Antithetic multilevel particle system sampling method for McKean-Vlasov SDEs

Łukasz Szpruch$^1$ and Alvin Tse$^1$

$^1$School of Mathematics, University of Edinburgh

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 $\Phi : \mathcal{P}_2(\mathbb{R}^d) \rightarrow \mathbb{R}$. 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 law $\mu_N = \frac{1}{N} \sum_{i=1}^{N} \delta_{X_i}$, where a) $(X_i)_{i=1}^{N}$ is a sequence of i.i.d samples from $\mu$ 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 $\mu_N$ to its limit $\mu$ for linear functionals of measure (i.e. $F(\mu) = \int_{\mathbb{R}^d} f(x)\mu(dx)$ for some function $f : \mathbb{R}^d \rightarrow \mathbb{R}$) 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 provides sharp error bounds. However, for general non-linear functionals of measure $\Phi : \mathcal{P}_2(\mathbb{R}^d) \rightarrow \mathbb{R}$, $\Phi(\mu_N)$ 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 $F(\mu) = R(\int_{\mathbb{R}^d} f(x)\mu(dx))$, with $R : \mathbb{R} \rightarrow \mathbb{R}$ being nonlinear, the multilevel Monte-Carlo (MLMC) [30, 20] and antithetic multilevel Monte-Carlo (A-MLMC) [21] estimators are more efficient than $F(\mu_N)$. In this work, we study the general case of functionals of measure, which are sufficiently 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 (i.e matching
i.i.d case and linear functions of measures) estimates of mean-square error\(^1\) \(\mathbb{E}[(\Phi(\mu) - \mathcal{A})^2]\). As already mentioned, 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 \(\mathcal{A}\)). Fix \(L \in \mathbb{N}_+\) and sequences \(\{N_\ell\}_{\ell=0}^L\) and \(\{M_\ell\}_{\ell=0}^L\) of non-decreasing and non-increasing natural numbers, respectively. The classical MLMC estimator is given by
\[
\mathcal{A}_{\text{MLMC}} := \frac{1}{M_0} \sum_{\theta=1}^{M_0} \Phi(\mu_{N_0,(\theta),(0)}) + \sum_{\ell=1}^L \left[ \frac{1}{M_\ell} \sum_{\theta=1}^{M_\ell} \left[ \Phi(\mu_{N_\ell,(\theta),(\ell)}) - \Phi(\mu_{N_{\ell-1},(\theta),(\ell)}) \right] \right],
\]
where \(\mu_{N_\ell,(\theta),(\ell)}\) is the empirical measure corresponding to each of the \(\sum_{\ell=0}^L M_\ell\) independent clouds of particles indexed by \(\ell \in \{0, \ldots, L\}\) and \(\theta \in \{1, \ldots, M_\ell\}\). In essence, MLMC breaks down the simulation of \(\mathbb{E}[\Phi(\mu_{N_L})]\) into a sequence of approximations of \(\mathbb{E}[\Phi(\mu_{N_\ell})]\), \(\ell = 0, \ldots, L\), with increasing accuracy, but also with increasing cost. If the variance between successive approximations converges to zero as the level increases, then 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 on high levels with low accuracy (and at a high cost). The idea has been independently developed by Giles and Heinrich [19, 25, 27] (see also 2-level Monte-Carlo of Kebaier [27]) in the context of temporal approximation of SDEs and parametric integration.

The second estimator that we consider in this paper is A-MLMC \(^2\)
\[
\mathcal{A}_{\text{A-MLMC}} := \frac{1}{M_0} \sum_{\theta=1}^{M_0} \Phi(\mu_{N_0,(\theta),(0)}) + \sum_{\ell=1}^L \left[ \frac{1}{M_\ell} \sum_{\theta=1}^{M_\ell} \left[ \Phi(\mu_{N_\ell,(\theta),(\ell)}) - \frac{1}{2} \left( \Phi(\mu_{N_\ell,(1),(\ell)}) + \Phi(\mu_{N_\ell,(2),(\ell)}) \right) \right] \right],
\]
where \(\mu_{N_\ell,(\theta),(\ell)}\) is defined as before, but we introduced an antithetic pair \((\mu_{N_\ell,(1),(\theta),(\ell)}, \mu_{N_\ell,(2),(\theta),(\ell)})\) with the property that \(\mathbb{E}[\Phi(\mu_{N_\ell,(1),(\theta),(\ell)})] = \mathbb{E}[\Phi(\mu_{N_\ell,(2),(\theta),(\ell)})] = \mathbb{E}[\Phi(\mu_{N_{\ell-1},(\theta),(\ell)})]\). This property ensures that \(\mathbb{E}[\mathcal{A}_{\text{MLMC}}] = \mathbb{E}[\mathcal{A}_{\text{A-MLMC}}]\), but as we demonstrate in this paper, \(\mathcal{A}_{\text{A-MLMC}}\) is more efficient by an order of magnitude. To see why that might be the case, we consider the following simple example.

**Example 1.1.** Consider \(\Phi(\mu) := \int_{\mathbb{R}^d} F(x) \mu(dx)\), where \(F : \mathbb{R}^d \to \mathbb{R}\) has linear growth. Let \(\mu_N := N^{-1} \sum_{i=1}^N \delta_{X_i}\) be the empirical law of \(N\) independent samples \(\{X_i\}_{i=1}^N\) from \(\mu \in \mathcal{P}_2(\mathbb{R}^d)\). We already observed that \(\mathbb{E}[\mathcal{A}_{\text{MLMC}}] = \mathbb{E}[\mathcal{A}_{\text{A-MLMC}}]\). The postulated independence conditions imply that
\[
\text{Var}[\mathcal{A}_{\text{MLMC}}] = \frac{\text{Var}[\Phi(\mu_{N_0})]}{M_0} + \sum_{\ell=1}^L \frac{\text{Var}[\Phi(\mu_{N_\ell}) - \Phi(\mu_{N_{\ell-1}})]}{M_\ell},
\]
\(^1\)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.
\(^2\)In subsequent sections, we denote \(\mathcal{A}_{\text{A-MLMC}}\) to be the A-MLMC estimator without time discretisation and \(\mathcal{A}_{\text{A-MLMC}_t}\) to be the A-MLMC estimator with Euler time discretisation.
On the other hand, 
\[
\text{Var}[\mathcal{A}^{\text{A-MLMC}}] = \frac{\text{Var}[\Phi(\mu_{N_0})]}{M_0} + \sum_{\ell=1}^L \frac{\text{Var}[\Phi(\mu_{N_\ell}) - \frac{1}{2}(\Phi(\mu_{N_{\ell}(1)}) + \Phi(\mu_{N_{\ell}(2)}))]}{M_\ell}.
\]

It is clear that the efficiency of this algorithm hinges on good coupling estimates that result in small variances across levels \( \ell \). Set \( N_\ell := 2N_{\ell-1} \). For \( \mathcal{A}^{\text{MLMC}} \), we have
\[
\text{Var}[\Phi(\mu_{N_\ell}) - \Phi(\mu_{N_{\ell-1}})] = \text{Var}\left[\left(\frac{1}{N_\ell} - \frac{1}{N_{\ell-1}}\right) \sum_{i=1}^{N_{\ell-1}} F(X_i) + \frac{1}{N_\ell} \sum_{i=N_{\ell-1}+1}^{N_\ell} F(X_i)\right] = \left(\frac{1}{N_\ell} - \frac{1}{N_{\ell-1}}\right)^2 \sum_{i=1}^{N_{\ell-1}} \text{Var}[F(X_i)] + \left(\frac{1}{N_\ell}\right)^2 \sum_{i=N_{\ell-1}+1}^{N_\ell} \text{Var}[F(X_i)] = O\left(\frac{1}{N_\ell}\right).
\]

On the other hand, for A-MLMC, we take \( \{X_i\}_{i=1}^{N_{\ell}} \cup \{X_i\}_{i=N_{\ell-1}+1}^{N_\ell} = \{X_i\}_{i=1}^{N_\ell} \) and construct corresponding empirical measures
\[
\mu_{N_\ell} := N_{\ell}^{-1} \sum_{i=1}^{N_\ell} \delta_{X_i}, \quad \mu_{N_{\ell}(1)} := N_{\ell-1}^{-1} \sum_{i=1}^{N_{\ell-1}} \delta_{X_i}, \quad \mu_{N_{\ell}(2)} := N_{\ell-1}^{-1} \sum_{i=N_{\ell-1}+1}^{N_\ell} \delta_{X_i}.
\]

Therefore, the variance of the antithetic difference is reduced to
\[
\text{Var}\left[\Phi(\mu_{N_\ell}) - \frac{1}{2}(\Phi(\mu_{N_{\ell}(1)}) + \Phi(\mu_{N_{\ell}(2)}))\right] = 0.
\]

The above example is indeed a very special case. This work explores regularity conditions of functionals \( \Phi \) 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 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 L-derivatives that will be used in other theorems.) Theorem 2.5 shows that if \( \mu \) has finite eighth moment and \( \Phi \in \mathcal{M}_k^L \), then
\[
\text{Var}\left[\Phi(\mu_{N_\ell}) - \frac{1}{2}(\Phi(\mu_{N_{\ell}(1)}) + \Phi(\mu_{N_{\ell}(2)}))\right] = O\left(\frac{1}{N_\ell^2}\right).
\] 
(1.1)

By Theorem 2.11 in [11], we also have
\[
|\mathbb{E}[\Phi(\mu_{N_\ell})] - \Phi(\mu)| \leq O\left(\frac{1}{N_\ell}\right).
\] 
(1.2)

Finally, since the empirical measures \( \mu_{N_\ell}, \mu_{N_{\ell}(1)} \) and \( \mu_{N_{\ell}(2)} \) correspond to i.i.d. random variables, the cost of simulating the antithetic difference is given by
\[
\text{Cost}\left[\Phi(\mu_{N_\ell}) - \frac{1}{2}(\Phi(\mu_{N_{\ell}(1)}) + \Phi(\mu_{N_{\ell}(2)}))\right] = O(N_\ell).
\] 
(1.3)
Hence, by combining (1.1), (1.2) and (1.3), the well-known result of M. Giles in Theorem 1 of [12] concludes that the complexity corresponding to the estimator $A^{A-MLMC}$ is reduced to $O(e^{-2})$ for a mean-square error of $O(\varepsilon^2)$.

We stress that our bound (1.1) is dimension independent, which is not common in the literature. For example, if we only assume that $\Phi$ 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, 37, 23] that study the problem from the perspective of Monge-Ampère PDEs. On the other hand, recently, authors [14, Lem. 5.10] observed that if the functional $\Phi$ 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).

1.1 A-MLMC for Interacting Diffusions

The second situation we treat in this work concerns estimates of propagation-of-chaos type for the system of 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 $b : \mathbb{R}^d \times \mathcal{P}_2(\mathbb{R}^d) \rightarrow \mathbb{R}^d$, $\sigma : \mathbb{R}^d \times \mathcal{P}_2(\mathbb{R}^d) \rightarrow \mathbb{R}^d \otimes \mathbb{R}^d$, and consider the corresponding McKV-SDE given by

$$
\begin{aligned}
\text{d}X_t &= \xi + \int_0^t b(X_s, \mu_s^X) \text{d}s + \int_0^t \sigma(X_s, \mu_s^X) \text{d}W_s, \quad t \in [0, T], \\
\mu_s^X &= \text{Law}(X_s),
\end{aligned}
$$

where $\xi \sim \nu \in \mathcal{P}_2(\mathbb{R}^d)$. Note that $\{X_t\}_{t \in [0, T]}$ is not necessarily a Markov process. Nonetheless using Itô’s formula with $P \in C_b^0(\mathbb{R}^d)$, one can derive corresponding nonlinear Kolmogorov-Fokker-Planck equation

$$
\partial_t (\mu_t, P) = \langle \mu_t, \frac{1}{2} \sum_{i,j=1}^d \partial^2_{x_i,x_j} P(.) \sigma T_{ij}(\cdot, \mu_t) + \sum_{i=1}^d \partial_x P(.) b_i(\cdot, \mu_t) \rangle,
$$

where $\langle m, F \rangle := \int_{\mathbb{R}^d} F(y) m(dy)$. The theory of propagation of chaos, [36], shows that (1.4) arises as a limiting equation of the system of interacting diffusions (particles) $\{Y_{t,i}^{i,N}\}_{i=1,...,N}$ on $(\mathbb{R}^d)^N$ given by

$$
\begin{aligned}
Y_{t,i}^{i,N} &= \xi_i + \int_0^t b(Y_{s,i}^{i,N}, \mu_s^{Y_{i,N}}) \text{d}s + \int_0^t \sigma(Y_{s,i}^{i,N}, \mu_s^{Y_{i,N}}) \text{d}W_s^i, & 1 \leq i \leq N, \quad t \in [0, T], \\
\dot{\mu}_s^{Y_{i,N}} := \frac{1}{N} \sum_{i=1}^N \delta_{Y_{s}^{i,N}},
\end{aligned}
$$

where $W^i, 1 \leq i \leq N$, are independent $d$-dimensional Brownian motions and $\xi_i, 1 \leq i \leq N$, are i.i.d. random variables with law $\nu \in \mathcal{P}_2(\mathbb{R}^d)$. We refer the reader to [18, 36, 32] for the classical
results in this direction and to [26, 2, 17, 33, 29] 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 (1.4) are linear in measure and globally Lipschitz continuous, [36] showed that $W_2(\mathcal{L}(Y^{i,N}_t),\mathcal{L}(X_t)) = O(N^{-1/2})$. 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 $\sigma$ in terms of L-derivatives, we have a strong error bound in fourth moment that is dimension-independent. (See Theorem 2.4.)

To lift the idea of A-MLMC from the i.i.d. setting to interacting diffusions, for each $\ell$, again assuming that $2N_{\ell-1} = N_\ell$, we take $\{\xi^i, W^i\}_{i=1}^{N_\ell-1} \cup \{\tilde{\xi}^i, \tilde{W}^i\}_{i=N_\ell-1+1} = \{\xi^i, W^i\}_{i=1}^{N_\ell}$ and build a particle system with $N_\ell$ particles and two corresponding sub-particle systems with $N_{\ell-1}$ particles each. We remark that idea of antithetic MLMC is not new. A variant of the method was developed and analysed in [22] to avoid the problem of simulating Lévy areas for the Milstein scheme for the approximation of SDEs. An encouraging numerical study of A-MLMC in the context of McKean-Vlasov SDEs recently appeared in [24]. In our case, the main challenge is to show that we can get a ‘good’ estimate on the variance for each $\ell$ of the A-MLMC estimator. Unlike the i.i.d. setting considered above, these particles are not independent. Our analysis relies heavily on the calculus on $\mathcal{P}_d[0,T]$ for more recent theory. Most of the results in [5] and [10]. This line of research has been recently explored in [28, Ch. 9] and [34, Th. 2.1] to obtain results of quantitative propagation of chaos for a general family of particle systems. A similar research programme, but in the context of mean-field games with a common noise, has been successfully undertaken in [7].

In the case of McKV-SDEs, the mean-square error corresponding to estimator $\mathcal{A}$ is given by $\mathbb{E}[(\Phi(\mu_T^Y) - \mathcal{A})^2]$. We now recall some of the observations from [11]. To achieve a mean-square error of $O(\epsilon^2)$ in the approximation, by standard Monte-Carlo, the number of interactions is of the order $O(\epsilon^{-4})$.

Unlike the i.i.d. case, the number of generated random processes is not a good proxy for the actual cost of the estimator, as particles are interacting and hence the cost of simulating an $N$-particle system is $N^2$. We introduce the estimator

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

corresponding to an ensemble of particle systems, where $\mu_T^{N,\theta}$ denotes the empirical measure of the particles obtained for each i.i.d. sample $\theta \in \{1, \ldots, M\}$.

By introducing ensembles of particles, the number of interactions is of the order $O(\epsilon^{-3})$. Then, by introducing Romberg extrapolation to the ensembles of particles, the number of interactions can be reduced to the order $O(\epsilon^{-2-1/k})$, [11, Sec 1.1 and Th 2.17], under the assumption that $b$, $\sigma$ and
the test function $\Phi$ are in $\mathcal{M}_{2k+1}$. The A-MLMC estimator $\mathcal{A}^{A}$-MLMC, achieves (almost) optimal order of interactions, whilst only requiring $b, \sigma$ and $\Phi$ are in $\mathcal{M}_4$. The table below provides detailed comparison among all just described methods.

Number of interactions (without time discretisation)

\[
\begin{align*}
O(\epsilon^{-3}), & \quad \text{for ensembles of particles}, \\
O(\epsilon^{-2-1/k}), & \quad \text{with Romberg extrapolation, for } b, \sigma \text{ and } \Phi \text{ in } \mathcal{M}_{2k+1}, \\
O(\epsilon^{-2} (\log(\epsilon))^2), & \quad \text{with Antithetic MLMC estimator } \mathcal{A}^{A}$-MLMC, for $b, \sigma$ and $\Phi$ in $\mathcal{M}_4.
\end{align*}
\] (1.6)

Finally, to obtain a fully implementable algorithm, one needs to study time discretisation of (1.4). We work with an Euler scheme as $[3, 4]$. 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

\[
\begin{align*}
Z_i^{i,N,h} &= \xi_i + \int_0^t b(Z_{\eta(r)}, \mu_{\eta(r)}) \, dr + \int_0^t \sigma(Z_{\eta(r)}, \mu_{\eta(r)}) \, dW_r, \\
\mu_{z_{N,h}} &= \frac{1}{N} \sum_{i=1}^N \delta_{Z_i^{i,N,h}}.
\end{align*}
\] (1.7)

Similar to above, we work with the estimator

\[
Q_{M,N,h} := \frac{1}{M} \sum_{\theta=1}^M \Phi(\mu_{T}^{Z_{N,h}(\theta)}).
\]

Note that we can write

\[
\Phi(\mu_{T}^{Z_{N,h}(\theta)}) - \Phi(\mu_{T}^{X}) = \left( \Phi(\mu_{T}^{Y_{N}(\theta)}) - \Phi(\mu_{T}^{X}) \right) + \left( \Phi(\mu_{T}^{Z_{N,h}(\theta)}) - \Phi(\mu_{T}^{Y_{N}(\theta)}) \right).
\]

Therefore, the additional step in the analysis involving time-discretisation relies on controlling the discretisation error between $\Phi(\mu_{T}^{Z_{N,h}(\theta)})$ and $\Phi(\mu_{T}^{Y_{N}(\theta)})$. This type of analysis is performed in Lemma 4.1 and Theorem B.3. The key challenge is to obtain estimates of such discretisation errors (both strong and weak) that are uniform in $N$. It is then straightforward to observe from Lemma 4.1 and Theorem B.3 that the number of interactions for achieving a mean-square-error of order $O(\epsilon^2)$ using the direct approach of Monte-Carlo simulation by ensembles of particles is $O(\epsilon^{-4})$. As in [11], this analysis with time discretisation can be done with Romberg extrapolation, for which the number of interactions becomes $O(\epsilon^{-3-1/k})$, if $b, \sigma$ and $\Phi$ are in $\mathcal{M}_{2k+1}$. Finally, Theorem 4.3 in Section 4 proves that, by using an Euler time-discretisation, the number of interactions upon applying antithetic MLMC is $O(\epsilon^{-3})$, if $b, \sigma$ and $\Phi$ are in $\mathcal{M}_4$. As a summary,

Number of interactions (with time discretisation)

\[
\begin{align*}
O(\epsilon^{-4}), & \quad \text{for ensembles of particles}, \\
O(\epsilon^{-3-1/k}), & \quad \text{with Romberg extrapolation, for } b, \sigma \text{ and } \Phi \text{ in } \mathcal{M}_{2k+1}, \\
O(\epsilon^{-3}), & \quad \text{with Antithetic MLMC estimator } \mathcal{A}^{A}$-MLMC, for $b, \sigma$ and $\Phi$ in $\mathcal{M}_4.
\end{align*}
\] (1.8)
Here is an outline of the main results of the article. Firstly, Theorem 2.4 proves a dimension-independent rate of uniform strong propagation of chaos for sufficiently smooth drift and diffusion functions. This is a considerable generalisation from [36], which assumes the drift and diffusion functions to be linear in measure. Secondly, Theorem 2.5 generalises the result in [19] (Section 9) from functionals in measure of the form (2.21) to general functionals in measure. As for the antithetic MLMC algorithm, Theorem 3.2 in Section 3 proves that, if it is possible to simulate (1.5) directly without time discretisation, then the computational complexity upon applying antithetic MLMC can be improved to $O(\epsilon^{-2}(\log \epsilon)^2)$. Finally, Theorem 4.3 in Section 4 proves that, by using an Euler time-discretisation, the computational complexity upon applying antithetic MLMC is $O(\epsilon^{-3})$, which is still a considerable improvement compared to direct Monte-Carlo simulation.

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, $L^2(\xi)$ denotes the law of $\xi$, for any square-integrable random variable $\xi$. 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 $\epsilon$. Finally, unless otherwise specified, $C$ denotes a generic constant that does not depend on $N, h$ or $\epsilon$, whose value may vary from line to line.

Since this work relies heavily on the theory of differentiation in measure developed by P. Lions in his course at Collège de France [31], the reader is directed to Appendices A and B for further details.

2 Dimension-independent rate of uniform strong propagation of chaos and $L^2$ estimate of antithetic difference for i.i.d. random variables

We begin this section with the following lemma on the $W_2$ metric.

Lemma 2.1. Let $\eta \in \mathbb{R}^d$ and $m \in \mathcal{P}_2(\mathbb{R}^d)$. Then

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

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,

$$W_2\left(\frac{1}{N} \delta_\eta + \frac{N-1}{N} m, m\right) \leq \mathbb{E}[|X - Y|^2]$$

$$= \mathbb{E}[|X - Y|^2|\Omega'] \mathbb{P}(\Omega') + \mathbb{E}[|X - Y|^2|\Omega']^c \mathbb{P}((\Omega')^c)$$

$$= \frac{1}{N} \mathbb{E}[|\eta - Y|^2]$$

7
\[
\leq \frac{2}{N}(|\eta|^2 + \mathbb{E}[|Y|^2]).
\]

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 more general, 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]. 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 2.3. The technique of the following proof is also adopted in the proof of Theorem 2.5.

**Lemma 2.2.** 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

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

**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

\[
U(m^N) - U(m_0) = \int_0^1 \int_{\mathbb{R}^d} \frac{\delta U}{\delta m}(\lambda m^N + (1 - \lambda)m_0, v) (m^N - m_0)(dv) d\lambda
\]

\[
= \frac{1}{N} \sum_{i=1}^N \int_0^1 \varphi^i_{\lambda} d\lambda,
\]

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

\[
\varphi^i_{\lambda} = \frac{\delta U}{\delta m}(\lambda m^N + (1 - \lambda)m_0, \zeta_i) - \mathbb{E}\left[\frac{\delta U}{\delta m}(\lambda m^N + (1 - \lambda)m_0, \zeta_i)\right].
\] (2.1)

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

\[
\mathbb{E}[(\varphi^i_{\lambda})^4 + (\varphi^j_{\lambda})^2(\varphi^j_{\lambda})^2 + \varphi^i_{\lambda}(\varphi^j_{\lambda})^3] \leq C\left\| \frac{\partial U}{\partial \mu} \right\|_{\infty}^4 \mathbb{E}[|\zeta_i|^4].
\] (2.2)

We have the estimate

\[
\mathbb{E}[|U(m^N) - U(m_0)|^4] \leq \frac{1}{N^4} \int_0^1 \mathbb{E}\left[\left( \sum_{i=1}^N \varphi^i_{\lambda} \right)^4 \right] d\lambda
\]
By the bound on $\delta_i$ for any distinct $i_1, i_2, i_3$, we define $m^{N,-(i_1,i_2,i_3)} := \frac{1}{N-3} \sum_{\ell \neq i_1, i_2, i_3} \delta_{\varsigma_\ell}$, which implies that
\[
m^N - m^{N,-(i_1,i_2,i_3)} = \frac{1}{N} (\delta_{\varsigma_1} + \delta_{\varsigma_2} + \delta_{\varsigma_3}) - \frac{3}{N(N-3)} \sum_{\ell \neq i_1, i_2, i_3} \delta_{\varsigma_\ell}.
\]

By the definition of second-order linear functional derivatives, we observe that
\[
\begin{align*}
\frac{\delta U}{\delta m}(\lambda m^N + (1 - \lambda)m_0, \varsigma_i) &- \frac{\delta U}{\delta m}(\lambda m^{N,-(i_1,i_2,i_3)} + (1 - \lambda)m_0, \varsigma_i) \\
&= \int_0^1 \int_{\mathbb{R}^d} \frac{\delta^2 U}{\delta m^2} (s\lambda m^N + (1 - s)\lambda m^{N,-(i_1,i_2,i_3)} + (1 - \lambda)m_0, \varsigma_i, v) \left(m^N - m^{N,-(i_1,i_2,i_3)}\right) (dv) \, ds \\
&= \int_0^1 \frac{1}{N} \left[ \sum_{\ell = i_1, i_2, i_3} \frac{\delta^2 U}{\delta m^2} (s\lambda m^N + (1 - s)\lambda m^{N,-(i_1,i_2,i_3)} + (1 - \lambda)m_0, \varsigma_i, \varsigma_\ell) \\
&\quad - \frac{3}{N-3} \sum_{\ell \neq i_1, i_2, i_3} \frac{\delta^2 U}{\delta m^2} (s\lambda m^N + (1 - s)\lambda m^{N,-(i_1,i_2,i_3)} + (1 - \lambda)m_0, \varsigma_i, \varsigma_\ell) \right] \, ds.
\end{align*}
\]

By the bound on $\frac{\delta^2 U}{\delta m^2}$ in Proposition A.5,
\[
\mathbb{E} \left| \frac{\delta U}{\delta m}(\lambda m^N + (1 - \lambda)m_0, \varsigma_i) - \frac{\delta U}{\delta m}(\lambda m^{N,-(i_1,i_2,i_3)} + (1 - \lambda)m_0, \varsigma_i) \right|^4 \leq \frac{C}{N^4} \|\partial^2 U\|_\infty^4 \mathbb{E}[|\varsigma_1|^8].
\]

Similarly, by applying the same argument to the second term in (2.1), we obtain that
\[
\mathbb{E} \left| \frac{\delta^2 U}{\delta m^2}(\lambda m^N + (1 - \lambda)m_0, \tilde{\varsigma}) \right|^4 \leq \frac{C}{N^4} \|\partial^2 U\|_\infty^4 \mathbb{E}[|\varsigma_1|^8],
\]

which implies that
\[
\mathbb{E} |\phi^1_\lambda - \phi^{1,-(i_1,i_2,i_3)}_\lambda|^4 \leq \frac{C}{N^4} \|\partial^2 U\|_\infty^4 \mathbb{E}[|\varsigma_1|^8],
\]

where
\[
\phi^{1,-(i_1,i_2,i_3)}_\lambda = \frac{\delta U}{\delta m}(\lambda m^{N,-(i_1,i_2,i_3)} + (1 - \lambda)m_0, \tilde{\varsigma}_i) - \mathbb{E} \left[ \frac{\delta U}{\delta m}(\lambda m^{N,-(i_1,i_2,i_3)} + (1 - \lambda)m_0, \tilde{\varsigma}_i) \right].
\]
Finally, by writing \( \varphi^i_\lambda = (\varphi^i_\lambda - \varphi^{i,-(i_1,i_2,i_3)}_\lambda) + \varphi^{i,-(i_1,i_2,i_3)}_\lambda \) and applying the generalised Hölder’s inequality to (2.2) and (2.5),

\[
\sum_{i_1,i_2,i_3 \text{ distinct}} \mathbb{E}\left[ \varphi^i_\lambda \varphi^{i_2}_\lambda \varphi^{i_3}_\lambda \right] 
\leq \sum_{i_1,i_2,i_3 \text{ distinct}} \frac{C}{N}(1 + \|\partial_\mu U\|_\infty^4)(1 + \|\partial^2_\mu U\|_\infty^4)\mathbb{E}[|\zeta_1|^8] + \mathbb{E}\left[ \varphi^{i_1,-(i_1,i_2,i_3)}_\lambda \varphi^{i_2,-(i_1,i_2,i_3)}_\lambda (\varphi^{i_3,-(i_1,i_2,i_3)}_\lambda)^2 \right] 
\leq CN^2(1 + \|\partial_\mu U\|_\infty^4)(1 + \|\partial^2_\mu U\|_\infty^4)\mathbb{E}[|\zeta_1|^8] + \sum_{i_1,i_2,i_3 \text{ distinct}} \mathbb{E}\left[ \varphi^{i_1,-(i_1,i_2,i_3)}_\lambda \varphi^{i_2,-(i_1,i_2,i_3)}_\lambda (\varphi^{i_3,-(i_1,i_2,i_3)}_\lambda)^2 \right].
\] 

(2.7)

Let \( \mathcal{F}^{-1} \) be the \( \sigma \)-algebra generated by \( \zeta_1, \ldots, \zeta_N \) except \( \zeta_i \). Since \( \zeta_1, \ldots, \zeta_N \) are independent, for any distinct \( i_1, i_2, i_3 \),

\[
\mathbb{E}\left[ \varphi^{i_1,-(i_1,i_2,i_3)}_\lambda \varphi^{i_2,-(i_1,i_2,i_3)}_\lambda (\varphi^{i_3,-(i_1,i_2,i_3)}_\lambda)^2 \right] = \mathbb{E}\left[ \varphi^{i_2,-(i_1,i_2,i_3)}_\lambda (\varphi^{i_3,-(i_1,i_2,i_3)}_\lambda)^2 \right] \mathbb{E}\left[ \varphi^{i_1,-(i_1,i_2,i_3)}_\lambda \left| \mathcal{F}^{-1} \right. \right] = 0,
\] 

which implies that

\[
\sum_{i_1,i_2,i_3} \mathbb{E}\left[ \varphi^i_\lambda \varphi^{i_2}_\lambda (\varphi^{i_3}_\lambda)^2 \right] \leq CN^2(1 + \|\partial_\mu U\|_\infty^4)(1 + \|\partial^2_\mu U\|_\infty^4)\mathbb{E}[|\zeta_1|^8].
\] 

(2.9)

Next, we define analogously the notation \( \varphi^{i,-(i_1,i_2,i_3,i_4)} \) as (2.6). As above, by applying the generalised Hölder’s inequality to (2.2) and (2.5), followed by a similar reasoning as (2.8), we have

\[
\sum_{i_1,i_2,i_3,i_4 \text{ distinct}} \mathbb{E}\left[ \varphi^i_\lambda \varphi^{i_2}_\lambda \varphi^{i_3}_\lambda \varphi^{i_4}_\lambda \right] 
\leq \sum_{i_1,i_2,i_3,i_4 \text{ distinct}} \frac{C}{N^2}(1 + \|\partial_\mu U\|_\infty^4)(1 + \|\partial^2_\mu U\|_\infty^4)\mathbb{E}[|\zeta_1|^8] + \mathbb{E}\left[ \sum_{j=1}^4 (\varphi^i_\lambda - \varphi^{j,-(i_1,i_2,i_3,i_4)}_\lambda) \prod_{k=1\atop k \neq j}^4 \varphi^{i_k,-(i_1,i_2,i_3,i_4)}_\lambda \right] 
\leq CN^2(1 + \|\partial_\mu U\|_\infty^4)(1 + \|\partial^2_\mu U\|_\infty^4)\mathbb{E}[|\zeta_1|^8] + \sum_{i_1,i_2,i_3,i_4 \text{ distinct}} \mathbb{E}\left[ \sum_{j=1}^4 (\varphi^i_\lambda - \varphi^{j,-(i_1,i_2,i_3,i_4)}_\lambda) \prod_{k=1\atop k \neq j}^4 \varphi^{i_k,-(i_1,i_2,i_3,i_4)}_\lambda \right].
\] 

(2.10)

Note that (2.5) only gives a growth in the order of \( O(N^3) \) for the final term in (2.10), therefore it is insufficient.
By (2.4) followed by an application of the definition of third order linear functional derivatives, we have

\[
\frac{\delta U}{\delta m}(\lambda m^N + (1 - \lambda)m_0, \zeta_i) - \frac{\delta U}{\delta m}(\lambda m^{N,-(i_1,i_2,i_3,i_4)} + (1 - \lambda)m_0, \zeta_i) = \frac{1}{N} \left[ \sum_{\ell = 1, i_1, i_2, i_3, i_4}^N \frac{\delta^2 U}{\delta m^2}\right. \\
\left. \left( \lambda m^N, -(i_1,i_2,i_3,i_4) + (1 - \lambda)m_0, \zeta_i, \zeta_{\ell} \right) - \frac{4}{N - 4} \sum_{\ell \neq 1, i_1, i_2, i_3, i_4} \frac{\delta^2 U}{\delta m^2}\right. \\
\left. \left( \lambda m^N, -(i_1,i_2,i_3,i_4) + (1 - \lambda)m_0, \zeta_i, \zeta_{\ell}, \zeta_{\ell'} \right) \right] + \varepsilon_N^{i,-(i_1,i_2,i_3,i_4)},
\] (2.11)

where

\[
\varepsilon_N^{i,-(i_1,i_2,i_3,i_4)} = \int_0^1 \frac{s\lambda N}{s} \left[ \sum_{\ell = 1, i_1, i_2, i_3, i_4}^N \left[ \sum_{\ell' = 1, i_1, i_2, i_3, i_4}^N \frac{\delta^3 U}{\delta m^3}\right. \\
\left. \left( ts \lambda m^N + (1 - ts)\lambda m^{N,-(i_1,i_2,i_3,i_4)} + (1 - \lambda)m_0, \zeta_i, \zeta_{\ell}, \zeta_{\ell'} \right) \right] \right] \bigg|_0^1 \bigg] \, ds,
\]

which implies that

\[
E|\varepsilon_N^{i,-(i_1,i_2,i_3,i_4)}|^4 \leq \frac{C}{N^8}\|\partial_{\rho}^3 U\|_{s}\|\zeta_1\|_{12}^4,
\]

by the bound on \(\frac{\delta^3 U}{\delta m^3}\) in Proposition A.5. Repeating the same argument to the other term in (2.1) gives

\[
\varphi^{i_1}_\lambda - \varphi^{i_1}_\lambda^{-(i_1,i_2,i_3,i_4)} = \int_{\mathbb{R}^2} \frac{\delta^2 U}{\delta m^2}\left( \lambda m^{N,-(i_1,i_2,i_3,i_4)} + (1 - \lambda)m_0, \zeta_i, v \right) \left( m^N - m^{N,-(i_1,i_2,i_3,i_4)} \right) \, dv \\
\left. - \mathbb{E} \left[ \int_{\mathbb{R}^4} \frac{\delta^2 U}{\delta m^2}\left( \lambda m^{N,-(i_1,i_2,i_3,i_4)} + (1 - \lambda)m_0, \zeta, v \right) \left( m^N - m^{N,-(i_1,i_2,i_3,i_4)} \right) \, dv \right] + \varepsilon_N^{i_1,-(i_1,i_2,i_3,i_4)},
\]

where

\[
E|\varepsilon_N^{i_1,-(i_1,i_2,i_3,i_4)}|^4 \leq \frac{C}{N^8}\|\partial_{\rho}^3 U\|_{s}\|\zeta_1\|_{12}^4.
\] (2.12)

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

\[
\varphi^{i_1}_\lambda - \varphi^{i_1}_\lambda^{-(i_1,i_2,i_3,i_4)} - \varepsilon_N^{i_1,-(i_1,i_2,i_3,i_4)} = \sum_{j=2}^4 F_j\left( (\zeta_r)_{r \neq i_1, \ldots, i_4}, \zeta_{i_1}, \zeta_{i_2} \right),
\]

11
for some measurable functions $F_2, F_3, F_4 : (\mathbb{R}^d)^{N-2} \to \mathbb{R}$. Therefore,

$$E \left[ \left( \varphi^2_{\lambda} - \varphi^4_{\lambda} \right) \left( \varphi^2_{\lambda} - \varphi^4_{\lambda} \right) \right] = 0.$$ 

Applying the generalised Hölder’s inequality to (2.12) and (2.2) gives

$$E \left[ \left( \varphi^2_{\lambda} - \varphi^4_{\lambda} \right) \right] \leq \frac{C}{N^2} \| \partial^2 U \|_\infty \left( E[|\zeta_1|^{12}] \right)^{1/4} \| \partial_\mu U \|_\infty^3 \left( E[|\zeta_1|^{4}] \right)^{3/4} \leq \frac{C}{N^2} \left( 1 + \| \partial_\mu U \|_\infty^4 \right) \left( 1 + \| \partial^2 U \|_\infty^4 \right) (1 + E[|\zeta_1|^{12}]).$$

By the same reasoning, we can show that

$$\sum_{i_1,i_2,i_3,i_4} E \left[ \left( \varphi^2_{\lambda} - \varphi^4_{\lambda} \right) \right] \leq C N^2 \left( 1 + \| \partial_\mu U \|_\infty^4 \right) \left( 1 + \| \partial^2 U \|_\infty^4 \right) (1 + E[|\zeta_1|^{12}]).$$

We conclude the result by combining (2.3), (2.9), (2.10) and (2.13).

We now introduce a mean-field coupling of the particle system (1.5) by

$$dX^i_t = \xi_i + \int_0^t b(X^i_s, \mu^X_s) \, ds + \int_0^t \sigma(X^i_s, \mu^X_s) \, dW^i_s, \quad 1 \leq i \leq N, \quad t \in [0,T],$$

$$\mu^X_{\lambda,N} := \frac{1}{N} \sum_{i=1}^N \delta_{X^i_s}.$$  

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

- $b$ and $\sigma$ are Lipschitz continuous with respect to the Euclidean norm and the $W_2$ norm, \hspace{1cm} (Lip)
- $\Phi$ is Lipschitz continuous with respect to the $W_2$ norm,

and that the initial law $\nu$ satisfies

$$\int_{\mathbb{R}^d} |x|^{12} \nu(dx) < +\infty.$$  \hspace{1cm} (Int)

Note that (Lip) guarantees strong existence and uniqueness of (1.4) and (1.5). The following proposition is essential to the proofs of Theorem 2.4 and Theorem 3.1.
Proposition 2.3. Assume (Lip) and (Int). Suppose that \( \varphi \in \mathcal{M}_d(\mathbb{R}^d \times \mathcal{P}_2(\mathbb{R}^d)) \). Then

\[
\frac{1}{N} \sum_{i=1}^{N} \sup_{t \in [0,T]} \mathbb{E} \left| \varphi(X_i^t, \mu^N_t) - \varphi(X_i^t, \mu_t^N) \right|^4 \leq \frac{C}{N^2},
\]

for some constant \( C > 0 \).

Proof.

\[
\frac{1}{N} \sum_{i=1}^{N} \sup_{t \in [0,T]} \mathbb{E} \left[ \varphi \left( X_i^t, \frac{1}{N} \sum_{j=1}^{N} \delta_{X_j^t} \right) - \varphi(X_i^t, \mu_t^N) \right]^4 \\
= \frac{1}{N} \sum_{i=1}^{N} \sup_{t \in [0,T]} \mathbb{E} \left[ \left| \varphi \left( \eta, \frac{1}{N} \delta \eta + \frac{N-1}{N} \sum_{1 \leq j \leq N, j \neq i} \delta_{X_j^t} \right) - \varphi(\eta, \mu_t^N) \right|^4 \right]_{\eta = X_i^t} \\
\leq \frac{8}{N} \sum_{i=1}^{N} \sup_{t \in [0,T]} \mathbb{E} \left[ \left| \varphi \left( \eta, \frac{1}{N} \delta \eta + \frac{N-1}{N} \sum_{1 \leq j \leq N, j \neq i} \delta_{X_j^t} \right) - \varphi(\eta, \mu_t^N) \right|^4 \right]_{\eta = X_i^t} \\
\quad + \frac{8}{N} \sum_{i=1}^{N} \sup_{t \in [0,T]} \mathbb{E} \left[ \left| \varphi \left( \eta, \frac{1}{N} \sum_{1 \leq j \leq N, j \neq i} \delta_{X_j^t} \right) - \varphi(\eta, \mu_t^N) \right|^4 \right]_{\eta = X_i^t} \\
= : \Pi_1 + \Pi_2.
\]

By Lemma 2.1,

\[
\Pi_1 \leq \frac{8}{N} \sum_{i=1}^{N} \sup_{t \in [0,T]} \mathbb{E} \left[ \frac{4}{N^2} \left( |X_i^t|^2 + \frac{1}{N-1} \sum_{1 \leq j \leq N, j \neq i} |X_j^t|^2 \right) \right] \lesssim \frac{1}{N^2}.
\] (2.15)

By the assumption on \( \varphi \), we observe that for any \( \eta \in \mathbb{R}^d \), the uniform bounds on \( \partial_\mu \varphi(\eta, \cdot), \partial^2_{\mu} \varphi(\eta, \cdot) \) and \( \partial^3_{\mu} \varphi(\eta, \cdot) \) do not depend on \( \eta \). Finally, since \( b \) and \( \sigma \) are Lipschitz and \( \mathbb{E}[|\xi|^{12}] < +\infty \), we have \( \sup_{t \in [0,T]} \mathbb{E}[|X_t|^2] < +\infty \). Therefore, Lemma 2.2 implies that

\[
\Pi_2 \lesssim \frac{1}{(N-1)^2} \prod_{i=1}^{3} \left( 1 + \sup_{\eta \in \mathbb{R}^d} \| \partial_\mu \varphi(\eta, \cdot)\|_{\infty} \right) \left( 1 + \sup_{t \in [0,T]} \int_{\mathbb{R}^d} |y|^{12} \mu_t^X(dy) \right).
\] (2.16)

A combination of (2.15) and (2.16) yields the result. \( \square \)
Note that Proposition 2.3 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 [35] gives us a rate of convergence of

$$\mathbb{E}\left[ \sup_{t \in [0,T]} W_2(\mu_t^{X}, \mu_t^{X,N})^2 \right] \leq \frac{C}{N^{2/d+8}}. \quad (2.17)$$

There are results in the literature that give a slightly better rate of convergence of the $W_2$ norm of empirical measures of i.i.d. random variables. However, they are not for i.i.d. processes and are still dimensionally dependent.

The following result gives a uniform rate of strong propagation of chaos between the particle system (1.5) and its coupled mean-field limit (2.14), under the assumption that $b$ and $\sigma$ are sufficiently smooth. Let $C_T := C([0,T], \mathbb{R}^d)$ be the space of continuous functions from $[0,T]$ to $\mathbb{R}^d$ equipped with the supremum norm and $W_{C_T,2}$ be the 2-Wasserstein metric on $C_T$.

**Theorem 2.4** (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

$$\mathbb{E}\left[ W_{C_T,2}(\mu^{Y,N}, \mu^{X,N})^4 \right] \leq \mathbb{E}\left[ \frac{1}{N} \sum_{i=1}^{N} \left( \sup_{t \in [0,T]} |X_i^t - Y_i^{t,N}|^4 \right) \right] \leq \frac{C}{N^2}, \quad (2.18)$$

for some constant $C > 0$.

**Proof.** By the Hölder and Buckholder-Davis-Gundy inequalities, estimating the $L^4$ difference between (1.5) and (2.14) gives

$$\mathbb{E}\left[ \sup_{s \in [0,t]} |X_s^i - Y_s^{i,N}|^4 \right] \leq C \left( \int_0^t \mathbb{E} |b(X_s^i, \mu_s^X) - b(Y_s^{i,N}, \mu_s^{Y,N})|^4 ds + \int_0^t \mathbb{E} \|\sigma(X_s^i, \mu_s^X) - \sigma(Y_s^{i,N}, \mu_s^{Y,N})\|^4 ds \right), \quad (2.18)$$

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

$$\mathbb{E}\left[ \sup_{s \in [0,t]} |X_s^i - Y_s^{i,N}|^4 \right] \leq C \left( \int_0^t \mathbb{E} \left[ \sup_{u \in [0,s]} |X_u^i - Y_u^{i,N}|^4 \right] ds + \int_0^t \mathbb{E} |b(X_s^i, \mu_s^X) - b(X_s^{i,N}, \mu_s^{Y,N})|^4 ds + \int_0^t \mathbb{E} \|\sigma(X_s^i, \mu_s^X) - \sigma(X_s^{i,N}, \mu_s^{Y,N})\|^4 ds \right),$$

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

$$\frac{1}{N} \sum_{i=1}^{N} \mathbb{E}\left[ \sup_{s \in [0,t]} |X_s^i - Y_s^{i,N}|^4 \right] \leq C \left( \int_0^t \frac{1}{N} \sum_{i=1}^{N} \mathbb{E} \left[ \sup_{u \in [0,s]} |X_u^i - Y_u^{i,N}|^4 \right] ds \right)$$
\[ + \int_0^t \frac{1}{N} \sum_{i=1}^N \mathbb{E}|b(X^i_s, \mu^X_s) - b(X^i_s, \mu^Y,N_s)|^4 \, ds \]
\[ + \int_0^t \frac{1}{N} \sum_{i=1}^N \mathbb{E}||\sigma(X^i_s, \mu^X_s) - \sigma(X^i_s, \mu^Y,N_s)||^4 \, ds \]. \tag{2.19} \]

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

\[ \mathbb{E}[W_2(\mu^{X,N}_s, \mu^{Y,N}_s)^4] \leq \left[ \left( \frac{1}{N} \sum_{i=1}^N \mathbb{E}|Y^{i,N}_s - X^{i}_s|^2 \right)^2 \right] \leq \frac{1}{N} \sum_{i=1}^N \mathbb{E}\left[ \sup_{u \in [0,s]} |X^i_u - Y^{i,N}_u|^4 \right]. \tag{2.20} \]

A combination of (2.19) and (2.20) gives

\[
\frac{1}{N} \sum_{i=1}^N \mathbb{E}\left[ \sup_{s \in [0,t]} |X^i_s - Y^{i,N}_s|^4 \right] \leq C \left( \int_0^t \frac{1}{N} \sum_{i=1}^N \mathbb{E}\left[ \sup_{u \in [0,s]} |X^i_u - Y^{i,N}_u|^4 \right] \right) ds
\]
\[ + \int_0^t \frac{1}{N} \sum_{i=1}^N \sup_{u \in [0,s]} \mathbb{E}|b(X^i_u, \mu^X_u) - b(X^i_u, \mu^Y,N_u)|^4 \, ds 
\]
\[ + \int_0^t \frac{1}{N} \sum_{i=1}^N \sup_{u \in [0,s]} \mathbb{E}||\sigma(X^i_u, \mu^X_u) - \sigma(X^i_u, \mu^Y,N_u)||^4 \, ds \].

Therefore, by Proposition 2.3 and Gronwall’s inequality, we have

\[ \frac{1}{N} \sum_{i=1}^N \mathbb{E}\left[ \sup_{s \in [0,T]} |X^i_s - Y^{i,N}_s|^4 \right] \leq C \frac{1}{N^2}, \]

for every \( t \in [0, T] \). \( \square \)

We now recall, from Section 9 in [19], that the the second moment of the antithetic difference (see (3.1) for the definition of \( \mu^{Y,2N,(1)} \) and \( \mu^{Y,2N,(2)} \)) given by

\[ U(\mu^{Y,2N}_0) - \frac{1}{2}(U(\mu^{Y,2N,(1)}_0) + U(\mu^{Y,2N,(2)}_0)) \]

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

\[ U(\mu) := F \left( \int_{\mathbb{R}^d} G(x) \, \mu(dx) \right), \tag{2.21} \]

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 is a generalisation of this result.
Theorem 2.5 (Antithetic error on the initial conditions). Suppose that $\nu \in \mathcal{P}_b(\mathbb{R}^d)$ and $U \in \mathcal{M}_4^L(\mathcal{P}_2(\mathbb{R}^d))$. Then there exists a constant $C > 0$ such that

$$
\mathbb{E}[U(\mu_0^{Y;2N}) - \frac{1}{2}(U(\mu_0^{Y;2N,(1)}) + U(\mu_0^{Y;2N,(2)}))]^2 \leq \frac{C}{N^2}.
$$

Proof. For simplicity of notations, let

$$
\mu_{2N} := \mu_0^{Y;2N}, \quad \mu_{2N,(1)} := \mu_0^{Y;2N,(1)}, \quad \mu_{2N,(2)} := \mu_0^{Y;2N,(2)}.
$$

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

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

We define

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

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

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

This yields

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

(2.22)

Similarly,

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

(2.23)

and

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

(2.24)

Computing the difference of (2.22) with the arithmetic average of (2.23) and (2.24) gives

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

16
\[-\frac{1}{2} \int_0^1 (1-t) \left[ \int_{\mathbb{R}^{2d}} \frac{\delta^2 U}{\delta m^2} (m_t^{2N,1}(1))(y) \left( \mu_{2N,(1)} - \nu \right)^{\otimes 2}(dy) \right] dt \]
\[-\frac{1}{2} \int_0^1 (1-t) \left[ \int_{\mathbb{R}^{2d}} \frac{\delta^2 U}{\delta m^2} (m_t^{2N,2}(1))(y) \left( \mu_{2N,(2)} - \nu \right)^{\otimes 2}(dy) \right] dt. \]

(2.25)

The rest of the proof is very similar to the proof of Lemma 2.2. It suffices to consider only the first term in (2.25). The other two terms can be handled in a similar way. We rewrite

\[
\int_{\mathbb{R}^d} \frac{\delta^2 U}{\delta m^2} (m_t^{2N}(1))(y) \left( \mu_{2N} - \nu \right)^{\otimes 2}(dy)
\]

\[
= \int_{\mathbb{R}^d} \left[ \frac{1}{2N} \sum_{i=1}^{2N} \frac{\delta^2 U}{\delta m^2} (m_t^{2N})(\xi_i, y_2) - \int_{\mathbb{R}^d} \frac{\delta^2 U}{\delta m^2} (m_t^{2N})(z, y_2) \nu(dz) \right] (\mu_{2N} - \nu)(dy_2)
\]

\[
= \frac{1}{(2N)^2} \sum_{i,j=1}^{2N} \frac{\delta^2 U}{\delta m^2} (m_t^{2N})(\xi_i, \xi_j) - \frac{1}{2N} \sum_{j=1}^{2N} \int_{\mathbb{R}^d} \frac{\delta^2 U}{\delta m^2} (m_t^{2N})(z, \xi_j) \nu(dz)
\]

\[-\frac{1}{2N} \sum_{i=1}^{2N} \int_{\mathbb{R}^d} \frac{\delta^2 U}{\delta m^2} (m_t^{2N})(\xi_i, z) \nu(dz) + \int_{\mathbb{R}^d} \int_{\mathbb{R}^d} \frac{\delta^2 U}{\delta m^2} (m_t^{2N})(z, z') \nu(dz) \nu(dz')
\]

\[
= \frac{1}{(2N)^2} \sum_{i,j=1}^{2N} \varphi_t^{(i,j)}, \tag{2.26}
\]

where

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

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

Next, we observe that

\[
\frac{1}{N^2} \sum_{i,j=1}^{2N} \varphi_t^{(i,j)} \leq \frac{1}{N^2} \sum_{i_1,j_1,i_2,j_2 \in \{1,\ldots,2N\}} \text{E} \left[ \varphi_t^{(i_1,j_1)} \varphi_t^{(i_2,j_2)} \right] + \sum_{i_1,j_1,i_2,j_2 \in \{1,\ldots,2N\}} \text{E} \left[ \varphi_t^{(i_1,j_1)} \varphi_t^{(i_2,j_2)} \right].
\]

(2.27)

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 2.2, we define

\[
\varphi_t^{(i,j),-(i_1,j_1,j_2)}
\]

17
By the Cauchy-Schwarz inequality and the bound on $\delta m$ in (A.11) (see (2.4) for details), we have

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

Then, we write

$$\mathbb{E}\left[\varphi_t^{(i_1,j_1)}\varphi_t^{(i_1,j_2)}\varphi_t^{(i_1,j_3)}\varphi_t^{(i_1,j_4)}\right] = \mathbb{E}\left[(\varphi_t^{(i_1,j_1)})^2 - \varphi_t^{(i_1,j_1),-(i_1,j_1,j_2)}\right]$$

$$+ \mathbb{E}\left[(\varphi_t^{(i_1,j_1),-(i_1,j_1,j_2)})^2 - \varphi_t^{(i_1,j_1),-(i_1,j_1,j_2)-,(i_1,j_1,j_3,j_4)}\right]$$

$$+ \mathbb{E}\left[\varphi_t^{(i_1,j_1),-(i_1,j_1,j_2),-(i_1,j_1,j_3,j_4)}\varphi_t^{(i_1,j_1),-(i_1,j_1,j_2),-(i_1,j_1,j_3,j_4)}\right].$$

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

$$\mathbb{E}\left[\varphi_t^{(i_1,j_1),-(i_1,j_1,j_2),-(i_1,j_1,j_3,j_4)}\varphi_t^{(i_1,j_2),-(i_1,j_1,j_2),-(i_1,j_1,j_3,j_4)}\right] = \mathbb{E}\left[\varphi_t^{(i_1,j_1),-(i_1,j_1,j_2)}\mathbb{E}\left[\varphi_t^{(i_1,j_2),-(i_1,j_1,j_2)}\mathcal{F}^{-j_2}\right]\right] = 0.$$  

Therefore,

$$\frac{1}{N^4} \sum_{i_1,j_1,j_2,j_3 \in \{1, \ldots, 2N\} \text{ exactly two of } i_1,j_1,j_2,j_3 \text{ are identical}} \mathbb{E}\left[\varphi_t^{(i_1,j_1),-(i_1,j_1,j_2)}\varphi_t^{(i_1,j_2),-(i_1,j_1,j_2)}\right] \leq \frac{1}{N^2}. \tag{2.29}$$

Finally, we consider the case where $i_1, j_1, i_2, j_2$ are mutually distinct. We define $\varphi_t^{(i,j),-(i_1,j_1,j_2,j_3)}$ analogously, as the definition of $\varphi_t^{(i,j),-(i_1,j_1,j_2,j_3)}$ in (2.28). As above, we write

$$\mathbb{E}\left[\varphi_t^{(i_1,j_1),-(i_1,j_1,j_2)}\varphi_t^{(i_1,j_2),-(i_1,j_1,j_2)}\varphi_t^{(i_1,j_3),-(i_1,j_1,j_2,j_3)}\varphi_t^{(i_1,j_4),-(i_1,j_1,j_2,j_3)}\right]$$

$$= \mathbb{E}\left[(\varphi_t^{(i_1,j_1),-(i_1,j_1,j_2)})^2 - \varphi_t^{(i_1,j_1),-(i_1,j_1,j_2)-,(i_1,j_1,j_2,j_3,j_4)}\right]$$

$$+ \mathbb{E}\left[(\varphi_t^{(i_1,j_1),-(i_1,j_1,j_2),-(i_1,j_1,j_2,j_3,j_4)})^2 - \varphi_t^{(i_1,j_1),-(i_1,j_1,j_2),-(i_1,j_1,j_2,j_3,j_4)-,(i_1,j_1,j_2,j_3,j_4)}\right]$$

$$+ \mathbb{E}\left[\varphi_t^{(i_1,j_1),-(i_1,j_1,j_2),-(i_1,j_1,j_2,j_3,j_4)}\varphi_t^{(i_1,j_1),-(i_1,j_1,j_2),-(i_1,j_1,j_2,j_3,j_4)}\right].$$
\[ + \mathbb{E} \left[ \varphi_t^{(1_1,1_1),-(1_1,1_1,1_1,1_1,1_1,1_1,1_1,1_1,1_1,1_1,1_1,1_1,1_1,1_1)} \right]. \]

As before, we have
\[ \mathbb{E} |\varphi_t^{(1,j)} - \varphi_t^{(1,j),-(i_1,j_1,i_2,j_2)}|^2 \lesssim \frac{1}{N^2} \]
and hence
\[ \mathbb{E} \left( \varphi_t^{(1,j_1)} - \varphi_t^{(1,j_1),-(i_1,j_1,i_2,j_2)} \right)^2 \lesssim \frac{1}{N^2}, \quad (2.30) \]
by the Cauchy-Schwarz inequality. By the same argument as in the proof of Lemma 2.2 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 (2.11) and (2.12) for details), we obtain that
\[ \varphi_t^{(1,j_1)} - \varphi_t^{(1,j_1),-(i_1,j_1,i_2,j_2)} = F_1((\xi_r)_{r \neq i_1,j_1,i_2,j_2}) \xi_{i_1}, \xi_{j_1}, \xi_{j_2}) + F_2((\xi_r)_{r \neq i_1,j_1,i_2,j_2}) \xi_{i_1}, \xi_{j_1}, \xi_{j_2}) \]
for some measurable functions \( F_1, F_2 : \mathbb{R}^d N \to \mathbb{R} \), where
\[ \mathbb{E} \left[ \varphi_t^{(1,j_1),-(i_1,j_1,i_2,j_2)} \right]^2 \lesssim \frac{1}{N^2}. \]

By a similar conditioning argument as the proof of Lemma 2.2,
\[ \mathbb{E} \left[ \varphi_t^{(1,j_1)} - \varphi_t^{(1,j_1),-(i_1,j_1,i_2,j_2)} \right]^2 \lesssim \frac{1}{N^2}. \]

Similarly,
\[ \mathbb{E} \left( \varphi_t^{(1,j_1),-(i_1,j_1,i_2,j_2)} \right)^2 \lesssim \frac{1}{N^2}. \]

By the same conditioning argument,
\[ \mathbb{E} \left[ \varphi_t^{(1,j_1),-(i_1,j_1,i_2,j_2)} \right]^2 = \mathbb{E} \left[ \varphi_t^{(1,j_1),-(i_1,j_1,i_2,j_2)} \right]^2 = 0. \]

A combination of (2.30), (2.31), (2.32) and (2.33) implies that
\[ \frac{1}{N^4} \sum_{i_1,i_2,i_3,i_4,i_5,i_6,i_7,i_8,i_9,i_{10},i_{11},i_{12},i_{13},i_{14},i_{15},i_{16},i_{17},i_{18},i_{19},i_{20}} \mathbb{E} \left[ \varphi_t^{(1,j_1),-(i_1,j_1,i_2,j_2)} \right] \lesssim \frac{1}{N^2}. \]

19
Finally, a combination of (2.26), (2.27), (2.29) and (2.34) implies that
\[
\mathbb{E} \left| \int_{\mathbb{R}^{2d}} \frac{\partial^2 U}{\partial m^2} (m_t\Phi(y)) (\mu_{2N} - \nu)^\otimes 2(dy) \right|^2 \lesssim \frac{1}{N^2}.
\]

3 Antithetic MLMC without time discretisation

We begin this section by elaborating on the idea of multilevel Monte-Carlo simulation that was discussed in the introduction. For each level \( \ell \), we approximate \( \mathbb{E}[\Phi(Y_{N,1}^\ell)] \) by a standard Monte-Carlo estimator. Subsequently, we combine this approximation with the antithetic trick, which involves estimating the second random variables of the differences in the telescopic sum by the arithmetic average of two sub-particle systems. For simplicity, we set
\[
N_\ell := 2^\ell, \quad \ell \in \{0, \ldots, L\}.
\]
We also set the two sub-particle systems to have the same number of particles. More precisely, we define the pair of sub-particle systems as \( \{Y_{i,2N}^\ell\}_{i=1}^{2N} \) as
\[
Y_{i,2N}^{\ell,1} = \xi_i + \int_0^t b(Y_{r,2N}^{\ell,1}, t) \, dr + \int_0^t \sigma(Y_{r,2N}^{\ell,1}, t) \, dW_r^i, \quad 1 \leq i \leq N,
\]
\[
Y_{i,2N}^{\ell,2} = \xi_i + \int_0^t b(Y_{r,2N}^{\ell,2}, t) \, dr + \int_0^t \sigma(Y_{r,2N}^{\ell,2}, t) \, dW_r^i, \quad N + 1 \leq i \leq 2N,
\]
where
\[
\mu_{r,Y_{i,2N}^{\ell,1}} := \frac{1}{N} \sum_{i=1}^N \delta_{Y_{i,2N}^{\ell,1}, t} \quad \text{and} \quad \mu_{r,Y_{i,2N}^{\ell,2}} := \frac{1}{N} \sum_{i=N+1}^{2N} \delta_{Y_{i,2N}^{\ell,2}, t}.
\]
Therefore, we define the theoretical MLMC estimator (without time discretisation) as
\[
A^{\text{MLMC}} := \frac{1}{M_0} \sum_{\theta=1}^{M_0} \Phi(y, N_0), (0) \]
\[
+ \sum_{\ell=1}^L \frac{1}{M_\ell} \sum_{\theta=1}^{M_\ell} \left[ \Phi(y, N_\ell), (0) \right] - \frac{1}{2} \left[ \Phi(y, N_\ell), (1) \right] + \Phi(y, N_\ell), (2) \right]\right],
\]
where \( \mu_{r,Y_{i,2N}^{\ell,1}}, \mu_{r,Y_{i,2N}^{\ell,2}} \) and \( \mu_{r,Y_{i,2N}^{\ell,1}}, \mu_{r,Y_{i,2N}^{\ell,2}} \) are defined similarly as \( \mu_{r,Y_{i,2N}^{\ell,1}}, \mu_{r,Y_{i,2N}^{\ell,1}} \) and \( \mu_{r,Y_{i,2N}^{\ell,2}}, \mu_{r,Y_{i,2N}^{\ell,2}} \) 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 \( \ell, \theta \)) 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 \( i, \ell \) and \( \theta \).
The following theorem states that the variance of the antithetic difference in (3.2) converges in $N$ in the rate $O(1/N^2)$. In the proof, Proposition 2.3 and Theorem 2.4 provide us with the necessary estimates when we revert to the mean-field limit.

**Theorem 3.1** (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

$$\var[\Phi(\mu_T^{Y,2N}) - \frac{1}{2}(\Phi(\mu_T^{Y,2N,(1)}) + \Phi(\mu_T^{Y,2N,(2)}))] \leq \mathbb{E}\left|\Phi(\mu_T^{Y,2N}) - \frac{1}{2}(\Phi(\mu_T^{Y,2N,(1)}) + \Phi(\mu_T^{Y,2N,(2)}))\right|^2 \leq \frac{C}{N^2},$$

where $C$ is a constant that depends on $\Phi, b, \sigma$ and $T$, but does not depend on $N$.

**Proof.** We begin by recalling the representation obtained in (B.7)

$$\Phi(\mu_T^{Y,2N}) - \Phi(\mu_T^Y) = (\mathcal{V}(0, \mu_0^{Y,2N}) - \mathcal{V}(0, \nu)) + \int_0^T \frac{1}{2} \left[\frac{1}{N^2} \sum_{i=1}^N \text{Tr}\left(a(Y_s^{i,N}, Y_s^{Y,N}) \frac{\partial \mathcal{V}}{\partial \mu}(s, \mu_s^{Y,N})(Y_s^{i,N}, Y_s^{Y,N})\right)\right] ds + \frac{1}{N} \sum_{i=1}^N \int_0^T \sigma(Y_s^{i,N}, \mu_s^{Y,N})^T \frac{\partial \mathcal{V}}{\partial \mu}(s, \mu_s^{Y,N})(Y_s^{i,N}) \cdot dW_s^i.$$

Hence,

$$\Phi(\mu_T^{Y,2N}) - \frac{1}{2}(\Phi(\mu_T^{Y,2N,(1)}) + \Phi(\mu_T^{Y,2N,(2)})) = \mathcal{A} + \mathcal{B} + \mathcal{J},$$

where

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

$$\mathcal{B} := \int_0^T \frac{1}{2} \left[\frac{1}{(2N)^2} \sum_{i=1}^{2N} \text{Tr}\left(a(Y_s^{i,N}, \mu_s^{Y,N}) \frac{\partial \mathcal{V}}{\partial \mu}(s, \mu_s^{Y,N})(Y_s^{i,2N}, Y_s^{Y,2N})\right)\right]$$

$$- \frac{1}{2N^2} \sum_{i=1}^N \text{Tr}\left(a(Y_s^{i,2N,(1)}, \mu_s^{Y,2N,(1)}) \frac{\partial \mathcal{V}}{\partial \mu}(s, \mu_s^{Y,2N,(1)})(Y_s^{i,2N,(1)}, Y_s^{Y,2N,(1)})\right)$$

$$+ \sum_{i=N+1}^{2N} \text{Tr}\left(a(Y_s^{i,2N,(2)}, \mu_s^{Y,2N,(2)}) \frac{\partial \mathcal{V}}{\partial \mu}(s, \mu_s^{Y,2N,(2)})(Y_s^{i,2N,(2)}, Y_s^{Y,2N,(2)})\right) ds$$

and

$$\mathcal{J} := \sum_{i=1}^{2N} \int_0^T \frac{1}{2N} \frac{\partial \mathcal{V}}{\partial \mu}(s, \mu_s^{Y,2N})(Y_s^{i,2N})^T \sigma(Y_s^{i,2N}, \mu_s^{Y,2N}) dW_s^i.$$
\[-\frac{1}{2N} \left( \sum_{i=1}^{N} \int_{0}^{T} \partial_{\mu} \mathcal{V} \left( \mu Y_{s}^{i,2N,(1)} \right) (Y_{s}^{i,2N,(1)})^{T} \sigma (Y_{s}^{i,2N,(1)}, \mu Y_{s}^{i,2N,(1)}) dW_{s}^{i} \right) + \sum_{i=N+1}^{2N} \int_{0}^{T} \partial_{\mu} \mathcal{V} \left( s, \mu Y_{s}^{2N,(2)} \right) (Y_{s}^{i,2N,(2)})^{T} \sigma (Y_{s}^{i,2N,(2)}, \mu Y_{s}^{i,2N,(2)}) dW_{s}^{i} \right].\]

By the assumptions on \( b, \sigma \) and \( \Phi \), it follows from Theorem B.1 that \( \mathcal{V} \in \mathcal{M}_{4}(\mathbb{R}^{d}) \). We can therefore see that
\[ E[\mathcal{G}] \lesssim 1/N^2. \]

In particular, \( \mathcal{V}(0, \cdot) \in \mathcal{M}_{4}(\mathcal{P}_2(\mathbb{R}^{d})) \). Therefore, by Theorem 2.5, we obtain that
\[ E[\mathcal{G}] \lesssim 1/N^2. \]

Hence, it remains to show that \( E[\mathcal{G}] \lesssim 1/N^2 \). 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 \( E[\mathcal{G}] \) as
\[ E[\mathcal{G}] = E \left[ \left( \frac{1}{2N} \sum_{i=1}^{N} \int_{0}^{T} \Sigma(s, Y_{s}^{i,2N,(1)}, \mu Y_{s}^{i,2N}) - \Sigma(s, Y_{s}^{i,2N,(1)}, \mu Y_{s}^{i,2N,(1)}) dW_{s}^{i} \right)^{2} \right] + E \left[ \left( \frac{1}{2N} \sum_{i=N+1}^{2N} \int_{0}^{T} \Sigma(s, Y_{s}^{i,2N,(2)}, \mu Y_{s}^{i,2N}) - \Sigma(s, Y_{s}^{i,2N,(2)}, \mu Y_{s}^{i,2N,(2)}) dW_{s}^{i} \right)^{2} \right]. \]

Using the independence of the Brownian motions and Itô’s isometry,
\[ E \left[ \left( \frac{1}{2N} \sum_{i=1}^{N} \int_{0}^{T} \Sigma(s, Y_{s}^{i,2N,(1)}, \mu Y_{s}^{i,2N}) - \Sigma(s, Y_{s}^{i,2N,(1)}, \mu Y_{s}^{i,2N,(1)}) dW_{s}^{i} \right)^{2} \right] \]
\[ = \frac{1}{4N^2} \sum_{i=1}^{N} E \left[ \left( \int_{0}^{T} \Sigma(s, Y_{s}^{i,2N,(1)}, \mu Y_{s}^{i,2N}) - \Sigma(s, Y_{s}^{i,2N,(1)}, \mu Y_{s}^{i,2N,(1)}) dW_{s}^{i} \right)^{2} \right] \]
\[ = \frac{1}{4N^2} \sum_{i=1}^{N} \int_{0}^{T} E \left[ \left\| \Sigma(s, Y_{s}^{i,2N,(1)}, \mu Y_{s}^{i,2N}) - \Sigma(s, Y_{s}^{i,2N,(1)}, \mu Y_{s}^{i,2N,(1)}) \right\|^{2} ds. \]

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 \( \sigma \) is Lipschitz continuous. By Theorem 2.4,
\[ \sup_{t \in [0, T]} E \left[ \left\| \Sigma(t, Y_{t}^{i,2N}, \mu Y_{t}^{i,2N}) - \Sigma(t, X_{t}^{i}, \mu X_{t}^{i,2N}) \right\|^{2} \right] \]
\[ = \sup_{t \in [0, T]} E \left[ \left\| \partial_{\mu} \mathcal{V}(t, \mu Y_{t}^{i,2N}) (Y_{t}^{i,2N})^{T} \sigma(Y_{t}^{i,2N}, \mu Y_{t}^{i,2N}) - \partial_{\mu} \mathcal{V} Y_{t}^{i,2N}(X_{t}^{i})^{T} \sigma(X_{t}^{i}, \mu X_{t}^{i,2N}) \right\|^{2} \right] \]
\[ \lesssim \sup_{t \in [0, T]} E \left[ \left\| \partial_{\mu} \mathcal{V}(t, \mu Y_{t}^{i,2N}) (Y_{t}^{i,2N})^{T} (\sigma(Y_{t}^{i,2N}, \mu Y_{t}^{i,2N}) - \sigma(X_{t}^{i}, \mu X_{t}^{i,2N})) \right\|^{2} \right]. \]
A combination of (3.3), (3.4), (3.5) and (3.6) gives
\[
\mathbb{E}\left[\left(\frac{1}{2N} \sum_{i=1}^{N} \int_{0}^{T} \Sigma(s, Y_{s,i,2N}, \mu_{s,2N}) - \Sigma(s, Y_{s,i,2N,1}, \mu_{s,2N,1}) dW_{s}^{i}\right)^{2}\right] \lesssim \frac{1}{N^{2}}.
\]

Similarly,
\[
\mathbb{E}\left[\left(\frac{1}{2N} \sum_{i=N+1}^{2N} \int_{0}^{T} \Sigma(s, Y_{s,i,2N}, \mu_{s,2N}) - \Sigma(s, Y_{s,i,2N,2}, \mu_{s,2N,2}) dW_{s}^{i}\right)^{2}\right] \lesssim \frac{1}{N^{2}}.
\]
Consequently, \( E[\mathcal{S}^2] \leq \frac{1}{N^2}. \)

We now perform an analysis on the complexity of this algorithm. Recall that, by Theorem B.2,
\[
|E[\Phi(\mu_T^{Y,N})] - \Phi(\mu_T^X)| \leq \frac{C}{N^\ell},
\]
(i)

Moreover, by Theorem 3.1, we have
\[
\text{Var}\left[\Phi(\mu_T^{Y,N_r,\ell,1}) - \frac{1}{2}\left(\Phi(\mu_T^{Y,N_r,\ell,1,1}) + \Phi(\mu_T^{Y,N_r,\ell,1,2})\right)\right] \leq \frac{C}{N^{2\ell}},
\]
(ii)

Since simulating particle systems of \( N \) particles requires \( N^2 \) operations in general, the cost function of the antithetic difference is bounded by
\[
\text{Cost}\left[\Phi(\mu_T^{Y,N_r,\ell,1}) - \frac{1}{2}\left(\Phi(\mu_T^{Y,N_r,\ell,1,1}) + \Phi(\mu_T^{Y,N_r,\ell,1,2})\right)\right] \leq CN^2\ell.
\]
(iii)

Properties (i) to (iii) allow us to conclude the complexity of the algorithm.

**Theorem 3.2** (Complexity of theoretical antithetic MLMC). Assume (Int). Suppose that \( b, \sigma \in \mathcal{M}_d(\mathbb{R}^d \times \mathcal{P}_2(\mathbb{R}^d)) \) and \( \Phi \in \mathcal{M}_d(\mathcal{P}_2(\mathbb{R}^d)). \) 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 root-mean-square error of \( \mathcal{A}^{A-\text{MLMC}} \) is bounded by
\[
\left(E\left[\left(\mathcal{A}^{A-\text{MLMC}} - \Phi(\mu^X)^2\right)\right]\right)^{1/2} \leq C_1 \epsilon
\]
and the computational cost of \( \mathcal{A}^{A-\text{MLMC}} \) is bounded by
\[
\text{Cost}(\mathcal{A}^{A-\text{MLMC}}) \leq C_2 \epsilon^{-2}(\log \epsilon)^2.
\]

**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 4.3) will be presented in detail for completeness.

\section{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
\[
N_\ell := 2^\ell, \quad h_\ell := \frac{T}{N_\ell}, \quad \ell \in \{0, \ldots, L\}.
\]

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
\[
Z_{i,2N,1,h} = \xi_i + \int_0^t b\left(Z_{\eta(r),1}^{i,2N,1,h}, \mu_{\eta(r)}\right) dr + \int_0^t \sigma\left(Z_{\eta(r),1}^{i,2N,1,h}, \mu_{\eta(r)}\right) dW_r, \quad 1 \leq i \leq N,
\]
\[
Z_{i,2N,2,h} = \xi_i + \int_0^t b\left(Z_{\eta(r),2}^{i,2N,2,h}, \mu_{\eta(r)}\right) dr + \int_0^t \sigma\left(Z_{\eta(r),2}^{i,2N,2,h}, \mu_{\eta(r)}\right) dW_r, \quad N + 1 \leq i \leq 2N,
\]
where
\[
\mu^Z_{r,2N,(1),h} := \frac{1}{N} \sum_{i=1}^{N} \delta_{\varphi_{i,2N,(1),h}} \quad \text{and} \quad \mu^Z_{r,2N,(2),h} := \frac{1}{\sum_{i=N+1}^{2N} \delta_{\varphi_{i,2N,(2),h}}}.
\]

Therefore, we define the MLMC estimator with time discretisation as
\[
\mathcal{A}^{\text{MLMC},t} := \frac{1}{M_0} \sum_{\theta=1}^{M_0} \Phi(\mu^Z_{T,h_0,0,(\theta)}(0)) + \sum_{\ell=1}^{L} \left[ \frac{1}{M_\ell} \sum_{\theta=1}^{M_\ell} \left[ \Phi(\mu^Z_{T,h_\ell,0,(\ell)}(0)) - \frac{1}{2} \left( \Phi(\mu^Z_{T,h_\ell,0,(1),2h_\ell,(\ell)}(0)) + \Phi(\mu^Z_{T,h_\ell,0,(2),2h_\ell,(\ell)) (0)) \right) \right] \right],
\]

(4.1)

where \(\mu^Z_{T,h_\ell,0,(\ell)}(0), \mu^Z_{T,h_\ell,0,(1),2h_\ell,(\ell)}(0), \mu^Z_{T,h_\ell,0,(2),2h_\ell,(\ell)) (0)\) are defined similarly as \(\mu^Z_{T,h_\ell,0,(\ell)}, \mu^Z_{T,h_\ell,0,(1),2h_\ell,(\ell)}, \mu^Z_{T,h_\ell,0,(2),2h_\ell,(\ell)) (0)\) respectively, but correspond to the \(\sum L \) independent clouds of particles indexed by \(\ell \in \{0, \ldots, L\}\) and \(\theta \in \{1, \ldots, M_\ell\}\). Each cloud (indexed by \(\ell, \theta\)) 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 \(i, \ell\) and \(\theta\).

To prove the analogue of Theorem 3.1 with time discretisation, we need the following lemma that provides a strong error bound between the particle system (1.5) and the Euler scheme (1.7). 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

\[
\sup_{N \in \mathbb{N}} \sup_{u \in [0,T]} \mathbb{E} \left[ \frac{1}{N} \sum_{i=1}^{N} |Y^{i,N}_u|^2 \right] < +\infty, \quad \sup_{N \in \mathbb{N}} \sup_{u \in [0,T]} \mathbb{E} \left[ \frac{1}{N} \sum_{i=1}^{N} |Z^{i,N,h}_u|^2 \right] < +\infty,
\]

(4.2)

for some \(C > 0\).

**Lemma 4.1.** Suppose that \(b \in \mathcal{M}_2(\mathbb{R}^d \times \mathcal{P}_2(\mathbb{R}^d))\) and \(\sigma\) is constant. Then

\[
\sup_{N \in \mathbb{N}} \sup_{u \in [0,T]} \mathbb{E} \left[ W^2(\mu^Y_s,\mu^Z_s)^2 \right] \leq C h^2,
\]

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,

\[
(Y^{i,N}_t - Z^{i,N,h}_t)^2 = 2 \int_0^t (Y^{i,N}_s - Z^{i,N,h}_s) \left( b(Y^{i,N}_s, \mu^Y_s) - b(Z^{i,N,h}_s, \mu^Z_s) \right) ds.
\]

Take \(0 \leq t' \leq t \leq T\). Then

\[
\frac{1}{N} \sum_{i=1}^{N} \mathbb{E} \left[ (Y^{i,N}_t - Z^{i,N,h}_t)^2 \right] = \frac{2}{N} \sum_{i=1}^{N} \mathbb{E} \left[ \int_0^{t'} (Y^{i,N}_s - Z^{i,N,h}_s) \left( b(Y^{i,N}_s, \mu^Y_s) - b(Z^{i,N,h}_s, \mu^Z_s) \right) ds \right]
\]

25
We first bound the first term of (4.3).

\[
\frac{2}{N} \sum_{i=1}^{N} \mathbb{E} \left[ \int_0^{t'} (Y_s^{i,N} - Z_s^{i,N,h}) \left( b(Y_s^{i,N}, \mu_s^{i,N}) - b(Z_s^{i,N,h}, \mu_s^{i,N,h}) \right) ds \right] \leq C \mathbb{E} \left[ \frac{1}{N} \sum_{i=1}^{N} \left( |Y_s^{i,N} - Z_s^{i,N,h}| + \left( \frac{1}{N} \sum_{j=1}^{N} |Y_s^{j,N} - Z_s^{j,N,h}|^2 \right)^{1/2} \right) ds \right]
\leq C \mathbb{E} \left[ \frac{1}{N} \sum_{i=1}^{N} |Y_u^{i,N} - Z_u^{i,N,h}|^2 \right] ds \leq C \int_0^{t'} \sup_{u \in [0,s]} \left[ \frac{1}{N} \sum_{i=1}^{N} \mathbb{E} |Y_u^{i,N} - Z_u^{i,N,h}|^2 \right] ds.
\]

To bound the second term of (4.3), we proceed as in the proof of Theorem B.3 by applying Itô's formula to the process

\[
\left\{ (Y_s^{i,N} - Z_s^{i,N,h})(b(Z_s^{i,N,h}, \mu_s^{i,N,h}) - b(Z_t^{i,N,h}, \mu_t^{i,N,h})) \right\}_{s \geq t_0},
\]

which gives

\[
(Y_s^{i,N} - Z_s^{i,N,h})(b(Z_s^{i,N,h}, \mu_s^{i,N,h}) - b(Z_t^{i,N,h}, \mu_t^{i,N,h})) = \int_{t_0}^{s} (b(Z_u^{i,N,h}, \mu_u^{i,N,h}) - b(Z_t^{i,N,h}, \mu_t^{i,N,h})) dY_u^{i,N} - Z_u^{i,N,h} = \int_{t_0}^{s} \left( \frac{1}{N} \partial_\mu b(Z_u^{i,N,h}, \mu_u^{i,N,h})(Z_u^{i,N,h}) \right) dZ_u^{i,N,h} + \frac{1}{2} \sum_{j \neq i} \int_{t_0}^{s} \left( Y_u^{i,N} - Z_u^{i,N,h} \right) \left( \frac{1}{N} \partial_\mu \partial_\mu b(Z_u^{i,N,h}, \mu_u^{i,N,h})(Z_u^{i,N,h}) \right) dZ_u^{i,N,h} + \frac{1}{2} \int_{t_0}^{s} \left( \frac{1}{N} \partial_\mu b(Z_u^{i,N,h}, \mu_u^{i,N,h})(Z_u^{i,N,h}) \right) d\left\langle Z_u^{i,N,h} \right\rangle + \frac{1}{N^2} \partial_\mu^2 b(Z_u^{i,N,h}, \mu_u^{i,N,h})(Z_u^{i,N,h}, Z_u^{i,N,h}) d\left\langle Z_u^{i,N,h} \right\rangle + \frac{1}{N^2} \partial_\mu^2 b(Z_u^{i,N,h}, \mu_u^{i,N,h})(Z_u^{i,N,h}, Z_u^{i,N,h}) + \frac{2}{N} \partial_\mu b(Z_u^{i,N,h}, \mu_u^{i,N,h})(Z_u^{i,N,h}) + \partial_\mu^2 b(Z_u^{i,N,h}, \mu_u^{i,N,h}) d\left\langle Z_u^{i,N,h} \right\rangle.
\]
Putting \( t_0 = \eta(s) \), taking average of \( i \) from 1 to \( N \), taking expectation and rewriting terms, we have

\[
\frac{1}{N} \sum_{i=1}^{N} E \left( \left( Y^{i,N}_{s} - Z^{i,N,h}_{s} \right) \left( b(Z^{i,N,h}_{s}, \mu^{Z,N,h}_{s}) - b(Z^{i,N,h}_{\eta(s)}, \mu^{Z,N,h}_{\eta(s)}) \right) \right) = I_1 + I_2,
\]

where

\[
I_1 := \frac{1}{N} \sum_{i=1}^{N} E \left[ \int_{\eta(s)}^{s} \left( b(Z^{i,N,h}_{u}, \mu^{Z,N,h}_{s})(Z^{j,N,h}_{u} - Z^{j,N,h}_{\eta(s)}) \right) \left( b(Y^{i,N}_{u}, \mu^{Y,N}_{s}) - b(Y^{i,N}_{\eta(s)}, \mu^{Y,N}_{\eta(s)}) \right) du \right]
\]

and

\[
I_2 := \frac{1}{N} \sum_{i=1}^{N} E \left[ \int_{\eta(s)}^{s} \left( Y^{i,N}_{u} - Z^{i,N,h}_{u} \right) D^{i}_{u} du \right],
\]

where

\[
D^{i}_{u} := \frac{1}{N} \sum_{j=1}^{N} \left( \partial_{\mu} b(Z^{i,N,h}_{u}, \mu^{Z,N,h}_{s})(Z^{j,N,h}_{u} - Z^{j,N,h}_{\eta(s)}) \right) + \partial_{\mu}^2 b(Z^{i,N,h}_{u}, \mu^{Z,N,h}_{s})(Z^{i,N,h}_{u} - Z^{i,N,h}_{\eta(s)})
\]
\[
+ \frac{1}{2}\sigma^2 \sum_{j=1}^{N} \left( \frac{1}{N^2} \partial_{\mu}^2 b(Z^{i,N,h}_{u}, \mu^{Z,N,h}_{s})(Z^{j,N,h}_{u} - Z^{j,N,h}_{\eta(s)}) + \frac{1}{N} \partial_{\mu} \partial_{\mu} b(Z^{i,N,h}_{u}, \mu^{Z,N,h}_{s})(Z^{j,N,h}_{u} - Z^{j,N,h}_{\eta(s)}) \right)
\]
\[
+ \frac{1}{2}\sigma^2 \left( \frac{2}{N} \partial_{\mu} \partial_{\mu} b(Z^{i,N,h}_{u}, \mu^{Z,N,h}_{s})(Z^{i,N,h}_{u} - Z^{i,N,h}_{\eta(s)}) + \partial_{\mu}^2 b(Z^{i,N,h}_{u}, \mu^{Z,N,h}_{s}) \right).
\]

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

\[
\frac{1}{N} \sum_{i=1}^{N} E|D^{i}_{u}|^2 \leq C \left( 1 + \frac{1}{N} \sum_{i=1}^{N} E|Z^{i,N,h}_{\eta(u)}|^2 \right).
\]

Then, by (4.2),

\[
\sup_{u \in [0,T]} \left[ \frac{1}{N} \sum_{i=1}^{N} E|D^{i}_{u}|^2 \right] \leq C.
\]

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

\[
I_2 \leq \int_{\eta(s)}^{s} \left( \frac{1}{N} \sum_{i=1}^{N} E|Y^{i,N}_{u} - Z^{i,N,h}_{u}|^2 \right)^{1/2} \left( \frac{1}{N} \sum_{i=1}^{N} E|D^{i}_{u}|^2 \right)^{1/2} du
\]
\[
\leq C \left( \sup_{u \in [0,s]} \frac{1}{N} \sum_{i=1}^{N} E|Y^{i,N}_{u} - Z^{i,N,h}_{u}|^2 \right)^{1/2} h
\]
It is clear that

\[ \sum_{i=1}^{N} \mathbb{E} \left[ \int_{\eta(s)}^{s} \left( b(Z_{u}^{i,N,h}, \mu_{u}^{Z,N,h}) - b(Z_{\eta(s)}^{i,N,h}, \mu_{\eta(s)}^{Z,N,h}) \right) (b(Y_{u}^{i,N}, \mu_{Y}^{i,N}) - b(Z_{u}^{i,N,h}, \mu_{u}^{Z,N,h})) \, du \right] \leq C \left( \frac{1}{2} \sup_{u \in [0,T]} \frac{1}{N} \sum_{i=1}^{N} \mathbb{E} |Y_{u}^{i,N} - Z_{u}^{i,N,h}|^2 + \frac{1}{2} h^2 \right). \]  

(4.5)

Next, we rewrite \( I_1 \) as

\[ I_1 = \frac{1}{N} \sum_{i=1}^{N} \mathbb{E} \left[ \int_{\eta(s)}^{s} \left( b(Z_{u}^{i,N,h}, \mu_{u}^{Z,N,h}) - b(Z_{\eta(s)}^{i,N,h}, \mu_{\eta(s)}^{Z,N,h}) \right) (b(Y_{u}^{i,N}, \mu_{Y}^{i,N}) - b(Z_{u}^{i,N,h}, \mu_{u}^{Z,N,h})) \, du \right] \]

\[ + \frac{1}{N} \sum_{i=1}^{N} \mathbb{E} \left[ \int_{\eta(s)}^{s} \left( b(Z_{u}^{i,N,h}, \mu_{u}^{Z,N,h}) - b(Z_{\eta(s)}^{i,N,h}, \mu_{\eta(s)}^{Z,N,h}) \right)^2 \, du \right]. \]

It is clear that

\[ \frac{1}{N} \sum_{i=1}^{N} \mathbb{E} \left[ \int_{\eta(s)}^{s} \left( b(Z_{u}^{i,N,h}, \mu_{u}^{Z,N,h}) - b(Z_{\eta(s)}^{i,N,h}, \mu_{\eta(s)}^{Z,N,h}) \right)^2 \, du \right] \leq C h^2. \]

(4.6)

By the Cauchy-Schwarz inequality and (Lip), the first term of \( I_1 \) is bounded by

\[ \frac{1}{N} \sum_{i=1}^{N} \mathbb{E} \left[ \int_{\eta(s)}^{s} \left( \mathbb{E} \left| b(Z_{u}^{i,N,h}, \mu_{u}^{Z,N,h}) - b(Z_{\eta(s)}^{i,N,h}, \mu_{\eta(s)}^{Z,N,h}) \right|^2 \right)^{1/2} \right] \]

\[ \leq \frac{1}{N} \sum_{i=1}^{N} \mathbb{E} \left[ \int_{\eta(s)}^{s} \left( \mathbb{E} \left| b(Y_{u}^{i,N}, \mu_{Y}^{i,N}) - b(Z_{u}^{i,N,h}, \mu_{u}^{Z,N,h}) \right|^2 \right)^{1/2} \right] \] \[ \leq \frac{1}{N} \sum_{i=1}^{N} C \sqrt{h} \int_{\eta(s)}^{s} \left( \mathbb{E} \left| Y_{u}^{i,N} - Z_{u}^{i,N,h} \right|^2 \right)^{1/2} \, du \] \[ \leq \frac{1}{N} \sum_{i=1}^{N} C \sqrt{h} \int_{\eta(s)}^{s} \left( \mathbb{E} \left| Y_{u}^{i,N} - Z_{u}^{i,N,h} \right|^2 \right)^{1/2} \, du \] \[ \leq \frac{2}{N} \sum_{i=1}^{N} C \frac{h}{2} \int_{\eta(s)}^{s} \left( \mathbb{E} \left| Y_{u}^{i,N} - Z_{u}^{i,N,h} \right|^2 \right)^{1/2} \, du \]

\[ \leq 2 C \left( \sup_{u \in [0,T]} \mathbb{E} \left( Y_{u}^{i,N} - Z_{u}^{i,N,h} \right)^2 \right)^{1/2} \]

\[ \leq C \left( h^3 + \sup_{u \in [0,T]} \mathbb{E} \left( Y_{u}^{i,N} - Z_{u}^{i,N,h} \right)^2 \right). \]

(4.7)

A combination of (4.3), (4.4), (4.5), (4.6) and (4.7) gives

\[ \sup_{u \in [0,T]} \left[ \frac{1}{N} \sum_{i=1}^{N} \mathbb{E} \left| Y_{u}^{i,N} - Z_{u}^{i,N,h} \right|^2 \right] \leq C \left( \int_{0}^{t} \sup_{u \in [0,s]} \left[ \frac{1}{N} \sum_{i=1}^{N} \mathbb{E} \left| Y_{u}^{i,N} - Z_{u}^{i,N,h} \right|^2 \right] \, ds + h^2 \right), \quad \forall t \in [0,T], \]
which implies by Gronwall’s inequality that
\[
\sup_{u \in [0,T]} \left[ \frac{1}{N} \sum_{i=1}^{N} \mathbb{E}(Y_{u}^{i,N} - Z_{u}^{i,N,h})^2 \right] \leq C h^2.
\]
Since the constant \( C \) does not depend on \( N \), we conclude that
\[
\sup_{N \in \mathbb{N}} \sup_{s \in [0,T]} \mathbb{E}\left[ W_2(\mu_{s}^{Y,N}, \mu_{s}^{Z,N,h})^2 \right] \leq \sup_{N \in \mathbb{N}} \sup_{s \in [0,T]} \left[ \frac{1}{N} \sum_{i=1}^{N} \mathbb{E}(Y_{u}^{i,N} - Z_{u}^{i,N,h})^2 \right] \leq C h^2.
\]

A combination of Lemma 4.1 and Theorem 3.1 immediately gives the following result.

**Theorem 4.2** (Variance of antithetic difference). Assume (Int). Suppose that \( b \in \mathcal{M}_1(\mathbb{R}^d \times \mathcal{P}_2(\mathbb{R}^d)) \) and \( \Phi \in \mathcal{M}_1(\mathcal{P}_2(\mathbb{R}^d)) \). Moreover, suppose that \( \sigma \) is constant. Then
\[
\text{Var}\left[ \Phi(\mu_T^{Z,N,h}) - \frac{1}{2}(\Phi(\mu_T^{Z,N,(1),2h}) + \Phi(\mu_T^{Z,N,(2),2h})) \right] 
\leq \mathbb{E}\left[ \Phi(\mu_T^{Z,N,h}) - \frac{1}{2}(\Phi(\mu_T^{Z,N,(1),2h}) + \Phi(\mu_T^{Z,N,(2),2h})) \right]^2 \leq C\left( \frac{1}{N^2} + h^2 \right),
\]
where \( C \) is a constant that depends on \( \Phi \), \( b \), \( \sigma \) and \( T \), but does not depend on \( N \) or \( h \).

As before, we perform an analysis on the complexity of this algorithm. 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 4.2, we have
\[
\text{Var}\left[ \Phi(\mu_T^{Z,N_\ell,h_\ell,(\theta),(\ell)}) - \frac{1}{2}(\Phi(\mu_T^{Z,N_\ell,(1),2h_\ell,(\theta),(\ell)}) + \Phi(\mu_T^{Z,N_\ell,(2),2h_\ell,(\theta),(\ell)}) \right] \leq \frac{C}{N_\ell^2}. \tag{II}
\]
Since simulating particle systems of \( N \) particles with \( p \) timesteps requires \( N^2p \) operations in general, the cost function of the antithetic difference is bounded by
\[
\text{Cost}\left[ \Phi(\mu_T^{Z,N_\ell,h_\ell,(\theta),(\ell)}) - \frac{1}{2}(\Phi(\mu_T^{Z,N_\ell,(1),2h_\ell,(\theta),(\ell)}) + \Phi(\mu_T^{Z,N_\ell,(2),2h_\ell,(\theta),(\ell)}) \right] \leq CN_\ell^3. \tag{III}
\]

**Theorem 4.3** (Complexity of antithetic MLMC with time discretisation). Assume (Int). Suppose that \( b \in \mathcal{M}_1(\mathbb{R}^d \times \mathcal{P}_2(\mathbb{R}^d)) \) and \( \Phi \in \mathcal{M}_1(\mathcal{P}_2(\mathbb{R}^d)) \). Moreover, suppose that \( \sigma \) 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 root-mean-square error of \( \mathcal{A}_A^{\text{MLMC},\ell} \) is bounded by
\[
\left( \mathbb{E}\left[ (\mathcal{A}_A^{\text{MLMC},\ell} - \Phi(\mu_X^Y))^2 \right] \right)^{1/2} \leq C_1 \epsilon
\]
and the computational cost of \( \mathcal{A}_A^{\text{MLMC},\ell} \) is bounded by
\[
\text{Cost}(\mathcal{A}_A^{\text{MLMC},\ell}) \leq C_2 \epsilon^{-3}.
\]
\textbf{Proof.} As in Theorem 3.2, 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
\[
L := \lceil \log_2(\sqrt{2}\epsilon^{-1}) \rceil, \quad M_\ell := \lceil 2\epsilon^{-2}2^{L/2}(1-2^{-1/2})^{-1}2^{-5\ell/2} \rceil, \quad \ell \in \{0, \ldots, L\}.
\]
The mean-square error decomposes as
\[
\text{Mean-square error} = \text{Var}(A^{A^-\text{MLMC},t}) + (E(A^{A^-\text{MLMC},t}) - \Phi(\mu_T^X))^2.
\]
By the choice of \( L \), \( 2^{-L} \leq \frac{\epsilon}{\sqrt{2}} \). Therefore, by Property (I),
\[
|E(A^{A^-\text{MLMC},t}) - \Phi(\mu_T^X)|^2 = |E\{\Phi(\mu_T^{Z, N_L, h_L})\} - \Phi(\mu_T^X)|^2 \leq \left( \frac{C}{N_L} \right)^2 = (C2^{-L})^2 \leq C^2\left( \frac{\epsilon^2}{2} \right). \tag{4.8}
\]
On the other hand, by Property (II) and the choice of \( \{M_\ell\}_{\ell=0}^L \),
\[
\text{Var}(A^{A^-\text{MLMC},t}) \leq \sum_{\ell=0}^L \frac{1}{M_\ell^2} \left[ \sum_{\theta=1}^{M_\ell} \frac{C}{N_L^2} \right] \leq \sum_{\ell=0}^L \frac{C}{M_\ell} 2^{-2\ell} \leq \sum_{\ell=0}^L C2^{-2\ell}(2^{-1}\epsilon^22^{-L/2}(1-2^{-1/2})2^{5\ell/2})
\]
\[
= C2^{-1}\epsilon^22^{-L/2}(1-2^{-1/2})\sum_{\ell=0}^L 2^{\ell/2}
\]
\[
< \frac{1}{2}C\epsilon^2.
\]
This verifies that the mean-square error is bounded by \( \frac{1}{2}(C^2 + C)\epsilon^2 \). Next, we note that
\[
M_\ell \leq 2\epsilon^{-2}2^{L/2}(1-2^{-1/2})^{-1}2^{-5\ell/2} + 1
\]
and hence, by Property (III),
\[
\text{Cost}(A^{A^-\text{MLMC},t}) \leq C\left( \sum_{\ell=0}^L 2\epsilon^{-2}2^{L/2}(1-2^{-1/2})^{-1}2^{-5\ell/2}2^{3\ell} + \sum_{\ell=0}^L 2^{3\ell} \right). \tag{4.9}
\]
Note that the choice of \( L \) implies that \( 2^L \leq 2\sqrt{2}\epsilon^{-1} \).
\[
\sum_{\ell=0}^L 2\epsilon^{-2}2^{L/2}(1-2^{-1/2})^{-1}2^{-5\ell/2}2^{3\ell} = 2\epsilon^{-2}2^{L/2}(1-2^{-1/2})^{-1} \sum_{\ell=0}^L 2^{\ell/2}
\]
\[
< 2\epsilon^{-2}2^{L/2}(1-2^{-1/2})^{-1}\left(2^{L/2}(1-2^{-1/2})^{-1}\right)
\]
\[
= 2\epsilon^{-2}2^L(1-2^{-1/2})^{-2}
\]
\[
< 4\sqrt{2}(1-2^{-1/2})^{-2}\epsilon^{-3}. \tag{4.10}
\]
Similarly,
\[
\sum_{\ell=0}^{L} 2^{3\ell} \leq \frac{2^{3L}}{1-2^{-3}} \leq \frac{(2\sqrt{2})^3}{1-2^{-3}} \epsilon^{-3}. \tag{4.11}
\]

A combination of (4.9), (4.10) and (4.11) finally gives
\[
\text{Cost}(A_{\text{MLMC}}) \leq C \left( 4\sqrt{2}(1 - 2^{-1/2})^{-2} + \frac{(2\sqrt{2})^3}{1-2^{-3}} \right) \epsilon^{-3}.
\]

\[\square\]

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 \( K \subset \mathcal{P}_2(\mathbb{R}^d), \ y \mapsto \frac{\delta U}{\delta m}(m, y) \) has at most quadratic growth in \( y \) uniformly in \( m \in K \),
- 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 \( K \subset \mathcal{P}_2(\mathbb{R}^d), (y, y') \mapsto \frac{\delta^p U}{\delta m^p}(m, y, y') \) has at most quadratic growth in \( (y, y') \) uniformly in \( m \in K \),
- for any \( m, m' \in \mathcal{P}_2(\mathbb{R}^d) \),
\[
\frac{\delta^{p-1} U}{\delta m^{p-1}}(m', y) - \frac{\delta^{p-1} U}{\delta m^{p-1}}(m, y) = \int_0^1 \int_{\mathbb{R}^d} \frac{\delta^p U}{\delta m^p}((1-s)m + sm', y, y') (m' - m)(dy') \, ds,
\]
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
\[
\frac{\delta^p U}{\delta m^p}(m, y_1, \ldots, y_p) = 0, \quad \text{if } y_i = 0 \text{ for some } i \in \{1, \ldots, p\}.
\] (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 \(C^1\), 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 [31]. 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 \(\partial_\mu\) and the derivation in the non-measure arguments. Namely, at order 2, one considers
\[
P_2(\mathbb{R}^d) \times \mathbb{R}^d \ni (m, y) \mapsto \partial_\mu \partial_\mu U(m, y) \quad \text{and} \quad P_2(\mathbb{R}^d) \times \mathbb{R}^d \times \mathbb{R}^d \ni (m, y, y') \mapsto \partial_\mu^2 U(m, y, y').
\]

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 \(k \geq 2\), \((i_1, \ldots, i_k) \in \{1, \ldots, d\}^k\) and \(x_1, \ldots, x_k \in \mathbb{R}^d\), the function \(\partial_\mu^k f: P_2(\mathbb{R}^d) \times (\mathbb{R}^d)^k \to (\mathbb{R}^d)^{\otimes k}\) is defined by
\[
\left(\partial_\mu^k f(\mu, x_1, \ldots, x_k)\right)_{(i_1, \ldots, i_k)} := \left(\partial_\mu \left(\partial_\mu^{k-1} f(\cdot, x_1, \ldots, x_{k-1})\right)_{(i_1, \ldots, i_{k-1})}\right)_{i_k}(\mu, x_k),
\] (A.3)

and its corresponding mixed derivatives in space \(\partial_{v_1}^{\ell_1} \ldots \partial_{v_k}^{\ell_k} \partial_\mu^k f: P_2(\mathbb{R}^d) \times (\mathbb{R}^d)^k \to (\mathbb{R}^d)^{\otimes (k + \ell_1 + \ldots + \ell_k)}\) are defined by
\[
\left(\partial_{v_1}^{\ell_1} \ldots \partial_{v_k}^{\ell_k} \partial_\mu^k f(\mu, x_1, \ldots, x_k)\right)_{(i_1, \ldots, i_k)} := \frac{\partial^{\ell_k}}{\partial x_k^{\ell_k}} \ldots \frac{\partial^{\ell_1}}{\partial x_1^{\ell_1}} \left[\left(\partial_\mu^k f(\mu, x_1, \ldots, x_k)\right)_{(i_1, \ldots, i_k)}\right], \quad \ell_1 \ldots \ell_k \in \mathbb{N} \cup \{0\}.
\] (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].

32
Definition A.2 (Multi-index notation). Let $n, \ell$ be non-negative integers. Also, let $\beta = (\beta_1, \ldots, \beta_n)$ be an $n$-dimensional vector of non-negative integers. Then we call any ordered tuple of the form $(n, \ell, \beta)$ or $(n, \beta)$ a multi-index. For a function $f : \mathbb{R}^d \times \mathcal{P}_2(\mathbb{R}^d) \to \mathbb{R}$, the derivative $D^{(n,\ell,\beta)}f(x, \mu, v_1, \ldots, v_n)$ is defined as

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

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

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

if this derivative is well-defined. Finally, we also define the order $3 \max |(n, \ell, \beta)|$ (resp. $|n, \beta|$ ) by

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

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 $k$th order differentiable functions).**

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

(a) $$|D^{(n,\ell,\beta)}b(x, \mu, v_1, \ldots, v_n)| \leq C, \quad |D^{(n,\ell,\beta)}\sigma(x, \mu, v_1, \ldots, v_n)| \leq C, \tag{A.6}$$

(b) $$|D^{(n,\ell,\beta)}b(x, \mu, v_1, \ldots, v_n) - D^{(n,\ell,\beta)}b(x', \mu', v_1, \ldots, v_n')| \leq C \left( |x - x'| + \sum_{i=1}^{n} |v_i - v_i'| + W_2(\mu, \mu') \right),$$

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

for any $x, x', v_1, v_1', \ldots, v_n, v_n' \in \mathbb{R}^d$ 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 $D^{(n,\beta)}\Phi(\mu, v_1, \ldots, v_n)$ exists for every multi-index $(n, \beta)$ such that $|(n, \beta)| \leq k$ and

3 We do not consider ‘zeroth’ order derivatives in our definition, i.e. at least one of $n, \beta_1, \ldots, \beta_n$ and $\ell$ must be non-zero, for every multi-index $(n, \ell, (\beta_1, \ldots, \beta_n))$.
(a) \[ |D^{(n,\beta)}\Phi(\mu, v_1, \ldots, v_n)| \leq C, \] \hspace{1cm} \text{(A.8)}

(b) \[ |D^{(n,\beta)}\Phi(\mu, v_1, \ldots, v_n) - D^{(n,\beta)}\Phi(\mu', v_1', \ldots, v_n')| \leq C \left( \sum_{i=1}^{n} |v_i - v_i'| + W_2(\mu, \mu') \right), \] \hspace{1cm} \text{(A.9)}

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

(iii) A function \( \Phi : [0, T] \times \mathcal{P}_2(\mathbb{R}^d) \to \mathbb{R} \) is said to be in \( \mathcal{M}_k([0, T] \times \mathcal{P}_2(\mathbb{R}^d)) \), if \( \Phi(\cdot, \mu) \) is in \( C^1([0, T]) \), for each \( \mu \in \mathcal{P}_2(\mathbb{R}^d) \) and \( \Phi(s, \cdot) \in \mathcal{M}_k(\mathcal{P}_2(\mathbb{R}^d)) \), for each \( s \in [0, T] \), where the \( L^\infty \) and Lipschitz bounds of the derivatives of \( \Phi(s, \cdot) \) 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. \cite{7} for the correspondence up to order 2,

\[ \partial^m \mu U(\cdot) = \partial_{y_n} \frac{\delta}{\delta m} \ldots \partial_{y_1} \frac{\delta}{\delta m} U(\cdot) = \partial_{y_n} \ldots \partial_{y_1} \frac{\delta^n}{\delta m^n} U(\cdot), \] \hspace{1cm} \text{(A.10)}

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

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

\[ \left| \frac{\delta^k U}{\delta m^k}(m, y_1, \ldots, y_k) \right| \leq C \left( |y_1|^k + \ldots |y_k|^k \right), \] \hspace{1cm} \text{(A.11)}

for some constant \( C > 0 \) that does not depend on \( m \) and \( y_1, \ldots, y_k \).

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

\[ \left| \frac{\delta^k U}{\delta m^k}(m, y_1, \ldots, y_k) \right| \leq \frac{\sqrt{d}}{k} \| \partial^k \mu \|_{\infty}(|y_1|^k + \ldots |y_k|^k). \] \hspace{1cm} \text{(A.12)}

Consequently, \( U \in \mathcal{M}^L_k(\mathcal{P}_2(\mathbb{R}^d)) \).
Appendix: Weak error analysis

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

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

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 \( \eta \), we define

\[
X_{s,\eta}^t = \eta + \int_s^t b(X_r^{s,\eta}, \mathcal{L}(X_r^{s,\eta})) \, dr + \int_s^t \sigma(X_r^{s,\eta}, \mathcal{L}(X_r^{s,\eta})) \, dW_r, \quad t \in [s, T].
\]

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

\[
\mathcal{V}(s, \mathcal{L}(\eta)) = \Phi(\mathcal{L}(X_s^{s,\eta})),
\]

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

\[
\begin{aligned}
\frac{\partial}{\partial s} \mathcal{V}(s, \mu) &+ \int_{\mathbb{R}^d} \left[ \partial_x \mathcal{V}(s, \mu)(x)b(x, \mu) + \frac{1}{2} \text{Tr}(\partial_x \partial_x \mathcal{V}(s, \mu)(x)a(x, \mu)) \right] \mu(dx) = 0, \quad s \in (0, T), \\
\mathcal{V}(T, \mu) &\equiv \Phi(\mu),
\end{aligned}
\]

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

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

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

\[
\mathcal{V}(s, x_1, \ldots, x_N) := \mathcal{V}
\left(s, \frac{1}{N} \sum_{i=1}^N \delta_{x_i} \right).
\]

Proposition 3.1 of [10] allows us to conclude that \( \mathcal{V} \) 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 \( \mathcal{V} \).

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

\[
\mathcal{V}(s, \mathcal{L}(X_s^{0,\xi})) = \Phi(\mathcal{L}(X_s^{0,\xi})) = \Phi(\mathcal{L}(X_T^{0,\xi})).
\]
Hence, this function is constant in time \( s \in [0, T] \). In particular, by the terminal condition, we have

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

By the terminal condition for the PDE, we notice that

\[
\Phi(\mu_Y^N, T) = \mathcal{V}(T, \mu_Y^N).
\]

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

\[
\Phi(\mu_Y^N, T) - \Phi(\mu_X^T) = \mathcal{V}(T, \mu_Y^N) - \mathcal{V}(0, \nu) = \left( \mathcal{V}(T, \mu_Y^N) - \mathcal{V}(0, \mu_0^N) \right) + \left( \mathcal{V}(0, \mu_0^N) - \mathcal{V}(0, \nu) \right). \tag{B.5}
\]

This decomposition enables us to prove the following result.

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

\[
\left| \mathbb{E}[\Phi(\mu_Y^N, T)] - \Phi(\mu_X^T) \right| \leq \frac{C}{N}, \tag{B.6}
\]

where \( C \) is a constant that depends on \( \Phi, b, \sigma \) 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 \( \sigma \), the standard Itô’s formula is applicable to \( V \) by Proposition 3.1 of [10]. Let \( x = (x_1, \ldots, x_N) \). Moreover, we know from this theorem that

\[
\frac{\partial V}{\partial x_i}(s, x) = \frac{1}{N} \partial_x \mathcal{V} \left( s, \frac{1}{N} \sum_{j=1}^N \delta_{x_j} \right)(x_i)
\]

and

\[
\frac{\partial^2 V}{\partial x_i^2}(s, x) = \frac{1}{N} \partial_x \partial_x \mathcal{V} \left( s, \frac{1}{N} \sum_{j=1}^N \delta_{x_j} \right)(x_i) + \frac{1}{N^2} \partial_x^2 \mathcal{V} \left( s, \frac{1}{N} \sum_{j=1}^N \delta_{x_j} \right)(x_i, x_i).
\]
for any \( s \in [0, T] \), \( x_1, \ldots, x_N \in \mathbb{R}^d \). Let \( Y^N := (Y^{1,N}, \ldots, Y^{N,N}) \). Then

\[
\mathcal{V}(T, \mu_Y^{Y,N}) - \mathcal{V}(0, \mu_Y^{Y,N}) = \mathcal{V}(T, \mu_Y^{Y,N}) - V(0, \mu_Y^Y)
\]

\[
= \left[ \int_0^T \frac{\partial \mathcal{V}}{\partial s}(s, \mu_Y^{Y,N}) + \sum_{i=1}^N \frac{\partial \mathcal{V}}{\partial x_i}(s, \mu_Y^{Y,N}) b(Y^{i,N}_s, \mu_s^{Y,N}) + \frac{1}{2} \text{Tr} \left( a(Y^{i,N}_s, \mu_s^{Y,N}) \sum_{i=1}^N \frac{\partial^2 \mathcal{V}}{\partial x_i^2}(s, \mu_Y^{Y,N}) \right) \right] \, ds
\]

\[+
\sum_{i=1}^N \int_0^T \sigma(Y^{i,N}_s, \mu_s^{Y,N})^T \frac{\partial \mathcal{V}}{\partial x_i}(s, \mu_Y^{Y,N}) \, dW_i^s
\]

\[=
\int_0^T \partial_s \mathcal{V}(s, \mu_Y^{Y,N}) + \sum_{i=1}^N \left[ \frac{1}{N} \partial \mathcal{V}(s, \mu_s^{Y,N})(Y^{i,N}_s) b(Y^{i,N}_s, \mu_s^{Y,N})
\]

\[+
\frac{1}{2} \text{Tr} \left( a(Y^{i,N}_s, \mu_s^{Y,N}) \left( \frac{1}{N} \partial \mathcal{V}(s, \mu_s^{Y,N})(Y^{i,N}_s) + \frac{1}{N^2} \partial^2 \mathcal{V}(s, \mu_s^{Y,N})(Y^{i,N}_s, Y^{i,N}_s) \right) \right) \right] \, ds
\]

\[+
\frac{1}{N} \sum_{i=1}^N \int_0^T \sigma(Y^{i,N}_s, \mu_s^{Y,N})^T \partial \mathcal{V}(s, \mu_s^{Y,N})(Y^{i,N}_s) \, dW_i^s.
\]

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

\[
\Phi(\mu_T^{Y,N}) - \Phi(\mu_T^X) = \left( \mathcal{V}(0, \mu_0^{Y,N}) - \mathcal{V}(0, \nu) \right)
\]

\[+
\int_0^T \frac{1}{2} \left[ \frac{1}{N^2} \sum_{i=1}^N \text{Tr} \left( a(Y^{i,N}_s, \mu_s^{Y,N}) \partial^2 \mathcal{V}(s, \mu_s^{Y,N})(Y^{i,N}_s, Y^{i,N}_s) \right) \right] \, ds
\]

\[+
\frac{1}{N} \sum_{i=1}^N \int_0^T \sigma(Y^{i,N}_s, \mu_s^{Y,N})^T \partial \mathcal{V}(s, \mu_s^{Y,N})(Y^{i,N}_s) \, dW_i^s.
\]

(B.7)

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

\[|\mathbb{E}(\mathcal{V}(0, \mu_0^{Y,N}) - \mathcal{V}(0, \nu))| \leq \frac{C}{N}.
\]

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

The next theorem concerns the weak error between (1.4) and (1.7).

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

\[|\mathbb{E}(\Phi(\mu_T^{Z,N,h})) - \Phi(\mu_T^X)| \leq C \left( \frac{1}{N} + h \right),
\]

(B.8)

where \( C \) is a constant that depends on \( \Phi, b, \sigma \) 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 \( Z^{N,h} := (Z^{1,N,h}, \ldots , Z^{N,N,h}) \). As before, by Lemma 2.5 and Theorem 2.11 from [11],

\[
\left| \mathbb{E}(V(0, \mu^0Z^{N,h}) - V(0, \nu)) \right| \leq \frac{C}{N^2}
\]

Next, by the previous analysis, we observe that

\[
\begin{align*}
&\left( \Phi(\mu^Z_{T, N,h}) - \Phi(\mu^Z_{T}) \right) - \left( V(0, \mu^0Z^{N,h}) - V(0, \mu^0Z^{N,h}) \right) \\
= &\left( V(T, \mu^Z_{T, N,h}) - V(0, \mu^0Z^{N,h}) \right) \\
= &\left[ \int_0^T \frac{\partial V}{\partial s}(s, Z^N_s, h) + \sum_{i=1}^N \frac{\partial V}{\partial x_i}(s, Z^N_s, h)b(Z^i_{N, h}, Z^N_s, h) + \frac{1}{2} \text{Tr} \left( a(Z^i_{N, h}, Z^N_s, h) + \frac{1}{N} \sum_{i=1}^N \frac{\partial^2 V}{\partial x_i^2}(s, Z^N_s, h) \right) \right] \\
&\int_0^T \int_0^T \frac{\partial V}{\partial x_i}(s, Z^N_s, h)T \sigma(Z^i_{N, h}, Z^N_s, h)dW_s \\
\end{align*}
\]

\[
\begin{align*}
&\left[ \int_0^T \frac{\partial V}{\partial s}(s, Z^N_s, h) + \sum_{i=1}^N \frac{\partial V}{\partial x_i}(s, Z^N_s, h)b(Z^i_{N, h}, Z^N_s, h) + \frac{1}{2} \text{Tr} \left( a(Z^i_{N, h}, Z^N_s, h) + \frac{1}{N} \sum_{i=1}^N \frac{\partial^2 V}{\partial x_i^2}(s, Z^N_s, h) \right) \right] ds \\
&\int_0^T \int_0^T \frac{\partial V}{\partial x_i}(s, Z^N_s, h)T \sigma(Z^i_{N, h}, Z^N_s, h)dW_s \\
\end{align*}
\]

\[
\begin{align*}
&\left[ \int_0^T \frac{\partial V}{\partial s}(s, Z^N_s, h) + \sum_{i=1}^N \frac{\partial V}{\partial x_i}(s, Z^N_s, h)b(Z^i_{N, h}, Z^N_s, h) + \frac{1}{2} \text{Tr} \left( a(Z^i_{N, h}, Z^N_s, h) + \frac{1}{N} \sum_{i=1}^N \frac{\partial^2 V}{\partial x_i^2}(s, Z^N_s, h) \right) \right] ds \\
&\int_0^T \int_0^T \frac{\partial V}{\partial x_i}(s, Z^N_s, h)T \sigma(Z^i_{N, h}, Z^N_s, h)dW_s \\
\end{align*}
\]

\[
\begin{align*}
&\left[ \int_0^T \frac{\partial V}{\partial s}(s, Z^N_s, h) + \sum_{i=1}^N \frac{\partial V}{\partial x_i}(s, Z^N_s, h)b(Z^i_{N, h}, Z^N_s, h) + \frac{1}{2} \text{Tr} \left( a(Z^i_{N, h}, Z^N_s, h) + \frac{1}{N} \sum_{i=1}^N \frac{\partial^2 V}{\partial x_i^2}(s, Z^N_s, h) \right) \right] ds \\
&\int_0^T \int_0^T \frac{\partial V}{\partial x_i}(s, Z^N_s, h)T \sigma(Z^i_{N, h}, Z^N_s, h)dW_s \\
\end{align*}
\]

\[
\begin{align*}
&\left[ \int_0^T \frac{\partial V}{\partial s}(s, Z^N_s, h) + \sum_{i=1}^N \frac{\partial V}{\partial x_i}(s, Z^N_s, h)b(Z^i_{N, h}, Z^N_s, h) + \frac{1}{2} \text{Tr} \left( a(Z^i_{N, h}, Z^N_s, h) + \frac{1}{N} \sum_{i=1}^N \frac{\partial^2 V}{\partial x_i^2}(s, Z^N_s, h) \right) \right] ds \\
&\int_0^T \int_0^T \frac{\partial V}{\partial x_i}(s, Z^N_s, h)T \sigma(Z^i_{N, h}, Z^N_s, h)dW_s \\
\end{align*}
\]

\[38\]
\[ \begin{align*}
+ \frac{1}{2} \text{Tr} \left( a(Z^{i,N,h}_{\eta(s)}; \mu_{\eta(s)}, \mu_s^{Z,N,h}) - a(Z^{i,N,h}_{\eta(s)}; \mu_{\eta(s)}, \mu_s^{Z,N,h}) \right) \frac{1}{N} (\partial_t \partial_s \mathcal{V}(s, \mu_s^{Z,N,h})(Z^{i,N,h}_s) \\
- \partial_t \partial_s \mathcal{V}(s, \mu_{\eta(s)})(Z^{i,N,h}_{\eta(s)}))\right) \\
+ \frac{1}{2} \text{Tr} \left( a(Z^{i,N,h}_{\eta(s)}; \mu_{\eta(s)}), \frac{1}{N^2} \partial^2_s \mathcal{V}(s, \mu_s^{Z,N,h})(Z^{i,N,h}_{s}, Z^{i,N,h}_{s}) \right) ds \\
+ \int_0^T \sum_{j=1}^N \frac{1}{N} \partial_s \mathcal{V}(s, \mu_s^{Z,N,h})(Z^{i,N,h}_s) T \sigma(Z^{i,N,h}_{\eta(s)}; \mu_{\eta(s)}) dW^i_s. \tag{B.9}
\end{align*} \]

Let \( \{\mathcal{F}_t\}_{t \in [0,T]} \) be the filtration generated by \( W^1, \ldots, W^N \). Then, by the Itô’s formula, for each \( k \in \{1, \ldots, d\} \),

\[ \begin{align*}
\mathbb{E}\left[ b_k(Z^{i,N,h}_{\eta(s)}; \mu_{\eta(s)}, \mu_s^{Z,N,h}) - b_k(Z^{i,N,h}_{\eta(s)}; \mu_{\eta(s)}, \mu_s^{Z,N,h}) \big| \mathcal{F}_{\eta(s)} \right] \\
= -\mathbb{E} \left[ \int_{\eta(s)}^t \left( \partial_s b_k(Z^{i,N,h}_r, \mu_r^{Z,N,h}) + \frac{1}{N} \partial_s b_k(Z^{i,N,h}_r, \mu_r^{Z,N,h})(Z^{i,N,h}_r) \right) dW^i_r \\
+ \sum_{j \neq i} \frac{1}{N} \partial_s b_k(Z^{i,N,h}_r, \mu_r^{Z,N,h})(Z^{j,N,h}_r) dZ^{j,N,h}_r \\
+ \int_{\eta(s)}^t \text{Tr} \left( \left( \frac{1}{N} \partial_s b_k(Z^{i,N,h}_r, \mu_r^{Z,N,h})(Z^{i,N,h}_r) + \frac{2}{N} \partial_s b_k(Z^{i,N,h}_r, \mu_r^{Z,N,h})(Z^{i,N,h}_r) \right) d\left( Z^{i,N,h}_r \right) \right) \\
+ \sum_{j \neq i} \frac{1}{N} \partial^2_s b_k(Z^{i,N,h}_r, \mu_r^{Z,N,h})(Z^{j,N,h}_r, Z^{j,N,h}_r) d\left( Z^{j,N,h}_r \right) \right) \big| \mathcal{F}_{\eta(s)} \right] \\
= -\mathbb{E} \left[ \int_{\eta(s)}^t \sum_{j=1}^N \frac{1}{N} \partial_s b_k(Z^{i,N,h}_r, \mu_r^{Z,N,h})(Z^{j,N,h}_r) b(Z^{j,N,h}_{\eta(r)}; \mu_{\eta(r)}, \mu_s^{Z,N,h}) dr \\
+ \int_{\eta(s)}^t \partial_s b_k(Z^{i,N,h}_r, \mu_r^{Z,N,h}) b(Z^{i,N,h}_{\eta(r)}; \mu_{\eta(r)}, \mu_s^{Z,N,h}) dr \\
+ \sum_{j=1}^N \int_{\eta(s)}^t \frac{1}{N} \partial_s b_k(Z^{i,N,h}_r, \mu_r^{Z,N,h})(Z^{j,N,h}_r) T \sigma(Z^{j,N,h}_{\eta(r)}; \mu_{\eta(r)}) dW^j_r \\
+ \int_{\eta(s)}^t \frac{1}{N} \partial_s b_k(Z^{i,N,h}_r, \mu_r^{Z,N,h}) T \sigma(Z^{i,N,h}_{\eta(r)}; \mu_{\eta(r)}) dW^i_r \right].
\end{align*} \]
\[
+ \sum_{j=1}^{N} \int_{s}^{T} \text{Tr} \left( \left( \frac{1}{N} \partial_{v} \partial_{\mu} b_{k}(Z_{r}^{i,N,h}, \mu_{r}^{Z,N,h})(Z_{r}^{j,N,h}) \right) a(Z_{r}^{i,N,h}, \mu_{r}^{Z,N,h}) \right) ds \\
+ \frac{1}{N^{2}} \partial_{\mu}^{2} b_{k}(Z_{r}^{i,N,h}, \mu_{r}^{Z,N,h})(Z_{r}^{i,N,h}, Z_{r}^{j,N,h}) a(Z_{r}^{i,N,h}, \mu_{r}^{Z,N,h}) dr \\
+ \int_{s}^{T} \text{Tr} \left( \left( \partial_{x}^{2} b_{k}(Z_{r}^{i,N,h}, \mu_{r}^{Z,N,h}) + \frac{2}{N} \partial_{x} \partial_{\mu} b_{k}(Z_{r}^{i,N,h}, \mu_{r}^{Z,N,h})(Z_{r}^{i,N,h}) a(Z_{r}^{i,N,h}, \mu_{r}^{Z,N,h}) \right) dr \right) \left[ F_{q(s)} \right] \\
= - \int_{s}^{T} \sum_{j=1}^{N} \text{Tr} \left( \left( \frac{1}{N} \partial_{v} \partial_{\mu} b_{k}(Z_{r}^{i,N,h}, \mu_{r}^{Z,N,h})(Z_{r}^{j,N,h}) b(Z_{r}^{i,N,h}, \mu_{r}^{Z,N,h}) \right) ds \right) \\
+ \frac{1}{N^{2}} \partial_{\mu}^{2} b_{k}(Z_{r}^{i,N,h}, \mu_{r}^{Z,N,h})(Z_{r}^{i,N,h}, Z_{r}^{j,N,h}) a(Z_{r}^{i,N,h}, \mu_{r}^{Z,N,h}) \\
+ \text{Tr} \left( \left( \partial_{x}^{2} b_{k}(Z_{r}^{i,N,h}, \mu_{r}^{Z,N,h}) + \frac{2}{N} \partial_{x} \partial_{\mu} b_{k}(Z_{r}^{i,N,h}, \mu_{r}^{Z,N,h})(Z_{r}^{i,N,h}) a(Z_{r}^{i,N,h}, \mu_{r}^{Z,N,h}) \right) dr \right) \left[ F_{q(s)} \right] \\
\]

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

\[
\int_{0}^{T} \sum_{i=1}^{N} \mathbb{E} \left[ \frac{1}{N} \partial_{\mu} \mathcal{V}(s, \mu_{r}^{Z,N,h})(Z_{s}^{i,N,h}, \mu_{r}^{Z,N,h}) \right] ds
= \int_{0}^{T} \frac{1}{N} \sum_{i=1}^{N} \sum_{k=1}^{d} \mathbb{E} \left[ \left( \partial_{\mu} \mathcal{V}(s, \mu_{r}^{Z,N,h})(Z_{s}^{i,N,h}) \right) \right] \mathbb{E} \left[ \left( b_{k}(Z_{s}^{i,N,h}, \mu_{r}^{Z,N,h}) - b_{k}(Z_{s}^{i,N,h}, \mu_{r}^{Z,N,h}) \right) \right] \left[ F_{q(s)} \right] ds
= - \int_{s}^{T} \sum_{j=1}^{N} \text{Tr} \left( \left( \frac{1}{N} \partial_{v} \partial_{\mu} b_{k}(Z_{r}^{i,N,h}, \mu_{r}^{Z,N,h})(Z_{r}^{j,N,h}) b(Z_{r}^{i,N,h}, \mu_{r}^{Z,N,h}) + \partial_{x} b_{k}(Z_{r}^{i,N,h}, \mu_{r}^{Z,N,h})(Z_{r}^{i,N,h}) b(Z_{r}^{i,N,h}, \mu_{r}^{Z,N,h}) \right) \right) dr ds.
\]
Finally, by (4.2) and the fact that $V \in \mathcal{M}_2([0, T] \times \mathcal{P}_2(\mathbb{R}^d))$, we have

$$\left| \int_0^T \sum_{i=1}^N \mathbb{E} \left[ \frac{1}{N} \partial_\mu V(s, \mu^{Z,N,h}_\eta(s)) (Z^i,N,h) (b(Z^i,N,h, \mu^{Z,N,h}_\eta(s)) - b(Z^i,N,h, \mu^{Z,N,h}_s)) \right] \right| ds \leq C h.$$

Similarly, upon taking expectation, the second term of (B.9) is bounded by $C h$ and the third and fourth terms of (B.9) are also bounded by $C h$ by the Cauchy-Schwarz inequality. This completes the proof. 

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] Francois Delarue, Daniel Lacker, and Kavita Ramanan. From the master equation to mean field game limit theory: A central limit theorem. *arXiv preprint arXiv:1804.08542*, 2018.

[15] Steffen Dereich, Michael Scheutzow, and Reik Schottstedt. Constructive quantization: Approximation by empirical measures. In *Annales de l'IHP Probabilités et statistiques*, volume 49, 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 methods. *Acta Numerica*, 24:259–328, 2015.

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

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

[22] Michael B Giles and Lukasz Szpruch. Antithetic multilevel Monte Carlo estimation for multi-dimensional SDEs without Lévy area simulation. *The Annals of Applied Probability*, 24(4):1585–1620, 2014.

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

[27] 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.

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

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

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

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

[32] 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.

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

[34] 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.

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

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

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