CONVERGENCE OF THE RANDOM BATCH METHOD FOR INTERACTING PARTICLES WITH DISPARATE SPECIES AND WEIGHTS*

SHI JIN†, LEI LI‡, AND JIAN-GUO LIU§

Abstract. We consider in this work the convergence of the random batch method proposed in our previous work [Jin et al., *J. Comput. Phys.*, 400(2020), 108877] for interacting particles to the case of disparate species and weights. We show that the strong error is of $O(\sqrt{\tau})$ while the weak error is of $O(\tau)$ where τ is the time step between two random divisions of batches. Both types of convergence are uniform in N, the number of particles. The proof of strong convergence follows closely the proof in [Jin et al., *J. Comput. Phys.*, 400(2020), 108877] for indistinguishable particles, but there are still some differences: Since there is no exchangeability now, we have to use a certain weighted average of the errors; some refined auxiliary lemmas have to be proved compared with our previous work. To show that the weak convergence of empirical measure is uniform in N, certain sharp estimates for the derivatives of the backward equations have been used. The weak convergence analysis is also illustrating for the convergence of the Random Batch Method for N-body Liouville equations.

Key words. interacting particle systems, random batch, empirical measure, multispecies, semigroup, N-body Liouville equation

AMS subject classifications. 65C20, 37M05

DOI. 10.1137/20M1327641

1. Introduction. Interacting particle systems are ubiquitous in nature, for example, molecules in fluids [19], plasma [7], galaxy in universe. In addition, many collective behaviors in natural and social sciences are due to interacting individuals, and examples include swarming [45, 11, 10, 15], flocking [14, 23, 1], and chemotaxis [24, 5] and consensus clusters in opinion dynamics [39]. In many models for these phenomenon, individual particles can have weights. For example, in the point vortex model [12, 22, 33], the "particles" correspond to different point vortices with different strength and they interact each other through a Hamiltonian system. To approximate the nonlinear Fokker–Planck equation for interacting particles, like the Keller–Segel equation, one can use interacting particles with different masses to approximate the dynamics and then compute the empirical density more efficiently [13, 34]. There may also be several species, where the particles may have different features; for example, in the microscopic description of the Poisson–Boltzmann equation, particles with different charges interact with each other through Coulomb forces [2, 31].

†School of Mathematical Sciences, Institute of Natural Sciences, MOE-LSC, Shanghai Jiao Tong University, Shanghai, 200240, P. R. China (shijin-m@sjtu.edu.cn).

‡Corresponding author. School of Mathematical Sciences, Institute of Natural Sciences, MOE-LSC, Shanghai Jiao Tong University, Shanghai, 200240, P. R. China. (leili2010@sjtu.edu.cn).

§Department of Mathematics and Department of Physics, Duke University, Durham, NC 27708 USA (jliu@phy.duke.edu).

The systems mentioned above can be written as the second order ordinary differential equation (ODE) and/or stochastic differential equation (SDE) system for $1 \leq i \leq N$

$$dX^i = V^i \, dt,$$

(1.1)
$$dV^i = \frac{1}{N-1} \sum_{j=1, j \neq i}^{N} q_j F(X^i - X^j) \, dt - \gamma V^i \, dt + \sigma \, dW^i,$$

or the first order ODE/SDE system

(1.2)
$$dX^i = \frac{1}{N-1} \sum_{j=1: j \neq i}^{N} q_j F(X^i - X^j) \, dt + \sigma \, dW^i, \quad i = 1, \cdots, N,$$

where q_j's are the weights (they can be mass, charges, etc). Note that (1.1) contains the Hamiltonian case when the force F is conservative and $\gamma = \sigma = 0$. If one discretizes (1.1) or (1.2) directly, the computational cost per time step is $O(N^2)$, which is undesired for large N. In the case of fast enough decaying interactions, the fast multipole method (FMM) [43] can reduce the complexity per iteration to $O(N)$. However, the implementation is not easy, as it needs some advanced data structures.

The authors proposed in [25] a simple random algorithm, called the random batch method (RBM), to reduce the computation cost per time step from $O(N^2)$ to $O(N)$ for indistinguishable particles (thus with the same weight). The idea was to apply the "mini-batch" idea, famous for its application in the so-called stochastic gradient descent (SGD) [42, 8, 9] in machine learning. Later on, the "mini-batch" was used to develop a Markov chain Monte Carlo method, called the stochastic gradient Langevin dynamics (SGLD), by Welling and Teh for Bayesian inference [46]. The random binary collisions were also used to simulate the Boltzmann equation [6, 40, 4] or the mean-field equation for flocking [1]. How to design the mini-batch strategy depends on the specific applications. For interacting particle systems, the strategy in [25] is to do random grouping at each time interval and then let the particles interact within the groups on that small time interval. Compared with FMM, the accuracy is lower (halfth order in time step), but RBM is simpler to implement and is valid for more general potentials (e.g., the stochastic variational gradient descent (SVGD) ODE [35, 29]). Intuitively, the method converges because there is time average in time, and thus the convergence is like that in the law of large number, but in time (see [25] for a more detailed explanation). Moreover, RBM converges in the regime of the mean field limit ([44, 20, 28]). In fact, the method is asymptotic-preserving regarding the mean-field limit, which means the algorithm can approximate the one-marginal distribution with error bound independent of N.

The error defined in [25] has a clean form and the proof there was simplified due to the exchangeability of the particles. If the particles carry weights, they are not exchangeable. Even so, it is clear that RBM can be equally well applied for interacting particles in multispecies and with weights (see Algorithm 1 below). The goal of this paper is to analyze the convergence of RBM for interacting particles with weights or multispecies. We will discuss both strong and weak convergences, which rely on different techniques so that the error estimates are independent of N. The proof of strong convergence follows closely the proof in [25] for indistinguishable particles, but there are still some differences: since there is no exchangeability, we have to use a kind of weighted average of the errors; some refined auxiliary lemmas have to be used here, compared with those in [25]. The weak convergence is for the empirical

measures instead of the one marginal distribution as in [25]. The proof of the weak convergence is based on the semigroup generated by the backward equations and the one generated by RBM (see [18] for the semigroup techniques used for SGD). Some sharp estimates for the derivatives of the backward equations have to be done so that the weak convergence is uniform in N. The weak convergence analysis is also illustrating for the convergence of RBM for N-body Liouville equations.

The rest of the paper is organized as follows. In section 2, we introduce some basic notations and give a detailed description of RBM. Section 3 is devoted to the proof of strong convergence. In section 4, we show the weak first order convergence, where a key proposition for the estimates of derivatives of the backward equation is needed.

2. Setup and notations. We consider interacting particles with weights. Since in practice, there can be external field, and also there are some applications where the interaction kernels depend on the two specific particles (for example, the SVGD ODE [29]), we consider in general the following first order system

$$(2.1) \quad dX^i = b(X^i)\,dt + \frac{1}{N-1}\sum_{j:j\neq i} m_j K_{ij}(X^i, X^j)\,dt + \sigma\,dW^i, \quad i = 1, 2, \ldots, N.$$

The argument in this paper can be generalized to second order systems without difficulty, which we omit. In (2.1), $b(\cdot)$ is the external force field, $\{W^i\}'s$ are some given independent d dimensional Wiener processes (the standard Brownian motions) and we impose

$$(2.2) \qquad\qquad\qquad\qquad m_j \geq 0.$$

Note that this model includes the cases for particles with multispecies since the signs of the interaction can be included into K_{ij} and for (1.2) $m_j = |q_j|$.

For notational convenience, we define

$$(2.3) \qquad\qquad F_i(x) := b(x_i) + \frac{1}{N-1}\sum_{j:j\neq i} m_j K_{ij}(x_i, x_j),$$

where

$$x := (x_1, \ldots, x_N) \in \mathbb{R}^{Nd}.$$

The (random) empirical probability measure corresponding to (2.1) is given by

$$(2.4) \qquad\qquad \mu^N(t) := \frac{1}{N}\sum_{j=1}^{N} \omega_j \delta(x - X^j(t)),$$

where

$$\omega_j = \frac{Nm_j}{\sum_j m_j}.$$

Assumption 2.1. We assume there are positive constants A, M independent of N such that

$$(2.5) \qquad\qquad \frac{1}{N}\sum_{j=1}^{N} m_j = M, \quad \max_j |m_j| \leq A.$$

With the assumption above, we find

$$(2.6) \qquad\qquad\qquad\qquad \omega_j = O(1).$$

Below are some assumptions on the external field and interaction kernels.

Assumption 2.2. Moreover, we assume $b(\cdot)$ is one-sided Lipschitz:

$$(2.7) \qquad (z_1 - z_2) \cdot (b(z_1) - b(z_2)) \le \beta |z_1 - z_2|^2$$

for some constant β and that $b, \nabla b$ have polynomial growth

$$(2.8) \qquad |b(z)| + |\nabla b| \le C(1 + |z|)^q.$$

The functions $K_{ij}(\cdot, \cdot)$ and their derivatives up to second order are uniformly bounded in $1 \le i, j \le N$.

2.1. RBM and mathematical setup. We now give some detailed explanation of RBM proposed in [25], when applied on (2.1). Suppose the computational interval is $[0, T]$. We pick a small time step τ and define the discrete time grids

$$(2.9) \qquad t_k := k\tau, \quad k \in \mathbb{N}.$$

The number of iteration for the algorithm is

$$(2.10) \qquad N_T := \left\lceil \frac{T}{\tau} \right\rceil.$$

At each time grid t_k, we divide the

$$N = np$$

particles into n small batches with equal size p ($p \ll N$, often $p = 2$) randomly. We have assumed p divides N for convenience. Denote the n batches by $\mathcal{C}_q, q = 1, \ldots, n$, and then each particle only interacts particles within its own batch. The detail is shown in Algorithm 1. Clearly, each iteration contains two main steps: (1) Randomly dividing the particles into n batches (implemented by random permuation, costing $O(N)$ [17]); (2) particles interact inside batches only.

Algorithm 1 RBM

1: **for** k in $1 : \lceil T/\tau \rceil$ **do**
2:     Divide $\{1, 2, \ldots, pn\}$ into n batches randomly.
3:     **for** each batch $\mathcal{C}_q$ **do**
4:         Update $\tilde{X}^i$'s ($i \in \mathcal{C}_q$) by solving the following SDE with $t \in [t_{k-1}, t_k)$.

$$(2.11) \qquad d\tilde{X}^i = b(\tilde{X}^i)dt + \frac{1}{p-1} \sum_{j \in \mathcal{C}_q, j \ne i} m_j K_{ij}(\tilde{X}^i, \tilde{X}^j)dt + \sigma dW^i.$$

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

Above, the Wiener process W^i (Brownian motion) used is the same as in (2.1).

Remark 2.1. For particles with weights, it is desirable to get some random batches by importance sampling. This is left for future study.

We denote $\mathcal{C}_q^{(k)}$, $1 \le q \le n$ the batches at t_k, and

$$(2.12) \qquad \mathcal{C}^{(k)} := \{\mathcal{C}_1^{(k)}, \ldots, \mathcal{C}_n^{(k)}\}$$

will denote the random division of batches at t_k. It is standard by the Kolmogorov extension theorem [16] that there exists a probability space $(\Omega, \mathcal{F}, \mathbb{P})$ so that the random variables $\{X_0^i, W^i, \mathcal{C}^{(k)} : 1 \le i \le N, k \ge 0\}$ are on this probability space, and they are all independent. As usual, we use $\mathbb{E}$ to denote the integration on Ω with respect the probability measure $\mathbb{P}$. For the convenience of the analysis, we introduce the $L^2(\mathbb{P})$ norm as

$$\|v\| = \sqrt{\mathbb{E}|v|^2}. \tag{2.13}$$

Define the filtration $\{\mathcal{F}_k\}_{k \ge 0}$ by

$$\mathcal{F}_{k-1} = \sigma(X_0^i, W^i(t), \mathcal{C}^{(j)}; t \le t_{k-1}, j \le k-1). \tag{2.14}$$

In other words, $\mathcal{F}_{k-1}$ is the σ-algebra generated by the initial values X_0^i, $W^i(t)$, and $\mathcal{C}^{(j)}$ for $i = 1, \ldots, N$, $t \le t_{k-1}$ and $j \le k-1$. Hence, $\mathcal{F}_{k-1}$ contains the information of how batches are constructed for $t \in [t_{k-1}, t_k)$. We also introduce the filtration $\{\mathcal{G}_k\}_{k \ge 0}$ by

$$\mathcal{G}_{k-1} = \sigma(X_0^i, W^i(t), \mathcal{C}^{(j)}; t \le t_{k-1}, j \le k-2). \tag{2.15}$$

If we use $\sigma(\mathcal{C}^{(k-1)})$ to mean the σ-algebra generated by $\mathcal{C}^{(k-1)}$, the random division of batches at t_{k-1}, then $\mathcal{F}_{k-1} = \sigma(\mathcal{G}_{k-1} \cup \sigma(\mathcal{C}^{(k-1)}))$.

For further discussion, given some random batches $\mathcal{C}$, we define the random variables

$$I_{ij} = \begin{cases} 1, & i, j \text{ are in the same batch,} \\ 0, & \text{otherwise,} \end{cases} \quad 1 \le i, j \le N. \tag{2.16}$$

We will focus on the approximation error of $\tilde{X}$ for X for $t \in [0, T]$. In particular, we define the error process

$$Z^i = \tilde{X}^i - X^i, \quad i = 1, \ldots, N. \tag{2.17}$$

2.2. A comment about interacting particles with multispecies. In applications, the most important cases where particles carry weights are the multispecies cases. For example, when we simulate the microscopic particles for the Poisson–Boltzmann equations [2, 31], we need to consider charged particles with different valences, in particular

$$dX^i = \frac{1}{N-1} \sum_{j=1:j \neq i}^{N} Q_j z_i z_j F(X^i - X^j) \, dt + \sigma \, dW^i, \quad i = 1, \ldots, N, \tag{2.18}$$

where $z_i = \pm 1$ represents whether the charge is positive or negative and $Q_j \ge 0$ is the absolute value of the charges. In this case, we can define

$$K_{ij}(x, y) = z_i z_j F(x - y), \tag{2.19}$$

and this reduces to (2.1).

Also, people may care about different densities for different species. Similar to (2.4), one can compute the empirical measures for all these species separately. The empirical measure (2.4) is then a mixture of them. Hence, the model (2.1) is rich enough to include interacting particles with disparate species and weights. Below, we will address these uniformly under the framework of (2.1).

3. The strong convergence. Since there is no exchangeability, we have to use weighted average of the errors. We consider

$$(3.1) \qquad J(t) := \frac{1}{2N} \sum_{i=1}^{N} m_i \mathbb{E} |\tilde{X}^i - X^i|^2,$$

which can be used for the strong convergence. As a common convention, the "$\frac{1}{2}$" prefactor is used for energies of quadratic forms. Recently, a certain quantum correspondence of this error has been used by Golse et al. to prove the convergence of RBM for N-body Schrödinger equations [21].

THEOREM 3.1. *Suppose Assumptions* 2.1–2.2 *hold. Then, for step size smaller than some constant independent of N, it holds that*

$$(3.2) \qquad \sup_{t \le T} J(t) \le C(T) \left(\frac{\tau}{p-1} + \tau^2 \right)$$

for some constant $C(T) > 0$ independent of N, p. Hence, the strong error is controlled as

$$(3.3) \qquad \sup_{t \le T} \sqrt{J(t)} \lesssim \sqrt{\frac{\tau}{p-1} + \tau^2}.$$

Define the error of the random approximation for the interacting force by

$$(3.4) \qquad \begin{aligned} \chi_i(x) &:= \frac{1}{p-1} \sum_{j \in \mathcal{C}_\theta, j \ne i} m_j K_{ij}(x_i, x_j) - \frac{1}{N-1} \sum_{j:j \ne i} m_j K_{ij}(x_i, x_j) \\ &= \frac{1}{p-1} \sum_{j:j \ne i} I_{ij} m_j K_{ij}(x_i, x_j) - \frac{1}{N-1} \sum_{j:j \ne i} m_j K_{ij}(x_i, x_j), \end{aligned}$$

where $x_i \in \mathbb{R}^d$, $\mathcal{C}_\theta (1 \le \theta \le n)$ is the random batch that contains x_i in a random division, and again $x = (x_1, \ldots, x_N) \in \mathbb{R}^{Nd}$.

We have the following facts.

LEMMA 3.2. *For $i \ne j$, it holds that*

$$(3.5) \qquad \mathbb{E} I_{ij} = \frac{p-1}{N-1},$$

and for distinct i, j, ℓ, it holds that

$$(3.6) \qquad \mathbb{E} I_{ij} I_{i\ell} = \frac{(p-1)(p-2)}{(N-1)(N-2)}.$$

They have been proved in the proof of [25, Lemma 3.1] so we omit the proofs here. Using Lemma 3.2, one obtains the following consistency of RBM.

LEMMA 3.3. *For given $x = (x_1, \ldots, x_N) \in \mathbb{R}^{Nd}$, it holds that*

$$(3.7) \qquad \mathbb{E} \chi_i(x) = 0.$$

Moreover, the second moment is given by

$$(3.8) \qquad \mathbb{E} |\chi_i(x)|^2 = \left(\frac{1}{p-1} - \frac{1}{N-1} \right) \Lambda_i(x),$$

where

$$(3.9) \qquad \Lambda_i(x) := \frac{1}{N-2} \sum_{j:j \ne i} \left| m_j K_{ij}(x_i, x_j) - \frac{1}{N-1} \sum_{\ell:\ell \ne i} m_\ell K_{i\ell}(x_i, x_\ell) \right|^2.$$

Though slightly different from [25, Lemma 3.1], the proof is essentially the same, and we omit.

We move to some important additional estimates.

LEMMA 3.4. *Let X^i and $\tilde{X}^i$ be solutions to* (2.1) *and* (2.11), *respectively. Suppose Assumptions* 2.1–2.2 *hold. Then,*

$$\text{(3.10)} \qquad \sup_{t \leq T}(\mathbb{E}|X^i(t)|^q + \mathbb{E}|\tilde{X}^i(t)|^q) \leq C_q.$$

Besides, for any $k > 0$ and $q \geq 2$,

$$\text{(3.11)} \qquad \sup_{t \in [t_{k-1}, t_k)} \left| \mathbb{E}(|\tilde{X}^i(t)|^q|\mathcal{F}_{k-1}) \right| \leq C_1 |\tilde{X}^i(t_{k-1})|^q + C_2$$

holds almost surely.

Moreover, almost surely, it holds that

$$\text{(3.12)} \qquad \begin{aligned} |\mathbb{E}(\tilde{X}^i(t) - \tilde{X}^i(t_{k-1})|\mathcal{F}_{k-1})| &\leq C(1 + |\tilde{X}^i(t_{k-1})|^q)\tau, \\ \left| \mathbb{E}\left(|\tilde{X}^i(t) - \tilde{X}^i(t_{k-1})|^2|\mathcal{F}_{k-1}\right) \right| &\leq C(1 + |\tilde{X}^i(t_{k-1})|^q)\tau. \end{aligned}$$

Note that the second equation in (3.12) is different from that in [25]. In fact, the proof in [25] has a small gap, and one needs this refined estimate to fill in that gap as well ([25], page 13, from line 17 to line 19, the variable $\mathbb{E}(|\delta\tilde{X}^i|^2 + |\delta\tilde{X}^j|^2|\mathcal{F}_{m-1})$ is not independent of $\mathcal{C}_\theta$; to get line 19, we need this refined version). Though the proof is not hard, we still attach it in Appendix A for a reference.

The following lemma is an improved version of [25, Lemma 3.2], which is very important to establish the strong convergence for particles with weights considered in this paper.

LEMMA 3.5. *Fix $i \in \{1, \ldots, N\}$. Let $\mathcal{C}_\theta$ be the random batch of size p that contains i in the random division. Let Y_j $(1 \leq j \leq N)$ be N random variables (or random vectors) that are independent of $\mathcal{C}_\theta$. Then, for $p \geq 2$,*

$$\text{(3.13)} \qquad \left\| \frac{1}{p-1} \sum_{j \in \mathcal{C}_\theta, j \neq i} Y_j \right\| \leq \left(\frac{1}{N-1} \sum_{j:j \neq i} \|Y_j\|^2 \right)^{1/2}.$$

Proof. By the definition and independence,

$$\begin{aligned} \left\| \frac{1}{p-1} \sum_{j \in \mathcal{C}_\theta, j \neq i} Y_j \right\|^2 &= \frac{1}{(p-1)^2} \mathbb{E}\left| \sum_{j:j \neq i} I_{ij} Y_j \right|^2 \\ &= \frac{1}{(p-1)^2} \sum_{j,\ell:j \neq i, \ell \neq i} \mathbb{E}(I_{ij}I_{i\ell})\mathbb{E}(Y_j \cdot Y_\ell) \\ &\leq \frac{1}{(p-1)^2} \left[\sum_{j,\ell:j \neq i, \ell \neq i, j \neq \ell} \frac{(p-1)(p-2)}{(N-1)(N-2)} \|Y_j\|\|Y_\ell\| \right. \\ &\qquad \left. + \sum_{j:j \neq i} \mathbb{E}I_{ij}\|Y_j\|^2 \right] \\ &=: R_1 + R_2. \end{aligned}$$

Note that the independence is used in the second equality. The first inequality is due to Lemma 3.2.

It is easy to calculate

$$R_1 \le \frac{p-2}{(p-1)(N-1)(N-2)} \sum_{j,\ell:j\neq i,\ell\neq i,j\neq\ell} \left(\frac{1}{2}\|Y_j\|^2 + \frac{1}{2}\|Y_\ell\|^2 \right)$$
$$= \frac{p-2}{(p-1)(N-1)} \sum_{j:j\neq i} \|Y_j\|^2,$$

while

$$R_2 = \frac{1}{(p-1)(N-1)} \sum_{j:j\neq i} \|Y_j\|^2.$$

Hence,

$$R_1 + R_2 \le \frac{1}{N-1} \sum_{j:j\neq i} \|Y_j\|^2.$$

The claim thus follows. □

We now consider the error process defined in (2.17). The derivative of Z^i is clearly given by

(3.14)
$$\frac{d}{dt}Z^i = [b(\tilde{X}^i) - b(X^i)] + \frac{1}{p-1} \sum_{j\in\mathcal{C}_\theta, j\neq i} m_j K_{ij}(\tilde{X}^i, \tilde{X}^j) - \frac{1}{N-1} \sum_{j:j\neq i} m_j K_{ij}(X^i, X^j),$$

where again $\mathcal{C}_\theta$ is the random batch in $\mathcal{C}$ that contains i.

Define

(3.15)
$$\delta K_{ij}(t) := K_{ij}(\tilde{X}^i(t), \tilde{X}^j(t)) - K_{ij}(X^i(t), X^j(t)).$$

The right-hand side of (3.14) can then be written as

(3.16)
$$\frac{d}{dt}Z^i = [b(\tilde{X}^i) - b(X^i)] + \frac{1}{N-1} \sum_{j:j\neq i} m_j \delta K_{ij} + \chi_i(\tilde{X})$$
$$= [b(\tilde{X}^i) - b(X^i)] + \frac{1}{p-1} \sum_{j\in\mathcal{C}_\theta, j\neq i} m_j \delta K_{ij} + \chi_i(X).$$

With this, one can obtain the following simple lemma.

LEMMA 3.6. *For $t \in [t_{k-1}, t_k)$,*

(3.17)
$$\|Z^i(t) - Z^i(t_{k-1})\| \le C\tau.$$

Also, almost surely,

(3.18)
$$|Z^i(t)| \le |Z^i(t_{k-1})|(1 + C\tau) + C\tau.$$

Proof. Integrating (3.16) from t_{k-1} to t, since b has polynomial growth, the claim for $\|Z^i(t) - Z^i(t_{k-1})\|$ is then an easy consequence of the q-moment estimates in Lemma 3.4.

Dotting (3.16) with Z^i and using the one-sided Lipschitz condition in Assumption 2.2, one has

$$\frac{1}{2}\frac{d}{dt}|Z^i|^2 \leq C|Z^i|^2 + C|Z^i|.$$

Hence, almost surely, it holds that

$$\frac{d}{dt}|Z^i| \leq C|Z^i| + C.$$

By Grönwall's inequality, the second claim then follows (for τ less than some constant independent of N). $\qquad\square$

We now give the proof of the strong convergence in Theorem 3.1.

Proof of Theorem 3.1. First,

$$\frac{dJ(t)}{dt} = \frac{1}{N}\sum_{i=1}^{N} m_i \mathbb{E}\Big\{Z^i \cdot [(F_i(\tilde{X}) - F_i(X)) + \chi_i(\tilde{X})]\Big\}.$$

The first term is easy to bound by the one-sided Lipschitz condition of b and the conditions of K_{ij} in Assumption 2.2:

$$\frac{1}{N}\sum_{i=1}^{N} m_i \mathbb{E}\Big\{Z^i \cdot (F_i(\tilde{X}) - F_i(X))\Big\} \leq \frac{C}{N}\sum_i m_i \mathbb{E}|Z^i|^2$$
$$+ \frac{C}{N(N-1)}\sum_{i\neq j} m_i m_j \mathbb{E}(|Z^i|^2 + |Z^i||Z^j|).$$

This is clearly bounded by $J(t)$.

Now, we focus on the second term. The technique is the same as in our previous work [25], but some special modifications are needed for our problem here:

$$\frac{1}{N}\sum_{i=1}^{N} m_i \mathbb{E}\Big\{Z^i(t) \cdot \chi_i(\tilde{X}(t))\Big\} = \frac{1}{N}\sum_{i=1}^{N} m_i \mathbb{E}\Big\{Z^i(t_{k-1}) \cdot \chi_i(\tilde{X}(t))\Big\}$$
$$+ \frac{1}{N}\sum_{i=1}^{N} m_i \mathbb{E}\Big\{(Z^i(t) - Z^i(t_{k-1})) \cdot \chi_i(\tilde{X}(t))\Big\}$$
$$=: I_1 + I_2.$$

Step 1—Estimate of I_1.

For I_1, using the consistency result in Lemma 3.3, one has

$$I_1 = \frac{1}{N}\sum_{i=1}^{N} m_i \mathbb{E}\left\{Z^i(t_{k-1}) \cdot \Big[\chi_i(\tilde{X}(t)) - \chi_i(\tilde{X}(t_{k-1}))\Big]\right\}.$$

In fact,

$$\mathbb{E}\left\{Z^i(t_{k-1}) \cdot \chi_i(\tilde{X}(t_{k-1}))\right\} = \mathbb{E}\left\{Z^i(t_{k-1}) \cdot \mathbb{E}\Big[\chi_i(\tilde{X}(t_{k-1}))|\mathcal{G}_{k-1}\Big]\right\} = \mathbb{E}0 = 0.$$

This is the only place where the σ-algebra $\mathcal{G}_{k-1}$ is used.

Note that

$$
\mathbb{E}\left\{Z^i(t_{k-1}) \cdot \left[\chi_i(\tilde{X}(t)) - \chi_i(\tilde{X}(t_{k-1}))\right]\right\}
$$
$$
= \mathbb{E}\left(Z^i(t_{k-1}) \cdot \mathbb{E}(\chi_i(\tilde{X}(t)) - \chi_i(\tilde{X}(t_{k-1}))|\mathcal{F}_{k-1})\right)
$$
$$
\leq C\|Z^i(t_{k-1})\| \left\|\mathbb{E}[\chi_i(\tilde{X}(t)) - \chi_i(\tilde{X}(t_{k-1}))|\mathcal{F}_{k-1}]\right\|.
$$

Using the definition of χ_i (3.4), one has

$$
\mathbb{E}\left[\chi_i(\tilde{X}(s)) - \chi_i(\tilde{X}(t_{k-1}))|\mathcal{F}_{k-1}\right]
$$
$$
= \frac{1}{p-1} \sum_{j \in \mathcal{C}_\theta, j \neq i} m_j \mathbb{E}(\delta \tilde{K}^{ij}|\mathcal{F}_{k-1}) - \frac{1}{N-1} \sum_{j: j \neq i} m_j \mathbb{E}(\delta \tilde{K}^{ij}|\mathcal{F}_{k-1}),
$$

where

$$
\delta \tilde{K}^{ij} = K_{ij}(\tilde{X}^i(s), \tilde{X}^j(s)) - K_{ij}(\tilde{X}^i(t_{k-1}), \tilde{X}^j(t_{k-1})).
$$

We first estimate $\mathbb{E}(\delta \tilde{K}^{ij}|\mathcal{F}_{k-1})$. Denote $\delta \tilde{X}^j := \tilde{X}^j(s) - \tilde{X}^j(t_{k-1})$. Performing Taylor expansion around t_{k-1}, one has

$$
\delta \tilde{K}^{ij} = (\nabla_{x_i} K_{ij}|_{t_{k-1}} \cdot \delta \tilde{X}^i + \nabla_{x_j} K_{ij}|_{t_{k-1}} \cdot \delta \tilde{X}^j) + \frac{1}{2} M : [\delta \tilde{X}^i, \delta \tilde{X}^j] \otimes [\delta \tilde{X}^i, \delta \tilde{X}^j],
$$

with M being a random variable (tensor) bounded by $\|\nabla^2 K_{ij}\|_\infty$. By (3.12), one finds that

$$
|\mathbb{E}(\delta \tilde{K}^{ij}|\mathcal{F}_{k-1})| \leq C(1 + |\tilde{X}^i(t_{k-1})|^q + |\tilde{X}^j(t_{k-1})|^q)\tau.
$$

The right-hand side is independent of $\mathcal{C}_\theta$, and this is the place where we need the almost surely bound (3.12). Applying Lemma 3.5, one has

$$
\left\|\frac{1}{p-1} \sum_{j \in \mathcal{C}_\theta, j \neq i} m_j \mathbb{E}(\delta \tilde{K}^{ij}|\mathcal{F}_{k-1})\right\| \leq C\tau \left(1 + \left(\frac{1}{N-1} \sum_{j: j \neq i} \||\tilde{X}^j(t_{k-1})|^{q_1}\|^2\right)^{1/2}\right) \leq C\tau.
$$

The term $\|\frac{1}{N-1} \sum_{j: j \neq i} m_j \mathbb{E}(\delta \tilde{K}^{ij}|\mathcal{F}_{k-1})\|$ is much easier to estimate, and it is also bounded by $C\tau$.

Hence

$$
\mathbb{E}Z^i(t_{k-1}) \cdot \left[\chi_i(\tilde{X}(t)) - \chi_i(\tilde{X}(t_{k-1}))\right] \leq C\|Z^i(t_{k-1})\|\tau \leq C\|Z^i(t)\|\tau + C\tau^2.
$$

Then,

$$
I_1 \leq \left(\frac{C}{N} \sum_{i=1}^N m_i \|Z^i(t)\|\tau\right) + C\tau^2 \leq \delta \frac{1}{N} \sum_{i=1}^N m_i \|Z^i(t)\|^2 + C(\delta)\tau^2.
$$

Step 2—Estimate of I_2.

We decompose

$$
I_2 = \frac{1}{N} \sum_{i=1}^N m_i \mathbb{E}(Z^i(t) - Z^i(t_{k-1})) \cdot [\chi_i(\tilde{X}(t)) - \chi_i(X(t))]
$$
$$
+ \frac{1}{N} \sum_{i=1}^N m_i \mathbb{E}(Z^i(t) - Z^i(t_{k-1})) \cdot \chi_i(X(t)) =: I_{21} + I_{22}.
$$

We first consider I_{21}. Clearly,

$$I_{21} \leq \frac{1}{N} \sum_{i=1}^{N} m_i C\tau \|\chi_i(\tilde{X}(t)) - \chi_i(X(t))\|.$$

Note that

(3.19)

$$\|\chi_i(\tilde{X}(t)) - \chi_i(X(t))\| \leq \left\| \frac{1}{p-1} \sum_{j \in \mathcal{C}_\theta, j \neq i} m_j \delta K_{ij}(t) \right\| + \left\| \frac{1}{N-1} \sum_{j:j \neq i} m_j \delta K_{ij}(t) \right\|.$$

Since K_{ij} is Lipschitz continuous with Lipschitz constant L independent of i, j,

$$|\delta K_{ij}(t)| \leq L(|Z^i(t)| + |Z^j(t)|),$$

and one thus has

(3.20) $$\left\| \frac{1}{p-1} \sum_{j \in \mathcal{C}_\theta, j \neq i} m_j \delta K_{ij}(t) \right\| \leq L \left(\|Z^i(t)\| + \left\| \frac{1}{p-1} \sum_{j \in \mathcal{C}_\theta, j \neq i} m_j |Z^j(t)| \right\| \right).$$

Note that $Z^j(t)$ depends on $\mathcal{C}_\theta$, and we cannot apply Lemma 3.5. Instead, by Lemma 3.6, one has that

$$|Z^i(t)| \leq C|Z^i(t_{k-1})| + C\tau.$$

Since $Z^i(t_{k-1})$ is independent of $\mathcal{C}_\theta$, Lemma 3.5 then gives us that

(3.21)
$$\left\| \frac{1}{p-1} \sum_{j \in \mathcal{C}_\theta, j \neq i} m_j |Z^j(t)| \right\| \leq C \left(\frac{1}{N-1} \sum_{j:j \neq i} m_j^2 \|Z^j(t_{k-1})\|^2 \right)^{1/2} + C\tau$$

$$\leq C \left(\frac{1}{N-1} \sum_{j:j \neq i} m_j \|Z^j(t_{k-1})\|^2 \right)^{1/2} + C\tau.$$

Remark 3.7. This is where we need Lemma 3.5, a refined version of [25, Lemma 3.2]. If using [25, Lemma 3.2], one controls $\|\frac{1}{p-1} \sum_{j \in \mathcal{C}_\theta, j \neq i} m_j |Z^j(t_{k-1})|\|$ by $\sup_j m_j \|Z^j(t_{k-1})\|$, which is not enough to close the estimate.

It is straightforward to obtain

$$\frac{1}{N} \sum_{i=1}^{N} m_i \tau \left(\frac{1}{N-1} \sum_{j:j \neq i} m_j \|Z^j(t_{k-1})\|^2 \right)^{1/2} \leq \delta \frac{1}{N} \sum_i m_i \|Z^i(t_{k-1})\|^2 + C(\delta)\tau^2,$$

where Assumption 2.1 has been used, and

$$\frac{1}{N} \sum_{i=1}^{N} m_i \tau \|Z^i(t)\| \leq \delta \frac{1}{N} \sum_{i=1}^{N} m_i \|Z^i(t)\|^2 + C(\delta)\tau^2.$$

Also, the second term in (3.19) is similarly estimated (but much simpler), so we omit the details. Moreover,

$$(3.22) \qquad \|Z^i(t_{k-1})\|^2 \le (\|Z^i(t)\| + C\tau)^2 \le 2\|Z^i(t)\|^2 + 2C^2\tau^2.$$

With these, we find (with δ redefined)

$$I_{21} \le \delta \frac{1}{N} \sum_i m_i \|Z^i(t)\|^2 + C(\delta)\tau^2.$$

We now consider I_{22}. We first recall

$$(3.23) \qquad \frac{d}{dt} Z^i = [b(\tilde{X}^i) - b(X^i)] + \frac{1}{p-1} \sum_{j \in \mathcal{C}_\theta, j \ne i} m_j \delta K_{ij}(t) + \chi_i(X(t)).$$

Note

$$|b(\tilde{X}^i) - b(X^i)| \le \int_0^1 |(\tilde{X}^i - X^i) \cdot \nabla b((1-\sigma)\tilde{X}^i + \sigma X^i)| \, d\sigma,$$

and $\nabla b((1-\sigma)\tilde{X}^i + \sigma X^i)$ is controlled by $C(|\tilde{X}^i|^q + |X^i|^q)$ for some $q > 0$. Hence,

$$\mathbb{E}|b(\tilde{X}^i) - b(X^i)\|\chi_i(X(t'))| \le \|\chi_i(X(t'))\|_\infty \|\tilde{X}^i - X^i\|(\mathbb{E}(|\tilde{X}^i|^{q_1} + |X^i|^{q_1})^2)^{1/2}$$
$$\le C\|Z^i(t)\|,$$

by (3.10).

Integrating (3.23) in time over $[t_{k-1}, t]$, then dotting with $\chi_i(X(t))$, and taking the expectation, one gets

$$(3.24) \quad \left|\mathbb{E}((Z^i(t) - Z^i(t_{k-1})) \cdot \chi_{m,i}(X(t)))\right| \le C \int_{t_{k-1}}^t \|Z^i(s)\| ds$$

$$+ \int_{t_{k-1}}^t \mathbb{E}\left[\left(\frac{1}{p-1} \sum_{j \in \mathcal{C}_\theta, j \ne i} m_j \delta K_{ij}(s)\right) \cdot \chi_i(X(t))\right] ds + \mathbb{E} \int_{t_{k-1}}^t \chi_i(X(s)) \cdot \chi_i(X(t)) \, ds.$$

Applying (3.20) and (3.21), the second term on the right-hand side of (3.24) is bounded by

$$C \left(\frac{1}{N-1} \sum_{j: j \ne i} m_j^2 \|Z^j(t_{k-1})\|^2\right)^{1/2} \tau + C\tau^2.$$

Similarly as we estimate I_{21}, this is controlled by $\delta \frac{1}{N} \sum_i m_i \|Z^i(t_{k-1})\|^2 + C(\delta)\tau^2$. The last term on the right hand of (3.24) is controlled by Lemma 3.3 (in particular, (3.8)):

$$\left[\frac{1}{p-1} - \frac{1}{N-1}\right] \|\Lambda_i\|_\infty \tau.$$

Therefore,

$$I_{22} \le \delta \frac{1}{N} \sum_i m_i \|Z^i(t_{k-1})\|^2 + C(\delta)\tau^2 + \left[\frac{1}{p-1} - \frac{1}{N-1}\right] \|\Lambda_i\|_\infty \tau.$$

Finally, taking all those estimates together and noting (3.22), one has the following estimate:

$$\frac{d}{dt} J \le CJ + C(\delta)\tau^2 + \left[\frac{1}{p-1} - \frac{1}{N-1}\right] \|\Lambda_i\|_\infty \tau.$$

The claim then follows by Grönwall's lemma. □

Remark 3.8. The strong convergence can imply the convergence of marginal distribution for indistinguishable particles. See [25] for more details. The weak convergence discussed below, however, is for empirical measure (2.4).

4. The weak convergence and RBM for backward equations. In practice, one may be more interested in the distributions of the particles instead of the trajectories of $X^i(t)$. Hence, the error (3.1) is not suitable for this purpose, and we seek to study the distributions, and the weak convergence, as commonly used in the numerical SDE literature [38, 27]. Roughly speaking, we say a family of measures ν^λ converges to some measure ν weakly under some limit $\lambda \to \lambda_0$ if for any suitable test function φ, it holds that

$$\int \varphi \, d\nu^\lambda =: \langle \nu^\lambda, \varphi \rangle \to \langle \nu, \varphi \rangle := \int \varphi \, d\nu, \quad \lambda \to \lambda_0.$$

For our problem, we consider the empirical measure corresponding to (2.11)

$$(4.1) \qquad \mu^{N,\tau}(t) := \frac{1}{N} \sum_{j=1}^{N} \omega_j \delta(x - \tilde{X}^j(t)),$$

where again

$$\omega_j = \frac{N m_j}{\sum_j m_j}.$$

Now, we care about the convergence of RBM in weak sense. Both $\mu^{N,\tau}$ defined in (4.1) and μ^N in (4.1) are random measures defined on Borel sets that are subsets of $\mathbb{R}^d$. Taking into consideration the distribution of the random measures, they finally give some measures on $\mathbb{R}^d$ by

$$\nu^{N,\tau}(E) := \mathbb{E}\mu^{N,\tau}(E), \quad \nu^N(E) := \mathbb{E}\mu^N(E),$$

for all $E \subset \mathbb{R}^d$ that are Borel measurable. The convergence of RBM in weak sense then reduces to the weak convergence of $\nu^{N,\tau}$ to ν^N as $\tau \to 0$. Hence, to show that the empirical measures given by X and $\tilde{X}$ are close in law, we pick a test function $\varphi \in C_b^\infty(\mathbb{R}^d)$ and hope to show that the weak error defined below is small:

$$(4.2) \qquad E_k := \left| \frac{1}{N} \sum_{i=1}^{N} \omega_i \mathbb{E}\varphi(\tilde{X}^i(k\tau)) - \frac{1}{N} \sum_{i=1}^{N} \omega_i \mathbb{E}\varphi(X^i(k\tau)) \right|.$$

Remark 4.1. Traditionally, the test functions used for schemes of numerical SDEs are those with polynomial growth at infinity [37, 38]. We used C_b^∞ as the test functions as done nowadays [3, 32], which will induce a weaker topology that disregards high order moments. If one has the corresponding moments control, the convergence using these two types of test functions will be the same.

For the weak convergence, we need some different assumptions on b and K_{ij}.

Assumption 4.1. The functions b, K_{ij} are C^4 and the derivatives of b and K_{ij} up to order 4 are uniformly bounded (uniform in i, j).

Remark 4.2. Proof of weak convergence using Assumption 2.2 instead of 4.1 should also be completed. Using Assumption 4.1 makes the proof based on semi-group technique elegant (see the details below).

We now state the main theorem for the weak convergence of RBM.

THEOREM 4.3. *Under Assumptions* 2.1 *and* 4.1, *the RBM system* (2.11) *converges weakly with first order to* (2.1), *in terms of the empirical measures* (4.1) *and* (2.4). *In particular, the weak error defined in* (4.2) *satisfies*

$$\sup_{k:k\tau \leq T} E_k \leq C\tau, \tag{4.3}$$

where $C = C(\varphi, T)$ *is independent of* N, τ *but depends on* φ *and* T.

Usually, in numerical SDEs, to prove the weak convergence, one makes use of the backward Kolmogorov equation ("backward equation" for short) which is defined in the same Euclidean space where the process takes values. For our problem, we need to lift the Euclidean space from $\mathbb{R}^d$ to $\mathbb{R}^{Nd}$. In particular, define

$$u(x,t) := \frac{1}{N} \sum_{i=1}^{N} \omega_i \mathbb{E}[\varphi(X^i(t))|X(0) = x], \tag{4.4}$$

where $x \in \mathbb{R}^{Nd}$ and X is the solution to (2.1). Then, we make use of this function to study the weak convergence. This function u satisfies the following backward equation [41]

$$\partial_t u = \mathcal{L}u := \sum_{i=1}^{N} \left[b(x_i) + \frac{1}{N-1} \sum_{j:j\neq i} m_j K_{ij}(x_i, x_j) \right] \cdot \nabla_{x_i} u + \sum_{i=1}^{N} \frac{1}{2}\sigma^2 \Delta_{x_i} u. \tag{4.5}$$

The operator $\mathcal{L}$ is called the generator of the ODE/SDE (2.1). The Laplacian Δ_{x_i} is given by

$$\Delta_{x_i} := \sum_{j=1}^{d} \partial_{x_i^{(j)} x_i^{(j)}}, \quad x_i = (x_i^{(1)}, \ldots, x_i^{(d)}) \in \mathbb{R}^d. \tag{4.6}$$

The solution semigroup for (4.5) will be denoted by

$$e^{t\mathcal{L}}u(\cdot, 0) := u(x, t). \tag{4.7}$$

By the well-known property of backward equation [41, 30], one has

$$\|u(\cdot, t)\|_\infty = \|e^{t\mathcal{L}}u(\cdot, 0)\|_\infty \leq \|u(\cdot, 0)\|_\infty. \tag{4.8}$$

The function u is defined on $\mathbb{R}^{Nd}$, and naive estimates of the norms for the derivatives will depend on N. This is not sufficient for us to show the N independence of weak convergence for the empirical measure. In fact, it is clear that

$$\|u\|_\infty \leq C \tag{4.9}$$

independent of N. The following proposition then gives the crucial estimates of the derivatives of u.

PROPOSITION 4.4. *For any* i *and* $j \neq i$, *one has*

$$\|\nabla_{x_i} u\|_\infty + \|\nabla_{x_i}^2 u\|_\infty + \|\nabla_{x_i}^3 u\|_\infty + \|\nabla_{x_i}^4 u\|_\infty \leq C(T)\frac{1}{N}, \tag{4.10}$$

and

$$\|\nabla_{x_i} \nabla_{x_j} u\|_\infty + \|\nabla_{x_i} \nabla_{x_j}^2 u\|_\infty + \|\nabla_{x_i}^2 \nabla_{x_j}^2 u\|_\infty \leq C(T)\frac{1}{N^2}. \tag{4.11}$$

Here, $\nabla^2_{x_i}$ is the Hessian matrix. Similarly, $\nabla^4_{x_i} u$ is the fourth order tensor with derivatives of the form $(\prod^4_{k=1} \partial_{x_i^{(j_k)}}) u$ and $j_k \in \{1, \ldots, d\}$. The norm $\|\nabla^4_{x_i} u\|_\infty$ is understood as

$$\sup_x \left[\sum_{j_1, j_2, j_3, j_4} \left(\left(\prod^4_{k=1} \partial_{x_i^{(j_k)}} \right) u \right)^2 \right]^{1/2}.$$

The proof is kind of tedious and is given in Appendix B.

For further discussion, we introduce the generator corresponding to RBM. For $t \in (t_{k-1}, t_k]$

$$\mathcal{L}_C := \sum^N_{i=1} \left[b(x_i) + \frac{1}{p-1} \sum_{j:j\neq i} I^{(k-1)}_{ij} m_j K_{ij}(x_i, x_j) \right] \cdot \nabla_{x_i} + \sum^N_{i=1} \frac{1}{2} \sigma^2 \Delta_{x_i}.$$

We recall that I_{ij} is the indicator for i, j being in the same batch, and we use $I^{(k-1)}_{ij}$ to mean the indicator corresponding to $\mathcal{C}^{(k-1)}$.

The importance of Proposition 4.4 is that one can bound $\mathcal{L}^i u$ uniformly in N, where $\mathcal{L}^i$ means the composition of $\mathcal{L}$ for i times (if $i = 0$, it is the identity operator). This is crucial for establishing the weak convergence uniformly in N. In particular, we have the following.

LEMMA 4.5. *For the function u defined in* (4.4)*, it holds that for $i = 0, 1, 2$ that*

(4.12) $$\|\mathcal{L}^i u\| \leq C, \quad \|\mathcal{L}^i_\mathcal{C} u\| \leq C.$$

Moreover,

(4.13) $$\mathbb{E}\mathcal{L}_\mathcal{C} = \mathcal{L}.$$

Proof. The first assertion is a corollary of Proposition 4.4. The second claim is the one proved in Lemma 3.3. □

To go further, we need to introduce a semigroup associated to RBM. Consider the transition $\tilde{X}(t_{k-1}) \to \tilde{X}(t_k)$; it is clear that this transition gives a time homogeneous Markov chain. Define the operator $\mathcal{S}^{(k)} : C_b(\mathbb{R}^{Nd}) \to C_b(\mathbb{R}^{Nd})$ as

(4.14) $$\mathcal{S}^{(k)} \phi := \mathbb{E}[\phi(\tilde{X}(k\tau)) | \tilde{X}(0) = x].$$

Below, we sometimes use $\mathbb{E}_x$ to mean $\mathbb{E}(\cdot | \tilde{X}(0) = x)$ for convenience. Then, by the Markov property [16], it can be shown that

(4.15) $$\mathcal{S}^{(k)} = \mathcal{S}^k := \mathcal{S} \circ \cdots \circ \mathcal{S}.$$

In fact, for any test function ϕ, we let $u^k := \mathcal{S}^{(k)} \phi$. Then, it holds that

$$\begin{aligned} u^k(x) &= \mathbb{E}[\phi(\tilde{X}(t_k)) | \tilde{X}(0) = x] \\ &= \mathbb{E}[\mathbb{E}[\phi(\tilde{X}(t_k)) | \tilde{X}(t_1)] | \tilde{X}(0) = x] \\ &= \mathbb{E}[u^{k-1}(\tilde{X}(t_1)) | \tilde{X}(0) = x], \end{aligned}$$

where the third equality holds by the Markov property. This means

$$\mathcal{S}^{(k)} = \mathcal{S} \circ \mathcal{S}^{(k-1)}.$$

This then shows that $\{\mathcal{S}^{(k)} : k \geq 0\}$ forms a semigroup. Clearly, $\mathcal{S}^{(k)}$ is nonexpansive in $L^\infty(\mathbb{R}^{Nd})$; i.e.,

(4.16) $$\|\mathcal{S}^{(k)}\phi\|_\infty \leq \|\phi\|_\infty.$$

Our goal is then to show that the semigroup $\{\mathcal{S}^{(k)} : k \geq 0\}$ is close to $e^{t_k\mathcal{L}}$ weakly on test functions of the form

$$\phi(x) = \frac{1}{N}\sum_{i=1}^N \omega_i\varphi(x_i), \quad \varphi \in C_b^\infty(\mathbb{R}^d),$$

instead of all test functions $\phi \in C_b^\infty(\mathbb{R}^{Nd})$ for our purpose.

Remark 4.6. The operator $\mathcal{S}^{(k)}$ is standard for Markov chains, and see [36, section 3.4.2]. Using the semigroup $\{\mathcal{S}^{(k)}\}$ to show weak convergence of Markov chains to time-continuous stochastic processes is not so common in literature. One of such examples is [3] where the semigroup technique is used implicitly for a weak scheme of SDEs. In [18], the semigroup generated by SGD was used explicitly to show its approximation to a time-continuous diffusion process. Using the semigroup $\mathcal{S}^{(k)}$ generated by RBM in this work to show the weak convergence shares similarity with the proof in [18] for SGD, but there is a big difference. Here, we aim to prove the convergence of empirical measures independent of N, so we need to lift the test functions from $\mathbb{R}^d$ to $\mathbb{R}^{Nd}$, and thus the test functions in $\mathbb{R}^{Nd}$ is a subclass of $C_b^\infty(\mathbb{R}^{Nd})$. We also need to control the derivatives of the backward equations for such test functions independent of N.

We now give the proof of the weak convergence.

Proof of Theorem 4.3. The weak error in (4.2) is to compare u in (4.4) and $\mathcal{S}^n\phi$ for test function

(4.17) $$\phi(x) = \frac{1}{N}\sum_{i=1}^N \omega_i\varphi(x_i).$$

By (4.15) ($\mathcal{S}^{(k)} = \mathcal{S}^k$), we need to estimate

$$\mathcal{S}^n\phi(x) - u(x, n\tau) = \sum_{k=1}^n \Big[\mathcal{S}^k u(x, (n-k)\tau) - \mathcal{S}^{k-1}u(x, (n-k+1)\tau)\Big].$$

Clearly, by (4.16), one has

$$|\mathcal{S}^n\phi(x) - u(x, n\tau)| \leq \sum_{k=1}^n \|\mathcal{S}u(x, (n-k)\tau) - u(x, (n-k+1)\tau)\|_\infty.$$

By the definition of S, one has

$$\begin{aligned}
\mathcal{S}u(x, (n-k)\tau) &= \mathbb{E}_x u(\tilde{X}(\tau), (n-k)\tau) \\
&= \mathbb{E}_\mathcal{C}\mathbb{E}_x(u(\tilde{X}(\tau), (n-k)\tau)|\mathcal{C}) \\
&= \mathbb{E}_\mathcal{C} e^{\tau\mathcal{L}_\mathcal{C}}u(x, (n-k)\tau).
\end{aligned}$$

The second equality means that we fix a division of batches to compute the expectation with respect to the Brownian motions, and then average out about the random

batches. For the last equality, we recall that $e^{\tau \mathcal{L}_c}$ means the solution semigroup by $\partial_t v = \mathcal{L}_\mathcal{C} v$.

Note

$$e^{\tau \mathcal{L}_c} u(x, (n-k)\tau) = u(x, (n-k)\tau) + \tau \mathcal{L}_\mathcal{C} u(x, (n-k)\tau)$$
$$+ \int_0^\tau (\tau - z) \mathcal{L}_\mathcal{C}^2 e^{z \mathcal{L}_c} u(x, (n-k)\tau) \, dz.$$

For the remainder term, by Lemma 4.5, it holds for every partition of batches that

$$\|\mathcal{L}_\mathcal{C}^2 e^{z \mathcal{L}_c} u(x, (n-k)\tau)\|_\infty = \|e^{z \mathcal{L}_c} \mathcal{L}_\mathcal{C}^2 u(x, (n-k)\tau)\|_\infty$$
$$\leq \|\mathcal{L}_\mathcal{C}^2 u(x, (n-k)\tau)\|_\infty \leq C.$$

Hence,

$$\|\mathcal{S} u(x, (n-k)\tau) - u(x, (n-k)\tau) - \mathbb{E}\tau \mathcal{L}_\mathcal{C} u(x, (n-k)\tau)\|_\infty \leq C\tau^2,$$

with C independent of N. Here, we used the fact that $e^{\tau \mathcal{L}_c}$ is nonexpansive in L^∞, similar as in (4.8).

Similarly,

$$u(x, (n-k+1)\tau) = e^{\tau \mathcal{L}} u(x, (n-k)\tau) = u(x, (n-k)\tau)$$
$$+ \tau \mathcal{L} u(x, (n-k)\tau) + \int_0^\tau (\tau - z) \mathcal{L}^2 e^{z \mathcal{L}} u(x, (n-k)\tau) \, dz.$$

The remainder term is again bounded by $C\tau^2$. Since

$$\mathbb{E}\mathcal{L}_\mathcal{C} = \mathcal{L},$$

one finds that

$$\|\mathcal{S} u(x, (n-k)\tau) - u(x, (n-k+1)\tau)\|_\infty \leq C\tau^2,$$

with C independent of N.

Hence,

$$|\mathcal{S}^n \phi(x) - u(x, n\tau)| \leq \sum_{k=1}^n \|\mathcal{S} u(x, (n-k)\tau) - u(x, (n-k+1)\tau)\|_\infty \leq Cn\tau^2 \leq C(T)\tau,$$

because $n\tau \leq T$. The claim is then proved. $\quad\square$

Remark 4.7. This result is reminiscent of the results by Golse et al. [21], where the average of one marginal density matrices for N body quantum system has been used for the convergence of RBM, and the convergence rate is also $O(\tau)$ under a certain weak norm.

The backward equation for the random batch system is given by

$$(4.18) \quad \partial_t \tilde{u} = \sum_{i=1}^N \left[b(x_i) + \frac{1}{p-1} \sum_{j:j \neq i} I_{ij}^{(k-1)} m_j K_{ij}(x_i, x_j) \right] \cdot \nabla_{x_i} \tilde{u} + \sum_{i=1}^N \frac{1}{2} \sigma^2 \Delta_{x_i} \tilde{u}.$$

Hence, Theorem 4.3 in fact says the following result when we apply RBM to backward equations or Liouville equations ($\sigma = 0$).

THEOREM 4.8. *Consider the backward equation* (4.5) *and its corresponding equation using RBM* (4.18). *If the initial value is given by*

$$u|_{t=0} = \tilde{u}|_{t=0} = \frac{1}{N} \sum_{i=1}^{N} \omega_i \varphi(x_i), \tag{4.19}$$

then it holds that

$$\sup_{t \le T} \|u(t) - \tilde{u}(t)\|_\infty \le C(T)\tau. \tag{4.20}$$

Remark 4.9. For general initial data, the approximation in L^∞ given by RBM for backward equation (Liouville equation when $\sigma = 0$) cannot be uniform in N.

Remark 4.10. The situation can be different if one considers the Liouville equation of second order system for the density distribution [26]. Clearly, in this case, we cannot use the L^∞ norm to gauge the difference. Instead, a certain weak norm for the combination of one marginals should be considered as in [21]. This is left for the future.

Appendix A. Proof of Lemma 3.4.

Proof. The first part is similar as in [25] except that for $b(\cdot)$ terms, we need to use the one-sided Lipschitz condition. Hence, we omit the proof.

We now prove (3.12). Consider a realization so that the equation is written as

$$d\tilde{X}^i(t) = b(\tilde{X}^i) \, dt + \frac{1}{p-1} \sum_{j \in \mathcal{C}_\theta, j \ne i} m_j K_{ij}(\tilde{X}^i, \tilde{X}^j) + \sigma dB^i,$$

where $\mathcal{C}_\theta$ again is the random batch that contains i from the random division at t_{k-1}, i.e., $\mathcal{C}^{(k-1)}$. It follows that

$$\mathbb{E}\left(\tilde{X}^i(t) - \tilde{X}^i(t_{k-1}) \Big| \mathcal{F}_{k-1}\right) = \int_{t_{k-1}}^{t} \mathbb{E}(b(\tilde{X}^i)|\mathcal{F}_{k-1}) \, ds$$
$$+ \int_{t_{k-1}}^{t} \mathbb{E}\left(\frac{1}{p-1} \sum_{j \in \mathcal{C}_\theta, j \ne i} m_j K_{ij}(\tilde{X}^i, \tilde{X}^j) \Big| \mathcal{F}_{k-1}\right) ds.$$

Note that K is bounded and $|b(x)| \le C(1 + |x|^q)$ for some $q > 0$. Together with (3.11), this implies the first estimate in (3.12).

For the second equation in (3.12), Itô's formula implies that

$$\frac{d}{dt} \mathbb{E}\left[|\tilde{X}^i(t) - \tilde{X}^i(t_{k-1})|^2 | \mathcal{F}_{k-1}\right]$$
$$= 2\mathbb{E}\left[\left(\tilde{X}^i(t) - \tilde{X}^i(t_{k-1})\right) \cdot \left(b(\tilde{X}^i) + \frac{1}{p-1} \sum_{j \in \mathcal{C}_\theta, j \ne i} m_j K_{ij}(\tilde{X}^i, \tilde{X}^j)\right) \Big| \mathcal{F}_{k-1}\right] + \sigma^2 d.$$

Note that

$$\left(\tilde{X}^i(t) - \tilde{X}^i(t_{k-1})\right) \cdot b(\tilde{X}^i) \le \beta |\tilde{X}^i(t) - \tilde{X}^i(t_{k-1})|^2 + \left(\tilde{X}^i(t) - \tilde{X}^i(t_{k-1})\right) \cdot b(\tilde{X}^i(t_{k-1}))$$
$$\le C|\tilde{X}^i(t) - \tilde{X}^i(t_{k-1})|^2 + C(1 + |\tilde{X}^i(t_{k-1})|^q).$$

Hence,

$$\frac{d}{dt}\mathbb{E}\left[|\tilde{X}^i(t)-\tilde{X}^i(t_{k-1})|^2|\mathcal{F}_{k-1}\right] \leq C\mathbb{E}\left[|\tilde{X}^i(t)-\tilde{X}^i(t_{k-1})|^2|\mathcal{F}_{k-1}\right]+C(1+|\tilde{X}^i(t_{k-1})|^q).$$

The claim then follows. □

Appendix B. Proof of Proposition 4.4.
Step 1—Estimates of $\nabla_{x_i}u$.
Taking ∇_{x_ℓ} in (4.5), one has

(B.1) $$\partial_t\nabla_{x_\ell}u = \mathcal{L}\nabla_{x_\ell}u + f_\ell(x)$$

with

(B.2) $$f_\ell(x,t) = \left[\nabla_{x_\ell}b(x_\ell) + \frac{1}{N-1}\sum_{j=1}^N m_j\nabla_{x_\ell}K_{\ell j}(x_\ell,x_j)\right] \cdot \nabla_{x_\ell}u$$
$$+ \frac{1}{N-1}\sum_{j=1}^N m_\ell\nabla_{x_\ell}K_{j\ell}(x_j,x_\ell) \cdot \nabla_{x_j}u.$$

Here $\nabla_{x_\ell}K_{\ell j}$ is a second order tensor, and we use the convention that $(A \cdot v)_i := \sum_j A_{ij}v_j$.
Hence,

$$\nabla_{x_\ell}u(t) = e^{t\mathcal{L}}\nabla_{x_\ell}u(0) + \int_0^t e^{(t-s)\mathcal{L}}f_\ell(x,s)\,ds.$$

Since the semigroup $e^{t\mathcal{L}}$ is nonexpansive in L^∞ as in (4.8), we find

$$\|\nabla_{x_\ell}u(t)\|_\infty \leq \|\nabla_{x_\ell}u(0)\|_\infty + \int_0^t \|f_\ell(x,s)\|_\infty\,ds.$$

However, by (B.2), the linear transform from $[\nabla_{x_1}u,\ldots,\nabla_{x_N}u]$ to $[f_1,\ldots,f_N]$ is a block matrix with the L^∞ norm bounded. This is because all the off-diagonal blocks are of order $O(\frac{1}{N})$. Hence,

$$\|f_\ell(x,s)\|_\infty \leq C\max_\ell\|\nabla_{x_\ell}u(t)\|_\infty =: a(t).$$

Therefore,

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

This means

$$\sup_{t\leq T}a(t) \leq C(T)a(0).$$

Clearly since

$$u(0) = \frac{1}{N}\sum_i\omega_i\varphi(x_i) \Rightarrow \|\nabla_{x_i}u(0)\|_\infty \leq C\frac{1}{N},$$

the estimate for $\nabla_{x_i}u$ follows by Grönwall's lemma.
Step 2—Second order derivatives.
We first compute the $\nabla^2_{x_i}$ derivatives (the x_i Hessian).

(B.3) $$\partial_t\nabla^2_{x_\ell}u = \mathcal{L}\nabla^2_{x_\ell}u + \tilde{f}_\ell,$$

with

$$\tilde{f}_\ell = 2\nabla_{x_\ell} b \cdot \nabla^2_{x_\ell} u + \frac{1}{N-1} \sum_j 2m_j \nabla_{x_\ell} K_{\ell j}(x_\ell, x_j) \cdot \nabla^2_{x_\ell} u$$

$$+ \frac{1}{N-1} \sum_{i:i\neq\ell} 2m_\ell \nabla_{x_\ell} K_{i\ell}(x_i, x_\ell) \cdot \nabla_{x_i} \nabla_{x_\ell} u$$

(B.4)

$$+ \left(2\nabla_{x_\ell} b + \frac{1}{N-1} \sum_j m_j \nabla^2_{x_\ell} K_{\ell j}(x_\ell, x_j) \right) \cdot \nabla_{x_\ell} u$$

$$+ \frac{1}{N-1} \sum_i m_\ell \nabla^2_{x_\ell} K_{i\ell}(x_i, x_\ell) \cdot \nabla_{x_i} u.$$

Again, $\nabla_{x_i} b$ for a vector field b is a second order tensor with $(\nabla_{x_i} b)_{rs} = \partial_{x_i^{(r)}} b^{(s)}$. For the dot product between tensors: $A \cdot B$ means the contraction between the last index of A and the first index of B. For example, we use the convention that

$$(\nabla_{x_\ell} K_{i\ell}(x_i, x_\ell) \cdot \nabla_{x_i} \nabla_{x_\ell} u)_{rs} := \sum_{q=1}^d \partial_{x_\ell^{(r)}} K_{i\ell}^{(q)} \partial_{x_i^{(q)}} \partial_{x_\ell^{(s)}} u.$$

Thus,

$$\tilde{f}_\ell = \tilde{A}_\ell^1 \cdot \nabla^2_{x_\ell} u + \frac{1}{N-1} \sum_i \tilde{A}_{i\ell}^2 \cdot \nabla_{x_i} \nabla_{x_\ell} u$$

$$+ \tilde{A}_\ell^3 \cdot \nabla_\ell u + \frac{1}{N-1} \sum_i \tilde{A}_{i\ell}^4 \cdot \nabla_{x_i} u.$$

Here, all the $\tilde{A}$ tensors are bounded independent of N. There are $O(N)$ terms in the summation in the first line, and thus the linear transform is again bounded in $L^\infty \to L^\infty$ independent of N. The terms on the second line is clearly controlled as $C\frac{1}{N}$ by the estimates of $\nabla_{x_i} u$.

Similarly, one can compute for $\ell \neq q$ that

(B.5) $$\partial_t \nabla_{x_\ell} \nabla_{x_q} u = \mathcal{L} \nabla_{x_\ell} \nabla_{x_q} u + g_{\ell q},$$

where the expression of $g_{\ell q}$ is complicated but it is of the following form

(B.6) $$g_{\ell q} = A_{\ell m}^1 \cdot \nabla_{x_\ell} \nabla_{x_q} u + \frac{1}{N-1} \sum_j (A_{j\ell}^2 \cdot \nabla_{x_j} \nabla_{x_\ell} u + A_{jq}^3 \cdot \nabla_{x_j} \nabla_{x_q} u)$$

$$+ \frac{1}{N-1} (m_\ell \nabla_{x_\ell} \nabla_{x_q} K_{\ell q}(x_\ell, x_q) \cdot \nabla_\ell u + m_q \nabla_{x_q} \nabla_{x_\ell} K_{q\ell}(x_q, x_\ell) \cdot \nabla_{x_q} u).$$

Here, A^k's are second order tensors that are made up of the derivatives of b, K_{ij}, and thus bounded by constants independent of N. Note that there are $O(N)$ terms in the summation in the first line, and therefore the linear transform is again bounded in $L^\infty \to L^\infty$ independent of N. The terms on the second line is clearly controlled as $C\frac{1}{N^2}$ by the estimates of $\nabla_{x_i} u$.

We first of all consider all the second order derivatives $\nabla_{x_\ell} \nabla_{x_q}$ where ℓ can be equal or not equal to q. By similar argument as for $\nabla_{x_i} u$, one can get the estimate

(B.7)

$$\|\nabla_{x_\ell} \nabla_{x_q} u(t)\|_\infty \leq \|\nabla_{x_\ell} \nabla_{x_q} u(0)\|_\infty + C \int_0^t \max_{\ell,q} \|\nabla_{x_\ell} \nabla_{x_q} u(s)\|_\infty \, ds + C(T)\frac{1}{N}.$$

Moreover,

$$\|\nabla_{x_\ell}^2 u(0)\|_\infty \le C\frac{1}{N}, \quad \|\nabla_{x_\ell}\nabla_{x_q} u(0)\|_\infty = 0.$$

Then, Grönwall's inequality tells us that

$$(\text{B.8}) \qquad \max_{\ell,q} \|\nabla_{x_\ell}\nabla_{x_q} u(t)\|_\infty \le C(T)\frac{1}{N}.$$

Now, we focus on (B.5) for $\ell \ne q$. To do this, we need to separate out the $j = \ell$ and $j = q$ terms in (B.6), which are of order $O(\frac{1}{N})$, and thus with the prefactor $1/(N-1)$, they contribute $O(1/N^2)$ to $g_{\ell q}$. Hence, one has for $\ell \ne q$ that

(B.9)

$$\|\nabla_{x_\ell}\nabla_{x_q} u(t)\|_\infty \le \|\nabla_{x_\ell}\nabla_{x_q} u(0)\|_\infty + C\int_0^t \max_{\ell \ne q}\|\nabla_{x_\ell}\nabla_{x_q} u(s)\|_\infty\,ds + C(T)\frac{1}{N^2}.$$

Since $\|\nabla_{x_\ell}\nabla_{x_q} u(0)\|_\infty = 0$ for $\ell \ne q$, one then has

$$(\text{B.10}) \qquad \max_{\ell \ne q}\|\nabla_{x_\ell}\nabla_{x_q} u(t)\|_\infty \le C\int_0^t \max_{\ell \ne q}\|\nabla_{x_\ell}\nabla_{x_q} u(s)\|_\infty\,ds + C(T)\frac{1}{N^2}.$$

Grönwall's lemma then tells us that

$$(\text{B.11}) \qquad \max_{\ell \ne q}\|\nabla_{x_\ell}\nabla_{x_q} u(t)\|_\infty \le C\frac{1}{N^2}.$$

Step 3—Higher order derivatives.

The higher order derivatives can be similarly estimated using induction. For these proofs, some derivatives that are not listed in Proposition 4.4 should be involved. For example, for third order derivatives, one should expect

$$\|\nabla_{x_i}\nabla_{x_j}\nabla_{x_k} u\|_\infty = O(1/N^3)$$

for distinct i, j, k. These proofs are tedious but the essential ideas are the same as we prove the claims for the Hessian. We omit the details.

REFERENCES

[1] G. ALBI AND L. PARESCHI, *Binary interaction algorithms for the simulation of flocking and swarming dynamics*, Multiscale Model. Sim., 11 (2013), pp. 1–29.

[2] G. ALLAIRE, J.-F. DUFRÊCHE, A. MIKELIĆ, AND A. PIATNITSKI, *Asymptotic analysis of the Poisson–Boltzmann equation describing electrokinetics in porous media*, Nonlinearity, 26 (2013), p. 881.

[3] D. F. ANDERSON AND J. C. MATTINGLY, *A weak trapezoidal method for a class of stochastic differential equations*, Commun. Math. Sci., 9 (2011), pp. 301–318.

[4] H. BABOVSKY AND R. ILLNER, *A convergence proof for Nanbu's simulation method for the full Boltzmann equation*, SIAM J. Numer. Anal., 26 (1989), pp. 45–65.

[5] A. L. BERTOZZI, J. B. GARNETT, AND T. LAURENT, *Characterization of radially symmetric finite time blowup in multidimensional aggregation equations*, SIAM J. Math. Anal., 44 (2012), pp. 651–681.

[6] G. A. BIRD, *Approach to translational equilibrium in a rigid sphere gas*, Phys. Fluids, 6 (1963), pp. 1518–1519.

[7] C. K. BIRDSALL AND A. B. LANGDON, *Plasma Physics Via Computer Simulation*, CRC Press, 2004.

[8] L. BOTTOU, *Online learning and stochastic approximations*, On-line Learning in Neural Networks, 17 (1998), p. 142.

[9] S. BUBECK, *Convex optimization: Algorithms and complexity*, Found. Trends. Mach. Learn., 8 (2015), pp. 231–357.

[10] E. CARLEN, P. DEGOND, AND B. WENNBERG, *Kinetic limits for pair-interaction driven master equations and biological swarm models*, Math. Models Methods Appl. Sci., 23 (2013), pp. 1339–1376.

[11] J. A. CARRILLO, L. PARESCHI, AND M. ZANELLA, *Particle based gPC methods for mean-field models of swarming with uncertainty*, Commun. Comput. Phys., 25 (2019).

[12] A. J. CHORIN, *Numerical study of slightly viscous flow*, J. Fluid Mech., 57 (1973), pp. 785–796.

[13] K. CRAIG AND A. L. BERTOZZI, *A blob method for the aggregation equation*, Math. Comp., 85 (2016), pp. 1681–1717.

[14] F. CUCKER AND S. SMALE, *Emergent behavior in flocks*, IEEE Trans. Automat. Control, 52 (2007), pp. 852–862.

[15] P. DEGOND, J.-G. LIU, AND R. L. PEGO, *Coagulation–fragmentation model for animal group-size statistics*, J. Nonlinear Sci., 27 (2017), pp. 379–424.

[16] R. DURRETT, *Probability: Theory and Examples*, Cambridge Series in Statistical and Probabilistic Mathematics, 4 ed., Cambridge University Press, 2010, https://doi.org/10.1017/CBO9780511779398.

[17] R. DURSTENFELD, *Algorithm 235: Random permutation*, Commun. ACM, 7 (1964), p. 420.

[18] Y. FENG, L. LI, AND J.-G. LIU, *Semi-groups of stochastic gradient descent and online principal component analysis: Properties and diffusion approximations*, Commun. Math. Sci., 16 (2018).

[19] D. FRENKEL AND B. SMIT, *Understanding Molecular Simulation: From Algorithms to Applications*, vol. 1, Elsevier, 2001.

[20] A. GEORGES, G. KOTLIAR, W. KRAUTH, AND M. J. ROZENBERG, *Dynamical mean-field theory of strongly correlated fermion systems and the limit of infinite dimensions*, Rev. Modern Phys., 68 (1996), p. 13.

[21] F. GOLSE, S. JIN, AND T. PAUL, *The random batch method for n-body quantum dynamics*, Found. Comput. Math., to appear.

[22] K. E. GUSTAFSON AND J. A. SETHIAN, *Vortex Methods and Vortex Motion*, vol. 19, SIAM, 1991.

[23] S.-Y. HA AND J.-G. LIU, *A simple proof of the Cucker-Smale flocking dynamics and mean-field limit*, Commun. Math. Sci., 7 (2009), pp. 297–325.

[24] D. HORSTMANN, *From 1970 until present: the Keller-Segel model in chemotaxis and its consequences*, Jahresber. Dtsch. Math.-Ver., 105 (2003), pp. 103–165.

[25] S. JIN, L. LI, AND J.-G. LIU, *Random Batch methods (RBM) for interacting particle systems*, J. Comput. Phys., 400 (2020), 108877.

[26] S. JIN AND X. WEN, *Hamiltonian-preserving schemes for the Liouville equation with discontinuous potentials*, Commun. Math. Sci., 3 (2005), pp. 285–315.

[27] P. E. KLOEDEN AND E. PLATEN, *Numerical Solution of Stochastic Differential Equations*, vol. 23, Springer Science & Business Media, 2013.

[28] J.-M. LASRY AND P.-L. LIONS, *Mean field games*, Jpn. J. Math., 2 (2007), pp. 229–260.

[29] L. LI, Y. LI, J.-G. LIU, Z. LIU, AND J. LU, *A stochastic version of Stein variational gradient descent for efficient sampling*, Comm. App. Math. Comp. Sci., 15 (2020), pp. 37–63.

[30] L. LI AND J.-G. LIU, *Large time behaviors of upwind schemes and B-schemes for Fokker-Planck equations on $\mathbb{R}$ by jump processes*, Math. Comp., 89 (2020), pp. 2283–2320.

[31] L. LI, J.-G. LIU, AND Y. TANG, *A direct simulation approach for the Poisson-Boltzmann equation using the Random Batch Method*, preprint, 2020.

[32] L. LI, J. LU, J. C. MATTINGLY, AND L. WANG, *Numerical methods for stochastic differential equations based on Gaussian mixture*, Commun. Math. Sci., to appear.

[33] J.-G. LIU AND Z. XIN, *Convergence of the point vortex method for 2-D vortex sheet*, Math. Comp., 70 (2001), pp. 595–606.

[34] J.-G. LIU AND R. YANG, *A random particle blob method for the Keller-Segel equation and convergence analysis*, Math. Comp., 86 (2017), pp. 725–745.

[35] Q. LIU AND D. WANG, *Stein variational gradient descent: A general purpose bayesian inference algorithm*, in Advances in Neural Information Processing Systems, Barcelona, Spain, 2016, pp. 2378–2386.

[36] S. MEYN AND R. L. TWEEDIE, *Markov Chains and Stochastic Stability*, 2nd ed., Cambridge University Press, Cambridge, 2009.

[37] G. MILSTEIN, *Weak approximation of solutions of systems of stochastic differential equations*, Theory Probab. Appl., 30 (1986), pp. 750–766.

[38] G. N. MILSTEIN AND M. V. TRETYAKOV, *Stochastic numerics for mathematical physics*, Springer Science & Business Media, 2013.

[39] S. Motsch and E. Tadmor, *Heterophilious dynamics enhances consensus*, SIAM Rev., 56 (2014), pp. 577–621.

[40] K. Nanbu, *Direct simulation scheme derived from the Boltzmann equation. i. monocomponent gases*, J. Phys. Soc. Japan, 49 (1980), pp. 2042–2049.

[41] B. Øksendal, *Stochastic Differential Equations: An Introduction with Applications*, 6th ed., Springer, Berlin, Heidelberg, 2003.

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

[43] V. Rokhlin, *Rapid solution of integral equations of classical potential theory*, J. Comput. Phys., 60 (1985), pp. 187–207.

[44] H. E. Stanley, *Phase transitions and critical phenomena*, Clarendon Press, Oxford, 1971.

[45] T. Vicsek, A. Czirók, E. Ben-Jacob, I. Cohen, and O. Shochet, *Novel type of phase transition in a system of self-driven particles*, Phys. Rev. Lett., 75 (1995), p. 1226.

[46] M. Welling and Y. W. Teh, *Bayesian learning via stochastic gradient Langevin dynamics*, in Proceedings of the 28th International Conference on Machine Learning (ICML-11), 2011, pp. 681–688.