ORIGINAL PAPER

Multilevel importance sampling for rare events associated with the McKean–Vlasov equation

Nadhir Ben Rached[1] · Abdul-Lateef Haji-Ali[2] · Shyam Mohan Subbiah Pillai[3] · Raúl Tempone[4,5]

Abstract

This work combines multilevel Monte Carlo with importance sampling to estimate rare-event quantities that can be expressed as the expectation of a Lipschitz observable of the solution to a broad class of McKean–Vlasov stochastic differential equations. We extend the double loop Monte Carlo (DLMC) estimator introduced in this context in Ben Rached et al. (Stat Comput, 2024. https://doi.org/10.1007/s11222-024-10497-3) to the multilevel setting. We formulate a novel multilevel DLMC estimator and perform a comprehensive cost-error analysis yielding new and improved complexity results. Crucially, we devise an antithetic sampler to estimate level differences guaranteeing reduced computational complexity for the multilevel DLMC estimator compared with the single-level DLMC estimator. To address rare events, we apply the importance sampling scheme, obtained via stochastic optimal control in Ben Rached et al. (2024), over all levels of the multilevel DLMC estimator. Combining importance sampling and multilevel DLMC reduces computational complexity by one order and drastically reduces the associated constant compared to the single-level DLMC estimator without importance sampling. We illustrate the effectiveness of the proposed multilevel DLMC estimator on the Kuramoto model from statistical physics with Lipschitz observables, confirming the reduced complexity from $\mathcal{O}(\text{TOL}_r^{-4})$ for the single-level DLMC estimator to $\mathcal{O}(\text{TOL}_r^{-3})$ while providing a feasible estimate of rare-event quantities up to prescribed relative error tolerance TOL_r.

✉ Shyam Mohan Subbiah Pillai
subbiah@uq.rwth-aachen.de

Nadhir Ben Rached
N.BenRached@leeds.ac.uk

Abdul-Lateef Haji-Ali
A.HajiAli@hw.ac.uk

Raúl Tempone
raul.tempone@kaust.edu.sa; tempone@uq.rwth-aachen.de

[1] Department of Statistics, School of Mathematics, University of Leeds, Leeds, UK

[2] Department of Actuarial Mathematics and Statistics, School of Mathematical and Computer Sciences, Heriot-Watt University, Edinburgh, UK

[3] Chair of Mathematics for Uncertainty Quantification, Department of Mathematics, RWTH Aachen University, Aachen, Germany

[4] Computer, Electrical and Mathematical Sciences and Engineering Division (CEMSE), King Abdullah University of Science and Technology (KAUST), Thuwal, Saudi Arabia

[5] Alexander von Humboldt Professor in Mathematics for Uncertainty Quantification, RWTH Aachen University, Aachen, Germany

1 Introduction

We consider the estimation of rare-event quantities expressed as an expectation of some observable of the solution to

a broad class of McKean–Vlasov stochastic differential equations (MV-SDEs). In particular, we develop a computationally efficient multilevel Monte Carlo (MLMC) estimator for $\mathbb{E}[G(X(T))]$, where $G : \mathbb{R}^d \rightarrow \mathbb{R}$ is a Lipschitz function and $X : [0, T] \times \Omega \rightarrow \mathbb{R}^d$ is the MV-SDE process up to a finite terminal time T. We consider G to be a rare-event observable, meaning that the support of G lies deep in the tails of the distribution of $X(T)$. MV-SDEs are a special class of SDEs whose drift and diffusion coefficients are a function of the law of the solution (McKean Jr 1966). Such equations arise from the mean-field behavior of stochastic interacting particle systems used in diverse applications (Haji Ali 2012; Erban and Haskovec 2011; Acebrón et al. 2005; Forrester 2015; Dobramysl et al. 2016; Bush et al. 2011). Significant recent literature has addressed the analysis (Mishura and Veretennikov 2020; Buckdahn et al. 2017; Crisan and McMurray 2018) and numerical treatment (Haji-Ali and Tempone 2018; dos Reis et al. 2022; Szpruch et al. 2019; Crisan and McMurray 2019) of MV-SDEs. MV-SDEs are often approximated using stochastic P-particle systems, sets of P coupled d-dimensional SDEs that approach a mean-field limit as the number of particles tends to infinity (Sznitman 1991). The associated Kolmogorov backward equation is Pd-dimensional; hence, Monte Carlo (MC) methods are used to approximate expectations associated with particle systems. MC methods using Euler–Maruyama time-discretized particle systems for bounded, Lipschitz drift/diffusion coefficients have been proposed for smooth, nonrare observables with a complexity of $\mathcal{O}(\text{TOL}^{-4})$ for a prescribed error tolerance TOL (Ogawa 1992; Haji-Ali and Tempone 2018; Li et al. 2023).

MLMC was introduced as an improvement to MC for SDEs in Giles (2008). MLMC is based on generalized control variates and improves efficiency when an approximation of the solution is computed based on a discretization parameter (Giles 2015). Most MLMC simulations are performed at cheaper, coarse levels, with relatively few simulations applied at the costlier, fine levels. MLMC for particle systems has been widely investigated (Haji-Ali and Tempone 2018; Rosin et al. 2014; Szpruch and Tse 2019; Bujok et al. 2015). In particular, an MLMC estimator with a partitioning sampler achieved a computational complexity of $\mathcal{O}(\text{TOL}^{-3})$ for smooth observables and bounded, Lipschitz drift/diffusion coefficients (Haji-Ali and Tempone 2018). However, MC and MLMC methods become extremely expensive in the context of rare events due to the 'blowing up' of the constant associated with the estimator complexity as the event becomes rarer (Kroese et al. 2013). This motivates using

importance sampling as a variance reduction technique to overcome the failure of standard MC and MLMC in the rare-event regime (Kroese et al. 2013).

Importance sampling for MV-SDEs has been studied in dos Reis et al. (2023), Ben Rached et al. (2024). The decoupling approach developed by dos Reis et al. (2023) defines a modified, decoupled MV-SDE with coefficients computed using a realization of the MV-SDE law estimated beforehand using a stochastic particle system. A change of measure is applied to the decoupled MV-SDE, decoupling importance sampling from the law estimation. The theory of large deviations and Pontryagin principle were employed in dos Reis et al. (2023) to obtain a deterministic, time-dependent control minimizing a proxy for the estimator variance. Ben Rached et al. (2024) employed the same decoupling approach to define a double loop MC (DLMC) estimator and employ stochastic optimal control to derive a time- and state-dependent control minimizing the variance of the importance sampling estimator. In Ben Rached et al. (2024), an adaptive DLMC algorithm with a complexity of $\mathcal{O}(\text{TOL}^{-4})$ was developed, the same as that for the MC estimator for nonrare observables in Haji-Ali and Tempone (2018), while enabling feasible estimates for rare-event probabilities. The development of such an importance sampling scheme using stochastic optimal control theory has been proposed before in other contexts, including standard SDEs (Hartmann et al. 2017, 2018; Zhang et al. 2014), stochastic reaction networks (Ben Hammouda et al. 2023), and discrete-time continuous-space Markov chains (Ben Amar et al. 2023; Dupuis and Wang 2004).

We combine importance sampling with MLMC to reduce the relative estimator variance in the rare-event regime by extending the results in Ben Rached et al. (2024) to the multilevel setting. Combining importance sampling with MLMC has been previously explored in other contexts, including standard SDEs (Ben Alaya et al. 2023; Kebaier and Lelong 2018; Fang and Giles 2019) and stochastic reaction networks (Ben Hammouda et al. 2020). We extend the DLMC estimator from Ben Rached et al. (2024) to the multilevel setting by introducing a multilevel DLMC estimator along with a detailed error and complexity analysis. We boost the efficiency of the multilevel DLMC estimator by developing a highly correlated, antithetic sampler for level differences (Haji-Ali and Tempone 2018; Giles and Szpruch 2014). We show reduced computational complexity using the multilevel DLMC estimator compared with the DLMC estimator for MV-SDEs. Then, we propose an importance sampling scheme for this estimator based on variance minimization using stochastic optimal control theory (Ben Rached et al. 2024) to address rare events and apply the obtained control on all levels. Contributions of this paper are summarized as follows.

- We extend the DLMC estimator introduced by Ben Rached et al. (2024) to the multilevel setting and propose a multilevel DLMC estimator for the decoupling approach (dos Reis et al. 2023) for MV-SDEs. We include a detailed discussion on the bias and variance of the proposed estimator and devise a complexity theorem, displaying improved complexity compared with the naïve DLMC estimator. We also formulate a robust multilevel DLMC algorithm to determine optimal parameters adaptively.
- We develop naïve and antithetic correlated samplers for level differences in the multilevel DLMC estimator. Numerical simulations confirm increased variance convergence rates for level-difference estimators using the antithetic sampler compared with the naïve one, leading to improved multilevel DLMC estimator complexity.
- We propose combining importance sampling with the multilevel DLMC estimator to address rare events. We employ the time- and state-dependent control developed by Ben Rached et al. (2024) for all levels in the multilevel DLMC estimator. Numerical simulations confirm a significant variance reduction in the multilevel DLMC estimator due to this importance sampling scheme, improving estimator complexity from $\mathcal{O}(\text{TOL}_\text{r}^{-4})$ in Ben Rached et al. (2024) for the Kuramoto model with Lipschitz observables to $\mathcal{O}(\text{TOL}_\text{r}^{-3})$ in the multilevel setting while allowing feasible rare-event quantity estimation up to the prescribed relative error tolerance TOL_r.

The remainder of this paper is structured as follows. In Sect. 2, we introduce the MV-SDE and associated notation, motivate MC methods to estimate expectations associated with its solution and set forth the problem to be solved. In Sect. 3, we introduce the decoupling approach for MV-SDEs (dos Reis et al. 2023) and formulate a DLMC estimator. Next, we state the optimal importance sampling control for the decoupled MV-SDE derived using stochastic optimal control and introduce the DLMC estimator with importance sampling from Ben Rached et al. (2024) in Sect. 4. Then, we introduce the novel multilevel DLMC estimator in Sect. 5, develop an antithetic sampler for it, and derive new complexity results for the estimator. We combine the multilevel DLMC estimator with the proposed importance sampling

scheme and develop an adaptive multilevel DLMC algorithm that feasibly estimates rare-event quantities associated with MV-SDEs. Finally, we apply the proposed methods to the Kuramoto model from statistical physics in Sect. 6 and numerically verify all assumptions in this work and the derived complexity rates for the multilevel DLMC estimator for two observables.

2 McKean–Vlasov stochastic differential equation

In this work, we consider a broad class of McKean–Vlasov equations that arise from the mean-field limit of stochastic interacting particle systems with pairwise interaction kernels (Sznitman 1991). We consider the probability space $\{\Omega, \mathcal{F}, \{\mathcal{F}_t\}_{t\geq 0}, P\}$, where $\mathcal{F}_t$ is the filtration of a standard Wiener process. For functions $b : \mathbb{R}^d \times \mathbb{R} \longrightarrow \mathbb{R}^d$, $\sigma : \mathbb{R}^d \times \mathbb{R} \longrightarrow \mathbb{R}^{d\times d}$, $\kappa_1 : \mathbb{R}^d \times \mathbb{R}^d \longrightarrow \mathbb{R}$, and $\kappa_2 : \mathbb{R}^d \times \mathbb{R}^d \longrightarrow \mathbb{R}$, we consider the following Itô SDE for the McKean–Vlasov stochastic process $X : [0, T] \times \Omega \to \mathbb{R}^d$:

$$
\begin{cases}
\mathrm{d}X(t) = b\left(X(t), \int_{\mathbb{R}^d} \kappa_1(X(t), x)\mu_t(\mathrm{d}x)\right)\mathrm{d}t \\
\qquad\quad + \sigma\left(X(t), \int_{\mathbb{R}^d} \kappa_2(X(t), x)\mu_t(\mathrm{d}x)\right)\mathrm{d}W(t), \quad t > 0 \\
X(0) = x_0 \sim \mu_0 \in \mathcal{P}(\mathbb{R}^d),
\end{cases}
\tag{1}
$$

where $W : [0, T] \times \Omega \to \mathbb{R}^d$ is a standard d-dimensional Wiener process with mutually independent components. $\mu_t \in \mathcal{P}(\mathbb{R}^d)$ is the probability distribution of $X(t)$, where $\mathcal{P}(\mathbb{R}^d)$ is the space of probability measures on $\mathbb{R}^d$. $x_0 \in \mathbb{R}^d$ is a random initial state with distribution $\mu_0 \in \mathcal{P}(\mathbb{R}^d)$. The functions b and σ are called drift and diffusion functions/coefficients, respectively. The existence and uniqueness of solutions to (1) follows from the results in Mishura and Veretennikov (2020), Crisan and Xiong (2010), Sznitman (1991), Hammersley et al. (2021), under the assumptions therein. These involve certain differentiability and boundedness conditions on b, σ, κ_1, and κ_2. The time evolution of μ_t is given by the multidimensional Fokker–Planck partial differential equation (PDE):

$$\begin{cases} -\dfrac{\partial \mu(s, x; t, y)}{\partial s} - \sum_{i=1}^{d} \dfrac{\partial}{\partial x_i} \left(b_i \left(x, \int_{\mathbb{R}^d} \kappa_1(x, z) \mu(s, z; t, y) \mathrm{d}z \right) \mu(s, x; t, y) \right) \\ \qquad + \sum_{i=1}^{d} \sum_{j=1}^{d} \dfrac{1}{2} \dfrac{\partial^2}{\partial x_i \partial x_j} \left(\sum_{k=1}^{d} \sigma_{ik} \sigma_{jk} \left(x, \int_{\mathbb{R}^d} \kappa_2(x, z) \mu(s, z; t, y) \mathrm{d}z \right) \mu(s, x; t, y) \right) \\ = 0, (s, x) \in (t, \infty) \times \mathbb{R}^d \, \mu(t, x; t, y) = \delta_y(x), \end{cases} \tag{2}$$

where $\mu(s, x; t, y)$ denotes the conditional distribution of $X(s)$ given that $X(t) = y$, and $\delta_y(\cdot)$ denotes the Dirac measure at y. Equation (2) is a nonlinear integral PDE with nonlocal terms. Solving such an equation using classical numerical methods up to relative error tolerances can be computationally prohibitive, particularly in higher dimensions ($d \gg 1$).

A strong approximation of the solution to the above class of MV-SDEs can be obtained by solving a system of P exchangeable Itô SDEs, also known as a stochastic interacting particle system, with pairwise interaction kernels comprising P particles (Sznitman 1991). For $p = 1, \ldots, P$, we have the following SDE for the process $X_p^P : [0, T] \times \Omega \to \mathbb{R}^d$:

$$\begin{cases} \mathrm{d}X_p^P(t) = b \left(X_p^P(t), \dfrac{1}{P} \sum_{j=1}^{P} \kappa_1(X_p^P(t), X_j^P(t)) \right) \mathrm{d}t \\ \quad + \sigma \left(X_p^P(t), \dfrac{1}{P} \sum_{j=1}^{P} \kappa_2(X_p^P(t), X_j^P(t)) \right) \mathrm{d}W_p(t), \quad t > 0 \\ X_p^P(0) = (x_0)_p \sim \mu_0 \in \mathcal{P}(\mathbb{R}^d), \end{cases} \tag{3}$$

where $\{(x_0)_p\}_{p=1}^{P}$ are independent and identically distributed (i.i.d.) random variables sampled from the initial distribution μ_0, and $\{W_p\}_{p=1}^{P}$ are mutually independent d-dimensional Wiener processes also independent of $\{(x_0)_p\}_{p=1}^{P}$. Equation (3) approximates the mean-field distribution μ_t from (1) using an empirical distribution based on particles $\{X_p^P\}_{p=1}^{P}$:

$$\mu_t(\mathrm{d}x) \approx \mu_t^P(\mathrm{d}x) = \frac{1}{P} \sum_{j=1}^{P} \delta_{X_j^P(t)}(\mathrm{d}x), \tag{4}$$

where particles $\{X_p^P\}_{p=1}^{P}$ are identically distributed but not mutually independent due to pairwise interaction kernels in the drift and diffusion coefficients. The strong convergence of particle systems follows from the results given in Bossy and

Talay (1997, 1996), Méléard (1996), under further assumptions on the drift and diffusion coefficients therein. The high dimensionality of the Fokker–Planck equation, satisfied by the joint probability density of the particle system, motivates the use of MC methods, which do not suffer from the curse of dimensionality.

2.1 Example: fully connected Kuramoto model for synchronized oscillators

We focus on a one-dimensional (1D) example of the MV-SDE in (1), called the Kuramoto model, which describes synchronization in statistical physics to help model the behavior of large sets of coupled oscillators. This model has widespread applications in chemical and biological systems (Acebrón et al. 2005), neuroscience (Cumin and Unsworth 2007), and oscillating flame dynamics (Forrester 2015). In particular, the Kuramoto model is a system of P fully connected, synchronized oscillators. We consider the following Itô SDE for the process $X_p^P : [0, T] \times \Omega \to \mathbb{R}$:

$$\begin{cases} \mathrm{d}X_p^P(t) = \left(\xi_p + \dfrac{1}{P} \sum_{q=1}^{P} \sin \left(X_p^P(t) - X_q^P(t) \right) \right) \mathrm{d}t + \sigma \mathrm{d}W_p(t), \quad t > 0 \\ X_p^P(0) = (x_0)_p \sim \mu_0 \in \mathcal{P}(\mathbb{R}), \end{cases} \tag{5}$$

where $\{\xi_p\}_{p=1}^{P}$ denotes i.i.d. random variables sampled from a prescribed distribution. The diffusion $\sigma \in \mathbb{R}$ is constant, and $\{(x_0)_p\}_{p=1}^{P}$ represents i.i.d. random variables sampled from a prescribed distribution μ_0. In addition, $\{W_p\}_{p=1}^{P}$ represents mutually independent 1D Wiener processes, and $\{\xi_p\}_{p=1}^{P}$, $\{(x_0)_p\}_{p=1}^{P}$, and $\{W_p\}_{p=1}^{P}$ are mutually independent. This coupled particle system reaches the mean-field limit as the number of oscillators tends to infinity. In this limit, each particle satisfies the following MV-SDE:

$$\begin{cases} \mathrm{d}X(t) = \left(\xi + \int_{\mathbb{R}} \sin(X(t) - x) \mu_t(\mathrm{d}x) \right) \mathrm{d}t + \sigma \mathrm{d}W(t), t > 0 \\ X(0) = x_0 \sim \mu_0 \in \mathcal{P}(\mathbb{R}), \end{cases} \tag{6}$$

where $X(t)$ denotes the state of each particle at time t, ξ represents a random variable sampled from some prescribed

distribution, and μ_t is the probability distribution of $X(t)$. This example is used throughout this work as a test case for the proposed MC algorithms.

2.2 Problem setting

We let $T > 0$ be some finite terminal time and $X : [0, T] \times \Omega \to \mathbb{R}^d$ denote the McKean-Vlasov process (1). We let $G : \mathbb{R}^d \longrightarrow \mathbb{R}$ be a given Lipschitz rare-event observable. The objective is to build a computationally efficient MLMC estimator $\mathcal{A}_{\mathrm{MLMC}}$ for $\mathbb{E}[G(X(T))]$ with a given relative tolerance $\mathrm{TOL_r} > 0$ that satisfies

$$\mathbb{P}\left[\frac{|\mathcal{A}_{\mathrm{MLMC}} - \mathbb{E}[G(X(T))]|}{|\mathbb{E}[G(X(T))]|} \geq \mathrm{TOL_r}\right] \leq \nu, \tag{7}$$

for a given confidence level determined by $0 < \nu \ll 1$. The high dimensionality of the Kolmogorov backward equation corresponding to the stochastic particle system (3) makes numerical solutions of $\mathbb{E}[G(X(T))]$ up to some $\mathrm{TOL_r}$ computationally infeasible. This motivates employing MC methods to overcome the curse of dimensionality. Combining MC with variance reduction techniques, such as importance sampling, is required to produce feasible rare-event estimates.

In Ben Rached et al. (2024), we introduced the DLMC estimator, based on a decoupling approach (dos Reis et al. 2023) to provide a simple importance sampling scheme implementation minimizing the estimator variance. The current paper extends the DLMC estimator to the multilevel setting, achieving better complexity than $\mathcal{O}(\mathrm{TOL_r^{-4}})$ from the single-level DLMC estimator. Section 3 introduces the decoupling approach for MV-SDEs and associated notation before introducing the DLMC estimator.

3 Double loop Monte Carlo estimator using a decoupling approach

The decoupling approach was developed for importance sampling in MV-SDEs (dos Reis et al. 2023; Ben Rached et al. 2024), where the idea is to approximate the MV-SDE law empirically as in (4), use the approximation as input to define a decoupled MV-SDE and apply a change of measure to it. We decouple the computation of the MV-SDE law and the change in probability measure required for importance sampling. First, we introduce the general decoupling approach.

3.1 Decoupling approach for McKean–Vlasov stochastic differential equation

The decoupling approach in Ben Rached et al. (2024), dos Reis et al. (2023) comprises the following steps.

1. We approximate the MV-SDE law $\{\mu_t : t \in [0, T]\}$ using the empirical measure $\{\mu_t^P : t \in [0, T]\}$ in (4) using particles $\{X_p^P(t) : t \in [0, T]\}_{p=1}^P$ satisfying (3).

2. Given $\{\mu_t^P : t \in [0, T]\}$, we define the decoupled MV-SDE for the process $\bar{X}^P : [0, T] \times \Omega \to \mathbb{R}^d$ as

$$\begin{cases} \mathrm{d}\bar{X}^P(t) = b\left(\bar{X}^P(t), \frac{1}{P}\sum_{j=1}^P \kappa_1(\bar{X}^P(t), X_j^P(t))\right)\mathrm{d}t \\ \quad + \sigma\left(\bar{X}^P(t), \frac{1}{P}\sum_{j=1}^P \kappa_2(\bar{X}^P(t), X_j^P(t))\right)\mathrm{d}\bar{W}(t), t \in [0, T] \\ \bar{X}^P(0) = \bar{x}_0 \sim \mu_0, \quad \bar{x}_0 \in \mathbb{R}^d, \end{cases} \tag{8}$$

where superscript P indicates that the drift and diffusion functions in (8) are computed using $\{\mu_t^P : t \in [0, T]\}$ derived from the stochastic P-particle system. Drift and diffusion coefficients b and σ are the same as defined in Sect. 2. In addition, $\bar{W} : [0, T] \times \mathbb{R}^d \to \mathbb{R}^d$ is a standard d-dimensional Wiener process independent of the Wiener processes $\{W_p\}_{p=1}^P$ used in (3), and $\bar{x}_0 \in \mathbb{R}^d$ is a random initial state sampled from μ_0 as defined in (1) and is independent from $\{(x_0)_p\}_{p=1}^P$ in (3). Thus, (8) is a standard SDE with random coefficients.

3. We introduce a copy space (see dos Reis et al. (2023)) to distinguish the decoupled MV-SDE (8) from the stochastic P-particle system. We suppose (3) is defined on the probability space $(\Omega, \mathcal{F}, \mathbb{P})$. We define a copy space $(\bar{\Omega}, \bar{\mathcal{F}}, \bar{\mathbb{P}})$; hence, we define (8) on the product space $(\Omega, \mathcal{F}, \mathbb{P}) \times (\bar{\Omega}, \bar{\mathcal{F}}, \bar{\mathbb{P}})$. Thus, $\mathbb{P}$ is a probability measure generated by the randomness of $\{\mu_t^P : t \in [0, T]\}$, and $\bar{\mathbb{P}}$ denotes the measure generated by the randomness of the Wiener process driving (8) conditioned on $\{\mu_t^P : t \in [0, T]\}$.

4. We approximate the quantity of interest as

$$\mathbb{E}[G(X(T))] \approx \mathbb{E}_{\mathbb{P}\otimes\bar{\mathbb{P}}}\left[G(\bar{X}^P(T))\right]$$
$$= \mathbb{E}_{\mathbb{P}}\left[\mathbb{E}_{\bar{\mathbb{P}}}\left[G(\bar{X}^P(T)) \mid \{\mu_t^P : t \in [0, T]\}\right]\right]. \tag{9}$$

Henceforth, we omit the probability measure above to simplify the notation such that $\mathbb{E}\left[G(\bar{X}^P(T))\right]$ means the expectation is taken with respect to all randomness in the decoupled MV-SDE (8). We estimate the inner expectation $\mathbb{E}\left[G(\bar{X}^P(T)) \mid \{\mu_t^P : t \in [0, T]\}\right]$ for a given $\{\mu_t^P : t \in [0, T]\}$ and then estimate the outer expectation using MC sampling over different realizations of $\{\mu_t^P : t \in [0, T]\}$.

The inner expectation $\mathbb{E}\left[G(\bar{X}^P(T)) \mid \{\mu_t^P : t \in [0, T]\}\right]$ can be obtained using the Kolmogorov backward equation for

the decoupled MV-SDE (8). Obtaining an analytical solution to the Kolmogorov backward equation is not always possible, and conventional numerical methods do not handle relative error tolerances, which are relevant in the rare-event regime. This motivates MC methods coupled with importance sampling, even for the 1D case, to estimate the nested expectation (9) in the rare-event regime. We approximate the nested expectation (9) using a DLMC estimator, as in Ben Rached et al. (2024). The general outline of the DLMC estimator is given in Algorithm 1. In the following, we use the notation $\omega_{p:P}^{(i)} \stackrel{\text{def}}{=} \left(\omega_q^{(i)} \right)_{q=p}^P$, where for each q, $\omega_q^{(i)}$ denotes the ith sample of the set of underlying random variables that are used in calculating the empirical law μ_t^P. Hence, $\mu_t^P(\omega_{1:P}^{(i)})$ denotes the ith realization of the empirical law (4). We let $\bar{\omega}^{(i)}$ denote the ith realization of random variables driving the decoupled MV-SDE dynamics (8) conditioned on a realization of the empirical law.

Algorithm 1: Outline of the double loop Monte Carlo algorithm for decoupled McKean–Vlasov stochastic differential equation

---

Inputs: P, M_1, M_2;
for $i = 1, \ldots, M_1$ **do**
  Generate $\left\{ \mu_t^P(\omega_{1:P}^{(i)}) : t \in [0, T] \right\}$ using (3),(4);
  for $j = 1, \ldots, M_2$ **do**
    Given $\left\{ \mu_t^P(\omega_{1:P}^{(i)}) : t \in [0, T] \right\}$, generate sample path of
    (8) using $\bar{\omega}^{(j)}$;
    Compute $G\left(\bar{X}^P(T) \right) \left(\omega_{1:P}^{(i)}, \bar{\omega}^{(j)} \right)$;
  end
end
Approximate $\mathbb{E}\left[G\left(X(T) \right) \right]$ by
$\frac{1}{M_1} \sum_{i=1}^{M_1} \frac{1}{M_2} \sum_{j=1}^{M_2} G\left(\bar{X}^P(T) \right) \left(\omega_{1:P}^{(i)}, \bar{\omega}^{(j)} \right)$;

The decoupled MV-SDE (8) for the given empirical law $\left\{ \mu_t^P : t \in [0, T] \right\}$ is a standard SDE, making it possible to use stochastic optimal control to derive an optimal change of measure minimizing the variance of the estimator of the inner expectation $\mathbb{E}\left[G\left(\bar{X}^P(T) \right) \mid \left\{ \mu_t^P : t \in [0, T] \right\} \right]$, as formulated in previous studies (Hartmann et al. 2017, 2018; Zhang et al. 2014). Such an importance sampling scheme for the decoupled MV-SDE was derived in Ben Rached et al. (2024) and is summarized in Sect. 4.

4 Importance sampling for the decoupled McKean–Vlasov stochastic differential equation

This section applies stochastic optimal control theory to obtain the optimal change of measure for the decoupled

MV-SDE (8). Then, we incorporate the above importance sampling scheme to the DLMC Algorithm 1, and formulate the DLMC estimator with importance sampling.

4.1 Optimal importance sampling control for decoupled McKean–Vlasov stochastic differential equation

Ben Rached et al. (2024) derived an optimal change of measure for the decoupled MV-SDE, minimizing the MC estimator variance based on stochastic optimal control. This was based on the well-known Girsanov theorem for change of measure (Oksendal 2013) for standard SDEs. We recall the main results here. First, we formulate the Hamilton–Jacobi–Bellman control PDE that provides optimal control for the decoupled MV-SDE. We introduce the following notation: $\langle \cdot, \cdot \rangle$ is the Euclidean dot product between two functions in $\mathbb{R}^d$, $\nabla \cdot$ denotes the gradient vector of a scalar function, $\nabla^2 \cdot$ represents the Hessian matrix of a scalar function, $\cdot : \cdot$ indicates the Frobenius inner product between two matrix-valued functions, $\|\cdot\|$ is the Euclidean norm of a function in $\mathbb{R}^d$ and $C^k(A, B)$ denotes the set of real-valued bounded continuous functions from domain A to set B with $k \geq 1$ bounded continuous derivatives on B.

Proposition 1 (Hamilton–Jacobi–Bellman PDE for decoupled MV-SDE (Ben Rached et al. 2024)) *Let the process $\bar{X}^P$ follow the dynamics (8). We consider the following Itô SDE for the controlled process $\bar{X}_\zeta^P : [0, T] \times \Omega \to \mathbb{R}^d$ with control $\zeta : [0, T] \times \mathbb{R}^d \to \mathbb{R}^d$:*

$$
\begin{cases}
\mathrm{d}\bar{X}_\zeta^P(t) = \left(b\left(\bar{X}_\zeta^P(t), \frac{1}{P} \sum_{j=1}^P \kappa_1(\bar{X}_\zeta^P(t), X_j^P(t)) \right) \right. \\
\left. + \sigma\left(\bar{X}_\zeta^P(t), \frac{1}{P} \sum_{j=1}^P \kappa_2(\bar{X}_\zeta^P(t), X_j^P(t)) \right) \zeta(t, \bar{X}_\zeta^P(t)) \right) \mathrm{d}t \\
+ \sigma\left(\bar{X}_\zeta^P(t), \frac{1}{P} \sum_{j=1}^P \kappa_2(\bar{X}_\zeta^P(t), X_j^P(t)) \right) \mathrm{d}W(t), 0 < t < T \\
\bar{X}_\zeta^P(0) = \bar{X}^P(0) = \bar{x}_0 \sim \mu_0.
\end{cases}
\tag{10}
$$

where (3) is used to compute $\{\mu_t^P : t \in [0, T]\}$ in (8) and (10). The value function $u : [0, T] \times \mathbb{R}^d \to \mathbb{R}^d$ minimizing the second moment (for the derivation, see Ben Rached et al. (2024)) of the DLMC estimator with importance sampling is written as

$$u(t, x) = \min_{\zeta \in \mathcal{Z}} \mathbb{E}\left[G^2(\bar{X}_\zeta^P(T)) \exp\left\{ -\int_t^T \|\zeta(s, \bar{X}_\zeta^P(s))\|^2 ds \right.\right.$$
$$\left. -2\int_t^T \langle \zeta(s, \bar{X}_\zeta^P(s)), dW(s) \rangle \right\}$$
$$\left| \; \bar{X}_\zeta^P(t) = x, \{\mu_t^P : t \in [0, T]\} \right].$$
$$(11)$$

Here $\mathcal{Z} = \left\{ f : f \in C^1\left([0, T] \times \mathbb{R}^d, \mathbb{R}^d\right) \right\}$ is the set of admissible deterministic d-dimensional Markov controls. Assume b and σ are sufficiently regular such that u has bounded and continuous derivatives up to first order in time and second order in space and $u(t, x) \neq 0 \; \forall(t, x) \in [0, T] \times \mathbb{R}^d$. We define $\gamma : [0, T] \times \mathbb{R}^d \to \mathbb{R}^d$, such that $u(t, x) = \exp{-2\gamma(t, x)}$. Then, γ satisfies the nonlinear Hamilton–Jacobi–Bellman equation:

$$\begin{cases} \dfrac{\partial \gamma}{\partial t} + \langle b\left(x, \dfrac{1}{P}\sum_{j=1}^P \kappa_1(x, X_j^P(t))\right), \nabla\gamma\rangle \\ \quad + \dfrac{1}{2}\nabla^2\gamma : \left(\sigma\sigma^T\right)\left(x, \dfrac{1}{P}\sum_{j=1}^P \kappa_2(x, X_j^P(t))\right) \\ \quad - \dfrac{1}{4}\|\sigma^T\nabla\gamma\left(x, \dfrac{1}{P}\sum_{j=1}^P \kappa_2(x, X_j^P(t))\right)\|^2 = 0, (t,x) \in [0,T) \times \mathbb{R}^d \\ \gamma(T, x) = -\log|G(x)|, \quad x \in \mathbb{R}^d, \end{cases} \quad (12)$$

with optimal control

$$\zeta^*(t, x) = -\sigma^T\left(x, \frac{1}{P}\sum_{j=1}^P \kappa_2(x, X_j^P(t))\right)\nabla\gamma(t, x),$$
$$(13)$$

minimizing the second moment.

Proof See Appendix B in Ben Rached et al. (2024). □

Previous studies (Ben Rached et al. 2024; Awad et al. 2013) have demonstrated that (12) leads to a zero-variance estimator of the inner expectation $\mathbb{E}\left[G\left(\bar{X}^P(T)\right)\right]$ $| \left\{\mu_t^P : t \in [0, T]\right\}$, provided $G(\cdot)$ does not change sign. Using the transformation $u(t, x) = v^2(t, x)$, we can recover the linear Kolmogorov backward equation for the dynamics (8) for a given $\left\{\mu_t^P : t \in [0, T]\right\}$.

$$\begin{cases} \dfrac{\partial v}{\partial t} + \langle b\left(x, \dfrac{1}{P}\sum_{j=1}^P \kappa_1(x, X_j^P(t))\right), \nabla v\rangle \\ \quad + \dfrac{1}{2}\nabla^2 v : \left(\sigma\sigma^T\right)\left(x, \dfrac{1}{P}\sum_{j=1}^P \kappa_2(x, X_j^P(t))\right) = 0, (t,x) \in [0,T) \times \mathbb{R}^d \\ v(T, x) = |G(x)|, \quad x \in \mathbb{R}^d, \end{cases} \quad (14)$$

with optimal control

$$\zeta^*(t, x) = \sigma^T\left(x, \frac{1}{P}\sum_{j=1}^P \kappa_2(x, X_j^P(t))\right)\nabla\log v(t, x).$$
$$(15)$$

Remark 1 In Proposition 1, we control the decoupled McKean–Vlasov process $\bar{X}^P$ (whose dynamics are given in 8) instead of the particles X_p^P from the interacting particle system (3) because the optimal control problem would be d-dimensional instead of Pd-dimensional.

We require a realization of the empirical law from the stochastic P-particle system (3) to obtain the control using (14) and (15). In practice, we obtain a time-discretized version of the empirical law using the Euler–Maruyama scheme. To avoid computing the optimal control for each $\{\mu_t^P : t \in [0, T]\}$ realization in the DLMC estimator, we independently obtain a sufficiently accurate empirical law realization off-line using a sufficiently large number of particles and time steps (see (Ben Rached et al. 2024), Algorithm 2). This approach is motivated by the convergence of the empirical law to the MV-SDE law $\{\mu_t : t \in [0, T]\}$ as the number of particles and time steps tend to infinity. This approach has two main advantages: we do not need to solve the original KBE (14) for the value function u to satisfy relative tolerance TOL_r, and we don't require high accuracy for the solution to ζ^* for the purpose of importance sampling. Section 6 confirms that the deterministic control thus obtained is sufficient to ensure variance reduction in the proposed estimator.

Remark 2 As a proof of concept, we numerically solve the 1D ($d = 1$) control PDE arising from the Kuramoto model (5) using finite differences and extend the solution to the entire domain using linear interpolation. Using such a method to solve (14) in higher dimensions ($d \gg 1$) is computationally expensive due to the curse of dimensionality. In such cases, model reduction techniques (Hartmann et al. 2016, 2015) or solving the minimization problem (11) using stochastic gradient methods (Hartmann et al. 2017) may help. We do not focus on building efficient methods to solve the control PDE in this work.

4.2 Double loop Monte Carlo estimator with importance sampling

We briefly outline the single-level DLMC estimator for a given importance sampling control ζ obtained off-line by solving (15) (for more details, see (Ben Rached et al. 2024)).

1. We consider the discretization $0 = t_0 < t_1 < t_2 < \cdots < t_N = T$ of the time domain $[0, T]$ with N equal time

steps of the particle system (3) (i.e., $t_n = n\Delta t$, $n = 0, 1, \ldots, N$ and $\Delta t = T/N$). The discretized version of particle X_p^P corresponding to (3) with P particles is denoted by $X_p^{P|N}$. Let $\omega_{1:P}$ denote the P underlying sets of random variables used to generate a realization of $X_p^{P|N}$ from (3).

2. We define the discrete law obtained from the time-discretized particle system by $\mu^{P|N}$ as

$$\mu^{P|N}(t_n) = \frac{1}{P} \sum_{p=1}^{P} \delta_{X_p^{P|N}(t_n)}, \quad \forall n = 0, \ldots, N. \quad (16)$$

Then, we define a time-continuous extension of the empirical law by extending the time-discrete stochastic particle system to all $t \in [0, T]$ using the continuous-time forward Euler-Maruyama extension.

3. Given the approximate law $\mu^{P|N}$ from (16) and control $\zeta : [0, T] \times \mathbb{R}^d \to \mathbb{R}^d$, we generate sample paths $\{\bar{X}_\zeta^{P|N}(t) : t \in [0, T]\}$ of the controlled decoupled MV-SDE (10). Let $\bar{\omega}$ denote the set of random variables to generate one of the above sample paths.

4. We consider the same discretization of the time domain $[0, T]$ as the particle system (3) for the controlled decoupled MV-SDE (10) with N equal time steps. We define $\{\bar{X}_\zeta^{P|N}(t_n)\}_{n=1}^{N}$ as the Euler–Maruyama time-discretized version of $\bar{X}_\zeta^{P|N}$, the decoupled MV-SDE process (10) defined using empirical law $\mu^{P|N}$ (16).

5. Thus, we can express the quantity of interest with importance sampling as follows:

$$\mathbb{E}\left[G(\bar{X}^{P|N}(T))\right] = \mathbb{E}\left[G(\bar{X}_\zeta^{P|N}(T))\mathbb{L}^{P|N}\right]. \quad (17)$$

This expectation is approximated using the DLMC estimator $\mathcal{A}_{\mathrm{MC}}$ from Ben Rached et al. (2024).

$$\mathcal{A}_{\mathrm{MC}} = \frac{1}{M_1} \sum_{i=1}^{M_1} \frac{1}{M_2} \sum_{j=1}^{M_2} G\left(\bar{X}_\zeta^{P|N}(T)\right) \mathbb{L}^{P|N}\left(\omega_{1:P}^{(i)}, \bar{\omega}^{(j)}\right). \quad (18)$$

where the likelihood factor (for detailed derivation, see Ben Rached et al. (2024)) is

$$\mathbb{L}^{P|N}\left(\omega_{1:P}^{(i)}, \bar{\omega}^{(j)}\right) = \prod_{n=0}^{N-1} \exp\left\{ -\frac{1}{2}\Delta t \|\zeta(t_n, \bar{X}_\zeta^{P|N}(t_n))\|^2 \right.$$
$$\left. - \langle \Delta W(t_n), \zeta(t_n, \bar{X}_\zeta^{P|N}(t_n))\rangle \right\}\left(\omega_{1:P}^{(i)}, \bar{\omega}^{(j)}\right). \quad (19)$$

Here, M_1 is the number of realizations of $\mu^{P|N}$ in the DLMC estimator, and $\omega_{1:P}^{(i)}$ denotes the ith realization of $\omega_{1:P}$. For each realization of $\mu^{P|N}$, M_2 is

the number of sample paths for the decoupled MV-SDE, and $\bar{\omega}^{(j)}$ denotes the jth realization of $\bar{\omega}$. Further $\Delta W(t_n) \sim \mathcal{N}(0, \sqrt{\Delta t}\mathbb{I}_d)$ are the Wiener increments driving the dynamics of the time-discretized decoupled MV-SDE (10).

Remark 3 The MC estimator for smooth, nonrare observables based on the particle system approximation introduced by Haji-Ali and Tempone (2018) has a computational complexity of $\mathcal{O}(\mathrm{TOL}^{-4})$ for a given absolute error tolerance TOL. However, the constant associated with this complexity substantially increases in the rare-event regime (Kroese et al. 2013). This problem can be overcome using the above importance sampling scheme. Ben Rached et al. (2024) demonstrates that the DLMC estimator (18) with importance sampling has a complexity of $\mathcal{O}(\mathrm{TOL}_r^{-4})$ for estimating rare-event probabilities up to the prescribed relative error tolerance TOL_r. Additionally, importance sampling ensures that the constant associated with the complexity of this DLMC estimator (18) is significantly reduced, enabling a feasible computation of rare-event probabilities.

Section 5 extends this estimator to the multilevel setting to obtain better complexity.

5 Multilevel double loop Monte Carlo estimator with importance sampling

5.1 Multilevel double loop Monte Carlo estimator for decoupled McKean–Vlasov stochastic differential equation

Two discretization parameters (P and N) approximate the solution to the decoupled MV-SDE. For MLMC purposes, we introduce the parameter (level) ℓ that couples both discretization parameters. The geometric sequence of levels is defined given the parameter τ. For $\ell = 0, 1, \ldots, L$, let:

$$P_\ell = P_0\tau^\ell,$$
$$N_\ell = N_0\tau^\ell. \quad (20)$$

We set $G = G(X(T))$, and its corresponding discretization at level ℓ is $G_\ell = G(\bar{X}^{P_\ell|N_\ell}(T))$, where $X : [0, T] \times \Omega \to \mathbb{R}^d$ is the MV-SDE process (1) and $\bar{X}^{P_\ell|N_\ell}$ satisfies (8). The MLMC concept (Giles 2015) uses a telescoping sum for level $L \in \mathbb{N}$:

$$\mathbb{E}[G_L] = \sum_{\ell=0}^{L} \mathbb{E}[G_\ell - G_{\ell-1}] \quad, G_{-1} = 0. \quad (21)$$

The current work approximates each expectation in (21) independently using the DLMC method, creating the multi-level DLMC estimator:

$$\mathcal{A}_{\mathrm{MLMC}} = \sum_{\ell=0}^{L} \frac{1}{M_{1,\ell}} \sum_{i=1}^{M_{1,\ell}} \frac{1}{M_{2,\ell}} \sum_{j=1}^{M_{2,\ell}} (G_\ell - \mathcal{G}_{\ell-1}) \left(\omega_{1:P_\ell}^{(\ell,i)}, \bar{\omega}^{(\ell,j)} \right). \tag{22}$$

where $\mathcal{G}_{\ell-1}$ is a random variable correlated to G_ℓ such that $\mathcal{G}_{-1} = 0$ and $\mathbb{E}\left[\mathcal{G}_{\ell-1}\right] = \mathbb{E}\left[G_{\ell-1}\right]$, ensuring $\mathbb{E}\left[\mathcal{A}_{\mathrm{MLMC}}\right] = \mathbb{E}\left[G_L\right]$. In addition, $\omega_{1:P_\ell}^{(\ell,i)}$ refers to the ith realization of the P_ℓ underlying sets of 8random variables used to estimate $\mu^{P_\ell|N_\ell}$ at level ℓ, and $\bar{\omega}^{(\ell,j)}$ denotes the jth realization of the random variables in (8) for the given $\mu^{P_\ell|N_\ell}$ realization at level ℓ. Samples of $G_\ell(\omega_{1:P_\ell}^{(\ell,i)}, \bar{\omega}^{(\ell,j)})$ and $\mathcal{G}_{\ell-1}(\omega_{1:P_\ell}^{(\ell,i)}, \bar{\omega}^{(\ell,j)})$ must be sufficiently correlated to ensure that the multilevel DLMC estimator has better complexity than the single-level DLMC estimator. We explore two correlated samplers motivated by the MLMC estimators in Haji-Ali and Tempone (2018).

1. *Naïve Sampler.* We use the first $P_{\ell-1}$ random variables out of each set of P_ℓ random variables to obtain empirical law $\mu_{\ell-1} = \mu^{P_{\ell-1}|N_{\ell-1}}$ to generate a sample of $\mathcal{G}_{\ell-1}$ using (3). Given $\mu_{\ell-1}$, we solve (8) at level $\ell - 1$ using the same $\bar{\omega}$ as for level ℓ and compute the quantity of interest:

$$\begin{aligned}\mathcal{G}_{\ell-1}\left(\omega_{1:P_\ell}^{(\ell,i)}, \bar{\omega}^{(\ell,j)}\right) &= \bar{\mathcal{G}}_{\ell-1}\left(\omega_{1:P_\ell}^{(\ell,i)}, \bar{\omega}^{(\ell,j)}\right) \\ &\stackrel{\mathrm{def}}{=} G_{\ell-1}\left(\omega_{1:P_{\ell-1}}^{(\ell,i)}, \bar{\omega}^{(\ell,j)}\right). \end{aligned} \tag{23}$$

2. *Antithetic Sampler.* We split the P_ℓ sets of random variables into τ i.i.d. groups of size $P_{\ell-1}$ each. Then, to generate a sample of $\mathcal{G}_{\ell-1}$, we use each group to independently simulate (3) and obtain empirical law $\mu_{\ell-1}^{(a)}$ for $a = 1, \ldots, \tau$. Given $\mu_{\ell-1}^{(a)}$, we solve (8) at level $\ell - 1$ independently for each a using the same $\bar{\omega}$ as for level ℓ. The quantity of interest is computed for each group and averaged over the τ groups:

$$\begin{aligned}\mathcal{G}_{\ell-1}(\omega_{1:P_\ell}^{(\ell,i)}, \bar{\omega}^{(\ell,j)}) &= \hat{\mathcal{G}}_{\ell-1}\left(\omega_{1:P_\ell}^{(\ell,i)}, \bar{\omega}^{(\ell,j)}\right) \\ &\stackrel{\mathrm{def}}{=} \frac{1}{\tau} \sum_{a=1}^{\tau} G_{\ell-1}\left(\omega_{(a-1)P_{\ell-1}+1:aP_{\ell-1}}^{(\ell,i)}, \bar{\omega}^{(\ell,j)}\right). \end{aligned} \tag{24}$$

Section 6 numerically investigates the effects of these two correlation schemes on the variance convergence for level differences.

5.1.1 Error analysis

We aim to build an efficient multilevel DLMC estimator that satisfies (7). We can bound the global relative error of $\mathcal{A}_{\mathrm{MLMC}}$

as follows:

$$\frac{|\mathbb{E}\left[G\right] - \mathcal{A}_{\mathrm{MLMC}}|}{|\mathbb{E}\left[G\right]|} \leq \underbrace{\frac{|\mathbb{E}\left[G\right] - \mathbb{E}\left[G_L\right]|}{|\mathbb{E}\left[G\right]|}}_{=\epsilon_b, \ \mathrm{Relative \ bias}} + \underbrace{\frac{|\mathbb{E}\left[G_L\right] - \mathcal{A}_{\mathrm{MLMC}}|}{|\mathbb{E}\left[G\right]|}}_{=\epsilon_s}, \tag{25}$$

We impose more restrictive error constraints than (7),

$$\epsilon_b \leq \theta \mathrm{TOL_r} |\mathbb{E}\left[G\right]|, \tag{26}$$

$$\mathbb{P}[\epsilon_s \geq (1-\theta)\mathrm{TOL_r}|\mathbb{E}\left[G\right]|] \leq \nu, \tag{27}$$

for a given tolerance splitting parameter $\theta \in (0, 1)$. We make the following assumption for the bias.

Assumption 1 *(Multilevel DLMC Bias)* There exists a constant $\tilde{\alpha} > 0$ such that

$$\epsilon_b = \frac{|\mathbb{E}\left[G\right] - \mathbb{E}\left[G_\ell\right]|}{|\mathbb{E}\left[G\right]|} \lesssim \tau^{-\tilde{\alpha}\ell},$$

where $\lesssim$ indicates that a constant C exists independent of ℓ, such that $\epsilon_b \leq C\tau^{-\tilde{\alpha}\ell}$, and constant $\tilde{\alpha}$ denotes the bias convergence rate with respect to level ℓ. Section 6 verifies this assumption and determines $\tilde{\alpha}$ numerically for the Kuramoto model. Assumption 1 is motivated by the weak convergence with respect to the number of particles (Kolokoltsov and Troeva 2019) and with respect to the number of time steps (Kloeden and Platen 1992). In order to use these bounds in this work, we assume the drift/diffusion coefficients satisfy further regularity and boundedness conditions entailed in Kolokoltsov and Troeva (2019), Kloeden and Platen (1992). We set the level L to satisfy the bias constraint (26). The statistical error constraint (27) can be rewritten as follows:

$$\begin{aligned}&\mathbb{P}[\epsilon_s \geq (1-\theta)\mathrm{TOL_r}|\mathbb{E}\left[G\right]|] \\ &= \mathbb{P}\left[\frac{\epsilon_s}{\sqrt{\mathrm{Var}\left[\mathcal{A}_{\mathrm{MLMC}}\right]}} \geq \frac{(1-\theta)\mathrm{TOL_r}|\mathbb{E}\left[G\right]|}{\sqrt{\mathrm{Var}\left[\mathcal{A}_{\mathrm{MLMC}}\right]}}\right] \leq \nu. \end{aligned} \tag{28}$$

By assuming the normality (at least asymptotically) of the estimator $\mathcal{A}_{\mathrm{MLMC}}$, we obtain the following condition for the estimator variance:

$$\mathrm{Var}\left[\mathcal{A}_{\mathrm{MLMC}}\right] \leq \left(\frac{(1-\theta)\mathrm{TOL_r}|\mathbb{E}\left[G\right]|}{C_\nu}\right)^2, \tag{29}$$

where C_ν is the $\left(1 - \frac{\nu}{2}\right)$-quantile for the standard normal distribution. The asymptotic normality of MLMC estimators can be demonstrated using the Lindeberg–Feller central limit

theorem (Ash and Doléans-Dade 2000). The proposed estimator variance can be expressed as

$$\text{Var}\left[\mathcal{A}_{\text{MLMC}}\right] = \sum_{\ell=0}^{L} \frac{1}{M_{1,\ell}} \text{Var}\left[\frac{1}{M_{2,\ell}} \sum_{j=1}^{M_{2,\ell}} (\Delta G_\ell)^{(1,j)}\right], \tag{30}$$

where $(\Delta G_\ell)^{(i,j)} = (G_\ell - \mathcal{G}_{\ell-1})\left(\omega_{1:P_\ell}^{(\ell,i)}, \bar{\omega}^{(\ell,j)}\right)$. Using the law of total variance,

$$\text{Var}[\mathcal{A}_{\text{MLMC}}] = \sum_{\ell=0}^{L} \frac{1}{M_{1,\ell}} \left(\underbrace{\text{Var}[\mathbb{E}[\Delta G_\ell \mid \{\mu_\ell, \mu_{\ell-1}\}]]}_{=V_{1,\ell}}\right.$$

$$\left. + \frac{1}{M_{2,\ell}} \underbrace{\mathbb{E}[\text{Var}[\Delta G_\ell \mid \{\mu_\ell, \mu_{\ell-1}\}]]}_{V_{2,\ell}}\right)$$

$$= \sum_{\ell=0}^{L} \left(\frac{V_{1,\ell}}{M_{1,\ell}} + \frac{V_{2,\ell}}{M_{1,\ell} M_{2,\ell}}\right), \tag{31}$$

where laws $\{\mu_\ell, \mu_{\ell-1}\}$ are coupled by the same sets of random variables $\omega_{1:P_\ell}^{(\ell,\cdot)}$ as described in the naïve (23) and antithetic (24) samplers. We make the following assumptions on $V_{1,\ell}$ and $V_{2,\ell}$.

Assumption 2 *(Multilevel DLMC variance)* There exist constants $\tilde{w} > 0$ and $\tilde{s} > 0$ such that,

$$V_{1,\ell} = \text{Var}[\mathbb{E}[\Delta G_\ell \mid \{\mu_\ell, \mu_{\ell-1}\}]] \lesssim \tau^{-\tilde{w}\ell}, \tag{32}$$

$$V_{2,\ell} = \mathbb{E}[\text{Var}[\Delta G_\ell \mid \{\mu_\ell, \mu_{\ell-1}\}]] \lesssim \tau^{-\tilde{s}\ell}. \tag{33}$$

where constants $\tilde{w}$ and $\tilde{s}$ are the convergence rates for $V_{1,\ell}$ and $V_{2,\ell}$, respectively, with respect to ℓ. Section 6 numerically determines these constants for the Kuramoto model.

5.1.2 Complexity analysis

We can express the total computational cost of the proposed multilevel DLMC estimator using the cost of the DLMC estimator (18),

$$\mathcal{W}[\mathcal{A}_{\text{MLMC}}] \lesssim \sum_{\ell=0}^{L} \left(M_{1,\ell} P_\ell^{1+\gamma_p} N_\ell^{\gamma_n} + M_{1,\ell} M_{2,\ell} P_\ell^{\gamma_p} N_\ell^{\gamma_n}\right). \tag{34}$$

$\gamma_p > 0$ and $\gamma_n > 0$ are the computational complexity rates of the empirical measure computation in the drift and diffusion coefficients and the time discretization scheme, respectively (see Appendix A). Let some level L satisfy the bias constraint (26). We aim to compute the optimal parameters

$\{M_{1,\ell}, M_{2,\ell}\}_{\ell=0}^{L}$ satisfying (29), which can be posed as the following optimization problem:

$$\begin{cases} \arg\min_{\{M_{1,\ell}, M_{2,\ell}\}_{\ell=0}^{L}} \sum_{\ell=0}^{L} \left(M_{1,\ell} P_\ell^{1+\gamma_p} N_\ell^{\gamma_n} + M_{1,\ell} M_{2,\ell} P_\ell^{\gamma_p} N_\ell^{\gamma_n}\right) \\ \text{s.t. } C_\nu^2 \left(\sum_{\ell=0}^{L} \left(\frac{V_{1,\ell}}{M_{1,\ell}} + \frac{V_{2,\ell}}{M_{1,\ell} M_{2,\ell}}\right)\right) \approx (1-\theta)^2 \text{TOL}_r^2 |\mathbb{E}[G]|^2 \end{cases} \tag{35}$$

The arguments in (35) in $\mathbb{R}^+$ that minimize the objective is found using the Lagrangian multiplier method $\forall \ell = 0, \ldots, L$.

$$\mathcal{M}_{1,\ell} = \frac{C_\nu^2}{(1-\theta)^2 \text{TOL}_r^2 |\mathbb{E}[G]|^2} \times$$

$$\frac{\sqrt{V_{1,\ell}}}{\sqrt{P_\ell^{1+\gamma_p} N_\ell^{\gamma_n}}} \left(\sum_{j=0}^{L} \sqrt{P_j^{\gamma_p} N_j^{\gamma_n}} (\sqrt{V_{1,j} P_j} + \sqrt{V_{2,j}})\right),$$

$$\tilde{\mathcal{M}}_\ell = \mathcal{M}_{1,\ell} \mathcal{M}_{2,\ell} = \frac{C_\nu^2}{(1-\theta)^2 \text{TOL}_r^2 |\mathbb{E}[G]|^2} \frac{\sqrt{V_{2,\ell}}}{\sqrt{P_\ell^{\gamma_p} N_\ell^{\gamma_n}}} \times$$

$$\left(\sum_{j=0}^{L} \sqrt{P_j^{\gamma_p} N_j^{\gamma_n}} (\sqrt{V_{1,j} P_j} + \sqrt{V_{2,j}})\right). \tag{36}$$

In practice, we only use natural numbers for $\{M_{1,\ell}, M_{2,\ell}\}_{\ell=0}^{L}$. Hence, we use the following quasi-optimal solution to (35):

$$M_{1,\ell} = \lceil \mathcal{M}_{1,\ell} \rceil, \quad M_{2,\ell} = \left\lceil \frac{\tilde{\mathcal{M}}_\ell}{\lceil \mathcal{M}_{1,\ell} \rceil} \right\rceil. \tag{37}$$

Note that $M_{1,\ell}$ required to satisfy TOL_r scales with the factor $\frac{1}{|\mathbb{E}[G]|^2}$, implying that DLMC without importance sampling can quickly become computationally expensive for rare events. We obtain the optimal computational cost for the proposed multilevel DLMC estimator using (37).

Theorem 1 *(Optimal multilevel DLMC complexity) Let G_ℓ be an approximation for the random variable G, for every $\ell \in \mathbb{N}$, and $G_\ell^{(i,j)} \equiv G_\ell(\omega_{1:P_\ell}^{(\ell,i)}, \bar{\omega}^{(\ell,j)})$ be a sample of G_ℓ. Consider the multilevel DLMC estimator (22) with $G_{-1}^{(i,j)} = 0$. Let Assumptions 1 and 2 hold. Let the constants τ, $\tilde{\alpha}$, $\tilde{w}$, $\tilde{s}$, γ_p, and $\gamma_n > 0$ from Assumptions 1 and 2 be such that $\tilde{\alpha} \geq \frac{1}{2} \min(\tilde{w}, 1+\tilde{s}, 1+\gamma_p+\gamma_n)$.*

Then, for any $\text{TOL}_r < 1/e$, there exists an optimal L and sequences $\{M_{1,\ell}\}_{\ell=0}^{L}$ and $\{M_{2,\ell}\}_{\ell=0}^{L}$ such that

$$\mathbb{P}\left[\frac{|\mathcal{A}_{MLMC}(L) - \mathbb{E}[G]|}{|\mathbb{E}[G]|} \geq \text{TOL}_r\right] \leq \nu, \tag{38}$$

and

$$\mathcal{W}[\mathcal{A}_{MLMC}] \lesssim TOL_r^{-2-\max\left(0, \frac{1+\gamma_n+\gamma_p-\tilde{w}}{\tilde{\alpha}}, \frac{\gamma_n+\gamma_p-\tilde{s}}{\tilde{\alpha}}\right)} \left(\log TOL_r^{-1}\right)^2 \mathcal{J}, \tag{39}$$

where

$$\mathcal{J} = \begin{cases} 1, & \text{if } \min(\tilde{w}, 1+\tilde{s}) = 1 + \gamma_p + \gamma_n, \\ 0, & \text{else.} \end{cases}$$

Proof See Appendix B. □

Remark 4 The complexity rate in (39) is independent of θ and τ or the dimension, d, of the MV-SDE. These only affect the associated constant in (39). The optimization of the parameters θ and τ in the context of general MLMC methods has been investigated (Haji-Ali et al. 2016b; Collier et al. 2015).

Remark 5 In many applications, the variance and second moments of the level differences in the multilevel DLMC estimator are of the same order. In this case, we can demonstrate that $V_{1,\ell} \leq \mathbb{E}\left[\mathbb{E}\left[\Delta G_\ell^2 \mid \{\mu_\ell, \mu_{\ell-1}\}\right]\right] \approx V_{2,\ell}$, implying $\tilde{w} \geq \tilde{s}$.

Remark 6 In the non-rare event context, TOL_r in (38) and (39) is easily replaced by $\frac{TOL}{|\mathbb{E}[G]|}$, where TOL is the required absolute error tolerance.

Remark 7 *(Kuramoto model)* $\gamma_p = 1$ corresponds to a naïve empirical mean estimation method and $\gamma_n = 1$ corresponds to the Euler–Maruyama scheme with a uniform time grid with time step $\frac{T}{N_\ell}$. Then, $\tilde{\alpha} = 1$ due to the standard rates of weak convergence with respect to P_ℓ (Kolokoltsov and Troeva 2019; Ben Rached et al. 2024) and N_ℓ using the Euler–Maruyama scheme (Kloeden and Platen 1992). We obtain better complexity than $\mathcal{O}(TOL_r^{-4})$ for the single-level DLMC estimator when $\tilde{s}+1 \geq \tilde{w}$ and $\tilde{w} > 1$. For this example with $G(x) = \cos(x)$, using the antithetic sampler ensures $\{\tilde{w}, \tilde{s}\} = \{2, 2\}$, leading to $\mathcal{O}(TOL_r^{-3})$ complexity. In contrast, the naïve sampler results in $\{\tilde{w}, \tilde{s}\} = \{1, 1\}$, leading to a complexity of $\mathcal{O}(TOL_r^{-4})$ (i.e., the same as the single-level DLMC estimator). Section 6 presents these outcomes numerically.

Section 5.2 devises an importance sampling scheme for the proposed multilevel DLMC estimator to address rare events associated with MV-SDEs.

5.2 Importance sampling scheme for the multilevel DLMC estimator for the decoupled MV-SDE

We propose the following method to couple importance sampling with the proposed multilevel DLMC estimator. We obtain one importance sampling control ζ off-line by solving the control PDE (14) derived in Sect. 4 using one realization of the stochastic particle system with a large number of particles $\bar{P}$ and time steps $\bar{N}$. We apply the same control across all levels $\ell = 0, \dots, L$ in the proposed multilevel DLMC estimator (22). Thus, we can rewrite the quantity of interest as

$$\mathbb{E}[G_L] = \sum_{\ell=0}^{L} \mathbb{E}[G_\ell - G_{\ell-1}]$$

$$= \sum_{\ell=0}^{L} \mathbb{E}\left[G_\ell^\zeta \mathbb{L}_\ell - G_{\ell-1}^\zeta \mathbb{L}_{\ell-1}\right], \tag{40}$$

where

$$G_\ell^\zeta = G(\bar{X}_\zeta^{P_\ell|N_\ell}(T)), \tag{41}$$

$$\mathbb{L}_\ell = \prod_{n=0}^{N_\ell-1} \exp\left\{-\frac{1}{2}\Delta t_\ell \|\zeta\left(t_{n,\ell}, \bar{X}_\zeta^{P_\ell|N_\ell}(t_{n,\ell})\right)\|^2 \right.$$
$$\left. - \langle \Delta W(t_{n,\ell}), \zeta\left(t_{n,\ell}, \bar{X}_\zeta^{P_\ell|N_\ell}(t_{n,\ell})\right)\rangle\right\}; \tag{42}$$

$\left\{\bar{X}_\zeta^{P_\ell|N_\ell}(t_{n,\ell})\right\}_{n=0}^{N_\ell}$ is the time-discretized controlled, decoupled MV-SDE sample path at level ℓ (see Sect. 4.2). $\mathbb{L}_\ell$ is the likelihood factor at level ℓ, and $\Delta t_\ell = \frac{T}{N_\ell}$ is the uniform time step of the Euler–Maruyama scheme for the decoupled MV-SDE at level ℓ. $\{\Delta W(t_{n,\ell})\}_{n=0}^{N_\ell} \sim \mathcal{N}(0, \sqrt{\Delta t_\ell}\mathbb{I}_d)$ are Wiener increments driving the dynamics of coarse and fine time-discretized paths of the decoupled MV-SDE at level ℓ. We define the proposed multilevel DLMC estimator with importance sampling as follows:

$$\mathbb{E}[G_L] \approx \mathcal{A}_{MLMC}^{IS}$$
$$= \sum_{\ell=0}^{L} \frac{1}{M_{1,\ell}} \sum_{i=1}^{M_{1,\ell}} \frac{1}{M_{2,\ell}} \sum_{j=1}^{M_{2,\ell}} \left(G_\ell^{IS} - \mathcal{G}_{\ell-1}^{IS}\right)\left(\omega_{1:P_\ell}^{(\ell,i)}, \bar{\omega}^{(\ell,j)}\right), \tag{43}$$

where $G_\ell^{IS}\left(\omega_{1:P_\ell}^{(\ell,i)}, \bar{\omega}^{(\ell,j)}\right) = G_\ell^\zeta \mathbb{L}_\ell\left(\omega_{1:P_\ell}^{(\ell,i)}, \bar{\omega}^{(\ell,j)}\right)$.

Remark 8 The complexity of the multilevel DLMC estimator with the above importance sampling scheme remains the same as in (39) because the optimal control problem (14) is solved once and we do not include its cost in the complexity.

Remark 9 This is a natural extension to the importance sampling scheme previously developed for the single-level estimator (Ben Rached et al. 2024). The optimal control ζ minimizes the single-level estimator variance of the conditional expectation $\mathbb{E}[G_\ell \mid \mu_\ell]$; thus, we expect a variance reduction for the level differences estimator. However, the optimal control ζ derived in Sect. 4.1 minimizes $\text{Var}[G_\ell]$ and not the variance of the MC estimator of $\mathbb{E}[\Delta G_\ell]$. Optimally,

we must determine a control that minimizes $\mathrm{Var}\left[\Delta G_\ell\right]$ at each level ℓ of the multilevel DLMC estimator, which we leave for future work.

Algorithm 4 in Appendix C implements the proposed importance sampling scheme in the level-difference estimator for the proposed multilevel DLMC method and can be easily modified for any other correlated sampler, such as the naïve sampler. Next, we build an adaptive multilevel DLMC algorithm that sequentially chooses parameters L, $\{M_{1,\ell}, M_{2,\ell}\}_{\ell=0}^{L}$ satisfying constraints (26) and (27). The bias and variances $V_{1,\ell}$ and $V_{2,\ell}$ corresponding to level ℓ must be estimated cheaply and robustly to develop such an algorithm.

5.3 Adaptive multilevel double loop Monte Carlo algorithm with importance sampling

5.3.1 Estimating Bias at level ℓ

The bias for level ℓ can be approximated using Richardson extrapolation (Lemaire and Pagès 2017):

$$|\mathbb{E}\left[G\right] - \mathbb{E}\left[G_\ell\right]| \approx \left(1 - \tau^{-\tilde{\alpha}}\right)^{-1} |\mathbb{E}\left[G_{\ell+1} - G_\ell\right]|. \quad (44)$$

Then, we use Algorithm 4 with at least $\underline{M}_1$ and $\underline{M}_2$ samples to obtain a DLMC estimation of the bias. To ensure robust bias estimates at all levels, we actually use $\hat{M}_1$ and $\hat{M}_2$ samples in Algorithm 4 defined as follows:

$$\hat{M}_1 = \max(M_{1,\ell}, \underline{M}_1),$$
$$\hat{M}_2 = \max(M_{2,\ell}, \underline{M}_2). \quad (45)$$

To ensure reliable bias estimates at higher levels ($\ell > 3$), we compare the DLMC bias estimator with the extrapolated bias from two previous levels using Assumption 1:

$$|\mathbb{E}\left[G\right] - \mathbb{E}\left[G_\ell\right]| \approx \max\left(\frac{|\mathbb{E}\left[\Delta G_{\ell+1}\right]|}{1 - \tau^{-\tilde{\alpha}}}, \frac{|\mathbb{E}\left[G\right] - \mathbb{E}\left[G_{\ell-1}\right]|}{\tau^{\tilde{\alpha}}}, \frac{|\mathbb{E}\left[G\right] - \mathbb{E}\left[G_{\ell-2}\right]|}{\tau^{2\tilde{\alpha}}}\right), \quad \ell > 2. \quad (46)$$

5.3.2 Estimating $V_{1,\ell}$, $V_{2,\ell}$ at Level ℓ

To compute the optimal number of samples required to satisfy the statistical error constraint (29) using (37), we require cheap and robust empirical estimates of the variance terms $V_{1,\ell}$ and $V_{2,\ell}$ for each level ℓ. For this, we apply the DLMC algorithm (Algorithm 5 in Appendix D) with appropriately chosen values of $\tilde{M}_1$ and $\tilde{M}_2$. Algorithm 5 could become computationally expensive at higher levels.

We exploit Assumption 2 to avoid this overload and extrapolate variances for higher levels. For levels $\ell > 3$,

$$V_{1,\ell} = \max\left(\frac{V_{1,\ell-1}}{\tau^{\tilde{w}}}, \frac{V_{1,\ell-2}}{\tau^{2\tilde{w}}}\right),$$
$$V_{2,\ell} = \max\left(\frac{V_{2,\ell-1}}{\tau^{\tilde{s}}}, \frac{V_{2,\ell-2}}{\tau^{2\tilde{s}}}\right). \quad (47)$$

5.3.3 Relative error control

To control the relative bias and statistical errors, we require a heuristic estimate of the quantity of interest $\mathbb{E}\left[G\right]$ itself. This estimate is continuously updated at each level L. At level $L = 0$, we use the DLMC algorithm (Algorithm 4) with $\bar{M}_1$ and $\bar{M}_2$ samples to obtain an initial robust but cheap estimate of $\mathbb{E}\left[G\right]$. For subsequent levels, we apply the multilevel DLMC estimator (22) with optimal values for $\{M_{1,\ell}, M_{2,\ell}\}_{\ell=0}^{L}$ to update the estimate. Putting all this together, we propose the adaptive multilevel DLMC algorithm (Algorithm 2) for rare-event observables in the MV-SDE context. The IS control ζ in Algorithm 2 is obtained off-line by generating one realization of the empirical law $\mu^{\bar{P}|\bar{N}}$ with large $\bar{P}$ and $\bar{N}$ and then numerically solving control equation (14) given $\mu^{\bar{P}|\bar{N}}$.

Algorithm 2: Adaptive multilevel DLMC algorithm with importance sampling

Input: P_0, N_0, $\mathrm{TOL_r}$, $\zeta(\cdot, \cdot)$, $\{\bar{M}_1, \bar{M}_2\}$, $\{\tilde{M}_1, \tilde{M}_2\}$, $\{\underline{M}_1, \underline{M}_2\}$;
Estimate $\bar{G} = \mathbb{E}\left[G_0\right]$ with P_0, N_0, $\bar{M}_1$, $\bar{M}_2$, $\zeta(\cdot, \cdot)$ using **Algorithm 4**;
Estimate and store $V_{1,0}, V_{2,0}$ with P_0, N_0, $\bar{M}_1$, $\bar{M}_2$, $\zeta(\cdot, \cdot)$ using **Algorithm 5**;
Set $L = 1$;
while *Bias* $> \theta\mathrm{TOL_r}\bar{G}$ **do**
    $P_L = P_0 2^L$, $N_L = N_0 2^L$;
    Estimate and store $V_{1,L}$, $V_{2,L}$ with P_L, N_L, $\tilde{M}_1$, $\tilde{M}_2$, $\zeta(\cdot, \cdot)$ using **Algorithm 5**;
    Compute optimal $\{M_{1,\ell}, M_{2,\ell}\}_{\ell=0}^{L}$ using (37);
    Estimate bias using (44) with P_{L+1}, N_{L+1}, $\hat{M}_1$, $\hat{M}_2$, $\zeta(\cdot, \cdot)$ using (45) and **Algorithm 4**;
    Reevaluate $\bar{G} = \mathbb{E}\left[G_0\right] + \sum_{\ell=1}^{L} \mathbb{E}\left[\Delta G_\ell\right]$ with $\{M_{1,\ell}, M_{2,\ell}\}_{\ell=0}^{L}$, $\zeta(\cdot, \cdot)$ using **Algorithm 4** for each ℓ;
    $L \longleftarrow L + 1$;
end
$\mathcal{A}_{\mathrm{MLMC}} = \bar{G}$.

---

6 Numerical results

This section provides numerical evidence for the assumptions and rates of computational complexity derived in Sect. 5. The results outlined below focus on the Kuramoto model (see Sect. 2.1) with the following settings: $\sigma = 0.4$, $T = 1$, $(x_0)_p \sim \mathcal{N}(0, 0.2)$, and $\xi_p \sim \mathcal{U}(-0.2, 0.2)$ for all $p =$

1, . . . , P. We set the parameters as follows: $\tau = 2$, $\theta = 0.5$, and $\nu = 0.05$. The complexity rates ($\gamma_p = 1$ and $\gamma_n = 1$) are explained in Remark 7. Moreover, we set the hierarchies for the multilevel DLMC estimator as

$$P_\ell = 5 \times 2^\ell, \quad N_\ell = 4 \times 2^\ell. \tag{48}$$

We implement the proposed multilevel DLMC method for nonrare and rare-event observables and investigate the computational complexity compared with the single-level DLMC estimator.

6.1 Objective function $G(x) = \cos(x)$

First, we numerically verify Assumptions 1 and 2 for the smooth, nonrare observable $G(x) = \cos(x)$ without importance sampling to determine constants $\tilde{\alpha}, \tilde{w}, \tilde{s}$ for the Kuramoto model. Figure 1 presents the estimated bias (44) using Algorithm 4, and $V_{1,\ell}$ and $V_{2,\ell}$ with respect to ℓ using Algorithm 5. Thus, Assumptions 1 and 2 are verified with $\tilde{\alpha} = 1$ and $\{\tilde{w}, \tilde{s}\} = \{1, 1\}$ for the naïve sampler (23) and with $\{\tilde{w}, \tilde{s}\} = \{2, 2\}$ for the antithetic sampler (24). Improved variance convergence rates for the antithetic sampler imply a complexity of $\mathcal{O}(\text{TOL}^{-3})$ for the proposed multilevel DLMC estimator, compared with $\mathcal{O}(\text{TOL}^{-4})$ for the naïve sampler (see Theorem 1) to achieve a prescribed absolute error tolerance TOL. Thus, we use the antithetic sampler in the proposed adaptive algorithm (Algorithm 2).

Next, we implement the proposed multilevel DLMC algorithm (Algorithm 2) with inputs $\{\bar{M}_1, \bar{M}_2\} = \{1000, 100\}$, $\{\tilde{M}_1, \tilde{M}_2\} = \{25, 1000\}$, and $\{\underline{M}_1, \underline{M}_2\} = \{100, 50\}$ for this observable with tuned parameters. In this case, importance sampling is not required because this is not a rare-event observable, i.e. we set $\zeta(t, x) = 0, \forall (t, x) \in [0, T] \times \mathbb{R}^d$. Figure 2 depicts the results of Algorithm 2 in this setting to numerically verify the complexity rates obtained from Theorem 1. Figure 2a illustrates the exact multilevel DLMC estimator error, estimated using a reference multilevel DLMC approximation with TOL $= 10^{-4}$, for separate runs of Algorithm 2 for different prescribed absolute error tolerances TOL. The adaptive multilevel DLMC algorithm successfully satisfies the error constraint (38) at the 95% confidence level (corresponding to $\nu = 0.05$). Figure 2b presents the number of levels L required to satisfy the bias constraint in (26) for each of the separate runs of Algorithm 2 for different prescribed error tolerances TOL. According to Theorem 1, the number of levels should increase by $\mathcal{O}(\log(\text{TOL}^{-1}))$. Figure 2c displays the average computational runtime for Algorithm 2 for various error tolerances. The runtimes in Fig. 2c include the cost of estimating the bias, $V_{1,\ell}$ and $V_{2,\ell}$, in Algorithm 2. The runtimes for sufficiently small tolerances follow the predicted $\mathcal{O}(\text{TOL}^{-3})$ rate from Theorem 1. Figure 2d indicates the average estimated computational cost of the multilevel DLMC estimator for various TOL values. This estimated computational cost is computed using (34):

Computational cost[$\mathcal{A}_{\text{MLMC}}$]

$$\approx \sum_{\ell=0}^{L} \left(M_{1,\ell} P_\ell^{1+\gamma_p} N_\ell^{\gamma_n} + M_{1,\ell} M_{2,\ell} P_\ell^{\gamma_p} N_\ell^{\gamma_n} \right). \tag{49}$$

Figure 2c and d verify that the proposed multilevel DLMC estimator with the antithetic sampler outperforms the single-level DLMC estimator, achieving one order of complexity reduction from $\mathcal{O}(\text{TOL}^{-4})$ to $\mathcal{O}(\text{TOL}^{-3})$.

6.2 Rare-event objective function

To test the importance sampling scheme, we consider the Kuramoto model with the following Lipschitz rare-event observable $G(x) = \Psi(x - K)$ for a sufficiently large threshold $K \in \mathbb{R}$, where

$$\Psi(x) = \begin{cases} 0 & , \quad x < -0.5 \\ 0.5 + x & , \quad -0.5 < x < 0.5 \\ 1 & , \quad x > 0.5 \end{cases} \tag{50}$$

We use the importance sampling scheme introduced in Sect. 5.2 with importance sampling control ζ obtained by solving (14) numerically using finite differences and linear interpolation throughout the domain. First, we verify the variance reduction in the level-difference estimators using this ζ in two numerical experiments, whose results are depicted in Fig. 3.

Note that ζ is the optimal importance sampling control for the decoupled MV-SDE (8) and not the particle system (3). With this scheme, we reduce the variance of the inner expectation in (9). Consequently, we assess the variance reduction in the MC estimator of the inner expectation in the first experiment. Experiment 1 investigates variance reduction on the DLMC estimator of $\mathbb{E}\left[\Delta G_\ell \mid \mu^{\bar{P}|\bar{N}} \right]$. We run DLMC Algorithm 4 with $M_1 = 1$ and $\ell = 3$ for different values of M_2. To generate Fig. 3a, we obtained the importance sampling control ζ using one realization of the empirical law $\mu^{\bar{P}|\bar{N}}$ with $\bar{P} = 200$ particles and $\bar{N} = 100$ time steps. Figure 3a compares squared coefficients of variation for the DLMC estimator with and without importance sampling versus the number of sample paths M. The results verify that the importance sampling scheme reduces the squared coefficient of variation approximately 100-fold. In Experiment 2, we verify the variance reduction for the DLMC estimator of $\mathbb{E}[\Delta G_\ell]$ with importance sampling using DLMC Algorithm 4 with $\ell = 3$, $M_1 = 10^3$ for different values of M_2. To generate Fig. 3b, we obtained importance sampling
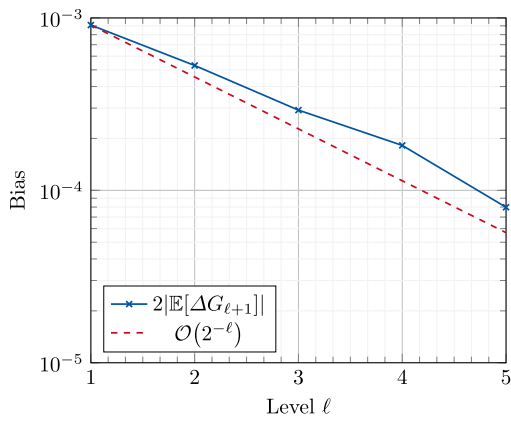
(a) Assumption 1: Double loop Monte Carlo bias estimator using Algorithm 4 with inputs $M_1 = 10^3$ and $M_2 = 10^3$ with respect to level ℓ.
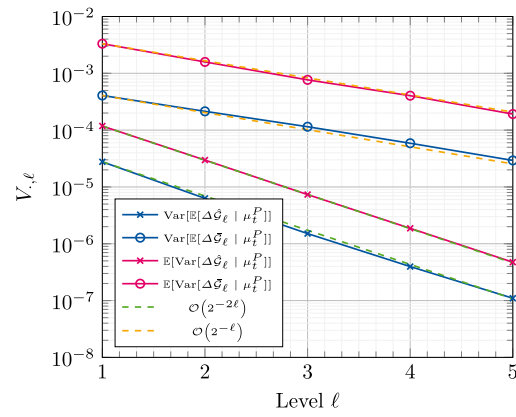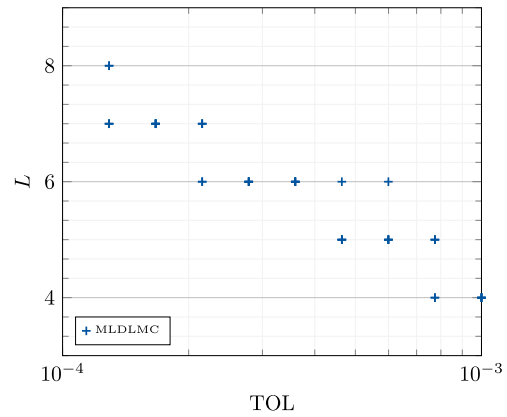
(b) Assumption 2: Double loop Monte Carlo estimates for $V_{1,\ell}$ and $V_{2,\ell}$ using Algorithm 5 with inputs $M_1 = 10^2$ and $M_2 = 10^4$ with respect to level ℓ.
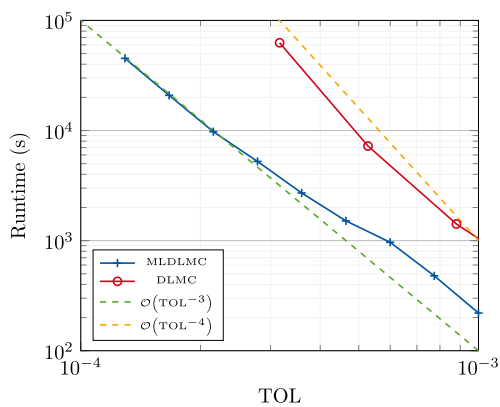
Fig. 1 Convergence rates of level differences using antithetic ($\hat{\mathcal{G}}$) and naïve ($\bar{\mathcal{G}}$) samplers for the Kuramoto model with $G(x) = \cos x$
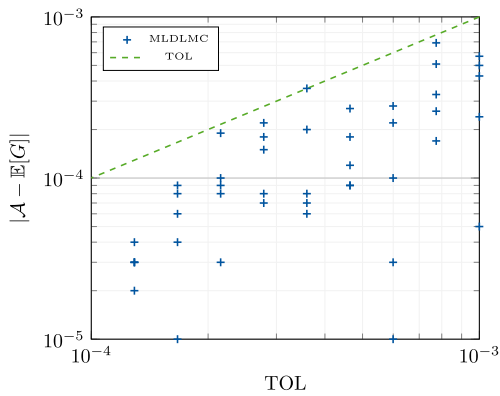
(a) Global estimator error with respect to tolerance TOL

(b) Estimator level L required to satisfy the tolerance TOL.

(c) Average computational runtime with respect to the tolerance TOL.
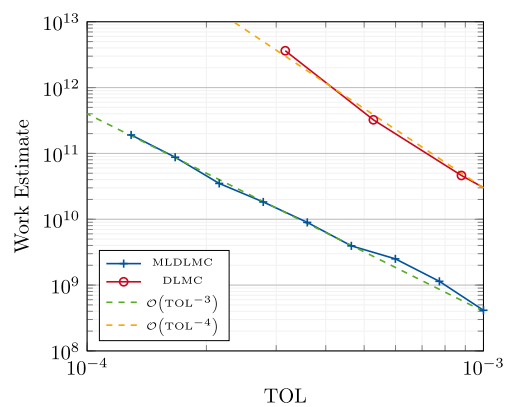
(d) Average computational cost estimate with respect to the tolerance TOL

Fig. 2 Algorithm 2 applied to the Kuramoto model for $G(x) = \cos x$. (MLDLMC: multilevel double loop Monte Carlo)
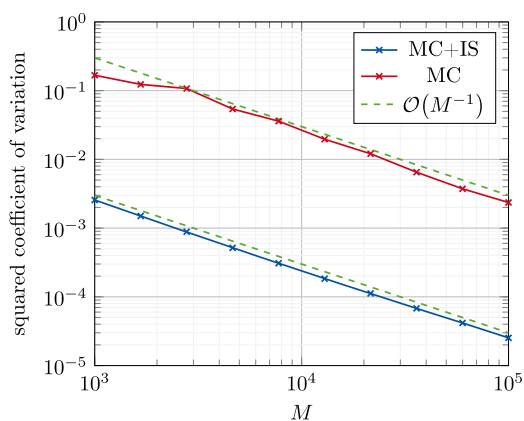
(a) Experiment 1: Squared coefficient of variation of the double loop Monte Carlo estimator for $\mathbb{E}\left[\Delta G_3 \mid \mu^{\bar{P}|\bar{N}}\right]$ with and without importance sampling with respect to the number of sample paths M.
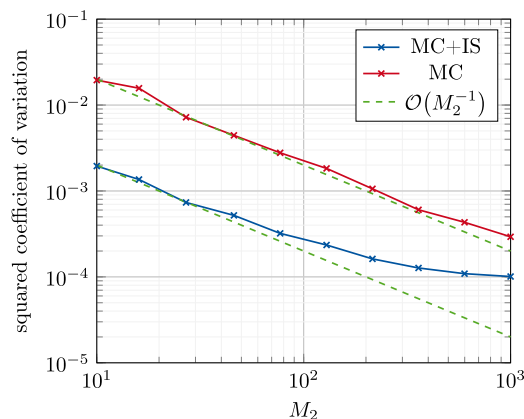
(b) Experiment 2: Squared coefficient of variation of the double loop Monte Carlo estimator for $\mathbb{E}[\Delta G_3]$ with and without importance sampling with respect to the number of sample paths in inner loop M_2.

Fig. 3 Numerical experiments verifying variance reduction in the double loop Monte Carlo estimator for level differences using importance sampling on the Kuramoto model for $G(x) = \Psi(x - K)$ with $K = 2.5$

control ζ by solving control Equation (14) off-line using an independent realization of the empirical law $\mu^{\bar{P}|\bar{N}}$ with $\bar{P} = 1000$ particles and $\bar{N} = 100$ time steps. The results verify a significantly reduced squared coefficient of variation (approximately one order of magnitude) with importance sampling. The estimator variance of $\mathbb{E}[\Delta G_\ell]$ is given by $\frac{V_{1,\ell}}{M_{1,\ell}} + \frac{V_{2,\ell}}{M_{1,\ell}M_{2,\ell}}$ (30). We notice convergence of the squared coefficient of variation as the second term vanishes with a large M_2 value.

We numerically verify Assumptions 1 and 2 for this rare-event observable to determine constants $\tilde{\alpha}$, $\tilde{w}$, and $\tilde{s}$ for the Kuramoto model with importance sampling. Figure 4 verifies Assumptions 1 and 2 with $\tilde{\alpha} = 1$ and $\{\tilde{w}, \tilde{s}\} = \{2, 1\}$ for the antithetic sampler (24). These rates imply a complexity of $\mathcal{O}(\text{TOL}_r^{-3})$ for the proposed multilevel DLMC estimator with importance sampling according to Theorem 1.

We implement the proposed adaptive multilevel DLMC algorithm (Algorithm 2) with inputs $\{\bar{M}_1, \bar{M}_2\} = \{1000, 100\}$, $\{\tilde{M}_1, \tilde{M}_2\} = \{25, 100\}$, and $\{\underline{M}_1, \underline{M}_2\} = \{100, 50\}$ using importance sampling for the 1D Kuramoto model. We use $\bar{P} = 1000$ particles and $\bar{N} = 100$ time steps to independently obtain one empirical realization of $\mu^{\bar{P}|\bar{N}}$ to compute importance sampling control ζ. Figure 5 presents the results of Algorithm 2 in this setting and numerically verifies the complexity rates obtained in Theorem 1. We observe that $K = 2.5$ corresponds to an expected value $\approx 3.2 \times 10^{-3}$, $K = 2$ corresponds to an expected value of $\approx 2.3 \times 10^{-2}$, and $K = 3$ corresponds to an expected value of $\approx 3.1 \times 10^{-4}$, all sufficiently rare events.

Figure 5a, b, and c correspond to $K = 2.5$. Figure 5a presents exact relative error for the proposed multilevel DLMC estimator, estimated using a reference multilevel DLMC approximation with $\text{TOL}_r = 1\%$, for multiple runs of Algorithm 2 with various prescribed relative error tolerances. Figure 5b verifies that the multilevel DLMC estimator with importance sampling satisfies error constraints with 95% confidence (corresponding to $\nu = 0.05$). The runtimes in Fig. 5b include the estimation time of the bias, $V_{1,\ell}$ and $V_{2,\ell}$, in adaptive Algorithm 2. Figure 5b confirms that the average computational runtime closely follows the predicted theoretical rate $\mathcal{O}(\text{TOL}_r^{-3})$ for the entire range of relative tolerances. Figure 5c depicts the average computational cost estimate for the multilevel DLMC estimator over the prescribed TOL_r values. The cost estimate is computed using (49). Figure 5b and c both display a complexity of $\mathcal{O}(\text{TOL}_r^{-3})$ for the multilevel DLMC estimator with importance sampling and the antithetic sampler, achieving one order complexity reduction compared with the single-level DLMC estimator with the same importance sampling scheme. Figure 5d plots the average work estimate for various threshold values K over a range of relative error tolerances. This verifies that the complexity rates are independent of parameter K.

7 Conclusion

Under certain assumptions that can be numerically verified, we have theoretically and numerically demonstrated the improvement of multilevel DLMC compared with the
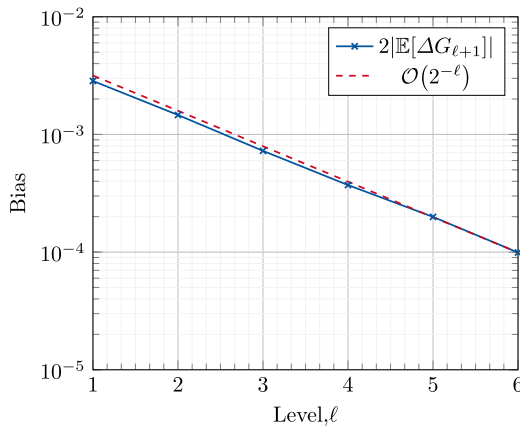
(a) Assumption 1 for $G(x) = \Psi(x - K)$: Double loop Monte Carlo bias estimator using Algorithm 4 with inputs $M_1 = 10^2$ and $M_2 = 10^4$ with respect to ℓ.
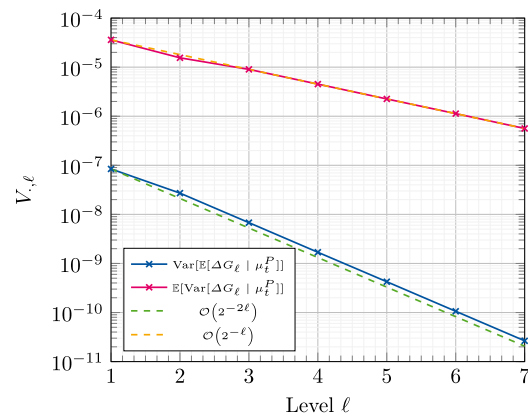
(b) Assumption 2 for $G(x) = \Psi(x-K)$: Double loop Monte Carlo estimator for $V_{1,\ell}, V_{2,\ell}$ using Algorithm 5 with inputs $M_1 = 10^2$ and $M_2 = 10^4$ with respect to ℓ.

Fig. 4 Convergence rates of level differences using the antithetic estimator ($\hat{\mathcal{G}}$) for the Kuramoto model with $G(x) = \Psi(x - K)$

single-level DLMC estimator used to approximate rare-event quantities of interest expressed as an expectation of a Lipschitz observable of the solution to stochastic particle systems in the mean-field limit. We used the importance sampling scheme introduced in Ben Rached et al. (2024) for all level-difference estimators in the proposed multilevel DLMC estimator and verified substantial variance reduction numerically. The proposed novel multilevel DLMC estimator achieved a complexity of $\mathcal{O}(\text{TOL}_r^{-3})$ for the treated example, one order less than the single-level DLMC estimator for the prescribed relative error tolerance TOL_r. Integrating the importance sampling scheme into the MLMC estimator ensured that the constant associated with its complexity also reduced significantly compared with the MLMC estimator for smooth, nonrare observables introduced in Haji-Ali and Tempone (2018).

Future studies could include extending the importance sampling scheme to higher-dimensional problems, using model reduction techniques or stochastic gradient-based learning methods to solve the associated higher-dimensional stochastic optimal control problem (see Remark 2). The importance sampling scheme could be further improved by solving an optimal control problem minimizing the level-difference estimator variance rather than the single-level DLMC estimator (see Remark 9). The multilevel DLMC algorithm could be optimized for determining the optimal parameters τ and θ (Haji-Ali et al. 2016b) or integrating a continuation MLMC algorithm (Collier et al. 2015) (see Remark 4). The present analysis could be extended to numerically address non-Lipschitz rare-event observables, such as the indicator function, to compute rare-event probabilities. Multiple discretization parameters for the decoupled

MV-SDE (P, N) suggest extending the current work to a multi-index Monte Carlo (Haji-Ali et al. 2016a; Haji-Ali and Tempone 2018) setting to further reduce computational complexity.

## A Computational cost of the multilevel double loop Monte Carlo estimator (22)

For a given level ℓ, we estimate $\mathbb{E}[\Delta G_\ell] = \mathbb{E}[G_\ell - G_{\ell-1}]$. First, we derive the computational cost of estimating $\mathbb{E}[G_\ell]$ using DLMC Algorithm 3. We consider the discretization $0 = t_0 < t_1 < t_2 < \ldots < t_{N_\ell} = T$ of the time domain $[0, T]$ with N_ℓ equal time steps (i.e., $t_n = n\Delta t$, $n = 0, 1, \ldots, N_\ell$ and $\Delta t = T/N_\ell$). First, we simulate the particle system at level ℓ. For each particle $p \in 1, \ldots, P_\ell$ and each time step $n \in 1, \ldots, N_\ell$, the Euler–Maruyama time-stepping scheme is written as

$$X_p^{P_\ell|N_\ell}(t_{n+1}) = X_p^{P_\ell|N_\ell}(t_n)$$
$$+ b\left(X_p^{P_\ell|N_\ell}(t_n), \frac{1}{P_\ell}\sum_{j=1}^{P_\ell}\kappa_1\left(X_p^{P_\ell|N_\ell}(t_n), X_j^{P_\ell|N_\ell}(t_n)\right)\right)\Delta t$$
$$+ \sigma\left(X_p^{P_\ell|N_\ell}(t_n), \frac{1}{P_\ell}\sum_{j=1}^{P_\ell}\kappa_2\left(X_p^{P_\ell|N_\ell}(t_n), X_j^{P_\ell|N_\ell}(t_n)\right)\right)$$
$$\Delta W(t_n), \quad n = 1, \ldots, N_\ell,$$
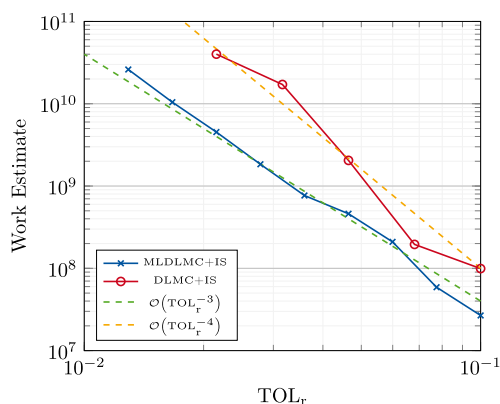$$X_p^{P_\ell|N_\ell}(t_0) \sim \mu_0. \tag{51}$$

The computational cost per time step per particle is $\mathcal{O}(P_\ell^{\gamma_p})$ because the cost of computing the empirical average
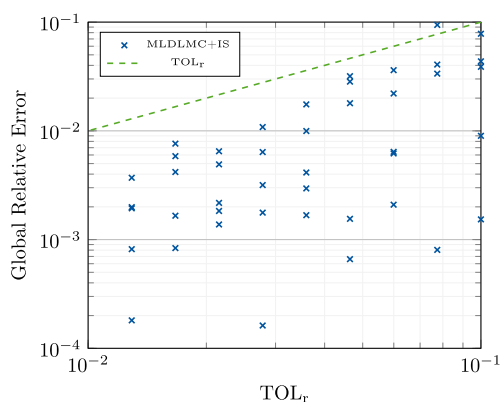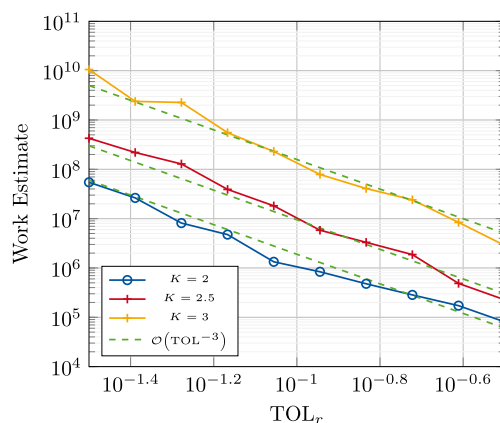
(a) Relative error of the estimator with respect to the relative tolerance TOL_r for $K = 2.5$.
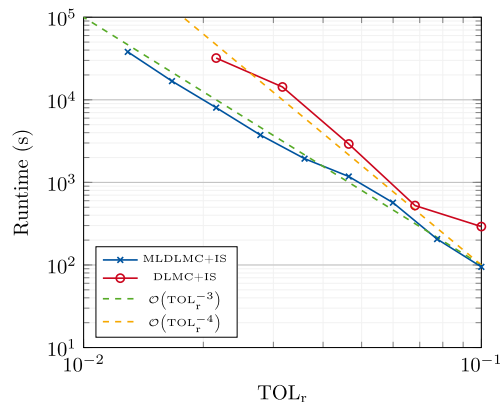
(b) Average computational runtime with respect to the relative tolerance TOL_r for $K = 2.5$.

(c) Average computational cost estimate with respect to the relative tolerance TOL_r for $K = 2.5$.

(d) Average computational cost estimate with respect to the relative tolerance TOL_r for different threshold (K) values.

Fig. 5 Algorithm 2 applied to the Kuramoto model for $G(x) = \Psi(x - K)$

Algorithm 3: General double loop Monte Carlo algorithm for decoupled McKean–Vlasov Stochastic Differential Equation

---

Inputs: P, N, M_1, M_2;
for $m_1 = 1, \ldots, M_1$ **do**
    Generate $\mu^{P|N}\left(\omega_{1:P}^{(m_1)}\right)$ with P-particle system and N time steps using (51);
    for $m_2 = 1, \ldots, M_2$ **do**
        Given $\mu^{P|N}\left(\omega_{1:P}^{(m_1)}\right)$, generate sample path of decoupled MV-SDE with N time steps with $\bar{\omega}^{(m_2)}$ using (52);
        Compute $G\left(\bar{X}^{P|N}(T)\right)\left(\omega_{1:P}^{(m_1)}, \bar{\omega}^{(m_2)}\right)$;
    end
end
Approximate $\mathbb{E}\left[G\left(\bar{X}^{P|N}(T)\right)\right]$ by
$\frac{1}{M_1}\sum_{m_1=1}^{M_1}\frac{1}{M_2}\sum_{m_2=1}^{M_2}G\left(\bar{X}^{P|N}(T)\right)\left(\omega_{1:P}^{(m_1)}, \bar{\omega}^{(m_2)}\right)$;

$\frac{1}{P_\ell}\sum_{j=1}^{P_\ell}\kappa\left(X_p^{P_\ell|N_\ell}(t_n), X_j^{P_\ell|N_\ell}(t_n)\right)$ in the drift and diffusion coefficients is assumed to be $\mathcal{O}(P_\ell^{\gamma_p})$ for $\gamma_p > 0$. This cost is $\mathcal{O}(P_\ell)$ for a naïve method. Hence, the computational cost of simulating a P_ℓ particle system once using N_ℓ time steps with scheme (51) is $\mathcal{O}(N_\ell P_\ell^{1+\gamma_p})$. For a given realization of the particle system, we simulate the decoupled MV-SDE (8) using the Euler–Maruyama scheme with the same time discretization as above. For each time step $n \in 1, \ldots, N_\ell$, the time-stepping scheme is written as

$$\bar{X}_\zeta^{P_\ell|N_\ell}(t_{n+1}) = \bar{X}_\zeta^{P_\ell|N_\ell}(t_n)$$
$$+ b\left(\bar{X}_\zeta^{P_\ell|N_\ell}(t_n), \frac{1}{P_\ell}\sum_{j=1}^{P_\ell}\kappa_1\left(\bar{X}_\zeta^{P_\ell|N_\ell}(t_n), X_j^{P_\ell|N_\ell}(t_n)\right)\right)\Delta t$$
$$+ \sigma\left(\bar{X}_\zeta^{P_\ell|N_\ell}(t_n), \frac{1}{P_\ell}\sum_{j=1}^{P_\ell}\kappa_2\left(\bar{X}_\zeta^{P_\ell|N_\ell}(t_n), X_j^{P_\ell|N_\ell}(t_n)\right)\right)$$

$$\Delta \bar{W}(t_n), \quad n = 1, \ldots, N_\ell,$$
$$\bar{X}_\zeta^{P_\ell | N_\ell}(t_0) \sim \mu_0. \tag{52}$$

The computational cost per time step is again $\mathcal{O}(P_\ell^{\gamma_p})$ due to the cost of computing the empirical average in the drift and diffusion coefficients. Hence, the computational cost of simulating the decoupled MV-SDE (8) using the above scheme (52) is $\mathcal{O}(N_\ell P_\ell^{\gamma_p})$. Thus, the computational complexity of DLMC in Algorithm 3 per level can be written as follows:

$$\mathcal{W}_\ell = \mathcal{O}(M_{1,\ell} \left(N_\ell P_\ell^{1+\gamma_p} + M_{2,\ell} \left(N_\ell P_\ell^{\gamma_p} \right) \right))$$
$$= \mathcal{O}(M_{1,\ell} N_\ell P_\ell^{1+\gamma_p} + M_{1,\ell} M_{2,\ell} N_\ell P_\ell^{\gamma_p}). \tag{53}$$

We obtain the expression in (34) by generalizing this cost estimate for any time-stepping scheme (with a computational complexity of $\mathcal{O}(N_\ell^{\gamma_n})$ instead of $\mathcal{O}(N_\ell)$) and then summing over all levels.

B Proof of Theorem 1

For given level $L \in \mathbb{N}$, (37) provides the optimal number of samples to satisfy the variance constraint in (29) for the proposed multilevel DLMC estimator. We bound the multilevel DLMC estimator (22) cost as follows:

$$\mathcal{W}[\mathcal{A}_{\text{MLMC}}] \lesssim \sum_{\ell=0}^L \left(M_{1,\ell} P_\ell^{1+\gamma_p} N_\ell^{\gamma_n} + M_{1,\ell} M_{2,\ell} P_\ell^{\gamma_p} N_\ell^{\gamma_n} \right)$$

$$\lesssim \sum_{\ell=0}^L \left((\mathcal{M}_{1,\ell} + 1) P_\ell^{1+\gamma_p} N_\ell^{\gamma_n} \right.$$

$$\left. + (\mathcal{M}_{1,\ell} + 1) \left(\frac{\tilde{\mathcal{M}}_\ell}{\lceil \mathcal{M}_{1,\ell} \rceil} + 1 \right) P_\ell^{\gamma_p} N_\ell^{\gamma_n} \right)$$

$$\leq \underbrace{\sum_{\ell=0}^L \left(\mathcal{M}_{1,\ell} P_\ell^{1+\gamma_p} N_\ell^{\gamma_n} + \tilde{\mathcal{M}}_\ell P_\ell^{\gamma_p} N_\ell^{\gamma_n} \right)}_{=W_1, \text{objective function of (35)}}$$

$$+ \underbrace{\sum_{\ell=0}^L \left(P_\ell^{1+\gamma_p} N_\ell^{\gamma_n} + P_\ell^{\gamma_p} N_\ell^{\gamma_n} \right)}_{W_2, \text{cost of one realization per level}} + \underbrace{\sum_{\ell=0}^L \mathcal{M}_{1,\ell} P_\ell^{\gamma_p} N_\ell^{\gamma_n}}_{=W_3}$$

$$+ \underbrace{\sum_{\ell=0}^L \frac{\tilde{\mathcal{M}}_\ell}{\lceil \mathcal{M}_{1,\ell} \rceil} P_\ell^{\gamma_p} N_\ell^{\gamma_n}}_{=W_4}. \tag{54}$$

Because $P_\ell > 1$ and $\gamma_p > 0$, W_3 is always dominated by the first term in W_1 (i.e., $\sum_{\ell=0}^L \mathcal{M}_{1,\ell} P_\ell^{1+\gamma_p} N_\ell^{\gamma_n}$). We analyze

each term individually. By substituting (36) in W_1,

$$W_1 = \frac{C_\nu^2}{(1-\theta)^2 \text{TOL}_r^2 |\mathbb{E}[G]|^2}$$
$$\times \left(\sum_{\ell=0}^L \sqrt{P_\ell^{\gamma_p} N_\ell^{\gamma_n}} (\sqrt{V_{1,\ell} P_\ell} + \sqrt{V_{2,\ell}}) \right)^2. \tag{55}$$

By selecting level L that satisfies (26) and using Assumption 1,

$$|\mathbb{E}[G - G_L]| \leq C_b \tau^{-\tilde{\alpha} L} \approx \theta \text{TOL}_r |\mathbb{E}[G]|, \tag{56}$$

and hence,

$$L = \left\lceil \frac{1}{\tilde{\alpha}} \log \left(\frac{C_b}{\theta |\mathbb{E}[G]|} \text{TOL}_r^{-1} \right) \right\rceil. \tag{57}$$

Using the hierarchies (20) and Assumption 2 in (55),

$$W_1 \lesssim \text{TOL}_r^{-2} \left(\sum_{\ell=0}^L \tau^{\frac{1+\gamma_p+\gamma_n-\tilde{w}}{2}\ell} + \tau^{\frac{\gamma_p+\gamma_n-\tilde{s}}{2}\ell} \right)^2. \tag{58}$$

The summation in (58) has two terms; thus, we have the following two cases:

- **Case 1:** $\tilde{s} + 1 \geq \tilde{w}$, i.e., the first term dominates, and W_1 can be expressed as

$$W_1 \lesssim \text{TOL}_r^{-2} \left(\sum_{\ell=0}^L \tau^{\frac{1+\gamma_p+\gamma_n-\tilde{w}}{2}\ell} \right)^2, \tag{59}$$

and simplified using (57) and the sum of a geometric series,

$$W_1 \lesssim \begin{cases} \text{TOL}_r^{-2} & , \tilde{w} > 1 + \gamma_p + \gamma_n \\ \text{TOL}_r^{-2}(\log \text{TOL}_r^{-1})^2 & , \tilde{w} = 1 + \gamma_p + \gamma_n \\ \text{TOL}_r^{-2-\left(\frac{1+\gamma_p+\gamma_n-\tilde{w}}{\tilde{\alpha}}\right)} & , \tilde{w} < 1 + \gamma_p + \gamma_n \end{cases} \tag{60}$$

which can be expressed more compactly:

$$W_1 \lesssim \text{TOL}_r^{-2-\max\left(0, \frac{1+\gamma_p+\gamma_n-\tilde{w}}{\tilde{\alpha}}\right)} (\log \text{TOL}_r^{-1})^{2\mathcal{J}_1}, \tag{61}$$

where

$$\mathcal{J}_1 = \begin{cases} 1, & \text{if } \tilde{w} = 1 + \gamma_p + \gamma_n, \\ 0, & \text{else.} \end{cases}$$

- **Case 2:** $\tilde{s} + 1 < \tilde{w}$, i.e., the second term dominates, and express W_1 as

$$W_1 \lesssim \text{TOL}_r^{-2} \left(\sum_{\ell=0}^{L} \beta^{\frac{\gamma_p + \gamma_n - \tilde{s}}{2} \ell} \right)^2, \tag{62}$$

and simplify using the sum of a geometric series,

$$W_1 \lesssim \begin{cases} \text{TOL}_r^{-2} & , \tilde{s} > \gamma_p + \gamma_n \\ \text{TOL}_r^{-2} (\log \text{TOL}_r^{-1})^2 & , \tilde{s} = \gamma_p + \gamma_n \\ \text{TOL}_r^{-2 - \left(\frac{\gamma_p + \gamma_n - \tilde{s}}{\tilde{\alpha}} \right)} & , \tilde{s} < \gamma_p + \gamma_n \end{cases} \tag{63}$$

which can be expressed more compactly:

$$W_1 \lesssim \text{TOL}_r^{-2 - \max\left(0, \frac{\gamma_p + \gamma_n - \tilde{s}}{\tilde{\alpha}} \right)} (\log \text{TOL}_r^{-1})^2 \mathcal{J}_2, \tag{64}$$

where

$$\mathcal{J}_2 = \begin{cases} 1, & \text{if } \tilde{s} = \gamma_p + \gamma_n, \\ 0, & \text{else.} \end{cases}$$

In this case, W_3 is of a lower order than the first term in W_1; therefore, W_1 dominates as $\text{TOL}_r \to 0$.

Using (20), for W_2,

$$W_2 \lesssim \text{TOL}_r^{-\left(\frac{1 + \gamma_p + \gamma_n}{\tilde{\alpha}} \right)}. \tag{65}$$

Next, we examine W_4:

$$W_4 = \sum_{\ell=0}^{L} \frac{\tilde{\mathcal{M}}_\ell}{\lceil \mathcal{M}_{1,\ell} \rceil} P_\ell^{\gamma_p} N_\ell^{\gamma_n}$$

$$\leq \sum_{\ell=0}^{L} \frac{\tilde{\mathcal{M}}_\ell}{\max\{1, \mathcal{M}_{1,\ell}\}} P_\ell^{\gamma_p} N_\ell^{\gamma_n}$$

$$= \sum_{\ell=0}^{L} \min\{\tilde{\mathcal{M}}_\ell P_\ell^{\gamma_p} N_\ell^{\gamma_n}, \frac{\tilde{\mathcal{M}}_\ell}{\mathcal{M}_{1,\ell}} P_\ell^{\gamma_p} N_\ell^{\gamma_n}\}$$

$$\leq \sum_{\ell=0}^{L} \tilde{\mathcal{M}}_\ell P_\ell^{\gamma_p} N_\ell^{\gamma_n}. \tag{66}$$

W_4 has the same or lower order than that of W_1. Next, we must ensure that W_1 dominates W_2 as $\text{TOL}_r \to 0$ for the proposed multilevel DLMC method to be feasible. Comparing (65) to W_1 for the cases, the following condition ensures W_1 is the dominant term:

$$\tilde{\alpha} \geq \frac{1}{2} \min(\tilde{w}, 1 + \tilde{s}, 1 + \gamma_p + \gamma_n). \tag{67}$$

Thus, (67), (61), and (64) complete the proof.

C Estimating level differences for the multilevel double loop Monte Carlo estimator

---

Algorithm 4: Importance sampling scheme to estimate $\mathbb{E}[\Delta G_\ell]$ using antithetic sampler

---

Inputs: $\ell, M_1, M_2, \zeta(\cdot, \cdot)$;

for $m_1 = 1, \ldots, M_1$ **do**

    Generate $\mu_\ell(\omega_{1:P_\ell}^{(\ell,m_1)})$ using (16);

    for $a = 1, \ldots, \tau$ **do**

        Generate $\mu_{\ell-1}^{(a)}(\omega_{(a-1)P_{\ell-1}+1:aP_{\ell-1}}^{(\ell,m_1)})$ using (16);

    end

    for $m_2 = 1, \ldots, M_2$ **do**

        Given $\mu_\ell(\omega_{1:P_\ell}^{(\ell,m_1)})$ and $\zeta(\cdot, \cdot)$, generate sample path of (10) at level ℓ with $\bar{\omega}^{(\ell,m_2)}$;

        Compute $G_\ell^{\text{IS}}(\omega_{1:P_\ell}^{(\ell,m_1)}, \bar{\omega}^{(\ell,m_2)})$;

        for $a = 1, \ldots, \tau$ **do**

            Given $\mu_{\ell-1}^{(a)}(\omega_{(a-1)P_{\ell-1}+1:aP_{\ell-1}}^{(\ell,m_1)})$ and $\zeta(\cdot, \cdot)$, generate sample path of (10) at level $\ell - 1$ with $\bar{\omega}^{(\ell,m_2)}$;

            Compute $G_{\ell-1}^{\text{IS}}(\omega_{(a-1)P_{\ell-1}+1:aP_{\ell-1}}^{(\ell,m_1)}, \bar{\omega}^{(\ell,m_2)})$;

        end

        $\hat{\mathcal{G}}_{\ell-1}^{\text{IS}}(\omega_{1:P_\ell}^{(\ell,m_1)}, \bar{\omega}^{(\ell,m_2)}) = \frac{1}{\tau} \sum_{a=1}^{\tau} G_{\ell-1}^{\text{IS}}(\omega_{(a-1)P_{\ell-1}+1:aP_{\ell-1}}^{(\ell,m_1)}, \bar{\omega}^{(\ell,m_2)})$;

    end

    $\Delta G_\ell^{(m_1,m_2)} = (G_\ell^{\text{IS}} - \hat{\mathcal{G}}_{\ell-1}^{\text{IS}})(\omega_{1:P_\ell}^{(\ell,m_1)}, \bar{\omega}^{(\ell,m_2)})$;

end

Approximate $\mathbb{E}[G_\ell - G_{\ell-1}]$ by $\frac{1}{M_1} \sum_{m_1=1}^{M_1} \frac{1}{M_2} \sum_{m_2=1}^{M_2} \Delta G_\ell^{(m_1,m_2)}$;

D Estimating variances for the adaptive multilevel double loop Monte Carlo algorithm

---

Algorithm 5: Estimating $V_{1,\ell}$ and $V_{2,\ell}$ for adaptive multilevel DLMC

Inputs: $\ell, M_1, M_2, \zeta(\cdot, \cdot)$;

for $m_1 = 1, \ldots, M_1$ **do**

    Generate $\mu_\ell(\omega_{1:P_\ell}^{(\ell,m_1)})$ using (16);

    for $a = 1, \ldots, \tau$ **do**

        Generate $\mu_{\ell-1}^{(a)}(\omega_{(a-1)P_{\ell-1}+1:aP_{\ell-1}}^{(\ell,m_1)})$ using (16);

    end

    for $m_2 = 1, \ldots, M_2$ **do**

        Given $\mu_\ell(\omega_{1:P_\ell}^{(\ell,m_1)})$ and $\zeta(\cdot, \cdot)$, generate sample path of (8) at level ℓ with $\bar{\omega}^{(\ell,m_2)}$;

        Compute $G_\ell^{\mathrm{IS}}(\omega_{1:P_\ell}^{(\ell,m_1)}, \bar{\omega}^{(\ell,m_2)})$;

        for $a = 1, \ldots, \tau$ **do**

            Given $\mu_{\ell-1}^{(a)}(\omega_{(a-1)P_{\ell-1}+1:aP_{\ell-1}}^{(\ell,m_1)})$ and $\zeta(\cdot, \cdot)$, generate sample path of (8) at level $\ell - 1$ with $\bar{\omega}^{(\ell,m_2)}$;

            Compute $G_{\ell-1}^{\mathrm{IS}}(\omega_{(a-1)P_{\ell-1}+1:aP_{\ell-1}}^{(\ell,m_1)}, \bar{\omega}^{(\ell,m_2)})$;

        end

        $\hat{\mathcal{G}}_{\ell-1}^{\mathrm{IS}}(\omega_{1:P_\ell}^{(\ell,m_1)}, \bar{\omega}^{(\ell,m_2)}) = \frac{1}{\tau} \sum_{a=1}^{\tau} G_{\ell-1}^{\mathrm{IS}}(\omega_{(a-1)P_{\ell-1}+1:aP_{\ell-1}}^{(\ell,m_1)}, \bar{\omega}^{(\ell,m_2)})$;

    end

    $\Delta G_\ell^{(m_1,m_2)} = (G_\ell^{\mathrm{IS}} - \hat{\mathcal{G}}_{\ell-1}^{\mathrm{IS}})(\omega_{1:P_\ell}^{(\ell,m_1)}, \bar{\omega}^{(\ell,m_2)})$;

    Approximate $\mathbb{E}\left[\Delta G_\ell \mid \{\mu_\ell, \mu_{\ell-1}\}(\omega_{1:P_\ell}^{(\ell,m_1)})\right]$ by $\frac{1}{M_2} \sum_{m_2=1}^{M_2} \Delta G_\ell^{(m_1,m_2)}$;

    Approximate $\mathrm{Var}\left[\Delta G_\ell \mid \{\mu_\ell, \mu_{\ell-1}\}(\omega_{1:P_\ell}^{(\ell,m_1)})\right]$ by sample variance of $\left\{\Delta G_\ell^{(m_1,m_2)}\right\}_{m_2=1}^{M_2}$;

end

Approximate $V_{1,\ell}$ by sample variance of $\left\{\mathbb{E}\left[\Delta G_\ell \mid \{\mu_\ell, \mu_{\ell-1}\}(\omega_{1:P_\ell}^{(\ell,m_1)})\right]\right\}_{m_1=1}^{M_1}$;

Approximate $V_{2,\ell}$ by $\frac{1}{M_1} \sum_{m_1=1}^{M_1} \mathrm{Var}\left[\Delta G_\ell \mid \{\mu_\ell, \mu_{\ell-1}\}(\omega_{1:P_\ell}^{(\ell,m_1)})\right]$.

Declarations

References

Acebrón, J.A., Bonilla, L.L., Pérez, C.J., Vicente, F.R., Spigler, R.: The Kuramoto model: a simple paradigm for synchronization phenomena. Rev. Mod. Phys. **77**(1), 137 (2005)

Ash, R.B., Doléans-Dade, C.A.: Probability and Measure Theory. Academic Press, Cambridge (2000)

Awad, H.P., Glynn, P.W., Rubinstein, R.Y.: Zero-variance importance sampling estimators for Markov process expectations. Math. Oper. Res. **38**(2), 358–388 (2013)

Alaya, M.B., Hajji, K., Kebaier, A.: Adaptive importance sampling for multilevel Monte Carlo Euler method. Stochastics **95**(2), 303–327 (2023)

Amar, E.B., Rached, N.B., Haji-Ali, A.-L., Tempone, R.: State-dependent importance sampling for estimating expectations of functionals of sums of independent random variables. Stat. Comput. **33**(2), 40 (2023)

Ben Hammouda, C., Ben Rached, N., Tempone, R.: Importance sampling for a robust and efficient multilevel Monte Carlo estimator for stochastic reaction networks. Stat. Comput. **30**, 1665–1689 (2020)

Ben Hammouda, C., Ben Rached, N., Tempone, R., Wiechert, S.: Learning-based importance sampling via stochastic optimal control for stochastic reaction networks. Stat. Comput. **33**(3), 58 (2023)

Ben Rached, N., Haji-Ali, A.L., Subbiah Pillai, S.M., Tempone. R.: Double-loop importance sampling for McKean–Vlasov stochastic differential equation. Stat. Comput. (2024). https://doi.org/10.1007/s11222-024-10497-3

Bossy, M., Talay, D.: Convergence rate for the approximation of the limit law of weakly interacting particles: application to the Burgers equation. Ann. Appl. Probab. **6**(3), 818–861 (1996)

Bossy, M., Talay, D.: A stochastic particle method for the McKean-Vlasov and the Burgers equation. Math. Comput. **66**(217), 157–192 (1997)

Buckdahn, R., Li, J., Peng, S., Rainer, C.: Mean-field stochastic differential equations and associated PDEs (2017)

Bujok, K., Hambly, B.M., Reisinger, C.: Multilevel simulation of functionals of Bernoulli random variables with application to basket credit derivatives. Methodol. Comput. Appl. Probab. **17**, 579–604 (2015)

Bush, N., Hambly, B.M., Haworth, H., Jin, L., Reisinger, C.: Stochastic evolution equations in portfolio credit modelling. SIAM J. Financ. Math. **2**(1), 627–664 (2011)

Collier, N., Haji-Ali, A.-L., Nobile, F., Von Schwerin, E., Tempone, R.: A continuation multilevel Monte Carlo algorithm. BIT Numer. Math. **55**, 399–432 (2015)

Crisan, D., McMurray, E.: Smoothing properties of McKean-Vlasov SDEs. Probab. Theory Relat. Fields **171**, 97–148 (2018)

Crisan, D., McMurray, E.: Cubature on Wiener space for McKean-Vlasov SDEs with smooth scalar interaction. Ann. Appl. Probab. **29**(1), 130–177 (2019)

Crisan, D., Xiong, J.: Approximate McKean-Vlasov representations for a class of SPDEs. Stoch. Int. J. Probab. Stoch. Process. **82**(1), 53–68 (2010)

Cumin, D., Unsworth, C.P.: Generalising the Kuramoto model for the study of neuronal synchronisation in the brain. Phys. D **226**(2), 181–196 (2007)

Dobramysl, U., Rüdiger, S., Erban, R.: Particle-based multiscale modeling of calcium puff dynamics. Multiscale Model. Simul. **14**(3), 997–1016 (2016)

dos Reis, G., Engelhardt, S., Smith, G.: Simulation of McKean-Vlasov SDEs with super-linear growth. IMA J. Numer. Anal. **42**(1), 874–922 (2022)

dos Reis, G., Smith, G., Tankov, P.: Importance sampling for McKean-Vlasov SDEs. Appl. Math. Comput. **453**, 128078 (2023)

Dupuis, P., Wang, H.: Importance sampling, large deviations, and differential games. Stoch. Int. J. Probab. Stoch. Process. **76**(6), 481–508 (2004)

Erban, R., Haskovec, J. From individual to collective behaviour of coupled velocity jump processes: a locust example. arXiv preprint arXiv:1104.2584 (2011)

Fang, W., Giles, M.B.: Multilevel Monte Carlo method for ergodic SDEs without contractivity. J. Math. Anal. Appl. **476**(1), 149–176 (2019)

Forrester, D.M.: Arrays of coupled chemical oscillators. Sci. Rep. **5**(1), 16994 (2015)

Giles, M.B.: Multilevel Monte Carlo path simulation. Oper. Res. **56**(3), 607–617 (2008)

Giles, M.B.: Multilevel Monte Carlo methods. Acta Numer. **24**, 259–328 (2015)

Giles, M.B., Szpruch, L.: Antithetic multilevel Monte Carlo estimation for multi-dimensional SDEs without Lévy area simulation (2014)

Haji Ali, A.L.: Pedestrian flow in the mean field limit (2012)

Haji-Ali, A.-L., Tempone, R.: Multilevel and multi-index Monte Carlo methods for the McKean-Vlasov equation. Stat. Comput. **28**, 923–935 (2018)

Haji-Ali, A.-L., Nobile, F., Tempone, R.: Multi-index Monte Carlo: when sparsity meets sampling. Numer. Math. **132**, 767–806 (2016)

Haji-Ali, A.-L., Nobile, F., von Schwerin, E., Tempone, R.: Optimization of mesh hierarchies in multilevel Monte Carlo samplers. Stoch. Partial Differ. Equ. Anal. Comput. **4**(1), 76–112 (2016)

Hammersley, W.R., Šiška, D., Szpruch, Ł.: Weak existence and uniqueness for McKean–Vlasov SDEs with common noise (2021)

Hartmann, C., Schütte, C., Zhang, W.: Projection-based algorithms for optimal control and importance sampling of diffusions (2015)

Hartmann, C., Schütte, C., Zhang, W.: Model reduction algorithms for optimal control and importance sampling of diffusions. Nonlinearity **29**(8), 2298 (2016)

Hartmann, C., Richter, L., Schütte, C., Zhang, W.: Variational characterization of free energy: theory and algorithms. Entropy **19**(11), 626 (2017)

Hartmann, C., Schütte, C., Weber, M., Zhang, W.: Importance sampling in path space for diffusion processes with slow-fast variables. Probab. Theory Relat. Fields **170**, 177–228 (2018)

Kebaier, A., Lelong, J.: Coupling importance sampling and multilevel Monte Carlo using sample average approximation. Methodol. Comput. Appl. Probab. **20**, 611–641 (2018)

Kloeden, P.E., Platen, E.: Numerical Solution of Stochastic Differential Equations. Springer, Berlin (1992)

Kolokoltsov, V.N., Troeva, M.: On mean field games with common noise and Mckean-Vlasov SPDEs. Stoch. Anal. Appl. **37**(4), 522–549 (2019)

Kroese, D.P., Taimre, T., Botev, Z.I.: Handbook of Monte Carlo Methods. John Wiley & Sons, Hoboken (2013)

Lemaire, V., Pagès, G.: Multilevel Richardson–Romberg extrapolation (2017)

Li, Y., Mao, X., Song, Q., Fuke, W., Yin, G.: Strong convergence of Euler-Maruyama schemes for McKean-Vlasov stochastic differential equations under local Lipschitz conditions of state variables. IMA J. Numer. Anal. **43**(2), 1001–1035 (2023)

McKean Jr, H.P.: A class of Markov processes associated with nonlinear parabolic equations. Proc. Natl. Acad. Sci. **56**(6), 1907–1911 (1966)

Méléard, S.: Asymptotic behaviour of some interacting particle systems; McKean-Vlasov and Boltzmann models. In: Talay, D., Tubaro, L. (eds.) Probabilistic Models for Nonlinear Partial Differential Equations, vol. 1627, pp. 42–95. Springer, Berlin (1996)

Mishura, Y., Veretennikov, A.: Existence and uniqueness theorems for solutions of McKean-Vlasov stochastic equations. Theory Probab. Math. Stat. **103**, 59–101 (2020)

Ogawa, S.: Monte Carlo simulation of nonlinear diffusion processes. Jpn. J. Ind. Appl. Math. **9**, 25–33 (1992)

Oksendal, B.: Stochastic Differential Equations: An Introduction with Applications. Springer Science & Business Media, Berlin (2013)

Rosin, M.S., Ricketson, L.F., Dimits, A.M., Caflisch, R.E., Cohen, B.I.: Multilevel Monte Carlo simulation of coulomb collisions. J. Comput. Phys. **274**, 140–157 (2014)

Sznitman, A.S.: Topics in propagation of chaos. In: Lecture Notes in Mathematics, pp. 165–251 (1991)

Szpruch, Ł., Tse, A.: Antithetic multilevel particle system sampling method for McKean-Vlasov SDEs. arXiv preprint arXiv:1903.07063 (2019)

Szpruch, L., Tan, S., Tse, A.: Iterative multilevel particle approximation for McKean-Vlasov SDEs. Ann. Appl. Probab. **29**(4), 2230–2265 (2019)

Zhang, W., Wang, H., Hartmann, C., Weber, M., Schütte, C.: Applications of the cross-entropy method to importance sampling and optimal control of diffusions. SIAM J. Sci. Comput. **36**(6), A2654–A2672 (2014)