Random Batch Methods (RBM) for interacting particle systems

Shi Jin [*1], Lei Li[†2], and Jian-Guo Liu[‡3]

[1,2]School of Mathematical Sciences, Institute of Natural Sciences, MOE-LSC, Shanghai Jiao Tong University, Shanghai, 200240, P. R. China.
[3]Department of Mathematics and Department of Physics, Duke University, Durham, NC 27708, USA.

Abstract

We develop Random Batch Methods for interacting particle systems with large number of particles. These methods use small but random batches for particle interactions, thus the computational cost is reduced from $O(N^2)$ per time step to $O(N)$, for a system with N particles with binary interactions. On one hand, these methods are efficient Asymptotic-Preserving schemes for the underlying particle systems, allowing N-independent time steps and also capture, in the $N \to \infty$ limit, the solution of the mean field limit which are nonlinear Fokker-Planck equations; on the other hand, the stochastic processes generated by the algorithms can also be regarded as new models for the underlying problems. For one of the methods, we give a particle number independent error estimate under some special interactions. Then, we apply these methods to some representative problems in mathematics, physics, social and data sciences, including the Dyson Brownian motion from random matrix theory, Thomson's problem, distribution of wealth, opinion dynamics and clustering. Numerical results show that the methods can capture both the transient solutions and the global equilibrium in these problems.

1 Introduction

In natural and social sciences, there are many collective behaviors resulted from a huge number of interrelated individuals. Examples include swarming or synchronization described by the Vicsek model [1], flocking in school of fishes [2], groups of birds [3], chemotaxis of bacteria [4], consensus clusters in opinion dynamics [5], to name a few.

We are interested in first order systems of the following form

$$dX^i = -\nabla V(X^i)\,dt + \sum_{k:k\le J} \alpha_{N,k} \sum_{\mathscr{C}\subset\{1,\dots,N\}:i\in\mathscr{C},|\mathscr{C}|=k} K^i(\mathscr{C})\,dt + \sigma dB^i,\ i=1,\dots,N, \tag{1.1}$$

where J is independent of N and $\alpha_{N,k}$ are some constants. In other words, the interaction acting on each particle is the superposition of the ones for which the number of particles involved is independent of N. Here, $X^i \in \mathbb{R}^d$ are the labels for particles, $-\nabla V(\cdot)$ is some given external field, $K^i(\cdot)$'s are the interaction kernels and $\{B^i\}$'s are independent standard Brownian motions. Such systems could be the overdamped limit of Langevin equations, where dX^i comes from the friction term so that X^i is the location of the ith particle and the terms on the right hand side are forces. Of course, X^i can also have other interpretations

[*]shijin-m@sjtu.edu.cn
[†]leili2010@sjtu.edu.cn
[‡]jliu@phy.duke.edu

depending on applications, like the velocity [3], the opinion or wealth (see section 5). We will loosely call X^i's the locations of particles, $-\nabla V$ the external force and $K(\cdot)$ the interacting forces for convenience in this paper, unless there are clear meanings. If $\sigma = 0$, there is no diffusion effect. If $\sigma > 0$, we have diffusion, and the equation is a stochastic differential equation (SDE) in Itô's sense [6]. We will mostly focus on $J = 2$, so that the systems are of binary interactions (i.e. each particle interacts with others separately). System (1.1) in the case of binary interaction can be written as

$$dX^i = -\nabla V(X^i)\,dt + \frac{1}{N-1}\sum_{j:j\neq i} K(X^i - X^j)\,dt + \sigma dB^i,\ i = 1,\dots,N. \tag{1.2}$$

We will present our algorithms and analysis for binary interactions (system (1.2)). Extensions to (1.1) are straightforward, though more involved (see section 2.3 and Remark 3.2). Although there are examples of many body interactions like Kac interaction in spin glasses [7, 8] for (1.1), we remark that the binary interaction is much more important and ubiquitous in nature, like Coulomb's interaction between electrons and nuclei, due to the superposition principles for forces. As another remark, we have assumed the additive noise for simplicity to illustrate our methodology. However, the discussion in this paper also applies for multiplicative noise (i.e. σ depends on x). (We in fact have one such example in Section 5.2.1.)

There may be two things we care about regarding (1.1) and (1.2), depending on applications. The first thing is the dynamics, for example, in the opinion dynamics, we would like to know how the consensus of opinions are developed. Another thing is the equilibrium distribution of the N particles. If N is large, simulation of (1.1) and (1.2) is very expensive since for each time step during the evolution, the computational cost is of $O(N^J)$. One possible method for studying large and complex systems where small individuals interact with each other is the mean field approximation [9, 10, 11]. In this approximation, the effect of surrounding particles is approximated by a consistent averaged force field so that one obtains a one body problem. For example, regarding (1.2) that we will focus on, in the $N \to \infty$ regime, the distribution μ of the particles formally satisfies the following nonlinear Fokker-Planck equation (see [12, 13, 14, 15, 16, 17, 18])

$$\partial_t \mu = \nabla \cdot (\nabla V(x)\mu) - \nabla \cdot (\mu K * \mu) + \frac{1}{2}\sigma^2 \Delta \mu. \tag{1.3}$$

It is expected that μ is close to the empirical measure for (1.2)

$$\mu_N(t) := \frac{1}{N}\sum_{i=1}^{N} \delta(x - X^i(t)), \tag{1.4}$$

which is a random measure. In fact, under certain assumptions on V and K, it can be shown that the uniform mean field limit holds ([19]). Both the marginal distribution of X^1 in (1.2), denoted as $\mu_N^{(1)}$, and (1.4) are close to μ.

On the other hand, one often cares more about the mean field equations like (1.3) and its invariant measure π, but the mean field equation and π are hard or expensive to solve or compute. Using the mean field approximation, one can generate some artificial particle systems of the form (1.1) or (1.2). In this sense, the interacting particle systems (1.1) and (1.2) are Monte-Carlo particle methods for the mean field equation and π. Clearly, when N (the number of numerical particles) is large, one still wants to reduce the computational cost. Hence, no matter whether studying system (1.2) (and (1.1)) is due to its own interest or due to numerical simulation of the mean field equation, it is highly desirable to design some efficient algorithms to solve the particle system (1.2).

Nowadays, in the era of big data, many stochastic algorithms have been developed to reduce the computational cost while keep certain accuracy. Hence, one may borrow some ideas from these areas for the physical problems. The stochastic gradient descent (SGD) algorithm is developed to reduce the computational cost and for better exploring the high dimensional parameter space [20, 21] for supervised learning [22]. In SGD, a small batch

of samples are chosen each time to form the noisy gradient. A similar algorithm is the stochastic coordinate descent where only a few coordinates are updated each time [23, 24]. The idea of random batches also appears in the stochastic gradient Langevin dynamics (a Markov Chain Monte Carlo method (MCMC)), which can be applied for Bayesian inference [25, 26, 27]. Pretty much like SGD, the method uses random batch of data to update the parameters one needs to estimate. Besides, many random algorithms also appear for matrices, such as constructing approximate matrix decompositions by randomized Singular Value Decompositions [28].

Motivated by the stochastic algorithms in machine learning and statistics, we develop Random Batch Methods (RBM) for interacting particle systems (1.2) and (1.1). The idea is quite simple: for a *small* duration of time, one randomly divides the particles into *small* batches and the interactions are only turned on inside each batch (see Section 2 for details). Numerical experiments in later sections show that RBMs work quite well for a wide range of applications with binary interactions from physical, biological to data sciences. These algorithms reduce the computational cost per time step from $O(N^2)$ to $O(N)$ for binary interactions (1.2) ($O(N^J)$ to $O(N)$ for (1.1)). They not only recover the equilibrium well in variaous problems, but also approximate *dynamics* of measures with very singular interaction kernels (see for example section 4.2). Moreover in section 3, under suitable conditions on the external potential, interaction force and batch size, we prove for one of the RBMs that the numerical error has a bound that only depends on the time step, *independent* of N and time. The key for the proof is Lemma 3.1, which guarantees that on average the random force is consistent with the full interaction.

The methodology of randomly choosing some objects has also been widely used in kinetic theory community. For example, the Direct Simulation Monte Carlo (DSMC) method proposed by Bird ([29, 30]) uses randomly chosen simulation molecules to solve the Boltzmann equation with finite Knudsen number (see the work by Wagner [31] for the proof of convergence). Moreover, Nanbu's simulation method directly derived from the Boltzmann equation has gained success and been proved to converge [32, 33]. In [34], biological swarm models with random pair interactions have been studied and proven to converge to Kac-Boltzmann equation. In the reverse way to [34], the work by Albi and Pareschi ([35]) uses stochastic binary interaction algorithms for flocking and swarming dynamics. The algorithms in [35] are intrinsically doing similar things as our RBMs do; their Algorithm 4.3 is particularly alike RBM-1 (see section 2) with $p = 2$. See a more recent application by Carrillo et al [36]. In this sense, our RBMs are generalizations of symmetric Nanbu (Algorithm 4.3) in [35]. However, the numerical particle system in [35] was motivated by the mean-field limit, in a way similar to the DSMC methods of binary collisions for the Boltzmann equation. It was aimed at solving the mean-field equation (corresponding to the right vertical line in Fig. 1). RBMs are motivated by the idea of mini-batch methods from machine learning. They are Monte-Carlo type approximations directly to the (physical) particle systems (1.2). They correspond to the left horizontal line in Fig. 1. Moreover, one of the RBMs will be proved (under special conditions) to converge, when time steps go to zero, to the particle system (1.2). The construction of RBMs, as well as the proof of their convergence, are obtained *without the knowledge of the mean-field limit*. Of course when $N \to \infty$ its density measure also converges to the mean-field limit equation, as will be proven using the Wasserstein distance. Therefore, the RBMs are a class of Asymptotic-Preserving schemes for particle systems (1.2) or (1.1), in the sense of [37] (see Fig. 1). Lastly, the idea of turning on interactions inside the batches can be easily extended. For example, one can change the batch size to adjust the noise levels. Moreover, they can also be extended to (1.1) involving more complex interactions.

The rest of the paper is organized as follows. In section 2, we propose the Random Batch Method without replacement (RBM-1) and with replacement (RBM-r), and give some discussions. In section 3, we obtain an error estimate, which only depends on time step but not on N and time, of the marginals for RBM-1 in the Wasserstein distance under some assumptions for the external and interacting forces. Though performed for binary interactions, the analysis should work for (1.1) as well. In section 4, we do numerical tests to verify the theoretic results in section 3 and use the Dyson Brownian motion to compare

the RBMs. In particular, in section 4.2, we compute the law of the Dyson Brownian motions from random matrix theory where the interaction kernel is *singular*. Our method can caputre the evolution of distribution and the equilibrium semicircle law, and the two algorithms (RBM-1 and RBM-r) give comparable results. In section 5, we use the RBMs to solve some interesting application problems with binary interactions. In section 5.1, we focus on Thomson's problems where we solve the dynamics on sphere. In section 5.2, we design stochastic dynamics for evolution of wealth and opinions. Lastly, we apply our randomized algorithms for efficient clusters in section 5.3. The paper is concluded in section 6.

2 The Random Batch Methods

In this section, we propose the RBMs by using random batches for the summation of the interacting force in (1.2). The extensions to (1.1) should be similar but are more involved, and we will give some discussion about this in section 2.3. For the setup, we pick a short duration of time τ and consider the discrete time

$$t_m = m\tau. \tag{2.1}$$

Suppose we compute up to time T and the number of iteration for the stochastic algorithm is

$$N_T = \left\lceil \frac{T}{\tau} \right\rceil. \tag{2.2}$$

Clearly to simulate the ODE system (1.2) directly, the complexity is $O(N_T N^2)$. If N is large, this is expensive. Motivated by the stochastic algorithms in the machine learning community, we will use a randomized strategy.

2.1 The first algorithm (RBM-1)

A natural idea is that at each time t_m, we divide the $N = np$ particles into n small batches with size p ($p \ll N$, often $p = 2$) randomly, denoted by $\mathcal{C}_q, q = 1, \dots, n$, and then interact particles within each batch (For conveience, we have assumed that p divides N. In general, the last batch does not have to have size p.) The algorithm is called RBM-1 (shown in 1). Each iteration contains two main steps: (1) Randomly shuffling and dividing the particles into n batches; (2) evolving with interactions only turned on inside batches.

Algorithm 1 (RBM-1)

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

$$dX^i = -\nabla V(X^i)dt + \frac{1}{p-1} \sum_{j \in \mathcal{C}_q, j \neq i} K(X^i - X^j)dt + \sigma dB^i. \tag{2.3}$$

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

Clearly, the update equation (2.3) can be rewritten as

$$dX^i = -\nabla V(X^i)dt + \frac{1}{N-1} \sum_{j:j \neq i} K(X^j - X^i)dt + \sigma dB^i + \chi_{m,i}(X(t))\, dt, \tag{2.4}$$

where

$$\chi_{m,i}(X(t)) = \frac{1}{p-1} \sum_{j \in \mathcal{C}_q, j \neq i} K(X^i - X^j) - \frac{1}{N-1} \sum_{j:j \neq i} K(X^i - X^j). \tag{2.5}$$

For a given $x = (x^1, \dots, x^N) \in \mathbb{R}^{Nd}$ that is independent of the random division, we have (see Lemma 3.1 for the proof)

$$\mathbb{E}\chi_{m,i}(x) = 0. \tag{2.6}$$

This is a key observation which eventually leads to the convergence of the algorithms in expectation. As a remark, the position $X = (X^1, \dots, X^N)$ itself depends on the random division. Hence, in general,

$$\mathbb{E}\chi_{m,i}(X(t)) \neq 0.$$

Regarding the complexity, note that random division into n batches of equal size can be implemented using random permutation. The latter can be realized in $O(N)$ operations by Durstenfld's modern revision of Fisher-Yates shuffle algorithm [38]. (In MATLAB, one can use 'randperm(N)' to generate a random permutation. Then, the first p elements are considered to be in the first batch, the second p elements are in the second batch, etc). Hence, if one is to simulate up to time T, the complexity is $O(pN_TN)$. If $p = 2$, the complexity is $O(N_TN)$. The cost is reduced significantly.

RBM-1 is in spirit similar to the stochastic gradient descent in machine learning ([20, 21]). Recently, there are some analysis of SGD in the mathematical viewpoints and applications to physical problems [39, 40, 41, 42]. RBM-1 can be used both for simulating the evolution of the measure (1.4) or (1.3) and for sampling from the equilibrium state π.

2.2 The Random Batch Method with replacement

RBM-1 requires the random division, and the elements in different batches are different. This is in fact the sampling without replacement. If one allows replacement, we have RBM-r$'$.

Algorithm 2a (RBM-r$'$)

1: **for** m in $1 : [T/\tau]$ **do**
2: **for** k from 1 to N/p **do**
3: Pick a set $\mathcal{C}_k$ of size p randomly with replacement.
4: Update X^i's ($i \in \mathcal{C}_k$) by solving the following SDE for time τ.

$$\begin{cases} dY^i = -\nabla V(Y^i)dt + \dfrac{1}{p-1}\displaystyle\sum_{j\in\mathcal{C}_k, j\neq i} K(Y^i - Y^j)dt + \sigma dB^i, \\ Y^i(0) = X^i, \end{cases} \tag{2.7}$$

i.e., solve (2.7) with initial values $Y^i(0) = X^i$, and set $X^i \leftarrow Y^i(\tau)$.
5: **end for**
6: **end for**

RBM-r$'$ can be reformulated as Random Batch Method with replacement (RBM-r) that has some flavor of the stochastic coordinate descent method [23, 24]. Here, the pseudo-time s is introduced for convenience and $s_m = m\tau$. Roughly, t_m corresponds to $s_{mN/p}$.

Algorithm 2b (RBM-r)

1: **for** m in $1 : [T/\tau] * (N/p)$ **do**
2: Pick a set $\mathcal{C}$ of size p randomly.
3: Update X^i's ($i \in \mathcal{C}$) by solving the following with pseudo-time $s \in [s_{m-1}, s_m)$.

$$dX^i = -\nabla V(X^i)\, ds + \frac{1}{p-1}\sum_{j\in\mathcal{C}, j\neq i} K(X^i - X^j)\, ds + \sigma\, dB^i. \tag{2.8}$$

4: **end for**

For such type of methods, we expect to have

$$\tilde{N}_T = \frac{1}{p} N N_T \tag{2.9}$$

iterations to get comparable behaviors. However, each step is very cheap: one only needs $O(p^2)$ work. Hence, we still expect the complexity to be $O(pNN_T)$.

Though the same as RBM-r$'$, RBM-r does not have explicit concept of time, since the positions of the particles are not changing simultaneously. Intuitively, RBM-r can only give sampling for the invariant measure for the nonlinear Fokker-Planck equation (1.2) while unable to simulate the dynamics of the distribution. However, as RBM-r$'$ indicates, N/p iterations correspond to time τ and thus a single step in RBM-1, so it might still approximate the evolution of distributions. In fact, the example later confirms this and $m(N/p)$ iterations indeed give acceptable approximation for the distribution at time $m\tau$. As a last comment, we may sometimes choose random time step; for example $\tau \sim Exp(\Delta t)$ (the exponential distribution with parameter Δt) such that $\mathbb{E}\tau = \Delta t$. Intuitively, this may help to increase the noise level and avoid being trapped in some local minimizers of V.

2.3 Some discussions about the algorithms

In this subsection we make some discussion about our algorithms and complementary remarks.

1. For system (1.1), the RBMs can be similarly developed. The batch size should be larger than or equal to J ($p \geq J$). After the random batch is chosen, the interactions only happen inside the randomly chosen batch. One should also adjust the coefficients, such that an analogy of Lemma 3.1 holds. In other words, the expectation of the random forces should equal to those in (1.1). The complexity clearly is reduced from $O(N^J)$ to $O(N)$.

2. The sizes of batch do not have to be equal. One can even choose the sizes randomly. One can also adjust the batch size to adjust the noise level.

3. After the batches are obtained, one can do parallel computing to update the positions of X for RBM-1. However, one has to do regrouping after one time interval which stops the parallelism. There is no big efficiency improvement unless solving the SDEs is expensive. Regarding complexity, there is no big difference in complexity. For RBM-r, to get p indices, one needs to generate p random numbers from $[0, 1]$, and the total complexity is $O(N)$ for N/p iterations. For RBM-1, as in Durstenfld's implementation of random permutation, one needs to generate N random numbers from $[0, 1]$ while do some swapping operations for each element. (This may be different from the intuition established from drawing balls from a bag. For randomly dividing N balls, one can choose p from N balls first, then p from $N - p$, ... It seems that the complexity becomes smaller and smaller as the total number of balls is becoming smaller and smaller. This is indeed not the case because in computer one can use arrays to store the 'balls' and one only needs to generate the indices without touching these 'balls'.)

   In summary, there is no big difference in complexity for the two algorithms. The advantage of RBM-1 might be its ability for parallelism during one interval while the advantage of RBM-r is its simplicity so that it is likely more flexible for extensions.

4. In Lemma 3.1, for x independent of the random division, we prove that

$$\mathrm{Var}(\chi_{m,i}(x)) = \left(\frac{1}{p-1} - \frac{1}{N-1}\right)\Lambda(x),$$

   where $\Lambda(x)$ is independent of p. This means for larger p, the variance is smaller and the noise level is lower. This noise somehow reflects the fluctuation of the empirical measure $\tilde{\mu}_N^{(1)}$ around $\mu_N^{(1)}$.

5. If K is a singular forcing (like Coulomb), we can do *splitting* method and have

$$\begin{aligned} dX^i &= \frac{1}{p-1}\sum_{j\in\mathcal{C}, j\neq i} K(X^i - X^j)\,dt, \\ dX^i &= -\nabla V(X^i)\,dt + \sigma dB^i. \end{aligned} \tag{2.10}$$

   If $p = 2$, the singular forcing term can be often solved *analytically*. This is another advantage of the stochastic algorithm: for the N-particle system, if the forcing K is singular so that the problem is stiff, explicit scheme needs very small time step while implicit scheme is hard to invert. Using the stochastic algorithm plus time splitting, the evolution can be solved *exactly*, thus avoiding stability constraint.

6. Suppose one aims at the law of particles in (1.2) with very large N. By mean field limit, one can choose to solve (1.3) directly. A possible way is to use particle method for (1.3) where the number of particles is much smaller than N and the masses for the particles in the numerical method can even be different (the particle blob method [43]). For the interacting particle system of this numerical purpose, our randomized algorithms also apply well. In this regard, we provided an efficient numerical particle method for the mean-field (or kinetic) equations. In fact this is exactly the starting point of the binary algorithms in [35].

7. To better simulate the invariant measure in the case of $\sigma = 0$, one may add the Brownian motion $\varepsilon_N dB^i$ where ε_N decreases with N. For example, we may set

$$\varepsilon_N = \frac{1}{N^\gamma},\ \gamma > 0.$$

8. If the initial distribution is far from the equilibrium and we aim to get the global equilibrium, one may need many iterations for convergence to the equilibrium. A possible way is to use the Gibbs distribution corresponding to V

$$\nu(dx) \propto \exp(-V(x))\,dx$$

   for initialization to reduce the number of iterations. This distribution can be sampled using the Markov chain Monte Carlo (MCMC) methods [44, 45]. Of course, for special V such as quadratic functions, the initial distribution can be sampled directly without MCMC.

3 An error analysis for RBM-1

In this section, we study RBM-1 proposed in Section 2.1. In particular, we will check how close it is to the fully coupled system (1.2). We leave the study of RBM-r to the future. However, as commented above, we expect RBM-r to also work when RBM-1 works. Recall that each iteraction consists of random division and evolution. The mechanism that makes RBMs work is Lemma 3.1 and small step size. The philosophy is as following. Suppose that the system has certain chaotic property. When τ is small enough, the accumulative behavior along many time steps will be roughly comparable to the average behavior, which is (1.2) thanks to Lemma 3.1. This is similar to Law of Large Numbers, but on time.

We assume the following conditions on the confining and interacting potentials:

Assumption 3.1. Suppose V is strongly convex on $\mathbb{R}^d$ so that $x \mapsto V(x) - \frac{r}{2}|x|^2$ is convex, and ∇V, $\nabla^2 V$ have polynomial growth (i.e. $|\nabla V(x)| + |\nabla^2 V(x)| \le C(1 + |x|^q)$ for some $q > 0$). Assume $K(\cdot)$ is bounded, Lipschitz on $\mathbb{R}^d$ with Lipschitz constant L and has bounded second order derivatives. Moreover,

$$r > 2L. \tag{3.1}$$

The condition $r > 2L$ is to ensure that the evolution group for the deterministic part of (1.2) is a contraction. In particular, if

$$\dot{X}^i = -\nabla V(X^i) + \frac{1}{N-1}\sum_{j:j\neq i} K(X^i - X^j),\ \dot{Y}^i = -\nabla V(Y^i) + \frac{1}{N-1}\sum_{j:j\neq i} K(Y^i - Y^j),$$

then

$$\frac{d}{dt}\sum_{i=1}^N |X^i - Y^i|^2 \leq -(r-2L)\sum_{i=1}^N |X^i - Y^i|^2. \tag{3.2}$$

Of course, for such systems, one must have $\sigma > 0$ to have a nontrivial equilibrium. Otherwise, all the particles will go to a single point.

Our goal is to prove that the distribution generated by RBM-1 is close to the distribution of the marginal distribution of the N particle system in the Wasserstein distance. We recall that the Wasserstein-2 distance is given by [46]

$$W_2(\mu,\nu) = \left(\inf_{\gamma\in\Pi(\mu,\nu)}\int_{\mathbb{R}^d\times\mathbb{R}^d} |x-y|^2 d\gamma\right)^{1/2}, \tag{3.3}$$

where $\Pi(\mu,\nu)$ means all the joint distributions whose marginal distributions are μ and ν respectively.

In order to achieve this goal, we consider the synchronous coupling between (1.2) and (2.3). We denote X^i the solution obtained by the N interacting particle system while $\tilde{X}^i$ the solution obtained by RBM-1. Correspondingly, $\tilde{B}^i$ will denote the Brownian motion used in RBM-1. Note that both (1.2) and (2.3) have exchangeability. This means, for example, the joint distribution of X^i's is symmetric. Consequently, $\tilde{X}^i$ and $\tilde{X}^j$ (for all i and j) have the same distribution. We construct the coupling as follows. For $i = 1, 2, \ldots, N$,

$$\begin{aligned} &\tilde{X}^i(0) = X^i(0) =: X_0^i \sim \nu, \\ &B^i(t) = \tilde{B}^i(t). \end{aligned} \tag{3.4}$$

With this coupling, we define the error process

$$Z^i(t) := \tilde{X}^i(t) - X^i(t). \tag{3.5}$$

Let ξ_{m-1} denote the random division of batches at t_{m-1}, and define

$$\mathcal{F}_{m-1} = \sigma(X_0^i, B^i(t), \xi_j; t \leq t_{m-1}, j \leq m-1). \tag{3.6}$$

In other words, $\mathcal{F}_{m-1}$ is the σ-algebra generated by the initial values X_0^i $(i = 1, \ldots, N)$, $B^i(t)$, $t \leq t_{m-1}$, and ξ_j, $j \leq m-1$. Hence, $\mathcal{F}_{m-1}$ contains the information for how batches are constructed for $t \in [t_{m-1}, t_m)$. We also introduce

$$\mathcal{G}_{m-1} = \sigma(X_0^i, B^i(t), \xi_j; t \leq t_{m-1}, j \leq m-2). \tag{3.7}$$

If we use $\sigma(\xi_m)$ to mean the σ-algebra generated by ξ_m, then $\mathcal{F}_{m-1} = \sigma(\mathcal{G}_{m-1} \cup \sigma(\xi_{m-1}))$. Throughout this section, C will denote generic constants whose concrete value can change from line to line. We use $\|\cdot\|$ to represent the L^2 norm:

$$\|v\| = \sqrt{\mathbb{E}|v|^2}. \tag{3.8}$$

We now state the theorem.

Theorem 3.1. *Suppose Assumption 3.1 holds. With the coupling constructed above,*

$$\sup_{t\geq 0}\|Z^1(t)\| \leq C\sqrt{\frac{\tau}{p-1} + \tau^2}, \tag{3.9}$$

where C is independent of N, p and t. Consequently, let $\mu_N^{(1)}(t)$ be the first marginal for (1.2) and $\tilde{\mu}_N^{(1)}$ be the first marginal for system (2.3). Then

$$\sup_{t\geq 0} W_2(\mu_N^{(1)}(t), \tilde{\mu}_N^{(1)}(t)) \leq C\sqrt{\frac{\tau}{p-1} + \tau^2} \leq C\sqrt{\tau}. \tag{3.10}$$

We need some preparation for the proof. Lemma 3.1 is a type of consistency lemma for RBMs while Lemma 3.2 somehow shows the stability of RBM-1.

Lemma 3.1. *Consider $p \geq 2$ and a given fixed $x = (x^1, \dots, x^N) \in \mathbb{R}^{Nd}$. Then, for all i,*

$$\mathbb{E}\chi_{m,i}(x) = 0, \tag{3.11}$$

where the expectation is taken with respect to the random division of batches. Moreover, the variance is given by

$$\mathrm{Var}(\chi_{m,i}(x)) = \left(\frac{1}{p-1} - \frac{1}{N-1}\right)\Lambda(x), \tag{3.12}$$

where

$$\Lambda(x) := \frac{1}{N-2}\sum_{j:j\neq i}\left(K(x^i - x^j) - \frac{1}{N-1}\sum_{k:k\neq i} K(x^i - x^k)\right)^2. \tag{3.13}$$

Proof. We use $I(i,j) = 1$ to indicate that i, j are in the same batch. We rewrite

$$f_{m,i} := \frac{1}{p-1}\sum_{j:\in\mathcal{C}_q, j\neq i} K(x^i - x^j) = \frac{1}{p-1}\sum_{j:j\neq i} K(x^i - x^j)1_{I(i,j)=1}.$$

We first show that

$$\mathbb{E}1_{I(i,j)=1} = \frac{p-1}{N-1}.$$

As is well-known, there are

$$M(n) := \frac{(pn)!}{(p!)^n n!}$$

ways of dividing pn distinguishable objects into n batches of size p.

For a given (i,j) pair, to compute $\mathbb{E}1_{I(i,j)=1}$, the probability that i, j are in the same batch, we compute the number of ways to group (i,j) together. We first choose $p-2$ objects from $np-2$ to group with i, j and then form a batch. Then, divide the remaining into $n-1$ batches. Hence, the number of ways to make i, j in the same batch is $\binom{np-2}{p-2}M(n-1)$ and thus

$$\mathbb{E}1_{I(i,j)=1} = \frac{\binom{np-2}{p-2}M(n-1)}{M(n)} = \frac{p-1}{N-1}.$$

This then proves that

$$\mathbb{E}f_{m,i} = \frac{1}{N-1}\sum_{j:j\neq i} K(x^i - x^j),$$

and thus the claim for the expectation (3.11) follows.

Now, let us consider the variance. We first compute the second moment of $\xi_{m,i}$:

$$\begin{aligned}\mathbb{E}f_{m,i}^2 =& \frac{1}{(p-1)^2}\sum_{j:j\neq i}(K(x^i - x^j))^2\mathbb{P}(I(i,j)=1)\\ &+ \frac{1}{(p-1)^2}\sum_{i,j:j\neq k, j\neq i, k\neq i} K(x^i - x^j)K(x^i - x^k)\mathbb{P}(I(i,j)I(j,k)=1).\end{aligned}$$

By similar argument,

$$\mathbb{P}(I(i,j)I(j,k)=1) = \frac{\binom{np-3}{p-3}M(n-1)}{M(n)} = \frac{(p-1)(p-2)}{(N-1)(N-2)}.$$

Hence,

$$\begin{aligned}\mathrm{Var}(\chi_{m,i}(x)) &= \mathbb{E}f_{m,i}^2 - (\mathbb{E}f_{m,i})^2 \\ &= \left(\frac{1}{p-1} - \frac{1}{N-1}\right)\Big(\frac{1}{N-1}\sum_{j\neq i}(K(x^i-x^j))^2 \\ &\quad - \frac{1}{(N-1)(N-2)}\sum_{i,j:j\neq k,j\neq i,k\neq i} K(x^i-x^j)K(x^i-x^k)\Big) \\ &= \left(\frac{1}{p-1} - \frac{1}{N-1}\right)\frac{1}{N-2}\sum_{j:j\neq i}\left(K(x^i-x^j) - \frac{1}{N-1}\sum_{k:k\neq i}K(x^i-x^k)\right)^2.\end{aligned}$$

□

Lemma 3.2. *Suppose the coupling constructed in (3.4) and Assumption 3.1 hold. Then, for any $q \geq 2$, there exists a constant C_q independent of N such that for any i*

$$\sup_{t\geq 0}(\mathbb{E}|X^i(t)|^q + \mathbb{E}|\tilde{X}^i(t)|^q) \leq C_q. \tag{3.14}$$

Moreover, for $t \in [t_{m-1}, t_m)$,

$$\begin{aligned}&\|Z^i(t) - Z^i(t_{m-1})\| \leq C\tau, \\ &|\mathbb{E}(Z^i(t) - Z^i(t_{m-1})\chi_{m,i}(\tilde{X}(t_{m-1})))| \leq \|\Lambda\|_\infty \frac{\tau}{p-1} + C\left(\tau^2 + \sup_{s\in[t_{m-1},t]}\|Z^i\|\tau\right),\end{aligned} \tag{3.15}$$

where C is independent of N, p and m.

Proof. Consider system (1.2) first. By Itô's calculus,

$$\begin{aligned}\frac{d}{dt}\mathbb{E}|X^i|^q = q\mathbb{E}|X^i|^{q-2}&\left(-X^i\cdot\nabla V(X^i) + \frac{1}{N-1}\sum_{j:j\neq i}X^i\cdot K(X^i-X^j)\right) \\ &+ \frac{1}{2}q(q+d-2)\sigma^2\mathbb{E}|X^i|^{q-2}.\end{aligned}$$

Note that

$$X^i\cdot\nabla V(X^i) = (X^i-0)\cdot(\nabla V(X^i)-\nabla V(0)) + X^i\cdot\nabla V(0) \geq r|X^i|^2 + X^i\cdot\nabla V(0).$$

Recalling also that K is bounded,

$$\frac{d}{dt}\mathbb{E}|X^i|^q \leq -qr\mathbb{E}|X^i|^q + q(\|K\|_\infty + |\nabla V(0)|)\mathbb{E}|X^i|^{q-1} + \frac{1}{2}q(q+d-2)\sigma^2\mathbb{E}|X^i|^{q-2}. \tag{3.16}$$

By Young's inequality,

$$\mathbb{E}(|X^i|^{q-1}) \leq \frac{(q-1)\nu}{q}\mathbb{E}(|X^i|^q) + \frac{1}{q\nu^{q-1}},$$

the second term on the right hand side of (3.16) is therefore controlled. If $q = 2$, the last term on the right hand side of (3.16) is controlled by a constant. Otherwise, one can apply Young's inequality similarly to control it with $\mathbb{E}(|X^i|^q)$. Clearly, when choosing ν fixed but small enough, $\mathbb{E}|X_i|^q$ can be uniformly bounded in time for any $q \geq 2$, and the bound is independent of N.

For $\tilde{X}^i$, we first consider a given random division so that the equation is given by

$$d\tilde{X}^i = -\nabla V(\tilde{X}^i)dt + \frac{1}{p-1}\sum_{j\in\mathcal{C}_\theta, j\neq i}K(\tilde{X}^i-\tilde{X}^j)\,dt + \sigma dB^i,$$

where $\mathcal{C}_\theta$ is the random batch that contains i from the random division at t_{m-1}, or ξ_{m-1}. Now, consider that $t \in [t_{m-1}, t_m)$. Conditioning on $\mathcal{F}_{m-1}$ and applying Itô's calculus on $[t_{m-1}, t_m)$, one also has

$$\frac{d}{dt}\mathbb{E}(|\tilde{X}^i|^q|\mathcal{F}_{m-1}) = q\mathbb{E}\Big[|\tilde{X}^i|^{q-2}(-\tilde{X}^i \cdot \nabla V(\tilde{X}^i) + \tilde{X}^i \cdot \frac{1}{p-1}\sum_{j\in\mathcal{C}_\theta, j\neq i} K(\tilde{X}^i - \tilde{X}^j)))|\mathcal{F}_{m-1}\Big] + \frac{1}{2}q(q+d-2)\sigma^2\mathbb{E}(|\tilde{X}^i|^{q-2}|\mathcal{F}_{m-1}).$$

Using similar estimates,

$$\frac{d}{dt}\mathbb{E}(|\tilde{X}^i|^q|\mathcal{F}_{m-1}) \leq -r_1\mathbb{E}(|\tilde{X}^i|^q|\mathcal{F}_{m-1}) + C_1,$$

for some $r_1 > 0$ and constant C_1 that are deterministic. Taking expectation about the randomness in $\mathcal{F}_{m-1}$ on both sides, one then obtains the same inequality as for X. The claim for $\tilde{X}$ also follows.

Now, by the coupling, Z^i satisfies on $t \in [t_{m-1}, t_m)$

$$dZ^i = -(\nabla V(\tilde{X}^i) - \nabla V(X^i))dt + \frac{1}{p-1}\sum_{j\in\mathcal{C}_\theta, j\neq i} K(\tilde{X}^i - \tilde{X}^j)\,dt - \frac{1}{N-1}\sum_{j:j\neq i} K(X^i - X^j)\,dt.$$

Since ∇V has polynomial growth, the claim for $\|Z^i(t) - Z^i(t_{m-1})\|$ is then an easy consequence of the q-moment estimates just proved.

Moreover, since

$$|\nabla V(\tilde{X}^i) - \nabla V(X^i)| \leq \int_0^1 |(\tilde{X}^i - X^i) \cdot \nabla^2 V((1-z)\tilde{X}^i + zX^i)|\,dz,$$

we find that this controlled by $C(|\tilde{X}^i|^{q_1} + |X^i|^{q_1})$ for some $q_1 > 0$. Hence,

$$\mathbb{E}((Z^i(t) - Z^i(t_{m-1}))\cdot\chi_{m,i}(\tilde{X}(t_{m-1}))) \leq (\sup_{s\in[t_{m-1},t]}\|Z^i(s)\|)C(\|K\|_\infty + L)\tau + \mathbb{E}\int_{t_{m-1}}^t \chi_{m,i}(\tilde{X}(s))\cdot\chi_{m,i}(\tilde{X}(t_{m-1}))\,ds.$$

For the second term,

$$\mathbb{E}\chi_{m,i}(\tilde{X}(s))\cdot\chi_{m,i}(\tilde{X}(t_{m-1})) = \mathbb{E}\Big((\chi_{m,i}(\tilde{X}(s)) - \chi_{m,i}(\tilde{X}(t_{m-1})))\cdot\chi_{m,i}(\tilde{X}(t_{m-1}))\Big) + \mathbb{E}(\chi(\tilde{X}(t_{m-1})))^2 \leq C\|K\|_\infty L\tau + \mathbb{E}(\mathbb{E}(\chi(\tilde{X}(t_{m-1}))^2|\mathcal{G}_{m-1})). \quad (3.17)$$

Applying Lemma 3.1,

$$\mathbb{E}(\mathbb{E}(\chi(\tilde{X}(t_{m-1}))^2|\mathcal{G}_{m-1})) = (\frac{1}{p-1} - \frac{1}{N-1})\mathbb{E}(\Lambda(\tilde{X}(t_{m-1}))) \leq \frac{1}{p-1}\|\Lambda\|_\infty,$$

the claim follows. □

Note that conditioning on the random division ξ_{m-1}, the indices are not symmetric. We now provide estimates regarding the conditional expectations.

Lemma 3.3. *Consider process $\tilde{X}$. For $t \in [t_{m-1}, t_m)$ and all i, it holds that*

$$\|\mathbb{E}(\tilde{X}^i(t) - \tilde{X}^i(t_{m-1})|\mathcal{F}_{m-1})\| \leq C\tau, \quad (3.18)$$

and

$$\|\mathbb{E}(|\tilde{X}^i(t) - \tilde{X}^i(t_{m-1})|^2|\mathcal{F}_{m-1})\| \leq C\tau, \quad (3.19)$$

where C is independent of ξ_m, m and N.

Proof. By Lemma 3.2, one has

$$\mathbb{E}|\tilde{X}^i(t)|^q \le C, \ \forall i,$$

and C is independent of N, m.

Now, consider an experiment so that the equation is written as

$$d\tilde{X}^i(t) = -\nabla V(\tilde{X}^i)\,dt + \frac{1}{p-1}\sum_{j\in\mathcal{C}_\theta, j\neq i} K(\tilde{X}^i - \tilde{X}^j) + \sigma dB^i,$$

where $\mathcal{C}_\theta$ again is the random batch that contains i from the random division at t_{m-1}, or ξ_{m-1}. It follows that

$$\begin{aligned}\mathbb{E}(\tilde{X}^i(t) - \tilde{X}^i(t_{m-1})|\mathcal{F}_{m-1}) = &-\int_{t_{m-1}}^t \mathbb{E}(\nabla V(\tilde{X}^i)|\mathcal{F}_{m-1})\,ds \\ &+ \int_{t_{m-1}}^t \mathbb{E}\left(\frac{1}{p-1}\sum_{j\in\mathcal{C}_\theta, j\neq i} K(\tilde{X}^i - \tilde{X}^j)|\mathcal{F}_{m-1}\right) ds.\end{aligned}$$

Note that K is bounded and $|\nabla V(x)| \le C(1+|x|^q)$ for some $q > 0$. Hölder's inequality for conditional expectations gives

$$\begin{aligned}\|\mathbb{E}(\nabla V(\tilde{X}^i)|\mathcal{F}_{m-1})\| &= (\mathbb{E}|\mathbb{E}(\nabla V(\tilde{X}^i)|\mathcal{F}_{m-1})|^2)^{1/2} \\ &\le \Big(\mathbb{E}\Big(\mathbb{E}(|\nabla V(\tilde{X}^i)|^2|\mathcal{F}_{m-1})\Big)\Big)^{1/2} = \|\nabla V(\tilde{X}^i)\|.\end{aligned}$$

Together with Lemma 3.2, this implies the first claim.

For the second claim, Itô's formula implies that

$$\begin{aligned}&\frac{d}{dt}\mathbb{E}\left[(\tilde{X}^i(t) - \tilde{X}^i(t_{m-1})^2|\mathcal{F}_{m-1}\right] \\ &= 2\mathbb{E}\left[\Big(\tilde{X}^i(t) - \tilde{X}^i(t_{m-1})\Big)\cdot(-\nabla V(\tilde{X}^i) + \frac{1}{p-1}\sum_{j\in\mathcal{C}_\theta, j\neq i} K(\tilde{X}^i - \tilde{X}^j))|\mathcal{F}_{m-1}\right] + \sigma^2.\end{aligned}$$

Similarly,

$$\left\|\mathbb{E}\left[\Big(\tilde{X}^i(t) - \tilde{X}^i(t_{m-1})\Big)\cdot\left(-\nabla V(\tilde{X}^i) + \frac{1}{p-1}\sum_{j\in\mathcal{C}_\theta, j\neq i} K(\tilde{X}^i - \tilde{X}^j)\right)|\mathcal{F}_{m-1}\right]\right\|$$

is uniformly bounded by Hölder's inequality for conditional expectations. The second claim also follows. □

Now, we are ready to prove Theorem 3.1.

Proof of Theorem 3.1. With the coupling (3.4), the continuous process Z^i satisfies for $t \in [t_{m-1}, t_m)$ that

$$\begin{aligned}dZ^i = &-(\nabla V(\tilde{X}^i) - \nabla V(X^i))dt \\ &+ \frac{1}{N-1}\sum_{j:j\neq i}(K(\tilde{X}^i - \tilde{X}^j) - K(X^i - X^j))dt + \chi_{m,i}(\tilde{X}(t))dt.\end{aligned}$$

Using the strong convexity of V and Lipscthitz continuity of K,

$$\frac{1}{2}d\mathbb{E}\sum_{i=1}^N (Z^i)^2 \le -(r-2L)\mathbb{E}\sum_{i=1}^N (Z^i)^2\,dt + \sum_{i=1}^N \mathbb{E}Z^i(t)\cdot\chi_{m,i}(\tilde{X}(t))\,dt,$$

where $\tilde{X} = (\tilde{X}^1, \dots, \tilde{X}^N) \in \mathbb{R}^{Nd}$. Due to the exchangeability,

$$R(t) := \frac{1}{N}\sum_{i=1}^{N} \mathbb{E} Z^i(t) \cdot \chi_{m,i}(\tilde{X}(t)) = \mathbb{E} Z^1(t) \cdot \chi_{m,1}(\tilde{X}(t)).$$

For notational convenience, we introduce

$$u(t) := \frac{1}{N}\mathbb{E}\sum_{i=1}^{N}(Z^i)^2 = \|Z^1(t)\|^2.$$

Then,

$$\frac{d}{dt}u(t) \le -2(r-2L)u(t) + 2R(t).$$

We now estimate $R(t)$. We rewrite it as

$$\begin{aligned} R(t) =& \mathbb{E} Z^1(t_{m-1}) \cdot \chi_{m,1}(\tilde{X}(t_{m-1})) + \mathbb{E}(Z^1(t) - Z^1(t_{m-1})) \cdot \chi_{m,1}(\tilde{X}(t_{m-1})) \\ &+ \mathbb{E} Z^1(t_{m-1}) \cdot (\chi_{m,1}(\tilde{X}(t)) - \chi_{m,1}(\tilde{X}(t_{m-1}))) \\ &+ \mathbb{E}(Z^1(t) - Z^1(t_{m-1})) \cdot (\chi_{m,1}(\tilde{X}(t)) - \chi_{m,1}(\tilde{X}(t_{m-1}))) \\ =:& I_1 + I_2 + I_3 + I_4. \end{aligned} \tag{3.20}$$

By Lemma 3.1, we have

$$I_1 = \mathbb{E}\left(Z^1(t_{m-1}) \cdot \mathbb{E}(\chi_{m,1}(\tilde{X}(t_{m-1}))|\mathcal{G}_{m-1})\right) = 0.$$

This is the consistency that ensures convergence. Note that $\mathcal{G}_{m-1}$ does not contain the information of random division at t_{m-1} and this is why we can get $I_1 = 0$.

By Lemma 3.2, the second term on the right hand side of (3.20) is bounded by

$$I_2 \le C\frac{\tau}{p-1} + C\|Z^1(t)\|\tau + C\tau^2,$$

where C is independent of N and p.

By Lemma 3.2 again,

$$I_4 \le \|Z^1 - Z^1(t_{m-1})\|\|\chi_{m,1}(\tilde{X}(t)) - \chi_{m,1}(\tilde{X}(t_{m-1}))\| \le C\tau^2,$$

where C is again independent of N and p.

For I_3 in (3.20), we denote

$$\delta\tilde{X}(t) = \tilde{X}(t) - \tilde{X}(t_{m-1}).$$

Note that

$$I_3 = \mathbb{E} Z^1(t_{m-1}) \cdot \nabla\chi_{m,1}(\tilde{X}(t_{m-1})) \cdot \delta\tilde{X}(t) + \frac{1}{2}\mathbb{E} Z^1(t_{m-1}) \cdot (\delta\tilde{X}^{\otimes 2}(t) : \nabla^2\chi_{m,1}(\eta_m)), \tag{3.21}$$

where η_m is some point between $\tilde{X}(t)$ and $\tilde{X}(t_{m-1})$ by the Lagrange Remainder Theorem.

Let us recall that

$$\chi_{m,1}(x) = \frac{1}{p-1}\sum_{j\in\mathcal{C}_\theta, j\neq i} K(x^1 - x^j) - \frac{1}{N-1}\sum_{k:k\neq 1} K(x^1 - x^k). \tag{3.22}$$

Conditioning on $\mathcal{F}_{m-1}$ and applying Hölder's inequality, one has

$$\begin{aligned} &|\mathbb{E} Z^1(t_{m-1}) \cdot \nabla\chi_{m,1}(\tilde{X}(t_{m-1})) \cdot \delta\tilde{X}(t)| \\ &\quad = |\mathbb{E}\Big(Z^1(t_{m-1}) \cdot \mathbb{E}(\nabla\chi_{m,1}(\tilde{X}(t_{m-1})) \cdot \delta\tilde{X}(t)|\mathcal{F}_{m-1})\Big)| \\ &\quad \le \|Z^1(t_{m-1})\|\|\mathbb{E}(\nabla\chi_{m,1}(\tilde{X}(t_{m-1})) \cdot \delta\tilde{X}(t)|\mathcal{F}_{m-1})\| \end{aligned}$$

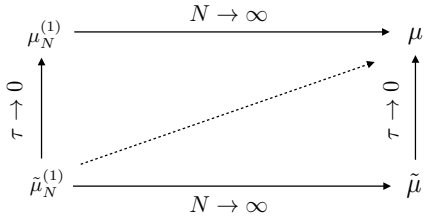

Figure 1: Illustration for the convergence of the RBMs

Applying Lemma 3.3 (equation (3.18)) and Lemma 3.2, this is controlled by

$$C\|Z^1(t_{m-1})\|\tau \le C\|Z^1(t)\| + C\tau^2.$$

The last term in (3.21) can be estimated using Lemma 3.3 (equation (3.19)) directly.

Hence,

$$R(t) \le C\|Z^1(t)\|\tau + C\frac{\tau}{p-1} + C\tau^2$$

and thus

$$\frac{d}{dt}u(t) \le -2(r-2L)u(t) + C\frac{\tau}{p-1} + C\sqrt{u(t)}\tau + C\tau^2.$$

This equality therefore implies that

$$\sup_{t\ge 0}\mathbb{E}|\tilde{X}^1 - X^1|^2 = \sup_{t\ge 0}u(t) \le C\frac{\tau}{p-1} + C\tau^2.$$

The closeness between $\mu_N^{(1)}$ and $\tilde{\mu}_N^{(1)}$ in W_2 distance is a simple application of definition (3.3). □

Finally, it is straightforward to conclude that RBM-1 can approximate the mean field measure (solution to (1.3)) since the mean field result is well known (see, for example, [19]). In fact, as shown in Fig. 1, the triangle inequality yields the following.

Corollary 3.1. *Suppose Assumption 3.1 holds, then*

$$W_2(\tilde{\mu}_N^{(1)}(t), \mu(t)) \le C(\sqrt{\tau} + N^{-1/2+\varepsilon}) \tag{3.23}$$

for any $\varepsilon > 0$.

The error bound between $\tilde{\mu}_N^{(1)}$ and $\mu_N^{(1)}$ is $C\sqrt{\tau}$ uniformly in time with C independent of N. The error bound between $\mu_N^{(1)}$ and μ is given by $CN^{-1/2+\varepsilon}$ ([19]). As shown in the figure, if one takes $N \to \infty$ first, one may get some measure $\tilde{\mu}(t)$, which is the mean field limit of the RBM-1 method. Analyzing this measure will be an interesting problem for the future.

As shown in Fig. 1, the RMBs are a class of Asymptotic-Preserving schemes [37] for particle system (1.2), in that they approximate the particle system (1.2) with an error of $O(\tau)$ independent of N, and as $N \to \infty$, they become good approximations to the limiting mean-field equation (1.3).

Remark 3.1. We also point out that the constant C in Theorem 3.1 stays bounded as $\sigma \to 0$. The proof in Theorem 3.1 is valid for $\sigma = 0$, when there is no Brownian motion. In this case, the first part in Theorem 3.1 in fact says that we can approximate the trajectories of the particles. This is a much stronger approximation.

Remark 3.2. The analysis here should also work for (1.1), provided some analogies of Lemma 3.1 hold (as discussed in section 2.3). For techanical needs, one will assume r large enough so that the semigroup given by the deterministic flow is again a contraction. The detailed analysis is quite involved and we choose not to do it here.

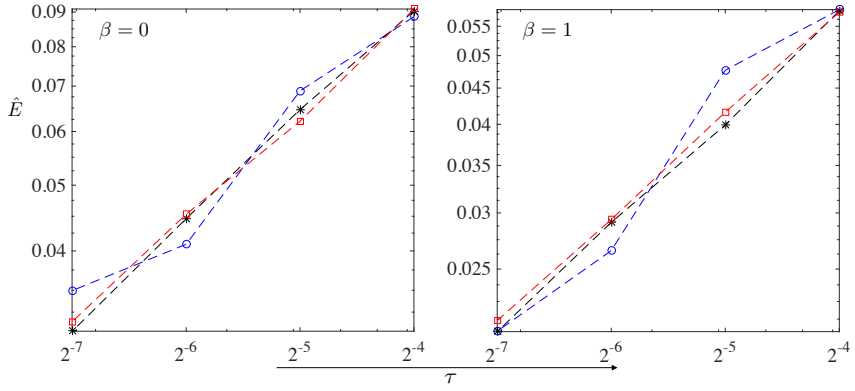

Figure 2: Error of RBM-1 versus time step. Blue circle is for $N = 50$, black star is for $N = 500$, red square is for $N = 2000$.

Remark 3.3. If the deterministic flow does not have the contraction property, we may only be able to prove the convergence of our methods on finite time interval $[0, T]$. Moreover, the mean field limit by Dobrushin's estimate holds on finite time interval [47, 16, 14]. Consequently, it is possible to show that

$$W_2(\tilde{\mu}_N^{(1)}(t), \mu(t)) \leq C(T)(\sqrt{\tau} + N^{-1/2}), \ \forall t \in [0, T], \tag{3.24}$$

for general interaction potentials, where the coefficient depends on T now.

4 Numerical verification

In this section, we run some numerical tests to evaluate the RBMs and verify our theory in section 3. The first example is a simple artificial example to test the dependence of the errors on N and τ. The second example is the Dyson Brownian motion.

4.1 A simple test example

We now use the following simple test example to check how the error in RBM-1 depends on N, τ and T (T is the time point where we compute the numerical solutions). Here, the spatial dimension is 1 ($d = 1$)

$$\dot{X}^i = -\beta X^i + \frac{1}{N-1} \sum_{j:j\neq i} \frac{X^i - X^j}{1 + |X^i - X^j|^2}. \tag{4.1}$$

The interaction is clearly smooth, bounded and with bounded derivatives. Moreover, it has a long-range interaction.

In principle, to evaluate $E(T) = \sqrt{\mathbb{E}|\tilde{X}^1(T) - X^1(T)|^2}$, we need to run many independent experiments and use empirical mean for the approximation. Doing this is clearly very expensive. Alternatively, we only run one experiment and use

$$\hat{E}(T) := \sqrt{\frac{1}{N} \sum_{i=1}^{N} |\tilde{X}^i(T) - X^i(T)|^2} \tag{4.2}$$

to approximate $E(T)$.

In Fig. 2, we show the numerical results for $T = 1$. The initial distribution is taken from

$$\rho_0(x) = \frac{\sqrt{4 - x^2}}{2\pi},$$

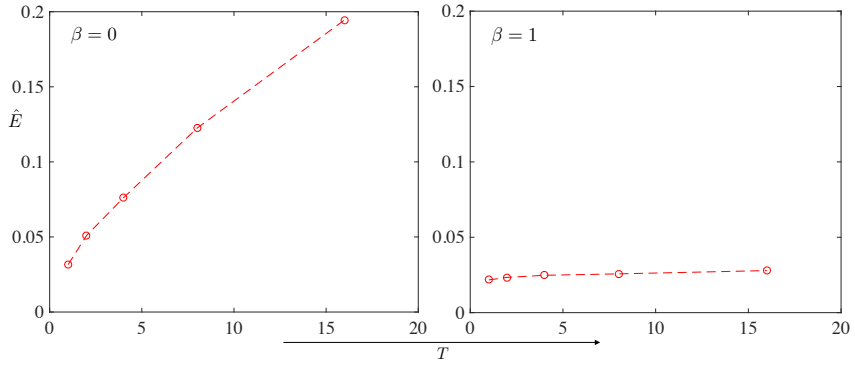

Figure 3: Error for RBM-1 versus time.

by the Metropolis-Hastings MCMC algorithm [44]. The reference solution $X^i(T)$ is obtained by solving the fully coupled system using the forward Euler scheme with $\tau = 2^{-15}$. The solution $\tilde{X}^i(T)$ is generated by RBM-1 with $p = 2$. Each step is solved by the forward Euler method with τ from 2^{-7} to 2^{-4}. We considered $N = 50, 500, 2000$ respectively. We plot the error $\hat{E}(T)$ versus τ for these three N values. The first picture in Fig. 2 is for $\beta = 0$ while the second picture in Fig. 2 is for $\beta = 1$. Clearly, the error is insensitive to the change of N. When N is small, like $N = 50$, the fluctuation in the error is kind of clear. When N is large, in the log-log scale, the curve is already close to straight lines with slope approximately 0.5, meaning that the error indeed decays like $\sqrt{\tau}$.

In Fig. 3, we take $N = 500$. The reference solution is again computed by solving the fully coupled system using the forward Euler scheme with $\tau = 2^{-15}$. The algorithm is performed by taking $\tau = 2^{-7}$. If there is confining potential, the error stays bounded as T increases. However, if there is no confining potential, the error clearly grows, consistent with Remark 3.3. This is indeed natural even for usual ODE discretization for the fully coupled system (1.2). In fact, if there is no confining potential, the numerical error grows with T for the forward Euler method.

4.2 The Dyson Brownian motion

Now, we consider a typical example in random matrix theory [48, 49] to test the difference between RBM-1 and RBM-r. The random matrix we consider is a Hermitian matrix valued Ornstein-Ulenbeck process

$$dA = -\beta A + \frac{1}{\sqrt{N}} dB, \tag{4.3}$$

where the matrix B is a Hermitian matrix consisting of some Brownian motions. In particular, the diagonal elements are independent standard Brownian motions. The off-diagonal elements in the upper triangular half are of the form $\frac{1}{\sqrt{2}}(B_R + iB_I)$ where B_R and B_I are independent standard Brownian motions. The lower triangular half elements are determined using the Hermitian property. By Itô's calculus [50, 48, 49], it can be shown that the eigenvalues of A satisfy the following system of SDEs ($1 \le j \le N$), called the Dyson Brownian motion:

$$d\lambda_j(t) = -\beta\lambda_j(t) + \frac{1}{N} \sum_{k:k\neq j} \frac{1}{\lambda_j - \lambda_k} dt + \frac{1}{\sqrt{N}} dB_j, \tag{4.4}$$

where $\{B_j\}$'s are independent standard Brownian motions. The Brownian motion effect is small when N is large. This system therefore should have similar effects as system (1.2) with $\sigma = 0$. The limiting equation for $N \to \infty$ is given by [51]

$$\partial_t \rho(x,t) + \partial_x(\rho(u - \beta x)) = 0, \;\; u(x,t) = \pi(H\rho)(x,t), \tag{4.5}$$

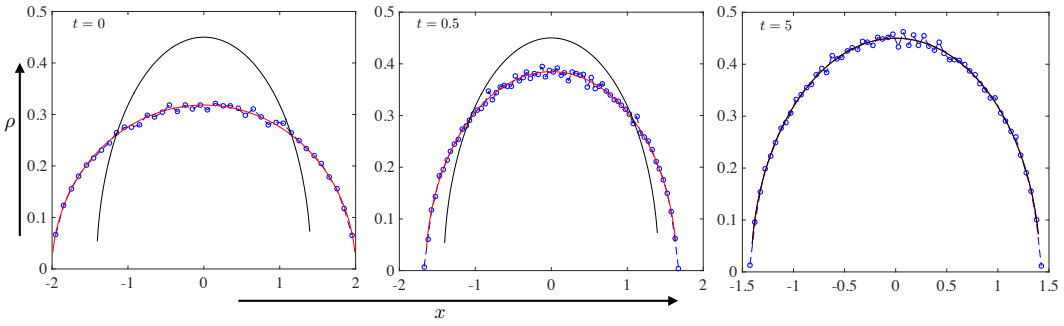

Figure 4: The RMB-1 simulation of the Dyson Brownian motion. The empirical densities at various times are plotted. The red curve is the density distribution predicted by the analytic solution (4.8). The black curve is the equilibrium semicircle law (4.7).

where ρ is the density for λ as $N \to \infty$, $H(\cdot)$ is the Hilbert transform on $\mathbb{R}$, and $\pi = 3.14\ldots$ is the circumference ratio.

Below we consider

$$\beta = 1. \tag{4.6}$$

It is shown that the corresponding limiting equation (4.5) has an invariant measure, given by the semicircle law:

$$\rho(x) = \frac{1}{\pi}\sqrt{2 - x^2} \tag{4.7}$$

To numerically test the behavior of our methods, we note an analytic solution to the limiting equation (4.5)

$$\rho(x,t) = \frac{\sqrt{2\sigma(t) - x^2}}{\sigma(t)\pi}, \ \sigma(t) = 1 + e^{-2t}. \tag{4.8}$$

For each iteration, since the force is singular, we adopted the splitting strategy mentioned in section 2.3. In particular, we define

$$X^{ij} := X^i - X^j. \tag{4.9}$$

The SDE solving step in RBM-1 is given by

- $$Y_m^i = \frac{1}{2}(X_{m-1}^i + X_{m-1}^j) + \mathrm{sgn}(X_{m-1}^{ij})\sqrt{|X_{m-1}^{ij}|^2 + 4\tau},$$
$$Y_m^j = \frac{1}{2}(X_{m-1}^i + X_{m-1}^j) - \mathrm{sgn}(X_{m-1}^{ij})\sqrt{|X_{m-1}^{ij}|^2 + 4\tau}.$$
- $$X_m^i = Y_m^i - \tau Y_m^i + \sqrt{\frac{\tau}{N}}z^i, \ X_m^j = Y^j(t_m) - \tau Y_m^j + \sqrt{\frac{\tau}{N}}z^j.$$

Here, $z^i, z^j \sim \mathcal{N}(0,1)$.

In Fig. 4, we show the numerical results using RBM-1 in section 2.1. The initial data (setting $t = 0$ in (4.8)) are sampled using the Metropolis Hastings algorithm [44]. We plot the results at $t = 0.5$ and $t = 5$. The number of particles is $N = 10^5$ while we use $\tau = 10^{-3}$ for time step. As can be seen, RBM-1 can successfully recover the evolution of distribution and the equilibrium semicircle law (4.7), as desired. In Fig. 5, the results of RBM-r are shown. Again, we take $N = 10^5$ and $\tau = 10^{-3}$. Within one iteration, the same

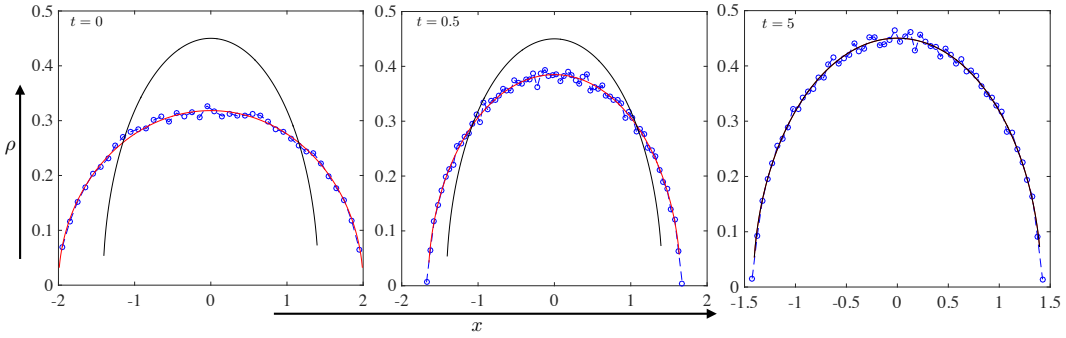

Figure 5: The RBM-r simulation of Dyson Brownian motion. The 'time' is regarded as $\tau = 10^{-3}$ for $N/2$ iterations. The red curve is the density distribution predicted by analytic solution (4.8). The black curve is the equilibrium semicircle law (4.7).

splitting scheme above is used. We find that RBM-r indeed has comparable results with RBM-1. Though RBM-r seemingly cannot simulate the dynamics of the distributions, $N/2$ iterations in fact has comparable behavior for time τ. This interesting observation confirms that RBM-r can capture the dynamics for some examples.

Since both stochastic algorithms give similar behaviors, in later examples, we only use one of them to implement for each example. If we care more about the dynamical behavior, we use RBM-1 (see the two examples in Section 5.2). Otherwise, we use RBM-r. (In fact, the two algorithms do not show significant difference, even for evolutional problems.)

5 Applications

In this section, we apply RBMs to some examples from physics, social and data sciences. On one hand, the positive results give more supports to the algorithms; on the other hand, for some applications, the stochastic algorithms can be regarded as new models for the underlying problems.

5.1 Charged particles on sphere

The traditional Thomson problem is to determine the stable configuration of N electrons on a sphere. When N becomes large, this could lead to the so-called spherical crystals ([52, 53, 54]). The configuration may have some meta-states (local minimizers of the energy surface). When the number of particles is large, the spherical crystals have defects due to the topology of the sphere [53, 54].

In the $N \to \infty$ limit, hopefully, we will have a continuous distribution of charges on the sphere $\rho(\cdot)$. The problem then corresponds to determining ρ such that the energy

$$E(\rho) = \frac{1}{2} \iint_{S\times S} \frac{1}{|x-y|} \rho(x)\rho(y)\, dS_x dS_y \tag{5.1}$$

is minimized. It is unclear how the energies corresponding to local minimizers are distributed (if there are any).

Regarding charges with surface densities, let us make a mathematical remark. Suppose S is a surface that divides the whole space $\mathbb{R}^d$ into two halves. Assume there is a continuous distribution of charges on S with density ρ. Let $\varphi^{\pm}(x)$ be the limits of the potential on the two sides of S, and $\varphi(x) := \varphi^+(x) + \varphi^-(x)$. Then, one has

$$(-\Delta)_S^{1/2}\varphi + s(\varphi) = \rho, \tag{5.2}$$

where $(-\Delta)_S^{1/2}$ is the $1/2$ fractional Laplacian on S and $s(\varphi)$ is some pseudo-differential operators with a symbol of degree lower than 1. In other words, to the leading order, the

1/2 fractional Laplacian of φ equals ρ. In the case that S is a plane or a circle in 2D plane, $s(\varphi) = 0$. In general, $s(\varphi) \neq 0$. In fact, by the jump condition of electric fields,

$$\rho = E^+ \cdot n^+ + E^- \cdot n^- = -\frac{\partial \varphi^+}{\partial n^+} - \frac{\partial \varphi^-}{\partial n^-}. \tag{5.3}$$

It is well-known that the Dirichlet to Neumann operator $\mathcal{L}$ is related to the 1/2-fractional Laplacian by

$$\mathcal{L}\varphi = -(-\Delta)_S^{1/2} f + r(f), \tag{5.4}$$

where n is the normal vector pointing into the side where the harmonic extension is performed and $r(f)$ includes lower terms. In the case that S is a plane $r(f) = 0$ [55]. (In fact, if S is the base of cylinders, $\mathcal{L} = -(-\Delta)_S^{1/2}$ as well. See [56].) In the case of the unit circle in 2D plane, one can refer to [57]. For spheres in higher dimensions, $s \neq 0$. With (5.4) and (5.3), (5.2) follows.

Remark 5.1. For the unit circle in 2D plane, one has a simple way to see (5.4) with $r(\varphi) = 0$. Let $f(z)$ be analytic inside Ω such that $\varphi = \mathrm{Re}(f(z))$. On the boundary, $z = e^{i\theta}$ $f^- = \sum_{k\geq 0} \hat{f}_k e^{ik\theta}$, so that $f(z) = \sum_{k\geq 0} \hat{f}_k z^n$ for z inside the disk. Clearly,

$$\frac{\partial \varphi^-}{\partial n^-} = \mathrm{Re}\left(\frac{\partial f}{\partial n^-}\right) = -\mathrm{Re}(zf'(z))|_{z\in\partial\Omega} = -\mathrm{Re}\left(\sum_{k\geq 0} |k| \hat{f}_k e^{ik\theta}\right).$$

The claim is true for the interior domain. For the exterior domain, one has $\sum_{k\leq 0}$ and $\frac{\partial \varphi^+}{\partial n^+} = \mathrm{Re}(\frac{\partial f}{\partial n^+}) = \mathrm{Re}(zf'(z))|_{z\in\partial\Omega}$. The claim again holds.

Interacting particle systems on sphere can be realized experimentally by beads on water droplets immersed in oil [58]. By adjusting the environmental solution, the interacting kernel $K(\cdot)$ can also be changed, which does not have to be the Coulomb interaction. For such systems, the particles clearly have heat exchange with the enviroment so that the interacting particle system may be described by certain Langevin equations on sphere:

$$\begin{aligned} dX^i &= V^i\, dt, \\ mdV^i &= -\gamma V^i\, dt + \mathbb{P}_S \left(\frac{1}{N}\sum_{j\neq i} F(X^i - X^j)\right) dt + \sqrt{2D} dB_S^i \end{aligned} \tag{5.5}$$

Here, $\mathbb{P}_S$ is the projection onto the tangent space of the sphere while B_S^i is the spherical Brownian motion to guarantee that the particle stays on the sphere. For theories of SDEs on manifolds, one may refer to [59]. Here, D and γ must be related as in the classical fluctuation-dissipation theorem [60].

In the overdamped limit and with suitable scaling, we then have interacting particle system on sphere

$$dX^i = \mathbb{P}_S \left(\frac{1}{N}\sum_{j\neq i} F(X^i - X^j)\right) dt + \sqrt{2D_1} dB_S^i \tag{5.6}$$

Numerically discretizing such SDEs on sphere is an interesting topic which we leave for the future. In this work, we consider the Coulomb interaction with $\sigma_1 = 0$, and use RBM-r as the stochastic strategy. The following simple scheme for the SDE on sphere is then adopted.

- Randomly picking two indices. Then, solve the following for time $t \in [t_{m-1}, t_m)$

$$dX^i = \sum_{j:I(i,j)=1} \frac{X^i - X^j}{|X^i - X^j|^3}\, dt, \tag{5.7}$$

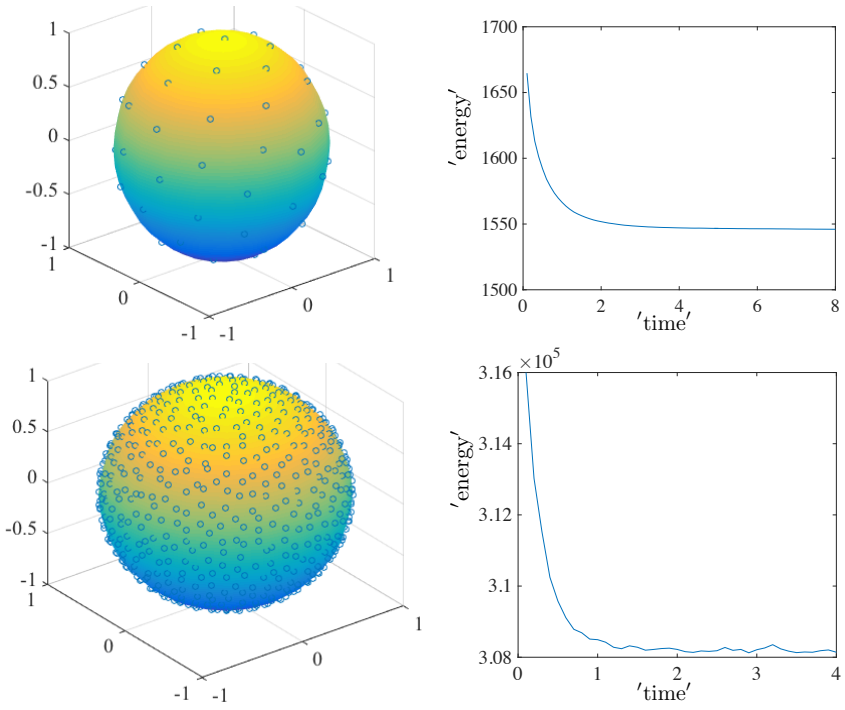

Figure 6: Charged particles on sphere. The first row is for $N = 60$ while the second row corresponds to $N = 800$. The first column shows the distributions at the end of simulation while the second column shows how the energy changes with 'time'.

where $I(i, j) = 1$ means that i, j are in the same batch. This can be solved analytically. In particular, defining

$$\hat{v}_m = (X^i_{m-1} - X^j_{m-1})/|X^i_{m-1} - X^j_{m-1}|, \tag{5.8}$$

one then has:

$$\begin{aligned} X^i_m &= (X^i_{m-1} + X^j_{m-1}) + \hat{v}_m(|X^i_{m-1} - X^j_{m-1}|^3 + 6\tau)^{1/3}, \\ X^j_m &= (X^i_{m-1} + X^j_{m-1}) - \hat{v}_m(|X^i_{m-1} - X^j_{m-1}|^3 + 6\tau)^{1/3} \end{aligned} \tag{5.9}$$

- Project the obtained points back to the sphere by dividing its magnitude.

The reason for setting $\sigma_1 = 0$ is that we would like to explore energy stable configurations. We desire low temperature regime for the ground state. Besides, the stochastic algorithm also introduces randomness so that we still have chance to get out of the local minimizers.

To check whether the method can give the desired ground state approximately, we randomly choose initial points on the sphere and run the above stochastic algorithms for enough iterations. We do many experiments and check whether we always obtain the same final energy level.

In Fig. 6, we show the numerical results in two experiments. The number of particles are chosen as $N = 60$ and $N = 800$ respectively. The initial points are chosen randomly. The time step is chosen as $\tau = 10^{-4}$. As before, we regard the 'time' to be τ after $N/2$ iterations. For $N = 60$, we see that in the eventual near stable configuration, each particle has 5 or 6 neighbors, and this agrees with the known results by physicists [53, 54]. This configuration is quite different from the fullerene C_{60} structure which is induced by the special properties of Carbon atoms. For the $N = 800$ case, the particles are roughly distributed uniformly. For both figures, there is only one stable energy level during the whole process. This means the system was only trapped in the final stable configuration.

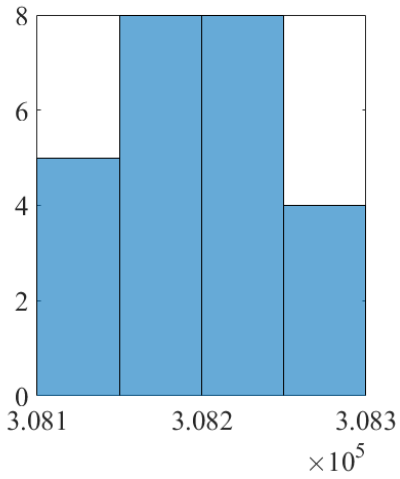

Figure 7: Energy statistics. This histogram shows the terminal energy for 25 experiments. The terminal 'time' is $T = 3$.

To check whether there are other possible stable energy configurations, we collect in Fig. 7 the energies for $N = 800$ after $3 * (N/2)/\tau$ iterations ($T = 3$) in 25 experiments. The simulation shows that one can find the ground state of the configuration almost surely using the stochastic algorithm and it is not easy to be trapped in local minimizers, if there are any. As studied by the physicists, there are many energy levels for the Thomson sphere. However, the numerical results here seem to suggest that the stochastic algorithms can obtain the ground state with high probability and the local minimizers of the energy landscape probably has small energy barriers. Maybe, some interesting phenomena happen for large N's which needs further investigation.

5.2 Two examples from economics and social science

In this section, we apply RBM-1 for two important models in social sciences, namely the evolution of wealth [61] and opinion dynmaics [5]. The obtained stochastic processes not only are algorithms for the original models, but also can be viewed as new models which consider the fact that only a few individuals commute during a short time.

5.2.1 Stochastic dynamics of wealth

We consider the model proposed by Degond et al. [61], which tries to understand the evolution of N market agents with two attributes: the economic configuration X^i and its wealth Y^j.

$$\begin{aligned} \dot{X}^i &= V(X^i, Y^i), \\ dY^i &= -\frac{1}{N}\sum_{k:k\neq i} \xi_{ik}\Psi(|X^i - X^k|)\partial_y\phi(Y^i - Y^k)\,dt + \sqrt{2D}Y^i dB^i. \end{aligned} \tag{5.10}$$

The first equation describes the evolution of the economic configuration, which is driven by the local Nash equilibrium and it is related to mean-field games [62, 11]. The second equation describes the evolution of the wealth, which contains two mechanisms: the trading model proposed by Bouchaud and Mezart [63], and the geometric Brownian motion in finance proposed by Bachelier in 1900 [64]. The quantity $\sqrt{2D}$ is the volatility. The function ϕ is the trading interaction potential, while $\xi_{ik}\Psi(|X^i - X^k|)$ is the trading frequency. Often one assumes that ξ_{ik} depends on the number of trading agents in the economic neighborhoods

of i and k:

$$\xi_{ik} = \xi(\frac{\rho^{i,\Psi} + \rho^{k,\Psi}}{2}),\ \rho^{i,\Psi} = \frac{1}{N-1}\sum_{\ell \neq j} \Psi(|X^\ell - X^i|). \tag{5.11}$$

The mean field Fokker-Planck equation is given by

$$\partial_t f + \partial_x(V(x,y)f) + \partial_y(Ff) = D\partial_{yy}(y^2 f) \tag{5.12}$$

where

$$F(x,y,t) = -\int_{x'\geq 0, y'\geq 0} \xi(\frac{1}{2}(\rho^\Psi(x,t) + \rho^\Psi(x',t)))\Psi(|x-x'|) \times \partial_y\phi(y-y')f(x',y',t)\,dx'dy', \tag{5.13}$$

and

$$\rho^\Psi(x,t) = \int_{x'>0, y'>0} \Psi(|x-x'|)f(x',y',t)dx'dy'.$$

Now, if one considers the homogeneous case where the wealth dynamics is independent of the position in the economic configuration space, then Ψ is a constant. In this case, the dynamics is reduced to the interacting particle system, except that one has *multiplicative* noise

$$dY^i = -\frac{\kappa}{N}\sum_{k:k\neq i} \partial_y\phi(Y^i - Y^k)\,dt + \sqrt{2D}Y^i dB^i, \tag{5.14}$$

where $\kappa := \Psi\xi\left(\frac{1}{2}(\rho^{i\Psi} + \rho^{k,\Psi})\right)$ is now a constant. The mean field equation is now given by

$$\partial_t\rho(y) = -\partial_y(F(y)\rho(y)) + D\partial_{yy}(y^2\rho(y)), \tag{5.15}$$

where

$$F(y) = -\kappa\int_{y\geq 0} \partial_y\phi(y-y')\rho(y')dy'.$$

The equilibrium distribution is given by

$$\rho_\infty(y) \propto \exp(-\frac{\alpha(y)}{D}),$$

where α satisfies

$$\partial_y\alpha(y) = -\frac{1}{y^2}F(y) + \frac{2D}{y}.$$

We now apply RBM-1 with $p = 2$ to (5.14) and have for $t \in [t_{m-1}, t_m)$

$$dY^i = -\kappa\partial_y\phi(Y^i - Y^\theta)\,dt + \sqrt{2D}Y^i dB^i,\ i = 1,\ldots,N, \tag{5.16}$$

where θ is a random index that is grouped with i in the random division. In some sense, the stochastic dynamics described by this algorithm can model what is happening in the real world: each agent only trades with a small number of random agents at a time. Then, after some time interval, the agents trade with others. Hence, (5.16) is not just an aglorithm but also it can be viewed as a *new model*.

For numerical test, choose the quadratic trading interaction as in [61, section 3.4]

$$\phi(y) = \frac{1}{2}y^2.$$

This interaction function may not be practical as it increases with y (intuitively, as $y \to \infty$, it should go to zero). The good thing is that with this interaction function, one can find the equilibirum distribution of wealth for (5.14) as

$$\rho_\infty(y) = \frac{(\kappa\eta/D)^{\kappa/D+1}}{\Gamma(\kappa/D+1)}y^{-(2+\kappa/D)}\exp(-\frac{\kappa\eta}{Dy})1_{y>0}. \tag{5.17}$$

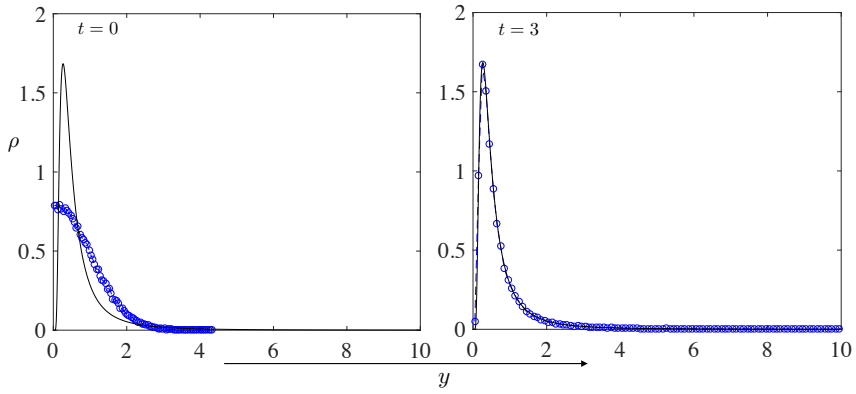

Figure 8: Wealth distribution for stochastic dynamics

This distribution is the inverse Gamma distribution and agrees with the Pareto power law for large y. Here, η is the mean wealth.

Now, we take $\kappa = D = 1$ and consider the random dynamics (5.16):

$$dY^i = -(Y^i - Y^\theta)\,dt + \sqrt{2}Y^i dB^i,\ i = 1, \dots, N.$$

In Fig. 8, we plot the empirical distribution of the wealth with $N = 10^5$ agents. We choose $\tau = 10^{-3}$ and do the simulation to $T = 3$. The SDE again is solved by the splitting strategy. Note that the splitting scheme preserves the mean wealth. For the test, we choose initial data as $X^i = |Y^i|$ with $Y^i \sim \mathcal{N}(0, 1)$. The reference curve is (5.17) with

$$\eta = \frac{\sqrt{2}}{\sqrt{\pi}}. \tag{5.18}$$

Clearly, the numerical results agree perfectly with the expected wealth distribution.

5.2.2 Stochastic opinion dynamics

In this section, we consider some stochastic revisions of the opinion dynamics in [5], where the following two models are mentioned

$$\frac{d}{dt}X^i = \alpha\frac{1}{N}\sum_{j\neq i}\phi(|X^j - X^i|)(X^j - X^i) \tag{5.19a}$$

and

$$\frac{d}{dt}X^i = \alpha\sum_{j\neq i}\frac{\phi(|X^j - X^i|)}{\sum_k \phi(|X^k - X^i|)}(X^j - X^i). \tag{5.19b}$$

Here, ϕ is called the influence function. These models are introduced for the emergence of consensus of opinions.

Here, (5.19b) is not convenient for stochastic algorithms because of the denominator. Instead, we consider RBM-1 applied to (5.19a), and have the following:

$$\frac{d}{dt}X^i = \alpha\phi(|X^\theta - X^i|)(X^\theta - X^i) + \varepsilon_N dB^i, \tag{5.20}$$

where θ, again as before, is a random index that is fixed for $t \in [t_{m-1}, t_m)$. If $\varepsilon_N = 0$, this is the stochastic algorithm of (5.19a) directly. The parameter ε_N is to represent the random fluctuation on its opinions. If the number of agents is large, we believe this should be small. In fact, this stochastic algorithm seems closer to what is happening in the world:

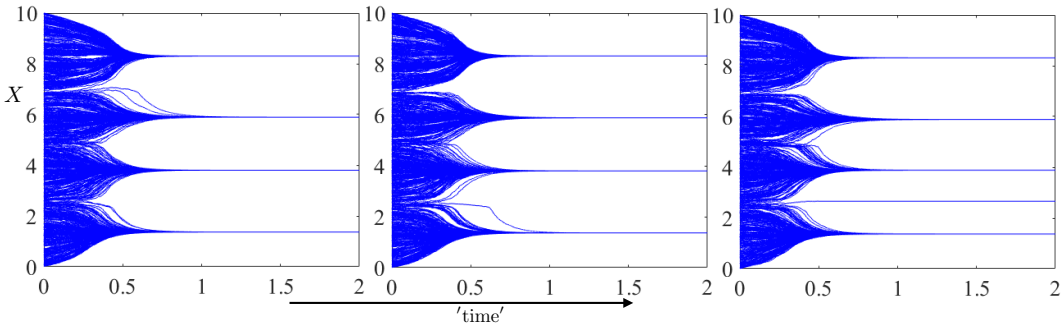

Figure 9: Stochastic opinion dynamics versus time. Plots of three experiments of the stochastic dynamics with the same initial data. No Brownian motion ($\varepsilon_N = 0$).

each person only talks to one or few people at a time. Combining its previous opinion, it forms a new opinion.

Let us consider the influence function

$$\phi(r) = \chi_{[0,1]} \tag{5.21}$$

as in [5, Section 3.3].

In Fig. 9, we show the numerical results. The figure shows the results of three experiments with the same initial data. We choose $N = 10^3$ and $\tau = 10^{-4}$, $\alpha = 40$. Clearly, the final stationary consensus is sensitive with respect to initial distribution. With the same initial data, though the dynamics is stochastic, the main behavior is the same for the three experiments. There are four main clusters of consensus. However, interestingly, in some experiments (like the third picture), there are may be some individuals that do not belong to any cluster, which seems to be the case in real world: some individuals are isolated at the early stage, and after the main clusters of consensus form, they are not affected by these groups since they are so different. The randomness introduced by the algorithms does not quite affect the main clusters of consensus, and only a few individuals might behave differently due to the randomness. After certain time, when the clusters of consensus are formed, the randomness does not play any roles any more: the individuals only talk to members in their own clusters.

In Fig. 10, we show the numerical results for the stochastic opinion dynamics with Brownian motion $\varepsilon_N = \frac{1}{N^{1/3}}$. The N, τ, α parameters are the same. The evolution of clusters of consensus is roughly the same with or without the Brownian motion. However, Brownian motions indeed introduce fluctuation of opionions within the clusters. This means that the fluctuation is not very important when the main clusters of opinions are formed.

5.3 Clustering through interacting particle system

In this section, we consider using the interacting particle system (1.2) for clustering, as discussed in [65]. The idea is like this: consider N particles with a given adjacent matrix $A = (a_{ij})$, $a_{ij} \geq 0$. Then, we construct the interacting particle system as

$$\frac{d}{dt}X^i = \alpha \frac{1}{N-1} \sum_{j:j\neq i} (a_{ij} - \beta)(X^j - X^i) \tag{5.22}$$

where α and β are some parameters. This is designed such that the particles with positive $a_{ij} - \beta$ attract with each other so that they tend to gather together, while those with negative $a_{ij} - \beta$ repel each other so that they separate. The hoping is that the intrinsic clusters will emerge automatically.

With the RBM-r, the computational cost is significantly reduced and this then becomes a practical method. Each time, we pick a random set $\mathcal{C}$ that contains $p = 2$ elements. The

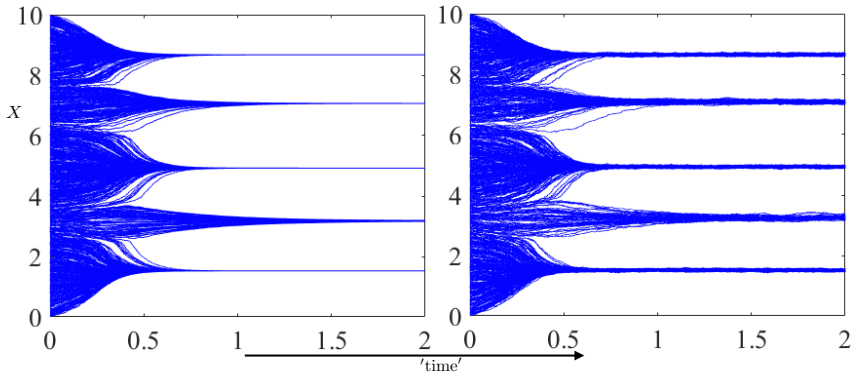

Figure 10: Opinion dynamics versus time. The first figure has no Brownian motion. The second is with $\varepsilon_N = N^{-1/3}$. The two figures are with the same initial data.

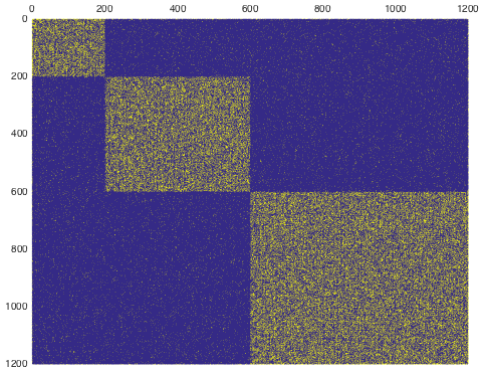

Figure 11: Clustering: an adjacent matrix for the stochastic block model

dynamics we consider is then

$$\frac{d}{dt}X^i = \alpha(a_{i\theta} - \beta)(X^\theta - X^i) \tag{5.23}$$

where θ is again the random index that is constant on $[t_{m-1}, t_m)$. Clearly, the random batch can be picked from those with nonzeros of a_{ij} only to improve the efficiency.

The update formula is

$$\begin{aligned} X^i_m &= [(X^i_{m-1} + X^j_{m-1}) + (X^i_{m-1} - X^j_{m-1})\exp(-2\alpha(a_{ij} - \beta)\tau)]/2; \\ X^j_m &= [(X^i_{m-1} + X^j_{m-1}) - (X^i_{m-1} - X^j_{m-1})\exp(-2\alpha(a_{ij} - \beta)\tau)]/2. \end{aligned}$$

5.3.1 cluster for stochastic block model

Let us consider the stochastic block model ([66, 67]). The model is like this: suppose there are k clusters. For i, j in the same cluster, $\mathbb{P}(a_{ij} = 1) = p$ and $\mathbb{P}(a_{ij} = 0) = 1-p$. Otherwise, $\mathbb{P}(a_{ij} = 1) = q$ and $\mathbb{P}(a_{ij} = 0) = 1 - q$. The entries a_{ij} are assumed to be independent. We assume that we only know the adjacent matrix in one experiment (if we know the matrices for several experiments, we can then combine them to get more accurate clusters). Clearly, the adjacent matrix is noisy. We are going to test whether or not we can still recover the clusters using the noisy adjacent matrix.

In Fig. 11, we show the adjacent matrix from one experiment. In this example, we have $N = 1200$ particles. The first 200 particles are designed to be in the first cluster, the next

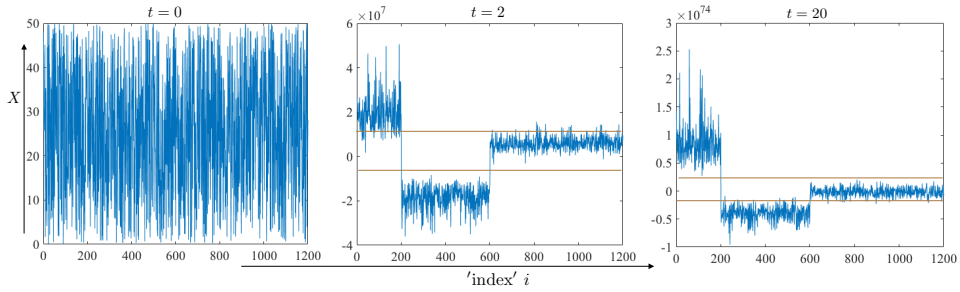

Figure 12: Clustering using a noisy adjacent matrix from the stochastic block model. The plots show the positions of the particles at different 'time' points. The desired clusters can be recovered with few mistakes.

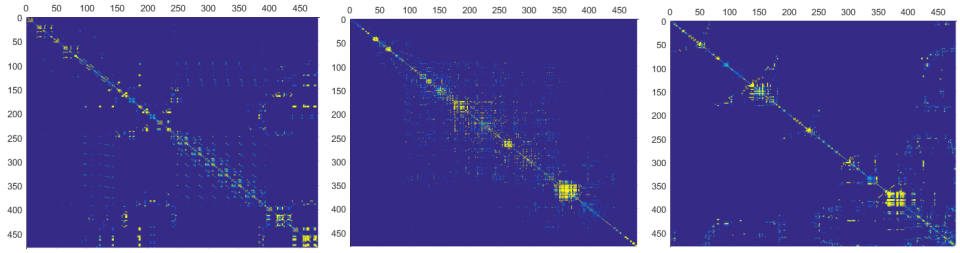

Figure 13: A sparse adjacent matrix: clustering and reordering

400 are in the second cluster and the last 600 are in the third cluster. The probabilities are chosen as

$$p = 0.7,\ q = 0.3.$$

In the matrix, yellow dots mean 1 while blue dots mean 0. As expected, most of the entries in the off-diagonal blocks are 0 with some yellow 'dust' scattered inside them. Most of the entries in the diagonal blocks are 1 with blue dots inside them.

In the experiment, we set $\beta = \frac{1}{2}$ and $\alpha = 40$, $\tau = 10^{-3}$. We initialize their positions randomly on $[0, 50]$. The numerical results in an experiment are shown in Fig. 12. Again the "time" is regarded as τ after $N/2$ iterations. From the figure, it is clear that the clusters can be recovered correctly though the adjacent matrix is noisy.

5.3.2 Reordering for sparse matrix

As another example, let us consider reordering sparse matrices as a byproduct of the clustering. The point is that large a_{ij} entry tends to group the two indices together. If we use the terminal X_i's to sort, the reordered matrix can have large entries near the diagonal. If there are several distinct clusters, we will then have diagonal block matrix.

Consider the matrix given by

$$A_1 = BB^T + I,$$

where B is the 'west0479' matrix, which is a sparse matrix in the standard database of MATLAB. Consequently, A_1 is a sparse matrix. Since matrix A_1 can have negative entries, we define $A = (a_{ij})$ with $a_{ij} = |A_1(i, j)|$ to get a suitable adjacent matrix. Since A is sparse, we do sampling over the nonzero entries only.

Fig. 13 shows the matrices $\min(P, 10)$ ("min" here means the entry-wise minimum), where the meaning of P is as follows: in the first figure, P is A; in the second figure, P is the reordered matrix using our strategy at $T = 5$; P in the third figure is the matrix with approximated minimal degree ordering. Clearly, our strategy gathers all the big entries near the diagonal, which means particles with strong interactions indeed form clusters.

Regarding the ordering resulted from our clustering strategy, it might not be good using the criterion from sparse matrix theory. For example, in the Cholesky decomposition, there are 37896 nonzero entries. Meanwhile, by the approximated minimal degree ordering, there are 14493 nonzero entries in the Cholesky decomposition. Anyhow, clustering big entries near diagonals may be advantageous for some applications.

6 Conclusions

We have developed Random Batch Methods for interacting particle systems with large number of particles and they reduce the computational cost significantly for $N (N \gg 1)$ particles from $O(N^J)$ to $O(N)$ per time step. For RBM-1, the method without replacement, we have given a particle number independent error estimate under some special interactions. We have applied these methods to some representative problems of binary interactions in math, physics, social science and data science, and numerical results have supported our theory and expectations. The random algorithms are powerful for systems with large number of individuals and high dimensions.

As well accepted, in stochastic gradient descent, adding momentum could help to find flatter minimizers and improve results. In other words, the Langevin dynamics seems better for optimization and sampling [68, 69]. Hence, considering the interacting particles with mass might be better for sampling the invariant measure of the nonlinear Fokker-Planck equation. This is left for future research. Besides, there are many interesting projections ahead, for example, proof of convergences for more general external and interacting potentials, and for RMBs with replacements. It is also interesting to develop similar particle methods for the mean field equations, whenever they are available, as was done in [35] but for more general mean field equations.

Acknowledgement

S. Jin was partially supported by the NSFC grant No. 31571071. The work of L. Li was partially supported by a starting up grant from Shanghai Jiao Tong University. The work of J.-G. Liu was partially supported by KI-Net NSF RNMS11-07444 and NSF DMS-1812573.

References

[1] 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.

[2] 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.

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

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

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

[6] B. Øksendal. *Stochastic differential equations: an introduction with applications*. Springer, Berlin, Heidelberg, sixth edition, 2003.

[7] J. Fröhlich and B. Zegarlinski. Some comments on the Sherrington-Kirkpatrick model of spin glasses. *Communications in Mathematical Physics*, 112(4):553–566, 1987.

[8] S. Franz and F. L. Toninelli. Finite-range spin glasses in the Kac limit: free energy and local observables. *Journal of Physics A: Mathematical and General*, 37(30):7433, 2004.

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

[10] 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.

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

[12] M. Kac. Foundations of kinetic theory. In *Proceedings of the Third Berkeley Symposium on Mathematical Statistics and Probability*, volume 3, pages 171–197. University of California Press Berkeley and Los Angeles, California, 1956.

[13] H. P. McKean. Propagation of chaos for a class of non-linear parabolic equations. *Stochastic Differential Equations (Lecture Series in Differential Equations, Session 7, Catholic Univ., 1967)*, pages 41–57, 1967.

[14] S Benachour, B Roynette, D Talay, and P Vallois. Nonlinear self-stabilizing processes–i existence, invariant probability, propagation of chaos. *Stochastic processes and their applications*, 75(2):173–201, 1998.

[15] J.-G. Liu and R. Yang. Propagation of chaos for large Brownian particle system with coulomb interaction. *Research in the Mathematical Sciences*, 3(1):40, 2016.

[16] P.-E. Jabin and Z. Wang. Quantitative estimates of propagation of chaos for stochastic systems with $W^{-1,\infty}$ kernels. *Inventiones mathematicae*, to appear.

[17] S. Serfaty and M. Duerinckx. Mean field limit for Coulomb flows. *arXiv preprint arXiv:1803.08345*, 2018.

[18] L. Li, J.-G. Liu, and P. Yu. On mean field limit for Brownian particles with Coulomb interaction in 3D. *arXiv preprint arXiv:1811.07797*, 2018.

[19] 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.

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

[21] S. Bubeck. Convex optimization: Algorithms and complexity. *Foundations and Trends® in Machine Learning*, 8(3-4):231–357, 2015.

[22] Jürgen Schmidhuber. Deep learning in neural networks: An overview. *Neural networks*, 61:85–117, 2015.

[23] Y. Nesterov. Efficiency of coordinate descent methods on huge-scale optimization problems. *SIAM Journal on Optimization*, 22(2):341–362, 2012.

[24] S. J. Wright. Coordinate descent algorithms. *Mathematical Programming*, 151(1):3–34, 2015.

[25] 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.

[26] Y.-A. Ma, T. Chen, and E. Fox. A complete recipe for stochastic gradient mcmc. In *Advances in Neural Information Processing Systems*, pages 2917–2925, 2015.

[27] T. Nagapetyan, A. B. Duncan, L. Hasenclever, S. J. Vollmer, L. Szpruch, and K. Zygalakis. The true cost of stochastic gradient Langevin dynamics. *arXiv preprint arXiv:1706.02692*, 2017.

[28] N. Halko, P.-G. Martinsson, and J. A. Tropp. Finding structure with randomness: Probabilistic algorithms for constructing approximate matrix decompositions. *SIAM review*, 53(2):217–288, 2011.

[29] G. A. Bird. Approach to translational equilibrium in a rigid sphere gas. *The Physics of Fluids*, 6(10):1518–1519, 1963.

[30] G. A. Bird. *Molecular Gas Dynamics and the Direct Simulation Monte Carlo of Gas Flows*. Clarendon, Oxford, 1994.

[31] W. Wagner. A convergence proof for bird's direct simulation monte carlo method for the boltzmann equation. *Journal of Statistical Physics*, 66(3-4):1011–1044, 1992.

[32] Kenichi Nanbu. Direct simulation scheme derived from the boltzmann equation. i. monocomponent gases. *Journal of the Physical Society of Japan*, 49(5):2042–2049, 1980.

[33] Hans Babovsky and Reinhard Illner. A convergence proof for nanbu¡¯s simulation method for the full boltzmann equation. *SIAM journal on numerical analysis*, 26(1):45–65, 1989.

[34] 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.

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

[36] 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.

[37] S. Jin. Asymptotic preserving (AP) schemes for multiscale kinetic and hyperbolic equations: a review. *Lecture notes for summer school on methods and models of kinetic theory (M&MKT), Porto Ercole (Grosseto, Italy)*, pages 177–216, 2010.

[38] R. Durstenfeld. Algorithm 235: random permutation. *Communications of the ACM*, 7(7):420, 1964.

[39] Q. Li, C. Tai, and W. E. Stochastic modified equations and adaptive stochastic gradient algorithms. In *International Conference on Machine Learning*, pages 2101–2110, 2017.

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

[41] W. Hu, C. J. Li, L. Li, and J.-G. Liu. On the diffusion approximation of nonconvex stochastic gradient descent. *Ann. Math. Sci. Appl.*, 2018. arXiv preprint arXiv:1705.07562.

[42] K. Chen, Q. Li, and J.-G. Liu. Online learning in optical tomography: a stochastic approach. *Inverse Problems*, 34(7):075010, 2018.

[43] J.-G. Liu and R. Yang. A random particle blob method for the Keller-Segel equation and convergence analysis. *Mathematics of Computation*, 86(304):725–745, 2017.

[44] W. K. Hastings. *Monte Carlo Sampling Methods Using Markov Chains and Their Applications*. Oxford University Press, 1970.

[45] W. R Gilks, S. Richardson, and D. Spiegelhalter. *Markov chain Monte Carlo in practice*. Chapman and Hall/CRC, 1995.

[46] F. Santambrogio. Optimal transport for applied mathematicians. *Birkäuser, NY*, pages 99–102, 2015.

[47] F. Golse. The mean-field limit for the dynamics of large particle systems. *Journées équations aux dérivées partielles*, 9:1–47, 2003.

[48] T. Tao. *Topics in random matrix theory*, volume 132. American Mathematical Soc., 2012.

[49] L. Erdos and H.-T. Yau. Dynamical approach to random matrix theory. *Courant Lecture Notes in Mathematics*, 28, 2017.

[50] F. J. Dyson. A Brownian-motion model for the eigenvalues of a random matrix. *Journal of Mathematical Physics*, 3(6):1191–1198, 1962.

[51] Y. Gao, Y. Gao, and J.-G. Liu. Large time behavior, bi-Hamiltonian structure and kinetic formulation for complex Burgers equation. *arXiv preprint arXiv:1812.09699*, 2018.

[52] M. Bowick, A. Cacciuto, D. R. Nelson, and A. Travesset. Crystalline order on a sphere and the generalized Thomson problem. *Physical Review Letters*, 89(18):185502, 2002.

[53] Z. Yao. Dressed active particles in spherical crystals. *Soft matter*, 12(33):7020–7027, 2016.

[54] Z. Yao. Topological vacancies in spherical crystals. *Soft matter*, 13(35):5905–5910, 2017.

[55] L. Caffarelli and L. Silvestre. An extension problem related to the fractional Laplacian. *Communications in partial differential equations*, 32(8):1245–1260, 2007.

[56] X. Cabré and J. Tan. Positive solutions of nonlinear problems involving the square root of the Laplacian. *Advances in Mathematics*, 224(5):2052–2093, 2010.

[57] P. L. De Nápoli and P. R. Stinga. Fractional Laplacians on the sphere, the Minakshisundaram zeta function and semigroups. *arXiv preprint arXiv:1709.00448*, 2017.

[58] AR Bausch, MJ Bowick, A Cacciuto, AD Dinsmore, MF Hsu, DR Nelson, MG Nikolaides, A Travesset, and DA Weitz. Grain boundary scars and spherical crystallography. *Science*, 299(5613):1716–1718, 2003.

[59] E. P. Hsu. *Stochastic analysis on manifolds*, volume 38. American Mathematical Soc., 2002.

[60] H. B. Callen and T. A. Welton. Irreversibility and generalized noise. *Physical Review*, 83(1):34, 1951.

[61] P. Degond, J.-G. Liu, and C. Ringhofer. Evolution of the distribution of wealth in an economic environment driven by local Nash equilibria. *Journal of Statistical Physics*, 154(3):751–780, 2014.

[62] P. Degond, J.-G. Liu, and C. Ringhofer. Large-scale dynamics of mean-field games driven by local Nash equilibria. *Journal of Nonlinear Science*, 24(1):93–115, 2014.

[63] J.-P. Bouchaud and M. Mézard. Wealth condensation in a simple model of economy. *Physica A: Statistical Mechanics and its Applications*, 282(3-4):536–545, 2000.

[64] L Bachelier. Théorie de la spéculation. In *Annales scientifiques de l'École Normale Supérieure*, volume 17, pages 21–86, 1900.

[65] D. Fang, S.-Y. Ha, and S. Jin. Emergent behaviors of the Cucker-Smale ensemble under attractive-repulsive couplings and Rayleigh frictions. *arXiv preprint arXiv:1808.03980*, 2018.

[66] P. W Holland, K. B. Laskey, and S. Leinhardt. Stochastic blockmodels: First steps. *Social networks*, 5(2):109–137, 1983.

[67] A. Damle, V. Minden, and L. Ying. Robust and efficient multi-way spectral clustering. *arXiv preprint arXiv:1609.08251*, 2016.

[68] T. Lelievre and G. Stoltz. Partial differential equations and stochastic methods in molecular dynamics. *Acta Numerica*, 25:681–880, 2016.

[69] X. Cheng, N. S. Chatterji, P. L. Bartlett, and M. I. Jordan. Underdamped Langevin MCMC: A non-asymptotic analysis. *arXiv preprint arXiv:1707.03663*, 2017.