MULTISCALE MODEL. SIMUL.
Vol. 20, No. 2, pp. 741-768

ON THE RANDOM BATCH METHOD FOR SECOND ORDER INTERACTING PARTICLE SYSTEMS*

SHI JIN†, LEI LI‡, AND YIQUN SUN§

Abstract. We investigate several important issues regarding the random batch method (RBM) for second order interacting particle systems. We first show the uniform-in-time strong convergence of RBM for second order systems under suitable contraction conditions when the interaction kernels are regular. Second, we propose some variants of RBM for second order systems that can have singular interaction kernels via a kernel splitting strategy, and investigate numerically the application of such methods to molecular dynamics.

Key words. second order systems, hypocoercivity, random batch, kernel splitting, molecular dynamics

MSC codes. 65C20, 37M05

DOI. 10.1137/20M1383069

1. Introduction.

A significant number of important phenomena in physical, social, and biological sciences are described at the microscopic level by interacting particle systems, which exhibit interesting features. Examples include fluids and plasma [21, 6], swarming [51, 11, 10, 17], chemotaxis [28, 3], flocking [16, 25, 1], synchronization [14, 24], and consensus [45], to name a few. These interacting particle systems can be described in general by the first order systems

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

or the second order systems

$$\begin{aligned} dX^i &= V^i\,dt, \\ dV^i &= \left[b(X^i) + \alpha_N \sum_{j:j\neq i} K(X^i - X^j) - \gamma V^i \right] dt + \sigma\,dW^i. \end{aligned} \tag{1.2}$$

Here, $X^i \in \mathbb{R}^d$ are the labels for the particles, and $b(\cdot)$ is some given external field. The stochastic processes $\{W^i\}_{i=1}^N$ are i.i.d. Wiener processes, or the standard Brownian

*Received by the editors November 30, 2020; accepted for publication (in revised form) February 10, 2022; published electronically June 24, 2022.
https://doi.org/10.1137/20M1383069
Funding: This work is supported by the National Key R&D Program of China, Project 2020YFA0712000, and by Shanghai Science and Technology Commission grant 21JC1403700. The fist author was partially supported by NSFC grant 12031013. The second author was partially sponsored by NSFC grant 11901389 and by Shanghai Sailing Program 19YF1421300. The third author was partially supported by the National Key Research and Development Program 2017YFB0202902.

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

‡Corresponding author. School of Mathematical Sciences, Institute of Natural Sciences, Qing Yuan Research Institute, MOE-LSC, Shanghai Jiao Tong University, Shanghai, 200240, People's Republic of China (leili2010@sjtu.edu.cn).

§School of Business, Shanghai Dianji University, Shanghai, 201306, People's Republic of China (yqsun5-c@my.cityu.edu.hk).

motions. We will loosely call X^i the "locations" or "positions," and V^i the velocities of the particles, though the specific meaning can be different in different situations. The function $K(\cdot) : \mathbb{R}^d \to \mathbb{R}^d$ is the interaction kernel. If $\gamma = \sigma = 0$ and $b = -\nabla U$ for some potential U, one has a Hamiltonian system like the one for electrons in plasma [53]. For the molecules in the heat bath [35, 9], X^i and V^i are the physical positions and velocities, described by the underdamped Langevin equations, where σ and γ satisfy the so-called fluctuation-dissipation relation

$$\sigma = \sqrt{2\gamma/\beta}, \tag{1.3}$$

where β 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 of the second order systems (1.2) with the fluctuation-dissipation relation satisfied.

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) [48] is able to reduce the complexity to $\mathcal{O}(N)$ for fast enough decaying interactions. However, the prefactor in the linear scaling could be large and the implementation of FMM is nontrivial. A simple random algorithm, called the random batch method (RBM), was proposed in [31] to reduce the computation cost per time step from $\mathcal{O}(N^2)$ to $\mathcal{O}(N)$, based on the simple "mini-batch" idea. The "random mini-batch" idea is famous for its application in the so-called stochastic gradient descent (SGD) [47, 7, 8] for machine learning problems. The idea was also used for Markov chain Monte Carlo methods like the stochastic gradient Langevin dynamics (SGLD) by Welling and Teh [54] and the random batch Monte Carlo methods [42], and also for the computation of the mean-field flocking model [1, 11] motivated by Nanbu's algorithm of the direct simulation Monte Carlo method [5, 46, 2]. 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 interacting particle systems in [31], this is realized by random grouping and then allowing the particles to interact only within the groups for each small time subinterval. Compared with FMM, the accuracy of RBM is lower, but RBM is much simpler to implement and is valid for more general potentials (e.g., the SVGD ODE [40]). In addition, the prefactor in the linear scaling is small (see section 4.2 for more discussions). 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 (see [31] for a more detailed explanation). Hence, one may understand such methods as certain Monte Carlo methods. If there is mixing and ergodicity for the systems, the simulation can converge well.

RBM for interacting particle systems has been used or extended in various directions, from sampling [40, 42, 33] to molecular dynamics [32, 41] and control of synchronization [4, 36]. RBM has been shown to converge for finite time intervals if the interaction kernels are good enough [40, 31], and in particular an error analysis for deterministic Newton-type second order systems is given in the appendix of [31]. Moreover, a convergence result of RBM for the N-body Schrödinger equation is also obtained in [23]. For long time behaviors, it is expected that the method works for systems that have ergodicity and mixing properties, like systems in contact with a heat bath and that converge to equilibria. Previous rigorous studies of such systems mainly focus on first order systems due to good contraction and mixing properties [31, 30]. Second order systems are, however, more common in nature, especially systems in contact with a heat bath that are very important for molecular dynamics [21].

Whether RBM can be applied directly to obtain good results for direct molecular dynamics simulation needs careful study both in theory and in practice. For closed systems that are Hamiltonian, like the particle systems for the Vlasov–Poisson equations (in this case $\gamma = \sigma = 0$), RBM may be applied to get correct simulation for finite time, but the long time behavior is not clear for these systems. Hence, in this work we mainly focus on systems that are in contact with a heat bath.

In this work, our goal is two-fold. First, we aim to prove rigorously that RBM converges for large times with certain contraction conditions for second order systems (1.2). The theoretic analysis is possible for systems with regular interaction kernels while it becomes challenging for singular kernels. The singular kernels, however, are very common in applications like in molecular dynamics simulations. In fact, not only is analysis challenging for singular kernels, but direct application of RBM could also bring in numerical instability. Hence, as a second goal, we aim to combine the random grouping with the kernel splitting strategies as in [43, 26, 42] so that RBM could be practically applied for systems with singular kernels (see section 2.2) and thus applied to molecular dynamics simulations (see section 4).

Now let us discuss the regimes to consider. In the mean field limit regime [49, 22, 37], one chooses

$$\alpha_N = \frac{1}{N-1} \tag{1.4}$$

so that as $N \to \infty$ the empirical distribution $\mu^{(N)} := N^{-1} \sum_{i=1}^{N} \delta(x - X^i) \otimes \delta(v - V^i)$ converges almost surely under the weak topology to the solutions of the limiting PDE

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

The particle system (1.2) can also be regarded as a numerical particle method for solving the mean field PDE (1.5). Examples of such PDEs include the granular media equations [12] and the Vlasov equations for which $\gamma = \sigma = 0$ [53]. In [31], it has been shown that RBM is asymptotic-preserving for first order systems regarding the mean-field limit, which means the algorithm can approximate the one-marginal distribution with error bound independent of N. Below, we show in section 3 that RBM is also asymptotic-preserving regarding the mean-field limit for second order systems under suitable conditions.

In the molecular dynamics simulations, one chooses $\alpha_N = 1$, and the equations are basically given by

$$\begin{aligned} dX^i &= V^i \, dt, \\ dV^i &= \left[-\sum_{j:j\neq i} \nabla \phi(X^i - X^j) \right] dt + d\xi^i. \end{aligned} \tag{1.6}$$

Here, $\phi(\cdot)$ is the interaction potential and $d\xi^i$ means the interaction with the environment that changes the momentum, which we will discuss in section 4.1. Depending on how to model the coupling to the heat batch, one may choose different thermostats like the Andersen thermostat, the Langevin dynamics, or the Nosé–Hoover thermostat, etc., so different expressions for $d\xi^i$ can be used (see section 4.1). Though $\alpha_N = 1$ is often chosen for molecular dynamics, one may do time and spatial rescalings to match the mean field regime $\alpha_N = 1/(N-1)$ factor. However, the scaling is not crucial for simulation of molecular dynamics (see the discussion in section 4.2); hence, in this

molecular dynamics regime with $\alpha_N = 1$, we will apply RBM directly when it has benefits without scaling it to the mean-field regime.

The rest of the paper is organized as follows. In section 2, we give a brief introduction to RBM and introduce the potential splitting so that RBM can be applied for systems with singular interaction kernels. In section 3, we establish the long time strong error estimate for regular kernels under certain contraction conditions. We provide some discussions on the details on applying RBM with kernel splitting to simulations of molecular dynamics in section 4. Some numerical experiments are performed in section 5 to verify the claims and validate the methods.

2. The algorithms. In this section, we give a detailed explanation of RBM for second order interacting particle systems and then propose the variants of RBM for singular interaction kernels via a kernel splitting strategy. The analysis, which is currently possible for regular kernels, will be given in section 3. The details of implementation of RBM with kernel splitting for one of the important applications of second order systems with singular kernels, the molecular dynamics simulations, will be given in section 4.

Let us briefly explain the random grouping strategy for RBM in [31] that realizes the mini-batch idea for interacting particle systems. Let $T > 0$ be the simulation time, and choose a time step $\tau > 0$. Pick a batch size $p \ll N$, $p \geq 2$ 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 $t_k := k\tau$, $k \in \mathbb{N}$. 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.

2.1. RBM for regular kernels. Applying the above strategy to the second order system (1.2) with interacting forces that do not have singularity yields the method as shown in Algorithm 2.1.

The method shown in Algorithm 2.1 shares some similarity with the stochastic gradient Hamiltonian Monte Carlo (SGHMC) method with friction proposed in [13, sections 3.2–3.3], which is a Markov chain Monte Carlo method for Bayesian inference and machine learning. The difference is that the method shown in Algorithm 2.1 uses random grouping for interacting particles, while SGHMC uses random samples to compute the approximating gradients; i.e., the ways to implement mini-batch are

Algorithm 2.1 (RBM for (1.2)).

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

$$\begin{aligned} dX^i &= V^i\, dt, \\ dV^i &= \left[b(X^i) + \frac{\alpha_N(N-1)}{p-1} \sum_{j \in \mathcal{C}_q, j \neq i} K(X^i - X^j) - \gamma V^i \right] dt + \sigma\, dW^i. \end{aligned} \tag{2.1}$$

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

different. The SGHMC in [13] is a sampling method and the momentum will be resampled occasionally. Since the underdamped Langevin system does not satisfy the detailed balance, using it as a block for the Markov chain may result in some systematic error. The method shown in Algorithm 2.1 is a direct simulation approach for the underdamped Langevin equation, so it can be used for dynamical simulation to capture the transition behaviors approximately, and it can also be used for sampling from the equilibrium.

Remark 2.1. Despite the difference mentioned above, we remark that the random grouping strategy can be viewed as a particular stochastic gradient as in [47, 54] when $K(x) = -\nabla\phi(x)$. In fact, for this case, we introduce the Nd-dimensional vector $\mathfrak{X} := (X^1, \dots, X^N)$ and consider the full interacting energy corresponding to $\alpha_N = 1/(N-1)$:

$$E(\mathfrak{X}) := \frac{1}{2(N-1)} \sum_{i \neq j} \phi(X^i - X^j). \tag{2.2}$$

In [47, 54], the stochastic gradient can be computed by choosing any subset of terms in the sum (2.2). For random grouping in [31], one is only allowed to choose the summands in a particular way. For a given set of random batches $\mathcal{C} = \{\mathcal{C}_1, \dots, \mathcal{C}_n\}$, one may use the random variable

$$\tilde{E}(\mathfrak{X}) := \frac{1}{2(p-1)} \sum_{q=1}^{n} \sum_{k,\ell \in \mathcal{C}_q} \phi(X^k - X^\ell) \tag{2.3}$$

to approximate $E(\mathfrak{X})$, and using its gradient for the dynamics leads to the random grouping in [31] in the case $K = -\nabla\phi$, though RBM in [31] applies to more general kernels.

2.2. RBM with kernel splitting. If the interaction kernel K is singular at $x = 0$, which is often the case in applications like the molecular dynamics simulations [15, 21], direct discretization of the equations in Algorithm 2.1 can lead to numerical instability. For first order systems, in the case $p = 2$, one may take advantage of the time-splitting method to accurately solve the singular part to eliminate the instability [31, 41]. For second order systems or first order systems with $p \geq 3$, the time splitting trick does not apply anymore, and applying RBM directly leads to poor results. To resolve this issue, we adopt the splitting strategy in [43, 26, 42].

In fact, one decomposes the interacting force K into two parts:

$$K(x) = K_1(x) + K_2(x). \tag{2.4}$$

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. The part $K_2(x)$ is a bounded smooth function. With this decomposition, we then apply RBM to the K_2 part only. The resulted method is shown in Algorithm 2.2. Now, if the particles do not accumulate and the local density remains low, the summation in K_1 can be done practically in $\mathcal{O}(1)$ time for given i due to the short range. This can be implemented using data structures like cell-list [21, Appendix F]. In many applications, K_1 is a repulsive force. One may put the particles on a lattice at $t = 0$ as in a solid, and the repulsive term K_1 could forbid them from getting too close so that the particles could stay separated in simulations. Hence, the cost per time step is practically $\mathcal{O}(N)$ though not theoretically guaranteed.

Algorithm 2.2 RBM with splitting for (1.1) and (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** k in $1 : [T/\tau]$ **do**
3: Divide $\{1, 2, \ldots, N = pn\}$ into n batches randomly.
4: **for** each batch $\mathcal{C}_q$ **do**
5: Update X^i's ($i \in \mathcal{C}_q$) by solving for $t \in [t_{k-1}, t_k)$

$$
\begin{aligned}
dX^i = \Bigg(& b(X^i) + \alpha_N \sum_{j:j\neq i} K_1(X^i - X^j) \\
& + \frac{\alpha_N (N-1)}{p-1} \sum_{j\in\mathcal{C}_q, j\neq i} K_2(X^i - X^j) \Bigg) dt + \sigma\, dW^i,
\end{aligned} \tag{2.5}
$$

or

$$
\begin{aligned}
dX^i &= V^i\, dt, \\
dV^i &= \left[b(X^i) + \alpha_N \sum_{j:j\neq i} K_1(X^i - X^j) - \gamma V^i \right] dt \\
&\quad + \frac{\alpha_N (N-1)}{p-1} \sum_{j\in\mathcal{C}_q, j\neq i} K_2(X^i - X^j)\, dt + \sigma\, dW^i.
\end{aligned} \tag{2.6}
$$

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

Since K_2 is bounded, RBM can be applied well due to the boundedness of variance, without introducing too much error. Moreover, we are not applying RBM on the K_1 part, which is a repulsive force as in many applications, so that the particles will not get too close. Hence, in practical simulations, the singularity in K_1 may not be seen and one may use a reasonable time step for the numerical simulations. We discuss some implementation details of the application to the molecular dynamics simulations in section 4 and show some numerical results in section 5.2.

3. A strong convergence analysis. In this section, we perform a strong convergence analysis of RBM for the second order systems (1.2) in the mean field regime (i.e., $\alpha_N = 1/(N-1)$). For the theoretic analysis, we only consider regular interaction kernels. In the appendix of [31], a convergence analysis of RBM for second order systems was given, but only for finite time horizon. In this work, we aim to establish the long time error control under certain conditions. The proof of such results largely makes use of the underlying contraction property for the underdamped Langevin equations [44, 20]. Note that, due to the degeneracy of the noise terms, the contraction should be proved by suitably chosen variables and Lyapunov functions.

For notational convenience, we denote (X_i, V_i) to be the solutions given by (1.2). We denote by $(\tilde{X}_i, \tilde{V}_i)$ the solutions given by the RBM process (2.1). We again use the synchronization coupling as in [31, 30]:

$$X^i(0) = \tilde{X}^i(0) \sim \mu_0,\ W^i = \tilde{W}^i. \tag{3.1}$$

Let $\mathcal{C}_q^{(k)}$ be the batches at t_k where $1 \le q \le n$. Define

$$\mathcal{C}^{(k)} := \left\{\mathcal{C}_1^{(k)}, \dots, \mathcal{C}_n^{(k)}\right\} \tag{3.2}$$

to be the random division of batches at t_k. By the Kolmogorov extension theorem [19], there exists a probability space $(\Omega, \mathcal{F}, \mathbb{P})$ such that the random variables $\{X_0^i, W^i, \mathcal{C}^{(k)} : 1 \le i \le N, k \ge 0\}$ are all defined on this probability space and are independent. Then $\mathbb{E}$ corresponds to the integration on Ω with respect to the probability measure $\mathbb{P}$. We introduce the $L^2(\cdot)$ norm

$$\|v\| = \sqrt{\mathbb{E}|v|^2}, \tag{3.3}$$

and the filtration $\{\mathcal{F}_k\}_{k\ge 0}$ is given by

$$\mathcal{F}_{k-1} = \sigma(X_0^i, W^i(t), \mathcal{C}^{(j)}; t \le t_{k-1}, j \le k-1). \tag{3.4}$$

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

For finite time intervals, the convergence of RBM is straightforward, as shown below in Proposition 3.1. The proof is similar to that of the results in [31, 30], and so we omit it.

PROPOSITION 3.1. *Let $b(\cdot)$ be Lipschitz continuous, and $|\nabla^2 b|$ has polynomial growth. The interaction kernel K is Lipschitz continuous. Then*

$$\sup_{t\in[0,T]} \sqrt{\mathbb{E}|\tilde{X}^1 - X^1|^2 + \mathbb{E}|\tilde{V}^1 - V^1|^2} \le C(T)\sqrt{\frac{\tau}{p-1} + \tau^2}, \tag{3.5}$$

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

Below, we consider the error estimate for long times. This is important if one uses RBM as a sampling method for the invariant measure of (1.2). The main challenge in establishing the uniform-in-time error control compared with the proof in [31] is determining how to make use of the contraction property for second order random systems. The following conditions will give certain contraction property for the second order systems, which can be utilized by considering new variables as in (3.20).

Assumption 3.1. Suppose $b = -\nabla U$ for some $U \in C^2(\mathbb{R}^d)$ that is bounded below (i.e., $\inf_x U(x) > -\infty$), and there exist $\lambda_M \ge \lambda_m > 0$ such that the eigenvalues of $H := \nabla^2 U$ satisfy

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

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

$$\gamma > \sqrt{\lambda_M + 2L},\ \lambda_m > 2L. \tag{3.6}$$

Remark 3.2. The assumptions here are a little different from those for first order systems (see [31, 30]): (1) b is assumed to be Lipschitz instead of one-sided Lipschitz;

(2) we are not assuming the second derivatives of K to be bounded, as there is no white noise in the equations for X_i so trajectories of X_i's are much smoother; (3) the second requirement in (3.6) is the same as the contraction assumption for first order systems in [31], while the first requirement in (3.6) is something new for second order systems to ensure the contraction property.

Under the assumptions above, we are able to establish the following uniform strong convergence estimate.

THEOREM 3.3. *Under Assumption* 3.1 *and the coupling* (3.1), *the solutions to* (1.2) *and* (2.1) *satisfy*

$$\sup_{t\geq 0}\sqrt{\mathbb{E}|\tilde{X}^1(t)-X^1(t)|^2+\mathbb{E}|\tilde{V}^1(t)-V^1(t)|^2}\leq C\sqrt{\frac{\tau}{p-1}+\tau^2}, \tag{3.7}$$

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

We give some lemmas which will be useful later. Below, the generic constants (like C and C_q, etc.) are independent of time t and particle number N unless explicitly stated otherwise. Their concrete meanings may change from line to line.

Denote

$$\mathfrak{X}=(X^1,\dots,X^N),\quad \tilde{\mathfrak{X}}=(\tilde{X}^1,\dots,\tilde{X}^N). \tag{3.8}$$

For a given configuration $\mathtt{x}:=(x^1,\dots,x^N)\in\mathbb{R}^{Nd}$, denote the random batches of $\{1,\dots,N\}$ as $\mathcal{C}:=\{\mathcal{C}_1,\dots,\mathcal{C}_n\}$. We introduce the random variables I_{ij} to indicate whether the two particles are in the same batch or not:

$$I_{ij}=\begin{cases}1 & \exists\ \mathcal{C}_q,\{i,j\}\subset\mathcal{C}_q,\\ 0 & \text{otherwise,}\end{cases}\qquad 1\leq i,j\leq N. \tag{3.9}$$

Accordingly, we use $I_{ij}^{(k)}$ to indicate whether particles i and j are in the same batch or not at t_k, corresponding to $\mathcal{C}^{(k)}$. Moreover, we will set

$$\theta(i):=q$$

for which $i\in\mathcal{C}_q$.

The auxiliary results, Lemmas 3.4–3.6 below, are in [30], and we omit their proofs.

LEMMA 3.4. *For $i\neq j$, it holds that*

$$\mathbb{E}I_{ij}=\frac{p-1}{N-1}, \tag{3.10}$$

and for distinct i,j,ℓ, it holds that

$$\mathbb{P}(I_{ij}I_{i\ell}=1)=\mathbb{E}I_{ij}I_{i\ell}=\frac{(p-1)(p-2)}{(N-1)(N-2)}. \tag{3.11}$$

For a given configuration $\mathtt{x}=(x^1,\dots,x^N)\in\mathbb{R}^{Nd}$, we introduce the error of the interacting force for the ith particle.

$$\chi_i(\mathtt{x}):=\frac{1}{p-1}\sum_{j\in\mathcal{C}_{\theta(i)}}K(x^i-x^j)-\frac{1}{N-1}\sum_{j:j\neq i}K(x^i-x^j). \tag{3.12}$$

LEMMA 3.5. *It holds that*

$$\mathbb{E}\chi_i(\mathtt{x}) = 0. \tag{3.13}$$

Moreover, the second moment is given by

$$\mathbb{E}|\chi_i(\mathtt{x})|^2 = \left(\frac{1}{p-1} - \frac{1}{N-1}\right)\Lambda_i(\mathtt{x}), \tag{3.14}$$

where

$$\Lambda_i(\mathtt{x}) := \frac{1}{N-2}\sum_{j:j\neq i}\left|K(x^i - x^j) - \frac{1}{N-1}\sum_{\ell:\ell\neq i}K(x^i - x^\ell)\right|^2. \tag{3.15}$$

LEMMA 3.6. *Fix $i \in \{1,\dots,N\}$. Let $\theta(i)$ be defined as above. Let Y_j $(1 \le j \le N)$ be N random variables (or random vectors) that are independent of $\mathcal{C}$. Then, for $p \ge 2$,*

$$\left\|\frac{1}{p-1}\sum_{j\in\mathcal{C}_{\theta(i)},j\neq i} Y_j\right\| \le \left(\frac{1}{N-1}\sum_{j:j\neq i}\|Y_j\|^2\right)^{1/2}. \tag{3.16}$$

Below, we establish some moment estimates so that we can establish the stability for the RBM and thus prove convergence.

LEMMA 3.7. *Under Assumption* 3.1, *for $q \ge 1$, it holds for some C_q independent of N that*

$$\sup_{t>0}\left(\mathbb{E}(|X^i(t)|^q + |V^i(t)|^q) + \mathbb{E}(|\tilde{X}^i(t)|^q + |\tilde{V}^i(t)|^q)\right) \le C_q. \tag{3.17}$$

In addition, for any $k > 0$ and $q \ge 2$,

$$\sup_{t\in[t_{k-1},t_k)}\left|\mathbb{E}(|\tilde{X}^i(t)|^q + |\tilde{V}^i(t)|^q|\mathcal{F}_{k-1})\right| \le C(1 + |\tilde{X}^i(t_{k-1})|^q + |\tilde{V}^i(t_{k-1})|^q) \tag{3.18}$$

holds almost surely, where C is independent of N.

Proof. Here, we show the moment bounds for $(\tilde{X}^i, \tilde{V}^i)$ only, as the estimates for the moments of (X^i, V^i) are similar (and easier).

Following [44], section 3], we consider the Lyapunov function

$$\ell(\tilde{X}^i, \tilde{V}^i) := \frac{1}{2}(|\tilde{X}^i|^2 + |\tilde{X}^i + \alpha\tilde{V}^i|^2) + \alpha^2 U(\tilde{X}^i).$$

By Assumption 3.1, we can assume without loss of generality that

$$\inf_x U(x) = 0.$$

Due to Assumption 3.1, the second moments of $\tilde{X}^i$ and $\tilde{V}^i$ can be controlled easily by this Lyapunov function.

Then, by Itô's formula, for any $r \ge 1$, and for $t \in [t_{k-1}, t_k)$,

$$\frac{d}{dt}\mathbb{E}\left[[\ell(\tilde{X}^i, \tilde{V}^i)]^r|\mathcal{F}_{k-1}\right] = \mathbb{E}[\mathcal{L}[\ell(\tilde{X}^i, \tilde{V}^i)]^r|\mathcal{F}_{k-1}],$$

where $\mathcal{L}$ is the generator for the SDE (2.1) given by

(3.19)
$$\mathcal{L} = \sum_{i=1}^{N} v^i \cdot \nabla_{x^i} + \sum_{i=1}^{N} \left(-\nabla U(x^i) + \frac{1}{p-1} \sum_{j \in \mathcal{C}_{\theta(i)}^{(k-1)}, j \neq i} K(x^i - x^j) - \gamma v^i \right) \cdot \nabla_{v^i} + \frac{1}{2} \sum_{i=1}^{N} \sigma^2 \Delta_{v_i},$$

where clearly $\mathcal{C}_{\theta(i)}^{(k-1)}$ indicates the batch that particle i belongs to at t_{k-1}. Note that $x^i \in \mathbb{R}^d$ and $v^i \in \mathbb{R}^d$. Direct computation reveals that

$$\mathcal{L}[\ell(x^i, v^i)]^r = r[\ell(x^i, v^i)]^{r-1} \mathcal{L}\ell(x^i, v^i) + \frac{\sigma^2}{2} r(r-1)[\ell(x^i, v^i)]^{r-2} |\nabla_{v^i} \ell|^2.$$

Clearly, $|\nabla_{v^i} \ell|^2 \leq \chi(\alpha) \ell(x^i, v^i)$ for some number χ depending on α. That means

$$\frac{\sigma^2}{2} r(r-1)[\ell(x^i, v^i)]^{r-2} |\nabla_{v^i} \ell|^2 \leq C[\ell(x^i, v^i)]^{r-1}.$$

The power $r-1$ indicates that this term can be controlled without difficulty. Moreover,

$$\begin{aligned}
&\mathcal{L}\ell(x^i, v^i) \\
&\quad = v^i \cdot \left(x^i + (x^i + \alpha v^i) + \alpha^2 \nabla U \right) \\
&\qquad + \left[-\nabla U(x^i) - \gamma v^i + \frac{1}{p-1} \sum_{j \in \mathcal{C}_{\theta(i)}^{(k-1)}, j \neq i} K(x^i - x^j) \right] \cdot \alpha(x^i + \alpha v^i) + \frac{1}{2} \sigma^2 \alpha^2 d \\
&\quad = -\alpha x^i \cdot \nabla U(x^i) + (\alpha - \alpha^2 \gamma) |v^i|^2 + (2 - \alpha\gamma) x^i \cdot v^i \\
&\qquad + \frac{\alpha}{p-1} x^i \cdot \sum_{j \in \mathcal{C}_{\theta(i)}^{(k-1)}, j \neq i} K(x^i - x^j) + \frac{\alpha^2}{p-1} v^i \cdot \sum_{j \in \mathcal{C}_{\theta(i)}^{(k-1)}, j \neq i} K(x^i - x^j) + \frac{1}{2} \sigma^2 \alpha^2 d.
\end{aligned}$$

Taking $\alpha = 2/\gamma$, one finds that

$$\mathcal{L}\ell(x^i, v^i) \leq -\frac{2}{\gamma} \lambda_m |x^i|^2 - \frac{2}{\gamma} |v^i|^2 + C(|x^i| + |v^i| + 1) \leq -\beta \ell(x^i, v^i) + C$$

for some $\beta > 0$. Hence,

$$\frac{d}{dt} \mathbb{E}\left[[\ell(\tilde{X}^i, \tilde{V}^i)]^r | \mathcal{F}_{k-1} \right] \leq -r\beta \mathbb{E}\left[[\ell(\tilde{X}^i, \tilde{V}^i)]^r | \mathcal{F}_{k-1} \right] + C\mathbb{E}\left[[\ell(\tilde{X}^i, \tilde{V}^i)]^{r-1} | \mathcal{F}_{k-1} \right].$$

Using the fact that $U(x^i) \leq C(1 + |x^i|^2)$ (since $\|\nabla^2 U\|_2 \leq \lambda_M$), (3.18) then follows easily.

Similarly, taking full expectation leads to

$$\frac{d}{dt} \mathbb{E}\left[[\ell(\tilde{X}^i, \tilde{V}^i)]^r \right] \leq -r\beta \mathbb{E}\left[[\ell(\tilde{X}^i, \tilde{V}^i)]^r \right] + C\mathbb{E}\left[[\ell(\tilde{X}^i, \tilde{V}^i)]^{r-1} \right].$$

The moment control (3.17) then follows for $q \geq 2$. The moments for $q \in [1, 2)$ are controlled by the $q = 2$ moment by the Hölder inequality. ∎

Next, we introduce some notation for better presentation. First of all, due to the degeneracy of the white noise in the equations for X_i's, the generator of the underdamped Langevin does not have the ellipticity. The proof of the ergodicity for the corresponding Fokker–Planck equation relies on the hypocoercivity [52, 18], and one needs to use the transport term to compensate for the degeneracy. In terms of particle formulation (SDEs), one may consider the following variables to compensate for the degeneracy (see [44], section 3] or [20]):

$$Z^i := \tilde{X}^i - X^i, \quad \hat{Z}^i := \tilde{X}^i - X^i + \alpha(\tilde{V}^i - V^i). \tag{3.20}$$

Moreover, we denote

$$\delta K_{ij}(t) := K(\tilde{X}^i(t) - \tilde{X}^j(t)) - K(X^i(t) - X^j(t)). \tag{3.21}$$

Using this notation, we can conveniently write

$$\begin{aligned} &\frac{1}{p-1}\sum_{j\in\mathcal{C}_{\theta(i)},j\neq i} K(\tilde{X}^i - \tilde{X}^j) - \frac{1}{N-1}\sum_{j\neq i} K(X^i - X^j) \\ &= \frac{1}{N-1}\sum_{j\neq i}\delta K_{ij} + \chi_i(\tilde{\mathfrak{X}}) \\ &= \frac{1}{p-1}\sum_{j\in\mathcal{C}_{\theta(i)},j\neq i}\delta K_{ij} + \chi_i(\mathfrak{X}). \end{aligned} \tag{3.22}$$

By the definition of $\hat{Z}^i$, for $t \in [t_{k-1}, t_k)$,

$$\frac{d}{dt}\hat{Z}^i = (1-\alpha\gamma)(\tilde{V}^i(t) - V^i(t)) - \alpha\Big(\nabla U(\tilde{X}^i) - \nabla U(X^i)\Big) + \frac{1}{p-1}\sum_{j\in\mathcal{C}^{(k-1)}_{\theta(i)},j\neq i}\delta K_{ij} + \chi_i(\mathfrak{X}).$$

Lemma 3.8. *Suppose Assumption* 3.1 *holds. For* $t \in [t_{k-1}, t_k)$,

$$\|Z^i(t) - Z^i(t_{k-1})\| + \|\hat{Z}^i(t) - \hat{Z}^i(t_{k-1})\| \le C\tau. \tag{3.23}$$

Also, almost surely, it holds that for $\tau \le 1$,

$$|Z^i(t)| + |\hat{Z}^i(t)| \le (|Z^i(t_{k-1})| + |\hat{Z}^i(t_{k-1})|)(1 + C\tau) + C\tau. \tag{3.24}$$

Proof. The first part is an easy consequence of the moment control in (3.17). Direct computation shows that for $t \in [t_{k-1}, t_k)$

$$\begin{aligned} &\frac{d}{dt}|Z^i| \le \frac{1}{\alpha}|\hat{Z}^i - Z^i|, \\ &\frac{d}{dt}|\hat{Z}^i| \le \left|\frac{(1-\alpha\gamma)(\hat{Z}^i - Z^i)}{\alpha} + \alpha\Bigg(- \nabla U(\tilde{X}^i) + \nabla U(X^i) \right. \\ &\qquad\qquad \left. + \frac{1}{p-1}\sum_{j\in\mathcal{C}^{(k-1)}_{\theta(i)},j\neq i}\delta K_{ij} + \chi_i(\mathfrak{X})\Bigg)\right|. \end{aligned}$$

Hence, one has

$$\frac{d}{dt}(|Z^i| + |\hat{Z}^i|) \le C(|Z^i| + |\hat{Z}^i|) + C.$$

The claim then follows. ∎

Proof of Theorem 3.3. We remark again that in this proof, C is a generic constant that is independent of time t and particle number N, and the concrete meaning can change from line to line.

Note that we aim to estimate how the quantity evolves:

$$u(t) := \frac{1}{N}\sum_{i=1}^{N}(\mathbb{E}|\tilde{X}^i - X^i|^2 + \mathbb{E}|V^i - \tilde{V}^i|^2) = \mathbb{E}|\tilde{X}^1 - X^1|^2 + \mathbb{E}|\tilde{V}^1 - V^1|^2. \tag{3.25}$$

Direct estimation of this quantity is not easy. As mentioned above already, we consider the following motivated by [20]:

$$J(t) := \frac{1}{2}(\mathbb{E}|Z^1|^2 + \mathbb{E}|\hat{Z}^1|^2) = \frac{1}{2}\left[\mathbb{E}|\tilde{X}^1 - X^1|^2 + \mathbb{E}|\tilde{X}^1 - X^1 + \alpha(\tilde{V}^1 - V^1)|^2\right], \tag{3.26}$$

where α is to be determined later. Then J is equivalent to u but it can be treated more easily, as we shall see below. Below, we consider $t \in [t_{k-1}, t_k)$.

Step 1. Contraction. Recall

$$\frac{1}{p-1}\sum_{j\in\mathcal{C}_{\theta(1)}, j\neq 1} K(\tilde{X}^1 - \tilde{X}^j) - \frac{1}{N-1}\sum_{j\neq 1} K(X^1 - X^j) = \frac{1}{N-1}\sum_{j\neq 1}\delta K_{1j} + \chi_1(\tilde{\mathfrak{X}}),$$

where $\mathcal{C}_{\theta(1)}$ is the random batch that contains particle 1, and define

$$B(\tilde{X}^1, X^1) := \int_0^1 \nabla^2 U(s\tilde{X}^1 + (1-s)X^1)\, ds. \tag{3.27}$$

Direct computation yields

$$\frac{d}{dt}J = \alpha^{-1}\mathbb{E}Z^1\cdot(\hat{Z}^1 - Z^1) + \alpha\mathbb{E}\hat{Z}^1\cdot\bigg(\alpha^{-2}(\hat{Z}^1 - Z^1) - B\cdot Z^1 + \frac{1}{N-1}\sum_{j\neq 1}\delta K_{1j} + \chi_1(\tilde{\mathfrak{X}}) - \frac{\gamma}{\alpha}(\hat{Z}^1 - Z^1)\bigg).$$

By symmetry, $\|Z^j\| = \|Z^1\|$, and thus

$$\mathbb{E}\hat{Z}^1\cdot\frac{1}{N-1}\sum_{j\neq 1}\delta K_{1j} \le L(\|\hat{Z}^1\|\|Z^1\| + \sum_{j\neq 1}\|\hat{Z}^1\|\|Z^j\|) \le 2LJ(t).$$

Then

$$\dot{J} \le -\mathbb{E}\left([Z^1, \hat{Z}^1]^T\begin{bmatrix}\alpha^{-1}I_d & \frac{1}{2}(\alpha B - \gamma I_d)\\ \frac{1}{2}(\alpha B - \gamma I_d) & \gamma - \alpha^{-1}I_d\end{bmatrix}\begin{bmatrix}Z^1\\ \hat{Z}^1\end{bmatrix}\right) + 2\alpha LJ(t) + \alpha\mathbb{E}\hat{Z}^1\cdot\chi_1(\tilde{\mathfrak{X}}). \tag{3.28}$$

Let the eigenvalues of B be $\tilde{\lambda}_i$, which are bounded below and above by λ_m, λ_M, respectively. The eigenvalues of the matrix in (3.28) are given by

$$\mu_{i,\pm} = \frac{1}{2}\left(\gamma \pm \sqrt{\gamma^2 + 4\alpha^{-1}(\alpha^{-1} - \gamma) + (\alpha\tilde{\lambda}_i - \gamma)^2}\right).$$

Choosing $\alpha = 2/\gamma$, all the eigenvalues are $\{\tilde{\lambda}_i/\gamma, \gamma - \tilde{\lambda}_i/\gamma\}_{i=1}^d$. Hence,

$$\dot{J} \le -\frac{2}{\gamma}\left[\min(\lambda_m - 2L, \gamma^2 - (\lambda_M + 2L))\right] J(t) + \frac{2}{\gamma}\mathbb{E}\hat{Z}\cdot\chi_1(\tilde{\mathfrak{X}}).$$

Under Assumption 3.1, the coefficient of $J(t)$ on the right-hand side is negative so that the Langevin dynamics has the contraction property.

Step 2. The local error estimate. We now estimate the local error term $\mathbb{E}\hat{Z}^1\cdot\chi_1(\tilde{\mathfrak{X}}(t))$, which we decompose as

$$\begin{aligned}\mathbb{E}\hat{Z}^1\cdot\chi_1(\tilde{\mathfrak{X}}(t)) &= \mathbb{E}\hat{Z}^1(t_{k-1})\cdot\chi_1(\tilde{\mathfrak{X}}(t)) + \mathbb{E}[\hat{Z}^1(t) - \hat{Z}^1(t_{k-1})]\cdot\chi_1(\tilde{\mathfrak{X}}(t))\\ &=: I_1 + I_2.\end{aligned}$$

Substep 2.1. Estimation of I_1. Using the consistency result in Lemma 3.5,

$$\mathbb{E}\hat{Z}^1(t_{k-1})\cdot\chi_1(\tilde{\mathfrak{X}}(t_{k-1})) = 0,$$

and thus one has

$$\begin{aligned}I_1 &= \mathbb{E}\hat{Z}^1(t_{k-1})\cdot[\chi_1(\tilde{\mathfrak{X}}(t)) - \chi_1(\tilde{\mathfrak{X}}(t_{k-1}))]\\ &= \mathbb{E}\left(\hat{Z}^1(t_{k-1})\cdot\mathbb{E}[\chi_1(\tilde{\mathfrak{X}}(t)) - \chi_1(\tilde{\mathfrak{X}}(t_{k-1}))|\mathcal{F}_{k-1}]\right)\\ &\le \|\hat{Z}^1(t_{k-1})\|\left\|\mathbb{E}[\chi_1(\tilde{\mathfrak{X}}(t)) - \chi_1(\tilde{\mathfrak{X}}(t_{k-1}))|\mathcal{F}_{k-1}]\right\|.\end{aligned}$$

Introducing

$$\begin{aligned}&\delta\tilde{K}^{1j} := K(\tilde{X}^1(t) - \tilde{X}^j(t)) - K(\tilde{X}^1(t_{k-1}) - \tilde{X}^j(t_{k-1})),\\ &\delta\tilde{X}^j = \tilde{X}^j(t) - \tilde{X}^j(t_{k-1}),\end{aligned}$$

one has

$$\begin{aligned}&|\mathbb{E}[\chi_1(\tilde{X}(t)) - \chi_1(\tilde{X}(t_{k-1}))|\mathcal{F}_{k-1}]|\\ &\le \frac{1}{p-1}\sum_{j\in\mathcal{C}_{\theta(1)}^{(k-1)}, j\neq 1}|\mathbb{E}(\delta\tilde{K}^{1j}(t)|\mathcal{F}_{k-1})| + \frac{1}{N-1}\sum_{j\neq 1}|\mathbb{E}(\delta\tilde{K}^{1j}(t)|\mathcal{F}_{k-1})|.\end{aligned} \tag{3.29}$$

Note that $\delta\tilde{K}^{1j}$ is different from δK_{1j} in (3.21).

Now, we estimate

$$\begin{aligned}&\left\|\frac{1}{p-1}\sum_{j\in\mathcal{C}_{\theta(1)}^{(k-1)}, j\neq 1}|\mathbb{E}(\delta\tilde{K}^{1j}(t)|\mathcal{F}_{k-1})|\right\|\\ &\le L\|\mathbb{E}(\delta\tilde{X}^1|\mathcal{F}_{k-1})\| + \frac{L}{p-1}\left\|\sum_{j\in\mathcal{C}_{\theta(1)}^{(k-1)}, j\neq 1}|\mathbb{E}(\delta\tilde{X}^j|\mathcal{F}_{k-1})|\right\|.\end{aligned}$$

Since $\delta\tilde{X}^j = \int_{t_{k-1}}^t \tilde{V}^j(s)\,ds$, we find easily that

$$\begin{aligned}|\mathbb{E}(\delta\tilde{X}^j|\mathcal{F}_{k-1})| &\le \int_{t_{k-1}}^t \mathbb{E}(|\tilde{V}^j(s)||\mathcal{F}_{k-1})\,ds\\ &\le \int_{t_{k-1}}^s \sqrt{\mathbb{E}(|\tilde{V}^j(s)|^2|\mathcal{F}_{k-1})}\,ds\\ &\le C\left(\sqrt{1 + |\tilde{X}^j(t_{k-1})|^2 + |\tilde{V}^j(t_{k-1})|^2}\right)\tau.\end{aligned}$$

Now, since $\sqrt{1+|\tilde{X}^j(t_{k-1})|^2+|\tilde{V}^j(t_{k-1})|^2}$ is independent of $\mathcal{C}^{(k-1)}$, applying Lemma 3.6, one has

$$\left\| \frac{L}{p-1} \sum_{j\in\mathcal{C}^{(k-1)}_{\theta(1)}, j\neq 1} \sqrt{1+|\tilde{X}^j(t_{k-1})|^2+|\tilde{V}^j(t_{k-1})|^2} \right\| \leq C.$$

The other term in (3.29) is similar, but much simpler.

Hence, we find

$$I_1 \leq C\|\hat{Z}^1(t_{k-1})\|\tau \leq C\|\hat{Z}(t)\|\tau + C\tau^2.$$

Substep 2.2. Estimate of I_2. Now, we estimate I_2:

$$\begin{aligned} I_2 &= \mathbb{E}[\hat{Z}^1(t)-\hat{Z}^1(t_{k-1})]\cdot\chi_1(\tilde{\mathfrak{X}}) \\ &= \mathbb{E}[\hat{Z}^1(t)-\hat{Z}^1(t_{k-1})]\cdot\chi_1(\mathfrak{X}) + \mathbb{E}[\hat{Z}^1(t)-\hat{Z}^1(t_{k-1})]\cdot[\chi_1(\tilde{\mathfrak{X}})-\chi_1(\mathfrak{X})] \\ &=: I_{21}+I_{22}. \end{aligned}$$

For I_{21}, we recall the matrix B defined in (3.27), and its spectral radius by Assumption 3.1 is controlled by

$$\rho(B(\tilde{X}^1, X^1)) \leq \lambda_M.$$

Then one has

(3.30)

$$\begin{aligned} \hat{Z}^1(t)-\hat{Z}^1(t_{k-1}) = &\int_{t_{k-1}}^t (\alpha^{-1}-\gamma)(\hat{Z}^1-Z^1)ds - \alpha\int_{t_{k-1}}^t Z^1\cdot B(\tilde{X}^1(s),X^1(s))ds \\ &+ \int_{t_{k-1}}^t \frac{1}{p-1}\sum_{j\in\mathcal{C}^{(k-1)}_{\theta(1)}, j\neq 1} \delta K_{1j}\, ds + \int_{t_{k-1}}^t \chi_1(\mathfrak{X}(s))\, ds. \end{aligned}$$

Noting Lemma 3.8, $\alpha^{-1}-\gamma = -\gamma/2$, and that $\|\chi_i\|_\infty \leq 2\|K\|_\infty$, one has

$$\mathbb{E}\left(\int_{t_{k-1}}^t \frac{1}{2}\gamma(\hat{Z}^1-Z^1)ds - \alpha\int_{t_{k-1}}^t Z^1\cdot B(\tilde{X}^1,X^1)ds\right)\cdot\chi_1(\mathfrak{X}(t)) \leq C\sqrt{J(t)}\tau + C\tau^2,$$

where we used $\mathbb{E}Z^1\cdot B(\tilde{X}^1(s),X^1(s))\cdot\chi_1(\mathfrak{X}(t)) \leq C\lambda_M\|Z^1(s)\| \leq C\|Z^1(t)\| + C\tau$.

For the third term in (3.30) dotted with χ_i, one has

$$\begin{aligned} &\mathbb{E}\left[\left(\int_{t_{k-1}}^t \frac{1}{p-1}\sum_{j\in\mathcal{C}^{(k-1)}_{\theta(1)}, j\neq 1}\delta K_{1j}\, ds\right)\cdot\chi_1(\mathfrak{X}(t))\right] \\ &\qquad \leq \|\chi_1\|_\infty \sup_{s\in[t_{k-1},t_k)}\left\|\frac{1}{p-1}\sum_{j\in\mathcal{C}^{(k-1)}_{\theta(1)}, j\neq 1}\delta K_{1j}(s)\right\|\tau. \end{aligned}$$

Clearly,

$$\left\|\frac{1}{p-1}\sum_{j\in\mathcal{C}^{(k-1)}_{\theta(1)}, j\neq 1}\delta K_{1j}(s)\right\| \leq L\left(\|Z^1(s)\| + \left\|\frac{1}{p-1}\sum_{j\in\mathcal{C}^{(k-1)}_{\theta(1)}, j\neq 1}|Z^j(s)|\right\|\right). \tag{3.31}$$

By Lemma 3.8, one has almost surely that

$$|Z^j(s)| \le |Z^j(t_{k-1})| + C\tau,$$

and thus

$$\left\| \frac{1}{p-1} \sum_{j \in \mathcal{C}^{(k-1)}_{\theta(1)}, j \neq 1} |Z^j(s)| \right\| \le \left\| \frac{1}{p-1} \sum_{j \in \mathcal{C}^{(k-1)}_{\theta(1)}, j \neq 1} |Z^j(t_{k-1})| \right\| + C\tau.$$

Since $\{|Z^j(t_{k-1})|\}$'s are independent of $\mathcal{C}^{(k-1)}$, Lemma 3.6 then gives

$$\begin{aligned} \left\| \frac{1}{p-1} \sum_{j \in \mathcal{C}^{(k-1)}_{\theta(1)}, j \neq 1} |Z^j(t_{k-1})| \right\| &\le \left(\frac{1}{N-1} \sum_{j \neq i} \|Z^j(t_{k-1})\|^2 \right)^{1/2} \\ &= \|Z^1(t_{k-1})\|. \end{aligned}$$

Hence, one in fact has

$$\begin{aligned} \left\| \frac{1}{p-1} \sum_{j \in \mathcal{C}^{(k-1)}_{\theta(1)}, j \neq 1} \delta K_{1j}(s) \right\| &\le 2L\|Z^1(t_{k-1})\| + C\tau \\ &\le 2L\|Z^1(t)\| + C\tau. \end{aligned} \tag{3.32}$$

The fourth term can be easily controlled using Lemma 3.5 so that

$$\mathbb{E} \int_{t_{k-1}}^{t} \chi_1(\mathfrak{X}(s)) \cdot \chi_1(\mathfrak{X}(t))\, ds \le \Big(\frac{1}{p-1} - \frac{1}{N-1} \Big) \|\Lambda_1\|_\infty \tau.$$

Now, we move to the estimate of I_{22} term, which is much easier. In fact,

$$I_{22} = \mathbb{E}[\hat{Z}^1(t) - \hat{Z}^1(t_{k-1})] \cdot [\chi_1(\tilde{\mathfrak{X}}) - \chi_1(\mathfrak{X})] \le C\tau \|\chi_1(\tilde{\mathfrak{X}}) - \chi_1(\mathfrak{X})\|,$$

where $\|\hat{Z}^1(t) - \hat{Z}^1(t_{k-1})\| \le C\tau$ by Lemma 3.8.

Now, by the definition of χ_i,

$$\|\chi_1(\tilde{\mathfrak{X}}) - \chi_1(\mathfrak{X})\| \le \left\| \frac{1}{p-1} \sum_{j \in \mathcal{C}^{(k-1)}_{\theta(1)}, j \neq 1} \delta K_{1j}(s) \right\| + \left\| \frac{1}{N-1} \sum_{j : j \neq 1} \delta K_{1j}(s) \right\|.$$

The first term has been estimated in (3.32). The second term is easily bounded with the same bound as in (3.32). Hence, I_2 is controlled as

$$I_2 \le C\sqrt{J(t)}\tau + C\tau^2 + \frac{1}{p-1}\|\Lambda_1\|_\infty \tau.$$

Step 3. Final error estimate. Combining all the estimates above, one has for any $k \ge 1$ and $t \in [t_{k-1}, t_k)$ (and hence all $t \ge 0$) that

$$\dot{J} \le -\frac{2}{\gamma} \left[\min(\lambda_m - 2L, \gamma^2 - (\lambda_M + 2L)) \right] J(t) + \frac{C}{\gamma} \left(\sqrt{J(t)}\tau + \tau^2 + \frac{1}{p-1}\tau \right). \tag{3.33}$$

Applying Grönwall's inequality then gives the desired result. ∎

4. Applications to molecular dynamics simulations. In this section, we discuss the details of RBM with kernel splitting for one of the important applications of the second order systems with singular kernels: molecular dynamics simulations.

Molecular dynamics (MD) refers to computer simulation of atoms and molecules, and the goals include but are not limited to understanding proteins, large biomolecules, etc.; computing the statistics of the distribution; and investigating the properties of solids and fluids [15, 21]. In this section, we will only consider pairwise interactions. In standard MD simulations, there could be interactions involving three or more bodies like torques. The application of RBM to such interactions should also be possible, as mentioned in [31], and is left for future exploration.

Consider N "molecules" (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:

$$\begin{aligned} dX^i &= V^i\,dt, \\ dV^i &= \left[-\sum_{j:j\neq i} \nabla\phi_{ij}(X^i - X^j)\right] dt + d\xi^i. \end{aligned} \tag{4.1}$$

Here, $\phi_{ij}(\cdot)$ is the interaction potential and $d\xi^i$ means some other possible terms that change the momentum, which we will discuss below. Typical examples include the Coulomb potentials,

$$\phi_{ij}(x) = \frac{q_i q_j}{r}, \tag{4.2}$$

where q_i is the charge for the ith particle and $r = |x|$, and the Lennard–Jones potential

$$\phi_{ij}(x) = \phi(x) = 4\left(\frac{1}{r^{12}} - \frac{1}{r^6}\right). \tag{4.3}$$

In this work, we will consider only the Lennard–Jones potential only so that

$$\phi_{ij}(x) \equiv \phi(x), \quad K(x) = -\nabla\phi(x).$$

The generalization to general cases should be straightforward but notationally heavy.

By modeling the ions or molecules as point particles in simulations, both potentials of the forms (4.2) and (4.3) can exist for ions, while (4.3) may be the main interaction for charge-neutral molecules. We remark, however, that Lennard–Jones interaction intrinsically also arises from the electromagnetic interactions between charges inside the molecules (especially the electron clouds), so they are intrinsically the same type of interaction. Of course, there can be other types of interaction that are not mentioned here.

To model solids or fluids with large volumes, one often uses a box with length L, equipped with the periodic conditions for the simulations.

4.1. Coupling with a heat bath. To model the interaction between molecules with a heat bath, one may consider some thermostats so that the temperature of the system can be controlled at a given value. Typical thermostats include the Andersen thermostat, the Langevin thermostat, and the Nosé–Hoover thermostat [21].

In the Andersen thermostat [21, section 6.1.1], one does the simulation for

$$d\xi^i = 0,$$

but a particle can collide with the heat bath each time. In particular, assume the collision frequency is ν, so in a duration of time $t \ll 1$ the chance that a collision has happened is given by the exponential distribution

$$1 - \exp(-\nu t) \approx \nu t, \quad t \ll 1. \tag{4.4}$$

If a collision happens, the new velocity is then sampled from the Maxwellian distribution with temperature β^{-1} (i.e., the normal distribution $\mathcal{N}(0, \beta^{-1})$).

Since the potential $\phi_{ij}(x) \equiv \phi(x)$ is often singular at $x = 0$, we need to split the interaction kernel (or the potential) for the simulation and apply RBM for the long-range but smooth part. Hence, Algorithm 2.2 can be applied when evolving the dynamics. To discretize the equations in Algorithm 2.2, one may in principle use a smaller time step Δt than τ. However, since the random batch system has already introduced a strong error of the magnitude $\sqrt{\tau}$, it seems that there is no need to use smaller time steps for the simulation. Of course, if one regards the random batch system as a new model and discretizes it, one may consider using smaller time steps. Below and in section 5, we will always use $\Delta t = \tau$ as the time step for discretization.

Here is some subtlety. If we discretize (2.6) using some second order integration methods like the Verlet method [21, section 4.3.1], we need to evaluate the forces at t_k^- and t_k^+. The force at t_k^- corresponds to the batches for $[t_{k-1}, t_k)$, while the force at t_k^+ corresponds to batches for $[t_k, t_{k+1})$. This is not quite efficient, as one needs to evaluate the force at t_k twice, so for practical implementation, we evaluate the forces only at t_k^+. Then we introduce a variant of the Verlet scheme, where the velocity is updated using

$$V_{k+1}^i = V_k^i + \frac{1}{2}\left[F_{k^+}^i + F_{(k+1)^+}^i\right]\tau. \tag{4.5}$$

In fact, it is known that the Verlet scheme is equivalent to the leapfrog scheme ((4.3.1)–(4.3.2) in [21, section 4.3.1]) and (4.5) corresponds to applying RBM to the leapfrog scheme. We will call the corresponding algorithm "Andersen-RBM," which is shown in Algorithm 4.1. Note that any reasonable separation that makes K_1 short-ranged and K_2 regular should be fine for Algorithm 4.1.

In the underdamped Langevin dynamics, one chooses

$$d\xi^i = -\gamma V^i\, dt + \sqrt{\frac{2\gamma}{\beta}}\, dW^i,$$

so that the "fluctuation-dissipation relation" is satisfied and the system will evolve to the equilibrium with the correct temperature. It is well known that the invariant measure of such systems is given by the Gibbs distribution

$$\pi(\mathtt{x}, \mathtt{v}) \propto \exp\left(-\beta\left(\frac{1}{2}\sum_{i=1}^N |v^i|^2 + E(\mathtt{x})\right)\right),$$

Algorithm 4.1 Andersen-RBM.

1: Split $K =: K_1 + K_2$: K_1 has short range, while K_2 has long range but is regular. Take a temperature β^{-1}.
2: Sample X^i, V^i for all i.
3: Obtain a set of random batches. For each i, find the batch $\mathcal{C}_{\theta(i)}$ where i lives, and compute:

$$F^i = \sum_{j:j\neq i} K_1(X^i - X^j) + \frac{N-1}{p-1} \sum_{j\in\mathcal{C}_{\theta(i)}, j\neq i} K_2(X^i - X^j). \tag{4.6}$$

4: **for** $k = 1, 2, \cdots$ **do**
5: For each i, generate $\zeta_i \sim \mathcal{U}[0,1]$, the uniform distribution on $[0,1]$. If $\zeta_i \leq 1 - \exp(-\nu\tau)$, sample $V_i \sim \mathcal{N}(0, \beta^{-1})$.
6: Update the positions:

$$X^i \leftarrow X^i + V^i\tau + \frac{1}{2}F^i\tau^2. \tag{4.7}$$

7: Set $F_o^i \leftarrow F^i$ for all i.
8: Obtain a new set of random batches, and compute the forces for all particles $i \in \{1, \cdots, N\}$ as in (4.6).
9: Update the velocities for all i:

$$V^i \leftarrow V^i + \frac{1}{2}[F_o^i + F^i]\tau.$$

10: **end for**

where $\mathtt{x} = (x^1, \ldots, x^N) \in \mathbb{R}^{Nd}$ and $\mathtt{v} = (v^1, \ldots, v^N) \in \mathbb{R}^{Nd}$. Algorithm 2.2 can be applied directly for Langevin dynamics, and one possible way to discretize (2.6) is the "BAOAB" splitting scheme proposed in [39, 38]:

$$\begin{aligned}
V_{k+\frac{1}{2}} &= V_k + \frac{1}{2}F_k\tau, \quad (B)\\
X_{k+\frac{1}{2}} &= X_k + \frac{1}{2}V_{k+\frac{1}{2}}\tau, \quad (A)\\
\hat{V}_{k+\frac{1}{2}} &= c_1 V_{k+\frac{1}{2}} + c_2 R_{k+1}, \quad (O)\\
X_{k+1} &= X_{k+\frac{1}{2}} + \frac{1}{2}\hat{V}_{k+\frac{1}{2}}\tau, \quad (A)\\
V_{k+1} &= \hat{V}_{k+\frac{1}{2}} + \frac{1}{2}F_{k+1}\tau, \quad (B)
\end{aligned} \tag{4.8}$$

where F_k means the force computed at time $t_k = k\tau$ and $\{R_{k+1}\}_{k\geq 0}$ are independent standard normal random vectors. The coefficients $c_1 = e^{-\gamma\tau}$, $c_2 = \sqrt{(1-c_1^2)/\beta}$, where γ is the friction coefficient. Clearly, for second order schemes like this BAOAB splitting scheme, the forces F_k should be computed by applying RBM on the K_2 part. Again, there can be two possible forces at t_k depending on whether the batches for $[t_{k-1}, t_k)$ or the batches for $[t_k, t_{k+1})$ are being used. Similarly as in Algorithm 4.1, we use the forces at t_k^+. The resulting algorithm is similar to Algorithm 4.1 so we omit it. The resulting discretized scheme will be called "Langevin-RBM," which does not

correspond to the discretization of (2.6) directly, but we believe there is no significant difference.

In this work, we adopt the Andersen or the underdamped Langevin thermostat in simulations later in section 5. We remark that RBM can potentially increase temperature of the system by $\mathbb{E}|\chi(\mathbf{x})|^2\tau/(2\gamma) \sim \tau/(p\gamma)$ due to the variance in the force computation. This numerical heating is not good if one wants to obtain some accurate results. Hence, a thermostat that can achieve better temperature control may be more desirable. Larger friction coefficients or smaller time steps could reduce this numerical heating effect. The Nosé–Hoover thermostat [21, section 6.1.2] sometimes can lead to better temperature control in some systems. In this work, we choose not to apply such sophisticated thermostats; the Andersen and Langevin thermostats can already yield acceptable results, as seen in section 5.

4.2. Discussion. Below, we first discuss the benefits of RBM in the molecular regime and the choice of the splitting $K = K_1 + K_2$. Suppose that r_0 is the effective range of K_1 (i.e., when $|x| \gg r_0$, the effects of K_1 can be neglected). Hence, to enjoy the benefits of RBM, we may desire to choose r_0 so that there are only $\mathcal{O}(1)$ particles in the ball $B(X^i, r_0)$ centered at a typical particle X^i.

- For kernels whose range covers effectively only a few particles (like Lennard–Jones fluid with low density), one can pick r_0 large enough. In this case, by the fast decay of the potential, $\sum_{j:j\neq i} K_2(X^i - X^j)$ is negligible. Using the random approximation $\frac{N-1}{p-1}\sum_{j\in\mathcal{C}_{\theta(i)},j\neq i} K_2(X^i - X^j)$ is not quite necessary. For short-range potentials with $\mathcal{O}(1)$ density, though one may make use of the rapid decay of the potential to make the full simulation cheaper, RBM with splitting can still speed up the simulation for such cases since the batch size p can be smaller than the effective number of neighbors (see Remark 5.1 for Lennard–Jones potentials).
- If the range of K is comparable to the size of the simulation domain (like the Lennard–Jones fluid with period box length $L = 1$ and long-range interactions like the Coulomb potentials) or the density is not low, then each particle can feel the effects from a significant number of other particles. In this case, we pick r_0 small so that $B(X^i, r_0)$ contains $\mathcal{O}(1)$ particles. RBM can speed up computation per iteration. Moreover, we also expect that the time step needed for RBM will be comparable to the time step for the full simulation. RBM will indeed save computational cost for these cases.
- Compared with the mean-field regime where RBM is asymptotic-preserving and the variance is controlled uniformly in N, the variance scales like $\mathcal{O}(N^2)$ in the molecular regime. In fact, the factor $(N-1)/(p-1)$ could make the random variable $\frac{N-1}{p-1}\sum_{j\in\mathcal{C}_{\theta(i)},j\neq i} K_2(X^i - X^j)$ differ a lot in magnitude if K_2 changes a lot in magnitude. For example, if one applies RBM to Lennard–Jones fluid with high density where one chooses r_0 very small, then $|K_2|$ changes from a large value to a small value from r_0 to $L/2$. Then, applying RBM with small batch size like $p = 2$ could result in noticeable effects like the numerical heating in molecular dynamics simulations. One has to take some actions like increasing p, decreasing τ, or other advanced techniques to reduce such effects (see section 5.2).
- The fast multipole method [48] is among the popular methods that can also compute the long-range interactions efficiently in $\mathcal{O}(N)$ scaling. However, since the implementation needs to keep the data structures, the prefactor in the linear scaling can be large and the implementation is nontrivial. RBM

with small batch sizes can have a small prefactor in the linear scaling and the implementation is easy. Hence, if the accuracy desired is not so high, RBMs can be preferable due to the efficiency.

Since RBM is asymptotic-preserving in the mean-field limit and the variance is controlled uniformly in N with a factor $1/(N-1)$ in the interaction term, one may be curious whether we should do scalings to have such a factor in the molecular dynamics simulation using RBM. For example, if the interacting forces are homogeneous (like the Coulomb interaction) in space, one may also zoom out in both time and space so that the prefactor $1/(N-1)$ could appear. Another direct way to obtain the $1/(N-1)$ factor is to do the time scaling $\tilde{t} = (N-1)t$, which corresponds to zooming in time. We remark that there is no intrinsic change in the physics due to the scaling, so applying RBM in the original regime (like a molecular regime) does the same thing. Scaling, however, changes the step sizes allowed. For example, if one does the time scaling $\tilde{t} = (N-1)t$, one takes time steps of order $\mathcal{O}(1)$ for the new time variable $\tilde{t}$ while one takes time steps of order $\mathcal{O}(1/N)$ in the original molecular regime. This small step size restriction is not due to RBM. In fact, for the full simulation, the step size also has to be small due to the summation of $N-1$ terms. Since there is no intrinsic difference by scaling, we will apply RBM directly in the molecular regime when it has benefits, as discussed.

5. Numerical experiments. In this section, we perform some numerical experiments to verify the theoretic claims in section 3 and validate RBM with kernel splitting (in particular, Andersen-RBM and Langevin-RBM) via the molecular dynamics simulations for Lennard–Jones fluids. In all the simulations in this section, the time step for discretization is the same as the time step in the RBM, i.e., $\Delta t = \tau$. All the numerical experiments in this section are performed via MATLAB R2020a on a Mac Pro laptop with Intel i5-6360U CPU @ 2 GHz and 8 GB memory.

5.1. A simple illustrative example. First, we consider an underdamped Langevin equation for $(X^i, V^i) \in \mathbb{R} \times \mathbb{R}$. This example is mainly designed to verify that Algorithm 2.1 works for regular kernels and confirm the theoretical results in section 3. In particular, we consider the following interacting particle system on $\mathbb{R}$ for $i = 1, \dots, N$:

$$\begin{aligned} dX^i &= V^i\, dt, \\ dV^i &= -\lambda X^i dt + \frac{1}{N-1} \sum_{j:j\neq i} \frac{X^i - X^j}{1+|X^i - X^j|^2} dt - \gamma V^i dt + \sqrt{2\gamma/\beta}\, dW^i. \end{aligned} \tag{5.1}$$

The kernel

$$K(x) = \frac{x}{1+|x|^2}$$

satisfies $|K| \le \frac{1}{2}$ and $|K'| \le 1$.

Below in the simulations we take

$$\lambda = \gamma = 2.5$$

so that the conditions in Theorem 3.3 hold, and the temperature is taken as $\beta^{-1} = 1$. The discretization will be the BAOAB scheme (4.8). The initial positions X_0^i's are sampled i.i.d. from $\mathcal{U}[-0.5, 0.5]$ (the uniform distribution on $[-0.5, 0.5]$), while the initial velocities are also sampled from $\mathcal{U}[-0.5, 0.5]$ but with the empirical mean

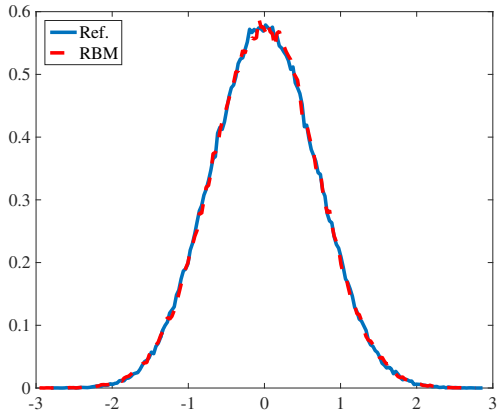

FIG. 1. *The equilibrium density distribution by RBM (red dashed line) for* (5.1) *with* $N = 500$, $p = 2$, *and* $\tau = 0.02$. *The blue solid line is the reference distribution by the full simulation without RBM with step size* $\tau = 10^{-3}$.

subtracted $V_0^i \leftarrow V_0^i - \bar{V}_0$ and then the magnitude rescaled such that the average of $(V_0^i)^2$ is the temperature (i.e., $N^{-1}\sum_i (V_0^i)^2 = \beta^{-1}$). For RBM simulations in this example, we always take batch size

$$p = 2.$$

To verify the effectiveness of RBM, we first do the simulation for $N = 500$ particles and check the computed equilibrium distribution. The system after time $t = 50$ is regarded to be in the equilibrium. Hence, we collect the $\{X^i\}$'s from many iterations after $t = 50$ as the samples. In Figure 1, we show the results by RBM where the Langevin equations are discretized by the BAOAB with step size $\tau = 0.02$. We collect the $N = 500$ particles as some samples with a time gap 0.5 (or 25 iterations for $\tau = 0.02$) up to $t = 300$. Hence, there are $500 * (300 - 50)/0.5 = 2.5 \times 10^5$ sample points to reduce the random fluctuation in Monte Carlo approximations. The reference distribution is plotted using samples at the same time points in the full simulation (i.e., running the Langevin dynamics (5.1) using BAOAB scheme without RBM) with a step size $\tilde{\tau} = 0.001$. Clearly, the equilibrium distribution density is recovered by RBM with good accuracy.

To confirm the sampling correctness quantitatively, we compute the relative weak errors:

$$\mathrm{err}_w = \left| \frac{\sum_{i=1}^{N_s} f(X^i)}{N_s} - \frac{\sum_{i=1}^{\bar{N}_s} f(\hat{X}^i)}{\bar{N}_s} \right| \Big/ \frac{\sum_{i=1}^{\bar{N}_s} f(\hat{X}^i)}{\bar{N}_s}$$

for various test functions f, where $\hat{X}^i$'s are the corresponding computed reference solutions. Here, N_s and $\bar{N}_s$ are the numbers of samples for RBM and reference, respectively. In particular, we again run the simulation for $N = 500$ with step sizes taken as $\tau = 1, 2^{-1}, \ldots, 2^{-3}$. The samples are again taken with a time gap 0.5 from $t = 50$ to $t = 300$. The samples for the reference are computed using the full simulation with the BAOAB scheme and step size $\tilde{\tau} = 2^{-10}$. The results are listed in Table 1, where we take $f(x) = e^{2x}, x^2, 1/((x - 0.1)^2 + 0.001), 1/(1 + x^2)$. Clearly, the weak error in fact tends to zero as we decrease τ, which means RBM indeed can recover the

TABLE 1
The weak errors using RBM for equilibrium distribution of (5.1) *with* $N = 500$.

	$\tau = 1$	$\tau = 2^{-1}$	$\tau = 2^{-2}$	$\tau = 2^{-3}$
e^{2x}	0.1098	0.0785	0.0337	9.229×10^{-5}
x^2	0.0256	0.0252	0.0051	2.7128×10^{-4}
$\frac{1}{(x-0.1)^2+0.001}$	0.0046	0.0016	0.0152	0.0014
$\frac{1}{1+x^2}$	0.0045	0.0049	6.5742×10^{-4}	0.0016

equilibrium distribution. Due to the Monte Carlo fluctuation, the weak convergence order (which should be order 1 or the weak error scales like $\mathcal{O}(\tau)$ motivated by the results in [30]) is not quite evident in Table 1 (only $f(x) = e^{2x}$ shows first order convergence, while other test functions already give very small errors even for $\tau = 1$).

The weak error results above indicate that the equilibrium can be correctly captured by RBM as well, consistent with the claim in the theorem that the error control is uniform in time.

To verify the strong convergence order claimed in Theorem 3.3, we consider the relative strong errors:

$$\text{err}_s = \sqrt{\frac{\sum_{i=1}^N (X^i - \hat{X}^i)^2}{N} \Big/ \frac{\sum_{i=1}^N (\hat{X}^i)^2}{N}}, \tag{5.2}$$

where X^i's are numerical solutions by RBM at $T = 2$ and $\hat{X}^i$'s are the reference solutions. The results are shown in Figure 2. Figure 2 (a) compares RBM with full simulation using the BAOAB scheme for $N = 2000$, while Figure 2 (b) compares the convergence results of RBM for sizes $N = 100, 500, 2000$. The reference solution is obtained using the full batch (the original particle system) using BAOAB with step size $\tau = 2^{-10}$. The same Brownian motions are used for the reference solution and RBM solution (i.e., the realizations of Brownian motions used for the reference solution are stored and then applied for the RBM simulation) for the strong solution. The simulation results above indicate that RBM can indeed obtain the 1/2 strong order for the underdamped Langevin equations with regular kernels, agreeing with our theory in Theorem 3.3.

Lastly, let us take a look at the efficiency. Figure 3 shows the CPU costs needed by RBM compared to the full simulation for system (5.1) with $N = 2000$ at time $T = 2$, both implemented using the BAOAB scheme. Figure 3 (a) clearly shows that the cost of RBM scales like $\mathcal{O}(N)$, while the full simulation scales like $\mathcal{O}(N^2)$ as the system size increases. Note that for small sizes like $N \lesssim 500$, such scalings are not evident due to the matrix operations in MATLAB. Figure 3 (b) plots the strong error defined in (5.2) against the CPU cost. Clearly, RBM has better efficiency for the regime we consider here.

5.2. The Lennard–Jones fluid. In this section, we test RBM with splitting, especially the methods discussed in section 4, on the Lennard–Jones fluids to validate these methods. In fact, our experience indicates that applying Algorithm 2.1 directly to such systems truly yields numerical instability, so RBM with splitting is desired for such applications. The potential ϕ in (4.1) is then given by ($x \in \mathbb{R}^3$) in this setting:

$$\phi(x) = 4\left(\frac{1}{r^{12}} - \frac{1}{r^6}\right), \quad r = |x|. \tag{5.3}$$

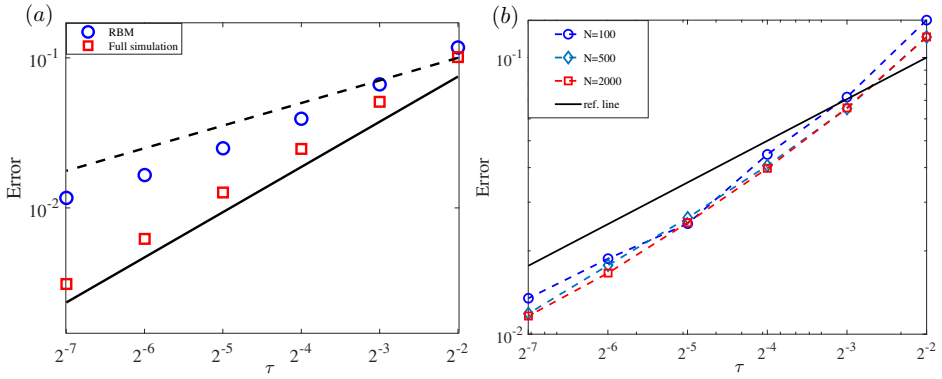

FIG. 2. *The strong errors* err_s *at time* $T = 2$ *vs. time step size* τ *of RBM for* (5.1). (a) *Comparison between RBM and full simulation for* $N = 2000$. *The black solid line is* $E = 0.3\tau$, *while the black dashed line is* $E = 0.2\tau^{1/2}$ *for reference.* (b) *RBM with different system sizes. The black solid line is* $E = 0.2\tau^{1/2}$ *for reference. The convergence order of RBM for strong error is* 1/2.

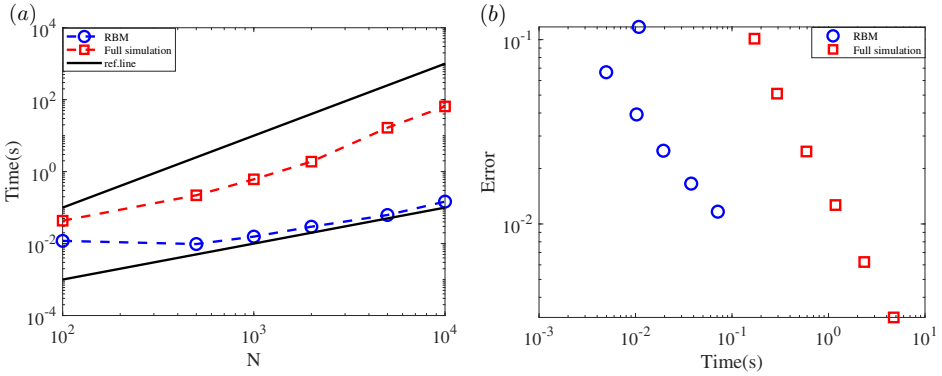

FIG. 3. *Comparison of the CPU costs between RBM and full simulation for* (5.1) *for solutions at* $T = 2$. (a) *The CPU cost versus system size. The solid lines are the curves* $\mathrm{Time} = 10^{-5}N^2$ *and* $\mathrm{Time} = 10^{-4}N$, *respectively.* (b) *The strong error versus CPU cost for* $N = 2000$.

As mentioned in section 4, the periodic boxes are used to approximate the fluids of large extent. Let L be the length of the box. With the periodic setting, a particle interacts with not only another particle but also its periodic images. Thanks to the fast decaying properties of the Lennard–Jones potential, one can pick a cutoff length r_c so that the interaction between two particles (including particle-image interaction) with distance larger than r_c will be treated in a mean-field fashion (see [21, Chapter 3] for more details). Following [21, Chapter 3], we choose

$$r_c = L/2.$$

With the cutoff mentioned here, the pressure formula is given approximately by (see [21, section 3.4], [42, section 4.2])

$$P = \frac{\rho}{\beta} + \frac{8}{V}\sum_{i=1}^{N}\sum_{j:j>i,\tilde{r}_{ij}<r_c}(2\tilde{r}_{ij}^{-12} - \tilde{r}_{ij}^{-6}) + \frac{16}{3}\pi\rho^2\left[\frac{2}{3}\left(\frac{1}{r_c}\right)^9 - \left(\frac{1}{r_c}\right)^3\right], \tag{5.4}$$

where β^{-1} is the scaled temperature, $V = L^3$ is the volume, and we have introduced

$$\tilde{r}_{ij} = |\vec{r}_{ij} + \vec{n}L|, \quad \vec{r}_{ij} = x_i - x_j$$

for some suitable three-dimensional integer vector $\vec{n}$ so that $|\vec{r}_{ij} + \vec{n}L|$ is minimized. Since $r_c = L/2$, there is at most one image of particle j (including itself) that falls into $B(x_i, r_c)$. Hence, when implementing the methods, the forces between particles are computed using the nearest image (see [21, section 4.2.2]).

To apply the methods in section 4, in all the simulations below, we take $r_0 = \sqrt[6]{2}$, where ϕ has a minimum (the force is repulsion if $r < r_0$, while the force is attraction for $r > r_0$) and split the potential ϕ:

$$\phi(x) =: \phi_1(x) + \phi_2(x), \tag{5.5}$$

where

$$\phi_1(x) = \begin{cases} 4\left(\frac{1}{r^{12}} - \frac{1}{r^6}\right) + 1, & 0 < r < \sqrt[6]{2}, \\ 0, & r \geq \sqrt[6]{2}, \end{cases} \tag{5.6}$$

and

$$\phi_2(x) = \begin{cases} -1, & 0 < r < \sqrt[6]{2}, \\ 4(\frac{1}{r^{12}} - \frac{1}{r^6}), & r \geq \sqrt[6]{2}. \end{cases} \tag{5.7}$$

The force $K = -\nabla\phi$ is split correspondingly. That means the part of interaction force for $r \leq \sqrt[6]{2}$ is regarded to have short range and the part for $r > r_0 = \sqrt[6]{2}$ is regarded to have long range. The long-range parts $(-\nabla\phi_2)$ will be computed using RBM. Note that the threshold $r_0 = \sqrt[6]{2}$ is different from the cutoff $r_c = L/2$ above. The cutoff r_c above means that the molecular interactions are computed explicitly only for $r \leq r_c$, while the ones for $r \geq r_c$ are treated in a mean-field fashion.

Remark 5.1. With the cutoff $r_c = L/2$, direct simulation has a complexity $\mathcal{O}(N^2)$ per time step. Since the Lennard–Jones potential decays fast and the density considered in this section is $\mathcal{O}(1)$, one may use smaller cutoff r_c to reduce the complexity to roughly linear and obtain roughly correct results. However, our experience indicates that applying RBM with $r_0 = \sqrt[6]{2}$ still saves time as the batch size p (which is 2 in the experiments below) is much less than the effective number of particles that interact with one particular particle.

For the simulation, the temperature is taken to be $\beta^{-1} = 2$ and the length of the box is set as

$$L = (N/\rho)^{1/3}$$

for a given density ρ. The particles are initially put on the cubic lattice with grid size $L/N^{1/3}$. The initial velocities are randomly chosen from uniform distribution $\mathcal{U}^3[-0.5, 0.5]$, and then shifted and rescaled so that the instantaneous temperature matches the desired value (i.e., $N^{-1}\sum_i |V_0^i|^2 = 3\beta^{-1}$).

For the thermostats, we use both the Andersen thermostat and the underdamped Langevin dynamics for simplicity, as indicated in section 4.1, and the resulting schemes are Andersen-RBM and Langevin-RBM, as already explained in section 4.1. The batch size is taken as $p = 2$ for all the experiments here.

We first run the simulations with the collision coefficient $\nu = 10$ for Andersen-RBM and the friction coefficient $\gamma = 10$ for Langevin-RBM. The simulation before time $T = 50$ is regarded as the burn-in phase, and we compute the pressure using the viral formulation (5.4) at a given time point (after $T = 50$). We compute 10^5 such pressures (each for one iteration) and then take the average as the computed value. The computed values using $N = 100$ and $N = 500$ are shown in Figure 4 for various densities. The reference curve (black solid line) is the fitting curves in [34]. The results show that RBM with splitting strategy (5.5)–(5.6) can work reasonably well for the Lennard–Jones fluid in the considered regime. Another observation from Figure 4 is that when $N = 500$ the extra variance brought by RBM can result in noticeable numerical heating and thus greater pressure (the variance depends on N in the molecular regime; see the discussion in section 4.2). Moreover, our experience indicates that direct application of RBM without splitting (Algorithm 2.1) to the Lennard–Jones potential (5.3) indeed results in numerical instability.

To reduce the numerical heating as shown in Figure 4 (b) or increase the temperature control ability, we try two strategies. The first strategy is to decrease the step size as the iteration goes on to decrease the numerical heating since the effective temperature rise is like $\mathbb{E}|\chi(\mathbf{x})|^2 * \tau/(2\gamma)$. This idea is similar to the one in simulated annealing [50, 29]. The second strategy is to increase the friction coefficient (i.e., the ν and γ in the Andersen thermostat and Langevin dynamics, respectively). Increasing the collision or friction coefficient clearly makes the system relax faster to the quasi-equilibrium, but it may also bring in some unphysical effects [27, 21]. We show the numerical results in Figure 5. Figure 5 (a) shows the results using the first strategy (i.e., decreasing step size), where we take

$$\tau_k = 0.001/\log(k+1).$$

The reason to take such τ_k is that we do not desire the step size to decrease too quickly. Figure 5 (b) shows the results using the second strategy (i.e., using a larger friction coefficient so that the temperature control is better) where we take $\gamma = \nu = 50$. Clearly, after these two approaches are applied, the numerical heating effects

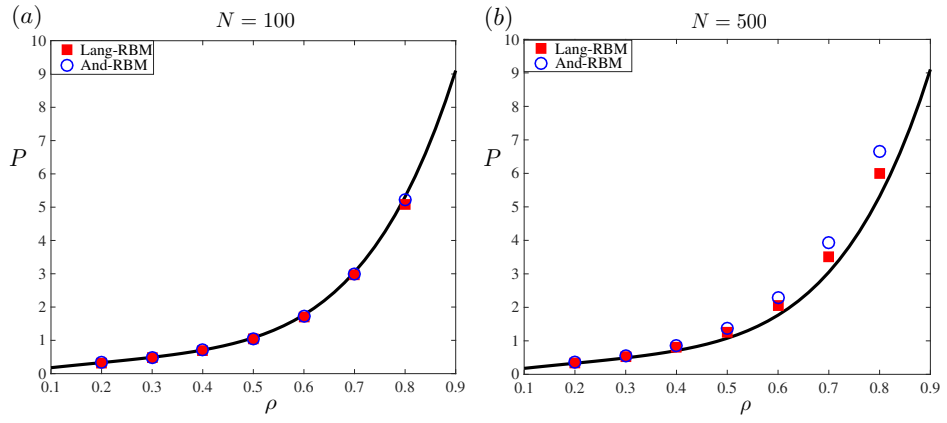

FIG. 4. *The pressure computed by Andersen-RBM and Langevin-RBM for Lennard–Jones fluid. The black solid line is the reference fitting curve in* [34]. *The blue circles are the pressure computed by Andersen-RBM, and the red squares denote the pressure computed by Langevin-RBM.* $\nu = 10$*;* $\gamma = 10$*; the time is* $T = 50$*; the step size is* $\tau = 0.001$.

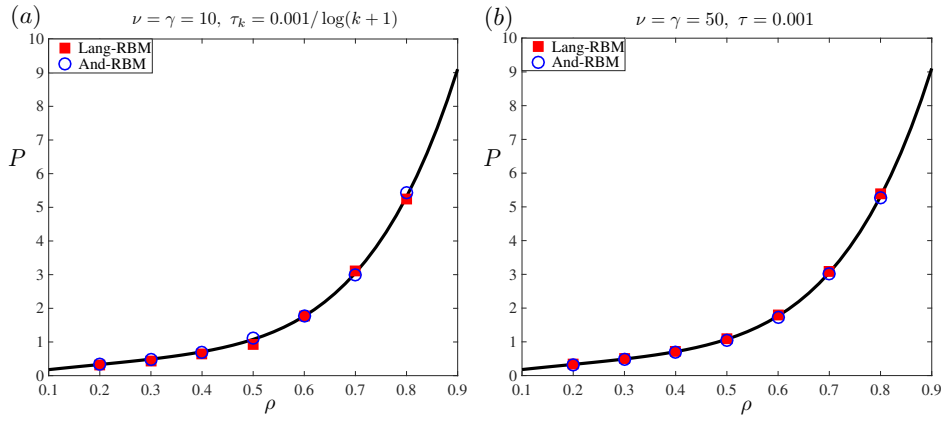

FIG. 5. *The pressure obtained by Andersen-RBM and Langevin-RBM using two strategies to reduce numerical heating for Lennard–Jones fluid with* $N = 500$. *The blue circles are those by Andersen-RBM, while the red squares are by Langevin-RBM.*

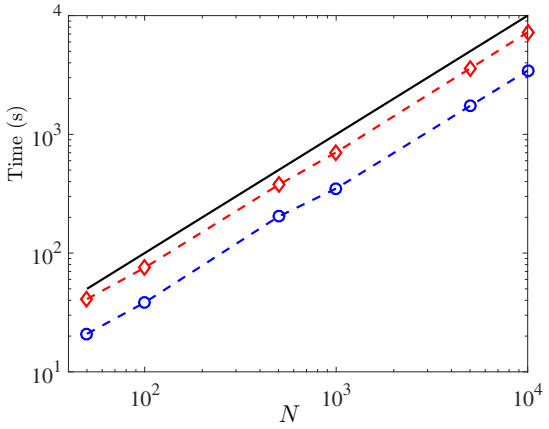

FIG. 6. *The CPU time vs. size of the system for Andersen-RBM (blue circles) and Langevin-RBM (red diamonds). Clearly, the computational time scales linearly with the size of the system for both methods*

are reduced significantly, and the correct equation of state is obtained. As another possible thermostat for better temperature control, one may consider the Nosé–Hoover thermostat [21, Chapter 6].

Lastly, we validate the claim that the complexity of our algorithm is $\mathcal{O}(N)$ in Figure 6, where the CPU time is plotted versus the size of the Lennard–Jones system. The simulation is performed up to time 30 with step size $\tau = 2^{-10}$ for systems with density $\rho = 0.5$. Clearly, both Andersen-RBM and Langevin-RBM scale linearly with the size of the system, and this result thus verified our claim.

Acknowledgments. L.L. would like to thank Prof. Xiantao Li for suggestions on Lennard–Jones fluid simulations. The authors would like to thank the anonymous referees for helpful suggestions.

REFERENCES

[1] G. Albi and L. Pareschi, *Binary interaction algorithms for the simulation of flocking and swarming dynamics*, Multiscale Model. Simul., 11 (2013), pp. 1–29, https://doi.org/10.1137/120868748.

[2] H. Babovsky and R. Illner, *A convergence proof for Nanbu's simulation method for the full Boltzmann equation*, SIAM J. Numer. Anal., 26 (1989), pp. 45–65, https://doi.org/10.1137/0726004.

[3] A. L. Bertozzi, J. B. Garnett, and T. Laurent, *Characterization of radially symmetric finite time blowup in multidimensional aggregation equations*, SIAM J. Math. Anal., 44 (2012), pp. 651–681, https://doi.org/10.1137/11081986X.

[4] U. Biccari and E. Zuazua, *A Stochastic Approach to the Synchronization of Coupled Oscillators*, preprint, https://arxiv.org/abs/2002.04472, 2020.

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

[6] C. K. Birdsall and A. B. Langdon, *Plasma Physics via Computer Simulation*, CRC Press, 2004.

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

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

[9] H. B. Callen and T. A. Welton, *Irreversibility and generalized noise*, Phys. Rev., 83 (1951), pp. 34.

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

[11] J. A. Carrillo, L. Pareschi and M. Zanella, *Particle based gPC methods for mean-field models of swarming with uncertainty*, Commun. Comput. Phys., 25 (2019

[12] P. Cattiaux, A. Guillin and F. Malrieu, *Probabilistic approach for granular media equations in the non-uniformly convex case*, Probab. Theory Relat. Fields, 140 (2008), pp. 19–40.

[13] T. Chen, E. Fox and C. Guestrin. *Stochastic gradient Hamiltonian Monte Carlo*. In: International Conference on Machine Learning, 2014, pp. 1683–1691.

[14] Y.-P. Choi, S.-Y. Ha and S.-B. Yun, *Complete synchronization of Kuramoto oscillators with finite inertia*, Phys. D Nonlinear Phenom., 240 (2011), pp. 32–44.

[15] 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.

[16] F. Cucker and S. Smale, *Emergent behavior in flocks*, IEEE Trans. Automat. Control, 52 (2007), pp. 852–862.

[17] P. Degond, J.-G. Liu and R. L. Pego, *Coagulation-fragmentation model for animal group-size statistics*, J. Nonlinear Sci., 27 (2017), pp. 379–424.

[18] J. Dolbeault, C. Mouhot and C. Schmeiser, *Hypocoercivity for linear kinetic equations conserving mass*, Trans. Am. Math. Soc., 367 (2015), pp. 3807–3828.

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

[20] A. Eberle, A. Guillin and R. Zimmer, *Couplings and quantitative contraction rates for Langevin dynamics*, Ann. Probab., 47 (2019), pp. 1982–2010.

[21] D. Frenkel and B. Smit, *Understanding Molecular Simulation: From Algorithms to Applications*, vol. 1. Elsevier, 2001.

[22] A. Georges, G. Kotliar, W. Krauth and M. J. Rozenberg, *Dynamical mean-field theory of strongly correlated fermion systems and the limit of infinite dimensions*, Rev. Mod. Phys., 68 (1996), pp. 13.

[23] F. Golse, S. Jin and T. Paul, *The random batch method for n-body quantum dynamics*. arXiv preprint arXiv:1912.07424, 2019.

[24] S.-Y. Ha and Z. Li, *Complete synchronization of Kuramoto oscillators with hierarchical leadership*, Commun. Math. Sci., 12 (2014), pp. 485–508.

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

[26] B. Hetenyi, K. Bernacki and B. J. Berne, *Multiple "time step" Monte Carlo*, J. Chem. Phys., 117 (2002), pp. 8203–8207.

[27] W. G. Hoover, *Canonical dynamics: equilibrium phase-space distributions*, Phys. Rev. A, 31 (1985), pp. 1695.

[28] D. Horstmann, *From 1970 until present: the Keller-Segel model in chemotaxis and its consequences*, Jahresber. Dtsch. Math.-Ver., 105 (2003), pp. 103–165.
[29] C.-R. Hwang and S.-J. Sheu, *Large-time behavior of perturbed diffusion Markov processes with applications to the second eigenvalue problem for Fokker-Planck operators and simulated annealing*, Acta. Appl. Math., 19 (1990), pp. 253–295.
[30] S. Jin, L. Li and J.-G. Liu, *Convergence of Random Batch Method for interacting particles with disparate species and weights*, SIAM J. Numer. Anal., 59 (2021), pp. 746–768.
[31] S. Jin, L. Li and J.-G. Liu, *Random Batch methods (RBM) for interacting particle systems*, J. Comput. Phys., 400 (2020), pp. 108877.
[32] S. Jin, L. Li, Z. Xu and Y. Zhao, *A random batch Ewald method for particle systems with Coulomb interactions*, SIAM J. Sci. Comput., 43 (2021), pp. B937–B960.
[33] S. Jin and X. Li, *Random batch algorithms for quantum Monte Carlo simulations*, Comm. Comp. Phys., 28 (2020), pp. 1907–1936.
[34] J. K. Johnson, J. A. Zollweg and K. E. Gubbins, *The Lennard-Jones equation of state revisited*, Mol. Phys., 78 (1993), pp. 591–618.
[35] K. Kawasaki, *Simple derivations of generalized linear and nonlinear Langevin equations*, J. Phys. A Math. Nucl. Gen., 6 (1973), pp. 1289.
[36] D. Ko and E. Zuazua, *Model predictive control with random batch methods for a guiding problem*. arXiv preprint arXiv:2004.14834, 2020.
[37] J.-M. Lasry and P.-L. Lions, *Mean field games*, Jpn. J. Math., 2 (2007), pp. 229–260.
[38] B. Leimkuhler and C. Matthews, *Rational construction of stochastic numerical methods for molecular sampling*, Appl. Math. Res. eXpress, 2013 (2013), pp. 34–56.
[39] B. Leimkuhler and C. Matthews, *Robust and efficient configurational molecular sampling via Langevin dynamics*, J. Chem. Phys., 138 (2013), pp. 05B601˙1.
[40] L. Li, Y. Li, J.-G. Liu, Z. Liu and J. Lu, *A stochastic version of Stein variational gradient descent for efficient sampling*, Commun. App. Math. Comp. Sci., 15 (2020), pp. 37–63.
[41] 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.
[42] L. Li, Z. Xu and Y. Zhao, *A random-batch Monte Carlo method for many-body systems with singular kernels*, SIAM J. Sci. Comput., 42 (2020), pp. A1486–A1509.
[43] 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*, J. Chem. Phys., 108 (1998), pp. 3383–3385.
[44] J. C. Mattingly, A. M. Stuart and D. J. Higham, *Ergodicity for SDEs and approximations: locally Lipschitz vector fields and degenerate noise*, Stoch. Process Appl., 101 (2002), pp. 185–232.
[45] S. Motsch and E. Tadmor, *Heterophilious dynamics enhances consensus*, SIAM Rev., 56 (2014), pp. 577–621.
[46] K. Nanbu, *Direct simulation scheme derived from the Boltzmann equation. i. monocomponent gases*, J. Phys. Soc. Jpn., 49 (1980), pp. 2042–2049.
[47] H. Robbins and S. Monro (1951). *A stochastic approximation method, The annals of mathematical statistics*, 400–407.
[48] V. Rokhlin, *Rapid solution of integral equations of classical potential theory*, J. Comput. Phys., 60 (1985), pp. 187–207.
[49] H. E. Stanley, *Phase Transitions and Critical Phenomena*, Clarendon Press, Oxford, 1971.
[50] P. J. M. van Laarhoven and E. H. L. Aarts, Simulated Annealing: Theory and Applications, D. Reidel Publishing Co, Dordrecht, 1987.
[51] T. Vicsek, A. Czirók, E. Ben-Jacob, I. Cohen and O. Shochet, *Novel type of phase transition in a system of self-driven particles*, Phys. Rev. Lett., 75 (1995), pp. 1226.
[52] C. Villani, Hypocoercivity (Memoirs of the American Mathematical Society), vol. 202. American Mathematical Society, 2009.
[53] A. A. Vlasov, *Many-Particle Theory and Its Application to Plasma*, Gordon and Breach Science Publishers, Inc., New York, 1961.
[54] M. Welling and Y. W. Teh. *Bayesian learning via stochastic gradient Langevin dynamics*. In: Proceedings of the 28th International Conference on Machine Learning (ICML-11), 2011, pp. 681–688.