Antithetic multilevel sampling method for non-linear functionals of measure

Łukasz Szpruch *[1,2] and Alvin Tse[1]

[1]School of Mathematics, University of Edinburgh
[2]The Alan Turing Institute, London

Abstract

Let $\mu \in \mathcal{P}_2(\mathbb{R}^d)$, where $\mathcal{P}_2(\mathbb{R}^d)$ denotes the space of square integrable probability measures, and consider a Borel-measurable function $\Phi : \mathcal{P}_2(\mathbb{R}^d) \to \mathbb{R}$. In this paper we develop Antithetic Monte Carlo estimator (A-MLMC) for $\Phi(\mu)$, which achieves sharp error bound under mild regularity assumptions. The estimator takes as input the empirical laws $\mu^N = \frac{1}{N}\sum_{i=1}^N \delta_{X_i}$, where a) $(X_i)_{i=1}^N$ is a sequence of i.i.d samples from μ or b) $(X_i)_{i=1}^N$ is a system of interacting particles (diffusions) corresponding to a McKean-Vlasov stochastic differential equation (McKV-SDE). Each case requires a separate analysis. For a mean-field particle system, we also consider the empirical law induced by its Euler discretisation which gives a fully implementable algorithm. As by-products of our analysis, we establish a dimension-independent rate of uniform *strong propagation of chaos*, as well as an L^2 estimate of the antithetic difference for i.i.d. random variables corresponding to general functionals defined on the space of probability measures.

1 Introduction

The convergence of the empirical law μ^N to its limit μ for linear functionals of measure (i.e. $F(\mu) = \int_{\mathbb{R}^d} f(x)\mu(dx)$ for some function $f : \mathbb{R}^d \to \mathbb{R}$) is rather well understood in the literature. Indeed, $F(\mu^N)$ is an unbiased estimator of $F(\mu)$ and in the i.i.d. case, the classical central limit theorems provide sharp error bounds. However, for general non-linear functionals of measure $\Phi : \mathcal{P}_2(\mathbb{R}^d) \to \mathbb{R}$, $\Phi(\mu^N)$ is typically a biased estimator of $\Phi(\mu)$ and hence when seeking an optimal estimator, more sophisticated techniques are needed. For example, in the context of nested Monte Carlo estimators, with $F(\mu) = R(\int_{\mathbb{R}^d} f(x)\mu(dx))$ and $R : \mathbb{R} \to \mathbb{R}$ being nonlinear, the multilevel Monte Carlo (MLMC) [32, 21] and antithetic multilevel Monte Carlo (A-MLMC) [22] estimators are more efficient than $F(\mu^N)$. In this work, we study the general case of functionals of measure, which are sufficiently

*This was work has been supported by The Alan Turing Institute under the Engineering and Physical Sciences Research Council grant EP/N510129/1.

smooth in an appropriate sense. Most importantly, we do not rely on specific structural assumptions imposed on $\Phi(\mu)$. Our goal is to find an estimator $\mathcal{A}$ that approximates $\Phi(\mu)$. We are interested in sharp estimates of its corresponding mean-square error[1] $\mathbb{E}[(\Phi(\mu) - \mathcal{A})^2]$. The multilevel Monte Carlo approach provides a very efficient strategy when one aims to find an implementable algorithm that achieves a sharp upper bound for the mean-square error for a given computational cost (in the i.i.d case, cost can be defined as the number of random numbers needed to be generated to compute $\mathcal{A}$). We shall consider the analysis of mean-square error via antithetic Monte Carlo techniques in two different cases: i.i.d. samples and interacting diffusions. It is also worth mentioning that the rates of convergence studied in this work are independent of the dimension, under sufficient regularity of the functions concerned. If we only assume that Φ is Lipschitz continuous with respect to the Wasserstein distance, i.e, there exists a constant $C > 0$ such that $|\Phi(\mu) - \Phi(\nu)| \leq CW_2(\mu, \nu)$, for all $\mu, \nu \in \mathcal{P}_2(\mathbb{R}^d)$, one could bound $|\Phi(\mu) - \mathbb{E}\Phi(\mu_N)|$ by $\mathbb{E}W_2(\mu, \mu_N)$. Consequently, following [16] or [15], the rate of convergence in the number of samples N deteriorates as the dimension d increases. We also refer the reader to recent works [1, 39, 23] that study the problem from the perspective of Monge-Ampére PDEs. On the other hand, recently, authors in [14] (in Lemma 5.10) observed that if the functional Φ is twice-differentiable with respect to the functional derivative (see Appendix A for its definition), then one can obtain a dimension-independent bound for the strong error $\mathbb{E}|\Phi(\mu) - \Phi(\mu_N)|^4$, which is of order $O(N^{-1/2})$ (as expected by CLT).

Recap of MLMC. Fix $\Phi : \mathcal{P}_2(\mathbb{R}^d) \to \mathbb{R}$ and $L > 0$. We seek to approximate $\Phi(\mu)$. In order to do so, we approximate measure μ with measures corresponding to different levels $\ell \in \{0, \dots, L\}$, where each ℓ corresponds to some level of approximation. As ℓ increases, the bias decreases and the corresponding computational cost increases. One therefore faces an optimisation problem and tries to obtain the minimal computation cost for a given accuracy (or, equivalently, to minimise the bias for a fixed computational budget). It turns out, perhaps surprisingly, that the MLMC estimator, which consists of a hierarchy of biased approximations, can achieve computational efficiency (or optimal scaling of variance) of vanilla Monte Carlo built from directly accessible unbiased samples [19, 20].

Let us consider a sequence of probability spaces $(\Omega^\ell, \mathcal{F}^\ell, \mathbb{P}^\ell)_{\ell=1}^L$. For each level ℓ, we consider M_ℓ copies of $(\Omega^\ell, \mathcal{F}^\ell, \mathbb{P}^\ell)$, that is $(\Omega^{\ell,\theta}, \mathcal{F}^{\ell,\theta}, \mathbb{P}^{\ell.\theta})_{\theta=1}^{M_\ell}$. For each level $\ell \in \{0, \dots, L\}$ and cloud $\theta \in \{1, \dots, M_\ell\}$, we generate N_ℓ samples (depending on the two cases: i.i.d. or interacting diffusion) and construct empirical measure $\mu^{N_\ell,\theta,\ell}$. For simplicity, we set $N_\ell = 2^\ell$ and let $(M_\ell)_{\ell=0}^L$ be a sequence of non-increasing natural numbers. The MLMC estimator is defined as

$$\mathcal{A}^{\text{MLMC}} := \frac{1}{M_0} \sum_{\theta=1}^{M_0} \Phi(\mu^{N_0,\theta,0}) + \sum_{\ell=1}^{L} \left[\frac{1}{M_\ell} \sum_{\theta=1}^{M_\ell} \left[\Phi(\mu^{N_\ell,\theta,\ell}) - \Phi(\mu^{N_{\ell-1},\theta,\ell}) \right] \right]. \tag{1.1}$$

Note that, for $\ell \in \{1, \dots, L\}$, $\mathbb{E}\big[\Phi(\mu^{N_{\ell-1},\theta,\ell})\big] = \mathbb{E}\big[\Phi(\mu^{N_{\ell-1},\theta,\ell-1})\big]$. Therefore, taking the expectation on both sides of (1.1) gives a telescopic sum, which simplifies to

$$\mathbb{E}\big[\mathcal{A}^{\text{MLMC}}\big] = \mathbb{E}\big[\Phi(\mu^{N_L,1,L})\big].$$

[1] We look at the mean-square error for simplicity, but a similar computation could be done to verify the Lindeberg condition and produce CLT with an appropriate scaling.

If the variance between successive approximations converges to zero as the level increases, MLMC reduces the computational cost of simulation by carefully combining many simulations on low levels with low accuracy (at a corresponding low cost); with relatively few simulations on high levels with low accuracy (and at a high cost). The idea has been independently developed by Giles and Heinrich [20, 25, 29] (see also 2-level Monte Carlo of Kebaier [29]) in the context of temporal approximation of SDEs and parametric integration.

The key challenge in developing MLMC estimators is the construction of suitable coupling between $(\mu^{N_\ell,\theta,\ell}, \mu^{N_{\ell-1},\theta,\ell})$ that ensures a quick decay of the variance of the estimator across the levels. With this in mind, in this work, we develop an antithetic extension of the MLMC algorithm (A-MLMC) defined by

$$\begin{aligned}\mathcal{A}^{\text{A-MLMC}} &:= \frac{1}{M_0}\sum_{\theta=1}^{M_0} \Phi(\mu^{N_0,\theta,0}) \\ &+ \sum_{\ell=1}^{L}\Bigg[\frac{1}{M_\ell}\sum_{\theta=1}^{M_\ell}\Big[\underbrace{\Phi(\mu^{N_\ell,\theta,\ell})}_{:=\Phi^{f,\ell}} - \underbrace{\frac{1}{2}\Big(\Phi(\mu^{N_\ell,(1),\theta,\ell}) + \Phi(\mu^{N_\ell,(2),\theta,\ell})\Big)}_{:=\Phi^{c,\ell}}\Big]\Bigg], \end{aligned} \tag{1.2}$$

where $\mu^{N_\ell,(1),\theta,\ell}$ and $\mu^{N_\ell,(2),\theta,\ell}$ correspond respectively to empirical measures corresponding to the first half and second half of N_ℓ samples (to be defined precisely later for the two cases separately), such that $\mathbb{E}\big[\Phi(\mu^{N_\ell,\theta,\ell})\big] = \mathbb{E}\big[\Phi(\mu^{N_\ell,(1),\theta,\ell})\big] = \mathbb{E}\big[\Phi(\mu^{N_\ell,(2),\theta,\ell})\big]$. Let $\text{Cost}\big(\Phi^{f,\ell} - \Phi^{c,\ell}\big)$ denote the computational cost of computing the estimator $\Phi^{f,\ell} - \Phi^{c,\ell}$. Theorem 1 in [12] gives a result concerning the complexity of antithetic MLMC. Let $\alpha, \beta, \gamma \in \mathbb{R}$ be positive constants such that $\alpha \geq \frac{1}{2}\min\{\beta,\gamma\}$. Suppose that

$$i)\ \big|\mathbb{E}\big[\Phi^{f,\ell} - \Phi^{c,\ell}\big]\big| \lesssim N_\ell^{-\alpha}, \quad ii)\ \text{Var}\big[\Phi^{f,\ell} - \Phi^{c,\ell}\big] \lesssim N_\ell^{-\beta}, \quad iii)\ \text{Cost}\big(\Phi^{f,\ell} - \Phi^{c,\ell}\big) \lesssim N_\ell^{\gamma}. \tag{1.3}$$

Then there exist L and a sequence $(M_\ell)_{\ell=1}^L$ such that estimator (1.2) has a bound

$$\mathbb{E}[(\Phi(\mu) - \mathcal{A}^{\text{A-MLMC}})^2] \lesssim \epsilon^2 ,$$

with a computational cost of $\mathcal{A}^{\text{A-MLMC}}$ of order ϵ^{-2} provided $\beta > \gamma$. The core part of this work is to establish that indeed the condition $\beta > \gamma$ holds. We also remark that ϵ^{-2} is the computational cost of a Monte Carlo estimator that achieves mean-square-error of the order ϵ^2 and has access to unbiased i.i.d. samples.

Notations. Throughout this article, we denote the Hilbert-Schmidt norm of any matrix by $\|\cdot\|$ and denote the standard Euclidean inner product $x \cdot y$ by xy. Also, $\mathscr{L}(\xi)$ denotes the law of ξ, for any square-integrable random variable ξ. $\mathcal{P}_k(\mathbb{R}^d)$ denotes the set of probability measures with finite kth moment. $C^k_{b,\text{Lip}}((\mathbb{R}^d)^\ell)$ denotes the set of all functions from $(\mathbb{R}^d)^\ell$ to $\mathbb{R}$ that are in C^k with bounded and Lipschitz partial derivatives up to and including order k. For any $a, b \geq 0$, we denote by $a \lesssim b$ if $a \leq Cb$, for some constant $C > 0$ that does not depend on N, h or ϵ. Finally, unless otherwise specified, C denotes a generic constant that does not depend on N, h or ϵ, whose value may vary from line to line. For any probability space $(\Omega, \mathcal{F}, \mathbb{P})$ with expectation operator $\mathbb{E}$ and a random variable ξ, we construct a separate probability space $(\widetilde{\Omega}, \widetilde{\mathcal{F}}, \widetilde{\mathbb{P}})$ with expectation operator $\widetilde{\mathbb{E}}$ and a random variable $\widetilde{\xi}$ with the same law as ξ.

2 Main results

2.1 A-MLMC for i.i.d samples

For every $0 \le \ell \le L$ and $1 \le \theta \le M_\ell$, we consider i.i.d. random variables $\{X_{i,\ell,\theta}\}_{1\le i\le N_\ell}$ with law $\mu \in \mathcal{P}_2(\mathbb{R}^d)$. The corresponding antithetic MLMC estimator for $\Phi(\mu)$ is given by (1.2), where

$$\mu^{N_\ell,\theta,\ell} := \frac{1}{N_\ell}\sum_{i=1}^{N_\ell} \delta_{X_{i,\ell,\theta}}$$

is the empirical measure corresponding to each of the $\sum_{\ell=0}^{L} M_\ell$ independent clouds of samples indexed by $\theta \in \{1,\dots,M_\ell\}$ and $\ell \in \{0,\dots,L\}$. Moreover,

$$\mu^{N_\ell,(1),\theta,\ell} := \frac{1}{N_\ell/2}\sum_{i=1}^{N_\ell/2} \delta_{X_{i,\ell,\theta}}, \qquad \mu^{N_\ell,(2),\theta,\ell} := \frac{1}{N_\ell/2}\sum_{i=N_\ell/2+1}^{N_\ell} \delta_{X_{i,\ell,\theta}}.$$

We illustrate the power of antithetic MLMC for i.i.d. samples through the following example in which there is a linear dependence on the measure in Φ. For simplicity of notations, we set

$$\mu^N := \mu^{N,1,0}, \qquad \mu^{N,(1)} := \mu^{N,(1),1,0}, \qquad \mu^{N,(2)} := \mu^{N,(2),1,0}. \tag{2.1}$$

Example 2.1. Consider $\Phi(\mu) := \int_{\mathbb{R}^d} F(x)\mu(dx)$, where $F : \mathbb{R}^d \to \mathbb{R}$ has linear growth. We already observed that $\mathbb{E}[\mathcal{A}^{\text{MLMC}}] = \mathbb{E}[\mathcal{A}^{\text{A- MLMC}}]$. The postulated independence conditions imply that

$$\mathbb{V}ar[\mathcal{A}^{\text{MLMC}}] = \frac{\mathbb{V}ar[\Phi(\mu^{N_0})]}{M_0} + \sum_{\ell=1}^{L} \frac{\mathbb{V}ar[\Phi(\mu^{N_\ell}) - \Phi(\mu^{N_{\ell-1}})]}{M_\ell}.$$

On the other hand,

$$\mathbb{V}ar[\mathcal{A}^{\text{A-MLMC}}] = \frac{\mathbb{V}ar[\Phi(\mu^{N_0})]}{M_0} + \sum_{\ell=1}^{L} \frac{\mathbb{V}ar\Big[\Phi(\mu^{N_\ell}) - \frac{1}{2}\big(\Phi(\mu^{N_\ell,(1)}) + \Phi(\mu^{N_\ell,(2)})\big)\Big]}{M_\ell}.$$

It is clear that the efficiency of this algorithm hinges on good coupling estimates that result in small variances across levels ℓ. Set $N_\ell := 2N_{\ell-1}$. For $\mathcal{A}^{\text{MLMC}}$, we have

$$\begin{aligned}\mathbb{V}ar[\Phi(\mu^{N_\ell}) - \Phi(\mu^{N_{\ell-1}})] &= \mathbb{V}ar\left[\left(\frac{1}{N_\ell} - \frac{1}{N_{\ell-1}}\right)\sum_{i=1}^{N_{\ell-1}} F(X_i) + \frac{1}{N_\ell}\sum_{i=N_{\ell-1}+1}^{N_\ell} F(X_i)\right] \\ &= \left(\frac{1}{N_\ell} - \frac{1}{N_{\ell-1}}\right)^2 \sum_{i=1}^{N_{\ell-1}} \mathbb{V}ar[F(X_i)] + \left(\frac{1}{N_\ell}\right)^2 \sum_{i=N_{\ell-1}+1}^{N_\ell} \mathbb{V}ar[F(X_i)] = O(\frac{1}{N_\ell}).\end{aligned}$$

On the other hand, for A-MLMC, we have

$$\mathbb{V}ar\Big[\Phi(\mu^{N_\ell}) - \frac{1}{2}\big(\Phi(\mu^{N_\ell,(1)}) + \Phi(\mu^{N_\ell,(2)})\big)\Big] = 0.$$

2.1.1 Complexity analysis of A-MLMC for i.i.d. samples

The above example is indeed a very special case. This work explores regularity conditions of functionals Φ that lead to a reduction in variance of the antithetic difference for general functions of measures. This result is formulated in terms of the class $\mathcal{M}_k^L$ of k times differentiable functions in the sense of linear functional derivatives. (See Definition A.4 for its precise meaning. See also Definition A.3 for the class $\mathcal{M}_k$ of k times differentiable functions in the sense of L-derivatives that will be used in other theorems).

Recall that, from Section 9 in [20], the second moment of the antithetic difference given by

$$U(\mu^{2N}) - \frac{1}{2}\big(U(\mu^{2N,(1)}) + U(\mu^{2N,(2)})\big)$$

converges to 0 in the rate $O(1/N^2)$, for functions $U : \mathcal{P}_2(\mathbb{R}^d) \to \mathbb{R}$ of the form

$$U(\mu) := F\bigg(\int_{\mathbb{R}^d} G(x)\,\mu(dx)\bigg), \tag{2.2}$$

where $G : \mathbb{R}^d \to \mathbb{R}$ is an integrable function and $F : \mathbb{R} \to \mathbb{R}$ is a twice-differentiable function with bounded derivatives. The following theorem generalises the structure of U from the aforementioned result, under a set of different conditions.

Theorem 2.2 (Theorem 3.3, antithetic error on i.i.d. random variables)**.** *Suppose that $\mu \in \mathcal{P}_8(\mathbb{R}^d)$ and $U \in \mathcal{M}_4^L(\mathcal{P}_2(\mathbb{R}^d))$. Then there exists a constant $C > 0$ such that*

$$\mathbb{E}\big|U(\mu^{2N}) - \frac{1}{2}\big(U(\mu^{2N,(1)}) + U(\mu^{2N,(2)})\big)\big|^2 \leq \frac{C}{N^2}.$$

The proof of the theorem can be found in Section 3. Theorem 3.3 allow us to establish bounds on the variance in (1.3), that is we have

$$\mathbb{V}ar\Big[\Phi(\mu^{N_\ell}) - \frac{1}{2}\big(\Phi(\mu^{N_\ell,(1)}) + \Phi(\mu^{N_\ell,(2)})\big)\Big] = O(\frac{1}{N_\ell^2}). \tag{2.3}$$

By Theorem 2.11 in [11], under the same hypothesis on Φ, we also have

$$|\mathbb{E}[\Phi(\mu^{N_\ell})] - \Phi(\mu)| \leq O(\frac{1}{N_\ell}). \tag{2.4}$$

Finally, since the empirical measures μ^{N_ℓ}, $\mu^{N_\ell,(1)}$ and $\mu^{N_\ell,(2)}$ correspond to i.i.d. random variables, the cost of simulating the antithetic difference is given by

$$\text{Cost}\Big[\Phi(\mu^{N_\ell}) - \frac{1}{2}\big(\Phi(\mu^{N_\ell,(1)}) + \Phi(\mu^{N_\ell,(2)})\big)\Big] = O(N_\ell). \tag{2.5}$$

Hence, by combining (2.3), (2.4) and (2.5), along with Theorem 1 in [12], we have the following result regarding the complexity of the antithetic MLMC estimator.

Theorem 2.3. *Let $\mu \in \mathcal{P}_8(\mathbb{R}^d)$ and $\Phi \in \mathcal{M}_4^L$. Then, for the mean-square error $\mathbb{E}[(\Phi(\mu) - \mathcal{A}^{\text{A-MLMC}})^2]$ to be of the order $O(\epsilon^2)$, there exist L and a sequence $(M_\ell)_{\ell=1}^L$ such that the computational cost of $\mathcal{A}^{\text{A-MLMC}}$ is of the order $O(\epsilon^{-2})$.*

2.2 A-MLMC for interacting diffusions

The second situation we treat in this work concerns estimates of propagation-of-chaos type for the system of McKean-Vlasov SDEs (McKV-SDEs). Building on regularity results recently obtained in [11], we extend the analysis of the i.i.d. case presented above to interacting particle systems. To be more precise, fix $T > 0$ and let $\{W_t\}_{t\in[0,T]}$ be a d-dimensional Brownian motion on a filtered probability space $(\Omega, \{\mathcal{F}_t\}_t, \mathcal{F}, \mathbb{P})$. Next, we consider functions $b : \mathbb{R}^d \times \mathcal{P}_2(\mathbb{R}^d) \to \mathbb{R}^d$, $\sigma : \mathbb{R}^d \times \mathcal{P}_2(\mathbb{R}^d) \to \mathbb{R}^d \otimes \mathbb{R}^d$, and consider the corresponding McKV-SDE given by

$$\begin{cases} dX_t = \xi + \int_0^t b(X_s, \mu_s^X)\, ds + \int_0^t \sigma(X_s, \mu_s^X)\, dW_s, \qquad t \in [0,T], \\ \mu_s^X := \text{Law}(X_s), \end{cases} \tag{2.6}$$

where $\xi \sim \nu \in \mathcal{P}_2(\mathbb{R}^d)$. The theory of propagation of chaos, [38], shows that (2.6) arises as a limiting equation of the system of interacting diffusions (particles) $\{Y_t^{i,N}\}_{i=1,\ldots,N}$ on $(\mathbb{R}^d)^N$ given by

$$\begin{cases} Y_t^{i,N} = \xi_i + \int_0^t b(Y_s^{i,N}, \mu_s^N)\, ds + \int_0^t \sigma(Y_s^{i,N}, \mu_s^N)\, dW_s^i, \quad 1 \le i \le N, \quad t \in [0,T], \\ \mu_s^N := \frac{1}{N}\sum_{i=1}^N \delta_{Y_s^{i,N}}, \end{cases} \tag{2.7}$$

where $W^i, 1 \le i \le N$, are independent d-dimensional Brownian motions and $\xi_i, 1 \le i \le N$, are i.i.d. random variables with law $\nu \in \mathcal{P}_2(\mathbb{R}^d)$. We refer the reader to [18, 38, 34] for the classical results in this direction and to [28, 2, 17, 35, 31] for more recent theory. Most of the results in the literature provide non-quantitative propagation of chaos with a few notable exceptions. In the case where the coefficients of (2.6) are linear in measure and globally Lipschitz continuous, [38] showed that $W_2(\mathscr{L}(Y_t^{i,N}), \mathscr{L}(X_t)) = O(N^{-1/2})$. We refer to Sznitman's result as strong propagation of chaos. Note that, in this work, we treat the case of McKean-Vlasov SDEs with coefficients with a general dependence in measure. In the case of Lipschitz continuous dependence in measure in the 2-Wasserstein metric, the rate of strong propagation of chaos deteriorates with the dimension d, [8, Ch. 1]. We demonstrate that under regularity assumptions on b and σ in terms of L-derivatives, we have a strong error bound in fourth moment that is dimension-independent:

Theorem 2.4 (Theorem 4.2, uniform strong propagation of chaos). *Assume* (Int). *Suppose that* $b, \sigma \in \mathcal{M}_3(\mathbb{R}^d \times \mathcal{P}_2(\mathbb{R}^d))$. *Then*

$$\mathbb{E}\Big[W_{\mathcal{C}_T,2}(\mu^N, \mu^{X,N})^4\Big] \le \mathbb{E}\Big[\frac{1}{N}\sum_{i=1}^N \Big(\sup_{t\in[0,T]} \big|X_t^i - Y_t^{i,N}\big|^4\Big)\Big] \le \frac{C}{N^2},$$

for some constant $C > 0$.

In the case of interacting diffusions (McKV-SDEs), we wish to consider the corresponding antithetic MLMC estimator to estimate $\Phi(\mu_T^X)$. As before, we set $N_\ell := 2^\ell$. For each particle system

$\{Y^{i,2N}\}_{1\leq i\leq 2N}$, we define two sub-particle systems to have the same number of particles, each defined by

$$Y_t^{i,2N,(1)} = \xi_i + \int_0^t b\left(Y_r^{i,2N,(1)}, \mu_r^{2N,(1)}\right) dr + \int_0^t \sigma\left(Y_r^{i,2N,(1)}, \mu_r^{2N,(1)}\right) dW_r^i, \quad 1 \leq i \leq N,$$
$$Y_t^{i,2N,(2)} = \xi_i + \int_0^t b\left(Y_r^{i,2N,(2)}, \mu_r^{2N,(2)}\right) dr + \int_0^t \sigma\left(Y_r^{i,2N,(2)}, \mu_r^{2N,(2)}\right) dW_r^i, \quad N+1 \leq i \leq 2N,$$

where

$$\mu_r^{2N,(1)} := \frac{1}{N}\sum_{i=1}^{N} \delta_{Y_r^{i,2N,(1)}} \qquad \text{and} \qquad \mu_r^{2N,(2)} := \frac{1}{N}\sum_{i=N+1}^{2N} \delta_{Y_r^{i,2N,(2)}}. \tag{2.8}$$

Unlike the i.i.d. setting considered above, these particles are not independent. The corresponding antithetic MLMC estimator for $\Phi(\mu)$ is again given by (1.2), where $\mu_T^{N_\ell,\theta,\ell}$, $\mu_T^{N_\ell,(1),\theta,\ell}$ and $\mu_T^{N_\ell,(2),\theta,\ell}$ are defined similarly as $\mu_T^{N_\ell}$, $\mu_T^{N_\ell,(1)}$ and $\mu_T^{N_\ell,(2)}$ respectively, but correspond to the copies of probability spaces indexed by $\ell \in \{0,\dots,L\}$ and $\theta \in \{1,\dots,M_\ell\}$. Each probability space (indexed by ℓ, θ) supports particles with initial conditions $\xi_{i,\ell,\theta}$, $i \in \{1,\dots,N_\ell\}$, driven by Brownian motions $W^{i,\ell,\theta}$, $i \in \{1,\dots,N_\ell\}$.

2.2.1 Complexity analysis of A-MLMC for interacting diffusions

Our analysis of complexity relies heavily on the calculus on $(\mathcal{P}_2(\mathbb{R}^d), W_2)$ and we follow the approach presented by P. Lions in his course at Collège de France [33] (redacted by Cardaliaguet [6]). The important object in our study, similar to [7], is the PDE written on the space $[0,T]\times\mathcal{P}_2(\mathbb{R}^d)$, which corresponds to the lifted semigroup and comes from the Itô's formula of functionals of measures established in [5] and [10]. This line of research has been recently explored in [30, Ch. 9] and [36, Th. 2.1] to obtain results of quantitative propagation of chaos for a general family of particle systems. We shall adopt the notion of L-derivatives (see Appendix A), as well as the notion of the class $\mathcal{M}_k$ of kth order differentiable functions in the sense of L-derivatives (see Definition A.3).

To proceed with the analysis of complexity for interacting diffusions, we first define the notions of *pth order interactions* and *order of interactions*.

Definition 2.5 (Interacting kernels with pth order interactions)**.** b and σ are said to be of *pth order interactions* if they take the forms

$$b_i(x,\mu) = \int_{\mathbb{R}^d}\cdots\int_{\mathbb{R}^d} \overline{b_i}(x,y_1,\dots,y_p)\,\mu(dy_1)\dots\mu(dy_p), \tag{2.9}$$
$$\sigma_{i,j}(x,\mu) = \int_{\mathbb{R}^d}\cdots\int_{\mathbb{R}^d} \overline{\sigma_{i,j}}(x,y_1,\dots,y_p)\,\mu(dy_1)\dots\mu(dy_p), \tag{2.10}$$

where $\overline{b_i}, \overline{\sigma_{i,j}} : (\mathbb{R}^d)^{p+1} \to \mathbb{R}$ are continuous functions, for each $i,j \in \{1,\dots,d\}$.

Hence, in any particle system with b and σ given by (2.9) and (2.10), each particle interacts with N^p other particles. Since there are N particles in total, the number of interactions of the entire system is N^{p+1}.

For $i \in \{1, \dots, k\}$, let $S_k := \{Y^{i,N_k}\}_{1 \le i \le N_k}$ denote an interacting particle system. Then the *order of interactions* of an estimator composed of particle systems $S_1, \dots, S_k$ is defined by

$$\text{Order of interactions of estimator} := \sum_{j=1}^{k} \Big[\text{Order of interactions of particle system } S_j \Big].$$

In the subsequent analysis of this section, we assume that b and σ have the forms (2.9) and (2.10) respectively. We now compare the A-MLMC estimator with the ensemble estimator studied in [11]. The ensemble estimator corresponds to

$$Q_{M,N} := \frac{1}{M} \sum_{\theta=1}^{M} \Phi(\mu_T^{N,\theta}),$$

where $\mu_T^{N,\theta}$ denotes the empirical measure of the particles obtained for each i.i.d. cloud $\theta \in \{1, \dots, M\}$. If $\Phi \in \mathcal{M}_3$, $\overline{b_i} \in C^3_{b,\text{Lip}}((\mathbb{R}^d)^{p+1})$ (see Section 2.3 for its definition) and that $\overline{\sigma_{i,j}}$ belongs to $C^3_{b,\text{Lip}}((\mathbb{R}^d)^{p+1})$ and is uniformly bounded, then it follows by [11] that the number of interactions is of the order $O(\epsilon^{-2-p})$. By introducing Romberg extrapolation to the ensembles of particles [11, Sec 1.1 and Th 2.17], the number of interactions can be reduced to the order $O(\epsilon^{-2-p/k})$, under the assumption that $\Phi \in \mathcal{M}_{2k+1}$, $\overline{b_i} \in C^{2k+1}_{b,\text{Lip}}((\mathbb{R}^d)^{p+1})$ and that $\overline{\sigma_{i,j}}$ belongs to $C^{2k+1}_{b,\text{Lip}}((\mathbb{R}^d)^{p+1})$ and is uniformly bounded. It is proven in Theorem 5.2 that the A-MLMC estimator $\mathcal{A}^{\text{A-MLMC}}$ (almost) achieves an optimal order of interactions (for $p = 1$), whilst only requiring $\Phi \in \mathcal{M}_4$, $\overline{b_i} \in C^4_{b,\text{Lip}}((\mathbb{R}^d)^{p+1})$ and $\overline{\sigma_{i,j}} \in C^4_{b,\text{Lip}}((\mathbb{R}^d)^{p+1})$. The table below provides detailed comparison among the aforementioned methods.

Table 1: Comparison of the order of interactions for different estimators

Estimator	Order of interactions	Regularity assumption of $\overline{b_i}$	Regularity assumption of $\overline{\sigma_{i,j}}$	Regularity assumption of Φ
Ensembles of particles	$O(\epsilon^{-2-p})$	$C^3_{b,\text{Lip}}((\mathbb{R}^d)^{p+1})$	$C^3_{b,\text{Lip}}((\mathbb{R}^d)^{p+1})$ and uniformly bounded	$\mathcal{M}_3$
Romberg extrapolation (from [11, Sec 1.1 and Th 2.17])	$O(\epsilon^{-2-p/k})$	$C^{2k+1}_{b,\text{Lip}}((\mathbb{R}^d)^{p+1})$	$C^{2k+1}_{b,\text{Lip}}((\mathbb{R}^d)^{p+1})$ and uniformly bounded	$\mathcal{M}_{2k+1}$
Antithetic MLMC (for estimator (1.2)) in Theorem 5.2	$O(\epsilon^{-2}(\log \epsilon)^2)$, for $p = 1$, $O(\epsilon^{-1-p})$, for $p > 1$.	$C^4_{b,\text{Lip}}((\mathbb{R}^d)^{p+1})$	$C^4_{b,\text{Lip}}((\mathbb{R}^d)^{p+1})$	$\mathcal{M}_4$

The following theorems are the two main results concerning the antithetic MLMC estimator for interacting particle systems. The first theorem gives an analogue of Theorem 2.2 in the case of interacting particle systems.

Theorem 2.6 (Theorem 5.1, variance of antithetic difference)**.** *Assume* (Int). *Suppose that* $b, \sigma \in \mathcal{M}_4(\mathbb{R}^d \times \mathcal{P}_2(\mathbb{R}^d))$ *and* $\Phi \in \mathcal{M}_4(\mathcal{P}_2(\mathbb{R}^d))$. *Then*

$$\mathit{Var}\Big[\Phi(\mu_T^{2N}) - \frac{1}{2}\big(\Phi(\mu_T^{2N,(1)}) + \Phi(\mu_T^{2N,(2)})\big)\Big] \leq \frac{C}{N^2},$$

where C *is a constant that depends on* Φ*,* b*,* σ *and* T*, but does not depend on* N*.*

The following theorem gives a quantitative estimate on the order of interactions of the antithetic MLMC estimator.

Theorem 2.7 (Theorem 5.2)**.** *Assume* (Int). *Suppose that* b *and* σ *are of the forms* (2.9) *and* (2.10) *respectively. Furthermore, suppose that* $b, \sigma \in \mathcal{M}_4(\mathbb{R}^d \times \mathcal{P}_2(\mathbb{R}^d))$ *and* $\Phi \in \mathcal{M}_4(\mathcal{P}_2(\mathbb{R}^d))$*. Then there exist constants* $C_1, C_2 > 0$ *such that for any* $\epsilon < e^{-1}$*, there exist a value* L *and a sequence* $\{M_\ell\}_{\ell=0}^L$ *such that the mean-square error of* $\mathcal{A}^{\text{A- MLMC}}$ *(given by* (1.2)*) is bounded by*

$$\mathbb{E}\big[\big(\mathcal{A}^{\text{A- MLMC}} - \Phi(\mu_T^X)\big)^2\big] \leq C_1 \epsilon^2$$

and the order of interactions of $\mathcal{A}^{\text{A- MLMC}}$ *is bounded by*

$$\text{Order of interactions}\left(\mathcal{A}^{\text{A- MLMC}}\right) \leq \begin{cases} C_2 \epsilon^{-2} (\log \epsilon)^2, & p = 1, \\ C_2 \epsilon^{-1-p}, & p > 1. \end{cases}$$

For practical purposes, time discretisation is generally needed to simulate SDEs. We consider the time discretisation of (2.6), as in seminal papers by Bossy and Talay [3, 4], by working with an Euler scheme. Take partition $\{t_k\}_k$ of $[0, T]$, with $t_k - t_{k-1} = h$ and define $\eta(t) := t_k$ if $t \in [t_k, t_{k+1})$. The continuous Euler scheme reads

$$Z_t^{i,N,h} = Z_{t_k}^{i,N,h} + b(Z_{\eta(t)}^{i,N,h}, \mu_{\eta(t)}^{Z,N,h})(t - t_k) + \sigma(Z_{\eta(t)}^{i,N,h}, \mu_{\eta(t)}^{Z,N,h})(W_t^i - W_{t_k}^i). \tag{2.11}$$

Section 6 extends the antithetic MLMC estimator to include time discretisation (along with its complexity analysis), so as to be implementable. The numerical simulations of the algorithm in Section 6 can be found in [24].

2.3 Outline of the paper

Here is an outline of the main results of the article.

Section 3 establishes the result on the complexity of A-MLMC in Section 2.1.1 by proving Theorem 3.3, which concerns the antithetic difference in the i.i.d. case. Theorem 3.3 generalises the result in [20] (Section 9) from functionals in measure of the form (2.2) to general functionals in measure, which ultimately allows us to prove the complexity result: Theorem 2.3.

Subsequently, in Section 4, Theorem 4.2 proves a dimension-independent rate of uniform strong propagation of chaos for sufficiently smooth drift and diffusion functions. This is a considerable generalisation from [38], which assumes the drift and diffusion functions to be linear in measure.

In Section 5, we show from Theorem 5.2 that, under sufficient regular conditions on b and σ (having forms (2.9) and (2.10) respectively), the order of interactions of $\mathcal{A}^{\text{A- MLMC}}$ (given by (1.2)) is bounded by $C\epsilon^{-2}(\log \epsilon)^2$, for $p = 1$; $C\epsilon^{-1-p}$, for $p > 1$.

Finally, in Section 6, we apply an Euler time-discretisation to the A-MLMC method. In Theorem 6.3, we show that, under sufficient regular conditions on b and σ (having forms (2.9) and (2.10) respectively), the computational complexity of estimator (6.1) is bounded by $C\epsilon^{-2-p}$, where $p \geq 1$, which is still a considerable improvement compared to direct Monte-Carlo simulation.

Since this work relies heavily on the theory of differentiation in measure developed by P. Lions in his course at Collège de France [33], the reader is directed to Appendices A and B for further details.

3 A-MLMC for i.i.d. random variables

The main result in this section is Theorem 3.3, from which we can prove Theorem 2.3. We begin this section with the following lemma on the W_2 metric.

Lemma 3.1. *Let $\eta \in \mathbb{R}^d$ and $m \in \mathcal{P}_2(\mathbb{R}^d)$. Then*

$$W_2\Big(\frac{1}{N}\delta_\eta + \frac{N-1}{N} m, m\Big)^2 \leq \frac{2}{N}\bigg(|\eta|^2 + \int_{\mathbb{R}^d} |x|^2 \, m(dx)\bigg).$$

Proof. Let Y be a random variable with law m and let $\Omega' \in \mathcal{F}$ be a measurable event that is independent of $\sigma(Y)$, with probability $\frac{N-1}{N}$. Let X be a random variable defined by

$$X(\omega) := \begin{cases} Y(\omega), & \omega \in \Omega', \\ \eta, & \omega \notin \Omega'. \end{cases}$$

Then the law of X is $\frac{1}{N}\delta_\eta + \frac{N-1}{N} m$. Therefore, by the definition of the 2-Wasserstein metric,

$$\begin{aligned} W_2\Big(\frac{1}{N}\delta_\eta + \frac{N-1}{N} m, m\Big)^2 &\leq \mathbb{E}\big[|X-Y|^2\big] \\ &= \mathbb{E}\big[|X-Y|^2\big|\Omega'\big]\mathbb{P}(\Omega') + \mathbb{E}\big[|X-Y|^2\big|(\Omega')^c\big]\mathbb{P}((\Omega')^c) \\ &= \frac{1}{N}\mathbb{E}[|\eta - Y|^2] \\ &\leq \frac{2}{N}(|\eta|^2 + \mathbb{E}[|Y|^2]). \end{aligned}$$

□

For any functional from $\mathcal{P}_2(\mathbb{R}^d)$ to $\mathbb{R}$, the following lemma gives a bound on the error between the value of empirical measures under the functional and its limiting law under the functional. It relies on the regularity conditions stipulated in Proposition A.5. The proof of the following lemma is similar to Lemma 5.10 in [14]. However, the following result is slightly different in terms of hypotheses, as the first and second order linear functional derivatives are only of linear and quadratic growth respectively

(Proposition A.5), whereas they are assumed to be uniformly bounded and W_1-Lipschitz continuous in Lemma 5.10 of [14]. In return, we require a much higher moment (12 v.s. 4 in Lemma 5.10 of [14]). The following result is stated in a way with a constant that does not depend on the functional of measure, nor on the limiting law, so that it is useful with the relevant conditioning argument in the proof of Proposition 4.1. The technique of the following proof is also adopted in the proof of Theorem 3.3.

Lemma 3.2. *Let $U \in \mathcal{M}_3(\mathcal{P}_2(\mathbb{R}^d))$. Let $m_0 \in \mathcal{P}_{12}(\mathbb{R}^d)$ and $m^N = \frac{1}{N}\sum_{i=1}^N \delta_{\zeta_i}$, where $\zeta_1, \ldots, \zeta_N$ are i.i.d samples with law m_0. Then there exists a constant $C > 0$ (which does not depend on U, $\zeta_1, \ldots, \zeta_N$ and m_0) such that*

$$\mathbb{E}\big[\big|U(m^N) - U(m_0)\big|^4\big] \leq \frac{C}{N^2}\prod_{i=1}^3\Big(1 + \|\partial_\mu^i U\|_\infty^4\Big)\Big(1 + \int_{\mathbb{R}^d}|x|^{12}\, m_0(dx)\Big).$$

Proof. In this proof, C denotes an absolute constant that does not depend on U, $\zeta_1, \ldots, \zeta_N$ and m_0, whose value may vary from line to line. By the definition of linear functional derivatives, we have

$$\begin{aligned} U(m^N) - U(m_0) &= \int_0^1 \int_{\mathbb{R}^d} \frac{\delta U}{\delta m}(\lambda m^N + (1-\lambda)m_0, v)\,(m^N - m_0)(dv)\,d\lambda \\ &= \frac{1}{N}\sum_{i=1}^N \int_0^1 \varphi_\lambda^i\,d\lambda, \end{aligned}$$

where, for $i \in \{1, \ldots, N\}$ and $\lambda \in [0,1]$,

$$\varphi_\lambda^i = \frac{\delta U}{\delta m}(\lambda m^N + (1-\lambda)m_0, \zeta_i) - \widetilde{\mathbb{E}}\Big[\frac{\delta U}{\delta m}(\lambda m^N + (1-\lambda)m_0, \tilde{\zeta})\Big]. \tag{3.1}$$

By the bound on $\frac{\delta U}{\delta m}$ in Proposition A.5, we know that for distinct $i, j \in \{1, \ldots, N\}$,

$$\mathbb{E}\big[(\varphi_\lambda^i)^4 + (\varphi_\lambda^i)^2(\varphi_\lambda^j)^2 + \varphi_\lambda^i(\varphi_\lambda^j)^3\big] \leq C\|\partial_\mu U\|_\infty^4\mathbb{E}[|\zeta_1|^4]. \tag{3.2}$$

We have the estimate

$$\begin{aligned} \mathbb{E}\big[\big|U(m^N) - U(m_0)\big|^4\big] &\leq \frac{1}{N^4}\int_0^1 \mathbb{E}\Big[\Big(\sum_{i=1}^N \varphi_\lambda^i\Big)^4\Big]\,d\lambda \\ &\leq C\Big(\frac{1}{N^2}\|\partial_\mu U\|_\infty^4\mathbb{E}[|\zeta_1|^4] \\ &\quad + \frac{1}{N^4}\int_0^1 \mathbb{E}\Big[\sum_{i_1,i_2,i_3 \text{ distinct}} \varphi_\lambda^{i_1}\varphi_\lambda^{i_2}(\varphi_\lambda^{i_3})^2 + \sum_{\substack{i_1,i_2,i_3,i_4 \\ \text{distinct}}} \varphi_\lambda^{i_1}\varphi_\lambda^{i_2}\varphi_\lambda^{i_3}\varphi_\lambda^{i_4}\Big]\,d\lambda\Big). \end{aligned} \tag{3.3}$$

For any distinct i_1, i_2, i_3, we define $m^{N,-(i_1,i_2,i_3)} := \frac{1}{N-3}\sum_{\ell\neq i_1,i_2,i_3}\delta_{\zeta_\ell}$, which implies that

$$m^N - m^{N,-(i_1,i_2,i_3)} = \frac{1}{N}(\delta_{\zeta_{i_1}} + \delta_{\zeta_{i_2}} + \delta_{\zeta_{i_3}}) - \frac{3}{N(N-3)}\sum_{\ell\neq i_1,i_2,i_3}\delta_{\zeta_\ell}.$$

By the definition of second-order linear functional derivatives, we observe that

$$\begin{aligned}
&\frac{\delta U}{\delta m}(\lambda m^N + (1-\lambda)m_0, \zeta_i) - \frac{\delta U}{\delta m}(\lambda m^{N,-(i_1,i_2,i_3)} + (1-\lambda)m_0, \zeta_i) \\
= &\int_0^1 \int_{\mathbb{R}^d} \frac{\delta^2 U}{\delta m^2}\Big(s\lambda m^N + (1-s)\lambda m^{N,-(i_1,i_2,i_3)} + (1-\lambda)m_0, \zeta_i, v\Big)\,(m^N - m^{N,-(i_1,i_2,i_3)})(dv)\,ds \\
= &\int_0^1 \frac{1}{N}\Big[\sum_{\ell=i_1,i_2,i_3} \frac{\delta^2 U}{\delta m^2}\Big(s\lambda m^N + (1-s)\lambda m^{N,-(i_1,i_2,i_3)} + (1-\lambda)m_0, \zeta_i, \zeta_\ell\Big) \\
&-\frac{3}{N-3}\sum_{\ell\neq i_1,i_2,i_3} \frac{\delta^2 U}{\delta m^2}\Big(s\lambda m^N + (1-s)\lambda m^{N,-(i_1,i_2,i_3)} + (1-\lambda)m_0, \zeta_i, \zeta_\ell\Big)\Big]\,ds. \qquad (3.4)
\end{aligned}$$

By the bound on $\frac{\delta^2 U}{\delta m^2}$ in Proposition A.5,

$$\mathbb{E}\Big|\frac{\delta U}{\delta m}(\lambda m^N + (1-\lambda)m_0, \zeta_i) - \frac{\delta U}{\delta m}(\lambda m^{N,-(i_1,i_2,i_3)} + (1-\lambda)m_0, \zeta_i)\Big|^4 \leq \frac{C}{N^4}\|\partial_\mu^2 U\|_\infty^4 \mathbb{E}[|\zeta_1|^8].$$

Similarly, by applying the same argument to the second term in (3.1), we obtain that

$$\mathbb{E}\Big|\widetilde{\mathbb{E}}\Big[\frac{\delta U}{\delta m}(\lambda m^N + (1-\lambda)m_0, \tilde{\zeta})\Big] - \widetilde{\mathbb{E}}\Big[\frac{\delta U}{\delta m}(\lambda m^{N,-(i_1,i_2,i_3)} + (1-\lambda)m_0, \tilde{\zeta})\Big]\Big|^4 \leq \frac{C}{N^4}\|\partial_\mu^2 U\|_\infty^4 \mathbb{E}[|\zeta_1|^8],$$

which implies that

$$\mathbb{E}|\varphi_\lambda^i - \varphi_\lambda^{i,-(i_1,i_2,i_3)}|^4 \leq \frac{C}{N^4}\|\partial_\mu^2 U\|_\infty^4 \mathbb{E}[|\zeta_1|^8], \qquad (3.5)$$

where

$$\varphi_\lambda^{i,-(i_1,i_2,i_3)} = \frac{\delta U}{\delta m}(\lambda m^{N,-(i_1,i_2,i_3)} + (1-\lambda)m_0, \zeta_i) - \widetilde{\mathbb{E}}\Big[\frac{\delta U}{\delta m}(\lambda m^{N,-(i_1,i_2,i_3)} + (1-\lambda)m_0, \tilde{\zeta})\Big]. \qquad (3.6)$$

Finally, by writing $\varphi_\lambda^i = (\varphi_\lambda^i - \varphi_\lambda^{i,-(i_1,i_2,i_3)}) + \varphi_\lambda^{i,-(i_1,i_2,i_3)}$ and applying the generalised Hölder's inequality to (3.2) and (3.5),

$$\begin{aligned}
&\sum_{i_1,i_2,i_3 \text{ distinct}} \mathbb{E}\Big[\varphi_\lambda^{i_1}\varphi_\lambda^{i_2}(\varphi_\lambda^{i_3})^2\Big] \\
\leq &\sum_{i_1,i_2,i_3 \text{ distinct}} \Big[\frac{C}{N}(1 + \|\partial_\mu U\|_\infty^4)(1 + \|\partial_\mu^2 U\|_\infty^4)\mathbb{E}[|\zeta_1|^8] + \mathbb{E}\big[\varphi_\lambda^{i_1,-(i_1,i_2,i_3)}\varphi_\lambda^{i_2,-(i_1,i_2,i_3)}\big(\varphi_\lambda^{i_3,-(i_1,i_2,i_3)}\big)^2\big]\Big]
\end{aligned}$$

$$\leq CN^2(1+\|\partial_\mu U\|_\infty^4)(1+\|\partial_\mu^2 U\|_\infty^4)\mathbb{E}[|\zeta_1|^8] + \sum_{\substack{i_1,i_2,i_3 \\ \text{distinct}}} \mathbb{E}\big[\varphi_\lambda^{i_1,-(i_1,i_2,i_3)}\varphi_\lambda^{i_2,-(i_1,i_2,i_3)}\big(\varphi_\lambda^{i_3,-(i_1,i_2,i_3)}\big)^2\big]. \tag{3.7}$$

Let $\mathcal{F}^{-i}$ be the σ-algebra generated by $\zeta_1,\ldots,\zeta_N$ except ζ_i. Since $\zeta_1,\ldots,\zeta_N$ are independent, for any distinct i_1,i_2,i_3,

$$\mathbb{E}\big[\varphi_\lambda^{i_1,-(i_1,i_2,i_3)}\varphi_\lambda^{i_2,-(i_1,i_2,i_3)}\big(\varphi_\lambda^{i_3,-(i_1,i_2,i_3)}\big)^2\big] = \mathbb{E}\big[\varphi_\lambda^{i_2,-(i_1,i_2,i_3)}\big(\varphi_\lambda^{i_3,-(i_1,i_2,i_3)}\big)^2\mathbb{E}\big[\varphi_\lambda^{i_1,-(i_1,i_2,i_3)}\big|\mathcal{F}^{-i_1}\big]\big] = 0, \tag{3.8}$$

which implies that

$$\sum_{i_1,i_2,i_3 \text{ distinct}} \mathbb{E}\Big[\varphi_\lambda^{i_1}\varphi_\lambda^{i_2}(\varphi_\lambda^{i_3})^2\Big] \leq CN^2(1+\|\partial_\mu U\|_\infty^4)(1+\|\partial_\mu^2 U\|_\infty^4)\mathbb{E}[|\zeta_1|^8]. \tag{3.9}$$

Next, we define analogously the notation $\varphi^{i,-(i_1,i_2,i_3,i_4)}$ as (3.6). As above, by applying the generalised Hölder's inequality to (3.2) and (3.5), followed by a similar reasoning as (3.8), we have

$$\begin{aligned}
&\sum_{\substack{i_1,i_2,i_3,i_4 \\ \text{distinct}}} \mathbb{E}\Big[\varphi_\lambda^{i_1}\varphi_\lambda^{i_2}\varphi_\lambda^{i_3}\varphi_\lambda^{i_4}\Big] \\
\leq\; &\sum_{\substack{i_1,i_2,i_3,i_4 \\ \text{distinct}}} \Bigg[\frac{C}{N^2}(1+\|\partial_\mu U\|_\infty^4)(1+\|\partial_\mu^2 U\|_\infty^4)\mathbb{E}[|\zeta_1|^8] + \mathbb{E}\Bigg[\sum_{j=1}^4\big(\varphi_\lambda^{i_j}-\varphi_\lambda^{i_j,-(i_1,i_2,i_3,i_4)}\big)\prod_{\substack{k=1 \\ k\neq j}}^4 \varphi_\lambda^{i_k,-(i_1,i_2,i_3,i_4)}\Bigg] \\
&+\mathbb{E}\big[\varphi_\lambda^{i_1,-(i_1,i_2,i_3,i_4)}\varphi_\lambda^{i_2,-(i_1,i_2,i_3,i_4)}\varphi_\lambda^{i_3,-(i_1,i_2,i_3,i_4)}\varphi_\lambda^{i_4,-(i_1,i_2,i_3,i_4)}\big]\Bigg] \\
\leq\; &CN^2(1+\|\partial_\mu U\|_\infty^4)(1+\|\partial_\mu^2 U\|_\infty^4)\mathbb{E}[|\zeta_1|^8] + \sum_{\substack{i_1,i_2,i_3,i_4 \\ \text{distinct}}} \mathbb{E}\Bigg[\sum_{j=1}^4\big(\varphi_\lambda^{i_j}-\varphi_\lambda^{i_j,-(i_1,i_2,i_3,i_4)}\big)\prod_{\substack{k=1 \\ k\neq j}}^4 \varphi_\lambda^{i_k,-(i_1,i_2,i_3,i_4)}\Bigg].
\end{aligned} \tag{3.10}$$

Note that (3.5) only gives a growth in the order of $O(N^3)$ for the final term in (3.10), therefore it is insufficient.

By (3.4) followed by an application of the definition of third order linear functional derivatives, we have

$$\begin{aligned}
&\frac{\delta U}{\delta m}(\lambda m^N+(1-\lambda)m_0,\zeta_i) - \frac{\delta U}{\delta m}(\lambda m^{N,-(i_1,i_2,i_3,i_4)}+(1-\lambda)m_0,\zeta_i) \\
=\; &\frac{1}{N}\Bigg[\sum_{\ell=i_1,i_2,i_3,i_4}\frac{\delta^2 U}{\delta m^2}\Big(\lambda m^{N,-(i_1,i_2,i_3,i_4)}+(1-\lambda)m_0,\zeta_i,\zeta_\ell\Big) \\
&-\frac{4}{N-4}\sum_{\ell\neq i_1,i_2,i_3,i_4}\frac{\delta^2 U}{\delta m^2}\Big(\lambda m^{N,-(i_1,i_2,i_3,i_4)}+(1-\lambda)m_0,\zeta_i,\zeta_\ell\Big)\Bigg] + \varepsilon_N^{i,-(i_1,i_2,i_3,i_4)},
\end{aligned} \tag{3.11}$$

where

$$
\begin{aligned}
&\varepsilon_N^{i,-(i_1,i_2,i_3,i_4)}\\
=\;& \int_0^1 \frac{s\lambda}{N^2}\Bigg[\sum_{\ell=i_1,i_2,i_3,i_4}\int_0^1\Big[\sum_{\ell'=i_1,i_2,i_3,i_4}\frac{\delta^3 U}{\delta m^3}\Big(ts\lambda m^N+(1-ts)\lambda m^{N,-(i_1,i_2,i_3,i_4)}+(1-\lambda)m_0,\\
&\zeta_i,\zeta_\ell,\zeta_{\ell'}\Big)-\frac{4}{N-4}\sum_{\ell'\neq i_1,i_2,i_3,i_4}\frac{\delta^3 U}{\delta m^3}\Big(ts\lambda m^N+(1-ts)\lambda m^{N,-(i_1,i_2,i_3,i_4)}+(1-\lambda)m_0,\zeta_i,\zeta_\ell,\zeta_{\ell'}\Big)\Big]\,dt\\
&-\frac{4}{N-4}\sum_{\ell\neq i_1,i_2,i_3,i_4}\int_0^1\Big[\sum_{\ell'=i_1,i_2,i_3,i_4}\frac{\delta^3 U}{\delta m^3}\Big(ts\lambda m^N+(1-ts)\lambda m^{N,-(i_1,i_2,i_3,i_4)}+(1-\lambda)m_0,\\
&\zeta_i,\zeta_\ell,\zeta_{\ell'}\Big)-\frac{4}{N-4}\sum_{\ell'\neq i_1,i_2,i_3,i_4}\frac{\delta^3 U}{\delta m^3}\Big(ts\lambda m^N+(1-ts)\lambda m^{N,-(i_1,i_2,i_3,i_4)}+(1-\lambda)m_0,\zeta_i,\zeta_\ell,\zeta_{\ell'}\Big)\Big]\,dt\Bigg]\,ds,
\end{aligned}
$$

which implies that

$$
\mathbb{E}|\varepsilon_N^{i,-(i_1,i_2,i_3,i_4)}|^4\le\frac{C}{N^8}\|\partial_\mu^3 U\|_\infty^4\mathbb{E}[|\zeta_1|^{12}],
$$

by the bound on $\frac{\delta^3 U}{\delta m^3}$ in Proposition A.5. Repeating the same argument to the other term in (3.1) gives

$$
\begin{aligned}
&\varphi_\lambda^i-\varphi_\lambda^{i,-(i_1,i_2,i_3,i_4)}\\
=\;&\int_{\mathbb{R}^d}\frac{\delta^2 U}{\delta m^2}\Big(\lambda m^{N,-(i_1,i_2,i_3,i_4)}+(1-\lambda)m_0,\zeta_i,v\Big)\,(m^N-m^{N,-(i_1,i_2,i_3,i_4)})(dv)\\
&-\tilde{\mathbb{E}}\Big[\int_{\mathbb{R}^d}\frac{\delta^2 U}{\delta m^2}\Big(\lambda m^{N,-(i_1,i_2,i_3,i_4)}+(1-\lambda)m_0,\tilde\zeta,v\Big)\,(m^N-m^{N,-(i_1,i_2,i_3,i_4)})(dv)\Big]+\tilde\varepsilon_N^{i,-(i_1,i_2,i_3,i_4)},
\end{aligned}
$$

where

$$
\mathbb{E}|\tilde\varepsilon_N^{i,-(i_1,i_2,i_3,i_4)}|^4\le\frac{C}{N^8}\|\partial_\mu^3 U\|_\infty^4\mathbb{E}[|\zeta_1|^{12}]. \tag{3.12}
$$

Note that we can write the difference $\varphi_\lambda^{i_1}-\varphi_\lambda^{i_1,-(i_1,i_2,i_3,i_4)}-\tilde\varepsilon_N^{i_1,-(i_1,i_2,i_3,i_4)}$ as

$$
\varphi_\lambda^{i_1}-\varphi_\lambda^{i_1,-(i_1,i_2,i_3,i_4)}-\tilde\varepsilon_N^{i_1,-(i_1,i_2,i_3,i_4)}=\sum_{j=2}^4 F_j((\zeta_r)_{r\neq i_1,\dots,i_4},\zeta_{i_1},\zeta_{i_j}),
$$

for some measurable functions $F_2,F_3,F_4:(\mathbb{R}^d)^{N-2}\to\mathbb{R}$. Therefore,

$$
\begin{aligned}
&\mathbb{E}\Big[\Big(\varphi_\lambda^{i_1}-\varphi_\lambda^{i_1,-(i_1,i_2,i_3,i_4)}-\tilde\varepsilon_N^{i_1,-(i_1,i_2,i_3,i_4)}\Big)\varphi_\lambda^{i_2,-(i_1,i_2,i_3,i_4)}\varphi_\lambda^{i_3,-(i_1,i_2,i_3,i_4)}\varphi_\lambda^{i_4,-(i_1,i_2,i_3,i_4)}\Big]\\
=\;&\mathbb{E}\Big[\Big(\sum_{j\in\{3,4\}}F_j((\zeta_r)_{r\neq i_1,\dots,i_4},\zeta_{i_1},\zeta_{i_j})\Big)\varphi_\lambda^{i_3,-(i_1,i_2,i_3,i_4)}\varphi_\lambda^{i_4,-(i_1,i_2,i_3,i_4)}\mathbb{E}\Big[\varphi_\lambda^{i_2,-(i_1,i_2,i_3,i_4)}\Big|\mathcal{F}^{-i_2}\Big]\Big]
\end{aligned}
$$

$$+\mathbb{E}\Big[F_2\big((\zeta_r)_{r\neq i_1,\dots,i_4},\zeta_{i_1},\zeta_{i_2}\big)\varphi_\lambda^{i_2,-(i_1,i_2,i_3,i_4)}\varphi_\lambda^{i_4,-(i_1,i_2,i_3,i_4)}\mathbb{E}\Big[\varphi_\lambda^{i_3,-(i_1,i_2,i_3,i_4)}\Big|\mathcal{F}^{-i_3}\Big]\Big]=0.$$

Applying the generalised Hölder's inequality to (3.12) and (3.2) gives

$$\begin{aligned}&\mathbb{E}\Big[\Big(\varphi_\lambda^{i_1}-\varphi_\lambda^{i_1,-(i_1,i_2,i_3,i_4)}\Big)\varphi_\lambda^{i_2,-(i_1,i_2,i_3,i_4)}\varphi_\lambda^{i_3,-(i_1,i_2,i_3,i_4)}\varphi_\lambda^{i_4,-(i_1,i_2,i_3,i_4)}\Big]\\ \leq\;&\frac{C}{N^2}\|\partial_\mu^3 U\|_\infty\big(\mathbb{E}[|\zeta_1|^{12}]\big)^{1/4}\|\partial_\mu U\|_\infty^3\big(\mathbb{E}[|\zeta_1|^4]\big)^{3/4}\leq\frac{C}{N^2}\Big(1+\|\partial_\mu U\|_\infty^4\Big)\Big(1+\|\partial_\mu^3 U\|_\infty^4\Big)(1+\mathbb{E}[|\zeta_1|^{12}]).\end{aligned}$$

By the same reasoning, we can show that

$$\sum_{\substack{i_1,i_2,i_3,i_4\\ \text{distinct}}}\mathbb{E}\Big[\sum_{j=1}^4(\varphi_\lambda^{i_j}-\varphi_\lambda^{i_j,-(i_1,i_2,i_3,i_4)})\prod_{\substack{k=1\\k\neq j}}^4\varphi_\lambda^{i_k,-(i_1,i_2,i_3,i_4)}\Big]\leq CN^2\Big(1+\|\partial_\mu U\|_\infty^4\Big)\Big(1+\|\partial_\mu^3 U\|_\infty^4\Big)(1+\mathbb{E}[|\zeta_1|^{12}]).\tag{3.13}$$

We conclude the result by combining (3.3), (3.9), (3.10) and (3.13). □

Theorem 3.3 (Antithetic error on i.i.d. random variables)**.** *Suppose that $\mu\in\mathcal{P}_8(\mathbb{R}^d)$ and $U\in\mathcal{M}_4^L(\mathcal{P}_2(\mathbb{R}^d))$. Then there exists a constant $C>0$ such that*

$$\mathbb{E}\big|U(\mu^{2N})-\frac{1}{2}\big(U(\mu^{2N,(1)})+U(\mu^{2N,(2)})\big)\big|^2\leq\frac{C}{N^2}.$$

Proof of theorem 3.3. For simplicity of notations, let

$$\mu_{2N}:=\mu^{2N},\quad \mu_{2N,(1)}:=\mu^{2N,(1)},\quad \mu_{2N,(2)}:=\mu^{2N,(2)}.$$

For every $t\in[0,1]$, let

$$m_t^{2N}:=(1-t)\mu+t\mu_{2N},\quad m_t^{2N,(1)}:=(1-t)\mu+t\mu_{2N,(1)},\quad m_t^{2N,(2)}:=(1-t)\mu+t\mu_{2N,(2)}.$$

We define

$$[0,1]\ni t\mapsto f(t)=U\big((1-t)\mu+t\mu_{2N}\big)=U\big(\mu+t(\mu_{2N}-\mu)\big)\in\mathbb{R}$$

and apply Taylor-Lagrange formula to f up to order 2, namely

$$f(1)-f(0)=f^{'}(0)+\int_0^1(1-t)f^{(2)}(t)\,dt.$$

This yields

$$U(\mu_{2N})-U(\mu)=\int_{\mathbb{R}^d}\frac{\delta U}{\delta m}(\mu)(\mathbf{y})\,(\mu_{2N}-\mu)(d\mathbf{y})+\int_0^1(1-t)\Big[\int_{\mathbb{R}^{2d}}\frac{\delta^2 U}{\delta m^2}(m_t^{2N})(\mathbf{y})\,(\mu_{2N}-\mu)^{\otimes 2}(d\mathbf{y})\Big]\,dt.\tag{3.14}$$

Similarly,

$$U(\mu_{2N,(1)}) - U(\mu) = \int_{\mathbb{R}^d} \frac{\delta U}{\delta m}(\mu)(\mathbf{y})\,(\mu_{2N,(1)} - \mu)(d\mathbf{y}) + \int_0^1 (1-t)\left[\int_{\mathbb{R}^{2d}} \frac{\delta^2 U}{\delta m^2}(m_t^{2N,(1)})(\mathbf{y})\,(\mu_{2N,(1)} - \mu)^{\otimes 2}(d\mathbf{y})\right] dt \tag{3.15}$$

and

$$U(\mu_{2N,(2)}) - U(\mu) = \int_{\mathbb{R}^d} \frac{\delta U}{\delta m}(\mu)(\mathbf{y})\,(\mu_{2N,(2)} - \mu)(d\mathbf{y}) + \int_0^1 (1-t)\left[\int_{\mathbb{R}^{2d}} \frac{\delta^2 U}{\delta m^2}(m_t^{2N,(2)})(\mathbf{y})\,(\mu_{2N,(2)} - \mu)^{\otimes 2}(d\mathbf{y})\right] dt. \tag{3.16}$$

Computing the difference of (3.14) with the arithmetic average of (3.15) and (3.16) gives

$$\begin{aligned} U(\mu_{2N}) - \frac{1}{2}\big(U(\mu_{2N,(1)}) + U(\mu_{2N,(2)})\big) &= \int_0^1 (1-t)\left[\int_{\mathbb{R}^{2d}} \frac{\delta^2 U}{\delta m^2}(m_t^{2N})(\mathbf{y})\,(\mu_{2N} - \mu)^{\otimes 2}(d\mathbf{y})\right] dt \\ &\quad -\frac{1}{2}\int_0^1 (1-t)\left[\int_{\mathbb{R}^{2d}} \frac{\delta^2 U}{\delta m^2}(m_t^{2N,(1)})(\mathbf{y})\,(\mu_{2N,(1)} - \mu)^{\otimes 2}(d\mathbf{y})\right] dt \\ &\quad -\frac{1}{2}\int_0^1 (1-t)\left[\int_{\mathbb{R}^{2d}} \frac{\delta^2 U}{\delta m^2}(m_t^{2N,(2)})(\mathbf{y})\,(\mu_{2N,(2)} - \mu)^{\otimes 2}(d\mathbf{y})\right] dt. \end{aligned} \tag{3.17}$$

The rest of the proof is very similar to the proof of Lemma 3.2. It suffices to consider only the first term in (3.17). The other two terms can be handled in a similar way. We rewrite

$$\begin{aligned} &\int_{\mathbb{R}^{2d}} \frac{\delta^2 U}{\delta m^2}(m_t^{2N})(\mathbf{y})\,(\mu_{2N} - \mu)^{\otimes 2}(d\mathbf{y}) \\ &= \int_{\mathbb{R}^d} \left[\frac{1}{2N}\sum_{i=1}^{2N} \frac{\delta^2 U}{\delta m^2}(m_t^{2N})(\xi_i, y_2) - \int_{\mathbb{R}^d} \frac{\delta^2 U}{\delta m^2}(m_t^{2N})(z, y_2)\,\mu(dz)\right](\mu_{2N} - \mu)(dy_2) \\ &= \frac{1}{(2N)^2}\sum_{i,j=1}^{2N} \frac{\delta^2 U}{\delta m^2}(m_t^{2N})(\xi_i, \xi_j) - \frac{1}{2N}\sum_{j=1}^{2N}\int_{\mathbb{R}^d} \frac{\delta^2 U}{\delta m^2}(m_t^{2N})(z, \xi_j)\,\mu(dz) \\ &\quad -\frac{1}{2N}\sum_{i=1}^{2N}\int_{\mathbb{R}^d} \frac{\delta^2 U}{\delta m^2}(m_t^{2N})(\xi_i, z)\,\mu(dz) + \int_{\mathbb{R}^d}\int_{\mathbb{R}^d} \frac{\delta^2 U}{\delta m^2}(m_t^{2N})(z, z')\,\mu(dz)\,\mu(dz') \\ &= \frac{1}{(2N)^2}\sum_{i,j=1}^{2N} \varphi_t^{(i,j)}, \end{aligned} \tag{3.18}$$

where

$$\varphi_t^{(i,j)} := \frac{\delta^2 U}{\delta m^2}(m_t^{2N})(\xi_i, \xi_j) - \int_{\mathbb{R}^d} \frac{\delta^2 U}{\delta m^2}(m_t^{2N})(z, \xi_j)\,\mu(dz)$$

$$-\int_{\mathbb{R}^d}\frac{\delta^2 U}{\delta m^2}(m_t^{2N})(\xi_i,z)\,\mu(dz)+\int_{\mathbb{R}^d}\int_{\mathbb{R}^d}\frac{\delta^2 U}{\delta m^2}(m_t^{2N})(z,z')\,\mu(dz)\,\mu(dz').$$

Next, we observe that

$$\begin{aligned}
&\mathbb{E}\left|\frac{1}{(2N)^2}\sum_{i,j=1}^{2N}\varphi_t^{(i,j)}\right|^2\\
\lesssim\;&\frac{1}{N^2}+\frac{1}{N^4}\Bigg[\sum_{\substack{i_1,j_1,i_2,j_2\in\{1,\dots,2N\}\\ \text{exactly two of } i_1,j_1,i_2,j_2 \text{ are identical}}}\mathbb{E}\Big[\varphi_t^{(i_1,j_1)}\varphi_t^{(i_2,j_2)}\Big]+\sum_{\substack{i_1,j_1,i_2,j_2\in\{1,\dots,2N\}\\ i_1,j_1,i_2,j_2 \text{ are distinct}}}\mathbb{E}\Big[\varphi_t^{(i_1,j_1)}\varphi_t^{(i_2,j_2)}\Big]\Bigg].
\end{aligned}\tag{3.19}$$

We first consider the case where exactly two of i_1,i_2,j_1,j_2 are identical. Without loss of generality, suppose that $i_1=i_2$. As in the proof of Lemma 3.2, we define

$$\begin{aligned}
&\varphi_t^{(i,j),-(i_1,j_1,j_2)}\\
:=\;&\frac{\delta^2 U}{\delta m^2}(m_t^{2N,-(i_1,j_1,j_2)})(\xi_i,\xi_j)-\int_{\mathbb{R}^d}\frac{\delta^2 U}{\delta m^2}(m_t^{2N,-(i_1,j_1,j_2)})(z,\xi_j)\,\mu(dz)\\
&-\int_{\mathbb{R}^d}\frac{\delta^2 U}{\delta m^2}(m_t^{2N,-(i_1,j_1,j_2)})(\xi_i,z)\,\mu(dz)+\int_{\mathbb{R}^d}\int_{\mathbb{R}^d}\frac{\delta^2 U}{\delta m^2}(m_t^{2N,-(i_1,j_1,j_2)})(z,z')\,\mu(dz)\,\mu(dz'),
\end{aligned}\tag{3.20}$$

where

$$m_t^{2N,-(i_1,j_1,j_2)}:=(1-t)\mu+t\Bigg[\frac{1}{2N-3}\sum_{\substack{1\le\ell\le 2N\\ \ell\notin\{i_1,j_1,j_2\}}}\delta_{\xi_\ell}\Bigg].$$

By the same argument as in the proof of Lemma 3.2, along with the bound on $\frac{\delta^3 U}{\delta m^3}$ in (A.11) (see (3.4) for details), we have

$$\mathbb{E}|\varphi_t^{(i,j)}-\varphi_t^{(i,j),-(i_1,j_1,j_2)}|^2\lesssim\frac{1}{N^2}.$$

Then, we write

$$\begin{aligned}
\mathbb{E}\Big[\varphi_t^{(i_1,j_1)}\varphi_t^{(i_1,j_2)}\Big]\;=\;&\mathbb{E}\Big[(\varphi_t^{(i_1,j_1)}-\varphi_t^{(i_1,j_1),-(i_1,j_1,j_2)})(\varphi_t^{(i_1,j_2)}-\varphi_t^{(i_1,j_2),-(i_1,j_1,j_2)})\Big]\\
&+\mathbb{E}\Big[(\varphi_t^{(i_1,j_1)}-\varphi_t^{(i_1,j_1),-(i_1,j_1,j_2)})\varphi_t^{(i_1,j_2),-(i_1,j_1,j_2)}\Big]\\
&+\mathbb{E}\Big[\varphi_t^{(i_1,j_1),-(i_1,j_1,j_2)}(\varphi_t^{(i_1,j_2)}-\varphi_t^{(i_1,j_2),-(i_1,j_1,j_2)})\Big]\\
&+\mathbb{E}\Big[\varphi_t^{(i_1,j_1),-(i_1,j_1,j_2)}\varphi_t^{(i_1,j_2),-(i_1,j_1,j_2)}\Big].
\end{aligned}$$

By the Cauchy-Schwarz inequality and the bound on $\frac{\delta^2 U}{\delta m^2}$ in (A.11), the first three terms converge to 0 in the order $O(1/N)$. Let $\mathcal{F}^{-i}$ be the σ-algebra generated by $\xi_1, \ldots, \xi_N$ except ξ_i. Then

$$\mathbb{E}\Big[\varphi_t^{(i_1,j_1),-(i_1,j_1,j_2)}\varphi_t^{(i_1,j_2),-(i_1,j_1,j_2)}\Big] = \mathbb{E}\Big[\varphi_t^{(i_1,j_1),-(i_1,j_1,j_2)}\mathbb{E}\Big[\varphi_t^{(i_1,j_2),-(i_1,j_1,j_2)}\Big|\mathcal{F}^{-j_2}\Big]\Big] = 0.$$

Therefore,

$$\frac{1}{N^4}\sum_{\substack{i_1,j_1,i_2,j_2\in\{1,\ldots,2N\} \\ \text{exactly two of } i_1,j_1,i_2,j_2 \text{ are identical}}} \mathbb{E}\Big[\varphi_t^{(i_1,j_1)}\varphi_t^{(i_2,j_2)}\Big] \lesssim \frac{1}{N^2}. \tag{3.21}$$

Finally, we consider the case where i_1, j_1, i_2, j_2 are mutually distinct. We define $\varphi_t^{(i,j),-(i_1,j_1,i_2,j_2)}$ analogously, as the definition of $\varphi_t^{(i,j),-(i_1,j_1,j_2)}$ in (3.20). As above, we write

$$\begin{aligned}
\mathbb{E}\Big[\varphi_t^{(i_1,j_1)}\varphi_t^{(i_2,j_2)}\Big] &= \mathbb{E}\Big[(\varphi_t^{(i_1,j_1)} - \varphi_t^{(i_1,j_1),-(i_1,j_1,i_2,j_2)})(\varphi_t^{(i_2,j_2)} - \varphi_t^{(i_2,j_2),-(i_1,j_1,i_2,j_2)})\Big] \\
&\quad + \mathbb{E}\Big[(\varphi_t^{(i_1,j_1)} - \varphi_t^{(i_1,j_1),-(i_1,j_1,i_2,j_2)})\varphi_t^{(i_2,j_2),-(i_1,j_1,i_2,j_2)}\Big] \\
&\quad + \mathbb{E}\Big[\varphi_t^{(i_1,j_1),-(i_1,j_1,i_2,j_2)}(\varphi_t^{(i_2,j_2)} - \varphi_t^{(i_2,j_2),-(i_1,j_1,i_2,j_2)})\Big] \\
&\quad + \mathbb{E}\Big[\varphi_t^{(i_1,j_1),-(i_1,j_1,i_2,j_2)}\varphi_t^{(i_2,j_2),-(i_1,j_1,i_2,j_2)}\Big].
\end{aligned}$$

As before, we have

$$\mathbb{E}|\varphi_t^{(i,j)} - \varphi_t^{(i,j),-(i_1,j_1,i_2,j_2)}|^2 \lesssim \frac{1}{N^2}$$

and hence

$$\mathbb{E}\Big|(\varphi_t^{(i_1,j_1)} - \varphi_t^{(i_1,j_1),-(i_1,j_1,i_2,j_2)})(\varphi_t^{(i_2,j_2)} - \varphi_t^{(i_2,j_2),-(i_1,j_1,i_2,j_2)})\Big| \lesssim \frac{1}{N^2}, \tag{3.22}$$

by the Cauchy-Schwarz inequality. By the same argument as in the proof of Lemma 3.2 through considering the fourth order linear functional derivative of U, along with the bound on $\frac{\delta^4 U}{\delta m^4}$ in (A.11) (see (3.11) and (3.12) for details), we obtain that

$$\begin{aligned}
&\varphi_t^{(i_1,j_1)} - \varphi_t^{(i_1,j_1),-(i_1,j_1,i_2,j_2)} \\
= \;& F_1((\xi_r)_{r\neq i_1,j_1,i_2,j_2}, \xi_{i_1}, \xi_{j_1}, \xi_{i_2}) + F_2((\xi_r)_{r\neq i_1,j_1,i_2,j_2}, \xi_{i_1}, \xi_{j_1}, \xi_{j_2}) + \tilde{\varepsilon}_N^{(i_1,j_1),-(i_1,j_1,i_2,j_2)},
\end{aligned}$$

for some measurable functions $F_1, F_2 : (\mathbb{R}^d)^{2N-1} \to \mathbb{R}$, where

$$\mathbb{E}\Big|\tilde{\varepsilon}_N^{(i_1,j_1),-(i_1,j_1,i_2,j_2)}\Big|^2 \lesssim \frac{1}{N^4}.$$

By a similar conditioning argument as the proof of Lemma 3.2,

$$\begin{aligned}
&\mathbb{E}\Big[\Big(\varphi_t^{(i_1,j_1)} - \varphi_t^{(i_1,j_1),-(i_1,j_1,i_2,j_2)} - \tilde{\varepsilon}_N^{(i_1,j_1),-(i_1,j_1,i_2,j_2)}\Big)\varphi_t^{(i_2,j_2),-(i_1,j_1,i_2,j_2)}\Big] \\
= \;& \mathbb{E}\Big[F_1((\xi_r)_{r\neq i_1,j_1,i_2,j_2}, \xi_{i_1}, \xi_{j_1}, \xi_{i_2})\mathbb{E}\Big[\varphi_t^{(i_2,j_2),-(i_1,j_1,i_2,j_2)}\Big|\mathcal{F}^{-j_2}\Big]\Big]
\end{aligned}$$

$$+\mathbb{E}\Big[F_2((\xi_r)_{r\neq i_1,j_1,i_2,j_2},\xi_{i_1},\xi_{j_1},\xi_{j_2})\mathbb{E}\Big[\varphi_t^{(i_2,j_2),-(i_1,j_1,i_2,j_2)}\Big|\mathcal{F}^{-i_2}\Big]\Big]=0,$$

which implies, by the Cauchy-Schwarz inequality and the bound on $\frac{\delta^2 U}{\delta m^2}$ in (A.11), that

$$\mathbb{E}\Big|\Big(\varphi_t^{(i_1,j_1)}-\varphi_t^{(i_1,j_1),-(i_1,j_1,i_2,j_2)}\Big)\varphi_t^{(i_2,j_2),-(i_1,j_1,i_2,j_2)}\Big|\lesssim\frac{1}{N^2}. \tag{3.23}$$

Similarly,

$$\mathbb{E}\Big|\varphi_t^{(i_1,j_1),-(i_1,j_1,i_2,j_2)}(\varphi_t^{(i_2,j_2)}-\varphi_t^{(i_2,j_2),-(i_1,j_1,i_2,j_2)})\Big|\lesssim\frac{1}{N^2}. \tag{3.24}$$

By the same conditioning argument,

$$\mathbb{E}\Big[\varphi_t^{(i_1,j_1),-(i_1,j_1,i_2,j_2)}\varphi_t^{(i_2,j_2),-(i_1,j_1,i_2,j_2)}\Big]=\mathbb{E}\Big[\varphi_t^{(i_1,j_1),-(i_1,j_1,i_2,j_2)}\mathbb{E}\Big[\varphi_t^{(i_2,j_2),-(i_1,j_1,i_2,j_2)}\Big|\mathcal{F}^{-i_2}\Big]\Big]=0. \tag{3.25}$$

A combination of (3.22), (3.23), (3.24) and (3.25) implies that

$$\frac{1}{N^4}\sum_{\substack{i_1,j_1,i_2,j_2\in\{1,\dots,2N\}\\ i_1,j_1,i_2,j_2\text{ are distinct}}}\mathbb{E}\Big[\varphi_t^{(i_1,j_1)}\varphi_t^{(i_2,j_2)}\Big]\lesssim\frac{1}{N^2}. \tag{3.26}$$

Finally, a combination of (3.18), (3.19), (3.21) and (3.26) implies that

$$\mathbb{E}\bigg|\int_{\mathbb{R}^{2d}}\frac{\delta^2 U}{\delta m^2}(m_t^{2N})(\mathbf{y})\,(\mu_{2N}-\mu)^{\otimes 2}(d\mathbf{y})\bigg|^2\lesssim\frac{1}{N^2}.$$

□

4 Dimension-independent rate of uniform strong propagation of chaos

We now introduce a mean-field coupling of the particle system (2.7) by

$$\begin{cases} dX_t^i=\xi_i+\int_0^t b(X_s^i,\mu_s^X)\,ds+\int_0^t\sigma(X_s^i,\mu_s^X)\,dW_s^i, \quad 1\le i\le N,\quad t\in[0,T],\\ \mu_s^{X,N}:=\frac{1}{N}\sum_{i=1}^N\delta_{X_s^i}.\end{cases} \tag{4.1}$$

The following two assumptions are adopted in most results. We assume that

$$\begin{cases} b\text{ and }\sigma\text{ are Lipschitz continuous with respect to the Euclidean norm and the }W_2\text{ norm},\\ \Phi\text{ is Lipschitz continuous with respect to the }W_2\text{ norm},\end{cases} \tag{Lip}$$

and that the initial law ν satisfies

$$\int_{\mathbb{R}^d}|x|^{12}\,\nu(dx)<+\infty. \tag{Int}$$

Note that (Lip) guarantees strong existence and uniqueness of (2.6) and (2.7). The following proposition is essential to the proofs of Theorem 4.2 and Theorem 5.1.

Proposition 4.1. *Suppose that b and σ admit linear growth in the spatial and measure components. Suppose that $\varphi \in \mathcal{M}_3(\mathbb{R}^d \times \mathcal{P}_2(\mathbb{R}^d))$. Assume* (Int)*. Then*

$$\frac{1}{N}\sum_{i=1}^{N} \sup_{t\in[0,T]} \mathbb{E}\Big|\varphi(X_t^i, \mu_t^{X,N}) - \varphi(X_t^i, \mu_t^X)\Big|^4 \le \frac{C}{N^2},$$

for some constant $C > 0$.

Proof.

$$\begin{aligned}
& \frac{1}{N}\sum_{i=1}^{N} \sup_{t\in[0,T]} \mathbb{E}\Bigg[\Bigg|\varphi\Big(X_t^i, \frac{1}{N}\sum_{j=1}^{N}\delta_{X_t^j}\Big) - \varphi(X_t^i, \mu_t^X)\Bigg|^4\Bigg] \\
= \;& \frac{1}{N}\sum_{i=1}^{N} \sup_{t\in[0,T]} \mathbb{E}\Bigg[\mathbb{E}\Bigg[\Bigg|\varphi\Big(\eta, \frac{1}{N}\delta_\eta + \frac{N-1}{N}\cdot\frac{1}{N-1}\sum_{\substack{1\le j\le N\\ j\ne i}}\delta_{X_t^j}\Big) - \varphi(\eta, \mu_t^X)\Bigg|^4\Bigg]\Bigg|_{\eta=X_t^i}\Bigg] \\
\le \;& \frac{8}{N}\sum_{i=1}^{N} \sup_{t\in[0,T]} \mathbb{E}\Bigg[\mathbb{E}\Bigg[\Bigg|\varphi\Big(\eta, \frac{1}{N}\delta_\eta + \frac{N-1}{N}\cdot\frac{1}{N-1}\sum_{\substack{1\le j\le N\\ j\ne i}}\delta_{X_t^j}\Big) \\
& -\varphi\Big(\eta, \frac{1}{N-1}\sum_{\substack{1\le j\le N\\ j\ne i}}\delta_{X_t^j}\Big)\Bigg|^4\Bigg]\Bigg|_{\eta=X_t^i}\Bigg] \\
& +\frac{8}{N}\sum_{i=1}^{N} \sup_{t\in[0,T]} \mathbb{E}\Bigg[\mathbb{E}\Bigg[\Bigg|\varphi\Big(\eta, \frac{1}{N-1}\sum_{\substack{1\le j\le N\\ j\ne i}}\delta_{X_t^j}\Big) - \varphi(\eta, \mu_t^X)\Bigg|^4\Bigg]\Bigg|_{\eta=X_t^i}\Bigg] \\
=: \;& \Pi_1 + \Pi_2.
\end{aligned}$$

By Lemma 3.1, using the same type of estimate as (4.5), we have

$$\Pi_1 \le \frac{8}{N}\sum_{i=1}^{N} \sup_{t\in[0,T]} \mathbb{E}\Bigg[\frac{4}{N^2}\Big(|X_t^i|^2 + \frac{1}{N-1}\sum_{\substack{1\le j\le N\\ j\ne i}}|X_t^j|^2\Big)^2\Bigg] \lesssim \frac{1}{N^2}. \tag{4.2}$$

By the assumption on φ, we observe that for any $\eta \in \mathbb{R}^d$, the uniform bounds on $\partial_\mu\varphi(\eta,\cdot)$, $\partial_\mu^2\varphi(\eta,\cdot)$ and $\partial_\mu^3\varphi(\eta,\cdot)$ do not depend on η. Finally, since b and σ are of linear growth in the spatial and measure components and $\mathbb{E}[|\xi|^{12}] < +\infty$, we have $\sup_{t\in[0,T]}\mathbb{E}[|X_t|^{12}] < +\infty$. Therefore, Lemma 3.2 implies that

$$\Pi_2 \lesssim \frac{1}{(N-1)^2}\prod_{i=1}^{3}\Big(1 + \sup_{\eta\in\mathbb{R}^d}\|\partial_\mu^i\varphi(\eta,\cdot)\|_\infty^4\Big)\Big(1 + \sup_{t\in[0,T]}\int_{\mathbb{R}^d}|y|^{12}\,\mu_t^X(dy)\Big). \tag{4.3}$$

A combination of (4.2) and (4.3) yields the result. □

Note that Proposition 4.1 allows us to completely bypass the consideration of the Wasserstein distance between empirical measures and their limiting law. Assuming (Lip) and (Int), Theorem 10.2.7 in [37] gives us a rate of convergence of

$$\mathbb{E}\Big[\sup_{t\in[0,T]} W_2\big(\mu_t^X,\mu_t^{X,N}\big)^2\Big] \le \frac{C}{N^{2/(d+8)}}. \tag{4.4}$$

The following result gives a uniform rate of strong propagation of chaos between the particle system (2.7) and its coupled mean-field limit (4.1), under the assumption that b and σ are sufficiently smooth in the sense of L-derivatives. This is a different set of sufficient conditions compared to the existing results in the literature with the same rate, such as Lemma 5.1 in [14], Theorem 1 in [26] and [27].

Let $\mathcal{C}_T := C([0,T],\mathbb{R}^d)$ be the space of continuous functions from $[0,T]$ to $\mathbb{R}^d$ equipped with the supremum norm and $W_{\mathcal{C}_T,2}$ be the 2-Wasserstein metric on $\mathcal{C}_T$.

Theorem 4.2 (Uniform strong propagation of chaos). *Assume* (Int). *Suppose that* $b,\sigma\in\mathcal{M}_3(\mathbb{R}^d\times\mathcal{P}_2(\mathbb{R}^d))$. *Then*

$$\mathbb{E}\Big[W_{\mathcal{C}_T,2}\big(\mu^N,\mu^{X,N}\big)^4\Big] \le \mathbb{E}\Big[\frac{1}{N}\sum_{i=1}^N\Big(\sup_{t\in[0,T]}|X_t^i-Y_t^{i,N}|^4\Big)\Big] \le \frac{C}{N^2},$$

for some constant $C>0$.

Proof. By the Hölder and Buckholder-Davis-Gundy inequalities, estimating the L^4 difference between (2.7) and (4.1) gives

$$\begin{aligned}\mathbb{E}\Big[\sup_{s\in[0,t]}\big|X_s^i-Y_s^{i,N}\big|^4\Big] \;\le\; & C\Big(\int_0^t \mathbb{E}|b(X_s^i,\mu_s^X)-b(Y_s^{i,N},\mu_s^N)|^4\,ds \\ & +\int_0^t \mathbb{E}\|\sigma(X_s^i,\mu_s^X)-\sigma(Y_s^{i,N},\mu_s^N)\|^4\,ds\Big),\end{aligned} \tag{4.5}$$

for every $t\in[0,T]$. By Lipschitz continuity of b and σ,

$$\begin{aligned}\mathbb{E}\Big[\sup_{s\in[0,t]}\big|X_s^i-Y_s^{i,N}\big|^4\Big] \;\le\; & C\Big(\int_0^t \mathbb{E}\Big[\sup_{u\in[0,s]}\big|X_u^i-Y_u^{i,N}\big|^4\Big]\,ds+\int_0^t \mathbb{E}|b(X_s^i,\mu_s^X)-b(X_s^i,\mu_s^N)|^4\,ds \\ & +\int_0^t \mathbb{E}\|\sigma(X_s^i,\mu_s^X)-\sigma(X_s^i,\mu_s^N)\|^4\,ds\Big),\end{aligned}$$

for every $t\in[0,T]$, which gives, upon taking average over i,

$$\frac{1}{N}\sum_{i=1}^N\mathbb{E}\Big[\sup_{s\in[0,t]}\big|X_s^i-Y_s^{i,N}\big|^4\Big] \;\le\; C\Big(\int_0^t\frac{1}{N}\sum_{i=1}^N\mathbb{E}\Big[\sup_{u\in[0,s]}\big|X_u^i-Y_u^{i,N}\big|^4\Big]\,ds$$

$$+\int_0^t \frac{1}{N}\sum_{i=1}^N \mathbb{E}|b(X_s^i,\mu_s^X)-b(X_s^i,\mu_s^N)|^4\,ds$$

$$+\int_0^t \frac{1}{N}\sum_{i=1}^N \mathbb{E}\|\sigma(X_s^i,\mu_s^X)-\sigma(X_s^i,\mu_s^N)\|^4\,ds\Big). \tag{4.6}$$

Also, the empirical measure of the particles can be replaced by the empirical measure of the coupled system by the bound

$$\mathbb{E}\big[W_2(\mu_s^{X,N},\mu_s^N)^4\big] \le \bigg[\Big(\frac{1}{N}\sum_{i=1}^N \mathbb{E}\big|Y_s^{i,N}-X_s^i\big|^2\Big)^2\bigg] \le \frac{1}{N}\sum_{i=1}^N \mathbb{E}\Big[\sup_{u\in[0,s]}\big|X_u^i-Y_u^{i,N}\big|^4\Big]. \tag{4.7}$$

A combination of (4.6) and (4.7) gives

$$\begin{aligned}\frac{1}{N}\sum_{i=1}^N \mathbb{E}\Big[\sup_{s\in[0,t]}\big|X_s^i-Y_s^{i,N}\big|^4\Big] \quad\le\quad & C\Big(\int_0^t \frac{1}{N}\sum_{i=1}^N \mathbb{E}\Big[\sup_{u\in[0,s]}\big|X_u^i-Y_u^{i,N}\big|^4\Big]\,ds \\ &+\int_0^t \frac{1}{N}\sum_{i=1}^N \sup_{u\in[0,s]}\mathbb{E}|b(X_u^i,\mu_u^X)-b(X_u^i,\mu_u^{X,N})|^4\,ds \\ &+\int_0^t \frac{1}{N}\sum_{i=1}^N \sup_{u\in[0,s]}\mathbb{E}\|\sigma(X_u^i,\mu_u^X)-\sigma(X_u^i,\mu_u^{X,N})\|^4\,ds\Big).\end{aligned}$$

Therefore, by Proposition 4.1 and Gronwall's inequality, we have

$$\frac{1}{N}\sum_{i=1}^N \mathbb{E}\Big[\sup_{s\in[0,T]}\big|X_s^i-Y_s^{i,N}\big|^4\Big] \le \frac{C}{N^2},$$

for every $t\in[0,T]$. □

5 Antithetic MLMC without time discretisation

The main aim of this section is to prove the complexity of the antithetic MLMC estimator, via the following theorem, which states that the variance of the antithetic difference in (1.2) converges in N in the rate $O(1/N^2)$. In the proof, Proposition 4.1 and Theorem 4.2 provide us with the necessary estimates when we revert to the mean-field limit.

Theorem 5.1 (Variance of antithetic difference). *Assume* (Int). *Suppose that* $b,\sigma\in\mathcal{M}_4\big(\mathbb{R}^d\times\mathcal{P}_2(\mathbb{R}^d)\big)$ *and* $\Phi\in\mathcal{M}_4\big(\mathcal{P}_2(\mathbb{R}^d)\big)$. *Then*

$$\mathit{Var}\Big[\Phi(\mu_T^{2N})-\frac{1}{2}\big(\Phi(\mu_T^{2N,(1)})+\Phi(\mu_T^{2N,(2)})\big)\Big]\le\frac{C}{N^2},$$

where C is a constant that depends on Φ, b, σ and T, but does not depend on N.

Proof of Theorem 5.1. The main techniques in the proof depend on the function $\mathcal{V} : [0,T] \times \mathcal{P}_2(\mathbb{R}^d) \to \mathbb{R}$, which is defined in (B.2) by

$$\mathcal{V}(s, \mathscr{L}(\eta)) = \Phi\big(\mathscr{L}(X_T^{s,\eta})\big).$$

Another crucial ingredient in the proof is (B.7), which represents the difference $\Phi(\mu_T^N) - \Phi(\mu_T^X)$ as

$$\begin{aligned}
\Phi(\mu_T^N) - \Phi(\mu_T^X) \quad &= \quad \big(\mathcal{V}(0,\mu_0^N) - \mathcal{V}(0,\nu)\big) \\
&\quad + \int_0^T \frac{1}{2}\left[\frac{1}{N^2}\sum_{i=1}^N \operatorname{Tr}\Big(a\big(Y_s^{i,N},\mu_s^N\big)\partial_\mu^2\mathcal{V}\big(s,\mu_s^N\big)(Y_s^{i,N},Y_s^{i,N})\Big)\right] ds \\
&\quad + \frac{1}{N}\sum_{i=1}^N \int_0^T \sigma(Y_s^{i,N},\mu_s^N)^T\partial_\mu\mathcal{V}\big(s,\mu_s^N\big)(Y_s^{i,N}) \cdot dW_s^i.
\end{aligned}$$

Hence,

$$\Phi(\mu_T^{2N}) - \frac{1}{2}\big(\Phi(\mu_T^{2N,(1)}) + \Phi(\mu_T^{2N,(2)})\big) = \mathscr{A} + \mathscr{D} + \mathscr{S},$$

where

$$\mathscr{A} := \mathcal{V}(0,\mu_0^{2N}) - \frac{1}{2}\big(\mathcal{V}(0,\mu_0^{2N,(1)}) + \mathcal{V}(0,\mu_0^{2N,(2)})\big),$$

$$\begin{aligned}
\mathscr{D} \quad := \quad &\int_0^T \frac{1}{2}\left[\frac{1}{(2N)^2}\sum_{i=1}^{2N} \operatorname{Tr}\Big(a\big(Y_s^{i,2N},\mu_s^{2N}\big)\partial_\mu^2\mathcal{V}\big(s,\mu_s^{2N}\big)(Y_s^{i,2N},Y_s^{i,2N})\Big)\right] \\
&- \frac{1}{2N^2}\left[\sum_{i=1}^N \operatorname{Tr}\Big(a\big(Y_s^{i,2N,(1)},\mu_s^{2N,(1)}\big)\partial_\mu^2\mathcal{V}\big(s,\mu_s^{2N,(1)}\big)(Y_s^{i,2N,(1)},Y_s^{i,2N,(1)})\Big)\right. \\
&\left. + \sum_{i=N+1}^{2N} \operatorname{Tr}\Big(a\big(Y_s^{i,2N,(2)},\mu_s^{2N,(2)}\big)\partial_\mu^2\mathcal{V}\big(s,\mu_s^{2N,(2)}\big)(Y_s^{i,2N,(2)},Y_s^{i,2N,(2)})\Big)\right] ds
\end{aligned}$$

and

$$\begin{aligned}
\mathscr{S} \quad := \quad &\sum_{i=1}^{2N}\int_0^T \frac{1}{2N}\partial_\mu\mathcal{V}\big(s,\mu_s^{2N}\big)(Y_s^{i,2N})^T\sigma(Y_s^{i,2N},\mu_s^{2N})dW_s^i \\
&- \frac{1}{2N}\Big(\sum_{i=1}^N\int_0^T \partial_\mu\mathcal{V}\big(\mu_s^{2N,(1)}\big)(Y_s^{i,2N,(1)})^T\sigma(Y_s^{i,2N,(1)},\mu_s^{2N,(1)})dW_s^i \\
&+ \sum_{i=N+1}^{2N}\int_0^T \partial_\mu\mathcal{V}\big(s,\mu_s^{2N,(2)}\big)(Y_s^{i,2N,(2)})^T\sigma(Y_s^{i,2N,(2)},\mu_s^{2N,(2)})dW_s^i\Big).
\end{aligned}$$

By the assumptions on b, σ and Φ, it follows from Theorem B.1 that $\mathcal{V} \in \mathcal{M}_4\big([0,T]\times\mathcal{P}_2(\mathbb{R}^d)\big)$. We can therefore see that

$$\mathbb{E}[\mathscr{D}^2] \lesssim 1/N^2.$$

In particular, $\mathcal{V}(0,\cdot)\in\mathcal{M}_4\big(\mathcal{P}_2(\mathbb{R}^d)\big)$. Therefore, by Theorem 3.3, we obtain that

$$\mathbb{E}[\mathscr{A}^2]\lesssim 1/N^2.$$

Hence, it remains to show that $\mathbb{E}(\mathscr{S}^2)\lesssim 1/N^2$. Define $\Sigma(t,x,\mu):=\partial_\mu\mathcal{V}(t,\mu)(x)^T\sigma(x,\mu)$. By the independence of the Brownian motions, we first rewrite $\mathbb{E}[\mathscr{S}^2]$ as

$$\begin{aligned}\mathbb{E}[\mathscr{S}^2] &= \mathbb{E}\Bigg[\Bigg(\frac{1}{2N}\sum_{i=1}^{N}\int_0^T\Sigma(s,Y_s^{i,2N},\mu_s^{2N})-\Sigma(s,Y_s^{i,2N,(1)},\mu_s^{2N,(1)})dW_s^i\Bigg)^2\Bigg]\\ &\quad+\mathbb{E}\Bigg[\Bigg(\frac{1}{2N}\sum_{i=N+1}^{2N}\int_0^T\Sigma(s,Y_s^{i,2N},\mu_s^{2N})-\Sigma(s,Y_s^{i,2N,(2)},\mu_s^{2N,(2)})dW_s^i\Bigg)^2\Bigg].\end{aligned}$$

Using the independence of the Brownian motions and Itô's isometry,

$$\begin{aligned}&\mathbb{E}\Bigg[\Bigg(\frac{1}{2N}\sum_{i=1}^{N}\int_0^T\Sigma(s,Y_s^{i,2N},\mu_s^{2N})-\Sigma(s,Y_s^{i,2N,(1)},\mu_s^{2N,(1)})dW_s^i\Bigg)^2\Bigg]\\ &=\frac{1}{4N^2}\sum_{i=1}^{N}\mathbb{E}\Bigg[\Bigg(\int_0^T\Sigma(s,Y_s^{i,2N},\mu_s^{2N})-\Sigma(s,Y_s^{i,2N,(1)},\mu_s^{2N,(1)})dW_s^i\Bigg)^2\Bigg]\\ &=\frac{1}{4N^2}\sum_{i=1}^{N}\int_0^T\mathbb{E}\Big[\Big\|\Sigma(s,Y_s^{i,2N},\mu_s^{2N})-\Sigma(s,Y_s^{i,2N,(1)},\mu_s^{2N,(1)})\Big\|^2\Big]\,ds.\end{aligned}$$

Note that $\mathcal{V}\in\mathcal{M}_4([0,T]\times\mathcal{P}_2(\mathbb{R}^d))$. Therefore, $\partial_\mu\mathcal{V}$ is Lipschitz continuous and uniformly bounded. Also, note that σ is Lipschitz continuous. By Theorem 4.2,

$$\begin{aligned}&\sup_{t\in[0,T]}\mathbb{E}\big[\big\|\Sigma(t,Y_t^{i,2N},\mu_t^{2N})-\Sigma(t,X_t^i,\mu_t^{X,2N})\big\|^2\big]\\ =\;&\sup_{t\in[0,T]}\mathbb{E}\big[\big\|\partial_\mu\mathcal{V}(t,\mu_t^{2N})(Y_t^{i,2N})^T\sigma(Y_t^{i,2N},\mu_t^{2N})-\partial_\mu\mathcal{V}(t,\mu_t^{X,2N})(X_t^i)^T\sigma(X_t^i,\mu_t^{X,2N})\big\|^2\big]\\ \lesssim\;&\sup_{t\in[0,T]}\mathbb{E}\big[\big\|\partial_\mu\mathcal{V}(t,\mu_t^{2N})(Y_t^{i,2N})^T\big(\sigma(Y_t^{i,2N},\mu_t^{2N})-\sigma(X_t^i,\mu_t^{X,2N})\big)\big\|^2\big]\\ &+\sup_{t\in[0,T]}\mathbb{E}\big[\big\|\big(\partial_\mu\mathcal{V}(t,\mu_t^{2N})(Y_t^{i,2N})^T-\partial_\mu\mathcal{V}(t,\mu_t^{X,2N})(X_t^i)^T\big)\sigma(X_t^i,\mu_t^{X,2N})\big\|^2\big]\\ \lesssim\;&\sup_{t\in[0,T]}\mathbb{E}\big[\big\|\partial_\mu\mathcal{V}(t,\mu_t^{2N})(Y_t^{i,2N})^T\big(\sigma(Y_t^{i,2N},\mu_t^{2N})-\sigma(X_t^i,\mu_t^{X,2N})\big)\big\|^2\big]\\ &+\sup_{t\in[0,T]}\Big(\mathbb{E}\big[\big\|\big(\partial_\mu\mathcal{V}(t,\mu_t^{2N})(Y_t^{i,2N})-\partial_\mu\mathcal{V}(t,\mu_t^{X,2N})(X_t^i)\big\|^4\big]\Big)^{1/2}\Big(\mathbb{E}\big[\big\|\sigma(X_t^i,\mu_t^{X,2N})\big\|^4\big]\Big)^{1/2}\\ \lesssim\;&\sup_{t\in[0,T]}\mathbb{E}\big[|Y_t^{i,2N}-X_t^i|^2\big]+\frac{1}{2N}\sum_{j=1}^{2N}\sup_{t\in[0,T]}\mathbb{E}\big[|Y_t^{j,2N}-X_t^j|^2\big]\end{aligned}$$

$$+\Big(\sup_{t\in[0,T]}\mathbb{E}\big[|Y_t^{i,2N}-X_t^i|^4\big]\Big)^{1/2}+\Big(\frac{1}{2N}\sum_{j=1}^{2N}\sup_{t\in[0,T]}\mathbb{E}\big[|Y_t^{j,2N}-X_t^j|^4\big]\Big)^{1/2}\lesssim\frac{1}{N}. \tag{5.1}$$

Similarly, we can show that

$$\sup_{t\in[0,T]}\mathbb{E}\big[\big\|\Sigma(t,Y_t^{i,2N,(1)},\mu_t^{2N,(1)})-\Sigma(t,X_t^i,\mu_t^{X,N})\big\|^2\big]\lesssim\frac{1}{N}. \tag{5.2}$$

Next, we apply Proposition 4.1 to σ and $\partial_\mu\mathcal{V}(t.\cdot)(\cdot)$. (Note that the constant C in Proposition 4.1 corresponding to $\varphi=\partial_\mu\mathcal{V}(t,\cdot)(\cdot)$ does not depend on time, since the first, second and third order derivatives in measure of this function are uniformly bounded in time.) By a similar calculation as (5.1), we obtain that

$$\begin{aligned}
&\sup_{t\in[0,T]}\mathbb{E}\big[\big\|\Sigma(t,X_t^i,\mu_t^{X,2N})-\Sigma(t,X_t^i,\mu_t^X)\big\|^2\big]\\
\lesssim\ &\sup_{t\in[0,T]}\mathbb{E}\big[\big\|\partial_\mu\mathcal{V}(t,\mu_t^{X,2N})(X_t^i)^T\big(\sigma(X_t^i,\mu_t^{X,2N})-\sigma(X_t^i,\mu_t^X)\big)\big\|^2\big]\\
&+\sup_{t\in[0,T]}\Big(\mathbb{E}\big[\big\|\big(\partial_\mu\mathcal{V}(t,\mu_t^{X,2N})(X_t^i)-\partial_\mu\mathcal{V}(t,\mu_t^X)(X_t^i)\big\|^4\big]\Big)^{1/2}\Big(\mathbb{E}\big[\big\|\sigma(X_t^i,\mu_t^X)\big\|^4\big]\Big)^{1/2}\lesssim\frac{1}{N}.
\end{aligned} \tag{5.3}$$

Similarly,

$$\sup_{t\in[0,T]}\mathbb{E}\big[\big\|\Sigma(t,X_t^i,\mu_t^{X,N})-\Sigma(t,X_t^i,\mu_t^X)\big\|^2\big]\lesssim\frac{1}{N}. \tag{5.4}$$

A combination of (5.1), (5.2), (5.3) and (5.4) gives

$$\mathbb{E}\Big[\Big(\frac{1}{2N}\sum_{i=1}^{N}\int_0^T\Sigma(s,Y_s^{i,2N},\mu_s^{2N})-\Sigma(s,Y_s^{i,2N,(1)},\mu_s^{2N,(1)})dW_s^i\Big)^2\Big]\lesssim\frac{1}{N^2}.$$

Similarly,

$$\mathbb{E}\Big[\Big(\frac{1}{2N}\sum_{i=N+1}^{2N}\int_0^T\Sigma(s,Y_s^{i,2N},\mu_s^{2N})-\Sigma(s,Y_s^{i,2N,(2)},\mu_s^{2N,(2)})dW_s^i\Big)^2\Big]\lesssim\frac{1}{N^2}.$$

Consequently, $\mathbb{E}[\mathscr{S}^2]\lesssim\frac{1}{N^2}$. □

We now perform an analysis on the order of interactions of this algorithm by assuming that b and σ are of the forms (2.9) and (2.10) respectively. Recall that, by Theorem B.2,

$$|\mathbb{E}[\Phi(\mu_T^{N_\ell})]-\Phi(\mu_T^X)|\leq\frac{C}{N_\ell}. \tag{i}$$

Moreover, by Theorem 5.1, we have

$$\mathrm{Var}\Big[\Phi(\mu_T^{N_\ell,\theta,\ell}) - \frac{1}{2}\Big(\Phi(\mu_T^{N_\ell,(1),\theta,\ell}) + \Phi(\mu_T^{N_\ell,(2),\theta,\ell})\Big)\Big] \le \frac{C}{N_\ell^2}. \tag{ii}$$

By Definition 2.5, the order of interactions of the antithetic difference is bounded by

$$\text{Order of interactions}\Big[\Phi(\mu_T^{N_\ell,\theta,\ell}) - \frac{1}{2}\Big(\Phi(\mu_T^{N_\ell,(1),\theta,\ell}) + \Phi(\mu_T^{N_\ell,(2),\theta,\ell})\Big)\Big] \le C N_\ell^{p+1}. \tag{iii}$$

Properties (i) to (iii) allow us to conclude the order of interactions of the theoretical antithetic MLMC estimator.

Theorem 5.2 (Order of interactions of the theoretical antithetic MLMC estimator (1.2))**.** *Assume* (Int). *Suppose that b and σ are of the forms* (2.9) *and* (2.10) *respectively. Furthermore, suppose that $b, \sigma \in \mathcal{M}_4(\mathbb{R}^d \times \mathcal{P}_2(\mathbb{R}^d))$ and $\Phi \in \mathcal{M}_4(\mathcal{P}_2(\mathbb{R}^d))$. Then there exist constants $C_1, C_2 > 0$ such that for any $\epsilon < e^{-1}$, there exist a value L and a sequence $\{M_\ell\}_{\ell=0}^L$ such that the mean-square error of $\mathcal{A}^{\text{A- MLMC}}$ (given by* (1.2)*) is bounded by*

$$\mathbb{E}\big[\big(\mathcal{A}^{\text{A- MLMC}} - \Phi(\mu_T^X)\big)^2\big] \le C_1 \epsilon^2$$

and the order of interactions of $\mathcal{A}^{\text{A- MLMC}}$ is bounded by

$$\text{Order of interactions}\,\big(\mathcal{A}^{\text{A- MLMC}}\big) \le \begin{cases} C_2 \epsilon^{-2}(\log \epsilon)^2, & p = 1, \\ C_2 \epsilon^{-1-p}, & p > 1. \end{cases}$$

Proof. The proof of this theorem is almost identical to the proof of Theorem 1 in [12] and is therefore omitted. Nonetheless, the proof for the complexity of the antithetic MLMC estimator with time discretisation (Theorem 6.3) will be presented in detail for completeness. □

6 Antithetic MLMC with Euler time discretisation

In this section, we construct an MLMC estimator in the same way as the previous section, but with time discretisation. We set

$$N_\ell := 2^\ell, \qquad h_\ell := \frac{T}{N_\ell}, \qquad \ell \in \{0, \dots, L\}.$$

We also set the two sub-particle systems to have the same number of particles. We define the pair of sub-particle systems to $\{Z^{i,2N,h}\}_{i=1}^{2N}$ as

$$Z_t^{i,2N,(1),h} = \xi_i + \int_0^t b\Big(Z_{\eta(r)}^{i,2N,(1),h}, \mu_{\eta(r)}^{Z,2N,(1),h}\Big)\,dr + \int_0^t \sigma\Big(Z_{\eta(r)}^{i,2N,(1),h}, \mu_{\eta(r)}^{Z,2N,(1),h}\Big)\,dW_r^i, \quad 1 \le i \le N,$$
$$Z_t^{i,2N,(2),h} = \xi_i + \int_0^t b\Big(Z_{\eta(r)}^{i,2N,(2),h}, \mu_{\eta(r)}^{Z,2N,(2),h}\Big)\,dr + \int_0^t \sigma\Big(Z_{\eta(r)}^{i,2N,(2),h}, \mu_{\eta(r)}^{Z,2N,(2),h}\Big)\,dW_r^i, \quad N+1 \le i \le 2N,$$

where

$$\mu_r^{Z,2N,(1),h} := \frac{1}{N}\sum_{i=1}^{N} \delta_{Z_r^{i,2N,(1),h}} \qquad \text{and} \qquad \mu_r^{Z,2N,(2),h} := \frac{1}{N}\sum_{i=N+1}^{2N} \delta_{Z_r^{i,2N,(2),h}}.$$

Therefore, we define the MLMC estimator with time discretisation as

$$\begin{aligned}\mathcal{A}^{\text{A- MLMC},t} \quad &:= \quad \frac{1}{M_0}\sum_{\theta=1}^{M_0} \Phi(\mu_T^{Z,N_0,h_0,\theta,0}) \\ &+ \sum_{\ell=1}^{L}\left[\frac{1}{M_\ell}\sum_{\theta=1}^{M_\ell}\left[\Phi(\mu_T^{Z,N_\ell,h_\ell,\theta,\ell}) - \frac{1}{2}\Big(\Phi(\mu_T^{Z,N_\ell,(1),2h_\ell,\theta,\ell}) + \Phi(\mu_T^{Z,N_\ell,(2),2h_\ell,\theta,\ell})\Big)\right]\right], \quad (6.1)\end{aligned}$$

where $\mu_T^{Z,N_\ell,h_\ell,\theta,\ell}$, $\mu_T^{Z,N_\ell,(1),2h_\ell,\theta,\ell}$ and $\mu_T^{Z,N_\ell,(2),2h_\ell,\theta,\ell}$ are defined similarly as μ_T^{Z,N_ℓ,h_ℓ}, $\mu_T^{Z,N_\ell,(1),2h_\ell}$, and $\mu_T^{Z,N_\ell,(2),2h_\ell}$ respectively, but correspond to the $\sum_{\ell=0}^{L} M_\ell$ independent clouds of particles indexed by $\ell \in \{0,\ldots,L\}$ and $\theta \in \{1,\ldots,M_\ell\}$. Each cloud (indexed by ℓ, θ) has particles with initial conditions $\xi_{i,\ell,\theta}$, $i \in \{1,\ldots,N_\ell\}$, driven by Brownian motions $W^{i,\ell,\theta}$, $i \in \{1,\ldots,N_\ell\}$, where $\{\xi_{i,\ell,\theta}\}$ and $\{W^{i,\ell,\theta}\}$ are independent over i, ℓ and θ.

To prove the analogue of Theorem 5.1 with time discretisation, we need the following lemma that provides a strong error bound between the particle system (2.7) and the Euler scheme (2.11). Since we require a higher-order approximation in time discretisation, we restrict ourselves to the case of constant diffusion, in order to avoid the complication of introducing the Milstein scheme of time discretisation. Note that, under (Lip), it follows by a standard Gronwall-type argument that

$$\sup_{N\in\mathbb{N}}\sup_{u\in[0,T]} \mathbb{E}\left[\frac{1}{N}\sum_{i=1}^{N}|Y_u^{i,N}|^2\right] < +\infty, \quad \sup_{N\in\mathbb{N}}\sup_{u\in[0,T]} \mathbb{E}\left[\frac{1}{N}\sum_{i=1}^{N}|Z_{\eta(u)}^{i,N,h}|^2\right] < +\infty, \qquad (6.2)$$

for some $C > 0$.

Lemma 6.1. *Suppose that $b \in \mathcal{M}_2(\mathbb{R}^d \times \mathcal{P}_2(\mathbb{R}^d))$ and σ is constant. Then*

$$\sup_{N\in\mathbb{N}}\sup_{s\in[0,T]} \mathbb{E}\left[W_2(\mu_s^N, \mu_s^{Z,N,h})^2\right] \leq Ch^2,$$

for some constant C that does not depend on h.

Proof. The proof is presented in dimension one, for simplicity of notations. By Itô's formula,

$$(Y_t^{i,N} - Z_t^{i,N,h})^2 = 2\int_0^t (Y_s^{i,N} - Z_s^{i,N,h})\big(b(Y_s^{i,N}, \mu_s^N) - b(Z_{\eta(s)}^{i,N,h}, \mu_{\eta(s)}^{Z,N,h})\big)\, ds.$$

Take $0 \leq t' \leq t \leq T$. Then

$$\frac{1}{N}\sum_{i=1}^{N}\mathbb{E}(Y_{t'}^{i,N} - Z_{t'}^{i,N,h})^2 \quad = \quad \frac{2}{N}\sum_{i=1}^{N}\mathbb{E}\left[\int_0^{t'} (Y_s^{i,N} - Z_s^{i,N,h})\big(b(Y_s^{i,N}, \mu_s^N) - b(Z_s^{i,N,h}, \mu_s^{Z,N,h})\big)\, ds\right]$$

$$+\frac{2}{N}\sum_{i=1}^{N}\mathbb{E}\Big[\int_0^{t'}(Y_s^{i,N}-Z_s^{i,N,h})\big(b(Z_s^{i,N,h},\mu_s^{Z,N,h})-b(Z_{\eta(s)}^{i,N,h},\mu_{\eta(s)}^{Z,N,h})\big)\,ds\Big]. \tag{6.3}$$

We first bound the first term of (6.3).

$$\begin{aligned}
&\frac{2}{N}\sum_{i=1}^{N}\mathbb{E}\Big[\int_0^{t'}(Y_s^{i,N}-Z_s^{i,N,h})\big(b(Y_s^{i,N},\mu_s^{N})-b(Z_s^{i,N,h},\mu_s^{Z,N,h})\big)\,ds\Big]\\
\leq\;& C\mathbb{E}\Big[\frac{1}{N}\sum_{i=1}^{N}\int_0^{t'}|Y_s^{i,N}-Z_s^{i,N,h}|\Big(|Y_s^{i,N}-Z_s^{i,N,h}|+\Big(\frac{1}{N}\sum_{j=1}^{N}|Y_s^{i,N}-Z_s^{i,N,h}|^2\Big)^{1/2}\Big)\,ds\Big]\\
\leq\;& \frac{C}{N}\sum_{i=1}^{N}\int_0^{t'}\mathbb{E}|Y_s^{i,N}-Z_s^{i,N,h}|^2\,ds\leq C\int_0^{t}\sup_{u\in[0,s]}\Big[\frac{1}{N}\sum_{i=1}^{N}\mathbb{E}|Y_u^{i,N}-Z_u^{i,N,h}|^2\Big]\,ds. \qquad (6.4)
\end{aligned}$$

To bound the second term of (6.3), we proceed as in the proof of Theorem B.3 (through Proposition 3.1 of [10] that relates real derivatives to L-derivatives) by applying Itô's formula to the process

$$\Big\{(Y_s^{i,N}-Z_s^{i,N,h})\big(b(Z_s^{i,N,h},\mu_s^{Z,N,h})-b(Z_{t_0}^{i,N,h},\mu_{t_0}^{Z,N,h})\big)\Big\}_{s\geq t_0},$$

which gives

$$\begin{aligned}
&(Y_s^{i,N}-Z_s^{i,N,h})\big(b(Z_s^{i,N,h},\mu_s^{Z,N,h})-b(Z_{t_0}^{i,N,h},\mu_{t_0}^{Z,N,h})\big)\\
=\;&\int_{t_0}^{s}\big(b(Z_u^{i,N,h},\mu_u^{Z,N,h})-b(Z_{t_0}^{i,N,h},\mu_{t_0}^{Z,N,h})\big)\,d(Y_u^{i,N}-Z_u^{i,N,h})\\
&+\sum_{j\neq i}\int_{t_0}^{s}(Y_u^{i,N}-Z_u^{i,N,h})\Big(\frac{1}{N}\partial_\mu b(Z_u^{i,N,h},\mu_u^{Z,N,h})(Z_u^{j,N,h})\Big)\,dZ_u^{j,N,h}\\
&+\int_{t_0}^{s}(Y_u^{i,N}-Z_u^{i,N,h})\Big(\frac{1}{N}\partial_\mu b(Z_u^{i,N,h},\mu_u^{Z,N,h})(Z_u^{i,N,h})+\partial_x b(Z_u^{i,N,h},\mu_u^{Z,N,h})\Big)\,dZ_u^{i,N,h}\\
&+\frac{1}{2}\sum_{j\neq i}\int_{t_0}^{s}(Y_u^{i,N}-Z_u^{i,N,h})\Big(\frac{1}{N}\partial_v\partial_\mu b(Z_u^{i,N,h},\mu_u^{Z,N,h})(Z_u^{j,N,h})\\
&+\frac{1}{N^2}\partial_\mu^2 b(Z_u^{i,N,h},\mu_u^{Z,N,h})(Z_u^{j,N,h},Z_u^{j,N,h})\Big)\,d\left\langle Z^{j,N,h}\right\rangle_u\\
&+\frac{1}{2}\int_{t_0}^{s}(Y_u^{i,N}-Z_u^{i,N,h})\Big(\frac{1}{N}\partial_v\partial_\mu b(Z_u^{i,N,h},\mu_u^{Z,N,h})(Z_u^{i,N,h})\\
&+\frac{1}{N^2}\partial_\mu^2 b(Z_u^{i,N,h},\mu_u^{Z,N,h})(Z_u^{i,N,h},Z_u^{i,N,h})+\frac{2}{N}\partial_x\partial_\mu b(Z_u^{i,N,h},\mu_u^{Z,N,h})(Z_u^{i,N,h})\\
&+\partial_x^2 b(Z_u^{i,N,h},\mu_u^{Z,N,h})\Big)\,d\left\langle Z^{i,N,h}\right\rangle_u.
\end{aligned}$$

Putting $t_0 = \eta(s)$, taking average of i from 1 to N, taking expectation and rewriting terms, we have

$$\frac{1}{N}\sum_{i=1}^{N}\mathbb{E}\Big[(Y_s^{i,N} - Z_s^{i,N,h})\big(b(Z_s^{i,N,h},\mu_s^{Z,N,h}) - b(Z_{\eta(s)}^{i,N,h},\mu_{\eta(s)}^{Z,N,h})\big)\Big] = \mathcal{I}_1 + \mathcal{I}_2,$$

where

$$\mathcal{I}_1 := \frac{1}{N}\sum_{i=1}^{N}\mathbb{E}\Big[\int_{\eta(s)}^{s}\big(b(Z_u^{i,N,h},\mu_u^{Z,N,h}) - b(Z_{\eta(s)}^{i,N,h},\mu_{\eta(s)}^{Z,N,h})\big)\big(b(Y_u^{i,N},\mu_u^{N}) - b(Z_{\eta(s)}^{i,N,h},\mu_{\eta(s)}^{Z,N,h})\big)\,du\Big]$$

and

$$\mathcal{I}_2 := \frac{1}{N}\sum_{i=1}^{N}\mathbb{E}\Big[\int_{\eta(s)}^{s}(Y_u^{i,N} - Z_u^{i,N,h})\mathcal{D}_u^i\,du\Big],$$

where

$$\begin{aligned}
\mathcal{D}_u^i \;&:=\; \frac{1}{N}\sum_{j=1}^{N}\Big(\partial_\mu b(Z_u^{i,N,h},\mu_u^{Z,N,h})(Z_u^{j,N,h})b(Z_{\eta(u)}^{j,N,h},\mu_{\eta(u)}^{Z,N,h})\Big) + \partial_x b(Z_u^{i,N,h},\mu_u^{Z,N,h})b(Z_{\eta(u)}^{i,N,h},\mu_{\eta(u)}^{Z,N,h})\\
&+\frac{1}{2}\sigma^2\sum_{j=1}^{N}\Big(\frac{1}{N^2}\partial_\mu^2 b(Z_u^{i,N,h},\mu_u^{Z,N,h})(Z_u^{j,N,h},Z_u^{j,N,h}) + \frac{1}{N}\partial_v\partial_\mu b(Z_u^{i,N,h},\mu_u^{Z,N,h})(Z_u^{j,N,h})\Big)\\
&+\frac{1}{2}\sigma^2\Big(\frac{2}{N}\partial_x\partial_\mu b(Z_u^{i,N,h},\mu_u^{Z,N,h})(Z_u^{i,N,h}) + \partial_x^2 b(Z_u^{i,N,h},\mu_u^{Z,N,h})\Big).
\end{aligned}$$

By the hypothesis on b, all derivatives of b are uniformly bounded. Moreover, by (Lip), b has linear growth in space and measure. Therefore,

$$\frac{1}{N}\sum_{i=1}^{N}\mathbb{E}|\mathcal{D}_u^i|^2 \le C\Big(1 + \frac{1}{N}\sum_{i=1}^{N}\mathbb{E}|Z_{\eta(u)}^{i,N,h}|^2\Big).$$

Then, by (6.2),

$$\sup_{u\in[0,T]}\Big[\frac{1}{N}\sum_{i=1}^{N}\mathbb{E}|\mathcal{D}_u^i|^2\Big] \le C.$$

By first applying the Cauchy-Schwarz inequality to the expectation operator and then to the sum,

$$\begin{aligned}
\mathcal{I}_2 \;&\le\; \int_{\eta(s)}^{s}\Big(\frac{1}{N}\sum_{i=1}^{N}\mathbb{E}|Y_u^{i,N} - Z_u^{i,N,h}|^2\Big)^{1/2}\Big(\frac{1}{N}\sum_{i=1}^{N}\mathbb{E}|\mathcal{D}_u^i|^2\Big)^{1/2}du\\
&\le\; C\Big(\sup_{u\in[0,s]}\frac{1}{N}\sum_{i=1}^{N}\mathbb{E}|Y_u^{i,N} - Z_u^{i,N,h}|^2\Big)^{1/2}h
\end{aligned}$$

$$\leq C\Big(\frac{1}{2}\sup_{u\in[0,s]}\frac{1}{N}\sum_{i=1}^{N}\mathbb{E}|Y_u^{i,N}-Z_u^{i,N,h}|^2+\frac{1}{2}h^2\Big). \tag{6.5}$$

Next, we rewrite $\mathcal{I}_1$ as

$$\begin{aligned}\mathcal{I}_1 &= \frac{1}{N}\sum_{i=1}^{N}\mathbb{E}\Big[\int_{\eta(s)}^{s}\big(b(Z_u^{i,N,h},\mu_u^{Z,N,h})-b(Z_{\eta(s)}^{i,N,h},\mu_{\eta(s)}^{Z,N,h})\big)\big(b(Y_u^{i,N},\mu_u^{N})-b(Z_u^{i,N,h},\mu_u^{Z,N,h})\big)\,du\Big]\\ &\quad+\frac{1}{N}\sum_{i=1}^{N}\mathbb{E}\Big[\int_{\eta(s)}^{s}\big(b(Z_u^{i,N,h},\mu_u^{Z,N,h})-b(Z_{\eta(s)}^{i,N,h},\mu_{\eta(s)}^{Z,N,h})\big)^2\,du\Big].\end{aligned}$$

It is clear that

$$\frac{1}{N}\sum_{i=1}^{N}\mathbb{E}\Big[\int_{\eta(s)}^{s}\big(b(Z_u^{i,N,h},\mu_u^{Z,N,h})-b(Z_{\eta(s)}^{i,N,h},\mu_{\eta(s)}^{Z,N,h})\big)^2\,du\Big]\leq Ch^2. \tag{6.6}$$

By the Cauchy-Schwarz inequality and (Lip), the first term of $\mathcal{I}_1$ is bounded by

$$\begin{aligned}&\frac{1}{N}\sum_{i=1}^{N}\mathbb{E}\Big[\int_{\eta(s)}^{s}\big(b(Z_u^{i,N,h},\mu_u^{Z,N,h})-b(Z_{\eta(s)}^{i,N,h},\mu_{\eta(s)}^{Z,N,h})\big)\big(b(Y_u^{i,N},\mu_u^{N})-b(Z_u^{i,N,h},\mu_u^{Z,N,h})\big)\,du\Big]\\ &\leq \frac{1}{N}\sum_{i=1}^{N}\int_{\eta(s)}^{s}\Big(\mathbb{E}\Big|b(Z_u^{i,N,h},\mu_u^{Z,N,h})-b(Z_{\eta(s)}^{i,N,h},\mu_{\eta(s)}^{Z,N,h})\Big|^2\Big)^{1/2}\\ &\qquad\Big(\mathbb{E}\Big|b(Y_u^{i,N},\mu_u^{N})-b(Z_u^{i,N,h},\mu_u^{Z,N,h})\Big|^2\Big)^{1/2}du\\ &\leq \frac{1}{N}\sum_{i=1}^{N}C\sqrt{h}\int_{\eta(s)}^{s}\Big(\mathbb{E}\Big|b(Y_u^{i,N},\mu_u^{N})-b(Z_u^{i,N,h},\mu_u^{Z,N,h})\Big|^2\Big)^{1/2}du\\ &\leq \frac{1}{N}\sum_{i=1}^{N}C\sqrt{h}\int_{\eta(s)}^{s}\Big(\mathbb{E}|Y_u^{i,N}-Z_u^{i,N,h}|^2+\frac{1}{N}\sum_{j=1}^{N}\mathbb{E}|Y_u^{j,N}-Z_u^{j,N,h}|^2\Big)^{1/2}du\\ &\leq \frac{2}{N}\sum_{i=1}^{N}C\sqrt{h}\int_{\eta(s)}^{s}(\mathbb{E}|Y_u^{i,N}-Z_u^{i,N,h}|^2)^{1/2}\,du\\ &\leq 2Ch^{3/2}\Big[\sup_{u\in[0,s]}\Big(\frac{1}{N}\sum_{i=1}^{N}\mathbb{E}|Y_u^{i,N}-Z_u^{i,N,h}|^2\Big)\Big]^{1/2}\\ &\leq C\Big(h^3+\sup_{u\in[0,s]}\Big(\frac{1}{N}\sum_{i=1}^{N}\mathbb{E}|Y_u^{i,N}-Z_u^{i,N,h}|^2\Big)\Big).\end{aligned} \tag{6.7}$$

A combination of (6.3), (6.4), (6.5), (6.6) and (6.7) gives

$$\sup_{u\in[0,t]}\Big[\frac{1}{N}\sum_{i=1}^{N}\mathbb{E}(Y_u^{i,N}-Z_u^{i,N,h})^2\Big]\leq C\Big(\int_0^t\sup_{u\in[0,s]}\Big[\frac{1}{N}\sum_{i=1}^{N}\mathbb{E}|Y_u^{i,N}-Z_u^{i,N,h}|^2\Big]ds+h^2\Big),\quad\forall t\in[0,T],$$

which implies by Gronwall's inequality that

$$\sup_{u\in[0,T]}\left[\frac{1}{N}\sum_{i=1}^{N}\mathbb{E}(Y_u^{i,N}-Z_u^{i,N,h})^2\right]\leq Ch^2.$$

Since the constant C does not depend on N, we conclude that

$$\sup_{N\in\mathbb{N}}\sup_{s\in[0,T]}\mathbb{E}\big[W_2(\mu_s^N,\mu_s^{Z,N,h})^2\big]\leq\sup_{N\in\mathbb{N}}\sup_{s\in[0,T]}\left[\frac{1}{N}\sum_{i=1}^{N}\mathbb{E}(Y_s^{i,N}-Z_s^{i,N,h})^2\right]\leq Ch^2.$$

□

A combination of Lemma 6.1 and Theorem 5.1 immediately gives the following result.

Theorem 6.2 (Variance of antithetic difference)**.** *Assume* (Int)*. Suppose that* $b\in\mathcal{M}_4\big(\mathbb{R}^d\times\mathcal{P}_2(\mathbb{R}^d)\big)$ *and* $\Phi\in\mathcal{M}_4\big(\mathcal{P}_2(\mathbb{R}^d)\big)$*. Moreover, suppose that* σ *is constant. Then*

$$\mathit{Var}\Big[\Phi(\mu_T^{Z,N,h})-\frac{1}{2}\big(\Phi(\mu_T^{Z,N,(1),2h})+\Phi(\mu_T^{Z,N,(2),2h})\big)\Big]\leq C\Big(\frac{1}{N^2}+h^2\Big),$$

where C *is a constant that depends on* Φ*,* b*,* σ *and* T*, but does not depend on* N *or* h*.*

For an estimator involving an Euler numerical scheme with discretisation step h, its *computational complexity* is defined by

$$\text{Computational complexity}:=h^{-1}\Big(\text{Order of interactions of estimator}\Big).$$

As before, we perform an analysis on the complexity of this algorithm by assuming that b is of the form (2.9) and that σ is constant. By Theorem B.3, since $h_\ell=\frac{T}{N_\ell}$,

$$|\mathbb{E}[\Phi(\mu_T^{Z,N_\ell,h_\ell})]-\Phi(\mu_T^X)|\leq\frac{C}{N_\ell}. \tag{I}$$

Moreover, by Theorem 6.2, we have

$$\text{Var}\Big[\Phi(\mu_T^{Z,N_\ell,h_\ell,\theta,\ell})-\frac{1}{2}\Big(\Phi(\mu_T^{Z,N_\ell,(1),2h_\ell,\theta,\ell})+\Phi(\mu_T^{Z,N_\ell,(2),2h_\ell,\theta,\ell})\Big)\Big]\leq\frac{C}{N_\ell^2}. \tag{II}$$

Finally, by Definition 2.5, the complexity of the antithetic difference is bounded by

$$\text{Complexity}\Big[\Phi(\mu_T^{Z,N_\ell,h_\ell,\theta,\ell})-\frac{1}{2}\Big(\Phi(\mu_T^{Z,N_\ell,(1),2h_\ell,\theta,\ell})+\Phi(\mu_T^{Z,N_\ell,(2),2h_\ell,\theta,\ell})\Big)\Big]\leq CN_\ell^{p+2}. \tag{III}$$

Theorem 6.3 (Complexity of antithetic MLMC with time discretisation for estimator (6.1))**.** *Assume* (Int)*. Suppose that b is of the form* (2.9)*. Furthermore, suppose that $b \in \mathcal{M}_4\big(\mathbb{R}^d \times \mathcal{P}_2(\mathbb{R}^d)\big)$, $\Phi \in \mathcal{M}_4\big(\mathcal{P}_2(\mathbb{R}^d)\big)$ and σ is constant. Then there exist constants $C_1, C_2 > 0$ such that for any $\epsilon < e^{-1}$, there exist a value L and a sequence $\{M_\ell\}_{\ell=0}^L$ such that the mean-square error of $\mathcal{A}^{\text{A- MLMC},t}$ is bounded by*

$$\mathbb{E}\big[\big(\mathcal{A}^{\text{A- MLMC},t} - \Phi(\mu_T^X)\big)^2\big] \leq C_1 \epsilon^2$$

and the complexity of $\mathcal{A}^{\text{A- MLMC},t}$ is bounded by

$$\text{Complexity}\left(\mathcal{A}^{\text{A- MLMC},t}\right) \leq C_2 \epsilon^{-2-p}.$$

Proof. As in Theorem 5.2, the proof of this theorem is also almost identical to the proof of Theorem 1 in [12]. Nonetheless, we present the proof with explicit expressions for L and $\{M_\ell\}_{\ell=0}^L$ so that practitioners can implement this algorithm easily. Set

$$L := \lceil \log_2(\sqrt{2}\epsilon^{-1}) \rceil, \qquad M_\ell := \lceil 2\epsilon^{-2} 2^{pL/2}(1-2^{-p/2})^{-1} 2^{-(p+4)\ell/2} \rceil, \ \ell \in \{0, \dots, L\}.$$

By standard decomposition of the mean-square error, we have

$$\text{Mean-square error} = \text{Var}(\mathcal{A}^{\text{A- MLMC},t}) + (\mathbb{E}(\mathcal{A}^{\text{A- MLMC},t}) - \Phi(\mu_T^X))^2.$$

By the choice of L, $2^{-L} \leq \frac{\epsilon}{\sqrt{2}}$. Therefore, by Property (I),

$$|\mathbb{E}(\mathcal{A}^{\text{A- MLMC},t}) - \Phi(\mu_T^X)|^2 = |\mathbb{E}[\Phi(\mu_T^{Z,N_L,h_L})] - \Phi(\mu_T^X)|^2 \leq \left(\frac{C}{N_L}\right)^2 = (C2^{-L})^2 \leq C^2\left(\frac{\epsilon^2}{2}\right). \tag{6.8}$$

On the other hand, by Property (II) and the choice of $\{M_\ell\}_{\ell=0}^L$,

$$\begin{aligned}
\text{Var}(\mathcal{A}^{\text{A- MLMC},t}) \leq \sum_{\ell=0}^{L} \frac{1}{M_\ell^2}\left[\sum_{\theta=1}^{M_\ell} \frac{C}{N_\ell^2}\right] \leq \sum_{\ell=0}^{L} \frac{C}{M_\ell} 2^{-2\ell} \quad &\leq \quad \sum_{\ell=0}^{L} C 2^{-2\ell}\left(2^{-1}\epsilon^2 2^{-pL/2}(1-2^{-p/2}) 2^{(p+4)\ell/2}\right) \\
&= \quad C2^{-1}\epsilon^2 2^{-pL/2}(1-2^{-p/2}) \sum_{\ell=0}^{L} 2^{p\ell/2} \\
&< \quad \frac{1}{2} C\epsilon^2.
\end{aligned}$$

This verifies that the mean-square error is bounded by $\frac{1}{2}(C^2 + C)\epsilon^2$. Next, we note that

$$M_\ell \leq 2\epsilon^{-2} 2^{pL/2}(1-2^{-p/2})^{-1} 2^{-(p+4)\ell/2} + 1$$

and hence, by Property (III),

$$\text{Complexity}(\mathcal{A}^{\text{A- MLMC},t}) \leq C\left(\sum_{\ell=0}^{L} 2\epsilon^{-2} 2^{pL/2}(1-2^{-p/2})^{-1} 2^{-(p+4)\ell/2} 2^{(p+2)\ell} + \sum_{\ell=0}^{L} 2^{(p+2)\ell}\right). \tag{6.9}$$

Note that the choice of L implies that $2^L \leq 2\sqrt{2}\epsilon^{-1}$.

$$\begin{aligned}\sum_{\ell=0}^{L} 2\epsilon^{-2}2^{pL/2}(1-2^{-p/2})^{-1}2^{-(p+4)\ell/2}2^{(p+2)\ell} &= 2\epsilon^{-2}2^{pL/2}(1-2^{-p/2})^{-1}\sum_{\ell=0}^{L}2^{p\ell/2} \\ &< 2\epsilon^{-2}2^{pL/2}(1-2^{-p/2})^{-1}\Big(2^{pL/2}(1-2^{-p/2})^{-1}\Big) \\ &= 2\epsilon^{-2}2^{pL}(1-2^{-p/2})^{-2} \\ &\leq 2\big(2\sqrt{2}\big)^p(1-2^{-p/2})^{-2}\epsilon^{-2-p}. \end{aligned} \tag{6.10}$$

Similarly,

$$\sum_{\ell=0}^{L} 2^{(p+2)\ell} \leq \frac{2^{(p+2)L}}{1-2^{-(p+2)}} \leq \frac{(2\sqrt{2})^{p+2}}{1-2^{-(p+2)}}\epsilon^{-(p+2)}. \tag{6.11}$$

A combination of (6.9), (6.10) and (6.11) finally gives

$$\text{Complexity}(\mathcal{A}^{\text{A- MLMC},t}) \leq C\left(2\big(2\sqrt{2}\big)^p(1-2^{-p/2})^{-2} + \frac{(2\sqrt{2})^{p+2}}{1-2^{-(p+2)}}\right)\epsilon^{-2-p}.$$

□

A Appendix: A review of linear functional derivatives and L-derivatives

Our method of proof is based on the theory of calculus on the Wasserstein space. A substantial portion of the appendix is extracted from a recent work [11]. We make an intensive use of the so-called "L-derivatives" and "linear functional derivatives" that we recall now, following essentially [7]. We also introduce higher-order versions of these derivatives as they are needed in the proofs.

Linear functional derivatives

A continuous function $\frac{\delta U}{\delta m} : \mathcal{P}_2(\mathbb{R}^d) \times \mathbb{R}^d \to \mathbb{R}$ is said to be the *linear functional derivative* of $U : \mathcal{P}_2(\mathbb{R}^d) \to \mathbb{R}$, if

- for any bounded set $\mathcal{K} \subset \mathcal{P}_2(\mathbb{R}^d)$, $y \mapsto \frac{\delta U}{\delta m}(m, y)$ has at most quadratic growth in y uniformly in $m \in \mathcal{K}$,
- for any $m, m' \in \mathcal{P}_2(\mathbb{R}^d)$,

$$U(m') - U(m) = \int_0^1 \int_{\mathbb{R}^d} \frac{\delta U}{\delta m}((1-s)m + sm', y)\,(m'-m)(dy)\,ds. \tag{A.1}$$

For the purpose of our work, we need to introduce derivatives at any order $p \geq 1$.

Definition A.1. For any $p \geq 1$, the p-th order linear functional of the function U is a continuous function from $\frac{\delta^p U}{\delta m^p} : \mathcal{P}_2(\mathbb{R}^d) \times (\mathbb{R}^d)^{p-1} \times \mathbb{R}^d \to \mathbb{R}$ satisfying

- for any bounded set $\mathcal{K} \subset \mathcal{P}_2(\mathbb{R}^d)$, $(y, y') \mapsto \frac{\delta^p U}{\delta m^p}(m, y, y')$ has at most quadratic growth in (y, y') uniformly in $m \in \mathcal{K}$,
- for any $m, m' \in \mathcal{P}_2(\mathbb{R}^d)$,

$$\frac{\delta^{p-1} U}{\delta m^{p-1}}(m', y) - \frac{\delta^{p-1} U}{\delta m^{p-1}}(m, y) = \int_0^1 \int_{\mathbb{R}^d} \frac{\delta^p U}{\delta m^p}((1-s)m + sm', y, y')\,(m' - m)(\mathrm{d}y')\,\mathrm{d}s,$$

provided that the $(p-1)$-th order derivative is well defined.

The above derivatives are defined up to an additive constant via (A.1). They are normalised by

$$\frac{\delta^p U}{\delta m^p}(m, y_1, \dots, y_p) = 0, \quad \text{if } y_i = 0 \text{ for some } i \in \{1, \dots, p\}. \tag{A.2}$$

L-derivatives

The above notion of linear functional derivatives is not enough for our work. We shall need to consider further derivatives in the non-measure argument of the derivative function.

If the function $y \mapsto \frac{\delta U}{\delta m}(m, y)$ is of class $\mathcal{C}^1$, we consider the *intrinsic* derivative of U that we denote

$$\partial_\mu U(m, y) := \partial_y \frac{\delta U}{\delta m}(m, y)\ .$$

The notation is borrowed from the literature on mean field games and corresponds to the notion of “L-derivative” introduced by P.-L. Lions in his lectures at Collège de France [33]. Traditionally, it is introduced by considering a lift on an L^2 space of the function U and using the Fréchet differentiability of this lift on this Hilbert space. The equivalence between the two notions is proved in [9, Tome I, Chapter 5], where the link with the notion of derivatives used in optimal transport theory is also made.

In this context, higher order derivatives are introduced by iterating the operator ∂_μ and the derivation in the non-measure arguments. Namely, at order 2, one considers

$$\mathcal{P}_2(\mathbb{R}^d) \times \mathbb{R}^d \ni (m, y) \mapsto \partial_y \partial_\mu U(m, y) \text{ and } \mathcal{P}_2(\mathbb{R}^d) \times \mathbb{R}^d \times \mathbb{R}^d \ni (m, y, y') \mapsto \partial^2_\mu U(m, y, y')\ .$$

Inspired by the work [13], for any $k \in \mathbb{N}$, we formally define the higher order derivatives in measures through the following iteration (provided that they actually exist): for any $k \geq 2$, $(i_1, \dots, i_k) \in \{1, \dots, d\}^k$ and $x_1, \dots, x_k \in \mathbb{R}^d$, the function $\partial^k_\mu f : \mathcal{P}_2(\mathbb{R}^d) \times (\mathbb{R}^d)^k \to (\mathbb{R}^d)^{\otimes k}$ is defined by

$$\Big(\partial^k_\mu f(\mu, x_1, \dots, x_k)\Big)_{(i_1, \dots, i_k)} := \left(\partial_\mu\left(\Big(\partial^{k-1}_\mu f(\cdot, x_1, \dots, x_{k-1})\Big)_{(i_1, \dots, i_{k-1})}\right)(\mu, x_k)\right)_{i_k}, \tag{A.3}$$

and its corresponding mixed derivatives in space $\partial_{v_k}^{\ell_k} \dots \partial_{v_1}^{\ell_1} \partial_\mu^k f : \mathcal{P}_2(\mathbb{R}^d) \times (\mathbb{R}^d)^k \to (\mathbb{R}^d)^{\otimes(k+\ell_1+\dots\ell_k)}$ are defined by

$$\left(\partial_{v_k}^{\ell_k} \dots \partial_{v_1}^{\ell_1} \partial_\mu^k f(\mu, x_1, \dots, x_k)\right)_{(i_1,\dots,i_k)} := \frac{\partial^{\ell_k}}{\partial x_k^{\ell_k}} \cdots \frac{\partial^{\ell_1}}{\partial x_1^{\ell_1}} \left[\left(\partial_\mu^k f(\mu, x_1, \dots, x_k)\right)_{(i_1,\dots,i_k)}\right], \quad \ell_1 \dots \ell_k \in \mathbb{N} \cup \{0\}. \tag{A.4}$$

Since this notation for higher order derivatives in measure is quite cumbersome, we introduce the following multi-index notation for brevity. This notation was first proposed in [13].

Definition A.2 (Multi-index notation)**.** Let n, ℓ be non-negative integers. Also, let $\boldsymbol{\beta} = (\beta_1, \dots, \beta_n)$ be an n-dimensional vector of non-negative integers. Then we call any ordered tuple of the form $(n, \ell, \boldsymbol{\beta})$ or $(n, \boldsymbol{\beta})$ a *multi-index*. For a function $f : \mathbb{R}^d \times \mathcal{P}_2(\mathbb{R}^d) \to \mathbb{R}$, the derivative $D^{(n,\ell,\boldsymbol{\beta})} f(x, \mu, v_1, \dots, v_n)$ is defined as

$$D^{(n,\ell,\boldsymbol{\beta})} f(x, \mu, v_1, \dots, v_n) := \partial_{v_n}^{\beta_n} \dots \partial_{v_1}^{\beta_1} \partial_x^\ell \partial_\mu^n f(x, \mu, v_1, \dots, v_n)$$

if this derivative is well-defined. For any function $\Phi : \mathcal{P}_2(\mathbb{R}^d) \to \mathbb{R}$, we define

$$D^{(n,\boldsymbol{\beta})} \Phi(\mu, v_1, \dots, v_n) := \partial_{v_n}^{\beta_n} \dots \partial_{v_1}^{\beta_1} \partial_\mu^n \Phi(\mu, v_1, \dots, v_n),$$

if this derivative is well-defined. Finally, we also define the *order* [2] $|(n, \ell, \boldsymbol{\beta})|$ (resp. $|(n, \boldsymbol{\beta})|$) by

$$|(n, \ell, \boldsymbol{\beta})| := n + \beta_1 + \dots \beta_n + \ell, \qquad |(n, \boldsymbol{\beta})| := n + \beta_1 + \dots \beta_n. \tag{A.5}$$

In our proofs, we aim to formulate sufficient conditions purely in terms of regularity of the drift and diffusion functions, as well as the test function. A class $\mathcal{M}_k$ of regularity in differentiating measures is proposed.

Definition A.3 (Class $\mathcal{M}_k$ of kth order differentiable functions)**.**

(i) The functions b and σ belong to class $\mathcal{M}_k(\mathbb{R}^d \times \mathcal{P}_2(\mathbb{R}^d))$, if the derivatives $D^{(n,\ell,\boldsymbol{\beta})} b(x, \mu, v_1, \dots, v_n)$ and $D^{(n,\ell,\boldsymbol{\beta})} \sigma(x, \mu, v_1, \dots, v_n)$ exist for every multi-index $(n, \ell, \boldsymbol{\beta})$ such that $|(n, \ell, \boldsymbol{\beta})| \le k$ and

(a)
$$\left|D^{(n,\ell,\boldsymbol{\beta})} b(x, \mu, v_1, \dots, v_n)\right| \le C, \qquad \left|D^{(n,\ell,\boldsymbol{\beta})} \sigma(x, \mu, v_1, \dots, v_n)\right| \le C, \tag{A.6}$$

(b)
$$\begin{aligned} &\left|D^{(n,\ell,\boldsymbol{\beta})} b(x, \mu, v_1, \dots, v_n) - D^{(n,\ell,\boldsymbol{\beta})} b(x', \mu', v_1', \dots, v_n')\right| \\ \le\ & C\left(|x - x'| + \sum_{i=1}^n |v_i - v_i'| + W_2(\mu, \mu')\right), \end{aligned}$$

[2] We do not consider ‘zeroth’ order derivatives in our definition, i.e. at least one of n, $\beta_1, \dots, \beta_n$ and ℓ must be non-zero, for every multi-index $(n, \ell, (\beta_1, \dots, \beta_n))$.

$$\Big|D^{(n,\ell,\boldsymbol{\beta})}\sigma(x,\mu,v_1,\dots,v_n) - D^{(n,\ell,\boldsymbol{\beta})}\sigma(x',\mu',v_1',\dots,v_n')\Big|$$
$$\leq \; C\bigg(|x-x'| + \sum_{i=1}^{n}|v_i - v_i'| + W_2(\mu,\mu')\bigg), \tag{A.7}$$

for any $x, x', v_1, v_1', \dots, v_n, v_n' \in \mathbb{R}^d$ and $\mu, \mu' \in \mathcal{P}_2(\mathbb{R}^d)$, for some constant $C > 0$.

(ii) Any function $\Phi : \mathcal{P}_2(\mathbb{R}^d) \to \mathbb{R}$ is said to be in $\mathcal{M}_k(\mathcal{P}_2(\mathbb{R}^d))$, if $D^{(n,\boldsymbol{\beta})}\Phi(\mu, v_1, \dots, v_n)$ exists for every multi-index $(n,\boldsymbol{\beta})$ such that $|(n,\boldsymbol{\beta})| \leq k$ and

(a)
$$\Big|D^{(n,\boldsymbol{\beta})}\Phi(\mu,v_1,\dots,v_n)\Big| \leq C, \tag{A.8}$$

(b)
$$\Big|D^{(n,\boldsymbol{\beta})}\Phi(\mu,v_1,\dots,v_n) - D^{(n,\boldsymbol{\beta})}\Phi(\mu',v_1',\dots,v_n')\Big|$$
$$\leq \; C\bigg(\sum_{i=1}^{n}|v_i - v_i'| + W_2(\mu,\mu')\bigg), \tag{A.9}$$

for any $v_1, v_1', \dots, v_n, v_n' \in \mathbb{R}^d$ and $\mu, \mu' \in \mathcal{P}_2(\mathbb{R}^d)$, for some constant $C > 0$.

(iii) A function $\mathcal{V} : [0,T] \times \mathcal{P}_2(\mathbb{R}^d) \to \mathbb{R}$ is said to be in $\mathcal{M}_k\big([0,T]\times\mathcal{P}_2(\mathbb{R}^d)\big)$, if $\mathcal{V}(\cdot,\mu)$ is in $C^1([0,T])$, for each $\mu \in \mathcal{P}_2(\mathbb{R}^d)$ and $\mathcal{V}(s,\cdot) \in \mathcal{M}_k\big(\mathcal{P}_2(\mathbb{R}^d)\big)$, for each $s \in [0,T]$, where the L^∞ and Lipschitz bounds of the derivatives of $\mathcal{V}(s,\cdot)$ are uniform in time, i.e. they only depend on T.

As for the first order case, we can establish the following relationship with linear functional derivatives, see e.g. [7] for the correspondence up to order 2,

$$\partial_\mu^n U(\cdot) = \partial_{y_n}\frac{\delta}{\delta m}\dots\partial_{y_1}\frac{\delta}{\delta m}U(\cdot) = \partial_{y_n}\dots\partial_{y_1}\frac{\delta^n}{\delta m^n}U(\cdot)\,, \tag{A.10}$$

provided one of the two derivatives is well-defined. The following proposition (Lemma 2.5 from [11]) relates regularity of L-derivatives with that of linear functional derivatives. We first define class $\mathcal{M}_k^L$ that characterises kth order linear functional derivatives.

Definition A.4 (Class $\mathcal{M}_k^L$ of kth order differentiable functions in linear functional derivatives)**.** A function $U : \mathcal{P}_2(\mathbb{R}^d) \to \mathbb{R}$ is said to be in class $\mathcal{M}_k^L(\mathcal{P}_2(\mathbb{R}^d))$ if it is k times differentiable in the sense of linear functional derivatives and satisfies

$$\left|\frac{\delta^k U}{\delta m^k}(m,y_1,\dots,y_k)\right| \leq C\big(|y_1|^k + \dots |y_k|^k\big), \tag{A.11}$$

for some constant $C > 0$ that does not depend on m and $y_1, \dots, y_k$.

Proposition A.5 (Lemma 2.5 from [11])**.** *Suppose that $U \in \mathcal{M}_k(\mathcal{P}_2(\mathbb{R}^d))$. Then*

$$\left|\frac{\delta^k U}{\delta m^k}(m,y_1,\dots,y_k)\right| \leq \frac{(\sqrt{d})^k}{k}\|\partial_\mu^k U\|_\infty\big(|y_1|^k + \dots |y_k|^k\big). \tag{A.12}$$

Consequently, $U \in \mathcal{M}_k^L(\mathcal{P}_2(\mathbb{R}^d))$.

B Appendix: Weak error analysis

In this section, we consider the following weak errors of the form

$$\left|\Phi(\mu_T^X) - \mathbb{E}[\Phi(\mu_T^N)]\right| \qquad \text{and} \qquad \left|\Phi(\mu_T^X) - \mathbb{E}[\Phi(\mu_T^{Z,N,h})]\right|,$$

for functionals $\Phi : \mathcal{P}_2(\mathbb{R}^d) \to \mathbb{R}$. The method of analysis follows from the work [11]. For any square-integrable random variable η, we define

$$X_t^{s,\eta} = \eta + \int_s^t b(X_r^{s,\eta}, \mathscr{L}(X_r^{s,\eta}))\, dr + \int_s^t \sigma(X_r^{s,\eta}, \mathscr{L}(X_r^{s,\eta}))\, dW_r, \quad t \in [s,T]. \tag{B.1}$$

A starting point of our investigation is the Feynman-Kac theorem for functionals of measures established in Theorem 7.2 of [5] (for the case $k = 2$). The generalisation to $k > 2$ is done in Theorem 2.15 of [11]. Note that the condition $\mathcal{M}_1(\mathbb{R}^d \times \mathcal{P}_2(\mathbb{R}^d))$ automatically implies (Lip).

Theorem B.1. *Let $k \geq 2$ be an integer. Suppose that $b, \sigma \in \mathcal{M}_k(\mathbb{R}^d \times \mathcal{P}_2(\mathbb{R}^d))$. We consider a function $\mathcal{V} : [0,T] \times \mathcal{P}_2(\mathbb{R}^d) \to \mathbb{R}$ defined by*

$$\mathcal{V}(s, \mathscr{L}(\eta)) = \Phi\big(\mathscr{L}(X_T^{s,\eta})\big), \tag{B.2}$$

for some function $\Phi : \mathcal{P}_2(\mathbb{R}^d) \to \mathbb{R}$ in $\mathcal{M}_k(\mathcal{P}_2(\mathbb{R}^d))$. Then $\mathcal{V} \in \mathcal{M}_k([0,T] \times \mathcal{P}_2(\mathbb{R}^d))$ and satisfies the PDE

$$\begin{cases} \partial_s \mathcal{V}(s,\mu) + \int_{\mathbb{R}^d} \left[\partial_\mu \mathcal{V}(s,\mu)(x) b(x,\mu) + \frac{1}{2} Tr\big(\partial_v \partial_\mu \mathcal{V}(s,\mu)(x) a(x,\mu)\big)\right] \mu(dx) = 0, \quad s \in (0,T), \\ \mathcal{V}(T,\mu) = \Phi(\mu), \end{cases} \tag{B.3}$$

where $a = (a_{i,k})_{1 \leq i,k \leq d} : \mathbb{R}^d \times \mathcal{P}_2(\mathbb{R}^d) \to \mathbb{R}^d \otimes \mathbb{R}^d$ denotes the diffusion operator

$$a_{i,k}(x,\mu) := \sum_{j=1}^m \sigma_{i,j}(x,\mu) \sigma_{k,j}(x,\mu), \qquad \forall x \in \mathbb{R}^d, \quad \forall \mu \in \mathcal{P}_2(\mathbb{R}^d).$$

We make the following observations before starting the main proof. The finite dimensional projection $V : [0,T] \times (\mathbb{R}^d)^N \to \mathbb{R}$ is defined by

$$V(s, x_1, \ldots, x_N) := \mathcal{V}\bigg(s, \frac{1}{N} \sum_{i=1}^N \delta_{x_i}\bigg). \tag{B.4}$$

Proposition 3.1 of [10] allows us to conclude that V is differentiable in the time component and twice-differentiable in the space components. Hence it is legitimate to apply the classical Itô's formula to V.

Next, by the flow property of (B.1) (see equation (3.5) in [5]), we observe that for any $s \in [0,T]$,

$$\mathcal{V}(s, \mathscr{L}(X_s^{0,\xi})) = \Phi\Big(\mathscr{L}(X_T^{s, X_s^{0,\xi}})\Big) = \Phi(\mathscr{L}(X_T^{0,\xi})).$$

Hence, this function is constant in time $s \in [0,T]$. In particular, by the terminal condition, we have

$$\Phi\big(\mu_T^X\big) = \Phi\big(\mathscr{L}(X_T^{0,\xi})\big) = \mathcal{V}(T, \mathscr{L}(X_T^{0,\xi})) = \mathcal{V}(0, \mathscr{L}(\xi)) = \mathcal{V}(0, \nu).$$

By the terminal condition for the PDE, we notice that

$$\Phi(\mu_T^N) = \mathcal{V}(T, \mu_T^N).$$

Therefore, the error between the particle system and the McKean-Vlasov limit decomposes as

$$\begin{aligned} \Phi(\mu_T^N) - \Phi(\mu_T^X) &= \mathcal{V}(T, \mu_T^N) - \mathcal{V}(0, \nu) \\ &= \big(\mathcal{V}(T, \mu_T^N) - \mathcal{V}(0, \mu_0^N)\big) + \big(\mathcal{V}(0, \mu_0^N) - \mathcal{V}(0, \nu)\big). \end{aligned} \tag{B.5}$$

This decomposition enables us to prove the following result.

Theorem B.2. *Suppose that $b, \sigma \in \mathcal{M}_2\big(\mathbb{R}^d \times \mathcal{P}_2(\mathbb{R}^d)\big)$ and $\Phi \in \mathcal{M}_2\big(\mathcal{P}_2(\mathbb{R}^d)\big)$. Then the weak error in the particle approximation satisfies*

$$\Big|\mathbb{E}[\Phi(\mu_T^N)] - \Phi(\mu_T^X)\Big| \leq \frac{C}{N}, \tag{B.6}$$

where C is a constant that depends on Φ, b, σ and T, but does not depend on N.

Proof. We first recall the definition of V defined in (B.4). By the assumptions on b and σ, the standard Itô's formula is applicable to V by Proposition 3.1 of [10]. Let $\mathbf{x} = (x_1, \ldots, x_N)$. Moreover, we know from this theorem that

$$\frac{\partial V}{\partial x_i}(s, \mathbf{x}) = \frac{1}{N} \partial_\mu \mathcal{V}\bigg(s, \frac{1}{N} \sum_{j=1}^N \delta_{x_j}\bigg)(x_i)$$

and

$$\frac{\partial^2 V}{\partial x_i^2}(s, \mathbf{x}) = \frac{1}{N} \partial_v \partial_\mu \mathcal{V}\bigg(s, \frac{1}{N} \sum_{j=1}^N \delta_{x_j}\bigg)(x_i) + \frac{1}{N^2} \partial_\mu^2 \mathcal{V}\bigg(s, \frac{1}{N} \sum_{j=1}^N \delta_{x_j}\bigg)(x_i, x_i),$$

for any $s \in [0,T]$, $x_1, \dots, x_N \in \mathbb{R}^d$. Let $\mathbf{Y}^N := (Y^{1,N}, \dots, Y^{N,N})$. Then

$$
\begin{aligned}
&\mathcal{V}(T,\mu_T^N) - \mathcal{V}(0,\mu_0^N) = V(T,\mathbf{Y}_T^N) - V(0,\mathbf{Y}_0^N) \\
&= \Big[\int_0^T \frac{\partial V}{\partial s}(s,\mathbf{Y}_s^N) + \sum_{i=1}^N \frac{\partial V}{\partial x_i}(s,\mathbf{Y}_s^N) b(Y_s^{i,N},\mu_s^N) + \frac{1}{2}\mathrm{Tr}\Big(a(Y_s^{i,N},\mu_s^N) \sum_{i=1}^N \frac{\partial^2 V}{\partial x_i^2}(s,\mathbf{Y}_s^N)\Big)\, ds \Big] \\
&+ \sum_{i=1}^N \int_0^T \sigma(Y_s^{i,N},\mu_s^N)^T \frac{\partial V}{\partial x_i}(s,\mathbf{Y}_s^N) \cdot dW_s^i \\
&= \int_0^T \partial_s \mathcal{V}(s,\mu_s^N) + \sum_{i=1}^N \Big[\frac{1}{N}\partial_\mu \mathcal{V}(s,\mu_s^N)(Y_s^{i,N}) b(Y_s^{i,N},\mu_s^N) \\
&+ \frac{1}{2}\mathrm{Tr}\Big(a(Y_s^{i,N},\mu_s^N)\Big(\frac{1}{N}\partial_v\partial_\mu \mathcal{V}(s,\mu_s^N)(Y_s^{i,N}) + \frac{1}{N^2}\partial_\mu^2 \mathcal{V}(s,\mu_s^N)(Y_s^{i,N},Y_s^{i,N})\Big)\Big)\Big]\, ds \\
&+ \frac{1}{N}\sum_{i=1}^N \int_0^T \sigma(Y_s^{i,N},\mu_s^N)^T \partial_\mu \mathcal{V}(s,\mu_s^N)(Y_s^{i,N}) \cdot dW_s^i.
\end{aligned}
$$

By (B.5) and PDE (B.3) evaluated at $(s,\mu_s^N)_{s\in[0,T]}$, the expression simplifies to

$$
\begin{aligned}
\Phi(\mu_T^N) - \Phi(\mu_T^X) \;=\; & \big(\mathcal{V}(0,\mu_0^N) - \mathcal{V}(0,\nu)\big) \\
& + \int_0^T \frac{1}{2}\Big[\frac{1}{N^2}\sum_{i=1}^N \mathrm{Tr}\Big(a(Y_s^{i,N},\mu_s^N)\partial_\mu^2 \mathcal{V}(s,\mu_s^N)(Y_s^{i,N},Y_s^{i,N})\Big)\Big]\, ds \\
& + \frac{1}{N}\sum_{i=1}^N \int_0^T \sigma(Y_s^{i,N},\mu_s^N)^T \partial_\mu \mathcal{V}(s,\mu_s^N)(Y_s^{i,N}) \cdot dW_s^i.
\end{aligned} \tag{B.7}
$$

It follows from Lemma 2.5 and Theorem 2.11 from [11] that

$$\big|\mathbb{E}\big(\mathcal{V}(0,\mu_0^N) - \mathcal{V}(0,\nu)\big)\big| \le \frac{C}{N}.$$

Taking expectation on both sides of (B.7) completes the proof. □

The next theorem concerns the weak error between (2.6) and (2.11).

Theorem B.3. *Suppose that $b \in \mathcal{M}_2\big(\mathbb{R}^d \times \mathcal{P}_2(\mathbb{R}^d)\big)$, $\Phi \in \mathcal{M}_2\big(\mathcal{P}_2(\mathbb{R}^d)\big)$ and that σ is constant. Then the weak error in the particle approximation with Euler scheme satisfies*

$$\Big|\mathbb{E}[\Phi(\mu_T^{Z,N,h})] - \Phi(\mu_T^X)\Big| \le C\Big(\frac{1}{N} + h\Big), \tag{B.8}$$

where C is a constant that depends on Φ, b, σ and T, but does not depend on N or h.

Proof. The main idea of the proof is identical to the previous theorem, with the extra complication of time discretisation. Let $\mathbf{Z}^{N,h} := (Z^{1,N,h}, \dots, Z^{N,N,h})$. As before, by Lemma 2.5 and Theorem 2.11 from [11],

$$\left|\mathbb{E}\big(\mathcal{V}(0, \mu_0^{Z,N,h}) - \mathcal{V}(0,\nu)\big)\right| \leq \frac{C}{N}.$$

Since σ is constant, we adopt the existing convention that the diffusion matrix a is defined by $a = \sigma\sigma^T$. Next, by the previous analysis, we observe that

$$\begin{aligned}
& \big(\Phi(\mu_T^{Z,N,h}) - \Phi(\mu_T^X)\big) - \big(\mathcal{V}(0, \mu_0^{Z,N,h}) - \mathcal{V}(0,\nu)\big) \\
= \; & \mathcal{V}(T, \mu_T^{Z,N,h}) - \mathcal{V}(0, \mu_0^{Z,N,h}) \\
= \; & V(T, \mathbf{Z}_T^{N,h}) - V(0, \mathbf{Z}_0^{N,h}) \\
= \; & \bigg[\int_0^T \frac{\partial V}{\partial s}(s, \mathbf{Z}_s^{N,h}) + \sum_{i=1}^N \frac{\partial V}{\partial x_i}(s, \mathbf{Z}_s^{N,h}) b\big(Z_{\eta(s)}^{i,N,h}, \mu_{\eta(s)}^{Z,N,h}\big) + \frac{1}{2}\mathrm{Tr}\bigg(a \sum_{i=1}^N \frac{\partial^2 V}{\partial x_i^2}(s, \mathbf{Z}_s^{N,h})\bigg)\, ds\bigg] \\
& + \int_0^T \sum_{i=1}^N \frac{\partial V}{\partial x_i}(s, \mathbf{Z}_s^{N,h})^T \sigma \, dW_s^i \\
= \; & \int_0^T \partial_s \mathcal{V}\big(s, \mu_s^{Z,N,h}\big) + \sum_{i=1}^N \bigg[\frac{1}{N}\partial_\mu \mathcal{V}\big(s, \mu_s^{Z,N,h}\big)(Z_s^{i,N,h}) b\big(Z_{\eta(s)}^{i,N,h}, \mu_{\eta(s)}^{Z,N,h}\big) \\
& + \frac{1}{2}\mathrm{Tr}\bigg(a\bigg(\frac{1}{N}\partial_v \partial_\mu \mathcal{V}\big(s, \mu_s^{Z,N,h}\big)(Z_s^{i,N,h}) + \frac{1}{N^2}\partial_\mu^2 \mathcal{V}\big(s, \mu_s^{Z,N,h}\big)(Z_s^{i,N,h}, Z_s^{i,N,h})\bigg)\bigg)\bigg]\, ds \\
& + \int_0^T \sum_{i=1}^N \frac{1}{N}\partial_\mu \mathcal{V}\big(s, \mu_s^{Z,N,h}\big)(Z_s^{i,N,h})^T \sigma \, dW_s^i \\
= \; & \int_0^T \sum_{i=1}^N \bigg[\frac{1}{N}\partial_\mu \mathcal{V}\big(s, \mu_s^{Z,N,h}\big)(Z_s^{i,N,h})\big(b\big(Z_{\eta(s)}^{i,N,h}, \mu_{\eta(s)}^{Z,N,h}\big) - b\big(Z_s^{i,N,h}, \mu_s^{Z,N,h}\big)\big) \\
& + \frac{1}{2}\mathrm{Tr}\bigg(a \frac{1}{N^2}\partial_\mu^2 \mathcal{V}\big(s, \mu_s^{Z,N,h}\big)(Z_s^{i,N,h}, Z_s^{i,N,h})\bigg)\bigg]\, ds + \int_0^T \sum_{i=1}^N \frac{1}{N}\partial_\mu \mathcal{V}\big(s, \mu_s^{Z,N,h}\big)(Z_s^{i,N,h})^T \sigma \, dW_s^i.
\end{aligned} \tag{B.9}$$

The second term of (B.9) clearly converges to zero in the rate $O(1/N)$ upon taking expectation. The third term of (B.9) becomes zero upon taking expectation. It remains to deal with the first term of (B.9). Let $\{\mathcal{F}_t\}_{t\in[0,T]}$ be the filtration generated by $W^1, \dots, W^N$. Then, by the Itô's formula, for each $k \in \{1, \dots, d\}$,

$$\begin{aligned}
& \mathbb{E}\Big[b_k\big(Z_{\eta(s)}^{i,N,h}, \mu_{\eta(s)}^{Z,N,h}\big) - b_k\big(Z_s^{i,N,h}, \mu_s^{Z,N,h}\big)\Big|\mathcal{F}_{\eta(s)}\Big] \\
= \; & -\mathbb{E}\bigg[\int_{\eta(s)}^s \bigg(\partial_x b_k(Z_r^{i,N,h}, \mu_r^{Z,N,h}) + \frac{1}{N}\partial_\mu b_k(Z_r^{i,N,h}, \mu_r^{Z,N,h})(Z_r^{i,N,h})\bigg) \cdot dZ_r^{i,N,h}
\end{aligned}$$

$$
\begin{aligned}
& +\sum_{j\neq i}\int_{\eta(s)}^{s}\frac{1}{N}\partial_\mu b_k\big(Z_r^{i,N,h},\mu_r^{Z,N,h}\big)(Z_r^{j,N,h})\cdot dZ_r^{j,N,h} \\
& +\int_{\eta(s)}^{s}\mathrm{Tr}\Bigg(\bigg(\partial_x^2 b_k(Z_r^{i,N,h},\mu_r^{Z,N,h})+\frac{2}{N}\partial_x\partial_\mu b_k(Z_r^{i,N,h},\mu_r^{Z,N,h})(Z_r^{i,N,h}) \\
& +\frac{1}{N}\partial_v\partial_\mu b_k(Z_r^{i,N,h},\mu_r^{Z,N,h})(Z_r^{i,N,h})+\frac{1}{N^2}\partial_\mu^2 b_k(Z_r^{i,N,h},\mu_r^{Z,N,h})(Z_r^{i,N,h},Z_r^{i,N,h})\bigg)\,d\left\langle Z^{i,N,h}\right\rangle_r\Bigg) \\
& +\sum_{j\neq i}\int_{\eta(s)}^{s}\mathrm{Tr}\Bigg(\bigg(\frac{1}{N}\partial_v\partial_\mu b_k(Z_r^{i,N,h},\mu_r^{Z,N,h})(Z_r^{j,N,h}) \\
& +\frac{1}{N^2}\partial_\mu^2 b_k(Z_r^{i,N,h},\mu_r^{Z,N,h})(Z_r^{j,N,h},Z_r^{j,N,h})\bigg)\,d\left\langle Z^{j,N,h}\right\rangle_r\Bigg)\bigg|\mathcal{F}_{\eta(s)}\bigg] \\
= \;& -\mathbb{E}\bigg[\int_{\eta(s)}^{s}\sum_{j=1}^{N}\frac{1}{N}\partial_\mu b_k\big(Z_r^{i,N,h},\mu_r^{Z,N,h}\big)(Z_r^{j,N,h})\,b(Z_{\eta(r)}^{j,N,h},\mu_{\eta(r)}^{Z,N,h})\,dr \\
& +\int_{\eta(s)}^{s}\partial_x b_k\big(Z_r^{i,N,h},\mu_r^{Z,N,h}\big)\,b(Z_{\eta(r)}^{i,N,h},\mu_{\eta(r)}^{Z,N,h})\,dr \\
& +\sum_{j=1}^{N}\int_{\eta(s)}^{s}\frac{1}{N}\partial_\mu b_k\big(Z_r^{i,N,h},\mu_r^{Z,N,h}\big)(Z_r^{j,N,h})^T\,\sigma\,dW_r^j+\frac{1}{N}\int_{\eta(s)}^{s}\partial_x b_k\big(Z_r^{i,N,h},\mu_r^{Z,N,h}\big)^T\,\sigma\,dW_r^i \\
& +\sum_{j=1}^{N}\int_{\eta(s)}^{s}\mathrm{Tr}\Bigg(a\bigg(\frac{1}{N}\partial_v\partial_\mu b_k(Z_r^{i,N,h},\mu_r^{Z,N,h})(Z_r^{j,N,h}) \\
& +\frac{1}{N^2}\partial_\mu^2 b_k(Z_r^{i,N,h},\mu_r^{Z,N,h})(Z_r^{j,N,h},Z_r^{j,N,h})\bigg)\Bigg)\,dr \\
& +\int_{\eta(s)}^{s}\mathrm{Tr}\Bigg(a\bigg(\partial_x^2 b_k(Z_r^{i,N,h},\mu_r^{Z,N,h})+\frac{2}{N}\partial_x\partial_\mu b_k(Z_r^{i,N,h},\mu_r^{Z,N,h})(Z_r^{i,N,h})\bigg)\Bigg)\,dr\bigg|\mathcal{F}_{\eta(s)}\bigg] \\
= \;& -\int_{\eta(s)}^{s}\mathbb{E}\bigg[\sum_{j=1}^{N}\frac{1}{N}\partial_\mu b_k\big(Z_r^{i,N,h},\mu_r^{Z,N,h}\big)(Z_r^{j,N,h})\,b(Z_{\eta(r)}^{j,N,h},\mu_{\eta(r)}^{Z,N,h}) \\
& +\partial_x b_k\big(Z_r^{i,N,h},\mu_r^{Z,N,h}\big)\,b(Z_{\eta(r)}^{i,N,h},\mu_{\eta(r)}^{Z,N,h}) \\
& +\sum_{j=1}^{N}\mathrm{Tr}\Bigg(a\bigg(\frac{1}{N}\partial_v\partial_\mu b_k(Z_r^{i,N,h},\mu_r^{Z,N,h})(Z_r^{j,N,h})+\frac{1}{N^2}\partial_\mu^2 b_k(Z_r^{i,N,h},\mu_r^{Z,N,h})(Z_r^{j,N,h},Z_r^{j,N,h})\bigg)\Bigg) \\
& +\mathrm{Tr}\Bigg(a\bigg(\partial_x^2 b_k(Z_r^{i,N,h},\mu_r^{Z,N,h})+\frac{2}{N}\partial_x\partial_\mu b_k(Z_r^{i,N,h},\mu_r^{Z,N,h})(Z_r^{i,N,h})\bigg)\Bigg)\bigg|\mathcal{F}_{\eta(s)}\bigg]\,dr.
\end{aligned}
\tag{B.10}
$$

Hence, upon taking expectation, by (B.10), the first term of (B.9) can be rewritten as

$$
\begin{aligned}
&\int_0^T \sum_{i=1}^N \mathbb{E}\Big[\frac{1}{N}\partial_\mu \mathcal{V}(s,\mu_s^{Z,N,h})(Z_s^{i,N,h})\big(b(Z_{\eta(s)}^{i,N,h},\mu_{\eta(s)}^{Z,N,h}) - b(Z_s^{i,N,h},\mu_s^{Z,N,h})\big)\Big]\, ds \\
= \quad & \int_0^T \frac{1}{N}\sum_{i=1}^N \sum_{k=1}^d \mathbb{E}\Big[\Big(\partial_\mu \mathcal{V}(s,\mu_s^{Z,N,h})(Z_s^{i,N,h})\Big)_k \mathbb{E}\Big[\big(b_k(Z_{\eta(s)}^{i,N,h},\mu_{\eta(s)}^{Z,N,h}) - b_k(Z_s^{i,N,h},\mu_s^{Z,N,h})\big)\Big|\mathcal{F}_{\eta(s)}\Big]\Big]\, ds \\
= \quad & -\int_0^T \int_{\eta(s)}^s \frac{1}{N}\sum_{i=1}^N \sum_{k=1}^d \mathbb{E}\Big[\Big(\partial_\mu \mathcal{V}(s,\mu_s^{Z,N,h})(Z_s^{i,N,h})\Big)_k \times \\
&\Big[\sum_{j=1}^N \frac{1}{N}\partial_\mu b_k(Z_r^{i,N,h},\mu_r^{Z,N,h})(Z_r^{j,N,h})\, b(Z_{\eta(r)}^{j,N,h},\mu_{\eta(r)}^{Z,N,h}) + \partial_x b_k(Z_r^{i,N,h},\mu_r^{Z,N,h})\, b(Z_{\eta(r)}^{i,N,h},\mu_{\eta(r)}^{Z,N,h}) \\
&+\sum_{j=1}^N \mathrm{Tr}\Big(a\Big(\frac{1}{N}\partial_v\partial_\mu b_k(Z_r^{i,N,h},\mu_r^{Z,N,h})(Z_r^{j,N,h}) + \frac{1}{N^2}\partial_\mu^2 b_k(Z_r^{i,N,h},\mu_r^{Z,N,h})(Z_r^{j,N,h},Z_r^{j,N,h})\Big)\Big) \\
&+\mathrm{Tr}\Big(a\Big(\partial_x^2 b_k(Z_r^{i,N,h},\mu_r^{Z,N,h}) + \frac{2}{N}\partial_x\partial_\mu b_k(Z_r^{i,N,h},\mu_r^{Z,N,h})(Z_r^{i,N,h})\Big)\Big)\Big]\Big]\, dr\, ds.
\end{aligned}
$$

Finally, by (6.2) and the fact that $\mathcal{V} \in \mathcal{M}_2([0,T]\times\mathcal{P}_2(\mathbb{R}^d))$, we have

$$
\Big|\int_0^T \sum_{i=1}^N \mathbb{E}\Big[\frac{1}{N}\partial_\mu \mathcal{V}(s,\mu_s^{Z,N,h})(Z_s^{i,N,h})\big(b(Z_{\eta(s)}^{i,N,h},\mu_{\eta(s)}^{Z,N,h}) - b(Z_s^{i,N,h},\mu_s^{Z,N,h})\big)\Big]\, ds\Big| \le Ch.
$$

□

References

[1] Luigi Ambrosio, Federico Stra, and Dario Trevisan. A PDE approach to a 2-dimensional matching problem. *Probability Theory and Related Fields*, pages 1–45, 2018.

[2] Mireille Bossy, Jean-François Jabir, and Denis Talay. On conditional McKean Lagrangian stochastic models. *Probability theory and related fields*, 151(1-2):319–351, 2011.

[3] Mireille Bossy and Denis 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):818–861, 1996.

[4] Mireille Bossy and Denis Talay. A stochastic particle method for the McKean-Vlasov and the Burgers equation. *Mathematics of Computation of the American Mathematical Society*, 66(217):157–192, 1997.

[5] Rainer Buckdahn, Juan Li, Shige Peng, and Catherine Rainer. Mean-field stochastic differential equations and associated PDEs. *The Annals of Probability*, 45(2):824–878, 2017.

[6] Pierre Cardaliaguet. Notes on mean field games. Technical report, Technical report, 2010.

[7] Pierre Cardaliaguet, François Delarue, Jean-Michel Lasry, and Pierre-Louis Lions. The master equation and the convergence problem in mean field games. *arXiv preprint arXiv:1509.02505*, 2015.

[8] René Carmona. *Lectures on BSDEs, stochastic control, and stochastic differential games with financial applications*, volume 1. SIAM, 2016.

[9] Rene Carmona and Francois Delarue. *Probabilistic theory of mean field games with applications I: Mean Field FBSDEs, Control, and Games*. Springer, 2017.

[10] Jean-François Chassagneux, Dan Crisan, and François Delarue. A probabilistic approach to classical solutions of the master equation for large population equilibria. *arXiv preprint arXiv:1411.3009*, 2014.

[11] Jean-François Chassagneux, Lukasz Szpruch, and Alvin Tse. Weak quantitative propagation of chaos via differential calculus on the space of measures. *arXiv preprint arXiv:1901.02556*, 2019.

[12] K Andrew Cliffe, Mike B Giles, Robert Scheichl, and Aretha L Teckentrup. Multilevel Monte Carlo methods and applications to elliptic PDEs with random coefficients. *Computing and Visualization in Science*, 14(1):3, 2011.

[13] Dan Crisan and Eamon McMurray. Smoothing properties of McKean–Vlasov SDEs. *Probability Theory and Related Fields*, pages 1–52, 2017.

[14] François Delarue, Daniel Lacker, Kavita Ramanan, et al. From the master equation to mean field game limit theory: A central limit theorem. *Electronic Journal of Probability*, 24, 2019.

[15] Steffen Dereich, Michael Scheutzow, and Reik Schottstedt. Constructive quantization: Approximation by empirical measures. In *Annales de l'IHP Probabilités et statistiques*, pages 1183–1203, 2013.

[16] Nicolas Fournier and Arnaud Guillin. On the rate of convergence in Wasserstein distance of the empirical measure. *Probability Theory and Related Fields*, 162(3-4):707–738, 2015.

[17] Nicolas Fournier and Maxime Hauray. Propagation of chaos for the Landau equation with moderately soft potentials. *The Annals of Probability*, 44(6):3581–3660, 2016.

[18] Jürgen Gärtner. On the Mckean-Vlasov limit for interacting diffusions. *Mathematische Nachrichten*, 137(1):197–248, 1988.

[19] Michael B Giles. Multilevel Monte Carlo path simulation. *Operations Research*, 56(3):607–617, 2008.

[20] Michael B Giles. Multilevel Monte Carlo methods. *Acta Numerica*, 24:259–328, 2015.

[21] Michael B Giles and Takashi Goda. Decision-making under uncertainty: using MLMC for efficient estimation of EVPPI. *Statistics and Computing*, pages 1–13, 2018.

[22] Michael B Giles and Abdul-Lateef Haji-Ali. Multilevel nested simulation for efficient risk estimation. *arXiv preprint arXiv:1802.05016*, 2018.

[23] Michael Goldman, Martin Huesmann, and Felix Otto. A large-scale regularity theory for the Monge-Ampère equation with rough data and application to the optimal matching problem. *arXiv preprint arXiv:1808.09250*, 2018.

[24] Abdul-Lateef Haji-Ali and Raúl Tempone. Multilevel and Multi-index Monte Carlo methods for the Mckean–Vlasov equation. *Statistics and Computing*, 28(4):923–935, 2018.

[25] Stefan Heinrich. Multilevel Monte Carlo methods. In *Large-scale scientific computing*, pages 58–67. Springer, 2001.

[26] Pierre-Emmanuel Jabin and Zhenfu Wang. Quantitative estimates of propagation of chaos for stochastic systems with $W^{-1,\infty}$ kernels. *Inventiones mathematicae*, 214(1):523–591, 2018.

[27] Jean-Francois Jabir. Rate of propagation of chaos for diffusive stochastic particle systems via Girsanov transformation. *arXiv preprint arXiv:1907.09096*, 2019.

[28] Benjamin Jourdain, Sylvie Méléard, and Wojbor Woyczynski. Nonlinear SDEs driven by Lévy processes and related PDEs. *arXiv preprint arXiv:0707.2723*, 2007.

[29] Ahmed Kebaier. Statistical Romberg extrapolation: a new variance reduction method and applications to option pricing. *The Annals of Applied Probability*, 15(4):2681–2705, 2005.

[30] Vassili N Kolokoltsov. *Nonlinear Markov processes and kinetic equations*, volume 182. Cambridge University Press, 2010.

[31] Daniel Lacker. On a strong form of propagation of chaos for mckean-vlasov equations. *Electronic Communications in Probability*, 23, 2018.

[32] Vincent Lemaire and Gilles Pagès. Multilevel Richardson–Romberg extrapolation. *Bernoulli*, 23(4A):2643–2692, 2017.

[33] PL Lions. Cours au Collège de France: Théorie des jeux à champs moyens, 2014.

[34] Sylvie Méléard. Asymptotic behaviour of some interacting particle systems; McKean-Vlasov and Boltzmann models. In *Probabilistic models for nonlinear partial differential equations*, pages 42–95. Springer, 1996.

[35] Stéphane Mischler and Clément Mouhot. Kac's program in kinetic theory. *Inventiones mathematicae*, 193(1):1–147, 2013.

[36] Stéphane Mischler, Clément Mouhot, and Bernt Wennberg. A new approach to quantitative propagation of chaos for drift, diffusion and jump processes. *Probability Theory and Related Fields*, 161(1-2):1–59, 2015.

[37] Svetlozar T Rachev and Ludger Rüschendorf. *Mass Transportation Problems: Volume II: Applications*, volume 2. Springer Science & Business Media, 1998.

[38] Alain-Sol Sznitman. *Topics in propagation of chaos*. Springer, 1991.

[39] Michel Talagrand. Scaling and non-standard matching theorems. *Comptes Rendus Mathematique*, 356(6):692–695, 2018.