Mathematical Models and Methods in Applied Sciences
Vol. 33, No. 1 (2023) 67–102

DOI: 10.1142/S0218202523500033

Ergodicity and long-time behavior of the Random Batch Method for interacting particle systems

Shi Jin

School of Mathematical Sciences,
Institute of Natural Sciences, MOE-LSC,
Shanghai Jiao Tong University,
Shanghai 200240, P. R. China
shijin-m@sjtu.edu.cn

Lei Li

School of Mathematical Sciences,
Institute of Natural Sciences, MOE-LSC,
Qing Yuan Research Institute,
Shanghai Jiao Tong University,
Shanghai 200240, P. R. China
leili2010@sjtu.edu.cn

Xuda Ye[*] and Zhennan Zhou[†]

Beijing International Center for Mathematical Research,
Peking University, Beijing 100871, P. R. China
[*]*abneryepku@pku.edu.cn*
[†]*zhennan@bicmr.pku.edu.cn*

Received 11 February 2022
Revised 10 July 2022
Accepted 27 November 2022
Published 20 January 2023
Communicated by F. Flandoli

We study the geometric ergodicity and the long-time behavior of the Random Batch Method for interacting particle systems, which exhibits superior numerical performance in recent large-scale scientific computing experiments. We show that for both the interacting particle system (IPS) and the random batch interacting particle system (RB–IPS), the distribution laws converge to their respective invariant distributions exponentially, and the convergence rate does not depend on the number of particles N, the time step τ for batch divisions or the batch size p. Moreover, the Wasserstein-1 distance between the invariant distributions of the IPS and the RB–IPS is bounded by $O(\sqrt{\tau})$, showing that the RB–IPS can be used to sample the invariant distribution of the IPS accurately with greatly reduced computational cost.

Keywords: Random batch method; interacting particle system; geometric ergodicity; reflection coupling; strong error estimation.

AMS Subject Classification: 65C20, 37M05

[†]Corresponding author.

1. Introduction

Simulation of large-size dynamical systems has always been a computational bottleneck in optimization and stochastic sampling. One of the main difficulties is that the complexity of updating in a single time step is extremely high, which is often beyond the linear scaling with respect to the size of the system. In the past years, various approximate simulation methods have been developed to reduce the computational cost with tolerable numerical error, for example, the stochastic gradient descent (SGD)[4] and the stochastic gradient Langevin dynamics (SGLD).[40] These methods have been widely used in machine learning for efficient simulation, and one may refer to Refs. 5, 10, 32, 36, 43 for the corresponding error analysis.

In this work, we focus on the interacting particle system (IPS), which is of vital importance in computational physics[16, 19] and computational chemistry.[17, 29] The study of their mean-field limits has also been of significant research interest.[3, 18, 24] Consider a system of N particles represented by a collection of position variables $X_t = \{X_t^i\}_{i=1}^N$, where is $X_t^i \in \mathbb{R}^d$ the position of the ith particle. The system of particles X_t is evolved by the overdamped Langevin dynamics:

$$\mathrm{d}X_t^i = b(X_t^i)\mathrm{d}t + \frac{1}{N-1}\sum_{j\neq i} K(X_t^i - X_t^j)\mathrm{d}t + \sigma \mathrm{d}W_t^i, \quad i = 1, \dots, N. \tag{1.1}$$

Here, $b(\cdot) : \mathbb{R}^d \to \mathbb{R}^d$ is the drift force, $K(\cdot) : \mathbb{R}^d \to \mathbb{R}^d$ is the interaction force, $\sigma > 0$ is a scalar constant, and $W_t = \{W_t^i\}_{i=1}^N$ denotes N independent standard Wiener processes in $\mathbb{R}^d$.

With certain additional assumptions on the parameters, there exists an invariant distribution $\pi \in \mathcal{P}(\mathbb{R}^{Nd})$ associated with the IPS (1.1), and thus (1.1) can be utilized to produce samples of π by time integration. In fact, if the drift force $b(x) = -\nabla U(x)$ and the interaction force $K(x) = -\nabla V(x)$ for some potential functions $U(x), V(x)$ with $\sigma = \sqrt{2}$ and $V(x)$ being even, then the invariant distribution π can be explicitly expressed as

$$\pi(\mathrm{d}x) \propto \exp\left(-\sum_{i=1}^N U(x^i) - \frac{1}{N-1}\sum_{1\le i<j\le N} V(x^i - x^j)\right)\mathrm{d}x. \tag{1.2}$$

To simulate the IPS (1.1) numerically, one has to discretize (1.1) in time and applies numerical integration in each time step. For an IPS of N particles, it requires $O(N^2)$ complexity to compute all the interaction forces $\{K(X_t^i - X_t^j)\}_{i\neq j}$, hence the computational cost per time step is $O(N^2)$, which results in inefficiency of the simulation. Therefore, it is desirable to apply an approximate simulation method which is able to reduce the computational cost and still produce reliable samples of the invariant distribution π.

The Random Batch Method (RBM) proposed in Ref. 25 is a simple random algorithm to reduce the computational cost per time step from $O(N^2)$ to $O(N)$. As supported by extensive numerical tests,[27, 31, 41] the RBM is not only an efficient algorithm for the evolution of the system, it also preserves the invariant distribution

π in an approximate sense, thus can be used to obtain statistical samples of the invariant distribution of the IPS (1.1). Yet, theoretical justification for the sampling accuracy of the RBM is still lacking.

The idea of the RBM is illustrated as follows. Let $\tau > 0$ be the time step for batch divisions and define $t_n := n\tau$. For each $n \geq 0$, let the index set $\{1, \dots, N\}$ be randomly divided into q batches $\mathcal{D} = \{\mathcal{C}_1, \dots, \mathcal{C}_q\}$, where each batch $\mathcal{C} \in \mathcal{D}$ has size $p = N/q$. The IPS (1.1) within the time interval $t \in [t_n, t_{n+1})$ is approximated as the SDE of another particle system $\tilde{X}_t = \{\tilde{X}_t\}_{i=1}^N$ in $\mathbb{R}^{Nd}$, given by

$$\mathrm{d}\tilde{X}_t^i = b(\tilde{X}_t^i)\mathrm{d}t + \frac{1}{p-1}\sum_{j\neq i, j\in\mathcal{C}} K(\tilde{X}_t^i - \tilde{X}_t^j)\mathrm{d}t + \sigma \mathrm{d}W_t^i, \quad i \in \mathcal{C}, \quad t \in [t_n, t_{n+1}), \tag{1.3}$$

where $\mathcal{C} \in \mathcal{D}$ is the batch that contains i. For the next time interval, the previous division $\mathcal{D}$ is discarded and another random division $\mathcal{D}'$ is employed for the dynamics (1.3). We point out that the RBM is not only a numerical method for the IPS (1.1), it is also a stochastic model for interacting particle systems, in which particles interact, within each time interval of length τ, with a small number $(p-1)$ of particles. In the following, the dynamical system (1.3) will be referred to as the random batch interacting particle system (RB–IPS), as a comparison to the IPS (1.1). For the convenience of analysis, assume both (1.1) and (1.3) are exactly integrated in time, and thus there is no error due to numerical discretization.

If one numerically integrates (1.1) and (1.3) in each time step, then the RB–IPS reduces the computational cost per time step from $O(N^2)$ to $O(Np)$, because one only needs to compute the interaction forces within each batch $\mathcal{C}$ to update (1.3) in a single time step. Since the batch $\mathcal{C}$ is expected to capture the binary interactions in the IPS (1.1), the least choice of the batch size is $p = 2$.

The goal of this paper is to answer: does the RB–IPS (1.3) produce accurate samples of the invariant distribution π. Specifically, our question is two-fold:

(1) Does the RB–IPS (1.3) has an invariant distribution $\tilde{\pi} \in \mathcal{P}(\mathbb{R}^{Nd})$?
(2) If so, what is the difference between the invariant distributions π and $\tilde{\pi}$?

In general, the analysis of invariant distributions (which is for the long-time behavior) of the stochastic process, is more challenging than the analysis of strong and weak error in the finite time. The strong and weak error analysis for the RB–IPS (1.3) has been systematically studied in Ref. 26, while the theoretical understanding of the invariant distribution is very limited, except in a random batch consensus model.[21] Intuitively, we expect the trajectory $\tilde{X}_t$ generated by the RB–IPS (1.3) is a good approximation to X_t generated by the IPS (1.1), since the RB–IPS (1.3) provides an unbiased approximation of the interaction forces:

$$\mathbb{E}\left(\frac{1}{p-1}\sum_{j\neq i, j\in\mathcal{C}} K(x^i - x^j)\right) = \frac{1}{N-1}\sum_{j\neq i} K(x^i - x^j), \quad \forall\, x \in \mathbb{R}^{Nd}, \tag{1.4}$$

where i is a fixed index in $\{1, \ldots, N\}$, and the remaining $(p-1)$ elements of the batch $\mathcal{C}$ are randomly chosen from $\{1, \ldots, N\} \backslash \{i\}$. The unbiased feature (1.4) of the RB–IPS (1.3) is very similar to the SGD and the SGLD. Unfortunately, (1.4) is not sufficient to tell the long-time behavior of the RB–IPS (1.3). Essentially, we lack the knowledge of the geometric ergodicty.

The geometric ergodicity of a general stochastic process depicts how fast the distribution law converges to the invariant distribution. For the overdamped and the underdamped Langevin dynamics, the classical approaches to derive geometric ergodicity include the hypocoercivity method,[28, 29, 38] functional inequalities[2, 20] and the Harris ergodic theorem.[9, 22, 23, 33, 35] However, it is not clear how these approaches could be applied to the RB–IPS (1.3). The main difficulty within the RB–IPS (1.3) is that, the structure of the SDE varies in different time steps, preventing direct analysis of the infinitesimal generator.

Recently, the reflection coupling[12, 13] has been employed to prove the geometric ergodicity of the overdamped Langevin dynamics, which is purely probabilistic and is rather different from the classical PDE approaches. In this paper, we aim to adopt the reflection coupling to prove the geometric ergodicity of the RB–IPS (1.3). The basic idea of reflection coupling is to couple the Wiener processes of two dynamics X_t, Y_t in a specially designed regime, and prove their distance $\mathbb{E}[\rho(X_t, Y_t)]$ decays exponentially in time. In particular, the reflection coupling does not require the strong convexity of the potential function $U(x)$ in (1.2). So far, the reflection coupling has been employed to prove the geometric ergodicity of a large variety of dynamical systems: first-order IPS,[11, 13, 15] underdamped Langevin dynamics,[14] Hamiltonian Monte Carlo[7, 8] and the Andersen dynamics.[6] In particular, it has been proved in Ref. 13 that the convergence rate of the IPS (1.1) does not depend on the number of particles N.

The geometric ergodicity together with the Banach fixed point theorem yields the existence of the invariant distribution $\tilde{\pi}$ of the RB–IPS (1.3), thus answers the first question. For the second question, we shall employ the triangle inequality framework described below to estimate the difference between π and $\tilde{\pi}$. Denote the transition kernels of the IPS (1.1) and the RB–IPS (1.3) by p_t and $\tilde{p}_t$, respectively. After choosing a distance function $d(\cdot, \cdot)$ of probability distributions, the estimate of $d(\pi, \tilde{\pi})$ relies on two key conclusions:

(1) **Geometric ergodicity.** For the RB–IPS (1.3), there exist $C_1, c > 0$ such that

$$d(\mu\tilde{p}_t, \nu\tilde{p}_t) \leq C_1 e^{-ct} d(\mu, \nu), \quad \forall\, t \geq 0 \tag{1.5}$$

for any probability distributions μ, ν in $\mathbb{R}^{Nd}$. The geometric ergodicity can be derived using the reflection coupling.

(2) **Finite-time error estimation.** Roughly speaking, we aim to prove

$$\sup_{0 \leq t \leq T} d(\nu p_t, \nu\tilde{p}_t) \leq C_2(T)\tau^{\alpha} \tag{1.6}$$

for given initial distribution ν and some exponent $\alpha > 0$, and $C_2(T)$ depends on simulation time T. The strong error estimation derived in Ref. 26 implies (1.6) with $d(\cdot,\cdot)$ being the Wasserstein-1 distance and $\alpha = 1/2$.

Using these conclusions, $d(\pi,\tilde{\pi})$ can be estimated as follows. For any $t \geq 0$, one has the triangle inequality

$$\begin{aligned} d(\pi,\tilde{\pi}) &= d(\pi p_t, \tilde{\pi}\tilde{p}_t) \\ &\leq d(\pi p_t, \pi\tilde{p}_t) + d(\pi\tilde{p}_t, \tilde{\pi}\tilde{p}_t) \\ &\leq d(\pi p_t, \pi\tilde{p}_t) + C_1 e^{-ct} d(\pi,\tilde{\pi}). \end{aligned} \tag{1.7}$$

By choosing t satisfying $C_1 e^{-ct} = 1/2$, one obtains

$$d(\pi,\tilde{\pi}) \leq 2\cdot d(\pi p_t, \pi\tilde{p}_t) \leq 2C_2(t)\tau^\alpha. \tag{1.8}$$

The triangle inequality (1.7), inspired from Refs. 33, 34 and 37, is the key step in our framework of estimating $d(\mu,\tilde{\mu})$. The logic behind this framework is simple: geometric ergodicity and finite time error estimation imply error in invariant distributions.

Our main result of this paper is briefly described below. Under appropriate dissipation conditions on the drift force $b(\cdot)$ and the interaction force $K(\cdot)$, the RB–IPS (1.3) has geometric ergodicity and the convergence rate does not depend on the number of particles N, the time step τ or the batch size p. Also, the Wasserstein-1 distance (defined in (3.3)) between π and $\tilde{\pi}$ is estimated as

$$\mathcal{W}_1(\pi,\tilde{\pi}) \leq C\sqrt{\frac{\tau}{p-1} + \tau^2}, \tag{1.9}$$

where the constant C does not depend on N, τ, p. We would like to point out that our result shows that the RBM, even as an approximate simulation method of the invariant distribution — which corresponds to the steady state of the system — has a convergence rate *independent of* N.

To utilize the RB–IPS (1.3) as a practical algorithm, one has to employ time-discretization, e.g. the Euler–Maruyama scheme. In terms of the time step τ, the error due to random batch divisions is $O(\sqrt{\tau})$ (see Theorem 3.2), while the error due to time-discretization is no greater than $O(\sqrt{\tau})$ (order of strong error). Therefore, one may simply choose the time step in the random batch divisions and time-discretization to be exactly the same. The error analysis of the resulting discrete-time RB–IPS can be found in a subsequent work.[42]

The paper is organized as follows. Section 2 proves the geometric ergodicity of both the IPS (1.1) and the RB–IPS (1.3). Section 3 proves of existence of invariant distributions $\pi, \tilde{\pi}$ and the strong error estimation in finite time, then estimates the difference between the invariant distributions $\pi, \tilde{\pi}$ of the IPS and the RB–IPS. Section 4 briefly summarizes the result in this paper.

2. Geometric Ergodicity of RB–IPS

In this section, we prove the geometric ergodicity of the RB–IPS (1.3), and the main technique is the reflection coupling.[12, 13] Following the methodology of Ref. 13, we first study the geometric ergodicity of a general multiparticle system: the product model, then apply the results to the IPS (1.1) and the RB–IPS (1.3).

The product model refers to the stochastic process of the particle system $X_t = \{X_t^i\}_{i=1}^N$ in $\mathbb{R}^{Nd}$, which is given by the SDE

$$\mathrm{d}X_t^i = b^i(X_t)\mathrm{d}t + \sigma \mathrm{d}W_t, \quad i = 1, \ldots, N. \tag{2.1}$$

Here, $b^i(\cdot) : \mathbb{R}^{Nd} \to \mathbb{R}^d$ is the total force exerted on the ith particle. The product model (2.1) is so named because it is defined on the product space $\mathbb{R}^{Nd} = \otimes_{i=1}^N \mathbb{R}^d$. Assume $b^i(x)$ is given by

$$b^i(x) = b(x^i) + \gamma^i(x), \quad i = 1, \ldots, N, \tag{2.2}$$

where $b(\cdot) : \mathbb{R}^d \to \mathbb{R}^d$ is the drift force and $\gamma^i(\cdot) : \mathbb{R}^{Nd} \to \mathbb{R}^d$ is the perturbation exerted on each particle. Formally, the IPS (1.1) and the RB–IPS (1.3) can be unified in the product model (2.1). In fact, the product model directly becomes the IPS (1.1) by choosing

$$\gamma^i(x) = \frac{1}{N-1} \sum_{j \neq i} K(x^i - x^j), \quad i = 1, \ldots, N. \tag{2.3}$$

Within each time interval $[t_n, t_{n+1})$, the RB–IPS (1.3) can be viewed as the product model with

$$\gamma^i(x) = \frac{1}{p-1} \sum_{j \neq i, j \in \mathcal{C}} K(x^i - x^j), \quad i \in \mathcal{C}, \tag{2.4}$$

where $\mathcal{C}$ is the batch that contains i. Note that $\gamma^i(x)$ in the RB–IPS (1.3) varies in every time step due to the use of random batches, but we have suppressed the appearance of such dependence for simplicity.

In the following, we shall use the notation $X_t = \{X_t^i\}_{i=1}^N$ to represent both the IPS (1.1) and the product model (2.1), and the notation $\tilde{X}_t = \{\tilde{X}_t^i\}_{i=1}^N$ to represent the RB–IPS (1.3). Using the same notation for the IPS (1.1) and the product model (2.1) will not be ambiguous since the two dynamics are directly related by (2.3).

2.1. *Product model*

We prove the geometric ergodicity of the product model (2.1) using the reflection coupling. Basically, we shall show that the transition kernel p_t of the product model is contractive, i.e. for some $c > 0$ it holds that

$$d(\mu p_t, \nu p_t) \leq e^{-ct} d(\mu, \nu) \tag{2.5}$$

for any probability distributions μ, ν in $\mathbb{R}^{Nd}$. The constant c is also referred to as the *contraction rate* of the dynamics. The contractivity (2.5) can be achieved by considering a coupled dynamics $\{(X_t, Y_t)\}_{t\geq 0}$ in $\mathbb{R}^{Nd}\times\mathbb{R}^{Nd}$, which is described as:

(1) The initial values $X_0 \sim \mu$ and $Y_0 \sim \nu$ (not necessarily independent).
(2) Both $\{X_t\}_{t\geq 0}$ and $\{Y_t\}_{t\geq 0}$ are weak solutions to the product model (2.1).
(3) X_t, Y_t are driven by two Wiener processes W_t^X, W_t^Y, respectively, while W_t^X, W_t^Y are coupled in a specific regime.

The coupled dynamics $\{(X_t, Y_t)\}_{t\geq 0}$ can also be written as the SDE

$$\begin{cases} \mathrm{d}X_t^i = b^i(X_t)\mathrm{d}t + \sigma \mathrm{d}W_t^{X,i} \\ \mathrm{d}Y_t^i = b^i(Y_t)\mathrm{d}t + \sigma \mathrm{d}W_t^{Y,i} \end{cases} \quad i = 1, \ldots, N, \tag{2.6}$$

where $W_t^{X,i}, W_t^{Y,i}$ are the ith arguments of the Wiener processes W_t^X, W_t^Y in $\mathbb{R}^{Nd}$. If one proves for some distance function $\rho(\cdot,\cdot)$ in $\mathbb{R}^{Nd}\times\mathbb{R}^{Nd}$, the expectation $\mathbb{E}[\rho(X_t, Y_t)]$ has exponential decay in time, i.e. for some $c > 0$ it holds that

$$\mathbb{E}[\rho(X_t, Y_t)] \leq e^{-ct}\mathbb{E}[\rho(X_0, Y_0)], \tag{2.7}$$

then the contractivity (2.5) holds with $d(\cdot,\cdot)$ being the Wasserstein distance

$$d(\mu,\nu) := \inf_{\gamma\in\Pi(\mu,\nu)} \int_{\mathbb{R}^{Nd}\times\mathbb{R}^{Nd}} \rho(x, y)\gamma(\mathrm{d}x\mathrm{d}y), \tag{2.8}$$

where $\Pi(\mu,\nu)$ is the set of joint distributions in $\mathbb{R}^{Nd}\times\mathbb{R}^{Nd}$ with marginal distributions μ,ν. The concept of Wasserstein distance has been widely adopted in optimal transport,[1, 39] where $\Pi(\mu,\nu)$ is known as the set of transport plans.

In the definition of the coupled dynamics $\{(X_t, Y_t)\}_{t\geq 0}$, we expect that the coupling scheme between the Wiener processes W_t^X, W_t^Y attracts X_t, Y_t together so that (2.7) holds. Note that the coupling scheme between W_t^X, W_t^Y does not impact the fact that $X_t \sim \mu p_t$ and $Y_t \sim \nu p_t$, as long as one fixes the initial distributions μ,ν. In other words, the choice of the coupling scheme between W_t^X, W_t^Y is flexible in the proof of contractivity (2.5). Therefore, our goal is to find an appropriate coupling scheme between W_t^X, W_t^Y so that the contractivity (2.7) holds.

The simplest coupling scheme is $W_t^X = W_t^Y$, which is also known as the *synchronous coupling*.[12] The synchronous coupling can be used to prove the contractivity (2.7) when $b(x) = -\nabla U(x)$ and $U(x)$ is strongly convex. Unfortunately, the synchronous coupling cannot directly apply when $U(x)$ is not convex.

Another choice is the *reflection coupling*. In Ref. 12, the reflection coupling is used to prove the contractivity of the overdamped Langevin dynamics of a single particle. Later in Ref. 13, this approach is used to prove the contractivity of the product model (2.1). In this paper, we shall review the reflection coupling for the product model and extend the results to the IPS (1.1) and the RB–IPS (1.3).

Consider the coupling scheme for the product model (2.1) introduced in Ref. 13. In the product model (2.1), each pair of particles $\{(X_t^i, Y_t^i)\}_{t\geq 0}$ is evolved by

$$\begin{cases} \mathrm{d}X_t^i = b^i(X_t)\mathrm{d}t + \sigma(\mathrm{rc}(Z_t^i)\mathrm{d}W_t^i + \mathrm{sc}(Z_t^i)\mathrm{d}\tilde{W}_t^i), \\ \mathrm{d}Y_t^i = b^i(Y_t)\mathrm{d}t + \sigma(\mathrm{rc}(Z_t^i)(I - 2e_t^i(e_t^i)^{\mathrm{T}})\mathrm{d}W_t^i + \mathrm{sc}(Z_t^i)\mathrm{d}\tilde{W}_t^i), \end{cases} \tag{2.9}$$

where $Z_t^i = X_t^i - Y_t^i$, $e_t^i = Z_t^i/|Z_t^i|$, and $\{W_t^i\}_{i=1}^N, \{\tilde{W}_t^i\}_{i=1}^N$ are independent Wiener processes in $\mathbb{R}^d$. Besides, $\mathrm{rc}(z), \mathrm{sc}(z)$ are smooth functions satisfying

$$\mathrm{rc}^2(z) + \mathrm{sc}^2(z) = 1, \quad \forall\, z \in \mathbb{R}^d \tag{2.10}$$

with $\mathrm{rc}(z) = 0$ for $|z| \leq \delta/2$ and $\mathrm{rc}(z) = 1$ for $|z| \geq \delta$. Clearly, "rc" denotes the reflection coupling and "sc" denotes the synchronous coupling. For each $i \in \{1, \ldots, N\}$, the dynamics $X_t^i, Y_t^i \in \mathbb{R}^d$ are driven by the stochastic processes

$$W_t^{X,i} = \int_0^t \big(\mathrm{rc}(Z_s^i)\mathrm{d}W_s^i + \mathrm{sc}(Z_s^i)\mathrm{d}\tilde{W}_s^i\big), \tag{2.11}$$

$$W_t^{Y,i} = \int_0^t \big(\mathrm{rc}(Z_s^i)(I - 2e_s^i(e_s^i)^{\mathrm{T}})\mathrm{d}W_s^i + \mathrm{sc}(Z_s^i)\mathrm{d}\tilde{W}_s^i\big), \tag{2.12}$$

respectively. We present some intuitive explanations of the coupled dynamics (2.9):

(1) The coupled dynamics (2.9) is a mixture of the synchronous coupling $(\mathrm{d}\tilde{W}_t^i)$ and the reflection coupling $(\mathrm{d}W_t^i)$. The matrix $I - 2e_t^i(e_t^i)^{\mathrm{T}} \in \mathbb{R}^{d\times d}$ is the reflection transform with respect to the normal plane of e_t^i, which is the reason the $(\mathrm{d}W_t^i)$ part is called *reflection coupling*.
(2) By Levy's characterization,[30] the normalizing condition (2.10) ensures that both $W_t^{X,i}, W_t^{Y,i}$ are standard Wiener processes in $\mathbb{R}^d$. Therefore, both dynamics X_t, Y_t are weak solutions to the product model (2.1).
(3) $\delta > 0$ is a free parameter in the definition of the coupled dynamics (2.9). Since $Z_t^i = X_t^i - Y_t^i$ is the relative displacement between X_t^i, Y_t^i, we have:

    (a) When $|Z_t^i| \geq \delta$, $\mathrm{rc}(Z_t^i) \equiv 1$ and (2.9) is fully reflection coupling.
    (b) When $|Z_t^i| \leq \delta/2$, (2.9) degenerates to fully synchronous coupling.

    When δ is sufficiently small, we expect that $\mathrm{rc}(z)$ is close to the constant function 1 and thus the reflection coupling dominates the coupled dynamics (2.9).

Remark 2.1. In Ref. 12, the coupling scheme for a single particle is fully reflection coupling, i.e. $\mathrm{rc}(z) \equiv 1$. However, if we simply choose $\mathrm{rc}(z) \equiv 1$ in the product model (2.1), it is inconvenient to define the coupled dynamics after the occurrence of $Z_t^i = 0$. Also as indicated in Ref. 13, it is difficult to make the proof of contractivity rigorous when $\mathrm{rc}(z) \equiv 1$.

From the coupled dynamics (2.9), the displacement Z_t^i satisfies the SDE

$$\mathrm{d}Z_t^i = (b^i(X_t) - b^i(Y_t))\mathrm{d}t + 2\sigma\mathrm{rc}(Z_t^i)|Z_t^i|^{-1}Z_t^i\mathrm{d}B_t^i, \tag{2.13}$$

where B_t^i is the 1D Wiener process defined by

$$B_t^i = \int_0^t (e_s^i)^{\mathrm{T}}\mathrm{d}W_s^i. \tag{2.14}$$

The synchronous coupling $(\mathrm{d}\tilde{W}_t^i)$ vanishes in (2.13), and the diffusion coefficient $\sigma\mathrm{rc}(Z_t^i)$ comes from the reflection coupling $(\mathrm{d}W_t^i)$. In the following proof, the diffusion term $\sigma\mathrm{rc}(Z_t^i)$ is the main reason that brings the two particles X_t^i, Y_t^i together. Let $r_t^i = |Z_t^i|$, then r_t^i satisfies the SDE

$$\mathrm{d}r_t^i = (r_t^i)^{-1} Z_t^i \cdot (b^i(X_t) - b^i(Y_t))\mathrm{d}t + 2\sigma\mathrm{rc}(Z_t^i)\mathrm{d}B_t^i. \tag{2.15}$$

Choosing a distance function $f(r) \in C^2[0, +\infty)$, by Itô's formula, one obtains

$$\begin{aligned}\mathrm{d}f(r_t^i) = 2\sigma\mathrm{rc}(Z_t^i) f'(r_t^i)\mathrm{d}B_t^i + \big((r_t^i)^{-1} Z_t^i \cdot (b^i(X_t) - b^i(Y_t)) f'(r_t^i) \\ + 2\sigma^2\mathrm{rc}^2(Z_t^i) f''(r_t^i)\big)\mathrm{d}t,\end{aligned} \tag{2.16}$$

hence the rate of change for $\mathbb{E}[f(r_t^i)]$ is

$$\frac{\mathrm{d}}{\mathrm{d}t}\mathbb{E}[f(r_t^i)] = \mathbb{E}\big((r_t^i)^{-1} Z_t^i \cdot (b^i(X_t) - b^i(Y_t)) f'(r_t^i) + 2\sigma^2\mathrm{rc}^2(Z_t^i) f''(r_t^i)\big). \tag{2.17}$$

Now define the distance $\rho(\cdot,\cdot)$ between the systems $X_t, Y_t \in \mathbb{R}^{Nd}$ by

$$\rho(X_t, Y_t) := \frac{1}{N}\sum_{i=1}^N f(r_t^i), \tag{2.18}$$

then the rate of change for $\mathbb{E}[\rho(X_t, Y_t)]$ is completely given by (2.17), (2.18).

In order to prove $\mathbb{E}[\rho(X_t, Y_t)]$ has exponential decay in time, we impose some technical assumptions on the drift forces $\{b^i(x)\}_{i=1}^N$. The distance function $f(r)$ will also be chosen according to these assumptions. Since each $b^i(x) = b(x^i) + \gamma^i(x)$, we only need to consider the drift force $b(\cdot)$ and the perturbation $\gamma^i(\cdot)$.

For the drift force $b(\cdot): \mathbb{R}^d \to \mathbb{R}^d$, suppose there is a function $\kappa(r)$ satisfying

$$\kappa(r) \le \inf\left\{-\frac{2}{\sigma^2}\frac{(x-y)\cdot(b(x)-b(y))}{|x-y|^2} : x, y \in \mathbb{R}^d, |x-y| = r\right\}. \tag{2.19}$$

Roughly speaking, when $b(x) = -\nabla U(x)$, the function $\kappa(r)$ depicts the convexity of the potential function $U(x)$. If the Hessian $\nabla^2 U(x)$ stays positive definite outside a finite spherical region, then $\kappa(r)$ is positive for sufficiently large r. Therefore, it is reasonable to require the asymptotic positivity of $\kappa(r)$.

Assumption 2.1. The function $\kappa(r)$ defined in (2.19) satisfies

(1) $\kappa(r)$ is continuous for $r \in (0, +\infty)$;
(2) $\kappa(r)$ has a lower bound for $r \in (0, +\infty)$;
(3) $\underline{\lim}_{r\to\infty} \kappa(r) > 0$.

For the perturbation $\gamma^i(\cdot): \mathbb{R}^{Nd} \to \mathbb{R}^d$, assume the Lipschitz condition holds.

Assumption 2.2. There exists a constant L such that

$$\sum_{i=1}^N |\gamma^i(x) - \gamma^i(y)| \le L\sum_{i=1}^N |x^i - y^i|, \quad \forall\, x, y \in \mathbb{R}^{Nd}. \tag{2.20}$$

Remark 2.2. If a non-continuous function $\kappa(r)$ satisfies (2.19) and the latter two conditions in Assumption 2.1, we can find another continuous function $\bar{\kappa}(r) \le \kappa(r)$ which satisfies all conditions in Assumption 2.1. Therefore, the continuity of $\kappa(r)$ is not an essential condition in Assumption 2.1. We assume the continuity of $\kappa(r)$ merely for technical convenience.

Remark 2.3. Assumption 2.1 can also be interpreted as the dissipation condition. In fact, the asymptotic positivity of $\kappa(r)$ implies there exist $A, B > 0$ such that

$$-x \cdot b(x) \ge A|x|^2 - B, \tag{2.21}$$

which is commonly adopted in the study of geometric ergodicity.[14, 20]

Following Refs. 12 and 13, we choose the distance function $f(r)$ according to the following lemma.

Lemma 2.1. *If the function* $\kappa(r)$ *defined in* (2.19) *satisfies Assumption* 2.1, *then there exists a function* $f(r)$ *defined in* $r \in [0, +\infty)$ *such that*

(1) $f(0) = 0$, *and* $f(r)$ *is concave and strictly increasing in* $[0, +\infty)$;
(2) $f(r) \in C^2[0, +\infty)$ *and there exists a constant* $c_0 > 0$ *such that*

$$f''(r) - \frac{1}{4} r\kappa(r) f'(r) \le -\frac{c_0}{2} f(r), \quad \forall\, r \ge 0. \tag{2.22}$$

(3) *There exists a constant* $\varphi_0 > 0$ *such that*

$$\frac{\varphi_0}{4} r \le f(r) \le r, \quad \forall\, r \ge 0. \tag{2.23}$$

The constants c_0, φ_0 *only depend on the function* $\kappa(r)$.

The proof of Lemma 2.1 is in Appendix A.

The positivity of c_0 in (2.22) is essential in the proof of contractivity of the product model (2.1). Figure 1 is an example of the distance function $f(r)$ in the case $\kappa(r) = \max\{r/2\sqrt{2} - 1, 1\}$. The graphs of $\kappa(r)$ and $f(r)$ are shown.

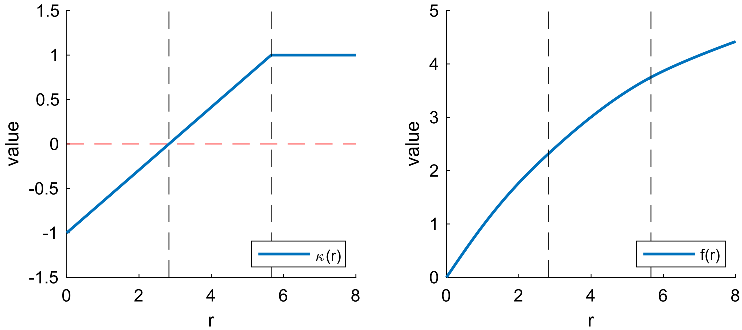

Fig. 1. (Color online) Graphs of $\kappa(r)$ (left) and $f(r)$ (right), where $f(r)$ is defined according to Lemma 2.1.

In Fig. 1, we observe that the distance function $f(r)$ is concave for small r and almost linear for large r. Here is an intuitive explanation how this feature of $f(r)$ is related to the inequality (2.22). When $r_t^i = |X_t^i - Y_t^i|$ is large, the particles X_t^i, Y_t^i are attracted together due to the dissipation of the drift force $b(x)$. When r_t^i is small, the concavity of $f(r)$ makes the quantity $f(r_t^i)$ more sensitive to the decreasing rather than increasing of the relative distance r_t^i, and thus we can expect the decreasing of $\mathbb{E}[f(r_t^i)]$ even without the global convexity.

Using the distance function $f(r)$ defined in Lemma 2.1, we are able to estimate the rate of change for $\mathbb{E}[\rho(X_t, Y_t)]$. The following lemma is a key step to derive the estimation.

Lemma 2.2. *Under Assumptions* 2.1 *and* 2.2, *let $f(r)$ be the distance function given in Lemma* 2.1. *Given $\delta > 0$, let $\mathrm{rc}(z)$ be a smooth continuous function with $|\mathrm{rc}(z)| \leq 1$ and $\mathrm{rc}(z) = 1$ for $|z| \geq \delta$. If the constant L in Assumption* 2.2 *satisfies*

$$L < \frac{c_0 \varphi_0 \sigma^2}{8},$$

then the following inequality holds with $c := c_0\sigma^2/2$,

$$\begin{aligned}\sum_{i=1}^N &\left((r^i)^{-1} Z^i \cdot (b^i(X) - b^i(Y)) f'(r^i) + 2\sigma^2 \mathrm{rc}^2(Z^i) f''(r^i)\right) \\ &\leq N m(\delta) - c \sum_{i=1}^N f(r^i),\end{aligned} \tag{2.24}$$

where $X, Y \in \mathbb{R}^{Nd}$, $Z = X - Y$, $r^i = |Z^i|$ and $m(\delta)$ is defined by

$$m(\delta) = \frac{\sigma^2}{2} \sup_{r<\delta} (r\kappa(r)^-) + c_0 \sigma^2 \delta. \tag{2.25}$$

Here $x^- = -\min\{x, 0\}$ denotes the negative part of $x \in \mathbb{R}$.

The proof of Lemma 2.2 is in Appendix A, and is similar to the proof of Theorem 7 of Ref. 13. (2.24) is also the condition of Lemma 5 of Ref. 13.

Remark 2.4. We have some remarks on Lemma 2.2.

(1) δ and $\mathrm{rc}(z)$ in Lemma 2.2 correspond to parameters in the coupled dynamics (2.9). The additional term $m(\delta)$ appears in the right-hand side of (2.24) because $\mathrm{rc}(z)$ is not identical to 1, i.e. we are not using the fully reflection coupling. By Assumption 2.1, $\kappa(r)^-$ is bounded for $r \in (0, +\infty)$, and thus $\lim_{\delta\to 0} m(\delta) = 0$.
(2) The distance function $f(r)$, the upper bound of L and the contraction rate c are all independent of δ, thus we may pass δ to the limit 0 without changing the value of c.
(3) Compared to (2.22), the left-hand side of (2.24) involves both the drift force $b(x^i)$ and the perturbation $\gamma^i(x)$. According to Lemma 2.1, the dissipation of

the drift force is essential to produce the negative coefficient $-c_0$ in the right-hand side of (2.22). Therefore, the perturbation should be moderately small so that the coefficient in (2.24) is still negative (see (A.21) in the proof of Lemma 2.2). In fact, if the perturbation is too large, the convergence rate of the IPS (1.1) can depend on N, and the mean-field dynamics can have multiple invariant distributions.[11]

Using Lemma 2.2, we can obtain the contractivity of the coupled dynamics (2.9).

Lemma 2.3. *Under Assumptions* 2.1 *and* 2.2, *let $f(r)$ be the distance function defined in Lemma* 2.1, *and $c := c_0\sigma^2/2$. If the constant L in Assumption* 2.2 *satisfies*

$$L < \frac{c_0\varphi_0\sigma^2}{8},$$

then for $\rho_t := \rho(X_t, Y_t)$ defined in (2.18), *one has*

$$\frac{\mathrm{d}}{\mathrm{d}t}\mathbb{E}[\rho_t] \leq m(\delta) - c\cdot\mathbb{E}[\rho_t], \quad \forall\, t \geq 0, \tag{2.26}$$

where $m(\delta)$ is defined in (2.25).

Proof. Since $\rho_t = \sum_{i=1}^N f(r_t^i)/N$, one has

$$\frac{\mathrm{d}}{\mathrm{d}t}\mathbb{E}[\rho_t] = \frac{1}{N}\sum_{i=1}^N \frac{\mathrm{d}}{\mathrm{d}t}\mathbb{E}[f(r_t^i)]. \tag{2.27}$$

Using (2.17), one obtains

$$\frac{\mathrm{d}}{\mathrm{d}t}\mathbb{E}[\rho_t] = \frac{1}{N}\sum_{i=1}^N \mathbb{E}\big((r_t^i)^{-1}Z_t^i\cdot(b^i(X_t) - b^i(Y_t))f'(r_t^i) + 2\sigma^2\mathrm{rc}^2(Z_t^i)f''(r_t^i)\big). \tag{2.28}$$

Applying the estimate in Lemma 2.2, one obtains

$$\frac{\mathrm{d}}{\mathrm{d}t}\mathbb{E}[\rho_t] \leq m(\delta) - c\cdot\mathbb{E}[\rho_t], \tag{2.29}$$

which is exactly the desired result. □

Integrating (2.26) in the time interval $[0, t)$ gives

$$\mathbb{E}[\rho_t] \leq e^{-ct}\mathbb{E}[\rho_0] + \frac{m(\delta)(1 - e^{-ct})}{c}, \quad \forall\, t \geq 0, \tag{2.30}$$

which can be used to derive the contractivity for the probability distributions.

To describe the probability distributions precisely, introduce the following terminologies. Let $\mathcal{P}_1$ be the set of probability distributions in $\mathbb{R}^{Nd}$ with finite first-order moments, i.e.

$$\mathcal{P}_1 = \left\{\mu \text{ is a probability distribution in } \mathbb{R}^{Nd} : \sum_{i=1}^N \int_{\mathbb{R}^{Nd}} |x^i|\mu(\mathrm{d}x) < +\infty\right\}. \tag{2.31}$$

For probability distributions $\mu,\nu\in\mathcal{P}_1$, define the normalized Wasserstein distances

$$\mathcal{W}_1(\mu,\nu)=\inf_{\gamma\in\Pi(\mu,\nu)}\int_{\mathbb{R}^{Nd}\times\mathbb{R}^{Nd}}\left(\frac{1}{N}\sum_{i=1}^N|x^i-y^i|\right)\gamma(\mathrm{d}x\mathrm{d}y), \tag{2.32}$$

$$\mathcal{W}_f(\mu,\nu)=\inf_{\gamma\in\Pi(\mu,\nu)}\int_{\mathbb{R}^{Nd}\times\mathbb{R}^{Nd}}\left(\frac{1}{N}\sum_{i=1}^N f(|x^i-y^i|)\right)\gamma(\mathrm{d}x\mathrm{d}y). \tag{2.33}$$

Then $(\mathcal{P}_1,\mathcal{W}_1(\cdot,\cdot))$ is a complete metric space (see Theorem 6.18 of Ref. 39). Note that $f(r)$ does not satisfy the triangle inequality, $\mathcal{W}_f$ is a semimetric rather than a metric. Nevertheless, Lemma 2.1 implies $\mathcal{W}_1$ and $\mathcal{W}_f$ that are equivalent in the sense

$$\frac{\varphi_0}{4}\mathcal{W}_1(\mu,\nu)\le\mathcal{W}_f(\mu,\nu)\le\mathcal{W}_1(\mu,\nu),\quad\forall\,\mu,\nu\in\mathcal{P}(\mathbb{R}^{Nd}). \tag{2.34}$$

Using the estimate (2.30), one obtains the following result.

Theorem 2.1. *Under Assumptions* 2.1 *and* 2.2, *let $f(r)$ be the distance function defined in Lemma* 2.1, *and $c:=c_0\sigma^2/2$. Let p_t be the transition kernel of the product model* (2.1). *If the Lipschitz constant L in Assumption* 2.2 *satisfies*

$$L<\frac{c_0\varphi_0\sigma^2}{8},$$

then one has

$$\mathcal{W}_f(\mu p_t,\nu p_t)\le e^{-ct}\mathcal{W}_f(\mu,\nu),\quad\forall\,t\ge 0 \tag{2.35}$$

for any probability distributions $\mu,\nu\in\mathcal{P}_1$.

The proof of Theorem 2.1 is similar to the proof of Theorem 7 of Ref. 13.

Proof. For given distributions $\mu,\nu\in\mathcal{P}_1$, let $\gamma\in\Pi(\mu,\nu)$ satisfies

$$\int_{\mathbb{R}^{Nd}}\left(\frac{1}{N}\sum_{i=1}^N f(|x^i-y^i|)\right)\gamma(\mathrm{d}x\mathrm{d}y)\le\mathcal{W}_f(\mu,\nu)+\varepsilon, \tag{2.36}$$

where $\varepsilon>0$ is an arbitrary small constant. Let $\{(X_t,Y_t)\}_{t\ge0}$ be evolved by the coupled dynamics (2.9) with the initial value $(X_0,Y_0)\sim\gamma$, then $X_t\sim\mu p_t$ and $Y_t\sim\nu p_t$. From the inequality (2.30) one obtains

$$\begin{aligned}\mathbb{E}[\rho(X_t,Y_t)]&\le e^{-ct}\mathbb{E}[\rho(X_0,Y_0)]+\frac{m(\delta)(1-e^{-ct})}{c}\\&\le e^{-ct}\mathcal{W}_f(\mu,\nu)+\frac{m(\delta)(1-e^{-ct})}{c}+\varepsilon.\end{aligned} \tag{2.37}$$

Using the definition of $\mathcal{W}_f$,

$$\mathbb{E}[\rho(X_t,Y_t)]\ge\inf_{\gamma\in\Pi(\mu p_t,\nu p_t)}\int\left(\frac{1}{N}\sum_{i=1}^N f(|x^i-y^i|)\right)\gamma(\mathrm{d}x\mathrm{d}y)=\mathcal{W}_f(\mu p_t,\nu p_t), \tag{2.38}$$

hence one obtains

$$\mathcal{W}_f(\mu p_t, \nu p_t) \le e^{-ct}\mathcal{W}_f(\mu,\nu) + \frac{m(\delta)(1-e^{-ct})}{c} + \varepsilon. \tag{2.39}$$

Note that the evolution of μp_t and νp_t does not depend on the coupling scheme, we can directly pass δ and ε to 0 and obtain

$$\mathcal{W}_f(\mu p_t, \nu p_t) \le e^{-ct}\mathcal{W}_f(\mu,\nu), \tag{2.40}$$

which is exactly the contractivity we need. □

2.2. *Exact dynamics*: *IPS*

We apply Theorem 2.1 to derive the geometric ergodicity of the IPS (1.1). For the IPS (1.1), the perturbation $\gamma^i(\cdot):\mathbb{R}^{Nd}\to\mathbb{R}^d$ is given by (2.3). Suppose L_K is the Lipschitz constant of the interaction $K(\cdot):\mathbb{R}^d\to\mathbb{R}^d$, then for any $x,y\in\mathbb{R}^{Nd}$,

$$\begin{aligned}|\gamma^i(x)-\gamma^i(y)| &\le \frac{1}{N-1}\sum_{j\neq i}|K(x^i-x^j)-K(y^i-y^j)| \\ &\le \frac{L_K}{N-1}\sum_{j\neq i}\big(|x^i-y^i|+|x^j-y^j|\big).\end{aligned}$$

Summation over $i\in\{1,\ldots,N\}$ gives

$$\sum_{i=1}^N|\gamma^i(x)-\gamma^i(y)| \le 2L_K\sum_{i=1}^N|x^i-y^i|. \tag{2.41}$$

Hence Assumption 2.2 holds with the constant $L=2L_K$. In terms of the interaction force $K(\cdot)$, we may replace Assumption 2.2 by the following one.

Assumption 2.3. There exists a constant L_K such that

$$\max\{|K(x)|,|\nabla K(x)|,|\nabla^2K(x)|\}\le L_K,\quad \forall\, x\in\mathbb{R}^d. \tag{2.42}$$

Remark 2.5. Assumption 2.3 is stronger than Assumption 2.2 because we require not only $\nabla K(\cdot)$ but also $K(\cdot)$ and $\nabla^2K(\cdot)$ to be uniformly bounded. The boundedness of $K(\cdot)$ and $\nabla^2K(\cdot)$ is not necessary to prove the geometric ergodicity, but will be useful in the strong error estimation in Sec. 3.

For completeness, we explicitly write the coupling scheme for the IPS (1.1). The coupled dynamics $\{(X_t,Y_t)\}_{t\ge0}$ in $\mathbb{R}^{Nd}\times\mathbb{R}^{Nd}$ is given by

$$\begin{cases}\mathrm{d}X_t^i = b(X_t^i)\mathrm{d}t + \dfrac{1}{N-1}\displaystyle\sum_{j\neq i}K(X_t^i-X_t^j)\mathrm{d}t \\ \qquad\quad +\sigma\big(\mathrm{rc}(Z_t^i)\mathrm{d}W_t^i+\mathrm{sc}(Z_t^i)\mathrm{d}\tilde{W}_t^i\big), \\ \mathrm{d}Y_t^i = b^i(Y_t)\mathrm{d}t + \dfrac{1}{N-1}\displaystyle\sum_{j\neq i}K(Y_t^i-Y_t^j)\mathrm{d}t \\ \qquad\quad +\sigma\big(\mathrm{rc}(Z_t^i)(I-2e_t^i(e_t^i)^{\mathrm{T}})\mathrm{d}W_t^i+\mathrm{sc}(Z_t^i)\mathrm{d}\tilde{W}_t^i\big),\end{cases} \tag{2.43}$$

for $i=1,\ldots,N$. Theorem 2.1 then immediately implies the following.

Theorem 2.2. *Under Assumptions* 2.1 *and* 2.3, *let* $f(r)$ *be the distance function defined in Lemma* 2.1, *and* $c := c_0\sigma^2/2$. *Let* p_t *be the transition kernel of the IPS* (1.1). *If the constant* L_K *in Assumption* 2.3 *satisfies*

$$L_K < \frac{c_0\varphi_0\sigma^2}{16},$$

then we have

$$\mathcal{W}_f(\mu p_t, \nu p_t) \le e^{-ct}\mathcal{W}_f(\mu,\nu), \quad \forall\, t \ge 0 \tag{2.44}$$

for any probability distributions $\mu, \nu \in \mathcal{P}_1$.

Theorem 2.2 is similar to Corollary 9 of Ref. 13. An important observation from Theorem 2.2 is that both the contraction rate c and the upper bound of L_K do not depend on the number of particles N. Also, for any initial distribution $\nu \in \mathcal{P}_1$, νp_t converges to the invariant distribution π exponentially.

Corollary 2.1. *Under Assumption* 2.1 *and* 2.3, *let* $f(r)$ *be the distance function defined in Lemma* 2.1, *and* $c := c_0\sigma^2/2$. *Let* p_t *be the transition kernel of the IPS* (1.1), *and* $\pi \in \mathcal{P}_1$ *be the invariant distribution. If the constant* L_K *in Assumption* 2.3 *satisfies*

$$L_K < \frac{c_0\varphi_0\sigma^2}{16},$$

then one has

$$\mathcal{W}_f(\nu p_t, \pi) \le e^{-ct}\mathcal{W}_f(\nu,\pi), \quad \forall\, t \ge 0 \tag{2.45}$$

for any probability distribution $\nu \in \mathcal{P}_1$.

The existence of the invariant distribution π will be later proved in Theorem 3.1.

2.3. *Random batch dynamics*: *RB–IPS*

We prove the geometric ergodicity of the RB–IPS (1.3) using the reflection coupling. Unfortunately, Theorem 2.2 cannot be directly applied since the perturbation $\gamma^i(x)$ changes its expression in different time steps. In the following, the proof of the contractivity for the RB–IPS (1.3) will be mainly based on Lemma 2.3. Also, it is necessary to clarify the coupled dynamics for the RB–IPS (1.3).

Suppose at the instant t_n, the division $\mathcal{D}_n = \{\mathcal{C}_1, \ldots, \mathcal{C}_q\}$ is randomly generated, then the perturbation $\gamma^i(x)$ within the time interval $[t_n, t_{n+1})$ is given by (2.4). It is easy to verify

$$\begin{aligned}
|\gamma^i(x) - \gamma^i(y)| &\le \frac{1}{p-1}\sum_{j\ne i, j\in\mathcal{C}} |K(x^i - x^j) - K(y^i - y^j)| \\
&\le \frac{L_K}{p-1}\sum_{j\ne i, j\in\mathcal{C}} (|x^i - y^i| + |x^j - y^j|),
\end{aligned}$$

where $\mathcal{C} \in \mathcal{D}_n$ is the batch that contains i. Summation over $i \in \mathcal{C}$ gives

$$\sum_{i\in\mathcal{C}} |\gamma^i(x) - \gamma^i(y)| \leq 2L_K \sum_{i\in\mathcal{C}} |x^i - y^i|. \tag{2.46}$$

Summation over $\mathcal{C} \in \{\mathcal{C}_1, \dots, \mathcal{C}_q\}$ gives

$$\sum_{i=1}^N |\gamma^i(x) - \gamma^i(y)| \leq 2L_K \sum_{i=1}^N |x^i - y^i|. \tag{2.47}$$

Hence Assumption 2.3 still holds with $L = 2L_K$. In a similar way, define the coupled dynamics for the RB–IPS (1.3) as follows.

Fix the parameter $\delta > 0$ and let the smooth functions $\mathrm{rc}(z), \mathrm{sc}(z)$ be defined as in (2.10). At each time step t_n, suppose the division $\mathcal{D}_n$ is randomly generated, and the coupled dynamics $\{(\tilde{X}_t, \tilde{Y}_t)\}_{t\geq 0}$ in $\mathbb{R}^{Nd} \times \mathbb{R}^{Nd}$ within the time interval $[t_n, t_{n+1})$ is defined by

$$\begin{cases} \mathrm{d}\tilde{X}_t^i = b(\tilde{X}_t^i)\mathrm{d}t + \dfrac{1}{p-1} \displaystyle\sum_{j\neq i, j\in\mathcal{C}} K(\tilde{X}_t^i - \tilde{X}_t^j)\mathrm{d}t \\ \qquad\quad + \sigma\big(\mathrm{rc}(\tilde{Z}_t^i)\mathrm{d}W_t^i + \mathrm{sc}(\tilde{Z}_t^i)\mathrm{d}\tilde{W}_t^i\big), \\ \mathrm{d}\tilde{Y}_t^i = b^i(\tilde{Y}_t)\mathrm{d}t + \dfrac{1}{p-1} \displaystyle\sum_{j\neq i, j\in\mathcal{C}} K(\tilde{Y}_t^i - \tilde{Y}_t^j)\mathrm{d}t \\ \qquad\quad + \sigma\big(\mathrm{rc}(\tilde{Z}_t^i)(I - 2e_t^i(e_t^i)^{\mathrm{T}})\mathrm{d}W_t^i + \mathrm{sc}(\tilde{Z}_t^i)\mathrm{d}\tilde{W}_t^i\big), \end{cases} \tag{2.48}$$

for $i \in \mathcal{C}$ and $\mathcal{C} \in \mathcal{D}_n$, where $\tilde{Z}_t^i = \tilde{X}_t^i - \tilde{Y}_t^i$ and $e_i = \tilde{Z}_t^i/|\tilde{Z}_t^i|$. For convenience, define the filtration of the coupled dynamics (2.48) by

$$\mathcal{G}_n = \sigma((\tilde{X}_0, \tilde{Y}_0), \{W_s\}_{0\leq s\leq t_n}, \{\tilde{W}_s\}_{0\leq s\leq t_n}, \{\mathcal{D}_k\}_{0\leq k\leq n}). \tag{2.49}$$

That is, $\mathcal{G}_n$ is determined by the joint distribution of $(\tilde{X}_0, \tilde{Y}_0)$ in $\mathbb{R}^{Nd} \times \mathbb{R}^{Nd}$, Wiener processes $W_t, \tilde{W}_t$ before t_n, and the batch divisions in the first $n+1$ time steps. Under the condition of $\mathcal{G}_n$, the division $\mathcal{D}_n$ within the time step $[t_n, t_{n+1})$ is determined, and the coupled dynamics of $(\tilde{X}_t, \tilde{Y}_t)$ is exactly given by (2.48).

We still choose the distance function $f(r)$ according to Lemma 2.1, and the distance between $\tilde{X}_t, \tilde{Y}_t \in \mathbb{R}^{Nd}$ is defined by

$$\rho(\tilde{X}_t, \tilde{Y}_t) = \frac{1}{N}\sum_{i=1}^N f(\tilde{r}_t^i), \tag{2.50}$$

where $\tilde{r}_t^i = |\tilde{Z}_t^i|$. Similar to Lemma 2.3, we may derive the contractivity for the coupled dynamics (2.48), but only in the time interval $[t_n, t_{n+1})$ and under the condition of fixed $\mathcal{G}_n$.

Corollary 2.2. *Under Assumptions* 2.1 *and* 2.3, *let $f(r)$ be the distance function in Lemma* 2.1, *and $c := c_0\sigma^2/2$. If the constant L_K in Assumption* 2.3 *satisfies*

$$L_K < \frac{c_0\varphi_0\sigma^2}{16},$$

then under the condition of fixed $\mathcal{G}_n$, *for* $\tilde{\rho}_t := \rho(\tilde{X}_t, \tilde{Y}_t)$ *defined in* (2.50), *one has*

$$\frac{\mathrm{d}}{\mathrm{d}t}\mathbb{E}[\tilde{\rho}_t|\mathcal{G}_n] \leq m(\delta) - c \cdot \mathbb{E}[\tilde{\rho}_t|\mathcal{G}_n], \quad t \in [t_n, t_{n+1}). \tag{2.51}$$

Here, the expectation $\mathbb{E}[\tilde{\rho}_t|\mathcal{G}_n]$ only involves the Wiener processes $W_t, \tilde{W}_t$ in the time interval $[t_n, t_{n+1})$. The trajectories of $\{(\tilde{X}_t, \tilde{Y}_t)\}_{0\leq t\leq t_n}$ and the batch divisions $\{D_k\}_{0\leq k\leq n}$ have been included in the filtration $\mathcal{G}_n$ in (2.49). Corollary 2.2 can be directly derived from Lemma 2.3 since Assumption 2.2 holds with $L = 2L_K$. Taking the expectation over the filtration $\mathcal{G}_n$, one obtains

$$\frac{\mathrm{d}}{\mathrm{d}t}\mathbb{E}[\tilde{\rho}_t] \leq m(\delta) - c \cdot \mathbb{E}[\tilde{\rho}_t], \quad t \in [t_n, t_{n+1}). \tag{2.52}$$

Integrating (2.52) in the time interval $[t_n, t_{n+1})$ gives

$$\mathbb{E}[\tilde{\rho}_{(n+1)\tau}] \leq e^{-c\tau}\mathbb{E}[\tilde{\rho}_{n\tau}] + \frac{m(\delta)(1-e^{-c\tau})}{c}, \quad \forall n \geq 0. \tag{2.53}$$

Induction on (2.53) for the first n time steps gives

$$\mathbb{E}[\tilde{\rho}_{n\tau}] \leq e^{-cn\tau}\mathbb{E}[\tilde{\rho}_0] + \frac{m(\delta)(1-e^{-cn\tau})}{c}, \quad \forall n \geq 0. \tag{2.54}$$

Let $\tilde{p}_t$ be the transition kernel of the RB–IPS (1.3). Given the probability distributions $\mu, \nu \in \mathcal{P}_1$, suppose the initial values $\tilde{X}_0 \sim \mu, \tilde{Y}_0 \sim \nu$, then $\tilde{X}_{n\tau} \sim \mu\tilde{p}_{n\tau}, \tilde{Y}_{n\tau} \sim \nu\tilde{p}_{n\tau}$. Clearly, (2.54) implies

$$\mathcal{W}_f(\mu\tilde{p}_{n\tau}, \nu\tilde{p}_{n\tau}) \leq e^{-cn\tau}\mathcal{W}_f(\mu,\nu) + \frac{m(\delta)(1-e^{-nc\tau})}{c}, \quad \forall n \geq 0. \tag{2.55}$$

A crucial observation of (2.55) is that the evolution of the distributions $\{\mu\tilde{p}_{n\tau}\}_{n\geq 0}$ and $\{\nu\tilde{p}_{n\tau}\}_{n\geq 0}$ does not depend on the coupling scheme, in particular, the free parameter $\delta > 0$. Therefore, one may pass the limit $\delta \to 0$ in (2.55) to obtain

$$\mathcal{W}_f(\mu\tilde{p}_{n\tau}, \nu\tilde{p}_{n\tau}) \leq e^{-cn\tau}\mathcal{W}_f(\mu,\nu), \quad \forall n \geq 0. \tag{2.56}$$

Concluding the deduction above, we obtain the following.

Theorem 2.3. *Under Assumptions* 2.1 *and* 2.3, *let* $f(r)$ *be the distance function defined in Lemma* 2.1, *and* $c := c_0\sigma^2/2$. *Let* $\tilde{p}_t$ *be the transition kernel of the RB–IPS* (1.3). *If the constant* L_K *in Assumption* 2.3 *satisfies*

$$L_K < \frac{c_0\varphi_0\sigma^2}{16}, \tag{2.57}$$

then one has

$$\mathcal{W}_f(\mu\tilde{p}_{n\tau}, \nu\tilde{p}_{n\tau}) \leq e^{-cn\tau}\mathcal{W}_f(\mu,\nu), \quad \forall n \geq 0 \tag{2.58}$$

for any probability distributions $\mu, \nu \in \mathcal{P}_1$.

Theorem 2.3 is a random batch version of Theorem 2.2. The contraction rate c is a constant of order 1 and does not depend on the number of particles N, the batch size p or the time step τ.

Remark 2.6. The continuous-time dynamics RB–IPS $\{\tilde{X}_t\}_{t\geq 0}$ is not a time-homogeneous Markov process, since the random divisions are determined at different time steps. However, $\{\tilde{X}_{n\tau}\}_{n\geq 0}$ is a time-homogeneous Markov chain, and the transition kernels $\{\tilde{p}_{n\tau}\}_{n\geq 0}$ forms a semi-group, i.e. for any probability distribution $\nu \in \mathcal{P}(\mathbb{R}^{Nd})$ and integers $n, m \geq 0$, we have $\nu\tilde{p}_{(n+m)\tau} = (\nu p_{n\tau})\tilde{p}_{m\tau}$.

Similar to Corollary 2.1, we can prove that for any initial distribution $\nu \in \mathcal{P}_1$, $\nu\tilde{p}_{n\tau}$ converges to the invariant distribution $\tilde{\pi}$ exponentially.

Corollary 2.3. *Under Assumptions* 2.1 *and* 2.3, *let $f(r)$ be the distance function defined in Lemma* 2.1, *and $c := c_0\sigma^2/2$. Let $\tilde{p}_t$ be the transition kernel of the RB–IPS* (1.3), *and $\tilde{\pi} \in \mathcal{P}_1$ be the invariant distribution. If the constant L_K in Assumption* 2.3 *satisfies*

$$L_K < \frac{c_0\varphi_0\sigma^2}{16},$$

then one has

$$\mathcal{W}_f(\nu\tilde{p}_{n\tau}, \tilde{\pi}) \leq e^{-cn\tau}\mathcal{W}_f(\nu, \tilde{\pi}), \quad \forall n \geq 0 \tag{2.59}$$

for any probability distribution $\nu \in \mathcal{P}_1$.

The existence of the invariant distribution $\tilde{\pi}$ will be later proved in Theorem 3.1.

3. Error Estimation of Invariant Distributions

In this section, we measure the difference between the invariant distributions $\pi, \tilde{\pi}$ of the IPS (1.1) and the RB–IPS (1.3). We shall prove the following results:

(1) **Existence of invariant distributions.** The IPS (1.1) has an invariant distribution $\pi \in \mathcal{P}_1$, and the RB–IPS (1.3) has an invariant distribution $\tilde{\pi} \in \mathcal{P}_1$. This is a direct corollary of the geometric ergodicity proved in Sec. 2 using the Banach fixed point theorem.
(2) **Strong error estimation in finite time.** Using the strong error estimation,[26] for given initial distribution ν, the distance between νp_t and $\nu\tilde{p}_t$ can be bounded by $O(\tau^{\frac{1}{2}})$, where $p_t, \tilde{p}_t$ are the transition kernels of the IPS (1.1) and the RB–IPS (1.3), respectively.
(3) **Error estimation of invariant distributions.** Combining the geometric ergodicity and the strong error estimation in finite time, we are able to estimate the difference between the invariant distributions $\pi, \tilde{\pi}$, using the triangle inequality described in the Introduction.

3.1. *Characterization of invariant distributions*

We prove the existence of the invariant distributions for the IPS (1.1) and the RB–IPS (1.3) and estimate their first-order moments. The proof is accomplished by the Banach fixed point theorem on the space $\mathcal{P}_1$ of probability distributions, where we have defined in (2.31). Such strategy has previously appeared in Ref. 13, which proves the existence of the invariant distribution π of the IPS (1.1). We extend this strategy to prove the existence of invariant distribution $\tilde{\pi}$ of the RB–IPS (1.3).

To begin with, we show that the distributions νp_t and $\nu \tilde{p}_t$ always have finite first-order moments.

Lemma 3.1. (moment) *Under Assumptions* 2.1 *and* 2.3, *there exists a constant* D *such that if the constant* L_K *in Assumption* 2.3 *satisfies*

$$L_K < \frac{c_0 \varphi_0 \sigma^2}{16},$$

then for any probability distribution $\nu \in \mathcal{P}_1$,

(i) $\nu p_t \in \mathcal{P}_1$ *for any real number* $t \geq 0$, *and*

$$\varlimsup_{t\to\infty} \int_{\mathbb{R}^{Nd}} \left(\frac{1}{N} \sum_{i=1}^N |x^i| \right) (\nu p_t)(\mathrm{d}x) \leq D; \tag{3.1}$$

(ii) $\nu \tilde{p}_{n\tau} \in \mathcal{P}_1$ *for any integer* $n \geq 0$, *and*

$$\varlimsup_{n\to\infty} \int_{\mathbb{R}^{Nd}} \left(\frac{1}{N} \sum_{i=1}^N |x^i| \right) (\nu \tilde{p}_{n\tau})(\mathrm{d}x) \leq D. \tag{3.2}$$

The constant D *does not depend on the number of particles* N, *the time step* τ, *the batch size* p *or the initial distribution* ν.

The proof of Lemma 3.1 is in Appendix A. The asymptotic positivity of the function $\kappa(r)$ in Assumption 2.1 is crucial to bound the moments of νp_t and $\nu \tilde{p}_t$ uniformly in time.

Remark 3.1. As we shall see in strong error estimation, we can also obtain the αth order moment estimation which is uniform in time for a general constant $\alpha \geq 2$.

Using the contractivity obtained in Sec. 2, we derive the existence of the invariant distributions.

Theorem 3.1. *Under Assumptions* 2.1 *and* 2.3, *if the constant* L_K *in Assumption* 2.3 *satisfies*

$$L_K < \frac{c_0 \varphi_0 \sigma^2}{16},$$

then

(i) *The Markov process* $\{X_t\}_{t\geq 0}$ *evolved by the IPS* (1.1) *has a unique invariant distribution* $\pi \in \mathcal{P}_1$.

(ii) *The Markov chain* $\{\tilde{X}_{n\tau}\}_{n\geq 0}$ *evolved by the RB–IPS* (1.3) *has a unique invariant distribution* $\tilde{\pi} \in \mathcal{P}_1$.

The proof below is similar to the proof of Corollary 3 of Ref. 13.

Proof. (i) Note that the Wasserstein distance $\mathcal{W}_f$ is equivalent to the standard $\mathcal{W}_1$-distance

$$\mathcal{W}_1(\mu,\nu) = \inf_{\gamma\in\Pi(\mu,\nu)} \int_{\mathbb{R}^{Nd}\times\mathbb{R}^{Nd}} \left(\frac{1}{N}\sum_{i=1}^N |x^i - y^i|\right)\gamma(\mathrm{d}x\mathrm{d}y). \tag{3.3}$$

From Theorem 2.2, there exists a constant $C > 0$ such that

$$\mathcal{W}_1(\mu p_t, \nu p_t) \leq Ce^{-ct}\cdot \mathcal{W}_1(\mu,\nu), \quad \forall\, t\geq 0 \tag{3.4}$$

for all distributions $\mu,\nu\in\mathcal{P}_1$. Then there exists $T>0$ such that $Ce^{-cT} = 1/2$ and

$$\mathcal{W}_1(\mu p_T, \nu p_T) \leq \frac{1}{2}\mathcal{W}_1(\mu,\nu). \tag{3.5}$$

Hence the mapping $\nu \mapsto \nu p_T$ is contractive in the complete metric space $\mathcal{P}_1$. From the Banach fixed point theorem, this mapping has a fixed point $\pi_0 \in \mathcal{P}_1$, i.e.

$$\pi_0 = \pi_0 p_T. \tag{3.6}$$

Define the measure in $\mathbb{R}^{Nd}$ by

$$\pi = \frac{1}{T}\int_0^T \pi_0 p_s \mathrm{d}s, \tag{3.7}$$

then π is a probability distribution in $\mathbb{R}^{Nd}$ and $\pi\in\mathcal{P}_1$ from Lemma 3.1. From the Markov property of the transition kernel $(p_t)_{t\geq 0}$, for any $t\geq 0$ we have

$$\pi p_t = \frac{1}{T}\int_0^T (\pi_0 p_s)p_t \mathrm{d}s = \frac{1}{T}\int_0^T \pi_0 p_{s+t}\mathrm{d}s. \tag{3.8}$$

Since the family of distributions $\{\pi_0 p_t\}_{t\geq 0}$ has the period T, we have

$$\pi p_t = \frac{1}{T}\int_0^T \pi_0 p_s \mathrm{d}s = \pi. \tag{3.9}$$

Therefore, π is the invariant distribution of the Markov process $\{X_t\}_{t\geq 0}$. The uniqueness of π follows from the contractivity in Theorem 2.2.

(ii) For given $\tau > 0$, there exists a constant $C > 0$ such that

$$\mathcal{W}_1(\mu\tilde{p}_{n\tau}, \nu\tilde{p}_{n\tau}) \leq Ce^{-nc\tau}\cdot\mathcal{W}_1(\mu,\nu), \tag{3.10}$$

then one can choose an integer $N\in\mathbb{N}$ such that $Ce^{-Nc\tau}\leq 1/2$, and

$$\mathcal{W}_1(\mu\tilde{p}_{N\tau}, \nu\tilde{p}_{N\tau}) \leq \frac{1}{2}\mathcal{W}_1(\mu,\nu), \tag{3.11}$$

so that the mapping $\nu\mapsto\nu\tilde{p}_{N\tau}$ is contractive. From the Banach fixed point theorem, this mapping has a fixed point $\tilde{\pi}_0\in\mathcal{P}_1$, i.e.

$$\tilde{\pi}_0 = \tilde{\pi}_0\tilde{p}_{N\tau}. \tag{3.12}$$

Define the probability distribution in $\mathbb{R}^{Nd}$ by

$$\tilde{\pi} = \frac{1}{N}\sum_{k=0}^{N-1} \tilde{\pi}_0 \tilde{p}_{k\tau}, \tag{3.13}$$

then from Lemma 3.1 $\tilde{\pi} \in \mathcal{P}_1$. From the Markov property of the transition kernel $(\tilde{p}_{n\tau})_{n\ge 0}$, for any $n \ge 0$ one has

$$\tilde{\pi}\tilde{p}_{n\tau} = \frac{1}{N}\sum_{k=0}^{N-1} (\tilde{\pi}_0 \tilde{p}_{k\tau})\tilde{p}_{n\tau} = \frac{1}{N}\sum_{k=0}^{N-1} \tilde{\pi}_0 \tilde{p}_{k\tau} = \tilde{\pi}. \tag{3.14}$$

Therefore, $\tilde{\pi}$ is the invariant distribution of the Markov chain $\{\tilde{X}_{n\tau}\}_{n\ge 0}$. The uniqueness of $\tilde{\pi}$ follows from the contractivity in Theorem 2.3. □

By choosing π to be the invariant distribution in Theorem 2.2, we have

$$\mathcal{W}_f(\pi, \nu p_t) \le e^{-ct}\mathcal{W}_f(\pi, \nu), \quad \forall t \ge 0, \tag{3.15}$$

which implies νp_t converges to π exponentially in the sense of the Wasserstein distance $\mathcal{W}_f$. Similarly, $\nu\tilde{p}_{n\tau}$ converges to $\tilde{\pi}$ exponentially. Since $f(r)$ is equivalent to the Euclidean norm, Lemma 3.1 directly implies $\pi, \tilde{\pi}$ have the following first-moment estimation.

Corollary 3.1. *Under Assumptions* 2.1 *and* 2.3, *there exists a constant D such that if the constant L_K in Assumption* 2.3 *satisfies*

$$L_K < \frac{c_0\varphi_0\sigma^2}{16},$$

then

$$\int_{\mathbb{R}^{Nd}} \left(\frac{1}{N}\sum_{i=1}^{N} |x^i|\right)\pi(\mathrm{d}x), \quad \int_{\mathbb{R}^{Nd}} \left(\frac{1}{N}\sum_{i=1}^{N} |x^i|\right)\tilde{\pi}(\mathrm{d}x) \le D. \tag{3.16}$$

The constant D does not depend on the number of particles N, the time step τ or the batch size p.

Although the invariant distribution $\tilde{\pi}$ of the RB–IPS (1.3) depends on the time step τ, the constant D in Corollary 3.1 is independent of τ. This means the estimate of the first-order moments of $\tilde{\pi}$ is uniform in τ.

Remark 3.2. The Banach fixed point theorem in the metric space $\mathcal{P}_1$ only implies $\pi, \tilde{\pi}$ have finite first-order moments, and does not guarantee $\pi, \tilde{\pi}$ have higher order moments, despite the fact that νp_t and $\nu\tilde{p}_t$ have finite αth order moments for any $\alpha \ge 2$ (see Lemma 3.3 of this paper).

3.2. *Strong error estimation in finite time*

In stochastic analysis, the strong error relates to the trajectory difference between two stochastic processes. Suppose the IPS X_t and the RB–IPS $\tilde{X}_t$ are driven by the same Wiener process W_t in $\mathbb{R}^{Nd}$, and the initial state $X_0 = \tilde{X}_0$ is sampled from the same distribution $\nu \in \mathcal{P}_1$. In other words, X_t and $\tilde{X}_t$ are coupled in the synchronous coupling scheme, and the only difference is the random batch approximation of the interaction forces. Define the strong error between the trajectories X_t and $\tilde{X}_t$ by

$$J(t) = \frac{1}{2N}\sum_{i=1}^N \mathbb{E}|\tilde{X}_t^i - X_t^i|^2, \quad t \geq 0. \tag{3.17}$$

We aim to estimate $J(t)$ in a finite interval $t \in [0,T]$, and derive the upper bound of $J(t)$ in terms of τ. Except for Assumptions 2.1 and 2.3, we additionally require the following.

Assumption 3.1. There exist constants $C > 0$ and $q \geq 2$ such that

$$\max\{|b(x)|, |\nabla b(x)|\} \leq C(|x|+1)^q, \quad \forall\, x \in \mathbb{R}^d. \tag{3.18}$$

Remark 3.3. The requirement $q \geq 2$ in Assumption 3.1 is merely for technical convenience.

To analyze $J(t)$ is different time steps, define the filtration

$$\mathcal{F}_n = \sigma(\nu, \{W_t\}_{t\leq t_n}, \{\mathcal{D}_k\}_{0\leq k\leq n}). \tag{3.19}$$

That is, $\mathcal{F}_n$ is determined by the initial distribution ν, the Wiener process W_t before t_n and the divisions $\mathcal{D}_k$ in the first $n+1$ time steps. Under the condition of $\mathcal{F}_n$, the RB–IPS (1.3) in the time interval $[t_n, t_{n+1})$ is evolved by (1.3). Now we have the following estimate of the αth order moments.

Lemma 3.2. *Under Assumptions* 2.1 *and* 2.3, *for any given constant* $\alpha \geq 2$, *there exist positive constants* C, β *depending on* α *such that for any* $i \in \{1, \ldots, N\}$,

$$\frac{\mathrm{d}}{\mathrm{d}t}\mathbb{E}|X_t^i|^\alpha \leq -\beta \cdot \mathbb{E}|X_t^i|^\alpha + C, \quad \forall\, t \geq 0, \tag{3.20}$$

and

$$\frac{\mathrm{d}}{\mathrm{d}t}\mathbb{E}\big(|\tilde{X}_t^i|^\alpha \big| \mathcal{F}_n\big) \leq -\beta \cdot \mathbb{E}\big(|\tilde{X}_t^i|^\alpha \big| \mathcal{F}_n\big) + C, \quad t \in [t_n, t_{n+1}). \tag{3.21}$$

The constants C, β *do not depend on the number of particles* N, *the time step* τ *or the batch size* p.

The proof of Lemma 3.2 is in Appendix A, and is similar to Lemma 3.3 of Ref. 25. The asymptotic positivity of the function $\kappa(r)$ in Assumption 2.1 is essential to produce the negative coefficient $-\beta$ in (3.20), (3.21). By Lemma 3.2, we immediately deduce that both $X_t, \tilde{X}_t$ have finite αth order moments which are uniform in time.

Lemma 3.3. *Under Assumptions* 2.1 *and* 2.3, *for any given constant* $\alpha \geq 2$, *if there exists a constant* M *such that the initial distribution* ν *satisfies*

$$\max_{1\leq i\leq N}\int_{\mathbb{R}^{Nd}}|x^i|^\alpha \nu(\mathrm{d}x) \leq M,$$

then there exists a constant C *depending on* M, α *such that*

$$\sup_{t\geq 0}\mathbb{E}|X_t^i|^\alpha \leq C, \quad \sup_{t\geq 0}\mathbb{E}|\tilde{X}_t^i|^\alpha \leq C. \tag{3.22}$$

The constant C *does not depend on the number of particles* N, *the time step* τ *or the batch size* p.

Remark 3.4. The constant C in Lemma 3.3 depends on the moments of the initial distribution ν. If one wishes C to be truly independent of N, the moment upper bound M should be also independent of N. In particular, if one chooses the ν to be the Dirac distribution at the origin, the constant M is simply zero.

The following strong error estimation of the RB–IPS (1.3) is exactly the same with the results of Ref. 26, thus we only state their main theorem here. The detailed proof can be seen in Theorem 3.1 of Ref. 26.

Theorem 3.2. *Under Assumptions* 2.1, 2.3 *and* 3.1, *if there exists a constant* M *such that the initial distribution* ν *satisfies*

$$\max_{1\leq i\leq N}\int_{\mathbb{R}^{Nd}}|x^i|^{2q} \nu(\mathrm{d}x) \leq M,$$

then for any $T>0$, *there exists a constant* C *depending on* T *and* M *such that*

$$\sup_{0\leq t\leq T} J(t) \leq C\left(\frac{\tau}{p-1}+\tau^2\right). \tag{3.23}$$

The constant C *does not depend on the number of particles* N, *the time step* τ *or the batch size* p.

Remark 3.5. In the statement of Theorem 3.1 of Ref. 26, the assumptions on the moments of the initial distribution $\nu \in \mathcal{P}(\mathbb{R}^{Nd})$ are not explicitly specified. In the proof of Theorem 3.1 of Ref. 26, when $b(x)$ satisfies the polynomial growth condition as in Assumption 3.1, the finiteness of the $2q$th order moments is enough to yield the estimation of $J(t)$ in (3.23). We clarify the assumptions on the moments of ν in the statement of Theorem 3.2 of this paper.

Now we can estimate the Wasserstein distance $\mathcal{W}_1(\nu p_t, \nu\tilde{p}_t)$ using the estimate of $J(t)$, where ν is the initial distribution, and $p_t, \tilde{p}_t$ are the transition kernels of the IPS (1.1) and the RB–IPS (1.3). Recall that normalized $\mathcal{W}_1$-distance between two probability distributions $\mu, \nu \in \mathcal{P}_1$ is defined by

$$\mathcal{W}_1(\mu, \nu) = \inf_{\gamma \in \Pi(\mu,\nu)} \int_{\mathbb{R}^{Nd} \times \mathbb{R}^{Nd}} \left(\frac{1}{N} \sum_{i=1}^{N} |x^i - y^i| \right) \gamma(\mathrm{d}x\mathrm{d}y),$$

hence if we choose γ to be the synchronous coupling (driven by the same Wiener process W_t), the Wasserstein distance $\mathcal{W}_1(\nu p_t, \nu \tilde{p}_t)$ can be bounded by

$$\begin{aligned}
\mathcal{W}_1(\nu p_t, \nu \tilde{p}_t) &\le \mathbb{E}\left(\frac{1}{N} \sum_{i=1}^{N} |X^i - \tilde{X}_t^i| \right) \\
&\le \sqrt{\mathbb{E}\left(\frac{1}{N} \sum_{i=1}^{N} |X_t^i - \tilde{X}_t^i| \right)^2} \\
&\le \sqrt{\frac{1}{N} \sum_{i=1}^{N} \mathbb{E}|X_t^i - \tilde{X}_t^i|^2} = \sqrt{2J(t)},
\end{aligned}$$

that is, $\mathcal{W}_1(\nu p_t, \nu \tilde{p}_t) \le \sqrt{2J(t)}$. Therefore, the estimate of $J(t)$ immediately implies the following.

Corollary 3.2. *Under Assumptions* 2.1, 2.3 *and* 3.1, *if there exists a constant M such that the initial distribution ν satisfies*

$$\max_{1 \le i \le N} \int_{\mathbb{R}^{Nd}} |x^i|^{2q} \nu(\mathrm{d}x) \le M,$$

then for any $T > 0$, there exists a constant C depending on T and M such that

$$\sup_{0 \le t \le T} \mathcal{W}_1(\nu p_t, \nu \tilde{p}_t) \le C \sqrt{\frac{\tau}{p-1} + \tau^2}. \tag{3.24}$$

The constant C does not depend on the number of particles N, the time step τ or the batch size p.

When the batch size p is small, $\sqrt{\tau/(p-1)}$ dominates the Wasserstein error $\mathcal{W}_1(\nu p_t, \nu \tilde{p}_t)$. In this sense, the Wasserstein error $\mathcal{W}_1(\nu p_t, \nu \tilde{p}_t)$ has at least half-order convergence with respect to the time step τ.

3.3. *Estimate of $\mathcal{W}_1(\pi, \tilde{\pi})$*

Now we estimate $\mathcal{W}_1(\pi, \tilde{\pi})$ using the results derived in the previous sections.

Theorem 3.3. *Under Assumptions* 2.1, 2.3 *and* 3.1, *there exists a constant C such that if the constant L_K in Assumption* 2.3 *satisfies*

$$L_K < \frac{c_0 \varphi_0 \sigma^2}{16},$$

then the invariant distributions $\pi, \tilde{\pi}$ of the IPS (1.1) *and the RB–IPS* (1.3) *satisfy*

$$\mathcal{W}_1(\pi, \tilde{\pi}) \le C\sqrt{\frac{\tau}{p-1} + \tau^2}. \tag{3.25}$$

The constant C does not depend on the number of particles N, the time step τ or the batch size p.

The proof of Theorem 3.3 is basically the triangle inequality described in the Introduction, but with minor difference.

Proof. For convenience, denote the first-order moment of $\nu \in \mathcal{P}_1$ by

$$\mathcal{M}_1(\nu) = \int_{\mathbb{R}^{Nd}} \left(\frac{1}{N} \sum_{i=1}^{N} |x^i| \right) \nu(\mathrm{d}x), \tag{3.26}$$

then by Corollary 3.1 $\mathcal{M}_1(\pi), \mathcal{M}_1(\tilde{\pi}) \le D$. Hence it always holds that

$$\mathcal{W}_1(\pi, \tilde{\pi}) \le \mathcal{M}_1(\pi) + \mathcal{M}_1(\tilde{\pi}) \le 2D, \tag{3.27}$$

and we may assume $\tau < D$ in the following proof. Let ν_0 be the distribution in $\mathbb{R}^{Nd}$ with all the N particles frozen at origin, then the $2q$th order moment of ν_0 is 0. By Lemma 3.3, there exists a constant M such that

$$\sup_{t\ge 0} \left\{ \max_{1\le i\le N} \int_{\mathbb{R}^{Nd}} |x^i|^{2q} (\nu_0 p_t)(\mathrm{d}x) \right\} \le M. \tag{3.28}$$

That is to say, the $2q$th order moment of $\nu_0 p_t$ is always no greater than M.

Instead of directly measuring the distance $\mathcal{W}_1(\pi, \tilde{\pi})$, we fix a constant $T > 0$ and consider the distance $\mathcal{W}_1(\nu_0 p_T, \tilde{\pi})$. By Theorem 2.3, there exists a constant C such that for any $n \ge 0$,

$$\begin{aligned}
\mathcal{W}_1(\nu_0 p_T, \tilde{\pi}) &= \mathcal{W}_1(\nu_0 p_T, \tilde{\pi}\tilde{p}_{n\tau}) \\
&\le \mathcal{W}_1(\nu_0 p_T \tilde{p}_{n\tau}, \tilde{\pi}\tilde{p}_{n\tau}) + \mathcal{W}_1(\nu_0 p_T, \nu_0 p_T \tilde{p}_{n\tau}) \\
&\le Ce^{-cn\tau} \mathcal{W}_1(\nu_0 p_T, \tilde{\pi}) + \mathcal{W}_1(\nu_0 p_T, \nu_0 p_T \tilde{p}_{n\tau}).
\end{aligned}$$

For given value of $\tau < D$, if one chooses the integer n to be

$$n = \left\lceil \frac{\log(2C)}{c\tau} \right\rceil, \tag{3.29}$$

then $Ce^{-cn\tau} \le \frac{1}{2}$ and

$$n\tau \le \left(\frac{\log(2C)}{c\tau} + 1 \right) \tau \le \frac{\log(2C)}{c} + D, \tag{3.30}$$

hence $n\tau$ has an upper bound. For this chosen n one has

$$\begin{aligned}
\mathcal{W}_1(\nu_0 p_T, \tilde{\pi}) &\le 2 \cdot \mathcal{W}_1(\nu_0 p_T, \nu_0 p_T \tilde{p}_{n\tau}) \\
&\le 2 \cdot \mathcal{W}_1(\nu_0 p_T, \nu_0 p_T p_{n\tau}) + 2 \cdot \mathcal{W}_1(\nu_0 p_T p_{n\tau}, \nu_0 p_T \tilde{p}_{n\tau}) \\
&\le Ce^{-cT} \mathcal{W}_1(\nu_0, \nu_0 p_{n\tau}) + 2 \cdot \mathcal{W}_1(\nu_0 p_T p_{n\tau}, \nu_0 p_T \tilde{p}_{n\tau}).
\end{aligned}$$

Passing to the limit $T \to \infty$ gives

$$\mathcal{W}_1(\pi, \tilde{\pi}) \leq 2 \varlimsup_{T\to\infty} \mathcal{W}_1(\nu_0 p_T p_{n\tau}, \nu_0 p_T \tilde{p}_{n\tau}). \tag{3.31}$$

Note that $\nu_0 p_T$ always has finite $2q$th order moments, hence by Corollary 3.2,

$$\mathcal{W}_1(\nu_0 p_T p_{n\tau}, \nu_0 p_T \tilde{p}_{n\tau}) \leq C\sqrt{\frac{\tau}{p-1} + \tau^2}, \quad \forall\, T > 0, \tag{3.32}$$

where the constant C does not depend on N, τ, p or the choice of T. Combining (3.31) and (3.32) we obtain the estimate of $\mathcal{W}_1(\mu, \tilde{\mu})$:

$$\mathcal{W}_1(\pi, \tilde{\pi}) \leq C\sqrt{\frac{\tau}{p-1} + \tau^2}, \tag{3.33}$$

which is exactly the result we need. □

Remark 3.6. We have some remarks on Theorem 3.3.

(1) In general, the $\mathcal{W}_1$-distance between the invariant distributions $\pi, \tilde{\pi}$ is of order $O(\sqrt{\tau})$, regardless of the batch size p used in the RB–IPS (1.3). Nevertheless, increasing the batch size p shall reduce the value of $\mathcal{W}_1(\pi, \tilde{\pi})$ as in (3.25).
(2) We estimate the distance $\mathcal{W}_1(\nu_0 p_T, \tilde{\pi})$ instead of $\mathcal{W}_1(\pi, \tilde{\pi})$ because it is nontrivial to prove the invariant distributions $\pi, \tilde{\pi} \in \mathcal{P}_1$ has finite $2q$th order moments. Therefore, we use a series of distributions $\{\nu_0 p_T\}_{T\geq 0}$ to approximate π, where the moments of $\nu_0 p_T$ can be easily derived.
(3) In this framework, the order of accuracy in the estimation of $\mathcal{W}_1(\pi, \tilde{\pi})$ cannot be greater than the order of the strong error. It is still an open question whether it is possible to apply the weak error estimation instead of the strong one in this framework. The main difficulty comes from the fact that we can only derive the geometric ergodicity in the sense of the Wasserstein distance, which is stronger than the weak error.

Combining Corollary 2.3 and Theorem 3.3, we immediately obtain the following.

Corollary 3.3. *Under Assumptions* 2.1, 2.3 *and* 3.1, *let and* $c := c_0\sigma^2/2$ *and* $\tilde{p}_t$ *be the transition kernel of the RB–IPS* (1.3). *There exists a constant C such that if the constant L_K in Assumption* 2.3 *satisfies*

$$L_K < \frac{c_0 \varphi_0 \sigma^2}{16},$$

then one has

$$\mathcal{W}_1(\nu\tilde{p}_{n\tau}, \pi) \leq Ce^{-cn\tau}\mathcal{W}_1(\nu, \pi) + C\sqrt{\frac{\tau}{p-1} + \tau^2}, \quad \forall\, n \geq 0 \tag{3.34}$$

for any probability distribution $\nu \in \mathcal{P}_1$. The constant C does not depend on the number of particles N, the time step τ or the batch size p.

Corollary 3.3 reveals the long-time sampling accuracy of the RB–IPS (1.3). The sampling error of the RB–IPS (1.3) consists of two parts: one part is the exponential convergence toward the equilibrium; the other part is the asymptotic error of the invariant distributions. Both parts do not depend on the number of particles N.

Remark 3.7. We can also estimate the sampling accuracy of the RB–IPS (1.3) in the sense of weak error. Suppose the test function $\phi : \mathbb{R}^{Nd} \to \mathbb{R}$ is given by

$$\phi(x^1, \ldots, x^N) = \frac{1}{N} \sum_{i=1}^{N} \bar{\phi}(x^i), \tag{3.35}$$

where $\bar{\phi} : \mathbb{R}^d \to \mathbb{R}$ is an one-particle observable function with Lipschitz constant $\bar{L}$. Then for any probability distributions $\mu, \nu \in \mathcal{P}_1$, one has

$$|\langle\phi\rangle_\mu - \langle\phi\rangle_\nu| \leq \bar{L} \cdot \mathcal{W}_1(\mu, \nu), \tag{3.36}$$

where $\langle\phi\rangle_\mu := \int_{\mathbb{R}^{Nd}} \phi \mathrm{d}\mu$ denotes the statistical average of ϕ in the distribution μ. Hence (3.34) implies the weak error estimate

$$|\langle\phi\rangle_{\nu\tilde{p}_{n\tau}} - \langle\phi\rangle_\pi| \leq C\bar{L}\left(e^{-cn\tau}\mathcal{W}_1(\nu, \pi) + \sqrt{\frac{\tau}{p-1} + \tau^2}\right), \quad \forall n \geq 0. \tag{3.37}$$

In particular, the constants $C, \bar{L}$ and the convergence rate c are independent of N.

4. Conclusion

In this paper, we have investigated the long-time behavior of the RB–IPS (1.3). We have proved the ergodicity of the RB–IPS (1.3) and showed that the $\mathcal{W}_1$-distance between the invariant distributions is $O(\sqrt{\tau})$, which quantitatively characterizes the long-time sampling error of the RB–IPS (1.3). Our future work shall focus on the error analysis of the time average estimator.

Appendix A. Proof of Main Results

Proof of Lemma 2.1. Under Assumption 2.1, define the constants $R_0, R_1 \geq 0$ by

$$R_0 := \inf\{R \geq 0 : \kappa(r) \geq 0, \ \forall\, r \geq R\}, \tag{A.1}$$

$$R_1 := \inf\{R \geq R_0 : \kappa(r)R(R - R_0) \geq 16, \forall\, r \geq R\}. \tag{A.2}$$

The existence of R_0, R_1 is guaranteed by the asymptotic positivity of $\kappa(r)$. Also, one has $\kappa(r) \geq 0$ for $r \geq R_0$ and $\kappa(r)R_1(R_1 - R_0) \geq 16$ for $r \geq R_1$. Given the function $\kappa(r)$, define the auxiliary functions $\varphi(r), \Phi(r), g(r)$ by

$$\varphi(r) = \exp\left(-\frac{1}{4}\int_0^r s\kappa(s)^- \mathrm{d}s\right), \quad \Phi(r) = \int_0^r \varphi(s)\mathrm{d}s, \tag{A.3}$$

$$g(r) = \begin{cases} 1 - \dfrac{1}{2}\displaystyle\int_0^r \frac{\Phi(s)}{\varphi(s)}\mathrm{d}s \Big/ \int_0^{R_1} \frac{\Phi(s)}{\varphi(s)}\mathrm{d}s, & r \leq R_1, \\ \dfrac{1}{2} - \dfrac{\eta(r - R_1)}{1 + 4\eta(r - R_1)}, & r > R_1, \end{cases} \tag{A.4}$$

where $x^- = -\min\{x,0\}$ is the negative part of $x \in \mathbb{R}$ and the constant $\eta > 0$ is defined by

$$\eta = -g'(R_1) = \frac{1}{2}\frac{\Phi(R_1)}{\varphi(R_1)}\Big/\int_0^{R_1}\frac{\Phi(s)}{\varphi(s)}\mathrm{d}s. \tag{A.5}$$

The choice of η in (A.5) ensures that $g(r)$ is differentiable at $r = R_1$. Finally, the distance function $f(r)$ is defined as

$$f(r) = \int_0^r \varphi(s)g(s)\mathrm{d}s. \tag{A.6}$$

The only difference between Eqs. (A.3)–(A.6) and the construction of $f(r)$ in Ref. 12 is the definition of $g(r)$ for $r > R_1$. In our choice, $g(r)$ is differentiable at $r = R_1$ so that $f(r)$ is always twice differentiable, while in the original proof $f(r) \in C^1$ and $f'(r)$ is absolutely continuous. From Eqs. (A.3)–(A.6), it is easy to verify the following properties of the functions $f(r), \varphi(r), \Phi(r), g(r)$:

(1) $0 < \varphi(r) \le 1$, $\frac{1}{4} \le g(r) \le 1$. $\varphi(0) = g(0) = 1$. $\Phi(0) = 0$.
(2) The derivatives of φ and g are given by

$$\begin{aligned}\varphi'(r) &= -\frac{1}{4}r\kappa(r)^-\varphi(r),\\ g'(r) &= -\frac{1}{2}\frac{\Phi(r)}{\varphi(r)}\Big/\int_0^{R_1}\frac{\Phi(s)}{\varphi(s)}\mathrm{d}s, \quad 0 \le r \le R_1.\end{aligned} \tag{A.7}$$

Hence $\varphi'(0) = g'(0) = 0$ and $\varphi'(r) \le 0, g'(r) \le 0$ for all $r \ge 0$.
(3) The second derivative of $f(r)$ is given by

$$f''(r) = \varphi(r)g'(r) + \varphi'(r)g(r) \le 0, \tag{A.8}$$

which implies $f(r)$ is concave for all $r \ge 0$.
(4) When $r > R_0$,

$$\varphi(r) \equiv \varphi_0 := \exp\left(-\frac{1}{4}\int_0^{R_0} s\kappa(s)^-\mathrm{d}s\right). \tag{A.9}$$

Since $\varphi(r) \ge \varphi_0$ and $g(r) \ge \frac{1}{4}$ for all $r \ge 0$, one obtains the estimate

$$f'(r) = \varphi(r)g(r) \ge \frac{\varphi_0}{4}, \tag{A.10}$$

which implies $f(r) \ge \frac{\varphi_0}{4}r$ for all $r \ge 0$.
(5) Since $g(r) \le 1$,

$$\Phi(r) = \int_0^r \varphi(s)\mathrm{d}s \ge \int_0^r \varphi(s)g(s)\mathrm{d}s = f(r). \tag{A.11}$$

From $\Phi''(r) = \varphi'(r) \le 0$, $\Phi(r)$ is also concave for $r \in [0, +\infty)$.

Now we prove the inequality (2.22) with the constant c_0 defined by

$$\frac{1}{c_0} = \int_0^{R_1}\frac{\Phi(s)}{\varphi(s)}\mathrm{d}s. \tag{A.12}$$

(1) When $r \le R_1$, using $f(r) \le \Phi(r)$,

$$\begin{aligned} f''(r) &= \varphi'(r)g(r) + \varphi(r)g'(r) \\ &= -\frac{1}{4}r\kappa(r)^-\varphi(r)g(r) - \frac{1}{2}\Phi(r)\Big/\int_0^{R_1}\frac{\Phi(s)}{\varphi(s)}\mathrm{d}s \\ &\le \frac{1}{4}r\kappa(r)f'(r) - \frac{1}{2}f(r)\Big/\int_0^{R_1}\frac{\Phi(s)}{\varphi(s)}\mathrm{d}s, \end{aligned}$$

hence (2.22) holds with c_0 defined in (A.12).

(2) When $r > R_1$, $f'(r) \ge \varphi_0/4$, $f''(r) \le 0$. Hence by the definition of R_1 and the concavity of $\Phi(r)$ with $\Phi(0) = 0$, one has

$$\begin{aligned} f''(r) - \frac{1}{4}r\kappa(r)f'(r) &\le -\frac{1}{16}r\kappa(r)\varphi_0 \le -\frac{\varphi_0}{R_1 - R_0}\frac{r}{R_1} \\ &\le -\frac{\varphi_0}{R_1 - R_0}\frac{\Phi(r)}{\Phi(R_1)}. \end{aligned} \tag{A.13}$$

Since $\varphi(r) \equiv \varphi_0$ for $r \ge R_0$, $\Phi(r)$ is linear in r, i.e.

$$\Phi(r) = \Phi(R_0) + (r - R_0)\varphi_0, \quad r \ge R_0. \tag{A.14}$$

In particular, $\Phi(R_1) = \Phi(R_0) + (R_1 - R_0)\varphi_0$, hence

$$\begin{aligned} \int_{R_0}^{R_1}\frac{\Phi(s)}{\varphi(s)}\mathrm{d}s &= \frac{\Phi(R_0)}{\varphi_0}(R_1 - R_0) + \frac{1}{2}(R_1 - R_0)^2 \\ &\ge \frac{1}{2}(R_1 - R_0)\frac{\Phi(R_1)}{\varphi_0}. \end{aligned} \tag{A.15}$$

Combining (A.13) and (A.15) one obtains

$$f''(r) - \frac{1}{4}r\kappa(r)f'(r) \le -\frac{1}{2}\Phi(r)\Big/\int_{R_0}^{R_1}\frac{\Phi(s)}{\varphi(s)}\mathrm{d}s \le -\frac{1}{2}f(r)\Big/\int_0^{R_1}\frac{\Phi(s)}{\varphi(s)}\mathrm{d}s, \tag{A.16}$$

hence (2.22) holds with c_0 defined in (A.12).

It is easy to see $\frac{\varphi_0}{4}r \le f(r) \le r$ for all $r \ge 0$. □

Proof of Lemma 2.2. Using $b^i(x) = b(x^i) + \gamma^i(x)$, the left-hand side of (2.24) is written as $I = I_1 + I_2 + I_3$,

$$I_1 = \sum_{i=1}^N (r^i)^{-1}Z^i \cdot (b(X^i) - b(Y^i))f'(r^i),$$

$$I_2 = \sum_{i=1}^N (r^i)^{-1}Z^i \cdot (\gamma^i(X) - \gamma^i(Y))f'(r^i),$$

$$I_3 = 2\sigma^2\sum_{i=1}^N \mathrm{rc}^2(Z^i)f''(r^i).$$

Now we estimate I_1, I_2, I_3, respectively.

- Estimate I_1: By the definition of $\kappa(r)$ in (2.19),

$$I_1 \leq -\frac{\sigma^2}{2}\sum_{i=1}^N r^i\kappa(r^i)f'(r^i). \tag{A.17}$$

- Estimate I_2: Using the Lipschitz condition in Assumption 2.2 and $f(r) \geq \varphi_0 r/4$,

$$I_2 \leq \sum_{i=1}^N |\gamma^i(X) - \gamma^i(Y)| \leq L\sum_{i=1}^N r^i \leq \frac{4L}{\varphi_0}\sum_{i=1}^N f(r^i). \tag{A.18}$$

- Estimate I_3: Using the estimation of $f''(r)$ in (2.22),

$$\begin{aligned} I_3 &\leq \frac{\sigma^2}{2}\sum_{i=1}^N r^i\kappa(r^i)\mathrm{rc}^2(Z^i)f'(r^i) - c_0\sigma^2\sum_{i=1}^N \mathrm{rc}^2(Z^i)f(r^i) \\ &= \frac{\sigma^2}{2}\sum_{i=1}^N r^i\kappa(r^i)f'(r^i) - c_0\sigma^2\sum_{i=1}^N f(r^i) \\ &\quad \underbrace{-\frac{\sigma^2}{2}\sum_{i=1}^N r^i\kappa(r^i)(1-\mathrm{rc}^2(Z^i))f'(r^i)}_{I_{31}} + \underbrace{c_0\sigma^2\sum_{i=1}^N(1-\mathrm{rc}^2(Z^i))f(r^i)}_{I_{32}}. \end{aligned} \tag{A.19}$$

We estimate I_{31} and I_{32} in (A.19).

— Estimate I_{31}: Note that $1 - \mathrm{rc}^2(Z^i) = 0$ if $r_i \geq \delta$, thus

$$\begin{aligned} I_{31} &= -\frac{\sigma^2}{2}\sum_{i=1}^N r^i\kappa(r^i)(1-\mathrm{rc}^2(Z^i))f'(r^i) \\ &\leq \frac{\sigma^2}{2}\sum_{i:r^i<\delta} r^i\kappa(r^i)^- f'(r^i) \\ &\leq \frac{\sigma^2}{2}\sum_{i:r^i<\delta} r^i\kappa(r^i)^- \\ &\leq \frac{N\sigma^2}{2}\sup_{r<\delta}(r\kappa(r)^-). \end{aligned}$$

— Estimate I_{32}: In a similar way, using $f(r) \leq r$ one obtains

$$\begin{aligned} I_{32} &= c_0\sigma^2\sum_{i=1}^N(1-\mathrm{rc}^2(Z^i))f(r^i) \\ &\leq c_0\sigma^2\sum_{i:r^i<\delta} f(r^i) \\ &\leq c_0 N\sigma^2\delta. \end{aligned}$$

From the definition of $m(\delta)$ in (2.25), one obtains the estimate of I_3:

$$I_3 \leq \frac{\sigma^2}{2} \sum_{i=1}^N r^i \kappa(r^i) f'(r^i) - c_0 \sigma^2 \sum_{i=1}^N f(r^i) + Nm(\delta). \tag{A.20}$$

Summation over the estimates (A.17), (A.18), (A.20) of I_1, I_2, I_3 gives

$$I \leq -\left(c_0 \sigma^2 - \frac{4L}{\varphi_0} \right) \sum_{i=1}^N f(r^i) + Nm(\delta). \tag{A.21}$$

When the Lipschitz constant $L < c_0 \varphi_0 \sigma^2 / 8$, one has

$$I \leq -\frac{c_0 \sigma^2}{2} \sum_{i=1}^N f(r^i) + Nm(\delta) = Nm(\delta) - c \sum_{i=1}^N f(r^i), \tag{A.22}$$

which is exactly the result we need. □

Proof of Lemma 3.1. Consider the stochastic processes X_t and $\tilde{X}_t$ evolved by the IPS (1.1) and the RB–IPS (1.3), respectively, with the initial distribution $\nu \in \mathcal{P}_1$. For convenience, we unify (1.1), (1.3) in the form of the product model (2.1).

(i) By choosing a smooth function $f(x) = \sqrt{|x|^2 + 1}$, each $f(X_t^i)$ satisfies the SDE

$$\mathrm{d} f(X_t^i) = b^i(X_t) \cdot \nabla f(X_t^i) \mathrm{d}t + \frac{\sigma^2}{2} \Delta f(X_t^i) \mathrm{d}t + \nabla f(X_t^i) \cdot \sigma \mathrm{d}W_t^i, \tag{A.23}$$

where $\Delta = \nabla \cdot \nabla$ is the Laplacian operator in $\mathbb{R}^d$. Taking the expectation, one obtains

$$\frac{\mathrm{d}}{\mathrm{d}t} \mathbb{E}[f(X_t^i)] = \mathbb{E}\left(b^i(X_t) \cdot \nabla f(X_t^i) + \frac{\sigma^2}{2} \Delta f(X_t^i) \right). \tag{A.24}$$

Note that the first and second derivatives of $f(x)$ and the perturbation $\gamma^i(x)$ are uniformly bounded (we have assumed $K(\cdot)$ to be bounded in Assumption 2.3), for each $i \in \{1, \ldots, N\}$ there is

$$\frac{\mathrm{d}}{\mathrm{d}t} \mathbb{E}[f(X_t^i)] \leq \mathbb{E}\big(b(X_t^i) \cdot \nabla f(X_t^i) \big) + C = \mathbb{E}\left(\frac{b(X_t^i) \cdot X_t^i}{\sqrt{|X_t^i|^2 + 1}} \right) + C. \tag{A.25}$$

Under Assumption 2.1, we claim that there exists constants $C, \beta > 0$ such that

$$\frac{x \cdot b(x)}{\sqrt{|x|^2 + 1}} \leq C - \beta \sqrt{|x|^2 + 1}, \quad \forall\, x \in \mathbb{R}^d. \tag{A.26}$$

In fact, from $\kappa(r)^- = 0$ for $r \geq R_0$, one has

$$\begin{aligned} x \cdot b(x) &\leq x \cdot b(0) - \frac{\sigma^2}{2} \kappa(r) |x|^2 \\ &= x \cdot b(0) - \frac{\sigma^2}{2} \kappa(r)^+ |x|^2 + \frac{\sigma^2}{2} \kappa(r)^- |x|^2 \end{aligned}$$

$$\leq x \cdot b(0) - \frac{\sigma^2}{2}\kappa(r)^+|x|^2 + \frac{\sigma^2 R_0}{2}\kappa(r)^-|x|$$

$$\leq C|x| - \frac{\sigma^2}{2}\kappa(r)^+|x|^2,$$

where $x^+ = \max\{x, 0\}$ denotes the positive part of $x \in \mathbb{R}$. Thus (A.26) holds true. Combining (A.26) and (A.25) yields

$$\frac{\mathrm{d}}{\mathrm{d}t}\mathbb{E}[f(X_t^i)] \leq C - \beta \cdot \mathbb{E}[f(X_t^i)], \quad i = 1, \ldots, N. \tag{A.27}$$

For the IPS $\{X_t\}_{t\geq 0}$, define

$$m(t) = \mathbb{E}\left(\frac{1}{N}\sum_{i=1}^N f(X_t^i)\right), \quad \forall\, t \geq 0. \tag{A.28}$$

Since ν is the initial distribution of X_0, clearly $m(t)$ is an upper bound of

$$\int_{\mathbb{R}^{Nd}} \left(\frac{1}{N}\sum_{i=1}^N |x^i|\right)(\nu p_t)(\mathrm{d}x). \tag{A.29}$$

Summation over $i \in \{1, \ldots, N\}$ in (A.27) gives

$$m'(t) \leq C - \beta \cdot m(t). \tag{A.30}$$

Hence $m(t)$ is finite for all $t \geq 0$, and by Gronwall's inequality,

$$\overline{\lim_{t\to\infty}} \int_{\mathbb{R}^{Nd}} \left(\frac{1}{N}\sum_{i=1}^N |x^i|\right)(\nu p_t)(\mathrm{d}x) \leq \overline{\lim_{t\to\infty}}\, m(t) \leq \frac{C}{\beta}. \tag{A.31}$$

Now one may just take $D = C/\beta$ in Lemma 3.1.

(ii) The moment estimate for the RB–IPS (1.3) can be derived in a similar way. For convenience, define the filtration $\mathcal{F}_n$ by

$$\mathcal{F}_n = \sigma(\nu, \{W_s\}_{0\leq s\leq t_n}, \{\mathcal{D}_k\}_{0\leq k\leq n}). \tag{A.32}$$

That is, $\mathcal{F}_n$ is determined by the initial distribution ν of $\tilde{X}_0$, the Wiener process W_t before time t_n, and the divisions in the first $n+1$ time steps. For the RB–IPS $\{\tilde{X}_t\}_{t\geq 0}$, define

$$\tilde{m}(t) = \mathbb{E}\left(\frac{1}{N}\sum_{i=1}^N f(\tilde{X}_t^i)\right), \quad \forall\, t \geq 0. \tag{A.33}$$

Under the condition of the filtration $\mathcal{F}_n$, define

$$\tilde{m}(t|\mathcal{F}_n) := \mathbb{E}\left(\frac{1}{N}\sum_{i=1}^N f(\tilde{X}_t^i)\middle|\mathcal{F}_n\right), \quad t \in [t_n, t_{n+1}). \tag{A.34}$$

With fixed division $\mathcal{D}_n$ of the index set $\{1, \ldots, N\}$, $\tilde{X}_t$ in the time interval $[t_n, t_{n+1})$ is evolved by the RB–IPS (1.3), and Assumption 2.2 still holds true with the constant $L = 2L_K$. Therefore, similarly with (A.30), one obtains

$$\tilde{m}'(t|\mathcal{F}_n) \leq C - \beta \cdot \tilde{m}(t|\mathcal{F}_n), \quad \forall\, t \in [t_n, t_{n+1}). \tag{A.35}$$

Taking the expectation over $\mathcal{F}_n$ in (A.35) gives

$$\tilde{m}'(t) \le C - \beta \cdot \tilde{m}(t), \quad t \in [t_n, t_{n+1}). \tag{A.36}$$

Integrating (A.36) in the time interval $[t_n, t_{n+1})$ gives

$$\tilde{m}((n+1)\tau) \le e^{-\beta\tau}\tilde{m}(n\tau) + \frac{C}{\beta}(1 - e^{-\beta\tau}), \quad \forall\, n \ge 0. \tag{A.37}$$

Hence $\tilde{m}(t)$ is finite for all integers $n \ge 0$, and by Gronwall's inequality,

$$\overline{\lim_{n\to\infty}}\, \tilde{m}(n\tau) \le \frac{C}{\beta}. \tag{A.38}$$

Now one may just take $D = C/\beta$ in Lemma 3.1. $\square$

Proof of Lemma 3.2. We first estimate $\mathbb{E}|X_t^i|^\alpha$ for the IPS (1.1). By Itô calculus,

$$\begin{aligned}\frac{\mathrm{d}}{\mathrm{d}t}\mathbb{E}|X_t^i|^\alpha &= \alpha \cdot \mathbb{E}\big\{|X_t^i|^{\alpha-2}\big(X_t^i \cdot b(X_t^i) + X_t^i \cdot \gamma^i(X_t)\big)\big\} \\ &\quad + \frac{1}{2}\alpha(\alpha + d - 2)\sigma^2\mathbb{E}|X_t^i|^{\alpha-2},\end{aligned} \tag{A.39}$$

where the perturbation $\gamma^i(x)$ is given by (2.3). By the definition of $\kappa(r)$, one has

$$-x \cdot (b(x) - b(0)) \ge \frac{\sigma^2}{2}\kappa(|x|)|x|^2, \quad \forall\, x \in \mathbb{R}^d. \tag{A.40}$$

Hence the drift force part in (A.39) is bounded by

$$|X_t^i|^{\alpha-2}X_t^i \cdot b(X_t^i) \le C|X_t^i|^{\alpha-1} - \frac{\sigma^2}{2}\kappa(|X_t^i|)|X_t^i|^\alpha. \tag{A.41}$$

Since $\gamma^i(x)$ is uniformly bounded according to Assumption 2.3, the perturbation part in (A.39) is bounded by

$$|X_t^i|^{\alpha-2}X_t^i \cdot \gamma^i(X_t) \le C|X_t^i|^{\alpha-1}. \tag{A.42}$$

Combining (A.41) and (A.42), from (A.39) one deduces that

$$\frac{\mathrm{d}}{\mathrm{d}t}\mathbb{E}|X_t^i|^\alpha \le -\frac{\alpha\sigma^2}{2}\mathbb{E}\big(\kappa(|X_t^i|)|X_t^i|^\alpha\big) + C\big(\mathbb{E}|X_t^i|^{\alpha-1} + \mathbb{E}|X_t^i|^{\alpha-2}\big). \tag{A.43}$$

Since $\kappa(r) \ge \delta$ for $r \ge R_0$ and $\kappa(r)$ has a lower bound for $r > 0$, one has

$$-\kappa(r)r^\alpha = (\delta - \kappa(r))r^\alpha - \delta r^\alpha \le C - \delta r^\alpha, \quad \forall\, r \ge 0, \tag{A.44}$$

which implies

$$-\mathbb{E}\big(\kappa(|X_t^i|)|X_t^i|^\alpha\big) \le C - \delta \cdot \mathbb{E}|X_t^i|^\alpha. \tag{A.45}$$

Therefore by choosing $c = \alpha\sigma^2\delta/2$, one has

$$\frac{\mathrm{d}}{\mathrm{d}t}\mathbb{E}|X_t^i|^\alpha \le -c \cdot \mathbb{E}|X_t^i|^\alpha + C(\mathbb{E}|X_t^i|^{\alpha-1} + \mathbb{E}|X_t^i|^{\alpha-2} + 1). \tag{A.46}$$

Using interpolation inequality, $\mathbb{E}|X_t^i|^{\alpha-1}$ and $\mathbb{E}|X_t^i|^{\alpha-2}$ can be bounded by $\mathbb{E}|X^i|^\alpha$ plus constant. Therefore, (A.46) implies

$$\frac{\mathrm{d}}{\mathrm{d}t}\mathbb{E}|X_t^i|^\alpha \leq -\frac{c}{2}\cdot \mathbb{E}|X_t^i|^\alpha + C \tag{A.47}$$

for some constant C, which is exactly the result we need. For the RB–IPS (1.3), the perturbation $\gamma^i(x)$ given by (2.4) is bounded by $2L_K$, thus the proof still works for the RB–IPS (1.3) in the time interval $[t_n, t_{n+1})$ under the condition of $\mathcal{F}_n$. $\square$

Acknowledgments

Z. Zhou was partially supported by the National Key R&D Program of China, Project Number 2021YFA1001200 and the NSFC Grant No. 12171013. Both S. Jin and L. Li were partially supported by Shanghai Municipal Science and Technology Major Project No. 2021SHZDZX0102 and NSFC Grant No. 12031013. In addition, S. Jin was partially supported by the Science and Technology Commission of Shanghai Municipality Grant No. 20JC1414100, and L. Li was partially sponsored by the Strategic Priority Research Program of Chinese Academy of Sciences, Grant No. XDA25010403, and NSFC 11901389.

References

1. L. Ambrosio, N. Gigli and G. Savaré, *Gradient Flows*: *In Metric Spaces and in the Space of Probability Measures* (Springer Science & Business Media, 2005).
2. D. Bakry *et al.*, *Analysis and Geometry of Markov Diffusion Operators*, Vol. 103 (Springer, 2014).
3. C. Bordenave, D. McDonald and A. Proutiere, A particle system in interaction with a rapidly varying environment: Mean field limits and applications, arXiv: math/0701363.
4. L. Bottou, Large-scale machine learning with stochastic gradient descent, in *Proc. COMPSTAT'2010* (Springer, 2010), pp. 177–186.
5. L. Bottou, Stochastic gradient descent tricks, in *Neural Networks*: *Tricks of the Trade* (Springer, 2012), pp. 421–436.
6. N. Bou-Rabee and A. Eberle, Couplings for Andersen dynamics, arXiv:2009.14239.
7. N. Bou-Rabee and A. Eberle, Two-scale coupling for preconditioned Hamiltonian Monte Carlo in infinite dimensions, *Stocha. Partial Differential Equations*: *Anal. Comput.* **9** (2021) 207–242.
8. N. Bou-Rabee, A. Eberle and R. Zimmer, Coupling and convergence for Hamiltonian Monte Carlo, *Ann. Appl. Probab.* **30** (2020) 1209–1250.
9. K. S. Chan and H. Tong, On the use of the deterministic Lyapunov function for the ergodicity of stochastic difference equations, *Adv. Appl. Probab.* **17** (1985) 666–678.
10. C. Chen, N. Ding and L. Carin, On the convergence of stochastic gradient MCMC algorithms with high-order integrators, *Adv. Neural Inform. Process. Syst.* **28** (2015) 2278–2286.
11. A. Durmus, A. Eberle, A. Guillin and R. Zimmer, An elementary approach to uniform in time propagation of chaos, *Proc. Amer. Math. Soc.* **148** (2020) 5387–5398.
12. A. Eberle, Reflection coupling and Wasserstein contractivity without convexity, *C. R. Math.* **349** (2011) 1101–1104.

13. A. Eberle, Reflection couplings and contraction rates for diffusions, *Probab. Theory Relat. Fields* **166** (2016) 851–886.
14. A. Eberle, A. Guillin and R. Zimmer, Couplings and quantitative contraction rates for Langevin dynamics, *Ann. Probab.* **47** (2019) 1982–2010.
15. A. Eberle, A. Guillin and R. Zimmer, Quantitative Harris-type theorems for diffusions and McKean–Vlasov processes, *Trans. Amer. Math. Soc.* **371** (2019) 7135–7173.
16. H. Fehske, R. Schneider and A. Weiße, *Computational Many-Particle Physics*, Vol. 739 (Springer, 2007).
17. D. Frenkel and B. Smit, *Understanding Molecular Simulation*: *From Algorithms to Applications*, Vol. 1 (Elsevier, 2001).
18. F. Golse, The mean-field limit for the dynamics of large particle systems, *Journées Équations aux Dérivées Partielles* **9** (2003) 1–47.
19. L. Greengard and V. Rokhlin, A fast algorithm for particle simulations, *J. Comput. Phys.* **73** (1987) 325–348.
20. A. Guillin, W. Liu, L. Wu and C. Zhang, Uniform poincar $\{\backslash' e\}$ and logarithmic sobolev inequalities for mean field particles systems, arXiv:1909.07051.
21. S.-Y. Ha, S. Jin, D. Kim and D. Ko, Convergence toward equilibrium of the first-order consensus model with random batch interactions, *J. Differential Equations* **302** (2021) 585–616.
22. M. Hairer and J. C. Mattingly, *Yet Another Look at Harris' Ergodic Theorem for Markov Chains* (Springer, 2011), pp. 109–117.
23. D. P. Herzog and J. C. Mattingly, Ergodicity and Lyapunov functions for Langevin dynamics with singular potentials, *Commun. Pure Appl. Math.* **72** (2019) 2231–2255.
24. P.-E. Jabin and Z. Wang, Mean field limit for stochastic particle systems, in *Active Particles*, *Volume* 1 (Springer, 2017), pp. 379–402.
25. S. Jin, L. Li and J.-G. Liu, Random batch methods (RBM) for interacting particle systems, *J. Comput. Phys.* **400** (2020) 108877.
26. S. Jin, L. Li and J.-G. Liu, Convergence of the random batch method for interacting particles with disparate species and weights, *SIAM J. Numer. Anal.* **59** (2021) 746–768.
27. 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) B937–B960.
28. B. Leimkuhler, M. Sachs and G. Stoltz, Hypocoercivity properties of adaptive Langevin dynamics, *SIAM J. Appl. Math.* **80** (2020) 1197–1222.
29. T. Lelievre and G. Stoltz, Partial differential equations and stochastic methods in molecular dynamics, *Acta Numer.* **25** (2016) 681–880.
30. P. Lévy, Sur certains processus stochastiques homogènes, *Compositio Mathematica* **7** (1940) 283–339.
31. 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) A1486–A1509.
32. X. Li and F. Orabona, On the convergence of stochastic gradient descent with adaptive stepsizes, in *22nd Int. Conf. on Artificial Intelligence and Statistics* (PMLR, 2019), pp. 983–992.
33. 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) 185–232.
34. J. C. Mattingly, A. M. Stuart and M. V. Tretyakov, Convergence of numerical time-averaging and stationary measures via Poisson equations, *SIAM J. Numer. Anal.* **48** (2010) 552–577.

35. S. P. Meyn and R. L. Tweedie, *Markov Chains and Stochastic Stability* (Springer Science & Business Media, 2012).
36. O. Shamir and T. Zhang, Stochastic gradient descent for non-smooth optimization: Convergence results and optimal averaging schemes, in *International Conference on Machine Learning* (PMLR, 2013), pp. 71–79.
37. T. Shardlow and A. M. Stuart, A perturbation theory for ergodic Markov chains and application to numerical approximations, *SIAM J. Numer. Anal.* **37** (2000) 1120–1137.
38. C. Villani, *Hypocoercivity* (Amer. Math. Soc., 2009), pp. 949–951.
39. C. Villani, *Optimal Transport*: *Old and New*, Vol. 338 (Springer, 2009).
40. M. Welling and Y. W. Teh, Bayesian learning via stochastic gradient Langevin dynamics, in *Proc. 28th Int. Conf. on Machine Learning* (Citeseer, 2011), pp. 681–688.
41. X. Ye and Z. Zhou, Efficient sampling of thermal averages of interacting quantum particle systems with random batches, *J. Chem. Phys.* **154** (2021) 204106.
42. X. Ye and Z. Zhou, Error analysis of time-discrete random batch method for interacting particle systems and associated mean-field limits, arXiv:2206.02166.
43. D. Zou, P. Xu and Q. Gu, Faster convergence of stochastic gradient Langevin dynamics for non-log-concave sampling, in *Uncertainty in Artificial Intelligence* (PMLR, 2021), pp. 1152–1162.