

## MULTILEVEL MONTE CARLO FOR STOCHASTIC DIFFERENTIAL EQUATIONS WITH SMALL NOISE\*

DAVID F. ANDERSON<sup>†</sup>, DESMOND J. HIGHAM<sup>‡</sup>, AND YU SUN<sup>†</sup>

**Abstract.** We consider the problem of numerically estimating expectations of solutions to stochastic differential equations driven by Brownian motions in the commonly occurring 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. Under the assumptions we make on the underlying model, the multilevel method combined with Euler–Maruyama is often found to be the most efficient option. Moreover, under a wide range of scalings the multilevel method is found to give the same asymptotic complexity that would arise in the idealized case where we have access to exact samples of the required distribution at a cost of  $O(1)$  per sample. A key step in our analysis is to analyze the variance between two coupled paths directly, as opposed to their  $L^2$  distance. Careful simulations are provided to illustrate the asymptotic results.

**Key words.** multilevel Monte Carlo, small noise, stochastic differential equations, Monte Carlo

**AMS subject classifications.** 60H35, 65C05, 65C30

**DOI.** 10.1137/15M1024664

**1. Introduction.** In many modeling and simulation contexts it has proved useful to parametrize the diffusion coefficient of a stochastic differential equation (SDE) and study the small noise case. In particular, diffusion and linear noise approximations to jump processes arise naturally under the “thermodynamic limit” in biochemistry and cell biology [4, 5, 12, 13]. Researchers in econometrics and finance may represent market microstructure noise as small scale diffusion, and the task of calibrating model parameters then gives rise to small noise SDE simulations; see, for example, [8, 31], with a more general overview in [25]. In computational fluid dynamics, small noise SDEs are used as a means to incorporate thermal fluctuations into traditional models in the “weak fluctuation regime” [9, section V]. In several other application areas, including ecology, circuit simulation, microbiology, neuroscience, and population dynamics [6, 10, 19, 24, 27, 28, 29, 30], the small noise limit is of interest from the perspective of understanding properties of physical models. The small noise regime has also been investigated as a means to validate conclusions drawn from analytical or heuristic arguments, especially with regard to long-time stability properties [16, 26].

From the perspective of computer simulation, many customized numerical methods have been developed for small noise SDEs with the aim of improving efficiency by exploiting the structure; see [22, Chapter 3] for an overview. In this work, we focus on the problem of numerically estimating expectations of solutions to small noise SDEs via Monte Carlo and multilevel Monte Carlo methods. In particular, we show

---

\*Received by the editors June 8, 2015; accepted for publication (in revised form) December 23, 2015; published electronically March 3, 2016.

<http://www.siam.org/journals/sinum/54-2/M102466.html>

<sup>†</sup>Department of Mathematics, University of Wisconsin, Madison, WI 53706 (anderson@math.wisc.edu, ysun@math.wisc.edu). The research of these authors was supported by grant NSF-DMS-1318832. The research of the first author was also supported by Army Research Office grant W911NF-14-1-0401.

<sup>‡</sup>Department of Mathematics and Statistics, University of Strathclyde, Glasgow G1 1XH, UK (d.j.higham@maths.strath.ac.uk). The research of this author was supported by a Royal Society/Wolfson Research Merit Award.

that under a range of scalings the standard Euler–Maruyama method combined with the usual multilevel Monte Carlo method of Giles [11] yields the same complexity that would arise if we had access to exact samples of the required distribution at a cost of  $O(1)$  per sample. So, in this well-defined setting, customized methods are not necessary.

Let  $(\Omega, \mathcal{F}, \{\mathcal{F}_t\}_{t \geq 0}, P)$  be a filtered probability space satisfying the usual conditions (it is complete and right-continuous), and let  $W(t) = (W_1(t), W_2(t), \dots, W_m(t))$  be an  $m$ -dimensional standard Wiener process under  $\{\mathcal{F}_t\}_{t \geq 0}$ . Let  $\varepsilon \in (0, 1)$  be a small parameter and let  $D^\varepsilon$  be the solution to the Itô SDE

$$(1.1) \quad D^\varepsilon(t) = D(0) + \int_0^t \mu(D^\varepsilon(s))ds + \varepsilon \int_0^t \sigma(D^\varepsilon(s))dW(s),$$

where  $\mu : \mathbb{R}^d \rightarrow \mathbb{R}^d$  and  $\sigma : \mathbb{R}^d \rightarrow \mathbb{R}^{d \times m}$  are continuous functions satisfying further assumptions detailed below.

Let  $f : \mathbb{R}^d \rightarrow \mathbb{R}$  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 [20, 21], and (ii) multilevel Monte Carlo methods combined with Euler–Maruyama [11]. We will show that in the small noise setting the  $L^2$  bounds on the difference between exact and approximate processes that are already in the literature [20] do not provide sharp estimates for the variance between two coupled paths; an analogous issue was previously addressed in the jump process setting [3]. Our main effort is therefore directed at analyzing the variance between two coupled paths in the small noise setting.

To get a feel for the best possible result, we note that in the idealized case where realizations of  $f(D^\varepsilon(T))$  could be generated with a single numerical calculation, and where  $\text{Var}(f(D^\varepsilon(T))) = O(\varepsilon^2)$ , then the computational complexity of solving the problem via Monte Carlo would be  $O(\varepsilon^2 \delta^{-2} + 1)$ , where the “+1” recognizes the fact that at least one realization must be produced. We show in this work that when  $\delta \geq e^{-\frac{1}{\varepsilon}}$  the multilevel Monte Carlo method of Giles [11] combined with a standard implementation of Euler–Maruyama solves the problem with a computational complexity of  $O(\varepsilon^2 \delta^{-2} + \delta^{-1})$ , which is the same as in the idealized case when  $\delta \leq \varepsilon^2$ , and is otherwise equal to the complexity of the standard Euler method applied to an ordinary differential equation. We will show that when  $\delta < e^{-\frac{1}{\varepsilon}}$  the multilevel Monte Carlo method combined with Euler–Maruyama solves the problem with a complexity of  $O(\varepsilon^4 \delta^{-2} \log(1/\delta)^2)$ .

We also demonstrate below that when  $\delta < \varepsilon^2$ , methods customized to the small noise setting combined with standard Monte Carlo can sometimes be more efficient than the multilevel Monte Carlo method combined with standard Euler–Maruyama. This occurs because in the regime  $\delta < \varepsilon^2$  the majority of the required work falls on accurately computing the drift in (1.1), and not due to the randomness of the process.

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:*

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

and

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

Under the above assumptions, the SDE (1.1) is known to have a unique strong solution (see, for example, Theorem 3.1 on page 51 in [18]). We also note that when these assumptions are violated the multilevel Monte Carlo method may fail, but the performance can be recovered by modifying the Euler–Maruyama discretization [14].

**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  $D_h^\varepsilon$  be the solution to

$$(1.2) \quad D_h^\varepsilon(t) = D(0) + \int_0^t \mu(D_h^\varepsilon(\eta_h(s))) ds + \varepsilon \int_0^t \sigma(D_h^\varepsilon(\eta_h(s))) dW(s),$$

where  $\eta_h(s) \stackrel{\text{def}}{=} \lfloor s/h \rfloor h$  for  $s \geq 0$ . It is straightforward to see that the solution to (1.2) 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 [17].

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 of this paper. The result bounds the variance between two coupled process; both are generated via (1.2), though they have different time discretization parameters. See the beginning of section 3 for more details related to the coupling.

**THEOREM 1.1.** *Suppose the functions  $\mu$  and  $\sigma$  satisfy our running assumptions and that  $T > 0$  and  $\varepsilon \in (0, 1)$ . Suppose further that  $D_{h_\ell}^\varepsilon(t)$  and  $D_{h_{\ell-1}}^\varepsilon(t)$  satisfy (1.2) with time discretization parameters  $h_\ell = T \cdot M^{-\ell}$  and  $h_{\ell-1} = T \cdot M^{-(\ell-1)}$ , respectively, where  $M \geq 2$  is a positive integer, and that these two processes are constructed with the same realization of Brownian motions. Assume that  $f : \mathbb{R}^d \rightarrow \mathbb{R}$  has continuous second derivative and there exists a constant  $C_L$  such that*

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

Then, for  $\ell \geq 1$ ,

$$(1.3) \quad \max_{0 \leq n \leq M^{\ell-1}} \text{Var}(f(D_{h_\ell}^\varepsilon(t_n)) - f(D_{h_{\ell-1}}^\varepsilon(t_n))) \leq \bar{C}_1 h_{\ell-1}^2 \varepsilon^2 + \bar{C}_2 h_{\ell-1} \varepsilon^4,$$

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

In the context of analyzing the classical mean-squared error, it was shown by Milstein and Tretyakov in [20] that under the same assumptions as in Theorem 1.1,

$$(1.4) \quad \mathbb{E}[|f(D^\varepsilon(T)) - f(D_h^\varepsilon(T))|^2] = O(h^2 + h\varepsilon^4),$$

where  $D^\varepsilon$  is the solution to (1.1). We note that the  $O(h^2)$  term cannot be avoided when we analyze the mean-squared error because the underlying deterministic Euler method is first order. From the mean-squared error bound (1.4) we can deduce that for some  $C_1, C_2 > 0$ , we have  $\max_{0 \leq n \leq M^{\ell-1}} \text{Var}(f(D_{h_\ell}^\varepsilon(t_n)) - f(D_{h_{\ell-1}}^\varepsilon(t_n))) \leq C_1 h_{\ell-1}^2 + C_2 h_{\ell-1} \varepsilon^4$ , where, again,  $t_n = n \cdot h_{\ell-1}$ . Theorem 1.1 sharpens this bound considerably, showing that the overall variance scales favorably with  $\varepsilon$ , 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  $D_h^\varepsilon$  is generated by a numerical scheme (not necessarily (1.2)) for which the bias of the discretization method satisfies

$$(2.1) \quad |\mathbb{E}[f(D_h^\varepsilon(T))] - \mathbb{E}[f(D^\varepsilon(T))]| = O(h^p + \varepsilon^r h^q),$$

where  $q \leq p$  and  $r \geq 0$  (see [21], where some such methods are provided). In order to ensure that the bias (2.1) is of order  $\delta$ , we require that

$$(2.2) \quad h = O(\min(\delta^{1/p}, \delta^{1/q} \varepsilon^{-r/q})).$$

Under our running assumptions and assuming that Lemma 3.2 below, which applies to Euler–Maruyama, holds for these customized methods, we find

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

where  $z_h$  is the Euler solution to the associated deterministic model obtained when  $\varepsilon$  is set to 0 in (1.1); see (3.2). 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  $i$ th independent realization of the process  $D_h^\varepsilon$ , has a variance that is  $O(n^{-1} \varepsilon^2)$ . To produce an overall estimator variance of  $O(\delta^2)$ , we require that  $n = O(\varepsilon^2 \delta^{-2} + 1)$ , where the “+1” captures the requirement that at least one path must be generated. 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

$$(2.3) \quad O((\varepsilon^2 \delta^{-2} + 1)h^{-1}) = O\left(\frac{\varepsilon^2 \delta^{-2} + 1}{\min(\delta^{1/p}, \delta^{1/q} \varepsilon^{-r/q})}\right).$$

For example, with the Euler–Maruyama scheme (1.2) we have that  $p = q = 1, r = 0$ , yielding a bias of  $O(h)$  in (2.1). In this case we select  $h = O(\delta)$  and find a computational complexity of  $O(\varepsilon^2 \delta^{-3} + \delta^{-1})$ .

To see how a customized method may be beneficial, consider the case  $\delta = O(\varepsilon^\rho)$  for  $\rho \in (1, 2)$ . We may select a method with  $p = 2, r = 2, q = 1$  (see section 5 of [21]), in which case (2.2) gives  $h = O(\varepsilon^{\rho/2})$ , and (2.3) yields a computational complexity of order  $O(\varepsilon^{2-\frac{5}{2}\rho})$ ; see subsection 2.3 for further details.

**2.2. Euler-based multilevel Monte Carlo.** Here we specify and analyze an Euler–Maruyama-based multilevel Monte Carlo method. We follow the original framework of Giles [11].

For a fixed positive integer  $M \geq 2$  we let  $h_\ell = T \cdot M^{-\ell}$  for  $\ell \in \{0, \dots, L\}$ . Reasonable choices for  $M$  include  $M \in \{2, 3, 4, 5, 6, 7\}$ , and  $L$  is determined below. For each  $\ell \in \{0, 1, \dots, L\}$ , let  $D_{h_\ell}^\varepsilon$  denote the approximate process generated by (1.2) with a step size of  $h_\ell$ . 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 depending only upon  $h_L$ . As mentioned in [21], 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  $L = O(\log(1/\delta))$ . We now let

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

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

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

which is the usual multilevel Monte Carlo estimator [11]. Set

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

By Theorem 1.1, we have  $\delta_{\varepsilon, \ell} = O(h_\ell^2 \varepsilon^2 + h_\ell \varepsilon^4)$  under a wide array of circumstances. Also note that  $\delta_{\varepsilon, 0} = \text{Var}(f(D_0^\varepsilon)) = O(\varepsilon^2)$ .

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

$$(2.4) \quad \underset{n_\ell}{\text{minimize}} \quad \sum_{\ell=0}^L n_\ell C_\ell,$$

$$(2.5) \quad \text{subject to} \quad \sum_{\ell=0}^L \frac{\delta_{\varepsilon, \ell}}{n_\ell} = \delta^2.$$

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

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

By taking a derivative with respect to  $n_\ell$  we obtain

$$(2.6) \quad n_\ell = \sqrt{\frac{\lambda}{K} \delta_{\varepsilon, \ell} h_\ell} \quad \text{for } \ell \in \{0, 1, 2, \dots, L\}$$

for some  $\lambda \geq 0$ . Plugging (2.6) into (2.5) yields

$$(2.7) \quad \sum_{\ell=0}^L \sqrt{\frac{\delta_{\varepsilon, \ell}}{h_\ell}} = \sqrt{\frac{\lambda}{K}} \cdot \delta^2,$$

and hence by Theorem 1.1 there is a  $C > 0$  for which

$$(2.8) \quad \sqrt{\frac{\lambda}{K}} = \delta^{-2} \sum_{\ell=0}^L \sqrt{\frac{\delta_{\varepsilon, \ell}}{h_\ell}} \leq C \delta^{-2} \sum_{\ell=0}^L \sqrt{h_\ell \varepsilon^2 + \varepsilon^4} \leq \tilde{C} \delta^{-2} (\varepsilon + \varepsilon^2 L),$$

where in the final inequality we used that  $\sqrt{a+b} \leq \sqrt{a} + \sqrt{b}$  for nonnegative  $a, b$ , and  $\tilde{C}$  is a new constant. Recall that  $L = O(\log(1/\delta))$ . Hence, if  $\delta \geq e^{-\frac{1}{\varepsilon}}$ , which is equivalent to  $\varepsilon^2 L \leq \varepsilon$ , then  $\frac{\lambda}{K} = O(\delta^{-4}\varepsilon^2)$ . Plugging this back into (2.6), and recognizing that we must have  $n_\ell \geq 1$ , yields

$$n_\ell = O(\delta^{-2}\varepsilon\sqrt{\delta_{\varepsilon,\ell}h_\ell} + 1).$$

Thus, we see that in this case of  $\delta \geq e^{-\frac{1}{\varepsilon}}$  the overall computational complexity is

$$(2.9) \quad \sum_{\ell=0}^L n_\ell K \cdot h_\ell^{-1} = O\left(\sum_{\ell=0}^L \delta^{-2}\varepsilon\sqrt{\delta_{\varepsilon,\ell}h_\ell^{-1}} + \sum_{\ell=0}^L h_\ell^{-1}\right) = O(\varepsilon^2\delta^{-2} + \delta^{-1}).$$

If  $\delta < e^{-\frac{1}{\varepsilon}}$ , in which case  $\varepsilon^2 L > \varepsilon$ , then  $\frac{\lambda}{K} = O(\delta^{-4}\varepsilon^4 L^2)$ , and (2.6) yields

$$n_\ell = O\left(\varepsilon^2\delta^{-2}L\sqrt{\delta_{\varepsilon,\ell}h_\ell}\right).$$

Since  $\delta^{-2}\varepsilon^2L\sqrt{\delta_{\varepsilon,\ell}h_\ell} \geq 0.89$  when  $\delta < e^{-\frac{1}{\varepsilon}}$ , we see that the usual “+1” term is not necessary in the expression above. We may now conclude that the overall computational complexity under the assumption  $\delta < e^{-\frac{1}{\varepsilon}}$  is

$$(2.10) \quad \sum_{\ell=0}^L n_\ell K \cdot h_\ell^{-1} = O(\varepsilon^4\delta^{-2}\log(1/\delta)^2).$$

**2.3. Comparisons.** There are multiple scaling regimes to consider. We begin with  $\delta < e^{-\frac{1}{\varepsilon}}$ , which represents a severe accuracy requirement. Under this assumption, the computational complexity of multilevel Monte Carlo with Euler–Maruyama is given by (2.10), whereas the complexity (2.3) required for methods tailored to the small noise setting is

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

Hence, so long as

$$(2.11) \quad \varepsilon^2 \log(1/\delta)^2 < \delta^{-\frac{1}{p}} + \delta^{-\frac{1}{q}}\varepsilon^{r/q},$$

multilevel Monte Carlo combined with Euler–Maruyama is most efficient, and there is no need to utilize customized methods. To get a sense of the restriction (2.11), we note that if  $p = 2$ , then (2.11) holds so long as  $\varepsilon < 0.65$ , and if  $p = 4$ , then (2.11) holds so long as  $\varepsilon < 0.33$ . In fact, under the further assumption that  $\delta \approx e^{-\frac{1}{\varepsilon}}$  we see that (2.10) is  $O(\varepsilon^2\delta^{-2})$ , the same—asymptotically in the parameters  $\delta$  or  $\varepsilon$ —as in the situation where we can generate independent realizations of  $f(D^\varepsilon(T))$  exactly in a single step. As  $\delta$  decreases below this threshold, the ratio between (2.10) and the complexity in the idealized setting considered in the introduction grows like  $\log(1/\delta)^2$ , as is common in the multilevel setting:

$$\frac{\varepsilon^4\delta^{-2}\log(1/\delta)^2}{\varepsilon^2\delta^{-2}} = \varepsilon^2 \log(1/\delta)^2.$$

Turning to the case  $\delta \geq e^{-\frac{1}{\varepsilon}}$ , there are two relevant subcases to consider. First, in the regime  $\delta \leq \varepsilon^2$  we have  $\varepsilon^2\delta^{-2} \geq \delta^{-1}$  and the complexity (2.9) is of order  $O(\varepsilon^2\delta^{-2})$ . This bound compares favorably with the bound  $O(\varepsilon^2\delta^{-3})$  that we derived

in subsection 2.2 for standard Monte Carlo with Euler–Maruyama and allows us to carry through a conclusion that applies to general SDEs [11]: multilevel Monte Carlo can improve on the complexity of standard Monte Carlo by a factor  $\delta^{-1}$ , where  $\delta$  is the required accuracy. Moreover, and still under the assumption that  $\delta \leq \varepsilon^2$ , the complexity  $O(\varepsilon^2\delta^{-2})$  is uniformly superior to the complexity (2.3) required for methods tailored to the small noise setting. Hence, we may conclude that when  $\delta \leq \varepsilon^2$ , there is no need to use such tailored methods. Finally, following the discussion in section 1 and in the paragraph above, we note that this multilevel Euler computational complexity is the same—asymptotically in the parameters  $\delta$  or  $\varepsilon$ —as in the situation where we can generate independent realizations of  $f(D^\varepsilon(T))$  exactly in a single step.

The last case to consider is  $\delta > \varepsilon^2$ . Now the complexity (2.9) is of order  $O(\delta^{-1})$ , the same as Euler’s method applied to an ordinary differential equation. In this case, well-selected customized methods can be asymptotically more efficient than multilevel Monte Carlo combined with standard Euler–Maruyama. For example, and following the discussion at the end of section 2.1, if  $\delta = \varepsilon^\rho$  for some  $\rho \in (1, 2)$ , then the multilevel method with Euler–Maruyama requires a complexity of order  $O(\varepsilon^{-\rho})$ . However, a customized method with  $p = 2, r = 2, q = 1$  requires a complexity of order  $O(\varepsilon^{2-\frac{5}{3}\rho})$ . Hence, the customized method is superior when  $\rho \in (1, \frac{4}{3})$ .

Finally, it is tempting to think that the computational complexity of the multilevel scheme found above can be heuristically derived in the following manner. Start with a continuous time Markov chain model which satisfies a scaling so that (1.1) is a natural diffusion approximation of the jump process. Next, use the results of [3], which are related to the variance between two coupled paths of the jump process, to infer the proper scaling in the diffusive regime. Somewhat surprisingly, this heuristic does not work and leads to overly pessimistic results. We delay a deeper discussion of this issue until section 4.1, where we address this issue both analytically and computationally.

**3. Proof of Theorem 1.1.** Throughout this section, we assume the conditions of Theorem 1.1 are met with positive integer  $M$  fixed.

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

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

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

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

$$(3.1) \quad D_{h_\ell}^\varepsilon(t_n^{k+1}) = D_{h_\ell}^\varepsilon(t_n^k) + \mu(D_{h_\ell}^\varepsilon(t_n^k))h_\ell + \varepsilon\sqrt{h_\ell}\sigma(D_{h_\ell}^\varepsilon(t_n^k))W_n^k,$$

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 (3.1) 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. Some of the proofs can be found in the online supplementary material.

LEMMA 3.1. *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.* See the online supplementary material. □

Let  $z_h$  be the deterministic solution to

$$(3.2) \quad z_h(t) = D(0) + \int_0^t \mu(z_h(\eta_h(s)))ds,$$

which is an Euler approximation to the ordinary differential equation obtained by setting  $\varepsilon = 0$  in (1.1).

LEMMA 3.2. *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.* See the online supplementary material. □

LEMMA 3.3.

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

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

*Proof.* Iterating (3.1) yields

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

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 3.1 and recalling that  $k \leq M$ . □



LEMMA 3.4.

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

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

*Proof.* Iterating (3.1) yields

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

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

$$\begin{aligned} & \mathbb{E}[|D_{h_\ell}^\varepsilon(t_n^k) - D_{h_\ell}^\varepsilon(t_n)|^4] \\ & \leq 8M^3 \sum_{i=0}^{k-1} \mathbb{E}[|\mu(D_{h_\ell}^\varepsilon(t_n^i)) h_\ell|^4] + 8\varepsilon^4 h_\ell^2 \mathbb{E} \left[ \left| \sum_{i=0}^{k-1} \sigma(D_{h_\ell}^\varepsilon(t_n^i)) W_n^i \right|^4 \right] \\ & \leq C(a, b, T, D(0)) M^4 h_\ell^4 + 2048 \varepsilon^4 h_\ell^2 \left( \sum_{i=0}^{M-1} \sum_{j=1}^m \|\sigma^j(D_{h_\ell}^\varepsilon(t_n^i))\|_{L^4(\Omega, \mathbb{R}^d)}^2 \right)^2 \\ & \leq C(a, b, T, D(0))^4 M^4 h_\ell^4 + 2048 b \varepsilon^4 M^2 h_\ell^2 m^2 (2 + 2 \max_{0 \leq i \leq M-1} \|D_{h_\ell}^\varepsilon(t_n^i)\|_{L^4(\Omega, \mathbb{R}^d)}^4) \\ & \leq C(a, b, T, D(0))^4 M^4 h_\ell^4 + C_2 \varepsilon^4 M^2 h_\ell^2, \end{aligned}$$

where the second inequality follows from Lemma 3.1 and [15, Lemma 3.8], the last inequality follows from Lemma 3.1, and  $C_1$  and  $C_2$  are constants depending only on  $a, b, T, m$ , and  $D(0)$ .  $\square$

The following is a Taylor expansion of the drift coefficient.

LEMMA 3.5. *Let  $\mu_i(x)$  be the  $i$ th component of  $\mu(x)$ . Then*

$$(3.3) \quad \mu_i(D_{h_\ell}^\varepsilon(t_n^k)) - \mu_i(D_{h_\ell}^\varepsilon(t_n)) = A_k + B_k + E_k,$$

where

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

$$(3.4) \quad B_k := \nabla \mu_i(D_{h_\ell}^\varepsilon(t_n)) \cdot \left( \varepsilon \sqrt{h_\ell} \sum_{j=0}^{k-1} \sigma(D_{h_\ell}^\varepsilon(t_n^j)) W_n^j \right),$$

and

$$\begin{aligned} E_k := & \left( \int_0^1 \int_0^s [\nabla^2 \mu_i(D_{h_\ell}^\varepsilon(t_n) + r(D_{h_\ell}^\varepsilon(t_n^k) - D_{h_\ell}^\varepsilon(t_n))) (D_{h_\ell}^\varepsilon(t_n^k) - D_{h_\ell}^\varepsilon(t_n))] dr ds \right) \\ & \cdot \left( \varepsilon \sqrt{h_\ell} \sum_{j=0}^{k-1} \sigma(D_{h_\ell}^\varepsilon(t_n^j)) W_n^j \right). \end{aligned}$$

*Proof.* See the online supplementary material.  $\square$

The following result is similar to the  $L^2$  bound found in [20] in the case where the numerical discretization method is Euler–Maruyama.

LEMMA 3.6.

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

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

$$\begin{aligned} D_{\tilde{h}_\ell}^\varepsilon(t_{n+1}) - D_{\tilde{h}_{\ell-1}}^\varepsilon(t_{n+1}) &= D_{\tilde{h}_\ell}^\varepsilon(t_n) - D_{\tilde{h}_{\ell-1}}^\varepsilon(t_n) \\ &\quad + h_\ell \sum_{k=0}^{M-1} (\mu(D_{\tilde{h}_\ell}^\varepsilon(t_n^k)) - \mu(D_{\tilde{h}_{\ell-1}}^\varepsilon(t_n))) \\ &\quad + \varepsilon \sqrt{h_\ell} \sum_{k=0}^{M-1} (\sigma(D_{\tilde{h}_\ell}^\varepsilon(t_n^k)) - \sigma(D_{\tilde{h}_{\ell-1}}^\varepsilon(t_n))) W_n^k \\ &= D_{\tilde{h}_\ell}^\varepsilon(t_n) - D_{\tilde{h}_{\ell-1}}^\varepsilon(t_n) + h_\ell \sum_{k=0}^{M-1} (\mu(D_{\tilde{h}_\ell}^\varepsilon(t_n^k)) - \mu(D_{\tilde{h}_\ell}^\varepsilon(t_n))) \\ &\quad + h_\ell \sum_{k=0}^{M-1} (\mu(D_{\tilde{h}_\ell}^\varepsilon(t_n)) - \mu(D_{\tilde{h}_{\ell-1}}^\varepsilon(t_n))) \\ &\quad + \varepsilon \sqrt{h_\ell} \sum_{k=0}^{M-1} (\sigma(D_{\tilde{h}_\ell}^\varepsilon(t_n^k)) - \sigma(D_{\tilde{h}_\ell}^\varepsilon(t_n))) W_n^k \\ &\quad + \varepsilon \sqrt{h_\ell} \sum_{k=0}^{M-1} (\sigma(D_{\tilde{h}_\ell}^\varepsilon(t_n)) - \sigma(D_{\tilde{h}_{\ell-1}}^\varepsilon(t_n))) W_n^k, \end{aligned}$$

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

$$\begin{aligned} |D_{\tilde{h}_\ell}^\varepsilon(t_{n+1}) - D_{\tilde{h}_{\ell-1}}^\varepsilon(t_{n+1})|^2 &\leq |D_{\tilde{h}_\ell}^\varepsilon(t_n) - D_{\tilde{h}_{\ell-1}}^\varepsilon(t_n)|^2 \\ &\quad + 4h_\ell^2 \left| \sum_{k=0}^{M-1} (\mu(D_{\tilde{h}_\ell}^\varepsilon(t_n^k)) - \mu(D_{\tilde{h}_\ell}^\varepsilon(t_n))) \right|^2 \\ &\quad + 4h_\ell^2 \left| \sum_{k=0}^{M-1} (\mu(D_{\tilde{h}_\ell}^\varepsilon(t_n)) - \mu(D_{\tilde{h}_{\ell-1}}^\varepsilon(t_n))) \right|^2 \\ &\quad + 4 \left| \varepsilon \sqrt{h_\ell} \sum_{k=0}^{M-1} (\sigma(D_{\tilde{h}_\ell}^\varepsilon(t_n^k)) - \sigma(D_{\tilde{h}_\ell}^\varepsilon(t_n))) W_n^k \right|^2 \\ &\quad + 4 \left| \varepsilon \sqrt{h_\ell} \sum_{k=0}^{M-1} (\sigma(D_{\tilde{h}_\ell}^\varepsilon(t_n)) - \sigma(D_{\tilde{h}_{\ell-1}}^\varepsilon(t_n))) W_n^k \right|^2 \\ &\quad + 2h_\ell \sum_{k=0}^{M-1} \langle D_{\tilde{h}_\ell}^\varepsilon(t_n) - D_{\tilde{h}_{\ell-1}}^\varepsilon(t_n), \mu(D_{\tilde{h}_\ell}^\varepsilon(t_n^k)) - \mu(D_{\tilde{h}_\ell}^\varepsilon(t_n)) \rangle \end{aligned}$$

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

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

$$\begin{aligned}
 &\mathbb{E}[|D_{h_\ell}^\varepsilon(t_{n+1}) - D_{h_{\ell-1}}^\varepsilon(t_{n+1})|^2] \\
 &\leq \mathbb{E}[|D_{h_\ell}^\varepsilon(t_n) - D_{h_{\ell-1}}^\varepsilon(t_n)|^2] + 4Mh_\ell^2 \sum_{k=0}^{M-1} \mathbb{E}[|\mu(D_{h_\ell}^\varepsilon(t_n^k)) - \mu(D_{h_\ell}^\varepsilon(t_n))|^2] \\
 &\quad + 4Mh_\ell^2 \sum_{k=0}^{M-1} \mathbb{E}[|\mu(D_{h_\ell}^\varepsilon(t_n)) - \mu(D_{h_{\ell-1}}^\varepsilon(t_n))|^2] \\
 &\quad + 4\varepsilon^2h_\ell \sum_{k=0}^{M-1} \mathbb{E}[|(\sigma(D_{h_\ell}^\varepsilon(t_n^k)) - \sigma(D_{h_\ell}^\varepsilon(t_n)))W_n^k|^2] \\
 &\quad + 4\varepsilon^2h_\ell \sum_{k=0}^{M-1} \mathbb{E}[|(\sigma(D_{h_\ell}^\varepsilon(t_n)) - \sigma(D_{h_{\ell-1}}^\varepsilon(t_n)))W_n^k|^2] \\
 &\quad + 2h_\ell \sum_{k=0}^{M-1} \mathbb{E}[\langle D_{h_\ell}^\varepsilon(t_n) - D_{h_{\ell-1}}^\varepsilon(t_n), \mu(D_{h_\ell}^\varepsilon(t_n^k)) - \mu(D_{h_\ell}^\varepsilon(t_n)) \rangle] \\
 &\quad + 2h_\ell \sum_{k=0}^{M-1} \mathbb{E}[\langle D_{h_\ell}^\varepsilon(t_n) - D_{h_{\ell-1}}^\varepsilon(t_n), \mu(D_{h_\ell}^\varepsilon(t_n)) - \mu(D_{h_{\ell-1}}^\varepsilon(t_n)) \rangle],
 \end{aligned}$$

where we used that  $W_n^k$  is independent of  $D_{h_\ell}^\varepsilon(t_n)$ ,  $D_{h_{\ell-1}}^\varepsilon(t_n)$ , and  $D_{h_\ell}^\varepsilon(t_n^k)$ . Hence, by Lemma 3.4, there are positive constants  $C_1$  and  $C_2$  that depend only on  $a, b, T, m, D(0)$ , such that

$$\begin{aligned}
 &\mathbb{E}[|D_{h_\ell}^\varepsilon(t_{n+1}) - D_{h_{\ell-1}}^\varepsilon(t_{n+1})|^2] \\
 &\leq \mathbb{E}[|D_{h_\ell}^\varepsilon(t_n) - D_{h_{\ell-1}}^\varepsilon(t_n)|^2] + 4aC_1M^4h_\ell^4 + 4aC_2\varepsilon^2M^3h_\ell^3 \\
 &\quad + 4aMh_\ell^2 \sum_{k=0}^{M-1} \mathbb{E}[|D_{h_\ell}^\varepsilon(t_n) - D_{h_{\ell-1}}^\varepsilon(t_n)|^2] + 4bC_1\varepsilon^2M^3h_\ell^3 + 4bC_2\varepsilon^4M^2h_\ell^2 \\
 &\quad + 4b\varepsilon^2h_\ell \sum_{k=0}^{M-1} \mathbb{E}[|D_{h_\ell}^\varepsilon(t_n) - D_{h_{\ell-1}}^\varepsilon(t_n)|^2] \\
 &\quad + 2h_\ell \sum_{k=0}^{M-1} \mathbb{E}[\langle D_{h_\ell}^\varepsilon(t_n) - D_{h_{\ell-1}}^\varepsilon(t_n), \mu(D_{h_\ell}^\varepsilon(t_n^k)) - \mu(D_{h_\ell}^\varepsilon(t_n)) \rangle] \\
 &\quad + 2h_\ell\sqrt{a}M\mathbb{E}[|D_{h_\ell}^\varepsilon(t_n) - D_{h_{\ell-1}}^\varepsilon(t_n)|^2],
 \end{aligned}$$

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

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

$$\begin{aligned}
(3.5) \quad &\leq \mathbb{E}[|D_{h_\ell}^\varepsilon(t_n) - D_{h_{\ell-1}}^\varepsilon(t_n)|^2] \\
&\quad + (2\sqrt{a} + 4aMh_\ell + \varepsilon^2 4b)Mh_\ell \mathbb{E}[|D_{h_\ell}^\varepsilon(t_n) - D_{h_{\ell-1}}^\varepsilon(t_n)|^2] \\
&\quad + 4aC_1M^4h_\ell^4 + 4aC_2\varepsilon^2M^3h_\ell^3 + 4bC_1\varepsilon^2M^3h_\ell^3 + 4bC_2\varepsilon^4M^2h_\ell^2 \\
(3.6) \quad &\quad + 2h_\ell \sum_{k=0}^{M-1} \mathbb{E}[\langle D_{h_\ell}^\varepsilon(t_n) - D_{h_{\ell-1}}^\varepsilon(t_n), \mu(D_{h_\ell}^\varepsilon(t_n^k)) - \mu(D_{h_\ell}^\varepsilon(t_n)) \rangle].
\end{aligned}$$

We turn to the term (3.6). Applying Lemma 3.5, we know

$$\mu_i(D_{h_\ell}^\varepsilon(t_n^k)) - \mu_i(D_{h_\ell}^\varepsilon(t_n)) = A_k + B_k + E_k.$$

Also we notice

$$\mathbb{E}[|A_k|^2] \leq K_1M^2h_\ell^2,$$

where  $K_1$  is a constant that depends only on  $a, b, T, m, D(0)$ . Utilizing Lemmas 3.1 and 3.4,

$$\begin{aligned}
(3.7) \quad &\mathbb{E}[|E_k|^2] \leq ah_\ell\varepsilon^2 \mathbb{E} \left[ |D_{h_\ell}^\varepsilon(t_n^k) - D_{h_\ell}^\varepsilon(t_n)|^2 \sum_{j=0}^{k-1} |\sigma(D_{h_\ell}^\varepsilon(t_n^j))W_n^j|^2 \right] \\
&\leq ah_\ell\varepsilon^2 (\mathbb{E}[|D_{h_\ell}^\varepsilon(t_n^k) - D_{h_\ell}^\varepsilon(t_n)|^4])^{1/2} \left( k \sum_{j=0}^{k-1} \mathbb{E}[|\sigma(D_{h_\ell}^\varepsilon(t_n^j))W_n^j|^4] \right)^{1/2} \\
&\leq K_2M^3h_\ell^3\varepsilon^2 + K_3M^2h_\ell^2\varepsilon^4,
\end{aligned}$$

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

$$\begin{aligned}
(3.8) \quad &2h_\ell \sum_{k=0}^{M-1} \mathbb{E}[\langle D_{h_\ell}^\varepsilon(t_n) - D_{h_{\ell-1}}^\varepsilon(t_n), \mu(D_{h_\ell}^\varepsilon(t_n^k)) - \mu(D_{h_\ell}^\varepsilon(t_n)) \rangle] \\
&= 2h_\ell \sum_{k=0}^{M-1} \mathbb{E}[\langle D_{h_\ell}^\varepsilon(t_n) - D_{h_{\ell-1}}^\varepsilon(t_n), A_k \rangle] \\
&\quad + 2h_\ell \sum_{k=0}^{M-1} \mathbb{E}[\langle D_{h_\ell}^\varepsilon(t_n) - D_{h_{\ell-1}}^\varepsilon(t_n), B_k \rangle] \\
&\quad + 2h_\ell \sum_{k=0}^{M-1} \mathbb{E}[\langle D_{h_\ell}^\varepsilon(t_n) - D_{h_{\ell-1}}^\varepsilon(t_n), E_k \rangle] \\
&\leq 2Mh_\ell \mathbb{E}[|D_{h_\ell}^\varepsilon(t_n) - D_{h_{\ell-1}}^\varepsilon(t_n)|^2] + h_\ell \sum_{k=0}^{M-1} \mathbb{E}[|A_k|^2] + h_\ell \sum_{k=0}^{M-1} \mathbb{E}[|E_k|^2] \\
&\leq 2Mh_\ell \mathbb{E}[|D_{h_\ell}^\varepsilon(t_n) - D_{h_{\ell-1}}^\varepsilon(t_n)|^2] + K_1M^3h_\ell^3 + K_2M^4h_\ell^4\varepsilon^2 + K_3M^3h_\ell^3\varepsilon^4,
\end{aligned}$$

where the first inequality follows from (i) the observation that the expectation (3.8) is zero, (ii) the Cauchy–Schwarz inequality, and (iii) the inequality  $2ab \leq a^2 + b^2$ . Combining all the estimates above, we find

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

$$\begin{aligned}
 &\leq \mathbb{E} \left[ |D_{h_\ell}^\varepsilon(t_n) - D_{h_{\ell-1}}^\varepsilon(t_n)|^2 \right] \\
 &\quad + (2 + 2\sqrt{a} + 4aMh_\ell + 4b\varepsilon^2)Mh_\ell \mathbb{E} \left[ |D_{h_\ell}^\varepsilon(t_n) - D_{h_{\ell-1}}^\varepsilon(t_n)|^2 \right] \\
 &\quad + 4aC_1M^4h_\ell^4 + 4aC_2\varepsilon^2M^3h_\ell^3 + 4bC_1\varepsilon^2M^3h_\ell^3 + 4bC_2\varepsilon^4M^2h_\ell^2 \\
 &\quad + K_1M^3h_\ell^3 + K_2M^4h_\ell^4\varepsilon^2 + K_3M^3h_\ell^3\varepsilon^4.
 \end{aligned}$$

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

We are now ready to prove our main result.

*Proof of Theorem 1.1.* Following [3], we first prove the result in the case that  $f(x) = x_i$  for some  $i \in \{1, \dots, d\}$ . We have that for  $n \leq M^{\ell-1} - 1$ ,

$$\begin{aligned}
 &[D_{h_\ell}^\varepsilon(t_{n+1}) - D_{h_{\ell-1}}^\varepsilon(t_{n+1})]_i = [D_{h_\ell}^\varepsilon(t_n) - D_{h_{\ell-1}}^\varepsilon(t_n)]_i \\
 &\quad + h_\ell \sum_{k=0}^{M-1} (\mu_i(D_{h_\ell}^\varepsilon(t_n^k)) - \mu_i(D_{h_\ell}^\varepsilon(t_n))) + h_\ell \sum_{k=0}^{M-1} (\mu_i(D_{h_\ell}^\varepsilon(t_n)) - \mu_i(D_{h_{\ell-1}}^\varepsilon(t_n))) \\
 &\quad + \varepsilon \sqrt{h_\ell} \sum_{k=0}^{M-1} (\sigma_i(D_{h_\ell}^\varepsilon(t_n^k)) - \sigma_i(D_{h_\ell}^\varepsilon(t_n))) W_n^k \\
 &\quad + \varepsilon \sqrt{h_\ell} \sum_{k=0}^{M-1} (\sigma_i(D_{h_\ell}^\varepsilon(t_n)) - \sigma_i(D_{h_{\ell-1}}^\varepsilon(t_n))) W_n^k,
 \end{aligned}$$

where  $\mu_i$  is the  $i$ th component of  $\mu$  and  $\sigma_i$  is the  $i$ th row of  $\sigma$ . As a result, and after some manipulation,

$$\begin{aligned}
 &\text{Var}([D_{h_\ell}^\varepsilon(t_{n+1}) - D_{h_{\ell-1}}^\varepsilon(t_{n+1})]_i) \leq (1 + Mh_\ell) \text{Var}([D_{h_\ell}^\varepsilon(t_n) - D_{h_{\ell-1}}^\varepsilon(t_n)]_i) \\
 (3.9) \quad &\quad + 4h_\ell^2M \sum_{k=0}^{M-1} \text{Var}(\mu_i(D_{h_\ell}^\varepsilon(t_n^k)) - \mu_i(D_{h_\ell}^\varepsilon(t_n)))
 \end{aligned}$$

$$(3.10) \quad + (4Mh_\ell + 1)Mh_\ell \text{Var}(\mu_i(D_{h_\ell}^\varepsilon(t_n)) - \mu_i(D_{h_{\ell-1}}^\varepsilon(t_n)))$$

$$(3.11) \quad + 4\varepsilon^2h_\ell \sum_{k=0}^{M-1} \text{Var}((\sigma_i(D_{h_\ell}^\varepsilon(t_n^k)) - \sigma_i(D_{h_\ell}^\varepsilon(t_n)))W_n^k)$$

$$(3.12) \quad + 4\varepsilon^2h_\ell \sum_{k=0}^{M-1} \text{Var}((\sigma_i(D_{h_\ell}^\varepsilon(t_n)) - \sigma_i(D_{h_{\ell-1}}^\varepsilon(t_n)))W_n^k)$$

$$(3.13) \quad + 2\text{Cov} \left( [D_{h_\ell}^\varepsilon(t_n) - D_{h_{\ell-1}}^\varepsilon(t_n)]_i, h_\ell \sum_{k=0}^{M-1} (\mu_i(D_{h_\ell}^\varepsilon(t_n^k)) - \mu_i(D_{h_\ell}^\varepsilon(t_n))) \right).$$

We must bound each expression on the right-hand side in order to apply Gronwall's inequality. We first consider (3.11). Lemma 3.4 implies that

$$\begin{aligned}
 \sum_{k=0}^{M-1} \text{Var}((\sigma_i(D_{h_\ell}^\varepsilon(t_n^k)) - \sigma_i(D_{h_\ell}^\varepsilon(t_n)))W_n^k) &\leq \sum_{k=0}^{M-1} \mathbb{E}[|\sigma_i(D_{h_\ell}^\varepsilon(t_n^k)) - \sigma_i(D_{h_\ell}^\varepsilon(t_n))|^2] \\
 &\leq Mb(c_1M^2h_\ell^2 + c_2\varepsilon^2Mh_\ell).
 \end{aligned}$$

Similarly, by Lemma 3.6 we may bound (3.12), which also yields a dominant term,

$$\begin{aligned} & \sum_{k=0}^{M-1} \text{Var}((\sigma_i(D_{h_\ell}^\varepsilon(t_n)) - \sigma_i(D_{h_{\ell-1}}^\varepsilon(t_n)))W_n^k) \\ & \leq \sum_{k=0}^{M-1} \mathbb{E}[|\sigma_i(D_{h_\ell}^\varepsilon(t_n)) - \sigma_i(D_{h_{\ell-1}}^\varepsilon(t_n))|^2] \\ & \leq Mb(d_1 M^2 h_\ell^2 + d_2 \varepsilon^4 M h_\ell), \end{aligned}$$

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

Turning to (3.9), we have the following lemma.

LEMMA 3.7.

$$\text{Var}(\mu_i(D_{h_\ell}^\varepsilon(t_n^k)) - \mu_i(D_{h_\ell}^\varepsilon(t_n))) \leq CMh_\ell \varepsilon^2,$$

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

*Proof.* From Lemma 3 in the supplementary material (Taylor approximation), we have

$$(3.14) \quad \mu_i(D_{h_\ell}^\varepsilon(t_n^k)) - \mu_i(D_{h_\ell}^\varepsilon(t_n)) = \rho^k(t_n) \cdot (D_{h_\ell}^\varepsilon(t_n^k) - D_{h_\ell}^\varepsilon(t_n)),$$

where

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

In order to bound the right-hand side of (3.14), we will apply Lemma 2 in the supplementary material with  $A^{\varepsilon, h_\ell-1} = [\rho^k(t_n)]_j$  and  $B^{\varepsilon, h_\ell-1} = [D_{h_\ell}^\varepsilon(t_n^k) - D_{h_\ell}^\varepsilon(t_n)]_j$ . Hence, we must find appropriate bounds on these components.

We begin with  $B^{\varepsilon, h_\ell-1}$ . We use Lemmas 3.1 and 3.2 after iterating (3.1) to find

$$\begin{aligned} (3.15) \quad & \text{Var}([D_{h_\ell}^\varepsilon(t_n^k) - D_{h_\ell}^\varepsilon(t_n)]_j) \\ & \leq 2\text{Var}\left(h_\ell \sum_{r=0}^{k-1} \mu_j(D_{h_\ell}^\varepsilon(t_n^r))\right) + 2\text{Var}\left(\varepsilon \sqrt{h_\ell} \sum_{r=0}^{k-1} \sigma_j(D_{h_\ell}^\varepsilon(t_n^r))W_n^r\right) \\ & \leq 2h_\ell^2 \text{Var}\left(\sum_{r=0}^{k-1} (\mu_j(D_{h_\ell}^\varepsilon(t_n^r)) - \mu_j(z_{h_\ell}(t_n^r)))\right) + 2\varepsilon^2 h_\ell \mathbb{E}\left[\left|\sum_{r=0}^{k-1} \sigma_j(D_{h_\ell}^\varepsilon(t_n^r))W_n^r\right|^2\right] \\ & \leq C_1 M^2 h_\ell^2 \varepsilon^2 + C_2 M h_\ell \varepsilon^2, \end{aligned}$$

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

Turning to  $A^{\varepsilon, h_\ell-1}$ , we apply Lemma 1 in the supplementary material with  $X_1(s) = D_{h_\ell}^\varepsilon(s)$ ,  $X_2(s) = D_{h_\ell}^\varepsilon(\eta_{h_\ell}(s))$ ,  $x_1(s) = z_{h_\ell}(s)$ ,  $x_2(s) = z_{h_\ell}(\eta_{h_\ell}(s))$ , and  $u(x) = \nabla_j \mu_i(x)$  to obtain

$$\text{Var}([\rho^k(t_n)]_j) = \text{Var}\left(\int_0^1 [\nabla_j \mu_i(D_{h_\ell}^\varepsilon(t_n)) + r(D_{h_\ell}^\varepsilon(t_n^k) - D_{h_\ell}^\varepsilon(t_n))] dr\right) \leq K \varepsilon^2,$$

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

We may now combine Lemma 2 in the supplementary material with Lemma 3.3 to conclude

$$\begin{aligned} \text{Var}([\rho^k(t_n)]_j \cdot [D_{h_\ell}^\varepsilon(t_n^k) - D_{h_\ell}^\varepsilon(t_n)]_j) &\leq \hat{C}KM^2h_\ell^2\varepsilon^2 + 15a\text{Var}([D_{h_\ell}^\varepsilon(t_n^k) - D_{h_\ell}^\varepsilon(t_n)]_j) \\ &\leq (\hat{C}K + 15aC_1)M^2h_\ell^2\varepsilon^2 + 15C_2Mh_\ell\varepsilon^2 \leq \hat{C}_1Mh_\ell\varepsilon^2, \end{aligned}$$

where  $\hat{C}_1$  is positive and does not depend on  $\varepsilon$  and  $h_\ell$ , and we applied (3.15) in the second inequality.

Returning to (3.14), the above allows us to conclude

$$\text{Var}(\mu_i(D_{h_\ell}^\varepsilon(t_n^k)) - \mu_i(D_{h_\ell}^\varepsilon(t_n))) \leq d^2\hat{C}_1Mh_\ell\varepsilon^2. \quad \square$$

We now turn to the first term of (3.10).

LEMMA 3.8.

$$\begin{aligned} \text{Var}(\mu_i(D_{h_\ell}^\varepsilon(t_n)) - \mu_i(D_{h_{\ell-1}}^\varepsilon(t_n))) \\ \leq 15ad \sum_{j=1}^d \text{Var}([D_{h_\ell}^\varepsilon(t_n) - D_{h_{\ell-1}}^\varepsilon(t_n)]_j) + K_1M^2h_\ell^2\varepsilon^2 + K_2Mh_\ell\varepsilon^6, \end{aligned}$$

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

*Proof.* We first write

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

where

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

We will again apply Lemma 2 of the supplementary material to get the necessary bounds. Therefore, we let  $A^{\varepsilon, h} = [\rho(t_n)]_j$  and  $B^{\varepsilon, h} = [D_{h_\ell}^\varepsilon(t_n) - D_{h_{\ell-1}}^\varepsilon(t_n)]_j$ .

Letting  $X_1(s) = D_{h_\ell}^\varepsilon(s)$ ,  $X_2(s) = D_{h_{\ell-1}}^\varepsilon(s)$ ,  $x_1(s) = z_{h_\ell}(s)$ ,  $x_2(s) = z_{h_{\ell-1}}(s)$ , and  $u(x) = \nabla_j\mu_i(x)$  for an application of Lemma 1 in the supplementary material, 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 Lemma 3.6 and Lemma 2 from the supplementary material, we see there are positive constants  $K_1, K_2$  depending only on  $a, b, T, m, D(0)$ , such that

$$\begin{aligned} \text{Var}([\rho(t_n)]_j([D_{h_\ell}^\varepsilon(t_n) - D_{h_{\ell-1}}^\varepsilon(t_n)]_j)) \\ \leq K_1M^2h_\ell^2\varepsilon^2 + K_2Mh_\ell\varepsilon^6 + 15a\text{Var}([D_{h_\ell}^\varepsilon(t_n) - D_{h_{\ell-1}}^\varepsilon(t_n)]_j) \end{aligned}$$

and

$$\begin{aligned} \text{Var}(\mu_i(D_{h_\ell}^\varepsilon(t_n)) - \mu_i(D_{h_{\ell-1}}^\varepsilon(t_n))) \\ \leq 15ad \sum_{j=1}^d \text{Var}([D_{h_\ell}^\varepsilon(t_n) - D_{h_{\ell-1}}^\varepsilon(t_n)]_j) + d^2K_1M^2h_\ell^2\varepsilon^2 + d^2K_2Mh_\ell\varepsilon^6. \quad \square \end{aligned}$$

Finally, we turn to the term (3.13).

LEMMA 3.9.

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

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

*Proof.* As a result of combining (3.3) in Lemma 3.5 with

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

where we recall the definition of  $B_k$  in (3.4), we have

$$\begin{aligned} (3.16) \quad & \text{Cov} \left( [D_{h_\ell}^\varepsilon(t_n) - D_{h_{\ell-1}}^\varepsilon(t_n)]_i, h_\ell \sum_{k=0}^{M-1} (\mu_i(D_{h_\ell}^\varepsilon(t_n^k)) - \mu_i(D_{h_\ell}^\varepsilon(t_n))) \right) \\ & = \text{Cov} \left( [D_{h_\ell}^\varepsilon(t_n) - D_{h_{\ell-1}}^\varepsilon(t_n)]_i, h_\ell \sum_{k=0}^{M-1} (A_k + E_k) \right) \\ & \quad + \text{Cov} \left( [D_{h_\ell}^\varepsilon(t_n) - D_{h_{\ell-1}}^\varepsilon(t_n)]_i, h_\ell \sum_{k=0}^{M-1} B_k \right) \\ & \leq M h_\ell \text{Var}([D_{h_\ell}^\varepsilon(t_n) - D_{h_{\ell-1}}^\varepsilon(t_n)]_i) + \frac{1}{2} h_\ell \sum_{k=0}^{M-1} \text{Var}(A_k) + \frac{1}{2} h_\ell \sum_{k=0}^{M-1} \text{Var}(E_k). \end{aligned}$$

First we want to estimate  $\text{Var}(A_k)$ . Applying Lemma 2 from the supplementary material with

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

and

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

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

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



where we also use Lemma 3.2 for the last line. On the other hand, from (3.7)

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

Returning to (3.16), we see

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

Now we return to (3.9)–(3.13) and combine all the estimates above to conclude that there exist  $C_1, C_2$ , and  $C_3$  which depend only on  $a, b, T, m, d, D(0)$  such that

$$\begin{aligned} & \text{Var}([D_{h_\ell}^\varepsilon(t_{n+1}) - D_{h_{\ell-1}}^\varepsilon(t_{n+1})]_i) \\ & \leq \text{Var}[D_{h_\ell}^\varepsilon(t_n) - D_{h_{\ell-1}}^\varepsilon(t_n)]_i + C_1 M^3 h_\ell^3 \varepsilon^2 \\ & \quad + C_2 M^2 h_\ell^2 \varepsilon^4 + C_3 M h_\ell \sum_{j=1}^d \text{Var}([D_{h_\ell}^\varepsilon(t_n) - D_{h_{\ell-1}}^\varepsilon(t_n)]_j). \end{aligned}$$

Therefore,

$$\begin{aligned} & \max_{i=1,2,\dots,d} \text{Var}([D_{h_\ell}^\varepsilon(t_{n+1}) - D_{h_{\ell-1}}^\varepsilon(t_{n+1})]_i) \\ & \leq \max_{i=1,2,\dots,d} \text{Var}([D_{h_\ell}^\varepsilon(t_n) - D_{h_{\ell-1}}^\varepsilon(t_n)]_i) + C_1 M^3 h_\ell^3 \varepsilon^2 + C_2 M^2 h_\ell^2 \varepsilon^4 \\ & \quad + C_3 d M h_\ell \max_{i=1,2,\dots,d} \text{Var}([D_{h_\ell}^\varepsilon(t_n) - D_{h_{\ell-1}}^\varepsilon(t_n)]_i). \end{aligned}$$

Applying Gronwall’s lemma, we obtain

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

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

We have shown the result under the assumption that  $f(x) = x_i$ . To show the general case, note that from Lemma 3 in the supplementary material we have

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

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 1 in the supplementary material, which yields

$$\text{Var} \left( \int_0^1 [\nabla_j f(D_{h_\ell}^\varepsilon(t_n) + r(D_{h_\ell}^\varepsilon(t_n) - D_{h_{\ell-1}}^\varepsilon(t_n)))] dr \right) \leq K \varepsilon^2,$$

where  $K$  is a universal constant that depends on  $C_L, D, a, b, T, D(0)$ . Hence, by an application of Lemma 3.6, Lemma 2 from the supplementary material, and the work above, we obtain

$$\begin{aligned} \text{Var} & \left( \int_0^1 \nabla_j f(D_{h_\ell}^\varepsilon(t_n) + r(D_{h_\ell}^\varepsilon(t_n) - D_{h_\ell}^\varepsilon(t_n))) dr \cdot [D_{h_\ell}^\varepsilon(t_n) - D_{h_{\ell-1}}^\varepsilon(t_n)]_j \right) \\ & \leq K(d_1 M^2 h_\ell^2 + d_2 M h_\ell \varepsilon^4) \varepsilon^2 + 15dC_L^2 \text{Var}([D^\varepsilon(s) - D_{h_\ell}^\varepsilon(s)]_j) \\ & \leq (Kd_1 + 15dC_L^2 C_1) M^2 h_\ell^2 \varepsilon^2 + (15dC_L^2 C_2 + Kd_2) M h_\ell \varepsilon^4. \end{aligned}$$

Thus

$$\begin{aligned} \text{Var}(f(D_{h_\ell}^\varepsilon(t_n)) - f(D_{h_{\ell-1}}^\varepsilon(t_n))) & \leq d^2(Kd_1 + 15dC_L^2 C_1) M^2 h_\ell^2 \varepsilon^2 \\ & \quad + d^2(15dC_L^2 C_2 + Kd_2) h_\ell \varepsilon^4, \end{aligned}$$

giving the result. □

**4. Numerical examples and comparison with results related to jump processes.** In this section we provide numerical evidence for the sharpness of both Theorem 1.1 and the computational complexity analyses provided in sections 2.1 and 2.2. Further, we compare our results to those found in [3] for scaled Markov processes.

*Example 4.1.* 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 sharpness of the bound

$$\text{Var}(D_{h_\ell}^\varepsilon(t) - D_{h_{\ell-1}}^\varepsilon(t)) = O(h^2 \varepsilon^2 + h \varepsilon^4),$$

we fix one of  $h$  or  $\varepsilon$  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 bounds.

**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 (1.3). See Figure 1(a), where the log-log plot is consistent with the functional form

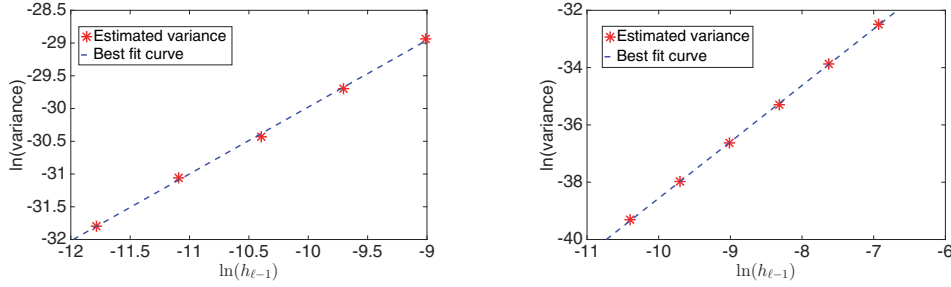
$$\text{Var}(D_{h_\ell}^\varepsilon(T) - D_{h_{\ell-1}}^\varepsilon(T)) = O(h_{\ell-1}).$$

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

$$h_{\ell-1} \in \{2^{-10}, 2^{-11}, 2^{-12}, 2^{-13}, 2^{-14}, 2^{-16}\}$$

to ensure  $h_{\ell-1} \geq \varepsilon^2$ . As a result,  $h_{\ell-1}^2 \varepsilon^2$  is likely to be the dominant term in (1.3). See Figure 1(b), where the log-log plot is consistent with the functional form

$$\text{Var}(D_{h_\ell}^\varepsilon(1) - D_{h_{\ell-1}}^\varepsilon(1)) = O(h_{\ell-1}^2).$$



(a)  $\varepsilon = 2^{-6}$  fixed while  $h$  is varied. The best fit curve is  $y = 1.02x - 19.76$ .

(b)  $\varepsilon = 2^{-10}$  fixed while  $h$  is varied. The best fit curve is  $y = 1.96x - 18.86$ .

FIG. 1. Log-log plots of  $\text{Var}(D_{h_\ell}^\varepsilon(1) - D_{h_{\ell-1}}^\varepsilon(1))$  with  $\varepsilon$  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.

**The exponent of  $\varepsilon$  in  $h\varepsilon^4$ .** We fix  $h_{\ell-1} = 2^{-19}$  and vary

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

to ensure  $h_{\ell-1} \leq \varepsilon^2$ . As a result,  $h_{\ell-1}\varepsilon^4$  is likely to be the dominant term in (1.3). See Figure 2(a), where the log-log plot is consistent with the functional form

$$\text{Var}(D_{h_\ell}^\varepsilon(1) - D_{h_{\ell-1}}^\varepsilon(1)) = O(\varepsilon^4).$$

**The exponent of  $\varepsilon$  in  $h^2\varepsilon^2$ .** We fix  $h_{\ell-1} = 2^{-9}$  and vary

$$\varepsilon \in \{2^{-6}, 2^{-7}, 2^{-8}, 2^{-9}, 2^{-10}, 2^{-11}\}$$

to ensure  $h_{\ell-1} \geq \varepsilon^2$ . As a result,  $h_{\ell-1}^2\varepsilon^2$  is likely to be the dominant term in (1.3). See Figure 2(b), where the log-log plot is consistent with the functional form

$$\text{Var}(D_{h_\ell}^\varepsilon(1) - D_{h_{\ell-1}}^\varepsilon(1)) = O(\varepsilon^2).$$

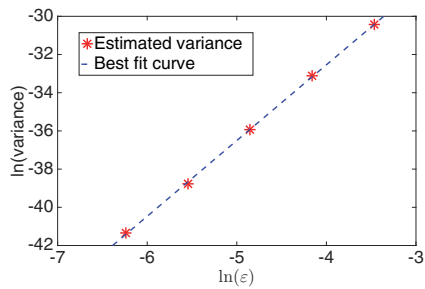
We turn to numerically demonstrating our conclusions related to the complexity of Euler-based multilevel Monte Carlo and the complexity of Euler-based standard Monte Carlo. We will measure complexity in two ways: by total number of random variables utilized and by required CPU time. Our implementation 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 sample trajectories in order to estimate  $\delta_{\varepsilon,\ell}$ , as defined in section 2.2. According to (2.6) and (2.8) we then selected

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

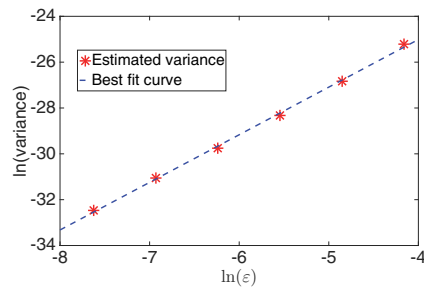
We implemented Euler’s method combined with standard Monte Carlo by selecting the number of paths by

$$N = \lceil \delta^{-2} \text{Var}(D_h^\varepsilon(1)) \rceil,$$

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

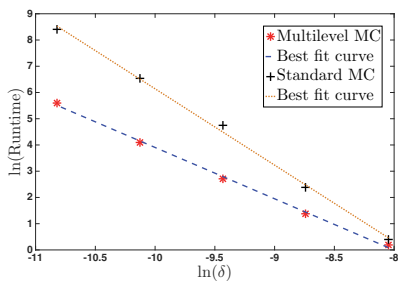


(a)  $h_{\ell-1} = 2^{-19}$  fixed while  $\varepsilon$  is varied. The best fit curve is  $y = 3.96x - 16.68$ .

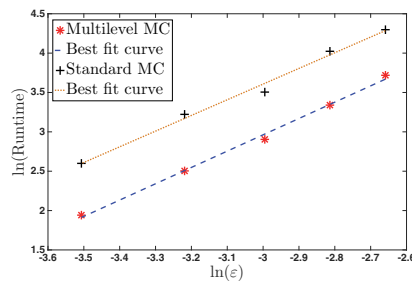


(b)  $h_{\ell-1} = 2^{-9}$  fixed while  $\varepsilon$  is varied. The best fit curve is  $y = 2.08x - 16.7$ .

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



(a)  $\varepsilon = 0.1$  held constant and  $\delta$  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  $\varepsilon$  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.

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

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

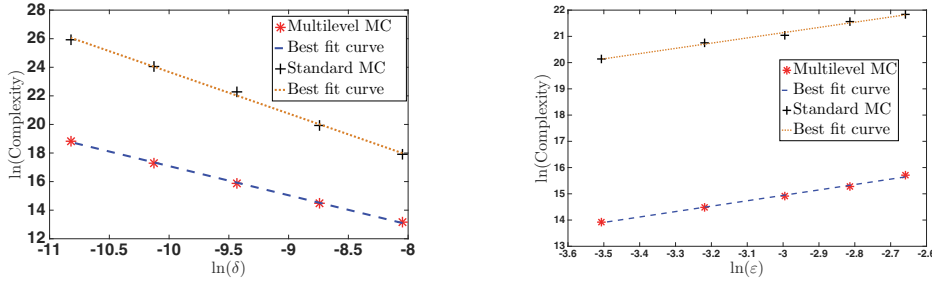
$$\delta \in \{0.00032, 0.00016, 0.00008, 0.00004, 0.00002\},$$

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 the standard Monte Carlo method is  $O(\delta^{-3})$  when  $\varepsilon$  is fixed. The Monte Carlo estimates which came from these simulations are detailed in Table 1. 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 Figures 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  $\varepsilon$  is varied as

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



(a)  $\epsilon = 0.1$  held constant and  $\delta$  varied. The best fit lines are  $y = -2.04x - 3.33$  for MLMC and  $y = -2.91x - 22.95$  for standard Monte Carlo.

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

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

TABLE 1

Result of Euler-based multilevel Monte Carlo and Euler-based Monte Carlo for fixed  $\epsilon = 0.1$  and varying  $\delta$ . The last two columns provide the standard deviations for the two estimators.

$\delta$	Mean-Euler	Mean-MLMC	SD-Euler	SD-MLMC
0.00032	0.367449	0.367944	0.000320	0.000305
0.00016	0.368028	0.367906	0.000160	0.000153
0.00008	0.367839	0.367891	0.000080	0.000077
0.00004	0.367941	0.367863	0.000040	0.000039
0.00002	0.367851	0.367883	0.000020	0.000020

which ensures  $\delta > e^{-\frac{1}{\epsilon}}$ . 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  $O(\epsilon^2)$  when  $\delta$  is fixed. The Monte Carlo estimates which came from these simulations are detailed in Table 2.  $\square$

**4.1. Comparison with results for continuous time Markov chains.** Diffusion processes with small noise structures of the form (1.1) often arise as approximations to continuous time Markov chains. Bounds on the variance between Euler approximations to such scaled jump processes can be found in [3]. Since the diffusion approximation is naturally related to the jump process model, it is tempting to believe that the analysis found in [3] can be utilized to infer the results presented in this paper. The following example and analysis shows that this intuition is incorrect.

*Example 4.2.* Consider a family of models, parametrized by  $N > 0$ , satisfying

$$(4.1) \quad X^N(t) = X^N(0) + \frac{1}{N} Y_1 \left( N \int_0^t X_1^N(s) \left( X_1^N(s) - \frac{1}{N} \right) ds \right) \begin{bmatrix} -2 \\ 1 \end{bmatrix} + \frac{1}{N} Y_2 \left( N \int_0^t X_2^N(s) ds \right) \begin{bmatrix} 2 \\ -1 \end{bmatrix}$$

with  $X^N(0) \in \frac{1}{N} \mathbb{Z}_{\geq 0}^2$  and  $Y_1, Y_2$  independent unit-rate Poisson processes. The process (4.1) can model the time evolution of the reaction network



TABLE 2

Results of Euler-based multilevel Monte Carlo and Euler-based Monte Carlo for fixed  $\delta = 2^{-14} \approx 0.000061$  and varying  $\varepsilon$ . The last two columns provide the standard deviations for the two estimators.

$\varepsilon$	Mean-Euler	Mean-MLMC	SD-Euler	SD-MLMC
0.07	0.367830	0.367834	0.000061	0.000059
0.06	0.367755	0.367920	0.000061	0.000059
0.05	0.367933	0.367819	0.000061	0.000059
0.04	0.367809	0.367856	0.000061	0.000059
0.03	0.367879	0.367925	0.000061	0.000059

in which two molecules of species  $A$  can combine to form a molecule of species  $B$ , and vice versa [4, 5]. The specific choice of scaling in (4.1) is called the *classical scaling* for biochemical processes [4, 5]. One representation for the continuous-in-time Euler–Maruyama approximation of the standard diffusion approximation to the model (4.1) is

$$\begin{aligned}
 D_h^N(t) &= D_h^N(0) + \int_0^t D_{h,1}^N(\eta_h(s))^2 ds \begin{bmatrix} -2 \\ 1 \end{bmatrix} + \int_0^t D_{h,2}^N(\eta_h(s)) ds \begin{bmatrix} 2 \\ -1 \end{bmatrix} \\
 (4.3) \quad &+ \varepsilon_N \int_0^t \sqrt{\max\{D_1^N(\eta_h(s))^2, 0\}} dW_1(s) \begin{bmatrix} -2 \\ 1 \end{bmatrix} \\
 &+ \varepsilon_N \int_0^t \sqrt{\max\{D_2^N(\eta_h(s)), 0\}} dW_2(s) \begin{bmatrix} 2 \\ -1 \end{bmatrix},
 \end{aligned}$$

where  $\varepsilon_N = \frac{1}{\sqrt{N}}$ , and  $W_1$  and  $W_2$  are independent Brownian motions [4, 5]. Let  $Z_h^N$  be an Euler approximation to (4.1), and let  $M > 0$  be some fixed positive integer. See [1] or [3] for a stochastic representation of  $Z_h^N$  that is similar to (4.3). For the relevant coupling between  $Z_{h_\ell}^N$  and  $Z_{h_{\ell-1}}^N$ , see [2, 3]. By Corollary 1 in [3], we have that for  $h_\ell = M^{-\ell}$

$$(4.4) \quad \text{Var}(Z_{h_\ell,1}^N(t) - Z_{h_{\ell-1},1}^N(t)) \leq D \cdot h_{\ell-1} N^{-1} = D \cdot h_{\ell-1} \varepsilon_N^2,$$

where the constant  $D$  does not depend upon  $N$ ,  $h_\ell$ , or  $h_{\ell-1}$ . Conversely, Theorem 1.1 allows us to conclude that

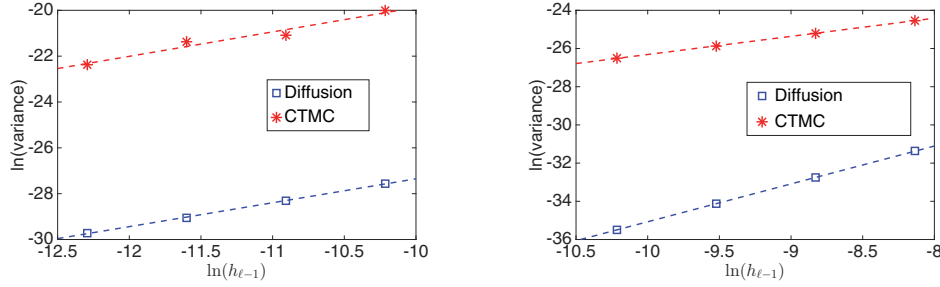
$$(4.5) \quad \text{Var}(D_{h_\ell,1}^N(t) - D_{h_{\ell-1},1}^N(t)) \leq C_1 \cdot h_{\ell-1}^2 \varepsilon_N^2 + C_2 \cdot h_{\ell-1} \varepsilon_N^4.$$

The key feature to note is that for  $h_{\ell-1} < 1$  and  $\varepsilon_N < 1$ , both of the terms  $h_{\ell-1}^2 \varepsilon_N^2$  and  $h_{\ell-1} \varepsilon_N^4$  are dominated by  $h_{\ell-1} \varepsilon_N^2$ . In fact,

$$\frac{h_{\ell-1}^2 \varepsilon_N^2 + h_{\ell-1} \varepsilon_N^4}{h_{\ell-1} \varepsilon_N^2} = h_{\ell-1} + \varepsilon_N^2,$$

showing a dramatic reduction in the variance when the coupled diffusion processes are considered, as opposed to the coupled jump processes.

In order to numerically demonstrate the bounds (4.4) and (4.5), we follow the numerical analysis performed in Example 4.1 by varying  $h_\ell$  and  $\varepsilon_N = \frac{1}{\sqrt{N}}$  in different scaling regimes in order to isolate the different possible exponents. For each of the numerical experiments performed we fixed a terminal time of  $T = 0.3$  and took an initial condition of  $(0.2, 0.2)$  for each model. As we also mentioned in Example 4.1,



(a)  $\varepsilon_N = 2^{-6}$  fixed while  $h_{\ell-1}$  is varied. The best fit curve for the data associated with the CTMC is  $y = 1.07x - 9.21$ , whereas the best fit curve for the data associated with the diffusion is  $y = 1.04x - 16.92$ .

(b)  $\varepsilon_N = 2^{-10}$  fixed while  $h_{\ell-1}$  is varied. The best fit curve for the data associated with the CTMC is  $y = 0.95x - 16.82$ , whereas the best fit curve for the data associated with the diffusion is  $y = 1.98x - 15.25$ .

FIG. 5. Log-log plots of  $\text{Var}(D_{h_{\ell},1}^N(T) - D_{h_{\ell-1},1}^N(T))$  and  $\text{Var}(Z_{h_{\ell},1}^N(T) - Z_{h_{\ell-1},1}^N(T))$  with  $\varepsilon_N$  held constant and  $h_{\ell-1}$  varied. The best fit curves for the data are overlain with dashed lines.

we emphasize that these experiments use extreme parameter choices solely for the purpose of testing the sharpness of the delicate asymptotic bounds.

**The exponent of  $h_{\ell-1}$  in  $h_{\ell-1}\varepsilon_N^4$ .** We fix  $N = 2^{12}$ , which corresponds with  $\varepsilon_N = 2^{-6}$ , and vary

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

to ensure  $h_{\ell-1} \leq \varepsilon_N^2$ . As a result,  $h_{\ell-1}\varepsilon_N^4$  is likely to be the dominant term in (4.5). See Figure 5(a), where the log-log plots are consistent with the functional forms

$$\text{Var}(D_{h_{\ell},1}^N(T) - D_{h_{\ell-1},1}^N(T)) = O(h_{\ell-1}), \quad \text{Var}(Z_{h_{\ell},1}^N(T) - Z_{h_{\ell-1},1}^N(T)) = O(h_{\ell-1}).$$

**The exponent of  $h$  in  $h^2\varepsilon^2$ .** We fix  $N = 2^{20}$ , which corresponds with  $\varepsilon_N = 2^{-10}$ , and vary

$$h_{\ell-1} \in \{2^{-10}, 2^{-11}, 2^{-12}, 2^{-13}\}$$

to ensure  $h_{\ell-1} \geq \varepsilon_N^2$ . As a result,  $h_{\ell-1}^2\varepsilon_N^2$  is likely to be the dominant term in (4.5). See Figure 5(b), where the log-log plots are consistent with the functional forms

$$\text{Var}(D_{h_{\ell},1}^N(T) - D_{h_{\ell-1},1}^N(T)) = O(h_{\ell-1}^2), \quad \text{Var}(Z_{h_{\ell},1}^N(T) - Z_{h_{\ell-1},1}^N(T)) = O(h_{\ell-1}).$$

**The exponent of  $\varepsilon$  in  $h\varepsilon^4$ .** We fix  $h_{\ell-1} = 2^{-12}$  and vary

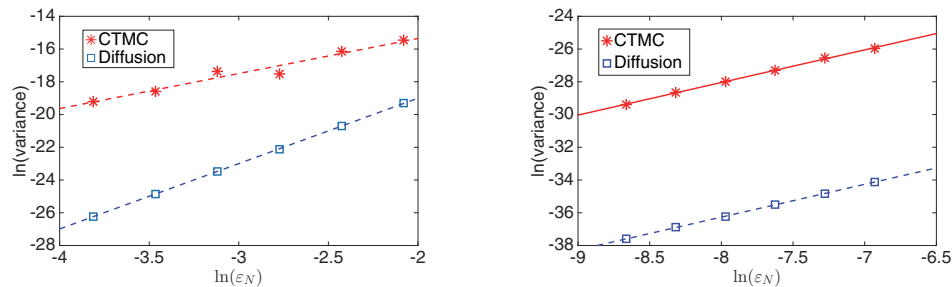
$$N^{-1} \in \{2^{-6}, 2^{-7}, 2^{-8}, 2^{-9}, 2^{-10}, 2^{-11}\}$$

to ensure  $h_{\ell-1} \leq \varepsilon_N^2 = N^{-1}$ . As a result,  $h_{\ell-1}\varepsilon_N^4$  is likely to be the dominant term in (4.5). See Figure 6(a), where the log-log plot is consistent with the functional forms

$$\text{Var}(D_{h_{\ell},1}^N(T) - D_{h_{\ell-1},1}^N(T)) = O(\varepsilon_N^4), \quad \text{Var}(Z_{h_{\ell},1}^N(T) - Z_{h_{\ell-1},1}^N(T)) = O(\varepsilon_N^2).$$

**The exponent of  $\varepsilon$  in  $h^2\varepsilon^2$ .** We fix  $h_{\ell-1} = 2^{-11}$  and vary

$$N \in \{2^{20}, 2^{21}, 2^{22}, 2^{23}, 2^{24}, 2^{25}\}$$



(a)  $h_{\ell-1} = 2^{-12}$  fixed while  $\varepsilon_N$  is varied. The best fit curve for the data associated with the CTMC is  $y = 2.14x - 11.07$ , whereas the best fit curve for the data associated with the diffusion is  $y = 3.99x - 11.02$ .

(b)  $h_{\ell-1} = 2^{-11}$  fixed while  $\varepsilon_N$  is varied. The best fit curve for the data associated with the CTMC is  $y = 1.99 - 12.11$ , whereas the best fit curve for the data associated with the diffusion is  $y = 2.00x - 20.25$ .

FIG. 6. Log-log plots of  $\text{Var}(D_{h_{\ell},1}^N(T) - D_{h_{\ell-1},1}^N(T))$  and  $\text{Var}(Z_{h_{\ell},1}^N(T) - Z_{h_{\ell-1},1}^N(T))$  with  $h_{\ell-1}$  held constant and  $\varepsilon_N$  varied. The best fit curves for all data are overlain with dashed lines.

to ensure  $h_{\ell-1} \geq \varepsilon_N^2 = N^{-1}$ . As a result,  $h_{\ell-1}^2 \varepsilon_N^2$  is likely to be the dominant term in (4.5). See Figure 6(b), where the log-log plot is consistent with the functional forms

$$\text{Var}(D_{h_{\ell},1}^N(T) - D_{h_{\ell-1},1}^N(T)) = O(\varepsilon_N^2), \quad \text{Var}(Z_{h_{\ell},1}^N(T) - Z_{h_{\ell-1},1}^N(T)) = O(\varepsilon_N^2).$$

**5. Summary.** This work focused 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. Hence, we analyzed this variance directly. Using  $\varepsilon$  to denote the small noise parameter and  $\delta$  to denote the required accuracy, we showed that when  $\delta \leq \varepsilon^2$  there is no benefit from using discretization methods that are customized for small noise. Moreover, so long as we also have  $\delta \geq e^{-\frac{1}{\varepsilon}}$ , a basic Euler–Maruyama discretization used in a multilevel setting leads to the same complexity that would arise in the idealized case where we had access to exact samples of the required distribution at a cost of  $O(1)$  per sample.

Interesting future work in this area includes the following:

- (i) Develop multilevel methods customized to the setting  $\delta > \varepsilon^2$ .
- (ii) Investigate whether the recently proposed techniques in [7] and [23] can be adapted to the small noise regime.

**Acknowledgment.** We thank a careful reviewer, whose comments allowed us to improve the presentation of this manuscript.

#### REFERENCES

- [1] D. F. ANDERSON, A. GANGULY, AND T. G. KURTZ, *Error analysis of tau-leap simulation methods*, Ann. Appl. Probab., 21 (2011), pp. 2226–2262.
- [2] D. F. ANDERSON AND D. J. HIGHAM, *Multilevel Monte Carlo for continuous time Markov chains, with applications in biochemical kinetics*, Multiscale Model. Simul., 10 (2012), pp. 146–179.
- [3] D. F. ANDERSON, D. J. HIGHAM, AND Y. SUN, *Complexity of multilevel Monte Carlo tau-leaping*, SIAM J. Numer. Anal., 52 (2014), pp. 3106–3127.
- [4] 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. Koepl, D. Densmore, G. Setti, and M. di Bernardo, eds., Springer, New York, 2011, pp. 3–42.
- [5] D. F. ANDERSON AND T. G. KURTZ, *Stochastic Analysis of Biochemical Systems*, Springer International Publishing, Cham, Switzerland, 2015.
  - [6] X. BARDINA, D. BASCOMPTE, C. ROVIRA, AND S. TINDEL, *An analysis of a stochastic model for bacteriophage systems*, *Math. Biosci.*, 241 (2013), pp. 99–108.
  - [7] D. BELOMESTNY AND T. NAGAPETYAN, *Variance Reduced Multilevel Path Simulation: Going beyond the Complexity  $\varepsilon^{-2}$* , preprint, arXiv:1412.4045, 2014.
  - [8] G. CONFORTI, S. DE MARCO, AND J.-D. DEUSCHEL, *On small-noise equations with degenerate limiting system arising from volatility models*, in *Large Deviations and Asymptotic Methods in Finance*, Springer Proceedings in Mathematics and Statistics, P. K. Friz, J. Gatheral, A. Gulisashvili, A. Jacquier, and J. Teichmann, eds., Springer Proc. Math. Stat. 110, Springer, Cham, Switzerland, 2015, pp. 473–505.
  - [9] S. DELONG, Y. SUN, B. E. GRIFFITH, E. VANDEN-ELJNDEN, AND A. DONEV, *Multiscale temporal integrators for fluctuating hydrodynamics*, *Phys. Rev. E*, 90 (2014), 063312.
  - [10] G. DENK AND R. WINKLER, *Modelling and simulation of transient noise in circuit simulation*, *Math. Comput. Model. Dyn. Syst.*, 13 (2007), pp. 383–394.
  - [11] M. B. GILES, *Multilevel Monte Carlo path simulation*, *Oper. Res.*, 56 (2008), pp. 607–617.
  - [12] D. T. GILLESPIE, *Markov Processes: An Introduction for Physical Scientists*, Academic Press, San Diego, 1991.
  - [13] D. T. GILLESPIE, *The chemical Langevin equation*, *J. Chem. Phys.*, 113 (2000), pp. 297–306.
  - [14] M. HUTZENTHALER, A. JENTZEN, AND P. E. KLOEDEN, *Divergence of the multilevel Monte Carlo method for nonlinear stochastic differential equations*, *Ann. Appl. Probab.*, 23 (2013), pp. 1913–1966.
  - [15] M. HUTZENTHALER, A. JENTZEN, AND P. E. KLOEDEN, *Strong convergence of an explicit numerical method for SDEs with nonglobally Lipschitz continuous coefficients*, *Ann. Appl. Probab.*, 22 (2012), pp. 1611–1641.
  - [16] G. G. IZÚS, R. R. DEZA, AND H. S. WIO, *Exact nonequilibrium potential for the Fitzhugh-Nagumo model in the excitable and bistable regimes*, *Phys. Rev. E*, 58 (1998), pp. 93–98.
  - [17] P. E. KLOEDEN AND E. PLATEN, *Numerical Solution of Stochastic Differential Equations*, *Appl. Math. (N.Y.)* 23, Springer-Verlag, Berlin, 1992.
  - [18] X. MAO, *Stochastic Differential Equations and Their Applications*, Horwood Publishing, Chichester, UK, 1997.
  - [19] X. MAO, G. MARION, AND E. RENSHAW, *Environmental noise suppresses explosion in population dynamics*, *Stochastic Process. Appl.*, 97 (2002), pp. 96–110.
  - [20] 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.
  - [21] G. N. MILSTEIN AND M. V. TRETYAKOV, *Numerical methods in the weak sense for stochastic differential equations with small noise*, *SIAM J. Numer. Anal.*, 34 (1997), pp. 2142–2167.
  - [22] G. N. MILSTEIN AND M. V. TRETYAKOV, *Stochastic Numerics for Mathematical Physics*, Springer-Verlag, Berlin, 2004.
  - [23] T. MÜLLER-GRONBACH AND L. YAROSLAVTSEVA, *Deterministic Quadrature Formulas for SDEs Based on Simplified Weak Ito-Taylor Steps*, Tech. Report 167, DFG-Schwerpunktprogramm 1324, 2014.
  - [24] J. OTWINOWSKI, S. TANASE-NICOLA, AND I. NEMENMAN, *Speeding up evolutionary search by small fitness fluctuations*, *J. Statist. Phys.*, 144 (2011), pp. 367–378.
  - [25] G. PAVLIOTIS AND A. M. STUART, *Parameter estimation for multiscale diffusions*, *J. Statist. Phys.*, 127 (2007), pp. 741–781.
  - [26] A. PIKOVSKY, M. ROSENBLUM, AND J. KURTHS, *Synchronization: A Universal Concept in Nonlinear Sciences*, Cambridge University Press, Cambridge, UK, 2001.
  - [27] B. A. SCHMERL AND M. D. McDONNELL, *Channel-noise-induced stochastic facilitation in an auditory brainstem neuron model*, *Phys. Rev. E*, 88 (2013), 052722.
  - [28] A. SHISHKIN AND D. POSTNOV, *Stochastic dynamics of Fitzhugh-Nagumo model near the canard explosion*, in *Proceedings of the 2003 International Conference on Physics and Control*, Vol. 2, 2003, pp. 649–653.
  - [29] M. SIEBER, H. MALCHOW, AND S. V. PETROVSKII, *Noise-induced suppression of periodic travelling waves in oscillatory reaction-diffusion systems*, *Proc. R. Soc. Lond. Ser. A Math. Phys. Eng. Sci.*, (2010), pp. 1903–1917.
  - [30] H. C. TUCKWELL, R. RODRIGUEZ, AND F. Y. M. WAN, *Determination of firing times for the stochastic Fitzhugh-Nagumo neuronal model*, *Neural Comput.*, 15 (2003), pp. 143–159.
  - [31] L. ZHANG, P. A. MYKLAND, AND Y. AÏT-SAHALIA, *A tale of two time scales: Determining integrated volatility with noisy high-frequency data*, *J. Amer. Statist. Assoc.*, 100 (2005), pp. 1394–1411.