Multilevel Monte Carlo Methods with Applications to Biochemical Models

By

Yu Sun

A dissertation submitted in partial fulfillment of the

requirements for the degree of

Doctor of Philosophy

(Mathematics)

at the

UNIVERSITY OF WISCONSIN – MADISON

2015

Date of final oral examination: December 16, 2015

The dissertation is approved by the following members of the Final Oral Committee:

David Anderson, Associate Professor, Mathematics

Gheorghe Craciun, Professor, Mathematics

Saverio Spagnolie, Assistant Professor, Mathematics

Bret Hanlon, Assistant Professor, Statistics

Benedek Valko, Associate Professor, Mathematics

Abstract

In this dissertation, we study multilevel Monte Carlo simulation methods and applications to different types of stochastic dynamical models. This dissertation focuses on three selected topics. Firstly, tau-leaping is a popular discretization method for generating approximate paths of continuous time, discrete space Markov chains, notably for biochemical reaction systems. To compute expected values in this context, an appropriate multilevel Monte Carlo form of tau-leaping has been shown to improve efficiency dramatically. The contribution for this topic is deriving new analytic results concerning the computational complexity of multilevel Monte Carlo tau-leaping that are significantly sharper than previous ones. The key feature of the analysis that allows for the sharper bounds is that when comparing relevant pairs of processes we analyze the variance of their difference directly rather than bounding via the second moment. Secondly, we consider the problem of numerically estimating expectations of solutions to stochastic differential equations driven by Brownian motions in the small noise regime. Our complexity analysis shows multilevel Monte Carlo can improve on the complexity of standard Monte Carlo by a factor ε, where ε is the desired accuracy. The take-home message is that, under reasonable assumptions, a basic Euler-Maruyama discretization leads to optimal asymptotic computational complexity when used in a multilevel setting. Thirdly, we analyze and compare the computational complexity of different simulation strategies for classically scaled continuous time Markov chains. We provide numerical examples demonstrating our main conclusions. The three topics have appeared in a series of three papers with one published [4], two submitted [6, 7].

Acknowledgements

I would like to thank all the people who have helped me, encouraged me during my Ph.D study.

First of all, I would like to thank my advisor, Prof. David F. Anderson for his guidance and patience in my Ph.D study and research.

Besides, I would like to thank the rest of my thesis committee: Prof. Gheorghe Craciun, Prof. Saverio Spagnolie, Prof. Bret Hanlon and Prof. Benedek Valko for their valuable comments on my thesis.

Last but not the least, I'm deeply grateful to my family who have absolutely supported my pursuit of Ph.D study aboard. Also I want to thank all my friends in Madison for their kindness.

List of Figures

List of Tables

Contents

Chapter 1

Introduction

1.1 Background

Stochastic Models have been gaining in popularity over the past few decades and are now used widely in various scientific fields. Growth of their use in the biological sciences has been particularly swift. In many of important circumstances stochastic models are being used in preference to deterministic ordinary differential equation models to describe biochemical reaction networks [34, 70, 68]. These stochastic models have given important insight into the behavior of biological systems, including gene regulatory systems, metabolism, and evolution [18, 53, 48, 49].

Numerical estimation of expectations of stochastically modeled biochemical reaction systems are the main focus of this thesis. Due to the degree of interactions in living cells, numerical simulation is often considered the only way to analyze a given system. A number of numerical methods have been developed to simulate biochemical reaction networks. Gillespie is known as a pioneer of this field for the development of the stochastic simulation algorithm (SSA) [27, 28] and the tau-leaping algorithm [31]. The SSA can be used to produce exact sample paths of stochastically modeled biochemical reaction systems and it is an unbiased method. Unfortunately in many realistic contexts the computational cost of performing standard Monte Carlo with SSA is still prohibitive

because many reactions can take place over time periods. Tau-leaping was developed to overcome the problem of long run time. Tau-leaping generates Euler-style approximate paths and is significantly faster than the SSA in a number of setting, of course tau-leaping is an approximate method and leads to a biased estimator for the estimation of expectations.

In 2008, Giles proposed the remarkable Multilevel Monte Carlo method (MLMC) to simulate diffusion processes [25]. MLMC has been able to reduce CPU times by orders of magnitude in comparison with standard Monte Carlo. A key ingredient of multilevel Monte Carlo is to combine paths at different discretization levels, using relatively fewer paths at the more expensive resolution scales. The accuracy of the final estimate relies on a recursive control variate construction, exploiting the fact that pairs of paths at neighboring resolution levels can be made to have a low variance.

Anderson and Higham adapted the multilevel philosophy for the case of tau-leaping in [3]. This required a novel simulation approach to generate tightly coupled pairs of Poisson-driven paths at neighboring resolutions. The resulting multilevel Monte Carlo tau-leaping algorithm is straightforward to implement and, unlike the standard multilevel methods utilized in the diffusion case, can be made unbiased by using an exact simulator at the most refined level. Computational complexity was analyzed in [3] for a generic system scaling.

In this thesis, we refine and develop the complexity analysis of the new multilevel Monte Carlo simulations for different processes, including continuous time Markov chains (CTMCs), stochastic differential equations driven by Brownian motions in small noise regime, and classically scaled population processes. Careful simulations are provided to illustrate the delicate asymptotic results.

1.2 Mathematical Model and Notations

Many modeling scenarios give rise to continuous-time, discrete-space Markov chains. Notable application areas include chemistry, systems biology, epidemiology, population dynamics, queuing theory and several branches of physics [22, 32, 63, 64, 71]. The main research target of this thesis will be continuous-time, discrete-space Markov chains.

For concreteness, we follow [3] by using the language of chemical kinetics. We consider a system with d constituent species, $\{S_1, \ldots, S_d\}$, taking part in $K < \infty$ reactions, with the kth such reaction written as

$$\sum_{i=1}^{d} \nu_{ki} S_i \to \sum_{i=1}^{d} \nu'_{ki} S_i.$$

Here, $\nu_{ki} \in \mathbb{Z}_{\geq 0}$ specifies the number of molecules of type S_i required as input for the kth reaction, and $\nu'_{ki} \in \mathbb{Z}_{\geq 0}$ specifies the number of molecules of type S_i produced during the kth reaction. If the kth reaction occurs at time t, the system is updated via

$$X(t) = X(t-) + \nu'_k - \nu_k,$$

where ν_k and ν'_k are the vectors whose ith components are ν_{ki} and ν'_{ki}, respectively, and $X_i(t)$ denotes the number of S_i molecules at time t. For notational compactness, we let

$$\zeta_k \stackrel{\text{def}}{=} \nu'_k - \nu_k,$$

which is commonly termed the *reaction vector* for the kth reaction. Then $X(t)$ satisfies

$$X(t) = X(0) + \sum_{k=1}^{K} R_k(t)\zeta_k,$$

where we have enumerated over the reactions, and $R_k(t)$ is the number of times the kth reaction has occurred by time t. By far the most widely used stochastic model for this

system assumes the existence of an *intensity*, or *propensity*, function $\lambda_k : \mathbb{R}^d \to \mathbb{R}$ for the kth reaction so that

$$R_k(t) = Y_k\left(\int_0^t \lambda_k(X(s))ds\right),$$

where the collection $\{Y_k\}$ are independent unit-rate Poisson processes. See, for example, [9, 44, 46]. Thus, $X(t)$ is the solution to the stochastic equation

$$X(t) = X(0) + \sum_k Y_k\left(\int_0^t \lambda_k(X(s))ds\right)\zeta_k. \tag{1.1}$$

We can also describe the model as a continuous time Markov chain with infinitesimal generator

$$(\mathcal{A}f)(x) = \sum_k \lambda_k(x)(f(x+\zeta_k) - f(x)) \tag{1.2}$$

where $f(x) : \mathbb{R}^d \to \mathbb{R}$ is chosen from a sufficiently large class of functions.

The standard choice for the intensity functions λ_k is that of mass-action kinetics, which assumes that for $x \in \mathbb{Z}^d_{\geq 0}$

$$\lambda_k(x) = \kappa_k \prod_{i=1}^{d} \frac{x_i!}{(x_i - \nu_{ki})!} 1_{\{x_i \geq \nu_{ki}\}}. \tag{1.3}$$

We note that the continuous-time Markov chain (1.1) is equivalent to the model derived by Gillespie [27, 28]. The corresponding forward Kolmogorov equation is often referred to as the Chemical Master Equation [71], and solving this very large ODE system forms the basis of an alternative computational approach that is appropriate when detailed information is required about the distribution of the solution process [40, 50].

1.3 Chapter Outline

In this dissertation, we study numerical simulation methods for stochastic models. The dissertation is structured as follows.

In the remainder of this chapter, we will introduce biochemical reaction network and the requisite mathematical model, and we will also summarize different numerical simulations methods that already exist.

In Chapter 2, we will derive new analytic results concerning the computational complexity of multilevel Monte Carlo tau-leaping that are significantly sharper than previously published results in [3].

In Chapter 3, we consider the problem of numerically estimating expectations of solutions to stochastic differential equations driven by Brownian motions in the small noise regime. We compare standard Monte Carlo methods with the multilevel Monte Carlo method. We find multilevel method combined with Euler-Maruyama is the most efficient option under the assumptions we make on the underlying model. We also provide simulations to illustrate the asymptotic results.

In Chapter 4, combining the conclusions obtained from Chapter 2 and Chapter 3, we analyze and compare the computational complexity of different simulation strategies for Monte Carlo in the setting of classically scaled population processes. This setting includes stochastically modeled biochemical systems. We consider the task of approximating the expected value of some function of the state of the system at a fixed time point. We study the use of standard Monte Carlo when samples are produced by exact simulation and by approximation with tau-leaping or an Euler- Maruyama discretization of a diffusion approximation. Appropriate modifications of recently proposed multilevel Monte Carlo algorithms are also studied for the tau-leaping and Euler-Maruyama approaches. We then introduce a novel asymptotic regime where the required accuracy is a function of the model scaling parameter. On the basis of this analysis we find that for this particular scaling a diffusion approximation offers little from a computational

standpoint. Instead, we find that multilevel tau-leaping, which combines exact and tau-leaped samples, is the most promising method. In particular, multilevel tau-leaping provides an unbiased estimate and, up to a logarithm term, is as efficient as a diffusion approximation combined with multilevel Monte Carlo. Our computational experiments confirm the effectiveness of the multilevel tau-leaping approach.

1.4 Monte Carlo Simulation Methods

In this section, we will briefly introduce various Monte Carlo Methods to simulate continuous time, discrete space Markov chains, notably for biochemical reaction systems.

1.4.1 Stochastic Simulation Algorithm

In 1976, Gillespie developed a computational method for the exact simulation of stochastically modeled biochemical reaction systems [27, 28].

Algorithm 1 (Gillespie Algorithm)**.** *Initialize. Set the initial number of molecules of each species and set* $t = 0$*. repeat the following until* $t \geq T$*:*

1. *Calculate the intensity function* $\lambda_i(X)$ *for each reaction and set* $\lambda_0(X) = \sum \lambda_i(X)$*.*
2. *Generate two independent uniform(0, 1) random numbers* r_1 *and* r_2*.*
3. *The time step to the next reaction is determined as* $\tau = \frac{1}{\lambda_0} ln(\frac{1}{r_1})$*.*
4. *The index of next reaction to fire is the smallest integer* j *satisfying* $\sum_{i=1}^{j} \lambda_i(X) > r_2\lambda_0(X)$*.*
5. *Update* $t = t + \tau$ *and* $X = X + \zeta_j$*.*

This procedure can be used to generate trajectories of CTMCs with generators $(\mathcal{A}f)$ given in (1.2) and became the standard algorithm for biochemical reaction networks.

1.4.2 Tau-Leaping Algorithm

Tau-leaping [31] is a computational method that generates Euler-style approximate paths for the continuous-time Markov chain (1.1). The basic idea is to hold the intensity functions fixed over a time interval $[t_n, t_n + h]$ at the values $\lambda_k(X(t_n))$, where $X(t_n)$ is the state of the system at time t_n, and, under this assumption, compute the number of times each reaction takes place over this period. As the waiting times for the reactions are exponentially distributed, this leads to the following algorithm, which simulates up to a time of $T > 0$. For $x \geq 0$ we will write Poisson(x) to denote a sample from the Poisson distribution with parameter x, with all such samples being independent of each other and of all other sources of randomness used.

Algorithm 2 (Euler tau-leaping)**.** *Fix $h > 0$. Set $Z_h(0) = x_0$, $t_0 = 0$, $n = 0$ and repeat the following until $t_{n+1} = T$:*

(i) Set $t_{n+1} = t_n + h$. If $t_{n+1} \geq T$, set $t_{n+1} = T$ and $h = T - t_n$.

(ii) For each k, let $\Lambda_k = Poisson(\lambda_k(Z_h(t_n))h)$.

(iii) Set $Z_h(t_{n+1}) = Z_h(t_n) + \sum_k \Lambda_k \zeta_k$.

(iv) Set $n \leftarrow n + 1$.

Analogously to (1.1), a path-wise representation of Euler tau-leaping defined for all $t \geq 0$ can be given through a random time change of Poisson processes:

$$Z_h(t) = Z_h(0) + \sum_k Y_k \left(\int_0^t \lambda_k(Z_h(\eta_h(s)))ds \right) \zeta_k, \tag{1.4}$$

where the Y_k are as before, and $\eta_h(s) \stackrel{\text{def}}{=} \left\lfloor \frac{s}{h} \right\rfloor h$. Thus, $Z_h(\eta_h(s)) = Z_h(t_n)$ if $t_n \leq s < t_{n+1}$.

Chapter 2

Complexity of Multilevel Monte Carlo Tau-Leaping

2.1 Introduction

This chapter focuses on the commonly arising task of computing an expected value of some feature of the solution of the process (1.1), for example the mean level of a chemical species at some specified time or the correlation between two population levels. We study the method proposed in [3], which combined ideas from tau-leaping and multilevel Monte Carlo in a manner that can dramatically improve computational complexity. Our main aim is to provide further analytical support for the method in the form of substantially sharper bounds for the variances of the coupled processes, and hence for the overall computational complexity of the resulting multilevel Monte Carlo estimator. The sharper analytic bounds will naturally inform implementation.

Tau-leaping proposed by Gillespie [31] is attractive when (a) many events occur over each subinterval, so that exact simulation would be expensive, and (b) the relative change in the system state is small over each subinterval, so that the discretization is accurate [31, 62]. However, in our case, where the aim is to combine sample paths in order to approximate an expected value, it should be kept in mind that

- we are concerned with the overall accuracy of the expected value approximation, not the accuracy of the individual paths,
- forming a Monte Carlo style sample average automatically introduces a statistical sampling error, so we should focus on carefully balancing sampling and discretization effects.

Our aim here is to refine the complexity analysis of [3]. We do this by directly estimating the variance between relevant pairs of processes, rather than bounding via the second moment. Second moment bounds arise naturally in the setting of a diffusion equation, where multilevel was first developed and where strong convergence results for numerical methods are readily available, but they are not optimal for the Poisson-driven jump systems that we consider here.

The next section sets up the notation, defines the simulation methods and discusses some issues that arise in developing realistic results. Section 2.2 presents the new analytical results. The consequent bounds on computational complexity are derived in Section 2.3. Section 2.4 then provides some numerical confirmation of the findings a nd Section 2.5 gives conclusions. Some technical results required for the analysis are collected in Appendix A.

In analyzing computational complexity, it is important to account for the fact that simulation becomes expensive when many reactions take place along a path, so some sort of "large system parameter" is required. Following [3, 11], we denote this parameter by N and scale the model by setting $X_i^N = N^{-\alpha_i} X_i$, where $\alpha_i \geq 0$ is chosen so that $X_i^N = O(1)$. The scaled variable then satisfies

$$X_i^N(t) = X_i^N(0) + \sum_k Y_k \left(\int_0^t \lambda_k(X(s)) ds \right) \zeta_{ki}^N,$$

where $\zeta_{ki}^N = N^{-\alpha_i}\zeta_{ki}$. Also following [3, 11], we set

$$\rho_k \stackrel{\text{def}}{=} \min\{\alpha_i \ : \ \zeta_{ki}^N \neq 0\},$$

so that $\zeta_{ki}^N = O(N^{-\rho_k})$, with the order achieved for at least one component of ζ_k^N (i.e. there can be j for which $\zeta_{kj}^N = o(N^{-\rho_k})$). We also set $\rho \stackrel{\text{def}}{=} \min\{\rho_k\} \geq 0$.

Accounting for the fact that the rate parameters of (1.3), κ_k, may also vary over several orders of magnitude, for each k there is an r_k for which $\lambda_k(X) = N^{r_k}\lambda_k^N(X^N)$ so that $\lambda_k^N(X^N)$ is $O(1)$. This yields the stochastic equation

$$X^N(t) = X^N(0) + \sum_k Y_k\left(\int_0^t N^{r_k}\lambda_k^N(X^N(s))ds\right)\zeta_k^N.$$

Define now

$$\gamma \stackrel{\text{def}}{=} \max_k\{r_k - \rho_k\},$$

which is the natural time-scale for the model. For $\gamma > 0$ the shortest timescale in the problem is much smaller than 1, and for $\gamma < 0$ it is much larger. Our analysis will be relevant to the case where $\gamma \leq 0$.

Setting $r_k = \gamma + c_k$, we arrive at the stochastic equation for our general scaled model,

$$X^N(t) = X^N(0) + \sum_k Y_k\left(N^\gamma\int_0^t N^{c_k}\lambda_k^N(X^N(s))ds\right)\zeta_k^N. \tag{2.1}$$

Note that, by construction, $c_k \leq \rho_k$. We may now regard

$$\overline{N} = N^\gamma\sum_k N^{c_k} \tag{2.2}$$

as an order of magnitude for the number of computations required to generate a single path using an exact algorithm. We refer to [3, 11] for further details and illustrative examples, and point out that the *classical scaling* is covered in this framework by taking $c_k \equiv \rho_k \equiv 1$ and $\gamma = 0$, see [2, 44].

Given a stepsize, h, the tau-leaping approximation for the scaled system (2.1) then takes the form

$$Z_h^N(t) = Z_h^N(0) + \sum_k Y_k \left(N^\gamma \int_0^t N^{c_k} \lambda_k^N(Z_h^N(\eta(s)))ds \right) \zeta_k^N, \tag{2.3}$$

where $\eta(s) \stackrel{\text{def}}{=} \left\lfloor \frac{s}{h} \right\rfloor h$. Note that we have defined the process over continuous time.

The multilevel tau-leaping method for approximating $\mathbb{E}[f(X^N(T))]$ uses levels $\ell = \ell_0, \ell_0+1, \ldots, L$, where both ℓ_0 and L are to be determined. The characteristic stepsize at level ℓ is given by $h_\ell = T \cdot M^{-\ell}$, for a fixed positive integer M, typically between 2 and 7. Using Z_ℓ^N to denote the tau-leaping process (2.3) generated with a step-size of h_ℓ, we consider the telescoping sum identity

$$\mathbb{E}[f(Z_L^N(T))] = \mathbb{E}[f(Z_{\ell_0}^N(T))] + \sum_{\ell=\ell_0+1}^{L} \mathbb{E}[f(Z_\ell^N(T)) - f(Z_{\ell-1}^N(T))]. \tag{2.4}$$

Exploiting the right-hand side of this identity, we introduce the estimators

$$Q_{\ell_0} \stackrel{\text{def}}{=} \frac{1}{n_0} \sum_{i=1}^{n_0} f(Z_{\ell_0,[i]}^N(T)), \quad \text{and} \quad Q_\ell \stackrel{\text{def}}{=} \frac{1}{n_\ell} \sum_{i=1}^{n_\ell} (f(Z_{\ell,[i]}^N(T)) - f(Z_{\ell-1,[i]}^N(T))), \tag{2.5}$$

where the pairs $Z_{\ell,[i]}^N$ and $Z_{\ell-1,[i]}^N$ are to be generated in such a way that $\mathsf{Var}(Q_\ell)$ is small. We will then let

$$Q_B \stackrel{\text{def}}{=} \sum_{\ell=\ell_0}^{L} Q_\ell, \tag{2.6}$$

be the unbiased estimator for $\mathbb{E}[f(Z_L^N(T))]$, which is therefore a biased estimator for the required quantity $\mathbb{E}[f(X^N(T))]$.

We emphasize at this stage that the number of levels, $L - \ell_0 + 1$, and the number of paths per level, n_0 and n_ℓ, have not yet been specified, and will be determined by the required accuracy.

In a similar manner, we may exploit the fact that exact simulation is possible and use the identity

$$\mathbb{E}[f(X^N(T))] = \mathbb{E}[f(X^N(T)) - f(Z_L^N(T))] + \sum_{\ell=\ell_0+1}^{L} \mathbb{E}[f(Z_\ell^N) - f(Z_{\ell-1}^N)] + \mathbb{E}[f(Z_{\ell_0}^N(T))].$$

We then define estimators for the three types of terms on the right-hand side via (2.5) and

$$Q_E \stackrel{\text{def}}{=} \frac{1}{n_E} \sum_{i=1}^{n_E} (f(X_{[i]}^N(T) - f(Z_{L,[i]}^N(T))), \tag{2.7}$$

leading to

$$Q_{UB} \stackrel{\text{def}}{=} Q_E + \sum_{\ell=\ell_0}^{L} Q_\ell, \tag{2.8}$$

which is then an *unbiased* estimator for $\mathbb{E}[f(X^N(T))]$.

In the multilevel framework, rather than single paths, we generally compute pairs of paths, and tight coupling is the key to success. Defining $a \wedge b \stackrel{\text{def}}{=} \min\{a, b\}$, the pair $(Z_\ell^N, Z_{\ell-1}^N)$ appearing in (2.5) is defined as the solution to the stochastic equation

$$\begin{aligned} Z_\ell^N(t) = Z_\ell^N(0) + \sum_k \zeta_k^N \Bigg[& Y_{k,1} \left(N^\gamma N^{c_k} \int_0^t \lambda_k^N(Z_\ell^N(\eta_\ell(s))) \wedge \lambda_k^N(Z_{\ell-1}^N(\eta_{\ell-1}(s)))ds \right) \\ & + Y_{k,2} \left(N^\gamma N^{c_k} \int_0^t [\lambda_k^N(Z_\ell^N(\eta_\ell(s))) - \lambda_k(Z_\ell^N(\eta_\ell(s))) \wedge \lambda_k(Z_{\ell-1}^N(\eta_{\ell-1}(s)))]ds \right) \Bigg], \end{aligned} \tag{2.9}$$

$$\begin{aligned} Z_{\ell-1}^N(t) = Z_{\ell-1}^N(0) + \sum_k \zeta_k^N \Bigg[& Y_{k,1} \left(N^\gamma N^{c_k} \int_0^t \lambda_k(Z_\ell^N(\eta_\ell(s))) \wedge \lambda_k(Z_{\ell-1}^N(\eta_{\ell-1}(s)))ds \right) \\ & + Y_{k,3} \left(N^\gamma N^{c_k} \int_0^t [\lambda_k(Z_{\ell-1}^N(\eta_{\ell-1}(s))) - \lambda_k(Z_\ell^N(\eta_\ell(s))) \wedge \lambda_k(Z_{\ell-1}^N(\eta_{\ell-1}(s)))]ds \right) \Bigg], \end{aligned} \tag{2.10}$$

where the $Y_{k,i},\ i \in \{1, 2, 3\}$, are independent, unit-rate Poisson processes, and for each ℓ, we define $\eta_\ell(s) \stackrel{\text{def}}{=} \lfloor s/h_\ell \rfloor h_\ell$.

Similarly, for (2.7), the pair (X^N, Z_L^N) is defined as the solution to the stochastic equation

$$\begin{aligned} X^N(t) =& X^N(0) + \sum_k Y_{k,1}\left(N^\gamma N^{c_k}\int_0^t \lambda_k^N(X^N(s)) \wedge \lambda_k^N(Z_L^N(\eta_L(s)))ds\right)\zeta_k^N \\ &+ \sum_k Y_{k,2}\left(N^\gamma N^{c_k}\int_0^t [\lambda_k^N(X^N(s)) - \lambda_k^N(X^N(s)) \wedge \lambda_k^N(Z_L^N(\eta_L(s)))]ds\right)\zeta_k^N, \end{aligned} \tag{2.11}$$

$$\begin{aligned} Z_L^N(t) =& Z_L^N(0) + \sum_k Y_{k,1}\left(N^\gamma N^{c_k}\int_0^t \lambda_k^N(X^N(s)) \wedge \lambda_k^N(Z_L^N(\eta_L(s)))ds\right)\zeta_k^N \\ &+ \sum_k Y_{k,3}\left(N^\gamma N^{c_k}\int_0^t [\lambda_k^N(Z_L^N(\eta_L(s))) - \lambda_k^N(X^N(s)) \wedge \lambda_k^N(Z_L^N(\eta_L(s)))]ds\right)\zeta_k^N. \end{aligned} \tag{2.12}$$

In practice, both the biased and the unbiased multilevel Monte Carlo methods described above can be implemented straightforwardly; full pseudo-code descriptions are given in [3]. In this setting, the exact paths X^N are essentially being simulated by the *next reaction method* [23], which has a natural relation to the random change of time representation (1.1); see [1].

To analyze the computational complexity of these methods, we assume that $\mathbb{E}[f(X^N(T))]$ is required to an accuracy of ϵ, in the sense of a confidence interval. Hence, we have in mind that ϵ is small. As discussed earlier, we also wish to regard the system parameter, N, as large. It is important, however, to keep in mind that we have a 'moving target.' As N increases the properties of the underlying system may vary. In particular, for the classical scaling, in the *thermodynamic limit* $N \to \infty$ the stochastic fluctuations become negligible [9, 44] and the system approaches a deterministic ODE. We are implicitly assuming that the problem is specified in a regime where fluctuations are of interest and the task is computationally challenging, so we regard ϵ as small and N as large without

committing ourselves to asymptotic limits. A key aim in our analysis is to capture the effect that the prescribed task can become less costly as N increases, in the sense that fluctuations decay.

The level-wise estimators Q_ℓ and Q_E appearing in (2.6) and (2.8) are independent, and hence their variances add. To make the overall variance achieve the desired value of $O(\epsilon^2)$, our aim is to choose the number of paths, n_ℓ, so that each level contributes equally to the overall variance. The goal, in terms of obtaining a good upper bound on the computational complexity of the method, is therefore to develop tight bounds on the variances of $f(Z^N_{\ell,[i]}(T)) - f(Z^N_{\ell-1,[i]}(T))$ and $f(X^N_{[i]}(T) - f(Z^N_{L,[i]}(T))$.

We note that multilevel was first developed and analyzed for diffusion processes [25]. Here the classical and well-studied concept of L^2 strong error of a numerical method can be used. Two paths that are both close to the underlying exact path in L^2 norm must also be close to each other, and the second moment trivially bounds the variance. This approach has proved useful in connection with Brownian motion, [24, 26, 33], and is optimal in the sense that complexity results can be derived that match the residual $O(\epsilon^{-2})$ cost that would remain if an expression for the exact solution were known. The L^2 approach was also used in [3] for the Possion-driven case that we consider here, but our aim is to show that improved bounds can be obtained from a more refined analysis that studies the variances directly.

2.2 Analyzing the Variance of the Coupled Processes

We begin by explicitly detailing the *running assumption* we make on the model. We will suppose that for each k, the intensity function for the normalized process $\lambda^N_k(x)$, along

with the components of its gradient and Hessian, are bounded.

Running assumption. *We suppose there is a constant $C > 0$ such that,*

$$\sum_k |\lambda_k^N(x)| \leq C, \quad \sum_k \left|\frac{\partial \lambda_k^N(x)}{\partial x_j}\right| \leq C, \text{ and } \sum_k \left|\frac{\partial^2 \lambda_k^N(x)}{\partial x_i \partial x_l}\right| \leq C \quad \text{for any } i,j,l = 1,...,d.$$

Hence, letting

$$F^N(x) \stackrel{def}{=} \sum_k N^\gamma N^{c_k} \lambda_k^N(x) \zeta_k^N,$$

we have

$$\left|\frac{\partial F_i^N}{\partial x_j}\right| \leq CN^\gamma \quad \text{and} \quad \left|\frac{\partial^2 F_i^N}{\partial x_k \partial x_l}\right| \leq CN^\gamma \quad \text{for any } i,j,k,l = 1,2,...,d.$$

We will also assume throughout that our initial condition is fixed. That is, $X^N(0) \equiv x(0) \in \mathbb{R}^d_{\geq 0}$ for all choices of N. □

We note that not all reaction networks satisfy the above running assumption. However, any model for which there is a $w \in \mathbb{R}^d_{>0}$ with $w \cdot \zeta_k \leq 0$ for all k will satisfy the running assumption so long as λ_k^N is defined to satisfy it outside the positive orthant. (Note that the τ-leap process may leave the positive orthant.) In particular, any process which satisfies a conservation relation will satisfy the running assumption. We further wish to emphasize that the boundedness assumption is being made on the *scaled* intensity function, and not the intensity function for the unscaled process. Since the scaled intensity function is explicitly constructed to satisfy $\lambda_k^N(x) = O(1)$ *in our region of interest*, it is not an unrealistic assumption.

Example 2.1. *Consider the family of models, indexed by N, with the following network, rate parameters, which are placed alongside the reaction arrows, and initial condition,*

$$2A \underset{1}{\overset{1/N}{\rightleftarrows}} B, \qquad X_1(0) = X_2(0) = M \cdot N,$$

where $X_1(t)$ and $X_2(t)$ denote the abundances of the A and B molecules, respectively, at time t, $M > 0$ is some constant, and only those N for which $M \cdot N \in \mathbb{Z}$ are considered. Note that $w = [1, 2]^T$ satisfies $w \cdot \zeta_k = 0$ for each $k \in \{1, 2\}$. Choosing $\alpha_i \equiv 1$, so that $\rho_k \equiv r_k \equiv c_k \equiv \rho = 1$, and $\gamma = 0$, the normalized process satisfies

$$\begin{aligned} X^N(t) = X^N(0) &+ \frac{1}{N} Y_1 \left(N \int_0^t \lambda_1^N(X^N(s))ds \right) \begin{bmatrix} -2 \\ 1 \end{bmatrix} \\ &+ \frac{1}{N} Y_2 \left(N \int_0^t \lambda_2^N(X^N(s))ds \right) \begin{bmatrix} 2 \\ -1 \end{bmatrix}, \end{aligned} \tag{2.13}$$

with

$$\lambda_1^N(x) = \begin{cases} x_1(x_1 - N^{-1}), & \text{if } x_1, x_2 \geq 0 \\ h_1^N(x), & \text{if } x_1 < 0 \text{ or } x_2 < 0 \end{cases},$$

where $h_1^N(x)$ is any function which ensures λ_1^N is differentiable at the boundary of the positive orthant and satisfies our running assumptions and

$$\lambda_2^N(x) = \begin{cases} x_2, & \text{if } x_1, x_2 \geq 0 \\ h_2^N(x), & \text{if } x_1 < 0 \text{ or } x_2 < 0 \end{cases},$$

where $h_2^N(x)$ is also chosen to ensure the satisfaction of our running assumptions. For example, we only require that $h_2^N(x) = 0$ when $x_2 = 0$, $h_2^N(x) = (3/2)\lfloor M \cdot N \rfloor$ when $x_1 = 0$, and $h_2^N(x)$ itself satisfies the running assumptions outside the positive orthant. Note that it is necessary to define each λ_k^N outside $\mathbb{Z}_{\geq 0}^2$ for the tau-leaping process, which does not satisfy the same constraints as the exact process. Also note that a bound of $9M^2$ can be used to satisfy our running assumption for this (non-linear) model.

2.2.1 Statements and proofs of our main results

We couple X^N and Z_h^N via (2.11) and (2.12) and are interested in the problem of estimating

$$\mathsf{Var}\left(f(X^N) - f(Z_h^N)\right),$$

for Lipschitz functions $f : \mathbb{R}^d \to \mathbb{R}$. Our main result is the following.

Theorem 2.2. *Suppose the model satisfies our running assumption and that X^N and Z_h^N satisfy the coupling (2.11) and (2.12). Assume that $f : \mathbb{R}^d \to \mathbb{R}$ has continuous second derivative and there exists a constant M such that*

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

Then, for $0 \leq t \leq T$,

$$\mathsf{Var}(f(X^N(t)) - f(Z_h^N(t))) \leq \bar{C}_1(N^\gamma, T, d, M)N^{-\rho}(N^\gamma h)^2 + \bar{C}_2(N^\gamma, T, d, M)N^{-\rho}N^\gamma h,$$

where $\bar{C}_1$ is defined as (2.28) and $\bar{C}_2$ is defined as (2.29). Moreover, the functional dependence of $\bar{C}_1$ and $\bar{C}_2$ on N and T is via the product $N^\gamma T$, so in the case of T fixed and $\gamma \leq 0$, the values of $\bar{C}_1$ and $\bar{C}_2$ may be bounded above uniformly in N, and so the variance is of order $O(N^{-\rho}h)$.

Remark 2.3. *If it is the case that X remains in a bounded region with a probability of one, for example if there is a $w \in \mathbb{R}^d_{>0}$ with $w \cdot \zeta_k \leq 0$ for all k, then the assumption pertaining to the derivative bounds of f is automatically satisfied for any smooth f.*

Remark 2.4. *While we refrain from taking limits as $h \to 0$, we are interested in the order of magnitude of the resulting bounds in both N and h. Hence, we neglect terms of*

the form h^2 when compared with h even though we may have, for example,

$$\sup_{T<\infty} \frac{\bar{C}_1(T)h^2}{\bar{C}_2(T)h} > 1.$$

The assumption that $\bar{C}_1 h^2 \ll \bar{C}_2 h$ may break down at the coarsest levels, where h_ℓ is presumably large. A finer analysis than that presented here, in which the constants $\bar{C}_1$ and $\bar{C}_2$ are estimated more carefully, will be required to study the effect of this assumption on the computational complexity bounds.

In the case $\gamma \leq 0$ the main results in [3], which considered only the L^2 norm of the difference of the coupled processes, present an upper bound on the variance of order $O(N^{-\rho}h + h^2)$. Theorem 2.2 sharpens this bound considerably since in many cases of interest h^2 is much larger than $hN^{-\rho}$.

An immediate corollary to Theorem 2.2 is that the coupled tau-leaping processes satisfy a similar bound.

Corollary 2.5. *Under the assumptions of Theorem 2.2, Z_ℓ^N and $Z_{\ell-1}^N$ in (2.9) and (2.10) satisfy, for $0 \leq t \leq T$,*

$$\mathsf{Var}(f(Z_\ell^N(t)) - f(Z_{\ell-1}^N(t))) \leq \bar{D}_1(N^\gamma, T, d, M)N^{-\rho}(N^\gamma h_\ell)^2 + \bar{D}_2(N^\gamma, T, d, M)N^{-\rho}N^\gamma h_\ell.$$

Where in the case of T fixed and $\gamma \leq 0$, the values of $\bar{D}_1(N^\gamma, T, d, M, C)$ and $\bar{D}_2(N^\gamma, T, d, M, C)$ may be bounded above uniformly in N, and so the variance is of order $O(N^{-\rho}h_\ell)$.

Proof. This is an immediate consequence of the following simple inequality

$$\begin{aligned}\mathsf{Var}(f(Z_\ell^N(t)) - f(Z_{\ell-1}^N(t))) &= \mathsf{Var}(f(Z_\ell^N(t)) - f(X^N(t)) + f(X^N(t)) - f(Z_{\ell-1}^N(t))) \\ &\leq 2\mathsf{Var}(f(Z_\ell^N(t)) - f(X^N(t))) + 2\mathsf{Var}(f(X^N(t)) - f(Z_{\ell-1}^N(t))),\end{aligned}$$

and Theorem 2.2. □

In order to prove Theorem 2.2, we first present some preliminary calculations on the variance of the processes X^N and Z_h^N, and some useful lemmas.

By our running assumption, we know that

$$|F_i^N(x) - F_i^N(y)| \leq \sqrt{d}CN^\gamma|x-y|,$$

for all x, y and $i = 1, 2, ..., d$, where $|\cdot|$ is the 2-norm. We let x^N be the solution to the *deterministic* equation

$$x^N(t) = x(0) + \int_0^t F^N(x^N(s))ds, \tag{2.14}$$

where $x(0)$ is the initial condition for each member of our family of processes.

Lemma 2.6. *Under our running assumption, for any $T > 0$ we have*

$$\mathbb{E}\left[\sup_{s\leq T}|X^N(s) - x^N(s)|^2\right] \leq \left(C_1 e^{C_2\cdot(TN^\gamma)^2}\right)\left(TN^\gamma N^{-\rho}\right),$$

where C_1 and C_2 are two positive constants that do not depend on N, γ, T.

Proof. Defining $\tilde{Y}_k(u) := Y(u) - u$ to be the centered Poisson process we have

$$X^N(t) - x^N(t) = \sum_k \tilde{Y}_k\left(N^\gamma N^{c_k}\int_0^t \lambda_k^N(X^N(s))ds\right)\zeta_k^N + \int_0^t F^N(X^N(s)) - F^N(x^N(s))ds,$$

and so using the trivial bound $(a+b)^2 \leq 2a^2 + 2b^2$ plus our running assumption,

$$\begin{aligned}
&|X^N(t) - x^N(t)|^2 \\
&\qquad \leq 2\left|\sum_k \tilde{Y}_k\left(N^\gamma N^{c_k}\int_0^t \lambda_k^N(X^N(s))ds\right)\zeta_k^N\right|^2 + 2C^2tdN^{2\gamma}\int_0^t |X^N(s) - x^N(s)|^2 ds.
\end{aligned}$$

Hence, by the Burkholder-Davis-Gundy inequality [16] and our running assumption

$$\begin{aligned}
&\mathbb{E}\left[\sup_{s\leq t}|X^N(s) - x^N(s)|^2\right] \\
&\quad \leq 2N^\gamma\sum_k |\zeta_k^N|^2\int_0^t N^{c_k}\mathbb{E}\left[\lambda_k^N(X^N(r))\right]dr + 2C^2tdN^{2\gamma}\int_0^t \mathbb{E}\left[\sup_{s\leq r}|X^N(s) - x^N(s)|^2\right]dr \\
&\quad \leq C_1 N^\gamma N^{-\rho}t + 2C^2tdN^{2\gamma}\int_0^t \mathbb{E}\left[\sup_{s\leq r}|X^N(s) - x^N(s)|^2\right]dr,
\end{aligned}$$

and the result follows from Gronwall's inequality with $C_2 = 2C^2 d$. □

Now we let z be the deterministic solution to

$$z_h^N(t) = x(0) + \int_0^t F^N(z_h^N(\eta(s)))ds, \tag{2.15}$$

which is the Euler approximate solution to the ODE (2.14).

Lemma 2.7. *Under our running assumptions, for any $T > 0$ and any $n \geq 0$ satisfying $nh \leq T$, we have*

$$\mathbb{E}\left[\sup_{m\leq n} |Z_h^N(mh) - z_h^N(mh)|^2\right] \leq \left(C_1 e^{C_2\cdot(TN^\gamma)^2}\right)\left(TN^\gamma N^{-\rho}\right).$$

where C_1 and C_2 are the same constants which appear in Lemma 2.6. Further,

$$\mathbb{E}\left[\sup_{s\leq T} |Z_h^N(s) - z_h^N(s)|^2\right] \leq \left(C_1 e^{2C_2\cdot(TN^\gamma)^2}\right)(TN^\gamma N^{-\rho}).$$

Proof. Following the proof of Lemma 2.6, we have

$$Z_h^N(t) - z_h^N(t) = \sum_k \tilde{Y}_k\left(N^\gamma N^{c_k}\int_0^t \lambda_k^N(Z_h^N(\eta(s)))ds\right)\zeta_k^N + \int_0^t F^N(Z_h^N(\eta(s))) - F^N(z_h^N(\eta(s)))ds,$$

and, again as a result of the Burkholder-Davis-Gundy inequality [16] and our running assumptions,

$$\begin{aligned}
&\mathbb{E}\left[\sup_{s\leq t}\left|Z_h^N(s) - z_h^N(s)\right|^2\right] \\
&\leq 2N^\gamma \sum_k |\zeta_k^N|^2 \int_0^t N^{c_k}\mathbb{E}\left[\lambda_k^N(Z_h^N(\eta(s)))\right]ds + 2C^2 tdN^{2\gamma}\int_0^t \mathbb{E}\left[\sup_{s\leq r}|Z_h^N(\eta(s)) - z_h^N(\eta(s))|^2\right]dr \\
&\leq C_1 N^\gamma N^{-\rho}t + C_2 tN^{2\gamma}\int_0^t \mathbb{E}\left[\sup_{s\leq r}|Z_h^N(\eta(s)) - z_h^N(\eta(s))|^2\right]dr,
\end{aligned} \tag{2.16}$$

where C_1 and C_2 are the same constants which appear in Lemma 2.6.

Using (2.16) with $t = nh$ with n an integer for which $nh \leq T$, we have

$$\mathbb{E}\left[\sup_{m\leq n}|Z_h^N(mh) - z_h^N(mh)|^2\right] \leq \mathbb{E}\left[\sup_{s\leq nh}|Z_h^N(s) - z_h^N(s)|^2\right]$$
$$\leq C_1 N^\gamma N^{-\rho} T + C_2 T N^{2\gamma} \sum_{i=0}^{n-1} h \cdot \mathbb{E}\left[\sup_{m\leq i}|Z_h^N(mh) - z_h^N(mh)|^2\right].$$

By the discrete version of Gronwall's inequality we see

$$\mathbb{E}\left[\sup_{m\leq n}|Z_h^N(mh) - z_h^N(mh)|^2\right] \leq \left(C_1 e^{C_2\cdot(TN^\gamma)^2} T\right)\left(N^\gamma N^{-\rho}\right).$$

Since n satisfying $nh \leq T$ was arbitrary, we return to (2.16) to conclude that for any $0 \leq s \leq T$

$$\mathbb{E}\left[\sup_{s\leq t}|Z_h^N(s) - z_h^N(s)|^2\right] \leq C_1 N^\gamma N^{-\rho} T + C_2 T^2 N^{2\gamma} C_1 e^{C_2 T^2 N^{2\gamma}} N^\gamma N^{-\rho} T.$$

Noting that $(1 + \alpha e^\alpha) \leq e^{2\alpha}$ for any $\alpha \geq 0$ completes the proof. □

Our proof of Theorem 2.2 will begin by considering the difference between X^N and Z_h^N. It is therefore useful to get bounds on the second moment of the quadratic variation of a certain martingale which will arise naturally. We therefore let

$$M^N(t) \stackrel{\text{def}}{=} \sum_k \tilde{Y}_{k,2}\left(N^\gamma N^{c_k}\int_0^t [\lambda_k^N(X^N(s)) - \lambda_k^N(X^N(s)) \wedge \lambda_k^N(Z_h^N(\eta(s)))]ds\right)\zeta_k^N$$
$$- \sum_k \tilde{Y}_{k,3}\left(N^\gamma N^{c_k}\int_0^t [\lambda_k^N(Z_h^N(\eta(s))) - \lambda_k^N(X^N(s)) \wedge \lambda_k^N(Z_h^N(\eta(s)))]ds\right)\zeta_k^N \tag{2.17}$$

where, again, $\tilde{Y}_k$ is the centered Poisson process. The proof of the following lemma can be found in [3].

Lemma 2.8. *Under our running assumption,*

$$\mathbb{E}[|M^N(t)|^2] \leq c_1 T e^{c_2 N^\gamma T} N^{2\gamma}(N^{-\rho}h),$$

where c_1, c_2 are independent of N, γ and T.

We are now ready to prove our main result.

Proof of Theorem 2.2. We first prove the result in the case that $f_i(x) = x_i$. We have

$$\begin{aligned} X_i^N(t) - Z_{h,i}^N(t) &= M_i^N(t) + \int_0^t F_i^N(X^N(s)) - F_i^N(Z_h^N(\eta(s)))ds \\ &= M_i^N(t) + \int_0^t F_i^N(X^N(s)) - F_i^N(Z_h^N(s))ds + \int_0^t F_i^N(Z_h^N(s)) - F_i^N(Z_h^N(\eta(s)))ds, \end{aligned} \tag{2.18}$$

where $M^N(t)$ is defined in (2.17).

We will prove the result by first considering the variance of the first and third terms of (2.18). Consideration of the second term will then lead naturally to an application of Gronwall's inequality.

To begin, we note that Lemma 2.8 implies that

$$\mathsf{Var}(M_i^N(t)) \leq c_1 T e^{c_2 N^\gamma T} N^{2\gamma}(N^{-\rho}h),$$

for some c_1 and c_2 which are positive and do not depend on N^γ and T.

Turning to the third term of (2.18), we have the following lemma.

Lemma 2.9.

$$\begin{aligned} &\mathsf{Var}\left(\int_0^t F_i^N(Z_h^N(s)) - F_i^N(Z_h^N(\eta(s)))ds\right) \\ &\qquad \leq d\hat{C}_1 \cdot (TN^\gamma)^2 N^{-\rho}(N^\gamma h)^2 + d\hat{C}_2 \cdot (TN^\gamma)^2 N^{-\rho} N^\gamma h, \end{aligned}$$

where $\hat{C}_1$ is a constant defined via (2.24), and which depends upon the product TN^γ, and $\hat{C}_2$ is a constant independent of N, γ and T.

Proof. From Lemma A.6 in the appendix, we have

$$F_i^N(Z_h^N(s)) - F_i^N(Z_h^N(\eta(s)) = \int_0^1 \nabla F_i^N(Z_h^N(\eta(s)) + r(Z_h^N(s) - Z_h^N(\eta(s))))dr \cdot (Z_h^N(s) - Z_h^N(\eta(s))). \tag{2.19}$$

In order to bound the right hand side of (2.19), we will apply Lemma A.2 in the appendix with $A^{N,h}$ as the jth component of $\int_0^1 \nabla F_i^N(Z_h^N(\eta(s)) + r(Z_h^N(s) - Z_h^N(\eta(s))))dr$ and $B^{N,h}$ as the jth component of $(Z_h^N(s) - Z_h^N(\eta(s)))$. Hence, we must find appropriate bounds on these components.

We begin with $Z_h^N(s) - Z_h^N(\eta(s))$. As

$$Z_h^N(s) = Z^N(0) + \sum_k Y_k \left(N^\gamma N^{c_k} \int_0^t \lambda_k^N(Z_h^N(\eta(s)))ds \right) \zeta_k^N, \tag{2.20}$$

we see

$$\begin{aligned} Z_h^N(s) - Z_h^N(\eta(s)) &= \sum_k Y_k \left(N^\gamma N^{c_k} \int_0^s \lambda_k^N(Z_h^N(\eta(s)))ds \right) \zeta_k^N \\ &\quad - \sum_k Y_k \left(N^\gamma N^{c_k} \int_0^{\eta(s)} \lambda_k^N(Z_h^N(\eta(s)))ds \right) \zeta_k^N \\ &\stackrel{d}{=} \sum_k Y_k' \left(N^\gamma N^{c_k} \int_{\eta(s)}^s \lambda_k^N(Z_h^N(\eta(s)))ds \right) \zeta_k^N, \end{aligned}$$

where the collection $\{Y_k'\}$ are independent unit-rate Poisson processes and the last equality is in the sense of distributions. This implies

$$\left|\mathbb{E}[Z_h^N(s) - Z_h^N(\eta(s))]\right| = \left|\mathbb{E}\left[\mathbb{E}[Z_h^N(s) - Z_h^N(\eta(s))|Z_h^N(\eta(s))]\right]\right| \leq CN^\gamma h.$$

Similarly, using Lemma 2.7 and the law of total variance, we may conclude

$$\begin{aligned}
\mathsf{Var}(Z^N_{h,j}(s) - Z^N_{h,j}(\eta(s))) &= \mathbb{E}\left[\mathsf{Var}(Z^N_{h,j}(s) - Z^N_{h,j}(\eta(s))) \mid Z^N_h(\eta(s)))\right] \\
&\quad + \mathsf{Var}\left(\mathbb{E}[Z^N_{h,j}(s) - Z^N_{h,j}(\eta(s))] \mid Z^N_h(\eta(s)))\right) \\
&= \mathbb{E}\left[\mathsf{Var}\left(\sum_k \tilde{Y}_k\left(N^\gamma N^{c_k}\int_{\eta(s)}^s \lambda^N_k(Z^N_h(\eta(s)))ds\right)\zeta^N_{kj} \;\middle|\; Z^N_h(\eta(s))\right)\right] \\
&\quad + \mathsf{Var}\left(\mathbb{E}\left[\sum_k \tilde{Y}_k\left(N^\gamma N^{c_k}\int_{\eta(s)}^s \lambda^N_k(Z^N_h(\eta(s)))ds\right)\zeta^N_{kj} \;\middle|\; Z^N_h(\eta(s))\right]\right) \\
&= \sum_k (s-\eta(s))N^\gamma N^{c_k}\mathbb{E}\left[\lambda^N_k(Z^N_h(\eta(s)))\right](\zeta^N_{kj})^2 \\
&\quad + \mathsf{Var}\left(\sum_k (s-\eta(s))N^\gamma N^{c_k}\lambda^N_k(Z^N_h(\eta(s)))\zeta^N_{kj}\right) \\
&\le CN^{-\rho}N^\gamma h + h^2N^{2\gamma}K\sum_k \mathsf{Var}\left(\lambda^N_k(Z^N_h(\eta(s)))\right) \\
&\le CN^{-\rho}N^\gamma h + h^2N^{2\gamma}K\sum_k \mathbb{E}\left[\left|\lambda^N_k(Z^N_h(\eta(s)) - \lambda^N_k(z^N_h(\eta(s)))\right|^2\right] \\
&\le CN^{-\rho}N^\gamma h + dK^2C^2\left(C_1e^{C_2\cdot(TN^\gamma)^2}\right)\left(TN^\gamma N^{-\rho}\right)h^2,
\end{aligned} \tag{2.21}$$

where we recall that K is the number of reaction channels, and Lemma 2.7 was utilized in the final inequality.

Turning to $\int_0^1 \nabla F^N_i(Z^N_h(\eta(s)) + r(Z^N_h(s) - Z^N_h(\eta(s))))dr$, we apply Lemma A.1 in the appendix with $X_1(s) = Z^N_h(s)$, $X_2(s) = Z^N_h(\eta(s))$, $x_1(s) = z^N_h(s)$, $x_2(s) = z^N_h(\eta(s))$ and $u(x) = [\nabla F^N_i(x)]_j$ to obtain

$$\mathsf{Var}\left(\int_0^1 [\nabla F^N_i(Z^N_h(\eta(s)) + r(Z^N_h(s) - Z^N_h(\eta(s)))]_j dr\right) \le N^{2\gamma}\bar{C}N^{-\rho},$$

where

$$\bar{C} = dC^2C_1e^{2C_2\cdot(TN^\gamma)^2}(TN^\gamma N^{-\rho}). \tag{2.22}$$

In order to apply Lemma A.2 with

$$A^{N,h} = \int_0^1 [\nabla F_i^N(Z_h^N(\eta(s)) + r(Z_h^N(s) - Z_h^N(\eta(s))))]_j dr$$

and $B^N = Z_{h,j}^N(s) - Z_{h,j}^N(\eta(s))$, we note $\left|[\nabla F_i^N]_j\right| \leq CN^\gamma$ from our running assumptions to see

$$\begin{aligned}
&\mathsf{Var}\left(\int_0^1 [\nabla F_i^N(Z_h^N(\eta(s)) + r(Z_h^N(s) - Z_h^N(\eta(s))))]_j dr \cdot (Z_{h,j}^N(s) - Z_{h,j}^N(\eta(s)))\right) \\
&\quad \leq 3C^2\bar{C}N^{2\gamma}N^{-\rho}(N^\gamma h)^2 + 15C^2N^{2\gamma}\mathsf{Var}(Z_{h,j}^N(s) - Z_{h,j}^N(\eta(s))) \\
&\quad \leq \hat{C}_1 N^{2\gamma}N^{-\rho}(N^\gamma h)^2 + \hat{C}_2 N^{2\gamma}N^{-\rho}N^\gamma h, \qquad (2.23)
\end{aligned}$$

where the final inequality follows from (2.21) with

$$\begin{aligned}
\hat{C}_1 &= 3C^2\bar{C} + 15dK^2C^4C_1 e^{C_2\cdot(TN^\gamma)^2}TN^\gamma, \\
\hat{C}_2 &= 15C^3. \qquad (2.24)
\end{aligned}$$

Returning to (2.19), the above allows us to conclude

$$\begin{aligned}
&\mathsf{Var}(F_i^N(Z_h^N(s)) - F_i^N(Z_h^N(\eta(s))) \\
&\qquad \leq d^2\hat{C}_1 N^{2\gamma}N^{-\rho}(N^\gamma h)^2 + d^2\hat{C}_2 N^{2\gamma}N^{-\rho}N^\gamma h.
\end{aligned}$$

Finally, by Lemma A.5 in the Appendix,

$$\begin{aligned}
&\mathsf{Var}\left(\int_0^t F_i^N(Z_h^N(s)) - F_i^N(Z_h^N(\eta(s))ds\right) \\
&\qquad \leq d^2T^2\hat{C}_1 N^{2\gamma}N^{-\rho}(N^\gamma h)^2 + d^2T^2\hat{C}_2 N^{2\gamma}N^{-\rho}N^\gamma h,
\end{aligned}$$

as desired. □

We now turn to the middle term of (2.18). We first write

$$F_i^N(X^N(s)) - F_i^N(Z_h^N(s)) = DF_i^N(s)\cdot(X^N(s) - Z_h^N(s)),$$

where

$$DF_i^N(s) = \int_0^1 \nabla F_i^N(Z_h^N(s) + r(X^N(s) - Z_h^N(s)))dr.$$

We will again apply Lemma A.2 to get the necessary bounds. Therefore, we let $A^N = [DF_i^N(s)]_j$ and $B^N = X_j^N(s) - Z_{h,j}^N(s)$.

Letting $X_1(s) = X^N(s)$, $X_2(s) = Z^N(s)$, $x_1(s) = x^N(s)$, $x_2(s) = z_h^N(s)$ and $u(x) = [\nabla F_i^N(x)]_j$ for an application of Lemma A.1, we have

$$\mathsf{Var}(A^N) \leq N^{2\gamma}\bar{C}N^{-\rho},$$

where $\bar{C}$ is defined in (2.22) and where we use our running assumption that $[\nabla F_i]_j$ is uniformly bounded by CN^γ. From [3, Lemma 3] we know

$$\mathbb{E}\left[|B^N|\right] \leq \tilde{c}_1(e^{\tilde{c}_2 N^\gamma T} - 1)N^\gamma h,$$

where $\tilde{c}_1$ and $\tilde{c}_2$ are constants independent of N, γ and T. Hence, applying Lemma A.2 we see

$$\begin{aligned}&\mathsf{Var}([DF_i^N(s)]_j(X_j^N(s) - Z_{h,j}^N(s))) \\ &\quad \leq 3N^{2\gamma}\bar{C}\tilde{c}_1^2(e^{\tilde{c}_2 N^\gamma T} - 1)^2 N^{-\rho}(N^\gamma h)^2 + 15C^2N^{2\gamma}\mathsf{Var}(X_j^N(s) - Z_{h,j}^N(s)),\end{aligned}$$

and

$$\mathsf{Var}(F_i^N(X^N(s)) - F_i^N(Z_h^N(s))) \leq 15C^2dN^{2\gamma}\sum_{j=1}^d \mathsf{Var}(X_j^N(s) - Z_{h,j}^N(s)) + d^2\bar{c}_1 N^{2\gamma}N^{-\rho}(N^\gamma h)^2,$$

where

$$\bar{c}_1 = 3\bar{C}\tilde{c}_1^2(e^{\tilde{c}_2 T N^\gamma} - 1)^2. \tag{2.25}$$

Finally returning to (2.18), we may combine all of the above to see

$$\begin{aligned}
\mathsf{Var}(X_i^N(t) - Z_{h,i}^N(t)) &\leq 3c_1 T e^{c_2 N^\gamma T} N^{2\gamma}(N^{-\rho}h) \\
&\quad + 3\left[15C^2 d N^{2\gamma} t \sum_{j=1}^{d} \int_0^t \mathsf{Var}(X_j^N(s) - Z_{h,j}^N(s))\, ds + d^2 \bar{c}_1 N^{2\gamma} N^{-\rho} (N^\gamma h)^2 T^2\right] \\
&\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad \text{(Lemma A.5)} \\
&\quad + 3\left[d^2 \hat{C}_1 \cdot (TN^\gamma)^2 N^{-\rho} (N^\gamma h)^2 + d^2 \hat{C}_2 \cdot (TN^\gamma)^2 N^{-\rho} N^\gamma h\right] \qquad \text{(Lemma 2.9)} \\
&\leq 3d^2 T^2 (\hat{C}_1 + \bar{c}_1) N^{2\gamma} N^{-\rho} (N^\gamma h)^2 + 3(d^2 T^2 \hat{C}_2 N^{2\gamma} + c_1 N^\gamma T e^{c_2 N^\gamma T}) N^{-\rho} N^\gamma h \\
&\quad + 45 C^2 T d N^{2\gamma} \sum_{j=1}^{d} \int_0^t \mathsf{Var}(X_j^N(s) - Z_{h,j}^N(s)) ds.
\end{aligned}$$

Thus, setting

$$g(t) = \max_{i \in \{1,\ldots,d\}} \{\mathsf{Var}([X^N(t)]_i - [Z_h^N(t)]_i)\},$$

we have

$$\begin{aligned}
g(t) &\leq 3d^2 T^2 (\hat{C}_1 + \bar{c}_1) N^{2\gamma} N^{-\rho} (N^\gamma h)^2 + 3(d^2 T^2 \hat{C}_2 N^{2\gamma} + c_1 N^\gamma T e^{c_2 N^\gamma T}) N^{-\rho} N^\gamma h \\
&\quad + 45 C^2 T d^2 N^{2\gamma} \int_0^t g(s) ds,
\end{aligned}$$

and the result under the assumption that $f_i(x) = x_i$ is shown by an application of Gronwall's inequality,

$$\begin{aligned}
\mathsf{Var}(X_i^N(s) - Z_{h,i}^N(s)) &\leq 3e^{45C^2T^2d^2N^{2\gamma}} d^2 T^2 (\hat{C}_1 + \bar{c}_1) N^{2\gamma} N^{-\rho} (N^\gamma h)^2 \\
&\qquad + 3e^{45C^2T^2d^2N^{2\gamma}} (d^2 T^2 \hat{C}_2 N^{2\gamma} + c_1 N^\gamma T e^{c_2 N^\gamma T}) N^{-\rho} N^\gamma h \\
&= \tilde{C}_1 \cdot N^{-\rho} (N^\gamma h)^2 + \tilde{C}_2 \cdot N^{-\rho} N^\gamma h, \qquad (2.26)
\end{aligned}$$

where $\tilde{C}_1$ and $\tilde{C}_2$ are defined by the above equality.

To show the general case, note that from Lemma A.6 in the appendix we have

$$f(X^N(t)) - f(Z_h^N(t)) = \int_0^1 \nabla f(Z_h^N(t) + r(X^N(t) - Z_h^N(t)))dr \cdot (X^N(t) - Z_h^N(t)).$$

We let $X_1(t) = X^N(t)$, $X_2(t) = Z_h^N(t)$, $x_1(t) = x^N(t)$, $x_2(t) = z_h^N(t)$ and $u(x) = [\nabla f(x)]_j$ for an application of Lemma A.1, which yields

$$\mathsf{Var}\left(\int_0^1 [\nabla f(Z_h^N(t) + r(X^N(t) - Z^N(t)))]_j dr\right) \leq dM^2 \bar{C} N^{-\rho}, \tag{2.27}$$

where M is the uniform bound of $\nabla_{ij} f(x)$ and the factor of d appears in the application of Lemma A.1 since $|[\nabla f(x)]_j - [\nabla f(y)]_j| \leq M\sqrt{d}|x - y|$ for each j. Hence, by an application of Lemma A.2 and the work above we see,

$$\begin{aligned}&\mathsf{Var}\left(\int_0^1 \nabla_j f(Z_h^N(t) + r(X^N(t) - Z_h^N(t)))dr \cdot (X_j^N(s) - Z_{h,j}^N(s))\right)\\ &\quad \leq dM^2 \bar{c}_1 N^{-\rho}(N^\gamma h)^2 + 15M^2 \mathsf{Var}(X_j^N(s) - Z_{h,j}^N(s)).\end{aligned}$$

Thus, utilizing (2.26), we have

$$\mathsf{Var}(f(X^N(t)) - f(Z_h^N(t))) \leq \bar{C}_1 \cdot N^{-\rho}(N^\gamma h)^2 + \bar{C}_2 \cdot N^{-\rho} N^\gamma h,$$

where

$$\bar{C}_1 = 15d^2 M^2 \tilde{C}_1 + d^3 m^2 \bar{c}_1, \tag{2.28}$$

and

$$\bar{C}_2 = 15d^2 M^2 \tilde{C}_2. \tag{2.29}$$

Note that the functional dependence of $\bar{C}_1$ and $\bar{C}_2$ on N and T is via the product $N^\gamma T$. □

2.3 Consequences for Complexity

Under the assumption $\gamma \leq 0$ (which holds, for example, in the classical scaling) and using the fact that $h_{\ell_0} \leq T$, Theorem 2.2 and Corollary 2.5 show that the variances in

the level-wise estimators Q_ℓ and Q_E appearing in (2.5) and (2.7) can be bounded as

$$\mathsf{Var}(Q_\ell) \le C \cdot \frac{N^{-\rho} N^\gamma h_\ell}{n_\ell} \qquad \text{and} \qquad \mathsf{Var}(Q_E) \le C \cdot \frac{N^{-\rho} N^\gamma h_L}{n_E}, \tag{2.30}$$

where C is a generic constant. Compared with the original analysis in [3, Theorems 1 and 2], the important refinement is deletion of the h_ℓ^2 terms found in the similar bounds of [3]. These h_ℓ^2 terms dominate in many cases, including when $h_\ell > N^{-1}$ and the system satisfies the classical scaling (recall that $\rho = 1$ in the classical scaling).

Further, in the basic inequality

$$\mathsf{Var}(f(Z_{\ell_0}^N(t))) \le \left(\sqrt{\mathsf{Var}(f(Z_{\ell_0}^N(t)) - f(X^N(t)))} + \sqrt{\mathsf{Var}(f(X^N(t)))} \right)^2$$

we may use Theorem 2.2 with $h = h_{\ell_0}$ to control $\mathsf{Var}(f(Z_{\ell_0}) - f(X^N)))$ and Lemma 2.6 to control $\mathsf{Var}(f(X^N))$, to find that

$$\mathsf{Var}(Q_0) \le C \cdot \frac{N^{-\rho} N^\gamma}{n_0}. \tag{2.31}$$

Taking variance in (2.6) it then follows from (2.30) and (2.31) that to leading order

$$\mathsf{Var}(Q_B) \le C N^{-\rho} N^\gamma \left(\frac{1}{n_0} + \sum_{\ell=\ell_0+1}^{L} \frac{h_\ell}{n_\ell} \right).$$

To achieve the required overall variance of ϵ^2, and to spread the variance budget fairly evenly across the levels, we may then use, to order of magnitude,

$$n_0 = N^{-\rho} N^\gamma (L - \ell_0 + 1) \epsilon^{-2} \qquad \text{and} \qquad n_\ell = N^{-\rho} N^\gamma (L - \ell_0 + 1) h_\ell \epsilon^{-2}. \tag{2.32}$$

For the biased estimator (2.6) we take $h_L = O(\epsilon)$ in order for the discretization error to be within our target accuracy. This gives $L = O(|\ln(\epsilon)|)$ levels. The computational cost of each pair of tau-leap paths is proportional to h_ℓ^{-1}, and hence the overall complexity

is of magnitude

$$n_0 h_{\ell_0}^{-1} + \sum_{\ell=\ell_0+1}^{L} n_\ell h_\ell^{-1} = N^{-\rho} N^\gamma \epsilon^{-2} \left((L - \ell_0 + 1) h_{\ell_0}^{-1} + (L - \ell_0 + 1)(L - \ell_0) \right). \quad (2.33)$$

Based on this analysis, which we recall is performed under the assumption that $\gamma \leq 0$, we take $\ell_0 = 0$. Doing so yields a computational complexity of leading order

$$N^{-\rho} N^\gamma \epsilon^{-2} \ln(\epsilon)^2.$$

Still assuming that $\ell_0 = 0$, for the unbiased estimator (2.8) we may take, to leading order of magnitude,

$$n_0 = N^{-\rho} N^\gamma (L+2) \epsilon^{-2}, \quad n_\ell = N^{-\rho} N^\gamma (L+2) h_\ell \epsilon^{-2}, \quad \text{and} \quad n_E = N^{-\rho} N^\gamma (L+2) h_L \epsilon^{-2}.$$

If we again use $h_L = O(\epsilon)$, then the computational complexity for the unbiased method is bounded in magnitude by

$$N^{-\rho} N^\gamma \epsilon^{-2} \ln(\epsilon)^2 + \max\{N^{-\rho} N^\gamma |\log(\epsilon)| \cdot h_L \epsilon^{-2}, 1\} \cdot \overline{N}, \quad (2.34)$$

where we recall that $\overline{N}$, defined in (2.2), is the cost of computing a sample path with the exact method, and we note that the maximum appears because some exact paths are necessarily computed for the unbiased method. We also note that we are implicitly assuming that h_L^{-1} is not appreciably larger than $\overline{N}$, which we believe is reasonable. Finally, while we computed the above under the assumption that $h_L = O(\epsilon)$, we kept $h_L \epsilon^{-2}$ in the second term of (2.34) instead of simply writing ϵ^{-1} in order to explicitly point out the dependence on each term.

We note that the complexity bounds derived in [3], which considered only the L^2 norm of the difference of the coupled processes, have another term of order $\max\{N^{2\gamma} h_L^2 \epsilon^{-2}, 1\} \cdot$

$\overline{N}$ added to (2.34). This term was often the dominating one, and has been removed by the direct analysis on the variance presented here.

To finish this section we point out that the analysis produces upper bounds on the computational complexity—in particular, the choices for the number of samples paths per level are sufficient to achieve the required accuracy, based on bounds on the individual variances and with the strategy of spreading the cost evenly between levels, but we have not shown that they are optimal. In practice, and as described more fully in [3], for a given problem, and with a small amount of further computation, it is possible to perform an initial optimization in order to choose these key parameters adaptively. Hence, practical performance may outstrip these complexity bounds.

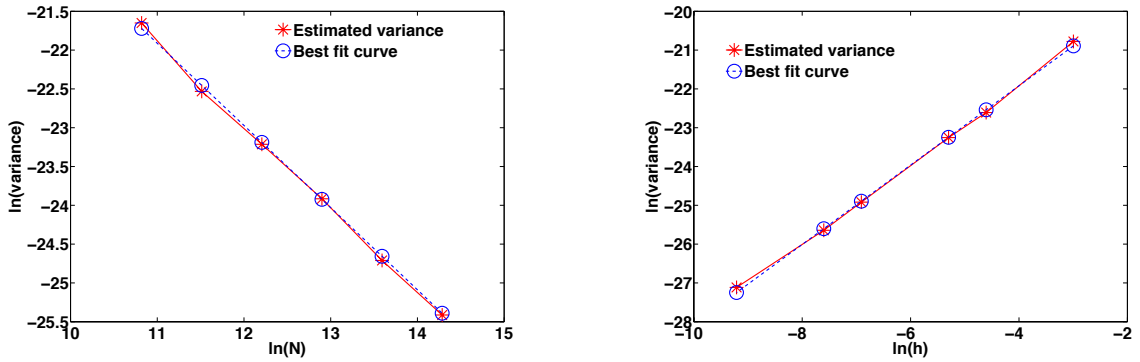
2.4 Computational Test

The efficiency of multilevel Monte Carlo tau-leaping was demonstrated computationally in [3], so we restrict ourselves here to testing the sharpness of our new analytical results on a simple non-linear model. We consider a particular case from the family of models presented in Example 2.1.

Example 2.10. *Consider the case where $M = 0.2$ in the model of Example 2.1, defined through (2.13). In Figure 1, we provide log-log plots of $\mathsf{Var}(X_1^N(T) - Z_{h,1}^N(T))$ for the coupled processes with $T = 0.3$, and varying values for N and h. The plots are consistent with the functional form*

$$\mathsf{Var}(X_1^N(t) - Z_{h,1}^N(t)) \approx CN^{-1}h, \tag{2.35}$$

matching the bound arising from Theorem 2.2. The best fit curve for the data, obtained



(a) Varying N with $h = 0.001$ fixed.

(b) Varying h with $N = 10^6$ fixed.

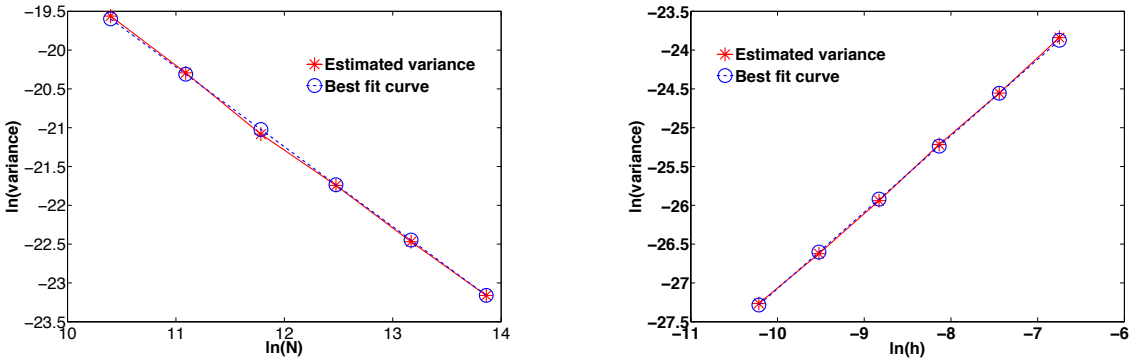
Figure 1: Log-log plots of $\mathsf{Var}(X_1^N(t) - Z_{h,1}^N(t))$ *for the model in Example 2.10. In (a),* h *is held constant while* N *is changed. In (b),* N *is fixed while* h *is varied. The best fit curve for all the data is overlain in the dashed blue line.*

by a least squares approximation and which is shown in each image, is $\mathsf{Var}(X_1^N(t) - Z_{h,1}^N(t)) \approx 0.0408 \cdot N^{-1.0588} h^{1.0228}$.

In Figure 2, we provide log-log plots of $\mathsf{Var}(Z_{\ell,1}^N(T) - Z_{\ell-1,1}^N(T))$ *for the coupled processes with* $T = 0.3$*, and varying values for* N *and* h_ℓ*. These plots also follow the functional form of (2.35), matching the bound arising from Corollary 2.5, with best fit curve of* $\mathsf{Var}(Z_{\ell,1}^N(T) - Z_{\ell-1,1}^N(T)) \approx 0.1038 \cdot N^{-1.0279} h_\ell^{0.9845}$.

2.5 Conclusions

The main contribution of this work is to add further theoretical support for multilevel Monte Carlo tau-leaping by developing new complexity bounds that behave well for large values of the system size parameter. To do this, we took the novel step of directly estimating the variance between pairs of paths, rather than proceeding via a mean-square convergence property. We also provided numerical support showing our estimates for



(a) Varying N with $h = 0.001$ fixed.

(b) Varying h with $N = 10^6$ fixed.

Figure 2: Log-log plots of $\mathsf{Var}(Z^N_{\ell,1}(t) - Z^N_{\ell-1,1}(t))$ *for the model in Example 2.10. In (a), h is held constant while N is changed. In (b), N is fixed while h is varied. The best fit curve for all the data is overlain in the dashed blue line.*

the variances are sharp.

Stochastic simulation of continuous time, discrete space, Markov chains is a bottleneck across a range of application areas, and there are many promising directions for further study of multilevel Monte Carlo in this context. In particular, specific instances of the very general scaling considered here could be used in order to develop more customized strategies, and complexity bounds, in suitable model classes; for example, where there is a known separation of scale.

The analysis presented is valid for $\gamma \leq 0$. For the case $\gamma > 0$, which is the regime of "stiff" systems, it is often possible to generate, via averaging techniques, a reduced model satisfying $\gamma \leq 0$. Taking this reduced model as the "finest level" in a MLMC framework is then a natural way to proceed in the construction of an efficient Monte Carlo method. This procedure was carried out successfully in Section 9 of [3] for an example of viral growth.

Chapter 3

Multilevel Monte Carlo for stochastic differential equations with small noise

3.1 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 [9, 10, 29, 30]. 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, [17, 72], with a more general overview in [60]. In computational fluid dynamics, small noise SDEs are used as a means to incorporate thermal fluctuations into traditional models in the "weak fluctuation regime" [20, Section V]. In several other application areas, including ecology, circuit simulation, microbiology, neuroscience and population dynamics, [12, 21, 52, 59, 65, 66, 67, 69], 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 [39, 61].

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 [57, 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 that under a range of scalings the standard Euler–Maruyama method combined with the usual multilevel Monte Carlo method of Giles [25] 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; 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 $\varepsilon \in (0, 1)$ be a small parameter and let D^ε be the solution to the following Itô SDE,

$$D^\varepsilon(t) = D(0) + \int_0^t \mu(D^\varepsilon(s))ds + \varepsilon \int_0^t \sigma(D^\varepsilon(s))dW(s), \tag{3.1}$$

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

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 [55, 56], and (ii) multilevel Monte Carlo methods combined with Euler-Maruyama [25]. 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 [55] do not provide sharp estimates for the variance between two coupled paths; an analogous issue was previously addressed in the jump process setting [4]. 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 if $\mathsf{Var}(f(D^\varepsilon(T))) = O(\varepsilon^2)$, 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 [25] 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 (3.1), 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:*

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

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 (3.1) is known to have a unique strong solution (see, for example, Theorem 3.1 on page 51 in [51]). 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 [35].

3.1.1 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

$$D_h^\varepsilon(t) = D(0) + \int_0^t \mu(D_h^\varepsilon(\eta_h(s)))\, ds + \varepsilon \int_0^t \sigma(D_h^\varepsilon(\eta_h(s)))\, dW(s), \tag{3.2}$$

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 (3.2) 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 [43].

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 thesis. The result bounds the variance between two coupled process; both are generated via (3.2), though they have different time discretization parameters. See the beginning of section 3.3 for more details related to the coupling.

Theorem 3.1. *Suppose the functions μ and σ satisfy our running assumptions and that $T > 0$ and $\varepsilon \in (0, 1)$. Suppose further that $D^{\varepsilon}_{h_\ell}(t)$ and $D^{\varepsilon}_{h_{\ell-1}}(t)$ satisfy (3.2) 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 $f : \mathbb{R}^d \to \mathbb{R}$ has continuous second derivative and there exists a constant C_L such that*

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

Then, for $\ell \geq 1$,

$$\max_{0 \leq n \leq M^{\ell-1}} \mathsf{Var}(f(D^{\varepsilon}_{h_\ell}(t_n)) - f(D^{\varepsilon}_{h_{\ell-1}}(t_n))) \leq \bar{C}_1 h^2_{\ell-1} \varepsilon^2 + \bar{C}_2 h_{\ell-1} \varepsilon^4, \tag{3.3}$$

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 analyzing the classical mean-square error, it was shown by Milstein and Tretyakov in [55] that under the same assumptions as in Theorem 3.1,

$$\mathbb{E}[|f(D^{\varepsilon}(T)) - f(D^{\varepsilon}_h(T))|^2] = O(h^2 + h\varepsilon^4), \tag{3.4}$$

where D^ε is the solution to (3.1). We note that the $O(h^2)$ term cannot be avoided when we analyze the mean-square error because the underlying deterministic Euler method is first order. From the mean-square error bound (3.4) we can deduce that for some $C_1, C_2 > 0$, we have $\max_{0\le n\le M^{\ell-1}} \mathsf{Var}(f(D^\varepsilon_{h_\ell}(t_n)) - f(D^\varepsilon_{h_{\ell-1}}(t_n))) \le C_1 h_{\ell-1}^2 + C_2 h_{\ell-1}\varepsilon^4$, where, again, $t_n = n \cdot h_{\ell-1}$. Theorem 3.1 sharpens this bound considerably, showing that the overall variance scales favorably with ε, even though the Euler–Maruyama method has not been customized to exploit the small noise property.

3.2 Complexity analysis

3.2.1 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^ε_h is generated by a numerical scheme (not necessarily (3.2)) for which the bias of the discretization method satisfies

$$|\mathbb{E}[f(D^\varepsilon_h(T))] - \mathbb{E}[f(D^\varepsilon(T))]| = O(h^p + \varepsilon^r h^q), \tag{3.5}$$

where $q \le p$ and $r \ge 0$ (see [56], where some such methods are provided). In order to ensure that the bias (3.5) is of order δ, we require that

$$h = O(\min(\delta^{1/p}, \delta^{1/q}\varepsilon^{-r/q})). \tag{3.6}$$

Under our running assumptions and assuming Lemma A.4 below, which applies to Euler-Maruyama, holds for these customized methods we find

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

where z_h is the Euler solution to the associated deterministic model obtained when ε is set to 0 in (3.1), see (3.19). 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 D_h^ε, 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

$$O((\varepsilon^2\delta^{-2}+1)h^{-1}) = O\left(\frac{\varepsilon^2\delta^{-2}+1}{\min(\delta^{1/p}, \delta^{1/q}\varepsilon^{-r/q})}\right). \tag{3.7}$$

For example, with the Euler-Maruyama scheme (3.2) we have that $p = q = 1, r = 0$ yielding a bias of $O(h)$ in (3.5). 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 [56]) in which case (3.6) gives $h = O(\varepsilon^{\rho/2})$, and (3.7) yields a computational complexity of order $O(\varepsilon^{2-\frac{5}{2}\rho})$; see subsection 3.2.3 for further details..

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 [25].

For a fixed positive integer $M \geq 2$ we let $h_\ell = T \cdot M^{-\ell}$ for $\ell \in \{0, \ldots, L\}$. Reasonable choices for M include $M \in \{2, 3, 4, 5, 6, 7\}$, and L is determined below. For each $\ell \in$

$\{0, 1, \ldots, L\}$, let $D^N_{h_\ell}$ denote the approximate process generated by (3.2) with a step size of h_ℓ. Note that

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

with the quality of the approximation only depending upon h_L. As mentioned in [56], 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

$$\widehat{Q}^N_0 \stackrel{\text{def}}{=} \frac{1}{n_0} \sum_{i=1}^{n_0} f(D^N_{h_0,[i]}(T)), \quad \text{and} \quad \widehat{Q}^N_\ell \stackrel{\text{def}}{=} \frac{1}{n_\ell} \sum_{i=1}^{n_\ell} (f(D^N_{h_\ell,[i]}(T)) - f(D^N_{h_{\ell-1},[i]}(T))),$$

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

$$\widehat{Q}^N \stackrel{\text{def}}{=} \widehat{Q}^N_0 + \sum_{\ell=1}^{L} \widehat{Q}^N_\ell,$$

which is the usual multilevel Monte Carlo estimator [25]. Set

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

By Theorem 3.1, we have $\delta_{\varepsilon,\ell} = O(h^2_\ell \varepsilon^2 + h_\ell \varepsilon^4)$ under a wide array of circumstances. Also note that $\delta_{\varepsilon,0} = \mathsf{Var}(f(D^\varepsilon_0)) = O(\varepsilon^2)$.

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 $\widehat{Q}^N$

no greater than order δ^2:

$$\underset{n_\ell}{\text{minimize}} \quad \sum_{\ell=0}^{L} n_\ell C_\ell, \tag{3.8}$$

$$\text{subject to} \quad \sum_{\ell=0}^{L} \frac{\delta_{\varepsilon,\ell}}{n_\ell} = \delta^2. \tag{3.9}$$

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

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

By taking a derivative with respect to n_ℓ we obtain,

$$n_\ell = \sqrt{\tfrac{\lambda}{K} \delta_{\varepsilon,\ell} h_\ell}, \qquad \text{for } \ell \in \{0, 1, 2, \ldots, L\} \tag{3.10}$$

for some $\lambda \geq 0$. Plugging (3.10) into (3.9) yields

$$\sum_{\ell=0}^{L} \sqrt{\frac{\delta_{\varepsilon,\ell}}{h_\ell}} = \sqrt{\tfrac{\lambda}{K}} \cdot \delta^2 \tag{3.11}$$

and hence, by Theorem 3.1 there is a $C > 0$ for which

$$\sqrt{\tfrac{\lambda}{K}} = \delta^{-2} \sum_{\ell=0}^{L} \sqrt{\frac{\delta_{\varepsilon,\ell}}{h_\ell}} \leq C\delta^{-2} \sum_{\ell=0}^{L} \sqrt{h_\ell \varepsilon^2 + \varepsilon^4} \leq \tilde{C}\delta^{-2}(\varepsilon + \varepsilon^2 L), \tag{3.12}$$

where in the final inequality we used that $\sqrt{a+b} \leq \sqrt{a} + \sqrt{b}$ for non-negative 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 (3.10), and recognizing that we must have $n_\ell \geq 1$, yields

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

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

$$\sum_{\ell=0}^{L} n_\ell K \cdot h_\ell^{-1} = O\left(\sum_{\ell=0}^{L} \delta^{-2}\varepsilon\sqrt{\delta_{\varepsilon,\ell} h_\ell^{-1}} + \sum_{\ell=0}^{L} h_\ell^{-1}\right) = O\left(\varepsilon^2\delta^{-2} + \delta^{-1}\right). \tag{3.13}$$

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 (3.10) yields

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

Since $\delta^{-2}\varepsilon^2 L\sqrt{\delta_{\varepsilon,\ell} h_\ell} \geq 0.89$ 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

$$\sum_{\ell=0}^{L} n_\ell K \cdot h_\ell^{-1} = O(\varepsilon^4\delta^{-2}\log(1/\delta)^2). \tag{3.14}$$

3.2.3 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 (3.14), whereas the complexity (3.7) required for methods tailored to the small noise setting is

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

Hence, so long as

$$\varepsilon^2\log(1/\delta)^2 < \delta^{-\frac{1}{p}} + \delta^{-\frac{1}{q}}\varepsilon^{r/q} \tag{3.15}$$

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 (3.15), we note that if $p = 2$, then (3.15) holds so long as $\varepsilon < 0.65$, and if $p = 4$, then (3.15) holds so

long as $\varepsilon < 0.33$. In fact, under the further assumption that $\delta \approx e^{-\frac{1}{\varepsilon}}$ we see that (3.14) is $O(\varepsilon^2\delta^{-2})$, the same —asymptotically in the parameters δ or ε—as in the situation where we can generate independent realizations of $f(D^\varepsilon(T))$ exactly in a single step. As δ decreases below this threshold, the ratio between (3.14) 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 (3.13) 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 3.2.2 for standard Monte Carlo with Euler–Maruyama, and allows us to carry through a conclusion that applies to general SDEs [25]: multilevel Monte Carlo can improve on the complexity of standard Monte Carlo by a factor δ^{-1}, where δ 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 (3.7) 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 4.1 and in the paragraph above, we note that this multilevel Euler computational complexity is the same—asymptotically in the parameters δ or ε—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 (3.13) 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 3.2.1, 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 $O(\varepsilon^{2-\frac{5}{2}\rho})$. 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 (3.1) is a natural diffusion approximation of the jump process. Next, use the results of [4], 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 3.4.1, where we address this issue both analytically and computationally.

3.3 Proof of Theorem 3.1

Throughout this section, we assume the conditions of Theorem 3.1 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

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

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

$$D^\varepsilon_{h_\ell}(t_n^{k+1}) = D^\varepsilon_{h_\ell}(t_n^k) + \mu(D^\varepsilon_{h_\ell}(t_n^k))h_\ell + \varepsilon\sqrt{h_\ell}\sigma(D^\varepsilon_{h_\ell}(t_n^k))W_n^k, \tag{3.16}$$

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 (3.16) implies

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

To simulate $D^\varepsilon_{h_{\ell-1}}$, we then use

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

We begin with a series of necessary lemmas.

Lemma 3.2. *For any $T > 0$ we have*

$$\mathbb{E}\left[\sup_{0\le s\le T} |D^\varepsilon_{h_\ell}(s)|^4\right] \le C,$$

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

Proof. For any $t > 0$,

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

Thus,

$$\sup_{0\le s\le t}|D^{\varepsilon}_{h_\ell}(s)|^4 \le 27|D(0)|^4 + 27t^3\int_0^t \sup_{0\le r\le s}|\mu(D^{\varepsilon}_{h_\ell}(\eta_{h_\ell}(r)))|^4 ds + 27\varepsilon^4 \sup_{0\le s\le t}\left|\int_0^s \sigma(D^{\varepsilon}_{h_\ell}(\eta_{h_\ell}(r)))dW(r)\right|^4, \tag{3.17}$$

since the right-hand-side is monotonically increasing in t. Applying the Burkholder-Davis-Gundy inequality [51] to the term (3.17) and taking expectations we get

$$\mathbb{E}\left[\sup_{0\le s\le t}|D^{\varepsilon}_{h_\ell}(s)|^4\right] \le 27|D(0)|^4 + 27t^3\int_0^t \mathbb{E}\left[\sup_{0\le r\le s}|\mu(D^{\varepsilon}_{h_\ell}(\eta_{h_\ell}(r)))|^4\right] ds + K(T)\varepsilon^4\int_0^t \mathbb{E}[|\sigma(D^{\varepsilon}_{h_\ell}(\eta_{h_\ell}(s)))|^4]ds, \tag{3.18}$$

where $K(T)$ is a generic constant only depending on T. Using (3.18) with $t = nh_\ell$ and $s = mh_\ell$, where n and m are nonnegative integers for which $mh_\ell \le nh_\ell \le t \le T$, we get

$$\begin{aligned}\mathbb{E}\left[\sup_{m\le n}|D^{\varepsilon}_{h_\ell}(mh_\ell)|^4\right] &\le 27|D(0)|^4 + 27t^3\sum_{i=0}^{n-1}\mathbb{E}\left[\sup_{m\le i}|\mu(D^{\varepsilon}_{h_\ell}(mh_\ell))|^4\right]h_\ell \\ &\quad + K(T)\varepsilon^4\sum_{i=0}^{n-1}\mathbb{E}\left[|\sigma(D^{\varepsilon}_{h_\ell}(ih_\ell))|^4\right]h_\ell \\ &\le 27|D(0)|^4 + 54a^2T^4 + K(T)b^2\varepsilon^4 \\ &\quad + (54a^2T^3 + K(T)b^2\varepsilon^4)\sum_{i=0}^{n-1}\mathbb{E}\left[\sup_{m\le i}|D^{\varepsilon}_{h_\ell}(mh_\ell)|^4\right]h_\ell,\end{aligned}$$

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

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

Now we return to (3.18) and, after applying the growth conditions pertaining to both μ and σ in our running assumption, conclude

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

for some new constant C.

Let z_h be the deterministic solution to

$$z_h(t) = D(0) + \int_0^t \mu(z_h(\eta_h(s)))ds, \tag{3.19}$$

which is an Euler approximation to the ODE obtained from (3.1) when ε is set to zero.

Lemma 3.3. *For any $T > 0$ we have*

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

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

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

$$\begin{aligned}|D^\varepsilon_{h_\ell}(t) - z_{h_\ell}(t)|^2 \le 2T \int_0^t |\mu(D^\varepsilon_{h_\ell}(\eta_{h_\ell}(s))) - \mu(z_{h_\ell}(\eta_{h_\ell}(s))|^2 ds \\ + 2\varepsilon^2 \left|\int_0^t \sigma(D^\varepsilon_{h_\ell}(\eta_{h_\ell}(s)))dW(s)\right|^2.\end{aligned}$$

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

$$\begin{aligned}&\mathbb{E}\left[\sup_{0\le s\le t} \left|D^\varepsilon_{h_\ell}(s) - z_{h_\ell}(s)\right|^2\right] \\ &\le 2aT \int_0^t \mathbb{E}\left[\sup_{0\le s\le r} |D^\varepsilon_{h_\ell}(\eta_{h_\ell}(s)) - z_{h_\ell}(\eta_{h_\ell}(s))|^2\right] dr + 8\varepsilon^2 \int_0^t \mathbb{E}[|\sigma(D^\varepsilon_{h_\ell}(\eta_{h_\ell}(s)))|^2]ds \\ &\le 8bT\mathbb{E}\left[\sup_{0\le s\le t} (1 + |D^\varepsilon_{h_\ell}(s)|^2)\right]\varepsilon^2 + 2aT \int_0^t \mathbb{E}\left[\sup_{0\le s\le r} |D^\varepsilon_{h_\ell}(\eta_{h_\ell}(s)) - z_{h_\ell}(\eta_{h_\ell}(s))|^2\right] dr.\end{aligned} \tag{3.20}$$

Specializing the above to $t = nh_\ell$ and $s = mh_\ell$, where n and m are nonnegative

integers for which $mh_\ell \le nh_\ell \le t \le T$, we get

$$\begin{aligned}
&\mathbb{E}\left[\sup_{m\le n}|D^\varepsilon_{h_\ell}(mh_\ell) - z_{h_\ell}(mh_\ell)|^2\right] \\
&\qquad \le 8bT\mathbb{E}\left[\sup_{0\le s\le t}(1+|D^\varepsilon_{h_\ell}(s)|^2)\right]\varepsilon^2 + 2aT\sum_{i=0}^{n-1} h_\ell \cdot \mathbb{E}\left[\sup_{m\le i}|D^\varepsilon_{h_\ell}(mh_\ell) - z_{h_\ell}(mh_\ell)|^2\right] \\
&\qquad \le 8bT(1+K)\varepsilon^2 + 2aT\sum_{i=0}^{n-1} h_\ell \cdot \mathbb{E}\left[\sup_{m\le i}|D^\varepsilon_{h_\ell}(mh_\ell) - z_{h_\ell}(mh_\ell)|^2\right],
\end{aligned}$$

for some $K = K(a,b,T,D(0))$, where the first inequality follows from (3.20) and the second utilizes Lemma 3.2.

By the discrete version of Gronwall's inequality we see

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

Since n satisfying $nh_\ell \le T$ was arbitrary, we return to (3.20) to conclude that for any $0 \le t \le T$

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

Lemma 3.4.

$$\max_{\substack{0\le n\le M^{\ell-1}\\ 1\le k\le M}} |\mathbb{E}[D^\varepsilon_{h_\ell}(t^k_n) - D^\varepsilon_{h_\ell}(t_n)]| \le CMh_\ell,$$

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

Proof. Iterating (3.16) yields

$$\begin{aligned}
\left|\mathbb{E}\left[D^\varepsilon_{h_\ell}(t^k_n) - D^\varepsilon_{h_\ell}(t_n)\right]\right| &\le \left|\mathbb{E}\left[\sum_{i=0}^{k-1}\mu(D^\varepsilon_{h_\ell}(t^i_n))h_\ell\right]\right| + \left|\mathbb{E}\left[\varepsilon\sqrt{h_\ell}\sum_{i=0}^{k-1}\sigma(D^\varepsilon_{h_\ell}(t^i_n))W^i_n\right]\right| \\
&\le \sum_{i=0}^{k-1}\mathbb{E}\left[|\mu(D^\varepsilon_{h_\ell}(t^i_n))h_\ell|\right] \\
&\le h_\ell\sqrt{a}\sum_{i=0}^{k-1}(1+\mathbb{E}[|D^\varepsilon_{h_\ell}(t^i_n)|]),
\end{aligned}$$

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 3.2 and recalling that $k \leq M$. □

Lemma 3.5.

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

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

Proof. Iterating (3.16) yields

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

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

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

where the second inequality follows from Lemma 3.2 and Lemma 3.8 in [38], the last inequality follows from Lemma 3.2, 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 3.6. *Let $\mu_i(x)$ be the ith component of $\mu(x)$, then*

$$\mu_i(D^\varepsilon_{h_\ell}(t^k_n)) - \mu_i(D^\varepsilon_{h_\ell}(t_n)) = A_k + B_k + E_k, \tag{3.21}$$

where

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

$$B_k := \nabla\mu_i(D^\varepsilon_{h_\ell}(t_n)) \cdot \left(\varepsilon\sqrt{h_\ell} \sum_{j=0}^{k-1} \sigma(D^\varepsilon_{h_\ell}(t^j_n))W^j_n\right), \tag{3.22}$$

and

$$E_k := \left(\int_0^1 \int_0^s \left[\nabla^2\mu_i(D^\varepsilon_{h_\ell}(t_n) + r(D^\varepsilon_{h_\ell}(t^k_n) - D^\varepsilon_{h_\ell}(t_n)))(D^\varepsilon_{h_\ell}(t^k_n) - D^\varepsilon_{h_\ell}(t_n))\right] dr ds\right) \cdot \left(\varepsilon\sqrt{h_\ell} \sum_{j=0}^{k-1} \sigma(D^\varepsilon_{h_\ell}(t^j_n))W^j_n\right).$$

Proof. Using Taylor's expansion (see Lemma A.6 in the appendix) we see

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

Applying a multidimensional version of Lemma A.6,

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

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 [55] in the case where the numerical discretization method is Euler–Maruyama.

Lemma 3.7.

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

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

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

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

some manipulation the above implies

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

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

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

where we used that W_n^k is independent from $D^{\varepsilon}_{h_\ell}(t_n), D^{\varepsilon}_{h_{\ell-1}}(t_n)$, and $D^{\varepsilon}_{h_\ell}(t_n^k)$. Hence, by Lemma 3.5, there are positive constants C_1 and C_2 that only depend on $a, b, T, m, D(0)$, such that

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

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

$$\begin{aligned}
&\mathbb{E}[|D^{\varepsilon}_{h_\ell}(t_{n+1}) - D^{\varepsilon}_{h_{\ell-1}}(t_{n+1})|^2] \\
&\quad \leq \mathbb{E}[|D^{\varepsilon}_{h_\ell}(t_n) - D^{\varepsilon}_{h_{\ell-1}}(t_n)|^2] + (2\sqrt{a} + 4aMh_\ell + \varepsilon^2 4b)Mh_\ell \mathbb{E}[|D^{\varepsilon}_{h_\ell}(t_n) - D^{\varepsilon}_{h_{\ell-1}}(t_n)|^2] \\
&\qquad + 4aC_1M^4h_\ell^4 + 4aC_2\varepsilon^2M^3h_\ell^3 + 4bC_1\varepsilon^2M^3h_\ell^3 + 4bC_2\varepsilon^4M^2h_\ell^2 \\
&\qquad + 2h_\ell \sum_{k=0}^{M-1} \mathbb{E}[\langle D^{\varepsilon}_{h_\ell}(t_n) - D^{\varepsilon}_{h_{\ell-1}}(t_n), \mu(D^{\varepsilon}_{h_\ell}(t_n^k)) - \mu(D^{\varepsilon}_{h_\ell}(t_n))\rangle]. \qquad (3.23)
\end{aligned}$$

We turn to the term (3.23). Applying Lemma 3.6, we know

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

Also we notice,

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

where K_1 is a constant that only depends on $a, b, T, m, D(0)$. Utilizing Lemmas 3.2 and 3.5

$$\begin{aligned}
\mathbb{E}[|E_k|^2] &\leq ah_\ell\varepsilon^2\mathbb{E}\left[|D^{\varepsilon}_{h_\ell}(t_n^k) - D^{\varepsilon}_{h_\ell}(t_n)|^2 \sum_{j=0}^{k-1} |\sigma(D^{\varepsilon}_{h_\ell}(t_n^j))W_n^j|^2\right] \\
&\leq ah_\ell\varepsilon^2 \left(\mathbb{E}[|D^{\varepsilon}_{h_\ell}(t_n^k) - D^{\varepsilon}_{h_\ell}(t_n)|^4]\right)^{1/2} \left(k\sum_{j=0}^{k-1} \mathbb{E}\left[|\sigma(D^{\varepsilon}_{h_\ell}(t_n^j))W_n^j|^4\right]\right)^{1/2} \qquad (3.24) \\
&\leq K_2M^3h_\ell^3\varepsilon^2 + K_3M^2h_\ell^2\varepsilon^4,
\end{aligned}$$

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

$$\begin{aligned}
&2h_\ell \sum_{k=0}^{M-1} \mathbb{E}[\langle D^\varepsilon_{h_\ell}(t_n) - D^\varepsilon_{h_{\ell-1}}(t_n), \mu(D^\varepsilon_{h_\ell}(t^k_n)) - \mu(D^\varepsilon_{h_\ell}(t_n))\rangle] \\
&\qquad = 2h_\ell \sum_{k=0}^{M-1} \mathbb{E}[\langle D^\varepsilon_{h_\ell}(t_n) - D^\varepsilon_{h_{\ell-1}}(t_n), A_k\rangle] \\
&\qquad\quad + 2h_\ell \sum_{k=0}^{M-1} \mathbb{E}[\langle D^\varepsilon_{h_\ell}(t_n) - D^\varepsilon_{h_{\ell-1}}(t_n), B_k\rangle] \qquad (3.25) \\
&\qquad\quad + 2h_\ell \sum_{k=0}^{M-1} \mathbb{E}[\langle D^\varepsilon_{h_\ell}(t_n) - D^\varepsilon_{h_{\ell-1}}(t_n), E_k\rangle] \\
&\qquad \le 2Mh_\ell \mathbb{E}[|D^\varepsilon_{h_\ell}(t_n) - D^\varepsilon_{h_{\ell-1}}(t_n)|^2] + h_\ell \sum_{k=0}^{M-1} \mathbb{E}[|A_k|^2] + h_\ell \sum_{k=0}^{M-1} \mathbb{E}[|E_k|^2] \\
&\qquad \le 2Mh_\ell \mathbb{E}[|D^\varepsilon_{h_\ell}(t_n) - D^\varepsilon_{h_{\ell-1}}(t_n)|^2] + K_1 M^3 h_\ell^3 + K_2 M^4 h_\ell^4 \varepsilon^2 + K_3 M^3 h_\ell^3 \varepsilon^4,
\end{aligned}$$

where the first inequality follows from: (i) the observation that the expectation (3.25) is zero, (ii) the Cauchy-Schwarz inequality, and (iii) the inequality $2ab \le a^2 + b^2$. Combining all the estimates above, we find

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

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

We are now ready to prove our main result.

Proof of Theorem 3.1. Following [4], we first prove the result in the case that $f(x) = x_i$ for some $i \in \{1, \dots, d\}$. We have that for $n \leq M^{\ell-1} - 1$,

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

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,

$$
\begin{aligned}
&\mathsf{Var}([D^{\varepsilon}_{h_\ell}(t_{n+1}) - D^{\varepsilon}_{h_{\ell-1}}(t_{n+1})]_i) \\
&\quad \leq (1 + Mh_\ell)\mathsf{Var}([D^{\varepsilon}_{h_\ell}(t_n) - D^{\varepsilon}_{h_{\ell-1}}(t_n)]_i) \\
&\qquad + 4h_\ell^2 M \sum_{k=0}^{M-1} \mathsf{Var}(\mu_i(D^{\varepsilon}_{h_\ell}(t_n^k)) - \mu_i(D^{\varepsilon}_{h_\ell}(t_n))) && (3.26) \\
&\qquad + (4Mh_\ell + 1)Mh_\ell \mathsf{Var}(\mu_i(D^{\varepsilon}_{h_\ell}(t_n)) - \mu_i(D^{\varepsilon}_{h_{\ell-1}}(t_n))) && (3.27) \\
&\qquad + 4\varepsilon^2 h_\ell \sum_{k=0}^{M-1} \mathsf{Var}((\sigma_i(D^{\varepsilon}_{h_\ell}(t_n^k)) - \sigma_i(D^{\varepsilon}_{h_\ell}(t_n)))W_n^k) && (3.28) \\
&\qquad + 4\varepsilon^2 h_\ell \sum_{k=0}^{M-1} \mathsf{Var}((\sigma_i(D^{\varepsilon}_{h_\ell}(t_n)) - \sigma_i(D^{\varepsilon}_{h_{\ell-1}}(t_n)))W_n^k) && (3.29) \\
&\qquad + 2\mathsf{Cov}([D^{\varepsilon}_{h_\ell}(t_n) - D^{\varepsilon}_{h_{\ell-1}}(t_n)]_i, h_\ell \sum_{k=0}^{M-1} (\mu_i(D^{\varepsilon}_{h_\ell}(t_n^k)) - \mu_i(D^{\varepsilon}_{h_\ell}(t_n)))). && (3.30)
\end{aligned}
$$

We must bound each expression on the right-hand side in order to apply Gronwall's inequality. We first consider (3.28), which leads to a dominant term. Lemma 3.5 implies

that

$$\sum_{k=0}^{M-1} \mathsf{Var}((\sigma_i(D^{\varepsilon}_{h_\ell}(t_n^k)) - \sigma_i(D^{\varepsilon}_{h_\ell}(t_n)))W_n^k) \leq \sum_{k=0}^{M-1} \mathbb{E}[|\sigma_i(D^{\varepsilon}_{h_\ell}(t_n^k)) - \sigma_i(D^{\varepsilon}_{h_\ell}(t_n)))|^2]$$
$$\leq Mb(c_1 M^2 h_\ell^2 + c_2 \varepsilon^2 M h_\ell).$$

Similarly, by Lemma 3.7 we may bound (3.29), which also yields a dominant term,

$$\sum_{k=0}^{M-1} \mathsf{Var}((\sigma_i(D^{\varepsilon}_{h_\ell}(t_n)) - \sigma_i(D^{\varepsilon}_{h_{\ell-1}}(t_n)))W_n^k) \leq \sum_{k=0}^{M-1} \mathbb{E}[|\sigma_i(D^{\varepsilon}_{h_\ell}(t_n)) - \sigma_i(D^{\varepsilon}_{h_{\ell-1}}(t_n)))|^2]$$
$$\leq Mb(d_1 M^2 h_\ell^2 + d_2 \varepsilon^4 M h_\ell),$$

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

Turning to (3.26), we have the following lemma.

Lemma 3.8.

$$\mathsf{Var}\left(\mu_i(D^{\varepsilon}_{h_\ell}(t_n^k)) - \mu_i(D^{\varepsilon}_{h_\ell}(t_n))\right) \leq CMh_\ell\varepsilon^2,$$

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

Proof. From Lemma A.6 in the appendix (Taylor approximation), we have

$$\mu_i(D^{\varepsilon}_{h_\ell}(t_n^k)) - \mu_i(D^{\varepsilon}_{h_\ell}(t_n)) = \rho^k(t_n) \cdot (D^{\varepsilon}_{h_\ell}(t_n^k) - D^{\varepsilon}_{h_\ell}(t_n)), \tag{3.31}$$

where

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

In order to bound the right hand side of (3.31), we will apply Lemma A.4 in the appendix with $A^{\varepsilon,h_{\ell-1}} = [\rho^k(t_n)]_j$ and $B^{\varepsilon,h_{\ell-1}} = [D^{\varepsilon}_{h_\ell}(t_n^k) - D^{\varepsilon}_{h_\ell}(t_n)]_j$. Hence, we must find appropriate bounds on these components.

We begin with $B^{\varepsilon,h_{\ell-1}}$. We use Lemmas 3.2 and A.4 after iterating (3.16) to find

$$\begin{aligned}
&\mathsf{Var}([D^{\varepsilon}_{h_\ell}(t^k_n) - D^{\varepsilon}_{h_\ell}(t_n)]_j) \\
&\quad \leq 2\mathsf{Var}\left(h_\ell \sum_{r=0}^{k-1} \mu_j(D^{\varepsilon}_{h_\ell}(t^r_n))\right) + 2\mathsf{Var}\left(\varepsilon\sqrt{h_\ell}\sum_{r=0}^{k-1}\sigma_j(D^{\varepsilon}_{h_\ell}(t^r_n))W^r_n\right) \\
&\quad \leq 2h^2_\ell \mathsf{Var}\left(\sum_{r=0}^{k-1}(\mu_j(D^{\varepsilon}_{h_\ell}(t^r_n)) - \mu_j(z_{h_\ell}(t^r_n)))\right) + 2\varepsilon^2 h_\ell \mathbb{E}\left[\left|\sum_{r=0}^{k-1}\sigma_j(D^{\varepsilon}_{h_\ell}(t^r_n))W^r_n\right|^2\right] \\
&\quad \leq C_1 M^2 h^2_\ell \varepsilon^2 + C_2 M h_\ell \varepsilon^2,
\end{aligned} \tag{3.32}$$

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 A.3 in the appendix with $X_1(s) = D^{\varepsilon}_{h_\ell}(s)$, $X_2(s) = D^{\varepsilon}_{h_\ell}(\eta_{h_\ell}(s))$, $x_1(s) = z_{h_\ell}(s)$, $x_2(s) = z_{h_\ell}(\eta_{h_\ell}(s))$ and $u(x) = \nabla_j \mu_i(x)$ to obtain

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

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

We may now combine Lemma A.4 with Lemma 3.4 to conclude

$$\begin{aligned}
\mathsf{Var}\left([\rho^k(t_n)]_j \cdot [D^{\varepsilon}_{h_\ell}(t^k_n) - D^{\varepsilon}_{h_\ell}(t_n)]_j\right) &\leq \hat{C}KM^2h^2_\ell\varepsilon^2 + 15a\mathsf{Var}([D^{\varepsilon}_{h_\ell}(t^k_n) - D^{\varepsilon}_{h_\ell}(t_n)]_j) \\
&\leq (\hat{C}K + 15aC_1)M^2h^2_\ell\varepsilon^2 + 15C_2Mh_\ell\varepsilon^2 \\
&\leq \hat{C}_1 M h_\ell \varepsilon^2,
\end{aligned}$$

where $\hat{C}_1$ is positive and does not depend on ε and h_ℓ, and we applied (3.32) in the second inequality.

Returning to (3.31), the above allows us to conclude

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

We now turn to the first term of (3.27).

Lemma 3.9.

$$\begin{aligned}&\mathsf{Var}\left(\mu_i(D^{\varepsilon}_{h_\ell}(t_n)) - \mu_i(D^{\varepsilon}_{h_{\ell-1}}(t_n))\right)\\ &\quad\leq 15ad\sum_{j=1}^{d}\mathsf{Var}([D^{\varepsilon}_{h_\ell}(t_n) - D^{\varepsilon}_{h_{\ell-1}}(t_n)]_j) + K_1M^2h_\ell^2\varepsilon^2 + K_2Mh_\ell\varepsilon^6,\end{aligned}$$

where K_1, K_2 are positive constants that 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 A.4 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 A.3, we have

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

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 3.7 and A.4 we see there are positive constants K_1, K_2 depending only on $a, b, T, m, D(0)$, such that,

$$\begin{aligned}&\mathsf{Var}([\rho(t_n)]_j([D^{\varepsilon}_{h_\ell}(t_n) - D^{\varepsilon}_{h_{\ell-1}}(t_n)]_j))\\ &\quad\leq K_1M^2h_\ell^2\varepsilon^2 + K_2Mh_\ell\varepsilon^6 + 15a\mathsf{Var}([D^{\varepsilon}_{h_\ell}(t_n) - D^{\varepsilon}_{h_{\ell-1}}(t_n)]_j,\end{aligned}$$

and

$$\begin{aligned}&\mathsf{Var}(\mu_i(D^\varepsilon_{h_\ell}(t_n)) - \mu_i(D^\varepsilon_{h_{\ell-1}}(t_n)))\\ &\quad\le 15ad\sum_{j=1}^{d}\mathsf{Var}([D^\varepsilon_{h_\ell}(t_n) - D^\varepsilon_{h_{\ell-1}}(t_n)]_j) + d^2K_1M^2h_\ell^2\varepsilon^2 + d^2K_2Mh_\ell\varepsilon^6.\end{aligned}$$

□

Finally, we turn to the term (3.30).

Lemma 3.10.

$$\begin{aligned}&\mathsf{Cov}\Big([D^\varepsilon_{h_\ell}(t_n) - D^\varepsilon_{h_{\ell-1}}(t_n)]_i, h_\ell\sum_{k=0}^{M-1}(\mu_i(D^\varepsilon_{h_\ell}(t_n^k)) - \mu_i(D^\varepsilon_{h_\ell}(t_n)))\Big)\\ &\qquad\le Mh_\ell\mathsf{Var}([D^\varepsilon_{h_\ell}(t_n) - D^\varepsilon_{h_{\ell-1}}(t_n)]_i) + K_1M^3h_\ell^3\varepsilon^2 + K_2M^5h_\ell^5\varepsilon^2 + K_3M^3h_\ell^3\varepsilon^4,\end{aligned}$$

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 (3.21) in Lemma 3.6 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 (3.22), we have

$$\begin{aligned}&\mathsf{Cov}\Big([D^\varepsilon_{h_\ell}(t_n) - D^\varepsilon_{h_{\ell-1}}(t_n)]_i, h_\ell\sum_{k=0}^{M-1}(\mu_i(D^\varepsilon_{h_\ell}(t_n^k)) - \mu_i(D^\varepsilon_{h_\ell}(t_n)))\Big)\\ &\qquad= \mathsf{Cov}\left([D^\varepsilon_{h_\ell}(t_n) - D^\varepsilon_{h_{\ell-1}}(t_n)]_i, h_\ell\sum_{k=0}^{M-1}(A_k + E_k)\right)\\ &\qquad\quad+ \mathsf{Cov}\left([D^\varepsilon_{h_\ell}(t_n) - D^\varepsilon_{h_{\ell-1}}(t_n)]_i, h_\ell\sum_{k=0}^{M-1}B_k\right)\\ &\qquad\le Mh_\ell\mathsf{Var}([D^\varepsilon_{h_\ell}(t_n) - D^\varepsilon_{h_{\ell-1}}(t_n)]_i) + \frac{1}{2}h_\ell\sum_{k=0}^{M-1}\mathsf{Var}\left(A_k\right) + \frac{1}{2}h_\ell\sum_{k=0}^{M-1}\mathsf{Var}\left(E_k\right).\end{aligned} \tag{3.33}$$

First we want to estimate $\mathsf{Var}\,(A_k)$. Applying Lemma A.4 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,

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

where we also use Lemma A.4 for the last line. On the other hand, from (3.24)

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

Returning to (3.33), we see,

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

□

Now we return to (3.26)–(3.30) 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

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

Therefore,

$$\begin{aligned}
&\max_{i=1,2,\cdots,d} \mathsf{Var}\left([D^{\varepsilon}_{h_\ell}(t_{n+1}) - D^{\varepsilon}_{h_{\ell-1}}(t_{n+1})]_i\right) \\
&\qquad \leq \max_{i=1,2,\cdots,d} \mathsf{Var}\left([D^{\varepsilon}_{h_\ell}(t_n) - D^{\varepsilon}_{h_{\ell-1}}(t_n)]_i\right) + C_1 M^3 h_\ell^3 \varepsilon^2 + C_2 M^2 h_\ell^2 \varepsilon^4 \\
&\qquad\qquad + C_3 d M h_\ell \max_{i=1,2,\cdots,d} \mathsf{Var}\left([D^{\varepsilon}_{h_\ell}(t_n) - D^{\varepsilon}_{h_{\ell-1}}(t_n)]_i\right).
\end{aligned}$$

Applying Gronwall's lemma, we obtain,

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

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 A.6 in the appendix we have

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

We let $X_1(t) = D^{\varepsilon}_{h_\ell}(t)$, $X_2(t) = D^{\varepsilon}_{h_{\ell-1}}(t)$, $x_1(t) = z_{h_\ell}(t)$, $x_2(t) = z_{h_{\ell-1}}(t)$ and $u(x) = \nabla_j f(x)$ in an application of Lemma A.3 which yields

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

where K is a universal constant that depends on $C_L, D, a, b, T, D(0)$. Hence, by an application of Lemmas 3.7 and A.4 and the work above we see,

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

Thus

$$\mathsf{Var}(f(D^{\varepsilon}_{h_\ell}(t_n)) - f(D^{\varepsilon}_{h_{\ell-1}}(t_n))) \leq d^2(Kd_1 + 15dC_L^2C_1)M^2h_\ell^2\varepsilon^2 + d^2(15dC_L^2C_2 + Kd_2)h_\ell\varepsilon^4,$$

giving the result. □

3.4 Numerical examples and comparison with results related to jump processes

In this section we provide numerical evidence for the sharpness of both Theorem 3.1 and the computational complexity analyses provided in sections 3.2.1 and 3.2.2. Further, we compare our results to those found in [4] for scaled Markov processes.

Example 3.11. *We consider the following simple one dimensional model,*

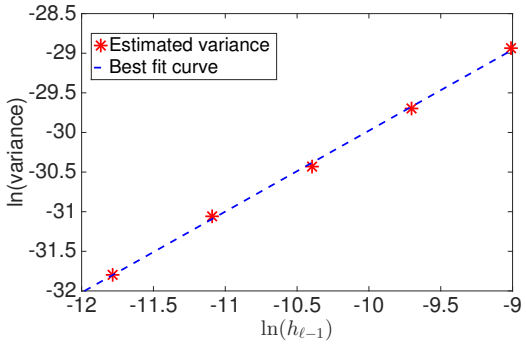
$$D^{\varepsilon}(t) = 1 - \int_0^t D^{\varepsilon}(s)ds + \varepsilon \int_0^t D^{\varepsilon}(s)dW(s),$$

where we simulate until $T = 1$.

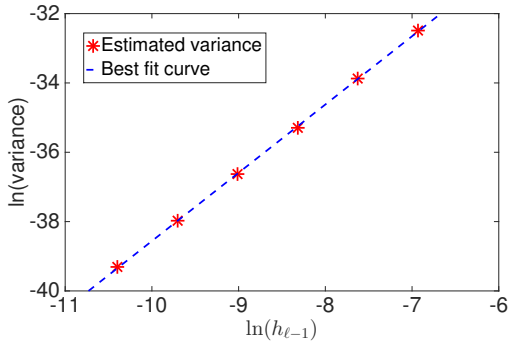
To gather evidence in support of the sharpness of the bound $\mathsf{Var}(D^{\varepsilon}_{h_\ell}(t) - D^{\varepsilon}_{h_{\ell-1}}(t)) = O(h^2\varepsilon^2 + h\varepsilon^4)$, *we fix one of* h *or* ε *in different scaling regimes and vary the other parameter in order to generate log-log plots. We 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 bounds.*

The exponent of h ***in*** $h\varepsilon^4$**.** *We fix* $\varepsilon = 2^{-6}$ *and vary*

$$h_{\ell-1} \in \{2^{-13}, 2^{-14}, 2^{-15}, 2^{-16}, 2^{-17}, 2^{-18}\}$$



(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 $y = 1.96x - 18.86$.

Figure 3: 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.*

to ensure $h_{\ell-1} \leq \varepsilon^2$*. As a result,* $h_{\ell-1}\varepsilon^4$ *is likely to be the dominant term in* (3.3)*. See Figure 3(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* $\varepsilon = 2^{-10}$ *and vary*

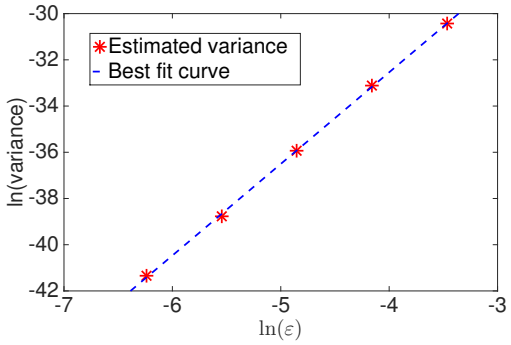
$$h_{\ell-1} \in \{2^{-10}, 2^{-11}, 2^{-12}, 2^{-13}, 2^{-14}, 2^{-16}\}$$

to ensure $h_{\ell-1} \geq \varepsilon^2$*. As a result,* $h^2_{\ell-1}\varepsilon^2$ *is likely to be the dominant term in* (3.3)*. See Figure 3(b), where the log-log plot is consistent with the functional form*

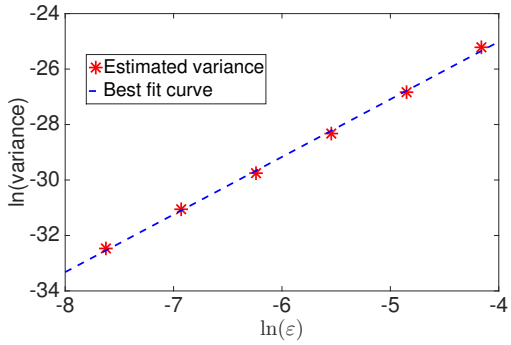
$$\mathsf{Var}(D^{\varepsilon}_{h_\ell}(1) - D^{\varepsilon}_{h_{\ell-1}}(1)) = O(h^2_{\ell-1}).$$

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

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



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



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

Figure 4: 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 1,000 independent samples.*

to ensure $h_{\ell-1} \leq \varepsilon^2$*. As a result,* $h_{\ell-1}\varepsilon^4$ *is likely to be the dominant term in* (3.3)*. See Figure 4(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* $h_{\ell-1} = 2^{-9}$ *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,* $h^2_{\ell-1}\varepsilon^2$ *is likely to be the dominant term in* (3.3)*. See Figure 4(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 sample trajectories in order to estimate* $\delta_{\varepsilon,\ell}$, *as defined in section 3.2.2. According to (3.10) and (3.12) 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.*

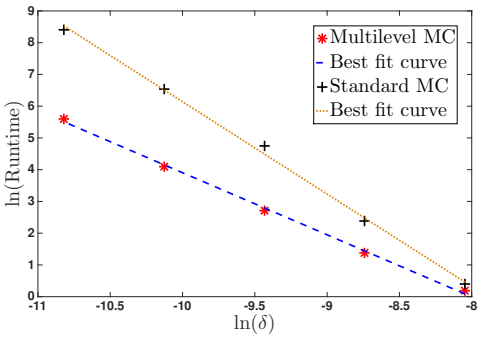
In Figures 5(a) and 6(a), we provide log-log plots of runtime (in seconds) and complexity (quantified by the total number of random variables utilized) for our implementation of multilevel and standard Monte Carlo with $\varepsilon = 0.1$ *fixed and*

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

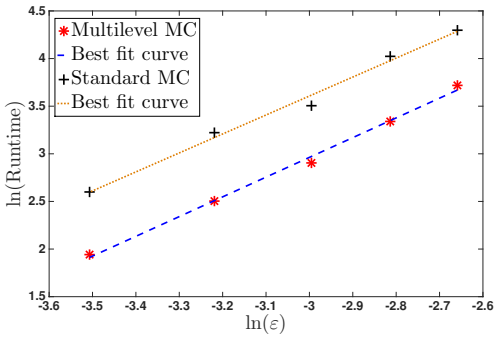
which ensures $\delta > \frac{1}{3} e^{-\frac{1}{\varepsilon}}$ *(see section 3.2.2). The best fit curves are consistent with the conclusion that the computational complexity of the Euler based multilevel Monte Carlo method is* $O(\delta^{-2})$ *while that of standard Monte Carlo method is* $O(\delta^{-3})$ *when* ε *is fixed. The Monte Carlo estimates which came from these simulations are detailed in 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 Figure 5(b) and 6(b), we provide similar log-log plots of runtime and computational complexity for Euler based multilevel Monte Carlo and standard Monte Carlo when $\delta =$

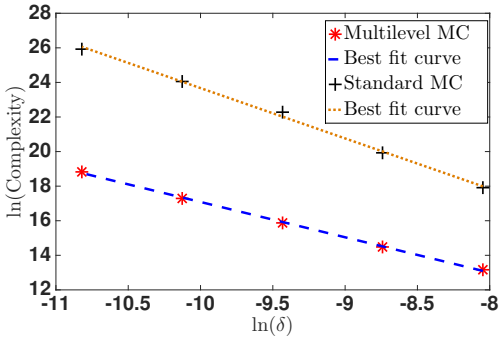


(a) $\varepsilon = 0.1$ held constant and δ varied. The best fit lines are $y = -1.96x - 15.65$ for Euler based MLMC and $y = -2.91x - 5.44$ for standard Euler based Monte Carlo.

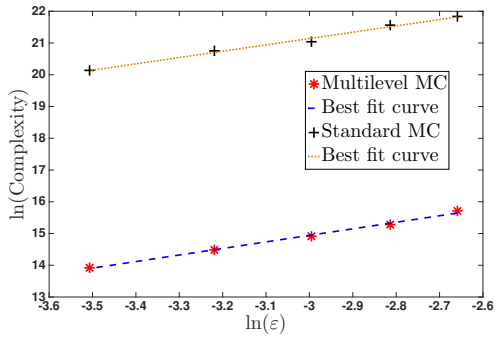


(b) $\delta = 2^{-14}$ held constant and ε varied. The best fit lines are $y = 2.07x + 9.18$ for Euler based MLMC and $y = 1.99x + 9.58$ for standard Euler based Monte Carlo.

Figure 5: Log-log plots of runtime (in seconds) for both multilevel and standard Euler based Monte Carlo.



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

2^{-14} *is fixed and* ε *is varied as*

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

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 Table 2.* □

3.4.1 Comparison with results for continuous time Markov chains

Diffusion processes with small noise structures of the form (3.1) 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 [4]. Since the diffusion approximation is naturally related to the jump process model, it is tempting to believe that the analysis found in [4] can be utilized to infer the results presented in this thesis. The following example and analysis shows that this intuition is incorrect

Example 3.12. *Consider a family of continuous time Markov chain models, parametrized*

δ	*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

Table 1: Result of Euler based multilevel Monte Carlo and Euler based Monte Carlo for fixed $\varepsilon = 0.1$ *and varying* δ. *The last two columns provide the standard deviations for the two estimators.*

ε	*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

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

by $N > 0$*, satisfying*

$$X^N(t) = X^N(0) + \frac{1}{N} Y_1 \left(N \int_0^t X_1^N(s)(X_1^N(s) - \tfrac{1}{N}) ds \right) \begin{bmatrix} -2 \\ 1 \end{bmatrix} + \frac{1}{N} Y_2 \left(N \int_0^t X_2^N(s) ds \right) \begin{bmatrix} 2 \\ -1 \end{bmatrix} \tag{3.34}$$

with $X^N(0) \in \frac{1}{N}\mathbb{Z}^2_{\geq 0}$ *and* Y_1, Y_2 *independent unit-rate Poisson processes. The process* (3.34) *can model the time evolution of the reaction network*

$$2A \underset{1}{\overset{1/N}{\rightleftarrows}} B, \tag{3.35}$$

in which two molecules of species A *can combine to form a molecule of species* B*, and vice versa [9, 10]. The specific choice of scaling in* (3.34) *is called the classical scaling for biochemical processes [9, 10]. One representation for the continuous in time Euler-Maruyama approximation of the standard diffusion approximation to the model* (3.34)

is

$$D_h^N(t) = D_h^N(0) + \int_0^t D_{h,1}^N(\eta_h(s))^2 ds \begin{bmatrix} -2 \\ 1 \end{bmatrix} + \int_0^t D_{h,2}^N(\eta_h(s)) ds \begin{bmatrix} 2 \\ -1 \end{bmatrix}$$
$$+ \varepsilon_N \int_0^t \sqrt{\max\{D_1^N(\eta_h(s))^2, 0\}}\, dW_1(s) \begin{bmatrix} -2 \\ 1 \end{bmatrix} \tag{3.36}$$
$$+ \varepsilon_N \int_0^t \sqrt{\max\{D_2^N(\eta_h(s)), 0\}}\, dW_2(s) \begin{bmatrix} 2 \\ -1 \end{bmatrix},$$

where $\varepsilon_N = \frac{1}{\sqrt{N}}$, W_1 *and* W_2 *are independent Brownian motions [9, 10]. Let* Z_h^N *be an Euler approximation to (3.34) and let* $M > 0$ *be some fixed positive integer. See [2] or [4] for a stochastic representation of* Z_h^N *that is similar to (3.36), and for the relevant coupling between* $Z_{h_\ell}^N$ *and* $Z_{h_{\ell-1}}^N$*. By Corollary 1 in [4], we have that for* $h_\ell = M^{-\ell}$

$$Var(Z_{h_\ell,1}^N(t) - Z_{h_{\ell-1},1}^N(t)) \leq D \cdot h_{\ell-1} N^{-1} = D \cdot h_{\ell-1} \varepsilon_N^2, \tag{3.37}$$

where the constant D *does not depend upon* N, h_ℓ *, or* $h_{\ell-1}$*. Conversely, Theorem 3.1 allows us to conclude that*

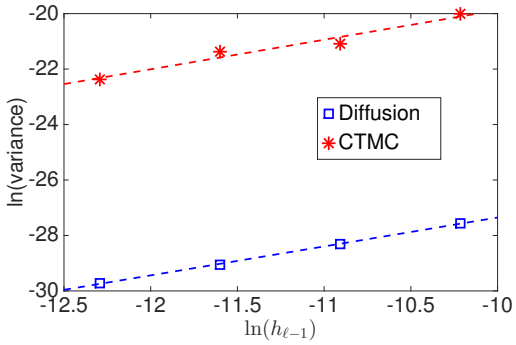
$$Var(D_{h_\ell,1}^N(t) - D_{h_{\ell-1},1}^N(t)) \leq C_1 \cdot h_{\ell-1}^2 \varepsilon_N^2 + C_2 \cdot h_{\ell-1} \varepsilon_N^4. \tag{3.38}$$

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,*

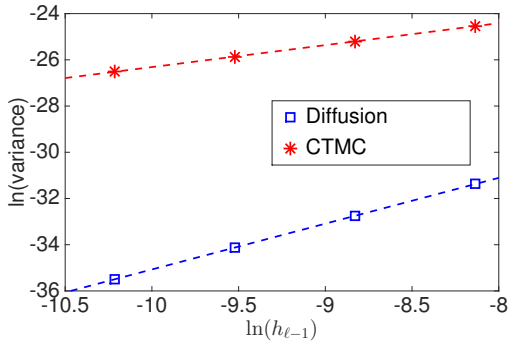
$$\frac{h_{\ell-1}^2 \varepsilon_N^2 + h_{\ell-1} \varepsilon_N^4}{h_{\ell-1} \varepsilon_N^2} = h_{\ell-1} + \varepsilon_N^2,$$

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 (3.37) and (3.38), we follow the numerical analysis performed in Example 3.11 by varying h_ℓ *and* $\varepsilon_N = \frac{1}{\sqrt{N}}$ *in different*



(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 $y = 1.98x - 15.25$.

Figure 7: Log-log plots of $\mathsf{Var}(D^N_{h_\ell,1}(T) - D^N_{h_{\ell-1},1}(T))$ *and* $\mathsf{Var}(Z^N_{h_\ell,1}(T) - Z^N_{h_{\ell-1},1}(T))$ *with* ε_N *held constant and* $h_{\ell-1}$ *varied. The best fit curves for the data are overlain with dashed lines.*

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 3.11, 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 (3.38). See Figure 11(a), where the log-log plots are consistent with the functional forms*

$$\mathsf{Var}(D^N_{h_\ell,1}(T) - D^N_{h_{\ell-1},1}(T)) = O(h_{\ell-1}), \quad \mathsf{Var}(Z^N_{h_\ell,1}(T) - Z^N_{h_{\ell-1},1}(T)) = O(h_{\ell-1}).$$

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* (3.38). *See Figure 11(b), where the log-log plots are consistent with the functional forms*

$$\mathsf{Var}(D^N_{h_\ell,1}(T) - D^N_{h_{\ell-1},1}(T)) = O(h_{\ell-1}^2), \quad \mathsf{Var}(Z^N_{h_\ell,1}(T) - Z^N_{h_{\ell-1},1}(T)) = O(h_{\ell-1}).$$

The exponent of ε ***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* (3.38). *See Figure 8(a), where the log-log plot is consistent with the functional form*

$$\mathsf{Var}(D^N_{h_\ell,1}(T) - D^N_{h_{\ell-1},1}(T)) = O(\varepsilon_N^4), \quad \mathsf{Var}(Z^N_{h_\ell,1}(T) - Z^N_{h_{\ell-1},1}(T)) = O(\varepsilon_N^2)$$

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

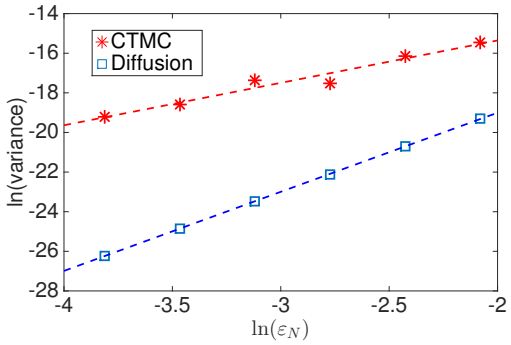
$$N \in \{2^{20}, 2^{21}, 2^{22}, 2^{23}, 2^{24}, 2^{25}\}$$

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* (3.38). *See Figure 8(b), where the log-log plot is consistent with the functional form*

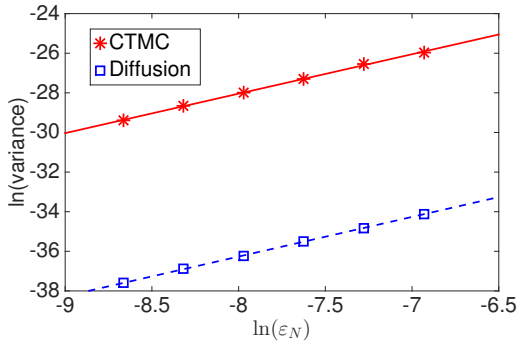
$$\mathsf{Var}(D^N_{h_\ell,1}(T) - D^N_{h_{\ell-1},1}(T)) = O(\varepsilon_N^2), \quad \mathsf{Var}(Z^N_{h_\ell,1}(T) - Z^N_{h_{\ell-1},1}(T)) = O(\varepsilon_N^2).$$

3.5 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



(a) $h_{\ell-1} = 2^{-12}$ fixed while ε_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 ε_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 $y = 2.00x - 20.25$.

Figure 8: Log-log plots of $\mathsf{Var}(D^N_{h_\ell,1}(T) - D^N_{h_{\ell-1},1}(T))$ *and* $\mathsf{Var}(Z^N_{h_\ell,1}(T) - Z^N_{h_{\ell-1},1}(T))$ *with* $h_{\ell-1}$ *held constant and* ε_N *varied. The best fit curves for all data are overlain with dashed lines.*

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 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 [15] and [58] can be adapted to the small noise regime.

Chapter 4

Computational complexity analysis for Monte Carlo approximations of classically scaled population processes

4.1 Introduction

For some large $N_0 > 0$ we consider a continuous time Markov chain satisfying the stochastic equation

$$X^{N_0}(t) = X^{N_0}(0) + \sum_{k=1}^{K} \frac{1}{N_0} Y_k \left(N_0 \int_0^t \lambda_k(X^{N_0}(s))ds \right) \zeta_k, \tag{4.1}$$

where $X^{N_0}(t) \in \mathbb{R}^d$, $K < \infty$, the Y_k are independent unit Poisson processes and, for each k, $\zeta_k \in \mathbb{R}^d$ and $\lambda_k : \mathbb{R}^d \to \mathbb{R}_{\geq 0}$ are Lipschitz continuous with bounded second derivatives. We further assume that $\lambda_k(X^N(0)) = O(1)$. This particular scaling choice is often termed the "classical scaling" [10, 44, 45]. We consider the task of numerically approximating $\mathbb{E}[f(X^{N_0}(t))]$, in the sense of confidence intervals, to some fixed tolerance $\varepsilon_0 < 1$, where f satisfies mild regularity conditions.

The class of models of the form (4.1) satisfying the above assumptions has a long history in terms of modelling [13, 14, 19, 54], analysis [10, 44, 45] and computation [27, 28]. The framework covers many application areas, including population dynamics [63], queueing theory [64], and several branches of physics [22]. In recent years, chemical and biochemical kinetics models in systems biology [71] have been the driving force behind a resurgence of activity in algorithmic developments, including tau-leaping [31] and its multilevel extension [3, 4]. In this setting, the parameter N_0 in (4.1) can represent Avagadro's number multplied by the volume, and in this classical scaling, species are measured in moles per liter. More generally, however, N_0 can just be considered a large number, often of the order 100s or 1000s. Our aim here is to present a unified computational complexity analysis for a range of Monte Carlo based methods. This allows us to make what we believe are the first concrete conclusions pertaining to the relative merits of current methods in a practically relevant asymptotic regime. Of particular note is that our analysis suggests that a diffusion approximation offers very few advantages from a computational standpoint.

In section 4.2, we discuss some of the issues involved in quantifying computational complexity in the present setting, and introduce a novel scaling regime in which clear-cut comparisons can be made. Further, a high-level summary of our main conclusions is presented. In section 4.3, we summarize two relevant methods for approximating the model (4.1): the tau-leap discretization method, and the Langevin or Diffusion approximation. In section 4.4, we quantify the computational complexity of using exact simulation, tau-leaping, and simulation of the diffusion equation with standard Monte Carlo for approximating $E[f(X^{N_0}(t))]$ to a desired tolerance. Further, in subsection 4.4.2 we review the more recent multilevel methods and quantify the benefits of their

use in both the tau-leaping and diffusion scenarios. In section 4.5, we provide numerical examples demonstrating our main conclusions. In section 4.6, we close with some brief conclusions.

4.2 Scaling regime and summary of results

In order to motivate our analysis and computations, we begin with a brief, high-level, overview. In particular, we discuss the entries in Table 3, which summarizes the key results of this work. Full details are given later in the manuscript, however we point out here that the terms in Table 3 include assumptions on the variances of the constituent processes that will be made clear, and proven in a wide variety of cases, in the subsequent sections.

A natural approach to approximate the desired expectation is to simulate paths exactly, for example with the stochastic simulation algorithm [27, 28] or the next reaction method [1, 23], in order to obtain independent sample paths $\{X^{N_0}_{[i]}\}_{i=1}^{n}$ that can be combined into a sample average

$$\hat{\mu}_n = \frac{1}{n}\sum_{i=1}^{n} f(X^{N_0}_{[i]}). \tag{4.2}$$

This becomes problematic if the cost of each sample path is high—to follow a path exactly we must take account of each individual transition in the process. This is a serious issue when many jumps take place, which is the case when N_0 is large.

The essence of the Euler tau-leaping approach is to fix the system intensities over time intervals of length h, and thereby only require the generation of K Poisson random variables per time interval [31]. In order to analyse the benefit of tau-leaping, and related

methods, Anderson, Ganguly, and Kurtz [2] considered the limit $N \to \infty$ and $h \to 0$ with $N = h^{-\beta}$ for some $\beta > 0$. To see why such a limit is useful we note two facts:

- If, instead, we allow $N \to \infty$ with h fixed, then the stochastic fluctuations become negligible [9, 44]. In this *thermodynamic limit* the model reduces to a deterministic ODE, so a simple deterministic numerical method could be used.

- If, instead, we allow $h \to 0$ with N fixed then tau-leaping becomes arbitrarily inefficient. The "empty" waiting times between reactions, which have nonzero expected values, are being needlessly refined by the discretization method.

The relation $N = h^{-\beta}$ brings together the large system size effect (where exact simulation is expensive and tau-leaping offers a computational advantage) with the the small h effect (where the accuracy of tau-leaping can be analysed). This gives a realistic setting where the benefits of tau-leaping can be quantified. It may then be shown [2, Theorem 4.1] that the bias arising from Euler tau-leaping is $O(h) = O(N^{-\beta})$.

For our purposes, rather than the step size h of a particular approximate method, it is more natural to work in terms of the system size, N_0. First, let $\alpha > 0$ satisfy $\varepsilon_0 = N_0^{-\alpha}$. Next, consider the following family of models parameterized by $N \geq N_0$,

$$X^N(t) = X^N(0) + \sum_{k=1}^{K} \frac{1}{N} Y_k \left(N \int_0^t \lambda_k(X^N(s)) ds \right) \zeta_k. \tag{4.3}$$

We will study the asymptotic behavior, as $N \to \infty$, of the computational complexity required to approximate $\mathbb{E}[f(X^N(t))]$ to a tolerance of

$$\varepsilon_N = N^{-\alpha}. \tag{4.4}$$

We emphasize at this stage that we are no longer studying a fixed model. Instead we look at the family of models (4.3) parameterized through the system size N, and consider

the limit, as $N \to \infty$, of the computational complexity of the different methods under the accuracy requirement (4.4). The computed results then tell us, to leading order, the costs associated with solving our fixed problem (4.1) with accuracy requirement $N_0^{-\alpha}$.

4.2.1 Summary of results

To quantify computational complexity, we define the "cost-per-path" to be the number of random variables generated in the simulation of a single path. We argue in section 4.4.1 that the expected cost-per-path of exact simulation scales linearly with N. Tau-leaping, by contrast, has a cost-per-path of $O(h^{-1})$, and we need $h \sim \varepsilon_N$ to get a small enough bias in our standard Monte Carlo approximation. Since the random variables of interest, $f(X^N(t))$, generally have a variance that scales like $O(N^{-1})$, see [4], the costs of standard Monte Carlo methods then become:

Monte Carlo plus exact simulation: $O(N^{-1}\varepsilon_N^{-2} + 1)$ paths at a cost of $O(N)$ per path, totaling a computational complexity of $O(\varepsilon_N^{-2} + N)$ or $O(N^{2\alpha} + N)$.

Monte Carlo plus tau-leaping: $O(N^{-1}\varepsilon_N^{-2} + 1)$ paths at a cost of $O(\varepsilon_N^{-1})$ per path, totaling a computational complexity of $O(N^{-1}\varepsilon_N^{-3} + \varepsilon_N^{-1})$ or $O(N^{3\alpha-1} + N^{\alpha})$,

as summarized in the first two rows of Table 3. Note that the "+1" terms above account for the requirement that we cannot generate less than one path. In this regime, we see that tau-leaping is beneficial for $\alpha < 1$. This makes sense intuitively. If we ask for too much accuracy relative to the system size ($\alpha > 1$ in (4.4)) then tau-leaping's built-in bias outweighs its cheapness, or, equivalently, the required stepsize is so small that tau-leaping works harder than exact simulation.

Monte Carlo method	Complexity	unbiased?	Most efficient
MC + exact simulation	$O(N^{2\alpha} + N)$	Yes	Never
MC + tau-leaping	$O(N^{3\alpha-1} + N^{\alpha})$	No	Never
MC + midpt. or trap. tau-leap	$O(N^{2.5\alpha-1} + N^{\alpha/2})$	No	$\frac{1}{2} < \alpha \leq \frac{2}{3}$
MC + Euler for diff. approx.	$O(N^{3\alpha-1} + N^{\alpha})$	No	Never
MLMC + E-M for diff. approx.	$O(N^{2\alpha-1} + N^{\alpha})$	No	$\alpha \geq \frac{2}{3}$
biased MLMC tau-leaping	$O(N^{2\alpha-1}\log(N)^2 + N^{\alpha})$	No	$\alpha \geq \frac{2}{3}$
unbiased MLMC tau-leaping	$O(N^{2\alpha-1}\log(N)^2 + N)$	Yes	$\alpha \geq 1$

Table 3: Computational cost for different Monte Carlo methods, as $N \to \infty$. The final column indicates when each method is most efficient, in terms of the parameter α, up to factors involving logarithms.

Higher order alternatives to the original tau-leaping method [31] are available. For example, a mid-point discretization [2, Theorem 4.2] or a trapezoidal method [8] both achieve $O(h^2) = O(N^{-2\alpha})$ bias, which allows the overall complexity to be reduced to $O(N^{2.5\alpha-1} + N^{\alpha/2})$ in this regime. This makes midpoint or trapezoidal tau-leaping beneficial as compared with exact simulation for $\alpha < 2$.

As an alternative to tau-leap discretizations, we could replace the continuous-time Markov chain by a diffusion approximation and use a numerical stochastic differential equation (SDE) simulation method to generate approximate paths. Even assuming that such an approximation produces an unbiased and consistent estimator (which is not, in general, the case) we find that an Euler method with standard Monte Carlo gives the same cost as basic tau-leaping; see section 4.4.1.

Multilevel versions of Monte Carlo (MLMC) have been developed in both the Euler + diffusion setting [24, 25] and tau-leaping [3]. Under mild assumptions on f and the stochastic system, we find that a multilevel scheme can reduce the standard Monte

Carlo Euler/diffusion cost from $O(N^{3\alpha-1}+N^{\alpha})$ to $O(N^{2\alpha-1}\log(N)^2+N^{\alpha})$. Similarly, the biased version of the multilevel algorithm at the tau-leaping level from [3] can also give a complexity of $O(N^{2\alpha-1}\log(N)^2+N^{\alpha})$, and this is achieved without perturbing the basic model by taking a diffusion approximation. Further, an *unbiased* version of the multilevel algorithm yields a complexity of

$$O(N^{2\alpha-1}\log(N)^2+N)$$

which is equivalent to the biased version when $\alpha \geq 1$.

We also mention that a crude and inexpensive approximation to the required expected value can be computed by simply simulating the deterministic mass action ODE approximation to (4.1), which is often referred to as the reaction rate equation. Ignoring all fluctuations in this manner, using a perturbed model that no longer depends upon N, produces an approximation to the mean that has an error of $O(1/\sqrt{N})$, meeting the required $O(N^{-\alpha})$ accuracy requirement when $\alpha \leq \frac{1}{2}$. In this case, solving the deterministic ODE with a pth order numerical method scheme (such as a Runge–Kutta method) requires a stepsize with $h^p = \varepsilon_N$, so that $h = N^{-\alpha/p}$, giving an asymptotic computational cost of $O(h^{-1})$; that is, $O(N^{\alpha/p})$. Since the order p can be made arbitrarily large, we see that this approach has negligible cost. For this reason, we view $\alpha = \frac{1}{2}$ as a natural cut-off in the relationship (4.4); we are not concerned with $\alpha \leq \frac{1}{2}$ since in this regime the requested level of accuracy does not require fluctuations to be respected.

In addition to the asymptotic complexity counts in Table 3, another important feature of a method is the availability of computable *a posteriori* confidence interval information. As indicated in the table, two of the methods considered here, exact simulation with Monte Carlo and an appropriately constructed multilevel tau-leaping, are unbiased.

The sample mean, accompanied by an estimate of the overall variance, can then be delivered with a computable confidence interval. By contrast, the remaining methods in the table are biased: tau-leaping and Euler introduce discretization errors and the diffusion approximation perturbs the underlying model. Although the asymptotic leading order of these biases can be estimated, useful a posteriori upper bounds cannot be computed straightforwardly in general, making these approaches much less attractive for reliably achieving a target accuracy.

Based on the range of methods analysed here in an asymptotic regime that couples system size and target accuracy, two key messages are

- even assuming there is no bias to the underlying model, simulating at the level of the the diffusion approximation is only marginally advantageous,
- tau-leaping can offer advantages over exact simulation, and an appropriately designed version of multilevel tau-leaping (which combines exact and tau-leaped samples) offers an unbiased method that is efficient over a wide range of accuracy requirements.

4.3 Approximation methods

In this section, we briefly review two alternatives to exact simulation of (4.3).

4.3.1 Tau-Leaping

Tau-leaping [31] is a computational method that generates Euler-style approximate paths for the continuous-time Markov chain (4.3). The basic idea is to hold the intensity

functions fixed over a time interval $[t_n, t_n + h]$ at the values $\lambda_k(X^N(t_n))$, where $X^N(t_n)$ is the state of the system at time t_n, and, under this assumption, compute the number of times each reaction takes place over this period. As the waiting times for the reactions are exponentially distributed, this leads to the following algorithm, which simulates up to a time of $T > 0$. For $x \geq 0$ we will write Poisson(x) to denote a sample from the Poisson distribution with parameter x, with all such samples being independent of each other and of all other sources of randomness used.

Algorithm 3 (Euler tau-leaping)**.** *Fix $h > 0$. Set $Z_h^N(0) = x_0$, $t_0 = 0$, $n = 0$ and repeat the following until $t_{n+1} = T$:*

(*i*) *Set $t_{n+1} = t_n + h$. If $t_{n+1} \geq T$, set $t_{n+1} = T$ and $h = T - t_n$.*

(*ii*) *For each k, let $\Lambda_k = Poisson(\lambda_k(Z_h^N(t_n))h)$.*

(*iii*) *Set $Z_h^N(t_{n+1}) = Z_h^N(t_n) + \sum_k \Lambda_k \zeta_k$.*

(*iv*) *Set $n \leftarrow n + 1$.*

Analogously to (4.3), a path-wise representation of Euler tau-leaping defined for all $t \geq 0$ can be given through a random time change of Poisson processes:

$$Z_h^N(t) = Z_h^N(0) + \sum_k Y_k \left(\int_0^t \lambda_k(Z_h^N(\eta_h(s)))ds \right) \zeta_k, \tag{4.5}$$

where the Y_k are as before, and $\eta_h(s) \stackrel{\text{def}}{=} \left\lfloor \frac{s}{h} \right\rfloor h$. Thus, $Z_h^N(\eta_h(s)) = Z_h^N(t_n)$ if $t_n \leq s < t_{n+1}$.

4.3.2 Diffusion approximation

The tau-leaping algorithm utilizes a time-stepping method to directly approximate the underlying model (4.3). Alternatively, a diffusion approximation arises by perturbing the underlying model into one which can be discretized more efficiently.

Define the function F via

$$F(x) = \sum_k \lambda_k(x)\zeta_k.$$

By the functional central limit theorem,

$$\frac{1}{\sqrt{N}}\left[Y_k(Nu) - Nu\right] \approx W_k(u), \tag{4.6}$$

where W_k is a standard Brownian motion. Applying (4.6) to (4.3) yields

$$X^N(t) \approx X^N(0) + \int_0^t F(X^N(s))ds + \sum_k \frac{1}{\sqrt{N}} W_k\left(\int_0^t \lambda_k(X^N(s))ds\right)\zeta_k,$$

where the W_k are independent standard Brownian motions. This implies that X^N can be approximated by the process D^N satisfying

$$D^N(t) = D^N(0) + \int_0^t F(D^N(s))ds + \sum_k \frac{1}{\sqrt{N}} W_k\left(\int_0^t \lambda_k(D^N(s))ds\right)\zeta_k,$$

where $D^N(0) = X^N(0)$. An equivalent, and more prevalent, way to represent D^N is via the Itô representation

$$D^N(t) = D^N(0) + \int_0^t F(D^N(s))ds + \sum_k \frac{1}{\sqrt{N}}\zeta_k \int_0^t \sqrt{\lambda_k(D^N(s))}dW_k(s), \tag{4.7}$$

which is often written in the differential form

$$dD^N(t) = F(D^N(t))dt + \sum_k \frac{1}{\sqrt{N}}\zeta_k\sqrt{\lambda_k(D^N(s))}dW_k(s). \tag{4.8}$$

The SDE system (4.8) is known as a *Langevin* approximation in the biology and chemistry literature, and a *diffusion* approximation in probability [10, 71]. We note the following points.

- The diffusion coefficient, often termed the "noise" in the system, is $O(\frac{1}{\sqrt{N}})$, and hence, in our setting is small relative to the drift.

- The diffusion coefficient involves square roots. Hence, it is critical that the choice of intensity functions λ_k only take values in $\mathbb{R}_{\geq 0}$ on the domain of the solution. This is of particular importance in the population process setting where the solutions of the underlying model (4.3) naturally satisfy a non negativity constraint whereas the SDE solution paths cannot be guaranteed to remain non-negative in general. Therefore, in this setting of population processes one reasonable representation, of many, would be

$$dD^N(t) = F(D^N(t))dt + \sum_k \frac{1}{\sqrt{N}}\zeta_k \sqrt{[\lambda_k(D^N(s))]^+}dW_k(s), \tag{4.9}$$

  where $[x]^+ = \max\{x, 0\}$.

- The coefficients of the SDE are not globally Lipschitz in general, and hence standard convergence theory for numerical methods, such as that in [43], is not applicable. Examples of nonlinear SDEs for which standard Monte Carlo and multilevel Monte Carlo, when combined with Euler-Maruyama discretization, fail to produce a convergent algorithm have been pointed out in the literature [36, 37]. The question of which classes of reaction system lead to well-defined SDEs and which discretizations converge at the traditional rate therefore remains open.

In this work, to get a feel for the best possible computational complexity that can arise from the Langevin approximation, we will study the case where the function F is linear. In this case the first two moments of the diffusion approximation match those of the underlying continuous time Markov chain exactly [10]. We will find that even in

this idealized light, the asymptotic computational complexity of Euler-Maruyama on a diffusion approximation combined with either a standard or a multilevel implementation is only marginally better than the corresponding computational complexity bounds for multilevel tau-leaping. In particular, they differ only in a log factor. We note that we only make the linearity assumption on the diffusion model.

4.4 Complexity analysis

In this section we establish the results given in Table 3. In subsection 4.4.1, we derive the first four rows, whereas in subsection 4.4.2 we discuss the multilevel framework and establish rows five, six, and seven.

4.4.1 Complexity analysis of standard Monte Carlo approaches

Exact Sampling and Monte Carlo

Suppose that we compute exact samples from the process $X^N(t)$. The intensity functions scale like $O(N)$ and the expected holding time between reactions is $O(N^{-1})$. Hence the number of system updates required to generate a single path is $O(N)$. Letting

$$\delta_N = \mathsf{Var}(f(X^N(t)))$$

we require

$$n^{-1}\delta_N = O(\varepsilon_N^2) \implies n = O(\delta_N \varepsilon_N^{-2} + 1).$$

Thus, the total computational complexity of making the desired approximation is

$$O(nN) = O(\delta_N \varepsilon_N^{-2} N + N) = O(\delta_N N^{2\alpha+1} + N).$$

In [4] it is shown that $\mathsf{Var}(X_i^N(t)) = O(N^{-1})$, giving $\delta_N = N^{-1}$ and an overall complexity of $O(N^{2\alpha} + N)$, as given in the first row of Table 3.

Tau-leaping and Monte Carlo

Suppose now that we use n paths of the tau-leaping process (4.5) to construct the Monte Carlo estimator $\hat{\mu}_n$ for $\mathbb{E}[f(X^N(t))]$. We note that

$$\mathbb{E}[f(X^N(t))] - \hat{\mu}_n = (\mathbb{E}[f(X^N(t))] - \mathbb{E}[f(Z_h^N(t))]) + (\mathbb{E}[f(Z_h^N(t))] - \hat{\mu}_n). \tag{4.10}$$

Tau-leaping was studied in [2] in a regime of the form $h = O(\varepsilon_N)$. By constraining $h = O(\varepsilon_N)$ we can apply [2, Theorem 4.1] to give

$$\mathbb{E}[f(X^N(t))] - \mathbb{E}[f(Z_h^N(t))] = O(h) = O(\varepsilon_N),$$

for a wide class of functionals f, as required for the bias. Letting

$$\delta_{N,h} = \mathsf{Var}(f(Z_h^N(t)))$$

we again require $n = O(\delta_{N,h}\varepsilon_N^{-2} + 1)$ to control the statistical error. Since there are $O(h^{-1})$ operations per path generation, the total computational complexity for making the desired approximation is

$$O(nh^{-1}) = O(\delta_{N,h}\varepsilon_N^{-3} + \varepsilon_N^{-1}).$$

We know from [4] that $\mathsf{Var}(Z_{h,i}^N(t)) = O(N^{-1})$, giving an overall complexity of $O(N^{3\alpha-1} + N^{\alpha})$, as reported in the second row of Table 3.

Weakly second order extensions to the tau-leaping method can lower the computational complexity dramatically. For example, if we use the midpoint tau-leaping process

$\mathcal{Z}_h^N$ from [2], we can set $h = \sqrt{\varepsilon_N}$ and still achieve a bias of $O(\varepsilon_N)$. In particular, from [2, Theorem 4.2], we have

$$\mathbb{E}[f(X^N(t))] - \mathbb{E}[F(\mathcal{Z}_h^N(t))] = O(h^2),$$

for a wide class of f. By similar methods as in [4], it can be shown that $\mathcal{Z}_h^N$ has the same variance scaling as X_h^N and Z_h^N. Since we need $n = O(\delta_{N,h}\varepsilon_N^{-2} + 1)$ paths, the complexity is

$$O(n \cdot h^{-1}) = O(n \cdot \varepsilon_N^{-1/2}) = O(\delta_{N,h}\varepsilon_N^{-2.5} + \varepsilon_N^{-1/2}).$$

When $\delta_{N,h} = N^{-1}$ the resulting complexity is $O(N^{2.5\alpha-1} + N^{\alpha/2})$, as stated in the third row of Table 3. The same conclusion can also be drawn for the trapezoidal method in [8].

Diffusion approximation and Monte Carlo

In considering an Euler–Maruyama discretization of the diffusion approximation, it is important to keep in mind that we are studying a parameterized class of SDEs in the $N \to \infty$ limit, rather than a single SDE. However, because the drift remains $O(1)$, the weak error stays at the $O(h)$ level arising from the underlying deterministic Euler method; that is, Euler-Maruyama introduces a bias of $O(h)$, uniformly in N. So we need $h = O(\varepsilon_N)$ to keep the bias within our tolerance. Letting

$$\delta_{N,h} = \mathsf{Var}(f(D_h^N(t))),$$

we require $n = O(\delta_{N,h}\varepsilon_N^{-2} + 1)$ to control the statistical error. Since there are $O(h^{-1})$ operations per path generation, the total computational complexity for making the desired approximation is

$$O(nh^{-1}) = O(\delta_{N,h}\varepsilon_N^{-3} + \varepsilon_N^{-1}).$$

Since by standard methods we can conclude $\mathsf{Var}(D^N_{h,i}(t)) = O(N^{-1})$, we can again give an overall complexity of $O(N^{3\alpha-1} + N^{\alpha})$, as reported in the fourth row of Table 3.

4.4.2 Multilevel Monte Carlo and complexity analysis

In this section we study multilevel Monte Carlo approaches and derive the results summarized in rows five, six, and seven of Table 3.

Multilevel Monte Carlo and Diffusion Approximation

Here we specify and analyze an Euler-based multilevel method for the diffusion approximation, following the original framework of Giles [25].

For some fixed $M > 1$ we let $h_\ell = T \cdot M^{-\ell}$ for $\ell \in \{0, \dots, L\}$. Reasonable choices for M include $M \in \{2, 3, 4, \dots, 7\}$, and L is determined below. Let $D^N_{h_\ell}$ denote the approximate process generated by Euler-Maruyama applied to (4.8) with a step size of h_ℓ. As mentioned in section 4.4.1, the discretization has weak order one for a large class of functionals f, so we set $h_L = \varepsilon_N$, giving $L = O(|\log(\varepsilon_N)|)$, so that the finest level achieves the required order of magnitude for the bias.

Noting that

$$E[f(D^N_{h_L}(t))] = E[f(D^N_{h_0}(t))] + \sum_{\ell=1}^{L} E[f(D^N_{h_\ell}(t)) - f(D^N_{h_{\ell-1}}(t))], \tag{4.11}$$

we use i as an index over sample paths and let

$$\widehat{Q}^N_0 \overset{\text{def}}{=} \frac{1}{n_0} \sum_{i=1}^{n_0} f(D^N_{h_0,[i]}(t)), \quad \text{and} \quad \widehat{Q}^N_\ell \overset{\text{def}}{=} \frac{1}{n_\ell} \sum_{i=1}^{n_\ell} (f(D^N_{h_\ell,[i]}(t)) - f(D^N_{h_{\ell-1},[i]}(t))),$$

for $\ell = 1, \dots, L$, where n_0 and the different n_ℓ have yet to be determined. Note that the form of the estimator $\widehat{Q}^N_\ell$ above implies that the processes $D^N_{h_\ell}$ and $D^N_{h_{\ell-1}}$ will be

coupled, or constructed on the same probability space. We consider here the case when $(D^N_{h_\ell}, D^N_{h_{\ell-1}})$ are coupled in the usual way by using the same Brownian path in the generation of each of the marginal processes. Our (biased) estimator is then

$$\widehat{Q}^N \stackrel{\text{def}}{=} \widehat{Q}^N_0 + \sum_{\ell=1}^{L} \widehat{Q}^N_\ell .$$

Set

$$\delta_{N,\ell} = \mathsf{Var}(f(D^N_{h_\ell}(t)) - f(D^N_{h_{\ell-1}}(t))).$$

It is shown in [5] that $\delta_{N,\ell} = O(N^{-1}h_\ell^2 + N^{-2}h_\ell)$ under a wide array of conditions. Finally note that we also have $\mathsf{Var}(f(D^N_0(t))) = O(N^{-1})$. In [5] it is shown that under these circumstances, the computational complexity required is $O(\varepsilon_N^{-2}N^{-1} + \varepsilon_N^{-1})$. In the regime (4.4) this translates to $O(N^{2\alpha-1} + N^\alpha)$, as reported in the fifth row of Table 3.

Multilevel Monte Carlo and tau-leaping

The use of multilevel Monte Carlo with tau-leaping for continuous-time Markov chains of the form considered here was proposed in [3], where effective algorithms were devised. Complexity results were given in a non-asymptotic multi-scale setting, with followup results in [4]. Our aim here is to customize the approach in the scaling regime (4.4) and thereby develop easily interpretable complexity bounds that allow straightforward comparison with other methods. In this section $Z^N_{h_\ell}$ denotes a tau-leaping process generated with a step-size of $h_\ell = T \cdot M^\ell$, for $\ell \in \{0, \dots, L\}$.

A major step in [3] was to show that a coupling technique used for analytical purposes in [2, 47] can also form the basis of a practical simulation algorithm. Letting $Y_{k,i},\ i \in \{1, 2, 3\}$, denote independent, unit rate Poisson processes, we couple the exact

and approximate tau-leaping processes in the following way,

$$\begin{aligned} X^N(t) =& X^N(0) + \sum_k \frac{1}{N} Y_{k,1}\left(N\int_0^t \lambda_k(X^N(s)) \wedge \lambda_k(Z^N_{h_L}(\eta_L(s)))ds\right)\zeta_k \\ &+ \sum_k \frac{1}{N} Y_{k,2}\left(N\int_0^t [\lambda_k(X^N(s)) - \lambda_k(X^N(s)) \wedge \lambda_k(Z^N_{h_L}(\eta_L(s)))]ds\right)\zeta_k, \end{aligned} \tag{4.12}$$

$$\begin{aligned} Z^N_{h_L}(t) =& Z^N_{h_L}(0) + \sum_k \frac{1}{N} Y_{k,1}\left(N\int_0^t \lambda_k(X^N(s)) \wedge \lambda_k(Z^N_{h_L}(\eta_L(s)))ds\right)\zeta_k \\ &+ \sum_k \frac{1}{N} Y_{k,3}\left(N\int_0^t [\lambda_k(Z^N_{h_L}(\eta_L(s))) - \lambda_k(X^N(s)) \wedge \lambda_k(Z^N_{h_L}(\eta_L(s)))]ds\right)\zeta_k, \end{aligned} \tag{4.13}$$

where $a \wedge b$ denotes $\min\{a, b\}$ and $\eta_L(s) = \lfloor s/h_L \rfloor h_L$. Sample paths of (4.12)–(4.13) can be generated with a natural extension of the next reaction method or Gillespie's algorithm, see [3], and for $h_L \geq N^{-1}$ the complexity required for the generation of a realization $(X^N, Z^N_{h_L})$ remains at the $O(N)$ level. The coupling of two approximate processes, $Z^N_{h_\ell}$ and $Z^N_{h_{\ell-1}}$, takes the form

$$\begin{aligned} Z^N_{h_\ell}(t) =& Z^N_{h_\ell}(0) + \sum_k \frac{1}{N} Y_{k,1}\left(N\int_0^t \lambda_k(Z^N_{h_\ell}(\eta_\ell(s))) \wedge \lambda_k(Z^N_{h_{\ell-1}}(\eta_{\ell-1}(s)))ds\right)\zeta_k \\ &+ \sum_k \frac{1}{N} Y_{k,2}\left(N\int_0^t [\lambda_k(Z^N_{h_\ell}(\eta_\ell(s))) - \lambda_k(Z^N_{h_\ell}(\eta_\ell(s))) \wedge \lambda_k(Z^N_{h_{\ell-1}}(\eta_{\ell-1}(s)))]ds\right)\zeta_k, \end{aligned} \tag{4.14}$$

$$\begin{aligned} Z^N_{h_{\ell-1}}(t) =& Z^N_{h_{\ell-1}}(0) + \sum_k \frac{1}{N} Y_{k,1}\left(N\int_0^t \lambda_k(Z^N_{h_\ell}(\eta_\ell(s))) \wedge \lambda_k(Z^N_{h_{\ell-1}}(\eta_{\ell-1}(s)))ds\right)\zeta_k \\ &+ \sum_k \frac{1}{N} Y_{k,3}\left(N\int_0^t [\lambda_k(Z^N_{h_{\ell-1}}(\eta_{\ell-1}(s))) - \lambda_k(Z^N_{h_\ell}(\eta_\ell(s))) \wedge \lambda_k(Z^N_{h_{\ell-1}}(\eta_{\ell-1}(s)))]ds\right)\zeta_k, \end{aligned} \tag{4.15}$$

where $\eta_\ell(s) \stackrel{\text{def}}{=} \lfloor s/h_\ell \rfloor h_\ell$. The pair (4.14)–(4.15) can be sampled at the same $O(h_\ell^{-1})$ cost as a single tau-leaping path, see [3].

For L as yet to be determined, and noting the identity

$$\mathbb{E}[f(X^N(t))] = \mathbb{E}[f(X^N(t)) - f(Z_L^N(t))] + \sum_{\ell=1}^{L} \mathbb{E}[f(Z_{h_\ell}^N(t)) - f(Z_{h_{\ell-1}}^N(t))] + \mathbb{E}[f(Z_{h_0}^N(t))], \tag{4.16}$$

we define estimators for the three terms above via

$$\begin{aligned}
\widehat{Q}_E^N &\stackrel{\text{def}}{=} \frac{1}{n_E} \sum_{i=1}^{n_E} (f(X_{[i]}^N(t)) - f(Z_{h_L,[i]}^N(t))), \\
\widehat{Q}_\ell^N &\stackrel{\text{def}}{=} \frac{1}{n_\ell} \sum_{i=1}^{n_\ell} (f(Z_{h_\ell,[i]}^N(t)) - f(Z_{h_\ell-1,[i]}^N(t))), \quad \text{for } \ell \in \{1, \ldots, L\}, \\
\widehat{Q}_0^N &\stackrel{\text{def}}{=} \frac{1}{n_0} \sum_{i=1}^{n_0} f(Z_{h_0,[i]}^N(t)),
\end{aligned} \tag{4.17}$$

so that

$$\widehat{Q}^N \stackrel{\text{def}}{=} \widehat{Q}_E^N + \sum_{\ell=1}^{L} \widehat{Q}_\ell^N + \widehat{Q}_0^N \tag{4.18}$$

is an unbiased estimator for $\mathbb{E}[f(X^N(T))]$. Here, $\widehat{Q}_E^N$ uses the coupling (4.12)–(4.13) between exact paths and tau-leaped paths of stepsize h_L, $\widehat{Q}_\ell^N$ uses the coupling (4.14)–(4.15) between tau-leaped paths of stepsizes h_ℓ and $h_{\ell-1}$, and $\widehat{Q}_0^N$ involves single tau-leaped paths of stepize h_0. Note that the algorithm implicit in (4.18) produces an unbiased estimator, whereas the estimator is biased if $\widehat{Q}_E^N$ is left off, as will sometimes be desirable. Hence, we will refer to estimator $\widehat{Q}^N$ in (4.18) as the unbiased estimator, and will refer to

$$\widehat{Q}_B^N \stackrel{\text{def}}{=} \sum_{\ell=1}^{L} \widehat{Q}_\ell^N + \widehat{Q}_0^N \tag{4.19}$$

as the biased estimator. For both the biased and unbiased estimators, the number of paths at each level, n_0, n_ℓ and n_E, will be chosen to ensure an overall estimator variance of $O(\varepsilon_N^2)$.

Continuing the analysis, we assume that for $\ell \geq 0$

$$\mathsf{Var}(f(X^N(t)) - f(Z^N_{h_\ell}(t))) = O(\delta_{N,E}) \tag{4.20}$$

$$\mathsf{Var}(f(Z^N_{h_\ell}(t)) - f(Z^N_{h_{\ell-1}}(t))) = O(\delta_{N,\ell}) \tag{4.21}$$

$$\mathsf{Var}(f(Z^N_{h_0}(t))) = O(N^{-1}). \tag{4.22}$$

It is shown in [4] that $\delta_{N,\ell} = N^{-1}h_\ell$ for $\ell \geq 0$ under a wide array of circumstances, so for the remainder of this section we assume this scaling.

We consider the biased and unbiased versions of tau-leaping multilevel Monte Carlo separately.

Biased multilevel Monte Carlo tau-leaping

Here we consider the estimator $\widehat{Q}^N_B$ defined in (4.19). We first note that $E[f(X^N(t))] - E[f(Z^N_{h_L}(t))] = O(h_L)$ for a large class of functionals f, see [2]. Hence, we begin by choosing $L = O(\log(1/\varepsilon_N)) = O(\log(N))$ in order to control the bias.

For $\ell \in \{1, \ldots, L\}$, let C_ℓ be the number of random variables 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 find n_ℓ, $\ell \in \{0, \ldots, L\}$, we solve the following optimization problem, which ensures that the variance of $\widehat{Q}^N_B$ is no greater than order ε^2_N:

$$\underset{n_\ell}{\text{minimize}} \quad \sum_{\ell=0}^{L} n_\ell C_\ell, \tag{4.23}$$

$$\text{subject to} \quad \sum_{\ell=0}^{L} \frac{\delta_{N,\ell}}{n_\ell} = \varepsilon^2_N. \tag{4.24}$$

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

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

Taking derivatives with respect to n_ℓ and setting each derivative to zero yields,

$$n_\ell = \sqrt{\tfrac{\lambda}{K} \delta_{N,\ell} h_\ell}, \qquad \text{for } \ell \in \{0, 1, 2, \ldots, L\} \tag{4.25}$$

for some $\lambda \geq 0$. Plugging (4.25) into (4.24) gives us,

$$\sum_{\ell=0}^{L} \sqrt{\frac{\delta_{N,\ell}}{h_\ell}} = \sqrt{\tfrac{\lambda}{K}} \cdot \varepsilon_N^2 \tag{4.26}$$

and hence, by (4.21) and (4.22),

$$\sqrt{\tfrac{\lambda}{K}} = \sum_{\ell=0}^{L} \sqrt{\tfrac{\delta_{N,\ell}}{h_\ell}} \leq C L \varepsilon_N^{-2} N^{-1/2}, \tag{4.27}$$

where C is a constant. Noting that $L = O(\log(\varepsilon_N^{-1}))$, we have

$$\tfrac{\lambda}{K} = O\left(\varepsilon_N^{-4} \log\left(\varepsilon_N\right)^2 N^{-1} \right).$$

Plugging this back into (4.25), and recognizing that at least one path must be generated to achieve the desired accuracy, we find

$$n_\ell = O(\varepsilon_N^{-2} N^{-1} h_\ell L + 1).$$

Hence, the overall computational complexity is

$$\begin{aligned} \sum_{\ell=0}^{L} n_\ell K h_\ell^{-1} &= O\left(\sum_{\ell=0}^{L} \varepsilon_N^{-2} N^{-1} h_\ell L h_\ell^{-1} + \sum_{\ell=0}^{L} h_\ell^{-1} \right) \\ &= O\left(\varepsilon_N^{-2} N^{-1} \log(\varepsilon_N)^2 + \varepsilon_N^{-1} \right) \\ &= O\left(N^{2\alpha - 1} \log(N)^2 + N^\alpha \right), \end{aligned}$$

recovering row six of Table 3.

Note that the computational complexity reported for this biased version of multilevel Monte Carlo tau-leaping is, up to logarithms, the same as that for multilevel Monte Carlo on the diffusion approximation. However, none of the generous assumptions we made for the diffusion approximation, including that $E[f(X^N(t))] = E[f(D^N(t))]$, were required.

Unbiased multilevel Monte Carlo tau-leaping

The first observation to make is that the telescoping sum (4.16) implies that the method which utilizes $E[f(X^N(t)) - f(Z^N_{h_L}(t))]$ at the finest level is unbiased for *any* choice of h_L. That is, we are no longer constrained to choose $L = O(|\log(\varepsilon_N)|)$.

Assume (4.20) holds with $\delta_{N,E} = N^{-1}h_L$, and that $h_L \geq N^{-1}$. Let C_E be the average number of random variables required to generate a single pair of the coupled exact and tau-leaped processes when the tau-leap discretization is h_L.We know $C_E = O(N+h_L^{-1}) = O(N)$. To determine n_ℓ and n_E, we still solve an optimization problem,

$$\underset{n_\ell}{\text{minimize}} \quad \sum_{\ell=0}^{L} n_\ell C_\ell + n_L C_E, \tag{4.28}$$

$$\text{subject to} \quad \sum_{\ell=0}^{L} \frac{\delta_{N,\ell}}{n_\ell} + \frac{\delta_{N,E}}{n_E} = \varepsilon_N^2. \tag{4.29}$$

Using Lagrange multipliers again, we obtain,

$$n_\ell = \sqrt{\frac{\lambda \delta_{N,\ell}}{C_\ell}} \qquad \text{for } \ell \in \{0, 1, 2, \ldots, L\} \tag{4.30}$$

and

$$n_E = \sqrt{\frac{\lambda \delta_{N,E}}{C_E}}. \tag{4.31}$$

Plugging back into (4.29) and noting $C_\ell = O(h_\ell^{-1})$ and $C_E = O(N)$ yields

$$\sqrt{\lambda} = \varepsilon_N^{-2}\left(\sum_{\ell=0}^{L}\sqrt{\delta_{N,\ell}C_\ell} + \sqrt{\delta_{N,E}C_E}\right) \leq C(L\varepsilon_N^{-2}N^{-1/2} + \varepsilon_N^{-2}\sqrt{h_L}). \tag{4.32}$$

Therefore, plugging (4.32) back into (4.30) an (4.31) and noting $n_\ell \geq 1$ and $n_E \geq 1$, we get

$$n_\ell = \sqrt{\frac{\lambda\delta_{N,\ell}}{C_\ell}} + 1 = O\left(\left(L\varepsilon_N^{-2}N^{-1} + \varepsilon_N^{-2}\sqrt{\frac{h_L}{N}}\right)h_\ell + 1\right) \qquad \text{for } \ell \in \{0,1,2,\ldots,L\}$$

and

$$n_E = \sqrt{\frac{\lambda\delta_{N,\ell}}{C_\ell}} + 1 = O(L\varepsilon_N^{-2}N^{-3/2}h_L^{1/2} + \varepsilon_N^{-2}N^{-1}h_L + 1). \tag{4.33}$$

As a result the total complexity is

$$\begin{aligned} g(h_L) &= O(\varepsilon_N^{-2}N^{-1}L^2 + \varepsilon_N^{-2}\sqrt{\frac{h_L}{N}}L + h_L^{-1} + \varepsilon_N^{-2}\sqrt{\frac{h_L}{N}}L + \varepsilon_N^{-2}h_L + N) \\ &\leq O(\varepsilon_N^{-2}N^{-1}L^2 + 2\varepsilon_N^{-2}\sqrt{\frac{h_L}{N}}L + \varepsilon_N^{-2}h_L + 2N) && \text{(since } h_L^{-1} \leq N\text{)} \\ &= O(2\varepsilon_N^{-2}N^{-1}L^2 + 2\varepsilon_N^{-2}h_L + 2N). && \text{(using that } 2ab \leq a^2 + b^2\text{)} \end{aligned}$$

It is relatively easy to show that the last line above is minimized at

$$h_L = \frac{2}{N}\text{LambertW}\left(\frac{N}{2}\right) \approx \frac{2}{N}\log\left(\frac{N}{2}\right). \tag{4.34}$$

Hence, taking $h_L = O(N^{-1}\log(N))$, we have $\log(h_L)^2 = O(\log(N)^2)$ and this method achieves a total computational complexity of leading order

$$O(\varepsilon_N^{-2}N^{-1}\log(N)^2 + \varepsilon_N^{-2}N^{-1}\log(N) + N) = O(\varepsilon_N^{-2}N^{-1}\log(N)^2 + N) = O(N^{2\alpha-1}\log(N)^2 + N),$$

as reported in the last row of Table 3.

Note here that if we choose $h_L = \frac{1}{N}$ we get the same order of magnitude for the computational complexity. However the h_L in (4.34) is the optimized solution, meaning the leading order constant should be better and we will see this in Figure 11 and Figure 12 in the next section.

4.5 Computational results

In this section we provide numerical evidence for the sharpness of the computational complexity analyses provided in Table 3. We will measure complexity by total number of random variables utilized. We emphasize that these experiments use extreme parameter choices solely for the purpose of testing the sharpness of the delicate asymptotic bounds.

Example 4.1. *We consider the classically scaled stochastic model for the following reaction network (see [10])*

$$S_1 + S_2 \underset{k_2}{\overset{k_1/N}{\rightleftarrows}} S_3 \overset{k_3}{\to} S_2 + S_4.$$

Letting $X_i(t)$ give the number of molecules of species S_i at time t, and letting $X^N(t) = X(t)/N$, the stochastic equations are

$$X^N(t) = X^N(0) + \frac{1}{N}Y_1\left(Nk_1\int_0^t X_1^N(s)X_2^N(s)ds\right)\begin{bmatrix}-1\\-1\\1\\0\end{bmatrix}$$
$$+\frac{1}{N}Y_2\left(Nk_2\int_0^t X_3^N(s)ds\right)\begin{bmatrix}1\\1\\-1\\0\end{bmatrix}$$
$$+\frac{1}{N}Y_3\left(Nk_3\int_0^t X_3^N(s)ds\right)\begin{bmatrix}0\\1\\-1\\1\end{bmatrix},$$

where we assume $X^N(0) = O(1)$. *We implemented different Monte Carlo simulation methods for the estimation of* $\mathbb{E}[X_1^N(T)]$ *to an accuracy of* $\varepsilon_N = N^{-\alpha}$ *for both* $\alpha = 1$ *and* $\alpha = 5/4$. *Specifically, for each of the order one methods we chose a step size of size* $h = \varepsilon_N$ *and required the variance of the estimator to be* ε_N^2. *For midpoint tau-leaping, which has a weak order of two, we chose* $h = \sqrt{\varepsilon_N}$. *For the unbiased multilevel Monte Carlo method we chose the finest time-step according to* (4.34). *We do not provide estimates for Monte Carlo combined with exact simulation as those computations were too intensive to complete to the target accuracy.*

For our numerical example we chose $T = 1$ *and* $X(0) = \lceil N \cdot [0.2, 0.2, 0.2, 0.2]^T \rceil$ *with* $X^N(0) = X(0)/N$. *Finally, we chose* $k_1 = k_2 = k_3 = 1$ *as our rate constants. In Figure 9, we provide log-log plots of the computational complexity required to solve this problem for the different Monte Carlo methods to an accuracy of* $\varepsilon_N = N^{-1}$, *for each of*

$$N \in \{2^{13}, 2^{14}, 2^{15}, 2^{16}, 2^{17}\}.$$

In Figure 10, we provide log-log plots for the computational complexity required to solve this problem for the different methods to an accuracy of $\varepsilon_N = N^{-\frac{5}{4}}$, *for each of*

$$N \in \{2^9, 2^{10}, 2^{11}, 2^{12}, 2^{13}\}.$$

Tables 4 and 5 provide the estimator variances for the different Monte Carlo methods with $\varepsilon_N = N^{-1}$ *and* $\varepsilon_N = N^{-\frac{5}{4}}$, *respectively. The top line provides the target variances.*

The specifics of the implementations and results for the different Monte Carlo methods are detailed below.

Diffusion Approximation plus Monte Carlo. *We took a time step of size* $h = \varepsilon_N$ *to generate our independent samples. See Figure 9, where the best fit line is* $y = 1.94x -$

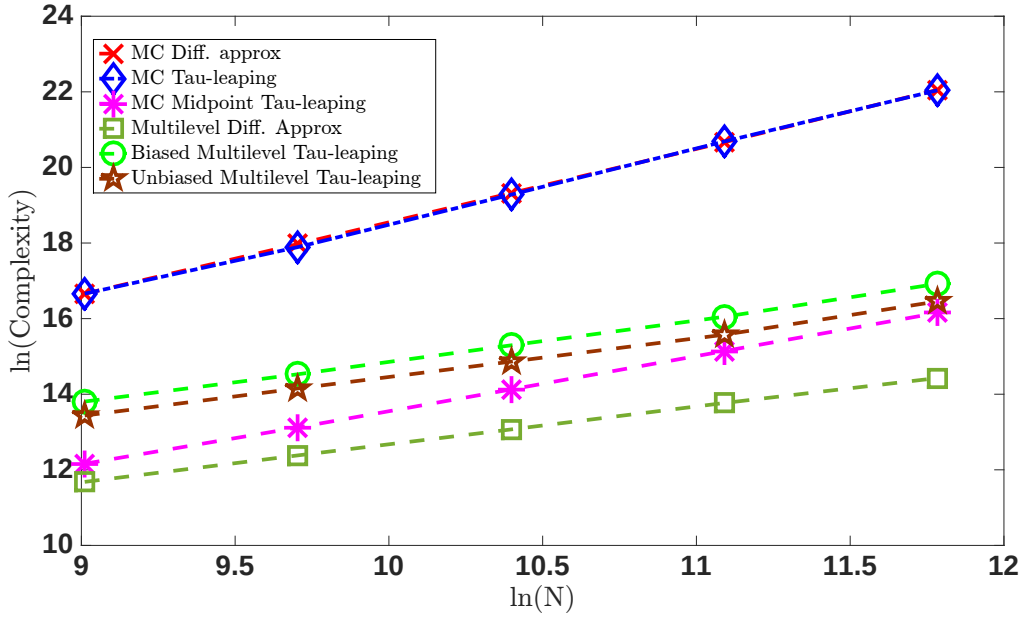


Figure 9: Log-log plots of the computational complexity for the different Monte Carlo methods with varying $N \in \{2^{13}, 2^{14}, 2^{15}, 2^{16}, 2^{17}\}$ *and* $\varepsilon_N = N^{-1}$.

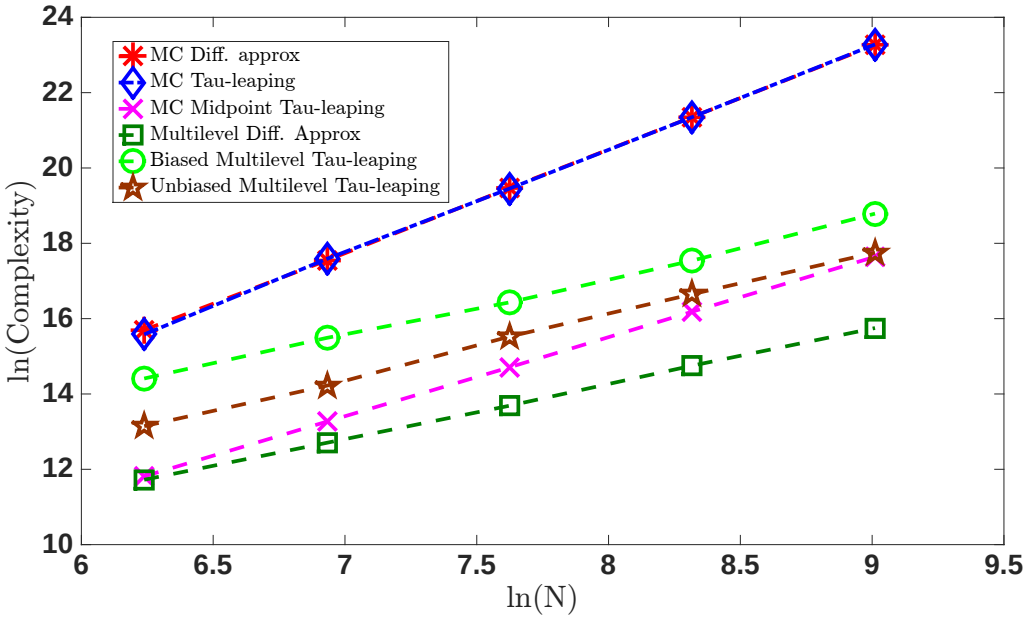


Figure 10: Log-log plots of the computational complexity for the different Monte Carlo methods with varying $N \in \{2^{9}, 2^{10}, 2^{11}, 2^{12}, 2^{13}\}$ *and* $\varepsilon_N = N^{-\frac{5}{4}}$.

Method	*estimator standard deviations*
$\varepsilon_N = N^{-1}$	$2^{-13}, 2^{-14}, 2^{-15}, 2^{-16}, 2^{-17}$
MC and Diff. approx	$2^{-13.10}, 2^{-14.02}, 2^{-15.02}, 2^{-16.01}, 2^{-17.00}$
MC and Tau-leaping	$2^{-13.09}, 2^{-14.01}, 2^{-15.01}, 2^{-16.01}, 2^{-17.00}$
MC and Midpoint Tau-leaping	$2^{-13.09}, 2^{-14.04}, 2^{-15.03}, 2^{-16.00}, 2^{-17.01}$
Multilevel Diff. approx	$2^{-13.20}, 2^{-14.15}, 2^{-15.11}, 2^{-16.09}, 2^{-17.07}$
Biased Multilevel Tau-leaping	$2^{-13.44}, 2^{-14.39}, 2^{-15.39}, 2^{-16.38}, 2^{-17.32}$
Unbiased Multilevel Tau-leaping	$2^{-13.29}, 2^{-14.28}, 2^{-15.26}, 2^{-16.21}, 2^{-17.18}$

Table 4: Actual estimator variances when $\varepsilon_N = N^{-1}$.

Method	*estimator standard deviations*
$\varepsilon_N = N^{-\frac{5}{4}}$	$2^{-11.25}, 2^{-12.50}, 2^{-13.75}, 2^{-15.00}, 2^{-16.25}$
MC and Diff. approx	$2^{-11.27}, 2^{-12.51}, 2^{-13.75}, 2^{-15.00}, 2^{-16.25}$
MC and Tau-leaping	$2^{-11.26}, 2^{-12.52}, 2^{-13.76}, 2^{-15.00}, 2^{-16.25}$
MC and Midpoint Tau-leaping	$2^{-11.26}, 2^{-12.52}, 2^{-13.76}, 2^{-15.00}, 2^{-16.25}$
Multilevel Diff. approx	$2^{-11.46}, 2^{-12.63}, 2^{-13.85}, 2^{-15.06}, 2^{-16.29}$
Biased Multilevel Tau-leaping	$2^{-11.62}, 2^{-12.81}, 2^{-13.99}, 2^{-15.19}, 2^{-16.41}$
Unbiased Multilevel Tau-leaping	$2^{-11.34}, 2^{-12.57}, 2^{-13.79}, 2^{-15.03}, 2^{-16.26}$

Table 5: Actual estimator variances when $\varepsilon_N = N^{-5/4}$.

0.88, *and Figure 10, where the best fit line is* $y = 2.73x - 1.37$, *which are consistent with Table 3.*

Monte Carlo Tau-Leaping. *We took a time step of size* $h = \varepsilon_N$ *to generate our independent samples. See Figure 9, where the best fit line is* $y = 1.96x - 1.02$, *and Figure 10, where the best fit line is* $y = 2.76x - 1.63$, *which are consistent with Table 3.*

Monte Carlo Midpoint Tau-Leaping. *We took a time step of size* $h = \sqrt{\varepsilon_N}$. *See Figure 9, where the best fit line is* $y = 1.44 - 0.86$, *and Figure 10, where the best fit line is* $y = 2.10x - 3.53$, *which are consistent with Table 3.*

Our implementation of the multilevel methods proceeded as follows. We chose $h_\ell = 2^{-\ell}$ *and for* $\varepsilon_N > 0$ *we fixed* $h_L = \varepsilon_N$ *and* $L = \lceil \log(h_L)/\log(2) \rceil$ *for the biased methods. For each level we generated* N_0 *independent sample trajectories in order to estimate* $\delta_{N,\ell}$, *as defined in section 4.3. Then we selected*

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

to ensure the overall variance is below the target ε_N^2.

Multi-Level Monte Carlo Diffusion Approximation *We used* $N_0 = 400$ *for our pre-calculation of the variances. See Figure 9, where the best fit line is* $y = 0.99x + 2.75$, *and Figure 10, where the best fit line is* $y = 1.45x + 2.61$, *which are consistent with Table 3.*

Multi-Level Monte Carlo Tau-Leaping. *We used* $N_0 = 100$ *for our pre-calculation of the variances. See Figure 9, where the best fit line is* $y = 1.12x + 3.70$, *and Figure 10, where the best fit line is* $y = 1.56x + 4.64$, *which are consistent with Table 3.*

Unbiased Tau-leaping multilevel Monte Carlo. *For our implementation of unbiased multilevlel tau-leaping, we set* $h_L = \frac{2}{N} LambertW\left(\frac{N}{2}\right)$ *and* $L = \lceil \log(h_L)/\log(2) \rceil$. *For each level we utilized* $N_0 = 100$ *independent sample trajectories in order to estimate* $\delta_{N,\ell}, C_\ell$ *and* $\delta_{N,E}, C_E$, *as defined in section 4.3. Then we then selected*

$$n_\ell = \left\lceil \varepsilon_N^{-2} \sqrt{\frac{\delta_{N,\ell}}{C_\ell}} \left(\sum_{\ell=0}^{L} \sqrt{\delta_{N,\ell} C_\ell} + \sqrt{\delta_{N,E} C_E} \right) \right\rceil + 1, \qquad \text{for } \ell \in \{0, 1, 2, \ldots, L\},$$

and

$$n_E = \left\lceil \varepsilon_N^{-2} \sqrt{\frac{\delta_{N,E}}{C_E}} \left(\sum_{\ell=0}^{L} \sqrt{\delta_{N,\ell} C_\ell} + \sqrt{\delta_{N,E} C_E} \right) \right\rceil + 1,$$

to ensure the overall estimator variance is below our target ε_N^2. *See Figure 9, where the best fit line is* $y = 1.08x + 3.71$, *and Figure 10, where the best fit line is* $y = 1.68x + 2.65$, *which are consistent with Table 3.*

We also used the unbiased tau-leaping multilevel Monte Carlo method with $h_L = N^{-1}$ *to estimate* $\mathbb{E}[X_1(1)]$ *to accuracy* $\varepsilon_N = N^{-\alpha}$, *for both* $\alpha = 1$ *and* $\alpha = 5/4$. *See Figures 11 and 12 for log-log plots of the required complexity when* $h_L = N^{-1}$ *and* $h_L = (\frac{2}{N}) LambertW(\frac{N}{2})$. *As predicted in section 4.4.2, the complexity required when* $h_L = (\frac{2}{N}) LambertW(\frac{N}{2})$ *is lower by some constant factor.*

4.6 Conclusions

Many researchers have observed in practice that approximation methods can lead to computational efficiency, relative to exact path simulation. However, *meaningful, rigorous justification* for *whether* and *under what circumstances* approximation methods offer computational benefit has proved elusive. Focusing on the classical scaling, we note that a useful analysis must resolve two issues:

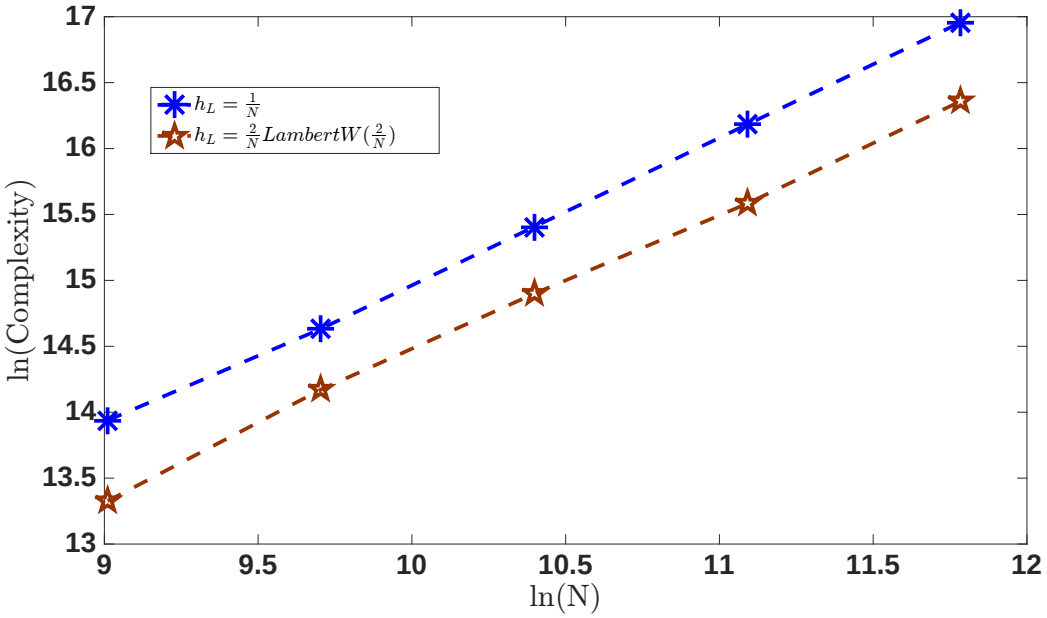


Figure 11: Complexity comparison of unbiased multilevel Monte Carlo tau-leaping when $h_L = \frac{1}{N}$ *and* $h_L = \frac{2}{N} LambertW(\frac{2}{N})$*, with* $\varepsilon_N = N^{-1}$.

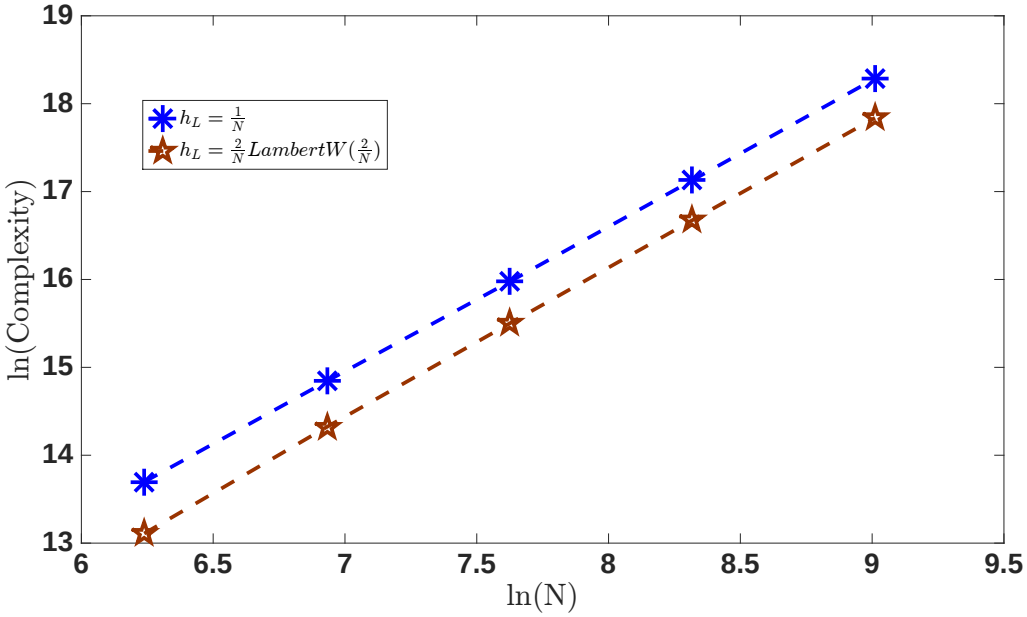


Figure 12: Complexity comparison of unbiased multilevel Monte Carlo tau-leaping when $h_L = \frac{1}{N}$ *and* $h_L = \frac{2}{N} LambertW(\frac{2}{N})$*, with* $\varepsilon_N = N^{-\frac{5}{4}}$.

(1) Computational complexity is most relevant for "large" problems, where many events take place. However, as the system size grows the problem converges to a simpler, deterministic limit that is cheap to solve.

(2) On a fixed problem, in the traditional numerical analysis setting where mesh size tends to zero, discretization methods become arbitrarily more expensive than exact simulation because the exact solution is piecewise constant.

In this work, we offer what we believe to be the first rigorous complexity analysis that allows for systematic comparison of simulation methods. The results, summarized in Table 3, apply under the classical scaling for a family of problems parametrized by the system size, N, with accuracy requirement $N^{-\alpha}$. In this regime, we can study performance on "large" problems when fluctuations are still relevant.

A simple conclusion from our analysis is that standard tau-leaping does offer a concrete advantage over exact simulation when the accuracy requirement is not too high, $\alpha < 1$; see the first two rows of Table 3. Also, "second order" midpoint or trapezoidal tau-leaping improves on exact simulation for $\alpha < 2$; row three of Table 3. Furthermore, in this framework, we were able to analyze the use of a diffusion, or Langevin, approximation and the multilevel Monte Carlo versions of tau-leaping and diffusion simulation. Our overall conclusion is that some form of tau-leaping is always worthwhile. For low accuracy ($\alpha < 2/3$), second order tau-leaping with standard Monte Carlo is the most efficient of the methods considered. At higher accuracy requirements, $\alpha > 2/3$, multilevel Monte Carlo tau-leaping is joint-best. Moreover, for all $\alpha > 1$ an unbiased version of multilevel Monte Carlo tau-leaping shares the lowest complexity level, making it our method of choice for high accuracy.

Possibilities for further research along the lines opened up by this work include:

- analyzing other methods within this framework, for example, (a) multilevel Monte Carlo for the diffusion approximation using discretization methods customized for small noise systems, or (b) methods that tackle the Chemical Master Equation directly using large scale deterministic ODE technology [41, 42],

- coupling the required accuracy to the system size in other scaling regimes, for example, to study specific problem classes with multiscale structure [11].

Appendix A

Appendix

We provide here some technical lemmas which were used in Section 2.2 and Section 3.3.

Lemma A.1. *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*

$$\sup_{s\leq T}\mathbb{E}\left[|X_1(s)-x_1(s)|^2\right]\leq \widehat{C}_1(N^\gamma T)N^{-\rho},\quad \sup_{s\leq T}\mathbb{E}\left[|X_2(s)-x_2(s)|^2\right]\leq \widehat{C}_2(N^\gamma T)N^{-\rho}, \tag{A.1}$$

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

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

Proof. First we know,

$$\begin{aligned}
&\mathsf{Var}\left(\int_0^1 u(X_2(s)+r(X_1(s)-X_2(s)))dr\right)\\
&\quad=\mathsf{Var}\left(\int_0^1 u(X_2(s)+r(X_1(s)-X_2(s)))-u(x_2(s)+r(x_1(s)-x_2(s)))dr\right)\\
&\quad\leq\mathbb{E}\left(\int_0^1 u(X_2(s)+r(X_1(s)-X_2(s)))-u(x_2(s)+r(x_1(s)-x_2(s)))dr\right)^2\\
&\quad\leq\int_0^1\mathbb{E}[u(X_2(s)+r(X_1(s)-X_2(s)))-u(x_2(s)+r(x_1(s)-x_2(s)))]^2dr.
\end{aligned}$$

Using that u is Lipschitz, we may continue

$$\begin{aligned}
&\mathsf{Var}\left(\int_0^1 u(X_2(s)+r(X_1(s)-X_2(s)))dr\right)\\
&\quad\le L^2\int_0^1 \mathbb{E}\left[|X_2(s)+r(X_1(s)-X_2(s))-(x_2(s)+r(x_1(s)-x_2(s)))|^2\right]dr\\
&\quad= L^2\int_0^1 \mathbb{E}\left[|r(X_1(s)-x_1(s))+(1-r)(X_2(s)-x_2(s))|^2\right]dr\\
&\quad\le L^2\int_0^1 r\mathbb{E}\left[|X_1(s)-x_1(s)|^2\right]+(1-r)\mathbb{E}\left[|X_2(s)-x_2(s)|^2\right]dr\\
&\quad\le L^2\max(\widehat{C}_1,\widehat{C}_2)N^{-\rho},
\end{aligned}$$

where the second to last inequality follows from convexity of the quadratic function, and the final inequality holds from applying (A.1). □

Lemma A.2. *Suppose that $A^{N,h}$ and $B^{N,h}$ are families of random variables determined by scaling parameters N and h. Further, suppose that there are $C_1>0, C_2>0$ and $C_3>0$ such that for all $N>0$ the following three conditions hold:*

1. $\mathsf{Var}(A^{N,h})\le C_1N^{-\rho}$ *uniformly in* h.
2. $|A^{N,h}|\le C_2N^{\gamma}$ *uniformly in* h.
3. $|\mathbb{E}[B^{N,h}]|\le C_3N^{\gamma}h$.

Then

$$\mathsf{Var}(A^{N,h}B^{N,h})\le 3C_3^2C_1N^{-\rho}(N^{\gamma}h)^2+15C_2^2N^{2\gamma}\mathsf{Var}(B^{N,h}).$$

Proof. Via a direct expansion, the variance of the product can be represented in the following manner

$$\begin{aligned}
\mathsf{Var}(A^{N,h}B^{N,h}) = \mathbb{E}[(\mathbb{E}[B^{N,h}])(A^{N,h}-\mathbb{E}[A^{N,h}])+(\mathbb{E}[A^{N,h}])(B^{N}-\mathbb{E}[B^{N,h}])\\
+(A^{N,h}-\mathbb{E}[A^{N,h}])(B^{N,h}-\mathbb{E}[B^{N,h}])-\mathbb{E}[(A^{N,h}-\mathbb{E}[A^{N,h}])(B^{N,h}-\mathbb{E}[B^{N,h}])]]^2.
\end{aligned}$$

Using the basic bound $(a+b+c)^2 \leq 3a^2+3b^2+3c^2$, we have

$$\begin{aligned}\mathsf{Var}(A^{N,h}B^{N,h}) \leq & 3(\mathbb{E}[B^{N,h}])^2\mathsf{Var}(A^{N,h}) + 3(\mathbb{E}[A^{N,h}])^2\mathsf{Var}(B^{N,h}) \\ & + 3\mathsf{Var}((A^{N,h}-\mathbb{E}[A^{N,h}])(B^{N,h}-\mathbb{E}[B^{N,h}])).\end{aligned}$$

Using our assumptions in the statement of the lemma, the following two inequalities are immediate

$$3(\mathbb{E}[B^{N,h}])^2\mathsf{Var}(A^{N,h}) \leq 3C_3^2C_1N^{-\rho}(N^\gamma h)^2, \tag{A.2}$$

and

$$3(\mathbb{E}[A^{N,h}])^2\mathsf{Var}(B^{N,h}) \leq 3C_2^2N^{2\gamma}\mathsf{Var}(B^{N,h}). \tag{A.3}$$

For the final term we bound the variance by the second moment to achieve

$$\begin{aligned}3\mathsf{Var}((A^{N,h}-\mathbb{E}[A^{N,h}])(B^{N,h}-\mathbb{E}[B^{N,h}])) &\leq 3\mathbb{E}((A^{N,h}-\mathbb{E}[A^{N,h}])(B^{N,h}-\mathbb{E}[B^{N,h}]))^2 \\ &\leq 12C_2^2N^{2\gamma}\mathsf{Var}(B^{N,h}).\end{aligned} \tag{A.4}$$

Combining (A.2), (A.3), and (A.4) gives the desired result. □

The two following lemmas are only slight perturbations of the two lemmas above. proof is therefore omitted.

Lemma A.3. *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 $u:\mathbb{R}^d\to\mathbb{R}$ is Lipschitz with Lipschitz constant C_L. Then,

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

Lemma A.4. *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 $\varepsilon \in (0,1)$ 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

$$\mathsf{Var}(A^{\varepsilon,h}B^{\varepsilon,h}) \leq 3C_3^2C_1h^2\varepsilon^2 + 15C_2^2\mathsf{Var}(B^{\varepsilon,h}).$$

Lemma A.5. *Let $Q(s)$ be a stochastic process for which $\sup_{s\in[a,b]} \mathsf{Var}(Q(s)) < \infty$. Then*

$$\mathsf{Var}\left(\int_a^b Q(s)ds\right) \leq (b-a)\int_a^b \mathsf{Var}(Q(s))ds.$$

Proof. The proof is straightforward.

$$\begin{aligned}\mathsf{Var}\left(\int_a^b Q(s)ds\right) &= \mathbb{E}\left(\int_a^b Q(s)ds - \mathbb{E}\left[\int_a^b Q(s)ds\right]\right)^2 = \mathbb{E}\left(\int_a^b (Q(s) - \mathbb{E}[Q(s)])ds\right)^2 \\ &\leq (b-a)\int_a^b \mathbb{E}\left[(Q(s) - \mathbb{E}Q(s))^2\right]ds = (b-a)\int_a^b \mathsf{Var}(Q(s))ds. \qquad \square\end{aligned}$$

Lemma A.6. *Let $f : \mathbb{R}^d \to \mathbb{R}$ have continuous first derivative. Then, for any $x, y \in \mathbb{R}^d$,*

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

Proof. Let $H(t) = f(tx + (1-t)y)$. Then $H'(t) = \nabla f(tx + (1-t)y) \cdot (x-y)$, and by the fundamental theorem of calculus, $H(1) = H(0) + \int_0^1 H'(s)ds$, which is equivalent to the statement of the lemma. $\square$

Bibliography

[1] D. F. Anderson, *A modified next reaction method for simulating chemical systems with time dependent propensities and delays*, J. Chem. Phys., 127 (2007), p. 214107.

[2] D. F. Anderson, A. Ganguly, and T. G. Kurtz, *Error analysis of tau-leap simulation methods*, Annals of Applied Probability, 21 (2011), pp. 2226–2262.

[3] D. F. Anderson and D. J. Higham, *Multilevel Monte-Carlo for continuous time Markov chains, with applications in biochemical kinetics*, SIAM: Multiscale Modeling and Simulation, 10 (2012), pp. 146 – 179.

[4] 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.

[5] D. F. Anderson, D. J. Higham, and Y. Sun, *Multilevel monte carlo for stochastic differential equations with small noise*, arXiv preprint arXiv:1412.3039, (2014).

[6] D. F. Anderson, D. J. Higham, and Y. Sun, *Computational complexity analysis for monte carlo approximations of classically scaled population processes*, submitted, (2015).

[7] ——, *Multilevel monte carlo for stochastic differential equations with small noise*, Accepted, (2015).

[8] D. F. Anderson and M. Koyama, *Weak error analysis of numerical methods*

for stochastic models of population processes, SIAM: Multiscale Modeling and Simulation, 10 (2012), pp. 1493–1524.

[9] 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.

[10] ——, *Stochastic analysis of biochemical systems*, Springer International Publishing, 2015.

[11] K. Ball, T. G. Kurtz, L. Popovic, and G. Rempala, *Asymptotic analysis of multiscale approximations to reaction networks*, Annals of Applied Probability, 16 (2006), pp. 1925–1961.

[12] X. Bardina, D. Bascompte, C. Rovira, and S. Tindel, *An analysis of a stochastic model for bacteriophage systems*, Mathematical Biosciences, 241 (2013), pp. 99–108.

[13] A. F. Bartholomay, *Stochastic models for chemical reactions. I. Theory of the unimolecular reaction process*, Bull. Math. Biophys., 20 (1958), pp. 175–190.

[14] ——, *Stochastic models for chemical reactions. II. The unimolecular rate constant*, Bull. Math. Biophys., 21 (1959), pp. 363–373.

[15] D. Belomestny and T. Nagapetyan, *Variance reduced multilevel path simulation: going beyond the complexity ε^{-2}*. arXiv:1412.4045, 2014.

[16] D. L. Burkholder, B. J. Davis, and R. F. Gundy, *Integral inequalities for convex functions of operators on martingales*, Proc. Sixth Berkeley Symp. on Math. Statist. and Prob., 2 (1972), pp. 223–240.

[17] G. Conforti, S. D. 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., vol. 110 of Proceedings in Mathematics and Statistics, Springer, 2015.

[18] H. De Jong, *Modeling and simulation of genetic regulatory systems: a literature review*, Journal of computational biology, 9 (2002), pp. 67–103.

[19] M. Delbrück, *Statistical fluctuations in autocatalytic reactions*, The Journal of Chemical Physics, 8 (1940), pp. 120–124.

[20] S. Delong, Y. Sun, B. E. Griffith, E. Vanden-Eijnden, and A. Donev, *Multiscale temporal integrators for fluctuating hydrodynamics*, Phys. Rev. E, 90 (2014), p. 063312.

[21] 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.

[22] C. W. Gardiner, *Handbook of Stochastic Methods: for Physics, Chemistry and the Natural Sciences*, Springer, Berlin, 2002.

[23] M. Gibson and J. Bruck, *Efficient exact stochastic simulation of chemical systems with many species and many channels*, J. Phys. Chem. A, 105 (2000), pp. 1876–1889.

[24] M. Giles, *Improved multilevel Monte Carlo convergence using the Milstein scheme*, in Monte Carlo and Quasi-Monte Carlo Methods 2006, A. Keller, S. Heinrich, and H. Niederreiter, eds., Springer-Verlag, 2007, pp. 343–358.

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

[26] M. Giles, D. J. Higham, and X. Mao, *Analysing multi-level Monte Carlo for options with non-globally Lipschitz payoff*, Finance and Stochastics, 13 (2009), pp. 403–413.

[27] D. T. Gillespie, *A general method for numerically simulating the stochastic time evolution of coupled chemical reactions*, J. Comput. Phys., 22 (1976), pp. 403–434.

[28] ——, *Exact stochastic simulation of coupled chemical reactions*, J. Phys. Chem., 81 (1977), pp. 2340–2361.

[29] D. T. Gillespie, *Markov Processes: An Introduction for Physical Scientists*, Academic Press, San Diego, 1991.

[30] ——, *The chemical Langevin equation*, J. Chem. Phys., 113 (2000), pp. 297–306.

[31] D. T. Gillespie, *Approximate accelerated simulation of chemically reacting systems*, J. Chem. Phys., 115 (2001), pp. 1716–1733.

[32] D. T. Gillespie, A. Hellander, and L. Petzold, *Perspective: Stochastic algorithms for chemical kinetics*, J. Chem. Phys., 138 (2013), p. 170901.

[33] D. J. Higham, X. Mao, M. Roj, Q. Song, and G. Yin, *Mean exit times and the multilevel Monte Carlo method*, SIAM/ASA Journal on Uncertainty Quantification, 1 (2013), pp. 2–18.

[34] S. Hoops, S. Sahle, R. Gauges, C. Lee, J. Pahle, N. Simus, M. Singhal, L. Xu, P. Mendes, and U. Kummer, *Copasi—a complex pathway simulator*, Bioinformatics, 22 (2006), pp. 3067–3074.

[35] M. Hutzenthaler, A. Jenstzen, and P. E. Kloeden, *Divergence of the multilevel Monte Carlo method for nonlinear stochastic differential equations*, Annals of Applied Probability, 23 (2013), pp. 1913–1966.

[36] M. Hutzenthaler, A. Jentzen, and P. E. Kloeden, *Divergence of the multilevel Monte Carlo method*, on arXiv, (2011).

[37] M. Hutzenthaler, A. Jentzen, and P. E. Kloeden, *Strong and weak divergence in finite time of Euler's method for stochastic differential equations with non-globally Lipschitz continuous coefficients*, Proceedings of the Royal Society A, 467 (2011), pp. 1563–1576.

[38] 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.

[39] 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.

[40] T. Jahnke, *On reduced models for the chemical master equation*, SIAM: Multiscale Modelling & Simulation, 9 (2011), pp. 1646–1676.

[41] ——, *On reduced models for the chemical master equation*, SIAM: Multiscale Modeling and Simulation, 9 (2011), pp. 1646–1676.

[42] V. Kazeev, M. Khammash, M. Nip, and C. Schwab, *Direct solution of the chemical master equation using quantized tensor trains*, PLOS Computational Biology, (2014).

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

[44] T. G. Kurtz, *The relationship between stochastic and deterministic models for chemical reactions*, J. Chem. Phys., 57 (1972), pp. 2976–2978.

[45] ——, *Strong approximation theorems for density dependent Markov chains*, Stoch. Proc. Appl., 6 (1978), pp. 223–240.

[46] ——, *Representations of Markov processes as multiparameter time changes*, Ann. Prob., 8 (1980), pp. 682–715.

[47] ——, *Representation and approximation of counting processes*, vol. 42 of Advances in filtering and optimal stochastic control, Springer, Berlin, 1982.

[48] B. Lehner, *Selection to minimise noise in living systems and its implications for the evolution of gene expression*, Molecular systems biology, 4 (2008), p. 170.

[49] M. Lynch and J. S. Conery, *The evolutionary fate and consequences of duplicate genes*, Science, 290 (2000), pp. 1151–1155.

[50] S. MacNamara, K. Burrage, and R. B. Sidje, *Multiscale modeling of chemical kinetics via the master equation*, SIAM: Multiscale Modeling & Simulation, 6 (2008), pp. 1146–1168.

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

[52] X. Mao, G. Marion, and E. Renshaw, *Environmental noise suppresses explosion in population dynamics*, Stochastic Process Appl., 97 (2002), pp. 96–110.

[53] H. H. McAdams and A. Arkin, *Stochastic mechanisms in gene expression*, Proceedings of the National Academy of Sciences, 94 (1997), pp. 814–819.

[54] D. A. McQuarrie, *Stochastic approach to chemical kinetics*, J. Appl. Prob., 4 (1967), pp. 413–478.

[55] 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.

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

[57] G. N. Milstein and M. V. Tretyakov, *Stochastic Numerics for Mathematical Physics*, Springer-Verlag, Berlin, 2004.

[58] T. Müller-Gronbach and L. Yaroslavtseva, *Deterministic quadrature formulas for SDEs based on simplified weak Ito-Taylor steps*, Tech. Rep. 167, DFG-Schwerpunktprogramm 1324, 2014.

[59] J. Otwinowski, S. Tanase-Nicola, and I. Nemenman, *Speeding up evolutionary search by small fitness fluctuations*, Journal of Statistical Physics, 144 (2011), pp. 367–378.

[60] G. Pavliotis and A. Stuart, *Parameter estimation for multiscale diffusions*, Journal of Statistical Physics, 127 (2007), pp. 741–781.

[61] A. Pikovsky, M. Rosenblum, and J. Kurths, *Synchronization: A Universal Concept in Nonlinear Sciences*, Cambridge University Press, Cambridge, 2001.

[62] M. Rathinam, L. R. Petzold, Y. Cao, and D. T. Gillespie, *Consistency and stability of tau-leaping schemes for chemical reaction systems*, SIAM: Multiscale Model. Simul., 3 (2005), pp. 867–895.

[63] E. Renshaw, *Stochastic Population Processes*, Oxford University Press, Oxford, 2011.

[64] S. M. Ross, *Simulation*, Academic Press, Burlington, MA, fourth ed., 2006.

[65] B. A. Schmerl and M. D. McDonnell, *Channel-noise-induced stochastic facilitation in an auditory brainstem neuron model*, Phys. Rev. E, 88 (2013), p. 052722.

[66] A. Shishkin and D. Postnov, *Stochastic dynamics of Fitzhugh-Nagumo model near the canard explosion*, in Physics and Control, 2003. Proceedings, vol. 2, 2003, pp. 649–653 vol.2.

[67] M. Sieber, H. Malchow, and S. V. Petrovskii, *Noise-induced suppression of periodic travelling waves in oscillatory reaction-diffusion systems*, Proc. R. Soc. A, (2010), pp. 1903–1917.

[68] T. Tian and K. Burrage, *Stochastic models for regulatory networks of the genetic toggle switch*, Proceedings of the National Academy of Sciences, 103 (2006), pp. 8372–8377.

[69] H. C. Tuckwell, R. Rodriguez, and F. Y. M. Wan, *Determination of firing times for the stochastic Fitzhugh-Nagumo neuronal model*, Neural Computation, 15 (2003), pp. 143–159.

[70] D. J. Wilkinson, *Stochastic modelling for quantitative description of heterogeneous biological systems*, Nature Reviews Genetics, 10 (2009), pp. 122–133.

[71] D. J. Wilkinson, *Stochastic Modelling for Systems Biology*, Chapman and Hall/CRC Press, second ed., 2011.

[72] L. Zhang, P. A. Mykland, and Y. Aït-Sahalia, *A tale of two time scales: Determining integrated volatility with noisy high-frequency data*, Journal of the American Statistical Association, 100 (2005), pp. 1394–1411.