

ZERO-INERTIA LIMIT: FROM PARTICLE SWARM OPTIMIZATION TO CONSENSUS-BASED OPTIMIZATION

CRISTINA CIPRIANI, HUI HUANG, AND JINNIAO QIU

ABSTRACT. Recently a continuous description of the particle swarm optimization (PSO) based on a system of stochastic differential equations was proposed by Grassi and Pareschi in [27] where the authors formally showed the link between PSO and the consensus-based optimization (CBO) through zero-inertia limit. This paper is devoted to solving this theoretical open problem proposed in [27] by providing a rigorous derivation of CBO from PSO through the limit of zero inertia, and a quantified convergence rate is obtained as well. The proofs are based on a probabilistic approach by investigating both the weak and strong convergence of the corresponding stochastic differential equations (SDEs) of McKean type in the continuous path space and the results are illustrated with some numerical examples.

Keywords: Swarm optimization, consensus-based optimization, Laplace's principle, tightness.

1. INTRODUCTION

Over the last decades, large systems of interacting particles are widely used in the investigation of complex systems that model collective behavior (or swarming), an area that has attracted a great deal of attention; see for instance [5, 14, 29, 46] and references therein. Such complex systems frequently appear in modeling phenomena such as biological swarms [17], crowd dynamics [6], self-assembly of nanoparticles [30], and opinion formation [46]. In the field of global optimization, similar particle models are also used in *metaheuristics* [1, 4, 8, 26], which provide empirically robust solutions to tackle hard optimization problems with fast algorithms. Metaheuristics are methods that orchestrate an interaction between local improvement procedures and global/high level strategies, and combine random and deterministic decisions, to create a process capable of escaping from local optima and performing a robust search of a solution space. In the sequel, we consider the following optimization problem

$$x^* \in \operatorname{argmin}_{x \in \mathbb{R}^d} \mathcal{E}(x), \quad (1.1)$$

where $\mathcal{E}(x) : \mathbb{R}^d \rightarrow \mathbb{R}$ is a given continuous cost function.

One noble example of metaheuristics is the so-called Particle Swarm Optimization (PSO), which was initially introduced to model the intelligent collective behavior of complex biological systems such as flocks of birds or schools of fish [38, 39, 51], and it is now widely recognized as an efficient method for tackling complex optimization problems [43, 49]. Certain convergence and stability analysis of PSO may be found, for instance, in [10, 48, 50, 53] and readers would be referred to [54] for a review on the PSO method and its variants and applications. The PSO method solves optimization problem (1.1) by considering a group of candidate solutions, which are represented by particles. Then the algorithm moves those particles in the search space according to certain mathematical relationships on the particle position and velocity. Each particle is driven to its best known local location, which is updated once the particles find better positions.

Date: April 7, 2022.

The authors would like to thank Prof. Lorenzo Pareschi for suggesting this problem to them. C. C. acknowledges the partial support of the project "Online Firestorms And Resentment Propagation On Social Media: Dynamics, Predictability and Mitigation" of the TUM Institute for Ethics in Artificial Intelligence. H. H. is partially supported by the Pacific Institute for the Mathematical Sciences (PIMS) postdoctoral fellowship. J. Q. is partially supported by the National Science and Engineering Research Council of Canada (NSERC) and by the start-up funds from the University of Calgary.

However, the mathematical understanding of PSO is still in its infancy. Recently Grassi and Pareschi [27] took a significant first step towards a mathematical theory for PSO based on a continuous description in the form of a system of stochastic differential equations:

$$\begin{cases} dX_t^{i,m} = V_t^{i,m} dt, \\ dV_t^{i,m} = -\frac{\gamma}{m} V_t^{i,m} dt + \frac{\lambda}{m} (X_t^\alpha(\rho^{N,m}) - X_t^{i,m}) dt + \frac{\sigma}{m} D(X_t^\alpha(\rho^{N,m}) - X_t^{i,m}) dB_t^i, \end{cases} \quad i = 1, \dots, N, \quad (1.2)$$

where the \mathbb{R}^d -valued functions $X_t^{i,m}$ and $V_t^{i,m}$ denote the position and velocity of the i -th particle at time t , $m > 0$ is the inertia weight, $\gamma = 1 - m \geq 0$ is the friction coefficient, $\lambda > 0$ is the acceleration coefficient, $\sigma > 0$ is the diffusion coefficient, and $\{(B_t^i)_{t \geq 0}\}_{i=1}^N$ are N independent d -dimensional Brownian motions. We also use the notations for the diagonal matrix

$$D(X_t) := \text{diag}\{(X_t)_1, \dots, (X_t)_d\} \in \mathbb{R}^{d \times d},$$

where $(X_t)_k$ is the k -th component of X_t , and the weighted average is given by

$$X_t^\alpha(\rho^{N,m}) := \frac{\int_{\mathbb{R}^d} x \omega_\alpha^\mathcal{E}(x) \rho^{N,m}(t, dx)}{\int_{\mathbb{R}^d} \omega_\alpha^\mathcal{E}(x) \rho^{N,m}(t, dx)} \quad (1.3)$$

with the empirical measure $\rho^{N,m}(t, dx) := \frac{1}{N} \sum_{i=1}^N \delta_{X_t^{i,m}}(dx)$. So we can rewrite

$$D(X_t^\alpha(\rho^{N,m}) - X_t^{i,m}) dB_t^i = \sum_{k=1}^d (X_t^\alpha(\rho^{N,m}) - X_t^{i,m})_k d(B_t^i)^k e_k, \quad (1.4)$$

where e_k is the unit vector in the k -th dimension for $k = 1, \dots, d$. Noise of the form (1.4) is called the anisotropic component-wise noise, which was used in [15, 23] to remove the dimensionality dependence for the CBO method. Furthermore, the initial data $(X_0^i, V_0^i)_{i=1}^N$ are independent and identically distributed (i.i.d.) with the common distribution $f_0 \in \mathcal{P}_4(\mathbb{R}^{2d})$, where $\mathcal{P}_4(\mathbb{R}^{2d})$ denotes the space of probability measures with finite fourth moment, endowed with the Wasserstein distance [3]. The choice of the weight function

$$\omega_\alpha^\mathcal{E}(x) := \exp(-\alpha \mathcal{E}(x))$$

comes from the well-known Laplace's principle [19, 45], a classical asymptotic method for integrals, which states that for any probability measure $\rho \in \mathcal{P}(\mathbb{R}^d)$, there holds

$$\lim_{\alpha \rightarrow \infty} \left(-\frac{1}{\alpha} \log \left(\int_{\mathbb{R}^d} \omega_\alpha^\mathcal{E}(x) \rho(dx) \right) \right) = \inf_{x \in \text{supp}(\rho)} \mathcal{E}(x). \quad (1.5)$$

Thus for α large enough, one expects that

$$X_t^\alpha(\rho^{N,m}) \approx \text{argmin} \{ \mathcal{E}(X_t^{1,m}), \dots, \mathcal{E}(X_t^{N,m}) \},$$

which means that $X_t^\alpha(\rho^{N,m})$ is a global best location at time t .

Before starting our analysis of the PSO dynamics (1.2), let us illustrate numerically the behavior of the dynamics for the benchmark Ackley function

$$\mathcal{E}(x) = -20 \exp \left(-\frac{0.2}{\sqrt{d}} |x - x^*| \right) - \exp \left(\frac{1}{d} \sum_{k=1}^d \cos(2\pi(x_k - x_k^*)) \right) + e + 20$$

in the case of $d = 2$, and with the global minimizer $x^* = (0, 0)^T$. In Figures 1 and 2, we initialize the particles with a normal distribution around x^* and then apply a discretization scheme (which will be explained in Section 4) to the system (1.2). We can see that all the particles successfully find the global minimizer x^* , and particles' velocity converges to zero.

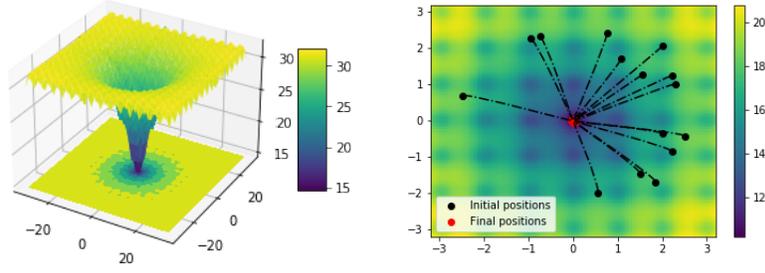


FIGURE 1. Left: the Ackley function for $d = 2$ with the unique global minimum at the point $x^* = (0, 0)^T$. Right: Particles trajectories of the PSO model (1.2) during the simulation for the 2- d Ackley function with the global minimizer x^* . The simulation parameters are: time discretization 0.01, number of particles 10^3 , $\lambda = 1$, $\sigma = \frac{1}{\sqrt{3}}$, $\alpha = 30$, $m = 0.1$. The initial data are sampled from a normal bi-dimensional distribution.

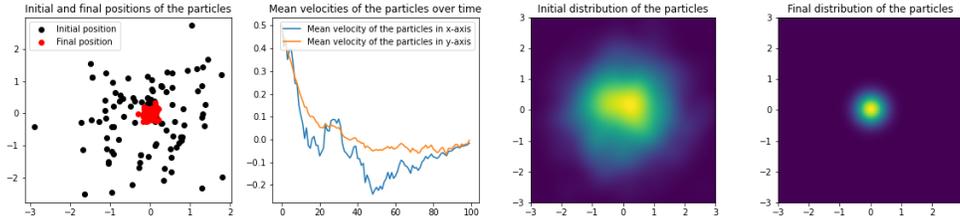


FIGURE 2. Application of the PSO dynamics (1.2) to the 2- d Ackley function $\mathcal{E}(x)$ with the global minimizer $x^* = (0, 0)^T$. Particles initially have a normal distribution around x^* . Then all particles converge to one point, the global minimizer x^* , and they stop moving eventually, i.e. velocity converges to zero. The simulation parameters are the ones described below Figure 1.

As it has been shown in [27], in the zero-inertia limit ($m \rightarrow 0$), one may expect to obtain the recent developed Consensus-Based Optimization (CBO) dynamics [13, 25, 28, 47] satisfying

$$dX_t^i = \lambda(X_t^\alpha(\rho^N) - X_t^i)dt + \sigma D(X_t^\alpha(\rho^N) - X_t^i)dB_t^i, \quad i = 1, \dots, N, \quad (1.6)$$

where

$$X_t^\alpha(\rho^N) := \frac{\int_{\mathbb{R}^d} x \omega_\alpha^\mathcal{E}(x) \rho^N(t, dx)}{\int_{\mathbb{R}^d} \omega_\alpha^\mathcal{E}(x) \rho^N(t, dx)} \quad \text{with} \quad \rho^N(t, dx) := \frac{1}{N} \sum_{i=1}^N \delta_{X_t^i}(dx).$$

It has been proved that CBO is a powerful and robust method to solve many interesting non-convex high-dimensional optimization problems in machine learning [15]. By now, CBO methods have also been generalized to optimization over manifolds [22–24, 40]. The objective of the present paper is to complete a theory gap suggested in [27] by providing a rigorous proof of the zero-inertia limit.

On the one hand, as $N \rightarrow \infty$, the mean-field limit results (see [9, 12, 31–33, 36, 52] for instance) indicate that our PSO dynamics (1.2) converges to the solutions of the following mean-field nonlinear McKean systems:

$$\begin{cases} d\bar{X}_t^m = \bar{V}_t^m dt, & (1.7a) \\ d\bar{V}_t^m = -\frac{\gamma}{m} \bar{V}_t^m dt + \frac{\lambda}{m} (X_t^\alpha(\rho^m) - \bar{X}_t^m) dt + \frac{\sigma}{m} D(X_t^\alpha(\rho^m) - \bar{X}_t^m) dB_t, & (1.7b) \end{cases}$$

where

$$X_t^\alpha(\rho^m) = \frac{\int_{\mathbb{R}^d} x \omega_\alpha^\mathcal{E}(x) \rho^m(t, dx)}{\int_{\mathbb{R}^d} \omega_\alpha^\mathcal{E}(x) \rho^m(t, dx)}, \quad \rho^m(t, x) = \int_{\mathbb{R}^d} f^m(t, x, dv), \quad (1.8)$$

and the initial data (\bar{X}_0, \bar{V}_0) is the same as in (1.2). Here $f^m(t, x, v)$ is the distribution of $(\bar{X}_t^m, \bar{V}_t^m)$ at time t , which makes the set of equations (1.7) nonlinear. We refer to [31] for a proof the well-posedness of the PSO particle system (1.2) and its mean-field dynamic (1.7). A direct application of the Itô-Doeblin formula yields that the law $f_t^m := f^m(t, \cdot, \cdot)$ at time t is a weak solution to the following nonlinear Vlasov-Fokker-Plank equation

$$\partial_t f_t^m + v \cdot \nabla_x f_t^m = \nabla_v \cdot \left(\frac{\gamma}{m} v f_t^m + \frac{\lambda}{m} (x - X_t^\alpha(\rho^m)) f_t + \frac{\sigma^2}{2m^2} [D(x - X_t^\alpha(\rho^m))]^2 \nabla_v f_t^m \right), \quad (1.9)$$

with the initial data $f_0^m(x, v) = \text{Law}(\bar{X}_0, \bar{V}_0)$. On the other hand, taking $N \rightarrow \infty$ in (1.6) leads to the mean-field CBO dynamic of the form

$$d\bar{X}_t = \lambda(X_t^\alpha(\rho) - \bar{X}_t)dt + \sigma D(X_t^\alpha(\rho) - \bar{X}_t)dB_t \quad (1.10)$$

with $\rho_t = \text{Law}(\bar{X}_t)$ satisfying the corresponding CBO equation

$$\partial_t \rho_t + \lambda \nabla_x \cdot (\rho_t(X_t^\alpha(\rho) - x)) = \frac{\sigma^2}{2} \sum_{j=1}^d \frac{\partial^2}{\partial x_j^2} \left(\rho_t(x_j - (X_t^\alpha(\rho))_j)^2 \right). \quad (1.11)$$

In this paper, we prove that in the zero-inertia limit, as $m \rightarrow 0^+$, the processes $\{\bar{X}^m\}$ satisfying SDEs (1.7) converge to the solution \bar{X} to the SDE (1.10) in the continuous path space $\mathcal{C}([0, T]; \mathbb{R}^d)$. A convergence rate is obtained and the generalization to the case with memory effects is also addressed. This is related to the study of the overdamped limit [16, 20, 41], or large friction limit [11, 21, 35] for Vlasov type equations. However, the nonlinear term $X_t^\alpha(\rho^m)$ here makes our model very different from theirs, which is nonstandard in the literature. Moreover, all of those results mentioned earlier are obtained through the investigation of PDEs like (1.9) and (1.11), while in the present paper we adopt a probabilistic approach by investigating the convergence of the non-Markovian stochastic processes $\{\bar{X}^m\}$ satisfying the SDE (1.7) to the solution $\{\bar{X}\}$ to SDE (1.10) in the continuous path space.

The rest of the paper is organized as follows: In Section 2 we verify the tightness of the PSO model (1.7) through Aldous criteria, which allows us to obtain the zero-inertia limit from the PSO model (1.7) towards the CBO model (1.10) as $m \rightarrow 0^+$; see Theorem 2.4. Then in Section 3 we generalize the result to the PSO model with memory effects of the local best positions. Lastly we conclude this paper in Section 4 by reporting a few instructive numerical experiments that aim to validating the zero-inertia limit.

2. ZERO-INERTIA LIMIT

Throughout this work, the letter C denotes a generic constant whose value may vary from line to line and its dependence on certain model parameters will be specified whenever needed. We start this section with the standing assumption on the cost function \mathcal{E} .

Assumption 1. The given cost function $\mathcal{E} : \mathbb{R}^d \rightarrow \mathbb{R}$ is locally Lipschitz continuous and satisfies the properties:

- (1) There exists some constant $L > 0$ such $|\mathcal{E}(x) - \mathcal{E}(y)| \leq L(|x| + |y|)|x - y|$ for all $x, y \in \mathbb{R}^d$;
- (2) \mathcal{E} is uniformly bounded, i.e. $-\infty < \underline{\mathcal{E}} := \inf \mathcal{E} \leq \mathcal{E} \leq \sup \mathcal{E} =: \bar{\mathcal{E}} < +\infty$, and define $C_{\alpha, \mathcal{E}} := e^{\alpha(\bar{\mathcal{E}} - \underline{\mathcal{E}})}$.

The following theorem gives the well-posedness of the mean-field PSO dynamics (1.7) whose proof is analogous to [31, Theorem 2.3] and [13, Theorem 3.1], and thus omitted.

Theorem 2.1. *Let Assumption 1 hold. If $(\bar{X}_0^m, \bar{V}_0^m) = (\bar{X}_0, \bar{V}_0)$ is distributed according to f_0 with $f_0 \in \mathcal{P}_4(\mathbb{R}^{2d})$, then for each $m \in (0, 1]$ and $T > 0$, the nonlinear SDE (1.7) admits a unique strong solution up to time T with the initial data $(\bar{X}_0^m, \bar{V}_0^m)$, and it holds further that*

$$\sup_{t \in [0, T]} \mathbb{E} \left[|\bar{X}_t^m|^4 + |\bar{V}_t^m|^4 \right] \leq e^{CT} \cdot \mathbb{E} [|\bar{X}_0|^4 + |\bar{V}_0|^4], \quad (2.1)$$

where C depends only on λ, m, σ , and $C_{\alpha, \varepsilon}$.

Solving (1.7b) for \bar{V}_t^m gives

$$\bar{V}_t^m = e^{-\frac{\gamma}{m}t} \left(\bar{V}_0 + \frac{\lambda}{m} \int_0^t e^{\frac{\gamma s}{m}} (X_s^\alpha(\rho^m) - \bar{X}_s^m) ds + \frac{\sigma}{m} \int_0^t e^{\frac{\gamma s}{m}} D(X_s^\alpha(\rho^m) - \bar{X}_s^m) dB_s \right),$$

which implies that

$$\begin{aligned} \bar{X}_t^m &= \bar{X}_0 + \int_0^t \bar{V}_\tau d\tau = \bar{X}_0 + \int_0^t e^{-\frac{\gamma}{m}\tau} \bar{V}_0 d\tau + \frac{\lambda}{m} \int_0^t e^{-\frac{\gamma}{m}\tau} \int_0^\tau e^{\frac{\gamma s}{m}} (X_s^\alpha(\rho^m) - \bar{X}_s^m) ds d\tau \\ &\quad + \frac{\sigma}{m} \int_0^t e^{-\frac{\gamma}{m}\tau} \int_0^\tau e^{\frac{\gamma s}{m}} D(X_s^\alpha(\rho^m) - \bar{X}_s^m) dB_s d\tau. \end{aligned} \quad (2.2)$$

Then \bar{X}_t^m has the law ρ_t^m for each $t \geq 0$.

For each $k \in \mathbb{N}^+$, denote by $\mathcal{C}([0, T]; \mathbb{R}^k)$ the space of all \mathbb{R}^k -valued continuous functions on $[0, T]$ equipped with the usual uniform norm $\|\cdot\|_0$. Each continuous stochastic process \bar{X}^m may be seen as a $\mathcal{C}([0, T]; \mathbb{R}^d)$ -valued random function and it induces a probability measure (or law, denoted by ρ^m) on $\mathcal{C}([0, T]; \mathbb{R}^d)$. We shall use the convergence in the space of probability measures on $\mathcal{C}([0, T]; \mathbb{R}^d)$. In what follows, we write $\bar{X}^m \rightharpoonup \bar{X}$ or $\rho^m \rightharpoonup \rho$ with ρ being the law of \bar{X} , if $\{\rho^m\}_{m>0}$, as a sequence of probability measures, converges weakly to ρ , i.e., for each bounded continuous functional Φ on $\mathcal{C}([0, T]; \mathbb{R}^d)$, there holds $\lim_{m \rightarrow 0^+} \mathbb{E} [\Phi(\bar{X}^m)] = \mathbb{E} [\Phi(\bar{X})]$. The weak convergence $\bar{X}^m \rightharpoonup \bar{X}$ is stronger than, and actually implies the convergence of $\{\rho_t^m\}_{m>0}$ to ρ_t with ρ_t being the law of \bar{X}_t for each $t \geq 0$, while the converse need not hold. Moreover, due to the separability and completeness of the space $\mathcal{C}([0, T]; \mathbb{R}^d)$, Prohorov's theorem implies that the relative compactness is equivalent to the tightness; see [7] for more details.

The proof of zero-inertia limit will proceed in two steps:

- The tightness of the sequence of probability distributions $\{\rho^m\}_{0 < m \leq 1}$ of $\{\bar{X}^m\}_{0 < m \leq 1}$ is justified by using Aldous tightness criteria.
- We will check that all the limit points of $\{\bar{X}^m\}_{0 < m \leq 1}$ as $m \rightarrow 0$ satisfy mean-field CBO dynamic (1.10) which in fact admits a unique strong solution.

For the sake of completeness, we recall a result concluded directly from the Aldous tightness criteria [37, Theorem 4.5]; it might have been existing somewhere which, however, we did not find, so its straightforward proof is sketched below.

Lemma 2.1. *Let $\{X^n\}_{n \in \mathbb{N}}$ be a sequence of random variables defined on a probability space $(\Omega, \mathcal{F}, \mathbb{P})$ and valued in $\mathcal{C}([0, T]; \mathbb{R}^d)$. The sequence of probability distributions $\{\mu_{X^n}\}_{n \in \mathbb{N}}$ of $\{X^n\}_{n \in \mathbb{N}}$ is tight if the following two conditions hold.*

(Con1) *For all $t \in [0, T]$, the set of distributions of X_t^n , denoted by $\{\mu_{X_t^n}\}_{n \in \mathbb{N}}$, is tight in \mathbb{R}^d .*

(Con2) *For all $\varepsilon > 0$, $\eta > 0$, there exists $\delta_0 > 0$ and $n_0 \in \mathbb{N}$ such that for all $n \geq n_0$ and for all $(\sigma(X_s^n; s \in [0, t]))_{t \geq 0}$ -stopping times β satisfying $0 \leq \beta + \delta_0 \leq T$, it holds that*

$$\sup_{\delta \in [0, \delta_0]} \mathbb{P} (|X_{\beta+\delta}^n - X_\beta^n| \geq \eta) \leq \varepsilon. \quad (2.3)$$

Sketched proof. First, we note that the Aldous tightness criteria is normally stated for the tightness of stochastic processes valued in the space of all càdlàg (right continuous with left limits) functions under the so-called Skorokhod topology (see [2, Theorem 1], [7, Theorem 16.10, Page 178] or [37, Theorem 4.5, Page 356]). In fact, by the assertion at the beginning of the proof of [2, Theorem 1], the condition (Con2) implies Hypothesis (A) therein (see [2, Pages 335 & 338]). Therefore, applying [2, Theorem 1] directly gives the tightness of $\{X^n\}_{n \in \mathbb{N}}$ in the space of all càdlàg functions under the Skorokhod topology.

Noteworthy, the condition (Con1) implies that

$$\lim_{\eta \rightarrow \infty} \limsup_{n \rightarrow \infty} \mathbb{P}(|X_t^n| > \eta) = 0, \quad \forall t \in [0, T], \quad (2.4)$$

Indeed, for each $t \in [0, T]$ and $\epsilon > 0$, the tightness in condition (Con1) indicates that there exists a compact set $\mathcal{K}_\epsilon \subset \mathbb{R}^d$ such that $\sup_{n \in \mathbb{N}} \mathbb{P}(X_t^n \notin \mathcal{K}_\epsilon) < \epsilon$. Choosing $\eta_\epsilon > 0$ to be so big that $\mathcal{K}_\epsilon \subset \{x \in \mathbb{R}^d : |x| \leq \eta_\epsilon\}$, we have $\sup_{n \in \mathbb{N}} \mathbb{P}(|X_t^n| > \eta_\epsilon) < \epsilon$, which by the arbitrariness of t and ϵ implies (2.4).

Further, by [37, Proposition 3.26 (i) & (iii), Page 351], because each X^n herein is valued in $\mathcal{C}([0, T]; \mathbb{R}^d)$ having continuous trajectories, we have the \mathcal{C} -tightness of $\{X^n\}_{n \in \mathbb{N}}$ which means the time-continuity of the paths of the limit(s), whereas this is still under the Skorokhod topology. However, in view of the equivalence of \mathcal{C} -tightness in [37, Proposition 3.26 (i) & (ii), Page 351], the \mathcal{C} -tightness implies further that for all $\epsilon > 0$ and $\eta > 0$, there are $N_0 \in \mathbb{N}$ and $\theta \in (0, T]$ such that

$$\sup_{n \geq N_0} \mathbb{P} \left(\sup_{r, s \in [0, T], |r-s| \leq \theta} |X_s^n - X_r^n| > \eta \right) \leq \epsilon. \quad (2.5)$$

Combining (2.4) and (2.5), we may use the tightness criteria for probability measures on $(\mathcal{C}([0, T]; \mathbb{R}^d), \|\cdot\|_0)$ in [7, Theorem 7.3, Page 82] to obtain the tightness of $\{\mu_{X^n}\}_{n \in \mathbb{N}}$ on $(\mathcal{C}([0, T]; \mathbb{R}^d), \|\cdot\|_0)$ that is endowed with the uniform norm. \square

Theorem 2.2 (Tightness). *Let Assumption 1 hold and $(X_t^m, V_t^m)_{t \in [0, T]}$ satisfy the system (1.7). For each countable subsequence $\{m_k\}_{k \in \mathbb{N}} \subset (0, 1]$ with $\lim_{k \rightarrow \infty} m_k = 0$, the sequence of probability distributions $\{\rho^{m_k}\}_{k \in \mathbb{N}}$ of $\{\bar{X}^{m_k}\}_{k \in \mathbb{N}}$ is tight on $(\mathcal{C}([0, T]; \mathbb{R}^d), \|\cdot\|_0)$.*

Proof. It is sufficient to justify conditions (Con1) and (Con2) in Lemma 2.1.

• *Step 1: Checking (Con1).* First, for $0 < m \leq \frac{1}{2}$, recalling (2.2), we have by Fubini's theorem (see [18, Theorem 4.33] for the stochastic version)

$$\begin{aligned} \bar{X}_t^m &= \bar{X}_0 + \int_0^t e^{-\frac{\gamma}{m}\tau} \bar{V}_0 d\tau + \frac{\lambda}{m} \int_0^t \int_0^\tau e^{-\frac{\gamma}{m}(\tau-s)} (X_s^\alpha(\rho^m) - \bar{X}_s^m) ds d\tau \\ &\quad + \frac{\sigma}{m} \int_0^t \int_0^\tau e^{-\frac{\gamma}{m}(\tau-s)} D(X_s^\alpha(\rho^m) - \bar{X}_s^m) dB_s d\tau \\ &= \bar{X}_0 + \int_0^t e^{-\frac{\gamma}{m}\tau} \bar{V}_0 d\tau + \frac{\lambda}{m} \int_0^t \int_s^t e^{-\frac{\gamma}{m}(\tau-s)} d\tau (X_s^\alpha(\rho^m) - \bar{X}_s^m) ds \\ &\quad + \frac{\sigma}{m} \int_0^t \int_s^t e^{-\frac{\gamma}{m}(\tau-s)} d\tau D(X_s^\alpha(\rho^m) - \bar{X}_s^m) dB_s \\ &= \bar{X}_0 + \frac{m}{\gamma} (1 - e^{-\frac{\gamma}{m}t}) \bar{V}_0 + \frac{\lambda}{\gamma} \int_0^t (1 - e^{-\frac{\gamma}{m}(t-s)}) (X_s^\alpha(\rho^m) - \bar{X}_s^m) ds \\ &\quad + \frac{\sigma}{\gamma} \int_0^t (1 - e^{-\frac{\gamma}{m}(t-s)}) D(X_s^\alpha(\rho^m) - \bar{X}_s^m) dB_s. \end{aligned} \quad (2.6)$$

Here, the time $t \in [0, T]$ is deterministic and the stochastic Fubini's theorem is applicable as there holds the following integrability:

$$\begin{aligned}
& \int_0^t \left(\mathbb{E} \left[\int_0^\tau e^{-\frac{2\gamma}{m}(\tau-s)} |X_s^\alpha(\rho^m) - \bar{X}_s^m|^2 ds \right] \right)^{\frac{1}{2}} d\tau \\
& \leq \int_0^t \left(\int_0^\tau e^{-\frac{2\gamma}{m}(\tau-s)} C \mathbb{E}[|\bar{X}_s^m|^2] ds \right)^{\frac{1}{2}} d\tau \leq (C \sup_{r \in [0, T]} \mathbb{E}[|\bar{X}_r^m|^2])^{\frac{1}{2}} \int_0^t \left(\int_0^\tau e^{-\frac{2\gamma}{m}(\tau-s)} ds \right)^{\frac{1}{2}} d\tau \\
& \leq (C \sup_{r \in [0, T]} \mathbb{E}[|\bar{X}_r^m|^2])^{\frac{1}{2}} T^{\frac{1}{2}} \left(\int_0^t \int_0^\tau e^{-\frac{2\gamma}{m}(\tau-s)} ds d\tau \right)^{\frac{1}{2}} = (C \sup_{r \in [0, T]} \mathbb{E}[|\bar{X}_r^m|^2])^{\frac{1}{2}} T^{\frac{1}{2}} \left(\int_0^t \frac{m}{2\gamma} (1 - e^{-\frac{2\gamma}{m}\tau}) d\tau \right)^{\frac{1}{2}} \\
& \leq (C \sup_{r \in [0, T]} \mathbb{E}[|\bar{X}_r^m|^2])^{\frac{1}{2}} T \left(\frac{m}{2\gamma} \right)^{\frac{1}{2}} < \infty,
\end{aligned}$$

where we have used the fact that for all real $p \geq 1$,

$$\begin{aligned}
\mathbb{E}[|X_t^\alpha(\rho^m) - \bar{X}_t^m|^p] &= \int_{\mathbb{R}^d} \left| \frac{\int_{\mathbb{R}^d} x \omega_\alpha^\mathcal{E}(x) \rho^m(t, dx)}{\int_{\mathbb{R}^d} \omega_\alpha^\mathcal{E}(x) \rho^m(t, dx)} - y \right|^p \rho^m(t, dy) = \int_{\mathbb{R}^d} \left| \frac{\int_{\mathbb{R}^d} (x - y) \omega_\alpha^\mathcal{E}(x) \rho^m(t, dx)}{\int_{\mathbb{R}^d} \omega_\alpha^\mathcal{E}(x) \rho^m(t, dx)} \right|^p \rho^m(t, dy) \\
&\leq \frac{\int_{\mathbb{R}^d} \int_{\mathbb{R}^d} |x - y|^p \omega_\alpha^\mathcal{E}(x) \rho^m(t, dx) \rho^m(t, dy)}{\int_{\mathbb{R}^d} \omega_\alpha^\mathcal{E}(x) \rho^m(t, dx)} \leq 2^p C_{\alpha, \mathcal{E}} \mathbb{E}[|\bar{X}_t^m|^p], \quad \forall t \in [0, T]. \tag{2.7}
\end{aligned}$$

Please note that in the stochastic integral

$$\frac{\sigma}{\gamma} \int_0^t (1 - e^{-\frac{\gamma}{m}(t-s)}) D(X_s^\alpha(\rho^m) - \bar{X}_s^m) dB_s,$$

the integrand is not just a function of s but also a function of t . Thus, the above stochastic integral is typically of the Volterra type and as a stochastic process of time t , it is not a local martingale except for certain trivial cases, so the BDG and Doob's inequalities for local martingales are not applicable herein. In what follows, we use the fact that for any sequence $\{a_i\}_{i=1}^n$ and $p \geq 1$, there holds

$$\left(\left| \sum_{i=1}^n a_i \right| \right)^p \leq n^{p-1} \sum_{i=1}^n |a_i|^p.$$

Note that the assumption on $0 < m \leq \frac{1}{2}$ ensures that $\gamma = 1 - m \in [\frac{1}{2}, 1)$, so $\frac{1}{\gamma}$ is well defined. It follows from Hölder's inequality that

$$|\bar{X}_t^m|^4 \leq 64 |\bar{X}_0|^4 + \frac{64m^4}{\gamma^4} |\bar{V}_0|^4 + \frac{64\lambda^4 t^3}{\gamma^4} \int_0^t |X_s^\alpha(\rho^m) - \bar{X}_s^m|^4 ds + \frac{64\sigma^4}{\gamma^4} \left| \int_0^t (1 - e^{-\frac{\gamma}{m}(t-s)}) D(X_s^\alpha(\rho^m) - \bar{X}_s^m) dB_s \right|^4. \tag{2.8}$$

Using the moment inequality for stochastic integrals as in [44, Theorem 7.1], we have

$$\begin{aligned}
& \mathbb{E} \left[\left| \int_0^t (1 - e^{-\frac{\gamma}{m}(t-s)}) D(X_s^\alpha(\rho^m) - \bar{X}_s^m) dB_s \right|^4 \right] \\
& \leq d^3 \mathbb{E} \left[\sum_{k=1}^d \left| \int_0^t (1 - e^{-\frac{\gamma}{m}(t-s)}) (X_s^\alpha(\rho^m) - \bar{X}_s^m)_k dB_s^k e_k \right|^4 \right] \\
& \leq 36d^3 t \int_0^t \mathbb{E} \left[\sum_{k=1}^d |(X_s^\alpha(\rho^m) - \bar{X}_s^m)_k|^4 \right] ds \leq 36d^3 t \int_0^t \mathbb{E} [|X_s^\alpha(\rho^m) - \bar{X}_s^m|^4] ds. \tag{2.9}
\end{aligned}$$

Thus,

$$\mathbb{E}[|\bar{X}_t^m|^4] \leq 64 \mathbb{E}[|\bar{X}_0|^4] + \frac{64m^4}{\gamma^4} \mathbb{E}[|\bar{V}_0|^4] + \frac{64(\lambda^4 t^3 + 36d^3 t \sigma^4)}{\gamma^4} \int_0^t \mathbb{E}[|X_s^\alpha(\rho^m) - \bar{X}_s^m|^4] ds.$$

Thus, we have

$$\mathbb{E}[|\bar{X}_t^m|^4] \leq 64\mathbb{E}[|\bar{X}_0|^4] + \frac{64m^4}{\gamma^4}\mathbb{E}[|\bar{V}_0|^4] + \frac{1024C_{\alpha,\varepsilon}(\lambda^4t^3 + 36d^3t\sigma^4)}{\gamma^4} \int_0^t \mathbb{E}[|\bar{X}_s^m|^4]ds.$$

Using Gronwall's inequality leads to

$$\mathbb{E}[|\bar{X}_t^m|^4] \leq \left(64\mathbb{E}[|\bar{X}_0|^4] + \frac{64m^4}{\gamma^4}\mathbb{E}[|\bar{V}_0|^4]\right) \exp\left(\frac{1024C_{\alpha,\varepsilon}(\lambda^4T^3 + 36d^3T\sigma^4)}{\gamma^4}T\right), \quad \forall t \in [0, T]. \quad (2.10)$$

Recalling $0 \leq m \leq \frac{1}{2}$ and $\frac{1}{\gamma} = \frac{1}{1-m} \leq 2$, from estimate (2.10) we obtain the boundedness:

$$\sup_{t \in [0, T]} \mathbb{E}[|\bar{X}_t^m|^4] \leq C(\mathbb{E}[|\bar{X}_0|^4], \mathbb{E}[|\bar{V}_0|^4], C_{\alpha,\varepsilon}, \lambda, d, \sigma, T). \quad (2.11)$$

Next, we consider the case when $\frac{1}{2} \leq m \leq 1$. It is obvious that

$$\begin{aligned} |\bar{X}_t^m|^4 &= |\bar{X}_0|^4 + 4 \int_0^t |\bar{X}_s^m|^2 \bar{X}_s^m \cdot \bar{V}_s^m ds \leq |\bar{X}_0|^4 + 2 \int_0^t |\bar{X}_s^m|^2 (|\bar{X}_s^m|^2 + |\bar{V}_s^m|^2) ds \\ &\leq |\bar{X}_0|^4 + 3 \int_0^t (|\bar{X}_s^m|^4 + |\bar{V}_s^m|^4) ds. \end{aligned} \quad (2.12)$$

Using (1.7b), arguments similar to (2.8)-(2.9) give

$$\begin{aligned} |\bar{V}_t^m|^4 &\leq 64|\bar{V}_0|^4 + \frac{64\gamma^4}{m^4} \left| \int_0^t \bar{V}_s^m ds \right|^4 + \frac{64\lambda^4}{m^4} \left| \int_0^t (X_s^\alpha(\rho^m) - \bar{X}_s^m) ds \right|^4 + \frac{64\sigma^4}{m^4} \left| \int_0^t D(X_s^\alpha(\rho^m) - \bar{X}_s^m) dB_s \right|^4 \\ &\leq 64|\bar{V}_0|^4 + \frac{64\gamma^4t^3}{m^4} \int_0^t |\bar{V}_s^m|^4 ds + \frac{64\lambda^4t^3}{m^4} \int_0^t |X_s^\alpha(\rho^m) - \bar{X}_s^m|^4 ds + \frac{64\sigma^4 \cdot 36d^3t}{m^4} \int_0^t |X_s^\alpha(\rho^m) - \bar{X}_s^m|^4 ds. \end{aligned} \quad (2.13)$$

Collecting estimates (2.13) and (2.12), and recalling the fact $\frac{1}{2} \leq m \leq 1$ and $0 \leq \gamma \leq 1$, we have

$$\begin{aligned} &\mathbb{E}[|\bar{X}_t^m|^4 + |\bar{V}_t^m|^4] \\ &\leq 64\mathbb{E}[|\bar{X}_0|^4 + |\bar{V}_0|^4] + C \int_0^t \mathbb{E}[|\bar{X}_s^m|^4 + |\bar{V}_s^m|^4] ds + C \int_0^t \mathbb{E}[|X_s^\alpha(\rho^m) - \bar{X}_s^m|^4] ds \\ &\leq 64\mathbb{E}[|\bar{X}_0|^4 + |\bar{V}_0|^4] + C(1 + 8C_{\alpha,\varepsilon}) \int_0^t \mathbb{E}[|\bar{X}_s^m|^4 + |\bar{V}_s^m|^4] ds, \end{aligned} \quad (2.14)$$

where the estimate (2.7) is used in the last inequality. Applying Gronwall's inequality yields that

$$\mathbb{E}[|\bar{X}_t^m|^4 + |\bar{V}_t^m|^4] \leq 64\mathbb{E}[|\bar{X}_0|^4 + |\bar{V}_0|^4] \exp(C \cdot (1 + 8C_{\alpha,\varepsilon})t), \quad \forall t \in [0, T]. \quad (2.15)$$

Finally, combining (2.11) and (2.15) yields that

$$\sup_{m \in (0, 1]} \sup_{t \in [0, T]} \mathbb{E}[|\bar{X}_t^m|^4] \leq C(\mathbb{E}[|\bar{X}_0|^4], \mathbb{E}[|\bar{V}_0|^4], C_{\alpha,\varepsilon}, \lambda, \sigma, d, T) =: C_1 \quad (2.16)$$

where the constant $C_1 > 0$ is independent of m . Therefore, for any $\varepsilon > 0$, there exists a compact subset $K_\varepsilon := \{x : |x|^4 \leq \frac{C_1}{\varepsilon}\}$ such that by Markov's inequality

$$\rho_t^m((K_\varepsilon)^c) = \mathbb{P}(|X_t^m|^4 > \frac{C_1}{\varepsilon}) \leq \frac{\varepsilon \mathbb{E}[|X_t^m|^4]}{C_1} \leq \varepsilon, \quad \forall m \in (0, 1]. \quad (2.17)$$

This means that for each $t \in [0, T]$, each countable subset of $\{\rho_t^m\}_{0 < m \leq 1}$ is tight, which verifies condition (Con1) in Lemma 2.1.

• *Step 2: Checking (Con2).* Let β be a $(\sigma(X_s^m; s \in [0, t]))_{t \geq 0}$ -stopping time such that $\beta + \delta_0 \leq T$. Without any loss of generality, we may assume that the concerned countable subsequence $\{m_k\}_{k \in \mathbb{N}} \subset [0, 1]$ satisfies $m_k \leq \frac{1}{2}$ for all $k \in \mathbb{N}$; thus, we may just consider the case of $0 < m \leq \frac{1}{2}$ which indicates $\frac{1}{2} \leq \gamma < 1$. Recall (2.2) and compute

$$\bar{X}_{\beta+\delta}^m - \bar{X}_\beta^m$$

$$\begin{aligned}
&= \int_{\beta}^{\beta+\delta} \bar{V}_{\tau} d\tau = \int_{\beta}^{\beta+\delta} e^{-\frac{\gamma}{m}\tau} \bar{V}_0 d\tau + \frac{\lambda}{m} \int_{\beta}^{\beta+\delta} \int_0^{\tau} e^{-\frac{\gamma}{m}(\tau-s)} (X_s^{\alpha}(\rho^m) - \bar{X}_s^m) ds d\tau \\
&\quad + \frac{\sigma}{m} \int_{\beta}^{\beta+\delta} e^{-\frac{\gamma\tau}{m}} \int_0^{\beta} e^{\frac{\gamma s}{m}} D(X_s^{\alpha}(\rho^m) - \bar{X}_s^m) dB_s d\tau + \frac{\sigma}{m} \int_{\beta}^{\beta+\delta} e^{-\frac{\gamma\tau}{m}} \int_{\beta}^{\tau} e^{\frac{\gamma s}{m}} D(X_s^{\alpha}(\rho^m) - \bar{X}_s^m) dB_s d\tau \\
&= \int_{\beta}^{\beta+\delta} e^{-\frac{\gamma}{m}\tau} \bar{V}_0 d\tau + \frac{\lambda}{m} \int_0^{\beta} \int_{\beta}^{\beta+\delta} e^{-\frac{\gamma}{m}(\tau-s)} d\tau (X_s^{\alpha}(\rho^m) - \bar{X}_s^m) ds + \frac{\lambda}{m} \int_{\beta}^{\beta+\delta} \int_s^{\beta+\delta} e^{-\frac{\gamma}{m}(\tau-s)} d\tau (X_s^{\alpha}(\rho^m) - \bar{X}_s^m) ds \\
&\quad + \frac{\sigma}{m} \int_{\beta}^{\beta+\delta} e^{-\frac{\gamma\tau}{m}} d\tau \int_0^{\beta} e^{\frac{\gamma s}{m}} D(X_s^{\alpha}(\rho^m) - \bar{X}_s^m) dB_s + \frac{\sigma}{m} \int_{\beta}^{\beta+\delta} \int_s^{\beta+\delta} e^{-\frac{\gamma}{m}(\tau-s)} d\tau D(X_s^{\alpha}(\rho^m) - \bar{X}_s^m) dB_s \\
&= \frac{m}{\gamma} (e^{-\frac{\gamma}{m}\beta} - e^{-\frac{\gamma}{m}(\beta+\delta)}) \bar{V}_0 \\
&\quad + \frac{\lambda}{\gamma} \int_0^{\beta} (e^{-\frac{\gamma}{m}(\beta-s)} - e^{-\frac{\gamma}{m}(\beta+\delta-s)}) (X_s^{\alpha}(\rho^m) - \bar{X}_s^m) ds + \frac{\lambda}{\gamma} \int_{\beta}^{\beta+\delta} (1 - e^{-\frac{\gamma}{m}(\beta+\delta-s)}) (X_s^{\alpha}(\rho^m) - \bar{X}_s^m) ds \\
&\quad + \frac{\sigma}{\gamma} (e^{-\frac{\gamma\beta}{m}} - e^{-\frac{\gamma(\beta+\delta)}{m}}) \int_0^{\beta} e^{\frac{\gamma s}{m}} D(X_s^{\alpha}(\rho^m) - \bar{X}_s^m) dB_s + \frac{\sigma}{\gamma} \int_{\beta}^{\beta+\delta} (1 - e^{-\frac{\gamma}{m}(\beta+\delta-s)}) D(X_s^{\alpha}(\rho^m) - \bar{X}_s^m) dB_s, \quad (2.18)
\end{aligned}$$

where the calculations have taken into account the fact that β is a stopping time.

Note that there holds $|e^{-x} - e^{-y}| \leq |x - y|^{\zeta} \wedge 1$ for all $x, y \in [0, \infty)$ and $\zeta \in [0, 1]$. Basic computations further indicate that for each $q \geq 1$, $\zeta \in [0, 1]$, and $\tau \in [0, T]$,

$$\begin{aligned}
\int_0^{\tau} \left| e^{-\frac{\gamma(\tau-s)}{m}} - e^{-\frac{\gamma(\tau+\delta-s)}{m}} \right|^q ds &\leq \int_0^{\tau} \left(e^{-\frac{\gamma(\tau-s)}{m}} - e^{-\frac{\gamma(\tau+\delta-s)}{m}} \right) ds = \frac{m}{\gamma} \left(1 - e^{-\frac{\gamma\delta}{m}} \right) - \frac{m}{\gamma} \left(e^{-\frac{\gamma\tau}{m}} - e^{-\frac{\gamma(\tau+\delta)}{m}} \right) \\
&\leq \frac{m}{\gamma} \cdot \left(\frac{\gamma\delta}{m} \right)^{\zeta} = \left(\frac{m}{\gamma} \right)^{1-\zeta} \delta^{\zeta}, \quad (2.19)
\end{aligned}$$

and obviously,

$$\int_{\beta}^{\beta+\delta} \left(1 - e^{-\frac{\gamma(\beta+\delta-s)}{m}} \right)^q ds \leq \int_{\beta}^{\beta+\delta} 1 ds = \delta.$$

Then, it is obvious that

$$\mathbb{E} \left[\left| \frac{m}{\gamma} (e^{-\frac{\gamma}{m}\beta} - e^{-\frac{\gamma}{m}(\beta+\delta)}) \bar{V}_0 \right|^2 \right] \leq \frac{m^2}{\gamma^2} \cdot \frac{\gamma^2 \delta^2}{m^2} (\mathbb{E}[|\bar{V}_0|^4])^{\frac{1}{2}} \leq \delta^2 (\mathbb{E}[|\bar{V}_0|^4])^{\frac{1}{2}}.$$

Next, it follows that

$$\begin{aligned}
&\mathbb{E} \left[\left| \int_0^{\beta} (e^{-\frac{\gamma}{m}(\beta-s)} - e^{-\frac{\gamma}{m}(\beta+\delta-s)}) (X_s^{\alpha}(\rho^m) - \bar{X}_s^m) ds \right|^2 \right] \\
&\leq \mathbb{E} \left[\int_0^{\beta} |e^{-\frac{\gamma}{m}(\beta-s)} - e^{-\frac{\gamma}{m}(\beta+\delta-s)}|^2 ds \cdot \int_0^{\beta} |X_s^{\alpha}(\rho^m) - \bar{X}_s^m|^2 ds \right] \\
&\leq \delta \cdot T \sup_{s \in [0, T]} (\mathbb{E}[|X_s^{\alpha}(\rho^m) - \bar{X}_s^m|^4])^{1/2},
\end{aligned}$$

and analogously,

$$\begin{aligned}
\mathbb{E} \left[\left| \int_{\beta}^{\beta+\delta} (1 - e^{-\frac{\gamma}{m}(\beta+\delta-s)}) (X_s^{\alpha}(\rho^m) - \bar{X}_s^m) ds \right|^2 \right] &\leq \mathbb{E} \left[\int_{\beta}^{\beta+\delta} \left(1 - e^{-\frac{\gamma(\beta+\delta-s)}{m}} \right)^2 ds \cdot \int_{\beta}^{\beta+\delta} |X_s^{\alpha}(\rho^m) - \bar{X}_s^m|^2 ds \right] \\
&\leq \delta \cdot \mathbb{E} \left[\int_{\beta}^{\beta+\delta} |X_s^{\alpha}(\rho^m) - \bar{X}_s^m|^2 ds \right] \\
&\leq \delta \cdot T \sup_{s \in [0, T]} (\mathbb{E}[|X_s^{\alpha}(\rho^m) - \bar{X}_s^m|^4])^{1/2}, .
\end{aligned}$$

Further, applying Itô's isometry gives

$$\mathbb{E} \left[\left| \int_{\beta}^{\beta+\delta} (1 - e^{-\frac{\gamma}{m}(\beta+\delta-s)}) D(X_s^{\alpha}(\rho^m) - \bar{X}_s^m) dB_s \right|^2 \right]$$

$$\begin{aligned}
&= \mathbb{E} \left[\int_{\beta}^{\beta+\delta} |1 - e^{-\frac{\gamma}{m}(\beta+\delta-s)}|^2 |X_s^\alpha(\rho^m) - \bar{X}_s^m|^2 ds \right] \\
&\leq \left(\mathbb{E} \left[\int_{\beta}^{\beta+\delta} |1 - e^{-\frac{\gamma}{m}(\beta+\delta-s)}|^4 ds \right] \right)^{1/2} \cdot \left(\mathbb{E} \left[\int_0^T |X_s^\alpha(\rho^m) - \bar{X}_s^m|^4 ds \right] \right)^{1/2} \\
&\leq (\delta T)^{1/2} \left(\sup_{s \in [0, T]} \mathbb{E} \left[|X_s^\alpha(\rho^m) - \bar{X}_s^m|^4 \right] \right)^{1/2}.
\end{aligned}$$

Particularly, let us look at

$$Z_t^{m, \delta} := (e^{-\frac{\gamma t}{m}} - e^{-\frac{\gamma(t+\delta)}{m}}) \int_0^t e^{\frac{\gamma s}{m}} D(X_s^\alpha(\rho^m) - \bar{X}_s^m) dB_s, \quad t \in [0, T],$$

and try to derive an estimate on $Z_\beta^{m, \delta}$. Here, we note that the main difficulty in estimating $Z_\beta^{m, \delta}$ arises from the fact that for a general stopping time β (in fact, unless β is $\sigma(\bar{X}_0, \bar{V}_0)$ -measurable in our setting), the multiplier $(e^{-\frac{\gamma \beta}{m}} - e^{-\frac{\gamma(\beta+\delta)}{m}})$ cannot enter the stochastic integral due to the nonanticipativity required for Itô integrals and associated moment estimates. Basic calculations as above yield that

$$\begin{aligned}
\mathbb{E} \left[\int_0^T |Z_t^{m, \delta}|^4 dt \right] &= \mathbb{E} \left[\int_0^T \left| \int_0^t (e^{-\frac{\gamma(t-s)}{m}} - e^{-\frac{\gamma(t+\delta-s)}{m}}) D(X_s^\alpha(\rho^m) - \bar{X}_s^m) dB_s \right|^4 dt \right] \\
&\leq C \int_0^T \mathbb{E} \left[\left(\int_0^t (e^{-\frac{\gamma(t-s)}{m}} - e^{-\frac{\gamma(t+\delta-s)}{m}}) |X_s^\alpha(\rho^m) - \bar{X}_s^m|^2 ds \right)^2 \right] dt \\
&\leq C \int_0^T \int_0^t (e^{-\frac{\gamma(t-s)}{m}} - e^{-\frac{\gamma(t+\delta-s)}{m}})^4 ds \cdot \mathbb{E} \left[\int_0^t |X_s^\alpha(\rho^m) - \bar{X}_s^m|^4 ds \right] dt \\
&\leq C \delta \int_0^T \mathbb{E} \left[\int_0^t |X_s^\alpha(\rho^m) - \bar{X}_s^m|^4 ds \right] dt \\
&\leq C \delta \sup_{s \in [0, T]} \mathbb{E} \left[|X_s^\alpha(\rho^m) - \bar{X}_s^m|^4 \right], \tag{2.20}
\end{aligned}$$

where we have used the so-called Burkholder-Davis-Gundy (BDG) inequality for martingales and the estimate (2.19) with $\zeta = 1$ and the constant C is independent of m and δ . Thus, the process

$$M_t^{m, \delta} := \int_0^t (Z_s^{m, \delta})' D(X_s^\alpha(\rho^m) - \bar{X}_s^m) dB_s, \quad t \in [0, T],$$

is a square-integrable continuous martingale; indeed, Doob's martingale inequality gives

$$\begin{aligned}
\mathbb{E} \left[\max_{t \in [0, T]} |M_t^{m, \delta}|^2 \right] &\leq C \mathbb{E} \left[\int_0^T |(Z_s^{m, \delta})' D(X_s^\alpha(\rho^m) - \bar{X}_s^m)|^2 ds \right] \\
&\leq C \left(\mathbb{E} \left[\int_0^T |Z_s^{m, \delta}|^4 ds \right] \mathbb{E} \left[\int_0^T |X_s^\alpha(\rho^m) - \bar{X}_s^m|^4 ds \right] \right)^{1/2} \\
&\leq C \delta^{1/2} \sup_{s \in [0, T]} \mathbb{E} \left[|X_s^\alpha(\rho^m) - \bar{X}_s^m|^4 \right], \tag{2.21}
\end{aligned}$$

with C being independent of m and δ . On the other hand, it is easy to see that $Z^{m, \delta}$ satisfies the following SDE

$$dZ_t^{m, \delta} = -\frac{\gamma}{m} Z_t^{m, \delta} dt + (1 - e^{-\frac{\gamma \delta}{m}}) D(X_t^\alpha(\rho^m) - \bar{X}_t^m) dB_t, \quad t > 0; \quad Z_0^{m, \delta} = 0.$$

By Itô-Doebelin formula, it holds that for all $t \in [0, T]$,

$$|Z_t^{m, \delta}|^2 = \int_0^t \left| (e^{-\frac{\gamma(t-s)}{m}} - e^{-\frac{\gamma(t+\delta-s)}{m}}) (X_s^\alpha(\rho^m) - \bar{X}_s^m) \right|^2 ds + 2 \int_0^t e^{-\frac{2\gamma(t-s)}{m}} (1 - e^{-\frac{\gamma \delta}{m}}) dM_s^{m, \delta}$$

$$\begin{aligned}
&\leq \left(\int_0^t \left| e^{-\frac{\gamma(t-s)}{m}} - e^{-\frac{\gamma(t+\delta-s)}{m}} \right|^4 ds \int_0^t \left| X_s^\alpha(\rho^m) - \bar{X}_s^m \right|^4 ds \right)^{1/2} + 2(1 - e^{-\frac{\gamma\delta}{m}}) M_t^{m,\delta} \\
&\quad - 4 \int_0^t \frac{\gamma}{m} e^{-\frac{2\gamma(t-s)}{m}} (1 - e^{-\frac{\gamma\delta}{m}}) M_s^{m,\delta} ds \\
&\leq \left(\delta \int_0^T \left| X_s^\alpha(\rho^m) - \bar{X}_s^m \right|^4 ds \right)^{1/2} + 2 |M_t^{m,\delta}| + \frac{4\gamma}{m} \max_{s \in [0, T]} |M_s^{m,\delta}| \int_0^t e^{-\frac{2\gamma(t-s)}{m}} ds \\
&\leq \left(\delta \int_0^T \left| X_s^\alpha(\rho^m) - \bar{X}_s^m \right|^4 ds \right)^{1/2} + 4 \max_{s \in [0, T]} |M_s^{m,\delta}|, \quad \text{a.s.}, \tag{2.22}
\end{aligned}$$

where the integration by parts formula is applied to the stochastic integral in the first line and in the second inequality, we used estimate (2.19) with ζ equal to 1. Combined with (2.21), it yields that

$$\mathbb{E} \left[\left| Z_\beta^{m,\delta} \right|^2 \right] \leq \mathbb{E} \left[\max_{t \in [0, T]} \left| Z_t^{m,\delta} \right|^2 \right] \leq C \left(\delta^{1/2} + \delta^{1/4} \right) \left(\sup_{s \in [0, T]} \mathbb{E} \left[\left| X_s^\alpha(\rho^m) - \bar{X}_s^m \right|^4 \right] \right)^{1/2},$$

where the constant C is independent of β, m , and δ .

Therefore, summing up the above estimates and recalling $0 < m \leq \frac{1}{2}$, $\frac{1}{\gamma} \leq 2$, and the relations (2.7) and (2.16), we arrive at

$$\begin{aligned}
\mathbb{E} \left[\left| \bar{X}_{\beta+\delta}^m - \bar{X}_\beta^m \right|^2 \right] &\leq 5\delta^2 (\mathbb{E} \left[\left| \bar{V}_0 \right|^4 \right])^{\frac{1}{2}} + \frac{5}{\gamma^2} \cdot \left(\lambda^2 \delta T + \sigma^2 C (\delta^{1/2} + \delta^{1/4}) \right) \sup_{s \in [0, T]} \left(\mathbb{E} \left[\left| X_s^\alpha(\rho^m) - \bar{X}_s^m \right|^4 \right] \right)^{1/2} \\
&\leq C \left(\mathbb{E} \left[\left| \bar{X}_0 \right|^4 \right], \mathbb{E} \left[\left| \bar{V}_0 \right|^4 \right], C_{\alpha, \varepsilon}, \lambda, \sigma, d, T \right) \left(\delta^{\frac{1}{4}} + \delta^{\frac{1}{2}} + \delta + \delta^2 \right).
\end{aligned}$$

Hence, for any $\varepsilon > 0$, $\eta > 0$, there exists some $\delta_0 > 0$ such that for all $0 < m \leq \frac{1}{2}$ it holds that

$$\sup_{\delta \in [0, \delta_0]} \mathbb{P} \left(\left| \bar{X}_{\beta+\delta}^m - \bar{X}_\beta^m \right|^2 \geq \eta \right) \leq \sup_{\delta \in [0, \delta_0]} \frac{\mathbb{E} \left[\left| \bar{X}_{\beta+\delta}^m - \bar{X}_\beta^m \right|^2 \right]}{\eta} \leq \varepsilon. \tag{2.23}$$

This justifies condition *Con2* in Lemma 2.1. \square

In a similar way to (2.7), it holds that for all real $p \geq 1$ and $\tau \in [0, T]$,

$$\begin{aligned}
\mathbb{E} \left[\sup_{t \in [0, \tau]} \left| X_t^\alpha(\rho^m) - \bar{X}_t^m \right|^p \right] &= \mathbb{E} \left[\sup_{t \in [0, \tau]} \left| \frac{\int_{\mathbb{R}^d} (x - \bar{X}_t^m) \omega_\alpha^\varepsilon(x) \rho^m(t, dx)}{\int_{\mathbb{R}^d} \omega_\alpha^\varepsilon(x) \rho^m(t, dx)} \right|^p \right] \\
&\leq \mathbb{E} \left[\sup_{t \in [0, \tau]} \frac{\int_{\mathbb{R}^d} |x - \bar{X}_t^m|^p \omega_\alpha^\varepsilon(x) \rho^m(t, dx)}{\int_{\mathbb{R}^d} \omega_\alpha^\varepsilon(x) \rho^m(t, dx)} \right] \\
&\leq 2^{p-1} C_{\alpha, \varepsilon} \mathbb{E} \left[\sup_{t \in [0, \tau]} \left(\mathbb{E} \left[\left| \bar{X}_t^m \right|^p \right] + \left| \bar{X}_t^m \right|^p \right) \right] \\
&\leq 2^p C_{\alpha, \varepsilon} \mathbb{E} \left[\sup_{t \in [0, \tau]} \left| \bar{X}_t^m \right|^p \right]. \tag{2.24}
\end{aligned}$$

Furthermore, the obtained uniform boundedness (2.16) may be strengthened as follows.

Corollary 2.3. *Let Assumption 1 hold and $(X_t^m, V_t^m)_{t \in [0, T]}$ satisfy the system (1.7) in Theorem 2.2. We have the following uniform boundedness*

$$\sup_{m \in (0, 1]} \mathbb{E} \left[\max_{t \in [0, T]} \left| \bar{X}_t^m \right|^4 \right] \leq C \left(\mathbb{E} \left[\left| \bar{X}_0 \right|^4 \right], \mathbb{E} \left[\left| \bar{V}_0 \right|^4 \right], C_{\alpha, \varepsilon}, \lambda, \sigma, d, T \right) < \infty. \tag{2.25}$$

Proof. For the unique strong solution (\bar{X}^m, \bar{V}^m) to SDE (1.7), applying the standard martingale inequalities (e.g., see [34, Chapter III, Page 110]) to the stochastic integrals in SDE (1.7), we may use the estimate (2.7) with standard arguments (see [42, Section 5 of Chapter 2] for instance) to conclude that the solution (\bar{X}^m, \bar{V}^m) has continuous trajectories and satisfies

$$\mathbb{E} \left[\max_{t \in [0, T]} \left(|\bar{X}_t^m|^4 + |\bar{V}_t^m|^4 \right) \right] \leq C \mathbb{E} [|\bar{X}_0|^4 + |\bar{V}_0|^4], \quad (2.26)$$

where the constant C depends only on λ, m, σ, T , and $C_{\alpha, \varepsilon}$. Also, we recall the uniform boundedness (2.16).

When $0 < m \leq \frac{1}{2}$, we may start with (2.8) and consider the involved stochastic integral:

$$\int_0^t (1 - e^{-\frac{\gamma}{m}(t-s)}) D(X_s^\alpha(\rho^m) - \bar{X}_s^m) dB_s =: \int_0^t D(X_s^\alpha(\rho^m) - \bar{X}_s^m) dB_s - \theta_t, \quad t \in [0, T].$$

The Burkholder-Davis-Gundy (BDG) inequality gives that

$$\mathbb{E} \left[\max_{\tau \in [0, t]} \left| \int_0^\tau D(X_s^\alpha(\rho^m) - \bar{X}_s^m) dB_s \right|^4 \right] \leq C \mathbb{E} \left[\left(\int_0^t |X_s^\alpha(\rho^m) - \bar{X}_s^m|^2 ds \right)^2 \right] \leq CT^2 C_1 < \infty. \quad (2.27)$$

On the other hand, it is obvious that $\theta_t = \int_0^t e^{-\frac{\gamma}{m}(t-s)} D(X_s^\alpha(\rho^m) - \bar{X}_s^m) dB_s$ satisfies uniquely the SDE:

$$d\theta_t = -\frac{\gamma}{m} \theta_t dt + D(X_t^\alpha(\rho^m) - \bar{X}_t^m) dB_t, \quad t > 0; \quad \theta_0 = 0. \quad (2.28)$$

Straightforward calculation implies that

$$\begin{aligned} \mathbb{E} [|\theta_t|^4] &= \mathbb{E} \left[\left| \int_0^t e^{-\frac{\gamma(t-s)}{m}} D(X_s^\alpha(\rho^m) - \bar{X}_s^m) dB_s \right|^4 \right] \\ &\leq C \mathbb{E} \left[\left(\int_0^t e^{-\frac{\gamma(t-s)}{m}} |X_s^\alpha(\rho^m) - \bar{X}_s^m|^2 ds \right)^2 \right] \\ (\text{by (2.7)}) &\leq C \int_0^t e^{-\frac{4\gamma(t-s)}{m}} ds \cdot \int_0^t \mathbb{E} [|\bar{X}_s^m|^4] ds \\ (\text{by (2.16)}) &\leq C m, \quad \forall t \in [0, T]. \end{aligned} \quad (2.29)$$

Set

$$\xi_t = \int_0^t |\theta_s|^2 (\theta_s)' D(X_s^\alpha(\rho^m) - \bar{X}_s^m) dB_s, \quad t \in [0, T]$$

with $(\theta_s)'$ being the transpose of θ_s . Then for each $\varepsilon \in (0, 1)$, we have

$$\begin{aligned} 8 \mathbb{E} \left[\max_{s \in [0, t]} |\xi_s| \right] &\leq C \mathbb{E} \left[\left(\int_0^t |\theta_s|^6 |X_s^\alpha(\rho^m) - \bar{X}_s^m|^2 ds \right)^{1/2} \right] \\ (\text{by (2.24)}) &\leq \frac{C}{\sqrt{\varepsilon}} \mathbb{E} \left[\int_0^t |\theta_s|^4 ds \right] + \frac{1}{4} \mathbb{E} \left[\max_{s \in [0, t]} |\theta_s|^4 \right] + \varepsilon \mathbb{E} \left[\max_{s \in [0, t]} |\bar{X}_s^m|^4 \right] \\ (\text{by (2.29)}) &\leq \frac{C}{\sqrt{\varepsilon}} + \frac{1}{4} \mathbb{E} \left[\max_{s \in [0, t]} |\theta_s|^4 \right] + \varepsilon \mathbb{E} \left[\max_{s \in [0, t]} |\bar{X}_s^m|^4 \right], \quad \forall t \in [0, T]. \end{aligned} \quad (2.30)$$

By Itô-Doeblin formula, it holds that for all $t \in [0, T]$,

$$\begin{aligned} |\theta_t|^4 &= \int_0^t 2e^{-\frac{4\gamma(t-s)}{m}} \left(|\theta_s|^2 |X_s^\alpha(\rho^m) - \bar{X}_s^m|^2 + 2|\theta_s' D(X_s^\alpha(\rho^m) - \bar{X}_s^m)|^2 \right) ds + 4 \int_0^t e^{-\frac{4\gamma(t-s)}{m}} d\xi_s \\ &\leq \frac{C}{\varepsilon} \left(\int_0^t e^{-\frac{4\gamma(t-s)}{m}} ds \right)^2 \max_{s \in [0, t]} |\theta_s|^4 + \frac{\varepsilon}{16C_{\alpha, \varepsilon}} \max_{s \in [0, t]} |X_s^\alpha(\rho^m) - \bar{X}_s^m|^4 + 4\xi_t - \frac{16\gamma}{m} \int_0^t e^{-\frac{4\gamma(t-s)}{m}} \xi_s ds \\ &\leq \frac{\tilde{C}m^2}{\varepsilon} \cdot \max_{s \in [0, t]} |\theta_s|^4 + \frac{\varepsilon}{16C_{\alpha, \varepsilon}} \max_{s \in [0, t]} |X_s^\alpha(\rho^m) - \bar{X}_s^m|^4 + 8 \max_{s \in [0, t]} |\xi_s|, \quad \text{a.s.}, \end{aligned} \quad (2.31)$$

where $\tilde{C} > 0$ is independent of (t, m, ϵ) . Choosing $\tilde{m} = \frac{\epsilon^{1/2}}{2\tilde{C}^{1/2}} \wedge \frac{1}{2}$ so that $\frac{\tilde{C}m^2}{\epsilon} \leq \frac{1}{4}$ and combining (2.24), (2.30) and (2.31), we have

$$\mathbb{E} \left[\max_{s \in [0, t]} |\theta_s|^4 \right] \leq \frac{C}{\sqrt{\epsilon}} + 4\epsilon \mathbb{E} \left[\max_{s \in [0, t]} |\overline{X}_s^m|^4 \right], \quad \forall m \in (0, \tilde{m}], \quad (2.32)$$

which together with (2.27) and (2.16) inserted into (2.8) implies that

$$\mathbb{E} \left[\max_{s \in [0, t]} |\overline{X}_s^m|^4 \right] \leq C + \frac{C}{\sqrt{\epsilon}} + C\epsilon \mathbb{E} \left[\max_{s \in [0, t]} |\overline{X}_s^m|^4 \right], \quad \forall t \in [0, T], \quad (2.33)$$

where the constant $C > 0$ is independent of (t, m, ϵ) . Therefore, by choosing $\epsilon = \frac{1}{2C}$ and setting \tilde{m} accordingly, we obtain the desired uniform boundedness for $m \in (0, \tilde{m}]$. When $\tilde{m} < m < 1$, the proof is standard (see [42, Section 5 of Chapter 2] for instance) and it is omitted. \square

Next we shall identify the limit process, before which we recall a lemma on the stability estimate of the nonlinear term $X^\alpha(\rho)$.

Lemma 2.2. [13, Lemma 3.2] *Assume that $\rho, \hat{\rho} \in \mathcal{P}_4(\mathbb{R}^d)$. Then the following stability estimate holds*

$$|X^\alpha(\rho) - X^\alpha(\hat{\rho})| \leq CW_2(\rho, \hat{\rho}), \quad (2.34)$$

where W_2 is the 2-Wasserstein distance, and C depends only on $\alpha, L, \int_{\mathbb{R}^d} |x|^4 \rho(dx)$, and $\int_{\mathbb{R}^d} |x|^4 \hat{\rho}(dx)$.

Theorem 2.4 (Zero-inertia limit). *Let Assumption 1 hold and $(X_t^m, V_t^m)_{t \in [0, T]}$ satisfy the system (1.7). Then as $m \rightarrow 0^+$, the sequence of stochastic processes $\{\overline{X}^m\}_{0 < m \leq 1}$ converge weakly to \overline{X} , which is the unique solution to the following SDE:*

$$\overline{X}_t = \overline{X}_0 + \lambda \int_0^t (X_s^\alpha(\rho) - \overline{X}_s) ds + \sigma \int_0^t D(X_s^\alpha(\rho) - \overline{X}_s) dB_s. \quad (2.35)$$

Moreover, we have the following convergences:

$$\mathbb{E} \left[\max_{t \in [0, T]} |\overline{X}_t^m - \overline{X}_t|^2 \right] \leq C \sqrt{m}, \quad \max_{t \in [0, T]} \mathbb{E} \left[|\overline{X}_t^m - \overline{X}_t|^2 \right] \leq C m, \quad (2.36)$$

where the constant C depends on $\mathbb{E}[|\overline{X}_0|^4], \mathbb{E}[|\overline{V}_0|^4], C_{\alpha, \mathcal{E}}, \lambda, \sigma, d$, and T .

Remark 2.1. By (2.36), it follows from the definition of Wasserstein distance that

$$\sup_{t \in [0, T]} W_2^2(\rho_t^m, \rho_t) \leq \max_{t \in [0, T]} \mathbb{E} \left[|\overline{X}_t^m - \overline{X}_t|^2 \right] \leq C m, \quad (2.37)$$

which is consistent with the result obtained in [16, Theorem 1.3], where the authors obtained a quantified overdamped limit (with the same rate m) of the singular Vlasov-Poisson-Fokker-Planck system to the aggregation-diffusion equation. Besides, the obtained (strong) convergence of \overline{X}^m to \overline{X} in the path space $\mathcal{C}([0, T]; \mathbb{R}^d)$ implies and is obviously stronger than the convergence of $\{\rho_t^m\}_{m > 0}$ to ρ_t for each time $t \geq 0$.

Proof. By Theorem 2.2, each subsequence $\{\overline{X}^{m_k}\}_{k \in \mathbb{N}}$ with $m_k \leq 1/2$ converging to 0 as $k \rightarrow \infty$ admits a subsequence (denoted w.l.o.g. by itself) that converges weakly. By Skorokhod's representation theorem (see [7, Theorem 6.7 on page 70]), we may find a common probability space $(\Omega, \mathcal{F}, \mathbb{P})$ on which the quadruple $\{(\overline{X}_0, \overline{X}^{m_k}, \overline{X}, B)\}_{k \in \mathbb{N}}$ converge to some $(\widehat{X}_0, \widehat{X}, \overline{X}, B)$ as random variables valued in $\mathbb{R}^d \times \mathcal{C}([0, T]; \mathbb{R}^{3d})$ almost surely. Here B is an identical d -dimensional Wiener process on $(\Omega, \mathcal{F}, \mathbb{P})$. In particular, we have

$$\mathbb{P} \left(\lim_{k \rightarrow \infty} \|\overline{X}^{m_k} - \widehat{X}\|_0 = 0 \right) = 1, \quad (2.38)$$

where we recall the uniform norm $\|\overline{X}^{m_k} - \widehat{X}\|_0 = \max_{t \in [0, T]} |\overline{X}_t^{m_k} - \widehat{X}_t|$. It is obvious that $\overline{X}_0 = \widehat{X}_0$ a.s. In the following, we shall verify that the limit \widehat{X} is indeed the unique solution \overline{X} to SDE (2.35).

Recall that the unique strong solution to an SDE may be regarded as a function of the initial value and the driving Wiener process (see [34, Chapter IV]). Thus, due to the uniqueness and existence of strong solution to SDE (2.6) in Theorem 2.1, on the above probability space $(\Omega, \mathcal{F}, \mathbb{P})$ we must have

$$\begin{aligned} \bar{X}_t^{m_k} &= \bar{X}_0 + \frac{m_k}{\gamma} (1 - e^{-\frac{\gamma}{m_k} t}) \bar{V}_0 + \frac{\lambda}{\gamma} \int_0^t (1 - e^{-\frac{\gamma}{m_k} (t-s)}) (X_s^\alpha(\rho^{m_k}) - \bar{X}_s^{m_k}) ds \\ &\quad + \frac{\sigma}{\gamma} \int_0^t (1 - e^{-\frac{\gamma}{m_k} (t-s)}) D(X_s^\alpha(\rho^{m_k}) - \bar{X}_s^{m_k}) dB_s. \end{aligned} \quad (2.39)$$

By the estimates in Corollary 2.3 and Fatou's lemma, there exists a constant C_2 such that

$$\sup_{k \in \mathbb{N}} \mathbb{E} \left[\max_{t \in [0, T]} |\bar{X}_t^{m_k}|^4 \right] + \mathbb{E} \left[\max_{t \in [0, T]} |\hat{X}_t|^4 \right] \leq C_2 := C(\mathbb{E}[|\bar{X}_0|^4], \mathbb{E}[|\bar{V}_0|^4], C_{\alpha, \varepsilon}, \lambda, \sigma, d, T) < \infty. \quad (2.40)$$

As a straightforward consequence of the above boundedness, it holds that

$$\sup_{k \in \mathbb{N}, t \in [0, T]} \mathbb{P}(|\bar{X}_t^{m_k} - \hat{X}_t| > A) \leq \frac{2^4 C_2}{A^4}, \quad \forall A > 0. \quad (2.41)$$

Thus, the dominated convergence theorem gives that for each $A > 0$,

$$\begin{aligned} &\limsup_{k \rightarrow \infty} \mathbb{E} \left[\int_0^T |\bar{X}_t^{m_k} - \hat{X}_t|^2 dt \right] \\ &\leq \limsup_{k \rightarrow \infty} \left(\mathbb{E} \left[\int_0^T |\bar{X}_t^{m_k} - \hat{X}_t|^2 \wedge A^2 dt \right] + \mathbb{E} \left[\int_0^T |\bar{X}_t^{m_k} - \hat{X}_t|^2 1_{\{|\bar{X}_t^{m_k} - \hat{X}_t| > A\}} dt \right] \right) \\ &\leq \limsup_{k \rightarrow \infty} \mathbb{E} \left[\int_0^T |\bar{X}_t^{m_k} - \hat{X}_t|^2 \wedge A^2 dt \right] + T \cdot \sup_{k \in \mathbb{N}} \left(\mathbb{E} \left[\max_{t \in [0, T]} |\bar{X}_t^{m_k} - \hat{X}_t|^4 \right] \right)^{1/2} \left| \mathbb{P}(|\bar{X}_t^{m_k} - \hat{X}_t| > A) \right|^{1/2} \\ &\leq \limsup_{k \rightarrow \infty} \mathbb{E} \left[\int_0^T |\bar{X}_t^{m_k} - \hat{X}_t|^2 \wedge A^2 dt \right] + \frac{2^4 C_2 T}{A^2} \\ &= \frac{2^4 C_2 T}{A^2}, \end{aligned}$$

which by the arbitrariness of $A > 0$ indicates that

$$\lim_{k \rightarrow \infty} \mathbb{E} \left[\int_0^T |\bar{X}_t^{m_k} - \hat{X}_t|^2 dt \right] = 0. \quad (2.42)$$

Letting $\rho(t, dx)$ be the probability distribution of \hat{X}_t for $t \in [0, T]$, we have

$$|X_t^\alpha(\rho)| = \left| \frac{\int_{\mathbb{R}^d} x \omega_\alpha^\varepsilon(x) \rho(t, dx)}{\int_{\mathbb{R}^d} \omega_\alpha^\varepsilon(x) \rho(t, dx)} \right| \leq C_{\alpha, \varepsilon} \int_{\mathbb{R}^d} |x| \rho(t, dx) \leq C_{\alpha, \varepsilon} (\mathbb{E}[|\hat{X}_t|^4])^{\frac{1}{4}},$$

and

$$\sup_{k \in \mathbb{N}} \sup_{t \in [0, T]} |X_t^\alpha(\rho^{m_k})| \leq C_{\alpha, \varepsilon} (C_2)^{\frac{1}{4}}, \quad \text{and} \quad \sup_{t \in [0, T]} |X_t^\alpha(\rho)| \leq C_{\alpha, \varepsilon} (C_2)^{\frac{1}{4}}. \quad (2.43)$$

Then we compute the limit of (2.39) term by term. By Lemma 2.2, it holds that

$$|X_t^\alpha(\rho^{m_k}) - X_t^\alpha(\rho)|^2 \leq CW_2^2(\rho_t^{m_k}, \rho_t) \leq C \mathbb{E}[|\bar{X}_t^{m_k} - \hat{X}_t|^2],$$

and thus, by recalling that $1 \geq \gamma = 1 - m_k \geq \frac{1}{2}$, we have for each $t \in [0, T]$,

$$\begin{aligned} &\mathbb{E} \left[\max_{\tau \in [0, t]} \left| \frac{\lambda}{\gamma} \int_0^\tau (1 - e^{-\frac{\gamma}{m_k} (t-s)}) (X_s^\alpha(\rho^{m_k}) - \bar{X}_s^{m_k}) ds - \lambda \int_0^\tau (X_s^\alpha(\rho) - \hat{X}_s) ds \right|^2 \right] \\ &\leq 2t \mathbb{E} \left[\frac{\lambda}{1 - m_k} \int_0^t \left| (1 - e^{-\frac{1 - m_k}{m_k} (t-s)}) (X_s^\alpha(\rho^{m_k}) - X_s^\alpha(\rho) + \hat{X}_s - \bar{X}_s^{m_k}) \right|^2 ds \right] \end{aligned}$$

$$\begin{aligned}
& + 2t\mathbb{E} \left[\lambda \int_0^t \left| \left(\frac{1 - e^{-\frac{1-m_k}{m_k}(t-s)}}}{1-m_k} - 1 \right) (X_s^\alpha(\rho) - \widehat{X}_s) \right|^2 ds \right] \\
& \leq C\mathbb{E} \left[\int_0^t |\widehat{X}_s - \overline{X}_s^{m_k}|^2 ds \right] + C\lambda^2 \int_0^t \left| \frac{1 - e^{-\frac{1-m_k}{m_k}(t-s)}}{1-m_k} - 1 \right|^2 ds \cdot \mathbb{E} \left[\int_0^T |X_s^\alpha(\rho) - \widehat{X}_s|^2 ds \right] \\
& \leq C\mathbb{E} \left[\int_0^t |\widehat{X}_s - \overline{X}_s^{m_k}|^2 ds \right] + C \int_0^t \left| \frac{1 - e^{-\frac{1-m_k}{m_k}(t-s)} - (1-m_k)}{1-m_k} \right|^2 ds \\
& \leq C\mathbb{E} \left[\int_0^t |\widehat{X}_s - \overline{X}_s^{m_k}|^2 ds \right] + C \int_0^t \left(|m_k|^2 + e^{-\frac{2(1-m_k)}{m_k}(t-s)} \right) ds \\
& \leq C\mathbb{E} \left[\int_0^t |\widehat{X}_s - \overline{X}_s^{m_k}|^2 ds \right] + C \left(t|m_k|^2 + \frac{m_k}{2(1-m_k)} \right), \tag{2.44}
\end{aligned}$$

where the constants C s are independent of (k, t) .

For the stochastic integrals, using Doob's inequality gives

$$\begin{aligned}
& \mathbb{E} \left[\max_{\tau \in [0, t]} \left| \frac{\sigma}{\gamma} \int_0^\tau D(X_s^\alpha(\rho^{m_k}) - \overline{X}_s^{m_k}) dB_s - \sigma \int_0^\tau D(X_s^\alpha(\rho) - \widehat{X}_s) dB_s \right|^2 \right] \\
& \leq C\mathbb{E} \left[\int_0^t \left| \frac{1}{\gamma} (X_s^\alpha(\rho^{m_k}) - \overline{X}_s^{m_k}) - (X_s^\alpha(\rho) - \widehat{X}_s) \right|^2 ds \right] \\
& \leq C\mathbb{E} \left[\int_0^t \frac{1}{\gamma^2} \left| (X_s^\alpha(\rho^{m_k}) - \overline{X}_s^{m_k}) - (X_s^\alpha(\rho) - \widehat{X}_s) \right|^2 ds \right] + C\mathbb{E} \left[\int_0^t \left| \left(\frac{1}{\gamma} - 1 \right) (X_s^\alpha(\rho) - \widehat{X}_s) \right|^2 ds \right] \\
& \leq C\mathbb{E} \left[\int_0^t |\overline{X}_s^{m_k} - \widehat{X}_s|^2 ds \right] + C|m_k|^2, \quad \forall t \in [0, T]. \tag{2.45}
\end{aligned}$$

Meanwhile, putting

$$Z_t^k := \frac{\sigma}{\gamma} \int_0^t e^{-\frac{\gamma}{m_k}(t-s)} D(X_s^\alpha(\rho^{m_k}) - \overline{X}_s^{m_k}) dB_s, \quad t \in [0, T],$$

we use Itô's isometry and the relations (2.24) and (2.40) to obtain

$$\begin{aligned}
\max_{t \in [0, T]} \mathbb{E} [|Z_t^k|^2] & = \max_{t \in [0, T]} \mathbb{E} \left[\frac{\sigma^2}{\gamma^2} \int_0^t e^{-\frac{2\gamma}{m_k}(t-s)} |X_s^\alpha(\rho^{m_k}) - \overline{X}_s^{m_k}|^2 ds \right] \\
& \leq \max_{t \in [0, T]} \frac{\sigma^2}{\gamma^2} \left(\sup_{s \in [0, T]} \mathbb{E} \left[|X_s^\alpha(\rho^{m_k}) - \overline{X}_s^{m_k}|^4 \right] \right)^{1/2} \int_0^t e^{-\frac{2\gamma}{m_k}(t-s)} ds \\
& \leq C m_k, \tag{2.46}
\end{aligned}$$

where C is independent of (m_k, t) . Then the process

$$\xi_t := \int_0^t (Z_s^k)' D(X_s^\alpha(\rho^{m_k}) - \overline{X}_s^{m_k}) dB_s, \quad t \in [0, T] \tag{2.47}$$

with $(Z_s^k)'$ being the transpose of Z_s^k , defines a continuous martingale with

$$\begin{aligned}
\mathbb{E} \left[\max_{s \in [0, T]} |\xi_s| \right] & \leq C\mathbb{E} \left[\left(\int_0^T |Z_s^k|^2 |X_s^\alpha(\rho^{m_k}) - \overline{X}_s^{m_k}|^2 ds \right)^{1/2} \right] \\
& \leq C \left(\mathbb{E} \left[\int_0^T |Z_s^k|^2 ds \right] \right)^{1/2} \cdot \left(\mathbb{E} \left[\sup_{s \in [0, T]} |X_s^\alpha(\rho^{m_k}) - \overline{X}_s^{m_k}|^4 \right] \right)^{1/4} \\
\text{(by (2.24), (2.40), and (2.46))} & \leq C |m_k|^{1/2}. \tag{2.48}
\end{aligned}$$

Using the Itô-Doebelin formula as in (2.31) gives that

$$|Z_t^k|^2 = \int_0^t e^{-\frac{2\gamma(t-s)}{m_k}} \left| X_s^\alpha(\rho^{m_k}) - \bar{X}_s^{m_k} \right|^2 ds + 2 \int_0^t e^{-\frac{2\gamma(t-s)}{m_k}} d\xi_s \quad (2.49)$$

$$\begin{aligned} &\leq C \int_0^t e^{-\frac{2\gamma(t-s)}{m_k}} ds \cdot \sup_{s \in [0,t]} \left| X_s^\alpha(\rho^{m_k}) - \bar{X}_s^{m_k} \right|^2 + 2\xi_t - \frac{4\gamma}{m_k} \int_0^t e^{-\frac{2\gamma(t-s)}{m_k}} \xi_s ds \\ &\leq C m_k \sup_{s \in [0,t]} \left| X_s^\alpha(\rho^{m_k}) - \bar{X}_s^{m_k} \right|^2 + 4 \max_{s \in [0,t]} |\xi_s|, \quad \text{a.s.} \end{aligned} \quad (2.50)$$

Putting (2.24), (2.40), (2.48), and (2.50) together, we obtain that

$$\mathbb{E} \left[\max_{\tau \in [0,t]} |Z_\tau^k|^2 \right] \leq C \sqrt{m_k}. \quad (2.51)$$

In addition, it is obvious that

$$\left| \frac{m_k}{\gamma} (1 - e^{-\frac{\gamma}{m_k} t}) \bar{V}_0 \right| \leq C m_k |\bar{V}_0|. \quad (2.52)$$

As the constants C s are independent of (k, t) , combining the estimates (2.44), (2.45), (2.51) and (2.52), letting k tend to infinity on both sides of (2.39) and recalling $\frac{1}{2} \geq m_k \rightarrow 0^+$ and the relation (2.42), we have

$$\hat{X}_t = \bar{X}_0 + \lambda \int_0^t (X_s^\alpha(\rho) - \hat{X}_s) ds + \sigma \int_0^t D(X_s^\alpha(\rho) - \hat{X}_s) dB_s.$$

Therefore, the limit \hat{X} turns out to be a (strong) solution to SDE (2.35). Meanwhile, in view of the continuity of $X^\alpha(\rho)$ in Lemma 2.2, we can easily verify the uniqueness of the strong solution with standard arguments as in [13, Theorem 3.1]. Thus, we must have $\hat{X} = \bar{X}$ that is the unique strong solution to SDE (2.35) with $\mathbb{E} [\max_{t \in [0,T]} |\bar{X}_t|^4] \leq C_2$. Further, due to the arbitrariness of the subsequence $\{\bar{X}^{m_k}\}_{k \in \mathbb{N}}$, we conclude that as $m \rightarrow 0^+$, the sequence of stochastic processes $\{\bar{X}^m\}_{0 < m \leq 1}$ converge weakly to the unique solution \bar{X} to SDE (2.35).

Finally, to measure the distance between \bar{X}^m and the limit $\hat{X} = \bar{X}$, we may replace \hat{X} with \bar{X} in the calculations (2.44), (2.45), (2.51), and (2.52) and arrive at

$$\mathbb{E} \left[\max_{\tau \in [0,t]} |\bar{X}_\tau^m - \bar{X}_\tau|^2 \right] \leq C \int_0^t \mathbb{E}[|\bar{X}_s^m - \bar{X}_s|^2] ds + C \sqrt{m}, \quad t \in [0, T],$$

where the constant C is independent of (m, t) ; if we adopt the estimate (2.46) instead of (2.51) in the above, it holds that

$$\mathbb{E} \left[|\bar{X}_t^m - \bar{X}_t|^2 \right] \leq C \int_0^t \mathbb{E}[|\bar{X}_s^m - \bar{X}_s|^2] ds + C m, \quad t \in [0, T],$$

with C independent of (m, t) . By Gronwall's inequality, we may have

$$\mathbb{E} \left[\max_{t \in [0,T]} |\bar{X}_t^m - \bar{X}_t|^2 \right] \leq C \sqrt{m}, \quad \max_{t \in [0,T]} \mathbb{E} \left[|\bar{X}_t^m - \bar{X}_t|^2 \right] \leq C m, \quad (2.53)$$

where the constants C s depend on $\mathbb{E}[|\bar{X}_0|^4], \mathbb{E}[|\bar{V}_0|^4], C_{\alpha,\varepsilon}, \lambda, \sigma, d$, and T . \square

Remark 2.2. When proving the convergence of $\{\bar{X}^m\}$ satisfying SDEs (1.7) to the solution \bar{X} of (1.10), we cannot expect the convergence of the associated velocity processes $\{\bar{V}^m\}$ due to the indifferenciability of the limit $\{\bar{X}_t\}_{t \geq 0}$ with respect to time t if $\sigma \neq 0$. Therefore, we do not investigate the convergence of the joint Markovian process $\{(\bar{X}^m, \bar{V}^m)\}$ and consider instead solely the process $\{\bar{X}^m\}$ which satisfies a stochastic equation (2.39) of Volterra type, being path-dependent and thus non-Markovian. This non-Markovianity prevents us from using the usual techniques for weak convergence with martingale problems but prompts us to identify the limit by measuring directly the distance between \bar{X}^{m_k} and $\hat{X} = \bar{X}$ in the above proof.

Regarding to the rate of convergence in (2.53), the L^2 -estimate of the uniform norm of the paths of $\overline{X}^m - \overline{X}$ is on the order of \sqrt{m} while the L^2 -estimate for each time $t \in [0, T]$ has order of m . Such a difference stems from the relation (2.49) which also reads

$$|Z_t^k|^2 = \int_0^t e^{-\frac{2\gamma(t-s)}{m_k}} \left| X_s^\alpha(\rho^{m_k}) - \overline{X}_s^{m_k} \right|^2 ds + 2 \int_0^t e^{-\frac{2\gamma(t-s)}{m_k}} (Z_s^k)' D(X_s^\alpha(\rho^{m_k}) - \overline{X}_s^{m_k}) dB_s.$$

Here the stochastic integral has mean zero and taking expectation on both sides can directly give us the order of m as seen in the first term on the right hand side of (2.50); however, when considering the uniform norm, we cannot ignore the stochastic integral and its *linear* dependence on $(Z_s^k)_{s \in [0, T]}$ leads to the order \sqrt{m} . In addition, in the above proof, because both \overline{X}^{m_k} and \overline{X} are the strong solutions to associated SDEs with pathwise uniqueness and a strong solution is a functional of the initial value and the driving Wiener process (see [34, Chapter IV]), the following convergence

$$\mathbb{P} \left(\lim_{k \rightarrow \infty} \|\overline{X}^{m_k} - \overline{X}\|_0 = 0 \right) = 1, \quad (2.54)$$

actually holds for strong solutions to SDEs (1.7) and (2.35) on any given probability space $(\Omega, \mathcal{F}, \mathbb{P})$ equipped with augmented filtration generated by a d -dimensional Wiener process B . Indeed, some experiments conducted at the end of Section 4 (see Figures 9, 10, and 11) numerically support the convergences established here.

3. GENERALIZATION TO THE CASE WITH MEMORY EFFECTS

In [27], the authors considered a PSO model which involves the memory of the local best positions, and it is of the form

$$d\overline{X}_t^m = \overline{V}_t^m dt, \quad (3.1)$$

$$d\overline{Y}_t^m = \nu (\overline{X}_t^m - \overline{Y}_t^m) S^\beta (\overline{X}_t^m, \overline{Y}_t^m) dt, \quad (3.2)$$

$$\begin{aligned} d\overline{V}_t^m &= -\frac{\gamma}{m} \overline{V}_t^m dt + \frac{\lambda_1}{m} (\overline{Y}_t^m - \overline{X}_t^m) dt + \frac{\lambda_2}{m} (Y_t^\alpha(\overline{\rho}^m) - \overline{X}_t^m) dt \\ &\quad + \frac{\sigma_1}{m} D(\overline{Y}_t^m - \overline{X}_t^m) dB_t^1 + \frac{\sigma_2}{m} D(Y_t^\alpha(\overline{\rho}^m) - \overline{X}_t^m) dB_t^2, \end{aligned} \quad (3.3)$$

where B^1 and B^2 are two mutually independent d -dimensional Wiener processes, and similarly to the previous section, we introduce the following regularization of the global best position

$$Y_t^\alpha(\overline{\rho}^m) = \frac{\int_{\mathbb{R}^d} y \omega_\alpha(y) \overline{\rho}^m(t, dy)}{\int_{\mathbb{R}^d} \omega_\alpha(y) \overline{\rho}^m(t, dy)}, \quad \overline{\rho}^m(t, y) = \iint_{\mathbb{R}^d \times \mathbb{R}^d} f^m(t, dx, y, dv). \quad (3.4)$$

Here the equation (3.2) of \overline{Y}^m is the time continuous approximation to the evolution of the local best position, and S^β with $\beta \gg 1$ is the hyperbolic tangent $S^\beta(x, y) = 1 + \tanh(\beta(\mathcal{E}(y) - \mathcal{E}(x)))$. The corresponding mean-field PSO equation is

$$\begin{aligned} \partial_t f_t^m + v \cdot \nabla_x f_t^m + \nabla_y \cdot (\nu(x - y) S^\beta(x, y) f_t^m) &= \nabla_v \cdot \left(\frac{\gamma}{m} v f_t^m + \frac{\lambda_1}{m} (x - y) f_t^m + \frac{\lambda_2}{m} (x - Y_t^\alpha(\overline{\rho}^m)) f_t^m \right. \\ &\quad \left. + \left(\frac{\sigma_2^2}{2m^2} D(x - Y_t^\alpha(\overline{\rho}^m))^2 + \frac{\sigma_1^2}{2m^2} D(x - y)^2 \right) \nabla_v f_t^m \right). \end{aligned} \quad (3.5)$$

We want to prove that the zero-inertia limit ($m \rightarrow 0$) leads to the following mean-field CBO dynamic

$$\begin{cases} \overline{X}_t = \overline{X}_0 + \lambda_1 \int_0^t (\overline{Y}_s - \overline{X}_s) ds + \sigma_1 \int_0^t D(\overline{Y}_s - \overline{X}_s) dB_s^1 + \lambda_2 \int_0^t (Y_s^\alpha(\overline{\rho}) - \overline{X}_s) ds + \sigma_2 \int_0^t D(Y_s^\alpha(\overline{\rho}) - \overline{X}_s) dB_s^2, \\ \overline{Y}_t = \overline{Y}_0 + \nu \int_0^t (\overline{X}_s - \overline{Y}_s) S^\beta(\overline{X}_s, \overline{Y}_s) ds, \end{cases}$$

and its corresponding partial differential equation is

$$\partial_t \rho_t + \nabla_x \cdot (\lambda_1(y - x) + \lambda_2(Y_t^\alpha(\overline{\rho}) - x) \rho_t) + \nabla_y \cdot (\nu(x - y) S^\beta(x, y)) \rho_t$$

$$= \frac{1}{2} \sum_{j=1}^d \frac{\partial^2}{\partial x_j^2} \left(\rho_t (\sigma_1^2 (x-y)_j^2 + \sigma_2^2 (x - Y_t^\alpha(\bar{\rho}))_j^2) \right), \quad (3.6)$$

where $\bar{\rho}(t, y) = \int_{\mathbb{R}^d} \rho(t, dx, y)$.

Since the proof of the zero-inertia limit for the PSO dynamics with memory effects follows similar arguments as developed in the previous section and no essential innovation is needed to be explained, we only sketch the proof for the tightness.

Theorem 3.1 (Tightness). *Let Assumption 1 hold and $(\bar{X}_t^m, \bar{Y}_t^m, \bar{V}_t^m)_{t \in [0, T]}$ satisfy the system (3.1)-(3.3). For each countable subsequence $\{m_k\}_{k \in \mathbb{N}} \subset [0, 1]$ with $\lim_{k \rightarrow \infty} m_k = 0$, the sequence of probability distributions $\{\rho^{m_k}\}_{k \in \mathbb{N}}$ of $\{(\bar{X}^{m_k}, \bar{Y}^{m_k})\}_{k \in \mathbb{N}}$ is tight.*

Proof. The proof is similar to Theorem 2.2.

• *Step 1: Checking (Con1).* For $0 < m \leq \frac{1}{2}$, we first solve (3.3) for \bar{V}^m and obtain

$$\begin{aligned} \bar{V}_t^m &= e^{-\frac{\gamma}{m}t} \bar{V}_0 + \frac{\lambda_1}{m} \int_0^t e^{-\frac{\gamma}{m}(t-s)} (\bar{Y}_s^m - \bar{X}_s^m) ds + \frac{\sigma_1}{m} \int_0^t e^{-\frac{\gamma}{m}(t-s)} D(\bar{Y}_s^m - \bar{X}_s^m) dB_s^1 \\ &\quad + \frac{\lambda_2}{m} \int_0^t e^{-\frac{\gamma}{m}(t-s)} (Y_s^\alpha(\bar{\rho}^m) - \bar{X}_s^m) ds + \frac{\sigma_2}{m} \int_0^t e^{-\frac{\gamma}{m}(t-s)} D(Y_s^\alpha(\bar{\rho}^m) - \bar{X}_s^m) dB_s^2. \end{aligned}$$

Here $\bar{\rho}^m(t, y) = \int_{\mathbb{R}^d} \rho^m(t, dx, y)$. By Fubini's theorem, similar arguments as in (2.6) yield that

$$\begin{aligned} \bar{X}_t^m &= \bar{X}_0 + \frac{m}{\gamma} (1 - e^{-\frac{\gamma}{m}t}) \bar{V}_0 + \frac{\lambda_1}{\gamma} \int_0^t (1 - e^{-\frac{\gamma}{m}(t-s)}) (\bar{Y}_s^m - \bar{X}_s^m) ds + \frac{\sigma_1}{\gamma} \int_0^t (1 - e^{-\frac{\gamma}{m}(t-s)}) D(\bar{Y}_s^m - \bar{X}_s^m) dB_s^1 \\ &\quad + \frac{\lambda_2}{\gamma} \int_0^t (1 - e^{-\frac{\gamma}{m}(t-s)}) (Y_s^\alpha(\bar{\rho}^m) - \bar{X}_s^m) ds + \frac{\sigma_2}{\gamma} \int_0^t (1 - e^{-\frac{\gamma}{m}(t-s)}) D(Y_s^\alpha(\bar{\rho}^m) - \bar{X}_s^m) dB_s^2. \end{aligned} \quad (3.7)$$

Following the same computations as in Theorem 2.2 gives

$$\mathbb{E}[|\bar{X}_t^m|^4] \leq C \mathbb{E}[|\bar{X}_0|^4 + |\bar{V}_0|^4] + C \int_0^t \mathbb{E}[|Y_s^\alpha(\bar{\rho}^m) - \bar{X}_s^m|^4] ds + C \int_0^t \mathbb{E}[|\bar{Y}_s^m - \bar{X}_s^m|^4] ds,$$

where C depends only on $\lambda_1, \sigma_2, \lambda_2, \sigma_2, d$, and T . Put $\tilde{\rho}^m(t, x) = \int_{\mathbb{R}^d} \rho^m(t, x, dy)$. In a similar way to (2.7) we have

$$\begin{aligned} \mathbb{E}[|Y_t^\alpha(\bar{\rho}^m) - \bar{X}_t^m|^4] &= \int_{\mathbb{R}^d} \left| \frac{\int_{\mathbb{R}^d} y \omega_\alpha^\mathcal{E}(y) \bar{\rho}^m(t, dy)}{\int_{\mathbb{R}^d} \omega_\alpha^\mathcal{E}(y) \bar{\rho}^m(t, dy)} - x \right|^4 \tilde{\rho}^m(t, dx) = \int_{\mathbb{R}^d} \left| \frac{\int_{\mathbb{R}^d} (y-x) \omega_\alpha^\mathcal{E}(y) \bar{\rho}^m(t, dy)}{\int_{\mathbb{R}^d} \omega_\alpha^\mathcal{E}(y) \bar{\rho}^m(t, dy)} \right|^4 \tilde{\rho}^m(t, dx) \\ &\leq \frac{\int_{\mathbb{R}^d} \int_{\mathbb{R}^d} |x-y|^4 \omega_\alpha^\mathcal{E}(y) \bar{\rho}^m(t, dy) \tilde{\rho}^m(t, dx)}{\int_{\mathbb{R}^d} \omega_\alpha^\mathcal{E}(y) \bar{\rho}^m(t, dy)} \leq 8C_{\alpha, \mathcal{E}} \mathbb{E}[|\bar{X}_t^m|^4 + |\bar{Y}_t^m|^4]. \end{aligned}$$

Thus it yields that

$$\mathbb{E}[|\bar{X}_t^m|^4] \leq C \mathbb{E}[|\bar{X}_0|^4 + |\bar{V}_0|^4] + C \int_0^t \mathbb{E}[|\bar{Y}_s^m|^4 + |\bar{X}_s^m|^4] ds, \quad (3.8)$$

where C depends only on $\lambda_1, \sigma_2, \lambda_2, \sigma_2, d, T$ and $C_{\alpha, \mathcal{E}}$.

Recall that

$$\bar{Y}_t^m = \bar{Y}_0 + \nu \int_0^t (\bar{X}_s^m - \bar{Y}_s^m) S^\beta(\bar{X}_s^m, \bar{Y}_s^m) ds$$

with $S^\beta(x, y) = 1 + \tanh(\beta(\mathcal{E}(y) - \mathcal{E}(x)))$. Using the fact that $|S^\beta| \leq 2$ then it follows

$$\mathbb{E}[|\bar{Y}_t^m|^4] \leq C \mathbb{E}[|\bar{Y}_0|^4] + C \int_0^t \mathbb{E}[|\bar{Y}_s^m|^4 + |\bar{X}_s^m|^4] ds,$$

where C depends only on ν and T . This together with (3.8) implies

$$\mathbb{E}[|\bar{X}_t^m|^4 + |\bar{Y}_t^m|^4] \leq C \mathbb{E}[|\bar{X}_0|^4 + |\bar{Y}_0|^4 + |\bar{V}_0|^4] + C \int_0^t \mathbb{E}[|\bar{Y}_s^m|^4 + |\bar{X}_s^m|^4] ds.$$

By Gronwall's inequality, it yields that

$$\sup_{t \in [0, T]} \mathbb{E}[|\bar{X}_t^m|^4 + |\bar{Y}_t^m|^4] \leq C(\mathbb{E}[|\bar{X}_0|^4 + |\bar{Y}_0|^4 + |\bar{V}_0|^4], \lambda_1, \sigma_2, \lambda_2, \sigma_2, d, T, C_{\alpha, \varepsilon}, \nu), \quad (3.9)$$

which verifies (Con1) for the case of $0 < m \leq \frac{1}{2}$. We omit the discussions for the case of $\frac{1}{2} < m \leq 1$.

• *Step 2: Checking (Con2).* Let β be a $\sigma(X_s^m; s \in [0, T])$ -stopping time such that $\beta + \delta_0 \leq T$. Set $m_0 = \frac{1}{2}$ w.l.o.g.. Then for all $0 < m \leq m_0$, one has $\frac{1}{2} \leq \gamma < 1$. Similar to (2.18), one has

$$\begin{aligned} & \bar{X}_{\beta+\delta}^m - \bar{X}_\beta^m \\ &= \frac{m}{\gamma} (e^{-\frac{\gamma}{m}\beta} - e^{-\frac{\gamma}{m}(\beta+\delta)}) \bar{V}_0 \\ &+ \frac{\lambda_2}{\gamma} \int_0^\beta (e^{-\frac{\gamma}{m}(\beta-s)} - e^{-\frac{\gamma}{m}(\beta+\delta-s)}) (Y_s^\alpha(\bar{\rho}^m) - \bar{X}_s^m) ds + \frac{\lambda_2}{\gamma} \int_\beta^{\beta+\delta} (1 - e^{-\frac{\gamma}{m}(\beta+\delta-s)}) (Y_s^\alpha(\bar{\rho}^m) - \bar{X}_s^m) ds \\ &+ \frac{\sigma_2}{\gamma} \int_0^\beta (e^{-\frac{\gamma}{m}(\beta-s)} - e^{-\frac{\gamma}{m}(\beta+\delta-s)}) D(Y_s^\alpha(\bar{\rho}^m) - \bar{X}_s^m) dB_s^2 + \frac{\sigma_2}{\gamma} \int_\beta^{\beta+\delta} (1 - e^{-\frac{\gamma}{m}(\beta+\delta-s)}) D(Y_s^\alpha(\bar{\rho}^m) - \bar{X}_s^m) dB_s^2 \\ &+ \frac{\lambda_1}{\gamma} \int_0^\beta (e^{-\frac{\gamma}{m}(\beta-s)} - e^{-\frac{\gamma}{m}(\beta+\delta-s)}) (\bar{Y}_s^m - \bar{X}_s^m) ds + \frac{\lambda_1}{\gamma} \int_\beta^{\beta+\delta} (1 - e^{-\frac{\gamma}{m}(\beta+\delta-s)}) (\bar{Y}_s^m - \bar{X}_s^m) ds \\ &+ \frac{\sigma_1}{\gamma} \int_0^\beta (e^{-\frac{\gamma}{m}(\beta-s)} - e^{-\frac{\gamma}{m}(\beta+\delta-s)}) D(\bar{Y}_s^m - \bar{X}_s^m) dB_s^1 + \frac{\sigma_1}{\gamma} \int_\beta^{\beta+\delta} (1 - e^{-\frac{\gamma}{m}(\beta+\delta-s)}) D(\bar{Y}_s^m - \bar{X}_s^m) dB_s^1. \end{aligned} \quad (3.10)$$

Using the estimate (3.9) it follows from the same computations as in Theorem 2.2 that

$$\mathbb{E}[|\bar{X}_{\beta+\delta}^m - \bar{X}_\beta^m|^2] \leq C(\delta^{\frac{1}{4}} + \delta^2), \quad (3.11)$$

where C depends only on $\mathbb{E}[|\bar{X}_0|^4 + |\bar{Y}_0|^4 + |\bar{V}_0|^4], \lambda_1, \sigma_2, \lambda_2, \sigma_2, T, C_{\alpha, \varepsilon}$, and ν .

Having a look at

$$\bar{Y}_{\beta+\delta}^m - \bar{Y}_\beta^m = \nu \int_\beta^{\beta+\delta} (\bar{X}_s^m - \bar{Y}_s^m) S^\beta(\bar{X}_s^m, \bar{Y}_s^m) ds,$$

we have

$$|\bar{Y}_{\beta+\delta}^m - \bar{Y}_\beta^m|^2 \leq \nu^2 \delta \int_0^T |\bar{X}_s^m - \bar{Y}_s^m|^2 dt.$$

By estimate (3.9), we have

$$\mathbb{E}[|\bar{Y}_{\beta+\delta}^m - \bar{Y}_\beta^m|^2] \leq \nu^2 \delta \int_0^T \mathbb{E}[|\bar{X}_s^m - \bar{Y}_s^m|^4]^{\frac{1}{2}} dt \leq C\delta, \quad (3.12)$$

where C depends on $\mathbb{E}[|\bar{X}_0|^4 + |\bar{Y}_0|^4 + |\bar{V}_0|^4], \lambda_1, \sigma_2, \lambda_2, \sigma_2, T, C_{\alpha, \varepsilon}$ and ν . This together with (3.11) justifies (Con2). \square

Let us recall

$$\begin{aligned} \bar{X}_t^m &= \bar{X}_0 + \frac{m}{\gamma} (1 - e^{-\frac{\gamma}{m}t}) \bar{V}_0 + \frac{\lambda_1}{\gamma} \int_0^t (1 - e^{-\frac{\gamma}{m}(t-s)}) (\bar{Y}_s^m - \bar{X}_s^m) ds + \frac{\sigma_1}{\gamma} \int_0^t (1 - e^{-\frac{\gamma}{m}(t-s)}) D(\bar{Y}_s^m - \bar{X}_s^m) dB_s^1 \\ &+ \frac{\lambda_2}{\gamma} \int_0^t (1 - e^{-\frac{\gamma}{m}(t-s)}) (Y_s^\alpha(\bar{\rho}^m) - \bar{X}_s^m) ds + \frac{\sigma_2}{\gamma} \int_0^t (1 - e^{-\frac{\gamma}{m}(t-s)}) D(Y_s^\alpha(\bar{\rho}^m) - \bar{X}_s^m) dB_s^2 \end{aligned} \quad (3.13)$$

and

$$\bar{Y}_t^m = \bar{Y}_0 + \nu \int_0^t (\bar{X}_s^m - \bar{Y}_s^m) S^\beta(\bar{X}_s^m, \bar{Y}_s^m) ds. \quad (3.14)$$

Then following the lines of the proof in Theorem 2.4, one can easily obtain the following theorem.

Theorem 3.2 (Zero-inertia limit). *Let Assumption 1 hold and $(\bar{X}_t^m, \bar{Y}_t^m)_{t \in [0, T]}$ satisfy the system (3.13)–(3.14). Then as $m \rightarrow 0^+$, the sequence of stochastic processes $\{(\bar{X}^m, \bar{Y}^m)\}_{0 < m \leq 1}$ converge weakly to (\bar{X}, \bar{Y}) which is the unique solution to the following coupled SDE:*

$$\begin{aligned} \bar{X}_t &= \bar{X}_0 + \lambda_1 \int_0^t (\bar{Y}_s - \bar{X}_s) ds + \sigma_1 \int_0^t D(\bar{Y}_s - \bar{X}_s) dB_s^1 + \lambda_2 \int_0^t (Y_s^\alpha(\bar{\rho}) - \bar{X}_s) ds + \sigma_2 \int_0^t D(Y_s^\alpha(\bar{\rho}) - \bar{X}_s) dB_s^2, \\ \bar{Y}_t &= \bar{Y}_0 + \nu \int_0^t (\bar{X}_s - \bar{Y}_s) S^\beta(\bar{X}_s, \bar{Y}_s) ds. \end{aligned}$$

Moreover, we have the following convergences

$$\mathbb{E} \left[\max_{t \in [0, T]} \left(|\bar{X}_t^m - \bar{X}_t|^2 + |\bar{Y}_t^m - \bar{Y}_t|^2 \right) \right] \leq C \sqrt{m}, \quad \max_{t \in [0, T]} \mathbb{E} \left[|\bar{X}_t^m - \bar{X}_t|^2 + |\bar{Y}_t^m - \bar{Y}_t|^2 \right] \leq C m. \quad (3.15)$$

where the constants C s depend on $\mathbb{E}[|\bar{X}_0|^4 + |\bar{Y}_0|^4 + |\bar{V}_0|^4]$, $\lambda_1, \sigma_2, \lambda_2, \sigma_2, \beta, T, C_{\alpha, \varepsilon}, d$, and ν .

4. NUMERICAL EXAMPLES ON THE ZERO-INERTIA LIMIT

We conclude this paper with a few instructive numerical experiments on validating the zero-inertia limit. We will focus on the mono-dimensional case since it allows us to see more clearly how the distribution of particles evolves in time depending on the inertia parameter m , and hence show the zero-inertia limit. Different benchmark functions have been used and tested, but we will report here the case of the Ackley function shown in Figure 3. Following the same structure of the paper, we will first analyze the case without memory effect and then we will generalize as in Section 3 to the case with memory. Extensive discussions on other numerical implementations and experiments are presented in [27].

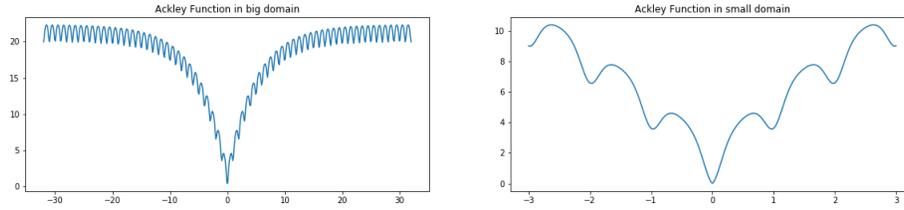


FIGURE 3. Ackley function in a big (left) and small (right) domain with its many local minima.

4.1. Small inertia limit without memory. Given the system of stochastic differential equations in (1.2), the particle system can be solved by using a semi-implicit discretization scheme

$$\begin{cases} X_{n+1}^{i,m} = X_n^{i,m} + \Delta t V_{n+1}^{i,m}, \\ V_{n+1}^{i,m} = \frac{m}{m+\gamma\Delta t} V_n^{i,m} + \frac{\lambda\Delta t}{m+\gamma\Delta t} (X_n^{\alpha,m} - X_n^{i,m}) + \frac{\sigma\sqrt{\Delta t}}{m+\gamma\Delta t} D(X_n^{\alpha,m} - X_n^{i,m}) \theta_n^i, \quad i = 1, \dots, N, \end{cases} \quad (4.1)$$

where $X_n^{i,m}$ and $V_n^{i,m}$ are, respectively, the position and velocity of the i -th particle at the discrete time $n\Delta t$ with Δt being the time discretization, and the diagonal matrix $D(X_n^{\alpha,m} - X_n^{i,m})$ simply coincides with $X_n^{\alpha,m} - X_n^{i,m}$ as we are considering the mono-dimensional case. Moreover, $X_n^{\alpha,m}$ is defined as in (1.3) and $\theta_n^i \sim \mathcal{N}(0, 1) \forall i, n$. We compare this particle system with the CBO dynamic of the form (1.10), which can be solved using the Euler-Maruyama scheme

$$X_{n+1}^i = X_n^i + \Delta t \lambda (X_n^\alpha - X_n^i) + \sqrt{\Delta t} \sigma (X_n^\alpha - X_n^i) \theta_n^i. \quad (4.2)$$

As already mentioned, we consider the minimization of the Ackley function with minimum at $x = 0$ and, starting from the same initial distribution of particles, we solve the PSO system (4.1) for different inertia

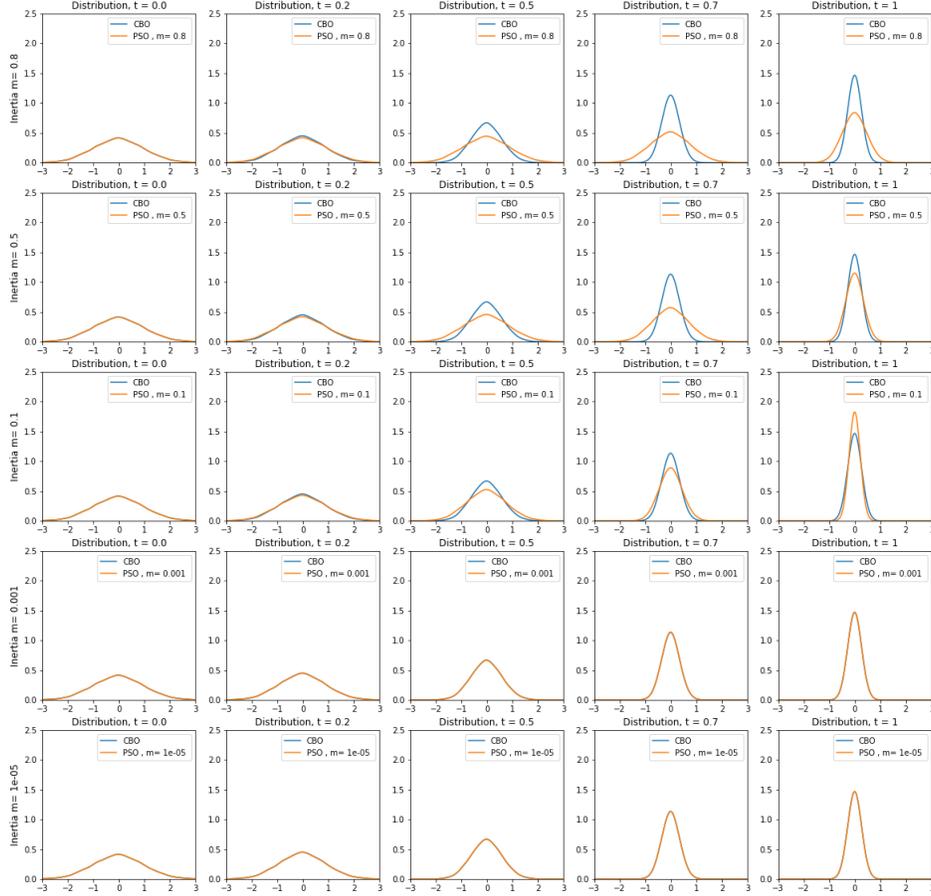


FIGURE 4. Comparison of the CBO (4.2) and PSO (4.1) dynamics for different inertia values (which are changing over the rows) and at many time steps (changing over columns), starting from a normal distribution.

values. Then, we compare the evolution of the distribution of particles with the one of the particles moving according to the CBO system (4.2). In order to be able to compare the results, we fix the parameters $\lambda = 1$, $\sigma = \frac{1}{\sqrt{3}}$ and $\alpha = 30$, while θ_n^i are sampled from $\mathcal{N}(0, 1)$ and fixed for each $i = 1, \dots, N$ and $n \in [0, T/\Delta t]$. Moreover, T is set to 1 and the time discretization is $\Delta t = 0.01$, with a total number of particles $N = 10^4$.

Figure 4 shows in each row the evolution of the CBO distribution and the one of the PSO system with m fixed that is decreasing over rows. The initial particles are always sampled from the same distribution, which is in this case a Gaussian centered in 0 with variance 1. Clearly, the PSO system with $m = 0.8$ leads to the correct minimum in 0 at the final time step $t = 1$, but the distribution of the particles is different from the one of the CBO. While, for any $t \in [0, T]$, if the inertia value is decreased to 0.1, or even to 0.001, the two distributions, namely the one obtained via CBO and the PSO one, are indistinguishable, as the last

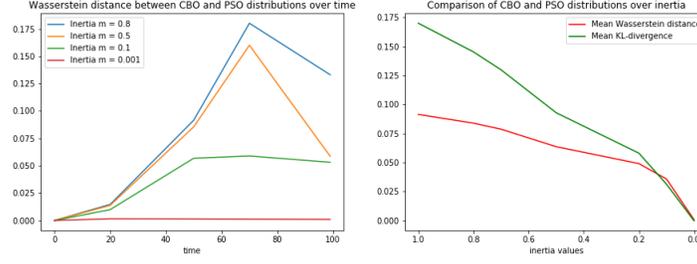


FIGURE 5. Left: Wasserstein distance over time ; Right: Wasserstein distance and Kullback-Leibler divergence (in mean) over the inertia values.

two rows of Figure 4 show.

These considerations are confirmed in Figure 5 where we compare the distributions obtained in Figure 4 using the Wasserstein 2 distance between the CBO distribution and the PSO distribution. On the left of Figure 5, the Wasserstein distance is plotted for each time step. Moreover, since we want to show the influence of the inertia parameter, we take the mean value of the Wasserstein distance over all time steps and plot it as a function of the inertia values. This is shown on the right of Figure 5 where we also add the mean value of the Kullback-Leibler divergence since the latter is a well-known measure used to compare distributions, especially in statistics. Moreover, since it is necessary to start with an initial distribution that is close to the global minimizer, we also try to see what happens when the initial distribution is a uniform distribution between -3 and 3 and compare the evolution of its particles according to the CBO and PSO dynamics, with varying inertia parameters. The result is shown in Figure 6. In this case, the difference between the CBO distribution and the one of the PSO dynamics is way higher in the case of big inertia value, but, as before, goes to zero as soon as m converges to 0.

4.2. Small inertia limit with memory effect. The PSO model which involves the memory of the local and global best positions, underlying (3.1)–(3.3), can similarly be solved via

$$\begin{cases} X_{n+1}^{i,m} = X_n^{i,m} + \Delta t V_{n+1}^{i,m}, & i = 1, \dots, N \\ Y_{n+1}^{i,m} = Y_n^{i,m} + \nu \Delta t (X_{n+1}^{i,m} - Y_n^{i,m}) S^\beta(X_{n+1}^{i,m}, Y_n^{i,m}), \\ V_{n+1}^{i,m} = \frac{m}{m+\gamma\Delta t} V_n^{i,m} + \frac{\lambda_1 \Delta t}{m+\gamma\Delta t} (Y_n^{i,m} - X_n^{i,m}) + \frac{\lambda_2 \Delta t}{m+\gamma\Delta t} (Y_n^{\alpha,m} - X_n^{i,m}) \\ \quad + \frac{\sigma_1 \sqrt{\Delta t}}{m+\gamma\Delta t} D(Y_n^{i,m} - X_n^{i,m}) \theta_n^{1,i} + \frac{\sigma_2 \sqrt{\Delta t}}{m+\gamma\Delta t} D(Y_n^{\alpha,m} - X_n^{i,m}) \theta_n^{2,i}, \end{cases} \quad (4.3)$$

where $Y_n^{i,m}$ is the local best that the i -th particle has memory of, and $Y_n^{\alpha,m}$ is the regularized global best, defined as in (3.4). Clearly, the corresponding CBO dynamics is the following

$$\begin{cases} X_n^i = X_n^i + \lambda_1 \Delta t (Y_n^i - X_n^i) + \lambda_2 \Delta t (Y_n^\alpha - X_n^i) + \sigma_1 \sqrt{\Delta t} D(Y_n^i - X_n^i) \theta_n^{1,i} + \sigma_2 \sqrt{\Delta t} (Y_n^\alpha - X_n^i) \theta_n^{2,i} \\ Y_n^i = Y_n^i + \nu \Delta t (X_n^i - Y_n^i) S^\beta(X_n^i, Y_n^i) \end{cases} \quad (4.4)$$

Once again, since we want to show the convergence of the PSO distribution obtained from (4.3) with a small inertia value to the one attained via the CBO system (4.4), we need to set some of the parameters to the same values in order to be able to compare the results. Their values are the following

$$\lambda_1 = \lambda_2 = 1 \quad \sigma_1 = \sigma_2 = \frac{1}{\sqrt{3}} \quad \alpha = 30 \quad \beta = 30 \quad \nu = \frac{1}{2} \quad (4.5)$$

and, as before, the effect of the Brownian motion leads to $\theta_n^{1,i}, \theta_n^{2,i}$ which are sampled from a normal distribution and set to a fixed value $\forall i = 1, \dots, N$ and $\forall n \in [0, T/\Delta t]$. The time discretization and the number of

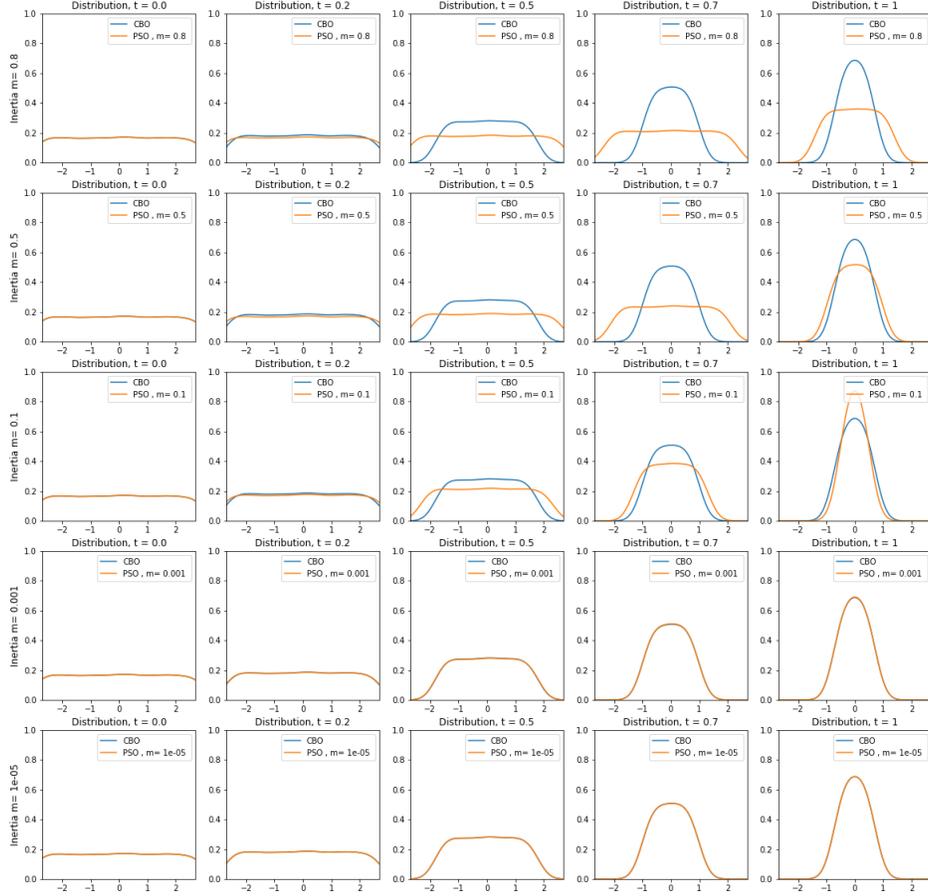


FIGURE 6. Evolution of an initial uniform distribution according to CBO (4.2) and PSO (4.1) dynamics and their comparison for different time steps (on the columns) and different inertia values (on the rows).

particles are set to the same values as in the case without memory, namely $T = 1$, $\Delta t = 0.01$, and $N = 10^4$. The difference with the previous case is that now the distribution of which we want to show convergence, is actually a function of both the particles' position and their local best and, as such, it is bi-dimensional. The result of the evolution of the particles according to the CBO dynamics (4.4) is shown in the first row of Figure 7, while the second row represents the evolution according to the PSO dynamics (4.3) with a big inertia parameter, e.g. $m = 0.8$. This is compared with the evolution presented in the third row in which the inertia is set to a very low value, e.g. $m = 0.001$.

In the case without memory, it is easy to see the convergence at each time step of the PSO system with small inertia to the CBO, but it's not as clear now that the distribution we are interested in is bi-dimensional. To be able to compare the evolution and to check how similar the distributions are at every time step, we

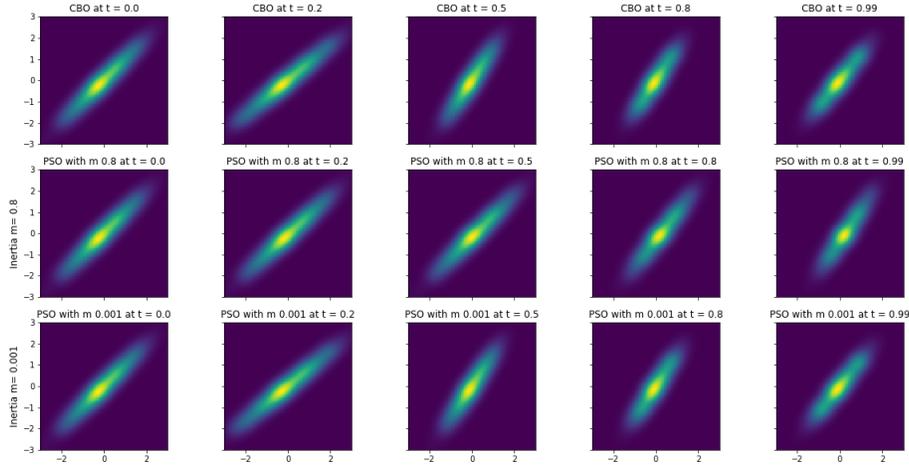


FIGURE 7. First Row: evolution in time of the initial gaussian distribution according to CBO dynamics (4.4); Second Row: evolution in time of the initial Gaussian distribution according to PSO dynamics (4.3) with $m = 0.8$; Third Row: evolution produced by the PSO dynamics (4.3) with $m = 0.001$.

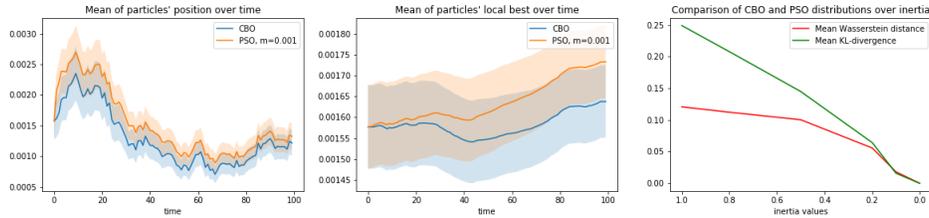


FIGURE 8. Left: mean particles' position with their standard deviation; Center: mean particles' local best with the standard deviation too; Right: plot of the Wasserstein distance and Kullback–Leibler divergence between the distribution obtained via CBO dynamics (4.4) and the one obtained through PSO system (4.3) with small inertia value, i.e. 0.001.

show in Figure 8 different plots: we look at the particles' mean positions (left) and their local best (center) for each time step and, in both cases, show their standard deviation as a colored area around the mean. It is interesting to see how the particles are moving and where they are attracted to, especially because the pattern of the PSO dynamics with small inertia is the same as the one of CBO. Finally, on the right plot of Figure 8, we show how both our similarity measures, namely the Wasserstein distance and the Kullback–Leibler divergence, are decreasing along with the inertia parameter, validating the small inertia limit also in the general case with memory.

Finally, we want to numerically investigate the convergence in path space given in Theorem 2.4 and Remark 2.2. Figure 9 shows on the left the trajectory of a random particle which is moving according to the CBO system (4.2) and the PSO system (4.1) with different inertia values. Using all the $N = 10^4$ particles,

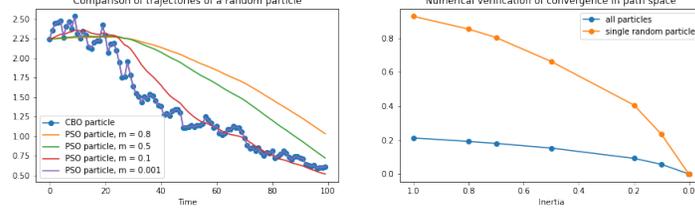


FIGURE 9. Left: given a random particle, its trajectory according to CBO (4.2) is compared with the ones obtained via PSO (4.2) with different inertia values ; Right: verification of limit (4.6), i.e. plot of $\frac{1}{N} \sum_{i=1}^N \max_{n \in [0, T/\Delta t]} |X_n^{i,m} - X_n^i|^2$ for every decreasing value of m (in blue) and of the quantity $\max_{n \in [0, T/\Delta t]} |X_n^{i_0,m} - X_n^{i_0}|^2$ for the same random particles considered on the left (orange color).

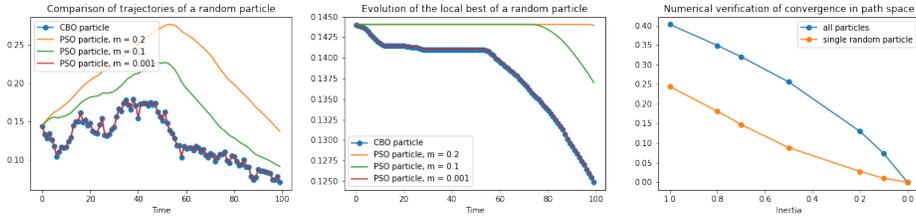


FIGURE 10. Left: random particle's trajectories according to CBO (4.4) and PSO (4.3) with decreasing inertia values; Center: random particle's memory of the local best according to CBO (4.4) and PSO (4.3) with decreasing inertia values; Right: verification of limit of $\frac{1}{N} \sum_{i=1}^N \max_{n \in [0, T/\Delta t]} (|X_n^{i,m} - X_n^i|^2 + |Y_n^{i,m} - Y_n^i|^2)$ for every decreasing value of m (in blue) and of the quantity $\max_{n \in [0, T/\Delta t]} (|X_n^{i_0,m} - X_n^{i_0}|^2 + |Y_n^{i_0,m} - Y_n^{i_0}|^2)$ for the same random particles considered on the left (orange color).

we take

$$\lim_{k \rightarrow \infty} \frac{1}{N} \sum_{i=1}^N \max_{n \in [0, T/\Delta t]} |X_n^{i,m} - X_n^i|^2, \quad (4.6)$$

namely the squared L^2 -norm of the vector containing the sup-norm in time of the distances between the CBO and PSO solutions. On the right of Figure 9, the numerical verification of formula (2.54) is presented: for every inertia value m , the quantity $\max_{n \in [0, T/\Delta t]} |X_n^{i_0,m} - X_n^{i_0}|^2$ is plotted, where i_0 indicates the same random particles whose trajectories are plotted on the left of Figure 9. Moreover, the quantity in (4.6) is also plotted for every inertia value m and the convergence to the limit 0 is clear.

Both the plots in Figure 9 represent the particles moving according to the discrete systems (4.1), (4.2) in the case without memory effect. In the more general case with memory, the systems (4.1), (4.2) become the ones in (4.3) and (4.4), but the convergence in path space still holds and it's numerically shown in Figure 10. Here, not only the trajectories of a random particle are plotted (on the left), but also the evolution of the local bests according to both the discretized systems and with different inertia values are shown in the central image of Figure 10. Once again, even in the case with memory effect, the strong convergence in path space is shown for the random particle previously considered and also for all of the particles in terms of the quantity $\frac{1}{N} \sum_{i=1}^N \max_{n \in [0, T/\Delta t]} (|X_n^{i,m} - X_n^i|^2 + |Y_n^{i,m} - Y_n^i|^2)$, on the right of Figure 10.

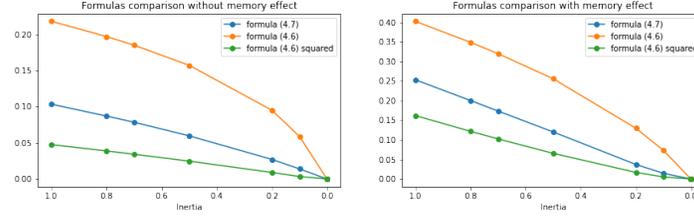


FIGURE 11. Left: case without memory, comparison of $\frac{1}{N} \sum_{i=1}^N \max_{n \in [0, T/\Delta t]} |X_n^{i,m} - X_n^i|^2$ (orange) and $\max_{n \in [0, T/\Delta t]} \frac{1}{N} \sum_{i=1}^N |X_n^{i,m} - X_n^i|^2$ (blue) for every inertia value, where the curve $\left(\frac{1}{N} \sum_{i=1}^N \max_{n \in [0, T/\Delta t]} |X_n^{i,m} - X_n^i|^2\right)^2$ (green) is also plotted as a function of m ; Right: case with memory, comparison of quantity $\frac{1}{N} \sum_{i=1}^N \max_{n \in [0, T/\Delta t]} \left(|X_n^{i,m} - X_n^i|^2 + |Y_n^{i,m} - Y_n^i|^2\right)$ (orange) with $\max_{n \in [0, T/\Delta t]} \frac{1}{N} \sum_{i=1}^N \left(|X_n^{i,m} - X_n^i|^2 + |Y_n^{i,m} - Y_n^i|^2\right)$ (blue) for every decreasing value of m , where the square $\left(\frac{1}{N} \sum_{i=1}^N \max_{n \in [0, T/\Delta t]} \left(|X_n^{i,m} - X_n^i|^2 + |Y_n^{i,m} - Y_n^i|^2\right)\right)^2$ (green) is also plotted.

Figure 11 shows a comparison between L^2 -estimate of the form (4.6) and the one of the following form

$$\lim_{k \rightarrow \infty} \max_{n \in [0, T/\Delta t]} \frac{1}{N} \sum_{i=1}^N |X_n^{i,m} - X_n^i|^2, \quad (4.7)$$

in order to check their convergence rate. On the left, both the quantities in formula (4.6) and (4.7) are plotted for every inertia value in the case without memory effect. While on the right, the corresponding formulas (indicated in the caption of Figure 11) are plotted for the case with memory effect. In both cases, our numerical experiments indicate that just like the rate of convergence given in (2.53) (resp. (3.15)), the L^2 -estimate of the uniform norm of the paths of $\bar{X}^m - \bar{X}$ (resp. $(\bar{X}^m - \bar{X}, \bar{Y}_t^m - \bar{Y})$) is on the order of \sqrt{m} while the L^2 -estimate for each time $t \in [0, T]$ has order of m .

REFERENCES

- [1] Emile Aarts and Jan Korst, *Simulated annealing and boltzmann machines: A stochastic approach to combinatorial optimization and neural computing*, John Wiley & Sons, Inc., New York, NY, USA, 1989.
- [2] David Aldous, *Stopping times and tightness*, Ann. Probab. (1978), 335–340.
- [3] Luigi Ambrosio, Nicola Gigli, and Giuseppe Savaré, *Gradient flows: In metric spaces and in the space of probability measures*, Springer Science & Business Media, 2008.
- [4] Thomas Back, David B. Fogel, and Zbigniew Michalewicz (eds.), *Handbook of evolutionary computation*, 1st ed., IOP Publishing Ltd., Bristol, UK, UK, 1997.
- [5] Nicola Bellomo, Abdelghani Bellouquid, and Damian Knopoff, *From the microscale to collective crowd dynamics*, Multiscale Modeling & Simulation **11** (2013), no. 3, 943–963.
- [6] Nicola Bellomo and Christian Dogbe, *On the modeling of traffic and crowds: A survey of models, speculations, and perspectives*, SIAM review **53** (2011), no. 3, 409–463.
- [7] Patrick Billingsley, *Convergence of probability measures*, John Wiley & Sons, 1999.
- [8] Christian Blum and Andrea Roli, *Metaheuristics in combinatorial optimization: Overview and conceptual comparison*, ACM Comput. Surv. **35** (September 2003), no. 3, 268–308.
- [9] François Bolley, José A Canizo, and José A Carrillo, *Stochastic mean-field limit: non-Lipschitz forces and swarming*, Mathematical Models and Methods in Applied Sciences **21** (2011), no. 11, 2179–2210.
- [10] Vianney Bruned, André Mas, and Sylvain Wlodarczyk, *Weak convergence of particle swarm optimization*, arXiv preprint arXiv:1811.04924 (2018).
- [11] José A Carrillo and Young-Pil Choi, *Quantitative error estimates for the large friction limit of Vlasov equation with nonlocal forces*, Annales de l’institut henri poincaré, analyse non linéaire, 2020, pp. 925–954.

- [12] José A Carrillo, Young-Pil Choi, and Samir Salem, *Propagation of chaos for the Vlasov–Poisson–Fokker–Planck equation with a polynomial cut-off*, Communications in Contemporary Mathematics **21** (2019), no. 04, 1850039.
- [13] José A Carrillo, Young-Pil Choi, Claudia Totzeck, and Oliver Tse, *An analytical framework for consensus-based global optimization method*, Mathematical Models and Methods in Applied Sciences **28** (2018), no. 06, 1037–1066.
- [14] José A Carrillo, Massimo Fornasier, Jesús Rosado, and Giuseppe Toscani, *Asymptotic flocking dynamics for the kinetic Cucker–Smale model*, SIAM Journal on Mathematical Analysis **42** (2010), no. 1, 218–236.
- [15] José A Carrillo, Shi Jin, Lei Li, and Yuhua Zhu, *A consensus-based global optimization method for high dimensional machine learning problems*, ESAIM: Control, Optimisation and Calculus of Variations (2019).
- [16] Young-Pil Choi and Oliver Tse, *Quantified overdamped limit for kinetic vlasov-fokker-planck equations with singular interaction forces*, arXiv preprint arXiv:2012.00422 (2020).
- [17] Felipe Cucker and Steve Smale, *Emergent behavior in flocks*, IEEE Transactions on automatic control **52** (2007), no. 5, 852–862.
- [18] Giuseppe Da Prato and Jerzy Zabczyk, *Stochastic equations in infinite dimensions*, Cambridge university press, 2014.
- [19] Amir Dembo and Ofer Zeitouni, *Large deviations techniques and applications*, Springer-Verlag Berlin Heidelberg, 2010.
- [20] Manh Hong Duong, Agnes Lamacz, Mark A Peletier, and Upanshu Sharma, *Variational approach to coarse-graining of generalized gradient flows*, Calculus of variations and partial differential equations **56** (2017), no. 4, 1–65.
- [21] RC Fetecau and Weiran Sun, *First-order aggregation models and zero inertia limits*, Journal of Differential Equations **259** (2015), no. 11, 6774–6802.
- [22] Massimo Fornasier, Hui Huang, Lorenzo Pareschi, and Philippe Sünnen, *Consensus-based optimization on hypersurfaces: Well-posedness and mean-field limit*, Mathematical Models and Methods in Applied Sciences **30** (2020), no. 14, 2725–2751.
- [23] Massimo Fornasier, Hui Huang, Lorenzo Pareschi, and Philippe Sünnen, *Anisotropic diffusion in consensus-based optimization on the sphere*, arXiv preprint arXiv:2104.00420 (2021).
- [24] Massimo Fornasier, Hui Huang, Lorenzo Pareschi, and Philippe Sünnen, *Consensus-based optimization on the sphere: Convergence to global minimizers and machine learning*, Journal of Machine Learning Research (2021).
- [25] Massimo Fornasier, Timo Klock, and Konstantin Riedl, *Consensus-based optimization methods converge globally in mean-field law*, arXiv preprint arXiv:2103.15130 (2021).
- [26] Michel Gendreau and Jean-Yves Potvin, *Handbook of metaheuristics*, 2nd ed., Springer Publishing Company, Incorporated, 2010.
- [27] Sara Grassi and Lorenzo Pareschi, *From particle swarm optimization to consensus based optimization: stochastic modeling and mean-field limit*, Mathematical Models and Methods in Applied Sciences (To appear).
- [28] Seung-Yeal Ha, Shi Jin, and Doheon Kim, *Convergence of a first-order consensus-based global optimization algorithm*, Mathematical Models and Methods in Applied Sciences **30** (2020), no. 12, 2417–2444.
- [29] Seung-Yeal Ha and Eitan Tadmor, *From particle to kinetic and hydrodynamic descriptions of flocking*, Kinetic & Related Models **1** (2008), no. 3, 415.
- [30] Darryl D Holm and Vakhtang Putkaradze, *Formation of clumps and patches in self-aggregation of finite-size particles*, Physica D: Nonlinear Phenomena **220** (2006), no. 2, 183–196.
- [31] Hui Huang, *A note on the mean-field limit for the particle swarm optimization*, Applied Mathematics Letters (2021), 107133.
- [32] Hui Huang, Jian-Guo Liu, and Peter Pickl, *On the mean-field limit for the Vlasov–Poisson–Fokker–Planck system*, Journal of Statistical Physics **181** (2020), no. 5, 1915–1965.
- [33] Hui Huang and Jinniao Qiu, *On the mean-field limit for the consensus-based optimization*, arXiv preprint arXiv:2105.12919 (2021).
- [34] Nobuyuki Ikeda and Shinzo Watanabe, *Stochastic differential equations and diffusion processes*, Elsevier, 1989.
- [35] Pierre-Emmanuel Jabin, *Macroscopic limit of Vlasov type equations with friction*, Annales de l’institut henri poincare (c) non linear analysis, 2000, pp. 651–672.
- [36] Pierre-Emmanuel Jabin and Zhenfu Wang, *Mean field limit for stochastic particle systems*, Active particles, volume 1, 2017, pp. 379–402.
- [37] Jean Jacod and Albert Shiryaev, *Limit theorems for stochastic processes*, Vol. 288, Springer Science & Business Media, 2002.
- [38] James Kennedy, *The particle swarm: social adaptation of knowledge*, Proceedings of 1997 IEEE International Conference on Evolutionary Computation, 1997, pp. 303–308.
- [39] James Kennedy and Russell Eberhart, *Particle swarm optimization*, Proceedings of 1995 IEEE International Conference on Neural Networks, 1995, pp. 1942–1948.

- [40] Jeongho Kim, Myeongju Kang, Dohyun Kim, Seung-Yeal Ha, and Insoon Yang, *A stochastic consensus method for non-convex optimization on the Stiefel manifold*, 2020 59th IEEE conference on decision and control (cdc), 2020, pp. 1050–1057.
- [41] Hendrik Anthony Kramers, *Brownian motion in a field of force and the diffusion model of chemical reactions*, *Physica* **7** (1940), no. 4, 284–304.
- [42] Nikolaj Vladimirovič Krylov, *Controlled diffusion processes*, Vol. 14, Springer Science & Business Media, 2008.
- [43] Shih-Wei Lin, Kuo-Ching Ying, Shih-Chieh Chen, and Zne-Jung Lee, *Particle swarm optimization for parameter determination and feature selection of support vector machines*, *Expert systems with applications* **35** (2008), no. 4, 1817–1824.
- [44] Xuerong Mao, *Stochastic differential equations and applications*, Elsevier, 2007.
- [45] Peter David Miller, *Applied asymptotic analysis*, Vol. 75, American Mathematical Soc., 2006.
- [46] Sebastien Motsch and Eitan Tadmor, *Heterophilious dynamics enhances consensus*, *SIAM review* **56** (2014), no. 4, 577–621.
- [47] René Pinnau, Claudia Totzeck, Oliver Tse, and Stephan Martin, *A consensus-based model for global optimization and its mean-field limit*, *Mathematical Models and Methods in Applied Sciences* **27** (2017), no. 01, 183–204.
- [48] Riccardo Poli, *Mean and variance of the sampling distribution of particle swarm optimizers during stagnation*, *IEEE Transactions on Evolutionary Computation* **13** (2009), no. 4, 712–721.
- [49] Riccardo Poli, James Kennedy, and Tim Blackwell, *Particle swarm optimization*, *Swarm intelligence* **1** (2007), no. 1, 33–57.
- [50] Manuel Schmitt and Rolf Wanka, *Particle swarm optimization almost surely finds local optima*, *Theoretical Computer Science* **561** (2015), 57–72.
- [51] Yuhui Shi and Russell Eberhart, *A modified particle swarm optimizer*, *Proceedings of 1998 IEEE international conference on evolutionary computation*, 1998, pp. 69–73.
- [52] Alain-Sol Sznitman, *Topics in propagation of chaos*, *Ecole d’été de probabilités de Saint-Flour XIX—1989*, 1991, pp. 165–251.
- [53] Frans Van den Bergh and Andries Petrus Engelbrecht, *A convergence proof for the particle swarm optimiser*, *Fundamenta Informaticae* **105** (2010), no. 4, 341–374.
- [54] Yudong Zhang, Shuihua Wang, and Genlin Ji, *A comprehensive survey on particle swarm optimization algorithm and its applications*, *Mathematical problems in engineering* **2015** (2015).

DEPARTMENT OF MATHEMATICS, TECHNICAL UNIVERSITY MUNICH, MUNICH, GERMANY
Email address: cristina.cipriani@ma.tum.de

DEPARTMENT OF MATHEMATICS AND STATISTICS, UNIVERSITY OF CALGARY, CALGARY, CANADA
Email address: hui.huang1@ucalgary.ca

DEPARTMENT OF MATHEMATICS AND STATISTICS, UNIVERSITY OF CALGARY, CALGARY, CANADA
Email address: jinniao.qiu@ucalgary.ca