THE UNIVERSITY *of* EDINBURGH

Edinburgh Research Explorer

Multilevel Monte Carlo for stochastic differential equations with small noise

Link:
Link to publication record in Edinburgh Research Explorer

Document Version:
Publisher's PDF, also known as Version of record

OPEN ACCESS

Multilevel Monte Carlo for stochastic differential equations with small noise

David F. Anderson,*   Desmond J. Higham,†   Yu Sun‡

December 10, 2014

Abstract

We consider the problem of numerically estimating expectations of solutions to stochastic differential equations driven by Brownian motions in the small noise regime. We consider (i) standard Monte Carlo methods combined with numerical discretization algorithms tailored to the small noise setting, and (ii) a multilevel Monte Carlo method combined with a standard Euler-Maruyama implementation. The multilevel method combined with Euler-Maruyama is found to be the most efficient option under the assumptions we make on the underlying model. Further, under a wide range of scalings the multilevel method is found to be optimal in the sense that it has the same asymptotic computational complexity that arises from Monte Carlo with direct sampling from the exact distribution — something that is typically impossible to do. The variance between two coupled paths, as opposed to the L^2 distance, is directly analyzed in order to provide sharp estimates in the multilevel setting.

1 Introduction

We study the problem of numerically estimating expectations of solutions to stochastic differential equations (SDEs) with small noise via Monte Carlo and multilevel Monte Carlo methods. Such models arise in a number of areas including electrical circuit simulation [3], modeling of signal propagation in neurons [4], and biochemistry [2]. Our particular motivation comes from biochemistry and cell biology where the diffusion and linear noise approximations to the usual jump models are SDEs with small noise [2].

Let $(\Omega, \mathcal{F}, \{\mathcal{F}_t\}_{t\geq 0}, P)$ be a filtered probability space satisfying the usual conditions; i.e. the filtration is complete and right-continuous. Let $W(t) = (W_1(t), W_2(t), \ldots, W_m(t))$ be

an m-dimensional standard Wiener processes under $\{\mathcal{F}_t\}_{t\geq 0}$. Let $\varepsilon > 0$ be a small parameter and let D^ε be the solution to the following Ito SDE,

$$D^\varepsilon(t) = D(0) + \int_0^t \mu(D^\varepsilon(s))ds + \varepsilon \int_0^t \sigma(D^\varepsilon(s))dW(s), \tag{1}$$

where $\mu : \mathbb{R}^d \to \mathbb{R}^d$ and $\sigma : \mathbb{R}^d \to \mathbb{R}^{d\times m}$ are continuous functions.

Let $f : \mathbb{R}^d \to \mathbb{R}$ have bounded first and second partial derivatives and let $T > 0$ be a fixed positive number. We are interested in the problem of numerically estimating $\mathbb{E}[f(D^\varepsilon(T))]$ to an accuracy of $\delta > 0$ in the sense of confidence intervals. In particular, we study the computational complexity required to solve this problem utilizing both (i) standard Monte Carlo methods combined with discretization methods tailored to the small noise setting [10, 9], and (ii) multilevel Monte Carlo methods combined with Euler-Maruyama [5]. The L^2 bounds on the difference between exact and approximate processes that are already in the literature [9] do not provide sharp estimates for the variance between two coupled paths; an analogous issue was previously addressed in the jump process setting [1]. Our main effort is therefore directed at analysing the variance between two coupled paths in the small noise setting.

We note that if realizations of $f(D^\varepsilon(T))$ could be generated with a single numerical calculation, and if $\mathsf{Var}(f(D^\varepsilon(T))) = O(\varepsilon^2)$, then the computational complexity of solving the problem would be $O(\varepsilon^2\delta^{-2})$. We show below that the multilevel Monte Carlo method combined with a standard implementation of Euler-Maruyama solves the problem with this same optimal computational complexity.

We make the following regularity assumption throughout the manuscript.

Running assumption. *We suppose there are constants $a, b > 0$ such that for all $x, y \in \mathbb{R}^d$ the following inequalities hold:*

$$|\nabla\mu(x)|^2 \vee |\nabla^2\mu(x)|^2 \leq a,$$

and

$$|\mu(x) - \mu(y)|^2 \leq a|x - y|^2, \quad |\sigma(x) - \sigma(y)|^2 \leq b|x - y|^2,$$

and

$$|\mu(x)|^2 \leq a(1 + |x|^2), \quad |\sigma(x)|^2 \leq b(1 + |x|^2).$$

Under the above assumptions, the SDE (1) is known to have a unique strong solution (see, for example, Theorem 3.1 on page 51 in [8]).

1.1   Euler-Maruyama and a statement of main mathematical result

We provide a continuous version of the Euler-Maruyama discretization method. Let $h > 0$ and let D_h^ε be the solution to

$$D_h^\varepsilon(t) = D(0) + \int_0^t \mu(D_h^\varepsilon(\eta(s))) \, ds + \varepsilon \int_0^t \sigma(D_h^\varepsilon(\eta_h(s))) \, dW(s), \tag{2}$$

where $\eta(s) \stackrel{\text{def}}{=} \lfloor s/h \rfloor h$. It is straightforward to see that the solution to (2) restricted to the set of times $\{0, h, 2h, \dots\}$ has the same distribution as the discrete time process generated by the usual Euler-Maruyama method [7].

In order to understand the computational complexity of the multilevel scheme, we need sharp estimates for the variance between two coupled paths. The following provides such an estimate and is the main theorem provided in this manuscript. The result bounds the variance between two coupled process; both are generated via (2), though they have different time discretization parameters. See the beginning of section 3 for more details related to the coupling.

Theorem 1. *Suppose the functions μ and σ satisfy our running assumptions and that $T > 0$. Suppose further that $D_{h_\ell}^\varepsilon(t)$ and $D_{h_{\ell-1}}^\varepsilon(t)$ satisfy (2) with time discretization parameters $h_\ell = T \cdot M^{-\ell}$ and $h_{\ell-1} = T \cdot M^{-(\ell-1)}$, respectively, where M is a positive integer, and that these two processes are constructed with the same realization of Brownian motions. Assume that $f : \mathbb{R}^d \to \mathbb{R}$ has continuous second derivative and there exists a constant C_L such that*

$$\left\| \frac{\partial f}{\partial x_i} \right\|_\infty \leq C_L \quad and \quad \left\| \frac{\partial^2 f}{\partial x_i \partial x_j} \right\|_\infty \leq C_L \quad for \ any \ i, j = 1, 2, ..., d.$$

Then,

$$\max_{0 \leq n \leq M^{\ell-1}} \mathsf{Var}(f(D_{h_\ell}^\varepsilon(t_n)) - f(D_{h_{\ell-1}}^\varepsilon(t_n))) \leq \bar{C}_1 h_{\ell-1}^2 \varepsilon^2 + \bar{C}_2 h_{\ell-1} \varepsilon^4, \tag{3}$$

where $t_n = n \cdot h_{\ell-1}$, and $\bar{C}_1$ and $\bar{C}_2$ are positive constants only depending on $a, b, d, m, T, D(0)$ and C_L.

In the context of analysing the classical mean-square error, it was shown by Milstein and Tretyakov in [9] that under the same assumptions as in Theorem 1,

$$\mathbb{E}[|f(D^\varepsilon(T)) - f(D_h^\varepsilon(T))|^2] = O(h^2 + h\varepsilon^4), \tag{4}$$

where D^ε is the solution to (1). The bound (4) implies that for some $C_1, C_2 > 0$ we have $\max_{0 \leq n \leq M^{\ell-1}} \mathsf{Var}(f(D_{h_\ell}^\varepsilon(t_n)) - f(D_{h_{\ell-1}}^\varepsilon(t_n))) \leq C_1 h_{\ell-1}^2 + C_2 h_{\ell-1} \varepsilon^4$, where, again, $t_n = n \cdot h_{\ell-1}$. Theorem 1 sharpens this bound considerably, showing that the overall variance scales favourably with ε, even though the Euler–Maruyama method has not been customized to exploit the small noise property.

2 Complexity Analysis

2.1 Standard Monte Carlo methods

As a basis for comparison, we first analyze the complexity of standard Monte Carlo with a general discretization method.

Suppose D_h^ε is generated by a numerical scheme (not necessarily (2)) for which the bias of the discretization method satisfies

$$|\mathbb{E}[f(D_h^\varepsilon(T))] - \mathbb{E}[f(D^\varepsilon(T))]| = O(h^p + \varepsilon^r h^q). \tag{5}$$

3

where $q < p$. In order to ensure that the bias (5) is smaller than δ, we require that

$$h = O(\max(\delta^{1/p}, \delta^{1/q}\varepsilon^{-r/q})).$$

Under our running assumptions and generalizing the analysis in Lemma 2 below, which applies to Euler-Maruyama, implies

$$\mathsf{Var}(f(D_h^\varepsilon(T))) = \mathsf{Var}(f(D_h^\varepsilon(T)) - f(z_h(T))) \le C\mathbb{E}\left[\sup_{s \le T}|D_h^\varepsilon(s) - z_h(s)|^2\right] = O(\varepsilon^2),$$

where z_h is the Euler solution to the associated deterministic model obtained when ε is set to 0 in (1), see (13). Thus, the standard Monte Carlo estimator

$$\mathbb{E}[f(D^\varepsilon(T))] \approx \mathbb{E}[f(D_h^\varepsilon(T))] \approx \frac{1}{n}\sum_{i=1}^n f(D_{h,[i]}^\varepsilon(T)),$$

where $D_{h,[i]}^\varepsilon$ is the ith independent realization of the process, has a variance that is $O(n^{-1}\varepsilon^2)$. Because we require an overall estimator variance of $O(\delta^2)$, we require that $n = O(\varepsilon^2\delta^{-2})$. Assuming that the cost of generating a single path of the scheme scales like h^{-1}, we obtain an upper bound on the overall computational complexity of order

$$O(\varepsilon^2\delta^{-2}h^{-1}) = O\left(\frac{\delta^{-2}\varepsilon^2}{\max(\delta^{1/p}, \delta^{1/q}\varepsilon^{-r/q})}\right).$$

For example, for the Euler-Maruyama scheme (2), we have that $p = q = 1, r = 0$ yielding a bias of $O(h)$ in (5). In this case we select $h = O(\delta)$, and find a computational complexity of $O(\varepsilon^2\delta^{-3})$.

2.2 Euler-based multilevel Monte Carlo

Here we specify and analyze an Euler-Maruyama based multilevel Monte Carlo method for the diffusion approximation. We follow the original framework of Giles [5].

For a fixed positive integer M we let $h_\ell = T \cdot M^{-\ell}$ for $\ell \in \{0, \ldots, L\}$. Reasonable choices for M include $M \in \{2, 3, 4, \ldots, 7\}$, and L is determined below. For each $\ell \in \{0, 1, \ldots, L\}$, let $D_{h_\ell}^\varepsilon$ denote the approximate process generated by (2) with a step size of h_ℓ. Note that

$$\mathbb{E}[f(D^\varepsilon(T))] \approx \mathbb{E}[f(D_{h_L}^\varepsilon(T))] = \mathbb{E}[f(D_{h_0}^\varepsilon(T))] + \sum_{\ell=1}^L \mathbb{E}[f(D_{h_\ell}^\varepsilon(T)) - f(D_{h_{\ell-1}}^\varepsilon(T))],$$

with the quality of the approximation only depending upon h_L. As mentioned in [9], the Euler discretization has a weak order of one in the present setting for a large class of functionals f. Hence, we set $h_L = \delta$ in order for the bias to be $O(\delta)$. This choice yields $L = O(|\log(\delta)|)$. We now let

$$\widehat{Q}_0^\varepsilon \overset{\text{def}}{=} \frac{1}{n_0}\sum_{i=1}^{n_0} f(D_{h_0,[i]}^\varepsilon(T)), \quad \text{and} \quad \widehat{Q}_\ell^\varepsilon \overset{\text{def}}{=} \frac{1}{n_\ell}\sum_{i=1}^{n_\ell}(f(D_{h_\ell,[i]}^\varepsilon(T)) - f(D_{h_{\ell-1},[i]}^\varepsilon(T))),$$

for $\ell = 1, \ldots, L$, where n_0 and the different n_ℓ have yet to be determined. Our estimator is then

$$\widehat{Q}^\varepsilon \overset{\text{def}}{=} \widehat{Q}_0^\varepsilon + \sum_{\ell=1}^{L} \widehat{Q}_\ell^\varepsilon.$$

Set

$$\delta_{\varepsilon,\ell} = \mathsf{Var}(f(D_{h_\ell}^\varepsilon(T)) - f(D_{h_{\ell-1}}^\varepsilon(T))).$$

By Theorem 1, we have $\delta_{\varepsilon,\ell} = O(h_\ell^2 \varepsilon^2 + h_\ell \varepsilon^4)$ under a wide array of circumstances. Also note that $\delta_{\varepsilon,0} = \mathsf{Var}(f(D_0^\varepsilon)) = O(\varepsilon^2)$.

For $\ell \in \{1, \ldots, L\}$, let C_ℓ be the computational complexity required to generate a single pair of coupled trajectories at level ℓ. Let C_0 be the computational complexity required to generate a single trajectory at the coarsest level. To be concrete, we set C_ℓ to be the number of random variables required to generate the requisite path. To determine n_ℓ, we solve the following optimization problem, which ensures a total variance of $\widehat{Q}^\varepsilon$ no greater than δ^2:

$$\underset{n_\ell}{\text{minimize}} \quad \sum_{\ell=0}^{L} n_\ell C_\ell,$$

$$\text{subject to} \quad \sum_{\ell=0}^{L} \frac{\delta_{\varepsilon,\ell}}{n_\ell} = \delta^2.$$

We use Lagrange multipliers. Since $C_\ell = h_\ell^{-1}$, the optimization problem above is solved at solutions to

$$\nabla_{n_0, \ldots, n_L, \lambda} \left(\sum_{\ell=0}^{L} n_\ell h_\ell^{-1} + \lambda \left(\sum_{\ell=0}^{L} \frac{\delta_{\varepsilon,\ell}}{n_\ell} - \delta^2 \right) \right) = 0.$$

By taking a derivative with respect to n_ℓ we obtain,

$$n_\ell = \sqrt{\lambda \delta_{\varepsilon,\ell} h_\ell}, \qquad \text{for } \ell \in \{0, 1, 2, \ldots, L\} \tag{6}$$

and some $\lambda \geq 0$. Thus,

$$\sum_{\ell=0}^{L} \sqrt{\frac{\delta_{\varepsilon,\ell}}{h_\ell}} = \sqrt{\lambda} \delta^2$$

and, by Theorem 1,

$$\sqrt{\lambda} = \delta^{-2} \sum_{\ell=0}^{L} \sqrt{\frac{\delta_{\varepsilon,\ell}}{h_\ell}} \leq C \delta^{-2} \sum_{\ell=0}^{L} \sqrt{h_\ell \varepsilon^2 + \varepsilon^4} \leq C(\delta^{-2} \varepsilon + \delta^{-2} \varepsilon^2 L). \tag{7}$$

Recall that $L = O(\log \frac{1}{\delta})$. Hence, if $\delta \geq e^{-\frac{1}{\varepsilon}}$, which implies $\varepsilon^2 L \leq \varepsilon$, then

$$\lambda = O(\delta^{-4} \varepsilon^2).$$

In this case the overall computational complexity is

$$\sum_{\ell=0}^{L} n_\ell h_\ell^{-1} = \sum_{\ell=0}^{L} \sqrt{\lambda} \sqrt{\frac{\delta_{\varepsilon,\ell}}{h_\ell}} = \delta^2 \lambda = O(\delta^{-2} \varepsilon^2). \tag{8}$$

If $\delta < e^{-\frac{1}{\varepsilon}}$, in which case $\varepsilon^2 L > \varepsilon$, then

$$\lambda = O(\delta^{-4}\varepsilon^4 L^2),$$

and the overall computational complexity becomes

$$O(\delta^{-2}\varepsilon^4|\log(\delta)|^2). \tag{9}$$

2.3 Comparison

The regime $\delta < e^{-\frac{1}{\varepsilon}}$ in which we derived (9) represents an extremely severe accuracy request that is unlikely to be relevant in practice. Hence, we focus on the complexity bound $O(\delta^{-2}\varepsilon^2)$ in (8). This bound compares favourably with the bound $O(\varepsilon^2\delta^{-3})$ that we derived in subsection 2.2 for standard Monte Carlo with Euler–Maruyama, allowing us to carry through a conclusion that applies to general SDEs [5]: multilevel Monte Carlo can improve on the complexity of standard Monte Carlo by a factor δ^{-1}, where δ is the required accuracy. Furthermore, following the discussion in subsection 1, we note that this multilevel Euler computational complexity is optimal in that we cannot do better—asymptotically in the parameters δ or ε—even if we could generate independent realizations of $f(D^\varepsilon(T))$ exactly in a single step.

3 Proof of Theorem 1

Throughout this section, we assume the conditions of Theorem 1 are met with positive integer M fixed.

The coupling of the two approximate processes, $D^\varepsilon_{h_\ell}(t)$ and $D^\varepsilon_{h_{\ell-1}}(t)$, takes the form

$$D^\varepsilon_{h_\ell}(t) = D(0) + \int_0^t \mu(D^\varepsilon_{h_\ell}(\eta(s)))\,ds + \varepsilon\int_0^t \sigma(D^\varepsilon_{h_\ell}(\eta_{h_\ell}(s)))\,dW(s),$$

$$D^\varepsilon_{h_{\ell-1}}(t) = D(0) + \int_0^t \mu(D^\varepsilon_{h_{\ell-1}}(\eta(s)))\,ds + \varepsilon\int_0^t \sigma(D^\varepsilon_{h_{\ell-1}}(\eta_{h_{\ell-1}}(s)))\,dW(s).$$

For $n \in \{0, 1, \ldots, M^{\ell-1}\}$ and $k \in \{0, \ldots, M\}$ let

$$t_n = nh_{\ell-1}, \quad \text{and} \quad t_n^k = nh_{\ell-1} + kh_\ell.$$

Note that for each n we have
$$t_n^0 = t_n, \quad t_n^M = t_{n+1}.$$

We use the following discretization scheme to simulate the coupling above. First, for each $n \in \{0, 1, \ldots, M^{\ell-1}\}$ and $k \in \{0, \ldots, M-1\}$, let

$$D^\varepsilon_{h_\ell}(t_n^{k+1}) = D^\varepsilon_{h_\ell}(t_n^k) + \mu(D^\varepsilon_{h_\ell}(t_n^k))h_\ell + \varepsilon\sqrt{h_\ell}\sigma(D^\varepsilon_{h_\ell}(t_n^k))W_n^k, \tag{10}$$

6

where the random vector $W_n^k \in \mathbb{R}^m$ has independent components (from each other and all previous random variables), and each component is distributed as $N(0,1)$. Note that (10) implies

$$D_{h_\ell}^\varepsilon(t_{n+1}) = D_{h_\ell}^\varepsilon(t_n) + \sum_{k=0}^{M-1} \mu(D_{h_\ell}^\varepsilon(t_n^k))h_\ell + \varepsilon\sqrt{h_\ell}\sum_{k=0}^{M-1}\sigma(D_{h_\ell}^\varepsilon(t_n^k))W_n^k.$$

To simulate $D_{h_{\ell-1}}^\varepsilon$, we then use

$$D_{h_{\ell-1}}^\varepsilon(t_{n+1}) = D_{h_{\ell-1}}^\varepsilon(t_n) + \mu(D_{h_{\ell-1}}^\varepsilon(t_n))h_{\ell-1} + \varepsilon\sqrt{h_{\ell-1}}\sigma(D_{h_{\ell-1}}^\varepsilon(t_n))\sum_{k=0}^{M-1}W_n^k.$$

We begin with a series of necessary lemmas.

Lemma 1. *For any $T > 0$ we have*

$$\mathbb{E}\left[\sup_{0\leq s\leq T}|D_{h_\ell}^\varepsilon(s)|^4\right] \leq C.$$

for some $C = C(a,b,T,D(0))$.

Proof. For any $t > 0$,

$$|D_{h_\ell}^\varepsilon(t)|^4 \leq 27|D(0)|^4 + 27\left|\int_0^t \mu(D_{h_\ell}^\varepsilon(\eta(s)))ds\right|^4 + 27\varepsilon^4\left|\int_0^t \sigma(D_{h_\ell}^\varepsilon(\eta(s)))dW(s)\right|^4.$$

Thus,

$$\sup_{0\leq s\leq t}|D_{h_\ell}^\varepsilon(s)|^4 \leq 27|D(0)|^4 + 27t^3\int_0^t \sup_{0\leq r\leq s}|\mu(D_{h_\ell}^\varepsilon(\eta(r)))|^4 ds$$
$$+ 27\varepsilon^4 \sup_{0\leq s\leq t}\left|\int_0^s \sigma(D_{h_\ell}^\varepsilon(\eta(r)))dW(r)\right|^4, \tag{11}$$

since the right-hand-side is monotonically increasing in t. Applying the Burkholder-Davis-Gundy inequality [8] to the term (11) and taking expectations we get

$$\mathbb{E}\left[\sup_{0\leq s\leq t}|D_{h_\ell}^\varepsilon(s)|^4\right] \leq 27|D(0)|^4 + 27t^3\int_0^t \mathbb{E}\left[\sup_{0\leq r\leq s}|\mu(D_{h_\ell}^\varepsilon(\eta(r)))|^4\right]ds$$
$$+ K(T)\varepsilon^4\int_0^t \mathbb{E}[|\sigma(D_{h_\ell}^\varepsilon(\eta(s)))|^4]ds, \tag{12}$$

where $K(T)$ is a generic constant only depending on T. Using (12) with $t = nh_\ell$ and $s = mh_\ell$,

where n and m are nonnegative integers for which $mh_\ell \leq nh_\ell \leq t \leq T$, we get

$$\mathbb{E}\left[\sup_{m\leq n} |D_{h_\ell}^\varepsilon(mh_\ell)|^4\right] \leq 27|D(0)|^4 + 27t^3 \sum_{i=0}^{n-1} \mathbb{E}\left[\sup_{m\leq i} |\mu(D_{h_\ell}^\varepsilon(mh_\ell))|^4\right] h_\ell$$

$$+ K(T)\varepsilon^4 \sum_{i=0}^{n-1} \mathbb{E}\left[|\sigma(D_{h_\ell}^\varepsilon(ih_\ell))|^4\right] h_\ell$$

$$\leq 27|D(0)|^4 + 54a^2T^4 + K(T)b^2\varepsilon^4$$

$$+ (54a^2T^3 + K(T)b^2\varepsilon^4) \sum_{i=0}^{n-1} \mathbb{E}\left[\sup_{m\leq i} |D_{h_\ell}^\varepsilon(mh_\ell)|^2\right] h_\ell,$$

where in the final inequality we applied the growth conditions for both μ and σ found in the running assumption. We then use the discrete version of Gronwall's Lemma to obtain

$$\mathbb{E}\left[\sup_{m\leq n} |D_{h_\ell}^\varepsilon(mh)|^4\right] \leq C_1(a,b,T,D(0)).$$

Now we return to (12) and, after applying the growth conditions pertaining to both μ and σ in our running assumption, conclude

$$\mathbb{E}\left[\sup_{0\leq s\leq T} |D_{h_\ell}^\varepsilon(s)|^4\right] \leq C(a,b,T,D(0)),$$

for some new constant C. $\qquad\square$

Let z be the deterministic solution to

$$z_h(t) = D(0) + \int_0^t \mu(z_h(\eta(s)))ds, \tag{13}$$

which is an Euler approximation to the ODE obtained from (1) when ε is set to zero.

Lemma 2. *For any $T > 0$ we have*

$$\mathbb{E}\left[\sup_{0\leq s\leq T} |D_{h_\ell}^\varepsilon(s) - z_{h_\ell}(s)|^2\right] \leq C\varepsilon^2,$$

for some $C = C(a,b,T,D(0))$.

Proof. For $t \leq T$, we have

$$|D_{h_\ell}^\varepsilon(t) - z_{h_\ell}(t)|^2 \leq 2T\int_0^t |\mu(D_{h_\ell}^\varepsilon(\eta(s))) - \mu(z_{h_\ell}(\eta(s))|^2 ds + 2\varepsilon^2\left|\int_0^t \sigma(D_{h_\ell}^\varepsilon(\eta(s)))dW(s)\right|^2.$$

As a result of the Burkholder-Davis-Gundy inequality and our running assumptions,

$$\mathbb{E}\left[\sup_{0\leq s\leq t} \left|D_{h_\ell}^\varepsilon(s) - z_{h_\ell}(s)\right|^2\right]$$

$$\leq 2aT\int_0^t \mathbb{E}\left[\sup_{0\leq s\leq r} |D_{h_\ell}^\varepsilon(\eta(s)) - z_{h_\ell}(\eta(s))|^2\right] dr + 8\varepsilon^2 \int_0^t \mathbb{E}[|\sigma(D_{h_\ell}^\varepsilon(\eta(s)))|^2]ds \tag{14}$$

$$\leq 8bT\mathbb{E}\left[\sup_{0\leq s\leq t} (1 + |D_{h_\ell}^\varepsilon(s)|^2)\right]\varepsilon^2 + 2aT\int_0^t \mathbb{E}\left[\sup_{0\leq s\leq r} |D_{h_\ell}^\varepsilon(\eta(s)) - z_{h_\ell}(\eta(s))|^2\right] dr.$$

8

Specializing the above to $t = nh_\ell$ and $s = mh_\ell$, where n and m are nonnegative integers for which $mh_\ell \leq nh_\ell \leq t \leq T$, we get

$$\mathbb{E}\left[\sup_{m \leq n} |D_{h_\ell}^\varepsilon(mh_\ell) - z_{h_\ell}(mh_\ell)|^2\right]$$

$$\leq 8bT\mathbb{E}\left[\sup_{0 \leq s \leq t}(1 + |D_{h_\ell}^\varepsilon(s)|^2)\right]\varepsilon^2 + 2aT\sum_{i=0}^{n-1} h_\ell \cdot \mathbb{E}\left[\sup_{m \leq i}|D_{h_\ell}^\varepsilon(mh_\ell) - z_{h_\ell}(mh_\ell)|^2\right]$$

$$\leq 8bT(1 + K)\varepsilon^2 + 2aT\sum_{i=0}^{n-1} h_\ell \cdot \mathbb{E}\left[\sup_{m \leq i}|D_{h_\ell}^\varepsilon(mh_\ell) - z_{h_\ell}(mh_\ell)|^2\right],$$

for some $K = K(a, b, T, D(0))$, where the first inequality follows from (14) and the second utilizes Lemma 1.

By the discrete version of Gronwall's inequality we see

$$\mathbb{E}\left[\sup_{m \leq n}|D_{h_\ell}^\varepsilon(mh_\ell) - z_{h_\ell}(mh_\ell)|^2\right] \leq (8bT(1 + K))e^{2aT^2}\varepsilon^2.$$

Since n satisfying $nh_\ell \leq T$ was arbitrary, we return to (14) to conclude that for any $0 \leq t \leq T$

$$\mathbb{E}\left[\sup_{s \leq t}|D_{h_\ell}^\varepsilon(s) - z_{h_\ell}(s)|^2\right] \leq C(a, b, T, D(0))\varepsilon^2. \qquad \square$$

Lemma 3.
$$\max_{\substack{0 \leq n \leq M^{\ell-1} \\ 1 \leq k \leq M}} |\mathbb{E}[D_{h_\ell}^\varepsilon(t_n^k) - D_{h_\ell}^\varepsilon(t_n)]| \leq CMh_\ell,$$

where C is a positive constant that only depends on $a, b, T, m, D(0)$.

Proof. Iterating (10) yields

$$\left|\mathbb{E}\left[D_{h_\ell}^\varepsilon(t_n^k) - D_{h_\ell}^\varepsilon(t_n)\right]\right| \leq \left|\mathbb{E}\left[\sum_{i=0}^{k-1}\mu(D_{h_\ell}^\varepsilon(t_n^i))h_\ell\right]\right| + \left|\mathbb{E}\left[\varepsilon\sqrt{h_\ell}\sum_{i=0}^{k-1}\sigma(D_{h_\ell}^\varepsilon(t_n^i))W_n^i\right]\right|$$

$$\leq \sum_{i=0}^{k-1}\mathbb{E}\left[|\mu(D_{h_\ell}^\varepsilon(t_n^i))h_\ell|\right]$$

$$\leq h_\ell\sqrt{a}\sum_{i=0}^{k-1}(1 + \mathbb{E}[|D_{h_\ell}^\varepsilon(t_n^i)|]),$$

where the first inequality is simply the triangle inequality, the second follows from the triangle inequality combined with the observation that the expectations of the diffusion terms are zero, and the third inequality follows from our running assumptions. The proof is completed by using Lemma 1 and recalling that $k \leq M$. $\qquad \square$

Lemma 4.
$$\max_{\substack{0 \leq n \leq M^{\ell-1} \\ 1 \leq k \leq M}} \mathbb{E}[|D_{h_\ell}^\varepsilon(t_n^k) - D_{h_\ell}^\varepsilon(t_n)|^4|] \leq C_1 M^4 h_\ell^4 + C_2\varepsilon^4 M^2 h_\ell^2,$$

where C_1 and C_2 are positive constants that only depend on $a, b, T, m, D(0)$.

Proof. Iterating (10) yields

$$D^\varepsilon_{h_\ell}(t^k_n) - D^\varepsilon_{h_\ell}(t_n) = \sum_{i=0}^{k-1} \mu(D^\varepsilon_{h_\ell}(t^i_n))h_\ell + \varepsilon\sqrt{h_\ell} \sum_{i=0}^{k-1} \sigma(D^\varepsilon_{h_\ell}(t^i_n))W^i_n.$$

Denoting $\|X\|_{L^4(\Omega,\mathbb{R}^d)} = (\mathbb{E}[|X|^4])^{1/4}$ and σ^j to be the jth column of σ, we use the inequality $(a+b)^4 \leq 8a^4 + 8b^4$ to conclude

$$\mathbb{E}\big[|D^\varepsilon_{h_\ell}(t^k_n) - D^\varepsilon_{h_\ell}(t_n)|^4\big] \leq 8M^3 \sum_{i=0}^{k-1} \mathbb{E}\big[|\mu(D^\varepsilon_{h_\ell}(t^i_n))h_\ell|^4\big] + 8\varepsilon^4 h_\ell^2 \mathbb{E}\left[\left|\sum_{i=0}^{k-1} \sigma(D^\varepsilon_{h_\ell}(t^i_n))W^i_n\right|^4\right]$$

$$\leq C(a,b,T,D(0))M^4 h_\ell^4 + 2048\varepsilon^4 h_\ell^2 \left(\sum_{i=0}^{M-1} \sum_{j=1}^{m} \|\sigma^j(D^\varepsilon_{h_\ell}(t^i_n))\|^2_{L^4(\Omega,\mathbb{R}^d)}\right)^2$$

$$\leq C(a,b,T,D(0))^4 M^4 h_\ell^4 + 2048b\varepsilon^4 M^2 h_\ell^2 m^2 (2 + 2 \max_{0 \leq i \leq M-1} \|D^\varepsilon_{h_\ell}(t^i_n)\|^4_{L^4(\Omega,\mathbb{R}^d)})$$

$$\leq C(a,b,T,D(0))^4 M^4 h_\ell^4 + C_2 \varepsilon^4 M^2 h_\ell^2,$$

where the second inequality follows from Lemma 1 and Lemma 3.8 in [6], the last inequality follows from Lemma 1, and C_1 and C_2 are constants only depending on $a, b, T, m, D(0)$. □

The following is a Taylor expansion of the drift coefficient.

Lemma 5. *Let $\mu_i(x)$ be the ith component of $\mu(x)$, then*

$$\mu_i(D^\varepsilon_{h_\ell}(t^k_n)) - \mu_i(D^\varepsilon_{h_\ell}(t_n)) = A_k + B_k + E_k, \tag{15}$$

where

$$A_k := \int_0^1 \big[\nabla\mu_i(D^\varepsilon_{h_\ell}(t_n) + s(D^\varepsilon_{h_\ell}(t^k_n) - D^\varepsilon_{h_\ell}(t_n)))\big]\, ds \cdot \left(h_\ell \sum_{j=0}^{k-1} \mu(D^\varepsilon_{h_\ell}(t^j_n))\right),$$

$$B_k := \nabla\mu_i(D^\varepsilon_{h_\ell}(t_n)) \cdot \left(\varepsilon\sqrt{h_\ell} \sum_{j=0}^{k-1} \sigma(D^\varepsilon_{h_\ell}(t^j_n))W^j_n\right), \tag{16}$$

and

$$E_k := \left(\int_0^1 \int_0^s \big[\nabla^2\mu_i(D^\varepsilon_{h_\ell}(t_n) + r(D^\varepsilon_{h_\ell}(t^k_n) - D^\varepsilon_{h_\ell}(t_n)))(D^\varepsilon_{h_\ell}(t^k_n) - D^\varepsilon_{h_\ell}(t_n))\big]\, dr ds\right)$$

$$\cdot \left(\varepsilon\sqrt{h_\ell} \sum_{j=0}^{k-1} \sigma(D^\varepsilon_{h_\ell}(t^j_n))W^j_n\right).$$

10

Proof. Using Taylor's expansion (see Lemma 12 in the appendix) we see

$$\mu_i(D^\varepsilon_{h_\ell}(t^k_n)) - \mu_i(D^\varepsilon_{h_\ell}(t_n))$$

$$= \int_0^1 \left[\nabla\mu_i(D^\varepsilon_{h_\ell}(t_n) + s(D^\varepsilon_{h_\ell}(t^k_n) - D^\varepsilon_{h_\ell}(t_n)))\right] ds \cdot (D^\varepsilon_{h_\ell}(t^k_n) - D^\varepsilon_{h_\ell}(t_n))$$

$$= \int_0^1 \left[\nabla\mu_i(D^\varepsilon_{h_\ell}(t_n) + s(D^\varepsilon_{h_\ell}(t^k_n) - D^\varepsilon_{h_\ell}(t_n)))\right] ds \cdot \left(h_\ell \sum_{j=0}^{k-1} \mu(D^\varepsilon_{h_\ell}(t^j_n))\right)$$

$$+ \int_0^1 \left[\nabla\mu_i(D^\varepsilon_{h_\ell}(t_n) + s(D^\varepsilon_{h_\ell}(t^k_n) - D^\varepsilon_{h_\ell}(t_n)))\right] ds \cdot \left(\varepsilon\sqrt{h_\ell} \sum_{j=0}^{k-1} \sigma(D^\varepsilon_{h_\ell}(t^j_n))W^j_n\right).$$

Applying a multidimensional version of Lemma 12,

$$\int_0^1 [\nabla\mu_i(D^\varepsilon_{h_\ell}(t_n) + s(D^\varepsilon_{h_\ell}(t^k_n) - D^\varepsilon_{h_\ell}(t_n)))]ds \cdot \left(\varepsilon\sqrt{h_\ell} \sum_{j=0}^{k-1} \sigma(D^\varepsilon_{h_\ell}(t^j_n))W^j_n\right)$$

$$= \nabla\mu_i(D^\varepsilon_{h_\ell}(t_n)) \cdot \left(\varepsilon\sqrt{h_\ell} \sum_{j=0}^{k-1} \sigma(D^\varepsilon_{h_\ell}(t^j_n))W^j_n\right)$$

$$+ \left(\int_0^1 \int_0^s \left[H(\mu_i)(D^\varepsilon_{h_\ell}(t_n) + r(D^\varepsilon_{h_\ell}(t^k_n) - D^\varepsilon_{h_\ell}(t_n)))(D^\varepsilon_{h_\ell}(t^k_n) - D^\varepsilon_{h_\ell}(t_n))\right] drds\right)$$

$$\cdot \left(\varepsilon\sqrt{h_\ell} \sum_{j=0}^{k-1} \sigma(D^\varepsilon_{h_\ell}(t^j_n))W^j_n\right).$$

Therefore,

$$\mu_i(D^\varepsilon_{h_\ell}(t^k_n)) - \mu_i(D^\varepsilon_{h_\ell}(t_n))$$

$$= \int_0^1 [\nabla\mu_i(D^\varepsilon_{h_\ell}(t_n) + s(D^\varepsilon_{h_\ell}(t^k_n) - D^\varepsilon_{h_\ell}(t_n)))]ds \cdot \left(h_\ell \sum_{j=0}^{k-1} \mu(D^\varepsilon_{h_\ell}(t^j_n))\right)$$

$$+ \nabla\mu_i(D^\varepsilon_{h_\ell}(t_n)) \cdot \left(\varepsilon\sqrt{h_\ell} \sum_{j=0}^{k-1} \sigma(D^\varepsilon_{h_\ell}(t^j_n))W^j_n\right)$$

$$+ \left(\int_0^1 \int_0^s [\nabla^2\mu_i(D^\varepsilon_{h_\ell}(t_n) + r(D^\varepsilon_{h_\ell}(t^k_n) - D^\varepsilon_{h_\ell}(t_n)))(D^\varepsilon_{h_\ell}(t^k_n) - D^\varepsilon_{h_\ell}(t_n))]drds\right)$$

$$\cdot \left(\varepsilon\sqrt{h_\ell} \sum_{j=0}^{k-1} \sigma(D^\varepsilon_{h_\ell}(t^j_n))W^j_n\right)$$

$$= A_k + B_k + E_k. \qquad \square$$

The following result is similar to the L^2 bound found in [9] in the case where the numerical discretization method is Euler–Maruyama.

Lemma 6.

$$\max_{0 \le n \le M^{\ell-1}} \mathbb{E}[|D^\varepsilon_{h_\ell}(t_n) - D^\varepsilon_{h_{\ell-1}}(t_n)|^2] \le d_1 h^2_{\ell-1} + d_2\varepsilon^4 h_{\ell-1},$$

where d_1 and d_2 are positive constants that depend on $a, b, T, m, D(0)$.

Proof. For $n \leq M^{\ell-1} - 1$ we have

$$D_{h_\ell}^\varepsilon(t_{n+1}) - D_{h_{\ell-1}}^\varepsilon(t_{n+1}) = D_{h_\ell}^\varepsilon(t_n) - D_{h_{\ell-1}}^\varepsilon(t_n) + h_\ell \sum_{k=0}^{M-1} (\mu(D_{h_\ell}^\varepsilon(t_n^k)) - \mu(D_{h_{\ell-1}}^\varepsilon(t_n)))$$

$$+ \varepsilon\sqrt{h_\ell} \sum_{k=0}^{M-1} (\sigma(D_{h_\ell}^\varepsilon(t_n^k)) - \sigma(D_{h_{\ell-1}}^\varepsilon(t_n))) W_n^k$$

$$= D_{h_\ell}^\varepsilon(t_n) - D_{h_{\ell-1}}^\varepsilon(t_n) + h_\ell \sum_{k=0}^{M-1} (\mu(D_{h_\ell}^\varepsilon(t_n^k)) - \mu(D_{h_\ell}^\varepsilon(t_n)))$$

$$+ h_\ell \sum_{k=0}^{M-1} (\mu(D_{h_\ell}^\varepsilon(t_n)) - \mu(D_{h_{\ell-1}}^\varepsilon(t_n)))$$

$$+ \varepsilon\sqrt{h_\ell} \sum_{k=0}^{M-1} (\sigma(D_{h_\ell}^\varepsilon(t_n^k)) - \sigma(D_{h_\ell}^\varepsilon(t_n))) W_n^k$$

$$+ \varepsilon\sqrt{h_\ell} \sum_{k=0}^{M-1} (\sigma(D_{h_\ell}^\varepsilon(t_n)) - \sigma(D_{h_{\ell-1}}^\varepsilon(t_n))) W_n^k,$$

where the final equality simply comes from adding and subtracting some terms. After some manipulation the above implies

$$|D_{h_\ell}^\varepsilon(t_{n+1}) - D_{h_{\ell-1}}^\varepsilon(t_{n+1})|^2 \leq |D_{h_\ell}^\varepsilon(t_n) - D_{h_{\ell-1}}^\varepsilon(t_n)|^2 + 4h_\ell^2 | \sum_{k=0}^{M-1} (\mu(D_{h_\ell}^\varepsilon(t_n^k)) - \mu(D_{h_\ell}^\varepsilon(t_n)))|^2$$

$$+ 4h_\ell^2 | \sum_{k=0}^{M-1} (\mu(D_{h_\ell}^\varepsilon(t_n)) - \mu(D_{h_{\ell-1}}^\varepsilon(t_n)))|^2$$

$$+ 4| \varepsilon\sqrt{h_\ell} \sum_{k=0}^{M-1} (\sigma(D_{h_\ell}^\varepsilon(t_n^k)) - \sigma(D_{h_\ell}^\varepsilon(t_n))) W_n^k|^2$$

$$+ 4| \varepsilon\sqrt{h_\ell} \sum_{k=0}^{M-1} (\sigma(D_{h_\ell}^\varepsilon(t_n)) - \sigma(D_{h_{\ell-1}}^\varepsilon(t_n))) W_n^k|^2$$

$$+ 2h_\ell \sum_{k=0}^{M-1} \langle D_{h_\ell}^\varepsilon(t_n) - D_{h_{\ell-1}}^\varepsilon(t_n), \mu(D_{h_\ell}^\varepsilon(t_n^k)) - \mu(D_{h_\ell}^\varepsilon(t_n)) \rangle$$

$$+ 2h_\ell \sum_{k=0}^{M-1} \langle D_{h_\ell}^\varepsilon(t_n) - D_{h_{\ell-1}}^\varepsilon(t_n), \mu(D_{h_\ell}^\varepsilon(t_n)) - \mu(D_{h_{\ell-1}}^\varepsilon(t_n)) \rangle$$

$$+ 2\varepsilon\sqrt{h_\ell} \sum_{k=0}^{M-1} \langle D_{h_\ell}^\varepsilon(t_n) - D_{h_{\ell-1}}^\varepsilon(t_n), (\sigma(D_{h_\ell}^\varepsilon(t_n^k)) - \sigma(D_{h_\ell}^\varepsilon(t_n))) W_n^k \rangle$$

$$+ 2\varepsilon\sqrt{h_\ell} \sum_{k=0}^{M-1} \langle D_{h_\ell}^\varepsilon(t_n) - D_{h_{\ell-1}}^\varepsilon(t_n), (\sigma(D_{h_\ell}^\varepsilon(t_n)) - \sigma(D_{h_{\ell-1}}^\varepsilon(t_n))) W_n^k \rangle,$$

where $\langle u, v \rangle$ denotes the inner product of u and v. Therefore,

$$\mathbb{E}[|D_{h_\ell}^\varepsilon(t_{n+1}) - D_{h_{\ell-1}}^\varepsilon(t_{n+1})|^2]$$

$$\leq \mathbb{E}[|D_{h_\ell}^\varepsilon(t_n) - D_{h_{\ell-1}}^\varepsilon(t_n)|^2] + 4Mh_\ell^2 \sum_{k=0}^{M-1} \mathbb{E}[|\mu(D_{h_\ell}^\varepsilon(t_n^k)) - \mu(D_{h_\ell}^\varepsilon(t_n))|^2]$$

$$+ 4Mh_\ell^2 \sum_{k=0}^{M-1} \mathbb{E}[|\mu(D_{h_\ell}^\varepsilon(t_n)) - \mu(D_{h_{\ell-1}}^\varepsilon(t_n))|^2]$$

$$+ 4\varepsilon^2 h_\ell \sum_{k=0}^{M-1} \mathbb{E}[|(\sigma(D_{h_\ell}^\varepsilon(t_n^k)) - \sigma(D_{h_\ell}^\varepsilon(t_n)))W_n^k|^2]$$

$$+ 4\varepsilon^2 h_\ell \sum_{k=0}^{M-1} \mathbb{E}[|(\sigma(D_{h_\ell}^\varepsilon(t_n)) - \sigma(D_{h_{\ell-1}}^\varepsilon(t_n)))W_n^k|^2]$$

$$+ 2h_\ell \sum_{k=0}^{M-1} \mathbb{E}[\langle D_{h_\ell}^\varepsilon(t_n) - D_{h_{\ell-1}}^\varepsilon(t_n), \mu(D_{h_\ell}^\varepsilon(t_n^k)) - \mu(D_{h_\ell}^\varepsilon(t_n))\rangle]$$

$$+ 2h_\ell \sum_{k=0}^{M-1} \mathbb{E}[\langle D_{h_\ell}^\varepsilon(t_n) - D_{h_{\ell-1}}^\varepsilon(t_n), \mu(D_{h_\ell}^\varepsilon(t_n)) - \mu(D_{h_{\ell-1}}^\varepsilon(t_n))\rangle],$$

where we used that W_n^k is independent from $D_{h_\ell}^\varepsilon(t_n), D_{h_{\ell-1}}^\varepsilon(t_n)$, and $D_{h_\ell}^\varepsilon(t_n^k)$. Hence, by Lemma 4, there are positive constants C_1 and C_2 that only depend on $a, b, T, m, D(0)$, such that

$$\mathbb{E}[|D_{h_\ell}^\varepsilon(t_{n+1}) - D_{h_{\ell-1}}^\varepsilon(t_{n+1})|^2]$$

$$\leq \mathbb{E}[|D_{h_\ell}^\varepsilon(t_n) - D_{h_{\ell-1}}^\varepsilon(t_n)|^2] + 4aC_1 M^4 h_\ell^4 + 4aC_2\varepsilon^2 M^3 h_\ell^3$$

$$+ 4aMh_\ell^2 \sum_{k=0}^{M-1} \mathbb{E}[|D_{h_\ell}^\varepsilon(t_n) - D_{h_{\ell-1}}^\varepsilon(t_n)|^2] + 4bC_1\varepsilon^2 M^3 h_\ell^3 + 4bC_2\varepsilon^4 M^2 h_\ell^2$$

$$+ 4b\varepsilon^2 h_\ell \sum_{k=0}^{M-1} \mathbb{E}[|D_{h_\ell}^\varepsilon(t_n) - D_{h_{\ell-1}}^\varepsilon(t_n)|^2]$$

$$+ 2h_\ell \sum_{k=0}^{M-1} \mathbb{E}[\langle D_{h_\ell}^\varepsilon(t_n) - D_{h_{\ell-1}}^\varepsilon(t_n), \mu(D_{h_\ell}^\varepsilon(t_n^k)) - \mu(D_{h_\ell}^\varepsilon(t_n))\rangle]$$

$$+ 2h_\ell \sqrt{a} M \mathbb{E}[|D_{h_\ell}^\varepsilon(t_n) - D_{h_{\ell-1}}^\varepsilon(t_n)|^2],$$

where the final term follows from the Cauchy-Schwarz inequality. Continuing,

$$\mathbb{E}[|D_{h_\ell}^\varepsilon(t_{n+1}) - D_{h_{\ell-1}}^\varepsilon(t_{n+1})|^2]$$

$$\leq \mathbb{E}[|D_{h_\ell}^\varepsilon(t_n) - D_{h_{\ell-1}}^\varepsilon(t_n)|^2] + (2\sqrt{a} + 4aMh_\ell + \varepsilon^2 4b)Mh_\ell \mathbb{E}[|D_{h_\ell}^\varepsilon(t_n) - D_{h_{\ell-1}}^\varepsilon(t_n)|^2]$$

$$+ 4aC_1 M^4 h_\ell^4 + 4aC_2\varepsilon^2 M^3 h_\ell^3 + 4bC_1\varepsilon^2 M^3 h_\ell^3 + 4bC_2\varepsilon^4 M^2 h_\ell^2$$

$$+ 2h_\ell \sum_{k=0}^{M-1} \mathbb{E}[\langle D_{h_\ell}^\varepsilon(t_n) - D_{h_{\ell-1}}^\varepsilon(t_n), \mu(D_{h_\ell}^\varepsilon(t_n^k)) - \mu(D_{h_\ell}^\varepsilon(t_n))\rangle]. \tag{17}$$

13

We turn to the term (17). Applying Lemma 5, we know

$$\mu_i(D^\varepsilon_{h_\ell}(t^k_n)) - \mu_i(D^\varepsilon_{h_\ell}(t_n)) = A_k + B_k + E_k.$$

Also we notice,

$$\mathbb{E}[|A_k|^2] \le K_1 M^2 h^2_\ell,$$

where K_1 is a constant that only depends on $a, b, T, m, D(0)$. Utilizing Lemmas 1 and 4

$$
\begin{aligned}
\mathbb{E}[|E_k|^2] &\le a h_\ell \varepsilon^2 \mathbb{E}\left[|D^\varepsilon_{h_\ell}(t^k_n) - D^\varepsilon_{h_\ell}(t_n)|^2 \sum_{j=0}^{k-1} |\sigma(D^\varepsilon_{h_\ell}(t^j_n))W^j_n|^2\right] \\
&\le a h_\ell \varepsilon^2 \left(\mathbb{E}[|D^\varepsilon_{h_\ell}(t^k_n) - D^\varepsilon_{h_\ell}(t_n)|^4]\right)^{1/2} \left(k \sum_{j=0}^{k-1} \mathbb{E}\left[|\sigma(D^\varepsilon_{h_\ell}(t^j_n))W^j_n|^4\right]\right)^{1/2} \\
&\le K_2 M^3 h^3_\ell \varepsilon^2 + K_3 M^2 h^2_\ell \varepsilon^4,
\end{aligned}
\tag{18}
$$

where K_2 and K_3 are constants depending only on $a, b, T, m, D(0)$. As a result,

$$
\begin{aligned}
2h_\ell \sum_{k=0}^{M-1} &\mathbb{E}[\langle D^\varepsilon_{h_\ell}(t_n) - D^\varepsilon_{h_{\ell-1}}(t_n), \mu(D^\varepsilon_{h_\ell}(t^k_n)) - \mu(D^\varepsilon_{h_\ell}(t_n))\rangle] \\
&= 2h_\ell \sum_{k=0}^{M-1} \mathbb{E}[\langle D^\varepsilon_{h_\ell}(t_n) - D^\varepsilon_{h_{\ell-1}}(t_n), A_k\rangle] \\
&\quad + 2h_\ell \sum_{k=0}^{M-1} \mathbb{E}[\langle D^\varepsilon_{h_\ell}(t_n) - D^\varepsilon_{h_{\ell-1}}(t_n), B_k\rangle] \\
&\quad + 2h_\ell \sum_{k=0}^{M-1} \mathbb{E}[\langle D^\varepsilon_{h_\ell}(t_n) - D^\varepsilon_{h_{\ell-1}}(t_n), E_k\rangle] \\
&\le 2Mh_\ell \mathbb{E}[|D^\varepsilon_{h_\ell}(t_n) - D^\varepsilon_{h_{\ell-1}}(t_n)|^2] + h_\ell \sum_{k=0}^{M-1} \mathbb{E}[|A_k|^2] + h_\ell \sum_{k=0}^{M-1} \mathbb{E}[|E_k|^2] \\
&\le 2Mh_\ell \mathbb{E}[|D^\varepsilon_{h_\ell}(t_n) - D^\varepsilon_{h_{\ell-1}}(t_n)|^2] + K_1 M^3 h^3_\ell + K_2 M^4 h^4_\ell \varepsilon^2 + K_3 M^3 h^3_\ell \varepsilon^4,
\end{aligned}
\tag{19}
$$

where the first inequality follows from: (i) the observation that the expectation (19) is zero, (ii) the Cauchy-Schwarz inequality, and (iii) the inequality $2ab \le a^2 + b^2$. Combining all the estimates above, we find

$$
\begin{aligned}
\mathbb{E}[|D^\varepsilon_{h_\ell}&(t_{n+1}) - D^\varepsilon_{h_{\ell-1}}(t_{n+1})|^2] \\
&\le \mathbb{E}\left[|D^\varepsilon_{h_\ell}(t_n) - D^\varepsilon_{h_{\ell-1}}(t_n)|^2\right] \\
&\quad + (2 + 2\sqrt{a} + 4aMh_\ell + 4b\varepsilon^2)Mh_\ell \mathbb{E}\left[|D^\varepsilon_{h_\ell}(t_n) - D^\varepsilon_{h_{\ell-1}}(t_n)|^2\right] \\
&\quad + 4aC_1 M^4 h^4_\ell + 4aC_2 \varepsilon^2 M^3 h^3_\ell + 4bC_1 \varepsilon^2 M^3 h^3_\ell + 4bC_2 \varepsilon^4 M^2 h^2_\ell \\
&\quad + K_1 M^3 h^3_\ell + K_2 M^4 h^4_\ell \varepsilon^2 + K_3 M^3 h^3_\ell \varepsilon^4.
\end{aligned}
$$

Noting that the dominant terms above are of order $h^2_{\ell-1}\epsilon^4$ and $h^3_{\ell-1}$, an application of Gronwall's inequality completes the proof. $\square$

14

We are now ready to prove our main result.

Proof of Theorem 1. Following [1], we first prove the result in the case that $f(x) = x_i$ for some $i \in \{1, \ldots, d\}$. We have that for $n \leq M^{\ell-1} - 1$,

$$
[D^\varepsilon_{h_\ell}(t_{n+1}) - D^\varepsilon_{h_{\ell-1}}(t_{n+1})]_i
$$

$$
= [D^\varepsilon_{h_\ell}(t_n) - D^\varepsilon_{h_{\ell-1}}(t_n)]_i + h_\ell \sum_{k=0}^{M-1} (\mu_i(D^\varepsilon_{h_\ell}(t^k_n)) - \mu_i(D^\varepsilon_{h_\ell}(t_n)))
$$

$$
+ h_\ell \sum_{k=0}^{M-1} (\mu_i(D^\varepsilon_{h_\ell}(t_n)) - \mu_i(D^\varepsilon_{h_{\ell-1}}(t_n))) + \varepsilon\sqrt{h_\ell} \sum_{k=0}^{M-1} (\sigma_i(D^\varepsilon_{h_\ell}(t^k_n)) - \sigma_i(D^\varepsilon_{h_\ell}(t_n)))W^k_n
$$

$$
+ \varepsilon\sqrt{h_\ell} \sum_{k=0}^{M-1} (\sigma_i D^\varepsilon_{h_\ell}(t_n)) - \sigma_i(D^\varepsilon_{h_{\ell-1}}(t_n)))W^k_n,
$$

where μ_i is the ith component of μ and σ_i is the ith row of σ. As a result, and after some manipulation,

$$
\mathsf{Var}([D^\varepsilon_{h_\ell}(t_{n+1}) - D^\varepsilon_{h_{\ell-1}}(t_{n+1})]_i)
$$

$$
\leq (1 + Mh_\ell)\mathsf{Var}([D^\varepsilon_{h_\ell}(t_n) - D^\varepsilon_{h_{\ell-1}}(t_n)]_i)
$$

$$
+ 4h^2_\ell M \sum_{k=0}^{M-1} \mathsf{Var}(\mu_i(D^\varepsilon_{h_\ell}(t^k_n)) - \mu_i(D^\varepsilon_{h_\ell}(t_n))) \tag{20}
$$

$$
+ (4Mh_\ell + 1)Mh_\ell \mathsf{Var}(\mu_i(D^\varepsilon_{h_\ell}(t_n)) - \mu_i(D^\varepsilon_{h_{\ell-1}}(t_n))) \tag{21}
$$

$$
+ 4\varepsilon^2 h_\ell \sum_{k=0}^{M-1} \mathsf{Var}((\sigma_i(D^\varepsilon_{h_\ell}(t^k_n)) - \sigma_i(D^\varepsilon_{h_\ell}(t_n)))W^k_n) \tag{22}
$$

$$
+ 4\varepsilon^2 h_\ell \sum_{k=0}^{M-1} \mathsf{Var}((\sigma_i(D^\varepsilon_{h_\ell}(t_n)) - \sigma_i(D^\varepsilon_{h_{\ell-1}}(t_n)))W^k_n) \tag{23}
$$

$$
+ 2\mathsf{Cov}([D^\varepsilon_{h_\ell}(t_n) - D^\varepsilon_{h_{\ell-1}}(t_n)]_i, h_\ell \sum_{k=0}^{M-1} (\mu_i(D^\varepsilon_{h_\ell}(t^k_n)) - \mu_i(D^\varepsilon_{h_\ell}(t_n)))). \tag{24}
$$

We must bound each of the terms above in order to apply Gronwall's inequality. We first consider (22), which leads to a dominant factor. Lemma 4 implies that

$$
\sum_{k=0}^{M-1} \mathsf{Var}((\sigma_i(D^\varepsilon_{h_\ell}(t^k_n)) - \sigma_i(D^\varepsilon_{h_\ell}(t_n)))W^k_n) \leq \sum_{k=0}^{M-1} \mathbb{E}[|\sigma_i(D^\varepsilon_{h_\ell}(t^k_n)) - \sigma_i(D^\varepsilon_{h_\ell}(t_n)))|^2]
$$

$$
\leq Mb(c_1 M^2 h^2_\ell + c_2 \varepsilon^2 Mh_\ell).
$$

Similarly, by Lemma 6 we may bound (23), which also yields a dominant factor,

$$
\sum_{k=0}^{M-1} \mathsf{Var}((\sigma_i(D^\varepsilon_{h_\ell}(t_n)) - \sigma_i(D^\varepsilon_{h_{\ell-1}}(t_n)))W^k_n) \leq \sum_{k=0}^{M-1} \mathbb{E}[|\sigma_i(D^\varepsilon_{h_\ell}(t_n)) - \sigma_i(D^\varepsilon_{h_{\ell-1}}(t_n)))|^2]
$$

$$
\leq Mb(d_1 M^2 h^2_\ell + d_2 \varepsilon^4 Mh_\ell),
$$

15

where c_1, c_2, d_1 and d_2 are positive constants only depending on $a, b, T, m, D(0)$.

Turning to (20), we have the following lemma.

Lemma 7.

$$\mathsf{Var}\left(\mu_i(D^\varepsilon_{h_\ell}(t^k_n)) - \mu_i(D^\varepsilon_{h_\ell}(t_n))\right) \le CMh_\ell\varepsilon^2,$$

where C is a positive constant that only depends on $a, b, T, d, m, D(0)$.

Proof. From Lemma 12 in the appendix (Taylor approximation), we have

$$\mu_i(D^\varepsilon_{h_\ell}(t^k_n)) - \mu_i(D^\varepsilon_{h_\ell}(t_n)) = \rho^k(t_n) \cdot (D^\varepsilon_{h_\ell}(t^k_n) - D^\varepsilon_{h_\ell}(t_n)), \tag{25}$$

where

$$\rho^k(t_n) = \int_0^1 \left[\nabla\mu_i(D^\varepsilon_{h_\ell}(t_n)) + r(D^\varepsilon_{h_\ell}(t^k_n) - D^\varepsilon_{h_\ell}(t_n))\right] dr.$$

In order to bound the right hand side of (25), we will apply Lemma 11 in the appendix with $A^{\varepsilon,h_{\ell-1}} = [\rho^k(t_n)]_j$ and $B^{\varepsilon,h_{\ell-1}} = [D^\varepsilon_{h_\ell}(t^k_n) - D^\varepsilon_{h_\ell}(t_n)]_j$. Hence, we must find appropriate bounds on these components.

We begin with $B^{\varepsilon,h_{\ell-1}}$. We use Lemmas 1 and 2 after iterating (10) to find

$$\mathsf{Var}([D^\varepsilon_{h_\ell}(t^k_n) - D^\varepsilon_{h_\ell}(t_n)]_j)$$

$$\le 2\mathsf{Var}\left(h_\ell \sum_{r=0}^{k-1} \mu_j(D^\varepsilon_{h_\ell}(t^r_n))\right) + 2\mathsf{Var}\left(\varepsilon\sqrt{h_\ell}\sum_{r=0}^{k-1}\sigma_j(D^\varepsilon_{h_\ell}(t^r_n))W^r_n\right)$$

$$\le 2h_\ell^2\mathsf{Var}\left(\sum_{r=0}^{k-1}(\mu_j(D^\varepsilon_{h_\ell}(t^r_n)) - \mu_j(z_{h_\ell}(t^r_n)))\right) + 2\varepsilon^2 h_\ell\mathbb{E}\left[\left|\sum_{r=0}^{k-1}\sigma_j(D^\varepsilon_{h_\ell}(t^r_n))W^r_n\right|^2\right] \tag{26}$$

$$\le C_1 M^2 h_\ell^2\varepsilon^2 + C_2 M h_\ell\varepsilon^2,$$

where C_1 and C_2 are positive constants that only depend on $a, b, T, m, D(0)$.

Turning to $A^{\varepsilon,h_{\ell-1}}$, we apply Lemma 10 in the appendix with $X_1(s) = D^\varepsilon_{h_\ell}(s)$, $X_2(s) = D^\varepsilon_{h_\ell}(\eta(s))$, $x_1(s) = z_h(s)$, $x_2(s) = z_h(\eta(s))$ and $u(x) = \nabla_j\mu_i(x)$ to obtain

$$\mathsf{Var}([\rho^k(t_n)]_j) = \mathsf{Var}\left(\int_0^1 [\nabla_j\mu_i(D^\varepsilon_{h_\ell}(t_n) + r(D^\varepsilon_{h_\ell}(t^k_n) - D^\varepsilon_{h_\ell}(t_n)))]dr\right) \le K\varepsilon^2,$$

where K is positive constant that only depends on $a, b, T, m, D(0)$.

We may now combine Lemma 11 with Lemma 3 to conclude

$$\mathsf{Var}\left([\rho^k(t_n)]_j \cdot [D^\varepsilon_{h_\ell}(t^k_n) - D^\varepsilon_{h_\ell}(t_n)]_j\right) \le \hat{C}KM^2 h_\ell^2\varepsilon^2 + 15a\mathsf{Var}([D^\varepsilon_{h_\ell}(t^k_n) - D^\varepsilon_{h_\ell}(t_n)]_j)$$

$$\le (\hat{C}K + 15aC_1)M^2 h_\ell^2\varepsilon^2 + 15C_2 M h_\ell\varepsilon^2$$

$$\le \hat{C}_1 M h_\ell\varepsilon^2,$$

where $\hat{C}_1$ is positive and does not depend on ε and h_ℓ, and we applied (26) in the second inequality.

Returning to (25), the above allows us to conclude

$$\mathsf{Var}(\mu_i(D^\varepsilon_{h_\ell}(t^k_n)) - \mu_i(D^\varepsilon_{h_\ell}(t_n))) \le d^2\hat{C}_1 M h_\ell\varepsilon^2. \qquad \square$$

We now turn to the first term of (21).

Lemma 8.

$$\text{Var}\left(\mu_i(D^\varepsilon_{h_\ell}(t_n)) - \mu_i(D^\varepsilon_{h_{\ell-1}}(t_n))\right)$$

$$\leq 15ad\sum_{j=1}^{d}\text{Var}([D^\varepsilon_{h_\ell}(t_n) - D^\varepsilon_{h_{\ell-1}}(t_n)]_j) + K_1 M^2 h_\ell^2 \varepsilon^2 + K_2 M h_\ell \varepsilon^6,$$

where K_1, K_2 are positive constants that only depend on $a, b, T, d, m, D(0)$.

Proof. We first write

$$\mu_i(D^\varepsilon_{h_\ell}(t_n)) - \mu_i(D^\varepsilon_{h_{\ell-1}}(t_n)) = \rho(t_n) \cdot (D^\varepsilon_{h_\ell}(t_n) - D^\varepsilon_{h_{\ell-1}}(t_n)),$$

where

$$\rho(t_n) = \int_0^1 [\nabla\mu_i(D^\varepsilon_{h_{\ell-1}}(t_n) + r(D^\varepsilon_{h_\ell}(t_n) - D^\varepsilon_{h_{\ell-1}}(t_n)))]dr.$$

We will again apply Lemma 11 to get the necessary bounds. Therefore, we let $A^{\varepsilon,h} = [\rho(t_n)]_j$ and $B^{\varepsilon,h} = [D^\varepsilon_{h_\ell}(t_n) - D^\varepsilon_{h_{\ell-1}}(t_n)]_j$.

Letting $X_1(s) = D^\varepsilon_{h_\ell}(s)$, $X_2(s) = D^\varepsilon_{h_{\ell-1}}(s)$, $x_1(s) = z_{h_\ell}(s)$, $x_2(s) = z_{h_{\ell-1}}(s)$ and $u(x) = \nabla_j\mu_i(x)$ for an application of Lemma 10, we have

$$\text{Var}(A^{\varepsilon,h}) \leq K\varepsilon^2,$$

for some $K(a, b, T, m, D(0))$, where we recall the running assumption that $|[\nabla\mu_i]_j|^2$ is uniformly bounded by a. Hence, applying Lemmas 6 and 11 we see there are positive constants K_1, K_2 depending only on $a, b, T, m, D(0)$, such that,

$$\text{Var}([\rho(t_n)]_j([D^\varepsilon_{h_\ell}(t_n) - D^\varepsilon_{h_{\ell-1}}(t_n)]_j))$$
$$\leq K_1 M^2 h_\ell^2 \varepsilon^2 + K_2 M h_\ell \varepsilon^6 + 15a\text{Var}([D^\varepsilon_{h_\ell}(t_n) - D^\varepsilon_{h_{\ell-1}}(t_n)]_j,$$

and

$$\text{Var}(\mu_i(D^\varepsilon_{h_\ell}(t_n)) - \mu_i(D^\varepsilon_{h_{\ell-1}}(t_n)))$$

$$\leq 15ad\sum_{j=1}^{d}\text{Var}([D^\varepsilon_{h_\ell}(t_n) - D^\varepsilon_{h_{\ell-1}}(t_n)]_j) + d^2 K_1 M^2 h_\ell^2 \varepsilon^2 + d^2 K_2 M h_\ell \varepsilon^6.$$

$\square$

Finally, we turn to the term (24).

Lemma 9.

$$\text{Cov}\left([D^\varepsilon_{h_\ell}(t_n) - D^\varepsilon_{h_{\ell-1}}(t_n)]_i, h_\ell\sum_{k=0}^{M-1}(\mu_i(D^\varepsilon_{h_\ell}(t^k_n)) - \mu_i(D^\varepsilon_{h_\ell}(t_n)))\right)$$

$$\leq Mh_\ell\text{Var}([D^\varepsilon_{h_\ell}(t_n) - D^\varepsilon_{h_{\ell-1}}(t_n)]_i) + K_1 M^3 h_\ell^3 \varepsilon^2 + K_2 M^5 h_\ell^5 \varepsilon^2 + K_3 M^3 h_\ell^3 \varepsilon^4,$$

where K_1, K_2 and K_3 are positive constants that only depend on $a, b, T, m, D(0)$.

Proof. As a result of combining (15) in Lemma 5 with

$$\mathsf{Cov}\left([D^{\varepsilon}_{h_\ell}(t_n) - D^{\varepsilon}_{h_{\ell-1}}(t_n)]_i, h_\ell \sum_{k=0}^{M-1} B_k\right) = 0,$$

where we recall the definition of B_k in (16), we have

$$
\begin{aligned}
\mathsf{Cov}&\left([D^{\varepsilon}_{h_\ell}(t_n) - D^{\varepsilon}_{h_{\ell-1}}(t_n)]_i, h_\ell \sum_{k=0}^{M-1} (\mu_i(D^{\varepsilon}_{h_\ell}(t_n^k)) - \mu_i(D^{\varepsilon}_{h_\ell}(t_n)))\right) \\
&= \mathsf{Cov}\left([D^{\varepsilon}_{h_\ell}(t_n) - D^{\varepsilon}_{h_{\ell-1}}(t_n)]_i, h_\ell \sum_{k=0}^{M-1} (A_k + E_k)\right) \\
&\quad + \mathsf{Cov}\left([D^{\varepsilon}_{h_\ell}(t_n) - D^{\varepsilon}_{h_{\ell-1}}(t_n)]_i, h_\ell \sum_{k=0}^{M-1} B_k\right) \\
&\leq M h_\ell \mathsf{Var}([D^{\varepsilon}_{h_\ell}(t_n) - D^{\varepsilon}_{h_{\ell-1}}(t_n)]_i) + \frac{1}{2} h_\ell \sum_{k=0}^{M-1} \mathsf{Var}(A_k) + \frac{1}{2} h_\ell \sum_{k=0}^{M-1} \mathsf{Var}(E_k).
\end{aligned}
\tag{27}
$$

First we want to estimate $\mathsf{Var}(A_k)$. Applying Lemma 11 with

$$A^{\varepsilon, h_{\ell-1}} = \int_0^1 \left[\nabla_j \mu_i(D^{\varepsilon}_{h_\ell}(t_n) + r(D^{\varepsilon}_{h_\ell}(t_n^k) - D^{\varepsilon}_{h_\ell}(t_n)))\right] dr$$

and

$$B^{\varepsilon, h_{\ell-1}} = h_\ell \sum_{r=0}^{k-1} \mu_j(D^{\varepsilon}_{h_\ell}(t_n^r)),$$

we can get for some $K_1(a, b, T, m, d, D(0))$ that may change from line to line,

$$
\begin{aligned}
\mathsf{Var}(A_k) &\leq K_1 M^2 h_\ell^2 \varepsilon^2 + 15 a d \sum_{i=1}^{d} \mathsf{Var}\left(h_\ell \sum_{r=0}^{k-1} \mu_i(D^{\varepsilon}_{h_\ell}(t_n^r))\right) \\
&\leq K_1 M^2 h_\ell^2 \varepsilon^2 + 15 a d \sum_{i=1}^{d} \mathbb{E}\left[h_\ell^2 \left(\sum_{r=0}^{k-1} \mu_i(D^{\varepsilon}_{h_\ell}(t_n^r)) - \mu_j(z_{h_\ell}(t_n^r))\right)^2\right] \\
&\leq K_1 M^2 h_\ell^2 \varepsilon^2,
\end{aligned}
$$

where we also use Lemma 2 for the last line. On the other hand, from (18)

$$\mathsf{Var}(E_k) \leq \mathbb{E}[|E_k|^2] \leq K_2 M^4 h_\ell^4 \varepsilon^2 + K_3 M^2 h_\ell^2 \varepsilon^4.$$

Returning to (27), we see,

$$
\begin{aligned}
\mathsf{Cov}&\left([D^{\varepsilon}_{h_\ell}(t_n) - D^{\varepsilon}_{h_{\ell-1}}(t_n)]_i, h_\ell \sum_{k=0}^{M-1} (\mu_i(D^{\varepsilon}_{h_\ell}(t_n^k)) - \mu_i(D^{\varepsilon}_{h_\ell}(t_n)))\right) \\
&\leq M h_\ell \sum_{j=1}^{d} \mathsf{Var}([D^{\varepsilon}_{h_\ell}(t_n) - D^{\varepsilon}_{h_{\ell-1}}(t_n)]_j) + \frac{1}{2} K_1 M^3 h_\ell^3 \varepsilon^2 + \frac{1}{2} K_2 M^5 h_\ell^5 \varepsilon^2 + \frac{1}{2} K_3 M^3 h_\ell^3 \varepsilon^4.
\end{aligned}
$$

$\square$

Now we return to (20)–(24) and combine all the estimates above to conclude that there exist C_1, C_2, and C_3 which only depend on $a, b, T, m, d, D(0)$ such that

$$\mathsf{Var}\left([D_{h_\ell}^\varepsilon(t_{n+1}) - D_{h_{\ell-1}}^\varepsilon(t_{n+1})]_i\right) \leq \mathsf{Var}[D_{h_\ell}^\varepsilon(t_n) - D_{h_{\ell-1}}^\varepsilon(t_n)]_i + C_1 M^3 h_\ell^3 \varepsilon^2 + C_2 M^2 h_\ell^2 \varepsilon^4$$

$$+ C_3 M h_\ell \sum_{j=1}^d \mathsf{Var}([D_{h_\ell}^\varepsilon(t_n) - D_{h_{\ell-1}}^\varepsilon(t_n)]_j).$$

Therefore,

$$\max_{i=1,2,\cdots,d} \mathsf{Var}\left([D_{h_\ell}^\varepsilon(t_{n+1}) - D_{h_{\ell-1}}^\varepsilon(t_{n+1})]_i\right)$$

$$\leq \max_{i=1,2,\cdots,d} \mathsf{Var}\left([D_{h_\ell}^\varepsilon(t_n) - D_{h_{\ell-1}}^\varepsilon(t_n)]_i\right) + C_1 M^3 h_\ell^3 \varepsilon^2 + C_2 M^2 h_\ell^2 \varepsilon^4$$

$$+ C_3 d M h_\ell \max_{i=1,2,\cdots,d} \mathsf{Var}\left([D_{h_\ell}^\varepsilon(t_n) - D_{h_{\ell-1}}^\varepsilon(t_n)]_i\right).$$

Applying Gronwall's lemma, we obtain,

$$\max_{0 \leq n \leq M^{\ell-1}} \max_{1 \leq i \leq d} \mathsf{Var}\left([D_{h_\ell}^\varepsilon(t_n) - D_{h_{\ell-1}}^\varepsilon(t_n)]_i\right) \leq C_1 M^2 h_\ell^2 \varepsilon^2 + C_2 M h_\ell \varepsilon^4,$$

where C_1 and C_2 are some universal constants which only depend on $a, b, T, m, d, D(0)$.

We have shown the result under the assumption that $f(x) = x_i$. To show the general case, note that from Lemma 12 in the appendix we have

$$f(D_{h_\ell}^\varepsilon(t_n)) - f(D_{h_{\ell-1}}^\varepsilon(t_n)))$$

$$= \int_0^1 [\nabla f(D_{h_\ell}^\varepsilon(t_n) + r(D_{h_\ell}^\varepsilon(t_n) - D_{h_\ell}^\varepsilon(t_n)))] dr \cdot (D_{h_\ell}^\varepsilon(t_n) - D_{h_{\ell-1}}^\varepsilon(t_n)).$$

We let $X_1(t) = D_{h_\ell}^\varepsilon(t)$, $X_2(t) = D_{h_{\ell-1}}^\varepsilon(t)$, $x_1(t) = z_{h_\ell}(t)$, $x_2(t) = z_{h_{\ell-1}}(t)$ and $u(x) = \nabla_j f(x)$ in an application of Lemma 10 which yields

$$\mathsf{Var}\left(\int_0^1 [\nabla_j f(D_{h_\ell}^\varepsilon(t_n) + r(D_{h_\ell}^\varepsilon(t_n) - D_{h_\ell}^\varepsilon(t_n)))] dr\right) \leq K \varepsilon^2,$$
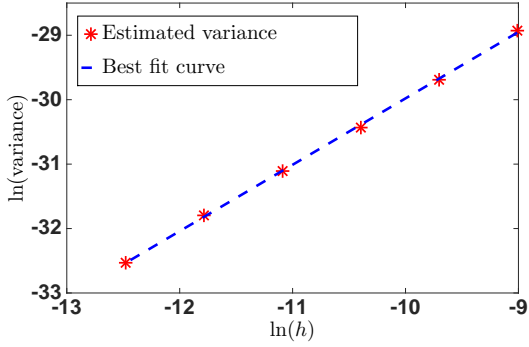
where K is a universal constant that depends on $C_L, D, a, b, T, D(0)$. Hence, by an application of Lemmas 6 and 11 and the work above we see,

$$\mathsf{Var}\left(\int_0^1 \nabla_j f(D_{h_\ell}^\varepsilon(t_n) + r(D_{h_\ell}^\varepsilon(t_n) - D_{h_\ell}^\varepsilon(t_n))) dr \cdot [D_{h_\ell}^\varepsilon(t_n) - D_{h_{\ell-1}}^\varepsilon(t_n)]_j\right)$$

$$\leq K(d_1 M^2 h_\ell^2 + d_2 M h_\ell \varepsilon^4) \varepsilon^2 + 15 d C_L^2 \mathsf{Var}([D^\varepsilon(s) - D_{h_\ell}^\varepsilon(s)]_j)$$

$$\leq (Kd_1 + 15 d C_L^2 C_1) M^2 h_\ell^2 \varepsilon^2 + (15 d C_L^2 C_2 + K d_2) M h_\ell \varepsilon^4.$$
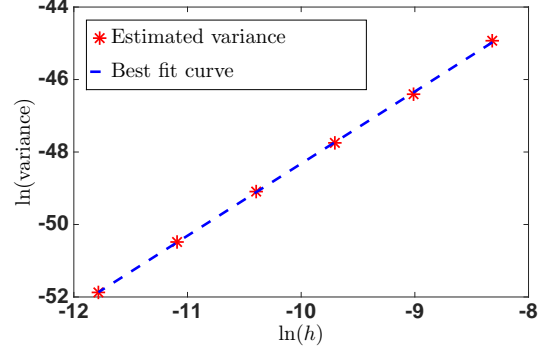
Thus

$$\mathsf{Var}(f(D_{h_\ell}^\varepsilon(t_n)) - f(D_{h_{\ell-1}}^\varepsilon(t_n))) \leq d^2 (Kd_1 + 15 d C_L^2 C_1) M^2 h_\ell^2 \varepsilon^2 + d^2 (15 d C_L^2 C_2 + K d_2) h_\ell \varepsilon^4,$$

giving the result. $\qquad \square$

(a) $\varepsilon = 2^{-6}$ fixed while h is varied. The best fit curve is $y = 1.03x - 19.68$.

(b) $\varepsilon = 2^{-17}$ fixed while h is varied. The best fit curve is $y = 1.99x - 28.43$.

Figure 1: Log-log plots of $\mathsf{Var}(D^\varepsilon_{h_\ell}(1) - D^\varepsilon_{h_{\ell-1}}(1))$ with ε held constant and $h_{\ell-1}$ varied. The best fit curves for all data are overlain in the dashed blue line. Each data point in (a) was generated using 2,000 independent samples and each data point in (b) was generated using 5,000 independent samples.

4 Numerical Test

In this section we provide numerical evidence for the sharpness of both Theorem 1 and the computational complexity analyses provided in sections 2.1 and 2.2. We consider the following simple one dimensional model,

$$D^\varepsilon(t) = 1 - \int_0^t D^\varepsilon(s)ds + \varepsilon \int_0^t D^\varepsilon(s)dW(s),$$

where we simulate until $T = 1$.

To gather evidence in support of the sharpnesss of the bound $\mathsf{Var}(D^\varepsilon_{h_\ell}(t) - D^\varepsilon_{h_{\ell-1}}(t)) = O(h^2\varepsilon^2 + h\varepsilon^4)$, we fix one of h or ε in different scaling regimes and vary the other parameter in order to generate log-log plots. We note that there are four exponents to discover, and so four log-log plots are used. Note also that $h^2\varepsilon^2$ is the dominant term in $h^2\varepsilon^2 + h\varepsilon^4$ if and only if $h \geq \varepsilon^2$. We emphasize that these experiments use extreme parameter choices solely for the purpose of testing the sharpness of the delicate asymptotic bound.

The exponent of h in $h\varepsilon^4$. We fix $\varepsilon = 2^{-6}$ and vary
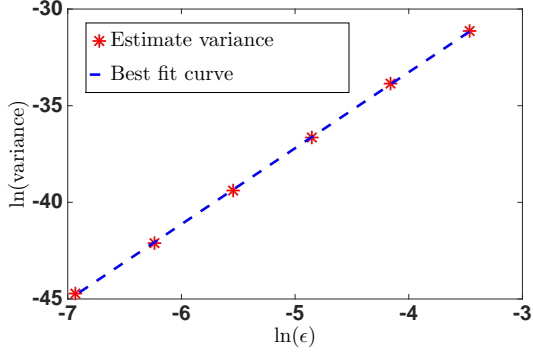
$$h_{\ell-1} \in \{2^{-13}, 2^{-14}, 2^{-15}, 2^{-16}, 2^{-17}, 2^{-18}\}$$

to ensure $h_{\ell-1} \leq \varepsilon^2$. As a result, $h_{\ell-1}\varepsilon^4$ is likely to be the dominant term in (3). See Figure 1(a), where the log-log plot is consistent with the functional form
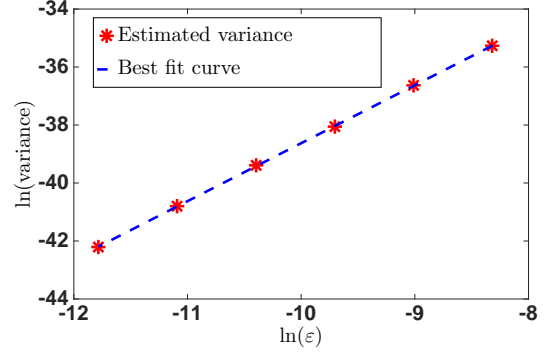
$$\mathsf{Var}(D^\varepsilon_{h_\ell}(T) - D^\varepsilon_{h_{\ell-1}}(T)) = O(h_{\ell-1}).$$

The exponent of h in $h^2\varepsilon^2$. We fix $\varepsilon = 2^{-17}$ and vary

$$h_{\ell-1} \in \{2^{-12}, 2^{-13}, 2^{-14}, 2^{-15}, 2^{-16}, 2^{-17}\}$$

20

(a) $h_{\ell-1} = 2^{-20}$ fixed while ε is varied. The best fit curve is $y = 3.93x - 17.54$.

(b) $h_{\ell-1} = 2^{-10}$ fixed while ε is varied. The best fit curve is $y = 2.00x - 18.62$.

Figure 2: Log-log plots of $\mathsf{Var}(D^\varepsilon_{h_\ell}(1) - D^\varepsilon_{h_{\ell-1}}(1))$ with $h_{\ell-1}$ held constant and ε varied. The best fit curves for all data are overlain in the dashed blue line. Each data point was generated using 2,000 independent samples.

to ensure $h_{\ell-1} \geq \varepsilon^2$. As a result, $h^2_{\ell-1}\varepsilon^2$ is likely to be the dominant term in (3). See Figure 1(b), where the log-log plot is consistent with the functional form

$$\mathsf{Var}(D^\varepsilon_{h_\ell}(1) - D^\varepsilon_{h_{\ell-1}}(1)) = O(h^2_{\ell-1}).$$

The exponent of ε in $h\varepsilon^4$. We fix $h_{\ell-1} = 2^{-20}$ and vary

$$\varepsilon \in \{2^{-5}, 2^{-6}, 2^{-7}, 2^{-8}, 2^{-9}, 2^{-10}\}$$

to ensure $h_{\ell-1} \leq \varepsilon^2$. As a result, $h_{\ell-1}\varepsilon^4$ is likely to be the dominant term in (3). See Figure 2(a), where the log-log plot is consistent with the functional form

$$\mathsf{Var}(D^\varepsilon_{h_\ell}(1) - D^\varepsilon_{h_{\ell-1}}(1)) = O(\varepsilon^4).$$

**The exponent of ε in $h^2\varepsilon^2$.** We fix $h_{\ell-1} = 2^{-10}$ and vary

$$\varepsilon \in \{2^{-12}, 2^{-13}, 2^{-14}, 2^{-15}, 2^{-16}, 2^{-17}\}$$

to ensure $h_{\ell-1} \geq \varepsilon^2$. As a result, $h^2_{\ell-1}\varepsilon^2$ is likely to be the dominant term in (3). See Figure 2(b), where the log-log plot is consistent with the functional form

$$\mathsf{Var}(D^\varepsilon_{h_\ell}(1) - D^\varepsilon_{h_{\ell-1}}(1)) = O(\varepsilon^2).$$

We turn to numerically demonstrating our conclusions related to the complexity of Euler based multilevel Monte Carlo and the complexity of Euler based standard Monte Carlo. We will measure complexity in two ways, by total number of random variables utilized and by required CPU time. Our implementation of MLMC proceeded as follows. We chose $h_\ell = 2^{-\ell}$ and for each $\delta > 0$ we set $L = \lceil \log(\delta)/\log(2) \rceil$. For each level we generated 200 independent
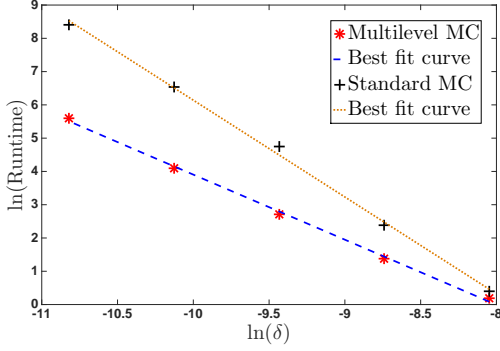
(a) $\varepsilon = 0.1$ held constant and δ varied. The best fit lines are $y = -1.96x - 15.65$ for Euler based MLMC and $y = -2.91x - 5.44$ for standard Euler based Monte Carlo.
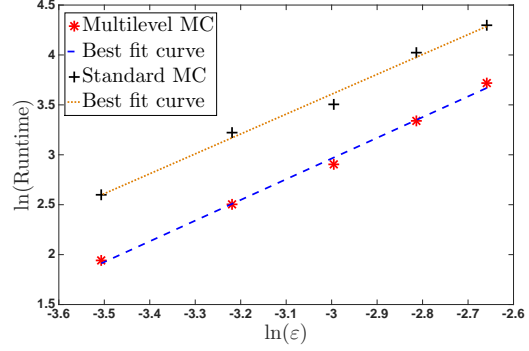
(b) $\delta = 2^{-14}$ held constant and ε varied. The best fit lines are $y = 2.07x+9.18$ for Euler based MLMC and $y = 1.99x + 9.58$ for standard Euler based Monte Carlo.

Figure 3: Log-log plots of runtime (in seconds) for both multilevel and standard Euler based Monte Carlo.

sample trajectories in order to estimate $\delta_{\epsilon,\ell}$, as defined in section 2.2. According to (6) and (7) we then selected

$$n_\ell = \left\lceil \delta^{-2} \sqrt{\delta_{\varepsilon,\ell} h_\ell} \sum_{j=0}^{L} \sqrt{\frac{\delta_{\varepsilon,j}}{h_j}} \right\rceil, \qquad \text{for } \ell \in \{0, 1, 2, \ldots, L\}.$$

We implemented Euler's method combined with standard Monte Carlo by selecting the number of paths by

$$N = \left\lceil \delta^{-2} \mathsf{Var}(D_h^\varepsilon(1)) \right\rceil$$

where $h = 2^{-L}$ and the parameter $\mathsf{Var}(D_h^\varepsilon(1))$ was estimated using 500 independent realizations of the relevant processes.

In Figures 3(a) and 4(a), we provide log-log plots of runtime (in seconds) and complexity (quantified by the total number of random variables utilized) for our implementation of multilevel and standard Monte Carlo with $\varepsilon = 0.1$ fixed and

$$\delta \in \{0.00032, 0.00016, 0.00008, 0.00004, 0.00002\},$$

which ensures $\delta > \frac{1}{3}e^{-\frac{1}{\varepsilon}}$ (see section 2.2). The best fit curves are consistent with the conclusion that the computational complexity of the Euler based multilevel Monte Carlo method is $O(\delta^{-2})$ while that of standard Monte Carlo method is $O(\delta^{-3})$ when ε is fixed. The Monte Carlo estimates which came from these simulations are detailed in tables 1 and 2. Notice that $\mathbb{E}[D^\varepsilon(1)]$ can be found explicitly in this case,
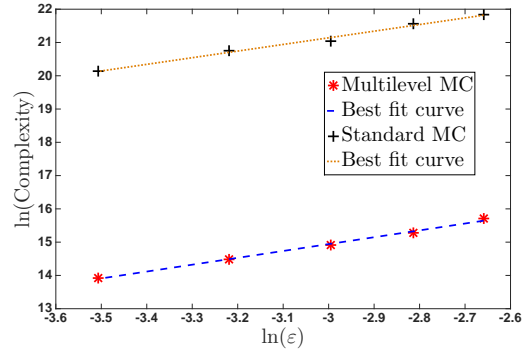
$$\mathbb{E}[D^\varepsilon(1)] = e^{-1} \approx 0.3678794.$$

In Figure 3(b) and 4(b), we provide similar log-log plots of runtime and computational complexity for Euler based multilevel Monte Carlo and standard Monte Carlo when $\delta = 2^{-14}$ is fixed and ε is varied as

$$\varepsilon \in \{0.07, 0.06, 0.05, 0.04, 0.03\},$$

22
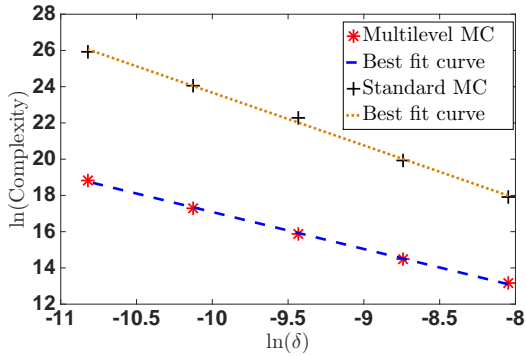
(a) $\varepsilon = 0.1$ held constant and δ varied. The best fit lines are $y = -2.04x - 3.33$ for MLMC and $y = -2.91x - 22.95$ for standard MC.

(b) $\delta = 2^{-14}$ held constant and ε varied. The best fit lines are $y = 2.06x + 21.13$ for MLMC and $y = 1.99x + 27.12$ for standard MC.

Figure 4: Log-log plots of computational complexity, quantified by the number of random variables used.

δ	Mean	Standard deviation of estimator
0.00032	0.367944	0.000305
0.00016	0.367906	0.000153
0.00008	0.367891	0.000077
0.00004	0.367863	0.000039
0.00002	0.367883	0.000020

Table 1: Result of Euler based multilevel Monte Carlo for fixed $\varepsilon = 0.1$ and varying δ.

δ	Mean	Standard deviation of estimator
0.00032	0.367449	0.000320
0.00016	0.368028	0.000160
0.00008	0.367839	0.000080
0.00004	0.367941	0.000040
0.00002	0.367851	0.000020

Table 2: Result of Euler based standard Monte Carlo for fixed $\varepsilon = 0.1$ and varying δ.

23

ε	Mean	Standard deviation of estimator
0.07	0.367834	0.000059
0.06	0.367920	0.000059
0.05	0.367819	0.000059
0.04	0.367856	0.000059
0.03	0.367925	0.000059

Table 3: Results of Euler based multilevel Monte Carlo for fixed $\delta = 2^{-14} \approx 0.000061$ and varying ε.

ε	Mean	Standard deviation of estimator
0.07	0.367830	0.000061
0.06	0.367755	0.000061
0.05	0.367933	0.000061
0.04	0.367809	0.000061
0.03	0.367879	0.000061

Table 4: Results of Euler based standard Monte Carlo for fixed $\delta = 2^{-14} \approx 0.000061$ and varying ε.

which ensures $\delta > e^{-\frac{1}{\varepsilon}}$. The best fit curves are again consistent with the conclusion that the complexity of Euler based multilevel Monte Carlo and standard Monte Carlo Methods are both $O(\varepsilon^{-2})$ when δ is fixed. The Monte Carlo estimates which came from these simulations are detailed in tables 3 and 4.

5   Summary

This work focussed on Monte Carlo methods for approximating expectations arising from SDEs with small noise. Our motivation was that for the highly effective multilevel approach, the classical strong error measure is less relevant than the variance between coupled pairs of paths at different discretization levels. By analyzing this variance directly, we showed that, under reasonable assumptions, a basic Euler–Maruyama discretization leads to optimal asymptotic computational complexity when used in a multilevel setting.

A   Some Technical Lemmas

We provide here some technical lemmas which were used in section 3.

The following is Lemma 5 in the appendix of [1].

Lemma 10. *Suppose $X_1(s)$ and $X_2(s)$ are stochastic processes on $\mathbb{R}^d$ and that $x_1(s)$ and $x_2(s)$ are deterministic processes on $\mathbb{R}^d$. Further, suppose that*

$$\sup_{s \leq T} \mathbb{E}\left[|X_1(s) - x_1(s)|^2\right] \leq \widehat{C}_1(T)\varepsilon^2, \quad \sup_{s \leq T} \mathbb{E}\left[|X_2(s) - x_2(s)|^2\right] \leq \widehat{C}_2(T)\varepsilon^2,$$

for some $\widehat{C}_1, \widehat{C}_2$ depending upon T. Assume that $u : \mathbb{R}^d \rightarrow \mathbb{R}$ is Lipschitz with Lipschitz constant C_L. Then,

$$\sup_{s \leq T} \mathsf{Var}\left(\int_0^1 u(X_2(s) + r(X_1(s) - X_2(s)))dr \right) \leq C_L^2 \max(\widehat{C}_1, \widehat{C}_2)\varepsilon^2.$$

The following lemma is only a slight perturbation of Lemma 6 in [1]. A proof is therefore omitted.

Lemma 11. *Suppose that $A^{\varepsilon,h}$ and $B^{\varepsilon,h}$ are families of random variables determined by scaling parameters ε and h. Further, suppose that there are $C_1 > 0, C_2 > 0$ and $C_3 > 0$ such that for all $\varepsilon > 0$ the following three conditions hold:*

  1. $\mathsf{Var}(A^{\varepsilon,h}) \leq C_1\varepsilon^2$ uniformly in h.

  2. $|A^{\varepsilon,h}| \leq C_2$ uniformly in h.

  3. $|\mathbb{E}[B^{\varepsilon,h}]| \leq C_3 h$.

Then
$$\mathsf{Var}(A^{\varepsilon,h} B^{\varepsilon,h}) \leq 3C_3^2 C_1 h^2 \varepsilon^2 + 15 C_2^2 \mathsf{Var}(B^{\varepsilon,h}).$$

The following lemma is standard, but is included for completeness.

Lemma 12. *Let $f : \mathbb{R}^d \rightarrow \mathbb{R}$ have continuous first derivative. Then, for any $x, y \in \mathbb{R}^d$,*

$$f(x) = f(y) + \int_0^1 \nabla f(sx + (1-s)y)ds \cdot (x - y).$$

References

[1] D. F. ANDERSON, D. J. HIGHAM, AND Y. SUN, *Complexity analysis of multilevel Monte Carlo tau-leaping.* Accepted for publication to SIAM Journal on Numerical Analysis, 2014.

[2] D. F. ANDERSON AND T. G. KURTZ, *Continuous time Markov chain models for chemical reaction networks*, in Design and Analysis of Biomolecular Circuits: Engineering Approaches to Systems and Synthetic Biology, H. Koeppl, D. Densmore, G. Setti, and M. di Bernardo, eds., Springer, 2011, pp. 3–42.

[3] G. DENK AND R. WINKLER, *Modelling and simulation of transient noise in circuit simulation*, Mathematical and Computer Modelling of Dynamical Systems, 13 (2007), pp. 383–394.

[4] A. A. FAISAL, J. A. WHITE, AND S. B. LAUGHLIN, *Ion-channel noise places limits on the miniaturization of the brain's wiring*, Current Biology, 15 (2005), pp. 1143–1149.

[5] M. GILES, *Multilevel Monte Carlo path simulation*, Operations Research, 56 (2008), pp. 607–617.

[6] M. Hutzenthaler, A. Jentzen, and P. E. Kloeden, *Strong convergence of an explicit numerical method for SDEs with nonglobally Lipschitz continuous coefficients*, The Annals of Applied Probability, 22 (2012), pp. 1611–1641.

[7] P. E. Kloeden and E. Platen, *Numerical Solution of Stochastic Differential Equations*, vol. 23 of Applications of Mathematics (New York), Springer-Verlag, Berlin, 1992.

[8] X. Mao, *Stochastic Differential Equations and Their Applications*, Horwood Publishing Ltd., Chichester, 1997.

[9] G. N. Milstein and M. V. Tretyakov, *Mean-square numerical methods for stochastic differential equations with small noises*, SIAM J. Sci. Comput., 18 (1997), pp. 1067–1087.

[10] ——, *Numerical methods in the weak sense for stochastic differential equations with small noise*, SIAM Journal on Numerical Analysis, 34 (1997), pp. 2142–2167.