R}^m$, f and g satisfy the contractive Lipschitz condition:*

$$\langle x-y, f(x)-f(y)\rangle + \frac{p^*-1}{2}\, \|g(x)-g(y)\|^2 \leq -\lambda\, \|x-y\|^2, \tag{16}$$

Note that 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 4 (SDE contractivity). *If the SDE satisfies Assumption 9 and for some fixed $p^* \in (1,\infty)$, then for $p \in (0,p^*]$ 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 that, $\forall\, t > 0$,*

$$\mathbb{E}\left[\|X_t - Y_t\|^p\right] \leq e^{-\lambda p t}\,\mathbb{E}\left[\|X_0 - Y_0\|^p\right].$$

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

Theorem 6 (Strong convergence order in infinite time interval). *If the SDE satisfies Assumption 9, and the timestep function h^δ satisfies Assumption 3 with h satisfying Assumption 8, then for all $p \in (0,p^*]$ there exists a constant C_p such that, $\forall t \geq 0$,*

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

For the infinite time interval, we can show that the expected number of timesteps per path is linear in T, which is the same as for uniform timesteps.

Lemma 5 (Number of timesteps). *If the SDE satisfies Assumption 9, and the timestep function h^δ satisfies Assumption 3, with $h(x)$ satisfying Assumption 5 and Assumption 8, then for all $T, p \in (0,\infty)$ there exists a constant c_p such that*

$$\mathbb{E}\left[(N_T - 1)^p\right] \leq c_p\,T^p\,\delta^{-p}.$$

where N_T is again the number of timesteps required by a path approximation.

First order strong convergence is also 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 10 (Enhanced contractive Lipschitz properties) *f satisfies Assumption 9 and in addition, f is differentiable, and f and ∇f satisfy the polynomial growth Lipschitz condition 12.*

Theorem 7 (Strong convergence for Langevin SDEs in infinite time interval). *If $m = d$, $g \equiv I_m$, f satisfies Assumption 10, and the timestep function h^δ satisfies Assumption 3 and 8, then for all $p \in (0,\infty)$ there exists a constant C_p such that, $\forall\, t \geq 0$,*

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

3 Multi-Level Monte Carlo in infinite time interval

We are interested in the problem of approximating:

$$\pi(\varphi) := \mathbb{E}_\pi \varphi = \int_{\mathbb{R}^m} \varphi(x)\pi(\mathrm{d}x), \quad \varphi \in L^1(\pi),$$

where π is the invariant measure of the SDE (1). Numerically, we can approximate this quantity by simulating $\mathbb{E}[\varphi(X_T)]$ for a sufficiently large T. In the following subsections, we will introduce our adaptive multilevel Monte Carlo algorithm and its numerical analysis.

3.1 Algorithm

To estimate $\mathbb{E}[\varphi(X_T)]$, the simplest Monte Carlo estimator is

$$\frac{1}{N}\sum_{n=1}^{N}\varphi(\widehat{X}_T^{(n)}),$$

where $\widehat{X}_T^{(n)}$ is the terminal value of the nth numerical path in the time interval $[0,T]$ using a suitable adaptive function h^δ. It can be extended to Multilevel Monte Carlo by using non-nested timesteps [7]. Consider the identity

$$\mathbb{E}[\varphi_L] = \mathbb{E}[\varphi_0] + \sum_{\ell=1}^{L}\mathbb{E}[\varphi_\ell - \varphi_{\ell-1}], \tag{17}$$

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

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

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 $\delta = M^{-\ell}$.

Unlike the standard MLMC with fixed time interval $[0,T]$, we now allow different levels to have a different length of time interval T_ℓ, satisfying $0 < T_0 < T_1 < \ldots < T_\ell < \ldots < T_L = T$, which means that as 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 [8] 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 terminating 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 the 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 4, two solutions to the SDE satisfying Assumption 9, 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 has the identity (17) 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

$$\widehat{Y} \triangleq \frac{1}{N_0}\sum_{n=1}^{N_0} \varphi(\widehat{X}_0^{(n,0)}) + \sum_{\ell=1}^{L}\left\{\frac{1}{N_\ell}\sum_{n=1}^{N_\ell}\left(\varphi(\widehat{X}_0^{(n,\ell)}) - \varphi(\widehat{X}_0^{(n,\ell-1)})\right)\right\}, \tag{18}$$

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

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

```
t := −T_ℓ; t^c := −T_{ℓ−1}; t^f := −T_ℓ;
h^c := 0; h^f := 0;
ΔW^c := 0; ΔW^f := 0;
X̂^c = x_0; X̂^f = x_0;
while t < 0 do
    t_old := t;
    t := min(t^c, t^f);
    ΔW := N(0, t − t_old);
    ΔW^c := ΔW^c + ΔW;
    if t = −T_{ℓ−1} then
        ΔW^c := 0;
    end
    ΔW^f := ΔW^f + ΔW;
    if t = t^c then
        update coarse path X̂^c using h^c and ΔW^c;
        compute new adapted coarse path timestep h^c = h^{2δ}(X̂^c);
        h^c := min(h^c, −t^c);
        t^c := t^c + h^c;
        ΔW^c := 0;
    end
    if t = t^f then
        update fine path X̂^f using h^f and ΔW^f;
        compute new adapted fine path timestep h^f = h^δ(X̂^f);
        h^f := min(h^f, −t^f);
        t^f := t^f + h^f;
        ΔW^f := 0;
    end
end
Result: X̂^f − X̂^c
```

3.2 Numerical analysis

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

Lemma 6 (Exponential convergence). *If the SDE satisfies Assumption 7 and Assumption 9, and φ satisfies the Lipschitz condition: there exists a constant $\kappa > 0$ such that*

$$\|\varphi(x) - \varphi(y)\| \leq \kappa \|x - y\|, \tag{19}$$

then there exists a constant $\mu > 0$ depending on x_0, κ and C_1 in Lemma 3 such that

$$|\mathbb{E}\left[\varphi(X_t) - \pi(\varphi)\right]| \leq \mu\, e^{-\lambda t}. \tag{20}$$

With this, we can bound the variance of the MLMC correction for each level.

Lemma 7 (Variance of MLMC corrections for bounded volatility). *If φ satisfies the Lipschitz condition* (19)*, the SDE satisfies Assumption 9 and the timestep function h^δ satisfies Assumption 3 with $\delta = M^{-\ell}$ 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*

$$V_\ell \leq c_1 M^{-\ell} + c_2\, e^{-2\lambda T_{\ell-1}}. \tag{21}$$

Note that if we set $T_\ell = \frac{\log M}{2\lambda}(\ell+1)$, then $V_\ell \leq (c_1 + c_2) M^{-\ell}$, which has the same magnitude order as the standard MLMC. In some cases, λ needs to be estimated numerically through Lemma 6. N_ℓ can be optimized following the same approach in the MLMC theorem in [6].

Theorem 8 (MLMC for infinite time interval). *If φ satisfies the Lipschitz condition* (19)*, the SDE satisfies Assumption 9 and the timestep function h^δ satisfies Assumption 3 with $\delta = M^{-\ell}$ for each level, then by choosing suitable T_ℓ, N_ℓ for each level ℓ, there exists a constant c_3 such that the MLMC estimator (18) has a mean square error (MSE) with bound*

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

and a computational cost $\boldsymbol{C}$ with bound

$$\mathbb{E}\left[\boldsymbol{C}\right] \leq c_3\, \varepsilon^{-2} |\log \varepsilon|^3.$$

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

Theorem 9 (Langevin SDEs). *If φ satisfies the Lipschitz condition* (19)*, and for the SDE, $m = d$, $g \equiv I_m$, f satisfies Assumption 10, and the timestep function h^δ satisfies Assumption 3 with $\delta = M^{-\ell}$ for each level, then for each level ℓ, there exist constants c_1 and c_2 such that*

$$V_l \leq c_1 M^{-2\ell} + c_2 e^{-2\lambda T_{\ell-1}}. \tag{22}$$

By choosing suitable $T_\ell = \frac{\log M}{\lambda}(\ell+1)$ and N_ℓ for each level ℓ in the MLMC estimator (18) such that it achieves the MSE bound ε^2, there exists a constant c_3 such that

$$\mathbb{E}[\boldsymbol{C}] \leq c_3 \varepsilon^{-2}.$$

Note that the choice of T_ℓ for Langevin equation is different from the one for SDEs with bounded volatility. In other words, the strong convergence result and the contractive convergence rate λ determine T_ℓ.

4 Examples and Numerical Results

In this section we first discuss some example SDEs with non-globally Lipschitz drift, then present the numerical result for finite time interval and its extension to infinite time interval.

For scalar SDEs, the drift is often of the form

$$f(x) \approx -c\,\mathrm{sign}(x)\,|x|^q, \quad \text{as } |x| \to \infty \tag{23}$$

for some constants $c>0$, $q>1$. Therefore, as $|x|\to\infty$, the maximum stable timestep satisfying Assumption 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

$$h(x) = \min\left(T, c^{-1}|x|^{1-q}\right), \quad h^\delta(x) = \delta\, h(x). \tag{24}$$

For example, the Ginzburg-Landau equation, which describes a phase transition from the theory of superconductivity [13, 19], is

$$\mathrm{d}X_t = \left((\eta + \tfrac{1}{2}\sigma^2)X_t - \lambda X_t^3\right)\mathrm{d}t + \sigma X_t\,\mathrm{d}W_t,$$

where $\eta \geq 0$, $\lambda, \sigma > 0$. The drift and volatility satisfy Assumptions 1 and 4, and therefore all of the theory is applicable, with a suitable choice for $h^\delta(x)$, based on (23) and (24), being

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

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

$$\langle x, f(x)\rangle \leq -\beta\,\|x\|\,\|f(x)\|,$$

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

$$h(x) = \min(T, \|x\|/\|f(x)\|).$$

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

$$h(x)=\min(T,\gamma\|x\|^2/\|f(x)\|^2), \tag{25}$$

for some $\gamma>0$. For example, the Stochastic Lorenz equation, which is a three-dimensional system modelling convection rolls in the atmosphere [12], is

$$\begin{aligned}
\mathrm{d}X_t^{(1)} &= \left(\alpha_1 X_t^{(2)}-\alpha_1 X_t^{(1)}\right)\mathrm{d}t+\beta_1 X_t^{(1)}\mathrm{d}W_t^{(1)}\\
\mathrm{d}X_t^{(2)} &= \left(\alpha_2 X_t^{(1)}-X_t^{(2)}-X_t^{(1)}X_t^{(3)}\right)\mathrm{d}t+\beta_2 X_t^{(2)}\mathrm{d}W_t^{(2)}\\
\mathrm{d}X_t^{(3)} &= \left(X_t^{(1)}X_t^{(2)}-\alpha_3 X_t^{(3)}\right)\mathrm{d}t+\beta_3 X_t^{(3)}\mathrm{d}W_t^{(3)}
\end{aligned}$$

where $\alpha_1,\alpha_2,\alpha_3,\beta_1,\beta_2,\beta_3>0$. 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 $h(x)=\min(T,\gamma\|x\|^{-2})$, for any $\gamma>0$. However, the drift does not satisfy the one-sided Lipschitz condition, and therefore the theory on the order of strong convergence is not applicable.

All the adaptive functions above satisfy the Assumptions 2 and 5. Other example applications include the stochastic Verhulst equation and a large class of Langevin equations.

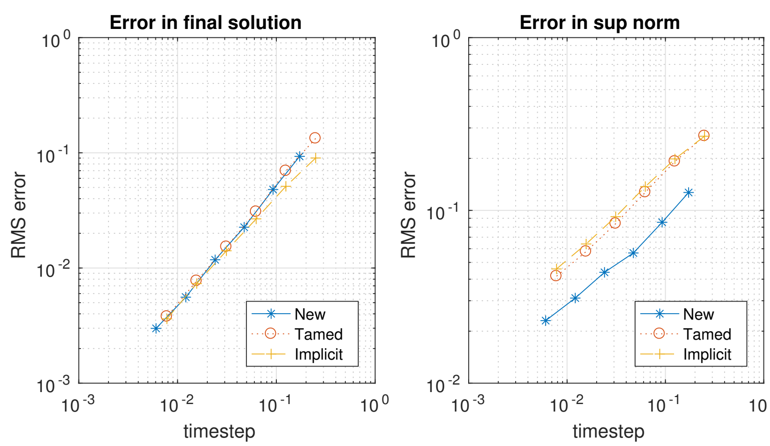

Fig. 1 Numerical results for finite time interval

The testcase taken from [14] is

$$\mathrm{d}X_t=-X_t-X_t^3\,\mathrm{d}t+\mathrm{d}W_t,\quad x_0=1,$$

with $T=1$. The three methods tested are the Tamed Euler scheme, the implicit Euler scheme, and the new Euler scheme with adaptive timestep. We can set $h_{max}=1$, $M=2$ and choose the adaptive function h, h^δ to be

$$h(x) = \frac{\max(1,|x|)}{\max(1,|x+x^3|)}, \;\; h^{\delta}(x) = 2^{-\ell} h(x).$$

Figure 1 shows the the root-mean-square error plotted against the average timestep. 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 2 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 [14], 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.

Next, we extend it to adaptive MLMC for the infinite time interval, since it also satisfies the dissipative condition (5) and the contractive condition (16). Our interest is to compute $\pi(\varphi)$ where $\varphi(x) = \|x\|$ satisfies a Lipschitz condition.

First we need to determine T_ℓ for each level. By differentiating drift f we know $\lambda \geq 1$ and choose λ to be 1 in our numerical scheme to simulate a sufficiently long time interval and control the truncation error. Then we choose

$$T_\ell = \log 2\,(\ell+1).$$

The variance result (22) for the Langevin equation is illustrated in Figure 2. The exponential part dominates the variance at the beginning, so the variance decays exponentially. As time increase, the $M^{-2\ell}$ term becomes the major part of the variance and the variance stops decreasing.

For level 10, we have $T_{10} = 7.62$ and the variance already stopped decreasing since $T = 5$ as shown in the Figure 2, which shows that the setting of T_ℓ is sufficient. Then, all the convergence results are the same as the standard MLMC and our algorithm works well. For more detail, see [3].

5 Conclusion

The central conclusion from this paper is that by using an adaptive timestep it is possible to make the Euler-Maruyama approximation stable for SDEs with a globally Lipschitz volatility and a drift which is not globally Lipschitz but is locally Lipschitz and satisfies a one-sided linear growth condition. 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. For ergodic SDEs satisfying the dissipative and contractive condition, we have shown that the moments and strong error of the numerical solutions are bounded and independent of time T. Moreover, we extend this adaptive scheme to MLMC for the infinite time interval by allowing different lengths of time intervals and carefully coupling the fine path and

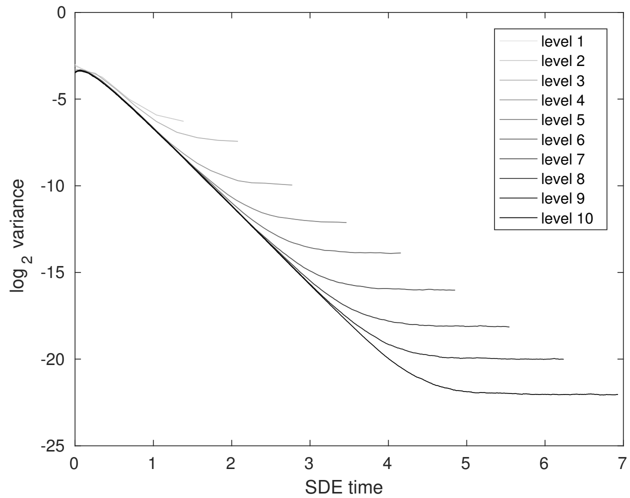

Fig. 2 Variance of corrections on each level ℓ

coarse path in each level ℓ. All the schemes work well and numerical experiments support the theoretical results.

References

1. J.W. Barrett and E. Süli. Existence of global weak solutions to some regularized kinetic models for dilute polymers. *SIAM Multiscale Modelling and Simulation*, 6(2):506–546, 2007.
2. W Fang and M.B. Giles. Adaptive Euler-Maruyama method for SDEs with non-globally Lipschitz drift: Part I, finite time interval. *arXiv preprint arXiv:1609.08101*, 2016.
3. W Fang and M.B. Giles. Adaptive Euler-Maruyama method for SDEs with non-globally Lipschitz drift: Part II, infinite time interval. *arXiv preprint arXiv:1703.06743*, 2017.
4. J.G. Gaines and T.J. Lyons. Variable step size control in the numerical solution of stochastic differential equations. *SIAM Journal on Applied Mathematics*, 57(5):1455–1484, 1997.
5. M.B. Giles. Multilevel Monte Carlo path simulation. *Operations Research*, 56(3):607–617, 2008.
6. M.B. Giles. Multilevel Monte Carlo methods. *Acta Numerica*, 24:259, 2015.
7. M.B. Giles, C. Lester, and J. Whittle. Non-nested adaptive timesteps in multilevel Monte Carlo computations. 2016.
8. P.W. Glynn, C. Rhee, et al. Exact estimation for Markov chain equilibrium expectations. *Journal of Applied Probability*, 51:377–389, 2014.
9. N.R. Hansen. Geometric ergodicity of discrete-time approximations to multivariate diffusions. *Bernoulli*, pages 725–743, 2003.
10. D.J. Higham, X. Mao, and A.M. Stuart. Strong convergence of Euler-type methods for nonlinear stochastic differential equations. *SIAM Journal on Numerical Analysis*, 40(3):1041–1063, 2002.
11. N. Hofmann, T. Müller-Gronbach, and K. Ritter. The optimal discretization of stochastic differential equations. *Journal of Complexity*, 17(1):117–153, 2001.

12. M. Hutzenthaler and A. Jentzen. *Numerical approximations of stochastic differential equations with non-globally Lipschitz continuous coefficients*, volume 236. American Mathematical Society, 2015.
13. M. Hutzenthaler, A. Jentzen, and P.E. Kloeden. Strong and weak divergence in finite time of Euler's method for stochastic differential equations with non-globally Lipschitz continuous coefficients. *Proceedings of the Royal Society of London A: Mathematical, Physical and Engineering Sciences*, 467(2130):1563–1576, 2011.
14. M. Hutzenthaler, A. Jentzen, and P.E. Kloeden. Strong convergence of an explicit numerical method for SDEs with nonglobally Lipschitz continuous coefficients. *Annals of Applied Probability*, 22(4):1611–1641, 2012.
15. H. Lamba. An adaptive timestepping algorithm for stochastic differential equations. *Journal of Computational and Applied Mathematics*, 161(2):417–430, 2003.
16. H. Lamba, J.C. Mattingly, and A.M. Stuart. An adaptive Euler-Maruyama scheme for SDEs: convergence and stability. *Journal of Numerical Analysis*, 27:479–506, 2007.
17. V. Lemaire. An adaptive scheme for the approximation of dissipative systems. *Stochastic Processes and their Applications*, 117(10):1491–1518, 2007.
18. Cónall Kelly and Gabriel J Lord. Adaptive timestepping strategies for nonlinear stochastic systems. *arXiv preprint arXiv:1610.04003*, 2016.
19. P.E. Kloeden and E. Platen. *Numerical Solution of Stochastic Differential Equations*. Springer, Berlin, 1992.
20. X. Mao. *Stochastic Differential Equations and Applications*. Horwood Publishers Ltd., 1997.
21. X. Mao. The truncated Euler-Maruyama method for stochastic differential equations. *Journal of Computational and Applied Mathematics*, 290:370–384, 2015.
22. X. Mao and L. Szpruch. Strong convergence and stability of implicit numerical methods for stochastic differential equations with non-globally Lipschitz continuous coefficients. *Journal of Computational and Applied Mathematics*, 238:14–28, 2013.
23. J.C. Mattingly, A.M. Stuart, and D.J. Higham. Ergodicity for SDEs and approximations: locally Lipschitz vector fields and degenerate noise. *Stochastic processes and their applications*, 101(2):185–232, 2002.
24. S. Mauthner. Step size control in the numerical solution of stochastic differential equations. *Journal of Computational and Applied Mathematics*, 100(1):93–109, 1998.
25. S.P. Meyn and R.L. Tweedie. Stability of Markovian processes III: Foster-Lyapunov criteria for continuous-time processes. *Advances in Applied Probability*, pages 518–548, 1993.
26. G.N. Milstein and M.V. Tretyakov. Numerical integration of stochastic differential equations with nonglobally Lipschitz coefficients. *SIAM Journal on Numerical Analysis*, 43(3):1139–1154, 2005.
27. G.N. Milstein and M.V. Tretyakov. Computing ergodic limits for Langevin equations. *Physica D: Nonlinear Phenomena*, 229(1):81–95, 2007.
28. T. Müller-Gronbach. Optimal pointwise approximation of SDEs based on Brownian motion at discrete points. *Annals of Applied Probability*, pages 1605–1642, 2004.
29. G.O. Roberts and R.L. Tweedie. Exponential convergence of Langevin distributions and their discrete approximations. *Bernoulli*, pages 341–363, 1996.
30. C. Soize. *The Fokker-Planck equation for stochastic dynamical systems and its explicit steady state solutions*, volume 17. World Scientific, 1994.
31. D. Talay. Second-order discretization schemes of stochastic differential systems for the computation of the invariant law. *Stochastics: An International Journal of Probability and Stochastic Processes*, 29(1):13–36, 1990.
32. X. Wang and S. Gan. The tamed Milstein method for commutative stochastic differential equations with non-globally Lipschitz continuous coefficients. *Journal of Difference Equations and Applications*, 19(3):466–490, 2013.