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 [25] 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 [25] 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 the weak 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 behaviour (or swarming), an area that has attracted a great deal of attention; see for instance [4, 12, 27, 40] and references therein. Such complex systems frequently appear in modeling phenomena such as biological swarms [15], crowd dynamics [5], self-assembly of nanoparticles [28], and opinion formation [40]. In the field of global optimization, similar particle models are also used in *metaheuristics* [1, 3, 7, 24], 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

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

where $\mathcal{E}(x):\ \mathbb{R}^d \to \mathbb{R}$ 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 [33, 34, 43], and it is now widely recognized as an efficient method for tackling complex optimization problems [37, 42]. 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. However the mathematical understanding of PSO is still in its infancy. Recently Grassi and

Date: April 15, 2021.

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) postdoc 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.

Pareschi [25] 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:

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

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

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

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

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

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

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

where e_k is the unit vector in the k-th dimension for $k = 1, \ldots, d$. 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 forth moment, endowed with the Wasserstein distance [2]. The choice of the weight function

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

comes from the well-known Laplace's principle [17, 39], a classical asymptotic method for integrals, which states that for any probability measure $\rho \in \mathcal{P}(\mathbb{R}^d)$, there holds

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

Thus for α large enough, one expects that

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

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

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

in the case of $d = 2$, and with the global minimizer $x^* = (0,0)^T$. In Figure 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.

As it has been shown in [25], in the zero-inertia limit ($m \to 0$), one may expect to obtain the recent developed Consensus Based Optimization (CBO) dynamics [11, 23, 26, 41] satisfying

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

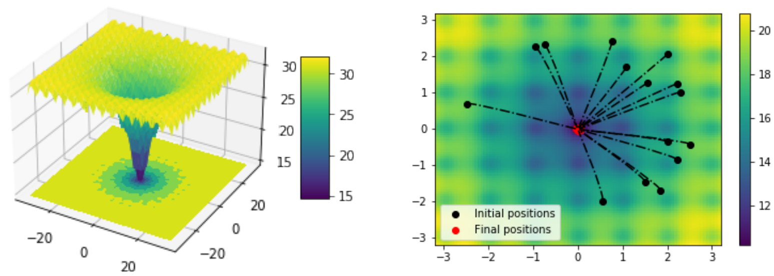

FIGURE 1. Left: the Ackley function for $d = 2$ with the unique global minimum at the point $x^* = (0,0)^T$. Right: Particles trajectories of the PSO model (1.2) along the simulation for the 2-d Ackley function with the global minimizer x^*. The simulation parameters are: time discretization 0.01, number 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.

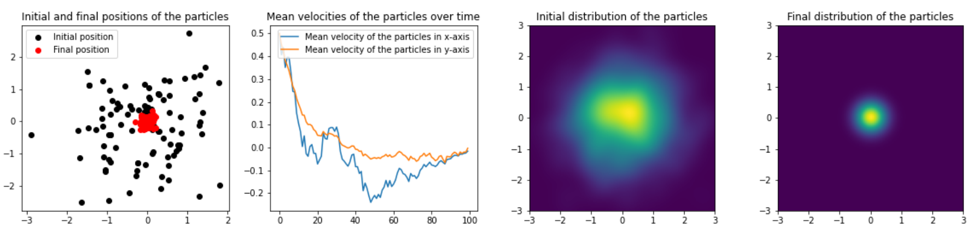

FIGURE 2. Application of the PSO dynamics (1.2) to the 2-d Ackley function $\mathcal{E}(x)$ with the global minimizer $x^* = (0,0)^T$. Particles initially have a normal distribution around x^*. Then all particles converge to one point, the global minimizer x^*, and they stop moving eventually, i.e. velocity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

where

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

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 [13]. By now, CBO methods have also been generalized to optimization over manifolds [20–22, 35]. The objective of the present paper is to complete a theory gap suggested in [25] by providing a rigorous proof of the zero-inertia limit.

On the one hand, as $N \to \infty$, the mean-field limit results (see [8, 10, 29, 30, 32, 44] for instance) indicate that our PSO dynamics (1.2) converge to the solutions of following mean-field nonlinear Mckean systems:

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

where

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

and the initial data $(\overline{X}_0, \overline{V}_0)$ is the same as in (1.2). Here $f^m(t,x,v)$ is the distribution of $(\overline{X}_t^m, \overline{V}_t^m)$ at time t , which makes the set of equations (1.7) nonlinear. We refer to [29] for a proof the well-posedness of 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

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

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

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

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

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

In this paper, we prove that in the zero-inertia limit, as $m\to 0^+$, the processes $\{\overline{X}^m\}$ satisfying SDEs (1.7) converge weakly to the solution $\overline{X}$ to SDE (1.10) in the continuous path space $\mathcal{C}([0,T];\mathbb{R}^d)$. A convergence rate is obtained and the generalizations to cases with memory effects are also addressed. This is related to the study of the overdamped limit [14, 18, 36], or large friction limit [9, 19, 31] for Vlasov type equations. However, the nolinear 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 weak convergence of the non-Markovian stochastic processes $\{\overline{X}^m\}$ satisfying SDE (1.7) to the solution $\{\overline{X}\}$ 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 $m\to 0$; see Theorem 2.3. 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 on 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 $\mathcal{E}:\mathbb{R}^d\to\mathbb{R}$ is locally Lipschitz continuous and satisfies the properties:

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

The following theorem gives the well-posedness of the mean-field PSO and CBO dynamics (1.7) and (1.10) whose proofs are analogous to [29, Theorem 2.3] and [11, Theorem 3.1], and thus omitted.

Theorem 2.1. *Let Assumption 1 hold. For each $T>0$, there hold the following assertions.*

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

$$\sup_{t\in[0,T]}\mathbb{E}\left[|\overline{X}_t^m|^4 + |\overline{V}_t^m|^4\right]\le e^{CT}\cdot\mathbb{E}\left[|\overline{X}_0|^4 + |\overline{V}_0|^4\right], \tag{2.1}$$

where C depends only on $\lambda, m, \sigma,$ and $C_{\alpha,\mathcal{E}}$.

(ii) If $\overline{X}_0$ is distributed according to $\rho_0 \in \mathcal{P}_4(\mathbb{R}^d)$, then SDE (1.10) admits a unique solution up to time T with the initial data $\overline{X}_0$ and it holds further that

$$\sup_{t\in[0,T]} \mathbb{E}\left[|\overline{X}_t|^4\right] \leq e^{CT}\cdot \mathbb{E}\left[|\overline{X}_0|^4\right] , \tag{2.2}$$

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

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

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

which implies that

$$\begin{aligned}\overline{X}_t^m = \overline{X}_0 + \int_0^t \overline{V}_\tau d\tau &= \overline{X}_0 + \int_0^t e^{-\frac{\gamma}{m}\tau}\overline{V}_0 d\tau + \frac{\lambda}{m}\int_0^t\int_0^\tau e^{-\frac{\gamma}{m}(\tau-s)}(X_s^\alpha(\rho^m) - \overline{X}_s^m)dsd\tau \\ &+ \frac{\sigma}{m}\int_0^t\int_0^\tau e^{-\frac{\gamma}{m}(\tau-s)}D(X_s^\alpha(\rho^m) - \overline{X}_s^m)dB_s d\tau\,. \end{aligned}\tag{2.3}$$

Then $\overline{X}_t^m$ has the law ρ_t^m for each $t\geq 0$. Denote by $\mathcal{C}([0,T];\mathbb{R}^d)$ the space of all $\mathbb{R}^d$-valued continuous functions on $[0,T]$ equipped with the usual uniform norm $\|\cdot\|_0$. Each continuous stochastic process $\overline{X}^m$ 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 weak convergence in the space of probability measures on $\mathcal{C}([0,T];\mathbb{R}^d)$. In what follows, we write $\overline{X}^m \rightharpoonup \overline{X}$ or $\rho^m \rightharpoonup \rho$ with ρ being the law of $\overline{X}$, if $\{\rho^m\}_{m>0}$, as a sequence of probability measures on $\mathcal{C}([0,T];\mathbb{R}^d)$, converges weakly to ρ, i.e., for each bounded continuous functional Φ on $\mathcal{C}([0,T];\mathbb{R}^d)$, there holds $\lim_{m\to 0^+}\mathbb{E}\left[\Phi(\overline{X}^m)\right] = \mathbb{E}\left[\Phi(\overline{X})\right]$. The weak convergence $\overline{X}^m \rightharpoonup \overline{X}$ is stronger than and actually implies the convergence of $\{\rho_t^m\}_{m>0}$ to ρ_t with ρ_t being the law of $\overline{X}_t$ 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 [6] 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 $\{\overline{X}^m\}_{0<m\leq 1}$ is justified by using Aldous tightness criteria.
- We will check that all the limit points of $\{\overline{X}^m\}_{0<m\leq 1}$ as $m\to 0$ satisfy mean-field CBO dynamic (1.10) which in fact admits a unique solution.

For the sake of completeness, let us recall Aldous tightness criteria.

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 on $\mathcal{C}([0,T];\mathbb{R}^d)$ if the following two conditions hold.*

($Con1$) For all $t\geq 0$, 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 discrete-valued $\sigma(X_s^n; s\in[0,T])$-stopping times β such that $0\leq \beta+\delta_0\leq T$,

$$\sup_{\delta\in[0,\delta_0]}\mathbb{P}\left(|X_{\beta+\delta}^n - X_\beta^n|\geq \eta\right)\leq \varepsilon\,. \tag{2.4}$$

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 $\lim_{k\to\infty} m_k = 0$, the sequence of probability distributions $\{\rho^{m_k}\}_{k\in\mathbb{N}}$ of $\{\overline{X}^{m_k}\}_{k\in\mathbb{N}}$ is tight.*

Proof. By Lemma 2.1, it is sufficient to justify conditions $(Con1)$ and $(Con2)$ in Aldous tightness criteria .

• *Step 1: Checking* $(Con1)$. First, for $0 < m \leq \frac{1}{2}$, recalling (2.3), we have by Fubini's theorem (see [16, Theorem 4.33] for the stochastic version)

$$
\begin{aligned}
\overline{X}_t^m &= \overline{X}_0 + \int_0^t e^{-\frac{\gamma}{m}\tau}\overline{V}_0 d\tau + \frac{\lambda}{m}\int_0^t\int_0^\tau e^{-\frac{\gamma}{m}(\tau-s)}(X_s^\alpha(\rho^m) - \overline{X}_s^m)dsd\tau \\
&\quad + \frac{\sigma}{m}\int_0^t\int_0^\tau e^{-\frac{\gamma}{m}(\tau-s)}D(X_s^\alpha(\rho^m) - \overline{X}_s^m)dB_s d\tau \\
&= \overline{X}_0 + \int_0^t e^{-\frac{\gamma}{m}\tau}\overline{V}_0 d\tau + \frac{\lambda}{m}\int_0^t\int_s^t e^{-\frac{\gamma}{m}(\tau-s)}d\tau(X_s^\alpha(\rho^m) - \overline{X}_s^m)ds \\
&\quad + \frac{\sigma}{m}\int_0^t\int_s^t e^{-\frac{\gamma}{m}(\tau-s)}d\tau D(X_s^\alpha(\rho^m) - \overline{X}_s^m)dB_s \\
&= \overline{X}_0 + \frac{m}{\gamma}(1-e^{-\frac{\gamma}{m}t})\overline{V}_0 + \frac{\lambda}{\gamma}\int_0^t(1-e^{-\frac{\gamma}{m}(t-s)})(X_s^\alpha(\rho^m) - \overline{X}_s^m)ds \\
&\quad + \frac{\sigma}{\gamma}\int_0^t(1-e^{-\frac{\gamma}{m}(t-s)})D(X_s^\alpha(\rho^m) - \overline{X}_s^m)dB_s\,.
\end{aligned}
\tag{2.5}
$$

Here 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

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

where we have used the fact that for any sequence $\{a_i\}_{i=1}^n \geq 0$ and $p \geq 2$, there holds

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

Using the moment inequality for stochastic integrals as in [38, Theorem 7.1] yields that

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

Thus,

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

Notice that

$$
\begin{aligned}
\mathbb{E}[|X_t^\alpha(\rho^m) - \overline{X}_t^m|^4] &= \int_{\mathbb{R}^d}\left|\frac{\int_{\mathbb{R}^d}x\omega_\alpha^{\mathcal{E}}(x)\rho^m(t,dx)}{\int_{\mathbb{R}^d}\omega_\alpha^{\mathcal{E}}(x)\rho^m(t,dx)} - y\right|^4\rho^m(t,dy) = \int_{\mathbb{R}^d}\left|\frac{\int_{\mathbb{R}^d}(x-y)\omega_\alpha^{\mathcal{E}}(x)\rho^m(t,dx)}{\int_{\mathbb{R}^d}\omega_\alpha^{\mathcal{E}}(x)\rho^m(t,dx)}\right|^4\rho^m(t,dy) \\
&\leq \frac{\int_{\mathbb{R}^d}\int_{\mathbb{R}^d}|x-y|^4\omega_\alpha^{\mathcal{E}}(x)\rho^m(t,dx)\rho^m(t,dy)}{\int_{\mathbb{R}^d}\omega_\alpha^{\mathcal{E}}(x)\rho^m(t,dx)} \leq 16C_{\alpha,\mathcal{E}}\mathbb{E}[|\overline{X}_t^m|^4]\,,
\end{aligned}
\tag{2.6}
$$

where Jensen's inequality is applied in the first inequality. Thus we have

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

Using Gronwall's inequality leads to

$$\mathbb{E}[|\overline{X}_t^m|^4] \leq \left(64\mathbb{E}[|\overline{X}_0|^4] + \frac{64m^4}{\gamma^4}\mathbb{E}[|\overline{V}_0|^4]\right)\exp\left(\frac{1024C_{\alpha,\mathcal{E}}(\lambda^4T^3 + 36d^3T\sigma^4)}{\gamma^4}T\right), \quad t \in [0,T]. \tag{2.7}$$

Recalling $0 \leq m \leq \frac{1}{2}$ and $\frac{1}{\gamma} = \frac{1}{1-m} \leq 2$, from estimate (2.7) we obtain the boundedness:

$$\mathbb{E}[|\overline{X}_t^m|^4] \leq C(\mathbb{E}[|\overline{X}_0|^4], \mathbb{E}[|\overline{V}_0|^4], C_{\alpha,\mathcal{E}}, \lambda, d, \sigma, T). \tag{2.8}$$

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

$$\begin{aligned}|\overline{X}_t^m|^4 = |\overline{X}_0|^4 + 4\int_0^t |\overline{X}_s^m|^2\overline{X}_s^m \cdot \overline{V}_s^m ds &\leq |\overline{X}_0^m|^4 + 8\int_0^t |\overline{X}_s^m|^2(|\overline{X}_s^m|^2 + |\overline{V}_s^m|^2)ds \\ &\leq |\overline{X}_0^m|^4 + C\int_0^t (|\overline{X}_s^m|^4 + |\overline{V}_s^m|^4)ds. \end{aligned}\tag{2.9}$$

Applying Itô-Doeblin formula to (1.7b) gives

$$\begin{aligned}|\overline{V}_t^m|^4 &= |\overline{V}_0|^4 + \frac{4\lambda}{m}\int_0^t |\overline{V}_s^m|^2\overline{V}_s^m \cdot (X_s^\alpha(\rho^m) - \overline{X}_s^m)ds + \frac{4\sigma}{m}\int_0^t |\overline{V}_s^m|^2\overline{V}_s^m \cdot D(X_s^\alpha(\rho^m) - \overline{X}_s^m)dB_s \\ &\quad + \frac{6\sigma^2}{m^2}\int_0^t |\overline{V}_s^m|^2|X_s^\alpha(\rho^m) - \overline{X}_s^m|^2 ds - \int_0^t \frac{4\gamma}{m}|\overline{V}_s^m|^4 ds \\ &\leq |\overline{V}_0|^4 + C\left(\frac{\lambda}{m} + \frac{\sigma^2}{m^2}\right)\int_0^t |\overline{V}_s^m|^4 ds + C\left(\frac{\lambda}{m} + \frac{\sigma^2}{m^2}\right)\int_0^t |X_s^\alpha(\rho^m) - \overline{X}_s^m|^4 ds \\ &\quad + \frac{4\sigma}{m}\int_0^t |\overline{V}_s^m|^2\overline{V}_s^m \cdot D(X_s^\alpha(\rho^m) - \overline{X}_s^m)dB_s. \end{aligned}\tag{2.10}$$

Collecting estimates (2.10) and (2.9) and recalling $\frac{1}{2} \leq m \leq 1$, we have

$$\begin{aligned}&\mathbb{E}[|\overline{X}_t^m|^4 + |\overline{V}_t^m|^4] \\ \leq&\mathbb{E}[|\overline{X}_0|^4 + |\overline{V}_0|^4] + C\int_0^t \mathbb{E}[|\overline{X}_s^m|^4 + |\overline{V}_s^m|^4]ds + C\int_0^t \mathbb{E}[|X_s^\alpha(\rho^m) - \overline{X}_s^m|^4]ds \\ \leq&\mathbb{E}[|\overline{X}_0|^4 + |\overline{V}_0|^4] + C(1 + 8C_{\alpha,\mathcal{E}})\int_0^t \mathbb{E}[|\overline{X}_s^m|^4 + |\overline{V}_s^m|^4]ds, \end{aligned}\tag{2.11}$$

where the estimate (2.6) is used in the last inequality. Applying Gronwall's inequality yields that

$$\mathbb{E}[|\overline{X}_t^m|^4 + |\overline{V}_t^m|^4] \leq \mathbb{E}[|\overline{X}_0|^4 + |\overline{V}_0|^4]\exp\left(C \cdot (1 + 8C_{\alpha,\mathcal{E}})t\right), \quad t \in [0,T]. \tag{2.12}$$

Finally, combining (2.8) and (2.12) yields that

$$\sup_{m\in(0,1]}\sup_{t\in[0,T]} \mathbb{E}[|\overline{X}_t^m|^4] \leq C(\mathbb{E}[|\overline{X}_0|^4], \mathbb{E}[|\overline{V}_0|^4], C_{\alpha,\mathcal{E}}, \lambda, \sigma, d, T) =: C_1 \tag{2.13}$$

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

$$\rho_t^m((K_\varepsilon)^c) = \mathbb{P}(|X_t^m|^4 > \frac{C_1}{\varepsilon}) \leq \frac{\varepsilon\mathbb{E}[|X_t^m|^4]}{C_1} \leq \varepsilon, \quad \forall\ 0 < m \leq 1. \tag{2.14}$$

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 $\sigma(X_s^m; s \in [0,T])$-stopping time with discrete values 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 (2.3) and compute

$$\begin{aligned}
&\overline{X}^m_{\beta+\delta} - \overline{X}^m_\beta \\
&= \int_\beta^{\beta+\delta} \overline{V}_\tau d\tau = \int_\beta^{\beta+\delta} e^{-\frac{\gamma}{m}\tau}\overline{V}_0 d\tau + \frac{\lambda}{m}\int_\beta^{\beta+\delta}\int_0^\tau e^{-\frac{\gamma}{m}(\tau-s)}(X_s^\alpha(\rho^m) - \overline{X}_s^m)dsd\tau \\
&\quad + \frac{\sigma}{m}\int_\beta^{\beta+\delta}\int_0^\tau e^{-\frac{\gamma}{m}(\tau-s)}D(X_s^\alpha(\rho^m) - \overline{X}_s^m)dB_s d\tau \\
&= \int_\beta^{\beta+\delta} e^{-\frac{\gamma}{m}\tau}\overline{V}_0 d\tau + \frac{\lambda}{m}\int_0^\beta\int_\beta^{\beta+\delta} e^{-\frac{\gamma}{m}(\tau-s)}d\tau(X_s^\alpha(\rho^m) - \overline{X}_s^m)ds + \frac{\lambda}{m}\int_\beta^{\beta+\delta}\int_s^{\beta+\delta} e^{-\frac{\gamma}{m}(\tau-s)}d\tau(X_s^\alpha(\rho^m) - \overline{X}_s^m)ds \\
&\quad + \frac{\sigma}{m}\int_0^\beta\int_\beta^{\beta+\delta} e^{-\frac{\gamma}{m}(\tau-s)}d\tau D(X_s^\alpha(\rho^m) - \overline{X}_s^m)dB_s + \frac{\sigma}{m}\int_\beta^{\beta+\delta}\int_s^{\beta+\delta} e^{-\frac{\gamma}{m}(\tau-s)}d\tau D(X_s^\alpha(\rho^m) - \overline{X}_s^m)dB_s \\
&= \frac{m}{\gamma}(e^{-\frac{\gamma}{m}\beta} - e^{-\frac{\gamma}{m}(\beta+\delta)})\overline{V}_0 \\
&\quad + \frac{\lambda}{\gamma}\int_0^\beta (e^{-\frac{\gamma}{m}(\beta-s)} - e^{-\frac{\gamma}{m}(\beta+\delta-s)})(X_s^\alpha(\rho^m) - \overline{X}_s^m)ds + \frac{\lambda}{\gamma}\int_\beta^{\beta+\delta}(1 - e^{-\frac{\gamma}{m}(\beta+\delta-s)})(X_s^\alpha(\rho^m) - \overline{X}_s^m)ds \\
&\quad + \frac{\sigma}{\gamma}\int_0^\beta (e^{-\frac{\gamma}{m}(\beta-s)} - e^{-\frac{\gamma}{m}(\beta+\delta-s)})D(X_s^\alpha(\rho^m) - \overline{X}_s^m)dB_s + \frac{\sigma}{\gamma}\int_\beta^{\beta+\delta}(1 - e^{-\frac{\gamma}{m}(\beta+\delta-s)})D(X_s^\alpha(\rho^m) - \overline{X}_s^m)dB_s\,.
\end{aligned} \tag{2.15}$$

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

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

and in particular,

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

Then, it is obvious that

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

Next, it follows that

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

and analogously,

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

Further, applying Itô's isometry gives

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

and analogously,

$$\mathbb{E}\left[\left|\int_\beta^{\beta+\delta} (1 - e^{-\frac{\gamma}{m}(\beta+\delta-s)})D(X_s^\alpha(\rho^m) - \overline{X}_s^m)dB_s\right|^2\right] \leq d\delta^{1/2}\left(T \sup_{s\in[0,T]} \mathbb{E}\left[|X_s^\alpha(\rho^m) - \overline{X}_s^m|^4\right]\right)^{1/2}.$$

Therefore, summing up the above estimates and recalling $0 < m \leq m_0 = \frac{1}{2}$, $\frac{1}{\gamma} \leq 2$, and the relations (2.6) and (2.13), we arrive at

$$\begin{aligned}
\mathbb{E}[|\overline{X}_{\beta+\delta}^m - \overline{X}_\beta^m|^2] &\leq \frac{5}{\gamma^2}\delta^2(\mathbb{E}[|\overline{V}_0|^4])^{\frac{1}{2}} + \frac{10}{\gamma^2} \cdot \left(\lambda^2 \delta T + \sigma^2 d\,(\delta T)^{1/2}\right) \sup_{s\in[0,T]} \left(\mathbb{E}\left[|X_s^\alpha(\rho^m) - \overline{X}_s^m|^4\right]\right)^{1/2} \\
&\leq C\left(\mathbb{E}[|\overline{X}_0|^4], \mathbb{E}[|\overline{V}_0|^4], C_{\alpha,\mathcal{E}}, \lambda, \sigma, d, T\right)\left(\delta^{\frac{1}{2}} + \delta + \delta^2\right).
\end{aligned}$$

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

$$\sup_{\delta\in[0,\delta_0]} \mathbb{P}(|\overline{X}_{\beta+\delta}^m - \overline{X}_\beta^m|^2 \geq \eta) \leq \sup_{\delta\in[0,\delta_0]} \frac{\mathbb{E}[|\overline{X}_{\beta+\delta}^m - \overline{X}_\beta^m|^2]}{\eta} \leq \varepsilon\,. \tag{2.16}$$

This justifies condition $Con2$ in Lemma 2.1. □

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

$$|X^\alpha(\rho) - X^\alpha(\widehat{\rho})| \leq CW_2(\rho, \widehat{\rho})\,, \tag{2.17}$$

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

Theorem 2.3 (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 $m \to 0^+$, 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:*

$$\overline{X}_t = \overline{X}_0 + \lambda \int_0^t (X_s^\alpha(\rho) - \overline{X}_s)ds + \sigma \int_0^t D(X_s^\alpha(\rho) - \overline{X}_s)dB_s\,. \tag{2.18}$$

Moreover it holds that

$$\sup_{t\in[0,T]} \mathbb{E}[|\overline{X}_t^m - \overline{X}_t|^2] \leq C\,m\,, \tag{2.19}$$

where the constant C depends only 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. It follows from the definition of Wasserstein distance that

$$\sup_{t\in[0,T]} W_2^2(\rho_t^m,\rho_t) \le \sup_{t\in[0,T]} \mathbb{E}[|\overline{X}_t^m - \overline{X}_t|^2] \le C\, m\,, \tag{2.20}$$

which in a way is consistent with the result obtained in [14, 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 weak convergence of $\overline{X}^m \rightharpoonup \overline{X}$ is in the path space $\mathcal{C}([0,T];\mathbb{R}^d)$, which implies and is obviously stronger than the convergence of $\{\rho_t^m\}_{m>0}$ to ρ_t for each time $t\ge 0$.

Proof. By Theorem 2.2, each subsequence $\{\overline{X}^{m_k}\}_{k\in\mathbb{N}}$ with $m_0 \le 1/2$ and m_k converging to 0 as $k\to\infty$ admits a subsequence (denoted w.l.o.g. by itself) that converges weakly. By Skorokhod's lemma (see [6, Theorem 6.7 on page 70]), we may find a common probability space $(\Omega,\mathcal{F},\mathbb{P})$ on which the processes $\{\overline{X}^{m_k}\}_{k\in\mathbb{N}}$ converge to some process $\widehat{X}$ as random variables valued in $\mathcal{C}([0,T];\mathbb{R}^d)$ almost surely. In particular, we have

$$\mathbb{P}\left(\lim_{k\to\infty} |\overline{X}_t^{m_k} - \widehat{X}_t| = 0\right) = 1,\ \forall\, t\in[0,T]. \tag{2.21}$$

We shall verify that the limit $\widehat{X}$ is indeed the unique solution $\overline{X}$ to SDE (2.18).

Recall the SDE satisfied by $\overline{X}^{m_k}$ in (2.5)

$$\begin{aligned}\overline{X}_t^{m_k} &= \overline{X}_0 + \frac{m_k}{\gamma}(1-e^{-\frac{\gamma}{m_k}t})\overline{V}_0 + \frac{\lambda}{\gamma}\int_0^t (1-e^{-\frac{\gamma}{m_k}(t-s)})(X_s^\alpha(\rho^{m_k}) - \overline{X}_s^{m_k})ds \\ &\quad + \frac{\sigma}{\gamma}\int_0^t (1-e^{-\frac{\gamma}{m_k}(t-s)})D(X_s^\alpha(\rho^{m_k}) - \overline{X}_s^{m_k})dB_s\,. \end{aligned}\tag{2.22}$$

By the estimates in (2.13) and (ii) of Theorem 2.1, there exists a constant C_2 being independent of m_k such that

$$\sup_{k\in\mathbb{N}}\sup_{t\in[0,T]} \mathbb{E}\left[|\overline{X}_t^{m_k}|^4\right] + \sup_{t\in[0,T]}\mathbb{E}\left[\left|\overline{X}_t\right|^4\right] \le C_2 := C(\mathbb{E}[|\overline{X}_0|^4],\mathbb{E}[|\overline{V}_0|^4], C_{\alpha,\mathcal{E}},\lambda,\sigma,d,T) < \infty. \tag{2.23}$$

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

$$|X_t^\alpha(\rho)| = \left|\frac{\int_{\mathbb{R}^d} x\omega_\alpha^{\mathcal{E}}(x)\rho(t,dx)}{\int_{\mathbb{R}^d}\omega_\alpha^{\mathcal{E}}(x)\rho(t,dx)}\right| \le C_{\alpha,\mathcal{E}}\int_{\mathbb{R}^d}|x|\rho(t,dx) \le C_{\alpha,\mathcal{E}}(\mathbb{E}[|\overline{X}_t|^4])^{\frac14},$$

and

$$\sup_{k\in\mathbb{N}}\sup_{t\in[0,T]} |X_t^\alpha(\rho^{m_k})| \le C_{\alpha,\mathcal{E}}(C_2)^{\frac14}, \quad \text{and} \quad \sup_{t\in[0,T]}|X_t^\alpha(\rho)| \le C_{\alpha,\mathcal{E}}(C_2)^{\frac14}\,. \tag{2.24}$$

Then we compare the SDEs (2.18) and (2.22) term by term. By Lemma 2.2, we have

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

and thus,

$$
\begin{aligned}
&\mathbb{E}\left[\left|\frac{\lambda}{\gamma}\int_0^t (1-e^{-\frac{\gamma}{m_k}(t-s)})(X_s^\alpha(\rho_s^{m_k})-\overline{X}_s^{m_k})ds-\lambda\int_0^t (X_s^\alpha(\rho_s)-\overline{X}_s)ds\right|^2\right]\\
&\le 2\mathbb{E}\left[\left|\frac{\lambda}{1-m_k}\int_0^t (1-e^{-\frac{1-m_k}{m_k}(t-s)})(X_s^\alpha(\rho_s^{m_k})-X_s^\alpha(\rho_s)+\overline{X}_s-\overline{X}_s^{m_k})ds\right|^2\right]\\
&\quad+2\mathbb{E}\left[\left|\lambda\int_0^t\left(\frac{1-e^{-\frac{1-m_k}{m_k}(t-s)}}{1-m_k}-1\right)(X_s^\alpha(\rho_s)-\overline{X}_s)ds\right|^2\right]\\
&\le C\mathbb{E}\left[\int_0^t\left|\overline{X}_s-\overline{X}_s^{m_k}\right|^2ds\right]+C\lambda^2\int_0^t\left|\frac{1-e^{-\frac{1-m_k}{m_k}(t-s)}}{1-m_k}-1\right|^2ds\cdot\mathbb{E}\left[\int_0^T\left|X_s^\alpha(\rho_s)-\overline{X}_s\right|^2ds\right]\\
&\le C\mathbb{E}\left[\int_0^t\left|\overline{X}_s-\overline{X}_s^{m_k}\right|^2ds\right]+C\int_0^t\left|\frac{1-e^{-\frac{1-m_k}{m_k}(t-s)}-(1-m_k)}{1-m_k}\right|^2ds\\
&\le C\mathbb{E}\left[\int_0^t\left|\overline{X}_s-\overline{X}_s^{m_k}\right|^2ds\right]+C\int_0^t\left(|m_k|^2+e^{-\frac{2(1-m_k)}{m_k}(t-s)}\right)ds\\
&\le C\mathbb{E}\left[\int_0^t\left|\overline{X}_s-\overline{X}_s^{m_k}\right|^2ds\right]+C\left(t\,|m_k|^2+\frac{m_k}{2(1-m_k)}\right),
\end{aligned}\tag{2.25}
$$

where the constants Cs are independent of k. For the stochastic integrals, it holds analogously that

$$
\begin{aligned}
&\mathbb{E}\left[\left|\frac{\sigma}{\gamma}\int_0^t (1-e^{-\frac{\gamma}{m_k}(t-s)})D(X_s^\alpha(\rho_s^{m_k})-\overline{X}_s^{m_k})dB_s-\sigma\int_0^t D(X_s^\alpha(\rho_s)-\overline{X}_s)dB_s\right|^2\right]\\
&\le d\sigma^2\sum_{n=1}^d\mathbb{E}\left[\left|\frac{1}{\gamma}\int_0^t (1-e^{-\frac{\gamma}{m_k}(t-s)})(X_s^\alpha(\rho_s^{m_k})-\overline{X}_s^{m_k})_n dB_s^n e_n-\int_0^t (X_s^\alpha(\rho_s)-\overline{X}_s)_n dB_s^n e_n\right|^2\right]\\
&= d\sigma^2\sum_{n=1}^d\mathbb{E}\left[\int_0^t\left|\frac{1-e^{-\frac{\gamma}{m_k}(t-s)}}{\gamma}(X_s^\alpha(\rho_s^{m_k})-\overline{X}_s^{m_k})_n-(X_s^\alpha(\rho_s)-\overline{X}_s)_n\right|^2ds\right]\\
&\le 2d\sigma^2\sum_{n=1}^d\mathbb{E}\left[\int_0^t\left|\frac{1-e^{-\frac{\gamma}{m_k}(t-s)}}{\gamma}\Big((X_s^\alpha(\rho_s^{m_k})-\overline{X}_s^{m_k})_n-(X_s^\alpha(\rho_s)-\overline{X}_s)_n\Big)\right|^2ds\right]\\
&\quad+2d\sigma^2\sum_{n=1}^d\mathbb{E}\left[\int_0^t\left|\left(\frac{1-e^{-\frac{\gamma}{m_k}(t-s)}}{\gamma}-1\right)(X_s^\alpha(\rho_s)-\overline{X}_s)_n\right|^2ds\right]\\
&\le C\mathbb{E}\left[\int_0^t\left|\overline{X}_s^{m_k}-\overline{X}_s\right|^2ds\right]+2d\sigma^2\sup_{s\in[0,t]}\mathbb{E}\left[\left|(X_s^\alpha(\rho_s)-\overline{X}_s)\right|^2\right]\cdot\int_0^t\left|\left(\frac{1-e^{-\frac{\gamma}{m_k}(t-s)}}{\gamma}-1\right)\right|^2ds\\
&\le C\mathbb{E}\left[\int_0^t\left|\overline{X}_s-\overline{X}_s^{m_k}\right|^2ds\right]+C\left(t\,|m_k|^2+\frac{m_k}{2(1-m_k)}\right).
\end{aligned}\tag{2.26}
$$

In addition, it is obvious that

$$
\left|\frac{m_k}{\gamma}(1-e^{-\frac{\gamma}{m_k}t})\overline{V}_0\right|\le Cm_k\left|\overline{V}_0\right|. \tag{2.27}
$$

Therefore, recalling $m_k \le \frac{1}{2}$, combining the estimates (2.25)-(2.27) and subtracting both sides of SDEs (2.18) from those of (2.22), we have

$$\mathbb{E}[|\overline{X}_t^{m_k} - \overline{X}_t|^2] \le C \int_0^t \mathbb{E}[|\overline{X}_s^{m_k} - \overline{X}_s|^2] ds + C\, m_k, \quad t \in [0, T].$$

By Gronwall's inequality it implies that

$$\sup_{t\in[0,T]} \mathbb{E}[|\overline{X}_t^{m_k} - \overline{X}_t|^2] \le C m_k \to 0, \text{ as } k \to \infty\,, \tag{2.28}$$

where C depends only on $\mathbb{E}[|\overline{X}_0|^4], \mathbb{E}[|\overline{V}_0|^4], C_{\alpha,\mathcal{E}}, \lambda, \sigma, d,$ and T. In view of both the convergences (2.21) and (2.28), we must have $\widehat{X} = \overline{X}$. Finally, due to the arbitrariness of the subsequence $\{\overline{X}^{m_k}\}_{k\in\mathbb{N}}$ and the uniqueness of $\overline{X}$, we conclude that as $m \to 0^+$, the sequence of stochastic processes $\{\overline{X}^m\}_{0<m\le 1}$ converge weakly to the unique solution $\overline{X}$ to SDE (2.18), with the estimate (2.19) following in the same way as (2.28). □

Remark 2.2. When proving the convergence of $\{\overline{X}^m\}$ satisfying SDEs (1.7) to the solution $\overline{X}$ of (1.10), we cannot expect the convergence of the associated velocity processes $\{\overline{V}^m\}$ due to the indifferentiability of the limit $\{\overline{X}_t\}_{t\ge 0}$ with respect to time t if $\sigma \ne 0$. Therefore, we do not investigate convergence of the joint Markovian process $\{(\overline{X}^m, \overline{V}^m)\}$ and consider instead solely the process $\{\overline{X}^m\}$ which satisfies a stochastic integral equation (2.22) 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 $\overline{X}^{m_k}$ and $\overline{X}$ in the above proof.

3. Generalization to the case with memory effects

In [25], the authors considered a PSO model which involves the memory of the local best positions, and it is of the form

$$d\overline{X}_t^m = \overline{V}_t^m dt, \tag{3.1}$$

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

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

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

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

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 hyperbolic tangent $S^\beta(x,y) = \tanh(\beta(\mathcal{E}(x) - \mathcal{E}(y))$. The corresponding mean-field PSO equation is

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

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

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

and its corresponding partial differential equation is

$$\begin{aligned}&\partial_t\rho_t+\nabla_x\cdot(\lambda_1(y-x)+\lambda_2(Y_t^\alpha(\overline{\rho})-x)\rho_t)+\nabla_y\cdot\Big(\nu(x-y)S^\beta(x,y))\rho_t\Big)\\ =&\frac{1}{2}\sum_{j=1}^d\frac{\partial^2}{\partial x_j^2}\left(\rho_t(\sigma_1^2(x-y)_j^2+\sigma_2^2(x-Y_t^\alpha(\overline{\rho}))_j^2)\right)\,,\end{aligned}\tag{3.6}$$

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

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

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

Here $\overline{\rho}^m(t,y)=\int_{\mathbb{R}^d}\rho^m(t,dx,y)$. By Fubini's theorem, similar arguments as in (2.5) yield that

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

Following the same computations as in Theorem 2.2 gives

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

where C depends only on $\lambda_1,\sigma_2,\lambda_2,\sigma_2,d$, and T. Put $\widetilde{\rho}^m(t,x)=\int_{\mathbb{R}^d}\rho^m(t,x,dy)$. In a similar way to (2.6) we have

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

Thus it yields that

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

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

Recall that

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

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

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

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

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

By Gronwall's inequality it yields that

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

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

• *Step 2: Checking* $(Con2)$. Let β be a $\sigma(X_s^m;s\in[0,T])$-stopping time with discrete values such that $\beta+\delta_0\le T$. Set $m_0=\frac{1}{2}$ w.l.o.g.. Then for all $0<m\le m_0$, one has $\frac{1}{2}\le\gamma<1$. Similar to (2.15), one has

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

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

$$\mathbb{E}[|\overline{X}_{\beta+\delta}^m-\overline{X}_\beta^m|^2]\le C\left(\delta^{\frac{1}{2}}+\delta+\delta^2\right)\,, \tag{3.11}$$

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

Having a look at

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

we have

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

By estimate (3.9), we have

$$\mathbb{E}[|\overline{Y}_{\beta+\delta}^m-\overline{Y}_\beta^m|^2]\le\nu^2\delta\int_0^T\mathbb{E}[|\overline{X}_s^m-\overline{Y}_s^m|^4]^{\frac{1}{2}}dt\le C\delta\,, \tag{3.12}$$

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

Let us recall

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

and

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

Then following the lines of the proof in Theorem 3.2, one can easily obtain

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

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

Moreover it holds that

$$\sup_{t\in[0,T]} \mathbb{E}\left[\left|\overline{X}_t^m - \overline{X}_t\right|^2 + \left|\overline{Y}_t^m - \overline{Y}_t\right|^2\right] \le C\, m\,, \tag{3.15}$$

where the constant C *depends only on* $\mathbb{E}[|\overline{X}_0|^4 + |\overline{Y}_0|^4 + |\overline{V}_0|^4], \lambda_1, \sigma_2, \lambda_2, \sigma_2, d, \beta, T, C_{\alpha,\mathcal{E}},$ *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 [25].

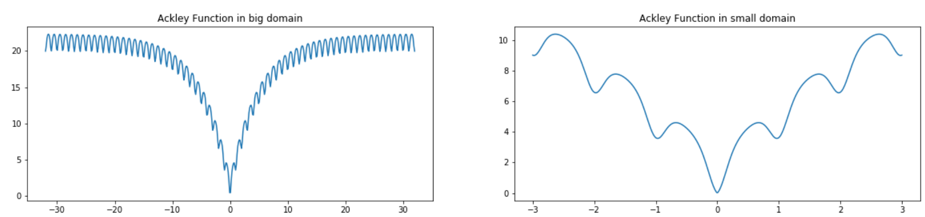

FIGURE 3. Ackley function in a big (left) and small (right) domain with its many local minima.

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

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

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

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

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 values. Then, we compare the evolution of the distribution of particles with the one of the particles moving according to 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 $i = 1, ..., N$ 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 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

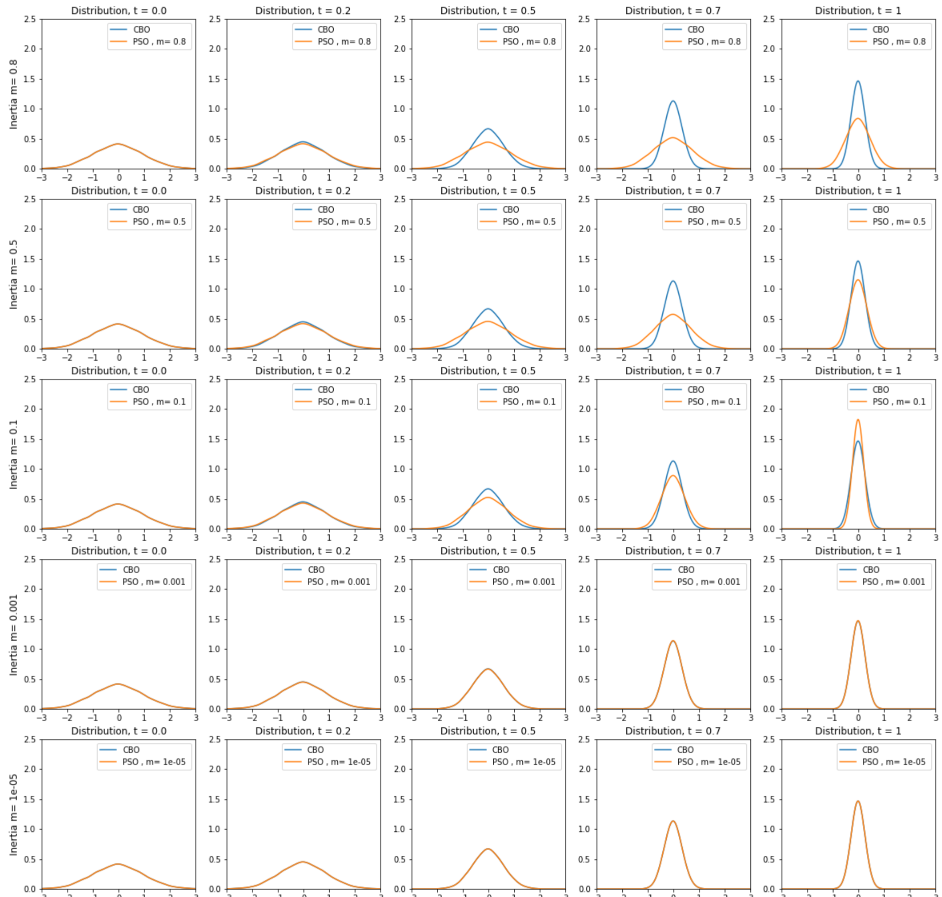

FIGURE 4. Comparison of the CBO (4.2) and PSO (4.1) dynamics for different inertia values (which are changing over the rows) and at many time steps (changing over columns), starting from a normal distribution.

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.

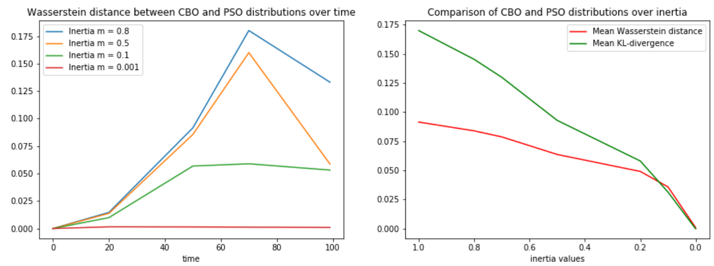

FIGURE 5. Left: Wasserstein distance over time ; Right: Wasserstein distance and Kullback-Leibler divergence (in mean) over the inertia values.

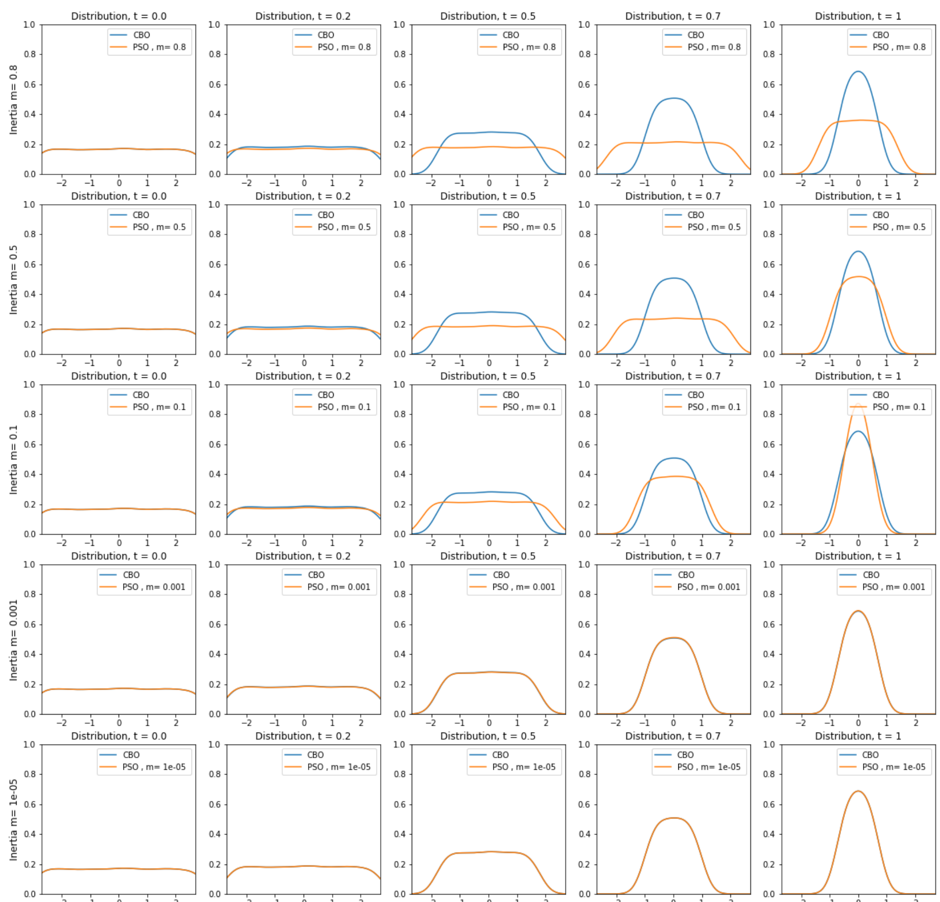

FIGURE 6. Evolution of an initial uniform distribution according to CBO (4.2) and PSO (4.1) dynamics and their comparison for different time steps (on the columns) and different inertia values (on the rows).

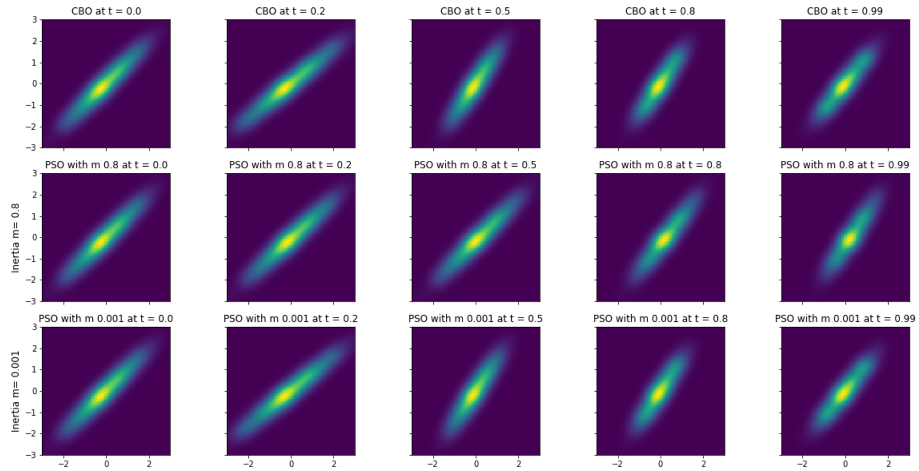

FIGURE 7. First Row: evolution in time of the initial gaussian distribution according to CBO dynamics (4.4); Second Row: evolution in time of the initial Gaussian distribution according to PSO dynamics (4.3) with $m = 0.8$; Third Row: evolution produced by the PSO dynamics (4.3) with $m = 0.001$.

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

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

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

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

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

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

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 $\forall i = 1, .., N$ and $\forall n \in [0, T/\Delta t]$. The time discretization and number of 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

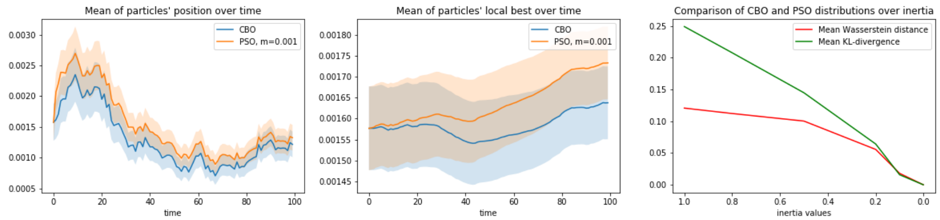

Figure 8. Left: mean particles' position with their standard deviation; Center: mean particles' local best with the standard deviation too; Right: plot of the Wasserstein distance and Kullback–Leibler divergence between the distribution obteined via CBO dynamics (4.4) and the one obtained through PSO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

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 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.

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] Luigi Ambrosio, Nicola Gigli, and Giuseppe Savaré, *Gradient flows: In metric spaces and in the space of probability measures*, Springer Science & Business Media, 2008.

[3] Thomas Back, David B. Fogel, and Zbigniew Michalewicz (eds.), *Handbook of evolutionary computation*, 1st ed., IOP Publishing Ltd., Bristol, UK, UK, 1997.

[4] Nicola Bellomo, Abdelghani Bellouquid, and Damian Knopoff, *From the microscale to collective crowd dynamics*, Multiscale Modeling & Simulation **11** (2013), no. 3, 943–963.

[5] 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.

[6] Patrick Billingsley, *Convergence of probability measures*, John Wiley & Sons, 1999.

[7] Christian Blum and Andrea Roli, *Metaheuristics in combinatorial optimization: Overview and conceptual comparison*, ACM Comput. Surv. **35** (September 2003), no. 3, 268–308.

[8] 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.

[9] 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 c, analyse non linéaire, 2020, pp. 925–954.

[10] 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.

[11] 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.

[12] 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.

[13] 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).

[14] 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).

[15] Felipe Cucker and Steve Smale, *Emergent behavior in flocks*, IEEE Transactions on automatic control **52** (2007), no. 5, 852–862.

[16] Giuseppe Da Prato and Jerzy Zabczyk, *Stochastic equations in infinite dimensions*, Cambridge university press, 2014.

[17] Amir Dembo and Ofer Zeitouni, *Large deviations techniques and applications*, Springer-Verlag Berlin Heidelberg, 2010.

[18] 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.

[19] RC Fetecau and Weiran Sun, *First-order aggregation models and zero inertia limits*, Journal of Differential Equations **259** (2015), no. 11, 6774–6802.

[20] 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.

[21] Massimo Fornasier, Hui Huang, Lorenzo Pareschi, and Philippe Sünnen, *Consensus-based optimization on the sphere II: Convergence to global mininizers and machine learning*, arXiv:2001.11988v3 (2020).

[22] Massimo Fornasier, Hui Huang, Lorenzo Pareschi, and Philippe Sünnen, *Anisotropic diffusion in consensus-based optimization on the sphere*, arXiv preprint arXiv:2104.00420 (2021).

[23] Massimo Fornasier, Timo Klock, and Konstantin Riedl, *Consensus-based optimization methods converge globally in mean-field law*, arXiv preprint arXiv:2103.15130 (2021).

[24] Michel Gendreau and Jean-Yves Potvin, *Handbook of metaheuristics*, 2nd ed., Springer Publishing Company, Incorporated, 2010.

[25] 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).

[26] 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.

[27] Seung-Yeal Ha and Eitan Tadmor, *From particle to kinetic and hydrodynamic descriptions of flocking*, Kinetic & Related Models **1** (2008), no. 3, 415.

[28] 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.

[29] Hui Huang, *A note on the mean-field limit for the particle swarm optimization*, Applied Mathematics Letters (2021), 107133.

[30] 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.

[31] 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.

[32] Pierre-Emmanuel Jabin and Zhenfu Wang, *Mean field limit for stochastic particle systems*, Active particles, volume 1, 2017, pp. 379–402.

[33] James Kennedy, *The particle swarm: social adaptation of knowledge*, Proceedings of 1997 IEEE International Conference on Evolutionary Computation, 1997, pp. 303–308.

[34] James Kennedy and Russell Eberhart, *Particle swarm optimization*, Proceedings of 1995 IEEE International Conference on Neural Networks, 1995, pp. 1942–1948.

[35] Jeongho Kim, Myeongju Kang, Dohyun Kim, Seung-Yeal Ha, and Insoon Yang, *A stochastic consensus method for nonconvex optimization on the Stiefel manifold*, 2020 59th ieee conference on decision and control (cdc), 2020, pp. 1050–1057.

[36] Hendrik Anthony Kramers, *Brownian motion in a field of force and the diffusion model of chemical reactions*, Physica **7** (1940), no. 4, 284–304.

[37] 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.

[38] Xuerong Mao, *Stochastic differential equations and applications*, Elsevier, 2007.

[39] Peter David Miller, *Applied asymptotic analysis*, Vol. 75, American Mathematical Soc., 2006.

[40] Sebastien Motsch and Eitan Tadmor, *Heterophilious dynamics enhances consensus*, SIAM review **56** (2014), no. 4, 577–621.

[41] 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.

[42] Riccardo Poli, James Kennedy, and Tim Blackwell, *Particle swarm optimization*, Swarm intelligence **1** (2007), no. 1, 33–57.

[43] Yuhui Shi and Russell Eberhart, *A modified particle swarm optimizer*, Proceedings of 1998 IEEE international conference on evolutionary computation, 1998, pp. 69–73.

[44] Alain-Sol Sznitman, *Topics in propagation of chaos*, Ecole d'été de probabilités de Saint-Flour XIX—1989, 1991, pp. 165–251.

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