



Deposited via The University of Leeds.

White Rose Research Online URL for this paper:

<https://eprints.whiterose.ac.uk/id/eprint/245165/>

Version: Accepted Version

Article:

Rached, N.B., Haji-Ali, A.-L., Pillai, S.M.S. et al. (2027) Multi-index importance sampling for McKean–Vlasov stochastic differential equations. *Journal of Computational and Applied Mathematics*, 490. 117988. ISSN: 0377-0427

<https://doi.org/10.1016/j.cam.2026.117988>

This is an author produced version of an article published in the *Journal of Computational and Applied Mathematics*, made available via the University of Leeds Research Outputs Policy under the terms of the Creative Commons Attribution License (CC-BY), which permits unrestricted use, distribution and reproduction in any medium, provided the original work is properly cited.

Reuse

This article is distributed under the terms of the Creative Commons Attribution (CC BY) licence. This licence allows you to distribute, remix, tweak, and build upon the work, even commercially, as long as you credit the authors for the original work. More information and the full terms of the licence here:

<https://creativecommons.org/licenses/>

Takedown

If you consider content in White Rose Research Online to be in breach of UK law, please notify us by emailing eprints@whiterose.ac.uk including the URL of the record and the reason for the withdrawal request.

Multi-index importance sampling for McKean–Vlasov stochastic differential equations

Nadhir Ben Rached^a, Abdul-Lateef Haji-Ali^b, Shyam Mohan Subbiah Pillai^{c,*},
Raúl Tempone^c

^a*Department of Statistics, School of Mathematics, University of Leeds, Leeds, LS2 9JT, United Kingdom*

^b*Department of Actuarial Mathematics and Statistics, School of Mathematical and Computer Sciences, Heriot-Watt University, Edinburgh, EH14 4AS, United Kingdom*

^c*Computer, Electrical and Mathematical Sciences & Engineering Division (CEMSE), King Abdullah University of Science and Technology (KAUST), Thuwal, Saudi Arabia*

Abstract

This work addresses the estimation of rare-event quantities expressed as expectations of smooth observables of solutions to a broad class of McKean–Vlasov stochastic differential equations (MV-SDEs). Building on the double loop Monte Carlo (DLMC) method with stochastic optimal control-based importance sampling (IS) introduced by [1], this work extends this framework to the multi-index Monte Carlo (MIMC) setting. The resulting multi-index DLMC estimator mitigates the explosion of the coefficient of variation for rare event quantities. Moreover, it exploits the sampling efficiency of MIMC by leveraging the propagation of chaos to ensure mixed-difference variances vanish in the mean-field limit. The complexity analysis relies on assumptions on mixed-difference bias and variance decay, similar to standard MIMC assumptions. Although not rigorously proved, this work presents strong numerical evidence in support of these assumptions. The primary contribution of this work is the novel numerical integration of the MIMC method with IS for MV-SDEs. This approach reduces the computational complexity from $\mathcal{O}(\text{TOL}_r^{-4})$ for the DLMC estimator to $\mathcal{O}(\text{TOL}_r^{-2}(\log \text{TOL}_r^{-1})^2)$, enabling an accurate estimation of rare-event quantities within a prescribed relative error tolerance TOL_r . Numerical experiments on the Kuramoto model from statistical physics demonstrate computational savings of several orders of magnitude for the multi-index DLMC estimator with IS, compared with the standard Monte Carlo (MC) method.

Keywords: McKean–Vlasov stochastic differential equation, importance sampling, multi-index Monte Carlo, decoupling approach, double loop Monte Carlo.

2020 MSC: 60H35, 65C05, 65C30, 65C35.

*Corresponding author, shyam.pillai@kaust.edu.sa

1. Introduction

McKean–Vlasov stochastic differential equations (MV-SDEs) are a class of SDEs whose drift and diffusion depend on the law of the solution [2]. These SDEs arise as mean-field limits of stochastic interacting particle systems with applications in pedestrian modeling [3], animal behavior [4], biology [5], finance [6], and chemistry [7]. This work focuses on building a computationally efficient Monte Carlo (MC) estimator for $\mathbb{E}[G(X(T))]$, where $X : [0, T] \times \Omega \rightarrow \mathbb{R}^d$ denotes the McKean–Vlasov process, and $G : \mathbb{R}^d \rightarrow \mathbb{R}$ represents a sufficiently smooth rare-event observable, meaning $\mathbb{E}[G(X(T))]$ is dominated by contributions from the tail of the law of X . For example, $G(x)$ can be a mollified version of $\mathbb{1}_{\{x > K\}}$, for some large threshold K . The standard MC method is prohibitively inefficient in this regime because the probability of sampling trajectories that contribute to the estimator is minimal.

Typically, explicit solutions to MV-SDEs are unavailable. These SDEs are first approximated using a stochastic interacting particle system [8], comprising P coupled d -dimensional SDEs. This system is generally not solvable in closed form; hence, its trajectories are numerically approximated via a time-stepping scheme with N time steps. Thus, the approximation of $\mathbb{E}[G(X(T))]$ for MV-SDEs involves two parameters: the numbers of particles P and time steps N . Standard MC estimators based on the above P -particle approximation and the Euler–Maruyama time-stepping method [9, 10] achieve $\mathcal{O}(\text{TOL}^{-4})$ complexity for smooth, nonrare observables under bounded, Lipschitz coefficients. However, for rare events, the cost prohibitively increases owing to the exploding coefficient of variation problem [11]. The coefficient of variation of the estimator explodes as the event becomes rarer, making standard MC intractable. To overcome this problem, importance sampling (IS) via a measure change in the path space is applied as a standard variance reduction technique [11]. This technique involves transforming the underlying probability measure to sample rare trajectories and appropriately reweight them to ensure an unbiased estimator.

The initial studies of IS for MV-SDEs have appeared in the works by [12] and [1]. A naive application of IS to the MV-SDE fails because, under a new measure, the process’s law is no longer consistent with the original dynamics of the MV-SDE. Hence, [12] proposed a decoupling approach, where the law in the drift/diffusion of the MV-SDE is replaced by an empirical realization of the law precomputed from a particle system, after which a measure change is applied. Building on this idea, [1] introduced a double-loop MC (DLMC) estimator and employed stochastic optimal control to design a time- and path-dependent IS control, reaching a complexity of $\mathcal{O}(\text{TOL}_r^{-4})$, matching the standard MC method for nonrare observables [9]. Recently, the multilevel MC method was combined with this IS scheme [13], reducing the complexity further compared with the DLMC estimator.

The multi-index MC (MIMC) method, introduced by [14], extends the multilevel MC method by exploiting mixed refinement in multiple discretization directions (in this work, particles P and time steps N). The MIMC method relies on mixed regularity with respect to these directions, further reducing the

computational cost. In the MV-SDE setting, as $P \rightarrow \infty$, the errors from finite particle approximation and time discretization become coupled, an effect of the propagation of chaos [8] that cancels out leading-order error terms, which is leveraged in the multi-index complexity analysis. Although the MIMC method has been applied to smooth, nonrare observables of particle systems [9], the goal of this work is to combine the MIMC method with the IS scheme from [1]. This work presents an estimator developed for rare-event expectations of MV-SDEs that surpasses the multilevel DLMC estimator [13]. Although the multilevel MC method with IS has been studied in other contexts (e.g., standard SDEs [15], ergodic SDEs [16], stochastic reaction networks [17], and diffusions in finance [18]), this work is the first integration of IS with the MIMC method to the best of our knowledge. The primary contributions of this work are as follows.

1. This work extends the DLMC estimator of [1] to the multi-index setting, proposing a multi-index DLMC estimator for MV-SDEs. Until now, variance reduction for rare events in MV-SDEs had only been explored in single [12, 1] and multilevel MC [13] settings. This work integrates IS with MIMC for MV-SDEs. This work applies the IS control from [1] across all multi-indices and introduces a strongly coupled antithetic sampler to control the rare-event variance of the mixed-difference estimators.
2. This work derives a complexity result (Theorem 1) for MV-SDEs based on assumptions on the bias and variance of the estimator (see Assumptions 1-2) that are similar to standard mixed-difference assumptions made in MIMC. This work reveals that the proposed approach improves on the multilevel DLMC estimator, which achieves $\mathcal{O}(\text{TOL}_r^{-3})$ complexity, to an asymptotic complexity of $\mathcal{O}(\text{TOL}_r^{-2}(\log \text{TOL}_r^{-1})^2)$. This work presents strong numerical evidence in support of these assumptions on a simple MV-SDE, leaving further theoretical analysis to future work.
3. This work numerically investigates the influence of IS on variance reduction in the proposed multi-index DLMC estimator. For sufficiently smooth observables, the numerical simulations demonstrate a significant variance reduction on all mixed-difference estimators. In the Kuramoto test case (Section 2.1), this result improves the complexity from $\mathcal{O}(\text{TOL}_r^{-4})$ obtained in the work by [1] to $\mathcal{O}(\text{TOL}_r^{-2}(\log \text{TOL}_r^{-1})^2)$ in the multi-index setting (see Figure 5). This substantial improvement makes estimating rare-event quantities up to a relative error tolerance TOL_r feasible in practice. Moreover, this work numerically studies the effect of the smoothness of the observable on the mixed-difference bias and variance convergence rates. The experiments confirm that the assumptions hold for a sufficiently smooth G and demonstrate how the convergence degrades as G becomes less regular.

This paper is organised as follows. Section 2 introduces the MV-SDE, associated notation, and problem statement. Next, Section 3 reviews the decoupling approach in the work by [12], recalls the IS control for the decoupled MV-SDE

from [1], and formulates a DLMC estimator with IS. Then, Section 4 presents the novel integration of IS with the multi-index DLMC estimator and derives the complexity bounds, conditional on two key assumptions on mixed-difference bias and variance decay (Assumptions 1 and 2). Section 5 applies the method to the Kuramoto model from statistical physics to test the proposed approach and empirically verify Assumptions 1 and 2 for problems of varying regularity. The study confirms the theoretical complexity rates for the multi-index DLMC estimator, demonstrating the effectiveness of the proposed method in rare-event estimation for sufficiently smooth problems.

Notation: The notation $\langle \cdot, \cdot \rangle$ represents the Euclidean dot product on $\mathbb{R}^d$, and $\|\cdot\|$ denotes the corresponding norm. For a scalar function, ∇ and ∇^2 denote the gradient and Hessian, respectively. The notation $\cdot : \cdot$ signifies the Frobenius inner product between two matrix-valued functions. We denote by $C^k(A, B)$ the space of functions $f : A \rightarrow B$ that are k -times continuously differentiable on A . When $B = \mathbb{R}$, we simply write $C^k(A)$. For non-negative quantities a and b , $a \lesssim b$ indicates the existence of a constant $c > 0$ such that $a \leq cb$. Unless stated otherwise, c is independent of the discretisation and tolerance parameters, but may depend on fixed problem data.

2. McKean–Vlasov stochastic differential equation

This work considers a broad class of McKean–Vlasov equations arising from the mean-field limit of stochastic interacting particle systems with pairwise interaction kernels [8], defined on the probability space $\{\Omega, \mathcal{F}, \{\mathcal{F}_t\}_{t \geq 0}, \mathbb{P}\}$, where $\mathcal{F}_t$ represents the filtration of a standard Wiener process $\{W(t) : t \in [0, T]\}$. For measurable functions $b : \mathbb{R}^d \times \mathbb{R} \rightarrow \mathbb{R}^d$, $\sigma : \mathbb{R}^d \times \mathbb{R} \rightarrow \mathbb{R}^{d \times d}$, $\kappa_1 : \mathbb{R}^d \times \mathbb{R}^d \rightarrow \mathbb{R}$, and $\kappa_2 : \mathbb{R}^d \times \mathbb{R}^d \rightarrow \mathbb{R}$, this work studies the following Itô SDE for the McKean–Vlasov stochastic process $X : [0, T] \times \Omega \rightarrow \mathbb{R}^d$:

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

where $W : [0, T] \times \Omega \rightarrow \mathbb{R}^d$ represents a standard d -dimensional Wiener process with mutually independent components; $\mu_t \in \mathcal{P}(\mathbb{R}^d)$ denotes the law of $X(t)$, where $\mathcal{P}(\mathbb{R}^d)$ indicates the space of probability measures in $\mathbb{R}^d$; and $x_0 \in \mathbb{R}^d$ indicates a random initial state with the distribution $\mu_0 \in \mathcal{P}(\mathbb{R}^d)$. The functions b and σ represent the drift and diffusion functions/coefficients, respectively.

Standing assumptions: Throughout, unless stated otherwise, b and σ are globally Lipschitz in both variables with at most linear growth, κ_1 and κ_2 are bounded and globally Lipschitz, μ_0 has sufficiently high finite moments for the weak and strong error estimates invoked in this work, and the Wiener process and initial state in (1) are mutually independent. These assumptions ensure the

existence and uniqueness of solutions to (1) [8]. The well-posedness of (1) under substantially weaker assumptions is known [19, 20, 21, 22], but such problems are not the focus of this work. Additional smoothness assumptions required for the convergence rates stated in Assumptions 1 and 2 are discussed in Section 4.

The Fokker–Planck equation associated with (1) is a nonlinear integro-differential partial differential equation (PDE) containing nonlocal interaction terms [23]. Numerically solving this PDE to obtain solutions to (1) with a prescribed relative accuracy is computationally infeasible. Instead, this work employs strong approximations that converge in L^p to the solution of (1).

A strong approximation is given by a system of P exchangeable Itô SDEs, also known as a stochastic interacting particle system, with pairwise interaction kernels [8]. For $p = 1, \dots, P$, the state $X_p^P : [0, T] \times \Omega \rightarrow \mathbb{R}^d$ of particle p (out of P) follows the following SDE:

$$\begin{cases} dX_p^P(t) = b \left(X_p^P(t), \frac{1}{P} \sum_{j=1}^P \kappa_1(X_p^P(t), X_j^P(t)) \right) dt \\ \quad + \sigma \left(X_p^P(t), \frac{1}{P} \sum_{j=1}^P \kappa_2(X_p^P(t), X_j^P(t)) \right) dW_p(t), \quad t > 0 \\ X_p^P(0) = (x_0)_p \sim \mu_0 \in \mathcal{P}(\mathbb{R}^d), \end{cases} \quad (2)$$

where $\{(x_0)_p\}_{p=1}^P$ denotes independent and identically distributed (i.i.d.) random variables sampled from the initial distribution μ_0 , and $\{W_p\}_{p=1}^P$ represents mutually independent d -dimensional Wiener processes that are independent of $\{(x_0)_p\}_{p=1}^P$. Equation (2) approximates the law μ_t in (1) using an empirical law based on P particles $\{X_p^P(t)\}_{p=1}^P$:

$$\mu_t(dx) \approx \mu_t^P(dx) = \frac{1}{P} \sum_{j=1}^P \delta_{X_j^P(t)}(dx). \quad (3)$$

In the limit $P \rightarrow \infty$, μ_t^P converges to μ_t (propagation of chaos) under suitable conditions [8]. Strong convergence (in the L^p -sense) of the particle system (2) to the MV-SDE (1) has been well established [24, 25, 26]. 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. Motivating example: Kuramoto oscillator model

The method in this work is assessed on the simple, one-dimensional (1D) Kuramoto model [7, 27, 28], describing synchronization in statistical physics to model the behavior of large sets of fully connected, synchronized oscillators. The phase of oscillator p (out of P) is represented by the process $X_p^P : [0, T] \times \Omega \rightarrow \mathbb{R}$ with the following Itô SDE dynamics:

$$\begin{cases} dX_p^P(t) = \left(\xi_p + \frac{1}{P} \sum_{q=1}^P \sin(X_p^P(t) - X_q^P(t)) \right) dt + \sigma dW_p(t), & t > 0 \\ X_p^P(0) = (x_0)_p \sim \mu_0 \in \mathcal{P}(\mathbb{R}), \end{cases} \quad (4)$$

where $\{\xi_p\}_{p=1}^P$ denotes i.i.d. random variables sampled from a prescribed distribution and models the natural frequency of each oscillator. The diffusion $\sigma \in \mathbb{R}$ is constant, and $\{(x_0)_p\}_{p=1}^P$ represents i.i.d. random variables sampled from $\mu_0 \in \mathcal{P}(\mathbb{R})$. In addition, $\{W_p\}_{p=1}^P$ represents mutually independent 1D Wiener processes, and $\{\xi_p\}_{p=1}^P, \{(x_0)_p\}_{p=1}^P, \{W_p\}_{p=1}^P$ are mutually independent. The coupled particle system (4) strongly converges in the mean-field limit as $P \rightarrow \infty$, where each particle satisfies the following MV-SDE:

$$\begin{cases} dX(t) = \left(\xi + \int_{\mathbb{R}} \sin(X(t) - x) \mu_t(dx) \right) dt + \sigma dW(t), & t > 0 \\ X(0) = x_0 \sim \mu_0 \in \mathcal{P}(\mathbb{R}), \end{cases} \quad (5)$$

where $X(t)$ denotes the McKean–Vlasov process at time t , ξ represents a random variable sampled from a prescribed distribution, and μ_t indicates the law of $X(t)$. In the notation of (1), this example corresponds to $b(x, y) = \xi + y$, $\sigma(x, y) = \sigma$, $\kappa_1(x, y) = \sin(x - y)$ and $\kappa_2(x, y) = 0$, satisfying the standard conditions ensuring the well-posedness of (5). This model is used as a test-bed because it is a simple, well-understood 1D MV-SDE that can be tuned by adjusting the parameters, such as the coupling kernel, and the distribution of ξ , to control the rarity of the event of interest.

2.2. Problem setting

Let $T > 0$ be a finite terminal time and $X : [0, T] \times \Omega \rightarrow \mathbb{R}^d$ denote the McKean–Vlasov process (1). Let $G : \mathbb{R}^d \rightarrow \mathbb{R}$ be a given nonnegative scalar observable that satisfies $G(x) \geq 0$ for all $x \in \mathbb{R}^d$. The objective is to build a computationally efficient MC estimator $\mathcal{A}$ of $\mathbb{E}[G(X(T))]$ that satisfies a relative error tolerance $\text{TOL}_r > 0$ for a given confidence level determined by $0 < \nu \ll 1$ in the following sense:

$$\mathbb{P} \left[\frac{|\mathcal{A} - \mathbb{E}[G(X(T))]|}{|\mathbb{E}[G(X(T))]|} \geq \text{TOL}_r \right] \leq \nu. \quad (6)$$

This work aims to compute rare events associated with the final state $X(T)$, implying that the distribution of $G(X(T))$ places most of its probability mass deep in the tail of the law of $X(T)$. The computational feasibility of the vanilla MC method rapidly diminishes due to the exploding coefficient of variation [11]. Hence, the MC method must be combined with the IS variance reduction technique to produce feasible estimates of rare events. However, applying this technique to MV-SDEs raises a considerable difficulty. The drift and diffusion functions are linked to $\{\mu_t : t \in [0, T]\}$, the law of the McKean–Vlasov process.

Under any IS measure change (e.g., via Girsanov's theorem), μ_t is no longer the law of the modified process. This challenge is overcome using the decoupling approach introduced by [12] by transforming the MV-SDE (1) into a standard SDE by conditioning it on an empirical approximation of μ_t . In summary, computing rare-event quantities for MV-SDEs requires variance reduction (to control the coefficient of variation explosion) and careful handling of mean-field coupling under measure change. Section 3 introduces the decoupling approach for MV-SDEs that addresses these problems.

3. Double-loop importance sampling for McKean–Vlasov stochastic differential equations

The decoupling approach replaces the deterministic law $\{\mu_t : t \in [0, T]\}$ in (1) with an empirical approximation (3) using the interacting particle system (2). Given this precomputed empirical law, a decoupled MV-SDE is introduced, and an IS measure change is applied to it. Hence, the empirical estimation of the law $\{\mu_t : t \in [0, T]\}$ and the IS measure change are decoupled. The decoupling approach involves the following steps [1]:

1. The law $\{\mu_t : t \in [0, T]\}$ in (1) is approximated using the empirical measure $\{\mu_t^P : t \in [0, T]\}$ in (3) using one realization of the P -particle system (2) with particles $\{X_p^P\}_{p=1}^P$.
2. Given $\{\{X_p^P(t)\}_{p=1}^P \sim \mu_t^P : t \in [0, T]\}$, this work defines the decoupled McKean–Vlasov process $\bar{X}^P : [0, T] \times \Omega \rightarrow \mathbb{R}^d$ as the solution to the following Itô SDE:

$$\begin{cases} d\bar{X}^P(t) = b\left(\bar{X}^P(t), \frac{1}{P} \sum_{j=1}^P \kappa_1(\bar{X}^P(t), X_j^P(t))\right) dt \\ \quad + \sigma\left(\bar{X}^P(t), \frac{1}{P} \sum_{j=1}^P \kappa_2(\bar{X}^P(t), X_j^P(t))\right) d\bar{W}(t), \quad t \in [0, T] \\ \bar{X}^P(0) = \bar{x}_0 \sim \mu_0, \quad \bar{x}_0 \in \mathbb{R}^d, \end{cases} \quad (7)$$

where the superscript P indicates that the drift and diffusion functions in (7) are computed using $\{\mu_t^P : t \in [0, T]\}$ derived from the stochastic P -particle system $\{X_p^P\}_{p=1}^P$. In addition, $\bar{W}$ denotes a standard d -dimensional Wiener process independent of the Wiener processes $\{W_p\}_{p=1}^P$ employed in (2), and $\bar{x}_0 \in \mathbb{R}^d$ represents a random initial state sampled from μ_0 , independent of $\{(x_0)_p\}_{p=1}^P$ in (2). Thus, (7) is a standard SDE for a given realization of the empirical law $\{\mu_t^P : t \in [0, T]\}$.

3. Let the P -particle system (2) $\{X_p^P\}_{p=1}^P : [0, T] \times \Omega \rightarrow (\mathbb{R}^d)^P$ be defined on the probability space $(\Omega, \mathcal{F}, \mathbb{P})$ with $\mathbb{E}_{\mathbb{P}}$ denoting the expectation with respect to $\mathbb{P}$. For given $\{X_p^P\}_{p=1}^P$, this work defines $\bar{X}^P$ in (7) on the copy

space $(\tilde{\Omega}, \tilde{\mathcal{F}}, \tilde{\mathbb{P}})$. Hence, $\bar{X}^P$ is defined on the product space $(\Omega, \mathcal{F}, \mathbb{P}) \times (\tilde{\Omega}, \tilde{\mathcal{F}}, \tilde{\mathbb{P}})$ with $\mathbb{E}_{\mathbb{P} \otimes \tilde{\mathbb{P}}}$ denoting the expectation with respect to the product measure space $\mathbb{P} \otimes \tilde{\mathbb{P}}$.

4. The quantity of interest is approximated as the following nested expectation:

$$\begin{aligned} \mathbb{E}[G(X(T))] &\approx \mathbb{E}_{\mathbb{P} \otimes \tilde{\mathbb{P}}} [G(\bar{X}^P(T))] \\ &= \mathbb{E}_{\mathbb{P}} [\mathbb{E}_{\tilde{\mathbb{P}}} [G(\bar{X}^P(T)) \mid \{\mu_t^P : t \in [0, T]\}]] . \end{aligned} \quad (8)$$

Henceforth, $\mathbb{E}[G(\bar{X}^P(T))] \equiv \mathbb{E}_{\mathbb{P} \otimes \tilde{\mathbb{P}}} [G(\bar{X}^P(T))]$ for convenience.

As (7) is a standard SDE for a given empirical law $\{\mu_t^P : t \in [0, T]\}$, this work follows the procedure from [1] to derive the optimal IS measure change.

3.1. Importance sampling using stochastic optimal control for decoupled McKean–Vlasov stochastic differential equation

The PDE that yields the optimal control for the decoupled MV-SDE for a given empirical law $\{\mu_t^P : t \in [0, T]\}$ was formulated by [1]. This work defines the controlled process $\bar{X}_\zeta^P : [0, T] \times \Omega \rightarrow \mathbb{R}^d$ following the given controlled dynamics with the Markov control function $\zeta : [0, T] \times \mathbb{R}^d \rightarrow \mathbb{R}^d$:

$$\left\{ \begin{aligned} d\bar{X}_\zeta^P(t) &= \left(b\left(\bar{X}_\zeta^P(t), \frac{1}{P} \sum_{j=1}^P \kappa_1(\bar{X}_\zeta^P(t), X_j^P(t))\right) \right. \\ &\quad \left. + \sigma\left(\bar{X}_\zeta^P(t), \frac{1}{P} \sum_{j=1}^P \kappa_2(\bar{X}_\zeta^P(t), X_j^P(t))\right) \zeta(t, \bar{X}_\zeta^P(t)) \right) dt \\ &\quad + \sigma\left(\bar{X}_\zeta^P(t), \frac{1}{P} \sum_{j=1}^P \kappa_2(\bar{X}_\zeta^P(t), X_j^P(t))\right) dW(t), 0 < t < T \\ \bar{X}_\zeta^P(0) &= \bar{X}^P(0) = \bar{x}_0 \sim \mu_0. \end{aligned} \right. \quad (9)$$

This work applies (2) to obtain $\{\{X_p^P(t)\}_{p=1}^P \sim \mu_t^P : t \in [0, T]\}$ in (9). This work assumes that ζ satisfies Novikov's condition:

$$\mathbb{E} \left[\exp \left\{ \frac{1}{2} \int_0^T \|\zeta(s, \bar{X}_\zeta^P(s))\|^2 ds \right\} \mid \{\mu_t^P : t \in [0, T]\} \right] < \infty. \quad (10)$$

The value function $u : [0, T] \times \mathbb{R}^d \rightarrow \mathbb{R}$ that minimizes the second moment of the MC estimator of $\mathbb{E}[G(\bar{X}^P(T)) \mid \{\mu_t^P : t \in [0, T]\}]$ is defined as follows:

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

$$\left. -2 \int_t^T \langle \zeta(s, \bar{X}_\zeta^P(s)), dW(s) \rangle \right\} \Big| \bar{X}_\zeta^P(t) = x, \{\mu_t^P : t \in [0, T]\} \Big]. \quad (11)$$

$\mathcal{Z}$ represents the class of deterministic Markov feedback functions $\zeta : [0, T] \times \mathbb{R}^d \rightarrow \mathbb{R}^d$ that are sufficiently regular to ensure well-posedness of (9), satisfy Novikov's condition (10) conditional on the empirical law, and make the second moment in (11) finite. This work defines $v : [0, T] \times \mathbb{R}^d \rightarrow \mathbb{R}$ such that $u(t, x) = v^2(t, x)$. Then, v satisfies the following linear backward PDE:

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

with the following optimal control:

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

which minimizes the second moment in (11), for a given $\{\{X_p^P(t)\}_{p=1}^P \sim \mu_t^P : t \in [0, T]\}$. The corresponding proof is available in Appendix B in the work by [1]. When G vanishes for the indicator function at $t = T$, $\nabla \log v$ is interpreted only on $[0, T)$ where the parabolic smoothing of (12) yields $v > 0$ under non-degeneracy assumptions. In numerical experiments, we evaluate the control ζ^* only at time nodes less than T . The IS control ζ^* , which depends on the time and state, tilts the paths of (9) towards the rare-event domain and reduces the variance of the MC estimator. The likelihood resulting from the IS measure change is written as the following Radon–Nikodym derivative:

$$\mathbb{L}_{[0, T]}^P = \exp \left\{ - \int_0^T \langle \zeta(s, \bar{X}_\zeta^P(s)), dW(s) \rangle - \frac{1}{2} \int_0^T \|\zeta(s, \bar{X}_\zeta^P(s))\|^2 ds \right\}. \quad (14)$$

With this measure change, the nested expectation (8) is rewritten as follows:

$$\begin{aligned} \mathbb{E}[G(X(T))] &\approx \mathbb{E}_{\mathbb{P}} [\mathbb{E}_{\tilde{\mathbb{P}}} [G(\bar{X}^P(T)) \mid \{\mu_t^P : t \in [0, T]\}]] \\ &= \mathbb{E}_{\mathbb{P}} [\mathbb{E}_{\tilde{\mathbb{Q}}} [G(\bar{X}_\zeta^P(T)) \mathbb{L}_{[0, T]}^P \mid \{\mu_t^P : t \in [0, T]\}]], \end{aligned} \quad (15)$$

where $\tilde{\mathbb{Q}}$ denotes the shifted probability measure of the copy space, such that (9) is defined on $(\Omega, \mathcal{F}, \mathbb{P}) \times (\tilde{\Omega}, \tilde{\mathcal{F}}, \tilde{\mathbb{Q}})$. The nested expectation (15) is approximated using a DLMC estimator.

3.2. Double loop Monte Carlo estimator with importance sampling

The DLMC estimator for the nested expectation (15) is constructed as follows.

1. The empirical law $\{\mu_t^P : t \in [0, T]\}$ approximates μ_t in (1) using (2). In practice, a time-discretized empirical law with N time steps is obtained from the Euler–Maruyama discretization $\{X_p^{P|N}\}_{p=1}^P$ of the particle system. The Euler–Maruyama continuous-time extension of the time-discretized stochastic particle system $\{X_p^{P|N}\}_{p=1}^P$ is applied to extend the time-discrete empirical law to all $t \in [0, T]$. Then, the empirical law is defined as follows:

$$\mu^{P|N}(t) = \frac{1}{P} \sum_{j=1}^P \delta_{X_j^{P|N}(t)}, \quad \forall t \in [0, T]. \quad (16)$$

The P underlying sets of random variables employed to generate a realization of $\{X_p^{P|N}\}_{p=1}^P$ using (2) are denoted by $\omega_{1:P}$.

2. Given $\mu^{P|N}$ from (16), (12) is solved and (13) is employed to obtain the optimal Markov control function ζ for IS.
3. Given $\mu^{P|N}$ from (16) and the control function ζ , sample paths of the controlled decoupled MV-SDE (9) are generated. The Euler–Maruyama discretization of SDE (9) with N time steps is applied to obtain the approximate sample paths denoted by $\{\bar{X}_\zeta^{P|N}(t_n)\}_{n=1}^N$. The set of random variables employed to generate a sample path is denoted by $\tilde{\omega}$.
4. The DLMC estimator $\mathcal{A}_{\text{MC}}$ approximates the nested expectation (15) using the following nested sample average:

$$\mathcal{A}_{\text{MC}} = \frac{1}{M_1} \sum_{i=1}^{M_1} \frac{1}{M_2} \sum_{j=1}^{M_2} G\left(\bar{X}_\zeta^{P|N}(T)\right) \mathbb{L}^{P|N}\left(\omega_{1:P}^{(i)} \times \tilde{\omega}^{(i,j)}\right), \quad (17)$$

where $\mathbb{L}^{P|N}$ denotes the time-discretized approximation of the likelihood (14) defined as follows:

$$\mathbb{L}^{P|N} = \prod_{n=0}^{N-1} \exp \left\{ -\frac{1}{2} \Delta t \left\| \zeta(t_n, \bar{X}_\zeta^{P|N}(t_n)) \right\|^2 - \langle \Delta W(t_n), \zeta(t_n, \bar{X}_\zeta^{P|N}(t_n)) \rangle \right\}, \quad (18)$$

where M_1 denotes the number of independent realizations of $\mu^{P|N}$ and $\omega_{1:P}^{(i)}$ denotes the i^{th} realization of $\omega_{1:P}$. For each realization of $\mu^{P|N}$, M_2 represents the number of sample paths of the decoupled MV-SDE, and $\tilde{\omega}^{(i,j)}$ denotes the j^{th} realization of $\tilde{\omega}$ given $\omega_{1:P}^{(i)}$. Further, $\Delta W(t_n)$ represents the Wiener increment driving the dynamics of the time-discretized decoupled MV-SDE (9).

Algorithm 2 in [Appendix A](#) summarises the above approach as pseudocode. In the ideal double loop formulation, the PDE (12) should be numerically solved for each realisation of $\mu^{P|N}$. In Algorithm 2, we instead compute a fixed off-line control ζ from one sufficiently fine empirical law $\mu^{P|\bar{N}}$ and reuse the same admissible control for all outer loop samples. This does not affect the bias of the IS estimator for all discretised decoupled MV-SDE (9) simulations, but it makes the control suboptimal relative to the realisation-dependent optimiser in (13). The quality of this approximation is assessed numerically in Section 5. Under standard assumptions, this DLMC estimator with IS (17) achieves $\mathcal{O}(\text{TOL}_r^{-4})$ complexity for a relative error tolerance TOL_r [1], which is a significant improvement over the standard MC method that is infeasible. Although the DLMC estimator mitigates the explosive growth of the coefficient of variation and achieves the same complexity order as the standard MC method for nonrare observables, further improvements are possible. In particular, the DLMC method still samples all trajectories at a fixed high resolution (P, N) . In the work by [13], the DLMC estimator was extended to the multilevel MC setting. This extension allows for estimating the expectations of Lipschitz rare-event observables up to a relative tolerance of TOL_r while reducing the complexity to $\mathcal{O}(\text{TOL}_r^{-3})$ for the Kuramoto example (4). The presence of two discretization parameters, P and N , in MV-SDEs motivates the extension of this estimator to an MIMC setting to further reduce the complexity.

4. Multi-index double loop Monte Carlo method

This section aims to combine the multi-index technique with the stochastic optimal control-based IS scheme for rare events introduced in Section 3.1. This approach yields the novel multi-index DLMC estimator. Following the work by [14], this work introduces the following multi-index DLMC discretization. Two discretization parameters (P, N) are required to generate sample paths of the decoupled MV-SDE (9). The multi-index $\alpha = (\alpha_1, \alpha_2) \in \mathbb{N}^2$ defines the discretization parameters as follows:

$$P_{\alpha_1} = P_0 \times 2^{\alpha_1}, \quad N_{\alpha_2} = N_0 \times 2^{\alpha_2}, \quad (19)$$

where P_0 and N_0 represent the minimum number of particles and time steps, respectively, to generate the approximate sample paths of the decoupled MV-SDE (9). This work proposes the following method to couple IS with the MIMC estimator. The optimal control PDE (12) is solved using one realization of the stochastic particle system (2) with a large number of particles $\bar{P}$ and time steps $\bar{N}$, obtaining the approximate IS control function ζ using (13). The same policy ζ is applied to generate samples of the decoupled MV-SDE (9) at all discretization multi-indices α . Let $G_\alpha = G(\bar{X}_\zeta^{P_{\alpha_1}|N_{\alpha_2}}(T))\mathbb{L}^{P_{\alpha_1}|N_{\alpha_2}}$, where $\bar{X}_\zeta^{P_{\alpha_1}|N_{\alpha_2}}$ denotes the Euler–Maruyama discretized controlled decoupled MV-SDE process (9) conditioned on the empirical law $\mu^{P_{\alpha_1}|N_{\alpha_2}}$ (16). The discretized likelihood factor $\mathbb{L}^{P_{\alpha_1}|N_{\alpha_2}}$ is computed using (18). Following the work by [14], this

work defines the first-order mixed-difference operator in two dimensions for this setting:

$$\Delta^{\text{mix}} G_{\alpha} = (G_{(\alpha_1, \alpha_2)} - G_{(\alpha_1-1, \alpha_2)}) - (G_{(\alpha_1, \alpha_2-1)} - G_{(\alpha_1-1, \alpha_2-1)}). \quad (20)$$

Assuming that G_{α} converges to $G(X(T))$ (in L^1 -sense) as $\min(\alpha_1, \alpha_2) \rightarrow \infty$, the multi-index telescoping identity gives

$$\mathbb{E}[G(X(T))] = \lim_{L_1, L_2 \rightarrow \infty} \sum_{\alpha_1=0}^{L_1} \sum_{\alpha_2=0}^{L_2} \mathbb{E}[\Delta^{\text{mix}} G_{\alpha}] = \sum_{\alpha \in \mathbb{N}^2} \mathbb{E}[\Delta^{\text{mix}} G_{\alpha}], \quad (21)$$

where $G_{(-1,0)} = 0$, $G_{(0,-1)} = 0$, and $G_{(-1,-1)} = 0$. Let $\hat{\Delta} G_{\alpha}$ be a random variable such that $\mathbb{E}[\hat{\Delta} G_{\alpha}] = \mathbb{E}[\Delta^{\text{mix}} G_{\alpha}]$. In the trivial case, $\hat{\Delta} G_{\alpha} = \Delta^{\text{mix}} G_{\alpha}$. Alternatively, $\hat{\Delta} G_{\alpha}$ can be chosen such that $\text{Var}[\hat{\Delta} G_{\alpha}] \ll \text{Var}[\Delta^{\text{mix}} G_{\alpha}]$. Each expectation in (21) is approximated using a DLMC estimator, resulting in the following multi-index DLMC estimator:

$$\mathcal{A}_{\text{MIMC}}(\mathcal{I}) = \sum_{\alpha \in \mathcal{I}} \frac{1}{M_{1,\alpha}} \sum_{m_1=1}^{M_{1,\alpha}} \frac{1}{M_{2,\alpha}} \sum_{m_2=1}^{M_{2,\alpha}} \hat{\Delta} G_{\alpha} \left(\omega_{1:P_{\alpha_1}}^{(\alpha, m_1)} \times \tilde{\omega}^{(\alpha, m_1, m_2)} \right), \quad (22)$$

where $\mathcal{I} \subset \mathbb{N}^2$ represents an appropriately chosen index set, and $\{M_{1,\alpha}, M_{2,\alpha}\}$ denotes the integer numbers of samples in the inner and outer loops, respectively, of the DLMC estimator for each $\alpha \in \mathcal{I}$. Assumption 1 is used to control the truncation bias for the admissible index set $\mathcal{I}$. To control the variance of the mixed-differences, G_{α} must be tightly coupled with the three lower-index terms in (20). Based on the antithetic estimators in the studies by [9], [29], and [13], this coupling is achieved via an antithetic estimator $\hat{\Delta} G_{\alpha}$ that reuses a large portion of the random variables (particles and Wiener paths) between multi-indices α and their neighbors. The estimator $\hat{\Delta} G_{\alpha}$ is defined as follows:

$$\begin{aligned} \hat{\Delta} G_{\alpha} \left(\omega_{1:P_{\alpha_1}}^{(\alpha, m_1)} \times \tilde{\omega}^{(\alpha, m_1, m_2)} \right) = & \left((G_{(\alpha_1, \alpha_2)} - \mathcal{G}_{(\alpha_1-1, \alpha_2)}) \right. \\ & \left. - (G_{(\alpha_1, \alpha_2-1)} - \mathcal{G}_{(\alpha_1-1, \alpha_2-1)}) \right) \left(\omega_{1:P_{\alpha_1}}^{(\alpha, m_1)} \times \tilde{\omega}^{(\alpha, m_1, m_2)} \right), \end{aligned} \quad (23)$$

where $\mathcal{G}_{(\alpha_1-1, \alpha_2)}$ is highly correlated with $G_{(\alpha_1, \alpha_2)}$, which is defined as follows:

$$\mathcal{G}_{(\alpha_1-1, \alpha_2)} \left(\omega_{1:P_{\alpha_1}}^{(\alpha, m_1)} \times \tilde{\omega}^{(\alpha, m_1, m_2)} \right) = \frac{1}{2} \left(G_{(\alpha_1-1, \alpha_2)} \left(\omega_{1:P_{\alpha_1-1}}^{(\alpha, m_1)} \times \tilde{\omega}^{(\alpha, m_1, m_2)} \right) \right) \quad (24)$$

$$+ G_{(\alpha_1-1, \alpha_2)} \left(\omega_{P_{\alpha_1-1}+1:P_{\alpha_1}}^{(\alpha, m_1)} \times \tilde{\omega}^{(\alpha, m_1, m_2)} \right).$$

In (24), the P_{α_1} sets of random variables are split into two i.i.d. groups of size P_{α_1-1} and each group is employed to generate a realization of the empirical law independently. Given each realization of the empirical law, approximate sample paths of $\bar{X}_{\zeta}^{P_{\alpha_1-1}|N_{\alpha_2}}$ are generated using the same $\tilde{\omega}$ as for $G_{(\alpha_1, \alpha_2)}$ before averaging the quantity of interest over the two groups. Intuitively, six simulations in (23) are coupled at different discretizations in a carefully arranged way to yield a small variance.

4.1. Error analysis

The relative error of the multi-index DLMC estimator $\mathcal{A}_{\text{MIMC}}$ is bounded as follows:

$$\frac{|\mathbb{E}[G(X(T))] - \mathcal{A}_{\text{MIMC}}|}{|\mathbb{E}[G(X(T))]|} \leq \underbrace{\frac{|\mathbb{E}[G(X(T))] - \mathbb{E}[\mathcal{A}_{\text{MIMC}}]|}{|\mathbb{E}[G(X(T))]|}}_{=\epsilon_b, \text{ Relative bias}} + \underbrace{\frac{|\mathbb{E}[\mathcal{A}_{\text{MIMC}}] - \mathcal{A}_{\text{MIMC}}|}{|\mathbb{E}[G(X(T))]|}}_{=\epsilon_s, \text{ Relative statistical error}}. \quad (25)$$

Instead of satisfying a prescribed relative error tolerance TOL_r in the sense of (6), the accuracy is split between the relative bias and statistical errors using parameter $\theta \in (0, 1)$ and imposing the following stricter constraints for a confidence level determined by $0 < \nu \ll 1$:

$$\epsilon_b \leq (1 - \theta)\text{TOL}_r, \quad (26)$$

$$\mathbb{P}[\epsilon_s \geq \theta\text{TOL}_r] \leq \nu. \quad (27)$$

Throughout this work, θ is assumed to be given and fixed. This work makes the following assumptions regarding the first-order mixed-differences.

Assumption 1. (*Multiplicative bias decay*) For all $\alpha \in \mathcal{I}$, for constants $Q_B > 0$ and $b_1, b_2 > 0$, the absolute value of the expected value of $\hat{\Delta}G_{\alpha}$ satisfies the following:

$$\left| \mathbb{E}[\hat{\Delta}G_{\alpha}] \right| \leq Q_B P_{\alpha_1}^{-b_1} N_{\alpha_2}^{-b_2}. \quad (28)$$

Assumption 1 is motivated by the standard weak error expansion of the Euler–Maruyama time-discretization scheme [30] with respect to the number of time steps N , and the weak convergence of the interacting particle system approximation (2) with respect to the number of particles P [31, 32, 29, 33]. Based on [30], $b_2 = 1$ is expected for the Euler–Maruyama scheme. A recent analysis by [34] established a $\mathcal{O}(P^{-1})$ weak convergence rate of the particle system (2) to the MV-SDE (1), indicating that $b_1 = 1$ under stronger conditions on $b, \sigma, \kappa_1, \kappa_2, G$ than those required for well-posedness. However, the first-order mixed-difference behaves as a product of one-direction errors [14] only when the solution map $(P_{\alpha_1}, N_{\alpha_2}) \mapsto G_{\alpha}$ has mixed regularity with respect to P and N .

This regularity ensures that $\left| \mathbb{E} \left[\hat{\Delta} G_{\alpha} \right] \right|$ approximately decays as a product of the geometric factors in each direction (P and N), which is the mechanism behind sparse-grid/combination technique error bounds [35]. These rates can be numerically verified for a given problem. Section 5 numerically investigates how the regularity of G and κ_1 affects Assumption 1 for the Kuramoto model (4). A full theoretical analysis of the conditions on $b, \sigma, \kappa_1, \kappa_2$ and G ensuring Assumption 1 is the subject of ongoing work.

The index set $\mathcal{I} \subset \mathbb{N}^2$ is selected to satisfy the bias constraint in (26). The statistical error constraint in (27) is expressed as the following constraint on the variance of the estimator using the asymptotic normality of the multi-index estimator [14]:

$$\frac{\text{Var} [\mathcal{A}_{\text{MIMC}}(\mathcal{I})]}{|\mathbb{E} [G(X(T))]|^2} \leq \left(\frac{\theta \text{TOL}_r}{C_{\nu}} \right)^2, \quad (29)$$

where C_{ν} denotes the $(1 - \frac{\nu}{2})$ -quantile of the standard normal distribution. Due to the independence of the DLMC estimators corresponding to each α in (22), the variance of the multi-index DLMC estimator is written as follows:

$$\text{Var} [\mathcal{A}_{\text{MIMC}}(\mathcal{I})] = \sum_{\alpha \in \mathcal{I}} \frac{1}{M_{1,\alpha}} \text{Var} \left[\frac{1}{M_{2,\alpha}} \sum_{m_2=1}^{M_{2,\alpha}} \hat{\Delta} G_{\alpha} \left(\omega_{1:P_{\alpha_1}}^{(\alpha, \cdot)} \times \tilde{\omega}^{(\alpha, \cdot, m_2)} \right) \right]. \quad (30)$$

Owing to the law of total variance,

$$\begin{aligned} \text{Var} [\mathcal{A}_{\text{MIMC}}(\mathcal{I})] &= \sum_{\alpha \in \mathcal{I}} \frac{1}{M_{1,\alpha}} \left(\underbrace{\text{Var} \left[\mathbb{E} \left[\hat{\Delta} G_{\alpha} \mid \omega_{1:P_{\alpha_1}}^{(\alpha, \cdot)} \right] \right]}_{=V_{1,\alpha}} \right. \\ &\quad \left. + \frac{1}{M_{2,\alpha}} \mathbb{E} \left[\underbrace{\text{Var} \left[\hat{\Delta} G_{\alpha} \mid \omega_{1:P_{\alpha_1}}^{(\alpha, \cdot)} \right]}_{=V_{2,\alpha}} \right] \right) \\ &= \sum_{\alpha \in \mathcal{I}} \left(\frac{V_{1,\alpha}}{M_{1,\alpha}} + \frac{V_{2,\alpha}}{M_{1,\alpha} M_{2,\alpha}} \right), \end{aligned} \quad (31)$$

where $V_{1,\alpha}$ and $V_{2,\alpha}$ are conditioned on the discretized empirical law $\mu^{P_{\alpha_1} | N_{\alpha_2}}$, described by the set of random variables $\omega_{1:P_{\alpha_1}}^{(\alpha, \cdot)}$.

Assumption 2. (*Multiplicative variance decay*) For all $\alpha \in \mathcal{I}$, for the constants $w_i > 0$ and $s_i > 0$ for $i = 1, 2$, the variance terms $V_{1,\alpha}, V_{2,\alpha}$ defined in (31) satisfy

$$V_{1,\alpha} = \text{Var} \left[\mathbb{E} \left[\hat{\Delta} G_{\alpha} \mid \omega_{1:P_{\alpha_1}}^{(\alpha, \cdot)} \right] \right] \lesssim P_{\alpha_1}^{-w_1} N_{\alpha_2}^{-w_2}, \quad (32)$$

$$V_{2,\alpha} = \mathbb{E} \left[\text{Var} \left[\hat{\Delta} G_{\alpha} \mid \omega_{1:P_{\alpha_1}}^{(\alpha, \cdot)} \right] \right] \lesssim P_{\alpha_1}^{-s_1} N_{\alpha_2}^{-s_2}. \quad (33)$$

Assumption 2 is motivated by the propagation of chaos [8] results for MV-SDEs and the strongly coupled antithetic estimator developed in (23). The rates $w_2 = 2$ and $s_2 = 1$ are expected from the standard weak [30] and strong [36] convergence results of the Euler–Maruyama scheme. Under the sufficient regularity of $b, \sigma, \kappa_1, \kappa_2$, and G and antithetic coupling (24), the weak [34] and strong [29] convergence analysis of the particle system (2) indicate that $w_1 = 2$ and $s_1 = 2$. Under further mixed regularity, this coupling in the mixed-differences ensures that the particle approximation and time discretization errors combine multiplicatively in the variance bounds. With such a coupling scheme, the leading-order increments cancel in the mixed-difference so that $V_{1,\alpha}$ and $V_{2,\alpha}$ decay faster than the multilevel variances in the work by [13]. These rates can be numerically checked for a given problem (see Section 5 for the Kuramoto model). Section 5 numerically examines how the regularity of G and κ_1 influences Assumption 2 for the Kuramoto model (4). A rigorous proof of these mixed convergence rates would require combining the propagation of chaos results with multi-index coupling arguments in (23) and is left for future work.

4.2. Computational complexity

For each multi-index $\alpha \in \mathbb{N}^2$, $\mathbb{E}[\hat{\Delta}G_\alpha]$ is estimated using six evaluations of G_α for various combinations of discretization parameters (23). The dominant computational contribution corresponds to the multi-index $\alpha = (\alpha_1, \alpha_2)$. This work assumes a naive $\mathcal{O}(P)$ implementation to evaluate the empirical means in the drift and diffusion functions in (2) and (9). The computational cost to generate one realization of G_α comprises two parts: (i) generating a realization of the empirical law $\mu^{P_{\alpha_1}|N_{\alpha_2}}$ using the Euler–Maruyama scheme with N_{α_2} time steps, requiring $\mathcal{O}(N_{\alpha_2}P_{\alpha_1}^2)$ operations, and (ii) generating a sample path of the decoupled MV-SDE (9) using the Euler–Maruyama scheme with N_{α_2} time steps, requiring $\mathcal{O}(N_{\alpha_2}P_{\alpha_1})$ operations. The cost of computing the IS control ζ by numerically solving (12) is incurred only once, and the obtained IS control is employed uniformly across all multi-indices α ; hence, it is treated as an off-line preprocessing step. Thus, the total computational cost of the multi-index DLMC estimator can be bound as follows:

$$\mathcal{W}[\mathcal{A}_{\text{MIMC}}(\mathcal{I})] \lesssim \sum_{\alpha \in \mathcal{I}} M_{1,\alpha} N_{\alpha_2} P_{\alpha_1}^2 + M_{1,\alpha} M_{2,\alpha} N_{\alpha_2} P_{\alpha_1}. \quad (34)$$

Following the work by [14], the optimal index set for every value of $L > 0$ under Assumptions 1 and 2 can be written as follows:

$$\mathcal{I}(L) = \{\alpha \in \mathbb{N}^2 : (1 - \bar{s}_1 + 2b_1)\alpha_1 + (1 - \bar{s}_2 + 2b_2)\alpha_2 \leq L\}, \quad (35)$$

where $\bar{s}_1 = \min(w_1 - 1, s_1)$ and $\bar{s}_2 = \min(w_2, s_2)$. The following theorem presents the optimal computational complexity of the multi-index DLMC estimator with the above optimal index set (35).

Theorem 1. (*Optimal multi-index DLMC complexity*) Let $\theta \in (0, 1)$ and $\nu \in (0, 1)$ in (26) and (29) be fixed. Let Assumptions 1 and 2 hold, along with

the standard well-posedness assumptions for (1). Let $b_1, b_2, w_1, w_2, s_1, s_2 > 0$ be the constants in Assumptions 1 and 2. Let $\bar{L} \equiv \bar{L}(\text{TOL}_r)$ be chosen such that the bias constraint (26) holds for the optimal multi-index set $\mathcal{I}(\bar{L})$ defined in (35). Let $\bar{s}_1 = \min(w_1 - 1, s_1)$ and $\bar{s}_2 = \min(w_2, s_2)$. If $2b_1 \geq \bar{s}_1 - 1$ and $2b_2 \geq \bar{s}_2 - 1$, then the total computational cost of the multi-index DLMC estimator (22) $\mathcal{W}[\mathcal{A}_{\text{MIMC}}(\mathcal{I}(\bar{L}))]$, subject to the variance constraint (29), satisfies

$$\limsup_{\text{TOL}_r \downarrow 0} \frac{\mathcal{W}[\mathcal{A}_{\text{MIMC}}(\mathcal{I}(\bar{L}))]}{\text{TOL}_r^{-2-2\max(0,\varsigma)} (\log \text{TOL}_r^{-1})^\varrho} \leq C_{\text{work}} < \infty, \quad (36)$$

where,

$$\varsigma = \max\left(\frac{1 - \bar{s}_1}{2b_1}, \frac{1 - \bar{s}_2}{2b_2}\right), \quad \xi = \min(2b_1 - \bar{s}_1, 2b_2 - \bar{s}_2),$$

$$\varrho = \begin{cases} 0, & \varsigma < 0 \\ 2\aleph, & \varsigma = 0 \\ 1 + 2(\aleph - 1)(1 + \varsigma), & \varsigma > 0 \text{ and } \xi = 0, \\ 2(\aleph - 1)(1 + \varsigma), & \varsigma > 0 \text{ and } \xi > 0. \end{cases}, \quad \aleph = \begin{cases} 1, & \frac{1 - \bar{s}_1}{2b_1} \neq \frac{1 - \bar{s}_2}{2b_2}, \\ 2, & \frac{1 - \bar{s}_1}{2b_1} = \frac{1 - \bar{s}_2}{2b_2}. \end{cases}$$

Proof. See [Appendix B](#). □

For the Kuramoto model (4) with the C^∞ -mollified indicator observable $G(x) = \frac{1}{2}(1 + \tanh(3(x - K)))$, Section 5 numerically demonstrates that $b_1 = b_2 = 1$, $w_1 = w_2 = s_1 = 2$, and $s_2 = 1.5$. Using Theorem 1 with these values indicates that the complexity rate of the multi-index DLMC estimator is $\mathcal{O}(\text{TOL}_r^{-2} (\log \text{TOL}_r^{-1})^2)$ compared with that of $\mathcal{O}(\text{TOL}_r^{-3})$ for the multilevel DLMC estimator [13], as confirmed in Figure 5. This complexity matches the optimal rate predicted by multi-index theory for these convergence rates. Achieving a better asymptotic rate would likely require either higher-order methods or additional variance reduction beyond IS, which is left for future research.

Remark 1. (*On isotropic directions*) To demonstrate the superiority of the multi-index DLMC estimator over the multilevel DLMC estimator from the work by [13], this work examines the case in which $w_1 = w_2 = s_1 = s_2 = s > 0$ and $b_1 = b_2 = b > 0$. The expectations and variances of the mixed-differences converge at the same rate in the P and N directions. One such case is the Kuramoto setting (4) with the smooth, nonrare observable $G(x) = \cos(x)$. Theorem 1, as $\text{TOL}_r \rightarrow 0$, yields the following result:

$$\mathcal{W}[\mathcal{A}_{\text{MIMC}}] = \begin{cases} \mathcal{O}(\text{TOL}_r^{-2}), & s > 2 \\ \mathcal{O}(\text{TOL}_r^{-2} (\log \text{TOL}_r^{-1})^2), & s = 2. \\ \mathcal{O}(\text{TOL}_r^{-2 - \frac{2-s}{b}}), & s < 2 \end{cases} \quad (37)$$

In contrast, the multilevel DLMC estimator [13] has the following complexity

rates:

$$\mathcal{W}[\mathcal{A}_{\text{MLMC}}] = \begin{cases} \mathcal{O}(\text{TOL}_r^{-2}), & s > 3 \\ \mathcal{O}(\text{TOL}_r^{-2} (\log \text{TOL}_r^{-1})^2), & s = 3 \\ \mathcal{O}(\text{TOL}_r^{-2 - \frac{3-s}{b}}), & s < 3 \end{cases}. \quad (38)$$

Comparing (37) and (38), the multi-index DLMC estimator has the same or better complexity rates than the multilevel DLMC estimator in all regimes of the variance convergence rate s . However, the multi-index DLMC estimator requires mixed regularity (in the sense of Assumptions 1 and 2) with respect to P and N , whereas the multilevel DLMC estimator entails only ordinary regularity.

Remark 2. (*Computing probabilities*) The multi-index DLMC estimator (22) requires observables G with sufficient mixed regularity in P and N (refer to Assumptions 1 and 2). An extension to rare-event probabilities of the form $\alpha_K = \mathbb{P}[H(X(T)) > K]$, with $H : \mathbb{R}^d \rightarrow \mathbb{R}$ and K in the tail of $H(X(T))$, corresponds to the discontinuous indicator observable $G(x) = \mathbb{1}_{\{H(x) > K\}}$. The lack of regularity results in a poor expectation and variance decay of mixed-differences and suboptimal complexity (Section 5). A solution is to replace the indicator with a mollifier $\tilde{\mathbb{1}}_{\{H(x) > K\}}^\varepsilon \in C^\infty(\mathbb{R})$, where ε denotes the smoothing parameter. If $H(X(T))$ admits a density that is sufficiently smooth in a neighbourhood of K , and if the mollification kernel is symmetric in the sense of [37], then the absolute mollification bias satisfies

$$\left| \mathbb{E} \left[\mathbb{1}_{\{H(X(T)) > K\}} - \tilde{\mathbb{1}}_{\{H(X(T)) > K\}}^\varepsilon \right] \right| = \mathcal{O}(\varepsilon^2).$$

Thus, setting $\varepsilon = \mathcal{O}((\alpha_K \text{TOL}_r)^{0.5})$ achieves a relative error tolerance TOL_r [37, 38] and restores the complexity rates of Theorem 1. In practice, decreasing ε requires an adaptive time-stepping method [39] to maintain mixed-difference convergence rates because the mollified observable becomes increasingly steep as $\text{TOL}_r \rightarrow 0$. The systematic development of such mollification-adaptivity strategies is the subject of ongoing work.

Remark 3. (*High-dimensional problems*) For fixed dimension d of the MV-SDE, Theorem 1 gives a complexity exponent in TOL_r that is independent of d . The constants and the per-sample cost, however, do depend on d , and the off-line cost of computing the IS control ζ by numerically solving the PDE (12) scales exponentially with d using standard finite difference/element solvers. For higher-dimensional problems ($d \gg 1$), this step may become the dominant computational bottleneck. In such settings, the total computational cost should be viewed as $\mathcal{W}_{\text{PDE}} + \mathcal{W}[\mathcal{A}_{\text{MIMC}}]$, where $\mathcal{W}_{\text{PDE}}$ denotes the cost of numerically constructing the IS control ζ . There is an inherent trade-off between these two contributions. Increasing the accuracy of the PDE solve requires finer discretisations and thus increases $\mathcal{W}_{\text{PDE}}$. However, the resulting control is closer to the exact IS control (13), leading to a smaller estimator variance and consequently less multi-index DLMC samples required to achieve a prescribed relative tolerance. Conversely, a less accurate PDE solve reduces the offline cost but typi-

cally increases the online sampling cost. Developing a framework that jointly optimizes the PDE accuracy in high dimensions and the multi-index DLMC sampling parameters, similar to the work in [40], is left for future research. To mitigate the curse of dimensionality, one may employ model reduction [41], variational formulation [42], sparse tensor techniques [43], or machine learning approaches [44]. For the 1D Kuramoto model (4) considered in this work, we observe that $\mathcal{W}_{\text{PDE}} \ll \mathcal{W}[\mathcal{A}_{\text{MIMC}}]$ over the range of prescribed relative error tolerances, and can hence treat the PDE solve as a preprocessing step.

4.3. Adaptive algorithm

Algorithm 1 presents the adaptive multi-index DLMC algorithm employed to construct the optimal index set $\mathcal{I}$ (35) sequentially and determine the optimal number of samples $\{M_{1,\alpha}, M_{2,\alpha}\}_{\alpha \in \mathcal{I}}$ to satisfy the relative error constraints in (26) and (27). This algorithm is adapted from the MIMC algorithm in the work by [14]. The algorithm employs pilot computations to estimate the convergence rates in Assumptions 1 and 2 and allocates computational effort optimally across multi-indices via the optimal index set (35). The adaptive strategy starts with a small index set (corresponding to $L = 1$ in (35)), estimates the bias and variance contributions on those indices, and expands the index set until the bias criterion (26) is met. At each expansion, this algorithm also computes the optimal number of samples per multi-index according to (B.3). Therefore, the resulting computational cost of the adaptive algorithm scales as predicted by Theorem 1. The algorithm relies on accurate, low-cost estimates of the relative bias and variances $\{V_{1,\alpha}, V_{2,\alpha}\}_{\alpha \in \mathcal{I}}$. Small auxiliary simulations (Algorithms 3-5 in Appendix C) estimate the bias/variances to guide adaptivity. The cost of these estimations is negligible compared with the total computational work as $\text{TOL}_r \rightarrow 0$ because these estimations are performed with fixed sample sizes and interpolation techniques. Another critical feature of Algorithm 1 is that the IS control ζ is computed once off-line. A single realization of the empirical law $\mu^{\bar{P}|\bar{N}}$ is generated for large $\bar{P}$ and $\bar{N}$, after which the control PDE in (12), given $\mu^{\bar{P}|\bar{N}}$, is solved numerically to obtain ζ . Appendix C presents further details.

5. Numerical results

This section presents numerical evidence supporting the assumptions and complexity rates derived in Section 4 and numerically verifies the variance reduction achieved using the IS scheme. All experiments¹ were performed on the Kuramoto model (4) with the following parameters: $\sigma = 0.4$, $T = 1$, $x_0 \sim \mathcal{N}(0, 0.2)$, and $\xi \sim \mathcal{U}(-0.2, 0.2)$. This work sets the tolerance-splitting parameter to $\theta = 0.5$ and the prescribed confidence level to $\nu = 0.05$ and defines the following:

¹Numerical results were produced using MATLAB (v2022b) on an Intel Xeon Platinum 8360Y CPU (2.40 GHz).

Algorithm 1: Adaptive multi-index double loop Monte Carlo estimator with importance sampling

User provided input: $P_0, N_0, \text{TOL}_r, \nu, \theta, \{\bar{M}_1, \bar{M}_2\}, \{\tilde{M}_1, \tilde{M}_2\}$;
Input from the numerical PDE solver: IS control $\zeta(\cdot, \cdot)$;
Input from pilots: $\{b_1, b_2\}, \{w_1, w_2\}, \{s_1, s_2\}$;
Estimate $\bar{G} = \mathbb{E} [\hat{\Delta}G_{(0,0)}]$ using **Algorithm 3** with $\bar{M}_1, \bar{M}_2, \zeta(\cdot, \cdot)$;
Estimate and store $\{V_{1,\alpha}, V_{2,\alpha}\}$ for $\alpha \in [0, 1, 2] \times [0, 1, 2]$ using **Algorithm 4** with $\tilde{M}_1, \tilde{M}_2, \zeta(\cdot, \cdot)$;
Set $L = 1$ and initialize $\hat{e}_b = +\infty$;
repeat
 Generate index set $\mathcal{I}(L)$ from (35);
 Estimate and store $\{V_{1,\alpha}, V_{2,\alpha}\}_{\alpha \in \mathcal{I}(L)}$ using **Algorithm 5**;
 Compute the optimal sample sizes $\{M_{1,\alpha}, M_{2,\alpha}\}_{\alpha \in \mathcal{I}(L)}$ using (B.3);
 Reevaluate $\bar{G} = \sum_{\alpha \in \mathcal{I}(L)} \mathbb{E} [\hat{\Delta}G_\alpha]$ as per (22) with
 $\{M_{1,\alpha}, M_{2,\alpha}\}, \zeta(\cdot, \cdot)$ using **Algorithm 3** for each α ;
 $\hat{e}_b = \frac{1}{|\bar{G}|} \sum_{\alpha \in \partial \mathcal{I}(L)} |\mathbb{E} [\hat{\Delta}G_\alpha]|$;
 $L \leftarrow L + 1$;
until $\hat{e}_b > (1 - \theta)\text{TOL}_r$;
 $\mathcal{A}_{\text{MIMC}} = \bar{G}$.

$$P_{\alpha_1} = 5 \times 2^{\alpha_1}, \quad N_{\alpha_2} = 4 \times 2^{\alpha_2}. \quad (39)$$

The multi-index DLMC method was assessed on the indicator observable $G(x) = \mathbb{1}_{\{x > K\}}$ and on several mollified variants with varying regularities, where $K \in \mathbb{R}$ denotes a threshold parameter determining the event rarity. Each mollifier is specified by a smoothing parameter ε and is defined as follows:

$$G(x) = \begin{cases} 0, & x \leq K - \varepsilon \\ S\left(\frac{x-K+\varepsilon}{2\varepsilon}\right), & K - \varepsilon < x \leq K + \varepsilon, \\ 1, & x > K + \varepsilon \end{cases} \quad (40)$$

where S denotes a polynomial selected to enforce the desired regularity. Table 1 lists the explicit forms. This work also considers the mollifier $G(x) = \frac{1}{2} \left(1 + \tanh\left(\frac{x-K}{\varepsilon}\right)\right)$ where $G \in C^\infty(\mathbb{R})$. Figure 1 depicts all observables in this section. For each choice of observable G , the IS control ζ from (13) is computed by simulating a particle system (2) off-line with $\bar{P} = 1000$ particles and $\bar{N} = 100$ time steps. These values were selected for the control computation so that the suboptimality of the control is negligible. Numerical tests found that further increasing this $\bar{P}, \bar{N}$ did not noticeably alter the results, indicating that the choice is sufficient. The PDE (12) is numerically solved using finite differences with linear interpolation.

Mollifier regularity	$S(x)$
$C^0(\mathbb{R})$	x
$C^1(\mathbb{R})$	$-2x^3 + 3x^2$
$C^2(\mathbb{R})$	$6x^5 - 15x^4 + 10x^3$
$C^3(\mathbb{R})$	$-20x^7 + 70x^6 - 84x^5 + 35x^4$

Table 1: List of polynomial functions applied in (40) to obtain mollifiers of varying regularity.

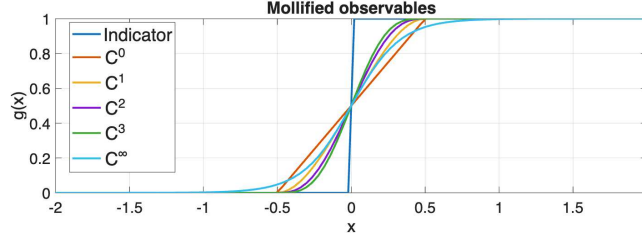


Figure 1: Rare-event observables G plotted for $\epsilon = 0.5$ and $K = 0$, including the indicator function and corresponding mollifiers of varying regularity.

5.1. Variance reduction

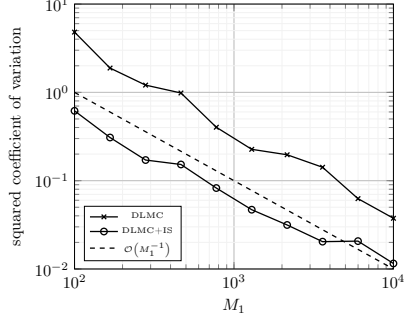
First, this work verifies variance reduction for DLMC estimators of $\mathbb{E}[\hat{\Delta}G_\alpha]$ for the C^∞ -mollifier with $\epsilon = \frac{1}{3}$ and $K = 3.5$, for which $\mathbb{E}[G(X(T))] \approx 2.04 \times 10^{-5}$. Figure 2a reports the squared coefficient of variation $(\frac{\text{Var}[\hat{\Delta}G_\alpha]}{|\mathbb{E}[\hat{\Delta}G_\alpha]|^2})$ as a function of M_1 with fixed M_2 for a given α , comparing DLMC with and without IS. For $\alpha = (2, 2)$, IS reduces the coefficient of variation by roughly an order of magnitude, while both estimators retain the expected $\mathcal{O}(M_1^{-1})$ decay. Then, this work examines the effect of the IS scheme across all multi-indices $\alpha \in \mathbb{N}^2$, via the following ratio:

$$\mathcal{R} = \frac{\text{Var}[\mathcal{A}_{\text{IS}}]}{\text{Var}[\mathcal{A}_{\text{MC}}]}, \quad (41)$$

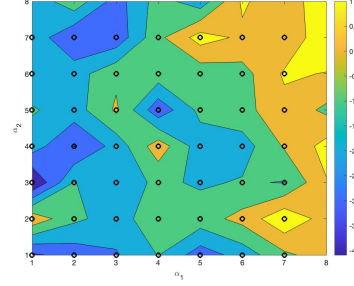
where $\mathcal{A}_{\text{IS}}$ and $\mathcal{A}_{\text{MC}}$ denote the DLMC estimators of $\mathbb{E}[\hat{\Delta}G_\alpha]$ with and without IS, respectively. Figure 2b presents a contour plot of $\mathcal{R}$ across various $\alpha \in \mathbb{N}^2$. Up to $\mathcal{O}(10^3)$ variance reduction was observed on the coarse multi-indices, demonstrating the effectiveness of the IS control ζ when combined with the multi-index DLMC estimator in addressing the explosive growth of the coefficient of variation. This variance reduction directly enables the success of the multi-index method for rare-event quantities.

5.2. Verifying Assumptions 1 and 2

Figure 3 presents numerical evidence for Assumptions 1 and 2 and estimates the convergence rates $\{b_1, b_2\}$, $\{w_1, w_2\}$, and $\{s_1, s_2\}$ for the Kuramoto example with the same C^∞ -mollifier as above. Figure 3a presents the decay of $|\mathbb{E}[\hat{\Delta}G_\alpha]|$, yielding rates $b_1 = 1$ and $b_2 = 1$. The figure also confirms that



(a) Squared coefficient of variation of the DLMC estimator for $\mathbb{E}[\hat{\Delta}G_{(2,2)}]$ with and without IS versus M_1 for fixed $M_2 = 100$.

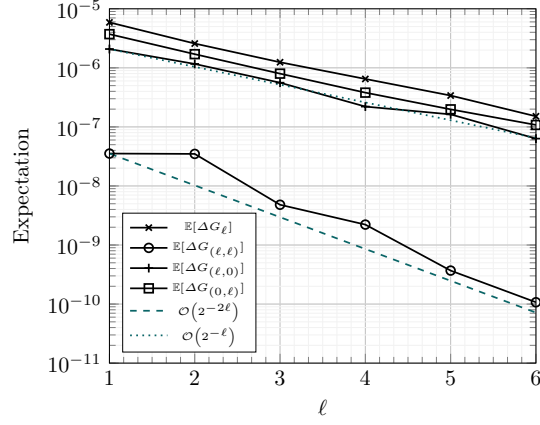


(b) Contour plot (log scale) of $\mathcal{R}$ for a set of $\alpha \in \mathbb{N}^2$ for the DLMC estimator of $\mathbb{E}[\hat{\Delta}G_\alpha]$, presenting significant variance reduction, especially in coarser multi-indices.

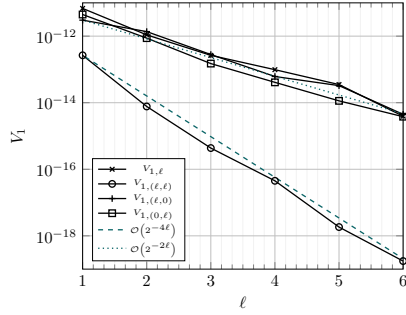
Figure 2: Variance reduction due to the IS scheme introduced in Section 3.2 for the Kuramoto model (4) with the C^∞ -mollifier observable with $\varepsilon = \frac{1}{3}$ and $K = 3.5$.

the mixed-difference expectation along the diagonal (ℓ, ℓ) converges at a rate of $b_1 + b_2 = 2$, demonstrating that it decays as the product of the single-direction errors. Figures 3b and 3c display the convergence of $V_{1,\alpha}$ and $V_{2,\alpha}$, respectively, yielding the rates $w_1 = 2$, $w_2 = 2$, $s_1 = 2$, and $s_2 = 1.5$. These results confirm the multiplicative decay of the mixed-difference variances. For comparison, we also plot the level differences $\mathbb{E}[\Delta G_\ell] = \mathbb{E}[G_\ell - G_{\ell-1}]$ from the multilevel DLMC estimator [13]. The mixed-differences display higher convergence rates than the level differences of the multilevel estimator. Figure 4 reveals that Assumptions 1 and 2 are satisfied for sufficiently fine discretizations.

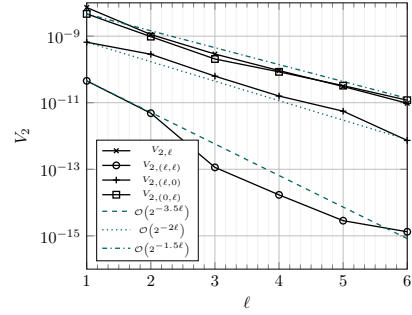
For further assessment of the validity of Assumptions 1 and 2, the same plots were repeated for observables G of varying regularity, ranging from the discontinuous indicator function to the C^3 -mollifier. Table 2 reports the numerical convergence rates in the P -direction $(\ell, 0)$, the N -direction $(0, \ell)$, and the diagonal (ℓ, ℓ) -direction. Appendix D presents the corresponding plots. For observables that are in $C^k(\mathbb{R})$ for $k \geq 2$, first-order bias decay occurs in both P and N directions, and second-order $V_{1,\alpha}$ decay occurs in P and N . For the Kuramoto example and the tested discretisation range, C^2 regularity appears sufficient to recover the multiplicative decay of the mixed-difference expectations and variances needed for Assumptions 1 and 2. The multiplicative bias and variance convergence rates degrade with the C^0 -mollifier, whereas slower convergence occurs in individual directions P and N for the indicator observable. Moreover, the numerical experiments demonstrate that the mixed-difference variances do not decay multiplicatively when replacing the C^∞ , bounded sinusoidal kernel in the Kuramoto model (4) with a C^0 , bounded triangular wave kernel, indicating that sufficient drift regularity is also necessary to satisfy Assumptions 1 and 2. This result confirms that sufficient smoothness in the MV-SDE (1) and observable G is necessary for the feasibility of the multi-index DLMC estimator. A theoretical explanation of this behaviour is the subject of ongoing work.



(a) Verification of (28) from Assumption 1 using Algorithm 3, where $M_1 = 10^3$ and $M_2 = 10^3$.

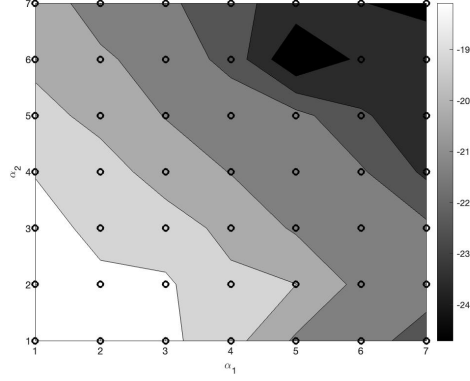


(b) Verification of (32) from Assumption 2 using Algorithm 4, where $M_1 = 10^2$ and $M_2 = 10^4$.

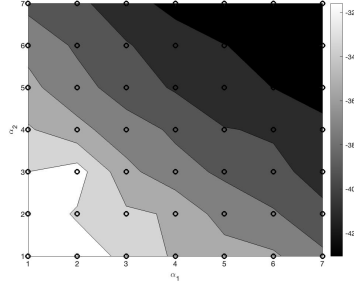


(c) Verification of (33) from Assumption 2 using Algorithm 4, where $M_1 = 10^2$ and $M_2 = 10^4$.

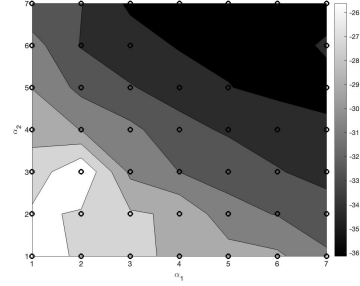
Figure 3: Kuramoto example, rate verification (C^∞ -mollifier observable with $K = 3.5$ and $\varepsilon = \frac{1}{3}$): Sample means and variances of mixed-differences in the multi-index DLMC estimator (22).



(a) Contour plots of the sample mean $|\mathbb{E}[\hat{\Delta G}_\alpha]|$ of the mixed-differences used in the multi-index DLMC estimator (22) obtained using Algorithm 3, where $M_1 = 10^3$ and $M_2 = 10^2$.



(b) Contour plots of sample $V_{1,\alpha}$ of the mixed-differences used in the multi-index DLMC estimator (22) obtained using Algorithm 4, where $M_1 = 10^2$ and $M_2 = 10^3$.



(c) Contour plots of sample $V_{2,\alpha}$ of the mixed-differences used in the multi-index DLMC estimator (22) obtained using Algorithm 4, where $M_1 = 10^2$ and $M_2 = 10^3$.

Figure 4: Kuramoto example, numerical rate verification (C^∞ -mollifier observable with $K = 3.5$ and $\varepsilon = \frac{1}{3}$): Contour plots of the sample mean and variances in the multi-index DLMC estimator (22). All plots numerically and asymptotically verify Assumptions 1 and 2.

According to Theorem 1, the mixed-difference convergence rates obtained for observables with C^2 -regularity (and higher) ensure that the optimal complexity of the multi-index DLMC estimator is $\mathcal{O}(\text{TOL}_r^{-2} (\log \text{TOL}_r^{-1})^2)$, which is one order better than the optimal work complexity of $\mathcal{O}(\text{TOL}_r^{-3})$ for the multilevel DLMC estimator [13]. The empirically observed rates $(b_i, w_i, s_i)_{i=1,2}$ validate Assumptions 1 and 2 for sufficiently smooth cases, supporting the use of the complexity predictions of Theorem 1.

$G(x)$	$\mathbb{E}$	$\hat{\Delta}G_\alpha$		$V_{1,\alpha}$			$V_{2,\alpha}$		
	(ℓ, ℓ)	$(\ell, 0)$	$(0, \ell)$	(ℓ, ℓ)	$(\ell, 0)$	$(0, \ell)$	(ℓ, ℓ)	$(\ell, 0)$	$(0, \ell)$
$\mathbb{1}_{\{x>K\}}$	0.7	1.1	0.3	1.2	1.4	1	1.1	0.6	0.6
C^0 -mollifier	1.8	1	1.6	3.1	2.1	1.7	2	1.8	0.8
C^1 -mollifier	1.9	1	1	3.6	2.1	1.8	2.5	2	1
C^2 -mollifier	2	1	0.8	3.7	2.1	1.8	2.6	2	1
C^3 -mollifier	2.2	1	0.8	3.8	2.1	1.8	2.6	2	1
C^∞ -mollifier	2	1	1	4	2	2	3.5	2	1.5

Table 2: Comparison of expectation and variance convergence rates of mixed-differences for the Kuramoto model (4) with mollifiers of varying regularity of the indicator function $G(x) = \mathbb{1}_{\{x>K\}}$ for $K = 2$ and $\varepsilon = 0.5$.

5.3. Computational complexity: Multi-index vs. multilevel

Figure 5 illustrates the results of running the adaptive multi-index DLMC Algorithm 1 for the C^∞ -mollifier with $K = 3.5$ and $\varepsilon = \frac{1}{3}$, comparing them with the results from the adaptive multilevel algorithm in the work by [13]. $K = 3.5$ corresponds to $\mathbb{E}[G(X(T))] \approx 2.04 \times 10^{-5}$, a sufficiently rare event. This work sets $\{\bar{M}_1, \bar{M}_2\} = \{10^3, 10^2\}$ and $\{\tilde{M}_1, \tilde{M}_2\} = \{25, 100\}$ (Algorithm 1). Algorithm 1 and the corresponding multilevel algorithm were independently executed five times, and the combined results were plotted. Both algorithms apply the same IS control; the only difference is the hierarchy of discretizations (multi-index vs multilevel).

Figure 5a displays the exact relative error achieved using the multi-index and multilevel DLMC estimators for various relative tolerances TOL_r , estimated using a reference multi-index DLMC approximation with $\text{TOL}_r = 1\%$. Each marker represents a separate iteration of the corresponding adaptive algorithm. Figure 5a illustrates the effectiveness of IS in estimating a rare-event quantity within a relative error tolerance, which is infeasible with the standard MC method. The result confirms that the multi-index and multilevel estimators with IS satisfy TOL_r , in the sense of (6).

Figure 5b displays the maximum discretization level for the number of particles and time steps to ensure that the multilevel and multi-index estimators satisfy the relative bias constraint (26). For $\text{TOL}_r = 1.67\%$, the multilevel DLMC estimator required $P = 640$ particles and $N = 512$ time steps, whereas the multi-index DLMC estimator increased to $P = 640$ particles and $N = 1024$.

time steps with a lower net computational work. This result highlights the efficiency of the MIMC method over the multilevel MC method in distributing computational work across multi-indices, while satisfying the same relative accuracy.

Figure 5c presents the average computational runtime for Algorithm 1 for various relative error tolerances, confirming the asymptotic complexity rates derived in Theorem 1. The multilevel DLMC runtime scales with a power of roughly -3 , whereas the multi-index DLMC scales at nearly -2 . Moreover, the multi-index estimator not only asymptotically outperforms the multilevel method, but the absolute runtime is significantly lower for all tested tolerances due to the considerable IS variance reduction (smaller constant in complexity).

Figure 5d presents the average computational cost for both estimators for different values of TOL_r , computed using the following computational cost model (34):

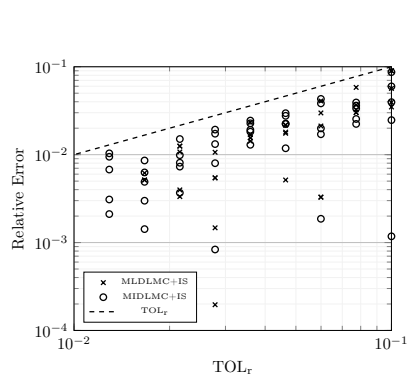
$$\text{Computational Cost}[\mathcal{A}_{\text{MIMC}}(\mathcal{I}(L))] \approx \sum_{\alpha \in \mathcal{I}(L)} (M_{1,\alpha} N_{\alpha_2} P_{\alpha_1}^2 + M_{1,\alpha} M_{2,\alpha} N_{\alpha_2} P_{\alpha_1}) \cdot \quad (42)$$

For $\text{TOL}_r = 2.78\%$, the multi-index DLMC estimator required $\sim 2 \times 10^9$ operations on average whereas the multilevel estimator required $\sim 10^{10}$ operations. For perspective, a direct MC simulation (without IS or multilevel/multi-index MC) requires $\sim 10^{15}$ operations to achieve the same relative accuracy given the event rarity.

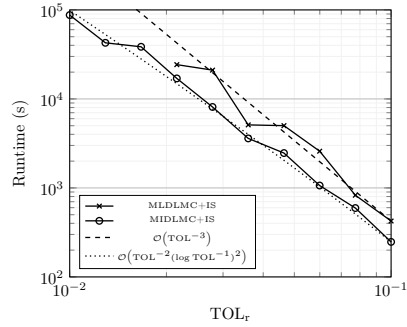
In conclusion, this numerical study confirms the theoretical convergence rates and complexity derived in Theorem 1. The slopes observed in Figures 5c and 5d match the predictions of Theorem 1, lending credence to Assumptions 1 and 2 for this problem. Figures 5c and 5d support the superiority of the proposed multi-index DLMC estimator over the multilevel DLMC estimator, with the incorporated IS scheme enabling both estimators to estimate rare-event quantities feasibly. The multilevel DLMC failed to produce estimates of the quantity of interest up to $\text{TOL}_r = 1\%$, whereas the multi-index algorithm achieved estimates within a fixed computational budget. Moreover, the multi-index estimator demonstrates a complexity reduction of about one order of magnitude (up to logarithmic terms) compared with the multilevel estimator, enabling the computation of rare-event quantities in MV-SDEs that were previously infeasible. The trade-off is the upfront computation of the IS control and the need for careful coupling of mixed-differences, but the results demonstrate that this overhead is justified for rare-events.

6. Conclusion

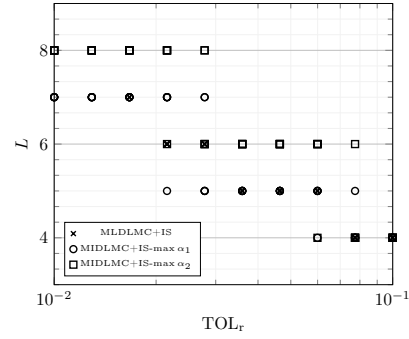
This work presents theoretical and numerical evidence under verifiable assumptions that the multi-index DLMC estimator outperforms the multilevel DLMC estimator [13] for rare-event quantities in MV-SDEs. These quantities are expressed as expectations of sufficiently smooth observables of solutions to



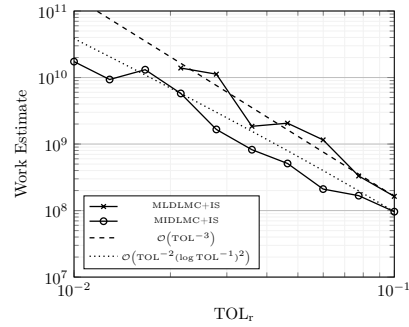
(a) Estimated relative error for multilevel and multi-index DLMC estimators with IS.



(c) Average computational runtime for different values of TOL_r .



(b) Maximum discretization level of the number of time steps and particles for the multi-index and multilevel DLMC estimators for different tolerances.



(d) Average computational cost estimate (42) for different values of TOL_r .

Figure 5: Kuramoto example, adaptive DLMC algorithms with IS (C^∞ -mollifier observable with $K = 3.5$ and $\varepsilon = \frac{1}{3}$): Comparing multilevel and multi-index estimators. Here, MLDLMC+IS denotes multilevel DLMC with IS and MIDLMC+IS denotes multi-index DLMC with IS.

stochastic particle systems in the mean-field limit. This work applies the same IS scheme in the work by [1] at every multi-index in the proposed multi-index DLMC estimator. The numerical results confirmed a significant variance reduction on all multi-indices. For the considered example, the multi-index DLMC estimator achieved a complexity of $\mathcal{O}(\text{TOL}_r^{-2}(\log \text{TOL}_r^{-1})^2)$, an order of improvement over the multilevel DLMC estimator, where TOL_r represents the relative error tolerance. Not only did the asymptotic exponent improve, but the constant prefactor is much smaller than that in a comparable MIMC method for a nonrare observable [9]. This improvement enables the computation of rare-event quantities with much less computational cost, enabling robust uncertainty quantification analysis for MV-SDEs, which was previously impractical.

Although this work presents numerical results for the Kuramoto model, the proposed method is quite general and readily extends to other MV-SDEs, such as animal flocking models [45] and social opinion dynamics [46]. The critical requirements are sufficient regularity and a tractable numerical solution to the control PDE (12). Similar computational gains are expected in these settings, although verifying Assumptions 1 and 2 for each case is critical. Future work will focus on rigorously analyzing these mixed regularity assumptions to derive sufficient conditions on $b, \sigma, \kappa_1, \kappa_2$, and G . Extending the method to nonsmooth (C^0 or lower) rare-event observables appears viable (as evidenced by the partial results with mollifiers; see Remark 2), although an adaptive mollification or other variance control strategy may be necessary for discontinuous observables. Although the complexity bound (36) does not depend on d , solving the control PDE (12) in very high dimensions ($d > 10$) is impractical. In such cases, model reduction techniques or stochastic optimization could be explored to obtain an effective IS control (see Remark 3). Finally, optimizing the multi-index DLMC parameters (e.g., by determining the optimal θ [47] or integrating a continuation-type multi-index algorithm [48]) could yield additional computational savings.

Appendix A. Double loop Monte Carlo algorithm with importance sampling

Algorithm 2: Double loop Monte Carlo algorithm with importance sampling for MV-SDEs

Off-line:

Generate one realization of law $\{X_p^{\bar{P}|\bar{N}}\}_{p=1}^{\bar{P}} \sim \mu^{\bar{P}|\bar{N}}$ with the $\bar{P}$ -particle system and $\bar{N}$ time steps with a large value of $\bar{P}, \bar{N}$;
 Given $\mu^{\bar{P}|\bar{N}}$, solve the PDE (12) to obtain IS control ζ ;

Inputs: P, N, M_1, M_2, ζ ;

for $i = 1, \dots, M_1$ **do**

Generate realization of random variables $\omega_{1:P}^{(i)}$;
 Generate a realization of law $\{X_p^{P|N}\}_{p=1}^P \sim \mu^{P|N}(\omega_{1:P}^{(i)})$ with the P -particle system and N time steps;
for $j = 1, \dots, M_2$ **do**
 Generate realization of random variables $\tilde{\omega}^{(i,j)}$;
 Given $\mu^{P|N}(\omega_{1:P}^{(i)})$ and ζ , generate a sample path of (9) with N time steps;
 Compute $G(\bar{X}_\zeta^{P|N}(T))(\omega_{1:P}^{(i)} \times \tilde{\omega}^{(i,j)})$;
 Compute $\mathbb{L}^{P|N}(\omega_{1:P}^{(i)} \times \tilde{\omega}^{(i,j)})$ using (18);
end

Approximate $\mathbb{E}\left[G(\bar{X}_\zeta^{P|N}(T))\mathbb{L}^{P|N} \mid \mu^{P|N}(\omega_{1:P}^{(i)})\right]$ by
 $\frac{1}{M_2} \sum_{j=1}^{M_2} G(\bar{X}_\zeta^{P|N}(T))\mathbb{L}^{P|N}(\omega_{1:P}^{(i)} \times \tilde{\omega}^{(i,j)})$;

end

Approximate $\mathbb{E}[G(\bar{X}^{P|N}(T))]$ by
 $\frac{1}{M_1} \sum_{i=1}^{M_1} \frac{1}{M_2} \sum_{j=1}^{M_2} G(\bar{X}_\zeta^{P|N}(T))\mathbb{L}^{P|N}(\omega_{1:P}^{(i)} \times \tilde{\omega}^{(i,j)})$;

Appendix B. Proof of Theorem 1

For a given index set $\mathcal{I}$, the optimal $M_{1,\alpha} \in \mathbb{R}_+$ and $M_{2,\alpha} \in \mathbb{R}_+$ are determined for all $\alpha \in \mathcal{I}$, minimizing the total computational cost (34) subject to the statistical error constraint (29):

$$\begin{cases} \min_{\{M_{1,\alpha}, M_{2,\alpha}\}_{\alpha \in \mathcal{I}}} \sum_{\alpha \in \mathcal{I}} M_{1,\alpha} N_{\alpha_2} P_{\alpha_1}^2 + M_{1,\alpha} M_{2,\alpha} N_{\alpha_2} P_{\alpha_1} \\ \text{s.t.} \quad \left(\sum_{\alpha \in \mathcal{I}} \frac{V_{1,\alpha}}{M_{1,\alpha}} + \frac{V_{2,\alpha}}{M_{1,\alpha} M_{2,\alpha}} \right) \approx \left(\frac{\theta \text{TOL}_r \mathbb{E}[G(X(T))]}{C_\nu} \right)^2. \end{cases} \quad (\text{B.1})$$

The solution to (B.1) in $\mathbb{R}_+$ using the Lagrange multiplier method for constrained optimization yields the following:

$$\begin{aligned}\mathcal{M}_{1,\alpha} &= \left(\frac{C_\nu}{\theta \text{TOL}_r \mathbb{E}[G(X(T))]} \right)^2 \sqrt{\frac{V_{1,\alpha}}{N_{\alpha_2} P_{\alpha_1}^2}} \sum_{\beta \in \mathcal{I}} \left(\sqrt{V_{1,\beta} N_{\beta_2} P_{\beta_1}^2} + \sqrt{V_{2,\beta} N_{\beta_2} P_{\beta_1}} \right), \\ \tilde{\mathcal{M}}_\alpha &= \mathcal{M}_{1,\alpha} \mathcal{M}_{2,\alpha} \\ &= \left(\frac{C_\nu}{\theta \text{TOL}_r \mathbb{E}[G(X(T))]} \right)^2 \sqrt{\frac{V_{2,\alpha}}{N_{\alpha_2} P_{\alpha_1}}} \sum_{\beta \in \mathcal{I}} \left(\sqrt{V_{1,\beta} N_{\beta_2} P_{\beta_1}^2} + \sqrt{V_{2,\beta} N_{\beta_2} P_{\beta_1}} \right).\end{aligned}\tag{B.2}$$

In practice, natural numbers are employed for $\{\mathcal{M}_{1,\alpha}, \mathcal{M}_{2,\alpha}\}_{\alpha \in \mathcal{I}}$. Therefore, to guarantee at least one $\mathcal{M}_{1,\alpha}$ and $\mathcal{M}_{2,\alpha}$ for each α , the following quasi-optimal solution to (B.1) is employed:

$$M_{1,\alpha} = \lceil \mathcal{M}_{1,\alpha} \rceil, \quad M_{2,\alpha} = \left\lceil \frac{\tilde{\mathcal{M}}_\alpha}{\lceil \mathcal{M}_{1,\alpha} \rceil} \right\rceil.\tag{B.3}$$

Using (B.3), the computational cost of the estimator is bound as follows:

$$\begin{aligned}\mathcal{W}[\mathcal{A}_{\text{MIMC}}(\mathcal{I})] &\lesssim \sum_{\alpha \in \mathcal{I}} \left((\mathcal{M}_{1,\alpha} + 1) N_{\alpha_2} P_{\alpha_1}^2 + (\mathcal{M}_{1,\alpha} + 1) \left(\frac{\tilde{\mathcal{M}}_\alpha}{\lceil \mathcal{M}_{1,\alpha} \rceil} + 1 \right) N_{\alpha_2} P_{\alpha_1} \right) \\ &\leq \underbrace{\sum_{\alpha \in \mathcal{I}} \left(\mathcal{M}_{1,\alpha} N_{\alpha_2} P_{\alpha_1}^2 + \tilde{\mathcal{M}}_\alpha N_{\alpha_2} P_{\alpha_1} \right)}_{=W_1(\mathcal{I})} \\ &\quad + \underbrace{\sum_{\alpha \in \mathcal{I}} (P_{\alpha_1}^2 N_{\alpha_2} + P_{\alpha_1} N_{\alpha_2})}_{=W_2(\mathcal{I}), \text{ cost of one sample per multi-index}} \\ &\quad + \underbrace{\sum_{\alpha \in \mathcal{I}} \mathcal{M}_{1,\alpha} N_{\alpha_2} P_{\alpha_1}}_{=W_3(\mathcal{I})} + \underbrace{\sum_{\alpha \in \mathcal{I}} \frac{\tilde{\mathcal{M}}_\alpha}{\lceil \mathcal{M}_{1,\alpha} \rceil} N_{\alpha_2} P_{\alpha_1}}_{=W_4(\mathcal{I})}.\end{aligned}\tag{B.4}$$

Applying $P_{\alpha_1} \geq 1$ reveals that

$$\begin{aligned}W_3(\mathcal{I}) &= \sum_{\alpha \in \mathcal{I}} \mathcal{M}_{1,\alpha} N_{\alpha_2} P_{\alpha_1} \leq \sum_{\alpha \in \mathcal{I}} \mathcal{M}_{1,\alpha} N_{\alpha_2} P_{\alpha_1}^2 \leq W_1(\mathcal{I}), \\ W_4(\mathcal{I}) &= \sum_{\alpha \in \mathcal{I}} \frac{\tilde{\mathcal{M}}_\alpha}{\lceil \mathcal{M}_{1,\alpha} \rceil} N_{\alpha_2} P_{\alpha_1} \leq \sum_{\alpha \in \mathcal{I}} \frac{\tilde{\mathcal{M}}_\alpha}{\max(1, \mathcal{M}_{1,\alpha})} N_{\alpha_2} P_{\alpha_1} \\ &\leq \sum_{\alpha \in \mathcal{I}} \tilde{\mathcal{M}}_\alpha N_{\alpha_2} P_{\alpha_1} \leq W_1(\mathcal{I}).\end{aligned}$$

Hence, (B.4) is rewritten as follows:

$$\mathcal{W}[\mathcal{A}_{\text{MIMC}}(\mathcal{I})] \lesssim W_1(\mathcal{I}) + W_2(\mathcal{I}). \quad (\text{B.5})$$

Substituting (B.2) in (B.5) yields

$$\begin{aligned} \mathcal{W}[\mathcal{A}_{\text{MIMC}}(\mathcal{I})] &\lesssim \left(\frac{C_\nu}{\theta \text{TOL}_r \mathbb{E}[G(X(T))]} \right)^2 \left(\sum_{\alpha \in \mathcal{I}} \sqrt{V_{1,\alpha} N_{\alpha_2} P_{\alpha_1}^2} + \sqrt{V_{2,\alpha} N_{\alpha_2} P_{\alpha_1}} \right)^2 \\ &\quad + \sum_{\alpha \in \mathcal{I}} (P_{\alpha_1}^2 N_{\alpha_2} + P_{\alpha_1} N_{\alpha_2}). \end{aligned} \quad (\text{B.6})$$

Under Assumptions 1 and 2, the total computational cost (B.6) to reach a relative error tolerance TOL_r is estimated as follows:

$$\begin{aligned} \mathcal{W}[\mathcal{A}_{\text{MIMC}}(\mathcal{I})] &\lesssim \left(\frac{C_\nu}{\theta \text{TOL}_r \mathbb{E}[G(X(T))]} \right)^2 \left(\sum_{\alpha \in \mathcal{I}} \left(2^{\frac{\alpha_1(2-w_1)+\alpha_2(1-w_2)}{2}} + 2^{\frac{\alpha_1(1-s_1)+\alpha_2(1-s_2)}{2}} \right) \right)^2 \\ &\quad + \sum_{\alpha \in \mathcal{I}} 2^{2\alpha_1+\alpha_2} + 2^{\alpha_1+\alpha_2} \\ &\leq \left(\frac{2C_\nu}{\theta \text{TOL}_r \mathbb{E}[G(X(T))]} \right)^2 \left(\sum_{\alpha \in \mathcal{I}} \left(2^{\frac{\alpha_1(1-\min(w_1-1, s_1))+\alpha_2(1-\min(w_2, s_2))}{2}} \right) \right)^2 \\ &\quad + 2 \sum_{\alpha \in \mathcal{I}} 2^{2\alpha_1+\alpha_2}. \end{aligned} \quad (\text{B.7})$$

The computational cost of the estimator is bounded as follows by defining $\bar{s}_1 = \min(w_1 - 1, s_1)$ and $\bar{s}_2 = \min(w_2, s_2)$:

$$\begin{aligned} \mathcal{W}[\mathcal{A}_{\text{MIMC}}(\mathcal{I})] &\lesssim \left(\frac{2C_\nu}{\theta \text{TOL}_r \mathbb{E}[G(X(T))]} \right)^2 \left(\underbrace{\sum_{\alpha \in \mathcal{I}} \left(2^{\frac{\alpha_1(1-\bar{s}_1)+\alpha_2(1-\bar{s}_2)}{2}} \right)}_{=\tilde{W}_1(\mathcal{I})} \right)^2 \\ &\quad + 2 \underbrace{\sum_{\alpha \in \mathcal{I}} 2^{2\alpha_1+\alpha_2}}_{=\tilde{W}_2(\mathcal{I})}, \end{aligned} \quad (\text{B.8})$$

where $\tilde{W}_2(\mathcal{I})$ represents the total computational cost required to generate one sample for each multi-index $\alpha \in \mathcal{I}$, which is the minimum cost for a DLMC estimator. It does not dominate the first term in the bound (B.7).

Next, this work justifies the choice for the set of multi-indices $\mathcal{I}$ that minimizes $\mathcal{W}[\mathcal{A}_{\text{MIMC}}(\mathcal{I})]$. First, the following bound is obtained for the relative bias of the multi-index DLMC estimator:

$$\epsilon_b(\mathcal{I}) = \frac{1}{|\mathbb{E}[G(X(T))]|} \left| \sum_{\alpha \notin \mathcal{I}} \mathbb{E}[\hat{\Delta}G_{\alpha}] \right| \leq \frac{1}{|\mathbb{E}[G(X(T))]|} \sum_{\alpha \notin \mathcal{I}} |\mathbb{E}[\hat{\Delta}G_{\alpha}]|. \quad (\text{B.9})$$

Then, $\mathcal{I}$ is set to satisfy the following relative bias constraint:

$$\frac{1}{|\mathbb{E}[G(X(T))]|} \sum_{\alpha \notin \mathcal{I}} |\mathbb{E}[\hat{\Delta}G_{\alpha}]| \leq (1 - \theta) \text{TOL}_r. \quad (\text{B.10})$$

Using Assumption 1, (B.10) is written as follows:

$$\tilde{B}(\mathcal{I}) := \sum_{\alpha \notin \mathcal{I}} 2^{-\alpha_1 b_1 - \alpha_2 b_2} \leq \frac{(1 - \theta) \text{TOL}_r |\mathbb{E}[G(X(T))]|}{Q_B}. \quad (\text{B.11})$$

The solution to the following optimization problem to obtain the quasi-optimal set $\mathcal{I}$:

$$\min_{\mathcal{I} \subset \mathbb{N}^2} \tilde{W}_1(\mathcal{I}) \quad \text{such that} \quad \tilde{B}(\mathcal{I}) \leq \frac{(1 - \theta) \text{TOL}_r |\mathbb{E}[G(X(T))]|}{Q_B}. \quad (\text{B.12})$$

The method in the work by [14] demonstrates that (35) is the optimal index set in the sense of (B.12). However, the conditions $2b_1 \geq \bar{s}_1 - 1$ and $2b_2 \geq \bar{s}_2 - 1$ must be satisfied to allow admissible index sets. Next, using the optimal index set (35) yields the following:

$$\tilde{W}_1(\mathcal{I}(L)) = \sum_{\alpha \in \mathbb{N}^2: \delta \cdot \alpha \leq L} \exp(\mathbf{g} \cdot \alpha), \quad (\text{B.13})$$

where $\mathbf{g} = \log 2 \left[\frac{1-\bar{s}_1}{2}, \frac{1-\bar{s}_2}{2} \right]$ and $\delta = \frac{\log 2}{C_{\delta}} \left[\frac{1-\bar{s}_1}{2} + b_1, \frac{1-\bar{s}_2}{2} + b_2 \right]$.

In addition, $C_{\delta} = \log 2 \left(\frac{1-\bar{s}_1}{2} + b_1 + \frac{1-\bar{s}_2}{2} + b_2 \right)$ represents the normalizing constant, ensuring that $0 \leq \delta_1, \delta_2 \leq 1$ and $\delta_1 + \delta_2 = 1$. For the optimal index set $\mathcal{I}(L)$ defined by (35), computational cost bound (B.13) and the bias relation (B.11), the asymptotic MIMC complexity estimates of [14] directly imply Theorem 1, thus concluding the proof. This computation is omitted here for ease of presentation.

Appendix C. Algorithm 1 implementation details

Algorithm 3 outlines the implementation of the IS scheme presented in Section 4 to estimate $\mathbb{E}[\hat{\Delta}G_{\alpha}]$ for each $\alpha \in \mathcal{I}$, which is necessary for the multi-index DLMC estimator (22).

Algorithm 3: Importance sampling scheme to estimate $\mathbb{E} [\hat{\Delta}G_{\alpha}]$

Inputs: $\alpha, M_1, M_2, \zeta(\cdot, \cdot)$;
for $m_1 = 1, \dots, M_1$ **do**
 Generate a realization of random variables $\omega_{1:P_{\alpha_1}}^{(\alpha, m_1)}$;
 Using $\omega_{1:P_{\alpha_1}}^{(\alpha, m_1)}$, generate a realization of empirical law μ for all
 required discretizations, as per (23) by generating sample paths of
 the particle system (2);
 for $m_2 = 1, \dots, M_2$ **do**
 Generate a realization of random variables $\tilde{\omega}^{(\alpha, m_1, m_2)}$;
 Using $\tilde{\omega}^{(\alpha, m_1, m_2)}$, generate a sample path of (9) for all required
 discretizations as per (23) with control ζ ;
 Compute $\hat{\Delta}G_{\alpha}(\omega_{1:P_{\alpha_1}}^{(\alpha, m_1)} \times \tilde{\omega}^{(\alpha, m_1, m_2)})$ using (23);
 end
end
Approximate $\mathbb{E} [\hat{\Delta}G_{\alpha}]$ by
 $\frac{1}{M_1} \sum_{m_1=1}^{M_1} \frac{1}{M_2} \sum_{m_2=1}^{M_2} \hat{\Delta}G_{\alpha}(\omega_{1:P_{\alpha_1}}^{(\alpha, m_1)} \times \tilde{\omega}^{(\alpha, m_1, m_2)})$;

Appendix C.1. Estimating the bias of index set $\mathcal{I}$

In Algorithm 1, to determine the optimal index set $\mathcal{I}(L)$ using (35), the rates $\{b_1, b_2\}$, $\{w_1, w_2\}$, and $\{s_1, s_2\}$ associated with Assumptions 1 and 2 must be estimated. Reliable pilot runs are necessary to obtain these estimates. However, such pilot runs are unnecessary for the adaptive DLMC algorithm [1] or multi-level DLMC algorithm [13]. The optimal index set in (35) is defined using the parameter (threshold) L . Given that $\mathcal{I}(L) \subset \mathcal{I}(L+1)$, the following heuristic estimate [14] is applied for the absolute bias corresponding to index set $\mathcal{I}(L)$.

$$|\mathbb{E}[G - \mathcal{A}_{\text{MIMC}}(\mathcal{I}(L))]| \approx \sum_{\alpha \in \partial \mathcal{I}(L)} |\mathbb{E}[\Delta^{\text{mix}} G_{\alpha}]|, \quad (\text{C.1})$$

where $\partial \mathcal{I}(L)$ denotes the outer boundary of the index set $\mathcal{I}(L)$, which is defined as follows:

$$\partial \mathcal{I}(L) = \{\alpha \in \mathcal{I}(L) : \alpha + (1, 0) \notin \mathcal{I}(L) \text{ or } \alpha + (0, 1) \notin \mathcal{I}(L)\}. \quad (\text{C.2})$$

Because $\partial \mathcal{I}(L) \subset \mathcal{I}(L)$, the computed nested averages are employed using the optimal $\{M_{1,\alpha}, M_{2,\alpha}\}_{\alpha \in \partial \mathcal{I}(L)}$ in (22) to estimate each expectation in (C.1).

Appendix C.2. Estimating $\{V_{1,\alpha}, V_{2,\alpha}\}$

Computationally inexpensive, robust empirical estimates of $\{V_{1,\alpha}, V_{2,\alpha}\}$ for all $\alpha \in \mathcal{I}(L)$ are necessary to compute the optimal number of samples to satisfy the variance constraint in (29) of the estimator using (B.3). Therefore, the

DLMC method in Algorithm 4 is employed with appropriately selected values of $\bar{M}_1, \bar{M}_2$.

Algorithm 4: Estimating $\{V_{1,\alpha}, V_{2,\alpha}\}$ for the adaptive multi-index double loop Monte Carlo method

Inputs: $\alpha, M_1, M_2, \zeta(\cdot, \cdot)$;
for $m_1 = 1, \dots, M_1$ **do**
 Generate a realization of random variables $\omega_{1:P_{\alpha_1}}^{(\alpha, m_1)}$;
 Using $\omega_{1:P_{\alpha_1}}^{(\alpha, m_1)}$, generate a realization of empirical law μ for all required discretizations, as per (23) by generating sample paths of the particle system (2);
 for $m_2 = 1, \dots, M_2$ **do**
 Generate a realization of random variables $\tilde{\omega}^{(\alpha, m_1, m_2)}$;
 Using $\tilde{\omega}^{(\alpha, m_1, m_2)}$, generate a sample path of (9) for all required discretizations, as per (23) with control ζ ;
 Compute $\hat{\Delta}G_{\alpha}(\omega_{1:P_{\alpha_1}}^{(\alpha, m_1)} \times \tilde{\omega}^{(\alpha, m_1, m_2)})$ using (23);
 end
 Approximate $\mathbb{E} [\hat{\Delta}G_{\alpha} | \omega_{1:P_{\alpha_1}}^{(\alpha, m_1)}]$ by
 $\frac{1}{M_2} \sum_{m_2=1}^{M_2} \hat{\Delta}G_{\alpha}(\omega_{1:P_{\alpha_1}}^{(\alpha, m_1)} \times \tilde{\omega}^{(\alpha, m_1, m_2)})$;
 Approximate $\text{Var} [\hat{\Delta}G_{\alpha} | \omega_{1:P_{\alpha_1}}^{(\alpha, m_1)}]$ by the sample variance of
 $\left\{ \hat{\Delta}G_{\alpha}(\omega_{1:P_{\alpha_1}}^{(\alpha, m_1)} \times \tilde{\omega}^{(\alpha, m_1, m_2)}) \right\}_{m_2=1}^{M_2}$;
end
 Approximate $V_{1,\alpha}$ by the sample variance of $\left\{ \mathbb{E} [\hat{\Delta}G_{\alpha} | \omega_{1:P_{\alpha_1}}^{(\alpha, m_1)}] \right\}_{m_1=1}^{M_1}$;
 Approximate $V_{2,\alpha}$ by $\frac{1}{M_1} \sum_{m_1=1}^{M_1} \text{Var} [\hat{\Delta}G_{\alpha} | \omega_{1:P_{\alpha_1}}^{(\alpha, m_1)}]$.

Estimating $\{V_{1,\alpha}, V_{2,\alpha}\}$ for all $\alpha \in \mathcal{I}(L)$ using Algorithm 4 can become a computational burden when $\mathcal{I}(L)$ is even moderately large. To ease this computational overload, Assumption 2 is applied using an extrapolation approach to estimate $\{V_{1,\alpha}, V_{2,\alpha}\}$ for deeper multi-indices. Algorithm 4 is applied solely to estimate $\{V_{1,\alpha}, V_{2,\alpha}\}$ for the small full tensor index set $\{0, 1, 2\} \times \{0, 1, 2\}$. Then, the extrapolation Algorithm 5 is employed using Assumption 2 to estimate $\{V_{1,\alpha}, V_{2,\alpha}\}$ for the remaining multi-indices. To alleviate the computational burden further, this work only estimates $\{V_{1,\alpha}, V_{2,\alpha}\}$ for the newly added multi-indices in each iteration (i.e., for $\alpha \in \mathcal{I}(L+1) - \mathcal{I}(L)$). As $\mathcal{I}(L) \subset \mathcal{I}(L+1)$ by construction, the estimates of $\{V_{1,\alpha}, V_{2,\alpha}\}$ can be reused for multi-indices from previous iterations.

Appendix C.3. Relative error control

The adaptive algorithm requires a heuristic estimate of the quantity of interest $\mathbb{E}[G(X(T))]$ to meet the relative error constraints in (26) and (29) for a

Algorithm 5: Sample variance extrapolation for the adaptive multi-index double loop Monte Carlo method

Inputs: $\mathcal{I}, \{\bar{M}_1, \bar{M}_2\}, (w_1, w_2), (s_1, s_2)$;
 Estimate $\{V_{1,\alpha}, V_{2,\alpha}\}$ for $\alpha \in \{0, 1, 2\} \times \{0, 1, 2\}$ using Algorithm 4
 with $\{\bar{M}_1, \bar{M}_2\}$;
for $\alpha \in \mathcal{I}$ **do**
 if $\alpha_1 = 0, 1$ **then**
 $V_{1,\alpha} = \max\left(\frac{V_{1,\alpha-(0,1)}}{2^{w_2}}, \frac{V_{1,\alpha-(0,2)}}{2^{2w_2}}\right)$;
 $V_{2,\alpha} = \max\left(\frac{V_{2,\alpha-(0,1)}}{2^{s_2}}, \frac{V_{2,\alpha-(0,2)}}{2^{2s_2}}\right)$;
 else if $\alpha_2 = 0, 1$ **then**
 $V_{1,\alpha} = \max\left(\frac{V_{1,\alpha-(1,0)}}{2^{w_1}}, \frac{V_{1,\alpha-(2,0)}}{2^{2w_1}}\right)$;
 $V_{2,\alpha} = \max\left(\frac{V_{2,\alpha-(1,0)}}{2^{s_1}}, \frac{V_{2,\alpha-(2,0)}}{2^{2s_1}}\right)$;
 else
 $V_{1,\alpha} = \max\left(\frac{V_{1,\alpha-(0,1)}}{2^{w_2}}, \frac{V_{1,\alpha-(1,0)}}{2^{w_1}}\right)$;
 $V_{2,\alpha} = \max\left(\frac{V_{2,\alpha-(0,1)}}{2^{s_2}}, \frac{V_{2,\alpha-(1,0)}}{2^{s_1}}\right)$;
 end
end

given relative error tolerance TOL_r . In the algorithm, this estimate is updated at each iteration L . For $L = 0$, Algorithm 3 with appropriately selected $\bar{M}_1$ and $\bar{M}_2$ provides an initial estimate of $\bar{G}$ for $\mathbb{E}[G_{(0,0)}]$. In subsequent iterations, the multi-index estimator in (22) with optimal values of $\{M_{1,\alpha}, M_{2,\alpha}\}_{\alpha \in \mathcal{I}(L)}$ updates $\bar{G}$ and the absolute error tolerances. Combining these components, the adaptive multi-index DLMC Algorithm 1 is developed to evaluate rare-event observables associated with MV-SDEs.

Appendix D. Further numerical results

Figures D.6, D.7, and D.8 corroborate the numerical rates depicted in Table 2. The corresponding plots for the indicator function are omitted because they were erratic, which is expected, as a discontinuous observable does not satisfy the necessary regularity for Assumption 2. This work also refrains from producing the corresponding plots for the Kuramoto model with a less regular triangular wave kernel, although these plots can be produced upon request.

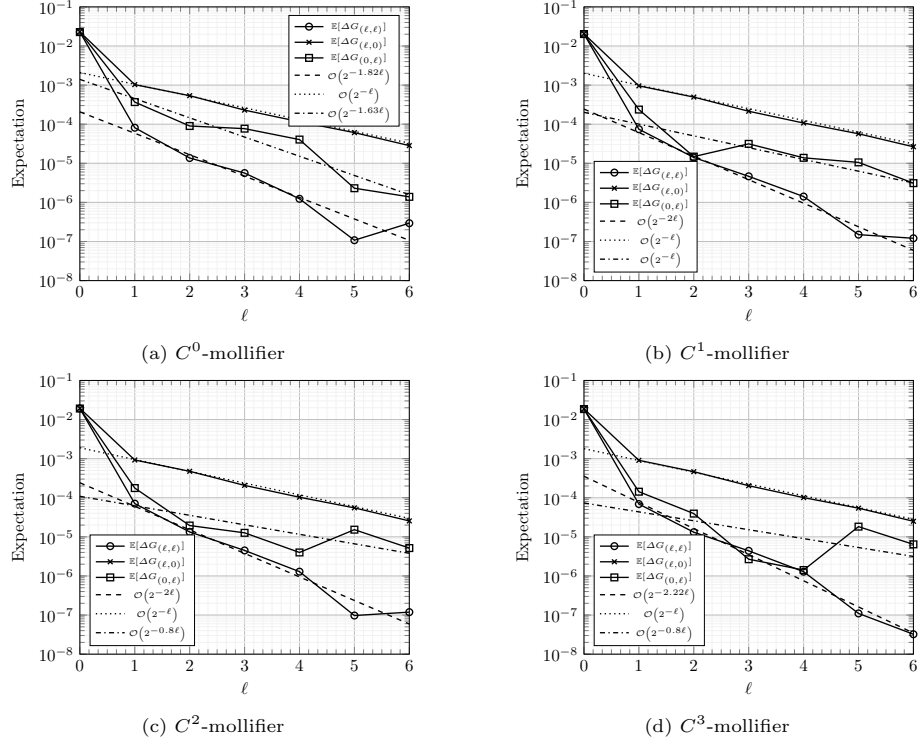


Figure D.6: Kuramoto example, numerical rate verification: Mixed-difference expectation convergence for mollifiers of varying regularity of the indicator function for $K = 2$ and mollification parameter $\varepsilon = \frac{1}{2}$. Plots indicate that C^1 -regularity suffices for multiplicative convergence of mixed-difference expectations, satisfying Assumption 1.

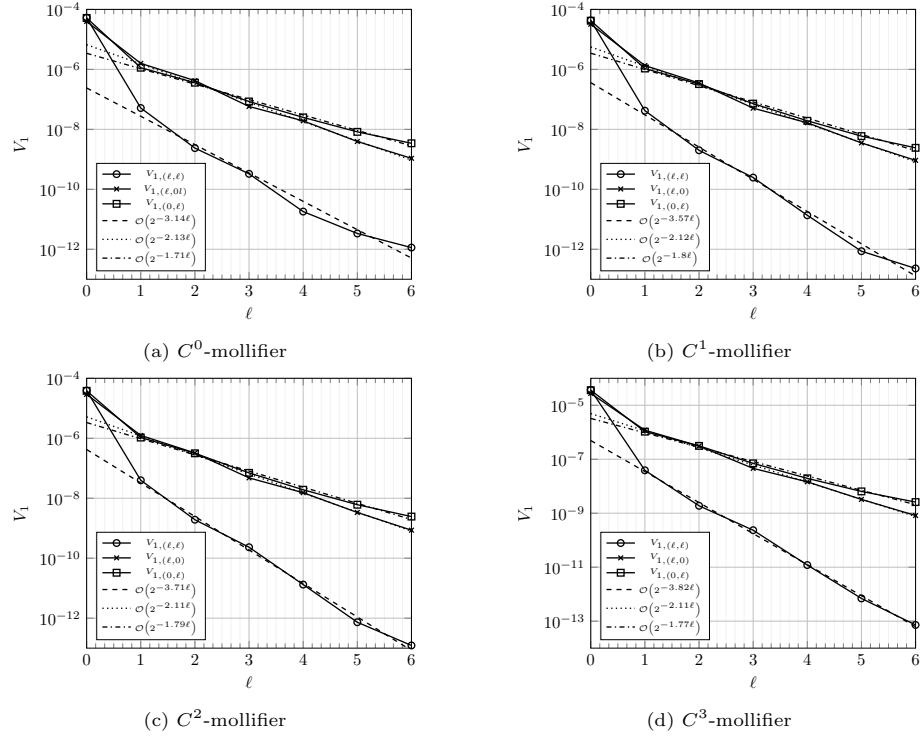


Figure D.7: Kuramoto example, numerical rate verification: Mixed-difference $V_{1,\alpha}$ convergence for mollifiers of varying regularity of the indicator function for $K = 2$ and mollification parameter $\varepsilon = \frac{1}{2}$. Plots indicate that C^2 -regularity suffices for multiplicative convergence of mixed-difference variances, satisfying Assumption 2.

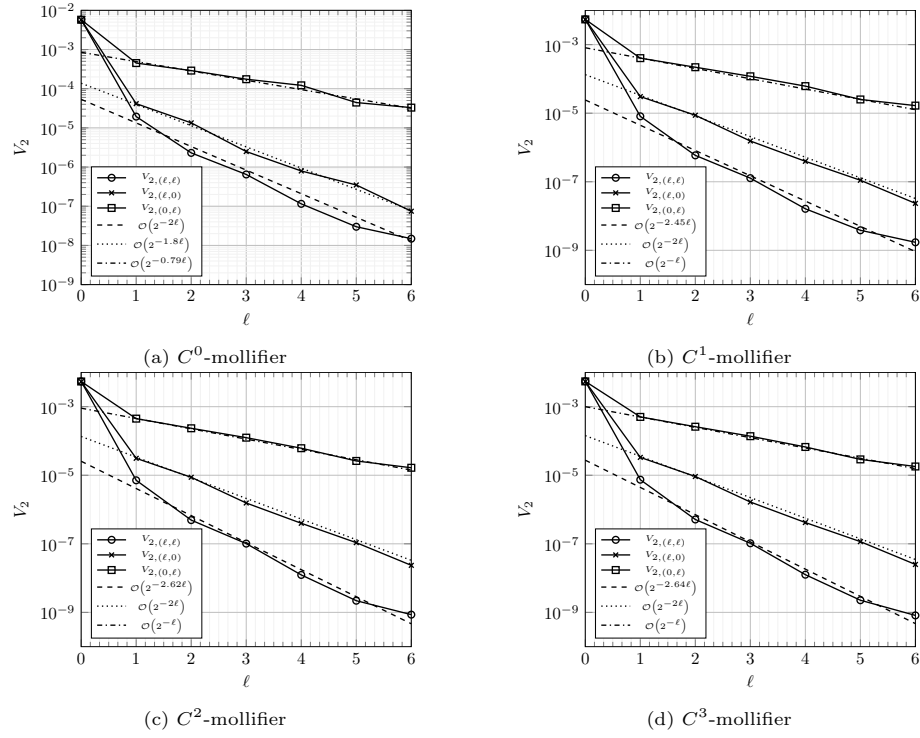


Figure D.8: Kuramoto example, numerical rate verification: Mixed-difference $V_{2,\alpha}$ convergence for mollifiers of varying regularity of the indicator function for $K = 2$ and mollification parameter $\varepsilon = \frac{1}{2}$. Plots indicate that C^2 -regularity suffices for multiplicative convergence of mixed-difference variances, satisfying Assumption 2.

Declaration of generative AI and AI-assisted technologies in the manuscript preparation process: During the preparation of this work, the authors used Anthropic Claude and OpenAI ChatGPT as editing tools. The authors reviewed and edited the output as needed and take full responsibility for the content of the published article.

References

- [1] N. Ben Rached, A.-L. Haji-Ali, S. M. Subbiah Pillai, R. Tempone, Double-loop importance sampling for McKean–Vlasov stochastic differential equation, *Statistics and Computing* 34 (6) (2024) 197.
- [2] H. P. McKean Jr, A class of Markov processes associated with nonlinear parabolic equations, *Proceedings of the National Academy of Sciences* 56 (6) (1966) 1907–1911.
- [3] D. Helbing, P. Molnár, Social force model for pedestrian dynamics, *Physical Review E* 51 (5) (1995) 4282–4286.
- [4] R. Erban, J. Haskovec, From individual to collective behaviour of coupled velocity jump processes: A locust example, *Kinetic and Related Models* 5 (4) (2012) 817–842.
- [5] U. Dobramysl, S. Rüdiger, R. Erban, Particle-based multiscale modeling of calcium puff dynamics, *Multiscale Modeling & Simulation* 14 (3) (2016) 997–1016.
- [6] N. Bush, B. M. Hambly, H. Haworth, L. Jin, C. Reisinger, Stochastic evolution equations in portfolio credit modelling, *SIAM Journal on Financial Mathematics* 2 (1) (2011) 627–664.
- [7] J. A. Acebrón, L. L. Bonilla, C. J. P. Vicente, F. Ritort, R. Spigler, The Kuramoto model: A simple paradigm for synchronization phenomena, *Reviews of modern physics* 77 (1) (2005) 137.
- [8] A.-S. Sznitman, Topics in propagation of chaos, *Lecture notes in mathematics* (1991) 165–251.
- [9] A.-L. Haji-Ali, R. Tempone, Multilevel and multi-index Monte Carlo methods for the McKean–Vlasov equation, *Statistics and Computing* 28 (2018) 923–935.
- [10] Y. Li, X. Mao, Q. Song, F. Wu, G. Yin, Strong convergence of Euler–Maruyama schemes for McKean–Vlasov stochastic differential equations under local Lipschitz conditions of state variables, *IMA Journal of Numerical Analysis* 43 (2) (2023) 1001–1035.
- [11] D. P. Kroese, T. Taimre, Z. I. Botev, *Handbook of Monte Carlo methods*, John Wiley & Sons, 2013.

- [12] G. dos Reis, G. Smith, P. Tankov, Importance sampling for McKean-Vlasov SDEs, *Applied Mathematics and Computation* 453 (2023) 128078.
- [13] N. Ben Rached, A.-L. Haji-Ali, S. M. Subbiah Pillai, R. Tempone, Multilevel importance sampling for rare events associated with the McKean-Vlasov equation, *Statistics and Computing* 35 (1) (2024) 1.
- [14] A.-L. Haji-Ali, F. Nobile, R. Tempone, Multi-index Monte Carlo: when sparsity meets sampling, *Numerische Mathematik* 132 (2016) 767–806.
- [15] A. Kebaier, J. Lelong, Coupling importance sampling and multilevel Monte Carlo using sample average approximation, *Methodology and Computing in Applied Probability* 20 (2018) 611–641.
- [16] W. Fang, M. B. Giles, Multilevel Monte Carlo method for ergodic SDEs without contractivity, *Journal of Mathematical Analysis and Applications* 476 (1) (2019) 149–176.
- [17] C. Ben Hammouda, N. Ben Rached, R. Tempone, Importance sampling for a robust and efficient multilevel Monte Carlo estimator for stochastic reaction networks, *Statistics and Computing* 30 (2020) 1665–1689.
- [18] M. Ben Alaya, K. Hajji, A. Kebaier, Adaptive importance sampling for multilevel Monte Carlo Euler method, *Stochastics* 95 (2) (2023) 303–327.
- [19] Y. Mishura, A. Veretennikov, Existence and uniqueness theorems for solutions of McKean–Vlasov stochastic equations, *Theory of Probability and Mathematical Statistics* 103 (2020) 59–101.
- [20] W. R. P. Hammersley, D. Šiška, L. Szpruch, Weak existence and uniqueness for mckean–vlasov sdes with common noise, *The Annals of Probability* 49 (2) (2021) 527–555.
- [21] P. E. Chaudru de Raynal, Strong well posedness of mckean–vlasov stochastic differential equations with hölder drift, *Stochastic Processes and their Applications* 130 (1) (2020) 79–107.
- [22] P.-E. Chaudru de Raynal, N. Frikha, Well-posedness for some non-linear sdes and related pde on the wasserstein space, *Journal de Mathématiques Pures et Appliquées* 159 (2022) 1–167.
- [23] R. Buckdahn, J. Li, S. Peng, C. Rainer, Mean-field stochastic differential equations and associated PDEs, *The Annals of Probability* 45 (2) (2017) 824–878.
- [24] S. Méléard, Asymptotic behaviour of some interacting particle systems; McKean-Vlasov and Boltzmann models, in: D. Talay, L. Tubaro (Eds.), *Probabilistic Models for Nonlinear Partial Differential Equations*, Vol. 1627, Springer, 1996, pp. 42–95.

- [25] M. Bossy, D. Talay, Convergence rate for the approximation of the limit law of weakly interacting particles: application to the Burgers equation, *The Annals of Applied Probability* 6 (3) (1996) 818–861.
- [26] M. Bossy, D. Talay, A stochastic particle method for the McKean-Vlasov and the Burgers equation, *Mathematics of computation* 66 (217) (1997) 157–192.
- [27] D. Cumin, C. Unsworth, Generalising the Kuramoto model for the study of neuronal synchronisation in the brain, *Physica D: Nonlinear Phenomena* 226 (2) (2007) 181–196.
- [28] D. M. Forrester, Arrays of coupled chemical oscillators, *Scientific reports* 5 (1) (2015) 16994.
- [29] Ł. Szpruch, A. Tse, Antithetic multilevel sampling method for nonlinear functionals of measure, *The Annals of Applied Probability* 31 (3) (2021) 1100–1139.
- [30] D. Talay, L. Tubaro, Expansion of the global error for numerical schemes solving stochastic differential equations, *Stochastic Analysis and Applications* 8 (4) (1990) 483–509.
- [31] S. Mischler, C. Mouhot, B. Wennberg, A new approach to quantitative propagation of chaos for drift, diffusion and jump processes, *Probability Theory and Related Fields* 161 (1) (2015) 1–59.
- [32] O. Bencheikh, B. Jourdain, Bias behaviour and antithetic sampling in mean-field particle approximations of SDEs nonlinear in the sense of McKean, *ESAIM: ProcS* 65 (2019) 219–235.
- [33] J.-F. Chassagneux, Ł. Szpruch, A. Tse, Weak quantitative propagation of chaos via differential calculus on the space of measures, *The Annals of Applied Probability* 32 (3) (2022) 1929–1969.
- [34] A.-L. Haji-Ali, H. Hoel, R. Tempone, Weak convergence analysis in the particle limit of the McKean–Vlasov equation using stochastic flows of particle systems, *IMA Journal of Applied Mathematics* (2025) hxaf015.
- [35] C. Pflaum, A. Zhou, Error analysis of the combination technique, *Numerische Mathematik* 84 (2) (1999) 327–350.
- [36] P. E. Kloeden, E. Platen, *Numerical Solution of Stochastic Differential Equations*, Springer, Berlin, 1992. doi:<https://doi.org/10.1007/978-3-662-12616-5>.
- [37] M. B. Giles, T. Nagapetyan, K. Ritter, Multilevel Monte Carlo approximation of distribution functions and densities, *SIAM/ASA Journal on Uncertainty Quantification* 3 (1) (2015) 267–295. doi:[10.1137/140960086](https://doi.org/10.1137/140960086).

- [38] A. Hinrichs, P. Kritzer, F. Pillichshammer (Eds.), *MLMC Techniques for Discontinuous Functions*, Springer International Publishing, Cham, 2024.
- [39] A. Szepessy, R. Tempone, G. E. Zouraris, Adaptive weak approximation of stochastic differential equations, *Communications on Pure and Applied Mathematics* 54 (10) (2001) 1169–1214.
- [40] E. Ben Amar, N. Ben Rached, R. Tempone, Hierarchical importance sampling for estimating occupation time for SDE solutions, *arXiv preprint arXiv:2509.13950* (2025).
- [41] C. Hartmann, C. Schütte, W. Zhang, Model reduction algorithms for optimal control and importance sampling of diffusions, *Nonlinearity* 29 (8) (2016) 2298.
- [42] C. Hartmann, O. Kebiri, L. Neureither, L. Richter, Variational approach to rare event simulation using least-squares regression, *Chaos: An Interdisciplinary Journal of Nonlinear Science* 29 (6) (2019) 063107.
- [43] B. N. Khoromskij, Tensor numerical methods for multidimensional PDEs: theoretical analysis and initial applications, *ESAIM: Proc.* 48 (2015) 1–28.
- [44] J. Han, A. Jentzen, W. E, Solving high-dimensional partial differential equations using deep learning, *Proceedings of the National Academy of Sciences* 115 (34) (2018) 8505–8510.
- [45] F. Cucker, S. Smale, On the mathematics of emergence, *Japanese Journal of Mathematics* 2 (1) (2007) 197–227.
- [46] P. Degond, J.-G. Liu, S. Merino-Aceituno, T. Tardiveau, Continuum dynamics of the intention field under weakly cohesive social interaction, *Mathematical Models and Methods in Applied Sciences* 27 (01) (2016) 159–182.
- [47] A.-L. Haji-Ali, F. Nobile, E. von Schwerin, R. Tempone, Optimization of mesh hierarchies in multilevel Monte Carlo samplers, *Stochastics and Partial Differential Equations Analysis and Computations* 4 (1) (2016) 76–112.
- [48] N. Collier, A.-L. Haji-Ali, F. Nobile, E. Von Schwerin, R. Tempone, A continuation multilevel Monte Carlo algorithm, *BIT Numerical Mathematics* 55 (2015) 399–432.