CONVERGENCE ANALYSIS OF AN EXPLICIT METHOD AND ITS RANDOM BATCH APPROXIMATION FOR THE MCKEAN–VLASOV EQUATIONS WITH NON-GLOBALLY LIPSCHITZ CONDITIONS

Qian Guo[1], Jie He[2,*] and Lei Li[3]

Abstract. In this paper, we present a numerical approach to solve the McKean–Vlasov equations, which are distribution-dependent stochastic differential equations, under some non-globally Lipschitz conditions for both the drift and diffusion coefficients. We establish a propagation of chaos result, based on which the McKean–Vlasov equation is approximated by an interacting particle system. A truncated Euler scheme is then proposed for the interacting particle system allowing for a Khasminskii-type condition on the coefficients. To reduce the computational cost, the random batch approximation proposed in [S. Jin, L. Li and J. Liu, *J. Comput. Phys.* **400** (2020) 108877.] is extended to the interacting particle system where the interaction could take place in the diffusion term. An almost half order of convergence is proved in L^p sense. Numerical tests are performed to verify the theoretical results.

Mathematics Subject Classification. 65C20, 65C30, 60H35.

1. Introduction

The McKean–Vlasov equations are some distribution dependent stochastic differential equations (SDEs) that are first introduced by H. McKean [29] to describe the Vlasov system in dynamics theory and mean field stochastic differential equations. They have been widely studied due to their important applications in nonlinear equations and finance [6, 31, 36].

In this paper we consider the following McKean–Vlasov SDE:

$$X_t = X_0 + \int_0^t a\left(X_s, \mathcal{L}\left(X_s\right)\right) ds + \int_0^t b\left(X_s, \mathcal{L}\left(X_s\right)\right) dW_s, \tag{1.1}$$

where $\mathcal{L}(X_s)$ indicates the law of X_s, *i.e.*, $\mu_s := \mathcal{L}(X_s)$ satisfies

$$\mu_s(E) = \mathbb{P}(X_s \in E)$$

[1] Department of Mathematics, Shanghai Normal University, Shanghai 200234, P.R. China.

[2] Department of Mathematics, Jiangsu Second Normal University, Nanjing 210013, P.R. China.

[3] School of Mathematical Sciences, Institute of Natural Sciences, MOE-LSC, DCI joint team, Shanghai Jiao Tong University, 200240 Shanghai, P.R. China.

*Corresponding author: 1000497941@smail.shnu.edu.cn

with any Borel $E \subset \mathbb{R}^d$. Here, $X_0 \in \mathbb{R}^d$ is the initial value of the SDE sampled from an initial distribution μ_0, $a : \mathbb{R}^d \times \mathcal{P}_2(\mathbb{R}^d) \to \mathbb{R}^d$, $b : \mathbb{R}^d \times \mathcal{P}_2(\mathbb{R}^d) \to \mathbb{R}^{d \times m'}$ are some given fields and W. is an m'-dimensional standard Wiener processes so W_s is the process at time s. Here, $\mathcal{P}_2$ is defined as follows. The set of all probability measures on $\mathbb{R}^d$ is denoted by $\mathcal{P}(\mathbb{R}^d)$ while the notation $\mathcal{P}_p$ indexed by some $p \geq 1$ means the set of probability measures having finite p-moment, *i.e.*,

$$\mathcal{P}_p\left(\mathbb{R}^d\right) := \left\{\mu \in \mathcal{P}\left(\mathbb{R}^d\right) : \int_{\mathbb{R}^d} |x|^p \mu(dx) < \infty\right\}. \tag{1.2}$$

Clearly, $\mathcal{P}_p \subset \mathcal{P}_2$ for $p \geq 2$. We consider the dynamics for all $t \in (0, \infty)$. Later for numerical approximation, we will fix one, but arbitrary, terminal time T to study the convergence. Different from the classical SDEs [26], the coefficients of McKean–Vlasov SDEs depend on the law of the current variable X_s so the dynamics of the law $\mathcal{L}(X)$ is nonlinear [23].

The existence and uniqueness for strong solutions of (1.1) have been established in [34] under some linear growth and Lyapunov-type conditions. Later the existence and uniqueness for weak and strong solutions of the McKean–Vlasov SDEs have been studied by some authors under various conditions and we refer the reader to [4, 16, 30] for more details. To our knowledge, there is little literature that reports the existence and uniqueness results on the solution of a McKean–Vlasov SDE with super-linear coefficients. In [11], the authors showed that the McKean–Vlasov SDE has a unique solution when the drift coefficient satisfies a one-sided Lipschitz condition, but the diffusion term is still linearly growing. It is worthwhile to mention that a recent result that proves that the McKean–Vlasov SDE admits a unique solution when both drift and diffusion coefficients are super-linear w.r.t. the state [24].

As is well known, the first step to simulate the McKean–Vlasov SDEs is usually to approximate the true measure $\mathcal{L}(X_t)$ by the empirical measure. That is, a particle system is adopted to simulate the McKean–Vlasov SDE following the theory of propagation of chaos [34]. The state of the particle $i \in 1, \ldots, N$ in the symmetric system of SDEs coupled in a mean field scaling is then given by

$$X_t^i = X_0^i + \int_0^t a\left(X_s^i, \mu_s^X\right) ds + \int_0^t b\left(X_s^i, \mu_s^X\right) dW_s^i, \tag{1.3}$$

where

$$\mu_t^X(\cdot) := \frac{1}{N} \sum_{i=1}^N \delta_{X_t^i}(\cdot).$$

Here, δ_x denotes the Dirac measure at point x, $X_0^i \in \mathbb{R}^d$ are the initial values of the SDEs and the processes W_s^i $(1 \leq i \leq N)$ are N independent copies of the m'-dimensional standard Wiener processes. We always assume that X_0^i are i.i.d. sampled from some initial law μ_0.

Assuming the coefficients of (1.1) satisfy global Lipschitz condition w.r.t. the state and measure, the convergence of an Euler scheme applied to the particle system has been shown in [7]. In [15], multilevel and multi-index Monte Carlo methods have been proposed for the Euler schemes applied to the McKean-Vlasov equation with global Lipschitz conditions. Bao and Huang [2] studied propagation of chaos and convergence rate of the tamed Euler–Maruyama scheme for McKean–Vlasov SDEs by using associated weakly interacting particle systems, where the drift or diffusion term is Hlder continuous. Then the convergence of Milstein schemes, taming the drift by a one-sided Lipschitz condition, for a time-discretized interacting particle system was discussed in [3]. Li *et al.* [25] proved strong convergence of the Euler–Maruyama schemes for approximating McKean–Vlasov SDEs under local Lipschitz conditions w.r.t. the state variable. Under a Khasminskii-type monotonicity condition, instead of imposing a one-sided and global Lipschitz condition on the drift and diffusion coefficient, respectively, two explicit schemes were proposed in [24] for an interacting particle system and the propagation of chaos property of associated McKean–Vlasov equation was also studied therein. Hoeksema *et al.* [17] proved a large deviation principle for the empirical measure of a general system of mean-field interacting diffusions with

a singular drift and show convergence to the associated McKean–Vlasov equation. Rached *et al.* introduced in [32] a decoupling approach for McKean–Vlasov SDEs, which approximates the McKean–Vlasov law in the coefficients.

Clearly, the computational cost for the numerical schemes is $\mathcal{O}(N^2)$ per time step if naively implemented, which is expensive in applications. To save the computational cost, a projection-based particle method was proposed in [5] for solving McKean–Vlasov stochastic differential equations. Recently, a random batch method(RBM) has been introduced in [20] to simulate the interacting particle systems efficiently, which can reduce the computational cost from $\mathcal{O}(N^2)$ per time step to $\mathcal{O}(N)$ with additive noise.

In this paper, we first extend the theory of propagation of chaos for a highly nonlinear McKean–Vlasov equation [24] to the case of $p \geq 2$. We then adopt the truncated Euler scheme, which is different from the tamed scheme in [24], to numerically solve (1.3) for approximation of the McKean–Vlasov SDE. The main contribution of this paper in the numerical analysis is twofold. Firstly, we provide a strong convergence analysis of a truncated Euler scheme to numerically solve the McKean–Vlasov SDEs under a Khasminskii-type monotonicity condition w.r.t. the state. Secondly, we extend the aforementioned random batch approximation to improve the efficiency of the truncated method, and establish the error estimate for the approximation. Different from the results in [20], the diffusion coefficient in the interacting particle systems considered in this paper could contain weakly interacting terms so that the random batch approximation also takes place in the diffusion.

The remainder of this article is organized as follows: In Section 2, some assumptions are introduced and the propagation of chaos property is obtained. In Section 3, we employ a truncated Euler method to solve the interacting particle systems and prove that the convergence order is almost one-half. Then in Section 4, we present the convergence and efficiency of the numerical method by integrating the RBM into the truncated Euler scheme. In the last section (Sect. 5), a numerical experiment is given to verify the theoretical results.

2. Mathematical preliminaries

We first introduce some notations and basic definitions for later sections. The Euclidean norm of a d-dimensional vector is denoted by $|\cdot|$ and the Hilbert-Schmidt norm of a $d \times m$-matrix is denoted by $||\cdot||$, where we recall the Hilbert-Schmidt norm is given by $\|A\|^2 = \mathrm{tr}(A^T A)$ for a real matrix A. If G is a set, its indicator function $\mathcal{I}_G$ is given by

$$\mathcal{I}_G(x) = \begin{cases} 1, & x \in G, \\ 0, & \text{otherwise.} \end{cases}$$

Moreover, $a \wedge b := \min(a, b)$ and $a \vee b := \max(a, b)$. The notation $u \otimes v$ for $u \in \mathbb{R}^d$ and $v \in \mathbb{R}^d$ means the tensor product of u and v, namely, $(u \otimes v)_{ij} = u_i v_j$. Throughout this article, $C > 0$ is a generic constant that might change its value from line to line. We use the notation $A \lesssim B$ to mean $A \leq CB$ for some generic constant C that is independent of the parameters been focused on.

Recall (1.2) for $\mathcal{P}_p$. For $p \geq 1$ and any $\mu, \nu \in \mathcal{P}_p(\mathbb{R}^d)$, the p-Wasserstein distance is defined by

$$\mathcal{W}_p(\mu, \nu) := \left(\inf_{\pi \in \Pi(\mu, \nu)} \int_{\mathbb{R}^d \times \mathbb{R}^d} |x - y|^p \pi(dx, dy) \right)^{1/p},$$

where $\Pi(\mu, \nu)$ is the set of couplings of μ and ν (*i.e.*, joint distributions with marginals to be μ and ν respectively). Clearly, $\mathcal{P}_p(\mathbb{R}^d)$ is a Polish space under the p-Wasserstein metric.

Regarding the $\mathcal{W}_p$ distance, we have the following simple observations.

Lemma 2.1. *Let $p \geq 1$. For any probability measure μ,*

$$\mathcal{W}_p^p(\mu, \delta_0) = \int |y|^p \mu(dy).$$

For two empirical measures $\mu_1 = \frac{1}{N} \sum_{i=1}^{N} \delta_{Y^{1,i}}$ and $\mu_2 = \frac{1}{N} \sum_{i=1}^{N} \delta_{Y^{2,i}}$. Then,

$$\mathcal{W}_p^p(\mu_1, \mu_2) \le \frac{1}{N} \sum_{i=1}^{N} |Y^{1,i} - Y^{2,i}|^p.$$

The proof for the first claim with $p = 2$ can be found, for example, in Lemma 2.3 of [11], and the proof for general $p \ge 1$ can be similarly performed. The second follows from constructing a simple transport plan $\pi(dx, dy) = \frac{1}{N} \sum_{i=1}^{N} \delta_{Y^{1i}}(dx) \otimes \delta_{Y^{2i}}(dy)$.

We will make the following technical assumptions on the coefficients, which are essentially the continuity and growth conditions.

Assumption 2.2. *There exists a constant $\gamma \ge 0$ such that*

$$|a(x, \mu) - a(\bar{x}, \bar{\mu})| \vee \|b(x, \mu) - b(\bar{x}, \bar{\mu})\| \le C[(1 + |x|^\gamma + |\bar{x}|^\gamma)|x - \bar{x}| + \mathcal{W}_2(\mu, \bar{\mu})], \tag{2.1}$$

for all $x, \bar{x} \in \mathbb{R}^d$ and $\mu, \bar{\mu} \in \mathcal{P}_2(\mathbb{R}^d)$.

Assumption 2.3. *There exists a pair of constants $L > 0$ and $p_0 > \max(2\gamma, 4)$ such that the coefficients $a(\cdot, \cdot) : \mathbb{R}^d \times \mathcal{P}_2(\mathbb{R}^d) \to \mathbb{R}^d$ and $b(\cdot, \cdot) : \mathbb{R}^d \times \mathcal{P}_2(\mathbb{R}^d) \to \mathbb{R}^{d \times m'}$ satisfy the initial condition $|a(0, \mu)| + |b(0, \mu)| \le L \mathcal{W}_2(\mu, \delta_0)$ and*

$$(x - \bar{x}) \cdot (a(x, \mu) - a(\bar{x}, \bar{\mu})) + \frac{1}{2}(p_0 - 1) \|b(x, \mu) - b(\bar{x}, \bar{\mu})\|^2 \le L \left\{ |x - \bar{x}|^2 + \mathcal{W}_2^2(\mu, \bar{\mu}) \right\}, \tag{2.2}$$

for all $x, \bar{x} \in \mathbb{R}^d$ and $\mu, \bar{\mu} \in \mathcal{P}_2(\mathbb{R}^d)$.

The assumptions here also appeared in [24]. We note that the Lipschitz continuity for μ is uniform. A typical example is $a(x, \mu) = f(x) + \int k(x, y)\mu(dy)$ where k is globally Lipschitz. As another comment, for condition (2.2) to hold, b is not necessarily globally Lipschitz in x. In fact, as long as the growth can be controlled by the confining effect in a, the condition still holds. One example is $a = x - x^3$ and $b = x^2$ for $x \in \mathbb{R}$.

The lower bound 2γ for p_0 in Assumption 2.3 is essential in our proof and is used in the proof of Theorem 3.4 where we need to bound $pq_1\gamma$ moment for some $p \ge 2$ and $q_1 > 1$. The lower bound 4 for p_0, however, is not essential for the convergence. It only ensures the existence of $p \in [2, p_0/2]$ so the error estimates given below in Proposition 2.5 and Theorem 3.4 can be cleaner (otherwise, there would be a term like $N^{-(q-p)/q}$).

The assumptions made above clearly imply the following growth conditions of the coefficients:

– For any $q < p_0$ where p_0 is the parameter in Assumption 2.3, there exists a constant $L > 0$ such that

$$x \cdot a(x, \mu) + \frac{1}{2}(q - 1) \|b(x, \mu)\|^2 \le L \left\{ (1 + |x|)^2 + \mathcal{W}_2^2(\mu, \delta_0) \right\} \tag{2.3}$$

for all $x \in \mathbb{R}^d$ and $\mu \in \mathcal{P}_2(\mathbb{R}^d)$.

In fact, taking $\bar{x} = 0$ and $\bar{\mu} = \delta_0$ in Assumption 2.3, one has

$$x \cdot (a(x, \mu) - a(0, \delta_0)) + \frac{1}{2}(p_0 - 1) \|b(x, \mu) - b(0, \delta_0)\|^2 \le L(|x|^2 + \mathcal{W}_2^2(\mu, \delta_0)).$$

Moreover, since $\|b(x, \mu)\|^2 \le \|b(x, \mu) - b(0, \delta_0)\|^2 + 2 \|b(x, \mu) - b(0, \delta_0)\| \|b(0, \delta_0)\| + \|b(0, \delta_0)\|^2$ and $2uv \le \delta u^2 + \frac{1}{\delta} v^2$ for any $u, v \in \mathbb{R}$ and $\delta > 0$, one has

$$\frac{1}{2}(q - 1) \|b(x, \mu)\|^2 \le \frac{1}{2}(p_0 - 1) \|b(x, \mu) - b(0, \delta_0)\|^2 + \frac{(q-1)(p_0-1)}{2(p_0-q)} \|b(0, \delta_0)\|^2.$$

Combining these two simple estimates, the conclusion then follows.

– For the same constant γ as in Assumption 2.2, it holds that

$$|a(x,\mu)| \vee \|b(x,\mu)\| \leq C[(1 + |x|^{\gamma+1}|) + \mathcal{W}_2(\mu, \delta_0)], \tag{2.4}$$

for all $x \in \mathbb{R}^d$ and $\mu \in \mathcal{P}_2(\mathbb{R}^d)$.

Corresponding to (1.3), we consider $\bar{X}_t^i$ which solves (1.1) but with $\bar{X}_0^i = X_0^i$ and the same Brownian motions. In other words, $\bar{X}_t^i$'s satisfy the following

$$\bar{X}_t^i = X_0^i + \int_0^t a\left(\bar{X}_s^i, \mathcal{L}\left(\bar{X}_s^i\right)\right) ds + \int_0^t b\left(\bar{X}_s^i, \mathcal{L}\left(\bar{X}_s^i\right)\right) dW_s^i \tag{2.5}$$

almost surely for any $t \in [0,T]$ and $i \in 1, \cdots, N$. Here, W_t^i is the same process as in (1.3) and $X_0^i \in \mathbb{R}^d$ are the initial values of the SDEs. Clearly, $\{\bar{X}_t^i\}$ are i.i.d. with the law being $\mathcal{L}(X_t)$ for the McKean–Vlasov SDE (1.1).

For the convenience of the discussion, we denote the moments of μ for $q \geq 1$ by

$$M_q(\mu) := \int_{\mathbb{R}^d} |x|^q \mu(dx).$$

Hence, $M_q(\mu_t) = \mathbb{E}[|\bar{X}_t^i|^q]$ for any i.

By standard techniques, one may show the following bounds on the moments (see, for example, [9], Thm. 1). For the convenience of the readers, we provide a short sketch in Appendix A.

Proposition 2.4. Let (2.3) hold for some $q \geq 2$ and $\mu_0 \in \mathcal{P}_q$. Then for any $T > 0$, there exists C depending on T and q but independent of N and i such that

$$\sup_{0 \leq t \leq T} \mathbb{E}[|X_t^i|^q + |\bar{X}_t^i|^q] \leq C\left(1 + \mathbb{E}[|X_0^i|^q]\right). \tag{2.6}$$

The following result claims that the trajectories of $\bar{X}_t^i$ and X_t^i become identical as $N \to \infty$ almost surely. This implies the propagation of chaos for the interacting particle system so that the mean field limit to the McKean SDE holds.

Proposition 2.5 (Propagation of Chaos). *Let Assumption 2.3 be satisfied. If for some $p \in [2, p_0/2)$ the initial law has finite p-moment (i.e., $\mu_0 \in \mathcal{P}_p \subset \mathcal{P}_2$), then it holds that*

$$\sup_{i \in \{1,\ldots,N\}} \sup_{t \in [0,T]} \mathbb{E}[|X_t^i - \bar{X}_t^i|^p] \leq C \begin{cases} N^{-1/2}, & \text{if } p > d/2, \\ N^{-1/2}\log(1+N), & \text{if } p = d/2, \\ N^{-p/d}, & \text{if } p \in [2, d/2), \end{cases} \tag{2.7}$$

where the constant $C > 0$ depends on T, p and p_0 but does not depend on N.

Proof. From equations (1.3), (2.5), we have

$$\bar{X}_t^i - X_t^i = \int_0^t a\left(\bar{X}_s^i, \mathcal{L}\left(\bar{X}_s^i\right)\right) - a\left(X_s^i, \mu_s^X\right) ds + \int_0^t b\left(\bar{X}_s^i, \mathcal{L}\left(\bar{X}_s^i\right)\right) - b\left(X_s^i, \mu_s^X\right) dW_s^i. \tag{2.8}$$

Using Itô's formula,

$$\mathbb{E}[|\bar{X}_t^i - X_t^i|^p] \leq p\mathbb{E}\left[\int_0^t |\bar{X}_s^i - X_s^i|^{p-2}\left(\bar{X}_s^i - X_s^i\right) \cdot \left(a\left(\bar{X}_s^i, \mathcal{L}\left(\bar{X}_s^i\right)\right) - a\left(X_s^i, \mu_s^X\right)\right) ds\right]$$

$$+ \frac{p(p-1)}{2}\mathbb{E}\left[\int_0^t |\bar{X}_s^i - X_s^i|^{p-2}\|b\left(\bar{X}_s^i, \mathcal{L}\left(\bar{X}_s^i\right)\right) - b\left(X_s^i, \mu_s^X\right)\|^2 ds\right].$$

Due to Assumption 2.3, one has

$$\mathbb{E}[|\bar{X}_t^i - X_t^i|]^p \le L\mathbb{E}\left[\int_0^t |\left(\bar{X}_s^i - X_s^i\right)|^{p-2}[\mathcal{W}_2^2\left(\mathcal{L}\left(\bar{X}_s^i\right), \mu_s^X\right) + |\bar{X}_s^i - X_s^i|^2]ds\right]$$

$$\le L\mathbb{E}\left[\int_0^t |\left(\bar{X}_s^i - X_s^i\right)|^p ds\right] + L\mathbb{E}\left[\int_0^t |\left(\bar{X}_s^i - X_s^i\right)|^{p-2}\mathcal{W}_2^2\left(\mathcal{L}\left(\bar{X}_s^i\right), \mu_s^X\right)ds\right]$$

$$\le C\mathbb{E}\left[\int_0^t |(\bar{X}_s^i - X_s^i)|^p ds\right] + C\mathbb{E}\left[\int_0^t \mathcal{W}_2^p\left(\mathcal{L}\left(\bar{X}_s^i\right), \mu_s^X\right)ds\right],$$

for any $t \in [0, T]$. The last inequality is due to the Young's inequality.

Denote $\mu_s^{\bar{X}} := \frac{1}{N}\sum_{j=1}^N \delta_{\bar{X}_t^j}$. By Lemma 2.1, one has

$$\mathcal{W}_2^p\left(\mathcal{L}\left(\bar{X}_s^i\right), \mu_s^X\right) = \left[\mathcal{W}_2^2\left(\mathcal{L}\left(\bar{X}_s^i\right), \mu_s^X\right)\right]^{p/2}$$

$$\le \left[2\mathcal{W}_2(\mu_s^{\bar{X}}, \mu_s^X)^2 + 2\mathcal{W}_2^2(\mu_s^{\bar{X}}, \mathcal{L}(\bar{X}_s^1))\right]^{p/2}$$

$$\le C\left[\frac{1}{N}\sum_{j=1}^N |\bar{X}_s^j - X_s^j|^2\right]^{p/2} + C\mathcal{W}_2^p\left(\frac{1}{N}\sum_{j=1}^N \delta_{\bar{X}_t^j}, \mathcal{L}\left(\bar{X}_s^1\right)\right).$$

Consequently, one obtains

$$\mathbb{E}[|\bar{X}_t^i - X_t^i|^p] \le C\int_0^t \mathbb{E}[|(\bar{X}_s^i - X_s^i)|^p]ds + C\int_0^t \mathbb{E}\left[\frac{1}{N}\sum_{j=1}^N |\bar{X}_s^j - X_s^j|^2\right]^{p/2} ds$$

$$+ C\int_0^t \mathbb{E}\left[\mathcal{W}_2^p\left(\frac{1}{N}\sum_{j=1}^N \delta_{\bar{X}_t^j}, \mathcal{L}\left(\bar{X}_s^1\right)\right)\right]ds$$

$$\le C\int_0^t \mathbb{E}[|\bar{X}_s^i - X_s^i|^p]ds + C\int_0^t \mathbb{E}\left[\mathcal{W}_p^p\left(\frac{1}{N}\sum_{j=1}^N \delta_{\bar{X}_t^j}, \mathcal{L}\left(\bar{X}_s^1\right)\right)\right]ds.$$

In the last inequality, we have applied $\mathcal{W}_2(\mu, \nu) \le \mathcal{W}_p(\mu, \nu)$. Then, by the Minkowski inequality,

$$\mathbb{E}\left[\frac{1}{N}\sum_{j=1}^N |\bar{X}_s^j - X_s^j|^2\right]^{p/2} \le \left[\frac{1}{N}\sum_{j=1}^N \||\bar{X}_s^j - X_s^j|^2\|_{L^{p/2}}\right]^{p/2} = \left[\frac{1}{N}\sum_{j=1}^N (\mathbb{E}[|\bar{X}_s^j - X_s^j|^p])^{2/p}\right]^{p/2} = \mathbb{E}[|\bar{X}_s^i - X_s^i|^p].$$

The last equality here holds due to the symmetry.

The term $\mathbb{E}[\mathcal{W}_p^p(\frac{1}{N}\sum_{j=1}^N \delta_{\bar{X}_t^j}, \mathcal{L}\left(\bar{X}_s^1\right))]$ is controlled by the Wasserstein distance estimate for the empirical measures with i.i.d. samples in Theorem 1 of [13]. Here, since $p_0 > 2p$, one can choose the moment q such that $q > 2p$ and the terms $N^{-(q-p)/q}$ in Theorem 1 of [13] can be removed for simplicity. Then, applying Grönwall's inequality completes the proof. □

We remark that the requirement $p < p_0/2$ is only used such that the term $N^{-(q-p)/q}$ is absent for simplicity. Clearly, if $p \in [p_0/2, p_0)$, there is still convergence. Another remark is that the rate of the propagation of chaos here is like $N^{-1/(2p)}$ under L^p norm when p is large. This low rate is due to the usage of Wasserstein distances to gauge the continuity on the distribution dependence in the coefficients. If the dependence on the distribution is through some statistics, the rate may be improved to $N^{-1/2}$ due to the central limit theorem. Note that though the particles are dependent for finite N, the rate of central limit theorem is expected to hold in the large N regime due to propagation of chaos, though rigorous proof needs careful estimates (see, for example, [35] regarding the rate of convergence for statistics).

3. A truncated Euler scheme

The drift and diffusion coefficients can be unbounded. To overcome this difficulty, we make use of the truncation techniques [14, 27, 28]. For this purpose, we choose a strictly increasing continuous function φ such that

$$\sup_{|x| \leq u} \left(|a(x)| \vee \|b(x)\| \right) \leq \varphi(u).$$

The inverse function φ^{-1} is strictly increasing continuous function from $[\varphi(0), +\infty)$ to $\mathbb{R}_+$. Firstly, define a uniform time discretization grid $T_{N_T} : 0 = t_0 < t_1 < \cdots < t_{N_T} = T$, where $t_m = m\Delta$ with the step size $\Delta = T/N_T$ and $N_T \in \mathbb{N}$ is the number of subintervals. Take $\Delta^* \in (0, 1]$, which is meant to be the largest step size, and a strictly decreasing function $h : (0, \Delta^*] \longrightarrow (0, +\infty)$ such that

$$h(\Delta^*) \geq \varphi(1), \quad \lim_{\Delta \to 0} h(\Delta) = \infty \quad \text{and} \quad \sup_{\Delta \in (0, \Delta^*]} \Delta^{1/4} h(\Delta) < \infty.$$

One possible example is $h(\Delta) = \Delta^{-\frac{\epsilon}{2}}$ and ϵ is any value in $(0, 1/4)$. Another possible example is $h(\Delta) = |\log \Delta|$. For a given step size $\Delta \in (0, 1]$, define the truncated functions by

$$a^\Delta(x, \mu) = a\left(\left(|x| \wedge \varphi^{-1}(h(\Delta)) \right) \frac{x}{|x|}, \quad \mu \right), \quad b^\Delta(x, \mu) = b\left(\left(|x| \wedge \varphi^{-1}(h(\Delta)) \right) \frac{x}{|x|}, \quad \mu \right).$$

Similar to that in [27], the truncated coefficients satisfy the following properties.

Lemma 3.1. *It holds that*

$$|a^\Delta(x, \mu)| \vee \|b^\Delta(x, \mu)\| \leq \varphi(\varphi^{-1}(h(\Delta))) = h(\Delta), \tag{3.1}$$

and for any $q \in [2, p_0)$ where p_0 is parameter in Assumption 2.3, there exists a constant $L > 0$ such that

$$x \cdot a^\Delta(x, \mu) + \frac{1}{2}(q-1) \left\| b^\Delta(x, \mu) \right\|^2 \leq L \left\{ (1 + |x|)^2 + \mathcal{W}_2^2(\mu, \delta_0) \right\}, \quad \forall x \in \mathbb{R}^d, \mu \in \mathcal{P}_2(\mathbb{R}^d). \tag{3.2}$$

Proof. Here, we only need to verify the Khasminskii-type condition (3.2).

When $|x| \leq \varphi^{-1}(h(\Delta))$, it is simply (2.3). For $|x| > \varphi^{-1}(h(\Delta))$, due to the dependence on μ, unlike the argument in [27], we start by (2.2) with

$$x \cdot a(x, \mu) + \frac{1}{2}(p_0 - 1) \|b(x, \mu) - b(0, \mu)\|^2 \leq x \cdot a(0, \mu) + L|x|^2.$$

Denote

$$\bar{x} = \varphi^{-1}(h(\Delta)) \frac{x}{|x|}.$$

Then, it follows that

$$\begin{aligned} x \cdot a^\Delta(x, \mu) + \frac{1}{2}(p_0 - 1) \left\| b^\Delta(x, \mu) - b^\Delta(0, \mu) \right\|^2 &= x \cdot a(\bar{x}, \mu) + \frac{1}{2}(p_0 - 1) \|b(\bar{x}, \mu) - b(0, \mu)\|^2 \\ &\leq \frac{|x|}{\varphi^{-1}(h(\Delta))} \left[\bar{x} \cdot a(\bar{x}, \mu) + \frac{1}{2}(p_0 - 1) \|b(\bar{x}, \mu) - b(0, \mu)\|^2 \right] \\ &\leq \frac{|x|}{\varphi^{-1}(h(\Delta))} \left(\bar{x} \cdot a(0, \mu) + L(\varphi^{-1}(h(\Delta)))^2 \right) \\ &= x \cdot a(0, \mu) + L|x|\varphi^{-1}(h(\Delta)) \\ &\leq x \cdot a(0, \mu) + L|x|^2. \end{aligned}$$

Then a similar argument as in deriving (2.3) gives the desired result. □

3.1. The truncated scheme

The numerical approximation of X_t^i at t_m is denoted by $X_{t_m}^{i,\Delta}$. With the truncated coefficients, the numerical solutions $X_{t_m}^{i,\Delta}$ are then generated by the basic Euler-Maruyama scheme

$$X_{t_{m+1}}^{i,\Delta} = X_{t_m}^{i,\Delta} + a^\Delta \left(X_{t_m}^{i,\Delta}, \mu_{t_m}^{X,\Delta} \right) \Delta + b^\Delta \left(X_{t_m}^{i,\Delta}, \mu_{t_m}^{X,\Delta} \right) \Delta W_{t_m}^i, \tag{3.3}$$

where $\Delta W_{t_m}^i = W_{t_{m+1}}^i - W_{t_m}^i$, the initial value $X_0^{i,\Delta} = X_0^i$ and

$$\mu_{t_m}^{X,\Delta}(\cdot) := \frac{1}{N} \sum_i \delta_{X_{t_m}^{i,\Delta}}(\cdot).$$

The scheme (3.3) is the basic truncated scheme studied in this work.

We introduce two versions of extension of the numerical solution at the discrete time points to $t \in [0, \infty)$. The first is the piecewise constant extension given by

$$\hat{X}_t^{i,\Delta} = X_{t_m}^{i,\Delta}, \quad t_m \le t < t_{m+1},$$

and the second is the continuous extension of the truncated Euler-Maruyama method defined by

$$X_t^{i,\Delta} = X_{t_m}^{i,\Delta} + \int_{t_m}^t a^\Delta \left(\hat{X}_s^{i,\Delta}, \hat{\mu}_s^{X,\Delta} \right) ds + \int_{t_m}^t b^\Delta \left(\hat{X}_s^{i,\Delta}, \hat{\mu}_s^{X,\Delta} \right) dW_s^i. \tag{3.4}$$

Here, the initial value $\hat{X}_0^{i,\Delta} = X_0^i$ and

$$\hat{\mu}_t^{X,\Delta}(\cdot) := \frac{1}{N} \sum_{i=1}^N \delta_{\hat{X}_t^{i,\Delta}}(\cdot).$$

3.2. Analysis of the truncated scheme

We present some preliminary properties of the numerical solutions. We first give some moments estimates of the numerical solution to ensure certain stability. The first lemma below is to control the difference between the two versions of extension.

Lemma 3.2. *For any $\Delta \in (0, \Delta^*]$, any $t \ge 0$ and any $p \ge 2$,*

$$\mathbb{E}[|X_t^{i,\Delta} - \hat{X}_t^{i,\Delta}|^p] \le C\Delta^{p/2} h^p(\Delta), \tag{3.5}$$

where C is a positive constant independent of Δ.

Proof. By the definitions (3.4) , one has

$$X_t^{i,\Delta} - \hat{X}_t^{i,\Delta} = \int_{t_m}^t a^\Delta \left(\hat{X}_s^{i,\Delta}, \hat{\mu}_s^{X,\Delta} \right) ds + \int_{t_m}^t b^\Delta \left(\hat{X}_s^{i,\Delta}, \hat{\mu}_s^{X,\Delta} \right) dW_s^i.$$

Using Hölder's inequality, Burkholder-Davis-Gundy inequality and (3.1), we have

$$\mathbb{E}[|X_t^{i,\Delta} - \hat{X}_t^{i,\Delta}|^p] \le C\mathbb{E}\left[\left| \int_{t_m}^t a^\Delta \left(\hat{X}_s^{i,\Delta}, \hat{\mu}_s^{X,\Delta} \right) ds \right|^p + \left| \int_{t_m}^t b^\Delta \left(\hat{X}_s^{i,\Delta}, \hat{\mu}_s^{X,\Delta} \right) dW_s^i \right|^p \right]$$

$$\le C\Delta^{p-1}\mathbb{E}\left[\int_{t_m}^t \left| a^\Delta \left(\hat{X}_s^{i,\Delta}, \hat{\mu}_s^{X,\Delta} \right) \right|^p ds \right] + C\Delta^{(p-2)/2}\mathbb{E}\left[\int_{t_m}^t \left\| b^\Delta \left(\hat{X}_s^{i,\Delta}, \hat{\mu}_s^{X,\Delta} \right) \right\|^p ds \right]$$

$$\le C\Delta^{p-1}\mathbb{E}\left[\int_{t_m}^t h^p(\Delta) ds \right] + C\Delta^{(p-2)/2}\mathbb{E}\left[\int_{t_m}^t h^p(\Delta) ds \right]$$

$$\le C\Delta^{p/2} h^p(\Delta).$$

$\square$

Proposition 3.3. *Suppose* (3.2) *holds and* $\mu_0 \in \mathcal{P}_q \subset \mathcal{P}_2$. *Then for any* $\Delta \in (0, \Delta^*]$, *any* $T > 0$, *it holds that*

$$\sup_{0 \leq t \leq T} \mathbb{E}[|X_t^{i,\Delta}|^q] \leq C \left(1 + \mathbb{E}[|X_0^i|^q] \right), \tag{3.6}$$

where C *is a positive constant dependent on* T *and* q *but independent of* Δ.

Proof. Recall (3.4):

$$dX_t^{i,\Delta} = a^\Delta \left(\hat{X}_t^{i,\Delta}, \hat{\mu}_t^{X,\Delta} \right) dt + b^\Delta \left(\hat{X}_t^{i,\Delta}, \hat{\mu}_t^{X,\Delta} \right) dW_t^i.$$

By the Itô's formula, one has

$$\frac{d}{dt}\mathbb{E}[|X_t^{i,\Delta}|^q] \leq q\mathbb{E}\left[|X_t^{i,\Delta}|^{q-2} \left(X_t^{i,\Delta} \cdot a^\Delta \left(\hat{X}_t^{i,\Delta}, \hat{\mu}_t^{X,\Delta} \right) \right) \right] + q\mathbb{E}\left[\frac{q-1}{2}|X_t^{i,\Delta}|^{q-2} \left\| b^\Delta \left(\hat{X}_t^{i,\Delta}, \hat{\mu}_t^{X,\Delta} \right) \right\|^2 \right].$$

Writing $X_t^{i,\Delta} = (X_t^{i,\Delta} - \hat{X}_t^{i,\Delta}) + \hat{X}_t^{i,\Delta}$, one has according to (3.2) that

$$\left(X_t^{i,\Delta} \cdot a^\Delta \left(\hat{X}_t^{i,\Delta}, \hat{\mu}_t^{X,\Delta} \right) \right) + \frac{q-1}{2} \left\| b^\Delta \left(\hat{X}_t^{i,\Delta}, \hat{\mu}_t^{X,\Delta} \right) \right\|^2$$

$$\leq \left(X_t^{i,\Delta} - \hat{X}_t^{i,\Delta} \right) \cdot a^\Delta \left(\hat{X}_t^{i,\Delta}, \hat{\mu}_t^{X,\Delta} \right) + L \left[(1 + |\hat{X}_t^{i,\Delta}|)^2 + \mathcal{W}_2^2(\hat{\mu}_t^{X,\Delta}, \delta_0) \right].$$

Hence,

$$\frac{d}{dt}\mathbb{E}[|X_t^{i,\Delta}|^q] \leq C\mathbb{E}\left[|X_t^{i,\Delta}|^{q-2}((1 + |\hat{X}_t^{i,\Delta}|)^2 + \mathcal{W}_2^2(\hat{\mu}_t^{X,\Delta}, \delta_0)) \right] + \mathbb{E}\left[|X_t^{i,\Delta} - \hat{X}_t^{i,\Delta}|^{q/2} \left| a^\Delta \left(\hat{X}_t^{i,\Delta}, \hat{\mu}_t^{X,\Delta} \right) \right|^{q/2} \right]$$

$$\leq C(1 + \mathbb{E}[|X_t^{i,\Delta}|^q]) + C\mathbb{E}\left[|X_t^{i,\Delta} - \hat{X}_t^{i,\Delta}|^{q/2}|a^\Delta \left(\hat{X}_t^{i,\Delta}, \hat{\mu}_t^{X,\Delta} \right)|^{q/2} \right] + \mathbb{E}\left[\left(\frac{1}{N}\sum_{j=1}^N |\hat{X}_t^{j,\Delta}|^2 \right)^{q/2} \right],$$

where we applied Lemma 2.1.
By Lemma 3.2, it is clear that

$$\mathbb{E}\left[|X_t^{i,\Delta} - \hat{X}_t^{i,\Delta}|^{q/2}|a^\Delta \left(\hat{X}_t^{i,\Delta}, \hat{\mu}_t^{X,\Delta} \right)|^{q/2} \right] \lesssim h^{q/2}\mathbb{E}\left[|X_t^{i,\Delta} - \hat{X}_t^{i,\Delta}|^{q/2} \right] \lesssim \Delta^{q/4}h^q(\Delta).$$

By the Minkowski inequality and the symmetry

$$\mathbb{E}\left(\frac{1}{N}\sum_{j=1}^N |\hat{X}_t^{i,\Delta}|^2 \right)^{q/2} \leq \left(\frac{1}{N}\sum_{j=1}^N (\mathbb{E}[|\hat{X}_t^{j,\Delta}|^q])^{2/q} \right)^{q/2} = \mathbb{E}[|\hat{X}_t^{i,\Delta}|^q], \quad \forall i.$$

It follows from the fact $|\hat{X}_t| \leq \sup_{s \leq t} |X_s^{i,\Delta}|$ by definition that

$$\mathbb{E}[|X_t^{i,\Delta}|^q] \leq (\mathbb{E}[|X_0^i|^q] + C_T) + C\int_0^t \sup_{0 \leq u \leq s} \mathbb{E}[|X_u^{i,\Delta}|^q]ds.$$

By the monotonticy of the bound on the right hand side, one has

$$\sup_{0 \leq s \leq t} \mathbb{E}[|X_s^{i,\Delta}|^q] \leq (\mathbb{E}[|X_0^i|^q] + C_T) + C\int_0^t \sup_{0 \leq u \leq s} \mathbb{E}[|X_u^{i,\Delta}|^q]ds.$$

Application of Grönwall's inequality then completes the proof. $\qquad\square$

Below, we prove the convergence of the truncated scheme. Define two stopping times for particle i

$$\tau_i := \inf\{t \geq 0, |\bar{X}_t^i| \geq \varphi^{-1}(h(\Delta))\}$$

and

$$\rho_i := \inf\{t \geq 0, |X_t^{i,\Delta}| \geq \varphi^{-1}(h(\Delta))\}.$$

Let $\theta_i = \tau_i \wedge \rho_i$. The estimates in Propostion 3.3 ensures that $\theta_i \to +\infty$ a.s. as $\Delta \to 0$. Clearly, before θ_i, the truncated method reduces to the usual Euler–Maruyama scheme.

Below, we make use of these stopping times to obtain the convergence of the truncated scheme, and the following is our first main result.

Theorem 3.4 (Rate of Convergence). *Suppose Assumptions 2.2 and 2.3 hold and $\mu_0 \in \mathcal{P}_{p_0}$. Let $\{X_t^i\}$ be the solution to the N-particle system* (1.3) *and $X_t^{i,\Delta}$ be the solution to the truncated scheme* (3.4)*. Denote $\eta(\Delta) := 1/(\varphi^{-1} \circ h(\Delta))$. Then, for $p \in [2, p_0/2)$ with $p\gamma < p_0$, it holds for every $q < p_0$ that*

$$\sup_{0 \leq t \leq T} \mathbb{E}[|X_t^i - X_t^{i,\Delta}|^p] \leq C_T \Delta^{p/2} h^p(\Delta) + C_T \eta(\Delta)^{q-p}, \tag{3.7}$$

where C_T is a positive constant dependent on p, q, T but independent of N and Δ. Consequently, one has for $q < p_0$,

$$\sup_{t \leq T} \mathbb{E}[|\bar{X}_t^i - X_t^{i,\Delta}|^p] \leq C \begin{cases} N^{-1/2} + \Delta^{p/2} h(\Delta)^p + \eta(\Delta)^{q-p}, & \text{if } p > d/2, \\ N^{-1/2} \log(1+N) + \Delta^{p/2} h(\Delta)^p + \eta(\Delta)^{q-p}, & \text{if } p = d/2, \\ N^{-p/d} + \Delta^{p/2} h(\Delta)^p + \eta(\Delta)^{q-p}, & \text{if } p \in [2, d/2), \end{cases} \tag{3.8}$$

where C depends on p, q, T but is independent of N and Δ. Hence, it holds that

$$\lim_{N \to \infty, \Delta \to 0} \sup_{t \leq T} \left(\mathbb{E}[|\bar{X}_t^i - X_t^{i,\Delta}|^p] \right)^{1/p} = 0.$$

Proof. For $0 \leq s \leq t \wedge \theta_i$, the coefficients a and b that particle i sees agree with the truncated ones, a^Δ and b^Δ. Hence, by the Itô's formula,

$$\mathbb{E}[|X_{t \wedge \theta_i}^i - X_{t \wedge \theta_i}^{i,\Delta}|^p] \leq p\mathbb{E}\left[\int_0^{t \wedge \theta_i} |X_s^i - X_s^{i,\Delta}|^{p-2} \left(X_s^i - X_s^{i,\Delta}\right) \cdot \left(a(X_s^i, \mu_s^X) - a(\hat{X}_s^{i,\Delta}, \hat{\mu}_s^{X,\Delta})\right) ds \right]$$

$$+ \frac{p(p-1)}{2} \mathbb{E}\left[\int_0^{t \wedge \theta_i} |X_s^i - X_s^{i,\Delta}|^{p-2} \|b(X_s^i, \mu_s^X) - b(\hat{X}_s^{i,\Delta}, \hat{\mu}_s^{X,\Delta})\|^2 ds \right],$$

By simple splitting,

$$\left(X_s^i - X_s^{i,\Delta}\right) \cdot \left(a(X_s^i, \mu_s^X) - a(\hat{X}_s^{i,\Delta}, \hat{\mu}_s^{X,\Delta})\right) + \frac{p-1}{2} \|b(X_s^i, \mu_s^X) - b(\hat{X}_s^{i,\Delta}, \hat{\mu}_s^{X,\Delta})\|^2 \leq A_1 + A_2,$$

with

$$A_1 = \left(X_s^i - X_s^{i,\Delta}\right) \cdot \left(a(X_s^i, \mu_s^X) - a(X_s^{i,\Delta}, \mu_s^{X,\Delta})\right) + \frac{p'-1}{2} \|b(X_s^i, \mu_s^X) - b(X_s^{i,\Delta}, \mu_s^{X,\Delta})\|^2$$

and

$$A_2 = \left(X_s^i - X_s^{i,\Delta}\right) \cdot \left(a(X_s^{i,\Delta}, \mu_s^{X,\Delta}) - a(\hat{X}_s^{i,\Delta}, \hat{\mu}_s^{X,\Delta})\right) + C(p', p) \|b(X_s^{i,\Delta}, \mu_s^{X,\Delta}) - b(\hat{X}_s^{i,\Delta}, \hat{\mu}_s^{X,\Delta})\|^2.$$

Here, we can take any $p' \in (p, p_0)$.

According to Assumption 2.3.

$$A_1 \leq C \left(|X_s^i - X_s^{i,\Delta}|^2 + \mathcal{W}_2^2(\mu_s^X, \mu_s^{X,\Delta}) \right).$$

The terms in A_2 can be controlled using Assumption 2.2:

$$A_2 \lesssim |X_s^i - X_s^{i,\Delta}|^2 + (1 + |X_s^{i,\Delta}|^\gamma + |\hat{X}_s^{i,\Delta}|^\gamma)^2 |X_s^{i,\Delta} - \hat{X}_s^{i,\Delta}|^2 + \mathcal{W}_2^2(\mu_s^{X,\Delta}, \hat{\mu}_s^{X,\Delta}).$$

Applying Young's inequality, one has

$$\mathbb{E}[|X_{t\wedge\theta_i}^i - X_{t\wedge\theta_i}^{i,\Delta}|^p] \lesssim \left\{ \mathbb{E}\left[\int_0^{t\wedge\theta_i} |X_s^i - X_s^{i,\Delta}|^p ds \right] + \mathbb{E}\left[\int_0^{t\wedge\theta_i} \mathcal{W}_2^p(\mu_s^X, \mu_s^{X,\Delta}) ds \right] \right\}$$
$$+ \mathbb{E}\left[\int_0^{t\wedge\theta_i} (1 + |X_s^{i,\Delta}|^\gamma + |\hat{X}_s^{i,\Delta}|^\gamma)^p |X_s^{i,\Delta} - \hat{X}_s^{i,\Delta}|^p ds \right]$$
$$+ \mathbb{E}\left[\int_0^{t\wedge\theta_i} \mathcal{W}_2^p(\mu_s^{X,\Delta}, \hat{\mu}_s^{X,\Delta}) ds \right] =: J_1 + J_2 + J_3.$$

By Lemma 2.1, the Minkowski inequality, and Lemma 3.2

$$J_3 \leq \mathbb{E}\left[\int_0^t \left(\frac{1}{N} \sum_{j=1}^N \left| X_s^{j,\Delta} - \hat{X}_s^{j,\Delta} \right|^2 \right)^{p/2} ds \right] \leq \int_0^t \left(\frac{1}{N} \sum_{j=1}^N (\mathbb{E}[|X_s^{j,\Delta} - \hat{X}_s^{j,\Delta}|^p])^{2/p} \right)^{p/2} ds \lesssim \Delta^{p/2} h^p(\Delta).$$

For J_2, by the Hölder inequality and Lemma 3.2 and Proposition 3.3, one has

$$J_2 \leq \mathbb{E}\left[\int_0^t (1 + |\hat{X}_s^{i,\Delta}|^\gamma + |X_s^{i,\Delta}|^\gamma)^p |X_s^{i,\Delta} - \hat{X}_s^{i,\Delta}|^p ds \right]$$
$$\lesssim \int_0^t \left(\mathbb{E}\left[(1 + |\hat{X}_s^{i,\Delta}|^\gamma + |X_s^{i,\Delta}|^\gamma)^{pq_1} \right] \right)^{1/q_1} \left(\mathbb{E}[|X_s^{i,\Delta} - \hat{X}_s^{i,\Delta}|^{p_1 p}] \right)^{1/p_1} \tag{3.9}$$
$$\leq C_T \Delta^{p/2} h^p(\Delta).$$

Here, $1/p_1 + 1/q_1 = 1$. Note that Lemma 3.2 holds for any $p_1 p \geq 2$ so we can choose p_1 large enough such that $p q_1 \gamma < p_0$. Such q_1 exists because $p\gamma < p_0$.

For J_1, we first note that

$$J_1 \leq \mathbb{E}\left[\int_0^t |X_{s\wedge\theta_i}^i - X_{s\wedge\theta_i}^{i,\Delta}|^p ds \right] + \mathbb{E}\left[\int_0^{t\wedge\theta_i} \mathcal{W}_2^p(\mu_s^X, \mu_s^{X,\Delta}) ds \right],$$

because for $s > \theta_i$, the integrand is nonnegative. Then again, by Lemma 2.1, the Minkowski inequality and the symmetry of the particles,

$$\mathbb{E}\left[\int_0^{t\wedge\theta_i} \mathcal{W}_2^p(\mu_s^X, \mu_s^{X,\Delta}) ds \right] \leq \mathbb{E}\left[\int_0^t \mathcal{W}_2^p(\mu_s^X, \mu_s^{X,\Delta}) ds \right]$$
$$\leq \int_0^t \left(\frac{1}{N} \sum_{j=1}^N (\mathbb{E}[|X_s^j - X_s^{j,\Delta}|^p])^{2/p} \right)^{p/2} ds$$
$$= \int_0^t \mathbb{E}[|X_s^i - X_s^{i,\Delta}|^p] ds.$$

Here, θ_i has been thrown away because it is not suitable for stopping other particles X_s^j.

Therefore,

$$\mathbb{E}[|X_{t\wedge\theta_i}^i - X_{t\wedge\theta_i}^{i,\Delta}|^p] \lesssim \int_0^t \mathbb{E}[|X_{s\wedge\theta_i}^i - X_{s\wedge\theta_i}^{i,\Delta}|^p]ds + \int_0^t \mathbb{E}[|X_s^i - X_s^{i,\Delta}|^p]ds + \Delta^{p/2}h^p(\Delta).$$

Note that for all $s \in [0,t]$,

$$\begin{aligned}
\mathbb{E}[|X_s^i - X_s^{i,\Delta}|^p] &= \mathbb{E}[|X_{s\wedge\theta_i}^i - X_{s\wedge\theta_i}^{i,\Delta}|^p\mathcal{I}_{\{\theta_i>T\}}] + \mathbb{E}[|X_s^i - X_s^{i,\Delta}|^p\mathcal{I}_{\{\tau_i\leq T \text{ or } \rho_i\leq T\}}]. \\
&\leq \mathbb{E}[|X_{s\wedge\theta_i}^i - X_{s\wedge\theta_i}^{i,\Delta}|^p] + \mathbb{E}[|X_s^i - X_s^{i,\Delta}|^p\mathcal{I}_{\{\tau_i\leq T \text{ or } \rho_i\leq T\}}].
\end{aligned} \tag{3.10}$$

Here, recall that $\mathcal{I}_G$ is the indicator function of the set G. By Young's inequality:

$$\mathbb{E}[|X_s^i - X_s^{i,\Delta}|^p\mathcal{I}_{\{\tau_i\leq T \text{ or } \rho_i\leq T\}}] \leq \frac{\delta}{p_2}\mathbb{E}[|X_s^i - X_s^{i,\Delta}|^{pp_2}] + \frac{1}{q_2\delta^{1/(p_2-1)}}\mathbb{P}(\tau_i \leq T \text{ or } \rho_i \leq T).$$

Where $p_2 > 1$ is some positive constant such that $pp_2 < p_0$. The expectation in the first term is then bounded. By the moment estimation of $\bar{X}_t^i$ and $X_t^{i,\Delta}$ and the Markov inequality

$$\mathbb{P}(\tau_i \leq T \text{ or } \rho_i \leq T) \leq \frac{\mathbb{E}[|\bar{X}_t^i|^q] + \mathbb{E}[|X_t^{i,\Delta}|^q]}{(\varphi^{-1}\circ h(\Delta))^q},$$

for $q < p_0$. Denote $\eta(\Delta) = (\varphi^{-1}\circ h(\Delta))^{-1}$. Balancing the terms, we may choose $pp_2 = q$ and $\delta \sim \delta^{-1/(p_2-1)}\eta(\Delta)^q$. Consequently,

$$\delta \sim \eta(\Delta)^{q-p}.$$

Therefore,

$$\mathbb{E}[|X_t^i - X_t^{i,\Delta}|^p] \lesssim \int_0^t \mathbb{E}[|X_s^i - X_s^{i,\Delta}|^p]ds + \Delta^{p/2}h^p(\Delta) + \eta(\Delta)^{q-p}.$$

Hence $\forall t \in [0,T]$,

$$\sup_{0\leq s\leq t}\mathbb{E}[|X_s^i - X_s^{i,\Delta}|^p] \leq \int_0^t \sup_{0\leq u\leq s}\mathbb{E}[|X_u^i - X_u^{i,\Delta}|^p]ds + C_T\Delta^{p/2}h^p(\Delta) + C_T\eta(\Delta)^{q-p}.$$

Applying the Grönwall inequality yields

$$\sup_{0\leq t\leq T}\mathbb{E}[|X_t^i - X_t^{i,\Delta}|^p] \leq C_T\Delta^{p/2}h^p(\Delta) + C_T\eta(\Delta)^{q-p}.$$

Combining with the propagation of chaos results (Prop. 2.5), one therefore obtains the eventual estimate as listed in the statement. $\square$

Now, we perform some discussions. First of all, we note that if b is globally Lipschitz and a is one-sided Lipschitz, then p_0 can be arbitrarily large. Then, $\eta(\Delta)$ can be chosen as polynomial of Δ. For large enough q, the last term can be dropped. Hence, one has

Corollary 3.5. *If b is globally Lipschitz and a is one-sided Lipschitz, then*

$$\sup_{t\leq T}\mathbb{E}[|\bar{X}_t^i - X_t^{i,\Delta}|^p] \leq C \begin{cases} N^{-1/2} + \Delta^{p/2}h(\Delta)^p, & \text{if } p > d/2, \\ N^{-1/2}\log(1+N) + \Delta^{p/2}h(\Delta)^p, & \text{if } p = d/2, \\ N^{-p/d} + \Delta^{p/2}h(\Delta)^p, & \text{if } p \in [2, d/2). \end{cases} \tag{3.11}$$

In this case, one may choose $h(\Delta)$ that grows slowly as $\Delta \to 0$ like $h(\Delta) \sim \Delta^{-\epsilon}$ for very small ϵ. This indicates that the order of the strong error can be arbitrarily close to one half.

However, if p_0 has an upper bound, choosing $h(\Delta) \sim \Delta^{-\epsilon}$ with small ϵ leads to big bound for the last term $\eta(\Delta)^{q/p-1}$. In this case, one needs to choose suitable h to optimize the rate.

4. The random batch approximation for the interacting particle systems and its error analysis

The computational cost for the simulation of the particle systems in (1.3) and (3.4) is high due to the $\mathcal{O}(N^2)$ complexity per time step. In this section, we extend the recently proposed RBM [18–21] to our particle systems and establish the approximation error estimate.

The RBM is based on the so-called "random mini-batch" idea, which is famous for its application in the stochastic gradient descent (SGD) [33] algorithm for optimization in machine learning. The key methodology of "mini-batch" is to find a cheap and unbiased random estimator using small subset of data/particles for the original quantity. How to design the random batch estimator clearly depend on the applications. In the RBM for particle systems, a random grouping strategy was proposed in [19–21], while an importance sampling in the Fourier space was proposed for the Random Batch Ewald method for molecular dynamics in [19, 22]. Note that for each single step, the random estimator has $\mathcal{O}(1)$ error for the quantity. The key reason for the method to converge is the error cancelation in time. It is this type of Law of Large Numbers (in time) that ensures convergence. A difference of the RBM method from SGD is that the method is designed to dynamical properties of the systems, not just for equilibrium distribution.

In the original work [20],

$$a(x, \mu) = f(x) + \int_{\mathbb{R}^d} k(x, y)\mu(dy), \quad b(x, \mu) \equiv \text{const.}$$

Under this setting, the interacting particle systems becomes

$$dX_t^i = f(X_t^i)\, dt + \frac{1}{N-1} \sum_{j \neq i} k(X_t^i, X_t^j) dt + b\, dW_t^i. \tag{4.1}$$

The initial values X_0^i are i.i.d. sampled from μ_0. Here, we have used $\frac{1}{N-1} \sum_{j \neq i} \delta_{X_t^i}$ to approximate μ, and there is no significance difference regarding the propagation of chaos.

Suppose we aim to do simulation until time $T > 0$. RBM does the following. Choose a batch size $P \ll N, P \geq 2$ that divides N. For the time step Δ and $t_m := m\Delta$, on each time subinterval $[t_m, t_{m+1})$, there are two steps: (1) at time grid t_m, we divide the N particles into $n := N/P$ groups (batches) randomly; (2) the particles evolve with interaction inside the batches only.

Algorithm 1 (RBM)**.**

1: **for** m in $1 : [T/\Delta]$ **do**
2:     Divide $\{1, 2, \ldots, N\}$ into $n = N/P$ batches randomly.
3:     **for** each batch $\mathcal{C}_q$ **do**
4:         Update $\tilde{X}_t^i$'s ($i \in \mathcal{C}_q$) by solving the following modified system with $t \in [t_m, t_{m+1})$.

$$d\tilde{X}_t^i = f(\tilde{X}_t^i)\, dt + \frac{1}{P-1} \sum_{j \in \mathcal{C}_q, j \neq i} k(\tilde{X}_t^i, \tilde{X}_t^j) dt + b\, dW_t^i. \tag{4.2}$$

5:     **end for**
6: **end for**

In the case above, the dependence in μ is linear in the drift. As proved in [20], for a fixed configuration, the random batch approximation in the drift is unbiased and there is no restriction on batch size P.

Below we extend the random batch approximation to general coefficients where the diffusion coefficient could also depend on μ and they could be nonlinear in μ. In general, if there is dependence of μ in b, the quadratic variation in the Itô's formula would cause a contributing term of the order $1/P$ to the mean square error (or

$\sqrt{1/P}$ to the error). Moreover, if the dependence on μ is nonlinear, the random approximation will no longer be unbiased for a fixed configuration. Consequently, it would also contribute an error that does not vanish as $\Delta \to 0$. All these would put restrictions on the batch size P. For example, $P \sim \Delta^{-1/2}$ to achieve half order for the mean square error. However, since the selection of batch size does not grow as the number of particles $N \to \infty$, the random batch approximation could still be beneficial.

We first consider a special case where the dependence on μ is linear to investigate how the random batch approximation in diffusion coefficient contributes to the error:

$$a(x,\mu) = f(x) + \int_{\mathbb{R}^d} k(x,y)\mu(dy), \quad b(x,\mu) = \int_{\mathbb{R}^d} \sigma(x,y)\mu(dy). \tag{4.3}$$

In particular, we investigate the bias introduced by the distribution dependence in the diffusion coefficient, which as we shall see, is already very different from [20] regarding the application of RBM.

Secondly, we would consider drifts where the dependence on μ is nonlinear to investigate the error introduced by the random batch approximation:

$$a(x,\mu) = f(x) + A\left(\int k(x,y)\mu(dy)\right), \quad b(x,\mu) \equiv \sigma(x). \tag{4.4}$$

Here, $A : \mathbb{R}^r \to \mathbb{R}^d$ for some $r \geq 1$ is a nonlinear field.

Lastly, we perform discussion on general cases. In principle, the analysis would be similar to the cases discussed. In this section, all the discussion on the random batch system will be performed by keeping the time continuous. Combining with the truncated scheme in the previous section, one will obtain the eventual method for numerical simulation.

4.1. The particle system and the random batch approximation

The interacting particle system (1.3) for (4.3) is given by

$$dX_t^i = f(X_t^i)\,dt + \frac{1}{N}\sum_{i=1}^{N} k(X_t^i, X_t^j)dt + \frac{1}{N}\sum_{i=1}^{N} \sigma(X_t^i, X_t^j)dW_t^i,$$

or

$$dX_t^i = f(X_t^i)\,dt + \frac{1}{N-1}\sum_{j\neq i} k(X_t^i, X_t^j)\,dt + \frac{1}{N-1}\sum_{j\neq i} \sigma(X_t^i, X_t^j)dW_t^i. \tag{4.5}$$

The initial values $X_0^i \in \mathbb{R}^d$ are the same as (1.3). The difference is whether we use $\frac{1}{N}\sum_j \delta_{X_t^j}$ or $\frac{1}{N-1}\sum_{j\neq i} \delta_{X_t^j}$ for the approximation. Note that there is no big difference regarding the propagation of chaos while $1/(N-1)$ is more convenient in notation for the random batch system. We will focus on (4.5) then. Similarly, for the case (4.4), the interacting particle system is then given by

$$dX_t^i = f(X_t^i)\,dt + A\left(\frac{1}{N-1}\sum_{j\neq i} k(X_t^i, X_t^j)\right)dt + \sigma(X_t^i)dW_t^i. \tag{4.6}$$

Recall that for the random batch approximation, we divide the $N = nP$ particles into n small batches with size $P \geq 2$ randomly. Let the random batches be $\mathcal{C}_{q(i)}$, $q = 1, \ldots, n$. We will use

$$\mathcal{C} := \{\mathcal{C}_1, \cdots, \mathcal{C}_n\}$$

to represent the division of random batches. Let $q(i)$ be the index q such that $i \in \mathcal{C}_q$. For the random batch approximation of (4.5), we generate two random batches $\mathcal{C}$ and $\mathcal{B}$ at t_m and run the following equation

$$d\tilde{X}_t^i = f(\tilde{X}_t^i)\,dt + \frac{1}{P-1}\sum_{j\in\mathcal{C}_{q_1(i)},j\neq i} k(\tilde{X}_t^i, \tilde{X}_t^j)dt + \frac{1}{P-1}\sum_{j\in\mathcal{B}_{q_2(i)},j\neq i} \sigma(\tilde{X}_t^i, \tilde{X}_t^j)dW_t^i, \quad t \in [t_m, t_{m+1}), \tag{4.7}$$

where $q_1(i)$ and $q_2(i)$ are used to distinguish the batchmates in the two different batches $\mathcal{B}$ and $\mathcal{C}$. And the initial values are $\tilde{X}_0^i = X_0^i$, the same as the particle system (1.3).

Correspondingly, the random batch approximation for (4.6)

$$d\tilde{X}_t^i = f(\tilde{X}_t^i)\,dt + A\left(\frac{1}{P-1}\sum_{j\in\mathcal{C}_{q_m(i)},j\neq i} k(\tilde{X}_t^i, \tilde{X}_t^j)\right) dt + \sigma(\tilde{X}_t^i)dW_t^i, \tag{4.8}$$

The initial values are $\tilde{X}_0^i = X_0^i$, the same as the particle system (1.3).

Our experience tells us that if N is not so small, there would be no significant difference if $\mathcal{C}$ and $\mathcal{B}$ are independent or the same. Hence, for the convenience of the notations and analysis, we will always assume from here on that

$$\mathcal{C} = \mathcal{B}.$$

Namely, the random batches used for the drift and the diffusion are the same. To distinguish the batches at different time intervals, we denote the random division of batches at t_m by

$$\mathcal{C}^{(m)} := \{\mathcal{C}_1^{(m)}, \cdots, \mathcal{C}_n^{(m)}\}.$$

Define the filtration $\{\mathcal{F}_m\}_{m\geq 0}$ by

$$\mathcal{F}_m = \sigma\left(\mathcal{C}^{(j)} : j \leq m\right).$$

This contains the information for the batches up to t_m. Also,

$$\mathcal{F} := \sigma(\cup_m \mathcal{F}_m)$$

is the σ-algebra for how batches are generated at all time points.

4.2. The linear dependence case

In this section, we consider the special cases where the dependence on the distribution is linear and analyze the random batch approximation (4.7). For this purpose, we need stronger assumptions imposed on the coefficients than the one in Assumption 2.3. These conditions are still non-globally Lipschitz but the essential part is Lipschitz. Specifically, we will investigate the problem with the following assumptions.

Assumption 4.1. $f : \mathbb{R}^d \to \mathbb{R}^d$ *is one-sided Lipschitz in the sense that*

$$(x - y) \cdot (f(x) - f(y)) \leq L|x - y|^2, \quad \forall x, y, \tag{4.9}$$

and the derivatives have polynomial growth. $k : \mathbb{R}^d \times \mathbb{R}^d \to \mathbb{R}^d$ *and* $\sigma : \mathbb{R}^d \times \mathbb{R}^d \to \mathbb{R}^{d\times m'}$ *are all Lipschitz continuous.*

Remark 4.2. The coefficient $b(x, \mu) = \int \sigma(x, y)\mu(dy)$ could also have a single term depending on x, like $\sigma(x, y) = g(x) + \sigma_1(x, y)$. In principle, $g(x)$ does not have to be globally Lipschitz. We assume the Lipschitz continuity of σ simply for the convenience of discussion.

Proposition 4.3. *Let Assumption 4.1 hold and* $\mu_0 \in \mathcal{P}_q$ *with* $q \geq 2$. *Let* X_t^j *be the solution to (4.5) and* $\tilde{X}_t^j$ *be the solution to (4.7) for* $j = 1, \cdots, N$. *Then conditioning on the sequence of random batches, one has*

$$\sup_{0\leq t\leq T} \max_{1\leq j\leq N} \mathbb{E}[|\tilde{X}_t^j|^q|\mathcal{F}] \leq C_q. \tag{4.10}$$

Similarly,

$$\sup_{0\leq t\leq T} \left(\mathbb{E}[|X_t^i|^q] + \mathbb{E}[|\tilde{X}_t^i|^q]\right) = \sup_{0\leq t\leq T} \left(\mathbb{E}[|X_t^1|^q] + \mathbb{E}[|\tilde{X}_t^1|^q]\right) \leq C_{q,T}, \forall i. \tag{4.11}$$

In both estimates, the constant $C_{q,T}$ *is independent of* N *and the sequence of batches.*

Proof. We only show the moment control for $\tilde{X}_t^i$ because the moment control for X_t^i is similar. The proof is essentially the same as Lemma 3.4 of [18]. Our approach is to show the boundedness of the moments for any given sequences of the batches, to avoid the interplay between randomness of the batches and the process $\tilde{X}_t$.

For $t \in [t_m, t_{m+1})$, by Itô's formula one has

$$\frac{d}{dt}\mathbb{E}[|\tilde{X}_t^i|^q|\mathcal{F}] = q\mathbb{E}\left[|\tilde{X}_t^i|^{q-2}\tilde{X}_t^i \cdot \left(f(\tilde{X}_t^i) + \frac{1}{P-1}\sum_{j\in\mathcal{C}_{q_m(i)},j\neq i}k(\tilde{X}_t^i,\tilde{X}_t^j)\right)\bigg|\mathcal{F}\right]$$

$$+\frac{1}{2}q\mathbb{E}\left[|\tilde{X}_t^i|^{q-2}\left(I + (q-2)\frac{\tilde{X}_t^i \otimes \tilde{X}_t^i}{|\tilde{X}_t^i|^2}\right) : \left(\frac{1}{P-1}\sum_{j\in\mathcal{C}_{q_m(i)},j\neq i}\sigma(\tilde{X}_t^i,\tilde{X}_t^j)\right)\left(\frac{1}{P-1}\sum_{j\in\mathcal{C}_{q_m(i)},j\neq i}\sigma(\tilde{X}_t^i,\tilde{X}_t^j)\right)^T\bigg|\mathcal{F}\right]$$

$$=: M_1 + M_2.$$

By Assumption 4.1,

$$\tilde{X}_t^i \cdot f(\tilde{X}_t^i) \leq L|\tilde{X}_t^i|^2 + |f(0)||\tilde{X}_t^i|,$$

$$|\tilde{X}_t^i|^{q-2}\tilde{X}_t^i \cdot k(\tilde{X}_t^i,\tilde{X}_t^j)) \leq L((2-q^{-1})|\tilde{X}_t^i|^q + q^{-1}|\tilde{X}_t^j|^q) + C|\tilde{X}_t^i|^{q-1}.$$

One thus has

$$M_1 \lesssim 1 + \mathbb{E}[|\tilde{X}_t^i|^q|\mathcal{F}] + \frac{1}{P-1}\sum_{j\in\mathcal{C}_{q_m(i)},j\neq i}\mathbb{E}(|\tilde{X}_t^j|^q|\mathcal{F}) \lesssim 1 + \max_{1\leq j\leq N}\mathbb{E}(|\tilde{X}_t^j|^q|\mathcal{F}).$$

Similarly,

$$M_2 \leq q(q-1)\frac{1}{(P-1)^2}\sum_{j\in\mathcal{C}_{q_m(i)},j'\in\mathcal{C}_{q_m(i)},j\neq i,j'\neq i}\mathbb{E}(|\tilde{X}_t^i|^{q-2}\|\sigma(\tilde{X}_t^i,\tilde{X}_t^j)\sigma(\tilde{X}_t^i,\tilde{X}_t^{j'})^T\||\mathcal{F}).$$

Again, using the Lipschitz continuity of σ and the Young's inequality like

$$|\tilde{X}_t^i|^{q-2}|\tilde{X}_t^j||\tilde{X}_t^{j'}| \leq \frac{q-2}{q}|\tilde{X}_t^i|^q + \frac{1}{q}|\tilde{X}_t^j|^q + \frac{1}{q}|\tilde{X}_t^{j'}|^q,$$

one will get the same bound as in M_1. Consequently, by gluing together all the estimates on different subintervals of the form $[t_m, t_{m+1})$, one has

$$\mathbb{E}[|\tilde{X}_t^i|^q|\mathcal{F}] \leq \mathbb{E}[|\tilde{X}_0^i|^q|\mathcal{F}] + C\int_0^t(1 + \max_{1\leq j\leq N}\mathbb{E}[|\tilde{X}_s^j|^q|\mathcal{F}])\,ds,$$

for any i. Hence, defining

$$a(t) := \max_{1\leq j\leq N}\mathbb{E}[|\tilde{X}_t^j|^q|\mathcal{F}],$$

one has the inequality

$$a(t) \leq a(0) + C\int_0^t a(s)\,ds.$$

Applying Grönwall's inequality yields the result for $\mathbb{E}[|\tilde{X}_t^j|^q|\mathcal{F}]$. Then, taking expectation over the random batches, the result for the moments also follows. $\square$

Below, for a given configuration, $\mathbf{x} := (x^1, x^2, \cdots, x^N) \in \mathbb{R}^{Nd}$. Define the random deviation in a random batch approximation, with interaction kernel k:

$$\chi_i(\mathbf{x}; k) := \frac{1}{P-1}\sum_{j\in\mathcal{C}_{q_m(i)},j\neq i}k(x^i,x^j) - \frac{1}{N-1}\sum_{j\neq i}k(x^i,x^j). \tag{4.12}$$

The following lemma, shown in Lemma 3.1 of [20], gives the basic consistency and stability features of the RBM.

Lemma 4.4. *The random deviation satisfies*

$$\mathbb{E}[\chi_i(\mathbf{x}; k)] = 0$$

and

$$\mathrm{var}\,\chi_i(\mathbf{x}; k) = \mathbb{E}[\|\chi_i(\mathbf{x}; k)\|^2] = \left(\frac{1}{P-1} - \frac{1}{N-1}\right)\Lambda_i(\mathbf{x}; k), \tag{4.13}$$

where

$$\Lambda_i(\mathbf{x}; k) := \frac{1}{N-2}\sum_{j:j\neq i}\left\|k(x^i, x^j) - \frac{1}{N-1}\sum_{j':j'\neq i}k(x^i, x^{j'})\right\|^2.$$

We remark that if k is a matrix-valued function, the norm in this lemma is the Hilbert-Schmidt norm. Let us briefly understand how the random batch approximation in the diffusion coefficient affects the error. If we define

$$Z_t^i := \tilde{X}_t^i - X_t^i. \tag{4.14}$$

Then, one finds that

$$dZ_t^i = (f(\tilde{X}_t^i) - f(X_t^i))\,dt + \left(\frac{1}{P-1}\sum_{j\in\mathcal{C}_{q_m(i)}, j\neq i}k(\tilde{X}_t^i, \tilde{X}_t^j) - \frac{1}{N-1}\sum_{j\neq i}k(X_t^i, X_t^j)\right)dt$$

$$+ \left(\frac{1}{P-1}\sum_{j\in\mathcal{C}_{q_m(i)}, j\neq i}\sigma(\tilde{X}_t^i, \tilde{X}_t^j) - \frac{1}{N-1}\sum_{j\neq i}\sigma(X_t^i, X_t^j)\right)dW_t^i.$$

Intuitively, if X_t^i is close to $\tilde{X}_t^i$, then the leading error would be

$$dZ_t^i \sim \chi_i(\mathbf{X}; k)dt + \chi_i(\mathbf{X}; \sigma)dW_t^i,$$

where $\mathbf{X} := (X^1, X^2, \cdots, X^N) \in \mathbb{R}^{Nd}$. Then, by Itô's formula,

$$d\mathbb{E}[|Z_t^i|^2] \sim 2\mathbb{E}[Z_t^i \cdot \chi_i(\mathbf{X}; k)]\,dt + \mathbb{E}[|\chi_i(\mathbf{X}; \sigma)|^2]\,dt.$$

The batch at t_m is independent of $Z_{t_m}^i$ so conditioning on $Z_{t_m}^i$, the expectation of χ_i is zero, then the term $Z_t^i \cdot \chi_i$ would contribute a term that vanishes as $\Delta \to 0$. The second term, however, is of order $1/P$. This would not vanish as $\Delta \to 0$ if we fix P. This term is the quadratic variation due to the Brownian motion. Clearly, when there is the distribution dependence in the diffusion coefficient, the effect of random batch would be very different from the ones discussed in [20].

The following result gives the basic estimate for the approximation error.

Theorem 4.5. *Let Assumption 4.1 hold.* $\mathbf{X}(t) := (X_t^1, \cdots, X_t^N)$ *are the solution of* (4.5) *and* $\tilde{\mathbf{X}}(t) := (\tilde{X}_t^1, \cdots, \tilde{X}_t^N)$ *are the solution of* (4.7). *Then, the error process defined in* (4.14) *satisfies*

$$\sup_{0\leq t\leq T}\max_i \mathbb{E}[|Z_t^i|^2] \leq C\left(\Delta^2 + \frac{\Lambda(k)\Delta}{P} + \frac{\Lambda(\sigma)}{P}\right), \tag{4.15}$$

where $\Lambda(k) := \sup_{0\leq t\leq T}\mathbb{E}[\Lambda_i(\mathbf{X}(t); k)] < \infty$ *and* $\Lambda(\sigma) := \sup_{0\leq t\leq T}\mathbb{E}[\Lambda_i(\mathbf{X}(t); \sigma)] < \infty$ *are independent of i and can be bounded independent of N. In particular,*

$$\sup_{0\leq t\leq T}\mathbb{E}[|Z_t^i|^2] \leq \begin{cases} C_T(\frac{\Delta}{P} + \Delta^2), & \sigma(x, y) \equiv \sigma(x), \\ C_T(\frac{1}{P} + \Delta^2), & otherwise. \end{cases} \tag{4.16}$$

Clearly, when $\sigma(x,y) \equiv \sigma(x)$, the result (4.16) is the same as in [20], though they assumed boundedness of the kernel $k(x,y)$. For general σ that may depend on y, a fixed batch size P clearly would not give convergence. Instead, we need to take P large to get convergence.

Corollary 4.6. *If one takes $P = \min(\Delta^{-\beta}, N)$ with $\beta \leq 1$, which stays finite as $N \to \infty$, the error is controlled as*

$$\sqrt{\mathbb{E}[|Z_t^i|^2]} \leq \begin{cases} C\Delta^{(\beta+1)/2}, & \sigma(x,y) \equiv \sigma(x), \\ C\Delta^{\beta/2}, & otherwise. \end{cases} \tag{4.17}$$

To prove Theorem 4.5, we need the following symmetry fact.

Lemma 4.7. *Let Z_t^i be given in (4.14), with X_0^i being i.i.d. sampled. Let $\varphi(\cdot)$ be a test function that has at most polynomial growth. Then, for $[t_m, t_{m+1})$, one has*

$$\mathbb{E}\left[\mathbb{E}\Big[\frac{1}{P}\sum_{j \in \mathcal{C}_{q_m(i)}} \varphi(Z_t^j)|\mathcal{F}_m\Big]\right] = \frac{1}{N}\sum_{j'=1}^{N} \mathbb{E}[\varphi(Z_t^{j'})] = \mathbb{E}[\varphi(Z_t^j)], \quad \forall j \in \{1, \cdots, N\}.$$

Similarly, for any particle i,

$$\mathbb{E}\left[\mathbb{E}\Big[\frac{1}{P-1}\sum_{j \in \mathcal{C}_{q_m(i)}, j \neq i} \varphi(Z_t^j)|\mathcal{F}_m\Big]\right] = \frac{1}{N-1}\sum_{j':j' \neq i} \mathbb{E}[\varphi(Z_t^{j'})] = \mathbb{E}[\varphi(Z_t^j)], \quad \forall j \in \{1, \cdots, N\}.$$

Proof. Note that $\mathbb{E}[\mathbb{E}[\cdot|\mathcal{F}_m]] = \mathbb{E}[\mathbb{E}[\cdot|\sigma(\mathcal{C}^{(m)})]]$. Then, by definition,

$$\mathbb{E}\left[\mathbb{E}\left[\frac{1}{P}\sum_{i \in \mathcal{C}_{q_m(i)}} \varphi(Z_t^i)|\sigma(\mathcal{C}^{(m)})\right]\right]$$

$$= \sum_{C_1,\cdots,C_n} \mathbb{P}(\mathcal{C}_1 = C_1, \cdots, \mathcal{C}_n = C_n)\frac{1}{P}\sum_{j'=1}^{N} I_{j' \in \mathcal{C}_{q_m(i)}}\mathbb{E}[\varphi(Z_t^{j'})|\mathcal{C}_1 = C_1, \cdots, \mathcal{C}_n = C_n]$$

$$= \frac{1}{P}\sum_{j'=1}^{N}\sum_{C_1,\cdots,C_n} \mathbb{P}(\mathcal{C}_1 = C_1, \cdots, \mathcal{C}_n = C_n)I_{j' \in \mathcal{C}_{q_m(i)}}\mathbb{E}[\varphi(Z_t^{j'})|\mathcal{C}_1 = C_1, \cdots, \mathcal{C}_n = C_n].$$

Here, we make of the symmetry. The point is that $\mathbb{E}[\varphi(Z^{j'}(t))|\mathcal{C}_1 = C_1, \cdots, \mathcal{C}_n = C_n]$ does not depend on the batch (the values $C_1, \cdots, C_n$). In other words, no matter what its batchmates are, this conditional expectation would be the same because the joint distribution of the particles is symmetric at t_m. Hence,

$$\sum_{C_1,\cdots,C_n} \mathbb{P}(\mathcal{C}_1 = C_1, \cdots, \mathcal{C}_n = C_n)I_{j' \in C_q}\mathbb{E}[\varphi(Z_t^{j'})|\mathcal{C}_1 = C_1, \cdots, \mathcal{C}_n = C_n]$$

$$= \mathbb{P}(j' \in \mathcal{C}_{q_m(i)})\mathbb{E}[\varphi(Z_t^{j'})] = \frac{P}{N}\mathbb{E}[\varphi(Z_t^{j'})].$$

The second claim is nearly the same. The only difference is that the batch $\mathcal{C}_{q_m(i)}$ used will vary. Let $\mathbf{1}_{ij'}$ be the indicator that particles i and j' are in the same batch. Then,

$$\sum_{C_1,\cdots,C_n} \mathbb{P}(\mathcal{C}_1 = C_1, \cdots, \mathcal{C}_n = C_n)\frac{1}{P-1}\sum_{j'=1}^{N} \mathbf{1}_{ij'}\mathbb{E}[\varphi(Z_t^{j'})|\mathcal{C}_1 = C_1, \cdots, \mathcal{C}_n = C_n]$$

$$= \frac{1}{P-1}\sum_{j'=1}^{N} \mathbb{P}(\mathbf{1}_{ij'} = 1)\sum_{C_1,\cdots,C_n} \mathbb{P}(\mathcal{C}_1 = C_1, \cdots, \mathcal{C}_n = C_n)\mathbb{E}[\varphi(Z_t^{j'})|\mathcal{C}_1 = C_1, \cdots, \mathcal{C}_n = C_n].$$

Applying Lemma 3.2 of [21] gives the result. $\square$

Remark 4.8. This lemma is different from Lemma 3.4 of [21], where the summands are required to be independent from the random batches. Here, we do not require the independence, but use the symmetry. Nevertheless the goal is the same: to deal with the random sum.

We rearrange the terms in dZ_t^i as

$$
dZ_t^i = \left(\int_0^1 \nabla f \left(X_t^i + \tau \left(\tilde{X}_t^i - X_t^i \right) \right) d\tau \right) \cdot Z_t^i
$$
$$
+ \frac{1}{P-1} \sum_{j \in \mathcal{C}_{q_m(i)}, j \neq i} \left(k \left(\tilde{X}_t^i, \tilde{X}_t^j \right) - k \left(X_t^i, X_t^j \right) \right) dt + \chi_i \left(\mathbf{X}(t); k \right) dt
$$
$$
+ \frac{1}{P-1} \sum_{j \in \mathcal{C}_{q_m(i)}, j \neq i} \left(\sigma \left(\tilde{X}_t^i, \tilde{X}_t^j \right) - \sigma \left(X_t^i, X_t^j \right) \right) dW_t^i + \chi_i \left(\mathbf{X}(t); \sigma \right) dW_t^i.
$$

Now, we can prove the main result.

Proof of Theorem 4.5. For $t \in [t_m, t_{m+1})$, by Itô's formula, one has

$$
\frac{d}{dt} \mathbb{E} \left[\left| Z_t^i \right|^2 | \mathcal{F}_m \right]
$$
$$
= \left(2\mathbb{E} \left[Z_t^i \cdot \left(f \left(X_t^i \right) - f(\tilde{X}_t^i) \right) \Big| \mathcal{F}_m \right] + 2\mathbb{E} \left[\frac{1}{P-1} \sum_{j \in \mathcal{C}_{q_m(i)}, j \neq i} Z_t^i \cdot \left(k \left(\tilde{X}_t^i, \tilde{X}_t^j \right) - k \left(X_t^i, X_t^j \right) \right) \Big| \mathcal{F}_m \right] \right)
$$
$$
+ 2\mathbb{E} \left[Z_t^i \cdot \chi_i \left(\mathbf{X}(t); k \right) | \mathcal{F}_m \right] + \mathbb{E} \left[\Sigma\Sigma^T | \mathcal{F}_m \right]
$$
$$
= S_1 + S_2 + S_3,
$$

where

$$
\Sigma = \frac{1}{P-1} \sum_{j \in \mathcal{C}_{q_m(i)}, j \neq i} \sigma(\tilde{X}_t^i, \tilde{X}_t^j) - \frac{1}{N-1} \sum_{j \neq i} \sigma(X_t^i, X_t^j).
$$

For S_1 term, using the one-sided Lipschitz condition of f and the Lipschitz continuity of k, one has

$$
S_1 \lesssim \mathbb{E} \left[\left| Z_t^i \right|^2 | \mathcal{F}_m \right] + \mathbb{E} \left[\frac{1}{P-1} \sum_{j \in \mathcal{C}_{q_m(i)}, j \neq i} |Z_t^i| \left(|Z_t^i| + |Z_t^j| \right) \Big| \mathcal{F}_m \right]
$$
$$
\lesssim \mathbb{E} \left[\left| Z_t^i \right|^2 | \mathcal{F}_m \right] + \mathbb{E} \left[\frac{1}{P-1} \sum_{j \in \mathcal{C}_{q_m(i)}, j \neq i} \left| Z_t^j \right|^2 \Big| \mathcal{F}_m \right].
$$

The S_3 term can be estimated similarly:

$$
S_3 \leq \mathbb{E} \left[\|\Sigma\|^2 | \mathcal{F}_m \right]
$$
$$
\leq 2\mathbb{E} \left[\left\| \frac{1}{P-1} \sum_{j \in \mathcal{C}_{q_m(i)}, j \neq i} \sigma(\tilde{X}_t^i, \tilde{X}_t^j) - \frac{1}{P-1} \sum_{j \in \mathcal{C}_{q_m(i)}, j \neq i} \sigma \left(X_t^i, X_t^j \right) \right\|^2 \Big| \mathcal{F}_m \right] + 2\mathbb{E} \left[\| \chi_i \left(\mathbf{X}(t); \sigma \right) \|^2 | \mathcal{F}_m \right]
$$
$$
=: S_{31} + S_{32}.
$$

Here, S_{32} is the main local error term arising from S_3, which will not vanish as $\Delta \to 0$. The S_{31} term can be estimated using the Lipschitz continuity of σ so that

$$S_{31} \lesssim \mathbb{E}\left[\left|Z_t^i\right|^2 |\mathcal{F}_m\right] + \mathbb{E}\left[\left|\frac{1}{P-1}\sum_{j\in\mathcal{C}_{q_m(i)},j\neq i}\left|Z_t^j\right|\right|^2 \Big|\mathcal{F}_m\right]$$

$$\leq \mathbb{E}\left[\left|Z_t^i\right|^2 |\mathcal{F}_m\right] + \mathbb{E}\left[\frac{1}{(P-1)^2}\sum_{j\in\mathcal{C}_{q_m(i)},j'\in\mathcal{C}_{q_m(i)},j,j'\neq i}\frac{1}{2}\left(\left|Z_t^j\right|^2 + \left|Z_t^{j'}\right|^2\right)\Big|\mathcal{F}_m\right]$$

$$= \mathbb{E}\left[\left|Z_t^i\right|^2 |\mathcal{F}_m\right] + \mathbb{E}\left[\frac{1}{P-1}\sum_{j\in\mathcal{C}_{q_m(i)},j\neq i}\left|Z_t^j\right|^2 \Big|\mathcal{F}_m\right].$$

Now, for S_2, due to Lemma 4.4 and the Lipschitz continuity of k, we break this term into the following.

$$S_2 \lesssim \left\{\mathbb{E}\left[Z_{t_m}^i \cdot \chi_i\left(\mathbf{X}(t);k\right)\Big|\mathcal{F}_m\right]\right\} + \left\{\mathbb{E}\left[\left|\chi_i\left(\mathbf{X}(t);k\right)\right|\int_{t_m}^t \left(1 + \left|\tilde{X}_s^i\right|^\gamma + |X_s^i|^\gamma\right)|Z_s^i|\,ds\Big|\mathcal{F}_m\right]\right.$$

$$\left.+\mathbb{E}\left[\left|\chi_i\left(\mathbf{X}(t);k\right)\right|\int_{t_m}^t \left(|Z_s^i| + \frac{1}{P-1}\sum_{j\in\mathcal{C}_{q_m(i)},j\neq i}|Z_s^j|\right)ds\Big|\mathcal{F}_m\right]\right\}$$

$$+ \left\{\mathbb{E}\left[\chi_i\left(\mathbf{X}(t);k\right)\int_{t_m}^t \frac{1}{P-1}\sum_{j\in\mathcal{C}_{q_m(i)},j\neq i}\left(\sigma\left(\tilde{X}_s^i,\tilde{X}_s^j\right) - \sigma\left(X_s^i,X_s^j\right)\right)dW_s^i\Big|\mathcal{F}_m\right]\right\}$$

$$+ \left\{\int_{t_m}^t \mathbb{E}\left[\chi_i\left(\mathbf{X}(s);k\right)\cdot\chi_i\left(\mathbf{X}(t);k\right)|\mathcal{F}_m\right]ds + \mathbb{E}\left[\chi_i\left(\mathbf{X}(t);k\right)\int_{t_m}^t \chi_i\left(\mathbf{X}(s);\sigma\right)dW_s^i\Big|\mathcal{F}_m\right]\right\}$$

$$=: S_{20} + S_{21} + S_{22} + S_{23}.$$

The terms around the brackets on the right side are called $S_{2,i}$, $i = 0, 1, 2, 3$ respectively for the convenience of notations. We will leave S_{20} as it is. Consider S_{21} first. The first term in S_{21} can be controlled using Proposition 4.3 by

$$\int_{t_m}^t \left(\mathbb{E}\left[\chi_i\left(\mathbf{X}(t);k\right)^4 |\mathcal{F}_m\right]\right)^{1/4}\left(\mathbb{E}\left[\left(1 + \left|\tilde{X}_s^i\right|^\gamma + |X_s^i|^\gamma\right)^4 |\mathcal{F}_m\right]\right)^{1/4}$$

$$\times \left(\mathbb{E}\left[\left|Z_s^i\right|^2 |\mathcal{F}_m\right]\right)^{1/2}ds \leq C\int_{t_m}^t \left(\mathbb{E}\left[\left|Z_s^i\right|^2 |\mathcal{F}_m\right]\right)^{1/2}ds.$$

The second term in S_{21} can be estimated similarly by

$$\int_{t_m}^t \left(\mathbb{E}\left[\left|Z_s^i\right|^2 |\mathcal{F}_m\right]\right)^{1/2}ds + \int_{t_m}^t \left(\mathbb{E}\left[\frac{1}{P-1}\sum_{j\in\mathcal{C}_{q_m(i)},j\neq i}\left|Z_s^j\right|^2 \Big|\mathcal{F}_m\right]\right)^{1/2}ds.$$

The S_{22} term can be estimated using the Lipschitz continuity of σ, Proposition 4.3 and the Itô's isometry by

$$\int_{t_m}^t \mathbb{E}\left[\left(|Z_s^i| + \frac{1}{P-1}\sum_{j\in\mathcal{C}_{q_m(i)},j\neq i}|Z_s^j|\right)^2 \Big|\mathcal{F}_m\right]^{1/2}ds$$

$$\leq \int_{t_m}^t \left(\mathbb{E}\left[|Z_s^i|^2|\mathcal{F}_m\right]\right)^{1/2}ds + \int_{t_m}^t \left(\mathbb{E}\left[\frac{1}{P-1}\sum_{j\in\mathcal{C}_{q_m(i)},j\neq i}\left|Z_s^j\right|^2 \Big|\mathcal{F}_m\right]\right)^{1/2}ds.$$

Here, S_{23} term contains essentially the local error term for random batch approximation arising from S_2. We may expect that this tends to zero as $\Delta \to 0$. This term can be simply estimated by

$$S_{23} \leq \frac{1}{2}\mathbb{E}\left[|\chi_i(\mathbf{X}(t);k)|^2 |\mathcal{F}_m\right]\Delta + \frac{1}{2}\int_{t_m}^t \mathbb{E}\left[|\chi_i(\mathbf{X}(s);k)|^2 |\mathcal{F}_m\right] ds$$
$$+ \mathbb{E}\left[|\chi_i(\mathbf{X}(t);k)|^2 |\mathcal{F}_m\right]^{1/2}\left(\int_{t_m}^t \mathbb{E}\left[\|\chi_i(\mathbf{X}(s);\sigma)\|^2 |\mathcal{F}_m\right] ds\right)^{1/2}.$$

Combining all these estimates together, we find

$$\mathbb{E}\left[\left|Z_t^i\right|^2 |\mathcal{F}_m\right]$$
$$\leq \mathbb{E}\left[\left|Z_{t_m}^i\right|^2 |\mathcal{F}_m\right] + C\int_{t_m}^t \mathbb{E}\left[\left|Z_s^i\right|^2 |\mathcal{F}_m\right] ds + C\int_{t_m}^t \mathbb{E}\left[\frac{1}{P-1}\sum_{j\in\mathcal{C}_{q_m(i)},j\neq i}\left|Z_s^j\right|^2 |\mathcal{F}_m\right] ds$$
$$+ C\int_{t_m}^t\int_{t_m}^s \left(\mathbb{E}\left[\left|Z_\tau^i\right|^2 |\mathcal{F}_m\right]\right)^{1/2} d\tau ds + \int_{t_m}^t\int_{t_m}^s \left(\mathbb{E}\left[\frac{1}{P-1}\sum_{j\in\mathcal{C}_{q_m(i)},j\neq i}\left|Z_\tau^j\right|^2 |\mathcal{F}_m\right]\right)^{1/2} d\tau ds$$
$$+ \int_{t_m}^t \mathbb{E}\left[Z_{t_m}^i \cdot \chi_i(\mathbf{X}(s);k) |\mathcal{F}_m\right] ds + \frac{\Delta}{2}\int_{t_m}^t \mathbb{E}\left[|\chi_i(\mathbf{X}(s);k)|^2 |\mathcal{F}_m\right] ds + \frac{1}{2}\int_{t_m}^t\int_{t_m}^s \mathbb{E}\left[|\chi_i(\mathbf{X}(\tau);k)|^2 |\mathcal{F}_m\right] d\tau ds$$
$$+ \int_{t_m}^t \mathbb{E}\left[|\chi_i(\mathbf{X}(s);k)|^2 |\mathcal{F}_m\right]^{1/2}\left(\int_{t_m}^s \mathbb{E}\left[\|\chi_i(\mathbf{X}(\tau);\sigma)\|^2 |\mathcal{F}_m\right] d\tau\right)^{1/2} ds + 2\int_{t_m}^t \mathbb{E}\left[\|\chi_i(\mathbf{X}(s);\sigma)\|^2 |\mathcal{F}_m\right] ds.$$

The double integrals can be estimated, for example, by

$$\int_{t_m}^t\int_{t_m}^s \left(\mathbb{E}\left[|Z_\tau^i|^2 |\mathcal{F}_m\right]\right)^{1/2} d\tau ds = \int_{t_m}^t (t-\tau)\left(\mathbb{E}\left[|Z_\tau^i|^2 |\mathcal{F}_m\right]\right)^{1/2} d\tau \leq \frac{\Delta^3}{6} + \frac{1}{2}\int_{t_m}^t \mathbb{E}\left[|Z_\tau^i|^2 |\mathcal{F}_m\right] d\tau,$$

and

$$\int_{t_m}^t \mathbb{E}\left[|\chi_i(\mathbf{X}(s);k)|^2 |\mathcal{F}_m\right]^{1/2}\left(\int_{t_m}^s \mathbb{E}\left[\|\chi_i(\mathbf{X}(\tau);\sigma)\|^2 |\mathcal{F}_m\right] d\tau\right)^{1/2} ds$$
$$\leq \frac{\Delta}{2}\int_{t_m}^t \mathbb{E}\left[|\chi_i(\mathbf{X}(s);k)|^2 |\mathcal{F}_m\right] ds + \frac{1}{2\Delta}\int_{t_m}^t (t-\tau)\mathbb{E}\left[\|\chi_i(\mathbf{X}(\tau);\sigma)\|^2 |\mathcal{F}_m\right] d\tau.$$

One actually has

$$\mathbb{E}\left[\left|Z_t^i\right|^2 |\mathcal{F}_m\right]$$
$$\leq \mathbb{E}\left[\left|Z_{t_m}^i\right|^2 |\mathcal{F}_m\right] + C\int_{t_m}^t \mathbb{E}\left[\left|Z_s^i\right|^2 |\mathcal{F}_m\right] ds + C\int_{t_m}^t \mathbb{E}\left[\frac{1}{P-1}\sum_{j\in\mathcal{C}_{q_m(i)},j\neq i}\left|Z_s^j\right|^2 |\mathcal{F}_m\right] ds$$
$$+ \int_{t_m}^t \mathbb{E}\left[Z_{t_m}^i \cdot \chi_i(\mathbf{X}(s);k) |\mathcal{F}_m\right] ds + C\Delta^3 + \frac{3\Delta}{2}\int_{t_m}^t \mathbb{E}\left[|\chi_i(\mathbf{X}(s);k)|^2 |\mathcal{F}_m\right] ds + \frac{5}{2}\int_{t_m}^t \mathbb{E}\left[\|\chi_i(\mathbf{X}(s);\sigma)\|^2 |\mathcal{F}_m\right] ds.$$

Due to Lemma 4.4, $\int_{t_m}^t \mathbb{E}(Z_{t_m}^i \cdot \chi_i(\mathbf{X}(s);k)|\mathcal{F}_m) ds = 0$. Regarding the term

$$\int_{t_m}^t \mathbb{E}\left[\frac{1}{P-1}\sum_{j\in\mathcal{C}_{q_m(i)},j\neq i}\left|Z_s^j\right|^2 |\mathcal{F}_m\right] ds$$

we note that the batch $\mathcal{C}^{(m)}$ is not independent of $|Z_t^j|$. To treat this, [20,21] considered the difference $|Z_t^j| - |Z_{t_m}^j|$. Here, we make use of the symmetry. By Lemma 4.7, taking expectation on both sides, one has

$$\mathbb{E}\left[\left|Z_t^i\right|^2\right] \leq \mathbb{E}\left[\left|Z_{t_m}^i\right|^2\right] + C\int_{t_m}^t \mathbb{E}\left[\left|Z_s^i\right|^2\right]ds + C\Delta^3 + \frac{3\Delta}{2P}\int_{t_m}^t \mathbb{E}[\Lambda_i\left(\mathbf{X}(s);k\right)]ds + \frac{5}{2P}\int_{t_m}^t \mathbb{E}[\Lambda_i\left(\mathbf{X}(s);\sigma\right)]ds.$$

Here, note that $\mathbf{X}(t)$ is independent of the random batches at t_m so that

$$\mathbb{E}\left[\mathbb{E}\left[\left|\chi_i\left(\mathbf{X}(s);k\right)\right|^2 |\mathcal{F}_m\right]\right] = \mathbb{E}\left[\left|\chi_i\left(\mathbf{X}(s);k\right)\right|^2\right] \leq \frac{1}{P}\mathbb{E}[\Lambda_i\left(\mathbf{X}(s);k\right)].$$

Finally, by iteration, one has

$$\mathbb{E}\left[\left|Z_t^i\right|^2\right] \leq C\int_0^t \mathbb{E}\left[\left|Z_s^i\right|^2\right]ds + C(m+1)\Delta^3 + \frac{3\Delta}{2P}\int_0^t \mathbb{E}[\Lambda_i\left(\mathbf{X}(s);k\right)]ds + \frac{5}{2P}\int_0^t \mathbb{E}[\Lambda_i\left(\mathbf{X}(s);\sigma\right)]ds.$$

Applying Grönwall's inequality (Thm. 1 in [12]) yields the result as $(m+1)\Delta^3 \leq T\Delta^2$. $\qquad\square$

4.3. The nonlinear case

In this subsection, we consider the analysis of the random batch approximation (4.8) for the nonlinear dependence case.

We will establish the approximation error estimate with the following assumption.

Assumption 4.9. *f is one-sided Lipschitz. $A(\cdot) : \mathbb{R}^r \to \mathbb{R}^d$, $k : \mathbb{R}^d \times \mathbb{R}^d \to \mathbb{R}^r$ and $\sigma : \mathbb{R}^d \to \mathbb{R}^{d\times m'}$ are all Lipschitz continuous and $\nabla^2 A$ is bounded.*

Following similar argument as Proposition 4.3, one can show that

Proposition 4.10. *Let Assumption 4.9 hold and $\mu_0 \in \mathcal{P}_q$ with $q \geq 2$. The process $\{X_t^i\}$ satisfies that*

$$\sup_{0\leq t\leq T} \mathbb{E}[|X_t^i|^q] = \sup_{0\leq t\leq T} \mathbb{E}[|X_t^1|^q] \leq C_{q,T}, \quad \forall i. \tag{4.18}$$

For the process $\tilde{X}_t^i$, conditioning on the sequence of random batches, one has

$$\sup_{0\leq t\leq T} \max_{1\leq j\leq N} \mathbb{E}[|\tilde{X}_t^j|^q|\mathcal{F}] \leq C_{q,T}, \tag{4.19}$$

where $C_{q,T}$ is independent of N and the sequence of batches.

The proof is essentially the same and we omit. Moreover, the symmetry Lemma 4.7 also holds for this case.

Below, let us focus on the error process $Z_t^i = \tilde{X}_t^i - X_t^i$. We find that

$$dZ_t^i = (f(\tilde{X}_t^i) - f(X_t^i))\,dt + \left(A\left(\frac{1}{P-1}\sum_{j\in\mathcal{C}_{q_m(i)},j\neq i} k(\tilde{X}_t^i,\tilde{X}_t^j)\right) - A\left(\frac{1}{P-1}\sum_{j\in\mathcal{C}_{q_m(i)},j\neq i} k(X_t^i,X_t^j)\right)\right)dt$$

$$+ \left(A\left(\frac{1}{P-1}\sum_{j\in\mathcal{C}_{q_m(i)},j\neq i} k(X_t^i,X_t^j)\right) - A\left(\frac{1}{N-1}\sum_{j\neq i} k(X_t^i,X_t^j)\right)\right)dt + (\sigma(\tilde{X}_t^i) - \sigma(X_t^i))\,dW_t^i.$$

Denote $F_i(\mathbf{X}(t)) := \frac{1}{N-1}\sum_{j\neq i} k(X_t^i,X_t^j)$. The main term that brings the difference is thus

$$A\Big(\chi_i(\mathbf{X}(t);k) + F_i(\mathbf{X}(t))\Big) - A(F_i(\mathbf{X}(t))) = \chi_i(\mathbf{X}(t);k)\cdot\nabla A(F_i(\mathbf{X}(t)))$$

$$+ \frac{1}{2}\chi_i(\mathbf{X}(t);k)\otimes\chi_i(\mathbf{X}(t);k) : \nabla^2 A(F_i(\mathbf{X}(t)) + \xi\chi_i).$$

where ξ is a constant in $(0, 1)$.

As one may expect, the linear term is fine and can be estimated similarly as in [20, 21]. The main difference is the quadratic term. In fact, this term will be of the order of the variance so that it contributes a term with order $1/P^2$ to the mean square error (thus $O(1/P)$ to the error).

We first consider the fourth moments of the random deviation:

Lemma 4.11. *There is a universal constant C such that for $N \geq 5$ that*

$$\mathbb{E}[|\chi_i(x;k)|^4] \leq C \frac{Q(x;k)}{P^2},$$

where

$$Q(x;k) = [M_1(x;k)^4 + M_2(x;k)M_1(x;k)^2 + M_2(x;k)^2 + M_3(x;k)M_1(x;k) + M_4(x;k)],$$
$$M_q(x;k) = \frac{1}{N} \sum_j |k(x_i, x_j)|^q, \quad q = 1, 2, 3, 4.$$

The proof of this lemma is tedious, and we provide its sketch in Appendix B.

Theorem 4.12. *Consider the nonlinear model* (4.6) *and recall $Z_t^i = \tilde{X}_t^i - X_t^i$. Let Assumption 4.9 hold. Then,*

$$\sup_{0 \leq t \leq T} \max_i \mathbb{E}[|Z_t^i|^2] \leq C \left(\Delta^2 + \frac{\Delta}{P} \Lambda(k) \right) + C(\|\nabla^2 A\|) \frac{Q(k)}{P^2}, \tag{4.20}$$

where $Q(k) = \sup_{0 \leq t \leq T} \mathbb{E}[Q(\mathbf{X}(t);k)] < +\infty$.

Proof. By Itô's formula, one has for $t \in [t_m, t_{m+1})$ that

$$\frac{d}{dt} \mathbb{E}\left[|Z_t^i|^2|\mathcal{F}_m\right]$$

$$\leq \left\{ 2\mathbb{E}\left[Z_t^i \cdot \left(f(\tilde{X}_t^i) - f(X_t^i)\right) \Big| \mathcal{F}_m \right] + 2\|\nabla A\|_\infty \mathbb{E}\left[\frac{|Z_t^i|}{P-1} \sum_{j \in \mathcal{C}_{q_m(i)}, j \neq i} \left|k\left(\tilde{X}_t^i, \tilde{X}_t^j\right) - k\left(X_t^i, X_t^j\right)\right| \Big| \mathcal{F}_m \right] \right\}$$

$$+ \left\{ 2\mathbb{E}\left[\chi_i\left(\mathbf{X}(t);k\right) \cdot \nabla A\left(F_i(\mathbf{X}(t))\right) \cdot Z_t^i|\mathcal{F}_m\right] + 2\mathbb{E}\left(\chi_i \otimes \chi_i : \nabla^2 A\left(F_i(\mathbf{X}(t)) + \xi\chi_i\right) \cdot Z_t^i|\mathcal{F}_m\right) \right\}$$

$$+ \left\{ \mathbb{E}\left[\mathrm{tr}\left(\left(\sigma(\tilde{X}_t^i) - \sigma(X_t^i)\right)\left(\sigma(\tilde{X}_t^i) - \sigma(X_t^i)\right)^T\right) \Big| \mathcal{F}_m \right] \right\}$$

$$=: S_1 + S_2 + S_3,$$

By Assumption 4.9, $S_1 + S_3$ can be simply controlled by

$$\mathbb{E}[|Z_t^i|^2|\mathcal{F}_m] + \mathbb{E}\left[\frac{1}{P-1} \sum_{j \in \mathcal{C}_{q_m(i)}, j \neq i} |Z_t^j|^2|\mathcal{F}_m\right]$$

Consider the S_2 term. The second term in S_2 will be bounded simply by

$$\mathbb{E}[|Z_t^i|^2|\mathcal{F}_m] + \mathbb{E}[|\chi_i(\mathbf{X}(t))|^4|\mathcal{F}_m].$$

For the first term in S_2, noticing that

$$\mathbb{E}\left[\chi_i(\mathbf{X}(t);k) \cdot \nabla A(F_i(\mathbf{X}(t))) \cdot Z_{t_m}^i\right] = 0,$$

one may again break $Z_t^i = Z_{t_m}^i + (Z_t^i - Z_{t_m}^i)$. The treatment for $Z_t^i - Z_{t_m}^i$ is tedious but is similar to the proof of Theorem 4.5. We omit the details. Eventually, one has

$$
\mathbb{E}\left[|Z_t^i|^2|\mathcal{F}_m\right] \leq \mathbb{E}\left[|Z_{t_m}^i|^2|\mathcal{F}_m\right] + C\int_{t_m}^t \mathbb{E}\left[|Z_s^i|^2|\mathcal{F}_m\right] ds + C\int_{t_m}^t \mathbb{E}\left[\frac{1}{P-1}\sum_{j\in\mathcal{C}_{q_m(i)},j\neq i}|Z_s^j|^2|\mathcal{F}_m\right] ds
$$

$$
+ \int_{t_m}^t \chi_i\left(\mathbf{X}(s);k\right)\cdot\nabla A\left(F_i\left(\mathbf{X}(s)\right)\right)\cdot Z_{t_m}^i ds + C\Delta^3
$$

$$
+ C\frac{\Delta}{2}\int_{t_m}^t \mathbb{E}\left[|\chi_i\left(\mathbf{X}(s);k\right)|^2|\mathcal{F}_m\right] ds + C(1+\Delta)\int_{t_m}^t \mathbb{E}\left[\left|\chi_i\left(\mathbf{X}(s);\sigma\right)\right|^4\Big|\mathcal{F}_m\right] ds.
$$

Taking expectation, using Lemma 4.4, the symmetry (Lem. 4.7 also holds for this case and the proof is the same) and Lemma 4.11, one has

$$
\mathbb{E}[|Z_t^i|^2] \leq \mathbb{E}[|Z_{t_m}^i|^2] + C\int_{t_m}^t \mathbb{E}[|Z_s^i|^2] ds + C\Delta^3 + C\int_{t_m}^t \frac{\Lambda(k)\Delta}{P} ds + C\int_{t_m}^t \left(\|\nabla^2 A\|\right)\frac{Q(k)}{P^2} ds.
$$

Then, similar as in the proof of Theorem 4.5, one can do iteration first and apply the Grönwall's inequality to obtain the final estimate. $\square$

Clearly, this error introduced by the nonlinearity is smaller compared to the one introduced by the random batch approximation in the noise term.

4.4. Discussion on more general cases

In this section, we discuss more general cases. In fact, the distribution dependence in the previous section does not include the cases when several statistical quantities are involved in the coefficients. For example, if the coefficient depends on the variance, then both the first and second moments will be involved. These cases could be given by

$$
a(x,\mu) = f(x) + A\left(\int k_1(x,y)\mu(dy),\cdots,\int k_p(x,y)\mu(dy)\right)
$$

and

$$
b(x,\mu) = B\left(\int \sigma_1(x,y)\mu(dy),\cdots,\int \sigma_q(x,y)\mu(dy)\right).
$$

Here, A, B, k_i, σ_j are all assumed to be Lipschitz and have bounded derivatives. f is allowed to be one-sided Lipschitz.

The random approximation for such systems can be similarly performed. There is no big difference for the error estimate, while the details could be tedious. One can find that the strong error is again like

$$
\mathbb{E}[|Z_t^i|^2] \leq \frac{C(\sigma)}{P} + \frac{C(k)}{P^2} + \frac{C\Delta}{P}.
$$

In other words, if there is law dependence in the coefficient of the noise, the mean square error would be like $\mathcal{O}(1/P)$. The mean error arising from the nonlinearity in the drift is like $\mathcal{O}(1/P^2)$. These do not vanish as the step size tends to zero. However, the estimates are independent of N so they could also save time if the required N is large. In the special case when the diffusion coefficient b is independent of μ and the drift linearly depends on μ, the RBM approximation has an error that vanishes in the $\Delta \to 0$ limit, which is exactly the one studied in [20].

5. Numerical example

In this section, we perform some numerical experiments to verify the theoretical analysis above. In many applications of McKean–Vlasov SDEs, the dependence on the distribution in the coefficients are simply through the expected value of some kernel with respect to the law, which is a simple example of the cooperative interaction. This might be due to the fact that many systems have a tendency to relax toward the center of gravity of the ensemble. See for example the model in [8]. Due to this reason, we will mainly consider examples of this form in this section.

Example 5.1. For linear cases, we consider the following one-dimensional SDE

$$dX(t) = \left(\lambda_1 X(t)\left(\lambda_2 - |X(t)|\right) + \int_{\mathbb{R}} (X(t) - y)\,\mu(dy)\right) dt + \left(\lambda_3 |X(t)|^{3/2} + \int_{\mathbb{R}} (X(t) - y)\,\mu(dy)\right) dW(t) \quad (5.1)$$

where λ_1, λ_2 and λ_3 are positive constants.

The corresponding system of interacting particles is given by

$$dX_t^{i,N} = \left(\lambda_1 X_t^{i,N}\left(\lambda_2 - |X_t^{i,N}|\right) + \frac{1}{N-1}\sum_{j\neq i}^{N}(X_t^{i,N} - X_t^{j,N})\right) dt + \left(\lambda_3|X_t^{i,N}|^{3/2} + \frac{1}{N-1}\sum_{j\neq i}^{N}(X_t^{i,N} - X_t^{j,N})\right) dW_t^i.$$
$$(5.2)$$

In the subsequent analysis, we fix the following parameters: $X_0^i = 1, \lambda_1 = 2.5, \lambda_2 = 1, \lambda_3 = 1$, for all $i \in \{1,\ldots,N\}$. We select a particle number of $N = 2^{12}$ and generate a total of 1000 sample paths (*i.e.*, conducting the simulation 1000 times) to evaluate the mean square error at $T = 1$. This evaluation is performed using four different step sizes: $\Delta = 2^p$ for $-10 \leq p \leq -7$. The mean square error, denoted by $\sqrt{\mathbb{E}[|X_t^{i,N} - X_t^{i,\Delta}|^2]}$, is calculated by approximating the expectation *via* Monte Carlo simulations, utilizing all the particles for all the sample paths. Additionally, we calculate 95% confidence intervals for each figure. We utilize the bootstrap method to estimate these confidence intervals for a given dataset. For each error plot, the process involve iterating over datasets generated by the errors of the numerical scheme with different step sizes. Each dataset is resampled 100 times to create new samples. Subsequently, the mean of each resampled dataset is computed, and the 95% confidence interval is determined based on these resampled means. These confidence intervals will be visually represented as error bars, positioned at the corresponding mean error values.

Figure 1 displays the resulting log-log error plot along with a reference line for slope one-half. The slopes of the two curves seem to line up nicely, which confirms the theoretical result in Theorem 3.4 when the truncated EM method applied to the interacting particle system corresponding to (5.1) directly. Here we choose the tamed Euler scheme with time step size $\Delta = 2^{-12}$ and the same number of particles $N = 2^{12}$ for producing a reference solution.

To validate Theorem 4.5, we also employ truncated type methods together with the random batch technique to do a simulation. From Corollary 4.6, we note that the order of convergence of our approach to approximating the interacting particle system related to the McKean–Vlasov SDE (5.1) is at least $\beta/2$.

Now we generate the reference solution using the truncated Euler method (3.4) with a step size of $\Delta = 2^{-12}$. Subsequently, with three different values of β chosen, we obtain three distinct batch sizes denoted by $P = 1/\Delta^\beta$. The numerical errors associated with these batch sizes are plotted in Figure 2. Hence, our findings align with the expected strong order of convergence stated in Corollary 4.6.

Considering that the main advantage of RBM is to save the computational cost, we compare the computation time of the truncated Euler method using a random batch technique and the counterpart without using the random batch technique. In Table 1, we denote the former by the notation 'TEMwRBM', and the notation 'TEM' represents the latter. Let us consider SDE (5.1) where there is a dependence in the diffusion coefficient. It is easy to find that the execution time significantly enlarges when increasing the number of particles. Here we
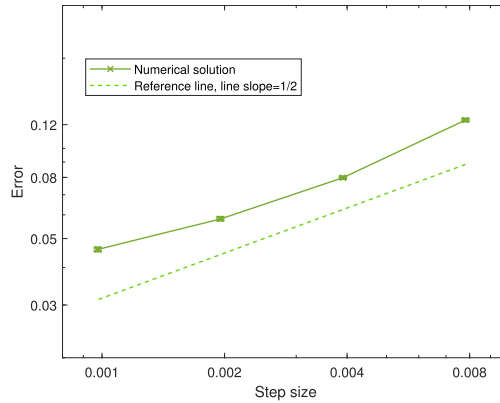
FIGURE 1. The convergence order of the truncated EM method applied to the interacting particle system related to (5.1) without using the random batch technique. The error bars indicate the 95% confidence interval.
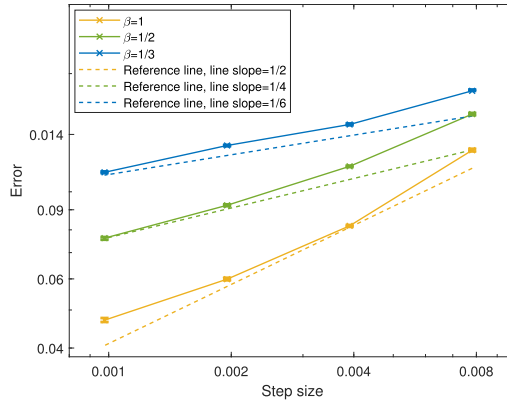

FIGURE 2. Convergence order of truncated EM method applied to the interacting particle system related to (5.1) with random batch technique for $\beta = 1, 1/2, 1/3$. The error bars indicate the 95% confidence interval.

fix the time step size $\Delta = 2^{-7}$. The 'TEM' and the 'TEMwRBM' run 1000 sample paths to obtain the average time consumption respectively.

Upon implementing the random batch technique, we can approximate the error, as stated in (4.16), using $\sqrt{1/P} \sim \epsilon$ for a given tolerance ϵ. In this scenario, $P = O(1/\epsilon^2)$ remains fixed, resulting in a cost per iteration of $O(NP)$. However, when $P = 1/\Delta^\beta$ is also required, the cost per iteration increases as the value of β grows. On the other hand, the computational cost presented in Table 1 corresponds to a fixed terminal time. Notably, adjusting β is equivalent to modifying Δ such that $\Delta = O(\epsilon^{2/\beta})$. Increasing the value of β results in a larger step size, consequently reducing the number of iterations. Thus, an optimal value of β exists in terms of the overall computational cost. However, we have yet to fully comprehend this phenomenon, and further investigation would be both intriguing and worthwhile.

Example 5.2. When there is no interaction in the diffusion part, (5.1) degenerates to

$$dX(t) = \left(\lambda_1 X(t) \left(\lambda_2 - |X(t)| \right) + \int_{\mathbb{R}} \left(X(t) - y \right) \mu(dy) \right) dt + \lambda_3 |X(t)|^{3/2} \, dW(t). \tag{5.3}$$

TABLE 1. Average execution time.

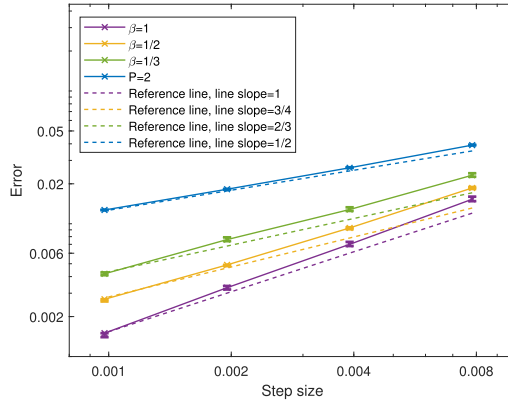| N	TEM	TEMwRBM		
		$P = 1/\Delta$	$P = 1/\sqrt{\Delta}$	$P = 1/\sqrt[3]{\Delta}$
2^{12}	4.8672s	1.3771s	0.9657s	0.9559s
2^{14}	84.1268s	5.6311s	3.9629s	4.1012s
2^{16}	1143.8016s	21.7272s	14.9615s	15.4843s

FIGURE 3. The convergence order of truncated Milstein method with a random batch size $P = 1/\Delta^\beta$ with $\beta = 1, 1/2, 1/3$ and $P = 2$ applied to the interactive particle system (5.4). The error bars indicate the 95% confidence interval.

The corresponding system of interacting particles is given by

$$dX_t^{i,N} = \left(\lambda_1 X_t^{i,N} \left(\lambda_2 - |X_t^{i,N}| \right) + \frac{1}{N-1} \sum_{j \neq i}^{N} (X_t^{i,N} - X_t^{j,N}) \right) dt + \lambda_3 |X_t^{i,N}|^{3/2} dW_t^i. \tag{5.4}$$

In what follows, we set parameters $X_0^i = 1, \lambda_1 = 2.5, \lambda_2 = 1, \lambda_3 = 1$, for all $i \in \{1, \ldots, N\}$. We choose particle number $N = 2^{12}$ and obtain 1000 sample paths (*i.e.*, repeat the simulation for 1000 times) to calculate the mean square error at $T = 1$ with four different step sizes: $\Delta = 2^p$ for $-10 \leq p \leq -7$.

The conclusion (4.16) implies that we can obtain the first-order convergence when approximating the interacting particle system associated to (5.3) by a random batch technique with $\beta = 1$ (the batch size $P = 1/\Delta$). At the same time strong convergence order of a numerical method should be one. Therefore we employ the truncated Milstein scheme [14] to verify the first result of the Corollary 4.6. The numerical results illustrated in Figure 3 confirm our theoretical result, where a dashed red line with a slope of one is provided for reference. Moreover, we conducted a comparison of the results with varying batch sizes. When considering $P = 1/\Delta^\beta$ with $\beta = 1, 1/2, 1/3$, as well as $P = 2$, as shown in Figure 3 and agrees with the theory in Corollary 4.6.

Example 5.3. For nonlinear cases, we consider the following SDE

$$dX(t) = \left(\lambda_1 X(t) \left(\lambda_2 - |X(t)| \right) + \sin \left(\int_{\mathbb{R}} (X(t) - y) \mu(dy) \right) \right) dt + \lambda_3 |X(t)|^{3/2} dW(t)$$

where λ_1, λ_2 and λ_3 are positive constants.
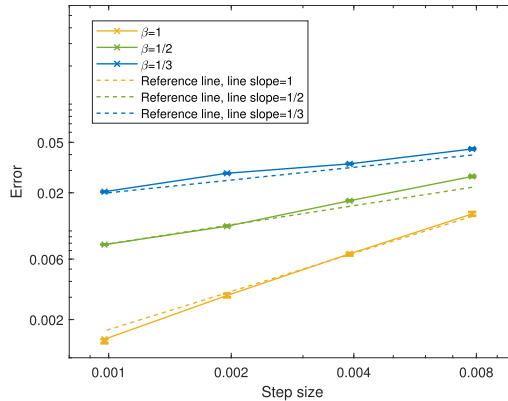
FIGURE 4. The convergence order of the truncated Milstein method and random batch technique with $\beta = 1, 1/2, 1/3$ in nonlinear cases. The error bars indicate the 95% confidence interval.

The corresponding system of interacting particles is given by

$$dX_t^{i,N} = \left(\lambda_1 X_t^{i,N} \left(\lambda_2 - |X_t^{i,N}| \right) + \sin \left(\frac{1}{N-1} \sum_{j \neq i}^{N} (X_t^{i,N} - X_t^{j,N}) \right) \right) dt + \lambda_3 |X_t^{i,N}|^{3/2} dW_t^i.$$

We set parameters $X_0^i = 1, \lambda_1 = 2.5, \lambda_2 = 1, \lambda_3 = 1$, for all $i \in \{1, \ldots, N\}$, and choose particle number $N = 2^{12}$ and obtain 1000 sample paths to calculate the mean square error at $T = 1$ with four different step sizes: $\Delta = 2^p$ for $-10 \leq p \leq -7$. Then we examine three distinct batch sizes represented by $P = 1/\Delta^\beta$. The numerical errors are illustrated in Figure 4, where three dashed red lines with slopes of 1/2, 1/4, and 1/6 are provided as reference. These results align with the expected strong order of convergence stated in Theorem 4.12.

Example 5.4. In this example, we consider a three-dimensional McKean–Vlasov SDE model, which is employed to describe neuron activity [1, 10], given as follows.

$$dX(t) = \int_0^t b(X(s), \mu) ds + \int_0^t \sigma(X(s), \mu) dW(s) \tag{5.5}$$

where

$$b(x, \mu) := \begin{pmatrix} x_1 - (x_1)^3/3 - x_2 + I - \int_{\mathbb{R}^3} J (x_1 - V_{rev}) z_3 \, d\mu(z) \\ c (x_1 + a - b x_2) \\ a_r \frac{T_{\max}(1-x_3)}{1+\exp(-\lambda(x_1 - V_T))} - a_d x_3 \end{pmatrix},$$

$$\sigma(x, \mu) := \begin{pmatrix} \sigma_{\text{ext}} & 0 - \int_{\mathbb{R}^3} \sigma_J (x_1 - V_{\text{rev}}) z_3 \, d\mu(z) \\ 0 & 0 & 0 \\ 0 & 1 & 0 \end{pmatrix}$$

for $x = (x_1, x_2, x_3)$, $z = (z_1, z_2, z_3)$. This is a stochastic model describing the firing dynamics of a network of spiking neurons. The integral terms in the functions b and σ capture the influence of the network's collective behavior, commonly referred to as the mean-field limit. This model is important for understanding how neurons interact and communicate with each other in the brain. The stochastic nature of the model accounts for the inherent randomness observed in the behavior of individual neurons, as discussed in [1].
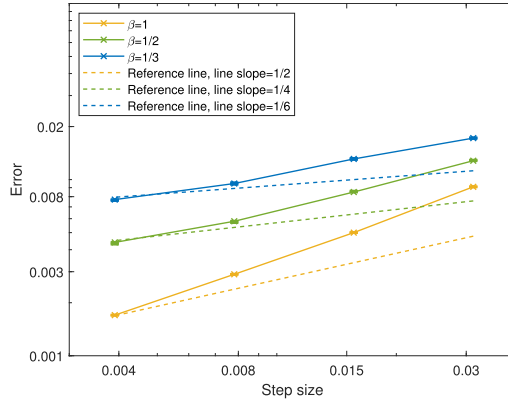
FIGURE 5. Convergence order of truncated EM method applied to the interacting particle system associated to (5.5) with random batch technique for $\beta = 1, 1/2, 1/3$. The error bars indicate the 95% confidence interval.

In our simulation, the initial value is chosen as

$$X_0 \sim \mathscr{N} \left(\begin{pmatrix} V_0 \\ w_0 \\ y_0 \end{pmatrix}, \begin{pmatrix} \sigma_{V_0} & 0 & 0 \\ 0 & \sigma_{w_0} & 0 \\ 0 & 0 & \sigma_{y_0} \end{pmatrix} \right)$$

and the parameters are set as follows

$$\begin{array}{llllll}
V_0 = 0 & \sigma_{V_0} = 0.4 & a = 0.7 & b = 0.8 & c = 0.08 & I = 0.5 & \sigma_{\text{ext}} = 0.5 \\
w_0 = 0.5 & \sigma_{w_0} = 0.4 & V_{\text{rev}} = 1 & a_r = 1 & a_d = 1 & T_{\max} = 1 & \lambda = 0.2 \\
y_0 = 0.3 & \sigma_{y_0} = 0.05 & J = 1 & \sigma_J = 0.2 & V_T = 2.
\end{array}$$

We choose particle number $N = 2^8$ and obtain 1000 sample paths to calculate the mean square error at $T = 1$ with four different step sizes: $\Delta = 2^p$ for $-8 \leq p \leq -5$. Here we choose the tamed Euler scheme with time step size $\Delta = 2^{-13}$ and the same number of particles $N = 2^8$ for producing a reference solution.

In Figure 5, we examine three distinct batch sizes represented by $P = 1/\Delta^\beta$, where three dashed red lines with slopes of 1/2, 1/4, and 1/6 are provided as reference. The numerical results roughly agree with the theory, and have slightly larger convergence rate compared to theory in the range observed. The error bars in the aforementioned figures indicate that the variance of the error data calculated by our method is small.

APPENDIX A. PROOF OF PROPOSITION 2.4

Proof. Here, we only illustrate the moments control for X_t^i. The one for $\bar{X}_t^i$ is similar and simpler.

For (2.5), using Itô's formula,

$$\mathbb{E}[|X_t^i|^q] \leq \mathbb{E}[|X_0^i|^q] + q\mathbb{E} \int_0^t |X_s^i|^{q-2} \langle X_s^i, a\left(X_s^i, \mu_s^X\right) \rangle ds + \frac{q(q-1)}{2} \mathbb{E} \int_0^t |X_s^i|^{q-2} \| b\left(X_s^i, \mu_s^X\right) \|^2 ds.$$

By (2.3), we get

$$\mathbb{E}[|X_t^i|^q] \leq \mathbb{E}[|X_0^i|^q] + qL\mathbb{E} \int_0^t |X_s^i|^{q-2}[(1 + |X_s^i|)^2 + \mathcal{W}_2^2\left(\mu_s^X, \delta_0\right)] ds.$$

Apply the Young's inequality and Lemma 2.1,

$$\mathbb{E}[|X_t^i|^q] \le C(1 + \mathbb{E}[|X_0^i|^q]) + C \int_0^t \mathbb{E}[|X_s^i|^q]ds + C \int_0^t \mathbb{E}\left(\frac{1}{N}\sum_{j=1}^N |X_s^j|^2\right)^{q/2} ds.$$

For the last term, we apply the Minkowski inequality and the symmetry consecutively,

$$\mathbb{E}\left(\frac{1}{N}\sum_{j=1}^N |X_s^j|^2\right)^{q/2} \le \left(\frac{1}{N}\sum_{j=1}^N \| |X_s^j|^2 \|_{L^{q/2}}\right)^{q/2} = \mathbb{E}[|X_s^i|^q].$$

Hence,

$$\mathbb{E}[|X_t^i|^q] \le C \int_0^t \mathbb{E}[|X_s^i|^q]ds + C(1 + \mathbb{E}[|X_0^i|^q]),$$

and applying the Grönwall's inequality finishes the estimate. $\qquad\square$

Appendix B. Proof of the auxilliary Lemma 4.11

Recall that $q(i)$ means the index of the batch that particle i falls into. Introduce the indicator function

$$I_{ij} = \begin{cases} 1 & q(i) = q(j), \\ 0 & \text{otherwise.} \end{cases}$$

Then, we have the simple lemma:

Lemma B.1. *Let $q < N$. Let $i, j_{j'}$, $1 \le j' \le q$ be all distinct. Then,*

$$\mathbb{E}\prod_{j'=1}^q I_{ij_{j'}} = \prod_{j'=1}^q \frac{P - j'}{N - j'}.$$

Proof. If $P \le q$, clearly both sides are zero so the inequality holds. Now, assume $P > q$.

The event for which $\prod_{j'=1}^q I_{ij_{j'}} = 1$ is when all the particles are in the same batch. The proof would be similar to the argument in Lemma 3.1 of [20]. There are

$$M(n) := \frac{(nP)!}{(P!)^n n!}$$

ways to divide $N = nP$ particles into n batches, without consideration of the order of batches. Then, the probability that $i, j_1, \cdots, j_q$ are in the same batch is given by

$$\frac{\binom{nP-q-1}{P-q-1}M(n-1)}{M(n)} = \prod_{j'=1}^q \frac{P - j'}{N - j'}.$$

$\qquad\square$

Proof of Lemma 4.11. Clearly,

$$\chi_i(\underline{x}; k) = \sum_{j=1, j \ne i}^N \left(\frac{1}{P-1}I_{ij} - \frac{1}{N-1}\right) k(x_i, x_j).$$

It follows that

$$|\chi_i(\mathbf{x};k)|^4 = \sum_{j_1 j_2 j_3 j_4, \neq i} \zeta(j_1, j_2, j_3, j_4) k(x_i, x_{j_1}) \cdot k(x_i, x_{j_2}) k(x_i, x_{j_3}) \cdot k(x_i, x_{j_4}).$$

Here, for matrices, $k(x_i, x_{j_1}) \cdot k(x_i, x_{j_2})$ means $\operatorname{tr}(k(x_i, x_{j_1}) k(x_i, x_{j_2})^T)$. The random variable ζ is given by

$$\zeta(j_1, j_2, j_3, j_4) = \frac{1}{(P-1)^4} \prod_{j'} I_{i,j_{j'}} - \frac{1}{(P-1)^3(N-1)}(I_{ij_1} I_{ij_2} I_{ij_3} + I_{ij_1} I_{ij_2} I_{ij_4} + I_{ij_1} I_{ij_3} I_{ij_4} + I_{ij_2} I_{ij_3} I_{ij_4})$$

$$+ \frac{1}{(P-1)^2(N-1)^2}(I_{ij_1} I_{ij_2} + I_{ij_1} I_{ij_3} + I_{ij_1} I_{ij_4} + I_{ij_2} I_{ij_3} + I_{ij_2} I_{ij_4} + I_{ij_3} I_{ij_4})$$

$$- \frac{1}{(P-1)(N-1)^3}(I_{ij_1} + I_{ij_2} + I_{ij_3} + I_{ij_4}) + \frac{1}{(N-1)^4}.$$

Here, we consider several cases.

Case 1. j_1, j_2, j_3, j_4 are all distinct. There are $(N-1)(N-2)(N-3)(N-4)$ such terms. By Lemma B.1, one has

$$\mathbb{E}\zeta(j_1, j_2, j_3, j_4) = \frac{(P-2)(P-3)(P-4)}{(P-1)^3 \prod_{j'=1}^{4}(N-j')} - \frac{4(P-2)(P-3)}{(P-1)^2(N-1)^2(N-2)(N-3)}$$

$$+ \frac{6(P-2)}{(P-1)(N-1)^3(N-2)} - \frac{3}{(N-1)^4}.$$

Adding these up, one has $Num/[(P-1)^3(N-1)^4(N-2)(N-3)(N-4)]$. We need to check the coefficients of $P^3 N^3, P^2 N^3$ and $P^3 N^2$ in the numerator Num. After some calculation, one finds that these are all zero. Hence,

$$\mathbb{E}\zeta(j_1, j_2, j_3, j_4) \leq C \frac{PN^3 + P^2 N^2 + P^3 N}{P^3 N^7} \leq C \frac{1}{P^2 N^4}.$$

It is then easy to see that these terms are controlled by $C \frac{1}{P^2} M_1(x;k)^4$.

Case 2. There are two indices that are the same. There are $6(N-1)(N-2)(N-3)$ such terms.

$$\mathbb{E}\zeta(j_1, j_2, j_3, j_4) = \frac{1}{(P-1)^4} \frac{(P-1)(P-2)(P-3)}{(N-1)(N-2)(N-3)}$$

$$- \frac{1}{(P-1)^3(N-1)} \left(2\frac{(P-1)(P-2)}{(N-1)(N-2)} + 2\frac{(P-1)(P-2)(P-3)}{(N-1)(N-2)(N-3)}\right)$$

$$+ \frac{1}{(P-1)^2(N-1)^2} \left(\frac{P-1}{N-1} + 5\frac{(P-1)(P-2)}{(N-1)(N-2)}\right) - \frac{3}{(N-1)^4}.$$

Similarly, by combining terms and checking the corresponding coefficients for some big terms, one can find that this is controlled by

$$C \frac{PN^2 + P^2 N}{P^3 N^5} + C \frac{P+1}{P^2 N^4} \leq C \frac{1}{P^2 N^3}.$$

Hence, this case is controlled by $C \frac{1}{P^2} M_1^2 M_2$.

Case 3. Two indices are the same while the other two are the same. There are $3(N-1)(N-2)$ such terms.

$$\mathbb{E}\zeta(j_1, j_2, j_3, j_4) = \frac{1}{(P-1)^4} \frac{(P-1)(P-2)}{(N-1)(N-2)} - \frac{4}{(P-1)^3(N-1)} \frac{(P-1)(P-2)}{(N-1)(N-2)}$$

$$+ O\left(\frac{1}{(P-1)^2(N-1)^2} + \frac{1}{(P-1)(N-1)^3}\right).$$

This is clearly controlled by C/P^2, and thus the sum of such terms is controlled by M_2^2/P^2.

Case 4. Three indices are the same. There are $4(N-1)(N-2)$ such terms.
This situation can be treated similarly as Case 3. This is controlled by $M_3 M_1/P^2$.

Case 5. All indices are the same. There are $N-1$ such terms. This is controlled by M_4/P^3.
These then add up to the estimate in the lemma.

□

References

[1] J. Baladron, D. Fasoli, O. Faugeras and J. Touboul, Mean-field description and propagation of chaos in networks of Hodgkin-Huxley and FitzHugh-Nagumo neurons. *J. Math. Neurosci.* **2** (2012) 10.

[2] J. Bao and X. Huang, Approximations of McKean–Vlasov stochastic differential equations with irregular coefficients. *J. Theor. Probab.* **35** (2022) 1187–1215.

[3] J. Bao, C. Reisinger, P. Ren and W. Stockinger, First-order convergence of Milstein schemes for McKean–Vlasov equations and interacting particle systems. *Proc. Roy. Soc. London Ser. A* **477** (2021) 20200258.

[4] M. Bauer, T. Meyer-Brandis and F. Proske, Strong solutions of mean-field stochastic differential equations with irregular drift. *Electron. J. Probab.* **23** (2018) 1–35.

[5] D. Belomestny and J. Schoenmakers, Projected particle methods for solving McKean–Vlasov stochastic differential equations. *SIAM J. Numer. Anal.* **56** (2018) 3169–3195.

[6] V.S. Borkar and K. Suresh Kumar, McKean–Vlasov limit in portfolio optimization. *Stoch. Anal. Appl.* **28** (2010) 884–906.

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

[8] D.A. Dawson, Critical dynamics and fluctuations for a mean-field model of cooperative behavior. *J. Stat. Phys.* **31** (1983) 29–85.

[9] S. Dereich, M. Scheutzow and R. Schottstedt, Constructive quantization: Approximation by empirical measures. *Ann. Inst. Henri Poincare Probab. Stat.* **49** (2013) 1183–1203.

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

[11] G. Dos Reis, W. Salkeld and J. Tugaut. Freidlin-Wentzell LDP in path space for McKean–Vlasov equations and the functional iterated logarithm law. *Ann. Appl. Probab.* **29** (2019) 1487–1540.

[12] S.S. Dragomir, Some gronwall type inequalities and applications. *Sci. Dir. Work. Pap.* (2003) 04.

[13] N. Fournier and A. Guillin, On the rate of convergence in Wasserstein distance of the empirical measure. *Probab. Theory Relat. Fields* **162** (2015) 707–738.

[14] Q. Guo, W. Liu, X. Mao and R. Yue, The truncated Milstein method for stochastic differential equations with commutative noise. *J. Comput. Appl. Math.* **338** (2018) 298–310.

[15] A.L. Haji-Ali and R. Tempone, Multilevel and multi-index monte carlo methods for the mckean–vlasov equation. *Stat. Comput.* **28** (2018) 923–935.

[16] A.L. Haji-Ali, Hå. Hoel and R. Tempone, A simple approach to proving the existence, uniqueness, and strong and weak convergence rates for a broad class of McKean–Vlasov equations. Preprint: arXiv:2101.00886 (2021).

[17] J. Hoeksema, T. Holding, M. Maurelli and O. Tse, Large deviations for singularly interacting diffusions. Preprint: arXiv:2002.01295 (2020).

[18] S. Jin and L. Li, On the mean field limit of the random batch method for interacting particle systems. *Sci. China Math.* **65** (2022) 169–202.

[19] S. Jin and L. Li, Random batch methods for classical and quantum interacting particle systems and statistical samplings. In Vol. 3 *Active Particles*. Springer (2022) 153–200.

[20] S. Jin, L. Li and J. Liu, Random batch methods (RBM) for interacting particle systems. *J. Comput. Phys.* **400** (2020) 108877.

[21] S. Jin, L. Li and J. Liu, Convergence of the random batch method for interacting particles with disparate species and weights. *SIAM J. Numer. Anal.* **59** (2021) 746–768.

[22] S. Jin, L. Li, Z. Xu and Y. Zhao, A random batch Ewald method for particle systems with Coulomb interactions. *SIAM J. Sci. Comput.* **43** (2021) B937–B960.

[23] V.N. Kolokoltsov. *Nonlinear Markov Processes and Kinetic Equations*, Vol. 182. Cambridge University Press (2010).

[24] C. Kumar, N.C. Reisinger and W. Stockinger, Well-posedness and tamed schemes for McKean–Vlasov equations with common noise. *Ann. Appl. Probab.* **32** (2022) 3283–3330.

[25] Y. Li, X. Mao, Q. Song, F. Wu and G. Yin, Strong convergence of Euler-Maruyama schemes for McKean–Vlasov stochastic differential equations under local Lipschitz conditions of state variables. *IMA J. Numer. Anal.* (2022).

[26] X. Mao, *Stochastic Differential Equations and Applications*. Elsevier (2007).

[27] X. Mao, The truncated Euler-Maruyama method for stochastic differential equations. *J. Comput. Appl. Math.* **290** (2015) 370–384.

[28] X. Mao, Convergence rates of the truncated Euler-Maruyama method for stochastic differential equations. *J. Comput. Appl. Math.* **296** (2016) 362–375.

[29] H.P. McKean Jr., A class of Markov processes associated with nonlinear parabolic equations. *Proc. Nat. Acad. Sci. USA* **56** (1966) 1907.

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

[31] H. Pham, Linear quadratic optimal control of conditional McKean–Vlasov equation with random coefficients and applications. *Probab. Uncertain. Quant. Risk* **1** (2016)1–26.

[32] N.B. Rached, A.L. Haji-Ali, S.M.S. Pillai and R. Tempone, Single level importance sampling for McKean–Vlasov stochastic differential equation. Preprint: arXiv:2207.06926 (2022).

[33] H. Robbins and S. Monro, A stochastic approximation method. *Ann. Math. Stat.* (1951) 400–407.

[34] A.S. Sznitman, Topics in propagation of chaos. In *Ecole d'été de probabilités de Saint-Flour XIX–1989*. Springer (1991) 165–251.

[35] Z. Wang, X. Zhao and R. Zhu, Gaussian fluctuations for interacting particle systems with singular kernels. Preprint: arXiv:2105.13201 (2021).

[36] J. Zhang, *Topics in McKean–Vlasov equations: Rank-based dynamics and Markovian projection with applications in finance and stochastic control*, Ph.D. thesis, Princeton University (2021).