On the mean field limit of Random Batch Method for interacting particle systems

Shi Jin *1 and Lei Li†2

1,2School of Mathematical Sciences, Institute of Natural Sciences, MOE-LSC, Shanghai Jiao Tong University, Shanghai, 200240, P. R. China.

Abstract

The Random Batch Method proposed in our previous work [Jin et al., J. Comput. Phys., 400(1), 2020] is not only a numerical method for interacting particle systems and its mean-field limit, but also can be viewed as a model of particle system in which particles interact, at discrete time, with randomly selected mini-batch of particles. In this paper we investigate the mean-field limit of this model as the number of particles $N \to \infty$. Unlike the propagation of chaos in the classical mean field limit for interacting particle systems, the chaos here is imposed at every discrete time. Despite this, we will not only justify this mean-field limit (discrete in time) but will also show that the limit, as the discrete time interval $\tau \to 0$, approaches to the solution of a nonlinear Fokker-Planck equation arising as the mean-field limit of the original interacting particle system, in Wasserstein-2 distance, namely, the two limits $\lim_{N\to\infty}$ and $\lim_{\tau\to 0}$ commute.

1 Introduction

Many physical, biological and social sciences phenomena, at the microscopic level, are described by interacting particle systems, for examples molecules in fluids [19], plasma [5], swarming [48, 9, 7, 13], chemotaxis [26, 4], flocking [12, 25, 1], synchronization [11, 24] and consensus [43]. We consider the following general first order systems

$$dX^i = b(X^i)\,dt + \frac{1}{N-1}\sum_{j:j\neq i} K(X^i - X^j)\,dt + \sqrt{2}\sigma\,dW^i,\ i = 1, 2, \cdots, N, \tag{1.1}$$

with the initial data X_0^i's i.i.d sampled from a common distribution μ_0. Here, we allow $\sigma = 0$ to include systems without noise.

As well-known, the mean field limit (i.e. $N \to \infty$) of (1.1) is given by

$$\partial_t \mu = -\nabla\cdot((b(x) + K * \mu)\mu) + \sigma^2\Delta\mu. \tag{1.2}$$

This implies that the empirical measure $\mu_N := \frac{1}{N}\sum_{i=1}^N \delta(x - X^i)$ converges weakly to μ almost surely and the one marginal distribution $\mu_N^{(1)} := \mathscr{L}(X^1)$, the law of X^1, converges to μ. See [10, 18, 15, 37] for some related models and proofs, though the setups in these works do not quite fit our problem as we allow $|b(\cdot)|$ to have polynomial growth. Recall that μ is in general a probability distribution and (1.2) is understood in the distributional sense. We will denote the solution operator to (1.2) by $\mathcal{S}$:

$$\mathcal{S}(\Delta)\mu(t_1) := \mu(t_1 + \Delta),\ \forall t_1 \geq 0, \Delta \geq 0. \tag{1.3}$$

*shijin-m@sjtu.edu.cn
†leili2010@sjtu.edu.cn

Clearly, $\{\mathcal{S}(t) : t \geq 0\}$ is a nonlinear semigroup.

Direct simulation of (1.1) costs $O(N^2)$ per time step, which is expensive. To reduce the computational cost, in [29], a random algorithm that uses random mini-batches, called the Random Batch Method (RBM), has been proposed to reduce the computation cost per time step from $O(N^2)$ to $O(N)$. The method has been applied to various problems with promising results [29, 35, 32, 36]. However, the understanding of the method is still limited, despite some theoretic proofs [29, 28]. The idea of using the "mini-batch" was inspired by the stochastic gradient descent (SGD) method [44, 6] in machine learning. The "mini-batch" was also used for Bayesian inference [50], and similar idea was used to simulate the mean-field equations for flocking [1]. How to apply the mini-batch depends on the specific problems. The strategy in [29] for interacting particle systems (1.1) is to do random grouping. Intuitively, the method converges due to certain time average in time, and thus the convergence is like the convergence in the Law of Large Number (in time). See [29] for more details. Compared with the Fast Multipole Method, the accuracy is lower (half order in time step), but RBM is simpler to implement and is valid for more general potentials ([35, 28]).

The RBM algorithm corresponding to (1.1) is shown in Algorithm 1. Suppose we aim to do simulation until time $T > 0$. We first choose a time step $\tau > 0$ and a batch size $p \ll N, p \geq 2$ that divides N. Define the discrete time grids $t_k := k\tau$, $k \in \mathbb{N}$. For each time subinterval $[t_{k-1}, t_k)$, there are two steps: (1) at time grid t_{k-1}, we divide the N particles into $n := N/p$ groups (batches) randomly; (2) the particles evolve with interaction inside the batches only. Here, we use the same symbols X^i without causing any confusion. The Wiener process W^i (Brownian motion) used in (1.4) is the same as in (1.1).

Algorithm 1 (RBM)

for k in $1 : [T/\tau]$ **do**
  Divide $\{1, 2, \dots, N\}$ into $n = N/p$ batches randomly.
  for each batch $\mathcal{C}_q$ **do**
    Update X^i's ($i \in \mathcal{C}_q$) by solving the following SDE with $t \in [t_{k-1}, t_k)$.

$$dX^i = b(X^i)dt + \frac{1}{p-1} \sum_{j \in \mathcal{C}_q, j \neq i} K(X^i - X^j)dt + \sqrt{2}\sigma\, dW^i. \tag{1.4}$$

  end for
end for

As pointed out in [29], RBM is asymptotic-preserving regarding the mean field limit $N \to \infty$ ([46, 21, 34]); namely, the error bound of the one marginal distribution can be made independent of N so that it can be used for large N as an efficient numerical particle method for the mean field nonlinear Fokker-Planck equation of (1.1). While RBM was introduced as a numerical method, it can also be viewed as a new model for the underlying particle system. A natural question, for both numerical and modeling interests, is: what is the limiting (mean field) dynamics as $N \to \infty$ for a fixed time step τ?

Intuitively, when $N \gg 1$, the probability that two chosen particles are correlated is very small. Hence, in the $N \to \infty$ limit, two chosen particles will be independent with probability 1, and the marginal distribution will be the mean field limit of RBM. Since the particles are exchangeable, the marginal distributions of them will be identical. Due to these observations, one can guess the following mean field limit for RBM as shown in Algorithm 2.

Algorithm 2 (Mean Field Dynamics of RBM (1.4))

1: $\tilde{\mu}(\cdot, t_0) = \mu_0$.
2: **for** $k \geq 0$ **do**
3: Let $\rho^{(p)}(\cdots, 0) = \tilde{\mu}(\cdot, t_k)^{\otimes p}$ be a probability measure on $(\mathbb{R}^d)^{\otimes p} \cong \mathbb{R}^{pd}$.
4: Evolve the measure $\rho^{(p)}$ to find $\rho^{(p)}(\cdots, \tau)$ by the following Fokker-Planck equation:

$$\partial_t \rho^{(p)} = -\sum_{i=1}^{p} \nabla_{x_i} \cdot \left(\Big[b(x_i) + \frac{1}{p-1} \sum_{j=1, j\neq i}^{p} K(x_i - x_j) \Big] \rho^{(p)} \right) + \sigma^2 \sum_{i=1}^{p} \Delta_{x_i} \rho^{(p)}. \quad (1.5)$$

5: Set

$$\tilde{\mu}(\cdot, t_{k+1}) := \int_{(\mathbb{R}^d)^{\otimes(p-1)}} \rho^{(p)}(\cdot, dy_2, \cdots, dy_p, \tau). \quad (1.6)$$

6: **end for**

The dynamics shown in Algorithm 2 naturally defines a nonlinear operator $\mathcal{G}_\infty : \mathbf{P}(\mathbb{R}^d) \to \mathbf{P}(\mathbb{R}^d)$ as

$$\tilde{\mu}(\cdot, t_{k+1}) =: \mathcal{G}_\infty(\tilde{\mu}(\cdot, t_k)). \quad (1.7)$$

This corresponds to the following SDE system for $t \in [t_k, t_{k+1})$

$$dY^i = b(Y^i)\, dt + \frac{1}{p-1} \sum_{j=1, j\neq i}^{p} K(Y^i - Y^j)\, dt + \sqrt{2}\sigma\, dW^i,\ i = 1, \cdots, p, \quad (1.8)$$

with $\{Y^i(t_k)\}$ drawn i.i.d from $\tilde{\mu}(\cdot, t_k)$. Then, $\tilde{\mu}(\cdot, t_{k+1}) = \mathscr{L}(Y^1(t_{k+1}^-))$, the law of $Y^1(t_{k+1}^-)$. Note that Y^i's all have the same distribution for any $t_k \leq t < t_{k+1}$. Without loss of generality, we will impose $Y^1(t_k^-) = Y^1(t_k^+)$. For other particles $i \neq 1$, $Y^i(t)$ in $[t_{k-1}, t_k)$ and $[t_k, t_{k+1})$ are independent and they are not continuous at t_k. In fact, in the $N \to \infty$ limit, $Y^i, i \neq 1$ correspond to different particles that interact with particle 1 as in Algorithm 1.

Hence, in the mean field limit, Random Batch Method is doing this: one starts with a chaotic configuration [1], the p particles evolve with interaction to each other. Then, at the starting point of the next time interval, one imposes the chaos so that the particles are independent again. This mean field limit is different from the standard mean field limit for system (1.1), given by (1.2): in the mean field limit of RBM, the chaos is imposed at every time step; in the classical mean field limit for interacting particle system, the chaos is propagated to later times. This mechanism allows the mean-field limit of RBM to achieve a higher convergence rate than the standard $N^{-1/2}$ convergence rate (at least N^{-1} as seen in section 3). In spite of the difference just mentioned, we will show that these two limiting dynamics are in fact close, In section 4, we will show that as $\tau \to 0$ the dynamics given by $\mathcal{G}_\infty$ can approximate that of the nonlinear Fokker-Planck equation (1.2). Thus, the two limits $\lim_{N\to\infty}$ and $\lim_{\tau\to 0}$ commute.

The argument in this paper for $t \leq T$ can be generalized to second order systems, which we omit, but one may see section 5 for some discussion. Of course, the argument for large time behavior can be different and this is left for future study.

The rest of the paper is organized as follows. We introduce the notations and give a brief review to Wasserstein distances in section 2. The mean field limit under Wasserstein-2 distance is shown in section 3. Section 4 is devoted to the discussion of the mean field dynamics of RBM. In particular, we show that it is close to the mean-field nonlinear Fokker-Planck equation. Some discussion is performed in section 5.

[1]By "chaotic configuration", we mean that there exists a one particle distribution f such that for any j, the j-marginal distribution is given by $\mu^{(j)} = f^{\otimes j}$. Such independence in a configuration is then loosely called "chaos". If the j-marginal distribution is more close to $f^{\otimes j}$ for some f, we loosely say "there is more chaos".

2 Preliminary and notations

In this section, we first introduce some assumptions and notations; then give a brief introduction to Wasserstein distances and prove some auxiliary results.

2.1 Mathematical setup of the problem

We first introduce several assumptions that will be used throughout the paper. In these assumptions, "being smooth" means that the functions are infinitely differentiable. Note that the conditions in these assumptions may be stronger than necessary.

Assumption 2.1. *The moments of the initial data are finite:*

$$\int_{\mathbb{R}^d} |x|^q \mu_0(dx) < \infty, \ \forall q \in [2, \infty). \tag{2.1}$$

One of the following two conditions will be used for the external fields and interaction kernels.

Assumption 2.2. *Assume $b(\cdot) : \mathbb{R}^d \to \mathbb{R}^d$ and $K(\cdot) : \mathbb{R}^d \to \mathbb{R}^d$ are smooth. Moreover, $b(\cdot)$ is one-sided Lipschitz:*

$$(z_1 - z_2) \cdot (b(z_1) - b(z_2)) \le \beta |z_1 - z_2|^2 \tag{2.2}$$

for some constant β, and K is Lipschitz continuous

$$|K(z_1) - K(z_2)| \le L|z_1 - z_2|.$$

Assumption 2.3. *The fields $b(\cdot) : \mathbb{R}^d \to \mathbb{R}^d$ and $K(\cdot) : \mathbb{R}^d \to \mathbb{R}^d$ are smooth. Moreover, $b(\cdot)$ is strongly confining:*

$$(z_1 - z_2) \cdot (b(z_1) - b(z_2)) \le -r|z_1 - z_2|^2 \tag{2.3}$$

for some constant $r > 0$, and K is Lipschitz continuous $|K(z_1) - K(z_2)| \le L|z_1 - z_2|$. The parameters r, L satisfy

$$r > 2L. \tag{2.4}$$

Remark 2.1. *Compared with our previous works, we are not assuming the boundedness of K in this paper to prove the mean-field limit and investigation of the limiting dynamics. The boundedness of K in our previous works is a simple condition to guarantee the boundedness of the variance of the random forces (though the boundedness of variance may also be proved without assuming boundedness of K).*

Denote $\mathcal{C}_q^{(k)}$ $(1 \le q \le n)$ the batches at t_k so that $\cup_q \mathcal{C}_q^{(k)} = \{1, \cdots, N\}$, and

$$\mathcal{C}^{(k)} := \{\mathcal{C}_1^{(k)}, \cdots, \mathcal{C}_n^{(k)}\} \tag{2.5}$$

will denote the random division of batches at t_k. By the Kolmogorov extension theorem [16], there exists a probability space $(\Omega, \mathcal{F}, \mathbb{P})$ such that the random variables $\{X_0^i, W^i, \mathcal{C}^{(k)} : 1 \le i \le N, k \ge 0\}$ are defined on this probability space and are all independent. We will 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:

$$\|v\| = \sqrt{\mathbb{E}|v|^2}. \tag{2.6}$$

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

$$\mathcal{F}_{k-1} = \sigma(X_0^i, W^i(t), \mathcal{C}^{(j)}; t \le t_{k-1}, j \le k-1). \tag{2.7}$$

Clearly, $\mathcal{F}_{k-1}$ is the σ-algebra generated by the initial values X_0^i $(i = 1, \dots, N)$, $B^i(t)$, $t \le t_{k-1}$, and $\mathcal{C}^{(j)}$, $j \le k-1$. Hence, $\mathcal{F}_{k-1}$ contains the information of how batches are constructed for $t \in [t_{k-1}, t_k)$.

2.2 A review of the Wasserstein distances

Consider a domain $O \subset \mathbb{R}^n$ where n is a positive integer. We denote $\mathbf{P}(O)$ the set of probability measures on O. Let $\mu, \nu \in \mathbf{P}(O)$ be two probability measures and $c : O \times O \to [0, \infty)$ be a cost function. One solves the following optimization problem for the optimal transport:

$$\min_{\gamma} \left\{ \int_{O\times O} c \, d\gamma \Big| \gamma \in \Pi(\mu, \nu) \right\},$$

where $\Pi(\mu, \nu)$ is the set of 'transport plans', i.e. a joint measure on $O \times O$ such that the marginal measures are μ and ν respectively. If there is a map $T : O \to O$ such that $(I \times T)_{\#}\mu$ minimizes the target function, then T is called an optimal transport map. Here, I is the identity map and

$$(I \times T)_{\#}\mu(E) := \mu((I \times T)^{-1}(E)), \ \forall E \subset O \times O, \text{ measurable}. \tag{2.8}$$

Choosing the particular cost function $c(x, y) = |x - y|^2$, one can define the so called Wasserstein-2 distance $W_2(\mu, \nu)$ as

$$W_2(\mu, \nu) = \left(\inf_{\gamma\in\Pi(\mu,\nu)} \int_{O\times O} |x - y|^2 d\gamma \right)^{1/2}. \tag{2.9}$$

It has been shown (see [3], [45, Chap. 5]) that the Wasserstein-2 distance between two probability measures μ and ν is also given by

$$W_2^2(\mu, \nu) = \min \left\{ \int_0^1 \|v\|^2_{L^2(\rho)} dt : \partial_t \rho + \nabla \cdot (\rho v) = 0, \rho|_{t=0} = \mu, \rho|_{t=1} = \nu \right\}, \tag{2.10}$$

where ρ is a (time-parametrized) nonnegative measure and

$$\|v\|^2_{L^2(\rho)} = \int_O |v|^2 \rho(dx).$$

Hence, v can be thought as the particle velocity for the optimal transport, as explained in [45, Chap. 5]. With this explanation, one can then understand $\mathbf{P}(O)$ equipped with W_2 distance as a Riemannian manifold so that the Fokker-Planck equations can be formulated as a class of gradient flows on this manifold (see, for example, [30], [49, Chap. 8]).

Below, we consider the effects of taking marginals for Wasserstein distances.

Definition 2.1. *Consider a Polish space $\mathcal{E}$. We say a probability measure $\bar{\mu} \in \mathbf{P}(\mathcal{E}^{\otimes p})$ is symmetric if for any permutation σ of $\{1, \cdots, p\}$ and any Borel measurable $E \subset \mathcal{E}^{\otimes p}$, one has $\bar{\mu}(E) = \bar{\mu}(\sigma(E))$, where*

$$\sigma(E) := \left\{ (x_{\sigma(1)}, \cdots, x_{\sigma(p)}) : (x_1, \cdots, x_p) \in E \right\}.$$

We have the following simple, but important, fact about Wasserstein-2 distances.

Lemma 2.1. *Let μ, ν be a probability measure on a domain $\mathcal{E} \subset \mathbb{R}^d$ for some positive integer d. Suppose $\bar{\mu}$ and $\bar{\nu}$ are two symmetric probability measures on $\mathcal{E}^{\otimes p}$ such that their one marginal distributions are μ, ν respectively. Then,*

$$W_2(\bar{\mu}, \bar{\nu}) \geq \sqrt{p} W_2(\mu, \nu).$$

If $\bar{\mu} = \mu^{\otimes p}$ and $\bar{\nu} = \nu^{\otimes p}$, the equality then holds.

Proof. By (2.9), for any $\epsilon > 0$, there is a plan $\tilde{\gamma}$ such that

$$\begin{aligned} W_2^2(\bar{\mu}, \bar{\nu}) + \epsilon &\geq \int_{\mathcal{E}^{\otimes p}\times\mathcal{E}^{\otimes p}} \sum_{i=1}^p |x_i - y_i|^2 \tilde{\gamma}(dx_1, \cdots, dx_p, dy_1, \cdots, dy_p) \\ &= \sum_{i=1}^p \int_{\mathcal{E}^{\otimes p}\times\mathcal{E}^{\otimes p}} |x_i - y_i|^2 \tilde{\gamma}(dx_1, \cdots, dx_p, dy_1, \cdots, dy_p) \\ &= \sum_{i=1}^p \int_{\mathcal{E}\times\mathcal{E}} |x_i - y_i|^2 \gamma_i(dx_i, dy_i), \end{aligned} \tag{2.11}$$

where $\gamma_i \in \Pi(\mu, \nu)$ due to the symmetry of $\bar{\mu}$ and $\bar{\nu}$. Hence, $W_2^2(\bar{\mu}, \bar{\nu}) + \epsilon \geq p W_2^2(\mu, \nu)$ and the first claim is proved.

For the claim regarding equality, pick $\gamma \in \Pi(\mu, \nu)$ such that

$$\int_{\mathcal{E}\times\mathcal{E}} |x_1 - y_1|^2 \gamma(dx_1, dy_1) \leq W_2^2(\mu, \nu) + \epsilon^2.$$

Since $\bar{\mu} = \mu^{\otimes p}$ and $\bar{\nu} = \nu^{\otimes p}$, we construct $\tilde{\gamma} \in \Pi(\tilde{\mu}, \tilde{\nu})$ with

$$d\tilde{\gamma} := \tilde{\gamma}(dx_1, \cdots, dx_p, dy_1, \cdots, dy_p) = \otimes_{i=1}^p \gamma(dx_i, dy_i).$$

Then

$$\begin{aligned} W_2^2(\tilde{\mu}, \tilde{\nu}) &\leq \int_{\mathcal{E}^{\otimes p}\times\mathcal{E}^{\otimes p}} (\sum_{i=1}^p |x_i - y_i|^2) d\tilde{\gamma} = \sum_{i=1}^p \int_{\mathcal{E}^{\otimes p}\times\mathcal{E}^{\otimes p}} |x_i - y_i|^2 d\tilde{\gamma} \\ &= p \int_{\mathcal{E}\times\mathcal{E}} |x_1 - y_1|^2 \gamma(dx_1, dy_1) \leq p W_2^2(\mu, \nu) + p\epsilon^2. \end{aligned}$$

Since ϵ is arbitrary, the claim is proved. □

Recall the Jordan decomposition for a signed measure $\mu = \mu^+ - \mu^-$, and $|\mu| := \mu^+ + \mu^-$. The total variation norm of the signed measure is defined by

$$\|\mu\|_{TV} := |\mu|(\mathcal{E}) = \mu^+(\mathcal{E}) + \mu^-(\mathcal{E}). \tag{2.12}$$

Below, we note a useful lemma that relates the total variation distance to the W_2 distance.

Lemma 2.2. *Let $\mu, \nu \in \mathbf{P}(\mathbb{R}^d)$ be two different probability measures. Let $\delta \geq 0$ and $\hat{\mu}$ be a measure such that $|\mu - \nu|(E) \leq \delta\, \hat{\mu}(E)$, for any Borel measurable E. Suppose $M_2 := \inf_{x_0} \int_{\mathbb{R}^d} |x - x_0|^2 \hat{\mu}(dx) < \infty$. Then*

$$W_2(\mu, \nu) \leq \sqrt{2\|\hat{\mu}\|_{TV} M_2 \delta}. \tag{2.13}$$

In particular, choosing $\delta = \|\mu - \nu\|_{TV}$, $\hat{\mu} := \frac{1}{\|\mu-\nu\|_{TV}} |\mu - \nu|$ yields

$$W_2(\mu, \nu) \leq \sqrt{2M_2} \|\mu - \nu\|_{TV}.$$

Proof. We consider $\mu_m := \mu \wedge \nu$, which is defined by

$$\mu_m(E) = \min(\mu(E), \nu(E)), \ \forall E \text{ measurable}.$$

Define two measures $\mu_1 := \mu - \mu_m$ and $\nu_1 := \nu - \nu_m$ whose meaning is evident. Then, one has

$$\|\mu - \nu\|_{TV} = \|\mu_1\|_{TV} + \|\nu_1\|_{TV}, \ \mu_1 + \nu_1 \leq \delta\, \hat{\mu}. \tag{2.14}$$

Construct the joint distribution

$$d\pi := \pi(dx, dy) = \mu_1(dx) \otimes \nu_1(dy) + Q_{\#}\mu_m(dx, dy),$$

with $Q(x) = (x, x)$ and $Q_{\#}$ is the standard pushforward map as in (2.8). Clearly, the marginal distributions of π are μ and ν respectively.

Then, fix $x_0 \in \mathbb{R}^d$.

$$\begin{aligned} \int_{\mathbb{R}^d\times\mathbb{R}^d} |x - y|^2 d\pi &= \int_{\mathbb{R}^d\times\mathbb{R}^d} |x - y|^2 \mu_1(dx) \otimes \nu_1(dy) \\ &\leq 2 \int_{\mathbb{R}^d\times\mathbb{R}^d} (|x - x_0|^2 + |y - x_0|^2) \mu_1(dx) \otimes \nu_1(dy) \\ &= 2 \left[\|\nu_1\|_{TV} \int_{\mathbb{R}^d} |x - x_0|^2 \mu_1(dx) + \int_{\mathbb{R}^d} |y - x_0|^2 \nu_1(dy) \right] \\ &\leq 2 \left(\left(\|\mu_1\|_{TV} + \|\nu_1\|_{TV} \right) \int_{\mathbb{R}^d} |x - x_0|^2 [\mu_1 + \nu_1](dx) \right). \end{aligned}$$

Note that there are some nonnegative cross terms like $\|\mu_1\|_{TV} \int |y - x_0|^2 \nu_1(dy)$ having been added. Since $\|\mu_1\|_{TV} + \|\nu_1\|_{TV} \leq \delta \|\hat{\mu}\|_{TV}$ and $\mu_1 + \nu_1 \leq \delta\, \hat{\mu}$, the claim follows by taking infimum on x_0. □

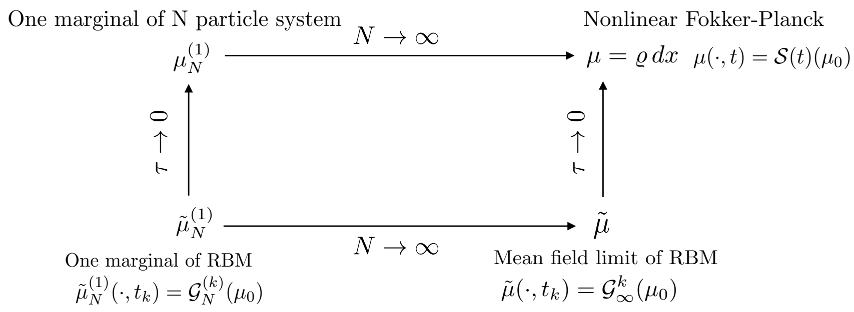

Figure 1: Illustration of the various operators and the asymptotic limits.

3 The mean field limit of RBM with τ fixed

Starting with μ_0, after k steps of the dynamics given in (1.7), one arrives at

$$\mathcal{G}_\infty^k(\mu_0) = \mathcal{G}_\infty \circ \cdots \circ \mathcal{G}_\infty(\mu_0), \ (k \text{ copies}),$$

which is expected to be the mean field limit of RBM after k steps. Corresponding to this, one may define the operator $\mathcal{G}_N^{(k)} : \mathbf{P}(\mathbb{R}^d) \to \mathbf{P}(\mathbb{R}^d)$ for RBM with N particles as follows. Let X_0^i's be i.i.d drawn from $\mu_0^{\otimes N}$, and consider (1.4). Define

$$\mathcal{G}_N^{(k)}(\mu_0) := \mathscr{L}(X^1(t_k)), \tag{3.1}$$

where recall that $\mathscr{L}(X^1)$ means the law of X^1, thus the one marginal distribution. Conditioning on a specific sequence of random batches, the particles are not exchangeable. However, when one considers the mixture of all possible sequences of random batches, the laws of the particles $X^i(t_k)$ $(1 \le i \le N)$ are identical. In Fig. 1, we illustrate these definitions and various limits.

The semigroup property is closely related to the Markovian property. For the $\mathcal{G}_\infty$ dynamics, knowing the marginal distribution of X^1 can fully determine the probability transition. However, knowing only the marginal distribution is not enough for $\mathcal{G}_N^{(k)}$ dynamics, and the joint distribution must be known. Hence, we remark that

Lemma 3.1. *$\{\mathcal{G}_\infty^k : k \ge 1\}$ forms a nonlinear semigroup while $\{\mathcal{G}_N^{(k)} : k \ge 1\}$ is not a semigroup.*

We first of all introduce some concepts. For each particle i, we define a sequence of lists $\{L_i^{(k)} : k \ge 0\}$ associated with i, given as follows:

(a) $L_i^{(0)} = \{i\}$.

(b) For $k \ge 1$, let $C_q^{(k-1)}$ be the batch that particle i stays in for $t \in [t_{k-1}, t_k)$. Then,

$$L_i^{(k)} = \cup_{j \in C_q^{(k-1)}} L_j^{(k-1)}. \tag{3.2}$$

Here, $L_i^{(k)}$ can be viewed as the particles that have impacted i for $t < t_k$. Clearly, a particle $i_1 \in L_i^{(k)}$ might not have been a batchmate of i. It could have been a batchmate of i_2, and then i_2 was a batchmate of i at some time. The important observation is that if $L_i^{(k)}$ and $L_j^{(k)}$ do not intersect for a given sequence of random batches, then particles i and j are independent at t_k^-. Note that we are not claiming all particles in $L_j^{(k)}$ are independent of those in $L_j^{(k)}$ at t_k^-. In fact, it is possible that some $i_1 \in L_i^{(k)}$ and $j_1 \in L_j^{(k)}$ are in the same batch on $[t_{k-1}, t_k)$. However, i_1 and j_1 must be independent at the times when they were added to the batches that eventually impact i, j at t_k^-. This motivates us to define the following.

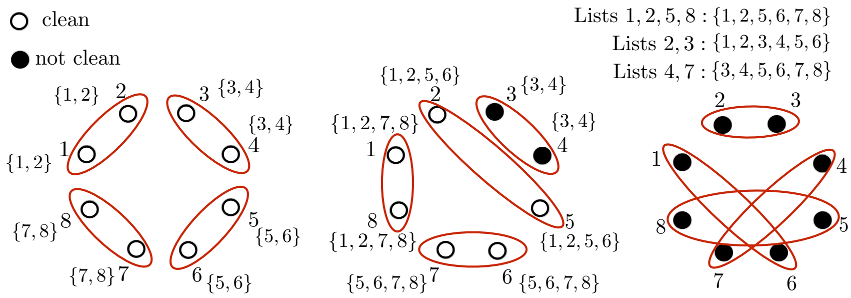

Figure 2: Illustration of the definitions of $L_i^{(k)}$ and particles being clean. The three pictures are for t_1^-, t_2^-, t_3^- respectively with $N=8$, $p=2$. The lists (i.e. $\{1,2\}, \{1,2,5,6\}$) indicate $L_i^{(k)}$ for the corresponding particles.

Definition 3.1. *We say particle i is clean on $[t_k, t_{k+1})$ if the batch $\mathcal{C}_q^{(k)}$ that contains i at t_k^+ satisfies the following: (1) any $j \in \mathcal{C}_q^{(k)}$ is clean at t_k^-; (2) any $j, \ell \in \mathcal{C}_q^{(k)}$ with $j \neq \ell$, $L_j^{(k)}$ and $L_\ell^{(k)}$ do not intersect.*

Fig. 2 gives the illustration for the definitions of $L_i^{(k)}$ and particles being clean. Plainly speaking, a particle i is "clean" at t_k^- if its batchmates at $t < t_k$ were mutually independent and independent to i when they interacted.

The proof of the following useful observation is a straightforward induction. We omit it. The symbol $|A|$ below for a set A means the cardinality of A.

Lemma 3.2. *Consider a fixed sequence of divisions of random batches $\{\mathcal{C}^{(\ell)}\}_{\ell \le k-1}$.*

(i) It holds that

$$|L_i^{(k)}| \le p^k,$$

and the particle i is clean at t_k^- if and only if the equality holds.

(ii) The distribution of X^i for a clean particle i at t_k^- is $\mathcal{G}_\infty^k(\mu_0)$.

Let A_k denote the set of particles that are clean at t_k^-. Then,

$$A_1 = A_0 = \{1, \cdots, N\}.$$

For $k \ge 2$, one has

$$A_k = \Big\{ i \in A_{k-1} : i \in \mathcal{C}_q^{(k-1)}, \forall j, \ell \in \mathcal{C}_q^{(k-1)}, j \neq \ell, \\ j \in A_{k-1}, \ell \in A_{k-1}, L_j^{(k-1)} \cap L_\ell^{(k-1)} = \emptyset \Big\}. \tag{3.3}$$

Denote

$$\epsilon_k := \mathbb{P}(1 \notin A_k). \tag{3.4}$$

Note that by symmetry, ϵ_k is also the probability that particle i is not clean. We state our main result.

Theorem 3.1. *It holds that*

$$W_2(\mathcal{G}_\infty^k(\mu_0), \mathcal{G}_N^k(\mu_0)) \le C \exp(\alpha t_k) \epsilon_k, \tag{3.5}$$

for some $\alpha > 0$. In the strong confinement case $\alpha = 0$.

To prove Theorem 3.1, we need some preparation. We first establish some moments estimates.

Lemma 3.3. *If the intial distribution satisfies Assumption* (2.1)*, then*

$$\sup_{k:k\tau\le T}\int_{\mathbb{R}^d}|x|^q\mathcal{G}_\infty^k(\mu_0)(dx)\le C(q)e^{\alpha T}. \tag{3.6}$$

In the strong confinement case (assumption 2.3*), one can take $\alpha=0$, i.e., the constant in the upper bound can be uniform in T.*

Proof. We note that $\{\mathcal{G}_\infty^k\}$ is a semigroup, so it suffices to estimate the growth of the moments in one step.

Consider (1.8). By Itô's calculus, one has

$$d\mathbb{E}|Y^i|^q=q\mathbb{E}|Y^i|^{q-2}Y^i\cdot\left[b(Y^i)+\frac{1}{p-1}\sum_{j=1,j\neq i}^{p}K(Y^i-Y^j)\right]dt \\ +\mathbb{E}q|Y^i|^{q-2}(d+q-2)\sigma^2\,dt. \tag{3.7}$$

Using the one-sided Lipschitz condition in Assumption 2.2, one has

$$Y^i\cdot b(Y^i)=(Y^i-0)\cdot(b(Y^i)-b(0))+Y^i\cdot b(0)\le\beta|Y^i|^2+C|Y^i|.$$

Similarly, since K is Lipschitz, one has $|K(Y^i-Y^j)|\le|K(0)|+L(|Y^i|+|Y^j|)$, and thus

$$Y^i\cdot K(Y^i-Y^j)\le|K(0)||Y^i|+L(|Y^i|^2+|Y^i||Y^j|).$$

It follows that

$$\mathbb{E}|Y^i|^{q-2}Y^i\cdot\left[b(Y^i)+\frac{1}{p-1}\sum_{j=1,j\neq i}^{p}K(Y^i-Y^j)\right] \\ \le(\beta+L)\mathbb{E}|Y^i|^q+C\mathbb{E}|Y^i|^{q-1}+\frac{1}{p-1}\sum_{j:j\neq i}\mathbb{E}|Y^i|^{q-1}|Y^j|.$$

By Young's inequality,

$$\mathbb{E}|Y^i|^{q-1}|Y^j|\le\frac{(q-1)\nu}{q}\mathbb{E}|Y^i|^q+\frac{\mathbb{E}|Y^j|^q}{q\nu^{q-1}}$$

for any $\nu>0$. In particular, one also has

$$\mathbb{E}|Y^i|^{q-1}\le\frac{(q-1)\nu}{q}\mathbb{E}|Y^i|^q+\frac{1}{q\nu^{q-1}}.$$

Similarly, using Young's inequality, $\mathbb{E}|Y^i|^{q-2}$ is also easily controlled by $\delta\mathbb{E}|Y^i|^q+C(\delta)$ for some small δ.

By the exchangeability so that $\mathbb{E}|Y^i|^q=\mathbb{E}|Y^j|^q$, one then has

$$\frac{d}{dt}\mathbb{E}|Y^i|^q\le q(\beta+2L+\delta)\mathbb{E}|Y^i|^q+C_2.$$

In the strong confinement case as in Assumption 2.3,

$$\mathbb{E}|Y^i|^{q-2}X^i\cdot\left[b(Y^i)+\frac{1}{p-1}\sum_{j=1,j\neq i}^{p}K(Y^i-Y^j)\right] \\ \le(-r+L)\mathbb{E}|Y^i|^q+C\mathbb{E}|Y^i|^{q-1}+\frac{L}{p-1}\sum_{j:j\neq i}\mathbb{E}|Y^i|^{q-1}|Y^j| \\ \le(-r+L+\frac{(q-1)L}{q})\mathbb{E}|Y^i|^q+\frac{L}{p-1}\sum_{j:j\neq i}\frac{1}{q}\mathbb{E}|Y^j|^q+\delta\mathbb{E}|Y^i|^q+C(\delta) \\ =(-r+2L)\mathbb{E}|Y^i|^q+\delta\mathbb{E}|Y^i|^q+C(\delta),$$

where δ is a sufficiently small but fixed number. The conclusions then follow easily. □

We also need the moment control for the Random Batch Method conditioning on any specific sequence of random batches.

Lemma 3.4. *Consider a fixed sequence of divisions of random batches $\{\mathcal{C}^{(\ell)}\}$. Again, consider the solutions $\{X^i(t)\}_{i=1}^N$ to (1.5). Then,*

$$\sup_{t\le T}\sup_i \mathbb{E}\left(|X^i|^q|\{\mathcal{C}^{(\ell)}\}\right) \le C(q)e^{\alpha T}. \tag{3.8}$$

In the strong confinement assumption 2.2,

$$\sup_{t\ge 0}\sup_i \mathbb{E}\left(|X^i|^q|\{\mathcal{C}^{(\ell)}\}\right) \le C(q), \tag{3.9}$$

where $C(q)$ and α do not depend on the specific sequence of divisions of random batches $\{\mathcal{C}^{(\ell)}\}$.

Proof. The proof follows the same line as that in Lemma 3.3. The difference is that there is no exchangeability now conditioning on the random batches.

Under Assumption 2.2 and using the similar estimates as in Lemma 3.3, one has for $t\in[t_k,t_{k+1}]$ that

$$\begin{aligned}\frac{d}{dt}\mathbb{E}(|X^i|^q|\{\mathcal{C}^{(\ell)}\}) \le q\left(\beta+L+\frac{(q-1)L}{q}+\delta\right)\mathbb{E}(|X^i|^q|\{\mathcal{C}^{(\ell)}\})\\ +\frac{qL}{p-1}\sum_{j:j\neq i}\frac{1}{q}\mathbb{E}(|X^j|^q|\{\mathcal{C}^{(\ell)}\})+C(\delta).\end{aligned} \tag{3.10}$$

Under Assumption 2.3, one then has for $t\in[t_k,t_{k+1}]$ that

$$\begin{aligned}\frac{d}{dt}\mathbb{E}(|X^i|^q|\{\mathcal{C}^{(\ell)}\}) \le q\left(-r+L+\frac{(q-1)L}{q}+\delta\right)\mathbb{E}(|X^i|^q|\{\mathcal{C}^{(\ell)}\})\\ +\frac{qL}{p-1}\sum_{j:j\neq i}\frac{1}{q}\mathbb{E}(|X^j|^q|\{\mathcal{C}^{(\ell)}\})+C(\delta).\end{aligned} \tag{3.11}$$

Next, based on (3.10), one easily finds

$$\begin{aligned}\mathbb{E}(|X^i(t)|^q|\{\mathcal{C}^{(\ell)}\})-\mathbb{E}(|X^i|^q(t_k)|\{\mathcal{C}^{(\ell)}\}) \le & q\left(\beta+L+\frac{(q-1)L}{q}+\delta\right)\int_{t_k}^t \mathbb{E}(|X^i(s)|^q|\{\mathcal{C}^{(\ell)}\})\,ds\\ &+L\int_{t_k}^t\max_{1\le i\le p}\mathbb{E}(|X^i(s)|^q|\{\mathcal{C}^{(\ell)}\})\,ds+\int_{t_k}^t C(\delta)\,ds.\end{aligned}$$

It follows that

$$a(t):=\max_{1\le i\le p}\mathbb{E}(|X^i|^q|\{\mathcal{C}^{(\ell)}\}) \tag{3.12}$$

satisfies

$$a(t)\le a(t_k)+q(\beta+2L+\delta)\int_{t_k}^t[a(s)+C(\delta)]\,ds.$$

Grönwall's inequality then yields the first claim with any $\alpha>\beta+2L$.

For (3.11), defining $r_1 := q\left(r-L-\frac{q-1}{q}L-\delta\right)>0$, one finds that

$$\begin{aligned}\mathbb{E}(|X^i(t)|^q|\{\mathcal{C}^{(\ell)}\}) \le \mathbb{E}(|X^i|^q(t_k)|\{\mathcal{C}^{(\ell)}\})e^{-r_1(t-t_k)}\\ +\int_{t_k}^t e^{-r_1(t-s)}\left[L\max_{1\le i\le p}\mathbb{E}(|X^i(s)|^q|\{\mathcal{C}^{(\ell)}\})+C(\delta)\right]ds.\end{aligned}$$

Hence, the function a defined in (3.12) satisfies

$$a(t)\le a(t_k)e^{-r_1(t-t_k)}+\int_{t_k}^t e^{-r_1(t-s)}\left[La(s)+C(\delta)\right]ds.$$

It can be shown easily that $a(t)$ is controlled by $b(t)$ which satisfies the following integral equality

$$b(t) = a(t_k)e^{-r_1(t-t_k)} + \int_{t_k}^{t} e^{-r_1(t-s)}\left[Lb(s) + C(\delta)\right] ds.$$

(One can perturb the initial data $a(t_k) \to a(t_k) + \epsilon$ for $b(\cdot)$ and then take $\epsilon \to 0$).

Then, one finds

$$b'(t) = (-r_1 + L)b(t) + C(\delta) = q(-r + 2L + \delta)b(t) + C(\delta),\ b(t_k) = a(t_k).$$

Hence,

$$a(t_{k+1}) \le b(t_{k+1}) \le a(t_k)e^{-q(r-2L+\delta)\tau} + \frac{C(\delta)}{q(r-2L-\delta)}(1 - e^{-q(r-2L+\delta)\tau}).$$

The second claim also follows. □

Now, we can prove the main theorem in this section.

Proof of Theorem 3.1. First of all, it is clear by Lemma 3.2 that

$$\mathcal{G}_N^k(\mu_0) = \mathbb{P}(1 \in A_k)\mathcal{G}_\infty^k(\mu_0) + \mathbb{P}(1 \notin A_k)\nu_k, \tag{3.13}$$

for some probability measure ν_k. ν_k is the mixture of the conditional marginal distribution of X^1, each is the one conditioning on some particular sequence of batches.

By (3.13), it holds that

$$|\mathcal{G}_\infty^k(\mu_0) - \mathcal{G}_N^k(\mu_0)| \le (1 - \mathbb{P}(1 \in A_k))\mathcal{G}_\infty^k(\mu_0) + \mathbb{P}(1 \notin A_k)\nu_k = \epsilon_k(\mathcal{G}_\infty^k(\mu_0) + \nu_k). \tag{3.14}$$

Therefore, the total variation distance between the two measures is controlled by

$$\|\mathcal{G}_\infty^k(\mu_0) - \mathcal{G}_N^k(\mu_0)\|_{TV} \le (1 - \mathbb{P}(1 \in A_k)) + \mathbb{P}(1 \notin A_k) = 2\epsilon_k. \tag{3.15}$$

By Lemma 3.4, for each sequence of batches, one has

$$\sup_i \mathbb{E}(|X^i|^q|\mathcal{C}^{(\ell)}) \le C(q)e^{\alpha t}.$$

Hence, it holds that

$$\int_{\mathbb{R}^d} |x|^q \nu_k(dx) \le C(q)e^{\alpha t}. \tag{3.16}$$

In the case of strong confinement, $\alpha = 0$. Similarly, by Lemma 3.3, $\mathcal{G}_\infty^k(\mu_0)$ has the same moment control.

Application of Lemma 2.2 then yields the desired result. □

Lastly, we close up the estimate.

Theorem 3.2. *For any fixed k, it holds that*

$$\lim_{N\to\infty} \epsilon_k = 0. \tag{3.17}$$

Proof. First of all, clearly, we have

$$\epsilon_0 = \epsilon_1 = 1 - 1 = 0.$$

Now, we do induction on k. Assume

$$\lim_{N\to\infty} \epsilon_k = 0.$$

Consider the batches for $t_k \to t_{k+1}^-$. Assume the batch for particle 1 is $\mathcal{C}_q^{(k)}$. Denote

$$B_k = \Big\{\forall j, \ell \in \mathcal{C}_q^{(k)}, j \neq \ell, L_j^{(k)} \cap L_\ell^{(k)} = \emptyset\Big\}.$$

Let $\mathcal{B} = \mathcal{C}_q^{(k)} \setminus \{1\}$ be the set of other particles that share the same batch with particle 1. Then, by definition of A_{k+1},

$$\mathbb{P}(1 \in A_{k+1}) = \sum_{j_1,\cdots,j_{p-1}} \mathbb{P}(\mathcal{B} = \{j_1, \cdots, j_{p-1}\})\times \\ \mathbb{P}(B_k \cap \{1 \in A_k\} \cap_{\ell=1}^{p-1} \{j_\ell \in A_k\} | \mathcal{B} = \{j_1, \cdots, j_{p-1}\}). \quad (3.18)$$

Denote $E := \{\mathcal{B} = \{j_1, \cdots, j_{p-1}\}\}$, where we omit the dependence in $j_\ell, 1 \le \ell \le p-1$ for notational convenience. Conditioning on $B_k \cap E$ (i.e., provided that the event $B_k \cap E$ happens), whether the particles are clean or not are independent. Hence,

$$\begin{aligned}\mathbb{P}(B_k \cap \{1 \in A_k\} \cap_{\ell=1}^{p-1} \{j_\ell \in A_k\}|E) &= \mathbb{P}(B_k|E)\mathbb{P}(\{1 \in A_k\} \cap_{\ell=1}^{p-1} \{j_\ell \in A_k\}|E, B_k) \\ &= \mathbb{P}(B_k|E) \prod_{\ell=1}^{p} \mathbb{P}(j_\ell \in A_k|E, B_k),\end{aligned}$$

where we have set $j_p = 1$. Moreover,

$$\mathbb{P}(j_\ell \in A_k|E, B_k) = \frac{\mathbb{P}(\{j_\ell \in A_k\} \cap E \cap B_k)}{\mathbb{P}(E \cap B_k)} = \frac{\mathbb{P}(\{1 \in A_k\} \cap E \cap B_k)}{\mathbb{P}(E \cap B_k)} = \frac{\mathbb{P}(\{1 \in A_k\} \cap B_k)}{\mathbb{P}(B_k)}.$$

The second and the last equalities are due to symmetry. For the last equality, $\mathbb{P}(\{\mathcal{B} = \{j_1, \cdots, j_{p-1}\}\} \cap B_k)$ should be equal for all possible $j_1, \cdots, j_{p-1}$, and the same is true for the numerator. This actually is a kind of independence. Hence, eventually due to the fact

$$\sum_{j_1,\cdots,j_{p-1}} \mathbb{P}(\mathcal{B} = \{j_1, \cdots, j_{p-1}\})\mathbb{P}(B_k|\mathcal{B} = \{j_1, \cdots, j_{p-1}\}) = \mathbb{P}(B_k),$$

one has

$$1 - \epsilon_{k+1} = \mathbb{P}(1 \in A_{k+1}) \ge \mathbb{P}(B_k)(1 - \epsilon_k/\mathbb{P}(B_k))^p.$$

Hence, it suffices to show

$$\lim_{N\to\infty} \mathbb{P}(B_k) = 1.$$

To get an estimate for this, we think the following way to construct $L_i^{(k)}$: one starts with $L_i := \{i\}$ and repeat the following for k times:

(1) Set $L_{tmp} \leftarrow L_i$ and $A = \emptyset$.

(2) Loop the following until L_{tmp} is empty.

   (a) Pick a particle $i_1 \in L_{tmp}$, then choose $p-1$ particles from $\{1, \cdots, N\} \setminus A \cup \{i_1\}$ denoted by $\{i_2, \cdots, i_p\}$.
   (b) Set $L_i \leftarrow L_i \cup \{i_2, \cdots, i_p\}$.
   (c) Set $A \leftarrow A \cup \{i_1, i_2, \cdots, i_p\}$.
   (d) Set $L_{tmp} \leftarrow L_{tmp} \setminus \{i_1, i_2, \cdots, i_p\}$.

In the above, we are actually looking back from t_{k-1}. In the jth iteration, we are constructing batches at t_{k-j}. Hence, this is an equivalent way to construct $L_i^{(k)}$.

Now, we estimate $\mathbb{P}(B_k)$ by constructing the lists $L_{j_\ell}^{(k)} : 1 \le \ell \le p$ for $j_\ell \in C_q^{(k)}$ using the above procedure. Consider that the lists for $j_1, \cdots, j_{\ell-1}$ have been constructed, which have included at most $(\ell-1)p^k$ particles. Now, for $L_{j_\ell}^{(k)}$ not to intersect with the previous lists, one has to choose particles from $\{1,\cdots,N\}\backslash[\cup_{z=1}^{\ell-1}L_{j_z}^{(k)}\cup A\cup\{i_1\}]$ in 2(a) step. Conditioning on the specific choices of $L_{j_\ell}^{(k)}: 1\le \ell\le p$ and A with

$$N_1 := |L_{j_\ell}^{(k)}\cup A\cup\{i_1\}|,\ N_2 := |A|,$$

this conditional probability is controlled below by

$$\frac{\binom{N-N_1}{p-1}}{\binom{N-1-N_2}{p-1}} \ge \frac{\binom{N-1-\ell p^k}{p-1}}{\binom{N-1}{p-1}}.$$

Hence, as $N\to\infty$,

$$\mathbb{P}(B_k) \ge \prod_{\ell=1}^{p}\left[\frac{\binom{N-1-\ell p^k}{p-1}}{\binom{N-1}{p-1}}\right]^k = 1 - O(N^{-1}).$$

Hence, the claim follows and actually $\epsilon_k \le C(p,k)N^{-1}$ for some $C(p,k) > 0$. □

Remark 3.1. *Clearly, the current argument of the mean field limit relies on the fact that two particles are unlikely to be related in RBM when $N\to\infty$ for finite iterations. This is not enough to get the mean field limit independent of τ. In particular, for fixed N, $\epsilon_k\to 1$ as $k\to\infty$. As pointed in [29], RBM works due to the averaging effect in time. The regime we consider here (finite iterations and $N\to\infty$) is clearly far before the averaging effect in time comes into play. Hence, if one wants to consider the mean field limit uniform in τ (the averaging mechanism can take effect), one must consider carefully how the correlation decays as k grows when two particles are not totally clean to each other. The study of this creation of chaos will be left for the future.*

4 Properties of the limiting dynamics

We consider the limit dynamics given by the operator $\mathcal{G}_\infty$ (defined in (1.7)) and its approximation to the dynamics of the nonlinear Fokker-Planck equation (1.2), the mean-field limit of the interacting particle system (1.1).

As proved in [29], the error between the one marginal distribution of the RBM particle system (1.5) and that of (1.1) are close independent of N under W_2 distance (the left side in Fig. 1). Combining the mean field result in section 3 and taking $N\to\infty$, one sees that the dyanmics of $\mathcal{G}_\infty$ is close to that of (1.2) (the right side in Fig. 1). In other words, the two limits $\lim_{N\to\infty}$ and $\lim_{\tau\to 0}$ commute.

However, a direct application of the strong mean square error in [29] gives an upper bound $O(\sqrt{\tau})$ for the W_2 distance corresponding to the left side in Fig. 1, and thus the right side in Fig. 1 after taking $N\to\infty$. It is shown in [28] (though for $b(\cdot)$ being bounded) that the weak error is $O(\tau)$. The W_2 distance is a kind of weak topology as it measures the closeness between distributions instead of the trajectories of particles. Hence, the sharp upper bound for the W_2 distance between these two marginal distributions is believed to be $O(\tau)$, even for unbounded $b(\cdot)$. Below, we aim to prove these.

4.1 Stability of the limiting dynamics

In this section, we study the stability and contraction properties of the nonlinear operator $\mathcal{G}_\infty$, for the limiting dynamics.

Proposition 4.1. *Under Assumption 2.2, the nonlinear operator $\mathcal{G}_\infty$ satisfies*

$$W_2(\mathcal{G}_\infty(\mu_1),\mathcal{G}_\infty(\mu_2)) \le e^{\alpha\tau/2}W_2(\mu_1,\mu_2),\ \ \mu_i\in\mathbf{P}(\mathbb{R}^d),\ i=1,2. \tag{4.1}$$

The operator $\mathcal{G}_\infty$ is a contraction in W_2 under Assumption 2.3:

$$W_2(\mathcal{G}_\infty(\mu_1),\mathcal{G}_\infty(\mu_2)) \le e^{-(r-2L)\tau}W_2(\mu_1,\mu_2), \tag{4.2}$$

so that $\mathcal{G}_\infty$ has a unique invariant measure π_τ and it holds for any μ_0 that

$$W_2(\mathcal{G}_\infty^n(\mu_0), \pi_\tau) \le e^{-(r-2L)} W_2(\mu_0, \pi_\tau). \tag{4.3}$$

Proof. Consider two copies of (1.8): one is

$$dY_1^i = b(Y_1^i)\,dt + \frac{1}{p-1}\sum_{j=1, j\neq i}^{p} K(Y_1^i - Y_1^j)\,dt + \sqrt{2}\sigma\,dW^i,\ i = 1, \cdots, p, \tag{4.4}$$

with $(Y_1^1(0), \cdots, Y_1^p(0))$ being drawn from $\mu_1^{\otimes p}$; the other one is

$$dY_2^i = b(Y_2^i)\,dt + \frac{1}{p-1}\sum_{j=1, j\neq i}^{p} K(Y_2^i - Y_2^j)\,dt + \sqrt{2}\sigma\,dW^i,\ i = 1, \cdots, p, \tag{4.5}$$

with $(Y_2^1(0), \cdots, Y_2^p(0))$ being drawn from $\mu_2^{\otimes p}$.

For any $\epsilon > 0$, we choose the coupling such that the joint distribution $\tilde{\gamma}$ between $(Y_1^1(0), \cdots, Y_1^p(0))$ and $(Y_2^1(0), \cdots, Y_2^p(0))$ satisfies

$$\mathbb{E}\sum_{i=1}^{p} |Y_1^i(0) - Y_2^i(0)|^2 = \int_{(\mathbb{R}^d)^{\otimes p}} \sum_{i=1}^{p} |x_i - y_i|^2 d\tilde{\gamma} \le W_2^2(\mu_1^{\otimes p}, \mu_2^{\otimes p}) + \epsilon,$$

and that the Brownian motions for the two systems are the same.

By Lemma 2.1,

$$pW_2^2(\mu_1, \mu_2) = W_2^2(\mu_1^{\otimes p}, \mu_2^{\otimes p}),$$

and that

$$pW_2^2(\mathcal{G}_\infty(\mu_1), \mathcal{G}_\infty(\mu_2)) \le \mathbb{E}\sum_i |Y_1^i(\tau) - Y_2^i(\tau)|^2. \tag{4.6}$$

Hence, to show the claims, it suffices to show that the second moments of the SDE system (1.8) is stable. In fact, it can be computed directly that under Assumption 2.2

$$\frac{d}{dt}\mathbb{E}\sum_{i=1}^{p} |Y_1^i - Y_2^i|^2 \le (\beta + 2L)\sum_{i=1}^{p} \mathbb{E}|Y_1^i - Y_2^i|^2,$$

and that under Assumption 2.3

$$\frac{d}{dt}\mathbb{E}\sum_{i=1}^{p} |Y_1^i - Y_2^i|^2 \le (-r + 2L)\sum_{i=1}^{p} \mathbb{E}|Y_1^i - Y_2^i|^2.$$

Then, the first two assertions have been proved.

The last claim follows then from the standard contraction mapping theorem [23, Chap. 1].

□

4.2 Basic properties of the nonlinear Fokker-Planck equation

We establish several basic results to (1.2) using a stronger version of Assumption 2.1:

Assumption 4.1. *The measure μ_0 has a density ϱ_0 that is smooth with finite moments $\int_{\mathbb{R}^d} |x|^q \varrho_0\,dx < \infty$, $\forall q \ge 2$, and the entropy is finite*

$$H(\mu_0) := \int_{\mathbb{R}^d} \varrho_0 \log \varrho_0\,dx < \infty. \tag{4.7}$$

Above, for the entropy, if $\varrho_0(x) = 0$ at some point x, we define $\varrho_0(x)\log\varrho_0(x) = 0$. We also introduce the following assumption on the growth rate of derivatives of b and K, which will be used for some claims below.

Assumption 4.2. *The function b and its derivatives have polynomial growth. The derivatives of K with order at least 2, i.e. $D^\alpha K$ with $|\alpha| \geq 2$, have polynomial growth.*

Based on these conditions, equation (1.2) can be formulated in terms of the density of μ:

$$\begin{aligned}&\partial_t \varrho = -\nabla \cdot ((b(x) + K * \varrho)\varrho) + \sigma^2 \Delta \varrho,\\&\varrho(0) = \varrho_0.\end{aligned} \tag{4.8}$$

Then, a weak solution to (4.8) corresponds to a measure solution $\mu = \varrho\, dx$ to (1.2), where the weak solution is defined as follows.

Definition 4.1. *We say $\varrho \in L^\infty([0,T]; L^1(\mathbb{R}^d))$ is a weak solution to (4.8), if $\varrho\, dx \in C([0,T]; \mathbf{P}(\mathbb{R}^d))$ where $\mathbf{P}(\mathbb{R}^d)$ is equipped with the weak topology, and for any $\varphi \in C_c^\infty(\mathbb{R}^d)$, it holds for any $t \leq T$ that*

$$\begin{aligned}&\int_{\mathbb{R}^d} \varrho(x,t)\varphi(x)\, dx - \int_{\mathbb{R}^d} \varrho_0(x)\varphi(x)\, dx\\&\quad = \int_0^t \int_{\mathbb{R}^d} \nabla\varphi(x) \cdot (b(x) + K * \varrho)\, \rho(x,s) dx ds + \sigma^2 \int_0^t \int_{\mathbb{R}^d} \Delta\varphi(x)\varrho(x,s)\, dx ds.\end{aligned} \tag{4.9}$$

Note that the test function used here does not depend on time variable, so we require the integral equation to hold for any $t \leq T$. Due to the relation between (4.8) and (1.2), we will not distinguish the measure and its density. For example, we will use $\mathcal{G}_\infty(\varrho_0)$ to mean the nonlinear semigroup acting on the measure μ_0, and will use $W_2(\varrho, \nu)$ to mean the Wasserstein-2 distance between $\mu = \varrho\, dx$ and another measure ν.

We have the following regarding the well-posedness of the nonlinear Fokker-Planck equation (4.8).

Proposition 4.2. *Let Assumption 2.2 or Assumption 2.3 hold, and also $|b| + |\nabla b| \leq C(1 + |x|^q)$ for some C, q. Fix any $T > 0$. Assume the initial data ϱ_0 satisfies Assumption 4.1. Then, the nonlinear Fokker-Planck equation (4.8) has a unique weak solution satisfying $\sup_{0\leq t\leq T} \int_{\mathbb{R}^d} |x| \varrho\, dx < \infty$. Moreover, this solution is a strong solution and is smooth together with the moment control:*

$$\sup_{0\leq t\leq T} \int_{\mathbb{R}^d} |x|^q \varrho\, dx \leq C(q,T) \int_{\mathbb{R}^d} |x|^q \varrho_0\, dx. \tag{4.10}$$

Besides, under Assumption 2.3, the moments are uniformly bounded in t, i.e., the constants $C(q,T)$ above can be made independent on T. Moreover, $\mu = \varrho\, dx$ will converge in W_2 to an invariant measure π exponentially as $t \to \infty$.

There are many works on similar models in literature, and see [40, 8, 10, 2] as a few of examples. However, in our case, b and K are not bounded and b can have polynomial growth at infinity, so the proofs in these works do not quite fit our setting here. For example, in the work of [8, 10], $b = -\nabla V$ and they require $\nabla V \cdot x \geq C$ for some constant while we allow $b \cdot x \leq \beta |x|^2$; also the requirements on the kernel $K(\cdot)$ also do not quite match the setup here. In the work [2], a certain class of nonlinear Fokker-Planck equations have been studied via the approach of Crandall and Liggett for m-accretive operators in $L^1(\mathbb{R}^d)$, but the approach cannot be applied directly to our case here. Due to these reasons, we attach a proof of Proposition 4.2 in Appendix A for a reference.

In proving the uniqueness of the solution to (1.2) in Appendix A, we have also in fact proved the following mean-field limit:

Proposition 4.3. *With the same assumptions of Proposition 4.2, one has*

$$\sup_{0\leq t\leq T} W_2(\varrho, \mu_N^{(1)}) \leq \frac{C(T)}{\sqrt{N}}, \tag{4.11}$$

where $\mu_N^{(1)}$ is the one marginal distribution of the interacting particle system (1.1). Moreover, if Assumption 2.3 holds, the constant $C(T)$ can be independent of T.

As long as the existence and uniqueness of the solutions to the nonlinear Fokker-Planck equation have been established, one can regard

$$\bar{K}(x,t) := \int_{\mathbb{R}^d} K(x-y)\varrho(y,t)\,dt, \tag{4.12}$$

as known, and the properties of ϱ can be studied via the *linear* Fokker-Planck equation

$$\partial_t \varrho = -\nabla\cdot[(b(x)+\bar{K}(x,t))\varrho] + \sigma^2\Delta\varrho. \tag{4.13}$$

By the moment estimates of ϱ, $\bar{K}(0,t)$ is bounded by the first moment of ϱ and it is Lipschitz continuous with uniform Lipschitz constant L. We consider the time continuity of $\bar{K}$.

Lemma 4.1. *Under Assumption 2.2, 4.1, 4.2, we have for any $\Delta t\in[0,\tau]$,*

$$\left|\bar{K}(x,t+\Delta t)-\bar{K}(x,t)\right| \le C(M_q(\varrho(t)))(1+|x|^q)\tau, \tag{4.14}$$

for some $q>1$, where $M_q(\varrho(t))$ means the q-moment of ϱ at t. Moreover, if Assumption 2.2 is replaced by Assumption 2.3, $C(M_q(\varrho(t)))$ has an upper bound independent of time t.

Proof. It can be computed directly that

$$\begin{aligned}\partial_t\bar{K}(x,t) &= \int_{\mathbb{R}^d} K(x-y)\{-\nabla\cdot[(b(y)+K*\varrho)\varrho]+\sigma^2\Delta_y\varrho\}dy\\ &= -\int_{\mathbb{R}^d}(b(y)+K*\varrho)\varrho\cdot(\nabla K)(x-y)\,dy + \int_{\mathbb{R}^d}\sigma^2(\Delta K)(x-y)\varrho\,dy.\end{aligned}$$

Since b has polynomial growth and ∇K is bounded, then

$$\begin{aligned}&\left|-\int_{\mathbb{R}^d}(b(y)+K*\varrho)\varrho\cdot(\nabla K)(x-y)\,dy\right|\\ &\le C\int_{\mathbb{R}^d}(1+|y|^q)\varrho\,dy + C\iint_{\mathbb{R}^d\times\mathbb{R}^d}|K(x-y)|\varrho(x)\varrho(y)\,dxdy.\end{aligned}$$

This is controlled by the moments of ϱ.

Moreover, since ΔK has polynomial growth,

$$\begin{aligned}\left|\int_{\mathbb{R}^d}\sigma^2(\Delta K)(x-y)\varrho\,dy\right| &\le \sigma^2 C\int_{\mathbb{R}^d}(1+|x-y|^q)\varrho\,dy\\ &\le C\left(1+\int_{\mathbb{R}^d}(|x|^q+|y|^q)\varrho\,dy\right)\le C(1+|x|^q),\end{aligned}$$

where C depends on the moments of ϱ.

Using the results in Proposition 4.2, the moments on $[t,t+\Delta t]$ can be controlled by the one at t. Since τ is a fixed small number, we omit the dependence in τ for the amplification constant, the claims then follow. □

Before further discussion, we first establish some auxilliaury results regarding the following linear Fokker-Planck equation

$$\partial_t f = -\nabla\cdot(b_1(x,t)f)\,dt + \sigma^2\Delta f =: \mathcal{L}^*_{b_1}(f). \tag{4.15}$$

We will assume $b_1(x,t)$ satisfies

$$(x-y)\cdot(b_1(x,t)-b(y,t)) \le \beta_1|x-y|^2. \tag{4.16}$$

We say b_1 satisfies the strong confinement condition if $\beta_1<0$. We also denote $S_{s,t}$ the solution operator from time s to time t:

$$f_t =: S_{s,t}f_s. \tag{4.17}$$

Study of the parabolic equation (4.15) with bounded drift b_1 or drift with linear growth is classical (see for example [33]). However, the study of the cases with drifts that may have polynomial growth seems limited, especially the properties of the fundamental solutions, which we need below (see Lemma 4.3 and Proposition 4.4). Below, we will show several results for our needs.

Lemma 4.2. *Consider equation (4.15), where b_1 satisfies (4.16). Also, assume the derivatives of $b_1(x,t)$ have polynomial growth and $\sup_{t\ge 0}|b(0,t)|<\infty$. Then,*

(i) For any $g\in L^1(\mathbb{R}^n)$, one has

$$\sup_{\Delta t\le T}\int_{\mathbb{R}^d}|x|^q|S_{t,t+\Delta t}g|\,dx\le C(T)\int_{\mathbb{R}^d}|x|^q|g(x)|\,dx.$$

(ii) If b_1 satisfies the strong confinement condition $\beta_1<0$, $C(T)$ in item (i) can be made independent of T. Moreover, when $\sigma>0$ and $\int_{\mathbb{R}^d}g\,dx=0$, $\beta_1<0$ implies that

$$\int_{\mathbb{R}^d}|x|^q|S_{t,t+\Delta t}g|\,dx\le P(M_{q_1}(|g|))e^{-\delta\Delta t},$$

where $\delta>0$ is independent of g, $P(\cdot)$ is some polynomial, $q_1>q$ is some suitable number, and $M_{q_1}(|g|)$ means the q_1-moment of $|g|$.

(iii) In the case b_1 does not depend on time so that $S_{s,t}=e^{(t-s)\mathcal{L}_{b_1}^*}$, one also has

$$\sup_{\Delta t\le T}\int_{\mathbb{R}^d}|x|^q|(\mathcal{L}_{b_1}^*)^m S_{t,t+\Delta t}g|\,dx\le C(T)\int_{\mathbb{R}^d}|x|^q|(\mathcal{L}_{b_1}^*)^m g|\,dx.$$

Proof. For (i), one decomposes $g=: g^+-g^-$. Then, one can normalize $g^\pm$ to probability densities. Then, $S_{t,t+\Delta t}g=(S_{t,t+\Delta t}g^+)-(S_{t,t+\Delta t}g^-)$ with each of them being nonnegative. Following similar approaches of Step 1 in Appendix A, we can show that the moments of $S_{t,t+\Delta t}g^\pm$ can be controlled by those of $g^\pm$. Hence, the moments of $S_{t,t+\Delta t}g$ has the desired estimates. We skip the details.

Regarding (ii), we first note that the q moments of $S_{t,t_1}g$ can be uniformly controlled by moments of $|g|$, due to similar reasons. Then, one can consider the measures $\mu^\pm := \frac{1}{\|g^\pm\|_{L^1}}S_{t,t+\Delta t}g^\pm$. Using standard techniques of Markov chains (see [39, Appendix A] and [41, Chapters 15-16]), one can show that

$$\|\mu^+-\mu^-\|_{TV}\le P_1(M_{q_1}(|\mu|))e^{-\delta'\Delta t}\Rightarrow \|S_{t,t+\Delta t}g\|_{TV}\le P(M_{q_1}(|g|))e^{-\delta'\Delta t},$$

for some $q_1>q$ and polynomials $P_1(\cdot)$, $P(\cdot)$. Then

$$\begin{aligned}\int_{\mathbb{R}^d}|x|^q|S_{t,t+\Delta t}g|(dx)&=\frac{1}{2}\|g\|_{L^1}\int|x|^q|\mu^+-\mu^1|(dx)\\&\le C\|g\|_{L^1}\sqrt{\int|x|^{2q}|\mu^+-\mu^1|(dx)}\sqrt{\|\mu^+-\mu^1\|_{TV}}\\&=C\sqrt{\int_{\mathbb{R}^d}|x|^{2q}|S_{t,t+\Delta t}g|(dx)}\sqrt{\|S_{t,t+\Delta t}g\|_{TV}}.\end{aligned}$$

Further, due to

$$\sqrt{\int_{\mathbb{R}^d}|x|^{2q}|S_{t,t+\Delta t}g|(dx)}\le\frac{1}{2}(1+M_{2q}(|g|)),$$

one can then choose another q_1 large enough such that claims in (ii) hold.

For (iii), we just note that

$$\partial_t((\mathcal{L}_{b_1}^*)^m f)=\mathcal{L}_{b_1}^*((\mathcal{L}_{b_1}^*)^m f).$$

Then, we apply the property of $e^{t\mathcal{L}_{b_1}^*}$ proved in the first part (i). □

Remark 4.1. *For (ii), if $\sigma = 0$, even if the strong confinement condition is satisfied, $\|\mu^+ - \mu^-\|_{TV}$ may not decay. However, we believe when $b_1(x,t) \to b_\infty(x)$, then*

$$\int_{\mathbb{R}^d} |x - x_*|^q |S_{t,t+\Delta t} g|\, dx \le C(M_q(|g|), M_{q_1}(g)) e^{-\delta \Delta t}$$

still holds for the limiting point x_* of the trajectories. We do not explore this in this work.

It is well-known that the linear equation (4.15) has a transition density $\Phi(x,t;y,s)$ solving (4.15) for $t > s$ with initial data $\Phi(x,s;y,s) = \delta(x-y)$. Then,

$$(S_{s,t} g)(x) = \int_{\mathbb{R}^d} \Phi(x,t;y,s) g(y)\, dy. \tag{4.18}$$

Hence, the property of Φ is important.

Lemma 4.3. *Consider equation (4.15) with $\sigma > 0$, and b_1 satisfying (4.16). Also, assume the derivatives of $b_1(x,t)$ have polynomial growth and $\sup_{t\ge 0} |b_1(0,t)| < \infty$. Then, for all $0 \le s < t \le T$, we have*

$$\int_{\mathbb{R}^d} (1 + |x|^q) |\nabla_y \Phi(x,t;y,s)|\, dx \le C(T) P(|y|)(1 + (t-s)^{-1/2}), \tag{4.19}$$

for some polynomial $P(\cdot)$. If $\beta_1 < 0$,

$$\int_{\mathbb{R}^d} (1 + |x|^q) |\nabla_y \Phi(x,t;y,s)|\, dx \le C P(|y|)(1 + (t-s)^{-1/2}) e^{-\delta(t-s)} \tag{4.20}$$

for some $\delta > 0$.

Proof of Lemma 4.3 is tedious, and we defer it to Appendix B. Below, we aim to consider the moments of the derivatives of ϱ. Now, we recall the standard multi-index notation used in PDE community:

$$D^\alpha := \prod_{j=1}^d \partial_j^{\alpha^j}, \ \alpha = (\alpha^1, \cdots, \alpha^d). \tag{4.21}$$

The length of the index is defined as $|\alpha| := \sum_{j=1}^d \alpha^j$.

The following proposition is helpful for our estimates later.

Proposition 4.4. *Let Assumptions 2.2, 4.1, and 4.2 hold. Then, for any multi-index α, it holds that*

$$\sup_{t \le T} \int_{\mathbb{R}^d} (1 + |x|^q) |D^\alpha \varrho| dx \le C(\alpha, q, T). \tag{4.22}$$

If $\sigma > 0$ and Assumption 2.3 holds, then

$$\sup_{t > 0} \int_{\mathbb{R}^d} (1 + |x|^q) |D^\alpha \varrho| dx < \infty. \tag{4.23}$$

Proof. We set

$$b_1(x,t) := b(x) + \bar{K}(x,t),$$

which is regarded as known (since existence and uniqueness of ϱ have been established).

In the case $\sigma = 0$, consider the characteristics satisfying

$$\dot{Z} = b(Z), \ Z(0;y) = y.$$

Using the one-sided Lipschitz condition in Assumption 2.2, one has $v \cdot \nabla b_1(x,t) \cdot v \le \beta_1 |v|^2$ for any v, x. With this and induction, one can show that $|\partial_{y_i} Z| \le C e^{\beta_1 t}$ and $D_y^\alpha Z(t;y)$

is controlled by polynomials of $|y|$ for higher order α. Using $\varrho = Z_{\#}\varrho_0$, the claim can be proved. We omit the details.

Now, we focus on $\sigma > 0$. We do by induction on the derivatives of ϱ. Let $\ell = |\alpha|$. We know already that the claim holds for $\ell = 0$.

Suppose the claim is true for $\ell - 1$ with $\ell \geq 1$. Now, we consider ℓ. One can see that

$$\partial_t D^\alpha \varrho = -\nabla \cdot (b_1(x,t) D^\alpha \varrho) + \sigma^2 \Delta D^\alpha \varrho + \sum_{|\beta| \leq \ell - 1} C_\beta \nabla \cdot [f_\beta(x) D^\beta \varrho].$$

Here, f_β are some functions with polynomial growth. Then, we have

$$D^\alpha \varrho = S_{0,t} D^\alpha \varrho_0 - \int_0^t \sum_{|\beta| \leq \ell - 1} C_\beta \nabla_y \Phi(x,t;y,s) \cdot (f_\beta(y) D^\beta \varrho(y,s))\, ds.$$

The claim follows by a direct application of Lemma 4.3 and the induction assumption. □

4.3 Approximation of the limiting dynamics to the nonlinear Fokker-Planck equation

To get a feeling how close the dynamics given by $\mathcal{G}_\infty$ (the mean field limit of RBM) is to the nonlinear Fokker-Planck equation (1.2), we consider (1.5). Recall that $\rho^{(p)}(\cdots, 0) = \tilde{\mu}(\cdot, t_k)^{\otimes p}$, with order τ error, (1.5) is approximated as

$$\partial_t \rho^{(p)} = -\sum_{i=1}^p \nabla_{x_i} \cdot \left(\left[b(x_i) + \frac{1}{p-1} \sum_{j: j \neq i} K(x_i - x_j) \right] \prod_{j=1}^p \tilde{\mu}(x_j, t_k) \right) + \sigma^2 \sum_{i=1}^p \Delta_{x_i} \rho^{(p)} + O(\tau). \tag{4.24}$$

Since we are curious about how the marginal distribution is evolving, one may take the integrals on $x_2, \cdots, x_p$ and have:

$$\partial_t \tilde{\rho} = -\nabla_{x_1} \cdot ([b(x_1) + K * \tilde{\mu}(\cdot, t_k)] \tilde{\mu}(x_1, t_k)) + \sigma^2 \Delta_{x_1} \tilde{\rho} + O(\tau).$$

Since $\tilde{\rho} := \int \rho^{(p)}\, dx_2 \cdots dx_p$ is equal to $\tilde{\mu}(\cdot, t_k)$ initially, one finds that this is close to (1.2) already. Thus, one expects that the overall error between $\mathcal{G}_\infty^k(\varrho_0)$ and $\varrho(k\tau)$ is like $O(\tau)$.

We now state the main results in this section.

Theorem 4.1. *Let ϱ be the solution to the nonlinear Fokker-Planck equation (4.8). Suppose Assumptions 2.2, 4.1 and 4.2 hold. Then,*

$$\sup_{n: n\tau \leq T} W_2(\mathcal{G}_\infty^n(\varrho_0), \varrho(n\tau)) \leq C(T)\tau. \tag{4.25}$$

If Assumption 2.3 is assumed in place of Assumption 2.2 and also $\sigma > 0$, then

$$\sup_{n \geq 0} W_2(\mathcal{G}_\infty^n(\varrho_0), \varrho(n\tau)) \leq C\tau. \tag{4.26}$$

Consequently, the invariant measures (see Proposition 4.1 and Proposition 4.2 for the related notations) satisfy

$$W_2(\pi_\tau, \pi) \leq C\tau. \tag{4.27}$$

Below, we aim to prove Theorem 4.1. We first establish the one-step error and then give the global estimate.

Define

$$M_{q,\ell}^{(k)} := \sum_{|\alpha| \leq \ell} \int_{\mathbb{R}^d} (1 + |x|^q) |D^\alpha \varrho(x, t_k))|\, dx, \tag{4.28}$$

which is is the moments of $|D^\alpha \varrho(\cdot, t_k)|$ for $|\alpha| \leq \ell$ (see (4.21) for the multi-index notation). In fact, we have the following result provided that ϱ is smooth enough.

Lemma 4.4. *Suppose Assumptions 2.2 and 4.2 hold. Let $t_k \le T-\tau$. Then,*

$$W_2(\mathcal{G}_\infty(\varrho(\cdot,t_k)), \varrho(\cdot,t_{k+1})) \le g(M_{q,4}^{(k)})\tau^2,$$

for some $q>1$ and nondecreasing function $g(\cdot)$, where $M_{q,4}^{(k)}$ is defined in (4.28).

Proof. For the notational convenience in this proof, we denote, only in this proof,

$$\varrho \equiv \varrho(\cdot, t_k).$$

If ϱ elsewhere is used, we adopt $\varrho(\cdot,t)$ directly.

Step 1– Consider the SDE corresponding to the nonlinear Fokker-Planck equation (4.8):

$$dX = b(X)\,dt + K(\cdot) * \varrho(\cdot,t)(X)\,dt + \sqrt{2}\sigma dW.$$

Denote $\bar{K}(X) := \int_{\mathbb{R}^d} K(X-z)\varrho(z;t_k)\,dz$, then we have

$$dX = b(X)\,dt + \bar{K}(X) + \sqrt{2}\sigma dW + R, \tag{4.29}$$

where, by a similary calculation as in the proof of Lemma 4.1,

$$|R| \le C(M_{q_1,0})(1+|X|^q)\tau,$$

for some $q_1 > 1$. In fact, C depends on the moments of $\varrho(\cdot,t)$ for $t\in[t_k,t_{k+1}]$, which can be controlled by the ones at t_k.

We show that the law of X is close in W_2 to the law generated by the following SDE:

$$d\hat{X} = b(\hat{X})\,dt + \bar{K}(\hat{X}) + \sqrt{2}\sigma dW. \tag{4.30}$$

To do this, we estimate the mean square error between X and $\hat{X}$ under the synchronization coupling (i.e. using the same Brownian motion). In fact,

$$\frac{d}{dt}\mathbb{E}|X-\hat{X}|^2 \le C\mathbb{E}|X-\hat{X}|^2 + C\sqrt{\mathbb{E}|X-\hat{X}|^2}\|R\|.$$

This means

$$\frac{d}{dt}\sqrt{\mathbb{E}|X-\hat{X}|^2} \le C\sqrt{\mathbb{E}|X-\hat{X}|^2} + C(M_{q,0})\tau. \tag{4.31}$$

Denote (recall that $\mathscr{L}$ means the law of a random variable)

$$\tilde{\mathcal{S}}(\varrho) := \mathscr{L}(\hat{X}(\tau)). \tag{4.32}$$

Then, applying Grönwall's lemma on (4.31) yields

$$W_2(\varrho(\cdot,t_{k+1}), \tilde{\mathcal{S}}(\varrho)) \le C(M_{q,0})\tau^2.$$

Step 2– Compare $\tilde{\mathcal{S}}(\varrho)$ with $\mathcal{G}_\infty(\varrho)$.

We compare the law of $\hat{X}$ in (4.30) (i.e. $\tilde{\mathcal{S}}(\varrho)$) with the law of Y^1 (i.e. $\mathcal{G}_\infty(\varrho)$) given by

$$dY^i = b(Y^i)\,dt + \frac{1}{p-1}\sum_{j=1, j\ne i}^{p} K(Y^i - Y^j)\,dt + \sqrt{2}\sigma\,dW^i,\ i=1,\cdots,p, \tag{4.33}$$

with the initial data drawn from $\varrho^{\otimes p}$. The main strategy is to use Lemma 2.2, so we need to estimate the difference of these two distributions and control the second moments of this difference.

It is clear that $\tilde{\mathcal{S}}(\varrho) = e^{\tau\hat{\mathcal{L}}^*}\varrho$, where $\hat{\mathcal{L}}^*$ is given by

$$\begin{aligned}\hat{\mathcal{L}}^*(\rho)(x) &:= -\nabla\cdot\Big([b(x)+\int_{\mathbb{R}^d}K(x-x_2)\rho(x_2)dx_2]\rho(x)\Big)+\sigma^2\Delta_x\rho(x)\\ &= -\int_{\mathbb{R}^d}dx_2\rho(x_2)[\nabla\cdot([b(x)+K(x-x_2)]\rho(x))+\sigma^2\Delta_x\rho(x)].\end{aligned} \tag{4.34}$$

Denote the Fokker-Planck operator for the evolution of $(Y^1,\cdots,Y^p)$ by

$$\bar{\mathcal{L}}^* := -\sum_{i=1}^p\nabla_{x_i}\cdot([b(x_i)+\frac{1}{p-1}\sum_{j:j\neq i}K(x_i-x_j)]\cdot)+\sigma^2\sum_{i=1}^p\Delta_{x_i}.$$

Then, the law of Y^1 at τ is given by

$$\mathcal{G}_\infty(\varrho) = \int_{(\mathbb{R}^d)^{p-1}} e^{\tau\bar{\mathcal{L}}^*}(\prod_{i=1}^p\varrho(x_i))\,dx_2\cdots dx_p. \tag{4.35}$$

First note

$$\tilde{S}(\varrho)(x) = \varrho(x)+\tau\hat{\mathcal{L}}^*\varrho(x)+\frac{1}{2}\int_0^\tau(\tau-s)(\hat{\mathcal{L}}^*)^2e^{s\hat{\mathcal{L}}^*}\varrho\,ds, \tag{4.36}$$

while

$$\begin{aligned}\mathcal{G}_\infty(\varrho)(x_1) &= \int_{(\mathbb{R}^d)^{p-1}}\prod_{i=1}^p\varrho(x_i)\,dx_2\cdots dx_p+\tau\int_{(\mathbb{R}^d)^{p-1}}\bar{\mathcal{L}}^*\prod_{i=1}^p\varrho(x_i)\,dx_2\cdots dx_p\\ &\quad+\frac{1}{2}\int_0^\tau(\tau-s)\int_{(\mathbb{R}^d)^{p-1}}(\bar{\mathcal{L}}^*)^2e^{s\bar{\mathcal{L}}^*}(\prod_{i=1}^p\varrho(x_i))\,dx_2\cdots dx_p\,ds.\end{aligned} \tag{4.37}$$

The first line of (4.37) is reduced to

$$\begin{aligned}&\varrho(x_1)-\tau\int_{(\mathbb{R}^d)^{p-1}}\nabla_{x_1}\cdot\left([b(x_1)+\frac{1}{p-1}\sum_{j:j\neq 1}K(x_1-x_j)]\prod_{i=1}^p\varrho(x_i)\right)dx_2\cdots dx_p\\ &= \varrho(x_1)-\tau\nabla_{x_1}\cdot\Big([b(x_1)+\int_{\mathbb{R}^d}K(x_1-y)\varrho(y)dy]\varrho(x_1)\Big) = \varrho(x_1)+\tau\hat{\mathcal{L}}^*\varrho(x_1).\end{aligned} \tag{4.38}$$

Hence, we find

$$\begin{aligned}&|\tilde{S}(\varrho)(x)-\mathcal{G}_\infty(\varrho)(x)|\\ &\le \frac{1}{2}\tau^2\left[\sup_{0\le s\le\tau}|(\hat{\mathcal{L}}^*)^2e^{s\hat{\mathcal{L}}^*}\varrho|+\sup_{0\le s\le\tau}\left|\int_{(\mathbb{R}^d)^{p-1}}(\bar{\mathcal{L}}^*)^2e^{s\bar{\mathcal{L}}^*}(\prod_{i=1}^p\varrho(x_i))\,dx_2\cdots dx_p\right|\right].\end{aligned} \tag{4.39}$$

Noticing that both $\hat{\mathcal{L}}^*$ and $\bar{\mathcal{L}}^*$ are constant operators, one has by Lemma 4.2 (iii) that for some $q>1$,

$$\int_{\mathbb{R}^d}(1+|x|^2)|\tilde{S}(\varrho)(x)-\mathcal{G}_\infty(\varrho)(x)|\,dx \le C\tau^2 M_{q,4}.$$

To illustrate how this is estimated, we take the second term as an example:

$$\begin{aligned}&\int_{\mathbb{R}^d}(1+|x|^2)\sup_{0\le s\le\tau}\left|\int_{(\mathbb{R}^d)^{p-1}}(\bar{\mathcal{L}}^*)^2e^{s\bar{\mathcal{L}}^*}(\prod_{i=1}^p\varrho(x_i))\,dx_2\cdots dx_p\right|dx\\ &\le \sup_{0\le s\le\tau}\int_{(\mathbb{R}^d)^p}(1+|x_1|^2)|(\bar{\mathcal{L}}^*)^2e^{s\bar{\mathcal{L}}^*}(\prod_{i=1}^p\varrho(x_i))|\,dx_1\cdots dx_p\\ &= \sup_{0\le s\le\tau}\int_{(\mathbb{R}^d)^p}(1+\frac{1}{p}\sum_i|x_i|^2)|(\bar{\mathcal{L}}^*)^2e^{s\bar{\mathcal{L}}^*}(\prod_{i=1}^p\varrho(x_i))|\,dx_1\cdots dx_p\\ &\le \sup_{0\le s\le\tau}\int_{(\mathbb{R}^d)^p}(1+\frac{1}{p}\sum_i|x_i|^2)|(\bar{\mathcal{L}}^*)^2(\prod_{i=1}^p\varrho(x_i))|\,dx_1\cdots dx_p.\end{aligned}$$

This is controlled by $M_{q,4}$. Note that the dependence in τ for the constant $C(\tau)$ in Lemma 4.2 has been omitted since $\tau \lesssim O(1)$.

By Lemma 2.2 (one may take $\delta = 1$; both $\|\hat{\mu}\|$ and M_2 are controlled by $C\tau^2 M_{q,4}$), the claim is then proved.

Lastly, the constants $C(M_{q0})$ and $C(M_{q,4})$ clearly have an upper bound $g(M_{q,4})$ with g nondecreasing, defined on $[0, \infty)$. □

Remark 4.2. *In the proof above, estimating $W_2(\tilde{\mathcal{S}}(\varrho), \mathcal{G}_\infty(\varrho))$ by the mean square error between $\hat{X}$ and Y^1 will not be enough to get the desired $C\tau^2$ bound (see Section 5.2).*

With the key one-step estimate established in Lemma 4.4 above, we can now finish the proof of Theorem 4.1.

Proof of Theorem 4.1. By the semigroup property of $\{\mathcal{G}_\infty^k\}$, one can find easily that

$$W_2(\mathcal{G}_\infty^n(\varrho_0), \varrho(n\tau)) \le \sum_{m=0} W_2(\mathcal{G}_\infty^{n-m+1}(\varrho((m-1)\tau)), \mathcal{G}_\infty^{n-m}(\varrho(m\tau)))$$

By Proposition 4.1, under Assumption 2.2 and Assumption 4.2, one has for $n\tau \le T$ that

$$\begin{aligned}&\sum_{m=1}^{n} W_2(\mathcal{G}_\infty^{n-m+1}(\varrho((m-1)\tau)), \mathcal{G}_\infty^{n-m}(\varrho(m\tau)))\\ &\le \sum_{m=1}^{n} e^{\alpha(n-m)\tau/2} W_2(\mathcal{G}_\infty(\varrho((m-1)\tau)), \varrho(m\tau)).\end{aligned}$$

Combining Proposition 4.4 and Lemma 4.4, $W_2(\mathcal{G}_\infty(\varrho((m-1)\tau)), \varrho(m\tau)) \le C(T)\tau^2$ and thus the claim follows.

Under Assumption 2.3 and Assumption 4.2, using Proposition 4.1, the above estimates can be changed by replacing α with $-(r - 2L)$. Hence, the conclusions follow easily. □

5 Some helpful discussions

In this section, we perform some helpful discussions to deepen the understanding and extend the result to second order interacting particle systems.

5.1 The mean field limit for $\tau \ll 1$

As $\tau \to 0$, the mean field limit of the Random Batch Method tends to (i.e. the limit for $\lim_{\tau\to 0}\lim_{N\to\infty}$) the SDE

$$dY = b(Y)\,dt + \frac{1}{p-1}\sum_{j=1}^{p-1} K(Y - Y_j)\,dt + \sqrt{2}\sigma\,dW, \tag{5.1}$$

with $Y_j \sim \mathscr{L}(Y)$ being i.i.d and $\{Y_j(s_i)\}$'s are independent for different time points s_i. Theorem 4.1 essentially tells us that the law of this SDE obeys the same nonlinear Fokker-Planck equation (1.2) which was satisfied by the law of the following seemingly different SDE

$$dX = b(X)\,dt + \left(\int_{\mathbb{R}^d} K(X - y)\varrho(y,t)\,dy\right) dt + \sqrt{2}\sigma\,dW,\ \ \varrho(x,t)\,dx = \mathscr{L}(X(t)). \tag{5.2}$$

See Fig. 3 for illustration (compare with Fig. 1).

To understand this, we consider for a small but fixed τ, and consider the following SDEs with the densities to compute the forces frozen at t_k:

$$\begin{aligned}d\hat{Y} &= b(Y)\,dt + \frac{1}{p-1}\sum_{j=1}^{p-1} K(\hat{Y} - Y_0^j)\,dt + \sqrt{2}\sigma dW,\ \ Y_0^j \sim \mathscr{L}(Y(t_k)),\\ d\hat{X} &= b(\hat{X})\,dt + \left(\int_{\mathbb{R}^d} K(\hat{X} - y)\varrho(y, t_k)\,dy\right) dt + \sqrt{2}\sigma\,dW.\end{aligned} \tag{5.3}$$

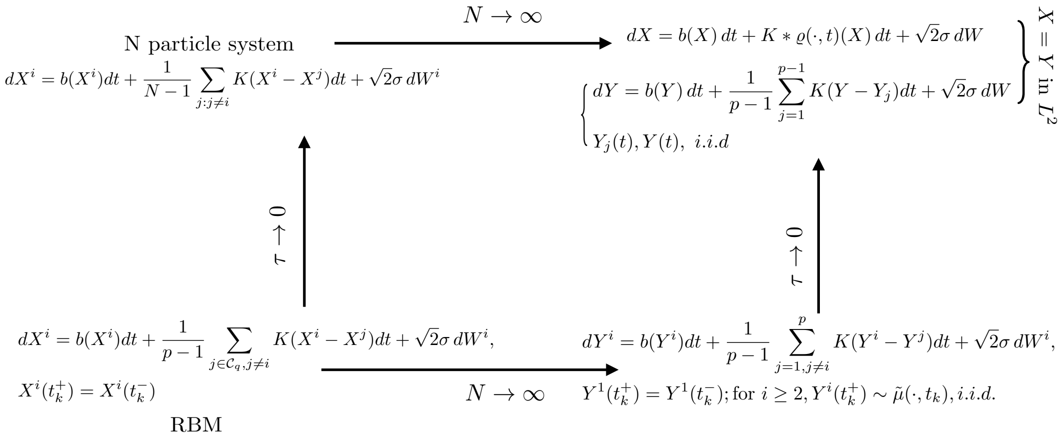

Figure 3: Illustration of the various SDEs in different regime.

The probability density for the former at $t_k + \tau$ is $\int_{\mathbb{R}^d} dy \varrho(y, t_k) e^{\tau \mathcal{L}_y^*} \varrho(\cdot, t_k)$, where

$$\mathcal{L}_y^* = -\nabla \cdot ([b(x) + K(x-y)]\cdot) + \sigma^2 \Delta_x,$$

while the probability density for the latter is $e^{\tau \hat{\mathcal{L}}^*} \varrho(\cdot, t_k)$ with

$$\hat{\mathcal{L}}^* = -\nabla \cdot \left(\left[b(x) + \int_{\mathbb{R}^d} K(x - x_2)\varrho(x_2)dx_2 \right] \cdot \right) + \sigma^2 \Delta_x = \int_{\mathbb{R}^d} dy \varrho(y, t_k) \mathcal{L}_y^*.$$

Clearly, to the leading order, the changing rates of the probability densities are the same.

In Fig. 3 we have made a stronger claim that the X and Y processes in the right-upper corner are equal in L^2, instead of "equal in law", if the Brownian motions W used are the same. To see this, one may compute

$$\frac{d}{dt}\mathbb{E}|X-Y|^2 = 2\mathbb{E}(X-Y)\cdot(b(X)-b(Y)) + 2\mathbb{E}(X-Y)\cdot(K*\varrho(\cdot,t)(X) - \frac{1}{p-1}\sum_{j=1}^{p-1} K(Y-Y_j)).$$

Since $Y_j(t)$ is independent of $Y(t)$ and $X(t)$, one has

$$\mathbb{E}(X-Y)\cdot(K*\varrho(\cdot,t)(X) - \frac{1}{p-1}\sum_{j=1}^{p-1} K(Y-Y_j)) = \mathbb{E}(X-Y)\cdot(K*\varrho(\cdot,t)(X) - K*\bar{\varrho}(\cdot,t)(Y))$$

where $\bar{\varrho}$ is the law of Y. Taking $\tau \to 0$ in Theorem 4.1, $\bar{\varrho} = \varrho$. Hence, one actually has $\frac{d}{dt}\mathbb{E}|X-Y|^2 \leq 2(\beta + L)\mathbb{E}|X-Y|^2$. Hence, $X = Y$ in L^2.

5.2 Regarding the approximation in Lemma 4.4

Usually, the Wasserstein distance was estimated using mean square error. A natural question is therefore whether one can estimate the Wasserstein distance in Lemma 4.4 via the mean square error.

Below, we discuss this by an approximating problem, with the force expressions frozen, i.e., (5.3). Here, we assume the Brownian motions used are the same. The values Y_0^j are drawn from $\varrho(\cdot)^{\otimes p}$.

We compute that

$$\frac{d}{dt}\mathbb{E}|\hat{X} - \hat{Y}|^2 = \mathbb{E}(\hat{X} - \hat{Y}) \cdot (b(\hat{X}) - b(\hat{Y})) + D,$$

where

$$D = \mathbb{E}(\hat{X} - \hat{Y}) \cdot \left(\bar{K}(\hat{X}) - \frac{1}{p-1}\sum_{j=1}^{p-1} K(\hat{Y} - Y_0^j) \right).$$

Clearly, for fixed x,

$$\mathbb{E}\frac{1}{p-1}\sum_{j=1}^{p-1} K(x - Y_0^j) = \bar{K}(x). \tag{5.4}$$

Hence, if $\hat{Y}$ is independent of Y_0^j's, then this term can be controlled as

$$\mathbb{E}(\hat{X}-\hat{Y})\cdot(\bar{K}(\hat{X})-\bar{K}(\hat{Y})) \le C\mathbb{E}|\hat{X}-\hat{Y}|^2.$$

One is thus tempted to believe that even though that $\hat{Y}$ is not independent of Y_0^j, one can do Itô-Taylor expansion and the extra term is small enough, which can yields the desired error.

Unfortunately, if one is going to do the Itô-Taylor expansion in $\hat{Y}$, one may find that $D = O(\tau)$. In fact,

$$\begin{aligned}&(\hat{X}-\hat{Y})\cdot\left(\bar{K}(\hat{X}) - \frac{1}{p-1}\sum_{j=1}^{p-1}K(\hat{Y}-Y_0^j)\right)\\ &= \int_0^t \left(\bar{K}(\hat{X}(s)) - \frac{1}{p-1}\sum_{j=1}^{p-1}K(\hat{Y}(s)-Y_0^j)\right)\cdot\left(\bar{K}(\hat{X}(t)) - \frac{1}{p-1}\sum_{j=1}^{p-1}K(\hat{Y}(t)-Y_0^j)\right)ds.\end{aligned}$$

If we take expectation, the variance of the random force $\frac{1}{p-1}\sum_{j=1}^{p-1}K(x-X_0^j)$ appears, which gives $D = O(\tau)$. Hence, this estimate is not good and the mean square error is only like $\sqrt{\mathbb{E}|\hat{X}-\hat{Y}|^2} = O(\tau)$. This means that the consistency (5.4) brings no benefit for this mean square error!

Intrinsically, the mean square error above is roughly comparable to

$$\int \varrho(z_1)\cdots\varrho(z_j)W_2^2(e^{\tau\hat{\mathcal{L}}^*}\varrho, e^{t\mathcal{L}^*_{z_1,\cdot,z_j}}\varrho)\,dz_1\cdots dz_j.$$

What we care about is $W_2^2\left(e^{\tau\hat{\mathcal{L}}^*}\varrho, \int\varrho(z_1)\cdots\varrho(z_j)e^{t\mathcal{L}^*_{z_1,\cdot,z_j}}\varrho\,dz_1\cdots dz_j\right)$. The former involves the variance introduced by the random force while the latter makes use the consistency (5.4). This is why we did not use the mean square error to prove Lemma 4.4.

5.3 Approximation using weak convergence

The weak convergence is another popular gauge of the convergence of probability $\mathcal{G}_\infty^k(\varrho_0)$ to $\varrho(k\tau)$ [42, 31].

Pick a test function φ, using a consistency condition similar to (4.38), it is not very hard to show

$$\left|\int_{\mathbb{R}^d}\varphi\mathcal{S}(\tau)(\mu)(dx) - \int_{\mathbb{R}^d}\varphi\mathcal{G}_\infty(\mu)(dx)\right| \le C\tau^2, \tag{5.5}$$

for any μ, where we recall $\mathcal{S}(t)$ is the evolution operator for (1.2) . Hence, the one-step error is easy to control for weak convergence. However, the difficulty is to get a certain stability property of the nonlinear dynamics under the weak topology. That means, if two measures are close in the weak topology at some time, then let them evolve under $\mathcal{G}_\infty$ for k times, one needs them to be close. Consider

$$U^n(x) := \int_{\mathbb{R}^d}\varphi\mathcal{G}_\infty^n(\delta(y-x))\,dy.$$

Unlike the linear case (see [17]), it is hard to write U^n as some operator acting on U^{n-1} due to the nonlinearity of $\mathcal{G}_\infty$. Proving the stability of this nonlinear dynamics under weak topology seems challenging, and this is why we chose the Wasserstein metric.

5.4 A remark for second order systems

As shown in [29], the Random Batch Method applied equally well to second order systems on finite time interval. Repeating the proof here, one can show that similar mean field limit holds for second order systems when $t \in [0, T]$. In particular, let us consider the models for swarming and flocking considered in [1]

$$\begin{aligned} \dot{x}_i &= v_i, \\ \dot{v}_i &= \frac{1}{N}\sum_j H_\alpha(x_i, x_j, v_i)(v_j - v_i). \end{aligned} \tag{5.6}$$

Here, $H_\alpha(\cdot,\cdot,\cdot)$ is some function modeling the interactions between particles. The mean field limit of (5.6) for $t \in [0, T]$ takes the following form (rigorous justification needs some assumptions on H_α; see [27])

$$\begin{aligned} &\partial_t f + \nabla_x \cdot (vf) + \nabla_v \cdot (\xi(f) f) = 0, \\ &\xi(f) = \int_{\mathbb{R}^{2d}} H_\alpha(x, y, v)(w - v) f(y, w, t)\, dw dy. \end{aligned} \tag{5.7}$$

Albi and Pareschi in [1] developed some stochastic binary interaction algorithms for the dynamics. The symmetric Nanbu algorithm (Algorithm 4.3) is like Random Batch Method when $p = 2$ and the Random Batch Method can be viewed as generalization of this Nanbu algorithm. When applying Random Batch Method to the particle system and consider $N \gg 1$, the dynamics is expected to be close to the following limiting dynamics:

Algorithm 3 (Mean Field Dynamics of RBM for flocking dynamics (5.6))

1: From t_k to t_{k+1}, the distribution f_k will be transformed into $f_{k+1} = \mathcal{Q}_\infty(f_k)$ as follows.
2: Let $f^{(p)}(\cdots, 0) = f(\cdot, \cdot, t_k)^{\otimes p}$ be a probability measure on $(\mathbb{R}^{2d})^{\otimes p} \cong \mathbb{R}^{2pd}$.
3: Evolve $f^{(p)}$ by time τ according to the following:

$$\begin{aligned} &\partial_t f^{(p)} + \sum_{i=1}^p \nabla_{x_i} \cdot (v_i f^{(p)}) + \sum_{i=1}^p \nabla_{v_i} \cdot (\xi_i f^{(p)}) = 0, \\ &\xi_i = \frac{1}{p-1} \sum_{j: j \neq i} H_\alpha(x_i, x_j, v_i)(v_j - v_i). \end{aligned} \tag{5.8}$$

4: Set

$$f_{k+1} = \mathcal{Q}_\infty(f_k) := \int_{(\mathbb{R}^{2d})^{\otimes(p-1)}} f^{(p)}(\cdot, dy_2, \cdots, dy_p; \cdot, dv_2, \cdots, dv_p; \tau). \tag{5.9}$$

We expect that this nonlinear operator will approximate the nonlinear kinetic equation (5.7). In this sense, we believe the $N \to \infty$ limit of the [1, Algorithm 4.3] will be an analogue of the dynamics $\mathcal{Q}_\infty$ given in Algorithm 3.

6 Conclusions

We first identified and justified in this work the mean field limit of RBM for fixed step size τ. Then, we showed that this mean field limit is close to that of the N particle system, though the chaos arises differently in these two dynamics. The current argument of the mean field limit relies on the fact that two particles are unlikely to be related in RBM when $N \to \infty$ for finite iterations. Hence, this argument cannot given a uniform in τ bound for the speed of the mean field limit. It will be an interesting topic to investigate how mixing and chaos can be created in RBM after two particles in a batch are separated, so that one may give a convergence speed independent of τ.

Acknowledgement

S. Jin was partially supported by the NSFC grant No. 31571071. The work of L. Li was partially sponsored by NSFC 11901389, 11971314, and Shanghai Sailing Program 19YF1421300. L. Li is grateful to Haitao Wang for discussion on fundamental solutions of parabolic equations with unbounded drifts.

A Proof of Proposition 4.2

Step 1–A priori estimates on moments and entropy

We first perform a priori estimates on the moments. Fix $q \geq 2$.

$$\begin{aligned}\partial_t \int_{\mathbb{R}^d} |x|^q \varrho\, dx &= \int_{\mathbb{R}^d} |x|^q \{-\nabla \cdot [(b(x) + K * \varrho)\varrho]\}\, dx + \int_{\mathbb{R}^d} |x|^q \sigma^2 \Delta \varrho\, dx \\ &= \int_{\mathbb{R}^d} q|x|^{q-2} x \cdot b(x) \varrho\, dx + \iint_{\mathbb{R}^d\times\mathbb{R}^d} q|x|^{q-2} x \cdot K(x-y)\varrho(x)\varrho(y)\, dxdy \\ &\quad + \sigma^2 \int_{\mathbb{R}^d} q(q-2+d)|x|^{q-2}\varrho\, dx =: I_1 + I_2 + I_3.\end{aligned}$$

For I_2, one has

$$\begin{aligned}\iint_{\mathbb{R}^d\times\mathbb{R}^d} q|x|^{q-2} x \cdot K(x-y)\varrho(x)\varrho(y)\, dxdy &\leq q \iint_{\mathbb{R}^d\times\mathbb{R}^d} |x|^{q-2} x \cdot K(0)\varrho(x)\varrho(y)\, dxdy \\ &\quad + qL \iint_{\mathbb{R}^d\times\mathbb{R}^d} |x|^{q-1}(|x|+|y|)\varrho(x)\varrho(y)\, dxdy.\end{aligned}$$

By Young's inequality,

$$q \iint_{\mathbb{R}^d\times\mathbb{R}^d} |x|^{q-2} x \cdot K(0)\varrho(x)\varrho(y)\, dxdy \leq \delta \int_{\mathbb{R}^d} |x|^q \varrho\, dx + C(\delta).$$

Also, Young's inequality implies that $|x|^{q-1}|y| \leq \frac{q-1}{q}|x|^q + \frac{1}{q}|y|^q$. Hence,

$$I_2 \leq q(2L+\delta) \int_{\mathbb{R}^d} |x|^q \varrho\, dx + C(\delta).$$

If $q = 2$, I_3 is a constant. Otherwise if $q > 2$, one can use Young's inequality and

$$I_3 \leq \delta \int_{\mathbb{R}^d} |x|^q \varrho\, dx + C(\delta).$$

For I_1, under Assumption 2.2, one has

$$\begin{aligned}I_1 &= \int_{\mathbb{R}^d} q|x|^{q-2} x \cdot (b(x) - b(0))\varrho\, dx + \int_{\mathbb{R}^d} q|x|^{q-2} x \cdot b(0)\varrho\, dx \\ &\leq \beta q \int_{\mathbb{R}^d} |x|^q \varrho\, dx + C \int_{\mathbb{R}^d} |x|^{q-1}\varrho\, dx.\end{aligned}$$

Hence,

$$I_1 + I_2 + I_3 \leq q(\beta + 2L + \delta) \int_{\mathbb{R}^d} |x|^q \varrho\, dx + C(\delta),$$

where the concrete meaning of δ and $C(\delta)$ have changed. Using Grönwall inequality, the moments can be controlled.

Now, we perform a priori estimates on the entropy. Multiply $1 + \log \varrho$ on both sides and integrate:

$$\frac{d}{dt} \int_{\mathbb{R}^d} \varrho \log \varrho\, dx = -\int_{\mathbb{R}^d} \varrho(x) \nabla \cdot (b(x) + (K * \varrho)(x))\, dx - 4\sigma^2 \int_{\mathbb{R}^d} |\nabla \sqrt{\varrho}|^2\, dx.$$

By the moment control, the first term is bounded on $[0, T]$. Hence, the entropy can be controlled.

As a remark, in the case $\sigma = 0$, ϱ could be zero at some points. In this case $1 + \log \varrho$ is not a good test function. This issue will be explained further in Step 2.

Step 2–Existence in $L^\infty(0, T; L^1(\mathbb{R}^d)) \cap C([0, T]; \mathbf{P}(\mathbb{R}^d))$

Take a smooth function $\chi \in C_c[0, \infty)$ that is 1 in $[0, 1]$ and zero on $[2, \infty)$. Consider the following approximating equation

$$\begin{aligned}
&\partial_t \rho_N = -\nabla \cdot (b(x)\chi(x/N)\rho_N) - \nabla \cdot (\rho_N (K * \rho_N)) + \Delta \rho_N, \\
&\varrho_N|_{t=0} = \varrho_0.
\end{aligned}$$

Now, $b(x)\chi(x/N)$ and K are Lipschitz functions and $b(x)\chi(x/N)$ is bounded (compactly supported). The existence of a smooth solution is clear (see, for example, Appendix A in [8]). Performing similar estimates as in Step 1, we have

$$\sup_N \sup_{0 \le t \le T} \int_{\mathbb{R}^d} |x|^2 \varrho_N \, dx \le C(T)$$

and

$$\sup_N \sup_{0 \le t \le T} \int_{\mathbb{R}^d} \varrho_N \log \varrho_N \, dx \le C(T).$$

Note that for the entropy, the zeros of ϱ_N may make $1 + \log(\varrho_N)$ an invalid test function. We instead multiply

$$\frac{\varrho_N}{\varrho_N + \epsilon} + \log(\varrho_N + \epsilon)$$

as the test function for $\epsilon > 0$. Then, the left hand side becomes $\frac{d}{dt} \int \varrho_N \log(\varrho_N + \epsilon) \, dx$ (note that $\epsilon \to \varrho \log(\varrho + \epsilon)$ is non-decreasing so later one can take $\epsilon \to 0$ to get desired entropy control). For the right hand side, we note

$$\nabla \left[\frac{\varrho_N}{\varrho_N + \epsilon} + \log(\varrho_N + \epsilon) \right] = \frac{(\varrho_N + 2\epsilon)\nabla \varrho_N}{(\varrho_N + \epsilon)^2}.$$

For the transport term,

$$b(x)\chi(\frac{x}{N}) \frac{\varrho_N(\varrho_N + 2\epsilon)\nabla \varrho_N}{(\varrho_N + \epsilon)^2} = (b(x)\chi(x/N)) \cdot \nabla \varrho_N - \epsilon^2 (b(x)\chi(x/N)) \cdot \nabla(\frac{1}{\varrho_N + \epsilon}).$$

Doing integration by parts and sending $\epsilon \to 0$ first, the second term here will vanish. Through this way, a prior estimate on the entropy can be justified for this approximating sequence.

The moment estimates imply that $\{\varrho_N \, dx\}$ is tight while the entropy estimates imply that $\{\varrho_N\}$ is uniformly integrable. By Dunford-Pettis theorem, ϱ_N converges weakly to some $\varrho \in L^1_{loc}([0, T] \times \mathbb{R}^d)$ and $\varrho \, dx \in C([0, T]; \mathbf{P}(\mathbb{R}^d))$. Moreover, with the moment control and the uniform integrability

$$\int_{\mathbb{R}^d} K(x - y)\varrho_N(y) \, dy \to \int_{\mathbb{R}^d} K(x - y)\varrho(y) \, dy$$

pointwise and actually uniformly on compact sets. With this, then one can easily verify that ϱ is a desired weak solution, with the corresponding moment control. This will further imply that $\varrho \in L^\infty([0, T]; L^1(\mathbb{R}^d))$.

Step 3–Uniqueness and smoothness of the solution

We now aim to prove the uniqueness. We divide this step into two sub-steps.

Step 3.1–The weak solution is a strong solution

Let ϱ be such a weak solution with

$$\sup_{0\le t\le T}\int_{\mathbb{R}^d}|x|\varrho\,dx < C(T).$$

Then, $\bar{K}(x,t) := K * \varrho$ is a smooth function (since K is smooth) and

$$|\bar{K}(0)| \le \left|\int_{\mathbb{R}^d} K(x)\varrho(x)\,dx\right| \le |K(0)| + LC(T).$$

Moreover, it is easy to see that $\bar{K}(x,t)$ is also Lipschitz with the Lipschitz constant bounded by L.

We claim that for a given ϱ, the solution to

$$\begin{aligned}&\partial_t u = -\nabla\cdot(b(x)u + \bar{K}(x,t)u) + \sigma^2\Delta u,\\ &u|_{t=0} = \varrho_0,\end{aligned}$$

is unique and thus must be ϱ. In fact, the existence can be justified by the following SDE as its law is a weak solution

$$dX = (b(X) + \bar{K}(X,t))\,dt + \sqrt{2}\sigma\,dW,\ \ X_0 \sim \varrho_0\,dx.$$

For the well-posedness of such SDEs, one can refer to [38, Chap 2, Theorem 3.5], and also see a recent work with weaker assumptions [47]. Regarding the uniqueness, one considers the difference of two such solutions $u_i, i = 1, 2$

$$\partial_t(u_1 - u_2) = -\nabla\cdot([b(x) + \bar{K}(x,t)](u_1 - u_2)) + \sigma^2\Delta(u_1 - u_2).$$

We then multiply $h_\epsilon(u_1 - u_2) := h((u_1 - u_2)/\epsilon)$ on both sides and take integral. Here, $h(\cdot)$ is an odd function that increases monotonely from -1 to 1 on $[-1, 1]$. It is 1 on $[1,\infty)$. Hence, $h(\cdot/\epsilon)$ is some approximation for the sign function.

Then,

$$\frac{d}{dt}\int_{\mathbb{R}^d} H_\epsilon(u_1 - u_2)\,dx \le \int_{\mathbb{R}^d} h'\left(\frac{u_1 - u_2}{\epsilon}\right)\frac{u_1 - u_2}{\epsilon}(b(x) + \bar{K}(x,t))\cdot\nabla(u_1 - u_2)\,dx,$$

where $H_\epsilon(u) = \int_0^u h_\epsilon(s)\,ds$. The right hand side goes to zero when $\epsilon \to 0$, because $h'(\frac{u_1-u_2}{\epsilon})\frac{u_1-u_2}{\epsilon}$ is bounded and nonzero only on $|u_1 - u_2| \le \epsilon$. Also, $H_\epsilon(u_1 - u_2) \to |u_1 - u_2|$ as $\epsilon \to 0$. Hence, the claim is shown and thus

$$u = \varrho.$$

By the theory of the *linear* PDEs, $u = \varrho$ is in fact a strong solution and smooth. For the general theory of linear parabolic equations, one may refer to [20].

Step 3.2–The uniqueness of the nonlinear Fokker-Planck equation

For the uniqueness of the nonlinear Fokker-Planck equation, we cannot use the technique in Step 3.1 as we show uniqueness for the linear PDE, as the term $K * \varrho$ involves the solution ϱ itself. Also, the classical Dobrushin's estimate [14, 22] cannot be used because the flow map is not well-defined before we show the uniqueness of ϱ.

Instead, we use the interacting particle system for mean-field limit and show that any weak solution is close to the one marginal distribution of the N-particle system. This then will result in the uniqueness.

Fix *any* weak solution of the nonlinear Fokker-Planck equation. Consider the following SDEs

$$dX^i = b(X^i)\,dt + (K * \varrho)(X^i)\,dt + \sqrt{2}\sigma\,dW^i,\ \ i = 1,\cdots,N. \tag{A.1}$$

According to the argument in Step 3.1, the law of each X^i is exactly the weak solution ϱ used to convolve with K. Moreover, these X^i's are independent.

Now, consider the interacting praticle system

$$dY^i = b(Y^i)\,dt + \frac{1}{N-1}\sum_{j:j\neq i} K(Y^i - Y^j)\,dt + \sqrt{2}\sigma\,dW^i,\ i = 1,\cdots,N. \tag{A.2}$$

The next step is to use the technique in the proof of [10, Theorem 3.1]. We compute for fixed i,

$$\frac{1}{2}\frac{d}{dt}\mathbb{E}|X^i - Y^i|^2 = \mathbb{E}(X^i - Y^i)\cdot(b(X^i) - b(Y^i)) + \mathbb{E}(X^i - Y^i)\cdot\left(\bar{K}(X^i,t) - \frac{1}{N-1}\sum_{j:j\neq i} K(Y^i - Y^j)\right). \tag{A.3}$$

The first term is controlled by $\beta\mathbb{E}|X^i - Y^i|^2$. The second term is split as

$$\begin{aligned}
&\mathbb{E}(X_i - Y_i)\cdot\left(\bar{K}(X_i,t) - \frac{1}{N-1}\sum_{j:j\neq i} K(Y_i - Y_j)\right)\\
&= \mathbb{E}(X_i - Y_i)\cdot(\bar{K}(X_i,t) - \frac{1}{N-1}\sum_{j:j\neq i} K(X_i - X_j))\\
&+ \mathbb{E}(X_i - Y_i)\cdot(\frac{1}{N-1}\sum_{j:j\neq i} K(X_i - X_j) - \frac{1}{N-1}\sum_{j:j\neq i} K(Y_i - Y_j)) =: D_1 + D_2.
\end{aligned}$$

The term D_2 is easily controlled by $2L\mathbb{E}|X_i - Y_i|^2$ by the exchangeability. For D_1, one can control it as

$$D_1 \le \sqrt{\mathbb{E}|X_i - Y_i|^2}\sqrt{\mathbb{E}\left|\bar{K}(X_i,t) - \frac{1}{N-1}\sum_{j:j\neq i} K(X_i - X_j)\right|^2}.$$

However,

$$\begin{aligned}
&\mathbb{E}\left|\bar{K}(X_i,t) - \frac{1}{N-1}\sum_{j:j\neq i} K(X_i - X_j)\right|^2\\
&= \frac{1}{(N-1)^2}\sum_{j,k:j\neq i,k\neq i}\mathbb{E}(\bar{K}(X_i,t) - K(X_i - X_j))(\bar{K}(X_i,t) - K(X_i - X_k)).
\end{aligned}$$

By independence, the terms for $j \neq k$ are zero. Hence, only $N-1$ terms will survive. This means

$$D_1 \le \sqrt{\mathbb{E}|X_i - Y_i|^2}\frac{C_1(T,\varrho)}{\sqrt{N-1}}.$$

Moreover, $C_1(T,\varrho)$ will have an upper bound that is independent of T if Assumption 2.3 holds.

By Grönwall's inequality,

$$\sqrt{\mathbb{E}|X_i - Y_i|^2} \le C(T,\varrho)\frac{1}{\sqrt{N-1}}.$$

Hence, for any two weak solutions ϱ_1, ϱ_2, we have

$$\sup_{0\le t\le T} W_2(\varrho_1,\varrho_2) \le [C(T,\varrho_1) + C(T,\varrho_2)]\frac{1}{\sqrt{N-1}}.$$

Taking $N \to \infty$ yields the uniqueness of the solutions to the nonlinear Fokker-Planck equation.

Step 4–Strong confinement
Under Assumption 2.3, one in fact has

$$I_1 + I_2 + I_3 \le q(-r + 2L + \delta) \int_{\mathbb{R}^d} |x|^q \varrho \, dx + C(\delta).$$

The assertions about moments have then been proved with application of Grönwall's inequality.

Under this condition, the estimate of D_1 term in Step 3 can also be independent of T, because of this uniform moment control. Hence, the mean field limit can be uniform in T.

Lastly, to show the convergence of ϱ as $t \to \infty$, we consider two different initial data $\varrho_{j,0}$ where $j = 1, 2$. Then, one can consider (A.2) with these two initial data. Pick the coupling between $Y_1^i(0)$ and $Y_2^i(0)$ such that

$$\mathbb{E}|Y_1^i(0) - Y_2^i(0)|^2 \le W_2(\varrho_{1,0}, \varrho_{2,0})^2 + \epsilon, \ \forall i = 1, \cdots, N.$$

Then, by similar computation,

$$\frac{d}{dt}\mathbb{E}|Y_1^i(t) - Y_2^i(t)|^2 \le 2(-r + 2L)\mathbb{E}|Y_1^i(t) - Y_2^i(t)|^2.$$

Taking $N \to \infty$, $\mathscr{L}(Y_j^i(t)) \to \varrho_j(t), j = 1, 2$. Hence, the evolutional nonlinear semigroup for the nonlinear Fokker-Planck equation is a contraction

$$W_2(\varrho_1(t), \varrho_2(t)) \le W_2(\varrho_{1,0}, \varrho_{2,0}) e^{-(r-2L)t}.$$

Thus, the last claim follows.

B Proof of Lemma 4.3

Since $\sigma > 0$, without loss of generality, we will assume

$$\sigma \equiv 1.$$

We first fix $s \ge 0$. Consider the trajectory determined by

$$\partial_t Z(t; y, s) = b(Z, t), \ Z(s; y, s) = y. \tag{B.1}$$

Then, one has

$$\frac{1}{2}\frac{d}{dt}|Z|^2 \le \beta_1 |Z|^2 + C|Z|$$

as $b(0, t)$ is bounded. Hence,

$$\frac{d}{dt}|Z| \le \beta_1 |Z| + C.$$

This means

$$|Z| \le |y| e^{\beta_1(t-s)} + C \int_s^t e^{\beta_1(t-s)} \, ds. \tag{B.2}$$

Moreover, (4.16) implies that

$$v \cdot \nabla b_1(x, t) \cdot v \le \beta_1 |v|^2, \ \forall v, x \in \mathbb{R}^d, t \ge 0.$$

Consequently

$$|\nabla_y Z| \le \sqrt{d} e^{\beta_1(t-s)}, \tag{B.3}$$

uniform in y, where $|A| := \sqrt{\sum_{ij} A_{ij}^2}$ is the matrix Frobenius norm.

Assume without loss of generality $|x| \ge |y|$. Clearly,

$$|b_1(x,t) - b_1(y,t)| \le |x-y| \left| \int_0^1 \nabla b_1(x\theta + y(1-\theta), t) d\theta \right|.$$

Due to the assumption of polynomial growth of derivatives of b_1,

$$|\nabla b_1(xz + y(1-z), t)| \le C(1 + |x\theta + y(1-\theta)|^q).$$

If $|y| \le \frac{1}{2}|x|$, then $|x\theta + y(1-\theta)| \le \frac{3}{2}|x| \le 3|x-y|$. Otherwise, we bound this by a polynomial of $|y|$ directly. Hence,

$$|b_1(x,t) - b_1(y,t)| \le \min(P_1(|x|), P_1(|y|))|x-y| + P_2(|x-y|)|x-y| \tag{B.4}$$

for some polynomials P_1, P_2.

We denote

$$\Phi_0(x,t;y,s) := \frac{1}{(2\pi(t-s))^{d/2}} \exp\left(-\frac{|x - Z(t;y,s)|^2}{2(t-s)}\right). \tag{B.5}$$

Below, we establish an important lemma indicating that Φ_0 is the main term of Φ, and Lemma 4.3 will follow easily.

Lemma B.1. *It holds that*

$$\Phi(x,t;y,s) = \Phi_0(x,t;y,s) + u(x,t;y,s), \tag{B.6}$$

where u satisfies

$$\int_{\mathbb{R}^d} (1 + |x|^q)|\nabla_y u|\, dx \le h(t-s)P(|y|), \tag{B.7}$$

for some polynomial $P(\cdot)$, some nondecreasing function $h(\cdot)$.

Moreover, if $\beta_1 < 0$, $h(t-s)$ can be taken as

$$h(t-s) = Ce^{-\delta_1(t-s)} \tag{B.8}$$

for some $\delta_1 > 0$.

Proof. It is not hard to verify

$$\partial_t \Phi_0 + \nabla_x \cdot (b_1(x,t)\Phi_0) - \Delta_x \Phi_0 = \nabla_x \cdot ([b_1(x,t) - b(Z,t)]\Phi_0). \tag{B.9}$$

Hence, letting $u = \Phi - \Phi_0$, one finds

$$\begin{aligned} &\partial_t u + \nabla_x \cdot (b_1(x,t)u) - \Delta_x u = -\nabla_x \cdot ([b_1(x,t) - b(Z,t)]\Phi_0), \\ &u|_{t=s} = 0. \end{aligned} \tag{B.10}$$

Letting

$$v := \partial_{y_i} u,$$

one has

$$\begin{aligned} &\partial_t v + \nabla_x \cdot (b_1(x,t)v) - \Delta_x v = R, \\ &u|_{t=s} = 0, \end{aligned} \tag{B.11}$$

where

$$R = \nabla_x \cdot b_1(x,t)\nabla\Phi_0 \cdot \partial_{y_i} Z + \partial_{y_i} Z \cdot \nabla b_1(x,t) \cdot \nabla\Phi_0 + (b_1(x,t) - b_1(Z,t)) \cdot \nabla^2\Phi_0 \cdot \partial_{y_i} Z.$$

Writing $\nabla_x \cdot b_1(x,t) = [\nabla_x \cdot b_1(x,t) - \nabla \cdot b_1(Z,t)] + \nabla \cdot b_1(Z,t)$, it is not hard to see (using also (B.3) and (B.4))

$$|R| \le P(|Z|) \frac{1}{(t-s)^{(d+1)/2}} \exp\left(-\frac{\gamma|x-Z|^2}{2(t-s)}\right) e^{\beta_1(t-s)}$$

for some polynomial P and $\gamma \in (0,1)$.

We then find

$$v = \int_s^t S_{\lambda,t} R\, d\lambda.$$

Below, we use $h_i(\cdot)$ to denote some nondecreasing functions defined on $[0,\infty)$. By Lemma 4.2, one has

$$\int_{\mathbb{R}^d} (1+|x|^q)|v|\, dx \le h_1(t-s)\int_s^t \int_{\mathbb{R}^d} (1+|x|^q)|R(x,\lambda)|\, dx d\lambda.$$

Clearly,

$$\int_{\mathbb{R}^d} (1+|x|^q)\frac{1}{(t-s)^{(d+1)/2}}\exp(-\frac{\delta|x-Z|^2}{2(t-s)})\, dx \le C\frac{1+(t-s)^{q/2}}{\sqrt{t-s}}(1+|Z|^q).$$

Moreover, by the stability of trajectory of Z (B.2), $P(|Z|) \le h_2(t-s)P(|y|)$. Hence,

$$\int_{\mathbb{R}^d} (1+|x|^q)|v|\, dx \le h_3(t-s)P(|y|)\int_s^t \frac{1}{\sqrt{\lambda - s}} d\lambda.$$

If $\beta_1 < 0$, we consider $t \ge s+1$ and

$$v = \int_s^t S_{\lambda,t} R\, d\lambda = S_{(t+s)/2,t}\int_s^{\frac{t+s}{2}} S_{\lambda,(t+s)/2} R\, d\lambda + \int_{(t+s)/2}^t S_{\lambda,t} R\, d\lambda. \tag{B.12}$$

The second term is like

$$\begin{aligned}\int_{\mathbb{R}^d} (1+|x|^q)|v|\, dx &\le C\int_{(s+t)/2}^t \int_{\mathbb{R}^d} (1+|x|^q)|R(x,\lambda)|\, dx d\lambda \\ &\le C\int_{(s+t)/2}^t e^{\beta_1(\lambda - s)}P(|Z|)\int_{\mathbb{R}^d} \frac{1+|x|^q}{(t-s)^{(d+1)/2}}\exp(-\frac{\delta|x-Z|^2}{2(t-s)})\, dx d\lambda.\end{aligned}$$

This is easily controlled by $P(|y|)e^{-\delta_1(t-s)}$ for some polynomial P and $\delta_1 > 0$ (recall (B.2)).

For the first term in (B.12), we note $\int_s^{\frac{t+s}{2}} S_{\lambda,(t+s)/2} R\, d\lambda \in L^1$, and

$$\int_{\mathbb{R}^d}\int_s^{\frac{t+s}{2}} S_{\lambda,(t+s)/2} R\, d\lambda\, dx = 0$$

since $\int R(x,\lambda)\, dx = 0$ for all λ. Hence, statement (ii) in Lemma 4.2 implies that

$$\begin{aligned}&\int_{\mathbb{R}^d} (1+|x|^q)\left|S_{(t+s)/2,t}\int_s^{\frac{t+s}{2}} S_{\lambda,(t+s)/2} R\, d\lambda\right|\, dx \\ &\qquad\le e^{-\delta(t-s)/2}P\left(M_{q_1}\left(\left|\int_s^{\frac{t+s}{2}} S_{\lambda,(t+s)/2} R\, d\lambda\right|\right)\right).\end{aligned}$$

For the inside

$$M_{q_1}\left(\left|\int_s^{\frac{t+s}{2}} S_{\lambda,(t+s)/2} R\, d\lambda\right|\right) \le C\int_s^{(t+s)/2}\int_{\mathbb{R}^d} (1+|x|^{q_1})|R|\, dx d\lambda,$$

where C is independent of time as $\beta_1 < 0$. As has been proved, the integral here is controlled by products of polynomials of $|y|$, $|t-s|$. Hence, the first term is also controlled similarly. ☐

As soon as Lemma B.1 is proved, Lemma 4.3 is very straightforward, as $|\nabla_y Z| \le Ce^{\beta_1(t-s)}$.

References

[1] G. Albi and L. Pareschi. Binary interaction algorithms for the simulation of flocking and swarming dynamics. *Multiscale Modeling & Simulation*, 11(1):1–29, 2013.

[2] V. Barbu and M. Röckner. From nonlinear Fokker-Planck equations to solutions of distribution dependent SDE. *arXiv preprint arXiv:1808.10706*, 2018.

[3] J.-D. Benamou and Y. Brenier. A computational fluid mechanics solution to the Monge-Kantorovich mass transfer problem. *Numer. Math.*, 84(3):375–393, 2000.

[4] 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(2):651–681, 2012.

[5] C. K. Birdsall and A. B. Langdon. *Plasma physics via computer simulation*. CRC press, 2004.

[6] L. Bottou. Online learning and stochastic approximations. *On-line learning in neural networks*, 17(9):142, 1998.

[7] E. Carlen, P. Degond, and B. Wennberg. Kinetic limits for pair-interaction driven master equations and biological swarm models. *Mathematical Models and Methods in Applied Sciences*, 23(07):1339–1376, 2013.

[8] J. A. Carrillo, R. J. McCann, and C. Villani. Kinetic equilibration rates for granular media and related equations: entropy dissipation and mass transportation estimates. *Revista Matematica Iberoamericana*, 19(3):971–1018, 2003.

[9] J. A. Carrillo, L. Pareschi, and M. Zanella. Particle based gPC methods for mean-field models of swarming with uncertainty. *Communications in Computational Physics*, 25(2), 2019.

[10] P. Cattiaux, A. Guillin, and F. Malrieu. Probabilistic approach for granular media equations in the non-uniformly convex case. *Probability theory and related fields*, 140(1-2):19–40, 2008.

[11] Y.-P. Choi, S.-Y. Ha, and S.-B. Yun. Complete synchronization of Kuramoto oscillators with finite inertia. *Physica D: Nonlinear Phenomena*, 240(1):32–44, 2011.

[12] F. Cucker and S. Smale. Emergent behavior in flocks. *IEEE Transactions on automatic control*, 52(5):852–862, 2007.

[13] P. Degond, J.-G. Liu, and R. L. Pego. Coagulation–fragmentation model for animal group-size statistics. *Journal of Nonlinear Science*, 27(2):379–424, 2017.

[14] R. L. Dobrushin. Vlasov equations. *Functional Analysis and Its Applications*, 13(2):115–123, 1979.

[15] A. Durmus, A. Eberle, A. Guillin, and R. Zimmer. An elementary approach to uniform in time propagation of chaos. *arXiv preprint arXiv:1805.11387*, 2018.

[16] R. Durrett. *Probability: Theory and Examples*. Cambridge Series in Statistical and Probabilistic Mathematics. Cambridge University Press, 4 edition, 2010.

[17] Y. Feng, L. Li, and J.-G. Liu. Semi-groups of stochastic gradient descent and on-line principal component analysis: properties and diffusion approximations. *Commun. Math. Sci.*, 16(3), 2018.

[18] N. Fournier, M. Hauray, and S. Mischler. Propagation of chaos for the 2D viscous vortex model. *J. Eur. Math. Soc.*, 16(7):1423–1466, 2014.

[19] D. Frenkel and B. Smit. *Understanding molecular simulation: from algorithms to applications*, volume 1. Elsevier, 2001.

[20] A. Friedman. *Partial differential equations of parabolic type*. Courier Dover Publications, 2008.

[21] 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. *Reviews of Modern Physics*, 68(1):13, 1996.

[22] François Golse. On the dynamics of large particle systems in the mean field limit. In *Macroscopic and large scale phenomena: coarse graining, mean field limits and ergodicity*, pages 1–144. Springer, 2016.

[23] A. Granas and J. Dugundji. *Fixed point theory*. Springer Science and Business Media, 2013.

[24] S.-Y. Ha and Z. Li. Complete synchronization of Kuramoto oscillators with hierarchical leadership. *Communications in Mathematical Sciences*, 12(3):485–508, 2014.

[25] S.-Y. Ha and J.-G. Liu. A simple proof of the Cucker-Smale flocking dynamics and mean-field limit. *Commun. Math. Sci.*, 7(2):297–325, 2009.

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

[27] P. E. Jabin. A review of the mean field limits for Vlasov equations. *Kinetic and Related Models*, 7(4):661–711, 2020.

[28] S. Jin, L. Li, and J.-G. Liu. Convergence of Random Batch Method for interacting particles with disparate species and weights. *arXiv preprint arXiv:2003.11257*, 2020.

[29] S. Jin, L. Li, and J.-G. Liu. Random Batch methods (RBM) for interacting particle systems. *Journal of Computational Physics*, 400:108877, 2020.

[30] R. Jordan, D. Kinderlehrer, and F. Otto. The variational formulation of the Fokker–Planck equation. *SIAM J. Math. Anal.*, 29(1):1–17, 1998.

[31] P. E. Kloeden and E. Platen. *Numerical solution of stochastic differential equations*, volume 23. Springer Science & Business Media, 2013.

[32] D. Ko and E. Zuazua. Model predictive control with Random Batch Methods for a guiding problem. *arXiv preprint arXiv:2004.14834*, 2020.

[33] O. A. Ladyženskaja, V. A. Solonnikov, and N. N. Ural'ceva. *Linear and quasi-linear equations of parabolic type*, volume 23. American Mathematical Soc., 1988.

[34] J.-M. Lasry and P.-L. Lions. Mean field games. *Japanese journal of mathematics*, 2(1):229–260, 2007.

[35] 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.*, 2020.

[36] L. Li, J.-G. Liu, and Y. Tang. A direct simulation approach for the Poisson-Boltzmann equation using the Random Batch Method. *arXiv preprint arXiv:2004.05614*, 2020.

[37] L. Li, J.-G. Liu, and P. Yu. On mean field limit for Brownian particles with Coulomb interaction in 3D. *J. Math. Phys.*, 60(111501), 2019.

[38] X. Mao. *Stochastic Differential Equations and Applications*. Horwood, Chichester, UK, 1997.

[39] J. C. Mattingly, A. M. Stuart, and D. J. Higham. Ergodicity for SDEs and approximations: locally lipschitz vector fields and degenerate noise. *Stochastic processes and their applications*, 10(2):185–232, 2002.

[40] H. P. McKean. A class of Markov processes associated with nonlinear parabolic equations. *Proceedings of the National Academy of Sciences of the United States of America*, 56(6):1907, 1966.

[41] S. Meyn and R. L. Tweedie. *Markov chains and stochastic stability.* Cambridge University Press, Cambridge, second edition, 2009. With a prologue by Peter W. Glynn.

[42] G. N. Milstein and M. V. Tretyakov. *Stochastic numerics for mathematical physics.* Springer Science & Business Media, 2013.

[43] S. Motsch and E. Tadmor. Heterophilious dynamics enhances consensus. *SIAM review*, 56(4):577–621, 2014.

[44] H. Robbins and S. Monro. A stochastic approximation method. *The annals of mathematical statistics*, pages 400–407, 1951.

[45] F. Santambrogio. *Optimal transport for applied mathematicians*, volume 87. Springer, 2015.

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

[47] D. Trevisan. Well-posedness of multidimensional diffusion processes with weakly differentiable coefficients. *Electronic Journal of Probability*, 21, 2016.

[48] T. Vicsek, A. Czirók, E. Ben-Jacob, I. Cohen, and O. Shochet. Novel type of phase transition in a system of self-driven particles. *Physical review letters*, 75(6):1226, 1995.

[49] C. Villani. *Topics in optimal transportation.* Number 58. American Mathematical Soc., 2003.

[50] 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)*, pages 681–688, 2011.