

# Random Batch Method with Momentum Correction

Yanshun Zhao<sup>a</sup>, Jingrun Chen<sup>b,\*</sup> and Zhiwen Zhang<sup>c,\*</sup>

<sup>a</sup>*School of Mathematical Sciences, University of Science and Technology of China, Hefei, 230026, People's Republic of China*

<sup>b</sup>*School of Mathematical Sciences and Suzhou Institute for Advanced Research, University of Science and Technology of China, Suzhou 215132, People's Republic of China*

<sup>c</sup>*School of Mathematical Sciences, University of Hong Kong, Pokfulam, Hong Kong*

---

## ARTICLE INFO

### Keywords:

Interacting particle systems  
Fast algorithm  
Random batch method  
Momentum

MSC: 65F10, 65K05, 65Y20

## ABSTRACT

The Random Batch Method (RBM) is an effective technique to reduce the computational complexity when solving certain stochastic differential problems (SDEs) involving interacting particles. It can transform the computational complexity from  $O(N^2)$  to  $O(N)$ , where  $N$  represents the number of particles. However, the traditional RBM can only be effectively applied to interacting particle systems with relatively smooth kernel functions to achieve satisfactory results. To address the issue of non-convergence of the RBM in particle interaction systems with significant singularities, we propose some enhanced methods to make the modified algorithm more applicable. The idea for improvement primarily revolves around a momentum-like correction, and we refer to the enhanced algorithm as the Random Batch Method with Momentum Correction (RBM-M). We provide a theoretical proof to control the error of the algorithm, which ensures that under ideal conditions it has a smaller error than the original algorithm. Finally, numerical experiments have demonstrated the effectiveness of the RBM-M algorithm.

---

## 1. Introduction

### 1.1. Interacting particle systems

Interacting particle systems are a class of stochastic models that have gained significant attention in the realm of applied mathematics and theoretical physics. These systems, comprised of a large number of individual particles or agents, exhibit complex and often emergent behaviors that arise from the intricate interactions among the particles[1, 2]. In general, the interacting particle systems can be represented by the following equation:

$$dX^i = -\nabla V(X^i) dt + \frac{1}{N-1} \sum_{j \neq i} K(X^i - X^j) dt + \sigma dB^i. \quad (1.1)$$

In this stochastic differential equation,  $K(\cdot)$  is the kernel of interacting particle system,  $V$  represents the external flow independent of the interacting particle systems,  $X^i$  represents some property of the particle, such as its position, and  $B$  is a stochastic disturbance while the diffusion term  $\sigma$  is represents the strength of the disturbance.

At the core of interacting particle systems lies the concept of randomness and probability. Each particle or agent within the system is governed by a set of stochastic differential equations, which capture the dynamic evolution of the system over time. These equations incorporate the interactions between the particles, enabling the modeling of a wide range of phenomena, from the diffusion of molecules in a fluid to the collective behavior of social networks.

The mathematical analysis of interacting particle systems requires an in-depth understanding of probability theory, stochastic processes, and partial differential equations. In addition to their theoretical significance, interacting particle systems are widely used in many fields, such as physics, chemistry, biology, economics, and social sciences [3, 4, 5, 6]. From modeling the spread of outbreaks to understanding financial market dynamics, these systems have proven to be powerful frameworks for capturing the complex interactions in the world around us.

Due to the significant theoretical and practical implications of interacting particle systems, finding effective solutions to these systems is of great importance. In other words, devising accurate and efficient numerical methods for solving interacting particle systems is a problem worthy of in-depth investigation.

---

\*Corresponding author

✉ zhangzw@hku.hk (Z. Zhang)

ORCID(s):

Upon initial consideration, the solution to this problem may appear to be a straightforward numerical approach to solving the underlying stochastic differential equations, which can be addressed using techniques such as the Euler method [7] or Runge-Kutta method [8]. However, further examination reveals certain challenges. Examining equation (1.1), we can observe that if we attempt to solve it directly using numerical methods, the computational complexity will scale quadratically with the number of particles,  $N$ , in the interacting particle system. This is because for each particle, the equation accounts for the interactions with all other particles, resulting in a complexity of  $O(N)$ . Given the  $N$  particles in the system, the overall computational complexity becomes  $O(N^2)$ .

For interacting particle systems with a large number of particles, this quadratic computational complexity poses a significant challenge, as it leads to lengthy simulation times. Therefore, if we aim to solve large-scale interacting particle systems in industrial applications, it may not be appropriate to directly apply methods like the Euler or Runge-Kutta techniques to obtain numerical solutions to the stochastic differential equations. Instead, we need to develop improved algorithms that can reduce the computational complexity as much as possible while maintaining the necessary accuracy.

## 1.2. Related work

To mitigate the computational complexity inherent in simulating interacting particle systems, various methods have been proposed by previous researchers. For instance, the Fast Multipole Method (FMM) has been a widely adopted technique since the late 20th century [9]. The FMM provides a means to reduce the computational complexity to  $O(N \log N)$ , or even  $O(N)$  in certain cases, by exploiting the inherent hierarchical structure of the problem. The core idea of the FMM is to partition the computational domain into a hierarchical tree-like structure, where each node in the tree represents a cluster of particles or objects. Through the use of a series of multipole expansions and local Taylor series expansions, the algorithm is able to efficiently approximate the long-range interactions between distant clusters, while accurately computing the short-range interactions between nearby particles. The implementation of the FMM involves several key steps, including the construction of the hierarchical tree structure, the computation of multipole and local Taylor series expansions, and the efficient traversal of the tree to accumulate the contributions from distant clusters. The algorithm also employs various optimization techniques, such as adaptive refinement of the tree structure and the utilization of parallel computing architectures, to further enhance its computational performance.

Additionally, the Tree Code method is another fast algorithm for calculating multi-body interactions. This method also utilizes a hierarchical tree structure to represent the particle distribution, and can significantly reduce the computational complexity by aggregating the contributions from distant particles [10].

Furthermore, there exist mesh-based methods, such as the Particle-Mesh Method [11], which transform the continuous space into a grid and perform calculations on the grid. This approach can leverage fast Fourier transform and other techniques to efficiently compute long-range interactions, thereby substantially reducing the computational complexity.

In recent years, a novel algorithm known as the Random Batch Method (RBM) has gained significant attention from many researchers due to its simplicity of implementation and impressive computational efficiency [12, 13]. The core principle underlying RBM is to divide the particle system into smaller, randomly selected batches, and then perform the computations on these batches rather than the entire system. This approach effectively mitigates the computational complexity from the traditional  $O(N^2)$  scaling to a more manageable  $O(N)$  or even better scaling. The key advantage of RBM lies in its ability to approximate the long-range interactions between distant particles by considering only the nearby particles within each randomly selected batch. Furthermore, RBM can be easily parallelized, and this parallelization capability serves to further enhance the computational performance of the algorithm, making it well-suited for simulating large-scale particle systems. Notably, the implementation of RBM is much simpler compared to the more sophisticated Fast Multipole Method (FMM) and tree code methods. This unique combination of simplicity and high computational efficiency has rendered RBM a popular choice among researchers and practitioners across a wide range of applications, including fluid mechanics, plasma physics, and molecular dynamics.

## 1.3. Advantage of Algorithm

However, despite the good results achieved by the above fast algorithms, there are still some small limitations. Therefore, this paper proposes the RBM with Momentum Correction (RBM-M) algorithm, which is an evolution of the standard RBM algorithm and offers several advantages. The main contributions of this work are as follows:

1. The RBM-M algorithm shares a similar underlying concept with the RBM algorithm, making it relatively simple to understand and implement.

2. In the face of particle systems with relatively large singularities, the RBM-M algorithm can achieve higher accuracy compared to the standard RBM (see theorem 3.2), while maintaining a similar computational time. Numerical experiments demonstrate that the error of RBM-M can be reduced to half that of RBM in some cases (Detail in table 3).
3. The paper provides a theoretical analysis, proving an upper bound on the error of the RBM-M algorithm. It also analyzes in detail the reasons why the error is smaller than that of the RBM algorithm. In the other hand, from a probabilistic perspective, the statistical properties of RBM-M are demonstrated (Detail in the theorem 3.4, the theorem 3.5 and the figure 7).
4. The RBM-M algorithm introduces a hyperparameter  $\beta$ , which allows for the flexible control of the correction strength. As the singularity of the system increases,  $\beta$  can be adjusted to obtain higher accuracy. Consequently, the RBM-M algorithm exhibits improved robustness and flexibility, making it suitable for a wider range of example systems.

In Section 2, we will be devoted to our proposed algorithm, a momentum-based modification of the random batch method, henceforth referred to as RBM-M. Then we will provide a rigorous analysis of the error characteristics of this RBM-M approach in Section 3. Section 4 will then detail numerical experiments conducted to validate the efficacy of the conclusions and algorithms associated with our RBM-M method. Finally, Section 5 will discuss potential future research directions and offer concluding remarks.

## 2. RBM-M

### 2.1. The deficiency of RBM

While the Random Batch Method (RBM) offers numerous advantages, it also possesses a certain limitation. Specifically, in order to obtain more accurate results, it is necessary to impose certain restrictions on the interacting particle system to ensure that the system's singularity is not excessively large. The following conclusions provide an analysis of the error associated with RBM under specific conditions [12].

**Proposition 2.1.** *Suppose  $V$  is strongly convex on  $\mathbb{R}^d$  so that  $x \mapsto V(x) - \frac{1}{2}|x|^2$  is convex, and  $\nabla V, \nabla^2 V$  have polynomial growth (i.e.,  $|\nabla V(x)| + |\nabla^2 V(x)| \leq C(1 + |x|^q)$  for some  $q > 0$ ). Assume  $K(\cdot)$  is bounded, Lipschitz on  $\mathbb{R}^d$  with Lipschitz constant  $L$  and has bounded second order derivatives. Then the error of RBM:*

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

It can be observed that in order to minimize the numerical error of the RBM as the time discretization step size  $\tau$  approaches zero, it is necessary to ensure that the kernel function  $K$  of the interacting particle system is relatively smooth, and the external flow  $V$  also satisfies certain properties.

However, not all interacting particle systems possess these desirable properties. Furthermore, when the singularity of the particle system is relatively large, the error associated with RBM can become substantial, and it may even be the case that the method does not converge at all. Therefore, the challenge of improving the algorithm to enhance the stability of RBM and obtain accurate results in the presence of particle systems with relatively large singularities is a problem that merits further investigation.

### 2.2. Regularization of kernel

When the RBM was first introduced, certain restrictions were imposed on the kernel function of the interacting particle systems in order to ensure the numerical solution converges to the true solution. Since these restrictions on the kernel function limit the scope of applicability, it would be desirable to explore ways to relax these constraints, thereby expanding the range of problems that RBM can effectively address [14, 15].

The way is:

$$K_\delta(z) = K(z) \frac{|z|^2}{|z|^2 + \delta^2}. \quad (2.1)$$

This is a very simple modification, but it can really achieve good results. After this regulation, some kernel are polished so that it satisfies the convergent condition, like Keller-Segal kernel[16] and Biot–Savart kernel[17]. Taking the KS kernel as an example.

For the KS algorithm, the kernel function  $K$  in two-dimensional space is:

$$K(z) = C \cdot \frac{z}{\|z\|^2}.$$

Specifically, taking a 2D example:

For the x-direction:

$$K_1(b - a) = \frac{C \cdot (b_x - a_x)}{\|b - a\|^2}.$$

For the y-direction:

$$K_2(b - a) = \frac{C \cdot (b_y - a_y)}{\|b - a\|^2}.$$

Where  $a$  and  $b$  represent the positions of the two examples respectively, and their corresponding angle symbols indicate whether the component is in the x- direction or the y-direction, and  $C$  is a constant.

For this kernel function, there is a singularity at  $z = 0$ . Hence, a regularization method is adopted:

$$K_\delta(z) = K(z) \cdot \frac{\|z\|^2}{\|z\|^2 + \delta^2} = C \cdot \frac{z}{\|z\|^2 + \delta^2}.$$

Therefore, the new kernel function  $K_\delta(z)$  is continuous on  $\mathbb{R}^d$ , and obviously  $K_\delta(z)$  is globally stable, hence  $K_\delta(z)$  is Lipschitz continuous on  $\mathbb{R}^d$ . Then the interacting particle systems satisfies the requirements of the proposition 2.1, so its error can be guaranteed by the proposition.

### 2.3. Correct algorithm with momentum

Based on the introduction of previous related work, we have gained a certain understanding of the RBM. The key idea behind RBM is to leverage the concept of randomness to process all particles in batches. Guided by this principle, small and randomly selected batches of particle interactions are utilized, thereby reducing the computational cost of the n-particle system for binary interactions from  $O(N^2)$  per time step to  $O(N)$  per time step.

Carefully examining the algorithmic steps of RBM, one can observe that its underlying concept is highly similar to that of Stochastic Gradient Descent (SGD) [18, 19]. Both methods involve processing the entire dataset by randomly selecting a portion of the samples, with the role of this partial sample approximating that of the entire dataset. Using this random subset instead of the full dataset can significantly reduce the computational complexity.

However, there are certain issues associated with the use of SGD. One of the primary shortcomings is the inherent noise in the gradient estimates. Since the gradient is computed using a single data point or a small batch of data points, the resulting gradient can be noisy and may not accurately represent the true gradient of the entire dataset. This can lead to oscillations in the parameter updates and slow down the convergence of the algorithm, particularly in the later stages of the optimization process. To address this issue, various techniques have been proposed, such as the use of momentum. Momentum-based methods, like the Nesterov Accelerated Gradient (NAG) algorithm, incorporate the previous gradient updates to smooth out the parameter updates and accelerate the convergence [20, 21].

Upon a careful examination of the challenges encountered with the RBM, it has been observed that the algorithm's accuracy is diminished when applied to interacting particle systems with relatively large singularities. We postulate that this issue arises from a similar shortcoming as that of Stochastic Gradient Descent (SGD) algorithms. Specifically, since the particle interactions are computed using small batches of data points, the resulting interactions may not accurately represent the true interactions of the entire particle system. This can lead to inaccurate calculation of the interaction terms when the governing equations are updated, consequently introducing substantial errors.

Drawing inspiration from the advancements in momentum correction for SGD, we have explored the feasibility of incorporating momentum into the RBM algorithm to rectify its numerical performance. Accordingly, we propose a

variant, named RBM with Momentum Correction (RBM-M), which aims to address the limitations observed in the original RBM formulation:

---

**Algorithm 1** RBM-M

---

$K(X^i - X^j) = 0$  when time=0  
for  $m$  in  $1:[T/\tau]$  do  
    Divide  $\{1, 2, \dots, pn\}$  into  $n$  batches randomly.  
    for each batch  $C_q$  do  
        Update  $X^i$ 's ( $i \in C_q$ ) by solving the following SDE with  $t \in [t_{m-1}, t_m)$   
         $K_{cor}(X^i - X^j) = \beta K_{t-1}(X^i - X^j) + (1 - \beta)K_t(X^i - X^j)$   
         $dX^i = -\nabla V(X^i) dt + \frac{1}{p-1} \sum_{j \in C_q, j \neq i} K_{cor}(X^i - X^j) dt + \sigma dB^i$   
    end for  
end for

---

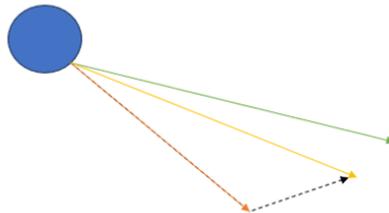
The RBM-M is designed to address the challenges associated with interacting particle systems, where the evolution time spans the interval  $[0, T]$ , and the discrete time step is denoted as  $\tau$ . The algorithmic flow can be outlined as follows: For the solution of the governing equations at each time step, the total  $N$  particles are divided into  $n$  groups, each containing  $p$  particles (if the division is not evenly distributed, the final group may have fewer than  $p$  particles). For each element within the group  $C_p$ , the local interactions are computed and used as a proxy for the global interactions to solve the governing equations. After solving the equations for one time step, the particle grouping is reorganized, and the aforementioned process is repeated until the final time  $T$  is reached. The interactions calculated within this algorithmic framework serve as the cornerstone for the RBM's approach to modeling interacting particle systems.

$$K_{cor}(X^i - X^j) = \beta K_{t-1}(X^i - X^j) + (1 - \beta)K_t(X^i - X^j). \quad (2.2)$$

The RBM-M employs an exponential weighted averaging technique, where each update step comprises two components: the interaction of the current time instance and the particle dynamics from the previous time step. The hyperparameter  $\beta$ , which can be determined empirically, plays a crucial role in this process.

When  $\beta$  is smaller, the correction effect is less pronounced, and the update equation primarily considers the information from the current time step. Conversely, as  $\beta$  increases, the impact of momentum becomes more significant, leading to a stronger corrective effect. However, this also introduces the potential for additional errors.

By performing the  $k$ -th evolution, the  $K_{cor}$  term incorporates information from all the preceding  $k$  evolutions, thereby enhancing the stability and robustness of the algorithm. This approach can be visualized more intuitively through the figure 1 provided. In the diagram, from top to bottom, the first green line represents the real interaction,



**Figure 1:** RBM-M algorithm diagram.

the second yellow line represents the RBM-M calculation, and the third orange line represents the RBM calculation. The gray dashed line represents the momentum correction effect, that is, the improvement of RBM by RBM-M.

**Remark 2.1.** In the RBM-M, the hyperparameter  $\beta$  exhibits a distinct role compared to Stochastic Gradient Descent (SGD) with momentum. In RBM-M, the particle position changes at each step, leading to a different distribution at every time step. Consequently, the previous information can introduce new errors in addition to the correction. To mitigate

the impact of these introduced errors, the hyperparameter  $\beta$  in RBM-M is required to be relatively small, typically less than 0.1, as further analyzed in the theoretical section. In contrast, SGD with momentum operates on a consistent data distribution, where the gradient of each update is theoretically subject to the same distribution. Therefore, to enhance the algorithm's stability, the updates in SGD with momentum are made in small increments, i.e.,  $\beta$  is set to a large value, and the newly added information is only modified by a small amplitude.

**Remark 2.2.** Compared to the original RBM, RBM-M introduces additional computations when calculating the interaction. However, this extra information is already available, and no additional data is required. Equivalently, the extra computation per evolution in RBM-M amounts to two multiplications and one addition, which can be considered negligible. As a result, the actual computational complexity of RBM-M remains  $O(N)$ , similar to RBM.

### 3. Theoretical analysis of RBM-M

The RBM-M algorithm has been described in greater detail. Now, we will provide a proof of the algorithm's performance.

#### 3.1. The error of RBM-M

First, we will denote the true solution as  $E$ . Additionally, we will represent the solution obtained by the RBM algorithm as  $u_p^i$  and the solution obtained by the RBM-M algorithm as  $\tilde{u}_p^i$ , where the subscript  $i$  denotes the  $i$ -th iteration, and the subscript  $p$  represents the number of particles per group.

Second, let us review the algorithmic process of RBM-M:

$$\tilde{u}_p^0 = u_p^0, \quad (3.1)$$

$$\tilde{u}_p^i = \beta \tilde{u}_p^{i-1} + (1 - \beta) u_p^i (i \geq 1). \quad (3.2)$$

And through some calculations related to exponential weighted averaging we can know that:

$$\begin{aligned} \tilde{u}_p^n &= \beta \tilde{u}_p^{n-1} + (1 - \beta) u_p^n \\ &= \beta(\beta \tilde{u}_p^{n-2} + (1 - \beta) u_p^{n-1}) + (1 - \beta) u_p^n \\ &= \beta^n u_p^0 + \beta^{n-1} (1 - \beta) u_p^1 + \dots + \beta(1 - \beta) u_p^{n-1} + (1 - \beta) u_p^n \\ &= \beta^n u_p^0 + (1 - \beta) \sum_{i=1}^n \beta^{n-i} u_p^i. \end{aligned}$$

Next, we will estimate the error of the RBM-M algorithm. Before that, we should note that based on our previous assumptions, the parameters  $\beta$  and  $\tau$  are both very small. Then, we will present some propositions that will be used to prove the error bound.

First, we will analyze the errors in adjacent time steps of the equation:

**Proposition 3.1.** Assuming the conditions proposition 2.1 in are met, then

$$\| u_p^t - u_p^{t-1} \| \leq C\tau, \quad t \in (t_{i-1}, t_i). \quad (3.3)$$

The proposition can be proved directly by the boundedness of kernel function.

With the preliminary preparation, the error bounds of RBM-M can be proved:

**Theorem 3.2.** Suppose  $V$  is strongly convex on  $\mathbb{R}^d$  so that  $x \mapsto V(x) - \frac{1}{2}|x|^2$  is convex, and  $\nabla V, \nabla^2 V$  have polynomial growth (i.e.,  $|\nabla V(x)| + |\nabla^2 V(x)| \leq C(1 + |x|^q)$  for some  $q > 0$ ). Assume  $K(\cdot)$  is bounded, Lipschitz on  $\mathbb{R}^d$  with Lipschitz constant  $L$  and has bounded second order derivatives. Then the error of RBM-M:

$$\| E - \tilde{u}_p^n \| \leq C(1 - \beta^2) \sqrt{\frac{\tau}{p-1} + \tau^2} + C\beta(1 - \beta)\tau + O(\beta^2) \lesssim C(1 - \beta^2) \sqrt{\tau}. \quad (3.4)$$

*Proof.* Substitute the definition of  $\tilde{u}_p^n$ :

$$\| E - \tilde{u}_p^n \| \leq \| E - (\beta^n u_p^0 + (1 - \beta) \sum_{i=1}^n \beta^{n-i} u_p^i) \| .$$

$\beta$  is vary small so that the high-order terms of  $\beta$  can be almost ignored:

$$\begin{aligned} & \| E - (\beta^n u_p^0 + (1 - \beta) \sum_{i=1}^n \beta^{n-i} u_p^i) \| \\ & \leq \| (1 - \beta^2)E - (1 - \beta)\beta u_{p1}^{n-1} - (1 - \beta)u_{p2}^n \| + O(\beta^2), \end{aligned}$$

where  $p1$  and  $p2$  are two different groups in RBM.

Let  $u_{p1}^{n-1}$  continue running time  $\tau$ , considering  $u_{p1}^n$ :

$$\begin{aligned} & \| (1 - \beta^2)E - (1 - \beta)\beta u_{p1}^{n-1} - (1 - \beta)u_{p2}^n \| + O(\beta^2) \\ & = \| (1 - \beta^2)E - (1 - \beta)\beta u_{p1}^n - (1 - \beta)u_{p2}^n + (1 - \beta)\beta(u_{p1}^n - u_{p1}^{n-1}) \| + O(\beta^2) \\ & \leq \beta(1 - \beta) \| E - u_{p1}^n \| + (1 - \beta) \| E - u_{p2}^n \| + \beta(1 - \beta) \| u_{p1}^n - u_{p1}^{n-1} \| + O(\beta^2). \end{aligned}$$

By proposition 2.1:

$$\| E - u_{p1}^n \| \leq C \sqrt{\frac{\tau}{p-1} + \tau^2},$$

$$\| E - u_{p2}^n \| \leq C \sqrt{\frac{\tau}{p-1} + \tau^2}.$$

By proposition 3.1:

$$\| u_{p1}^n - u_{p1}^{n-1} \| \leq C\tau.$$

Note that compare to  $\tau$ ,  $\tau^2$  is very small, so:

$$\begin{aligned} & \| E - \tilde{u}_p^n \| \\ & \leq C(1 - \beta) \sqrt{\frac{\tau}{p-1} + \tau^2} + C\beta(1 - \beta) \sqrt{\frac{\tau}{p-1} + \tau^2} + C\beta(1 - \beta)\tau + O(\beta^2) \\ & \leq C(1 - \beta^2) \sqrt{\frac{\tau}{p-1} + \tau^2} + C\beta(1 - \beta)\tau + O(\beta^2) \\ & \lesssim C(1 - \beta^2) \sqrt{\tau}. \end{aligned}$$

□

We have mathematically proven that the error of the RBM-M is **smaller than** that of the RBM.

The error of the RBM-M can be decomposed into two terms. The first term of error (3.4) is:

$$(1 - \beta^2) \sqrt{\frac{\tau}{p-1} + \tau^2},$$

which represents the main part of the error and corresponds to the error of the RBM. The second term of error (3.4) is:

$$\beta(1 - \beta)\tau + O(\beta^2),$$

which represents the additional error introduced by the momentum term.

Based on the form of the error expression, we hypothesize that the RBM-M algorithm exhibits better performance when the kernel of the RBM has a relatively large singularity. The rationale is that when the kernel has a large singularity, the error of the RBM ( $\sqrt{\frac{\tau}{p-1} + \tau^2}$ ) is significant. In this case, the effect caused by the coefficient  $(1 - \beta^2)$  becomes more prominent. However, the additional term  $(\beta(1 - \beta)\tau + O(\beta^2))$  has little relevance to the singularity of the kernel. As long as  $\beta$  becomes larger, this error term also becomes larger. Therefore, the overall error of the RBM-M is a balanced result. When the singularity of the kernel is large, the advantages brought by the momentum correction outweigh the disadvantages caused by the additional error term, enabling the RBM-M to achieve a more accurate solution. We plan to validate this hypothesis through numerical experiments in the subsequent analysis.

### 3.2. Good statistical properties about variance and expectations

On the other hand, we can explain the advantages of the RBM-M algorithm from a probabilistic perspective. We first consider the situation of the RBM, for which the following theoretical analysis [12] has been provided, with its expectation and variance respectively:

**Proposition 3.3.** Consider  $p \geq 2$  and a given fixed  $x = (x^1, \dots, x^N) \in \mathbb{R}^d$ . Denote:

$$X_{m,i}(x) = \frac{1}{p-1} \sum_{j \in C_i, j \neq i} K(x^i - x^j) - \frac{1}{N-1} \sum_{j \neq i} K(x^i - x^j), \quad (3.5)$$

Then, for all  $i$ ,

$$\mathbb{E}X_{m,i}(x) = 0, \quad (3.6)$$

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

$$\text{Var}(X_{m,i}(x)) = \mathbb{E}[X_{m,i}(x)]^2 = \left( \frac{1}{p-1} - \frac{1}{N-1} \right) \Lambda_i(t), \quad (3.7)$$

where

$$\Lambda_i(t) = \frac{1}{N-2} \sum_{j \neq i} |K(x^i - x^j) - \frac{1}{N-1} \sum_{k \neq i} K(x^i - x^k)|^2. \quad (3.8)$$

Then we consider RBM-M, denoting:

$$\tilde{K}(x_0^i - x_0^j) = K(x_0^i - x_0^j), \quad (3.9)$$

$$\tilde{K}(x_n^i - x_n^j) = \beta \tilde{K}(x_{n-1}^i - x_{n-1}^j) + (1 - \beta)K(x_n^i - x_n^j), \quad (n \geq 1). \quad (3.10)$$

while  $x^i, x^j$  means different particles, subscript 'n' means n-th time.

Let  $\tilde{K}$  is the exponential weighted averaging of interacting particles term:

$$\begin{aligned} \tilde{K}_n^i &= \beta \tilde{K}_{n-1}^i + \frac{1-\beta}{p-1} \sum_{j \in C_n, j \neq i} K(x_n^i - x_n^j) \\ &= \frac{\beta^n}{p-1} \sum_{j \in C_0, j \neq i} K(x_0^i - x_0^j) + \frac{1-\beta}{p-1} \sum_{s=1}^n \beta^{n-s} \sum_{j \in C_s, j \neq i} K(x_s^i - x_s^j). \end{aligned}$$

while  $C_s$  means different groups.

Define:

$$Y_{n,i}(x) = \tilde{K}_n^i - \frac{1}{N-1} \sum_{i \neq j} K(x_n^i - x_n^j). \quad (3.11)$$

Next, we calculate and prove some statistical properties of RBM-M. The first is unbiasedness, which has the following theorem:

**Theorem 3.4.** *The interacting particles term of RBM-M is asymptotic unbiased estimation of true interacting particles term, which means that:*

$$\lim_{\substack{\tau \rightarrow 0 \\ \beta \rightarrow 0}} \mathbb{E}(Y_{n,i}(x)) = 0. \quad (3.12)$$

*Proof.* By the mathematical definition of expectation, we know that

$$\mathbb{E}(X_{n,i}(x)) = \mathbb{E}\left(\frac{1}{p-1} \sum_{j \in C, j \neq i} K(x^i - x^j) - \frac{1}{N-1} \sum_{j \neq i} K(x^i - x^j)\right) = 0.$$

Then we have:

$$\begin{aligned} \|\mathbb{E}(Y_{n,i}(x))\| &\leq \beta^n \|\mathbb{E}\left(\frac{1}{p-1} \sum_{j \in C_0, j \neq i} K(x_0^i - x_0^j) - \frac{1}{N-1} \sum_{j \neq i} K(x_n^i - x_n^j)\right)\| \\ &+ \dots + \beta(1-\beta) \|\mathbb{E}\left(\frac{1}{p-1} \sum_{j \in C_{n-1}, j \neq i} K(x_{n-1}^i - x_{n-1}^j) - \frac{1}{N-1} \sum_{j \neq i} K(x_n^i - x_n^j)\right)\| \\ &+ (1-\beta) \|\mathbb{E}\left(\frac{1}{p-1} \sum_{j \in C_n, j \neq i} K(x_n^i - x_n^j) - \frac{1}{N-1} \sum_{j \neq i} K(x_n^i - x_n^j)\right)\| \\ &\leq \frac{\beta^n}{N-1} \sum_{j \neq i} \|K(x_0^i - x_0^j) - K(x_n^i - x_n^j)\| \\ &+ \dots + \frac{\beta(1-\beta)}{N-1} \sum_{j \neq i} \|K(x_{n-1}^i - x_{n-1}^j) - K(x_n^i - x_n^j)\| \\ &+ \frac{(1-\beta)}{N-1} \sum_{j \neq i} \|K(x_n^i - x_n^j) - K(x_n^i - x_n^j)\|. \end{aligned} \quad (3.13)$$

By proposition 3.1 and the Lipschitz of K and  $\|x\|$  is bounded, we can get it easily:

$$\|K(x_n^i - x_n^j) - K(x_0^i - x_0^j)\| \leq Cn\tau. \quad (3.14)$$

Substitute the equation (3.14) into the equation (3.13):

$$\begin{aligned} \|\mathbb{E}(Y_{n,i}(x))\| &\leq \frac{\beta^n}{N-1} C_1 n\tau + \frac{\beta^{n-1}(1-\beta)}{N-1} C_2 (n-1)\tau + \dots + \frac{\beta(1-\beta)}{N-1} C_{n-1} \tau \\ &\leq \frac{C\tau}{N-1} (n\beta^n + (n-1)\beta^{n-1} + \dots + \beta). \end{aligned} \quad (3.15)$$

Using misalignment subtraction:

$$n\beta^n + (n-1)\beta^{n-1} + \dots + \beta = \frac{\beta - \beta^{n+1}}{(1-\beta)^2} - \frac{n\beta^{n+1}}{1-\beta}. \quad (3.16)$$

Substitute the equation (3.16) into the equation (3.15):

$$\|\mathbb{E}(Y_{n,i}(x))\| \leq \frac{C\tau}{N-1} \left( \frac{\beta - \beta^{n+1}}{(1-\beta)^2} - \frac{n\beta^{n+1}}{1-\beta} \right) \leq \frac{C\beta\tau}{(1-\beta)^2(N-1)}. \quad (3.17)$$

So we can get:

$$\lim_{\substack{\tau \rightarrow 0 \\ \beta \rightarrow 0}} \mathbb{E}(Y_{n,i}(x)) = 0.$$

□

Then there are some properties of variance that prove that the variance of RBM-M is more stable:

**Theorem 3.5.** Assume when  $t_1 \neq t_2$ ,  $K(x_{t_1}^i - x_{t_1}^j)$  and  $K(x_{t_2}^i - x_{t_2}^j)$  are independent (it is reasonable for the re grouping of different time periods to be independent of each other), and  $\tau \rightarrow 0, \beta \rightarrow 0$ , then we have the following estimation formula:

$$\text{Var}(Y_{n,i}(x)) < \left( \frac{(1-\beta)^2}{1-\beta^2} + o(\beta^{2n+1}) \right) \text{Var}(X_{n,i}) + 2\tau \frac{\beta^2}{(1-\beta^2)^2} \lesssim \frac{(1-\beta)^2}{1-\beta^2} \text{Var}(X_{n,i}), \quad (3.18)$$

while  $\text{Var}(X_{n,i})$  is RBM's variance.

*Proof.* According to the definition of variance:

$$\text{Var}(Y_{n,i}) = \mathbb{E}(Y_{n,i}^2) - (\mathbb{E}(Y_{n,i}))^2,$$

processing them separately:

$$\begin{aligned} \mathbb{E}(Y_{n,i}^2) &= \mathbb{E} \left| \tilde{K}_n^i - \frac{1}{N-1} \sum_{i \neq j} K(x_n^i - x_n^j) \right|^2 \\ &= \mathbb{E} \left| \tilde{K}_n^i \right|^2 - \frac{2}{N-1} \sum_{i \neq j} K(x_n^i - x_n^j) \mathbb{E} \left| \tilde{K}_n^i \right| + \frac{1}{(N-1)^2} \sum_{i \neq j} K(x_n^i - x_n^j) \sum_{i \neq k} K(x_n^i - x_n^k), \\ (\mathbb{E}(Y_{n,i}))^2 &= \left( \mathbb{E} \left( \tilde{K}_n^i - \frac{1}{N-1} \sum_{i \neq j} K(x_n^i - x_n^j) \right) \right)^2 \\ &= (\mathbb{E}(\tilde{K}_n^i))^2 - \frac{2}{N-1} \sum_{i \neq j} K(x_n^i - x_n^j) \mathbb{E} \left| \tilde{K}_n^i \right| + \frac{1}{(N-1)^2} \sum_{i \neq j} K(x_n^i - x_n^j) \sum_{i \neq k} K(x_n^i - x_n^k). \end{aligned}$$

Rewrite the variance based on the two calculation of expectations:

$$\begin{aligned} \text{Var}(Y_{n,i}) &= \mathbb{E} \left| \tilde{K}_n^i \right|^2 - (\mathbb{E}(\tilde{K}_n^i))^2 \\ &= \mathbb{E} \left| \frac{\beta^n}{p-1} \sum_{j \in C_0, j \neq i} K(x_0^i - x_0^j) + \frac{1-\beta}{p-1} \sum_{s=1}^n \beta^{n-s} \sum_{j \in C_s, j \neq i} K(x_s^i - x_s^j) \right|^2 \\ &\quad - \left( \mathbb{E} \left( \frac{\beta^n}{p-1} \sum_{j \in C_0, j \neq i} K(x_0^i - x_0^j) + \frac{1-\beta}{p-1} \sum_{s=1}^n \beta^{n-s} \sum_{j \in C_s, j \neq i} K(x_s^i - x_s^j) \right) \right)^2. \end{aligned}$$

Use the independence of  $K(x_{t_1}^i - x_{t_1}^j)$  and  $K(x_{t_2}^i - x_{t_2}^j)$  ( $t_1 \neq t_2$ ):

$$\begin{aligned} \text{Var}(Y_{n,i}) &= \frac{\beta^{2n}}{(p-1)^2} \left( \mathbb{E} \left| \sum_{j \in C_0, j \neq i} K(x_0^i - x_0^j) \right|^2 - \left( \frac{1}{N-1} \sum_{j \neq i} K(x_0^i - x_0^j) \right)^2 \right) \\ &\quad + \frac{(1-\beta)^2}{(p-1)^2} \sum_{s=1}^n \beta^{2(n-s)} \sum_{j \in C_s, j \neq i} \left( \mathbb{E} \left| \sum_{j \in C_0, j \neq i} K(x_s^i - x_s^j) \right|^2 - \left( \frac{1}{N-1} \sum_{j \neq i} K(x_s^i - x_s^j) \right)^2 \right). \end{aligned}$$

Denote

$$X_{s,i}(x) = \frac{1}{p-1} \sum_{j \in C_i, j \neq i} K(x_s^i - x_s^j) - \frac{1}{N-1} \sum_{j \neq i} K(x_s^i - x_s^j).$$

Note that  $X_{n,i}(x)$  is the variance of RBM, then:

$$| \text{Var}(Y_{n,i}) | = | \beta^{2n} \text{Var}(X_{0,i}) + \beta^{2(n-1)}(1-\beta)^2 \text{Var}(X_{1,i}) + \dots + \beta^2(1-\beta)^2 \text{Var}(X_{n-1,i}) + (1-\beta)^2 \text{Var}(X_{n,i}) |,$$

because  $K(\cdot)$  is bound,so:

$$| K(x_0^i - x_0^j) - \frac{1}{N-1} \sum_{k_2 \neq i} K(x_0^i - x_0^{k_2}) | \leq | K(x_n^i - x_n^j) - \frac{1}{N-1} \sum_{k_1 \neq i} K(x_n^i - x_n^{k_1}) | + 2n\tau,$$

hence:

$$\text{Var}(X_{0,i}) \leq \text{Var}(X_{n,i}) + 2n\tau C,$$

add up the variances of all different time steps:

$$\begin{aligned} \text{Var}(Y_{n,i}) &< ((1-\beta)^2 + \beta^2(1-\beta)^2 + \dots + \beta^{2n-2}(1-\beta)^2 + \beta^{2n}) \text{Var}(X_{n,i}) \\ &\quad + C\tau(2n\beta^{2n} + 2(n-1)\beta^{2n-2} + \dots + \beta^2(1-\beta)^2) \\ &\leq \left( \frac{(1-\beta)^2}{1-\beta^2} + \frac{2\beta^{2n+1}}{1+\beta} \right) \text{Var}(X_{n,i}) + \frac{2C\beta^2\tau}{(1-\beta^2)^2}. \end{aligned}$$

So when  $\tau \rightarrow 0$  and ignoring higher-order term of  $\beta$ ,we can get:

$$\text{Var}(Y_{n,i}) < \frac{(1-\beta)^2}{1-\beta^2} \text{Var}(X_{n,i}).$$

□

Through the above theoretical analysis, we know that when the hyperparameter  $\beta$  is appropriately selected, the error of the RBM-M algorithm is smaller than that of the RBM, and it exhibits better properties in terms of statistics and probability. Next, we will conduct some numerical experiments to verify our theoretical conclusions from the numerical perspective.

### 3.3. The error estimation of rescaled relative entropy

The joint distribution of all particles simulated by RBM-M ( $\tilde{\rho}_t^N$ ) can be controlled by numerically solving the corresponding joint distribution ( $\rho_t^N$ ) for all particles. In order to prove this, the particle association law obtained by solving RBM ( $\tilde{\rho}_t^N$ ) is introduced as an auxiliary proof of the intermediate result. What this part needs to estimate is the error about the rescaled relative entropy between RBM-M and reference solution is:

$$\mathcal{H}_N(\tilde{\rho}_t^N | \rho_t^N) = \frac{1}{N} \int_{\mathbb{R}^N} \tilde{\rho}_t^N \log \frac{\tilde{\rho}_t^N}{\rho_t^N}. \quad (3.19)$$

This part of the proof is mainly referenced from [22, 23], using the solution and corrected kernel of RBM-M to replace the solution and kernel of RBM.

For convenience, the velocity field  $-\nabla V(x)$  is denoted as  $b(x)$ . For the joint distribution  $\rho_t^N$  that N particle systems obey at time t, the Fokker-Planck equation is satisfied:

$$\partial_t \rho_t^N + \sum_{i=1}^N \operatorname{div}_{x_i} (\rho_t^N (b(x_i) + K * \rho_t(x_i))) = \sum_{i=1}^N \sigma \Delta_{x_i} \rho_t^N. \quad (3.20)$$

For the original RBM and RBM-M solutions, Euler scheme is generally used to solve SDE to simulate the system evolution. In order to make the distribution smoother, the solution process is processed continuously and the following equation is obtained:

$$X_t = X_{T_k} + (t - T_k)b(X_{T_k}) + (t - T_k)K^{C_k}(X_{T_k}) + \sqrt{2\sigma}(W_{T_{k+1}} - W_{T_k}), \quad t \in (T_k, T_{k+1}). \quad (3.21)$$

In this case, t can continuously take values from  $(T_k, T_{k+1})$  for the smooth simulation of each time.

For RBM,

$$K^{C_k}(X_t^i) := \bar{K}_t^i = \frac{1}{p-1} \sum_{j \in C_k, j \neq i} K(X_t^i - X_t^j), \quad (3.22)$$

and for RBM-M,

$$K^{C_k}(X_t^i) := \tilde{K}_t^i = \frac{1}{p-1} \sum_{j \in C_k, j \neq i} K_{cor}(X_t^i - X_t^j) = \frac{1}{p-1} \sum_{j \in C_k, j \neq i} (\beta K(X_t^i - X_t^j) + (1 - \beta)K_{cor}(X_{t-1}^i - X_{t-1}^j)). \quad (3.23)$$

For continuously solved formula (3.21), the continuous solution distributions obtained by RBM and RBM-M are denoted as  $\bar{\rho}_t^N$  and  $\tilde{\rho}_t^N$ . And for them, the Liouville equation holds[22]:

**Proposition 3.6.** Denote by  $\bar{\rho}_t^{N,C}$  the probability density function of  $X_t = (X_t^1, \dots, X_t^N)$  for  $t \in [T_k, T_{k+1})$ . Then the following Liouville equation holds:

$$\partial_t \bar{\rho}_t^{N,C} + \sum_{i=1}^N \operatorname{div}_{x_i} \left( \bar{\rho}_t^{N,C} \left( \bar{b}_t^{C,i}(x) + \bar{K}_t^{C,i}(x) \right) \right) = \sum_{i=1}^N \sigma \Delta_{x_i} \bar{\rho}_t^{N,C}, \quad (3.24)$$

where

$$\bar{b}_t^{C,i}(x) = \mathbb{E} \left[ b \left( X_{T_k}^i \right) \mid X_t = x, C \right], \quad t \in [T_k, T_{k+1}), \quad (3.25)$$

and

$$\bar{K}_t^{C,i}(x) := \mathbb{E} \left[ \bar{K}_t^i \mid X_t = x, C \right], \quad t \in [T_k, T_{k+1}). \quad (3.26)$$

Here,  $x = (x_1, \dots, x_n) \in \mathbb{R}^{Nd}$ .

In the same way, using the same proof method, we can get the following lemma of RBM-M:

**Lemma 3.1.** Denote by  $\tilde{\rho}_t^{N,C}$  the probability density function of  $X_t = (X_t^1, \dots, X_t^N)$  for  $t \in [T_k, T_{k+1})$ . Then the following Liouville equation holds:

$$\partial_t \tilde{\rho}_t^{N,C} + \sum_{i=1}^N \operatorname{div}_{x_i} \left( \tilde{\rho}_t^{N,C} \left( \tilde{b}_t^{C,i}(x) + \tilde{K}_t^{C,i}(x) \right) \right) = \sum_{i=1}^N \sigma \Delta_{x_i} \tilde{\rho}_t^{N,C}, \quad (3.27)$$

where

$$\tilde{b}_t^{C,i}(x) = \mathbb{E} \left[ b \left( X_{T_k}^i \right) \mid X_t = x, C \right], \quad t \in [T_k, T_{k+1}), \quad (3.28)$$

and

$$\overline{\tilde{K}_t^{C,i}(x)} := \mathbb{E} [\tilde{K}_t^i | X_t = x, C], \quad t \in [T_k, T_{k+1}). \quad (3.29)$$

To prove (3.19), let's start with Fisher information, which is defined by:

$$\mathcal{I}(\rho) = \int \rho |\nabla \rho|^2 dx, \quad (3.30)$$

And, we know that Fisher information satisfies the following inequality:

$$\frac{1}{\mathcal{I}(p * q)} \geq \frac{1}{\mathcal{I}(p)} + \frac{1}{\mathcal{I}(q)}, \quad (3.31)$$

Define a definite mapping:

$$\tilde{\psi}_\tau^{C_k}(\mathbf{x}) := \mathbf{x} + \tau (\mathbf{b}(\mathbf{x}) + \mathbf{K}^{C_k}(\mathbf{x})), \quad (3.32)$$

where  $\mathbf{x}$ ,  $\mathbf{b}(\mathbf{x})$ ,  $\mathbf{K}^{C_k}(\mathbf{x})$  represents the union form of all corresponding variables respectively, i.e

$$\begin{aligned} \mathbf{x} &= (x_1, \dots, x_n) \in \mathbb{R}^{Nd}, \\ \mathbf{b}(\mathbf{x}) &= (b(x_1), \dots, b(x_n))^T \in \mathbb{R}^{Nd}, \\ \tilde{\mathbf{K}}^{C_k}(\mathbf{x}) &= (\tilde{K}^{C_k}(x_1), \dots, \tilde{K}^{C_k}(x_n))^T \in \mathbb{R}^{Nd}. \end{aligned}$$

Let  $\tilde{p}_k(\cdot)$  be the density of random variable  $\tilde{\mathbf{Z}}_k = \tilde{\psi}_\tau^{C_k}(\mathbf{X}_k)$ , then through the conversion of variables, we can get that:

$$\tilde{p}_k(\mathbf{z}) = \frac{\tilde{\varrho}_{T_k}^{N,C}(\mathbf{x})}{\det(\nabla \tilde{\psi}_\tau^{C_k} \mathbf{x})}.$$

Let  $q_\tau$  denote the  $Nd$ -dimensional Gaussian distribution  $\mathcal{N}(0, 2\sigma\tau\mathbf{I}_{Nd})$ . By (3.21) and (3.32), we have that  $\tilde{\varrho}_{T_k}^{N,C}(\mathbf{x}) = p_k(\mathbf{x}) * q_\tau$ . Note that for RBM-M, the velocity field is the same as for RBM. Moreover, if  $K(\cdot)$  is Lipschitz, then  $K_{cor}(\cdot)$  is also Lipschitz. Then we have the following estimate for  $\mathcal{I}(\tilde{\varrho}_{T_k}^{N,C})$  upper bound:

$$\mathcal{I}(\tilde{\varrho}_{T_k}^{N,C}) \leq \max \left( \mathcal{I}(\rho_0^N), \frac{1 + \tau(r+L)}{(1 - \tau(r+L))^2} MN, \frac{Nd(r+L)(3 + \tau(r+L))}{2\sigma} \right). \quad (3.33)$$

For the specific proof process, see [22]. Recall (3.21), the  $\tilde{\varrho}_t^{N,C}$  is the probability density of the RBM for a given sequence of batches  $\mathbf{C} := (C_0, C_1, \dots, C_k, \dots)$ , so that  $\tilde{\rho}_{T_k}^N = \mathbb{E}_{\mathbf{C}} [\tilde{\varrho}_{T_k}^{N,C}]$ . Moreover, by the Markov property, we are able to define

$$\tilde{\rho}_t^{N,C_k} := \mathbb{E} [\tilde{\varrho}_t^{N,C} | C_i, i \geq k], \quad t \in [T_k, T_{k+1}).$$

Next is a concrete analysis of (3.19), which is calculated the time derivative.

**Theorem 3.7.** *Suppose:*

- (a) the field  $b: \mathbb{R}^d \rightarrow \mathbb{R}^d$  and the interaction kernel  $K(\cdot)$  are Lipschitz,
- (b) the Hessians of  $b$  and  $K$  have at most polynomial growth:

$$|\nabla^2 b(x)| \leq C(1 + |x|)^q, \quad |\nabla^2 K(x)| \leq \tilde{C}(1 + |x|)^q.$$

(c)  $\overline{K}^{C,i} - F$  is uniformly bounded, where:

$$F(x_i) = \frac{1}{N} \sum_{j=1}^N K(x_i - x_j), \overline{K}_t^{C,i}(x) = \frac{1}{p} \sum_{j \in C_i} K(x_i - x_j),$$

$$\text{esssup}_C \left\| \overline{K}_t^{C,i} - F \right\|_{L^\infty(\mathbb{R}^d)} < +\infty,$$

then, we have,

$$\sup_t \mathcal{H}_N(\tilde{\rho}_t^N | \rho_t^N) \leq \mathcal{H}_N(\tilde{\rho}_t^0 | \rho_t^0) + C_1 \tau^2 + \frac{C_2}{N}.$$

*Proof.* First we switch the order of integration and derivation, then replace  $\partial_t \tilde{\rho}_t^N$  and  $\partial_t \rho_t^N$  with (3.24) and (3.20), finally we use the Green formula:

$$\begin{aligned} \frac{d}{dt} \mathcal{H}_N(\tilde{\rho}_t^N | \rho_t^N) &= \frac{1}{N} \frac{d(\int_{\mathbb{R}^d} \tilde{\rho}_t^N \log \frac{\tilde{\rho}_t^N}{\rho_t^N})}{dt} \\ &= \frac{1}{N} \int_{\mathbb{R}^d} \partial_t \tilde{\rho}_t^N \log \frac{\tilde{\rho}_t^N}{\rho_t^N} dx + \frac{1}{N} \int_{\mathbb{R}^d} \tilde{\rho}_t^N \left( \frac{\partial_t \tilde{\rho}_t^N}{\tilde{\rho}_t^N} - \frac{\partial_t \rho_t^N}{\rho_t^N} \right) dx \\ &= \frac{1}{N} \int_{\mathbb{R}^d} (\partial_t \tilde{\rho}_t^N) \left( \log \frac{\tilde{\rho}_t^N}{\rho_t^N} + 1 \right) dx + \frac{1}{N} \int_{\mathbb{R}^d} (\partial_t \rho_t^N) \left( -\frac{\tilde{\rho}_t^N}{\rho_t^N} \right) dx \\ &= \frac{1}{N} \sum_{i=1}^N \int_{\mathbb{R}^d} - \left( \mathbb{E}_{C_k} \left[ \text{div}_{x_i} \left( \tilde{\rho}_t^{N,C_k} \left( \tilde{b}_t^{C_k,i}(\mathbf{x}) + \overline{K}_t^{C_k,i}(\mathbf{x}) \right) \right) - \sigma \Delta_{x_i} \tilde{\rho}_t^N \right] \right) \left( \log \frac{\tilde{\rho}_t^N}{\rho_t^N} + 1 \right) dx \\ &\quad + \frac{1}{N} \sum_{i=1}^N \int_{\mathbb{R}^d} \left( -\text{div}_{x_i} (\rho_t^N (b(x_i) + K * \rho_t(x_i))) + \sigma \Delta_{x_i} \rho_t^N \right) \left( -\frac{\tilde{\rho}_t^N}{\rho_t^N} \right) dx \\ &= \frac{1}{N} \sum_{i=1}^N \int_{\mathbb{R}^d} \left( \mathbb{E}_{C_k} \left( \tilde{\rho}_t^{N,C_k} \left( \tilde{b}_t^{C_k,i}(\mathbf{x}) + \overline{K}_t^{C_k,i}(\mathbf{x}) \right) \right) - \sigma \text{div}_{x_i} \tilde{\rho}_t^N \right) \cdot \left( \nabla_{x_i} \log \frac{\tilde{\rho}_t^N}{\rho_t^N} \right) dx \\ &\quad + \frac{1}{N} \sum_{i=1}^N \int_{\mathbb{R}^d} \left( \rho_t^N (b(x_i) + K * \rho_t(x_i)) - \sigma \text{div}_{x_i} \rho_t^N \right) \cdot \left( -\nabla_{x_i} \frac{\tilde{\rho}_t^N}{\rho_t^N} \right) dx. \end{aligned}$$

Taking note of  $\nabla_{x_i} \log \frac{\tilde{\rho}_t^N}{\rho_t^N} = \frac{\tilde{\rho}_t^N}{\rho_t^N} \nabla_{x_i} \frac{\tilde{\rho}_t^N}{\rho_t^N}$  and introducing auxiliary variables  $\tilde{K}_t^i$  and  $F(x_i) = \frac{1}{N} \sum_{j \neq i} K(x_i - x_j)$ , the above equation is equivalent to

$$\frac{d}{dt} \mathcal{H}_N(\tilde{\rho}_t^N | \rho_t^N) = \frac{1}{N} \sum_{i=1}^N \int_{\mathbb{R}^d} \mathbb{E}_{C_k} \left( \tilde{\rho}_t^{N,C_k} \left( \tilde{b}_t^{C_k,i}(\mathbf{x}) - b(x_i) \right) \right) \cdot \nabla_{x_i} \log \frac{\tilde{\rho}_t^N}{\rho_t^N} dx \quad (3.34)$$

$$+ \frac{1}{N} \sum_{i=1}^N \int_{\mathbb{R}^d} \mathbb{E}_{C_k} \left( \tilde{\rho}_t^{N,C_k} \overline{K}_t^{C_k,i}(\mathbf{x}) - \tilde{\rho}_t^{N,C_k} \tilde{K}_t^i(x_i) \right) \cdot \nabla_{x_i} \log \frac{\tilde{\rho}_t^N}{\rho_t^N} dx \quad (3.35)$$

$$+ \frac{1}{N} \sum_{i=1}^N \int_{\mathbb{R}^d} \mathbb{E}_{C_k} \left( \tilde{\rho}_t^{N,C_k} \tilde{K}_t^i(x_i) - \tilde{\rho}_t^N F(x_i) \right) \cdot \nabla_{x_i} \log \frac{\tilde{\rho}_t^N}{\rho_t^N} dx \quad (3.36)$$

$$+ \frac{1}{N} \sum_{i=1}^N \int_{\mathbb{R}^d} (F(x_i) - K * \rho_t(x_i)) \tilde{\rho}_t^N \cdot \nabla_{x_i} \log \frac{\tilde{\rho}_t^N}{\rho_t^N} dx \quad (3.37)$$

$$-\frac{\sigma}{N} \sum_{i=1}^N \int_{\mathbb{R}^{Nd}} \tilde{\rho}_t^N \left| \nabla_{x_i} \log \frac{\tilde{\rho}_t^N}{\rho_t^N} \right|^2 d\mathbf{x}. \quad (3.38)$$

Among them, unlike [22], there is no  $\mathbb{E}_C (\tilde{K}_t^i) = F(x_i)$ , i.e.  $\mathbb{E}_C (\tilde{K}_t^i)$  is no longer an unbiased estimate of  $F(x_i)$ , so extra processing is required here. Consider another batch  $\tilde{C}_k$  independent of  $C_k$ , so,

$$\begin{aligned} & \frac{1}{N} \sum_{i=1}^N \int_{\mathbb{R}^{Nd}} \mathbb{E}_{C_k} \left( \tilde{\rho}_t^{N, C_k} \tilde{K}_t^i(x_i) - \tilde{\rho}_t^N F(x_i) \right) \cdot \nabla_{x_i} \log \frac{\tilde{\rho}_t^N}{\rho_t^N} d\mathbf{x} \\ &= \frac{1}{N} \sum_{i=1}^N \int_{\mathbb{R}^{Nd}} \left( \mathbb{E}_{C_k, \tilde{C}_k} \left[ \left( \tilde{\rho}_t^{N, C_k} - \tilde{\rho}_t^{N, \tilde{C}_k} \right) \left( \tilde{K}_t^i(x_i) - F(x_i) \right) \right] + \mathbb{E}_{C_k} \left[ \tilde{K}_t^i(x_i) - F(x_i) \right] \tilde{\rho}_t^N \right) \cdot \nabla_{x_i} \log \frac{\tilde{\rho}_t^N}{\rho_t^N} d\mathbf{x}. \end{aligned}$$

Take note of  $ab \leq ra^2 + \frac{1}{4r}b^2$ , apply it then,

$$\begin{aligned} & \int_{\mathbb{R}^{Nd}} \mathbb{E}_{C_k, \tilde{C}_k} \left[ \left( \tilde{K}_t^i(x_i) - F(x_i) \right) \left( \tilde{\rho}_t^{N, C_k} - \tilde{\rho}_t^{N, \tilde{C}_k} \right) \right] \cdot \nabla_{x_i} \log \frac{\tilde{\rho}_t^N}{\rho_t^N} d\mathbf{x} \\ & \leq \frac{2}{\sigma} \mathbb{E}_{C_k, \tilde{C}_k} \left[ \int_{\mathbb{R}^{Nd}} \left| \tilde{K}_t^i(x_i) - F(x_i) \right|^2 \frac{\left| \tilde{\rho}_t^{N, C_k} - \tilde{\rho}_t^{N, \tilde{C}_k} \right|^2}{\tilde{\rho}_t^{N, \tilde{C}_k}} d\mathbf{x} \right] + \frac{\sigma}{8} \int_{\mathbb{R}^{Nd}} \tilde{\rho}_t^N \left| \nabla_{x_i} \log \frac{\tilde{\rho}_t^N}{\rho_t^N} \right|^2 d\mathbf{x}, \end{aligned}$$

$$\begin{aligned} & \int_{\mathbb{R}^{Nd}} \mathbb{E}_{C_k} \left[ F(x_i) - \tilde{K}_t^i(x_i) \right] \tilde{\rho}_t^N \cdot \nabla_{x_i} \log \frac{\tilde{\rho}_t^N}{\rho_t^N} d\mathbf{x} \\ & \leq \frac{2}{\sigma} \int_{\mathbb{R}^{Nd}} \left( \mathbb{E}_{C_k} \left[ \tilde{K}_t^i(x_i) - F(x_i) \right] \right)^2 \tilde{\rho}_t^N d\mathbf{x} + \frac{\sigma}{8} \int_{\mathbb{R}^{Nd}} \tilde{\rho}_t^N \left| \nabla_{x_i} \log \frac{\tilde{\rho}_t^N}{\rho_t^N} \right|^2 d\mathbf{x}. \end{aligned}$$

Apply (3.17) of theorem 3.4, we have,

$$\int_{\mathbb{R}^{Nd}} \left( \mathbb{E}_{C_k} \left[ \tilde{K}_t^i(x_i) - F(x_i) \right] \right)^2 \tilde{\rho}_t^N d\mathbf{x} \leq c\beta^2\tau^2.$$

where  $c$  is a constant. Then, consider the Log-Sobolev inequality,

$$\mathbb{E}(f \log f) - (\mathbb{E}f) \log(\mathbb{E}f) \leq \frac{1}{2} \mathbb{E}(f |\nabla \log f|^2). \quad (3.39)$$

apply (3.39) so that,

$$\begin{aligned} & \int_{\mathbb{R}^{Nd}} \tilde{\rho}_t^N \left| \nabla_{x_i} \log \frac{\tilde{\rho}_t^N}{\rho_t^N} \right|^2 d\mathbf{x} = \mathbb{E}_{\rho_t^N} \left( \frac{\tilde{\rho}_t^N}{\rho_t^N} \left| \nabla_{x_i} \log \frac{\tilde{\rho}_t^N}{\rho_t^N} \right|^2 \right) \\ & \geq 2\mathbb{E}_{\rho_t^N} \left( \frac{\tilde{\rho}_t^N}{\rho_t^N} \log \frac{\tilde{\rho}_t^N}{\rho_t^N} \right) - 2\mathbb{E}_{\rho_t^N} \left( \frac{\tilde{\rho}_t^N}{\rho_t^N} \right) \log \mathbb{E}_{\rho_t^N} \left( \frac{\tilde{\rho}_t^N}{\rho_t^N} \right) \geq c\mathcal{H}_N(\tilde{\rho}_t^N | \rho_t^N). \end{aligned}$$

The same method is applied to the other items in (3.34), and the following results can be obtained by integrating them:

$$\frac{d}{dt} \mathcal{H}_N(\tilde{\rho}_t^N | \rho_t^N) \leq -\frac{\sigma}{2C_{LS}} \mathcal{H}_N(\tilde{\rho}_t^N | \rho_t^N) + \frac{2}{\sigma N} \sum_{i=1}^N \mathbb{E}_{C_k} \int_{\mathbb{R}^d} |\tilde{b}_t^i(x) - b(x_i)|^2 \tilde{\rho}_t^{N, C_k} d\mathbf{x}$$

$$\begin{aligned}
& + \frac{2}{\sigma N} \sum_{i=1}^N \mathbb{E}_{C_k} \left[ \int_{\mathbb{R}^d} \left| \overline{\tilde{K}_t^{C_k, i}}(x_i) - \tilde{K}_t^i(x_i) \right|^2 \tilde{\rho}_t^{N, C_k} dx \right] \\
& + \frac{2}{\sigma N} \sum_{i=1}^N \mathbb{E}_{C_k, \tilde{C}_k} \left[ \int_{\mathbb{R}^{Nd}} \left| \tilde{K}_t^i(x_i) - F(x_i) \right|^2 \frac{\left| \tilde{\rho}_t^{N, C_k} - \tilde{\rho}_t^{N, \tilde{C}_k} \right|^2}{\tilde{\rho}_t^{N, \tilde{C}_k}} dx \right] \\
& + \frac{2}{\sigma N} \sum_{i=1}^N \int_{\mathbb{R}^{Nd}} \tilde{\rho}_t^N \left| F(x_i) - K * \rho_t(x_i) \right|^2 dx + c\beta^2 \tau^2 \\
& := -\frac{\sigma}{2C_{LS}} H_N \left( \tilde{\rho}_t^N | \rho_t^N \right) + J_1 + J_2 + J_3 + J_4 + c\beta^2 \tau^2,
\end{aligned}$$

where  $C_{LS}$  and  $c$  are positive constants. Since RBM and RBM-M are almost identical except for kernel, the estimation of the same term like  $J_1$  and  $J_4$  can be directly referred to in the analysis [22]. The estimates of several terms related to  $\tilde{K}(\cdot)$  are given here. For  $J_2$ , we have  $\tilde{K}_t^i(x_i) = \mathbb{E} \left[ \tilde{K}_t^i(\tilde{X}_t^i) \mid \tilde{\mathbf{X}}_t = \mathbf{x}, C_k \right]$  and apply Taylor's expansion for any  $t \in [T_k, T_{k+1})$ :

$$\begin{aligned}
\overline{\tilde{K}_t^{C_k, i}}(x_i) - \tilde{K}_t^i(x_i) &= \mathbb{E} \left[ \tilde{K}_t^i(\tilde{X}_{T_k}^i) - \tilde{K}_t^i(\tilde{X}_t^i) \mid \tilde{\mathbf{X}}_t = \mathbf{x}, C_k \right] \\
&= \mathbb{E} \left[ \tilde{X}_{T_k}^i - \tilde{X}_t^i \mid \tilde{\mathbf{X}}_t = \mathbf{x}, C_k \right] \cdot \nabla_{x_i} \tilde{K}_t^i(x_i) + \hat{r}_t(x_i),
\end{aligned}$$

where the remainder takes the form

$$\hat{r}_t(x_i) := \frac{1}{2} \mathbb{E} \left[ \left( \tilde{X}_{T_k}^i - \tilde{X}_t^i \right)^{\otimes 2} : \int_0^1 \nabla_{x_i}^2 \tilde{K}_t^i \left( (1-s)\tilde{X}_t^i + s\tilde{X}_{T_k}^i \right) ds \mid \tilde{\mathbf{X}}_t = \mathbf{x}, C_k \right].$$

Note that the Hessians of  $K$  have at most polynomial growth through, ignore the higher order of  $\beta$  and  $\sup_{t \geq 0} \mathbb{E} \left( |\tilde{X}_t^i|^p \mid \mathcal{F}_C \right) \leq C_p$  through [24]:

$$\begin{aligned}
\left| \hat{r}_t(x_i) \right| &\leq \int_0^1 \mathbb{E} \left[ \left| \nabla_{x_i}^2 \tilde{K}_t^i \left( (1-s)\tilde{X}_t^i + s\tilde{X}_{T_k}^i \right) \right| \cdot \left| \tilde{X}_{T_k}^i - \tilde{X}_t^i \right|^2 \mid \tilde{\mathbf{X}}_t = \mathbf{x}, C_k \right] ds \\
&\lesssim C \mathbb{E} \left[ \left| \beta \left( 1 + |\tilde{X}_t^i|^p + |\tilde{X}_{T_k}^i|^p \right) + \beta(1-\beta) \left( 1 + |\tilde{X}_{t-1}^i|^p + |\tilde{X}_{T_k-1}^i|^p \right) \right| \cdot \left| \tilde{X}_{T_k}^i - \tilde{X}_t^i \right|^2 \mid \tilde{\mathbf{X}}_t = \mathbf{x}, C_k \right] \\
&\leq \tilde{C} \mathbb{E} \left[ \left| \tilde{X}_{T_k}^i - \tilde{X}_t^i \right|^2 \mid \tilde{\mathbf{X}}_t = \mathbf{x}, C_k \right].
\end{aligned}$$

The remaining part of  $J_2$  has nothing to do with kernel, and the conclusion can be drawn by referring to the proof of [22]:

$$J_2 \leq \tilde{c}' \tau^2 \left( 3 + \frac{1}{N} \mathcal{I} \left( \tilde{\rho}_{T_k}^N \right) \right).$$

Then for the  $J_3$ , through the assumption that  $|\overline{\tilde{K}_t^i} - F(x_i)|$  is bounded and kernel  $K(\cdot)$  is Lipschitz continuous:

$$\left| \tilde{K}_t^i(x_i) - F(x_i) \right| \leq \left| \overline{\tilde{K}_t^i}(x_i) - F(x_i) \right| + \left| \overline{\tilde{K}_t^i}(x_i) - \tilde{K}_t^i(x_i) \right| \leq C_{bound} + C_L L.$$

**Table 1**  
The relationship between the error and the time step.

$\tau \times 10^{-3}$	error
16	0.0188
4	0.0102
1	0.0051
0.25	0.0025

which means  $|\tilde{K}_t^i - F(x_i)|$  is also bounded, then we only need to estimate  $\mathbb{E}_{C_k, \tilde{C}_k} \left[ \int_{\mathbb{R}^{Nd}} \frac{|\tilde{\rho}_t^{N, C_k} - \tilde{\rho}_t^{N, \tilde{C}_k}|^2}{\tilde{\rho}_t^{N, \varepsilon_k}} d\mathbf{x} \right]$  which has been given by [22]. So that we have

$$J_3 \leq \tilde{c}'' \tau^2 \left( 1 + \frac{1}{N} \mathcal{I}(\tilde{\rho}_{T_k}^N) \right).$$

Finally, combining all of the above results can be obtained

$$\frac{d}{dt} \mathcal{H}_N(\tilde{\rho}_t^N | \rho_t^N) \leq -C_0 \mathcal{H}_N(\tilde{\rho}_t^N | \rho_t^N) + C_1 \tau^2 + \frac{C_2}{N},$$

then

$$\sup_t \mathcal{H}_N(\tilde{\rho}_t^N | \rho_t^N) \leq \mathcal{H}_N(\tilde{\rho}_t^0 | \rho_t^0) + C_1 \tau^2 + \frac{C_2}{N}.$$

□

## 4. Numerical experiment

This section mainly demonstrates a set of numerical experiments to help verify the theories presented in the previous section. In the following numerical experiments, regardless of whether the RBM or RBM-M algorithm is used, the total number of particles  $N$  in the interactive particle system is 10,000, and the number of particles  $p$  in each group is 360. This setting takes into account the computational efficiency, and the numerical instability caused by a small number of particles in each group can be avoided.

The following numerical experiments can correspond to the previous theoretical analysis one by one, including the convergence order of the kernel function after regularizing, the error of RBM-M and the good statistical properties of RBM-M. In addition, the theoretical results are supplemented to some extent by numerical experiments.

### 4.1. The error is proportional to $\sqrt{\tau}$

According to the theorem 3.2 that we gave the description, the error of RBM-M and discrete time step  $\tau$  should be proportional.

We select biot-savart to verify this conclusion. In the corresponding interacting particle systems, the corresponding kernel function is[25]:

$$K(z) = \frac{z^\perp}{\|z\|^2}.$$

Here  $z^\perp$  is the orthogonal complement of  $z$ . At the same time, for the sake of simplicity, the diffusion term  $\sigma$  of the equation is selected to obey the standard normal distribution, and the applied flow  $V$  of the system is 0.

The numerical results are recorded in table 1, and the result of numerical experiment that we do can prove it greatly:

It can be seen that as the time step of the dispersion becomes  $\frac{1}{4}$  of the original, the error of the algorithm basically becomes half of the original. This shows that  $\sqrt{\tau}$  is proportional to error, which is exactly what our theory tells us.

## 4.2. Compare the error between RBM-M and RBM

In this section, we will primarily focus on comparing the algorithmic accuracy of RBM-M and RBM under various scenarios (primarily involving different kernel functions, representing different interacting particle systems). Some of these interacting particle systems are derived from real physical equations, while others are self-constructed systems with significant singularity, which are specifically designed to verify the stability of the algorithms. The comparison will include both first-order and higher-order systems. By examining several numerical examples, we can comprehensively evaluate the performance between the two algorithms.

As an illustrative example, we will choose a second-order interacting particle system. The specific form of this second-order system is:

$$dX_t^{i,N} = V_t^{i,N} dt, \quad (4.1)$$

$$dV_t^i = \frac{1}{N} \sum_{j \neq i} K(|X_t^j - X_t^i|) (V_t^j - V_t^i) dt + \sigma dB_t^i, \quad (4.2)$$

where the kernel function  $k$  has the concrete form:

$$K(|X_t^j - X_t^i|) = \frac{X_t^j - X_t^i}{1 + |X_t^j - X_t^i|^2}. \quad (4.3)$$

In this second-order system, the variable  $x$  can be interpreted as the position of the particle, and the variable  $v$  can be interpreted as the velocity of the particle. This provides a clear physical interpretation of the entire system. The system is a concrete representation of Newton's second law of motion. Equation (4.1) shows that the position of the particle is determined by the product of velocity and time, while Equation (4.2) can be understood as the effect of an external force on the particle, namely the acceleration, and the velocity is determined by the acceleration.

We set the initial conditions of the stochastic differential equation (SDE) such that the initial values of  $X$  are distributed uniformly within a unit circle, while the initial values of  $V$  are set to 0. Additionally, the evolution time is set to  $T = 0.02s$ , and the discrete time step is  $\tau = 0.001s$ . In this case, we will compare the real solution of the interacting particle system, the numerical solution of RBM, and the numerical solution of RBM-M (with a control parameter  $\beta = 0.01$ ). The results are shown in the figure 2 and 3.

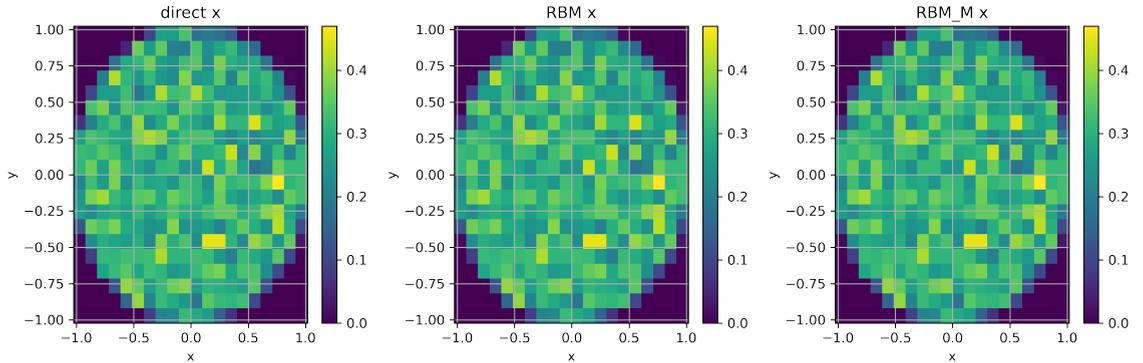


Figure 2: Two order system-x.

It can be intuitively observed from the above figure that both RBM and RBM-M have achieved good results.

In addition to intuitively comparing the results from the figure above, the specific  $L_2$  values of RBM and RBM-M are also listed here. The  $L_2$  error here is defined as:

$$l = \sqrt{\sum_{i=1}^N (x^i - x_{true}^i)^2}. \quad (4.4)$$

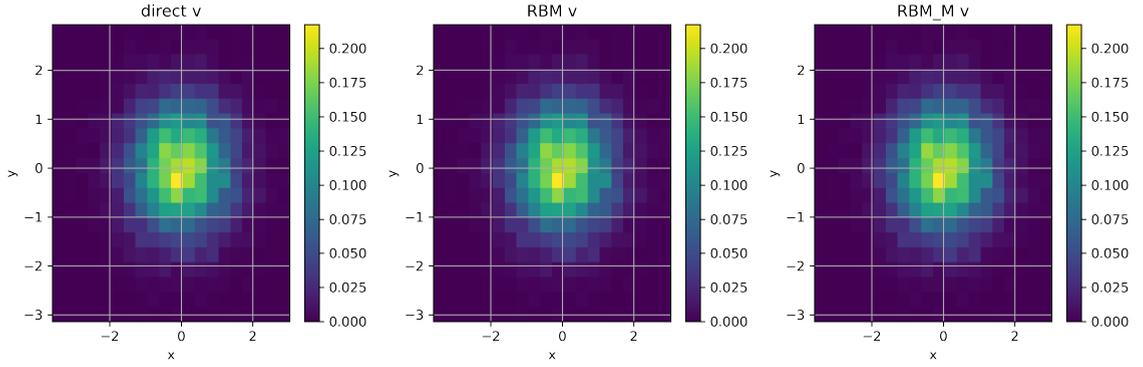


Figure 3: Two order system-v.

Here  $x^i$  represents the numerical value and  $x_{true}^i$  represents the real solution. Under this definition, the specific numerical result is: the error of RBM is  $4.489 \times 10^{-2}$ , and the error of RBM-M is  $4.419 \times 10^{-2}$ . In this example, although the error of RBM-M is slightly lower than that of RBM, the effect is not significant. We speculate that because this second-order interacting particle system is relatively smooth, the correction brought by the introduction of momentum is not substantial. At the same time, the introduction of momentum may also introduce some errors, so the value of the control parameter  $\beta$  is small in this case. However, from another perspective, the performance of RBM-M should be no worse than RBM, provided that the parameter  $\beta$  is reasonably selected.

To further demonstrate that the RBM-M algorithm can indeed achieve good results when the interacting particle system has a large singularity, we construct the following system:

$$dX^i = -\nabla V(X^i) dt + \frac{1}{N-1} \sum_{j \neq i} K(X^i - X^j) dt + \sigma dB^i, \quad (4.5)$$

where kernel function is:

$$K_x(s_i - s) = \frac{x_i - x}{\cosh(\|s_i - s\|^2)}, \quad (4.6)$$

$$K_y(s_i - s) = \frac{\cosh(y_i - y)}{\|s_i - s\|^2}. \quad (4.7)$$

here  $s = (x, y)$  represents the position of the particle, and  $K_x(\cdot)$  and  $K_y(\cdot)$  represent the components of the kernel function in the X-axis and y-direction, respectively. The additional set system flow is:

$$-\nabla V(s_i) = \begin{pmatrix} 0 \\ \cos(x_i) \end{pmatrix}. \quad (4.8)$$

Due to the large singularity of this system, RBM-M theoretically needs stronger correction effect, so the control parameter  $\beta=0.1$  here has a larger value. The remaining settings are consistent with the above second-order system, i.e., the initial value is that the initial values of  $X$  are distributed uniformly in a unit circle, the evolution time of termination is  $T = 0.02s$ , and the discrete time step is  $\tau = 0.001s$ . So the standard solution, the solution of RBM and the solution of RBM-M are respectively in figure 4.

There is still no difference between the two numerical algorithms. Similarly, according to the definition of formula (4.4), compare the  $L_2$  error of the value, where the error of RBM is  $7.946 \times 10^{-2}$ , and the error of RBM-M is only  $4.691 \times 10^{-2}$ . In this example, the difference between the two algorithms is still obvious, and the effect of RBM-M is obviously better than that of RBM.

In addition, it should be noted that in this example, if the original equation is solved, when  $N = 10,000$  particles are selected, it takes  $54.48s$  to solve, while with the same number of particles, it only takes  $4.09s$  to use RBM and  $4.45s$

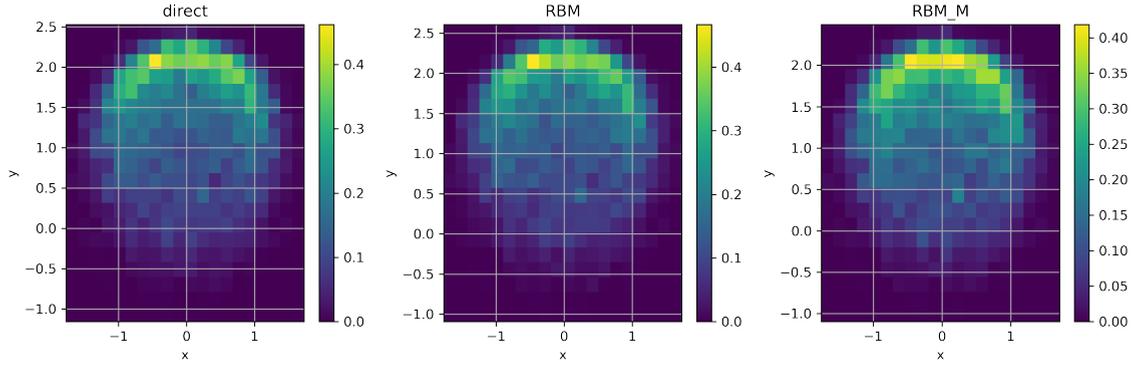


Figure 4: A case of large singularity.

Table 2  
The effect of parameter  $\beta$  on the algorithm.

$\beta$	RBM error	RBM-M error
0.04	$8.029 \times 10^{-2}$	$4.529 \times 10^{-2}$
0.06	$7.984 \times 10^{-2}$	$3.449 \times 10^{-2}$
0.08	$8.057 \times 10^{-2}$	$3.519 \times 10^{-2}$
0.10	$7.946 \times 10^{-2}$	$4.695 \times 10^{-2}$
0.12	$8.041 \times 10^{-2}$	$6.423 \times 10^{-2}$

to use RBM-M. This shows that both RBM and RBM-M can significantly reduce the computation time, and RBM-M has little additional time compared to RBM.

At the same time, with this example, the influence of control parameter  $\beta$  on the correction effect of RBM-M algorithm can be explored. We set the value of  $\beta$  from  $\{0.04, 0.06, 0.08, 0.1, 0.12\}$ , and take the selected beta as the parameter of RBM-M to compare the accuracy of RBM-M algorithm under different parameter conditions, the results are in table 2.

As can be seen from the above table, with the increase of control parameter  $\beta$  from 0.04 to 0.12, the error of RBM-M changes to the original 56.41%, **43.19%**, **43.67%**, 59.09% and 79.88% respectively, which is, when  $\beta$  is selected reasonably, the error can be reduced to less than half of the original. Due to the large singularity of the system, the corrective effect of the algorithm is significant. Therefore, in the beginning, as  $\beta$  increases, the performance of the RBM-M algorithm improves gradually. However, as  $\beta$  continues to increase, the corrective effect reaches saturation, and the positive impact of further increasing  $\beta$  becomes limited, while the introduced error becomes larger and larger, resulting in a deterioration of the final outcome. In summary, when the system singularity is small, the selected  $\beta$  should be small; and when the system singularity is large, the selected  $\beta$  should be large. In addition, we also tested on other systems, The kernel functions employed in the various systems are as follows:

K1: Biot-Savart kernel:

$$K(z) = \frac{z^\perp}{\|z\|}.$$

K2: two order system:

$$dX_t^{i,N} = V_t^{i,N} dt,$$

$$dV_t^i = \frac{1}{N} \sum_{j \neq i} K(|X_t^j - X_t^i|) (V_t^j - V_t^i) dt + \sigma dB_t^i.$$

**Table 3**  
**Error contrast between RBM and RBM-M.**

kernel	RBM error	RBM-M error
1	$8.0033 \times 10^{-3}$	$7.8905 \times 10^{-3}$
2	$4.849 \times 10^{-2}$	$4.819 \times 10^{-2}$
3	$7.817 \times 10^{-2}$	$7.792 \times 10^{-2}$
4	$7.946 \times 10^{-2}$	$4.691 \times 10^{-2}$
5	$3.589 \times 10^{-1}$	$2.342 \times 10^{-1}$
6	$4.289 \times 10^{-1}$	$2.423 \times 10^{-1}$

K3: repulsive-attractive Morse potential[26]:

$$K(\mathbf{x}) = C_r e^{-\frac{\|\mathbf{x}\|}{l_r}} - C_a e^{-\frac{\|\mathbf{x}\|}{l_a}}.$$

K4: Let the particle position be set as  $s_i = (x_i, y_i)$

x-dimension:

$$\frac{x_i - x}{\cosh(\|s_i - s\|^2)},$$

y-dimension:

$$\frac{e^{-(y_i - y)^2}}{\|s_i - s\|}.$$

K5: Let the particle position be set as  $s_i = (x_i, y_i)$

x-dimension:

$$\frac{\sinh(x_i - x)}{\|s_i - s\|^2},$$

y-dimension:

$$\frac{\cosh(y_i - y)}{\|s_i - s\|^2}.$$

K6: Let the particle position be set as  $s_i = (x_i, y_i)$

$$K(z) = \frac{\ln(-\exp(z^2) + 1)}{\tan(\|z\|^2)/10 + 2\pi}.$$

The corresponding  $L_2$  error is shown in the table 3.

The first column represents the different kinds of kernels, the second column represents the error of RBM solution, and the third column represents the error of RBM-M solution. It is important to note that due to the different final results evolved from various equations (some distributions are more concentrated, while others are more diffuse), comparing the absolute errors between different equations is not very meaningful. The focus should be on comparing the errors of the two algorithms.

It can be clearly observed that as the kernel singularity increases, the effect of momentum correction becomes more pronounced. This suggests that the local momentum algorithm has significant effects in certain situations.

### 4.3. Probabilistic explanation that RBM-M is more stable than RBM

Now, let us consider characterizing the Singularity of kernels. We define a family of similar kernels as:

$$f(x) = \begin{cases} -\frac{1}{x-1} & x < 1 - \alpha \\ \frac{1}{\alpha^2} |x-1| & 1 - \alpha < x < 1 + \alpha, \\ \frac{1}{x-1} & x > \alpha + 1 \end{cases} \quad (4.9)$$

where  $\alpha$  is a parameter to measure the steepness of a function, and  $\alpha$  is smaller, the function is more steep (when selecting a function, the reason for not taking 0 as the center is to avoid some division of 0 that may lead to a decrease in numerical accuracy). Figure 5 is the figure of  $f(x)$ :

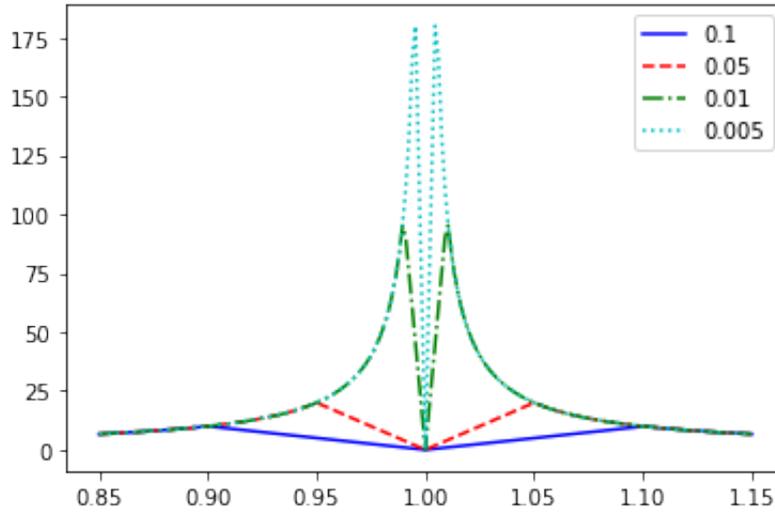


Figure 5: The steepness of different  $\alpha$ .

If  $\alpha \rightarrow 0$ , the function  $f(x)$  becomes a discontinuous function. In this situation, there is a significant degree of singularity associated with  $f(x)$ . If we were to use the standard RBM approach in this case, the error would be quite large. However, as demonstrated by the proof presented earlier, the RBM-M algorithm outperforms the standard RBM when the function exhibits a high degree of singularity. Let us now compare the effects of these two approaches in figure 6:

The figure above compares the errors between the two methods. For each data point represented in the figure, the reported error is the average of 10 repeated calculations, in order to mitigate the effects of random factors. The closer the  $x$ -axis direction is to 1, the smaller the  $\alpha$ , which means the function is more steep and has greater singularity. From the graph, it can be observed that as  $\alpha$  decreases, the singularity of the function increases, and the error of the standard RBM becomes increasingly larger than that of the RBM-M method.

The specific data is in table 4.

In this numerical experiment, we set the coefficient of the RBM-M method,  $\beta$ , to 0.1. When the function is flatter, we use a smaller  $\beta$ , and when the function is more steep, we use a larger  $\beta$  (the final difference has decreased because the singularity of the function is too large, and a larger  $\beta$  should be selected). All of these choices follow the theorem 3.4.

Moreover, we can explain the main idea of RBM and RBM-M from a sampling perspective. (3.5) and (3.11) have explained that the core concept is to use the distribution of a few particles to estimate the overall distribution of all particles. When the singularity of the kernel is relatively small, and the distribution is relatively flat, it is easier to estimate. However, if the kernel singularity is relatively high, as mentioned earlier, it becomes more challenging to estimate accurately. The figure 7 illustrates the sampling processes of two distinct methods. In each time step, the RBM

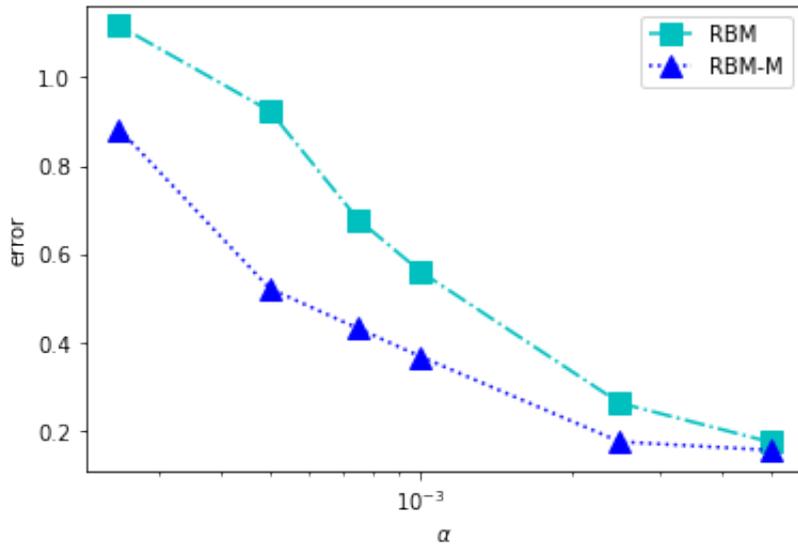


Figure 6: Comparison of errors between RBM and RBM-M.

Table 4

Comparison of errors of two algorithms for functions with different singularity.

$\alpha$	RBM	RBM-M	advantage
$5 \times 10^{-3}$	0.1743	0.1562	0.0181
$2.5 \times 10^{-3}$	0.2627	0.1752	0.0875
$1 \times 10^{-3}$	0.5601	0.3679	0.1923
$7.5 \times 10^{-4}$	0.6784	0.4317	0.2467
$5 \times 10^{-4}$	0.9236	0.5218	0.3418
$2.5 \times 10^{-4}$	1.1161	0.8827	0.2835

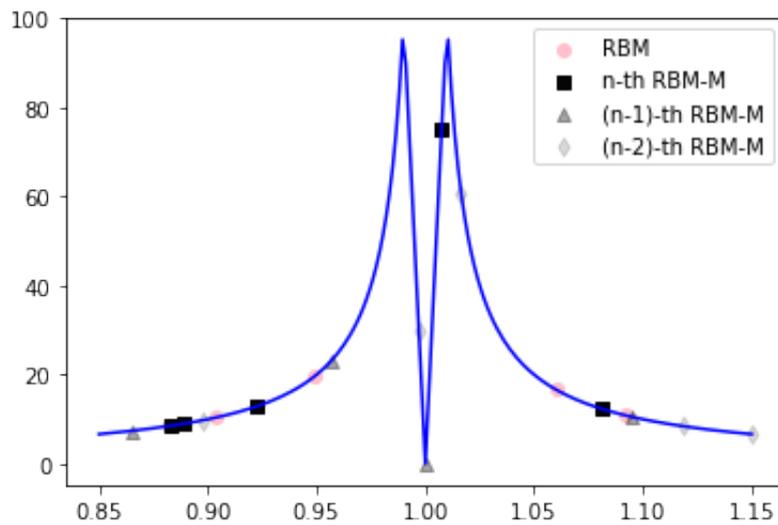


Figure 7: The sampling of RBM and RBM-M.

requires a single sampling operation, while the RBM-M performs multiple sampling iterations due to its consideration of multiple time steps. The pink circular data points represent the particles sampled by the RBM at a specific time step, whereas the black data points depict the samples generated by the RBM-M. Since each step in the RBM-M process influences the subsequent evolution of the system, it is analogous to having a greater number of data points being sampled. The transparency of the black data points in the figure indicates the intensity of the influence of the data on the RBM-M results. The earlier the evolution time, the more transparent the corresponding data points, and the effects of the system's previous evolution will gradually diminish. This visual representation highlights the key differences between the sampling approaches of the RBM and RBM-M methods, underscoring the importance of considering multiple time steps in the RBM-M approach and the resulting impact on the data's influence on the overall system evolution.

The above figure illustrates the sampling process of the two methods. In each time step, the standard RBM needs to sample only once, while the RBM-M samples multiple times by considering multiple time steps. If we directly solve the stochastic differential equations (SDEs), we will definitely sample the steep parts of the function. However, if we use the standard RBM, our sampling may not necessarily represent the entire distribution: if the steep part of the function is not extracted during sampling, the  $|\frac{1}{p-1} \sum_{j \in C_i, j \neq i} K(x^i - x^j)|$  will be slightly small. Even if the steep part of the function is successfully extracted during sampling, the large outlier will only be weakened by the coefficient  $\frac{1}{p-1}$ , when it should be accurately using  $\frac{1}{N-1}$ , so the  $|\frac{1}{p-1} \sum_{j \in C_i, j \neq i} K(x^i - x^j)|$  will be slightly large. Since the RBM-M uses exponential average weighting to consider more particles, it can to some extent avoid the errors caused by sampling.

Furthermore, it can be inferred from (3.18) that the variance of RBM-M is slightly smaller than that of the standard RBM, and both should be (asymptotically) unbiased estimates in an ideal state, so the RBM-M is more likely to converge to a good result.

## 5. Conclusions and future work

The RBM-M algorithm is an extension of the RBM algorithm that introduces a momentum correction. We have carried out a detailed theoretical derivation of RBM-M, including error measurement, expectation, kl divergence, etc., and have supported it with corresponding numerical experiments to prove the correctness of our theory. In addition, for the proposed momentum correction, we have designed a hyperparameter  $\beta$  to control the intensity of the added momentum in response to the different cases of different particle systems. Finally, through hard work, we have demonstrated that the RBM-M can achieve better results than the standard RBM when the interacting particle system has exhibited a relatively large singularity, and the value of the control parameter  $\beta$  should be relatively large in such cases. Conversely, when the singularity of the interacting particle system is relatively small, the value of the control parameter  $\beta$  should be relatively small.

Furthermore, there is potential to expand the application domain of this algorithm to make it suitable for a wider range of particle systems. For example, other algorithms such as Adam [27] and RMSProp [28] could be considered for the correction of RBM. Simultaneously, identifying the optimal parameter  $\beta$  is also a valuable area of research.

## Data availability

No data was used for the research described in the article.

## Acknowledgement

JC was partially supported by the National Natural Science Foundation of China (Project 12425113). ZZ was partially supported by the National Natural Science Foundation of China (Project 92470103), the Hong Kong RGC grant (Project 17304324), an R&D Funding Scheme from the HKU-SCF FinTech Academy, the Outstanding Young Researcher Award of HKU (2020-21), and Seed Funding for Strategic Interdisciplinary Research Scheme 2021/22 (HKU).

## References

- [1] P. Clifford, A. Sudbury, A model for spatial conflict, *Biometrika* 60 (3) (1973) 581–588. doi:10.1093/biomet/60.3.581.
- [2] J. M. Swart, A course in interacting particle systems, arXiv: Probability (2017).

- [3] P. Degond, M. Engel, Coagulation–fragmentation model for animal group-size statistics, *Journal of Nonlinear Science* 27 (2015) 379–424.
- [4] T. Vicsek, A. Czirók, E. Ben-Jacob, I. Cohen, O. Sochet, Novel type of phase transition in a system of self-driven particles., *Physical review letters* 75 6 (1995) 1226–1229.
- [5] J. A. Carrillo, S. Lisini, E. Mainini, Uniqueness for keller-segel-type chemotaxis models, *Discrete and Continuous Dynamical Systems* 34 (2012) 1319–1338.
- [6] S. Motsch, E. Tadmor, Heterophilous dynamics enhances consensus, *SIAM Rev.* 56 (2013) 577–621.
- [7] A. Ferreiro-Castilla, A. E. Kyprianou, R. Scheichl, An euler-poisson scheme for lévy driven sdes, *arXiv: Probability* (2013).
- [8] J. Hong, C. Huang, X. Wang, Strong convergence rate of runge–kutta methods and simplified step- $n$  euler schemes for sdes driven by fractional brownian motions, *arXiv: Numerical Analysis* (2017).
- [9] L. Greengard, V. Rokhlin, A fast algorithm for particle simulations, *Journal of Computational Physics* 73 (2) (1987) 325–348. doi:[https://doi.org/10.1016/0021-9991\(87\)90140-9](https://doi.org/10.1016/0021-9991(87)90140-9).
- [10] J. S. Bagla, Treepm: A code for cosmological n-body simulations, *Journal of Astrophysics and Astronomy* 23 (1999) 185–196.
- [11] Y. Yao, J. Capecelatro, An accurate particle-mesh method for simulating charged particles in wall-bounded flows, *Powder Technology* 387 (2021) 239–250. doi:<https://doi.org/10.1016/j.powtec.2021.04.012>.
- [12] S. Jin, L. Li, J.-G. Liu, Random batch methods (rbm) for interacting particle systems, *Journal of Computational Physics* 400 (2020) 108877. doi:<https://doi.org/10.1016/j.jcp.2019.108877>.
- [13] S. Jin, L. Li, Random batch methods for classical and quantum interacting particle systems and statistical samplings, *ArXiv abs/2104.04337* (2021).
- [14] Z. Wang, J. X. Xin, Z. Zhang, Deepparticle: learning invariant measure by a deep neural network minimizing wasserstein distance on data generated from an interacting particle method, *J. Comput. Phys.* 464 (2021) 111309.
- [15] Z. Wang, J. X. Xin, Z. Zhang, A deepparticle method for learning and generating aggregation patterns in multi-dimensional keller-segel chemotaxis systems, *ArXiv abs/2209.00109* (2022).
- [16] J. Liu, L. Wang, Z. jiang Zhou, Positivity-preserving and asymptotic preserving method for 2d keller-segal equations, *Math. Comput.* 87 (2016) 1165–1189.
- [17] M. Eisterer, The significance of solutions of the inverse biot–savart problem in thick superconductors, *Superconductor Science and Technology* 18 (2004) S58 – S62.
- [18] S. Ruder, An overview of gradient descent optimization algorithms, *ArXiv abs/1609.04747* (2016).
- [19] X. Chen, Z. S. Wu, M. Hong, Understanding gradient clipping in private sgd: A geometric perspective, *ArXiv abs/2006.15429* (2020).
- [20] J. Lin, C. Song, K. He, L. Wang, J. E. Hopcroft, Nesterov accelerated gradient and scale invariance for improving transferability of adversarial examples, *ArXiv abs/1908.06281* (2019).
- [21] L. Kong, M. Tao, Quantitative convergences of lie group momentum optimizers, *ArXiv abs/2405.20390* (2024).
- [22] Z. Huang, S. Jin, L. Li, Mean field error estimate of the random batch method for large interacting particle system, *ArXiv abs/2403.08336* (2024).
- [23] L. Li, Y. Wang, A sharp uniform-in-time error estimate for stochastic gradient langevin dynamics, *ArXiv abs/2207.09304* (2022).
- [24] S. Jin, L. Li, On the mean field limit of the random batch method for interacting particle systems, *Science China Mathematics* 65 (2020) 169 – 202.
- [25] L.-P. Chaintron, A. Diez, Propagation of chaos: A review of models, methods and applications. i. models and methods, *Kinetic and Related Models* (2022).
- [26] J. A. Carrillo, Y. Huang, S. Martin, Explicit flock solutions for quasi-morse potentials, *European Journal of Applied Mathematics* 25 (2013) 553 – 578.
- [27] D. P. Kingma, J. Ba, Adam: A method for stochastic optimization, *CoRR abs/1412.6980* (2014).
- [28] D. Xu, S. Zhang, H. Zhang, D. P. Mandic, Convergence of the rmsprop deep learning method with penalty for nonconvex optimization, *Neural Networks* 139 (2021) 17–23. doi:<https://doi.org/10.1016/j.neunet.2021.02.011>.