

# Random Batch Methods for classical and quantum interacting particle systems and statistical samplings

Shi Jin <sup>\*1</sup> and Lei Li <sup>†1</sup>

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

## Abstract

We review the Random Batch Methods (RBM) for interacting particle systems consisting of  $N$ -particles, with  $N$  being large. The computational cost of such systems is of  $\mathcal{O}(N^2)$ , which is prohibitively expensive. The RBM methods use small but random batches so the computational cost is reduced, per time step, to  $\mathcal{O}(N)$ . In this article we discuss these methods for both classical and quantum systems, the corresponding theory, and applications from molecular dynamics, statistical samplings, to agent-based models for collective behavior, and quantum Monte-Carlo methods.

## Contents

<b>1</b>	<b>Introduction</b>	<b>2</b>
<b>2</b>	<b>The RBM methods</b>	<b>3</b>
2.1	The RBM algorithms . . . . .	3
2.2	Convergence analysis . . . . .	4
2.3	An illustrating example: wealth evolution . . . . .	6
<b>3</b>	<b>The mean-field limit</b>	<b>7</b>
<b>4</b>	<b>Molecular dynamics</b>	<b>10</b>
4.1	RBM with kernel splitting . . . . .	12
4.2	Random Batch Ewald: an importance sampling in the Fourier space . . . . .	12
<b>5</b>	<b>Statistical sampling</b>	<b>16</b>
5.1	Random Batch Monte Carlo for many body systems . . . . .	17
5.2	RBM-SVGD: a stochastic version of stein variational gradient descent . . . . .	19
<b>6</b>	<b>Agent-based models for collective dynamics</b>	<b>21</b>
6.1	The Cucker-Smale model . . . . .	22
6.2	Consensus models . . . . .	23
<b>7</b>	<b>Quantum dynamics</b>	<b>24</b>
7.1	A theoretical result on the $N$ -body Schrödinger equation . . . . .	24
7.1.1	Mathematical Setting and Main Result . . . . .	25
7.2	Quantum Monte-Carlo methods . . . . .	27
7.2.1	The Random Batch Method for VMC . . . . .	29
7.2.2	The Random Batch Method for DMC . . . . .	30

---

\*shijin-m@sjtu.edu.cn

†leili2010@sjtu.edu.cn

# 1 Introduction

Interacting particle systems arise in a variety of important phenomena in physical, social, and biological sciences. They usually take the form of Newton’s second law that governs the interactions of  $N$ -particles under interacting forces that vary depending on different applications. Such systems are important in physics—from electrostatics to astrophysics, in chemistry and material sciences—such as molecular dynamics, in biological and social sciences—such as agent based models in swarming [93, 16, 15, 21], chemotaxis [51, 8], flocking [19, 46, 2], synchronization [17, 45] and consensus [81]).

These interacting particle systems can be described in general by the first order systems

$$d\mathbf{r}_i = b(\mathbf{r}_i) dt + \alpha_N \sum_{j:j \neq i} K(\mathbf{r}_i - \mathbf{r}_j) dt + \sigma d\mathbf{W}_i, \quad i = 1, 2, \dots, N, \quad (1.1)$$

or the second order systems

$$\begin{aligned} d\mathbf{r}_i &= \mathbf{v}_i dt, \\ d\mathbf{v}_i &= \left[ b(\mathbf{r}_i) + \alpha_N \sum_{j:j \neq i} K(\mathbf{r}_i - \mathbf{r}_j) - \gamma \mathbf{v}_i \right] dt + \sigma d\mathbf{W}_i. \end{aligned} \quad (1.2)$$

We use  $\mathbf{r}_i \in \mathbb{R}^d$  to denote the labels for the particles. We will loosely call  $\mathbf{r}_i$  the “locations” or “positions”, and  $\mathbf{v}_i$  the velocities of the particles, though the specific meaning can be different in different applications. The function  $K(\cdot)$  and  $b(\cdot) : \mathbb{R}^d \rightarrow \mathbb{R}^d$  are the interaction kernel and some given external field respectively. The stochastic processes  $\{\mathbf{W}^i\}_{i=1}^N$  are i.i.d. Wiener processes, or the standard Brownian motions. If  $\gamma = \sigma = 0$  and  $b = -\nabla V$  for some potential  $V$ , one has a Hamiltonian system in classical mechanics. For the molecules in the heat bath [61, 14],  $\mathbf{r}_i$  and  $\mathbf{v}_i$  are the physical positions and velocities, described by the underdamped Langevin equations, where  $\sigma$  and  $\gamma$  satisfy the so-called “fluctuation-dissipation relation”

$$\sigma = \frac{\tilde{E}}{2\gamma/\beta}, \quad (1.3)$$

where  $\beta$  is the inverse of the temperature (we assume all the quantities are scaled and hence dimensionless so that the Boltzmann constant is absent). The first order system (1.1) can be viewed as the overdamped limit (when  $\gamma \rightarrow \infty$  and the time rescaled as  $\gamma t$ ) of the second order systems (1.2).

In the case  $\alpha_N = \frac{1}{N-1}$ , as  $N \rightarrow \infty$ , the dynamics of the so-called mean field limit of (1.1) is given by the nonlinear Fokker-Planck equation [40, 77]

$$\partial_t \mu = -\nabla \cdot ((b(x) + K * \mu)\mu) + \frac{1}{2} \sigma^2 \Delta \mu, \quad (1.4)$$

where  $\mu(dx) \in \mathbf{P}(\mathbb{R}^d)$ . The regime  $\alpha_N = c/N + o(1)$  is thus naturally called the mean-field regime. Correspondingly, the mean-field limit of the second order system (1.2) is

$$\partial_t f = -\nabla_x \cdot (vf) - \nabla_v \cdot ((b(x) + K *_x f - \gamma v)f) + \frac{1}{2} \sigma^2 \Delta_v f, \quad (1.5)$$

where  $f(dx, dv) \in \mathbf{P}(\mathbb{R}^d \times \mathbb{R}^d)$  and  $*_x$  means that the convolution is performed only on the  $x$  variable.

If one directly discretizes (1.1) or (1.2), the computational cost per time step is  $\mathcal{O}(N^2)$ . This is undesired for large  $N$ . The Fast Multipole Method (FMM) [88] is able to reduce the complexity to  $\mathcal{O}(N)$  for fast enough decaying interactions. However, the implementation of FMM is quite involved. A simple random algorithm, called the Random Batch Method (RBM), has been proposed in [53] to reduce the computation cost per time step from  $\mathcal{O}(N^2)$  to  $\mathcal{O}(N)$ , based on the simple “random mini-batch” idea. Such an idea is famous for its application in the so-called stochastic gradient descent (SGD) [87, 11, 13] for machine learning problems. The idea was also used for Markov Chain Monte Carlo methods like the stochastic

gradient Langevin dynamics (SGLD) [96], and the computation of the mean-field flocking model [2, 16], motivated by Nanbu’s algorithm of the Direct Simulation Monte Carlo method [10, 82, 5].

The key behind the “mini-batch” idea is to find some cheap unbiased random estimator for the original quantity with the variance being controlled. Depending on the specific applications, the design can be different. For instances, the random batch strategy was proposed regarding general interacting particle systems in [53], while the importance sampling in the Fourier space was proposed for the Random Batch Ewald method for molecular dynamics in [56]. Compared with FMM, the accuracy of RBM is lower, but RBM is much simpler and is valid for more general potentials (e.g. the SVGD ODE [66]). The method converges due to the time average in time, and thus the convergence is like that in the Law of Large Numbers, but in time. For long time behaviors, the method works for systems that own ergodicity and mixing properties, like systems in contact with heat bath and converging to equilibria. A key difference from SGD or SGLD is that the RBM algorithms proposed are aiming to approximate and grasp the dynamical properties of the systems as well, not just to find the optimizer or equilibrium distribution.

RBM for interacting particle systems has been used or extended in various directions, from statistical sampling [66, 68, 57] to molecular dynamics [56, 67], control of synchronization [9, 64], and collective behavior of agent-based models [44, 43, 63]. RBM has been shown to converge for finite time interval if the interaction kernels are good enough [66, 53], and in particular an error estimate *uniformly* in  $N$  was first obtained in [53]. A convergence result of RBM for  $N$ -body Schrödinger equation was established in [41].

The goal of this review is to introduce the basics of the RBM, the fundamental theory for the convergence and error estimates, and various applications.

## 2 The RBM methods

In this section, we describe the RBM for general interacting particle systems introduced first in [53]. We use bold fonts (e.g.  $\mathbf{r}_i, \mathbf{x}_i, \mathbf{v}_i, \mathbf{u}_i$ ) and capital letters ( $X_i, Y_i$ ) to denote the quantities that are functions of time  $t$  associated with the particles, use usual letters like  $x_i, v_i$  to represent a point in the state space (often  $\mathbb{R}^d$ ), and use letters like  $\underline{x}, \underline{v}$  to represent quantities in the configurational space  $\mathbb{R}^{Nd}$ .

### 2.1 The RBM algorithms

Let  $T > 0$  be the simulation time, and choose a time step  $\Delta t > 0$ . Pick a batch size  $2 \leq p \ll N$  that divides  $N$  (RBM can also be applied if  $p$  does not divide  $N$ ; we assume this only for convenience). Consider the discrete time grids  $t_k := k\Delta t, k \in \mathbb{N}$ . For each subinterval  $[t_{k-1}, t_k)$ , the method has two substeps: (1) at  $t_{k-1}$ , divide the  $N$  particles into  $n := N/p$  groups (batches) randomly; (2) let the particles evolve with interaction only inside the batches.

The above procedure, when applied to the second order system (1.2), leads to Algorithm 1. The versions for first order systems is similar.

---

**Algorithm 1** (RBM for (1.2))

---

- 1: **for**  $m$  in  $1 : [T/\Delta t]$  **do**
- 2:     Divide  $\{1, 2, \dots, N = pn\}$  into  $n$  batches randomly.
- 3:     **for** each batch  $\mathcal{C}_q$  **do**
- 4:         Update  $\mathbf{r}_i, \mathbf{v}_i$  ( $i \in \mathcal{C}_q$ ) by solving for  $t \in [t_{m-1}, t_m]$  the following

$$\begin{aligned} d\mathbf{r}_i &= \mathbf{v}_i dt, \\ d\mathbf{v}_i &= \left[ b(\mathbf{r}_i) + \frac{\alpha_N(N-1)}{p-1} \sum_{j \in \mathcal{C}_q, j \neq i} K(\mathbf{r}_i - \mathbf{r}_j) - \gamma \mathbf{v}_i \right] dt + \sigma d\mathbf{W}_i. \end{aligned} \quad (2.1)$$

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

RBM requires the random division, and the elements in different batches are different. This is in fact the sampling without replacement. If one allows replacement, one has the following version of RBM.

---

**Algorithm 2** (RBM-r)

---

- 1: **for**  $m$  in  $1 : [T/\Delta t]$  **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  $\mathbf{r}_i$ 's ( $i \in \mathcal{C}_k$ ) by solving the following SDE for time  $\Delta t$ .

$$\begin{cases} d\mathbf{x}_i = \mathbf{u}_i dt, \\ d\mathbf{u}_i = \left[ b(\mathbf{x}_i) + \frac{\alpha_N(N-1)}{p-1} \sum_{j \in \mathcal{C}_k, j \neq i} K(\mathbf{x}_i - \mathbf{x}_j) - \gamma \mathbf{u}_i \right] dt + \sigma d\mathbf{W}_i, \\ \mathbf{x}_i(0) = \mathbf{r}_i, \quad \mathbf{u}_i(0) = \mathbf{v}_i, \end{cases} \quad (2.2)$$

i.e., solve (2.2) with initial values  $\mathbf{x}_i(0) = \mathbf{r}_i, \mathbf{u}_i(0) = \mathbf{v}_i$ , and set  $\mathbf{r}_i \leftarrow \mathbf{x}_i(\Delta t), \mathbf{v}_i \leftarrow \mathbf{u}_i(\Delta t)$ .

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

We now discuss the computational cost. Note that random division into  $n$  batches of equal size can be implemented using random permutation, which can be realized in  $\mathcal{O}(N)$  operations by Durstenfeld's modern revision of Fisher-Yates shuffle algorithm [28] (in MATLAB, one can use "randperm(N)"). After the permutation, one takes the first  $p$  elements to be in the first batch, the second  $p$  elements to be in the second batch, etc. The ODE solver per particle per time step (2.2) requires merely  $\mathcal{O}(p)$  operations, thus for all particles, each time step costs only  $\mathcal{O}(pN)$ . Since  $p \ll N$  the overall cost per time step is significantly reduced from  $\mathcal{O}(N^2)$ .

However, one might encounter the issue of having to use a much smaller time step—which could be of  $\mathcal{O}(N)$  times smaller—in the RBM implementation. For RBM to really gain significant efficiency, one needs  $\Delta t$  to be *independent* of  $N$ . This is justified by an error analysis to be presented in the next subsection.

## 2.2 Convergence analysis

In this subsection, we present the convergence results of RBM for the second order systems (1.2) in the mean field regime (i.e.,  $\alpha_N = 1/(N-1)$ ), which was given in [55]. We remark that the proof relies on the underlying contraction property of the second order systems under certain conditions ([76, 30]). Due to the degeneracy of the noise terms, the contraction should be proved by suitably chosen variables and Lyapunov functions, and we refer the readers to [55] for more details.

Denote  $(\tilde{\mathbf{r}}_i, \tilde{\mathbf{v}}_i)$  the solutions to the RBM process (2.1), and use the synchronization coupling as in [53, 54]:

$$\mathbf{r}_i(0) = \tilde{\mathbf{r}}_i(0) \sim \mu_0, \mathbf{W}_i = \tilde{\mathbf{W}}_i. \quad (2.3)$$

Let  $\mathcal{C}_q^{(k)}$  ( $1 \leq q \leq n$ ) be the batches at  $t_k$ , and define

$$\mathcal{C}^{(k)} := \{\mathcal{C}_1^{(k)}, \dots, \mathcal{C}_n^{(k)}\}, \quad (2.4)$$

to be the random division of batches at  $t_k$ . According to the Kolmogorov extension theorem [27], there exists a probability space  $(\Omega, \mathcal{F}, \mathbb{P})$  such that the random variables  $\{\mathbf{r}_0^i, \mathbf{W}^i, \mathcal{C}^{(k)} : 1 \leq i \leq N, k \geq 0\}$  are all defined on this probability space and are independent. Let  $\mathbb{E}$  denote the integration on  $\Omega$  with respect to the probability measure  $\mathbb{P}$ , and consider the  $L^2(\cdot)$  norm of a random variable

$$\|\zeta\| = \sqrt{\mathbb{E} \frac{|\zeta|^2}{|\zeta|^2}}. \quad (2.5)$$

For finite time interval, the convergence of RBM is as following.

**Theorem 2.1.** *Let  $b(\cdot)$  be Lipschitz continuous, and assume that  $|b|, |\nabla b|$  have polynomial growth, and the interaction kernel  $K$  is Lipschitz continuous. Then,*

$$\sup_{t \in [0, T]} \mathbb{E} \frac{|\tilde{\mathbf{r}}_i(t) - \mathbf{r}_i(t)|^2 + \mathbb{E} |\tilde{\mathbf{v}}_i(t) - \mathbf{v}_i(t)|^2}{|\zeta|^2} \leq C(T) \frac{\Delta t}{p-1} + \Delta t^2, \quad (2.6)$$

where  $C(T)$  is independent of  $N$ .

Often the long-time error estimates are important since one could use RBM as a sampling method for the invariant measure of (1.2) (see section 5). For this we need some additional contraction assumptions:

*Assumption 2.1.* Suppose  $b = -\nabla V$  for some  $V$  that is bounded from below (i.e.,  $\inf_x V(x) > -\infty$ ), and there exist  $\lambda_M \geq \lambda_m > 0$  such that the eigenvalues of  $H := \nabla^2 V$  satisfy

$$\lambda_m \leq \lambda_i(x) \leq \lambda_M, \quad \forall 1 \leq i \leq d, x \in \mathbb{R}^d.$$

The interaction kernel  $K$  is bounded and Lipschitz continuous. Moreover, the friction  $\gamma$  and the Lipschitz constant  $L$  of  $K(\cdot)$  satisfy

$$\gamma > \sqrt{\lambda_M + 2L}, \quad \lambda_m > 2L. \quad (2.7)$$

Then the following uniform strong convergence estimate holds:

**Theorem 2.2.** *Under Assumption 2.1 and the coupling (2.3), the solutions to (1.2) and (2.1) satisfy*

$$\sup_{t \geq 0} \mathbb{E} \frac{|\tilde{\mathbf{r}}_i(t) - \mathbf{r}_i(t)|^2 + \mathbb{E} |\tilde{\mathbf{v}}_i(t) - \mathbf{v}_i(t)|^2}{|\zeta|^2} \leq C \frac{\Delta t}{p-1} + \Delta t^2, \quad (2.8)$$

where the constant  $C$  does not depend on  $p$  and  $N$ .

Clearly, these error estimates imply that the RBM algorithms can also grasp the dynamical properties. The error estimates above are consequence of some intuitive results, which we summarize here (see [53]).

For given  $\underline{x} := (x_1, \dots, x_N) \in \mathbb{R}^{Nd}$ , introduce the error of the interacting force for the  $i$ th particle.

$$\chi_i(\underline{x}) := \frac{1}{p-1} \sum_{j \in \mathcal{C}} K(x_i - x_j) - \frac{1}{N-1} \sum_{j: j \neq i} K(x_i - x_j). \quad (2.9)$$

Here,  $\mathcal{C}$  is the random batch that contains  $i$  in a random division of the batches.

**Lemma 2.1.** Consider a configuration  $\underline{x}$  that is independent of the random division. Then,

$$\mathbb{E}\chi_i(\underline{x}) = 0. \quad (2.10)$$

Moreover, the (scalar) variance is given by

$$\text{Var}(\chi_i(\underline{x})) = \frac{1}{p-1} - \frac{1}{N-1} \Lambda_i(\underline{x}), \quad (2.11)$$

where

$$\Lambda_i(\underline{x}) := \frac{1}{N-2} \sum_{j:j \neq i} \left| K(x_i - x_j) - \frac{1}{N-1} \sum_{\ell:\ell \neq i} K(x_i - x_\ell) \right|^2. \quad (2.12)$$

Lemma 2.1 in fact lays the foundation of the convergence of RBM-type algorithms. The first claim implies that the random estimates of the interacting forces are unbalanced in the sense that the expectation is zero. This gives the consistency—in expected value—of the RBM approximation, although each random batch approximation  $\frac{1}{p-1} \sum_{j \in \mathcal{C}} K(x_i - x_j)$  to the true interacting force  $\frac{1}{N-1} \sum_{j:j \neq i} K(x_i - x_j)$  gives an  $\mathcal{O}(1)$  error (which is clear from  $\sqrt{\text{Var}(\chi_i(\underline{x}))} = \mathcal{O}(1)$ ). Being a Monte-Carlo like methods, the boundedness of the variance ensures the stability of the RBM methods as can be seen in the proof [55, 54]. The intrinsic mechanism why such type of methods work is the independent resampling in later time steps, and due to some averaging effect in time these  $\mathcal{O}(1)$  errors become small. This Law of Large Numbers type feature *in time* guarantees the convergence of RBMs (as indicated by the error bound  $\sqrt{|\text{Var}(\chi)|\tau} \sim \sqrt{|\text{Var}(\chi)|/N_T}$  in Theorems 2.1 and 2.2).

As another remark, the nonzero variance of the RBM approximation gives some effective noise into the system. This could bring in some “numerical heating” effects when RBM is applied for some interacting particle systems. When the system has some dissipation, or in contact with a heat bath as in section section 4, RBM approximation can be valid for long time and can capture the equilibrium.

In both Theorem 2.1 and Theorem 2.2, the error bound is independent of  $N$  so that the time step can be chosen independent of  $N$  for a fixed accuracy tolerance  $\varepsilon$ . Hence, for each time step, the cost of RBM is  $\mathcal{O}(1/N)$  of that for direct simulation, but it does not need to take  $\mathcal{O}(N)$  times longer to finish the computation. Such convergence results were first established for first order systems (1.1) [53] and then extended to disparate mass cases [54]. The weak convergence has also been discussed in [54].

### 2.3 An illustrating example: wealth evolution

To illustrate the algorithms, we consider the model proposed by Degond et. al. [22] for the evolution of  $N$  market agents with two attributes: the economic configuration  $X_i$  and its wealth  $Y_i$ .

$$\begin{aligned} \dot{X}_i &= V(X_i, Y_i), \\ dY_i &= -\frac{1}{N-1} \sum_{k:k \neq i} \xi_{ik} \Psi(|X_i - X_k|) \partial_y \phi(Y_i - Y_k) dt + \sqrt{2DY_i} dW_i. \end{aligned} \quad (2.13)$$

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 [65]. This model clearly is a suitable interacting particle system for which the RBM algorithms fit perfectly. Moreover, the RBM version of (2.13) can be viewed as a new model as one agent may only trade with a small number of random agents during a short time in the real world.

For numerical experiments, [53] considers the homogeneous case when the wealth dynamics is independent of the position in the economic configuration space. Then, the dynamics of the wealth is reduced the following

$$dY_i = -\frac{\kappa}{N-1} \sum_{k:k \neq i} \partial_y \phi(Y_i - Y_k) dt + \sqrt{2DY_i} dW_i. \quad (2.14)$$

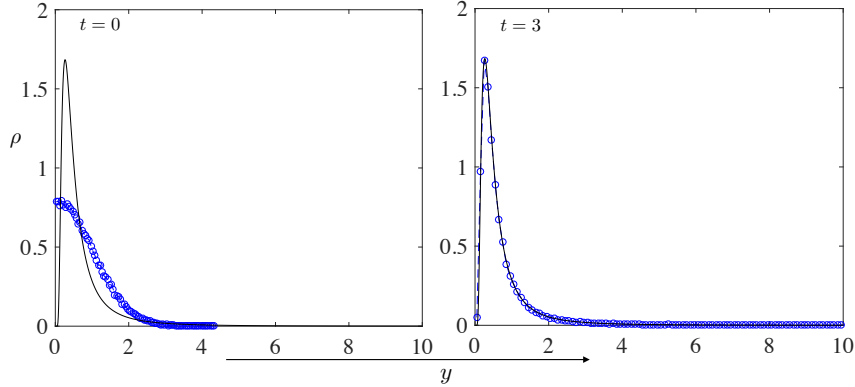


Figure 1: Wealth distribution obtained by RBM compared with the reference curve

The corresponding mean field dynamics has an equilibrium distribution given by

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

where  $\alpha$  satisfies

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

In Fig. 1, the empirical distribution of the wealth obtained by RBM for the case  $\phi(y) = \frac{1}{2}y^2$  is compared to the reference curve (an inverse Gamma distribution), which is

$$\rho_\infty(y) = \frac{(\kappa\eta/D)^{\kappa/D+1}}{\Gamma(\kappa/D+1)} y^{-(2+\kappa/D)} \exp \left[ -\frac{\kappa\eta}{Dy} \right] \mathbf{1}_{y>0}, \quad \eta = \frac{\sqrt{2}}{\sqrt{\pi}}.$$

Clearly, the distribution obtained by RBM agrees perfectly with the expected wealth distribution at  $t = 3$  already.

This example has two distinguished features: long range and multiplicative noises. We point out that although the convergence results presented in subsection 2.2 were for regular interacting potentials  $K$  and additive noises, as shown by this and more examples in [53], and those in later sections, the RBM algorithms are applicable to much broader classes of interacting particle systems, including long-range, singular (like the Lenard-Jones and Coulomb) potentials (see section 4 below), and with multiplicative noise.

### 3 The mean-field limit

It is known that the  $N$ -particle system (1.1) has the mean-field limit given by the Fokker-Planck equation (1.4). Namely, the empirical measure of the particle system (1.1) is close, in Wasserstein distance, to  $\mu$  in (1.4). Thus, when  $N$  is large, one may use the RBM as a numerical (particle method) for (1.4). Indeed, since the error bounds obtained in the precious section is independent of  $N$ , one could hope that when  $N \rightarrow \infty$ , the empirical measure of the RBM should be close to  $\mu$ . To justify this, one first needs to derive the mean field limit of the RBM, for fixed  $\Delta t$ , then compare it with (1.4). In addition, the RBM could be viewed as a random model for the underlying physics, hence it is also natural to ask what its mean field limit is.

The mean-field limit of the RBM for the general first order system (1.1), given below by Algorithm 3, was derived and proved in [52]. We summarize the results in this section.

---

**Algorithm 3** (RBM for first order systems)

---

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

$$d\mathbf{r}_i = b(\mathbf{r}_i)dt + \frac{1}{p-1} \sum_{j \in \mathcal{C}_q, j \neq i} K(\mathbf{r}_i - \mathbf{r}_j)dt + \sigma d\mathbf{W}^i. \quad (3.1)$$

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

Intuitively, when  $N \gg 1$ , the probability that two chosen particles are correlated is very small. Hence, in the  $N \rightarrow \infty$  limit, two chosen particles will be independent with probability 1. Due to the exchangeability, the marginal distributions of the particles will be identical. Based on this observation, the following mean field limit for RBM can be obtained for the one-particle distribution:

---

**Algorithm 4** (Mean Field Dynamics of RBM (3.1))

---

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

$$\partial_t \rho^{(p)} = - \sum_{i=1}^p \nabla_{x_i} \cdot \left[ b(x_i) + \frac{1}{p-1} \sum_{j=1, j \neq i}^p K(x_i - x_j) \right] \rho^{(p)} + \frac{1}{2} \sigma^2 \sum_{i=1}^p \Delta_{x_i} \rho^{(p)}. \quad (3.2)$$

- 5:     Set

$$\tilde{\mu}(\cdot, t_{k+1}) := \int_{(\mathbb{R}^d)^{\otimes (p-1)}} \rho^{(p)}(\cdot, dy_2, \dots, dy_p, \Delta t). \quad (3.3)$$

- 6: **end for**
- 

The dynamics in Algorithm 4 naturally gives a nonlinear operator  $\mathcal{G}_\infty : \mathbf{P}(\mathbb{R}^d) \rightarrow \mathbf{P}(\mathbb{R}^d)$  as

$$\tilde{\mu}(\cdot, t_{k+1}) =: \mathcal{G}_\infty(\tilde{\mu}(\cdot, t_k)). \quad (3.4)$$

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

$$d\mathbf{x}_i = b(\mathbf{x}_i) dt + \frac{1}{p-1} \sum_{j=1, j \neq i}^p K(\mathbf{x}_i - \mathbf{x}_j) dt + \sigma d\mathbf{W}_i, \quad i = 1, \dots, p, \quad (3.5)$$

with  $\{\mathbf{x}_i(t_k)\}$  drawn i.i.d from  $\tilde{\mu}(\cdot, t_k)$ . Then,  $\tilde{\mu}(\cdot, t_{k+1}) = \mathcal{L}(\mathbf{x}_1(t_{k+1}^-))$ , the law of  $\mathbf{x}_1(t_{k+1}^-)$ . Note that all  $\mathbf{x}_i$  have the same distribution for any  $t_k \leq t < t_{k+1}$ . Without loss of generality, we will impose  $\mathbf{x}_1(t_k^-) = \mathbf{x}_1(t_k^+)$ . For other particles  $i \neq 1$ ,  $\mathbf{x}_i(t)$  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  $N \rightarrow \infty$  limit,  $\mathbf{x}_i, i \neq 1$  at different subintervals correspond to different particles that interact with particle 1 as in Algorithm 3.

Hence, in the mean field limit of RBM, one starts with a chaotic configuration<sup>1</sup>, the  $p$

---

<sup>1</sup>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".



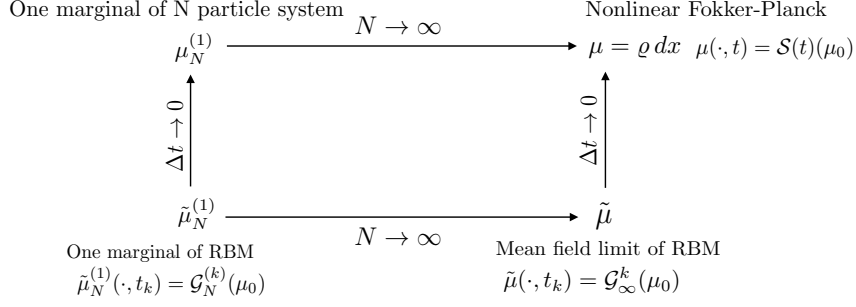


Figure 2: Illustration of the various operators and the asymptotic limits.

particles evolve by interacting with each other. Then, at the starting point of the next time interval, one imposes the chaos condition so that the particles are independent again.

In [52], this intuition has been justified rigorously for finitely many steps under the following assumptions.

*Assumption 3.1.* The moments of the initial data are finite:

$$\int_{\mathbb{R}^d} |x|^q \mu_0(dx) < \infty, \quad \forall q \in [2, \infty). \quad (3.6)$$

*Assumption 3.2.* Assume  $b(\cdot) : \mathbb{R}^d \rightarrow \mathbb{R}^d$  and  $K(\cdot) : \mathbb{R}^d \rightarrow \mathbb{R}^d$  satisfy the following conditions.

- $b(\cdot)$  is one-sided Lipschitz:

$$(z_1 - z_2) \cdot (b(z_1) - b(z_2)) \leq \beta |z_1 - z_2|^2 \quad (3.7)$$

for some constant  $\beta$ ;

- $K$  is Lipschitz continuous

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

Corresponding to the operator (3.4), one may define the operator  $\mathcal{G}_N^k : \mathbf{P}(\mathbb{R}^d) \rightarrow \mathbf{P}(\mathbb{R}^d)$  for RBM with  $N$  particles as follows. Let  $\mathbf{r}_i(0)$ 's be i.i.d drawn from  $\mu_0^{\otimes N}$ , and consider (3.1). Define

$$\mathcal{G}_N^k(\mu_0) := L(\mathbf{r}_1(t_k)). \quad (3.8)$$

Recall that  $L(\mathbf{r}_1)$  denotes the law of  $\mathbf{r}_1$ , 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  $\mathbf{r}_i(t_k)$  ( $1 \leq i \leq N$ ) are identical. In Fig.2, we illustrate these definitions and various limits. With these setup introduced, we may state the first main result in [52] as follows:

**Theorem 3.1.** *Under assumptions 3.1 and 3.2, for any fixed  $k$ , it holds that*

$$\lim_{N \rightarrow \infty} W_q(\mathcal{G}_\infty^k(\mu_0), \mathcal{G}_N^k(\mu_0)) = 0. \quad (3.9)$$

Here,  $W_q$  is the Wasserstein- $q$  distance [89]:

$$W_q(\mu, \nu) = \left( \inf_{\gamma \in \Pi(\mu, \nu)} \int_{\mathbb{R}^d \times \mathbb{R}^d} |x - y|^q d\gamma \right)^{1/q}, \quad (3.10)$$

where  $\Pi(\mu, \nu)$  is the set of “transport plans”, i.e. a joint measure on  $\mathbb{R}^d \times \mathbb{R}^d$  such that the marginal measures corresponding to  $x$  and  $y$  are  $\mu$  and  $\nu$  respectively.

The next question is whether the one marginal distribution  $\mu_N^{(1)} := \mathcal{L}(\mathbf{r}_1)$  of the RBM converges to  $\mu$ . Denote the solution operator to (1.4) by  $\mathcal{S}$ :

$$\mathcal{S}(\Delta)\mu(t_1) := \mu(t_1 + \Delta), \quad \forall t_1 \geq 0, \Delta \geq 0. \quad (3.11)$$

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

We make more technical assumptions here.

*Assumption 3.3.* The measure  $\mu_0$  has a density  $\varrho_0$  that is smooth with finite moments  $\int_{\mathbb{R}^d} |x|^q \varrho_0 dx < \infty, \forall q \geq 1$ , and the entropy is finite

$$H(\mu_0) := \int_{\mathbb{R}^d} \varrho_0 \log \varrho_0 dx < \infty. \quad (3.12)$$

If  $\varrho_0(x) = 0$  at some point  $x$ , one defines  $\varrho_0(x) \log \varrho_0(x) = 0$ . We also introduce the following assumption on the growth rate of derivatives of  $b$  and  $K$ , which will be used below.

*Assumption 3.4.* 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, it can be shown that  $\mu$  has a density  $\varrho(\cdot, t)$ . For convenience, we will not distinguish  $\mu$  from its density  $\varrho$ . Sometimes, one may also assume the strong confinement condition:

*Assumption 3.5.* The fields  $b(\cdot) : \mathbb{R}^d \rightarrow \mathbb{R}^d$  and  $K(\cdot) : \mathbb{R}^d \rightarrow \mathbb{R}^d$  are smooth. Moreover,  $b(\cdot)$  is strongly confining:

$$(z_1 - z_2) \cdot (b(z_1) - b(z_2)) \leq -r|z_1 - z_2|^2 \quad (3.13)$$

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

$$r > 2L. \quad (3.14)$$

With the assumptions stated, we can state the second main result in [52].

**Theorem 3.2.** *Suppose Assumptions 3.2, 3.3 and 3.4 hold. Then,*

$$\sup_{n:n\Delta t \leq T} W_1(\mathcal{G}_\infty^n(\varrho_0), \varrho(n\Delta t)) \leq C(T)\Delta t. \quad (3.15)$$

*If Assumption 3.5 is assumed in place of Assumption 3.2 and also  $\sigma > 0$ , then*

$$\sup_{n \geq 0} W_1(\mathcal{G}_\infty^n(\varrho_0), \varrho(n\Delta t)) \leq C\Delta t. \quad (3.16)$$

These theorems show that the dynamics given by  $\mathcal{G}_\infty$  can approximate that of the nonlinear Fokker-Planck equation (1.4), with the  $W_1$  distance to be of  $\mathcal{O}(\Delta t)$ . Thus, the two limits  $\lim_{N \rightarrow \infty}$  and  $\lim_{\Delta t \rightarrow 0}$  commute, as shown in Fig. 2.

## 4 Molecular dynamics

Molecular dynamics (MD) refers to computer simulation of atoms and molecules, and is among the most popular numerical methods to understand the dynamical and equilibrium properties of many-body particle systems in many areas such as chemical physics, soft materials and biophysics [18, 34, 33]. In this section, we discuss the relevant issues and the applications of RBM and its modifications.

Consider  $N$  ‘‘molecules’’ with masses  $m_i$ ’s (each might be a model for a real molecule or a numerical molecule that is a packet of many real molecules) that interact with each other. The equations of motion are given by

$$\begin{aligned} d\mathbf{r}_i &= \mathbf{v}_i dt, \\ m_i d\mathbf{v}_i &= \left[ - \sum_{j:j \neq i} \nabla \phi(\mathbf{r}_i - \mathbf{r}_j) \right] dt + d\xi_i. \end{aligned} \quad (4.1)$$

Here,  $\phi(\cdot)$  is the interaction potential and  $d\xi_i$  means some other possible terms that change the momentum,. Typical examples of the potential include the Coulomb potentials

$$\phi(x) = \frac{q_i q_j}{r},$$

where  $q_i$  is the charge for the  $i$ th particle and  $r = |x|$ , and the Lennard-Jones potential

$$\phi(x) = 4 \left( \frac{1}{r^{12}} - \frac{1}{r^6} \right).$$

Between ions, both types of potential exist and between charge-neutral molecules, the Lennard-Jones potential might be the main force (the Lennard-Jones interaction intrinsically also arises from the interactions between charges, so these two types are in fact both electromagnetic forces) [34, 33]. To model the solids or fluids with large volume, one often uses a box with length  $L$ , equipped with the periodic conditions for the simulations.

To model the interaction between the molecules with the heat bath, one may consider some thermostats so that the temperature of the system can be controlled at a given value. The thermostats are especially good for RBM approximations as the effective noise introduced by RBM approximation can be damped by the thermostats, reducing the “numerical heating” effects [56]. Typical thermostats include the Andersen thermostat, the Langevin thermostat and the Nosé-Hoover thermostat [34]. In the Andersen thermostat [34, section 6.1.1], one does the simulation for  $d\xi_i = 0$  between two time steps, but a particle can collide with the heat bath at each discrete time. Specifically, assume the collision frequency is  $\nu$ , so in a duration of time  $t \ll 1$  the chance that a collision has happened is given by the exponential distribution

$$1 - \exp(-\nu t) \approx \nu t, \quad t \ll 1.$$

If a collision happens, the new velocity is then sampled from the Maxwellian distribution with temperature  $T$  (i.e., the normal distribution  $\mathcal{N}(0, T)$ ). In the underdamped Langevin dynamics, one chooses

$$d\xi_i = -\gamma \mathbf{v}_i dt + \frac{\sqrt{2\gamma}}{\beta} d\mathbf{W}_i,$$

so that the “fluctuation-dissipation relation” is satisfied and the system will evolve to the equilibrium with the correct temperature  $T = \beta^{-1}$ . It is well-known that the invariant measure of such systems is given by the Gibbs distribution [70]

$$\pi(\underline{x}, \underline{v}) \propto \exp\left(-\beta\left(\frac{1}{2} \sum_{i=1}^N |v_i|^2 + U(\underline{x})\right)\right), \quad U(\underline{x}) = \frac{1}{2} \sum_{i,j:i \neq j} \phi(x_i - x_j),$$

where  $\underline{x} = (x_1, \dots, x_N) \in \mathbb{R}^{Nd}$  and  $\underline{v} = (v_1, \dots, v_N) \in \mathbb{R}^{Nd}$ . The Nosé-Hoover thermostat uses a Hamiltonian for an extended system of  $N$  particles plus an additional coordinate  $s$  ([83, 50]):

$$\mathcal{H}_{\text{NH}} = \sum_{i=1}^N \frac{|\tilde{\mathbf{p}}_i|^2}{2m_i s^2} + U(\{\mathbf{r}_i\}) + \frac{p_s^2}{2Q} + L \frac{\ln s}{\beta}.$$

Here,  $\tilde{\mathbf{p}}_i$  is the momentum of the  $i$ th particle. The microcanonical ensemble corresponding to this Hamiltonian reduces to the canonical ensemble for the real variables  $\mathbf{p}_i = \tilde{\mathbf{p}}_i/s$ . Hence, one may run the following deterministic ODEs, which are the Hamiltonian ODEs with Hamiltonian  $\mathcal{H}_{\text{NH}}$  in terms of the so-called real variables,

$$\begin{aligned} \dot{\mathbf{r}}_i &= \mathbf{p}_i, \\ \dot{\mathbf{p}}_i &= -\nabla_{\mathbf{r}_i} U - \xi \mathbf{p}_i, \\ \dot{\xi} &= \frac{1}{Q} \left( \sum_{i=1}^N \frac{|\mathbf{p}_i|^2}{m_i} - \frac{3N}{\beta} \right). \end{aligned}$$

The time average of the desired quantities will be the correct canonical ensemble average. As one can see, when the temperature of the system is different from  $T$ , the extra term  $-\xi \mathbf{p}_i$  will drive the system back to temperature  $T$ , thus it may give better behaviors for controlling the temperature.

## 4.1 RBM with kernel splitting

In molecular dynamics simulation, the interaction force kernel

$$K(x) := -\nabla\phi(x), \quad x \in \mathbb{R}^d,$$

is often singular at  $x = 0$ . Hence, the direct application of RBM could lead to poor results. To resolve this issue, one can adopt the splitting strategy in [75, 49], and decompose the interacting force  $K$  into two parts:

$$K(x) = K_1(x) + K_2(x). \quad (4.2)$$

Here,  $K_1$  has short range that vanishes for  $|x| \geq r_0$  where  $r_0$  is a certain cutoff chosen to be comparable to the mean distance of the particles.  $K_2(x)$  is a bounded smooth function. One then applies RBM to the  $K_2$  part only. The resulted method is shown in Algorithm 5. Now, the cost of summation in  $K_1$  is of  $\mathcal{O}(1)$  for each given  $i$  using data structures like Cell-List [34, Appendix F]. Since  $K_2$  is bounded, RBM can be applied well due to the boundedness of variance, without introducing too much error. Hence, the cost per time step is again  $\mathcal{O}(N)$ . For practical applications, one places the initial positions of the molecules on the grid of a lattice, and the repulsive force  $K_1$  will forbid the particles from getting too close so that the system is not too stiff.

---

### Algorithm 5 RBM with splitting for (1.2)

---

- 1: Split  $K =: K_1 + K_2$ , where  $K_1$  has short range, while  $K_2$  has long range but is smooth.
- 2: **for**  $m$  in  $1 : [T/\Delta t]$  **do**
- 3:     Divide  $\{1, 2, \dots, N = pn\}$  into  $n$  batches randomly.
- 4:     **for** each batch  $\mathcal{C}_q$  **do**
- 5:         Update  $(\mathbf{r}_i, \mathbf{v}_i)$ 's ( $i \in \mathcal{C}_q$ ) by solving for  $t \in [t_{m-1}, t_m)$

$$\begin{aligned} d\mathbf{r}_i &= \mathbf{v}_i dt, \\ d\mathbf{v}_i &= \left[ b(\mathbf{r}_i) + \alpha_N \sum_{j:j \neq i} K_1(\mathbf{r}_i - \mathbf{r}_j) - \gamma \mathbf{v}_i \right] dt \\ &\quad + \frac{\alpha_N(N-1)}{p-1} \sum_{j \in \mathcal{C}_q, j \neq i} K_2(\mathbf{r}_i - \mathbf{r}_j) dt + \sigma d\mathbf{W}_i. \end{aligned} \quad (4.3)$$

- 6:     **end for**
  - 7: **end for**
- 

Using this splitting strategy, one may apply RBM to the MD simulations with different thermostats. In Fig. 3, we show the numerical results from [55] for a Lennard-Jones fluid with temperature  $\beta^{-1} = 2$  and the length of box  $L = (N/\rho)^{1/3}$  for a given density  $\rho$ . The results are obtained using the Andersen thermostat and the Langevin thermostat respectively, with the splitting and RBM strategy, for particle number  $N = 500$ . The reference curves (black solid line) are the fitting curves in [59]. In the first figure, the decreasing step sizes  $\Delta t_k = 0.001/\log(k+1)$  are taken to reduce the numerical heating effect brought by RBM when the collision coefficient are not so big ( $\nu = \gamma = 10$ ). The results show that RBM with splitting strategy can work reasonably well for the Lennard-Jones fluid in the considered regime.

## 4.2 Random Batch Ewald: an importance sampling in the Fourier space

In the presence of long-range interactions such as the Coulomb interactions, the molecular dynamics simulations becomes computationally expensive for large  $N$ . A lot of effort has already been devoted to computing such long-range interactions efficiently. Some popular methods include lattice summation methods such as the particle-particle particle mesh Ewald

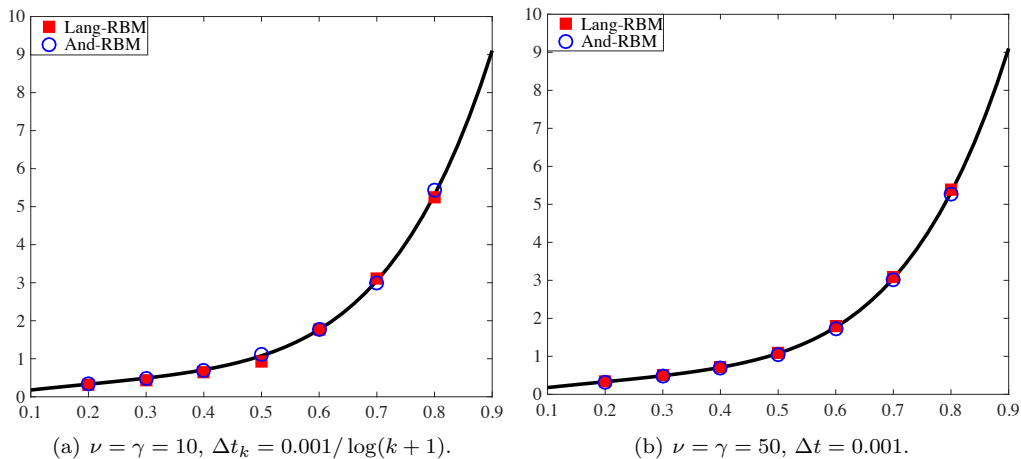


Figure 3: The pressure obtained by Andersen-RBM and Langevin-RBM for Lenard-Jones fluid with  $N = 500$ : the blue circles are those by Andersen-RBM while the red squares are by Langevin-RBM.

(PPPM) [74, 23], and multipole type methods such as treecode [7, 25] and fast multipole methods (FMM) [42, 99]. These methods can reduce the complexity per time step from  $O(N^2)$  to  $O(N \log N)$  or even  $O(N)$ , and have gained big success in practice. However, some issues still remain to be resolved, e.g., the prefactor in the linear scaling can be large, or the implementation can be nontrivial, or the scalability for parallel computing is not high.

In this section, we give a brief introduction to the recently proposed Random Batch Ewald (RBE) method for molecular dynamics simulations of particle systems with long-range Coulomb interactions, which achieves an  $O(N)$  complexity [56] with a high parallel efficiency. The RBE method is based on the Ewald splitting for the Coulomb kernel with a random “mini-batch” type technique applied in the Fourier series for the long-range part. Compared with PPPM where the Fast Fourier Transform is used to speed up the computation in the Fourier space, the RBE method uses random batch type technique to speed up the computation.

Consider  $N$  physical or numerical particles inside the periodic box with side length  $L$ , assumed to have net charge  $q_i$  ( $1 \leq i \leq N$ ) with the electroneutrality condition

$$\sum_{i=1}^N q_i = 0. \quad (4.4)$$

The forces are computed using  $\mathbf{F}_i = -\nabla_{\mathbf{r}_i} U$ , where  $U$  is the potential energy of the system. Since the Coulomb potential is of long range, with the periodic boundary condition, one must consider the images so that

$$U = \frac{1}{2} \sum_{\mathbf{n}}' \sum_{i,j=1}^N q_i q_j \frac{1}{|\mathbf{r}_{ij} + \mathbf{n}L|}, \quad (4.5)$$

where  $\mathbf{n} \in Z^3$  ranges over the three-dimensional integer vectors and  $\sum'$  is defined such that  $\mathbf{n} = 0$  is not included when  $i = j$ .

Due to the long-range nature of the Coulomb potential, the series (4.5) converges conditionally, thus a naive truncation would require a very large  $r$  to maintain the desired numerical accuracy. The classical Ewald summation separates the series into long-range smooth parts and short-range singular parts:

$$\frac{1}{r} = \frac{\text{erf}(\sqrt{\alpha}r)}{r} + \frac{\text{erfc}(\sqrt{\alpha}r)}{r}, \quad (4.6)$$

where  $\text{erf}(x) := \frac{2}{\sqrt{\pi}} \int_0^x \exp(-u^2) du$  is the error function and  $\text{erfc} = 1 - \text{erf}$ . Correspondingly,

$$U_1 = \frac{1}{2} \sum_{\mathbf{n}}' \sum_{i,j} q_i q_j \frac{\text{erf}(\sqrt{\alpha} |\mathbf{r}_{ij} + \mathbf{n}L|)}{|\mathbf{r}_{ij} + \mathbf{n}L|}, \quad (4.7)$$

$$U_2 = \frac{1}{2} \sum_{\mathbf{n}}' \sum_{i,j} q_i q_j \frac{\text{erfc}(\sqrt{\alpha} |\mathbf{r}_{ij} + \mathbf{n}L|)}{|\mathbf{r}_{ij} + \mathbf{n}L|}. \quad (4.8)$$

The computation of force can be done directly using

$$\mathbf{F}_i = -\nabla_{\mathbf{r}_i} U = -\nabla_{\mathbf{r}_i} U_1 - \nabla_{\mathbf{r}_i} U_2 =: \mathbf{F}_{i,1} + \mathbf{F}_{i,2}.$$

The second part  $\mathbf{F}_{i,2}$  corresponds to the short-range forces whose computational cost is relatively low, since, for each particle, one just needs to add a finite number of particles in its close neighbour. We now focus on the first part.

The slow decay of  $U_1$  in  $r$ , corresponding to the long-range, can be dealt with in the Fourier space thanks to its smoothness (see [34, Chap. 12]):

$$U_1 = \frac{2\pi}{V} \sum_{\mathbf{k} \neq 0} \frac{1}{k^2} |\rho(\mathbf{k})|^2 e^{-k^2/4\alpha} - \frac{\hat{E}}{\pi} \sum_{i=1}^N q_i^2, \quad (4.9)$$

where  $k = |\mathbf{k}|$  and  $\rho(\mathbf{k})$  is given by  $\rho(\mathbf{k}) := \sum_{i=1}^N q_i e^{i\mathbf{k} \cdot \mathbf{r}_i}$ . The divergent  $\mathbf{k} = 0$  term is usually neglected in simulations to represent that the periodic system is embedded in a conducting medium which is essential for simulating ionic systems. Then

$$\mathbf{F}_{i,1} = - \sum_{\mathbf{k} \neq 0} \frac{4\pi q_i \mathbf{k}}{V k^2} e^{-k^2/(4\alpha)} \text{Im}(e^{-i\mathbf{k} \cdot \mathbf{r}_i} \rho(\mathbf{k})), \quad (4.10)$$

where we recall  $\mathbf{r}_{ij} = \mathbf{r}_j - \mathbf{r}_i$ , pointing towards particle  $j$  from particle  $i$ . Note that the force  $\mathbf{F}_{i,1}$  is bounded for small  $\mathbf{k}$ . In fact,  $k \geq 2\pi/L$ , so  $Vk \geq 2\pi L^2$ . Let us consider the factor  $e^{-k^2/(4\alpha)}$ , and denote the sum of such factors by

$$S := \sum_{\mathbf{k} \neq 0} e^{-k^2/(4\alpha)} = H^3 - 1, \quad (4.11)$$

where

$$H := \sum_{m \in \mathbb{Z}} e^{-\pi^2 m^2 / (\alpha L^2)} = \frac{\hat{E}}{\pi} \sum_{m=-\infty}^{\infty} e^{-\alpha m^2 L^2} \approx \frac{\hat{E}}{\pi} \frac{\alpha L^2}{\pi} (1 + 2e^{-\alpha L^2}), \quad (4.12)$$

since often  $\alpha L^2 \gg 1$ . Hence,  $S$  is the sum for all three-dimensional vectors  $\mathbf{k}$  except 0. Then, one can regard the sum as an expectation over the probability distribution

$$P_{\mathbf{k}} := S^{-1} e^{-k^2/(4\alpha)}, \quad (4.13)$$

which, with  $\mathbf{k} \neq 0$ , is a discrete Gaussian distribution and can be sampled efficiently. For example, one can use the Metropolis-Hastings (MH) algorithm (see [48] for details) by choosing proposal samples from the continuous Gaussian  $\mathcal{N}(0, \alpha L^2 / (2\pi^2))$ , the normal distribution with mean zero and variance  $\alpha L^2 / (2\pi^2)$ . It should be emphasized that this sampling can be done *o ine*, before the iteration begins. Once the time evolution starts one just needs to randomly draw a few ( $p$ ) samples for each time step from this pre-sampled Gaussian sequence.

With this observation, the MD simulations can then be done via the random mini-batch approach with this importance sampling strategy. Specifically, one approximates the force  $\mathbf{F}_{i,1}$  in (4.10) by the following random variable:

$$\mathbf{F}_{i,1} \approx \mathbf{F}_{i,1}^* := - \sum_{\ell=1}^p \frac{S}{p} \frac{4\pi \mathbf{k}_\ell q_i}{V k_\ell^2} \text{Im}(e^{-i\mathbf{k}_\ell \cdot \mathbf{r}_i} \rho(\mathbf{k}_\ell)). \quad (4.14)$$

The corresponding algorithm is shown in Algorithm 6.

---

**Algorithm 6** (Random-batch Ewald)

---

- 1: Choose  $\alpha$ ,  $r_c$  and  $k_c$  (the cutoffs in real and Fourier spaces respectively),  $\Delta t$ , and batch size  $p$ . Initialize the positions and velocities of charges  $\mathbf{r}_i^0, \mathbf{v}_i^0$  for  $1 \leq i \leq N$ .
  - 2: Sample sufficient number of  $\mathbf{k} \sim e^{-k^2/(4\alpha)}$ ,  $\mathbf{k} \neq 0$  by the MH procedure to form a set  $\mathcal{K}$ .
  - 3: **for**  $n$  in  $1 : N$  **do**
  - 4:     Integrate Newton's equations (4.1) for time  $\Delta t$  with appropriate integration scheme and some appropriate thermostat. The Fourier parts of the Coulomb forces are computed using RBE force (4.14) with the  $p$  frequencies chosen from  $\mathcal{K}$  in order.
  - 5: **end for**
- 

Similar to the strategy in the PPPM, one may choose  $\alpha$  such that the time cost in real space is cheap and then speed up the computation in the Fourier space. Compared with PPPM, the only difference is that PPPM uses FFT while RBE uses random mini-batch to speed up the computation in the Fourier space. Hence, we make the same choice

$$\sqrt{\alpha} \sim \frac{N^{1/3}}{L} = \rho_r^{1/3},$$

which is inverse of the average distance between two numerical particles. The complexity for the real space part is  $\mathcal{O}(N)$ . By choosing *the same batch* of frequencies for all forces (4.14) (i.e., using the same  $\mathbf{k}_\ell$ ,  $1 \leq \ell \leq p$  for all  $\mathbf{F}_{i,1}^*$ ,  $1 \leq i \leq N$ ) in the same time step, the complexity per iteration for the frequency part is reduced to  $\mathcal{O}(pN)$ . This implies that the RBE method has linear complexity per time step if one chooses  $p = \mathcal{O}(1)$ .

To illustrate the performance of the RBE method, consider an electrolyte with monovalent binary ions (first example in [56]). In the reduced units ([34, section 3.2]), the dielectric constant is taken as  $\varepsilon = 1/4\pi$  so that the potential of a charge is  $\phi(r) = q/r$  and the temperature is  $T = \beta^{-1} = 1$ . Under the Debye-Hückel (DH) theory (linearized Poisson-Boltzmann equation), the charge potential outside one ion is given by

$$-\varepsilon\Delta\phi = \begin{cases} 0 & r < a \\ q\rho_{\infty,+}e^{-\beta q\phi} - q\rho_{\infty,-}e^{\beta q\phi} \approx \beta q^2\rho_r\phi, & r > a \end{cases}$$

where  $\rho_{\infty,+} = \rho_{\infty,-} = N/(2V)$  are the densities of the positive and negative ions at infinity, both being  $\rho_r/2$ . The parameter  $a$  is the effective diameter of the ions, which is related to the setting of the Lennard-Jones potential. In the simulations,  $a = 0.2$  and the setting of Lennard-Jones potential can be found in [56]. This approximation gives the net charge density  $\rho = -\varepsilon\Delta\phi$  for  $r \gg a$ ,

$$\ln(r\rho(r)) \approx -1.941r - 1.144.$$

The results in the left panel of Fig.4 were obtained by  $N = 300$  (i.e., 150 cation and anion particles respectively) in a periodic box with side length  $L = 10$ . The thermostat was Andersen's thermostat with collision frequency  $\nu = 3$ . These parameters are chosen such that they give comparable results. Clearly, all the three methods give correct results, agreeing with the curve predicted by the DH theory. Regarding the efficiency, the right figure shows the time consumed for different particle numbers inside the box with the same side length  $L = 10$ . Both the PPPM and RBE methods scale lineally with the particle numbers. However, even for batch size  $p = 100$ , the RBE method consumes much less time. The relative accuracies of the potential obtained by RBE against the PPPM are listed in Table 1, for different densities  $\rho_r = N/L^3$ . Clearly, the RBE method has the same level of accuracy compared with the PPPM method for the densities considered.

Next, in Fig. 5, the parallel efficiency of the PPPM and RBE methods from [69] for the all-atom simulation of pure water systems is shown. As can be seen, due to the reduction of communications for the particles, the RBE method gains better parallel efficiency. This parallel efficiency is more obvious when the number of particles is larger. In [69], the

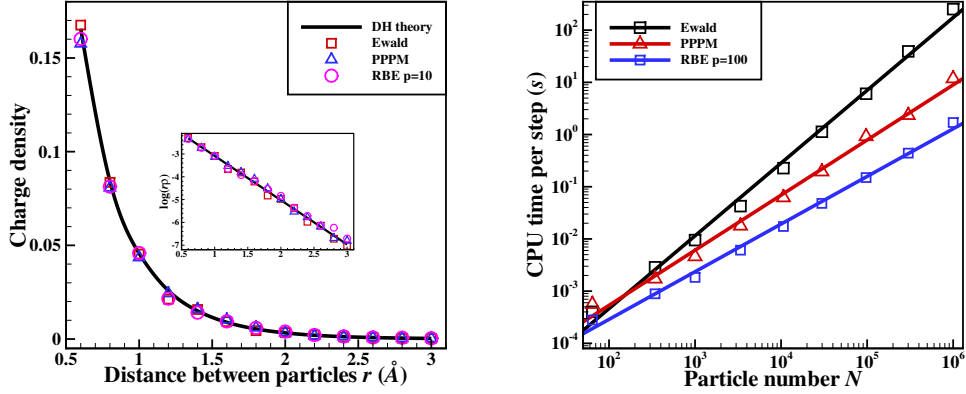


Figure 4: Comparison of the Ewald sum, the PPPM and the RBE methods

	$p = 10$	$p = 20$	$p = 50$	$p = 100$
$\rho_r = 0.1$	0.15%	0.13%	0.13%	0.08%
$\rho_r = 0.3$	0.10%	0.08%	0.04%	0.09%
$\rho_r = 1$	0.66%	0.18%	0.11%	0.04%
$\rho_r = 4$	7.83%	2.38%	0.71%	0.31%

Table 1: Relative error of potential energy for the RBE method against PPPM method with different densities and batch sizes.

simulation results of pure water system also indicate that the RBE type methods can not only sample from the equilibrium distribution, but also compute accurately the dynamical properties of the pure water systems.

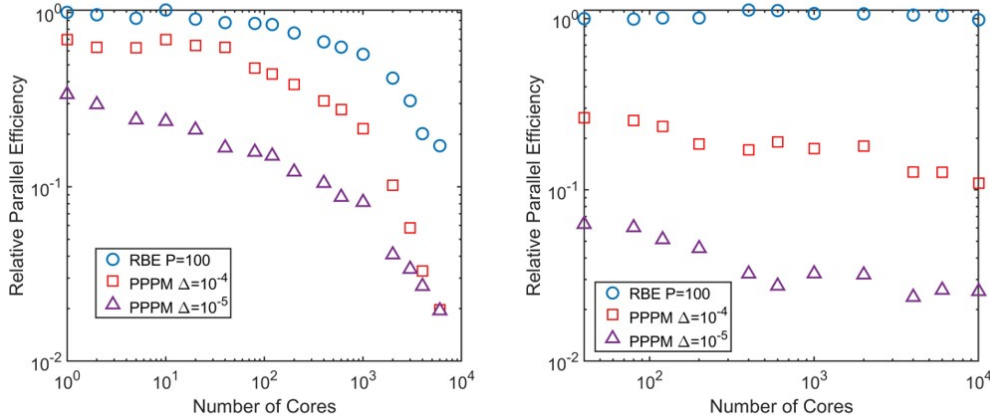


Figure 5: The parallel efficiency of the PPPM and the RBE methods for all-atom simulation of pure water system (Left)  $3 \times 10^5$  atoms; (Right)  $3 \times 10^7$  atoms

## 5 Statistical sampling

Sampling from a complicated or even unknown probability distribution is crucial in many applications, including numerical integration for statistics of many-body systems [34, 29], parameter estimation for Bayesian inference [91, 12] etc.. The methods that rely on random numbers for sampling and numerical simulations are generally called the Monte Carlo (MC)



methods [60, 29]. The law of large numbers [27] validates the usage of empirical measures for approximation of the complicated or unknown probability measure. By the central limit theorem [27], the error of the MC methods scales like  $\mathcal{O}(N^{-1/2})$  which is independent of the dimension  $d$ , hence the MC methods overcome the curse of dimensionality. The Markov Chain Monte Carlo (MCMC) methods [39, 35] are among the most popular MC methods. By constructing Markov chains that have the desired distributions to be the invariant measures, one can obtain samples from the desired distributions by recording the states of the Markov chains. A typical MCMC algorithm is the Metropolis-Hastings algorithm [79, 48].

Unlike the MCMC, the Stein variational Gradient method (proposed by Liu and Wang in [72]) belongs to the class of *particle based* variational inference sampling methods (see also [86, 20]). These methods update particles by solving optimization problems, and each iteration is expected to make progress toward the desired distribution. As a non-parametric variational inference method, SVGD gives a deterministic way to generate points that approximate the desired probability distribution by solving an ODE particle system, which displays different features from the Monte Carlo methods.

We describe in this section two sampling methods that use RBM to improve the efficiency. The first method, Random Batch Monte Carlo, is a fast MCMC that costs only  $\mathcal{O}(1)$  per iteration to sample from the Gibbs measures corresponding to many-body particle systems with singular interacting kernels. The second method, RBM-SVGd, is an interesting application of RBM to the Stein variational gradient descent ODE system, which is an interacting particle system.

## 5.1 Random Batch Monte Carlo for many body systems

Suppose that one wants to sample from the  $N$ -particle Gibbs distribution

$$\pi(\underline{x}) \propto \exp[-\beta H(\underline{x})], \quad (5.1)$$

with  $\underline{x} = (x_1, \dots, x_N) \in \mathbb{R}^{Nd}$  ( $x_i \in \mathbb{R}^d$ , and  $d \geq 1, d \in \mathbb{N}$ ),  $\beta$  being a positive constant, the  $N$ -body energy

$$H(\underline{x}) := \sum_{i=1}^N w_i V(x_i) + \sum_{i,j:i < j} w_i w_j \phi(x_i - x_j), \quad (5.2)$$

and  $V$  being the external potential assumed to be smooth. Here,  $w_i$ 's are the weights. In the molecular regime,  $w_i$ 's are often taken to be 1, while in the mean field regime [90, 37, 65], one may have  $w \sim N^{-1}$ .

In [68], Li et. al. proposed the Random Batch Monte Carlo method, which costs  $\mathcal{O}(1)$  per time step for sampling from equilibrium distributions (Gibbs measures) corresponding to particle systems with singular interacting kernels. Similarly to [75, 49] and the MD methods above, the interacting potential is decomposed into two parts

$$\phi(x) = \phi_1(x) + \phi_2(x), \quad (5.3)$$

where we suppose that  $\phi_1$  has long range but is smooth and bounded, while  $\phi_2$  is singular and of short range. The algorithm is based on the following splitting Monte Carlo, which is a special case of the Metropolis-Hastings algorithm:

Suppose there are  $N$  particles located at  $x_j$  for  $j = 1, \dots, N$ . Let us consider the following method for a Markovian jump.

*Step 1* / Randomly choose a particle  $i$ .

*Step 2* / Move the particle using  $\phi_1$  with overdamped Langevin equation:

$$d\mathbf{r}_i = - \frac{\nabla V(\mathbf{r}_i)}{w(N-1)} + \frac{1}{N-1} \sum_{j:j \neq i} \nabla \phi_1(\mathbf{r}_i - \mathbf{r}_j) dt + \frac{\tilde{Z}}{\hat{E}} \frac{2}{(N-1)w^2\beta} d\mathbf{W}_i, \quad (5.4)$$

$$\mathbf{r}_i(0) = x_i,$$

where  $x_j$ 's are fixed. Evolve this SDE with some time  $t > 0$  and obtain  $\mathbf{r}_i(t) \rightarrow x_i^*$  as a candidate position of particle  $i$  for the new sample.

*Step 3* / Use  $\phi_2$  to do the Metropolis rejection. Define

$$\text{acc}(x_i, x_i^*) = \min \left\{ 1, \exp \left[ -\beta \sum_{j:j \neq i} w^2(\phi_2(x_i^* - x_j) - \phi_2(x_i - x_j)) \right] \right\}. \quad (5.5)$$

With probability  $\text{acc}(x_i, x_i^*)$ , accept  $x_i^*$  and set

$$x_i \leftarrow x_i^*. \quad (5.6)$$

Otherwise,  $x_i$  is unchanged. Then, a new sample  $\{x_1, \dots, x_N\}$  is obtained for the Markov chain.

Note that the overdamped Langevin equation satisfies the detailed balance condition so the above algorithm is a special case of the Metropolis-Hastings algorithm, thus can correctly sample from the desired Gibbs distribution. Due to the short range of  $\phi_2$ , *Step 3* can be done in  $\mathcal{O}(1)$  operations using some standard data structures such as the cell list [34, Appendix F]. The idea is to use the random mini-batch approach to *Step 2*. Hence, one discretizes the SDE with the Euler-Maruyama scheme [62, 80]. The interaction force is approximated within the random mini batch idea. This gives the following algorithm.

---

**Algorithm 7** (Random-batch Monte Carlo algorithm)

---

- 1: Split  $\phi := \phi_1 + \phi_2$  such that  $\phi_1$  is smooth and with long range;  $\phi_2$  is with short range. Generate  $N$  initial particles; choose  $N_s$  (the total number of samples),  $p > 1$ ,  $m \geq 1$
- 2: **for**  $n$  in  $1 : N_s$  **do**
- 3:     Randomly pick an index  $i \in \{1, \dots, N\}$  with uniform probability
- 4:      $\mathbf{r}_i \leftarrow x_i$
- 5:     **for**  $k = 1, \dots, m$  **do**
- 6:         Choose  $\xi_k, \mathbf{z}_k \sim \mathcal{N}(0, I_d)$ ,  $\Delta t_k > 0$  and let,

$$\mathbf{r}_i \leftarrow \mathbf{r}_i - \Delta t_k \left[ \frac{\nabla V(\mathbf{r}_i)}{w(N-1)} + \frac{1}{p-1} \sum_{j \in \xi_k} \nabla \phi_1(\mathbf{r}_i - x_j) \right] + \frac{\hat{\mathbf{E}}}{(N-1)w^2\beta} \mathbf{z}_k$$

- 7:     **end for**
- 8:     Let  $x_i^* \leftarrow \mathbf{r}_i$ . Compute the following using cell list or other data structures:

$$\alpha = \min \left\{ 1, \exp \left[ -\beta \sum_{j:j \neq i} w^2(\phi_2(x_i^* - x_j) - \phi_2(x_i - x_j)) \right] \right\}$$

- 9:     Generate a random number  $\zeta$  from uniform distribution on  $[0, 1]$ . If  $\zeta \leq \alpha$ , set

$$x_i \leftarrow x_i^*$$

- 10: **end for**
- 

It has been proved in [68] that the mini-batch approximation has the an error control for the transition probability so that the method is correct with some systematic error. The computational cost is  $\mathcal{O}(1)$  for each iteration and the efficiency could be higher since there is no rejection in *Step 2*.

We now present a numerical result from [68] to illustrate the efficiency of RBMC. Consider the Dyson Brownian motion [31] :

$$d\lambda_j(t) = -\lambda_j(t) dt + \frac{1}{N-1} \sum_{k:k \neq j} \frac{1}{\lambda_j - \lambda_k} dt + \frac{1}{\sqrt{N-1}} dW_j, \quad j = 1, \dots, N, \quad (5.7)$$

where  $\{\lambda_j\}$ 's represent the eigenvalues of certain random matrices (compared with the original Dyson Brownian motion,  $N - 1$  instead of  $N$  is used in (5.7); there is little effect due to the replacement  $N \rightarrow N - 1$ ). In the limit  $N \rightarrow \infty$ , the distribution obeys the following nonlocal PDE

$$\partial_t \rho(x, t) + \partial_x(\rho(u - x)) = 0, \quad u(x, t) = \pi(H\rho)(x, t) = \text{p.v.} \int_{\mathbb{R}} \frac{\rho(y, t)}{x - y} dy, \quad (5.8)$$

where  $H(\cdot)$  is the Hilbert transform on  $\mathbb{R}$ ,  $\pi = 3.14 \dots$  is the circumference ratio and p.v. represents the Cauchy principal value. From this PDE, one finds that the limiting equation (5.8) has an invariant measure, given by the semicircle law:

$$\rho(x) = \frac{1}{\pi} \sqrt{2 - x^2}. \quad (5.9)$$

Fig.6 shows the sampling results of RBMC and MH methods for empirical measures with particles from the joint distribution

$$\pi(d\underline{x}) \propto \exp \left[ - \left( \frac{N-1}{2} \sum_i x_i^2 - \sum_{i < j} \ln |x_i - x_j| \right) \right],$$

which is the invariant measure for the interacting particle system (5.7). The empirical measure is expected to be close to the semicircle law when  $N$  is large enough. In the simulations, the particle number was fixed as  $N = 500$ . In the RBMC, the splitting was done for  $\ln r$  at  $r = 0.01$ , and the time step was chosen as  $\Delta t = 10^{-4}$ . The MH algorithm uses a certain Gaussian proposal for the random movement of a chosen particle. The left panel of Fig. 6 shows that both methods yield results that agree with the semicircle law reasonably well. The right panel plots the relative error with respect to the semicircle law versus CPU time. Clearly, the RBMC method only needs 10% of the time for the MH method to get the error tolerance considered.

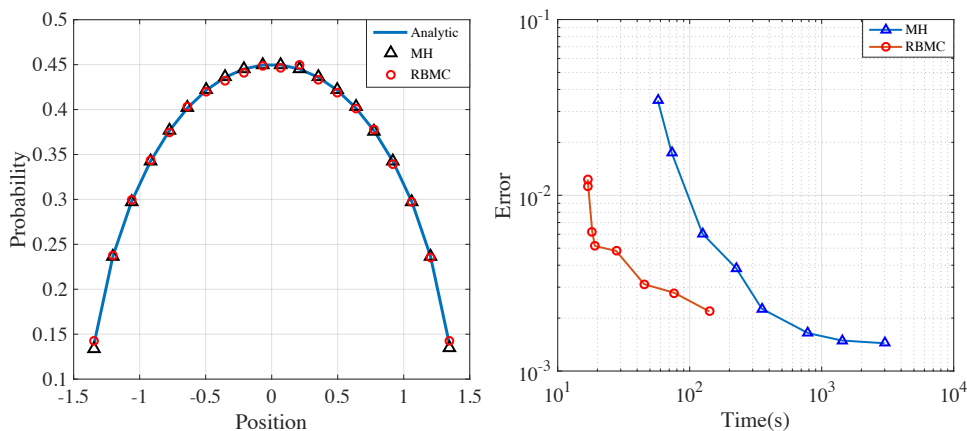


Figure 6: (Left) Empirical densities with  $1e7$  sampling iterations ( $1e7N$  sample points). The blue curve is the analytical curve given by the semicircle law (5.9). (Right) error versus CPU time.

## 5.2 RBM-SVGD: a stochastic version of stein variational gradient descent

Suppose that one is interested in some target probability distribution with density  $\pi(x)$  ( $x \in \mathbb{R}^d$ ). In SVGD, one sets  $V = -\log \pi$ , chooses some symmetric positive definite kernel  $\mathcal{K}(x, y)$  and solves the following ODE system for given initial points  $\{\mathbf{r}_i(0)\}_{i=1}^N$  (see [72, 71]):

$$\dot{\mathbf{r}}_i = \frac{1}{N} \sum_{j=1}^N \nabla_y \mathcal{K}(\mathbf{r}_i, \mathbf{r}_j) - \frac{1}{N} \sum_{j=1}^N \mathcal{K}(\mathbf{r}_i, \mathbf{r}_j) \nabla V(\mathbf{r}_j), \quad i = 1, \dots, N, \quad (5.10)$$

where  $N$  is the number of particles for the sampling purpose. The subindex “ $y$ ” in  $\nabla_y$  means that the gradient is taken with respect to the second variable in  $\mathcal{K}(\cdot, \cdot)$ ; i.e.  $\nabla_y \mathcal{K}(\mathbf{r}_i, \mathbf{r}_j) := \nabla_y \mathcal{K}(x, y)|_{(x,y)=(\mathbf{r}_i, \mathbf{r}_j)}$ . When  $t$  is large enough, the empirical measures constructed using  $\{\mathbf{r}_i(t)\}_{i=1}^N$  is expected to be close to  $\pi$ , i.e.

$$\frac{1}{N} \sum_{i=1}^N \delta(x - \mathbf{r}_i(t)) \approx \pi(x) dx, \quad t \gg 1.$$

SVGD provides consistent estimation for generic distributions as Monte Carlo methods do, but it seems to be more efficient than some Monte Carlo methods in practice level for approximating the desired measure, when the number of particles is small [72, 24]. Interestingly, it reduces to the maximum a posterior (MAP) method when  $N = 1$  [72].

The ODE system (5.10) clearly is an interacting particle system but now the interaction kernel is no longer translation invariant and is not symmetric. The kernel can even grow as  $|\mathbf{r}_i - \mathbf{r}_j| \rightarrow \infty$ . Clearly, for such systems, RBM is applicable. Applying the RBM to this special kernel and using any suitable ODE solvers, one gets a class of sampling algorithms, which is called RBM-SVGD in [66]. The discrete algorithm (with possible variant step size) is shown in Algorithm 8. Clearly, the complexity is  $\mathcal{O}(pN)$  for each iteration.

---

**Algorithm 8** RBM-SVGD

---

- 1: **for**  $k$  in  $0 : N_T - 1$  **do**
- 2:     Divide  $\{1, 2, \dots, pn\}$  into  $n$  batches randomly.
- 3:     **for** each batch  $\mathcal{C}_q$  **do**
- 4:         For all  $i \in \mathcal{C}_q$ ,

$$\mathbf{r}_i^{(k+1)} \leftarrow \mathbf{r}_i^{(k)} + \frac{1}{N} \left( \nabla_y \mathcal{K}(\mathbf{r}_i^{(k)}, \mathbf{r}_i^{(k)}) - \mathcal{K}(\mathbf{r}_i^{(k)}, \mathbf{r}_i^{(k)}) \nabla V(\mathbf{r}_i^{(k)}) \right) \eta_k + \Phi_{k,i} \eta_k,$$

where

$$\Phi_{k,i} = \frac{N-1}{N(p-1)} \sum_{j \in \mathcal{C}_q, j \neq i} \nabla_y \mathcal{K}(\mathbf{r}_i^{(k)}, \mathbf{r}_j^{(k)}) - \mathcal{K}(\mathbf{r}_i^{(k)}, \mathbf{r}_j^{(k)}) \nabla V(\mathbf{r}_j^{(k)}). \quad (5.11)$$

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

Here,  $N_T$  is the number of iterations and  $\{\eta_k\}$  is the sequence of time steps, which play the same role as learning rate in SGD [11, 13]. For some applications, one may simply set  $\eta_k = \eta \ll 1$  to be a constant and get relatively good results. However, in many high dimensional problems, choosing  $\eta_k$  to be constant may yield divergent sequences [87]. One may decrease  $\eta_k$  to obtain convergent data sequences. For example, one may simply choose  $\eta_k = 1/k$  as in SGD. Another frequently used strategy is the AdaGrad approach [26, 95].

We recall the gradient flow under the so-called “Stein metric” in the space of probability measures [71, 36]:

$$\partial_t \rho = \nabla \cdot \left( \rho \mathcal{K} * \left( \rho \nabla \frac{\delta E}{\delta \rho} \right) \right), \quad (5.12)$$

where  $\mathcal{K} * g = \int \mathcal{K}(x, y) g(y) dy$ . Consider taking the energy functional as the Kullback-Leibler (KL) divergence between  $\rho$  and the target distribution  $\pi$ , where KL divergence is also known as the relative entropy defined by

$$\text{KL}(\mu || \nu) = \mathbb{E}_{Y \sim \mu} \log \frac{d\mu}{d\nu}(Y). \quad (5.13)$$

Here  $\frac{d\mu}{d\nu}$  is the well-known Radon-Nikodym derivative. Then, equation (5.12) becomes

$$\partial_t \rho = \nabla \cdot (\rho \mathcal{K} * (\rho \nabla V + \nabla \rho)). \quad (5.14)$$

It is easy to see that  $\pi \propto \exp(-V)$  is invariant under this PDE. See [71, 73] for some relevant studies.

The above theory encounters difficulty for empirical measures because the KL divergence is simply infinity. One benefit of the of the ‘‘Stein metric’’ is that the gradient may be moved from  $\nabla\rho$  onto the kernel  $\mathcal{K}(x, y)$  so that the flow (5.12) becomes (5.10), which is then well-defined. In fact, if  $\{r_i\}$  solves the ODE system (5.10), then the corresponding empirical measure is a measure solution to (5.14) (see [73, Proposition 2.5]). Hence, one may reasonably expect that (5.10) will give approximation for the desired distribution  $\pi$ .

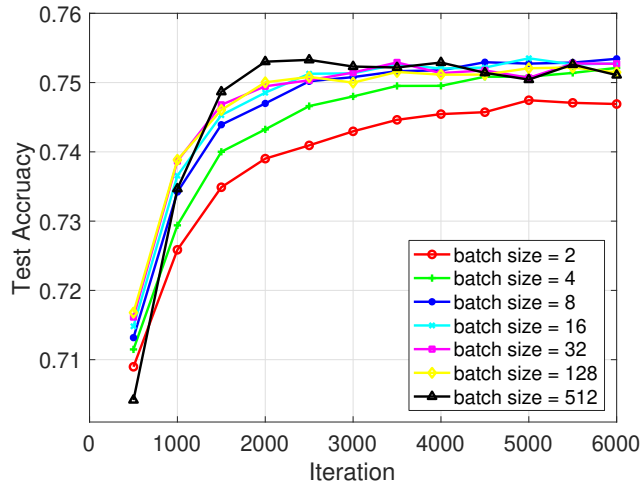


Figure 7: Test accuracy on the Covertypes dataset

For numerical illustration, we take an example from [66]. Consider the logistic regression for binary classification on the Covertypes dataset, with 581012 data points and 54 features [38]. The inference is applied on posterior  $p(x|D)$  with the parameter  $x = [w, \log \alpha]$  being of dimension 55. Here,  $D$  is 80% of the data and the remaining data were used for test. Figure 7 shows the performance of SVGD and RBM-SVGD with  $N = 512$  particles and kernel  $\mathcal{K}(x, y) = k(x - y)$  for a Gaussian kernel  $k(\cdot)$ . Clearly, RBM-SVGD gives comparable results with SVGD, both results being as good as some traditional methods.

Table 2 shows the CPU time and speedup of RBM-SVGD. Clearly, for comparable results, RBM-SVGD is more efficient.

Table 2: Average runtime of 6000 iterations

p	RBM-SVGD						SVG D
	2	4	8	16	32	128	512
Runtime(s)	8.59	11.24	16.28	26.15	21.66	19.42	47.01
Speedup	5.5x	4.2x	2.9x	1.8x	2.2x	2.4x	

## 6 Agent-based models for collective dynamics

Collective behaviors of self-propelled particles (agents) are ubiquitous in nature, for example, synchronous flashing of fireflies and pacemaker cells, swarming of fish, flocking of birds and herding of sheep. We refer to [1, 19, 81, 92, 98] for survey articles and related literature.

While the RBM was introduced as an efficient algorithm for interacting particle systems, one can also view it as a (random) model of the underlying problem, which takes into account only a small number of interactions randomly at discrete time steps. Two natural questions arise with such models: (a) How accurate are these ‘‘new’’ random models compared to the original, full batch models? (b) Do these random models still capture the main features of

the original model, such as the collective or long time behavior, and under what conditions? Here we review some recent results that address these issues for two representative problems, the Cucker-Smale model for flocking and the consensus model.

## 6.1 The Cucker-Smale model

We begin with the Cucker-Smale (CS) model [19]:

$$\begin{cases} \frac{d\mathbf{x}_i}{dt} = \mathbf{v}_i, & t > 0, \quad i = 1, \dots, N, \\ \frac{d\mathbf{v}_i}{dt} = \frac{\kappa}{N-1} \sum_{j:j \neq i} \psi(|\mathbf{x}_j - \mathbf{x}_i|)(\mathbf{v}_j - \mathbf{v}_i), \\ (\mathbf{x}_i(0), \mathbf{v}_i(0)) = (x_i^{in}, v_i^{in}), \end{cases} \quad (6.1)$$

where  $\mathbf{x}_i$  and  $\mathbf{v}_i$  are the position and velocity of the  $i$ -th CS particle, respectively,  $\kappa$  is the non-negative coupling strength and  $\psi$ , the communication weight measuring mutual interactions between agents, is positive, bounded, and Lipschitz continuous and satisfies the monotonicity conditions:

$$\begin{aligned} 0 \leq \psi(r) \leq \psi_M, \quad \forall r \geq 0, \quad \|\psi\|_{\text{Lip}} < \infty, \\ (\psi(r_1) - \psi(r_2))(r_1 - r_2) \leq 0, \quad r_1, r_2 \in \mathbb{R}_+. \end{aligned} \quad (6.2)$$

Here  $\psi_M > 0$  is a constant. The emergent dynamics of (6.1), flocking, in which all particles will eventually stay in a bounded domain with the same velocity, has been extensively studied in literature [46, 47].

Consider the RBM-approximation for (6.1):

$$\begin{cases} \frac{d\tilde{\mathbf{x}}_i}{dt} = \tilde{\mathbf{v}}_i, & t \in [t_{m-1}, t_m), \quad m = 1, 2, \dots, \\ \frac{d\tilde{\mathbf{v}}_i}{dt} = \frac{\kappa}{p-1} \sum_{j \in \mathcal{C}_i^{(k)}, j \neq i} \psi(|\tilde{\mathbf{x}}_j - \tilde{\mathbf{x}}_i|)(\tilde{\mathbf{v}}_j - \tilde{\mathbf{v}}_i), \\ (\tilde{\mathbf{x}}_i(0), \tilde{\mathbf{v}}_i(0)) = (x_i^{in}, v_i^{in}), \quad i = 1, \dots, N. \end{cases} \quad (6.3)$$

Assume that  $\psi$  is long-ranged:

$$1/\psi(r) = \mathcal{O}(r^\beta) \quad \text{as } r \rightarrow \infty \quad \text{for some } \beta \in [0, 1). \quad (6.4)$$

For example, one can take

$$\psi(r) = \frac{1}{(1+r^2)^{\beta/2}}, \quad \beta \in [0, 1).$$

Then [44] establishes the following emergence of a global flocking: there exist positive constants  $\tilde{x}_\infty$  and  $C$  such that

$$\begin{aligned} \sup_{0 \leq t < \infty} \mathbb{E} \left( \frac{1}{N^2} \sum_{i,j=1}^N |\tilde{\mathbf{x}}_i - \tilde{\mathbf{x}}_j|^2 \right) < \tilde{x}_\infty \quad \text{and} \\ \mathbb{E} \left( \frac{1}{N^2} \sum_{i,j=1}^N |\tilde{\mathbf{v}}_i - \tilde{\mathbf{v}}_j|^2 \right) \leq C \exp \left[ - \frac{C(p-1)}{(N-1)(1+\Delta t)} t(1+t)^{-\beta} \right], \end{aligned} \quad (6.5)$$

where  $C$  depends only on  $\psi$ ,  $\beta$ ,  $\kappa$  and the initial data.

Furthermore, the following uniform-in-time error estimate was also proved: when  $\psi$  has a positive lower bound  $\psi_0$ ,

$$\psi(r) \geq \psi_0 \quad \text{for } r \geq 0, \quad (6.6)$$

then

$$\begin{aligned} \mathbb{E} \left( \frac{1}{N} \sum_{i=1}^N |\tilde{\mathbf{v}}_i(t) - \mathbf{v}_i(t)|^2 \right) &\leq C \Delta t \left( \frac{1}{p-1} - \frac{1}{N-1} \right) + C \Delta t^2 \\ &\quad + C(1 + \Delta t) \exp(-\kappa \psi_0 t), \end{aligned} \quad (6.7)$$

where the dependency of the constant  $C$  is the same as in (6.5).

Note that the positive lower bound assumption (6.6) corresponds to the case of  $\beta = 0$  in the long-ranged communication (6.4). However, the third time-decaying term in the right-hand side of (6.7) is independent of  $p$  and  $N$ .

## 6.2 Consensus models

Let  $\mathbf{q}_i \in \mathbb{R}^d$ ,  $1 \leq i \leq N$  be a collection of agents that seek for a consensus, governed by the Cauchy problem:

$$\begin{cases} \frac{d\mathbf{q}_i}{dt} = \nu_i + \frac{\kappa}{N-1} \sum_{j \neq i} a_{ij} \Gamma(\mathbf{q}_j - \mathbf{q}_i), & t > 0, \\ \mathbf{q}_i(0) = \mathbf{q}_i^{in}, & i = 1, \dots, N, \end{cases} \quad (6.8)$$

where  $\kappa$  is a non-negative coupling strength and  $\nu_i$  is the intrinsic velocity of the  $i$ -th agent. Here  $\Gamma$  is an interaction function satisfying the following properties: there exists  $C_1 > 0$  such that

$$\Gamma \in \mathcal{C}^2(B_{C_1}(0)), \quad \Gamma(-q) = -\Gamma(q), \quad \forall q \in \overline{B_{C_1}(0)}. \quad (6.9)$$

Here  $B_r(x)$  is the open ball with radius  $r$  and center  $x$ . We assume, without loss of generality, that the total sum is zero:

$$\sum_{i=1}^N \nu_i = 0,$$

and the adjacency matrix  $(a_{ij})_{i,j=1}^N$  represents the network structure for interactions between agents satisfying symmetry and non-negative conditions:

$$a_{ij} = a_{ji} \geq 0, \quad 1 \leq i, j \leq N.$$

Note that the first term on the R.H.S. of (6.8) induces the "dispersion effect" due to the heterogeneity of  $\nu_i$ . The second term in the R.H.S. of (6.8), modeled by the convolution type consensus force, generates "concentration effect". The overall dynamics of (6.8) is determined by the competitions between dispersion and concentration.

Below we present the study on RBM to this problem in [63]. Consider the RBM-approximation where the interaction term is approximated by the random mini-batch at each time step. Then the relative state  $|\tilde{\mathbf{q}}_i - \tilde{\mathbf{q}}_j|$  for RBM approximation can be *unbounded* even if the original relative state  $|\mathbf{q}_i - \mathbf{q}_j|$  is uniformly bounded. Thus to balance dispersion and interaction in the RBM, one also needs to apply the RBM in the dispersion part as well. A sufficient framework leading to the uniform boundedness of relative states is to introduce suitable decomposition of the dispersion term  $\nu_i$  as a sum of  $N$ -dispersion terms  $\bar{\nu}_{ij}$ :

$$\bar{\nu}_{ij} = -\bar{\nu}_{ji}, \quad \nu_i = \frac{\kappa}{N-1} \sum_{j=1}^N \bar{\nu}_{ij}, \quad i, j = 1, \dots, N. \quad (6.10)$$

Then, the original Cauchy problem (6.8) is equivalent to the following problem:

$$\begin{cases} \frac{d\mathbf{q}_i}{dt} = \frac{\kappa}{N-1} \sum_{j \neq i} (\nu_{ij} + a_{ij} \Gamma(\mathbf{q}_j - \mathbf{q}_i)), & t > 0, \\ \mathbf{q}_i(0) = \mathbf{q}_i^{in}, & i = 1, \dots, N, \end{cases} \quad (6.11)$$

and the RBM samples dispersions and interactions proportionally,

$$\begin{cases} \frac{d\tilde{\mathbf{q}}_i}{dt} = \frac{\kappa}{p-1} \sum_{j \in \mathcal{C}_i^{(k)}, j \neq i} (\bar{v}_{ij} + a_{ij}\Gamma(\tilde{\mathbf{q}}_j - \tilde{\mathbf{q}}_i)), & t \in (t_k, t_{k+1}), \\ \tilde{\mathbf{q}}_i(0) = \mathbf{q}_i^{in}, & i = 1, \dots, N, k = 0, 1, 2, \dots \end{cases} \quad (6.12)$$

We first state the main result for the *one-dimensional* case. Assume that the coupling function  $\Gamma$  is strongly dissipative in the sense that

$$(\Gamma(q_1) - \Gamma(q)) \cdot (q_1 - q) \approx |q_1 - q|^2, \quad \forall q, q_1 \in [-C_1, C_1],$$

and also the full system (6.8) has an equilibrium  $\Phi = (\phi_1, \dots, \phi_N) \in (-C_1, C_1)^N$  with initial data sufficiently close to  $\Phi$ . The main result is the following uniform error estimate, under the condition that the underlying network topology is connected strongly enough:

$$\sup_{0 \leq t < \infty} \left[ \frac{1}{N} \sum_{i=1}^N \mathbb{E} |\tilde{\mathbf{q}}_i(t) - \mathbf{q}_i(t)|^2 \right] \lesssim \left[ \Delta t \frac{1}{p-1} - \frac{1}{N-1} + \Delta t^2 \right].$$

For the multi-dimensional setting with  $\mathbf{q}_i \in \mathbb{R}^d$ , the same error analysis can be obtained under one more extra assumption, which guarantees that the states  $Q := (\mathbf{q}_1, \dots, \mathbf{q}_N)$  and  $\tilde{Q} := (\tilde{\mathbf{q}}_1, \dots, \tilde{\mathbf{q}}_N)$  are confined in the symmetric interval.

Now, we give two main results on the emergent dynamics of (6.12) proved in [43]. Introduce two functionals for  $\tilde{Q} = (\tilde{\mathbf{q}}_1, \dots, \tilde{\mathbf{q}}_N)$ :

$$\mathcal{M}_2(\tilde{Q}) := \frac{1}{N} \sum_{j=1}^N |\tilde{\mathbf{q}}_j|^2, \quad \mathcal{D}(\tilde{Q}) := \max_{1 \leq i, j \leq N} |\tilde{\mathbf{q}}_i - \tilde{\mathbf{q}}_j|.$$

The first main result is concerned with the exponential decay of the second moment of  $\hat{q}_i^R$ : there exists a positive constant  $\Lambda_1 = \Lambda_1(N, P, \tau, \kappa, L_1)$  satisfying

$$\mathbb{E} \left( \mathcal{M}_2(\tilde{Q}(t)) \right) \leq e^{-\Lambda_1 t} \mathbb{E} \left( \mathcal{M}_2(\tilde{Q}(0)) \right), \quad t \geq 0.$$

The second main result deals with almost sure (a.s.) convergence of  $\tilde{Q}$ : there exists a positive constant  $\Lambda_2 = \Lambda_2(N, P, \tau, \kappa, L_1, L_2)$  such that

$$\mathcal{D}(\tilde{Q}(t)) \leq \mathcal{D}(\tilde{Q}(0)) C e^{-\Lambda_2 t}, \quad t \geq 0.$$

We remark that although the exponential decay rates in above results depend on  $N$ , numerical results in [43] show that the decay rates are in not sharp, and they are independent of  $N$ .

## 7 Quantum dynamics

In this section, we have a review of the applications of RBM to interacting particles in the quantum regime. In particular, we first present and comment on the convergence results of RBM applied to the  $N$ -body Schrödinger equation in [41], and then have a review of the application of RBM to quantum Monte Carlo (QMC) methods in [57].

### 7.1 A theoretical result on the $N$ -body Schrödinger equation

The first principle computation is based on solving for complex-valued wave function  $\Psi_N \equiv \Psi_N(t, x_1, \dots, x_N) \in \mathbb{C}$  of the  $N$ -body Schrödinger equation

$$i\partial_t \Psi_N(t, x_1, \dots, x_N) = \mathcal{H}_N \Psi_N(t, x_1, \dots, x_N), \quad \Psi_N|_{t=0} = \Psi_N^{in} \quad (7.1)$$



where  $t \geq 0$  is the time while  $x_m \in \mathbb{R}^d (1 \leq m \leq N)$  is the position of the  $m$ th particle,  $\mathcal{H}_N$  is the quantum Hamiltonian for  $N$  identical particles with unit mass:

$$\mathcal{H}_N := \sum_{m=1}^N -\frac{1}{2}\hbar^{-2}\Delta_{x_m} + \frac{1}{N-1} \sum_{1 \leq \ell < n \leq N} V(x_\ell - x_n), \quad (7.2)$$

while  $\hbar$  is the reduced Planck constant. The  $N$  particles in this system interact via a binary (real-valued) potential  $V$  assumed to be even, bounded and sufficiently regular (at least of class  $C^{1,1}$  on  $\mathbb{R}^d$ ). The coupling constant  $\frac{1}{N-1}$  is chosen in order to balance the summations in the kinetic energy (involving  $N$  terms) and in the potential energy (involving  $\frac{1}{2}N(N-1)$  terms).

When solving (7.1), the computation is exceedingly expensive due to the smallness of  $\hbar$ , which requires small time steps  $\Delta t$  and small mesh sizes of order  $\hbar$  for the convergence of the numerical scheme, due to the oscillation in the wave function  $\Psi_N$  with frequency of order  $1/\hbar$  (see [6, 58]). On top of this, any numerical scheme for (7.1) requires computing, at each time step, the sum of the interaction potential for each particle pair in the  $N$ -particle system, which needs  $\mathcal{O}(N^2)$  operations. The RBM described below reduces the computational cost to  $\mathcal{O}(N)$  per time-step.

Below we follow the presentation of [41]. Assume for simplicity that  $N \geq 2$  is an even integer. Let  $\sigma_1, \sigma_2, \dots, \sigma_j, \dots$  be a mutually independent and uniformly distributed random sequence of permutations. Each permutation  $\sigma \in \mathcal{S}_N$  defines a partition of  $\{1, \dots, N\}$  into  $N/2$  batches of two indices:

$$\{1, \dots, N\} = \bigcup_{k=1}^{N/2} \{\sigma(2k-1), \sigma(2k)\}.$$

Set

$$\mathbf{T}_t(\ell, n) := \begin{cases} 1 & \text{if } \{\ell, n\} = \{\sigma_{[\frac{t}{\Delta t}]+1}(2k-1), \sigma_{[\frac{t}{\Delta t}]+1}(2k)\} \text{ for some } k \leq \frac{N}{2}, \\ 0 & \text{otherwise,} \end{cases} \quad (7.3)$$

and consider the time-dependent random batch Hamiltonian

$$\mathcal{H}_N(t) := \sum_{m=1}^N -\frac{1}{2}\hbar^{-2}\Delta_{x_m} + \sum_{1 \leq \ell < n \leq N} \mathbf{T}_t(\ell, n)V(x_\ell - x_n). \quad (7.4)$$

The RBM then solves the random batch Schrödinger equation

$$i\hbar\partial_t \Psi_N(t, x_1, \dots, x_N) = \mathcal{H}_N(t)\Psi_N(t, x_1, \dots, x_N), \quad \Psi_N|_{t=0} = \Psi_N^{in}. \quad (7.5)$$

Clearly, for each time step the cost of computing the interaction potential is reduced from  $\mathcal{O}(N^2)$  to  $\mathcal{O}(N)$ .

As we have seen, RBM is known to converge in the case of classical dynamics. It is therefore natural to seek an error estimate for the quantum RBM method. The major difficulty here is to obtain an error estimate that is *independent* of  $\hbar$  and  $N$ .

### 7.1.1 Mathematical Setting and Main Result

It will be more convenient to carry out the analysis on the corresponding von Neumann equations

$$i\hbar\partial_t R_N(t) = \mathcal{H}_N R_N(t) - R_N(t)\mathcal{H}_N =: [\mathcal{H}_N, R_N(t)], \quad R_N(0) = R_N^{in}. \quad (7.6)$$

Here we denote  $\mathbb{H} := L^2(\mathbb{R}^d; \mathbb{C})$  and  $\mathbb{H}_N = \mathbb{H}^{\otimes N} \simeq L^2((\mathbb{R}^d)^N; \mathbb{C})$  for each  $N \geq 2$ . The algebra of bounded operators on  $\mathbb{H}$  is denoted by  $\mathcal{L}(\mathbb{H})$ , while  $\mathcal{L}^1(\mathbb{H}) \subset \mathcal{L}(\mathbb{H})$  and  $\mathcal{L}^2(\mathbb{H})$  are respectively the two-sided ideals of trace-class and Hilbert-Schmidt operators on  $\mathbb{H}$ . The

operator norm of  $A \in \mathcal{L}(H)$  is denoted  $\|A\|$ . A density operator on  $H$  is a trace-class operator  $R$  on  $H$  such that

$$R = R^* \geq 0 \quad \text{and} \quad \text{trace}_{\mathfrak{H}}(R) = 1.$$

The set of density operators on a separable Hilbert space  $H$  is henceforth denoted  $\mathcal{D}(H)$ .

The random batch von Neumann equation is

$$i\partial_t \tilde{R}_N(t) = [\mathcal{H}(t), \tilde{R}_N(t)], \quad \tilde{R}_N(0) = R_N^{in}. \quad (7.7)$$

In order to find an error estimate for the RBM that is independent of the particle number  $N$ , one first needs to define in terms of  $R_N(t)$  and  $\tilde{R}_N(t)$  *quantities of interest* to be compared that are *independent of  $N$* . A common practice when considering large systems of identical particles is to study the reduced density operators, which unfortunately does not lead to  $N$ -independent error estimates [41]. Assume that  $R_N^{in}$  has an integral kernel  $r^{in} \equiv r^{in}(x_1, \dots, x_N; y_1, \dots, y_N)$  satisfying the symmetry

$$r^{in}(x_1, \dots, x_N; y_1, \dots, y_N) = r^{in}(x_{\sigma(1)}, \dots, x_{\sigma(N)}; y_{\sigma(1)}, \dots, y_{\sigma(N)}) \quad (7.8)$$

for each permutation  $\sigma \in \mathfrak{S}_N$ . Then, for each  $t \geq 0$ , the  $N$ -body density operator  $R_N(t)$  solution of (7.6) satisfies the same symmetry, i.e. its integral kernel of the form  $r(t; x_1, \dots, x_N; y_1, \dots, y_N)$  also satisfies

$$r(t; x_1, \dots, x_N; y_1, \dots, y_N) = r(t; x_{\sigma(1)}, \dots, x_{\sigma(N)}; y_{\sigma(1)}, \dots, y_{\sigma(N)}) \quad (7.9)$$

for each permutation  $\sigma \in \mathfrak{S}_N$ . The 1-particle reduced density operator of  $R_N(t) \in \mathcal{D}(H_N)$  is  $R_{N,1}(t) \in \mathcal{D}(H)$  defined by the integral kernel

$$r_{\mathbf{1}}(t, x, y) := \int_{(\mathbb{R}^d)^{N-1}} r(t; x, z_2, \dots, z_N; y, z_2, \dots, z_N) dz_2 \dots dz_N. \quad (7.10)$$

Even if  $R_N^{in}$  satisfies the symmetry (7.8), in general  $\tilde{R}_N(t)$  does not satisfy the symmetry analogous to (7.9) for  $t > 0$  (with  $r$  replaced with  $\tilde{r}$ , an integral kernel for  $\tilde{R}_N(t)$ ) because the random batch potential

$$\sum_{1 \leq \ell < n \leq N} \mathbf{T}_t(\ell, n) V(x_\ell - x_n)$$

is not invariant under permutations of the particle labels. For that reason, the 1-particle reduced density operator of  $\tilde{R}_N(t)$  one needs is  $\tilde{R}_{N,1}(t) \in \mathcal{D}(H)$  defined for all  $t > 0$  by the integral kernel

$$\tilde{r}_{\mathbf{1}}(t, x, y) := \frac{1}{N} \sum_{j=1}^N \int_{(\mathbb{R}^d)^{N-1}} \tilde{r}(t; Z_{j,N}[x], Z_{j,N}[y]) d\hat{Z}_{j,N}, \quad (7.11)$$

with the notation

$$Z_{j,N}[x] := z_1, \dots, z_{j-1}, x, z_{j+1}, \dots, z_N, \quad d\hat{Z}_{j,N} = dz_1 \dots dz_{j-1} dz_{j+1} \dots dz_N.$$

(Obviously (7.11) holds with  $r_{\mathbf{1}}$  and  $r$  in the place of  $\tilde{r}_{\mathbf{1}}$  and  $\tilde{r}$  respectively because of the symmetry (7.9).)

We also need to introduce the Wigner functions of the density operators  $R_N(t)$  and  $\tilde{R}_N(t)$ . Let  $s \equiv s(x, y) \in L^2(\mathbb{R}^d \times \mathbb{R}^d)$  be an integral kernel of operator  $S \in \mathcal{L}^2(\mathcal{H})$ . Then the Wigner function of  $S$  is defined by the formula

$$W_{-}[S](x, \cdot) := \frac{1}{(2\pi)^d} \mathcal{F}(y \mapsto s(x + \frac{1}{2}\tilde{y}, x - \frac{1}{2}\tilde{y})) \quad \text{for a.e. } x \in \mathbb{R}^d, \quad (7.12)$$

where  $\mathcal{F}$  is the Fourier transform on  $L^2(\mathbb{R}^d)$ .

For each integer  $M \geq 1$ , we also introduce the dual norm

$$\|f\|_{-M} := \sup \left\{ \left| \iint_{\mathbb{R}^d \times \mathbb{R}^d} f(x, \xi) \overline{a(x, \xi)} dx d\xi \right| \mid \begin{array}{l} a \in C_c(\mathbb{R}^d \times \mathbb{R}^d), \text{ and} \\ \max_{\substack{|\alpha|, |\beta| \leq M \\ |\alpha| + |\beta| > 0}} \|\partial_x^\alpha \partial_\xi^\beta a\|_{L^\infty(\mathbb{R}^d \times \mathbb{R}^d)} \leq 1 \end{array} \right\}. \quad (7.13)$$

The main results in [41] is the following theorem.

**Theorem 7.1.** *Assume that  $N \geq 2$  and that  $V \in C(\mathbb{R}^d)$  is a real-valued function such that*

$$V(z) = V(-z) \text{ for all } z \in \mathbb{R}^d, \quad \lim_{|z| \rightarrow +\infty} V(z) = 0, \quad \text{and} \quad \int_{\mathbb{R}^d} (1 + |\omega|^2) |\mathcal{F}(V)(\omega)| d\omega < \infty.$$

*Let  $R_{N,1}(t)$  and  $\tilde{R}_{N,1}(t)$  be the single-particle reduced density operators defined in terms of  $R_N(t)$  and  $R_N^R(t)$  respectively by (7.10). Then there exists a constant  $\gamma_d > 0$  depending only on the dimension  $d$  of the configuration space such that, for each  $t > 0$ , one has*

$$\begin{aligned} & \| |W_-[\tilde{R}_{N,1}(t)] - W_-[R_{N,1}(t)]| \|_{-[d/2]-3} \\ & \leq 2\gamma_d \Delta t e^{6t \max(1, \sqrt{d}L(V))} \Lambda(V) (2 + 3t\Lambda(V) \max(1, \Delta t) + 4\sqrt{d}L(V)t\Delta t), \end{aligned} \quad (7.14)$$

where

$$L(V) := \frac{1}{(2\pi)^d} \int_{\mathbb{R}^d} |\omega|^2 |\hat{V}(\omega)| d\omega, \quad \Lambda(V) := \frac{1}{(2\pi)^d} \int_{\mathbb{R}^d} \sum_{\mu=1}^d |\omega^\mu| |\hat{V}(\omega)| d\omega,$$

with  $\omega^\nu$  the  $\nu$ -th component of  $\omega$ .

This error estimate gives an error *independent* of  $\sim$  and  $N$ . It was also pointed out in [41] that the error bound obtained in above theorem is small as  $\Delta t \rightarrow 0$ , *even for moderate values of  $N$*  for which the factor  $\frac{1}{N-1}$  is insignificant. Therefore the result applies to  $N$ -body quantum Hamiltonians *without the  $\frac{1}{N-1}$  normalization of the interaction potential*, as a simple corollary for each finite value of  $N \geq 2$ .

**Remark 7.1.** Note that the dual norm (7.13) is a kind of weak norm. The error bound  $\mathcal{O}(\Delta t)$  is consistent with the weak error estimate in [54].

## 7.2 Quantum Monte-Carlo methods

Computing the ground state energy of a many-body quantum system is a fundamental problem in chemistry. An important tool to determine the ground state energy and electron correlations is the quantum Monte Carlo (QMC) method [94, 4].

Consider the Hamiltonian,

$$\mathcal{H} = \sum_{i=1}^N -\frac{\hbar^2}{2m} \Delta_{x_i} + \sum_{i \neq j} W(x_i - x_j) + \sum_{i=1}^N V_{\text{ext}}(x_i). \quad (7.15)$$

Here  $V_{\text{ext}}$  is the external potential given by

$$V_{\text{ext}}(x_i) = \sum_{\alpha=1}^M U(x_i - R_\alpha), \quad (7.16)$$

where  $R_\alpha$ , for instance, can be the position of an atom.

Up to some global phase factor, the ground state takes real values and is nonnegative everywhere. The ground state and the corresponding eigenvalue can be obtained via the Rayleigh quotient,

$$E = \min_{\Phi_N} \frac{\int_{(\mathbb{R}^3)^N} \Phi_N \mathcal{H} \Phi_N d\mathbf{x}}{\int_{(\mathbb{R}^3)^N} |\Phi_N|^2 d\mathbf{x}}, \quad (7.17)$$

where the minimizer  $\Phi_N$  corresponds to the ground state wave function. The main computational challenge here is the curse of dimensionality due to the high dimensional integral.

In the variational Monte Carlo (VMC) framework, the ground state is approximated by selecting an appropriate ansatz  $\Phi_N \approx \Phi_0$ . Traditionally,  $\Phi_0$  is constructed using the one-body wave functions, by taking into the effect of particle correlations described by the Jastrow factors [32]. For example, in the Boson systems like the liquid Helium interacting with a graphite surface [78, 97, 84], the following ansatz has been proven successful,

$$\Phi_0 = e^{-J(\underline{x})} \prod_{i=1}^N \phi(x_i), \quad J(\underline{x}) = \frac{1}{2} \sum_{i,j:i \neq j} u(|x_i - x_j|), \quad u(r) = \left(\frac{a}{r}\right)^5 + \frac{b^2}{r^2 + c^2}. \quad (7.18)$$

The non-negative one-particle wave function is often taken as

$$\phi(x_i) = \sum_{\alpha=1}^M e^{-\theta(x_i - R_\alpha)}, \quad (7.19)$$

for some function  $\theta$ . This form has been used in [97] and the parameters were obtained by solving a one-dimensional Schrödinger equation. With the approximation of  $\Phi_N$  being fixed, the multi-dimensional integral is then interpreted as a statistical average. In fact, introducing the probability density function (PDF),

$$p(\underline{x}) \propto |\Phi_0(\underline{x})|^2, \quad (7.20)$$

the ground state energy is the average of  $E_{\text{tot}}$  under  $p(\underline{x})$ , where

$$E_{\text{tot}}(\underline{x}) = \frac{\mathcal{H}\Phi_0}{\Phi_0}. \quad (7.21)$$

Hence,  $E$  can be computed by a Monte Carlo procedure, and such a method is called the VMC, which is a typical QMC method.

In the VMC methods, the ground state is not updated. Instead, one may use another QMC method—the diffusion Monte Carlo (DMC) method [3, 85]—to compute the ground state and the energy. In particular, one solves a pseudo-time Schrödinger equation (TDSE) which is a parabolic equation [85]

$$\partial_t \Psi_N = (E_T - \mathcal{H}_N) \Psi_N. \quad (7.22)$$

Here,  $t$  represents a fictitious time. The energy shift  $E_T$  is adjusted on-the-fly based on the change of magnitude of the wave function. Instead of solving (7.22) directly, it is often more practical to find  $f(\underline{r}, t)$  with

$$f(\underline{x}, t) = \Psi_N(\underline{x}, t) \Phi_0(\underline{x}). \quad (7.23)$$

By choosing  $\Psi_N(\underline{x}, 0) = \Phi_0(\underline{x})$ ,  $f(\underline{x}, 0) = |\Phi_0|^2 \propto p(\underline{x})$ . Hence, a VMC method may be used to initialize  $f(\underline{x}, t)$ . Clearly,  $f$  solves the following differential equation [85],

$$\partial_t f = -\nabla \cdot \left( \frac{\underline{v}}{m} f \right) + \frac{\underline{v}^2}{2m} f - (E_T - E_{\text{tot}}(\underline{x})) f, \quad (7.24)$$

where  $\underline{v} = (v_1, \dots, v_N) \in \mathbb{R}^{Nd}$  and

$$v_i(\underline{x}) = \nabla \log \phi(x_i) - \sum_{j:j \neq i} \nabla_{x_i} u(|x_i - x_j|).$$

The average energy  $E(t)$  is then defined as a weighted average,

$$E(t) = \frac{\int_{(\mathbb{R}^3)^N} f(\underline{x}, t) E_{\text{tot}}(\underline{x}) d\underline{x}}{\int_{(\mathbb{R}^3)^N} f(\underline{x}, t) d\underline{x}}, \quad (7.25)$$

where the correctness can be seen by  $E(t) = \int \Psi_N \mathcal{H} \Phi_0 d\underline{x} / \int \Psi_N \Phi_0 d\underline{x}$ . If  $\Psi_N$  is close to the eigenstate, this will be close to  $E$ .

The key observation is that the dynamics (7.24) can be associated with a stochastic process, in which the particles are experiencing birth/death while driven by drift velocity and diffusion. This process can be implemented by a number of walkers together with birth/death processes [3, 85].

### 7.2.1 The Random Batch Method for VMC

With (7.18), the density (7.20) can be found as

$$p(\underline{x}) \propto e^{-2V}, \quad V = -\ln \Psi_0 = -\sum_i \log \phi(x_i) + \frac{1}{2} \sum_i \sum_{j \neq i} u(|x_i - x_j|), \quad (7.26)$$

and the total energy can be expressed as

$$E_{\text{tot}}(\underline{x}) = -\frac{\hbar^2}{2m} \Delta V - \frac{\hbar^2}{2m} \|\nabla V\|^2 + \sum_{i \neq j} W(x_i - x_j) + \sum_{i=1}^N \sum_{\alpha=1}^M U(x_i - R_\alpha). \quad (7.27)$$

To sample from  $p(\underline{x})$ , one may make use of the Markov chain Monte Carlo (MCMC) methods. Consider the over-damped Langevin dynamics,

$$d\mathbf{r}_i = \nabla \log \phi(\mathbf{r}_i) dt - \sum_{j: j \neq i} \nabla_{\mathbf{r}_i} u(|\mathbf{r}_i - \mathbf{r}_j|) dt + d\mathbf{W}_i(t), \quad 1 \leq i \leq N. \quad (7.28)$$

Under suitable conditions [76], the dynamical system with potential given by (7.26) is ergodic and the PDF  $p(\underline{x})$  in (7.26) is the unique equilibrium measure of (7.28). By the classical Euler-Maruyama method ([62]), the underdamped Langevin can be discretized to a Markov Chain:

$$\mathbf{r}_i(t + \Delta t) = \mathbf{r}_i(t) + \nabla \log \phi(\mathbf{r}_i) \Delta t - \sum_{j \neq i} \nabla_{\mathbf{r}_i} u(|\mathbf{r}_i(t) - \mathbf{r}_j(t)|) \Delta t + \Delta W_i, \quad 1 \leq i \leq N, \quad (7.29)$$

where  $\Delta W_i$  is again sampled from  $\mathcal{N}(0, \Delta t)$ . It is clear that  $\mathcal{O}(M + N)$  operations should be taken for each particle at each time step.

The cost of the above MCMC is high. The strategy in [57] is to apply a RBM strategy with replacement. In particular, at each step, one randomly picks two particles,  $i$  and  $j$ , and compute their interactions,  $\nabla_{\mathbf{r}_i} u(|\mathbf{r}_i - \mathbf{r}_j|)$ , then updates their positions as follows,

$$\begin{aligned} \mathbf{r}_i(t + \Delta t) &= \mathbf{r}_i(t) + \nabla \log \phi(\mathbf{r}_i) \Delta t + (N - 1) \nabla_{\mathbf{r}_i} u(|\mathbf{r}_i - \mathbf{r}_j|) \Delta t + \Delta W_i, \\ \mathbf{r}_j(t + \Delta t) &= \mathbf{r}_j(t) + \nabla \log \phi(\mathbf{r}_j) \Delta t + (N - 1) \nabla_{\mathbf{r}_j} u(|\mathbf{r}_i - \mathbf{r}_j|) \Delta t + \Delta W_j. \end{aligned} \quad (7.30)$$

For the one-body term  $\nabla \log \phi(\mathbf{r}_i)$ ,

$$\nabla \log \phi(\mathbf{r}_i) = \sum_{\alpha=1}^M -\nabla \theta(\mathbf{r}_i - R_\alpha) q_\alpha^i, \quad q_\alpha^i = \frac{e^{-\theta(\mathbf{r}_i - R_\alpha)}}{\sum_{\beta=1}^M e^{-\theta(\mathbf{r}_i - R_\beta)}}, \quad (7.31)$$

where the coefficients  $q_\alpha^i$ 's are non-negative and  $\sum_\alpha q_\alpha^i = 1$ . To reduce the cost, one may further use a direct Monte-Carlo method: pick *just one* term  $\alpha$  randomly. Specifically, assume that one starts with  $\alpha$  and computes  $e_{\text{old}} = \theta(\mathbf{r}_i - R_\alpha)$ , and then one randomly picks  $1 \leq \beta \leq M$ , and computes  $e_{\text{new}} = \theta(\mathbf{r}_i - R_\beta)$ .  $\beta$  is accepted with probability

$$p_{\text{acc}} \propto \exp[-(e_{\text{new}} - e_{\text{old}})]. \quad (7.32)$$

For the detailed algorithm see [57]. As a result of the random sampling of the one- and two-body interactions, updating the position of each particle *only requires*  $\mathcal{O}(1)$  operations per time-step. Another practical issue emerges when the interaction  $u(|x|)$  has a singularity

near zero. One can use the splitting idea as mentioned in section 4.1, i.e., applying RBM only to the long-range smooth part.

It was shown in [57] that the above random batch algorithm, when applied to one batch of two particles, has the same accuracy as the Euler-Maruyama method over a time step of  $2\Delta t/N$ . One full time step in Euler-Maruyama method corresponds to  $N/2$  such steps in the random batch algorithm. This corresponds to the random batch method with replacement.

We show a numerical experiment performed in [57] on  ${}^4\text{He}$  atoms interacting with a two-dimensional lattice. The CPU times taken to move the 300 Markov chains for 1000 steps were compared. In this comparison, the cost associated with the energy calculations was excluded in the random batch and Euler-Maruyama methods. From Table 3, one clearly sees that the RBM is more efficient than the Euler-Maruyama method. It is much more efficient than the random walk Metropolis-Hastings algorithm, mainly because the latter method requires the calculation of the energy at *every* step.

Table 3: CPU times (seconds) for several VMC methods.

	Random Walk Metropolis- Hastings	Euler- Maruyama	Random Batch
CPU time for a 1000-step sampling period	1503	469	54

### 7.2.2 The Random Batch Method for DMC

Viewing (7.24), one may consider an ensemble of  $L$  copies of the system, also known as walkers [3]. For each realization, one first solves the SDEs corresponding to the drift and diffusion, which is the same as the overdamped Langevin as in VMC up to a time scaling. Hence, the same Random Batch Algorithm in the VMC can be used for this part.

The relaxation term  $-(E_T - E_{\text{tot}})f$  is then done by using a birth/death process to determine whether a realization should be removed or duplicated. For each walker, one computes a weight factor,

$$w(t + \Delta t) = \exp \left[ \Delta t (E_T - \frac{1}{2}(E_{\text{tot}}(\mathbf{r}) + E_{\text{tot}}(\mathbf{r}')) \right]. \quad (7.33)$$

This weight determines how the walker should be removed or duplicated. See [57] for more details. The primary challenge is that computing the energy at each step requires  $\mathcal{O}((N + M)N)$  operations in order to update the position of  $N$  particles. To reduce this part of the computation cost, one rewrites the total energy as

$$E_{\text{tot}}(\mathbf{r}) = \sum_{i=1}^N E_1(\mathbf{r}_i) + \sum_{1 \leq i < j \leq N} E_2(\mathbf{r}_i, \mathbf{r}_j) + \sum_{1 \leq i < j < k \leq N} E_3(\mathbf{r}_i, \mathbf{r}_j, \mathbf{r}_k), \quad (7.34)$$

where

$$\begin{aligned} E_1(\mathbf{r}_i) &= -\frac{\tilde{\phantom{2}}}{2m} \nabla^2 \ln \phi(\mathbf{r}_i) - \frac{\tilde{\phantom{2}}}{2m} |\nabla \ln \phi(\mathbf{r}_i)|^2 + \sum_{\alpha=1}^M U(\mathbf{r}_i - R_\alpha), \\ E_2(\mathbf{r}_i, \mathbf{r}_j) &= -\frac{\tilde{\phantom{2}}}{m} \nabla^2 \ln u(r_{ij}) + \frac{\tilde{\phantom{2}}}{m} (\nabla \ln \phi(\mathbf{r}_i) - \nabla \ln \phi(\mathbf{r}_j)) \cdot \nabla u(r_{ij}) + \frac{\tilde{\phantom{2}}}{m} |\nabla u(r_{ij})|^2 + W(r_{ij}), \\ E_3(\mathbf{r}_i, \mathbf{r}_j, \mathbf{r}_k) &= \frac{\tilde{\phantom{2}}}{m} \left[ \nabla u(r_{ij}) \cdot \nabla u(r_{ik}) + \nabla u(r_{ji}) \cdot \nabla u(r_{jk}) + \nabla u(r_{ki}) \cdot \nabla u(r_{kj}) \right]. \end{aligned} \quad (7.35)$$

Here,  $\mathbf{r}_{ij} = \mathbf{r}_i - \mathbf{r}_j$  and  $r_{ij} = |\mathbf{r}_{ij}|$ . The three-body terms arise because of the  $\|\nabla V\|^2$  term in (7.27).

In the random batch algorithm proposed in [57], one randomly picks a batch  $C_I$  with three particles:  $C_I = \{i, j, k\}$ . One first updates the position of the three particles (drift and

diffusion) by solving the overdamped Langevin dynamics using the random batch algorithm with batch size 3. Then, one then defines a *local* energy,

$$\begin{aligned}
E_I(\mathbf{r}_i, \mathbf{r}_j, \mathbf{r}_k) &= E_1(\mathbf{r}_i) + E_1(\mathbf{r}_j) + E_1(\mathbf{r}_k) \\
&+ \frac{N-1}{2} \left[ E_2(\mathbf{r}_i, \mathbf{r}_j) + E_2(\mathbf{r}_j, \mathbf{r}_k) + E_2(\mathbf{r}_k, \mathbf{r}_i) \right], \\
&+ \frac{(N-1)(N-2)}{2} E_3(\mathbf{r}_i, \mathbf{r}_j, \mathbf{r}_k).
\end{aligned} \tag{7.36}$$

where in  $E_1$ , the sum  $\sum_{\alpha=1}^M$  can be further reduced by a mini-batch strategy. Computing this local energy is clearly  $\mathcal{O}(1)$ . To avoid frequent removal and duplication of walkers, the branching process is applied after  $N/3$  batches of particles are updated. In this case, the weight function is defined by collecting the local energy from each batch (denoted by  $I_m$  here),

$$w(\mathbf{r}) = \exp \left[ \Delta t (E_T - \dot{\mathbf{H}}_{\text{tot}}) \right], \quad \dot{\mathbf{H}}_{\text{tot}} = \sum_{m=1}^{N/3} E_{I_m}. \tag{7.37}$$

Because of the smallness of  $\Delta t$ , the expectation of  $w_I$  equals  $w(t + \Delta t)$  modulus an error of  $\mathcal{O}(\Delta t^2)$ . See [57] for the verification using the Green's functions.

The detailed algorithm can be found in [57] and we omit it here. Now we show a test of the RBM-DMC algorithm conducted in [57], which compares the results with the direct DMC method. For the initialization, a VMC method using the ansatz (7.18) for the wave function  $\Phi_0$  was first applied. The random walk Metropolis-Hastings Monte Carlo method is used in both methods so that they start at the same states. 300 ensembles are created by sub-sampling one sample out of every 500 steps from the VMC runs to avoid correlations among the ensembles. For both methods,  $\Delta t = 10^{-4}$  was used and 200,000 steps of simulations were run. The CPU run-time is recorded for various system sizes. More specifically, the system size is increased from the original 168 particles, to  $N = 378$ ,  $N = 672$  and  $N = 1050$  particles, and in each case, the direct DMC and the RBM-DMC were run for 1000 steps. As shown in Figure 8, the CPU time for the direct DMC method increases much more rapidly as  $N$  increases.

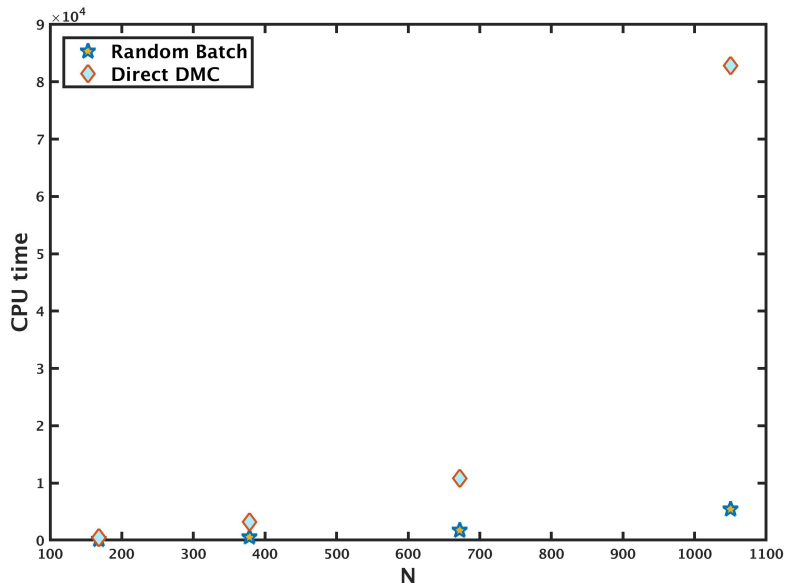


Figure 8: A comparison of the CPU runtime (in seconds) for running 1000 steps of DMC.

## Acknowledgement

S. Jin was partially supported by the NSFC grant No.12031013. The work of L. Li was partially sponsored by NSFC 11901389, 11971314, and Shanghai Sailing Program 19YF1421300. Both authors were also supported by Shanghai Science and Technology Commission Grant No. 20JC144100.

## References

- [1] G. Albi, N. Bellomo, L. Fermo, S-Y Ha, J. Kim, L. Pareschi, D. Poyato, and J. Soler. Vehicular traffic, crowds, and swarms: From kinetic theory and multiscale methods to applications and research perspectives. *Mathematical Models and Methods in Applied Sciences*, 29(10):1901–2005, 2019.
- [2] G. Albi and L. Pareschi. Binary interaction algorithms for the simulation of flocking and swarming dynamics. *Multiscale Modeling & Simulation*, 11(1):1–29, 2013.
- [3] J. B. Anderson. A random-walk simulation of the Schrödinger equation: H+3. *The Journal of Chemical Physics*, 63(4):1499–1503, 1975.
- [4] J. B. Anderson. *Quantum Monte Carlo: origins, development, applications*. Oxford University Press, 2007.
- [5] H. Babovsky and R. Illner. A convergence proof for Nanbu’s simulation method for the full Boltzmann equation. *SIAM journal on numerical analysis*, 26(1):45–65, 1989.
- [6] W. Bao, S. Jin, and P. A. Markowich. On time-splitting spectral approximations for the Schrödinger equation in the semiclassical regime. *Journal of Computational Physics*, 175(2):487–524, 2002.
- [7] J. Barnes and P. Hut. A hierarchical  $O(N \log N)$  force-calculation algorithm. *Nature*, 324:446–449, 1986.
- [8] 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.
- [9] U. Biccari and E. Zuazua. A stochastic approach to the synchronization of coupled oscillators. *Front. Energy Res.*, 8(115), 2020.
- [10] G. A. Bird. Approach to translational equilibrium in a rigid sphere gas. *The Physics of Fluids*, 6(10):1518–1519, 1963.
- [11] L. Bottou. Online learning and stochastic approximations. *On-line learning in neural networks*, 17(9):142, 1998.
- [12] George EP Box and George C Tiao. *Bayesian inference in statistical analysis*, volume 40. John Wiley & Sons, 2011.
- [13] S. Bubeck. Convex optimization: Algorithms and complexity. *Foundations and Trends® in Machine Learning*, 8(3-4):231–357, 2015.
- [14] H. B. Callen and T. A. Welton. Irreversibility and generalized noise. *Physical Review*, 83(1):34, 1951.
- [15] 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.
- [16] 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.



- [17] 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.
- [18] G. Ciccotti, D. Frenkel, and I. R. McDonald. *Simulation of liquids and solids: Molecular Dynamics and Monte Carlo Methods in Statistical Mechanics*. North-Holland, Amsterdam, 1987.
- [19] F. Cucker and S. Smale. Emergent behavior in flocks. *IEEE Transactions on automatic control*, 52(5):852–862, 2007.
- [20] B. Dai, N. He, H. Dai, and L. Song. Provable Bayesian inference via particle mirror descent. In *Artificial Intelligence and Statistics*, pages 985–994, 2016.
- [21] 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.
- [22] 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.
- [23] Markus Deserno and Christian Holm. How to mesh up Ewald sums. II. An accurate error estimate for the particle-particle particle-mesh algorithm. *The Journal of Chemical Physics*, 109(18):7694–7701, 1998.
- [24] G. Detommaso, T. Cui, Y. Marzouk, A. Spantini, and R. Scheichl. A Stein variational Newton method. In *Advances in Neural Information Processing Systems*, pages 9187–9197, 2018.
- [25] Z. H. Duan and R. Krasny. An Ewald summation based multipole method. *J. Chem. Phys.*, 113:3492–3495, 2000.
- [26] J. Duchi, E. Hazan, and Y. Singer. Adaptive subgradient methods for online learning and stochastic optimization. *Journal of Machine Learning Research*, 12(Jul):2121–2159, 2011.
- [27] R. Durrett. *Probability: Theory and Examples*. Cambridge Series in Statistical and Probabilistic Mathematics. Cambridge University Press, 4 edition, 2010.
- [28] R. Durstenfeld. Algorithm 235: random permutation. *Communications of the ACM*, 7(7):420, 1964.
- [29] Weinan E, Tiejun Li, and Eric Vanden-Eijnden. *Applied stochastic analysis*, volume 199. American Mathematical Soc., 2019.
- [30] A. Eberle, A. Guillin, and R. Zimmer. Couplings and quantitative contraction rates for Langevin dynamics. *The Annals of Probability*, 47(4):1982–2010, 2019.
- [31] L. Erdos and H.-T. Yau. Dynamical approach to random matrix theory. *Courant Lecture Notes in Mathematics*, 28, 2017.
- [32] WMC Foulkes, Lubos Mitas, RJ Needs, and Guna Rajagopal. Quantum monte carlo simulations of solids. *Reviews of Modern Physics*, 73(1):33, 2001.
- [33] R. H. French, V. A. Parsegian, R. Podgornik, R. F. Rajter, A. Jagota, J. Luo, D. Asthagiri, M. K. Chaudhury, Y.-M. Chiang, S. Granick, S. Kalinin, M. Kardar, R. Kjellander, D. C. Langreth, J. Lewis, S. Lustig, D. Wesolowski, J. S. Wettlaufer, W.-Y. Ching, M. Finnis, F. Houlihan, O. A. von Lilienfeld, C. J. van Oss, and T. Zemb. Long range interactions in nanoscale science. *Rev. Mod. Phys.*, 82(2):1887–1944, 2010.
- [34] D. Frenkel and B. Smit. *Understanding molecular simulation: from algorithms to applications*, volume 1. Elsevier, 2001.

- [35] D. Gamerman and H. F. Lopes. *Markov chain Monte Carlo: stochastic simulation for Bayesian inference*. Chapman and Hall/CRC, 2006.
- [36] Y. Gao and J.-G. Liu. A note on parametric Bayesian inference via gradient flows. *Annals of Mathematical Sciences and Applications*, 5(2):261–282, 2020.
- [37] 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.
- [38] S. Gershman, M. Hoffman, and D. Blei. Nonparametric variational inference. In *Proceedings of the 29th International Conference on Machine Learning*, pages 235–242, 2012.
- [39] W. R. Gilks, S. Richardson, and D. Spiegelhalter. *Markov chain Monte Carlo in practice*. Chapman and Hall/CRC, 1995.
- [40] F. Golse. The mean-field limit for the dynamics of large particle systems. *Journées equations aux dérivées partielles*, 9:1–47, 2003.
- [41] F. Golse, S. Jin, and T. Paul. The random batch method for  $n$ -body quantum dynamics. *J. Comp. Math.*, To appear.
- [42] L. Greengard and V. Rokhlin. A fast algorithm for particle simulations. *J. Comput. Phys.*, 73:325–348, 1987.
- [43] S.-Y. Ha, S. Jin, D. Kim, and D. Ko. Convergence toward equilibrium of the first-order consensus model with random bath interactions. preprint.
- [44] S.-Y. Ha, S. Jin, D. Kim, and D. Ko. Uniform-in-time error estimate of the random batch method for the Cucker-Smale model. *Math. Model Methods Appl. Sci.*, to appear.
- [45] S.-Y. Ha and Z. Li. Complete synchronization of Kuramoto oscillators with hierarchical leadership. *Communications in Mathematical Sciences*, 12(3):485–508, 2014.
- [46] 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.
- [47] S.-Y. Ha and E. Tadmor. From particle to kinetic and hydrodynamic descriptions of flocking. *arXiv preprint arXiv:0806.2182*, 2008.
- [48] W. K. Hastings. *Monte Carlo Sampling Methods Using Markov Chains and Their Applications*. Oxford University Press, 1970.
- [49] B. Hetenyi, K. Bernacki, and B. J. Berne. Multiple "time step" Monte Carlo. *J. Chem. Phys.*, 117(18):8203–8207, 2002.
- [50] W. G. Hoover. Canonical dynamics: Equilibrium phase-space distributions. *Physical review A*, 31(3):1695, 1985.
- [51] D. Horstmann. From 1970 until present: the Keller-Segel model in chemotaxis and its consequences. *Jahresber. Dtsch. Math.-Ver.*, 105:103–165, 2003.
- [52] S. Jin and L. Li. On the mean field limit of the Random Batch Method for interacting particle systems. *Science China Mathematics*, pages 1–34, 2021.
- [53] S. Jin, L. Li, and J.-G. Liu. Random Batch methods (RBM) for interacting particle systems. *Journal of Computational Physics*, 400:108877, 2020.
- [54] S. Jin, L. Li, and J.-G. Liu. Convergence of the random batch method for interacting particles with disparate species and weights. *SIAM Journal on Numerical Analysis*, 59(2):746–768, 2021.

- [55] S. Jin, L. Li, and Y. Sun. On the Random Batch Method for second order interacting particle systems. *arXiv preprint arXiv:2011.10778*, 2020.
- [56] S. Jin, L. Li, Z. Xu, and Y. Zhao. A random batch Ewald method for particle systems with Coulomb interactions. *arXiv preprint arXiv:2010.01559*, 2020.
- [57] S. Jin and X. Li. Random batch algorithms for quantum Monte Carlo simulations. *Commun. Comput. Phys.*, 28(5):1907–1936, 2020.
- [58] S. Jin, P. Markowich, and C. Sparber. Mathematical and computational methods for semiclassical Schrödinger equations. *Acta Numerica*, 20:121–209, 2011.
- [59] J. K. Johnson, J. A. Zollweg, and K. E. Gubbins. The Lennard-Jones equation of state revisited. *Molecular Physics*, 78(3):591–618, 1993.
- [60] M. H. Kalos and P. A. Whitlock. *Monte Carlo methods*. John Wiley & Sons, 2009.
- [61] K. Kawasaki. Simple derivations of generalized linear and nonlinear Langevin equations. *Journal of Physics A: Mathematical, Nuclear and General*, 6(9):1289, 1973.
- [62] P. E. Kloeden and E. Platen. *Numerical solution of stochastic differential equations*, volume 23. Springer Science & Business Media, 2013.
- [63] D. Ko, S.-Y. Ha, S. Jin, and D. Kim. Uniform error estimates for the Random Batch Method to the first-order consensus models with anti-symmetric interacting kernels. *Studies Appl. Math.*, to appear.
- [64] D. Ko and E. Zuazua. Model predictive control with random batch methods for a guiding problem. *Mathematical Models and Methods in Applied Sciences*, To appear. arXiv preprint arXiv:2004.14834.
- [65] J.-M. Lasry and P.-L. Lions. Mean field games. *Japanese journal of mathematics*, 2(1):229–260, 2007.
- [66] L. Li, Y. Li, J.-G. Liu, Z. Liu, and J. Lu. A stochastic version of Stein variational gradient descent for efficient sampling. *Communications in Applied Mathematics and Computational Science*, 15(1):37–63, 2020.
- [67] 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.
- [68] L. Li, Z. Xu, and Y. Zhao. A random-batch Monte Carlo method for many-body systems with singular kernels. *SIAM Journal on Scientific Computing*, 42(3):A1486–A1509, 2020.
- [69] J. Liang, P. Tan, Y. Zhao, H. Liu, L. Li, S. Jin, L. Hong, and Z. Xu. Linear scalable algorithm to calculate coulomb interaction. *preprint*, In preparation.
- [70] Evgenii Mikhailovich Lifshitz and Lev Petrovich Pitaevskii. *Statistical physics: theory of the condensed state*, volume 9. Elsevier, 2013.
- [71] Q. Liu. Stein variational gradient descent as gradient flow. In *Advances in neural information processing systems*, pages 3115–3123, 2017.
- [72] Q. Liu and D. Wang. Stein variational gradient descent: A general purpose bayesian inference algorithm. In *Advances In Neural Information Processing Systems*, pages 2378–2386, 2016.
- [73] J. Lu, Y. Lu, and J. Nolen. Scaling limit of the stein variational gradient descent: The mean field regime. *SIAM J. Math. Anal.*, 51(2):648–671, 2019.

- [74] B. A. Luty, M. E. Davis, I. G. Tironi, and W. F. Van Gunsteren. A comparison of particle-particle, particle-mesh and Ewald methods for calculating electrostatic interactions in periodic molecular systems. *Mol. Simul.*, 14:11–20, 1994.
- [75] M. G. Martin, B. Chen, and J. I. Siepmann. A novel Monte Carlo algorithm for polarizable force fields: application to a fluctuating charge model for water. *The Journal of chemical physics*, 108(9):3383–3385, 1998.
- [76] 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*, 101(2):185–232, 2002.
- [77] 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.
- [78] W. L. McMillan. Ground state of liquid  $he_4$ . *Physical Review*, 138(2A):A442, 1965.
- [79] N. Metropolis, A. W. Rosenbluth, M. N. Rosenbluth, A. H. Teller, and E. Teller. Equation of state calculations by fast computing machines. *J. Chem. Phys.*, 21(6):1087–1092, 1953.
- [80] G. N. Milstein and M. V. Tretyakov. *Stochastic numerics for mathematical physics*. Springer Science & Business Media, 2013.
- [81] S. Motsch and E. Tadmor. Heterophilious dynamics enhances consensus. *SIAM Review*, 56(4):577–621, 2014.
- [82] K. Nanbu. Direct simulation scheme derived from the Boltzmann equation. i. mono-component gases. *Journal of the Physical Society of Japan*, 49(5):2042–2049, 1980.
- [83] S. Nosé. A molecular dynamics method for simulations in the canonical ensemble. *Molecular physics*, 52(2):255–268, 1984.
- [84] T. Pang. Diffusion Monte Carlo: a powerful tool for studying quantum many-body systems. *American Journal of Physics*, 82(10):980–988, 2014.
- [85] P. J. Reynolds, D. M. Ceperley, B. J. Alder, and W. A. Lester Jr. Fixed-node quantum Monte Carlo for molecules. *The Journal of Chemical Physics*, 77(11):5593–5603, 1982.
- [86] D. J. Rezende and S. Mohamed. Variational inference with normalizing flows. In *International Conference on Machine Learning*, pages 1530–1538, 2015.
- [87] H. Robbins and S. Monro. A stochastic approximation method. *The Annals of Mathematical Statistics*, pages 400–407, 1951.
- [88] V. Rokhlin. Rapid solution of integral equations of classical potential theory. *Journal of computational physics*, 60(2):187–207, 1985.
- [89] F. Santambrogio. Optimal transport for applied mathematicians. *Birkhäuser, NY*, pages 99–102, 2015.
- [90] H. E. Stanley. *Phase transitions and critical phenomena*. Clarendon Press, Oxford, 1971.
- [91] Albert Tarantola. *Inverse problem theory and methods for model parameter estimation*. SIAM, 2005.
- [92] J. Toner and Y. Tu. Flocks, herds, and schools: A quantitative theory of flocking. *Physical review E*, 58(4):4828, 1998.
- [93] 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.

- [94] W. von der Linden. A quantum Monte Carlo approach to many-body physics. *Physics Reports*, 220(2-3):53–162, 1992.
- [95] R. Ward, X. Wu, and L. Bottou. Adagrad stepsizes: sharp convergence over nonconvex landscapes. In *International Conference on Machine Learning*, pages 6677–6686, 2019.
- [96] 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.
- [97] PA Whitlock, GV Chester, and B Krishnamachari. Monte carlo simulation of a helium film on graphite. *Physical Review B*, 58(13):8704, 1998.
- [98] A. T. Winfree. *The geometry of biological time*, volume 12. Springer Science & Business Media, 2001.
- [99] L. Ying, G. Biros, and D. Zorin. A kernel-independent adaptive fast multipole algorithm in two and three dimensions. *J. Comput. Phys.*, 196:591–626, 2004.