

## Error analysis of time-discrete random batch method for interacting particle systems and associated mean-field limits

XUDA YE AND ZHENNAN ZHOU\*

Beijing International Center for Mathematical Research, Peking University, Beijing 100871, China

\*Corresponding author: zhennan@bicmr.pku.edu.cn

[Received on 31 August 2022; revised on 25 February 2023]

The random batch method provides an efficient algorithm for computing statistical properties of a canonical ensemble of interacting particles. In this work, we study the error estimates of the fully discrete random batch method, especially in terms of approximating the invariant distribution. The triangle inequality framework employed in this paper is a convenient approach to estimate the long-time sampling error of the numerical methods. Using the triangle inequality framework, we show that the long-time error of the discrete random batch method is  $O(\sqrt{\tau} + e^{-\lambda t})$ , where  $\tau$  is the time step and  $\lambda$  is the convergence rate, which does not depend on the time step  $\tau$  or the number of particles  $N$ . Our results also apply to the McKean–Vlasov process, which is the mean-field limit of the interacting particle system as the number of particles  $N \rightarrow \infty$ .

**Keywords:** random batch method; interacting particle system; McKean–Vlasov process; mean-field limit; long-time error analysis.

### 1. Introduction

Simulation of large interacting particle systems (IPS) has always been an appealing research topic in computational physics (Frenkel & Smit, 2001; Golse, 2003) and computational chemistry (Lelievre & Stoltz, 2016; Li *et al.*, 2020a). It is not only because the IPS itself is an important model in molecular dynamics and quantum mechanics, but also because the IPS has a mathematically well-defined mean-field limit (Bordenave *et al.*, 2007; Leimkuhler & Matthews, 2015; Golse *et al.*, 2019; Jin & Li, 2022) as the number of particles tends to infinity. The mean-field dynamics of the IPS is a distribution-dependent Stochastic differential equation (SDE), also known as the McKean–Vlasov process (MVP), which has been frequently used in statistical physics to describe the ensemble behavior of a system of particles (Jabin & Wang, 2017; Bahlali *et al.*, 2020). In this paper, we focus on a simple IPS model, which is evolved by the overdamped Langevin dynamics with only pairwise interactions.

Consider the system of  $N$  particles in  $\mathbb{R}^d$  represented by a collection of position variables  $X_t = \{X_t^i\}_{i=1}^N$ , where the subscript  $t \geq 0$  denotes the evolution time and each particle  $X_t^i \in \mathbb{R}^d$  is evolved by the overdamped Langevin dynamics

$$dX_t^i = \left( b(X_t^i) + \frac{1}{N-1} \sum_{j \neq i} K(X_t^i - X_t^j) \right) dt + \sigma dW_t^i. \quad (1.1)$$

Here,  $b : \mathbb{R}^d \rightarrow \mathbb{R}^d$  is the drift force,  $K : \mathbb{R}^d \rightarrow \mathbb{R}^d$  is the interaction force,  $\sigma > 0$  is the diffusion coefficient and  $\{W_t^i\}_{i=1}^N$  are  $N$  independent Wiener processes in  $\mathbb{R}^d$ . Formally, the mean-field limit of

(1.1) as  $N \rightarrow \infty$  is the MVP represented by a single position variable  $\bar{X}_t \in \mathbb{R}^d$

$$\begin{aligned} d\bar{X}_t &= \left( b(\bar{X}_t) + \int_{\mathbb{R}^d} K(\bar{X}_t - z) \nu_t(dz) \right) dt + \sigma dW_t, \\ \nu_t &= \text{Law}(\bar{X}_t). \end{aligned} \quad (1.2)$$

Here,  $\text{Law}(\cdot)$  denotes the distribution law of a random variable, and  $W_t$  is the Wiener process in  $\mathbb{R}^d$ . The convergence mechanism of the IPS (1.1) towards the MVP (1.2) as  $N \rightarrow \infty$  in the finite time has been systematically studied in the theory of the propagation of chaos (Sznitman, 1991; Chaintron & Diez, 2022).

The goal of this paper is to study the sampling accuracy of the numerical methods for the IPS (1.1) and the MVP (1.2). To characterize the sampling accuracy of a numerical method at different time scales, it is reasonable to ask the following two questions:

1. In the finite-time level, does the method produce accurate trajectories?
2. In the long-time level, does the method sample the correct invariant distribution?

The invariant distributions of the IPS (1.1) and the MVP (1.2) are probability distributions in  $\mathbb{R}^{Nd}$  and  $\mathbb{R}^d$ , respectively, and remain unchanged under the evolution of the corresponding SDEs. Let the numerical method for the IPS (1.1) with the time step  $\tau$  produce a discrete-time trajectory  $\{\tilde{X}_n\}_{n \geq 0}$  in  $\mathbb{R}^{Nd}$ , where the subscript  $n$  is a non-negative integer representing the number of iterations. In the finite-time level, we expect  $\{\tilde{X}_n\}_{n \geq 0}$  to be close to the original IPS trajectory  $\{X_{n\tau}\}_{n \geq 0}$ . In the long-time level, for sufficiently large  $n$ , we expect the numerical distribution law  $\text{Law}(\tilde{X}_n)$  to be close to the invariant distribution of the IPS (1.1).

In this paper, we shall consider the Euler–Maruyama scheme, which provides a simple numerical method for the IPS (1.1). Fix the time step  $\tau$  and define  $t_n := n\tau$ , then the IPS (1.1) is approximated by a system of particles  $\tilde{X}_n = \{\tilde{X}_n^i\}_{i=1}^N$ , where each particle  $\tilde{X}_n^i \in \mathbb{R}^d$  in the time interval  $[t_n, t_{n+1})$  is updated by the following stochastic equation

$$\tilde{X}_{n+1}^i = \tilde{X}_n^i + \left( b(\tilde{X}_n^i) + \frac{1}{N-1} \sum_{j \neq i} K(\tilde{X}_n^i - \tilde{X}_n^j) \right) \tau + \sigma (W_{t_{n+1}}^i - W_{t_n}^i), \quad (1.3)$$

which we shall refer to as the discrete IPS thereafter. We note that the discrete IPS (1.3) is also known as the stochastic particle method (Bossy & Talay, 1997; Antonelli & Kohatsu-Higa, 2002), which can be applied in a wide class of MVPs, and the associated error analysis can be found in Mattingly *et al.* (2002); Malrieu (2003); Ding & Qiao (2021); Bao & Huang (2022). To update the discrete IPS (1.3) in a single time step, we need to compute all the pairwise interactions  $K(\tilde{X}_n^i - \tilde{X}_n^j)$ , hence the computational cost per time step is  $O(N^2)$ . Such huge complexity brings great burden when  $N$  is large.

The random batch method (RBM) proposed in Jin *et al.* (2021a) resolves the complexity burden in the discrete IPS (1.3) with a simple idea: for each time step, compute the interaction forces within small random batches. 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 the equal size  $p = N/q$  where the integer  $p \geq 2$ . We compute the interaction force between two particles only when their indices  $i, j$  belong to the same

batch. The discrete IPS (1.3) is then approximated by the discrete random batch interacting particle system (discrete RB-IPS), represented by a system of particles  $\tilde{Y}_n = \{\tilde{Y}_n^i\}_{i=1}^N$  in  $\mathbb{R}^{Nd}$ , where each particle  $\tilde{Y}_n^i \in \mathbb{R}^d$  is updated by

$$\tilde{Y}_{n+1}^i = \tilde{Y}_n^i + \left( b(\tilde{Y}_n^i) + \frac{1}{p-1} \sum_{j \neq i, j \in \mathcal{C}} K(\tilde{Y}_n^i - \tilde{Y}_n^j) \right) \tau + \sigma(W_{t_{n+1}}^i - W_{t_n}^i), \quad i \in \mathcal{C}. \quad (1.4)$$

Here,  $\mathcal{C} \in \mathcal{D}$  is the unique batch containing  $i$ . For the next time interval, the previous division  $\mathcal{D}$  is discarded and another random division is employed. At each time step, all the  $N$  particles are randomly divided into  $N/p$  batches. Since each batch of size  $p$  requires  $O(p^2)$  complexity to compute the interaction forces, the discrete RB-IPS (1.4) requires only  $O(Np)$  rather than  $O(N^2)$  complexity to compute the interaction forces in a time step, which is a significant advance in simulation efficiency.

Nowadays, the RBM has become a prominent simulation tool for large particle systems. It is not only a highly efficient numerical method for complex chemical systems (Ye & Zhou, 2021; Guillin *et al.*, 2022; Jin *et al.*, 2022; Li *et al.*, 2023), but also accelerates the particle ensemble methods (Hairer & Mattingly, 2011; Li *et al.*, 2020b; Carrillo *et al.*, 2021) for optimization or solving PDEs. There have been some theoretical results on the error analysis of the RBM, but they mainly focus on the continuous-time random batch interacting particle system (RB-IPS, defined in (3.1)). In the finite-time level, it has been proved in Jin *et al.* (2021b) that the strong and weak error are  $O(\sqrt{\tau})$  and  $O(\tau)$ , respectively; while in the long-time level, the authors of Klokov & Veretennikov (2006) have applied the reflection coupling Eberle (2016); Eberle *et al.* (2019) to show that the RB-IPS has uniform geometric ergodicity, and the Wasserstein-1 distance between the invariant distributions of the IPS and the RB-IPS is bounded by  $O(\sqrt{\tau})$ . However, the error analysis of the discrete RB-IPS (1.4) is not a direct consequence of the previous results, since the update schemes are changing in different time steps, and it is nontrivial to apply the perturbation theory (Shardlow & Stuart, 2000; Rudolf & Schweizer, 2018) or the Harris ergodic theorem (Mattingly *et al.*, 2010; Hauray & Mischler, 2014; Reis *et al.*, 2022) to obtain an explicit convergence rate towards the invariant distribution. Therefore, it is necessary to study the long-time error of the discrete RB-IPS (1.4), which is the main task of this paper.

The triangle inequality framework employed in this paper is our main technique to study the long-time sampling error. This framework agrees with the perturbation theory for ergodic Markov chains Shardlow & Stuart (2000); Rudolf & Schweizer (2018) and is directly inspired by the works of McKean (1967); Mattingly *et al.* (2010) and Eberle (2011), and can be conveniently applied in a wide class of numerical methods. For a given stochastic process and the corresponding numerical method, the triangle inequality framework is able to utilize the ergodicity of the original process and the finite-time error analysis to estimate the long-time error. Furthermore, with the triangle inequality framework, it is easy to produce an explicit convergence rate, which is independent of the time step  $\tau$  and other parameters. In particular, for the IPS (1.1) and its corresponding numerical method—the discrete RB-IPS (1.4), the convergence rate is independent of the number of particles  $N$ .

Before we elaborate the principle of the triangle inequality framework in Section 2, we state the main results of this paper. These results are proved by combining the triangle inequality framework and the error analysis results for the RB-IPS in Klokov & Veretennikov (2006); Jin *et al.* (2021b).

1. **(Theorem 3.7) The finite-time strong error is  $O(\sqrt{\tau})$ .**

When the IPS (1.1) and the discrete RB-IPS (1.4) are driven by the same initial value and the same Wiener processes, there exists a positive constant  $C = C(T)$  such that

$$\sup_{0 \leq n \leq T/\tau} \left( \frac{1}{N} \sum_{i=1}^N \mathbb{E} |X_{n\tau}^i - \tilde{Y}_n^i|^2 \right) \leq C\tau. \quad (1.5)$$

The constant  $C$  does not depend on  $N$ ,  $\tau$  or  $p$ .

2. **(Theorem 3.19) The long-time sampling error is  $O(\sqrt{\tau} + e^{-\lambda t})$ .**

When the interaction force  $K$  is moderately large, there exist constants  $C, \lambda > 0$  such that

$$\mathcal{W}_1(\pi, \text{Law}(\tilde{Y}_n)) \leq C\sqrt{\tau} + Ce^{-\lambda n\tau}, \quad \forall n \geq 0, \quad (1.6)$$

where  $\pi \in \mathcal{P}(\mathbb{R}^{Nd})$  is the invariant distribution of the IPS (1.1) and  $\mathcal{W}_1$  is the normalized Wasserstein-1 distance defined in (3.23). The constants  $C, \lambda$  do not depend on  $N, \tau$  or  $p$ .

In the long-time sampling error (1.6), the order of accuracy in the time step  $\tau$  may not be optimal. This is because we have used the strong error estimate (1.5) in the triangle inequality framework to prove (1.6) (see Section 2.3). Nevertheless, the convergence rate  $\lambda$  does not depend on the number of particles  $N$ , the time step  $\tau$  or the batch size  $p$ .

Using the results in the propagation of chaos Huang *et al.* (2006); Chainton & Diez (2022), we show that the discrete RB-IPS (1.4) is also a reliable numerical method for the MVP (1.2). Under appropriate conditions on the drift force  $b$  and the interaction force  $K$  (see Corollary 4.7), the invariant distribution of the MVP (1.2) is unique, and can be approximated by the empirical distribution of the discrete RB-IPS (1.4) by choosing the number of particles  $N$  sufficiently large.

1. **(Corollary 4.3) The finite-time strong error is  $O(\sqrt{\tau} + \frac{1}{\sqrt{N}})$ .**

When  $N$  duplicates of the MVP (1.2) and the discrete RB-IPS (1.4) are driven by the same initial value and the same Wiener processes, there exists a positive constant  $C = C(T)$  such that

$$\sup_{0 \leq n \leq T/\tau} \left( \frac{1}{N} \sum_{i=1}^N \mathbb{E} |\bar{X}_{n\tau}^i - \tilde{Y}_n^i|^2 \right) \leq C\tau + \frac{C}{N}, \quad (1.7)$$

where  $\{\bar{X}_t^i\}_{t \geq 0}$  is the  $i$ th duplicate of the MVP (1.2). The constant  $C$  does not depend on  $N, \tau$  or  $p$ .

2. **(Corollary 4.9) The long-time sampling error is  $O(\sqrt{\tau} + e^{-\lambda t} + \frac{1}{\sqrt{N}})$ .**

When the interaction force  $K$  is moderately large, there exist constants  $C, \lambda > 0$  such that

$$\mathbb{E}[\mathcal{W}_1(\bar{\pi}, \tilde{\mu}_{n\tau}^{\text{RB}})] \leq C\sqrt{\tau} + Ce^{-\lambda n\tau} + \frac{C}{\sqrt{N}}, \quad \forall n \geq 0, \quad (1.8)$$

where  $\bar{\pi} \in \mathcal{P}(\mathbb{R}^d)$  is the invariant distribution of the MVP (1.2), and  $\tilde{\mu}_{n\tau}^{\text{RB}}$  is the empirical measure of the  $N$ -particle system  $\{\tilde{Y}_n^i\}_{i=1}^N$ , i.e.,

$$\tilde{\mu}_{n\tau}^{\text{RB}}(x) = \frac{1}{N} \sum_{i=1}^N \delta(x - \tilde{Y}_n^i) \in \mathcal{P}(\mathbb{R}^d). \quad (1.9)$$

The constants  $C, \lambda$  do not depend on  $N, \tau$  or  $p$ .

The paper is organized as follows. In Section 2, we introduce the triangle inequality framework for estimating the long-time sampling error. In Section 3, we prove (1.5) and (1.6) for the IPS (1.1). In Section 4, we prove (1.7) and (1.8) for the MVP (1.2).

## 2. Triangle inequality for long-time error analysis

In general, the long-time error analysis of a numerical method is much more difficult than the finite-time error analysis, whose proof is standard and can be found in textbooks, e.g., Chapter 7.5 of [Weinan et al. \(2021\)](#). Nevertheless, [McKean \(1967\)](#); [Mattingly et al. \(2010\)](#) and [Eberle \(2011\)](#) employed a special strategy—which we refer to as the triangle inequality framework in this paper—to address the problem of the long-time error analysis. The idea of this framework is simple. In addition to the known results in the finite-time error analysis, one only needs the geometric ergodicity of the stochastic dynamics to perform the long-time error analysis. In short words, the geometric ergodicity with the finite-time error yields the long-time error.

In the rest part of this section, we first review the original approaches employed in [McKean \(1967\)](#); [Mattingly et al. \(2010\)](#); [Eberle \(2011\)](#) for the long-time error analysis. Motivated by their results, we propose a general lemma on the long-time error analysis. Finally, we demonstrate why the triangle inequality framework can be applied in a wide class of stochastic dynamics, including the discrete RB–IPS (1.4).

### 2.1 A historical review

The geometric ergodicity is the key property to describe the long-time behavior of a stochastic process, and is essential to build up the triangle inequality framework. For simplicity, consider the continuous-time Markov process  $\{X_t\}_{t \geq 0}$ , whose transition probability is  $(p_t)_{t \geq 0}$ . Let  $\mathcal{P}(\mathbb{R}^d)$  be the space of all probability distributions on  $\mathbb{R}^d$ , then for any  $\nu \in \mathcal{P}(\mathbb{R}^d)$ ,  $\nu p_t \in \mathcal{P}(\mathbb{R}^d)$  is the distribution law of  $X_t$  provided  $X_0 \sim \nu$ . Given the metric  $d(\cdot, \cdot)$  on  $\mathcal{P}(\mathbb{R}^d)$ , the stochastic process  $\{X_t\}_{t \geq 0}$  is said to have *geometric ergodicity*, if it has an invariant distribution  $\pi \in \mathcal{P}(\mathbb{R}^d)$ , and there exist positive constants  $C, \beta$  such that

$$d(\nu p_t, \pi) \leq C e^{-\beta t} d(\nu, \pi), \quad \forall \nu \in \mathcal{P}(\mathbb{R}^d). \quad (2.1)$$

In other words, the distribution law  $\nu p_t$  converges to the invariant distribution  $\pi$  exponentially, and  $\beta$  is the convergence rate. Readers interested in the detailed properties of the transition probabilities and the geometric ergodicity may refer to [dos Reis et al. \(2022\)](#).

Now consider another stochastic process  $\{\tilde{X}_t\}_{t \geq 0}$  with transition probability  $(\tilde{p}_t)_{t \geq 0}$ , which can be viewed as an approximation to the original process  $\{X_t\}_{t \geq 0}$ . For example,  $\{X_t\}_{t \geq 0}$  is the solution to an

SDE, while  $\{\tilde{X}_t\}_{t \geq 0}$  is given by the Euler–Maruyama scheme. To characterize the long-time error of  $\{\tilde{X}_t\}_{t \geq 0}$ , the following two questions are proposed in McKean (1967):

1. Does  $\{\tilde{X}_t\}_{t \geq 0}$  has a unique invariant distribution  $\tilde{\pi} \in \mathcal{P}(\mathbb{R}^d)$ ?
2. If so, what is the difference between  $\pi$  and  $\tilde{\pi}$ ?

The first question can be directly addressed by the Harris ergodic theorem (Fehske *et al.*, 2007; Mattingly *et al.*, 2010; Hauray & Mischler, 2014; Cañizo & Mischler, 2021). For the second question, a special triangle inequality was adopted in McKean (1967) to estimate the difference between  $\pi$  and  $\tilde{\pi}$ . Under the same metric  $d(\cdot, \cdot)$ , assume the *finite-time difference relation* between the distribution laws  $\nu p_t, \nu \tilde{p}_t$  is known, that is, for any  $T > 0$  there exists a constant  $\varepsilon(T)$  such that

$$\sup_{0 \leq t \leq T} d(\nu p_t, \nu \tilde{p}_t) \leq \varepsilon(T), \quad \forall \nu \in \mathcal{P}(\mathbb{R}^d). \quad (2.2)$$

Here,  $T$  is a reference evolution time of the processes  $\{X_t\}_{t \geq 0}, \{\tilde{X}_t\}_{t \geq 0}$ , and  $(p_t)_{t \geq 0}, (\tilde{p}_t)_{t \geq 0}$  are the corresponding transition probabilities. If we choose the metric  $d(\cdot, \cdot)$  to be the Wasserstein-1 distance, and derive the finite-time difference relation (2.2) from the standard strong error estimate, then the error bound  $\varepsilon(T)$  is approximated by

$$\varepsilon(T) \approx O(e^{CT} \sqrt{\tau}), \quad (2.3)$$

where  $\tau > 0$  is the time step used in time discretization. (2.3) implies that  $\varepsilon(T)$  grows exponentially with the evolution time  $T$ , and  $\varepsilon(T)$  is bounded by  $O(\sqrt{\tau})$  with a fixed  $T$ .

Provided the geometric ergodicity (2.1) and the finite-time difference relation (2.2), we can now use the triangle inequality to estimate  $d(\pi, \tilde{\pi})$ . In fact, for any  $T > 0$ , we have

$$\begin{aligned} d(\pi, \tilde{\pi}) &= d(\pi p_T, \tilde{\pi} \tilde{p}_T) \\ &\leq d(\pi p_T, \tilde{\pi} p_T) + d(\tilde{\pi} p_T, \tilde{\pi} \tilde{p}_T) \\ &\leq C e^{-\beta T} d(\pi, \tilde{\pi}) + \varepsilon(T). \end{aligned} \quad (2.4)$$

Hence, if we choose  $T = T_0$  in (2.4) to satisfy  $C e^{-\beta T_0} = 1/2$ , then

$$d(\pi, \tilde{\pi}) \leq 2\varepsilon(T_0), \quad (2.5)$$

which measures the difference between the invariant distributions  $\pi$  and  $\tilde{\pi}$ . Since  $T_0$  is a fixed value, we have  $\varepsilon(T_0) \approx O(\sqrt{\tau})$ , hence we obtain  $d(\pi, \tilde{\pi}) \leq O(\sqrt{\tau})$ .

The triangle inequality used in (2.4) is essentially the same as Remark 6.3 of McKean (1967), and also previously appeared in Shardlow & Stuart (2000); Mattingly *et al.* (2010). The benefit of the triangle inequality (2.4) is obvious: it does not require the ergodicity of the approximation  $\{\tilde{X}_t\}_{t \geq 0}$  to estimate the difference between  $\pi$  and  $\tilde{\pi}$ . It only requires the geometric ergodicity of the original process  $\{X_t\}_{t \geq 0}$ , and the finite-time difference relation (2.2). The drawback of the triangle inequality (2.4) is that it does not tell how fast the distribution law of  $\{\tilde{X}_t\}_{t \geq 0}$  converges to the invariant distribution  $\tilde{\pi} \in \mathcal{P}(\mathbb{R}^d)$ . Although the Harris ergodic theorem ensures that  $\nu \tilde{p}_t$  converges to  $\tilde{\pi}$  exponentially (McKean, 1967;

Mattingly *et al.*, 2010), it is usually difficult to make the convergence rate independent of the time step  $\tau$  (see Theorem 7.3 of Mattingly *et al.* (2010) for example).

In a recent paper (Eberle, 2011), the authors have utilized the geometric ergodicity and the triangle inequality to estimate the long-time sampling error of a given numerical method. Instead of calculating the difference between invariant distributions  $d(\pi, \tilde{\pi})$  directly, one turns to estimate  $d(v\tilde{p}_t, \pi)$  for large  $t$ , that is, the difference between the numerical distribution law  $v\tilde{p}_t$  and the true invariant distribution  $\pi$ . For large  $t$ ,  $d(v\tilde{p}_t, \pi)$  can be interpreted as the long-time sampling error of the approximation  $\{\tilde{X}_t\}_{t \geq 0}$ . Also, one avoids computing the numerical invariant distribution  $\tilde{\pi}$  directly. Although the proof strategies used in McKean (1967); Eberle (2011) are quite different, it is clear that the triangle inequality plays an important role in estimating the long-time sampling error.

Based on the original triangle inequality adopted in McKean (1967), and the idea of using  $d(v\tilde{p}_t, \pi)$  instead of  $d(\pi, \tilde{\pi})$  in Eberle (2011), we describe the triangle inequality framework in the next subsection. By choosing the metric  $d(\cdot, \cdot)$  to be the Wasserstein-1 distance, we expect the long-time sampling error  $d(v\tilde{p}_t, \pi)$  is bounded by

$$d(v\tilde{p}_t, \pi) \leq O(\sqrt{\tau} + e^{-\lambda t}), \quad \forall t > 0, \quad (2.6)$$

where the constant  $\lambda > 0$  does not depend on the time step  $\tau$ . Clearly,  $d(v\tilde{p}_t, \pi)$  consists of two parts: the finite-time strong error  $O(\sqrt{\tau})$  and the exponential convergence part  $O(e^{-\lambda t})$ . Although  $\lambda$  does not indicate the geometric ergodicity of the approximation  $\{\tilde{X}_t\}_{t \geq 0}$  itself, it does reveal the fact that the sampling efficiency of  $\{\tilde{X}_t\}_{t \geq 0}$  can be uniform in the time step  $\tau$ .

We summarize the major differences between our work and the results in McKean (1967); Mattingly *et al.* (2010); Eberle (2011). First, our work considers the numerical methods for the IPS (1.1), which is a multi-particle system rather than a single particle. The geometric ergodicity of the IPS (1.1) is guaranteed by the reflection coupling (Eberle, 2016; Eberle *et al.*, 2019), while their results mainly rely on the Harris ergodic theorem. This also leads to a difference in the choice of the metric  $d(\cdot, \cdot)$ : we shall always employ the normalized Wasserstein-1 distance, while their results mainly involve the weighted total variation (Hauray & Mischler, 2014). Second, the numerical method in our work involves the random batch approximations, which means the update schemes are randomly changing in different time steps, and thus the discrete RB-IPS (1.4) is more complicated than the standard Euler-Maruyama scheme (1.3). Finally, the triangle inequality used in this work is a variant of (2.4) in McKean (1967) rather than the one used in Eberle (2011).

## 2.2 Main lemma for the long-time error estimate

We state the main lemma for the long-time error estimate, which is the key conclusion of the triangle inequality framework.

**LEMMA 2.1** Let  $\{X_t\}_{t \geq 0}, \{\tilde{X}_t\}_{t \geq 0}$  be stochastic processes in  $\mathbb{R}^d$  with transition probabilities  $(p_t)_{t \geq 0}, (\tilde{p}_t)_{t \geq 0}$ . Given the metric  $d(\cdot, \cdot)$  on  $\mathcal{P}(\mathbb{R}^d)$ , assume  $(p_t)_{t \geq 0}$  has an invariant distribution  $\pi \in \mathcal{P}(\mathbb{R}^d)$  and there exist constants  $C, \beta > 0$  such that

$$d(vp_t, \pi) \leq Ce^{-\beta t} d(v, \pi), \quad \forall v \in \mathcal{P}(\mathbb{R}^d);$$

and for any  $T > 0$ , there exists a constant  $\varepsilon(T)$  such that

$$\sup_{0 \leq t \leq T} d(v\tilde{p}_t, vp_t) \leq \varepsilon(T), \quad \forall v \in \mathcal{P}(\mathbb{R}^d).$$



Then there exist constants  $T_0, \lambda > 0$  such that

$$d(v\tilde{p}_t, \pi) \leq 2\varepsilon(T_0) + 2M_0 e^{-\lambda t}, \quad \forall t \geq 0, \quad (2.7)$$

where  $M_0 := \sup_{s \in [0, T_0]} d(v\tilde{p}_s, \pi)$ .

*Proof.* We still estimate  $d(v\tilde{p}_t, \pi)$  using the triangle inequality. For any  $T > 0$  and  $t \geq T$ ,

$$\begin{aligned} d(v\tilde{p}_t, \pi) &\leq d(v\tilde{p}_{t-T}\tilde{p}_T, v\tilde{p}_{t-T}p_T) + d(v\tilde{p}_{t-T}p_T, \pi p_T) \\ &\leq \varepsilon(T) + Ce^{-\beta T} d(v\tilde{p}_{t-T}, \pi). \end{aligned}$$

By choosing  $T = T_0$  such that  $Ce^{-\beta T_0} = 1/2$ , we have

$$d(v\tilde{p}_t, \pi) \leq \varepsilon(T_0) + \frac{1}{2}d(v\tilde{p}_{t-T_0}, \pi), \quad \forall t \geq T_0. \quad (2.8)$$

By induction on the integer  $n \geq 0$ , we obtain

$$d(v\tilde{p}_t, \pi) \leq 2\left(1 - \frac{1}{2^n}\right)\varepsilon(T_0) + \frac{1}{2^n}d(v\tilde{p}_{t-nT_0}, \pi), \quad \forall t \geq nT_0. \quad (2.9)$$

For any  $t \in [0, +\infty)$ , there exists a unique integer  $n \geq 0$  such that  $t \in [nT_0, (n+1)T_0)$ . Then

$$d(v\tilde{p}_t, \pi) \leq 2\varepsilon(T_0) + 2^{1-t/T_0} \sup_{s \in [0, T_0]} d(v\tilde{p}_s, \pi), \quad (2.10)$$

which implies the long-time error estimate (2.7) with  $\lambda = \ln 2/T_0$ .  $\square$

The conditions in Lemma 2.1 are exactly the geometric ergodicity (2.1) and the finite-time difference relation (2.2), and the result (2.7) characterizes the long-time sampling error of the stochastic process  $\{\tilde{X}_t\}_{t \geq 0}$ . The triangle inequality used in (2.8) is essential in the proof of Lemma 2.1, which is the reason that Lemma 2.1 is referred to as the triangle inequality framework. In particular, when  $d(\cdot, \cdot)$  is the Wasserstein-1 distance,  $\varepsilon(T_0)$  is of order  $O(\sqrt{\tau})$ , and thus we recover the result in (2.6). Now we briefly summarize the pros and cons of the triangle inequality framework.

1. It requires only the geometric ergodicity of the original dynamics  $\{X_t\}_{t \geq 0}$ . The existence of the invariant distribution for  $\{\tilde{X}_t\}_{t \geq 0}$  is not required. This allows us to study the long-time behavior of a wide class of numerical methods, including the discrete IPS (1.3) and discrete RB-IPS (1.4). It is also possible to apply the triangle inequality framework in an SDE with nonconstant diffusions.
2. It produces an explicit convergence rate  $\lambda > 0$ , which can be easily made independent of the time step  $\tau$  and other parameters. In fact,  $\lambda$  is uniquely determined by the parameters  $C, \beta$  in the geometric ergodicity condition (2.1). In the discrete IPS (1.3) and the discrete RB-IPS (1.4),  $\lambda$  is independent of the number of particles  $N$ .



3. The geometric ergodicity condition (2.1) is very restrictive in the choice of the metric  $d(\cdot, \cdot)$ . Here, the metric  $d(\cdot, \cdot)$  must be symmetric in its two arguments and satisfy the triangle inequality. As a consequence, the convergence in entropy

$$H(\nu p_t | \pi) \leqslant C e^{-\beta t} H(\nu | \pi), \quad \forall \nu \in \mathcal{P}(\mathbb{R}^d) \quad (2.11)$$

cannot be used to prove the geometric ergodicity condition (2.1) because the relative entropy  $H(\cdot | \pi)$  is not symmetric, despite the fact that it is stronger than the Wasserstein distance (Talagrand's inequality; Bobkov *et al.*, 2001) and the total variation (Pinsker's inequality).

4. The finite-time difference relation (2.2) must be derived with a metric at least stronger than  $d$ , which might make the order of accuracy not optimal. For example, when  $d$  is the Wasserstein-1 distance, (2.2) can be naturally derived from the strong error estimate, but the order of accuracy is only  $O(\sqrt{\tau})$ . It is still challenging for the triangle inequality framework to yield better accuracy in the time step.

In short words, as long as the original dynamics,  $\{X_t\}_{t \geq 0}$  satisfies the geometric ergodicity condition (2.1) in a specific metric  $d(\cdot, \cdot)$ , and the finite-time error analysis is valid in the metric  $d(\cdot, \cdot)$ , then we can use the triangle inequality to estimate the long-time sampling error. For example,  $\{X_t\}_{t \geq 0}$  can be the IPS (1.1), the MVP (1.2) or the Hamiltonian Monte Carlo, where the geometric ergodicity of  $\{X_t\}_{t \geq 0}$  is guaranteed by the reflection coupling (Fehske *et al.*, 2007; Eberle *et al.*, 2019; Bou-Rabee *et al.*, 2020).

Finally, we remark that the triangle inequality framework is remotely reminiscent of the well-known Lax equivalence theorem in numerical analysis. Here, the geometric ergodicity serves as the stability and it helps translate the finite-time error estimate to the long-time error estimate without sacrificing the accuracy order.

### 2.3 Application in the interacting particle system

A significant advantage of the triangle inequality framework is that it naturally applies to the IPS (not necessarily in the form of (1.1)). When sampling an IPS, we expect the error bound to be independent of the number of particles  $N$ . This is in general a difficult problem in stochastic analysis, and even more in the case of the long-time sampling error. Nevertheless, the requirement of the uniform-in- $N$  error bound can be explicitly interpreted in the triangle inequality framework.

In order to make the long-time sampling error (2.7) independent of the number of particles  $N$ , we need to satisfy the following two conditions:

1. The finite-time error bound  $\varepsilon(T_0)$  is independent of  $N$  (for fixed  $T_0$ ).
2. The exponential convergence rate  $\beta$  is independent of  $N$ .

The first condition is relatively easy to obtain because  $\varepsilon(T_0)$  only relates to the finite-time error analysis. If the IPS has a mean-field limit as  $N \rightarrow \infty$ , the theory of the propagation of chaos usually provides a convenient tool to study  $\varepsilon(T_0)$ , see Malrieu (2003); Durmus & Moulines (2017); Bao & Huang (2022) for example.

The second condition is more demanding because it requires the IPS to have uniform geometric ergodicity in a specific metric  $d(\cdot, \cdot)$ . The Harris ergodic theorem is not suitable to prove the uniform ergodicity because it is difficult to quantify the minorization condition in high dimensions, Fehske *et al.* (2007). The uniform log-Sobolev inequality proved in Ha *et al.* (2021) has a uniform-in- $N$  convergence rate, but the relative entropy used to quantify the convergence is not a metric. Therefore, the most natural

TABLE 1 Notations of IPS, MVP, discrete IPS, RB-IPS and discrete RB-IPS

dynamics	notation	transition prob.	invariant dist.	equation
IPS	$X_t \in \mathbb{R}^{Nd}$	$(p_t)_{t \geq 0}$	$\pi \in \mathcal{P}(\mathbb{R}^{Nd})$	(1.1)
MVP	$\tilde{X}_t \in \mathbb{R}^d$	$(\tilde{p}_t)_{t \geq 0}$	$\tilde{\pi} \in \mathcal{P}(\mathbb{R}^d)$	(1.2)
discrete IPS	$\tilde{X}_n \in \mathbb{R}^{Nd}$	$(\tilde{p}_{n\tau})_{n \geq 0}$	–	(1.3)
RB-IPS	$Y_t \in \mathbb{R}^{Nd}$	$(q_{n\tau})_{n \geq 0}$	–	(3.1)
discrete RB-IPS	$\tilde{Y}_n \in \mathbb{R}^{Nd}$	$(\tilde{q}_{n\tau})_{n \geq 0}$	–	(1.4)

choice for the metric  $d(\cdot, \cdot)$  in the IPS is the Wasserstein distance, and the uniform geometric ergodicity can be verified by the reflection coupling (Eberle, 2016; Eberle *et al.*, 2019).

When both conditions are satisfied, we can use the triangle inequality framework to estimate the long-time sampling error of a large variety of numerical methods, although the order of accuracy is not optimal. In particular, for the first time, we show that the discrete RB-IPS (1.4), as a time discretization of the RBM, possesses a long-time error bound independent of the number of particles  $N$ .

### 3. Error analysis of discrete RB-IPS for IPS

In this section, we analyze the error of the discrete RB-IPS (1.4), as an approximation to the IPS (1.1). In Section 3.1, we derive the strong error in the finite time. In Section 3.2, we prove the uniform-in-time moment estimates for the discrete RB-IPS (1.4), which is necessary for the long-time error estimate. In Section 3.3, we briefly review the geometric ergodicity of the IPS (1.1) derived by the reflection coupling. In Section 3.4, we combine the results above with the triangle inequality framework to derive the long-time error in the normalized Wasserstein-1 distance.

For the convenience of analysis, we also introduce the continuous-time RB-IPS, which is represented by a system of particles  $Y_t = \{Y_t^i\}_{i=1}^N$  in  $\mathbb{R}^{Nd}$ , where each particle  $Y_t^i \in \mathbb{R}^{Nd}$  in the time interval  $[t_n, t_{n+1})$  is evolved by the following SDE

$$dY_t^i = \left( b(Y_t^i) + \frac{1}{p-1} \sum_{j \neq i, j \in \mathcal{C}} K(Y_t^i - Y_t^j) \right) dt + \sigma dW_t^i, \quad i \in \mathcal{C}, \quad t \in [t_n, t_{n+1}). \quad (3.1)$$

Here,  $\mathcal{D} = \{\mathcal{C}_1, \dots, \mathcal{C}_q\}$  is the batch division used in the time interval  $[t_n, t_{n+1})$ ; and for each  $i \in \{1, \dots, N\}$ ,  $\mathcal{C} \in \mathcal{D}$  is the unique batch that contains  $i$ . The error analysis for the RB-IPS (3.1) can be found in Klovov & Veretennikov (2006); Jin *et al.* (2021b).

We list in Table 1 the notations of all dynamics involved in this paper and their corresponding transition probabilities, invariant distributions and equation numbers.

Here, ‘–’ in the invariant distribution column means that the existence of such distribution is not required in our analysis.

#### 3.1 Strong error in finite time

The discrete RB-IPS (1.4) deviates from the IPS (1.1) for two reasons: time discretization and random batch divisions at each time step. Therefore, it is natural to analyze the impact of these two factors

separately. Among the four dynamics: IPS (1.1), discrete IPS (1.3), RB-IPS (3.1) and discrete RB-IPS (1.4), we focus on the following two types of strong error estimates.

1. Time discretization.

$$\text{discrete RB-IPS vs RB-IPS: } \sup_{0 \leq n \leq T/\tau} \left( \frac{1}{N} \sum_{i=1}^N \mathbb{E} |Y_{n\tau}^i - \tilde{Y}_n^i|^2 \right). \quad (3.2)$$

2. Random batch divisions.

$$\text{RB-IPS vs IPS: } \sup_{0 \leq n \leq T/\tau} \left( \frac{1}{N} \sum_{i=1}^N \mathbb{E} |X_{n\tau}^i - Y_{n\tau}^i|^2 \right). \quad (3.3)$$

Here, we assume the four dynamics (1.1), (1.3), (3.1) and (1.4) are in the synchronous coupling, i.e., they are driven by the same Wiener processes  $\{W_t^i\}_{i=1}^N$ , the same random batch divisions (if required) at each time step, and the same initial value  $X_0$ , where  $X_0$  is a random variable on  $\mathbb{R}^{Nd}$  with  $\text{Law}(X_0) = \nu$ . Note that the time discretization is the only reason that the discrete RB-IPS (1.4) deviates from the RB-IPS (3.1) because we impose the same random batch divisions for these two dynamics.

Once we obtain the strong errors (3.2) and (3.3), the strong error between the discrete RB-IPS (1.4) and the IPS (1.1) defined by

$$\text{discrete RB-IPS vs IPS: } \sup_{0 \leq n \leq T/\tau} \left( \frac{1}{N} \sum_{i=1}^N \mathbb{E} |X_{n\tau}^i - \tilde{Y}_n^i|^2 \right) \quad (3.4)$$

directly follows from the triangle inequality. In the following, we estimate (3.2) and (3.3), respectively.

Before we begin to estimate (3.2), it is convenient to introduce the strong error below

$$\text{discrete IPS vs IPS: } \sup_{0 \leq n \leq T/\tau} \left( \frac{1}{N} \sum_{i=1}^N \mathbb{E} |X_{n\tau}^i - \tilde{X}_n^i|^2 \right). \quad (3.5)$$

Since the both (3.2) and (3.5) origin from time discretization, we may apply similar methods to estimate (3.2) and (3.5). As in the standard routine in the strong error analysis, we impose the global Lipschitz and boundedness condition on the drift force  $b$  and the interaction force  $K$  as follows.

**ASSUMPTION 3.1** For the drift force  $b : \mathbb{R}^d \rightarrow \mathbb{R}^d$ , there exists a constant  $L_0$  such that

$$|b(x)| \leq L_0(|x| + 1), \quad |\nabla b(x)| \leq L_0, \quad \forall x \in \mathbb{R}^d. \quad (3.6)$$

For the interaction force  $K : \mathbb{R}^d \rightarrow \mathbb{R}^d$ , there exists a constant  $L_1$  such that

$$\max\{|K(x)|, |\nabla K(x)|, |\nabla^2 K(x)|\} \leq L_1, \quad \forall x \in \mathbb{R}^d. \quad (3.7)$$

In the IPS (1.1), define the perturbation force of the  $i$ th particle by

$$\gamma^i(x) := \frac{1}{N-1} \sum_{j \neq i} K(x^i - x^j), \quad \forall x \in \mathbb{R}^{Nd}. \quad (3.8)$$

Define the total force applied to the  $i$ th particle by  $b^i(x) = b(x^i) + \gamma^i(x)$ . Then the IPS (1.1) and the discrete IPS (1.3) can be simply written as

$$dX_t^i = b^i(X_t) dt + \sigma dW_t^i, \quad \tilde{X}_{n+1}^i = \tilde{X}_n^i + b^i(\tilde{X}_n) \tau + \sigma(W_{t_{n+1}}^i - W_{t_n}^i), \quad i = 1, \dots, N. \quad (3.9)$$

According to (3.7), it is easy to verify  $\gamma^i(x)$  is uniformly bounded by  $L_1$ , and

$$|\gamma^i(x) - \gamma^i(y)| \leq L_1 |x^i - y^i| + \frac{L_1}{N-1} \sum_{j \neq i} |x^j - y^j|. \quad (3.10)$$

Summation over  $i$  yields the global Lipschitz condition for the perturbation force

$$\sum_{i=1}^N |\gamma^i(x) - \gamma^i(y)| \leq 2L_1 \sum_{i=1}^N |x^i - y^i|, \quad \forall x, y \in \mathbb{R}^{Nd}. \quad (3.11)$$

In the RB-IPS (3.1), suppose the index set  $\{1, \dots, N\}$  is divided to  $\mathcal{D} = \{\mathcal{C}_1, \dots, \mathcal{C}_q\}$  to form the random batch dynamics in the time interval  $[t_n, t_{n+1})$ . In this case, we slightly abuse the notation and again define the perturbation force by

$$\gamma^i(x) = \frac{1}{p-1} \sum_{j \neq i, j \in \mathcal{C}} K(x^i - x^j), \quad \forall x \in \mathbb{R}^{Nd}, \quad (3.12)$$

then with the new total force  $b^i(x) = b(x^i) + \gamma^i(x)$ , the RB-IPS (3.1) and the discrete RB-IPS (1.4) are simply given by

$$dY_t^i = b^i(Y_t) dt + \sigma dW_t^i, \quad \tilde{Y}_{n+1}^i = \tilde{Y}_n^i + b^i(\tilde{Y}_n) \tau + \sigma(W_{t_{n+1}}^i - W_{t_n}^i), \quad i = 1, \dots, N. \quad (3.13)$$

Although (3.13) is very similar to (3.9), we stress that the dynamics of  $Y_t$  in (3.13) is valid only in the time step  $[t_n, t_{n+1})$  due to the random batch divisions, and the formulation of  $\gamma^i(x)$  varies in different time steps. Nevertheless,  $\gamma^i(x)$  is uniformly bounded by  $L_1$  regardless of the batch division  $\mathcal{D}$ . Also,

$$|\gamma^i(x) - \gamma^i(y)| \leq L_1 |x^i - y^i| + \frac{L_1}{p-1} \sum_{j \neq i, j \in \mathcal{C}} |x^j - y^j|. \quad (3.14)$$

Summation over  $i \in \mathcal{C}$  gives

$$\sum_{i \in \mathcal{C}} |\gamma^i(x) - \gamma^i(y)| \leq 2L_1 \sum_{i \in \mathcal{C}} |x^i - y^i|, \quad (3.15)$$

and summation over  $\mathcal{C} \in \mathcal{D}$  gives

$$\sum_{i=1}^N |\gamma^i(x) - \gamma^i(y)| \leq 2L_1 \sum_{i=1}^N |x^i - y^i|. \quad (3.16)$$

Therefore, the global Lipschitz condition still holds true for the random batch dynamics.

Based on the observation of  $\gamma^i(x)$  above, we can prove the following results.

LEMMA 3.2 Under Assumption 3.1, if there exists a constant  $M_2$  such that

$$\max_{1 \leq i \leq N} \mathbb{E}|X_0^i|^2 \leq M_2,$$

then there exists a constant  $C = C(L_0, L_1, M_2, T, \sigma)$  such that

$$\sup_{0 \leq t \leq T} \mathbb{E}|X_t^i|^2 \leq C, \quad \sup_{t \in [t_n, t_{n+1} \wedge T)} \mathbb{E}|X_t^i - X_{t_n}^i|^2 \leq C\tau, \quad (3.17)$$

and

$$\sup_{0 \leq t \leq T} \mathbb{E}|Y_t^i|^2 \leq C, \quad \sup_{t \in [t_n, t_{n+1} \wedge T)} \mathbb{E}|Y_t^i - Y_{t_n}^i|^2 \leq C\tau. \quad (3.18)$$

The proof of Lemma 3.2 is in Appendix A. The proof only requires the fact that  $|\gamma^i(x)| \leq L_1$ .

THEOREM 3.3 Under Assumption 3.1, if there exists a constant  $M_2$  such that

$$\max_{1 \leq i \leq N} \mathbb{E}|X_0^i|^2 \leq M_2,$$

then there exists a constant  $C = C(L_0, L_1, M_2, T, \sigma)$  such that

$$\sup_{0 \leq n \leq T/\tau} \left( \frac{1}{N} \sum_{i=1}^N \mathbb{E}|X_{n\tau}^i - \tilde{X}_n^i|^2 \right) \leq C\tau \quad (3.19)$$

and

$$\sup_{0 \leq n \leq T/\tau} \left( \frac{1}{N} \sum_{i=1}^N \mathbb{E}|Y_{n\tau}^i - \tilde{Y}_n^i|^2 \right) \leq C\tau. \quad (3.20)$$

The proof of Theorem 3.3 is similar to Lemma 7.21 of Weinan *et al.* (2021) and elaborated in Appendix A. The proof uses the fact that  $\gamma^i(x)$  is global Lipschitz.

REMARK 3.4 We have some remarks on Theorem 3.3.

1. If one employs a constant time step  $\tau$ , the global Lipschitz condition on the drift force  $b$  is necessary to ensure the stability of the numerical method. Even for an ergodic SDE, the Euler–Maruyama scheme can be unstable when  $b$  is not globally Lipschitz, see the example in Section

6.3 of Mattingly *et al.* (2010). If there is only local Lipschitz condition on  $b$ , the readers may refer to Malrieu (2003); Ding & Qiao (2021) for the discussion of other types of Euler–Maruyama schemes.

2. The constant  $C$  depends on the second moments of the initial distribution  $\nu \in \mathcal{P}(\mathbb{R}^{Nd})$ , which is characterized by the constant  $M_2$  in Theorem 3.3.

Next we compare the trajectory difference between the IPS (1.1) and the RB–IPS (3.1), which are both exactly integrated in the time interval  $[t_n, t_{n+1})$ . Recall that the strong error in this case is

$$\sup_{0 \leq n \leq T/\tau} \left( \frac{1}{N} \sum_{i=1}^N \mathbb{E} |X_{n\tau}^i - Y_{n\tau}^i|^2 \right),$$

where (1.1) and (3.1) are driven by the same Wiener processes  $\{W_t^i\}_{i=1}^N$  and the same initial random variable  $X_0 \sim \nu$ . The estimate of the strong error above directly follows Theorem 3.1 in Jin *et al.* (2021b), and we restate this result here.

**THEOREM 3.5** Under Assumption 3.1, if there exists a constant  $M_4$  such that

$$\max_{1 \leq i \leq N} \mathbb{E} |X_0^i|^4 \leq M_4,$$

then there exists a constant  $C = C(L_0, L_1, M_4, T, \sigma)$  such that

$$\sup_{0 \leq t \leq T} \left( \frac{1}{N} \sum_{i=1}^N \mathbb{E} |X_t^i - Y_t^i|^2 \right) \leq C \left( \frac{\tau}{p-1} + \tau^2 \right). \quad (3.21)$$

**REMARK 3.6** We have some remarks on Theorem 3.5:

1. Compared to Theorem 3.3, Theorem 3.5 requires the initial distribution  $\nu$  has finite fourth-order moments rather than second-order moments. This is because in the original proof in Jin *et al.* (2021b), the authors used the second-order Taylor expansion to estimate the  $L^2$  norm of

$$K(Y_t^i - Y_t^j) - K(Y_{n\tau}^i - Y_{n\tau}^j),$$

which naturally produces the fourth order moments.

2. If the linear growth condition of  $b(x)$  in (3.6) is replaced by  $|b(x)| \leq L_0(|x| + 1)^q$  for some  $q \geq 2$ , then the initial distribution  $\nu$  should have finite  $2q$ th order moments. In this paper, we only consider the case of  $q = 2$ .

Using Theorems 3.3 and 3.5, we can now estimate the strong error (3.4) between the discrete RB–IPS (1.4) and the IPS (1.1).

**THEOREM 3.7** Under Assumption 3.1, if there exists a constant  $M_4$  such that

$$\max_{1 \leq i \leq N} \mathbb{E} |X_0^i|^4 \leq M_4,$$

then there exists a constant  $C = C(L_0, L_1, M_4, T, \sigma)$  such that

$$\sup_{0 \leq n \leq T/\tau} \left( \frac{1}{N} \sum_{i=1}^N \mathbb{E} |X_{n\tau}^i - \tilde{Y}_n^i|^2 \right) \leq C\tau. \quad (3.22)$$

Theorem 3.7 implies that applying random batch divisions does not worsen the order of strong error. However, the constant  $C$  in Theorem 3.7 can be much larger than in Theorem 3.3.

A direct consequence of Theorem 3.7 is the finite-time error estimate of the discrete RB-IPS (1.4) in the Wasserstein-2 distance. For given distributions  $\mu, \nu \in \mathcal{P}(\mathbb{R}^{Nd})$ , the normalized Wasserstein- $p$  distance between  $\mu, \nu$  is defined by

$$\mathcal{W}_p^p(\mu, \nu) = \inf_{\gamma \in \Pi(\mu, \nu)} \int_{\mathbb{R}^{Nd} \times \mathbb{R}^{Nd}} \left( \frac{1}{N} \sum_{i=1}^N |x^i - y^i|^p \right) \gamma(dx dy), \quad (3.23)$$

where  $\Pi(\mu, \nu)$  is the transport plans between  $\mu$  and  $\nu$ . The readers may also refer to Chaintron & Diez (2022) for a thorough introduction to the normalized Wasserstein distance.

Let  $(p_t)_{t \geq 0}$ ,  $(\tilde{p}_{n\tau})_{n \geq 0}$ ,  $(q_{n\tau})_{n \geq 0}$  and  $(\tilde{q}_{n\tau})_{n \geq 0}$  be the transition probabilities of the four dynamics (1.1), (1.3), (3.1) and (1.4), respectively. Then for given initial distribution  $\nu \in \mathcal{P}(\mathbb{R}^{Nd})$ , the distribution laws of  $X_{n\tau}^i$ ,  $\tilde{X}_n^i$ ,  $Y_{n\tau}^i$  and  $\tilde{Y}_n^i$  are  $\nu p_{n\tau}$ ,  $\nu \tilde{p}_{n\tau}$ ,  $\nu q_{n\tau}$  and  $\nu \tilde{q}_{n\tau}$ , respectively (see the notations in Table 1). Here we note that  $(p_t)_{t \geq 0}$  defines a continuous-time Markov process, while  $(\tilde{p}_{n\tau})_{n \geq 0}$ ,  $(q_{n\tau})_{n \geq 0}$ ,  $(\tilde{q}_{n\tau})_{n \geq 0}$  only define discrete-time Markov chains because of the random batch divisions at each time step. Although, formally, the transition probability  $(q_t)_{t \geq 0}$  for the RB-IPS (3.1) can be defined for any  $t \geq 0$ ,  $(q_t)_{t \geq 0}$  does not form a Markov semigroup.

Now we have the  $\mathcal{W}_2$  error estimate for the discrete IPS (1.3) and the discrete RB-IPS (1.4).

**COROLLARY 3.8** Under Assumption 3.1, if there exists a constant  $M_4$  such that

$$\max_{1 \leq i \leq N} \int_{\mathbb{R}^{Nd}} |x^i|^4 \nu(dx) \leq M_4,$$

then there exists a constant  $C = C(L_0, L_1, M_4, T, \sigma)$  such that

$$\max \left\{ \sup_{0 \leq n \leq T/\tau} \mathcal{W}_2(\nu p_{n\tau}, \nu \tilde{p}_{n\tau}), \sup_{0 \leq n \leq T/\tau} \mathcal{W}_2(\nu p_{n\tau}, \nu \tilde{q}_{n\tau}) \right\} \leq C\sqrt{\tau}. \quad (3.24)$$

Note that the left-hand side (LHS) of (3.24) only involves the transition probabilities  $p_{n\tau}$ ,  $\tilde{p}_{n\tau}$ ,  $\tilde{q}_{n\tau}$ , and does not require the dynamics (1.1), (1.3) and (1.4) to be coupled. This is because the Wasserstein distance compares the distribution laws rather than trajectories.

### 3.2 Uniform-in-time moment estimate

To investigate the long-time behavior of the numerical methods, we need some preliminary results on the moment estimates. Under appropriate dissipation conditions, it can be proved that the discrete IPS (1.3) and the discrete RB-IPS (1.4) have uniform-in-time moment estimates.



ASSUMPTION 3.9 For the drift force  $b : \mathbb{R}^d \rightarrow \mathbb{R}^d$ , there exist constants  $\alpha, \theta > 0$  such that

$$-x \cdot b(x) \geq \alpha |x|^2 - \theta, \quad \forall x \in \mathbb{R}^d. \quad (3.25)$$

The following result is crucial to establish the recurrence relations of both  $\mathbb{E}|\tilde{X}_n^i|^4$  and  $\mathbb{E}|\tilde{Y}_n^i|^4$ .

LEMMA 3.10 Under Assumptions 3.1 and 3.9, let  $f(x, \tau) := x + b(x)\tau$  and  $\tau_0 := \min\{\alpha/(2L_0^2), 1/(2\alpha)\}$ .

1. There exists a constant  $C = C(\alpha, \theta)$  such that if  $\tau < \tau_0$ ,

$$|f(x, \tau)|^4 \leq (1 - \alpha\tau)|x|^4 + C\tau. \quad (3.26)$$

2. For any  $\gamma \in \mathbb{R}^d$  with  $|\gamma| \leq L_1$ , there exists a constant  $C = C(\alpha, \theta, L_1)$  such that if  $\tau < \tau_0$ ,

$$|f(x, \tau) + \gamma\tau|^4 \leq \left(1 - \frac{\alpha\tau}{2}\right)|x|^4 + C\tau. \quad (3.27)$$

The proof of Lemma 3.10 is elementary and is in Appendix A. In Lemma 3.10,  $f(x, \tau) = x + b(x)\tau$  can be viewed as a simplified Euler–Maruyama scheme, where the time step  $\tau$  is restricted to be smaller than  $\tau_0$  to ensure the stability. In the following, we shall always adopt  $\tau_0 := \min\{\alpha/(2L_0^2), 1/(2\alpha)\}$  as the upper bound of the time step  $\tau$ . Note that  $\tau_0$  is uniquely determined from Assumptions 3.1 and 3.9 and does not depend on  $N$ .

Using Lemma 3.10, we have the following uniform-in-time moments estimate for the discrete IPS (1.3) and the discrete RB–IPS (1.4).

THEOREM 3.11 Under Assumptions 3.1 and 3.9, if there exists a constant  $M_4$  such that

$$\max_{1 \leq i \leq N} \mathbb{E}|X_0^i|^4 \leq M_4,$$

and if the time step  $\tau$  satisfies

$$\tau < \min \left\{ \frac{\alpha}{2L_0^2}, \frac{1}{2\alpha} \right\},$$

then there exists a constant  $C = C(\alpha, \theta, L_1, \sigma)$  such that

$$\max \left\{ \sup_{n \geq 0} \mathbb{E}|\tilde{X}_n^i|^4, \sup_{n \geq 0} \mathbb{E}|\tilde{Y}_n^i|^4 \right\} \leq \max\{M_4, C\}, \quad i = 1, \dots, N. \quad (3.28)$$

The proof of Theorem 3.11 is in Appendix A.

Theorem 3.11 tells that when the time step  $\tau < \tau_0$ , the fourth-order moments of the discrete IPS (1.3) and the discrete RB–IPS (1.4) can be bounded uniformly in time.

REMARK 3.12 We have some remarks on Theorem 3.11.

1. We estimate the fourth-order moments of  $\tilde{X}_n^i$  and  $\tilde{Y}_n^i$  rather than the second-order moments because applying Theorem 3.7 requires the initial distribution to have finite fourth-order moments.

2. Utilization of the dissipation condition (3.25) is essential in the proof of Theorem 3.11. From the geometric perspective, the drift force  $b(x)$  pulls the particle  $x \in \mathbb{R}^d$  back when  $x$  is far from the origin, hence the particle system shall stay in a bounded region for most of the time, and the moments are bounded uniformly in time. It can also be proved that, if the initial distribution  $\nu$  has finite moments of order  $2m$  for positive some integer  $m \in \mathbb{N}$ , then  $\mathbb{E}|\tilde{X}_n^i|^{2m}$  and  $\mathbb{E}|\tilde{Y}_n^i|^{2m}$  are bounded uniformly in time. Readers may also refer to Proposition 5.1 of Benachour *et al.* (1998) or Lemma 5.2 of Mattingly *et al.* (2002).
3. The constant  $C = C(\alpha, \beta, L_1)$  in Theorem 3.11 does not depend on  $L_0$ , which is related to the boundedness of the drift force  $b$ . In other words, the moment upper bound is completely controlled by the dissipation condition (3.25).

### 3.3 Geometric ergodicity of IPS

In order to investigate the long-time behavior of the IPS (1.1) and its mean-field limit, the MVP (1.2), it is important that the distribution law of the IPS (1.1) converges to the equilibrium with a convergence rate  $\beta$  independent of the number of particles  $N$ . If the independence of  $\beta$  on  $N$  holds true, hopefully the distribution law of the MVP (1.2) also converges with the convergence rate  $\beta$ , which allows us to prove the geometric ergodicity of the nonlinear MVP (1.2). Therefore, a natural question in studying the ergodicity is to find the conditions ensuring the IPS (1.1) have a convergence rate independent of  $N$ .

On the one hand, the interaction force  $K$  needs to be moderately large to ensure the uniform-in- $N$  convergence rate. If the drift force  $b$  is not the gradient of a strongly convex function, it is well-known that the MVP (1.2) can have multiple invariant distributions when the interaction force  $K$  is too large, see Durmus *et al.* (2020, 2022) for example. In this case the IPS (1.1) must not have a convergence rate independent of  $N$ .

On the other hand, it is sufficient for the interaction force  $K$  to be moderately large to ensure the uniform-in- $N$  convergence rate. To our knowledge, two major approaches to derive the uniform geometric ergodicity of the IPS (1.1) are the log-Sobolev inequality (Ha *et al.*, 2021) and the reflection coupling technique (Fehske *et al.*, 2007; Eberle *et al.*, 2019). Under appropriate dissipation conditions, Ha *et al.* (2021) proves the ergodicity in the sense of relative entropy, while Fehske *et al.* (2007); Eberle *et al.* (2019) proves the ergodicity in the  $\mathcal{W}_1$  distance. Although the relative entropy is stronger than the  $\mathcal{W}_1$  distance, in this paper we shall use the  $\mathcal{W}_1$  distance because it is compatible with the triangle inequality framework.

In the following, we restate the result of the geometric ergodicity of the IPS (1.1) in the  $\mathcal{W}_1$  distance in Eberle *et al.* (2019). The dissipation of the drift force  $b$  is characterized by a function  $\kappa : (0, +\infty) \rightarrow \mathbb{R}$ , satisfying

$$\kappa(r) \leq \left\{ -\frac{2}{\sigma^2} \frac{(x-y) \cdot (b(x) - b(y))}{|x-y|^2} : x, y \in \mathbb{R}^d, |x-y| = r \right\}. \quad (3.29)$$

Assumption 3.9 is now replaced by the asymptotic positivity of  $\kappa(r)$ .

**ASSUMPTION 3.13** The function  $\kappa(r)$  defined in (3.29) satisfies

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

We note that Assumption 3.13 is stronger than Assumption 3.9. In fact, the asymptotic positivity of  $\kappa(r)$  implies that there exist positive constants  $\alpha, \beta > 0$  such that

$$r^2 \kappa(r) \geq \alpha r^2 - \beta, \quad \forall r > 0. \quad (3.30)$$

Then we easily obtain

$$-(x - y) \cdot (b(x) - b(y)) \leq \frac{\sigma^2}{2} (\alpha |x - y|^2 - \beta), \quad (3.31)$$

and thus (3.25) holds. Under Assumption 3.13, we can construct a concave function  $f : [0, +\infty) \rightarrow [0, +\infty)$ , satisfying the following.

LEMMA 3.14 Under Assumption 3.13, there exists a function  $f : [0, +\infty) \rightarrow [0, +\infty)$ , satisfying

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

$$f''(r) - \frac{1}{4} r \kappa(r) f'(r) \leq -\frac{c_0}{2} f(r), \quad \forall r \geq 0. \quad (3.32)$$

3. There exists a constant  $\varphi_0 > 0$  such that

$$\frac{\varphi_0}{4} r \leq f(r) \leq r. \quad (3.33)$$

The proof of Lemma 3.14 can be seen in Theorem 1 of Eberle *et al.* (2019) or Lemma 2.1 of Klokov & Veretennikov (2006). Although Lemma 3.14 serves as part of the proof of the geometric ergodicity for the IPS (1.1) and is not directly related to the topic of this paper, it does provide an explicit upper bound of the constant  $L_1$  in (3.7), which is used in the statement of the main theorem.

Define the space of probability distributions with finite first-order moments by

$$\mathcal{P}_1(\mathbb{R}^{Nd}) = \left\{ \nu \in \mathcal{P}(\mathbb{R}^{Nd}) : \max_{1 \leq i \leq N} \int_{\mathbb{R}^{Nd}} |x^i| \nu(dx) < +\infty \right\}. \quad (3.34)$$

Equipped with the normalized Wasserstein-1 distance,  $(\mathcal{P}_1(\mathbb{R}^{Nd}), \mathcal{W}_1)$  is a complete metric space. Now we have the following result of the geometric ergodicity for the IPS (1.1).

THEOREM 3.15 Under Assumptions 3.1 and 3.13, if the constant  $L_1$  in (3.7) satisfies

$$L_1 < \frac{c_0 \varphi_0 \sigma^2}{16},$$

then for  $\beta := c_0 \sigma^2 / 2$  there exists a positive constant  $C = C(\kappa, \sigma)$  such that

$$\mathcal{W}_1(\mu p_t, \nu p_t) \leq C e^{-\beta t} \mathcal{W}_1(\mu, \nu), \quad \forall t \geq 0 \quad (3.35)$$

for any probability distributions  $\mu, \nu \in \mathcal{P}_1(\mathbb{R}^{Nd})$ .

The proof of Theorem 3.15 can be seen in Corollary 2 of Eberle *et al.* (2019) or Theorem 2.2 of Klokov & Veretennikov (2006).

REMARK 3.16 We have some remarks on Theorem 3.15.

1. Using the reflection coupling technique, we can actually prove that the IPS (1.1) is contractive in the Wasserstein- $f$  distance:

$$\mathcal{W}_f(\mu p_t, \nu p_t) \leq e^{-\beta t} \mathcal{W}_f(\mu, \nu), \quad (3.36)$$

where  $\mathcal{W}_f(\cdot, \cdot)$  is the normalized Wasserstein-1 distance induced by the function  $f$ ,

$$\mathcal{W}_f(\mu, \nu) = \inf_{\gamma \in \Pi(\mu, \nu)} \int_{\mathbb{R}^{Nd} \times \mathbb{R}^{Nd}} \left( \frac{1}{N} \sum_{i=1}^N f(|x^i - y^i|) \right) \gamma(\mathrm{d}x \mathrm{d}y). \quad (3.37)$$

Since  $f(r)$  is equivalent to the usual Euclidean norm, (3.35) is a direct consequence of (3.37).

2. The explicit convergence rate  $\beta = c_0 \sigma^2 / 2$  and the upper bound  $c_0 \varphi_0 \sigma^2 / 16$  only depend on  $\kappa(r)$  and  $\sigma$ . In particular, these parameters do not depend on the number of particles  $N$ . Hence, the IPS (1.1) has an exponential convergence rate independent of  $N$ .
3. The positivity of the diffusion constant  $\sigma$  is essential in the proof by reflection coupling. In fact, for given interaction force  $K$ , the MVP (1.2) can be nonergodic if  $\sigma$  is too small (Del Moral & Tugaut, 2014).

Using Theorem 3.15, the existence and uniqueness of the invariant distribution  $\pi \in \mathcal{P}(\mathbb{R}^{Nd})$  can be derived using the Banach fixed point theorem.

COROLLARY 3.17 Under Assumptions 3.1 and 3.13, if the constant  $L_1$  in (3.7) satisfies

$$L_1 < \frac{c_0 \varphi_0 \sigma^2}{16},$$

then the IPS (1.1) has a unique invariant distribution  $\pi \in \mathcal{P}_1(\mathbb{R}^{Nd})$ , and for  $\beta := c_0 \sigma^2 / 2$  there exist a positive constant  $C = C(\kappa, \sigma)$  such that

$$\mathcal{W}_1(\nu p_t, \pi) \leq C e^{-\beta t} \mathcal{W}_1(\nu, \pi), \quad \forall t \geq 0 \quad (3.38)$$

for any probability distributions  $\nu \in \mathcal{P}_1(\mathbb{R}^{Nd})$ .

The proof of Corollary 3.17 can be seen in Corollary 3 of Eberle *et al.* (2019) or Theorem 3.1 of Klokov & Veretennikov (2006).

### 3.4 Wasserstein-1 error in long time

We estimate the long-time sampling error of the discrete IPS (1.3) and the discrete RB-IPS (1.4) in the  $\mathcal{W}_1$  distance using the triangle inequality and results in previous subsections. We begin with the following induction lemma, which can be viewed as a discrete version of Lemma 2.1.

LEMMA 3.18 Given  $m \in \mathbb{N}$ ,  $\varepsilon > 0$  and  $q \in (0, 1)$ . If a non-negative sequence  $\{a_n\}_{n \geq 0}$  satisfies

$$a_n \leq \varepsilon + qa_{n-m}, \quad \forall n \geq m, \quad (3.39)$$

then

$$a_n \leq \frac{\varepsilon}{1-q} + Mq^{\frac{n}{m}-1}, \quad \forall n \geq 0, \quad (3.40)$$

where  $M = \max_{0 \leq k \leq m-1} a_k$ .

The proof of Lemma 3.18 is in Appendix A. Lemma 3.18 implies that if  $q < 1$  in the recurrence relation (3.39), then the asymptotic form of  $a_n$  is  $O(\varepsilon)$  plus an exponential part.

Combining the finite-time difference relation (3.24) in Corollary 3.8 and the geometric ergodicity in Theorem 3.15, we employ the triangle inequality to estimate the long-time sampling error of the numerical methods (1.3) and (1.4). Recall that the transition probabilities of the discrete IPS (1.3) and the discrete RB-IPS (1.4) are  $\tilde{p}_{n\tau}$  and  $\tilde{q}_{n\tau}$ , respectively.

THEOREM 3.19 Under Assumptions 3.1 and 3.13, if there exists a constant  $M_4$  such that

$$\max_{1 \leq i \leq N} \int_{\mathbb{R}^{Nd}} |x^i|^4 \nu(dx) \leq M_4,$$

and the constant  $L_1$  in (3.7) and the time step  $\tau$  satisfy

$$L_1 < \frac{c_0 \varphi_0 \sigma^2}{16}, \quad \tau < \min \left\{ \frac{\alpha}{2L_0^2}, \frac{1}{2\alpha} \right\},$$

then there exist positive constants  $\lambda = \lambda(\kappa, L_0, \sigma)$  and  $C = C(\kappa, L_0, M_4, \sigma)$  such that

1. The transition probability  $(\tilde{p}_{n\tau})_{n \geq 0}$  of discrete IPS (1.3) satisfies

$$\mathcal{W}_1(v\tilde{p}_{n\tau}, \pi) \leq C\sqrt{\tau} + Ce^{-\lambda n\tau}, \quad \forall n \geq 0. \quad (3.41)$$

2. The transition probability  $(\tilde{q}_{n\tau})_{n \geq 0}$  of discrete RB-IPS (1.4) satisfies

$$\mathcal{W}_1(v\tilde{q}_{n\tau}, \pi) \leq C\sqrt{\tau} + Ce^{-\lambda n\tau}, \quad \forall n \geq 0. \quad (3.42)$$

*Proof.* For any given integers  $n \geq m$ , we have the following triangle inequality

$$\mathcal{W}_1(v\tilde{p}_{n\tau}, \pi) \leq \mathcal{W}_1(v\tilde{p}_{(n-m)\tau}\tilde{p}_{m\tau}, v\tilde{p}_{(n-m)\tau}p_{m\tau}) + \mathcal{W}_1(v\tilde{p}_{(n-m)\tau}p_{m\tau}, \pi p_{m\tau}). \quad (3.43)$$

By Theorem 3.11,  $v\tilde{p}_{(n-m)\tau}$  has uniform-in-time fourth order moment, i.e., there exists a constant  $M'_4 = M'_4(\kappa, M_4, \sigma)$  such that

$$\max_{1 \leq i \leq N} \left\{ \sup_{n \geq m} \int_{\mathbb{R}^{Nd}} |x^i|^4 (v\tilde{p}_{(n-m)\tau})(dx) \right\} \leq M'_4. \quad (3.44)$$

Hence, by Corollary 3.8, there exists a constant  $C_1 = C_1(\kappa, L_0, M_4, m\tau, \sigma)$  such that

$$\mathcal{W}_1(v\tilde{p}_{(n-m)\tau} \tilde{p}_{m\tau}, v\tilde{p}_{(n-m)\tau} p_{m\tau}) \leq C_1 \sqrt{\tau}, \quad \forall n \geq m. \quad (3.45)$$

Note that the constant  $C_1$  depends on the upper bound of  $m\tau$ , which is the evolution time of the IPS (1.1) and the discrete IPS (1.3). By Theorem 3.15, there exists a constant  $C_0 = C_0(\kappa, \sigma)$  such that

$$\mathcal{W}_1(v\tilde{p}_{(n-m)\tau} p_{m\tau}, \pi p_{m\tau}) \leq C_0 e^{-\beta m\tau} \mathcal{W}_1(v\tilde{p}_{(n-m)\tau}, \pi), \quad \forall n \geq m. \quad (3.46)$$

From (3.43), (3.45) and (3.46), we employ the triangle inequality and obtain

$$\mathcal{W}_1(v\tilde{p}_{n\tau}, \pi) \leq C_1 \sqrt{\tau} + C_0 e^{-\beta m\tau} \mathcal{W}_1(v_0 \tilde{p}_{(n-m)\tau}, \pi), \quad \forall n \geq m. \quad (3.47)$$

For given time step  $\tau > 0$ , we wish to choose  $m$  to satisfy  $C_0 e^{-\beta m\tau} = 1/e$ , so that Lemma 3.18 can be applied. However,  $m$  is restricted to be an integer, thus our choice is

$$m = \left\lceil \frac{\log C_0 + 1}{\beta \tau} \right\rceil. \quad (3.48)$$

It is easy to check  $m\tau$  has an upper bound independent of  $\tau$ ,

$$m\tau \leq \left( \frac{\log C_0 + 1}{\beta \tau} + 1 \right) \tau \leq \frac{\log C_0 + 1}{\beta} + \frac{1}{2\alpha}, \quad (3.49)$$

hence the constant  $C_1$  in (3.45) can be made independent of  $\tau$ , i.e.,  $C_1 = C_1(\kappa, L_0, M_4, \sigma)$ . Note that for this choice of  $m$  we have  $C_0 e^{-\beta m\tau} \leq 1/e$ , and (3.47) implies

$$\mathcal{W}_1(v\tilde{p}_{n\tau}, \pi) \leq C_1 \sqrt{\tau} + \frac{1}{e} \mathcal{W}_1(v_0 \tilde{p}_{(n-m)\tau}, \pi), \quad \forall n \geq m. \quad (3.50)$$

Applying Lemma 3.18 with  $a_n := \mathcal{W}_1(v_0 \tilde{p}_{n\tau}, \pi)$ , we have

$$\mathcal{W}_1(v\tilde{p}_{n\tau}, \pi) \leq C_1 \sqrt{\tau} + M_0 e^{1 - \frac{n}{m}}, \quad \forall n \geq 0, \quad (3.51)$$

where the constant

$$M_0 := \sup_{0 \leq k \leq m-1} \mathcal{W}_1(v\tilde{p}_{k\tau}, \pi) \leq \sup_{k \geq 0} \mathcal{W}_1(v\tilde{p}_{k\tau}, \pi). \quad (3.52)$$

Introduce the normalized moment for  $\nu \in \mathcal{P}_1(\mathbb{R}^{Nd})$  by

$$\mathcal{M}_1(\nu) = \int_{\mathbb{R}^{Nd}} \left( \frac{1}{N} \sum_{i=1}^N |x^i| \right) \nu(\mathrm{d}x), \quad (3.53)$$

then the  $\mathcal{W}_1$  distance in (3.52) is bounded by

$$\mathcal{W}_1(\nu \tilde{p}_{k\tau}, \pi) \leq \mathcal{M}_1(\nu \tilde{p}_{k\tau}) + \mathcal{M}_1(\pi). \quad (3.54)$$

On the one hand,  $\nu \tilde{p}_{k\tau}$  has uniform-in-time fourth-order moments, hence there exists a constant  $C_2 = C_2(\kappa, L_0, M_4, \sigma)$  such that

$$\sup_{k \geq 0} \mathcal{M}_1(\nu \tilde{p}_{k\tau}) \leq C_2. \quad (3.55)$$

On the other hand, by Lemma 3.1 of [Klokov & Veretennikov \(2006\)](#), for the invariant distribution  $\pi$  of the IPS (1.1), there exists a constant  $C_2 = C_2(\kappa, L_0, M_4, \sigma)$  such that

$$\mathcal{M}_1(\pi) \leq C_2. \quad (3.56)$$

Combining (3.51)–(3.56), we obtain

$$\mathcal{W}_1(\nu \tilde{p}_{n\tau}, \pi) \leq C_1 \sqrt{\tau} + C_2 e^{-\frac{n}{m}}, \quad \forall n \geq 0, \quad (3.57)$$

where both constants  $C_1, C_2$  only depend on  $\kappa, L_0, M_4, \sigma$ . Note that by the choice of  $m$

$$\frac{n}{m} \geq \frac{n}{\frac{\log C_0 + 1}{c\tau} + 1} \geq \frac{\beta n \tau}{\log C_0 + \beta/(2\alpha) + 1}, \quad (3.58)$$

hence by defining

$$\lambda := \frac{\beta}{\log C_0 + \beta/(2\alpha) + 1}, \quad (3.59)$$

there holds  $e^{-n/m} \leq e^{-\lambda n \tau}$ . Hence, (3.57) implies

$$\mathcal{W}_1(\nu \tilde{p}_{n\tau}, \pi) \leq C \sqrt{\tau} + C e^{-\lambda n \tau}, \quad \forall n \geq 0, \quad (3.60)$$

which is exactly the long-time sampling error. The proof for the discrete RB–IPS (1.4) is the same as above.  $\square$

Theorem 3.19 produces the long-time sampling error of the two numerical methods, the discrete IPS (1.3) and the discrete RB–IPS (1.4), in the  $\mathcal{W}_1$  distance. The error in (3.41) and (3.42) consists of two parts:  $C\sqrt{\tau}$  represents the bias between the invariant distribution  $\pi$  and the asymptotic limit of



$\nu\tilde{p}_{n\tau}$  or  $\nu\tilde{q}_{n\tau}$ , while  $Ce^{-\lambda n\tau}$  represents the exponential convergence of the numerical methods. Here the convergence rate  $\lambda = \lambda(\kappa, L_0, \sigma)$  can be different from the convergence rate  $\beta := c_0\sigma^2/2$  of the IPS (1.1). Still,  $\lambda$  is independent of the number of particles  $N$ , the time step  $\tau$ , the batch size  $p$  and the choice of the initial distribution  $\nu$ .

REMARK 3.20 We have some remarks on Theorem 3.19.

1. Assumption 3.9 is a corollary of Assumption 3.13, and the constants  $\alpha, \theta$  in (3.25) can be directly derived from Assumption 3.13.
2. The constant  $C$  in (3.42) depends on  $M_4$ , the fourth-order moments of initial distribution  $\nu$ . However, the convergence rate  $\lambda$  does not depend on  $M_4$ . In practical simulation, one may choose the initial distribution as the Dirac distribution centered at origin to sample the invariant distribution  $\pi$ , and in this case the dependence of  $C$  on  $M_4$  can be ignored.
3. Since we are studying the long-time behavior of the numerical methods, it is natural to ask: do the numerical methods (1.3) and (1.4) have invariant distributions? If so, does the convergence rate depends on  $N$ ? The existence of the invariant distributions can be proved by the Harris ergodic theorem under appropriate Lyapunov conditions, see Mattingly *et al.* (2010); Eberle (2011) Hauray & Mischler (2014);. However, the convergence rate derived from the Harris ergodic theorem is very implicit. Still, there are a few results which proved that the convergence rate of the numerical method can be independent of  $N$ , under global boundedness condition of the drift force  $b$  (Lacker, 2018), which are too strong for practical use. In this paper, we follow the idea in Eberle (2011) and avoid discussing the geometric ergodicity of the numerical methods themselves.

#### 4. Error analysis of discrete RB-IPS for MVP

In this section, we analyze the error of the discrete RB-IPS (3.1), which is a numerical approximation to the MVP (1.2). Thanks to the theory of propagation of chaos, we can easily extend our results in Section 3 for the IPS (1.1) to the case of the MVP (1.2). Nevertheless, we should be careful that the major difference between the IPS (1.1) and the MVP (1.2) is the *linearity* of the transition probability, as we illustrate follows.

The transition probability  $(p_t)_{t \geq 0}$  of the IPS (1.1) forms a linear semigroup, that is,

1. For any  $\nu \in \mathcal{P}(\mathbb{R}^{Nd})$ ,  $(\nu p_t)p_s = \nu p_{t+s}$ .
2. For any  $t > 0$ , the mapping  $\nu \mapsto \nu p_t$  is linear in  $\nu \in \mathcal{P}(\mathbb{R}^{Nd})$ .

Denote the transition probability of the MVP (1.2) by  $(\bar{p}_t)_{t \geq 0}$ , then for any  $\nu \in \mathcal{P}(\mathbb{R}^d)$ ,  $\nu\bar{p}_t$  is the distribution law of  $\bar{X}_t$  in the MVP (1.2). Although  $(\bar{p}_t)_{t \geq 0}$  still satisfies the semigroup property  $(\nu\bar{p}_t)\bar{p}_s = \nu\bar{p}_{t+s}$ ,  $(\bar{p}_t)_{t \geq 0}$  does not form a linear semigroup, because the MVP (1.2) is a distribution dependent SDE, and thus the mapping  $\nu \mapsto \nu\bar{p}_t$  is nonlinear. Readers may also refer to Wang (2018) for a complete guide to distribution-dependent SDEs and nonlinear semigroups.

##### 4.1 Strong error in finite time

To estimate the strong error between the IPS (1.1) in  $\mathbb{R}^{Nd}$  and the MVP (1.2) in  $\mathbb{R}^d$ , we need to define the synchronous coupling between (1.1) and (1.2). Given the initial distribution  $\nu \in \mathcal{P}(\mathbb{R}^d)$  and  $N$

independent Wiener processes  $\{W_t^i\}_{i=1}^N$ , the strong solution to the IPS (1.1) is

$$X_t^i = X_0^i + \int_0^t \left( b(X_s^i) + \frac{1}{N-1} \sum_{j \neq i} K(X_s^i - X_s^j) \right) ds + \sigma W_t^i, \quad i = 1, \dots, N, \quad (4.1)$$

where the initial value  $\{X_0^i\}_{i=1}^N$  are sampled from  $\nu$  independently. Introduce  $N$  duplicates of the MVP (1.2) represented by  $\{\bar{X}_t^i\}_{i=1}^N$ , where each  $\bar{X}_t^i$  is the strong solution to the SDE

$$\bar{X}_t^i = X_0^i + \int_0^t \left( b(\bar{X}_s^i) + (K * \text{Law}(\bar{X}_s^i))(\bar{X}_s^i) \right) ds + \sigma dW_t^i, \quad i = 1, \dots, N. \quad (4.2)$$

Here, ‘ $*$ ’ denotes the convolution of a density kernel with a probability distribution,

$$(K * \mu)(x) = \int_{\mathbb{R}^d} K(x - z) \mu(dz). \quad (4.3)$$

It can be observed from (4.1) and (4.2) that each  $\bar{X}_t^i$  uses the same initial value  $X_0^i \sim \nu$  and the same Wiener process  $W_t^i$  with  $X_t^i$ . The major difference between (4.1) and (4.2) is that the particles in  $\{X_t^i\}_{i=1}^N$  are interacting with each other, while the particles in  $\{\bar{X}_t^i\}_{i=1}^N$  are fully decoupled, i.e., the evolution of the  $N$  particles in  $\{\bar{X}_t^i\}_{i=1}^N$  is mutually independent.

The estimate of the strong error between (4.1) and (4.2) is a classical topic in the theory of propagation of chaos. McKean proved the following result in his celebrated work (Meyn & Tweedie, 2012).

**THEOREM 4.1** Under Assumption 3.1, there exists a constant  $C = C(L_0, L_1, T, \sigma)$  such that

$$\frac{1}{N} \sum_{i=1}^N \mathbb{E} \left[ \sup_{t \leq T} |X_t^i - \bar{X}_t^i|^2 \right] \leq \frac{C}{N}. \quad (4.4)$$

As in the synchronous coupling, the expectation is taken over the Wiener processes  $\{W_t^i\}_{i=1}^N$  in the time interval  $[0, T]$  and the random variables  $\{X_0^i\}_{i=1}^N$ . We note that the IPS (1.1) in this paper is slightly different from the original setting in Meyn & Tweedie (2012), where the perturbation force  $\gamma^i(x)$  is given by

$$\gamma^i(x) = \frac{1}{N} \sum_{j=1}^N K(x^i - x^j) \quad (4.5)$$

rather than

$$\gamma^i(x) = \frac{1}{N-1} \sum_{j \neq i} K(x^i - x^j). \quad (4.6)$$

This minor difference in the choice of  $\gamma^i$  does not impact the final result of propagation of chaos. The proof of Theorem 4.1 under the settings (4.5) and (4.6) can be found in Theorem 3.1 of Chaintron & Diez (2022) and Proposition 4.2 of Jin *et al.* (2020), respectively.

Combining Theorems 3.3 and 4.1, we directly obtain the strong error of the discrete IPS (1.3).

COROLLARY 4.2 Under Assumption 3.1, if there exists a constant  $M_2$  such that

$$\int_{\mathbb{R}^d} |x|^2 v(dx) \leq M_2,$$

then there exist constants  $C_1 = C_1(L_0, L_1, M_2, T, \sigma)$  and  $C_2 = C_2(L_0, L_1, T, \sigma)$  such that

$$\sup_{0 \leq n \leq T/\tau} \left( \frac{1}{N} \sum_{i=1}^N \mathbb{E} |\bar{X}_{n\tau}^i - \tilde{X}_n^i|^2 \right) \leq C_1 \tau + \frac{C_2}{N}. \quad (4.7)$$

Combining Theorems 3.7 and 4.1, we obtain the strong error of the discrete RB-IPS (1.4).

COROLLARY 4.3 Under Assumption 3.1, if there exists a constant  $M_4$  such that

$$\int_{\mathbb{R}^d} |x|^4 v(dx) \leq M_4,$$

then there exist constants  $C_1 = C_1(L_0, L_1, M_4, T, \sigma)$  and  $C_2 = C_2(L_0, L_1, T, \sigma)$  such that

$$\sup_{0 \leq n \leq T/\tau} \left( \frac{1}{N} \sum_{i=1}^N \mathbb{E} |\bar{X}_{n\tau}^i - \tilde{Y}_n^i|^2 \right) \leq C_1 \tau + \frac{C_2}{N}. \quad (4.8)$$

In the  $\mathcal{W}_2$  distance, the finite-time error of the discrete IPS (1.3) and the discrete RB-IPS (1.4) are estimated as follows.

COROLLARY 4.4 Under Assumption 3.1, if there exists a constant  $M_4$  such that

$$\int_{\mathbb{R}^d} |x|^4 v(dx) \leq M_4,$$

then there exists constant  $C_1 = C_1(L_0, L_1, M_4, T, \sigma)$  and  $C_2 = C_2(L_0, L_1, T, \sigma)$  such that

$$\max \left\{ \sup_{0 \leq n \leq T/\tau} \mathcal{W}_2(v^{\otimes N} \bar{p}_{n\tau}^{\otimes N}, v^{\otimes N} \tilde{p}_{n\tau}), \sup_{0 \leq n \leq T/\tau} \mathcal{W}_2(v^{\otimes N} \bar{p}_{n\tau}^{\otimes N}, v^{\otimes N} \tilde{q}_{n\tau}) \right\} \leq C_1 \sqrt{\tau} + \frac{C_2}{\sqrt{N}}. \quad (4.9)$$

Here,  $v^{\otimes N} \in \mathcal{P}(\mathbb{R}^{Nd})$  denotes the tensor product of the distribution  $v \in \mathcal{P}(\mathbb{R}^d)$ , and  $\bar{p}_i^{\otimes N}$  denotes the product of  $\bar{p}_i$  in  $\mathbb{R}^{Nd}$ . Recall that the  $N$  duplicates  $\{\tilde{X}_t^i\}_{i=1}^N$  of the MVP (1.2) are mutually independent, hence  $v^{\otimes N} \bar{p}_{n\tau}^{\otimes N} = (v \bar{p}_{n\tau})^{\otimes N}$ .

Let  $[\mu]_1 \in \mathcal{P}(\mathbb{R}^d)$  denote the marginal distribution of a symmetric distribution  $\mu \in \mathcal{P}(\mathbb{R}^{Nd})$  (see Definition 2.1 of Chainton & Diez, 2022). Note that (4.7) and (4.8) can be written as

$$\max \left\{ \sup_{0 \leq n \leq T/\tau} \mathbb{E} |\bar{X}_{n\tau}^1 - \tilde{X}_n^1|^2, \sup_{0 \leq n \leq T/\tau} \mathbb{E} |\bar{X}_{n\tau}^1 - \tilde{Y}_n^1|^2 \right\} \leq C_1 \tau + \frac{C_2}{N}, \quad (4.10)$$

hence in the sense of marginal distributions we have the following.

COROLLARY 4.5 Under Assumption 3.1, if there exists a constant  $M_4$  such that

$$\int_{\mathbb{R}^d} |x|^4 \nu(dx) \leq M_4,$$

then there exist constants  $C_1 = C_1(L_0, L_1, M_4, T, \sigma)$  and  $C_2 = C_2(L_0, L_1, T, \sigma)$  such that

$$\max \left\{ \sup_{0 \leq n \leq T/\tau} \mathcal{W}_2(v\bar{p}_{n\tau}, [v^{\otimes N} \tilde{p}_{n\tau} 1_1]), \sup_{0 \leq n \leq T/\tau} \mathcal{W}_2(v\bar{p}_{n\tau}, [v^{\otimes N} \tilde{q}_{n\tau} 1_1]) \right\} \leq C_1 \sqrt{\tau} + \frac{C_2}{\sqrt{N}}. \quad (4.11)$$

In Corollary 4.5,  $v\bar{p}_{n\tau}$  is the distribution law of the MVP (1.2) and does not depend on  $N$ . Hence, Corollary 4.5 implies that we can obtain the correct distribution law  $v\bar{p}_{n\tau} = \text{Law}(\bar{X}_{n\tau})$  by choosing  $N$  sufficiently large and  $\tau$  sufficiently small.

#### 4.2 Geometric ergodicity of MVP

It has been proved that when the interaction force  $K$  is moderately large, the IPS (1.1) has a convergence rate  $\beta$  uniform in the number of particles  $N$ . Since the MVP (1.2) is the mean-field limit of the IPS (1.1), it is natural to expect that the MVP (1.2) also has the convergence rate  $\beta$ . In fact, the geometric ergodicity of the MVP (1.2) can be directly derived from Theorem 3.15.

THEOREM 4.6 Under Assumptions 3.1 and 3.13, if the constant  $L_1$  in (3.7) satisfies

$$L_1 < \frac{c_0 \varphi_0 \sigma^2}{16},$$

then for  $\beta := c_0 \sigma^2 / 2$  there exist a positive constant  $C = C(\kappa, \sigma)$  such that

$$\mathcal{W}_1(\mu \bar{p}_t, \nu \bar{p}_t) \leq C e^{-\beta t} \mathcal{W}_1(\mu, \nu), \quad \forall t \geq 0 \quad (4.12)$$

for any probability distributions  $\mu, \nu \in \mathcal{P}_1(\mathbb{R}^d)$ .

The proof of Theorem 4.6 is in Appendix A. As a consequence, we can prove that the MVP (1.2) has a unique invariant distribution  $\bar{\pi} \in \mathcal{P}_1(\mathbb{R}^d)$ . Also, the  $\mathcal{W}_1$  distance between the invariant distributions  $\pi \in \mathcal{P}_1(\mathbb{R}^{Nd})$  and  $\bar{\pi} \in \mathcal{P}_1(\mathbb{R}^d)$  can be controlled.

COROLLARY 4.7 Under Assumptions 3.1 and 3.13, if the constant  $L_1$  in (3.7) satisfies

$$L_1 < \frac{c_0 \varphi_0 \sigma^2}{16},$$

then the invariant distribution  $\bar{\pi} \in \mathcal{P}_1(\mathbb{R}^d)$  of the MVP (1.2) is unique, and for  $\beta := c_0 \sigma^2 / 2$  there exists a positive constant  $C = C(\kappa, \sigma)$  such that

$$\mathcal{W}_1(v\bar{p}_t, \bar{\pi}) \leq C e^{-\beta t} \mathcal{W}_1(v, \bar{\pi}), \quad \forall t \geq 0 \quad (4.13)$$

for any  $v \in \mathcal{P}_1(\mathbb{R}^d)$ . Furthermore, there exists a constant  $C = C(\kappa, L_0, \sigma)$  such that

$$\mathcal{W}_1(\bar{\pi}^{\otimes N}, \pi) \leq \frac{C}{\sqrt{N}}. \quad (4.14)$$

The proof of Corollary 4.7 is in Appendix A. We note that (4.14) can also be viewed as the corollary of the uniform-in-time propagation of chaos, see Theorem 2 of Durmus *et al.* (2022) for example.

#### 4.3 Wasserstein-1 error in long time

Combining Theorem 3.19 and Corollary 4.7, we immediately obtain the following result of long-time sampling error of the discrete IPS (1.3) and the discrete RB-IPS (1.4).

**THEOREM 4.8** Under Assumptions 3.1 and 3.13, if there exists a constant  $M_4$  such that

$$\max_{1 \leq i \leq N} \int_{\mathbb{R}^{Nd}} |x^i|^4 v(dx) \leq M_4,$$

and the constant  $L_1$  in (3.7) and the time step  $\tau$  satisfy

$$L_1 < \frac{c_0 \varphi_0 \sigma^2}{16}, \quad \tau < \min \left\{ \frac{\alpha}{2L_0^2}, \frac{1}{2\alpha} \right\},$$

then there exist positive constants  $\lambda = \lambda(\kappa, L_0, \sigma)$ ,  $C_1 = C_1(\kappa, L_0, M_4, \sigma)$  and  $C_2 = C_2(\kappa, L_0, \sigma)$

1. The transition probability  $(\tilde{p}_{n\tau})_{n \geq 0}$  of discrete IPS (1.3) satisfies

$$\mathcal{W}_1(v\tilde{p}_{n\tau}, \bar{\pi}^{\otimes N}) \leq C_1 \sqrt{\tau} + C_1 e^{-\lambda n\tau} + \frac{C_2}{\sqrt{N}}, \quad \forall n \geq 0. \quad (4.15)$$

2. The transition probability  $(\tilde{q}_{n\tau})_{n \geq 0}$  of discrete RB-IPS (1.4) satisfies

$$\mathcal{W}_1(v\tilde{q}_{n\tau}, \bar{\pi}^{\otimes N}) \leq C_1 \sqrt{\tau} + C_1 e^{-\lambda n\tau} + \frac{C_2}{\sqrt{N}}, \quad \forall n \geq 0. \quad (4.16)$$

Using the theory of the propagation of chaos, we may translate the normalized Wasserstein-1 distance in  $\mathcal{P}(\mathbb{R}^{Nd})$  to the Wasserstein-1 distance in  $\mathcal{P}(\mathcal{P}(\mathbb{R}^d))$ . Denote the empirical distributions of the discrete IPS (1.3) and the discrete RB-IPS (1.4) by  $\tilde{\mu}_{n\tau} \in \mathcal{P}(\mathbb{R}^d)$  and  $\tilde{\mu}_{n\tau}^{\text{RB}} \in \mathcal{P}(\mathbb{R}^d)$ , i.e.,

$$\tilde{\mu}_{n\tau}(x) = \frac{1}{N} \sum_{i=1}^N \delta(x - \tilde{X}_{n\tau}^i) \in \mathcal{P}(\mathbb{R}^d), \quad \tilde{\mu}_{n\tau}^{\text{RB}}(x) = \frac{1}{N} \sum_{i=1}^N \delta(x - \tilde{Y}_{n\tau}^i) \in \mathcal{P}(\mathbb{R}^d). \quad (4.17)$$

Since  $\tilde{X}_{n\tau}, \tilde{Y}_{n\tau}$  are random variables with distribution laws  $v\tilde{p}_{n\tau}, v\tilde{q}_{n\tau}$ , the empirical distributions  $\tilde{\mu}_{n\tau}, \tilde{\mu}_{n\tau}^{\text{RB}}$  are actually random measures on  $\mathbb{R}^d$ , and thus their distribution laws  $\text{Law}(\tilde{\mu}_{n\tau}), \text{Law}(\tilde{\mu}_{n\tau}^{\text{RB}})$

can be identified as elements of  $\mathcal{P}(\mathcal{P}(\mathbb{R}^d))$ . By Proposition 2.14 of [Huang \*et al.\* \(2006\)](#), we have

$$\mathcal{W}_1(v\tilde{p}_{n\tau}, \bar{\pi}^{\otimes N}) = \mathcal{W}_1(\text{Law}(\tilde{\mu}_{n\tau}), \delta_{\bar{\pi}}), \quad \mathcal{W}_1(v\tilde{q}_{n\tau}, \bar{\pi}^{\otimes N}) = \mathcal{W}_1(\text{Law}(\tilde{\mu}_{n\tau}^{\text{RB}}), \delta_{\bar{\pi}}). \quad (4.18)$$

Here,  $\mathcal{W}_1$  on the LHS is the normalized Wasserstein-1 distance in  $\mathcal{P}(\mathbb{R}^{Nd})$  defined in (3.23), and  $\mathcal{W}_1$  on the right-hand side (RHS) is the Wasserstein-1 distance in  $\mathcal{P}(\mathcal{P}(\mathbb{R}^d))$  defined in Definition 3.5 of [Chaintron & Diez \(2022\)](#). Since  $\delta_{\bar{\pi}}$  is the Dirac measure in  $\mathcal{P}(\mathcal{P}(\mathbb{R}^d))$ , we have

$$\mathcal{W}_1(\text{Law}(\tilde{\mu}_{n\tau}), \delta_{\bar{\pi}}) = \mathbb{E}[\mathcal{W}_1(\tilde{\mu}_{n\tau}, \bar{\pi})], \quad \mathcal{W}_1(\text{Law}(\tilde{\mu}_{n\tau}^{\text{RB}}), \delta_{\bar{\pi}}) = \mathbb{E}[\mathcal{W}_1(\tilde{\mu}_{n\tau}^{\text{RB}}, \bar{\pi})]. \quad (4.19)$$

Concluding the discussion above, we have the following equivalent form of Theorem 4.8.

**COROLLARY 4.9** Under Assumptions 3.1 and 3.13, if there exists a constant  $M_4$  such that

$$\max_{1 \leq i \leq N} \int_{\mathbb{R}^{Nd}} |x^i|^4 v(dx) \leq M_4,$$

and the constant  $L_1$  in (3.7) and the time step  $\tau$  satisfy

$$L_1 < \frac{c_0 \varphi_0 \sigma^2}{16}, \quad \tau < \min \left\{ \frac{\alpha}{2L_0^2}, \frac{1}{2\alpha} \right\},$$

then there exist positive constants  $\lambda = \lambda(\kappa, L_0, \sigma)$ ,  $C_1 = C_1(\kappa, L_0, M_4, \sigma)$  and  $C_2 = C_2(\kappa, L_0, \sigma)$

$$\max \left\{ \mathbb{E}[\mathcal{W}_1(\tilde{\mu}_{n\tau}, \bar{\pi})], \mathbb{E}[\mathcal{W}_1(\tilde{\mu}_{n\tau}^{\text{RB}}, \bar{\pi})] \right\} \leq C_1 \sqrt{\tau} + C_1 e^{-\lambda n\tau} + \frac{C_2}{\sqrt{N}}, \quad \forall n \geq 0. \quad (4.20)$$

Corollary 4.9 characterizes the long-time sampling error of the numerical methods (1.3) and (1.4) for the MVP (1.2). The error terms in the RHS of (4.20) consist of three parts:

1.  $C_1 \sqrt{\tau}$ : time discretization and random batch divisions;
2.  $C_1 e^{-\lambda n\tau}$ : exponential convergence of the numerical method;
3.  $C_2/\sqrt{N}$ : uniform-in-time propagation of chaos.

If we aim to achieve  $O(\varepsilon)$  error in the  $\mathcal{W}_1$  distance, then the parameters of the numerical methods should be chosen as

$$N = O(\varepsilon^{-2}), \quad \tau = O(\varepsilon^2), \quad n\tau = O(\log \varepsilon^{-1}), \quad (4.21)$$

then for fixed batch size  $p$ , the complexity of the discrete IPS (1.3) and the discrete RB-IPS (1.4) is  $O(\varepsilon^{-6} \log \varepsilon^{-1})$  and  $O(\varepsilon^{-4} \log \varepsilon^{-1})$ , respectively. In this way, the discrete RB-IPS (1.4) consumes less complexity to achieve the desired error tolerance.

## 5. Conclusion

In this paper, we have employed the triangle inequality framework to study the long-time sampling error of the discrete RB-IPS (1.4), and showed that the discrete RB-IPS (1.4) is a reliable numerical approximation to the IPS (1.1) and the MVP (1.2). The triangle inequality framework is a flexible approach to estimate the long-time sampling error using the geometric ergodicity and the finite-time error analysis. It is expected that such an error analysis framework can be used to estimate the long-time sampling error of a wide class of numerical methods. However, this framework also has some limitations—it requires strict conditions on the geometric ergodicity, and is only compatible with the strong error analysis. In particular, when applied on a numerical method, the long-time sampling error is at most half-order in the time step.

## Acknowledgements

The authors would like to thank Prof. Zhenfu Wang and Lei Li for the helpful discussion on the propagation of chaos and Prof. Jian-Guo Liu for the discussion on the geometric ergodicity.

## Funding

NationalAU: Please write out in full funding agency abbreviation ‘National Natural Science Foundation of China’ in the funding section. Key R&D Program of China (2020YFA0712000 and 2021YFA1001200 to Z.Z.); NSFC (12031013 and 12171013 to Z.Z.).

## Conflict of interest

The authors declare that there is no conflict of interest in this research work.

## REFERENCES

- ANTONELLI, F. & KOHATSU-HIGA, A. (2002) Rate of convergence of a particle method to the solution of the McKean–Vlasov equation. *Ann. Appl. Probab.*, **12**, 423–476.
- BAHLALI, K., MEZERDI, M. A. & MEZERDI, B. (2020) Stability of McKean–Vlasov stochastic differential equations and applications. *Stoch. Dyn.*, **20**, 2050007.
- BAO, J. & HUANG, X. (2022) Approximations of McKean–Vlasov stochastic differential equations with irregular coefficients. *J. Theor. Probab.*, **35**, 1187–1215.
- BENACHOUR, S., ROYNETTE, B., TALAY, D. & VALLOIS, P. (1998) Nonlinear self-stabilizing processes—i existence, invariant probability, propagation of chaos. *Stoch. Process. Appl.*, **75**, 173–201.
- BOBKOV, S. G., GENTIL, I. & LEDOUX, M. (2001) Hypercontractivity of Hamilton–Jacobi equations. *J. Math. Pure. Appl.*, **80**, 669–696.
- BORDENAVE, C., McDONALD, D., & PROUTIERE, A. (2007) A particle system in interaction with a rapidly varying environment: mean field limits and applications. arXiv Preprint *math/0701363*.
- BOSSY, M. & TALAY, D. (1997) A stochastic particle method for the McKean–Vlasov and the Burgers equation. *Math. Comput.*, **66**, 157–193.
- BOU-RABEE, N., EBERLE, A. & ZIMMER, R. (2020) Coupling and convergence for hamiltonian Monte Carlo. *Ann. Appl. Probab.*, **30**, 1209–1250.
- CAÑIZO, J. A. & MISCHLER, S. (2021) Harris-type results on geometric and subgeometric convergence to equilibrium for stochastic semigroups. arXiv Preprint arXiv:2110.09650.
- CARRILLO, J. A., JIN, S. & TANG, Y. (2021) Random batch particle methods for the homogeneous Landau equation. arXiv Preprint arXiv:2110.06430.
- CHAINTRON, L.-P. & DIEZ, A. (2022) Propagation of chaos: a review of models, methods and applications. I. Models and methods.



- DEL MORAL, P. & TUGAUT, J. (2014) Uniform propagation of chaos for a class of inhomogeneous diffusions. *Technical Report*. Citeseer.
- DING, X. & QIAO, H. (2021) Euler–maruyama approximations for stochastic McKean–Vlasov equations with non-Lipschitz coefficients. *J. Theor. Probab.*, **34**, 1408–1425.
- DURMUS, A., EBERLE, A., GUILLIN, A. & ZIMMER, R. (2020) An elementary approach to uniform in time propagation of chaos. *Proc. Am. Math. Soc.*, **148**, 5387–5398.
- DURMUS, A., EBERLE, A., GUILLIN, A. & SCHUH, K. (2022) Sticky nonlinear SDEs and convergence of McKean–Vlasov equations without confinement. arXiv Preprint arXiv:2201.07652.
- DURMUS, A. & MOULINES, E. (2017) Nonasymptotic convergence analysis for the unadjusted Langevin algorithm. *Ann. Appl. Probab.*, **27**, 1551–1587.
- EBERLE, A. (2011) Reflection coupling and Wasserstein contractivity without convexity. *C. R. Math.*, **349**, 1101–1104.
- EBERLE, A. (2016) Reflection couplings and contraction rates for diffusions. *Probab. Theory Relat. Fields*, **166**, 851–886.
- EBERLE, A., GUILLIN, A. & ZIMMER, R. (2019) Quantitative Harris-type theorems for diffusions and McKean–Vlasov processes. *Trans. Am. Math. Soc.*, **371**, 7135–7173.
- FEHSKE, H., SCHNEIDER, R. & WEIßE, A. (2007) *Computational Many-Particle Physics*, vol. **739**. Berlin, Heidelberg: Springer.
- FRENKEL, D. & SMIT, B. (2001) *Understanding Molecular Simulation: From Algorithms to Applications*, vol. **1**, 2nd edn. San Diego: Academic Press.
- GOLSE, F. (2003) The mean-field limit for the dynamics of large particle systems. *Journ. Équ. Dériv. Partielles*, 1–47.
- GOLSE, F., JIN, S. & PAUL, T. (2019) The random batch method for  $n$ -body quantum dynamics. arXiv Preprint arXiv:1912.07424.
- GUILLIN, A., LIU, W., WU, L. & ZHANG, C. (2022) Uniform Poincaré and logarithmic Sobolev inequalities for mean field particle systems. *Ann. Appl. Probab.*, **32**, 1590–1614.
- HA, S.-Y., JIN, S., KIM, D. & KO, D. (2021) Convergence toward equilibrium of the first-order consensus model with random batch interactions. *J. Differ. Equ.*, **302**, 585–616.
- HAIRER, M. & MATTINGLY, J. C. (2011) Yet another look at Harris’ ergodic theorem for Markov chains. *Seminar on Stochastic Analysis, Random Fields and Applications VI*. Basel: Springer Basel, pp. 109–117.
- HAURAY, M. & MISCHLER, S. (2014) On Kac’s chaos and related problems. *J. Funct. Anal.*, **266**, 6055–6157.
- HUANG, M., MALHAMÉ, R. P. & CAINES, P. E. (2006) Large population stochastic dynamic games: closed-loop McKean–Vlasov systems and the Nash certainty equivalence principle. *Commun. Inf. Syst.*, **6**, 221–252.
- JABIN, P.-E. & WANG, Z. (2017) Mean field limit for stochastic particle systems. *Active Particles* (N. Bellomo, P. Degond, E. Tadmor eds). vol. **1**. Modeling and Simulation in Science, Engineering and Technology. Cham: Birkhäuser. [https://doi.org/10.1007/978-3-319-49996-3\\_10](https://doi.org/10.1007/978-3-319-49996-3_10)
- JIN, S. & LI, L. (2022) On the mean field limit of the random batch method for interacting particle systems. *Sci. China Math.*, **65**, 169–202.
- JIN, S., LI, L. & LIU, J.-G. (2020) Random batch methods (rbm) for interacting particle systems. *J. Comput. Phys.*, **400**, 108877.
- JIN, S., LI, L. & LIU, J.-G. (2021a) Convergence of the random batch method for interacting particles with disparate species and weights. *SIAM J. Numer. Anal.*, **59**, 746–768.
- JIN, S., LI, L., XU, Z. & ZHAO, Y. (2021b) A random batch Ewald method for particle systems with Coulomb interactions. *SIAM J. Sci. Comput.*, **43**, B937–B960.
- JIN, S., LI, L., YE, X. & ZHOU, Z. (2022) Ergodicity and long-time behavior of the random batch method for interacting particle systems. arXiv Preprint arXiv:2202.04952.
- KLOKOV, S. & VERETENNIKOV, A. Y. (2006) On mixing and convergence rates for a family of Markov processes approximating SDEs. *Random Oper. Stoch. Equ.*, **14**, 103–126.
- LACKER, D. (2018) Mean field games and interacting particle systems. *Preprint*.

- LEIMKUHLER, B. & MATTHEWS, C. (2015) Molecular dynamics. *Interdisciplinary Applied Mathematics*, vol. **39**. Cham: Springer International Publishing AG.
- LELIEVRE, T. & STOLTZ, G. (2016) Partial differential equations and stochastic methods in molecular dynamics. *Acta Numer.*, **25**, 681–880.
- LI, L., LIU, J.-G. & TANG, Y. (2020a) Some random batch particle methods for the Poisson–Nernst–Planck and Poisson–Boltzmann equations. arXiv Preprint arXiv:2004.05614.
- LI, L., XU, Z. & ZHAO, Y. (2020b) A random-batch Monte Carlo method for many-body systems with singular kernels. *SIAM J. Sci. Comput.*, **42**, A1486–A1509.
- LI, Y., MAO, X., SONG, Q., WU, F. & YIN, G. (2023) Strong convergence of Euler–Maruyama schemes for McKean–Vlasov stochastic differential equations under local Lipschitz conditions of state variables. *IMA J. Numer. Anal.*, **43**, 1001–1035.
- MALRIEU, F. (2003) Convergence to equilibrium for granular media equations and their Euler schemes. *Ann. Appl. Probab.*, **13**, 540–560.
- MATTINGLY, J. C., STUART, A. M. & HIGHAM, D. J. (2002) Ergodicity for SDEs and approximations: locally Lipschitz vector fields and degenerate noise. *Stoch. Process. Appl.*, **101**, 185–232.
- MATTINGLY, J. C., STUART, A. M. & TRETYAKOV, M. V. (2010) Convergence of numerical time-averaging and stationary measures via poisson equations. *SIAM J. Numer. Anal.*, **48**, 552–577.
- McKEAN, H. P. (1967) Propagation of chaos for a class of non-linear parabolic equations. *Stoch. Differ. Equ. (Lecture Series in Differential Equations, Session 7, Catholic Univ., 1967)*, 41–57.
- MEYN, S. P. & TWEEDIE, R. L. (2012) *Markov Chains and Stochastic Stability*. London: Springer.
- DOS REIS, G., ENGELHARDT, S. & SMITH, G. (2022) Simulation of McKean–Vlasov SDEs with super-linear growth. *IMA J. Numer. Anal.*, **42**, 874–922.
- RUDOLF, D. & SCHWEIZER, N. (2018) Bernoulli : official journal of the bernoulli society for mathematical statistics and probability. **24**, 2610–2639.
- SHARDLOW, T. & STUART, A. M. (2000) A perturbation theory for ergodic Markov chains and application to numerical approximations. *SIAM J. Numer. Anal.*, **37**, 1120–1137.
- SZNITMAN, A.-S. (1991) Topics in propagation of chaos. *Ecole d’été de Probabilités de Saint-Flour XIX—1989*. Berlin, Heidelberg: Springer, pp. 165–251.
- WANG, F.-Y. (2018) Distribution dependent SDEs for Landau type equations. *Stoch. Process. Appl.*, **128**, 595–621.
- WEINAN, E., LI, T. & VANDEN-EIJNDEN, E. (2021) *Applied Stochastic Analysis*, vol. **199**. Providence, Rhode Island: American Mathematical Society.
- YE, X. & ZHOU, Z. (2021) Efficient sampling of thermal averages of interacting quantum particle systems with random batches. *J. Chem. Phys.*, **154**, 204106.

## Appendix A. Additional proofs for Sections 3 and 4

*Proof of Lemma 3.2.* WLOG assume the time step  $\tau \leq T$ . Consider the IPS (1.1) first. By Ito’s formula,

$$d|X_t^i|^2 = 2X_t^i \cdot (b^i(X_t^i) dt + \sigma dW_t^i) + d\sigma^2 dt. \quad (\text{A.1})$$

Hence,

$$\begin{aligned} \mathbb{E}|X_t^i|^2 &= \mathbb{E}|X_0^i|^2 + 2 \int_0^t X_s^i \cdot b^i(X_s^i) ds + d\sigma^2 t \\ &= \mathbb{E}|X_0^i|^2 + 2 \int_0^t X_s^i \cdot (b(X_s^i) + \gamma^i(X_s^i)) ds + d\sigma^2 t. \end{aligned}$$

On the one hand,  $\gamma^i$  is uniformly bounded by  $L_1$ , hence

$$2 \int_0^t X_s^i \cdot \gamma^i(X_s) \, ds \leq 2L_1 \int_0^t |X_s^i| \, ds \leq L_1 \int_0^t |X_s^i|^2 \, ds + L_1 t. \quad (\text{A.2})$$

On the other hand, using the linear growth condition on  $b$ , one has

$$\begin{aligned} 2 \int_0^t X_s^i \cdot b(X_s^i) \, ds &\leq 2L_0 \int_0^t (|X_s^i|^2 + |X_s^i|) \, ds \\ &\leq L_0 \int_0^t (3|X_s^i|^2 + 1) \, ds \\ &\leq 3L_0 \int_0^t |X_s^i|^2 \, ds + L_0 t. \end{aligned} \quad (\text{A.3})$$

Using these inequalities, one obtains

$$2 \int_0^t X_s^i \cdot (b(X_s^i) + \gamma^i(X_s)) \, ds \leq (3L_0 + L_1) \int_0^t |X_s^i|^2 \, ds + (L_0 + L_1)t. \quad (\text{A.4})$$

Let  $L := 3L_0 + L_1 + d\sigma^2$ , then for any  $t \in [0, T]$ ,

$$\begin{aligned} \mathbb{E}|X_t^i|^2 &\leq \mathbb{E}|X_0^i|^2 + L \int_0^t |X_s^i|^2 \, ds + Lt \\ &\leq M + LT + L \int_0^t |X_s^i|^2 \, ds. \end{aligned}$$

Using the Gronwall's inequality (see the proof of Lemma 7.21 of [Weinan \*et al.\*, 2021](#)),

$$\mathbb{E}|X_t^i|^2 \leq (M + LT) \exp(LT), \quad t \in [0, T], \quad (\text{A.5})$$

which yields the first inequality of (3.17). For the second inequality of (3.17), use the SDE

$$X_t^i - X_{t_n}^i = \int_{t_n}^t b^i(X_s) \, ds + \sigma(W_t^i - W_{t_n}^i). \quad (\text{A.6})$$

Hence,

$$\mathbb{E}|X_t^i - X_{t_n}^i|^2 \leq 2\mathbb{E} \left| \int_{t_n}^t b^i(X_s) \, ds \right|^2 + 2d\sigma^2\tau \leq 2\tau \int_{t_n}^t \mathbb{E}|b^i(X_s)|^2 \, ds + 2d\sigma^2\tau. \quad (\text{A.7})$$

Using the linear growth condition

$$|b^i(x)| \leq |b(x^i)| + |\gamma^i(x)| \leq L_0(|x^i| + 1) + L_1 \leq L(|x^i| + 1), \quad (\text{A.8})$$

then one has  $|b^i(x)|^2 \leq 2L^2(|x^i|^2 + 1)$ , and thus

$$\begin{aligned} \mathbb{E}|X_t^i - X_{t_n}^i|^2 &\leq 4L^2\tau \int_{t_n}^t (\mathbb{E}|X_s^i|^2 + 1) \, ds + 2d\sigma^2\tau \\ &\leq 4L^2TC\tau + 2d\sigma^2\tau = C\tau, \end{aligned}$$

which is exactly the desired result. The proof above also holds true for the RB-IPS (3.1) because we only need to use  $|\gamma^i(x)| \leq L_1$  in each time step  $[t_n, t_{n+1})$ .  $\square$

*Proof of Theorem 3.3.* WLOG assume the time step  $\tau \leq T$ . Define the trajectory difference  $e_n^i = X_{t_n}^i - \tilde{X}_n^i$ . The IPS (1.1) and the discrete IPS (1.3) are given by

$$X_{t_{n+1}}^i = X_{t_n}^i + \int_{t_n}^{t_{n+1}} b^i(X_t) dt + \sigma W_\tau^i, \quad \tilde{X}_{n+1}^i = \tilde{X}_n^i + \int_{t_n}^{t_{n+1}} b^i(\tilde{X}_n) dt + \sigma W_\tau^i, \quad (\text{A.9})$$

where  $W_\tau^i := W_{t_{n+1}}^i - W_{t_n}^i \sim \mathcal{N}(0, \tau)$ . Then  $e_n^i$  satisfies the recurrence relation

$$e_{n+1}^i = e_n^i + \int_{t_n}^{t_{n+1}} (b^i(X_t) - b^i(\tilde{X}_n)) dt. \quad (\text{A.10})$$

Recall  $b^i(x) = b(x^i) + \gamma^i(x)$ . Squaring both sides of (A.10), we obtain

$$\begin{aligned} |e_{n+1}^i|^2 &\leq (1 + \tau)|e_n^i|^2 + \left(1 + \frac{1}{\tau}\right) \left(\int_{t_n}^{t_{n+1}} (b^i(X_t) - b^i(\tilde{X}_n)) dt\right)^2 \\ &\leq (1 + \tau)|e_n^i|^2 + (1 + \tau) \int_{t_n}^{t_{n+1}} |b^i(X_t) - b^i(\tilde{X}_n)|^2 dt \\ &\leq (1 + \tau)|e_n^i|^2 + 2(1 + \tau) \int_{t_n}^{t_{n+1}} |b(X_t^i) - b(\tilde{X}_n^i)|^2 dt + 2(1 + \tau) \int_{t_n}^{t_{n+1}} |\gamma^i(X_t) - \gamma^i(\tilde{X}_n)|^2 dt. \end{aligned}$$

On the one hand, the global Lipschitz condition of  $b$  implies

$$|b(X_t^i) - b(\tilde{X}_n^i)| \leq L_0 |X_t^i - \tilde{X}_n^i| \implies \int_{t_n}^{t_{n+1}} |b(X_t^i) - b(\tilde{X}_n^i)|^2 dt \leq L_0^2 \int_{t_n}^{t_{n+1}} |X_t^i - \tilde{X}_n^i|^2 dt. \quad (\text{A.11})$$

On the other hand, the boundedness of  $\gamma^i$  implies

$$\begin{aligned} |\gamma^i(X_t) - \gamma^i(\tilde{X}_n)| &\leq L_1 |X_t^i - \tilde{X}_n^i| + \frac{L_1}{N-1} \sum_{j \neq i} |X_t^j - \tilde{X}_n^j| \\ \implies |\gamma^i(X_t) - \gamma^i(\tilde{X}_n)|^2 &\leq 2L_1^2 |X_t^i - \tilde{X}_n^i|^2 + 2L_1^2 \left(\frac{1}{N-1} \sum_{j \neq i} |X_t^j - \tilde{X}_n^j|\right)^2 \\ &\leq 2L_1^2 |X_t^i - \tilde{X}_n^i|^2 + \frac{2L_1^2}{N-1} \sum_{j \neq i} |X_t^j - \tilde{X}_n^j|^2 \\ \implies \int_{t_n}^{t_{n+1}} |\gamma^i(X_t) - \gamma^i(\tilde{X}_n)|^2 dt &\leq 2L_1^2 \int_{t_n}^{t_{n+1}} |X_t^i - \tilde{X}_n^i|^2 dt + \frac{2L_1^2}{N-1} \sum_{j \neq i} \int_{t_n}^{t_{n+1}} |X_t^j - \tilde{X}_n^j|^2 dt. \quad (\text{A.12}) \end{aligned}$$

Combining (A.11) and (A.12),  $e_{n+1}^i$  can be bounded by

$$\begin{aligned} |e_{n+1}^i|^2 &\leq (1 + \tau)|e_n^i|^2 + (1 + \tau)(2L_0^2 + 4L_1^2) \int_{t_n}^{t_{n+1}} |X_t^i - \tilde{X}_n^i|^2 dt \\ &\quad + (1 + \tau) \frac{4L_1^2}{N-1} \sum_{j \neq i} \int_{t_n}^{t_{n+1}} |X_t^j - \tilde{X}_n^j|^2 dt. \end{aligned}$$

Summation over  $i$  gives

$$\sum_{i=1}^N |e_{n+1}^i|^2 \leq (1 + \tau) \sum_{i=1}^N |e_n^i|^2 + (1 + \tau)(2L_0^2 + 8L_1^2) \sum_{i=1}^N \int_{t_n}^{t_{n+1}} |X_t^i - \tilde{X}_n^i|^2 dt. \quad (\text{A.13})$$

Note that by the Cauchy's inequality

$$|X_t^i - \tilde{X}_n^i|^2 \leq 2|X_t^i - X_{t_n}^i|^2 + 2|X_{t_n}^i - \tilde{X}_n^i|^2, \quad (\text{A.14})$$

from Lemma 3.2, we have

$$\mathbb{E}|X_t^i - \tilde{X}_n^i|^2 \leq C\tau + 2\mathbb{E}|e_n^i|^2. \quad (\text{A.15})$$

Integrating (A.15) in the time interval  $[t_n, t_{n+1})$  gives

$$\int_{t_n}^{t_{n+1}} \mathbb{E}|X_t^i - \tilde{X}_n^i|^2 \leq C\tau^2 + 2\tau\mathbb{E}|e_n^i|^2. \quad (\text{A.16})$$

Taking the expectation in (A.13) gives

$$\begin{aligned} \sum_{i=1}^N \mathbb{E}|e_{n+1}^i|^2 &\leq (1 + \tau) \sum_{i=1}^N \mathbb{E}|e_n^i|^2 + C(1 + \tau) \left( N\tau^2 + \tau \sum_{i=1}^N \mathbb{E}|e_n^i|^2 \right) \\ &\leq (1 + C\tau) \sum_{i=1}^N \mathbb{E}|e_n^i|^2 + CN\tau^2. \end{aligned} \quad (\text{A.17})$$

Note that  $e_0^i \equiv 0$ , the discrete Gronwall's inequality thus gives

$$\frac{1}{N} \sum_{i=1}^N \mathbb{E}|e_n^i|^2 \leq \tau \left( (1 + C\tau)^n - 1 \right) \leq e^{CT} \tau = C\tau, \quad (\text{A.18})$$

which implies the strong error is bounded by  $C\tau$  for  $0 \leq n \leq T/\tau$ .

Now we turn to the random batch case. Let  $e_n^i = Y_{t_n}^i - \tilde{Y}_n^i$ , then  $e_n^i$  satisfies

$$|e_{n+1}^i|^2 \leq (1 + \tau)|e_n^i|^2 + 2(1 + \tau) \int_{t_n}^{t_{n+1}} |b(Y_t^i) - b(\tilde{Y}_n^i)|^2 dt + 2(1 + \tau) \int_{t_n}^{t_{n+1}} |\gamma^i(Y_t) - \gamma^i(\tilde{Y}_n)|^2 dt. \quad (\text{A.19})$$

Again we stress that the perturbation force  $\gamma^i(x)$  depends on the batch division  $\mathcal{D} = \{\mathcal{C}_1, \dots, \mathcal{C}_q\}$ . Regardless of the batch division in the time interval  $[t_n, t_{n+1})$ , we have the inequalities

$$\int_{t_n}^{t_{n+1}} |b(Y_t^i) - b(\tilde{Y}_n^i)|^2 \leq L_0^2 \int_{t_n}^{t_{n+1}} |Y_t^i - \tilde{Y}_n^i|^2 dt \quad (\text{A.20})$$

and

$$|\gamma^i(Y_t) - \gamma^i(\tilde{Y}_n)| \leq 2L_1^2 \int_{t_n}^{t_{n+1}} |Y_t^i - \tilde{Y}_n^i|^2 dt + \frac{2L_1^2}{p-1} \sum_{j \neq i, j \in \mathcal{C}} \int_{t_n}^{t_{n+1}} |Y_t^j - \tilde{Y}_n^j|^2 dt. \quad (\text{A.21})$$

Combining (A.20) and (A.21),  $e_{n+1}^i$  has the estimate

$$\begin{aligned} |e_{n+1}^i|^2 &\leq (1+\tau)|e_n^i|^2 + (1+\tau)(2L_0^2 + 4L_1^2) \int_{t_n}^{t_{n+1}} |Y_t^i - \tilde{Y}_n^i|^2 dt \\ &\quad + (1+\tau) \frac{4L_1^2}{p-1} \sum_{j \neq i, j \in \mathcal{C}} \int_{t_n}^{t_{n+1}} |Y_t^j - \tilde{Y}_n^j|^2 dt. \end{aligned}$$

Summation over  $i \in \mathcal{C}$  and  $\mathcal{C} \in \mathcal{D}$  recovers

$$\sum_{i=1}^N |e_{n+1}^i|^2 \leq (1+\tau) \sum_{i=1}^N |e_n^i|^2 + (1+\tau)(2L_0^2 + 8L_1^2) \sum_{i=1}^N \int_{t_n}^{t_{n+1}} |Y_t^i - \tilde{Y}_n^i|^2 dt. \quad (\text{A.22})$$

Using the same strategy in the case of the IPS (1.1), we have

$$\mathbb{E}|Y_t^i - \tilde{Y}_n^i|^2 \leq C\tau + 2\mathbb{E}|e_n^i|^2. \quad (\text{A.23})$$

Taking the expectation in (A.22) then gives

$$\begin{aligned} \sum_{i=1}^N \mathbb{E}|e_{n+1}^i|^2 &\leq (1+\tau) \sum_{i=1}^N \mathbb{E}|e_n^i|^2 + C(1+\tau) \left( N\tau^2 + \tau \sum_{i=1}^N \mathbb{E}|e_n^i|^2 \right) \\ &\leq (1+C\tau) \sum_{i=1}^N \mathbb{E}|e_n^i|^2 + CN\tau^2, \end{aligned} \quad (\text{A.24})$$

which is exactly the same with (A.17). The rest part of the proof is completely the same.  $\square$

*Proof of Lemma 3.10.* 1. First we estimate  $|f(x, \tau)|^2$ . Using Assumptions 3.1 and 3.9, we have

$$\begin{aligned} |f(x, \tau)|^2 &= |x + b(x)\tau|^2 \\ &= |x|^2 + 2x \cdot b(x)\tau + |b(x)|^2\tau^2 \\ &\leq |x|^2 + 2(\theta - \alpha|x|^2)\tau + 2L_0^2(|x|^2 + 1)\tau^2 \end{aligned} \quad (\text{A.25})$$

$$= (1 + 2L_0^2\tau^2 - 2\alpha\tau)|x|^2 + 2\theta\tau + 2L_0^2\tau^2. \quad (\text{A.26})$$

Since  $\tau < \alpha/(2L_0^2)$ , (A.26) implies

$$|f(x, \tau)|^2 \leq (1 - \alpha\tau)|x|^2 + (\alpha + 2\theta)\tau. \quad (\text{A.27})$$

To estimate  $|f(x, \tau)|^4$ , square both sides of (A.27) and utilize  $\tau < 1/(2\alpha)$ , then

$$\begin{aligned} |f(x, \tau)|^4 &\leq (1 - \alpha\tau)^2|x|^4 + 2(\alpha + 2\theta)|x|^2\tau + (\alpha + 2\theta)^2\tau^2 \\ &\leq (1 - \alpha\tau)^2|x|^4 + \left(k\tau|x|^4 + \frac{(\alpha + 2\theta)^2}{k}\tau\right) + (\alpha + 2\theta)^2\tau^2 \\ &\leq (1 - 2\alpha\tau + \alpha^2\tau^2 + k\tau)|x|^4 + \frac{(\alpha + 2\theta)^2}{k}\tau + (\alpha + 2\theta)^2\tau^2 \\ &= \left(1 - \frac{3\alpha\tau}{2} + k\tau\right)|x|^4 + O(\tau), \end{aligned} \quad (\text{A.28})$$

where  $k > 0$  is an  $O(1)$  parameter to be determined. Choose  $k = \alpha/2$  in (A.28), then

$$|f(x, \tau)|^4 \leq (1 - \alpha\tau)|x|^4 + O(\tau), \quad (\text{A.29})$$

hence (3.26) holds. 2. By direct calculation,

$$\begin{aligned} |f(x, \tau) + \gamma\tau|^2 &\leq |f(x, \tau)|^2 + 2|f(x, \tau)|L_1\tau + L_1^2\tau^2 \\ &\leq |f(x, \tau)|^2 + \left(|f(x, \tau)|^2 k\tau + \frac{L_1^2}{k}\tau\right) + L_1^2\tau^2 \\ &= (1 + k\tau)|f(x, \tau)|^2 + O(\tau), \end{aligned}$$

where  $k > 0$  is an  $O(1)$  parameter to be determined. By (3.26), we choose  $k = \alpha/2$  and

$$\begin{aligned} |f(x, \tau) + \gamma\tau|^2 &\leq \left(1 + \frac{\alpha\tau}{2}\right)|f(x, \tau)|^2 + O(\tau) \\ &\leq \left(1 + \frac{\alpha\tau}{2}\right)\left((1 - \alpha\tau)|x|^2 + O(\tau)\right) + O(\tau) \\ &\leq \left(1 - \frac{\alpha\tau}{2}\right)|x|^2 + O(\tau). \end{aligned} \quad (\text{A.30})$$

Squaring both sides of (A.30), one obtains

$$\begin{aligned} |f(x, \tau) + \gamma\tau|^4 &\leq \left(\left(1 - \frac{\alpha\tau}{2}\right)|x|^2 + C\tau\right)^2 \\ &= \left(1 - \alpha\tau + \frac{\alpha^2\tau^2}{4}\right)|x|^4 + 2C|x|^2\tau + O(\tau^2) \\ &\leq \left(1 - \frac{3\alpha\tau}{4}\right)|x|^4 + \left(k\tau|x|^4 + \frac{C^2\tau}{k}\right) + O(\tau^2) \\ &= \left(1 - \frac{3\alpha\tau}{4} + k\tau\right)|x|^4 + O(\tau), \end{aligned}$$

where  $k > 0$  is a  $O(1)$  parameter. By choosing  $k = \alpha/4$ , (3.27) becomes true.  $\square$

*Proof of Theorem 3.11.* The update scheme of the discrete IPS trajectory  $\tilde{X}_n^i$  is given by

$$\tilde{X}_{n+1}^i = \tilde{X}_n^i + b(\tilde{X}_n^i)\tau + \gamma^i(\tilde{X}_n^i)\tau + \sigma W_\tau^i, \quad (\text{A.31})$$

where  $W_\tau^i \sim \mathcal{N}(0, \tau)$ , and  $\gamma^i$  is defined in (3.8). With  $f(x, \tau) = x + b(x)\tau$ , we can write (A.31) as

$$\tilde{X}_{n+1}^i = f(\tilde{X}_n^i, \tau) + \gamma^i(\tilde{X}_n^i)\tau + \sigma W_\tau^i. \quad (\text{A.32})$$



Note that the random variable  $W_\tau^i$  is independent of  $\tilde{X}_n$ , we have

$$\begin{aligned}\mathbb{E}|\tilde{X}_{n+1}^i|^4 &= \mathbb{E}|f(\tilde{X}_n, \tau) + \gamma^i(\tilde{X}_n)\tau|^4 + 6\mathbb{E}|f(\tilde{X}_n, \tau) + \gamma^i(\tilde{X}_n)\tau|^2 \mathbb{E}|\sigma W_\tau^i|^2 + \mathbb{E}|\sigma W_\tau^i|^4 \\ &= \mathbb{E}|f(\tilde{X}_n, \tau) + \gamma^i(\tilde{X}_n)\tau|^4 + 6\mathbb{E}|f(\tilde{X}_n, \tau) + \gamma^i(\tilde{X}_n)\tau|^2 d\sigma^2\tau + 3d^2\sigma^4\tau^2 \\ &\leq \mathbb{E}|f(\tilde{X}_n, \tau) + \gamma^i(\tilde{X}_n)\tau|^4 + \left(k\tau \mathbb{E}|f(\tilde{X}_n, \tau) + \gamma^i(\tilde{X}_n)\tau|^4 + \frac{9d^2\sigma^4\tau^2}{k}\right) + 3d^2\sigma^4\tau^2 \\ &= (1 + k\tau)\mathbb{E}|f(\tilde{X}_n, \tau) + \gamma^i(\tilde{X}_n)\tau|^4 + O(\tau^2),\end{aligned}\quad (\text{A.33})$$

where  $k > 0$  is an  $O(1)$  parameter to be determined. Since  $\gamma^i(\tilde{X}_n)$  is uniformly bounded by  $L_1$ , by Lemma 3.10 we have

$$\mathbb{E}|f(\tilde{X}_n, \tau) + \gamma^i(\tilde{X}_n)\tau|^4 \leq \left(1 - \frac{\alpha\tau}{2}\right)\mathbb{E}|\tilde{X}_n^i|^4 + C\tau. \quad (\text{A.34})$$

Hence, (A.33) implies

$$\begin{aligned}\mathbb{E}|\tilde{X}_{n+1}^i|^4 &\leq (1 + k\tau)\left(\left(1 - \frac{\alpha\tau}{2}\right)\mathbb{E}|\tilde{X}_n^i|^4 + C\tau\right) + O(\tau^2) \\ &\leq \left(1 - \left(\frac{\alpha}{2} - k\right)\tau\right)\mathbb{E}|\tilde{X}_n^i|^4 + O(\tau).\end{aligned}\quad (\text{A.35})$$

Now we can choose  $k = \alpha/4$  in (A.35) to obtain

$$\mathbb{E}|\tilde{X}_{n+1}^i|^4 \leq \left(1 - \frac{\alpha\tau}{4}\right)\mathbb{E}|\tilde{X}_n^i|^4 + C\tau, \quad (\text{A.36})$$

and thus by the discrete Gronwall's inequality,

$$\sup_{n \geq 0} \mathbb{E}|\tilde{X}_n^i|^4 \leq \max\left\{M_4, \frac{4C}{\alpha}\right\}. \quad (\text{A.37})$$

For the discrete RB-IPS (1.4), the proof is completely the same because we still have  $|\gamma^i(x)| \leq L_1$ , and thus the recurrence relation

$$\mathbb{E}|\tilde{Y}_{n+1}^i|^4 \leq \left(1 - \frac{\alpha\tau}{4}\right)\mathbb{E}|\tilde{Y}_n^i|^4 + C\tau \quad (\text{A.38})$$

becomes true.  $\square$

*Proof of Lemma 3.18.* By induction on the integer  $s \geq 1$ , it is easy to verify if  $n \geq sm$ , then

$$a_n \leq \varepsilon \frac{1 - q^s}{1 - q} + q^s a_{n-sm}. \quad (\text{A.39})$$

For any integer  $n \geq 0$ , let  $n = sm + r$  for some integer  $s \geq 0$  and  $r \in \{0, 1, \dots, m-1\}$ . Then

$$a_n \leq \frac{\varepsilon}{1 - q} + Mq^s \leq \frac{\varepsilon}{1 - q} + Mq^{\frac{n}{m}-1}, \quad (\text{A.40})$$

yielding (3.40).  $\square$

*Proof of Theorem 4.6.* Given the probability distributions  $\mu, \nu \in \mathcal{P}(\mathbb{R}^d)$ , by Theorem 3.15 we have

$$\mathcal{W}_1(\mu^{\otimes N} p_t, \nu^{\otimes N} p_t) \leq Ce^{-\beta t} \mathcal{W}_1(\mu^{\otimes N}, \nu^{\otimes N}) = Ce^{-\beta t} \mathcal{W}_1(\mu, \nu). \quad (\text{A.41})$$

Here,  $(p_t)_{t \geq 0}$  is the semigroup of the IPS (1.1) in  $\mathbb{R}^{Nd}$ . Recall that  $\mathcal{W}_1$  on the LHS is the normalized Wasserstein-1 distance in  $\mathbb{R}^{Nd}$  defined in (3.23), and  $\mathcal{W}_1$  on the RHS is the classical Wasserstein-1 distance in  $\mathbb{R}^d$ . Using the triangle inequality, we have

$$\begin{aligned} \mathcal{W}_1(\mu \bar{p}_t, \nu \bar{p}_t) &= \mathcal{W}_1(\mu^{\otimes N} \bar{p}_t^{\otimes N}, \nu^{\otimes N} \bar{p}_t^{\otimes N}) \\ &\leq \mathcal{W}_1(\mu^{\otimes N} \bar{p}_t^{\otimes N}, \mu^{\otimes N} p_t) + \mathcal{W}_1(\nu^{\otimes N} \bar{p}_t^{\otimes N}, \nu^{\otimes N} p_t) + \mathcal{W}_1(\mu^{\otimes N} p_t, \nu^{\otimes N} p_t) \\ &\leq \mathcal{W}_1(\mu^{\otimes N} \bar{p}_t^{\otimes N}, \mu^{\otimes N} p_t) + \mathcal{W}_1(\nu^{\otimes N} \bar{p}_t^{\otimes N}, \nu^{\otimes N} p_t) + Ce^{-\beta t} \mathcal{W}_1(\mu, \nu). \end{aligned}$$

By Theorem 4.1, for given  $t > 0$ , there exists a constant  $C_0 = C_0(\kappa, L_0, \sigma, t)$  such that

$$\mathcal{W}_1(\mu^{\otimes N} \bar{p}_t^{\otimes N}, \mu^{\otimes N} p_t) \leq \frac{C_0}{N}. \quad (\text{A.42})$$

Hence, we obtain

$$\mathcal{W}_1(\mu \bar{p}_t, \nu \bar{p}_t) \leq \frac{2C_0}{\sqrt{N}} + Ce^{-\beta t} \mathcal{W}_1(\mu, \nu). \quad (\text{A.43})$$

Fix  $t > 0$  and let  $N \rightarrow \infty$ , we obtain

$$\mathcal{W}_1(\mu \bar{p}_t, \nu \bar{p}_t) \leq Ce^{-\beta t} \mathcal{W}_1(\mu, \nu), \quad (\text{A.44})$$

which is the desired result.  $\square$

*Proof of Corollary 4.7.* First we prove the existence of the invariant distribution  $\bar{\pi} \in \mathcal{P}_1(\mathbb{R}^d)$  of the MVP (1.2). Since  $(\bar{p}_t)_{t \geq 0}$  is a nonlinear semigroup, we cannot use the same technique as in the linear case. Our proof below is partially inspired from Theorem 5.1 of Cañizo & Mischler (2021). Choose the constant  $T$ , which satisfies  $Ce^{-\beta T} = 1/2$ , then we have

$$\mathcal{W}_1(\mu \bar{p}_T, \nu \bar{p}_T) \leq \frac{1}{2} \mathcal{W}_1(\mu, \nu) \quad (\text{A.45})$$

for any probability distributions  $\mu, \nu \in \mathcal{P}_1(\mathbb{R}^d)$ . Hence, the mapping  $\mu \mapsto \mu \bar{p}_T$  is contractive in the complete metric space  $(\mathcal{P}_1(\mathbb{R}^d), \mathcal{W}_1(\cdot, \cdot))$ . Using the Banach fixed point theorem, there exists a unique fixed point  $\bar{\pi} \in \mathcal{P}_1(\mathbb{R}^d)$  such that

$$\bar{\pi} \bar{p}_T = \bar{\pi}. \quad (\text{A.46})$$

Since  $(\bar{p}_t)_{t \geq 0}$  forms a semigroup, for any  $t \geq 0$ , we have

$$(\bar{\pi} \bar{p}_t) \bar{p}_T = \bar{\pi} \bar{p}_t, \quad (\text{A.47})$$

which implies  $\bar{\pi} \bar{p}_t \in \mathcal{P}_1(\mathbb{R}^d)$  is the invariant distribution of the operator  $\bar{p}_T$ . Due to the uniqueness of the invariant distribution  $\bar{\pi}$  for the operator  $\bar{p}_T$ , we obtain

$$\bar{\pi} \bar{p}_t = \bar{\pi}, \quad \forall t \geq 0, \quad (\text{A.48})$$

hence  $\bar{\pi} \in \mathcal{P}_1(\mathbb{R}^d)$  is the invariant distribution of the semigroup  $(\bar{p}_t)_{t \geq 0}$ .

Next we estimate the difference between the invariant distributions  $\pi, \bar{\pi} \in \mathcal{P}_1(\mathbb{R}^d)$ . We still choose the constant  $T$  according to  $Ce^{-\beta T} = 1/2$ . Using the triangle inequality, there exists a constant

$C = C(\kappa, L_0, \sigma)$  such that

$$\begin{aligned}\mathcal{W}_1(\bar{\pi}^{\otimes N}, \pi) &= \mathcal{W}_1(\bar{\pi}^{\otimes N} \bar{p}_T^{\otimes N}, \pi p_T) \\ &\leq \mathcal{W}_1(\bar{\pi}^{\otimes N} \bar{p}_T^{\otimes N}, \bar{\pi}^{\otimes N} p_T) + \mathcal{W}_1(\bar{\pi}^{\otimes N} p_T, \pi p_T) \\ &\leq \frac{C}{\sqrt{N}} + C e^{-\beta T} \mathcal{W}_1(\bar{\pi}^{\otimes N}, \pi) \\ &= \frac{C}{\sqrt{N}} + \frac{1}{2} \mathcal{W}_1(\bar{\pi}^{\otimes N}, \pi).\end{aligned}$$

Then  $\mathcal{W}_1(\bar{\pi}^{\otimes N}, \pi) \leq C/\sqrt{N}$ .

□