ADAPTIVE EULER-MARUYAMA METHOD FOR SDES WITH NON-GLOBALLY LIPSCHITZ DRIFT

By Wei Fang and Michael B. Giles

University of Oxford

This paper proposes an adaptive timestep construction for an Euler-Maruyama approximation of SDEs with non-globally Lipschitz drift. It is proved that if the timestep is bounded appropriately, then over a finite time interval the numerical approximation is stable, and the expected number of timesteps is finite. Furthermore, the order of strong convergence is the same as usual, i.e. order $\frac{1}{2}$ 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 distribution. 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 coefficient $f : \mathbb{R}^m \to \mathbb{R}^m$ and the diffusion coefficient $g : \mathbb{R}^m \to \mathbb{R}^{m \times d}$ are both globally Lipschitz. Under this assumption, there is well-established theory on the existence and uniqueness of strong solutions, and the numerical approximation $\widehat{X}_t$ obtained from the explicit 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 [18] to have a strong error which is $O(h^{1/2})$.

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

---

1

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

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 [22]). Furthermore (see Lemma 3.2 in [12]), 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 $\widehat{X}_t$ given by the uniform timestep explicit Euler-Maruyama discretization may not be stable. Indeed, for the SDE

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

it has been proved [15] 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 coefficient, see [12, 16, 17, 23, 24, 25, 29, 32, 39] and 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, especially the FENE (Finitely Extensible Nonlinear Elastic) model in [2]. To ensure the ergodicity, we assume that the SDEs have a locally Lipschitz drift $f :$ $\mathbb{R}^m \to \mathbb{R}^m$ satisfying the dissipative condition: for some $\alpha, \beta > 0$,

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

and a bounded and non-degenerate diffusion coefficient $g : \mathbb{R}^m \to \mathbb{R}^{m \times d}$. 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. Several different methodologies have been developed to estimate the expectation $\pi(\varphi)$.

First, we can compute the probability density function $\rho(x)$ of π by solving the corresponding stationary Fokker-Planck equation, see [36] and references therein. The second approach is based on the ergodicity of the SDEs:

$$\text{(4)} \qquad \lim_{T \to \infty} \frac{1}{T} \int_0^T \varphi(X_t)\,\mathrm{d}t \;=\; \pi(\varphi),\; a.s.,$$

where the limit does not depend on initial value x_0. This approach uses discretized numerical schemes to approximate the SDEs and requires the numerical solution $\widehat{X}_t$ to preserve the ergodicity. See [11, 20, 21, 26, 34, 38] and the references therein.

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

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

Then, for this fixed T, we can use all the methods mentioned in finite time analysis to estimate $\mathbb{E}[\varphi(X_T)]$. Milstein & Tretyakov [30] analyse the error of this kind of approach based on their quasi-symplectic method.

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 [6, 13, 19, 27, 31] 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 can avoid this problem, as proved by Lemaire[21] for the time-averaging approach. His adaptive construction has similarities to the one used in this paper.

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 [7, 8] is employed and non-nested timestepping is used to construct an adaptive MLMC [9]. Following the idea of Glynn and Rhee [10] 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 the finite and infinite time intervals are stated in Section 2. Section 3 introduces the MLMC schemes, and the relevant numerical experiments are provided in section 4. Section 5 has the proofs of the main theorems. Finally, section 6 concludes.

In this paper we consider both the finite time interval $[0, T]$ with $T > 0$ a fixed positive real number and the infinite time interval $[0, \infty)$. Let $(\Omega, \mathcal{F}, \mathbb{P})$ be a probability space with normal filtration $(\mathcal{F}_t)_{t \in [0, \infty)}$ for section 2 and $(\mathcal{F}_t)_{t \in (-\infty, 0]}$ 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 $\langle v, w \rangle \triangleq v_1 w_1 + v_2 w_2 + \ldots + v_m w_m$, 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 all $A \in \mathbb{R}^{m \times d}$.

2. Adaptive algorithm and theoretical results.

2.1. *Adaptive Euler-Maruyama method.* The proposed adaptive Euler-Maruyama discretisation is

$$(5) \qquad t_{n+1} = t_n + h_n, \quad \widehat{X}_{t_{n+1}} = \widehat{X}_{t_n} + f(\widehat{X}_{t_n})\, h_n + g(\widehat{X}_{t_n})\, \Delta W_n,$$

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 $\underline{t} \triangleq \max\{t_n : t_n \leq t\}$, $n_t \triangleq \max\{n : t_n \leq t\}$ for the nearest time point before time t, and its index. We define the piecewise constant interpolant process $\overline{X}_t = \widehat{X}_{\underline{t}}$ and also define the standard continuous interpolant [18] as

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

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

$$(6) \qquad \mathrm{d}\widehat{X}_t = f(\widehat{X}_{\underline{t}})\, \mathrm{d}t + g(\widehat{X}_{\underline{t}})\, \mathrm{d}W_t = f(\overline{X}_t)\, \mathrm{d}t + g(\overline{X}_t)\, \mathrm{d}W_t.$$

In the following 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. The main proofs are deferred to Section 6.

2.2. *Finite Time Interval.*

2.2.1. *Stability.*

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

(7) $$\|f(x)-f(y)\| + \|g(x)-g(y)\| \leq C_R \|x-y\|$$

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

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

and g satisfies the linear growth condition:

(9) $$\|g(x)\|^2 \leq \alpha \|x\|^2 + \beta.$$

Together, (8) and (9) imply the monotone condition

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

which is a key assumption in the analysis of Mao & Szpruch [25] and Mao [23] 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 $p>0$*

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

PROOF. The proof is given in Lemma 3.2 in [12]; the statement of that lemma makes stronger assumptions on f and g, corresponding to (12) and (13), but the proof only uses the conditions in Assumption 1. $\square$

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 satisfies the inequality*

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

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 (10), which is motivated by the requirements of the proof of the next theorem, is very similar to (8). Indeed, if (10) is satisfied then (8) 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 $\omega \in \Omega$, $\mathbb{P}(\exists N(\omega) < \infty \ s.t. \ t_{N(\omega)} \geq T) = 1$) 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*

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

PROOF. The proof is deferred to Section 6.1.   □

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 $0 < \delta \leq 1$, and consider the limit $\delta \to 0$.

Given a timestep function $h(x)$ which satisfies Assumption 2, ensuring stability as analyzed 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*

(11) $\delta \, \min(T, h(x)) \leq h^\delta(x) \leq \min(\delta \, T, h(x)),$

Examples of suitable $h(x)$ and $h^\delta(x)$ are given in Section 4. 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 with h satisfying Assumption 2, then for all $p > 0$*

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

PROOF. The proof is essentially identical to the uniform timestep Euler-Maruyama analysis in Theorem 2.2 in [12] by Higham, Mao & Stuart. The only change required by the use of an adaptive timestep is to note that

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

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

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:*

$$(12) \qquad \langle x - y, f(x) - f(y) \rangle \le \tfrac{1}{2}\alpha\|x - y\|^2,$$

and g satisfies the Lipschitz condition:

$$(13) \qquad \|g(x) - g(y)\|^2 \le \tfrac{1}{2}\alpha\|x - y\|^2.$$

In addition, f satisfies the polynomial growth Lipschitz condition

$$(14) \qquad \|f(x) - f(y)\| \le \left(\gamma\,(\|x\|^q + \|y\|^q) + \mu\right)\,\|x - y\|,$$

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

Note that setting $y = 0$ gives

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

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

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 with h satisfying Assumption 2, then for all $p > 0$ there exists a constant $C_{p,T}$ such that*

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

PROOF. The proof is deferred to Section 6.2.   □

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*

$$h(x) \ge \left(\xi\|x\|^q + \zeta\right)^{-1}.$$

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

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

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

PROOF. By Assumption 3 and Assumption 5, we have

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

Therefore, by Jensen's inequality, we obtain

$$(15) \qquad \mathbb{E}\left[(N_T - 1)^p\right] \le T^{p-1}\delta^{-p}\int_0^T \mathbb{E}\left[\left(\xi\|\overline{X}_t\|^q + \zeta + 1\right)^p\right]\mathrm{d}t$$

and the result is then an immediate consequence of Theorem 1.   □

The conclusion from Theorem 3 and Lemma 2 is that

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

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

First order strong convergence is achievable for the SDEs with uniform diffusion coefficient 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). *Assume f satisfies the Assumption 4 and in addition, f is differentiable, and f and ∇f satisfy the polynomial growth Lipschitz condition*

$$(16) \quad \|f(x) - f(y)\| + \|\nabla f(x) - \nabla f(y)\| \leq (\gamma (\|x\|^q + \|y\|^q) + \mu) \|x - y\|,$$

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

THEOREM 4 (Strong convergence for SDEs with uniform diffusion coefficient). *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*

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

PROOF. The proof is given in Theorem 4 in [3]   □

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 [18].

2.3. *Infinite Time Interval.* Now, we focus on a class of ergodic SDEs and show that the moment bounds and strong error bound are uniform in T under stronger assumptions.

2.3.1. *Stability.*

ASSUMPTION 7 (Dissipative condition). *f and g satisfy the locally Lipschitz condition (7) 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:*

$$(17) \qquad\qquad \langle x, f(x) \rangle \leq -\alpha \|x\|^2 + \beta,$$

and g is globally bounded and non-degenerate:

$$\|g(x)\|^2 \leq \beta. \tag{18}$$

Theorem 4.4 in [26] and Theorem 6.1 in [28] 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, $\forall\, t \geq 0$,*

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

PROOF. The result follows Proposition 3.1 (i) in [38]. $\qquad\square$

We now specify the critical assumption about the adaptive timestep for the 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*

$$\langle x, f(x)\rangle + \tfrac{1}{2}\, h(x)\, \|f(x)\|^2 \leq -\alpha\|x\|^2 + \beta. \tag{19}$$

Compared with the Assumption 2 in the finite time analysis, this assumption additionally bound h 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, $\forall t \geq 0$,*

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

PROOF. The proof is deferred to Section 6.3. $\qquad\square$

2.3.2. *Strong convergence.* To prove an order of strong convergence again requires new assumptions on f and g:

ASSUMPTION 9 (Contractive Lipschitz properties).   *For some fixed $p^* \in (2, \infty)$, there exist constants $\lambda, \eta > 0$ such that for all $x, y \in \mathbb{R}^m$, f and g satisfy the contractive Lipschitz condition:*

$$(20) \qquad \langle x - y, f(x) - f(y) \rangle + \frac{p^* - 1}{2} \|g(x) - g(y)\|^2 \leq -\lambda \|x - y\|^2,$$

and g satisfies the Lipschitz condition:

$$(21) \qquad \|g(x) - g(y)\|^2 \leq \eta \|x - y\|^2.$$

In addition, f satisfies the polynomial growth Lipschitz condition (14).

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 for some fixed $p^* \in (2, \infty)$, 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, satisfy that, $\forall \, t > 0$,*

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

PROOF.   First, we can define $e_t \triangleq X_t - Y_t$, and since X_t and Y_t are driven by the same Brownian motion, we get

$$de_t = (f(X_t) - f(Y_t)) \, dt + (g(X_t) - g(Y_t)) \, dW_t$$

By Itô's formula, we have for any $0 < t \leq T$,

$$
e^{\lambda p t} \|e_t\|^p - \|e_0\|^p \leq \int_0^t \lambda p \, e^{\lambda p s} \|e_s\|^p \, ds + \int_0^t p \langle e_s, f(X_s) - f(Y_s) \rangle e^{\lambda p s} \|e_s\|^{p-2} \, ds
$$
$$
+ \int_0^t \frac{p(p-1)}{2} \|g(X_s) - g(Y_s)\|^2 e^{\lambda p s} \|e_s\|^{p-2} ds
$$
$$
+ \int_0^t p \, e^{\lambda p s} \|e_s\|^{p-2} \langle e_s, (g(X_s) - g(Y_s)) \, dW_s \rangle.
$$

Therefore, by taking expectations on both sides and using the contractive Lipschitz property (20), we obtain that

$$\mathbb{E}\left[e^{\lambda p t} \|e_t\|^p\right] \leq \|e_0\|^p.$$

$\square$

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, $\forall t \geq 0$,*

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

PROOF.   The proof is deferred to Section 6.4.   □

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

LEMMA 5 (Number of timesteps).   *If the SDE satisfies Assumption 9, and the timestep function h^δ satisfies Assumption 3, with h 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.

PROOF.   Similar to the proof of Lemma 2, the uniform moment bound from Theorem 5 and the equation (15) give the result.   □

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

ASSUMPTION 10 (Enhanced contractive Lipschitz properties).   *Assume f satisfies Assumption 9 and in addition, f is differentiable, and f and ∇f satisfy the polynomial growth Lipschitz condition (16).*

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

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

PROOF.   The proof is given in Theorem 3 in [4].   □

3. Adaptive Multilevel Monte Carlo for invariant distributions.
We are interested in the problem of approximating:

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

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

3.1. *Algorithm.* To estimate $\mathbb{E}\left[\varphi(X_T)\right]$, 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 as explained in [9]. Consider the identity

$$(22) \qquad \mathbb{E}\left[\varphi_L\right] = \mathbb{E}\left[\varphi_0\right] + \sum_{\ell=1}^{L} \mathbb{E}\left[\varphi_\ell - \varphi_{\ell-1}\right],$$

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 fixed $M > 1$. 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 $0 < T_0 < T_1 < ... < T_\ell < ... < T_L = T$, 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 [10] to estimate the invariant measure of some Markov chains, we perform the coupling by starting a level ℓ fine path simulation at time $t_0^f = -T_\ell$ and a coarse path simulation at time $t_0^c = -T_{\ell-1}$ and terminate both paths at $t = 0$. Since the drift f and diffusion coefficient 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 and coarse paths 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 (22) 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

$$(23) \quad \widehat{Y} := \frac{1}{N_0} \sum_{n=1}^{N_0} \varphi(\widehat{X}_0^{(n,0)}) + \sum_{\ell=1}^{L} \left\{ \frac{1}{N_\ell} \sum_{n=1}^{N_\ell} \left(\varphi(\widehat{X}_0^{(n,\ell)}) - \varphi(\widehat{X}_0^{(n,\ell-1)}) \right) \right\},$$

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 $M = 2$.

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 Assumptions 1 and 9 and φ satisfies the polynomial growth Lipschitz condition (14), then there exists a constant $\mu_0 > 0$ depending on x_0 and constants in Lemma 3 and 4 such that*

$$(24) \qquad\qquad |\mathbb{E}\left[\varphi(X_t) - \pi(\varphi)\right]| \le \mu_0 \, e^{-\lambda t}.$$

PROOF. We can define a new random variable Y_0 which follows the invariant measure π, then the solution Y_t to the SDE with the initial value Y_0 will also follows the invariant measure for any $t > 0$. Therefore, by the

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

$t := -T_\ell$; $t^c := -T_{\ell-1}$; $t^f := -T_\ell$;
$h^c := 0$; $h^f := 0$;
$\Delta W^c := 0$; $\Delta W^f := 0$;
$\widehat{X}^c = x_0$; $\widehat{X}^f = x_0$;
while $t < 0$ **do**

    $t_{old} := t$;
    $t := \min(t^c, t^f)$;
    $\Delta W := N(0, t - t_{old})$;
    $\Delta W^c := \Delta W^c + \Delta W$;
    if $t = -T_{\ell-1}$ **then**
        $\Delta W^c := 0$;
    end
    $\Delta W^f := \Delta W^f + \Delta W$;
    if $t = t^c$ **then**

        update coarse path $\widehat{X}^c$ using h^c and ΔW^c;
        compute new adapted coarse path timestep $h^c = h^{2\delta}(\widehat{X}^c)$;
        $h^c := \min(h^c, -t^c)$;
        $t^c := t^c + h^c$;
        $\Delta W^c := 0$;
    end
    if $t = t^f$ **then**

        update fine path $\widehat{X}^f$ using h^f and ΔW^f;
        compute new adapted fine path timestep $h^f = h^{\delta}(\widehat{X}^f)$;
        $h^f := \min(h^f, -t^f)$;
        $t^f := t^f + h^f$;
        $\Delta W^f := 0$;
    end
end
Result: $\widehat{X}^f - \widehat{X}^c$

---

polynomial growth Lipschitz property of φ and Lemmas 3 and 4 and Hölder inequality, there exists constants $\mu_0, \mu_1 > 0$ such that

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

$$\square$$

Note that Assumption 9 is a sufficient condition for this Lemma. We use it here to show that the contractivity rate λ is a lower bound for the true

convergence rate λ^* and it is λ that determine the choice of T_ℓ shown in the following results.

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

LEMMA 7 (Variance of MLMC corrections for bounded diffusion coefficient). *If φ satisfies the polynomial growth Lipschitz condition* (14)*, the SDE satisfies Assumption 9 and the timestep function h^δ satisfies Assumption 3 with h satisfying Assumption 8 and $\delta = M^{-\ell}$ for each level, then for each level ℓ, there exist constants c_1 and c_2 such that the variance of correction $V_\ell := \mathbb{V}\left[\varphi(\widehat{X}_0^\ell) - \varphi(\widehat{X}_0^{\ell-1})\right]$ satisfies*

$$(25) \qquad V_\ell \le c_1\, M^{-\ell} + c_2\, \mathrm{e}^{-2\lambda T_{\ell-1}}.$$

PROOF. By the polynomial growth Lipschitz condition (14) of φ, Hölder inequality and Theorem 5, there exists a constant $\kappa > 0$ such that

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

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

Then, by Jensen's inequality, we obtain that

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

where

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

Theorem 6 implies that there exist a constant C_{p^*} which does not depend on T_ℓ such that

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

and Lemma 3 and Lemma 4 imply that there exists a constant C depending on x_0 and C_4 in Lemma 1 such that

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

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

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

$\square$

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

THEOREM 8 (MLMC for invariant measure).   *If φ satisfies the polynomial growth Lipschitz condition* (14), *the SDE satisfies Assumption* 9 *and the timestep function h^δ satisfies Assumption* 3 *with h satisfying Assumption* 8 *and $\delta = M^{-\ell}$ for each level, then by choosing suitable values for L and T_ℓ, N_ℓ for each level ℓ, there exists a constant c_3 such that the MLMC estimator* (23) *has a mean square error (MSE) with bound*

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

and an expected computational cost $\boldsymbol{C}_{MLMC}$ with bound

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

PROOF.   By Jensen's inequality, the mean square error can be decomposed into three parts:

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

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

If we set

(26)
$$
T_\ell = (\ell+1)\log M/2\lambda,
$$

then $V_\ell \leq (c_1 + c_2)M^{-\ell}$, which has the same order of magnitude as the variance bound for the standard MLMC theorem. Lemma 6 implies that

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

provided

$$(27) \qquad\qquad L \geq \left\lceil \frac{2|\log\varepsilon|}{\log M} + \frac{\log(6\mu_0^2)}{\log M} \right\rceil.$$

By Theorems 5 and 6, the polynomial growth Lipschitz condition (14) of φ and Hölder inequality, there exists constants $\kappa_1, \kappa_2 > 0$ such that

$$2\left|\mathbb{E}\left[\widehat{Y}\right] - \mathbb{E}\left[\varphi(X_{T_L})\right]\right|^2 = 2\left|\mathbb{E}\left[\varphi(\widehat{X}_{T_L}^L) - \varphi(X_{T_L})\right]\right|^2$$
$$\leq 2\kappa_1\,\mathbb{E}\left[\|\widehat{X}_{T_L}^L - X_{T_L}\|^4\right]^{1/2} \;\leq\; \kappa_2 M^{-L} \;\leq\; \frac{\varepsilon^2}{3},$$

provided

$$(28) \qquad\qquad L \geq \left\lceil \frac{2|\log\varepsilon|}{\log M} + \frac{\log(3\kappa_2)}{\log M} \right\rceil.$$

Combining the requirements (27) and (28), we choose to define

$$(29) \qquad\qquad L = \left\lceil \frac{2|\log\varepsilon|}{\log M} + \frac{\log\left(\max(6\mu_0^2, 3\kappa_2)\right)}{\log M} \right\rceil,$$

giving $L = O(|\log\varepsilon|)$ as $\varepsilon \to 0$. Therefore, we have $V_\ell = O(M^{-\ell})$ and $C_\ell = O(\ell M^\ell)$, where C_ℓ is the expected cost of a sample on level ℓ. Following the analysis in [8], choosing

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

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

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

Since

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

and

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

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

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

$\square$

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

THEOREM 9 (SDEs with uniform diffusion coefficient). *If φ satisfies the polynomial growth Lipschitz condition* (14), *and for the SDE, $m=d$, $g \equiv I_m$, f satisfies Assumption 6, and the timestep function h^{δ} satisfies Assumption 3 with h satisfying Assumption 8 and $\delta = M^{-\ell}$ for each level, then for each level ℓ, there exist constants c_1 and c_2 such that*

$$(30) \qquad V_{\ell} \leq c_1\, M^{-2\ell} + c_2\, \mathrm{e}^{-2\lambda T_{\ell-1}}.$$

Furthermore, by choosing suitable L, T_{ℓ} and N_{ℓ} for each level ℓ in the MLMC estimator (23), *one can achieve the MSE bound ε^2 at an expected computational cost bounded by*

$$\boldsymbol{C}_{MLMC} \leq c_3\,\varepsilon^{-2},$$

for some constant $c_3 > 0$.

PROOF. Following a similar argument to the proof of Lemma 7, Theorem 7 implies $V_{\ell} \leq c_1\, M^{-2\ell} + c_2\, \mathrm{e}^{-2\lambda T_{\ell-1}}$, and by choosing T_{ℓ} to be

$$(31) \qquad\qquad\qquad T_{\ell} = (\ell+1)\log M/\lambda,$$

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

$$C_{\ell} \ \leq \ C_0(\ell+1)M^{\ell} \ \leq \ C\, M^{(1+\epsilon)\ell}$$

for any $0 < \epsilon \ll 1$ and some $C > 0$. Therefore, the standard MLMC Theorem 1 in [8] is applicable with $\gamma < \beta$, giving an $O(\varepsilon^{-2})$ complexity. $\square$

Note that the choice of T_{ℓ} (31) for the equation with uniform diffusion coefficient is different from (26) for SDEs with bounded diffusion coefficient.

In other words, the strong convergence result and the contractive convergence rate λ together determine T_ℓ. In some cases, λ needs to be estimated numerically through Lemma 4. The difference in the variance convergence rate also affects the choice of M. Based on the analysis in [7], the optimal M for SDEs with general g is in the range $4 - 8$, while in the uniform diffusion coefficient case the optimal M is around 2.

4. Examples and numerical results. In this section we first give suggestions on the choice of adaptive function together with some example SDEs, then present numerical results for a finite time interval and their extension to the infinite time interval.

- **Scalar SDEs**. For any scalar SDE satisfying Assumption 1, we can choose the adaptive function:

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

- **Multi-dimensional SDEs**. For SDEs with a drift which, for some $\xi, \eta > 0$, satisfying the condition

(32) $$\langle x, f(x) \rangle \leq -\xi \|x\| \|f(x)\| + \eta,$$

one can use

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

Alternatively, if condition (32) is not satisfied, we can use

$$h^\delta(x) = \frac{\max(1, \|x\|^2)}{\max(1, \|f(x)\|^2)} \delta.$$

It is straightforward to check that these selected $h^\delta(x)$ satisfy Assumptions 3 and 5. These are only general suggestions and users can design a more specific and efficient adaptive function based on the applications. For example, consider the Ginzburg-Landau equation, which describes a phase transition from the theory of superconductivity [15, 18],

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

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

$$h^\delta(x) = \delta \, \min \left(T, \lambda^{-1} x^{-2} \right).$$
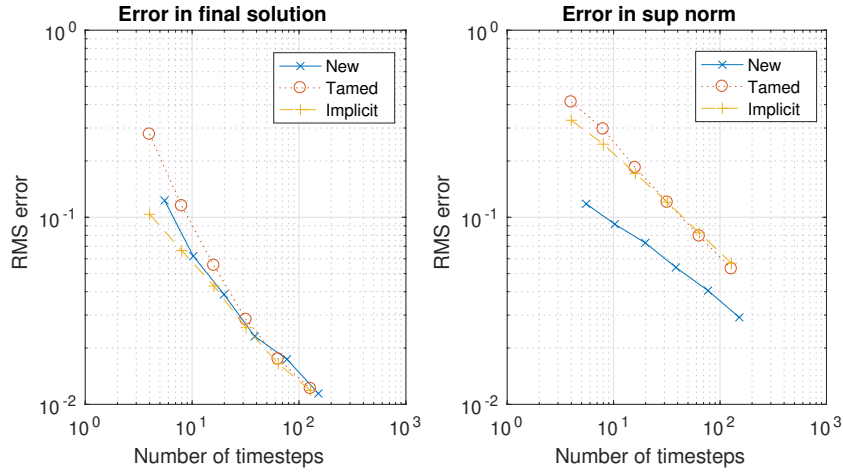
Fig 1. *Numerical results for finite time interval*

The second example is the Stochastic Lorenz equation, which is a three-dimensional system modeling convection rolls in the atmosphere [14]:

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

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.

The testcase taken from [16] 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

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

Figure 1 shows the the root-mean-square error plotted against the average number of timesteps. The plot on the left shows the error in $\widehat{X}_T$ at the terminal time, while the plot on the right shows the error in the maximum magnitude of the solution over the whole interval. 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 first order strong convergence. However, as reported in [16], 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 (8) and the contractive condition (20). Our interest is to compute $\pi(\varphi)$ where $\varphi(x) = (x + 1)^2$ satisfies a polynomial growth Lipschitz condition.

Since the probability density function π is

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

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

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

$$dY_t = -\left(1 + 3\,X_t^2\right) Y_t \, dt,$$

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

Figure 2 displays the variance of the multilevel correction on each level as a function of T; this is to be compared to the bound in result (30). The exponential part dominates the variance at the beginning, so the variance decays exponentially. As time increases, the $M^{-2\ell}$ term becomes the major part of the variance, and the variance stops decreasing.
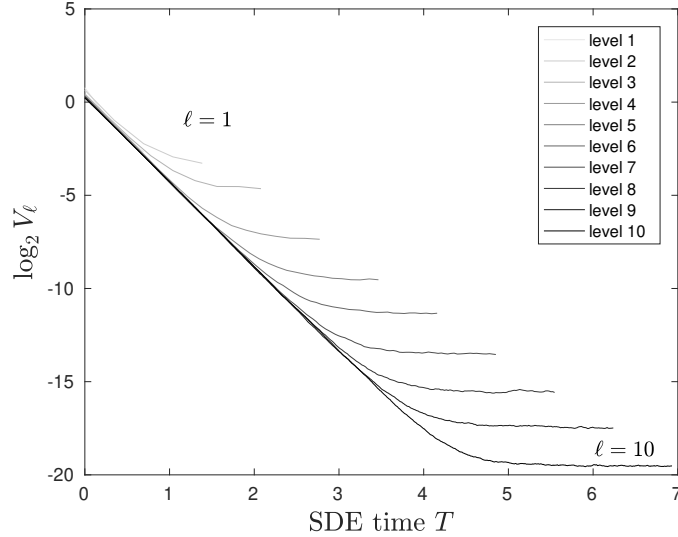
FIG 2. *Variance of corrections on each level ℓ*

Figure 3 presents the MLMC results. The top right plot shows first order convergence for the weak error and the top left plot shows second order convergence for the multilevel correction variance. Hence the computational cost for RMS accuracy ε is $O(\varepsilon^{-2})$ which is verified in the bottom right plot, while the bottom left plot shows the number of MLMC samples on each level as a function of the target accuracy. Here, we also compared our MLMC scheme with standard Monte Carlo (Standard MC) method directly simulating $\widehat{X}_{T_L}$, the adaptive scheme proposed by Lemaire using same step sequence as in example 7.1 in [21], and the MATLA in [34] with timestep $h = 0.1$. Both standard MC and adaptive scheme by Lemaire have the order $O(\varepsilon^{-3})$. MLMC and MATLA have the optimal complexity $O(\varepsilon^{-2})$. In this case, MATLA performs better due to the relative short mixing time and low correlations. However, MATLA and adaptive scheme by Lemaire only simulate one path and are difficult to perform parallel computing.

5. Extension to a larger class of ergodic SDEs.   In this section, we extend our adaptive scheme to a larger class of ergodic SDEs: the SDEs with negative Lyapunov exponent and then propose a new MLMC scheme with change of measure for the SDEs with positive Lyapunov exponent, that is chaotic system. For a detailed analysis of the Lyapunov exponent for the nonlinear stochastic system, please refer to Arnold [1].
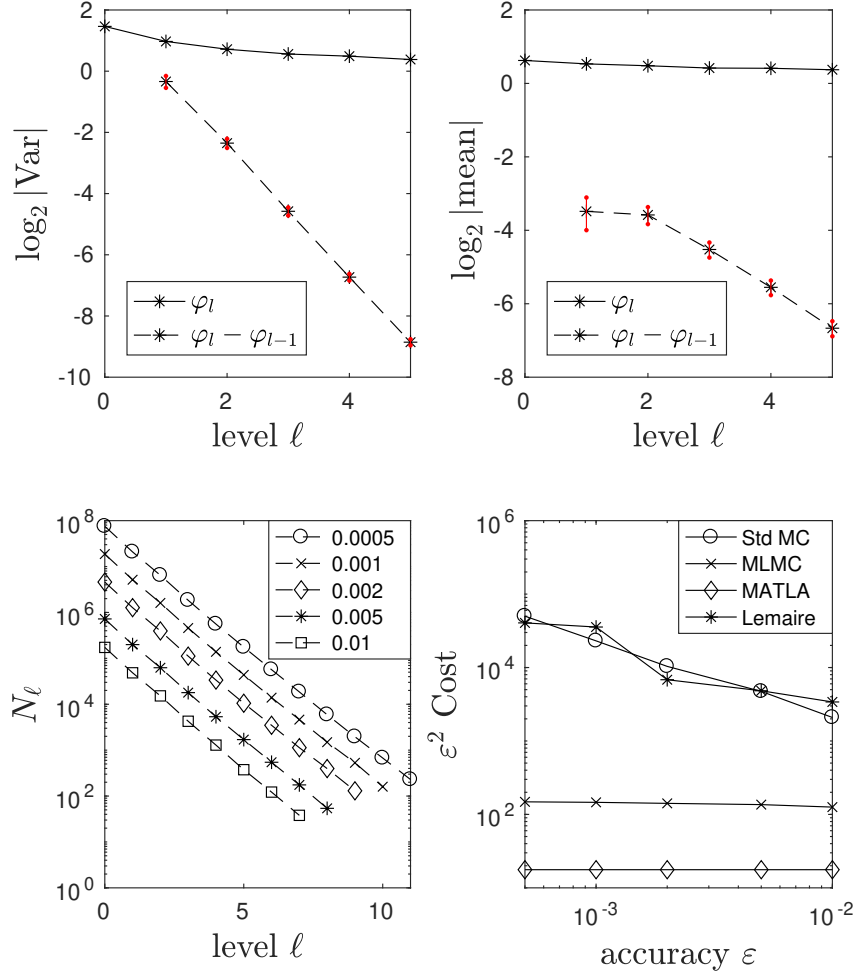
FIG 3. *Adaptive MLMC for invariant distribution*

5.1. *Systems with negative Lyapunov exponent.* In practice, the condition (20) in Assumption 9 is restrictive and means the SDE is contractive everywhere in the whole space. However, this condition is only a sufficient condition to make the adaptive MLMC work. Numerically, our adaptive MLMC does not need contractivity everywhere but does require contractivity in a global sense. Therefore, intuitively, our scheme works well for
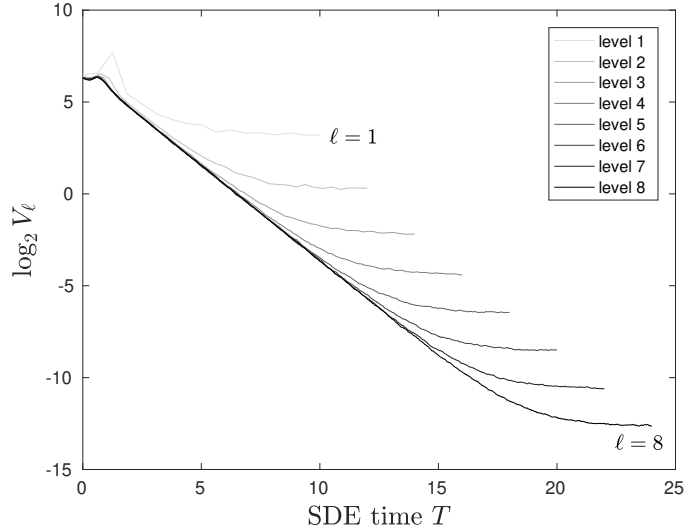
FIG 4. *Variance of corrections on each level ℓ for 100-dimensional double-well potential SDE*

systems with negative Lyapunov exponent. We present the numerical results for both a 100-dimensional SDE with double-well potential energy and the FENE model.

5.1.1. *Double-well potential energy.*   First, we apply adaptive MLMC for a 100-dimensional SDEs with double-well potential energy:

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

where $\mathbf{0}$ is the original point in $\mathbb{R}^{100}$. Our interest is to compute $\pi(\varphi)$ with $\phi(x) = \|x\|^2$ satisfying the polynomial growth Lipschitz condition. Although this SDE does not satisfies Assumption 9, it has negative Lyapunov exponent and the numerical estimation of the contractivity rate λ in Lemma 4 is 0.15. We choose T_ℓ based on equation (31).

Figure 4 shows the variance decays due to the contractivity and the first order strong convergence. The convergence results of adaptive MLMC and the comparison with other schemes are shown in Figure 5. We use MATLA with timestep $h = 0.02$ as the optimal scaling suggested in [33] and the adaptive scheme proposed by Lemaire using same step sequence as in example 7.1 in [21].
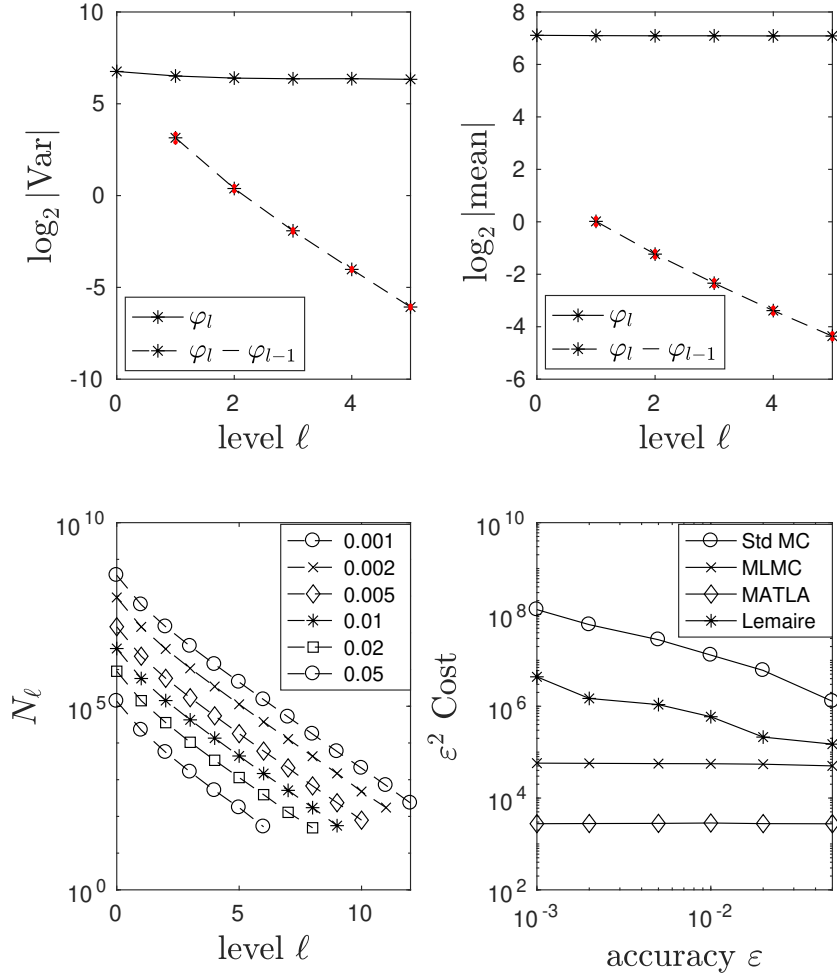
FIG 5. *Adaptive MLMC for invariant distribution (Double Well)*

5.1.2. *FENE model.* The FENE (Finitely Extensible Nonlinear Elastic) model is a Langevin equation describing the motion of a long-chained polymer in a liquid [2, 9]. The unusual feature of the FENE model is that the potential $V(x)$ becomes infinite for finite values of x. In the simplest case of a molecule with a single bond, X_t is three-dimensional and the SDE takes

the form

$$\mathrm{d}X_t = -\frac{16\,X_t}{1-\|X_t\|^2}\,\mathrm{d}t + \mathrm{d}W_t, \quad X_0 = 0,$$

which is defined on $\|X_t\| < 1$. The drift term ensures that $\|X_t\| < 1$ for all $t > 0$ almost surely. Also, it can be verified that $\langle x, f(x)\rangle \le 0$.

Because the SDE is not defined on all of $\mathbb{R}^3$, the theory in this paper is not applicable. However, it was one of the original motivations for the analysis in this paper, since it seems natural to use an adaptive timestep, taking smaller timestep as $\|\widehat{X}_t\|$ approaches 1, to maintain good accuracy, as the drift varies so rapidly near the boundary, and to greatly reduce the possibility of needing to clamp the computed solution to prevent it from crossing a numerical boundary at radius $1-\delta$ for some $\delta \ll 1$ [9].

Numerically we use the adaptive function h, h^δ to be

$$h(x) = (1-\|x\|)^2/8, \quad h^\delta(x) = 2^{-\ell}h(x).$$

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

The numerical results in Figure 6 show that the new scheme is considerably more accurate than either of the others, confirming that an adaptive timestep is desirable in this situation in which the drift varies enormously as $\|\widehat{X}_t\|$ approaches the maximum radius. Figure 7 shows that the adaptive MLMC also works well and achieve the optimal computational cost $O(\varepsilon^{-2})$ for the invariant measure computation. All other methods are not applicable here.

5.2. *Chaotic system.* For chaotic systems with positive Lyapunov exponent, for example, the stochastic Lorenz equation, our schemes will fail due to the loss of contractivity. In [5], we deal with these chaotic systems by introducing a coupling term between the coarse and fine paths, which leads to a change of measure and hence a Radon-Nikodym derivative in the Monte Carlo estimates. We give a brief outline of this shemem. Instead of considering the fine and coarse paths of the original SDEs under the same measure $\mathbb{P}$:

$$\begin{aligned} \mathrm{d}X_t^f &= f(X_t^f)\,\mathrm{d}t + \sigma\,\mathrm{d}W_t^{\mathbb{P}}, \\ \mathrm{d}X_t^c &= f(X_t^c)\,\mathrm{d}t + \sigma\,\mathrm{d}W_t^{\mathbb{P}}. \end{aligned}$$
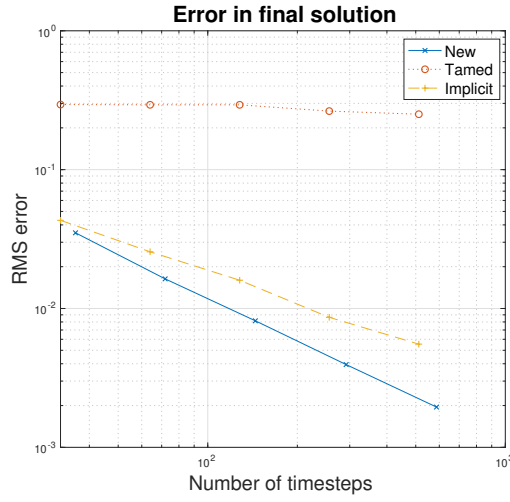
Fig 6. *Numerical results for FENE model in finite time interval*

We add a spring term with spring coefficient $S > 0$ for both fine path and coarse paths and consider both paths under different measures:

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

with

$$
\begin{aligned}
\mathrm{d}W_t^{\mathbb{Q}^f} &= \frac{S}{\sigma}(Y_t^c - Y_t^f)\,\mathrm{d}t + \mathrm{d}W_t^{\mathbb{P}}, \\
(33) \qquad \mathrm{d}W_t^{\mathbb{Q}^c} &= \frac{S}{\sigma}(Y_t^f - Y_t^c)\,\mathrm{d}t + \mathrm{d}W_t^{\mathbb{P}}.
\end{aligned}
$$

Therefore, under simulation measure $\mathbb{P}$, we obtain

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

The Girsanov theorem gives

$$
\begin{aligned}
\mathbb{E}^{\mathbb{P}}[\varphi(X_T^f)] - \mathbb{E}^{\mathbb{P}}[\varphi(X_T^c)] &= \mathbb{E}^{\mathbb{Q}^f}[\varphi(Y_T^f)] - \mathbb{E}^{\mathbb{Q}^c}[\varphi(Y_T^c)] \\
&= \mathbb{E}^{\mathbb{P}}\left[\varphi(Y_T^f)\frac{\mathrm{d}\mathbb{Q}^f}{\mathrm{d}\mathbb{P}_T} - \varphi(Y_T^c)\frac{\mathrm{d}\mathbb{Q}^c}{\mathrm{d}\mathbb{P}_T}\right],
\end{aligned}
$$
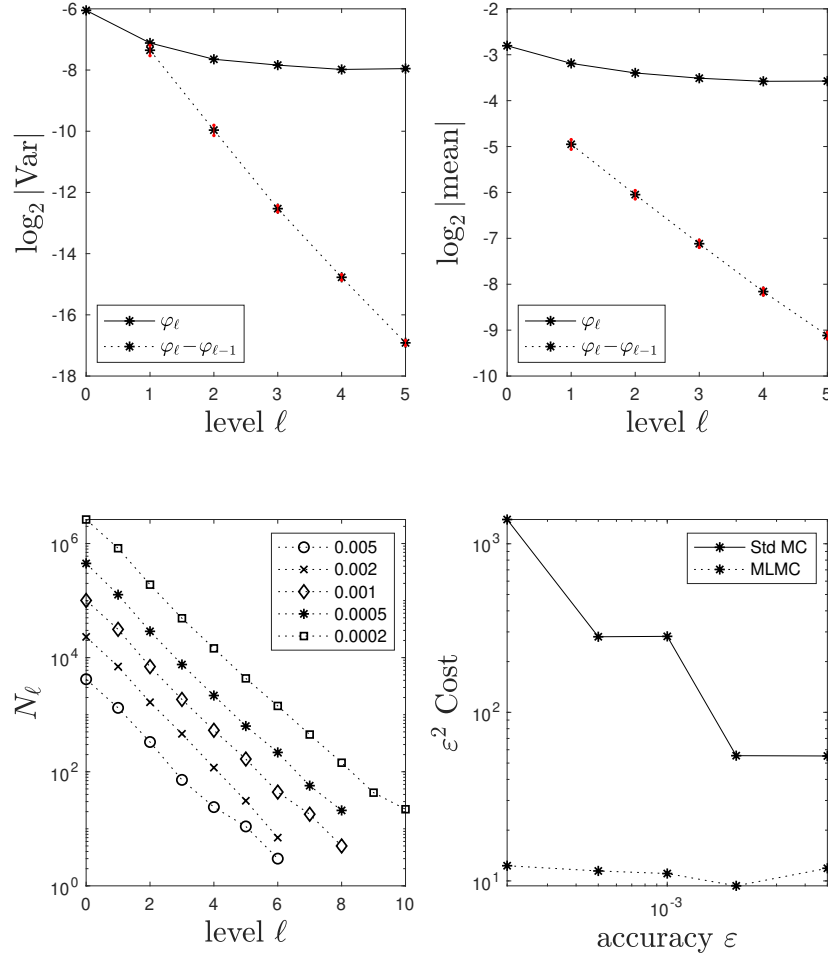
FIG 7. *Adaptive MLMC for invariant distribution (FENE)*

where $\frac{\mathrm{d}\mathbb{Q}^f}{\mathrm{d}\mathbb{P}_T}$ is the corresponding Radon-Nikodym derivative with following form:

$$\frac{\mathrm{d}\mathbb{Q}^f}{\mathrm{d}\mathbb{P}_T} = \exp\left(-\int_0^T \left\langle \frac{S}{\sigma}(Y_t^f - Y_t^c), \mathrm{d}W_t^{\mathbb{P}} \right\rangle - \frac{1}{2}\int_0^T \frac{S^2}{\sigma^2}\left\|Y_t^f - Y_t^c\right\|^2 \mathrm{d}t\right)$$

and $\frac{\mathrm{d}\mathbb{Q}^c}{\mathrm{d}\mathbb{P}_T}$ is similar.

The benefit of this technique is that under measure $\mathbb{P}$, we recover the contractivity between Y_t^c and Y_t^f using sufficiently large $S > 0$,

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

and the variance of the level estimator increases linearly in T instead of the exponential increase of standard MLMC. For the detailed numerical scheme, see section 2 in [5]. Note that it is not possible in general to use different simulation times T_ℓ on different multilevel levels as in the current MLMC scheme – instead the same simulation time T has to be used on all levels, with T being adjusted (automatically) to ensure the necessary weak convergence as the target error approaches zero.

6. Proofs. This section has the proofs of the four main theorems in this paper, two on stability, and two on the order of strong convergence.

6.1. *Theorem 1.*

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

The proof is given for $p \ge 4$; the result for $0 \le p < 4$ follows from Hölder's inequality.

Step 1: K-Scheme definition

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

$$(34) \qquad \widehat{X}_{t_{n+1}}^K = P_K\left(\widehat{X}_{t_n}^K + f(\widehat{X}_{t_n}^K)\,h_n + g(\widehat{X}_{t_n}^K)\,\Delta W_n\right),$$

where $P_K(Y) \triangleq \min(1, K/\|Y\|)\,Y$ and therefore $\|\widehat{X}_{t_n}^K\| \le K$, $\forall n$. The piece-wise constant approximation for intermediate times is again $\overline{X}_t^K = \widehat{X}_{\underline{t}}^K$, and the continuous approximation is

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

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

$$h_{min}^K \triangleq \inf_{\|x\| \le K} h(x) > 0.$$

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

Step 2: pth-moment of K-Scheme solution

$\|P_K(Y)\| \leq \|Y\|$, so if we define $\phi(x) \triangleq x + h(x)f(x)$, then (34) gives

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

Using condition (10) for $h(x)$ then gives

$$
\begin{aligned}
\|\widehat{X}^K_{t_{n+1}}\|^2 &\leq \|\widehat{X}^K_{t_n}\|^2 + 2\,\alpha\|\widehat{X}^K_{t_n}\|^2 h_n + 2\,\beta\,h_n \\
&\quad + 2\,\langle \phi(\widehat{X}^K_{t_n}), g(\widehat{X}^K_{t_n})\,\Delta W_n\rangle + \|g(\widehat{X}^K_{t_n})\,\Delta W_n\|^2.
\end{aligned}
$$
(35)

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

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

and therefore we obtain

$$
\begin{aligned}
\|\widehat{X}^K_t\|^2 &\leq \|\widehat{X}^K_{\underline{t}}\|^2 + 2\,\alpha\|\widehat{X}^K_{\underline{t}}\|^2(t-\underline{t}) + 2\,\beta\,(t-\underline{t}) \\
&\quad + 2\,\langle \widehat{X}^K_{\underline{t}} + f(\widehat{X}^K_{\underline{t}})\,(t-\underline{t}), g(\widehat{X}^K_{\underline{t}})\,(W_t - W_{\underline{t}})\rangle \\
&\quad + \|g(\widehat{X}^K_{\underline{t}})\,(W_t - W_{\underline{t}})\|^2.
\end{aligned}
$$
(37)

Summing (35) over multiple timesteps and then adding (37) gives

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

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

inequality, we obtain

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

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

For any $0 \leq t \leq T$ we take the supremum on both sides of inequality (38) and then take the expectation to obtain

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

where

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

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

$$
I_1 \leq \|X_0\|^p + (2\alpha)^{p/2} T^{p/2-1} \int_0^t \mathbb{E}\left[\sup_{0\leq u\leq s} \|\widehat{X}_u^K\|^p\right]\,\mathrm{d}s + (2\,\beta\,T)^{p/2}.
$$

For I_2, we begin by noting that due to condition (10), for $u < \underline{t}$ we have

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

and hence by Jensen's inequality

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

In addition, the linear growth condition (9) gives

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

and combining the last two equation, there exists a constant $c_{p,T}$ depending on p and T, in addition to α, β, such that

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

Then, by the Burkholder-Davis-Gundy inequality, there is a constant C_p such that

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

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

$$
\tag{39}
\mathbb{E}[\sup_{t_k\leq u\leq s} \|W_u - W_{t_k}\|^p \mid \mathcal{F}_{t_k}] = c_p\,(s-\underline{s})^{p/2}.
$$

Using Jensen's inequality and (39) with $s \equiv t_{k+1}$ so that $s - \underline{s} = h_k$,

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

Using condition (9), and Jensen's inequality, we then obtain

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

For I_4, using (36) and following the same argument as for I_2, there exists a constant $c_{p,T}$ depending on both p and T such that

$$\|\overline{X}_s^K + f(\overline{X}_s^K)(s-\underline{s})\|^{p/2} \|g(\overline{X}_s^K)\|^{p/2} \leq c_{p,T}\left(\|\overline{X}_s^K\|^p + 1 \right).$$

Therefore, again using (39),

$$
\begin{aligned}
I_4 &\leq 2^{p/2}\,\mathbb{E}\left[\sup_{0 \leq s \leq t} \left| \langle \overline{X}_s^K + f(\overline{X}_s^K)(s-\underline{s}), g(\overline{X}_s^K)\,(W_s - W_{\underline{s}}) \rangle \right|^{p/2} \right] \\[2mm]
&\leq c_{p,T}\, 2^{p/2}\, \mathbb{E}\left[\sum_{k=0}^{n_t-1} \left(\|\overline{X}_{t_k}^K\|^p + 1 \right) \sup_{t_k \leq s < t_{k+1}} \|(W_s - W_{\underline{s}})\|^{p/2} \right. \\[2mm]
&\qquad\qquad \left. + \left(\|\overline{X}_t^K\|^p + 1 \right) \sup_{\underline{t} \leq s \leq t} \|(W_s - W_{\underline{s}})\|^{p/2} \right] \\[2mm]
&\leq c_{p/2}\, c_{p,T}\, 2^{p/2}\, T^{p/4-1}\, \mathbb{E}\left[\sum_{k=0}^{n_t-1} \left(\|\overline{X}_{t_k}^K\|^p + 1 \right) h_k + \left(\|\overline{X}_t^K\|^p + 1 \right)(t-\underline{t}) \right] \\[2mm]
&\leq c_{p/2}\, c_{p,T}\, 2^{p/2}\, T^{p/4-1} \left(\int_0^t \mathbb{E}\left[\sup_{0 \leq u \leq s} \|\widehat{X}_u^K\|^p \right] \mathrm{d}s \;+\; T \right).
\end{aligned}
$$

Similarly, using the same definition for c_p, we have

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

Collecting together the bounds for I_1, I_2, I_3, I_4, I_5, we conclude that there exist constants $C_{p,T}^1$ and $C_{p,T}^2$ such that

$$\mathbb{E}\left[\sup_{0 \leq s \leq t} \|\widehat{X}_s^K\|^p \right] \leq C_{p,T}^1 + C_{p,T}^2 \int_0^t \mathbb{E}\left[\sup_{0 \leq u \leq s} \|\widehat{X}_u^K\|^p \right] \mathrm{d}s,$$

and Grönwall's inequality gives the result

$$\mathbb{E}\left[\sup_{0 \leq t \leq T} \|\widehat{X}_t^K\|^p \right] \leq C_{p,T}^1 \exp(C_{p,T}^2\,T) \triangleq C_{p,T} < \infty.$$

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

For any $\omega \in \Omega$, $\widehat{X}_t = \widehat{X}_t^K$ for all $0 \leq t \leq T$ if, and only if, $\sup_{0 \leq t \leq T} \|\widehat{X}_t\| \leq K$. Therefore, by the Markov inequality,

$$\mathbb{P}(\sup_{0 \leq t \leq T} \|\widehat{X}_t\| < K) = \mathbb{P}(\sup_{0 \leq t \leq T} \|\widehat{X}_t^K\| < K) \geq 1 - \mathbb{E}[\sup_{0 \leq t \leq T} \|\widehat{X}_t^K\|^4]/K^4 \to 1$$

as $K \to \infty$. Hence, almost surely, $\sup_{0 \leq t \leq T} \|\widehat{X}_t\| < \infty$ and T is attainable. Also,

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

and for $0 < K_1 \leq K_2$,

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

Therefore, by the Monotone Convergence Theorem,

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

$\square$

6.2. *Theorem 3.*

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

The proof is again given for $p \geq 4$; the result for $0 \leq p < 4$ follows from Hölder's inequality.

We start by combining the original SDE with (6) to obtain

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

and then by Itô's formula, together with $e_0 = 0$, we get

$$\|e_t\|^2 \leq 2\int_0^t \langle e_s, f(\widehat{X}_s) - f(X_s)\rangle\,\mathrm{d}s - 2\int_0^t \langle e_s, f(\widehat{X}_s) - f(\overline{X}_s)\rangle\,\mathrm{d}s$$
$$+ \int_0^t \|g(\overline{X}_s) - g(X_s)\|^2\,\mathrm{d}s + 2\int_0^t \langle e_s, (g(\overline{X}_s) - g(X_s))\,\mathrm{d}W_s\rangle.$$

Using the conditions in Assumption 4, (12) implies that

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

(14) implies that

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

where $L(x,y) \triangleq \gamma(\|x\|^q + \|y\|^q) + \mu$, and (13) gives

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

Hence,

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

and then by Jensen's inequality we obtain

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

Taking the supremum of each side, and then the expectation yields

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

By the Hölder inequality,

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

and $\mathbb{E}\left[\left(L(\widehat{X}_s, \overline{X}_s)^2 + \alpha\right)^p\right]$ is uniformly bounded on $[0, T]$ due to the stability property in Theorem 1.

In addition, by the Burkholder-Davis-Gundy inequality (which gives the constant C_p which depends only on p) followed by Jensen's inequality plus the Lipschitz condition for g, we obtain

$$\mathbb{E}\left[\sup_{0 \le s \le t}\left|\int_0^s \langle e_u, (g(\overline{X}_u) - g(X_u))\,\mathrm{d}W_u\rangle\right|^{p/2}\right]$$

$$\le\ C_p\,\mathbb{E}\left[\left(\int_0^t \|e_s\|^2 \|g(\overline{X}_s) - g(X_s)\|^2\,\mathrm{d}s\right)^{p/4}\right].$$

$$\le\ C_p\,T^{p/4-1}\,(\tfrac{1}{2}\alpha)^{p/4}\mathbb{E}\left[\int_0^t \|e_s\|^{p/2}\|\overline{X}_s - X_s\|^{p/2}\,\mathrm{d}s\right]$$

$$\le\ C_p\,T^{p/4-1}\,(\tfrac{1}{2}\alpha)^{p/4}\mathbb{E}\left[\int_0^t \tfrac{1}{2}\|e_s\|^p + \tfrac{1}{2}\|\overline{X}_s - X_s\|^p\,\mathrm{d}s\right]$$

$$\le\ C_p\,T^{p/4-1}\,(\tfrac{1}{2}\alpha)^{p/4}\mathbb{E}\left[\int_0^t (\tfrac{1}{2}+2^{p-2})\|e_s\|^p + 2^{p-2}\|\widehat{X}_s - \overline{X}_s\|^p\,\mathrm{d}s\right].$$

Hence, using $\mathbb{E}[\|\widehat{X}_s - \overline{X}_s\|^p] \le (\mathbb{E}[\|\widehat{X}_s - \overline{X}_s\|^{2p}])^{1/2}$, there are constants $C_{p,T}^1, C_{p,T}^2$ such that
(40)
$$\mathbb{E}\left[\sup_{0 \le s \le t}\|e_s\|^p\right] \le C_{p,T}^1 \int_0^t \mathbb{E}\left[\sup_{0 \le u \le s}\|e_u\|^p\right]\mathrm{d}s + C_{p,T}^2 \int_0^t \left(\mathbb{E}\left[\|\widehat{X}_s - \overline{X}_s\|^{2p}\right]\right)^{1/2}\mathrm{d}s.$$

For any $s \in [0, T]$, $\widehat{X}_s - \overline{X}_s = f(\widehat{X}_{\underline{s}})(s - \underline{s}) + g(\widehat{X}_{\underline{s}})(W_s - W_{\underline{s}})$, and hence, by a combination of Jensen and Hölder inequalities, we get

$$\mathbb{E}\left[\|\widehat{X}_s - \overline{X}_s\|^{2p}\right]\ \le\ 2^{2p-1}\left(\mathbb{E}\left[\|f(\widehat{X}_{\underline{s}})\|^{4p}\right]\mathbb{E}\left[(s - \underline{s})^{4p}\right]\right)^{1/2}$$

(41)
$$+\ 2^{2p-1}\left(\mathbb{E}\left[\|g(\widehat{X}_{\underline{s}})\|^{4p}\right]\mathbb{E}\left[\|W_s - W_{\underline{s}}\|^{4p}\right]\right)^{1/2}.$$

$\mathbb{E}[\|f(\widehat{X}_{\underline{s}})\|^{4p}]$ and $\mathbb{E}[\|g(\widehat{X}_{\underline{s}})\|^{4p}]$ are both uniformly bounded on $[0, T]$ due to stability and the polynomial bounds on the growth of f and g. Furthermore, we have $\mathbb{E}[(s - \underline{s})^{4p}] \le (\delta T)^{4p} \le \delta^{2p}T^{4p}$, and by standard results there is a constant c_p such that $\mathbb{E}[\|W_s - W_{\underline{s}}\|^{4p}] = \mathbb{E}[\,\mathbb{E}[\|W_s - W_{\underline{s}}\|^{4p}\,|\,\mathcal{F}_{\underline{s}}]\,] \le c_p(\delta T)^{2p}$. Hence, there exists a constant $C_{p,T}^3 > 0$ such that $\mathbb{E}[\,\|\widehat{X}_s - \overline{X}_s\|^{2p}] \le C_{p,T}^3\,\delta^p$, and therefore equation (40) gives us

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

and Grönwall's inequality then provides the final result.   □

6.3. *Theorem 5.*

PROOF. For simplicity, for $\alpha > 0$, we can define $\widehat{M}_t^{\alpha,p} \triangleq \sup_{0 \le s \le t} \mathrm{e}^{\alpha ps} \|\widehat{X}_s\|^p$, and $\overline{M}_t^{\alpha,p} \triangleq \sup_{0 \le s \le t} \mathrm{e}^{\alpha ps} \|\overline{X}_s\|^p$, which implies

$$\overline{M}_t^{\alpha,p} \le \mathrm{e}^{\alpha p h_{\max}} \widehat{M}_t^{\alpha,p}, \tag{42}$$

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

$$
\begin{aligned}
\int_0^t \mathrm{e}^{\gamma ps/2} \|\overline{X}_s\|^{p/2} \mathrm{d}s &\le \overline{M}_t^{\alpha,p/2} \int_0^t \mathrm{e}^{(\gamma-\alpha)ps/2} \mathrm{d}s \\
&\le \frac{2\mathrm{e}^{(\gamma-\alpha)pt/2}}{p\,(\gamma-\alpha)} \mathrm{e}^{\alpha p h_{\max}/2} \widehat{M}_t^{\alpha,p/2}.
\end{aligned}
\tag{43}
$$

provided $\gamma > \alpha > 0$. For $\widehat{M}_t^{\alpha,p}$, Young's inequality gives, for any $\xi > 0$,

$$\widehat{M}_t^{\alpha,p/2} \le \xi\, \widehat{M}_t^{\alpha,p} + \frac{1}{4\,\xi}. \tag{44}$$

By theorem 1, we know T is almost surely attainable. Therefore we can directly analyse our discretization scheme without the K truncation. The proof proceeds in three steps. First, we derive an upper bound for $\mathrm{e}^{\alpha pt} \|\widehat{X}_t\|^p$. Second, we show that the moments $\mathbb{E}[\widehat{M}_t^{\alpha,p}]$ and $\mathbb{E}[\overline{M}_t^{\alpha,p}]$ are each bounded by $C_p \mathrm{e}^{\alpha pt}$ where C_p is a constant which only depends on p, x_0, $h_{\max}$ and the constants α, β in Assumption 8. Finally, we get the uniform bound for $\mathbb{E}[\|\widehat{X}_t\|^p]$ and $\mathbb{E}[\|\overline{X}_t\|^p]$. The proof is given for $p \ge 4$; the result for $0 < p < 4$ follows from Hölder's inequality.

Step 1: If we define $\phi(x) \triangleq x + h(x)f(x)$, then (5) gives

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

Using condition (19) for h then gives

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

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

$$
\begin{aligned}
\mathrm{e}^{2\alpha t_{n+1}} \|\widehat{X}_{t_{n+1}}\|^2 &\le \mathrm{e}^{2\alpha t_n} \|\widehat{X}_{t_n}\|^2 + 2\mathrm{e}^{2\alpha(t_n+h_{\max})}\,\beta\,h_n + \mathrm{e}^{2\alpha(t_n+h_{\max})}\beta\|\Delta W_n\|^2 \\
&\quad + 2\,\mathrm{e}^{2\alpha t_{n+1}} \langle \phi(\widehat{X}_{t_n}), g(\widehat{X}_{t_n})\,\Delta W_n \rangle.
\end{aligned}
\tag{45}
$$

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

$$(46) \qquad \langle \widehat{X}_{\underline{t}}, f(\widehat{X}_{\underline{t}}) \rangle + \tfrac{1}{2}(t-\underline{t})\|f(\widehat{X}_{\underline{t}})\|^2 \leq -\alpha\|\widehat{X}_{\underline{t}}\|^2 + \beta,$$

and therefore we obtain

$$\begin{aligned}
\mathrm{e}^{2\alpha t}\|\widehat{X}_t\|^2 &\leq \mathrm{e}^{2\alpha \underline{t}}\|\widehat{X}_{\underline{t}}\|^2 \ + \ 2\mathrm{e}^{2\alpha(\underline{t}+h_{\max})}\beta\,(t-\underline{t}) + \mathrm{e}^{2\alpha(\underline{t}+h_{\max})}\beta\|W_t - W_{\underline{t}}\|^2 \\
(47) &\qquad + 2\,\mathrm{e}^{2\alpha t}\langle \phi(\widehat{X}_{\underline{t}}), g(\widehat{X}_{\underline{t}})\,(W_t - W_{\underline{t}})\rangle.
\end{aligned}$$

Summing (45) over multiple timesteps and then adding (47) gives

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

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

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

(49)

Step 2: For any $0 \leq t \leq T$, we take the supremum on both sides of inequality (49) and then take the expectation to obtain

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

where

$$\begin{aligned}
I_1 \ &= \ \|x_0\|^p + \left(2\beta \int_0^t \mathrm{e}^{2\alpha s}\,\mathrm{d}s \right)^{p/2}, \\
I_2 \ &= \ \mathbb{E}\left[\sup_{0\leq s\leq t} \left| 2\int_0^s \mathrm{e}^{2\alpha(\underline{u}+h(\overline{X}_u))}\langle \phi(\overline{X}_u), g(\overline{X}_u)\,\mathrm{d}W_u\rangle \right|^{p/2} \right],
\end{aligned}$$

$$
I_3 = \mathbb{E}\left[\left(\beta \sum_{k=0}^{n_t-1} e^{2\alpha t_k}\|\Delta W_k\|^2\right)^{p/2}\right],
$$

$$
I_4 = \mathbb{E}\left[\sup_{0\le s\le t}\left|2e^{2\alpha \underline{s}}\langle \overline{X}_s + f(\overline{X}_s)\,(s-\underline{s}), g(\overline{X}_s)(W_s-W_{\underline{s}})\rangle\right|^{p/2}\right],
$$

$$
I_5 = \mathbb{E}\left[\sup_{0\le s\le t}\beta^{p/2}e^{\alpha p\underline{s}}\|W_s-W_{\underline{s}}\|^p\right].
$$

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

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

By the Burkholder-Davis-Gundy inequality, there exist constants C_p^1 such that

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

Due to condition (19), for $u<\underline{t}$ we have

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

hence by Jensen's inequality and the boundedness condition (18) of g , we obtain

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

One variant of the Jensen's inequality is

$$
(50) \qquad \left|\int_0^t \Phi(s)\,e^{\gamma s}\,ds\right|^p \le \left(\int_0^t e^{\gamma s}\,ds\right)^{p-1}\int_0^t |\Phi(s)|^p e^{\gamma s}\,ds,
$$

for some function Φ. Therefore, using Jensen's inequality (50) with $\gamma=2\alpha$, followed by (43) with $\gamma=(1+4/p)\alpha$ and then (44) with $\xi=e^{-\alpha pt/2}\zeta$, there

exists a constant C_p^2 which is linearly dependent on ζ^{-1} such that

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

Using discrete version of Jensen's inequality (50) we obtain

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

where c_p is defined in equation (39).

In considering I_4, we start by observing that for $t_k\leq s<t_{k+1}$

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

In addition, using (46) and following the same argument as for I_2, we have

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

Therefore, combining the estimation (51), (43) with $\gamma=2\alpha$ and (44) with

$\xi = \mathrm{e}^{-\alpha pt/2}\zeta$, there exists C_p^3 which is linearly dependent on ζ^{-1} such that

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

Similarly, again using the same definition for c_p, we have

$$
I_5 \leq c_p\,\beta^{p/2}h_{\max}{}^{p/2-1}\mathrm{e}^{\alpha pt}/(\alpha p).
$$

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

$$
\mathbb{E}\left[\widehat{M}_t^{\alpha,p}\right] \leq \tfrac{1}{2}\,\mathbb{E}\left[\widehat{M}_t^{\alpha,p}\right] + C_p^4\|x_0\|^p + C_p^5\,\mathrm{e}^{\alpha pt},
$$

and hence

$$
\mathbb{E}\left[\widehat{M}_t^{\alpha,p}\right] \leq 2C_p^4\,\|x_0\|^p + 2C_p^5\,\mathrm{e}^{\alpha pt}.
$$

Step 3: Due to the definition of $\overline{M}_t^{\alpha,p}$ and inequality (42), for any $t \geq 0$,

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

and similarly

$$
\mathbb{E}\left[\|\widehat{X}_t\|^p\right] \;\leq\; \mathrm{e}^{-\alpha pt}\,\mathbb{E}\left[\widehat{M}_t^{\alpha,p}\right] \;\leq\; 2\,C_p^4\,\|x_0\|^p + 2C_p^5 \;<\; C_p.
$$

$\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\square$

6.4. *Theorem 6.*

PROOF. Similar to the subsection 6.2, we start by combining the original SDE (1) with (6) to obtain

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

and then by Itô's formula and Young's inequality, together with $e_0 = 0$, and λ, η as defined in Assumption 9, we get

$$\mathrm{e}^{\lambda p t/2}\|e_t\|^p \leq \int_0^t \frac{p\lambda}{2} \mathrm{e}^{\lambda p s/2}\|e_s\|^p \, \mathrm{d}s + \int_0^t p\langle e_s, f(\overline{X}_s) - f(X_s)\rangle \mathrm{e}^{\lambda p s/2}\|e_s\|^{p-2} \, \mathrm{d}s$$

$$+ \int_0^t \frac{p(p-1)}{2}\|g(\overline{X}_s) - g(X_s)\|^2 \mathrm{e}^{\lambda p s/2}\|e_s\|^{p-2} \, \mathrm{d}s$$

$$+ \int_0^t p \left\langle e_s, (g(\overline{X}_s) - g(X_s))\mathrm{e}^{\lambda p s/2} \|e_s\|^{p-2} \, \mathrm{d}W_s \right\rangle$$

$$\leq \int_0^t \frac{p\lambda}{2} \mathrm{e}^{\lambda p s/2}\|e_s\|^p \, \mathrm{d}s + \int_0^t p\langle e_s, f(\widehat{X}_s) - f(X_s)\rangle \mathrm{e}^{\lambda p s/2} \|e_s\|^{p-2} \, \mathrm{d}s$$

$$- \int_0^t p\langle e_s, f(\widehat{X}_s) - f(\overline{X}_s)\rangle \mathrm{e}^{\lambda p s/2} \|e_s\|^{p-2} \, \mathrm{d}s$$

$$+ p\int_0^t \left(\frac{p-1}{2} + \frac{\lambda}{4\eta} \right) \|g(\widehat{X}_s) - g(X_s)\|^2 \mathrm{e}^{\lambda p s/2} \|e_s\|^{p-2} \mathrm{d}s$$

$$+ p\int_0^t \left(\frac{p-1}{2} + \frac{\eta(p-1)^2}{\lambda} \right) \|g(\widehat{X}_s) - g(\overline{X}_s)\|^2 \mathrm{e}^{\lambda p s/2} \|e_s\|^{p-2} \mathrm{d}s$$

$$+ \int_0^t p \left\langle e_s, (g(\overline{X}_s) - g(X_s))\mathrm{e}^{\lambda p s/2} \|e_s\|^{p-2} \, \mathrm{d}W_s \right\rangle$$

Using the conditions in Assumption 9, (21) implies that

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

(20) and (21) imply that

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

(14) and Young inequality implies that

$$\left| \langle e_s, f(\widehat{X}_s) - f(\overline{X}_s)\rangle \right| \leq \|e_s\| L(\widehat{X}_s, \overline{X}_s) \|\widehat{X}_s - \overline{X}_s\|$$

$$\leq \frac{\lambda}{8}\|e_s\|^2 + \frac{2}{\lambda}L(\widehat{X}_s, \overline{X}_s)^2\|\widehat{X}_s - \overline{X}_s\|^2.$$

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

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

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

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

Taking the expectation of each side yields
(52)
$$
\mathbb{E}\left[\mathrm{e}^{\lambda p t/2}\|e_t\|^p\right] \leq 2\left(\frac{8(p-2)}{p\lambda}\right)^{p/2-1}\int_0^t \mathbb{E}\left[\hat{L}(\widehat{X}_s, \overline{X}_s)^{p/2}\|\widehat{X}_s - \overline{X}_s\|^p\right]\mathrm{e}^{\lambda p s/2}\mathrm{d}s.
$$

By the Hölder inequality,

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

and $\mathbb{E}\left[\hat{L}(\widehat{X}_s, \overline{X}_s)^p\right]$ can be bounded by a constant C_p^1 due to the stability property in Theorem 5. Then following the same analysis for $\mathbb{E}\left[\|\widehat{X}_s - \overline{X}_s\|^{2p}\right]$ in subsection 6.2 together with the uniform moments bound, there exists a constant C_p^2 such that

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

which provides the final result:

$$
\mathbb{E}\left[\|e_t\|^p\right] \leq \frac{4}{\lambda p}\left(\frac{8(p-2)}{p\lambda}\right)^{p/2-1}\sqrt{C_p^1 C_p^2}\ \delta^{p/2} \triangleq C_p\,\delta^{p/2},\ \forall\ t \geq 0
$$

$\square$

7. Conclusions and future work. 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 diffusion coefficient 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 SDEs with uniform diffusion coefficient, 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 coarse path in each level ℓ. All the schemes work well and numerical experiments support the theoretical results.

One direction for extension of the theory for finite time interval is to SDEs with a diffusion coefficient which is not globally Lipschitz, but instead satisfies the Khasminskii-type condition used by Mao & Szpruch [23, 25], Sabanis [35] and Szpruch & Zhāng [37]. Another possibility is to use a Lyapunov function $V(x)$ in place of $\|x\|^2$ in the stability analysis; this might enable one to prove stability and convergence for a larger set of SDEs.

Another extension direction for the theory in the infinite time interval is to address SDEs which don't satisfy the contractive property. Numerically, our scheme works well for all the dissipative systems with negative Lyapunov exponent as shown in subsection 5.1, but the numerical analysis needs to be done in the future. For the chaotic systems with positive Lyapunov exponent, a further paper [5] will address this challenge by using change of measure as outlined in subsection 5.2.

REFERENCES

[1] L. Arnold. Lyapunov exponents of nonlinear stochastic systems. In *Nonlinear Stochastic Dynamic Engineering Systems*, pages 181–201. Springer, 1988.

[2] 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.

[3] 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.

[4] 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.

[5] Wei Fang and Michael B Giles. Multilevel Monte Carlo method for ergodic SDEs without contractivity. *Journal of Mathematical Analysis and Applications*, 2019.

[6] 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.

[7] M.B. Giles. Multilevel Monte Carlo path simulation. *Operations Research*, 56(3):607–617, 2008.

[8] M.B. Giles. Multilevel Monte Carlo methods. *Acta Numerica*, 24:259–328, 2015.

[9] M.B. Giles, C. Lester, and J. Whittle. Non-nested adaptive timesteps in multilevel Monte Carlo computations. In R. Cools and D. Nuyens, editors, *Monte Carlo and Quasi-Monte Carlo Methods 2014*. Springer, 2016.

[10] P. Glynn and C. Rhee. Exact estimation for Markov chain equilibrium expectations. *Journal of Applied Probability*, 51:377–389, 2014.

[11] N.R. Hansen. Geometric ergodicity of discrete-time approximations to multivariate diffusions. *Bernoulli*, pages 725–743, 2003.

[12] 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.

[13] N. Hofmann, T. Müller-Gronbach, and K. Ritter. The optimal discretization of stochastic differential equations. *Journal of Complexity*, 17(1):117–153, 2001.

[14] M. Hutzenthaler and A. Jentzen. *Numerical approximations of stochastic differential equations with non-globally Lipschitz continuous coefficients*, volume 236. American Mathematical Society, 2015.

[15] 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.

[16] 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.

[17] Cónall Kelly and Gabriel J Lord. Adaptive time-stepping strategies for nonlinear stochastic systems. *IMA Journal of Numerical Analysis*, 2016.

[18] P.E. Kloeden and E. Platen. *Numerical Solution of Stochastic Differential Equations*. Springer, Berlin, 1992.

[19] H. Lamba. An adaptive timestepping algorithm for stochastic differential equations. *Journal of computational and applied mathematics*, 161(2):417–430, 2003.

[20] H. Lamba, J.C. Mattingly, and A.M. Stuart. An adaptive Euler-Maruyama scheme for SDEs: convergence and stability. *IMA Journal of Numerical Analysis*, 27(3):479–506, 2007.

[21] V. Lemaire. An adaptive scheme for the approximation of dissipative systems. *Stochastic Processes and their Applications*, 117(10):1491–1518, 2007.

[22] X. Mao. *Stochastic Differential Equations and Applications*. Horwood Publishers Ltd., 1997.

[23] X. Mao. The truncated Euler-Maruyama method for stochastic differential equations. *Journal of Computational and Applied Mathematics*, 290:370–384, 2015.

[24] X. Mao. Convergence rates of the truncated Euler-Maruyama method for stochastic differential equations. *Journal of Computational and Applied Mathematics*, 296:362–375, 2016.

[25] 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.

[26] 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.

[27] S. Mauthner. Step size control in the numerical solution of stochastic differential equations. *Journal of Computational and Applied Mathematics*, 100(1):93–109, 1998.

[28] 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.

[29] 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.

[30] G.N. Milstein and M.V. Tretyakov. Computing ergodic limits for Langevin equations. *Physica D: Nonlinear Phenomena*, 229(1):81–95, 2007.

[31] T. Müller-Gronbach. Optimal pointwise approximation of SDEs based on Brownian motion at discrete points. *Annals of Applied Probability*, pages 1605–1642, 2004.

[32] P. Protter, L. Qiu, and J. Martin. Asymptotic error distribution for the Euler scheme with locally Lipschitz coefficients. *arXiv preprint arXiv:1709.04480*, 2017.

[33] G.O. Roberts and J.S. Rosenthal. Optimal scaling of discrete approximations to Langevin diffusions. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 60(1):255–268, 1998.

[34] G.O. Roberts and R.L. Tweedie. Exponential convergence of Langevin distributions and their discrete approximations. *Bernoulli*, pages 341–363, 1996.

[35] S. Sabanis. Euler approximations with varying coefficients: the case of superlinearly growing diffusion coefficients. *The Annals of Applied Probability*, 26(4):2083–2105, 2016.

[36] C. Soize. *The Fokker-Planck equation for stochastic dynamical systems and its explicit steady state solutions*, volume 17. World Scientific, 1994.

[37] Ł. Szpruch and X. Zhāng. V-integrability, asymptotic stability and comparison property of explicit numerical schemes for non-linear SDEs. *Mathematics of Computation*, 87(310):755–783, 2018.

[38] 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.

[39] 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.

MATHEMATICAL INSTITUTE
UNIVERSITY OF OXFORD
OXFORD OX2 6GG
UNITED KINGDOM
E-MAIL: wei.fang@maths.ox.ac.uk
        mike.giles@maths.ox.ac.uk